Structural and functional characterization of the triplet acyl carrier protein in the curacin cluster and its interaction partners

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich

Biochemie, Chemie und Pharmazie (FB 14)

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Alena Ella-Louise Busche

aus Oldenburg

Frankfurt 2011

(D 30)

Vom Fachbereich Biochemie, Chemie und Pharmazie (FB14)
der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Dieter Steinhilber
Gutachter:	Prof. Dr. Volker Dötsch
	Prof. Dr. Clemens Glaubitz

Datum der Disputation:

25.01.2012

„Auch aus Steinen, die einem in den Weg gelegt werden, kann man Schönes bauen“

Johann Wolfgang von Goethe

Table of contents

1 Summary

According to the World Health Organization (WHO) bacterial resistance to antibiotic drug therapy is emerging as a major public health problem around the world. Infectious diseases seriously threaten the health and economy of all countries. Hence, the preservation of the effectiveness of antibiotics is a world wide priority. The key to preserving the power of antibiotics lies in maintaining their diversity. Many microorganisms are capable of producing these bioactive products, the so called antibiotics. Specifically in microorganisms, polyketide synthases (PKS) and non-ribosomal peptide synthases (NRPS) produce these natural bioactive compounds. Besides being used as antibiotics these non-ribosomal peptides and polyketides display an even broader spectrum of biological activities, e.g. as antivirals, immunosuppressants or in antitumor therapy. The wide functional spectrum of the peptides and ketides is due to their structural diversity. Mostly they are cyclic or branched cyclic compounds, containing non-proteinogenic amino acids, small heterocyclic rings and other unusual modifications such as epimerization, methylation, *N*-formylation or heterocyclization. It is has been shown that these modifications are important for biological activity, but little is known about their biosynthetic origin.

PKS and NRPS are multidomain protein assembly lines which function by sequentially elongating a growing polyketide or peptide chain by incorporating acyl units or amino acids, respectively. The growing product is attached via a thioester linkage to the 4'-phosphopantetheine (4'-Ppant) arm of a holo acyl carrier protein (ACP) in PKSs or holo peptidyl carrier protein (PCP) in NRPSs and is passed from one module to another along the chain of reaction centers. The modular arrangement makes PKS and NRPS systems an interesting target for protein engineering. More than 200 novel polyketide compounds have already been created by module swapping, gene deletion or other specific manipulations. Unfortunately, however, engineered PKS often fail to produce significant amounts of the desired products. Structural studies may faciliate yield improvement from engineered systems by providing a more complete understanding of the interface between the different domains. While some information about domain-domain interactions, involving the most common enzymatic modules, ketosynthase and acyltransferase, is starting to emerge, little is known about the interaction of ACP domains with other modifying enzymes such as methyltransferases, epimerases or halogenases.

To further improve the understanding of domain-domain interactions this work focuses on the curacin A assembly line. Curacin A, which exhibits anti-mitotic activity, is from the marine cyanobacterium *Lyngbya majuscula*. This outstanding natural product contains a cyclopropane ring, a thiazoline ring, an internal *cis* double bond and a terminal alkene. The biosynthesis of curacin A is performed by a 2.2 Mega Dalton (MDa) hybrid PKS-NRPS cluster. A 10-enzyme assembly catalyzes the formation of the cyclopropane moiety as the first building block of the final product. Interestingly, for these enzymes the substrate is presented by an unusual cluster of three consecutive ACPs ($ACP_{I,II,III}$). Little is known about the function of multiple ACPs which are supposed to increase the overall flux for enhanced production of secondary metabolites.

The first task in this work was to elucidate the structural effect of the triplet ACP repetition by nuclear magnetic resonance (NMR). The initial data show that the excised ACP_I, ACP_{II} or ACP_{III} proteins resulted in [^{15}N, ^{1}H]-TROSY spectra with strong chemical shift perturbations (CSPs), suggesting an effect on the structure. The triplet ACP domains display a high sequence identity (93-100%) making structural investigation using usual NMR techniques due to high peak overlap impossible. To enable the investigation of the triplet ACP in its native composition we developed a powerful method, the three fragment ligation. Segmental labeling allows incorporating isotopes into one single domain in its multidomain context. As a result we could prepare the triplet ACP with only one domain isotopically labeled and therefore assign the full length protein. In this way our method paved the way to study the structural effects of the triplet ACP repetition. We could show unexpectedly, that, despite the fact that the triplet repeat of CurA $ACP_{I,II,III}$ has a synergistic effect in the biosynthesis of CurA, the domains are structurally independent.

In the second part of this work, we studied the structure of the isolated ACP_I domain. Our results show that the CurA ACP_I undergoes no major conformational changes upon activation via phosphopantetheinylation and therefore contradicts the conformational switching model which has been proposed for PCPs. Further we report the NMR solution structures of holo-ACP_I and 3-hydroxyl-3-methylglutaryl (HMG)-ACP_I. Data obtained from filtered nuclear overhauser effect (NOE) experiments indicate that the substrate HMG is not sequestered but presented on the ACP surface.

In the third part of this work we focussed on the protein-protein interactions of the isolated ACP_I with its cognate interaction partners. We were especially interested in the interaction with the halogenase (Cur Hal), the first enzyme within the curacin A sub-cluster, acting on the initial hydroxyl-methyl-glutaryl (HMG) attached to ACP_I. Primarily we studied the interaction using NMR titration and fluorescence anisotropy measurements. Surprisingly no complex between ACP_I and Cur Hal could be detected. The combination of an activity assay using matrix-assisted laser desorption/ionization (MALDI) mass spectroscopy and mutational analysis revealed several amino acids of ACP_I that strongly decrease the activity of CurA Hal. Mapping these mutations according to their effect on the Cur Hal activity onto the structure of HMG-ACP_I displays that these amino acids surround the substrate and form a consecutive surface. These results suggest that this surface is important for Cur Hal recognition and selectivity. Our research presented herein is an excellent example for protein-protein interactions in PKS systems underlying a specific recognition process.

2 Zusammenfassung

„Strukturelle und funktionelle Charakterisierung des dreifachen „Acyl Carrier Proteins“ in Curacin A sowie dessen Protein-Protein Interaktionen“

Ziel dieser Arbeit ist es, die strukturelle Bedeutung einer ungewöhnlichen dreifachen Wiederholung von Acyl Träger Proteinen (von englisch *acyl carrier Protein*, ACP) am C-Terminus des Proteins CurA zu beschreiben. Auf der Basis dieser strukturellen Studie wurden des Weiteren die molekularen Mechanismen untersucht, die den Interaktionen zwischen den ACPs sowie deren Proteininteraktionen mit anderen Proteinen zugrunde liegen.

Nach Angaben der Weltgesundheitsorganisation (WHO) ist die steigende Resistenz von Krankheitserregern gegenüber Antibiotika ein zunehmendes Problem. Infektiöse Krankheiten sind eine Gefahr für die Gesundheit der Bevölkerung, zusätzlich ist zu Bedenken, dass durch die Behandlung und Bekämpfung dieser infektiösen Krankheiten nicht zu vernachlässigende Kosten entstehen. Aus diesem Grund ist es von großer Bedeutung, die Wirksamkeit von Antibiotika zu erhalten. Der Schlüssel zum Erhalt der Wirksamkeit liegt in der Vielfalt der auf dem Markt existierenden Antibiotika. Viele Mikroorganismen sind fähig, diese niedermolekularen Stoffwechselprodukte, wie Peptide oder Polyketide, herzustellen. Die Biosynthese wird von Megasynthetasen durchgeführt, genauer gesagt von nichtribosomalen Peptidsynthetasen (NRPS) oder Polyketidsynthetasen (PKS). Die entstehenden Moleküle besitzen oftmals neben ihrer antibakteriellen auch antivirale oder immununterdrückende Wirkung. Des Weiteren konnten einige dieser Peptide oder Polyketide, da sie die unkontrollierte Zellteilung hemmen, als potentielle Medikamente in der Krebstherapie gehandelt werden. Das große Wirkungsspektrum dieser biologischen Substanzen lässt sich durch ihre strukturelle Vielfalt erklären. Überwiegend sind diese Substanzen zyklisch oder besitzen Verzweigungen mit zyklischen Komponenten. Hinzu kommt, dass sie eine Vielzahl ungewöhnlicher und nichtproteinogener Aminosäuren synthetisieren und einbauen können. Viele dieser Peptide oder Ketide werden zusätzlich modifiziert. So finden zum Beispiel Epimerisierungen, Methylierungen, N-Formylierung und Heterozyklisierungen statt. Es wurde gezeigt, dass insbesondere diese ungewöhnlichen Modifizierungen oftmals für die Wirkung verantwortlich sind.

Nichtribosomale Peptidsynthetasen (NRPS) sowie Polyketidsynthetasen (PKS) sind lange Produktionsmaschinerien, die aus verschiedenen katalytischen Domänen zusammengesetzt sind. Dabei ergeben mehrere Domänen ein Modul, welches für den Einbau einer Aminosäure im Falle der NRPS beziehungsweise kurzkettiger Carbonsäuren bei den PKS zuständig ist. Das spätere Produkt wird so nach und nach iterativ zusammengesetzt. Im Zentrum dieser Synthese steht das Peptid Träger Protein (aus dem englischen *peptidyl carrier Protein*, PCP) bei nichtribosomalen

Peptidsynthetasen oder das ACP bei Polyketidsynthetasen. An dieses kleine Protein (10 kilo Dalton (kDa)) ist das wachsende Syntheseprodukt über einen Thioester an den Kofaktor 4'-Phosphopantethein kovalent gebunden. Das mit dem Kofaktor beladene ACP oder PCP wird als holo ACP/PCP bezeichnet und die mit Substrat beladenen Proteine als Acyl-ACP oder Peptidyl-PCP. Durch die Bindung des Substrates an den Kofaktor ist das wachsende Peptid oder Polyketid für andere Proteine leichter zugänglich und kann den jeweils aktiven Zentren der verschiedenen Enzyme für Modifizierungen oder zur Kettenverlängerung präsentiert werden. Das ACP übernimmt hierbei vermutlich die räumliche und zeitliche Koordination der Protein-Protein-/Substrat-Interaktion. Der modulare Aufbau von nichtribosomalen Peptid- und Polyketid-Synthetasen macht sie zu höchst interessanten Zielen für das Chemieingenieurwesen. Mehr als 200 neue Moleküle konnten bisher hergestellt werden, indem Module entfernt, ausgetauscht oder auf andere Art und Weise verändert wurden. Dennoch scheitert die Produktion dieser gewünschten Substanzen im großen Maßstab oftmals an der geringen Ausbeute. Aus diesem Grund ist es von großem Interesse, die von ACPs eingegangenen Protein-Protein Interaktionen, die der Biosynthese zu Grunde liegen, zu verstehen. Erst durch dieses Verständnis wird eine künstliche Veränderung der bestehenden Megasynthetasen, die dessen Produktivität erhöhen könnte, ermöglicht. In den letzten Jahren lieferten biochemische und strukturelle Untersuchungen immer mehr Erkenntnisse über die Funktionsweise und gegenseitige Erkennung verschiedener katalytischer Domänen in NRPS und PKS. Dazu gehören erste Erkenntnisse über die Interaktionen von ACPs mit den häufigsten Interaktionspartnern wie Acyltransferasen und Ketosynthetasen, beziehungsweise deren Pendants in den nichtribosomalen Peptidsynthetasen. Dennoch ist bis jetzt nur wenig bekannt über Protein Interaktionen mit den selteneren aber wichtigen Enzymen wie Methyltransferasen, Epimerasen oder Halogenasen, die essentielle Veränderungen vornehmen.

Um gerade diese Interaktionen besser zu verstehen, wurde in dieser Arbeit eine NRPS/PKS Hybrid-Megasynthetase untersucht, die Curacin A herstellt. Curacin A ist eine biologische Substanz, die aus dem maritimen Cyanobakterium *Lyngbya majuscula* gewonnen wurde. In Experimenten zeigte diese Substanz die Inhibition von ungehemmter Zellteilung, eine Aktivität, die Curacin A zu einem potentiellen Krebsmittel macht. Die Wirkweise von Curacin A ist auf einige besondere strukturelle Eigenarten wie den Cyclopropanring, den Thiazolinring, eine interne *Cis*-C-C-Doppelbinding sowie ein terminales Alken zurückzuführen. Zu Beginn der Biosynthese von Curacin A stehen 10 verschiedene Enzyme, die für die Herstellung des Cyclopropanrings in Curacin A verantwortlich sind. Diese 10 Enzyme sind in den Proteinen CurA bis CurF präsent. Je nach katalytischer Funktion können sie in drei Untergruppen unterteilt werden. Die Erste Gruppe besteht aus der Halogenase Cur Hal, welche in das Protein CurA eingebettet ist. Die zweite Gruppe ist die Hydroxymethyl-glutaryl (HMG) Kassette und umfasst eine dreifache ACP Widerholung am C-terminalen Ende von CurA ($ACP_{I,II,III}$), ein einzelstehendes ACP (CurB), ein Ketosynthetase (KS)

ähnliches Enmzym bezeichnet als CurC, ein HMG-Synthetase-(HCS) ähnliches Enzym CurD, die Dehydratase ECH_1 (CurE) und eine Decarboxylase ECH_2 am C-terminalen Ende von CurF. Der letzte Schritt der Cyclopropanbildung wird durch die als drittes zu benennende Komponente, die Enoyl-Reductase (ER), eingebettet in CurF katalysiert. Interessanterweise agieren all diese Enzyme, während das Substrat an die ungewöhnliche dreifach Wiederholung von ACPs am C-terminus von CurA gebunden ($ACP_{I,II,III}$) ist. Bisher ist wenig über solche multiplen ACPs bekannt, jedoch wurde kürzlich gezeigt, dass sie die Effizienz der aneinandergekoppelten enzymatischen Schritte zur Cyclopropanbildung erhöhen.

Im ersten Teil dieser Doktorarbeit wurde untersucht, welchen strukturellen Einfluss die ACPs innerhalb der dreifachen Wiederholung ($ACP_{I,II,III}$) gegenseitig auf sich ausüben. Zu diesem Zweck wurden zunächst [^{1}H,^{15}N]-TROSY-HSQC Spektren von $ACP_{I,II,III}$ sowie den isolierten ACPs aufgenommen. Die Isolierung jeder Domäne, ob ACP_I, ACP_{II} oder ACP_{III} hatte starke Veränderungen der chemischen Verschiebungen in den HSQC Spektren zur Folge. Solche chemischen Verschiebungen können auf eine Veränderung der Struktur hindeuten. Dadurch, dass die dreifachen ACPs sich in ihrer Sequenz nahezu identisch sind (95-100%), wurde eine strukturelle Untersuchung des nativen, dreifachen ACPs mit den gängigen Kernspinresonanz-Methoden (von engl. *nuclear magnetic resonance*, NMR) aufgrund des ausgeprägten Signalüberlapps unmöglich. Aus diesem Grund wurde im Rahmen dieser Arbeit mit der segmentellen Isotopenmarkierung (aus dem englischen von *segmental labeling*) eine neue Methode entwickelt, die es ermöglicht, die dreifachen ACPs in ihrer natürlichen Umgebung zu untersuchen. Die segmentelle Isotopenmarkierung erlaubt es, Isotope in nur ein Segment, zum Beispiel eine Domäne eines Proteins, einzubauen. Dies geschieht, indem man das Zielprotein in zwei Fragmente teilt. Jedes Fragment wird mit einem Intein fusioniert. Die so entstandenen Proteinkonstrukte können getrennt exprimiert und unterschiedlich markiert werden. Wenn man die so erhaltenen Proteine nun wieder zusammenfügt erkennen sich die Inteinesequenzen und schneiden sich selbst, unter der Ausbildung eines natürlichen Peptidbindung, autokatalytisch herraus. Bisher wurde die segmentelle Isotopenmarkierung nur auf Proteine mit zwei Domänen angewendet. In dieser Arbeit wurde das Protokoll auf Proteine mit drei Domänen ausgeweitet. Das entwickelte Protokoll ermöglichte es uns, jedes ACP der dreifachen ACP Widerholung ($ACP_{I,II,III}$) einzeln mit Isotopen zu kennzeichnen, ohne es aus seinem native Zusammenhang zu isolieren. Auf diese Art und Weise konnte letztendlich $ACP_{I,II,III}$ trotz der zuvor hohen Signalüberlappung, zugeordnet werden. Durch diese Zuordnung wurde es möglich zu erkennen, dass die starken chemischen Verschiebungen, die mit der Isolierung einer Domäne einhergehen, sich nur auf den N- und C-Terminus beschränken und nicht wie zunächst angenommen eine Veränderung der ACP Struktur indizieren. Unerwarteter Weise konnten wir auf diese Art zeigen, dass, obwohl die ACPs einen synergistischen Effekt haben, sie sich dennoch strukturell voneinander unabhängig verhalten.

Durch die Tatsache das sich die ACPs voneinander unabhängig verhalten, konnten wir unsere strukturellen Studien im zweiten Teil dieser Arbeit auf ein ACP (ACP_I) als Repräsentanten fokussieren. Zunächst wurde die *de novo* Struktur von holo-ACP_I mit NMR bestimmt. Von der Frage getrieben, ob das ACP durch die Aktivierung mit dem 4'-Phosphopantethein-Arm einen konformationelle Veränderung durchläuft, wurden Kern-Overhauser-Effekt (engl. *nuclear Overhauser effect,* NOE) Messungen von apo ACP und holo ACP durchgeführt. Der Vergleich der beiden Spektren zeigt, dass das Protein keine strukturellen Veränderungen durch die Aktivierung durchläuft. Somit ist das Modell der strukturellen Veränderung, welches besagt, dass mit der Aktivierung des ACPs einhergehende konformationelle Veränderungen neue Protein-Protein-Interaktionen begünstigen, nicht auf die hier untersuchten ACPs anwendbar. Um die Lokalisierung des Substrats zu bestimmen, wurde $^{15}N/^{13}C$-isotopenmarkiertes ACP_I *in vitro* mit dem unmarkierten Substrat Hydroxymethyl-glutaryl (HMG) beladen. Durch verschiedene NOE Filterexperimente konnten Distanzinformationen für das Substrat erhalten werden und ermöglichten so eine Strukturberechnung von HMG-ACP_I. Die Ergebnisse zeigen, dass das Substrat HMG nicht in einer hydrophoben Tasche gebunden sondern auf der Oberfläche des ACP präsentiert wird.

Der letzte Teil dieser Arbeit befasst sich tiefgehender mit den verschiedenen Proteininteraktionen, die CurA ACP_I ausübt. Die verschiedenen Enzyme (AT-KS Didomäne (CurA), Cur Hal (CurA), ACP (CurB), KS (CurC), HCS (CurD), ECH_1(CurE) und ECH_2 (CurF)) wurden exprimiert und im Falle von geringer Löslichkeit umkloniert und so mit Löslichkeit verstärkenden Proteinen (Ubiquitin (Ub), Maltose bindendes Protein (MBP)) fusioniert. Daraus resultierend stand uns ECH_1 für biochemische Studien und Cur Hal und ECH_2 für strukturelle und biochemische Studien zur Verfügung. Weder in NMR-Titrationen noch in Fluoreszenzanisotropiemessungen von ACP_I mit ECH_2 war eine spezifische Bindung der Proteine detektierbar. Im weiteren Verlauf der Protein-Protein-Interaktionsstudien lag insbesondere der Fokus auf der Interaktion zwischen ACP_I und der Halogenase Cur Hal. Cur Hal katalysiert die Chlorierung von 3-Hydroxy-3-Methyl-Glutaryl (HMG) welches eine der Ersten von einer Folge von Enzymreaktionen katalysiert, die zur Bildung des Cyclopropanrings führen. Es wurde bewiesen, dass Cur Hal nur die Chlorierung von HMG katalysiert, wenn dieses an eins der ACPs im Trippel ACP gebunden ist. Zunächst wurde in dieser Arbeit diese Interaktion mit NMR und Fluoreszenz Anisotropie untersucht. Allerdings konnte auf diese Art und Weise keine Komplexbildung zwischen den Proteinen festgestellt werden. Die Kombination eines Aktivitätstests zusammen mit Massenspektrometrie-Untersuchungen ermöglichte es zu zeigen, dass Cur Hal in der Tat keine Chlorierung von HMG vornimmt, wenn dieses an CurB ACP oder TycB1 PCP gebunden ist, sondern nur, wenn es von einem ACP des Trippel ACP ($ACP_{I,II,III}$) präsentiert wird. Darauf folgend wurden verschiedene Mutanten von ACP_I hergestellt, um die Interaktionsfläche der beiden Proteine zu bestimmen. Mittels Massenspektrometrie konnten mehrere ACP_I Mutanten

gefunden werden, die die Aktivität der Halogenase (Cur Hal) stark beeinträchtigen. Die Ergebnisse zeigen, dass in erster Linie Helix II sowie Helix III von ACP_I für dessen Erkennung durch Cur Hal wichtig sind.Bei der Kennzeichnung dieser Mutationen auf der zuvor gelösten Struktur von HMG-ACP_I, zeigte sich, dass das Substrat zusammen mit den für die Interaktion wichtigen Aminosäuren eine gemeinsame Oberfläche bildet. Zusätzlich konnte gezeigt werden, dass die mangelnde Interaktion nicht durch eine veränderte Lokalisierung des Substrats, sondern tatsächlich durch die Veränderung der Oberfläche ausgelöst wird.

Zusammengefasst konnte in dieser Arbeit zum Einen ein erfolgreiches Protokoll für die segmentelle Dreifragment Isotopenmarkierung erstellt werden, einer Methode, die es ermöglicht, auch andere Proteine in ihrer nativen Umgebung zu studieren. Des Weiteren war es uns möglich zu zeigen, dass die Oberfläche des ACPs zusammen mit dem Substrat HMG ist für die Erkennung durch Cur Hal wichtig ist. Die in dieser Arbeit gewonnenen Daten zeigen, dass die HMG-ACP_I/Hal Interaktion in der Curacin A Synthetase auf der Basis von spezifischen Erkennungsprozessen stattfindet. Dies ist ein weiteres Beispiel dafür, dass die Protein-Protein Interaktionen in PKS und NRPS sehr komplexen und spezifischen Erkennungsprozessen unterliegen.

3 Abbreviations

°	degree
∞	forever
µ	micro
α	alpha
A	ampère
Å	angstrom, 1 Å = 100pm= 1×10^{-10} m
aa	amino acid(s)
AcP	acetyl phosphate
ACP	acetyl carrier protein
ACPS	acetyl carrier protein synthase
App.	appendix
APS	ammoniumperoxodisulfate
AT	acyl transferase
ATP	adenosine triphosphate
bp	base pairs
β	beta
BSA	bovine serum albumin
c	concentration
C	celsius
CD	circular dichroism
CE-CF	continuous-exchange cell-free
CF	cell-free
CoA	coenzymeA
CV	column volume
Da	dalton
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DMSO	dimethyl sulfoxide
DNA	desoxyribonucleinacid
dNTP	desoxyribonucleitriphosphate
DTT	dithiothreitol
E. coli	*Escherischia coli*
ECH_1	dehydratase
ECH_2	decarboxylase
EDTA	ethylenediaminetetraacetic acid
e.g.	for example
EM	electron microscopy
ER	enoylreductase
ESI	electrospray ionisation
et al.	and others
EtBr	ethidium bromide
FAS	fatty acid synthases
Ferene S	3-(2-pyridyl)-5,6-bis(2-(5-furyl sulfonicacid))-1,2,4-triazine
FM	feeding mix
FPLC	fast performance liquid chromatography
Frct.	fraction
g	gram
h	hour(s)
Hal	halogenase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMG	3-hydroxy-3-methylglutaryl coenzymeA
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence

IDA	iminodiacetic acid
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranosid
α-KG	alpha-ketoglutarate
KS	ketosynthase
L	liter(s)
LB	Luria-Bertani
m	milli
M	molar (mol/L)
MALS	multi angle light scattering
MALDI	matrix-assisted laser desorption/ ionization
MBP	maltose binding protein
MS	mass spectrometry
mdeg	millidegree
mg	milligram(s)
min	minute
mL	milliliter(s)
MW	molecular weight
MWCO	molecular weight cut-off
n	nano
Ni-NTA	nickel-nitrilotriacetic acid
nL	nanoliter(s)
nm	nanometer
NMR	nuclear magnetic resonance
NRP	non ribosomal peptide
NRPS	non ribosomal peptide synthases
NTP	nucleotide triphosphate
ON	over night
PPT	pellet
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEP	phosphoenol pyruvate
PK	polyketide
PKS	polyketide sunthetase
RM	reaction mix
RNA	ribonucleic acid
4'- Ppant	4'- phosphopanthetheinyl
ppm	parts per million
PPTase	phosphopantetheineyl transferase
PK	polyketide
PKS	polyketidesynthetases
rpm	round per minute
RT	room temperature
s	second
S	supernatant
SEC	size exlusion chromatography
SDS	sodium dodecylsulfate
SNT	supernatant
Srf	surfactin
ST	sulfotransferase
T	thiolation domain stands for ACPs and PCPs
T7RNAP	T7 RNA polymerase
TCEP	tris-(2-carboxyethyl)phosphine hydrochloride
TE	thioesterase
TE II	thioesterase II
Tris	tris-(hydroxymethyl)methylamine

Triton X-100	polyethylene-glycol P-1,1,3,3- tetra-methyl-butylphenyl-ether
tRNA	transfer RNA
Tween 20	polyoxyethylene-sorbitane-monolaurate 20
Tyc	tyrocidin
Ub	ubiquitin
Ulp1	ubiquitin-like protein specific protease 1
WHO	world health organization

4 Introduction

Bacterial resistance to antibiotic drug therapy is emerging as a major public health problem around the world. Hence infectious diseases seriously threaten the health and economy of all countries, the preservation of the effectiveness of antibiotics is a world wide priority [2, 3]. The key to preserve the power of antibiotics lies in maintaining their diversity. The ability to synthesize such nonribosomal or polyketide compounds is widely spread among microorganisms and these small bioactive compounds find application in modern medicine. As broad as the spectrum of biological activities, is the structural diversity of these peptides and ketides (**Fig. 1**), which are mostly cyclic or branched cyclic compounds, containing non-proteinogenic amino acids, small heterocyclic rings and other unusual modifications such as epimerization, methylation, *N*-formylation.

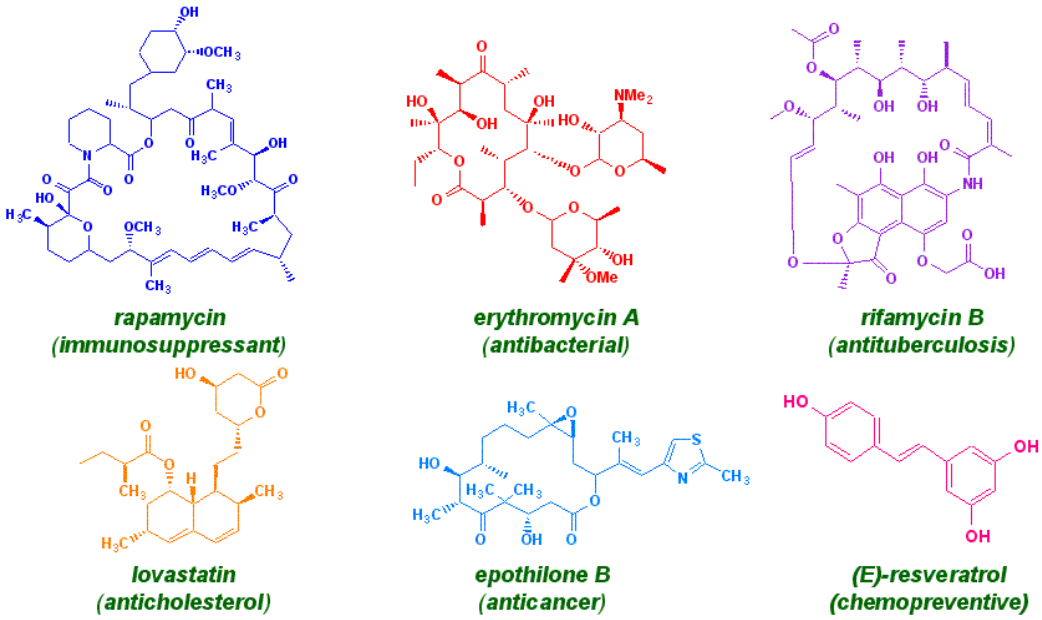


Fig. 1: Polyketide compounds and their function (picture from http://linux1.nii.res.in/~pksdb/polyketide.html).

These bioactive compounds are synthesized using polyketide synthases (PKSs) and nonribosomal peptide synthases (NRPSs). In these enzymes that are organized in large modular assembly lines each module incorporates a given monomer unit into the growing chain. This modular architecture is ideally suited for creating new assembly lines, producing new complex molecules with potential antibiotic, anti-cancer or anti-inflammatory functions. Actually, over two hundred novel polyketide compounds have been created by modul swapping, gene deletion or other specific manipulations; unfortunately engineered polyketide synthases (PKS) or nonribosomal peptide synthases (NRPSs) often fail to produce significant amounts of the desired products [4-6]. These results show that reengineering requires a detailed knowledge of the interaction between the individual domains. Recent investigations suppose that dynamic

processes and conformational switches play important roles for the communication between domains in nonribosomal pepdide synthases (NRPS) [7].

Decreasing effectiveness of antibiotics- The introduction of antibiotics into medical practice in the 1940s revolutionized human's ability to cure infectious diseases. Since their discovery, antibiotics have been extensively used for more than half a century in human and animal therapy and as animal growth promoters. Because each application of antibiotics inevitably selects for resistant bacteria, insensitive strains become more and more of a problem, especially pathogenic germs acquiring multiple resistances. Resistant strains were found against several antibiotics that are commonly used. According to the world health organization (WHO), in the last quarter of the last century organisms including *Haemophilus influenzae* and agents of sexually-transmitted infections, such as *Neisseria gonorrhoeae*, have emerged as worldwide multidrug-resistant threats. In the struggle of raising resistance, some antibiotics like vancomycin were regarded as last-resort antibiotics being efficient against already resistant pathogens like methicillin resistant *Staphylococcus aureus* strains. But the appearance of vancomycin-resistant *Staphylococcus aureus* in 1986 has raised concerns that even industrialized countries may be losing this antibiotic of last resort [8]. The only possibility to overcome the increasing number of resistant strains is the development of new agents with combinatorial diversity. Nevertheless, the number of new antibiotics introduced on the market decreased dramatically [9, 10]. Both trends: the increasing pathogen resistance and the decline in new agents necessarily require reinforced efforts to discover and develop new antibiotics.

Natural antibiotics, important elements of modern medicine- The ability to synthesize polyketides (PK) and nonribosomal peptides (NRP) that find application in modern medicine is widely spread among microorganisms. In organic chemistry these biosynthetic reactions are either too difficult, with low efficiency or too expansive to realize. To conclude: naturally occurring antibiotics have benefits due to structural features rarely found in synthetic libraries. Biochemical and genetic studies have unveiled the key principles of those nonribosomal peptide synthases (NRPS) and polyketide synthases (PKS). The modular organization of NRPSs and PKSs offers possibilities for a rearrangement of the biosynthetic pathways and therefore the possibility of new antibiotic biosynthesis [11].

4.1 The polyketide synthases (PKS)

4.1.1 Definition PKS I, II and III

The main domains in the PKS systems are the acyl carrier protein (ACP) or also called thiolation (T) domain, the acyl transferase domain (AT domain) and the ketosynthase domain (KS domain) which compose all together one minimal module and are responsible for the elongation of the growing acyl chain by one acylunit per module [12]. The mechanism and function of each domain will be described in detail later on (**4.1.3**). In the Type I PKS system these domains are connected in *cis,* the 6-deoxyerythronolide B synthethase (DEBS) is one of the best studied examples for the PKS I system. PKS I systems typically have one module for each round of chain elongation and modification. According to this definition, all so far characterised NRPS systems can be classified into Type I [13]. The Type II PKS systems have catalytic domains and carrier proteins on separate units, they interact in *trans*. Furthermore, there exists only one acyl carrier protein (ACP), one acyl transferase (AT) domain, and one ketosynthase (KS) domain which catalyze the elongation of the substrate in consecutive rounds. The chain grows then by two or three carbons in each cycle, tethered to the same ACP. KS and AT act then iteratively [14]. Some Type II PKSs in *Streptomyces* species are for example: actinorhodin, oxytetracycline, granaticin, frenolocin B [15]. Historically, those free standing ACPs were first detected in bacterial type II FAS (e.g. *Escherichia Coli*). Type III PKSs differ from Type I and II by their choice of substrate. They use acyl-coenzymeAs (acyl-CoAs) as activated monomers rather than acyl-S-pantetheinyl-ACP [16]. One example of a Type III PKS is the enzyme DpgA, which uses four malonyl-CoA molecules to produce 3,5-dihydroxyphenylacetyl-CoA (DPA-CoA) in the biosynthetic pathway to 3,5-dihydroxyphenylglycine (3,5-Dpg) for chloroeremomycin assembly [17]. This work focuses especially on the PKS I system.

4.1.2 General mechanism of biosynthesis

PKS are multimodular enzymatic assembly lines of a remarkable size up to 2 MDa [18]. Among several characterized PKSs, the 6-deoxyerythronolide B synthase (DEBS) has provided the paradigm for understanding the structure and function of the PKSs that are responsible for assembling complex polyketides [19]. In the PKS I assembly line each module incorporates one acyl unit into the growing acyl chain. Therefore the modules are also called extender modules. 6-deoxyerythronolide B synthase (DEBS) catalyzes the formation of 6-deoxyerythronolide (6-dEB). It consists of six of these extender modules. They are organized in doublets of extender modules in three large subunits with molecular weights above 300 kDa, called DEBS1, DEBS2 and DEBS3 (**Fig. 2**). Each extender module contains at least three essential domains: the 45 kDa ketosynthethase (KS) domain, the 50 kDa acyl transferase (AT) domain and the

10 kDa acyl carrier protein (ACP) also called thiolation (T) domain. The acyl transferase (AT) and ketosynthethase (KS) domains are the two catalytically active core domains. The AT domain selects the acyl units (either malonyl-CoA or methylmalonyl-CoA) as substrate and transfers the C3 or C4 acyl group to the ACP domain [20] (**Fig. 2**).

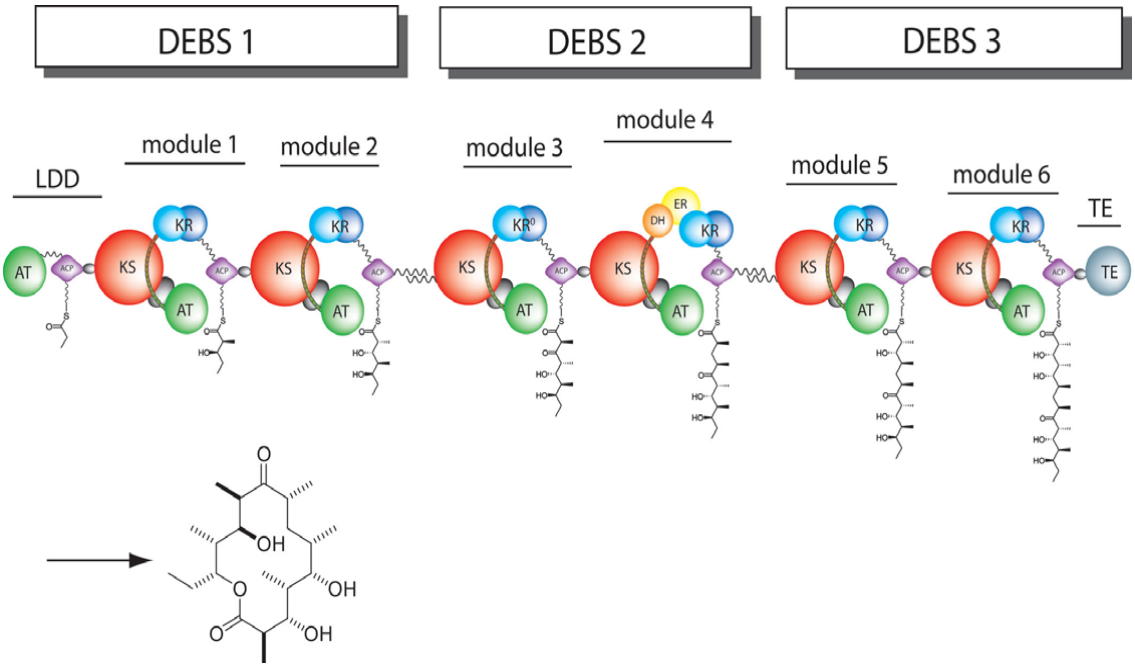


Fig. 2: Illustration of the 6-deoxyerythronolide B synthase (DEBS) PKS from the erythromycin biosynthetic pathway. DEBS is responsible for generation of the aglycone 6-deoxyerythronolide B (6-DEB). The domains are labeled according to the following abbreviations: Loading domain (LDD), Acytranferase (AT), Ketosynthase (KS), Acetyl Carrier Protein (ACP), Ketoreductase (KR), Dehydratase (DH), Enoylreductase (ER) and Thioesterase (TE) (from [21]).

The ACP needs to be post-translationally modified to be activated. Therefore the phosphopantetheinyl transferase (PPTase) attaches the cofactor 4'-phosphopantetheinyl (4'-Ppant) – which derives from coenzymeA (CoA) - to a highly conserved serine. After modification the ACP is transformed from its inactive "apo" to its active "holo" form [22] (**Fig. 3**).

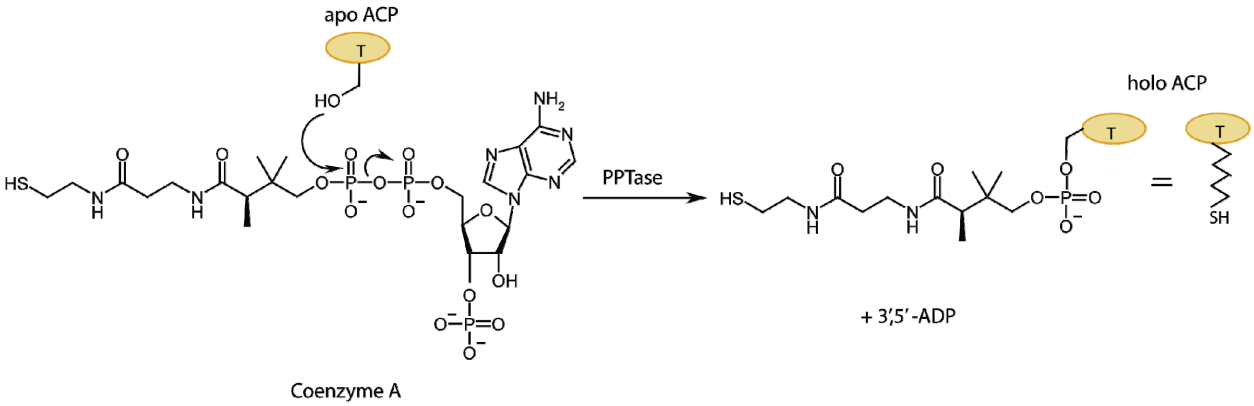


Fig. 3: Posttranslational modification of the ACP by the phosphopantetheineyltransferase (PPTase). Enzymes of this family transfer the phosphopantetheine (4'-Ppant) from coenzyme A (CoA) to a conserved serine of the ACP domain [22] (picture from [12]).

Substrates and intermediate acyl chains are then covalently linked to the thiolgroup of the 4'-phosphopantetheinyl (4'-Ppant) arm of the holo ACP, a reaction catalyzed by the AT domain.

The KS forms the C-C bond between the acyl groups bound to the up- and downstream located ACP. Therefore it elongates the growing acyl chain by two or three carbons [12].

4.1.3 Mechanism

The chain elongation

Polketides display a remarkable structural and functional diversity with respect to just a few simple building blocks: acyl-coenzymeA (acyl-CoA) thioesters such as acetyl-CoA, malonyl-CoA, methylmalonyl-CoA that derive from the pool of primary metabolites. The information for the polyketide product is encoded in the number, identity, and order of protein modules in the assembly line. The modules determine the monomers that are incorporated and the order in which they are condensed. The AT domain selects the appropriate carbon extender unit and transfers the units from acyl-CoA onto the phosphopantetheine (4'-Ppant) arm of the ACP (**Fig. 4**). This transthiolation step activates the substrate and is essential before any condensation can occur. Only starter units differ from the basic extender units, they can be as well the thioesters of propionyl-, and benzoyl-CoAs, or structural variants, such as malonamyl-CoA or methoxymalonyl-CoA [12].

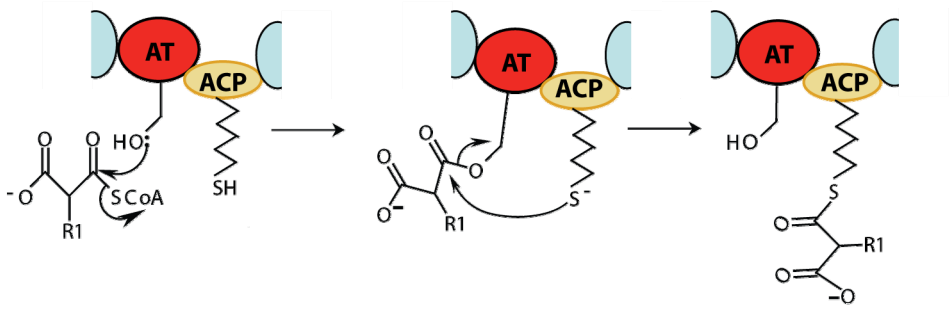


Fig. 4: The acyltransfer (AT) domain activity. During PK biosynthesis, the acyltransferase (AT) selects the appropriate acyl unit and transfers it to the ACP domain, leading to acyl-S-ACP (adapted from [12]).

The KS has two neighboring ACP domains, one upstream and one downstream. The acyl chain from the upstream located ACP domain gets transferred to a conserved serine of the KS domain [12, 23]. In the mean time the KS decarboxylates the (methyl)-malonyl thioester of the downstream ACP to yield the stabilized C2-carbanion (thioester enolate). The carbanion then attacks as a nucleophile in a claisen condensation the acyl-*S*-group which is bound via the serine-thiol to the KS (**Fig. 5**). The timed decarboxylation of the (methyl)-malonyl C3-carboxylate renders the C-C bondforming step irreversible. In this way in PKS the chain elongation is performed [12].

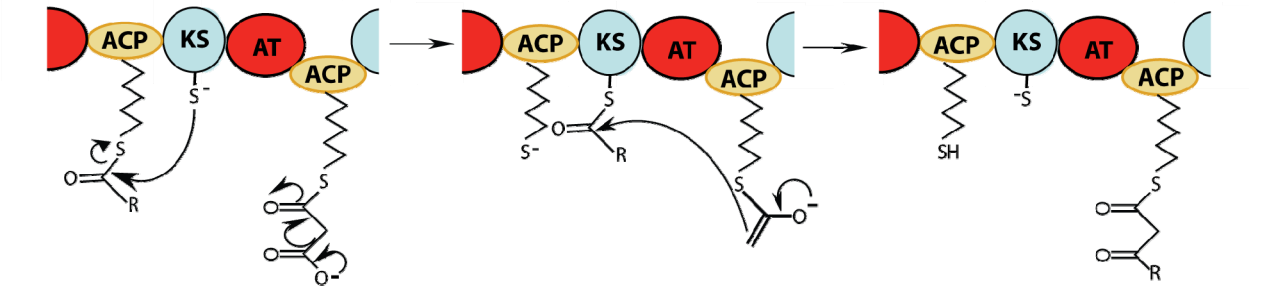


Fig. 5: The ketosynthase (KS) domain activity. During PK biosynthesis, decarboxylation of the downstream (methyl)-malonyl-S-ACP yields a nucleophilic thioester enolate. This attacks the upstream acyl-S-ACP thioester in a claisen condensation to form a C-C bond (adapted from [12]).

Further modifying enzymes

A deeper look at the DEBS (**Fig. 2**) biosynthesis uncovers further optional tailoring domains, including ketoreductases (KRs), dehydratases (DHs), and enoyl reductases (ERs). These domains act in the well established order KR → DH → ER. The tandem action of these three domains can perform the four electron reduction of β-C=O to the β-CH_2. The individual steps are the processing of the extender unit from a β-ketoacyl-S-ACP to yield a β-hydroxyacyl-S-ACP (KR action), followed by hydration to α, β-enoyl-S-ACP (DH action) and further on reducing to a acyl-S-ACP group (ER action) (**Fig. 6**) [20], [12].

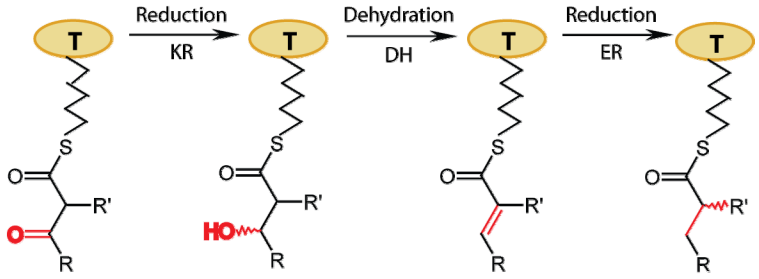


Fig. 6: Auxiliary PKS domains mediating ketoreduction (KR), dehydration (DH) and and enoylacyl reduction (ER) (adapted from [24] and [12])

These domains are unalterable in Type I and II fatty acid synthases (FAS) but can be mutated, unfunctional or deleted in Type I and II PKSs. In this case we obtain incomplete reduction giving raise to diverse chemical functionality and chemical reactivity [12]. Type I PKS systems contain starter units, so called initiation module and termination modules, which differ from the standard elongation modul and will be presented in the following.

Initiation module

Polyketide chain initiation is catalyzed by a “loading module” that, in all characterized type I PKSs, shares high similarity with the minimal extension module. A loading acyltransferase domain (AT_L) uses an acyl-CoA substrate to load an adjacent ACP domain (ACP_L). Typically, the AT_L substrate is α-carboxylated acyl-CoA (e.g.: malonyl-CoA), like the substrate for chain extension, and acyl transfer (e.g: actinorhodin PKS and tetracenomycin PKS) other systems might have preferences for methyl-malonyl. Substrate loading is followed by a decarboxylation

via a mutated "KS_Q" domain, which is catalytically inactive for condensation [25],[26],[27].

Alternatively, the KS domain may be missing entirely from the initiation/loading module at the N-terminus of the assembly line, like in the erythromycin PKS [25]. In that case, an acyl-CoA (propionyl- instead of malonyl- or methyl-malonyl-CoA) is selected by the initiating AT_L domain and loaded on the first ACP_L domain [20].

Termination module

The last module of a Type I PKS assembly line is responsible for the release of the full-length acyl chain from covalent tethering on the final ACP domain. This process of chain termination is typically catalyzed by a 35 kDa thioesterase (TE) domain, which can catalyze hydrolysis, generating the free acid (e.g. during *β*-lactam biosynthesis, or intramolecular capture by one of the –OH groups in the polyketide chain, generating a cyclic lactone) [28].

4.1.4 Structural insights

The ketosynthase (KS) and acyltranferase (AT) domain

As the model system of modular PKSs, DEBS has been studied genetically and biochemically over the past two decades. The crystal structure of a 194 kDa homodimeric fragment of the KS-AT didomain of DEBS module five has been determined by Chaitan Khosla's group at Stanford University [29]. This high resolution X-ray structure is the first didomain structure of a PKS module. It provides insights into the complex structural organization of the modular polyketide synthases machinery. The 194 kDa homodimeric fragment contains the full-length KS and AT domains as well as three flanking linkers: the N-terminal linker, an intervening KS-to-AT linker, and post-AT linker. The linker regions have been crucial for protein stability contrary to previous expectations that linkers only have a minor role and are mostly unfolded (**Fig. 7**) [29].

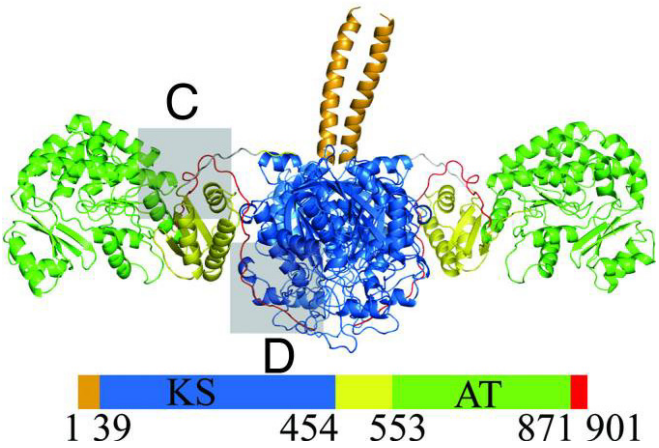


Fig. 7: X-Ray crystal structure of the ketosynthase and acyltransferase (KS-AT) didomain from DEBS module 5. The KS–AT protein forms a homodimer. In orange: N-terminal coiled coil linker domain; in blue: ketosynthase (KS) domain; in green, acyl transferase (AT) domain; in yellow: KS–to–AT linker; in red: post-AT linker peptide. The residues 458–465 lacked electron density and were therefore modeled and shown in gray. (C,D) Side view (Picture from [29])

The structure reveals that the KS domain adopts a αβαβα fold, the AT domain contains an α,β-hydrolase like core domain and an appended smaller ferredoxin like subdomain. The three linkers are also structurally well defined [29]. The KS-to-AT linker is formed by a three stranded β-sheet packed against two α-helices on one side, the post-AT linker wraps back over both the AT domain and the KS-to-AT linker so as to interact specifically with the KS domain. The N-terminal linker forms a coiled-coil structure. The crystal structure also reveals that the active site Cys199 residue of the KS domain is more than 80 Å away from the active site Ser642 residue of the AT domain. This distance is too large to be covered by the 4'-phosphopantetheine arm (4' Ppant) of the ACP domain measuring 18 Å, too large even when the 4'-Ppant arm is alternatively positioned and flexible. Thus, for domain-domain interaction, reorganization may be necessary for the ACP to interact successively with both the AT and the KS domains of this polyketide synthase module [21]. The overall structure is consistent with the 4.5 Å structure of the mammalian fatty acid synthase [29]. Similar to the SrfA-C structure of an NRPS system [30] these findings give a great insight into the overall domain structure; it also emphasizes the crucial role of non-conserved but structurally well defined linkers. But these findings do not give any insights into domain-domain interactions; those would be only possible if large interdomain movements would take place.

The acyl carrier protein (ACP)

General structure: The ACPs consist of 60-100 amino acids, their family is characterized by a four α-helices fold (I–IV) [31][32]. Each carrier protein has a conserved serine residue at the N-terminus of helix II within the conserved Asp-Ser-Leu (DSL) motife (**Fig. 8**). The four-helix bundle fold is conserved among ACPs from fatty acid synthases and polyketide synthases (**Fig. 9**). Nevertheless detailed comparison revealed that ACPs from polyketide biosynthetic pathways are more related to each other in tertiary fold than to their homologues from fatty acid biosynthetic pathways. PCPs and ACPs are known to share the same fold but they differ in their surface charge, ACPs are more acidic proteins [13].

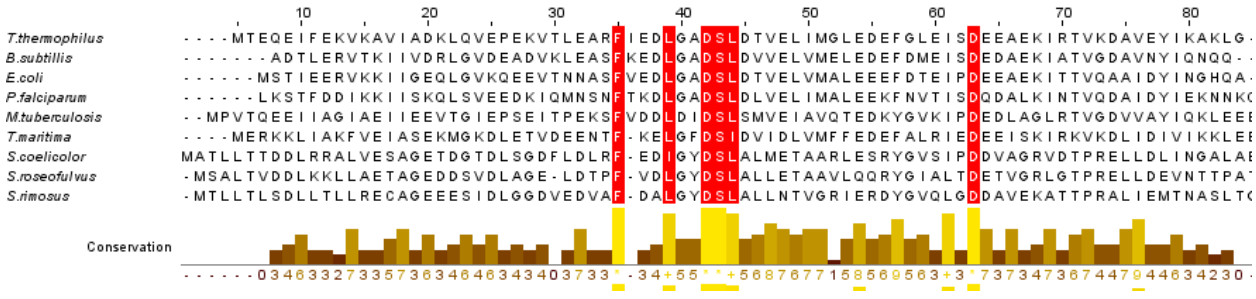


Fig. 8: Sequential alignment of acyl carrier proteins. The sequential alignment of ACPs with an averaged size of 80 amino acid residues from different organisms like the erythronolide ACP (DEBS) from *Saccharopolyspora erythraea, the* frenolicin ACP from *Streptomyces roseofulvus*, the oxatetracycline PKS ACP from *Streptomyces rimosus*. In red are indicated the highly conserved amino acid residues containing the conserved DSL motif. The alignment was generated with clustalW.

In nature the conserved serine residue in the DSL motife is subject to post-translational modification by acetyl carrier protein synthases (ACPS) [22]. ACPS belongs to the phosphopantetheinetransferases (PPtases) which transfer a 4'-Ppant moiety from coenzyme A (CoA) to the active site serine (**Fig. 3**). The activity of ACP depends on this conversion from the inactive apo to the active holo form. During the PKS biosynthesis different substrates are tethered to the thiolgroup of the 4'-Ppant arm. Studying the PKS and NRPS system means to understand the structural and functional relationship. The most prominent questions related to the ACP structure are as follows.

a) Is there a conformational switch between apo and holo form?

b) How do different substrates loaded on the 4'-Ppant arm modulate the ACP structures?

c) Is the ACP in the different states static or dynamic?

d) What is the impact on protein-protein interactions of the aspects mentioned above?

Reviewing the literature in the ACP field shows, that the answers to these questions strongly depend on the respective system. The individual questions will be discussed in detail in the following paragraph.

Conformational switch between apo and holo form- Evans et al. investigated the actinorhodin (act) ACP of *Streptomyces coelicolor* [33]. They solved high quality NMR structures of the holo and apo version. For the holo ACP chemical shift perturbations (CSPs) were observed in the 2D [^{1}H,^{15}N]-HSQC spectra for 10% of the peaks located on helix III and II- near to the 4'-Ppant attachment site. High resolution structures were calculated for the holo and the apo form and showed that the overall fold is the same.

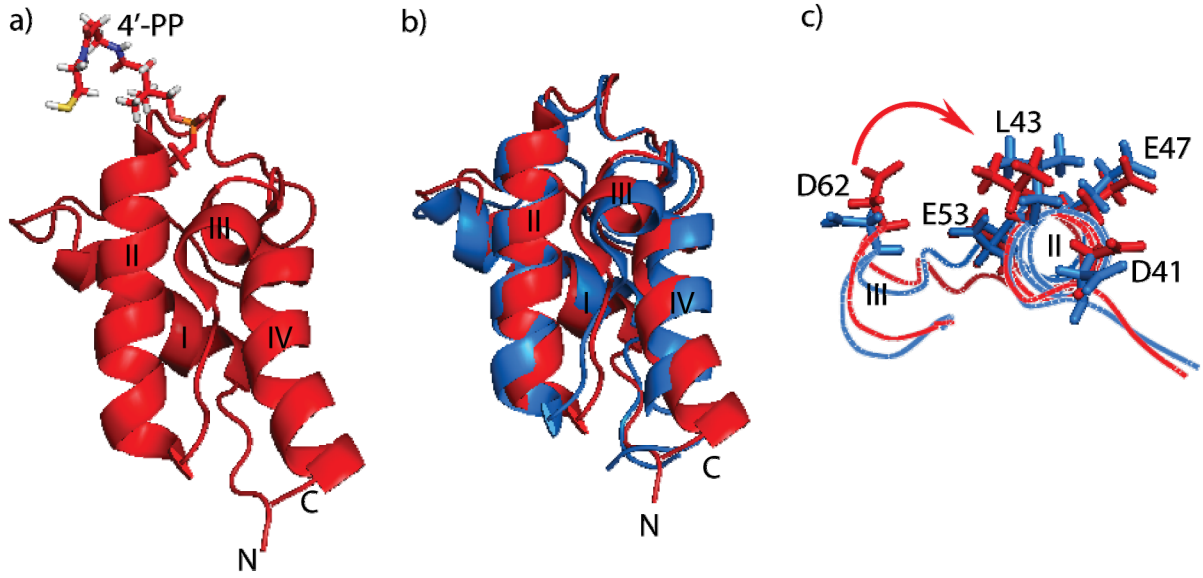


Fig. 9: The actinorhodin (act) ACP of *Streptomyces coelicolor* in its apo and holo state. (a) Cartoon view of holo act ACP from *Streptomyces coelicolor* (pdb: 2K0X). The phophopantheinyl arm (4'-Ppant) is presented in sticks. (b) Superimposition of the holo form (red) and the apo form (light blue; pdb: 2K0Y) which indicates the same overall fold. (c) Magnification of helix II and III of the superimposed structures with the detailed view of Asp41, Leu43, Glu53, and Asp62, indicating that helix III moves towards helix II for the holo form as Evan and colleagues acclaim (adapted from [33]).

Evans states that the difference between both structures is the movement of helix III towards helix II for the holo form which constricts the hydrophobic cleft [33](**Fig. 9**). Studies of the NMR holo ACP structures of the FAS ACP of *Bacillus subtillis* [34], FAS ACP of *Escherichia coli* [35] and the ACP of Type II FAS *Mycobacterium tuberculosis* [36] are in accordance with the act ACP studies in the fact that the prosthetic group is localized transiently at a hydrophobic pocket between helix II and helix III. But in all three ACP structural investigations NOE pattern did not reveal that the conformations of apo ACP and holo ACP would be different. Furthermore, in these cases no NOEs were found for the prosthetic group. This supposes that in case of any interactions of the 4'-Ppant with the vicinity, this being too transient for NMR detection. The researchers coincide that the 4'-Ppant arm has no strong interactions with the protein surface and switches back and forth between protein association and solvent exposure. Relaxation data for the *E. coli* FAS ACP support that the ACP conformation itself is quite flexible [35].

ACP structure modulation upon substrates loading- Previous NMR studies of different holo ACPs (FAS and PKS) such as: FAS ACP of *Bacillus subtillis* [34], FAS ACP of *Escherichia coli* [35], the ACP of Type II FAS *Mycobacterium tuberculosis* [36] and the PKS ACP of *Streptomyces coelicolor* [33] have shown that there are no strong interactions of the 4'-Ppant arm and the protein surface. No NOE's are observed which could precisely locate the 4'-Ppant arm. Roujeinikova and colleagues as well as Evans and colleagues, both studied ACPs in different ligand bound states. Interestingly both got slightly different results which might be related to the different systems which were studied and the different techniques used. Roujeinikova and colleagues studied by X-ray the *E. coli* FAS ACP in its buturyl bound state [37] and in its hexanoyl-, heptanoyl- and decanoyl-ACP bound states [38]. Interestingly for the buturyl ligand the electron density was quite weak and discontinous, which suggests that this part has only weak or no interactions with the ACP or that it is poorly ordered [37]. This is in a good agreement with Evans et al. For the acetyl, malonyl, 3-oxobutyl and 3,5-dioxohexyl ACP (PKS*, Streptomyces coelicolor)* they could only observe minimal or no interactions with the ACP [39]. This leads in both cases to the conclusion, that these kind of short length substrates are exposed to the solvent and in this way exposed to the interaction partners. Furthermore, the attachments of these short chain substrates do not affect the overall structure in both cases, for the *E. coli* ACP and for the *Streptomyces coelicolor* ACP. But helix II and III are slightly affected in their conformation. These effects are even more pronounced and different for the studied PKS and FAS ACPs for a growing chain length of the substrate.

For the FAS hexanoyl-, heptanoyl- and decanoyl-ACP of *E. coli* the analysis of the crystal structures reveal how the protein stabilizes the growing fatty acyl chain. For all these substrates the aliphatic head and the β-mercaptoethylamine moiety of the 4'-Ppant group is sequestered into the hydrophobic cavity in the core of the four-helix bundle. This results in the shielding of the thioester bond from the solvent and presumably prevents hydrolysis. The hydrophobic cavity

to which the substrate is bound in a linear fashion is formed like a tunnel parallel to helix αII and is narrowed to its end by helix IV [38] (**Fig. 10 a) b)**). Interestingly their analysis reveals that upon binding of acyl chains of up to ten carbon atoms a conformational rearrangement of three amino acids (Phe28, Ile54 and Ile72) occurs. This local rearrangement affects a small angular displacement of helix III leading to an expansion of the hydrophobic cavity. This structural plasticity allows a ligand-length dependant variation of the volume of the cavity of the FAS ACP of *E. coli* [38]. Studies on the spinach ACP also demonstrate that its fold undergoes only subtle further changes on binding of a longer (C18) acyl chain which induces large change in the size of the cavity (from 128 $Å^3$ to 305 $Å^3$) which can even accommodate longer acyl chains [31]. This was supposed not to be possible for the *E. coli* ACP [38]. Upon a growing acyl chain (butyryl-, hexanoyl-, and octanoyl-ACP) Evans and his colleagues could observe that in each case the acyl chain is clearly bound in a cleft which is also located between helices II and III of the PKS ACP. They see as well, in agreement with Roujeinikova and colleagues that the overall structures are superimposable for helix I, II and IV. Helix III is the only structural element which changes significantly, it moves away from helix II. Interestingly this movement results in a cavity which runs perpendicular to that observed in FAS ACP (**compare Fig. 10 a) and c) and b) and d)**). It is also broader and more open then in the FAS ACP. The difference becomes even more apparent in the way the substrate is bound to the cavity, the 4'-Ppant arm remains predominantly solvent exposed which results in a solvent exposed thioester bond [39](**Fig. 10**).

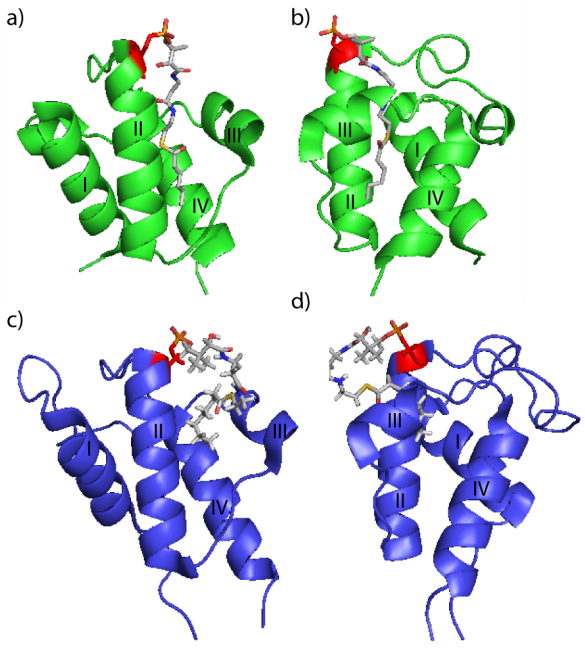


Fig. 10: Comparison of the substrate localization. (a) Cristal structure of the FAS hexanoyl ACP of *E. coli* (PDB: 2FAC [38]) presented in the front view b) the same protein rotated by 90°. c) Presentation of the NMR structure of the actinorhodin polyketide synthase octanoyl-ACP from *Streptomyces coelicolor* (PDB: 2KGC; [39]) in the front view d) the same protein rotated by 90°. In the FAS hexanoyl-ACP structure the substrate is embedded in a tunnel build between helix II and III. The substrate aligns in parallel to helix II. In the structure of the PKS act ACP the headgroup is bound into a cavity created between helix II and III but the 4'-Ppant arm itself stays solvent exposed.

Summarizing these two investigations we can say, that the Type II FAS ACP of *E. coli* binds their fatty acyl intermediates in a narrow cleft that runs parallel with helix II and helix IV. In contrast the binding cleft in the Type II PKS ACP of *Streptomyces coelicolor* runs perpendicular to helix II. Three specific key amino acid substitutions between *E. coli* FAS and act PKS ACP, namely, Thr39/Leu45, Leu42/Thr48, and Ala68/Leu74, alter the size, shape, and location of this cavity. This may contribute to the differential binding modes which were identified [39].

Dynamics of ACPs in different chemical states- Functional flexibility is often accommodated through conformational flexibility. This has been previously shown for the TycC3 PCP from the NRPS system. The PCP exists in its apo form in two different states, A and A/H. For the holo form it exists in a new equilibrium between the A/H conformation and a new conformation, the H state. The interaction with the PPtase SFP specifically takes place when the PCP is in the A state [7]. The enzyme thioesterase II (TEII) becomes active when the PCP is loaded with the wrong substrate-acetyl CoA (about 80% of the CoAs are acetylated). This substrate locks the PCP in its H state, the conformation that has shown to be the preferred one for TEII interaction [40]. So far a similar conformational exchange could not be shown for ACPs. One exception is the FAS spinach ACP, two distinct sets of peaks could be observed in a [^{1}H,^{15}N]-HSQC spectra for the holo form representing presumably one folded and one unfolded state. It could be shown that this conformational heterogeneity is reduced upon substrate loading with decanoate-(10:0) and stereate-(18:0) leading only to a few, neglectable double peaks in a [^{1}H,^{15}N]-HSQC [31]. Furthermore a conformational exchange between a mayor state (85%) and a minor state (15%) could be observed as well for the 3-oxobutyl loaded act ACP from the PKS system of *Streptomyces coelicolor*. The slight differences between those conformations are located on helix II, helix III and the connecting loop [33]. Similar to these investigations are the results obtained for the holo form of the frenolicin (fren) PKS ACP. Four residues, located in Loop two and helix III, were identified to exchange slowly on the NMR timescale (k_{ex}<103 s-1). Additional relaxation studies reveal further that loop I as well as the 4'-Ppant arm have low order parameters indicating that they undergo rapid and large amplitude motions [41]. A third PKS Type II example for a conformational exchange is the apo ACP from *Streptomyces rimosus*. Relaxation experiments have shown that the loop between helix I and II is flexible and that helix II, which is proposed as a site of protein-protein interaction, shows conformational exchange. These results are in agreement with the previous ones [42]. For most of the holo ACPs investigated, no NOE could be detected between the ACP and the 4'-Ppant arm. A fact which changes with the rising length of the acyl chains, longer chains lead to a fixation of the 4'-Ppant arm. An exception to this rule seems to be the Pf ACP from the Type II FAS of *Plasmodium falciparum*. Two conformations could be observed which are in a slow exchange with each other. For both conformations different NOE patterns could be derived exhibiting a difference in the orientation of the 4'-Ppant arm. The 4'-Ppant arm orientation is

accompanied by the loss of a minor helix II' (between helix II and helix III) giving rise to a longer loop connecting helix II and III [43].

ACPs and their protein-protein interactions- Many key domain-domain interactions involving carrier proteins are required during the production of the PKS or NRPS scaffold. For a typical cycle of PKS elongation, as described before, an ACP must interact with the AT domain, the upstream KS domain, eventually further modifying enzymes that perform tailoring of the resulting *α*-ketone and the downstream KS domain. All these events must take place in a well-timed manner. Leading to the question how these diverse interactions can be coordinated? How much of a carrier proteins surface is involved in each of these inter-domain interactions? Does each domain contact a separate face of the carrier domain? Is the swinging of the 4'-Ppant arm directed by the catalytic domains or through the carrier protein itself? It has been shown that for the *E. coli* FAS the substrate which is bound to the ACP is sequestered within the carrier. How does this way of transport enables individual FAS enzymes to discriminate between the different intermediates (e.g. β-keto, β-hydroxyl, acetoacetyl-, malonyl-ACP)? One major factor which makes these studies challenging is the high amount of protein often needed for structural studies. Another drawback for NMR studies is the often transient and weak nature of these interactions, in particular those involving the ACP. Efforts to structurally characterize these interactions via X-ray crystallography are limited due to the highly dynamic character of those interactions [44]. Nevertheless much research has been done and this chapter will try to summarize the insights obtained by NMR, X-ray, computer simulations and docking experiments and results obtained via biochemical approaches.

Phosphopantetheine transferases are essential to activate the ACPs in PKS systems and PCPs in NRPS systems. NMR investigations could enlighten the activation process for PCPs by SFPs and regeneration of misprimed PCPs by SrfTEII- all *cis*-acting enzymes. The apo TycC3 PCP exists in a conformational equilibrium called state A and A/H. SFP interactions take exclusively place with apo TycC3 in its A state. Titration experiments could narrow down the interactions to residues 30 to 55 of helix αII and the 4'-Ppant recognition site within the loop between helix I and II [7]. Furthermore it has been shown that the PCP interacts in its H state with SrfTEII. SrfTEII is a repair enzyme which removes acetyl groups from misprimed ACPs. This interaction could be localized on a long loop between helix II and IV. Interestingly, this loop can form a helix (helix III) in the inactive A/H state and after unfolding it is essential for the recognition by SrfTEII. The mechanism can therefore be interpreted as an allosteric regulation [7]. As discussed before for ACPs it was so far impossible to observe the same conformational plasticity [13]. Furthermore the 4'-Ppant arm is solvent exposed and flexible and not attached to the surface as for the TycC3 PCP [7]. It was possible to co-crystallize holo ACP from *B.subtillis* with ACPS and the result showed that the contacts between holo-ACP and ACPS are predominantly hydrophilic in nature with almost all of the interactions occuring between helix α1

of ACPS and helix III of ACP [45]. Due to different nomenclature this helix III corresponds to helix II in most other ACP structures which locates the active serine on its N-terminus. Interestingly SFP can modulate ACPs as well as PCPs and vice versa. Nevertheless for the ACPS it was shown that ACPs are the better substrates, hypothetically due to the different surface charge and the alteration of the electrostatic interactions [46]. Based on the crystal structure of the KS3-AT3 didomain from DEBS module 3 from Yeast FAS a docking for ACP3 and ACP2 was performed leading to the hypothesis that both interactions are also mediated via helix II [21]. In triketide lactone formation assays it could be shown that mutations on helix II of the different ACPs in DEBS have an effect on the specificity of the system. For example it was possible to change ACP2 by one single mutation (R59G) to be a bad interaction partner for KS-AT of module 2 but a good one for KS-AT of module 6 [47]. Docking experiments with the *E. coli* FAS proposed that FabH (a β-ketoacyl-ACP Synthase III) and ACP mainly interact via helix II [48]. Furthermore, it was shown that helix II plays also a major role in mediating the interactions with β-ketoacyl-acyl carrier protein reductase (FabG) [49]. In addition there is direct evidence from X-ray for electrostatic interactions between helix II of dodeconyl-ACP and FabI [50]. A region which has shown to be especially flexible; maybe due to electrostatic repulsion (14 acidic and no positive residues between positions 30 and 60) [51] [52] [53]. This flexibility is thought to be important in facilitating rapid association and dissociation of ACP with its various partner enzymes. In conclusion jelix II can be called a “recognition helix” for universal enzyme interaction [49]. Helix II could be further confirmed to be such a “recognition helix” for interactions of the FAS ACP from *Vibrio harveyi* with holo-ACP synthase (N-terminus of helix II, Asp35), acyl-ACP synthase (via C-terminus of helix II) and UDP-*N*-acetylglucosamine acyl-transferase (LpxA) (also via C-terminus of helix II) [54]. Via NMR titration experiments it could be proven that the negatively charged helix II from the FAS ACP of *Steptomyces coelicolor* interacts with malonyl-coenzymeA-ACP transacylase (MCAT) through the negatively charged helix II of ACP [55]. Finally interaction studies have been performed in the NRPS cluster producing enterobacterin. The interaction surface on EntB-ArCP for EntF consisted of one residue on helix II (M249) and two residues on helix III (F264 and A268). The observations support a model in which the downstream recognition is mediated by helix III which might work due to its structural flexibility and functional ability as a conformational switch [13]. All in all the research on interactions in PKS, FAS and NRPS systems maps interactions on helix II, whereas we have shown an additional conformational hot spot on helix III within PCPs from NRPS systems. Additionally, many interactions seem to be transient and have been assumed not to alter the structure significantly [33, 35, 36, 41].

4.1.5 Quaternary structure and the role of docking domains

A common feature of NRPS and PKS systems is that the modules are distributed among multiple protein subunits. As a consequence the intermodular transfer has to take place

between domains located on different polypeptides [56]. The formation of the correct product depends therefore not only on a specific ordering of the multienzymes, it must also prevent improper associations (**Fig. 12**)[57]. It has been shown, that the communication between different modules is linked to interactions between short linker regions (30-90 amino acids) located at the end of a module in both, NRPS and PKS systems. In NRPS systems these domains are called “communication domains” (COM-domains), in PKS domains these “interpolypeptide linkers” are called “docking domains” [57][58]. NRPS enzymes appear to be mostly monomeric and different polypetides associate via the COM-domains [59][58]. COM-domains are significantly shorter than the respective C- and N-terminal docking domains of PKSs, approximately 15–25 in comparison to 20–30 amino acids. They exhibit little mutual sequence conservation. Domain swapping experiments in the tyrocidine NRPS could clearly demonstrate that matching pairs of COM domains are both necessary and sufficient for the establishment of communication between partner NRPSs in *trans* [58,60]. In contrast to NRPS proteins, PKS proteins are functional homodimers [61,62]. Their association with other subunits is more complex. In the DEBS system it was shown that the interpolypeptide linkers- the docking domains- differ from the covalent linkers who link two domains with each other. The docking domains are hydrophobic whereas the polypeptide linkers are more hydrophilic [56]. NMR structural studies of the interpolypeptide linker in the erythromycin PKS system (DEBS) showed that these domains are α helical structured (**Fig. 11**). Interhelical contacts play a crucial role for the formation of homo and heterodimers and different recognition sequences promote the right positioning of each module [57].

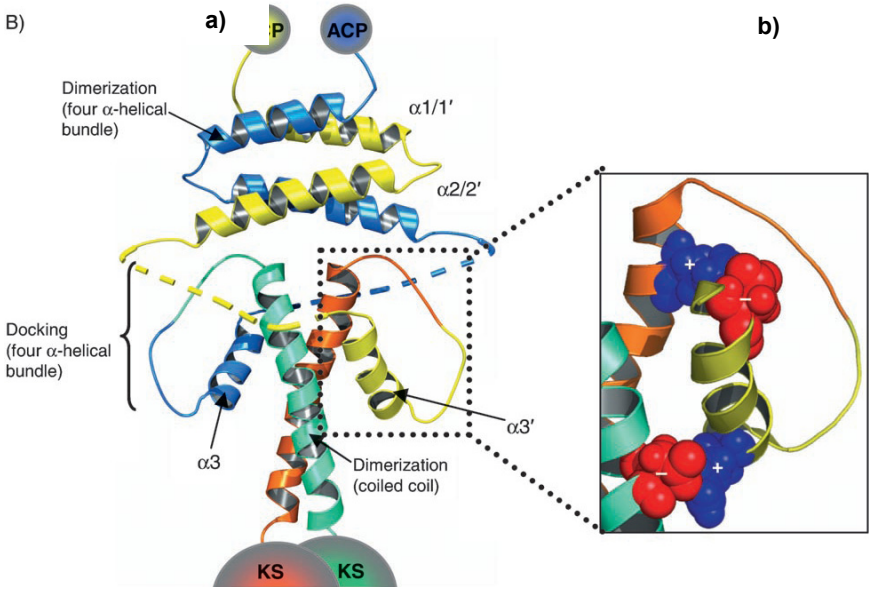


Fig. 11: Model of the structural basis for docking in modular PKS. a) Representation of the NMR solution structure of the DEBS docking complex. The dimeric C-terminal docking domain is shown in blue and yellow (three helices), while the dimeric N-terminal docking domain is shown in green and orange. The linker region between helices 2 and 3 is highly mobile, and therefore is represented as a dashed line in the structure. b) Magnification of the charged residues located at critical positions in the interface (from [1]).

The implementation of different docking domains assure the right positioning of the different subunits as drafted for the complete DEBS system (**Fig. 12**)[57]. With these different quaternary architectures for NRPS and PKS proteins, it has been unclear how molecular recognition should occur in mixed PKS and NRPS systems. The structural investigation of the N-terminal docking

domain of TubC (‘TubCdd’) from the tubulysin PKS-NRPS gives insight into a αββαα fold. These interactions seem to take place via a corresponding and specific electrostatic code [18][63].

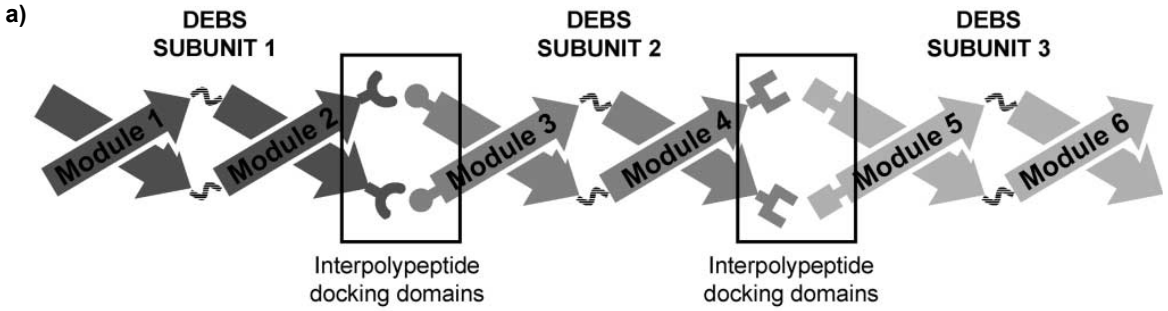


Fig. 12: Schematic organization of the erythromycin polyketide synthase. a) The three DEBS subunits are homidimeric. The individual polypeptide linkers, so called docking domains are twisted around each other in a “double-helical” structure, they assure the right positioning of each subunit (from [57]).

TubCdd is a homodimer and there is a reasonable assumption that the parent protein also oligomerizes. It might be that in hybrid PKS-NRPS systems also the NRPS part builds a homodimer, with the docking elements actively contributing to subunit self- association [18].

4.2 The curacin cluster

Curacin was originally purified as a major lipid component of the cyanobacterium *Lyngbya majuscula*, isolated in Curaçao. Curacin A has proven to be a potent inhibitor of cell growth and mitosis by binding rapidly and tightly at the colchicine site of tubulin. Curacin A contains a unique combination of functionalities in comparison to many other colchicine inhibitors including a *cis*-vinyl thiazoline heterocycle, a terminal alkene and cyclopropyl ring [64] (
Fig. 13).

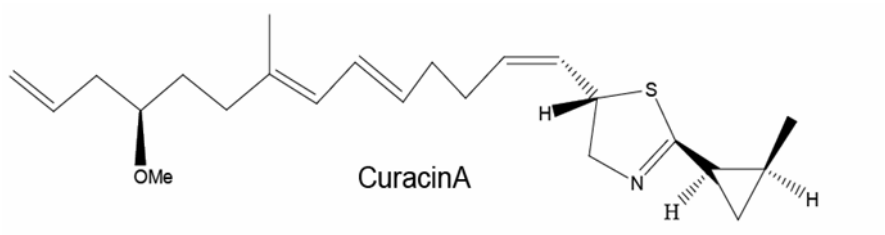


Fig. 13: The structure of curacin A. The unique combination of functionality is due to a *cis*-vinyl thiazoline heterocycle, a terminal alkene and cyclopropyl ring [65].

Six years later, in 2004, the first publications were made by Professor Sherman and his group investigating the biosynthesis of this putative cancer cell toxin. Feeding experiments allowed determining the metabolic origin of the different carbon atoms. Further cloning procedures of the 64 kb operon were undertaken in order to assign structures and functions to the individual domains [65]. The combination of bioinformatics, enzymological studies and feeding experiments allowed identifying the different enzymes in the Curacin A assembly line in the preceeding years (**Table 1 [next page]**).

Table 1: Description of the protein sequences encoded by the curacin biosynthetic gene cluster. Proteins of which the structure is solved are indicated in bold (adapted and modified from [65])

Protein	AA	Domains	Publication	Similar to Organism Id/Sim	Accession
CurA	2311	GNAT, ACP, KS, AT, **Hal**, ACP, ACP, ACP PKS	[65] [27] [66] [67]	StiC (*Stigmatella auratiaca*) 50%/ 67%	CAD19087
CurB	79	ACP		(*Bacillus subtilis*) 53%/ 72%	NP570904
CurC	408	PksF		(*Bacillus subtilis*) 40%/ 56%	NP389594
CurD	442	HCS HMG-CoA synthase		FAB-H-like condensation 44%/63%	AAN85526
CurE	254	ECH_1 enoyl-CoA hydratase/ isomerase	[68] [69]	enoyl-CoA hydratase/carnithine racemase (*Nostoc punctiforme*) 35%/53%	A48956
CurF	3195	C- term: **ECH_2**, ER, KS, AT N- term: UNK, ACP, C, A, PCP	[68] [70] [69]	C- term: mixed PKS/NRPS BarG *(L. majuscula*) 56%/71% N- term: UNK, ACP, C, A, PCP (*T. erythraeum*) 49%/68%	AAN32981 ZP00074378
CurG	1583	KS, AT KR, **DH**, ACP	[71]	CurG PKS NosB (*Nostoc* sp. GSV224) 51%/68%	AAF15892
CurH	2199	KS, AT, ER, **DH**, KR, ACP	[71]	StiF (*Stigmatella auratiaca*) 40%/56%	CAD19090
CurI	1656	KS, AT, **DH**, KR, ACP	[71]	StiB (*Stigmatella auratiaca*) 44%/60%	CAD19086
CurJ	2298	KS, AT, MT, **DH**, KR, ACP	[71]	StiC (*Stigmatella auratiaca*) 46%/64%	CAD19087
CurK	2232	KS, AT, ER, **DH**, KR, ACP	[71]	StiF (*Stigmatella auratiaca*) 40%/58%	CAD19090
CurL	1956	KS, AT, MT, KR, ACP		StiE (*Stigmatella auratiaca*) 38%/56%	CAD19089
CurM	2147	KS, AT, KR, ACP, **ST**, TE PKS/TE	[66]	StiB (*Stigmatella auratiaca*) 43%/60%	CAD19086
CurN	341	Hydrolase haloalkane dehalogenase		Hydrolase haloalkane dehalogenase 50%/64% hydrolase 39%/58%	P59336 NP820220

4.2.1 The mechanism of the curacin cluster

Curacin became a prominent example for a mixed PKS-NRPS system. So far different steps of the biosynthesis have been explored biochemically and will be presented in the following Chapters.

4.2.2 The initiation module

The initiation module is a loading tridomain. Similar loading modules were found in pederin [72], its structural analogs onnamide and theopederin [73], myxovirescin A [74], and rhizoxin [75]. These chain initiation modules are a structured, N-terminal regions spanning ~180 amino acids which code for three domains. There is an adapter (AR) domain which does not have any catalytical function but increases the stability and the efficiency of the system via protein-protein interactions. The catalytical active domain is the GCN5-related N-acetyltransferase (GNAT) domain with a bifunctional decarboxylase/S-acetyltransferase function. In a first step a decarboxylation of malonyl-coenzyme A (malonyl-CoA) to acetyl-CoA is performed followed by a direct S-acetyl transfer from acetyl-CoA to load an adjacent acyl carrier protein domain (ACP_L)- the third domain of the tridomain [27](**Fig. 14**).

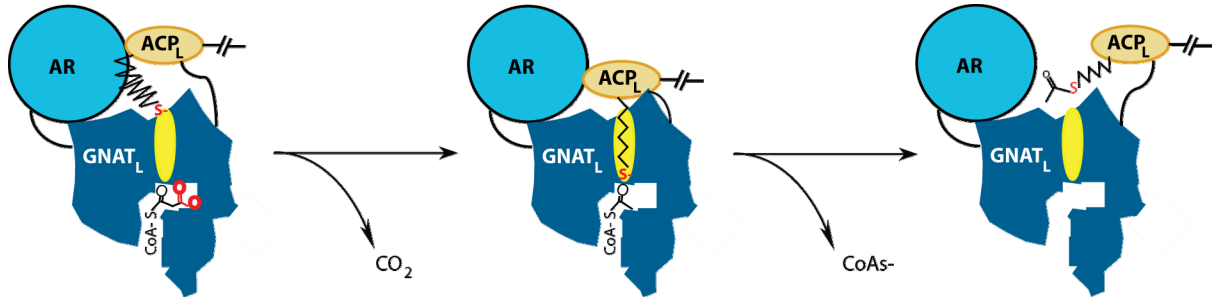


Fig. 14: Proposed mechanism of the CurA AR-GNAT$_L$-ACP$_L$ chain initiation module. Malonyl-CoA enters the CoA-binding tunnel (white tunnel) of the GNAT$_L$ domain (dark blue region) and GNAT$_L$ catalyzes decarboxylation to acetyl-CoA. The AR domain (light blue circle) directs the 4'-Ppant arm of ACP$_L$ into the ACP tunnel (yellow tunnel) for subsequent acetyl-group transfer via trans-thioesterification (picture adapted from [27]).

Formation of the cyclopropane moiety

A 10-enzyme assembly catalyzes the formation of the cyclopropane moiety. These catalytic domains and discrete enzymes span from CurA to CurF and are grouped into three subsets: the (1) halogenase (Cur Hal) embedded in CurA; (2) 3-hydroxy-3-methylglutaryl (HMG) enzyme cassette containing a tandem acyl carrier protein (ACP) tridomain ($ACP_{I,II,III}$) including ACP_I, ACP_{II} and ACP_{III}, a discrete ACP CurB, a ketosynthase-like enzymes (KS) CurC , the HMG-CoA synthase-like enzyme (HCS) CurD, the dehydratase (ECH_1) CurE, and the decarboxylase (ECH_2) embedded in CurF. Further on, there exists (3) the enoyl reductase domains (ERs) embedded in CurF. All these different catalytic domains are active when the substrate is tethered to the triplet $ACP_{I,II,III}$ [76](**Fig. 15 a)**). In the first steps of the biosynthesis the HMG cassette (malonyl-ACP_{IV} [malonyl-CurB], KS [CurC] and HCS [CurD]) introduce a β-branching and leads to the formation of HMG-ACP as illustrated (**Fig. 15 b)**). In detail, CurB is loaded with

malonyl by an AT domain, CurC a KS inactive for condensation, decarboxylates malonyl and activates in this way the carboxyl-group. HCS (CurD) catalyzes then the condensation of the CurB bound acetate unit with acetoacetyl- bound to ACP_I to form (*S*)-3-hydroxy-3-methylglutaryl-ACP (*S*-HMG-ACP) [65].

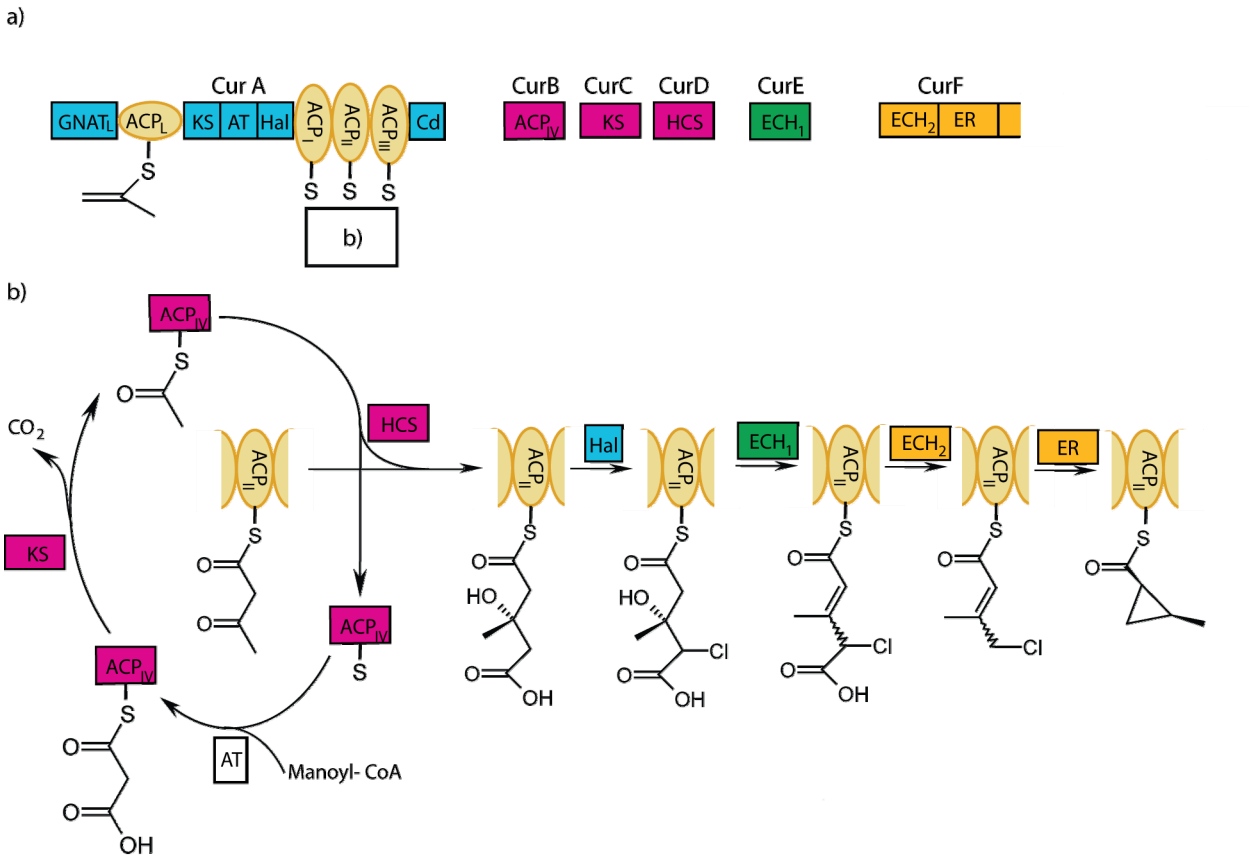


Fig. 15: The 10-enzyme assembly catalyzing the cyclopropane ring formation. a) The 10 enzymes which are involved in the cyclopropane ring are encoded on different proteins (CurA-CurF). b) Representation of the Biosynthesis of the cyclopropane formation. $GNAT_L$= loading module, KS= ketosynthethase, AT= acyltransferase, Hal= halogenase, ACP= acetyl- carrier- protein, HCS= HMG-CoA synthase-like enzyme, ECH_1= dehydratase, ECH_2= decarboxylase, ER= enoyl- reductase (adapted and modified from [69] and [76]).

The following reactions can be performed in different orders but as discussed by Gu and his colleagues the chance is quite high that the HMG formation is first followed by the chlorination step obtaining (*S*)-4-chloro-3-hydroxy-3-methylglutaryl-ACP_I via Cur Hal. As shown previously; ECH_1 catalyzes then the dehydration of 4-chloro-3-hydroxy-3-methylglutaryl-ACP_I to 4-chloro-3-methylglutaconyl-ACP_I, followed by ECH_2 decarboxylation to generate 4-chloro-3-methylcrotonyl-ACP_I. This is then the presumed precursor for (1*R*,2*S*)-2-methylcyclopropane-1-carboxyl-ACP_I (**Fig. 15 b)**) [69, 76].

The termination modul

In the termination module the sulfotransferase (ST) is coupled with a thioesterase domain (TE) domain. The catalytic events transform the β-hydroxyl of the penultimate chain elongation intermediate into a β-sulfate. B-sulfate is an excellent leaving group that is positioned chemically to facilitate decarboxylative elimination in the presence of the terminal carboxylate following TE-

mediated hydrolysis of the acylthioester. In this way the decarboxylative chain termination forms an unusual terminal olefin in the final product [66].

4.2.3 Structures of the curacin cluster

Initiation modul- the GCN5-related N-acetyltransferase (GNAT) domain

The crystal structures of $GNAT_L$ (CurA residues 219 to 439) could be solved at 1.95 Å (ligand-free form; PDB: 2REE) and 2.75 Å (acyl-CoA complex; PDB: 2REF) and is monomeric. $GNAT_L$ has the GNAT superfamily fold and consists of a central, mostly antiparallel β-sheet, flanked by α-helices, and is according to Gu and his colleagues most similar to serotonin N-acetyltransferase [27]. Due to the difference of the acyl and the apo form it could be shown that there are two distinct substrate tunnels for acyl-CoA and holo-ACP_L binding (**Fig. 16**). Further on it could be demonstrated by site-directed mutagenesis that His389 and Thr355, at the convergence of the CoA and ACP tunnels, participate in malonyl-CoA decarboxylation but not in acetyl-group transfer. Amino acids involved in the acetyl-group transfer could not be identified. These data led to a proposition for the mechanism of the biosynthesis initiation (**Fig. 16**) (compare **4.2.2**).

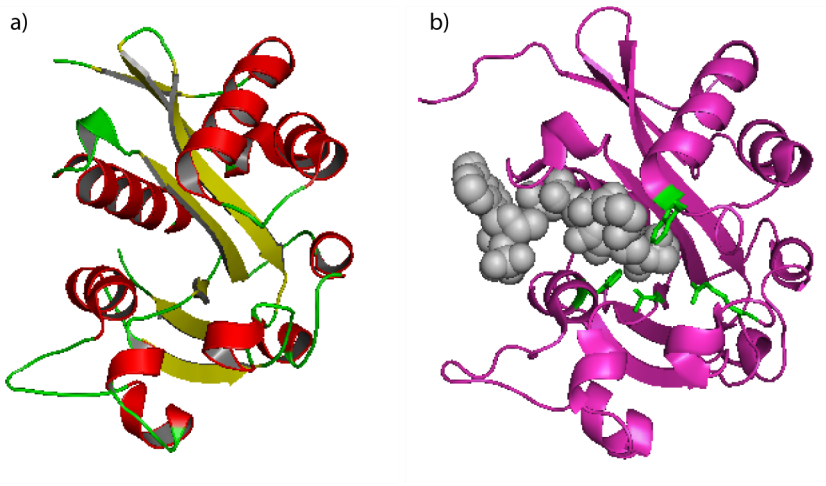


Fig. 16: The $GNAT_L$ domain a) Structure of $GNAT_L$ domain (PDB: 2REE). b) Structure of GNA_L with acetyl CoA tunnelling the protein (grey spheres) (PDB: 2REF). Residues Trp249, His389, Thr355, and Arg404 are shown (green-colored carbons).

The dehydratase domain (ECH_1)

ECH_1 is a dehydratase domain which catalyzes the formation of an α,β-double bond in the nascent polyketide intermediate. So far the structure of ECH_1 itself could not be solved due to the low solubility [71]. But the crystal structures of four other dehydratase domains present in the curacin cluster could be solved [71](PDBs entries: CurF-DH 3KG6; CurH-DH 3KG7; CurJ-DH 3KG8; CurK-DH 3KG9). The dehydratase domains (DH) are in solution as well as in the crystal structure dimers. DH domains have a double hot-dog fold in which the active site contains a histidine from the N-terminal hotdog and an aspartate from the C-terminal hotdog building together the active site [71, 77, 78](**Fig. 17**). Similarities between the four structures

and the reported structure of the Ery4 DH domain suggest that loading and unloading of intermediates is likely facilitated by reordering of mobile elements opening the active-site tunnel. It is reported that ECH_1 accepts HMG-CoA as substrate but has a 10* fold higher affinity towards the ACP bound substrate [68][69].

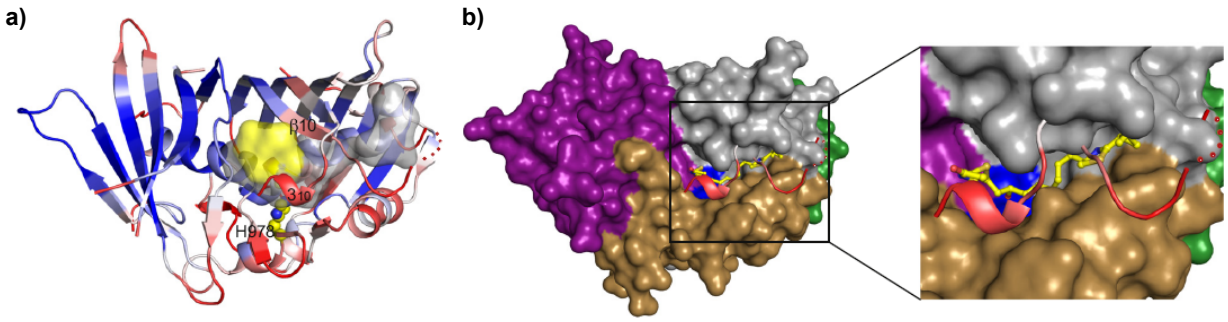


Fig. 17: Presentation of CurJ dehydrogenase (PDB 3KG8) a) The CurJ monomer is colored according to average atomic temperature factor per residue (B- factor). The lowest B- factors (blue) occur in the N-terminal hotdog fold, which forms the dimer interface. The highest B-factors are in residues covering the active- site tunnel (the loop connecting helix αHD2 and strand β10, and the C-terminal 3_{10} helix) and also in the linker connecting the two hotdog motifs. The substrate tunnel is colored in grey, and the active-site histidine is shown as spheres with yellow. b) Presentation of a surface diagram of CurJ DH monomer without the tunnel cover. The residues with a high B-factor (αHD2—β10 loop and C-terminal 3_{10} helix, shown in a) in red) were removed to allow visibility of the active site. The surface is colored by the structural motif (N-terminal hot dog in magenta; C-terminal hot dog in grey and the linking regions in brown and green). The CurJ DH substrate is modeled as sticks with yellow. Further on there is a more detail inset with the same orientation (Picture and description from [71]).

The decarboxylation domain (ECH_2)

This decarboxylase is encoded at the N-terminus of CurF. The N-terminal domain is a multifunctional protein which catalyzes decarboxylation of 3-methylglutaconyl-ACP to 3-methylcrotonyl-ACP. The product of the reaction is the postulated precursor of the cyclopropane ring of curacin A. The crystal structure of the CurF N-terminal ECH_2 domain establishes that the protein is a crotonase superfamily member. CurF ECH_2 has less than 20% sequence identity to identified members of the superfamily and the superfamily membership could only be established via the structure itself [70]. Reason is, that the crotonase superfamily consists of a wide variety of mechanistically diverse enzymes that exhibit various activities, such as hydratase [79], dehalogenase [80], decarboxylase [81][82], isomerase [82][83][84,85] (PDB 2F6Q), hydrolase [82], and carbon-carbon bond forming [82] or cleavage activity [86,87]. Due to this high functional diversity the superfamily lacks conserved catalytical residues. The crystal structure of ECH_2 is trimeric, however in solution CurF ECH_2 also displays concentration dependent dissociation indicating a dynamic equilibrium between trimeric and lower oligomeric states [70] (**Fig. 18**).

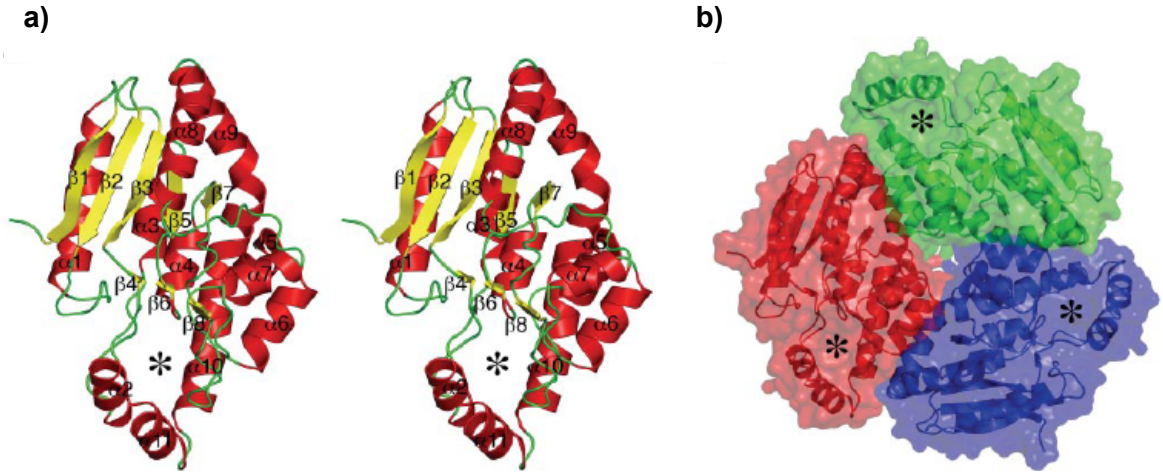


Fig. 18: Structure of CurF ECH_2. a) Presentation of the CurF ECH_2 domain as a stereo diagram. b) Trimeric structure of the CurF ECH_2 domain. In both representations the active site chamber is indicated by an *asterisk* (picture and subtitle from [70]).

The backbone amides of Ala78 and Gly118 form an oxyanion hole and the hydrophobic active site chamber includes only three polar side chains. Mutational analysis showed that product formation was more than 20-fold reduced for K86A, K86Q, H240A, and H240Q variants compared with wild type revealing Lys86 and His240 are critical for catalytic activity but not Tyr82. Thus it could be concluded that Lys86 and His240 play an important role in substrate binding or catalysis (

Fig**. 19**). His240 is proposed to stabilize the substrate carboxylate, and Lys86 is proposed to donate a proton to the C-4 position of the product. The structural arrangement might be the explanation for the specific regiochemistry in the formation of the key isopentenyl-ACP [70]. Further on Geders and his colleagues could show that CurF ECH_2 had a 20-fold preference for the ACP-linked substrate- a preference which has not been reported for any other member of the crotonase superfamily so far [70].

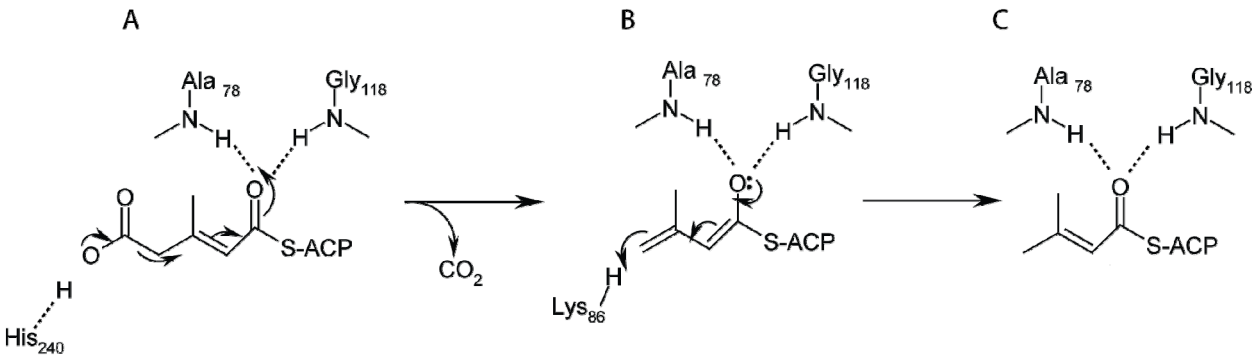


Fig. 19: The proposed mechanism for CurF ECH_2. During the hole mechanism is the oxyanion hole formed by the backbone amides of Ala78 and Gly118. His240 stabilizes the substrate carboxylate (A), and Lys86 donates a proton to the decarboxylated intermediate (B) to form the product (C) (Picture from [70]).

The halogenase (Cur Hal)

The Halogenase (Hal) is located at the C-terminus of CurA before the triplet ACP; it catalyzes a cryptic chlorination which leads to the cyclopropane ring formation in the synthesis of the natural

product curacin A. Cur Hal was originally annotated as phy due to its sequence similarity of 16% to the human oxygenase PhyH. Due to its function the name was changed later on. Hal is a nonheme Fe^{2+} and α-ketoglutarate (α-KG)-dependant enzyme which catalyzes under oxidising conditions the chlorination of inactivated carbon centres. The functional homologues SyrB2 from the syrongomycin pathway and CytC3 from the armentomycin pathway show a quite low sequence identity (13% and 16% sequence identity). It has been shown that these enzymes are able to catalyze the halogenation of an inactivated aliphatic carbon through a non-heme Fe-(IV)-O_5 intermediate ((SyrB2) [88][89][90]; (barbamide) [91][92][93]; (CytC3) [94]). The role and the timing for the chlorination in the curacin A synthesis could be established recently [69]. Cur Hal is specific for (*S*)-HMG-ACP as a substrate, once chlorinated the product reacts sequentially with the CurE ECH_1 dehydratase, the CurF ECH_2 decarboxylase and the enoylreductase (ER). Then a nucleophilic displacement will take place and the production of the cyclopropane ring forms the endpoint of the reaction series [69].

Structure- The protein Cur Hal was recently crystallized and spans the residues 1600-1919 which are designated as 1-1320 [67]. Hal is a dimer in solution as well as in the crystals. Crystals could be obtained for five different ligand states which are resumed in the following table (**Table 2**). All crystal structures have in common that the amino acids 319 and 320 are not visible (**Table 2**). It is to say that in the open conformation I and III (only Fe^{3+}) the lid region is not visible in the crystal structure.

Table 2: Summary of the different ligand states which could be crystalized (data from [67]).

Crystal Structure	I	II	III		VI
PDB	3NNJ	3NNF	3NNL		3NNM
Conformation	Open	closed	closed	open	Closed
Ligand State	None	Fe^{3+} α-KG Cl^- O_2 Formate	Fe^{3+} α-KG Cl^-	Fe^{3+}	Formate

All four halogenase crystal structure posses the cupin fold. The distantly related enzymes SyrB2 and CytC3 share the same fold which consists of a barrel-like structure formed by a major anti-parallel β-sheet ((β1, β6, β4, β8, β3, and β2). The active site is housed by the cupin barrel and forms specific interactions between α-KG and Fe with Cur Hal. Helices α5, α7, and α8 stabilize the dimer by hydrophobic interactions. The active site of each monomer is separated by 37 Å (not shown). Hal has with its homologues SyrB2 and CytC3 a lid in common, which is formed by

27 residues (**Fig.20 a)**, amino acids 40-66 form the lid, which is indicated in red). In the different crystal structures two conformations can be distinguished- the open conformation with no ligand and the closed conformation with α-KG, Fe, Cl^-. The data suggest that the conformational switch between them is triggered upon α-KG binding which in turn shows dramatic differences in the lid region: When α-KG is bound the active site is covered by the lid and the protein is in its "closed" conformation, without α-KG no lid region is visible in the crystals. The lid seems to become disordered and opens the active site ("open" conformation). Further on the structures differ radically for the loop connecting helices α9 and α10 (residues 274–284) (magenta and dark blue in **Fig.20 a)**). Finally, several secondary structures are in slightly different positions, including strand β3 and helices α2, α4, α5, α6, α9, and α10 [67]. These structural differences clearly distinguish the "open" and the "closed" state of the enzyme.

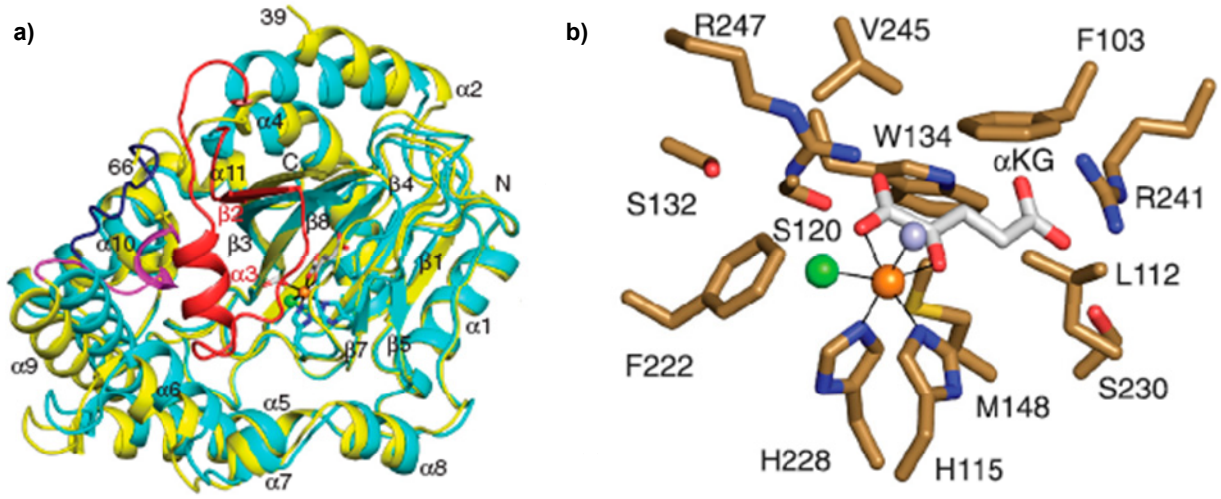


Fig.20: The crystal structure of the halogenase (Cur Ha)l (a) Overlay of the open (3NNJ) and closed form (3NNF) of Cur Hal. In the closed-form subunit (cyan), the 27 amino acid lid (colored in red) covers the active site. The α9–α10 N- loop (dark blue) is at the periphery. In the open form (yellow), the lid (residues 40–65) is missing, and the α9–α10 loop (magenta) is adjacent to the active site and to a 3_{10}-like helix preceding α4. Structural changes include shifts of β3 and helices α2, α4, α5, α6, α9, and α10. b) In the metal center (shown in sticks), iron (orange) coordinates two histidine side chains (cyan C), αKG (gray C), chloride (green) and formate (white C). Ligands are shown with atomic colors (red O, blue N) and coordination bonds in black (pictures from [67]).

Mechanism

The different ligand states are consistent with the mechanism in which decarboxylation of α- KG by iron bound dioxygen creates a ferryl-oxo (Fe4+-oxo) intermediate (

Fig. 21) [89, 90, 94]. Based on the mechanism proposed for SyrB2, CytC3 and the Cur Hal crystal structure, Khare and colleagues propose the following mechanism: The resting enzyme containing Fe^{2+} is in the open form excluding HMG. Upon binding of α-KG the enzyme closes the lid and triggers *S*-HMG binding. The substrate binding permits then dioxygen to coordinate the Fe^{2+} and triggers decarboxylation of α-KG by iron bound dioxygen which creates a chloro-oxo-ferryl intermediate. In the following the substrate subtracts hydrogen and the substrate radical will be resolved by chlorination [67].

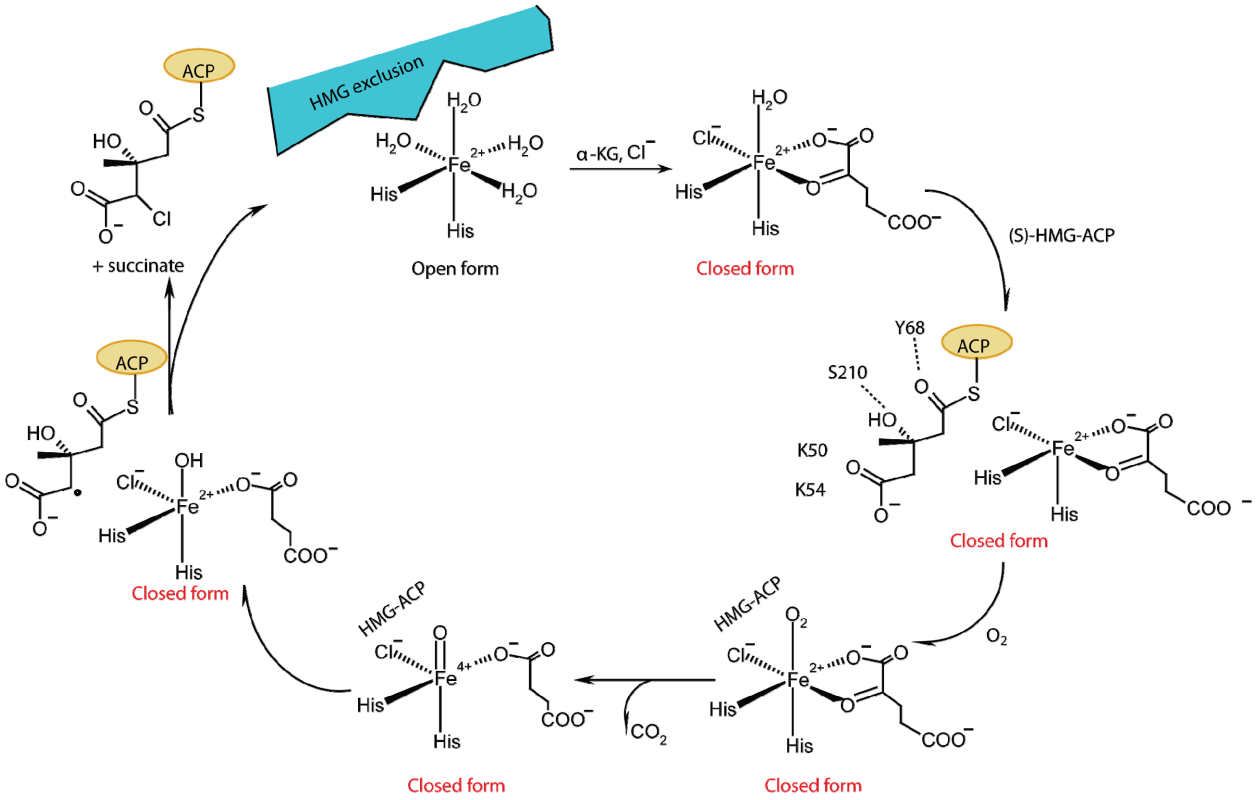


Fig. 21: The proposed Cur Hal mechanism including the role of the conformational change. As long as the enzyme is in the open form, (*S*)-HMG-ACP is excluded. The conformational switch from the open to the closed form is triggered upon binding of α-KG and Cl-binding. Once Cur Hal is in its closed form (*S*)-HMG-ACP can bind. Due to the substrate binding dioxygen coordinates the iron and triggers oxidative decarboxylation of α-KG. The intermediate chloro-oxo-ferryl abstracts a hydrogen atom from the substrate; the substrate radical is resolved by chlorination via a "rebound" mechanism (Picture adapted from [67]).

Based on modeling and alanine replacement it is supposed that the amino acids- Leu112, Lys48, Lys50, Lys54, Ser44, and Tyr68- are involved into the stereo-specific recognition and binding of *S*-HMG. Based on residue substitution it seems to be that Arg241 and Arg247 are essential for α-KG binding, Ser120 is important for Cl^-. Concerning the substrate it is important to say that the proximal carboxylate in HMG is a complication for Fe-dependant enzymes. Therefore it might be crucial to keep HMG out of the sphere of Fe^{2+} until the active site is assembled, due to the potential of the carboxylate to be a ligand itself. Therefore it might make sense to exclude HMG as long α-KG and Cl^- are not bound [67].

4.3 Nuclear magnetic resonance on mega-synthethases

Structural elucidation of NRPS and PKS modules is inevitable in order to get a crucial understanding of the structural determinants important for biosynthesis. During the last decades single domain architecture were determined by either X-ray or NMR structure elucidation. However, dissected domain structures only gave little knowledge of the mechanism underlying the domain interaction during NRP- and PKS-synthesis [95]. In the last years, it was possible to obtain high-resolution NMR and X-ray structures of didomains such as the homodimeric AT-KS fragment of DEBS module 5 or the SrfA-C domain [29, 30]. Both cases give an insight into the

structural arrangement but do not enlighten the domain-domain interactions. In both structures interdomain movements would be required to allow interactions of the domains with the 4'-Ppant arm of the ACP domain. PKS and NRPS systems contain large flexible linker regions which make crystallization of multidomain proteins difficult. It is possible to deduce mechanisms or even dynamics from different structures of the proteins in case it is possible to trap it in different moments which worked out so far for single domains such as CurA Hal [67].

Beside X-ray, NMR spectroscopy is a powerful tool that can be used to investigate structures, dynamics, and chemical kinetics and can overcome the restrictions X-ray suffers from. Although structure determination by NMR is limited to proteins with molecular weights less than 40-60 kDa new NMR techniques allow to investigate bigger proteins using for example stereo-array isotope labeling (SAIL) labeling [96-98] or thermophile homologues [99]. Beside structural investigations, NMR is an excellent tool to obtain complementary results for domain interactions by performing titration experiments and using fast mapping approaches for the domain-domain interactions [100]. Concerning the curacin triplet, a further challenge to face is the high symmetry of the protein leading to high overlay in the NMR spectra. One solution to this problem is to break the symmetry of the protein. Gaponenko and colleagues broke the symmetry of a dimer by attaching a paramagnetic probe resulting in improved peak distribution due to pseudocontact shifts [101]. In recent years, segmental labeling advanced as tool enabling the investigation of domain-domain interactions within the native environment and enabling the reduction of peak overlays of challenging proteins [129].

4.4 Segmental labeling

The DNA level of eukaryotes is distinguishable in two major divisions, the protein-coding sections called exons, and nonprotein-coding sections called introns. All genes begin with the protein-coding sequences, the exons, and have then a variable number of exons and introns alternating with each other. RNA splicing is the process in which the mRNA will be produced which codes for the final protein [102]. Inteins are as well coded on the mRNA level and fully translated into a protein sequence, the non-functional protein precursor. Protein splicing is a post-translational process which opened with its discovery 20 years ago a new chapter of protein-biochemistry [103, 104]. The intervening sequence called intein divides a functional protein (the extein) into an N- and C-terminal part (the N- and C-intein). This “precursor” protein undergoes an autocatalytic process in which the intein excises itself and fuses the two adjacent exteins via a covalent and natural peptide bond [105] (**Fig. 22**).

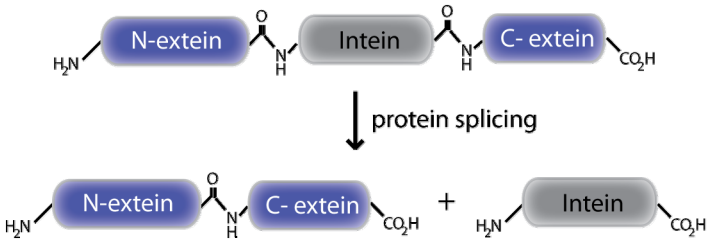


Fig. 22: Schematic representation of the protein splicing.

About 100 cases of protein splicing has been found in all three domains of life, archaea, bacteria and eukarya [105]. Mutagenesis studies in the 90s using an *in vitro* system revealed that the splicing process is i) catalyzed due to the amino acids contained in the intein ii) a intramolecular process [106] iii) needs no external energy source, cofactors or auxilaury enzymes [107].

4.4.1 Splicing mechanism

Protein splicing reaction involves four concerted steps: 1) N-S acylshift 2) *trans*-thioestherification 3) Asn-cyclization 4) S-N acylshift and possibly undesired side reactions [108]. The first step in the splicing mechanism is the formation of a linear ester intermediate by N–O or N–S acyl rearrangement involving the nucleophilic amino acid residue at the N-terminal splice junction (**Fig. 23 a)**) [109]. This acylshift occurs adjascent to a serine, threonine, or cysteine residue. The formation of this linear ester intermediate is an essential step in protein splicing. Mutations of the serine or cysteine to other amino acids completely block the protein splicing [110][107]. The following step is the transesterification. An intramolecular nucleophilic attack of a nucleophilic residue at the C-terminal splice junction occurs on the linear ester intermediate (**Fig. 23 b)**).For the next step an asparagine residue adjascent to the C-terminal

splice junction is essential. The asparagine undergoes a slow cyclization coupled with the cleavage of the ester intermediate resulting in an intein with an aminosuccinimide and two exteins with an ester bond (**Fig. 23 c)**). Finally the ester rearranges to a more stable amide bond and the aminosuccinimide spontaneous hydrolysis in uncatalyzed reactions (**Fig. 23 d)**).

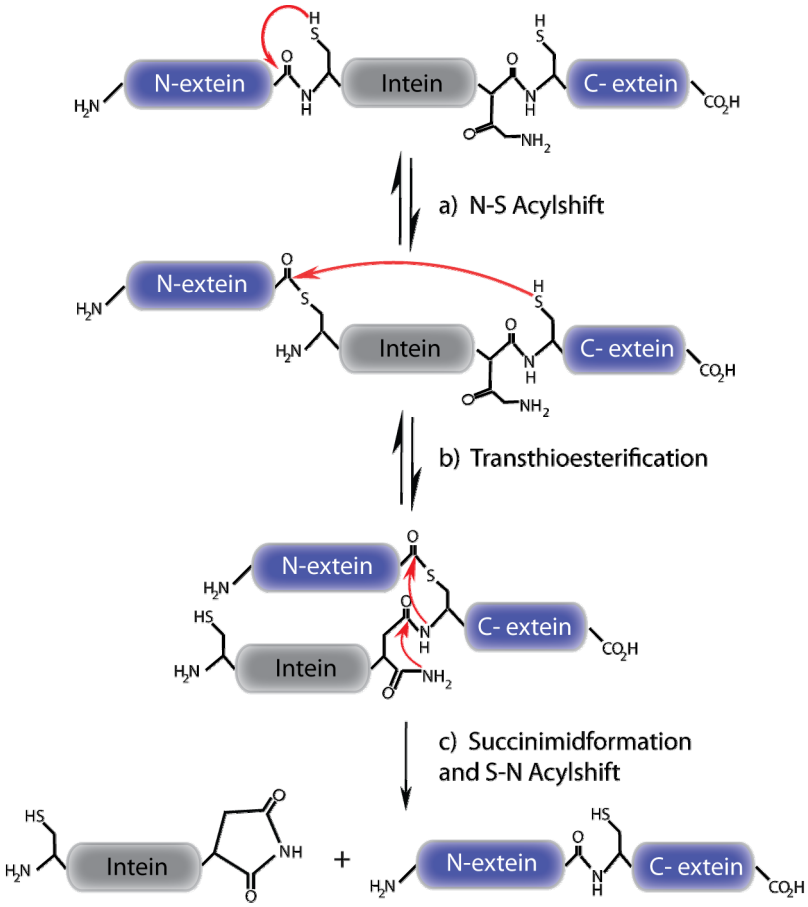


Fig. 23: The mechanism of protein splicing. The standard protein splicing mechanism is presented here with cysteines at both splice junctions. The mechanism is similar if the cysteine nucleophile is replaced by a threonine or serine but in that case esters are generated (adapted from [111]).

4.4.2 The nomenclature of inteins

In nature the host proteins for inteins are quite diverse in function; they range from metabolic enzymes over proteases to DNA and RNA polymerases, ribonucleotide reductases, and the vacuolar-type ATPase. However host proteins involved in DNA repair and replication seem to dominate [112]. The inteins are mostly found in highly conserved regions near to the active site [113]. Inteins are named after the organism and gene in which they are found. The organism name follows the same consensus as restriction enzymes and uses a three letter genus and species designation, followed by a strain designation, if necessary. The organism name is followed by an abbreviation of the extein gene name. If more than one intein is present in an extein gene, the inteins are given a numerical suffix starting from 5' to 3' or in order of their identification [114]. The *Pyrococcus furiosus* ribonucleoside-diphosphate reductase α subunit gene for example contains two inteins. 'Pfu' abbreviates the organism; the gene has been called

the 'RIR1' gene. Thus, these two inteins are called the Pfu RIR1- 1 intein and the Pfu RIR1- 2 intein.

The intein name refers to both, the gene and the intein protein. Therefore, as usual, the consensus is to italicize the gene name and to capitalize the first letter of the protein name [114]. Many inteins are bifunctional and contain an endonuclease activity; in this case the intein is also given a second name that follows the endonuclease naming conventions [115]. This name includes the prefix 'PI-', for protein insertion, the 3 letter organism abbreviation and a Roman numeral indicating the order of identification of the intein endonuclease in that organism. As a result the endonuclease names for the Pfu RIR1- 1 and Pfu RIR1- 2 inteins are PI- PfuI and PI- PfuII, respectively [114]. Inteins can be differentiated according to their size. The 300 to 600 amino acid large inteins contain a self-splicing domain and an endonuclease domain. Interestingly the deletion of the endonuclease domain from a large intein does not affect protein splicing [116-118]). The endonuclease domain might be essential for the spread of the inteins among the kingdoms of life. Inteins missing the endonuclease are much smaller and so called mini-inteins. They contain only the self-splicing domain and range from 150 to 200 amino acids. Split inteins are mini-inteins split into two units. Inteins, such as PI-PfuI and PI-PfuII from *Pyrococcus furious* have been artificially split into two fragments in order to obtain split inteins that can perform *trans*-splicing [119-121]. Later on two naturally split inteins were identified in DnaE, the catalytic subunit α of DNA polymerase III, in the cyanobacteria *Synechocystis sp.* strain PCC6803 (*Ssp*) and *Nostoc punctiforme* (*Npu*), both are capable of protein *trans*-splicing [114, 122, 123] (**Fig. 24**).

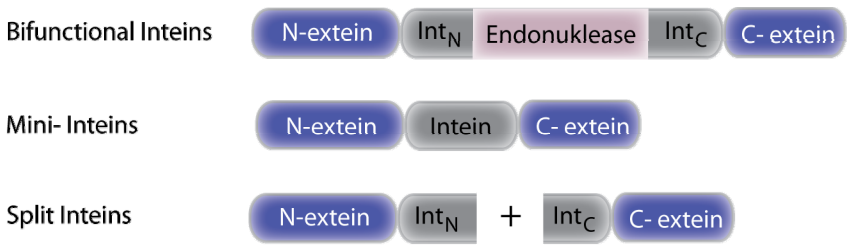


Fig. 24: Schematic presentation of the different sorts of inteins.

4.4.3 The split intein *Npu* DnaE and the protein *trans*-splicing

N- and C-terminal artificially split inteins can be reconstituted and promote after *in vitro* reconstitution protein splicing activity in *trans* [124-126]. By fusing proteins N- and C-terminally with inteins, a ligation of the target protein can be achieved [119, 123, 124] (**Fig. 25**). Limiting factors of the protein ligation using such artificially split inteins are the low yield and the required denaturating and reconstitution steps which make the procedure tedious. Further required denaturation and refolding protocols limit the ligation using artificially split inteins to protein targets which survive refolding [120, 126]. In contrast to artificially split inteins, naturally split

inteins such as *Ssp* DnaE and *Npu* DnaE have been used for spontaneous *trans*- splicing *in vivo* as well as *in vitro* without refolding procedures [114, 122, 127]. The usage of protein *trans*-splicing paved a way for many applications such as segmental labeling [119, 120, 126, 128, 129], protein cyclization [122, 130-132], *in vivo* protein engineering [133-136], site-specific chemical modifications [137] [138] and activation of genes in transgenic plants [139].

Analysis of the protein-splicing mechanism is quite complicated and the last step occurs spontaneously and is therefore difficult to analyze **(4.4.1)**. Split inteins first associate (**Fig. 25a)**) and then the usual splicing mechanism takes place via *trans*-splicing leading to the intein on the one hand and the ligated protein on the other hand **(4.4.1)**. For the application of *trans*-splicing it is important to mention that different side reactions may lead to either N- or C-terminal cleavage of the extein building end points of the reactions (**Fig. 25 b),d)).** The amount of side reactions depends on pH and temperature; thiol-induced cleavage depends on the intein and extein sequence [140].

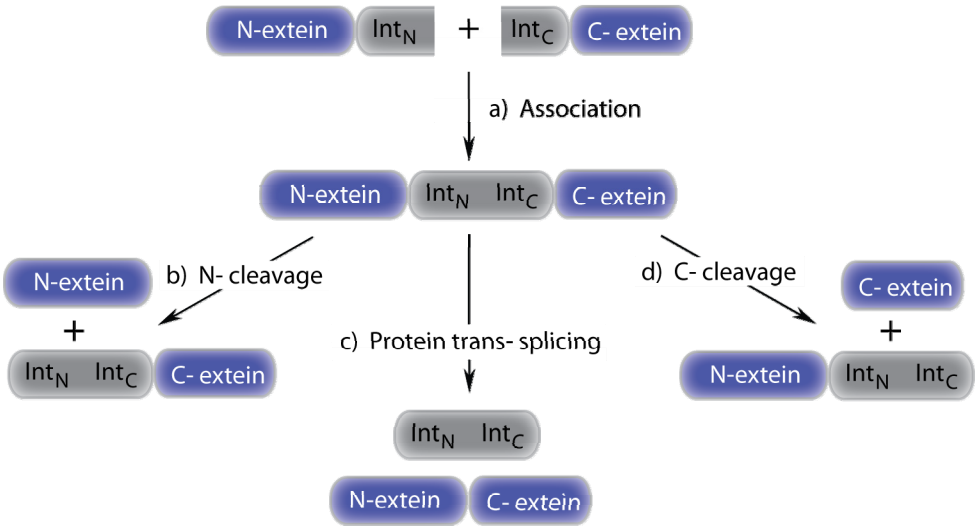


Fig. 25: The *trans*- splicing reaction mechanism (from [141]).

The robust splicing activity and high efficiency of the naturally split intein of the gene DnaE from *Nostoc punctiforme* (*Npu* DnaE) of over 98% and its tolerance of sequence variations at the splicing junctions made this protein especially interesting for further engineering [127]. For a broad application it is of interest to have short C-terminal inteins (C-intein or IntC) which can be synthesized chemically. The C-terminal fragment of the naturally split *Npu* DnaE intein consists of 36 residues (NpuInt$_{C36}$) (**Fig. 26 I)**). 36 Amino acids are too long for chemical synthesis; this length might disturb the solubility of the precursor fragment fused to the C-intein due to the relatively high hydrophobicity and the unfolded state of the C-terminal Intein. For Hideo and co-workers it was possible to identify a functional C-intein consisting of 15 residues (NpuInt$_{C15}$) (**Fig. 26 II))** by shortening the C-intein systematically [141, 142].

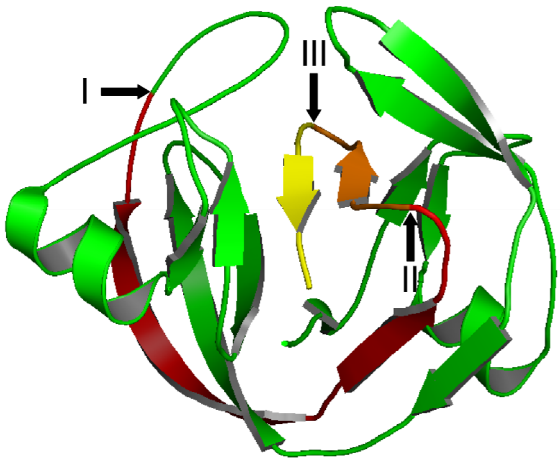


Fig. 26: Solution structure of the *Nostoc punctiforme* Intein (*Npu* DnaE)(PDB: 2KEQ [140]. The roman numbers indicate the different splicing sites I) indicates the WT natural splicing site separating the WT into the N-intein (*Npu*Int$_{N\ \Delta C36}$) and C-intein (*Npu*Int$_{C36}$ (red, orange and yellow)) II) indicates the artificial splicing site identified by a systematic approach separating the Intein into the N-intein (*Npu*Int$_{N\Delta C15}$) and C-intein (*Npu*Int$_{C15}$ (orange and yellow)) III) shows the splicing site identified by the structural approach splitting the Intein into the N-Intein (*Npu*Int$_{N\ \Delta C6}$) and C-intein (*Npu*Int$_{C6}$ in yellow) (Picture adapted from [140]).

Solving the solution structure of the *Npu* DnaE intein and determining localizations of conformational exchange allowed creating a further functional intein pair with only six amino acids for the C-terminus (NpuInt$_{C6}$) (**Fig. 26 III)**) [140]. Using these different *Npu* DnaE splicing variants it turned out that for the ligation of B1 domain of protein G (GB1) with itself, the intein pair *Npu*DnaE-Int$_{N123}$/*Npu*DnaE-Int$_{C15}$ shows the highest efficiency with 96%. In the same way different splicing variants for DnaE intein from *Synechocystis sp PCC6803* (*Ssp*DnaE) were prepared but there, the highest yield was obtained for the wt combination *Ssp*DnaE-Int$_{N123}$/ *Ssp*DnaE-Int$_{C36}$ (67%) in case of the ligation of GB1 with chitin binding domain (CBD) and the new functional split inteins displayed no practical use [141]. Using the two different splicing variants of *Npu*DnaE Intein, the wt *Npu*DnaE-Int$_{N102}$/*Npu*DnaE-Int$_{C36}$ and the engineered *Npu*DnaE-Int$_{N123}$/*Npu*DnaE-Int$_{C15}$, it was shown that ligation by protein *trans*-splicing does not only depend on the splicing junction sequence but also on the foreign extein sequence [141] (**Table 3**).

Table 3: The final yields of protein ligation by protein *trans*- splicing [141].

Intein	N-extein	C- extein	Yield (%) with 50mM DTT	Yield (%) with 0.5 mM TCEP
NpuDnaE -Int$_{N102}$/ Int$_{C36}$ (Wild type)	nSH3	cSH3	n.d	n.d
	GB1	cSH3	n.d	n.d
	nSH3	GB1	77+- 10	61+- 8
NpuDnaE -Int$_{N123}$/Int$_{C15}$	nSH3	cSH3	27+- 10	50+- 7
	Smt3	GB1	9+- 3	65+- 7

These studies revealed further, the effect of reducing conditions on the *trans*-splicing efficiency. Both the N- and C-intein contain cysteines which might form disulphide bonds if the proteins are not in reducing conditions. In case of *Ssp*DnaE intein 50mM dithiothreitol (DTT) blocks the

trans-splicing and shifts the reactions completely to the side reactions, the *trans*-cleavages (**Fig. 25 b)** and **d)**)[143]. In fact, besides being a sulfhydryl reductant dithiothreitol (DTT) can act as a nucleophile and therefore induces N- and C-cleavage. Replacing dithiothreitol (DTT) with trialkylphospine(tris-(2-carboxyethyl)-phosphine, TCEP), which is unreactive with cyteines, might improve the ligation in many cases for example for the ligation of Src homology 3 (nSH3) domain with cSH3 using the inteins *Npu*DnaE -Int_{N123}/Int_{C15} [141].

4.4.4 Segmental labeling for nuclear magnetic resonance

Segmental isotopic labeling is a method which was first established *in vitro* and allows incorporating isotopes selectively in regions or domains of the target protein to reduce the complexity of the spectra [119, 120, 144]. Iwai and his group established a convenient dual expression system which allows segmental isotopic labeling in *Escherichia Coli* and will be presented in the following part [128, 145]. In order to achieve the incorporation of NMR active isotopes only into one domain a dual vector system with two different tightly controlled inducible promoters for individual expression is indispensable. Additionally, the two vectors need two different antibiotic resistances and ORI's to ensure a co-existence and similar distribution during cell division. The vectors used for this dual expression for the segmental labeling of the proteins GB1-CBD were pJJDuet30 (RSF ori, Kan^R, T7 Promotor, lacI) which can be induced by the addition of IPTG to the media and the vector pSFBAD09 (ColE1 ori, Amp^R, araBAD) which underlies the tight induction by L-arabinose.

In the *in vivo* approach the preparation of an isotopically labeled sample takes place as following. For explanation, we take as example that the C-terminal extein, fused to the C-terminal intein, is under arabinose control and the N-terminal extein fused to the N-terminal intein is under T7 control. The exteins might be different domains, a tag and a domain or two different parts of a protein of interest. The *E. coli* cells containing both plasmids are grown in unlabeled media (**Fig. 27 a)**). At an appropriate OD the C-terminal extein fused to the C-terminal intein are induced by the addition of L-arabinose. The protein is then expressed for a certain amount of time, followed by media exchanged to ^{15}N-labeled medium by implementing a centrifugation step. An additional washing step allows incorporating ^{15}N afterwards to a level of 97% [128](**Fig. 27 b)**). During the media exchange residual arabinose gets metabolized and the arabinose promoter turns off. After adaptation to the ^{15}N-labeled media the N-terminal protein fused to the N-terminal intein can be induced with IPTG. During the expression the protein *trans*-splicing takes place (**Fig. 27 c)**). After the 2nd expression the segmental labeled protein can be purified as usual [128] (**Fig. 27 d)**). The dual expression system gives high flexibility, tags can be added to the precursors in order to increase solubility or for purification purposes, the order of expression can be changed by switching the vectors. In fact, it is also possible to label the protein during the first expression step under arabinose control [128].

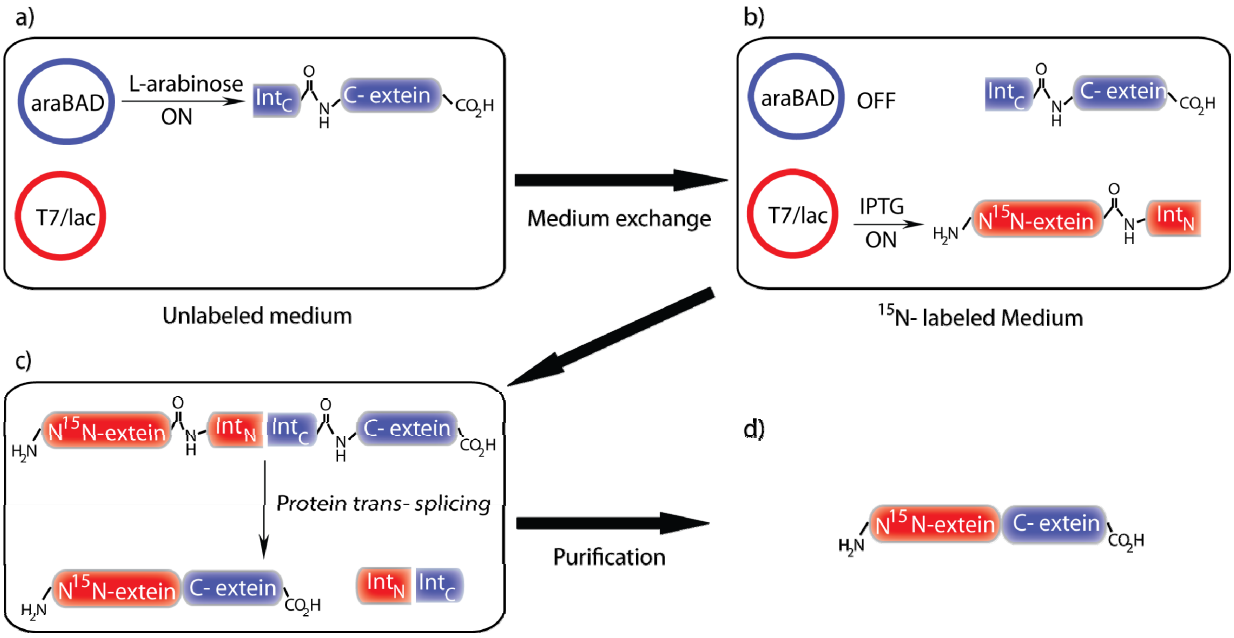


Fig. 27: Preparation of a segmental isotopic labeled sample using *in vivo trans*–splicing. (a) The C-terminal protein (C-extein) fused with inteinC can be expressed by addition of L-arabinose. b) The medium will then be exchanged by centrifugation step into ^{15}N-labeled medium. During adaptation to the new media remaining arabinose will be metabolized and the arabinose promoter stops activity. The N-terminal protein fused to the N-intein is then induced by IPTG producing the ^{15}N isotopically labeled N-terminal precursor. C) The two N- and C-terminal precursor associate and protein *trans*-splicing occurs. In this way the ^{15}N-labeled N-terminal extein fuses with the unlabeled C-terminal part. d) The fused protein can then be purified in the appropriate way (adapted from [128]).

4.4.5 Three fragment ligation

The two fragment ligation can be easily extended to three fragment ligation. In this way, if one uses an additional split intein that is able to catalyze spontaneous *trans*-splicing efficiently and specifically, even larger proteins can be accessed. For three fragment ligation it is important that the two Inteins are orthogonal with high splicing efficiency and without any cross-reactivity. This approach was elegantly solved using the two orthogonal split inteins, PI-PfuI and PI-PfuII [120]. However, the artificially split inteins of PI-PfuI and PI-PfuII required tedious refolding and optimization steps to restore the splicing activity due to insolubility of the precursors [120, 146]. Since refolding can also result in inactive proteins, avoiding any refolding steps would be highly beneficial. Hence, it is crucial to have split inteins with more robust and specific *trans*-splicing activity. The wild type Intein *Npu*DnaE-Int$_{N102}$/Int$_{C36}$ and the newly engineered splicing variant *Npu*DnaE-Int$_{N123}$/Int$_{C15}$, engineered by Iwai and collegues, are the key to this three fragment ligation without any tedious refolding steps [127, 129]) (**4.4.3**). The three fragment ligation using these intein pairs will be presented in detail in results and discussion (**7.2**).

5 Materials

5.1 Laboratory equipment

Name	Producer
Agarose gel electrophoresis system	Peqlab Biotechnologie GmbH, Erlangen (Germany)
ÄKTAprime plus	GE Healthcare Europe GmbH, München (Germany)
ÄKTApurifier	GE Healthcare Europe GmbH, München (Germany)
Amicon ultrafiltration membrane	Millipore Corporation, Billerica (USA)
Amicon stirrer unit	Millipore Corporation, Billerica (USA)
Amicon filter- centrifugation unit	Millipore Corporation, Billerica (USA)
Analytical balance CPA124S-OCE	Sartorius AG, Göttingen (Germany)
Autoclave	Tecnomara AG (Switzerland), Gettinge AB (Sweden)
Avance 500 MHz NMR Spektrometer	Bruker BioSpin GmbH, Rheinstetten (Germany)
Avance 900 MHz NMR Spektrometer	Bruker BioSpin GmbH, Rheinstetten (Germany)
Cell disruptor Constant Systems	IUL Instruments GmbH, Königswinter (Germany)
Centrifuge 5810 R	Eppendorf AG, Hamburg (Germany)
Centrifuge Centricon H-401	Kontron-Hermle, Gosheim (Germany)
Centrifuge Sorvall RC-5B	Sorvall Instruments, Bad-Homburg (Germany)
Centrifuge Sorvall RC-5C	Sorvall Instruments, Bad-Homburg (Germany)
Centrifuge rotor A 6.9	Kontron-Hermle, Gosheim (Germany)
Centrifuge rotor F28-50	Sorvall Instruments, Bad-Homburg (Germany)
Centrifuge rotor F34-6-38	Eppendorf AG, Hamburg (Germany)
Centrifuge rotor GS-3	Sorvall Instruments, Bad-Homburg (Germany)
Centrifuge rotor SS-34	Sorvall Instruments, Bad-Homburg (Germany)
Dialyse Cassette Slyde-A-Lyzer	Thermo Scientific, Rockford (USA)
Dialysemembrane	Spectrum Laboratories, Ranco Dominguez (USA)
Elektrophorese- system Mini-Protean Tetra, Cell	Bio-Rad Laboratories GmbH, München (Germany)
Fermenter, 10-Litre	Sartorius AG, Göttingen (Germany)
French press cell disruptor	Thermo Scientific, Rockford (USA)
Heating unit	VWR International GmbH, Darmstadt (Germany)
HPLC Pump PU-980	Jasco GmbH, Gross-Umstadt (Germany)
HPLC Detector MD-910	Jasco GmbH, Gross-Umstadt (Germany)
Incubation shaker	Infors AG, Bottmingen-Basel (Germany)
Incubation shaker	Innova 4330 New Brunswick Scientific GmbH, Nürtin (Germany)
Magnetic stirrer RCT basic	IKA-Werke GmbH & Co. KG, Staufen (Germany)
Membrane pump	VACUUBRAND GmbH & Co. KG, Wertheim (Germany)
NanoDrop DN 1000	Peqlab Biotechnologie, Erlangen (Germany)
Optilab rex (RI detector)	Wyatt Technology Europe GmbH, Dernbach (Germany)
PCR-Maschine TPersonal	Biometra biomedizinische Analytik, GmbH, Göttingen (Germany)
pH-Metre PHM210	Radiometer-Analytical SAS, Villurbanne (France)
Plate incubator	Memmert GmbH & Co. KG, Schwabach (Germany)
Portable Precesion balance	Sartorius AG, Göttingen (Germany)
Precesion balance PB3002	Mettler-Toledo GmbH, Gießen (Germany)
Power supply EPS 300	Pharmacia Biotech/GE Healthcare Europe GmbH, München (Germany)
Power supply Power PAC 3000	Bio-Rad Laboratories GmbH, München (Germany)

Shaker Promax 2020	Heidolph Instruments GmbH & Co., KG, Schwabach (Germany)
Sonificator	Labsonic U B. Braun Biotech International, Melsungen (Germany)
Spectrofluorometre FP- 6500	Jasco GmbH, Gross-Umstadt (Germany)
Spectropolarimetre J- 810	Jasco GmbH, Gross-Umstadt (Germany)
Table centrifuge Biofuge 13	Heraeus, Hanau (Germany)
Table centrifuge Mikro 22R	Andreas Hettich GmbH & Co. KG, Tuttlingen (Germany)
Waterbath shaker	New Brunswick Scientific Inc., Edison (USA)

5.2 Chromatography

Material/ Column	Producer
IMAC Sepharose 6 fast flow	GE Healthcare Europe GmbH, München (Germany)
Ni-NTA Superflow	Qiagen, Hilden (Germany)
OmniFit 25/100 column	GE Healthcare
HiLoad 1660 Superdex 200	GE Healthcare
Superdex 75 10/300 column	GE Healthcare
PerfectSil 250x 4,6 C18 column	MZ Analysetechnik GmbH, Mainz (Germany)
Superose 6 10/300 column	GE Healthcare
MonoQ 5/50GL	GE Healthcare
Amylose Resign	New England Biolabs GmbH, Frankfurt (Germany)

5.3 Chemicals

Name	Producer
Acetyl phosphate lithium potassium salt	Sigma-Aldrich Chemie GmbH (Germany)
Adenosine 5'-triphosphate (ATP)	Roche Diagnostics GmbH (Germany)
Amino acids (unlabeled)	Sigma-Aldrich Chemie GmbH (Germany)
Antifoam Y-30 emulsion	Sigma-Aldrich Chemie GmbH (Germany)
L-(+)-Arabinose	Carl Roth GmbH+Co.KG (Germany)
Bactotryptone	Carl Roth GmbH+Co.KG (Germany)
Bovine serum albumin Fraction V	Sigma-Aldrich Chemie GmbH (Germany)
Complete Protease Inhibitor Cocktail	Roche Diagnostics GmbH (Germany)
Dimethylsulfoxide	Carl Roth GmbH+Co.KG (Germany)
1,4-Dithiotreitol (DTT)	Carl Roth GmbH+Co.KG (Germany)
Ethidiumbromide	Carl Roth GmbH+Co.KG (Germany)
Folinic acid calcium salt	Sigma-Aldrich Chemie GmbH (Germany)
Gene ruler 100bp, 1kb DNA ladder	Fermentas GmbH (Germany)
Glucose monohydrate	Carl Roth GmbH+Co.KG (Germany)
^{13}C-Glucose monohydrate	Sigma-Aldrich Chemie GmbH (Germany)
HEPES	Carl Roth GmbH+Co.KG (Germany)
Imidazole	Carl Roth GmbH+Co.KG (Germany)
IPTG	Carl Roth GmbH+Co.KG (Germany)
KCl	Carl Roth GmbH+Co.KG (Germany)
α-Ketoglutarate	Sigma-Aldrich Chemie GmbH (Germany)
KH_2PO_4	Carl Roth GmbH+Co.KG (Germany)
K_2HPO_4	Carl Roth GmbH+Co.KG (Germany)
Magnesium acetate tetrahydrate	Sigma-Aldrich Chemie GmbH (Germany)
D-(+)- Maltose monohydrat	Sigma-Aldrich Chemie GmbH (Germany)
2-Mercaptoethanol	Carl Roth GmbH+Co.KG (Germany)
NaCl	Carl Roth GmbH+Co.KG (Germany)

NaN_3	Fluka, Sigma-Aldrich Chemie GmbH (Germany)
NH_4Cl	Sigma-Aldrich Chemie GmbH (Germany)
^{15}N-NH_4Cl	Sigma-Aldrich Chemie GmbH (Germany)
$(NH_4)_2Fe(SO_4)_2$* $6H_2O$	Carl Roth GmbH+Co.KG (Germany)
Peptone, tryptic digest	Carl Roth GmbH+Co.KG (Germany)
Phosphoenol pyruvic acid (PEP), monopotassium salt	Sigma-Aldrich Chemie GmbH (Germany)
Potassium acetate	Sigma-Aldrich Chemie GmbH (Germany)
Pyruvate kinase	Roche Diagnostics GmbH (Germany)
Restriction enzymes	New England Biolabs GmbH (Germany)
RiboLock RNase Inhibitor	Fermentas GmbH (Germany)
Rotisphorese Gel 30	Carl Roth GmbH+Co.KG (Germany)
Rotisphorese Gel 40	Carl Roth GmbH+Co.KG (Germany)
SDS	
T4 DNA-ligase	New England Biolabs GmbH (Germany)
TCEP	PIERCE, Thermo Fisher Scientific (Germany)
tRNA *E. coli* MRE 600	Roche Diagnostics GmbH (Germany)
Yeast extract	Carl Roth GmbH+Co.KG (Germany)

5.4 Bacterial strains

Strain	Features
NEB 5-alpha competent *E. coli* (NEB GmbH)	*F' proA$^+$B$^+$ lacI^q Δ(lacZ)M15 zzf::Tn10* (Tet^R)*/fhuA2 Δ (argF-lacZ)U169 phoA glnV44 f80 Δ (lacZ)M15 gyrA96 recA1 endA1 thi-1 hsdR17*
T7 Express competent *E. coli* (NEB GmbH)	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10*–Tet^S*)2 [dcm] R(zgb-210:: Tn10–Tets) endA1 Δ(mcrC-mmr)114::IS10*
M15 competent *E. coli* (Quiagen GmbH)	*Nal^S, Str^S, Rif^S, Thi$^-$, Lac$^-$, Ara$^+$, Gal$^+$, Mtl$^-$, F$^-$, RecA$^+$, Uvr$^+$, Lon$^+$,* contains mutations in *lac* and *mtl* genes, contains pREP4
Arctic Express (DE3) (Stratagene)	E. coli B *F– ompT hsdS(r_B^- m_B^-) dcm$^+$* Tet^r *gal λ* (DE3) *endA Hte* [*cpn10 cpn60* Gent^R]

5.5 Enzymes

Name	Producer
Phusion DNA Polymerase	New England Biolabs GmbH, Frankfurt (Germany)
Pfu Ultra II Polymerase	Stratagene, Agilent Technologies, Waldbronn (Germany)
DpnI Endonuklease	New England Biolabs GmbH, Frankfurt (Germany)
DNase I	Sigma-Aldrich Chemie GmbH, Taufkirchen (Germany)
RNase A	Sigma-Aldrich Chemie GmbH, Taufkirchen (Germany)
All restriction Enzymes	New England Biolabs GmbH, Frankfurt (Germany)

5.6 Kits

Name	Producer
NucleoSpin Plasmid	Macherey-Nagel GmbH & Co. KG, Düren (Germany)
NucleoBond PC 100	Macherey-Nagel GmbH & Co. KG, Düren (Germany)
QIAquick PCR Purification Kit	Qiagen GmbH, Hilden (Germany)
QIAquick gel extraction kit	Qiagen GmbH, Hilden (Germany)

5.7 Plasmids

Plasmid	Feature
pET-28a (+) (Novagen)	Kan^R; E. coli expression vector, T7 promotor and T7 terminator with lac operon, option for N- or C-terminal hexahistidine affinity tag.
pQE60 (Qiagen)	Amp^R; E. coli expression vector, T7 promotor and T7 terminator with lac operon, option for N- or C-terminal hexahistidine affinity tag.
pBH4	Amp^R; modifed pet E. coli expression vector
pRSF (Merck)	Kan^R; E. coli expression vector, contains T7 Promotor with lac operon and the T7 Terminator, N- terminal histidine tag with an enterokinase cleavage site possible as well as a C terminal S-tag possible, contains an RSF ori.
pMAL-c2x (New England Biolabs)	Amp^R; lacIq promoter and Ptac promotor; N- terminal MBP fusion protein; M13 Ori and pMB1 Ori
pBAD/ His (Invitrogen)	Amp^R; *ara*BAD promoter (PBAD); N-terminal polyhistidine tag with Enterokinase cleavage site; pBR322 origin
pCDF (Merck)	$Strep^R$; T7 promoptor and terminator with lac operon, a replication origin derived from CloDF13, N-terminal His Tag followed by an enterokinase (Ek) cleavage site and an optional C-terminal S- Tag sequence.
pRARE (Novagen)	Cam^R; pRARE derives from pACYC184 and codes for rare t- RNAs for the codons AUA, AGG, AGA, CUA, CCC, and GGA

5.8 Software

Name	Function
ASTRA 5.3.4.13	Processing and evaluation software for multi angle light scattering (MALS) data
Ape	Processing software for DNA sequencing chromatogram files
UNICORN 5.11	Software package for control, supervision of chromatograpy systems and processing of chromatograms
PyMOL	Molecular graphics software for 3D presentation and analysis of molecules and proteins
Molmol	Molecular graphics software for 3D presentation and analysis of molecules and proteins
Sparky	Sparky is a graphical NMR assignment and integration program for proteins, nucleic acids, and other polymers

Talos+	software for empirical prediction of phi and psi backbone torsion angles
Cyana	software for automatically NOE spectra assignment and structure calculation

5.9 Oligonucleotides

All primers were ordered from Biospring (BioSpring GmbH, Frankfurt).

5.10 Common buffers, media and reagents

All solutions are prepared with MilliQ water and kept at –20°C if not stated otherwise. Sterilization by filtering is performed using 0.22 µm filters.

Antibiotics:

Ampicillin stock, 1000-fold:	100 mg/mL Na^+- ampicillin salt in H2O.
Chloramphenicol, 1000-fold:	34 mg/mL salt in ETOH.
Gentamycin, 1000-fold:	20mg/ml gentamycin sulfate
Kanamycin stock, 1000-fold:	35 mg/mL kanamycin sulfate in H_2O

Inducers

L- Arabinose stock:	20% in H_2O.
IPTG stock:	1 M in H_2O.

Media

LB medium (1 liter):	10 g bactotryptone, 5 g yeast extract, 10 g NaCl. Adjust to pH to 7.2- 7.4 (NaOH). Sterilize by autoclaving.
2 x XYT medium:	16 g peptone, 10 g yeast extract, 5 g NaCl. Adjust to pH to 7.2- 7.4 (NaOH). Sterilize by autoclaving.
SOC medium (1 liter):	20 g bactotryptone, 5 g yeast extract, 0.5 g NaCl. Adjust to pH 7.0 with NaOH and sterilize by autoclaving. Add final concentrations of 25 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose solution (all filtersterilized).
M9 medium (1 liter):	8.6 g Na_2HPO_4 * 2 H_2O, 3 g KH_2PO_4, 5 g NaCl in 1 L H_2O, Adjust to pH to 7.2-7.4 (NaOH). Sterilize by autoclaving. After autoclaving add: 1.5 ml of autoclaved 1M $MgSO_4$, 150 µl of autoclaved 1 M $CaCl_2$, 3 ml Vitamin solution (Sigma MEM Vitamin solution 100X cat no M6895), 2 ml solution Q. Dissolve in H_2O 1 g NH_4Cl (^{15}N or unlabeled) and either 2 g of ^{13}C- Glucose or 4 g of ^{12}C- Glucose and add by filtersterilization.
Solution Q (1L):	8 ml of 5 M HCl, 5 g $FeCl_2$* 4 H_2O, 184 mg $CaCl_2$* 2 H_2O, 64 mg H_3BO_3, 18 mg $CoCl_2$* 6 H_2O, 4 mg $CuCl_2$* 2 H_2O, 340 mg $ZnCl_2$, 605 mg Na_2MoO_4* 2 H_2O, 40 mg $MnCl_2$* 4 H_2O

Reagents and buffers for cell-free expression

Protease inhibitor cocktail: 1.197 mg/mL AEBSF, 0.238 mg/mL leupeptin, 0.154 mg/mL bestatin, 0.098 mg/mL aprotinin, 0.089 mg/mL E-64, 0.034 mg/mL pepstatin.

Acetyl phosphate lithium potassium salt, AcP:

1 M stock solution, pH 7,0, adjusted with 10 M KOH.

Amino acid stock: final concentration of 8 mM of each amino acid in H_2O. Adjust pH to 8.0 with 10 M KOH. Suspensions are pipetted to the cell-free reaction set-up.

Complete® protease inhibitors cocktail with EDTA:

1 tablett with 1 ml H_2O gives a 50- fold stock solution.

DTT: 1M or 500 mM stock in H_2O.

Folinic acid calcium salt: 10 mg/mL stock in H_2O. Completely dissolve by incubation at 37°C.

HEPES: 2.4 M stock. Adjust to pH 8.0 with 10 M KOH.

Lithium chloride, LiCl: 14 M stock solution in distilled RNase free water. Sterilize by filtering. Store at RT.

Magnesium acetate tetrahydrate, Mg $(OAc)_2$:

1 M solution in H_2O.

NTP mix: 75-fold stock containing 90 mM ATP, 60 mM each GTP, CTP and UTP, pH 7.0 adjusted with NaOH.

Phospho(enol)pyruvic acid monopotassium salt, PEP:

1 M stock. Adjust to pH 7.0 with 10 M KOH.

Polyethylenglycol 8000, PEG8000:

40% (w/v) solution dissolved in H_2O by heating in hand warm prior to use.

Potassium acetate, KOAc: 4 M in H_2O.

Pyruvat kinase (Roche Diagnostics): Stock of 10 mg/mL. Commercial, no separate preparation necessary.

RiboLock™ RNase inhibitor: Stock of 40 U/µL. Commercial, no separate preparation necessary. (Fermentas)

Sodium acetate, NaOAc: 3 M stock in distilled RNase free water, pH 5.5 adjusted with NaOH. Sterilize by filtering. Store at 4°C.

Sodium azide, NaN_3: 3% and 10% (w/v) stocks in H_2O.

tRNA *E. coli* MRE 600: 40 mg/mL stock in H_2O.

Buffers for DNA-Agarose gels and SDS-PAGE

Agarose: e.g. 1% (w/v) agarose boiled in 1-fold TAE buffer. Store at RT.

Ammoniumperoxid sulfate, APS: 10% stock solution by dissolving 100 mg/mL ammonium-peroxid sulfate in H_2O. Store at 4°C.

Bradford (1L): 100 mg Coomassie Brilliant Blue G250, 50 ml Ethanol (99%), 100 ml Phosphorsäure (85%) in H_2O.

Coomassie brilliant blue-staining solution for SDS gels:

50% (v/v) ethanol (96%), 10% (v/v) acetic acid (100%) and 0.1% (w/v) Coomassie Brilliant Blue G250. Dissolve in H_2O and store at RT in a dark bottle to avoid exposure to light.

DNA-loading dye, 6-fold: 40% (w/v) sucrose, 0.25% (w/v) bromphenol blue and 0.25% xylene cyanol FF. Dissolve in H_2O and store at 4°C.

EtBr-stock solution: 10 mg/mL stock in H_2O.

SDS gel buffers:

Running buffer: 25 mM Tris-HCl, pH 8.0, 0.1% (w/v) SDS and 200 mM glycine. Store at RT.

Stacking gel buffer: 0.4% (w/v) SDS, 0.5 M Tris-HCl, pH 6.8. Store at 4°C.

Separating gel buffer: 0.4% (w/v) SDS, 1.5 M Tris-HCl, pH 8.9. Store at 4°C.

SDS-PAGE sample buffer, 5-fold: 25% (w/v) glycerol, 25% (v/v) β-mercaptoethanol, 7.5% (w/v) SDS, 0.1% (w/v) coomassie G250, 300 mM Tris- HCl, pH 6.8. Store at RT.

TAE buffer, 50-fold: Dissolve 2 M Tris , 5.7% acetic acid (v/v) and 50 mM EDTA in H_2O. Store at RT.

Tricine gel buffers:

Anode buffer, 10-fold: 1 M Tris-HCl, pH 8.9. Store at RT.

Cathode buffer, 10-fold: 1 M Tris, 1 M Tricine, 1% SDS (w/v). The pH should self-adjust to 8.25. Store at RT.

Gel buffer, 3-fold: 3 M Tris-HCl, pH 8.45, 0.3% SDS (w/v). Store at RT.

Protein purification and protein analysis

Ni-NTA/IDA-Buffer A: 50 mM NaPP, 300 mM NaCl, 10 mM Imidazol, pH=8

Ni-NTA/IDA-Buffer B: 50 mM NaPP, 300 mM NaCl, 600 mM Imidazol, pH=8

Ni-NTA/ IDA Stripping buffer:	500 mM NaCl, 100 mM EDTA, 20 mM TRIS, pH= 8; TRIS does not go into solution for pH<8
Ni-NTA/ IDA Loading buffer:	100 mM $NiSO_4$
Amylose buffer A:	20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, pH=7.8 add fresh 10 mM β-Mercaptoethanol
Amylose buffer B:	20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10mM Maltose, pH=7.8 add fresh 10 mM β-Mercaptoethanol

6 Methods

6.1 Standard methods of molecular biology and microbiology

6.1.1 Polymerase chain reaction (PCR)

Conventional PCR [147, 148] reactions were performed in 100 µl volume using Phusion High Fidelity DNA Polymerase from NEB to according to manufacturer's instructions. Primers were ordered from Biospring (Biospring, Frankfurt).

6.1.2 PCR purification and gel extraction

Restriction PCR purification and gel extraction were performed using the Qiagen QIAquick PCR Purification Kit and QIAquick Gel Extraction Kit according to manufacturer's instructions.

6.1.3 Restriction enzyme digest

Restriction enzymes were purchased from New England Biolabs (NEB). Enzyme concentration and buffer conditions were used as suggested by the manufacturer. Circular plasmid DNA was digested for 1 hr at the recommended temperature. PCR products were usually digested overnight at the recommended temperature.

6.1.4 Ligation of DNA fragment

For ligation of DNA fragment, T4-DNA-ligase was purchased from NEB. Ligation reactions were performed in a 10 µl volume containing 1 µl 10x T4-Ligase buffer, 1 µl digested backbone (40 ng), 7 µl digested insert (depending on the size and concentration of the insert, usually a molar ratio of 1:3 [vector to insert] was used), and 1 µl T4-DNA-ligase. The reaction was incubated overnight at 16°C and transformed the next day into bacteria.

6.1.5 Cloning

After PCR amplification of the gene of interest with gene- specific primers containing overhanging ends with restriction enzyme recognition sequences, the PCR product was purified and digested with the appropriate restriction enzymes. The vector backbone was as well digested with restriction enzymes and dephosphorylated using Antarctic Phosphatase from NEB according to the manufacturer's Instruction. Backbone and PCR products were separated by agarose gel electrophoresis and extracted from the gel. PCR product and backbone were ligated; the ligation mix was transformed into bacteria, and plated onto LB agarose plates. After overnight incubation at 37°C, colonies were picked, and grown in 5 ml liquid LB culture overnight. The next day, plasmid DNA was purified and analyzed by restriction digestion for insertion of the PCR product into the vector backbone. Sequences were verified by Sequencing.

6.1.6 Transformation of bacteria

For transformation competent bacteria (DH5α or BL21) [149] were incubated with circular plasmid DNA or ligation mixtures for 20 min on ice, heat-shocked for 1 min at 42°C, recovered on ice for 2 min, incubated with 400 µl SOC media at 37°C, centrifuged at 6000 rpm for 1 min and plated onto LB- Agar plates with appropriate antibiotics in 100-200 µl liquid. The LB- Agar plates were incubated overnight at 37°C.

6.1.7 Plasmid DNA preparation from bacteria

Plasmid DNA was prepared from bacteria using either the Nucleo Spin Plasmid Kit for mini preps or Nuclo Bond PC 100 for midi preps both from Macherey Nagel according to manufacturer's instructions.

6.1.8 Determining DNA Concentration

DNA concentrations were determined spectrometrically using the NanoDrop 1000 Spectrophotometer (Peqlab Biotechnologie GmbH).

6.1.9 Mutagenesis

Mutagenesis PCR was carried out using the either Pfu Turbo Polymerase or PfuUltra II Fusion HS DNA Polymerase from Stratagene according to manufacturer's instructions. Primers were ordered from Biospring (Biospring, Frankfurt). Mutations were verified by DNA sequencing.

6.1.10 DNA sequencing

DNA sequencing was performed by Sequence Laboratories Göttingen (Seqlab, Göttingen). Plasmid DNA and primers were submitted at the suggested concentration and volume.

6.1.11 SDS-PAGE

Proteins were analyzed by Tris-Glycine-SDS gels [150] using the Bio-Rad Mini Gel System. For the Bio-Rad Mini Gel system, gels were poured with 4% stacking (**Table 4**) and the percentage of the resolving gel was chosen according to the size of the protein of interest (**Table 4**). Protein samples were mixed 1: 5 with 5 x SDS sample buffer, incubated for 5 min at 95°C, and spun down for 1 min at 13000 rpm. Bio-Rad Mini Gels were run in home-made 1 x SDS-PAGE buffer. Proteins were separated at 100 V for 15 min followed by 200 V for 45 min with constant current (time varied with % of the gel).

Table 4: **Pipetting scheme for two 12 or 16% SDS separating gels and two 4% stacking gels.**

Compound	Seperating gel (12%)	Seperating gel (16%)	Compound	Stacking gel (4%)
H_2O	3.4ml	2.1ml	H_2O	6.1ml
1.5M Tris/HCl pH 8.8	2.5ml	2.5ml	0.5 M Tris/HCl pH 6.8	2.5ml
30% Acrylamide	4ml	5,3ml	30% Acrylamide	1.3ml
10% SDS	0.1ml	0.1ml	10% SDS	0.1ml
10% APS	0.05ml	0.05ml	10% APS	0.025ml
TEMED	0.005ml	0.05ml	TEMED	0.005ml
Summe	~10ml	~10ml		~10ml

As Tris-Tricine-PAGE gives a higher resolution proteins with a molecular weight less than 30 kDa ACPs and PCPs were analysed by 11% Tris-Tricine-SDS gels [151, 152] (**Table 5**). 11% Tris-Tricine-SDS gels were run in home-made 1 x cathode (inside the chamber) and anode buffer. Proteins were separated at 80 V for 15 min followed by 130 V for 50 min with constant current (time varied with % of the gel).

Table 5: Pipetting scheme for two 11% Tris- Tricine- SDS separating gels and two 4% stacking gels.

Compound	Seperating gel (11%)	Stacking gel (4%)
30% (w/v) Polyacrylamid (37,5:1)	4,4 ml	0,84 ml
Gelbuffer, 3-fold	4 ml	1,5 ml
Glycerol (98%)	1 ml	-
H_2O	2,6 ml	3,61 ml
APS (10%)	0,1 ml	0,045 ml
TEMED	0,001 ml	0,0045 ml
Summe	12 ml	6 ml

6.1.12 Coomassie brilliant blue staining

After SDS-PAGE gel electrophoresis, the gel was placed in Coomassie brilliant blue solution and incubated for 1hr at room temperature, with shaking. The staining solution was then exchanged against water and cooked in the microwave for 1min. After cooling the water was supplemented with 10% acetic acid and destained under shaking for 30min up to ON until the background was completely clear.

6.1.13 Cross linking

300 µl of mixture of HMG-ACP_I and Hal with 2, 8, 20, 40 µM of each protein were mixed with Glutaraldehyd leading to an end concentration of 0.125% and 0.25%. The samples were incubated at RT without any shaking. After 0, 3, 20 and 60 min 40 µl of sample were taken and

supplemeted with 10 µl reducing buffer (5X). Samples were directly frozen in N_2 to slow down the reaction. Samples were not boiled prior to gel running.

6.2 Proteinbiochemical methods

6.2.1 Determination of protein concentration

The protein concentrations of purified proteins were determined as a routine spectrophotometrically using a NanoDrop 1000 Spectrophotometer (Peqlab Biotechnologie GmbH, Erlangen, Germany) and then calculated using the extinction coefficients at 280 nm. The coefficients were determined using Protparam software and are given for the different target proteins which were used most frequently in **Table 6**.

Table 6: Calculated extinction coefficients at 280 nm for the most frequently used target proteins.

Protein	Vector	Tag	Number of amino acids	Molecular weight [g/mol]	ε280 [mg/ml] oxidized state
TycB1- PCP	pBH4	N- terminal H_6,TEV	105	11776.2	0.973
ECH_1	pet24b	C-terminal H_6	262	29412.8	1.169
ECH_2	pet28a	C-terminal H_6	250	27894.9	0.534
Hal	Pet28b	N-terminal H_6	351	40456.8	1.144
CurB	Pet28b	N-terminal H_6	99	10973.7	0.136
$ACP_{I,II,III}$	Pet28a	N-terminal H_6	324	36054.6	1.078
ACP_I	Pet28a	C-terminal H_6	110	12274.7	12274.7

Several proteins such as CurB suffered from a low extinction coefficient. In that case or to verify exact concentrations Bradford was used as an additional method to determine the concentration. Therefore 990 µl of Bradford solution were mixed with 10 µl of protein of interest in solution. After 3 min the sample was measured at 595 nm. The concentration was determined using the Bradford standard curve.

6.2.2 Concentrating the proteins

Concentrating the proteins took place with either an Amicon stirring unit or an amicon centrifugation unit according to the manufacturer's instructions. The pore size of the filter was chosen according to the size of the protein of interest. For ACPs and PCPs with a size ~10 kDa filter units with a cut of of 3 kDa were chosen for bigger proteins a cutoff of 10 kDa was used.

6.2.3 Protein expression

Expression of pet and pMal, pBH4 vectors

In general for protein expression fresh transformation was performed into competent cells of the bacterial strain of interest. Next day 100 mL LB medium supplemented with the appropriate antibiotic(s) ((pMal, pBH4) ampicillin 100 µg/L, (pet28a/b, pet24b) kanamycin 35 µg/L, pRARE

Chloramphenicol 20 µg/L) were inoculated with the transformants. After growth of 2-3 hrs at 37°C and 180 rpm 1 L of media in a 2 L flask, supplemented with the appropriate antibiotic(s), was inoculated with 10-20 ml of the fresh preculture. Cells were then grown at 37°C or 30°C to an OD of 0.5-1 and the temperature was reduced to the required temperature (15°C). Cells had to adapt to the new temperature for 20-30 min and were then induced with 0.4 mM IPTG. After expression the cells were harvested in the Kontron A6.9 for 10 min at 4°C at 6000 rpm. Cell pellets were resuspended in buffer A for Ni-purification and supplemented with protease inhibitor. In case of MBP-tagged proteins the cells were resuspended in amylose buffer A (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM β-Mercaptoethanol, pH 7.8). The different proteins differ in the optimal OD for induction, the temperature used for cell growth and induction and in the optimal time span for expression. Some proteins needed additional supplements in their media e.g the ACPs got modified for expressions longer than 10 hrs at 15°C. All these specialities follow in **Table 7**.

Table 7: Requirements for the expressions of the most frequently used pet constructs

Protein	Bacterial strain/ Supplementation	Media	Cell growth at T (°C)	OD	Expression at T (°C)	Expression (hrs)
$ACP_{I,II,III}$ constructs	BL21 0.05 mM Fe(II)	1L LB	37	0.5-1	15	7-10
Hal	BL21	1.2L XYT	30	1.0	15	13-15
ECH_2	BL21	1L LB	30	0.6	15	13-15
ECH_1	BL21	1L LB	30	0.6	15	13-15
CurB	BL21/pRARE 0.05 mM Fe(II)	1L LB	37	0.7	15	7-10

Labeling For ^{15}N and ^{13}C labeling the LB precultures were spun down at 6000 rpm for 5 min and solubilized in M9 before inoculation. For experiments which needed more than 98% incorporation of the isotopes precultures were done ON in labeled M9 media.

6.2.4 Cell free expression

Continous exchange cell free expression (CECF) expression reactions in P-CF mode were basically performed as previously described in detail [153, 154]. Analytical scale reactions were carried out in 24-well microplates in Mini-CECF reactors. Appropriate dialysis membranes, type

27/32, having a MWCO of 12-14 kDa (Roth, Karlsruhe, Germany) were used to separate reaction mix (RM) from feeding mix (FM). The RM volume was 55 µL and the FM volume 800 µl giving a ratio of 1:15. Optimal concentrations of Mg^{2+} ions for the S30 extract batch were analysed in analytical scale reactions in the P-CF mode. CF expressions were separated into precipitate (P-CF) and supernatant (SNT) by centrifugation at 18.000 x g for 10 min at 4°C. The soluble proteins Hal, ECH_2 and CurA $ACP_{I,II,III}$ were expected in the soluble fraction. Each reaction was performed in doublets. 2 µL of SNT and precipitate of P-CF over expressed protein were analysed by SDS polyacrylamide gel electrophoresis (SDS-PAGE). In **Table 8** is the pipetting sheme for the cell-free reaction of ECH_2 given as an example.

Table 8: Pipetting sheme for the analytical scale P-CF mode cell-free expression of ECH_2. The pipetting scheme is for two analytical scale reactions for reaction mix (RM) volumes of 55 µL and 800 µL of feeding mix (FM).

Compound	Concentration unit		unit	MM (µl)	FM (µl)	RM (µl)
	stock	final				
NaN3	10	0,05	%	8,6		
PEG8000	40	2	%	85,5		
KOAc	4000	150,8	mM	64,5		
Mg(OAc)2	1000	11,1	mM	19		
HEPES buffer	2500	100	mM	62,7		
Complete	50	1	X	34,2		
Folinic acid	10	0,1	mg/ml	17,1		
DTT	500	2	mM	6,8		
NTP mix	75	1	X	22,8		
PEP	1000	20	mM	34,2		
AcP	1000	20	mM	34,2		
AA-mix	4	0,5	mM	213,8		
RCWMDE-mix	16,7	1	mM	102,6		
Sum				**706**		
MM					660,5	45,40
S30 buffer	100	35	%		560	
AA-mix	4	0,55	mM		220	
Pyruvate kinase	10	0,04	mg/ml			0,44
tRNA E. coli	40	0,5	mg/ml			1,38
T7- RNAP	420	6	U/µl			1,57
RiboLockTM	40	0,3	U/µl			1,20
Template DNA	200	0,015	µg/ml			8,30
S30 extract	100	35	%			38,5
H2O					160	13,3
Sum					**1600,5**	**110,10**

6.2.5 Protein purification

General protocol for His-tagged proteins

For purification of histagged proteins, $ACP_{I,II,III}$ and truncations as well as mutants, CurB and TycB1, cell suspensions were thawed, a small amount of Lysozym added and disrupted either by french press (SLM Aminco Instruments Inc. (USA)) or a sonificator (Labsonic U B. Braun Biotech International, Melsungen). After centrifugation (17.000-19.000 rpm, 4°C, 30 min) the supernatant was carefully removed and in case of the recombinant proteins $ACP_{I,II,III}$ and truncations as well as mutants, CurB and TycB1 the recombinant protein was purified by Ni-NTA (Qiagen) affinity chromatography using a FPLC system (Amersham Pharmacia Biotech). Therefore the 5 ml Ni-column was equilibrated in Ni-NTA/IDA-buffer A (50 mM NaPP, 300 mM NaCl, 10 mM Imidazol pH 8.0), the supernatant was loaded using an external pump with a flow rate of 1 ml/min. After washing an increasing gradient of imidazole was employed (10 mM to 240 mM in 20 min, flow rate 3 ml/min). Fractions containing the recombinant protein were identified by SDS-PAGE analysis, combined, and dialysed against the modification buffer (20 mM Hepes, 50 mM NaCl, 20 mM $MgCl_2$, pH 7,5). The proteins were aliquoted, flash frozen in liquid nitrogen and stored at -80°C until usage.

Hal, ECH_2

The lysis of Cur Hal and ECH_2 was performed as described above for histagged proteins. After centrifugation, the recombinant protein present in the SNT was purified by Ni-IDA (GE Healthcare) affinity chromatography using a FPLC system (Amersham Pharmacia Biotech). The 10 ml Ni-IDA column was equilibrated in Ni-NTA/IDA-buffer A (50 mM NaPP, 300 mM NaCl, 10 mM Imidazol pH 8.0), the supernatant was loaded using an external pump with a flow rate of 1 ml/min. After loading the proteins were extensively washed with Ni-NTA/IDA-buffer A. For the purification of Hal the first wash was followed by a 50 mM and a 80 mM Imidazol washing step. Hal was then eluted using 400 mM Imidazol and dialyzed ON against the reconstitution buffer (50 mM PP and 200 mM NaCl, pH 7.5) or gelfiltrated on Superdex 200 with the same buffer. The Halogenase was then reconstituted anaerob following the reconstitution protocol as described under the corresponding section. Freezing of Cur Hal before reconstitution leads to precipitation of protein. ECH_2 was washed with 10 mM followed by 70 mM Imidazol wash and then eluted with 400 mM Imidazol and either dialyzed ON against 20 mM Tris, 100 mM NaCl, pH 7.5 or gelfiltrated on Superdex 200 with the same buffer.

MBP tagged Proteins

For purification of MBP tagged proteins, cell suspensions were thawed, a small amount of lysozym added and disrupted either by french press (SLM Aminco Instruments Inc. (USA)) or a sonificator (Labsonic U B. Braun Biotech International, Melsungen). After centrifugation (17.000-19.000 rpm, 4°C, 30 min) the supernatant was carefully removed and the recombinant protein

was purified via amylose resin. Therefore the column was equilibrated with amylose buffer A (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 10 mM β-Mercaptoethanol, pH 7.8) after supernatant loading (1 ml/min) using an external pump the protein was extensively washed and then eluted using amylose buffer B (20 mM Tris-HCl, 200 mM NaCl, 10 mM Maltose, 1 mM EDTA, 10 mM β-Mercaptoethanol, pH 7.8)

Anaerob reconstitution of Hal

For the reconstitution glass and buffers were exposed to nitrogen atmosphere in a tent for a minimum of three days, plastic had to be exposed to nitrogen atmosphere for a minimum of five days. After Ni-purification Cur Hal was either gelfiltrated or dialysed against 50 mM PP, 200 mM NaCl, pH=7.5. Then the protein was concentrated in a stirring amicon to concentrations of 12 mg or higher. The protein was transferred to a glass vial and oxygen removed by three alternating cycles of nitrogen aeration and vacuum application (More exact one cycle was composed of nitrogen aeration for 30 s followed by vacuuming for 1 min, this three times repeated and then 5 min vacuum were applied to the protein. Then the next cycle followed). In the tent the protein was then supplemented with 0.5 mM α-KG and 0.75 mM $Fe(NH_4)_2(SO_4)_2$ and stirred for 30 min at 25°C. The protein was then centrifuged to remove precipitated protein (10 min, 4°C at 10 krpm) and dialyzed against nitrogen treated buffer which was chilled in nitrogen environment for five days (buffer 20 mM Hepes, 50 mM NaCl, α-KG 0.5 mM (freshly added) pH=7). After that the protein was centrifuged again for 10 min at 10 000 rpm. Finally protein was at concentrations of approximately 100 µM and frozen in closed glas vials. Shock freezing of aliquots in eppendorf tubes in liquid nitrogen if proper closing was assured and stored at -80°C. Contact with air had to be prevented prior freezing because it led to precipitation and unfunctionality.

6.2.6 Size exclusion chromatography

SEC with prepacked columns (GE Healthcare Europe, München, Germany) was used in order to assess the homogeneity and the stability in diverse buffer systems of different proteins on a superose 6 column (10_300) at 16°C. Usually 100 µl with concentrations ranging from 50 µM-1 mM were loaded and run at 0.5 ml/min. Protein preparations were centrifuged at 13 000 x rpm for 10 min or filtered (0.2 µm) prior to loading. For preparative scale 5 ml of protein were loaded on a preparative superdex 200 (GE Healthcare Europe, München, Germany) and run at 1.5 ml/min at 4°C on an Äktapurifier station (GE Healthcare Europe, München, Germany). Absorption was monitored at 280 as well as at 260 nm during the run.

6.2.7 Modification of ACP constructs

The recombinant ACPs and PCP (200 µM) were artificially loaded with either coenzyme A (CoA) or analogs (Acetyl-, Malonyl-, HMG-CoA) by using the promiscuous *B. subtilis*

phosphopantetheinyl (Ppan) transferase Sfp (3 µM). The loading reaction was incubated at RT for 1 hr in buffered aqueous solution (pH 7.5, 20 mM HEPES, 50 mM NaCl, 20 mM $MgCl_2$). To ensure that complete loading was achieved the reaction mixture was analyzed directly by MALDI-MS.

6.2.8 Analysis of the iron content

The content of the iron cofactor for the halogenase was determined using the chromogen Ferene S assay as previously published [155]. To the corresponding halogenase (10 nmol in 800 µL dd. H_20) concentrated hydrochloric acid (100 µL) was added and incubated in an Eppendorf Thermomixer (10 min, 400 rpm, 25°C). Proteins were precipitated by addition of 80% trichloroacetic acid followed by centrifugation (13.000 rpm, 10 min, RT). The supernatant was transferred into an acid washed clean glass tube and 200 µL of a 45% (w/v) sodium acetate solution was added. 1.8 mL of the Ferene S solution (0.75 mM Ferene S, 10 mM L- ascorbic acid, 45% (w/v) sodium acetate) was admixed and 1 ml of the mixture was measured photometrically at α=593 nm. The obtained absorbance values were compared with the recorded calibration curve (**App. 1**), which was recorded analogously with distinct iron concentration. $FeCl_3$ was used as iron source and the amount was varied between 0 and 20 nmol.

6.3 Spectroscopical and spectrometrical methods

6.3.1 Nuclear magnetic resonance (NMR)

Structural studies and interactions studies

0.6 -1 mM solutions of ACP_I were prepared in 50 mM arginine/50 mM glutamate buffer at pH 6.8 containing 5% D_2O and 0.15 mM DSS (4, 4-dimethyl-4-silapentane-1-sulfonate). NMR spectra were collected at a temperature of 291 K on 950, 900, 800 and 600 MHz on Bruker Avance spectrometers equipped with cryogenic 5 mm z-axis gradient triple resonance probes and a 500 MHz Bruker Avance spectrometer equipped with 5 mm x,y,z-axis gradient triple resonance room temperature probe. For backbone and aliphatic side chain assignments [^{15}N,^{1}H]-TROSY [156] version of HSQC, HNCACO, HNCO, HNCACB, (H)C(CCO)NH-TOCSY, and H(CCCO)NH-TOCSY were recorded. Aromatic ring resonances were assigned with (H)CB(CGCC-TOCSY)H^{ar} experiments [157]. ^{15}N and ^{13}C separated 3D NOESY experiments were recorded with 70 ms mixing time. All spectra were processed with TopSpin (Bruker Biospin) and analysed with Sparky [158]. For the acyl chain binding studies, standard sensitivity-enhanced [^{15}N,^{1}H]-TROSY spectra were acquired. Steady state {^{1}H}^{15}N hetero NOE measurements were performed on a 600 MHz instrument using a TROSY-type pulse sequence [159].

To observe NOEs between the 4'-Ppant arm and HMG loaded 4'-Ppant arm and the ACP domain highly ^{13}C, ^{15}N enriched ACP_I was loaded with unlabeled ^{12}C-CoA or HMG-CoA. The protein was concentrated to 1 mM in 50 mM arginine/ 50 mM glutamate buffer at a pH of 6.8 and a 3D F1-$^{13}C/^{15}N$-filtered, F3-^{13}C-separated NOESY-HSQC was acquired at 800 MHz using a mixing time of 80 ms and filter delays of 93 Hz for $^1J_{NH}$ and to 125 Hz and 147 Hz for $^1J_{CH}$. For the assignment of the unlabeled cofactor a 2D F1/F2-$^{13}C/^{15}N$ double-filtered NOESY experiment [160] with 150 ms mixing time was acquired at 950 MHz. Fully $^{13}C/^{15}N$-labeled ACP_I was loaded with unlabeled ^{12}C-HMG-CoA and the protein was concentrated to 1 mM in 25 mM NaPi and 50 mM NaCl at a pH of 6.8 and measured at 291K.

6.3.2 Titration experiments

Interaction between the different ACP_I forms and their interaction partners (ECH_2, Hal) were investigated by NMR titration experiments. [^1H,^{15}N]-TROSY were recorded of ^{15}N-labeled protein alone (either HMG-ACP_I, apo ACP_I or apo TycB1) and then the unlabeled interaction partner (ECH_2, Hal) was added in several steps (([^{15}N]ACP)1:1(Interaction partner); 1:2, 1:4, 1:8 molar ratio) (Lower ratios were not necessary due to the weak and transient nature of the interaction). The amount of labeled protein ranged between 100-300 µM depending on the solubility of the interaction partner. Chemical shift changes were monitored by [^1H,^{15}N]-TROSY experiments, measured with spectral widths of 400 points and 1536 points in the ^{15}N and ^1H dimensions, respectively

6.3.3 Spectra analysis

NMR spectra were processed with Top Spin 2.1 (Bruker Biospin) and analysed using Sparky Version 3.114 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco).

Structural calculation

Backbone torsion angle restraints were obtained from chemical shift data using the TALOS+ algorithm [161]. NOE-based distance restraints from 3D ^{15}N-NOESY-HSQC and ^{13}C-NOESY HSQC spectra were assigned automatically by CYANA [162, 163], which was also used for the structure calculations by torsion angle dynamics [164]. The final structure calculations included 2852 distance restraints and 162 backbone torsion angle restraints (**App. 10**). 100 conformers were computed using 10000 torsion angle dynamics steps. The 20 conformers with the lowest target function values were subjected to restrained energy refinement with the program OPALP [165] using the AMBER force field [166]. The quality of the structures was checked by PROCHECK [167] and WHATCHECK [168].

For the structure calculations with HMG a model of HMG was created using the online version of the program CORINA for 3D structure generation (http://www.molecular-

networks.com/online_demos/corina_demo). HMG was attached covalently to the Ser1989 residue. For the structure calculation with CYANA [169], a residue consisting of a serine backbone connected to the HMG cofactor was added to the standard residue library and the dataset was complemented with additional NOE information for the cofactor. The resulting structures were subjected to restrained energy minimization using OPALp [165]. The necessary partial charges for HMG-serine were calculated using the PRODRG server [170].

Molecular imaging

Molecular graphic images were produced using pyMOL Molecular Graphic system (2007, Delano DeLano Scientific LLC)

Hetero NOE experiment

Steady state $\{^1H\}^{15}N$ hetero NOE measurements were performed on a 600 MHz instrument using a TROSY-type pulse sequence [159]. The sample was 0.4 mM ^{15}N-labeled holo CurA ACP_I-ACP_{II}-ACP-$_{III}$ in 50 mM arginine and 50 mM glutamate buffer at pH 6.8 and 291 K. For the measurements two standard $\{^1H\}^{15}N$-hetNOE-TROSY experiments were performed, one with presaturation of the protons and one control experiment without presaturation (reference spectra without hetNOE), both spectra were recorded in an interleaved manner. The spectra were recorded with spectral widths of 288 points and 1536 points in the ^{15}N and ^1H dimensions, respectively. For analysis peak heights were used instead of peak volumes due to the high overlay of the peaks. For hetNOE calculations the peak intensitities (I_{het}, I_{ref}) of both spectra were multiplied by 2^{NCproc} (where the NC proc is an internal scaling factor of bruker). Then the hetNOE value could be calculated by deviding the modified I_{het}/I_{ref}. The relaxation data obtained by the hetNOE measurements give evidence about the motional properties of the amide groups of the protein. Hetero NOE values of around 0.8 represent highly structured regions whereas hetero NOE <0.4 indicate unstructured, highly flexible regions.

Tract measurement

The rotational correlation time also called tumbling rate is a fundamental parameter to draw conclusions about the multimeric state; in addition it helps to draw conclusions about the relaxation behaviour. The TROSY for rotational correlation times (TRACT) is one fast method to determine τc value [171]. The tract method is based on the TROSY principle and allows determining τc without interference from dipole-dipole relaxation and chemical exchange induced relaxation. Therefore a series of one-dimensional experiments are recorded. This allows extracting the R_2 relaxation rate upon integration over the amide region. The obtained τc is an averaging of fast and slow relaxing components occurs during integration. The more precise τc can be achieved with a two-dimensional version of the TRACT experiment. This leads to the possibility to distinguish between highly flexible loop regions and highly structured regions which mostly display faster relaxation.

6.3.4 Circular dicroisme (CD) spectroscopy

Proteins were purified as described before. Concentrated proteins were dialysed at 4°C against high salt buffers and then diluted prior to CD measurements to 10 µM giving the buffer conditions below. CD spectrometry and melting curves were measured with a Jasco J-180 spectropolarimeter (Jasco Labortechnik, Gross-Umstadt, Germany) in the CD-buffer giving below. Experiments were carried out at standard sensitivity with a band width of 1 nm and a response of 1 s. The data pitch was set to 0.1 nm and a scanning speed of 100 nm/min was used. CD spectra were recorded from 190 to 250 nm at 20°C in a cuvette of 1 mm cell length. The data represent an average of three to five accumulations and were baseline corrected by subtraction of a buffer spectrum recorded under identical conditions. Melting curves were measured with a band width of 1 nm, a response of 1 s and a data pitch set to 0.1 °C at a wavelength of 208 nm with a slope of 2°C/min from 4°C to 95°C. For CD measurements the following conditions were used: ECH_1 (20 mM Na_2HPO_4, 50 mM NaCl at pH 6.5); ECH_2 (20 mM Na_2HPO_4, 50 mM NaCl at pH 7.4); ACPs (2 mM Hepes, 5 mM NaCl, pH= 7.5); Hal (10 mM Hepes, 25 mM NaCl, pH= 7.2)

6.3.5 Matrix-assisted laser desorption/ionization mass spectrometry (MALDI- MS)

Mass spectrometry was performed by Dr. Bahr in the group of Prof. Karas. Mass spectra were recorded on a MALDI-TOF mass analyzer (Voyager STR, Applied Biosystems, Framingham MA, USA) in linear positive ion mode. Samples were diluted with water to a concentration of 5-10 µM and 1 µL was mixed with 1 µL MALDI matrix (sDHB, Bruker, Germany) directly on the sample target. Spectra were accumulated over 50-100 laser shots.

6.3.6 Fluorescence anisotropy

Fluorescence anisotropy measurements with Hal, ECH_2 and fluorophore labeled ACP_I were performed using a Jasco spectrofluorometer FP-6500 (Jasco Labortechnik, Gross-Umstadt, Germany). ECH_2, Hal and ACP_IG89C as well as TycB1 PCP as a control were expressed and purified as described before. For the fluorescin labeling the cysteine mutants were Ni-purified as usual. After dialyzing them against the fluorescin labeling buffer (20 mM MES, 100 mM NaCl, pH= 6) the proteins were incubated with 0.5 mM TCEP at least two hours on ice. Then 10 molar- excess of fluorescin-malemide (AAT Bioquest, Sunnyvale (USA)) (dissolved in DMSO) were added and the protein incubated for 1 hr at RT. Afterwards the buffer was exchanged on a PD10 column against the anisotropy assay buffer (for Hal: 20 mM Hepes, 50 mM NaCl, pH= 7.5, 0.5 mM α-KG; for ECH_2: 20 mM Tris, 100 mM NaCl, pH= 7.5). Fractions were checked for labeled protein using Bradford. MALDI-MS was performed to check for the amount of labeled protein. Longer incubation times as well as higher pH turned out to lead to the attachment of 1-3 fluorophores and are therefore crucial. For the measurement excitation

and emission maxima of the f-ACP_I have been determined for 485 nm for 518 nm, respectively. The G-factor of the ligand was measured in assay buffer (Hal: 20 mM Hepes, 50 mM NaCl, pH= 7.5, 0.5 mM α-KG; for ECH_2: 20 mM Tris, 100 mM NaCl, pH= 7.5) with a value of 1.50. For saturation binding, 60 nM of f-apo-ACP_I or f-HMG-ACP_I was incubated with increasing concentrations of ECH_2 or Hal (0.5- 400 µM) for 1 hr at RT in a final volume of 500 µL. Anisotropy was determined as mean value of ten repetitions per data point. The Anisotropy of f-ACP_I without the interaction partner was subtracted from each point for normalization. Data were processed by non-linear regression curve fitting using Origin 8.5 (OriginLab Corporation, Northampton, MA 01060 USA).

6.3.7 Multi angle light scattering measurement (MALS)

Size exclusion chromatography (SEC) is well suited as basic means to analyse protein homogeneity as well as its stability over time. Nevertheless for proteins which have not a global fold the MW determination via SEC can be strongly influenced and gives therefore higher molecular masses. Multi angle light scattering (MALS) determines the absolute molecular mass as well as the oligomerization state by correlating molecular mass and concentration of the protein to the amount of scattered light. This technique does not take into account the shape of proteins [172]. Combined SEC-light scattering analysis was performed on a Superdex 75 5/150 column (GE Healthcare Europe, München, Germany) at a flow rate of 0.3 ml/min on a Jasco HPLC unit (Jasco Labor- und Datentechnik GmbH, Gross-Umstadt, Germany) connected to a light scattering detector measuring at three angles (mini Dawn Treos, Wyatt Technology Europe GmbH, Dernbach, Germany) and a refractive index detector (Optilab rex, Wyatt Technology Europe GmbH, Dernbach, Germany)). The column was equilibrated for at least 16 h with SEC buffer (20 mM HEPES pH 7.5, 50 mM NaCl, filtered through 0.1 µm pore size VVLP filters (Millipore, Eschborn, Germany). The Proteins were Ni-purified followed by gelfiltration. 20 µL of protein samples (4 mg/mL) were separated on the column. Data analysis was accomplished using the ASTRA software package 5.3.4.13 (Wyatt Technology Europe GmbH, Dernbach, Germany).

6.4 Analysis of functionality

6.4.1 ECH_1 and ECH_2 activity test

For the Activity test 5 µM ECH_1 and or 5 µM ECH_2 were mixed with 100 µM HMG-ACP_I or 500 µM HMG-CoA. In case of HMG-ACP_I as a substrate the reaction was incubated for 3.5 hrs at 25°C. Using HMG-CoA the reaction was incubated for 12 hrs at 37°C. The reactions were analysed using MALDI-MS. In case of HMG-CoA high performance liquid chromatography (HPLC) was used to characterize substrates and products by retention time on a chromatography column followed by mass spectrometric analysis. Reversed-phase (RP)

chromatography is based on hydrophobic interactions with the non-polar stationary phase (carbon, C18 or C4 coated silica gel). Elution is mediated by non-polar acetonitrile, which competes with the adsorbed analytical compounds for non-covalent binding positions. For HPLC analysis a PerfectSil 250x 4.6 C18 column was used in a Jasco HPLC System. The running buffers were H_2O with 0.1% TFA and acetonitril with 0.1% TFA. Prior to running the 500 µl samples were filtered with a 0.2 µm cellulosemembrane. The elution was done following the gradient described in **Table 9** with a flowrate of 1 ml/min at 25°C. Fractions of 500 µl were collected and subjected to electron spray ionization mass spectrometry (ESI- MS).

Table 9: Elution profile for the seperation of the products of the ECH_1 and ECH_2 activity test.

Time (min)	H_2O (%)	Acetonitril (%)
0	98	2
2	98	2
20	80	20
25	50	50
26	2	98
29	2	98
30	98	2
37	98	2

6.4.2 Cur Hal activity test

For the Halogenase activity test Cur hal was purified and reconstituted as described before. 100 µl of 100 µM HMG-ACP_I in 20 mM Hepes, 50 mM NaCl, 20 mM $MgCl_2$ pH 7.2 were supplemented with 0.5 mM α-KG and 75 µM Fe-II- SO_4 and 5 µM Hal. ($MgCl_2$ was not essential for the reaction but derived from the previous modification). The reaction was incubated at 25°C open and under shaking (800 rpm) for 10 min and 1 hr. The reaction products were analysed using MALDI-MS.

6.5 Segmental labeling

6.5.1 Construction of vectors

Plasmids for the expression of $ACP_{I, II, III}$ from CurA

The three consecutive ACP domains ($ACP_{I, II, III}$) located at the C-terminus of the polypeptide CurA were amplified from genomic DNA of *Lyngbya majuscula* strain 19L (a gift from Dr. C. Walsh, Harvard University, USA) using the oligonucleotides: #CurAT1T2T3-*NdeI*-F (5'-GGGCATATGG GCAGCAGCCA TCATCATCA), #CurAT1T2T3-*XhoI*-R (5'-GGCTCGAGTC ACAGCTTAGA ACCACCAGTA). The PCR product was digested with *NdeI* and *XhoI* and cloned into pET-28b(+) vector (Novagen), resulting in the new expression vector pet28-01.

Plasmid for three-fragment ligation of CurA

The first domain (ACP_I) of the three consecutive ACP domains of the CurA protein was amplified using the primers: #I1CurAT1-*NcoI*-F (5'-GGCCATGGCC ACTCCTCAGG TAAATCAAGT) and #I1CurAT1-*BamHI*-R (5'-CCGGATCCCG GTTTGGTACC CTGAGAGCTC A). For the construction of the N-terminal precursor protein (H_6-Smt3-ACP_I-*Npu*IntNΔC15), the PCR product was digested with *NcoI* and *BamHI* and cloned into pHYRSF66-36, resulting in the new expression vector pABRSF-1. pHYRSF66-36 was derived from pHYRSF49-36 by eliminating *NcoI* site using the two oligonucleotides #HK181 (5'-CTTTAATAAG GAGATATAACA TGGGCAGCAG) and #HK182 (5'-CTGCTGCCCA TGTTATATCT CCTTATTAAAG) and by replacing the *NdeI* site with *NcoI* site using the two oligonucleotides #HK183 (5'-CAGATTGGTG GTTCCATGGA GTACAAACTT ATCC) and #HK184 (5'-GGATAAGTTT GTACTCCATG GAACCACCAA TCTG). This plasmid expresses H_6-Smt3- ACP_I -*Npu*IntNΔC15 fusion protein upon induction with isopropyl-D-thiogalactoside (IPTG). The second domain (ACP_{II}) of the CurA protein was amplified using the primers: #I2CurAT2-*KpnI*-F (5'-CCGGTACCCA GCAATCTCTG AAAA) and #I2CurAT2-*HindIII*-R (5'-CCAAGCTTTT ATTTAGTGCC CTGAGAGGCC A). The PCR product was digested with *KpnI* and *HindIII* and cloned into the vector pSABAD120. The pSABAD120 was constructed by transferring the inserted gene in pSARSF120 into pSKBAD2 by using the restriction sites of NdeI and HindIII. The new expression vector pABBAD-02 expressed the fusion protein of the C-terminal 15 residues of NpuDnaE-IntC with ACP_{II} (H_6-*Npu*IntC15-ACP_{II}) for the ligation with ACP_I. In order to construct the central precursor protein for the three-fragment ligation, the plasmid pABRSF03 was constructed by transferring the gene of NpuIntC15 and ACP_{II} into pHYRSF49- 36 by using NdeI and BamHI sites. The resulting plasmid pABRSF-3 expresses H_6- Smt3-*Npu*IntC15-ACP_{II}-*Npu*IntNΔC36 fusion upon induction with IPTG. For the C-terminal precursor, the gene of the third ACP domain (ACP_{III}) of the CurA protein was amplified using the two primers #IBCurAT3-*KpnI*-F (5'-CCGGTACCAAAACCCTGCAG CCGCTGCC) and #IBCurAT3-*HindIII*-R (5'-GGAAGCTTTC ACAGCTTAGA ACCACCAGTA GCAGC). The PCR product was digested with KpnI and HindIII and cloned into the vector pHYRSF1-02. The new expression vector pABRSF-2 expressed the fusion protein of H_6-*Npu*IntC36-ACP_{III} under the control of T7 promoter. The plasmid pABBAD-15 was constructed by transferring the gene encoding *Npu*IntC36-ACP_{III} into pBAD vector using *NdeI* and *HindIII* sites. The resulting plasmid pABBAD-15 expresses the fusion protein of *Npu*IntC36-ACP_{III} upon induction with L-arabinose

6.5.2 Segmental isotopic labeling of a central domain of CurA ($ACP_{I,II,III}$)

Protein expression and purification of CurA ($ACP_{I,II,III}$) and H_6-Smt3-ACP_I-*Npu*IntNΔC15 apo-CurA ($ACP_{I,II,III}$) was expressed in *E. coli ER2566* using the plasmid pET28-01. 40 ml of overnight culture were spinned down at 1000×g and inoculated into 2 L ^{15}N labeled M9 medium supplemented with 25 mg/ml kanamycin and grown at 37°C. The M9 media contains 0.05 mM

$FeCl_2$ which inhibits the apo to holo conversion of the protein. The cells bearing the expression vector were grown to an OD of 0.6 at 37°C and then transferred to 20°C and induced with 0.4 mM IPTG at an OD of 0.7- 0.9. After 6 hours induction with IPTG, the cells were harvested. The protein was purified by affinity- chromatography using Ni-NTA superflow (QIAGEN). Total amount of purified protein was estimated to be 32 mg. Unlabeled H_6-Smt3-ACP_I-*Npu*IntNΔC15, which is a precursor protein for the three-fragment ligation, was expressed using the plasmid pABRSF-1 in 1.4 L of LB medium supplemented with 25 g/ml kanamycin and 0.05 mM $FeSO_4$. The *E. coli* ER2566 cells bearing pABRSF-1 were grown to an OD of 0.6 at 37°C and then transferred to 30°C, followed by the induction with a final concentration of 0.4 mM IPTG at an OD of 0.7-0.8. The cells were grown for another 6 hours and harvested by centrifugation at 9,000 ×g for 15 min. The pellet was frozen and stored at –80°C for further purification.

The first step *in vivo* protein ligation for central fragment labeling of CurA ($ACP_{I,II,III}$)

E. coli ER2566 bearing the two plasmids pABRSF-03 (H_6-Smt3-*Npu*IntC15-ACP_{II}-*Npu*IntNΔC36) and pABBAD-15 (NpuIntC36-ACP_{III}) were grown at 37°C in 2 L of LB medium supplemented with two antibiotics 100 µg/ml ampicillin and 25 µg/ml kanamycin and 0,05 mM $FeSO_4$. The C-terminal precursor protein (*Npu*IntC36-ACP_{III}) was first induced with 0.1% (w/v) L-arabinose at an OD of 0.4-0.5 for 4 hrs and transferred to another shaker adjusted to 30°C. The culture medium was replaced with 2 L of ^{15}N-labeled M9 medium supplemented with 25 µg/ml kanamycin, 0.05 mM $FeCl_2$ and a final concentration of 0.4 mM IPTG. The replacement of the medium was done by spinning down the cells at 900×g with a rotor SLA-3000. The cell pellets were gently resuspended with 100 ml of the labeled M9 medium and then span down once more at 900×g for 10 min. The cell pellets were then resuspended with 2 L of the ^{15}N-labeled M9 medium. After further 5 hours of the second induction, the cells were harvested at 9,000 ×g for 10 min and frozen at –80°C. 20 mg of the segmentally isotope-labeled ligated product of H_6-Smt3-*Npu*IntC15-ACP_{II}-ACP_{III} was obtained after IMAC purification.

The second step *in vitro* protein ligation for central fragment labeling of CurA ($ACP_{I,II,III}$)

The purified two precursor proteins, H_6-Smt3-*Npu*IntC15-ACP_{II}-ACP_{III} and H_6-Smt3-ACP_I-*Npu*IntNΔC15 were dialyzed against PBS, pH 7.4. About 20 mg of H_6-Smt3-*Npu*IntC15-ACP_{II}-ACP_{III} were mixed with a total amount of 5 mg H_6-Smt3-ACP_I-*Npu*IntNΔC15. The final concentration of H_6-Smt3-*Npu*IntC15-ACP_{II}-ACP_{III} was 50 µM in the reaction mixture. H_6-Smt3-ACP_I-*Npu*IntNΔC15 had a final concentration of 11 µM in the reaction mixture. The reaction mixture was incubated at room temperature in the presence of 0.5 mM TCEP for 10 hrs. In the reaction mixture yeast ubiquitin-like protein-specific protease 1 (Ulp1) was added for removal of the H_6-Smt3 fusion tag. After 10 hours the buffer was exchanged with 10 mM Tris, 500 mM NaCl, 0.5 mM EDTA, 0.5 mM TCEP using a centrifugal filter unit (MWCO 5000) (Amicon). In this ligation buffer (10 mM Tris, 500 mM NaCl, 0.5 mM EDTA, 0.5 mM TCEP), the reaction was continued for further 7 hrs at 25°C with agitation at 200 rpm. After the reaction, the H_6-Smt3

fusion tag was removed by IMAC. The flow through from IMAC was collected and concentrated by a centrifugal filter unit (MWCO 5000) (Amicon). The buffer was exchanged to 10 mM Na phosphate, pH 7.2 for further purification with ion exchange chromatography. The ligated product was purified by anion exchange chromatography using MonoQ 5/50GL. Fractions were collected and dialysed for over night against 5 L of 50 mM Na phosphate, 100 mM NaCl, pH 7.2. The protein concentrated to 500 µl with a centrifugal filter unit for NMR measurements. The total amount of segmentally ^{15}N-labeled ACP$_{I}$-[^{15}N]-ACP$_{II}$-ACP$_{III}$ was 1.7 mg.

6.5.3 Segmental isotopic labeling of ACP$_{I}$ domain in CurA (ACP$_{I,II,III}$)

Protein expression and purification of labeled H$_6$-Smt3-ACP$_{I}$-*Npu*IntNΔC15

The N-terminal precursor protein was expressed in *E. coli* ER2566 using the plasmid pABRSF-1. 40 ml of the 100 ml pre-culture were spun down at 1,000×g and inoculated into 2 L ^{15}N-labeled M9 medium supplemented with 25 µg/ml kanamycin and grown at 37°C. The M9 media contains 0.05 mM $FeCl_2$ which inhibits the apo to holo conversion of the protein. The cells bearing the expression vector were grown to an OD of 0.5 at 37°C and then transferred to 20°C and induced with 0.4 mM IPTG at an OD of 0.6-0.7. After 6 hrs induction with IPTG, the cells were harvested. The protein was purified by Immobilized Metal ion Affinity Chromatography (IMAC) using Ni-NTA superflow (QIAGEN). Total amount of purified protein was estimated to be 40 mg.

Co-expression and purification of unlabeled C-terminal precursor H$_6$-Smt3-*Npu*IntC15-ACP$_{II}$- ACP$_{III}$

E. coli ER2566 bearing the two plasmids pABRSF-3 (H$_6$-Smt3-*Npu*IntC15-ACP$_{II}$-*Npu*IntNΔC36) and pABBAD-15 (NpuIntC36-ACP$_{III}$) were grown at 37°C in 2 L of LB medium supplemented with two antibiotics 100 µg/ml ampicillin and 25 µg/ml kanamycin and 0.05 mM $FeSO_4$. At an OD of 0.4 cells were transferred to another shaker adjusted to 30°C. The C-terminal precursor protein (NpuIntC36-ACP$_{III}$) was first induced with 0.1% (w/v) L-arabinose at an OD of 0.5- 0.6 for 1 hr. Then the N-terminal precursor protein was induced with 0.4 mM IPTG. The two proteins were co-expressed for further 7 hrs. The cells were harvested at 9,000×g for 10 min and frozen at –80°C. 160 mg of the unlabeled, ligated product of H$_6$-Smt3-*Npu*IntC15-ACP$_{II}$-ACP$_{III}$ were obtained after IMAC purification.

The *in vitro* protein ligation for N-terminal ACP$_{I}$ domain labeling of CurA (ACP$_{I,II,III}$)

The purified two precursor proteins, H$_6$-Smt3-NpuIntC15-ACP$_{II}$-ACP$_{III}$ and ^{15}N-labeled H$_6$-Smt3-ACP$_{I}$-*Npu*IntNΔC15 were dialyzed against ligation buffer (10 mM Tris, 500 mM NaCl, 0.5 mM EDTA) for the second step *in vitro* ligation. About 42 mg of H$_6$-Smt3-*Npu*IntC15-ACP$_{II}$-ACP$_{III}$ were mixed with a total amount of 40 mg H$_6$-Smt3-ACP$_{I}$-*Npu*IntNΔC15. The final concentration of H$_6$-Smt3-*Npu*IntC15-ACP$_{II}$-ACP$_{III}$ and H$_6$-Smt3-ACP$_{I}$-*Npu*IntNΔC15 was 50 µM in the presence of 0.5 mM Tris (2-carboxyethyl)phosphine (TCEP). The reaction mixture was

incubated at room temperature for 13 hrs with agitation at 200 rpm. Then the yeast ubiquitin-like protein specific protease 1 (Ulp1) was added for removal of the H_6-Smt3 fusion tag and incubated for further 12 hrs. After the reaction, the H_6-Smt3 fusion tag was removed by IMAC. The flow through fraction from IMAC was collected and concentrated by a centrifugal filter unit (MWCO 10000) (Amicon). 30 mg of the ligated protein were obtained and the buffer was exchanged to 10 mM Na phosphate, pH 7.2 for further purification with ion exchange chromatography. The ligated product was purified by anion exchange chromatography using MonoQ 5/50GL. The fractions containing the ligated protein were collected and dialysed for over night against 5 L of 50 mM Na phosphate, 100 mM NaCl, pH 7.2. For the NMR measurements, the protein was concentrated to 0.4 mM using a centrifugal filter unit for NMR measurements. The total amount of segmentally [^{15}N, ^{15}C]-labeled [^{15}N/^{13}C]-ACP_I-ACP_{II}-ACP_{III} was 13 mg.

6.5.4 Nuclear magnetic resonance spectroscopy

All the NMR measurements were performed on either a Bruker Avance Spectrometer or Varian INOVA spectrometer at ^{1}H frequency of 600 MHz, which are equipped with cryogenic probe heads. The spectra were recorded with 80 mM sample for the segmentally isotope- labeled ACP_I-[^{15}N]ACP_{II}- ACP_{III} and with a 0.4 mM sample for uniformly ^{15}N-labeled CurA (ACP_I-ACP_{II}-ACP_{III}) at 18°C. The sample conditions were 50 mM Naphosphate, pH 7.2, 100 mM NaCl with 1 mM $^2H_{12}$- DTT.

7 Results

7.1 Expression and purification of CurA $ACP_{I,II,III}$ interaction partners

It is known that eight to ten different domains interact during biosynthesis with the triplet ACP (compare **Fig. 15**). For structural investigations we started to work on all enzymes of the assembly line and excluded during target evaluation the ones with too low solubility. The 3-hydroxy-3-methyl-gluconate (HMG) cassette (HCS (CurD), KS (CurC), ACP (CurB)) is responsible for producing the β-methylated hydroxymethyl-glutaryl (HMG). Attempts to increase the solubility of the HMG-ACP synthase (HCS) and the lone standing ketosynthase (KS) to turn them into target proteins for structural investigations are described briefly in **7.1.1**, the CurB ACP is described in **7.1.7**. Similarly, we tried to obtain the ketosynthase-acyltransferase (AT-KS) didomain soluble for interaction studies. These efforts are described in **7.1.2**. We optimized as well the expressions and purifications of the chlorinating enzyme Cur Hal (**7.1.6**) the dehydratase ECH_1 (**7.1.4**) and the decarboxylase ECH_2 (**7.1.5**), these three enzymes were used for further investigation covering NMR titration experiments, fluorescence anisotropy and activity assays all described starting from **7.4**.

7.1.1 The hydroxymethyl-gluconate cassette (HCS and KS)

The genes CurC (GenBank accession number: AAT70098.1) and CurD (GenBank accession number: AAT70099.1) encode homologs of the 3-hydroxyl-3-methyl-glutaryl-CoA synthase, the KS (CurC) and the HMG-CoA synthase (HCS, CurD), respectively. The HMG-CoA synthase acts *in trans* on the β-ketoacyl-S-ACP intermediate to add the C2-carbanion of acetyl-S-ACP into the ketone, the acetyl group is delivered by the single ACP CurB (**7.1.7**). The expression constructs of HCS and KS were obtained as a gift from the Sherman lab. The full length gene CurD coding for HCS was cloned into pet24b (using *NdeI* and *XhoI*) giving raise to a C-terminal His_6-tag. The full length gene CurC coding for the KS was cloned into pet28b (using NdeI and XhoI) giving raise to an N-terminal His_6-tag. The constructs were expressed at 15°C, 25°C and 37°C in standard BL21 *E. coli* cells and Arctic Express *E. coli* cells, coexpressing cold adapted chaperons. In all cases no soluble protein was be obtained. To improve solubility HCS and KS were cloned into a petM-60 vector (EMBL) (**App. 2**) in which the NusA leader was substituted with a modified ubiquitin between the *NcoI* and *BamHI* sites (**App. 3**). The final constructs contain ubiquitin as an N-terminal fusion protein with a linker containing His_6-tag and TEV cleavage site. We cloned both proteins into a modified pMAL-c2x vector giving an N-terminal fusion with the maltose binding protein (MBP) via a poly N-linker containing an additional His_6-tag and TEV cleavage site (**App. 4**). In both cases the protein was still expressed in inclusion bodies. In a next step we aimed for inclusion body expression (expressed for 4 hrs at 37° in BL21 *E. coli* cells) and Ni-purified them in denaturating conditions, followed by refolding (1:200

dilution into 50 mM NaPP, 400 mM Arg, 200 mM NaCl, 0.5 mM EDTA, 5 mM Gluthathion (red) and 0.5 mM Gluthathion (ox)). Protein quality was verified by 16% SDS-PAGE (**Fig. 28**). From 1L XYT 18 mg Ub-HCS and 36 mg Ub-KS were obtained with 20-30% impurities.

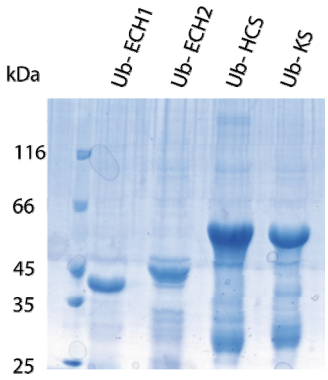


Fig. 28: SDS analysis of Ub- fusion proteins. Proteins were expressed in inclusion bodies and Ni-purified in a one step protocol in denaturating conditions.

7.1.2 The ketosynthase and acyltransferase didomain

The ketosynthase and acyltransferase (KS-AT) didomain of CurA and the KS and AT of DEBS modul 3 (PDB: 2QO3, **Fig. 29**) share 38% identity and 65% similarity (Emboss Needle alignment surfer). This alignment was used to exactly define the linker regions and the start and end of the domains. To further define the start and end point of a domain a multiple alignment was performed with DEBS3, DEBS5 (PDB: 2HG4) and huFAS (UniProtKB/Swiss-Prot-entry: P49327) (**App. 5**).

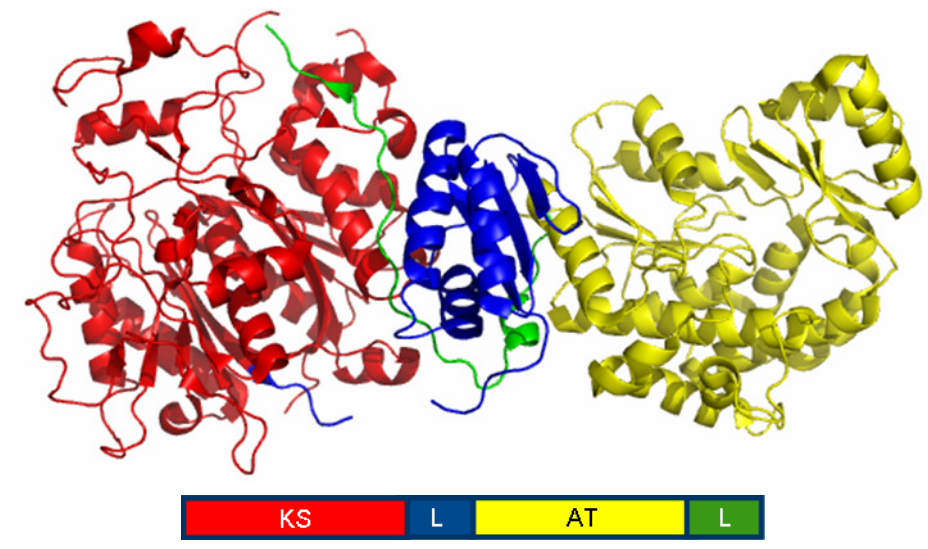


Fig. 29: Shematic presentation of one monomer of the DEBS3 KS-AT didomain (PDB: 2QO3) (adapted from [21])

To improve solubility, regions of low hydrophobicity were chosen for the N- and C-termini of the constructs. The hydrophobicty of the sequence was determined using protscale (http://www.expasy.ch/cgi-bin/protscale.pl). Interestingly, linker regions in the DEBS3 structure have shown to play an important structural role in activity and solubility [21]. Especially the post-

AT linker is required for KS catalyzed β-ketoacyl-ACP synthase activity and ACP recognition. Tang and colleagues obtained soluble constructs by keeping the KS-to-AT linker. The linkers mainly have hydrophobic interactions with catalytic domains and with each other. Therefore we decided to clone constructs containing the linkers despite the fact that they have quite high hydrophobicity and are less structured [21]. Taken together, all these aspects result in different core expression constructs summarized in **Table 10.** These core constructs were cloned in the pet28a (+) (Novagen) without any solubility tag but a C-terminal His_6-fusion, into a modified pMAL-c2x with MBP as an N-terminal solubility tag and in some cases into petM-60 (EMBL) with Ub as an N-terminal solubility tag. Some proteins did not express very well neither in pet28a (+) nor with MBP. Proteins were purified and analysed via analytical gelfiltration (on a Superose 6 column). Even in case of MBP fusion, most constructs showed mainly soluble aggregation. Gelfiltrations were performed with and without reducing agents (DTT up to 10 mM) and in different buffers containing compounds to prevent the protein from aggregation (e.g.: arginine, glutamate, sucrose). Nevertheless no constructs were useful to go on with structural studies due to aggregation appearing already at low concentration (50 µM).

Table 10: Table summerizing the different fusion constructs for AT and KS from CurA. PMAL stands for an N-terminal MBP fusion based on a modified pMAl-c2x vector, pet28a(+) for no solubility tag but a C-terminal His_6-fusion and petM-60 (EMBL) for an Ub His_6-fusion at the N-terminus. [1]indicates that the gelfiltratin profiles are attached (**App. 6**). The amino acids are indicated based on the pubmed entry of CurA (Accession: AAT70096)

Construct	Amino Acid	pMal (MBP-Fusion)	Pet28a (+)	Pet60 (Ub Fusion)
AT-L	1086- 1484	Gelfiltration: soluble aggregates	Gelfiltration: different multimers[1]	Expressed
L-AT-L	974- 1484	Weak expression	No expression	
L-AT	974- 1408	Gelfiltration: just MBP [1]	No expression	
AT	1086- 1408	Gelfiltration: just MBP	Expressed	
KS-L-AT-L	540-1484	Gelfiltration: soluble aggregates	Expressed	
KS	540- 980	Gelfiltration: soluble aggregates[1]	-	

7.1.3 A peptidyl carrier protein (PCP)

The TycB1 PCP is a peptidyl carrier protein (PCP) from the tyrocidin synthase II (Tyc) from *Brevibacillus brevis*. PCPs and ACPs share many common features– both shuttle the substrate via a thiolinkage- and share the general carrier fold, however differ in their overall charge: ACPs are acidic whereas PCPs are neutral [32, 46]. We were interested in the question whether this overall charge might interfere in the interaction once we replace CurA ACP_I with this PCP. In case it disturbs the interaction the PCP would present an ideal negative control. For expression purposes the TycB1 PCP was cloned into a pBH4 vector leading to a construct with a N-terminal His_6-tag followed by a TEV cleavage site and corresponding to the native aminoacids 968-1049 of TycB1 (Gene bank accession code: YP_002772269) (**App. 7**). The PCP was

expressed and Ni-purified according to the protocol for the ACPs with a yield of 20 mg purified protein per 1L LB (**Fig. 30a**). The quality of the protein was reviewed via Tricine gels. A double band was present for TycB1. The slightly bigger protein eluted earlier (Fraction 18-21), then the smaller one (33-36). To understand the reason for two different PCP populations both were subjected to mass spectrometry analysis. The protein sample from the fraction 18-21 for a ^{15}N-labeled sample gave a molecular mass of 11.777 kDa determined via matrix-assisted laser desorption/ionization (MALDI) measurements. The molecular weight for the full length protein is 11.776 kDa containing 141 nitrogens. In case of a ^{15}N-labeled sample this would give for the full length protein and 98% labeling 11.915 kDa. It is known that in *E. coli* methionine aminopeptidase might remove the initiator methionine depending on the nature of the adjacent residue– the smaller the 2nd aminoacid the better the substrate [173]. Taking this into account 11.777 kDa represents presumably N-terminal processed TycB1 with 99% incorporated ^{15}N (11.645kDa without N-terminal methionine and with 140 Nitrogens) (**App. 8 a)**.

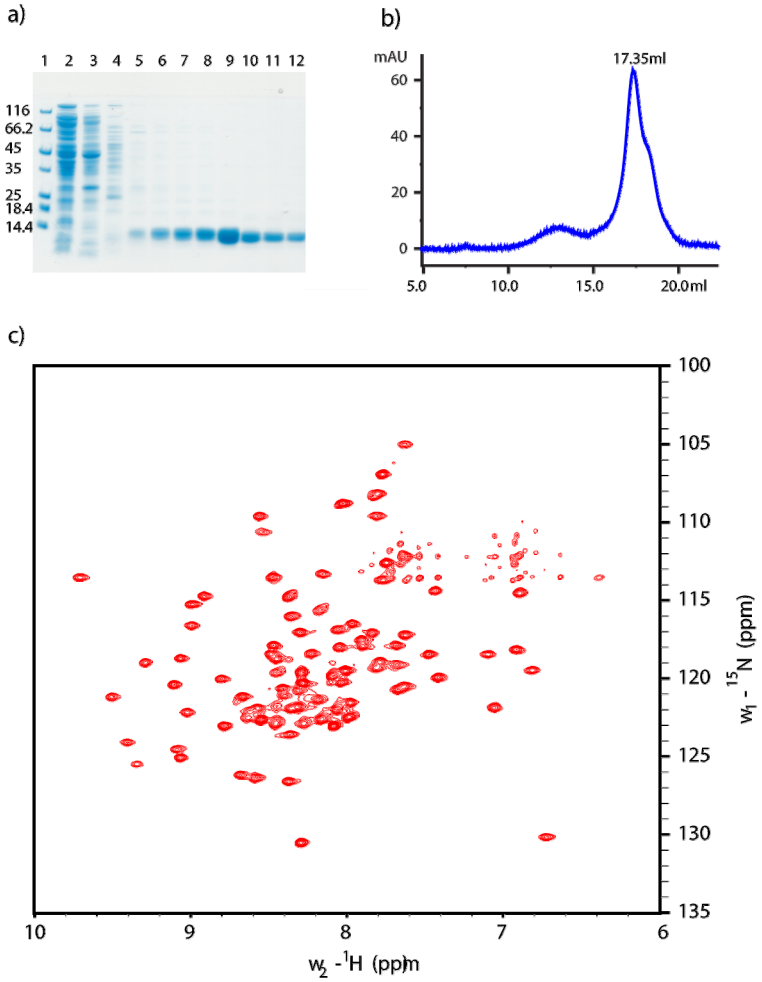


Fig. 30: Quality assessment of TycB1 PCP a) 11% Tricine gel of the Ni-NTA purified TycB1 with Line 1: marker; 2: before induction; 3: ppt; 4: snt; 5: frct. 11; 6: frct. 18; 7: fract.19; 8: frct. 20; 9: frct. 21; 10: frct. 26; 11: frct. 35; 12: frct. 38. b) Superose 6 run of TycB1 c) [^{1}H,^{15}N] TROSY of ^{15}N-labeled TycB1 PCP at 500 MHz in 20 mM Tris, 100 mM NaCl at pH 7.5.

A molecular mass of 10.782 kDa was obtained for the later fractions of the ^{15}N-labeled sample. This is a difference of 995 Da (**App. 8 b)**). Despite the difference in the molecular weight, the different fractions give nearly the same spectrum. Two peaks are not present which might be presumably glycines, an assumption based on their particular chemical shifts (**App. 8 b)**). The fact that the shortened protein and the longer protein, both stick to the Ni-column and the fact that glycines might be missing which are present at the C-terminus; all this taken together might indicate a C-terminal degradation or incomplete translation for a portion of TycB1. Nevertheless due to the small differences in the spectra and the irrelevance of the last C-terminal amino acids for function we combined the fractions for NMR titration experiments (**7.4.2**). The maximal concentration in our NMR buffer was 200 µM higher concentration gave raise to soluble aggregates. For quality assessment a gelfiltration on the semi analytical superose 6 was performed at 18°C. The fraction 18-21 of TycB1 PCP elute at 17.35 ml which corresponds to a molecular weight of 41 kDa (**Fig. 30b)**). The PCP is known not to dimerize, therefore the early elution is presumably due to a non globular behaviour of the protein. ^{15}N-TROSY experiments measured at 500 MHz of the combined fractions display a spectrum with 76 dispersed peaks with homogenous intensities. 94 amino acids are present in the TycB1 construct, of which the first 20 amino acids belong to the His_6-tag and the TEV cleavage site. These first 20 aminoacids are presumably unstructured and flexible and therefore do not appear in the [^{1}H,^{15}N]-TROSY spectrum.

7.1.4 The dehydratase (CurE- ECH_1)

ECH_1 (CurE) is a dehydratase domain from the curacin cluster described in **4.2.3**. The construct of ECH_1 was received as a gift from the Sherman lab. The construct in pet24b is composed of the native sequence spanning from amino acid 1 to amino acid 253 (GenBank accession code: AAT70100.1), C-terminally tagged with a His_6-tag. The protein was successfully expressed in BL21 at 15°C. High amounts of protein aggregates and stays in the pellet after sonification. The soluble part was successfully purified via Ni-NTA and was concentrated to 10 µM in the activity test buffer (30 mM Tris, 50 mM NaCl, pH=6.5) and CD buffer (20 mM Na_2HPO_4, 50 mM NaCl at pH 6.5) (**Fig. 31**). The protein precipitated at higher concentrations. This is also the case when ECH_1 was kept at 10 µM at 4°C for several days. Storage at -20°C turned out to be impossible, even at low concentrations. For quality assessment circular dichroism (CD) spectroscopy (**Fig. 31a)**), CD-melting curves (**Fig. 31b)**) and SDS-PAGE analysis (**Fig. 31c)**) were performed. The CD curve displays a folded protein with no unfolded content. The melting curve shows that ECH_1 has a melting temperature of $38.66 \pm 0{,}12$°C.

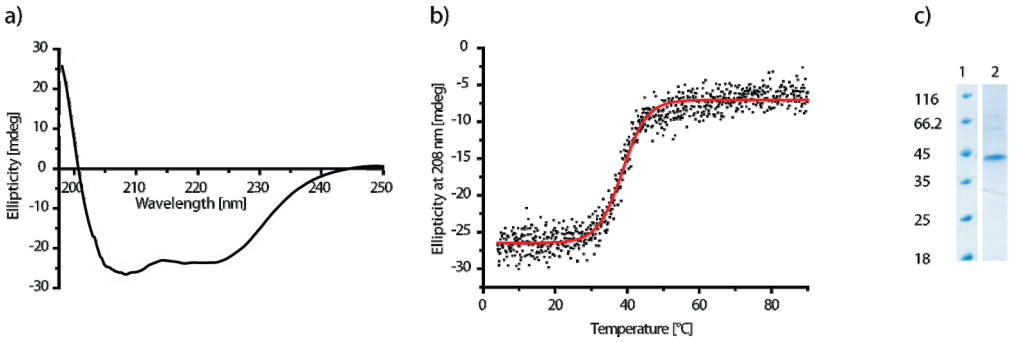


Fig. 31: Quality assessment of ECH_1. a) CD spectroscopy of ECH_1 b) CD melting curve of ECH_1 c) SDS-PAGE analysis of ECH_1

To overcome the low solubility of the protein and to make it available for structural studies we cloned it into a modified petM-60 (EMBL) to fuse it at the N-terminus with ubiquitin and cloned it into a modified pMal-c2x to obtain an N-terminal MBP fusion. Expressions were performed in BL21 and Arctic Express cells at 15°C. High amounts of the protein were mainly expressed in the insoluble fractions. Solubilty was not increased by the fusions. Therefore, we decided to use the pet24b construct described above and perform activity assays with ECH_1 (**7.4.1**).

7.1.5 The decarboxylase (CurF- ECH_2)

ECH_2 is located at the N-terminus of CurF and has decarboxylative activity [68] (as described in **4.2.3**). We received the CurF didomain ECH_2 with ER in *cis* as a gift from David Sherman. ECH_2 was then cloned into pet28a (+) containing a C-terminal His_6-tag. The protein is expressed with an N-terminal truncation and ranges from amino acid 17-257 of the CurF protein (Gene bank accession code: AAT70101.1). Geders and colleagues used this construct for crystalisation [70]. We expressed the protein in the BL21 *E. coli* strain at 15°C and purified it by Ni-NTA affinity chromatography as described in the method section. The purified protein was concentrated and the quality analysed using CD-spectroscopy (**Fig. 32 a) b)**) SDS-PAGE with comassie staining (**Fig. 32c**), and gelchromatography (**Fig. 32d)**). CD spectroscopy showed the spectra of a well folded protein with a dominant α-helical content. From the CD melting curve of ECH_2 the derived melting temperature was 45 ± 0.1°C. The protein was purified by Ni-NTA purification to a high purity with minor impurities (< 5%). Per 1L LB we obtained 20 mg purified ECH_2. In some cases we performed additional gelfiltration on a superdex 200 but no soluble aggregates were observed at any point. ECH_2 has a calculated molecular weight of 27 kDa, however, elutes on a superose 6 column at 17.1 ml corresponding to a molecular weight of 55 kDa indicating a dimeric state.

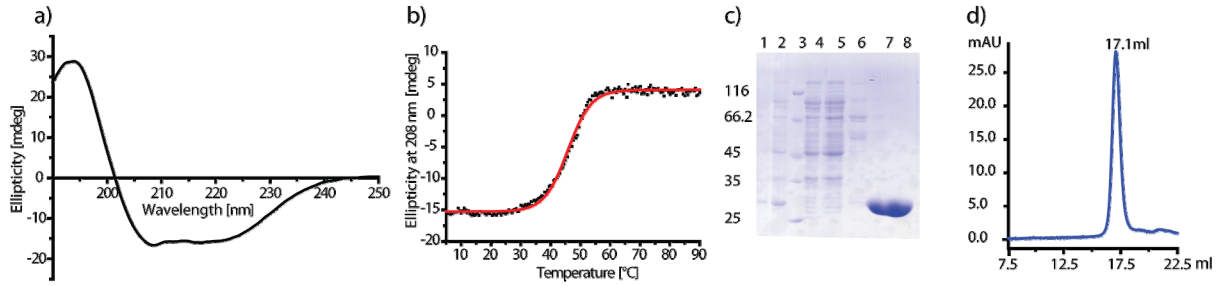


Fig. 32: Quality controls of the ECH_2-contruct (17-257) a) Circular dichroism (CD) spectroscopy from 190 nm to 250 nm of ECH_2 b) Denaturation curve using CD spectroscopy at 208 nm c) SDS-PAGE of the ECH_2 Ni-NTA purification: 1) ppt after lysis 2) snt after lysis 3) marker 4) flow through 5) 10 mM imidazol wash 6) 60 mM imidazol wash 7) 400 mM elution 8) after gelfiltration run d) gelfiltration run of ECH_2 on a superose 6 10_300 column. The protein elutes at 17.1 ml which corresponds to 55 kDa.

Different buffers were screened and in high salt buffer (50 mM PP, 200 mM NaCl, 1 mM DTT, pH 7.4) it is possible to concentrate ECH_2 to 900 µM. In this buffer, ECH_2 can be frozen when the protein is not concentrated higher than 50 µM and supplemented with 10% glycerol. For our experiments we prepared fresh ECH_2. In low salt buffers, which are necessary for NMR and CD, ECH_2 was concentrated to 500 µM and 560 µM (in 20 mM Tris or NaPP, 100 mM NaCl, 1 mM DTT, pH 7.5) but started then to precipitate over time (2 hrs at 4°C; 10 min RT). Therefore we tried Ub (petM-60) and MBP (modified pMAL-c2x) fusions to increase solubility. Additionally we elongated the C-terminus and removed the N-terminal truncation. All these constructs did not improve the stability. Therefore, we worked in this thesis with the pet24b ECH_2-construct (17-257) characterized above and used before in other studies [65, 68-70, 76]. Gelfiltrations were performed in different buffers with and without DTT (10 mM) all profiles show no soluble aggregation and a peak at 17.1 ml corresponding to a molecular weight of 55 kDa corresponding to a dimeric state. Geders and his colleagues obtained in their crystalls trimeric ECH_2, but they suppose a dimeric state for the protein in solution [70]. In our research this dimeric state was not influenced neither by high concentrations of ECH_2 (400 µM) nor by DTT (10 mM). One initial objective of this work was to map the interaction site between the ACP_I and the interaction partner ECH_2. For fast mapping approaches labelled amino acids that are at the end of biosynthetic pathways (Lys, Arg, His) can be added directly to the minimal medium without scrambling, when the protein of desire is expressed in *E. coli* BL21 cells. For other amino acids (such as Ala, Asp, Tyr, Phe, Leu, Ile and Val) the auxotrophic strains (DL39) that suppresses any cross-labeling can be used [100]. Another way to label specific amino acids is possible by using the cell free system. This approach allows more flexibility [174]. Therefore we tried cell free expression of ECH_2. An initial Mg^{2+} screen was performed at 25°C and 30°C ranging from 15-21 mM Mg to find the optimal Mg^{2+} concentration for ECH_2 expression (18 mM Mg^{2+} was the optimum for GFP expression in the cell free extract used). A very faint band is visible in the soluble part of 15 mM Mg^{2+} for 25°C and a little bit more at 30°C (Fig. 33 **a) and b)**).

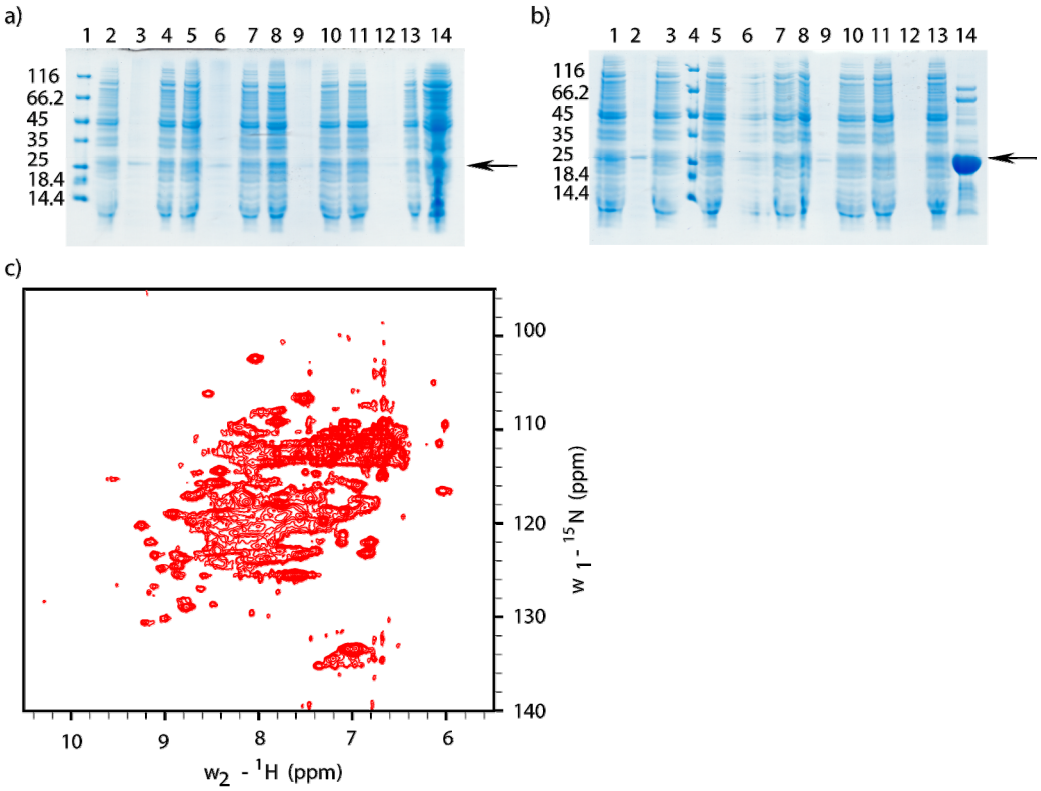


Fig. 33: ECH_2 cell free expression and [^{1}H,^{15}N]-TROSY measurements a) Magnesium screen at 25°C with Lane 1: marker; Lane 2: 15 mM reaction mix (RM); 3: 15 mM pellet (ppt); 4: 15 mM supernatant (snt); lane 5: 17 mM RM; 6: 17 mM ppt; 7: 17mM snt; lane 8: 19 mM RM; 9: 19 mM ppt; 10: 19 mM snt; Lane 11: 21 mM RM; 12: 21 mM ppt; 13: 21 mM snt; Lane 14: RM without an overexpression. b) Magnesium screen at 30°C with Lane 1: 15 mM RM; 2: 15 mM ppt; 3: 15 mM snt; 4: marker; 5: 17 mM RM; 6: 17 mM ppt; 7: 17 mM snt; 8: 19 mM RM; 9: 19 mM ppt; 10: 19 mM snt; Lane 11: 21 mM RM; 12: 21 mM ppt; 13: 21 mM snt; Lane 14: from *E. coli* purified ECH_2. c) [^{1}H,^{15}N]- TROSY of ^{15}N-labeled ECH_2 at 291K at 900 MHz in 20 mM Tris, 100 mM NaCl, 1 mM DTT at pH = 7.5**.**

The amounts of soluble protein obtained by cell free expression were too small and far away from 1-2 mg/ml, turning this method in case of ECH_2 to a non-reasonable approach. In a next step we expressed ECH_2 ^{15}N-labeled and recorded a [^{1}H,^{15}N]-TROSY to see how the protein behaves under NMR conditions (20 mM Tris, 100 mM NaCl, pH=7.5) (**Fig. 33 c)**). The peaks are not homogenously distributed and are of low intensity. The apparent molecular weight of 55 kDa leads to fast relaxion and a fast decline of signal intensity (**Fig. 33 c)**). In conclusion the fast mapping approach for ECH_2 is not feasable.

7.1.6 The halogenase (Cur Hal)

The halogenase Cur Hal is a dimeric protein which was crystalized in 2010 [67]. We received the construct as a gift from D. Sherman. Cur Hal was cloned into the vector pET28b with an N-terminal His_6-tag. The native protein seuence covers the amino acids 1599 to 1930 of CurA according to the sequence accessible at NCBI (GenBank accession code: AAT70096.1). The CD spectrum shows a folded protein with α-helical contents, which dominate the CD spectra (**Fig. 34 a)**). The protein has a melting temperature (T_M) of 42.7 ± 0.027°C (**Fig. 34 b)**). Cur Hal shows high affinity to the Ni-IDA column and was purified in a convenient three step protocol (1st wash 50 mM imidazol, 2nd wash 80 mM imidazol, elution with 400 mM imidazol). The protein

purity was monitored using SDS-PAGE and coomassie staining (**Fig. 34 c)**). For high purity of Cur Hal no further purification step was required. However gelfiltration was performed to assure a homogenous sample but skipped later on due to 50% protein loss. Therefore gelfiltration was only applied for quality assessment (**Fig. 34 d)**). Hal elutes on a superose 6 column at 16.6 ml. This corresponds to a molecular weight of 76 kDa, hal has a calculated molecular weight of 40.5 kDa therefore the apparent molecular weight obtained by gelfiltration corresponds to a dimeric state. This is in accordance with results obtained by Khare and colleagues [67]. Addition of 10 mM DTT did not affect the dimeric state.

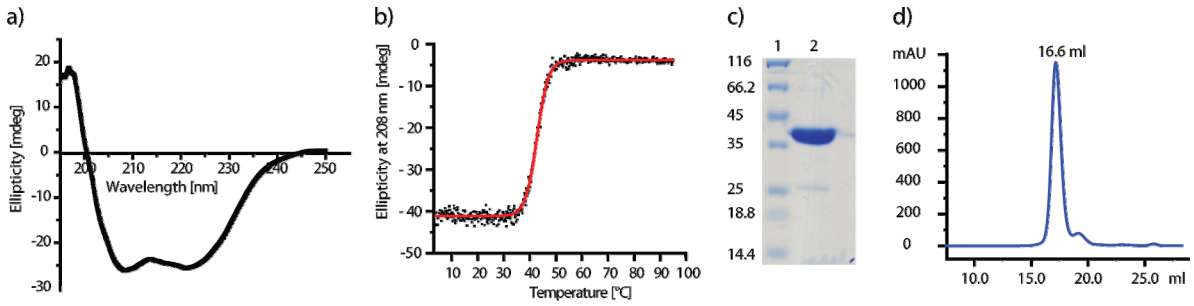


Fig. 34: Quality assessment of Cur Hal. a) CD curve b) CD melting curve c) SDS-PAGE d) Gelfiltration profile

The iron content of the non-heme iron (II) and α-KG dependent halogenase was determined by using the Ferene S [3-(2-pyridyl)-5,6-bis(2-(5-furyl-sulfonicacid))-1,2,4-triazine] dye (**6.2.8**) [155]. The Ferene S dye chelates iron and the complex has an absorption maximum at 593 nm which allows determination of the amount of the complex present in solution. A standard curve for calibration of the Ferene S-iron complex was obtained by using stock iron solutions and measuring the absorption at 593 nm versus a blank run (**App. 1**). For Hal the A593nm value was 0.199 ± 0,007. From this value the iron content was calculated using the equation of the linear regression of the calibration curve. This led to a calculated iron content of 5.81 nmol. As 13.83 nmol of the enzyme was used, this value corresponds to 42% iron content. The crystal structure reveals that each active monomer within the dimer contains one iron center for catalytic activity. The fact that 42% of the protein contains iron means that a maximum of 42% might be active not considering that parts of the iron might be oxidized and therefore inactive. Nevertheless keeping in mind that iron can be presumably replaced by other divalent cations during purification the presence of iron in 42% of the protein represents a quite good protein quality. We assessed the iron content as well from Cur Hal expressed in media supplemented with iron and the iron content was 40%. Therefore supplementing the media with extra iron is not accompanied by higher incorporation.

Cur Hal is a well behaving protein; it can be concentrated to 900 µM without aggregation and survives freezing and thawing as long as it is kept anaerob. Once it was in contact with oxygen it started to precipitate and lost activity. We obtained 3 mg of purified, not reconstituted Cur Hal from 1L LB. To improve expression yields we tried N-terminal MBP (modified pMAL-c2x) and

Ub-fusions (petM-60). In fact the expression was improved but the requirement of tag cleavage and further purification reduced the yield and turned the fusions inefficient (**Table 11**). Looman and colleagues showed that the 2nd codon following the AUG initiation codon has a strong effect on the expression yield of a modified lacZ gene in *E. coli* expression. In their study the codon AAA as a 2nd codon gave by far the best yield for the expression in *E. coli* [175]. We applied the same idea to our protein. Actually, by changing the 2nd codon from GGC to AAA at least doubled our protein yields (**Table 11**). In all experiments displayed in this thesis we finally worked then with the resulting mutant HalG2K.

Table 11: Comparison of different phy constructs in term of necessary purification and yield per 1L LB.

Construct	1st	2nd	3rd
Hal	Ni-NTA: 6mg/1L LB	Gefi: 3mg	/
Ub-Hal	Ni-NTA: 10 mg/1L LB	Tev and Rev. Ni-NTA: 2mg/1L LB	Gefi : 1mg/1L LB
MBP-Hal	Amylose: 8 mg/1L LB	Rev. Ni-Nta: 2 mg/1L LB	/
Hal G2K	Ni-NTA: 12 mg/1L LB	Gefi: 6 mg	/

The halogenase turned out to be an interesting target for the fast mapping approach and we were interested in evaluating the feasibility of this approach for Hal. In the Cell- free expression we obtained the best expression for Hal at 16 mM Mg^{2+} at 30°C for the extract used. Nevertheless, at this optimum 50% of the soluble expressed protein was precipitated and present in the insoluble fraction, the pellet (**Fig. 35a)**). Additionally we expressed Hal in *E. coli* DL39 cells. DL39 cells are auxotroph for the amino acids Ala, Asp, Tyr, Phe, Leu, Ile and Val. They have to be added to the minimal media and can be labeled without cross-labeling [100]. Unfortunately proteins that have a tendancy to aggregate during expression are even more difficult to express in DL39 cells. Nevertheless, we expressed Hal in DL39 cells and labeled Lys ^{15}N. The protein was purified and reconstituted and we recorded a [^{1}H,^{15}N]-TROSY spectra (**Fig. 35 b)**). A Cur Hal monomer contains 31 Lysines. In case of a symmetric dimer we would expect 31 signals in case of dissymetrie 62 in a [^{1}H,^{15}N]-TROSY. The recorded spectrum suffers from high peak overlap and inhomogenous intensities even when recorded at 900 MHz. Due to the dimeric state in solution the protein relaxes very fast allowing only 60 increments to be recorded and therefore restricts the resolution. To summerize Cur Hal turned out to be a well behaving protein for NMR and therefore not a feasible target for fast mapping approaches.

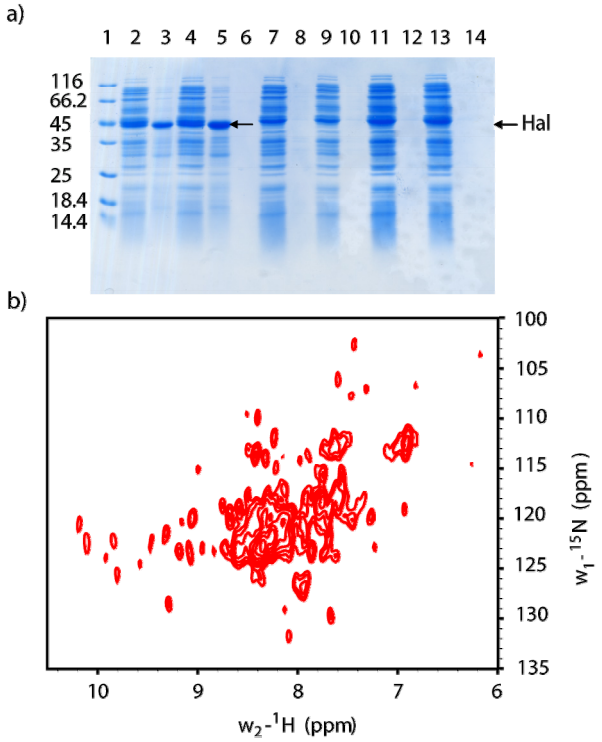


Fig. 35: Cell free expression of Cur Hal and [^{1}H,^{15}N]- TROSY a) Magnesium screen of cell free expression of Cur Hal. Magnesium screen at 30°C with lane 1: marker; lane 2: 15 mM supernatant (snt); lane 3: 15 mM pellet (ppt); lane 4: 16 mM snt; lane 5: 16 mM ppt; lane 6: empty; lane 7: 17 mM ppt; lane 8: 17 mM snt; 9: 18 mM ppt; 10: 18 mM snt; Lane 11: 19 mM ppt; lane 12: 19 mM snt; lane 13: 20 mM ppt; Lane 14: 20 mM snt. b) [^{1}H,^{15}N]- TROSY of ^{15}N-Lysine labeled Hal. Sample was 0.1 mM Hal in 20 mM Hepes, 71 mM Arg, 50 mM Glu, 0.5 mM α-KG at a pH of 7.5 and 291 K at 900 MHz. The sample was in anaerob conditions. The spectrum was recorded with TD1 60 and 512 numbers of scans. Hal contains 31 lysines; due to dimerization 31- 62 signals are expected. The peak intensities are not homogenous and clustered in the middle area.

In NMR titration experiments presented in detail later on, no interaction was observed between Hal and ACP_I in different chemical states (**7.4.3**). Marius Clore and his group showed that transient interaction can be detected by paramagnetic NMR [176, 177]. For these experiments paramagnetic probes must be attached to cysteines located near to the active site (maximal distance of 15 Å). To try this approach for interaction studies via NMR, we created several cysteine mutants with a cysteine in a range of 15 Å around the active site (Cys19Ala, Cys19Ser, Ile61Cys, His109Cys, Gly111Cys, Tyr274Cys, His168Cys and Ser178Cys). Unfortunately all mutants turned out to be insoluble.

Linker regions can enforce the interaction of two domains, especially in case of low K_Ds. We expressed the constructs Hal-$ACP_{I,II,III}$-C_D and Hal-ACP_I. Both constructs turned out to be insoluble. Due to the fact that the single constructs are soluble we tried to express the single domains fused to inteins for carrying out segmental labeling (compare with **4.4** and **7.2**). The idea behind was, that even if inteins are known to decrease solubility additional solubility tags might help the domain precursors to stay in solution. The Cur Hal protein which was expressed soluble containing the N-terminal sequence SISS (indicated in pink) therefore the same site was chosen as cleavage site for the intein fusions (C/N-Intein). After a fusion this linker region would keep some mutations which are presented in

Fig. 36. There the wild type sequence (WT) of the linker region and the sequence which would result from segmental labeling (SL) are compared.

```
WT              PHVQQTLAARMAAGAISFEETPSISSAPQTQQPLKTLQPL 40
SL              PHVQQTLAARMAAGAISFEETPSGSCFNGTQQPLKTLQPL 40
                ********************** *.   ***********
```

Fig. 36: Sequence of the Hal-ACP$_I$ linker site which was chosen for the segmental labeling approach. In pink is indicated the sequence of the Hal construct which was expressed soluble.

The individual domain precursors were not soluble despite the fact that they contained SMT3 as a solubility tag (Data not shown).

7.1.7 The single acyl carrier protein (CurB- ACP)

CurB is a lone standing ACP. It belongs to the HMG cassette together with HCS and KS (**4.2.1**) and is the substrate deliverer for the acetyl group (**Fig. 15**) necessary for introducing the β-branched methyl group. Therefore it is supposed to interact with the proteins KS (CurC) and HCS (CurD). We obtained the construct for CurB from D. Sherman. The full length gene of CurB (Gene bank accession code: AAT70097.1) was cloned in pet28b with an N-terminal His_6-tag giving raise to 99 amino acids whereof 20 are artificial due to the N-terminal tag [68]. We were not able to express CurB in BL21 cells as described before [68]. Therefore we first tagged the CurB construct N-terminally with Ub (petM-60) which also did not express in BL21 or Arctic Express cells. Further investigations showed that CurB is encoded with a high amount of rare codons (35 rare codons out of 99). Therefore we co-expressed the CurB construct with the vetor pRARE which codes for rare t-RNAs (NEB). As a result, both the Ub-tagged and non-tagged version of CurB was expressed and purified via Ni-NTA affinity chromatography in high amounts (**Fig. 37a)**) (Ub-CurB data not shown). For CurB we obtained 42 mg Ni-NTA purified protein per 1L LB for Ub-tagged CurB 41 mg. Due to no significant increase in expression for the Ub-tagged protein we worked with the N-terminally His_6-tagged version. At high concentrations such as 0.6 mM in 20 mM Hepes and 50 mM NaCl, pH 7.5, CurB is pretty unstable. It precipitates after 10 min. Acetyl-CurB is more stable than CurB. At pH 5 acetyl CurB is much more stable and can be kept at least for several hours. We performed buffer screens with acetyl CurB (pH 5: 20 mM NaAc, 50 mM Arg/Glu; pH 6: 20 mM MES, 50 mM Arg/Glu; pH 7: 20 mM Hepes 50 mM Arg/Glu; pH 8: 20 mM Hepes, 50 mM Arg/Glu) on a gelfiltration column (superose 6) and recorded [^{1}H,^{15}N]-TROSYs (**Fig. 37 b) and c)**).

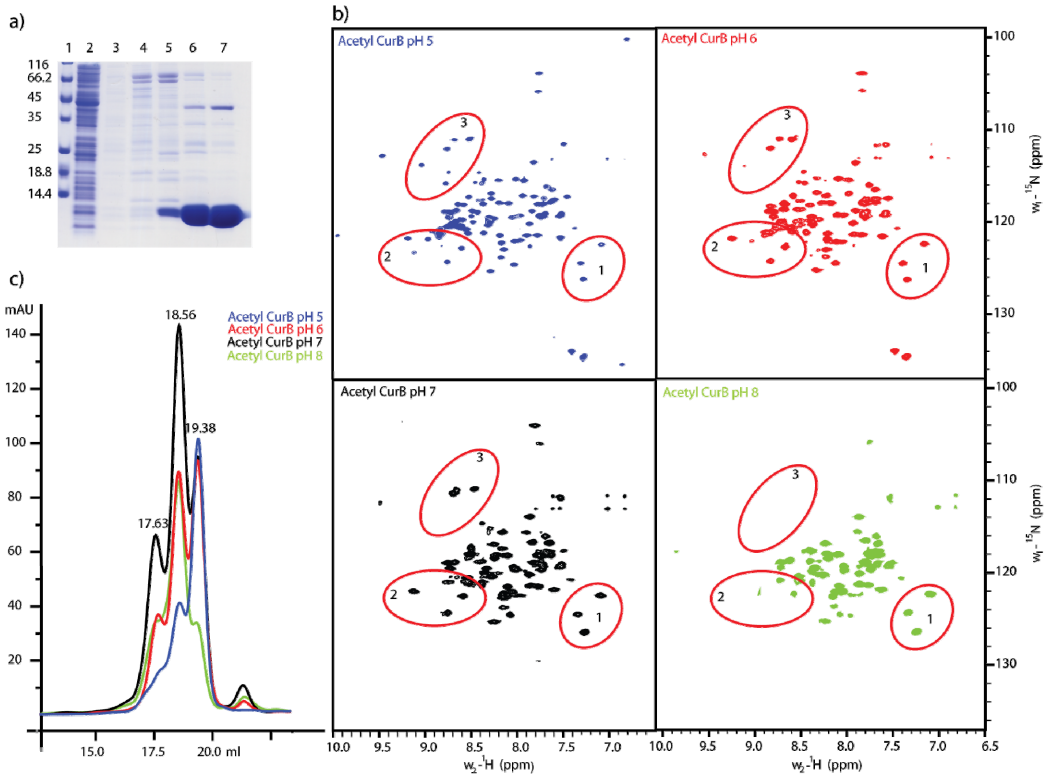


Fig. 37: PH dependant quality assessment of acetyl-CurB a) Tricine-PAGE b) [^{1}H, ^{15}N]-TROSY at different pH at 500 MHz. The different red circles indicate regions of interest, e.g. number 1 indicates peaks which are unaffected in terms of intensity by increasing the pH. In comparison region 2 and 3 present peaks which disappear with increasing pH c) Gelfiltration profile of acetyl CurB on a superpose 6 column at different pH.

At a pH of 5 we obtained the best spectra. With increasing pH signals start to disappear. It is known that at pH>7 amide proton signals have the tendancy to disappear due to the increased NH exchange of the backbone; but in case of acetyl-CurB this effect is already visible at pH 6. The gelfiltrations display that CurB exists in a pH dependant multimeric state. At pH 5 (blue) CurB is manly in its monomeric and dimeric state. With pH 6 (red) monomer, dimers and trimers are present. The higher the pH the more dimers and trimers are populated (**Fig. 37 b) and c)**). In order to confirm the mulitmeric state of acetyl-CurB, we performed size exclusion chromatography (SEC) multi angle light scattering (MALS) measurements at pH= 5. The protein eluted with one peak which we separated into a front and a back part for analysis. The front part appeared to be a monodispers, dimeric state and the back part confirmed the monomeric state. It is important to mention that 100 µM CurB was modified with 3 µM SFP. This SFP is an impurity, nevertheless the front part of the peak is not polydisper it is a homogenous dimer peak **(Fig. 38, Table 12**).

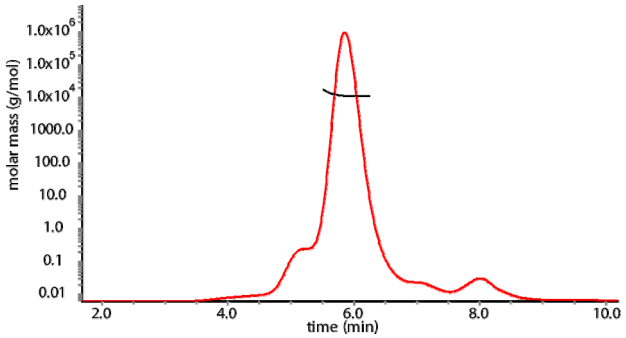


Fig. 38: SEC-MALS measurements of acetyl CurB at pH 5. For the measurement a Superdex 75 5/150 column was used with a Optilab rEX Refractive Index Detector and a Dawn Heleos II with a wavelength of 658 nm (Wyatt Technology). In the Picture are displayed the UV curves at 280 nm (red) and the determination of the molecular weight (black). On the y-axis is marked the determined molecular weight of the protein in g/mol.

Table 12: Results of the SEC-MALS measurement of acetyl CurB. For analysis the UV curve of the elution with the weight determination was divided into two parts **(Fig. 38)**. For each sample the averaged molecular weight of the eluted molecules is indicated.

Protein	Molecular weight front part [kDa]	error [kDa]	Molecular weight back part [kDa]	error [kDa]
acetyl CurB	27.46 (3%)	0,823	11.71 (6%)	0,35

7.1.8 $ACP_{I,II,III}$ and the isolated domains

The triplet ACP domains $ACP_{I,II,III}$ are located at the C-terminus of the CurA protein from *Lyngbya majuscula*. The initial structural studies of the triplet ACP (CurA-$ACP_{I,II,III}$) were done with a truncated version missing the C-terminal C_d-domain (C_d, CurA 2261–2311) [76]. During this study single domain constructs as well as double domain constructs were prepared in the pet28a (+) vector. The constructs contained either a C-terminal or an N-terminal His_6-tag. In **Table 13** the different ACP constructs used in this study are listed with the accoding yields after expression in labeled minimal media and after Ni-purification.

Table 13: ACP constructs and there yields after Ni-purification.

Construct	Amino Acid	His-tag	Yield in mg M9- Medium
$ACP_{I,II,III}$	1946- 2248	N-terminal	9,2
$ACP_{I,II,III}$-C_d	1946– 2311	N-terminal	11,5
ACP_I	1946- 2034	N-terminal	33.3
ACP_{II}	2055- 2143	N-terminal	24,48
ACP_{III}	2145- 2248	N-terminal	26,67
$ACP_{I,II}$	1946- 2151	C-terminal	8.13
$ACP_{II,III}$	2035- 2248	C-terminal	7.2

All constructs were expressed in BL21 in LB or M9 media at 15°C for 7-10 hrs, media was supplemented with 0.04 mM Fe-II-solution in order to prevent modifications by an endogenous PPTase [22]. Longer expressions in fact produced small amounts of holo-ACP proteins. All

proteins were purified by affinity chromatography using Ni-NTA. Due to the weak binding the proteins were purified via a gradient protocol and eluted around 70 mM imidazol. The purity of the proteins was monitored via SDS-PAGE analysis (for $ACP_{I,II,III}$) or Tricine gels (for single domains). The proteins were expressed as apo proteins and then modified according to the protocol in the material and method section (**6.2.7**). The modification was monitored using MALDI-MS analysis to assure 100% conversion and exclude impurities of holo Protein (e.g.: presented on apo to HMG conversion of ACP_I (**Fig. 39**).

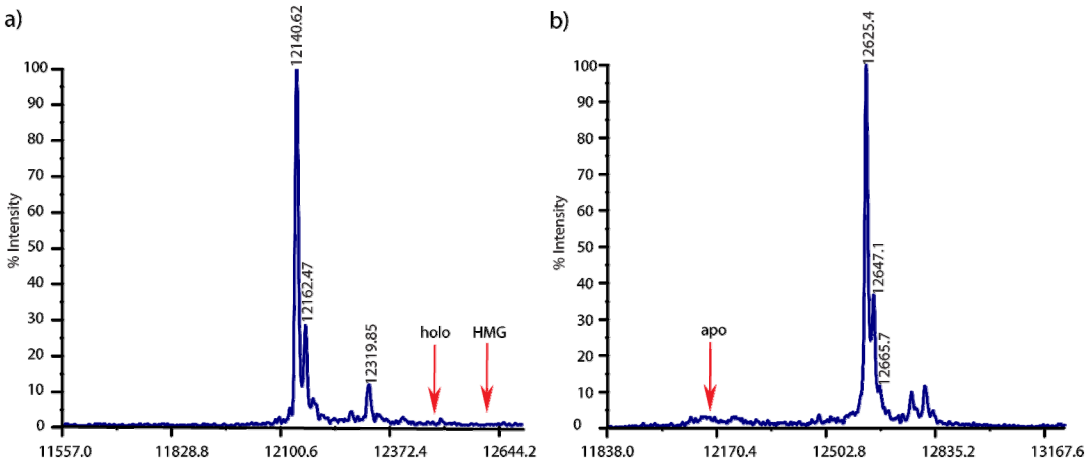


Fig. 39: MALDI-MS analysis of apo ACP_I and HMG-ACP_I after conversion. a) MALDI-MS spectra of apo ACP_I. No peak is observable corresponding to the mass of holo ACP_I or HMG-ACP_I b) MALDI-MS analysis of HMG-ACP_I. No mass is detected corresponding with the MW of apo ACP_I. The conversion took place ~100%.

For apo $ACP_{I,II,III}$ and apo $ACP_{I,II,III}$-Cd ^{15}N-labeled NMR samples were prepared in 50 mM arginine and 50 mM glutamate buffer with a pH of 6.8 and a [^1H,^{15}N]-TROSYs were recorded at 950 MHz and the same samples were run on a superose 6 column (**Fig. 40**). Apo $ACP_{I,II,III}$ which has a calculated molecular weight of 36 kDa elutes at 16.5 ml, which corresponds to a molecular weight of 81 kDa-2.25 bigger than expected. Apo $ACP_{I,II,III}$-C_d has a molecular weight of 44 kDa and elutes as 16.13 ml which corresponds to a molecular weight of 104 kDa- hence the protein runs 2.36 times bigger than expected. Despite the fact that $ACP_{I,II,III}$-C_d is 60 AA longer than $ACP_{I,II,III}$ no big differences are visible in the overlay of the [^1H,^{15}N]-TROSY spectrum for these proteins. In the region ~8.5 ppm many peaks are present with high overlay (**Fig. 40**). This is the region mainly occupied by less structured regions of the protein. Due to the high peak overlay it can not be excluded that some new peaks appear in this region.

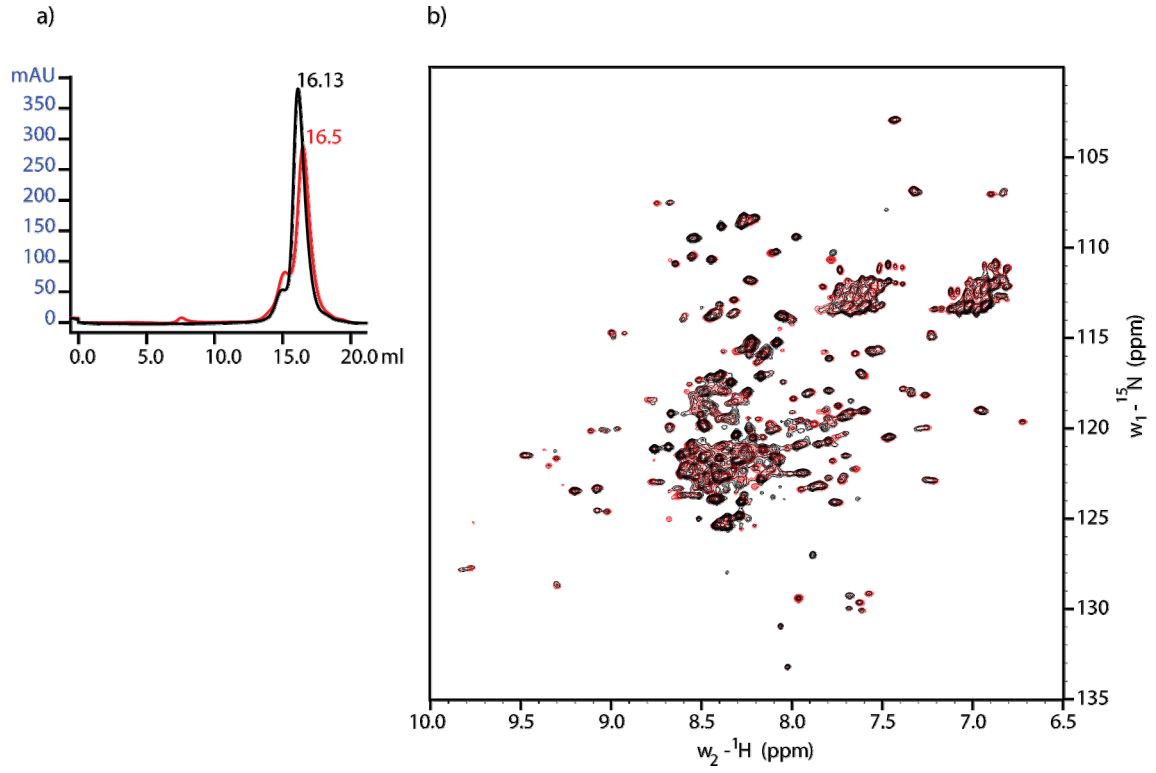


Fig. 40: Comparison of $ACP_{I,II,III}$ and $ACP_{I,II,III}$-C_d. a) Gelfiltration chromatography of apo $ACP_{I,II,III}$ (red) and $ACP_{I,II,III}$–C_d (black). b) Overlay of [^{1}H,^{15}N]-TROSY of apo $ACP_{I,II,III}$ (red) and $ACP_{I,II,III}$-C_d (black). With the C-terminal elongation some peaks become less intensive. No new peaks occur in the structured region.

In the following is the quality assessment of one isolated ACP_I presented. The CD spectra of ACP_I shows a high α-helical content (**Fig. 41 a)**). The melting temperature (T_M) of the protein is 62.8°C (**Fig. 41b)**). The quality of the purification was monitored usind Tricine gels from Schaegger [151] and does not show any impurities (**Fig. 41c)**). ACP_I elutes on a superpose 6 column at 18.51 ml corresponding to a molecular weight of 21 kDa-1.75 times more than the expected MW of 12 kDa and presumably a dimeric state. 10 mM DTT as reducing agent did not effect the gelfiltration profile.

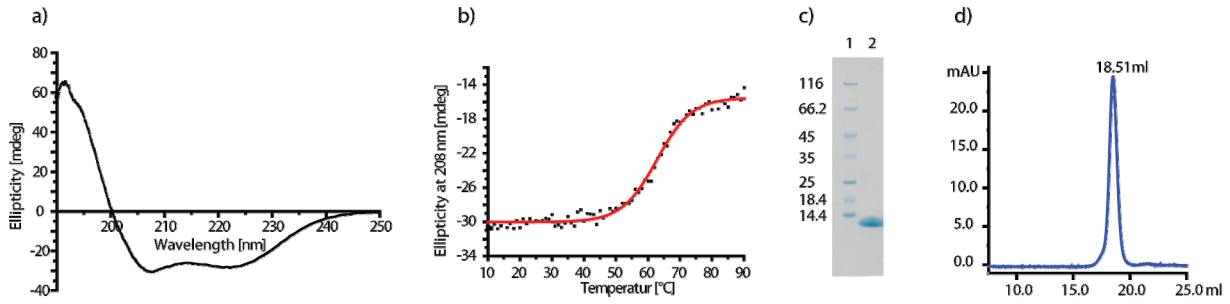


Fig. 41: : Quality assessment of ACP_I a) CD curve b) CD melting curve c) Tricine gel d) Gelfiltration profile on superose 6 10_300 column

Beside Crawford and Wong nobody reported any dimerization of ACP domains without a dimerization domain [36, 178]. The rotational correlation time (τ c) also called tumbling rate is a fundamental parameter to draw conclusions about the multimeric state in addition it helps to get insights into the relaxation behaviour of the protein of interest. The TROSY for rotational

correlation times (TRACT) is one fast method to determine τ c value [171]. We performed the two-dimensional version of the TRACT experiment for a more precise determination of τ c. TROSY and anti-TROSY components were recorded in an interleaved manner, and the mixing time Δ was incremented from 3ms to 141ms in 6ms steps. For data analysis, or each amino acid peak heights in each of these spectra were plotted versus the delay time and fitted exponentially. From this first order exponential fit, τ c was derived for each single amino acid. The derived τ c values were then plotted against the sequence (**Fig. 42**). The ACP_I is 110 aa long of which the first 21 are unstructured and should be removed from the tumbling rate calculation (**Fig. 42**). For a dimeric state we can calculate and expect a tumbling rate of 15.9 ns [171, 179].

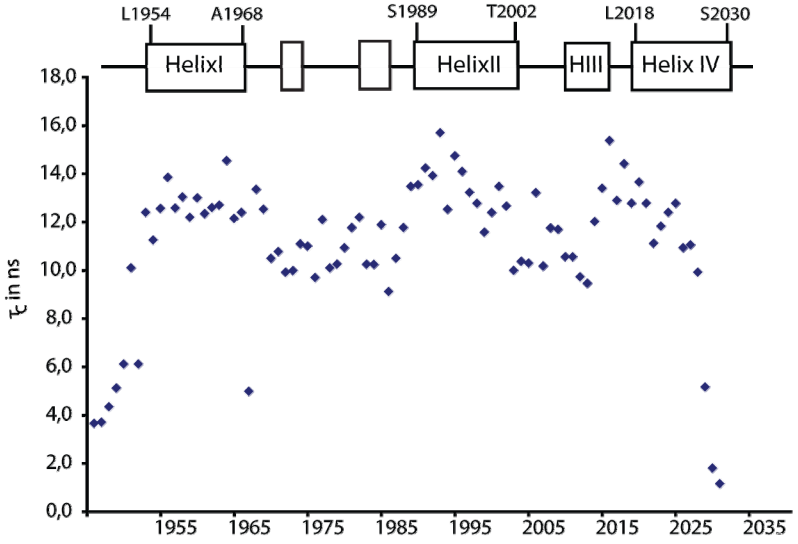


Fig. 42: TRACT measurements of apo ACP_I. The rotational correlation time τ c of ACP_I was determined by TRACT. Experiments were conducted as 2D versions at 291 K using fully ^{15}N-labeled ACP_I. Peak heights were analysed by Sparky yielding individual τ c values for each amino acid. The obtained values vary with the secondary structure. Unstructured linker regions at the C- and N-terminus have lower τ c values than residues in α-helical regions.

The experimental results for τ c are in an average 11.1 ns which is between a monomer and a dimer. In order to obtain a non ambigous result about the multimeric state of ACP_I we performed light scattering experiments. Light scattering experiments allow determining the averaged molecular weight of the molecules present in the detector. By coupling a gelfiltration column to the light scattering set up, the proteins can be separated by their size and then be directly analyzed by light scattering. In this way a molecular weight can be assigned to each peak. The technic allows as well to determine how disperse and homogenous a sample is. The combination of size exclusion chromatography and multiple angle light scattering is called SEC-MALS. Apo ACP_I and the different loading states (holo, acetyl, HMG) were applied to a Superdex 75 5/150 column and then measured via Optilab rEX Refractive Index Detector and a Dawn Heleos II Laser with a wavelength of 658 nm (Wyatt Technology). The obtained curves are displayed in **Fig. 43**. The peak elutes at ~6 min.

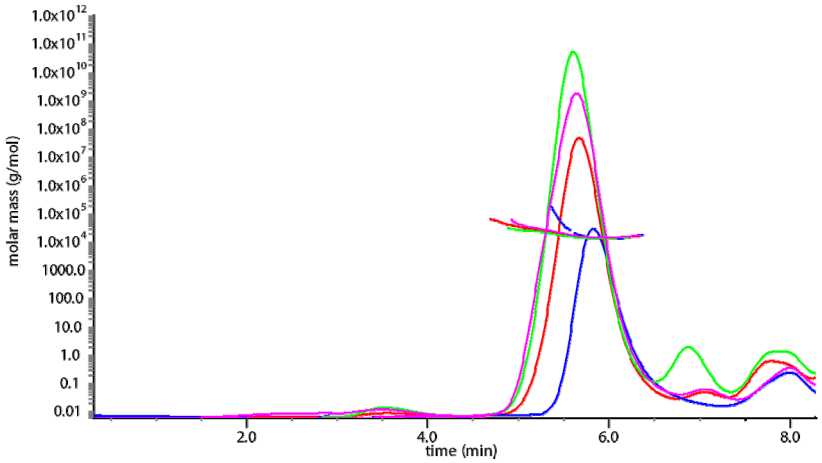


Fig. 43: SEC-MALS measurements of ACP$_I$ in different chemical states. For the measurement a Superdex 75 5/150 column was used with a Optilab rEX Refractive Index Detector and a Dawn Heleos II with a wavelength of 658 nm (Wyatt Technology). In the Picture are displayed the UV curves of apo ACP$_I$ (blue), holo ACP$_I$ (pink), acetyl ACP$_I$ (red) and HMG ACP$_I$ (green) at 280 nm and their determination of the molecular weight. On the y-axis is marked the determined molecular weight of the protein in g/mol.

The SEC-MALS measurements of the different loading states such as holo ACP$_I$, acetyl-ACP$_I$ and HMG-ACP$_I$ reveal that there is a mixture of at least two populations. Due to this fact the peak was distinguished into a front and a back part for separated analysis. For the unmodified apo ACP$_I$ no separation was necessary because only one population was present. In all cases a second peak was observed after 7 min, but the RI excluded this to be a protein. No further analyses were performed to observe what was responsible for the 7 min peak. The results for the analysis of the SEC-MALS measurements are summarized in **Table 14**. For the modified proteins we have a polydisperse solution which elutes with an average mass of 21 kDa. This might be a mixture of aggregation and a contamination of SFP due to the modification. In the back part proteins elute with a MW of 15 kDa which fits the monomeric state of ACP$_I$. Apo-ACP$_I$ only elutes with 15 kDa indicating that the ACP is monomeric and SFP modification might lead to some aggregation for holo, acetyl- and HMG-ACP$_I$.

Table 14: Result of the SEC-MALS measurement. For analysis the UV curve of the elution with the weight determination **(Fig. 43)** was divided into two parts. For each sample is indicated the averaged molecular weight of the eluted molecules.

Protein	Molecular weight Front part [kDa]	Error	Molecular weight Back part [kDa]	Error
Apo ACP$_I$			15,57	0,47
Holo ACP$_I$	21,64	0,22	15,58	0,62
Acetyl ACP$_I$	23,31	0,7	15	0,3
HMG ACP$_I$	20,34	0,3	14,74	0,29

7.2 Segmental labeling

NMR is the ideal tool to characterize the structure and dynamics of biological interesting proteins, improving the comprehension of complex biological problems. However, NMR spectroscopic analysis of multiple domain proteins, often suffer from high signal overlap. This

problem becomes even more profound when proteins contain recurring modular domains or highly disordered regions, like in the case of the of three consecutive ACP ($ACP_{I,II,III}$) which appear in the curacin A assembly line. High sequence identity (100-93%) among all three modules of this investigated protein leads to a severe peak overlap in NMR spectra, which would make NMR structure investigation impossible (**Fig. 44**).

```
T1              -TPQVNQVNLSEIKQVLKQQLAEALYTEESEIAEDQKFVDLGLDSIVGVEWTTTINQTYN 59
T2              -TPQVNQVNLSEIKQVLKQQLAEALYTEESEIAEDQKFVDLGLDSIVGVEWTTTINQTYN 59
T3              PLPQP-QVNLSEIKQVLKQQLAEALYTEESEIAEDQKFVDLGLDSIVGVEWTTTINQTYN 59
                  **  **************************************************

T1              LNLKATKLYDYPTLLELSGYIAQILSSQGTKPISSSSQTQQSLKTLQPLP 109
T2              LNLKATKLYDYPTLLELAAYIAQTLASQGTKPQVS----QQPLKTLQ--- 102
T3              LNLKATKLYDYPTLLELAPYIAQEIAATGGSKL----------------- 92
                ****************: **** ::: * .
```

Fig. 44: Clustal W multiple alignment of the triplet ACPs $ACP_{I,II,III}$ of CurA with each other.

Using segmental isotopic labeling to incorporate stable isotopes only into one desired domain of the three-domain protein solves the problem and allows distinguishing the different domains. We applied segmental isotopic labeling to individual domains of curacin A by a combination of *in vivo* and *in vitro* approaches of protein ligation. Segmental isotopic labeling turned out to be an essential tool for investigating the structure and interaction of the three consecutive ACP domains.

The central idea of our approach is to apply segmental isotopic labeling to the ACP_{II} middle domain in order to reduce the complexity of the spectrum but to keep the domain in its native context. One promising approach therefore is to ligate three polypeptide fragments using protein *trans*- splicing (PTS) with two split inteins (denoted in **Fig. 45** as Int1 and Int2). PTS does not require any cumbersome steps to create an N-terminal cysteine needed for expressed protein ligation (EPL) (**Fig. 45a)**) [120, 144, 180] (**4.4.3**). One of the main aspects to consider for this approach is that the two split inteins used, must be orthogonal without any cross-reactivity and should have a high splicing efficiency. Therefore DnaE inteins cannot be directly used because the domain or fragment inserted in the middle of the precursor protein containing both N- and C-intein (I_N/I_C) resulted in cyclization or polymerization (**Fig. 45 b)**) [122, 127, 128, 181, 182]. To solve this problem Hideo Iwai and his group exploited engineered *Npu*DnaE inteins with the same sequence but different split sites [141]. Shortening for example the C-intein prevented the formation of a functional intein with the unmodified N-intein (I_{N1}). The splicing activity, however, can be restored in reactions with a modified N-intein (I_{N2}) bearing the missing sequence (**Fig. 45 c) (4.4.5)**). We used these engineered inteins in order to set up an easy system and protocol for the segmental isotopic labeling of a middle domain. The results of this approach will be presented in detail on the next pages.

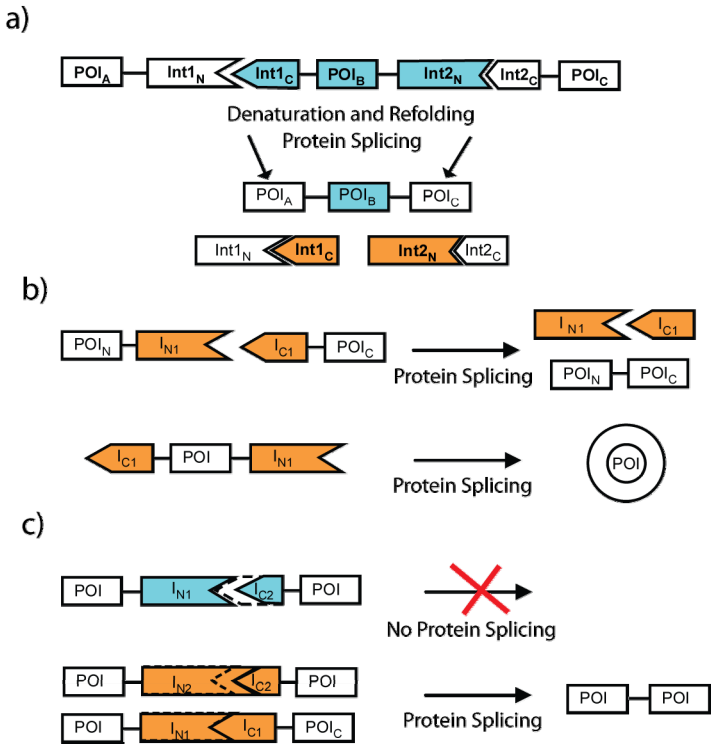


Fig. 45: Sheme for three fragment ligation a) Segmental isotopic labeling of a central region of protein of interest (POI) by three-fragment ligation using two orthogonal split inteins (Int1 and Int2). b) The direct use of inteins for the design of a central precursor protein can lead to cyclization reaction in *cis*. c) The solution for cyclization reaction. Elimination of protein splicing activity can be achieved by shortening the C-intein. The restoration of protein splicing activity can be realized by complementation with an elongated N-intein. Note, however, that a combination of I_{N2} and I_{C1} shows cross-reactivity and should therefore be avoided [129].

7.2.1 The ligation site

As mentioned in the introduction, the extein sequence has a high impact on the fact if the ligation reaction can take place. Several aspects play therefore a role:

a) Accessibility of the ligation site [111, 127]

b) Size of the amino acid prior to the active site [111, 183]

c) Compatibility of extein and intein amino acid composition [141].

In a first approach we decided to simplify the three fragment ligation to a two fragment ligation between ACP_I and $ACP_{II,III}$. The ligation site was chosen at the end of helix IV which was defined based on structure modeling. This first ligation version (V1) is presented in a sequence alignment (**Fig. 46** V1 in red). This first approach failed and will not be treated in more detail in this work. The ligation sites which were chosen in a second approach are indicated in green (**Fig. 46** V2 in green). The ligation sites where chosen to preserve as much as possible the native amino acid sequence. However for the ligation between ACP_I and ACP_{II} the sequence ISSSSQGT will be changed to GSCFNGT due to ligation and between ACP_{II} and ACP_{III} QVSQQPL will be changed to GSCFNGT after ligation. Nevertheless determining a suitable ligation site is the first important step which is then followed by several further optimization steps presented and explained in detail on the following pages.

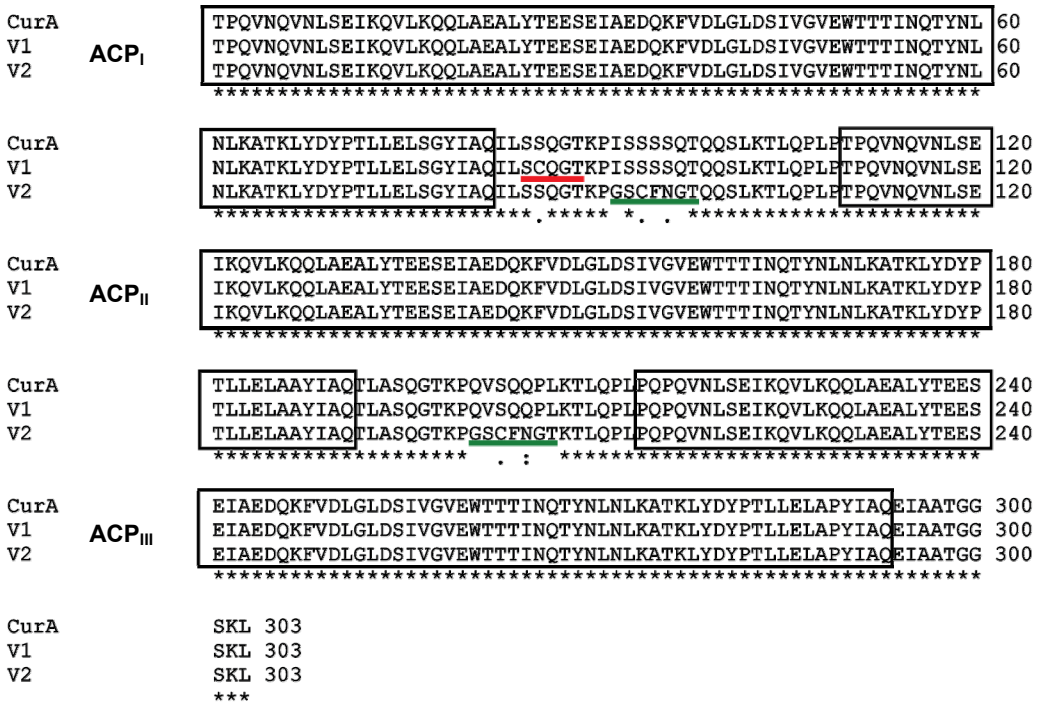


Fig. 46: Ligation sites used for segmental labeling. V1 is the ligation site used for the first, unsuccessful approach. V2 shows the sequence which originates if the two ligations between ACP_I and ACP_{II} and ACP_{II} and ACP_{III} are realized within the 2nd approach.

During the optimization process several constructs were created and are listed in **Table 15**. Cloning procedures are not described in detail for each of the plasmid but for some of them (pABRSF-1, pABRSF-3 and pABBAD-15 (**6.5.1**)), which were used for the labeling of the middle domain in the three fragment ligation (**7.2.6**) (Shematic presentation of these vectors in **Fig. 47**).

Table 15: Constructs which were cloned and used for three fragment ligation optimization.

Vector	Protein	Size (kDa)	Extinction	PI
pABRSF-1	H6SMT3 ACP_I Δ15	37,410	0,815	4,98
pABBAD-1	C15 ACP_{II}	13,496	0,960	4,67
pABBAD-2	H6C15 ACP_{II}	15,659	0,827	6,02
pABRSF-2				
pABCDF-1				
pABBAD-17	H6C36 ACP_{III}	17,502	0,91	6,43
pABBAD-11	H6 ACP_{II} Δ36	24,452	1,064	4,83
pABBAD-12	ACP_{II} Δ36	22,289	1,16	4,36
pABBAD-13	C15 ACP_{II} Δ36	25,575	1,02	4,49
pABBAD-14	H6C15 ACP_{II} Δ36	27,739	0,943	4,98
pABBAD-15	C36 ACP_{III}	15,339		5,15
pABBAD-16	C15 ACP_{III}	12,892		4,58
pABBAD-18	H6C15 ACP_{III}	15,055		5,84
pABRSF-3	H6SMT3c15 ACP_{II} Δ36	38,501	0,72	4,96
pABRSF-49-36	H6SMT3GB1Δ15	33,594	0,86	5,11
pABBAD-19	H6SMT3 C16 ACP_{II} Δ36	38,501	0,72	4,96

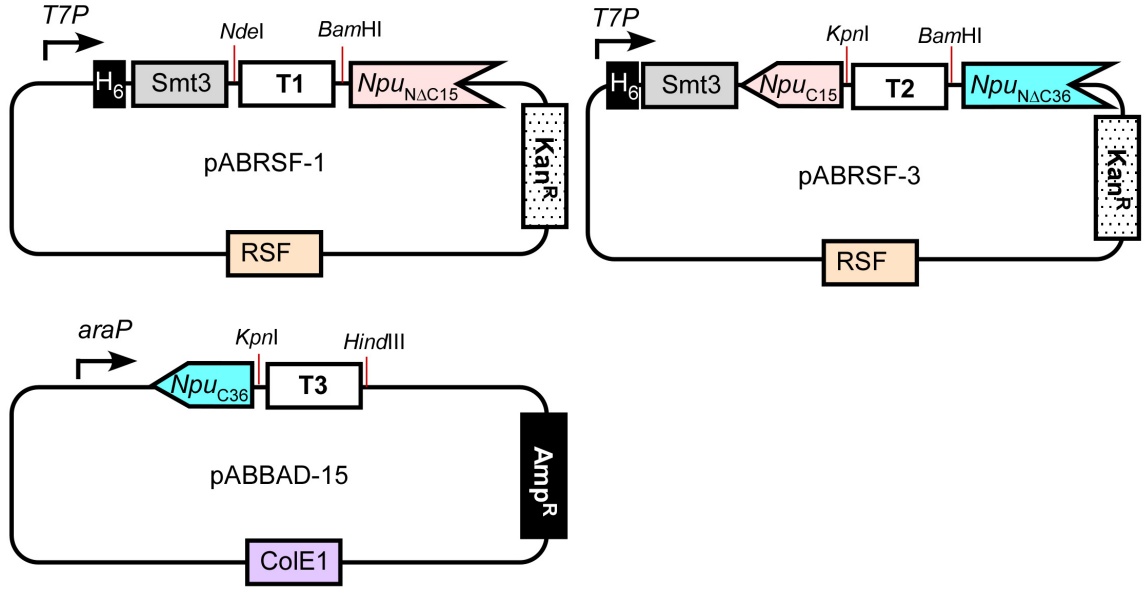


Fig. 47: Maps of the plasmids pABRSF-1, pABRSF-3 and pABBAD-15 used for segmental isotopic labeling of CurA $ACP_{I,II,III}$ by three fragment ligation. T1= ACP_I, T2= ACP_{II}, T3= ACP_{III}

7.2.2 Ligation between ACP_{II} and ACP_{III}

The ligation of two fragments can be evaluated in a fast approach by using the dual expression system that is explained in chapter **4.4.4**. In order to evaluate if the ligation between ACP_{II} and ACP_{III} occurs, we modified the protocol slightly from the protocol used for segmental labeling. In a first approach the fusion between ACP_{II} and ACP_{III} was the main focus. Therefore BL21 cells were co-transfected with the vector pABBAD-12 (ACP_{II}-ΔC36, Ara) and pABRSF-2 (H_6-c36-ACP_{III}). Cells were grown in 4 ml LB to an OD of 0.4 and then transferred to 30°C and 20°C after 20 min adaption the cells were induced using 0.04% Arabinose. After 1.5 hrs they were further induced by 0.4 mM IPTG.

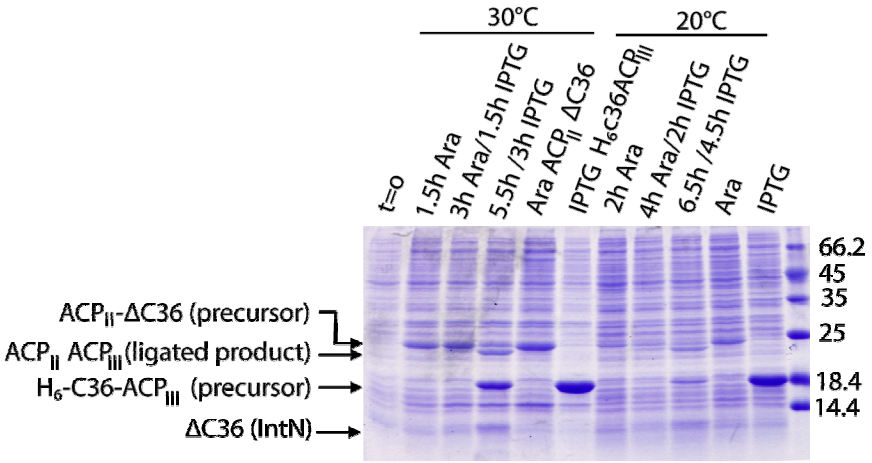


Fig. 48: 18% SDS gel showing the ligation between the proteins ACP_{II}-ΔC36 and H_6-C36-ACP_{III}. At 20°C and 30°C co-expression leads to ligation. The MW of the precursors and product as well as the intein is indicated on the left side with an arrow.

Samples were taken in a time course manner: before arabinose induction, after arabinose induction and after IPTG induction and analyzed using SDS-PAGE (**Fig. 48**). For the expressions at 30°C we observed after 1.5 hrs a band which occurs at the size corresponding to the size of ACP_{II}-ΔC36 (22, 29 kDa). After 3 hrs total expression (containing 1.5 hrs IPTG induction) nothing changed. After 5.5 hrs a band appeared corresponding to the size of the 2nd precursor (H_6-C36-ACP_{III}, 17.5 kDa). In addition a fainte band, smaller than 14 kDa appeared corresponding to the intein (ΔC36, 11.8 kDa) and indicates therefore, that a splicing reaction took place. After 5.5 hrs a 4th band which is slightly smaller then the precursor ACP_{II}-ΔC36 appeared corresponding to the product of the ligtion reaction. At 20°C the intein and the product gave a light band after 6.5 hrs. We can conclude that 1st the ligation between ACP_{II} and ACP_{III} took place and 2ndly that the temperature has a high impact on the reaction. 30°C seemed to be advantageous for the yield and therefore better suited for the ligation reaction.

7.2.3 Engineering the middle fragment

Cyclization

In order to perform a ligation of the middle fragment, ACP_{II} needs to be fused to inteins on its C- and N-terminus. We showed that ACP_{II} fused on its C-terminus to ΔC36 reacts with c36-ACP_{III} (**7.2.2**). Fusing ACP_{II} at its N-terminus to a intein which cross-reacts with ΔC36 might lead to cyclization (**4.4.5**) [131, 184]. In order to prove this hypothesis our cooperation partners created H_6-C36-ACP_{II}-ΔC36 and expressed it for 3 hrs, taking samples at the beginning, after one and three hrs (**Fig. 49**). After three hours a band appeared smaller than 14 kDa representing the intein, indicating that a splicing reaction took place. A second band appeared at a much smaller molecular weight representing the cyclized ACP_{II} which runs at a much smaller molecular weight then the uncyclized version (compare with **Fig. 48**). Even when co-expressed with H_6-C36-ACP_{III} no *trans*-ligation was observed (C36-ACP_{II}-ACP_{III}, ~24 kDa) which would be expected at slightly smaller size than H_6-C36-ACP_{II}-ΔC36 (~26 kDa) (**Fig. 49b)**). Our experiment shows that the intra-molecular reaction is entropically favoured, protein splicing in *cis* of ACP_{II} takes place immediately after protein synthesis. Therefore, cyclization by *cis*-splicing dominates the ligation (or polymerization) reaction by protein *trans*-splicing.

In our case using the same split inteins for the central fragment leads clearly to a circular form of the central target protein and *trans*-splicing is not detectable (<5%). This is an observation which has been made previously and could be strategically used to produce cyclized proteins to stabilize them [122, 130, 131]. For three fragment ligation, however, cyclization must be avoided.

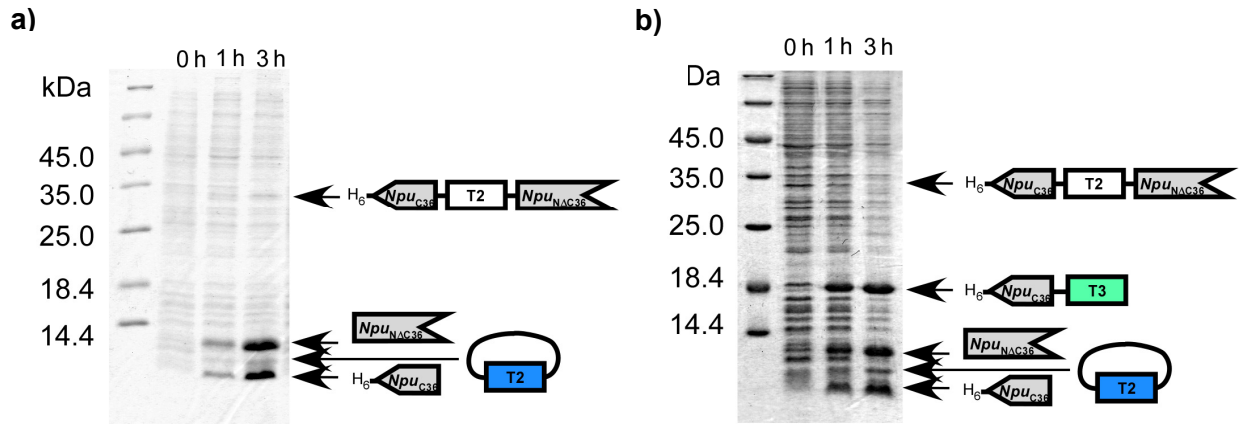


Fig. 49: Cyclization of ACP$_{II}$ a) *Cis*-splicing by C36-ACP$_{II}$-IntΔC36 resulting in the spontaneous protein splicing in *cis* after induction. b) Co-expression with ACP$_{III}$ using the same intein with the original site indicates little or no *trans*-splicing. Due to space limitations ACP$_{II}$ and ACP$_{III}$ are indicated by T2 and T3, respectively.

Solubility

In order to prevent our middle domain from this intra molecular reaction we used an N-terminal intein fusion which can not react with C-terminal intein fusion. The intein pair ΔC15 and C15 matches our requirements. ΔC15 is highly reactive with c15, but c15 can not react with ΔC36 because of the missing 20 aa which would be required for restoring the activity. First we were interested in evaluating the ligation efficiency of our new reaction partners. pABBAD-13 contains the ACP$_{II}$ construct (C15-ACP$_{II}$-ΔC36) and pABRSF-2 as before the ACP$_{III}$ construct (H$_6$-C36-ACP$_{III}$).The reactivity of the new constructs were monitored as described before (**7.2.2**). The first induction at an OD of 0.4 at 30°C (best temperature from **7.2.2**) was perfomed with 0.04% arabinose and 0.1% arabinose, followed by 0.4 mM IPTG after 1 hr. Time course analysis by SDS-PAGE reveal that we obtained in both cases a product band appearing below the precursor (C15-ACP$_{II}$-ΔC36, 25.5 kDa) (**Fig. 50 a)**). Therefore the N-terminal fusion of c15 does not prevent the reaction observed before between ACP$_{II}$ and ACP$_{III}$ in **7.2.2**. It is clearly shown that we obtain a stronger product band for the induction with a lower arabinose concentration.

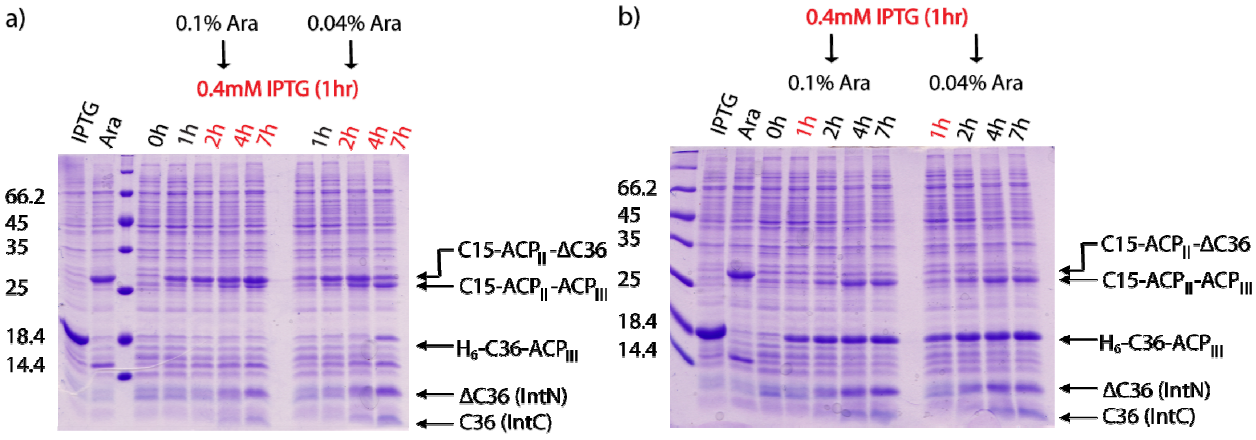


Fig. 50: Time course of c15-ACP$_{II}$-ΔC36 coexpression with H$_6$-c36-ACP$_{III}$. The ligation of the two precursors was performed. a) The expression was started with arabinose induction for the precursor c15-ACP$_{II}$-ΔC36 followed by IPTG induction for H$_6$-c36-ACP$_{III}$. b) The expression was started with IPTG induction for the precursor H$_6$-c36-ACP$_{III}$ followed by arabinose induction for c15-ACP$_{II}$-ΔC36.

However after 7 hrs we obtained beside the product, residual unreacted precursors. On one hand some amounts of the precursors might not have reacted due to the equilibrium of the reaction, on the other hand it is possible that the reaction was inhibited due to low solubility of one of the precursors or both. We repeated the experiment with inverse induction. In case the IPTG induced protein has a higher solubility we should obtain higher ligation efficiency. First we induced the cells with 0.4 mM IPTG followed by arabinose induction after 1 hr (**Fig. 50 b)**). In this case we obtained much more H_6-C36-ACP_{III} in the cells but no excess of C15-ACP_{II}-ΔC36, instead, only the band corresponding to the ligation product occured. We conclude that C15-ACP_{II}-ΔC36 protein suffers from low solubility. If expressed first, it is stored in inclusion bodies and only low amounts can react with the 2nd precursor. This might become a problem for the labeling procedure which requires longer expression times and a M9 medium which usually means more stress for the cells effcting even more the solubility of proteins in a negative way. Therefore we decided to use Smt3 as a solubility tag to overcome solubility problems.

7.2.4 Ligation of ACP_I and ACP_{II}

The amino acid composition in the linker region between ACP_I and ACP_{II} differs from ACP_{II} and ACP_{III} (**Fig. 46**). Before going for three fragment ligation we wanted to assure that the ligation between ACP_I and ACP_{II} takes place as well to elimate eventual problems which might occur later on. Therefore, we co-transformed the vector pABBAD-1 (c15-ACP_{II}) and pABRSF-1 (H_6-Smt3-ACP_I-ΔC15, 37.4 kDa) and induced the proteins first by arabinose 0.04% for 2 hrs followed by 0.4 mM IPTG. The ligated product H_6-Smt3-ACP_I-ACP_{II} (34.9 kDa) is very similar in size to the ACP_I-precursor and is therefore difficult to distinguish on a SDS gel. Nevertheless the fact that a reaction occurred, can be seen firs by the inteins which appear, secondly by the product band which appears slightly smaller then the ACP_I–precursor. No big difference is observable for the reaction performed at 20°C in comparison with the reaction performed at 30°C (**Fig. 51**).

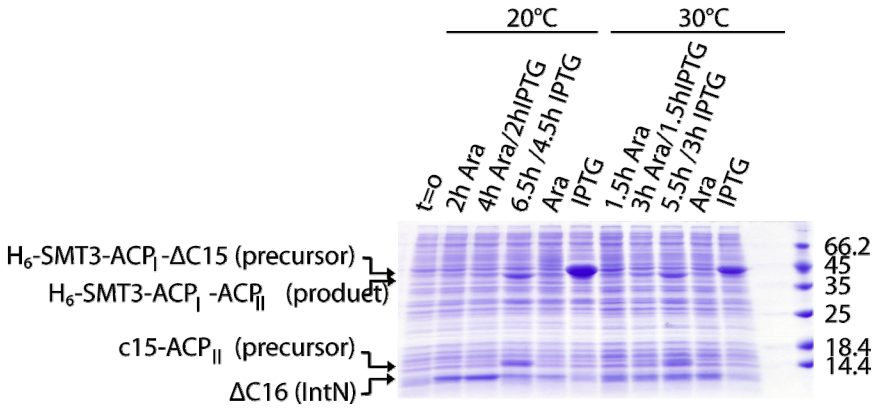


Fig. 51: Time course of c15-ACP_{II} co-expressed with H_6-SMT3-ACP_I-ΔC15. The ligation of the two precursors was performed at two different temperatures. The molecular weight of precursors and product is indicated by arrows on the left side.

7.2.5 One pot reaction

We showed that H_6-Smt3-ACP_I-ΔC15 reacts with C15-ACP_{II} (**7.2.4**) and furthermore that C15-ACP_{II}-ΔC36 reacts with H_6-C36-ACP_{III}. Furtermore it was of interest to perform a one pot reaction by co-expressing of H_6-Smt3-ACP_I-ΔC15, C15-ACP_{II}-ΔC36, and H_6- C36-ACP_{III}. Therefore, the proteins must be all coded on different vectors with different ORIs and different antibiotic resistences (pABBAD-13 [Amp^R, arabinose], pABRSF-1 [IPTG, Kan^R], pABCDF-1 [IPTG, $Strep^R$] (see **Table 15**)). In this case pABRSF-1 (H_6-Smt3-ACP_I-Δ15) and pABCDF-3 (H_6-C36-ACP_{III}) are both under IPTG control but can not cross-react due to the missing 20 amino acids. We performed the expression protocol as usual by varying the order of expression and the arabinose concentration (**Fig. 52**:). In all cases a band appeared at 30 kDa which was never observed before and is presumably a stress protein expressed as stress response to tzhe different antibiotics. In general the SDS gels (**Fig. 52**: **a) and b)**) looked comparable to the SDS gels from the ACP_{II} and ACP_{III} ligation (**Fig. 48**). An explanation herefore is that the biggest precursor H_6-Smt3-ACP_I-ΔC15 was not at all expressed. The ACP_{II} and ACP_{III} precursors are expressed and ligated as shown before (**7.2.2**).

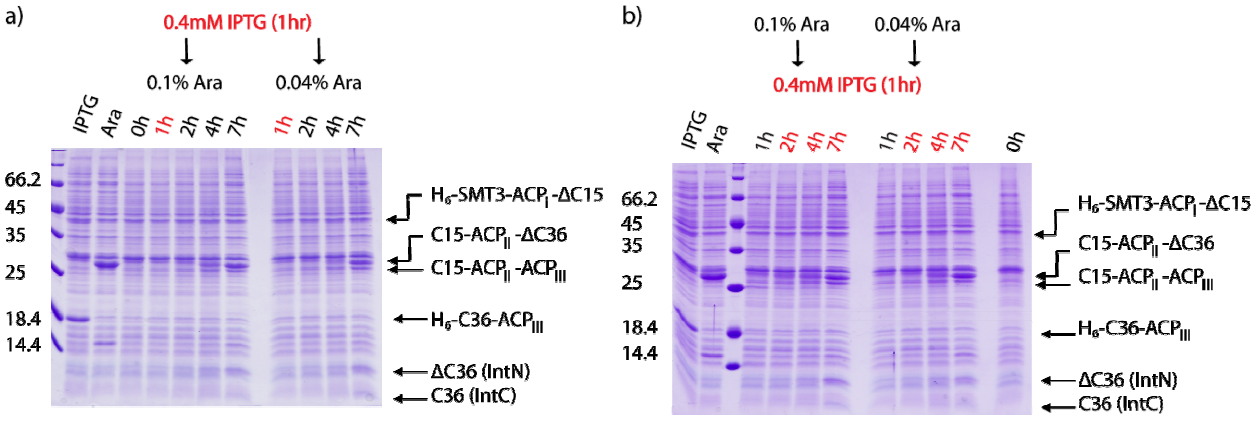


Fig. 52: Three fragment ligation by one pot reaction. The three plasmids expressing pABBAD-13 (c15-ACP_{II}-ΔC36), pABRSF-1 (H_6-Smt3-ACP_I-ΔC15), pABCDF-1 (H_6-c36-ACP_{III}) were co-transformed in BL21 a) Time course of the co-expression inducing first H_6-Smt3-ACP_I-ΔC15, H_6-c36-ACP_{III} by IPTG followed by c15-ACP_{II}-ΔC36 expression b) Time course of the co-expression inducing first c15-ACP_{II}-ΔC36 by arabinose induction followed by IPTG induced H_6-Smt3-ACP_I- ΔC15, H_6-c36-ACP_{III} expression.

7.2.6 The final three fragment ligation protocol

Based on the results shown in the previous chapters (**7.2.1- 7.2.5**) we developed the following strategy for the three fragment ligation for labeling the middle domain. The first ligation step between ACP_{II} and ACP_{III} was performed *in vivo* using the precursor H_6-Smt3-c15-ACP_{II}-*Npu*NΔC36 (pABRSF-3 (**Fig. 47**)) and *Npu*C36-ACP_{III} (pABBAD-15 (**Fig. 47**) on the basis of time delayed dual expression (**4.4.4**) followed by Ni-purification of the *NpuC*15-[^{15}N]ACP_{II}-ACP_{III} precursor. In a 2nd step, the *in vitro* ligation step the precursor from step one was mixed with H_6-Smt3-ACP_I-*Npu*NΔC15 which was expressed and purified before individually (**Fig. 47**). The *in*

vitro ligation was then followed by a ULP digestion and reverse Ni-purification leading to ACP_I-[^{15}N]ACP_{II}- ACP_{III} (**Fig. 53**).

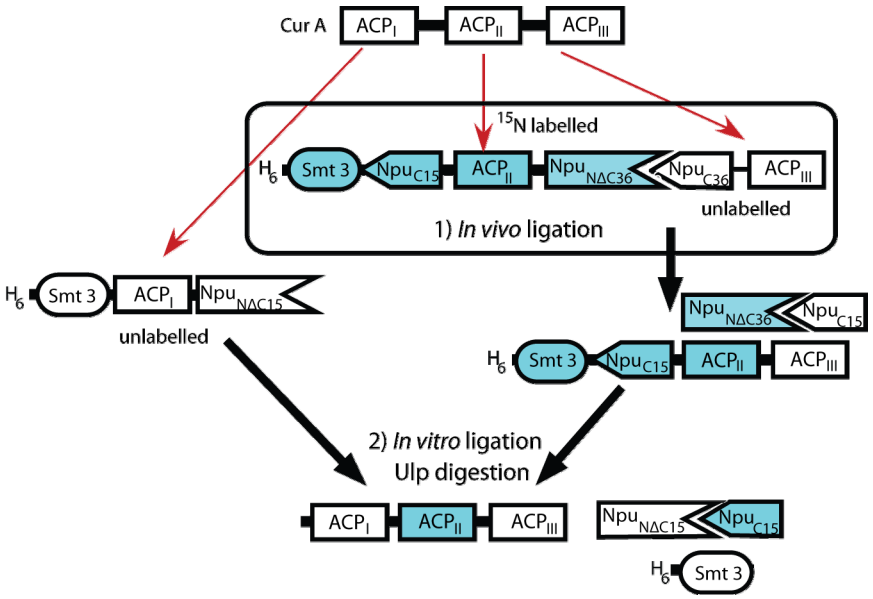


Fig. 53: Strategy for segmental isotopic labeling of a central domain in CurA.

In vivo reaction

For the *in vivo* reaction BL21 cells were co-transfected with the plasmids pABRSF-3 (H_6-Smt3-*Npu*C15-ACP_{II}-*Npu*NΔC36) and pABBAD-15 (*Npu*C36-ACP_{III}) and grown at 37°C in 2 L of LB medium supplemented with two antibiotics 100 µg/ml ampicillin and 25 µg/ml kanamycin and 0.04 mM Fe-II-SO_4. The C-terminal precursor protein (*Npu*C36-ACP_{III}) was first induced with 0.1% (w/v) L-arabinose at an OD of 0.4-0.5 after 20 min adjustment to to 30°C. After 4 hrs the culture medium was replaced with 2 L of ^{15}N-labeled M9 medium supplemented with 35 µg/ml kanamycin, 0.05 mM $FeCl_2$ and a final concentration of 0.4 mM IPTG. After further 5 hrs of the second induction, the cells were harvested and Ni-NTA purified later on. In case of the *in vivo* ligated product, it has accidentally the very same apparent size (37,87 kDa) as the precursor protein (H_6-Smt3-*Npu*C15-ACP_{II}-*Npu*NΔC36, 38,5 kDa) and therefore the proteins cannot be distinguished in SDS gel (18%), thus the gel of the *in vivo* ligation shown in **Fig. 54** for the segmental isotopic labeling provides little information for demonstrating the *in vivo* ligation. However, we know the ligation was successful due to a) the intein band which appeared and b) the decreased c36-ACP_{III} band and c) because it was tested without the Smt3 fusion as well (refer to **Fig. 50**).

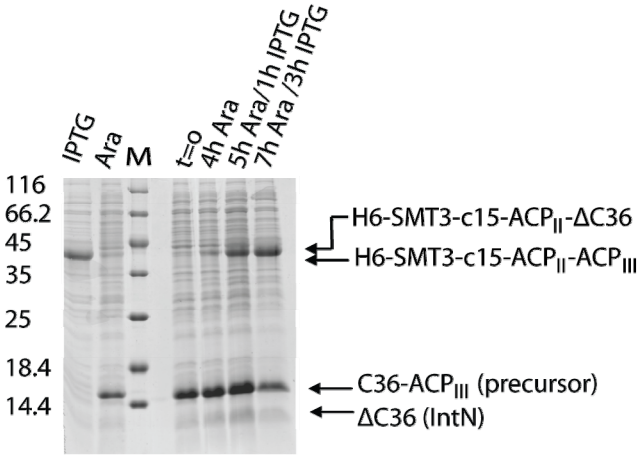


Fig. 54: The first step of the segmental labeling protocol. SDS gel analysis of the *in vivo* ligation between H_6-Smt3-c15-ACP_{II}-ΔC36 and c36-ACP_{III}. Ligation product (H_6-Smt3-c15-ACP_{II}-ACP_{III}) and precursor (H_6-Smt3-c15-ACP_{II}-ΔC36) are impossible to differentiate on the SDS gel due to the similar molecular weight. Less precursor of c36-ACP_{III} after 7 hrs expression and appearing intein bands indicate that a reaction takes place. The different proteins are indicated with arrows on the right side.

In vitro reaction

H_6-Smt3-ACP_I-*Npu*NΔC15 was expressed and purified individually. About 20 mg of H_6-Smt3-*Npu*C15-ACP_{II}-ACP_{III} was mixed with a total amount of 5 mg H_6-Smt3-ACP_I-*Npu*NΔC15. The final concentration of H_6-Smt3-ACP_{II}-ACP_{III} was 50 µM in the reaction mixture. H_6-Smt3-ACP_I-*Npu*NΔC15 had a final concentration of 11 µM in the reaction mixture. The reaction mixture was incubated at room temperature in PBS in the presence of 0.5 mM TCEP for 10 hrs. In the reaction mixture yeast ubiquitin-like protein-specific protease 1 (Ulp1) was added for removal of the H_6-Smt3 fusion tag. After 10 hrs the buffer was exchanged into the ligation buffer 10 mM Tris, 500 mM NaCl, 0.5 mM EDTA, 0.5 mM TCEP was continued for further 7 hours at 25°C with agitation at 200 rpm. Samples were taken for SDS analysis of the *in vitro* ligation (**Fig. 55**). After the reaction, the H_6-Smt3 fusion tag was removed by reverse Ni-purification. The ligation product was further purified with ion exchange chromatography (MonoQ). The total amount of segmentally ^{15}N-labeled ACP_I-[^{15}N]-ACP_{II}-ACP_{III} was 1.7 mg (more detilas about the protocol in **6.5.2**).

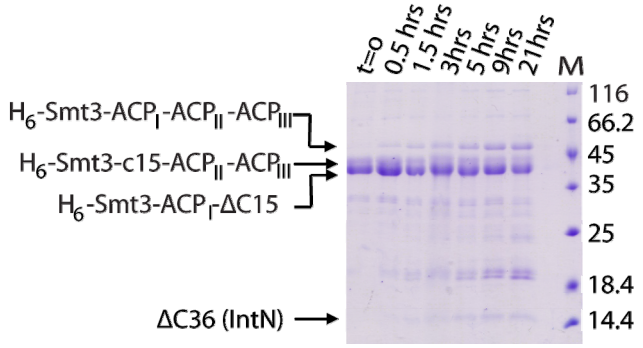


Fig. 55: Time course of the second *in vitro* step of the protein splicing reaction. 18% SDS Page shows the ligation between H_6-Smt3-ACP_I-ΔC15 and H_6-Smt3-ACP_{II}-ACP_{III}. A small band corresponding with the molecular weight of the product appears over time. The different proteins are idicated with arrows on the right side.

7.2.7 Evaluation of CurA $ACP_{I,II,III}$ containing the middle domain ^{15}N labeled

The different steps of the segmentally labeled protein have been observed via SDS-PAGE and the product was further verified via MALDI-MS. For 100% ^{15}N-labeled ACP_{II} domain in $ACP_{I,II,III}$ we would expect a molecular weight of 34.054 kDa. The measured molecular weight was 34.087 kDa. With respect to the high molecular weight for this technic and the peak broadening due to isotopic labeling the result matches the expected value (**Fig. 56**).

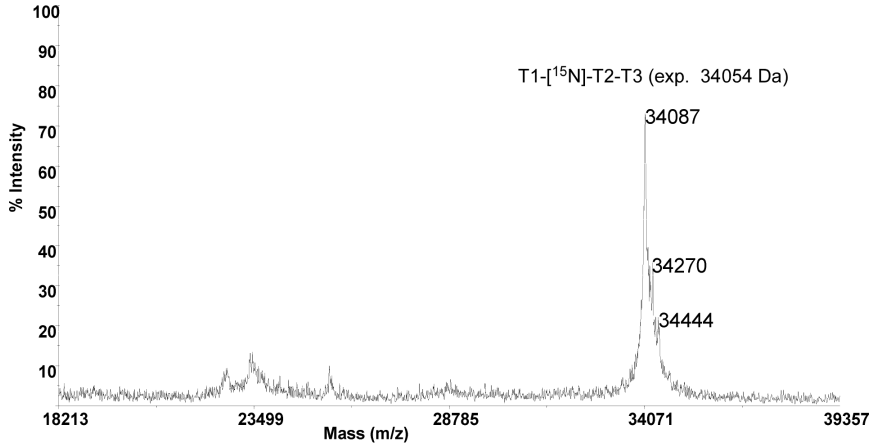


Fig. 56: Maldi-MS of the segmental Isotope-labeled ACP_I-[^{15}N]ACP_{II}-ACP_{III} protein.

We prepared different segmentally labeled samples in which the 1st (protocol **6.5.3**), then the 2nd and then the 3rd domain have been labeled ^{15}N. The overlay of the spectra shows that the peaks of the segmentally labeled domains perfectly match the peaks of the full length wt spectra. Slight differences are due to the modifications in the linker regions.

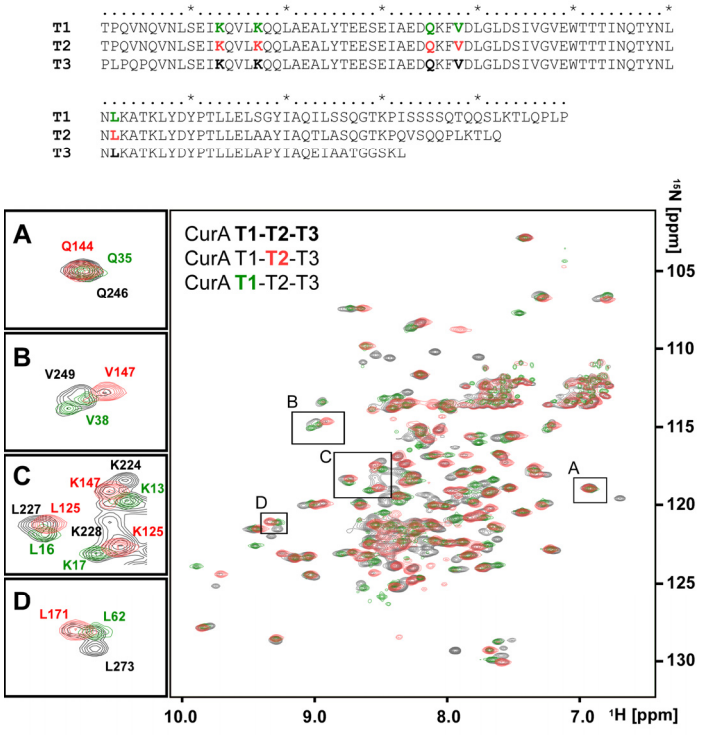


Fig. 57: Result of segmental labeling a) Primary structure of CurA containing three ACP domains. b) An overlay of [^{1}H,^{15}N]-TROSY spectra from uniformly ^{15}N-labeled CurA (ACP_I-ACP_{II}-ACP_{III}) (black) and centrally ^{15}N-labeled CurA (ACP_I-[^{15}N]ACP_{II}-ACP_{III}) (red), and N-terminally ^{15}N-labeled [^{15}N]ACP_I-ACP_{II}-ACP_{III} (green). Insets (A)–(D) are the magnifications of regions in the spectra. The assignments of a few residues are indicated.

Yields of the splicing reaction

Performing quantitative analysis of the yields of the splicing experiments with the ACP domains is difficult, due to the fact that the ligation product (37.876 kDa) of the first ligation steps has accidentally the same size as one of the precursors (38.501 kDa) such a quantitative analysis is in this case not possible. The yield for the 2nd *in vitro* ligation step, however, was determined to be 40%.

Isotope cross labeling

Isotope cross labeling might pose a major challenge for the three way ligation method using *in vivo* ligation as it is described. The protocol presented here is based on the effect that arabinose is used as an inducer for the first expression gets then metabolized and therefore can be effectively removed from the media upon media exchange [128, 142, 145]. A washing step further reduces the isotopic cross labeling. It was shown, that when applying the *in vivo* protocol described by us, the isotopic cross labeling is less than 5% [128, 142, 145]. To prove that in our particular triple ACP labeling experiment, however, cross labeling does not occur to any significant amount we have analyzed the corresponding NMR spectra. While many resonances of the individual domains severely overlap in the uniformly labeled sample, some resonances show chemical shift differences. For these amino acids the cross labeling percentage can be determined from the spectra of the uniformly and the segmentally labeled samples. These data are provided below and reveal a cross labeling of less than 5% (**Fig. 58 this page and next page**).

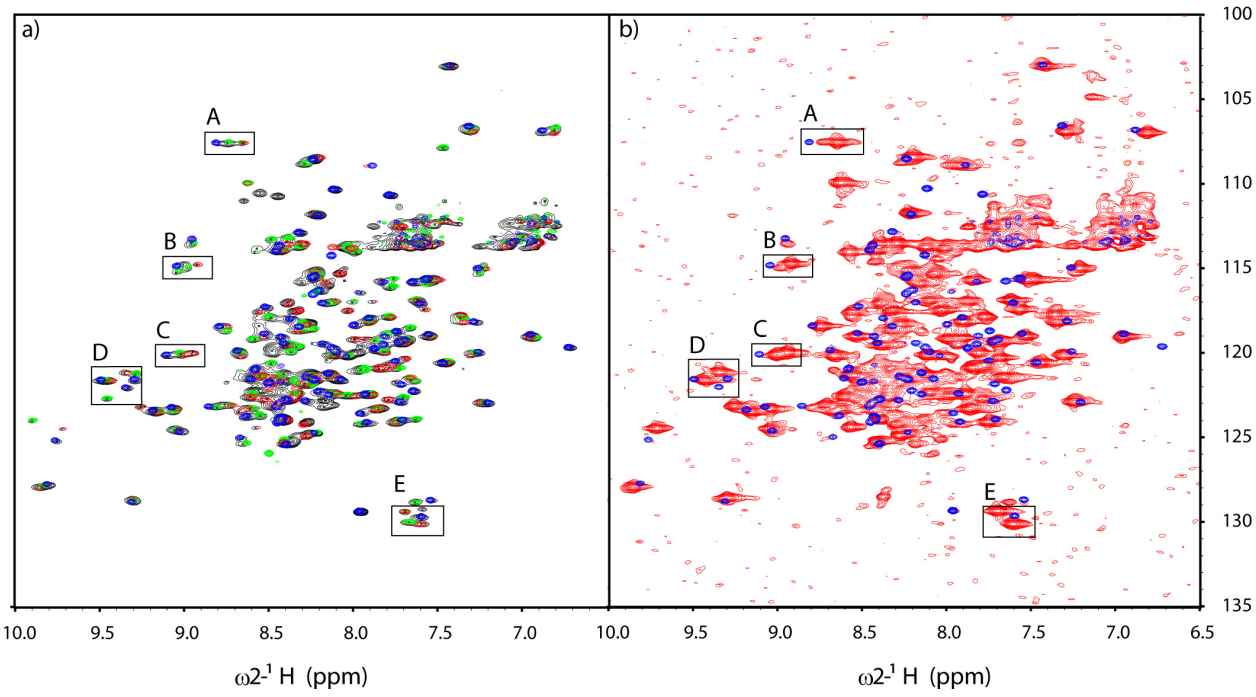
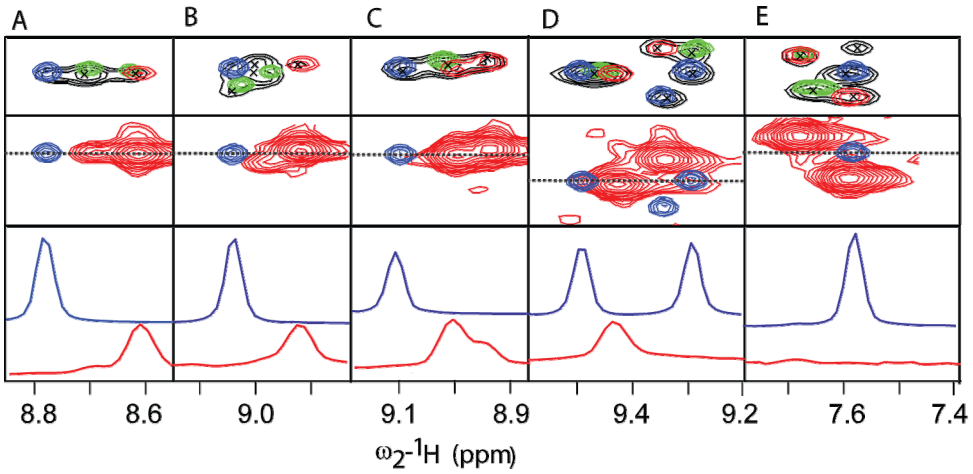


Fig. 58: Evaluation of cross labeling (a) shows an overlay of the HSQC spectra of the uniformly ^{15}N-labeled CurA (ACP_I-ACP_{II}-ACP_{III}) (black), of the central segmentally ^{15}N-labeled CurA (ACP_I-[^{15}N]ACP_{II}-ACP_{III}) (red), of the N-terminal segmentally ^{15}N-labeled [^{15}N]ACP_I-ACP_{II}-ACP_{III} CurA (green) and of the C-terminal segmentally ^{15}N-labeled ACP_I-ACP_{II}-[^{15}N]ACP_{III} CurA (blue). (b) Displays the central segmentally ^{15}N-labeled CurA (ACP_I-[^{15}N]ACP_{II}-ACP_{III})

(red) HSQC plotted at a lower level to show the background noise. Additionally the [^{1}H,^{15}N] TROSY spectra of the C-terminal, segmentally ^{15}N-labeled ACP$_{I}$-ACP$_{II}$-[^{15}N]ACP$_{III}$ (blue) is shown. In case of cross labeling red peaks would occur at the positions of the blue peaks. Insets (A-E) are magnifications of those regions indicated in (a) and (b). In addition, we show 1D slices for the C-terminal, segmentally ^{15}N-labeled ACP$_{I}$-ACP$_{II}$-[^{15}N]ACP$_{III}$ (blue) and the central, segmentally ^{15}N labeled CurA (ACP$_{I}$-[^{15}N]ACP$_{II}$-ACP$_{III}$) (red) taken at those positions that are highlighted with a dashed line. In the central segmentally ^{15}N-labeled CurA ACP$_{I}$-[^{15}N]ACP$_{II}$-ACP$_{III}$ (red) no peaks are detectable at the positions of the ACP$_{III}$ resonances. Therefore we can conclude that cross-labeling should be below the detection level of approximately 10%. These results confirm earlier investigations reported in [128, 145].



7.3 Characterization of the ACP domain

7.3.1 The triplet ACP$_{I}$-ACP$_{II}$-ACP$_{III}$

The triplet ACP domains are located at the C-terminus of the CurA protein from *Lyngbya majuscula*. ACP$_{I,II,III}$ (residues 1946-2248) further dimerizes through a C-terminal domain (C$_{d}$-domain, residues 2261–2311) [76]. Our initial objective was to evaluate whether the presence of the triplet ACP$_{I,II,III}$ domains or their dimer leads to domain-domain interactions that adopt a prefered conformation. [^{15}N,^{1}H]-TROSY spectra of the two proteins CurA-ACP$_{I,II,III}$ with (in black) and without C$_{d}$ (in red) were recorded and compared in **Fig. 40**. The absence of chemical shift differences indicates, that no conformational changes occur. Some peaks decrease in intensity upon adding the 60 aa C$_{d}$ domain. However, surprisingly we do not observe new peaks representing the C$_{d}$ domain. Due to dimerization the overall size increases from 36 kDa to 89 kDa [76], which affects the rotational correlation time of the protein and most likely broadens the resonances of the C$_{d}$ domain beyond the detection limit. Parts which tumble independently, such as the individual ACP domains are affected less by this increase in size (**Fig. 40**). The [^{1}H,^{15}N]-TROSY of the ACP$_{I,II,III}$ domain showed severe peak overlap due to the high sequence identity (93-100%) (**Fig. 44**). Therefore a distinction of the individual domains in a three domain construct was impossible using usual techniques such as combinatorial labeling techniques. The use of segmental labeling allowed us to distinguish the three domains from each other and assign the backbone of the full length construct [129] (**Fig. 57**). Isolating ACP$_{I}$, ACP$_{II}$ or ACP$_{III}$ from the three domain construct CurA-ACP$_{I,II,III}$ showed us in [^{1}H,^{15}N]-TROSY experiments strong chemical shift perturbations (CSP). The assignment of the full length construct enabled

us to identify, that these strong CSP only occur at the N- and C-terminus. Both parts are disconnected from their natural environment upon isolation, explaining the CSP (**Fig. 59 next page**).

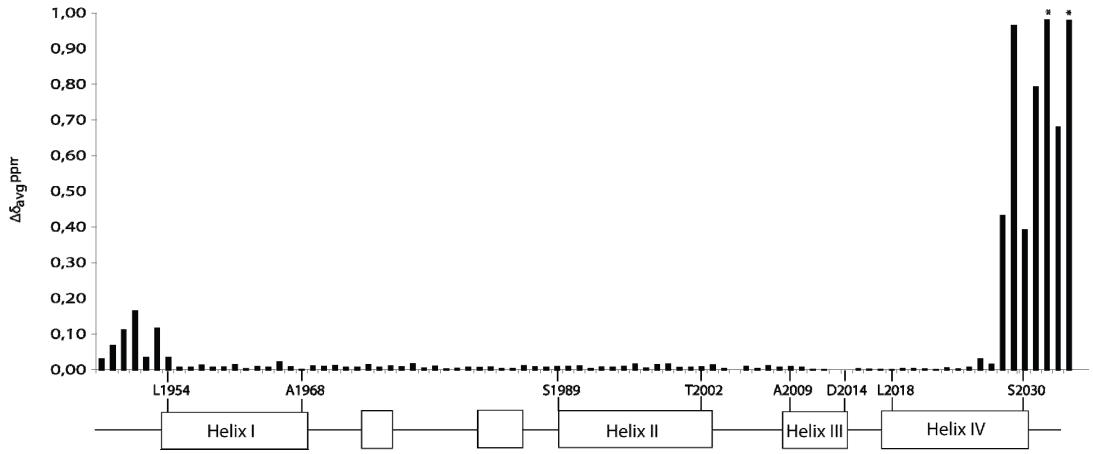


Fig. 59: A plot of chemical shift perturbations due to ACP_I isolation from CurA-$ACP_{I,II,III}$. Both proteins were in their holo forms. Due to presentational purposes the CSP for Q2032 and T2034- indicated with a star (*)- are not completely displayed. Their CSP is 1.95 ppm (Q2032) and 1.26 ppm (T2034)

These data support the aspect that there are no interactions between the individual domains. We performed additionally titration experiments of ACP_I (1946- 2043) with the didomain construct $ACP_{II,III}$ (2044-2248). No shifts were observed for ACP_I (Data not shown). In addition, no inter-domain NOEs were detected in 3D ^{13}C–edited/^{15}N-separated NOESY-HSQC experiments with a [^{15}N]ACP_I–[^{13}C]ACP_{II}-ACP_{III}, labeled version (data not shown). Taken together these data show that the ACP triple and didomains lack domain-domain interactions.

To investigate the dynamics of the ACP linker regions which were found to be structured in other systems [21, 29, 185] we performed {^1H}^{15}N-heteronuclear NOE measurements. Nearly all amino acids of the structured domains ACP_I, ACP_{II} and ACP_{III} was identified, the additional signals can therefore be associated with the linker region. Heteronuclear NOE values of around 0.8 represent highly structured regions (**Fig. 60 a) b) c)**) which is the case for the assigned amino acids within the ACP domains, whereas values around 0.2 were measured for the unassigned peaks, indicating that the linkers are unstructured and highly flexible (**Fig. 60 d)**).

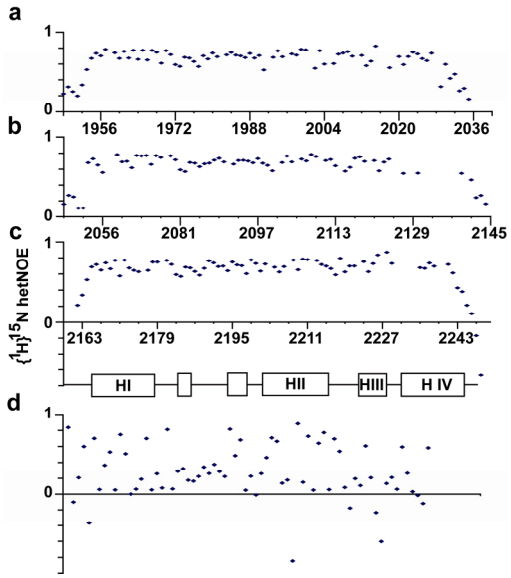


Fig. 60: Hetero-NOE measurements. Backbone $\{^1H\}^{15}N$ hetero-NOE values of ACP_I (a), ACP_{II} (b) and ACP_{III} (c) domains within the full length protein. Most values are around 0.8 reflecting a well defined structure. Towards the N- and C-terminus hetNOE values decrease continuously. The unassigned Peaks (d) are mostly located in the linker region between the domains. Those values are on average around 0.2 reflecting that these regions are most likely very flexible and unstructured.

7.3.2 Characterization of ACP_I

Taken together we showed in the previous chapter that the linker region between the ACP domains is unstructured and that the ACP domains behave independently from each other in $ACP_{I,II,III}$ as well as in the dimerized form of $ACP_{I,II,III}$-C_d. This leads to the conclusion that the interaction between Cur Hal and ACP domains can be studied using one representing ACP domain of the triplet, for example ACP_I (residues 1946-2043). The ACP_I domain was cloned into the pet28 vector with an N-terminal His_6-tag as recently described [69]. The histag showed to be structurally disordered and does not interact with the protein. Thus in the final sample the first 21 aa are non natural amino acids of which the first 11 amino acids including the histag are so flexible that they do not give raise to any signals in a $[^1H,^{15}N]$-TROSY. The following 10 aa give backbone NH signals but no NOEs, making them structurally undefined. The residues 22-110 are identical to the CurA ACP_I residues 1946-2043 (CurA, GenBank accession code AAT70096.1). Protein sample purity and quality assessment is described in **7.1.8** and **Fig. 41**. The protein was expressed during inhibition of the endogenous phosphopantetheinee transferase [22] and was *in vitro* modified using CoA and SFP. Expression and modification was verified via MALDI-MS and confirmed, that minimal 98% of the protein was in the apo form and minimal 98% were converted into the modified form (holo form) (**Fig. 39**). It has been shown that the 4'-Ppant arm is in presence of small substrates (acetyl, malonyl, buturyl) or without substrates solvent exposed and flexible and only interacts transiently with the ACP [34-36, 39]. In this work, we proceeded to determine the solution structure of holo ACP_I, HMG-ACP_I and the interaction site between HMG-ACP_I and Cur Hal.

7.3.3 Resonance assignment

The [^{1}H,^{15}N]-TROSY spectrum of ACP$_{I}$ is shown in **Fig. 62**. The spectrum is well resolved, with sufficient chemical-shift dispersion in both the ^{1}H and the ^{15}N dimension. Only in a few instances for the side chains the spectra suffered from peak overlap. Over 95% of the sequence-specific backbone and side chain assignments were achieved using the data from a series of triple-resonance NMR experiments described in materials and methods (**6.3.1**).

7.3.4 Quality of the calculated structure

Backbone torsion angle restraints were obtained from chemical shift data using the TALOS+ algorithm [161]. NOE-based distance restraints from 3D ^{15}N-NOESY-HSQC and ^{13}C-NOESY HSQC spectra were assigned automatically by CYANA [162, 163], which was also used for the structure calculations by torsion angle dynamics [164]. The final structure calculations included 2852 distance restraints and 162 backbone torsion angle restraints (**App. 10**). 100 conformers were computed using 10000 torsion angle dynamics steps. The 20 conformers with the lowest target function values were subjected to restrained energy refinement with the program OPALP [165] using the AMBER force field [166]. The quality of the structures was checked by PROCHECK [167] and WHATCHECK [168].

The structural statistics demonstrate good agreement of the structures with the experimental data, giving rise to no upper limit violations greater than 0.12 Å (**App. 9**). A measure of precision for this family of 20 structures is given by the superposition of residues Val1950–Gly2033 of the ensemble on the mean structure, with a backbone RMSD value of 0.24 Å and non- hydrogen atoms RMSD value of 0.54 Å. Ramachandran statistics (evaluated with PROCHECKNMR software v.3.5.4 fo residues Val1949–Thr2034 of the ensemble of 20 structures) [167] show 88.6% of the backbone torsion angle pairs occupying the most favored regions, 11.4% in the additionally allowed regions (**App. 9**).

7.3.5 Tertiary structure

The overall structure of holo-ACP$_{I}$ consists of a right-hand twisted bundle formed by four major α-helices (I,II,III,VI) connected by three loops. Helix II and VI are orientated parallel to each other. Helix I is turned around 45° to the orientation of helix II and IV. The third helix (III) is much shorter in size and lies nearly perpendicular to the helices II and IV. **Figure 61** shows the ribbon diagram of the minimized mean structure. The observed right-twisted helical bundle topology matches the conserved fold for the ACP and PCP family [7, 10, 34-36, 41-43, 52, 186, 187], with helix I (residues 1954–1968) running anti-parallel to helices II (residues 1989–2002), the perpendicular orientated helix III (residues 2009–2014) and the C-terminal helix IV (2018-2030). Helix I, II, IV are quite similar in size ranging from 14 (helix I) to 13 (helix II) and 12 (helix IV) amino acids in length. Helix III is comparably short ranging only over five amino acids.

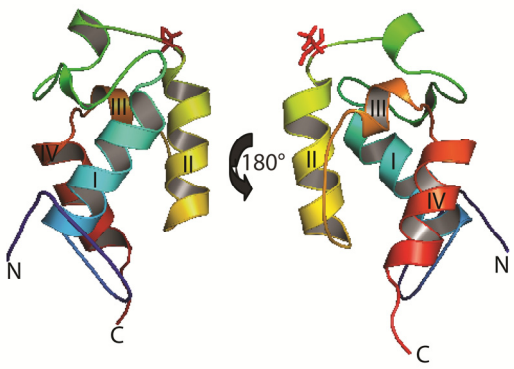


Fig. 61: Structure of holo ACP$_I$. Ribbon diagram of the averaged and minimized NMR structure of holo ACP$_I$ from *Lyngbya majuscula*. Helixes I–VI are colored light blue, yellow, orange-brown and red, respectively. The active ser1989 is indicated in red at the N-terminus of helix II

The long loop between helices I and II contains two structurally well defined areas (1973-1975; 1982-1985) the rest of the loop is less defined. Hydrophobic interactions stabilize the helical bundle, as evidenced by the NOE signals between core leucine, valine, alanine and tryptophan residues. For example, helix I and helix II are aligned by the NOEs between Leu1961 and Trp1985, Ala1968 and Trp1985. Furthermore we have a hydrogen bond between Phe1982 and Pro2016, which might stabilize the position of helix III.

7.3.6 Comparison of different chemical states

During the biosynthesis ACP$_I$ is loaded with different substrates. The unloaded state without any cofactor is called apo form. The apo form undergoes interaction with the phosphophanteinyl transferase and gets activated by being loaded with the 4'-Ppant arm, called holo state [188] (Compare chapter **4.1.2**). In the following step ACP$_I$ interacts with the upstream located acyltransferase domain which loads ACP$_I$ with malonyl. In the proceeding step the upstream located KS decarboxylates the malonyl to acetyl and couples the previous activated acetyl with the acetyl giving 3-oxo-butanoyl. Being loaded with this diketone the ACP interacts with the HMG-cassette (CurC, CurD, CurB) resulting in a modification of 3-oxo-butanoyl to 3-hydroxy-3-methylglutaryl (**compare Fig. 15**). We investigated some of these different situations measuring [^{1}H,^{15}N]-TROSY spectra of apo, holo, acetyl, 3-hydroxy-3-methylglutaryl (HMG) and 4-chloro-3-hydroxy-3-methylglutaryl (Cl-HMG) **(Fig. 62**). 4-chloro-3-hydroxy-3-methylglutaryl was obtained by a reaction of (*R,S*)3-hydroxy-3-methylglutaryl-ACP$_I$ with Cur Hal. Due to the stereo-specificity for (*S*)-HMG we obtained a 50% mixture of (*R*)-3-hydroxy-3-methylglutaryl-ACP$_I$ and (*S*)4-chloro-3-hydroxy-3-methylglutaryl-ACP$_I$. The loading state of all samples was checked by MALDI-MS (**Fig. 39**).

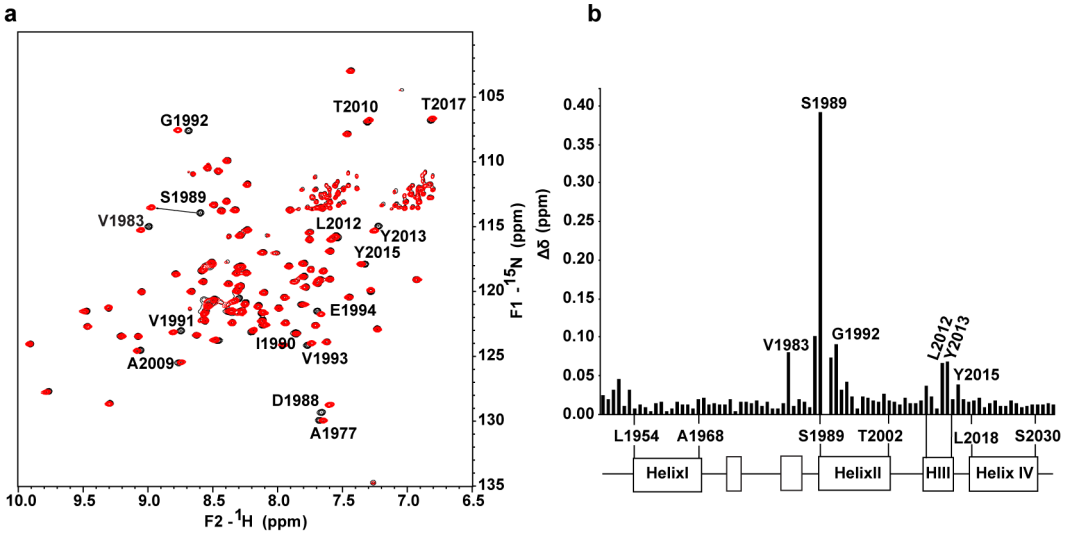


Fig. 62: Comparison of apo and holo ACP_I a) An overlay of [^{1}H,^{15}N]-TROSY spectra from uniformly ^{15}N-labeled apo ACP_I (black) and ^{15}N-labeled holo ACP_I (red). The assignments of the amino acids undergoing the strongest chemical shift perturbations are indicated. b) A plot of chemical shift perturbations of ACP_I holo compared to ACP_I apo. The secondary structure elements are indicated.

The analysis of the NMR spectra show, that the apo to holo conversion is accompanied by strong chemical shift perturbations (CSP). The amino acids around the active site S1989 show two different peaks for the apo and holo version such as D1988, S1989, V1991, G1992, V1993 which are located near to the active site S1989 and helix II and L2012, Y2013, Y2015 which are located on helix III (all >0.05ppm CSP). I1990 shows a peak in the holo form which is not present in the apo form (**Fig. 62)**. In order to prove if the chemical shift perturbations are due to a conformational exchange we recorded ^{15}N-NOESY-HSQC for apo and holo ACP_I. The comparison of the strip blots show that the NOEs for the apo and the holo forms are indistinguishable (**Fig. 63**). Therefore we can conclude that the chemical shift perturbations are due to transient interactions with the 4'-Ppant arm and we can exclude major conformational changes upon loading. Similar observations have been made in other systems as well [34-36, 189, 190]) (Compare **4.1.4**).

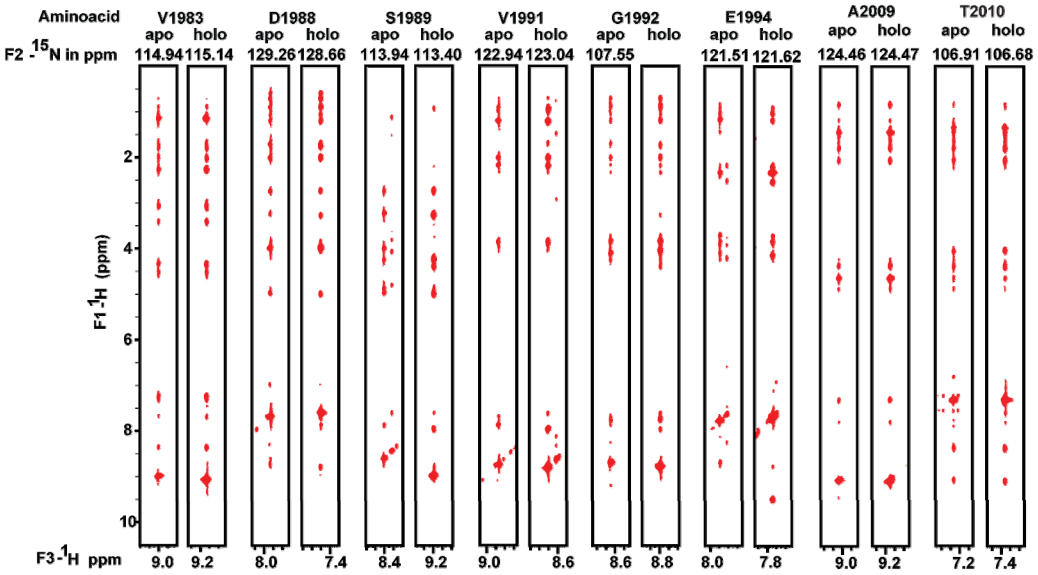


Fig. 63: Comparison of NOEs for apo and holo ACP_I. Observed NOE-pattern in a ^{15}N-separated NOESY-HSQC for the amino acids which undergo chemical shift perturbations (> 0.035ppm) for the apo and holo form of ACP_I.

The fact that we can exclude structural rearrangements, allows us to map the amino acids undergoing CSPs for the apo to holo conversion on the surface of the holo structure (**Fig. 65**). The amino acids which are involved in the transient interaction are located on the first loop towards the N-terminus of helix II (V1983), around the active site S1989 and helix II and helix III. The data show that the 4'-Ppant arm is transiently located in a hydrophobic pocket located between helix II and III (**Fig. 64 a) b) c)**). We were interested in the fact if different substrates have different impacts on the chemical shift perturbations (CSP). Therefore we calculated the chemical shift perturbations for acetyl, HMG, and Cl-HMG versus the holo form. In case of a strong interaction between the protein ACP_I and the substrate, CSP should be significant for the substrate and not the 4'-Ppant arm (**Fig. 64**). In all three cases we observe small CSPs, all affecting the same amino acids. V1983, D1988, I1990, A2009 are involved in all three cases. Interestingly these amino acids have smaller CSP for acetyl, higher CSP for HMG and the strongest CSP for Cl-HMG. The differences in CSP might be linked to the longer chain length and in case of Cl-HMG to the chlorid-Ion (**Fig. 64**). This data are in accordance with Evans and his colleagues who observed for acetyl, malonyl, 3-oxobutyl and 3,5-dioxohexyl ACP (PKS II, act ACP from *Streptomyces coelicolor)* only minimal or no interactions with the ACP giving small CSP [39]. But for a growing acyl chains such as butyryl-, hexanoyl-, and octanoyl-ACP they observed that in each case the acyl chain is clearly bound in a cleft which is also located between helices II and III of the PKS ACP which was also apparent due to stronger CSP [39].

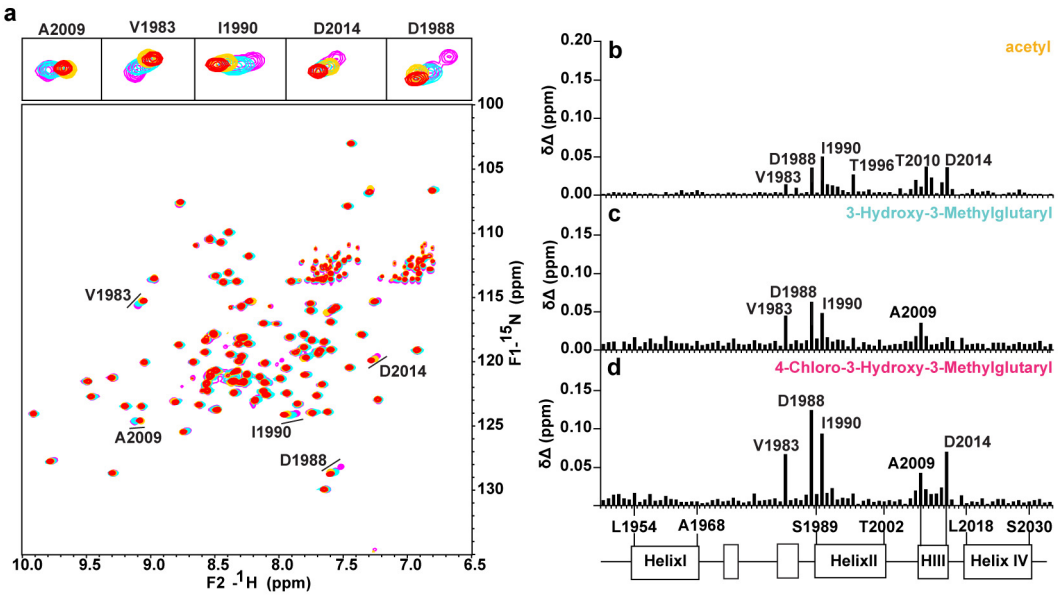


Fig. 64: Comparison of the different chemical states of ACP_I. a) An overlay of [^{15}N, ^{1}H] TROSY spectra of holo- ACP_I (red) and acetyl-ACP_I (yellow), HMG-ACP_I (light blue), Cl-HMG-ACP_I (magenta). Assignments of the amino acids undergoing the strongest chemical shift perturbations are indicated and their magnifications are presented on the top panel. The right hand panel presents plots of chemical shift differences between holo-ACP_I and acetyl-ACP_I (b), HMG-ACP_I (c), Cl-HMG-ACP_I (d). The secondary structure elements are indicated below.

When we map the CSP for apo to holo on the structure we can see that the cleft to which the 4'-Ppant arm is bound is framed by helix II and III (**Fig. 65**). The amino acids building this pocket are highly hydrophobic. Interestingly when we change the substrate, there are no further amino acids involved in substrate interactions. This means, the 4'-Ppant arm and its substrate under investigation will not occupy more space in the hydrophobic pocket than a smaller substrate. Nevertheless for the longer substrate, such as HMG and Cl-HMG the same amino acids show slightly stronger CSP assuming that with growing chain length and chloride modification the 4'- Ppant arm with substrate presumably has longer contact times with the surface. Interestingly Asp2014 is the only amino acid which is involved in addition to the binding of acetyl as well as Cl-HMG but not with HMG.

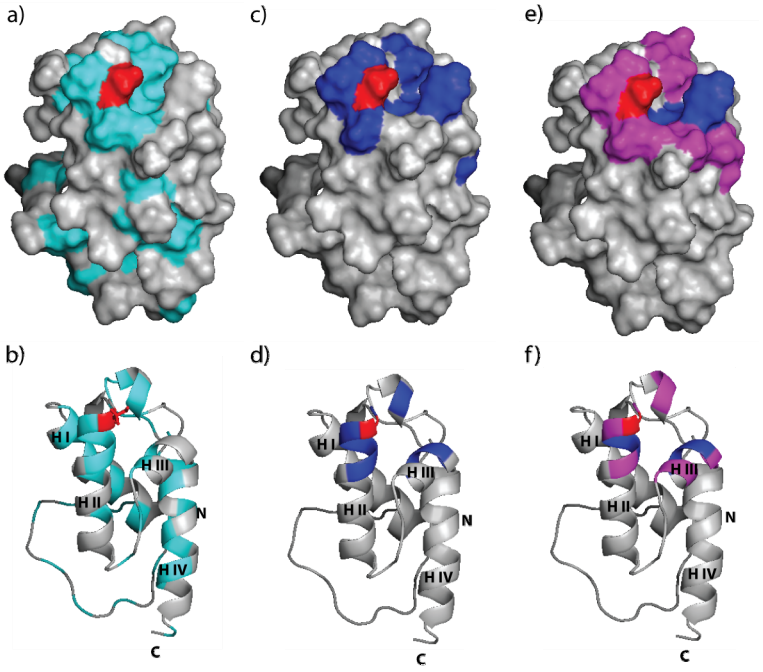


Fig. 65: Graphical presentation of ACP_I and the amino acids involved in substrate interaction a) and b) ribbon diagram and surface diagram of holo ACP_I with its hydrophobic amino acids (cyan); c) and d) The CSP for apo to holo conversion of ACP_I are mapped in dark blue (CSP > 0.05ppm), the active serine is colored in red. e) and f) The amino acids of of ACP_I which are involved in the transient interaction with acetyl, HMG and Cl-HMG are indicated in magenta (CSP >0.02ppm).

7.3.7 Localization of the cofactor with the substrate

For a detailed characterization of the interaction of ACP_I with Cur Hal the localization of the substrate on the surface has to be defined. We expressed $^{13}C/^{15}N$-labeled ACP_I and modified it *in vitro* with unlabeled CoA and (*R,S*)-HMG-CoA. NOEs between ^{12}CH groups of the attached cofactor and the ^{13}CH groups of the protein were detected in a 3D F1-$^{13}C/^{15}N$ filtered, F3-^{13}C-separated NOESY (**Fig. 66 a)**). Surprisingly, we obtained NOEs for both the holo and HMG state. In a next step we recorded F1/F2-$^{13}C/^{15}N$ double-filtered 2D-NOESY spectra to selectively detect the ^{12}CH-^{12}CH NOEs present within the 4'-Ppant arm loaded with HMG (**Fig. 66 c)**). Using the assignment of the 4'-Ppant arm loaded on TycC3 PCP [7] and combining it with the assignment of the 4'-Ppant chain loaded with hexanoyl in an ACP bound state from a mammalian FAS ACP [190]. This combination enabled us to finally assign the 4'-Ppant chain. The 1D spectra of CoA and HMG-CoA then gave rise to the unambiguous assignment of the protons of the HMG methyl group (66,67,68), which was used to assign these methyl groups in the ACP-bound form.

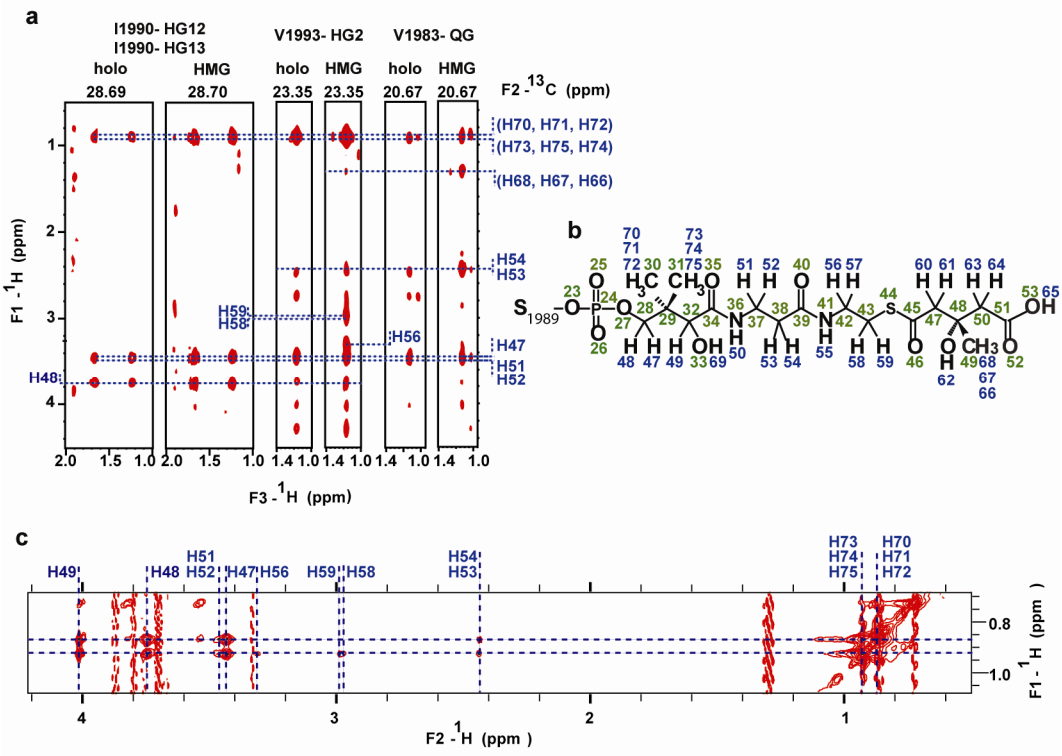


Fig. 66: Isotope filtered NOESY experiments for the structure determination of HMG-ACP$_I$ a) F1-F3 strips from a 3D F1-^{13}C/^{15}N filtered, F3-^{13}C-separated NOESY-HSQC of holo and HMG-ACP$_I$ filtering the ^{13}CH-^{15}NH NOEs. The protein was loaded *in vitro* and the substrate is therefore unlabeled giving strong NOEs from ^{12}CH- to the ^{13}CH groups of the protein surface. b) the cofactor *S*-HMG and its numeration; c) Expansion of the two-dimensional *F2*-filtered NOESY spectrum of ^{13}C,^{15}N-labeled HMG-ACP$_I$. The 4'-Ppant chain and the HMG group are unlabeled and give rise to strong NOE connectivities within the 4'-Ppant arm.

.

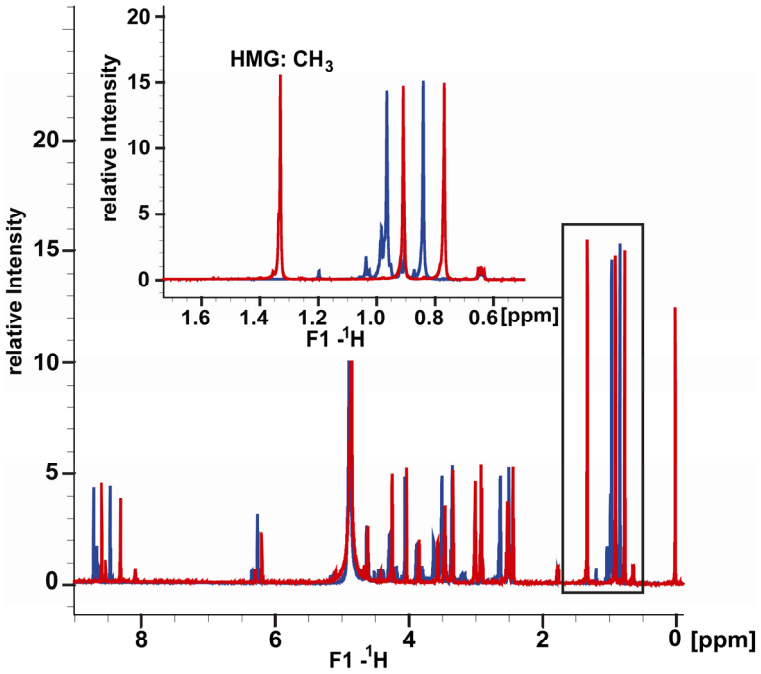


Fig. 67: Comparison of 1D spectra of HMG-CoA (red) and CoA (blue). The protons of the HMG-CoA methyl group (HMG:CH_3) was unambiguously assigned with a chemical shift of 1,35 ppm in the free form.

To calculate the structure of HMG-ACP_I we combined the restraints obtained for holo-ACP_I with the NOE-based distance restraints obtained for HMG, assuming that no major conformational changes occur between the different states. Structural statistics are given in Table **App. 9** . X-ray studies of the *E. coli* FAS II ACP demonstrated that starting from C4 chain length the substrate and the 4'-Ppant arm is sequestered into a hydrophobic pocket [37, 38]. These results were similar to the NMR findings for the PKS actinorhodin system of *Streptomyces Coelicor* [39]. For ACP_I from *Lyngbya majuscula* HMG which has a chain length of C5 is clearly not bound to a hydrophobic cleft. Instead it seems to switch between solvent exposure and surface attachment (**Fig. 68**).

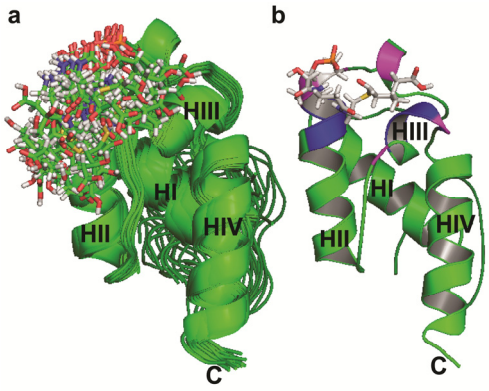


Fig. 68: Presentation of HMG- ACP_I a) Presentation of the ensemble of structures of (S)-HMG-ACP_I in a ribbon diagram. The 4'-Ppant arm with hydroxy-methyl-glutaryl (HMG) from all 20 structures is shown as well. b) Presentation of the complete mean structure of HMG-ACP_I. The amino acids which undergo chemical shift perturbations once the 4'-Ppant arm is added are labeled in blue, amino acids which undergo slight chemical shift perturbations when HMG is added to the 4'-Ppant arm are indicated in magenta (CSP >0.04ppm). The shifts are located around the attachment site of the substrate on helix II and helix III.

7.4 Interactions with ACP_I

7.4.1 Activity assay of ECH_1

It has been shown that ECH_1 accepts CurB and ACP_{II} as a substrate deliverer [68-70]. Furthermore it has been shown that ECH_1 accepts HMG-CoA as substrate but has a 10* fold higher affinity towards the ACP bound substrate [68] [69]. The activity test was performed using HMG-ACP_I (12.627 kDa) and HMG-CoA (911 Da). ECH_1 catalyzes dehydration which leads in both cases to a loss of 18 kDa (**Fig. 69 a) -d)**). Interestingly we observed product formation for HMG-ACP_I after 3.5 hrs at 25°C but for HMG-CoA we had to incubate it with more ECH_1 (8 µM instead of 5 µm) for a longer time (24 hrs). Nevertheless a direct comparison is not possible because higher amounts of HMG-CoA were used (500 µM) than for HMG-ACP_I. For the HMG-TycB1 no reaction product was observed.

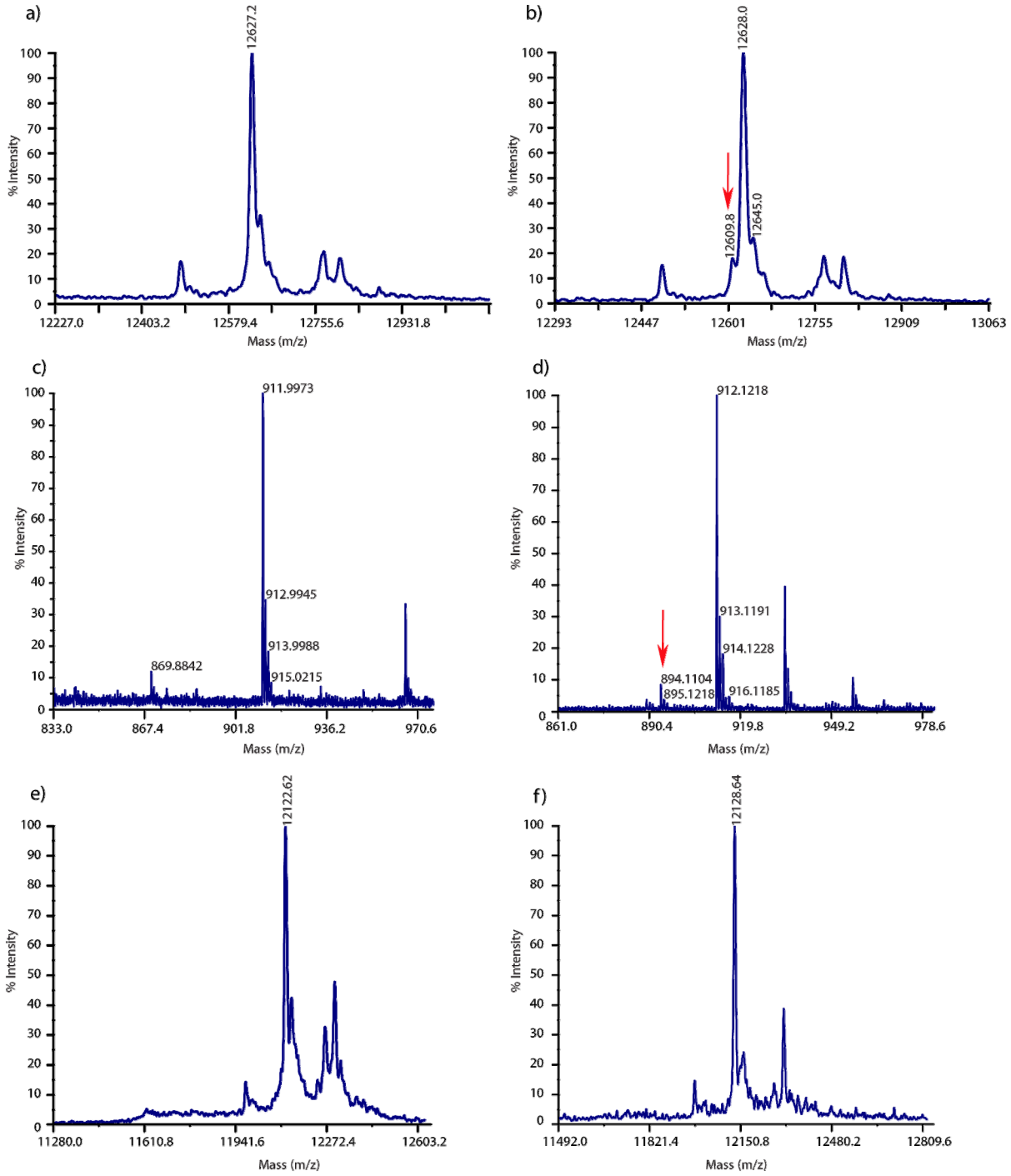


Fig. 69: Result of activity test of ECH_1. a,b) MALDI-MS of 100 µM HMG-ACP_I incubated with 5 µM ECH_1 for 3.5 hrs at 25°C. The product is indicated with a red arrow in b). c) and d) ESI-MS of 500 µM HMG-CoA incubated with 8 µM ECH_1 at 25°C for 24 hrs the product is indicated with a red arrow.e) and f) MALDI-MS of the same reaction as in a),b) with HMG-TycB1 but no product was built.

Reaction conditions for ECH_1 were optimized (buffer, time and temperature), either 24 hrs at 25°C or 37°C about 12 hrs (compare **Fig. 69**) turned out to be the best with about 10% product. Furthermore we wanted to see if we can separate the product methylglutaconyl-CoA of the ECH_1 catalyzd dehydration of the educt (*R,S*)-HMG-CoA. Therefore 500 µM HMG-CoA were incubated with ECH_1 (5 µM) at 37°C for 12hrs and then subjected to a HPLC (PerfectSil 250x4,6 C_{18}) (more details in **6.4.1**). The elution profile is presented in **Fig. 70**. Fractions of the different

peaks were analyzed via ESI-MS. ESI-MS analysis proved, that the molecular weight (MW) of the fraction corespending with peak 1 is identical with CoA, which builts up over the time due to hydrolosis. Peak 3 correspends with the MW of HMG-CoA and Peak 4 contains the product methylglutaconyl-CoA. Peak 2 was identified via ESI-MS but due to control runs it is either ECH_1 itself or a compound associated with the enzyme used.

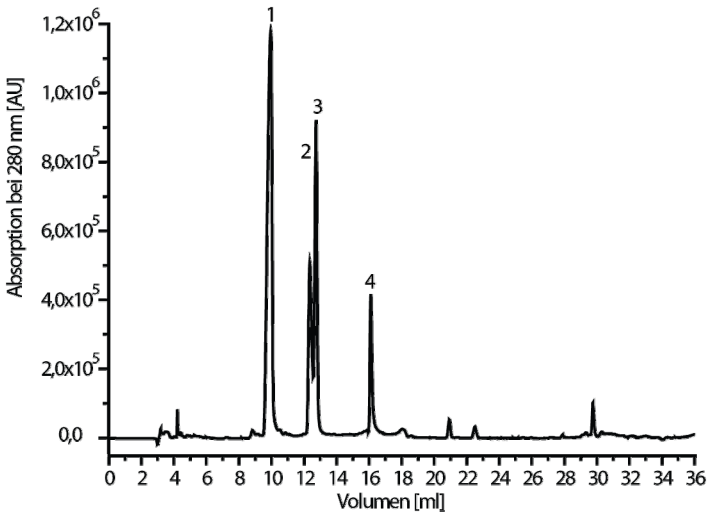


Fig. 70: Elution profile of the HPLC purification of methylglutaconyl-CoA. 500 µM HMG-CoA were incubated with ECH_1 (5 µM) at 37°C for 12 hrs and then subjected to a HPLC (PerfectSil 250x4,6 C18). ESI-MS analysis of the fractions revealed that peak 1 correspends with the molecular weight of CoA, 3= HMG-CoA, 4= methylglutaconyl-CoA

The experiment shows that the product of the ECH_1 reaction can be purified from the educts. The HPLC runs were repeated with different incubation times amd temperatures to further improve the conditions (data not shown), but the best conditions are the ones presented here (**Fig. 70**). In a next step the sample was further incubated (either in parallel or iteratively) with ECH_2. Unfortunately no further peak was obtained for the decarboxlated product methylcrotonyl (data not shown).

7.4.2 NMR Titration of ACP_I with ECH_2

For the NMR titration HMG-ACP_I, apo ACP_I and apo TycB1-PCP were titrated with ECH_2 in the ratios 1:0, 8:1, 1:1, 1:2, 1:4 und 1:8. The ACP/PCP concentration was not kept constant and chosen as high as possible in order to obtain best spectra quality. The titration experiments were performed at 20 mM Tris, 100 mM NaCl, pH= 7.5 at 500 MHz and 291 K. During the measurent protein precipitated and the ratios were therefore analyzed on SDS-PAGE after measurement (data not shown). Due to co-precipitation, the ratios up to 1:4 (ACP_I: ECH_2) did not change and was therefore analyzed (**Fig. 71**). For apo ACP_I the amino acids Asp1988, Ser 1989, Ile1990 and Gly1992 show slightly higher CSP then 0.02 ppm. For HMG-ACP_I titrated with ECH_2 we obtain slightly higher CSP than 0.02 ppm for Leu1954, Leu1969, Asp1988, Ser1989 and Gly1992 (**App. 12**). Ser1989 is the active center to which the cofactor is attached and the amino acids undergoing CSP a grouped around this active site in both cases for apo

ACP_I as well as HMG-ACP_I (**Fig. 71a)**). The PCP TycB1-PCP was not assigned at the compilation of this thesis. Hence all peaks are just numerated aribitrary from 1-84 independant of the position in the amino acids. For the apo PCP titration with ECH_2 we obtain two amino acids with CSP higher than 0.02 ppm (**Fig. 71b)**).

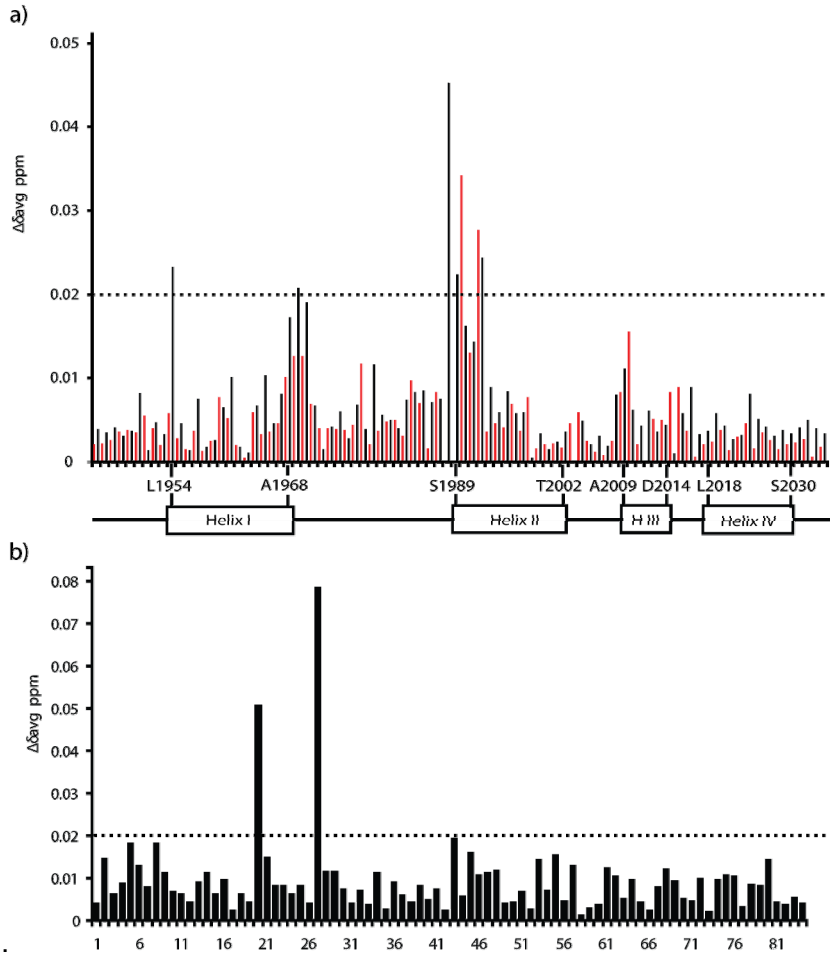


Fig. 71: A plot of chemical shift perturbations for ACP_I titration with ECH_2 a) of HMG-ACP_I with ECH_2 (in black) and apo ACP_I with ECH_2 (in red). Titrations were performed at 500 MHz. The CSP calculation is based the comparison of 1:0 spectra with 1:4 molar excess spectra. In the graphic below are indicated the secondary elements. b) plot of CSP of the apo TycB1 Titration with ECH_2. The CSP calculation is based the comparison of 1:0 spectra with 1:4 molar excess spectra recorded at a 500 machine.

7.4.3 Interaction of ACP_I and Hal

Hal activity and NMR Titrations

The curacin biosynthesis differentiates from other β-branching pathways in the introduction of a c-chlorination step on (*S*)-3-hydroxy-3-methylglutaryl (HMG) mediated by Cur Hal, a non-heme Fe(II), α-ketoglutarate-dependent enzyme [69]. Recent studies convinced that the chlorination step takes place before the dehydration catalyzed by ECH_1 and decarboxylation followed by ECH_2 leading to 4-chloro-3-methylcrotonyl-ACP. ECH_1 and ECH_2 are known to be promiscuous enzymes. ECH_1 for example uses either *S*-HMG- or Cl-(*S*)-HMG either bound to one of the triplet ACPs [69] or CurB [70] or even in a CoA form. To investigate the activity of hal we used an activity assay similar to the one described before [67, 69]. In this activity assay we loaded

ACP_I with (*R,S*)-HMG and detected substrate chlorination by MALD-MS. We showed that the halogenase which we expressed and purified was active and that it specifically interacts with ACP_I but not with CurB and not TycB1 PCP (**Fig. 72**). This specificity has been shown already previously; furthermore distinguishes hal between *R*- and *S*-HMG [67, 69] and accepts *S*-HMG, but as mentioned before, only in an ACP bound form. The ACP is necessarily a representative of the triplet ACP, located at the C-terminus of CurA. We were therefore interested in determining further molecular determants important for the specificity of the interaction. We expressed, purified and reconstituted hal anaerobe as described before [69][67] and proved by MALDI mass spectrometry its activity (**Fig. 72).** We attempted to map the specific interaction surface on ACP_I using NMR titration experiments. Therefore (*R,S*)-HMG-[^{15}N]ACP_I was titrated with active, unlabeled halogenase, but no chemical shift perturbations were observed. We repeated the same experiment with functional halogenase in non-reactive conditions: without Cl, α-KG, O_2 or missing all components. Additionally we tried to manipulate the halogenase using the unfunctional mutant R241A [67] or Fe^{3+} reconstituted, inactive halogenase. Finally we tried to use functional Halogenase but ACP_I loaded with different and no substrates such as apo, holo, acetyl and malonyl. In all cases, despite using a high excess of Hal to ACP (8:1), no chemical shift perturbations or line broadening effects were observed which would indicate any formation of an ACP/Hal complex. This result might propose that interaction might be limited to the 4'-Ppant arm or acyl chain, similar to the TE-ACP interaction observed in the DEBS PKS [191]. We performed the activity assay using ACP_I, CurB–ACP and TycB1 PCP as a substrate deliverer. Only in case of ACP_I all *S*-HMG-ACP_I (50% from *R,S*–HMG-ACP_I) was chlorinated after 10 min into Cl-*S*-HMG (**Fig. 72**). Even after 1 hr HMG-CurB and HMG-PCP did not show any chlorinated form (Data not shown). This result demonstrates clearly that Cur Hal distinguishes between the ACPs (ACP_I vs CurB), leading to the conclusion that an interface must be present, which can not be detected by NMR. In the assay the reaction is finished after 10 min using 100 µM HMG-ACP and 5 µM Hal (20:1 ratio) in NMR of course we use even higher ratios (1:8). A reason for not observing any interaction in NMR might be due to the fact that the reaction is performed to fast and the product Cl-HMG-ACP_I has too low affinity to Cur hal; too low affinity might also be the reason for all unreactive conditions/combinations tested. Presumably we have also too transient interactions to be detected by NMR.

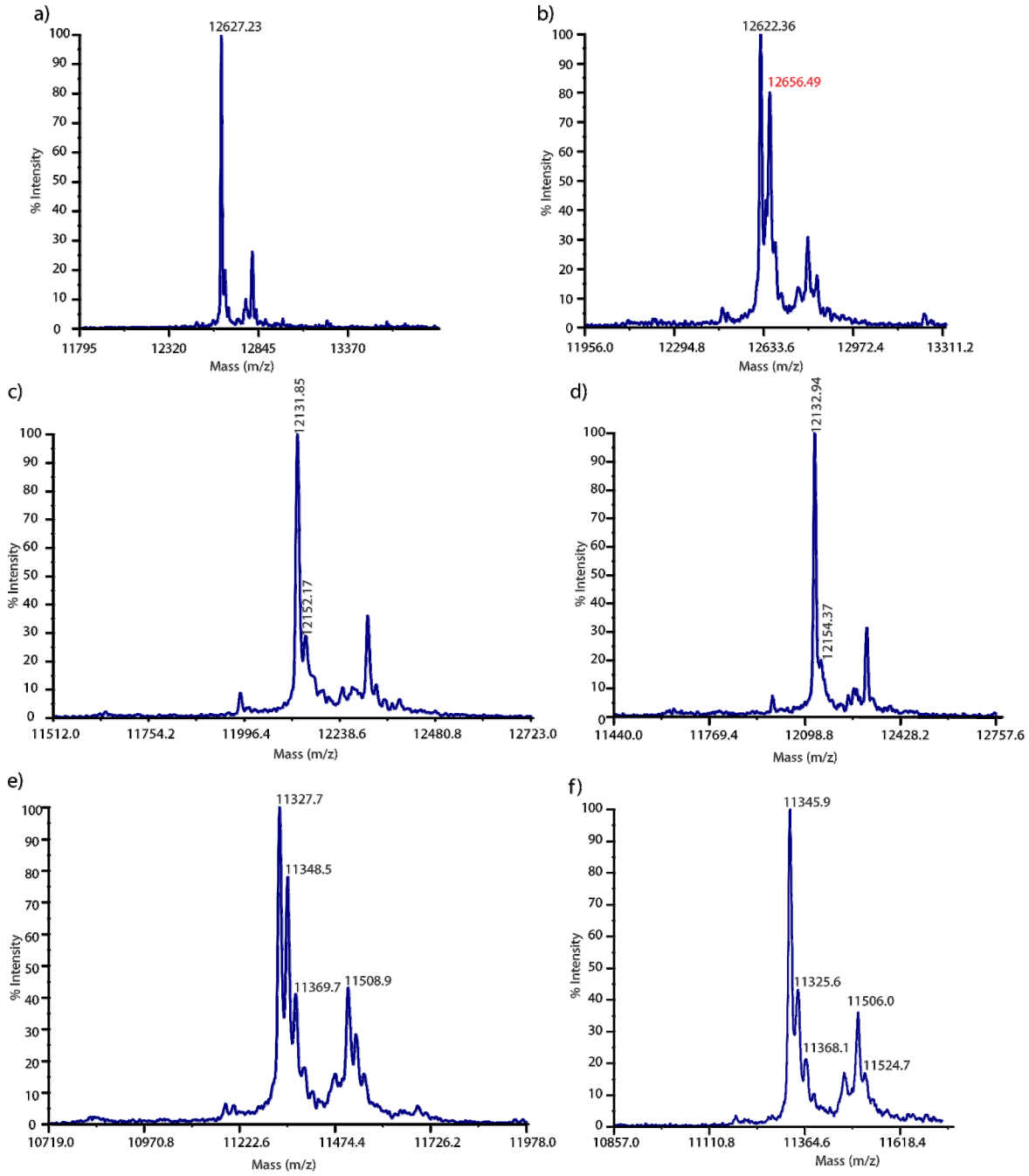


Fig. 72: MALDI-MS of reactions of HMG-ACP/PCP with Hal a) HMG-ACP$_I$ b) HMG-ACP$_I$ with Hal c) HMG-TycB1 PCP d) HMG-TycB1PCP with Hal e) HMG-CurB f) HMG-CurB with Hal

Fluorescence anisotropy of Cur Hal and ACP$_I$

Despite the fact that pH 7-7.5 is recommended for the reaction of malemide with cyteines we obtained in these conditions proteins with 2-3 fluorescin attached to it. We screened the pH (6-7.5) for one specific fluorescin labeling and obtained this in 20 mM MES, 100 mM NaCl at pH 6 (**App. 10**). Different ACP$_I$ cysteine mutants (S10C, S87C, G89C, T90C) were prepared and evaluated using CD spectroscopy. They all showed melting curves with a transition from folded to unfolded giving melting temperatures at 54°C (S87C), 50.5°C (G89C), 52.1°C (T90C) indicating less stability then the wt protein (T_M 62.8). We decidede to perform titration experiments with S10C and for a C-terminal fluorescin T90C, in both cases fluorescin should

not disturb in any interactions due to the long distance from the active site (35 Å). As a control, TycB1 PCP in its apo state was used, due to the fact that we showed via MS that the PCP loaded with HMG is not recognized as a substrate deliverer (**Fig. 73**). Fluorescence anisotropy measurements were done in the same conditions as the activity assay and each time the activity of the enzyme was proved in parallel by MS.

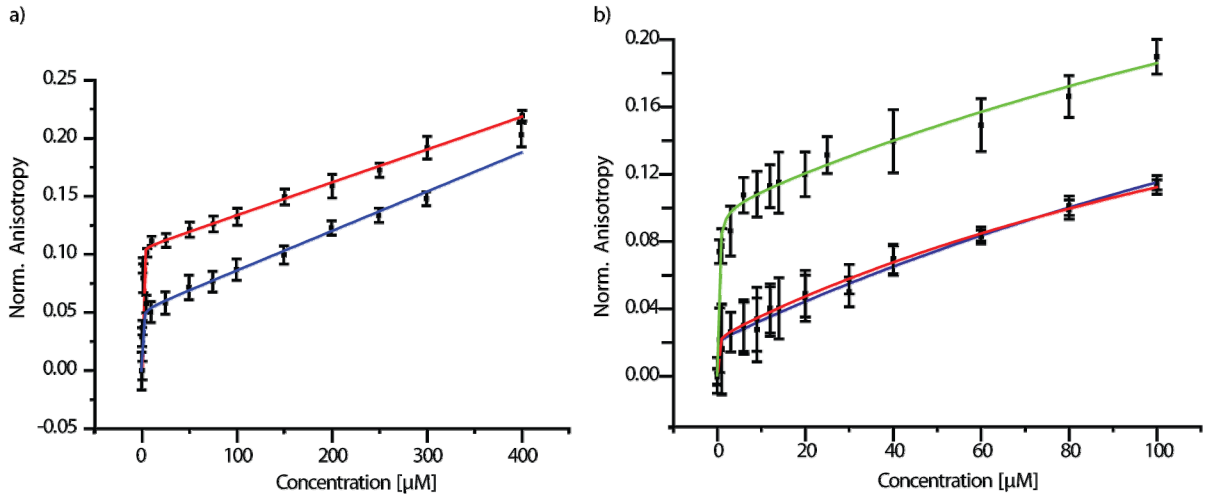


Fig. 73: Fluorescence anisotropy measurements of ACP with Cur Hal. a) Blots of the normalized anisotropy of HMG- ACP$_I$S10C (red) and apo (blue form) titrated with Hal b) Blots of the normalized anisotropy of HMG-ACP$_I$ T90C (red) and apo ACP$_I$T90C (blue) and TycB1 PCP in apo (green) form titrated with Hal. Solid lines show the fit to a two-binding-site model. Error bars show standard deviation.

The experiments were repeated twice with different batches of Cur Hal. The concentration is the monomeric concentration with regard to the fact that each monomer has its own active site. Data were fit according to a two side-binding model giving the two binding constants K_1 and K_2, all fits have a correlation coefficient (R^2) higher than 0.98. The Binding constants resulting from the fluorescence anisotropy measurement are listed below (**Table 16**).

Table 16: Summary table of the binding constants measured via fluorescence anisotropy

Sample	K_1 [µM]	Error [µM]	K_2 [µM]	Error [µM]
ACP$_I$S10C HMG	0,105	0,033	2,4 * 10^{16}	
ACP$_I$S10C apo	0,391	0,127	7,0 * 10^{16}	
ACP$_I$T90C HMG	0,11	0,073	204,154	21,76
ACP$_I$T90C apo	0,037	0,211	314,71	145,87
TycB1PCPapo	0,206	0,052	317,05	379, 18

KD values for the second binding event were 2.4 * 10^{16} , 7.0 * 10^{16}, 204.15± 21,76 µM, 314.71± 145.87 µM , 317.05± 379.18 µM respectively, suggesting that they represent unspecific binding. The first Kd is unexpected high (0.03-0.4 µM) for both the apo version, the HMG version and PCP suggesting that these values are based on hydrophobic binding of Hal to the fluorophore fluorescin instead of specific binding to the protein.

Activity assay and mutational analysis for Hal and ACP_I

To prove our hypothesis of the existence of a specific interaction surface, we went for mutational analysis. For the mutational approach we mutated solvent exposed, not conserved amino acids in HI, LI, HII, LII, HIII and HIV. Amino acids which are conserved between ACP_I and ACPs (involved in other protein-protein interactions) were taken as essential for the structural stability and therefore excluded for a mutational approach. An alignment between CurB and ACP_I highlights amino acids which might be important for a Hal-ACP interaction (**App. 11**). All mutants were expressed and purified and the activity test performed as described before. Due to using (*R,S*)-HMG-ACP_I as a substrate, activity was determined by assuming that a chlorination of 100% *S*-HMG, resulting in one peak unreacted *R*-HMG-ACP and one peak Cl-(*S*)-HMG-ACP, separated by 34 Da and of equal intensities; this would then correspond to 100% activity. This approach implies the assumption that both proteins display equal ionization capacities. The MALDI-MS raw data are presented in **App. 14** and the derived activities are listed in **Table 17**.

Table 17: Results from the activity test. [1]The activity is estimated based on the MALDI-MS results presented in **App. 14**. [2] Impurities were estimated based on the SDS gel, some mutants were less pure. Based on the mutants E1978P, D1979T, Q1980D and Q2001E, T2002D, N2004S we can see that impurities do not decrease activity. The SDS gel is presented in **App. 15**. The melting temperature was determined by CD spectroscopy. The curves for CD spectroscopy are presented in **App. 16**.

Mutation	Localization	Activity [%][1]	No of exp	Impurities [%][2]	T_M[3] [°C]
WT		85 ± 11	5	<5	64 ± 0,15
Q1959A, Q1963A, E1967A	H I	91 ± 1	2	<5	
E1973D, E1975P, A1977E	L I	Not soluble			
E1978P, D1979T, Q1980D	L I	82 ± 8	3	~15	
K1981S, V1982K, D1983K	L I	75 ± 5	3	<5	
D1988A	Active Site	29 ± 10	3	<5	57 ± 1,2
I1990A	Active Site	18 ± 7	3	<5	64 ± 0,2
V1993N, T1998M, T1999M	H II	70 ± 19	3	<5	65 ± 0,2
Q2001E, T2002D, N2004S	C- H II	92 ± 2,5	3	~30	
K2008P	L II	75 ± 7	3	<5	
A2009R	N- HIII	15 ± 12	3	<10	70 ± 2,5
T2010A	H III	37 ± 7	3	<5	63 ± 0,26
Y2013A	H III	35 ± 2,5	3	<5	61 ± 0,14
D2014A	C- HIII	59 ± 2	3	<5	58 ± 1,35
Y2015A	L IV	77	1	<10	
P2016K	L IV	90	1	<5	
L2019A, Q2027A, S2031A	H IV	37 ± 5	2	<5	

The mutations D1988A, I1990A neighboring the active site S1989 and A2009R on helix III had the strongest impact on Hal activity (activity: 29% ±10%; 18%±7%; 15%±12%, respectively). It is necessary to mention, that in these cases no product peaks were observed, which were in any way comparable with the other "active" proteins. Nevertheless, to ensure no underestimation of the activity we considered that a product peak might be covered by the first and second natrium peak (~+22 and ~+44 Da) leading to the assumption, that the peak at 34 Da away from the educt might be product. This leads in general to an overestimation rather than an underestimation of the activity. The single point mutations T2010A, Y2013A, D2014A on helix III decrease the activity to 37%±7%, 35%±2%, 59%±2%, respectively. CD curves of the less active mutants show that the mutations did not have any impact on the global fold. Whereas melting temperatures derived from CD melting curves show that some mutations lead to a less stable protein (**Table 17, App. 16**). To ensure that the mutations showing the strongest impact, are not related to structural perturbations we compared therefore [^{1}H,^{15}N]-TROSYs of the ^{15}N labeled mutants with the WT (**Fig. 74**).

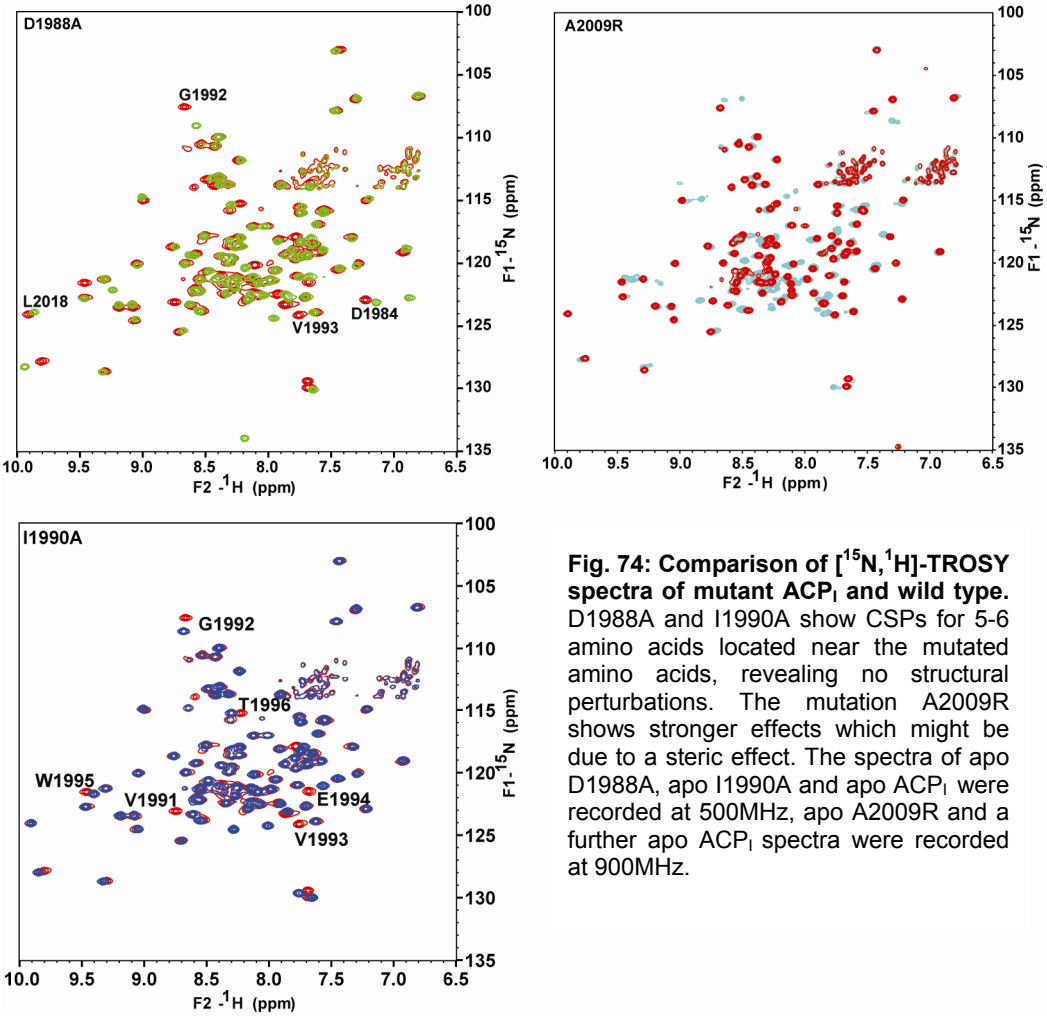


Fig. 74: Comparison of [^{15}N,^{1}H]-TROSY spectra of mutant ACP$_{I}$ and wild type. D1988A and I1990A show CSPs for 5-6 amino acids located near the mutated amino acids, revealing no structural perturbations. The mutation A2009R shows stronger effects which might be due to a steric effect. The spectra of apo D1988A, apo I1990A and apo ACP$_{I}$ were recorded at 500MHz, apo A2009R and a further apo ACP$_{I}$ spectra were recorded at 900MHz.

Actually the comparison of [^{1}H,^{15}N]-TROSY apo mutant spectra with the apo WT spectra, reveals for D1988A and I1990A CSPs for 5-6 amino acids located near to the mutated amino acids, revealing that the mutation does not effect the structure. The mutation A2009 leads to stronger CSP which might be due to sterical reasons. We were interested in the question if these mutations might lead to a displacement of the cofactor which then in return effects the Hal/ACP$_{I}$ interaction. For clarification we loaded the mutants D1988A, I1990A and A2009R with the 4'-Ppant arm, acetyl and HMG, mapped the CSP against the sequence and compared these results with the WT (**Fig. 75**).

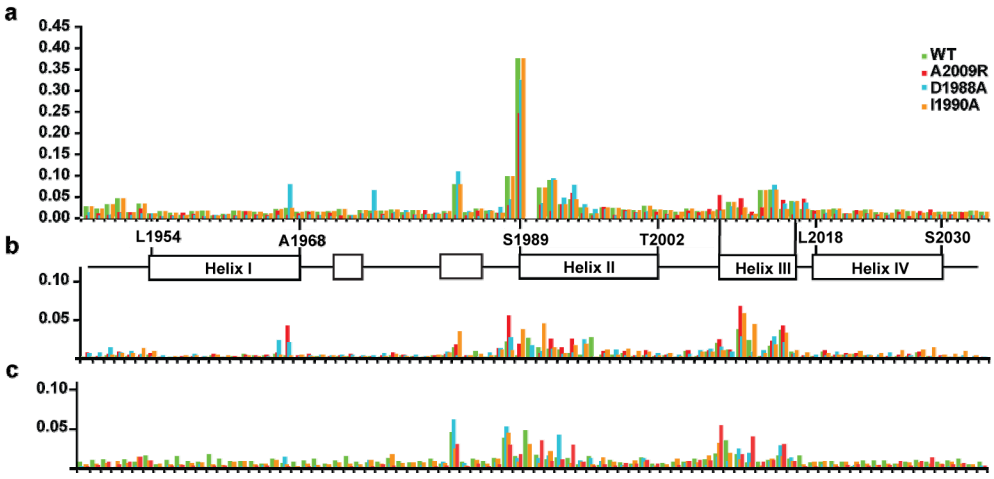


Fig. 75: Comparison of the different chemical states of ACP$_{I}$ WT (green), A2009R (red), D1988A (light blue) and I1990A (yellow). a) CSP for the apo to holo conversion b) loaded with acetyl and c) HMG. In all cases the same regions such as helix II and III are affected.

Interestingly, upon loading, the same amino acids undergo CSPs (around the active site Ser1989 and helix III), even for A2009R, supposing that the localization of the cofactor is not affected (**Fig. 75**). In the following we mapped the mutational effect according to the strength of activity decrease. Mutations which decreased the activity down to 0 to 30% are marked in red, this is the case for (D1988A and I1990A amino acids neighboring the active site Ser and A2009R located on helix III), mutations which decreased the activity to 30-70% were marked in magenta (e.g. the multiple mutant V1993N with T1998M and T1999M located on helix II, the single mutants T2010A, Y2013, D2014A all located on helix III) and mutants with minor to no effect and activities >70% were marked in yellow (**Fig. 76**). The mutations which affect the activity surround the substrate, building a nice, closed interface, which turned out to be important for Cur Hal recognition and selectivity.

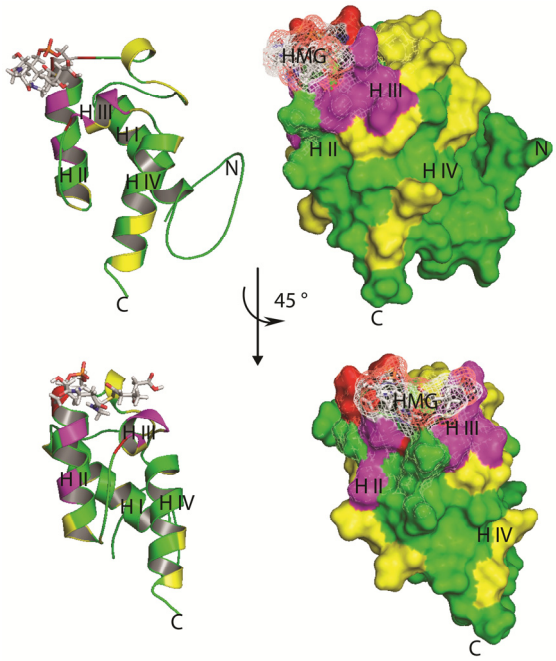


Fig. 76: Effect of ACP_I mutation on activity. According to the strength of activity decrease we applied a color code. Mutations which decreased the activity to 0-30% are marked in red this is the case for (D1988A and I1990A amino acids neighboring the active site Ser and A2009R located on helix III), mutations which decreased the activity to 30-70% were marked in magenta (e.g. the multiple mutant V1993N with T1998M and T1999M located on helix II, the single mutants T2010A, Y2013 D2014A all located on helix III) and minor to no effect with an activity of over 70% are marked in yellow.

8 Discussion

8.1 Optimization of constructs

The biosynthetic origin of bacterial polyketides and non-ribosomal peptides has been studied extensively in the passing years. These biochemical studies unravelled the biochemical rules leading to the synthesis of complex polyketides and polypeptides. These biochemical rules dictate for example the polyketide starter unit selection or amino acid selection, chain length control, and post-PKS processing. Nevertheless the manipulation of these megasynthethases in their native organism or related hosts is still problematic. In the past years several X-ray and NMR structures of essential NRPS and PKS domains have been solved [7, 10, 40, 192]. Several bi-domain structures have been solved as well in recent years, such as the X-ray structures of the DEBS AT-KS didomains [21, 29], the X-ray structure of the SrfA-C termination modul [30] and the X-ray structure of the PCP-C domain [193]. These structures gave a good idea about how the domains are organized, but still, little structural informations are available on how domains interact and communicate. This may be because some of the crystallized molecules may not represent the active domain orientations. NMR as a structural technique has the advantage to be able to detect the proteins in there soluble, active states. The drawback is that in solution, the large size and internal dynamics of multidomain fragments (> 35 kDa) make structure determination by nuclear magnetic resonance a challenge and require advanced technologies. The advances in NMR in recent years allowed working on increasing systems. Therefore the 37 kDa solution structure of the NRPS didomain PCP-TE from enterobactin synthase EntF gives insights into the interaction surface of those domains [194].

One major bottleneck for structural investigations of PKS and NRPS systems is the high amount of protein needed. Therefore, a first major step in a structural project is the evaluation of putative protein targets for structural studies, meaning essentially the assessment of protein quality and in most cases the improvement or enhancement of the protein solubility. We faced this bottleneck during our work with the targets KS and AT (CurA), HCS (CurD) and KS (CurC) which displayed low solubility (**7.1.1, 7.1.2**). Solubility problems can be distinguished into two major fields, proteins that express insoluble and accumulat in so-called inclusion bodies (e.g HCS and KS) and proteins which can be expressed solubly but which either precipitate or build soluble aggregates after purification, the latter was encountered during the work with AT and KS from CurA (**7.1.2**).

Several approaches are known to increase protein solubility, for example a) to reduce the rate of protein synthesis. This can be achieved by decreasing for example the temperature or the usage of a weaker promoter. b) To change the growth media by adding prosthetic groups, or glucose, or ethanol or polyols; c) To co-express chaperons or foldases e.g. by using Arctic Express cells; d) Using specific host strains with more oxidizing cytoplasmic environment such

as AD494 and Origami; e) Addition of fusion partners; f) *In vitro* denaturation and refolding. Concerning the addition of a solubility tag, numerous solubility tags are reported in the literature, such as MBP (Maltose-binding protein), GST (Glutathione-S-transferase) [195], Trx (Thioredoxin), NusA (N-Utilization substance), SUMO (Small ubiquitin-modifier), SET (Solubility-enhancing tag Synthetic), DsbC (Disulfide bond C), Skp (Seventeen kilodalton protein), T7PK (Phage T7 protein kinase), GB1 (Protein G B1 domain)) [196], but the majority of recent work has continued to focus on a few major players, notably the maltose-binding protein (MBP), the N-utilization substance A (NusA), thioredoxin (Trx) and glutathione-S-transferase (GST) [196]. GST, however, has turned out to be a bad solubility enhacer [197]. NusA and Trx have shown to be comparable to MBP in its capability of producing soluble proteins but without independent affinity functionality in case of NusA [196]. We decided in all cases where we encountered solubility problems to use MBP and Ub as a solubility tag. For AT and KS from CurA the N-terminal MBP-tag increased the expression and helped to gain soluble protein. For HCS and KS both MBP and Ub N-terminally tagged constructs turned out to be insoluble in the conditions tested (low temperature, coexpression with chaperons). Due to the reasons mentioned before replacing MBP by other solubility tags might not increase the solubility of HCS and KS. Instead of trying other solubility tags, it might be better to check the influence of the tag at the C-terminus. Further modification of the main construct by mutations and/or deletions of hydrophobic patches or unstructured regions might be even more promising in improving solubility. A drawback for this approach is, that mutations might have a negative impact on the interactions of these proteins. We did not try to modify the media (XYT) with prosthetic groups or additional glucose which might be worth trying. The expression of HCS (CurD) and KS (CurC) under a weaker promotor might also be worth a try (e.g trc instead of T7). Condsidering the biosynthesis of HCS and KS we assume that these proteins might build a complex, presumably with CurB as a third interplayer. An alternative approach to achieve the solubility of these proteins might be the co-expression of HCS and KS or HCS, KS and CurB. To achieve this, HCS and KS were cloned for example into the pstduet vector (Novagen). Such a construct would further surve for a co-expression with CurB. One main point to respect for co-expression is, that the vectors used, need two different Oris and antibiotic resistances and for more flexibility probably even different inducers. In case none of these options would help to express soluble protein refolding might be used. ACPs for example can be easily purified in denaturing conditions and followed by refolding without the 3D structure to be affected (Data not shown). We tried this approach in one shot by expressing Ub-HCS and Ub-KS in inclusion bodies. Then we separated soluble proteins from insoluble proteins and went directly for solubilization in urea followed by His_6-tag purification in denaturing conditions. After purification the protein was refolded. The purity of the proteins was quite low (~30% impurities) (**Fig. 28**). For future work purity would be easy to improve by washing the inclusion bodies with buffers containing

increasing amounts of urea or guanidinimhydrochlorid. So far, refolding worked for several challenging proteins to gain enough material for further studies [198-201]. It is important to mention that the smaller the protein, the easier is the refolding. HCS and KS are both quite complex domains with 45 kDa in size, this makes in fact refolding attempts challenging. In addition the proteins have 10 cysteines in case of HCS and 8 cysteines in case of KS which sets a further challenge to any refolding attempt.

As mentioned before AT and KS constructs from CurA was expressed solubly as MBP fusion proteins (**Table 10**), expect the construct L-AT-L (amino acids 974-1484 according to accession code entry AAT70096). But surprisingly in all cases the proteins showed soluble aggregation and no monomer peak on a gelfiltration column. To reduce the soluble aggregation of AT and KS constructs we tried altering the expression conditions by lowering the temperature, co-expressed the proteins with chaperones and added additives during protein isolation such as detergents (CHAPS), sugar (sucrose), arginine or glutamate. No one method has proven successful. The AT-L construct gives a surprising contradiction. AT-L was expressed and purified as MBP fusion protein but as well as a His_6-tag fusion protein. Surprisingly AT-L fused N-terminally to MBP shows on the superose 6 gelfiltration column soluble aggregation and elutes in the void volume (**App. 6b)**). Instead, AT-L with a His_6-tag was expressed -in little amounts- solubly and shows impoved behaviour versus the MBP fusion on a gelfiltration column by presenting a population of monomer to dimer, trimer and undefined multimers (**App. 6c)**). This result led to the hypothesis that MBP might promote in some cases soluble aggregation- a hypothesis which has not been checked. Furthermore, it might be of interest to see how the proteins behave without MBP. Due to the fact that all AT and KS constructs only express as MBP fusions an *in vivo* cleavage prior to purification might be reasonable. For this purpose, TEV protease would be co-expressed along with the fusion protein but in a time delayed manner as described by Kapust and Waugh [202]. Briefly, IPTG would be added to the log phase culture and the fusion protein would be allowed to accumulate for a period of time. Then, TEV protease expression (encoded on vector prk603) would be stimulated by adding anhydrotetracycline to the culture. For this purposes a special *E. coli* strain (e.g., DH5alphaPRO or BL21PRO cells from Clontech) producing the Tet repressor would be required.

To summerize, for structural studies and interaction studies of the triplet ACPs we expressed and purified the triplet ACP as well as individual domains in high amounts and purity. For biochemical assays we could prepare ECH_1, ECH_2 as well as Cur Hal. For ECH_2 and Cur Hal conditions were found in which NMR titration and fluorescence anisotropy were performed to investigate the interaction with ACP_I.

8.2 Segmental labeling

We demonstrated a novel scheme for three fragment ligation by using only one robust split DnaE intein without any refolding steps. The use of only well characterized *Npu*DnaE intein for three-fragment ligation was of practical advantage for obtaining higher yields. From 2 L *in vivo* reaction giving H_6-Smt3-*Npu*IntC15-ACP_{II}-ACP_{III} and 1 L H_6-Smt3-ACP_I-NpuIntNΔC15 we obtained 1.7 mg final ACP_I-[^{15}N]ACP_{II}-ACP_{III} (**7.2.7**). Following the same protocol we produced ACP_I-ACP_{II}-[^{15}N]ACP_{III} and [^{15}N]ACP_I-ACP_{II}-ACP_{III}. The three fragment ligation was the only way to distinguish the different domains in ACP_I-ACP_{II}-ACP_{III} (**Fig. 57**), as a consequence allowing us to proceed in investigating the structural effect of the ACP repetion. Three fragment ligation is an excellent tool to investigate inter-domain NOEs for example by applying 3D ^{13}C-edited/^{15}N-separated NOESY-HSQC experiments to a [^{15}N]ACP_I-[^{13}C]ACP_{II}-ACP_{III} protein sample. In our case no inter-domain NOEs were observed. Of course the absence of NOEs itself do not exclude interactions between the domains but NMR titration experiments of ACP_I with ACP_{II}-ACP_{II} did also not confirm any interaction between the domains (**7.3**). This taken together with the comparison of the chemical shifts of ACP_I alone and in the triplet proves that the domains behave independent of each other (**compare 7.3**). In fact the inter-domain linkers are 30 aas long and unstructured as presented in **Fig. 60**, even if the linker would be α-helical this would still make a interdomain distance of 48 Å. The NOE effect is sensitive to distances <5 Å.

The investigation of apoE3 is an example in which two fragment segmental labeling brought the break through for structure determination of this 299 aas long multi-domain protein. The full length backbone assignment was only possible due to segmental labeling [203, 204]. However for apoE3 NOE assignment was impossible due to severe peak overlap, only the application of segmental labeling, by preparing a sample in which the NT-domain was ^{2}H(99%)/^{15}N-labeled, whereas the CT-domain was only ^{15}N-labeled or ^{13}C/^{15}N-labeled allowed to distinguish NOEs by applying 3-D ^{15}N-edited NOESY experiments [205]. Another example for applying expressed protein ligation (EPL) to obtain interdomain NOEs is the study of the 40 kDa hHR23a protein with the purpose of studying the interaction between the N-terminal UBL domain and the C-terminal UBA and XPC domains [206, 207]. There, similar to us no inter-domain NOEs were detected. But different to us, chemical shift perturbation analysis of individually expressed domains were performed and confirmed an interaction of the UBL domain with the UBA domain [206]. The authors suggested that its structure is not rigidly locked into one conformation, which would explain the lack of inter-domain NOEs.

These examples show how powerful segmental labeling is especially for obtaining NOEs for multi-domain proteins for structure calculations, like in case of apoE3. The herein developed three fragment ligation is a tool to investigate the domain-domain interaction and the structure of even more complex proteins.

In the studies of the triplet ACP ($ACP_{I,II,III}$) we can exclude short range interactions between the domains, nevertheless based on our data, we can not exclude that there might be long range interactions between the ACPs. To exclude this, paramagnetic relaxation enhancement (PRE) experiments were performed which allow to measure distances up to 15 Å [208, 209]. Unfortunately an evaluation of these PRE effects was for ACP_I-ACP_{II}-ACP_{III} not possible (data not shown). The high overlap of the peaks made it impossible to evaluate the decrease of intensity. To overcome this problem it might be interesting to attach a spin label to a segmentally labeled sample. In theory, this approach would solve the overlay problem. One drawback is that segmentally labeled proteins contain one cysteine; this is present in the flexible ligation site. Preferentially the spin label should be attached in rigid regions. To overcome this, one idea was to attach the spin label to the cofactor CoA and to load it then on one domain prior to the *in vitro* step of segmental labeling protocol. The drawback is, that the *in vitro* step was usually performed for 13 hours in reducing environment which has a strong impact on the spin label. It has been shown, for example, that nitroxide spin labels are reduced in reducing conditions, 1 mM DTT for example reduces spin labels after 3 hrs to 70% and after 48 hrs to 6% of its original activity, TCEP in contrast is less aggressive, 1 mM TCEP reduces spin labels after 3 hrs to 86% and after 48 hrs to 53% of its original activity [210]. To sum up, results from PRE experiments would be of interest but these experiments were not interpretable for ACP_I-ACP_{II}-ACP_{III} in the way performed.

Our actual three fragment ligation protocol consists of an *in vivo* step and an *in vitro* step. A raising question was whether it is possible to further minimize the segmental labeling protocol by performing three fragment ligation in “one pot” (**7.2.5**). In theorie the “one pot” approach could be performed *in vivo* as well as *in vitro*. Both have pros and cons. As for the *in vitro* “one pot” approach two ligation steps required for three-fragment ligation can be performed in “one pot” by two orthogonal inteins, which could be convenient. However, considering the increased number of components that will be present in the “one pot” reaction, the purification steps by the “one pot” approach might be more labour-intensive as at least two spliced inteins, possible residual precursors, and by- products such as cleaved products have to be removed [140, 141]. In contrast, our stepwise approach with the combination of *in vivo and in vitro* ligation could simplify the purification step particularly when the ligation efficiency is very high.

As for *in vivo* approach, a splicing system that would allow NMR spectroscopists to carry out three way ligations within *E. coli* in a “one pot” reaction would indeed provide advantages – but also disadvantages. Such a one pot system could in principle be constructed in two different ways: The first one uses three different plasmids with three different promoters and three different antibiotics (as presented in **7.2.5**). In this system one would have to change from an unlabeled to a labeled and back to an unlabeled media, thus extending the expression time by additional ~4 hrs. This long expression time combined with the different antibiotics and further

stress from washing steps could easily reduce the overall yield of protein – if the bacteria would be able to survive and produce protein at all. For our “one pot” trial we co-expressed all proteins at the same time, only two of three proteins could be expressed (**Fig. 52**). This reveals that indeed three plasmids with three different antibiotics means at least in our case too much stress for the *E. coli* cells (**7.2.5**). To prevent the cells from this stress in a alternative approach one would place the DNA for the two unlabeled domains on one plasmid and thus reduce the number of plasmids, orthogonal promoters and antibiotics down to two. In this way the expression time will be reduced back to 8-9 hrs with only one media exchange and washing step as we used it in our current protocol. One major problem, however, would be the very high concentration of expressed proteins that would have to be stored before the splicing reaction could occur during the last phase of the expression protocol. This requirement could easily lead to intracellular aggregation, in particular for proteins that are not very soluble. We had to use additional solubility enhancement tags already in our two component *in vivo* ligation system with the ACPs (**7.2.4**). Finally, to obtain inter-domain contact information, NMR spectroscopists use NOE measurements between a ^{15}N- and a ^{13}C-labeled domain. In our ACP_I-ACP_{II}-ACP_{III} system we have produced for this purpose a sample with the ACP_I domain ^{15}N labeled and the ACP_{II} domain ^{13}C labeled while the ACP_{III} domain was unlabeled as initially mentioned (**7.3.1**). Such a labeling scheme could again only be realized with the three plasmid system and by changing the media twice. Overall, we believe, while conceptually and for some “easy” model systems a “one pot” reaction would be nice, the protocol that is presented here is more practical, more robust and also more flexible.

Taken together we demonstrated a novel scheme for three fragment ligation by using only one robust split DnaE intein without any refolding steps. The use of only well characterized *Npu*DnaE intein for three-fragment ligation was of practical advantage for obtaining higher yields. This new strategy can be generally applicable to segmental isotopic labeling of an arbitrary region in many other multidomain proteins for NMR spectroscopic studies. This simple and robust scheme for a central-fragment labeling could become a valuable tool for elucidating structure–function relationships of a domain or a region in intact proteins by NMR spectroscopy.

8.3 The ACP domains

Taken together we show in **7.3.1** that the linker region is unstructured and that the ACP domains behave independent from each other in ACP_I-ACP_{II}-ACP_{III} as well as in the dimerized form of ACP_I-ACP_{II}-ACP_{III} -C_d. We can conclude that the structure of one representative ACP domain correspends with the structure of the other ACP domains, isolated as well as in the full length construct. The interaction between Cur Hal and ACP domains can be therefore as well studied using one representing ACP domain of the triplet, for example ACP_I.

Analysis of the NMR spectra showed that the apo to holo conversion of ACP_I is accompanied by CSPs. The amino acids surrounding the active site Ser1989 show the

strongest perturbations between the apo and holo forms, including amino acids at the N-terminus of helix II and on helix III (all >0.05 ppm CSP) (**Fig. 62**). In addition ^{15}N-NOESY-HSQCs for apo and holo ACP_I were measured, however, comparison of the NOE pattern of amino acids undergoing strong chemical shift perturbations revealed no differences between both states excluding major conformational changes upon formation of the holo state. This is a further example- in addition to the FAS ACP of *Bacillus subtillis* [34], FAS ACP of *Escherichia Coli* [35] and the ACP of Type II FAS *Mycobacterium tuberculosis* [36]- of an ACP which contradicts the conformational switching model which has been proposed for enzymatic regulation for PCPs [7], the act ACP [33], the ACP from frenolicin [41] or the ACP from *Plasmodium falciparum* (PfACP) [43].

To calculate the structure of HMG-ACP_I we combined the restraints obtained for holo-ACP_I with the NOE-based distance restraints obtained for HMG, assuming that no major conformational changes occur between the different states due to the negligeable CSP between both states (**Fig. 64**). Structural statistics for holo ACP_I and HMG-ACP_I are given in **App. 9**. X-ray studies of the *E. coli* FAS II ACP demonstrated that starting from C4 chain length the substrate and the 4'-Ppant arm is sequestered into a hydrophobic pocket [38]. These results were similar to the NMR findings for the PKS actinorhodin system of *Streptomyces Coelicor* [189]. For ACP_I from *Lyngbya majuscula* HMG which has a chain length of C5 is clearly not bound to a hydrophobic cleft. Instead it seems to switch between solvent exposure and surface attachment (**Fig. 68**).

8.4 Interactions with ACP_I

The curacin biosynthesis differs from other β-branching pathways in the introduction of a γ-chlorination step on (*S*)-3-hydroxy-3-methylglutaryl mediated by Cur Hal, a non-heme Fe(II), α-ketoglutarate-dependent enzyme [69]. Recent studies showed that the chlorination step takes place before the dehydration catalyzed by ECH_1 and decarboxylation followed by ECH_2 [69]. ECH_1 and ECH_2 are known to be promiscuous enzymes. ECH_1 for example uses either *S*-HMG- or γ-Cl-(*S*)-HMG bound to one of the triplet ACPs or CurB or even in a CoA form [68, 70]. Interestingly ECH_1 accepts CoA substrate but for example does not accept the substrate once presented by a PCP (as presented in **7.4.1**). Also if direct interactions with the ACP are not important, the complete different surface charges of ACPs and PCPs most likely disrupt the interaction with ECH_1. ECH_1 was not soluble enough to further investigate these interactions using other biophysical methods, such as ITC, NMR or fluorescence anisotropy. Presumably ECH_1 interacts with the 4'-Ppant arm as it has been the case for other enzymes that can effectively interact with multiple, noncognate ACP domains such as the DEBS TE domain [191]. In case of higher solubility of ECH_1 this hypothesis could be checked by labeling isotopically the Ppant arm covalently linked to the ACP and performing titration experiments.

For ECH_2, the results from NMR titrations exclude direct interactions with the protein surface (**7.4.2**). We detected weak interactions (CSP~0.03 ppm) surrounding the 4'-Ppant arm but this negligeable effect was also observable for apo ACP and seems therefore to be more likely an artefact. It has been proposed that ECH_1 and ECH_2 , due to the fact that they accept CoA substrates, bind to the 4'-Ppant arm, thus facilitating rapid access and pairing of enzymes and substrates [70, 76]. Unfortunately this, at least for ECH_2 could not be supported by fluorescence anisotropy measurements, there, unspecific presumably hydrophobic interactions was shown with Kd>200µM for HMG-ACP_I and apo ACP_I (**App. 13**).

Overall our data demonstrate that Cur Hal interacts with ACP_I specifically, suggesting that it will also interact with the other two ACPs of the triplet cluster. Cur Hal recognizes an interface consisting mainly of the N-terminus of helix II, the 4'-Ppant arm with its substrate, and helix III. Substrate attached to a different carrier domain (either the CurB ACP or TycB1 PCP) is not recognized by the halogenase and does not result in product formation (**Fig. 72**). The importance of presenting the substrate as part of an entire substrate-ACP complex is also demonstrated by the fact that the isolated HMG-CoA does not react with the halogenase. According to this example, we can clearly say that the carrier protein domain represents and architectural scaffold with an information-rich surface. It can be distinguished by distinct catalytic partner proteins such as Cur Hal. In contrast to this high specificity of the halogenase the other enzymes, e.g. ECH_1 are less specific and also react with the isolated substrate-CoA. One reason for this specificity of the halogenase might be the radical mechanism used in the chlorination step which itself is most likely highly reactive and less specific. The requirement of presenting the correct substrate within a specific environment might be necessary to prevent unspecific chlorination or competitive hydroxylation reaction of substrates presented by other ACPs. Some of these ACPs, for example CurB, are stand alone ACPs, which can freely diffuse and should therefore be able to interact with the halogenase. Specificity of the other enzymes of the curacin assembly chain is most likely achieved in the standard way of specific substrate–enzyme interaction, therefore not requiring a specific ACP surface.

Surprisingly, it was not possible to directly detect the interactions between ACP_I and Cur Hal by NMR spectroscopy. Weak or transient interactions are difficult to detect using the standard NMR experiments due to the low population of the complex. One technique for the investigation of such transient interactions that has recently emerged is paramagnetic relaxation enhancement (PRE) measurements. Spin labels attached to one partner result in relaxation enhancement in the other partner even when the contact time is short and the complex not stable [176, 177, 211]. Using this technique the unspecific interaction of a sequence specific transcription factor (the homeodomain of HOXD9) could be detected when sliding along the non-cognate DNA to its destination sequence [176]. In addition, it was possible to demonstrate the existence of an ensemble of transient, non-specific encounter complexes between the

amino-terminal domain of enzyme I and the phosphocarrier protein HPr [212]. We were interested in applying the same PRE experiments to our system, in order to visualize low populated states. Unfortunately no spin label was attached to Cur Hal because removing the WT cysteine (Cys19Ala) as well as the insertion of new cysteines into Cur Hal (Cys19Ser, Ile61Cys, His109Cys, Gly111Cys, Tyr274Cys, His168Cys and Ser178Cys) resulted in precipitation.

Our results clearly show, that the individual triplet ACPs do not interact with each other. The question remains what is the biological function for this triplet assembly and its advantage over a single ACP copy. Previous studies have shown that the efficiency of the entire enzyme reaction increases with the number of ACP domains. It has been proposed that this clustering of ACP-linked substrates might result in binding of adjacent enzymes to corresponding substrates through non-specific enzyme–substrate interactions [76]. The reason for the existence of this triplet ACP cluster most likely has to be investigated at the level of the entire module by kinetic and enzymatic investigations and might be related to increasing the local concentration of potential substrate rather than distinct conformational and structural effects.

9 Outlook

The manipulation of PKS and/or NRPS systems for the production of new bioactive compounds is challenging. For this engineering approach it is important to decipher the degree of substrate promiscuity for enzymes that catalyze the transformation on the growing chain and even better to find ways to influence this substrate specificity. Additionally, it is important to understand the specificities of domain-domain or protein-domain interaction that occur during the biosynthesis. Almost all substrate precursors or intermediates are presented by ACPs or PCPs during the biosynthesis. To unravel the “rules” for recognition of these ACP or PCP domains is a major goal. Research in the past tried to identify the “rules” for recognition by structure determination using NMR or X-ray. Many insights could be obtained, especially in combination with biochemical approaches also using rational site-directed mutagenesis.

In this work we showed that Cur Hal recognizes an interface consisting mainly of the N-terminus of helix II, the 4’-Ppant arm with its substrate and helix III. The ACP represents an architectural scaffold with an information-rich surface and can be distinguished by Hal from other ACPs. Former publications support these findings by identifying amino acids surrounding the active site Ser and helix II as crucial for recognition leading to the name “recognition helix” [21, 45, 48-50, 54, 55]. We showed that helix III is essential for ACP_I and Cur Hal interaction. In other systems helix III is involved in conformational switches for example in the NRPS PCP (TycB3) [7] or the ACP from frenolicin [41] and the ACP from *Plasmodium falciparum* (PfACP) [43]. It has been demonstrated that helix III is recognized during interactions for example between the aryl carrier protein ArcP (EntB) and the elongation modul EntF ([213, 214]). Taken together ACPs seem to share the same overall recognition hot spots. Furthermore not only helix II can be declared to be a “recognition helix”, but helix III as well. It is now of interest to transform the knowledge accumulated from previous structural and biochemical investigations into novel engineering approaches, for example mutating a non cognate ACP (e.g CurB) to an ACP of native-like recognition by Cur Hal.

A further aspect which requires ongoing research is the dynamics of the NRPS and PKS systems. At present, publications about ACPs and PCPs are divided into two fields, with some ACPs/PCPs displaying conformational switches between apo and holo [41, 43, 213, 214] while others do not [34-36]. Certainly, dynamics and different conformations might play an important role in the recognition process and therefore the investigation of dynamics of PCPs and ACPs should be a goal of future research. In the past ACPs/ PCPs have been isolated from the full length protein for structural studies; interactions which natively take place in *cis* were investigated in *trans.* Segmental labeling proved to be a valuable tool for investigating domain-domain interactions in the native environment. It would be of great interest for the future research to study NRPS and PKS domains in their native environment focussing not only on interaction sites but also considering the importance and role of dynamics.

10 References

1. Weissman KJ, Muller R: **Protein-protein interactions in multienzyme megasynthetases**. *Chembiochem* 2008, **9**(6):826-848.
2. Bronzwaer SL, Buchholz U, Kool JL, Monen J, Schrijnemakers P: **EARSS activities and results: update**. *Euro Surveill* 2001, **6**(1):2-5.
3. **National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2003, issued August 2003**. *American journal of infection control* 2003, **31**(8):481-498.
4. Menzella HG, Carney JR, Santi DV: **Rational design and assembly of synthetic trimodular polyketide synthases**. *Chemistry & biology* 2007, **14**(2):143-151.
5. Menzella HG, Reeves CD: **Combinatorial biosynthesis for drug development**. *Current opinion in microbiology* 2007, **10**(3):238-245.
6. Kittendorf JD, Sherman DH: **Developing tools for engineering hybrid polyketide synthetic pathways**. *Current opinion in biotechnology* 2006, **17**(6):597-605.
7. Koglin A, Mofid MR, Loehr F, Schafer B, Rogov VV, Blum MM, Mittag T, Marahiel MA, Bernhard F, Doetsch V: **Conformational switches modulate protein interactions in peptide antibiotic synthetases**. *Science (New York, NY* 2006, **312**(5771):273-276.
8. WHO: **World Health report**. 2000.
9. von Nussbaum F, Brands M, Hinzen B, Weigand S, Habich D: **Antibacterial natural products in medicinal chemistry--exodus or revival?** *Angewandte Chemie (International ed* 2006, **45**(31):5072-5129.
10. Weber T, Baumgartner R, Renner C, Marahiel MA, Holak TA: **Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases**. *Structure* 2000, **8**(4):407-418.
11. Clardy J, Fischbach MA, Walsh CT: **New antibiotics from bacterial natural products**. *Nature biotechnology* 2006, **24**(12):1541-1550.
12. Fischbach MA, Walsh CT: **Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms**. *Chemical reviews* 2006, **106**(8):3468-3496.
13. Lai JR, Koglin A, Walsh CT: **Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis**. *Biochemistry* 2006, **45**(50):14869-14879.
14. Staunton J, Wilkinson B: **Combinatorial biosynthesis of polyketides and nonribosomal peptides**. *Current opinion in chemical biology* 2001, **5**(2):159-164.
15. Hopwood DA: **Genetic Contributions to Understanding Polyketide Synthases**. *Chemical reviews* 1997, **97**(7):2465-2498.
16. Austin MB, Noel JP: **The chalcone synthase superfamily of type III polyketide synthases**. *Natural product reports* 2003, **20**(1):79-110.
17. Tseng CC, Vaillancourt FH, Bruner SD, Walsh CT: **DpgC is a metal- and cofactor-free 3,5-dihydroxyphenylacetyl-CoA 1,2-dioxygenase in the vancomycin biosynthetic pathway**. *Chemistry & biology* 2004, **11**(9):1195-1203.
18. Richter CD, Nietlispach D, Broadhurst RW, Weissman KJ: **Multienzyme docking in hybrid megasynthetases**. *Nature chemical biology* 2008, **4**(1):75-81.
19. Weissman KJ: **Introduction to polyketide biosynthesis**. *Methods in enzymology* 2009, **459**:3-16.
20. Mortison JD, Kittendorf JD, Sherman DH: **Synthesis and biochemical analysis of complex chain-elongation intermediates for interrogation of molecular specificity in the erythromycin and pikromycin polyketide synthases**. *Journal of the American Chemical Society* 2009, **131**(43):15784-15793.
21. Tang Y, Chen AY, Kim CY, Cane DE, Khosla C: **Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase**. *Chemistry & biology* 2007, **14**(8):931-943.
22. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT: **A new enzyme superfamily - the phosphopantetheinyl transferases**. *Chemistry & biology* 1996, **3**(11):923-936.

23. Staunton J, Weissman KJ: **Polyketide biosynthesis: a millennium review**. *Natural product reports* 2001, **18**(4):380-416.
24. Cane DE, Walsh CT: **The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases**. *Chemistry & biology* 1999, **6**(12):R319-325.
25. Bisang C, Long PF, Cortes J, Westcott J, Crosby J, Matharu AL, Cox RJ, Simpson TJ, Staunton J, Leadlay PF: **A chain initiation factor common to both modular and aromatic polyketide synthases**. *Nature* 1999, **401**(6752):502-505.
26. Xue Y, Zhao L, Liu HW, Sherman DH: **A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(21):12111-12116.
27. Gu L, Geders TW, Wang B, Gerwick WH, Hakansson K, Smith JL, Sherman DH: **GNAT-like strategy for polyketide chain initiation**. *Science (New York, NY* 2007, **318**(5852):970-974.
28. Heathcote ML, Staunton J, Leadlay PF: **Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units**. *Chemistry & biology* 2001, **8**(2):207-220.
29. Tang Y, Kim CY, Mathews, II, Cane DE, Khosla C: **The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(30):11124-11129.
30. Tanovic A, Samel SA, Essen LO, Marahiel MA: **Crystal structure of the termination module of a nonribosomal peptide synthetase**. *Science (New York, NY* 2008, **321**(5889):659-663.
31. Zornetzer GA, Fox BG, Markley JL: **Solution structures of spinach acyl carrier protein with decanoate and stearate**. *Biochemistry* 2006, **45**(16):5217-5227.
32. Byers DM, Gong H: **Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family**. *Biochemistry and cell biology = Biochimie et biologie cellulaire* 2007, **85**(6):649-662.
33. Evans SE, Williams C, Arthur CJ, Burston SG, Simpson TJ, Crosby J, Crump MP: **An ACP structural switch: conformational differences between the apo and holo forms of the actinorhodin polyketide synthase acyl carrier protein**. *Chembiochem* 2008, **9**(15):2424-2432.
34. Xu GY, Tam A, Lin L, Hixon J, Fritz CC, Powers R: **Solution structure of B. subtilis acyl carrier protein**. *Structure* 2001, **9**(4):277-287.
35. Kim Y, Kovrigin EL, Eletr Z: **NMR studies of Escherichia coli acyl carrier protein: dynamic and structural differences of the apo- and holo-forms**. *Biochemical and biophysical research communications* 2006, **341**(3):776-783.
36. Wong HC, Liu G, Zhang YM, Rock CO, Zheng J: **The solution structure of acyl carrier protein from Mycobacterium tuberculosis**. *The Journal of biological chemistry* 2002, **277**(18):15874-15880.
37. Roujeinikova A, Baldock C, Simon WJ, Gilroy J, Baker PJ, Stuitje AR, Rice DW, Slabas AR, Rafferty JB: **X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site**. *Structure* 2002, **10**(6):825-835.
38. Roujeinikova A, Simon WJ, Gilroy J, Rice DW, Rafferty JB, Slabas AR: **Structural studies of fatty acyl-(acyl carrier protein) thioesters reveal a hydrophobic binding cavity that can expand to fit longer substrates**. *Journal of molecular biology* 2007, **365**(1):135-145.
39. Evans SE, Williams C, Arthur CJ, Ploskon E, Wattana-amorn P, Cox RJ, Crosby J, Willis CL, Simpson TJ, Crump MP: **Probing the Interactions of early polyketide intermediates with the Actinorhodin ACP from S. coelicolor A3(2)**. *Journal of molecular biology* 2009, **389**(3):511-528.

40. Koglin A, Loehr F, Bernhard F, Rogov VV, Frueh DP, Strieter ER, Mofid MR, Guentert P, Wagner G, Walsh CT *et al*: **Structural basis for the selectivity of the external thioesterase of the surfactin synthetase**. *Nature* 2008, **454**(7206):907-911.
41. Li Q, Khosla C, Puglisi JD, Liu CW: **Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein**. *Biochemistry* 2003, **42**(16):4648-4657.
42. Findlow SC, Winsor C, Simpson TJ, Crosby J, Crump MP: **Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from Streptomyces rimosus**. *Biochemistry* 2003, **42**(28):8423-8433.
43. Sharma AK, Sharma SK, Surolia A, Surolia N, Sarma SP: **Solution structures of conformationally equilibrium forms of holo-acyl carrier protein (PfACP) from Plasmodium falciparum provides insight into the mechanism of activation of ACPs**. *Biochemistry* 2006, **45**(22):6904-6916.
44. Kapur S, Worthington A, Tang Y, Cane DE, Burkart MD, Khosla C: **Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase**. *Bioorganic & medicinal chemistry letters* 2008, **18**(10):3034-3038.
45. Perrin MH, Grace CR, Riek R, Vale WW: **The three-dimensional structure of the N-terminal domain of corticotropin-releasing factor receptors: sushi domains and the B1 family of G protein-coupled receptors**. *Annals of the New York Academy of Sciences* 2006, **1070**:105-119.
46. Finking R, Mofid MR, Marahiel MA: **Mutational analysis of peptidyl carrier protein and acyl carrier protein synthase unveils residues involved in protein-protein recognition**. *Biochemistry* 2004, **43**(28):8946-8956.
47. Alekseyev VY, Liu CW, Cane DE, Puglisi JD, Khosla C: **Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase**. *Protein Sci* 2007, **16**(10):2093-2107.
48. Zhang YM, Rao MS, Heath RJ, Price AC, Olson AJ, Rock CO, White SW: **Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III**. *The Journal of biological chemistry* 2001, **276**(11):8231-8238.
49. Zhang YM, Wu B, Zheng J, Rock CO: **Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction**. *The Journal of biological chemistry* 2003, **278**(52):52935-52943.
50. Rafi S, Novichenok P, Kolappan S, Zhang X, Stratton CF, Rawat R, Kisker C, Simmerling C, Tonge PJ: **Structure of acyl carrier protein bound to FabI, the FASII enoyl reductase from Escherichia coli**. *The Journal of biological chemistry* 2006, **281**(51):39285-39293.
51. Kim Y, Ohlrogge JB, Prestegard JH: **Motional effects on NMR structural data. Comparison of spinach and Escherichia coli acyl carrier proteins**. *Biochemical pharmacology* 1990, **40**(1):7-13.
52. Kim Y, Prestegard JH: **A dynamic model for the structure of acyl carrier protein in solution**. *Biochemistry* 1989, **28**(22):8792-8797.
53. Andrec M, Hill RB, Prestegard JH: **Amide exchange rates in Escherichia coli acyl carrier protein: correlation with protein structure and dynamics**. *Protein Sci* 1995, **4**(5):983-993.
54. Gong H, Murphy A, McMaster CR, Byers DM: **Neutralization of acidic residues in helix II stabilizes the folded conformation of acyl carrier protein and variably alters its function with different enzymes**. *The Journal of biological chemistry* 2007, **282**(7):4494-4503.
55. Arthur CJ, Williams C, Pottage K, Ploskon E, Findlow SC, Burston SG, Simpson TJ, Crump MP, Crosby J: **Structure and malonyl CoA-ACP transacylase binding of streptomyces coelicolor fatty acid synthase acyl carrier protein**. *ACS chemical biology* 2009, **4**(8):625-636.
56. Gokhale RS, Tsuji SY, Cane DE, Khosla C: **Dissecting and exploiting intermodular communication in polyketide synthases**. *Science (New York, NY* 1999, **284**(5413):482-485.

57. Broadhurst RW, Nietlispach D, Wheatcroft MP, Leadlay PF, Weissman KJ: **The structure of docking domains in modular polyketide synthases**. *Chemistry & biology* 2003, **10**(8):723-731.
58. Hahn M, Stachelhaus T: **Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(44):15585-15590.
59. Sieber SA, Marahiel MA: **Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides**. *Journal of bacteriology* 2003, **185**(24):7036-7043.
60. Hahn M, Stachelhaus T: **Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(2):275-280.
61. Aparicio JF, Caffrey P, Marsden AF, Staunton J, Leadlay PF: **Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase**. *The Journal of biological chemistry* 1994, **269**(11):8524-8528.
62. Staunton J, Caffrey P, Aparicio JF, Roberts GA, Bethell SS, Leadlay PF: **Evidence for a double-helical structure for modular polyketide synthases**. *Nature structural biology* 1996, **3**(2):188-192.
63. Sandmann A, Sasse F, Muller R: **Identification and analysis of the core biosynthetic machinery of tubulysin, a potent cytotoxin with potential anticancer activity**. *Chemistry & biology* 2004, **11**(8):1071-1079.
64. Verdier-Pinard P, Lai JY, Yoo HD, Yu J, Marquez B, Nagle DG, Nambu M, White JD, Falck JR, Gerwick WH *et al*: **Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimitotic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells**. *Molecular pharmacology* 1998, **53**(1):62-76.
65. Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH: **Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula**. *Journal of natural products* 2004, **67**(8):1356-1367.
66. Gu L, Wang B, Kulkarni A, Gehret JJ, Lloyd KR, Gerwick L, Gerwick WH, Wipf P, Hakansson K, Smith JL *et al*: **Polyketide decarboxylative chain termination preceded by o-sulfonation in curacin a biosynthesis**. *Journal of the American Chemical Society* 2009, **131**(44):16033-16035.
67. Khare D, Wang B, Gu L, Razelun J, Sherman DH, Gerwick WH, Hakansson K, Smith JL: **Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(32):14099-14104.
68. Gu L, Jia J, Liu H, Hakansson K, Gerwick WH, Sherman DH: **Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair**. *Journal of the American Chemical Society* 2006, **128**(28):9014-9015.
69. Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, Hakansson K, Wipf P, Smith JL, Gerwick WH *et al*: **Metamorphic enzyme assembly in polyketide diversification**. *Nature* 2009, **459**(7247):731-735.
70. Geders TW, Gu L, Mowers JC, Liu H, Gerwick WH, Hakansson K, Sherman DH, Smith JL: **Crystal structure of the ECH2 catalytic domain of CurF from Lyngbya majuscula. Insights into a decarboxylase involved in polyketide chain beta-branching**. *The Journal of biological chemistry* 2007, **282**(49):35954-35963.
71. Akey DL, Razelun JR, Tehranisa J, Sherman DH, Gerwick WH, Smith JL: **Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway**. *Structure* 2010, **18**(1):94-105.

72. Piel J: **A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(22):14002-14007.
73. Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, Matsunaga S: **Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(46):16222-16227.
74. Simunovic V, Zapp J, Rachid S, Krug D, Meiser P, Muller R: **Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases**. *Chembiochem* 2006, **7**(8):1206-1220.
75. Partida-Martinez LP, Hertweck C: **A gene cluster encoding rhizoxin biosynthesis in "Burkholderia rhizoxina", the bacterial endosymbiont of the fungus Rhizopus microsporus**. *Chembiochem* 2007, **8**(1):41-45.
76. Gu L, Eisman EB, Dutta S, Franzmann TM, Walter S, Gerwick WH, Skiniotis G, Sherman DH: **Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect**. *Angewandte Chemie (International ed* 2011, **50**(12):2795-2798.
77. Keatinge-Clay A: **Crystal structure of the erythromycin polyketide synthase dehydratase**. *Journal of molecular biology* 2008, **384**(4):941-953.
78. Maier T, Jenni S, Ban N: **Architecture of mammalian fatty acid synthase at 4.5 A resolution**. *Science (New York, NY* 2006, **311**(5765):1258-1262.
79. Engel CK, Mathieu M, Zeelen JP, Hiltunen JK, Wierenga RK: **Crystal structure of enoyl-coenzyme A (CoA) hydratase at 2.5 angstroms resolution: a spiral fold defines the CoA-binding pocket**. *The EMBO journal* 1996, **15**(19):5135-5145.
80. Benning MM, Taylor KL, Liu RQ, Yang G, Xiang H, Wesenberg G, Dunaway-Mariano D, Holden HM: **Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 A resolution: an enzyme catalyst generated via adaptive mutation**. *Biochemistry* 1996, **35**(25):8103-8109.
81. Benning MM, Haller T, Gerlt JA, Holden HM: **New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from Escherichia coli**. *Biochemistry* 2000, **39**(16):4630-4639.
82. Sleeman MC, Sorensen JL, Batchelar ET, McDonough MA, Schofield CJ: **Structural and mechanistic studies on carboxymethylproline synthase (CarB), a unique member of the crotonase superfamily catalyzing the first step in carbapenem biosynthesis**. *The Journal of biological chemistry* 2005, **280**(41):34956-34965.
83. Hubbard PA, Yu W, Schulz H, Kim JJ: **Domain swapping in the low-similarity isomerase/hydratase superfamily: the crystal structure of rat mitochondrial Delta3, Delta2-enoyl-CoA isomerase**. *Protein Sci* 2005, **14**(6):1545-1555.
84. Mursula AM, van Aalten DM, Hiltunen JK, Wierenga RK: **The crystal structure of delta(3)-delta(2)-enoyl-CoA isomerase**. *Journal of molecular biology* 2001, **309**(4):845-853.
85. Partanen ST, Novikov DK, Popov AN, Mursula AM, Hiltunen JK, Wierenga RK: **The 1.3 A crystal structure of human mitochondrial Delta3-Delta2-enoyl-CoA isomerase shows a novel mode of binding for the fatty acyl group**. *Journal of molecular biology* 2004, **342**(4):1197-1208.
86. Leonard PM, Brzozowski AM, Lebedev A, Marshall CM, Smith DJ, Verma CS, Walton NJ, Grogan G: **The 1.8 A resolution structure of hydroxycinnamoyl-coenzyme A hydratase-lyase (HCHL) from Pseudomonas fluorescens, an enzyme that catalyses the transformation of feruloyl-coenzyme A to vanillin**. *Acta crystallographica* 2006, **62**(Pt 12):1494-1501.
87. Whittingham JL, Turkenburg JP, Verma CS, Walsh MA, Grogan G: **The 2-A crystal structure of 6-oxo camphor hydrolase. New structural diversity in the crotonase superfamily**. *The Journal of biological chemistry* 2003, **278**(3):1744-1750.

88. Blasiak LC, Vaillancourt FH, Walsh CT, Drennan CL: **Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis**. *Nature* 2006, **440**(7082):368-371.
89. Vaillancourt FH, Yin J, Walsh CT: **SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(29):10111-10116.
90. Vaillancourt FH, Yeh E, Vosburg DA, O'Connor SE, Walsh CT: **Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis**. *Nature* 2005, **436**(7054):1191-1194.
91. Galonic DP, Vaillancourt FH, Walsh CT: **Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis**. *Journal of the American Chemical Society* 2006, **128**(12):3900-3901.
92. Chang Z, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH: **The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit**. *Gene* 2002, **296**(1-2):235-247.
93. Flatt PM, O'Connell SJ, McPhail KL, Zeller G, Willis CL, Sherman DH, Gerwick WH: **Characterization of the initial enzymatic steps of barbamide biosynthesis**. *Journal of natural products* 2006, **69**(6):938-944.
94. Galonic DP, Barr EW, Walsh CT, Bollinger JM, Jr., Krebs C: **Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3**. *Nature chemical biology* 2007, **3**(2):113-116.
95. Challis GL, Naismith JH: **Structural aspects of non-ribosomal peptide biosynthesis**. *Current opinion in structural biology* 2004, **14**(6):748-756.
96. Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guentert P: **Optimal isotope labelling for NMR protein structure determinations**. *Nature* 2006, **440**(7080):52-57.
97. Takeda M, Chang CK, Ikeya T, Guentert P, Chang YH, Hsu YL, Huang TH, Kainosho M: **Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method**. *Journal of molecular biology* 2008, **380**(4):608-622.
98. Takeda M, Sugimori N, Torizawa T, Terauchi T, Ono AM, Yagi H, Yamaguchi Y, Kato K, Ikeya T, Jee J *et al*: **Structure of the putative 32 kDa myrosinase-binding protein from Arabidopsis (At3g16450.1) determined by SAIL-NMR**. *The FEBS journal* 2008, **275**(23):5873-5884.
99. Alberti E, Consonni R, Zetta L: **Applications of NMR to Thermostable Proteins** *Annual Reports on NMR Spectroscopy* 2003, **50**:121-161.
100. Reese ML, Doetsch V: **Fast mapping of protein-protein interfaces by NMR spectroscopy**. *Journal of the American Chemical Society* 2003, **125**(47):14250-14251.
101. Gaponenko V, Altieri AS, Li J, Byrd RA: **Breaking symmetry in the structure determination of (large) symmetric protein dimers**. *Journal of biomolecular NMR* 2002, **24**(2):143-148.
102. Malys N, McCarthy JE: **Translation initiation: variations in the mechanism can be anticipated**. *Cell Mol Life Sci* 2010, **68**(6):991-1003.
103. Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH: **Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase**. *Science (New York, NY* 1990, **250**(4981):651-657.
104. Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y: **Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae**. *The Journal of biological chemistry* 1990, **265**(12):6726-6733.
105. Paulus H: **Protein splicing and related forms of protein autoprocessing**. *Annual review of biochemistry* 2000, **69**:447-496.

106. Kawasaki M, Satow Y, Ohya Y, Anraku Y: **Protein splicing in the yeast Vma1 protozyme: evidence for an intramolecular reaction**. *FEBS letters* 1997, **412**(3):518-520.
107. Xu MQ, Perler FB: **The mechanism of protein splicing and its modulation by mutation**. *The EMBO journal* 1996, **15**(19):5146-5153.
108. Perler FB, Xu MQ, Paulus H: **Protein splicing and autoproteolysis mechanisms**. *Current opinion in chemical biology* 1997, **1**(3):292-299.
109. Shao Y, Xu MQ, Paulus H: **Protein splicing: evidence for an N-O acyl rearrangement as the initial step in the splicing process**. *Biochemistry* 1996, **35**(12):3810-3815.
110. Cooper AA, Chen YJ, Lindorfer MA, Stevens TH: **Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision**. *The EMBO journal* 1993, **12**(6):2575-2583.
111. Saleh L, Perler FB: **Protein splicing in cis and in trans**. *Chemical record (New York, NY* 2006, **6**(4):183-193.
112. Liu XQ: **Protein-splicing intein: Genetic mobility, origin, and evolution**. *Annual review of genetics* 2000, **34**:61-76.
113. Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E: **Inteins: structure, function, and evolution**. *Annual review of microbiology* 2002, **56**:263-287.
114. Perler FB: **InBase: the Intein Database**. *Nucleic acids research* 2002, **30**(1):383-384.
115. Belfort M, Roberts RJ: **Homing endonucleases: keeping the house in order**. *Nucleic acids research* 1997, **25**(17):3379-3388.
116. Chong S, Xu MQ: **Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs**. *The Journal of biological chemistry* 1997, **272**(25):15587-15590.
117. Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M: **Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(21):11466-11471.
118. Shingledecker K, Jiang SQ, Paulus H: **Molecular dissection of the Mycobacterium tuberculosis RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments**. *Gene* 1998, **207**(2):187-195.
119. Yamazaki T, Otomo T, Oda N, Kyogoku Y, Uegaki K, Ito N, Ishino Y, Nakamura aH: **Segmental Isotope Labeling for Protein NMR Using Peptide Splicing**. *Journal of the American Chemical Society* 1998, **120**(22):5591-5592.
120. Otomo T, Ito N, Kyogoku Y, Yamazaki T: **NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation**. *Biochemistry* 1999, **38**(49):16040-16044.
121. Giriat I, Muir TW: **Protein semi-synthesis in living cells**. *Journal of the American Chemical Society* 2003, **125**(24):7180-7181.
122. Evans TC, Jr., Martin D, Kolly R, Panne D, Sun L, Ghosh I, Chen L, Benner J, Liu XQ, Xu MQ: **Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803**. *The Journal of biological chemistry* 2000, **275**(13):9091-9094.
123. Wu H, Hu Z, Liu XQ: **Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(16):9226-9231.
124. Mills KV, Lew BM, Jiang S, Paulus H: **Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(7):3543-3548.
125. Southworth MW, Adam E, Panne D, Byer R, Kautz R, Perler FB: **Control of protein splicing by intein fragment reassembly**. *The EMBO journal* 1998, **17**(4):918-926.
126. Otomo T, Teruya K, Uegaki K, Yamazaki T, Kyogoku Y: **Improved segmental isotope labeling of proteins and application to a larger protein**. *Journal of biomolecular NMR* 1999, **14**(2):105-114.

127. Iwai H, Zuger S, Jin J, Tam PH: **Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme**. *FEBS letters* 2006, **580**(7):1853-1858.
128. Züger S, Iwai H: **Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies**. *Nature biotechnology* 2005, **23**(6):736-740.
129. Busche AE, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Doetsch V, Iwai H: **Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein**. *Angewandte Chemie (International ed* 2009, **48**(33):6128-6131.
130. Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ: **Production of cyclic peptides and proteins in vivo**. *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(24):13638-13643.
131. Iwai H, Lingel A, Pluckthun A: **Cyclic green fluorescent protein produced in vivo using an artificially split PI-Pful intein from Pyrococcus furiosus**. *The Journal of biological chemistry* 2001, **276**(19):16548-16554.
132. Williams NK, Prosselkov P, Liepinsh E, Line I, Sharipo A, Littler DR, Curmi PM, Otting G, Dixon NE: **In vivo protein cyclization promoted by a circularly permuted Synechocystis sp. PCC6803 DnaB mini-intein**. *The Journal of biological chemistry* 2002, **277**(10):7790-7798.
133. Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW: **Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo**. *Journal of the American Chemical Society* 2003, **125**(35):10561-10569.
134. Buskirk AR, Ong YC, Gartner ZJ, Liu DR: **Directed evolution of ligand dependence: small-molecule-activated protein splicing**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(29):10505-10510.
135. Ozawa T, Kaihara A, Sato M, Tachihara K, Umezawa Y: **Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing**. *Analytical chemistry* 2001, **73**(11):2516-2521.
136. Schwartz EC, Saez L, Young MW, Muir TW: **Post-translational enzyme activation in an animal via optimized conditional protein splicing**. *Nature chemical biology* 2007, **3**(1):50-54.
137. Kurpiers T, Mootz HD: **Regioselective cysteine bioconjugation by appending a labeled cystein tag to a protein by using protein splicing in trans**. *Angewandte Chemie (International ed* 2007, **46**(27):5234-5237.
138. Ludwig C, Pfeiff M, Linne U, Mootz HD: **Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing**. *Angewandte Chemie (International ed* 2006, **45**(31):5218-5221.
139. Chin HG, Kim GD, Marin I, Mersha F, Evans TC, Jr., Chen L, Xu MQ, Pradhan S: **Protein trans-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(8):4510-4515.
140. Oeemig JS, Aranko AS, Djupsjobacka J, Heinamaki K, Iwai H: **Solution structure of DnaE intein from Nostoc punctiforme: Structural basis for the design of a new split intein suitable for site-specific chemical modification**. *FEBS letters* 2009.
141. Aranko AS, Zuger S, Buchinger E, Iwai H: **In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins**. *PLoS ONE* 2009, **4**(4):e5185.
142. Muona M, Aranko AS, Raulinaitis V, Iwai H: **Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro**. *Nature protocols* 2010, **5**(3):574-587.
143. Martin DD, Xu MQ, Evans TC, Jr.: **Characterization of a naturally occurring trans-splicing intein from Synechocystis sp. PCC6803**. *Biochemistry* 2001, **40**(5):1393-1402.
144. Xu R, Ayers B, Cowburn D, Muir TW: **Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies**. *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(2):388-393.

145. Muona M, Aranko AS, Iwai H: **Segmental Isotopic Labelling of a Multidomain Protein by Protein Ligation by Protein Trans-Splicing**. *Chembiochem* 2008, **9**(18):2958-2961.
146. Yagi H, Tsujimoto T, Yamazaki T, Yoshida M, Akutsu H: **Conformational change of H+-ATPase beta monomer revealed on segmental isotope labeling NMR spectroscopy**. *Journal of the American Chemical Society* 2004, **126**(50):16632-16638.
147. Mullis KB, Faloona FA: **Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction**. *Methods in enzymology* 1987, **155**:335-350.
148. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: **Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase**. *Science (New York, NY* 1988, **239**(4839):487-491.
149. Inoue H, Nojima H, Okayama H: **High efficiency transformation of Escherichia coli with plasmids**. *Gene* 1990, **96**(1):23-28.
150. Laemmli UK, Beguin F, Gujer-Kellenberger G: **A factor preventing the major head protein of bacteriophage T4 from random aggregation**. *Journal of molecular biology* 1970, **47**(1):69-85.
151. Schagger H: **Tricine-SDS-PAGE**. *Nature protocols* 2006, **1**(1):16-22.
152. Schagger H, von Jagow G: **Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa**. *Analytical biochemistry* 1987, **166**(2):368-379.
153. Schwarz D, Junge F, Durst F, Frolich N, Schneider B, Reckel S, Sobhanifar S, Doetsch V, Bernhard F: **Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems**. *Nature protocols* 2007, **2**(11):2945-2957.
154. Schneider B, Junge F, Shirokov VA, Durst F, Schwarz D, Doetsch V, Bernhard F: **Membrane protein expression in cell-free systems**. *Methods in molecular biology (Clifton, NJ* 2010), **601**:165-186.
155. Haigler BE, Gibson DT: **Purification and properties of NADH-ferredoxinNAP reductase, a component of naphthalene dioxygenase from Pseudomonas sp. strain NCIB 9816**. *Journal of bacteriology* 1990, **172**(1):457-464.
156. Pervushin K, Riek R, Wider G, Wuthrich K: **Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(23):12366-12371.
157. Loehr F, Hansel R, Rogov VV, Doetsch V: **Improved pulse sequences for sequence specific assignment of aromatic proton resonances in proteins**. *Journal of biomolecular NMR* 2007, **37**(3):205-224.
158. Goddard TD, Kneller DG: **SPARKY 3**.
159. Ferrage F, Piserchio A, Cowburn D, Ghose R: **On the measurement of 15N-{1H} nuclear Overhauser effects**. *J Magn Reson* 2008, **192**(2):302-313.
160. Ogura K, Terasawa H, Inagaki F: **An improved double-tuned and isotope-filtered pulse scheme based on a pulsed field gradient and a wide-band inversion shaped pulse**. *Journal of biomolecular NMR* 1996, **8**(4):492-498.
161. Shen Y, Delaglio F, Cornilescu G, Bax A: **TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts**. *Journal of biomolecular NMR* 2009, **44**(4):213-223.
162. Guentert P: **Automated structure determination from NMR spectra**. *Eur Biophys J* 2009, **38**(2):129-143.
163. Herrmann T, Guentert P, Wuthrich K: **Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA**. *Journal of molecular biology* 2002, **319**(1):209-227.
164. Guentert P, Mumenthaler C, Wuthrich K: **Torsion angle dynamics for NMR structure calculation with the new program DYANA**. *Journal of molecular biology* 1997, **273**(1):283-298.

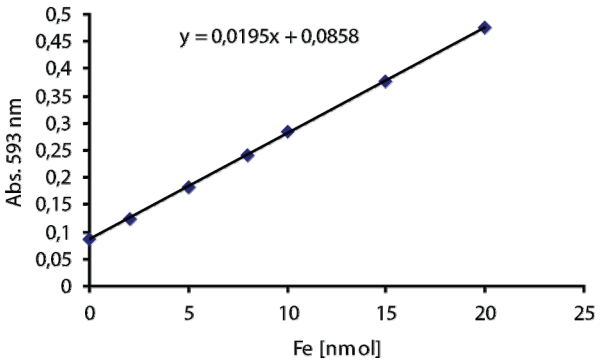
165. Koradi R, Billeter M, Güntert P: **Point-centered domain decomposition for parallel molecular dynamics simulation** *Computer Physics Communications* 2000, **124**(2-3):139-147
166. Ponder JW, Case DA: **Force fields for protein simulations**. *Advances in protein chemistry* 2003, **66**:27-85.
167. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM: **AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR**. *Journal of biomolecular NMR* 1996, **8**(4):477-486.
168. Hooft RW, Vriend G, Sander C, Abola EE: **Errors in protein structures**. *Nature* 1996, **381**(6580):272.
169. Guentert P: **Calculating protein structures from NMR data**. *Methods in molecular biology (Clifton, NJ* 1997), **60**:157-194.
170. Schuttelkopf AW, van Aalten DM: **PRODRG: a tool for high-throughput crystallography of protein-ligand complexes**. *Acta crystallographica* 2004, **60**(Pt 8):1355-1363.
171. Lee D, Hilty C, Wider G, Wuthrich K: **Effective rotational correlation times of proteins from NMR relaxation interference**. *J Magn Reson* 2006, **178**(1):72-76.
172. Slotboom DJ, Duurkens RH, Olieman K, Erkens GB: **Static light scattering to characterize membrane proteins in detergent solution**. *Methods (San Diego, Calif* 2008, **46**(2):73-82.
173. Bradshaw RA, Brickey WW, Walker KW: **N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families**. *Trends in biochemical sciences* 1998, **23**(7):263-267.
174. Reckel S, Sobhanifar S, Schneider B, Junge F, Schwarz D, Durst F, Loehr F, Guentert P, Bernhard F, Doetsch V: **Transmembrane segment enhanced labeling as a tool for the backbone assignment of alpha-helical membrane proteins**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(24):8262-8267.
175. Looman AC, Bodlaender J, Comstock LJ, Eaton D, Jhurani P, de Boer HA, van Knippenberg PH: **Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in Escherichia coli**. *The EMBO journal* 1987, **6**(8):2489-2492.
176. Iwahara J, Clore GM: **Detecting transient intermediates in macromolecular binding by paramagnetic NMR**. *Nature* 2006, **440**(7088):1227-1230.
177. Clore GM, Iwahara J: **Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes**. *Chemical reviews* 2009, **109**(9):4108-4139.
178. Crawford JM, Dancy BC, Hill EA, Udwary DW, Townsend CA: **Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(45):16728-16733.
179. Daragan VA, Ilyina EE, Fields CG, Fields GB, Mayo KH: **Backbone and side-chain dynamics of residues in a partially folded beta-sheet peptide from platelet factor-4**. *Protein Sci* 1997, **6**(2):355-363.
180. Shi J, Muir TW: **Development of a tandem protein trans-splicing system based on native and engineered split inteins**. *Journal of the American Chemical Society* 2005, **127**(17):6198-6206.
181. Dassa B, Amitai G, Caspi J, Schueler-Furman O, Pietrokovski S: **Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations**. *Biochemistry* 2007, **46**(1):322-330.
182. Volkmann G, Murphy PW, Rowland EE, Cronan JE, Jr., Liu XQ, Blouin C, Byers DM: **Intein-mediated cyclization of bacterial acyl carrier protein stabilizes its folded conformation but does not abolish function**. *The Journal of biological chemistry* 2010, **285**(12):8605-8614.

183. Muralidharan V, Muir TW: **Protein ligation: an enabling technology for the biophysical analysis of proteins**. *Nature methods* 2006, **3**(6):429-438.
184. Xu MQ, Evans TC, Jr.: **Intein-mediated ligation and cyclization of expressed proteins**. *Methods (San Diego, Calif* 2001), **24**(3):257-277.
185. Maier T, Leibundgut M, Boehringer D, Ban N: **Structure and function of eukaryotic fatty acid synthases**. *Quarterly reviews of biophysics* 2010, **43**(3):373-422.
186. Crump MP, Crosby J, Dempsey CE, Parkinson JA, Murray M, Hopwood DA, Simpson TJ: **Solution structure of the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2)**. *Biochemistry* 1997, **36**(20):6000-6008.
187. Volkman BF, Zhang Q, Debabov DV, Rivera E, Kresheck GC, Neuhaus FC: **Biosynthesis of D-alanyl-lipoteichoic acid: the tertiary structure of apo-D-alanyl carrier protein**. *Biochemistry* 2001, **40**(27):7964-7972.
188. Suo Z, Tseng CC, Walsh CT: **Purification, priming, and catalytic acylation of carrier protein domains in the polyketide synthase and nonribosomal peptidyl synthetase modules of the HMWP1 subunit of yersiniabactin synthetase**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(1):99-104.
189. Evans BS, Kelleher NL: **To cyclize or not to cyclize: catching enzyme evolution in the act**. *ACS chemical biology* 2009, **4**(7):495-497.
190. Ploskon E, Arthur CJ, Evans SE, Williams C, Crosby J, Simpson TJ, Crump MP: **A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains**. *The Journal of biological chemistry* 2008, **283**(1):518-528.
191. Tran L, Broadhurst RW, Tosin M, Cavalli A, Weissman KJ: **Insights into protein-protein and enzyme-substrate interactions in modular polyketide synthases**. *Chemistry & biology* 2010, **17**(7):705-716.
192. Conti E, Stachelhaus T, Marahiel MA, Brick P: **Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S**. *The EMBO journal* 1997, **16**(14):4174-4183.
193. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen LO: **Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase**. *Structure* 2007, **15**(7):781-792.
194. Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G: **Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase**. *Nature* 2008, **454**(7206):903-906.
195. Nygren PA, Stahl S, Uhlen M: **Engineering proteins to facilitate bioprocessing**. *Trends in biotechnology* 1994, **12**(5):184-188.
196. Esposito D, Chatterjee DK: **Enhancement of soluble protein expression through the use of fusion tags**. *Current opinion in biotechnology* 2006, **17**(4):353-358.
197. Dyson MR, Shadbolt SP, Vincent KJ, Perera RL, McCafferty J: **Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression**. *BMC biotechnology* 2004, **4**:32.
198. Lin L, Seehra J, Stahl ML: **High-throughput identification of refolding conditions for LXRbeta without a functional assay**. *Protein expression and purification* 2006, **47**(2):355-366.
199. Cheng Y, Patel DJ: **An efficient system for small protein expression and refolding**. *Biochemical and biophysical research communications* 2004, **317**(2):401-405.
200. Wang J, Chen W, Lu J, Lu S: **Overexpression and purification of recombinant atrial natriuretic peptide using hybrid fusion protein REF-ANP in Escherichia coli**. *Protein expression and purification* 2003, **28**(1):49-56.
201. Kojima S, Miyoshi K, Miura K: **Synthesis of a squash-type protease inhibitor by gene engineering and effects of replacements of conserved hydrophobic amino acid residues on its inhibitory activity**. *Protein engineering* 1996, **9**(12):1241-1246.
202. Kapust RB, Waugh DS: **Controlled intracellular processing of fusion proteins by TEV protease**. *Protein expression and purification* 2000, **19**(2):312-318.
203. Zhao W, Zhang Y, Cui C, Li Q, Wang J: **An efficient on-column expressed protein ligation strategy: application to segmental triple labeling of human apolipoprotein E3**. *Protein Sci* 2008, **17**(4):736-747.

204. Zhang Y, Chen J, Wang J: **A complete backbone spectral assignment of lipid-free human apolipoprotein E (apoE)**. *Biomolecular NMR assignments* 2008, **2**(2):207-210.
205. Chen J, Wang J: **A segmental labeling strategy for unambiguous determination of domain-domain interactions of large multi-domain proteins**. *Journal of biomolecular NMR* 2011, **50**(4):403-410.
206. Walters KJ, Dayie KT, Reece RJ, Ptashne M, Wagner G: **Structure and mobility of the PUT3 dimer**. *Nature structural biology* 1997, **4**(9):744-750.
207. Zwahlen C, Legault P, Vincent SJF, Greenblatt J, Konrat R, Kay LE: **Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage lambda N-Peptide/boxB RNA Complex**. *Journal of the American Chemical Society* 1997 **119**(29):6711-6721.
208. Jeschke G, Polyhach Y: **Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance**. *Phys Chem Chem Phys* 2007, **9**(16):1895-1910.
209. Gillespie JR, Shortle D: **Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels**. *Journal of molecular biology* 1997, **268**(1):158-169.
210. Getz EB, Xiao M, Chakrabarty T, Cooke R, Selvin PR: **A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry**. *Analytical biochemistry* 1999, **273**(1):73-80.
211. Iwahara J, Tang C, Marius Clore G: **Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules**. *J Magn Reson* 2007, **184**(2):185-195.
212. Tang C, Iwahara J, Clore GM: **Visualization of transient encounter complexes in protein-protein association**. *Nature* 2006, **444**(7117):383-386.
213. Lai JR, Fischbach MA, Liu DR, Walsh CT: **Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection**. *Journal of the American Chemical Society* 2006, **128**(34):11002-11003.
214. Lai JR, Fischbach MA, Liu DR, Walsh CT: **A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(14):5314-5319.

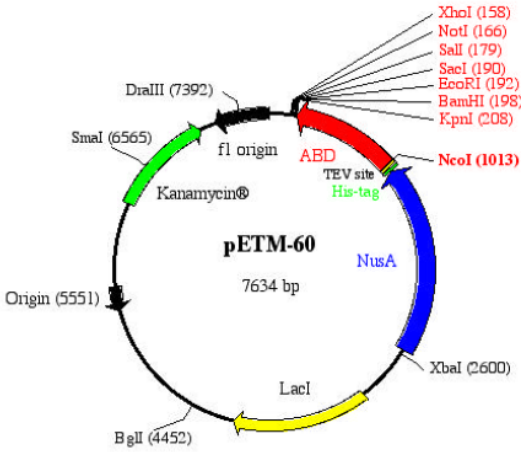
11 Appendix

11.1 Calibration curve for the Ferene S iron content determination



App. 1: Calibration curve for the Ferene S iron content determination. Iron amounts were varied between 0 and 20 nmol and measured versus Ferene S solution without added iron.

11.2 Vector map: petM-60 from EMBL

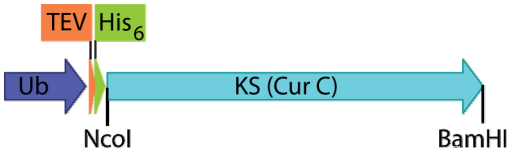


App. 2: petM- 60 vector from EMBL

The vector map was obtained from

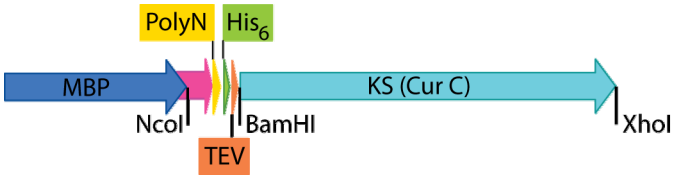
http://www.embl.de/pepcore/pepcore_services/cloning/choice_vector/ecoli/embl/)

11.3 Open reading frame (ORF) of Ub-KS (CurC) in modified petM-60



App. 3: Open reading frame (ORF) of Ub-KS (CurC) in modified petM-60

11.4 Open reading frame (ORF) of PolyN-TEV-KS (CurC) in modified pMAL-c2x



App. 4: Open reading frame (ORF) of the MBP-PolyN-His$_6$-TEV-KS (CurC) construct. The backbone of the vector is identical to pMAL-c2x

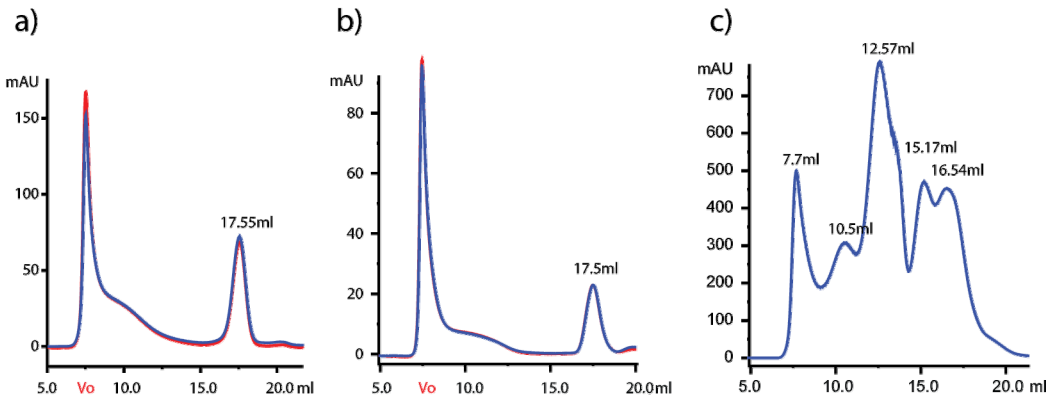
11.5 Sequence alignment of different KS-AT didomains

```
DEBS3           --------------------MELES-------------DPIAIVSMACRLPGGVNTPQRL 27
DEBS5           MSGDNGMTEEKLRRYLKRTVTELDSVTARLREVEHRAGEPIAIVGMACRFPGDVDSPESF 60
CurA            -----------------FSTAQLEN-------------GVAIIGMACRFPGDADSPQQY 29
huFAS           ------------------------------------MEEVVIAGMSGKLP-ESENLQEF 22
                                                     :.* .*: ::*   :. :

DEBS3           WELLREGGETLSGFPTDRGWDLARLHHPDPDNPGTSYVDKGGFLDDAAGFDAEFFGVSPR 87
DEBS5           WEFVSGGGDAIAEAPADRGWE------PDPD------ARLGGMLAAAGDFDAGFFGISPR 108
CurA            WSLLHDGIDGITEVPPTR-WDIEQYYHPEKNQPGKIANRYGGFLTEVDKFEPDFFRISAR 88
huFAS           WDNLIGGVDMVTD--DDRRWKAGLYGLPRRS----------GKLKDLSRFDASFFGVHPK 70
                *. :  * : ::     * *.      *  .           * *     *:. ** : .:

DEBS3           EAAAMDPQQRLLLETSWELVENAGIDPHSLRGTATGVFLGVAKFGYG-EDTAAAEDVEG- 145
DEBS5           EALAMDPQQRIMLEISWEALERAGHDPVSLRGSATGVFTGVGTVDYGPRPDEAPDEVLG- 167
CurA            EALYMDPQQRLLLEEHWKALEDAGINPESLSGTETGIFVGIAFHDYERLQDKYYQEQDLN 148
huFAS           QAHTMDPQLRLLLEVTYEAIVDGGINPDSLRGTHTGVWVGVSGSETSEALSRDPETLVG- 129
                :*  **** *::**  :: :  .* :* ** *: **:: *:.            :

DEBS3           -YSVTGVAPAVASGRISYTMGLEGPSISVDTACSSSLVALHLAVESLRKGESSMAVVGGA 204
DEBS5           -YVGTGTASSVASGRVAYCLGLEGPAMTVDTACSSGLTALHLAMESLRRDECGLALAGGV 226
CurA            IYFATGSSTAIGAGRLSYFFQLNGPSITVDTACSSSLSAVHLACQSIRNGECQLALASGV 208
huFAS           -YSMVGCQRAMMANRLSFFFDFRGPSIALDTACSSSLMALQNAYQAIHSGQCPAAIVGGI 188
                 *  .*   :: :.*::: : :.**::::******.* *:: * :::: .:.  *:..*

DEBS3           AVMATPGVFVDFSRQRALAADGRSKAFGAGADGFGFSEGVTLVLLERLSEARRNGHEVLA 264
DEBS5           TVMSSPGAFTEFRSQGGLAADGRCKPFSKAADGFGLAEGAGVLVLQRLSAARREGRPVLA 286
CurA            NLLLSPELSISFSQAGMLSPDGRCKTFDAAANGYVRSEGCGVVVLKSLKQAIADNDRILA 268
huFAS           NVLLKPNTSVQFLRLGMLSPEGTCKAFDTAGNGYCRSEGVVAVLLTKKSLARR----VYA 244
                 :: .*    .*     *:.:* .*.*. ..:*:  :**   ::*   . *      : *

DEBS3           VVRGSALNQDGASN-GLSAPSGPAQRRVIRQALESCGLEPGDVDAVEAHGTGTALGDPIE 323
DEBS5           VLRGSAVNQDGASN-GLTAPSGPAQQRVIRRALENAGVRAGDVDYVEAHGTGTRLGDPIE 345
CurA            VVRGTAINQDGASN-GLTAPNQSAQEAVLKRALSVAGVSANQISYVEAHGTGTSLGDPVE 327
huFAS           TILNAGTNTDGFKEQGVTFPSGDIQEQLIRSLYQSAGVAPESFEYIEAHGTGTKVGDPQE 304
                .: .:. * ** .: *:: *.   *. :::   . .*: . ... :******* :*** *

DEBS3           ANALLDTYGRDRDADRPLWLGSVKSNIGHTQAAAGVTGLLKVVLALRNGELPATLHVEEP 383
DEBS5           VHALLSTYGAERDPDDPLWIGSVKSNIGHTQAAAGVAGVMKAVLALRHGEMPRTLHFDEP 405
CurA            IKAIEAVYGKDRSADRPLMIGSVKTNIGHTEAAAGIAGLIKVVLSLQNQYIPPHLHWKQL 387
huFAS           LNGITRALCATR--QEPLLIGSTKSNMGHPEPASGLAALAKVLLSLEHGLWAPNLHFHSP 362
                 :.:  .    *  : ** :**.*:*:**.:.*:*::.: *.:*:*.:   .  ** ..

DEBS3           TPHVDWSSGGVALLAGNQPWRR--GERTRRAAVSAFGISGTNAHVIVEEAPEREHRETTA 441
DEBS5           SPQIEWDLGAVSVVSQARSWPA--GERPRRAGVSSFGISGTNAHVIVEEAPEADEPEPAP 463
CurA            NPYISLAGIPGVIPTEGKVWEQYEGDETRVAALSSFGFSGTNSHAIVEEAPEGRSQEPGA 447
huFAS           NPEIPALLDGRLQVVDQPLPVR-----GGNVGINSFGFGGSNVHIILRPNTQPPPAPAPH 417
                .* :                         ..:.:**:.*:* * *:.  .:     .

DEBS3           HDGR--------PVPLVVSARSTAALRAQAAQIAELLER-PDADLAGVGLGLATTRARHE 492
DEBS5           DSG---------PVPLVLSGRDEQAMRAQAGRLADHLAREPRNSLRDTGFTLATRRSAWE 514
CurA            RSQEPGARSQEPPYLLTVSAKKEQALKELVSSYQHHLETNPELELADVCYTANTGRADFN 507
huFAS           ATL---------PRLLRASGRTPEAVQKLLEQGLRHSQDLAFLSMLNDIAAVPATAMPFR 468
                             *  *  *.:   *::             .  .: .      :     .

DEBS3           HRAAVVASTREEAVRGLREIAAGAATADAVVEGVTEVDG-RNVVFLFPGQGSQWAGMGAE 551
DEBS5           HRAVVVG-DRDEALAGLRAVADG-RIADRTATGQARTR--RGVAMVFPGQGAQWQGMARD 570
CurA            HRLAIIATNPQQLTNKLRQYQAGEEESGVFSGQLSHSSSPAKVAFLFTGQGSQYVQMGRQ 567
huFAS           GYAVLGG---ERGGPEVQQVPAG---------------ERPLWFICSGMGTQWRGMGLS 509
                   .: .   :.    ::    *                  : :: .* *:*:  *. .
```

```
DEBS3           LLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAP----GAPGLDRVDVVQPVLFAVMVS 607
DEBS5           LLRESQVFADSIRDCERALAPHVDWSLTDLLS-------GARPLDRVDVVQPALFAVMVS 623
CurA            LYQTQPVFRQVLDQCDELLRPYLERPLLEVLYPQDTPNSNSYLLDQTAYTQPTLFALEYA 627
huFAS           LMR-LDRFRDSILRSDEAVKPFGLKVSQLLLS------TDESTFDDIVHSFVSLTAIQIG 562
                *      *   :  .:. : *        :*        .   :*        * *:  .

DEBS3           LAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRLMRSLSGE-GGMAA 666
DEBS5           LAALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLRRLGGQ-GGMAS 682
CurA            LCKLWESWGIKPQVVMGHSVGEYVAATIAGVLSLEDGLKLIALRGRLMQQLPAG-GEMVS 686
huFAS           LIDLLSCMGLRPDGIVGHSLGEVACGYADGCLSQEEAVLAAYWRGQCIKEAHLPPGAMAA 622
                *  *  . *:.*  ::*** ** ...   * *: *:.       *.: ::      * *.:

DEBS3           VALGEAAVRERLRPWQDRLSVAAVNGPRSVVVSGEPGALRAFSEDCAAEGIRVRDIDVD- 725
DEBS5           FGLGTEQAAERIGRFAGALSIASVNGPRSVVVAGESGPLDELIAECEAEAHKARRIPVD- 741
CurA            VMASKSQVKDAIANHTKQVTIAAINGPESVVISGEAGAIQAIVTKLESKLIKTKQLQVS- 745
huFAS           VGLSWEECKQRCP---PGVVPACHNSKDTVTISGPQAPVFEFVEQLRKEGVFAKEVRTGG 679
                .  .     :        :  *. *.  :*.::*  ..:  :  .   :   .: : ..

DEBS3           YASHSPQIERVREELLETTGDIAPRPAR------VTFHSTVESRSMDGTELDARYWYRNL 779
DEBS5           YASHSPQVESLREELLTELAGISPVSAD------VALYSTTTGQPIDTATMDTAYWYANL 795
CurA            HAFHSPLMTPMLAEFAAVAQQITYHQPR------IPVISNVTGTIADKSIATADYWVEHV 799
huFAS           MAFHSYFMEAIAPPLLQELKKVIREPKPRSARWLSTSIPEAQWHSSLARTSSAEYNVNNL 739
                 * **  :  :   :      :              .  . .           : *   ::

DEBS3           RETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAED-AVVVGSLHRDGGD 838
DEBS5           REQVRFQDATRQLAEAGFDAFVEVSPHPVLTVGIEATLDSALPADAGACVVGTLRRDRGG 855
CurA            VKPVRFVAGIKTLAEQDIRIFLEIGPKPVLLVMGRECLIGSK-----KIWLPSLRPGKPD 854
huFAS           VSPVLFQEALWHVPEH--AVVLEIAPHALLQAVLKRGLKPSC-----TIIPLMKKDHRDN 792
                 . * *  .   :.*     .:*:.*:.::    .  :  :             :    .

DEBS3           LSAFLRSMATAHVSGVDIRWDVALPGAAP-------FALPTYPFQRKRYWLQPAAPAAAS 891
DEBS5           LADFHTALGEAYAQGVEVDWSPAFADARP-------VELPVYPFQRQRYWLPIPTGGRAR 908
CurA            WLQMLQSLGQLYVQGVKVDWLGFYPDDAPQK-----VVLPTYPWQRKRYWISDLQQYKNK 909
huFAS           LEFFLAGIGRLHLSGIDANPNALFPPVEFPAPRGTPLISPLIKWDHSLAWDVPAAEDFPN 852
                   :  .:.  : .*:.       .           .  *   :::.  *

DEBS3           DELAYRSSSVDKLAAALEHHHHHH---------------------------------- 915
DEBS5           DE---------------DDDWRYQ---------------------------------- 917
CurA            GK---------------------------------------------------------- 911
huFAS           GSGSPSAAIYNIDTSSESPDHYLVDHTLDGRVLFPATGYLSIVWKTLARALGLGVEQLP 911
```

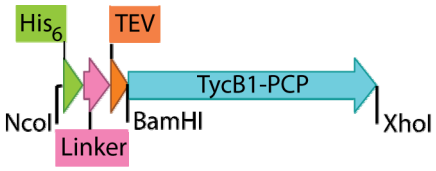
App. 5: Sequence alignment of different KS- AT didomains. The sequence alignment of CurA KS-AT, DEBS3 KS-AT, DEBS5 KS-AT, huFAS KS-AT didomain was performed using ClustalW. The beginning and end points of the domains are colored. KS in red and AT in yellow.

11.6 Gelfiltration profile for some constructs of the Cur KS-AT didomain



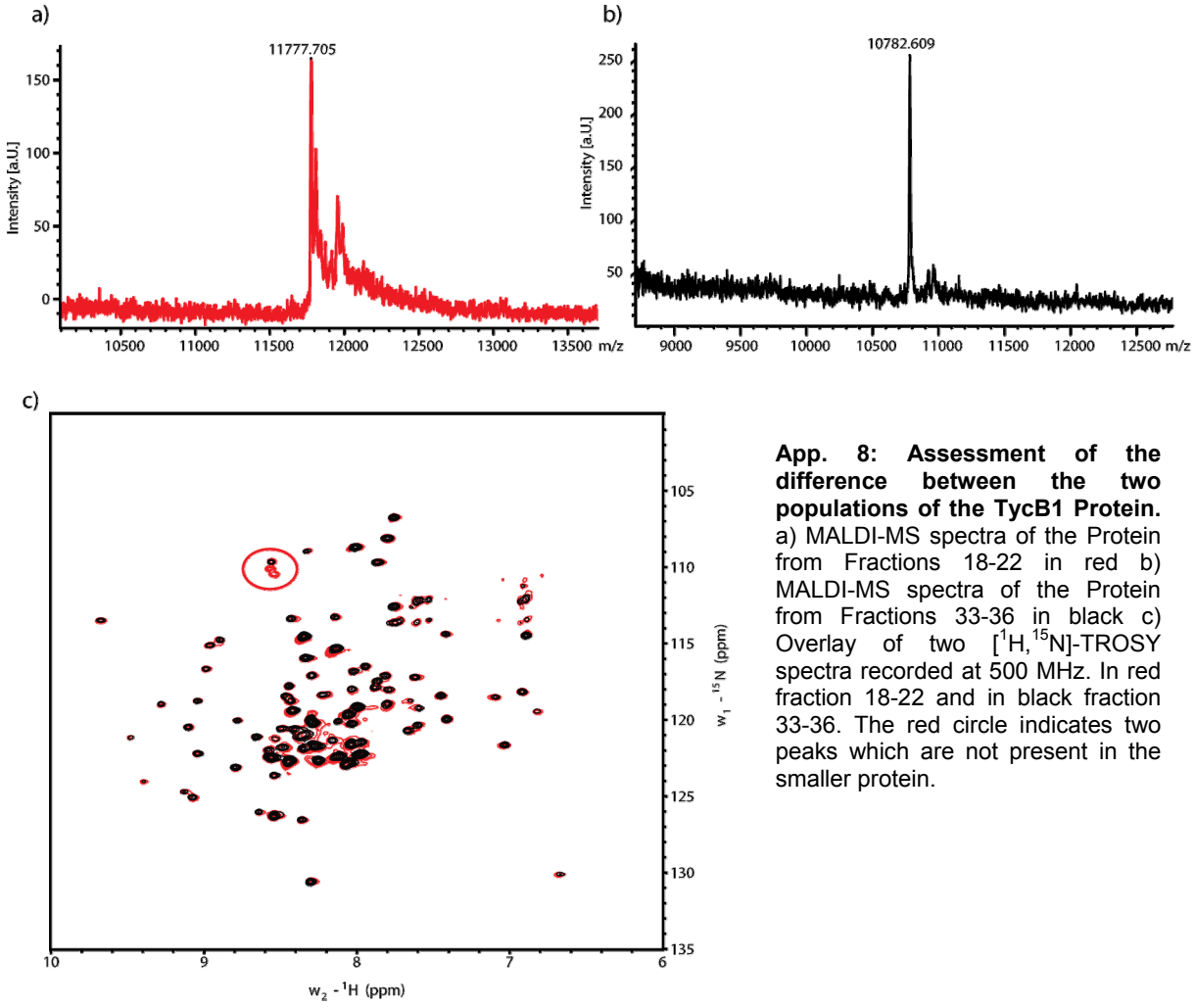
App. 6: Gelfiltration profile for some constructs of the Cur KS-AT didomain. All gelfiltrations were performed at 16°c on a superose 6 column. a) Overlay of gelfiltration profiles of MBP-KS, the proteinconcentration was 3 mg/ml with 10 mM DTT (blue) and without DTT (red). DTT has no impact on aggregation. b) Overlay of gelfiltration profiles of MBP-AT-L at a concentration of 1.3 mg/ml with 10 mM DTT (red) and without DTT (blue) c) gelfiltration profile of AT. The peak at 17.55 and 17.5 ml in a) and b) is isolated MBP which was proven by MALDI-MS.

11.7 Open reading frame of His_6-TEV-TycB1 PCP in pBH4



App. 7: Open reading frame (ORF) of the His_6-TEV-TycB1 PCP. The backbone of the vector is identical to pBH4

11.8 TycB1 PCP



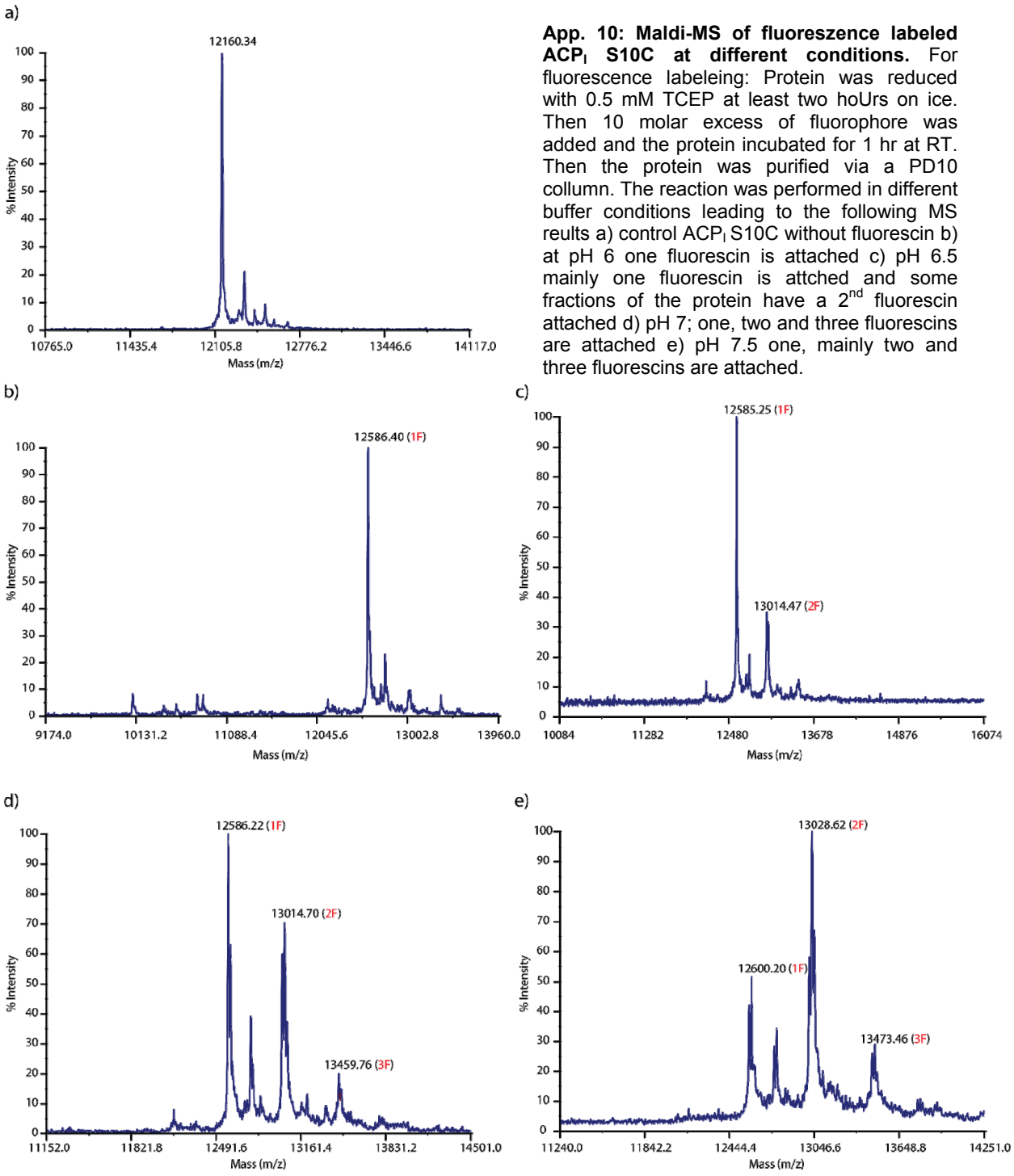
App. 8: Assessment of the difference between the two populations of the TycB1 Protein. a) MALDI-MS spectra of the Protein from Fractions 18-22 in red b) MALDI-MS spectra of the Protein from Fractions 33-36 in black c) Overlay of two [^{1}H,^{15}N]-TROSY spectra recorded at 500 MHz. In red fraction 18-22 and in black fraction 33-36. The red circle indicates two peaks which are not present in the smaller protein.

11.9 Structural statistics

App. 9: Statistics derived from Cyana for the structure of holo and (S)-hydroxyl-methylglutaryl (HMG)-ACP$_I$

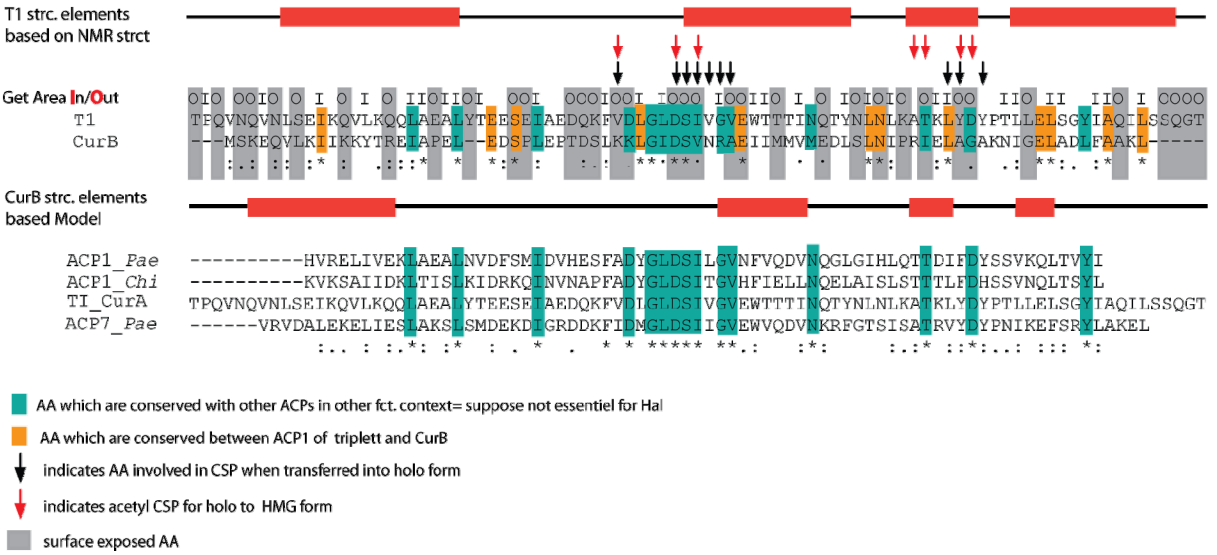
NMR restraints	Holo	HMG	Structural statistics	Holo	HMG
Number of peaks			**Restraint violations**		
^{13}C NOESY-HSQC	5260	5078	Max. Dihedral angle restraint violations (°)	/	/
^{15}N NOESY-HSQC	1557	1557	Max. distance restraint violations (A)	0.12	0.21
$^{13}C^{Arom}$ NOESY-HSQC	58	57	Cyana target function value (Å^2)	2.42	1.56
Hydroxy-methyl-glutaryl					
Intramolecular NOEs	-	86			
Intermolecular NOEs	-	37			
Distance restraints			**Precision for residues 1950-2033**		
Total NOE	2866	2923	**R.m.s.d to mean coordinates (Å)**		
Short range |i-j|<=1	1307	1349	Heavy atoms	0.54 ± 0.06	1.01 ± 0.29
Medium-range 1<|i-j|<5	826	819	Backbone	0.24 ± 0.04	0.11 ± 0.02
Long range |i-j|>=5	734	755	**Ramachandran plot staistics (%)**		
Hydrogen bonds	-		most favored	88.6	87.1
			additionally allowed	11.4	12.9
Dihedral angle restraints			generously allowed	0	0
Talos+ Φ/ Ψ	162	162	disallowed	0	0

11.10 Results of the optimization of the fluorescence labeling of ACP_I



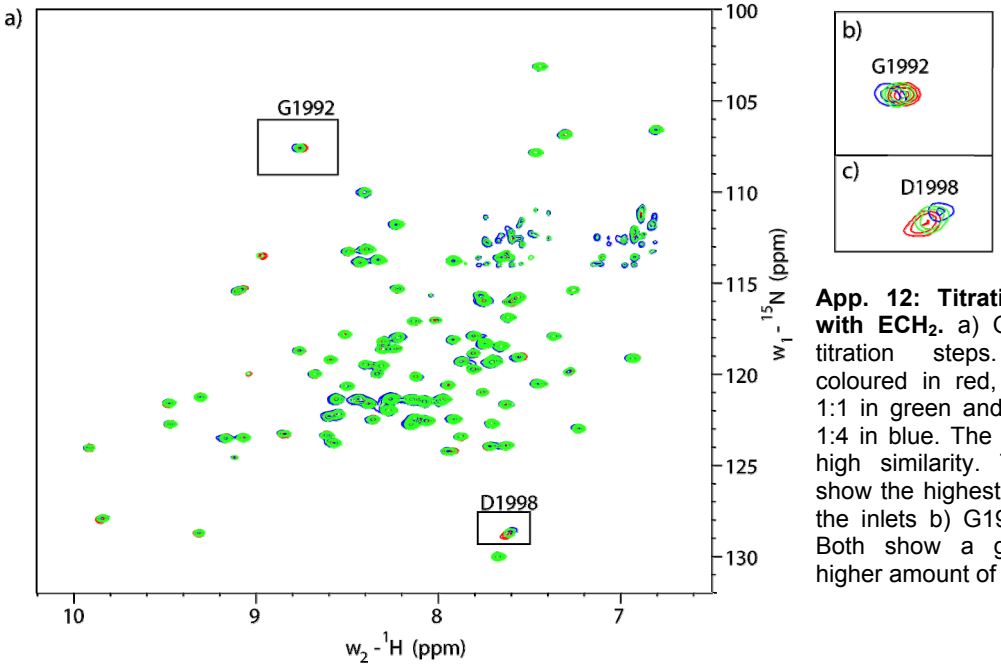
App. 10: Maldi-MS of fluoreszence labeled ACP_I S10C at different conditions. For fluorescence labeleing: Protein was reduced with 0.5 mM TCEP at least two hoUrs on ice. Then 10 molar excess of fluorophore was added and the protein incubated for 1 hr at RT. Then the protein was purified via a PD10 collumn. The reaction was performed in different buffer conditions leading to the following MS reults a) control ACP_I S10C without fluorescin b) at pH 6 one fluorescin is attached c) pH 6.5 mainly one fluorescin is attched and some fractions of the protein have a 2nd fluorescin attached d) pH 7; one, two and three fluorescins are attached e) pH 7.5 one, mainly two and three fluorescins are attached.

11.11 Basis for the rational mutation of ACP_I



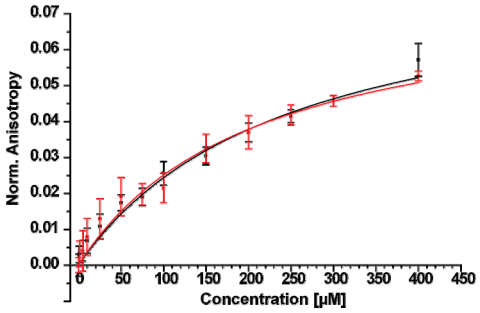
App. 11: Basis for the rational mutation of ACP_I. The alignement of ACP_I with ACP with complete other function indicates regions which are conserved between ACPs (cyan). The alignment between CurB and ACP_I shows further amino acids which might be relevant for the general ACP stability (orange). Amino acids which are solvent exposed (calculated with Get Area) and therefore target for mutations are indicate in grey. On the top and bottom of the alignement are indicated the structural elements from the calculated ACP_I structure and the modeled CurB structure. Arrows indicate the AA which underly CSP for the apo to holo switch with CSP > 0.05 ppm (black) or the holo to HMG conversion with CSP> 0.02 ppm (red).

11.12 NMR titration of HMG-ACP_I with ECH_2



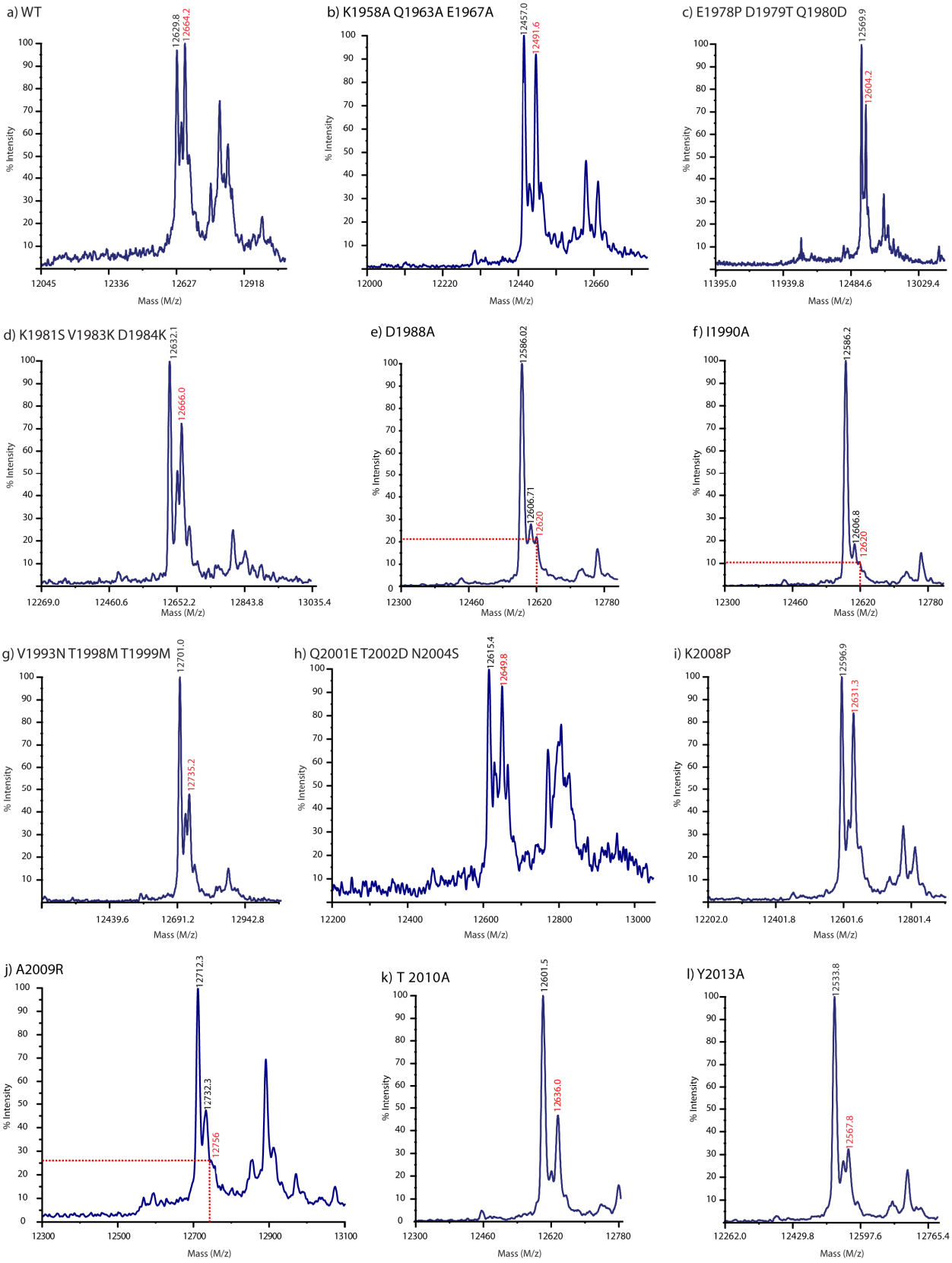
App. 12: Titration of HMG-ACP_I with ECH_2. a) Overlay of different titration steps. HMG-ACP_I is coloured in red, HMG-ACP_I: ECH_2 1:1 in green and HMG-ACP_I: ECH_2 1:4 in blue. The spectra match with high similarity. Two peaks which show the highest CSP are shown in the inlets b) G1992 and c) D1988. Both show a growing CSP with higher amount of ECH_2 added.

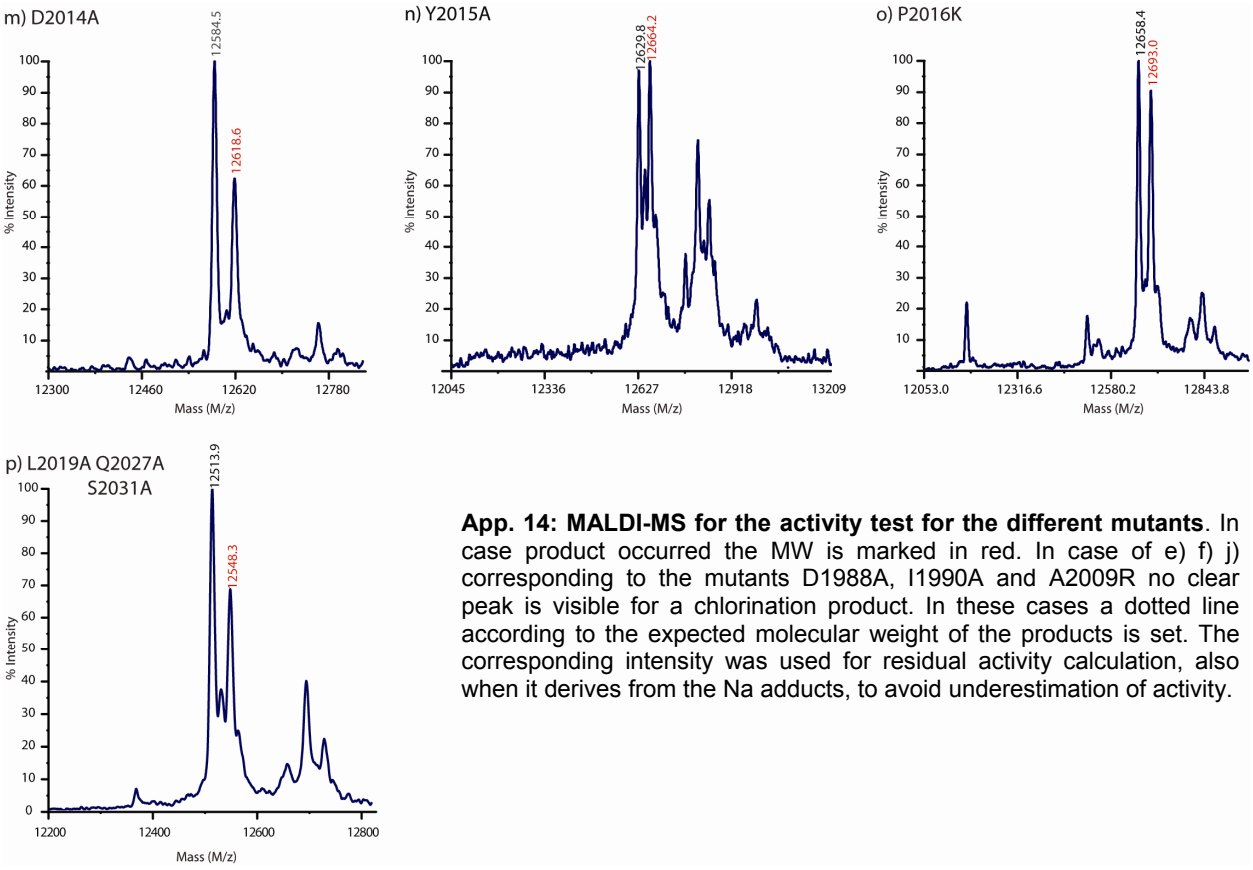
11.13 Fluorescence anisotropy of ACP_I with ECH_2



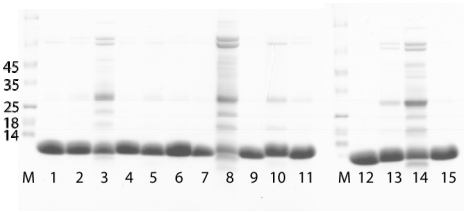
App. 13: Fluorescence anisotropy measurements of ACP with ECH_2. Blot of the normalized anisotropy of HMG-ACP_I S90C (black) and apo (red) both titrated with ECH_2. In both cases no saturation was reached. The binding constants (247µM for apo ACP_I and ECH_2 and 209µM for HMG-ACP_I and ECH_2) were derived by fitting a one-site-binding model present more likely unspecific hydrophobic interaction and no specific interaction for HMG-ACP_I with ECH_2.

11.14 Data for the activity test for different ACP_I mutants and Cur Hal

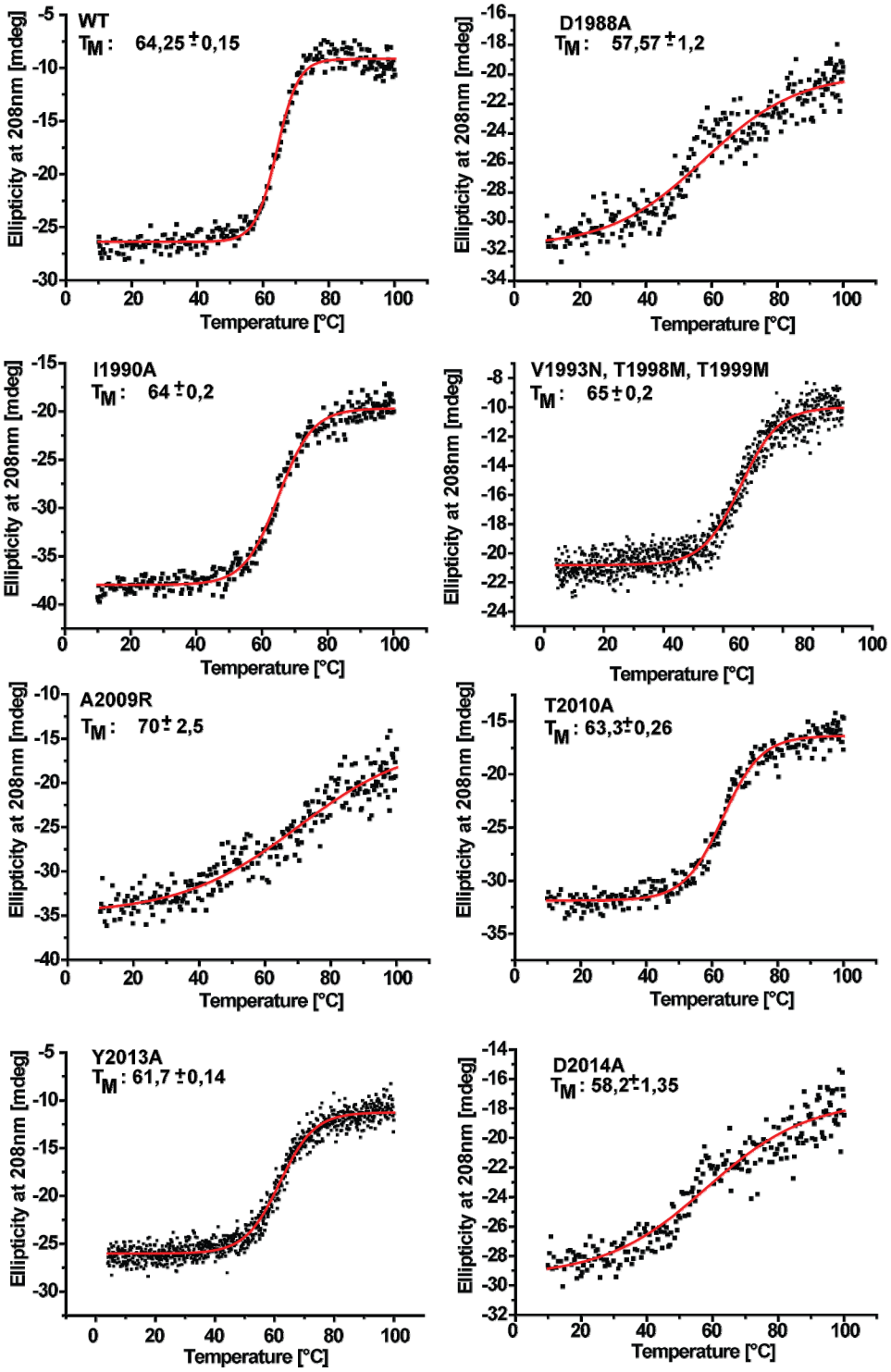




App. 14: MALDI-MS for the activity test for the different mutants. In case product occurred the MW is marked in red. In case of e) f) j) corresponding to the mutants D1988A, I1990A and A2009R no clear peak is visible for a chlorination product. In these cases a dotted line according to the expected molecular weight of the products is set. The corresponding intensity was used for residual activity calculation, also when it derives from the Na adducts, to avoid underestimation of activity.



App. 15: Tricine gel of ACP$_I$ mutants. 1: WT- ACP$_I$; 2: Q1958A, Q1963A, E1967A ; 3: E1978P, D1979T, Q1980D 4:K1981S, V1982K, D19835K;5: D1988A; 6: I1990A; 7: V1993N,T1998M, T1999M; 8: Q2001E, T2002D, N2004S; 9: K2008P; 10: A2009R; 11: T2010A; 12: Y2013A; 13: D2014A; 14: Y2015A; 15: L2019A.



App. 16: CD melting curves of the mutants displaying a decrease in activity.

Acknowledgements

First of all, I like to thank Prof. Volker Dötsch for being an excellent supervisor, his scientific guidance and constant generous support during my PhD thesis. I gratefully acknowledge his open-minded character and his support of new ideas. Paired with his enthusiasm about science, he generated an atmosphere of inspiration, which was of great benefit for my work. Most of all, I am grateful for the opportunity to spend a very inspiring time in Helsinki at the Institute of Biotechnology and got the chance to participate in several conference and the EMBO summer school.

I want to especially thank Dr. Frank Löhr. He introduced me into the field of NMR, I am grateful for his patience, explanations and all the NMR measurements he performed for me. This was a highly productive collaboration which made my PhD thesis possible.

I am much indebted to Dr. Hideo Iwai for welcoming me in his laboratory and for the constant support during my research stay in his group at the Institute of Biotechnology in Helsinki. The collaboration about segmental labeling was highly productive and I could highly benefit from his knowledge. In this context I want to thank as well Sesilja Aranko who was a helping hand in Hideo Iwai's lab in Helsinki. Together with Jesper Oemig, both created a highly productive and enjoyable atmosphere in the lab.

My ACP_I project was based on a close collaboration with Prof. Dr. Peter Güntert and Daniel Gottstein. I would like to thank them for their cooperativeness and many fruitful discussions.

Especially I am grateful to Prof. Dr. Clemens Glaubitz for introducing me into solid state NMR and for co-examining this thesis. I want to thank him for his support and helpful advice.

Furthermore I wanna thank my students Nina Ripin, Christopher Hein, Irina Pader and Florian Sochor who contributed to this work during their diploma thesis or long term internships. For a productive teamwork in the lab and helpful discussions I would like to thank Sina Reckel, Robert Hänsel, Jan Heering and Gregor Deutsch as well as the rest of the Dötsch lab. I could benefit a lot from their knowledge. In addition I want to thank them and the rest of the Dötsch group for creating an enjoyable and friendly atmosphere in the lab.

Further I am thankful to Robert Hänsel, Gregor Deutsch, Dr. Aisha Laguerre and Sebastian Kehrlößer for proof-reading this thesis. Also I would like to thank Birgit Schäfer, Manfred Strupf, Natalja Rogova and Sigrid Oguzer-Fachinger who are all responsible for the smooth flow in the lab. Together with all the other lab members they contributed significantly to a good working atmosphere. Many thanks also to my parents and family for their moral support throughout my education. I would like to thank my friends, my roommates and T7 for their encouragement and support outside the lab. I want to thank as well my fiancé Charbel for his encouragement and for being a delightful partner.

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Dissertation über „ Structural and functional characterization of the ACP triplet in the curacin cluster and its interaction partners “ selbstständig angefertigt und mich anderer Hilfsmittel als der in der Dissertation angegebenen nicht bedient habe. Alle Entlehnungen aus anderen Schriften sind mit Angabe der betreffenden Schrift gekennzeichnet. Ich erkläre weiterhin, dass ich mich bisher keiner Doktorprüfung unterzogen habe.

Frankfurt, den

_____

Alena Busche, Teichstraße 7, 60594 Frankfurt am Main
Cellphone: 0049 (0)176- 20496644, Email: alenabusche@gmx.de

Curiculum vitae

Personal details

Date of birth: **06. Decembre 1982**
Place of birth: **Oldenburg**
Nationality: **German**

Education

11/2007- 11/2011	**Goethe University of Frankfurt** pHD at the Institute of Biophysical Chemistry Supervisor: Prof. Dr. V. Dötsch	**Frankfurt, Germany**
01/2007 – 09/2007	**Harvard Medical School** Diploma thesis in the Department of Cell Biology, Supervisor: Prof. Dr. Frank McKeon	**Boston, USA**
09/2004 – 10/2007	**École Superieure de Biotechnologie Strasbourg (ESBS)** Master's degree in Biotechnology Grade: 15.07 ('very good')	**Strasbourg, France**
10/2002 – 07/2004	**Leibnitz University** Intermediate examen in Chemistry Grade: 1.0 ('excellent')	**Hanover, Germany**
09/1994 - 10/2002	**Gymnasium Lehrte, grammar school** German university entrance level	**Lehrte, Germany**

Professional experience

Since 11/2007	**University of Frankfurt, PhD** Institute of biophysical chemistry, Prof. Dr. V. Dötsch Thesis: "Structural and functional characterization of the acyl carrier protein triplet in the curacin cluster and its interaction partners"	**Frankfurt, Germany**
11/2009-12/2009	**University of Helsinki** Research stay at the Department of Biotechnology, Dr. H. Iwai Collaboration on segemntal labeling	**Helsinki, Finland**
01/2007– 10/2007	**Harvard Medical School** Diploma thesis in the Department of Cell Biology, Prof. Dr. F. McKeon Thesis: "Characterization of Astrotactin2 as p73 target"	**Boston, USA**

Alena Busche, Teichstraße 7, 60594 Frankfurt am Main
Cellphone: 0049 (0)176- 20496644, Email: alenabusche@gmx.de

06/2005 – 07/2005 **University of Amsterdam** **Amsterdam, Netherlands**
Internship at Swamerdame Institute for Life Science (SILS),
Dr. Remco Kort
Subject: Molecular Biology and Microbiology in Yeast

09/2004 - 12/2006 **University of Strasbourg** **Strasbourg, France**
University of Basel **Basel, Switzerland**
University of Freiburg **Freiburg, Germany**
University of Karlsruhe **Karlsruhe, Germany**
Internships of two months each, in

- Plant-biotechnology in Freiburg (Germany)
- Microbiology in Basel (Switzerland)
- Bio-engineering in Karlsruhe (Germany)
- Cell biology in Strasbourg (France)

Scholarships

08/2011 Conference scholarship, "XXIVth International Conference on Magnetic Resonance in Biological Systems (ICRMBS)"

01/2007 – 09/2007 Bourse d'Alsace

01/2007 – 09/2007 City of Strasbourg "bourse jumelles"

09/2005 – 09/2006 Baden Württemberg Fond

09/2004 - 09/2005 Erasmus

Additional Information

Languages German (native)
French (fluent)
English (fluent)

Hobbies: Climbing
Traveling (Europe, South-America, North-America, Lebanon, Australia)
Diving

List of publications

Busche AE, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Doetsch V, Iwai H: Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein**.** ***Angewandte Chemie (International ed*** **2009,** 48**(33):6128-6131.**

Busche AE, Gottstein D, Hein C, Ripin N, Pader I, Gu L, Walsh CT, Sherman DH, Löhr F, Güntert P, Dötsch V: Characterization of the interaction between an ACP domain and a halogenase in the curacin A polyketide synthetase. *submitted*

Structure of holo- ACP$_1$ from *Lyngbya majuscula*, *submitted* **accession code: 2liu**

Structure of HMG- ACP$_1$ from *Lyngbya majuscula, submitted* **accession code: 2liw**

Alena Busche, Teichstraße 7, 60594 Frankfurt am Main
Cellphone: 0049 (0)176- 20496644, Email: alenabusche@gmx.de

Talks

08/2010 **XXIVth International Conference on Magnetic Resonance in Biological Systems (ICRMBS)** **Cairns, Australia**
Talk: "Segmental Isotopic Labeling of a central domain in a multidomain Protein" Alena E.L. Busche, S.A. Aranko, F. Bernhard, V. Doetsch, H. Iwai

04/2010 **4th Center for Biomolecular Magnetic Resonance (BMRZ)- Cluster of Excellence (CEF) symposium** **Frankfurt, Germany**
Chair: „Magnetic resonance on Biomolecular complexes"

05/2008 **2nd Center for Biomolecular Magnetic Resonance (BMRZ)- Cluster of Excellence (CEF) symposium** **Frankfurt, Germany**
Talk: "Structural Dynamics of Peptidyl Carrier Domains in Nonribosomal Peptide S ynthetases (NRPS)"

Posters

08/2010 **Annual meeting of the European magnetic resonance Community (Euromar)** **Frankfurt, Germany**
Poster:"Characterization of the interaction between an ACP domain and a halogenase in the curacin A polyketide synthetase"

07/2009 **VIII European Symposium of the Protein Society** **Zuerich, Switzerland**
Poster:"Segmental Isotopic Labeling of a central domain by protein trans-splicing" Alena E.L. Busche, S.A. Aranko, C. Hein, M. Talebzadeh-Faroji, F. Bernhard, A. Koglin, C. T. Walsh,V. Doetsch, H. Iwai

08/2008 **EMBO course on Multidimensional NMR in Structural Biology** **Il Ciocco, Italy**
Poster: „Structural Dynamics of Peptidyl Carrier Domains in Nonribosomal Peptide Synthetases" Alexander Koglin, Mehdi Talebzadeh, Alena Busche, Frank Löhr, Frank Bernhard, Gerhard Wagner, Christopher T. Walsh , Mohamed A. Marahiel and Volker Dötsch