

Institut für Pharmazeutische Biologie Bonn

**Bioinformatische und *in-vitro*-Analyse
der *frs*-Biosynthesegencluster von
Cand. Burkholderia crenata und
*Chromobacterium vaccinii***

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Isabella Schamari

aus

Bonn

Bonn 2018

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

Erstgutachter: Frau Professor Dr. rer. nat. Gabriele König

Zweitgutachter: Herr Professor Dr. rer. nat. Werner Knöss

Tag der Promotion: 15.02.2019

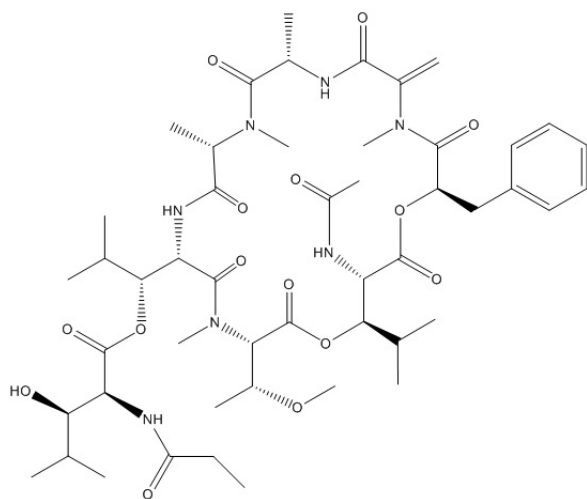
Erscheinungsjahr: 2019

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die zum Erfolg dieser Arbeit mit beigetragen haben, besonders bei Frau Professor Gabriele König, die mir das spannende Thema zur Bearbeitung überlassen hat. Mein Dank gilt auch Herrn Dr. Max Crüsemann, der durch seine detaillierte Recherche in Datenbanken Teilsequenzen des *cvfrs*-BGC gefunden hat und dadurch einen großen Beitrag dazu geliefert hat, dass die Arbeit in dieser Form überhaupt möglich war. An dieser Stelle danke ich auch Dr. René Richarz, der in mühevoller Kleinarbeit die vorhandenen Sequenzlücken im *cvfrs*-BGC geschlossen hat, was ebenfalls grundlegend zum Erfolg dieser Arbeit beigetragen hat. Natürlich möchte ich mich auch bei unserem Arbeitskreis bedanken, und insbesondere den Kolleginnen Kollegen, die mit in dieses Projekt involviert waren: Cornelia Hermes, Daniel Wirtz, Tatjana Reuter, Steffen Hild. Ein großer Dank gilt auch unserer Technischen Assistentin Emilie Goralski. Vielen Dank für die Starthilfe in unserem Arbeitskreis und in die Einführung der Methoden. Auch dem Rest des Arbeitskreises ein herzliches Dankeschön, für die anregenden Diskussionen und die vielen Ideen und Hilfestellungen. Ich danke auch der DFG für die Finanzierung unserer Forschergruppe und der Forschergruppe selber für den regen Austausch an Wissen. Vielen Dank an Marc Sylvester und Bernd Gehrig für die Hilfe bei den MALDI-TOF-MS-Messungen.

Zusammenfassung

FR900359 (FR) ist ein selektiver G_q -Proteininhibitor, der ursprünglich aus der Pflanze *Ardisia crenata* isoliert wurde. G-Proteine spielen eine zentrale Rolle in der Signalübertragung von G-Protein gekoppelten Rezeptoren (GPCR), die an vielen physiologischen Prozessen beteiligt sind. G-Proteine sind im Vergleich zu GPCRs relativ gering erforscht, da es nur wenig membran-gängige G-Protein-Modulatoren gibt. FR kann aufgrund seiner Eigenschaften diese Lücke schließen.



Das Depsipeptid FR besteht aus acht größtenteils ungewöhnlichen Aminosäuren, darunter drei Hydroxyleucine mit unterschiedlichen Seitenketten, N-Methyldehydroalanin, N,O-Dimethylthreonin und der Carbonsäure Phenyllaktat. Ein N-Methylalanin und Alanin sind auch Teil des Peptid-Zyklus.

FR ist ein zyklisches Depsipeptid, bestehend aus 8 Einheiten, von denen 7 Aminosäuren sind. Die meisten davon sind nicht-proteinogenen Ursprungs. Einen weiteren ungewöhnlichen Baustein bildet die Carbonsäure Phenyllaktat, die über eine Esterbindung in den Heterozyklus eingebaut ist. Das Biosynthesegencluster (BGC) von FR konnte in zwei unterschiedlichen Wirtsor-

ganismen lokalisiert werden, in *Candidatus Burkholderia crenata* und in *Chromobacterium vaccinii*. Obwohl *Cand. B. crenata* im Gegensatz zu *C. vaccinii* ein Endosymbiont ist, sind beide BGC strukturell nahezu identisch.

In dieser Arbeit wurden die beiden Nichtribosomalen Peptidsynthetase BGC aus *Cand. B. crenata* und *C. vaccinii* identifiziert und bioinformatisch betrachtet. Hierzu wurden die BGC vor allem auf Gemeinsamkeiten und Unterschiede hin untersucht. Des Weiteren wurden die bioinformatischen Grundlagen geschaffen, um einen neuartigen Dehydratisierungsmechanismus durch Kondensierungsdomänen (C-Domänen) zu untersuchen. Ebenfalls konnte für die letzte Adenylierungsdomäne (A-Domäne) aus FrsG (A8) eine für A-Domänen bis jetzt unbekannt Struktur postuliert werden, die sowohl zwei Methyltransferasen (MT-Domänen) als auch zwei Subdomänen enthält. Die A-Domänen aus FrsA (A1) und FrsD (A2) wurden anhand von *in-vitro*-Studien untersucht, um die Biosynthese der nicht-proteinogenen Aminosäure Hydroxyleucin während der FR-Biosynthese zu belegen.

Im ersten Teil der Arbeit konnte durch einen Vergleich der BGC auf DNA- und Protein-Ebene die Identität der BGC untereinander bestimmt werden. Hierbei wurde festgestellt, dass die BGC nicht nur untereinander sehr ähnlich sind, sondern sich auch innerhalb des jeweiligen BGC wiederholende Bereiche mit ausgesprochen hoher Identität befinden. Diese sind bei beiden BGC ähnlich groß und ihre Lokalisation ist übereinstimmend.

Eine weitere Besonderheit der *frs*-BGC ist die A-Domäne A8 mit zwei aufeinander folgenden MT-Transferasen. Durch die Analyse der Core-Bereiche der A-Domänen nach Stachelhaus konnte die ungewöhnliche Struktur der A-Domäne A8 aus FrsG genauer bestimmt werden, die nicht nur zwei MT-Domänen, sondern auch zwei A-Subdomänen enthält.

FR enthält die ungewöhnliche Aminosäure N-Methyldehydroalanin. Die *frs*-BGC kodieren aber für kein Enzym, das offensichtlich die Dehydratisierung von Serin zu Dehydroalanin katalysieren könnte. Durch Alignments und an-

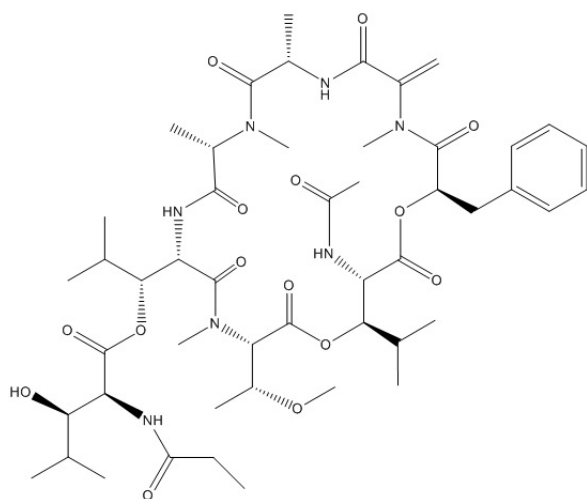
schließende Erstellung eines 3D-Modells der C-Domänen C6 und C8 konnte ein Hinweis auf einen neuen putativen Mechanismus für eine Dehydratisierung von Serin zu Dehydroalanin gefunden werden.

Im zweiten Teil der Arbeit wurden die beiden A-Domänen von FrsA und FrsD, und damit die Hydroxyleucin-Biosynthese mithilfe des gut etablierten A-Domänen-Assay nach Phelan et al. 2009 charakterisiert. Hierzu wurden diese A-Domänen heterolog exprimiert und in einem massenspektroskopisch basierten Assay auf ihre Spezifität getestet. Dabei konnte eine eindeutige Prävalenz beider A-Domänen für Leucin gezeigt werden. Diese Erkenntnis präzisiert die bioinformatische Analyse und zeigt, dass die Hydroxygruppe enzymatisch nach dem Beladen des Peptidyl Carrier Protein (PCP) mit Leucin eingefügt wird.

Insgesamt konnten durch diese Arbeit zahlreiche Einblicke und Grundlagen in die Biosynthese von FR erhalten werden, die bei der Modifizierung dieses Wirkstoffes helfen könnten. Neue Strukturvarianten von FR durch Bioengineering können dazu beitragen, die Mechanismen der Funktion von G-Proteinen zu entschlüsseln.

Abstract

FR900359 (FR) is a selective G_q-Protein-inhibitor which is originally isolated from the plant *Ardisia crenata*. G-Proteins play an important role in the signalling pathways of G-Protein coupled receptors (GPCR) which are involved in many physiological processes. Compared to GPCRs, there is relatively little known about G-Proteins because there are not so many G-Protein inhibitors that are able to pass membranes. FR can close this gap. In this work, the biosynthetic gene clusters (BGCs) of FR from *Candidatus* B. crenata and *Chromobacterium vaccinii* are bioinformatically analyzed and A domains are partially expressed to investigate their specificity.



The depsipeptide FR consists of eight mostly unusual amino acids, including three hydroxyleucines with different side chains: N-methyl-dehydroalanine, N,O-dimethyl-threonine and the carbonic acid phenyllactate. N-methyl-alanine and alanine are also part of the peptide.

FR is a cyclic depsipeptide containing 8 building blocks, 7 of which are of non-proteinogenic origin. One of them is the carbonic acid phenyllactate. The BGCs of FR are located in two hosts *Cand.* B. crenata and *C. vaccinii*. Even

though *Cand. B. crenata* is an endosymbiont, in contrast to *C. vaccinii*, the BGCs are almost identical.

In the first part of this work, both BGCs were examined for similarities and differences in their overall and detailed structure. Comparing DNA- and protein level, the identity of both BGCs had been determined. It could be shown that both BGC contain repeating sequences with very high identity in the respective BGC. Those are similar in both BGCs respecting their size and location.

Additionally, a novel structure of the A domain A8, which contains two methyltransferases as well as two A subdomains could be predicted. The identity of the BGCs with regard to each other was examined by comparisons on the DNA- and protein-level. Hereby it was shown that both BGC have the same architecture. It could also be shown that within one BGC, repeating sequences can be found which occur in the same way in both clusters.

Both *frs*-BGCs contain the A domain A8 with two methyltransferases which is unique for the BGCs of FR-analogues. By analysis of the core motifs of all A domains [Stachelhaus et al., 1999], the unusual structure of the A domain A8 from FrsG could be elucidated in detail. This A domain contains not only two methyltransferases but also two A subdomains.

FR contains the unusual amino acid N-methyl-dehydroalanine, but there is no obvious enzyme that is able to catalyze the dehydration from serine to dehydroalanine. By alignments and subsequent establishment of a 3D-model from the two very similar C domains C6 and C8, new hints on a putative dehydration mechanism from serine to dehydroalanine could be found.

In the second part of this work, *in vitro* experiments for the two A domains from FrsA and FrsD were performed in order to examine their substrate specificity. Therefore, these A domains were heterologously expressed and evaluated for their specificity by a mass spectrometry based assay. For both

A domain's clear preference for leucine was shown which specified the bioinformatic analysis. Consequently, the hydroxygroup is incorporated into leucine after having been activated and binded to the Peptidyl Carrier Protein (PCP). Through this work we gained insight into the biosynthesis of FR, which could help to modify FR by bioengineering, and to learn more about the mode of action and function of G-Proteins.

Inhaltsverzeichnis

1	Einleitung	5
1.1	Naturstoffe und deren Bedeutung für die Pharmazie	5
1.2	Das Depsipeptid FR900359	7
1.2.1	Quellen für FR900359	11
1.2.2	Die Biosyntheseegencluster <i>bcfrs</i> und <i>cvfrs</i> von FR . . .	14
1.3	Nicht ribosomale Peptidsynthetasen (NRPS)	16
1.4	Adenylierungsdomänen (A-Domänen)	17
1.5	Methyltransferase-Domänen (MT-Domänen)	22
1.6	Kondensierungsdomänen (C-Domänen)	24
1.7	Peptidyl Carrier Protein oder PCP-Domänen	29
1.8	MbtH-ähnliche Proteine	30
1.9	Thioesterasen (TE-Domänen)	33
1.10	Vorstufengerichtete Biosynthese von NRPs und Bioengineering von NRPS	33
2	Zielsetzung der Arbeit	39
3	Ergebnisse	40
3.1	Bioinformatische Analyse der Biosyntheseegencluster <i>bcfrs</i> und <i>cvfrs</i> von FR	40
3.1.1	Analyse der <i>bcfrs</i> - und <i>cvfrs</i> -Biosyntheseegencluster aus <i>Cand. B. crenata</i> und <i>C.vaccinii</i> mit antiSMASH . . .	40
3.1.2	Vergleichende Analyse der Gene des Biosyntheseegen- clusters aus <i>Cand. B. crenata</i> und <i>C.vaccinii</i>	46
3.1.3	Bioinformatische Analyse sich wiederholender Sequenz- bereiche in den BGC von <i>Cand. B. crenata</i> und <i>C.vaccinii</i>	49
3.2	Bioinformatische Analyse der in den <i>frs</i> -BGC von <i>Cand. B.</i> <i>crenata</i> und <i>C.vaccinii</i> kodierten A-Domänen	54
3.3	Bioinformatische Analyse der in den <i>frs</i> -BGC von <i>Cand. B.</i> <i>crenata</i> und <i>C.vaccinii</i> kodierten MT-Domänen	60
3.4	Bioinformatische Analyse der in den <i>frs</i> -BGC von <i>Cand. B.</i> <i>crenata</i> und <i>C.vaccinii</i> kodierten Kondensierungs-Domänen . .	62

3.4.1	Bioinformatische Analyse der PCP-Domänen und TE-Domänen	70
3.5	<i>In-vitro</i> -Untersuchungen von A-Domänen	73
3.5.1	Grundlagen des γ - $^{18}\text{O}_4$ -ATP-Austausch-Assay nach Phelan et al. für die A-Domänen von FrsA und FrsD	73
3.5.2	Klonierung und Expression der A-Domänen FrsA und FrsD	77
3.5.3	SDS-PAGE-Analytik der A-Domänen von FrsA und FrsD	79
3.5.4	Durchführung des γ - $^{18}\text{O}_4$ -ATP-Assay	81
3.5.5	Detektion und Auswertung des γ - $^{18}\text{O}_4$ -ATP-Austauschs mittels MALDI-TOF-MS	82
4	Diskussion	91
4.1	Pharmazeutische Relevanz von FR	91
4.2	Hypothese zur Biosynthese von FR	91
4.3	Vergleich der <i>frs</i> -BGC auf DNA- und Protein-Ebene	93
4.4	Sich wiederholende Sequenzabschnitte in beiden <i>frs</i> -BGC	93
4.5	Strukturelle Besonderheiten der A-Domäne A8 aus FrsG	96
4.6	Bioinformatische Betrachtungen zum Dehydratisierungsmechanismus der C-Domäne C6	98
4.7	Spezifität der A-Domänen aus FrsA und FrsD	99
4.8	Ausblick	100
5	Material und Methoden	102
5.1	Verwendete Datenbanken und Programme	102
5.2	Vektoren und Organismen	103
5.2.1	Vektoren	103
5.2.2	Organismen	104
5.3	Medien und Puffer	106
5.3.1	Medien	106
5.3.2	Proteinaufreinigungspuffer	107
5.3.3	Plasmidpräparationspuffer	108
5.3.4	Assaypuffer	109

5.3.5	Anoden- und Kathodenpuffer für die SDS-Analyse . . .	109
5.3.6	Weitere Puffer und Lösungen	110
5.3.7	Antibiotikastocks	111
5.4	Mikrobiologische Techniken	111
5.4.1	Stammhaltung in Kryokulturen	111
5.4.2	Kultivierung von Bakterien	111
5.4.3	Konzentrationsbestimmung von Bakterienkulturen . . .	112
5.4.4	Transformation von Bakterien	112
5.4.5	Blau-Weiß-Selektion	113
5.5	Molekularbiologische Methoden	114
5.5.1	Polymerase-Kettenreaktion (PCR)	114
5.5.2	Kolonie-PCR	116
5.5.3	Verwendete Primer	116
5.5.4	Reinigen von PCR-Produkten oder Aufkonzentrieren von DNA	118
5.5.5	DNA-Präparation	118
5.5.6	Agarose Gelelektrophorese	118
5.5.7	DNA-Isolierung aus Agarosegelen	119
5.5.8	Sequenzierung	119
5.5.9	Restriktionsspaltung von DNA	119
5.5.10	Dephosphorylierung von Vektoren	120
5.5.11	Ligation	120
5.5.12	Vektorkonstrukte	121
5.6	Proteinexpression, Proteinreinigung und Proteinanalytik . . .	124
5.6.1	Proteinexpression	124
5.6.2	Zellaufschluss	124
5.6.3	Aufreinigung mit His-Taq	124
5.6.4	Entsalzen und Aufkonzentrierung	125
5.6.5	Probenvorbereitung für die SDS-PAGE-Analyse	125
5.6.6	SDS-Polyacrylamidgelelektrophorese	125
5.6.7	Proteinkonzentrationsbestimmungen	127
5.6.8	A-Domänen-Assay	127
5.6.9	Massenspektrometrische Untersuchung von Proteinen .	128

5.7	Verwendete Chemikalien	128
5.8	Verwendete Enzyme	129
5.9	Verwendete Kits und Standards	130
5.10	Verwendete Geräte	131
	Abkürzungsverzeichnis	132
	Abbildungsverzeichnis	135
	Tabellenverzeichnis	144
	Literaturverzeichnis	148
	A Anhang	167

1 Einleitung

1.1 Naturstoffe und deren Bedeutung für die Pharmazie

Als Naturstoffe werden Metabolite bezeichnet, die von Organismen gebildet werden. Das bedeutet, dass sie tierischen, pflanzlichen oder mikrobiellen Ursprungs sind [Gordon et al., 1997]. Naturstoffe haben seit längerem eine heilkundliche Bedeutung für den Menschen. Zum einen, weil es lange Zeit keine synthetischen Wirkstoffe gab und zum anderen, weil viele dieser Stoffe von Pflanzen oder Bakterien für die eigene Abwehr gegen Schädlinge gebildet wurden und sich im Laufe der Zeit perfekt auf ihr „Target“ spezialisiert haben. In vielen Fällen führt eine kleine Modifikation des Naturstoffes sogar zum Wirkverlust. Auch heute ist ein Großteil der neu auf den Markt kommenden Medikamente Naturstoffe oder deren Derivate, darunter Krebsmedikamente, Antiinfektiva, Antidiabetika und andere [Harvey, 2008]. In den Jahren 1981-2014 machte der Anteil der zugelassenen unmodifizierten Naturstoffe an allen zugelassenen Wirkstoffen einen Anteil von 4% aus [Newman and Cragg, 2016].

Für die Produzenten haben Naturstoffe in erster Linie eine ökologische Bedeutung und werden als Abwehrstoffe produziert. Hier spielen auch Symbiosen eine große Rolle [Luo et al., 2014], d.h. viele Spezies produzieren diese Abwehrstoffe nicht selbst, sondern beherbergen Bakterien, die für sie diese Aufgabe übernehmen. Als Beispiel seien die im Wasser lebenden Ascidiaceae [Schmidt et al., 2012] oder Teredinidae [Elshahawi et al., 2013] genannt. Auch bei Pflanzen sind Symbiosen bekannt, wie im Falle von *Psychotria kirkii* und deren Kirkamid produzierendem Endosymbionten *Cand. Burkholderia kirkii* [Sieber et al., 2015].

Auch das Depsipeptid FR900359 (FR), Gegenstand der vorliegenden Arbeit (siehe Abb. 1), wird unter anderem von einem pflanzlichen Endosymbionten, *Cand. B. crenata*, gebildet [Carlier et al., 2016]. FR ist ein selektiver G_{q/11/14}-Protein-Inhibitor [Schrage et al., 2015]. Die in vielen Organismen vorhandene

Signalübertragung über G-Proteine [Link, 2016] führte zu der Überlegung, dass FR genau wie Kirkamid möglicherweise eine ökologische Funktion für die Pflanze *Ardisia crenata* (*A. crenata*) besitzt (siehe auch Kap. 1.2.1).

Peptidische Naturstoffe können von nichtribosomalen Peptidsynthetasen (NRPS) gebildet werden. Diese großen Multienzymkomplexe bilden vorwiegend kleine, oft zyklische Peptide, die anders als die ribosomal gebildeten Peptide nicht nur proteinogene, sondern auch sehr ungewöhnliche Aminosäuren bis hin zu Carbonsäuren in die Peptide integrieren (siehe Kap. 1.3 und folgende). Über molekularbiologische und bioinformatische Methoden konnten gute Einblicke in die Funktion von NRPS-Systemen gewonnen werden. Durch den modularen Aufbau von NRPS, auf den in Kapitel 1.10 weiter eingegangen wird, erhofft man sich neue Möglichkeiten, Wirkstoffe zu entwickeln, indem man diese Systeme genetisch verändert, um neue Naturstoffe zu produzieren (siehe auch Kap. 1.10).

1.2 Das Depsipeptid FR900359

Das zyklische Depsipeptid FR900359 wurde ursprünglich 1988 aus der tropischen Pflanze *A. crenata* gewonnen [Fujioka M., 1988]. FR besteht aus zehn Einheiten (siehe Abb. 1), wovon sieben Aminosäuren und drei Carbonsäuren sind. Zu den Aminosäuren zählen sechs nicht proteinogene Bausteine, wie die drei Hydroxyleucine, das N-Methylalanin, das N-Methyldehydroalanin und das N,O-Dimethylthreonin. Letztere zwei fallen hier besonders auf. Das sehr seltene N-Methyldehydroalanin ist in dieser Form nur für die Microcystine [Tillett et al., 2000], Sameuramid [Machida et al., 2018] und YM-254890 (YM) bekannt. Das N,O-Dimethylthreonin kommt bis jetzt nur im strukturell sehr ähnlichen YM-254890 (siehe Abb. 2) [Taniguchi et al., 2003] und im Sameuramid [Machida et al., 2018] vor. Des Weiteren sind drei Carbonsäuren enthalten: Phenyllaktat, Essigsäure und Propionsäure, die letzteren beiden an N-OH-Leucin gebunden. YM unterscheidet sich von FR lediglich in einem Aminosäuregrundbaustein und in den zwei Acyl-Seitenketten. Anstelle eines der N-acetyl-Hydroxyleucine ist hier ein N-Acetyl-Threonin in den Peptidzyklus eingebaut. Des Weiteren enthält YM einen N-Acetyl-Substituenten an der Hydroxyleucin-Seitenkette, wo FR eine N-Propionyleinheit besitzt (siehe Abb. 1 und 2).

Des Weiteren enthält FR drei Hydroxyleucine und zwei Alanine, von denen eines N-methyliert ist. Auch die Art der Bindungen im Peptidring ist teilweise ungewöhnlich. So sind das Phenyllaktat und das N,O-Dimethylthreonin über eine Esterbindung im Ring verknüpft, so dass zwei Esterbindungen benachbart sind. Außerdem gibt es cis-konfigurierte Peptidbindungen [Crüseemann et al., 2018].

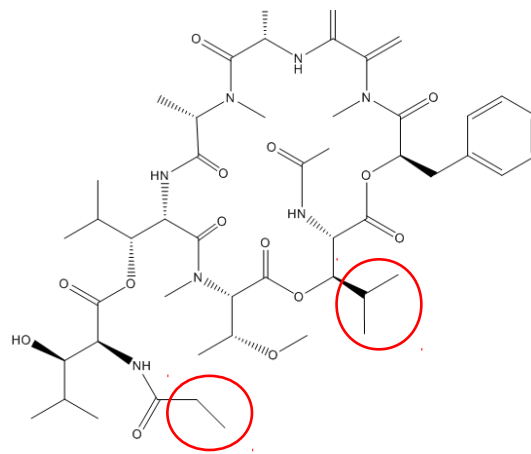


Abbildung 1: Das Depsipeptid FR besteht aus acht größtenteils ungewöhnlichen Aminosäuren, darunter drei Hydroxyleucine mit unterschiedlichen Seitenketten, N-Methyldehydroalanin, N,O-Dimethylthreonin und die Carbonsäure Phenyllaktat. Ein N-Methylalanin und Alanin sind auch Teil des Peptid-Zyklus. Die rot markierten Stellen zeigen die Unterschiede von YM und FR auf.

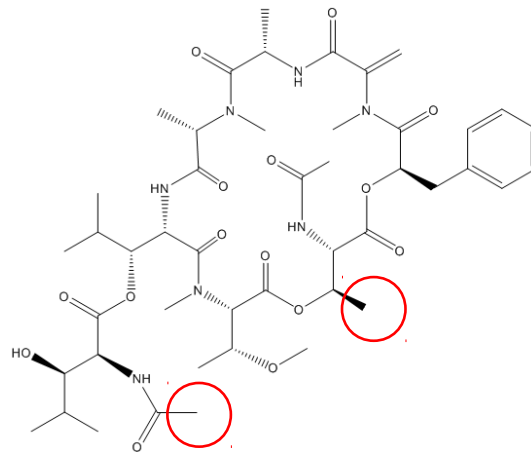


Abbildung 2: Struktur des mit FR verwandten YM-254890. YM unterscheidet sich lediglich dadurch, dass es eine Acetyl- anstatt einer Propionylseitenkette, und eine Methyl- anstatt einer Isopropylgruppe enthält. Die rot markierten Stellen zeigen die Unterschiede von YM und FR auf.

Aufgrund seiner ungewöhnlichen strukturellen Eigenschaften deutet vieles auf einen nicht-ribosomalen Ursprung von FR hin, was die Vermutung nahe

legt, dass es kein Metabolit von *A. crenata*, sondern bakteriellen Ursprungs ist (siehe auch Kap. 1.2.1). *A. crenata* enthält den Endosymbionten *Cand. Burkholderia crenata*, der vor allem in den Blatträndern in hoher Konzentration zu finden ist [Crüsemann et al., 2018]. Carlier et al. konnten das Genom von *Cand. Burkholderia crenata* sequenzieren und stellten fest, dass dieses evolutionär bedingt stark reduziert ist. Das Genom besteht aus einem 2,67 Mb großen Chromosom und zwei extrachromosomalen Plasmiden, eines davon 103 kb und das andere 78 kb groß. Auf dem größeren Plasmid konnte ein BGC für FR, d.h. das *bcfrs*-BGC gefunden werden [Carlier et al., 2016] [Crüsemann et al., 2018]. Die Zugehörigkeit und Vollständigkeit des *bcfrs*-BGC wurde durch heterologe Expression des BGC in *E. coli* bestätigt.

FR inhibiert selektiv heterotrimere Guaninnukleotid bindende Proteine (G-Proteine) der G_q -Familien (G_q , G_{11} , G_{14}) [Schrage et al., 2015]. G-Proteine sind Signalpeptide, die durch membranständige, G-Protein gekoppelte Rezeptoren (GPCRs) aktiviert werden. G-Proteine befinden sich auf der inneren Seite der Zellmembran und lösen von dort aus die Signale aus, die an der Außenseite der Zelle an den GPCRs ankommen [Lefkowitz, 2004] [Rosenbaum et al., 2009]. Ein Drittel aller Arzneistoffe setzt an GPCRs an [Overington et al., 2006] [Rask-Andersen et al., 2011]. GPCRs spielen unter anderem bei Asthma, allergischen Reaktionen, Krebs, Bluthochdruck oder metabolischen Prozessen eine zentrale Rolle [Deshpande and Penn, 2006] [Druey, 2009] [Dorsam and Gutkind, 2007] [Takefuji et al., 2012] [Blad et al., 2012] [Klepac, 2016]. Oft sind diese Signalwege sehr komplex. Es ist meist mehr als ein Rezeptor an der Krankheit beteiligt, was es schwierig macht, Signalwege über GPCRs zu modulieren. Trotz der Vielzahl an GPCR gibt es nur wenige G-Proteine, die die Signale der GPCRs im Zellinneren in einen Effekt übersetzen [Gudermann et al., 1996]. Daher ist es wichtig die Beteiligung der G-Proteine am Signalgeschehen besser zu verstehen, um noch gezielter Einfluss auf diese Signalwege nehmen zu können. Bis jetzt sind G-Proteine aufgrund mangelnder pharmakologischer Werkzeuge kaum erforscht. Einer der wenigen zuverlässigen G-Proteininhibitor ist Pertussistoxin (PTX) (z.B. [Hughes et al., 1984] [Sauliere et al., 2012]), welches aber nur für Forschungszwecke

zur Verfügung steht und keinerlei therapeutische Relevanz besitzt. Aufgrund der Membrangängigkeit von FR und seiner Eigenschaft, G_q-Proteine zu inhibieren, ist es gut geeignet, die komplexen Signalprozesse von G-Proteinen zu untersuchen [Crüsemann et al., 2018]. In dieser Hinsicht ist FR einzigartig, da es, abgesehen von seinem Struktur analogon YM, die einzige Substanz ist, die selektiv als hochpotenter Wirkstoff G_q-Proteine inhibieren kann [Nishimura et al., 2010].

1.2.1 Quellen für FR900359

Candidatus Burkholderia crenata

Candidatus Burkholderia crenata ist ein Endosymbiont aus der Klasse der β -*Proteobacteria*. Er lebt in Symbiose mit *A. crenata*. Alle Versuche, das Bakterium ohne Wirt zu kultivieren, scheiterten bis jetzt. Der Endosymbiont wird vermutlich vertikal auf die nächste Generation weitergegeben [Pinto-Carbo et al., 2018]. Ein weiteres Beispiel dieser Symbioseart findet man bei *Psychotria kirkii* und deren Endosymbiont *Cand. Burkholderia kirkii* [Sieber et al., 2015]. Carlier et al. gelang es 2015, das Genom von *Cand. B. crenata* zu sequenzieren, welches gegenwärtig als das kleinste *Burkholderia*-Genom gilt. Das Genom ist stark reduziert, was auf die evolutionäre Entwicklung als Endosymbiont zurückzuführen ist. Einige Gene mit möglicher Funktion für Sekundärmetabolite konnten allerdings eindeutig dem Genom von *Cand. B. crenata* zugeordnet werden, darunter das putative BGC von FR auf einem 103 kb großen Plasmid. Dieses BGC ist bis jetzt nur für diese *Burkholderia* Spezies beschrieben worden. Des Weiteren wurde ein 19 kb großes PKS-Operon auf einem 78 kb großen Plasmid gefunden, das für zwei putative Typ I *trans*-AT Polyketidsynthasen codiert [Carlier et al., 2016].

Ardisia crenata

A. crenata ist eine asiatische Pflanze aus der Familie der Primulaceae, die weit verbreitet als Schmuckpflanze Verwendung findet. An den Blatträndern befinden sich kleine Blatteinkerbungen, und die Pflanze bildet rote Früchte aus (siehe Abb. 3).



Abbildung 3: *A. crenata* hat Einkerbungen an den Blatträndern, in denen *Cand. B. crenata* als Endosymbiont eingebettet ist. Die Pflanze wird weit verbreitet als Schmuckpflanze verwendet und hat eine lange Tradition in der chinesischen Medizin.

Phytochemisch wurde die Pflanze 1986 von Fujioka und Kollegen untersucht. Aus dem methanolischen Extrakt von *A. crenata* wurde FR isoliert und Strukturaufklärung mittels $^1\text{H-NMR}$ und massenspektrometrischer Studien vorgenommen. Sie konnten ebenfalls schon zeigen, dass der methanolische Extrakt hemmende Wirkung auf die Blutplättchenaggregation hat und den Blutdruck senkt [Fujioka M., 1988]. Im Genus *Ardisia* gibt es bis jetzt dreißig Arten, für die eine Endosymbiose mit Bakterien belegt werden konnte, darunter auch *A. crenata*, die den Endosymbionten *Cand. B. crenata* in ihren Blatteinkerbungen enthält. Dort sind die Bakterien in einer Matrix aus Schleimstoffen eingebettet [Miller, 1990] [Lersten and Horner, 1976]. Die Weitergabe der Symbionten erfolgt vermutlich vertikal [Pinto-Carbo et al., 2018].

Chromobacterium vaccinii

Chromobacterium vaccinii (*C. vaccinii*) ist ein gramnegatives Bakterium, das ebenfalls aus der Klasse der β -Proteobakterien stammt. Das Bakterium ist freilebend und wurde von den Wurzeln von *Vaccinium macrocarpon* (Cranberry) isoliert [Soby et al., 2013]. Das Genom von *C. vaccinii* wurde teilweise sequenziert [Voing et al., 2015]. Die Sequenz enthielt Bruchstücke von NRPS-Genen, die denen von *Cand. B. crenata* ähneln. Die teilweise vorliegende Sequenz des FR-BGC wurde im Rahmen dieses Projekts mittels PCR-Lückenschluss vervollständigt. Aus der Fermentationskultur von *C. vaccinii* konnte ebenfalls FR isoliert werden.

1.2.2 Die Biosyntheseegencluster *bcfrs* und *cvfrs* von FR

Sowohl *Cand. B. crenata* als auch *C. vaccinii* sind in der Lage, das Depsipeptid FR zu bilden [Crüsemann et al., 2018]. Trotz ihrer sehr unterschiedlichen Habitate besitzen sie zwei strukturell identische BGC. Die BGC (siehe Abb. 4) von FR bestehen aus acht Genen. Diese bilden zwei NRPS-Systeme. Die erste NRPS ist durch das erste Gen *frsA* kodiert und enthält eine Starter-Kondensierungsdomäne (C-Domäne), eine Adenylierungsdomäne (A-Domäne), ein Peptidyl Carrier Protein (PCP) und eine Thioesterasedomäne (TE-Domäne). Zwischen beiden NRPS liegt ein Gen für ein MLP (MbtH like protein), gefolgt von *frsC*, einem Gen, das für eine Malatdehydrogenase kodiert. Es folgen vier weitere Gene mit insgesamt sieben weiteren NRPS-Modulen *frsD-frsG*. Insgesamt sind acht C-Domänen, acht A-Domänen, acht PCP-Domänen, vier Methyltransferasen (MT-Domänen), zwei TE-Domänen und eine Epimerisierungsdomäne (E-Domäne) vorhanden. Nach der zweiten NRPS folgt die β -Hydroxylase FrsH. Beide BGC haben zwar die gleiche Struktur, unterscheiden sich aber auf der DNA- und Proteinebene. Das BGC von *Cand. B. crenata* wird im Folgenden als *bcfrs* und das von *C. vaccinii* als *cvfrs* bezeichnet. Ein detaillierter Vergleich beider BGC erfolgt im Kap. 3.1 Bioinformatische Analyse der BGC *bcfrs* und *cvfrs* von FR.

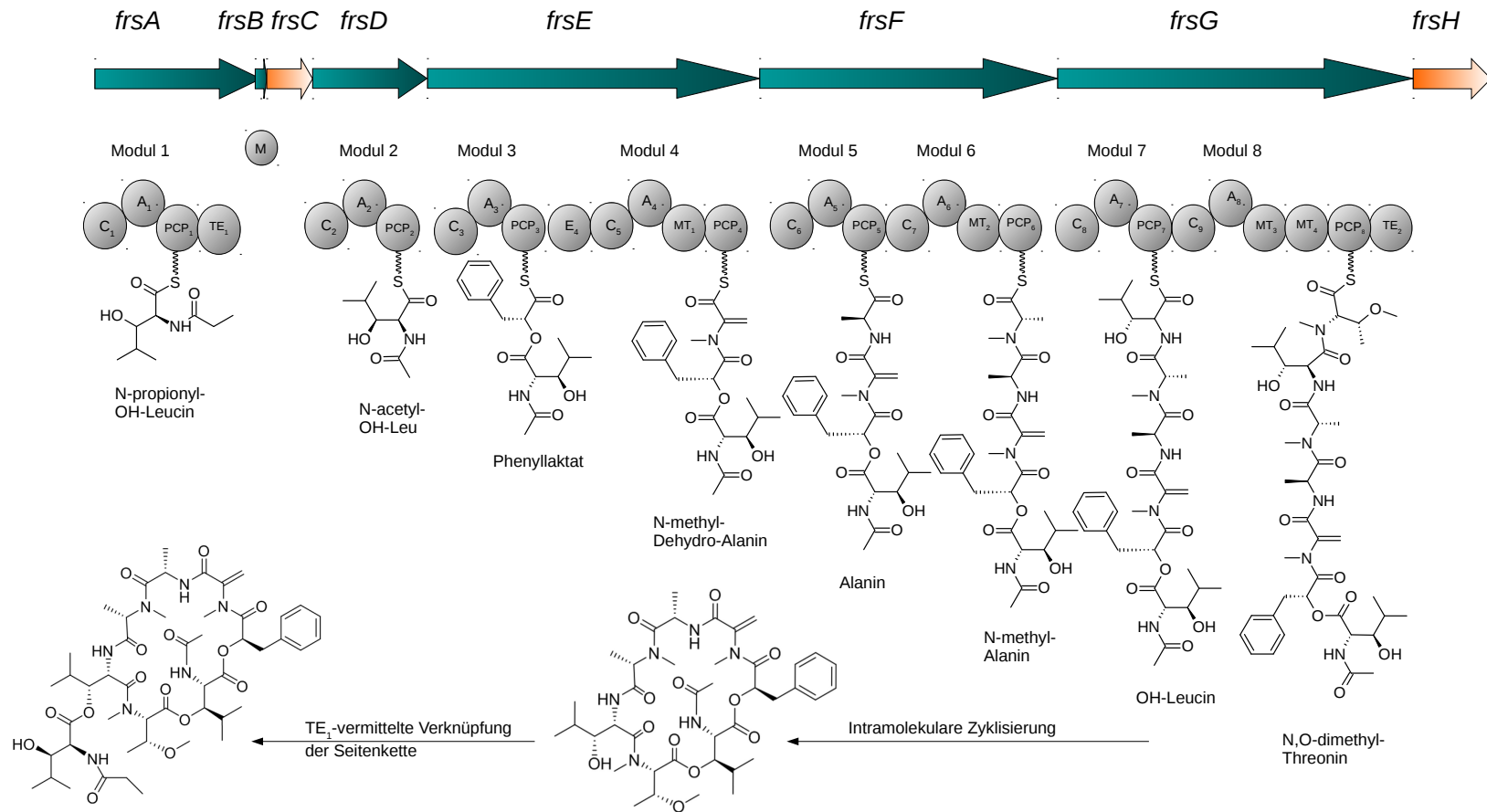


Abbildung 4: Hypothese zur Biosynthese von FR. Nach der intramolekularen Zyklisierung durch die endständige TE-Domäne aus FrsG, kommt es zur Übertragung der Hydroxyleucinseitenkette durch die TE-Domäne aus FrsA.

1.3 Nicht ribosomale Peptidsynthetasen (NRPS)

NRPS sind Multienzymkomplexe, die hauptsächlich in Mikroorganismen vorzufinden sind. An ihnen findet die Proteinbiosynthese kleiner Peptide statt, die oft zyklischer Natur sind und auch nicht-proteinogene Aminosäuren oder andere ungewöhnliche Bestandteile enthalten [Marahiel, 2009].

Eine NRPS besteht aus mehreren Domänen, die in Module zusammengefasst sind. Jedes Modul ist für die Aktivierung einer Aminosäure und deren Verknüpfung mit der wachsenden Peptidkette zuständig. Dieses Prinzip nennt sich Colinearitätsregel [Süßmuth, 2017]. Für ein Peptid wie FR, das aus acht Bausteinen besteht, braucht man also acht Module, d.h. die Anordnung der Module innerhalb der NRPS bestimmt die Primärstruktur des Peptids.

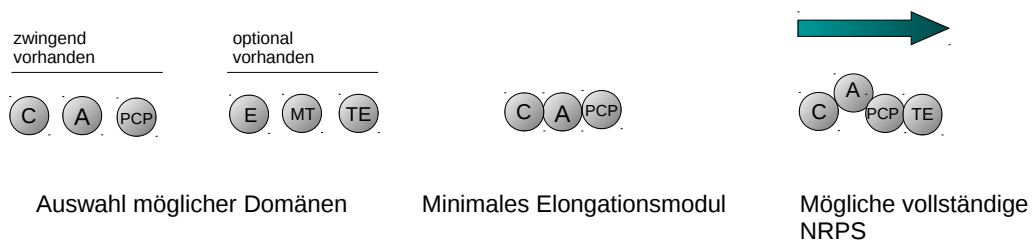


Abbildung 5: Aufbau einer NRPS: Sie besteht aus aneinandergesetzten Modulen, welche wiederum aus Domänen aufgebaut sind. Ein funktionsfähiges Modul benötigt zwingend eine A-, eine PCP- und eine C-Domäne. Optional können modifizierende Domänen wie MT- oder E-Domänen vorhanden sein. Oft werden NRPS mit einer TE-Domäne abgeschlossen. Die Abfolge der Domänen bestimmt die Struktur des Peptids.

Die Domänen sind die kleinsten funktionellen Einheiten der NRPS. Jede Domäne hat ihre eigene Funktion und zeigt auch isoliert vom Komplex Enzymaktivität. Für ein minimales Elongationsmodul sind mindestens drei Domänen nötig (siehe Abb. 5). Dieses besteht aus einer C-Domäne, einer A-Domäne, und einer PCP-Domäne [Marahiel et al., 1997]. Die Reihenfolge dieser Module bestimmt die Struktur des Peptids. Ein Startermodul besteht in der Regel aus zwei Domänen und fängt oft mit einer A-Domäne an, die die

erste Aminosäure aktiviert. Ein Elongationsmodul fängt mit einer C-Domäne an, um die vorangegangene aktivierte Aminosäure mit der nachfolgenden zu verknüpfen [Linne and Marahiel, 2000].

Innerhalb eines Moduls bildet die A-Domäne (siehe Kap. 1.4) eine Matrize für die jeweilige vorgesehene Aminosäure und aktiviert diese. Die PCP-Domäne (siehe Kap. 1.7) dient als Transportdomäne. Sie überträgt die wachsende Peptidkette auf die PCP-Domäne des nächsten Moduls. Die C-Domäne (siehe Kap. 1.6) kondensiert die Proteinkette mit der nächsten aktivierten Aminosäure. Optional können in einem Modul noch weitere Domänen vorhanden sein, wie zum Beispiel Epimerasen (siehe Kap. 1.6), die die Konfiguration der assemblierten Aminosäuren ändern oder MT-Domänen (siehe Kap. 1.5), die N-, oder O-Methylierungen vornehmen können. So können im Gegensatz zur ribosomalen Peptidsynthese eine Vielzahl ungewöhnlicher Aminosäuren in ein Protein eingebaut werden [Marahiel, 2009] [Süßmuth, 2017]. In den folgenden Kapiteln soll die Funktion der einzelnen Domänen genauer betrachtet werden.

1.4 Adenylierungsdomänen (A-Domänen)

Die A-Domänen initiieren die nichtribosomale Peptidsynthese. Im ersten Schritt wird das spezifische Substrat der A-Domäne zum Aminoacyladenylat aktiviert (siehe Abb. 6) und in einem zweiten Schritt das aktivierte Substrat auf den Phosphopantetheinarm der PCP-Domäne übertragen (siehe Abb. 7) [Süßmuth, 2017].

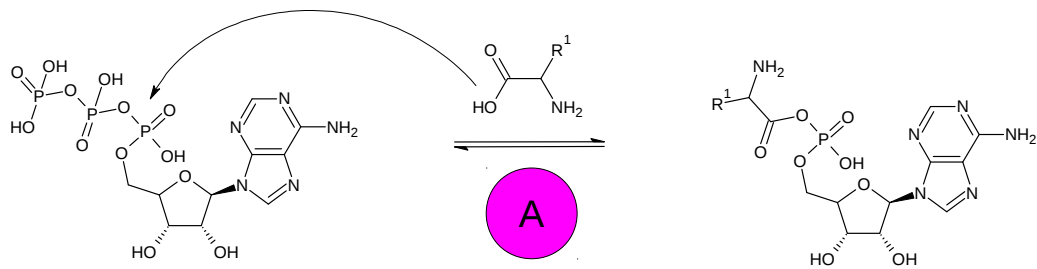


Abbildung 6: Adenylierungsreaktion von A-Domänen: Die selektierte Aminosäure wird mit Hilfe der A-Domäne unter Verbrauch von ATP zum Aminoacyladenylat aktiviert und Pyrophosphat wird abgespalten.

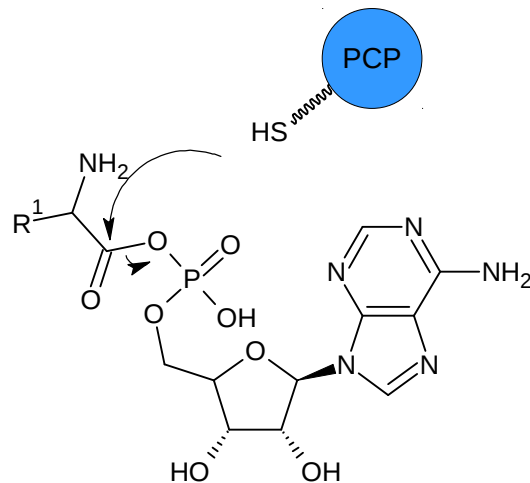


Abbildung 7: Thiolation: In einer zweiten Reaktion überträgt die A-Domäne die aktivierte Aminosäure auf den Phosphopantetheinarm der PCP-Domäne. Die PCP-Domäne dient als Transportdomäne für das entstehende Peptid.

A-Domänen gehören zur Enzymklasse der ANL Superfamilie der Adenylierungsenzyme (Acyl-CoA Synthetasen, NRPS-Adenylierungsdomänen und Luciferase Enzyme). Diese Enzymklasse katalysiert eine initiale Adenylierungsreaktion eines Carboxylatsubstrats unter ATP-Verbrauch. Mit diesen Enzymen teilt die A-Domäne vor allem eine hochkonservierte dreidimensionale Struktur. A-Domänen bestehen aus einer ca. 50 kDa N-terminalen Kerndomäne und einer ca. 10 kDa großen C-terminalen Subdomäne. Beide

Untereinheiten sind über eine flexible Gelenkregion miteinander verbunden [Conti et al., 1997].

Um diese dreidimensionale Struktur zu gewährleisten, haben A-Domänen trotz ihrer sehr heterogenen Aminosäuresequenzen sogenannte Core-Regionen. Diese Core-Regionen bilden hochkonservierte Aminosäuremotive innerhalb der Proteinsequenz der A-Domänen, die für alle A-Domänen identisch sind. Sie dienen als Orientierung bei der bioinformatischen Analyse der Sequenz. So liegen die Core-Regionen A1 bis A8 auf der größeren N-terminalen Kern-domäne und die Core-Regionen A9 bis A11 auf der kleineren C-terminalen A-Subdomäne [Conti et al., 1997] [Stachelhaus et al., 1999].

Nachdem mit der Kristallstruktur der A-Domäne von GrsA aus dem Gramicidin-BGC mit ihrem Substrat Phenylalanin die Struktur und die Substratbindungstasche der A-Domänen bekannt war [Conti et al., 1997], konnte aus Alignments von GrsA mit anderen A-Domänen die Bindungstasche für A-Domänen allgemein abgeleitet werden [Stachelhaus et al., 1999].

Die Substratbindungstasche der A-Domänen befindet sich zum Großteil in dem Bereich der Core-Regionen A4 und A5. Hier liegen neun der zehn Bindungsstellen. Eine weitere liegt innerhalb der Core-A10-Region auf der A-Subdomäne [Stachelhaus et al., 1999].

Die Aminosäuren in diesen zehn Positionen variieren von A-Domäne zu A-Domäne, je nachdem, welches Substrat von der A-Domäne aktiviert werden soll. Die Abfolge der Aminosäuren in diesen zehn Positionen legt also die Substratspezifität der A-Domänen fest. Daher werden diese zehn Aminosäuren auch als Stachelhauscode bezeichnet [Stachelhaus et al., 1999]. Position eins und zehn des Stachelhauscodes sind dabei weitestgehend konserviert. In Position eins befindet sich das Asp-235 und in Position zehn das Lys-517. Diese beiden Aminosäuren sind für die Koordination und Bindung der α -Aminogruppe und der Carboxygruppe der Substrataminosäure zuständig. Alle anderen Positionen variieren stark [Stachelhaus et al., 1999].

Die Selektivität einer A-Domäne ist nicht absolut. Die A-Domäne hat sozusagen eine gewisse Fehlertoleranz gegenüber ihrem Substrat. So können oft chemisch sehr ähnliche Aminosäuren wie z.B. Leucin und Isoleucin in unterschiedlichem Maß aktiviert werden. Um trotzdem am Ende der NRP-Synthese ein korrektes Peptid zu bekommen, hat die C-Domäne (siehe auch Kap. 1.6) desselben Moduls ebenfalls eine gewisse Substratselektivität und Korrekturfunktion. So kann es bei einer Fehlaktivierung zum Abbruch der NRP-Synthese kommen [Süßmuth, 2017].

Die Substrate der A-Domänen sind häufig proteinogene Aminosäuren oder auch Carbonsäuren. Die Vielzahl ungewöhnlicher Aminosäuren, die typischerweise in NRPs zu finden sind, entstehen nach oder vor der Assemblierung durch Modifikationen. Diese Modifikationen werden durch andere Domänenarten, wie MT-Domänen oder C-Domänen vorgenommen. Sie werden in den Kapiteln 1.5 und 1.6 noch ausführlicher betrachtet.

Die modifizierenden Enzymdomänen liegen oft C-terminal der A-Domänen. Einige von ihnen sind aber auch in die A-Domänen eingebettet. Dadurch entstehen Didomänen oder unterbrochene A-Domänen. Die Insertionspunkte dieser modifizierenden Domänen liegen sehr häufig zwischen den Core-Regionen der A-Domänen [Labby et al., 2015].

MT-Domänen (siehe auch Kap. 1.5) kommen häufig zwischen der Core-A8- und Core-A9-Region der A-Domäne vor, seltener auch zwischen der Core-A2- und Core-A3-Region, wie im Falle von Thiocoralin [Lombo et al., 2006]. Hier liegt die MT-Domäne aus TioN [Al-Mestarihi et al., 2014], die den seltenen Fall einer S-Methylierung darstellt, zwischen der Core-A2- und Core-A3-Region der A-Domäne. Im selben BGC befindet sich auch eine MT-Domäne, die eine zweifache Methylierung vornimmt [Mori et al., 2017]. Im BGC von Microcystin [Tillett et al., 2000] konnte eine zwischen der Core-A8- und Core-A9-Region unterbrochene A-Domäne in McyA gefunden werden, die genau wie bei den *frs*-BGC eine N-Methylierung des sehr seltenen Dehydroalanins vornimmt [Labby et al., 2015].

Weitere modifizierende Domänen, die in A-Domänen gefunden werden konnten, sind beispielsweise Reduktasedomänen, wie im Falle von Cereulid [Margarvey et al., 2006]. Diese liegt zwischen der Core-A8- und Core-A9-Region [Labby et al., 2015]. Im Falle von Myxothiazol [Silakowski et al., 1999] wurden Oxygenasedomänen zwischen der Core-A8- und Core-A9-Region identifiziert, von denen die Monooxygenasen in der Core-A4- und Core-A5-Region eingebettet waren [Labby et al., 2015].

A-Domänen sind ein Teil des großen NRPS-Enzymkomplexes, aber auch als isoliertes Enzym funktional. Eine Funktionsanalyse ist daher über den gut etablierten A-Domänenassay nach Phelan möglich [Phelan et al., 2009]. Dazu wird die A-Domäne mit dem zu testenden Substrat und $\gamma^{18}\text{O}_4\text{-ATP}$ inkubiert. Der $\gamma^{18}\text{O}_4\text{-ATP}$ -Verbrauch und das dadurch entstehende, leichtere $^{16}\text{O}_4\text{-ATP}$ kann detektiert werden. Bleibt die Adenylierungsreaktion, wie im Falle eines falschen Substrates, aus, wird das $\gamma^{18}\text{O}_4\text{-ATP}$ nicht verbraucht, und es kann kein $^{16}\text{O}_4\text{-ATP}$ detektiert werden (siehe auch Kap. 5.6.8).

Um voll funktionstüchtig zu sein, benötigen manche A-Domänen ein Helferprotein, das sogenannte MLP (siehe Kap. 1.8). Das Gen, welches für ein ca. 8 kDa großes Protein kodiert, ist Teil von vielen NRPS-BGC und wurde erstmals für ein BGC aus *Mycobacterium tuberculosis* beschrieben [Quadri et al., 1998]. MLPs haben Einfluss auf die Konformation, Löslichkeit, Spezifität und Funktion der A-Domänen [Schomer and Thomas, 2017]. Das MLP wird oft unabhängig von der A-Domäne exprimiert, bindet aber in einem stöchiometrischen Verhältnis von 1:1 mit der A-Domäne [Boll et al., 2011]. Im Falle von SlgN1, eines Hybrid-PKS/NRPS-Systems des Antibiotikums Streptolydigin [Deboer et al., 1955], kommt das MLP am N-terminalen Ende der A-Domäne als Fusionsprotein vor [Herbst et al., 2013].

Da A-Domänen die entscheidenden Enzyme sind, die für die Substratwahl der NRPSs verantwortlich sind, wurde viel unternommen, um die Funktionsweise und Konformationsänderungen von A-Domänen zu verstehen und

daraus Bioengineeringansätze zu entwickeln. Problematisch ist hier vor allem der strukturelle Erhalt und das Zusammenspiel der einzelnen Domänen, das reibungslos funktionieren muss. Im Kapitel 1.10 soll auf dieses Thema weiter eingegangen werden.

1.5 Methyltransferase-Domänen (MT-Domänen)

MT-Domänen sind ca. 45 kDa groß [Süßmuth, 2017] und für die Methylierung der Aminosäurebausteine während der Biosynthese von NRPs zuständig. Es existieren N-, O-, und C-methylierende Domänen [Ansari et al., 2008].

Die meisten MT-Domänen sind in die A-Domäne zwischen der Core-A8- und Core-A9-Region integriert. Es gibt aber auch MT-Domänen, die zwischen der Core-A2- und Core-A3-Region positioniert sind [Labby et al., 2015]. Die N-Methylierung ist die bekannteste Art der Methylierung und kommt nicht nur bei NRPS vor. Einen Überblick hierüber gibt der Review von Chatterjee et al [Chatterjee et al., 2013].

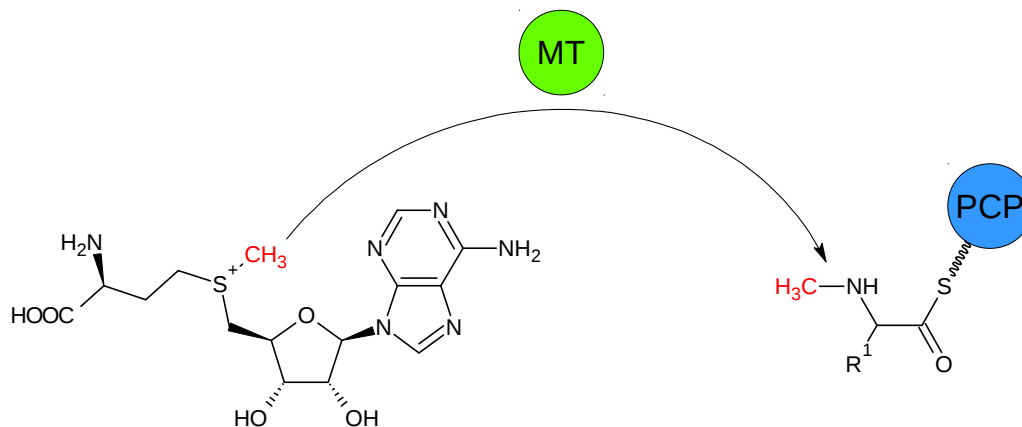


Abbildung 8: Bei der Methylierung wird eine Methylgruppe (rot) durch die MT-Domäne von SAM auf die Aminogruppe übertragen.

Bei der Methylierung von Aminosäuren in der NRP-Synthese wird, während die Aminosäure am PCP gebunden vorliegt, eine Methylgruppe von einem S-Adenosyl-methionin (SAM) als Methylgruppendonator auf die Ami-

nosäure übertragen und als Abgangsgruppe wird S-Adenosyl-L-homocystein abgespalten (siehe Abb. 8) [Ansari et al., 2008]. Das SAM wird in der MT-Domäne durch vier konservierte Bereiche komplexiert, darunter das hochkonservierte GxGxG-Motiv [Velkov and Lawen, 2003]. Die N-Methylierung wurde 2011 von Velkov et al. am Beispiel der Cyclosporin-Synthetase [Lawen and Zocher, 1990] genauer beschrieben. Mit Hilfe von N-Methyltransferaseinhibitoren wurde der Effekt einer Blockierung der MT-Domäne auf die Adenylierungsreaktion und Thioesterbildung sowie auf die Kondensationsreaktion untersucht. Die Bildung von Cyclosporin konnte durch Blockierung der Methylierungsreaktion vollständig unterbrochen werden [Velkov et al., 2011].

Ist eine Methylierung in der Biosyntheseabfolge vorgesehen, scheint sie für den Verlauf der Synthese essentiell zu sein. Bei der Actinomycin-Synthetase [Keller, 1987] wurde eine Valin aktivierende Domäne gegen eine N-Methyl-Valin-Domäne ausgetauscht. Anstatt Valin wurde Methylvalin eingebaut, die weitere Peptidsynthese wurde unterbrochen, da die folgenden Domänen beeinträchtigt waren [Schauwecker et al., 2000]. Bei dem N-methylierten Anthelmintikum PF1022A konnten Weckwerth et al. die komplette Synthese zellfrei durchführen. Hier brach die Synthese allerdings ab, als kein SAM für die Reaktion zugeführt wurde. Die Peptidsynthese ist in diesen Fällen an die Methylierung gebunden [Weckwerth et al., 2000].

Das PKS-NRPS-BGC des Yersiniabactin enthält eine C-Methyltransferase. Diese spaltet vermutlich den am C2 gebundenen Wasserstoff des PCP gebundenen Intermediates ab. Das daraus entstehende delokalisierte Carbanion greift wiederum die elektrophile Methylsulfoniumgruppe des ebenfalls gebundenen SAM an [Miller et al., 2001].

2018 gelang es Mori et al., eine Kristallstruktur von TioS, einer MT- unterbrochenen A-Domäne aus dem Thiocoralin-BGC, zu erstellen und zu analysieren. Hier konnte erstmals gezeigt werden, wie die dreidimensionale Struktur einer A-MT-Domäne aussieht. Die Domäne bildet ein hantelförmiges Kon-

strukt, dessen Enden die Hauptdomäne der A- und die MT-Domäne bilden. Die A-Subdomäne bildet das Verbindungsstück [Mori et al., 2018].

Im Falle des Thiocoralin-BGC konnten Mori et al. zwei interessante MT-Domänen finden. Einmal TioS, eine MT-Domäne, die in der Lage ist, eine Zweifachmethylierung an einem Stickstoff vorzunehmen, und TioN, der einzige Fall einer MT-Domäne, für die eine S-Methylierung zurzeit bekannt ist [Al-Mestarihi et al., 2014]. Die N-methylierende Domäne liegt zwischen der Core-A8- und Core-A9-Region und die S-methylierende Domäne zwischen der Core-A2- und Core-A3-Region [Mori et al., 2017].

Das Prinzip der eingebetteten MT-Domänen machten sich Lundy et al. zunutze, indem sie versuchten, eine monofunktionale A-Domäne ohne insertierte MT-Domäne mit jeweils zwei unterschiedlichen fremden MT-Domänen zu erstellen. Dazu fügten sie einmal die O-methylierende Domäne von KtzH und die N-methylierende Domäne von TioS zwischen die Core-A8- und Core-A9-Region ein. Die A-Domäne war danach in der Lage, spezifisch eine N- bzw. O-Methylierung vorzunehmen [Lundy et al., 2018]. Zwischen der Core-A8- und Core-A9-Region scheint also eine natürliche Schnittstelle für weitere Bioengineeringansätze zu sein.

1.6 Kondensierungsdomänen (C-Domänen)

C-Domänen sind sehr vielseitige Domänen innerhalb der NRPS-Komplexe. Sie sind ca. 450 Aminosäuren groß und sind N-terminal der A-Domänen angeordnet, um die PCP-gebundene Aminosäure mit der wachsenden Peptidkette zu verknüpfen [De Crecy-Lagard et al., 1995] [Marahiel et al., 1997].

Mit der ersten Kristallstrukturaufklärung einer C-Domäne von VibH (Vibriobactin-Synthetase) im Jahre 2002 konnte erstmals die Struktur der C-Domänen identifiziert und damit ihre Verwandtschaft zur Klasse der Chloramphenicol-Acetyltransferase-Superfamilie festgestellt werden. Sie bilden monomere Pseudodimere, bestehend aus zwei Unterdomänen, die eine V-förmige Tasche bilden, innerhalb derer die Kondensationsreaktion katalysiert wird

[Keating et al., 2002].

Mit der Kristallstruktur der C-Domäne von CDA konnten mögliche Konformationsänderungen während des Syntheszyklus am Computer simuliert werden [Bloudoff et al., 2013].

Eine weitere Klassifizierung der C-Domänen erfolgte 2007 auf Basis der phylogenetischen Verwandtschaftsverhältnisse [Rausch et al., 2007]. Allen C-Domänen war bis dahin das konservierte HHxxxDG Motiv gemeinsam. Man wusste, dass die Aufgaben der C-Domänen weitaus komplexer waren als nur die Verknüpfung zweier Aminosäuren. 2003 konnte die C5 C-Domäne der Tyrocidin-Synthetase eindeutig einer $^D C_L$ C-Domäne zugeordnet werden, also einer C-Domäne, die eine D-, mit einer L-konfigurierten Aminosäure verbindet [Clugston et al., 2003]. Rausch et al. konnten 2007 über eine phylogenetische Analyse weitere Subtypen von C-Domänen entschlüsseln und ihnen typische Sequenzmotive zuordnen, so dass von da an eine bioinformatische Vorhersage der Subtypen über einen "Hidden-Markov-Algorithmus" möglich war [Rausch et al., 2007].

Die C-Domänen lassen sich über diese phylogenetische Analyse in mehrere Unterklassen einteilen: Die $^L C_L$ -C-Domänen, die die Kondensation zweier L-Aminosäuren katalysieren, und die $^D C_L$ -C-Domänen, die die Kondensationsreaktion zwischen einer D- und einer L-Aminosäure katalysieren. Darüber hinaus gibt es Starter-C-Domänen, die stark verwandt sind mit den $^L C_L$ -C-Domänen, aber eine andere Substratspezifität aufweisen, da sie zum Beispiel Acylierungen katalysieren [Rausch et al., 2007]. Starter-C-Domänen sind in NRPS nicht zwangsläufig vorhanden, da eine Peptidbindung mit einem vorangehenden Peptid nicht nötig ist.

E-Domänen kommen ebenfalls aus der Klasse der C-Domänen. Sie agieren stark mit den C-Domänen zusammen [Linne and Marahiel, 2000]. Anhand der Kristallstruktur der E-Domäne der Tyrocidin-Synthetase konnte gezeigt werden, dass das ausschließlich bei E-Domänen konservierte Glu-882 als Katalysator wirkt und das His-743 im protonierten Zustand das Enolat-Intermediat

der Isomerisation stabilisiert [Samel et al., 2014]. Da in den Bakterienzellen vor allem L-Aminosäuren vorhanden sind, werden D-Aminosäuren eher selten direkt eingebaut, sondern aus Epimerisierung einer L-konfigurierten Aminosäure gebildet. Daher folgen auf E-Domänen im nächsten Modul in der Regel D C_L-C-Domänen [Luo et al., 2002].

Auch Zyklisierungsdomänen (Cy-Domänen) können der Klasse der C-Domänen zugeordnet werden. Sie teilen ebenso wie die E-domänen die Struktur der C-Domänen, allerdings ist hier das katalytische HHxxxDG-Motiv durch ein DxxxxDxxS-Motiv ersetzt, dessen Rolle bei der Katalyse nicht eindeutig bestimmt werden konnte [Dowling et al., 2016]. 2017 konnte der putative Zyklisierungsmechanismus anhand der Kristallstruktur der BmdB-Cy2 Domäne aus dem BGC der Bacillamid Synthetase gefunden werden. Dieser verläuft über eine Säure-Base-Katalyse mit anschließender Dehydratisierung. Anders als bei den C-Domänen spielen hier vor allem die konservierten Aminosäuren T1196 and D1226 eine Rolle [Bloudoff et al., 2017].

Darüber hinaus gibt es C-Domänen, die modifizierende Reaktionen katalysieren. Darunter befinden sich vermutlich Dehydratisierungsreaktionen, wie die Dehydroalaninbildung aus Serin (siehe Abb. 9).

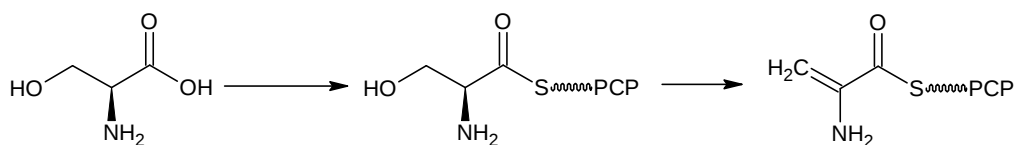


Abbildung 9: Putative Dehydroalaninbildung in NRPS: Dehydroalanin ist eine ungewöhnliche Aminosäure, die auch in NRPs nicht sehr oft vorkommt. Wahrscheinlich wird durch die vorangehende A-Domäne Serin aktiviert, welches in einem zweiten Reaktionsschritt putativ von der C-Domäne zu Dehydroalanin dehydratisiert wird.

Dehydroalanin ist eine für NRPs ungewöhnliche Aminosäure. Es gibt in der Literatur unter den NRPs lediglich zwei Peptide, bei denen eine Dehydroalaninbildung untersucht wurde. Das erste ist das Microcystin, welches Dehy-

droalanin als Baustein enthält. Man vermutet, dass die Dehydratisierung von Serin zu Dehydroalanin über die C-Domäne erfolgt, dies konnte aber noch nicht nachgewiesen werden [Tillett et al., 2000] [Crüsemann et al., 2018]. Das zweite Peptid ist Nocardicin, ein β -Lactam-Antibiotikum. Hier katalysiert die C-Domäne die Dehydratisierung von Serin zu Dehydroalanin als einem Intermediat der β -Lactam-Bildung. Die C-Domäne des Nocardicin-BGC enthält einen weiteren Histidinrest direkt N-terminal des HHxxxDG-Motivs, der vermutlich die β -Elimination von Wasser und eine anschließende β -Addition am Seryl-(dehydroalanyl)- β -Kohlenstoff katalysiert [Gaudelli et al., 2015].

Aufgrund der vielseitigen Reaktionen, die C-Domänen katalysieren, war ihre genaue Rolle in der NRP-Synthese lange unklar. 1997 kam in einem Review von Mahariel et al. schon das katalytische Motiv HHxxxDG zur Sprache, von dessen zweitem Histidinrest man eine katalytische Funktion als Base für eine nucleophile Reaktion in der Kondensationsreaktion zweier Aminosäuren vermutete [Marahiel et al., 1997]. Der genaue Kondensationsmechanismus war trotzdem lange unklar.

Der zweite Histidinrest des konservierten Motivs His-147 konnte 1998 von Stachelhaus et al. eindeutig dem katalytischen Zentrum zugeordnet werden. Eine Kondensationsreaktion fand ohne das His-147 nicht statt [Stachelhaus et al., 1998]. Eine weitere Mutationsanalyse diverser, aus 80 C-Domänen gefundener, konservierter Aminosäuren von Bergendahl et al. konnte das noch einmal bestätigen und ebenfalls eine katalytische Funktion des His-147 zeigen, sowie eine strukturelle Unterstützung der Kondensationsreaktion über das Arginin R62 und das Aspartat D151. Eine Mutation von diesen Aminosäuren bewirkte einen Ausfall der Kondensationsreaktion. Weitere Mutationen von konservierten Motiven führten vor allem zu unlöslichem Protein oder zu größeren strukturellen Änderungen im Protein [Bergendahl et al., 2002].

Der zweite Histidinrest des konservierten Motivs ist zwar essenziell, aber seine katalytische Funktion als Base wurde 2007 von Samel et al. für unwahr-

scheinlich erachtet. Da die Kondensationsreaktion eine weite pH-Toleranz hat, schlossen sie eine generelle Säure-Base-Katalyse aus. Sie vermuteten einen Mechanismus ähnlich dem der ribosomalen Peptidsynthese über elektrostatische Stabilisierung des Substrates in einem tetrahedralen Übergangszustand [Samel et al., 2007]. 2016 konnte dann durch Bloudoff et al. an der Kristallstruktur der C-Domäne des Calcium-Dependent Antibiotic CDA-C1 gezeigt werden, dass der Histidinrest vor allem dafür verantwortlich ist, die α -Aminogruppe des Substrates so in Position zu halten, dass ein nucleophiler Angriff erfolgen kann (siehe Abb. 10) [Bloudoff et al., 2016].

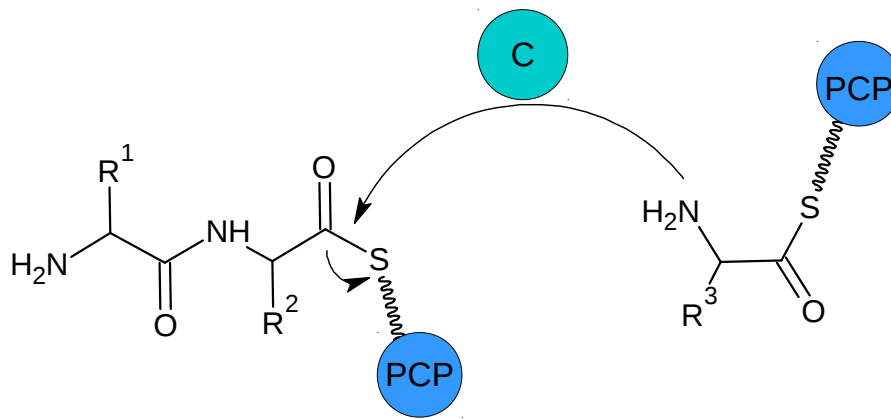


Abbildung 10: Kondensationsreaktion: Mit Hilfe der C-Domänen wird die Peptidbindung zwischen dem entstehenden Peptid des vorangehenden Moduls und der aktivierten Aminosäure des folgenden Moduls geknüpft.

C-Domänen katalysieren die Kondensationsreaktion und tragen in dieser Funktion auch zu einer Fehlerkorrektur durch von A-Domänen falsch aktivierten Substrataminosäuren bei. Sie enthalten zwei Bindungstaschen, eine an der N-terminalen Donorseite und eine stark stereoselektive auf der C-terminalen Akzeptorseite. Letztere weist ebenfalls eine Selektivität gegenüber der Seitenkette des Aminoacyl-Thioesters auf [Ehmann et al., 2000] [Lautru and Challis, 2004]. Die von der A-Domäne zur Verfügung gestellte aktivierte Aminosäure wird dadurch noch einmal "Korrektur gelesen". Die Donorseite ist gegenüber dem Substrat weitaus toleranter. Sowohl Größe als auch Konstitution des hier ankommenden Peptids scheinen keinen Einfluss auf eine

Kondensationsreaktion zu haben [Linne and Marahiel, 2000]. Die Selektivität an der Akzeptorseite scheint auch eine Rolle zu spielen, wenn es um die Initiation der Bildung eines NRPs geht. So kann zum Beispiel ein Elongationsmodul keine Initiationsreaktion ausführen. Durch Abtrennung der N-terminalen C-Domäne eines Elongationsmoduls konnte dieses in ein Initiationsmodul umgewandelt werden [Linne and Marahiel, 2000] [Luo et al., 2001] [Lautru and Challis, 2004] [Belshaw et al., 1999] [Ehmann et al., 2000].

2016 konnten Bloudoff et al. erstmals durch Punktmutation des S309G in der C-Domäne des Calcium-Dependent Antibiotic CDA-C1 die Substratspezifität der C-Domäne von Alanin zu Serin abändern [Bloudoff et al., 2016].

1.7 Peptidyl Carrier Protein oder PCP-Domänen

PCP-Domänen bilden flexible Domänen, die das Peptid während der NRP-Synthese zu den einzelnen Domänen transportieren. Sie bestehen aus 80-100 Aminosäuren, [Lai et al., 2006] die vier Helices bilden.

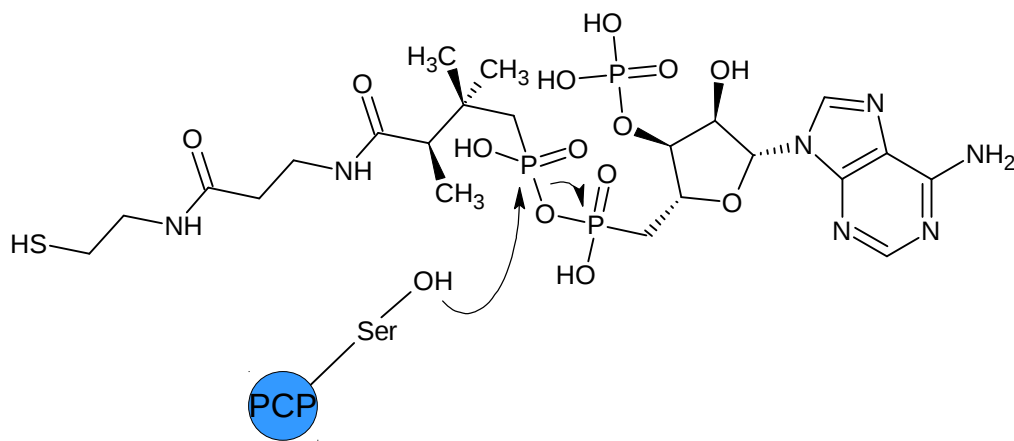


Abbildung 11: Phosphopantetheinylierung: An die inaktive PCP-Domäne wird unter Mitwirkung einer PPTase ein Phosphopantetheinrest an einen hochkonservierten Serinrest des Apo-PCP gebunden. In dieser Form ist die PCP-Domäne in der Lage, sowohl entstehende Peptide als auch aktivierte Aminosäuren kovalent zu binden.

Die vier helikalen Strukturen der PCP-Domänen haben auf der nach innen liegenden Seite hydrophobe Seitenketten und, zur äußeren Umgebung gerichtet, polare Seitenketten [Weber et al., 2000]. In dieser Form ist eine PCP-Domäne noch inaktiv. Sie wird durch Phosphopantetheinyl-Transferasen [Lambalot et al., 1996] in ihre aktive Form umgewandelt. Dabei wird ein Phosphopantetheinrest auf das Ser45 des konservierten Motivs GxxS übertragen [Weber et al., 2000]. Dieser Phosphopantetheinrest hat eine terminale reaktive SH-Gruppe [Abe et al., 2018], die für die kovalente Fixierung der Aminosäure verantwortlich ist (siehe Abb. 11). PCPs besitzen wahrscheinlich keine eigene Substratspezifität [Doekel and Marahiel, 2000].

Tan et al. konnten anhand der Kristallstruktur der A-PCP Domäne von McyG mit dem Intermediat L-Phe-AMP erste Einblicke in das Zusammenspiel von A-Domäne, PCP und deren katalytischen Regelkreis erhalten [Tan et al., 2015]. Owen et al. versuchten nach Austausch verschiedener PCP-Domänen im Indigoidin-Cluster, den Farbstoff weiterhin zu produzieren. Sie stellten fest, dass die Produktion des Indigoidins aufrecht erhalten werden konnte, sobald eine PCP-Domäne eingefügt wurde, die ursprünglich N-terminal einer TE-Domäne stand. Mit PCP-Domänen, die ursprünglich N-terminal einer C-Domäne lagen, kam die Produktion zum Erliegen. Es scheint also einen Unterschied zwischen PCP-Domänen, auf die eine C-Domäne folgt, und zwischen solchen, auf die eine TE-Domäne folgt, zu geben [Owen et al., 2016].

1.8 MbtH-ähnliche Proteine

MbtH ähnliche Proteine, auch MLPs (MbtH like proteins) genannt, sind Proteine, die eine sehr hohe Ähnlichkeit zum Protein MbtH aus dem Mycobactin-BGC aus *Mycobacterium tuberculosis* aufweisen [Quadri et al., 1998]. Gene für MLPs sind sehr oft in NRPS-BGC zu finden und besitzen eine Funktion als Helferprotein für manche A-Domänen.

MLP-Gene kodieren für ein ca. 8 kDa großes Peptid. Die Funktion dieses Pep-

tids war lange unklar, da eine NRP-Synthese oft trotz Knockout des MLP-Gens funktionierte. Dies wurde am Beispiel der Biosynthese des Glycopeptids Balhimycin gezeigt. Nach einem Knockout des MLP-Gens im Balhimycin-BGC wurde die Synthese des Wildtyps nicht beeinträchtigt [Stegmann et al., 2006]. Der Grund dafür war aber keineswegs, dass MLPs für die Biosynthese unwichtig sind, sondern, dass sie funktional gegeneinander austauschbar sind. So konnten Wolpert et al. an der Biosynthese von Clorobiocin zeigen, dass die Clorobiocinsynthese erst stoppte, als sämtliche im Wirtsorganismus vorhandene MLP-Gene inaktiviert wurden [Wolpert et al., 2007]. Zu einem ähnlichen Ergebnis kamen auch Lautru et al., indem sie herausfanden, dass nur ein MLP-Gen für die Produktion des Peptidsiderophores Coelichelin und dem Calcium-Dependent Antibiotic notwendig ist. Die MLPs waren gegeneinander austauschbar [Lautru et al., 2007]. In der Glidobactin Biosynthese konnte nur durch Coexpression des Initiationsmoduls GlbF mit MLP lösliches und aktives Protein gewonnen werden [Imker et al., 2010].

Zhang et al. und Felnagle et al. konnten zeitgleich nachweisen, dass MLPs in der Adenylierungsreaktion mancher A-Domänen benötigt werden [Zhang et al., 2010] [Felnagle et al., 2010]. Es gibt also A-Domänen, die ohne MLPs funktionieren und solche, die ein MLP benötigen. Boll und Kollegen änderten durch eine einfache Mutation des Leucin-383 zu Methionin eine MLP-abhängige A-Domäne in eine MLP-unabhängigen A-Domäne ab. Sie fanden heraus, dass A-Domäne und MLP in einem stöchiometrischen Verhältnis von 1:1 binden [Boll et al., 2011]. MLPs scheinen eine gewisse chaperonähnliche Eigenschaft zu besitzen, indem sie die Löslichkeit, Stabilität und Aktivität mancher A-Domänen beeinflussen [Heemstra et al., 2009].

2013 zeigten D. Herbst und Kollegen anhand der Kristallstruktur des Adenylierungsenzyms SlgN1 [Herbst et al., 2013], das in die Biosynthese des Antibiotikums Streptolydigin involviert ist, dass die Aminosäuren Trp-25, Trp-35 des MLP-Proteins und Ala-433 der A-Domäne für die Domäneninteraktion mit dem MLP verantwortlich sind. Das MLP hat keinen Kontakt zum Substrat der A-Domäne. 2016 wurde in einem ähnlichen Ansatz anhand der

NRPS EntF die komplette, hochdynamische Domänenarchitektur von EntF mit assoziiertem MLP YbdZ aufgeklärt [Miller et al., 2016]. Durch Kristallisation und Röntgenstrukturanalyse konnten Tarry et al. herausfinden, wie MLPs in der Gesamtstruktur der Domänen eingebettet sind [Tarry et al., 2017].

MLP-Proteine haben auch eine funktionelle Varianz. Schomer et al. tauschten das MLP YbdZ der Enterobactinsynthase gegen andere MLPs aus. Sie wiesen nach, dass die A-Domäne der Enterobactinsynthase aus EntF durch unterschiedliche MLPs unterschiedlich gut ihr Substrat aktivieren konnte. Das zeigte sich in starken Ausbeuteschwankungen in der Enterobactin-Produktion. MLPs scheinen also unterschiedlich in der Lage zu sein, die Löslichkeit oder die Katalyse der Adenylierungsreaktion zu beeinflussen. Die Komplexität der Interaktion ist bis jetzt nicht restlos geklärt [Schomer and Thomas, 2017].

1.9 Thioesterasen (TE-Domänen)

TE-Domänen sind die endständigen Domänen der NRPS. Sie sind ca. 30 kDa groß und agieren als α/β -Hydrolasen [Süßmuth, 2017]. Sie sorgen für die Produktfreisetzung des fertigen Peptids durch eine nucleophile Reaktion. Nach Bindung der Peptidkette an das Serin des aktiven Zentrums der TE-Domäne wird die Freisetzung des Produkts entweder durch Hydrolyse oder Aminolyse bei linearen Peptiden, oder auch durch intramolekulare Zyklisierung des Peptids erreicht (siehe Abb. 12) [Kopp and Marahiel, 2007] [Süßmuth, 2017].

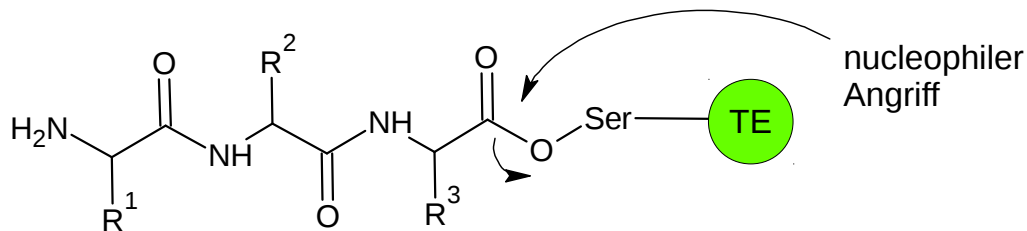


Abbildung 12: Durch Hydrolyse oder Aminolyse bei linearen Peptiden oder eine intramolekulare Zyklisierungsreaktion wird das fertige Peptid freigesetzt.

1.10 Vorstufengerichtete Biosynthese von NRPs und Bioengineering von NRPS

Der Wunsch, neue Naturstoffanaloga zu synthetisieren, führte im Laufe der Zeit dazu, modulare Biosynthesysteme, wie sie die NRPS-BGC darstellen, durch vorstufengerichtete Biosynthese, Mutasyntese und genetische Manipulation zu beeinflussen oder abzuändern.

Zu Beginn konzentrierte man sich jedoch auf die einfache Zufütterung von Vorstufen. Bei diesem Prinzip werden dem nativen Naturstoffproduzenten bestimmte Aminosäuren und andere Vorstufen zur Verfügung gestellt, die den

natürlichen Substraten ähneln, aber sich hinsichtlich der gewünschten Substituenten unterscheiden. Die ersten Versuche mit dieser Methode wurden schon durchgeführt, bevor man Gencluster kannte. Die vorstufengerichtete Synthese funktionierte am besten in Fällen, in denen das zugefütterte Substrat nur minimal anders war als das natürliche Substrat. So haben Beacco et al. beispielsweise ein chloriertes Alkaloidanalogon eines Ergopeptins durch Fütterung einer Kultur von *Claviceps purpurea* mit D,L-*p*-Chlorophenylalanin erhalten [Beacco et al., 1978]. Bister et al. tauschten das Chlorid im Fermentationsmedium von *Amycolatopsis balhimycina* [Wink et al., 2003] durch andere Halogensalze aus und erhielten unterschiedliche Halogenierungsmuster bei Balhimycin [Bister et al., 2003]. Auch bei Iturin A wurde mit der Vorstufe 3-Fluoro-L-Tyrosin ein fluoriertes Produkt erzeugt [Moran et al., 2009]. Bei Aureobasidin [Takesako et al., 1991], einem zyklischen Depsipeptid, konnten durch Zufütterung unterschiedlicher Aminosäuren sogar ganze Aminosäurebausteine in der Peptidsequenz ausgetauscht werden [Takesako et al., 1996]. Grüschow et al. zeigten 2009 am Beispiel des Antibiotikums Pacidamycin [Karwowski et al., 1989] [Chen et al., 1989] [Fernandes et al., 1989], dass durch Zufütterung von in unterschiedlichen Positionen substituiertem *Bromo*-Tryptophan abweichende Ausbeuten im Vergleich zur Produktion des nativen Pacidamycins erzielt werden konnten. Bei Zufütterung von Aminosäurevarianten, in denen sich der Brom-Substituent in Position zwei und sieben befand, war die Produktion teilweise höher als die des natürlichen Pacidamycins, wohingegen das Protein bei Substituenten in den Positionen vier, fünf und sechs in deutlich geringerem Ausmaß produziert wurde [Grüschow et al., 2009].

Auch *in vitro* konnten durch Fütterungsexperimente Analoga hergestellt werden. Durch Zugabe von L-Alanin, L-Threonin, L-Valin oder L-Norvalin anstatt L-2-Aminobuttersäure, zu einer Enzymfraktion des Rohextrakts von *Tolypocladium inflanatum* erhielt man unterschiedliche Cyclosporin-Derivate [Billich and Zocher, 1987]. Traber et al. benutzten dieses System dazu, in Fermentationskulturen diese Cyclosporine in ausreichender Menge zu synthetisieren, um Strukturaufklärung betreiben zu können [Traber et al., 1989].

Bei Enniatin, einem zyklischen Hexadepsipeptid, wurden *in vitro* durch Precursorzugabe in einer Reaktion mit der Enniatin-Synthetase und *in vivo*-Versuchen durch Fütterung einer Kultur von *Fusarium sambucinum* und *scirpi* diverse Enniatin-Analoga erzeugt. So konnten verschiedenen Aminosäuren gegeneinander ausgetauscht werden und die D-2-Hydroxyisovalerate gegen D-2-Hydroxybuttersäure und D-Laktat ausgetauscht werden [Krause et al., 2001]. Ein Review von Thiericke fasst die Thematik der vorstufengerichteten Biosynthese zusammen [Thiericke and Rohr, 1993].

Einen ähnlichen Ansatz wie die vorstufengerichtete Biosynthese verfolgt die Mutasyntese, bei der die Vorstufen nicht den nativen, sondern mutierten Naturstoffproduzenten zugefüttert werden. Auf diese Weise aktiviert das Enzym Vorstufen, die vorher nicht als Substrat akzeptiert wurden.

Beispielsweise konnten Ankerbauer et al. einen *Pseudomonas aeruginosa* Mutanten herstellen, dessen Biosynthesegen für die Salicylsäureproduktion ausgeknockt war. Diese Mutante war in der Lage, exogen zugefütterte Salicylsäureanaloga in Pyochelin einzubauen [Ankenbauer et al., 1991]. Das Glycopeptidantibiotikum Balhimycin konnte durch einen Knockout des β -Hydroxy-Tyrosinbildenden Gens und Zufütterung von 3-Fluoro- β -hydroxy-Tyrosin als Fluorbalhimycin erhalten werden [Weist et al., 2002]. Powell et al. knockten eine MT-Domäne im BGC des Calcium-Dependent Antibiotic aus und erhielten durch Zufütterung von dreifach methyliertem Substrat trotzdem ein vollständig methyliertes Calcium-Dependent Antibiotic. Des Weiteren wurde anstatt der dreifach methylierten Aminosäure eine dreifach fluoromethylierte Glutaminsäure erfolgreich zugefüttert [Powell et al., 2007]. Im Falle des Antitumorwirkstoffs Sibiromycin erzeugte man durch Knockout des Methyltransferasegens und anschließender Zufütterung von 4-Methylantranilsäure einen Wirkstoff, der weiterhin antitumorwirksam war, aber verringerte Kardiotoxizität aufwies [Yonemoto et al., 2012].

Sowohl die vorstufengerichtete Synthese, als auch die Mutasyntese funktionieren nur begrenzt, da die Substrate wenig variabel sind. Mit der detaillierten Aufklärung der BGCs versuchte man immer mehr auf genetischer Ebene

die BGC zu manipulieren. Hier sind die Ausbeuten oft nicht sehr hoch, da mit einem Engineering der NRPS die dreidimensionale Struktur der Enzyme verändert wird und diese oft in ihrer Funktion beeinträchtigt sind. Die Versuche reichen von einfachen Punktmutationen über die Rekombination von Sequenzen im aktiven Zentrum der A-Domäne bis hin zum Austausch von Domänen oder sogar Modulen [Süßmuth, 2017].

Crüseemann et al. konnten im Falle des Hormaomycins A-Domänen finden, die sich nur im Bereich der für die Spezifität kodierenden Sequenzen stark unterschieden, ansonsten aber eine hohe Identität aufwiesen. Diese natürlichen Rekombinationsschnittstellen wurden genutzt, um eine ca. 400 bp große Sequenz im kodierenden Bereich der A-Domäne auszutauschen, und damit die Substratspezifität der A-Domänen abzuändern [Crüseemann, 2013]. Bei Kries et al. wurde ähnlich wie bei Crüseemann et al. durch Austausch eines Sequenzabschnittes im Bereich des Stachelhauscodes der A-Domäne der *cyclo*(D-Val-Pro)-Synthetase mit der kodierenden Sequenz der Phenylalanin aktivierenden Domäne der Gramacidinsynthetase die Spezifität von Valin zu Phenylalanin abgeändert [Kries et al., 2015]. Ein relativ kleiner Eingriff in die NRPS der Tyrocidinsynthetase wurde von Villiers et al. vorgenommen. Durch Mutation des Stachelhauscodes erfolgte eine Abänderung der Spezifität der Phenylalanin aktivierenden A-Domäne aus TycA zu Alanin [Villiers and Hollfelder, 2011].

Nicht nur die A-Domänen spielen eine Rolle bei der Wahl des Substrates, sondern auch die C- und andere Domänen. Daher gab es zahlreiche Versuche, vor allem A-Domänen zusammen mit ihren PCP-, C-, oder anderen Domänen auszutauschen. So ist es beispielsweise gelungen, den endständigen Bereich der Daptomycin-NRPS mit dem von zwei sehr ähnlichen NRPS des Antibiotikums A54145 und des Calcium-Dependent Antibiotic (CDA) zu ersetzen. Während Daptomycin Kynurenin enthält, enthalten CDA und A54145 Tryptophan und entweder Isoleucin oder Valin. Zuerst wurde eine Knockout-Variante des Daptomycin-BGC hergestellt, dessen C-terminales Gen *dptD* fehlt, wodurch die Daptomycin Produktion zum Erliegen kam. Durch Co-

expression mit einem Vektor, der *dptD* enthält, konnte die Daptomycin-Produktion wieder hergestellt werden. Von den C-terminalen Genen des CDA-BGC und des A54145 wurden ebenfalls Vektoren erstellt, die zusammen mit dem Knockout-Vektor exprimiert wurden. Bei letzteren beiden wurden anstatt Kynurenin Tryptophan und Isoleucin/Valin in das Peptid eingebaut [Miao et al., 2006] [Coeffet-Le Gal et al., 2006].

Calcott et al. führten am Pyoverdin-BGC Studien zum Austausch von C- und A-Domänen durch. Hier versuchte man die Spezifität der Threonin aktivierenden C-terminalen A-Domäne des endständigen Moduls durch Austausch von entweder nur A-Domänen oder CA-Domänen abzuändern. Auch hier wurde, ähnlich wie in den Versuchen bei Daptomycin, das C-terminale Gen *pvdD* in ein separates Plasmid kloniert. Die A-, oder CA-Domäne des C-terminalen Moduls wurde entweder durch eine andere A-Domäne oder eine andere A-Domäne mit deren C-Domäne ersetzt. Die für den Austausch verwendeten A- und C-Domänen wurden aus anderen Modulen des Pyoverdin-BGC, oder aus homologen *Pseudomonas*-BGC genommen. Wurde die native A-Domäne gegen eine andere Threonin-aktivierende A-Domäne ausgetauscht, wurde weiterhin Pyoverdin produziert, wenn auch mit unterschiedlichen Ausbeuten. A-Domänen mit anderen Spezifitäten wurden durch dieses System nicht akzeptiert und produzierten ebenfalls ausschließlich Pyoverdin, allerdings nur in Spuren. Wurde mit der A-Domäne ebenfalls die C-Domäne ersetzt, konnte in zwei Fällen das Threonin einmal durch Serin und einmal durch Lysin ersetzt werden. In anderen Fällen war jedoch der Austausch der CA-Domäne erfolglos [Calcott et al., 2014].

Die Funktionalität einer NRPS wird insbesondere durch die dreidimensionale Struktur und das komplexe Zusammenspiel der einzelnen Domänen untereinander bestimmt. Werden durch Modifikationen die Struktur oder die Beweglichkeit einzelner Domänen stark verändert oder eingeschränkt, kann das zu einem Funktionsverlust der Enzyme führen [Winn et al., 2016]. Tan et al. zeigten durch die Kristallstrukturanalyse der A-PCP-Domäne aus McyG des Microcystin, wie A- und PCP-Domäne interagieren. So ist beispielsweise

die Core-A9-Region der A-Domäne an der Interaktion mit der PCP beteiligt [Tan et al., 2015]. Die Manipulation der Linkerregion zwischen PCP und C-Domänen des Daptomycin-BGC führten zu keiner größeren Beeinträchtigung der Daptomycin-Produktion [Doekel and Marahiel, 2000]. Eine weitere Kommunikationsstruktur der Domänen untereinander ist in den Kommunikationsdomänen (communication-mediating domains = COM-Domänen) zu sehen. Zwischen den Domänen befinden sich jeweils miteinander kompatible COM-Domänen, die miteinander interagieren können. Hier führten kleine Modifikationen innerhalb der Domänen-Sequenz dazu, dass keine Interaktion zwischen den COM-Domänen-Partnern mehr stattfand [Hahn and Stachelhaus, 2006].

Bozhüyük et al. gelang es, ein Grundprinzip aus den bisherigen Kenntnissen abzuleiten. Die Module wurden an neu definierten Linkerregionen separiert. Dadurch entstanden CA- oder CAT-Module, die gleich einem Baukastenprinzip aneinander gesetzt wurden, ohne die übergeordnete Domänenarchitektur zu beeinflussen. Mit diesem Ansatz ließen sich künstliche Peptide mit Ausbeuten zwischen 20 mg/L und 50 mg/L generieren [Bozhüyük et al., 2018].

Viele Bioengineeringansätze führen heute schon zu guten Ergebnissen. Es ist unerlässlich, die Komplexität der NRPS-Systeme zu verstehen, wenn man diese für eine erfolgreiche rekombinante Produktion von Naturstoffen nutzen will.

2 Zielsetzung der Arbeit

FR ist ein Naturstoff aus *A. crenata*, der von symbiontischen Bakterien dieser Pflanze gebildet wird. Aufgrund seiner einzigartigen Hemmung von G_q-Proteinen, welche maßgeblich an der Signaltransduktion von GPCRs beteiligt sind, ist die Aufklärung der Biosynthese als Basis für das Bioengineering von Analoga essenziell. Zu Beginn dieser Arbeit lag nur das BGC aus *Cand. B. crenata* vor. Erste bioinformatische Untersuchungen ließen vermuten, dass es sich um das BGC des FR-Peptids handelt.

Das Hauptziel dieser Arbeit ergab sich mit dem Fund des zweiten *frs*-BGC im Genom von *C. vaccinii*. Die BGC von *Cand. B. crenata* und von *C. vaccinii* sollten bioinformatisch analysiert und auf DNA- und Proteinebene verglichen werden. Daraus sollten sowohl Erkenntnisse über die Details der Biosynthese des FR-Peptids, als auch neue Erkenntnisse über mögliche Bioengineeringansätze gewonnen werden. Aus der detaillierten Strukturanalyse der BGCs sollten auch Hinweise auf mögliche evolutionäre Vorgänge in der Entstehung der BGCs gewonnen werden.

Ein zweites Ziel dieser Arbeit war es, die Spezifität der A-Domänen aus FrsA und FrsD durch Expression zu belegen und die OH-Leucin-Biosynthese während der FR-Biosynthese nachzuvollziehen. Dies war notwendig, da die erste bioinformatische Analyse der A-Domänen nicht erkennen ließ, ob OH-Leucin oder Leucin von der A-Domäne aktiviert wird. Dies sollte auch die Rolle von FrsH im *frs*-BGC beleuchten.

3 Ergebnisse

3.1 Bioinformatische Analyse der Biosynthesegencluster *bcfrs* und *cvfrs* von FR

Carlier et al. konnten 2016 das Genom des Endosymbionten von *A. crenata* sequenzieren. Auf einem Plasmid wurde ein 35,800 kb großes BGC (*bcfrs*-Gencluster) gefunden [Carlier et al., 2016]. Eine Analyse mit antiSMASH ergab, dass dieses BGC eine NRPS darstellt und aus acht Modulen und insgesamt 31 Domänen aufgebaut ist. Die NRPS weist eindeutige Kolinearität mit der Struktur von FR auf, die hier durch bioinformatische und *in-vitro*-Analyse bestätigt werden soll (siehe Abb. 4). Weiterhin wurden in unserer Arbeitsgruppe in einem anderen Bakterium Teile einer BGC-Sequenz gefunden, die eine große Ähnlichkeit mit dem *bcfrs*-Gencluster aus *Cand. B. crenata* aufweist. Dieses letztere Gencluster (*cvfrs*) stammt aus *C. vaccinii* und lag zu Beginn dieser Arbeit lückenhaft auf mehreren Contigs vor. Diese Lücken konnten durch uns mit Hilfe von PCR vollständig geschlossen werden. Im Folgenden sollen beide BGC bioinformatisch analysiert werden.

3.1.1 Analyse der *bcfrs*- und *cvfrs*-Biosynthesegencluster aus *Cand. B. crenata* und *C. vaccinii* mit antiSMASH

Die antiSMASH-Analyseplattform (antibiotics & Secondary Metabolite Analysis Shell) kann die BGC aller bekannten sekundären Metabolitklassen in Gensequenzen identifizieren und lokalisieren (siehe Kap. 5.1). Die Annotation erfolgt nach der Durchführung von Alignments mit bekannten BGC aus einer Datenbank [Medema et al., 2011]. Seit der Einführung wurde antiSMASH kontinuierlich weiterentwickelt. Die Uploadformate wurden erweitert und Subcluster, die an der Bildung von Baueinheiten beteiligt sind, können analysiert werden [Blin et al., 2013]. Es wurde der Cluster Finder Algorithmus entwickelt, um mögliche BGC unbekannter Typen sowie Ähnlichkeiten zwischen dem identifizierten BGC und schon bekannten BGC zu finden. Der Algorithmus basiert auf dem Hidden Markov Modell und auf der Annahme, dass auch BGC für unbekannte Metabolitklassen die gleichen Enzymfamili-

en beinhalten. Mittlerweile können auch aktive Zentren aufgrund bekannter Schlüssel motive lokalisiert und annotiert werden [Weber et al., 2015]. 2017 wurde antiSMASH um eine Datenbank erweitert, die von antiSMASH annotierte BGC zur Verfügung stellt [Blin et al., 2017b]. Der Review von Blin et al. dokumentiert nochmals alle Leistungsmerkmale von antiSMASH [Blin et al., 2017a].

Die Analyse der *frs*-Genclustersequenzen durch antiSMASH zeigt, dass beide NRPS-BGC den gleichen Aufbau und die gleiche Anzahl an Domänen haben. Beide BGC bestehen aus jeweils acht Genen, acht Modulen und 31 Domänen. Sie enthalten ebenfalls beide zwei TE-Domänen, was auf zwei NRPS-Systeme hindeutet. Die erste monomodulare NRPS FrsA ist für die Synthese des N-Propionyl-hydroxyleucin zuständig. Dies soll später anhand der bioinformatischen Analyse der C-Domänen und anhand von *in-vitro*-Untersuchungen der A-Domäne bestätigt werden. Diese Seitenkette des FR-Moleküls wird dann vermutlich unter Mitwirkung der TE₁-Domäne ähnlich wie in der Salinamid Biosynthese [Ray et al., 2016] auf das zirkuläre Heptapeptid übertragen, welches von der zweiten NRPS gebildet wird. Die zweite, weitaus größere NRPS mit jeweils sieben weiteren Elongationseinheiten bildet entsprechend unserer Biosynthesehypothese das Depsipeptid mit sieben weiteren Aminosäuren.

Verwendeter Algorithmus	FrsA	FrsD	FrsE 1	FrsE 2	FrsF 1	FrsF 2	FrsG 1	FrsG 2
Stachelhaus code	boh-d-leu d-leu leu	boh-d-leu d-leu leu	N/A	gln ser	N/A	N/A	boh-d-leu d-leu leu	nme-thr
NRPSPredictor 3 SVM	leu	leu	N/A	ser	ala	ala	leu	thr
pHMM	tyr	tyr	trp	ser	ala	ala	tyr	thr
SANDPUMA ensemble	leu	leu	N/A	ser	N/A	N/A	leu	thr
ID to nearest neighbour (%)	51.69	51.69	53.09	58.6	53.83	54.1	51.27	58.68

Tabelle 1: A-Domänen Spezifitätsvorhersage nach unterschiedlichen Algorithmen durch antiSMASH von *Cand. B. crenata*. Für die A-Domänen von FrsD-FrsG werden jeweils in gleicher Reihenfolge Leucin oder Hydroxyleucin (Leu/OH-Leu), Phenylalanin (Phe) oder Tryptophan (Trp), Serin (Ser), Alanin (Ala), Alanin, Leucin oder Hydroxyleucin und Threonin (Thr) vorhergesagt. Der pHMM-Algorithmus weicht bei FrsA, D und G1 von den Vorhersagen der anderen Algorithmen ab. Schwierigkeiten in der Vorhersage gibt es bei FrsE1. Das Substrat konnte hier nicht eindeutig identifiziert werden, wobei die Aromatizität erkannt wurde. Abkürzungen: boh-d-Leu= β -Hydroxyleucin; bht= Hydroxytyrosin [Blin et al., 2013]; N/A = keine Angabe

Verwendeter Algorithmus	FrsA	FrsD	FrsE 1	FrsE 2	FrsF 1	FrsF 2	FrsG 1	FrsG 2
Stachelhaus code	N/A	N/A	N/A	gln ser	ala gly	ala gly	N/A	thr
NRPSPredictor 3 SVM	leu	leu	phe	ser	ala	ala	leu	thr
pHMM	bht	bht	trp	ser	ala	ala	bht	thr
SANDPUMA ensemble	leu	leu	phe	ala	N/A	N/A	leu	thr
ID to nearest neighbour	49.24	49.24	56.09	60.47	56.24	57.14	49.24	59.78

Tabelle 2: A-Domänen Spezifitätsvorhersage nach unterschiedlichen Algorithmen durch antiSMASH von *C. vacci-
nii*. Für die A-Domänen von FrsD-FrsG werden jeweils in gleicher Reihenfolge Leucin (Leu), Phenylalanin (Phe)
oder Tryptophan (Trp), Serin (Ser), Alanin (Ala), Alanin, Leucin und Threonin (Thr) vorhergesagt. Der pHMM-
Algorithmus weicht bei FrsA, D und G1 von den Vorhersagen der anderen Algorithmen ab. Schwierigkeiten in der
Vorhersage gibt es bei FrsE1. Das Substrat konnte hier nicht eindeutig identifiziert werden, wobei die Aromatizität
erkannt wurde. Abkürzungen: bht= Hydroxytyrosin [Blin et al., 2013] ; N/A = keine Angabe

Vergleicht man bioinformatisch die Spezifität der A-Domänen, so ergibt sich für beide BGC eine ähnliche Prognose (siehe Tabellen 1 und 2). Für die A-Domänen von FrsD-FrsG werden jeweils in gleicher Reihenfolge Leucin oder Hydroxyleucin, Phenylalanin oder Tryptophan, Serin, Alanin, Alanin, Leucin oder Hydroxyleucin und Threonin vorhergesagt. Abweichungen von dieser Hypothese zeigt vor allem der pHMM-Algorithmus. In manchen Fällen konnte durch die Algorithmen keine Zuordnung für ein Substrat getroffen werden. Die meisten Substrate wurden von mindestens zwei der Algorithmen im Sinne der Hypothese bestätigt. Die einzige Abweichung zeigt sich bei FrsE1 (Modul 3). Anstatt Phenyllaktat, d.h. dem in FR vorhandenen Baustein, wurde hier Phenylalanin oder Tryptophan als Substrat erkannt. Eine Vorhersage scheint hier aufgrund mangelnder Datenlage noch nicht möglich zu sein. Allerdings wurde sowohl für das *bcfrs*-Gencluster von *Cand. B. crenata*, als auch für das *cvfrs*-Gencluster von *C. vaccinii* mit Phenylalanin und Tryptophan eine aromatische Aminosäure vorhergesagt, die dem Phenyllaktat sehr ähnlich ist.

Die zweite A-Domäne von FrsE (Modul 4) aktiviert laut antiSMASH-Analyse Serin. Das FR-Peptid enthält an dieser Stelle jedoch N-Methyl-Dehydroalanin. Eine ähnliche Konstellation findet sich in der Biosynthese für Microcystin wieder [Tillett et al., 2000], für das ebenfalls Serin als Substrat vorhergesagt wird. Da in den *frs*-BGC kein offensichtliches Enzym für die Bildung von Dehydroalanin aus Serin vorhanden ist, kann vermutet werden, dass ein ähnlicher Biosynthese-Mechanismus vorliegt [Crüsemann et al., 2018]. Für das Microcystin wird postuliert, dass unter Mitwirkung einer C-Domäne eine Dehydratisierung des Serins zu Dehydroalanin stattfindet [Tillett et al., 2000] [Crüsemann et al., 2018]. Dieser Mechanismus ist jedoch noch nicht belegt. Eine weitere bioinformatische Analyse der C-Domänen wird in Kapitel 3.4 vorgenommen.

FrsE und FrsF enthalten jeweils eine von vier MT-Domänen im BGC. Diese folgen auf die A-Domänen A4 und A6, wohingegen in FrsG zwei MT-Domänen vorhanden sind. Das entspricht genau dem Methylierungsmuster von FR, welches ein einfach methyliertes N-Methyldehydroalanin, ein

einfach methyliertes N-Methylalanin, sowie ein zweifach methyliertes N,O-Dimethylthreonin enthält. Man kann also vermuten, dass die Methylgruppen nach der Aktivierung der Aminosäure eingefügt werden, während die Aminosäure am PCP gebunden vorliegt, wie es auch für andere NRPS bereits beschrieben wurde [Marahiel et al., 1997] [Süßmuth, 2017]. Eine weitere bioinformatische Analyse der MT-Domänen beinhaltet Kapitel 3.3.

Zwischen den beiden NRPS-Genen liegt *frsB*, welches für ein MbtH ähnliches Protein codiert. Es folgt *frsC*, ein Gen für eine putative Malat-Dehydrogenase, und mit dem endständigen FrsH das einzige modifizierende Enzym im gesamten BGC. FrsC weist bioinformatisch eine hohe Ähnlichkeit zu einer Malat-Dehydrogenase und L-Laktat-Dehydrogenase auf [Minarik et al., 2002] [Wright et al., 2000]. Daher könnte es die Reduktion von Phenylpyruvat zu Phenyllaktat katalysieren (siehe Abb. 13) [Crüsemann et al., 2018].

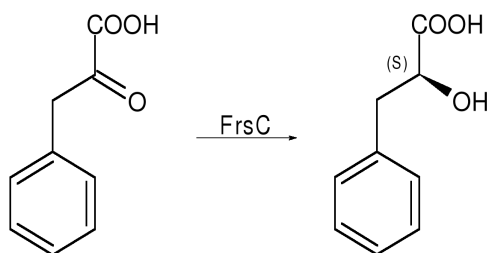


Abbildung 13: Hypothetische Reaktion von FrsC: FrsC katalysiert mutmaßlich die Reduktion von Phenylpyruvat zu Phenyllaktat.

FrsH weist bioinformatisch im aktiven Zentrum Ähnlichkeiten mit CmlA, einer eisenabhängigen β -Hydroxylase aus der Chloramphenicolbiosynthese, auf. Dort katalysiert CmlA die Hydroxylierung des PCP-gebundenen *para*-Aminophenylalanin [Makris et al., 2010]. Ein 3D-Modell von FrsH bestätigt die Ähnlichkeit mit CmlA [Crüsemann et al., 2018].

Die Analyse der A-Domänen durch antiSMASH zeigt, dass FrsA, FrsD und FrsG (Modul 7) Leucin oder Hydroxyleucin aktivieren. Das FR-Peptid ent-

hält Hydroxyleucin als Baustein. Die *in-vitro*-Analyse der A-Domänen (siehe Kap. 3.5) zeigt eine eindeutige Prävalenz der A-Domäne A1, A2 und A7 für Leucin. Daher muss vermutet werden, dass die Hydroxylierung nach der Aktivierung von Leucin am Leucin-PCP erfolgt. Hierfür wäre FrsH ein möglicher Kandidat (siehe Abb. 14).

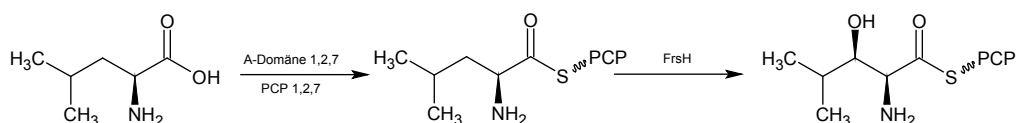


Abbildung 14: Hypothetische Reaktion von FrsH: FrsH weist bioinformatisch eine Ähnlichkeit mit einer β -Hydroxylase aus der Chloramphenicolbiosynthese auf. Dort katalysiert CmlA die Hydroxylierung des PCP-gebundenen *para*-Aminophenylalanin [Makris et al., 2010].

Vergleicht man die Ergebnisse der antiSMASH-Analyse mit der Struktur von FR, so zeigt sich, dass die große NRPS FrsD-FrsG mit hoher Wahrscheinlichkeit den 7-gliedrigen Zyklus von FR bildet, während FrsA höchstwahrscheinlich für die Bildung der Seitenkette und deren Verknüpfung mit dem zyklischen Teil von FR zuständig ist. Die Analyse bestätigt ebenfalls, dass beide BGC nahezu identisch sind, was in den nachfolgenden Kapiteln detaillierter beschrieben werden wird. Es kann davon ausgegangen werden, dass in beiden Fällen die Bildungseinheit für FR vorliegt.

3.1.2 Vergleichende Analyse der Gene des Biosynthesegenclusters aus *Cand. B. crenata* und *C. vaccinii*

Betrachtet man beide *frs*-BGC auf DNA- und Proteinebene, so sind sie hier ebenfalls auffallend ähnlich. Das *bcfrs*-BGC von *Cand. B. crenata* besteht aus 35.800 Basenpaaren, die Größe des *cvfrs*-BGC von *C. vaccinii* beträgt 35.851 Basenpaare. Beide unterscheiden sich damit hinsichtlich der Größe in nur 51 Basenpaaren. Auch beim Vergleich der einzelnen Gene (siehe Tabelle 3) sieht man, dass die Größen nahezu übereinstimmen. Die einzelnen Gene zeigen in Alignments (siehe A27-A44) sowohl auf DNA-Ebene als auch

auf Proteinebene zwischen 68 % und 85 % Identität (siehe Tabelle 3). Der GC-Gehalt der beiden BGC unterscheidet sich ebenfalls nur gering und liegt zwischen 44 % und 69 % (siehe Tabelle 3).

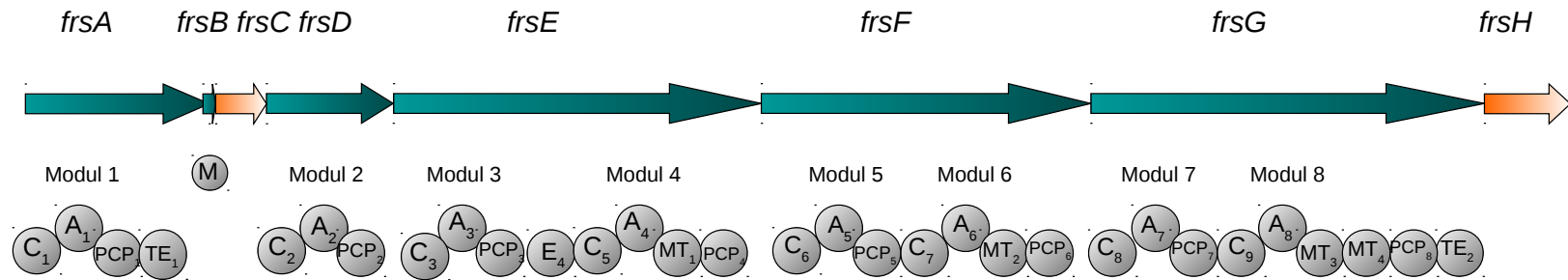
	FrsA	FrsB	FrsC	FrsD	FrsE	FrsF	FrsG	FrsH
DNA Burk	3768 bp	219 bp	987 bp	3078 bp	9048 bp	7560 bp	9411 bp	1599 bp
DNA Chro	3819 bp	219 bp	987 bp	3081 bp	9051 bp	7557 bp	9408 bp	1596 bp
Diff. bp.	51 bp	0 bp	0 bp	3 bp	3 bp	3 bp	3 bp	3 bp
Ident DNA [%]	70	72	68	70	70	73	72	77
GCBurk [%]	58	44	53	60	60	59	58	55
GCChro [%]	66	44	49	69	67	69	68	61
Prot Burk	1255 AS	72 AS	328 AS	1025 AS	3015 AS	2519 AS	3136 AS	532 AS
Prot Chro	1272 AS	72 AS	328 AS	1026 AS	3016 AS	2518 AS	3135 AS	531 AS
Ident Prot [%]	71	75	72	70	71	75	73	85

Tabelle 3: Vergleich der BGC von *Cand. B. crenata* und *C. vaccinii* auf DNA und Proteinebene. bp=Basenpaare, Diff. bp= Basenpaardifferenz der Gene beider BGC zueinander; AS=Aminosäure, Ident DNA=Identität der Gene von *C. vaccinii* und *Cand. B. crenata* zueinander in Prozent; Ident Prot=Identität von FrsA-FrsH von *C. vaccinii* und *Cand. B. crenata* auf Proteinebene zueinander in Prozent; GCBurk und GCChro=GC-Gehalt der Gene *frsA-frsH* von *C. vaccinii* und *Cand. B. crenata* in Prozent.

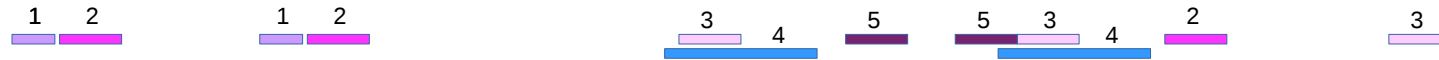
Die Größe der Gene unterscheidet sich in den NRPS-Genen nur um wenige Basenpaare, die des MLP-Gens und beider modifizierender Enzyme *frsC* und *frsH* sind sogar identisch. Die Identität der BGC auf DNA-Ebene beträgt bei den meisten Genen ca. 70 %. Mit 77 % Identität bildet das *frsH* hier eine Ausnahme vom Durchschnitt des BGC. *frsA* und *frsD-frsG* weisen bei *Cand. B. crenata* einen GC-Gehalt von ca. 60 % auf, während der GC-Gehalt von *C. vaccinii* hier mit ca. 70 % um 10 Prozentpunkte höher liegt. Der GC-Gehalt von *frsB* ist mit 44 % der niedrigste und bei beiden BGC identisch. Der GC-Gehalt der *frsH*-Gene liegt mit 55 % bei *Cand. B. crenata* und mit 61 % bei *C. vaccinii* niedriger als der GC-Gehalt der jeweiligen NRPS-Gene. Bei *frsC* liegen die GC-Gehalte von *Cand. B. crenata* mit 53 % und bei *C. vaccinii* mit 49 % nah beieinander. Hier hat, anders als im Rest des Genclusters, *Cand. B. crenata* einen leicht höheren GC-Gehalt als *C. vaccinii*.

3.1.3 Bioinformatische Analyse sich wiederholender Sequenzbereiche in den BGC von *Cand. B. crenata* und *C. vaccinii*

Sowohl das *bcfrs*-BGC aus *Cand. B. crenata*, als auch das *cvfrs*-BGC aus *C. vaccinii* zeigen Bereiche sich wiederholender Basen- bzw. Proteinsequenzen. Darunter versteht man Sequenzabschnitte, die in demselben BGC doppelt oder dreifach vorkommen und dies teilweise unabhängig von Domänen oder Gengrenzen. Um diese identischen Sequenzabschnitte genauer festzulegen, wurde im ganzen BGC nach identischen Aminosäuresequenzen gesucht. In Alignments wurden die Grenzen dieser Bereiche festgelegt (siehe Abb. 15, Alignments nicht gezeigt). Mit gleicher Farbe und mit gleicher Zahl markierte Bereiche sind innerhalb eines BGCs identisch.



Cand. Burkholderia crenata



Chromobacterium vaccinii

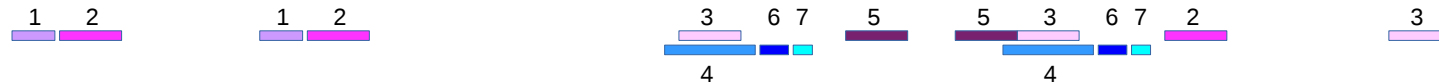


Abbildung 15: Der *frs*-BGC-Vergleich: In gleichen Farben dargestellte und gleich nummerierte Bereiche innerhalb eines BGCs kennzeichnen identische Sequenzabschnitte. Die Farbgebung wurde für beide BGC gleich gewählt. Beide BGC weisen sehr ähnliche sich wiederholende Sequenzabschnitte auf. Bei *C. vaccinii* sind Unterbrechungen zwischen den Bereichen 4, 6 und 7, während bei *Cand. B. crenata* im Übergang der Gene *frsE* zu *frsF* und *frsF* zu *frsG* durchgehend identische Bereiche sind.

In beiden *frs*-BGC wurden jeweils 5 Sequenzabschnitte gefunden, die häufiger als einmal vorkommen. Im *cvfrs*-BGC von *C. vaccinii* ist der Bereich Nr. 4 nicht durchgehend identisch und wurde daher in weitere Bereiche 6 und 7 unterteilt (siehe Abb. 15). Teilweise stimmen diese Sequenzabschnitte mit einzelnen Domänen überein, teilweise sind diese Bereiche auch über Domänen- oder Gengrenzen hinweg zu sehen. Tabellen 4 und 5 zeigen die Häufigkeit, mit der sich die Sequenzabschnitte wiederholen und über welchen Bereich sie sich im Gencluster erstrecken.

Bereich	bp	AS	Häufigkeit	Gene	abgedeckter funktionaler Enzymteil
1	1011 bp	337 AS	2	<i>frsA, frsD</i>	Starter C-Domäne
2	1632 bp	544 AS	3	<i>frsA, frsD frsG</i>	A-Domänen Leucin
3	1548 bp	516 AS	3	<i>frsE, frsF frsG</i>	MT-Domänen
4	1956 bp	652 AS	2	<i>frsE, frsF</i>	A-Domäne, MT-Domäne, PCP-Domäne
5	1467 bp	489 AS	2	<i>frsF</i>	letzter Teil d. C-Domäne und beide Alanin A-Domänen
6	630 bp	210 AS	2	<i>frsF frsG</i>	C-Domäne
7	351 bp	117 AS	2	<i>frsF frsG</i>	C-Domäne

Tabelle 4: Sequenzbereiche, die im *cvfrs*-BGC von *C. vaccinii* mehrfach vorkommen. Die Abkürzung bp zeigt die Länge der Bereiche in Basenpaaren an. AS zeigt die Länge des Bereichs in Aminosäuren. Bei Genen wurden alle Gene angegeben, die ganz oder zum Teil von diesen sich wiederholenden Bereichen abgedeckt werden.

52

Bereich	bp	AS	Häufigkeit	Gene	abgedeckter funktionaler Enzymteil
1	915 bp	305 AS	2	<i>frsA, frsD</i>	Starter C-Domäne
2	1695 bp	565 AS	3	<i>frsA, frsD, frsG</i>	A-Domänen Leucin
3	1578 bp	526 AS	3	<i>frsE, frsF frsG</i>	MT-Domänen
4	3070 bp	680/346 AS	2	<i>frsE, frsF, frsG</i>	Ende A-Domäne bis C-Domäne
5	1488 bp	496 AS	2	<i>frsF</i>	letzter Teil d. C-Domäne und beide Alanin A-Domänen

Tabelle 5: Sequenzbereiche, die im *bcfrs*-BGC von *Cand. B. crenata* mehrfach vorkommen. Die Abkürzung bp zeigt die Länge der Bereiche in Basenpaaren an. AS zeigt die Länge des Bereichs in Aminosäuren. Bei Genen wurden alle Gene angegeben, die ganz oder zum Teil von diesen sich wiederholenden Bereichen abgedeckt werden.

Die *frs*-BGC von *C. vaccinii* und *Cand. B. crenata* zeigen eine sehr ähnliche Beschaffenheit in Größe und Häufigkeit dieser sich wiederholenden Sequenzabschnitte. Eine Ausnahme bildet hier lediglich der bei *Cand. B. crenata* sehr große durchgängige Sequenzabschnitt 4, der bei *C. vaccinii* zweimal unterbrochen wird. Alle anderen innerhalb des jeweiligen BGC identischen Sequenzabschnitte unterscheiden sich bei beiden BGC kaum.

Sequenz 1 kodiert für einen großen Teil der Starter-C-Domäne sowohl von FrsA, als auch von FrsD und kommt damit zweimal in jedem BGC vor. Diese Sequenz grenzt direkt an den Abschnitt 2 an. Sequenz 2 kodiert für die A-Domänen, die Leucin aktivieren. Dieser Sequenzabschnitt kommt in beiden BGC dreimal vor und zwar in *frsA*, *frsD* und in Modul 7 von *frsG*. Sequenz 3 kodiert für die putativ N-methylierenden-Domänen und befindet sich in *frsE*, *frsF* und *frsG* und kommt ebenfalls dreimal im gesamten BGC vor. Sequenz 4 ist mit einer Größe von 3070 Basenpaaren bei *Cand. B. crenata* die mit Abstand größte zusammenhängende sich wiederholende Gensequenz. Sie kommt zweimal im *frs*-BGC von *Cand. B. crenata* vor und erstreckt sich jeweils vom letzten Teil der A-Domäne aus den Modulen 4 und 6 über die MT-Domäne, die PCP-Domäne und die C-Domäne des nächsten Gens (Modul 5 und 7). Sequenz 4 ist die einzige Sequenz, die über eine Gengrenze hinaus geht. Der Bereich kommt zweimal vor; einmal im Übergangsbereich von *frsE* zu *frsF* und im Übergangsbereich von *frsF* zu *frsG*. Bei *C. vaccinii* gibt es in dieser Region ebenfalls gleiche Sequenzabschnitte. Diese sind allerdings unterbrochen, einmal zwischen der PCP-Domäne der Module 4 und 6 und in der Mitte der C-Domäne 180 Basenpaare nach der Sequenz, die für das konservierte HHxxxDG Motiv kodiert. Der Sequenzabschnitt 5 kodiert für den C-terminalen Bereich der C-Domäne von FrsF und umfasst die A-Domänen bis kurz vor der Core-A9-Region (siehe Kap. 3.4). Die A-Subdomäne unterscheidet sich bei diesen zwei A-Domänen, sowohl bei *Cand. B. crenata* als auch bei *C. vaccinii* wieder. Die Bereiche 6 und 7 kommen nur bei *C. vaccinii* vor und entstehen dadurch, dass der Bereich der Sequenz 4, der bei *Cand. B. crenata* durchgehend ist, bei *C. vaccinii* unterbrochen ist. Sowohl im *bcfrs*-BGC von *Cand. B. crenata*, als auch im *cvfrs*-BGC von

C. vaccinii nehmen die sich wiederholenden Bereiche ca. 24 % des gesamten BGC ein.

3.2 Bioinformatische Analyse der in den *frs*-BGC von *Cand. B. crenata* und *C.vaccinii* kodierten A-Domänen

A-Domänen nehmen eine zentrale Stellung in der Biosynthese der nichtribosomalen Peptide ein. Durch die Spezifität jeder einzelnen A-Domäne wird die Reihenfolge der Aminosäuren im Peptid determiniert. Die Spezifität lässt sich bioinformatisch über den nichtribosomalen Code nach Stachelhaus vorhersagen (siehe auch Kap. 1.4) [Stachelhaus et al., 1999].

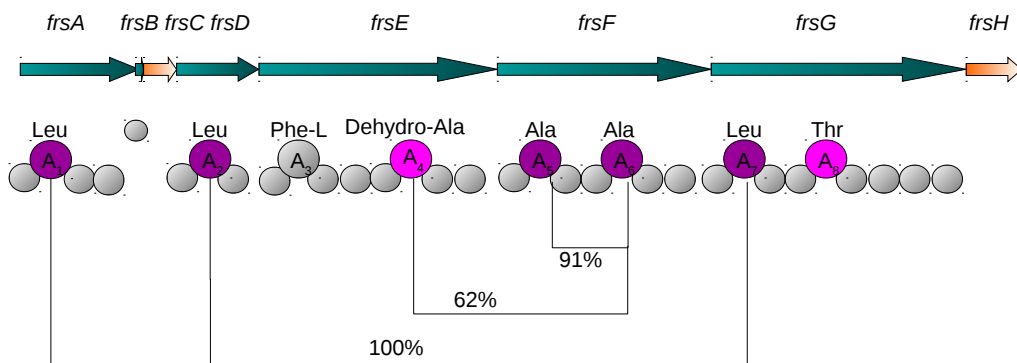


Abbildung 16: Vergleich der Identitäten der A-Domänen auf Proteinebene innerhalb des *cvfrs*-BGCs von *C. vaccinii*. Die Lokalisierung der A-Domänen ist für beide *frs*-BGC gleich. Die Identitäten in dieser Abbildung entsprechen denen von *C. vaccinii*.

Die A-Domänen wurden ebenfalls auf ihre Identitäten überprüft. Auch hier sind auffallend ähnliche Sequenzen zu finden. Die A-Domänen von FrsA (A1), FrsD (A2) und FrsG (A7) gleichen sich zu 100 %. Die beiden Alanin-aktivierenden Domänen von FrsF (A5 und A6) gleichen sich zu 91 % und die zweite Domäne von FrsE (A4), die putativ Serin aktiviert, ist ebenfalls noch zu 62 % mit der 2. A-Domäne von FrsF (A6) identisch. Alle anderen aus

FrsE1 (A3) und FrsG2 (A8) sind nur zu 30 % bis 55 % identisch miteinander (siehe Abb. 16, A8 und A10).

A-Domänen besitzen eine hochkonservierte dreidimensionale Struktur, und die Aminosäuren, die diese Struktur aufrecht erhalten, sind ebenfalls hochkonserviert. Die Substrate, die von A-Domänen selektiert und aktiviert werden, sind durch 10 Aminosäuren im aktiven Zentrum der A-Domäne festgelegt. Durch Alignments mit der A-Domäne von GrsA (siehe Kap. 1.4), für die die Aminosäuren des sogenannten nichtribosomalen Code erstmals definiert wurden, kann die Bindungstasche in anderen A-Domänen bestimmt werden. Die Aminosäurekombination in der Bindungstasche determiniert die Aminosäure, die im Rahmen der Biosynthese aktiviert werden soll.

Tabellen 6 und 7 zeigen den nichtribosomalen Code der in den *frs*-BGC kodierten A-Domänen von *Cand. B. crenata* und *C.vaccinii*. Die Codes in den beiden BGC unterscheiden sich in den fett gedruckten Aminosäuren.

A-Domäne	235	236	239	278	299	301	322	330	331	517
A1 Burk	D	A	M	L	V	G	A	V	C	K
A2 Burk	D	A	M	L	V	G	A	V	C	K
A3 Burk	G	A	F	V	M	A	G	V	C	K
A4 Burk	D	V	W	H	L	S	L	V	D	K
A5 Burk	D	V	F	S	V	A	I	V	Y	K
A6 Burk	D	V	F	S	V	A	I	V	Y	K
A7 Burk	D	A	M	L	V	G	A	V	C	K
A8 Burk	D	F	W	N	I	G	M	V	H	K

Tabelle 6: Nichtribosomaler Code der A-Domänen des *bcfrs*-BGC aus *Cand. B. crenata* für die A-Domänen A1-A8 (siehe auch Abb. 4, 16, A7 und A9).

A-Domäne	235	236	239	278	299	301	322	330	331	517
A1 Chro	D	A	M	L	V	G	A	A	C	K
A2 Chro	D	A	M	L	V	G	A	A	C	K
A3 Chro	G	A	F	V	M	A	G	V	C	K
A4 Chro	D	V	W	H	L	S	L	V	D	K
A5 Chro	D	V	F	S	V	A	I	V	Y	K
A6 Chro	D	V	F	S	V	A	I	V	Y	K
A7 Chro	D	A	M	L	V	G	A	A	C	K
A8 Chro	D	F	W	N	V	G	M	V	H	K

Tabelle 7: Nichtribosomaler Code der A-Domänen des *cvfrs*-BGC aus *C. vaccinii* für die A-Domänen A1-A8 (siehe auch Abb. 4, 16, A7 und A9).

Die Aminosäuren Asp-235 und Lys-517 zeigten im Modell von GrsA eine Bindung zur α -Aminogruppe und α -Carboxygruppe von Phenylalanin. Da diese funktionellen Gruppen in jeder Aminosäure gleich sind, unterscheiden sich die Codes an diesen Stellen bei den meisten A-Domänen kaum. Auch bei den A-Domänen der *frs*-BGC ist das Asp-235 und Lys-517 konserviert. Eine Ausnahme stellt hier die A-Domäne 3 dar, die in Position 235 anstatt einer Asparaginsäure ein Glycin enthält. Das lässt sich dadurch erklären, dass das putative Substrat der A-Domäne 3 eine Carbonsäure und keine Aminosäure ist und daher keine Aminogruppe über die Asparaginsäure koordiniert werden muss.

Das hochkonservierte Lys-517 von A-Domänen ist in der Core-A10-Region der A-Subdomäne zu finden (siehe Kap. 1.4). Bei vielen A-Domänen, auf die eine MT-Domäne folgt, befindet sich die MT-Domäne zwischen der Core-A8- und der Core-A9-Region [Labby et al., 2015]. Die Hauptdomäne und die Subdomäne werden also von der MT-Domäne geteilt. Da die *frs*-BGC bei den A-Domänen A4, A6, und A8 von MT-Domänen unterbrochen sind, musste für die Core-A9- und Core-A10-Region ein weiteres Alignment angefertigt werden, um die Position Lys-517 des nichtribosomalen Codes genau zu bestimmen.

Bei der Suche nach der Core-A10-Region in den BGC fiel auf, dass die Core-A10-Region neun und nicht acht mal im gesamten BGC vorhanden ist, wie es der Anzahl der A-Domänen entspricht. Bei einem Alignment dieser Sequenzbereiche (siehe Abb. 17 und 18) fiel auf, dass nicht nur die Core-A10-Region, sondern auch die Core-A9- und Core-A11-Region neunmal in beiden *frs*-BGC zu finden ist, bei den A-Domänen A1, A2, A3, A5 am Ende der A-Domäne, bei den A-Domänen A4, A6 jeweils nach der MT-Domäne. Im Modul der A-Domäne A8 sind zwei MT-Domänen vorhanden. Hier konnte die Core-A9, A10- und A11-Sequenz jeweils bei beiden MT-Domänen am Ende wiedergefunden werden (siehe Abb. 19).

```

GrsA Core9      QPYLCAYFVSEK--HIPLEQLRQFSSEELPTYMIPSYFIQLDKMPLTNSNGKIDRKQLPEPDLTFGMRVD--
FrsA Core9 Burk DTQLVGYYVTIRG--EVDGQALRRQVANWLP EYMPVPAVVLVLEELPRLPNGKLDHQALPAPEYTGKRYQRPR
FrsD Core9 Burk DTQLVGYYVTVRG--EVDGQALRRQVANWLP EYMPVPAVVLVLEELPRLPNGKLDHQALPAPEYTGKRYQRPR
FrsE1 Core9 Burk AKQLVGYYVVPKEGVMLEPRAMRRELAEHLTDHMVPAVLVELSALPRTPNGKLDRSALPAPVFTVTEGYREPR
FrsE2 Core9 Burk LKPLSGYVNNPA-NFEQFAAIRRYVGEQLPDYMPVPAALVLEGLPLTPNGKLDRRALPAPEFGTACYRAPG
FrsF1 Core9 Burk HKQLVGYYVLLDD-AAADGQMLRRFLSRHLP EYMPVSAVVVRAALPLTPNGKLDRKALPAPTFVSADGRAPR
FrsAF2 Core9 Burk LKPLSGYVNNPA-NFEQFAAIRRYVGEQLPDYMPVPAALVLEGLPLTPNGKLDRRALPAPEFGTACYRAPG
FrsG1 Core9 Burk DTQLVGYYVTVRG--EVDGQALRRQVANWLP EYMPVPAVVLVLEELPRLPNGKLNHQALPAPEYTGKRYQRPR
FrsG2 Core9 Burk AENTSLFWNCR---ADLLRDVRAANLRKRLPDYMQPNHMLLDLTFPLTPNGKLDRRALPAPEQAAMRVRDIE
FrsG3 Core9 Burk LKPLSGYVNNPA-NFEQFAAIRRYVGEQLPDYMPVPAALVLEGLPLTPNGKLDRRALPAPEFGTACYRAPG
      :                :*      . * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                Core A9                                Core A10 Core A11

```

Abbildung 17: Alignment der Core-A9- und Core-A10-Region aus *Cand. B. crenata*. Sequenzen FrsG2 und FrsG3 zeigen den Bereich der zwei Subdomänen der A-Domäne 8.

58

```

GrsA      EQPYLCAYFVS--EKHIPLEQLRQFSSEELPTYMIPSYFIQLDKMPLTNSNGKIDRKQLPEPDLT
FrsA Core9 Chro GDSQLVAYVAAV--GGVEGSELRRLAAGQLPEHMVPAAVVLESPLQLPNGKLDRKSLPAPEFG
FrsD Core9 Chro GDSQLVAYVAAV--GGVEGSELRRLAAGQLPEHMVPAAVVLESPLQLPNGKLDRKSLPAPEFG
FrsE1 Core9 Chro GVRQLVGYYVVAATEAKLDAQGLRRQLAEHLPEHMVPAALVELAALPRTPNGKFDRGALPVPTFA
FrsE2 Core9 Chro ALQPLSGYVNNPA-NFEQYAAIRRYALEQLPEYMPVPAIVLLDALPLTPNGKLDRRALPAPEFG
FrsF1 Core9 Chro GHKQLVGYYVVLNA-AAADAQALRRFLADQLPEHMVPAAMIGLDVLP LTPSGKLDRKALPAPDFA
FrsF2 Core9 Chro ALQPLSGYVNNPA-NFEQYAAIRRYALEQLPEYMPVPAIVLLDALPLTPNGKLDRRALPAPEFG
FrsG1 Core9 Chro GDSQLVAYVAAV--GGVEGSELRRLAAGQLPEHMVPAAVVLESPLQLPNGKLDRKSLPAPEFG
FrsG2 Core9 Chro KAEPAAVWRS---RSALLRDVQAELRSQLPDYMQPNHLVLLDAFPLTANGKLDRRALPTPEQA
FrsG3 Core9 Chro ALQPLSGYVNNPA-NFEQYAAIRRYALEQLPEYMPVPAIVLLDALPLTPNGKLDRRALPAPEFG
      :                ::      . * * * * * * * * * * * * * * * * * * * * * * * * * * *
                                Core A9                                Core A10 Core A11

```

Abbildung 18: Alignment der Core-A9- und Core-A10-Region aus *C. vaccinii*. Sequenzen FrsG2 und FrsG3 zeigen den Bereich der zwei Subdomänen der A-Domäne 8.

In den Alignments (siehe Abb. 17 und 18) wurden mit Hilfe der A-Domäne von GrsA die Core-Bereiche Core-A9 bis A11 festgelegt. Die lila markierten Bereiche zeigen die konservierten Bereiche. Das Lys-517 wurde gelb markiert. Die Sequenzen FrsG2 und FrsG3 zeigen die beiden Core-A9- und Core-A10-Bereiche der A-Domäne A8 aus FrsG. Das bedeutet, dass bei der A-Domäne A8 nicht nur die A-Domäne von den MT-Domänen unterbrochen ist, sondern die Core-A9- bis Core-A11-Bereiche für die A-Domäne A8 zweimal existieren, jeweils am Ende jeder MT-Domäne (siehe Abb. 19). Da diese Core-Regionen auf der A-Subdomäne lokalisiert sind, kann davon ausgegangen werden, dass für die A8-Domäne zwei Subdomänen existieren. Eine Modellierung ist aufgrund mangelnder Datenlage noch nicht möglich. Die A-Domäne mit zwei aufeinander folgenden MT-Domänen ist bisher einzigartig für die NRPS des FR und seiner Derivate.

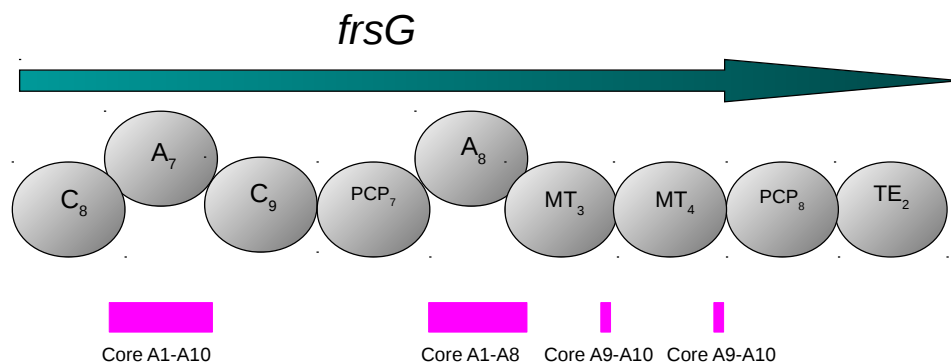


Abbildung 19: Die erste A-Domäne von FrsG aus Modul 7 ist eine einfache A-Domäne ohne MT-Domäne. Hier liegen die Core-A9- und Core-A10-Regionen unmittelbar nach der Core-A8-Region. Die A-Domäne ist nicht unterbrochen. Die zweite A-Domäne wird von zwei MT-Domänen unterbrochen. Die Core-A9 und Core-A10-Regionen, die sich in der A-Subdomäne befinden, existieren zweimal, jeweils am Ende jeder A-Domäne.

3.3 Bioinformatische Analyse der in den *frs*-BGC von *Cand. B. crenata* und *C.vaccinii* kodierten MT-Domänen

MT-Domänen sind für die Methylierung der von den A-Domänen aktivierten Substrate zuständig. Die MT-Domänen folgen unmittelbar auf die A-Domänen bzw. sind häufig zwischen der Core-A8- und Core-A9-Region der A-Domänen integriert (siehe Kap. 1.4).

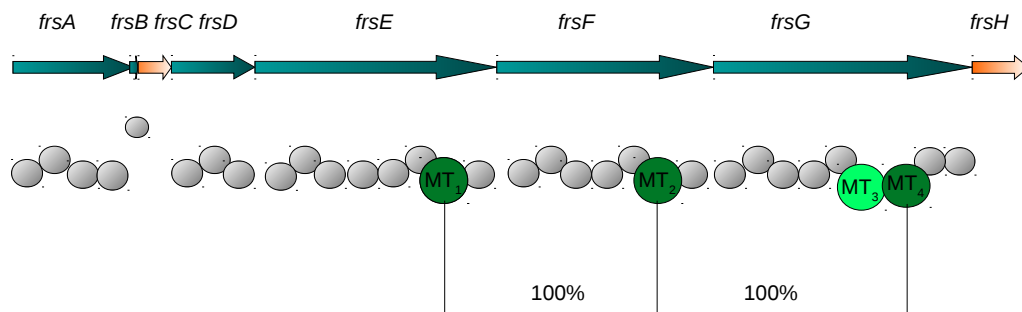


Abbildung 20: Vergleich der Identität der MT-Domänen innerhalb des BGCs von *C. vaccinii*.

Das *frs*-BGC enthält vier Gensequenzen, die für MT-Domänen kodieren. Die MT-Domänen 1, 2 und 4 von *C. vaccinii* sind in ihrer Proteinsequenz zu 100 % identisch, wohingegen die dritte MT-Domäne nur zu ca. 24-27 % Übereinstimmung zu den ersten aufweist (siehe Abb. 20 und A14). Die MT-Domänen, die durch das *bcfrs*-BGC von *Cand. B. crenata* kodiert werden, weisen ähnliche Identitäten auf (siehe A13).

Vergleicht man die MT-Domänen der *frs*-BGC untereinander, so liegt bei den drei MT-Domänen 1,2 und 4 die Übereinstimmung bei über 72 %, und bei den zwei stark abweichenden MT-Domänen MT 3 besteht untereinander auch eine Identität von 73 %. Das entspricht ungefähr der durchschnittlichen Identität der Proteinsequenzen, die durch die *frs*-BGC kodiert werden (siehe Abb. 21 und Tab. 3).

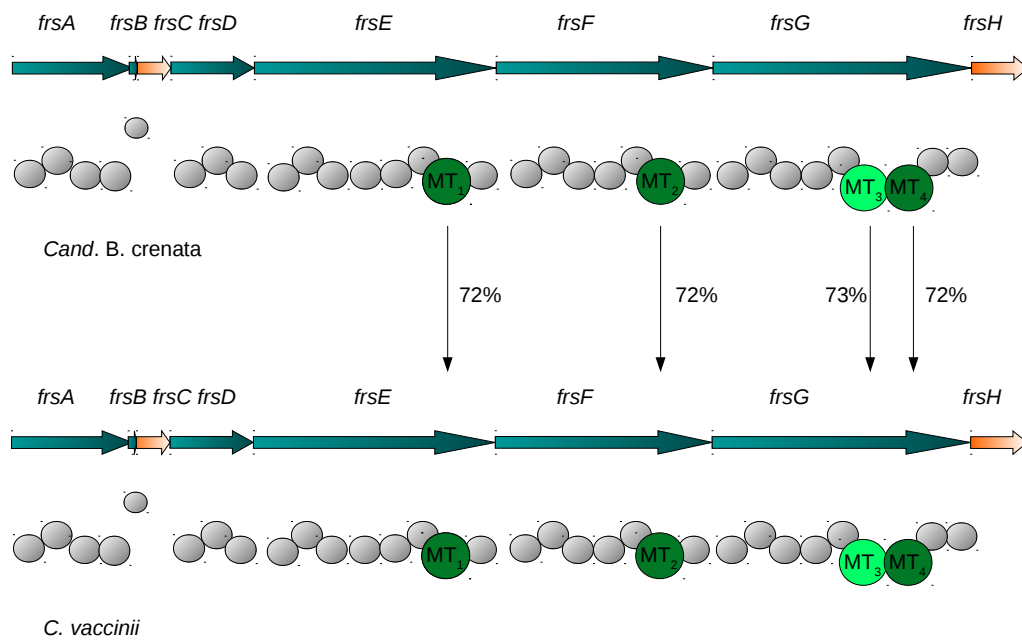


Abbildung 21: MT-Domänen im Vergleich der *frs*-BGC aus *Cand. B. crenata* und *C. vaccinii*.

Das Depsipeptid FR ist an vier Stellen methyliert. Dies betrifft den Baustein des N-Methyldehydroalanin, welcher vom Modul 4 in FrsE bereitgestellt wird, das N-Methylalanin des Moduls 4 aus FrsF und das zweifach methylierte N,O-Dimethylthreonin aus FrsG. Die erste MT-Domäne aus FrsG MT 3 unterscheidet sich in beiden BGC sehr stark von den anderen MT-Domänen MT 1,2 und 4. Letztere enthalten die Konsensussequenz VLEIGVGSGL im BGC von *Cand. B. crenata* und LLEIGVGSGL im BGC von *C. vaccinii*. Diese Sequenz ist für die Bindung von S-Adenosylmethionin (SAM) zuständig [Labby et al., 2015]. Es handelt sich hierbei höchstwahrscheinlich um die drei N-Methylierenden Domänen. Die erste MT-Domäne der A-Domäne A8 weicht mit 20 % Übereinstimmung mit den anderen drei MT-Domänen stark von den anderen Domänen ab. Es kann daher vermutet werden, dass es sich hierbei um die O-Methylierende Domäne handelt.

3.4 Bioinformatische Analyse der in den *frs*-BGC von *Cand. B. crenata* und *C. vaccinii* kodierten Kondensierungs-Domänen

C-Domänen katalysieren die Kondensationsreaktion der einzelnen NRP-Bausteine. Das *frs*-BGC kodiert insgesamt 9 C-Domänen, davon eine E-Domäne, die phylogenetisch mit den C-Domänen verwandt ist. Die C-Domänen C1 und C2 sind zu 91 % identisch, während zwischen C6 und C8 eine Identität von 86 % besteht. Alle übrigen C-Domänen (C3, C5, C7, C9) weisen lediglich eine Identität zwischen 53 % und 62 % auf (siehe Abb. 22 und A20). Die Identitäten der C-Domänen von *Cand. B. crenata* liegen in einem ähnlichen Bereich (siehe Abb. A22).

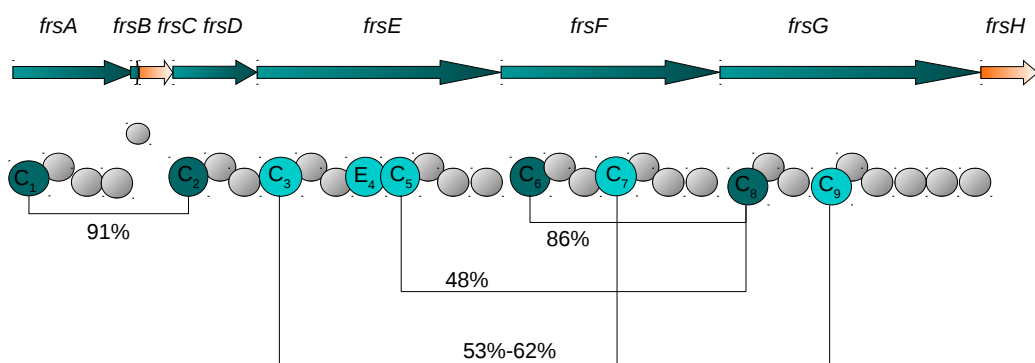


Abbildung 22: Vergleich der Identitäten der C-Domänen, welche im *cvfrs*-BGC von *C. vaccinii* kodiert sind.

Für die phylogenetische Analyse der C-Domänen aus *Cand. B. crenata* [Crüsemann et al., 2018] wurden bekannte C-Domänensequenzen aus der NaPDoS-Datenbank [Ziemert et al., 2012] mit den *befrs*-C-Domänensequenzen aligniert. Der phylogenetische Baum (siehe Abb. A1) wurde mit Hilfe von FastTree (Version: 2.1.7) erstellt [Price et al., 2009] (siehe Kap. 5.1). Die C-Domänen wurden entsprechend ihrer Typenzuordnung [Rausch et al., 2007] in unterschiedlichen Farben dargestellt.

Die meisten der durch die Analyse der C-Domänen vorhergesagten Eigen-

schaften [Rausch et al., 2007] korrelierten mit der Biosynthese-Hypothese. Die drei als ${}^L\text{C}_L$ ausgewiesenen C-Domänen in den Modulen 3, 6 und 8 katalysieren die Kondensationsreaktion zwischen zwei L-Aminosäuren. Die Epimerase-Domäne und D-Phenyllaktat-akzeptierende Domäne in Modul 4 entsprechen ebenfalls dem postulierten Biosynthesemechanismus. Auch die C-Domänen von FrsA und FrsD korrelieren wie erwartet mit der Klade der Starter-C-Domänen.

Die C-Domänen der Module 5 und 7 (C6 und C8) sind phylogenetisch mit den ${}^D\text{C}_L$ -Domänen verwandt, was der erwarteten Biosynthese nicht entspricht. Die C-Domäne C6 verknüpft das L-konfigurierte N-Methyl-Dehydroalanin mit einem L-konfigurierten Alanin, und die C8 verknüpft ein L-konfiguriertes Alanin mit L-Hydroxyleucin. ${}^D\text{C}_L$ -C-Domänen verknüpfen aber normalerweise eine D- mit einer L-Aminosäure. Eine dieser C-Domänen könnte an der Bildung von Dehydroalanin aus Serin beteiligt sein, vor allem da die *frs*-BGC kein weiteres Enzym enthalten, das diese Reaktion katalysieren könnte.

Bei den *frs*-BGC käme für eine solche Reaktion insbesondere die C-Domäne C6 in Frage. Mechanismen, bei denen eine C-Domäne für eine Dehydratisierung eines Serins zuständig sein könnte, wurden bereits in mehreren Fällen beschrieben. Für die Biosynthese von Microcystin wurde postuliert, dass eine Dehydratisierung über eine C-Domäne möglich wäre [Tillett et al., 2000] [Crüsemann et al., 2018], was aber bis jetzt noch nicht belegt werden konnte. Die phylogenetische Analyse der C-Domäne McyA aus dem BGC von Microcystin ergab, dass McyA eine eigene Klade mit anderen C-Domänen bildet, die möglicherweise für Dehydratisierungen verantwortlich sind. Die C6-Domäne der *frs*-BGC korreliert aber mit den ${}^D\text{C}_L$ -Domänen und nicht mit den dehydratisierenden C-Domänen. Ein Alignment (siehe Abb. A6) der McyA-C-Domäne mit den C6-Domänen der BGC ergab auch keinen Hinweis auf einen ähnlichen Mechanismus.

Ein anderer Dehydratisierungsmechanismus ist für die NRPS des Nocardicin beschrieben, welches eine ${}^L\text{C}_L$ -C-Domäne besitzt, die die Dehydratisierung

von Serin zu Dehydroalanin, einem Intermediat der β -Lactam Biosynthese, katalysiert [Gaudelli et al., 2015]. Das bedeutet, dass C_{DH}-Kondensierungsdomänen nicht die einzig möglichen Domänen für die Bildung von dehydratisierten Aminosäuren in NRPS sind. Ebenso wurden bei vielen *Pseudomonas Arten* keine C_{DH}-Domänen gefunden, trotz dehydratisierter Aminosäuren in den Produkten. Der Mechanismus der Dehydratisierung des Nocardicins in der Nocardicin-Biosynthese beruht wahrscheinlich auf einem Histidin-Rest direkt vor der Konsensussequenz HHxxxGL [Gaudelli et al., 2015]. Dieses Histidin ist in unserer C-Domäne ebenfalls nicht vorhanden. Es muss also einen anderen zugrundeliegenden Mechanismus geben.

Ein interessanter Hinweis ist, dass die C-Domänen C6 und C8 der beiden *frs*-BGC eine Identität von 85 % zueinander besitzen und damit über der durchschnittlichen Identität der C-Domänen innerhalb desselben BGCs liegen (siehe Abb. 22, A20 und A22). Beide haben ähnliche Substrate, nämlich N-Methyl-Serin für die C-Domäne C6 und N-Methylalanin für die C-Domäne C8, wobei C8 höchstwahrscheinlich keine Dehydratisierung durchführt. Möglich wäre, dass die C-Domäne C8 eine ähnliche Funktion hatte, die im Laufe der Evolution verloren gegangen ist.

Um herauszufinden, welche Unterschiede jeweils zwischen den C-Domänen C6 und C8 in beiden BGC bestehen, wurden innerhalb eines BGCs die C-Domäne C6 gegen C8 aligniert (siehe Abb. 23 und 24). Die Alignments zeigen, dass die C-Domänen im N-terminalen Bereich fast identisch sind, wohingegen sie sich im C-terminale Bereich deutlich unterscheiden. Das Alignment der *cvfrs*-Aminosäuresequenz von *C. vaccinii* (siehe Abb. 24) zeigt im mittleren Bereich des Proteins Abweichungen, die aber auch im Laufe der Evolution entstanden sein könnten. Trotz allem ist auch hier die Sequenz vor und nach diesem Bereich der Abweichung wieder identisch.


```

C.v. FrsF C6 DCL      MNPKKIEAAYPLSPLQKGF LFHAGYDLQSADSYVAQLFLDFEGELDGAAMRSAADALMRR
C.v. FrsG C8 DCL      MNPKKIEAAYPLSPLQKGF LFHAGYDLQSADSYVAQLFLDFEGELDGAAMRSAADALMRR
*****

C.v. FrsF C6 DCL      HANLRAGFVHPGGQEPVQVVLREVVAGWEERDWRGRNPQEAAEAQSAWQEADRERRFELS
C.v. FrsG C8 DCL      HANLRAGFVHPGGQEPVQVVLREVVAGWEERDWRGRNPQEAAEAQSAWQEADRERRFELS
*****

C.v. FrsF C6 DCL      QPPLLRFGWLR L PAGRSQLVVTYHHILLDGWSLPL LLEELLALYRAGGGDGLPEATPYS
C.v. FrsG C8 DCL      QPPLLRFGWLR L PAGRSQLVVTYHHILLDGWSLPL LLEELLALYRAGGGDGLPEATPYS
*****

C.v. FrsF C6 DCL      AYLGWLQERDRAAAACEAWGGYLEGLEGPTL LAQEGQADKGA AAAQARLSLELPAELTQAL
C.v. FrsG C8 DCL      AYLGWLQERDRAAAACEAWGGYLEGLEGPTL VSTGHPQDH - - AEQKQRAWRLPAELTQAL
*****:; * * * : . *****

C.v. FrsF C6 DCL      TRQARQQGVT LNTLLQAAWMLL GKLNL SRDVFVFGITVAGRPGELPGVERMIGLFINTVP
C.v. FrsG C8 DCL      TRQARQQGVT LNTLLQAAWMLL GKLNL SRDVFVFGITVAGRPGELPGVERMIGLFINTVP
*****

C.v. FrsF C6 DCL      VRLRWSAGETVAGLVGRLQREQAGLLDHQHLDLVEIQRLAGQRQLFDTL FIFENY PFD SQ
C.v. FrsG C8 DCL      VRLRWSAGETVAGLVGRLQREQAGLLDHQHLDLVEIQRLAGQRQLFDTL CVFENY PVDAA
*****:; ***** *

C.v. FrsF C6 DCL      AMAPDLGQASLRVSGG E Q E SHYPL T LMAVP E EMLSLYLSYDAQRFDKGVVQGLLTRFR
C.v. FrsG C8 DCL      AMEQPSGGLQLRGVSGG Y MTHYPL S L M I E P G PELKLN L T Y QPERFEPEAIERLSAQLT
** * . * * * * : : : * * * * * * * * * * * * * * * * * * * * * * * * * * *
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

C.v. FrsF C6 DCL      LLL EAVAAEPSRPVSDI ELLDEAERR
C.v. FrsG C8 DCL      RLLGVIAAEPSPVSDI E L-----
** . : *****

```

Abbildung 24: Alignment der C-Domänen C6 und C8 aus *C. vaccinii*. Vergleicht man dieses Alignment mit dem Alignment der C-Domänen C6 und C8 aus *Cand. B. crenata* (siehe Abb. 23), so stellt man fest, dass sich die C-Domänen C6 und C8 in beiden Alignments im vorderen Teil sehr ähnlich sind und im hinteren Bereich voneinander abweichen. Im Gegensatz zu *Cand. B. crenata* gibt es bei *C. vaccinii* einen Bereich in der Mitte der C-Domänen, in dem die Proteinsequenzen voneinander abweichen (siehe hellgrüner Bereich). Die farblich markierten Aminosäuren entsprechen den Aminosäuren, die sich bei *C. vaccinii* und *Cand. B. crenata* auf die gleiche Art unterscheiden. Unpolare Aminosäuren wurden gelb und polare ungeladene Aminosäuren hellblau gekennzeichnet. Positiv geladene Aminosäuren wurden grün markiert und negativ geladene Aminosäuren wurden dunkelblau gekennzeichnet.

Die große Abweichung in der Mitte der Sequenz von *C. vaccinii* (hellgrüner Bereich des Alignments Abb. 24) und die vereinzelt unterschiedlichen Aminosäuren bei *Cand. B. crenata* (siehe Abb. 23) scheinen zufällig zu sein, da sie jeweils nur bei *C. vaccinii* oder bei *Cand. B. crenata* vorkommen. Diese Proteinsequenzen sind daher für die Dehydroalaninbildung eher unwahrscheinliche

Kandidaten. Im hinteren Bereich der Alignments zeigen die Proteinsequenzen Abweichungen, die sich in 13 Positionen sowohl bei *C.vaccinii* als auch bei *Cand. B. crenata* in den gleichen Aminosäuren unterscheiden. Sowohl die Anzahl, die Position, als auch die Art der abweichenden Aminosäuren stimmen vollständig überein. Die Aminosäuren wurden im Alignment entsprechend ihrer chemischen Eigenschaften farbig dargestellt (siehe Abb. 23 und 24). Es ist sehr unwahrscheinlich, dass dieser Unterschied auf einer Zufälligkeit basiert. Diese 13 Aminosäuren könnten eine Rolle in der Bildung von Dehydroalanin aus Serin spielen.

Um zu überprüfen, ob diese Aminosäuren sich im aktiven Zentrum der C-Domäne befinden, wurde ein 3D-Modell mit „Phyre 2“ erstellt (siehe Abb. 25-27). Man erkennt das im blauen Bereich liegende konservierte HHxxxDG Motiv (siehe Abb. 25 und 27), das für die Verknüpfung von 2 Aminosäuren relevant ist [Marahiel et al., 1997]. In unmittelbarer Nähe im gelben Bereich befinden sich die 13 Aminosäuren, in denen sich die Sequenzen von der C6- und der C8-Domäne jeweils unterscheiden. Die Modelle zeigen, dass sich diese Aminosäuren am oberen Rand der V-förmigen Tasche der C-Domäne befinden (siehe Abb. 25-27).

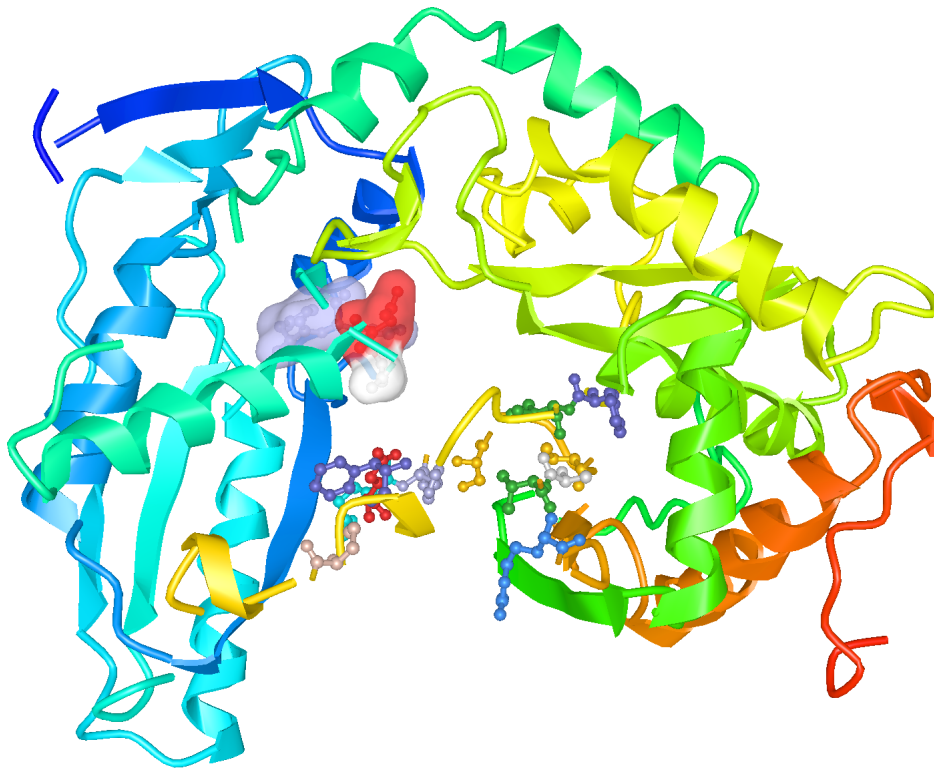


Abbildung 25: 3D-Modell der C6-Domäne von *C. vaccinii*, erstellt mit „Phyre 2“ ; seitliche Ansicht auf den Tunnel, der durch die beiden Subdomänen entsteht. Die Aminosäuresequenz der C-Domäne ist in Spektralfärbung dargestellt. Der blaue Bereich markiert den N-Terminus, der rote Bereich den C-terminalen Teil des Peptids. Die Aminosäuren des konservierten HDxxxDG-Motivs sind als Van-der-Waals Modell dargestellt, die putativ für die Dehydratisierung von Serin verantwortlichen Aminosäuren als „sticks-and-ball“-Modell. Bei einer seitlichen Ansicht auf die V-förmige Tasche, erkennt man, dass sich die Aminosäuren am oberen Rand der Tasche befinden.

Der blaue Bereich liegt am N-terminalen Ende, der rote Bereich am C-terminalen Ende der C-Domäne. Die Aminosäuren, die putativ an der Dehydratisierung des Serins beteiligt sein könnten, liegen in der Kluft zwischen den beiden Subdomänen und sind als Stäbchen markiert. Sie befinden sich fast alle in einer Ebene am oberen Rand der Tasche. Links oben durch van-der-Waals-Kräfte dargestellte Aminosäuren gehören zum konservierten HHxxxDG-Motiv (siehe Abb. 25). Eine Beschriftung wurde aufgrund der Übersichtlichkeit in dieser Perspektive weggelassen (siehe auch Abb. 26).

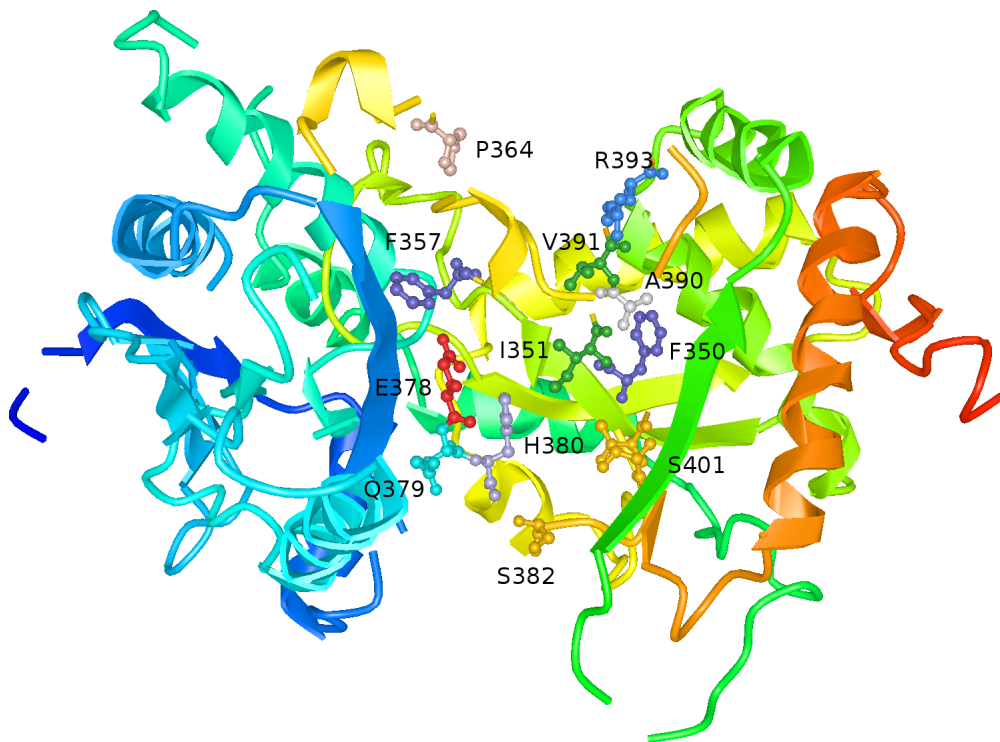


Abbildung 26: 3D-Modell der C6-Domäne von *C. vaccinii*, erstellt mit „Phyre 2“ ; Untersicht auf den Tunnel, der durch die beiden Subdomänen entsteht. Die Aminosäuresequenz der C-Domäne ist in Spektralfärbung dargestellt. Der blaue Bereich markiert den N-Terminus, der rote Bereich den C-terminalen Teil des Peptids. Die Aminosäuren des konservierten HDxxxDG-Motivs sind nicht dargestellt, die putativ für die Dehydratisierung von Serin verantwortlichen Aminosäuren sind als “sticks-and-ball“ -Modell dargestellt und mit Positionsangaben beschriftet.

Abbildung 26 zeigt die C6-Domäne von *C. vaccinii* in der Aufsicht auf die Kluft zwischen den beiden Subdomänen. Man sieht, dass die betreffenden 13 Aminosäuren über die komplette Länge des Tunnels verteilt sind. Abbildung 27 zeigt das konservierte HHxxxDG Motiv im Hintergrund.

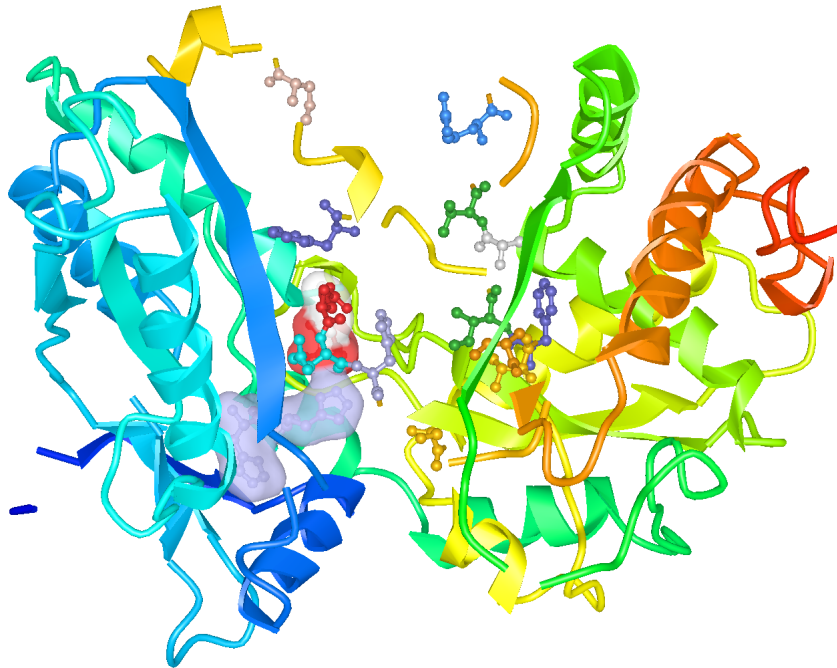


Abbildung 27: 3D-Modell der C6-Domäne von *C. vaccinii*, erstellt mit „Phyre 2“ ; Unteransicht auf den Tunnel, der durch die beiden Subdomänen entsteht. Die Aminosäuresequenz der C-Domäne ist in Spektralfärbung dargestellt. Der blaue Bereich markiert den N-Terminus, der rote Bereich den C-terminalen Teil des Peptids. Die Aminosäuren des konservierten HDxxxDG-Motivs sind im Hintergrund durch Van-der-Waals-Darstellung gekennzeichnet, die putativ für die Dehydratisierung von Serin verantwortlichen Aminosäuren sind als “sticks-and-ball“ -Modell dargestellt.

Der Dehydratisierungsmechanismus des Nocardicins [Gaudelli et al., 2015] beruht vermutlich auf einem Histidinrest, der direkt an der Dehydratisierung beteiligt ist. Im Fall der FR-Biosynthese befindet sich bei C6 ebenfalls ein Histidinrest und zwar in Position H380, der in C8 nicht vorkommt. Es könnte sich hierbei also um einen wichtigen Teil des aktiven Zentrum der dehydratisierenden C-Domäne handeln.

3.4.1 Bioinformatische Analyse der PCP-Domänen und TE-Domänen

PCP-Domänen haben in erster Linie eine Transportfunktion. Die an den Phosphopantetheinrest gebundenen Aminosäuren und Peptide werden über

die PCP-Domänen zu den katalytischen Zentren der anderen Domänen transportiert (siehe Kap. 1.7).

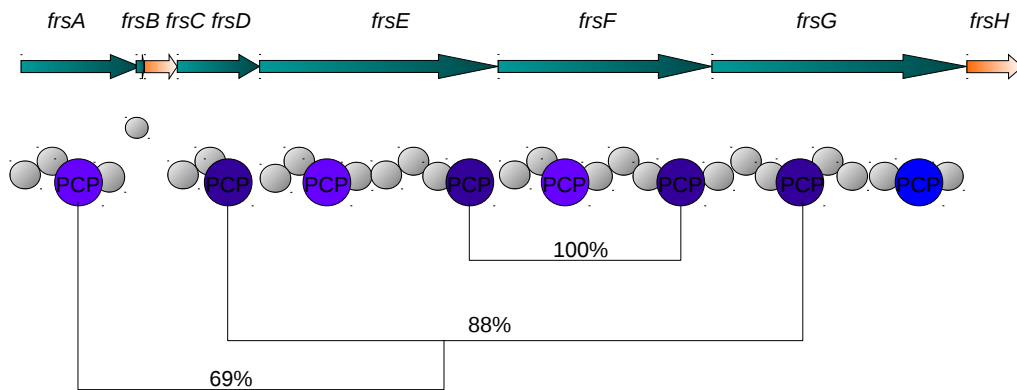


Abbildung 28: PCP-Domänen des *cvfrs*-BGCs. Die PCP-Domänen wurden blau markiert. PCP-Domänen, die sich stärker gleichen in dunkelblau, die anderen etwas heller. Die Prozente geben die prozentuale Identität der Domänen untereinander an.

Die endständigen PCP-Domänen von FrsE und FrsF sind zu 100 % identisch. Beide PCPs befinden sich nach einer A-Domäne, die ein N-Methyliertes Alanin bzw. Dehydroalanin eingefügt hat. Die PCP-Domänen von FrsD und FrsG1 weisen immerhin noch eine Identität von 88 % zueinander auf und folgen auf eine A-Domäne, die in der größeren NRPS die Hydroxyleucine eingefügt haben. Die PCP von FrsA zu FrsD und FrsG sind noch zu ca. 70 % identisch. Alle anderen gleichen sich noch zwischen 40 % und 70 % (siehe Abb. 28).

TE-Domänen stehen am Ende von NRPS. Sie sind für die Freisetzung und eventuelle Zyklisierung des Peptids verantwortlich (siehe Kap. 1.9).

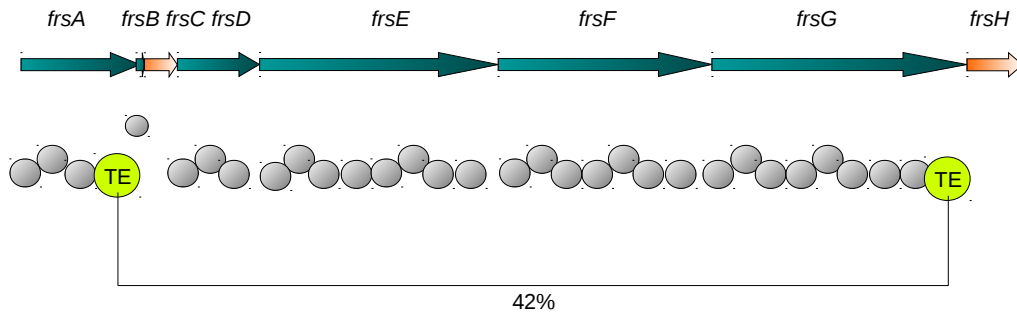


Abbildung 29: TE-Domänen des *cvfrs*-BGCs.

Von den TE-Domänen sind jeweils zwei in jedem BGC vorhanden. Die TE-Domänen sind nur zu 42% identisch und gleichen sich damit am wenigsten (siehe Abb. 29).

3.5 *In-vitro*-Untersuchungen von A-Domänen

3.5.1 Grundlagen des γ - $^{18}\text{O}_4$ -ATP-Austausch-Assay nach Phelan et al. für die A-Domänen von FrsA und FrsD

Die bioinformatische Analyse der A-Domänen mittels antiSMASH ergab für die A-Domänen für FrsA und FrsD eine Spezifität für Leucin oder OH-Leucin. Um die Spezifität der A-Domänen *in vitro* zu differenzieren, wurde für die entsprechenden A-Domänen aus dem *frs*-BGC in *Cand. B. crenata* der γ - $^{18}\text{O}_4$ -ATP-Austausch-Assay nach Phelan [Phelan et al., 2009] durchgeführt. Der Assay basiert auf der Adenylierungsreaktion der A-Domäne. Hierbei wird ein Substrat, welches häufig eine Aminosäure ist, von der A-Domäne aktiviert. Dabei reagiert die Carboxylgruppe der Aminosäure mit der Phosphatgruppe eines Adenosin-Triphosphats (ATP) unter Abspaltung von Pyrophosphat (PPI) zu einem Aminoacyladenylat (siehe Abb. 30).

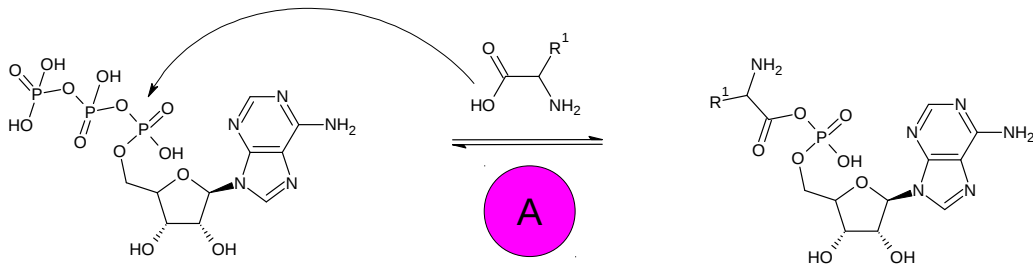


Abbildung 30: Die Adenylierungsreaktion von A-Domänen erfolgt unter ATP-Verbrauch. Diese Reaktion ist reversibel.

Die so aktivierte Aminosäure kann dann in einem weiteren Schritt auf den Phosphopantethein-Arm der PCP-Domäne geladen werden, um weitere Modifikationen zu unterlaufen und mit der entstehenden Peptidkette verknüpft zu werden (siehe Abb. 31 und Kap. 1.4).

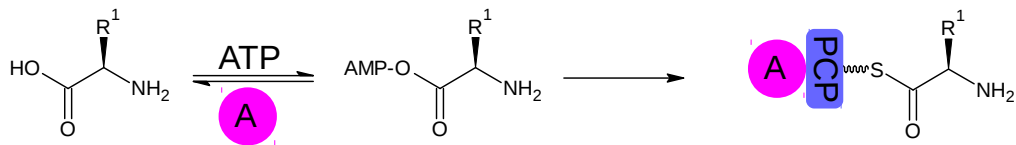


Abbildung 31: Die Aminosäure wird mit Hilfe der A-Domäne unter Verbrauch von ATP zum Aminoacyladenylat aktiviert, um im Anschluss auf den Phosphantethin-Arm der PCP-Domäne geladen zu werden.

Diese Adenylierungsreaktion ist reversibel. Auf der Rückreaktion des Aminoacyladenylates zu ATP und Aminosäure basiert der γ - $^{18}\text{O}_4$ -ATP-Austausch-Assay. Im Assay wird γ - $^{18}\text{O}_4$ -ATP verwendet, welches in der endständigen Phosphatgruppe vier Mal das schwerere $^{18}\text{O}_4$ -Isotop des Sauerstoffes enthält (siehe auch Abb. 32). Dieses wird während der Adenylierungsreaktion abgespalten. Das so entstandene Adenosin Monophosphat (AMP) wird im Anschluss größtenteils mit $^{16}\text{O}_4$ -PPI regeneriert, da dieses dem Reaktionsgemisch im Überschuss zugegeben wurde.

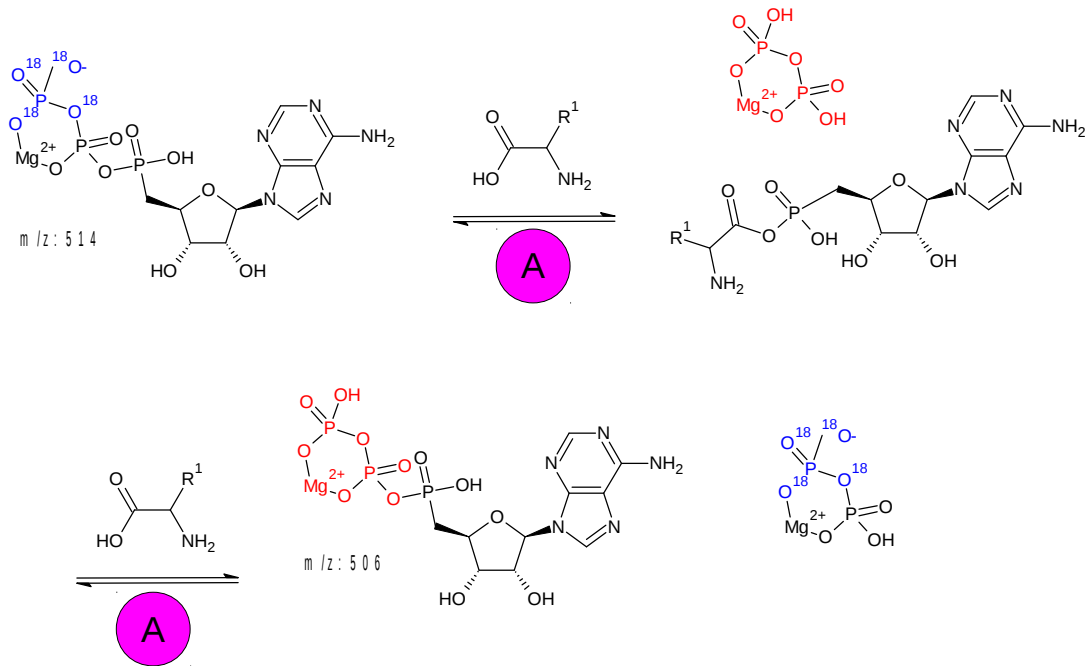


Abbildung 32: Austauschreaktion des A-Domänen Assay. Nach Abspaltung der Aminosäure vom AMP wird aus AMP und nicht markiertem PPI ein ATP gebildet, das kein ¹⁸O₄-Isotop mehr enthält.

Mit Hilfe einer Detektion über MALDI-TOF-MS (matrix assisted laser desorption/ionisation time-of-flight mass spectrometry) (siehe Kap. 5.6.9) lässt sich diese Umwandlung des schweren γ -¹⁸O₄-ATP zu dem leichteren ¹⁶O₄-ATP in einer Massenverschiebung nachweisen (siehe Abb. 32).

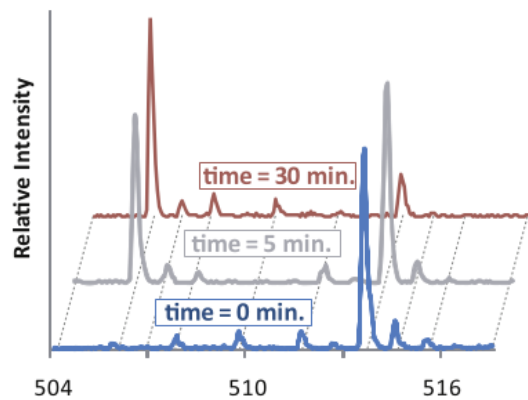


Abbildung 33: Allgemeine Darstellung einer MALDI-TOF-Analyse des Reaktionsproduktes einer A-Domäne, γ - $^{18}\text{O}_4$ -ATP, $^{16}\text{O}_4$ -ATP und der entsprechenden Aminosäure in Abhängigkeit von der Zeit. Bei der Regenerierung von AMP zu ATP wird $^{16}\text{O}_4$ -PPI verwendet. Es kommt zur Bildung des ATP mit einem $^{16}\text{O}_4$ -Isotop-PPI und in der Folge zu einer Detektion des m/z : 506 Peaks anstatt des m/z : 514 Peaks des γ - $^{18}\text{O}_4$ -ATP. Die Abbildung wurde aus Phelan et al. 2009 entnommen.

Für jede Aminosäure, auf die die A-Domäne getestet werden soll, wird ein Reaktionsansatz hergestellt (siehe Kap. 5.6.8), der die Aminosäure, γ - $^{18}\text{O}_4$ -ATP und PPI im Überschuss enthält und bei 25 °C inkubiert. Lässt sich im Anschluss $^{16}\text{O}_4$ -ATP als Peak bei m/z : 506 nachweisen, war die A-Domäne in der Lage, das Substrat zu aktivieren (siehe Abb. 33). Somit kann die Spezifität einer A-Domäne nachgewiesen werden.

Die A-Domänen aus FrsA und FrsD wurden neben anderen gängigen proteinogenen Aminosäuren mit den Aminosäuren Leucin, Hydroxyleucin und Isoleucin getestet, um die vermutete Spezifität für Leucin zu bestätigen und die Toleranz der A-Domäne gegenüber leucinähnlichen Aminosäuren zu untersuchen (siehe Abb. 34). Die bioinformatische Analyse gab sowohl Leucin als auch Hydroxyleucin als mögliche Substrate an.

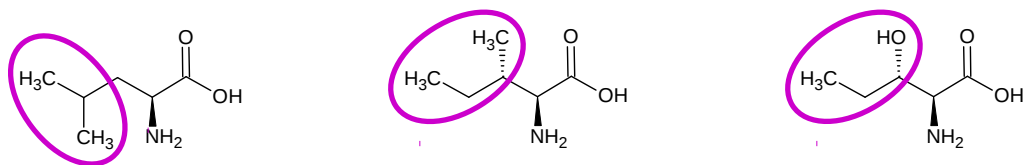


Abbildung 34: Strukturen von Leucin, Isoleucin und Hydroxyisoleucin. Um die Toleranz der A-Domänen in FrsA und FrsD zu testen, wurden u.a. diese strukturell ähnlichen Aminosäuren getestet.

3.5.2 Klonierung und Expression der A-Domänen FrsA und FrsD

Um die A-Domänen in funktionierender Form zu erhalten, wurden sie kloniert und in *E. coli* exprimiert. Dazu wurden Primersequenzen (siehe Tabellen 8 und 26) mit Clone Manager (siehe Kap 5.1) erstellt. Die Primer wurden so gewählt, dass sie sowohl die A-Domäne als auch die PCP-Domäne umfassen.

A-Domäne	Primer forwärts	Primer rückwärts	Länge [bp]	Enzym I	Enzym II
A1	Ac Adom 1 for.	Ac Adom 1 rev.	1671	BamHI	HindIII
A2	Ac Adom 2 for.	Ac Adom 2 rev.	1634	NotI	XhoI
MLP	MbtH for	MbtH rev	228	NdeI	XhoI

Tabelle 8: Verwendete Primer mit den dazugehörigen Schnittstellen.

Sowohl die PCR-Fragmente, die mit diesen Primern erstellt wurden, als auch die entsprechenden Vektoren wurden mit den für die Schnittstellen ausgewählten Enzymen (siehe Tabelle 8) wenn nötig über Nacht geschnitten und in einem molaren Verhältnis von 1:3 pET28a:PCR-Fragment über Nacht ligiert (siehe Kap. 5.5.11). Gleiches Vorgehen wurde für das MLP FrsB genutzt. Allerdings wurde hier der Coexpressionsvektor pCDFDuet verwendet. *E. coli* B121 (DE3) star Zellen wurden einmal nur mit dem Plasmid der jeweiligen A-PCP-Domäne und einmal mit der A-PCP-Domäne und dem MLP transformiert, um die entsprechenden Proteine zu exprimieren (siehe auch Kap.

5.5.12).

Dazu wurden TB-Medium Schikanekolben aus Übernachtskulturen angeimpft und bis zu einer OD_{600} von 1,2 kultiviert. Danach wurden die Kulturen auf Eis gekühlt und im Anschluss mit IPTG induziert. Die Kultur wurde über Nacht exprimiert. Am nächsten Tag wurden die Kulturen abzentrifugiert, mit Lysepuffer versetzt und mit Ultraschall-Impulsen aufgeschlossen. Zwischen den Pulsintervallen wurde die Kultur auf Eis gestellt, um ein Überhitzen zu verhindern. Die Zellsuspension wurde abzentrifugiert und der Überstand mit Ni-NTA-Agarose versetzt und auf Eis inkubiert. Das Pellet wurde für die SDS-Analyse aufbewahrt. Die Suspension wurde auf eine Polypropylen-Säule (Quiagen) gegeben und mit unterschiedlich konzentrierten Waschpuffern gewaschen. Anschließend wurde das Protein mit einem Elutionspuffer von der Säule gewaschen (siehe auch Kap. 5.6.4).

Fraktion	A1	A1 mit MLP	Fraktion	A2	A2 mit MLP
W I	15,30	10,44	W I	5,31	9,9
W II	1,24	0,63	W II	0,36	0,36
E I	0,89	0,58	E I	0,24	0,71
E II	0,24	0,20	E II	0,08	0,23
E III	0,34	0,10	E III	0,09	0,22
konz.	3	3,22	konz.	1,7	5,2

Tabelle 9: Proteinkonzentrationen der Wasch- und Elutionsfraktionen der exprimierten A-Domänen in [mg/mL]. Die Proben wurden auf einem Nanodrop-Gerät der Firma Thermo Scientific gemessen. WI und WII sind Waschschritte mit jeweils 25 mM und 50 mM Imidazolpuffer. EI, EII, EIII sind die Elutionen mit einem 300 mM Imidazolpuffer. Konz. entspricht der mit der Vivaspin 500 Säule aufkonzentrierten Fraktion.

Die Elutionsfraktionen wurden mit Hilfe einer PD10-Säule (GE) umgepuffert und entsalzt. Dazu wurde zunächst die Säule mit Assaypuffer gewaschen. Das Eluat wurden auf die Säule gegeben und im Anschluss mit Assaypuffer (Lösung 3; siehe auch Kap. 5.6.4) eluiert. Das Eluat wurde mit einer Vivaspin

500 Säule 30 MWCO von Satorius aufkonzentriert (siehe auch Kap. 5.6.4).

Von jeder Fraktion (Tabelle 9) wurden Proben für die SDS-Polyacrylamidgелеlektrophorese-Analytik (SDS-PAGE) entnommen, wobei das SDS-Gel mit Tris/Tricin-Puffern erstellt wurde (siehe Kap. 5.3.5) [Schägger, 2006].

3.5.3 SDS-PAGE-Analytik der A-Domänen von FrsA und FrsD

Von jeder Fraktion (siehe Tab. 9) wurden Proben für die SDS-PAGE-Analytik entnommen.

Protein	kDa
A1 mit Histag	63
A2 mit Histag	63
MLP	8

Tabelle 10: Zu erwartende Größe der exprimierten Proteine

Die MLPs bei 8 kDa konnten nicht eindeutig identifiziert werden. Die folgenden Gele zeigen die jeweilige A-Domänen auf der Höhe zwischen 60 kDa und 70 kDa (siehe Abb. 35 und 36). Das entspricht der erwarteten Größe (siehe Tab. 10), und die A-Domänen wurden für eine weitere Funktionsanalyse verwendet.

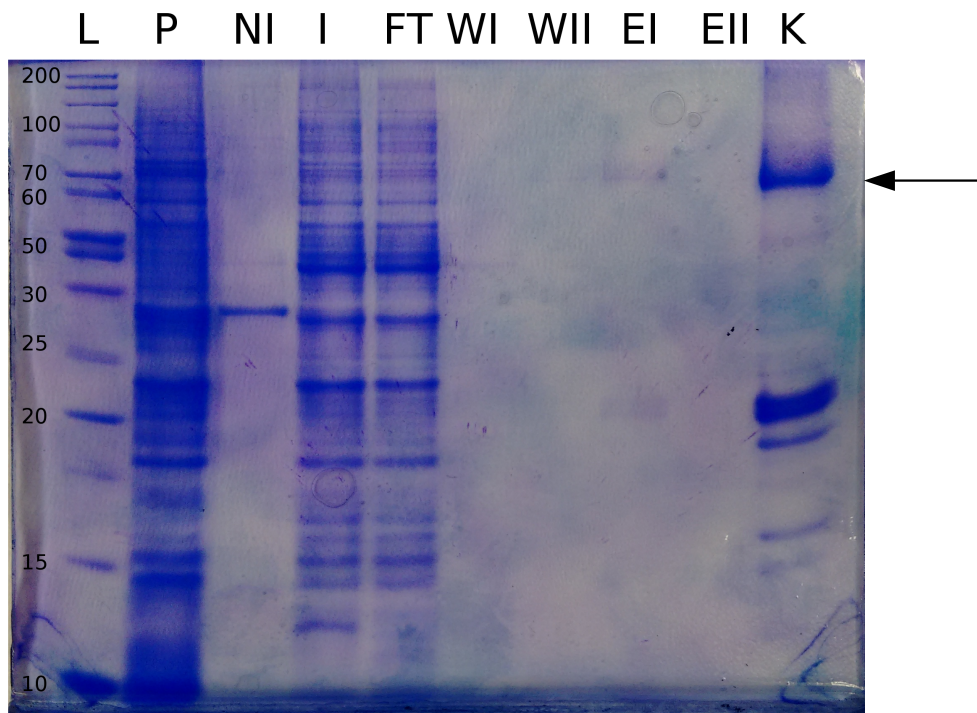


Abbildung 35: FrsA (A1-PCP) zusammen mit MLP exprimiert: Der Pfeil markiert die Position der A-Domäne. Das MLP kann nicht identifiziert werden. L=Ladder, P=Pellet, NI= nicht induzierte Fraktion, I= induzierte Fraktion vor Zugabe von Ni-NTA, FT= Filtrat nach Ni-NTA Entfernung, WI=Waschschritt I, WII= Waschschritt II, EI+EII=Elutionsfraktionen, K=vereinigte, aufkonzentrierte Elutionsfraktionen

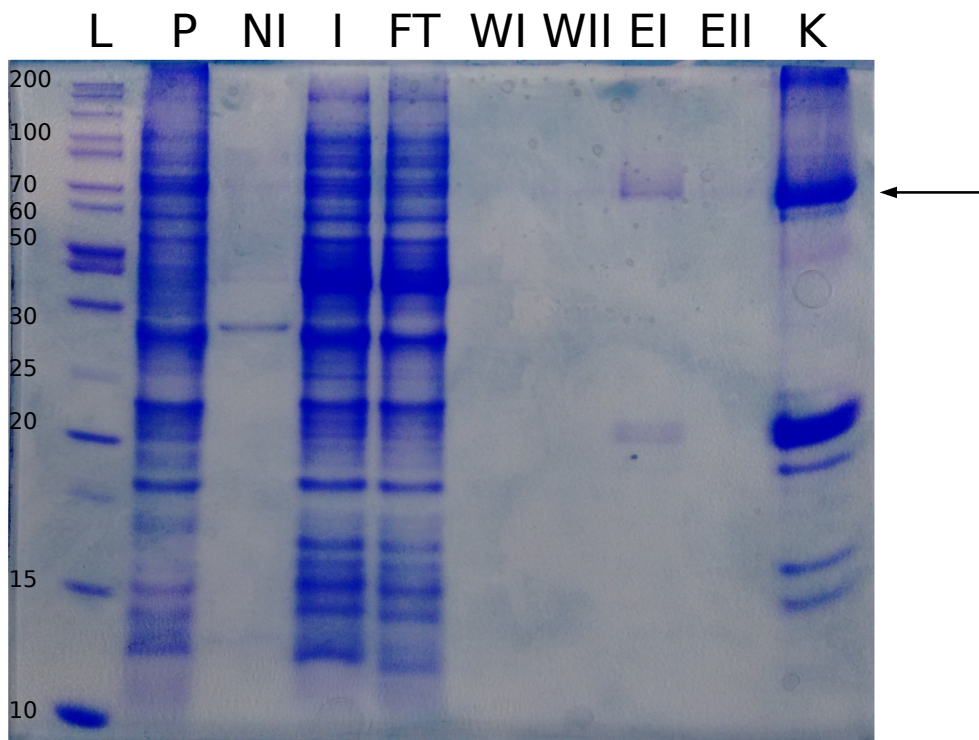


Abbildung 36: FrsD (A2-PCP) zusammen mit MLP exprimiert: Der Pfeil markiert die Position der A-Domäne. Das MLP kann nicht identifiziert werden. L=Ladder, P=Pellet, NI= nicht induzierte Fraktion, I= induzierte Fraktion vor Zugabe von Ni-NTA, FT=Filtrat nach Ni-NTA Entfernung, WI=Waschschrift I, WII= Waschschrift II, EI+EII=Elutionsfraktionen, K=vereinigte, aufkonzentrierte Elutionsfraktionen

Die A-Domänen konnten in ausreichender Menge exprimiert und eindeutig bestimmt werden. Das MLP konnte auf den Gelen nicht eindeutig identifiziert werden. In der aufkonzentrierten Fraktion sieht man Verunreinigungen, die trotz mehrmaligen Waschens nicht beseitigt werden konnten.

3.5.4 Durchführung des γ - $^{18}\text{O}_4$ -ATP-Assay

Die aus der vorangegangenen Expression erhaltenen A-Domänen wurden mit dem gut etablierten γ - $^{18}\text{O}_4$ -ATP-Assay nach Phelan [Phelan et al., 2009] auf ihre Funktion und Spezifität überprüft. Für den Assay wurden drei Lösungen hergestellt (siehe Kap. 5.3.4). Lösung 1 enthält die zu testende Aminosäu-

re (siehe Tab. 11 und 12), PPi und Tris-Puffer. Lösung 2 besteht aus γ - $^{18}\text{O}_4$ -ATP und Tris Puffer, während die dritte Lösung Tris, Glycerol und 1 mM DTT enthält. Von der 2. und 3. Lösung wurden jeweils 2 μL auf PCR-Tubes verteilt. Im Anschluss wurde jedes der PCR-Tubes mit 2 μL einer Aminosäure-Lösungen (Lösung 1) versetzt. Die Reaktionsansätze wurden inkubiert und im Anschluss mit einer Aminoacridin-Lösung in Aceton versetzt, welches später als Matrix für die MALDI-TOF-MS Messung diente. Im Anschluss wurden die Proben mithilfe einer MALDI-TOF-Messung analysiert (siehe auch Kap. 5.6.9).

3.5.5 Detektion und Auswertung des γ - $^{18}\text{O}_4$ -ATP-Austauschs mittels MALDI-TOF-MS

Sowohl für die A-Domäne A1 aus FrsA als auch für die A-Domäne A2 aus FrsD wurde bioinformatisch Leucin oder OH-Leucin als Substrat vorhergesagt. Diese Vorhersage sollte durch den Assay genauer bestimmt und eine Abhängigkeit der Aktivierung vom MLP überprüft werden. Des weiteren sollte eine eventuelle Substrattoleranz der A-Domänen gegenüber leucinähnlichen Aminosäuren betrachtet werden.

Im Folgenden sollen exemplarisch die MS-Spektren der 4 Reaktionsansätze der A-PCP-Domänen von FrsA und FrsD mit Leucin, sowohl mit als auch ohne MLP betrachtet werden. Die Spektren (siehe Abb. 37) zeigen in allen vier Reaktionsansätzen eine Aktivierung von Leucin und die damit verbundene Detektion des Massesignals bei m/z : 506. Das bedeutet, dass sowohl die A-Domäne A1 als auch die A-Domäne A2 Leucin unabhängig von einem MLP aktivieren können.

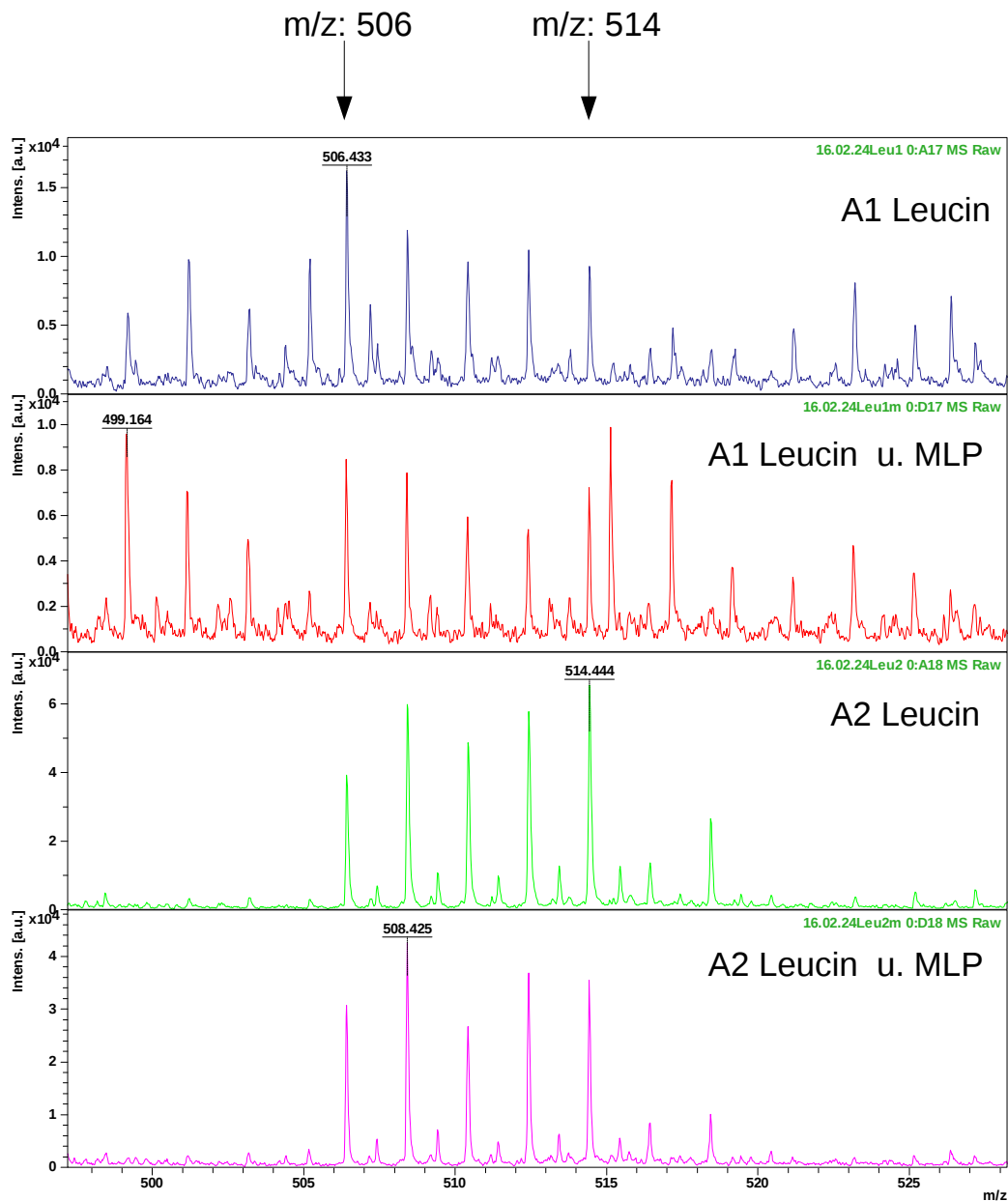


Abbildung 37: Massenspektrum des A-Domänen-Assay von FrsA (A1-PCP) und FrsD (A2-PCP), einmal mit und ohne MLP. Bei allen Reaktionsansätzen konnte das Massesignal des ¹⁶O₄-ATP-Peak bei *m/z*: 506 detektiert werden. In allen Fällen fand eine Aktivierung von Leucin statt. Beide A-Domänen sind ohne MLP funktionsfähig.

Die prozentuale Aktivierung aller verwendeter Aminosäuren (siehe Tabelle 11 und 12) wurde berechnet, um das Ausmaß der Aktivierung von Leucin zu bestimmen und eine Substrattoleranz der A-PCP-Domänen zu ermitteln. Um das Ausmaß der Rückreaktion des AMP zu $^{16}\text{O}_4\text{-ATP}$ zu berechnen und damit auch das Ausmaß der Aktivierungsreaktion, wurden die Integrale der im Spektrum vorhandenen Peaks aller ATP-Species verwendet (siehe auch Kap. 5.6.9). Das molare Verhältnis von $\gamma\text{-}^{18}\text{O}_4\text{-ATP}$ zu PPI beträgt 5:1. Daher wird bei der Auswertung davon ausgegangen, dass bei einer gemessenen Aktivierung von 83,33 % eine tatsächliche Aktivierung von 100 % stattgefunden hat. Für die prozentuale Aktivierung der Aminosäuren wurde das Verhältnis von $\gamma\text{-}^{18}\text{O}_4\text{-ATP}$ zur Summe aller gemessenen ATP-Peaks gebildet. Für die Berechnung wurden die jeweiligen Flächen unter den Kurven verwendet, die als Rohdaten aus der Messung entnommen wurden. Nachfolgende Formel diente zur Berechnung des tatsächlichen Austauschs von $\gamma\text{-}^{18}\text{O}_4\text{-ATP}$ zu $^{16}\text{O}_4\text{-ATP}$ [Phelan et al., 2009].

$$\text{Austausch [\%]} = (100/0.833) * ^{16}\text{O} / (^{18}\text{O} + ^{16}\text{O})$$

Aminosäure	A1 ohne MLP	A1 mit MLP	A2 ohne MLP	A2 mit MLP
Alanin	k. A.	8,3	0	0
Cystein	0	0	0	2,9
Glutaminsäure	0	2,3	0	0
Glycin	0	1,6	0	0
Histidin	0	0	0	2,4
Leucin	46,0	44,7	17,7	23,3
Isoleucin	28,1	19,8	13,6	21,1
Hydroxyleucin	9,6	0	0	7,5
Lysin	0	0	0	2,9
Methionin	0	0	0	2,4
Asparagin	k. A.	0	0	2,5
Prolin	0	0	0	0
Phenyllaktat	17,2	0.	0	13,4
Glutamin	0	2,6	0	3,2
Arginin	0	4,5	0	2,5
Serin	3,7	0	0	0
Threonin	0	k. A.	6,6	3,3
Tryptophan	0	0	3,2	4,5
Tyrosin	0	0	0	2,9

Tabelle 11: Berechneter absoluter γ - $^{18}\text{O}_4$ -ATP-Austausch durch verschiedene Aminosäuren im A-Domänen-Assay für FrsA (A1-PCP) und FrsD (A2-PCP). k.A.=es konnte keine Auswertung gemacht werden; 0=keine Aktivierung vorhanden

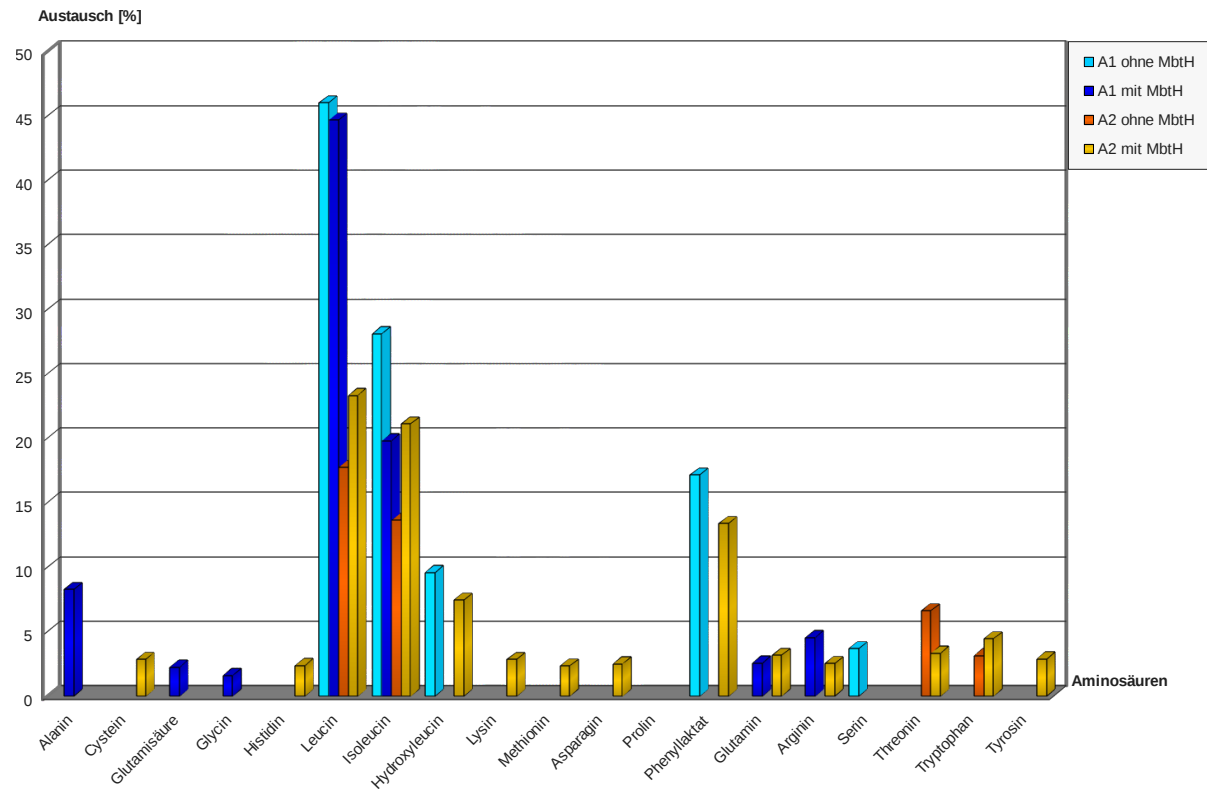


Abbildung 38: tatsächlicher prozentualer γ - $^{18}\text{O}_4$ -ATP-Austausch durch die A-Domänen von FrsA (A1-PCP) und FrsD (A2-PCP) in Anwesenheit der einzelnen Aminosäuren.

Die Auswertung der Assayreihe zeigt, dass sowohl die A1-PCP-Domäne, als auch die A2-PCP-Domäne in erster Linie für Leucin spezifisch sind (siehe Abb. 38 und 39). Es lässt sich ebenfalls eine geringe Substratspezifität für Isoleucin feststellen. Hydroxyleucin wird nicht in größerem Ausmaß aktiviert als der Rest der durchschnittlich gering aktivierten Aminosäuren. Das lässt sich durch die ähnliche Struktur von Leucin und Isoleucin erklären. Hydroxyleucin unterscheidet sich hingegen durch die OH-Gruppe sehr stark von den beiden Anderen (siehe Abb. 34).

Um zu bestimmen, in welchem Verhältnis die A-Domänen aus FrsA und FrsD eine Substrattoleranz gegenüber nicht nativen Substraten haben, wurde die prozentuale Aktivierung von Aminosäuren im Verhältnis zu Leucin (100 %) berechnet.

Aminosäure	A1 ohne MLP	A1 mit MLP	A2 ohne MLP	A2 mit MLP
Alanin	k. A.	18,5	0	0
Cystein	0	0	0	12,2
Glutaminsäure	0	5,1	0	0
Glycin	0	3,6	0	0
Histidin	0	0	0	10,2
Leucin	100	100	100	100
Isoleucin	61,1	44,3	76,9	90,5
Hydroxyleucin	20,9	0	0	32,0
Lysin	0	0	0	12,2
Methionin	0	0	0	10,1
Asparagin	k. A.	0	0	10,8
Prolin	0	0	0	0
Phenyllaktat	37,3	0	0	57,4
Glutamin	0	5,7	0	13,6
Arginin	0	10,2	0	10,9
Serin	8,1	0	0	0
Threonin	0	k. A.	37,3	14,3
Tryptophan	0	0	17,8	19,1
Tyrosin	0	0	0	12,4

Tabelle 12: Prozentualer γ - $^{18}\text{O}_4$ -ATP-Austausch durch die Aminosäuren im A-Domänen-Assay in Relation zur Leucin-Aktivierung von FrsA (A1-PCP) und FrsD (A2-PCP). k.A.=es konnte keine Auswertung gemacht werden; 0=keine Aktivierung vorhanden

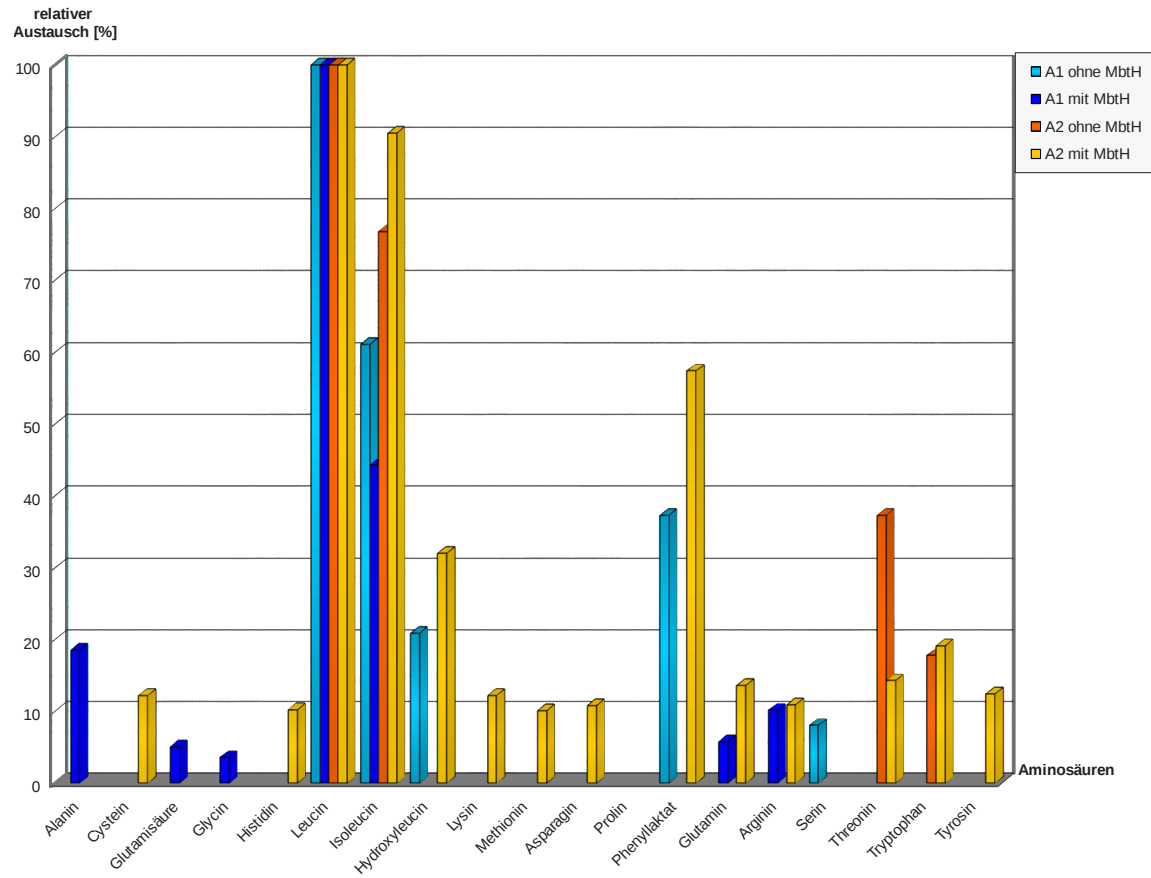


Abbildung 39: Prozentualer γ - $^{18}\text{O}_4$ -ATP-Austausch durch die A-Domänen FrsA (A1-PCP) und FrsD (A2-PCP) in Anwesenheit der einzelnen Aminosäuren bezogen auf Leucin.

Der γ - $^{18}\text{O}_4$ -ATP-Austausch-Assay präzisiert die vorhergehende bioinformatische Analyse. Die A-Domänen FrsA und FrsD aktivieren Leucin. Das Hydroxyleucin wird demnach nach der Aktivierung des Leucins gebildet. Der Assay zeigt ebenfalls, dass sowohl die A-Domänen aus FrsA, als auch FrsD ohne MLP in der Lage sind, Leucin zu aktivieren.

4 Diskussion

4.1 Pharmazeutische Relevanz von FR

Das Depsipeptid FR ist strukturell sehr ungewöhnlich und aufgrund seiner Eigenschaft, selektiv G_q-Proteine zu inhibieren, ein pharmakologisch wertvolles Werkzeug zur Erforschung G_q-Protein vermittelten Signalkaskaden [Schrage et al., 2015]. G-Proteine regulieren eine Vielzahl von Signalwegen in eukaryotischen Zellen und spielen eine zentrale Rolle bei vielen Erkrankungen, wie Asthma, Bluthochdruck, Krebs und vielen anderen [Deshpande and Penn, 2006] [Takefuji et al., 2012] [Dorsam and Gutkind, 2007]. Es gibt bis heute sehr wenig zuverlässige Substanzen, die G-Proteine modulieren können, membrangängig sind und dabei therapeutisches Potential haben [Smrcka, 2013]. FR und YM sind neben PTX die zur Zeit am meisten genutzten G-Protein Modulatoren. Lediglich YM und FR sind mögliche Therapeutika. YM war lange Zeit nicht kommerziell verfügbar und konnte nicht zu Forschungszwecken eingesetzt werden. Mittlerweile ist YM wieder im Handel, aber es gibt neue, noch nicht veröffentlichte Erkenntnisse aus der Forschergruppe FOR2372, die darauf hindeuten, dass sich YM und FR pharmakologisch unterscheiden und keineswegs gegeneinander austauschbar sind.

4.2 Hypothese zur Biosynthese von FR

Durch die vorliegende Arbeit konnte eine plausible Hypothese zur Biosynthese von FR erstellt werden (siehe Abb. 4). FR wird über eine nicht ribosomale Peptidsynthese gebildet. Das entsprechende *frs*-BGC kodiert für zwei TE-Domänen und besteht somit aus zwei NRPS, die beide an der Bildung von FR beteiligt sind. Die TE₂-Domäne der C-terminalen NRPS ist vermutlich für die Zyklisierung und Freisetzung des zyklischen Teils des Peptids verantwortlich. Die TE₁-Domäne der kleineren N-terminalen NRPS ist mutmaßlich für die intermolekulare Übertragung der Seitenkette zuständig. Ein Zusammenspiel zweier NRPS-Systeme, die voneinander durch eine TE-Domäne getrennt sind, ist selten. Eine ähnliche Konstellation kommt z.B. noch bei der Salinamid Biosynthese vor [Tan and Ma, 2008] [Ray et al., 2016]. Diese An-

ordnung der NRPS setzt eine gute Kommunikation zwischen den zwei NRPS voraus. Um diese Mechanismen besser zu verstehen, sind vor allem weitere funktionelle und kinetische Studien erforderlich.

Das *frs*-BGC enthält zwei modifizierende Enzyme, FrsC und FrsH, sowie FrsB, ein MLP. FrsC besitzt eine hohe Sequenzähnlichkeit mit einer Malat- bzw. L-Laktatdehydrogenase. FrsC ist daher ein möglicher Kandidat, der die Reduktion der Ketogruppe des Phenylpyruvates zu Phenyllaktat bewerkstelligen könnte. FrsH hat keine große Ähnlichkeit mit anderen Proteinen, es enthält aber dasselbe konservierte Motiv, wie CmlA, eine eisenabhängige β -Hydroxylase und könnte deshalb für die Übertragung von Hydroxy-Gruppen auf die Leucin-Bausteine verantwortlich sein [Crüsemann et al., 2018]. Damit sind alle Funktionen modifizierender Enzyme plausibel in die Biosynthesehypothese eingeordnet.

Bioinformatische und *in-vitro*-Untersuchungen dieser Arbeit dienten dazu, ein tieferes Verständnis der Biosynthese von FR zu erhalten. So konnten durch die antiSMASH-Analyse die Spezifitäten der A-Domänen begrenzt vorhergesagt werden. Eine *in-vitro*-Analyse der A-Domänen A1 und A2 aus FrsA und FrsD konnte schließlich die eindeutige Spezifität für Leucin beweisen. Diese detaillierten Einblicke in die FR-Biosynthese machen es künftig möglich, die Struktur von FR durch genetische Modifikation zu verändern. Dies ermöglicht es, ggf. pharmakodynamische Eigenschaften abzuändern, die Pharmakokinetik zu modifizieren, oder durch Einführen von funktionellen Gruppen nachfolgend semisynthetische Varianten herzustellen, oder FR mit Fluoreszenzmarkern zu versehen.

FR ist aus zwei unterschiedlichen natürlichen Quellen bekannt. Ursprünglich wurde es aus der tropischen Pflanze *A. crenata* isoliert, deren Endosymbiont *Cand. B. crenata* in der Lage ist, FR zu produzieren [Fujioka M., 1988] [Carrier et al., 2016]. Im Verlauf dieser Arbeit konnte *C. vaccinii* [Voing et al., 2015] als zweiter Produzent ausgemacht und dessen *cvfrs*-BGC vollständig erhalten werden. Durch Betrachtung sowohl der einzelnen BGC als auch den Vergleich der beiden *frs*-BGC *bcfrs* und *cvfrs* konnten außergewöhnliche Ein-

blicke in die Funktion beider Enzymsysteme erlangt werden, die im Folgenden weiter betrachtet werden sollen.

4.3 Vergleich der *frs*-BGC auf DNA- und Protein-Ebene

Die bioinformatische Betrachtung und der Vergleich der *frs*-BGC auf DNA- und Proteinebene ergaben grundlegende Erkenntnisse hinsichtlich deren Struktur (siehe Kap. 3.1). Hier konnte über die Größe der Gene gezeigt werden, dass sich die *frs*-BGC sehr ähnlich sind. Die Identität der Gene des *bcfrs*-BGC zum *cvfrs*-BGC beträgt sowohl auf DNA-, als auch auf Proteinebene um die 70 % und ist damit sehr hoch (siehe Tab. 3). Die Gene unterscheiden sich vor allem in ihrem GC-Gehalt (siehe Tab. 3). Wahrscheinlich wurde dieser im Laufe der Zeit an den natürlicherweise vorkommenden GC-Gehalt des Wirtes angepasst. Die Größe der einzelnen Gene unterscheidet sich kaum voneinander, sie weichen weniger als drei bp voneinander ab. Die Ausnahme bildet *frsA* mit der größten Abweichung. Hier ist *cvfrsA* um 51 bp größer als *bcfrsA*. Diese Abweichung ist hauptsächlich N-terminal bei beiden Genen lokalisiert (siehe Abb. A27).

4.4 Sich wiederholende Sequenzabschnitte in beiden *frs*-BGC

Die BGC *bcfrs* und *cvfrs* wurden jeweils auf Bereiche mit hoher Identität untersucht. Der Vergleich beider *frs*-BGC in Bezug auf die sich wiederholenden DNA- bzw. Aminosäuresequenzen konnte noch einmal die strukturelle Gemeinsamkeit beider BGC verdeutlichen (siehe Abb. 15). Die in beiden BGC gefundenen sich wiederholenden Bereiche geben einen möglichen Einblick in ihre Entstehung (siehe Abb. 15). Die Bildung neuer Gene wird von einer Vielzahl rekombinanter Ereignisse, wie Exon shuffling, Genduplikation, Retroposition, Insertionen einzelner Elemente oder Genfusion begünstigt. Durch den Selektionsdruck wird dieser Prozess beschleunigt [Long et al., 2003]. Betrachtet man die Sequenz beider BGC, so konnte gezeigt werden, dass beide BGCs ungewöhnlich viele sich wiederholender Sequenzen innerhalb des BGC

haben. Diese natürlich vorkommenden Rekombinationsstellen innerhalb der *frs*-BGC scheinen auf den ersten Blick willkürlich und liegen auch in Bereichen, in denen Gene enden und neu beginnen. Die Bereiche befinden sich teilweise an funktionellen Übergängen zwischen den einzelnen Domänen, tauchen an anderer Stelle aber in der Mitte einer Domäne auf (siehe Abb. 15).

Die Größe dieser sich wiederholenden Bereiche innerhalb beider *frs*-BGC reicht von 351 bp bis zu 3070 bp (siehe Tab. 4 und 5). Interessant ist hier, dass die Domänen, welche innerhalb der NRPS die gleiche oder eine ähnliche Funktion haben, durch identische DNA-Sequenzen codiert werden. So haben beispielsweise alle drei Leucin-aktivierenden A-Domänen die gleiche für sie kodierende DNA-Sequenz innerhalb eines BGC (siehe Abb. 15), so dass man vermuten kann, dass diese durch Verdopplung und Rekombination innerhalb des BGCs entstanden sind. Das gleiche gilt für die Alanin-aktivierenden A-Domänen, die N-MT-Domänen oder die beiden Starter-C-Domänen. Der größte gefundene sich wiederholende Bereich geht sogar über zwei Gene hinaus. Er umfasst den nicht-kodierenden Teil der A-Domäne ab der Core-A6-Region über eine MT-Domäne, die PCP-Domäne und einen Großteil der nächsten C-Domäne. Dieser Bereich wurde im N-Methyl-Dehydroalanin-Modul und im N-Methylalanin-Modul gefunden. Der kodierende Bereich der A-Domäne und der terminale C-Domänen-Abschnitt unterscheiden sich wieder, da hier andere funktionelle Eigenschaften verlangt werden. Die kodierenden Bereiche der A-Domäne A4 aus FrsE haben eine Spezifität für Serin und die A-Domäne A6 aus FrsF für Alanin.

Die an die Module 4 und 6 anschließenden C-Domänen, C6 und C8, werden ebenfalls bis zu dem C-terminalen Bereich durch gleiche DNA-Sequenzen codiert, ab dem sich die Funktion beider C-Domänen möglicherweise wieder unterscheidet. Bei Serin wird wahrscheinlich eine Dehydratisierung vorgenommen, das Alanin wird bis auf die anschließende Methylierung nicht modifiziert. Die DNA-Sequenzen, die für den Teil der C-Domänen kodieren, in dem sich die putativ für die Dehydratisierung verantwortlichen Aminosäuren befinden, weichen wieder voneinander ab. Die Rekombinationsstellen dieses

DNA-Abschnittes liegen in diesem Fall an einer Stelle, an der sich innerhalb der codierten Domäne C6 und C8 deren biochemische Funktionen unterscheiden.

Insgesamt scheint es, als wären für die Entstehung der BGC in Bereichen gleicher Funktion auf wenige vorhandene Sequenzen zurückgegriffen und nur diejenigen Bereiche aus einem anderen Genpool heraus entstanden, die eine andere Funktion erhalten haben (siehe Abb. 15). Dabei entsprechen die natürlichen Rekombinationsstellen nicht den Bereichen, die man aus molekularbiologischer oder aus proteinbiochemischer Sicht als funktionelle Einheit betrachten würde, wie beispielsweise eine DNA-Sequenz, die für eine vollständige A-Domäne oder eine vollständige C-Domäne codiert. Die Rekombinationsstellen liegen in der Mitte einer DNA-Sequenz, die für eine Domäne codiert. Das kann zum Einen ausschließlich durch zufällige Ereignisse passieren, oder es könnten sich im Laufe der Evolution innerhalb der Sequenzen Bereiche als sinnvoll herausgestellt haben, die öfter an bestimmten Stellen vorhanden sind. Mögliche Sequenzen wären dafür beispielsweise strukturbildende Elemente, wie die A-Core-Motive bei A-Domänen. Durch den Selektionsdruck könnten auf diese Art und Weise funktionale BGC entstehen. Das so beschriebene Szenario scheint eine effiziente Möglichkeit zur Entwicklung des *frs*-BGC zu sein. Die Tatsache, dass sich diese wiederholenden Bereiche in beiden BGC gleichen, macht es sehr wahrscheinlich, dass beide *frs*-BGC einen gemeinsamen Ursprung hatten. Nachfolgend wurden die *frs*-BGC schließlich auf *Cand. B. crenata* und *C. vaccinii* und ggf. weitere Bakterien übertragen. In diesem Zusammenhang wäre ein weiterer Vergleich mit den BGC der FR-Analoga YM und Sameuramid [Machida et al., 2018] interessant, die wahrscheinlich einen gemeinsamen Ursprung mit den FR-BGC haben.

Das Verständnis natürlich vorkommender Schnitt- und Rekombinationsstellen in BGCs kann Aufschluss über deren Phylogenie geben und darüber hinaus für Bioengineering-Strategien wichtig sein.

4.5 Strukturelle Besonderheiten der A-Domäne A8 aus FrsG

Die bioinformatische Untersuchung der A-Domänen A8 und der darauf folgenden zwei MT-Domänen ergab eine völlig neue, bislang nicht beschriebene Struktur einer A-Tridomäne (siehe Kap. 3.2). Die A-Domäne A8 ist die bis jetzt einzig bekannte A-Domäne, die zwei MT-Transferasen hat. Die erste ist wahrscheinlich eine O-methylierende-MT-Domäne, die zweite eine N-methylierende-MT-Domäne. Diese MT-Domänen liegen, wie bei den meisten A-Domänen, zwischen der Core-A8- und Core-A9-Region [Labby et al., 2015]. Interessant ist hier, dass auf beide MT-Domänen jeweils eine Core-A9- und Core-A10-Region folgt. Diese Core-Regionen liegen auf der A-Subdomäne [Conti et al., 1997] [Stachelhaus et al., 1999]. Das bedeutet, dass die A-Domäne A8 zwei Subdomänen enthält. Dieser Fall, der in beiden *frs*-BGC zu finden ist, ist bis jetzt einzigartig.

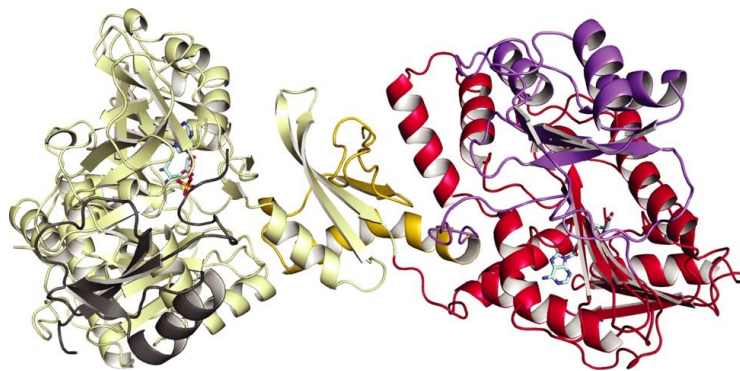


Abbildung 40: Kristallstrukturanalyse der A-MT-Domäne aus TioS des Thiocoralin-BGC. Die Abb. ist entnommen aus Mori et al. 2018. Der hellgelbe Bereich stellt die A-Core-Domäne dar, mit dem assoziierten MLP in grau. Der dunkelgelbe Bereich ist die A-Subdomäne. In rot der N-terminale katalytische Bereich der MT-Domäne und in violett die C-terminale SirA-like Region.

Für die A-MT-Domäne aus TioS des Thiocoralin-BGC (siehe Abb. 40) konnte 2018 die Kristallstruktur erstmals ermittelt [Mori et al., 2018] und damit die Beschaffenheit einer A-MT-Domäne gezeigt werden. Sie bildet eine

hantelförmige Struktur aus, dessen beide Enden einmal von der A-Domäne mit gebundenem MLP und von der MT-Domäne gebildet werden. Die A-Subdomäne liegt wie ein Verbindungsstück in der Mitte beider Domänen. Im Falle von TioS folgt auf die Subdomäne eine PCP-Domäne, die für die richtige Positionierung des Substrates in den aktiven Zentren beider Domänen sorgt. Im Falle der A-Domäne A8 aus FrsG folgt an dieser Stelle die zweite MT-Domäne mit einer weiteren A-Subdomäne (siehe Abb. 41).

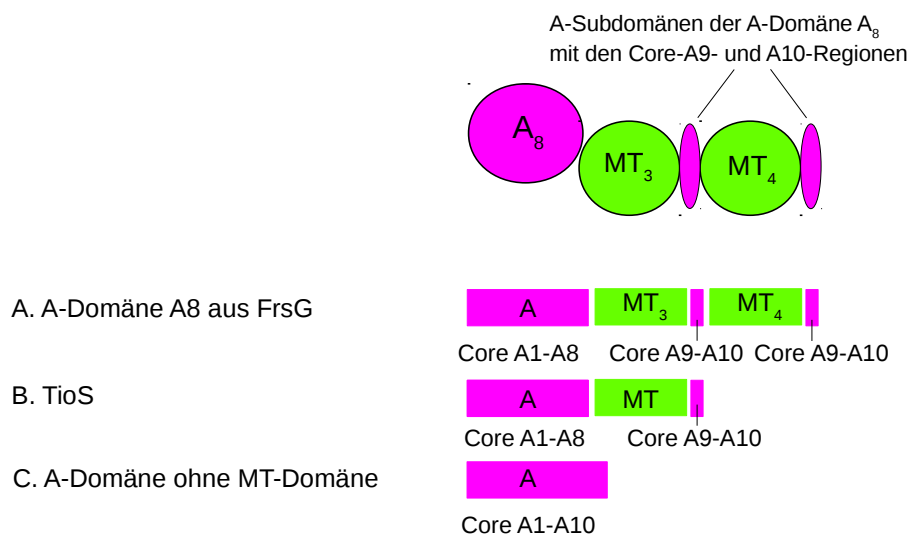


Abbildung 41: A. Schematische Darstellung der A-Domäne A8 aus FrsG mit zwei MT-Domänen, B. Vergleich mit TioS als Beispiel für eine A-Domäne mit einer MT-Domäne und C. Vergleich mit einer A-Domäne ohne MT-Domäne. Die Core-Regionen A9 und A10 sind auf den A-Subdomänen lokalisiert [Conti et al., 1997] [Stachelhaus et al., 1999].

Mori et al. vermuteten für TioS, dass, aufgrund der Distanz der aktiven Zentren von A- und MT-Domäne, beide Enzyme unabhängig voneinander arbeiten können. Eine weitere Subdomäne wie im Falle der *frs*-BGC wirkt allerdings die Frage auf, ob die A-Subdomänen in den Methylierungsvorgang involviert sein könnten. Möglicherweise dienen sie als Distanzhalter zwischen den Enzymen und hätte damit eventuell eine stabilisierende Eigenschaft auf

die Konformation. Eine weitere Möglichkeit besteht darin, dass die sehr bewegliche A-Subdomäne die Aminosäure nicht nur für den Adenylierungsvorgang, sondern auch für den Methylierungsvorgang in der richtigen Position hält. Das Lys-517 in der Core-A10-Region der A-Subdomäne koordiniert die Carboxygruppe, die in *trans* zur α -Aminogruppe und zur OH-Gruppe des Threonin steht. Sowohl die α -Aminogruppe, als auch die OH-Gruppe werden von den MT-Domänen methyliert. Ein Transport nur über die PCP-Domäne könnte an dieser Stelle nicht ausreichend sein, um die Reihenfolge der Reaktionen einzuhalten, daher wäre eine Stabilisierung über die A-Subdomäne ein plausibler Mechanismus. Wie hier das genaue Zusammenspiel der einzelnen Domänen aussieht, muss weiter untersucht werden.

4.6 Bioinformatische Betrachtungen zum Dehydratisierungsmechanismus der C-Domäne C6

Der Dehydroalanin-Baustein in FR wird vermutlich durch eine Dehydratisierung von Serin gebildet. Über die Analyse der C-Domänen des *bcfrs* und *cvfrs* konnten eindeutige Hinweise auf einen bisher unbekanntem Dehydratisierungsmechanismus gewonnen werden. C-Domänen katalysieren vielseitige Reaktionen in der NRPS-Maschinerie, vor allem die Kondensationsreaktion zweier Substrate, aber auch Dehydratisierungen und andere Modifikationen [Rausch et al., 2007] [Gaudelli et al., 2015]. Die C-Domänen beider *frs*-BGC weisen innerhalb eines BGC jeweils Identitäten zwischen 40 % und 60 % auf (siehe Abb. 22, A20 und A22). Diese sind gemessen am Durchschnitt (ca. 70 %) der *frs*-BGC eher gering. Nur die beiden Starter-C-Domänen sowie die C-Domänen C6 und C8 haben höhere Identitäten zueinander. Die Starter-C-Domänen ähneln sich mit um die 90 % Identität innerhalb eines BGCs am stärksten (siehe Abb. 22, A20 und A22), was aufgrund des gleichen Substrates nachvollzogen werden kann. Die überdurchschnittlich hohe Identität zwischen den C-Domänen C6 und C8 lässt sich damit aber nicht erklären. Trotz der hohen Identität zwischen diesen beiden C-Domänen müsste es aufgrund des funktionalen Unterschiedes auch signifikante Unterschiede in der Aminosäuresequenz geben. Die C-Domäne C6 könnte aufgrund der Struktur von

FR in eine Dehydratisierung involviert sein, die C-Domäne C8 nicht. Durch Alignments der C-Domänen C6 und C8 miteinander, jeweils für das *bcfrs*- und *cvfrs*-BGC (siehe Abb. 23 und 24), konnte gezeigt werden, dass beide C-Domänen im N-terminalen Bereich fast identisch sind, was der Grund für die hohe berechnete Identität ist. Im C-terminalen Bereich hingegen konnten Bereiche gefunden werden, in denen sich die C-Domänen C6 und C8 unterscheiden. Dabei fiel auf, dass sie sich in beiden BGC in 13 Aminosäuren auf identische Art und Weise unterscheiden (siehe Abb. 23 und 24).

An den bekannten Dehydratisierungsmechanismen, wie beispielsweise der in der Nocardicin-Biosynthese, sind Histidin-Reste an der Katalyse beteiligt [Gaudelli et al., 2015]. Die 13 in den *frs*-BGC gefundenen Aminosäuren der Domänen C6 beinhalten das Sequenzmotiv EQH, das in dieser Form auch in der C-Domäne des BGC von Microcystin (siehe Abb. A6) vorkommt, für die ebenfalls eine Dehydratisierungsfunktion postuliert wurde [Tillett et al., 2000] [Crüsemann et al., 2018]. Dieses Histidinmotiv ist in den C-Domänen C6 beider BGC vorhanden, aber nicht in denen der C-Domänen C8. Dieser Histidinrest könnte also eine Schlüsselrolle in der Dehydratisierung des Serin zu Dehydroalanin spielen. Ein 3D-Modell der C6-Domäne aus *C. vaccinii* konnte zeigen, dass sich diese 13 Aminosäuren inklusive der EQH-Sequenz im aktiven Zentrum der C-Domäne befinden und in dieser Position für die Katalyse einer solchen Reaktion in Frage kämen (siehe Abb. 25-27). Abschließend kann festgehalten werden, dass diese 13 Aminosäuren mit großer Wahrscheinlichkeit eine Rolle bei der Dehydratisierung des Serins zu Dehydroalanin spielen. Weitere *in-vitro*-Untersuchungen der C-Domänen sind an dieser Stelle nötig.

4.7 Spezifität der A-Domänen aus FrsA und FrsD

Die *in-vitro*-Untersuchung der A-Domänen FrsA und FrsD zeigte, dass diese Leucin als Substrat verwenden, wie es aufgrund der bioinformatischen Analyse vermutet wurde (siehe Tab. 1 und 2). Beide A-Domänen konnten erfolgreich exprimiert werden (siehe Kap. 3.5), allerdings konnte eine hohe

Reinheit der Proteine trotz Aufreinigung nicht erreicht werden. Es gab mehrere Verunreinigungen, die unter 20 kDa groß waren.

Das MLP wurde in der SDS-PAGE-Analyse nicht eindeutig detektiert. Ob das MLP nicht exprimiert wurde, aufgrund geringer Größe nicht detektierbar war, oder nicht mit der A-Domäne coeluiert wurde, kann nicht abschließend beurteilt werden. Trotzdem wurde im Assay durch beide A-Domänen Leucin aktiviert. Die Aktivierung von Leucin erfolgte wahrscheinlich unabhängig vom MLP. OH-Leucin wurde im Assay kaum aktiviert (siehe Abb. 38 und 39). Während der FR-Biosynthese erfolgt also zuerst eine Aktivierung von Leucin mit einer anschließenden Hydroxylierung, die potentiell von der β -Hydroxylase FrsH durchgeführt wird (siehe Abb. 38 und 39).

Die A-Domänen FrsA und FrsD wurden sowohl mit, als auch ohne MLP exprimiert. So konnte gezeigt werden, dass beide A-Domänen ohne MLP funktionsfähig sind. Ob eine andere A-Domäne der *frs*-BGC ein MLP benötigt, oder ob die MLP-kodierende Sequenz ein Relikt aus evolutionären Gründen sind, muss in weiteren Experimenten geklärt werden.

4.8 Ausblick

Die Biosynthese-Enzyme des nicht-ribosomalen Peptids FR werden durch die BGC *bcfrs* und *cvfrs* kodiert. Die bioinformatische Betrachtung beider BGC (siehe Kap. 3.1-3.4) und *in-vitro*-Studien der A-Domänen von FrsA und FrsD (siehe Kap. 3.5) erbrachten Hinweise auf neue Biosynthesemechanismen.

Die bis jetzt einzigartige A-Tridomäne-Struktur von A-Domäne A8 von *bcfrs* und *cvfrs* deutet darauf hin, dass die A-Subdomäne womöglich nicht nur eine Rolle bei der Adenylierungsreaktion spielt (siehe Kap. 3.2). Eine Analyse der Kristallstruktur wäre notwendig, um das strukturelle Zusammenspiel der A-Domäne mit den beiden A-Subdomänen und den integrierten MT-Domänen zu verstehen.

Der bioinformatische Vergleich beider BGC ermöglichte, die potentiell für den

Dehydratisierungsmechanismus, d.h. im Falle von FR die für die Umwandlung von Serin zu Dehydroalanin, wichtigen Aminosäuren der C-Domänen C6 zu finden (siehe Kap. 3.4). Dadurch kann künftig eine gezieltere Durchführung von *in-vitro*-Studien zu Dehydratisierungen in NRPS vorgenommen werden.

Eine interessante Möglichkeit wäre ein Vergleich der *frs*-BGC mit denen von YM-254890 und Sameuramid. Beides sind Strukturanaloga von FR, wobei YM-254890 ebenfalls in einer *Chromobacterium*-Spezies vorkommt. Sameuramid wurde aus marinen Ascidiaceae isoliert [Machida et al., 2018]. Die große Strukturähnlichkeit der Peptide macht einen gemeinsamen phylogenetischen Ursprung sehr wahrscheinlich. Insgesamt bieten die *frs*-BGCs, zusammen mit jenen für YM und Sameuramid, die Chance, die phylogenetische Entwicklung von BGCs nachvollziehen zu können. Wie genau sind diese BGC entstanden und wie hat sich durch Verdopplung einzelner Sequenzen ein System entwickeln können, das in komplexer Art und Weise aufeinander abgestimmt ist? Die Antwort auf solche Fragen könnte helfen, im Labor neue Bioengineeringansätze zu entwickeln.

5 Material und Methoden

5.1 Verwendete Datenbanken und Programme

FastTree: FastTree wurde für die Erstellung des phylogenetischen Baumes der C-Domänen genutzt. Dieser wurde in [Crüsemann et al., 2018] veröffentlicht [Price et al., 2009].

FigTree: FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) (Version 1.4.3) ist ein Programm für die grafische Darstellung von phylogenetischen Bäumen und wurde ebenfalls für die Erstellung des phylogenetischen Baumes der C-Domänen genutzt [Crüsemann et al., 2018].

Basic Local Alignment Search Tool (BLAST): [<https://blast.ncbi.nlm.nih.gov/Blast.cgi>] BLAST sucht Ähnlichkeiten in Sequenzen mithilfe von Datenbanken und gleicht diese ab. Dadurch können Funktionen von Proteinen vorhergesagt werden. Das Tool wird vom National Center for Biotechnology Information (NCBI) zur Verfügung gestellt.

Clustal omega: Clustal omega wurde verwendet, um größere Alignments durchzuführen, oder die Enden der sich wiederholenden Bereiche der Sequenzen festzulegen. [<https://www.ebi.ac.uk/Tools/msa/clustalo/>] Das Tool wird vom European Bioinformatics Institute (EBI), (EMBL) zur Verfügung gestellt.

antiSMASH: Die antiSMASH-Plattform wurde für die Annotationen der BGC *bcfrs* und *cvfrs* verwendet. Außerdem für die Substratspezifitätsvorhersage der A-Domänen. antiSMASH wird von vielen Forschungsinstituten und Universitäten unterstützt. [Medema et al., 2011]

Clonemanager: Clonemanager wurde zusammen mit Snapgene für Primerdesign und *in silico*-Klonierungen verwendet. Des Weiteren für die Berechnung des GC-Gehalts von Genen (Scientific & Educational Software).

Snappene Viewer: Snappene Viewer wurden zusammen mit Clonemanager für die Bearbeitung von Sequenzen und *in silico*-Klonierung verwendet (Snappene, Version: 4.1).

Phyre2: Zur Erstellung von Homologie 3D-Modellen wurde die Onlineplattform „Phyre 2“ verwendet. Modelle wurden für die C-Domänen abgeleitet [Kelley et al., 2015].

NaPDos: Die für den phylogenetischen Baum verwendeten C-Domänen wurden der NaPDos-Datenbank entnommen [Ziemert et al., 2012].

ACD/ChemSketch (Freeware2015): ChemSketch wurde verwendet, um Strukturformeln zu zeichnen.

DNA/RNA GC Content Calculator Endmemo:

[<http://www.endmemo.com/bio/gc.php>] Ein Onlinetool, das neben dem Clonemanager für die Berechnung des GC-Gehalts von Genen genutzt wurde.

MIBiG: [<https://mibig.secondarymetabolites.org/>] [Epstein et al., 2018] MIBiG ist eine Datenbank, in der gesammelte Annotationen und Metadaten diverser BGC und deren Produkte bereitgestellt werden.

5.2 Vektoren und Organismen

5.2.1 Vektoren

Für die Erstellung von Konstrukten wurden folgende Vektoren verwendet (siehe Tab: 13):

Vektor	Resistenz	Herkunft
pGEM-T	Ampicillin	Promega
pET28a	Kanamycin	Merck KGaA
pCDFDuet	Streptomycin	Merck KGaA
pCDFApra	Streptomycin Apramycin	AG König
19E BAC Vektor	Chloramphenicol	Aurelien Carlier [Carlier et al., 2016]

Tabelle 13: Verwendete Vektoren

Als Subklonierungsvektor wurde der pGEM-T-Vektor der Firma Promega verwendet, der für eine direkte Klonierung von PCR-Produkten mit A-Überhang geeignet ist und eine Blau-weiß-Selektion erlaubt. Für die Proteinexpression der A-Domänen wurde der pET28a Vektor der Firma Merck verwendet. Als Coexpressionsvektor für die Expression des MLPs wurde der pCDFDuet Vektor, ebenfalls der Firma Merck, verwendet, bei dem eine zusätzliche Apramycinresistenz eingefügt wurde (siehe Abb. A5).

5.2.2 Organismen

Für die Erstellung von Lager- und Expressionsstämmen wurden folgende Organismen verwendet (siehe Tab. 14):

Stamm	Genotyp	Herkunft
<i>E. coli</i> XL-1 blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac relA1 lac F' proAB lacIqZΔM15 Tn10 Tet^r</i>	Stratagen
<i>E. coli</i> B121 (DE3)	<i>F- ompT gal dcm lon hsd Sb (rB- mB-) λ(DE3 [lac lacUV5-T7 gene 1 ind 1 sam7 nin5])</i>	FA Invitrogen Live Technologies Corporation (Karlsruhe, Germany)
<i>E. coli</i> One shot B121 TM Star TM (DE3)	<i>F⁻ ompT hsdS_B (r_B⁻, m_B⁻) gal dcm rne131 (DE3)</i>	Fa. Thermo Fisher Scientific
<i>E. coli</i> α select silver	<i>F' deoR endA1 recA1 relA1 gyrA96 hsdR17 (r_k⁻, m_k⁺) supE44 thi-1 phoA Δ(lacZYA argF U169 Φ80 lacZΔM15λ⁻</i>	Fa. Bionline

Tabelle 14: Verwendete Organismen

Zur Lagerung der Konstrukte wurden entweder *E. coli* XL1 blue oder α -select silver verwendet. Zur Expression dienten *E. coli* B121 (DE3).

5.3 Medien und Puffer

In den folgenden Abschnitten werden die verwendeten Medien und Puffer beschrieben.

5.3.1 Medien

Zur Kultivierung der Bakterien wurden die in Tabelle 15 aufgeführten Medien verwendet. Für alle Medien wurde MilliQ-Wasser verwendet. Falls ein Bakterium ein Plasmid mit Resistenzgen trug, wurde das entsprechende Antibiotikum zu Beginn der Kultivierung dazu gegeben.

Medium	Bestandteil	Menge
LB-Medium [Bertani, 1951]	Hefeextrakt:	5 g/L
	Trypton:	10 g/L
	Natriumchlorid:	10 g/L
LB-Agarplatten	Hefeextrakt:	5 g/L
	Trypton:	10 g/L
	Natriumchlorid:	5 g/L
	Agar:	15 g/L
SOC-Medium	Hefeextrakt:	5 g/L
	Trypton:	20 g/L
	Natriumchlorid:	10 mM
	Kaliumchlorid:	2,5 mM
	Magnesiumchlorid:	10 mM
2xYT-Medium	Magnesiumsulfat:	10 mM
	NaCl	5 g/L
	Hefeextrakt	10 g/L
	Trypton	16 g/L
TB-Medium	Hefeextrakt:	24 g/L
	Trypton:	12 g/L
	Glycerol:	5 g/L
TB-10x-Salts	KH ₂ PO ₄	23,1 g/L
	K ₂ HPO ₄	125,4 g/L

Tabelle 15: Verwendete Medien

TB-Medium und TB-Salts wurden getrennt voneinander hergestellt, autoklaviert und erst kurz vor Gebrauch des Mediums 1:10 gemischt.

Alle Medien wurden mit MilliQ-Wasser hergestellt und im Anschluss autoklaviert. Die Salzlösung für das TB-Medium wurde nach dem Autoklavieren zugefügt.

5.3.2 Proteinaufreinigungspuffer

Für die Aufreinigung von Proteinen aus Expressionskonstrukten wurden folgende Puffer verwendet (siehe Tab. 16) [Qiagen, 2003].

Puffer	Bestandteil	Menge
Lysepuffer	NaH ₂ PO ₄	50 mM
	NaCl	300 mM
	Imidazol	10 mM
Waschpuffer 1	NaH ₂ PO ₄	50 mM
	NaCl	300 mM
	Imidazol	20 mM
Waschpuffer 2	NaH ₂ PO ₄	50 mM
	NaCl	300 mM
	Imidazol	30 mM
Elutionspuffer	NaH ₂ PO ₄	300 mM
	NaCl	300 mM
	Imidazol	30 mM

Tabelle 16: Proteinaufreinigungspuffer

5.3.3 Plasmidpräparationspuffer

Für die Präparation von Plasmid-DNA oder dem BAC-Vektor wurden folgende Puffer verwendet (siehe Tab. 17) [Quiagen, 2012]

Puffer	Bestandteil	Menge
P1	Tris-HCl (pH8)	50 mM
	EDTA (pH8)	10 mM
P2	NaOH	0,2 mM
	SDS	1 %
P3	Kaliumacetatlösung (pH5,5)	3 M

Tabelle 17: Plasmidpräparationspuffer

Wenn eine größere Menge DNA gewünscht war, oder der BAC-Vektor isoliert wurde, dienten die Puffer P1-P3 für die Präparation. 2 mL einer ÜNK wurden abzentrifugiert und das Pellet in 100 µL P1 resuspendiert. Die Zellsuspension

wurde mit 200 μL P2 versetzt und nicht länger als 5 min. inkubiert, bis eine Schleimbildung eintrat. Im Anschluss wurde mit 150 μL eiskaltem Puffer P3 neutralisiert. Der Ansatz wurde für 3-5 Minuten auf Eis gestellt. Die ausgeflockten Proteine und Zellreste wurden für 10 min bei voller Geschwindigkeit abzentrifugiert und der Überstand mit dem zweifachen Volumen Ethanol versetzt. Danach wurde die DNA für 10 Minuten bei voller Geschwindigkeit abzentrifugiert. Im Anschluss wurde das Pellet mit einer 70 %igen Ethanol-Lösung gewaschen, für 2 Minuten zentrifugiert und der Überstand restlos entfernt. Das Pellet wurde getrocknet, bis keine Ethanolreste mehr vorhanden waren. Das trockene Pellet wurde im Anschluss in mindestens 30 μL MilliQ-Wasser gelöst.

5.3.4 Assaypuffer

Für die Durchführung des Assays wurden folgende Puffer verwendet (siehe Tab. 18) [Phelan et al., 2009].

Puffer	Bestandteil	Menge
Lösung 1	Aminosäure	3 mM
	PPi	15 mM
	Tris (pH:7,5)	20 mM
Lösung 2	γ - ¹⁸ -O ₄ -ATP	3 mM
	MgCl	15 mM
	Tris (pH:7,5)	20 mM
Lösung 3	Tris (pH:7,5)	20 mM
	Glycerol	5 %
	DTT	1 mM

Tabelle 18: Assaypuffer

5.3.5 Anoden- und Kathodenpuffer für die SDS-Analyse

Für die SDS-Analyse von Proteinen wurden für die Anode und die Kathode verschiedene Puffer verwendet, um eine bessere Auftrennung der Proteine

im kleinen kDa Bereich zu erreichen (siehe Tab. 19) [Schägger, 2006]. Die Puffer wurden in den jeweiligen Kathodenbereich bzw. Anodenbereich der Gelkammer gefüllt.

Puffer	Bestandteil	Menge
Kathodenpuffer	Tris (pH 8,25)	100 mM
	Tricin	100 mM
Anodenpuffer	Tris (pH 8,9)	200 mM

Tabelle 19: Kathoden- und Anodenpuffer

5.3.6 Weitere Puffer und Lösungen

Für die Gelelektrophorese wurde TBE- oder TAE-Puffer verwendet. Dazu wurde der 10x oder 50x Puffer zu einem 1x Puffer verdünnt. Die Färbe- und Entfärbelösung diente zum Färben- und Entfärben von Proteingelen (siehe Tab. 20).

Puffer	Bestandteil	Menge
TE	Tris-HCl (pH 8)	100 mM
	EDTA	1 mM
10X TBE	Tris	0,89 M
	EDTA	0,02 M
	H ₃ BO ₃	0,87 M
50X TAE	EDTA-Na x 2H ₂ O	18,61 g/L
	Eisessig	60,05 g/L
	Tris	242,28 g/L
Färbelösung für Proteingele	Essigsäure	10 %
	Methanol	50 %
	Comassie brilliant blue R-250	0,005 %
	Wasser	40 %
Entfärbelösung für Proteingele	Essigsäure	10 %
	Methanol	30 %
	Wasser	60 %

Tabelle 20: Weitere Puffer und Lösungen, die verwendet wurden.

5.3.7 Antibiotikastocks

Die Antibiotika wurden in Antibiotikastocks bei $-20\text{ }^{\circ}\text{C}$ gelagert und bei Gebrauch im Verhältnis 1:1000 verdünnt. Als Lösungsmittel wurde Wasser verwendet. Bei Chloramphenicol wurde Ethanol als Lösungsmittel verwendet (siehe Tab. 21).

Antibiotikum	Konzentration
Ampicillin	100 mg/mL
Apramycin	100 mg/mL
Chloramphenicol	12 mg/mL
Kanamycin	60 mg/mL
Streptomycin	100 mg/mL

Tabelle 21: Für diese Arbeit verwendete Antibiotikastocks.

5.4 Mikrobiologische Techniken

In den folgenden Abschnitten werden die verwendeten Mikrobiologischen Techniken beschrieben.

5.4.1 Stammhaltung in Kryokulturen

Zur längerfristigen Lagerung wurden die Bakterienstämme als Kryokulturen bei $-80\text{ }^{\circ}\text{C}$ gelagert. Dazu wurden $800\text{ }\mu\text{L}$ einer Übernachtskultur (ÜNK) unter sterilen Bedingungen mit $800\text{ }\mu\text{L}$ Glycerin in ein Kryoröhrchen gefüllt und bei $-80\text{ }^{\circ}\text{C}$ gelagert.

5.4.2 Kultivierung von Bakterien

Die Bakterien wurden unter folgenden Bedingungen kultiviert (siehe Tab. 22).

Stamm	Medium	Temperatur	Schüttler- geschwindigkeit
<i>E. coli</i> X11 blue	LB	37 °C	200-220 RPM
<i>E. coli</i> α -select silver	LB	37 °C	200-220 RPM
<i>E. coli</i> B121 (DE3)	LB, TB	37 °C	200-220 RPM
<i>E. coli</i> One shot B121 TM Star TM (DE3)	LB, TB	37 °C	200-220 RPM

Tabelle 22: Bedingungen zur Kultivierung von Bakterien.

5.4.3 Konzentrationsbestimmung von Bakterienkulturen

Die Bakterienkulturen wurden über die optische Dichte der Bakterienkultur in Relation zum sterilen Medium photometrisch bestimmt. Hierzu wurde ein Nanodrop der Firma Thermo Scientific verwendet.

5.4.4 Transformation von Bakterien

Die Transformation erfolgte entweder mit Elektroporation oder mit Chemischer Transformation .

Herstellung elektrisch kompetenter Zellen: Aus einer ÜNK wurde ein 100 mL LB Erlenmeyerkolben mit 1 mL ÜNK angeimpft. Die Zellen wurden bei 37 °C und 220 RPM bis zu einer OD von 0,5 wachsen gelassen und bei 10.000-15.000 RPM abzentrifugiert. Im Anschluss wurden die Zellen drei mal mit 30 mL MilliQ-Wasser gewaschen und in 100-200 μ L Wasser aufgenommen. Elektrisch kompetente Zellen wurden in Aliquots von 75 μ L aufgeteilt und meistens sofort weiterverwendet oder manchmal nach Zugabe von 10 % Glycerol bei -80 °C eingefroren.

Elektroporation: 75 μ L Zellsuspension wurden bei Raumtemperatur mit 1-8 μ L Plasmid versetzt und in eine Elektroporationsküvette von 2 mm Durchmesser gegeben. Der Ansatz wurde 2-3 Minuten bei Raumtemperatur inkubiert. Dann wurde in einem Elektroporator der Firma Biorad mit dem

Programm Ec2 bei 2,5 kV eine kurzzeitige Spannung angelegt. Unmittelbar danach wurden 800 μ L SOC-Medium hinzugegeben und die Zellen für eine Stunde bei 37 °C inkubiert. Danach wurde auf Selektivagar ausplattiert und über Nacht bei 37 °C inkubiert.

Herstellung chemisch kompetenter Zellen: Für die Herstellung chemisch kompetenter Zellen wurden 3 mL LB-Medium mit 30 μ L einer Kryokultur angeimpft und über Nacht kultiviert. 2x 100 mL 2xYT-Medium wurden mit jeweils 1 mL ÜNK angeimpft. Die Kulturen wurden bei 37 °C ca. 2 Stunden auf OD₆₀₀: 0,3-0,4 wachsen gelassen. Die Kultur wurde auf 4 Fal-kontubes aufgeteilt und 10 Minuten bei 8000 RPM und 4 °C abzentrifugiert. Der Überstand wurde verworfen und die Pellets in jeweils 7,14 mL einer eiskalten CaCl₂/MgSO₄-Lösung (70 mM/20 mM) resuspendiert und 30 Minuten auf Eis inkubiert. Im Anschluss wurde 625 μ L 100 % Glycerol zugegeben und zu 50 μ L aliquotiert.

Chemische Transformation: 50-100 mL der tiefgefrorenen chemisch kompetenten Zellen wurden mit 1-5 μ L Plasmid oder bis zu 10 μ L bei Ligationen versetzt und für 20 Minuten auf Eis aufgetaut und inkubiert. Danach wurde für 90 Sekunden auf 42 °C erhitzt, und im Anschluss wurden die Zellen mit 800 mL LB-Medium versetzt und für eine Stunde bei 37 °C inkubiert. Danach wurde auf Selektivagar ausplattiert und über Nacht bei 37 °C inkubiert.

5.4.5 Blau-Weiß-Selektion

Um den Erfolg der Ligation einer PCR in den pGEM-T Vektor zu überprüfen, wurde vor dem Screening der Klone eine Blau-weiß-Selektion durchgeführt. Der besagte Vektor verfügt über ein *lacZ* Gen, welches genau in der MCS liegt. Sobald der Vektor ein Fragment aufnimmt, wird das *lacZ*-Gen unterbrochen, das offene Leseraster zerstört, und es kann keine funktionale β -Galaktosidase exprimiert werden. Das führt dazu, dass nach der Zugabe einer X-Gal-Lösung (4 mg/mL) bei den Kolonien, die ein nicht erfolgreich kloniertes Plasmid aufgenommen haben, eine funktionale β -Galaktosidase exprimiert wird. Diese spaltet das künstliche Substrat in Galaktose und einen

Indigofarbstoff, der negative Kolonien blau färbt. Wurde ein Insert erfolgreich kloniert und in ein Bakterium gebracht, ist die Zelle nicht mehr in der Lage, eine funktionierende β -Galaktosidase zu exprimieren und die Kolonien bleiben weiß gefärbt. Das reduziert den Screeningaufwand.

5.5 Molekularbiologische Methoden

In den folgenden Abschnitten werden die verwendeten Molekularbiologischen Techniken beschrieben.

5.5.1 Polymerase-Kettenreaktion (PCR)

Um DNA-Abschnitte zu amplifizieren, wurde eine PCR durchgeführt. Als Template für die A-Domänen und das MLP diente der BAC Clone 19E [Carrier et al., 2016]. Die PCR wurde ebenfalls für die Kontrolle der Subklonierungsvektoren und Expressionsvektoren verwendet. Hier wurden in der Regel erst einmal Kolonie-PCRs durchgeführt. Bei nicht eindeutigem Ergebnis, wurde die DNA isoliert und eine weitere PCR- oder eine Restriktionsanalyse folgte.

Als Reaktionsansatz für die PCR wurden folgende Mischungen verwendet (siehe Tab. 23 und 24).

Bestandteil	Volumen
Puffer	10 μ L
DMSO	2,5 μ L
MgCl (25 mM)	2,5 μ L
dNTP (10 mM)	1 μ L
Templat-DNA	0,5-1 μ L
DNA-Polymerase (5 U/ μ L)	0,25-0,5 μ L
Vorwärtsprimer	0,5 μ L
Rückwärtsprimer	0,5 μ L
MilliQ-H ₂ O	ad 50 μ L

Tabelle 23: Reaktionsansatz für die PCR mit *Taq*-Polymerase.

Bestandteil	Volumen
Puffer	5 μL
dNTP (10 mM)	1 μL
Templat-DNA	0,5-1 μL
DNA-Polymerase (5 U/ μL)	0,25 μL
Vorwärtsprimer	0,5 μL
Rückwärtsprimer	0,5 μL
MilliQ-H ₂ O	ad 50 μL

Tabelle 24: Reaktionsansatz für die PCR mit *Pfu*-Polymerase.

Es wurden zwei unterschiedliche PCR-Protokolle verwendet. Zur Erfolgskontrolle einer Klonierung wurde die *Taq*-Polymerase verwendet (siehe Tab. 23). Sollte ein PCR-Fragment in ein Plasmid kloniert werden, wurde die *Pfu*-Polymerase (siehe Tab. 24) verwendet, die eine niedrigere Fehlerrate aufweist als die *Taq*-Polymerase. Es wurden jeweilige Standardprotokolle der beiden Polymerasen verwendet, in denen die Annealingtemperatur für jeden Primer variiert wurde. Diese wurde für jeden Primer mit einer Gradienten-PCR festgestellt. Die Elongationszeiten richteten sich jeweils nach der Länge des zu amplifizierenden Produkts und der Art der verwendeten Polymerase. Wurde eine Kolonie-PCR-durchgeführt, wurde 1 μL Templat verwendet, ansonsten 0,5 μL .

Als Temperaturprogramm wurde folgendes Standardtemperaturprotokoll verwendet (siehe Tab. 25).

Schritt	Temperatur	Zeit	Widerholung
1	95 °C	5 min	
2	95 °C	1,30 min	
3	50-70 °C	2,30 min	
4	72 °C	2,30 min	2-4: 29x
5	72 °C	4 min	

Tabelle 25: Temperaturprogramm für die PCR.

Die Annealingtemperatur der Primer wurde durch Gradienten-PCR festgelegt. Die Elongationszeit wurde je nach Fragmentlänge und Geschwindigkeit der Polymerase angepasst.

5.5.2 Kolonie-PCR

Mussten viele Klone auf Plasmide gescreent werden, wurde die Kolonie-PCR verwendet. Dafür wurden die jeweiligen Klone von einer Agarplatte gepickt, in 10 μL Wasser gelöst und für 10 Minuten auf 95 °C erhitzt. Davon wurde 1-2 μL als Templat verwendet. Aus Flüssigkulturen wurden jeweils 2 μL entnommen und dem PCR-Ansatz direkt dazugegeben.

5.5.3 Verwendete Primer

Um die Primer für die Amplifizierung von DNA-Abschnitten zu bekommen, wurden die Primer mit Clone-Manager und Snap-Gene-Viewer designt und von der Fa. Eurofins Genomics synthetisiert (siehe Tab. 26).

Primer	Sequenz
A1 vorwärts	GGATCC GGGCCGGACGAACGCGTGAGCTATGG
A1 rückwärts	AAGCTT CTATTCATACGGATTCCAGCTCTGATTG
A2 vorwärts	GCGGCCG CACGAACGCGTGAGCTATGGCGAGTTG
A2 rückwärts	CTCGAGT TATAAGACAACCGCCAGCTCCG
FrsB vorwärts	CATATG AGCAATCCTTTTGATGATAAAGAC
FrsB rückwärts	CTCGAGT CATGTTTCATCCGATTGCATCGCATCGAT

Tabelle 26: Sequenzen der verwendeten Primer. Die jeweiligen Schnittstellen sind fett gedruckt.

5.5.4 Reinigen von PCR-Produkten oder Aufkonzentrieren von DNA

Zum Reinigen von PCR-Produkten oder zum Aufkonzentrieren von DNA wurde das Clean and Concentrator Kit der Firma Zymo Research verwendet. Dabei wurde nach den Vorgaben des Herstellers gearbeitet.

5.5.5 DNA-Präparation

Um Plasmide zu isolieren, wurde die *E. coli* Kolonie mit dem gewünschten Plasmid in 3 mL LB-Medium kultiviert und über Nacht bei 37 °C wachsen gelassen. Die Plasmide wurden entweder mit dem Pure yield Miniprep Kit von Promega aufgereinigt oder mit den Puffern P1-P3. Um den BAC zu isolieren oder größere Mengen Plasmid zu isolieren, wurde mit den Puffern P1-P3 gearbeitet. Dazu wurden Kulturen bis zu 50 mL abzentrifugiert und das Zellpellet in 100-200 µL Puffer P1 gelöst. Für den Zellaufschluss wurde 200-400 µL Puffer P2 dazugegeben, bis eine Schleimbildung eintrat, aber nur für 5 Minuten. Im Anschluss wurde der Ansatz mit 150-300 µL eiskaltem Puffer P3 neutralisiert. Gegebenenfalls wurde alles noch mal auf Eis stehen gelassen und im Anschluss abzentrifugiert. Die DNA im Überstand wurde mit doppeltem Volumen 100 % Ethanol ausgefällt und abzentrifugiert. Das Pellet wurde mit wenig 70 % Ethanol gewaschen und getrocknet, bis keine Ethanolrückstände mehr vorhanden waren. Im Anschluss wurde die DNA in mindestens 30 µL MilliQ-H₂O gelöst. Die Lagerung erfolgte bei -20 °C.

5.5.6 Agarose Gelelektrophorese

Zur Analytik von PCR-Fragmenten oder zur Aufreinigung von Restriktionen wurde Gelelektrophorese eingesetzt. In der Regel wurde dafür eine 1 %ige Agaroselösung in TAE-Puffer hergestellt. Für die kleineren Fragmente wie beispielsweise das MLP wurde eine 3 %ige Lösung hergestellt. Diese wurde in der Mikrowelle erhitzt, bis die Agarose vollständig gelöst war. Die Agarose wurde in eine Gelkammer gegossen. Nach Aushärten der Agarose wurde die Kammer mit TAE-Puffer befüllt und der Kamm entfernt. Die Proben wurden mit Ladebuffer versetzt und aufgetragen. Für analytische Zwecke wurden

2-5 μL für präparative Gele wurden zwischen 25 μL und 50 μL verwendet. Als Größenmarker wurde der Gene Ruler Mix der Firma Thermo Fisher Scientific verwendet. Um eine Auftrennung der Proteine zu erreichen, wurde eine Spannung von 90 V bis 100 V für 20-30 Minuten angelegt und das Gel anschließend in einer Ethidiumbromid-Lösung gefärbt. Die Auswertung erfolgte unter UV-Licht mit dem Intas Gel iX Imager.

5.5.7 DNA-Isolierung aus Agarosegelen

Für eine Weiterverarbeitung der DNA-Fragmente aus den Agarosegelen wurden die gewünschten Banden mit einem Skalpell ausgeschnitten, in ein 2 mL Eppendorf tube überführt und mit dem Qiagen Gel extraction Kit nach Herstellerprotokoll isoliert.

5.5.8 Sequenzierung

Ob die amplifizierten Fragmente von den Polymerasen korrekt abgelesen wurden, wurde überprüft, indem die einzelnen Konstrukte durch die Firma GATC Biotech sequenziert wurden. Hierfür wurde die Didesoxymethode oder Kettenabbruchmethode nach Sanger verwendet [Sanger et al., 1977]. Dafür wird in vier gleichen Reaktionsansätzen jeweils eines der Nukleotide als Didesoxy-Nucleotid beigemischt. Auf diese Weise erfolgt nach Einbau des Nucleotids an dieser Stelle ein Kettenabbruch. Im Anschluss werden die Fragmente über Gelelektrophorese aufgetrennt.

5.5.9 Restriktionsspaltung von DNA

Um die gewünschten Inserts in die Expressionsvektoren zu bringen oder um den Erfolg einer Klonierung zu testen, wurden die Vektoren mit den entsprechenden Restriktionsendonucleasen falls nötig über Nacht geschnitten. Restriktionsendonucleasen katalysieren den Schnitt der DNA an bestimmten palindromischen Sequenzen. In den meisten Fällen wurden zwei verschiedene Restriktionsenzyme verwendet. War das nicht der Fall, wurde der Vektor, der das Fragment aufnehmen sollte, dephosphoryliert, um Selbstligation zu

vermeiden. Die Restriktionsendonucleasen wurden nach Vorschrift des Herstellers verwendet. Eine Auftrennung der Fragmente erfolgte im Anschluss über Gelelektrophorese.

5.5.10 Dephosphorylierung von Vektoren

In Fällen, in denen eine Restriktion mit nur einem Enzym durchgeführt wurde, wurden Vektoren dephosphoryliert, um Selbstligation zu vermeiden. Dazu wurde dem Restriktionsansatz nach vollendeter Restriktion alkalische Phosphatase aus Kälberdarm zugegeben und für eine halbe Stunde inkubiert. Dadurch wurden die Phosphatgruppen am 5-Ende der Vektoren abgespalten. Im Anschluss wurde der Ansatz mit dem Clean and Concentrator Kit der Firma Zymo Research aufgereinigt.

5.5.11 Ligation

Um DNA-Fragmente in Vektoren einzubringen, wurde eine Ligation mit Promega T4 DNA Ligase über Nacht durchgeführt. Dadurch wurden die Diester Bindungen zwischen der 3'-Hydroxygruppe und 5'-Phosphat durch die T4-DNA-Ligase neu verknüpft. Das molare Verhältnis von Insert zu Vektor wurde 3:1 gewählt. Der Reaktionsansatz wurde in der Regel mit 10 μL angesetzt.

Bestandteil	Volumen
T4-DNA Ligase Puffer	1 μL
T4-DNA Ligase	1 μL
Vector:Insert DNA	1:3
MilliQ-Wasser	ad 10 μL

Tabelle 27: Ligationsansatz

Der Reaktionsansatz (siehe Tab. 27) wurde über Nacht bei 16 °C inkubiert und im Anschluss bei 65 °C für 10 Minuten inaktiviert.

5.5.12 Vektorstrukture

Folgende Konstrukte, Expressions- und Coexpressionsstämme wurden im Laufe dieser Arbeit erstellt oder verwendet (siehe Tab. 28 und 29):

Konstrukt	Vektor	Insert	Schnittstellen	Größe Insert [bp]
pSMART BAC clone 19E	pSMART BAC	komplettes <i>bcfrs</i> -BGC		35811
A1 clone E	pGEM-T	<i>frsA</i> A-PCP-Domäne <i>Cand. B. crenata</i>	BamHI, HindIII	1671
A2 clone 2e	pGEM-T	<i>frsD</i> A-PCP-Domäne <i>Cand. B. crenata</i>	NotI, XhoI	1627
pAC1	pET28a	<i>frsA</i> A-PCP-Domäne <i>Cand. B. crenata</i>	BamHI, HindIII	1671
pAC2	pET28a	<i>frsD</i> A-PCP-Domäne <i>Cand. B. crenata</i>	NotI, XhoI	1627
MLP 8	pCDFApra	<i>frsB</i> aus <i>Cand. B. crenata</i>	NdeI, XhoI	221

Tabelle 28: Konstrukte, die in dieser Arbeit verwendet wurden. Der pSMART BAC clone 19E wurde von Aurelien Carlier zur Verfügung gestellt [Carlier et al., 2016].

Wirt	Vektor I	Vektor II	Resistenzen
<i>E. coli</i> XL-1 blue	A1 clone E	-	Amp ^R
<i>E. coli</i> XL-1 blue	A2 clone 2e	-	Amp ^R
<i>E. coli</i> α select silver	pAC1	-	Kan ^R
<i>E. coli</i> α select silver	pAC2	-	Kan ^R
One shot B121 TM Star TM (DE3)	pAC1	-	Kan ^R
One shot B121 TM Star TM (DE3)	pAC2	-	Kan ^R
One shot B121 TM Star TM (DE3)	pAC1	MLP 8	Kan ^R Strep ^R Apra ^R
One shot B121 TM Star TM (DE3)	pAC2	MLP 8	Kan ^R Strep ^R Apra ^R

Tabelle 29: Expressionsstämme und Coexpressionsstämme, die in dieser Arbeit verwendet wurden.

Die A-Domänen und das FrsB wurden mittels PCR amplifiziert. Als Templat diente der BAC Vektor 19E. Die A-PCP-Domänen und das MLP wurden mittels pGEM-T subcloniert, positive Klone kultiviert, auf Mutationen überprüft, im Anschluss mit den entsprechenden Restriktionsenzymen geschnitten und in den Expressionsvektor ligiert. Das FrsB wurde mit NdeI aus dem pGEM-T-Vektor geschnitten. Im Anschluss wurde der Klon 8, der das FrsB in der richtigen Orientierung enthielt, ausgewählt.

5.6 Proteinexpression, Proteinreinigung und Proteinanalytik

5.6.1 Proteinexpression

Aus einer ÜNK wurden 5 mL entnommen und in 0,5 L TB-Medium in einen Schikanekolben gegeben. Die Zellen wurden auf OD_{600} : 1,2 wachsen gelassen. Die Kultur wurde im Anschluss für 15 Minuten auf Eis gekühlt, mit einer Endkonzentration von 0,5 mM IPTG induziert und bei 16 °C und 200 RPM über Nacht exprimiert. Am nächsten Tag wurde sie bei 4000 RPM und 4 °C abzentrifugiert und in 2,5 mL Lysepuffer pro g Zellpellet gelöst.

5.6.2 Zellaufschluss

Die Zellsuspension wurde möglichst auf Eis gehalten. Der Zellaufschluss erfolgte mit Ultraschall in einem Branson Sonifier 250 mit einer Leistung von 4.5 %. Die Suspension wurde für 10 Sekunden den Impulsen ausgesetzt. Danach erfolgte eine Pause auf Eis von ebenfalls 10 Sekunden, um ein Überhitzen der Proteine zu vermeiden. Dieser Vorgang wurde 10x wiederholt. Danach wurden die unlöslichen Zellfragmente bei 15000 g für 30 Minuten abzentrifugiert. Der Überstand wurde für eine weitere Aufreinigung mit Ni-NTA-Agarose verwendet. Das Pellet wurde für eine SDS-PAGE-Analyse verwahrt.

5.6.3 Aufreinigung mit His-Taq

Der Zellüberstand wurde mit 1 mL einer Ni-NTA-Agarose versetzt und für eine Stunde unter leichtem Schwenken inkubiert. Anschließend wurden die

Bestandteile der Suspension über eine Polypropylensäule der Fa. Qiagen voneinander getrennt, zunächst mit 4 mL eines 20 mM Imidazolpuffers gewaschen und danach mit 4-16 mL eines 30 mM Imidazolpuffers. Die Elution erfolgte mit einem 300 mM Imidazolpuffer in drei Elutionen. Das Endvolumen der Elution betrug 2,5 mL, um für den Entsalzungsprozess das entsprechende Volumen zu bekommen. Auch hier wurden aus jeder Fraktion 30 μ L Probe für die SDS-Analyse entnommen.

5.6.4 Entsalzen und Aufkonzentrierung

Zum Entsalzen wurden die drei Elutionsfraktionen auf eine mit 4x5 mL Assaylösung vorgewaschene PD10-Desalting column gegeben und mit 3 mL Assaypuffer wieder von der Säule gewaschen. Die entsalzte Proteinlösung wurde in Vivaspin 500 Zentrifugenröhrchen der Größe 30 (30 MWCO molecular weight cut-off) bis auf ein Volumen, das für SDS-PAGE-Analyse und Assay noch nötig war, aufkonzentriert.

5.6.5 Probenvorbereitung für die SDS-PAGE-Analyse

Von jeder Wasch- und Elutionsfraktion wurden 30 μ L in ein PCR-Tube gegeben und mit 10 μ L Nupage LDS Sample Buffer versetzt. Im Anschluss wurden die Proben für 10 Minuten bei 95 °C denaturiert.

5.6.6 SDS-Polyacrylamidgelelektrophorese

Für die Analytik der Proteinexpression wurde die SDS-PAGE-Analyse verwendet. Dazu wurden die vorbereiteten Proben aus Kapitel 5.6.5 auf ein Polyacrylamid Bisacrylamid Gel aufgetragen (siehe Tab. 30 und 31). Die Elektrophorese wurde mit dem XCell SureLOCK Mini-Cell System der Fa. Thermo Fisher Scientific durchgeführt. Zuerst wurde das Trenngel gegossen und mit Isopropanol überschichtet. Nach vollständiger Polymerisation des Trenngels wurde der Isopropanol entfernt und das Trenn- mit dem Sammelgel überschichtet. Nach vollständiger Polymerisation des Sammelgels wurden die Proben in die Taschen eingebracht und eine Spannung von 90 Volt angelegt, bis die Proben das Trenngel erreicht hatten. Danach wurde die Spannung auf

120 Volt erhöht. Im Anschluss wurden die Gele mit einer Coomassie-Blau-Färbung eingefärbt und analysiert (siehe Kap. 5.3.6). Dazu wurden sie in die Färbelösung gelegt und in der Mikrowelle erhitzt, bis zur leichten Blasenbildung. Danach wurde auf dem Schüttler bei leichter Bewegung für 5 Minuten inkubiert. Im Anschluss wurden die Gele mit Wasser abgespült und über Nacht in die Entfärbelösung gelegt. Um überschüssige Farbe aufzusaugen, wurden Zellstofftücher mit in die Entfärbelösung gelegt.

Bestandteil	Volumen
Acrylamid/Bisacrylamid	4000 μL
Wasser	3300 μL
Tris/HCl pH 6.8 (1 M)	2500 μL
SDS (10 %)	100 μL
APS (10 %)	100 μL
TEMED	4 μL

Tabelle 30: Trenngel

Bestandteil	Volumen
Acrylamid/Bisacrylamid	510 μL
Wasser	2040 μL
Tris/HCl pH 6.8 (1 M)	375 μL
SDS (10 %)	30 μL
APS (10 %)	30 μL
TEMED	3 μL

Tabelle 31: Sammelgel

Als SDS-Laufpuffer wurde das Tris/Tricin-System nach Schagger verwendet, welches eine höhere Auflösung im Bereich kleiner Proteine hat (siehe auch Kap. 5.3.5) [Schägger, 2006].

5.6.7 Proteinkonzentrationsbestimmungen

Die Konzentration der Proteinlösung und der einzelnen Fraktionen wurde in einem Nanodrop-Gerät der Firma Thermo Scientific gemessen. Als Referenz dienten die jeweilig verwendeten Puffer der einzelnen Wasch- und Elutionsfraktionen.

5.6.8 A-Domänen-Assay

Um die Substratspezifität der einzelnen A-Domänen zu bestimmen, wurde ein massenspektrometrisch basierter ATP-Austausch-Assay durchgeführt [Phelan et al., 2009]. Mit Hilfe von MALDI-TOF-MS wurde die Bildung von ATP aus γ - $^{18}\text{O}_4$ -ATP gemessen. In Anwesenheit des richtigen Substrates und ATP bildet die A-Domäne das entsprechende Substrat-Adenylat. Diese Reaktion unterliegt einem Gleichgewicht. Für den Assay wurde die Aminosäure zusammen mit γ - $^{18}\text{O}_4$ -ATP und einem Überschuss an PPI inkubiert. Für den Fall des richtigen Substrates wurde das γ - $^{18}\text{O}_4$ -ATP von der A-Domäne verbraucht, indem es unter Abspaltung der gelabelten Phosphatgruppe an das Substrat bindet. Bei Ausbleiben der folgenden Thiolierungsreaktion, wie sie normalerweise bei der NRPS stattfindet, reagiert das Aminoadenylat wieder zu ATP und der Aminosäure. Aufgrund des Überschusses an PPI wird nicht mehr das γ - $^{18}\text{O}_4$ -ATP gebildet, sondern ATP. Im MALDI-TOF-MS-Diagramm kommt es zu einem Shift vom m/z 514 γ - $^{18}\text{O}_4$ -ATP Peak zu einem m/z 506 ATP-Peak. Für den Assay wurden 3 Stock-Lösungen hergestellt (siehe auch Tab. 18, Kap. 5.3.4). Lösung 1: 3 mM der zu testenden Aminosäure, 15 mM PPI, 20 mM Tris (pH 7,5) Lösung 2: 3 mM γ - $^{18}\text{O}_4$ -ATP, 15 mM MgCl in 20 mM Tris Lösung 3: das aufkonzentrierte Enzym in 20 mM Tris mit 5 % Glycerol und 1 mM DTT. Von jeder Lösung wurden jeweils 2 μL in ein 200 μL PCR-Tube gegeben und bei 25 °C für 1 h inkubiert. Im Anschluss wurde den Reaktionsansätzen eine 10 %ige Lösung von 9-Aminoacridin in Aceton zugegeben. Das Aminoacridin dient als Matrix für die spätere MALDI-TOF-MS-Messung. Die MALDI-TOF-MS Messung wurde auf einem Bruker Autoflex durchgeführt. Dafür wurde 1 μL jeder Reaktion auf einen MALDI-Stahl-Target gegeben und im Negativmodus gemessen.

5.6.9 Massenspektrometrische Untersuchung von Proteinen

MALDI-TOF-MS: Für die Analytik des γ - $^{18}\text{O}_4$ -ATP-Austausch-Assays wurde eine MALDI-TOF-MS-Messung auf einem Bruker Autoflex durchgeführt. Die Detektion erfolgte im Negativmodus bei m/z 450 bis 1200. Die Spektren wurden mit dem Programm Flex analysis von Bruker analysiert. Der prozentuale Austausch wurde mit folgender Formel berechnet.

$$\text{Austausch [\%]} = (100/0,833) * ^{16}\text{O} / (^{18}\text{O} + ^{16}\text{O})$$

$(^{18}\text{O} + ^{16}\text{O})$ ist die Summe aller Integrale der ATP Peaks mit einer m/z von 506, 508, 510, 512, 514, 528, 530, 532, 534, 536. Das Integral bei m/z 506 entspricht dem Integral des $^{16}\text{O}_4$ -ATP. Da das molare Verhältnis von PPI zu γ - $^{18}\text{O}_4$ -ATP 5:1 beträgt, entspricht ein gemessener Austausch von 83,33 % einer Aktivierung von 100 %. Dadurch enthält die Formel den Ausgleichsfaktor 0,833.

5.7 Verwendete Chemikalien

Folgende Chemikalien wurden im Laufe dieser Arbeit verwendet (siehe Tab 32):

Chemikalie	Firma
Acrylamid	Roth Chemie GmbH (Karlsruhe, Deutschland)
Aceton	Roth Chemie GmbH (Karlsruhe, Deutschland)
Agar	Fluka Chemie GmbH (Buchs, Schweiz)
CaCl ₂ x 2H ₂ O	Merck KGaA (Darmstadt, Deutschland)
Chloroform	Roth Chemie GmbH (Karlsruhe, Deutschland)
DMSO	Roth Chemie GmbH (Karlsruhe, Deutschland)
Ethanol 99,8 %	Roth Chemie GmbH (Karlsruhe, Deutschland)
Ethidium Bromid	Roth Chemie GmbH (Karlsruhe, Deutschland)
Glycerol	Roth Chemie GmbH (Karlsruhe, Deutschland)
Hefeextrakt	Fluka Chemie GmbH (Buchs, Schweiz)
Imidazol	Roth Chemie GmbH (Karlsruhe, Deutschland)
Isopropanol	Roth Chemie GmbH (Karlsruhe, Deutschland)
KCl	Roth Chemie GmbH (Karlsruhe, Deutschland)
MgCl 25 mM aus <i>Taq</i> - Polymerase Kit	Promega GmbH (Mannheim, Deutschland)
MgCl ₂ x 6H ₂ O	Merck KGaA (Darmstadt, Deutschland)
MgSO ₄ x 7H ₂ O	Merck KGaA (Darmstadt, Deutschland)
NaCl	Merck KGaA (Darmstadt, Deutschland)
NaHCO ₃	Roth Chemie GmbH (Karlsruhe, Deutschland)
NaOH	Merck KGaA (Darmstadt, Deutschland)
N,N,N,N-Tetraethylen- diamin (TEMED)	Roth Chemie GmbH (Karlsruhe, Deutschland)
Ni-NTA-Agarose	Qiagen GmbH (Hilden, Deutschland)
pegGOLD Agarose	PEQLAB Biotechnologie GmbH (Erlangen, Deutschland)
SDS	Roth Chemie GmbH (Karlsruhe, Deutschland)
Tris	Roth Chemie GmbH (Karlsruhe, Deutschland)
Trypton	Roth Chemie GmbH (Karlsruhe, Deutschland)
γ - ¹⁸ O ₄ -ATP	Cambridge Isotope laboratories, Inc. (Andover, USA)

Tabelle 32: Verwendete Chemikalien.

5.8 Verwendete Enzyme

Folgende Enzyme wurden im Laufe dieser Arbeit verwendet (siehe Tab. 33):

Enzym	Firma
Restriktionsendonucleasen	New England Biolabs GmbH (Frankfurt, Deutschland)
Restriktionsendonucleasen	Promega GmbH (Mannheim, Deutschland)
<i>Taq</i> -Polymerase	Promega GmbH (Mannheim, Deutschland)
<i>Pfu</i> -Polymerase	Promega GmbH (Mannheim, Deutschland)
T4 Ligase	Promega GmbH (Mannheim, Deutschland)
Alkalische Phosphatase	

Tabelle 33: Verwendete Enzyme.

5.9 Verwendete Kits und Standards

Folgende Kits und Standards wurden im Laufe dieser Arbeit verwendet (siehe Tab. 34):

Kit	Firma
DNA Clean and Concentrator TM	Zymo Research Corporation (Freiburg, Deutschland)
dNTP	Promega GmbH (Mannheim, Deutschland)
Gene Ruler DNA Ladder Mix	Fermentas GmbH
Gene Ruler 1kb plus DNA ladder	Fermentas GmbH
NuPAGE LDS Sample Buffer (4X)	Thermo Fisher Scientific (Waltham, USA)
Page ruler Unstained	Thermo Fisher Scientific (Waltham, USA)
pGEM-T Vector System	Promega GmbH (Mannheim, Deutschland)
PureYield TM	Promega GmbH
Plasmid Miniprep System	(Mannheim, Deutschland)
QIAprep Spin Miniprep Kit	Qiagen GmbH (Hilden, Deutschland)
QIAquick Gel Extraction Kit	Qiagen GmbH (Hilden, Deutschland)
Vivaspin 500 30 MWCO	Satorius (Göttingen, Deutschland)
Wizard SV Gel and PCR clean-up Kit	Promega GmbH (Mannheim, Deutschland)

Tabelle 34: Verwendete Kits und Standards.

5.10 Verwendete Geräte

Folgende Geräte wurden im Laufe dieser Arbeit verwendet (siehe Tab. 35):

Gerät	Firma
Autoklav	Millipore GmbH (Schwalbach, Deutschland)
Branson Sonifier 250	G. Heinemann Ultraschall- und Labortechnik (Schwäbisch Gmünd, Deutschland)
Biometra T3000 Thermocycler	Biometra GmbH (Göttingen, Deutschland)
Zentrifuge Heraeus Biofuge fresco	Thermo Fisher Scientific (Waltham, USA)
Zentrifuge Heraeus Contifuge Stratos	Thermo Fisher Scientific (Waltham, USA)
Zentrifugengefäße (15/50 mL)	TPP AG (Trasadingen, Deutschland)
Eppendorf Zentrifuge 5415	Eppendorf (Hamburg, Deutschland)
Eppendorf Tubes (0,5, 1,5, 2 mL)	Eppendorf (Hamburg, Deutschland)
Gelkammer Horizon 58 und 11.14	Life Technologies (Karlsruhe, Deutschland)
Inkubator	Memmert GmbH und Co. KG (Staufen, Deutschland)
Inolab pH-Meter	WTW GmbH (Weilheim, Deutschland)
Intas iX imager	Intas Science Imaging Instruments GmbH (Göttingen, Deutschland)
Laminar Airflow Sterilbank BSB 4A (Hera Safe, Klasse II)	Heraeus, (Hanau, Deutschland)
Milli-Q Wassersystem	Millipore (Eschborn, Deutschland)
Multitron Inkubations-Schüttler	IKA Werke GmbH und Co. KG (Staufen, Deutschland)
Nalgene Kryoröhrchen	Nalgene Nunc international (Rochester, USA)
Waage (Satorius BL 3100)	Satorius AG (Göttingen, Deutschland)
Waage (Satorius BP 221S)	Satorius AG (Göttingen, Deutschland)

Tabelle 35: Verwendete Geräte.

Abkürzungsverzeichnis

A-Domäne	Adenylierungs-Domäne
AMP	Adenosinmonophosphat
ATP	Adenosintriphosphat
Ala	Alanin
AS	Aminosäure
Asp	Asparaginsäure
antiSMASH	antibiotic & secondary Metabolite Analysis Shell
ANL	Acyl-CoA Synthetasen, NRPS-Adenylierungsdomänen und Luciferase Enzyme
BLAST	Basic Local Alignment search Tool
bp	Basenpaare
BGC	Biosyntheseegencluster
<i>bcfrs</i>	Bezeichnung für das Biosyntheseegencluster von <i>Candidatus Burkholderia crenata</i>
C-Domäne	Kondensierungsdomäne
<i>cvfrs</i>	Bezeichnung für das Biosyntheseegencluster von <i>Chromobacterium vaccinii</i>
CDA	Calcium Dependent Antibiotic
COM-Domäne	communication-mediating domains
DMSO	Dimethylsulfoxid

DNA	Desoxyribonucleinsäure
Dha	Dehydroalanin
EDTA	Ethylendiamintetraessigsäure
FR	FR900359
<i>frs</i>	Bezeichnung für die FR-Biosynthesegencluster
G_qPCRs	G _q protein coupled receptors
Gly	Glycin
Gln	Glutamin
¹H-NMR	nuclear magnetic resonance spectroscopy
hHMM	hidden Markov Modell
IPTG	Isopropyl- β -D-thiogalaktosid
kDa	Kilodalton
Lys	Lysin
Leu	Leucin
MT-Domäne	Methyltransferase-Domäne
MALDI	Matrix assisted Laser Desorption/Ionisation
MLP	MbtH like Protein
MWCO	Molecular weight cut-off
Ni-NTA	Nickel Nitrilotriessigsäure
NRPS	Nichtribosomale Peptidsynthetase
PKS	Polyketid Synthetase
PCP	Peptidyl Carrier Protein

PCR	Polymerase Chain Reaction
PPi	Pyrophosphat
PTX	Pertussistoxin
SAM	S-Adenosyl-Methionin
SDS	Sodium dodecylsulfate
SDS-PAGE	sodium dodecylsulfate polyacrylamide gelelectrophoresis
Ser	Serin
TOF	Time of flight
Thr	Threonin
TEMED	Tetramethylethylendiamin
TE-Domäne	Thioesterase Domäne
ÜNK	Übernachtkultur
YM	YM-254890

Abbildungsverzeichnis

- 1 Das Depsipeptid FR besteht aus acht größtenteils ungewöhnlichen Aminosäuren, darunter drei Hydroxyleucine mit unterschiedlichen Seitenketten, N-Methyldehydroalanin, N,O-Dimethylthreonin und die Carbonsäure Phenyllaktat. Ein N-Methylalanin und Alanin sind auch Teil des Peptid-Zyklus. Die rot markierten Stellen zeigen die Unterschiede von YM und FR auf. 8
- 2 Struktur des mit FR verwandten YM-254890. YM unterscheidet sich lediglich dadurch, dass es eine Acetyl- anstatt einer Propionylseitenkette, und eine Methyl- anstatt einer Isopropylgruppe enthält. Die rot markierten Stellen zeigen die Unterschiede von YM und FR auf. 8
- 3 *A. crenata* hat Einkerbungen an den Blatträndern, in denen *Cand. B. crenata* als Endosymbiont eingebettet ist. Die Pflanze wird weit verbreitet als Schmuckpflanze verwendet und hat eine lange Tradition in der chinesischen Medizin. 12
- 4 Hypothese zur Biosynthese von FR. Nach der intramolekularen Zyklisierung durch die endständige TE-Domäne aus FrsG, kommt es zur Übertragung der Hydroxyleucinseitenkette durch die TE-Domäne aus FrsA. 15
- 5 Aufbau einer NRPS: Sie besteht aus aneinandergesetzten Modulen, welche wiederum aus Domänen aufgebaut sind. Ein funktionsfähiges Modul benötigt zwingend eine A-, eine PCP- und eine C-Domäne. Optional können modifizierende Domänen wie MT- oder E-Domänen vorhanden sein. Oft werden NRPS mit einer TE-Domäne abgeschlossen. Die Abfolge der Domänen bestimmt die Struktur des Peptids. 16
- 6 Adenylierungsreaktion von A-Domänen: Die selektierte Aminosäure wird mit Hilfe der A-Domäne unter Verbrauch von ATP zum Aminoacyladenylat aktiviert und Pyrophosphat wird abgespalten. 18

7	Thiolation: In einer zweiten Reaktion überträgt die A-Domäne die aktivierte Aminosäure auf den Phosphopantetheinarm der PCP-Domäne. Die PCP-Domäne dient als Transportdomäne für das entstehende Peptid.	18
8	Bei der Methylierung wird eine Methylgruppe (rot) durch die MT-Domäne von SAM auf die Aminogruppe übertragen. . . .	22
9	Putative Dehydroalaninbildung in NRPS: Dehydroalanin ist eine ungewöhnliche Aminosäure, die auch in NRPs nicht sehr oft vorkommt. Wahrscheinlich wird durch die vorangehende A-Domäne Serin aktiviert, welches in einem zweiten Reaktionsschritt putativ von der C-Domäne zu Dehydroalanin dehydratisiert wird.	26
10	Kondensationsreaktion: Mit Hilfe der C-Domänen wird die Peptidbindung zwischen dem entstehenden Peptid des vorangehenden Moduls und der aktivierten Aminosäure des folgenden Moduls geknüpft.	28
11	Phosphopantetheinylierung: An die inaktive PCP-Domäne wird unter Mitwirkung einer PPTase ein Phosphopantetheinrest an einen hochkonservierten Serinrest des Apo-PCP gebunden. In dieser Form ist die PCP-Domäne in der Lage, sowohl entstehende Peptide als auch aktivierte Aminosäuren kovalent zu binden.	29
12	Durch Hydrolyse oder Aminolyse bei linearen Peptiden oder eine intramolekulare Zyklisierungsreaktion wird das fertige Peptid freigesetzt.	33
13	Hypothetische Reaktion von FrsC: FrsC katalysiert mutmaßlich die Reduktion von Phenylpyruvat zu Phenyllaktat.	45
14	Hypothetische Reaktion von FrsH: FrsH weist bioinformatisch eine Ähnlichkeit mit einer β -Hydroxylase aus der Chloramphenicolbiosynthese auf. Dort katalysiert CmlA die Hydroxylierung des PCP-gebundenen <i>para</i> -Aminophenylalanin [Makris et al., 2010].	46

15	Der <i>frs</i> -BGC-Vergleich: In gleichen Farben dargestellte und gleich nummerierte Bereiche innerhalb eines BGCs kennzeichnen identische Sequenzabschnitte. Die Farbgebung wurde für beide BGC gleich gewählt. Beide BGC weisen sehr ähnliche sich wiederholende Sequenzabschnitte auf. Bei <i>C. vaccinii</i> sind Unterbrechungen zwischen den Bereichen 4, 6 und 7, während bei <i>Cand. B. crenata</i> im Übergang der Gene <i>frsE</i> zu <i>frsF</i> und <i>frsF</i> zu <i>frsG</i> durchgehend identische Bereiche sind.	50
16	Vergleich der Identitäten der A-Domänen auf Proteinebene innerhalb des <i>cvfrs</i> -BGCs von <i>C. vaccinii</i> . Die Lokalisierung der A-Domänen ist für beide <i>frs</i> -BGC gleich. Die Identitäten in dieser Abbildung entsprechen denen von <i>C. vaccinii</i>	54
17	Alignment der Core-A9- und Core-A10-Region aus <i>Cand. B. crenata</i> . Sequenzen FrsG2 und FrsG3 zeigen den Bereich der zwei Subdomänen der A-Domäne 8.	58
18	Alignment der Core-A9- und Core-A10-Region aus <i>C. vaccinii</i> . Sequenzen FrsG2 und FrsG3 zeigen den Bereich der zwei Subdomänen der A-Domäne 8.	58
19	Die erste A-Domäne von FrsG aus Modul 7 ist eine einfache A-Domäne ohne MT-Domäne. Hier liegen die Core-A9- und Core-A10-Regionen unmittelbar nach der Core-A8-Region. Die A-Domäne ist nicht unterbrochen. Die zweite A-Domäne wird von zwei MT-Domänen unterbrochen. Die Core-A9 und Core-A10-Regionen, die sich in der A-Subdomäne befinden, existieren zweimal, jeweils am Ende jeder A-Domäne.	59
20	Vergleich der Identität der MT-Domänen innerhalb des BGCs von <i>C. vaccinii</i>	60
21	MT-Domänen im Vergleich der <i>frs</i> -BGC aus <i>Cand. B. crenata</i> und <i>C. vaccinii</i>	61
22	Vergleich der Identitäten der C-Domänen, welche im <i>cvfrs</i> -BGC von <i>C. vaccinii</i> kodiert sind.	62

- 23 Alignment der C-Domänen C6 und C8 aus *Cand. B. crenata*. Vergleicht man dieses Alignment mit dem Alignment der C-Domänen C6 und C8 aus *C. vaccinii* (siehe Abb. 24), so stellt man fest, dass sich die C-Domänen C6 und C8 in beiden Alignments im vorderen Teil sehr ähnlich sind und im hinteren Bereich voneinander abweichen. Die farblich markierten Aminosäuren entsprechen den Aminosäuren, die sich bei *C. vaccinii* und *Cand. B. crenata* auf die gleiche Art unterscheiden. Unpolare Aminosäuren wurden gelb und polare ungeladene Aminosäuren hellblau gekennzeichnet. Positiv geladene Aminosäuren wurden grün gekennzeichnet. Negativ geladene Aminosäuren wurden dunkelblau markiert. 65
- 24 Alignment der C-Domänen C6 und C8 aus *C. vaccinii*. Vergleicht man dieses Alignment mit dem Alignment der C-Domänen C6 und C8 aus *Cand. B. crenata* (siehe Abb. 23), so stellt man fest, dass sich die C-Domänen C6 und C8 in beiden Alignments im vorderen Teil sehr ähnlich sind und im hinteren Bereich voneinander abweichen. Im Gegensatz zu *Cand. B. crenata* gibt es bei *C. vaccinii* einen Bereich in der Mitte der C-Domänen, in dem die Proteinsequenzen voneinander abweichen (siehe hellgrüner Bereich). Die farblich markierten Aminosäuren entsprechen den Aminosäuren, die sich bei *C. vaccinii* und *Cand. B. crenata* auf die gleiche Art unterscheiden. Unpolare Aminosäuren wurden gelb und polare ungeladene Aminosäuren hellblau gekennzeichnet. Positiv geladene Aminosäuren wurden grün markiert und negativ geladene Aminosäuren wurden dunkelblau gekennzeichnet. 66

- 25 3D-Modell der C6-Domäne von *C. vaccinii*, erstellt mit „Phyre 2“ ; seitliche Ansicht auf den Tunnel, der durch die beiden Subdomänen entsteht. Die Aminosäuresequenz der C-Domäne ist in Spektralfärbung dargestellt. Der blaue Bereich markiert den N-Terminus, der rote Bereich den C-terminalen Teil des Peptids. Die Aminosäuren des konservierten HDxxxDG-Motivs sind als Van-der-Waals Modell dargestellt, die putativ für die Dehydratisierung von Serin verantwortlichen Aminosäuren als „sticks-and-ball“ -Modell. Bei einer seitlichen Ansicht auf die V-förmige Tasche, erkennt man, dass sich die Aminosäuren am oberen Rand der Tasche befinden. 68
- 26 3D-Modell der C6-Domäne von *C. vaccinii*, erstellt mit „Phyre 2“ ; Unteransicht auf den Tunnel, der durch die beiden Subdomänen entsteht. Die Aminosäuresequenz der C-Domäne ist in Spektralfärbung dargestellt. Der blaue Bereich markiert den N-Terminus, der rote Bereich den C-terminalen Teil des Peptids. Die Aminosäuren des konservierten HDxxxDG-Motivs sind nicht dargestellt, die putativ für die Dehydratisierung von Serin verantwortlichen Aminosäuren sind als “sticks-and-ball“-Modell dargestellt und mit Positionsangaben beschriftet. . . . 69
- 27 3D-Modell der C6-Domäne von *C. vaccinii*, erstellt mit „Phyre 2“ ; Unteransicht auf den Tunnel, der durch die beiden Subdomänen entsteht. Die Aminosäuresequenz der C-Domäne ist in Spektralfärbung dargestellt. Der blaue Bereich markiert den N-Terminus, der rote Bereich den C-terminalen Teil des Peptids. Die Aminosäuren des konservierten HDxxxDG-Motivs sind im Hintergrund durch Van-der-Waals-Darstellung gekennzeichnet, die putativ für die Dehydratisierung von Serin verantwortlichen Aminosäuren sind als “sticks-and-ball“ -Modell dargestellt. 70

28	PCP-Domänen des <i>cvfrs</i> -BGCs. Die PCP-Domänen wurden blau markiert. PCP-Domänen, die sich stärker gleichen in dunkelblau, die anderen etwas heller. Die Prozepte geben die prozentuale Identität der Domänen untereinander an.	71
29	TE-Domänen des <i>cvfrs</i> -BGCs.	72
30	Die Adenylierungsreaktion von A-Domänen erfolgt unter ATP-Verbrauch. Diese Reaktion ist reversibel.	73
31	Die Aminosäure wird mit Hilfe der A-Domäne unter Verbrauch von ATP zum Aminoacyladenylat aktiviert, um im Anschluss auf den Phosphopantethein-Arm der PCP-Domäne geladen zu werden.	74
32	Austauschreaktion des A-Domänen Assay. Nach Abspaltung der Aminosäure vom AMP wird aus AMP und nicht markiertem PPI ein ATP gebildet, das kein $^{18}\text{O}_4$ -Isotop mehr enthält.	75
33	Allgemeine Darstellung einer MALDI-TOF-Analyse des Reaktionsproduktes einer A-Domäne, γ - $^{18}\text{O}_4$ -ATP, $^{16}\text{O}_4$ -ATP und der entsprechenden Aminosäure in Abhängigkeit von der Zeit. Bei der Regenerierung von AMP zu ATP wird $^{16}\text{O}_4$ -PPI verwendet. Es kommt zur Bildung des ATP mit einem $^{16}\text{O}_4$ -Isotop-PPI und in der Folge zu einer Detektion des m/z : 506 Peaks anstatt des m/z : 514 Peaks des γ - $^{18}\text{O}_4$ -ATP. Die Abbildung wurde aus Phelan et al. 2009 entnommen.	76
34	Strukturen von Leucin, Isoleucin und Hydroxyleucin. Um die Toleranz der A-Domänen in FrsA und FrsD zu testen, wurden u.a. diese strukturell ähnlichen Aminosäuren getestet.	77
35	FrsA (A1-PCP) zusammen mit MLP exprimiert: Der Pfeil markiert die Position der A-Domäne. Das MLP kann nicht identifiziert werden. L=Ladder, P=Pellet, NI= nicht induzierte Fraktion, I= induzierte Fraktion vor Zugabe von Ni-NTA, FT= Filtrat nach Ni-NTA Entfernung, WI= Waschschrift I, WII= Waschschrift II, EI+EII= Elutionsfraktionen, K= vereinigte, aufkonzentrierte Elutionsfraktionen	80

36	FrsD (A2-PCP) zusammen mit MLP exprimiert: Der Pfeil markiert die Position der A-Domäne. Das MLP kann nicht identifiziert werden. L=Ladder, P=Pellet, NI= nicht induzierte Fraktion, I= induzierte Fraktion vor Zugabe von Ni-NTA, FT=Filtrat nach Ni-NTA Entfernung, WI=Waschschritt I, WII= Waschschritt II, EI+EII=Elutionsfraktionen, K=vereinigte, aufkonzentrierte Elutionsfraktionen	81
37	Massenspektrum des A-Domänen-Assay von FrsA (A1-PCP) und FrsD (A2-PCP), einmal mit und ohne MLP. Bei allen Reaktionsansätzen konnte das Massesignal des $^{16}\text{O}_4\text{-ATP}$ -Peak bei m/z : 506 detektiert werden. In allen Fällen fand eine Aktivierung von Leucin statt. Beide A-Domänen sind ohne MLP funktionsfähig.	83
38	tatsächlicher prozentualer $\gamma\text{-}^{18}\text{O}_4\text{-ATP}$ -Austausch durch die A-Domänen von FrsA (A1-PCP) und FrsD (A2-PCP) in Anwesenheit der einzelnen Aminosäuren.	86
39	Prozentualer $\gamma\text{-}^{18}\text{O}_4\text{-ATP}$ -Austausch durch die A-Domänen FrsA (A1-PCP) und FrsD (A2-PCP) in Anwesenheit der einzelnen Aminosäuren bezogen auf Leucin.	89
40	Kristallstrukturanalyse der A-MT-Domäne aus TioS des Thio-coralin-BGC. Die Abb. ist entnommen aus Mori et al. 2018. Der hellgelbe Bereich stellt die A-Core-Domäne dar, mit dem assoziierten MLP in grau. Der dunkelgelbe Bereich ist die A-Subdomäne. In rot der N-terminale katalytische Bereich der MT-Domäne und in violett die C-terminale SirA-like Region. .	96
41	A. Schematische Darstellung der A-Domäne A8 aus FrsG mit zwei MT-Domänen, B. Vergleich mit TioS als Beispiel für eine A-Domäne mit einer MT-Domäne und C. Vergleich mit einer A-Domäne ohne MT-Domäne. Die Core-Regionen A9 und A10 sind auf den A-Subdomänen lokalisiert [Conti et al., 1997] [Stachelhaus et al., 1999].	97

A1	Phylogenetische Analyse der C-Domänen mit FastTree. Die Abbildung ist entnommen aus Crüsemann et al. 2018. Die Farben markieren die einzelnen C-Domänen Klassen. Die roten Umrandungen zeigen die C-Domänen des <i>bcfrs</i> -BGCs am Beispiel von <i>Cand. B. crenata</i>	169
A2	Vektorkarte A1-PCP in pET28a.	170
A3	Vektorkarte A2-PCP in pET28a.	171
A4	MLP (<i>frsB</i>) aus dem <i>bcfrs</i> -BGC in pCDFApra.	172
A5	Vektorkarte des pCDFApra. Der Vektor wurde von Luis Linares Otoyá aus dem pCDFDuett-Vektor erstellt.	173
A6	Alignment der C-Domänen aus McyA und der C-Domänen C6 aus dem <i>cvfrs</i> -BGC und dem <i>bcfrs</i> -BGC.	174
A7	Alignment der A-Domänen aus dem <i>bcfrs</i> -BGC.	176
A8	Percent identity Matrix des Alignments der A-Domänen aus dem <i>bcfrs</i> -BGC.	177
A9	Alignment der A-Domänen aus dem <i>cvfrs</i> -BGC.	179
A10	Percent Identity Matrix des Alignments der A-Domänen aus dem <i>cvfrs</i> -BGC.	180
A11	Alignment der MT-Domänen aus dem <i>bcfrs</i> -BGC.	182
A12	Alignment der MT-Domänen aus dem <i>cvfrs</i> -BGC.	183
A13	Percent Identity Matrix des Alignments der MT-Domänen aus dem <i>bcfrs</i> -BGC.	184
A14	Percent Identity Matrix des Alignments der MT-Domänen aus dem <i>cvfrs</i> -BGC.	185
A15	Alignment der N-MT-Domänen aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	187
A16	Percent Identity Matrix des Alignments der N-Methyltransferasen aus dem <i>bcfrs</i> -BGC und <i>cvfrs</i> -BGC.	188
A17	Alignment der O-MT-Domänen aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	189
A18	Percent Identity Matrix des Alignments der O-Methyltransferasen aus dem <i>bcfrs</i> -BGC und <i>cvfrs</i> -BGC.	189
A19	Alignment der C-Domänen aus dem <i>cvfrs</i> -BGC.	191

A20	Percent Identity Matrix des Alignments der C-Domänen aus dem <i>cvfrs</i> -BGC.	192
A21	Alignment der C-Domänen aus dem <i>bcfrs</i> -BGC.	194
A22	Percent Identity Matrix des Alignments der C-Domänen aus dem <i>bcfrs</i> -BGC.	195
A23	Alignment der PCP-Domänen aus dem <i>cvfrs</i> -BGC.	196
A24	Percent Identity Matrix des Alignments der PCP-Domänen aus dem <i>cvfrs</i> -BGC.	197
A25	Alignment der TE-Domänen aus dem <i>cvfrs</i> -BGC.	198
A26	Percent Identity Matrix des Alignments der TE-Domänen aus dem <i>cvfrs</i> -BGC.	198
A27	Alignment von <i>frsA</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	203
A28	Alignment von <i>frsB</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	203
A29	Alignment von <i>frsC</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	205
A30	Alignment von <i>frsD</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	209
A31	Alignment von <i>frsE</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	219
A32	Alignment von <i>frsF</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	227
A33	Alignment von <i>frsG</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	238
A34	Alignment von <i>frsH</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	240
A35	Percent Identity Matrices der Alignment von <i>frsA-frsH</i> aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	242
A36	Alignment von FrsA aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	244
A37	Alignment von FrsB aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	244
A38	Alignment von FrsC aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	245
A39	Alignment von FrsD aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	247
A40	Alignment von FrsE aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	251
A41	Alignment von FrsF aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	254
A42	Alignment von FrsG aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	258
A43	Alignment von FrsH aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	259
A44	Percent Identity Matrices der Alignments von FrsA-FrsH aus dem <i>bcfrs</i> -BGC und dem <i>cvfrs</i> -BGC.	261

Tabellenverzeichnis

- 1 A-Domänen Spezifitätsvorhersage nach unterschiedlichen Algorithmen durch antiSMASH von *Cand. B. crenata*. Für die A-Domänen von FrsD-FrsG werden jeweils in gleicher Reihenfolge Leucin oder Hydroxyleucin (Leu/OH-Leu), Phenylalanin (Phe) oder Tryptophan (Trp), Serin (Ser), Alanin (Ala), Alanin, Leucin oder Hydroxyleucin und Threonin (Thr) vorhergesagt. Der pHMM-Algorithmus weicht bei FrsA, D und G1 von den Vorhersagen der anderen Algorithmen ab. Schwierigkeiten in der Vorhersage gibt es bei FrsE1. Das Substrat konnte hier nicht eindeutig identifiziert werden, wobei die Aromatizität erkannt wurde. Abkürzungen: boh-d-Leu= β -Hydroxyleucin; bht= Hydroxytyrosin [Blin et al., 2013]; N/A = keine Angabe 42
- 2 A-Domänen Spezifitätsvorhersage nach unterschiedlichen Algorithmen durch antiSMASH von *C. vaccinii*. Für die A-Domänen von FrsD-FrsG werden jeweils in gleicher Reihenfolge Leucin (Leu), Phenylalanin (Phe) oder Tryptophan (Trp), Serin (Ser), Alanin (Ala), Alanin, Leucin und Threonin (Thr) vorhergesagt. Der pHMM-Algorithmus weicht bei FrsA, D und G1 von den Vorhersagen der anderen Algorithmen ab. Schwierigkeiten in der Vorhersage gibt es bei FrsE1. Das Substrat konnte hier nicht eindeutig identifiziert werden, wobei die Aromatizität erkannt wurde. Abkürzungen: bht= Hydroxytyrosin [Blin et al., 2013] ; N/A = keine Angabe 43

3	Vergleich der BGC von <i>Cand. B. crenata</i> und <i>C. vaccinii</i> auf DNA und Proteinebene. bp=Basenpaare, Diff. bp= Basenpaardifferenz der Gene beider BGC zueinander; AS=Aminosäure, Ident DNA=Identität der Gene von <i>C. vaccinii</i> und <i>Cand. B. crenata</i> zueinander in Prozent; Ident Prot=Identität von FrsA-FrsH von <i>C. vaccinii</i> und <i>Cand. B. crenata</i> auf Proteinebene zueinander in Prozent; GCBurk und GCChro=GC-Gehalt der Gene <i>frsA-frsH</i> von <i>C. vaccinii</i> und <i>Cand. B. crenata</i> in Prozent.	48
4	Sequenzbereiche, die im <i>cvfrs</i> -BGC von <i>C. vaccinii</i> mehrfach vorkommen. Die Abkürzung bp zeigt die Länge der Bereiche in Basenpaaren an. AS zeigt die Länge des Bereichs in Aminosäuren. Bei Genen wurden alle Gene angegeben, die ganz oder zum Teil von diesen sich wiederholenden Bereichen abgedeckt werden.	52
5	Sequenzbereiche, die im <i>bcfrs</i> -BGC von <i>Cand. B. crenata</i> mehrfach vorkommen. Die Abkürzung bp zeigt die Länge der Bereiche in Basenpaaren an. AS zeigt die Länge des Bereichs in Aminosäuren. Bei Genen wurden alle Gene angegeben, die ganz oder zum Teil von diesen sich wiederholenden Bereichen abgedeckt werden.	52
6	Nichtribosomaler Code der A-Domänen des <i>bcfrs</i> -BGC aus <i>Cand. B. crenata</i> für die A-Domänen A1-A8 (siehe auch Abb. 4, 16, A7 und A9).	56
7	Nichtribosomaler Code der A-Domänen des <i>cvfrs</i> -BGC aus <i>C. vaccinii</i> für die A-Domänen A1-A8 (siehe auch Abb. 4, 16, A7 und A9).	56
8	Verwendete Primer mit den dazugehörigen Schnittstellen. . . .	77

9	Proteinkonzentrationen der Wasch- und Elutionsfraktionen der exprimierten A-Domänen in [mg/mL]. Die Proben wurden auf einem Nanodrop-Gerät der Firma Thermo Scientific gemessen. WI und WII sind Waschschriffe mit jeweils 25 mM und 50 mM Imidazolpuffer. EI, EII, EIII sind die Elutionen mit einem 300 mM Imidazolpuffer. Konz. entspricht der mit der Vivaspin 500 Säule aufkonzentrierten Fraktion.	78
10	Zu erwartende Größe der exprimierten Proteine	79
11	Berechneter absoluter γ - $^{18}\text{O}_4$ -ATP-Austausch durch verschiedene Aminosäuren im A-Domänen-Assay für FrsA (A1-PCP) und FrsD (A2-PCP). k.A.=es konnte keine Auswertung gemacht werden; 0=keine Aktivierung vorhanden	85
12	Prozentualer γ - $^{18}\text{O}_4$ -ATP-Austausch durch die Aminosäuren im A-Domänen-Assay in Relation zur Leucin-Aktivierung von FrsA (A1-PCP) und FrsD (A2-PCP). k.A.=es konnte keine Auswertung gemacht werden; 0=keine Aktivierung vorhanden	88
13	Verwendete Vektoren	104
14	Verwendete Organismen	105
15	Verwendete Medien	107
16	Proteinaufreinigungspuffer	108
17	Plasmidpräparationspuffer	108
18	Assaypuffer	109
19	Kathoden- und Anodenpuffer	110
20	Weitere Puffer und Lösungen, die verwendet wurden.	110
21	Für diese Arbeit verwendete Antibiotikastocks.	111
22	Bedingungen zur Kultivierung von Bakterien.	112
23	Reaktionsansatz für die PCR mit <i>Taq</i> -Polymerase.	114
24	Reaktionsansatz für die PCR mit <i>Pfu</i> -Polymerase.	115
25	Temperaturprogramm für die PCR.	115
26	Sequenzen der verwendeten Primer. Die jeweiligen Schnittstellen sind fett gedruckt.	117
27	Ligationsansatz	120

28	Konstrukte, die in dieser Arbeit verwendet wurden. Der pS-MART BAC clone 19E wurde von Aurelien Carlier zur Verfügung gestellt [Carlier et al., 2016].	122
29	Expressionsstämme und Coexpressionsstämme, die in dieser Arbeit verwendet wurden.	123
30	Trenngel	126
31	Sammelgel	126
32	Verwendete Chemikalien.	129
33	Verwendete Enzyme.	130
34	Verwendete Kits und Standards.	130
35	Verwendete Geräte.	131

Literatur

- Abe, T., Kobayashi, K., Kawamura, S., Sakaguchi, T., Shiiba, K., and Kobayashi, M. (2018). Dipeptide synthesis by internal adenylation domains of a multidomain enzyme involved in nonribosomal peptide synthesis. *J. Gen. Appl. Microbiol.*
- Al-Mestarihi, A. H., Villamizar, G., Fernandez, J., Zolova, O. E., Lombo, F., and Garneau-Tsodikova, S. (2014). Adenylation and S-methylation of cysteine by the bifunctional enzyme TioN in thiocoraline biosynthesis. *J. Am. Chem. Soc.*, 136(49):17350–17354.
- Ankenbauer, R. G., Staley, A. L., Rinehart, K. L., and Cox, C. D. (1991). Mutasynthesis of siderophore analogues by *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.*, 88(5):1878–1882.
- Ansari, M. Z., Sharma, J., Gokhale, R. S., and Mohanty, D. (2008). In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC Bioinformatics*, 9:454.
- Beacco, E., Bianchi, M. L., Minghetti, A., and Spalla, C. (1978). Directed biosynthesis of analogues of ergot peptide alkaloids with *Claviceps purpurea*. *Experientia*, 34(10):1291–1293.
- Belshaw, P. J., Walsh, C. T., and Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, 284(5413):486–489.
- Bergendahl, V., Linne, U., and Marahiel, M. A. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.*, 269(2):620–629.
- Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.*, 62(3):293–300.
- Billich, A. and Zocher, R. (1987). Enzymatic synthesis of cyclosporin A. *J. Biol. Chem.*, 262(36):17258–17259.

- Bister, B., Bischoff, D., Nicholson, G. J., Stockert, S., Wink, J., Brunati, C., Donadio, S., Pelzer, S., Wohlleben, W., and Süßmuth, R. D. (2003). Bromobalhimycin and chlorobromobalhimycins—illuminating the potential of halogenases in glycopeptide antibiotic biosyntheses. *Chembiochem*, 4(7):658–662.
- Blad, C. C., Tang, C., and Offermanns, S. (2012). G protein-coupled receptors for energy metabolites as new therapeutic targets. *Nat Rev Drug Discov*, 11(8):603–619.
- Blin, K., Kim, H. U., Medema, M. H., and Weber, T. (2017a). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Brief. Bioinformatics*.
- Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. (2013). antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.*, 41(Web Server issue):W204–212.
- Blin, K., Medema, M. H., Kottmann, R., Lee, S. Y., and Weber, T. (2017b). The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.*, 45(D1):D555–D559.
- Bloudoff, K., Alonzo, D. A., and Schmeing, T. M. (2016). Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem Biol*, 23(3):331–339.
- Bloudoff, K., Fage, C. D., Marahiel, M. A., and Schmeing, T. M. (2017). Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc. Natl. Acad. Sci. U.S.A.*, 114(1):95–100.
- Bloudoff, K., Rodionov, D., and Schmeing, T. M. (2013). Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J. Mol. Biol.*, 425(17):3137–3150.

- Boll, B., Taubitz, T., and Heide, L. (2011). Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J. Biol. Chem.*, 286(42):36281–36290.
- Bozhüyük, K. A. J.yuek, K. A. J., Fleischhacker, F., Linck, A., Wesche, F., Tietze, A., Niesert, C. P., and Bode, H. B. (2018). De novo design and engineering of non-ribosomal peptide synthetases. *Nat Chem*, 10(3):275–281.
- Calcott, M. J., Owen, J. G., Lamont, I. L., and Ackerley, D. F. (2014). Biosynthesis of novel Pyoverdines by domain substitution in a nonribosomal peptide synthetase of *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.*, 80(18):5723–5731.
- Carrier, A., Fehr, L., Pinto-Carbo, M., Schäberle, T., Reher, R., Dessen, S., König, G., and Eberl, L. (2016). The genome analysis of *Candidatus Burkholderia crenata* reveals that secondary metabolism may be a key function of the *Ardisia crenata* leaf nodule symbiosis. *Environ. Microbiol.*, 18(8):2507–2522.
- Chatterjee, J., Rechenmacher, F., and Kessler, H. (2013). N-methylation of peptides and proteins: an important element for modulating biological functions. *Angew. Chem. Int. Ed. Engl.*, 52(1):254–269.
- Chen, R. H., Buko, A. M., Whittern, D. N., and McAlpine, J. B. (1989). Paucidamycins, a novel series of antibiotics with anti-*Pseudomonas aeruginosa* activity. II. Isolation and structural elucidation. *J. Antibiot.*, 42(4):512–520.
- Clugston, S. L., Sieber, S. A., Marahiel, M. A., and Walsh, C. T. (2003). Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. *Biochemistry*, 42(41):12095–12104.
- Coeffet-Le Gal, M. F., Thurston, L., Rich, P., Miao, V., and Baltz, R. H. (2006). Complementation of daptomycin dptA and dptD deletion mutati-

- ons in trans and production of hybrid lipopeptide antibiotics. *Microbiology (Reading, Engl.)*, 152(Pt 10):2993–3001.
- Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.*, 16(14):4174–4183.
- Crüseman, M. Kohlhaas, C. P. J. (2013). Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chemical Science*.
- Crüseman, M., Reher, R., Schamari, I., Brachmann, A. O., Ohbayashi, T., Kuschak, M., Malfacini, D., Seidinger, A., Pinto-Carbo, M., Richarz, R., Reuter, T., Kehraus, S., Hallab, A., Attwood, M., Schioth, H. B., Mergaert, P., Kikuchi, Y., Schäberle, T., Kostenis, E., Wenzel, D., E., M. C., Piel, J., Carlier, A., Eberl, L., and König, G. M. (2018). Heterologous Expression, Biosynthetic Studies, and Ecological Function of the Selective Gq-Signaling Inhibitor FR900359. *Angew. Chem. Int. Ed. Engl.*, 57(3):836–840.
- De Crecy-Lagard, V., Marliere, P., and Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *C. R. Acad. Sci. III, Sci. Vie*, 318(9):927–936.
- Deboer, C. und DIETZ, A., SAVAGE, G. M., and SILVER, W. S. (1955). Streptolydigin, a new antimicrobial antibiotic. I. Biologic studies of streptolydigin. *Antibiot Annu*, 3:886–892.
- Deshpande, D. A. and Penn, R. B. (2006). Targeting G protein-coupled receptor signaling in asthma. *Cell. Signal.*, 18(12):2105–2120.
- Doekel, S. and Marahiel, M. A. (2000). Dipeptide formation on engineered hybrid peptide synthetases. *Chem. Biol.*, 7(6):373–384.
- Dorsam, R. T. and Gutkind, J. S. (2007). G-protein-coupled receptors and cancer. *Nat. Rev. Cancer*, 7(2):79–94.

- Dowling, D. P., Kung, Y., Croft, A. K., Taghizadeh, K., Kelly, W. L., Walsh, C. T., and Drennan, C. L. (2016). Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U.S.A.*, 113(44):12432–12437.
- Druey, K. M. (2009). Regulation of G-protein-coupled signaling pathways in allergic inflammation. *Immunol. Res.*, 43(1-3):62–76.
- Ehmann, D. E., Trauger, J. W., Stachelhaus, T., and Walsh, C. T. (2000). Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem. Biol.*, 7(10):765–772.
- Elshahawi, S. I., Trindade-Silva, A. E., Hanora, A., Han, A. W., Flores, M. S., Vizzoni, V., Schrago, C. G., Soares, C. A., Concepcion, G. P., Distel, D. L., Schmidt, E. W., and Haygood, M. G. (2013). Boronated tartrolon antibiotic produced by symbiotic cellulose-degrading bacteria in shipworm gills. *Proc. Natl. Acad. Sci. U.S.A.*, 110(4):295–304.
- Epstein, S. C., Charkoudian, L. K., and Medema, M. H. (2018). A standardized workflow for submitting data to the Minimum Information about a Biosynthetic Gene cluster (MIBiG) repository: prospects for research-based educational experiences. *Stand Genomic Sci*, 13:16.
- Felnagle, E. A., Barkei, J. J., Park, H., Podevels, A. M., McMahon, M. D., Drott, D. W., and Thomas, M. G. (2010). MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*, 49(41):8815–8817.
- Fernandes, P. B., Swanson, R. N., Hardy, D. J., Hanson, C. W., Coen, L., Rasmussen, R. R., and Chen, R. H. (1989). Pacidamycins, a novel series of antibiotics with anti-*Pseudomonas aeruginosa* activity. III. Microbiologic profile. *J. Antibiot.*, 42(4):521–526.
- Fujioka M., Koda S., M. Y. (1988). Structure of FR900359, a cyclic depsipeptide from *ardisia crenata* sims. *J. Org. Chem.*

- Garg, N., Salazar-Ocampo, L. M., and van der Donk, W. A. (2013). In vitro activity of the nisin dehydratase NisB. *Proc. Natl. Acad. Sci. U.S.A.*, 110(18):7258–7263.
- Gaudelli, N. M., Long, D. H., and Townsend, C. A. (2015). \hat{I}^2 -Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature*, 520(7547):383–387.
- Gordon, M., Cragg, D. J., Newman, ., M., K., and Snader (1997). Natural products in drug discovery and development. *J. Nat. Prod.*
- Grüschow, S., Rackham, E. J., Elkins, B., Newill, P. L., Hill, L. M., and Goss, R. J. (2009). New pacidamycin antibiotics through precursor-directed biosynthesis. *Chembiochem*, 10(2):355–360.
- Gudermann, T., Kalkbrenner, F., and Schultz, G. (1996). Diversity and selectivity of receptor-G protein interaction. *Annu. Rev. Pharmacol. Toxicol.*, 36:429–459.
- Hahn, M. and Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc. Natl. Acad. Sci. U.S.A.*, 103(2):275–280.
- Harvey, A. L. (2008). Natural products in drug discovery. *Drug Discov. Today*, 13(19-20):894–901.
- Heemstra, J. R., Walsh, C. T., and Sattely, E. S. (2009). Enzymatic tailoring of ornithine in the biosynthesis of the *Rhizobium* cyclic trihydroxamate siderophore vicibactin. *J. Am. Chem. Soc.*, 131(42):15317–15329.
- Herbst, D. A., Boll, B., Zocher, G., Stehle, T., and Heide, L. (2013). Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.*, 288(3):1991–2003.
- Hughes, A. R., Martin, M. W., and Harden, T. K. (1984). Pertussis toxin differentiates between two mechanisms of attenuation of cyclic AMP

- accumulation by muscarinic cholinergic receptors. *Proc. Natl. Acad. Sci. U.S.A.*, 81(18):5680–5684.
- Imker, H. J., Krahn, D., Clerc, J., Kaiser, M., and Walsh, C. T. (2010). N-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem. Biol.*, 17(10):1077–1083.
- Karwowski, J. P., Jackson, M., Theriault, R. J., Chen, R. H., Barlow, G. J., and Maus, M. L. (1989). Pacidamycins, a novel series of antibiotics with anti-*Pseudomonas aeruginosa* activity. I. Taxonomy of the producing organism and fermentation. *J. Antibiot.*, 42(4):506–511.
- Keating, T. A., Marshall, C. G., Walsh, C. T., and Keating, A. E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.*, 9(7):522–526.
- Keller, U. (1987). Actinomycin synthetases. Multifunctional enzymes responsible for the synthesis of the peptide chains of actinomycin. *J. Biol. Chem.*, 262(12):5852–5856.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 10(6):845–858.
- Klepac, K. e. a. (2016). The Gq signalling pathway inhibits brown and beige adipose tissue. *Nat. Commun.*
- Kopp, F. and Marahiel, M. A. (2007). Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat Prod Rep*, 24(4):735–749.
- Krause, M., Lindemann, A., Glinski, M., Hornbogen, T., Bonse, G., Jeschke, P., Thielking, G., Gau, W., Kleinkauf, H., and Zocher, R. (2001). Directed biosynthesis of new enniatins. *J. Antibiot.*, 54(10):797–804.
- Kries, H., Niquille, D. L., and Hilvert, D. (2015). A subdomain swap strategy for reengineering nonribosomal peptides. *Chem. Biol.*, 22(5):640–648.

- Labby, K. J., Watsula, S. G., and Garneau-Tsodikova, S. (2015). Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat Prod Rep*, 32(5):641–653.
- Lai, J. R., Koglin, A., and Walsh, C. T. (2006). Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry*, 45(50):14869–14879.
- Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., and Walsh, C. T. (1996). A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.*, 3(11):923–936.
- Lautru, S. and Challis, G. L. (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology (Reading, Engl.)*, 150(Pt 6):1629–1636.
- Lautru, S., Oves-Costales, D., Pernodet, J. L., and Challis, G. L. (2007). MbtH-like protein-mediated cross-talk between non-ribosomal peptide antibiotic and siderophore biosynthetic pathways in *Streptomyces coelicolor* M145. *Microbiologie (Reading, Engl.)*, 153(Pt 5):1405–1412.
- Lawen, A. and Zocher, R. (1990). Cyclosporin synthetase. The most complex peptide synthesizing multienzyme polypeptide so far described. *J. Biol. Chem.*, 265(19):11355–11360.
- Lefkowitz, R. J. (2004). Historical review: a brief history and personal retrospective of seven-transmembrane receptors. *Trends Pharmacol. Sci.*, 25(8):413–422.
- Lersten, N. and Horner, H. (1976). Bacterial leaf nodule symbiosis in angiosperms with emphasis on rubiaceae and myrsinaceae. *Bot Rev.*
- Link, A. und Müller, C. E. (2016). G-Protein-Coupled Receptors: Sustained Signaling via Intracellular Megaplexes and Pathway-Specific Drugs. *Angew. Chem. Int. Ed. Engl.*, 55(52):15962–15964.

- Linne, U. and Marahiel, M. A. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry*, 39(34):10439–10447.
- Lombo, F., Velasco, A., Castro, A., de la Calle, F., Brana, A. F., Sanchez-Puelles, J. M., Mendez, C., and Salas, J. A. (2006). Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two streptomyces species. *Chembiochem*, 7(2):366–376.
- Long, M., Betran, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.*, 4(11):865–875.
- Lundy, T. A., Mori, S., and Garneau-Tsodikova, S. (2018). Engineering Bifunctional Enzymes Capable of Adenylating and Selectively Methylating the Side Chain or Core of Amino Acids. *ACS Synth Biol*, 7(2):399–404.
- Luo, L., Burkart, M. D., Stachelhaus, T., and Walsh, C. T. (2001). Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase. *J. Am. Chem. Soc.*, 123(45):11208–11218.
- Luo, L., Kohli, R. M., Onishi, M., Linne, U., Marahiel, M. A., and Walsh, C. T. (2002). Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B. *Biochemistry*, 41(29):9184–9196.
- Luo, Y., Cobb, R. E., and Zhao, H. (2014). Recent advances in natural product discovery. *Curr. Opin. Biotechnol.*, 30:230–237.
- Machida, K., Arai, D., Katsumata, R., Otsuka, S., Yamashita, J. K., Ye, T., Tang, S., Fusetani, N., and Nakao, Y. (2018). Sameuramide A, a new cyclic depsipeptide isolated from an ascidian of the family Didemnidae. *Bioorg. Med. Chem.*, 26(13):3852–3857.

- Magarvey, N. A., Ehling-Schulz, M., and Walsh, C. T. (2006). Characterization of the cereulide NRPS alpha-hydroxy acid specifying modules: activation of alpha-keto acids and chiral reduction on the assembly line. *J. Am. Chem. Soc.*, 128(33):10698–10699.
- Makris, T. M., Chakrabarti, M., Munck, E., and Lipscomb, J. D. (2010). A family of diiron monooxygenases catalyzing amino acid beta-hydroxylation in antibiotic biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, 107(35):15391–15396.
- Marahiel, M. A. (2009). Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J. Pept. Sci.*, 15(12):799–807.
- Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.*, 97(7):2651–2674.
- Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.*, 39(Web Server issue):W339–346.
- Miao, V., Coeffet-Le Gal, M. F., Nguyen, K., Brian, P., Penn, J., Whiting, A., Steele, J., Kau, D., Martin, S., Ford, R., Gibson, T., Bouchard, M., Wrigley, S. K., and Baltz, R. H. (2006). Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics. *Chem. Biol.*, 13(3):269–276.
- Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C., and Gulick, A. M. (2016). Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.*, 291(43):22559–22571.
- Miller, D. A., Walsh, C. T., and Luo, L. (2001). C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase. *J. Am. Chem. Soc.*, 123(34):8434–8435.

- Miller, I. (1990). Bacterial leaf nodule symbiosis. *Adv Bot Res.*
- Minarik, P., Tomaskova, N., Kollarova, M., and Antalík, M. (2002). Malate dehydrogenases—structure and function. *Gen. Physiol. Biophys.*, 21(3):257–265.
- Moran, S., Rai, D. K., Clark, B. R., and Murphy, C. D. (2009). Precursor-directed biosynthesis of fluorinated iturin A in *Bacillus* spp. *Org. Biomol. Chem.*, 7(4):644–646.
- Mori, S., Garzan, A., Tsodikov, O. V., and Garneau-Tsodikova, S. (2017). Deciphering Nature’s Intricate Way of N,S-Dimethylating l-Cysteine: Sequential Action of Two Bifunctional Adenylation Domains. *Biochemistry*, 56(46):6087–6097.
- Mori, S., Pang, A. H., Lundy, T. A., Garzan, A., Tsodikov, O. V., and Garneau-Tsodikova, S. (2018). Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nat. Chem. Biol.*
- Newman, D. J. and Cragg, G. M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.*, 79(3):629–661.
- Nishimura, A., Kitano, K., Takasaki, J., Taniguchi, M., Mizuno, N., Tago, K., Hakoshima, T., and Itoh, H. (2010). Structural basis for the specific inhibition of heterotrimeric Gq protein by a small molecule. *Proc. Natl. Acad. Sci. U.S.A.*, 107(31):13666–13671.
- Ortega, M. A., Hao, Y., Zhang, Q., Walker, M. C., van der Donk, W. A., and Nair, S. K. (2015). Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature*, 517(7535):509–512.
- Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006). How many drug targets are there? *Nat Rev Drug Discov*, 5(12):993–996.
- Owen, J. G., Calcott, M. J., Robins, K. J., and Ackerley, D. F. (2016). Generating Functional Recombinant NRPS Enzymes in the Laboratory Setting via Peptidyl Carrier Protein Engineering. *Cell Chem Biol*, 23(11):1395–1406.

- Phelan, V. V., Du, Y., McLean, J. A., and Bachmann, B. O. (2009). Adenylation enzyme characterization using gamma-(18)O(4)-ATP pyrophosphate exchange. *Chem. Biol.*, 16(5):473–478.
- Pinto-Carbo, M., Gademann, K., Eberl, L., and Carlier, A. (2018). Leaf nodule symbiosis: function and transmission of obligate bacterial endophytes. *Curr. Opin. Plant Biol.*, 44:23–31.
- Powell, A., Al Nakeeb, M., Wilkinson, B., and Micklefield, J. (2007). Precursor-directed biosynthesis of nonribosomal lipopeptides with modified glutamate residues. *Chem. Commun. (Camb.)*, (26):2683–2685.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.*
- Qiagen (2003). The QIAexpressionist.
- Quadri, L. E., Sello, J., Keating, T. A., Weinreb, P. H., and Walsh, C. T. (1998). Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.*, 5(11):631–645.
- Quiagen (2012). Quiagen Plasmid Purification Handbook 04/2012.
- Rask-Andersen, M., Almen, M. S., and Schioth, H. B. (2011). Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov*, 10(8):579–590.
- Rausch, C., Hoof, I., Weber, T., Wohlleben, W., and Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.*, 7:78.
- Ray, L., Yamanaka, K., and Moore, B. S. (2016). A Peptidyl-Transesterifying Type I Thioesterase in Salinamide Biosynthesis. *Angew. Chem. Int. Ed. Engl.*, 55(1):364–367.

- Rosenbaum, D. M., Rasmussen, S. G., and Kobilka, B. K. (2009). The structure and function of G-protein-coupled receptors. *Nature*, 459(7245):356–363.
- Samel, S. A., Czodrowski, P., and Essen, L. O. (2014). Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallogr. D Biol. Crystallogr.*, 70(Pt 5):1442–1452.
- Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A., and Essen, L. O. (2007). Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure*, 15(7):781–792.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, 74(12):5463–5467.
- Sauliere, A., Bellot, M., Paris, H., Denis, C., Finana, F., Hansen, J. T., Altie, M. F., Seguelas, M. H., Pathak, A., Hansen, J. L., Senard, J. M., and Gales, C. (2012). Deciphering biased-agonism complexity reveals a new active AT1 receptor entity. *Nat. Chem. Biol.*, 8(7):622–630.
- Schauwecker, F., Pfennig, F., Grammel, N., and Keller, U. (2000). Construction and in vitro analysis of a new bi-modular polypeptide synthetase for synthesis of N-methylated acyl peptides. *Chem. Biol.*, 7(4):287–297.
- Schmidt, E. W., Donia, M. S., McIntosh, J. A., Fricke, W. F., and Ravel, J. (2012). Origin and variation of tunicate secondary metabolites. *J. Nat. Prod.*, 75(2):295–304.
- Schomer, R. A. and Thomas, M. G. (2017). Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase. *Biochemistry*, 56(40):5380–5390.
- Schrage, R., Schmitz, A. L., Gaffal, E., Annala, S., Kehraus, S., Wenzel, D., Bullesbach, K. M., Bald, T., Inoue, A., Y., S., Galandrin, S., Shridhar, N., Hesse, M., Grundmann, M., Merten, N., Charpentier, T. H., Martz,

- M., Butcher, A. J., Slodczyk, T., Armando, S., Effer, M., Namkung, Y., Jenkins, L., Horn, V., Stossel, A., Dargatz, H., Tietze, D., Imhof, D., Gales, C., Drewke, C., E., M. C., Holzel, M., Milligan, G., Tobin, A. B., Gomez, J., Dohlman, H. G., Sondek, J., Harden, T. K., Bouvier, M., Laporte, S. A., Aoki, J., Fleischmann, B. K., Mohr, K., König, G. M., Tuting, T., and Kostenis, E. (2015). The experimental power of FR900359 to study Gq-regulated biological processes. *Nat Commun*, 6:10156.
- Schägger, H. (2006). Tricine-SDS-PAGE. *Nat Protoc*, 1(1):16–22.
- Sieber, S., Carlier, A., Neuburger, M., Grabenweger, G., Eberl, L., and Gademann, K. (2015). Isolation and Total Synthesis of Kirkamide, an Aminocyclitol from an Obligate Leaf Nodule Symbiont. *Angew. Chem. Int. Ed. Engl.*, 54(27):7968–7970.
- Silakowski, B., Schairer, H. U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S., and Müller, R. (1999). New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.*, 274(52):37391–37399.
- Smrcka, A. V. (2013). Molecular targeting of $G\alpha$ and $G\beta\gamma$ subunits: a potential approach for cancer therapeutics. *Trends Pharmacol. Sci.*, 34(5):290–298.
- Soby, S. D., Gadagkar, S. R., Contreras, C., and Caruso, F. L. (2013). Chromobacterium vaccinii sp. nov., isolated from native and cultivated cranberry (*Vaccinium macrocarpon* Ait.) bogs and irrigation ponds. *Int. J. Syst. Evol. Microbiol.*, 63(Pt 5):1840–1846.
- Stachelhaus, T., Mootz, H. D., Bergendahl, V., and Marahiel, M. A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.*, 273(35):22773–22781.
- Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.*, 6(8):493–505.

- Stegmann, E., Rausch, C., Stockert, S., Burkert, D., and Wohlleben, W. (2006). The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production. *FEMS Microbiol. Lett.*, 262(1):85–92.
- Süßmuth, R. D. und Mainz, A. (2017). Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew. Chem. Int. Ed. Engl.*, 56(14):3770–3821.
- Takefuji, M., Wirth, A., Lukasova, M., Takefuji, S., Boettger, T., Braun, T., Althoff, T., Offermanns, S., and Wettschureck, N. (2012). G(13)-mediated signaling pathway is required for pressure overload-induced cardiac remodeling and heart failure. *Circulation*, 126(16):1972–1982.
- Takesako, K., Ikai, K., Haruna, F., Endo, M., Shimanaka, K., Sono, E., Nakamura, T., Kato, I., and Yamaguchi, H. (1991). Aureobasidins, new antifungal antibiotics. Taxonomy, fermentation, isolation, and properties. *J. Antibiot.*, 44(9):919–924.
- Takesako, K., Mizutani, S., Sakakibara, H., Endo, M., Yoshikawa, Y., Masuda, T., Sono-Koyama, E., and Kato, I. (1996). Precursor directed biosynthesis of aureobasidins. *J. Antibiot.*, 49(7):676–681.
- Tan, L. and Ma, D. (2008). Total synthesis of salinamide A: a potent anti-inflammatory bicyclic depsipeptide. *Angew. Chem. Int. Ed. Engl.*, 47(19):3614–3617.
- Tan, X. F., Dai, Y. N., Zhou, K., Jiang, Y. L., Ren, Y. M., Chen, Y., and Zhou, C. Z. (2015). Structure of the adenylation-peptidyl carrier protein didomain of the *Microcystis aeruginosa* microcystin synthetase McyG. *Acta Crystallogr. D Biol. Crystallogr.*, 71(Pt 4):873–881.
- Taniguchi, M., Nagai, K., Arao, N., Kawasaki, T., Saito, T., Moritani, Y., Takasaki, J., Hayashi, K., Fujita, S., Suzuki, K., and Tsukamoto, S. (2003). YM-254890, a novel platelet aggregation inhibitor produced by *Chromobacterium* sp. QS3666. *J. Antibiot.*, 56(4):358–363.

- Tarry, M. J., Haque, A. S., Bui, K. H., and Schmeing, T. M. (2017). X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure*, 25(5):783–793.
- Thiericke, R. and Rohr, J. (1993). Biological variation of microbial metabolites by precursor-directed biosynthesis. *Nat Prod Rep*, 10(3):265–289.
- Tillett, D., Dittmann, E., Erhard, M., von Dohren, H., Borner, T., and Neilan, B. A. (2000). Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem. Biol.*, 7(10):753–764.
- Traber, R., Hofmann, H., and Kobel, H. (1989). Cyclosporins—new analogues by precursor directed biosynthesis. *J. Antibiot.*, 42(4):591–597.
- van der Donk, W. A. and Nair, S. K. (2014). Structure and mechanism of lanthipeptide biosynthetic enzymes. *Curr. Opin. Struct. Biol.*, 29:58–66.
- Velkov, T., Horne, J., Scanlon, M. J., Capuano, B., Yuriev, E., and Lawen, A. (2011). Characterization of the N-methyltransferase activities of the multifunctional polypeptide cyclosporin synthetase. *Chem. Biol.*, 18(4):464–475.
- Velkov, T. and Lawen, A. (2003). Mapping and molecular modeling of S-adenosyl-L-methionine binding sites in N-methyltransferase domains of the multifunctional polypeptide cyclosporin synthetase. *J. Biol. Chem.*, 278(2):1137–1148.
- Villiers, B. and Hollfelder, F. (2011). Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. *Chem. Biol.*, 18(10):1290–1299.
- Voing, K., Harrison, A., and Soby, S. D. (2015). Draft Genome Sequence of *Chromobacterium vaccinii*, a Potential Biocontrol Agent against Mosquito (*Aedes aegypti*) Larvae. *Genome Announc*, 3(3).
- Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure*, 8(4):407–418.

- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., and Medema, M. H. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.*, 43(W1):W237–243.
- Weckwerth, W., Miyamoto, K., Iinuma, K., Krause, M., Glinski, M., Storm, T., Bonse, G., Kleinkauf, H., and Zocher, R. (2000). Biosynthesis of PF1022A and related cyclooctadepsipeptides. *J. Biol. Chem.*, 275(23):17909–17915.
- Weist, S., Bister, B., Puk, O., Bischoff, D., Pelzer, S., Nicholson, G. J., Wohlleben, W., Jung, G., and Sussmuth, R. D. (2002). Fluorobalhimycin—a new chapter in glycopeptide antibiotic research. *Angew. Chem. Int. Ed. Engl.*, 41(18):3383–3385.
- Wink, J. M., Kroppenstedt, R. M., Ganguli, B. N., Nadkarni, S. R., Schumann, P., Seibert, G., and Stackebrandt, E. (2003). Three new antibiotic producing species of the genus *Amycolatopsis*, *Amycolatopsis balhimycina* sp. nov., *A. tolypomycina* sp. nov., *A. vancoresmycina* sp. nov., and description of *Amycolatopsis keratiniphila* subsp. *keratiniphila* subsp. nov. and *A. keratiniphila* subsp. *nogabecina* subsp. nov. *Syst. Appl. Microbiol.*, 26(1):38–46.
- Winn, M., Fyans, J. K., Zhuo, Y., and Micklefield, J. (2016). Recent advances in engineering nonribosomal peptide assembly lines. *Nat Prod Rep*, 33(2):317–347.
- Wolpert, M., Gust, B., Kammerer, B., and Heide, L. (2007). Effects of deletions of mbtH-like genes on clorobiocin biosynthesis in *Streptomyces coelicolor*. *Microbiology (Reading, Engl.)*, 153(Pt 5):1413–1423.
- Wright, S. K., Kish, M. M., and Viola, R. E. (2000). From malate dehydrogenase to phenyllactate dehydrogenase. Incorporation of unnatural amino acids to generate an improved enzyme-catalyzed activity. *J. Biol. Chem.*, 275(41):31689–31694.

- Yonemoto, I. T., Li, W., Khullar, A., Reixach, N., and Gerratana, B. (2012). Mutasythesis of a potent anticancer sibiromycin analogue. *ACS Chem. Biol.*, 7(6):973–977.
- Zhang, W., Heemstra, J. R., Walsh, C. T., and Imker, H. J. (2010). Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry*, 49(46):9946–9947.
- Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., and Jensen, P. R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE*, 7(3):e34064.

A Anhang

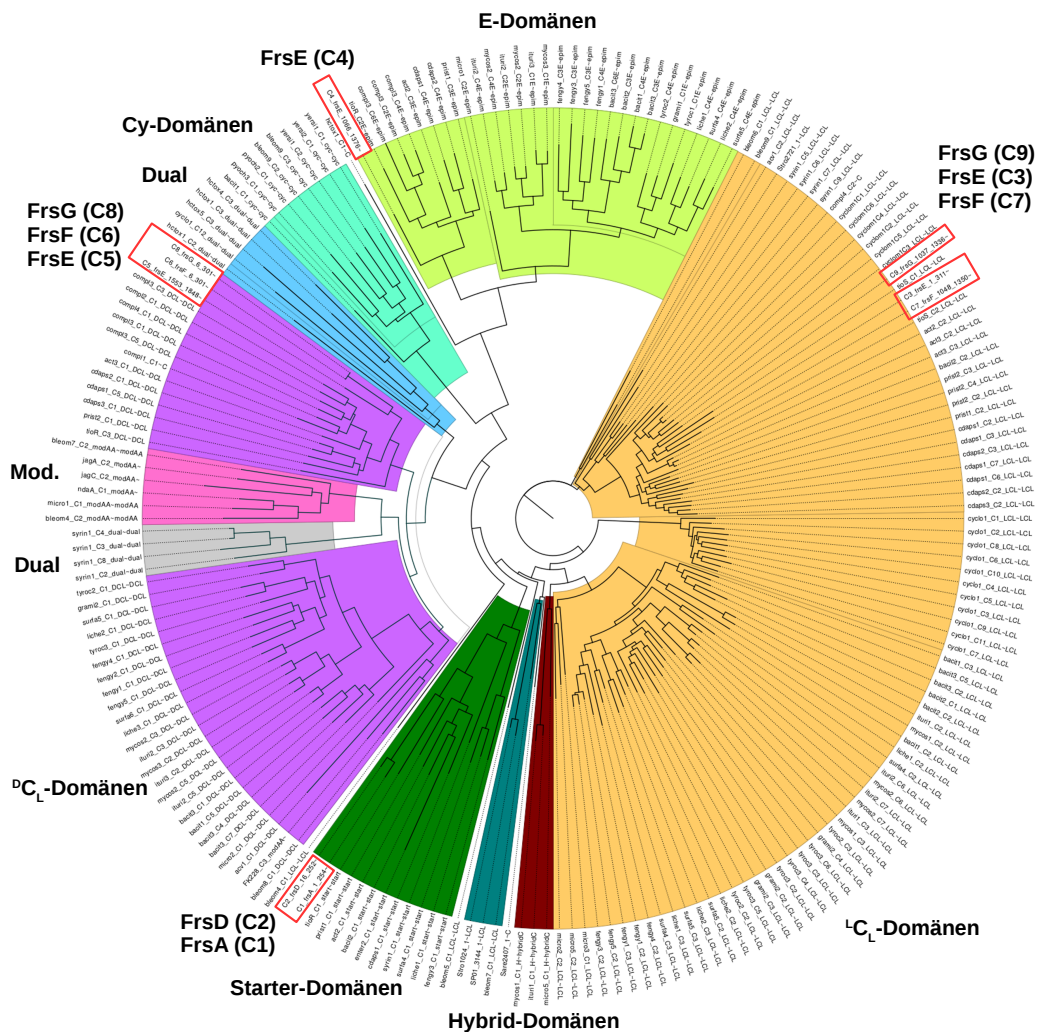


Abbildung A1: Phylogenetische Analyse der C-Domänen mit FastTree. Die Abbildung ist entnommen aus Crüsemann et al. 2018. Die Farben markieren die einzelnen C-Domänen Klassen. Die roten Umrandungen zeigen die C-Domänen des *bcfrs*-BGCs am Beispiel von *Cand. B. crenata*

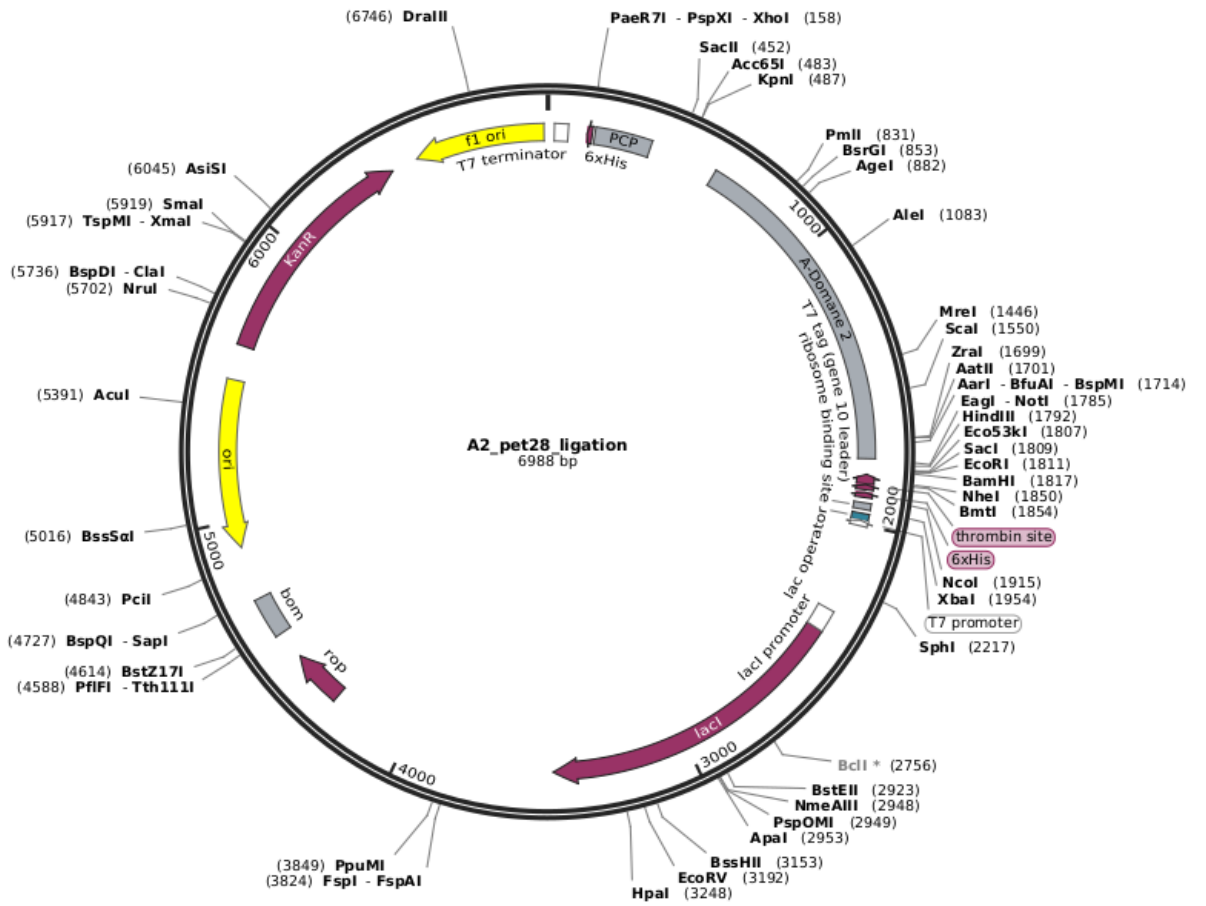


Abbildung A3: Vektorkarte A2-PCP in pET28a.

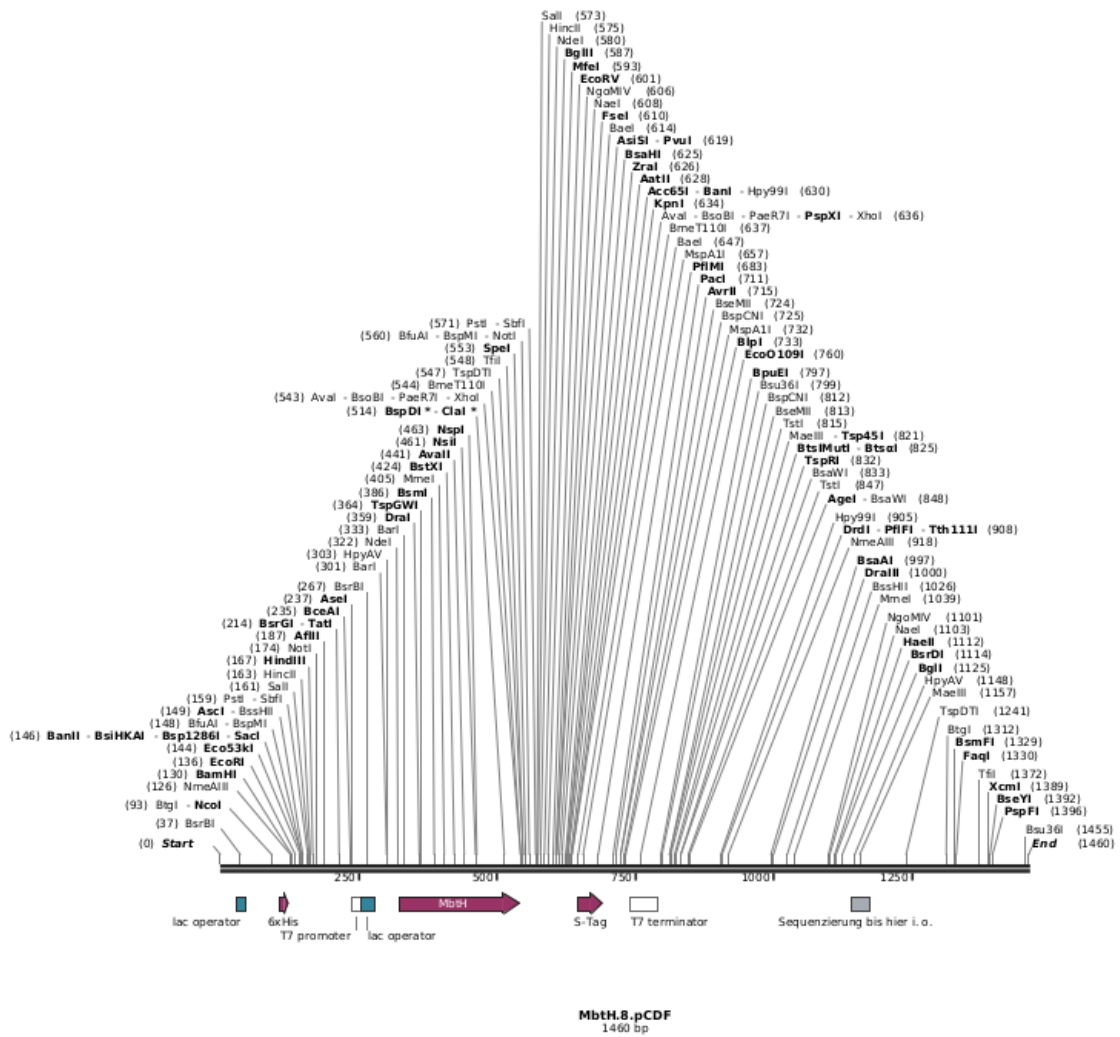


Abbildung A4: MLP (*frsB*) aus dem *bcfrs*-BGC in pCDFApra.

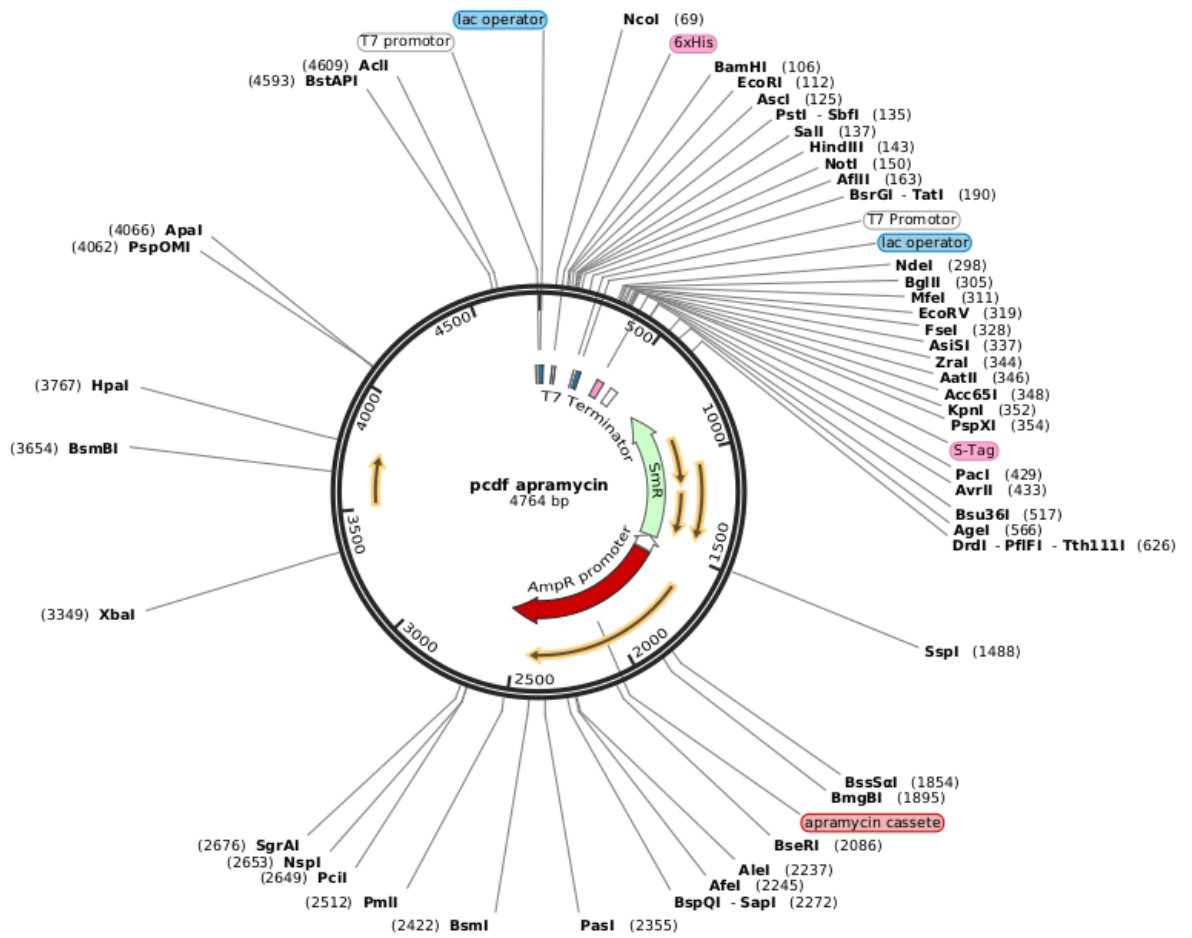


Abbildung A5: Vektorkarte des pCDFApra. Der Vektor wurde von Luis Linares Otoya aus dem pCDFDuet-Vektor erstellt.

Alignment der A-Domänen aus dem *bcfrs*-BGC

CLUSTAL 0(1.2.4) multiple sequence alignment

```
GrsA          -----MLNSSKSILIIHAQNKNGTHEEEQYLFVNNTKAEYPRDKTIHQLFEEQVSKR
A8_FrsG_2_Burk -----TVDDTIIGRFAEQVRRRA
A4_FrsE_2_Burk -----ANGIAQWNDTQHACPSADLAQLFERQVRLT
A5_FrsF_1_Burk -----LPEWFEAQVERT
A6_FrsF_2_Burk RFRTLLEAVVTDSS-CPIADIDLLTA-DE RQQLLEVNATDRPLPKVTLPEWFEAQVERT
A3_FrsE_1_Burk -----QAALT
A2_FrsD_Burk  -----AVPQ-SPIDTLPWLDE-SERRQLLEEWSGNALDLGEITLAE LFEVQATRQ
A1_FrsA_Burk  -----AVPQ-SPIDTLPWLDE-SERRQLLEEWSGNALDLGEITLAE LFEVQATRQ
A7_FrsG_1_Burk -----LAELFEAQATRQ
* .
```

```
GrsA          PNNVAIVCENEQLTYHELNVKANQLARIFIEKGIGKDTLVGIMMEKSIDLFIGILAVLKA
A8_FrsG_2_Burk PDATVLT YRGETLTCAELNARANRLARCL IADGAGPEELIAVALPRSIDMVVSL LAILKT
A4_FrsE_2_Burk PDASALTFGSQTL SYAVLNACANRLARWLLMHSIGPDDVVAVALPRSIDLV IALLAVVKS
A5_FrsF_1_Burk PTAMAVLCDEVALDYRELNCRANQIAHWLISQGVGPEQKVALALPRRVELLVAMLGV LKA
A6_FrsF_2_Burk PTAMAVLCDEVALDYRELNCRANQIAHWLISQGVGPEQKVALALPRRVELLVAMLGV LKA
A3_FrsE_1_Burk PEAVALAFGDEVLSY AELNKQANRLARMLVAAGLGPEGRVALAVPRSLDMVVAL LGVTKA
A2_FrsD_Burk  PNAVALEGPDERVSYGELDARANRLASHLQSLGVGPDVVVVVCLERSIDMVVAILGI AKA
A1_FrsA_Burk  PNAVALEGPDERVSYGELDARANRLASHLQSLGVGPDVVVGCLERSIDMVVAILGI AKA
A7_FrsG_1_Burk PNAVALEGPDGCVS YGELDARANRLASHLQSLGVGPDVVVGCLERSIDMVVAILGI AKA
* . : : * : * : * : . * : : : : : : : : : : : : : : : : * :
```

```
GrsA          GGAYVPIDIEYPKERIQYI LDDSQARMLLTQKHLVLIHNIQFNGQV---EIFEEDTIK
A8_FrsG_2_Burk GAAYLPM DPAYPLERLSFMLTD AKPRLLLANMETIRQVAEVS RVSTLALDCPELAQTLTR
A4_FrsE_2_Burk GAAYLPLDADYPRNRLDFMLTDARPRALLNGSMVEALSPAAGTQVLLDAP EWTAARNH
A5_FrsF_1_Burk GAAYLPLDPAYPAEHLMYMLVDAKPA CLLGLGDSIAALPDSNV-----RAWELDDEA
A6_FrsF_2_Burk GAAYLPLDPAYPAEHLMYMLVDAKPA CLLGLGDSIAALPDSNV-----RAWELDDEA
A3_FrsE_1_Burk GAAYLPLDP EYPAERLAYMLADAKPTLLMTVNAQLGSLSECA GIPVLALDADSVRAISQ
A2_FrsD_Burk  GAAYLPLFADPYPTERLAYMLTDSMAPVLLTESKQVERLPSYWGHLVKL-DR-----
A1_FrsA_Burk  GAAYLPLADPYPTERLAYMLTDSMAPVLLTESKQVERLPSYWGHLVKL-DR-----
A7_FrsG_1_Burk GAAYLPLADPYPTERLAYMLTDSMAPVLLTESKQVERLPSYWGHLVEL-DR-----
* . * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
```

```
GrsA          -----IREGTLNHVPSKSTDLAYVIYTSGTTGNPKGTMLEHKGISNLKVF FENSLN
A8_FrsG_2_Burk -----TDDSNLVP RPLRSDNAAYLIYTSGSTGAPKGVLI PHSNVLRLLDKTAHWFD
A4_FrsE_2_Burk LD-----DRDMV VTERKQPLRPLDAAYVIYTSGSTGLPKGVV NTHHGIVNRLTWMQSAYR
A5_FrsF_1_Burk VKQALVAQPQENPTPQRRLR AEHPAYVIYTSGSTGKPKGVEISQRSATDFVSWAHEAFG
A6_FrsF_2_Burk VKQALVAQPQENPTPQRRLR AEHPAYVIYTSGSTGKPKGVEISQRSATDFVSWAHEAFG
A3_FrsE_1_Burk MS-----GCNLVQSERLCP LQPQHPCVVIYTSGSTGRPKGVMVTHQGI VSLRASQIERFG
A2_FrsD_Burk  LD-----LSGQASSAPARALRPDHLAYVIYTSGSTGQPKGVAVSHAGLAGLVKSQEERFA
A1_FrsA_Burk  LD-----LSGQASSAPARALRPDHLAYVIYTSGSTGQPKGVAVSHAGLAGLVKSQEERFA
A7_FrsG_1_Burk LD-----LSWQASSAPARALRPDHLAYVIYTSGSTGQPKGVAVSHAGLAGLVKSQEERFA
: . : : * : * : * : * : * : * : * : * : * : * : * : * : * :
```

```
GrsA          VTEKDRIGQFASISFDASVWEMFMALLTGASLYIILKDTI-ND FVKFEQYINQKEITVIT
A8_FrsG_2_Burk FGPKDVWTLFHSYAFDFSVWEI WGALLTSGRLVVVPQTVS-LAPDEF LTLLEYEKVTILN
A4_FrsE_2_Burk LDASDVVLQKTPFSFDVSVW EFWPLLN GARLVMVAVPDGH-RDPAYLAELIQRQGVTTLH
A5_FrsF_1_Burk PDFTFGDVLATTSLSF DVSVF ELLAPLLCGGRVNLLRD-----LLVLGERSIERG-SLIS
A6_FrsF_2_Burk PDFTFGDVLATTSLSF DVSVF ELLAPLLCGGRVNLLRD-----LLVLGERSIERG-SLIS
A3_FrsE_1_Burk VSAESSVLQFASLSF GAALFEICTSLLTGARLVLVSSI KEALNVETMTALVTRHRLSHMV
A2_FrsD_Burk  VAGPVRVLQFASLSF DAAVM EILMAFCSSGRLVLPAA--GPLLGEQLEETLNRYAISHAL
A1_FrsA_Burk  VAGPVRVLQFASLSF DAAVM EILMAFCSSGRLVLPAA--GPLLGEQLEETLNRYAISHAL
A7_FrsG_1_Burk VAGPVRVLQFASLSF DAAVM EILMAFCSSGRLVLPAA--GPLLGEQLEKTLNRYAISHAL
: * . : : * : : . . : : : : : : : : : : : : : : : :
```

```

GrsA      LPPTYVVHLDPE-----RILSIQTLITAGSATSPSLVNK---WKE----KVTYIN
A8_FrsG_2_Burk QTPSAFYALMQAERHHGSNGGASLSLRRIIFGGEALDLSAL---QPWYQRHGDETTQLVN
A4_FrsE_2_Burk  FVPSMLDAFLNEPS-----SRQCLSLKRVLCSEVLSGNLA-ALQQHVL-----KRPLHN
A5_FrsF_1_Burk  AVPSVFAQLLQHGD-----LR--LDASTVVFAGEALPPELVEAVRRRWPF-----GCRAAN
A6_FrsF_2_Burk  AVPSVFAQLLQHGD-----LR--LDASTVVFAGEALPPELVEAVRRRWPF-----GCRVAN
A3_FrsE_1_Burk  VPPSALDTLCAAD-----RLPRTVRIMVAGEHCPAHLVER---WSA-----DRFMVN
A2_FrsD_Burk    IAPSALETVEAE-----VVPGLSTLVVGGGETCSGATAAS---WSQ-----GRRMVN
A1_FrsA_Burk    IAPSALETVEAE-----VVPGLSTLVVGGGETCSGATAAS---WSQ-----GRRMVN
A7_FrsG_1_Burk  IAPSALETVEAE-----VVPGLSTLVVGGGETCSGATAAS---WSQ-----GRRMVN
          *: . . .                :: :*.

GrsA      AYGPTETTICATTWATKE--TIGHSVPIGAPIQNTQIYIVDENLQLKSVGEAGELCTIGG
A8_FrsG_2_Burk MYGITETTIVHSYQPLDSGMCGRSNCNSLIGIGIPDLHLFLDDACLQVPVGGIGELYVGG
A4_FrsE_2_Burk  LYGPTEAADVTAHACDPN--DTGSSIQIGKPIWNTRIHVLDGLRPPVPLGVAGELYTIGG
A5_FrsF_1_Burk  IYGPTETTVYAIPTWLKEV--EGSRAPMIGRPLGNTPAYVLDGLRPPVPGVVGELYTIGG
A6_FrsF_2_Burk  IYGPTETTVYAIPTWLKEA--EGSRAPMIGRPLDNTRAYVLDGLRPPVPGVVGAGELYTIGG
A3_FrsE_1_Burk  GYGSSEVTVCA--TMSKPL--SGRALPPMGAPNANTRLYLLDAGMQPVPAGVMGELYVAG
A2_FrsD_Burk    AYGPTETVTCV--TMSKPL--SGSDKPKLGRPTLGAKLYVLDSTLQVPVPGVAGELYTAG
A1_FrsA_Burk    AYGPTETVTCV--TMSKPL--SGSDKPKLGRPTLGAKLYVLDSTLQVPVPGVAGELYTAG
A7_FrsG_1_Burk  AYGPTETVTCV--TMSKPL--SGSDKPKLGRPTLGAKLYVLDSTLQVPVPGVAGELYTAG
          ** :* : : .                :* . . : : * : * : * : *

GrsA      EGLARGYWKRPELTSQKFVDNPFVPGKLYKTGDQARWLSGDNIEYLRIDNQVKIRGHR
A8_FrsG_2_Burk AGLARGYHNRAASLTAERFVANPFASSKRMRYRTGDLARRNVNDIFEYHGRADQVVKVGRFR
A4_FrsE_2_Burk  TGLARGYLNRAAGLTAERFVANPYGEGERLYRSGDLARWNAEGELEYLGRLDQQLKIRGFR
A5_FrsF_1_Burk  AGLARGYLNRAAGLTAERFVANPYGEGERLYRSGDLARWNAEGELEYLGRVDQQLKIRGFR
A6_FrsF_2_Burk  TGLARGYLNRAAGLTAERFVANPYGEGERLYRSGDLARWNAEGELEYLGRVDQQLKIRGFR
A3_FrsE_1_Burk  EGLARGYLSRPGGLTAERFVANPFEPGWQMYRTGDLARRDIDGRDLYLGRVDHVKIRGFR
A2_FrsD_Burk    RGLARGYYQRPGLTAERFVANPYGKGERLYRSGDLARWSGEGELEYLGRVDQQLKIRGFR
A1_FrsA_Burk    RGLARGYYQRPGLTAERFVANPYGKGERLYRSGDLARWSGEGELEYLGRVDQQLKIRGFR
A7_FrsG_1_Burk  RGLARGYYQRPGLTAERFVANPYGKGERLYRSGDLARWSGEGELEYLGRVDQQLKIRGFR
          ***** . * ** : : * * : . : : * : * * : . . : : * * * : * : * : *

GrsA      VELEEVESILLKHMYSIETAVSVHKDHQEQPYLCAYFVSEKHI--PLEQLRQFSSEELPT
A8_FrsG_2_Burk  IELGEIETVLRQHPGVEDARVVVQTLHDNDCRLVAYLQPSERTAAPL-----RRWLKV
A4_FrsE_2_Burk  IEPGEIEAALCRHPLVSQAVVIAREDTPGHKQLVGYVVLDDAAAL-QRDAEDET-RQVQA
A5_FrsF_1_Burk  IEPGEIEAALCRHPLVSQAVVIAREDTPGHKQLVGYVVLDDAAAD-GQMLRRFLSRHLPE
A6_FrsF_2_Burk  IEPGEIEAALCRHPLVSQAVVIAREDTPGHKQLVGYVVLDDAAAL-QRDAEDET-RQVQA
A3_FrsE_1_Burk  IEPGEIEAALRQLPGVAQATVVAWEVPGAKQLVGYVVPKEGMLEPRAMRRELAEHLTD
A2_FrsD_Burk    IEPGEIETVLCQHPQLREAVVVSRT-NGRDTQLVGYVTVRGE--VDGQALRRQVANWLPE
A1_FrsA_Burk    IEPGEIETVLCQHPQVREAVVVSRT-NGRDTQLVGYVTIRGE--VDGQALRRQVANWLPE
A7_FrsG_1_Burk  IEPGEIETVLCQHPQVREAVVVSRT-NGRDTQLVGYVTVRGE--VDGQALRRQVANWLPE
          : * * : * : : : : *                * . .                . :

```

grün: Core-Regionen nach Stachelhaus
gelb: Stachelhauscode

Abbildung A7: Alignment der A-Domänen aus dem *befrs*-BGC.

Percent Identity Matrix des A-Domänenalignments aus dem *bcfrs*-BGC

 #
 # Percent Identity Matrix - created by Clustal2.1
 #
 #

177

1: GrsA	100.00	31.07	29.65	35.98	27.57	34.84	35.35	35.17	36.40
2: A8_FrsG_2_Burk	31.07	100.00	38.24	36.44	35.30	37.87	37.10	37.10	37.27
3: A4_FrsE_2_Burk	29.65	38.24	100.00	47.43	73.23	40.55	41.10	41.28	42.30
4: A5_FrsF_1_Burk	35.98	36.44	47.43	100.00	84.00	46.71	50.68	51.08	51.08
5: A6_FrsF_2_Burk	27.57	35.30	73.23	84.00	100.00	39.44	42.43	42.43	43.26
6: A3_FrsE_1_Burk	34.84	37.87	40.55	46.71	39.44	100.00	50.09	50.28	50.56
7: A2_FrsD_Burk	35.35	37.10	41.10	50.68	42.43	50.09	100.00	96.92	96.13
8: A1_FrsA_Burk	35.17	37.10	41.28	51.08	42.43	50.28	96.92	100.00	97.79
9: A7_FrsG_1_Burk	36.40	37.27	42.30	51.08	43.26	50.56	96.13	97.79	100.00

Abbildung A8: Percent identity Matrix des Alignments der A-Domänen aus dem *bcfrs*-BGC.

Alignment der A-Domänen aus dem *cvfrs*-BGC

CLUSTAL O(1.2.4) multiple sequence alignment

```

GrsA          -----ML-----NSSKSILIIHAQNKNGTHEEEQYLFVAVNNTKAEYPRDKTIHQ
FrsG2_Cro     -----SIVQR
FrsE1_Cro     -----VSGWNDSARPLE-AGTLACL
FrsG1_Cro     -----GVIAAEPSPVSDIELLDEAERRQLLDVWNRTGPDHG-QATFPQL
FrsA_Cro      -----AAEPSQPVSDIELLDEAERRQLLDVWNRTGPDHG-QATFPQL
FrsD_Cro      -----AAEPSQPVSDIELLDEAERRQLLDVWNRTGPDHG-QATFPQL
FrsE2_Cro     -----WLRSWNDTGHAYP-DGDLAQL
FrsF1_Cro     KGVVQGLLTRFRLLLEAVAAEPSRPVSDIELLDEAERRQVLIENWATERPSP-QATLPEL
FrsF2_Cro     --TVERLLTRFRLLLEAVAAEPSRPMVDIELLDEAERRQVLIENWATERPSP-QATLPEL
:

```

```

GrsA          FEEQVSKRPNNVAIVCENEQLTYHELNVKANQLARIFIEKGIGKDTLVGIMMEKSIDLFI
FrsG2_Cro     FEEQAGRAPQASALSFQGTLSYAEINARANRLAHLIARGAGPEDLIAALPRSLDLVV
FrsE1_Cro     FEAQAKASPDVALAFGSEQYSYAEIDRRANQLARTLAGAGIGPEDIVALAVPRSLDMVV
FrsG1_Cro     FETQAALTPHAVALSPDARLSYAEIDARANRLARHLQSLGVGADVLVIGICLERSIDMVV
FrsA_Cro      FETQAALTPHAVALSPDARLSYAEIDARANRLARHLQSLGVGADVLVIGICLERSIDMVV
FrsD_Cro      FETQAALTPHAVALSPDARLSYAEIDARANRLARHLQSLGVGADVLVIGICLERSIDMVV
FrsE2_Cro     FERQARLTPNAQALIFERQSLSYAALNARANQLCRVLLAHGVGPDVVAVALPRSELV
FrsF1_Cro     FEAQAARAPNAAALFCDGETLDYAEINRRANQLAHLIGQGVGPEQRVALALPRRAELLV
FrsF2_Cro     FEAQAARAPNAAALFCDGETLDYAEINRRANQLAHLIGQGVGPEQRVALALPRRAELLV
** * . * . * : * * : : * * : : * * : : : : : : : : : : : : : :

```

```

GrsA          GILAVLKAGGAYVPIDIEYPKERIQYILDDSQARMMLTQKHLVHLIHNIFQNGQVEIFEE
FrsG2_Cro     SLLAILKTGAAYLPLDPDYPAERLGFMLADAKPRLLLGHGEALRHLVPVDASTQ-ATALDD
FrsE1_Cro     ALLGVVKGAAAYLPLDPEYPRERLAHMLSDATPRLLLATSDTVGGLPFAFSGLR-VQVLD
FrsG1_Cro     AVLGALKSGAAYLPLSPEYPTERLAYMLGDSMAPVLLTDSAQVERLPS-YWGR-VVELDR
FrsA_Cro      AVLGALKSGAAYLPLSPEYPTERLAYMLGDSMAPVLLTDSAQVERLPS-YWGR-VVELDR
FrsD_Cro      AVLGVLKSGAAYLPLSPEYPTERLAYMLGDSMAPVLLTDSAQVERLPS-YWGR-VVELDR
FrsE2_Cro     ALLAVVKSGAAYLPLDADYPRERLDFMLADARPAVLLSNAAMAGILSPADGTR-LLSLDE
FrsF1_Cro     AVLGVLKSGAAYLPLDPAYPAERLAHMLADARPAACLLGLGDGIEA-LPDSGVA-CWRLLD
FrsF2_Cro     AVLGVLKSGAAYLPLDPAYPAERLAHMLADARPAACLLGLGDGIEA-LPDSGVA-CWRLLD
.:* . :*: .*: * . * * * : : * * : * * : * * : : : : : : : : : :

```

```

GrsA          DTI-----KIREGTLNHPVSKSTDLAYVIYTSGTTGNPKGTMLEHGKISNLKVFEE
FrsG2_Cro     AELGRELALAG--DGNPERARPLGADHAAYVIYTSGSTGQPKGVLPVPHRNVRLLDSTE
FrsE1_Cro     PAWREMPVARADGRPLAQRERTRPLLPQHVCVVIYTSGSTGKPKGVSVTHQGIASLRASQI
FrsG1_Cro     L-----DLDALPDSAPERALARAEHLAYVIYTSGSTGQPKGVAVSHAGLAGLAGSQT
FrsA_Cro      L-----DLDALPDSAPERALARAEHLAYVIYTSGSTGQPKGVAVSHAGLAGLAGSQT
FrsD_Cro      L-----DLDALPDSAPERALARAEHLAYVIYTSGSTGQPKGVAVSHAGLAGLAGSQT
FrsE2_Cro     PGLLSAQDGADSGDLAAGERRRRLRPQDAAYVIYTSGSTGKPKGVINTHQGIVNRLAWMQ
FrsF1_Cro     AALRQTLAAQ--PQSDPTPAQRRLRPEHPAYVIYTSGSSGQPKGVVISQRSAADFVDWAL
FrsF2_Cro     AALRQTLAAQ--PQSDPTPAQRRLRPEHPAYVIYTSGSSGQPKGVVISQRSAADFVDWAL
. . . * * * * * : : * * * : : .

```

```

GrsA          NSLNVTEKDRIGQFASISFDASVWEMFMALLTGASLYIILKDTI--NDFVKFEQYINQKE
FrsG2_Cro     RWFVFGAGDVWTLFHSYAFDFSVWEIWGALLSGGRLVVVPRAAV--QAPDEFLLALEREK
FrsE1_Cro     ERFVGSAGSSVLQFASLSFGAALFEVCMSLLVGARLVLVASAREAL-NAGAMAELARRHG
FrsG1_Cro     ERFALQGPTRVLQFASLSFDAAVMEMLMAFCSSGGRLVLPAAAGPLLGEQLLDT---LNRHE
FrsA_Cro      ERFALQGPTRVLQFASLSFDAAVMEMLMAFCSSGGRLVLPAAAGPLLGEQLLDT---LNRHE
FrsD_Cro      ERFALQGPTRVLQFASLSFDAAVMEMLMAFCSSGGRLVLPAAAGPLLGEQLLDT---LNRHE
FrsE2_Cro     SAYRLDASDTVLRKTPFSFDVSVWEFFWPLLEGARLVLAVPDGH--RDPAYLAALIQRRER
FrsF1_Cro     ASFGAETFADVLATTSLSDVSVFELLTPLSSGGRVRLRLDLELGERPISGG-----
FrsF2_Cro     ASFGAETFADVLATTSLSDVSVFELLTPLSSGGRVRLRLDLELGERPISGG-----
: * . : : * . : * . : :

```


Percent Identity Matrix des A-Domänenalignments aus dem *cvfrs*-BGC

```

#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: GrsA          100.00  31.71  35.62  36.35  36.65  37.02  31.16  34.35  31.27
2: FrsG2_Cro    31.71  100.00  43.69  43.75  43.48  43.48  44.18  46.46  46.64
3: FrsE1_Cro    35.62  43.69  100.00  52.84  52.84  53.03  46.73  51.37  46.72
4: FrsG1_Cro    36.35  43.75  52.84  100.00  100.00  99.82  45.56  54.73  50.29
5: FrsA_Cro     36.65  43.48  52.84  100.00  100.00  99.63  45.30  55.05  50.58
6: FrsD_Cro     37.02  43.48  53.03  99.82  99.63  100.00  45.49  55.24  50.78
7: FrsE2_Cro    31.16  44.18  46.73  45.56  45.30  45.49  100.00  52.52  61.97
8: FrsF1_Cro    34.35  46.46  51.37  54.73  55.05  55.24  52.52  100.00  91.24
9: FrsF2_Cro    31.27  46.64  46.72  50.29  50.58  50.78  61.97  91.24  100.00

```

Abbildung A10: Percent Identity Matrix des Alignments der A-Domänen aus dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

Acre_FrsG_MT_1      GRADQQVKVGRFRIELGEIETVLRQHPGVEDARVVVQTLHDNDCRLVAYLQPSE-----
Acre_FrsE_MT        -----EIEAALCRHPLVSQAVVIAREDTPGHKQLVGYVVLDDAAALQR
Acre_FrsF_MT        -----EIEAALCRHPLVSQAVVIAREDTPGHKQLVGYVVLDDAAALQR
Acre_FrsG_MT_2      -----EIEAALCRHPLVSQAVVIAREDPGHKQLVGYVVLDDAAALQR
                    ***.:.* :** *.:* *.:. . . :*.** : .:

Acre_FrsG_MT_1      --RTAAPLRRWLKVRRAEA-QGARAFELPNGMPIFHHTMETEFLYEEIFEDLVYFKH
Acre_FrsE_MT        DAEDETRQVQAWRQVYDTLYEAHCQPPFGEN--FGGWSSYDGGQPLPLEEMRAWRQATVE
Acre_FrsF_MT        DAEDETRQVQAWRQVYDTLYEAHCQPPFGEN--FGGWSSYDGGQPLPLEEMRAWRQATVE
Acre_FrsG_MT_2      DTEDETRQVQAWRQVYDTLYEAHCQPPFGEN--FGGWSSYDGGQPLPLEEMRAWRQATVE
                    : : * : * : : * : : . . : * * : * :

Acre_FrsG_MT_1      GIRLDDGACVFDVGANIGLFLMFLVFGQHCRNATIFAFEPPIPVFRTLTLNAEVHGDKVRLF
Acre_FrsE_MT        RIRELQPRRVLEIGVSGLLLAPLAPECEAYWGTDLSPTTIAVLERQLSEQTFGDRVRLF
Acre_FrsF_MT        RIRELQPRRVLEIGVSGLLLAPLAPECEAYWGTDLSPTTIAVLERQLSEQTFGDRVRLF
Acre_FrsG_MT_2      RIRELQPRRVLEIGVSGLLLAPLAPECEAYWGTDLSPTTIAVLERQLSEQTFGDRVRLF
                    ** : * : : * . ** : : . * . : * . . * . . . * * * * *

Acre_FrsG_MT_1      DCGLSDAVRQ-----ERFTFYPNDTLISTSR-----NSPEA--IRGMVKS---
Acre_FrsE_MT        ALGAHELLKLPENGFVIVINSVLQYFPNAAYLTEVIKQSLERLSPGGALYLGDVNRNAL
Acre_FrsF_MT        ALGAHELLKLPENGFVIVINSVLQYFPNAAYLTEVIKQSLERLSPGGALYLGDVNRNAL
Acre_FrsG_MT_2      ALGAHELKLPENGFVIVINSVLQYFPNAAYLTEVIKQSLERLSPGGALYLGDVNRNAL
                    * : : . : : * * : : * * . * * :

Acre_FrsG_MT_1      ---FL---INQHGNHSGDDAEVGEILLDERLA-----SQDYMCVLRTLSDVITEH---
Acre_FrsE_MT        LDVFATAVDLCQVGTET-DAASLRRRVEQRLLAEKELLVSPAFFSRRLRETLQPIGAVDIR
Acre_FrsF_MT        LDVFATAVDLCQVGTET-DAASLRRRVEQRLLAEKELLVSPAFFSRRLRETLQPIGAVDIR
Acre_FrsG_MT_2      LDVFATAVDLCQVGTET-DAASLRRRVEQRLLAEKELLVSPAFFSRRLRETLQPIGAVDIR
                    * : * * . : * * . : : * * * * * * * * * *

Acre_FrsG_MT_1      -----NVDRIDLLKI---DVENAEYNVLQGVESDWPKIRQLVMEV-----HDVD
Acre_FrsE_MT        LKRGEAVNELNRYRYEAVLYKEPCEAVSLSKVETQVWSDVGNMSACRTRLMQYASMMWRIE
Acre_FrsF_MT        LKRGEAVNELNRYRYEAVLYKEPCEAVSLSKVETQVWSDVGNMSACRTRLMQYASMMWRIE
Acre_FrsG_MT_2      LKRGEAVNELSRYRYEAVLYKEPCEAVSLSEVETQVWSDVGNMSACRTRLMQYAPMMWRIE
                    * : . : : * * * * . : . * . : : :

Acre_FrsG_MT_1      -----GR-----L-----RRIVDLLHRGYRVIHEQDRLL
Acre_FrsE_MT        GIPNARLHGEVLVTRRLKSEEGLASWRRWLDESGGVEPEDLYRLGAEVLYRVLVTSNQP
Acre_FrsF_MT        GIPNARLHGEVLVTRRLKSEEGLASWRRWLDESGGVEPEDLYRLGAEVLYRVLVTSNQP
Acre_FrsG_MT_2      GIPNARLHGEVLVTRRLKSEEGLASWRRWLDESGGVEPEDLYRLGAEVLYRVLVTSNQP
                    * . * . : * . * * * : . .

Acre_FrsG_MT_1      RDTAIHCLYAVHVS---DAEMRPGHHAENTSLFWN--CRADLLRDVRANLRKRLPDYMQ
Acre_FrsE_MT        D--HFDAVFLPEQQRGALDAVYRTPPVLPKPLSGYVNNPANFEQFAAIRRYVGEQLPDYMV
Acre_FrsF_MT        D--HFDAVFLPEQQRGALDAVYRTPPVLPKPLSGYVNNPANFEQFAAIRRYVGEQLPDYMV
Acre_FrsG_MT_2      D--HFDAVFLPEQQRGALDAVYRTPPVLPKPLSGYVNNPANFEQFAAIRRYVGEQLPDYMV
                    : : : : . * * * * : * : * . . : : * : : * * * *

Acre_FrsG_MT_1      PNHMVLLDTFPLTPNGKLDLRRALPAPFEQAAMRVRDIEPA-----
Acre_FrsE_MT        PAALVLLLEGLPLTPNGKLDLRRALPAPFEFGTACYRAPGSKQEKVLAQLFAEVLGLPQVGV
Acre_FrsF_MT        PAALVLLLEGLPLTPNGKLDLRRALPAPFEFGTACYRAPGSKQEKVLAQLFAEVLGLPQVGV
Acre_FrsG_MT_2      PAALVLLLEGLPLTPNGKLDLRRALPAPFEFGTACYRAPGSEQEKVLAQLFAEVLGLPQVGV
                    * : * * * : * * * * * * * * * * : *

```

```
Acre_FrsG_MT_1      -----  
Acre_FrsE_MT        DSFFDLGG  
Acre_FrsF_MT        DSFFDLGG  
Acre_FrsG_MT_2      DSFFDLGG
```

gelb: konservierte Regionen von MT-Domänen nach Labby et al..
rosa: Core A9-Core A11 Bereiche der A-Domänen.

Abbildung A11: Alignment der MT-Domänen aus dem *bcfrs*-BGC.

CLUSTAL 0(1.2.4) multiple sequence alignment

```

FrsG_MT1_Cro      GRADQQVKVGRGFRIEPEIETALRGHPGVEDARVVVKAMDGNDQRLIAYLLPSERAAA--
FrsF_MT_Cro      GRADQQLKVRGFRIEPEIEAALCRHPSVAQA AVIAREDTPGHKQLVGYVVLDDAAALQR
FrsE_MT_Cro      GRADQQLKVRGFRIEPEIEAALCRHPSVAQA AVIAREDTPGHKQLVGYVVLDDAAALQR
FrsG_MT2_Cro      -----EIEAALCRHPSVAQA AVIAREDTPGHKQLVGYVVLDDAAALQR
                  ***.* **.* : * : : : . : : * : : : *

FrsG_MT1_Cro      -----PLRRWLRAKSSVAQNAREIELPNGLPVFHHNAEETFLYEEIFEDRIYLKHG
FrsF_MT_Cro      DEDSESQRQVEAWQQVYDTLYDAHQH-QPFGENFGGWDSSYDQQLPLAQMREWRSATVER
FrsE_MT_Cro      DEDSESQRQVEAWQQVYDTLYDAHQH-QPFGENFGGWDSSYDQQLPLAQMREWRSATVER
FrsG_MT2_Cro      DEDSESQRQVEAWQQVYDTLYDAHQH-QPFGENFGGWDSSYDQQLPLAQMREWRSATVER
                  : . * : . : : : : : : : : : : : : : : : * * .

FrsG_MT1_Cro      IRLDNDACVFDVGANIGLFTLVFGQHCNATVFAFEPIPPVFGTSLNAAVHGGKVRFLD
FrsF_MT_Cro      IRELPRRLLLEIGVGSGLLLAPLADACEAYWGTDLSPATIAVLEKQLETQSCRDKVRLFA
FrsE_MT_Cro      IRELPRRLLLEIGVGSGLLLAPLADACEAYWGTDLSPATIAVLEKQLETQSCRDKVRLFA
FrsG_MT2_Cro      IRELPRRLLLEIGVGSGLLLAPLADACEAYWGTDLSPATIAVLEKQLETQSCRDKVRLFA
                  ** . : : : * . ** : : : * : * . . * : : . *****

FrsG_MT1_Cro      CGLSDAARQE-----TFTFYPNDTLISSSRNSAEATR-----RMV
FrsF_MT_Cro      LGAHELARLPAMRFDCIVINSVLQYFPNAAAYLGEVIEQALARLEAGGALYLGVRNLELL
FrsE_MT_Cro      LGAHELARLPAMRFDCIVINSVLQYFPNAAAYLGEVIEQALARLEAGGALYLGVRNLELL
FrsG_MT2_Cro      LGAHELARLPAMRFDCIVINSVLQYFPNAAAYLGEVIEQALARLEAGGALYLGVRNLELL
                  * : ** . : : : * : : . . : * * . : :

FrsG_MT1_Cro      KSFLINQ-----HGSDGAEVDELLEERL-----SSQYVCGLRSLSDVVAEHGVERI
FrsF_MT_Cro      PSFAAAVELRQSEPEVDAALQRRVSRQLLAEKELLAPDFFSRLRE---QLPQIGAVDI
FrsE_MT_Cro      PSFAAAVELRQSEPEVDAALQRRVSRQLLAEKELLAPDFFSRLRE---QLPQIGAVDI
FrsG_MT2_Cro      PSFAAAVELRQSEPEVDAALQRRVSRQLLAEKELLAPDFFSRLRE---QLPQIGAVDI
                  ** . : * . * : : . : : * : : : . . ** . : : * . *

FrsG_MT1_Cro      DLLKI----DVENAEDV-----LRGISDADWPKIRQLVMEVH--DVD-----
FrsF_MT_Cro      RLKRGEAANELNRYRYEVVLRKGPCQARSLASAAAEPWSSLGSLSACRERLSVGGDALRV
FrsE_MT_Cro      RLKRGEAANELNRYRYEVVLRKGPCQARSLASAAAEPWSSLGSLSACRERLSVGGDALRV
FrsG_MT2_Cro      RLKRGEAANELNRYRYEVVLRKGPCQARSLASAAAEPWSSLGSLSACRERLSVGGDALRV
                  * : : : : . * : * . * : : * . : * . . * .

FrsG_MT1_Cro      -----GRLA-----CIVELLRERGRYVIEHQDRL
FrsF_MT_Cro      TGVPNALLHGEEAAARELKAGGSPSALLARLDENGGVRPEALRRLGAELGWRMLATWSRQ
FrsE_MT_Cro      TGVPNALLHGEEAAARELKAGGSPSALLARLDENGGVRPEALRRLGAELGWRMLATWSRQ
FrsG_MT2_Cro      TGVPNALLHGEEAAARELKAGGSPSALLARLDENGGVRPEALRRLGAELGWRMLATWSRQ
                  * : . : * * * : : : . *

FrsG_MT1_Cro      LQNTAIHCLYAVHASCASASADA--VPKAEPAAVWRSRSA--LLRDVQAEQRSQLPDY
FrsF_MT_Cro      AGH--FDAVFV-RGEDGEALDGVYQPAGALQPLSGYVNNPANFEQYAAIRRYALEQLPEY
FrsE_MT_Cro      AGH--FDAVFV-RGEDGEALDGVYQPAGALQPLSGYVNNPANFEQYAAIRRYALEQLPEY
FrsG_MT2_Cro      AGH--FDAVFV-RGEDGEALDGVYQPAGALQPLSGYVNNPANFEQYAAIRRYALEQLPEY
                  : : : : : . . . * * . . . : * : : . . * : : . *****

FrsG_MT1_Cro      MQPNHLVLLDAFPLTANGKLDRRALPTPEQAALRTRDVPDPA-----
FrsF_MT_Cro      MVPAAIVLLDALPLTPNGKLDRRALPAPEFGGTGYRAPESERE-----
FrsE_MT_Cro      MVPAAIVLLDALPLTPNGKLDRRALPAPEFGGTGYRAPESEREQLLARLFGEVLGLPQVG
FrsG_MT2_Cro      MVPAAIVLLDALPLTPNGKLDRRALPAPEFGGTGYRAPESERE-----
                  * * : ***** : ** * : : : : : * : : * : :

```

gelb: konservierte Regionen von MT-Domänen nach Labby et al.

rosa: Core A9-Core A11 Bereiche der A-Domänen

Abbildung A12: Alignment der MT-Domänen aus dem *cvfrs*-BGC.

```
#  
#  
# Percent Identity Matrix MT-Domänen Cand. B. crenata - created by Clustal2.1  
#  
#  
1: Acre_FrsG_MT_1 100.00 24.36 24.36 24.36  
2: Acre_FrsE_MT 24.36 100.00 100.00 98.10  
3: Acre_FrsF_MT 24.36 100.00 100.00 98.10  
4: Acre_FrsG_MT_2 24.36 98.10 98.10 100.00
```

Abbildung A13: Percent Identity Matrix des Alignments der MT-Domänen aus dem *befrs*-BGC.

```
#  
#  
# Percent Identity Matrix MT-Domänen C. vaccinii - created by Clustal2.1  
#  
#  
1: FrsG_MT1_Cro 100.00 26.55 26.55 23.58  
2: FrsF_MT_Cro 26.55 100.00 100.00 100.00  
3: FrsE_MT_Cro 26.55 100.00 100.00 100.00  
4: FrsG_MT2_Cro 23.58 100.00 100.00 100.00
```

Abbildung A14: Percent Identity Matrix des Alignments der MT-Domänen aus dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

Acre_FrsE_MT      EIEAALCRHPLVSVAVVIAREDPGHKQLVGYVVLDDAAALQRDAEETRQVQAWRQVYD
Acre_FrsF_MT      EIEAALCRHPLVSVAVVIAREDPGHKQLVGYVVLDDAAALQRDAEETRQVQAWRQVYD
Acre_FrsG_MT_2    EIEAALCRHPLVSVAVVIAREDPGHKQLVGYVVLDDAAALQRDETRQVQAWRQVYD
FrsE_MT_Cro       EIEAALCRHPSVAQA AVIAREDPGHKQLVGYVVLDDAAALQRDEDSERQVEAWQQVYD
FrsF_MT_Cro       EIEAALCRHPSVAQA AVIAREDPGHKQLVGYVVLDDAAALQRDEDSERQVEAWQQVYD
FrsG_MT2_Cro      EIEAALCRHPSVAQA AVIAREDPGHKQLVGYVVLDDAAALQRDEDSERQVEAWQQVYD
***** *.*.*****.*****.*****.*****.*****.*****.*****.*****

Acre_FrsE_MT      TLYEAHCQQPFGENFGGWSSYDQGQPLPEEMRAWRQATVERIRELQPRRVL EIGVGSGL
Acre_FrsF_MT      TLYEAHCQQPFGENFGGWSSYDQGQPLPEEMRAWRQATVERIRELQPRRVL EIGVGSGL
Acre_FrsG_MT_2    TLYEAHCQQPFGENFGGWSSYDQGQPLPEEMRAWRQATVERIRELQPRRVL EIGVGSGL
FrsE_MT_Cro       TLYDAHQHQPFGENFGGWSSYDQGQPLPLAQMREWRSATVERIRELRPRRL EIGVGSGL
FrsF_MT_Cro       TLYDAHQHQPFGENFGGWSSYDQGQPLPLAQMREWRSATVERIRELRPRRL EIGVGSGL
FrsG_MT2_Cro      TLYDAHQHQPFGENFGGWSSYDQGQPLPLAQMREWRSATVERIRELRPRRL EIGVGSGL
***.*.*****.*****.*****.*.*.*****.*****.*****.*****

Acre_FrsE_MT      LLAPLAPECEAYWGTDLSPPTTIAVLERQLSEQTFGDRVRLFALGAHELLKLPENFDVIV
Acre_FrsF_MT      LLAPLAPECEAYWGTDLSPPTTIAVLERQLSEQTFGDRVRLFALGAHELLKLPENFDVIV
Acre_FrsG_MT_2    LLAPLAPECEAYWGTDLSPPTTIAVLERQLSEQTFGDRVRLFALGAHELLKLPENFDVIV
FrsE_MT_Cro       LLAPLADACEAYWGTDLSPATIAVLEKQLETQSCRDKVRLFALGAHELARLPAMRFDIV
FrsF_MT_Cro       LLAPLADACEAYWGTDLSPATIAVLEKQLETQSCRDKVRLFALGAHELARLPAMRFDIV
FrsG_MT2_Cro      LLAPLADACEAYWGTDLSPATIAVLEKQLETQSCRDKVRLFALGAHELARLPAMRFDIV
***** *****.*****.*.*.*****.*****.*.*.*****.*****

Acre_FrsE_MT      INSVLQYFPNAAYLTEVIKQSLERLSPGGALYLG DVRNLALLDVFATAVDLCQVGTETDA
Acre_FrsF_MT      INSVLQYFPNAAYLTEVIKQSLERLSPGGALYLG DVRNLALLDVFATAVDLCQVGTETDA
Acre_FrsG_MT_2    INSVLQYFPNAAYLTEVIKQSLERLSPGGALYLG DVRNLALLDVFATAVDLCQVGTETDA
FrsE_MT_Cro       INSVLQYFPNAAYLGEVIEQALARLEAGGALYLG DVRNLELLPSFAAAVELRQSEPEVDA
FrsF_MT_Cro       INSVLQYFPNAAYLGEVIEQALARLEAGGALYLG DVRNLELLPSFAAAVELRQSEPEVDA
FrsG_MT2_Cro      INSVLQYFPNAAYLGEVIEQALARLEAGGALYLG DVRNLELLPSFAAAVELRQSEPEVDA
***** *****.*.*.*****.*****.*****.*****.*****.*.*

Acre_FrsE_MT      ASLRRRVEQRLLAEKELLVSPAFFSRLRETLRPQIGAVDIRLKRGEAVNELNRYRYEAVLY
Acre_FrsF_MT      ASLRRRVEQRLLAEKELLVSPAFFSRLRETLRPQIGAVDIRLKRGEAVNELNRYRYEAVLY
Acre_FrsG_MT_2    ASLRRRVEQRLLAEKELLVSPAFFSRLRETLRPQIGAVDIRLKRGEAVNELSRRYEAVLY
FrsE_MT_Cro       AALQRRVSQRLLAEKELLAPDFFSRLREQLPQIGAVDIRLKRGEAANELNRYRYEVVLR
FrsF_MT_Cro       AALQRRVSQRLLAEKELLAPDFFSRLREQLPQIGAVDIRLKRGEAANELNRYRYEVVLR
FrsG_MT2_Cro      AALQRRVSQRLLAEKELLAPDFFSRLREQLPQIGAVDIRLKRGEAANELNRYRYEVVLR
*.*.***.*****.*.*.*****.*****.*****.*****.*****.*.*

Acre_FrsE_MT      KEPCEAVSLSKVETQTWSDVGNMSACRTRLMQYASMWRIEGIPNARLHGEVLVTRRLKSE
Acre_FrsF_MT      KEPCEAVSLSKVETQTWSDVGNMSACRTRLMQYASMWRIEGIPNARLHGEVLVTRRLKSE
Acre_FrsG_MT_2    KEPCEAVSLSEVETQTWSDVGNMSACRTRLMQYAPMWRIEGIPNARLHGEVLVTRRLKSE
FrsE_MT_Cro       KGPCQARSLASAAAEPWSSLGSLSACRERLSVGGDALRVTGVPNALHGEAAAARELKAG
FrsF_MT_Cro       KGPCQARSLASAAAEPWSSLGSLSACRERLSVGGDALRVTGVPNALHGEAAAARELKAG
FrsG_MT2_Cro      KGPCQARSLASAAAEPWSSLGSLSACRERLSVGGDALRVTGVPNALHGEAAAARELKAG
*.*.*.*.*.***.***.***.***.***.***.***.***.***.***.***.***.***.***.***

Acre_FrsE_MT      EGLASWRRWLDESGGVEPEDLYRLGAELGYRVVLTWSNQPDHFDAVFLPEQQRGALDAVY
Acre_FrsF_MT      EGLASWRRWLDESGGVEPEDLYRLGAELGYRVVLTWSNQPDHFDAVFLPEQQRGALDAVY
Acre_FrsG_MT_2    DGLASWRRWLDESGGVEPEDLYRLGAELGYRVVLTWSNQPDHFDAVFLPEQQRGALDAVY
FrsE_MT_Cro       GSPSALLARLDENGGVRPEALRRLGAELGWRMLATWSRQAGHFDAVFRGEGDGEALDGVY
FrsF_MT_Cro       GSPSALLARLDENGGVRPEALRRLGAELGWRMLATWSRQAGHFDAVFRGEGDGEALDGVY
FrsG_MT2_Cro      GSPSALLARLDENGGVRPEALRRLGAELGWRMLATWSRQAGHFDAVFRGEGDGEALDGVY
.:. *****.*.*.*****.*.*.*****.*.*.*****.*.*.*****.*.*

```

```

Acre_FrsE_MT      RTPPVLKPLSGYVNNPANFEQFAAIRRYVGEQLPDYMPAALVLLLEGLPLTPNGKLDLRRRA
Acre_FrsF_MT      RTPPVLKPLSGYVNNPANFEQFAAIRRYVGEQLPDYMPAALVLLLEGLPLTPNGKLDLRRRA
Acre_FrsG_MT_2    RTPPVLKPLSGYVNNPANFEQFAAIRRYVGEQLPDYMPAALVLLLEGLPLTPNGKLDLRRRA
FrsE_MT_Cro       QPAGALQPLSGYVNNPANFEQYAAIRRYALEQLPEYMPAAIVLLDALPLTPNGKLDLRRRA
FrsF_MT_Cro       QPAGALQPLSGYVNNPANFEQYAAIRRYALEQLPEYMPAAIVLLDALPLTPNGKLDLRRRA
FrsG_MT2_Cro      QPAGALQPLSGYVNNPANFEQYAAIRRYALEQLPEYMPAAIVLLDALPLTPNGKLDLRRRA
                  :   .*:*****:*****. ****:*:*:*:*:*:*:*:*:*:*:*:*:*:*
Acre_FrsE_MT      LPAPEFGTACYRAPGSKQE
Acre_FrsF_MT      LPAPEFGTACYRAPGSKQE
Acre_FrsG_MT_2    LPAPEFGTACYRAPGSEQE
FrsE_MT_Cro       LPAPEFGGTGYRAPESERE
FrsF_MT_Cro       LPAPEFGGTGYRAPESERE
FrsG_MT2_Cro      LPAPEFGGTGYRAPESERE
                  ***** : **** *:*

```

grün: S-Adenosylmethionin-Bindungsstelle

Abbildung A15: Alignment der N-MT-Domänen aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

```
#  
#  
# Percent Identity Matrix - created by Clustal2.1  
#  
#  
1: Acre_FrsE_MT      100.00  100.00  98.20  71.94  71.94  71.94  
2: Acre_FrsF_MT      100.00  100.00  98.20  71.94  71.94  71.94  
3: Acre_FrsG_MT_2    98.20   98.20 100.00  71.54  71.54  71.54  
4: FrsE_MT_Cro       71.94   71.94  71.54 100.00 100.00 100.00  
5: FrsF_MT_Cro       71.94   71.94  71.54 100.00 100.00 100.00  
6: FrsG_MT2_Cro      71.94   71.94  71.54 100.00 100.00 100.00
```

Abbildung A16: Percent Identity Matrix des Alignments der N-Methyltransferasen aus dem *bcfrs*-BGC und *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

FrsG1_Acre      GRADQQVKVRGFRIELGEIETVLRQHPGVEDARVVVQTLHDNDCRLVAYLQPSERTAAPL
FrsG1_Chro      GRADQQVKVRGFRIEPGEIETALRGHPGVEDARVVVKAMDGNDQRLIAYLLPSERAAAPL
*****
FrsG1_Acre      RRWLKVRRDAAEQGARAFELPNGMPIFHHTMETEFLYEEIFEDLVYFKHGIRLDDGACV
FrsG1_Chro      RRWLRAKSSVAAQNAREIELPNGLPVFHHNAAETEFLYEEIFEDRIYLLKHGIRLNDACV
*****
FrsG1_Acre      FDVGANIGLFLFVQGHCNATIFAFEPIPPVFRTLTLNAEVHGDVKVRLFDCGLSDAVRQ
FrsG1_Chro      FDVGANIGLFTLVFVQHCNATVFAFEPIPPVFGTSLNAAVHGGKVRFLFDCGLSDAARQ
*****
FrsG1_Acre      ERFTFYPNDTLISTSRNSPEAIRGMVKSFLINQHGNSGDDAEV GELLDERLASQDYMVCV
FrsG1_Chro      ETFTFYPNDTLISSSRNSAEATRRMVKSFLINQHGDSDGE--AVDELLEERLSSQQYVCG
*****
FrsG1_Acre      LRTLSDVITEHNVDRIDLKIDVENAEYNVLQGIVESDWPKIRQLVMEVHDVDGRLRRIV
FrsG1_Chro      LRSLSDVVAEHGVERIDLKIDVENAEYDVLRGISDADWPKIRQLVMEVHDVDGRLACIV
*****
FrsG1_Acre      DLLLHRGYRVIHEQDRLLRDTAIHCLYAVHVS RDAEMRPGHHAENTSLFWNCRADLLRDV
FrsG1_Chro      ELLRERGYRVIHEQDRLLQNTAIHCLYAVHASC GSASADAVPKAEPAAVWRSRALLRDV
*****
FrsG1_Acre      RANLRKRLPDYMQPNHMLLDTFPLTPNGKLD RRALPAPEQAAMRVRDIEPA
FrsG1_Chro      QAE LRSQLPDYMQPNHLVLLDAFPLTANGKLD RRALPTPEQAALRTRDVPDA
*****

```

Abbildung A17: Alignment der O-MT-Domänen aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

```

#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
      1: FrsG1_Acre_Burk  100.00   73.08
      2: FrsG1_Chro     73.08   100.00

```

Abbildung A18: Percent Identity Matrix des Alignments der O-Methyltransferasen aus dem *bcfrs*-BGC und *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

C.v._FrsE_C4_Epimer --PVESDRNERVDQACGDLPATP-----IIHWMLAAPPYRRFNQSLLRTPGGLKRDDL 53
C.v._FrsA_C1_start -----MKNSESPIHHFQASSAQLDVWISQEVSPN--LPNNIAEYLNLAGSLDAGLFL 50
C.v._FrsD_C2_start -----MEIWLAAQQLMPD--SPNNIAEYLNLAGSLDAGLFL 33
C.v._FrsE_C3_LCL -----MPIPEIEVLPLSYAQRRLWFTHRFNGPS-PTYNIPIALSLSGEPEQPALQ 49
C.v._FrsF_C7_LCL SGDGELRPPLRRLRPQTLPLSYAQQRLWFIHHLEGPS-PTYNIPVALRLSGALETDLVLE 59
C.v._FrsG_C9_LCL -DPAAVRPPLRQARPKRLPLSYAQQRLWFLYRFEGPS-STYNIPLALRLKGDHLPEALQ 58
C.v._FrsE_C5_DCL -----RPLEDLVPLSPLQKGLLFHGLYDPAGVDPYIERVTYQLEGLDPAAMK 48
C.v._FrsF_C6_DCL -----MNPKKIEAAYPLSPLQKGLFHAGYDLQSADSYVAQLFLDFEGELDGAAMR 51
C.v._FrsG_C8_DCL -----MNPKKIEAAYPLSPLQKGLFHAGYDLQSADSYVAQLFLDFEGELDGAAMR 51
                                     :                               * . :

C.v._FrsE_C4_Epimer AALQALLDHHDALRRLDAAASDEALLSIPPAGSVRAGDCLRRIDAVGRDEAEWHLLAR 113
C.v._FrsA_C1_start  QALSQVASESAELQYNFRHDGLQL---TKFRR--DDEGWEPDFIDVSTHGEP-EHAALRA 104
C.v._FrsD_C2_start  KTLRQVASETPALQVNFSDIEDGRP---CPVSR--VHEDWSPDFIDVSTHGEP-EHAALRA 87
C.v._FrsE_C3_LCL    AALRDVLARHESLRTLVESEEGAPAQHILPP--DAAGTFC--LAVQTAGSA-AEQASL 104
C.v._FrsF_C7_LCL    AALLDVVRRHESLRTVYVQSADDTPLQCIVLSV--EQLDDCR--LIVHEPLGE-TRLGAVL 114
C.v._FrsG_C9_LCL    QAVADIVERHESLRTVFVDNDGV-PEQRISQA--EQALP-S--LPRVEVADE-AELQRAL 111
C.v._FrsE_C5_DCL    RAAHGLLQRHGHLRACFVDLGGKGPVQIISQL--EALPWQDIDL SMLAGDER-QAMLAQI 105
C.v._FrsF_C6_DCL    SAADALMRRHANLRAGFVHPGGQEPVQVVLRE--VVAGWEERDWRGRNPQEA-AEAQSAW 108
C.v._FrsG_C8_DCL    SAADALMRRHANLRAGFVHPGGQEPVQVVLRE--VVAGWEERDWRGRNPQEA-AEAQSAW 108
                                     : : . *:

C.v._FrsE_C4_Epimer EGEEAAERLDCEAGLLLQAVWLDAGDEAGRLLLVIHHLAVDGVSWRVLLPDLQAWQAAS 173
C.v._FrsA_C1_start  MRERVEKPFDLARDALFRWTLIRLADERHIFCHVYHHIAMDGAGYVMLLQRIAEVYGALR 164
C.v._FrsD_C2_start  MRERVEKPFDLARDALFRWTLIRLADERHIFCHVYHHIAMDVAGYVMLLQRIAEVYGALR 147
C.v._FrsE_C3_LCL    -ETACRHCFDLSREMLRAGFLADGAEPVLLLLLHIIAADGDSLPLVLARDLELAYRSRL 163
C.v._FrsF_C7_LCL    -SEASRHCFDLSCEPLRAELFPLDSSGQVLLLLLHIIAGDGGSLPLVADDLAAAYEARC 173
C.v._FrsG_C9_LCL    -REAAEHRFDLSSETPLRCTLFRLGEQEWLLLLLHIIAGDGGSLPLGRELDAAYAARC 170
C.v._FrsE_C5_DCL    REQVRSQRFDLGRAPLLSFTLIRLAADRHLIMNNHII LLDGWSEPLLWRELMTYRNGG 165
C.v._FrsF_C6_DCL    QEADRERRFELSQPPLRFGLRLLPAGRSQLVVTYHHI LLDGWSPLLLLEELLALYRAGG 168
C.v._FrsG_C8_DCL    QEADRERRFELSQPPLRFGLRLLPAGRSQLVVTYHHI LLDGWSPLLLLEELLALYRAGG 168
                                     . :: : : : : ** : * . * : :

C.v._FrsE_C4_Epimer QDRAIALDPVGASFRLWLSHLGQEAR-----SSQREAEALHWKWAALAEED-PLL----GP 223
C.v._FrsA_C1_start  EGQPAP---ACGFADADAIVREEER--YRQSEQFAVDRAFQARSAELAT-----AEPP 213
C.v._FrsD_C2_start  EGQPAP---ACGFADADAIVREEER--YRQSEQFAVDRAFQARSAELAT-----AEPP 196
C.v._FrsE_C3_LCL    RGLAPEWRPLPVQYADYALWQQLLGDLEVPDLSAARQLVYVWREALRGMPDVLELPTDHP 223
C.v._FrsF_C7_LCL    QGREPAWRPLPVQYADYTLWQRQLLGNERDPDSLIGRFAYWKQALAGLPDQLCLPTDRP 233
C.v._FrsG_C9_LCL    QGRAPDWQPLAVQYADYTLWQRQLLGEENDPESL IAAQFAYWKRTLGAPEQLTLPTDRP 230
C.v._FrsE_C5_DCL    ---DLGAMPRTVTPYRDYLVWLGRRDH-----EADRQAWRHYLAELETPTLLSPEPP 213
C.v._FrsF_C6_DCL    ---GDGGLPEATPYSAYLWQLQERDR-----AAACEAWGGYLEGLEGPTLLAQEGQ 216
C.v._FrsG_C8_DCL    ---GDGGLPEATPYSAYLWQLQERDR-----AAACEAWGGYLEGLEGPTLVSTGHP 216
                                     :                               *

C.v._FrsE_C4_Epimer RPYDSARDATRTRQSLSLA----LPPAVTQALLQATARFHAHANDVLLTVFALAVAVWR 279
C.v._FrsA_C1_start  LP-----AADGPFLAFAQTAVIPEDACGGLRMTAER-LGVSQSRLLTAAIVAYFHRW- 264
C.v._FrsD_C2_start  LP-----AADGPFLAFAQTAVIPEDACGGLRMTAER-LGVSQSRLLTAAIVAYFHRW- 247
C.v._FrsE_C3_LCL    RPAVA----SHRGGKAPFA----LSAAAHARLKTAAQ-QGATLSMALQAGLAALLHRL- 273
C.v._FrsF_C7_LCL    RPAMA----SYQGDYCEFR----LDAETLKQLKTLARR-HGATLSMALQAAALAALLSRL- 283
C.v._FrsG_C9_LCL    RPAVA----SYRGRYLPFE----LDAGLHAELRRLARD-HDATLSMLLQAGLAALFTRL- 280
C.v._FrsE_C5_DCL    AA-----YVDQETYSLA----LPPALAQALARAARAE-LGITLNTLVQAGWGRVLACL- 260
C.v._FrsF_C6_DCL    ADKGA----AAAQARLSLE----LPAELTQALTRQARQ-QGVTLNTLLQAAWGMLLGLK- 266
C.v._FrsG_C8_DCL    QDH-----AEQQRRAWR----LPAELTQALTRQARQ-QGVTLNTLLQAAWGMLLGLK- 263
                                     :                               * * . : .

```



```

C.v._FrsE_C4_Epimer RRRLPEAPAELLFDLEGHGREARDTGIDLSRTVGVFTSLFPVRLALDCAGLDEALKGGDS 339
C.v._FrsA_C1_start -----GGQQEILFRLAVSARSDA-----TRHAPGHLAHLPLLASLP-----PRAS 305
C.v._FrsD_C2_start -----GGQQEILFRLAVSARSDA-----TRHAPGHLAHLPLLASLP-----PRAS 288
C.v._FrsE_C3_LCL -----GAGTDIAVGGLLAGRGEE----SLKELIGFFVNAWVLRVDVS-----GRPN 315
C.v._FrsF_C7_LCL -----GAGDDIPLGCPAIGRTDD----ALANLVGFFVNTWVLRVDTS-----GRPD 325
C.v._FrsG_C9_LCL -----GAGEDIPLGCGIAGRTDD----ALNDMVGFFVNSWVLRADTS-----GNPD 322
C.v._FrsE_C5_DCL -----TMSQDVVFGSNVAGRPAEL--GGIEDMVGLFINTVPLRVRWT-----RGES 304
C.v._FrsF_C6_DCL -----NLSRDVVFGITVAGRPGEL--PGVERMIGLFINTVPVRLRWS-----AGET 310
C.v._FrsG_C8_DCL -----NLSRDVVFGITVAGRPGEL--PGVERMIGLFINTVPVRLRWS-----AGET 307
      :: . . * * : :

C.v._FrsE_C4_Epimer LGRLLKSVKE-QLRAIPDRGMGFGLLRH-LNPGARGEL--AAL-SSPQIGFNYLGRFTAA 394
C.v._FrsA_C1_start LADIARQLDGEVERMRPHTRYRAEDIVRDQAGAGLGRGAQGPVINLMPFAYRFEFGACRV 365
C.v._FrsD_C2_start LADIARQLDGEVERMRPHTRYRAEDIVRDQAGAGLGRGAQGPVINLMPFAYRFEFGACRV 348
C.v._FrsE_C3_LCL FPALLRRVREQALQAYSHQDLPFEWLVEQINPARS--TSHHPLFQVALVLQNNLSADFH 373
C.v._FrsF_C7_LCL FATLLGRVRQALAAAYSHQDVPFERLVELLNPARS--ASHHPLFQVNLALQNNIFPLFRF 383
C.v._FrsG_C9_LCL FITLLGRVREQALAAAYHQDAPFERLVELINPSRS--SAHHPLFQVNLTLQNNALPDFRL 380
C.v._FrsE_C5_DCL IADMLVRLQAEQVGLLEHQYLDLAEIQE-----L--AGLGDLDFTVYAFENYPVFGDGG 356
C.v._FrsF_C6_DCL VAGLVGRLQREQAGLLDHQHLDLVEIQR-----L--AGQRQLFDTLFIFENYPFDSQAM 362
C.v._FrsG_C8_DCL VAGLVGRLQREQAGLLDHQHLDLVEIQR-----L--AGQRQLFDTCVFNYPVDAAM 359
      . : : . : . : .

C.v._FrsE_C4_Epimer EGEDWQPAGCAGIEGGQDPDMLPHLLSFNAQTLDREQGPHELHAIWSWAGELFDQEQIAE 454
C.v._FrsA_C1_start ESA-----HQLTVGVL-DTL---EVAV----HDRKNGDGLHLDLYASERGCPPPELRR 410
C.v._FrsD_C2_start ESA-----HQLTVGVL-DTL---EVAV----HDRKNGDGLHLDLYASERGCPPPELRR 393
C.v._FrsE_C3_LCL EGL---AVDQRLLTGTAKFDL---AFNLFERQDGDGRPQGLDGELEFASDLFDRAGAER 427
C.v._FrsF_C7_LCL SGR---EVSLETVGNRTAKFDL---FFNLFEMPSDEPQTQYLQGYVEYATELFDRAVER 437
C.v._FrsG_C9_LCL DGL---EVSLSPIEADTAKFDL---FFNLFEIFGLDQGAGGLQGGVEYATDLYEQATVEQ 434
C.v._FrsE_C5_DCL ASA-A-GPRVTGVSGG-S--TT---HYPLG---LIVNPQSGLSFLFSYRPDLYLPEDIQR 405
C.v._FrsF_C6_DCL APD-LGQASLRRVSGG-EQHES---HYPLT---LMAVPREMSLSLYSYDAQRFDKGVVQG 414
C.v._FrsG_C8_DCL EQP-SGGLQLRGVSGG-DRYMT---HYPLS---LMIEPGPELKLNIYQPERFEPEAIER 411
      : . *

C.v._FrsE_C4_Epimer LAQLWRQAAVALAEHASRPEAGGRSPSDLPLVGLEQT-----QIERLEAEY--- 500
C.v._FrsA_C1_start HALRLARFIVEAAAEPSPVSD-----IELLDEAERRQLLDVWNRTGPDHGQATFPQL 463
C.v._FrsD_C2_start HALRLARFIVEAAAEPSPVSD-----IELLDEAERRQLLDVWNRTGPDHGQATFPQL 446
C.v._FrsE_C3_LCL LVARFVRLLERWSMQPAHLIAE-----LDLDDDDERRQAV----- 462
C.v._FrsF_C7_LCL LLTRFRLLLEAVAAEPSRPMSD-----IELLDEAERR----- 469
C.v._FrsG_C9_LCL FMRHFQRLLRQVAAPVTAPIAA-----LDLSEPQSAEADCAARPAASDSI----- 480
C.v._FrsE_C5_DCL IAGYLQTLQAFADDPAQSVAG-----LELAPVEQAGWL----- 439
C.v._FrsF_C6_DCL LLTRFRLLLEAVAAEPSRPVSD-----IELLDEAERR----- 446
C.v._FrsG_C8_DCL LSAQLTRLLGVIAAEPSPVSD-----IEL----- 436
      : : : : *

```

Abbildung A19: Alignment der C-Domänen aus dem *cvfrs*-BGC.

```

#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: C.v._FrsE_C4_Epimerisation 100.00  17.67  17.83  21.00  24.22  22.05  20.62  21.11  21.80
2: C.v._FrsA_C1_starter      17.67 100.00  91.26  19.82  21.25  18.20  21.15  22.73  22.93
3: C.v._FrsD_C2_starter      17.83  91.26 100.00  19.09  22.36  18.69  20.84  21.14  21.32
4: C.v._FrsE_C3_LCL          21.00  19.82  19.09 100.00  57.30  53.04  24.07  25.86  26.00
5: C.v._FrsF_C7_LCL          24.22  21.25  22.36  57.30 100.00  62.23  23.26  30.14  26.64
6: C.v._FrsG_C9_LCL          22.05  18.20  18.69  53.04  62.23 100.00  23.95  25.92  25.82
7: C.v._FrsE_C5_DCL          20.62  21.15  20.84  24.07  23.26  23.95 100.00  46.68  48.02
8: C.v._FrsF_C6_DCL          21.11  22.73  21.14  25.86  30.14  25.92  46.68 100.00  85.55
9: C.v._FrsG_C8_DCL          21.80  22.93  21.32  26.00  26.64  25.82  48.02  85.55 100.00

```

Abbildung A20: Percent Identity Matrix des Alignments der C-Domänen aus dem *cvfrs*-BGC.


```

FrsE_Condensation_4_Epimer -QCMGHAPDALLFDLEGHGRETQDTAIDLRTVGVWFTSLFPVVRRLDAVLDLDDALGEGAS 338
FrsA_Condensation_1_start FRRWDGQ-NEMRFRLAVSARS--EVT--MQAPGYMAHALPLQASFTP-----HTS 289
FrsD_Condensation_2_start FRRWDGQ-NEMRFRLAVSARS--EVT--MQAPGNMAHALPLQASFTP-----RTS 287
FrsE_Condensation_3_LCL LYRLGAG-SDVVIGLLAGRN--DEA--LKDLIGFFVNAWVLRDLSG-----HPD 314
FrsF_Condensation_7_LCL LNRMGAG-DDIPLGSPFIAGRT--DDA--LINLVGFFVNTWVLRVDTSG-----RPD 325
FrsG_Condensation_9_LCL FTRLGVG-NDIPLGCGIAGRT--DEA--LGDLVGFVNTWILRADTSG-----DPD 324
FrsE_Condensation_5_DCL LGCLTTS-QDVMFGSNVAGRP--AELNGIEDMIGLFINTIPLRVRSR-----GES 301
FrsF_Condensation_6_DCL LGKLSSS-RDVVFGITVAGRP--GELPGVERMIGLFINTVPLRLRWRA-----GET 301
FrsG_Condensation_8_DCL LGKLSSS-RDVVFGITVAGRP--GELPGVERMIGLFINTVPLRLRWRA-----GET 301
      : :      *      *      : :
      : :      *      *      : :

FrsE_Condensation_4_Epimer LGRLKKSVEQL-HALPDRGLGFGLLRYLNQGTAAELA AHGQPQIGFNYL-----GRFA 391
FrsA_Condensation_1_start LADIVRQLDGEVRCMRPHIRYRAEDIVRDWASTGGVQGAQGP-VINIMPF-----SYAFD 343
FrsD_Condensation_2_start LADIVRQLDGEVRCMRPHIRYRAEDIVRDWASTGGVQGAQGP-VINIMPF-----SYAFD 343
FrsE_Condensation_3_LCL FHVLLRRVREQALQAYSHPDLPFEWLVQLNPI--RSTSYHP-LFQVVLVQLNNQRARFR 371
FrsF_Condensation_7_LCL FATLLGRVRSRALAAYAHQDVPFDRLVELLNPV--RSISHHP-LFQVNLALQNNVLPKFR 382
FrsG_Condensation_9_LCL FVTLLGRVREQALAAAHQDAPFERLVELINPV--RSSAHP-LFQVNLTLQNNALPEFR 381
FrsE_Condensation_5_DCL IGDVLKRIQSEQVDLLEHQYLD-----LVEIQ--SQASHRD-LFDSVYAFENYPVHAND 352
FrsF_Condensation_6_DCL VAELLERLQREQARLLEYQYLD-----LAEIQ--RLAGQRQ-LFDTLFIENYFPDAQA 352
FrsG_Condensation_8_DCL VVELLERLQREQARLLEYQYLD-----LAEIQ--RLAGQRQ-LFDTLCVFENYPVNAKA 352
      . : : : . : :
      . : : : . : :

FrsE_Condensation_4_Epimer ASEGGDWQLASDVGIEAGQDPEMPLPHPLSFD AHTLDRTHGPELTAIWSWGSSELFSSDEI 451
FrsA_Condensation_1_start F---GECRV TSAHQLTVG-----LLNAL--EVAVHDRKIGDGLHIDL YAPQACGSPVQL 392
FrsD_Condensation_2_start F---GECRV TSAHQLTVG-----VLNAL--EVAVHDRKIGDGLHIDL YAPQACGSPVQL 390
FrsE_Condensation_3_LCL L---GGLVVE-QQVLGTGTA-KFDLAFNL--FETMSDDGHPLGVTGDLEYACDLFDPPSA 424
FrsF_Condensation_7_LCL F---FGQAVA-LETVSSRTA-KFDLFFNL--CEMPADGSQPCYLHGYVEYASDLFNRDVT 435
FrsG_Condensation_9_LCL L---DGMQVS-LHRIVSDIA-KFDLFFNL--YETFDQDGRALGLQGA VEYACDLYERATV 434
FrsE_Condensation_5_DCL E---DEASGP-RVKVVS GG---STTHYPL--GLIVN---PQAGLSLLFSYR PDCYRRCDI 400
FrsF_Condensation_6_DCL I---APALGR TVLSRINGGE-QHDSHYPV--TLMAV---PRETLTLYLSYQSGRFEHGTM 403
FrsG_Condensation_8_DCL I---VQQDEGFGLRHISGGD-RYMTHYPL--SVMIE---PGERMTLNL IYRPASFDAAK- 402
      :
      :

FrsE_Condensation_4_Epimer AELAQLWQQAAMALAEHVTRPGAGGRTPSDLP LVLHQAQIEQLEVEYPRI----- 502
FrsA_Condensation_1_start QQHVRRLARFIEVATA----- 408
FrsD_Condensation_2_start QQHVRRLARFIEVAAA----- 406
FrsE_Condensation_3_LCL VRLAYRLSRLLEIWSAAPSQS-----IAIDLLE--RSEREQALLEWNATTRPLPALTL 476
FrsF_Condensation_7_LCL ERLLTRFRTLLEAVVTDSSCP-----IADIDLLT--ADERQQLLVEWNATDRPLPKVTL 487
FrsG_Condensation_9_LCL QRFIYHFVRLLEVVAAPTAA-----ISALDLS D--EAVASSPG-ALDLSASITVDDTI 485
FrsE_Condensation_5_DCL ERIAAYLQCVLEAFVDSQP-----IAQLDLLP--PEQ-ANGIAQWNDTQHACPSADL 451
FrsF_Condensation_6_DCL ENLLTRFRTLLEAVVTDSSCP-----IVDIDLLT--ADERQQLLVEWNATDRPLPKVTL 455
FrsG_Condensation_8_DCL -RLGAQLIRLLEAIATVPQSP-----IDTLPLWD--KSERRQLLEWWSGKALDSGEITL 453
      .

FrsE_Condensation_4_Epimer ----- 502
FrsA_Condensation_1_start ----- 408
FrsD_Condensation_2_start ----- 406
FrsE_Condensation_3_LCL AEA FETQAALTPEAVALAFGDEV- 499
FrsF_Condensation_7_LCL PEWF EAQVERTPTAMAVLCDEVAL 511
FrsG_Condensation_9_LCL IGRFAEQVRRAPDATVLT YRGETL 509
FrsE_Condensation_5_DCL AQLFERQVRLTPDASALTFGSQTL 475
FrsF_Condensation_6_DCL PEWF EAQVERTPTAMAVLCDEVAL 479
FrsG_Condensation_8_DCL AELFEAQATRQPNVALEGP DGCV 477

```

Abbildung A21: Alignment der C-Domänen aus dem *befrs*-BGC.

```

#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
1: FrsE_Condensation_4_Epimerisation 100.00  15.99  14.29  20.05  18.76  18.67  20.00  21.70  22.04
2: FrsA_Condensation_1_starter      15.99 100.00  82.51  18.00  15.25  17.59  19.27  17.14  18.02
3: FrsD_Condensation_2_starter      14.29  82.51 100.00  18.34  18.09  17.68  20.16  17.49  18.64
4: FrsE_Condensation_3_LCL          20.05  18.00  18.34 100.00  50.10  47.58  27.78  26.96  27.51
5: FrsF_Condensation_7_LCL          18.76  15.25  18.09  50.10 100.00  53.35  27.08  35.81  25.74
6: FrsG_Condensation_9_LCL          18.67  17.59  17.68  47.58  53.35 100.00  25.32  22.60  23.13
7: FrsE_Condensation_5_DCL          20.00  19.27  20.16  27.78  27.08  25.32 100.00  44.00  44.19
8: FrsF_Condensation_6_DCL          21.70  17.14  17.49  26.96  35.81  22.60  44.00 100.00  81.13
9: FrsG_Condensation_8_DCL          22.04  18.02  18.64  27.51  25.74  23.13  44.19  81.13 100.00

```

Abbildung A22: Percent Identity Matrix des Alignments der C-Domänen aus dem *befrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

FrsG_PCP8          -----QLLARLFGEVLGLPQVGLHDSFFDLGGHSLLATRLISRIRALFQVELPIRAL    52
FrsF_PCP5          -----DVLGGLFAETLGLPQVSIIDSSFFDLGGDSLLATRLISRIRVTRRELSIRTL    52
FrsA_PCP_interpro  QRPRNAQEEMLCGLFAEVLDMKVGGRGDSFFDLGGHSLLATRLIRRIRETLDVELSIRDL    60
FrsD_PCP_PKS/NRPS  -----EMLCGLFAEVLEVGSVGIIDSSFFDLGGHSLLATRLISRIRAALNVELPIRQL    52
FrsG_PCP7          -----EMLCGLFAEVLEVGSVGIIDSSFFDLGGHSLLATRLISRIRAALNVELSIREL    52
FrsE_PCP3          -----RALCQLYAQVLDLSKVGIDGFFDLGGDSISSIQLASRARKAG-WLLTPNQI     51
FrsE_PCP4          -----QLLARLFGEVLGLPQVGADDSFFDLGGDSIISIQLVSRARKAG-WRLTPRDV    51
FrsF_PCP6          -----QLLARLFGEVLGLPQVGADDSFFDLGGDSIISIQLVSRARKAG-WRLTPRDV    51
                    * . * : : . * : . * . * . * * * * * . * : : * * * * * * . :

FrsG_PCP8          FESPTVAGLARYI---- 65
FrsF_PCP5          FECPTVAALARQL---- 65
FrsA_PCP_interpro FEAPCVTELSRHIAEGG 77
FrsD_PCP_PKS/NRPS  FDLPSVAELLEVL---- 65
FrsG_PCP7          FMPTCAELAPRL---- 65
FrsE_PCP3          FRHPRVEDLAAEL---- 64
FrsE_PCP4          FQOPTVAALAQAI---- 64
FrsF_PCP6          FQOPTVAALAQAI---- 64
                    * * * * * :

```

Abbildung A23: Alignment der PCP-Domänen aus dem *cvfrs*-BGC.

```
#  
#  
# Percent Identity Matrix - created by Clustal2.1  
#  
#  
1: FrsG_PCP8          100.00  67.69  63.08  64.62  64.62  39.06  62.50  62.50  
2: FrsF_PCP5          67.69 100.00  63.08  64.62  67.69  45.31  54.69  54.69  
3: FrsA_PCP_interpro 63.08  63.08 100.00  69.23  69.23  40.62  43.75  43.75  
4: FrsD_PCP_PKS/NRPS 64.62  64.62  69.23 100.00  87.69  46.88  46.88  46.88  
5: FrsG_PCP7          64.62  67.69  69.23  87.69 100.00  45.31  48.44  48.44  
6: FrsE_PCP3          39.06  45.31  40.62  46.88  45.31 100.00  59.38  59.38  
7: FrsE_PCP4          62.50  54.69  43.75  46.88  48.44  59.38 100.00 100.00  
8: FrsF_PCP6          62.50  54.69  43.75  46.88  48.44  59.38 100.00 100.00
```

Abbildung A24: Percent Identity Matrix des Alignments der PCP-Domänen aus dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

TE      GDSKSPYQVLMPIRATGGRHPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMS 60
TE2     -----PPLFCIHPPGGCLSWTYVSLVRYLDAEQPIYGLQARGIDGQS 41
          ***** * * .*:*. * . :** ***** *****:* *

TE      ELAPSIPDMAADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILDT 120
TE2     EPASSIEAMAADYVAQIRGIQPHGPYLLGWSLGGNLAQAMASQLESMDQEVGLLFLLDS 101
          * * * * *****: ***.***.***.*****.*** :* *** :.....** :*:

TE      FPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHY 180
TE2     GPSPMHKD----DEMIEYPLFTKEFKNTFKFHVSETKMQAIFEVTKRHVELIRQSTTPVS 157
          * : :: .: * **:: :. : : :.*::::: : : *:: :

TE      GGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLEVDIECTHMDMMQRDVLKIIGPVLES 240
TE2     QGPALLFRATVPYDESTPLPPHAWNEYVKGDIEVHEVHCQHAQMNRIFMEQMGPVIER 217
          * ****: .** *.: : .:*.*:.*:***:.* * :* : :. : :***.*

TE      KLS      243
TE2     KL-      219
          **
    
```

Abbildung A25: Alignment der TE-Domänen aus dem *cvfrs*-BGC.

```

#
#
# Percent Identity Matrix - created by
# Clustal2.1
#
#
          1: TE      100.00  42.01
          2: TE2    42.01  100.00
    
```

Abbildung A26: Percent Identity Matrix des Alignments der TE-Domänen aus dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

frsA_Burk -----atgcagatggatata 15
frsA_Chro atgaaaaacagtgaatcgccaatccatcattttcaggcatcttcagcacagctggatgta 60
                *** ***** **

frsA_Burk tggatggttcagaaaatcgcgccgattatccaattgcatcgccgagtatctccatttt 75
frsA_Chro tggatttctcaggaagtccaccgaatctgccaacaatattgccgagtatctgaatctc 120
***** ** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk cgtggagcactggatgggcccatttttgggtgccttgactgcttgcaagggaggca 135
frsA_Chro gccggctcgttggatgctggattgtttctgcaggctttaagccaggctgccagtgagagc 180
** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk agcgagctgcggtgcaatttgaccacgatggcgtgaggctgatcaagtaccagagat 195
frsA_Chro gcggagctgcaatacaacttccgtcacgatggctccagttgaccaagtttcgctgagat 240
***** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk ttagccgaatgggtgccagatttcattgatgtttcgacagaggcaaacctgaggcaacg 255
frsA_Chro gatgaaggctgggagccggacttcacgatgatcgacgcacggcgagccggaacacgca 300
* * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk gctctgctgatcatcgcatcacaggtcgtaagtacgtcgacatgagaacggagcactg 315
frsA_Chro gccctgcgcgccatcgggagcgggtggagaaaccttcgatctggcggggagcgcgttg 360
** *** ***** * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk tttcgttggctgctgattcgtctcagcgacgagcatcatatattctttcatgcctatcat 375
frsA_Chro tttcgtggaccttgatccgcctggccgacgagcgccacatcttctgccatgtgtatcac 420
***** ** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk cacattgtgatggatggggtgggctatgtgctgctactcgagcgcgtggcggagggtgat 435
frsA_Chro cacatcgcgatggatgggcccggctatgtgatgctgctgcagcgcatagccgaggttac 480
***** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk cgcgcgctgcgctcggatttgccactgccagcatgctgcttgcctaccgctagcggcacc 495
frsA_Chro ggccgcgctcgggaaagccagccggcaccggcctgcggttgcggatgaggatgacatc 540
***** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk gtcgatgatgaggtccgttatcgtgctccgaacaatttgccgtggaccgccccttctgg 555
frsA_Chro gtccgcgaggaagagcgtaccgccagtcggagcagttcgcggtcgaccggcattctgg 600
*** ** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk caagcgcgccgcttgcaagccaagggcgaacccccttggcgttgccggcgagccg 615
frsA_Chro caagcgcgctcggccgagctggcgacggcgagccgccgctgccggcgggcagtgcccg 660
***** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk ttctcgctttcgccgaaagcgcagttattccggaggccggcaggctgagcctgaaagcg 675
frsA_Chro ttctggcgttcgccagacggcgtgattccggaagacgcctgaggcgggctgaggatg 720
***** ** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk gccgcagagcgtctgggtgatcgctttccgggttgtaagcgggccatcatcgcttac 735
frsA_Chro acggccgagcggctggcgtctcccagtcctgctgacagcagccatcgctcgcttat 780
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk ttccgtcgctgggacggccagaacagatgcggtttcggctggcgtgctggcccgcagc 795
frsA_Chro ttccatcgctgggcccagcaagagatctgttccggctggcgtatcggcgcgcagc 840
*** * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsA_Burk gaagtcacatgcaggcaccaggctatatggcacacgcgttgccgttgccagggagcttt 855
frsA_Chro gatgcgacgcacacgcgccggccacctggcgcatgcttggcgtgctggccagcctg 900
** * * * * * * * * * * * * * * * * * * * * * * * * * * *

```

frsA_Burk	acgccccacaccagtcttgccgatatcgtgcgccagctagacggcgaagtgcgctgcatg	915
frsA_Chro	ccgccgcgcccagctctggccgacatcgcgcgacagctggacggcgaggtggagcggatg **** * * ***** ** ** * * * * * * * * * * * * * * * * * * *	960
frsA_Burk	cgtccgcacattcgtctatcggggggaagacatcgttcgggactgggcatcgaccggcgga	975
frsA_Chro	cgccgcatacccgtatcgggctgaggacatcgtgcgcgaccaggccggtgccggtttg ***** * ***** ** * * * * * * * * * * * * * * * * * * *	1020
frsA_Burk	gtgcaggggtcacaggggcccgggtgatcaacatcatgccgttctcttatgcattcgacttt	1035
frsA_Chro	ggcgcggggcgacaggggctgtgatcaacctcatgccttttgctaccgcttcgagttt *	1080
frsA_Burk	ggcgaatgccgagtgacatccgcacatcaattgaccgtcggcttgctgaatgcgctggaa	1095
frsA_Chro	ggcgctgtcgcgtggagtcgcccacagctgaccgtcggcggtggacacgcctggaa **** *	1140
frsA_Burk	gttgcggtccacgatcgaagattggcgatggtctgcatatcgacctgtatgccccgaa	1155
frsA_Chro	gtggcgtgcacgacgcaagaacggtgacggcctccacctgatttgtacgcatcgag ***** ***** ***** *	1200
frsA_Burk	gcatcgggctcgcctgttcagttgcaacagcatgtcagaagactggcgcggtttcattgag	1215
frsA_Chro	cgcgctgcccgcccgaaccgctcggcgcatgccctgcggctggcccgttcacgctc *	1260
frsA_Burk	gtcgcgacggctgtgccacaagtcgcatcgacacactgccatggctggacgagtcggaa	1275
frsA_Chro	gaggcggcggcgagccgtcgcagccggttccgacatcgagctgctggacgagccgga *	1320
frsA_Burk	cggaggcaactgctggaggaaatggagtggaacgcactggatttggggcaaatcacgctg	1335
frsA_Chro	cgccggcaactgctggtcactggaaccgcaccggaccggaccacggccagggcaccttc ** ***** ** *	1380
frsA_Burk	gcggaactgttcgaagttcaggcaaccgctcagccgaatgccgtggcgctggagggccg	1395
frsA_Chro	ccgcaactgttcgaaacccaggggccctcaccgcacgcgctcgcgctggaagccc ** ***** ** *	1440
frsA_Burk	gacgaacgcgtgagctatggcgagttggacgcacgggctaaccgactggcgagtcacctg	1455
frsA_Chro	gagcccggctcagctatgccgaactggacggccgcgcaaccggctggcgcgccatctg **** *	1500
frsA_Burk	cagagtttggcgctggggccagatgtcgtggtcggggtgtgcttggagcgtcgcattgac	1515
frsA_Chro	caaagcctggcgctcggcgccgacgtgctggtcggcatctgcctggagcgtcgcattgac ** *	1560
frsA_Burk	atggtggtggcgatacttgggatcggcaaggcaggggctgcctatttggcctcgcgccg	1575
frsA_Chro	atggtggtcgggtgctggcgcgctgaagtccggcgccgctatctgcccgtgtcggcg ***** *	1620
frsA_Burk	gattaccgacggagcgcctggcctacatgctgacggattcgatggcgccagctacttta	1635
frsA_Chro	gagtaccgacggaaaggctggcctacatgctggcgactcgatggccccgctgctgctg ** ***** *	1680
frsA_Burk	accgagtcaaaacaggtcgaacgactgccgtcgtattggggcatctggtgaagctggat	1695
frsA_Chro	accgactcggcacaagtcgagcggctgccgtcgtattggggccgggtagtcgaactggac ***** ** *	1740
frsA_Burk	cgactcgatttgtcggggcagggcgtcagtgccggcgagcgtacgaccagatcac	1755
frsA_Chro	cggtcgcactggacgctcggcgacagcgcggcgaacggcgctgcccggcggagc ** ***** *	1800

frsA_Burk	ttggcgtatgtgatctatacctcgggctcgacgggccaacccaagggggtggcgggtcagc	1815
frsA_Chro	ctggcctatgtgatctacactcggctccaccggccaaccgaaggcggtggcgggtcagc **** * 1860	1860
frsA_Burk	cacgccggcttagctgggttggtaaaagtccaggaggagcggttcgcgggtggcggggccg	1875
frsA_Chro	cacgccggcctggccggcctggccggcagaccagacagcgggttcgcgctgcaaggccg ***** * 1920	1920
frsA_Burk	gtacgggttctgcagtttgcgctcgtgagttttgacgcagcggtaatggaaatccttatg	1935
frsA_Chro	acgcgggtgctgcaattcgcctcgtgagtttcgacgcggcggatggaaatgctgatg ***** * 1980	1980
frsA_Burk	gcgttctgcagtgccggctgggttggattgcccggcagcggggccgctgctgggtgagcag	1995
frsA_Chro	gccttctgcagcggcggcggcctgggtgctgcccggcggcggggccgctgctggcgaacag * 2040	2040
frsA_Burk	ctagaggagaccctgaatcgttacgcaatcagtcacgctgattgcaccatccgcgctg	2055
frsA_Chro	ctgctggacacgctgaaccgcaatgaaatagccacgcgctgatctcgcctcggcggctg * 2100	2100
frsA_Burk	gagacgggtgaagcagaggtcgtgcctggtctgagcacgctggtggtgggagggcagacc	2115
frsA_Chro	agcacggcagcggcgttggcgcggctcctgaggcggcgtggtggtggcggggaagcc * 2160	2160
frsA_Burk	tgttccggggcaacagcggcgtcgtggctcgcaggggaaggcgcattgtaacgcctacggc	2175
frsA_Chro	tgccgggcgcgacgggtggcggcctggctggcgggacggcggatggtgaacgcctacggc ** * 2220	2220
frsA_Burk	ccgaccgagggcagcgggtgtgtgtagcatgagcaagccgctgtcgggcagcgacaagccg	2235
frsA_Chro	ccgaccgagggcagcggcctgcgtgacgatgagcgagccgctgtcgggcagcggcgcggc ***** * 2280	2280
frsA_Burk	aagctgggtcgaccaacactgggggcaaagctgtatgtgctggatgacacgttgcaaccg	2295
frsA_Chro	aagctgggcgctccgacgcacaacgcggcgtgtacgtgctggatggcgcgctgcaactg ***** * 2340	2340
frsA_Burk	gttccgggtgggggtggcgggggaactgtacatcgcgggcccgggatggcagctgggtat	2355
frsA_Chro	gcgccgggtgggggtggcgggcgagctgtacatcgcggggcccggcgtggcgcgaggctat * 2400	2400
frsA_Burk	taccaacgaccggactgacggcggagcgggttcgtggcgaaccatattggcaaaggcga	2415
frsA_Chro	ctgaaccggcccgggctgacggcggagcggcttcgtggcgaatccgtacggagaggggtgag * 2460	2460
frsA_Burk	cggctgtatcggctcgggagatctggcacactggagcggggaaggagaactggagtatctg	2475
frsA_Chro	cggctgtaccgcagcggcgacctggcgggtggacggaagaaggcagctggaataacctg ***** * 2520	2520
frsA_Burk	ggccgggtggaccaacaactgaaatccgaggttttcgcatcgagccgggcgagatcgag	2535
frsA_Chro	ggcgcagcgaccagcaggtgaaggtgccccgtttccgcatcgagccgggcgagatcgaa ** * 2580	2580
frsA_Burk	acggtgtgtgccagcaccgcaggtgaggaggcgggtggtggtatcggcacaa--at	2592
frsA_Chro	gcgggtgctgaaccggcatccgcaagtgagccagtcgggtggtggtggcggcggcagagccag ***** * 2640	2640
frsA_Burk	ggccgggatcgcgaactggtgggttatgtcacgatacgtggcggaggtggatgggcagggc	2652
frsA_Chro	ggcggcgacagccagttggtggcgtacgtggcggcggcgtcggcggggtggaggggtcggag *** * 2700	2700


```

frsA_Burk      tcgcattaccgaggagatttgctgttgttccgctcaacgataccttatgcgagcagcgc 3612
frsA_Chro      tctcattatggtggcgatttgctgctatctcgctccttgattccatagccgaagacgcg 3660
                ** ***** * ** ***** * ** ***** ** ** ***** ** *****

frsA_Burk      ttgatgccgaaccagaaacctggagcccatatatttcagggatactggaagttcacgaa 3672
frsA_Chro      ctgatgccagaggcggatacatggcagccttattgtctggccaattggaagttcatgac 3720
                ***** * * ** * ** * ** ** * ** * * ** ***** * **

frsA_Burk      gttgaatgcactcatatggagatgatgcagaaggacgtaacgaaagtaattggccagatt 3732
frsA_Chro      atcgagtgcacacatatggacatgatgcaaagagatgttttgaaaataattggtcctgtt 3780
                * ** ***** ***** ***** * ** ** ***** ***** * **

frsA_Burk      ctgaaatcaaagctatcaaaaactgctgaaaagtag-- 3768
frsA_Chro      ctcgagtcaagttgtctgtcaccgctgtcaagcaata 3819
                ** * ** * ** * ** ** ***** * ** *

```

Abbildung A27: Alignment von *frsA* aus dem *bfrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

frsB_Burk      atgagcaatccttttgatgataaagacggaatttttaagtttggatgacgaaaac 60
frsB_Chro      atgagcaatccttttgatgataaagatggtgtgtttctggttctgatcaatgacgagaat 60
                ***** ***** ***** * ** * ** ** ** * ** * ***** **

frsB_Burk      cagcattcgcttggccggattacatcacctaccatctggttggaaacaccgtatttggaa 120
frsB_Chro      cagcattctctatggccggattatattgcggttcagcaggatggagccagagctttggc 120
                ***** * ** ***** ***** * ** * ** * ** * ** * ***** *

frsB_Burk      cccgcatcccgcgatgcatgtctggattacgttgaaaataactggacagatatgcgacc 180
frsB_Chro      cctgcctccaagcaaaccttgcttgattatatagaacaagtggactgatatgaggcct 180
                ** ** * ** * * ** * ** * ** * ** * ** * ** * ***** * **

frsB_Burk      aaaagcctgatcgatgcatgcaatcggatgaacatga 219
frsB_Chro      aaaagcctgattgatgcaatggagtgcgatgataaatga 219
                ***** ***** * ** * ** * ** * ** *

```

Abbildung A28: Alignment von *frsB* aus dem *bfrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

frsC_Burk      atgacaacgcccgttcgagtggtgatcctctggggcagttggtcgcattggtaaatagttg 60
frsC_Chro      atgaaaaatccggtccggtgtagctatcttctgggtgcggctggccgcattgccaaacttg 60
                **** *  **** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      ttgttcaatattgccagtgccagctggtggggagcgcagcctgtcgtgttgagcttg 120
frsC_Chro      ttgttcagtattgcaagtgccagctattgggtgatcagcaaccaatatcgtgagtttg 120
                ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      ctggaggcatcttctcgattaccgatgctaagcggcgtcttgatggagttgatgacggg 180
frsC_Chro      ctggagacgctgcccaggttgtaagtcttaatggcattgcgatggagctgcacgacgga 180
                ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      gcttttccattgctcgcgaatgtcgaatccatgatgatccgtggctagcgtttgacgg 240
frsC_Chro      gtttttccattacttgccggagttagggtgagcgtgatccatggcaggcttttgagggc 240
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      gcgactatgctgttctgatcagttcgccgcccacagctctgagcgaagcatcgaaaacg 300
frsC_Chro      gcgattatgttttctgatcagttcgccgctcgatagcctggccacggcatcaaagact 300
                ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      gcggatgccaggatggaaaatcatggcgctacatttgcggtacacggcagagcgctgaac 360
frsC_Chro      gctgatgagagaatggagcagcatggcaactcttctgctcttcatggcaaggcattgaat 360
                ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      ggcgttgccaatcgcgatgtcaaactactgggtggtcagcaatccagtcagctcaacgca 420
frsC_Chro      gatgtggctagccgcatgtcaagatactggttatcagcaatcccgtgatgatcaatgca 420
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      ctaacagtgaacgcaatgccccgaatctcgatccatcgtgtatatgccccgatgatgccc 480
frsC_Chro      ttgatggttcagcgcacgctccagatttgaattcgtcttctgatttctgcatgtagcgg 480
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      ttggatcacaaccgacgatgattgctggcgcaagggccgggttcatctctcggag 540
frsC_Chro      ctggatcataatcgcgctcatgctttattggcgcataaagctggagccagcctcgctgat 540
                ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      gtcagaaaggtaatgtctggggcaatcacagtcgcaccaataaccctgatttcgatcac 600
frsC_Chro      gtgaggaaagtgattgtttggggtaatcacagctctactcaataaccggatttttatcat 600
                ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      gcgaccattggtggaattccggcctgcaacctgatcagtcacgactggctccgccaggac 660
frsC_Chro      gccactattggcggagtccgctggatgccttgttgagaatgactggctgcatcaagtt 660
                ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      tccgtcgatatcgttagacaacgtgggtatgcagtgatcgatgcctatggcggcttgcca 720
frsC_Chro      tcaatggccttgtcaggcagcaggggtatgcagtgattgatgcttatggtgggttgcca 720
                ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      gccgcacctctgcagcaaaaggcggcgattgatcatatgccccgattgggtcttcggcacg 780
frsC_Chro      gcggcatcctcagcagctaaagcggccatcgatcatatgccccgactggatttttggaacc 780
                ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      cgtgaaggagattggacatccatgggggtatcttctgacggctcctatggtgttcccggc 840
frsC_Chro      agggacggcgattggacctgatgggtgttttgctggatgggtcattatggaatcccttct 840
                * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsC_Burk      ggaatattcttcggttaccctggtctcccatcaagggtgatttacatcgtcaagaat 900
frsC_Chro      ggaatcttcttgggtttccagttgtagccgatggtggacaggtgaatattgtgcaggga 900
                ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
    
```

```

frsC_Burk      ctgaggccgaatccgatcgttctggagaaaattcatctcagtcagatgagatttatcag 960
frsC_Chro      ttgcaaatttgcctgaaagattggaaaaatacatcacagtgccgacgaaatttatcga 960
                **                **** ***** ** * * *****

frsC_Burk      agatcaaaaaatttcaagctgattga      987
frsC_Chro      cgggtgcagcaattcaattgtttaa      987
                * * * * * ***** ** * * *

```

Abbildung A29: Alignment von *frsC* aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

frsD_Burk      atggagatatggttggcgcagcaattgaggaaggatctattcatcaatatagccgagtat 60
frsD-Chro      atggaaatatggctggcgcacagctgatgccggattcgccgaataatatggccgagtat 60
*****
frsD_Burk      ctfcagattgacggggtgttggacatcgactattcatggacgcgttacaccgagtgctg 120
frsD-Chro      ctgcatctttccgggccgttggatccagattgtttttcaaaaccttggcgcaggtcgcg 120
** * * * * *
frsD_Burk      caggaagccatagattacacgtcaatttctccggtcatgttgaccgtcctgtgcaattc 180
frsD-Chro      agcgagacgccggcatgacaggtcaattttccattgaggatggacggccttgcccggtt 180
* * * * *
frsD_Burk      ttgvcacaaagcaaggctgvcaccgagtttcatcgatgatcagvcagctgtatcct 240
frsD-Chro      agccgctccatgaggactggagctctgatttcatcgatgatcagvcacggcgagccg 240
* * * * *
frsD_Burk      ttctttgcagcacaacgggvcgatgagagaacttgcvcacagtccttcgacctggggvcag 300
frsD-Chro      gaacacgcagccctgvcgvcctatgvcgggagvcgggtggagaacccttcgatctggcvcgg 300
*****
frsD_Burk      gacvcactgtttcgttggvgcctgattcgtctcagvcagvcagcatcatatattctttcat 360
frsD-Chro      gacvcggtgtttcgttggaccttgatccgctggcvcagvcagvcacacatcttctgcact 360
*****
frsD_Burk      gcctatcatcacattgtgatggatgtggcgggctatgtgctgctactcagvcgcgtggcg 420
frsD-Chro      gtgatcaccacatcvcgatggatgtcvcgggctatgtgatgctgctcagvcgcgatagcc 420
* * * * *
frsD_Burk      gaggtatatcvcgcatgvcgctcggatttgcactgvcagcatgtcgtttcvcgtaccgct 480
frsD-Chro      gaggtttacggcvcgctgvcggaagvcagvcggcaccgvcctgvcggtttcvcgcatgvcg 480
*****
frsD_Burk      agvcccatcvcgatgatgaggtcvcgttatcgtgvcgvcgaacaatttgcvcgtggaccvc 540
frsD-Chro      gatvcccatcvcgvcgaggaagvcgctaccvcagctcvcgagvcagttcvcgvcgvcagvcg 540
*****
frsD_Burk      gccttctggaagvcgvcgvcggttgaagccaagvcggaacccctttgvcggtgtgvc 600
frsD-Chro      gcattctggaagvcgvcgctcvcgvcagctgvcgvcagvcgvcgvcgvcgvcgvcgvcgvc 600
** * * * * *
frsD_Burk      gvcgagvcggttctcvcgtttcvcggaagvcgvcagttattcvcgagvcgvcgvcgvcgvcgvc 660
frsD-Chro      gatgvcgvcggttctcvcggttvcgvcagvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 660
* * * * *
frsD_Burk      ctgaaagvcgvcgvcagvcgctcvcggtgatcvcgtttcvcggttvttaagvcgvcgvcctc 720
frsD-Chro      ctgvcgatgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 720
*** * * * *
frsD_Burk      atcvcgttacttvcgctcvcgtgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 780
frsD-Chro      gtcvcgttattcvcctcvcgtgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 780
*****
frsD_Burk      gvcvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 840
frsD-Chro      gvcvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 840
** * * * * *
frsD_Burk      gvcgagctttacvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 900
frsD-Chro      gccagvcctgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvcgvc 900
** * * * * *

```


frsD_Burk	cgctgcatgctccgcacatccgctatcgggcggaagacatcgttcgggactgggcatcg	960
frsD-Chro	gagcggatgctccgcataaccgctatcgggctgaggacatcgtcgcgaccaggccggt	960
	* ***** * ***** ** ***** ** * *	
frsD_Burk	accggcggagtgcagggtgcacaggggcccgtgatcaacatcatgccgttctcttatgca	1020
frsD-Chro	gccggtttggggcgcggggcgcaggggacctgatcaacctcatgccttttgcttaccgc	1020
	**** * ** *	
frsD_Burk	ttcgactttggcgaatgccgagtgcacatccgcacatcaattgaccgtcggcgtgctgaat	1080
frsD-Chro	ttcgagtttggcgccgtgcccgtggagtcgcccatcagctgaccgtcggcgtgctggac	1080
	***** ***** *	
frsD_Burk	gcgctggaagtgcgggtccacgatcgcaagattggtgatggtctgcatattgacctgat	1140
frsD-Chro	agcctggaagtggcggtgacgaccgcaagaacggtgacggcctccacctcgattgtac	1140
	***** ***** ***** ***** ***** * * * * * * * * * *	
frsD_Burk	gccccgaagcatgcggctcgcctgttcagttgcaacagcatgtcagaagactggcgcgt	1200
frsD-Chro	gcatccgagcgcggctgcccgaaccgctgcccggcagcgcctgcccggcggcgg	1200
	* *	
frsD_Burk	ttcattgaggtcgcggcggctgtgccacaaagtccgatcgacacactgccatggctggac	1260
frsD-Chro	ttcatcgtcaggcggcggcggagctgcagccgggtgtccgacatcgagctgctggac	1260
	***** *	
frsD_Burk	gagtcggaaccggaggcaactgctggaggaatggagtggcaacgcactggatttggcgaa	1320
frsD-Chro	gaggccgagcgcggcaactgctggtcgactggaaccgcaccggaccggaccagccag	1320
	*** *	
frsD_Burk	atcacgctggcggaactgttcgaggttcaggcaaccgctcagccgaatgccgtggcgtg	1380
frsD-Chro	gccaccttcccgaactgttcgaaaccaggcggcctcaccggcacgcccgtcgcgctg	1380
	*** *	
frsD_Burk	gaggggcccggacgaacgcgtgagctatggcgagttggacgcacggccaaccgactggca	1440
frsD-Chro	gaaagcccggacgcccggctcagctatgccgaactggacgcccgcgaaccggctggcg	1440
	** *	
frsD_Burk	agtcacctgcagagtttggcggtggggccagacgtcgtggtcgtagtgtgcctggagcgc	1500
frsD-Chro	cgccatctgcaaagcctggcgctggcgccgacgtcgtggtcggcatctgcctggagcgc	1500
	* *	
frsD_Burk	tcgattgacatggtggtggcgatacttgggatcgccaaggcaggggctgcctatttgccg	1560
frsD-Chro	tcgatcgacatggtggtcgcggtactggcggtgctgaagtccggcggcctatctgccc	1560
	***** ***** *	
frsD_Burk	ttcgcgccgattaccgaccgagcgcctggcctacatgctgacggattcgatggcggcca	1620
frsD-Chro	ctgtcgccggagtaccgaccggaacggctggcctacatgctggcgactcgatggcccc	1620
	* ***** ***** * * * * * * * * * * * * * * * * * *	
frsD_Burk	gtacttttaaccgagtcaaaacaggtcgaacgactgccgtcgtattggggcatctggtg	1680
frsD-Chro	gtgctgctgaccgactcggcacaagtgcagcggctgccgtcgtattggggccgggtagtc	1680
	** *	
frsD_Burk	aagctggatcgactcgatttgcggggcagggcgtcagtcgcggcgcgagcgtacga	1740
frsD-Chro	gaactggaccggctcgacctggagcctctgcccagcagcgcgggaacggggcgtgccc	1740
	* *	
frsD_Burk	ccagatcacttggcgtatgtgatctatacctcgggctcgacgggccaaccgaggggtg	1800
frsD-Chro	gccgagcacctggcctatgtgatctacacctcggctcaccggccaaccgagggcgtg	1800
	* *	

frsD_Burk	gcggtcagccacgccggccttagctgggttggtaaaaagtcaggaggagcgggttcgcggtg	1860
frsD-Chro	gcggtcagccacgccggcctggccggcctggccggcagccagacagcgggttcgcgctg ***** * ** * ** * * * * * * * * * * * *	1860
frsD_Burk	gccgggcccgttacgggttctgcagtttgcttcgctgagtttgacgcagcggtaatggaa	1920
frsD-Chro	caaggcccgcagcgggtgctgcaatcgccctcgctgagtttcgacgcggcgggtgatggaa *	1920
frsD_Burk	atccttatggcgttctgcagtgccggtcggttggattgcccggcagcggggccgctgctg	1980
frsD-Chro	atgctgatggccttctgcagcggcggcggcggcggctggcctgcccggcggcggggccgctg ** *	1980
frsD_Burk	ggtgagcagctagaggagaccctgaatcgttacgcaatcagtcacgctgatggacca	2040
frsD-Chro	ggcgaacagctgctggacacgctgaaccgccatgaaatcagccacgctgatcctgccg *	2040
frsD_Burk	tccgcgctggagacggtggaagcagaggtcgctgcctggctgagcacgctggtggtggga	2100
frsD-Chro	tcggcgctgagcacggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc *	2100
frsD_Burk	ggcgagacctgtccggggcaacagcggcgtcggtgctgcagggaggcgcatggtcaac	2160
frsD-Chro	ggggaagcctgccggggcgcagcgggtggcgccgtggtcggcgggacggcggatggtgaac *	2160
frsD_Burk	gcctacggcccgcagcgggtgacggtgctgcctgacgatgagcaagccgctgctgggcagc	2220
frsD-Chro	gcctacggctccgaccgagcgacggcctgctgacgatgagcagcggcgtgctccggcgac *	2220
frsD_Burk	gacaagcccgaagctgggtcgaccaactgggggcaagctgtatgctggtatgacagc	2280
frsD-Chro	ggcgcgccgaagctgggccctccagcgcacaacgcggcggctgtacgtgctggatggcggc *	2280
frsD_Burk	ttgcaaccggttccggtgggggtggcgggggaactgtacatcgccgggtcggggattggca	2340
frsD-Chro	ctgcaactggcgccgggtgggggtggcggggagcgtgtacatcgccggggccgggtggcg *	2340
frsD_Burk	cgctgggtattaccaacgaccggattgacggcggagcggttcggtggcgaaccataggc	2400
frsD-Chro	cgcggtatctgaaccggccggggcctgacggcggagcggctcggtggcgaatccgtacgga *	2400
frsD_Burk	aaaggcgaaccggtgtatcggctgggagatcggcagcggcggaggcgggaaggagaactg	2460
frsD-Chro	gagggtgagcggctgtaccgcagcggcgacctggcggcggaggcggagaaggcggagctg *	2460
frsD_Burk	gaatatctgggccgggtggaccaacaactgaaatccgaggttttcgcatcgagccgggc	2520
frsD-Chro	gaatacctggggcgcagcgaccagcaggtgaaggctcgggggttccgtatcgagccgggc *	2520
frsD_Burk	gagatcgagacgggtgtgtgcccagcaccgcagttaagggaggcgggtggtgtatcgcg	2580
frsD-Chro	gagatcgaagcgggtgtgaaccggcatccgcaagtgagccagctcgggtggtggtggcggc *	2580
frsD_Burk	acgaa--tggccgcgatacgaactgggtgggttatgtcacggtgctggtggcggggtggat	2637
frsD-Chro	cagagccagggcggcgacagccagttggtggcgtacgtggcggcggcgtggcggggtggag *	2640
frsD_Burk	gggcaggcgtacgcgctcaggtagcgaactggttggcagagtacatggtaccggcagctg	2697
frsD-Chro	gggtcggagctggcggcctggcggcggggcagctgccggagcagatggtgccggcggc *	2700

frsE_Burk	gtcaatgcctgggttttgccggacggatctgtccggccatcccgatTTTCatgtcctgttg	960
frsE_Chro	gtaaAtgcttggttattgcgagTggatgttccggccggccaactttccggcctTgttg	960
	** ***** **	
frsE_Burk	cgTcgtgtgcgagcaggcacttcaagcttacagTcaccagatctacccttcgagTgg	1020
frsE_Chro	cgccggTgcggaaCaggcgtTgcaagcctacagcctacaggattTgCCgttcgagTgg	1020
	** * ***** **	
frsE_Burk	ctggTcgagcagttgaatccgatccgtTccacttCgtatcatccgctgttccaggTggta	1080
frsE_Chro	ctggTgagcagatcaatccggcgtTccacttCgtatcatcctctatttcaggTgcg	1080
	***** ***** * *****	
frsE_Burk	ttggttttgagaacaatcagcgtgccgatTTTCgattgggCggactggTggTggagcag	1140
frsE_Chro	ctggTgctgcagaacaatcTgagCggacttccactggaaggctTgCCgttcgagTggcag	1140
	** * ***** * * * *	
frsE_Burk	caagTgcttggtacaggtacagcgaagTtcgacctgccttcaatctgTttgagacaatg	1200
frsE_Chro	cgctgctTggaacgggaaccgccaagTtcgatctggctTtcaatTgttcgagcggcag	1200
	* ***** * * * * * * *	
frsE_Burk	agcgagcagggcacccgctgggcgtgagggcgatctcgaatcgcagTgacTgttc	1260
frsE_Chro	gatggcagTggcggcgcagggtTggacggagaactggaatTcgcttcgagcctgTtc	1260
	* * * * * * * * * * * *	
frsE_Burk	gatcctcccagcggTtcgattggcctatcgctTtagccgcttactggaatctggtcc	1320
frsE_Chro	gaccgcccggcggagcggctggTggcgcgtTtTgctgTttgctggaacgatggTcc	1320
	** * * * * * * * * * * *	
frsE_Burk	gcagcaccgtcgagctcgattgccatccttgatttactggagagatcagagcggaaca	1380
frsE_Chro	atgcagcggcgcattTgatTgccgaactggatctattggacgagcagcgccggcag	1380
	** * * * * * * * * * * *	
frsE_Burk	gctttgctggagTggaacgcgacgactcggccactcccgcgtTgacgctggcagaagca	1440
frsE_Chro	cggtgagcggatggaacgatTcagccgTccgctggaggcgggacgctcgctgtTg	1440
	** * * * * * * * * * * *	
frsE_Burk	ttcgaacgcaggcggcgttacaccagaggcggtagcgtggcgtTtgccgatgaagta	1500
frsE_Chro	ttcgaagcccaggcaaggctTcgccgactcggtcgcgctggctTtggctcgagcag	1500
	***** * ***** * * * * * * *	
frsE_Burk	ctgagttacgtgaactaaacaacaggccaatcgctTggcacggatgctggtTgCCgCg	1560
frsE_Chro	tatagctatgCCgagctggaccggcgcgaaccagTtggccaggacgctggcgggcgc	1560
	* * * * * * * * * * * *	
frsE_Burk	ggTctcggcgggaaggcaggTcgcgtTggccgttccgagatcgctggatTggTcgTg	1620
frsE_Chro	ggcatcggcccggaggacatcgtcgcgtggcggtgccgcctcgtggacatgTgTgTg	1620
	** ***** * * * * * * *	
frsE_Burk	gctttgctcggcgtgactaaggcggcggcgtattTgCCgctagatccagatTccg	1680
frsE_Chro	gcgtTgctggggTggtcaaggcggcgtgcctatctgCCgctggaccccagatTccg	1680
	** ***** * * * * * * *	
frsE_Burk	gcagagcggctggcatacatgctTggcggatgcaaagccgacgctgtTgatgactgtcaac	1740
frsE_Chro	cgagagcggTtgcccataTgctgtccgacgccacgccgctTtgcTgctggcgacatcc	1740
	***** * * * * * * * * * * *	
frsE_Burk	gcgCagctcggtagcctgtccgaatgcgCgggataccggTgctggcgtggatgcggat	1800
frsE_Chro	gataccgTcggcggTtTgCCgcatTctcaggctTgcgggttcaggTgctggatgagccg	1800
	* ***** * * * * * * *	

frsE_Burk	tcggttcgggacgcgatttcgcagatgtcaggatgcaacttggttcaatcggaacggctg	1860
frsE_Chro	gcctggcgggagatggttgcccagcggacggcgccgcttggcgcagcagagcggacg	1860
	* * * * *	
frsE_Burk	tgtccattgcagcctcagcaccctgtttgcgtgatctatacctccggttcgacaggccgt	1920
frsE_Chro	cgccccttgttgccgcagcaccgggtttgcgtgatctatacctccgggtcaaccggcaag	1920
	* * * * *	
frsE_Burk	ccgaagggcgtgatggtgaccaccagggtatcgtcagcttacgtccagccaaatcgag	1980
frsE_Chro	cccaagggcgtgctcggtgaccatcagggcattgccagcctgcgcgcagccagatcgag	1980
	* * * * *	
frsE_Burk	cgattcggcgtgtccgctgaatcgagcgtgttgcaattcgcctccttgagcttcggcgcg	2040
frsE_Chro	cgtttcggcgtgttcgcccagcagcgtgttcgagcttcgcgtcagcttcggggc	2040
	* * * * *	
frsE_Burk	gcgctgttcgaaatttcacgtctctgttgacagggcgcgcctgggtgttgatccagc	2100
frsE_Chro	gcgctgttcgaaatttcacgtctctgttgacagggcgcgcctgggtgttgatccagc	2100
	* * * * *	
frsE_Burk	atcaagaagcgtgaacgttgagacaatgaccgcattggtgactcggcaccggttgagt	2160
frsE_Chro	gcccgcgagcgtgaatgccggagcattggcggagctggcgcggcggatggttgagc	2160
	* * * * *	
frsE_Burk	catatggtggtgccccatccgcgttgacacggttggtgcgcgaccgctctgctcgtacg	2220
frsE_Chro	catgtggtgctgccccgctggcgttgaggcattggcttcggagcggctgccgatagc	2220
	* * * * *	
frsE_Burk	gtccgcatcatggtggcagggtgagcattgtccggcccatctgggtgagcgttggtccgcg	2280
frsE_Chro	ctcggcatcatggtggcgggagcattgccggctcatctgcaggaacgctggtcggct	2280
	* * * * *	
frsE_Burk	gatcgattcatggttaacggttacggttcgtctgaagtaaccgatatgcgcgacgatgagt	2340
frsE_Chro	ggccggtgatggttaacggttacggttcgtctgaagtaaccgatatgcgcgacgatcagc	2340
	* * * * *	
frsE_Burk	cagcccctgtccggcgtgcactcccgcgatgggggcgccgaatgccaataccggactc	2400
frsE_Chro	ctgccgctgtccggcgtgcactcccgcgatgggggcgccgaatgccaataccggactc	2400
	* * * * *	
frsE_Burk	tacttgctggatgcaggatgcaaccctcccggcggcgtgatgggtgagctatacgtt	2460
frsE_Chro	tacgtgctggatgcaggatgcaaccctcccggcggcgtgatgggtgagctatacgtt	2460
	* * * * *	
frsE_Burk	gccggggaagggttggtcagggttatctgagccgccgggtcttaccgagcgtttt	2520
frsE_Chro	gcagcggatgggtggcgcggctatctgcgcgccgggctgagcggcagcgtttt	2520
	* * * * *	
frsE_Burk	gtcgccaatccattcgagcagggttgccagatgtaccggacggcgatctcgcacggcgc	2580
frsE_Chro	gtcgccaatccggttcgcccaggggaacggatgtatcgacccggagacctcggagggca	2580
	* * * * *	
frsE_Burk	gacatcgatggccgctcgaatatttgggacgcggtgacctcaggtcaagatccgaggt	2640
frsE_Chro	ggcggggacggcaggtcgaatatttgggacgcggtgacctcaggtcaagatccgaggt	2640
	* * * * *	
frsE_Burk	ttccggatcgaaccggctgagattgaggcggcgttcgccagctaccaggcgtggccaa	2700
frsE_Chro	ttccgatcgaaccgggaaatcagggtgtgttgccgcttgcggcgtggcgcag	2700
	* * * * *	

frsE_Burk	gcgactgtcgtagcttgggaggaggtaccgggggcaagcaactggttggctacgtcgtg	2760
frsE_Chro	gcgctgtggtcgcccgagagatgcgcccggcgtcaggcagctggtgggtatgtggtg *** ** *	2760
frsE_Burk	cccaaggaggcggttatgctggagccgcgccatgcgccgtgagttggccgagcacctg	2820
frsE_Chro	gcggcgacggaagcgaagctggatgcgcaaggtttgcggcgacagctggccgagcacctt *	2820
frsE_Burk	accgatcatatggttcccgcggtgctggctgagctatcggctttgccccgaacgccgaac	2880
frsE_Chro	cccgcacataggtgctgcagcgtgggtggagctggcgcccttgcgagacgccgaac *** *	2880
frsE_Burk	ggcaaacctgaccgcagtgccctgcctgcgccggtattcgtcacagaaggatatacgcgag	2940
frsE_Chro	gccaagtgcaccgcggtgcccctgcccagctggccacacctgcgcgagggctaccgccg ***** *	2940
frsE_Burk	ccacgcacggcaaaagagtgaccctatgtcaactgttcgcgcaagtactcgatttgcct	3000
frsE_Chro	cccgcacggcgaagagcggcgttcttgcagctgtacgccaggtactggattgttc ** *	3000
frsE_Burk	caagtaggtatcgacgacgccttctttgatctggcgggcgacagcatcagttccatccaa	3060
frsE_Chro	aaagtcggtatagacgatggttcttcttgcagctggcgggcgacagcatcagctcgatccag *** *	3060
frsE_Burk	ctggccagccgcgcgcaagcgtggctggcacgtgacgccgaaccagggtttcgcctat	3120
frsE_Chro	cttgcagccgcgcgcaagcgggtggctgctgacgccgaaccagatattccggcat ** *	3120
frsE_Burk	ccgcgctgagcaggatctggctacggtgatgtgcccgttgacaaatggggtggactctgtg	3180
frsE_Chro	ccccggctgaggacctggctgaggacttaagcctgctgagtcgaccggaacgagcgg ** *	3180
frsE_Burk	caagaggagccgtgtggtgacctgacgctgacgccgattctacgctggatggggaaaac	3240
frsE_Chro	gtgatcaggcctgcggcgtttgcccggcgacgccatcatccactggatgctggccgcg *	3240
frsE_Burk	ggcccatatcgtcttttccatcagtcgcaactggtcgtgcccgtctgggttgcggcgc	3300
frsE_Chro	ccgccgatcggcgttcaaccaatctcaactgctgcgacgccgggtgggctcaaacgg *	3300
frsE_Burk	gccgacttgcgtggcaatggtccaggcgtgctcgatcaccacgatcgttgcgcatgcgt	3360
frsE_Chro	gatgacctgctggccgctgcaggcgtgctggaccatcacgatgccttgcgctgcgt *	3360
frsE_Burk	ctacacgaggacgatggcgaagcgcg--catgacgattctaccggtgggtacgacgagg	3417
frsE_Chro	ttggacgcgccggccagcgtgagcgcgtgctatcgattccgcctgccgatcgggtcgg *	3420
frsE_Burk	gccgaggattgctgctgctatcgagatcgtgggagtcgatgccctcgagcgacagggtg	3477
frsE_Chro	gccggcgtatgctgcccggatcgacgcggctcggctcgggatgaggcggagtggcatctc *** *	3480
frsE_Burk	gtgctggcccgtgaaacggacgaggcaatctgcgtctggatagcgaatgtggtcgcactg	3537
frsE_Chro	ctattggcgcgggaaggggaagccgcccgaacggctggactgagggccggcttgcgtg *	3540
frsE_Burk	gtgcaagtgtgtgctagatgccgggttcggaagaaggggtggttacgtcttgcattcac	3597
frsE_Chro	ctgcaggcgtatggctggatgccggggacgaggccggcgtttattgctggtgatccat *** *	3600

frsE_Burk	catctggccgtggacggcgtgtcgtggcgtgtgttgctatcggattggcagcaagcttgg	3657
frsE_Chro	cacctggcggatggatggcgtttcctggcgcgtgctgttgccggatttgacagcaggcctgg ** ***** **	3660
frsE_Burk	gcgacgtgtgcgtggggtgcgctatttcgctcgatccggttggcagtcgttccgcaac	3717
frsE_Chro	caggcggcagccaggaccgtgcatcgcgtggaccggctcggggcctcatctccggcta * * * * *	3720
frsE_Burk	tgggcactgtgtttgcaacgggatgcacagtcctccgagcgcgaagccgagctggcgtac	3777
frsE_Chro	tggctcgtcatctggggcaagagcccggtcgtcgcagcgcgaagcgaactggcgcac *** * ** * * * * *	3780
frsE_Burk	tggctcgtcgatgctgtcgacgaccgatatgccgctcggacgacgggcattcgatccagcc	3837
frsE_Chro	tggaaatggcgcgtggcggcgaagatccgttgcctggcccgcgtacgattccgccc *** * * * * * * * * * *	3840
frsE_Burk	cgcgataccaccgaaaccaaacagagtctcagcctatcactgcccgtccggacgacccaa	3897
frsE_Chro	cgcgagccaccggaccggcaagcctgagcctggctttgccgcggcggcgaccgac ***** ***** ** * * * * * * * * * * * * * * * * *	3900
frsE_Burk	gcgttactgactcaggctgcaacgcgctttcatgcgagcgcgaacgatgtgttggtaacc	3957
frsE_Chro	gcgctgctgacgcagggcagccgctttccacgccatgcaacgatgtgctgttgacg *** *	3960
frsE_Burk	gtgtttgtgctcgccatggcagcttggcggcgccaatgcatgggccatgccccgatgcg	4017
frsE_Chro	gtgtttgccttgcggttgcggtttggcggcgccgctctgccgagcaccggcggag ***** ** *	4020
frsE_Burk	ttgctattcgacctggaagggcacggtcgcgagaccaggacaccgcatcgacctgtcg	4077
frsE_Chro	ttgttgttcgacctggaggccatggacggaggcgggatactggcatcgattgtcc *** *	4080
frsE_Burk	cgtagcgttggttggtttaccagctgttccgggtcgcagtgagactggatgccgtcgat	4137
frsE_Chro	gcaaccgtcggctggttcaccagcctgttccgggtcggctggcgtggatgagccggg ** *	4140
frsE_Burk	ctggacgacgcactaggggagggcgctagcctggcggctgtgctgaagtcggtaaaggag	4197
frsE_Chro	ctggacgaagcgtgaagggcgggacagcctggggcgctgtgctcaagtcggtgaaggaa ***** ** *	4200
frsE_Burk	caacttcatcgttgccggaccggggcctgggcttcggcttgcctccgctacttgaaccag	4257
frsE_Chro	cagttgcccggccatcccggaccggggcatgggttttggcttgcctgcgccatctgaatcc ** *	4260
frsE_Burk	ggtagcgtgcagagcttgcgcacacggacagccacaatcggtttcaactatcttggc	4317
frsE_Chro	ggcgcgcccggcagctggccgcttgagcagtcgcagatcggtttcaattaccttggc ** *	4320
frsE_Burk	cgctttggcgcacatcgaggggggcactggcaactggcctcagacgtgggcatcgaggcg	4377
frsE_Chro	cgtttcaccgccgcggagggggaggactggcagccggggctgcgccgcatcgaggcc ** *	4380
frsE_Burk	gggcaggatccggaatgccctgcccgcacatccgctgtcgtttgacgcacatccttggac	4437
frsE_Chro	ggacaggatccggacatgcccttgcgcacatctcttgcgttcaatgccagacgctggac ** ***** ***** ***** ***** ***** * * * * * * * * * * *	4440
frsE_Burk	cgaacacatggccggaattaaccgcatcttggctcggggcagcgaattgttctcctcg	4497
frsE_Chro	cgggagcaagggccggaactgcatgccatctggctcggggcggagagttgttcgatcag ** *	4500

frsE_Burk	gacgagattgccgaactggcacagttgtggcagcaggccgcaatggctttagctgagcat	4557
frsE_Chro	gagcagatcgcggagctggcccagttgtggcggcaggcggcgtcgcgctggcggagcac	4560
	** **** *	
frsE_Burk	gtgacaaggccggggcaggaggccgcacgcgctcggatctgccctcgttcacttgcac	4617
frsE_Chro	gcctcaaggcggagcggcggcggcgtcgcgctcggatctgccgctgggtgggtttggag	4620
	* *	
frsE_Burk	caggcgcagattgagcaactggagggtggaatatccgcggatcgaagaggcttggcgttg	4677
frsE_Chro	caaacgcaaatcagcggctggaggccagatcgcggcgtggaagatgtgcttccgttg	4680
	** **** *	
frsE_Burk	tctccgctgcagaagggactactgttccacggctctgtacgatccggcgggctgatccc	4737
frsE_Chro	tcgcccgtcaaaaggcttgttatccacggcctgtacgatccggcgggcttgatccc	4740
	** *	
frsE_Burk	tatgtcgcgagtgattgatctacgcgctagaaggcgaactcgcgcggctgcgctcaagca	4797
frsE_Chro	tatcgcgagtgatgacgtatcagctggaaggcgcgctgatccgcggcgtggaagg	4800
	** *	
frsE_Burk	gcggtacacggcctgctgctgcagcacagcaatcgcgcgctggttgcgttgatctgggc	4857
frsE_Chro	gccgcgatggcttgttgcagcggcgcagcaatcgcgcgctgcttgcgatttgggg	4860
	** *	
frsE_Burk	cgggagcagccagtgagggttatcgtgcccgtttcggcactacctggcaggaattgac	4917
frsE_Chro	aaaggccagccggtgcagatcattccagctggaggcgtgccatggcaggacatcgat	4920
	* *	
frsE_Burk	ttgtccatgctgggggaggacgagcagcaggcggacttgaacaaatgcaggaggaagac	4977
frsE_Chro	ttgtcgatgctggctggcgacgagcggcagccatgctggcgcagatcgcgagcaagt	4980
	* *	
frsE_Burk	cgatcatcagcggttcgatctgtctatgcgcgcttggcttgcgttctcgtgctgatccgactc	5037
frsE_Chro	cgcagccagcgtttcgattggggcgcgctccactgctgtccttcaccctgatccggctg	5040
	** *	
frsE_Burk	gctgtcgatcggcaccggctgatcatgagcaaccatcacatcctgctggatggttggctt	5097
frsE_Chro	gcgccgaccggcaccagctgatcatgaacaatcatcacatcctcctggatggttggctc	5100
	** *	
frsE_Burk	gggccactgttatggcgcgagttgatgaaatgtatcgaagcggcggcgtctgcgcgcc	5157
frsE_Chro	gagccctgttatggcgggagttgatgacgctgtaccgcaatggcggcgtctgggcgcg	5160
	* *	
frsE_Burk	ataccgcgctgacccatcgcgattatctggatggctggctaggcgcgatctggag	5217
frsE_Chro	atgccgcgctcagccttaccgggactacctgggtggctcggccggcgcgacctgaa	5220
	** *	
frsE_Burk	cccgaccggatggcctggcgcggatatttgcgcgatttggttaccggacgttgcgtggct	5277
frsE_Chro	gccgatcgcaggcttggcggcatatctcggcgaattggagacgctacgctgctgtgc	5280
	* *	
frsE_Burk	ccagccgcgccacggagtacgtgatccaggagacctatgaacggcgcttgcggcagca	5337
frsE_Chro	ccggagcggcggcggcctatgtcgatcaggaaacttacagcctcgccttgcggcagcg	5340
	** *	
frsE_Burk	ctggcgagcgggctgactgcgcttgcgaacagttgggggtgacgttgaataccgtgata	5397
frsE_Chro	ctggcgaagcgtggcggcccgcggcagagctggggatcactctgaataaccctggtc	5400
	* *	

frsF_Burk	ctccgcttgcgctggcggagcgggagagaccgtggcggagtgttggaaacgcttgaacga	951
frsF_Chro	gtacggctgcgatggctcggcgggcgagacggtagcgggcctggaggcggcgtgcagcga	960
	* ** *	
frsF_Burk	gagcaagcacgcttctggagtaccagtatctggatctggcagaaatccaacgtctggct	1011
frsF_Chro	gagcaggcgggattgtggaccaccagcatctggacctggaggaaatccagcggctggcc	1020
	* *	
frsF_Burk	ggacaacgcagcttctgcacacgctgttcatcttgcgagaactatccgcttcgacgcacag	1071
frsF_Chro	ggccagcgcagttgttcgacacctgtgttcatcttgcgagaactatcccttcgattcgcag	1080
	* *	
frsF_Burk	gcgattgcgcggcgtggggcggaccgtcttaagtagaatcaacgggggagaaacagcat	1131
frsF_Chro	gcatggcggcggattgggaccagcatcgcgcgcgctgagcggcgggagaaacat	1140
	* *	
frsF_Burk	gatagtcactatcccgtcaccttgatggcggcctcgcgaaacactgacgctgtacctt	1191
frsF_Chro	gaaagccactatccgctgacctgtgctgtgcgcgggagatgttgacctatactctg	1200
	* *	
frsF_Burk	tcttatcagagcggcgggttcgaacatgggacgatggaaaacctgctgacacgatttcgc	1251
frsF_Chro	tctacgatgcacagaggttcgacaaggtgtggtgcaaggcctgctgacgcggttcgcg	1260
	* *	
frsF_Burk	acattgctggaagcgggtggacagactcatcgtgcccgatgtgacatcgatttggctg	1311
frsF_Chro	ctgctgctggagggcggcggcggagcctcgcggcgggtgtccgacatcgagctgctg	1320
	* *	
frsF_Burk	acggcggatgaacggcagcagttgctggtcgaatggaatgccacggatcgcccattgccc	1371
frsF_Chro	gacgagccgagcggcggcaggtgtgatcgagtggaacgccaccgagcgtccatcgccg	1380
	* *	
frsF_Burk	aaggtcacgttgccggaatggttcgaggccaggtggaacgaacaccgacagcggatggcc	1431
frsF_Chro	caagccacctgcccggagctgttcgaagccaggcggcgcgcgcgaacggcggcggca	1440
	* *	
frsF_Burk	gtgctctgcgacgaagtgcgctcgattaccggaactcaattgccgagccaaccagatc	1491
frsF_Chro	ttgtctgcgacggcggagacgctggactacgccgaactgaaccggcgcgccaaccaactg	1500
	* *	
frsF_Burk	gcacactggttgatttcacaagggtcggaccggaacagaaggtggcgtggcgttgccg	1551
frsF_Chro	gcgcatggctgatcgggcagggcgtcggcccggagcagcgggtggcgtggccttgccg	1560
	* *	
frsF_Burk	cgccgggtggaattgctggtggcaatgctgggggtgtgaaggcagggcagcctacttg	1611
frsF_Chro	cgccgggcccagctgctggtcgcggtgctggcgtgctgaagtccggcggccgctatctg	1620
	* *	
frsF_Burk	ccattggaccggcctatccggcgaacatctgatgtacatgctggtcgatgccaaacct	1671
frsF_Chro	ccgctggatccggcctatccggcggagcgcctggcgcacatgctggccgacgccaggccg	1680
	* *	
frsF_Burk	gcatgcctgctcggctctggcgcagcattcggcgttgccggacagcaacgtgcccgcc	1731
frsF_Chro	gcctgcctgctggggctggcgcagcgcctcgaagcgtgcccggacagcggcgtggcttgc	1740
	* *	
frsF_Burk	tgggagctggacgacaagcgggtgaagcagggcactggtggcgcagcggcaagagaatccg	1791
frsF_Chro	tggcggctggatgacgcggcgttgcggcaaacgctggcggcgcagccgagagcgaccg	1800
	* *	

frsF_Burk	acgcctcaaccgacgcttacgcgcggagcatccggcttatgtgatctacacatcgggc	1851
frsF_Chro	acgccggcccaaccgctgctggccgaacatcccgcctatgtgatctacacctcggc	1860
	***** * ** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	tcgaccggcaagccgaagggcgtggagatttcgcagcgcagcgaacggacttcgtgagt	1911
frsF_Chro	tcctcggccagccaagggcgtggatctcccagcgcagcggccgacttcgtcgac	1920
	** ***** ** ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	tgggcgcacgagggcgttggcccggacagcttcggatggttagcgcaccacatcactc	1971
frsF_Chro	tggcgcgtggcgtccttcggcgcggaaacctttgccgacgtgctggccactcctcgctc	1980
	***** * * * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	agtttcgacgtatcagatattcgaacttttggcgcgctgctgtgagggtggccgagtgaa	2031
frsF_Chro	agtttcgacgtatcagatattcgaacttttggcgcgctgctgtgagggtggccgagtgaa	2040
	** ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ttgttgcgcgatcttctggtgctggcggaacgctcgatcgaagggggagtttgattagc	2091
frsF_Chro	ctgctgcgcgatctgctggagctggcgagcggccgatcagcggcggcagcctgatcagc	2100
	** ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gctgtgccttcggatattcgcgcaactgctacagcagcggtagctgagggtggacgcgagc	2151
frsF_Chro	gagggtgccttcggatattcgcgcaactgctacagcagcggtagctgagggtggacgcgagc	2160
	** ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	acggtggtgttcgcgggagggcgttggccgagaactggtggaggcagtgctgcgccgc	2211
frsF_Chro	acggtggtgttcgcgggagggcgttggccgagaactggtggaggcagtgctgcgccgc	2220
	** ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	tggccgggatgctgggcagcaaacatatttggtccgacagagaccagtgatgctgac	2271
frsF_Chro	tggccgggatgctgggcagcaaacatatttggtccgacagagaccagtgatgctgac	2280
	***** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ggacttggctgaaggaagtgaaggctcgcgcgcccgatgattggccggccgctgggc	2331
frsF_Chro	gggggctggctggaggcg- - - -ggcagcgcgcgcccgatagggccgcccgctgggc	2334
	** ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	aacacgcccgcctatgtgctggatgaaggcttgcgtccggtgccggtgggtgtgtgggt	2391
frsF_Chro	aataccagacctacgtgctggacgcgggtcgcgtccgcccgggtgggggtggcgggc	2394
	** * * * * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gagctgtacattggtggagcgggactggcgcgggatacttgaaccgcgcccgtgacg	2451
frsF_Chro	gagctgtacatcgcggggccgggctggcgcgggctatctgaaccggcccgggctgacg	2454
	***** * * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gcggagcggcttcgtggcgaaccgctatggcgaaggtgaacgactgtatcggtcggcgat	2511
frsF_Chro	gcggagcggcttcgtggcgaaccgctatggcgaaggtgaacgactgtatcggtcggcgat	2514
	***** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ttggcgcgttgaacgcagaaggtgaactggaatacctggccgggtggaccaacagtta	2571
frsF_Chro	ctggcgcgttggaccggaagaagggcagctggagtacctggccgggcccggaccagcagctg	2574
	***** * ** * * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	aagatccgaggttccgcatgaaccaggtgagatcgagcggcgctgtgtcgacatcca	2631
frsF_Chro	aaggtgcggggttccgcatcgagcgggagatcgagcggcgctgtgtcgacatcca	2634
	** * * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ttggttcacagcagtggtgattgcacgagaagacacgccggacataagcaattggtc	2691
frsF_Chro	tcggtggcgcagggcggggtgattgcgcgggaggacacgccgggcccacaagcagctggtg	2694
	* ** * * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	

frsF_Burk	gaacggcagcagttgctggtcgaatggaatgccacggatcgcccatgcccgaaggtcacg	4551
frsF_Chro	gagcggcggcaggtgttgatcgagtggaacgccaccgagcgtccatcgccgaagccacc	4554
	** **	
frsF_Burk	ttgccggaatggttcgagggcccaggtggaacgaacaccgacagcgatggccgtgctctgc	4611
frsF_Chro	ctgccggagctggttcgaagcccagggcgcgcgcgccaacgcggcggcattggtctctgc	4614
	***** ** ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gacgaagttcgctcgattaccgcgaactcaattgccgagccaaccagatcgcacactgg	4671
frsF_Chro	gacggcagacgctggactacgccgaactgaaccggcgcgccaaccaactggcgcattgg	4674
	**** * ** ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ttgatttcacaaggtgtcggaccggaacagaaggtggcgtggcgttgccgcgccgggtg	4731
frsF_Chro	ctgatcggcagggcgtcggcccagcagcgggtggcgtggccttgccgcgccgggccc	4734
	**** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gaattgctggtggcaatgctgggggtgttgaaggcaggggcagcctacttgccattggac	4791
frsF_Chro	gacgtgctggtcgcgggtgctggcgtgctgaagtcggcgcgcgctatctgcctggtg	4794
	** ***** ** ***** ** ***** * ** * ** ***** ** ** *	
frsF_Burk	ccggcctatccggcggaaacatctgatgtacatgctggtcgatgccaaacctgcatgcctg	4851
frsF_Chro	ccggcctatccggcggagcctggcgcacatgctggcgcagccagggcggcctgctg	4854
	***** ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ctcggctttggcgacagcattgcccgttgccggacagcaactgcccgcctgggagctg	4911
frsF_Chro	ctggggctggggcagggcatcgaagcgtgcccgcagcggcgtggcttgctggcggctg	4914
	** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gacgacgaagcgggtgaagcagggcactggtggcgcagccgcaagagaatccgacgcctcaa	4971
frsF_Chro	gatgacgcggcgttgccgcaaacgctggcggcgcagccgcagagcgaccggcggcggc	4974
	** ** ** ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ccgcgacgcttacgcgcggagcatccggccttatgtgatctacacatccggcctcgaccggc	5031
frsF_Chro	caacccggctgcccggcaaacatcccgcctatgtgatctacacctccggctcctccggc	5034
	* ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	aagccaagggcgtggagatttcgacgcagcgcgaacggacttcgtgagttggcgcac	5091
frsF_Chro	cagccaagggcgtggatctcccagcgcagcgcggcggcacttcgtcgactggcgcgtg	5094
	**** ***** ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gaggcgtttggcccggacagcttcggtgatgtgttagcgaccacatcactcagtttcgac	5151
frsF_Chro	gcgtccttcggcggaaacctttgccgacgtgctggccactacctcgctcagcttcgac	5154
	* * * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gtatcagtattcgaaacttttggcgcgctgctgtgcccggaggagtgatttggctgcgc	5211
frsF_Chro	gtctcgggtttgagctgctgacgccgctgttgagcggcggcgggtgcccgtgctgcgc	5214
	** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	gatcttctggtgctggcgaacgctcgatcgaagggggagtttgattagcgtgtgcct	5271
frsF_Chro	gatctgctggagctggcgcagcggcggatcagcggcggcagcctgatcagcgggtgccc	5274
	***** ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	tcggtattcgcgcaactgctacagcagcggacactcgggttgacgcgagcagcgggtggtg	5331
frsF_Chro	tcgggttcccccagctgttcagcatggggcgggtgctgctggaggcgaacacctgggtg	5334
	***** ** * ** * ** * ** * ** * ** * ** * ** *	
frsF_Burk	ttcgcggggcagggcgttcgccccagaactggtggaggcagtgcgctcggcgtggccggga	5391
frsF_Chro	ttcgcgggagaggcgtgcgccggagctggtggcggcgggtgcggcaacgctggccggag	5394
	***** ** * ** * ** * ** * ** * ** * ** * ** *	

frsF_Burk	tgtcgggtagcaaacatttatggtccgacagagaccacagtgtatcgatcggacttgg	5451
frsF_Chro	tgccgggtggcgaatatctacggaccgaccgaaaccacggatatatcgattgggggctgg ** ***** **	5454
frsF_Burk	ctgaaggaagcgggaaggctcgcgcgcgccgatgattggccgtccgctggacaacacgcgg	5511
frsF_Chro	ctggaggcg-----ggcagcgcgcgccgataggccgcccgctgggcaataccag *** ** *	5508
frsF_Burk	gcctatgtgctggatgaaggcttgcgtccggtgccggtgggtgttgcgggcgagctgtac	5571
frsF_Chro	acctacgtgctggacgcggcctgcgtccggcggctgggggtggcgggcgagctgtac **** ***** *	5568
frsF_Burk	attggtggaactggactggcgcggggataacttgaaccgcgcggcctgacggcggagcgg	5631
frsF_Chro	atcgcggggcgggctggcgcggctacgtacgtacgtacgtacgtacgtacgtacgtacgtac ** *	5628
frsF_Burk	ttcgtggcaaacccgtatggcgaaggtgaacgactgtatcggcggcgatttggcgcgt	5691
frsF_Chro	ttcgtggcgaatccgtacggagaggttgtagggtgtaccgcagcggcagctggcgcgg ***** ** *	5688
frsF_Burk	tggaacgcggaaggcgaactggagtacctgggcgggtggaccaacagttgaagatccga	5751
frsF_Chro	tggacggaagaaggcagctggagtacctgggcgggaccagcagctgaaggtgccc ***** *	5748
frsF_Burk	ggcttctgcattgaaccagggtgagatcgaggcggcgttgtgtcgacatccattggttca	5811
frsF_Chro	ggttccgtatcgagccgggcgagatcgaggcggcgttgcggcagcagctggcggcgg ** *	5808
frsF_Burk	caggcagtggtgattgcacgggaagacacgccggacataagcaattggtcgggtatgtg	5871
frsF_Chro	caggcggcgtgatgcgcgggagacacgccgggcccacaagcagctggtgggctacgtg ***** *	5868
frsF_Burk	gtgctggatgatgcccgggcattgcagcgcgacgcggaagacgagacgcgacaggttcag	5931
frsF_Chro	gtgctggacgacgcggcggcgtgcagcgcgacgaaagacagcgaatcgggcaggtggaa ***** ** *	5928
frsF_Burk	gcgtggcggcaggtgtacgacacgctgtacgaggcgcactgtcagcagccgttcggggag	5991
frsF_Chro	gcgtggcagcaggtctacgacaccctgtacgacgccaccagcaccagcccttcggcgag ***** *	5988
frsF_Burk	aacttcggcggatgggacagtagttacgacggcagccattgccgctggaagagatgctg	6051
frsF_Chro	aacttcggcggatgggacagcagctacgacggcagccgctgccgctggcgcagatgccc ***** *	6048
frsF_Burk	gcctggcggcaggcaacgggtggagcggattcgcgagttgcagccgcggcgggtactggaa	6111
frsF_Chro	gaatggcgcagcggcaccgctggagcggatccgcgaactgcgtccgcggcggctgctggaa * ***** ** *	6108
frsF_Burk	attggggtggaagcggcttgcgtgctggcgcggcggcggcggcggcggcggcggcggcgg	6171
frsF_Chro	atcggcgtcggcagcggcctgctgctggcgcggcggcggcggcggcggcggcggcggcgg ** *	6168
frsF_Burk	ggcaccgatctgtcaccgacgacgatcgagtgctggagcggcaactgtcagagcagacc	6231
frsF_Chro	ggcaccgacctgtcggcggcagcagatcgcggtactggagaacaactggaaccagctcc ***** ** *	6228
frsF_Burk	ttcggagaccgggtgaggctgttcgcgctaggggcgcatgaactgttgaagctgccagaa	6291
frsF_Chro	tggcgcgacaaggtgcgcctgttcgcgctgggcgcgcacgagctggcggcggctgccggcg *	6288

frsF_Burk	gggaacttcgatgtgatcgtgatcaattcgggttgcagtaacttcccgaacgcagcgtac	6351
frsF_Chro	atgcgtttcgactgcatcgtgatcaactcgggtgctgcagtaacttcccgaacgccgcttat	6348
	* ***** ***** ***** ***** ***** ***** ***** *****	
frsF_Burk	ctgacggaagtgattaagcagtcgctggagcgggtgtcaccgggaggcggcactgtatttg	6411
frsF_Chro	ctggcggaagtgatcgagcaggcgcctggcggcgtggaagcgggaggcggcgtgtacctg	6408
	*** ***** ***** ***** ***** * * ***** * ***** **	
frsF_Burk	ggggacgtgcgcaatttggcgtgctggatgtgttgcctacggcagtggtattgtgtcag	6471
frsF_Chro	ggcgacgtgcgcaatcctggagctgctgccagcttcgcggcggcgggtggagcgtgccgcag	6468
	** ***** * ***** * * * ***** * * *****	
frsF_Burk	gtgggcactgagacggatgcggcgagcttgcggcgtcagtgaggcagcgtttgtgtggca	6531
frsF_Chro	tcggagcggaggtcgagcggcgcctgcagcggcgggtgagcagcggcgtgtggcg	6528
	* * * * * * * * * * * * * * * * * * *	
frsF_Burk	gaaaaggaactgctggtgtcgccggcgttttccagccgggtgcgtgagacgttgcacag	6591
frsF_Chro	gagaaggaactgctgctgctggcggcagcttcttcagccggcgtgcgggaacagcgtgccgcag	6588
	** ***** * ***** * ***** * ***** * * ***** * * *****	
frsF_Burk	attggtgcagtggtatctgcctgaagcagggcagggcgtgaacgaattgaatcgttac	6651
frsF_Chro	atcgccgctggacatccgctgaagcgggagggcggcgaacgagcttaaccgctac	6648
	** * * * * * * * * * * * * * * * * * *	
frsF_Burk	cgctacgagcgggtactgtacaaggaaccgtgtgaagcgggtgctactgtcgaaggtggaa	6711
frsF_Chro	cgctacgaagtgtgtgctgcgaaggccctgccagggcgcctgctggcgtcggcggcg	6708
	***** * * * * * * * * * * * * * * * * * *	
frsF_Burk	acgcaaactggctcggacgtggggaatatgtcggcctgccggacaaggctgatgcagtat	6771
frsF_Chro	gcggagcctggctcgtcgtggcagcctgtcggcctgccggagcggctgtcggcggc	6768
	** * ***** ***** * ***** ***** ***** ***** ***** ***** * *****	
frsF_Burk	gcgtcgatgtggcggatagaaggaattccgaacgcgcgggtgcatggcaggtattggta	6831
frsF_Chro	ggcgacgcctgcgggtgacgggggtgccgaacgcgcctgctgcatggcgaagcggcggcg	6828
	* * * * * * * * * * * * * * * * * * *	
frsF_Burk	acgcggcggctgaagtcaggagggggttggcgtcgtggcagcgttggtggacgagtcg	6891
frsF_Chro	gcgcgggagctgaaggcggcgggttcgcctcggcgtcgtggcggcgtggcggcgtggacgagaac	6888
	***** ***** * * * * * * * * * * * * * * * * *	
frsF_Burk	ggcggggctcagccggaggatcgtgatcgtatggggggcggagctgggttatcgggtgggtg	6951
frsF_Chro	ggcggggctcagccggaggatcgtgatcgtatggggggcggagctgggttatcgggtgggtg	6948
	***** ***** * * * * * * * * * * * * * * * * *	
frsF_Burk	ttgacatggtcaaaccagccggatcatttcgatgcgggtgttctgcggagcagcagcgg	7011
frsF_Chro	gcgacgtggtcggcaggcggggcatttcgacgcgggtgtcgtgcggggagaggacggc	7008
	** ***** * * * * * * * * * * * * * * * * *	
frsF_Burk	ggtgcactggacgcgggtgtataggacgccaccgggttgaaccgttatcggggtatgtg	7071
frsF_Chro	gaagcgtggacgggggtgtaccagccggcggggcgttgcagccgctgtcgggctatgtg	7068
	* * ***** ***** * * * * * * * * * * * * * * *	
frsF_Burk	aacaaccggcgaatttcgaacagtttcagcaatcgcgggtatgtgggagaacaactg	7131
frsF_Chro	aacaaccggcgaacttcgagcagtagcggcgatccgtcgtacgcgttggagcagttg	7128
	***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** *****	
frsF_Burk	ccggattacatggtgccggcggcgtggtactgctggaggggtgacctgacaccaaac	7191
frsF_Chro	ccggagtacatggtgccggcggcgtggtactgctggaggggtgacctgacaccaaac	7188
	***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** ***** *****	

frsG_Burk	cacttggcgtatgtgatctatacctcgggttcgacgggccaaccaagggtggcggtc	1854
frsG_Chro	cacctggcctatgtgatctacacctcggctccaccggccaaccgaaggcgtggcggtc *** **	1860
frsG_Burk	agccacgcccggcttagctgggttggtaaaaagtcaggaggagcggttcgcggtggccggg	1914
frsG_Chro	agccacgcccggcctggccggcctggccggcagccagacagagcggttcgcgctgcaaggc ***** * ** *	1920
frsG_Burk	ccggtacgggttctgcagtttgctcgtgagtttgacgcagcggtaatggaaatcctt	1974
frsG_Chro	ccgacggggtgctgcaattcgcctcgtgagtttcgacgcggcggatgagaaatgctg *** **	1980
frsG_Burk	atggcgttctgcagtgccggctgggttggtatggccggcagcggggccgctgctgggtgag	2034
frsG_Chro	atggccttctgcagcggccggcggctgggtgctgcccggcggcggggcggcggcggcggaa ***** ** *	2040
frsG_Burk	cagctagagaagaccctgaatcggttacgcaatcagtcacgctgattgcaccatccgcg	2094
frsG_Chro	cagctgctggacacgctgaaccgcatgaaattagccacgcgctgactcgcgctggcg ***** *	2100
frsG_Burk	ctggagacggtggaagcagaggtcgtgcctggtctgagcacgctggtgggtgggagcggag	2154
frsG_Chro	ctgagaccgaggacggcggcgttggccggctcctgcggacgctggtgggtggcggggaa *** ** *	2160
frsG_Burk	acctgttccggggcaacagcggcgtcgtggtcgcagggaaaggcgcattggtcaacgcctac	2214
frsG_Chro	gcctgccggggcgcgacggtggcggcctggtcggcgggacggcggatggtgaacgcctac ***** *	2220
frsG_Burk	ggcccgaccgaggtgacggtgtgctgacgatgagcaagccgctgtcgggcagcgacaag	2274
frsG_Chro	ggtccgaccgagggcagggcctgctgacgatgagcagccgctgtcgggcagcgcgcg ** ***** ** *	2280
frsG_Burk	ccgaagctgggtcgaccaactgggggcaagctgtatgtgctggatagcacggttcaa	2334
frsG_Chro	ccgaagctgggcccgcgacgcaaacggcggctgtacgtgctggatggcgcgctgcaa ***** ** *	2340
frsG_Burk	ccggttccggtgggtgtggcgggggaactgtacatcgccggccgagattggcacgtggg	2394
frsG_Chro	ctggcggcgggtgggggtggcgggcgagctgtacatcgccggggccgggctggcgcggc * ** ***** ** *	2400
frsG_Burk	tattaccaacgaccggattgacggcggagcggttcgtggcgaaccatattggcaaaggc	2454
frsG_Chro	tatctgaaccggccgggctgacggcggagcgttcgtggcgaatccgtacggagaggt *** ** *	2460
frsG_Burk	gaacggctgtatcggtcgggagatctggcacgctggagcggggaaggagaactggaatat	2514
frsG_Chro	gagcggctgtaccgcagcggcgaacctggcgcggtggacggaagaaggcagctggaatac ** ***** ** *	2520
frsG_Burk	ctgggcccgggtggaccaacaactgaaatccgaggttttcgcatcagccgggagagatc	2574
frsG_Chro	ctggggcgcagcgaccagcaggtgaaggtgcggggtttccgtatcagccgggagagatc ***** ** *	2580
frsG_Burk	gagacgggtgtgtgcccagcaccggcaggtaagagaggcgggtggtggtatcgcgcacgaat	2634
frsG_Chro	gaagcgggtgctgaaccggcatccgcaagtgagccagtcggtggtggtggcggcagagc ** ***** ** *	2640
frsG_Burk	--ggccgcgatacgaactggtgggttatgtcacggtgctggcgggaggtggatgggag	2691
frsG_Chro	cagggcggcagcagccagttggtggcgtacgtggcggcggcggcgggggtggaggggtcg ***** ** *	2700

frsG_Burk	gcgctacgccgtcaggtagcgaactggttgccagagtacatggtaccggcagtggtgctg	2751
frsG_Chro	gagctgCGGCcctggCGGCggggcagctgCGGgacacatggtgCGGCggCGgtggtg	2760
	* ** *	
frsG_Burk	gtgcttgaggaattgccCGgtgCCgaacgggaagctgaatcatcaggcattgcccCG	2811
frsG_Chro	gtgctggaatcgctgCGcagttgCCgaacgggaagctggaccgcaagtcgctgCGgCG	2820
	***** *	
frsG_Burk	ccggaatacacgggcaagcgtaccaacgtccgCGcaatgctcaggaggaaatattgtgt	2871
frsG_Chro	ccggagtttggCGgctcgattatcagCGcCGcaacCGcaggaggaaatgctgtgc	2880
	***** *	
frsG_Burk	ggactgtttCGggaagtgctggaagtggaacagtggtgggatagatgacagtttttcgat	2931
frsG_Chro	gtgctgttcCGggaagtgcTggaagtgggagcgtCGggatagacgagtttcttcgat	2940
	* *	
frsG_Burk	ttggCGggtcattcgctgctggaacacggttgatcagCCgaattcCGccacaatgaac	2991
frsG_Chro	ctggCGggcactcgTtgcattcagCGcCGgtgatcagCGcatcCGgCGccttgaat	3000
	***** *	
frsG_Burk	atcgaaccctctattcgcatctgttcgacatgccgaccagtgccgagctcCGctccCG	3051
frsG_Chro	gtggagctgtcgatcCGgagctgttcgacatgccgacctgCGcCGaactgCGcCGgaga	3060
	* *	
frsG_Burk	ttgagcaggaagccgacggcagtgCGcaacctttgttgccacaagagCGccaaaacga	3111
frsG_Chro	ctgagCGgacCCgCGcggTgCGcCGcCGttgCagCGcagCGcCGgCGcCGgaagCGg	3120
	***** *	
frsG_Burk	ctgCGgttGctctatgcacaacagCGgttGgtttctgCaccggttcgaaggTccgagT	3171
frsG_Chro	ctgCcttGtcctatgCGcagcagCGgtGgttCctgtatcGttcgaaggCGcagC	3180
	***** *	
frsG_Burk	tccacctacaacatcccactCGccttcaaactgcaagggtcgttgacgtcgaagcGctg	3231
frsG_Chro	tccacctataacatccgctggccttgaggctgaaaggcGatctgCaccCGgaggcttG	3240
	***** *	
frsG_Burk	cgacgagCGctggaagacgtcgtggttcggcatgaaagcctgCGcagcatttctgTgga	3291
frsG_Chro	cagcagCGgtgCGcagacattgtggagCGcagatgaaagcctgCGcagcgttctgTcag	3300
	* *	
frsG_Burk	agCGaaggaatacccgagcagaaaatcttgctccagacgaggcgtgcttgccattacag	3351
frsG_Chro	aacgatggcgtgCCagagcagCGgatccagTccgCGgaacagCGcgtgCGctccttgCG	3360
	* *	
frsG_Burk	cttatcgataccaacgaggaaccggttagatatggcgttgcgtgaagcgtccgattac	3411
frsG_Chro	cgctcgaggtgCGcagcgaagCGgaactgCagCGgcatTgCGcgaagCGcCGgagcag	3420
	* *	
frsG_Burk	cacttcgatttGtcCGggagatccCGctGcgttGcagcgtgttccGcaagagTccag	3471
frsG_Chro	cgtttcgatttGtcGagcGagacCGcgtGcgttGcagcgtgttccGcctgCGcagcaa	3480
	* *	
frsG_Burk	gtgtggacactcttGctGctgatccaccatatcGcaggggatggCGgatccttGcttccg	3531
frsG_Chro	gagTgggtattGctGtTgtgatccaccatatcGcCGcagTggCGgttCcttGctgCG	3540
	* *	
frsG_Burk	ctgggCGcGacttgCCactGcctatGcCGcCGcagaaagggatTggatcCGgctTgg	3591
frsG_Chro	ctggCGcGGagctggacCGcctacGCGcCGcgtGcCaaggCGgCGcCGgactgg	3600
	***** *	

frsG_Burk	gcaccgctcccggtagcagtagccgactacaccttatggcaacgccagttgctcggcagt	3651
frsG_Chro	cagccgctggcgggtgcaatacgccgactacaccttggcagcgccagttgctgggagag ***** **	3660
frsG_Burk	gaaggcgatcctgacagcctgatatcgactcaatttgcgtattggaagcaaaatctggct	3711
frsG_Chro	gaaaacgatccggagagcttgatcgccgcccagttcgcttattggaagcgacgctggcg *** **	3720
frsG_Burk	ggcgcacaggaacaactgacgctgccgacggatcatccgagggccggcgtggcaagtta	3771
frsG_Chro	ggcgcgcccagcaattgacgctggccagccgctcctcgcccggcggtagcagctac ***** * **	3780
frsG_Burk	cgcgccactatctgccgttccagctggaggctgaactgcatagcgacttgcgccgtctc	3831
frsG_Chro	cgcgccgctatctgccatctgagctggacgcccgttgcattgccgaactgacgctg ***** **	3840
frsG_Burk	gcgcgcacgaacgacaccacgctgtcgatgctgttgaagcggccttggctgcgttgttc	3891
frsG_Chro	gcccgcgacgacgacccaccttgccatgctgctgcaggcgggctggcggcgtgtt ** **	3900
frsG_Burk	accggctcgggtgtgggaacgacattccgctaggctgcccatagcaggccgacccgac	3951
frsG_Chro	accggctcggcgcagggaggacatcccgtggctgcccatagcaggccgacccgac ***** **	3960
frsG_Burk	gaggcgcttggtgatctggcggatttttctgtaacacctggatattgcccggacact	4011
frsG_Chro	gatcgctcaacgacatggtcggcttctctgtaacacctggatattgcccggacact ** **	4020
frsG_Burk	tcgggggaccggacttcgtcacgctgctaggccgctgcccagcaagcgttggcagcc	4071
frsG_Chro	tcgggcaatccggatttcacacgctgctggccgggtgcccagcaggcctggcggc ***** **	4080
frsG_Burk	tacgccatcaagatgcgccatttgagcgttggttgagctcatcaaccggtagcatca	4131
frsG_Chro	tatgccatcaggatgcgccgttcgaacgcttggctgagttgatcaatccatcgcgctcc ** *****	4140
frsG_Burk	agtgccaccatccattgttcaggctcaatctgactttgcagaacaatgcgttggcagag	4191
frsG_Chro	agcggccaccaccttgttccaggctcaatctgactttgcagaacaatgcgttggcagag ** *****	4200
frsG_Burk	ttccgcctagatggcatgcaggtagcttgcattgtgtccgatattgccaagttc	4251
frsG_Chro	ttcaggctggatggcgtggaggttccctgtcgctatcgaggccgataaccgccaagttc *** * **	4260
frsG_Burk	gacctgttcttcaatctgtacgaaacctttgaccaagatggccgcccactgggattaca	4311
frsG_Chro	gacctgttcttcaatctgtacgaaacctttgacctggatggccagggcgggtgggctgcaa ***** **	4320
frsG_Burk	ggcgcggctgcaatagcgtgctgattgtacgagcgcgaacgggtacaacggtttatctac	4371
frsG_Chro	ggcgggtgctgagtagccaccgatctgtacgaacaggcgacgggtggagcaattcatgccc ** * **	4380
frsG_Burk	cacttcggttcggttattgcgtgaagtggtggcagcaccgacagcgcaatttctgcactg	4431
frsG_Chro	cacttcagcgcctttgcgccaggtggcggctgtgccgacggcggcggatcgcgccattg ***** **	4440
frsG_Burk	gatttctgtagcaggcagtcgcctcttcgcccggctgcccgtggtatttatcgccagcatc	4491
frsG_Chro	gatctgtccgagccgaaagcggccgaagcg-----gattgcccgcgcccggccg *** **	4491

frsG_Burk	acgcaactagtaaacatgtatggatcaccgaaacgaccgtacatgtgagctatcagccg	5451
frsG_Chro	ccgaagctggcgaacatgtacggcatcaccgaaaccacggtgcacgccagtaccaggcg ** * ** * ***** ** ***** ** * ** * ** * ** * ** * ** *	5448
frsG_Burk	ctcgacagcggaaatgtgtggttcccgtgcaatagtctgattggcattggatcccggat	5511
frsG_Chro	ctggaccgccgcttgtgccaagaggcggaacagcctgacggcgaggcgataccggac ** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	5508
frsG_Burk	ctgcatctgtttctgctggatgctgtctgcaacctgtgcccggtgggtggaattggcgaa	5571
frsG_Chro	ttgcggctgcacttgcctggaccgttggctgacggcgggtgccggcgggcgctgagcgag *** ** * ***** ** ***** ** ***** * ** * ** * ** * ** *	5568
frsG_Burk	ctgtacgtcggggggcggtctggcgcgcggttatcacaatcgggccagcctcacagcg	5631
frsG_Chro	ctgtacatcggcggcgccgcttgcgcgcggttatcgaaccggcggcggcgccgagcgcc ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	5628
frsG_Burk	gagcgattcgtagcgaatcccttcgcttcagcaaacggatgtaccggaccggcgatctg	5691
frsG_Chro	gagcgtttcatcgccaatcggcttgcgcggcggcgaacggatgtaccgagcggcgatctg ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	5688
frsG_Burk	gcacggcgcaacgtcaatgacatcttcaataccatggccgcgccgatcaacaggtcaag	5751
frsG_Chro	gcgcggcgaatgccgccccggcctggaataccaggccgagccgaccaggtcaag ** ***** * * * * ***** ***** ***** ** ***** ** *	5748
frsG_Burk	gtgcgcggttccgcattgagctaggcgaatcgaacgggtttgctgagcatccgggc	5811
frsG_Chro	gtcggggcttccgcattgagcggcggcgagatcgagaccgctgctgctggcattccggg ***** ** ***** ** * ** * ** * ** * ** * ** * ** * ** * ** *	5808
frsG_Burk	gtagaggacgcgagagttgtcgtgacagcgtgcacgacaacgactgtcgcttggttgct	5871
frsG_Chro	gtcgaagatgcccgctgctcgtcaaggcagatggatggcaacgatcagcgctgatcgcc ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	5868
frsG_Burk	tacctcagccgctcgaacgcactgcggcaccggttgcgtcgatggcttaaggtgccccg	5931
frsG_Chro	tatctgttgcctcggagcgcgccgcgccctgagacgctggctgcccccaaatcc ** * ** ***** ** * ** * ** * ** * ** * ** * ** * ** * ** *	5928
frsG_Burk	gatgcggaagcacaaggtgcgcggcattcaggttccgaacggcatgccgatctttcat	5991
frsG_Chro	agcgtcggcgcgagaacgcgggagatcgagctgcccaacggcctgctgtttccac * * ** * ** ***** ***** ***** ***** ***** * ** * ** *	5988
frsG_Burk	cacaacacgatggaaccgagttcctttacgaggaaatcttcgaggacctagtctatctc	6051
frsG_Chro	cataacgggctgaaaccgaattcctttacgaggagatctttgaggatcggatctatctc ** ** * ** ***** ***** ***** ***** ***** * ** * ** *	6048
frsG_Burk	aagcacggaatcagactggacgacggtgcctgctgcttttcgatgtcgggtgcaacattggc	6111
frsG_Chro	aaacacggcattcgtctggataacgatgcctgctgcttcgacgtcggcgccaacatcggc ** ***** * * ***** ** * ** * ** * ** * ** * ** * ** * ** *	6108
frsG_Burk	ctgtttatgctggttgcgtgggtcaacattgtcgcaatgccaccatcttcgcttcgagccg	6171
frsG_Chro	ttgttcacctgtttgtcggccagcattgcggcaatgccacggtgttcgcttcgaacc *** * ***** ** * ** * ** * ** * ** * ** * ** * ** * ** *	6168
frsG_Burk	attccaccggttttcgacgctgacgctcaatgccgaggtgcacggagacaaggtcagg	6231
frsG_Chro	attccccggttttcgacgctgctcgtgaacgcggcgttcatggaggaaggtccgc ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	6228
frsG_Burk	ttgttcgattgcccgttatccgatgcggtgagacaggagcgttttacttttaccgcaat	6291
frsG_Chro	ctgttcgattgcccgtgtcggacgcggccaggcaggaaccttcaccttctatccaac ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	6288

frsG_Burk	gggtggaacagtagttacgacggtcagccattgccgctggaagagatgctgcctggcgc	7251
frsG_Chro	ggctgggacagcagctacgacggccagccgctgccgctggcgcagatgcccgaatggcgc ** ** *	7242
frsG_Burk	caggcgacggtggagcggattcgcgagttgcagccgcgccgggtgctagagattggggtg	7311
frsG_Chro	agccaccctggagcggatccgcgaactgctccgcccggctgctggaatcggcgtc *	7302
frsG_Burk	gggagcggcttgctgctggcgcgctggcaccggagtgcgaagcctactggggcacggat	7371
frsG_Chro	ggcagcggcctgctgctggcgcgctggcggagcctgcgaagcctactggggcacggac *	7362
frsG_Burk	ctgtcaccaacgacgatcgcagtgctggagcggcaactgtcagagcagaccttcggagac	7431
frsG_Chro	ctgtcgcggcgcagatcgcggctactggagaacaactggaaccagctctgccgcgac *	7422
frsG_Burk	cgggtgaggtgttcgcgctaggggcacatgaactgtcgaagctgccagaaggaacttc	7491
frsG_Chro	aagtgccctgttcgcgctggcgcgacgagctggcgcggctgcgcggatgctttc *	7482
frsG_Burk	gatgtgatcgtgatcaattcgggtgtgcagttccccaacgcagctacctgacggaa	7551
frsG_Chro	gactgatcgtgatcaactcgggtgtgcagttccccaacgcggcttatctggcggaa *	7542
frsG_Burk	gtgattaagcagtcgctggagcgggtgtcaccgggcccggcactgtatttggggacgtg	7611
frsG_Chro	gtgatcagcagcgcctggcgcggctggaagcgggcccgcgctgtacctgggcagcgtg *	7602
frsG_Burk	cgcaatttggcgcgctgctggatgtgtttgctaccggcagtggtttgtgtcaggtggcact	7671
frsG_Chro	cgcaatctggagctgctgccagcttcggcggcgggtggagctgcccagctggagcgg *	7662
frsG_Burk	gagacggatgcccgcagcttggcgcgctgagtgagcagcgttgcctggcagaaaaagaa	7731
frsG_Chro	gagtcgacgcggcggcgtgcagcggcgggtgagccagcggctgctggcggagaagaa *	7722
frsG_Burk	ctgctggatcggcggcgttttcagccggtgtgcgtgagacgttgccacagattggtgca	7791
frsG_Chro	ctgctgctggcggcggacttcttcagccgctgcccggacagctgcccagatcggcgcg *	7782
frsG_Burk	gtggatattcgcctgaagcggggcagggcggatgaacgaattgagtcgttacgctacgag	7851
frsG_Chro	gtggacatccgcctgaagcggggcagggcggcgaacgagcttaaccgctaccgctacgaa *	7842
frsG_Burk	gcggtactgtacaaggagcctgtgaagcgggtgtcactgtcggaggtggaacgcagacg	7911
frsG_Chro	gtgggtgttcgcaagggcccgtgccagggcgcgtcgcctggcgtcggcggcggcggagccg *	7902
frsG_Burk	tggtcggagctggggaatatgtcggcctgcccggacaaggctgatgcagatgcccgatg	7971
frsG_Chro	tggtcgtcgtgggcagcctgtcggcctgcccggagcggctgctggctcggcggcagcgcg *	7962
frsG_Burk	tggcggatagaaggaaattccaacgcacggttgcatggcgaggatggtaacgcggcgg	8031
frsG_Chro	ctgcgggtgacgggggtgccaaacgcgctgctgcatggcgaagcggcggcggcggag *	8022
frsG_Burk	ctgaagtccaggatgggttggcgtcgtggcgacgtggcctggacgagtcggcggggtc	8091
frsG_Chro	ctgaaggcggggcgttcggcgtcggcgtcgtggcggcggcggcggcggcggcgggta *	8082

frsG_Burk	gagccggaagatctgtatcgattgggtgaggagctgggttatcgggtggtgttgacatgg	8151
frsG_Chro	cggccggaagcgtcgccgtctgggcgagctgggctggcggatgctggcgacgtgg ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8142
frsG_Burk	tcaaaccagccggatcatttcgatgcggtgtttctgccagagcagcagcggggtgactg	8211
frsG_Chro	tcgcggcaggcgggcatctcgacgcggtgttcgtgcgggagaggacggcgaagcgtg ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8202
frsG_Burk	gacgcggtgtataggacgccaccggtgttgaaccgttatcggggtatgtgaacaaccg	8271
frsG_Chro	gacgggtgtaccagcggcggggcgttcagccgctgtcgggctatgtgaacaaccg **** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8262
frsG_Burk	gcgaatttcgaacagttcgcagcaatacgcgcatatgtgggagaacaactgccggattac	8331
frsG_Chro	gcgaacttcgagcagtagcggcatccgtcgcacgcgttgaggacgttcgggagtagc ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8322
frsG_Burk	atggtgccggcagcgtggtactgctggaagggttgcactgacaccaaaccgggaagctg	8391
frsG_Chro	atggtgccggcggcgtcgtgttcgtggacgcgttgcgctgacgcgaacgggaagctg ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8382
frsG_Burk	gaccgccgcgctaccggcgccggagtttggcacggcatgctatcgagcgggggtca	8451
frsG_Chro	gaccgagggcgctgccggcgccggatctcggggagggctaccggcgccggaatcg ***** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8442
frsG_Burk	gagcaggagaaggtgctggcgcaactgtttgccgaagtgtgggttgccgaggtgggg	8511
frsG_Chro	gaacgggagcaattgctggcgcggtgttcggggaggtgctgggtcggcgaggtgggg ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8502
frsG_Burk	gtgaatgacagcttcttcgaccttggcggtcactccttgcggcaaccgctctgattagc	8571
frsG_Chro	ttgcacgacagcttcttcgatctggcgggcactccttgcggcaaccggtgatcagc ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8562
frsG_Burk	cggttcgctcgctgttccatctcatttgcgatccggaccttgcgagtcgccacc	8631
frsG_Chro	cgcattccgcgctgttccaggtcagctgccgatccgcgctgttcgaatcgccgagc ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8622
frsG_Burk	gtgccagactggctcgctacattgacgacggcgcaatcgacctcgattcgttcgaggtg	8691
frsG_Chro	gtggccgggtggcgcgctacatcgatgacggcgagatcgatgacgattccttcaatgtg ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8682
frsG_Burk	ctattaccgatcaaaccggagcggcaagcgtgctccgctgttctgcattcatccaggcggg	8751
frsG_Chro	ctgctgccgatcaaaccgacggcaagcggcggcggcgtgttctgcattcaccgggcggg ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8742
frsG_Burk	tgctgagttggacctacgttggattggtgctgtatctggatgccgagcaaccgatctac	8811
frsG_Chro	tgctgagctggacctatgtcagcctggtgctgctacctggatgaggagcagccatctat ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8802
frsG_Burk	ggcctgcaagcccaggcatcgatggcaagagtcaggtggcacttcgatcaggggcatg	8871
frsG_Chro	ggcctgcaagcggaggcatcgacggccagagcggcggcaagctcgattgaagccatg ***** ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8862
frsG_Burk	gcaagcgattacgtcgagcagatacgggattcagccgaagggcggctactacctgctt	8931
frsG_Chro	gccgcccattacgtggcccagatccgcccattcagccgacggcccttattacctgctg ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8922
frsG_Burk	ggctggtccttaggggcaacattgccaggagatggcctactcttggaaagccagaat	8991
frsG_Chro	ggctggtccttaggggcaacattgccaggagatggcctactcttggaaagccagaat ***** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** * ** *	8982

```

frsG_Burk      catgatgtgggctgctgattctgctcgattccggaccgtctcctatgcacaagaatgac 9051
frsG_Chro      caggagggtgggcttgctgttccctgctcgattccgggcccgtcgccgatgcacaaggacgat 9042
                ** ** ***** ***** * ***** ***** ** ***** * **

frsG_Burk      gagatgatcgagtatccgctgtttaccaaggaatttaggaaaaccttcaagttccatgtc 9111
frsG_Chro      gaaatgatcgagtaccgctgttcaccaaggagttaagaacaccttcaagttccatgtc 9102
                ** ***** ***** ***** ** * ** ***** *****

frsG_Burk      agcgaatcgaagatgcgggccatatttgaggtgagtaagcaccacattgaactgatcaag 9171
frsG_Chro      agcgaacgaagatgcaggcattttcgaggtgaccaaaccatgtcgaactgatccgg 9162
                ***** ***** ***** ** ***** ** * ** * ***** *

frsG_Burk      cagtctacgactcccgttagccggggcggtgccctgttgttccgggccacggtgccttac 9231
frsG_Chro      caatcgaccacgcccgtcagccaggcccagcgtgttgttccgggccacggtgccttac 9222
                ** ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsG_Burk      gacgaaaatacgccacttctgtcgccccacgcttggaaatgactatgtgaaaggggaaata 9291
frsG_Chro      gacgaatctacgccgtgctgccccgcacgtggaatgagtatgtgaaagggcagatc 9282
                ***** ***** ** * * * * * * * * * * * * * * * * * * * * *

frsG_Burk      gaaatacgcgacattcattgccaacacgcgcacatgaaccacgtcgaattcatggcgcag 9351
frsG_Chro      gaagtgcagaggtccattgccagcatgcgcaaatgaaccgaatcgaattcatggagcag 9342
                *** * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

frsG_Burk      atgggggaagtgatagagacgaaactggctgagctccatgcccgagccactcgatattaa 9411
frsG_Chro      atggggcgggtgatcgaacgcaagctggccacgctccacgaccagtccaccgccgtaat 9402
                ***** ***** * * * * * * * * * * * * * * * * * * * * *

frsG_Burk      ----- 9411
frsG_Chro      tcttga 9408

```

Abbildung A33: Alignment von *frsG* aus dem *befrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

frsH_Burk      atgatgcaagagaatgacgaagtttatctgcgtagccataccaagatcgagccgcttatac 60
frsH_Chro      --atgaccgtatccgataacgtattcctgcgcagccacaccaagatcgagccggtgatc 57
                *** * ** * ** * ***** ***** ***** * **

frsH_Burk      atgcgctggtatgcgtggtcacatctgatttctccggcccagcatgcatgaatatggcg 120
frsH_Chro      atgcgctggtatgcctgggcgacctggtatcgccagcccagcatgccttgaatatcgcc 117
                ***** * ** * ** * ** * ** ***** ***** * **

frsH_Burk      tttcgccactgccaatggtgaagtcggttctcgcgcgcccacggtgcacgaagcggt 180
frsH_Chro      tttcgccactgccaatgctcaagcttctcgcgcctcgcggcggtgcatgaagccg 177
                ***** * ** * ** * ** * ** ***** ***** *

frsH_Burk      tccaataatcggagatgctaggtggccttatctagaattgaagaaaagcgatcgcc 240
frsH_Chro      tccagcaatcggaaatgctcggcgggccttttctggaactgaagaaaagcgatcgcc 237
                **** ***** * ** * ** * ** * ** ***** ***** * **

frsH_Burk      gcggtaaaggcgctatggcaacagactcagcagcggcgaaaagatgatccagttggc 300
frsH_Chro      gcggtgaaggcgctatggcagcagaccagcagcggcgaggcaggcagatcgcttcg 297
                ***** ***** * ** * ** * ** * ** ***** * **

frsH_Burk      gaagcactgctgaaactggatcgccgtctgcaaaagagcgaagcgggctttagccttgat 360
frsH_Chro      gaggcgctgctggagctggatcgaggtcgcagcagagcagaccggcctcagcctcgac 357
                ** * ** ***** * ***** * ** * ** * ** * ** ***** **

frsH_Burk      catgtatatgccgagttgccgagcgcggttgaagggtggtcgaagtcagctacgacct 420
frsH_Chro      catactacgccgagctgccggagcgcgctgcagggcctggtggaagtcagctatgact 417
                *** * ** ***** * ** * ** * ** * ** ***** ***** * **

frsH_Burk      cacaaccatccgtcactgcggctgatcgaagaactcctttatctggaggactggatgaac 480
frsH_Chro      cataaccaccgctctcgcggctgatcgaggagctgctttatctggaggactgggtgg 477
                ** ***** ***** ***** * ** ***** ***** * **

frsH_Burk      gatgatgggaagagatcgctttcaatttaagcaaaaggaggaggagcgcgcttcttcac 540
frsH_Chro      ggcgctggacaggaatcgcttcagcttgacaaggaggaggagcgcgcttcttcac 537
                * * ** * ** * ** * ** * ** * ** ***** ***** *****

frsH_Burk      aacacaccgcgctcgatatgccggggcggtggtggctccactgccattcgtgacaag 600
frsH_Chro      aatcagccgagctcagcgcgcccgggagcagctggtggtgcccgtcgcgcttgcgag 597
                ** * ** ***** ***** ***** ***** * ** * ** * ** * **

frsH_Burk      cgattcgacttgctggccaccagtcggcttagccccgtctctcctcgggagctggccgat 660
frsH_Chro      cgcttcgatttgcttccggcagccgctgtcctcgggtgccttcagccagctggccgat 657
                ** ***** ***** * ** * ** * ** * ** * ** * ** * ** *****

frsH_Burk      gcgctggaatcccagcaaccxaaaggccggccttccgcatcacttaccacgacacct 720
frsH_Chro      gcgctggagattccggagatcagcgcggccttccgcaatatttaccacctcgcg 717
                ***** * ** * ** * ** * ** ***** ***** ***** * **

frsH_Burk      ccgagcgcgaataagccgacgtatcgggagacggcgtgagggtccgctactttggccat 780
frsH_Chro      ccgagcgcgaatgagcccagctacgaaggcagtgagggtgagctccgttatctcgccat 777
                ***** ***** * ** * ** * ** * ** ***** *****

frsH_Burk      gcttgcgtgctggtacagctcgcggaggttccgctactggtcagatccgttctgaactgg 840
frsH_Chro      gcttgcgtgctggttcagaccgagaggtatcggttctggtcagatcccttctgacctgg 837
                ***** ***** * ** * ** * ** * ** ***** ***** *****

frsH_Burk      gaccacaatacagaagagaagcgaactgactttttacgacctgcccggaccgaatcgattac 900
frsH_Chro      gaccatcagccggagcaggagcgttgaccttctacgatctgcccggaccatcagctac 897
                ***** * ** * ** * ** * ** ***** ***** ***** *****

```

frsH_Burk	gtgttcacactcacaacccatccggatcatttttagctgcaagcgctgttgaactgcgt	960
frsH_Chro	gttttcctgaccacaaccaccaggatcatttcagctgcaaggctttgctgcaattgcga	957
	* *	
frsH_Burk	aaccgaatcggccatattctggtgcccggtaataatggcaacaatttcgccgatccatcc	1020
frsH_Chro	ggccgatcggccacatcctggtcccgcgcaacaatggcaacaatttcgcgacccttcg	1017
	* *	
frsH_Burk	atgaaattgaccctgaaacggctcgggttcgacaacgtgatggtcatggacgaaatggcg	1080
frsH_Chro	atgaagctgacgctgaagcggctgggttttgacaatgtgatcgtgatggacgagatggcc	1077
	* *	
frsH_Burk	agcatcacctgcccagatggccgactggtaagcttgccgtcctatggtgagcattctgat	1140
frsH_Chro	gacatcacgctgcccggaccgcatggtgagcctgcccgtcctatggcgagcattccgac	1137
	* *	
frsH_Burk	ctgagcatcaccagcaagcagggcctgttcctttcgtgaaaggacgtactttcatgttt	1200
frsH_Chro	ctgagcatcaccagcaagcagggcctgttcctttcgtgaaaggacgtactttcatgttc	1197
	* *	
frsH_Burk	ctcggcattccgacgcaagaccgctgtgttgaccgtcgcacgttcgccaagtgggc	1260
frsH_Chro	ctcggcattccgacgcaagaccgctgtgttgaccgtcgcacgttcgccaagtgggc	1257
	* *	
frsH_Burk	agagtcgacaacctgttcattggcatggaatgtgacggggcaccgtaagtggctttat	1320
frsH_Chro	aagtggaacaacctgttcattggcatggaatgtgacggggcaccgtaagtggctttat	1317
	* *	
frsH_Burk	ggaccctatttgagcaaccctatcggctcgcaaggaagatgagtcgaggcgcttgcgagg	1380
frsH_Chro	ggaccctatttgagcaaccctatcggctcgcaaggaagatgagtcgaggcgcttgcgagg	1377
	* *	
frsH_Burk	tcgattgtgaacggccttgaggatgtgcaagagtggtggtgtcccagcgcgctgggt	1440
frsH_Chro	tcgattgtgagcgcgctggcgcacgtcagaggatgaggcctgcagccagggcgtgggt	1437
	* *	
frsH_Burk	tatgccatggggcaggagtcctggttccgctttgtcgtcggcttgagtagacacgccggac	1500
frsH_Chro	tacgcaatggggcaggaaatcctggttccgctttcgtggtgggcttgagtagacacccggac	1497
	* *	
frsH_Burk	aaaaagcagatcgtcgaatccgacgtgttcgtcgcgatcgttgccggcaagccggtctggat	1560
frsH_Chro	aagaagcagatcgtcgaatccgacaagtgtgacacgttgccgcaagccggcatggcg	1557
	* *	
frsH_Burk	ggagaacgtctacacggctgctgcaccgtgcttctatga	1599
frsH_Chro	gcgacgcttgcattggctgccaacccatgctgctgtag	1596
	* *	

Abbildung A34: Alignment von *frsH* aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: frsA_Burk 100.00 70.46
2: frsA_Chro 70.46 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: frsB_Burk 100.00 72.15
2: frsB_Chro 72.15 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: frsC_Burk 100.00 67.88
2: frsC_Chro 67.88 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: frsD_Burk 100.00 69.69
2: frsD-Chro 69.69 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: frsE_Burk 100.00 69.93
2: frsE_Chro 69.93 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: frsF_Burk 100.00 72.69
2: frsF_Chro 72.69 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
#       1: frsG_Burk   100.00   71.52
#       2: frsG_Chro   71.52   100.00
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
#       1: frsH_Burk   100.00   77.26
#       2: frsH_Chro   77.26   100.00
```

Abbildung A35: Percent Identity Matrices der Alignment von *frsA-frsH* aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

FrsA_Burk      -----MQMDIWMVQKIAADYPNCIAEYLHFRGALDGPFLGALHCVAREA 45
FrsA_Chro      MKNSESPIHHFQASSAQLDVVISQEVSPNLPNNIAEYLNLAGSLDAGLFLQALSQVASES 60
                *:*:*: *::: : ** *****: *:* . ** ** * * *:

FrsA_Burk      SELRCNLHHDGVRLIKYHRDLAEWVPDFIDVSTEANPEATALSIMRSQVVKVSDMRD DAL 105
FrsA_Chro      AELQYNFRHDGLQLTKFRRDDEGWEPDFIDVSTHGEPEHAALRAMRERVEKPFDLARDAL 120
                :*: *::*:*: * :*: * * *****..*: * * * : * * .*: ***

FrsA_Burk      FRWCLIRLSDEHHIFFHAYHHIVMDGVGYVLLLERVAEVYRALRSDLPACRFATASAI 165
FrsA_Chro      FRWTLIRLADERHIFCHVYHHIAMDGAGYVMLLQRIAEVYGALREGQPAPACGFADADAI 180
                *** ***:*:*: * * * * * * * * * * * * * * * * * * * * * * * *

FrsA_Burk      VDDEVYRASEQFAVDRAFWQARAALQAKAEPPLPLSGEPFLAFAESAVIPEAGRLSLKA 225
FrsA_Chro      VREEERYRQSEQFAVDRAFWQARS AELATAEPPLPAADGPF LAFAQTAVIPEDACGGLRM 240
                * : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsA_Burk      AAERLGVSLRLLSAAIIAYFRRWDGQNMFRFLAVSARSEVTMQAPGYMAHALPLQASF 285
FrsA_Chro      TAERLGVSQSRLTAAIVAYFHRWGQQEILFRLAVSARS DATRHAPGHLAHALPLLASL 300
                :***** ***:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:

FrsA_Burk      TPHTSLADIVRQLDGEVRCMRPHIRYRAEDIVRDWASTGGVQGAQGPVINIMPFSYAFDF 345
FrsA_Chro      PPRASLADIARQLDGEVERMRPHTRYRAEDIVRDQAGAGLGRGAQGPVINLMPFAYRFEF 360
                * : * * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsA_Burk      GECRVSAHQLTVGLLNALEVAVHDKRIGDGLHIDL YAPQACGSPVQLQQHVRRLARFIE 405
FrsA_Chro      GACRVSAHQLTVGVLDTLEVAVHDKRNGDGLHDL YASERGCPPPELRRHALRLARFIV 420
                * * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsA_Burk      VATAVPQSPIDTLPWLDESERRQLLEEWSGNALDLGEITLAE LFEVQATRPQNAVALEGP 465
FrsA_Chro      EAAAEPSQPVS DIELLDEAERRQLLDWNRTGPDHGQATFPQLFETQAALTPHAVALESP 480
                * : * * . * : . : * * * : * * * * : * . . * * : * : * * * : * : * * * * *

FrsA_Burk      DERVSYGELDARANRLASHLQSLGVGPDVVGVCLERSIDMVVA I LGIAKAGAAYLPLAP 525
FrsA_Chro      DARLSYAELDARANRLARHLQSLGVGADVLV GICLERSIDMVVAVLGALKSGAAYLPLSP 540
                * * . * * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsA_Burk      DYPTERLAYMLTDSMAPVLLTESKQVERLPSYWGH LVKLDRLDLSGQASSAPARALRPDH 585
FrsA_Chro      EYPTERLAYMLGDSMAPVLLTDSAQVERLPSY WGRVVELDRDLDALPDSAPERALARAEH 600
                :***** ***:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:

FrsA_Burk      LAYVIYTSGSTGQPKGVAVSHAGLAGLVKSQEERFAVAGPVRVLQFASLSFDAAVMEILM 645
FrsA_Chro      LAYVIYTSGSTGQPKGVAVSHAGLAGLAGS QTERFALQGPTRVLQFASLSFDAAVMEMLM 660
                ***** ***:*:*:*:*:*:*:*:*:*: * * * * * * * * * * * * * *

FrsA_Burk      AFCSGGRVLVPAAGPLLGEQLEETLNRYAISHALIAPSALETVEAEVVPGLSTLVGGGET 705
FrsA_Chro      AFCSGGRVLVPAAGPLLGEQLD LTNRHEISHALISPSALSTADAALAPVLR TLVVGGEA 720
                ***** ***:*:*:*:*:*:*:*:*:*: * * * * * * * * * * * * *

FrsA_Burk      CSGATAASWSQGRRMVNAYGPTTEATCVTMSKPLSGSDKPKLGRPTLGAKLYVLDSTLQP 765
FrsA_Chro      CPGATVAAWSAGRRMVNAYGPTTEATACVTM SEPLSGDGAPKLG RPTHNARLYVLDGALQL 780
                * * * * * : * * * * * * * * * * * * * * * * * * * * * * * *

FrsA_Burk      VPVGVAGELYIAGRGLARGYYQRPLTAERFVANPYGKERLYRSGDLAHWSGEGELEYL 825
FrsA_Chro      APVGVAGELYIAGRGLARGYLNRPGLTAERF VANPYGGERLYRSGDLARWTEEGELEYL 840
                .***** ***:*:*:*:*:*:*:*:*:*: * * * * * * * * * * * * *

FrsA_Burk      GRVDQQLKIRGFRIEPGEIETVLCQHPQVREAVVSR - TNGRDTQLVGYVTIRGEVDGQA 884
FrsA_Chro      GRSDQQVKVRGFRIEPGEIEAVLN RHPQVSQSVVVARQSQGGDSQLVAVVAVGGVEGSE 900
                ** * * * : * * * * * * * * * * * * * * * * * * * * * * * *

```

FrsA_Burk	LRRQVANWLP EYMPAVVLVLEELPRLPNGKLDHQALPAPEYTGKRYQRPRNAQEELCG	944
FrsA_Chro	LRRLAAGQLPEHMPAAVVLESPLQLPNGKLDKSLPAPEFGGSHYQRPRNAQEEMLCG	960
	*** . * . ***:***:*.***:***:*****:.;*****: *.:*****:***	
FrsA_Burk	LFAEVLEVGSVGIDDSFFDLGGHSLLATRLISRIRATMNIELSIRDLFECSSVASLSRHI	1004
FrsA_Chro	LFAEVLDMEKVGRGDSFFDLGGHSLLATRLIRRIRETLDVELSIRDLFEAPCVTELSRHI	1020
	*****:.; * * .*****:***** * * * *:.;*****: *.:*****	
FrsA_Burk	VIGNQSWNPYEVLMPIRASGNRHPLFCIHPEGGLGWSYIGLALHLDHLQPIYTLQARGLD	1064
FrsA_Chro	AEGGDSKSPYQVLMPIRATGGRHPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLD	1080
	. *.:* *.:*****:*.*****:*****:*****:*****:*****	
FrsA_Burk	GVSKLASSIRDMAADYIQQIRTVQPDGPYHLLGWSLGGVAQEMAVQLERADEKVALLAI	1124
FrsA_Chro	GMSELAPSPIDMAADYIEQIRSIQPNQPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAI	1140
	*.:** * * *****:***:.;***:*****:.;***:***** *.:*****	
FrsA_Burk	LDTFPIEILHDAMFGKQACAYDIFAQVVQDMHAMQIDEDRLKSMYQIGLNHMKITATFSS	1184
FrsA_Chro	LDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYQIGLNHMKITAAFSS	1200
	*****:*****:***:***: * * * * ***** *****:***	
FrsA_Burk	SHYRGDLLLFRSTIPYADDALMPQETWSPYISGILEVHEVECTHMEMMQKDVTKVIGQI	1244
FrsA_Chro	SHYGGDLLLFRSLIPYAEDALMPEADTWQPYLGGQLEVHDIECTHMDMMQRDVLKIIGPV	1260
	*** ***** * *.:*****: *.:**:* * * *.:*****:***: * *.:*	
FrsA_Burk	LKSKLSKTAEK* - 1255	
FrsA_Chro	LESKLSVTAVKQ* 1272	
	*.:*** * * *	

Abbildung A36: Alignment von FrsA aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

FrsB_Burk	MSNPFDDKDGIFKVLVNDENQHSLWPDYITVPSGWNTVFGPASRDACLDYENNWTDMRP	60
FrsB_Chro	MSNPFDDKDGVFLVLINDENQHSLWPDYIAPAGWSQSFQASKQTCLDYIETSWTDMRP	60
	*****:.* *.:*****:*****:***:***. *****:.;*****:*.*****	
FrsB_Burk	KSLIDAMQSD* 72	
FrsB_Chro	KSLIDAMECDK* 72	
	*****:.*.:*	

Abbildung A37: Alignment von FrsB aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

FrsD_Burk      MEIWLAAQLRKDLFINIAEYLQIDGVLIDIALFMDALHRVLQEAIVLHVNFSGHVDRPVQF 60
FrsD-Chro      MEIWLAAQLMPDSPNNIAEYLHLSGPLDPDLFFKTLRQVAETPALQVNFSDIEDGRPCPV 60
                ***** *      *****:.* ** *:.*:.*:.* :.*:.*:***** .** .

FrsD_Burk      LRTSKGCAPSFIDVSAQSDPFFAAQRAMRELAHSPFDLGQDALFRWCLIRLSDEHHIFFH 120
FrsD-Chro      SRVHEDWSPDFIDVSTHGEPEHAALRAMRERVEKPFDLARDALFRWTLIRLADERHIFCH 120
                * . :. :.*:*****:.* * ** ***** .. *****:***** *****:***:*** *

FrsD_Burk      AYHHIVMDVAGYVLLLERVAVYRALRSDLPLPACRFATASAIVDDEVYRASEQFAVDR 180
FrsD-Chro      VYHHIAMDVAGYVMLLQRIAEVYGALREGQPAPACGFADADAIVREEERYRQSEQFAVDR 180
                .***.*****:.*:.*:***** **.. * ** * * * .*** : * ** * *****

FrsD_Burk      AFWQARAALQAKAEPPLPLSGEPFLAFAESAUIPEAGRLSLKAAAERLGVSLRLLSAAI 240
FrsD-Chro      AFWQARSAELATAEPLPAADGPFLLAFAQTAVIPEDACGGLRMTAERLGVSQSRLLTAAI 240
                *****.* * .***** :. *****:.* ***** . :.* :***** *****:***

FrsD_Burk      IAYFRRWGQGNEMRFLAVSARSEVTMQAPGNMAHALPLQASFTPRTSLADIVRQLDGEV 300
FrsD-Chro      VAYFHRWGGQQEILFRLAVSARSDATRHAPGHLAHLPLLASLPPRASLADIARQLDGEV 300
                :***:***:*.*:.*: *****:.* * :***:.*:***** *.* :***:***.*****

FrsD_Burk      RCMRPHIRYRAEDIVRDWASTGGVQGAQGPVINIMPFSYAFDFGECRVTSAHQLTVGVLN 360
FrsD-Chro      ERMRPHTRYRAEDIVRDQAGAGLGRGAQGPVINLMPFAYRFEFGACRVESAHQLTVGVLD 360
                . *** ***** * :.* :*****:***:.* * :*** *****:

FrsD_Burk      ALEVAVHDKIGDGLHIDL YAPQACGSPVQLQHVRRLARFIEVAAAVPQSPIDTLPWLD 420
FrsD-Chro      TLEVAVHDKRNGDGLHLDL YASERGPPEPLRRHALRLARFIVEAAAEPSPVSDIELLD 420
                :***** *****:*** : * * :.* ***** ** * . :. : **

FrsD_Burk      ESERRQLLEEWSGNALDLGEITLAELEFEVQATRQPNVALEGPDERVSYGELDARANRLA 480
FrsD-Chro      EAERQLLVVDWNRTPGDHGQATFPQLFETQAALTPHAVALESPDARLSYAELDARANRLA 480
                *:***** :.* .. * * :.* :.*:***:.* :.*:***** ** * :.*:*****

FrsD_Burk      SHLQSLGVGPDVVVVVCLERSIDMVVAILGIAKAGAAYLPFAPDYPTERLAYMLTDSMAP 540
FrsD-Chro      RHLQSLGVGADVLVGICLERSIDMVVAVLVGLKSGAAYLPLSPEYPTERLAYMLGDSMAP 540
                ***** **.* :*****:***:.* * :*****:.*:***** *****

FrsD_Burk      VLLTESKQVERLPSYWGHLVKLDRLDLGQASSAPARALRPDHLAYVIYTSGSTGQPKGV 600
FrsD-Chro      VLLTDSAQVERLPSYWRVVELDRLDLDALPDSAPERALARAEHLAYVIYTSGSTGQPKGV 600
                ***** * *****:.*:***** .. ** * * * :*****:*****

FrsD_Burk      AVSHAGLAGLVKSQEERFAVAGPVRVLQFASLSFDAAVMEILMAFCSGGRLVLPAAGPLL 660
FrsD-Chro      AVSHAGLAGLAGSQTERTFALQGPTVRLQFASLSFDAAVMEILMAFCSGGRLVLPAAGPLL 660
                ***** . ** ***** . ** *****:*****:*****:*****

FrsD_Burk      GEQLEETLNRYAISHALIAPSALETVEAEVVPGLSTLVVGGETCSGATAASWSQGRMVN 720
FrsD-Chro      GEQLDNLNRHEISHALISPSALSTADAALAPVRLTVVGGEACPGATVAAWSAGRRMVN 720
                **** :***: *****:***:.* :.* :. * * *****:.* ** * :.* *****

FrsD_Burk      AYGPTTEVTVCVTMSKPLSGSDKPKLGRPTLGAKLYVLDSTLQPVVGVAGELYIAGRGLA 780
FrsD-Chro      AYGPTTEATACVTMSEPLSGDGAPKLGPRPTHNARLYVLDGALQLAPVGVAGELYIAGAGLA 780
                ***** * *****:***** .. ***** . :*****:.* ***** *****

FrsD_Burk      RGYQRPLTAERFVANPYGKGERLYRSGDLARWSGEGELEYLGRVDQQLKIRGFRIEPG 840
FrsD-Chro      RGYLNRPLTAERFVANPYGEGERYRSGDLARWTEEGELEYLGRSDQVQVVRGFRIEPG 840
                *** :*****:*****:*****:*****:***** ***** :.*:*****

FrsD_Burk      EIETVLCQHPQLREAVVSR-TNGRDTQLVGYYTVRGEVDGQALRRQVANWLPPEYMPVAV 899
FrsD-Chro      EIEAVLNRHPQVSQSVVVARQSQGDSQLVAYVAAVGGVEGSELRRLAAGQLPEHMPVAA 900
                ***:*** :.*: :.*:***:.* :.* * :***:.*:.* * :.*. *** .*. *****:*****
    
```



```

FrsD_Burk      VLVLEELPRLPNGKLDHQALPAPEYTGKRYQRPRNTQEEILCGLFAEVLEVNNVGIDDSF 959
FrsD-Chro      VVLESPLQLPNGKLDKSLPAPEFGGSHYQRPRNAQEEMLCGLFAEVLEVGSVGIDDSF 960
                *::***::**::*****:::*****: *::*****:***:*****::*****
                *::***::**::*****:::*****: *::*****:***:*****::*****

FrsD_Burk      FDLGGHSLLATRLVDRI RTTLGVLP IRTIFDQPTIAELAVVLPQCQA ITRPPLRARRRV 1019
FrsD-Chro      FDLGGHSLLATRLISRIRAALNVELPIRQLFDLPSVAELLEVL PQYQGAARPALQPRRRQ 1020
                *****:::***::*::***** :** *::*** ***** *::** *:: ***
                *****:::***::*::***** :** *::*** ***** *::** *:: ***

FrsD_Burk      ENSDWS* 1025
FrsD-Chro      QHAVRS* 1026
                ::: **

```

Abbildung A39: Alignment von FrsD aus dem *bcf*rs-BGC und dem *cvf*rs-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

FrsE_Burk      MQNSAIEVLPLSYAQRRLWFTHCIKGPSPAYNIPLMLRLAGKPAKEALQAALQDVLTRHE 60
FrsE_Chro      MPIPEIEVLPLSYAQRRLWFTHRFNGSPSTYNIPIALSLSGEPEQPALQAALRDVLARHE 60
*              ***** :.*****: * *:* * : *****:***:***

FrsE_Burk      SLRTLVCVEADDGPEMQHILPAQAVTVFRLETHVAASVAEQSEAVVEASRHCDFLSTEIPL 120
FrsE_Chro      SLRTLVESEEGAPAQHILPPDAAGTFCLAVQTAGSAAEQAASLETACRHCDFLSREMP 120
***** *::: * * ***** :* . * * :.:*.*.***: :. * .***** *:*

FrsE_Burk      RATLFLAEGAPLLLLLLHHIAADGDSLPLVAKNLEFAYLARHESRPEWSLLAVQYADY 180
FrsE_Chro      RAGLFLADGAEPVLLLLLLHHIAADGDSLPLVARDLELAYRSRLRGLAPEWRPLPVQYADY 180
** ***** * * :*****:*****:*****: * * . * * * * *

FrsE_Burk      TLWQREWLGNIGTADSPAHLQRYWRGALRGMPQVMALPTDRPRPPVATHRGGKVPFALP 240
FrsE_Chro      ALWQQELLGDLEVPDLSAARQLVYWRREALRGMPDVLELPTDHRPRAVASHRGGKAPFALS 240
:***.* *::. * * *.* * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      AAAHARLKTLAETEAVTLSMVLQAGLSALLYRLGAGSDVVIIGLLAGRNDKALDILGFF 300
FrsE_Chro      AAAHARLKTAAQQGATLSMALQAGLAALLHRLGAGTDIAVGGLLAGRGEESLKEILGFF 300
***** * .: .*****.*****:*****:*****:*****: * : * * * * *

FrsE_Burk      VNAWVLRDLSGHPDFHVLRRVREQALQAYSHPDLPFEWLVEQLNPIRSTSYHPLFQVV 360
FrsE_Chro      VNAWVLRVDVSGRPNFPALLRRVREQALQAYSHQDLPFEWLVEQINPARSTSHHPLFQVA 360
***** * : * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      LVLQNNQRARFRLGGLVVEQVLTGTAKFDLAFNLFETMSDDGHP LGVTGDLEYACDLF 420
FrsE_Chro      LVLQNNLSADFHLEGLAVDQRLLTGTAKFDLAFNLFERQDGDGRPQGLDGELEFASDLF 420
***** * * : * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      DPPSAVRLAYRLSRLEIWSAAPSQSIAILDLLERSEREQALLEWNATTRPLPALTAEA 480
FrsE_Chro      DRAGAERLVARFVRLLRWSMQPAHLIAELDLLDDERRQAVSGWNDSARP LEAGTLACL 480
* . * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      FETQAALTPEAVALAFGDEVLSYAE LNKQANRLARMLVAAGLGPEGRVALAVPRSLDMVV 540
FrsE_Chro      FEAQAKASPDVALAFGSEQSYAE LDRRANQLARTLAGAGIGPEDIVALAVPRSLDMVV 540
**.* * :.:***** * * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      ALLGVTKAGAAYLPDPEYPAERLAYMLADAKPTLLMTVNAQLGSLSECAGIPVLALDAD 600
FrsE_Chro      ALLGVVKAGAAYLPDPEYPRERLAHMLSDATPRLLLATSDTVGGLPAFSGLRVQLDEP 600
***** .***** *****:***.* * * * :.: * * * * * * * * *

FrsE_Burk      SVRDAISQMSGCNLVQSERLCPQPVCYIYTSGSTGRPKGVMVTHQGIVSLRASQIE 660
FrsE_Chro      AWREMVADGRPLAQRERTRPLLPQHPVCYIYTSGSTGKPKGVSVTHQGIASLRASQIE 660
: * : : : * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      RFGVSAESSVLQFASLSFGAALFEICTSLLTGARLVLVSSIKEALNVETMTALVTRHRLS 720
FrsE_Chro      RFGVSAGSSVLQFASLSFGAALFEVCMSSLVGARLVLVASAREALNAGAMAE LARRHGLS 720
***** *****:***** * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      H MVPPSALDTLCADRLPRTVRIMVAGEHCPAHLVERWSADRFMVNGYGSSEVTVCATMS 780
FrsE_Chro      HVVLPPSALEALASERLPDSLGMVAGEHCPAHLQERWSAGRLMVNGYGSSEVTVCATIS 780
* : * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      QPLSGRALPPMGAPNANTRLYLLDAGMQVPVAGVMGELYVAGEGLARGYLSRPLTAERF 840
FrsE_Chro      LPLSGRGAPPMGLPNANTRLYVLDAGLQVPVPGVPGELYIAGDGLARGYLRPGLS AERF 840
***** . * * * * * * * * * * * * * * * * * * * * * * * * * * *

FrsE_Burk      VANPFEPGQMYRTGDLARRDIDGRLDYLGRVDHQVKIRGFRIEPAEIEAALRQLPGVAQ 900
FrsE_Chro      VANPFAEGERYRTGDLARRGGDGRLEYLGRTDHQVKIRGFRIEPAEIEAVLRRLSGVAQ 900
***** * :*****. *****:*****.*****:*****:*****: * * * * *

```

FrsE_Burk	ATVVAWEVPGAKQLVGYVVPKEGVMLEPRAMRELAEHLTDHMVPAVLVELSALPRTPN	960
FrsE_Chro	AAVVAREDPAGVRQLVGYVVAATEAKLDAQGLRRQLAEHLPEHMVPAALVELAALPRTPN *:*:*:* *:*:*:*:*:*:*:* *:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*	960
FrsE_Burk	GKLDRSALPAPVFTVEGYREPRTAKECTLCQLFAQVLDLPQVGIDDAFFDLGGDSISSIQ	1020
FrsE_Chro	GKFDRGALPVPTFAAEGYRPPRTAEERALCQLYAQVLDLSKVGIDGDFDLGGDSISSIQ *:	1020
FrsE_Burk	LASRARKRGWHVTPNQVFRYPRVQDLATVMLPLTNGVDSVQEEPCGDLTLPILRWMWEN	1080
FrsE_Chro	LASRARKAGWLLTPNQIFRHPRVEDLAAELKPVESDRNERVDQACGDLPATPIIHWMLAA ***** * * :*****:*:*:*:*:*:*: : * : . . . : : ***** * : : * *	1080
FrsE_Burk	GPYRLFHQSQLLRAPSGLRRADLLAMVQALLDHHDALRMLRHEDDG - EARMTILPVGTTR	1139
FrsE_Chro	PPYRRFNQSQLLRTPGGLKRDDLALQALLDHHDALRRLDAAASDEALLSIPPAGSVR *** * :*****:*:*:* * * * * :*****:*:*:*:*:*:*:*:*:*:*:*:*:	1140
FrsE_Burk	AEDCVRRIEIVGVDVERQVVLARETDEAILRLDSECGRVQVWLDAGSEEGWLRLLVIH	1199
FrsE_Chro	AGDCLRRIDAVGRDEAEWHLLAREGEAAAERLDCEAGLLQAVWLDAGDEAGLLVIH * * * :*****: * * * * * : : * * * * * : * * * * * * * * * * * * * * * * *	1200
FrsE_Burk	HLAVDGVSWRVLLSDWQQAWADVVCVCAISLDPVGTFRNWLCLQRDAQSPQREAEALAY	1259
FrsE_Chro	HLAVDGVSWRVLLPDLQAWQAASQDRAIALDPVGASFRWLSHLGQEARSSQREAEALAH ***** * * * * * . . . * :*****:*:*:* * * * * : : * * * * * * * * * * :	1260
FrsE_Burk	WCSMLSTTDMPLGRRAFDPARDTTRTKQSLSLSPVRTTQALLTQAATRFHAQANDVLLT	1319
FrsE_Chro	WKWALAAEDPLLGPYPYDSARDATRQSLSLALPPAVTQALLTQATARFHAHANDVLLT * * * : *	1320
FrsE_Burk	VFVLAMAARRQCMGHAPDALLFDLEGHGRETQDTAIDLSRTVGFWSLFPVRVRLDAVD	1379
FrsE_Chro	VFALAVAVWRRRRLPEAPAEALLFDLEGHGREARDTGIDLSRTVGFWSLFPVRLALDCAG *:	1380
FrsE_Burk	LDDALGEGASLGRLLKSVKEQLHALPDRGLGFGLLRYLNQGTAAELAAHQPPQIGFNLYG	1439
FrsE_Chro	LDEALKGGDSLGRLLKSVKEQLRAIPDRGMGFGLLRHLNPGARGELAAALSSPQIGFNLYG *:*:* *	1440
FrsE_Burk	RFAASEGGDWQLASDVGIEAGQDPEMPLPHPLSFDHTLDRTHGPELTAIWSWGSSELFSS	1499
FrsE_Chro	RFTAEGEDWQPAGCAGIEGGQDPMPLPHLLSFNAQTLDREQPELHAIWSWAGELFDQ *:*:*:* *	1500
FrsE_Burk	DEIAELAQLWQQAAMALAEHVTRPGAGGRTPSDLPLVHLHQAQIEQLEVEYPRIEEVLP	1559
FrsE_Chro	EQIAELAQLWRQAVALAEHASRPEAGGRSPSDLPLVGLQEQIERLEAEYRPLEDVLPL : : *	1560
FrsE_Burk	SPLQKGLLFHGLYDPAGVDPYVERLIYALEGELDAGALKQAVHGLLQHSNLRACFVLDG	1619
FrsE_Chro	SPLQKGLLFHGLYDPAGVDPYIERVTYQLEGELDPAAMKRAAHGLLQRHGNLRACFVLDG ***** *	1620
FrsE_Burk	RGQPQVIVPLSALPWQEIIDL SMLGEDEQAVLEQMQEEDRHRQFDL SHAPLLSFLVIRL	1679
FrsE_Chro	KGQPQVQIISQLEALPWQIDIDL SMLAGDERQAMLAQIREQVRSQRFDLGRAPLLSFTLIRL : *	1680
FrsE_Burk	AVDRHRLIMNHHILLDGWSPLLWRELMKLYRSGGDLRAIPRVTPYRDYLDWLARRDLE	1739
FrsE_Chro	AADRHLQIMNHHILLDGWSEPLLWREMLTYRNGGDLGAMPRVTPYRDYLVWLGRRDHE *	1740
FrsE_Burk	PDRMAWRGYLRDLVTPTLLAPAAPTEYVIQETYERALPDALASGLTALAEQLGVTLNVTI	1799
FrsE_Chro	ADRQAWRHYLELETPTLLSPEPPAAYVDQETYSALPPALAQALAAARAEELGITLNTLV *	1800

FrsE_Burk	QGAWGRVLGCLTTSQDVMFGSNVAGRPAELNGIEDMIGLFIINTIPLRVRWSRGESIGDVL	1859
FrsE_Chro	QGAWGRVLACLTMSQDVFSGSNVAGRPAELGGIEDMVGLFINTVPLRVRWTRGESIADML ***** .*** *****.*****.*****.*****.*****.*****.*.*	1860
FrsE_Burk	KRIQSEQVDLLEHQYLDLVEIQSQASHRDLFDSVYAFENYPVHANDEDEASGRPVKVVSG	1919
FrsE_Chro	VRLQAEQVGLLEHQYLDLAEIQELAGLGDLDFTVYAFENYPVFDGGASAAGPRVTGVSG *.*.***.*****.***. * . *****.*****. . . . *.*.***. ***	1920
FrsE_Burk	GSTTHYPLGLIVNPQAGLSLLFSYRPDCYRRCDIERIAAYLQCVLEAFVDSTQPIAQLD	1979
FrsE_Chro	GSTTHYPLGLIVNPQAGLSLLFSYRPDCYRRCDIERIAAYLQCVLEAFVDSTQPIAQLD *****.*****.* * * * .***.***.*** * .*.*** * .*. * * *	1980
FrsE_Burk	LLPPEQANGIAQWNTQHACPSADLAQLFERQVRLTPDASALTFGSQTLASYAVLNACANR	2039
FrsE_Chro	LAPVEQAGWLRWNTDGHAYPDGDLAQLFERQARLTPNAQALIFERQSLSYAALNARANQ * * * * . : . * * * * * * * . * * * * * * * . * * * * * * * . * * * * * * *	2040
FrsE_Burk	LARWLLMHSIGPDDVVAVALPRSIDLVIALLAVVKSQAAYLPLDADYPRNRDLDFMLTDAR	2099
FrsE_Chro	LCRVLLAHGVPDDVVAVALPRSIELVVALLAVVKSQAAYLPLDADYPRNRDLDFMLTDAR * . * * * * . : * * * * * * * . * * * * * * * . * * * * * * * . * * * * * * *	2100
FrsE_Burk	PRALLTNGSMVEALSPAAGTQVLLLDAPewTAARNHLDDRDMVTERKQPLRPLDAAVVI	2159
FrsE_Chro	PAVLLSNAAMAGILSPADGTRLLSLDEPGLLSAQDGADSGDLAAGERRRRLRPQDAAYVI * . * . * . * . * . *	2160
FrsE_Burk	YTSGSTGLPKGVVNTHHGIVNRLTWMQSAYRLDASDVVLQKTPFSFDVSVWEFFWPLLNG	2219
FrsE_Chro	YTSGSTGKPKGVINTHQGI VNR L A W M Q S A Y R L D A S D T V L Q K T P F S F D V S V W E F F W P L L E G ***** * * * . * * * . *	2220
FrsE_Burk	ARLVMAVPDGHDPAYLAELIQRQGVTTLHFVPSMLDAFLNEPSSRQCLSLKRVLCSGEV	2279
FrsE_Chro	ARLVAVPDGHDPAYLAALIQRERVTTVHFVPSMLEVFLREAGSRQCLSLRRVLCSGEA * * * * . * . * * * * * * * .	2280
FrsE_Burk	LSGNLAALQQHVLKRPLHNLGYPTEAAVDVTAHACDPNDTGSSIQIGKPIWNTRIHVLE	2339
FrsE_Chro	LSGELSALHRQVLGSPHNLGYPTEAAVDVTAHACENGETGVSVPIGAPIWNTRIHVLE * * * . * . * . * . *	2340
FrsE_Burk	GLRPVPLGVAGELYIGGTGLARGYLNRAGLTAERFVANPYGEGERYRSGDLARWNAEGE	2399
FrsE_Chro	GLRPAPVGVAGELYIAGAGLARGYLNRPGLTAERFVANPYGEGERYRSGDLARWTEEGE * * * * . * . * * * * * * . * . *	2400
FrsE_Burk	LEYLGRLDQQLKIRGFRIEPEIEAALCRHPLVSQAQAVIAREDTPGHKQLVGVVLDAA	2459
FrsE_Chro	LEYLGRADQQLKVRGFRIEPEIEAALCRHPSVAQAQAVIAREDTPGHKQLVGVVLDAA ***** * * * * . *	2460
FrsE_Burk	ALQRDAEDETRQVQAWRQVYDTLYEAHCQPFGENFGGWDSSYDQGQPLPLEMRAWRQAT	2519
FrsE_Chro	ALQRDEDESERQVEAWQVYDTLYDAHQHPFGENFGGWDSSYDQGQPLPLAQMREWRSAT * * * * . : * . * * * * . * . *	2520
FrsE_Burk	VERIRELQPRRLEIGVGSGLLLAPLAPECEAYWGTDLSPATTIIVLERQLEQTFGDRVR	2579
FrsE_Chro	VERIRELRPRRLEIGVGSGLLLAPLADACEAYWGTDLSPATTIIVLEKQLETSQSCRDKVR ***** . * * . * . * * . * . * * *	2580
FrsE_Burk	LFALGAHELLKLPENFDVIVINSVLQYFPNAAYLTEVIKQSLERLSPGGALYLGDVRL	2639
FrsE_Chro	LFALGAHELARLPAMRFDCIVINSVLQYFPNAAYLGEVIEQALARLEAGGALYLGDVRL ***** * . * * . *	2640
FrsE_Burk	ALLDVFAVAVDLCQVGTETDAASLRRRVEQRLLAEKELLVSPAFFSRLRETLQPQIGAVDI	2699
FrsE_Chro	ELLPSFAAAVELRQSEPEVDAALQRRVSQRLLAEKELLAPDFFSRLREQLPQIGAVDI *	2700

FrsE_Burk	RLKRGEAVNELNRYRYEAVLYKEPCEAVSLSKVETQTSWSDVGNMSACRTRLMQYASMWRI	2759
FrsE_Chro	RLKRGEAAANELNRYRYEVVLRKGPQARSLASAAAEPWSSLGSLACRRLSVGGDALRV *****.*****.* * **:* **:. . : : * : . : * : * : * : * : * : * : * : *	2760
FrsE_Burk	EGIPNARLHGEVLVTRRLKSEEGLASWRRWLDESGGVEPEDLYRLGAELGYRVVLTWSNQ	2819
FrsE_Chro	TGVPNALLHGEEAAAARELKAGGSPSALLARLDENGGVRPEALRRLGAELGWRMLATWSRQ * : * * * * * * * * * * . : * : * * : . . : : * * * . * * * . * * * * * * * : * : * * * . *	2820
FrsE_Burk	PDHFDAVFLPEQQRGALDAVYRTPPVLKPLSGYVNNPANFEQFAAIRRYVGEQLPDYMVP	2879
FrsE_Chro	AGHFDAVFRGEDGEALDGVYQPAGALQPLSGYVNNPANFEQYAAIRRYALEQLPEYMVP .*****: . : * * * . * * : . : * : * * * * * * * * * * * : * * * * * . * * * : * * * *	2880
FrsE_Burk	AALVLLLEGLPLTPNGKLDLRRALPAPEFGTACYRAPGSKQEKVLAQLFAEVLGLPQVGDD	2939
FrsE_Chro	AAIVLLDALPLTPNGKLDLRRALPAPEFGGTGYRAPESEREQLLARLFGEVLGLPQVGADD * : * * * : . * * * * * * * * * * * * * * * * : * * * * * : * : * : * : * * * . * * * * * * * . *	2940
FrsE_Burk	SFFDLGGDSIISIQLVSRARKAGWQLTPRDIFQQPTIAALARLITPLGNEPATPASVAAL	2999
FrsE_Chro	SFFDLGGDSIISIQLVSRARKAGWRLTPRDVFQQPTVAALAQAITPLDDEPAAPAPAATP *****.*****.*****.*****: *****.*****.* * : * * * . * * * . * * :	3000
FrsE_Burk	RLDQDELDYLGSLYGS*	3015
FrsE_Chro	KLDQDELNYLGSLYES* :*****:***** * *	3016

Abbildung A40: Alignment von FrsE aus dem *bcfrs*-BGC und dem *cvfrs*-BGC.

CLUSTAL O(1.2.4) multiple sequence alignment

```

FrsG_Burk      MDRKKIETTYPLSPLQKGF LFHAGYDLAHADNYIAQLFLDFDGEVDAGLMRAAADTLIRR 60
FrsG_Chro      MNPKKIEAAYPLSPLQKGF LFHAGYDLQSADSYVAQLFLDFEGELDGAAMRSAADALMRR 60
                *:  ****:***** ** *:*****:*. * . *:***:*.**

FrsG_Burk      HANLRAGFVHPGGREPQVILREVAACWKEHDWRTEPLVRAAELQSAWHAEDRQRFFDLS 120
FrsG_Chro      HANLRAGFVHPGGQEPVQVVLREVAGWEERDWRGRNPQEAEEAQSAAWQEAADRERRFELS 120
                *****:*****.* *:*.*** . *** ***: **:*:*.*

FrsG_Burk      QPPLLRFGLWLRPEERTQLVLTYYHILLDGSWLPVLEELLTYRTQGDALSLPKTTPYS 180
FrsG_Chro      QPPLLRFGLWLRPAGRSQLVVTYHILLDGSWLPVLEELLALYRAGGGDGGGLPEATPYS 180
                ***** *.*:*****:*****:***: * . *:*:*.*

FrsG_Burk      TYLGLWLGQRDRASAQAWGDYLSGLEGPTLLARRSASEDQTQSKSLTLPIELTQALNQ 240
FrsG_Chro      AYLGLWQERDRAAAACEAWGGYLEGLEPTLVSTGHPQDHAEQKQRAWRLPAELTQALTRQ 240
                :***** ***:.* :***.*.*****: . . . *.: : ** *****:.*

FrsG_Burk      ARQQGVTLNLTLLQAAWGILLGKLSRRDVFVGITVAGRPGELPGVERMIGLFINTVPLRL 300
FrsG_Chro      ARQQGVTLNLTLLQAAWGMLLGKLNLSRDVFVGITVAGRPGELPGVERMIGLFINTVPLRL 300
                *****:*****.***** *****:*****:*****.*

FrsG_Burk      RWRAGETVVELLERLQREQARLLEYQYLDLAEIQRLAGQRQLFDTLCVFENYPVNAKAI 360
FrsG_Chro      RWSAGETVAGLVGRLQREQAGLLDQHLDLVEIQRLAGQRQLFDTLCVFENYPVDAAME 360
                ** ***** * . ***** ***:.*.***.*****:*****:*****.* * :

FrsG_Burk      QQDEGFGLRHISGGDRYMTYPLSVMIEPGERMTLNLIYRPASFDA - - AKRLGAQLIRLL 418
FrsG_Chro      QPSGGLQLRGVSGGDRYMTYPLSLMIEPPELKLNIYQPERFEPEAIERLSAQLTRLL 420
                * . * . ** :*****:*****. .:*****.* *: .*:*** **

FrsG_Burk      EAIATVPQSPIDTLPWLDKSERRQLLEEWSGKALDSGEITLAELEAQATRPNAVALEG 478
FrsG_Chro      GVIAAEPSPVSDIELLDEAERRQLLDVWNRTGPDHGQATFPQLFETQAALTPHAVALES 480
                .**: *.*.*: . ***:***** :*. . * *: * .:***:***: *:*:*.*

FrsG_Burk      PDGCVSYGELDARANRLASHLQSLGVGPDVVVGVCLERSIDMVVAI LGIAKAGAAYLPLA 538
FrsG_Chro      PDARLSYAELDARANRLARHLQSLGVGADVLVGICLERSIDMVVAI LGALKSGAAYLPLS 540
                ** . ** ***** ***** ***:*****:*** *.*:*****:

FrsG_Burk      PDYPTERLAYMLTDSMAPVLLTESKQVERLPSYWGHLVELDRLDLQASSAPARALRPD 598
FrsG_Chro      PEYPTERLAYMLGDSMAPVLLTDSAQVERLPSYWGRVVELDRLDLQASSAPERALARAE 600
                *.*:***** *****.* *****:*****. ***** ** * ** *

FrsG_Burk      HLAYVIYTSGSTGQPKGVAVSHAGLAGLVKSQEERFAVAGPVRVLQFASLSFDAAVMEIL 658
FrsG_Chro      HLAYVIYTSGSTGQPKGVAVSHAGLAGLAGSQTERTALQGPTRVLQFASLSFDAAVMEIL 660
                *****:*****. ** *****. ** *****:*****.*

FrsG_Burk      MAFCSGRLLVPAAGPLLGEQLEKTLNRYAISHALIAPSALETVEAEVVPGLSTLVVGG 718
FrsG_Chro      MAFCSGRLLVPAAGPLLGEQLLDTLNRHEISHALISPSALSTADAALAPVLR TLVVG 720
                *****:*****. *****:*****.* * . * * *****

FrsG_Burk      TCSGATAASWSQGRMVNAYGPTVTVCVTMSKPLSGSDPKLGRPTLGAKLYVLDSTLQ 778
FrsG_Chro      ACPGATVAAWSAGRRMVNAYGPTVTVCVTMSKPLSGSDPKLGRPTLNARLYVLDGALQ 780
                :* ** *.* ** *****.* *****:*****. ***** *.*:*****:***

FrsG_Burk      PVPVGAGELYIAGRGLARGYYQRPGLTAERFVANPYGKGERLYRSGDLARWSGEGELE 838
FrsG_Chro      LAPVGAGELYIAGAGLARGYLNRPGLTAERFVANPYGEGERYRSGDLARWTEEGELE 840
                .***** ***** :*****:*****:*****:*****:*****

FrsG_Burk      LGRVDQQLKIRGFRIEPEIETVLCQHPQVREAVVSR - TNGRDTQLVGYVTVRGEVDGQ 897
FrsG_Chro      LGRSDQVKVRFRIEPEIEAVLNHRPQVSQSVVARQSQGGDSQLVAYVAAVGVGEGS 900
                *** ***:.*:*****:*** :*** :***:*. :* *.*:***:*. * *:*.*
    
```


FrsG_Burk	TQLVNMYGITETTTHVHSYQPLDSGMCGRCSNLSLIGIGIPDLHLFLDACLQVPVGGIGE	1857
FrsG_Chro	PKLANMYGITETTTHVHSYQALDRRLCEEGRSLIGEAIPLRLHLLDRWLQVPVAGGVGE :*.*****.*** * :* . ***** .****.*** *****.***.***	1856
FrsG_Burk	LYVGGAGLARGYHNRAASLTAEFRVANPFASSKRMRYRTGDLARRNVNDIFEYHGRADQQVK	1917
FrsG_Chro	LYIGGAGLARGYLNRPLSARFIANPFAAGERMYRSGDLARRNAAGALEYQGRADQQVK **.****** ** .*:****.*****.:.****.*****. . .:*.*****	1916
FrsG_Burk	VRGFRIELGEIETVLRQHPGVEDARVVVQTLHDNDCRLVAYLQPSERTAAPLRRWLKVR	1977
FrsG_Chro	VRGFRIEPEIETALRGHPGVEDARVVVKAMDGNQRLIAYLLPSERAAAPLRRWLRAS ***** *****.* *****:..* **.* **.*:*****.:.	1976
FrsG_Burk	DAAEQGARAFELPNGMPIFHHTMETEFLYEEIFEDLVYFKHGIRLDDGACVFDVGANIG	2037
FrsG_Chro	SVAQAQNAEIELPNGLPVFHNAEETFLYEEIFEDRIYKHGIRLDDACVFDVGANIG . . *.** :****.**:****. ***** ***** :*:*****.:*****	2036
FrsG_Burk	LFMLFVQGHCNATIFAFEPVPPVFRITLNAEVHGDVKVRLFDCGLSDAVRQERFTFYPN	2097
FrsG_Chro	LFTLVFGQHCNATVFAFEPVPPVFTLSLNAAVHGGKVRVLFDCGLSDAAQETFTFYPN ** ***** **.****** **.* **.* ***** **.* *****	2096
FrsG_Burk	DTLISTSRNSPEAIRGMVKSFLINQHGNSGDDAEVGEELDERLASQDYMCVLRRLSDVI	2157
FrsG_Chro	DTLISSSRNSAEATRMMVKSFLINQHGSDGE--AVDELLEERLSSQYVCGLRSLSDVV *****.*** ** * *****.:* : *.*.*.***.***.***.***.***.***.	2154
FrsG_Burk	TEHNVDRIDLKIDVENAEYVNLQGIVESDWPKIRQLVMEVHDVDGRLRRIVDLLHRGY	2217
FrsG_Chro	AEHGVERIDLKIDVENAEYDVLRGISDADWPKIRQLVMEVHDVDGRLRRIVDLLHRGY :*.**.******.***.*** :* :*****.***.***.***.***.***.***.***.***	2214
FrsG_Burk	RVIHEQDRLLRDTAIHCLYAVHVSRAEMRPGHHAENTSLFWNCRADLLRDVRLRKRRL	2277
FrsG_Chro	RVIHEQDRLLQNTAIHCLYAVHASCASADAVPKAEPAAVWRSRSLRDVQAEVLRSL *****.:*****.* :. : :.***.***.***.***.***.***.***.***	2274
FrsG_Burk	PDYMQPNHMLLDTFPLTPNGKLDLRRALPAPEQAAMRVRDIEPAEIEAALCRHPLVSQAV	2337
FrsG_Chro	PDYMQPNHLVLLDAFPLTANGKLDLRRALTPPEQAALRTRDVPDPAEIEAALCRHPSVAQA *****.***.***.*** *****.***.***.***.***.***.***.***.***	2334
FrsG_Burk	VIAREDPGHKQLVGYVVLDDAAALQRDTEDETRQVQAWRQVYDTLYEAHCQQPFGENFG	2397
FrsG_Chro	VIAREDPGHKQLVGYVVLDDAAALQRDEDESERQVEAWQVYDTLYDAHQHPFGENFG *****.*****.*****.***.***.***.***.***.***.***.***.***.***.***	2394
FrsG_Burk	GWSSYDGGQLPLEEMRAWRQATVERIRELQPRRVLQVGVGSGLLLAPLAPECEAYWGT	2457
FrsG_Chro	GWSSYDGGQLPLAQMRWRSATVERIRELRPRRVLQVGVGSGLLLAPLADACEAYWGT **.******.*** ** * *****.***.***.***.***.***.***.***.***	2454
FrsG_Burk	LSPTTIAVLERQLSEQTFGDRVRLFALGAHELKSLPEGNFDVIVINSVLQYFPNAAYLTE	2517
FrsG_Chro	LSPATIAVLEKQLETQSCRDKVRLFALGAHELARLPAMRFDVIVINSVLQYFPNAAYLGE ***.***.***.*** * :* :*****.***.***.***.***.***.***.***.***	2514
FrsG_Burk	VIKQSLERLSPGGALYLDVNRNALLDVFATAVDLCQVGTETDAASLRRRVEQRLLAKE	2577
FrsG_Chro	VIEQALARLEAGGALYLDVNRNLELLPSFAAAVELRQSEPEVDAALQRRVSQRLLAKE **.* ** * *****.*** **.* **.* * .***.***.***.***.***.***	2574
FrsG_Burk	LLVSPAFFSRLRETLQIGAVDIRLKRGEAVNELSRYRYEAVLYKEPCEAVSLSEVETQT	2637
FrsG_Chro	LLLAPDFFSRLRETLQIGAVDIRLKRGEAAVELNRYRYEVVLRKGPQARSLASAAAEP **.* ***** *****.***.***.***.***.***.***.***.***.***.***.***	2634
FrsG_Burk	WSDVGNMSACRTRLMQYAPMWRIEIPNARLHGEVLVTRRLKSEDGLASWRRLDESGGV	2697
FrsG_Chro	WSSLGSLACRRLSVGGDALRVTGVPNALLHGEAAAARELKAGGSPSALLARLDENGGV **.* :* :*** ** * :* :***.***.***.***.***.***.***.***.***	2694


```
#  
#  
# Percent Identity Matrix - created by Clustal2.1  
#  
#
```

```
1: FrsA_Burk 100.00 71.47  
2: FrsA_Chro 71.47 100.00
```

```
#  
#  
# Percent Identity Matrix - created by Clustal2.1  
#  
#
```

```
1: FrsB_Burk 100.00 75.00  
2: FrsB_Chro 75.00 100.00
```

```
#  
#  
# Percent Identity Matrix - created by Clustal2.1  
#  
#
```

```
1: FrsC_Burk 100.00 72.26  
2: FrsC_Chro 72.26 100.00
```

```
#  
#  
# Percent Identity Matrix - created by Clustal2.1  
#  
#
```

```
1: FrsD_Burk 100.00 69.66  
2: FrsD-Chro 69.66 100.00
```

```
#  
#  
# Percent Identity Matrix - created by Clustal2.1  
#  
#
```

```
1: FrsE_Burk 100.00 71.04  
2: FrsE_Chro 71.04 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: FrsF_Burk 100.00 75.11
2: FrsF_Chro 75.11 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: FrsG_Burk 100.00 72.78
2: FrsG_Chro 72.78 100.00
```

```
#
#
# Percent Identity Matrix - created by Clustal2.1
#
#
```

```
1: FrsH_Burk 100.00 84.56
2: FrsH_Chro 84.56 100.00
```

Abbildung A44: Percent Identity Matrices der Alignments von FrsA-FrsH aus dem *befrs*-BGC und dem *cvfrs*-BGC.