

**Wechselwirkungen zwischen Adenylierungs- und Peptidyl Carrier Protein-  
Domänen in nicht-ribosomalen Peptidsynthetasen sowie  
biochemische und strukturelle Untersuchungen zu gespaltenen Inteinen**

Dissertation  
zur  
Erlangung des Doktorgrades  
der Naturwissenschaften  
(Dr. rer. nat.)

der Fakultät Chemie  
der Technischen Universität Dortmund  
vorgelegt von

Joachim Zettler  
aus Dietzenbach

Dortmund 2010

Die vorliegende Arbeit wurde unter Betreuung von Professor Dr. Henning Mootz im Zeitraum von Oktober 2006 bis März 2010 an der Fakultät Chemie - Chemische Biologie der Technischen Universität Dortmund durchgeführt.

Erstgutachter: Prof. Dr. Henning D. Mootz

Zweitgutachter: Prof. Dr. Martin Engelhard

Tag der Abgabe:

Tag der Disputation:

## Zusammenfassung

Die aktuelle Antibiotikaforschung basiert trotz großer Fortschritte auf dem Gebiet der kombinatorischen, organischen Synthese immer noch zu einem großen Teil auf Naturstoffbibliotheken bakteriellen oder pilzlichen Ursprungs. Innerhalb dieser Naturstoffe sind nichtribosomale Peptide (NRPs) eine enorm wichtige bioaktive Substanzklasse. Neben antibiotischen Eigenschaften können sie weiterhin immunsuppressive, antifungale und tumorunterdrückende Aktivität entfalten. NRPs werden von Multi-Domänen Proteinen den so genannten nichtribosomalen Peptidsynthetasen (NRPS) in einer Art Fließbandmechanismus aufgebaut. Zwar sind die individuellen NRPS-Domänen biochemisch und strukturell sehr gut untersucht, jedoch sind die für das detaillierte Verständnis der Naturstoffbiosynthese essentiellen Wechselwirkungen zwischen den einzelnen Domänen sowohl strukturell, als auch mechanistisch noch kaum verstanden.

In dieser Arbeit gelang der erste experimentelle Nachweis von Interaktionen zwischen einer Adenylierungs- mit einer *in cis* vorliegenden Peptidyl Carrier Protein-Domäne aus einem NRPS Initiationsmodul. Mit biochemischen Methoden, wie partiellen tryptischen Verdauen, Gelfiltrationschromatographie und chemischen Markierungsassays, wurde festgestellt, dass der 4'-Phosphopantetheinarm der PCP-Domäne sich in unterschiedlichen Positionen während des Katalysezyklus befindet und zwar abhängig vom Zustand der Adenylierungs-Domäne. Die produktive Wechselwirkung zwischen den beiden Domänen benötigt sowohl eine post-translational modifizierte holo-Peptidyl Carrier Protein-Domäne, als auch eine spezielle Konformation der Adenylierungs-Domäne, die Thioesterkonformation. Die hier durchgeführten Studien charakterisieren erstmalig die großen konformationellen Bewegungen in NRPS-Proteinen und sollten den Startpunkt für ein tiefgreifendes Verständnis des antibiotika-produzierenden Proteintemplats bilden.

Für die Untersuchung von Proteinstrukturen und -dynamiken mit spektroskopischen Methoden, wie der NMR- oder der Fluoreszenzspektroskopie, ist die Modifikation des zu untersuchenden Proteins mit biophysikalischen Sonden oft unvermeidlich. Für den gezielten Einbau dieser Sonden in Proteine sind, unter anderem, gespaltene Inteine ein wichtiges Werkzeug der Chemischen Biologie. Die von gespaltenen Inteinen vermittelte Protein *trans*-Spleißreaktion verbindet die N- und C-terminal zu den individuellen Inteinhälften vorliegenden Polypeptidsequenzen, die N- und C-Exteine, über eine native Peptidbindung und erlaubt damit den definierten Aufbau eines Proteins aus mehreren Segmenten.

Mit dem natürlich gespaltenen DnaE Intein aus *Nostoc punctiforme* (*Npu*) wurde in dieser Arbeit das bisher schnellste *trans*-spleißende Intein biochemisch mit gereinigten Proteinen charakterisiert. Neben der hohen monomolekularen Reaktionsrate von  $1,1 \pm 0,2 \text{ s}^{-1}$  bei  $37^\circ\text{C}$  zeigte es mit verschiedenen Exteinsequenzen robuste Spleißausbeuten von 50 - 90 % in dem Temperaturbereich von 6 bis  $37^\circ\text{C}$ . Als ein Grund für verminderte Spleißausbeuten konnte die Bildung von löslichen Aggregaten der individuellen Inteinhälften nachgewiesen werden. NMR-Untersuchungen ergaben, dass sowohl die natürlich gespaltenen individuellen Hälften des *Npu* DnaE Inteins, als auch die des künstlich gespaltenen *Synechocystis* species PCC6803 DnaB Inteins keine definierte Faltung ohne das Partnerprotein aufwiesen. Für erstere konnte bei Komplexbildung der Hälften und nach der Spleißreaktion der Übergang in eine gefaltete Konformation nachgewiesen werden. Zusätzlich wurde mit dem SPLICEFINDER-System eine Methode entwickelt, die es schnell und unkompliziert erlaubt parallel mehrere Insertionspositionen von gespaltenen Inteinen in Zielproteinen auf ihre Spleißaktivität zu überprüfen. Für eine Anwendung dieses Systems, der segmentellen Isotopenmarkierung von Proteinen für NMR-Untersuchungen, konnte ein einfaches Verfahren zur Bestimmung der Markierungseffizienz etabliert werden. Es ist anzunehmen, dass die in dieser Arbeit durchgeführten Studien zu gespaltenen Inteinen deren Einsatzbreite für die selektive Modifikation von Proteinen signifikant erhöhen werden.

## Summary

Modern antibiotic research relies, despite great efforts in combinatorial organic synthesis, still strongly on compound libraries with bacterial or fungal origin. One of the most prominent bio-active substance classes consists of the so called nonribosomal peptides (NRPs). They can possess besides antibacterial properties various other activities, like immuno- or tumour-suppressive and antifungal activities. NRP-production is catalyzed by multidomain enzymes, the nonribosomal peptide synthetases (NRPS), in an assembly line fashion via a protein-template directed mechanism. In contrast to the well investigated and structurally characterized individual NRPS domains, the essential domain-interactions during the natural product biosynthesis remain largely elusive.

This work presents the first experimental evidence for conformational changes in the cross-talk between an adenylation and an *in cis* peptidyl-carrier protein domain from an NRPS initiation module. Using partial tryptic digests, gel filtration chromatography, native PAGE and chemical labeling experiments it could be shown, that the 4'-phosphopantethein moiety of the peptidyl-carrier protein domain changes its position as a result of a conformational change in the adenylation domain. The productive interaction between the two domains requires the post-translationally modified holo-form of the peptidyl-carrier protein domain and a distinct conformation of the adenylation domain, the thioester conformation. The conducted studies characterize for the first time the large conformational movements in NRPS proteins and are a good starting point for an in-depth understanding of the antibiotics-producing protein template.

Fluorescence and NMR spectroscopy serve as important tools for the elucidation of protein structure and dynamics. They often rely on the incorporation of biophysical markers into the protein of interest. Split inteins are one method out of the Chemical Biology's tools box to achieve these selective protein modifications. The split intein mediated protein *trans*-splicing reaction fuses the N- and C-terminal flanking regions of the intein, the N- and the C-extein, together through a native peptide bond. This ligation reaction therefore allows building up a protein from differently derived and/or treated segments.

The naturally split DnaE intein from *Nostoc punctiforme* (*Npu*) was characterized for the first time in this work with purified proteins *in vitro*. It surprisingly possessed the highest apparent first-order rate constant reported for the protein *trans*-splicing reaction so far -  $(1.1 \pm 0.2) \times 10^{-2} \text{ s}^{-1}$  at 37°C. Furthermore, the reaction was high-yielding (50 - 90 %) with respect to different extein sequences and in the temperature range from 6 to 37°C. The formation of soluble aggregates of the individual *Npu* DnaE intein halves was identified as one reason for splicing yields below completion. NMR investigation established that the naturally split *Npu* DnaE as well as the halves of the artificially split *Synechocystis* species PCC6803 DnaB intein do not exhibit one defined folded structure and that they are intrinsically disordered without their partner protein. For the *Npu* DnaE intein it also could be shown via NMR that the individual halves fold into an ordered structure after the protein *trans*-splicing reaction, as well as upon complex formation. Additionally, the SPLICEFINDER system, a PCR-based method for the fast and easy screening of active split intein insertions in any target protein was developed. It shortens significantly the tedious cloning procedures to obtain the desired intein fusion genes. For one application of the SPLICEFINDER method, the segmental isotopical labeling of proteins for NMR studies, a new and easy procedure to determine the labeling efficiency was established. It is based on tryptic in-gel digests with subsequent MALDI-TOF MS analysis. In conclusion, the conducted studies on split inteins are expected to promote and strengthen the general use of these ligation catalysts in the area of selective protein modification.

## Publikationen

Der Hauptteil dieser Arbeit wurde oder wird in den folgenden Publikationen veröffentlicht:

### **Joachim Zettler\*, Vivien Schütz\* and Henning D. Mootz (2009)**

The naturally split *Npu* DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction

*FEBS Letters*, **583**, 909-914. (\* equally contributing authors)

### **Joachim Zettler and Henning D. Mootz (2010)**

Biochemical evidence for conformational changes in the cross-talk between adenylation and peptidyl-carrier protein domains of nonribosomal peptide synthetases

*FEBS Journal*, **277**, 1159-1171.

### **Joachim Zettler\*, Tim Sonntag\*, Simone Eppmann, Alena Busche, Volker Dötsch and Henning D. Mootz**

A PCR-based method to determine the splice activity of split inteins in target proteins

*Manuscript in preparation*. (\* equally contributing authors)

Weitere Publikationen:

### **Christina Ludwig, Dirk Schwarzer, Joachim Zettler, Daniel Garbe, Petra Janning, Claus Czeslik and Henning D. Mootz (2009)**

Semisynthesis of proteins using split inteins

*Methods in Enzymology*, **Volume 462**, Chapter 4, 77-96.

---

# Inhaltsverzeichnis

Zusammenfassung.....	I
Summary.....	II
Publikationen.....	III
Inhaltsverzeichnis.....	1
Abkürzungen.....	3
1 Einleitung.....	6
1.1 Die nicht-ribosomale Peptidsynthese.....	9
1.1.1 Naturstoffprodukte der nicht-ribosomalen Peptidsynthese.....	9
1.1.2 Die nicht-ribosomalen Peptidsynthetasen.....	12
1.1.3 Gramicidin S Biosynthese.....	14
1.1.4 Die Adenylierungsdomänen.....	18
1.1.5 Die PCP-Domänen.....	24
1.1.6 Domänenanordnung innerhalb eines Moduls.....	30
1.2 Inteine.....	35
1.2.1 Der Mechanismus der Protein-Spleißreaktion.....	39
1.2.2 Abhängigkeit der Protein-Spleißreaktion von heterologen Exteinsequenzen..	41
1.2.3 Gespaltene Inteine.....	42
1.2.4 Biochemische Anwendungen gespaltener Inteine.....	45
1.3 Aufgabenstellung.....	51
2 Material.....	52
2.1 Geräte und Apparaturen.....	52
2.2 Chemikalien, Enzyme und Verbrauchsmaterialien.....	53
2.3 Vektoren.....	54
2.4 Mikroorganismen.....	56
2.4.1 <i>E. coli</i> Top10.....	56
2.4.2 <i>E. coli</i> BL21 Gold (DE3).....	56
2.5 Medien.....	56
2.6 Puffer und Lösungen.....	57
3 Methoden.....	60
3.1 Konstruktion der Plasmide.....	60
3.2 Expression und Proteinreinigung.....	66
3.2.1 Heterologe Expression in <i>E. coli</i> Zellen.....	66
3.2.2 Zellaufschluss.....	67
3.2.3 Proteinreinigung.....	68
3.3 Biochemische Untersuchungen von NRPS-Proteinen.....	68
3.3.1 Partielle tryptische Verdaue von NRPS-Fragmenten.....	68
3.3.2 Native PAGE- und Gelfiltrations-Assays.....	69
3.3.3 Chemisches Markieren des Ppant-Arms.....	69
3.4 Kinetische Analyse der Protein trans-Spleißreaktion des Npu DnaE Inteins.....	70
3.4.1 Durchführung der PTS-Reaktion.....	70
3.4.2 Auswertung der kinetischen Daten.....	70
3.5 Die SPLICEFINDER-Methode.....	73
3.5.1 Testexpressionen.....	73
3.5.2 Western Blot Analyse.....	73
3.5.3 Präparation segmentell isotopenmarkierter ST-gpD-Trx-H <sub>6</sub> Proteine.....	74
3.5.4 Bestimmung der Markierungseffizienz von segmentell isotopenmarkierten Proteinen.....	75
3.6 Modifikationen von NRPS-Proteinen mit Fluorophoren.....	76

3.6.1	Semi-synthetisches PTS mit dem an Position 11 gespaltenen <i>Ssp</i> DnaB Intein..	76
3.6.2	Modifikation der Cysteinmutanten von GrsA A-PCP.....	76
3.6.3	Sfp-katalysierte Modifikation der PCP-Domäne .....	77
3.7	Allgemeine Techniken .....	77
3.7.1	Präparative Herstellung von holo-A-PCP .....	77
3.7.2	Photometrische Bestimmung der Proteinkonzentration.....	77
3.7.3	Massenspektrometrische Analyse von SDS-PAGE Gel-Proteinbanden.....	78
3.7.4	Festphasenpeptidsynthese .....	78
3.7.5	NMR-Spektroskopie.....	80
4	Ergebnisse .....	84
4.1	Wechselwirkungen zwischen NRPS A- und PCP-Domänen.....	84
4.1.1	Der Einfluss der Sulfamoyl-Inhibitoren auf die Konformation von GrsA A-PCP im Vergleich mit den Substraten ATP und L-Phe.....	86
4.1.2	Globale Konformationsanalyse von A-PCP Didomänenproteinen.....	91
4.1.3	Untersuchung der Position des Ppant-Arms in A-PCP Proteinen.....	94
4.2	Biochemische und strukturelle Untersuchungen zu gespaltenen Inteinen.....	102
4.2.1	Biochemische Charakterisierung des natürlich gespaltenen <i>Npu</i> DnaE Inteins ..	102
4.2.2	NMR-Untersuchungen von gespaltenen Inteinen .....	111
4.3	SPLICEFINDER - Schnelle Klonierung und Aktivitätsbestimmung von gespaltenen Inteinen in Zielproteinen .....	119
4.3.1	Prinzip des SPLICEFINDER-Systems.....	120
4.3.2	Das Modellprotein gpD-Trx.....	121
4.3.3	Überprüfung der segmentellen Isotopenmarkierung bei <i>in vivo</i> Expressionen....	127
4.4	Vorarbeiten zu spektroskopischen Untersuchungen von NRPS-Proteinen.....	133
4.4.1	Fluoreszenzspektroskopie .....	133
4.4.2	NMR-Spektroskopie.....	138
5	Diskussion und Ausblick.....	143
5.1	Dynamische Eigenschaften von Proteinen und Protein-Ligand Komplexen.....	143
5.1.1	Konformationelle Änderungen und Domäneninteraktionen in NRPS Proteinen.	146
5.1.2	Zukünftige Arbeiten an NRPS-Proteinen.....	150
5.2	Neue pharmazeutische Wirkstoffe durch Manipulation der NRPS-Biosynthesewege	153
5.3	Gespaltene Inteine zur selektiven Modifikation von Proteinen .....	159
5.3.1	Gespaltene Inteinhälften als Beispiele für nativ unstrukturierte Proteine.....	160
5.3.2	Die Geschwindigkeit der PTS-Reaktion .....	163
5.3.3	Exteinabhängigkeit.....	165
	Literatur.....	169
	Anhang .....	178
	Danksagung.....	187

---

## Abkürzungen

4-CBL	4-Chlorbenzoat CoA Ligase
ACP	engl. <i>Acyl carrier protein</i>
AcpS	engl. <i>Acyl carrier protein synthase</i>
A <sup>C</sup> -Subdomäne	C-terminale Subdomäne einer NRPS A-Domäne
ACV	δ-(α-Aminoadipyl)-L-cysteinyl-D-Valin
A-Domäne	Adenylierungs-Domäne in NRPS-Systemen
AMP	Adenosinmonophosphat
A <sup>N</sup> -Subdomäne	N-terminale Subdomäne einer NRPS A-Domäne
ANL	Proteinsuperfamilie der Acyl- und Aryl-CoA Synthetasen, NRPS A-Domänen und Luciferasen
ANS	1-Anilin-8-Naphthalen-sulfonat
ApCpp	Adenosin-5'-f-triphosphat
ArCP	engl. <i>Aryl Carrier Protein</i>
AS	Aminosäure(n)
AT	Acetyltransferase
ATP	Adenosintriphosphat
BA	Bromacetamid
ber.	berechnet
bp	Basenpaare
<i>bla</i>	β-Lactamase Gen
BSA	Rinderserumalbumin (engl. <i>bovine serum albumin</i> )
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
cAMP	3'-5'-zyklisches AMP
CBD	chitinbindende Domäne
CD	zirkulärer Dichroismus
C-Domäne	Kondensations-Domäne in NRPS-Systemen
CHCA	α-Cyano-4-hydroxizimtsäure
CIAP	Alkalische Phosphatase (engl. <i>calf intestine alkaline phosphatase</i> )
CoA	Coenzym A
CP	engl. <i>Carrier Protein</i>
CSA	chemische Verschiebungsanisotropie (engl. <i>chemical shift anisotropy</i> )
Da	Dalton
ddNTP	2', 3'-Didesoxynukleosid-5'-triphosphat
DH	Dehydratase
DHB	2,5-Dihydroxybenzoesäure
DIPEA	N,N-Diisopropylethylamin
DltA	D-Alanin-D-Alanyl Carrier Protein Ligase
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure
DnaB Intein	Intein aus der DNA-Helikase
DnaE Intein	Intein aus der α-Untereinheit der DNA Polymerase III
DTT	1,4-Dithio-D,L-threitol
<i>E. coli</i>	<i>Escherichia coli</i>
E-Domäne	Epimerisierungs-Domäne in NRPS-Systemen
EDT	1,2-Ethandithiol
EDTA	Ethylendiamintetraessigsäure

## Abkürzungen

---

eGFP	verbessertes grün-fluoreszierendes Protein (engl. <i>enhanced green fluorescent protein</i> )
EPL	engl. <i>Expressed Protein Ligation</i>
eq.	Äquivalente
ER	β-Enoylreduktase
FAS	Fettsäuresynthase
FID	engl. <i>free induction decay</i>
FKBP	FK506 bindendes Protein
Fl	Fluorescein
Fmoc	Fluorenyl-9-methyloxycarbonyl
FRET	Fluoreszenz-Resonanz-Energie-Transfer
gem.	gemessen
GFP	grün-fluoreszierendes Protein
Gl.	Gleichung
GOI	Gen von Interesse (engl. <i>gen of interest</i> )
gpD	Kopfprotein D des Bakteriophagen λ
GrsA	Gramicidin S Synthetase I
GrsB	Gramicidin S Synthetase II
GyrA	DNA Gyrase A
GyrB	DNA Gyrase B
HEPES	2-N'-[N-(2-Hydroxyethyl)-piperazinyl]-ethansulfonsäure
HINT	Hedgehog- und Intein-Proteinfamilie
H <sub>6</sub>	Hexahistidinsequenz
HPLC	Hochdruckflüssigkeitschromatographie (engl. <i>high performance liquid chromatography</i> )
Int <sup>C</sup>	C-terminale Hälfte eines gespaltenen Inteins
Int <sup>N</sup>	N-terminale Hälfte eines gespaltenen Inteins
IPTG	Isopropyl-β-D-thiogalactosid
<i>lacI</i>	Gen des Lac-Repressors
LB-Medium	Luri-Bertani-Medium
K <sub>D</sub>	Dissoziationskonstante
KR	β-Ketoacylreduktase
KS	β-Ketoacylsynthase
M	Mol pro Liter
MALDI	Matrix-unterstützte Laser-Desorptions Ionisierung (engl. <i>matrix assisted laser desorption ionisation</i> )
MAT	Malonyl-CoA/acetyl-CoA-ACP Transacylase
MBP	Maltose bindendes Protein
MCS	Multiple Klonierungsstelle (engl. <i>multiple cloning site</i> )
MDR	Multimedikamenten-resistent
MESNA	Natriumsalz der Mercaptoethansulfonsäure
MG	engl. <i>molten globule</i>
min	Minuten
mRNA	Boten-Ribonukleinsäure (engl. <i>messenger ribonucleic acid</i> )
MRSA	Methicillin resistente <i>Staphylococcus aureus</i>
MS	Massenspektrometrie
Mtt	Methyl-triphenylmethyl
<i>Mtu</i>	<i>Mycobakterium tuberculosis</i>
MW	Molekulargewicht (engl. <i>molecular weight</i> )
MWCO	Molmassenausschlussgröße (engl. <i>molecular weight cut off</i> )
<i>Mxe</i>	<i>Mycobakterium xenopi</i>

## Abkürzungen

---

<i>m/z</i>	Verhältnis Masse zu Ladung
NCL	native chemische Ligation
<i>Neq</i>	<i>Nanoarchaeum equitans</i>
NMR	Kernspinresonanz (engl. <i>nuclear magnetic resonance</i> )
<i>Npu</i>	<i>Nostoc punctiforme</i>
NRP	Nicht-ribosomal synthetisierte Peptide
NRPS	Nicht-ribosomale Peptidsynthetasen
NTA	Nitrilotriacetat
OD	Optische Dichte
ORF	offenes Leseraster (engl. <i>open reading frame</i> )
PAGE	Polyacrylamid-Gelelektrophorese
PCP-Domäne	Peptidyl Carrier Protein-Domäne in NRPS-Systemen
PCR	Polymerase-Kettenreaktion (engl. <i>polymerase chain reaction</i> )
PDB	Proteindatenbank
PKS	Polyketidsynthase
POI	Protein von Interesse (engl. <i>protein of interest</i> )
Ppant	4'-Phosphopantetheinyl
PP <sub>i</sub>	Pyrophosphat
PPTase	4'-Phosphopantetheinyltransferase
PTS	Protein <i>trans</i> -Spleißen
RC	engl. <i>random coil</i>
RNA	Ribonukleinsäure
rpm	Umdrehungen pro Minute
<i>Sc</i>	<i>Saccharomyces cerevisiae</i>
SDS	Natriumdodecylsulfat (engl. <i>sodium dodecylsulfate</i> )
Sfp	4'-Phosphopantetheinyltransferase aus dem Biosyntheseweg von Surfactin in <i>Bacillus subtilis</i>
sm	Einzelmolekül (engl. <i>single molecule</i> )
SPPS	Festphasenpeptidsynthese (engl. <i>solid phase peptid synthesis</i> )
<i>Ssp</i>	<i>Synechocystis</i> sp. Stamm PCC6803
ST	StrepTag II (Sequenz WSHPQFEK)
t <sub>1/2</sub>	Halbwertszeit
TAMRA	Tetramethylrhodamin
TCEP	Tris-(2-carboxyethyl)phosphin
TE-Domäne	Thioesterase-Domäne in NRPS-Systemen
TFA	Trifluoressigsäure
TIS	Triisopropylsilan
TOF	Flugzeit (engl. <i>time of flight</i> )
TR	TexasRed
Tris	Tris(hydroxymethyl)-aminomethan
tRNA	Transfer-Ribonukleinsäure
TROSY	engl. <i>transverse relaxation-optimised spectroscopy</i>
Trx	Thioredoxin
UV	Ultraviolettes Licht
v/v	Volumen pro Volumen
VMA	Vakuoläre ATPase
VRSA	Vancomycin resistente <i>Staphylococcus aureus</i>
w/v	Gewicht pro Volumen (engl. <i>weight per volume</i> )
w/w	Gewicht pro Gewicht (engl. <i>weight per weight</i> )

---

# 1 Einleitung

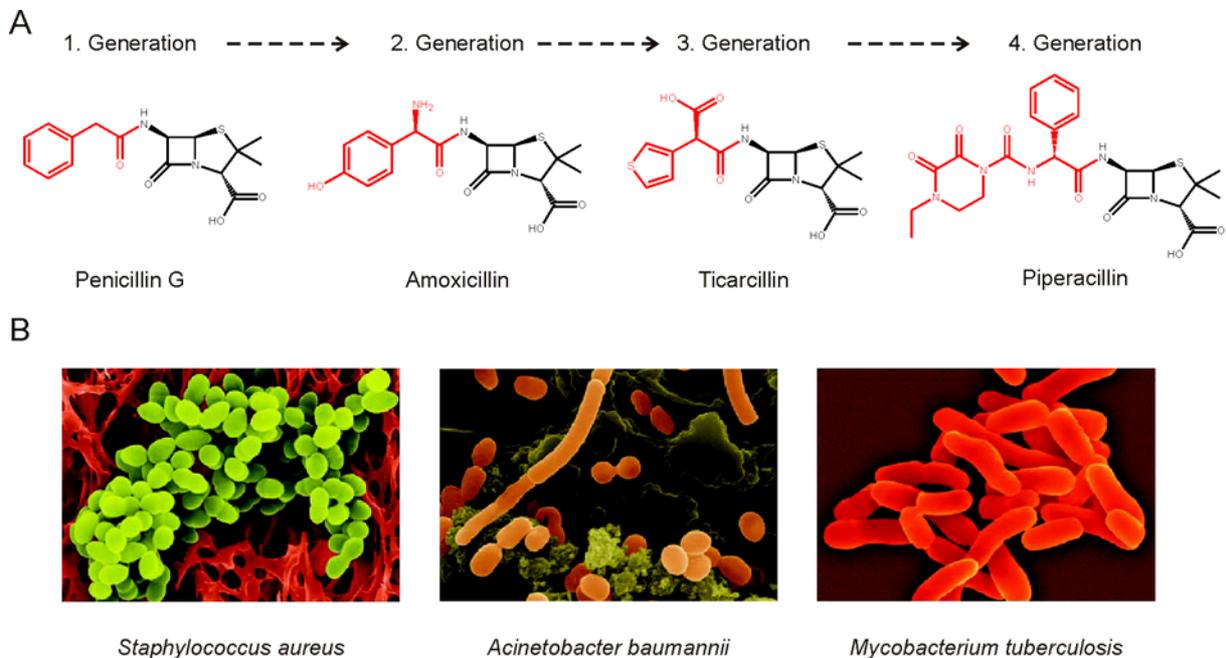
## Ein Überblick über die aktuelle Antibiotikaforschung

Durch Mikroorganismen ausgelöste Infektionskrankheiten sind weltweit die zweithäufigste Todesursache [1]. Zur Behandlung dieser Infektionskrankheiten werden seit Anfang der 1940er Jahre Antibiotika eingesetzt. Sie unterbinden die Ausbreitung (bakteriostatisch) oder bewirken das Absterben (bakteriozid) der pathogenen Mikroorganismen an den Infektionsherden. Ein klinisch erfolgreich einsetzbares Antibiotikum muss nicht nur effektiv die schädlichen Mikroorganismen bekämpfen, sondern sollte auch möglichst keine Nebenwirkungen beim Patienten auslösen. Somit können Antibiotika auf zellulärer Ebene nur Ziele angreifen, bei denen größere biochemische und strukturelle Unterschiede zwischen den Prokaryoten und den Eukaryoten bestehen. Die Angriffspunkte der meisten Antibiotika sind vor allem die prokaryotische Proteinbiosynthese, die DNA Replikation und Reparatur, sowie die Zellwandbiosynthese. Zum Beispiel inhibiert die Klasse der  $\beta$ -Lactame (unter anderem die Penicilline) die Enzyme, die für die Quervernetzung der Peptidoglykan-Zellwand verantwortlich sind. Vancomycin bindet an das Substrat dieser Quervernetzung und verhindert damit ebenfalls einen effektiven Aufbau einer stabilen Zellwand. Beide Substanzen resultieren in einer erleichterten Lyse des Bakteriums.

Der vermehrte und intensive Einsatz eines Antibiotikums führt zu einer dramatischen Erhöhung des Selektionsdrucks auf die Entwicklung von Resistenzen. Klinisch signifikante Resistenzen treten im Abstand von Monaten oder Jahren auf [2]. Für Penicillin G wurden die ersten resistenten Stämme innerhalb von zwei Jahren nach der Einführung des Medikaments beobachtet. Bakterien inaktivieren Antibiotika mit unterschiedlichen Mechanismen. So pumpen, z.B., spezielle Membranproteine eingedrungene Antibiotikamoleküle aktiv aus der Zelle heraus und verhindern so eine effektive, toxische Konzentration im Zellinnern. Eine verminderte Affinität des Antibiotikums und damit eine geringere Inhibierung kann durch eine Veränderung der Zielstruktur erreicht werden. Ein anderer Mechanismus betrifft die direkte Modifikation des Antibiotikums und damit dessen Inaktivierung. So sind Penicillin G resistente Stämme mit Hilfe von  $\beta$ -Lactamasen in der Lage den  $\beta$ -Lactamring des Moleküls zu hydrolysieren und es damit unschädlich zu machen.

Ein wichtiger Zweig in der Entwicklung neuer und wieder effektiver Antibiotika ist die Veränderung von bereits aktiven Grundgerüsten [3]. Diese Leitstrukturen basierten bis zur Jahrtausendwende hauptsächlich auf Naturstoffen, vor allem auf Stoffwechselprodukten von Bakterien und Pilzen. Exemplarisch soll die Entwicklung neuer Antibiotika am Beispiel der

Penicilline dargestellt werden [3]. Nach dem Auftreten von Resistenzen gegen Penicillin G (s. oben) wurden Stoffe entwickelt, die immer noch das gleiche Grundgerüst und damit den identischen Wirkmechanismus, wie das ursprünglichen Molekül besaßen, jedoch nicht mehr von den modifizierenden Enzymen der Bakterien erkannt wurden (s. Abbildung 1A). Ein verbreiteter Einsatz der Antibiotika der 2. Generation führte auf der Bakterienseite zu Mutationen in dem Gen der  $\beta$ -Lactamase, die schließlich auch die neuen Wirkstoffmoleküle enzymatisch inaktivieren konnte und damit den Bakterien Resistenz verlieh. Weitere veränderte Dekorationen am Penicillingrundgerüst brachten die Antibiotika der 3. Generation hervor (s. Abbildung 1A). Ein verstärkter Selektionsdruck bewirkte jedoch auch hier zwangsläufig die Ausbildung von Resistenzen.



**Abbildung 1 Antibiotika und ihre Zielorganismen.** A) Entwicklung von neuen Antibiotika ausgehend von einer Naturstoffleitstruktur (schwarz) am Beispiel der Penicilline; B) Beispiele für resistente Stämme von bakteriellen Pathogenen (Bilder entnommen aus [3]).

Die aktuelle Antibiotikaforschung sieht sich dem vermehrten Auftreten von multiresistenten Stämmen gegenüber, die sich als Folge von Mutationen oder Aufnahme resistenzvermittelnder DNA gebildet haben (s. Abbildung 1B). Methicillin- oder Vancomycin-resistente *Staphylococcus aureus* (MRSA oder VRSA) verursachen eine hohe Mortalitätsrate bei Klinikpatienten [4]. Multiresistente (MDR) oder panresistente (PDR) gram-negative Bakterien, wie z.B. *Acinetobacter baumannii*, sprechen auf eine Vielzahl von Antibiotikaklassen (Penicilline, Cephalosporine, Carbapeneme, Quinolone, Aminoglykoside, Tetracycline und Polymyxine) nicht mehr an [5]. Des Weiteren sind MDR-Stämme von *Mycobacterium tuberculosis* in den Entwicklungsländern auf dem Vormarsch [6]. Es ist

anzunehmen, dass eine zukünftige Bekämpfung von Infektionskrankheiten weiteren großen Forschungseinsatz verlangt und dass nur eine Kombination mehrerer unterschiedlicher Strategien erfolgreich sein kann. Die Verbesserung bekannter Antibiotikagrundgerüste wird weiterhin ein wichtiges Standbein der Forschung bleiben, jedoch sollte auch die voranschreitende Sequenzierung von bakteriellen Genomen die Identifikation von neuen Zielstrukturen in diesen Pathogenen erlauben. Das Testen großer Substanzbibliotheken gegen diese neu definierten Zielstrukturen ermöglichte in den Jahren nach der Jahrtausendwende das Auffinden neuer antibiotischer Leitverbindungen (Oxazolidinone, Lipopeptide, Mutilone) und scheint auch in Zukunft Erfolg versprechend zu sein. In der Therapie geht der Trend weg vom übermäßigen Gebrauch von Breitbandantibiotika, weil mit deren Einsatz eine schnelle Resistenzbildung begünstigt wird. Durch eine verbesserte Diagnose der Infektionserreger in einzelnen Patienten könnten selektivere Antibiotika einsetzbar werden. Diese würden aufgrund ihrer limitierten Einsetzbarkeit erst in längeren Zyklen die Entwicklung von Resistenzen begünstigen [1].

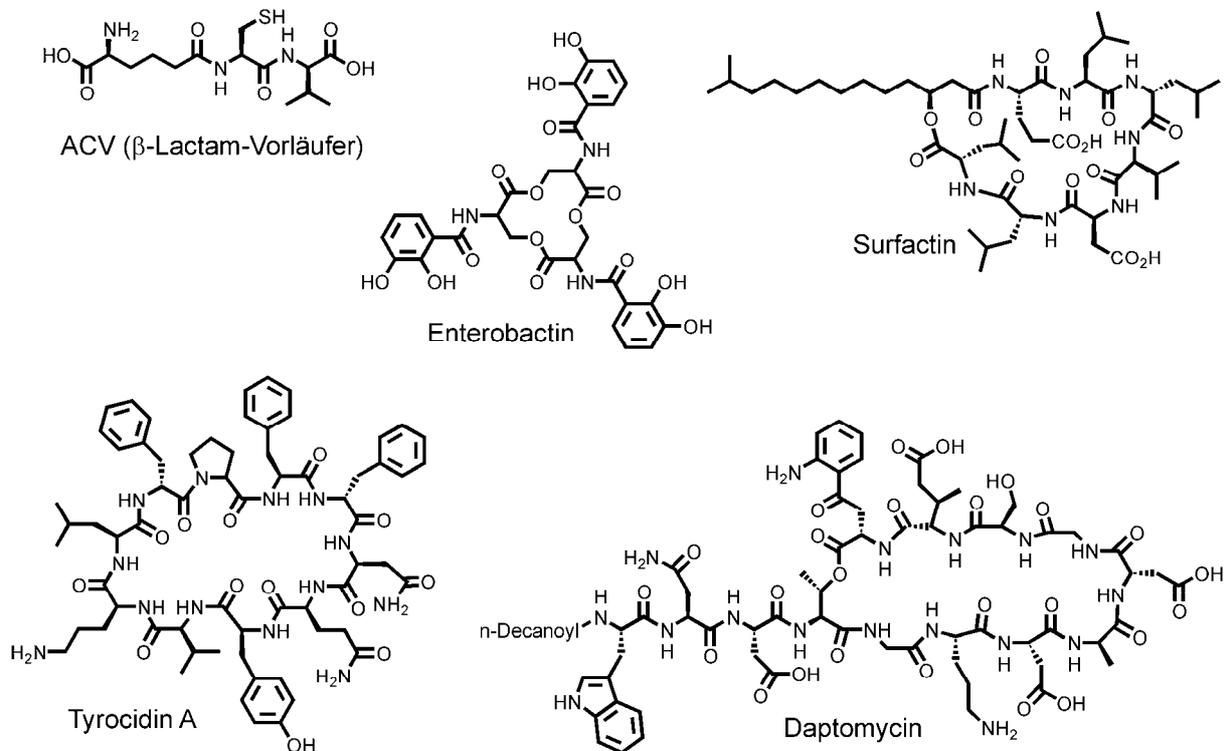
Eine Vielzahl von Peptidantibiotika wird von so genannten nicht-ribosomalen Peptidsynthetasen (NRPS) in Bakterien und Pilzen produziert. Unter diesen Sekundärmetaboliten findet man den Penicillin- und Cephalosporin-Vorläufer  $\delta$ -( $\alpha$ -Amino adipyl)-L-cysteinyl-D-Valin (ACV) [7], Tyrocidin A [8], Gramacidin S [9,10] und Daptomycin [11]. In der vorliegenden Arbeit wurden Untersuchungen zu Protein-Protein Wechselwirkungen in NRPS-Fragmenten durchgeführt. Sie dienen nicht nur dem reinen Verständnis dieser komplexen biochemischen Synthesemaschinerien, sondern ermöglichen vielleicht in Zukunft auch eine bessere Umprogrammierung derselben. Dies könnte eine effektive Quelle für die Herstellung neuer Substanzbibliotheken mit möglicherweise neuen, aktiven und klinisch einsetzbaren Antibiotika sein. In Folge wird in diesem Kapitel die nicht-ribosomale Peptidsynthese, mit einem Schwerpunkt auf den für diese Arbeit relevanten NRPS-Domänen, der Adenylierungs- und der Peptidyl Carrier Protein-Domäne, vorgestellt.

Detaillierte spektroskopische Untersuchungen zu Dynamiken von großen Proteinkomplexen, wie z.B. den in dieser Arbeit untersuchten NRPS-Proteinen, erfordern eine selektive Modifikation des Proteingerüsts mit biophysikalischen Sonden. Gespaltene Inteine sind ein wichtiges Werkzeug um diese gezielten Veränderungen an Proteinen zu erreichen. Ihnen ist der zweite Teil dieses Kapitels gewidmet.

## 1.1 Die nicht-ribosomale Peptidsynthese

### 1.1.1 Naturstoffprodukte der nicht-ribosomalen Peptidsynthese

Nicht-ribosomal synthetisierte Peptide (NRP) sind eine große Klasse von Naturstoffen bakteriellen und pilzlichen Ursprungs (einige Beispiele sind in Abbildung 2 dargestellt). Wie schon vom Namen abgeleitet werden kann, werden sie in einer alternativen Syntheseroute, die nicht mit der ribosomalen Peptid- und Proteinbiosynthese verwandt ist, hergestellt. Zwar erfolgt auch die Synthese der meist 3 bis 22 Aminosäuren großen Peptide vom N- zum C-Terminus, jedoch enden hier die Gemeinsamkeiten der beiden Synthesewege. Die Auswahl der Monomerbausteine ist bei der nicht-ribosomalen Peptidsynthese nicht auf die zwanzig proteinogenen Aminosäuren beschränkt. Neben Enantiomeren der proteinogenen Aminosäuren und unnatürlichen Aminosäuren finden sich auch  $\alpha$ -Hydroxy- und Carbonsäuren in den NRPs wieder. Weitere Dekorationen des Peptidgerüsts entstehen z. B. durch *O*-, *N*- und *C*-Methylierung, Acylierung, Formylierung, sowie Glykosylierung und Phosphorylierung.

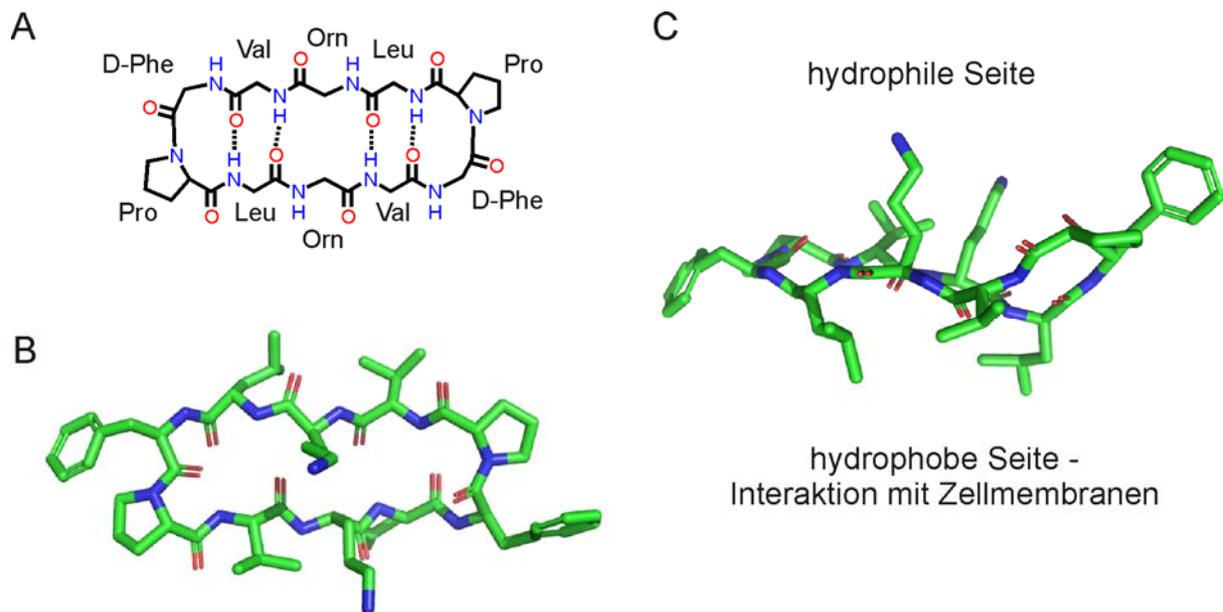


**Abbildung 2 Nicht-ribosomal hergestellte Naturstoffe.** Dargestellt sind  $\delta$ -( $\alpha$ -Aminoadipyl)-L-cysteinyl-D-Valin (ACV) (*Penicillium chrysogenum*), Enterobactin (*Escherichia coli*), Surfactin (*Bacillus subtilis*), Tyrocidin A (*Bacillus brevis*) und Daptomycin (*Streptomyces roseosporus*).

Ein weiteres Merkmal vieler NRPs ist ihre eingeschränkte konformationelle Flexibilität, die oft über Zyklisierung oder verzweigte Zyklisierung erreicht wird. Ihre rigide Struktur schützt die NRPs vor der hydrolytischen Spaltung durch Proteasen und ermöglicht ihnen wahrscheinlich die präzise Bindung an ihre Zielstrukturen und dadurch erst die biologische Aktivität [12]. Die Makrozyklisierung kann über Lakton-, Laktam-, Iminbildung oder über Quervernetzung erfolgen [13]. Die Eigenschaften der NRPs beschränken sich nicht nur auf die bereits oben erwähnten antibiotischen Aktivitäten. So findet man in dieser Naturstoffklasse Immunsuppressiva, wie Cyclosporin A [14] oder das als Biotensid wirkende und hämolytische Aktivität besitzende Surfactin [15]. Enterobactin [16] und Bacillibactin [17] sind Siderophore, die unter Eisenmangelbedingungen vom Produzentenstamm hergestellt und sekretiert werden [18,19].

Die in dieser Arbeit untersuchten NRPS-Proteinfragmente stammten aus dem Bodenbakterium *Bacillus brevis* (aktuelle Nomenklatur *Aneurinibacillus migulans*) und sind dort für die Biosynthese des kationischen, zyklischen Dekapeptids Gramacidin S verantwortlich [9,10]. Gramacidin S ist aus zwei identischen Pentapeptiden aufgebaut und besitzt die Primärstruktur cyclo-(D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro-Val-Orn-Leu). NMR Untersuchungen zeigten, dass Gramacidin S eine antiparallele  $\beta$ -Faltblattstruktur mit zwei  $\beta$ -Turns einnimmt [20]. Das Molekül ist amphiphil mit einer polaren Seite bestehend aus den kationischen Ornithin- und den D-Phenylalaninseitenketten. Die andere Seite des  $\beta$ -Faltblatts zeigt mit den Valin- und Leucinresten einen stark hydrophoben Charakter (s. Abbildung 3).

Gramacidin S ist ein potentes Breitbandantibiotikum, das sowohl gegen gram-positive und gram-negative Bakterien, als auch pathogene Pilze in geringen mikromolaren Konzentrationen inhibierend wirkt [21,22]. Die klinische Einsetzbarkeit ist jedoch begrenzt, da Gramacidin S ebenfalls die Lyse von roten Blutkörperchen induziert. Sowohl die hämolytische, als auch die antibiotische Wirkung gehen auf die destabilisierenden Eigenschaften von Gramacidin S auf Lipiddoppelschichten zurück. Es erhöht die Durchlässigkeit der Membran und zerstört deren Integrität [23-26]. Jedoch scheint das Antibiotikum ebenfalls ein Inhibitor für die Cytochrom *bd*-Typ Quinoloxidase in *E. coli* zu sein, so dass nicht ausgeschlossen werden kann, dass Gramacidin S seine antibiotische Wirkung nicht nur an der Lipiddoppelschicht, sondern auch am Proteinanteil der Membran entfaltet [24,27,28].



**Abbildung 3 Das Peptidantibiotikum Gramacidin S.** A) Schematische Struktur des Peptidrückgrats - die vier stabilisierenden Wasserstoffbrücken der  $\beta$ -Faltblattstruktur sind eingezeichnet; B) Struktur von Gramacidin S (entnommen aus der PDB-Datei 1TK2, Gramacidin S im Komplex mit der alkalischen Proteinase Savinase, Bhatt, V.S. *et al.*, zu veröffentlichen); C) Die Seitenansicht verdeutlicht den amphiphilen Charakter des Dekapeptids. Die fast planare Ringstruktur des Peptidrückgrats besitzt eine hydrophile (zwei kationische Ornithinreste) und eine hydrophobe Seite (Valin- und Leucinseitenketten). Über letztere kann Gramacidin S wahrscheinlich seine membran-destabilisierenden Eigenschaften entfalten (die Abbildung wurde mit dem Programm Pymol erstellt).

Lange Zeit wurde angenommen, dass eine Trennung der hämolytischen von der antibiotischen Wirkung bei der Klasse der kationischen antimikrobiellen Peptide nicht möglich ist. In den letzten Jahren wurde jedoch die Synthese von Gramacidin S analogen Substanzen berichtet, die trotz guter Wirksamkeit gegen Bakterien und Pilze rote Blutkörperchen intakt ließen. Diese Analoga zeichnen sich entweder durch Modifikationen an den Prolinresten aus [29] oder sind zyklische 14er Peptide, die ebenfalls eine  $\beta$ -Faltblattstruktur mit zwei  $\beta$ -Turns aufweisen [26,30]. Die Selektivität dieser kationischen Peptide wird wahrscheinlich durch den höheren Anteil anionischer Lipide in bakteriellen im Vergleich zu menschlichen Membranen erreicht. Auch schwächt Cholesterin in letzteren den lytischen Effekt von Gramacidin S und seiner Analoga ab [31]. Die Rückbesinnung auf schon lange bekannte, aktive Naturstoffe, die aus unterschiedlichen Gründen zum Zeitpunkt ihrer Entdeckung nicht als Therapeutika in Frage kamen, könnte damit ebenfalls ein wichtiges Standbein der zukünftigen Antibiotikaforschung werden. In Bezug auf Gramacidin S und seine Analoga ist noch anzumerken, dass aufgrund ihres Wirkmechanismus (Destabilisierung der bakteriellen Membran) eine Resistenzbildung möglicherweise erst verlangsamt eintritt.

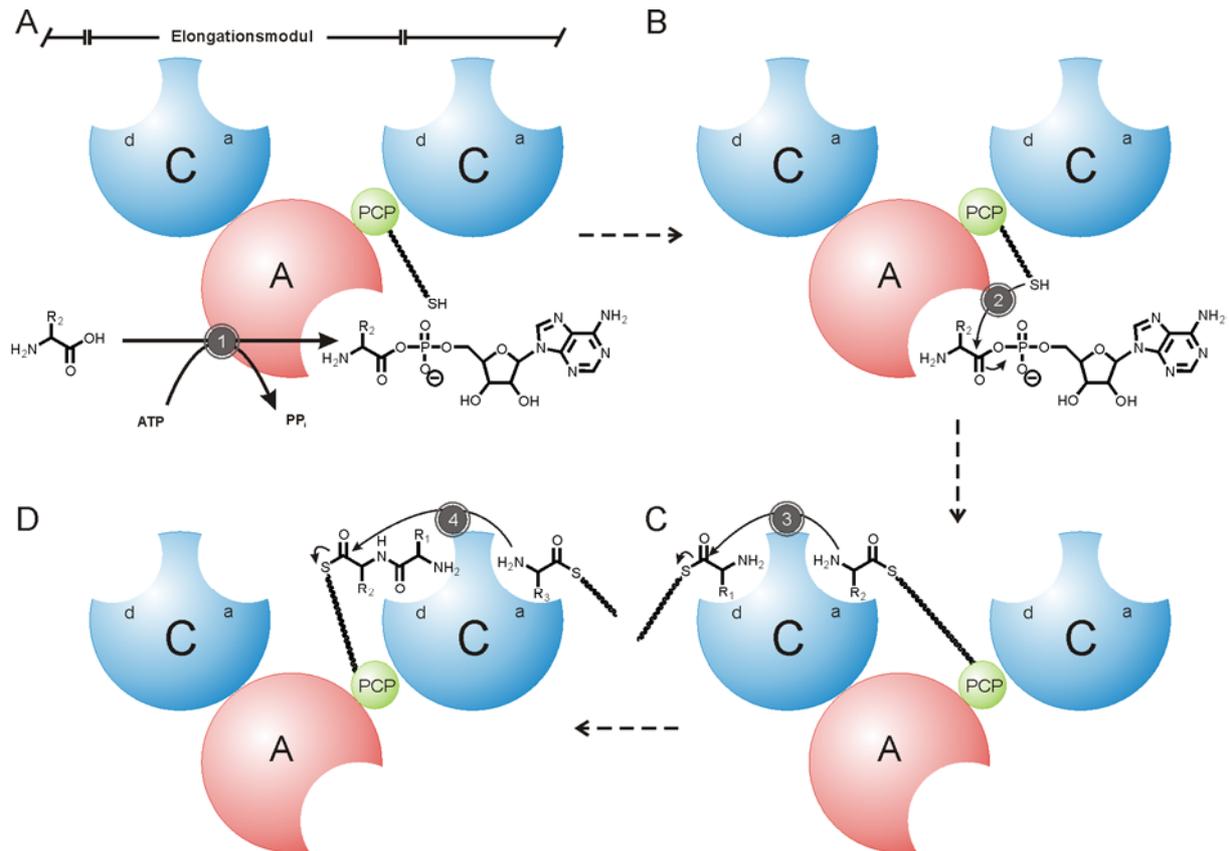
### 1.1.2 Die nicht-ribosomalen Peptidsynthetasen

Multi-Enzym-Komplexe, die so genannten nicht-ribosomalen Peptidsynthetasen (NRPS), sind für die Bildung der NRP in Bakterien und Pilzen verantwortlich. Sie sind modular aufgebaut, d.h. jedes Modul ist für den Einbau eines Monomerbausteins in das Peptidgerüst verantwortlich. Analog zu einer Fließbandproduktion aktiviert ein Modul einen spezifischen Monomerbaustein, bindet ihn, modifiziert ihn eventuell und verknüpft ihn mit dem Substrat des nächsten Moduls. So läuft die wachsende Kette auf dem Fließband der NRPS von einem Modul zum nächsten und wird dabei jeweils um einen Baustein verlängert. Wird bei der Synthese des NRP jedes Modul nur einmal benutzt, so ist es möglich, anhand der Anzahl und Anordnung der Module direkt auf die Primärstruktur des Naturstoffes zu schließen. Man spricht in diesem Fall von linearen NRPS-Systemen. In iterativen NRPS-Systemen wird das Proteintemplat mehrmals für die Bildung eines NRPs benutzt. Bei nicht-linearen Systemen ist die Abhängigkeit zwischen Modulstruktur und Peptidprodukt komplexer [32].

NRPS-Module lassen sich von der Biosyntheselogik in noch kleinere funktionelle Einheiten, die Domänen, einteilen. Ein Initiationsmodul am Anfang der Naturstoffsynthese besteht aus einer Adenylierungsdomäne (A-Domäne) und einer Peptidyl Carrier Protein-Domäne (PCP-Domäne, früher auch Thiolierungs- (T) Domäne genannt). In Elongationsmodulen ist zusätzlich eine Kondensationsdomäne (C-Domäne) vorhanden, so dass sich die Reihenfolge C-A-PCP in diesen Modulen ergibt. Die nicht-ribosomale Peptidsynthese wird in den meisten Fällen mit einer Thioesterasedomäne (TE-Domäne) beendet. Terminationsmodule besitzen damit den folgenden Domänenaufbau: C-A-PCP-TE. Zusätzliche Domänen zuständig für z.B. die Epimerisierung, Formylierung, Zyklisierung, Reduktion, Oxidation, Hydroxylierung [33] oder Methylierung der Substrate können innerhalb der einzelnen Module eingebaut sein. Abbildung 4 zeigt schematisch die, zur Verlängerung des Peptidgerüsts um einen Monomerbaustein, notwendigen Reaktionsschritte in einem Elongationsmodul.

Der Katalysezyklus beginnt mit der Substraterkennung in der Bindungstasche der A-Domäne. Die für eine Peptidbindungsknüpfung essentielle Aktivierung der Carbonsäurefunktion des Monomerbausteins wird durch die Bildung eines Acyladenylats unter Verbrauch von ATP erreicht (siehe Abbildung 4A). Das aktivierte Substrat wird unter AMP-Abspaltung auf den 4'-Phosphopantetheinylarm (Ppant-Arm) der PCP-Domäne übertragen und dort als Thioester gebunden. Dieser in Abbildung 4B dargestellte Reaktionsschritt benötigt eine produktive Domäneninteraktion zwischen der A- und der PCP-

Domäne. Die C-Domäne des Elongationsmoduls katalysiert schließlich die Kondensationsreaktion des gebundenen Monomerbausteins mit der Amino- bzw. Peptidgruppe des vorangegangenen Moduls (siehe Abbildung 4C) [34-36]. Nach einer Weitergabe der um einen Baustein verlängerten Peptidkette auf das nächste Modul (siehe Abbildung 4D) liegt der Ppant-Arm des Elongationsmoduls wieder frei vor und kann in einem weiteren Katalysezyklus erneut beladen werden.

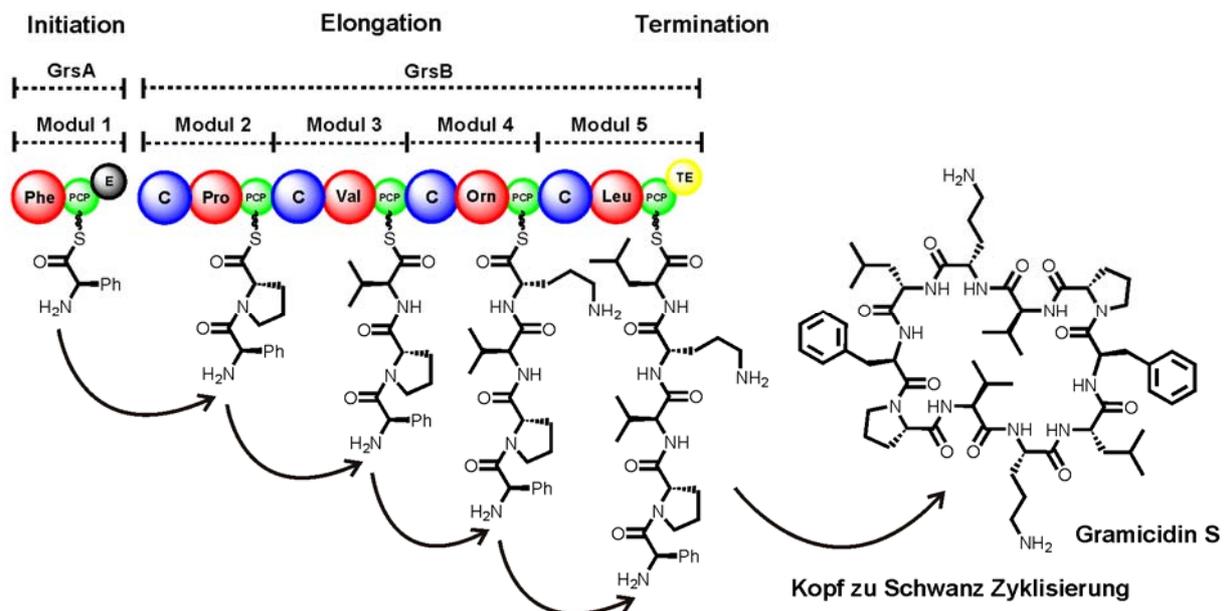


**Abbildung 4 Schematischer Reaktionsverlauf an einem Elongationsmodul (C-A-PCP) in NRPS.** A) Die A-Domäne aktiviert eine spezifische Aminosäure unter ATP-Verbrauch zum Aminoacyladenylat, B) Unter AMP-Abspaltung wird die aktivierte Aminosäure kovalent als Thioester an den 4'-Phosphopantetheinylarm der PCP-Domäne gebunden, C) Die Peptidbindungsknüpfung findet an der C-Domäne des Elongationsmoduls mit dem Peptidyl- oder Aminoacylthioester des vorangehenden, beladenen Moduls statt, D) Die um einen Baustein verlängerte Peptidkette wird unter Bildung einer weiteren Peptidbindung auf die PCP-Domäne des nächsten Moduls übertragen.

Am Beispiel von Gramicidin S soll im Folgenden die NRP-Biosynthese noch detaillierter beschrieben werden.

### 1.1.3 Gramicidin S Biosynthese

Normalerweise sind alle Gene, die in der Produktion eines NRP zusätzlich zu den eigentlichen NRPS-Genen benötigt werden (Gene zur Produktion der unnatürlichen Aminosäuren, Gene für die Modifikation des Naturstoffes nach dem NRPS Fließbandprozess, Resistenzgene, etc.), gebündelt in einem so genannten Gencluster. Dies erlaubt eine koordinierte Regulierung der Genexpression und resultiert sehr wahrscheinlich aus dem zwischen Bakterien häufig vorkommenden horizontalen Gentransfer. So kann in einem einzigen DNA-Transferereignis die vollständige biosynthetische Kapazität übertragen werden.

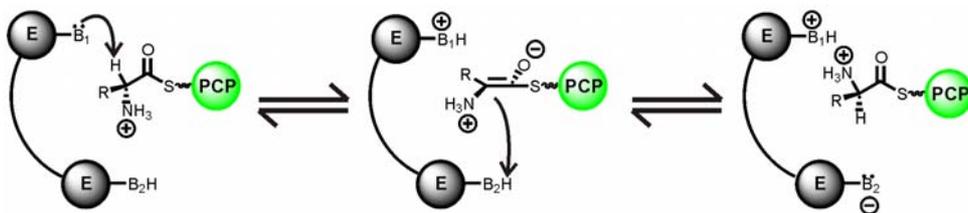


**Abbildung 5 NRPS-Biosynthese von Gramicidin S.** Die zwei Multi-Domänenproteine Gramicidin S Synthetase I (GrsA) und Gramicidin S Synthetase II (GrsB) sind in *Bacillus brevis* für die Synthese des Breitbandantibiotikums Gramicidin S verantwortlich. Eine doppelte Verwendung des fünf-modulgroßen NRPS-Fließbands führt nach einer Kopf zu Schwanz Dimerisierung zu dem zyclischen Zehnerpeptid. Als unnatürliche Bausteine sind in diesem NRP sowohl D-Phenylalanin, als auch Ornithin zu finden (Rot = Adenylierungsdomänen; Grün = PCP-Domänen, Blau = Kondensationsdomänen; Schwarz = Epimerisierungsdomäne; Gelb = Thioesterasedomäne).

Das Gramicidin S Operon aus *Bacillus brevis* besteht aus dem am 5'-Ende lokalisierte *grsT*-Gen (dieses Gen wurde noch nicht detailliert charakterisiert, jedoch zeigt es eine große Homologie zu Fettsäurethioesterasen), sowie den zwei Genen Gramicidin S Synthetase I (*grsA*) und Gramicidin S Synthetase II (*grsB*) [9,10]. Das Gen der Phosphopantetheinyltransferase *gsp* (diese Enzymklasse wird in Kapitel 1.1.5 vorgestellt) für dieses NRPS-System befindet sich ca. 320 bp in 5'-Richtung von *grsT* [37]. Hinweise auf resistenz-vermittelnde Gene wurden nicht gefunden. Das Genprodukt von *grsA*, das 127 kDa

große GrsA-Protein, besteht aus einem Initiationsmodul (A-PCP) mit einer zusätzlichen Epimerisierungsdomäne (E-Domäne) und ist für den Einbau von D-Phe in Gramicidin S verantwortlich. Vier weitere Module sind auf dem ~ 510 kDa großen GrsB-Protein zu finden - sie bauen Prolin, Valin, Ornithin und Leucin in das Peptidgerüst ein. Die Gramicidin S Biosynthese ist ein Paradebeispiel für ein iteratives NRPS-System. Jeweils zwei Durchgänge der fünf Module des Fließbandprozesses führen zu dem zyklischen Dekapeptid (s. Abbildung 5) [38].

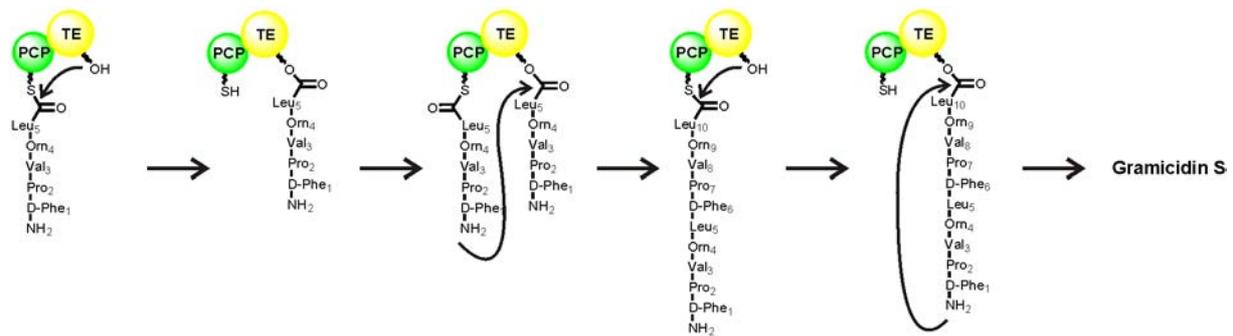
Am Anfang der Biosynthese steht die essentielle Umwandlung beider Gramicidin S Synthetasen von ihrer apo- in die holo-Form. Dazu wird die prosthetische Gruppe 4'-Phosphopantethein (Ppant), ausgehend von dem Vorläufermolekül CoA, enzymatisch auf spezifische Serinreste jeder PCP-Domäne übertragen (s. Abbildung 12). Somit stattet der Ppant-Arm die PCP-Domänen mit einer nukleophilen Sulphydrylgruppe aus, die es ermöglicht die von der A-Domäne aktivierten Aminosäuren kovalent als Thioester zu binden. Die GrsA A-Domäne aktiviert mit identischer Effektivität sowohl L-, wie auch D-Phenylalanin [39,40]. Das natürliche Substrat ist aufgrund der höheren Verfügbarkeit sehr wahrscheinlich die L-Aminosäure. Beide Aminoacyladenylat-Stereoisomere können auf die GrsA PCP-Domäne übertragen werden (s. Reaktionsschritte A und B in Abbildung 4). Die GrsA E-Domäne katalysiert nun durch De- und Reprotonierung des C $\alpha$ -Protons die Epimerisierung des an die PCP-Domäne gebundenen Thioesters (s. Abbildung 6). Unabhängig von der Konfiguration des durch die A-Domäne aktivierten Phenylalanins (D- oder L-) bildet sich sehr schnell ( $k_{\text{cat}} > 100 \text{ min}^{-1}$ ) ein leicht in Richtung des D-Isomers verschobenes Gleichgewicht (1,9:1) zwischen den beiden Phenylalaninthioestern aus. Obwohl der Mechanismus der E-Domäne noch nicht vollständig aufgeklärt ist, so sprechen einige Indizien [41] für einen Zwei-Basen-Mechanismus (für weitere Erklärungen s. Abbildung 6).



**Abbildung 6 Vorgeschlagener Mechanismus von NRPS E-Domänen (Stachelhaus).** Die Epimerisierung der L-Aminosäure in die korrespondierende D-Aminosäure findet auf der Stufe des an die PCP-Domäne gebundenen Aminoacylthioester statt. Untersuchungen deuten daraufhin, dass zwei unterschiedliche Säure/Base Funktionalitäten der E-Domäne die De- und Reprotonierung an der C $\alpha$ -Position katalysieren (Zwei-Basen-Mechanismus). Die katalytisch aktiven Reste der E-Domänen konnten jedoch noch nicht eindeutig identifiziert werden. Das Schema zeigt von links nach rechts die Umwandlung eines L- in einem D-Aminoacylthioester.

Obwohl der im Gleichgewicht vorliegende Anteil des an die GrsA PCP-Domäne gebundenen L-Phe-Thioester mit ca. 35 % relativ groß ist, findet eine Peptidbindungsknüpfung fast ausschließlich zwischen D-Phe und L-Pro des ersten GrsB Moduls (GrsB1) statt (~ 98%). Diese Stereoselektivität der Kondensationsreaktion wird hauptsächlich von der Donorposition der GrsB1 C-Domäne vermittelt, die das D-Konformer des Phe-S-PCP<sub>GrsA</sub> gegenüber dem L-Konformer bevorzugt [41,42]. Für die Akzeptorpositionen in NRPS C-Domänen wird angenommen, dass sie ihr Aminoacyl-S-PCP Substrat des eigenen Elongationsmoduls mit hoher Affinität ebenfalls stereoselektiv binden [42,43]. Erst nach der Peptidbindungsknüpfung kann die beladene PCP-Domäne mit weiteren katalytischen Zentren, z.B. der Donorposition des folgenden Moduls interagieren. Dies ist sehr wahrscheinlich die Grundlage für die Unidirektionalität der NRPS Biosynthese und verhindert eine verfrühte Initiation der Peptidsynthese an internen Modulen [42-44]. Den entsprechenden Reaktionsablauf findet man in der weiteren Synthese von Gramicidin S. Ausgehend von dem an die GrsB1 PCP-Domäne gebundenen D-Phe-Pro-Thioester wird nun das Peptid an den nächsten drei Modulen analog zu denen in Abbildung 4 dargestellten Schritten um die Aminosäuren Val, Orn und Leu verlängert.

Die finalen Reaktionsschritte, die Dimerisierung und Zyklisierung des Pentapeptides, werden von der PCP- und der Thioesterase-Domäne des Terminationsmoduls GrsB4 katalysiert (s. Abbildung 7). Dabei wird zunächst das lineare Pentapeptid vom Kofaktor Ppant der PCP-Domäne auf einen hochkonservierten Serinrest der TE-Domäne übertragen [45]. Anschließend ergibt ein weiterer Durchgang durch das NRPS-System die erneute Bildung des D-Phe-Pro-Val-Orn-Leu-Thioesters an der PCP-Domäne von GrsB4. Es gibt experimentelle Hinweise [38], dass nun die N-terminale Aminogruppe des an die PCP-Domäne gebundenen Pentapeptids die Esterfunktionalität der Thioesterasedomäne nukleophil angreift und damit das lineare Dekapeptid erzeugt. In einem weiteren Schritt wird dieses zurück auf den Serinrest der TE-Domäne übertragen. Die Abspaltung des Naturstoffes vom Proteintemplat erfolgt durch einen internen, nukleophilen Angriff der Peptidaminogruppe an der Esterfunktionalität (s. Abbildung 7). Die TE-Domäne katalysiert selektiv diese Reaktion und verhindert gleichzeitig eine mögliche Hydrolyse zum linearen Produkt, sowie die Angriffe anderer interner Nukleophile (wie z.B. der Ornithinseitenketten). Die Selektivität dieser Reaktion wird sehr wahrscheinlich durch die Vorfaltung des Peptidsubstrates begünstigt. Denn Untersuchungen ergaben, dass bereits der lineare Dekapeptidvorläufer eine  $\beta$ -Faltblattstruktur annimmt, und sich damit die für die Zyklisierung benötigten Gruppen bereits in räumlicher Nähe befinden [46,47].



**Abbildung 7** Vorgeschlagener Mechanismus der Dimerisierung und Zyklisierung von Gramacidin S am GrsB4 Terminationsmodul [38]. Das durch die fünf NRPS-Module von GrsA und GrsB erzeugte Pentapeptid wird vom Ppant-Arm der PCP-Domäne auf einen konservierten Serinrest der TE-Domäne transferiert. Nach einer erneuten Beladung der GrsB4 PCP-Domäne mit einem weiteren Pentapeptid bildet sich unter Peptidbindungsknüpfung das Dekapeptid an der PCP-Domäne und wird anschließend auf die TE-Domäne zurück übertragen. Eine als  $\beta$ -Faltblattstruktur vorliegende Vorfaltung des Peptids und eine Positionierung und Aktivierung der reaktiven Gruppen (N-terminale Aminogruppe des Peptids und Esterfunktion) durch die TE-Domäne erlauben im letzten Schritt eine effiziente und selektive Bildung des zyclischen Peptidantibiotikums Gramacidin S.

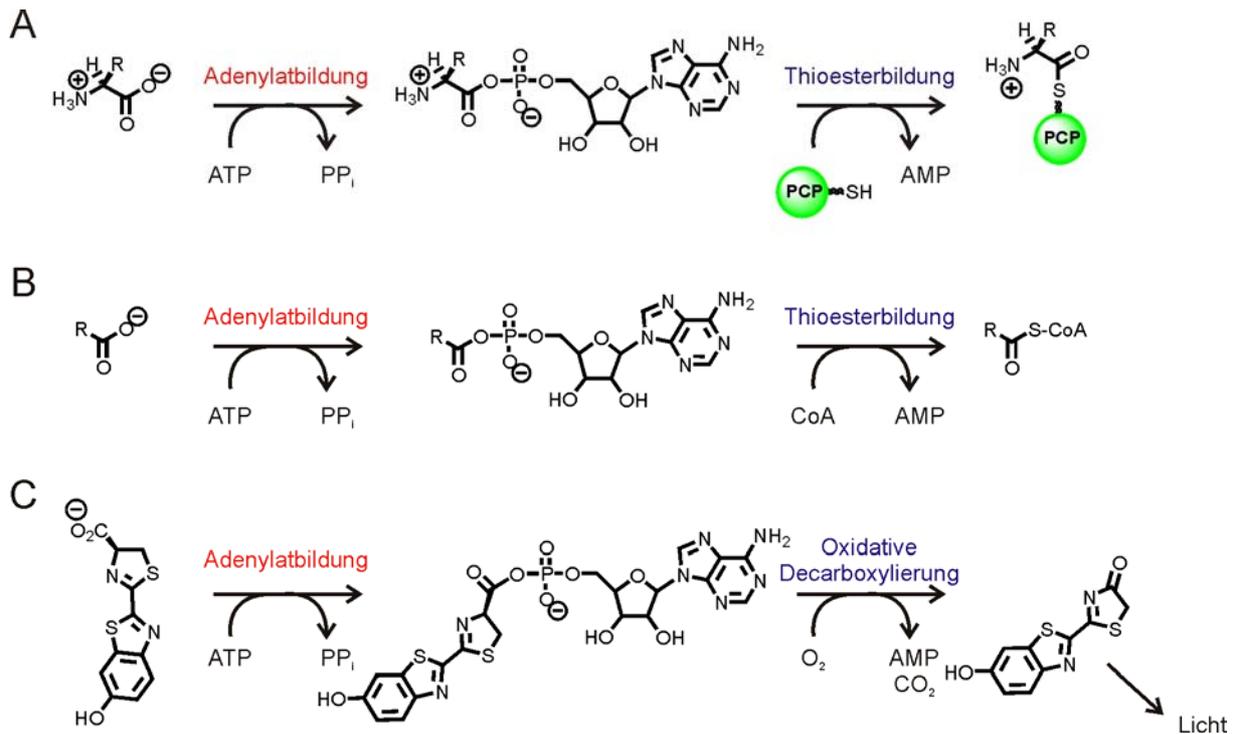
Die relativ einfache Biosynthese von Gramacidin S deckt bei Weitem nicht alle möglichen Varianten der NRPS-Biosynthesemaschinerien ab. So wurden oxidative Quervernetzungen und andere modifizierende Reaktionen, wie z.B. Glykosylierungen, die entweder noch am oder nach der Abspaltung vom NRPS-Templat erfolgen, in dieser Einleitung nicht angesprochen. Beispiele für solche NRP Naturstoffe sind die Antibiotika Vancomycin und Teicoplanin. Auch ist die große Klasse der Hybride aus NRPS und Polyketidsynthasen (PKS) nicht Gegenstand dieser Arbeit. Es sei hier auf eine Vielzahl an exzellenten Übersichtsartikeln zu diesen Naturstoffsynthesen hingewiesen [48-51].

Zu den weiteren Errungenschaften der NRPS-Forschung der letzten Jahrzehnte gehört die detaillierte biochemische und strukturelle Charakterisierung einzelner NRPS-Domänen. Ein Überblick über den Stand der Forschung in Bezug auf die für diese Arbeit relevanten Domänen, der Adenylierungs- und der PCP-Domäne, wird in den nächsten Kapiteln präsentiert.

### 1.1.4 Die Adenylierungsdomänen

Zur Knüpfung einer Peptidbindung ist die Aktivierung der Carbonsäurefunktionalität eines Aminosäurebausteins essentiell. Die Umwandlung der Carbonsäure in eine gute Abgangsgruppe erfolgt sowohl in der ribosomalen, wie auch in der nicht-ribosomalen Peptidsynthese durch die Bildung eines Acyladenylats unter ATP-Verbrauch. In NRPS-Systemen katalysieren Adenylierungsdomänen (A-Domänen) nicht nur diese Reaktion, sondern vermitteln zusätzlich den nächsten Reaktionsschritt, die Übertragung des aktivierten Substrats auf den Ppant-Kofaktor der PCP-Domäne des gleichen Moduls (s. Schritte 1 und 2 in Abbildung 4). Welchen Monomerbaustein ein Modul in das Peptidgerüst einbaut, wird durch die Spezifität der A-Domäne bestimmt. Aus einem Pool von möglichen Substraten in der Zelle selektiert die A-Domäne den passenden Baustein aus und aktiviert ihn.

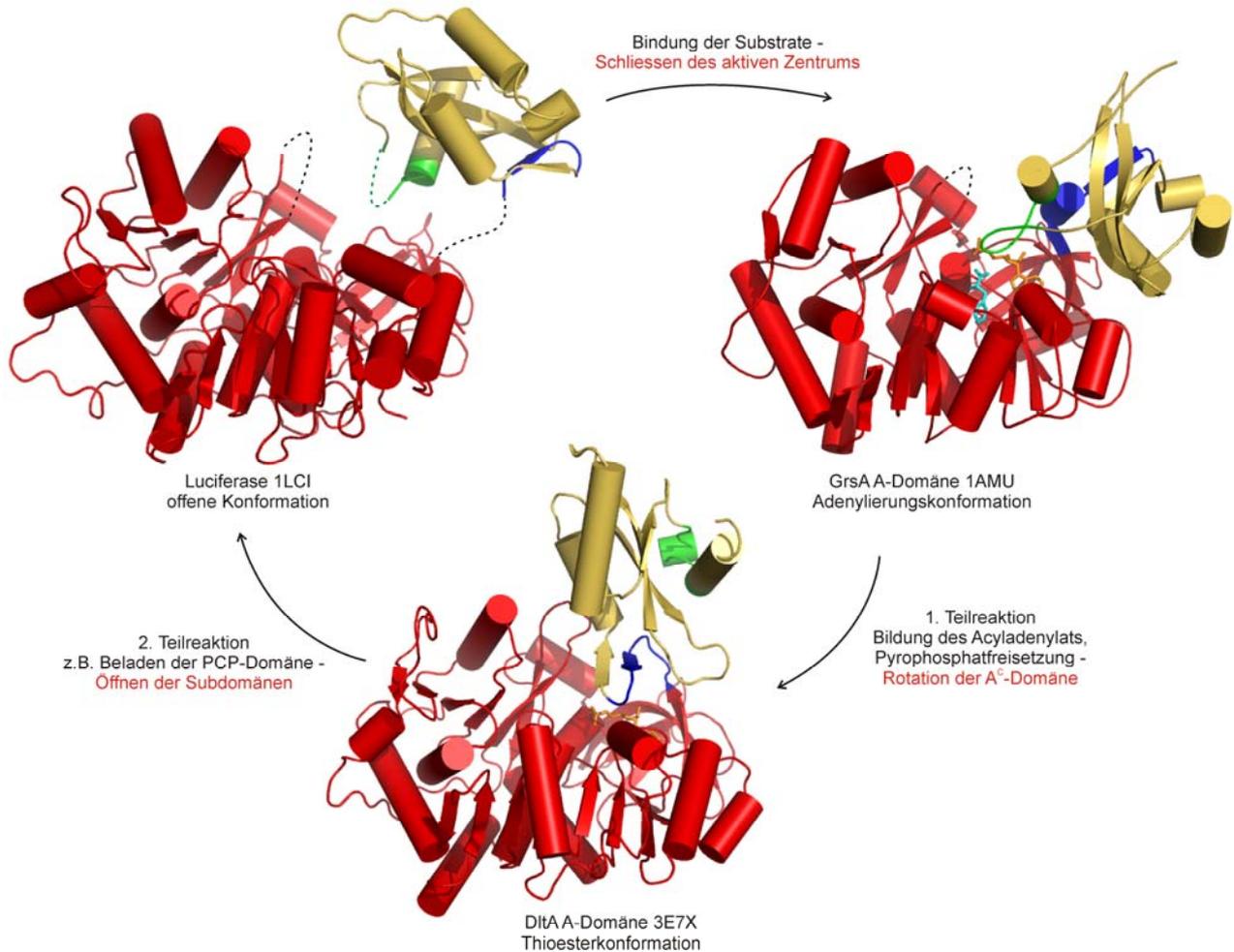
Trotz ähnlich katalysierten Reaktionen zeigen A-Domänen keine Homologie mit tRNA-Synthetasen der Klassen I und II. Jedoch werden sie einer anderen Proteinsuperfamilie, zusätzlich bestehend aus den Acyl- und Aryl-CoA Synthetasen und den Leuchtkeifer Luciferasen, zugeordnet. Diese Proteine werden unter dem Begriff ANL-Superfamilie zusammengefasst. Sie besitzen eine Sequenzidentität von ~ 20% und vergleichende Analysen identifizierten konservierte Sequenzmotive [52]. Von den zehn anhand der NRPS A-Domänen abgeleiteten Kernmotiven (A1-A10) sind neun innerhalb der gesamten Superfamilie konserviert [53]. Zusätzlich zu dieser Sequenzhomologie weisen die ANL-Enzyme eine große strukturelle Homologie auf [53]. Sie bestehen aus zwei Subdomänen, einem 400 - 500 Aminosäuren großen N-terminalen ( $A^N$ ) und einem 100-150 Reste großen C-terminalen Teil ( $A^C$ ). Die drei Enzymklassen katalysieren jeweils zwei Teilreaktionen, wobei die erste, die Aktivierung einer Carbonsäure zum Acyladenylat unter ATP Verbrauch, bei allen Proteinen gemeinsam vorkommt (s. Abbildung 8). Bei der zweiten Teilreaktion kommt es im Falle der A-Domänen und der Aryl- bzw. Acyl-CoA Synthetasen zu einem nukleophilen Angriff eines Ppant-Thiols auf den Kohlenstoff der Carbonsäuregruppe. Das Thiol ist entweder an *in cis* oder *in trans* interagierende PCP-Domänen gebunden oder kommt direkt von CoA und verdrängt unter Bildung eines Thioesters die sehr gute Abgangsgruppe AMP. Bei den Luciferasen kommt es nach der Luciferinaktivierung zur oxidativen Decarboxylierung des Substrats und das entstehende Oxyluciferin emittiert noch im aktiven Zentrum des Enzyms ein Photon [54].



**Abbildung 8 Reaktionen der ANL-Superfamilie.** A) NRPS A-Domänen; B) Aryl- und Acyl-CoA Synthetasen; C) Leuchtkäfer Luciferasen; Bei allen Enzymen dieser Superfamilie ist der erste Teilschritt, die Aktivierung des Substrats als Acyladenylat, identisch. Eine Thioesterbildung auf das Thiol einer Ppant-Gruppe schließt sich bei den NRPS A-Domänen und den Acyl- bzw. Aryl-CoA Synthetasen an. Luciferasen verwenden eine oxidative Decarboxylierung zur Erzeugung von Biolumineszenz.

Für alle drei Enzymklassen innerhalb der ANL-Superfamilie werden aus Gründen der Sequenzidentität und der strukturellen Homologien identische Katalysemeechanismen und konformationelle Änderungen während der Katalyse angenommen. Anhand von Kristallstrukturen wurde postuliert, dass die ANL-Superfamilie und damit auch die NRPS A-Domänen, drei unterschiedliche Konformationen während eines Katalysezyklus durchlaufen [53,55]. Sie unterscheiden sich hauptsächlich durch verschiedene Orientierung der Subdomänen ( $A^N$  und  $A^C$ ) zueinander. Abbildung 9 zeigt diese Konformationen am Beispiel von repräsentativen Strukturen von Mitgliedern der ANL-Superfamilie.

Der Katalysezyklus beginnt in einer offenen Struktur, in der die beiden Subdomänen fast keine Kontaktfläche miteinander besitzen. Der Zugang zum aktiven Zentrum, das zwischen den Subdomänen gebildet wird, ist somit frei für die Substratbindung. Ein gutes Beispiel für diese offene Konformation ist die Kristallstruktur der Luciferase aus *Photinus pyralis* [56] (s. Abbildung 9). Die Bindung der Substrate induziert wahrscheinlich den Übergang in die so genannte Adenylierungskonformation, eine geschlossene Konformation ohne Lösungsmittelzugang zum aktiven Zentrum.

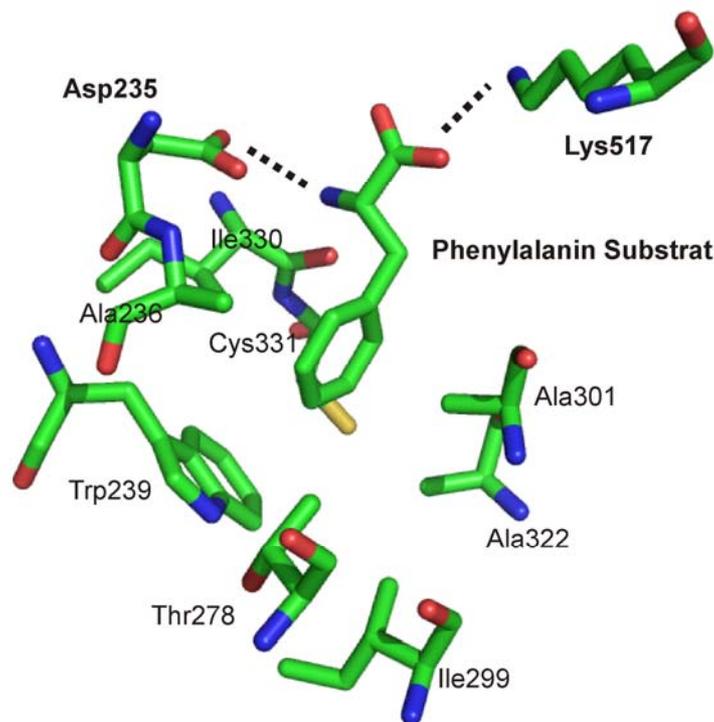


**Abbildung 9 Vorgeschlagener Katalysezyklus von Mitgliedern der ANL-Superfamilie.** Gezeigt sind die Kristallstrukturen der Luciferase aus *Photinus pyralis* [56] (pdb-Code 1LCI) und die beiden NRPS A-Domänen aus Gramicidin S Synthetase I (GrsA, *Bacillus brevis*) [57] (pdb-Code 1AMU) und die freistehende D-Alanin-D-Alanyl Carrier Protein Ligase DltA aus *Bacillus subtilis* [55] (pdb-Code 3E7X); In allen drei Strukturen ist die N-terminale Subdomäne (A<sup>N</sup>) in rot, die C-terminale Subdomäne (A<sup>C</sup>) in hellgelb und die konservierten Sequenzmotive A8 und A10 in blau, respektive in grün dargestellt; Bei fehlender Elektronendichte sind die Verbindungen durch gestrichelte Linien ergänzt; für Details siehe Text.

In Abbildung 9 ist dafür stellvertretend die GrsA A-Domäne gezeigt [57]. Sie wurde im Komplex zusammen mit Phenylalanin und AMP kristallisiert. Die beiden Substrate befinden sich im aktiven Zentrum, welches hauptsächlich von der A<sup>N</sup>-Subdomäne gebildet wird. Der einzige Rest in der A<sup>C</sup>-Subdomäne, der aktiv an der Positionierung der Substrate beteiligt ist, ist Lys517 (GrsA Nomenklatur). Es ist Teil des konservierten Sequenzmotivs A10 (in Abbildung 9 in grün dargestellt) und bildet Wasserstoffbrücken mit dem Carboxylatsauerstoff des Phenylalanins, sowie dem Ringsauerstoff und dem 5'-Sauerstoff der Ribose aus. Neueste Kristallstrukturen ordnen diesem Rest ebenfalls Salzbrücken mit den Phosphatgruppen des ATPs zu [58]. Die Wichtigkeit dieses Restes für die erste Teilreaktion, die Bildung des Acyladenylats, wurde durch Mutationsstudien an der Luciferase [59] und der Propionyl-CoA Synthetase PrpE [60] nachgewiesen. Bei der Letzteren konnte für die Mutanten Lys592Ala

und Lys592Glu (analoger Rest zu Lys517 in GrsA) bei Inkubation mit Propionat und ATP keine Bildung von Propionyl-AMP nachgewiesen werden. Interessanterweise konnten die Mutanten die zweite Teilreaktion, die Umwandlung von Propionyl-AMP in Propionyl-CoA, weiterhin katalysieren. Dies deutet daraufhin, dass das Motiv A10 im zweiten Reaktionsschritt nicht mehr essentiell ist.

Die Co-Kristallisation der GrsA A-Domäne mit Phenylalanin erlaubte die Bestimmung der Reste, die für die Bildung der Bindetasche des Aminosäuresubstrats verantwortlich sind [57]. Das sich in Motiv A4 der A<sup>N</sup>-Subdomäne befindende, hochkonservierte Asp235 (GrsA Nomenklatur) stabilisiert durch die negativ geladene Carboxylseitenkette den Aminoterminus des Substrates. Nur in A-Domänen, die keine Aminosäuren, und stattdessen z.B. Fettsäuren, aktivieren kann dieser Rest mutiert sein. Folgende Reste bilden die Bindetasche für die L-Phe Seitenkette in der GrsA A-Domäne: Ala236, Trp239, Thr278, Ile299, Ala301, Ala322, Ile330 und Cys331 (s. Abbildung 10).



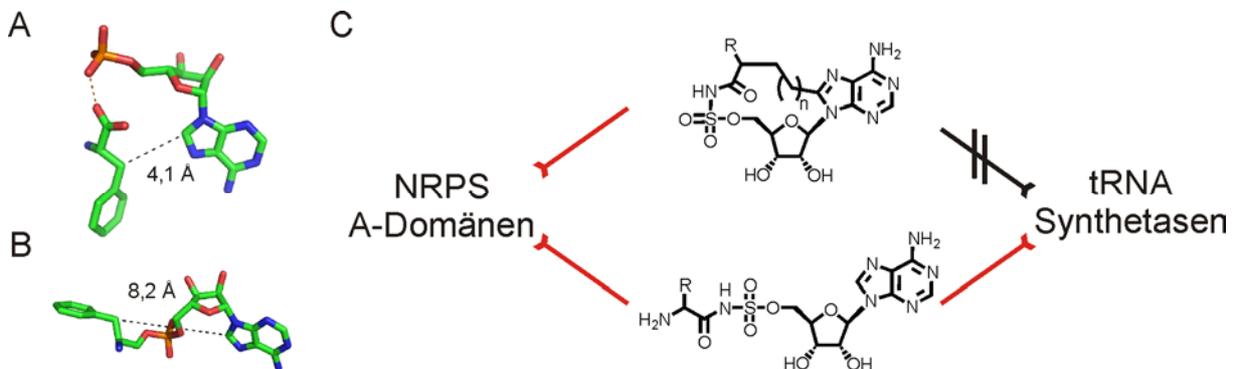
**Abbildung 10 Bindetasche des Phenylalaninsubstrates in der A-Domäne von GrsA (pdb-Code 1AMU).** Die innerhalb von NRPS A-Domänen hochkonservierten Reste Asp235 und Lys517 (GrsA Nummerierung) sind mit ihren Wechselwirkungen zu der Amino-, bzw. zu der Carboxylgruppe des Substrates eingezeichnet; die anderen mittel- bis hochvariablen Reste sind für die Selektivität der A-Domäne in Bezug auf die Seitenkette der zu aktivierenden Aminosäure verantwortlich [61].

Vergleichende Untersuchungen mit anderen NRPS A-Domänen bekannter Selektivität erlaubten das Aufstellen eines Spezifitätscodes für A-Domänen anhand dieser Reste [61]. Damit gelang es sowohl die Spezifität unbekannter A-Domänen vorherzusagen, als auch

durch Mutation von Resten in der Bindetasche die Spezifität einer A-Domäne zu verändern [61].

Sowohl der ATP-Bindungsmodus, als auch die ATP-Bindetasche weisen innerhalb der ANL-Superfamilie eine große Ähnlichkeit auf. Hochkonservierte Reste interagieren spezifisch mit den verschiedenen funktionellen Gruppen des ATPs. So bildet Asp413 (Motiv A7) mit einem oder beiden Ribosehydroxylgruppe Wasserstoffbrücken aus. Tyr326 aus dem Sequenzmotiv A5 macht  $\pi$ -Basenstacking mit dem Adeninring. Ebenfalls hochkonserviert ist das glycin-, serin- und threoninreiche Motiv A3, der so genannte p-Loop oder Walker A-Motiv [62]. Diese in Kristallstrukturen selten aufgelöste Region positioniert die  $\beta$ - und  $\gamma$ -Phosphatgruppen [58,63].

Nach der ersten Teilreaktion liegt das gebildete Adenylat in einer so genannten „cis“-Konformation im aktiven Zentrum der ANL-Enzyme vor (s. Abbildung 11A). Dieser Sachverhalt konnte elegant zur Synthese von Inhibitoren ausgenutzt werden, die selektiv nur NRPS A-Domänen inhibieren und nicht die korrespondierenden tRNA-Synthetasen. Die Inhibitoren basieren auf nicht-hydrolysierbaren Sulfoamoylanalogen der Aminoacyladenylate und sind in der „cis“-Konformation fixiert. Damit entsprechen sie nicht den in der „trans“-Konformation vorliegenden Aminoacyladenylaten der tRNA-Synthetasen und inhibieren letztere auch nicht (s. Abbildung 11) [64].



**Abbildung 11 Konformationen der Substratbindung und Inhibierung von adenylatbildenden Enzymen.** A) In der Kristallstruktur der GrsA A-Domäne [57] (1AMU) sind AMP und Phe in der „cis“-Konformation gebunden; B) In der Kristallstruktur der phenylalaninaktivierenden tRNA Synthetase aus *Thermus thermophilus* [65] (1B7Y) liegt das Adenylatanalogon in der gestreckten „trans“-Konformation vor; C) Nicht-hydrolysierbare Sulfoamoylanaloga der Aminoacyladenylate sind Inhibitoren der adenylatbildenden Enzyme. Das in der „cis“-Konformation fixierte Molekül (oben) inhibiert selektiv NRPS A-Domänen und keine tRNA Synthetasen. Der konformativ flexible Inhibitor (unten) kann sowohl die „cis“- als auch die „trans“-Konformation einnehmen, an beide Enzymklassen binden und damit auch beide inhibieren [64].

Für die zweite Teilreaktion in der ANL-Superfamilie wird eine weitere geschlossene Konformation der Proteine, die Thioesterkonformation, angenommen. Unter anderem wurde die D-Alanin-D-Alanyl Carrier Protein Ligase (DltA) aus *Bacillus subtilis* im Komplex mit

AMP in dieser Konformation kristallisiert [55]. Im Vergleich zur Adenylierungskonformation ist die C-terminale Subdomäne um  $\sim 140^\circ$  gedreht. Dies positioniert an Stelle des konservierten A10-Motivs, welches für die zweite Teilreaktion nicht mehr essentiell ist (s. oben), die Aminosäurereste des A8-Motivs im aktiven Zentrum der A-Domäne (vergleiche Abbildung 9). Mutationsstudien weisen vor allem dem zu Lys434 in GrsA analogen Rest eine wichtige katalytische Funktion zu. So zeigte die korrespondierende Alanin-Mutation in der *Photinus pyralis* Luciferase eine drastisch verminderte Reaktionsrate bei der oxidativen Decarboxylierung. Die Fähigkeit zur Adenylatbildung war bei dieser Mutante im Vergleich zum Wildtyp nicht eingeschränkt [66]. Während der Katalyse wird für diesen Lysinrest eine Stabilisierung des Adenylatintermediats durch Koordinierung der Sauerstoffatome der Carbonsäuregruppe und des Phosphats postuliert. Die Beteiligung weiterer Reste des A8-Motivs an der Katalyse der zweiten Teilreaktion wurde in weiteren Studien nachgewiesen (Gly437 analoger Rest in der Acetyl-CoA Synthetase [67] - Arg436 analoger Rest in EntE, einer NRPS A-Domäne aus dem Enterobactin-Biosyntheseweg [68]). Die Drehung der kleineren Subdomäne bringt nicht nur das A8-Motiv in das aktive Zentrum, sondern bildet zusätzlich einen Pantethein-Tunnel zwischen den beiden Subdomänen aus. Im Falle der A-Domänen kann so die Ppant-Gruppe der korrespondierenden holo-PCP Domäne im aktiven Zentrum positioniert werden. Zur Katalyse des nukleophilen Angriffs des Ppant-Thiols auf die Carbonsäuregruppe des Acyladenylats ist wahrscheinlich eine Deprotonierung des Thiols notwendig. Für diesen Teilschritt konnte bisher kein konservierter Rest gefunden werden. Neueste Untersuchungen postulieren den Einfluss eines Helix-Dipolmoments der Region A4 für die Deprotonierung [69].

Der zeitliche Ablauf der konformationellen Änderungen und der Substratbindungen wird wie folgt für NRPS A-Domänen postuliert [53,55]. Die offene Konformation geht nach Bindung von ATP und der Aminosäure in die Adenylierungskonformation über. In dieser Konformation wird die Bildung des Aminoacyladenylats katalysiert und das entstandene Pyrophosphat diffundiert schnell aus dem aktiven Zentrum heraus [70]. Die  $\beta$ - und  $\gamma$ -Phosphatgruppe des ATPs belegen in der Adenylierungskonformation einen Hohlraum, der in der Thioesterkonformation von dem A8 Motiv eingenommen wird. Dies, zusammen mit kinetischen Studien, deutet darauf hin, dass die Pyrophosphatabspaltung der Domänenalternierung vorangeht [58,63,71]. Die Bindung der holo-PCP-Domäne, oder zumindest der Ppant-Gruppe, an die A-Domäne scheint erst in der Thioesterkonformation möglich zu sein, da der Pantethein-Tunnel noch nicht in der Adenylierungskonformation vorhanden ist. Nach der Thioesterbildung und der Freisetzung von AMP und der beladenen

holo-PCP-Domäne geht die A-Domäne wahrscheinlich wieder in die offene Konformation zurück.

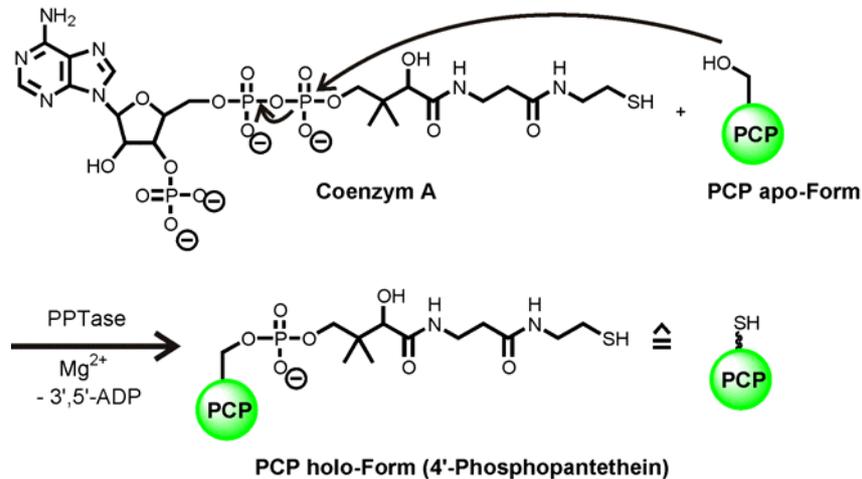
In der Literatur ist der Katalysezyklus der ANL-Enzyme mittels Domänenalternierung weitgehend akzeptiert [53,55], obwohl kein einziges Enzym in allen drei Konformationen kristallisiert werden konnte oder strukturelle Untersuchungen zu dem Reaktionsablauf in Lösung vorliegen. Ein erster biochemischer Hinweis gelang durch Mutation der Gelenkaminosäure Asp402 zu Prolin (Asp430 in GrsA), die zwischen den beiden Subdomänen A<sup>N</sup> und A<sup>C</sup> in der 4-Chlorobenzoat CoA Ligase (4-CBL) lokalisiert ist [72]. Diese Mutante kristallisierte in der Adenylierungskonformation und zeigte eine Reduktion in der katalytischen Effizienz der ersten Teilreaktion um den Faktor drei. Die zweite Teilreaktion war jedoch um den Faktor 1000 beeinträchtigt und die Autoren schlossen daraus, dass die Prolinmutation eine Rotation der A<sup>C</sup>-Subdomäne verhindert und die katalytischen Reste des A8 Motivs nicht in das aktive Zentrum gelangen können [72]. Yonus et al. [55] wiesen darauf hin, dass während der Domänenalternierung alle relevanten geladenen Aminosäuren ihre Partneraminosäuren austauschen. Sie postulierten nur geringe Energiedifferenzen zwischen den verschiedenen konformationellen Zuständen. Durch kleinere Veränderungen im aktiven Zentrum könnte die Umorientierung der Domänen ausgelöst werden.

Diesem profunden Wissen über die ANL-Superfamilie im Allgemeinen und über die NRPS A-Domänen im Speziellen steht ein relativ geringes Wissen über die Wechselwirkung der A-Domänen mit ihren korrespondierenden holo-PCP-Domänen gegenüber. So konnte bisher z.B. noch keine A-Domäne im interagierenden Komplex mit einer PCP-Domäne kristallisiert werden. Den NRPS PCP-Domänen und ihren Wechselwirkungen mit anderen Domänen ist das nächste Kapitel gewidmet.

### 1.1.5 Die PCP-Domänen

Der zentrale Ankerpunkt der Intermediate der NRP-Biosynthese sind die 80 bis 100 Aminosäuren großen Peptidyl Carrier Protein-Domänen (PCP-Domänen) der einzelnen NRPS-Module. Sie können zusammen mit den Acyl Carrier Proteinen (ACPs) aus der Fettsäure- und der Polyketidsynthese und den Aryl Carrier Proteinen (ArCP) strukturell einer Proteinfamilie zugeordnet werden und besitzen keine eigenständige katalytische Funktion. PCPs unterscheiden sich von ACPs durch ihre Nettoladung bei neutralem pH (PCPs sind neutral, ACPs sind negativ geladen). Die Mehrzahl der PCP-Domänen ist in größere

Multidomänenproteine integriert (s. z.B. die Biosynthese von Gramicidin S), jedoch findet man auch allein stehende PCP-Domänen. Allen Carrier-Proteinen gemein ist die post-translationale Addition einer Ppant-Gruppe ausgehend von CoA auf einen hoch konservierten Serinrest (s. Abbildung 12).



**Abbildung 12 Die „Priming“-Reaktion.** Eine Ppant-Transferase wandelt eine PCP-Domäne von der inaktiven apo- in die aktive holo-Form um. Dabei wird die Seitenkette eines hoch konservierten Serins der PCP-Domäne mit einem 4'-Phosphopantetheinrest ausgehend von CoA post-translational modifiziert.

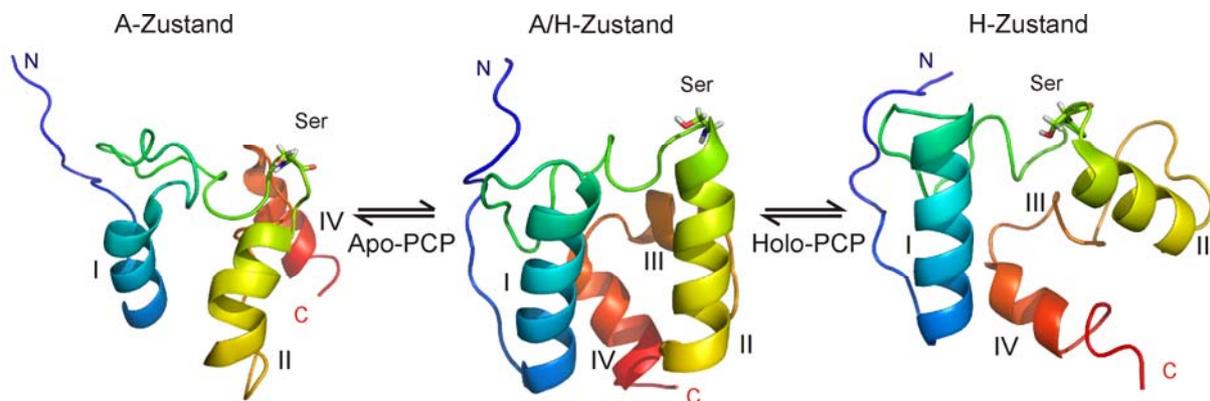
Diese apo- zu holo-Modifikation, die so genannte „Priming“-Reaktion, wird von der Enzymklasse der Ppant-Transferasen durchgeführt [73-75]. In Mikroorganismen, wie *Bacillus subtilis* oder *Mycobacterium tuberculosis*, gibt es eine Vielzahl an Carrier-Proteinen, die jedoch alle von nur zwei Ppant-Transferasen modifiziert werden [76]. So erkennt die Acyl carrier protein synthase (AcpS) aus *B. subtilis* spezifisch die ACPs der Fettsäuresynthese des primären Metabolismus [76]. In *M. tuberculosis* ist eine zweite Ppant-Transferase selektiv in der Lage alle Carrier-Proteine des sekundären Metabolismus umzuwandeln. Im Lichte einer besseren Regulation dieser beiden Stoffwechselsysteme ist die Selektivität der Ppant-Transferasen evolutionär sinnvoll zu erklären. Eine Ausnahme ist die neben AcpS zweite Ppant-Transferase in *B. subtilis* [75,76]. Sfp ist Teil des NRPS-Systems zur Produktion des Antibiotikums Sufactin und zeigt nicht nur eine große Toleranz auf Seiten der Carrier-Proteine, sondern ist ebenfalls in der Lage eine Vielzahl an Coenzym A-Derivaten zu übertragen. Das Enzym ist aufgrund dieser Eigenschaften zu einem sehr interessanten Werkzeug in der post-translationalen Modifikation von Proteinen geworden [77-80].

Wie in früheren Kapiteln bereits erwähnt, ist die wachsende Peptidkette während der NRP-Biosynthese permanent an die Thiofunktionalität der holo-PCP-Domänen gebunden. Erst durch die Abspaltung an den Thioesterase-Domänen erfolgt eine Freisetzung vom Proteintemplat. Die kovalente Bindung der Intermediate, im Falle der NRPS als Thioester,

besitzt mehrere konzeptionelle Vorteile. Das so genannte „substrate channeling“ verhindert die freie Diffusion der Intermediate innerhalb der Zelle und erhöht die lokale Konzentration der Substrate an den aktiven Zentren dramatisch. Eine Verbesserung der Reaktionsgeschwindigkeit im Vergleich zur freien Diffusion ist in diesen Systemen zu erwarten. Zwar sind die gemessenen Reaktionsraten von NRP-Biosynthesen *in vivo* und *in vitro* nicht beeindruckend ( $k_{\text{cat}}$  typischerweise im Bereich von  $0,1 - 10 \text{ min}^{-1}$ ) [81,82], sie wären ohne die kovalente Bindung an die PCPs jedoch bei weitem noch langsamer [83]. Ein weiterer Vorteil besteht in der Abschirmung der Intermediate vor dem Zugang des Lösungsmittels und damit eine effektive Reduktion von möglichen Nebenreaktionen [84,85]. Durch die Bindung an eine PCP-Domäne kann man nicht mehr alleine die wachsende Peptidkette als Substrat für die nächste Reaktion betrachten, sondern muss dies für die gesamte modifizierte Domäne annehmen. Aus den Wechselwirkungen der (beladenen) PCP-Domäne mit den anderen NRPS-Domänen resultiert sicherlich ein Großteil der Selektivität und Spezifität in diesen Systemen - dies zeigt die enorme Wichtigkeit dieser Interaktionen. Bevor auf diese Wechselwirkungen im Detail eingegangen wird, sei hier auf die generellen strukturellen Eigenschaften der PCP-Domänen hingewiesen.

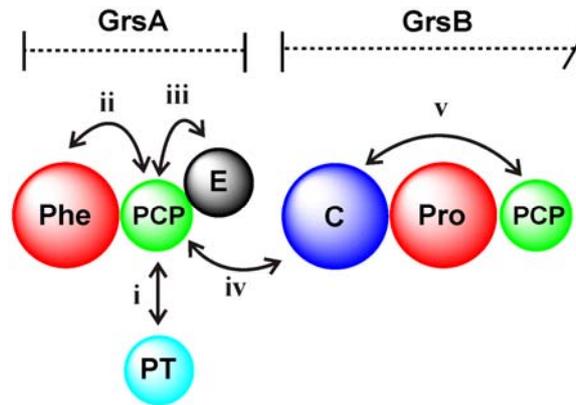
Aufgrund ihrer relativ geringen Größe lassen sich isolierte PCP-Domänen sehr gut mittels NMR-Spektroskopie in Lösung untersuchen [86,87]. Ebenfalls sind andere Repräsentanten der CP-Familie auf diese Weise analysiert worden [88,89]. Die Untersuchungen zeigten, dass die CP-Domänen aus einem drei- oder vier-Helix Bündel bestehen. Helix I liegt antiparallel zu den Helices II und IV vor. Zusätzlich werden letztere durch eine dritte Helix, eine kurze  $\alpha$ - oder  $3_{10}$ -Helix, verbunden. Im Durchschnitt lassen sich mehr als 50 % der Sekundärstrukturen in Carrier-Proteinen einem helikalen Charakter zuordnen. Schon früh wurde anhand von NMR-Daten postuliert, dass die CP-Domänen intrinsisch mobil sind und sie einem konformationellen Austausch unterliegen [90]. Für PCP-Domänen konnte dies am Beispiel der TycC3 PCP-Domäne (aus dem dritten Modul der Tyrocidin A Synthetase III) eindeutig bestätigt werden [87]. In Lösung wechselten sowohl die freistehende apo-, wie auch holo-PCP Domäne zwischen zwei unterschiedlichen stabilen Konformationen. Eine Struktur, der so genannte A/H-Zustand, war in beiden Proteinen mit Ausnahme des Ppant-Arms identisch (s. Abbildung 13). Bei beiden Formen wurde zusätzlich eine einzigartige Struktur ermittelt. So befand sich die apo-PCP Domäne im langsamen Austausch zwischen dem A/H- und dem A-Zustand, holo-PCP wechselte langsam zwischen dem A/H- und dem so genannten H-Zustand (s. Abbildung 13). Durch die konformationelle Dynamik ist es der PCP-Domäne möglich unterschiedlich beschaffene Proteinoberflächen zu präsentieren. Die Erkennung und Bindung

einer anderen Domäne erfordert das Vorhandensein einer spezifischen Kontaktfläche, die nur in einer der interkonvertierenden Strukturen gebildet ist. So wird bei Zugabe der Ppant-Transferase Sfp zu apo-PCP der A-Zustand selektiv gebunden und dadurch stabilisiert [87]. Analog bindet die Typ II Thioesterase an die fehlbeladene TycC3 PCP-Domäne im H-Zustand. Erkennungsbereiche sind die Loop III Region, der Loop II mit dem konservierten Serinrest und das N-terminalen Ende der Helix II. Nur im H-, und nicht im A/H-Zustand, bilden diese Elemente eine kontinuierliche Kontaktfläche [87,91]. Erwähnenswert ist noch, dass für die verwandten ACPs teilweise ein ungefalteter, intrinsisch unstrukturierter Zustand angenommen wird (für einen Übersichtsartikel s. [92]).



**Abbildung 13 NMR-Strukturen der apo- und holo-TycC3 PCP-Domäne [87].** Sowohl die apo-, als auch die holo-Form befinden sich im langsamen Austausch zwischen zwei Konformationen. Für apo-TycC3 PCP sind dies der A-Zustand (pdb-Code 2GDY) und der mit holo-PCP gemeinsame A/H-Zustand (pdb-Code 2GDW). Die, dem holo-Protein eigene Konformation ist der so genannte H-Zustand (pdb-Code 2GDX). In der Abbildung sind die Proteine vom N- zum C-Terminus von blau nach rot koloriert. Die Nummerierungen der Helices, sowie die Position des hoch konservierten Serinrestes am N-terminalen Ende der Helix II sind ebenfalls dargestellt.

Die Wechselwirkungen der PCP-Domänen mit katalytischen Partnerdomänen, wie z.B. A-C-, E- oder TE-Domänen, können in NRPS-Systemen entweder *in cis* (beide interagierenden Domänen auf einer Polypeptidkette) oder *in trans* (die Domänen liegen auf unterschiedlichen Proteinen vor) vonstatten gehen. Schematisch sind diese Wechselwirkungen am Beispiel der PCP-Domäne von GrsA in Abbildung 14 dargestellt. Jüngste Forschungsergebnisse lieferten erste Hinweise auf die genauen Interaktionsflächen auf Seiten der PCP-Domäne und ermöglichten die Bestimmung der für die korrekte Protein-Protein-Wechselwirkung essentiellen Aminosäurereste. Hier sind vorrangig zum einen die hochauflösende Strukturen eines PCP-TE Fragmentes [93] und eines kompletten Terminationsmoduls [94], sowie zum anderen die kombinatorischen Mutagenesestudien in der Enterobactin-Biosynthese zu nennen [68,95-98].



**Abbildung 14 Schematische Darstellung der Wechselwirkungen der GrsA und der GrsB1 PCP-Domäne mit anderen NRPS-Domänen.** Die PCP-Domäne des Initiationsmoduls muss sowohl Interaktionen *in trans* mit der Ppant-Transferase (i) oder der C-Domäne des folgenden Moduls (iv), als auch *in cis* mit der A- (ii) und der E-Domäne (iii) koordinieren. Neben weiteren modifizierenden Domänen sind in Elongationsmodulen ebenfalls noch Wechselwirkungen mit der C-Domäne des eigenen Moduls (v), sowie in Terminationsmodulen die Interaktion mit der TE-Domäne möglich (nicht gezeigt).

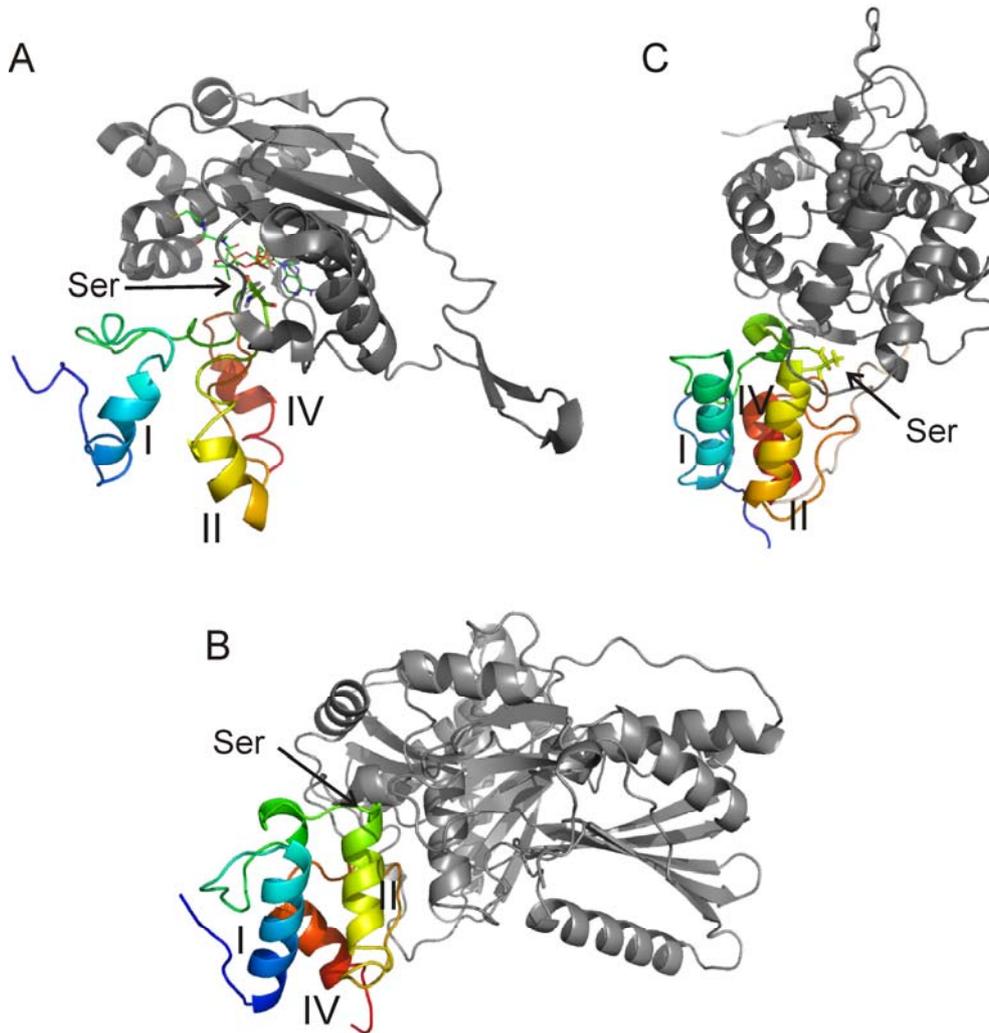
Auf die Interaktionen der apo-PCP Domäne mit den Ppant-Transferasen (Wechselwirkung (i) in Abbildung 14) wurde bereits weiter oben eingegangen. Zur besseren Visualisierung ist der modellierte Komplex des A-Zustandes von TycC3 PCP mit der Ppant-Transferase Sfp und CoA in Abbildung 15A dargestellt [87]. Die Erkennung scheint vor allem über Aminosäurereste in direkter Nähe zu dem zu modifizierenden Serin abzulaufen [95,98-100]. Die Identifikation einer definierten Kontaktfläche zu *in trans* oder *in cis* agierenden A-Domänen (Wechselwirkung (ii) in Abbildung 14) gestaltet sich schwieriger. Mutationsstudien zeigten, dass die Interaktion dieser Domänen eine große Anzahl an Veränderungen auf Seiten der PCP-Domäne toleriert. So wurde in dem zu GrsA analogen TycA Protein die gesamte Helix II der PCP-Domäne gegen die korrespondierende Helix des *Bacillus subtilis* ACPs der Fettsäuresynthase ausgetauscht. Jedoch fand weiterhin eine ungestörte Kommunikation zwischen den beiden Domänen statt [99]. Interessanterweise konnte in diesem chimären Protein die als Thioester gebundene Aminosäure nicht mehr korrekt epimerisiert werden, so dass anzunehmen ist, dass Helix II eine wichtige Rolle bei der Wechselwirkung mit der E-Domäne spielt (s. (iii) in Abbildung 14). Weiterhin wurden intensive Mutationsstudien an der ArCP Domäne des Proteins EntB aus der Enterobactin-Biosynthese durchgeführt. Auch hier zeigte sich, dass die Wechselwirkung zwischen der kognaten *in trans* arbeitenden A-Domäne (EntE) und der EntB ArCP Domäne sehr tolerant ist. Von neun getesteten Aminosäureresten in Loop-Regionen um das konservierte Serin herum zeigten nur zwei Mutanten eine verminderte Affinität zu EntE (die -5 (5 Reste N-terminal vom konservierten Serin) und die +18 Position (18 Reste C-terminal vom konservierten Serin)) [68]. Noch detaillierter analysierte die Gruppe um Walsh die identische ArCP Domäne (EntB). Hier wurden ~ 80 %

aller an der Oberfläche beteiligten Aminosäuren kombinatorisch mutiert. In Übereinstimmung mit früheren Ergebnissen [101] wurden die Reste -8 und -9 identifiziert. Zusammen mit dem Rest +24 aus der Helix III waren sie die einzigen Reste, denen eine essentielle Beteiligung an der Wechselwirkung mit der A-Domäne nachgewiesen werden konnte [98,102].

Für die Interaktion mit der C-Domäne des nächsten Moduls (Wechselwirkung (iv) in Abbildung 14) liegt eine Kristallstruktur eines Didomänenkonstruktes (PCP-C) vor [103]. Die beiden Domänen befinden sich jedoch nicht in einer produktiven Konformation, so dass keine Interaktionsoberfläche bestimmt werden konnte. Daten aus den oben genannten Mutationsstudien weisen jedoch den Resten +19 und +23 in Helix III einen wichtigen Einfluss auf diese Wechselwirkung zu [68,95-98]. Grafische Darstellungen der Interaktionen mit der C-Domäne des eigenen Moduls ((v) in Abbildung 14) und mit einer Typ I Thioesterase-Domäne sind in Abbildung 15B, respektive C gezeigt. Die strukturellen Untersuchungen wurden jeweils an inaktiven Mutanten der PCP-Domäne (Serin zu Alanin) durchgeführt [93,94].

Zusammenfassend kann man den aktuellen Forschungsstand zu den NRPS PCP-Domänen und ihren Interaktionen mit ihren Partnerdomänen wie folgt darstellen. Die PCP-Domänen sind intrinsisch mobil und können unterschiedliche Konformationen einnehmen. Bei Bindung an eine Partnerdomäne wird eine der Konformationen selektiert und stabilisiert. Die Erkennung erfolgt über Oberflächenmotive, deren essentielle Kontakte stark lokalisiert sind. Diese wichtigen Aminosäurereste auf den PCP-Domänen scheinen zwar je nach Partnerdomäne unterschiedlich zu sein, jedoch zeigen die hochauflösenden Komplexstrukturen, dass die vollständigen Kontaktflächen auch teilweise überlappen. Für ein vollständiges Verständnis der PCP-Dynamiken und Wechselwirkungen in NRPS Systemen sind jedoch noch weitere Forschungen notwendig, auch um das Ableiten von generellen Regeln der Domäneninteraktion für eine gezielte Manipulation der PCP-Domänen zu ermöglichen.

Der strukturellen Anordnung der Domänen innerhalb eines Moduls ist der letzte Abschnitts des ersten Teils der Einleitung gewidmet.



**Abbildung 15** Strukturen von PCP-Domänen im Komplex mit interagierenden NRPS-Domänen. A) Modellierter Komplex der apo-TycC3 PCP-Domäne mit der Ppant-Transferase Sfp und CoA (pdb-Code 2GE1) [87]; B) Ausschnitt aus der Kristallstruktur des Terminationsmoduls SrfA-C - gezeigt sind die PCP-Domäne und die C-Domäne des eigenen Moduls (pdb-Code 2VSQ) [94]; C) NMR-Struktur des Didomänenfragments PCP-TE aus EntF (pdb-Code 2R0Q) [93]; In der Abbildung sind die PCP-Domänen vom N- zum C-Terminus von blau nach rot und die interagierenden Domänen in grau koloriert. Die Nummerierungen der Helices, sowie die Position des in B) und C) mutierten Serinrestes sind gezeigt.

### 1.1.6 Domänenanordnung innerhalb eines Moduls

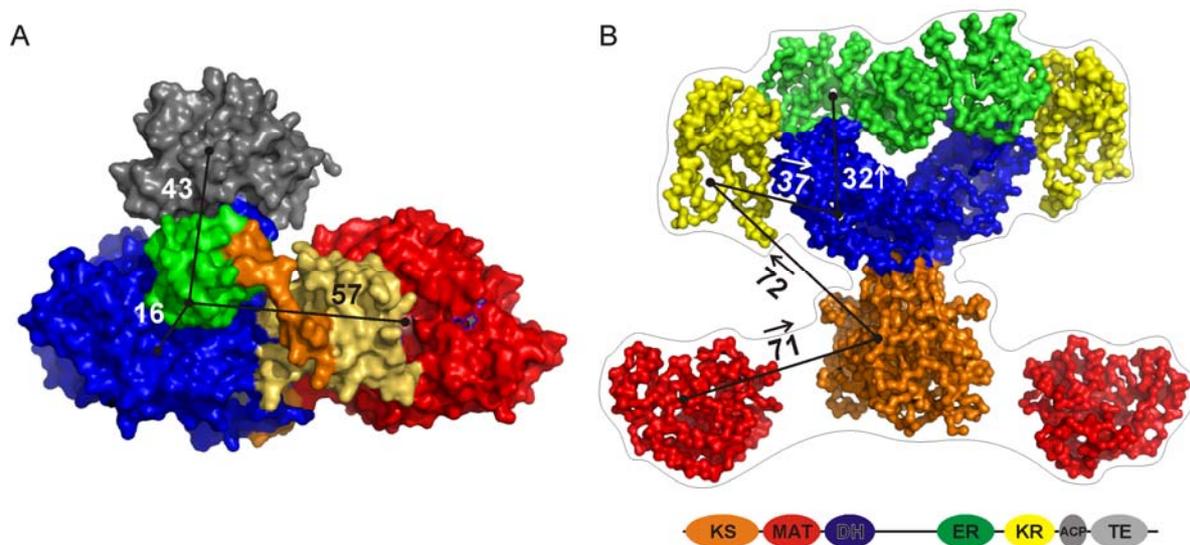
Damit die in Kapitel 1.1.5 beschriebenen Wechselwirkungen der PCP-Domänen mit den katalytischen Domänen überhaupt stattfinden können, müssen diese in räumliche Nähe zueinander kommen. Für *in trans* Interaktionen lässt sich ein einfaches Modell der kurzzeitigen Bindung und Dissoziation aufstellen. Interessanter wäre eine detaillierte Kenntnis über die dreidimensionale Organisation in einem Multi-Domänen- oder Multi-Modulprotein während der verschiedenen Stadien der Katalyse. Erste Hinweise lieferte eine 2008 veröffentlichte Kristallstruktur eines kompletten Terminationsmoduls, mit der Domänenzusammensetzung C-A-PCP-TE [94] (s. Abbildung 16A). In dem 144 kDa großen

SrfA-C Protein aus dem Surfactin-Biosyntheseweg besitzt die TE-Domäne keine nennenswerten Kontaktregionen mit den anderen Domänen. Ihre Faltung in diesem Multidomänenprotein ist identisch mit der Struktur der freistehenden TE-Domäne [104]. Jedoch stehen die drei anderen Domänen (C-A-PCP), die man auch als Prototyp eines Elongationsmoduls ansehen kann, in engem Kontakt. Hier bilden die C-Domäne und die A<sup>N</sup>-Subdomäne eine Plattform, auf der die A<sup>C</sup>-Subdomäne und die PCP-Domäne befestigt sind. Die 32 Aminosäuren lange Linkerregion zwischen der C- und der A-Domäne stabilisiert durch zahlreiche Kontakte die Interaktion beider Domänen. Die Autoren postulierten, dass aufgrund der starken Wechselwirkungen die C-A<sup>N</sup>-Plattform während des Katalysezyklus invariant bleibt [94]. Im Gegensatz dazu besitzt der 15 Aminosäure große Linkerbereich zwischen der A- und der PCP-Domäne kaum Kontakte zu den einzelnen Domänen.

Wie bereits im vorhergehenden Kapitel erwähnt befindet sich die PCP-Domäne im Kontakt mit der Akzeptorposition der C-Domäne (s. Abbildung 15B). Der Abstand des hoch konservierten PCP-Serinrestes (in diesem Protein wurde die Alaninmutante verwendet) zu dem aktiven Zentrum der C-Domäne beträgt 16 Å. Diese Distanz lässt sich von einem gestreckten Ppant-Arm (18 - 20 Å) überbrücken und damit ist in dieser Kristallstruktur wahrscheinlich die produktive Interaktion der PCP- mit der C-Domäne abgebildet. Im Laufe des Katalysezyklus müsste die PCP-Domäne jedoch noch mit dem aktiven Zentrum der A- (57 Å Abstand) und der TE-Domäne (43 Å Abstand) interagieren (s. Abbildung 16A). Diese Abstände sind zu groß, um sie durch eine konformationelle Änderung der PCP-Domäne, auch in Kombination mit einem Umklappen des Ppant-Arms zu überbrücken. Diese Daten sprechen gegen das früher aufgestellte „Modell des schwingenden Arms“, in dem die PCP-Domäne ausgehend von einer zentralen Position alle aktiven Zentren mit Hilfe des Ppant-Arms erreicht [105,106]. Vielmehr ist eine Bewegung der PCP-Domäne relativ zu der A- und der TE-Domäne zu erwarten. Wie diese PCP-Verschiebungen aussehen und welche Faktoren sie auslösen, lässt sich anhand der Kristallstruktur jedoch nicht ableiten [107].

Die biosynthetische Logik von NRPS-Multidomänenproteinen, d.h. ein zentrales Carrier-Protein, das die Intermediate der Synthese zwischen den unterschiedlichen katalytischen Zentren der anderen Domänen transportiert, findet man auch in den Fettsäuresynthesen (FAS). Die Domänenanordnung in den Multidomänenproteinen von FAS Typ I wurde ebenfalls bereits im Detail untersucht und hier wurde, analog zu den NRPS, das Auftreten von Carrier-Protein Verschiebungen relativ zu ihren katalytischen Domänen postuliert. Die Gruppe um Nenad Ban veröffentlichte 2006 die Kristallstrukturen sowohl einer Pilz-, als auch einer Säugetier-Fettsäuresynthase [108-110]. Obwohl das Proteinrückgrat nicht mit atomarer

Auflösung bestimmt werden konnte, gelang es doch die Position der katalytischen Zentren in den Komplexen zu lokalisieren. Obwohl beide Fettsäuresynthasen unterschiedliche architektonische Strukturen ausbilden, so waren in allen gebildeten Reaktionskammern die Abstände zwischen den aktiven Zentren zu groß, um von einer zentralen Carrier-Protein Domäne erreicht zu werden. Eine genaue Lokalisation der ACP-Domänen gelang aufgrund fehlender Elektronendichte nicht. Abbildung 16B zeigt die homodimere Struktur der Säugetier-Fettsäuresynthase mit den Abständen der katalytischen Zentren zueinander. Im Laufe eines Reaktionszyklus muss die ACP-Domäne eine produktive Interaktion mit der Malonyl-CoA/acetyl-CoA-ACP Transacylase (MAT), der  $\beta$ -Ketoacylsynthase (KS), der Ketoacylreduktase (KR), der Dehydratase (DH) und der  $\beta$ -Enoylreduktase (ER) eingehen. Am Ende der Synthese erfolgt die Abspaltung der Fettsäure von dem Multidomänenprotein an der Thioesterasedomäne (Elektronendichte ist ebenfalls nicht aufgelöst).



**Abbildung 16 Räumliche Anordnung der Domänen in NRPS und FAS Proteinen.** A) Kristallstruktur des Terminationsmoduls SrfA-C aus der Surfactinbiosynthese [94] (pdb-Code 2VSQ) bestehend aus der C-Domäne (blau), der valinaktivierenden A-Domäne ( $A^N$  rot und  $A^C$  hellgelb), der PCP-Domäne (grün) und der TE-Domäne (grau). Die Linkerbereiche zwischen den Domänen sind in orange koloriert. Ausgehend vom möglichen Anknüpfungspunkt des Ppant-Arms (S1003A Mutante) an der PCP-Domäne sind die Abstände (in Å) zu den aktiven Zentren der anderen Domänen gezeigt. Die Struktur stellt wahrscheinlich einen Schnappschuss der produktiven Domäneninteraktionen zwischen der C- und der PCP-Domäne dar. Der 18 - 20 Å lange Ppant-Arm der PCP-Domäne kann jedoch ohne eine relative Verschiebung der Domänen zueinander nicht die Zentren der A- (57 Å) und der TE- Domäne (43 Å) erreichen. B) Homodimere Kristallstruktur einer Säugetier-Fettsäuresynthase [108] (pdb-Code 2CF2). Das Monomer hat in der Primärsequenz die Domänenanordnung: Ketosynthase (KS, orange), Malonyl-CoA/acetyl-CoA-ACP Transacylase (MAT, rot), Dehydratase (DH, blau), Enoylreduktase (ER, grün), Ketoacylreduktase (KR, gelb), Acyl-Carrier Protein (ACP) und Thioesterase (TE). Die beiden letzten Domänen sind in der Kristallstruktur aufgrund fehlender Elektronendichte nicht aufgelöst. Die Abstände (in Å) zwischen den Stationen des ACPs während eines Katalysezyklus (MAT-KS-KR-DH-ER) sind nur für die linke Reaktionskammer graphisch dargestellt. Auch hier werden unter der Annahme einer stabilen Gerüststruktur Verschiebungen des Carrier Proteins relativ zu den aktiven Zentren postuliert [108].

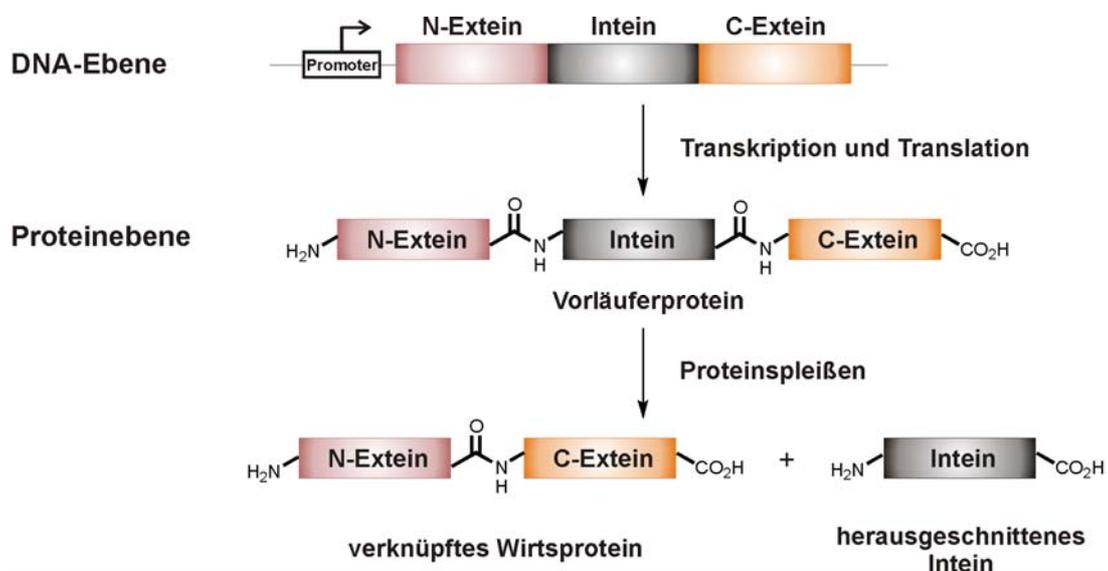
## **Limitationen im Verständnis der Koordination von Multidomänenproteinen und Methoden der chemischen Biologie zu deren Aufklärung**

Die in diesem Kapitel vorgestellten Kristallstrukturen von Multidomänenproteinen sind Meilensteine auf dem Weg zum Verständnis dieser hochkomplexen Biosyntheseapparate. Sie vermitteln ein „Bild“ von der räumlichen Organisation der einzelnen Domänen zueinander und bilden die Grundlage für weiterführende biochemische Experimente. Jedoch ermöglichen sie keinen Einblick in das dynamische Verhalten und die mit sehr großer Wahrscheinlichkeit auftretenden konformationellen Änderungen in diesen enormen Proteinen. Auch wenn es in Zukunft gelingen sollte, weitere NRPS-Module in unterschiedlichen Stadien der Katalyse, z.B. eine PCP-Domäne in Interaktion mit einer A-Domäne, zu kristallisieren, so wird dadurch nicht das „wirkliche“ Verhalten dieser Maschinerien in Lösung und damit ihre Wechselwirkung mit den Substraten zu entschlüsseln sein. Andere biophysikalische Methoden, die es ermöglichen ein zeitaufgelöstes Bild (beginnend bei Seitenkettenfluktuationen im Pikosekundenbereich und endend bei Domänenverschiebungen im Millisekunden- bis Sekundenbereich) von Proteinen in atomarer Auflösung zu erzielen, sind unter anderem die Einzelmolekülfluoreszenz- und die Kernspinresonanzspektroskopie (NMR). Sie besitzen jedoch im Vergleich zur Röntgenkristallographie den Nachteil, dass unnatürliche Sonden (Fluorophore oder NMR-aktive Kerne) in das zu untersuchende Protein eingebracht werden müssen. Im Laufe der letzten Jahrzehnte hat die noch junge Disziplin der Chemischen Biologie eine Vielzahl an Möglichkeiten entwickelt Proteine mittels chemischer Methoden selektiv zu verändern. Die Ansätze reichen von der Ausnutzung der intrinsischen Reaktivitäten von Aminosäureseitenketten und anderen Biokonjugationsmethoden über den Codon-basierten Einbau unnatürlicher Aminosäuren [111] bis zur chemischen Totalsynthese von Proteinen [112]. Mit dem Begriff der Proteinsemisynthese bezeichnet man seit Kurzem den Zusammenbau von Proteinen aus rekombinanten und synthetischen Proteinfragmenten. Zur Ligation zweier Proteinfragmente stehen unterschiedlichen Methoden, wie z.B. die Expressed Protein Ligation (EPL) [113,114], die Staudinger-Ligation [115], die reverse Proteolyse [116,117] oder die Sortase-vermittelte Ligation zur Verfügung [118,119]. Alle diese Techniken ermöglichen den gezielten Aufbau eines Proteins, auch mit biophysikalischen Sonden, jedoch ist eine detaillierte Beschreibung dieser Techniken nicht Gegenstand dieser Einleitung und es sei an dieser Stelle auf einen exzellenten Übersichtsartikel hingewiesen [120]. Eine weitere Ligationsmethode zur Verknüpfung zweier Protein- oder Peptidfragmente ist das so genannte Protein *trans*-Spleißen (PTS) mit Hilfe gespaltener Inteine. Im Laufe dieser Doktorarbeit wurden strukturelle und biochemische

Untersuchungen an gespaltenen Inteinen durchgeführt, auch um deren Einsetzbarkeit für die Analyse von Multidomänenproteinen, wie den NRPS, zu etablieren und zu erweitern. Das folgende Kapitel stellt Inteine im Allgemeinen, und die Besonderheiten und Anwendungsmöglichkeiten gespalteener Inteine im Speziellen vor.

## 1.2 Inteine

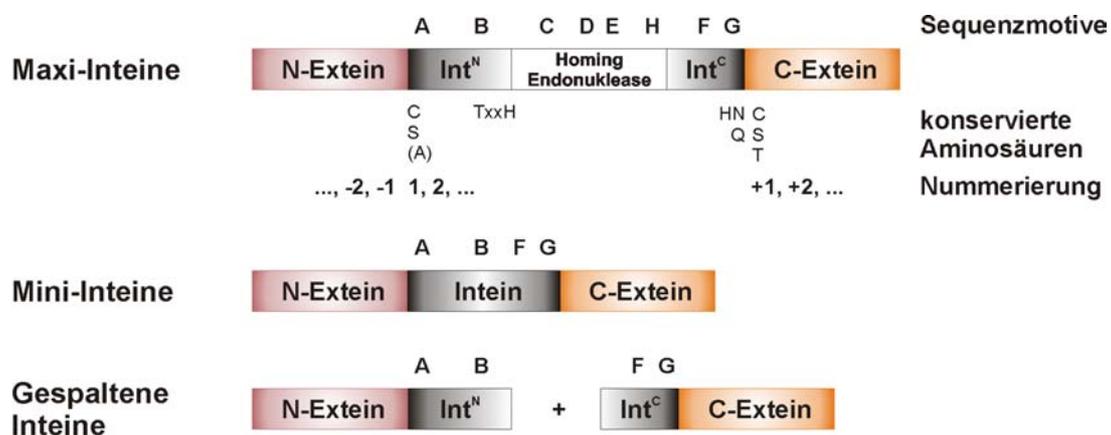
Polypeptidketten werden oftmals nicht kontinuierlich auf der DNA-Ebene kodiert. So können eukaryotische Gene so genannte Introns enthalten, die nach der Transkription durch RNA Spleißen entfernt werden. Die entstandene, gekürzte Boten-RNA (mRNA) kodiert schließlich für das vollständige, kontinuierlich translatierte Protein [121]. Im Gegensatz dazu werden andere DNA-Sequenzen, die für so genannte Inteine kodieren, zunächst sowohl vollständig transkribiert als auch translatiert (s. Abbildung 17). Erst auf Proteinebene findet eine durch das Intein vermittelte post-translationale Prozessierung, das Protein-Spleißen, statt. Dabei entfernt sich das Intein aus dem Vorläuferprotein durch Spaltung zweier Peptidbindungen und verknüpft dabei die angrenzenden Proteinsegmente, die so genannten Exteine, durch eine native Peptidbindung [122,123]. Bei dem in Abbildung 17 dargestellten Prozess kodiert also ein Gen für zwei unterschiedliche Genprodukte, zum einen die verknüpften Exteine (das aktive Wirtsprotein) und zum anderen das herausgeschnittene Intein. Diese bemerkenswerten Proteindomänen sind in allen Bereichen des Lebens, sowohl in Archaeabakterien, Eubakterien und Eukaryoten zu finden. Jedoch befindet sich unter den aktuell über 300 gefundenen Inteinen (s. Intein-Datenbank *InBase* [124]) noch kein Vertreter aus einem mehrzelligen höheren Organismus.



**Abbildung 17 Schematische Darstellung des Proteinspleißens.** Inteine sind Beispiele, in denen Polypeptidketten nicht kontinuierlich auf der DNA-Ebene kodiert werden. Nach Transkription und Translationprozessiert sich das entstandene Vorläuferprotein selbständig. Dabei wird das Intein durch Spaltung zweier Peptidbindungen entfernt. Simultan werden die beiden Exteinfragmente miteinander ligiert und es entsteht das aktive Wirtsprotein.

Intein-enhaltende Wirtspoteine besitzen diverse Funktionen. Es sind Stoffwechsellzyme, Proteasen, aber vor allem Proteine, die an der DNA-Replikation oder -Reparatur beteiligt sind. Ihnen allen ist gemein, dass die Inteine in hochkonservierten Bereichen oder nahe den aktiven Zentren zu finden sind [125]. Inteine werden sowohl nach ihrem Wirtorganismus als auch nach ihrem Wirtspotein benannt. So ist das *Ssp* DnaB Intein in die DnaB Helikase aus dem Cyanobakterium *Synechocystis* sp. strain PCC6803 integriert. Bei mehreren Inteinen innerhalb eines Wirtspoteins werden diese vom N-Terminus durchnummeriert (*Tfu* Pol-2 ist das zweite Intein in der DNA Polymerase aus *Thermococcus fumicolans*). Besitzen zwei Inteine aus verschiedenen Organismen die gleiche Insertionsstelle innerhalb eines Wirtspoteins, so spricht man von unterschiedlichen Inteinallelen [124].

Inteine lassen sich anhand ihres Aufbaus in drei große Klassen einteilen, die in Abbildung 18 schematisch dargestellt sind. Die üblicherweise 300 - 600 Aminosäure großen Maxi-Inteine enthalten neben den N- und C-terminalen Spleißbereichen (Int<sup>N</sup> bzw. Int<sup>C</sup>) zusätzlich eine integrierte Homing Endonuklease. Diese Domäne fehlt der Klasse der Mini-Inteine, welche deshalb nur ca. 150 - 200 Aminosäuren umfassen. Bei der dritten Klasse, den gespaltenen Inteinen, sind die N- und C-terminalen Spleißbereiche (100 - 150 bzw. 35 - 50 AS) auf zwei verschiedene Vorläuferproteine aufgeteilt.



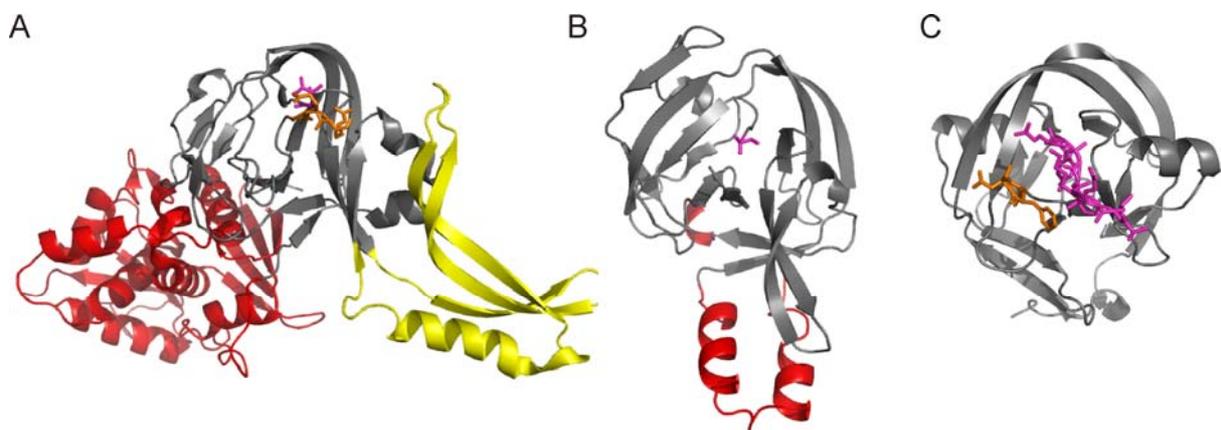
**Abbildung 18 Schematische Darstellung der drei Inteinklassen.** Die konservierten Sequenzmotive (nach [126]), die in diesen Bereichen hauptsächlich anzutreffenden Aminosäuren sowie deren Nummerierung sind gezeigt. Die dargestellten Sequenzmotive der Homing Endonuklease beziehen sich auf die Familie der DOD Homing Endonukleasen. Von den vier bisher bekannten Endonukleasefamilien findet sich diese am häufigsten in Inteinen.

Der evolutionäre Ursprung der Inteine ist weitgehend unbekannt. Eine Hypothese lautet, dass die Inteine aus einer Fusion zweier selbst-spaltender Proteine entstanden sein könnten [127]. In diesem Szenario hätte das Fusionsprotein anschließend im Laufe der Zeit zusätzlich zur Spaltung des N- und des C-Exteins die Fähigkeit zur Ligation der flankierenden Bereiche entwickelt. Für eine effiziente Verbreitung eines Inteingens im Genpool war wahrscheinlich

zusätzlich die Insertion einer Homing Endonuklease zwischen die N- und C-terminalen Spleißbereiche erforderlich [128,129]. Diese erlaubt es dem Intein nach dem Eindringen in einen anderen Organismus durch einen DNA-Doppelstrangbruch die Intein-freien Allele gegen das Intein enthaltende Gen auszutauschen (*intein homing*) [130]. Für ein nachträgliches Eindringen der Homing Endonukleasen in Inteingene spricht auch der Befund, dass verschiedene Endonukleasefamilien in Inteinen gefunden wurden [131]. Nach der Etablierung eines Inteingens im Genpool einer Spezies lastet kein Selektionsdruck auf der Funktionalität der Homing Endonuklease. Im Einklang damit wurden inaktive Varianten der Homing Endonuklease in Inteinen gefunden. Darüberhinaus wurde postuliert, dass die heutigen Mini-Inteine aus einer vollständigen Deletion dieser Domäne entstanden sind [131]. Eine genetische Umorganisation könnte zur Aufteilung der Spleißbereiche auf unterschiedliche chromosomale Bereiche und somit zum Entstehen der natürlich gespaltenen Inteine geführt haben [131]. Im Gegensatz zur Homing Endonuklease besteht für das Intein weiterhin ein Selektionsdruck zur Erhaltung der Spleißfähigkeit, da ansonsten kein aktives Wirtsprotein hergestellt werden kann, das in vielen Fällen essentiell für das Überleben des Organismus ist. Dass Inteine jedoch ihren Wirts-Organismen einen weiteren Selektionsvorteil, wie z.B. eine zusätzliche Regulation der Expression verschaffen, konnte bis heute nicht nachgewiesen werden. Man bezeichnet sie deshalb als parasitäre oder egoistische genetische Elemente [125,129].

Trotz einer sehr geringen Sequenzidentität besitzen alle spleißaktiven Bereiche der Inteine eine große strukturelle Homologie. Dieser Faltungstyp wird als HINT-Modul bezeichnet [132], da man ihn zusätzlich zu den Inteinen bei der C-terminalen Autoprozessierungsdomäne der Hedgehog-Proteine findet [132]. 12 - 14  $\beta$ -Faltblätter bilden eine hufeisenförmige Struktur mit einer pseudo-zweizähligen Achse. Abbildung 19 zeigt ein strukturelles Beispiel für jede der drei Inteinklassen. Die katalytischen Zentren befinden sich jeweils in der Mitte der Hufeisenstrukturen an den Enden zweier benachbarter antiparalleler  $\beta$ -Faltblätter. Am Beispiel des *Sce* VMA Inteins aus *Saccharomyces cerevisiae*, das wohl am besten charakterisierte Maxi-Intein, ist besonders gut zu erkennen, dass die Spleißbereiche eine eigene abgegrenzte Domäne bilden (graue Domäne in Abbildung 19A) [133]. Sowohl die Homing Endonuklease (rot), als auch das DNA Bindemotiv (gelb), sind in die Schleifen-Regionen zwischen zwei  $\beta$ -Faltblättern integriert und beeinträchtigen die Faltung nicht. Für mehrere Maxi-Inteine konnte gezeigt werden, dass das Entfernen der Endonuklease nicht die Spleißaktivität beeinträchtigt [134-136]. An der Position der Endonuklease befindet sich im

natürlichen Mini-Intein *Mxe* GyrA (s. Abbildung 19B) ein 51 Aminosäuren großer Linkerbereich (rot), welcher aus einer unvollständigen Deletion der Endonuklease hervorgegangen sein könnte. Die Kristallstruktur eines natürlich gespaltenen Inteins, dem *Ssp* DnaE Intein, ist in Abbildung 19C dargestellt. Für die Kristallisation wurden die N- und C-terminalen Spleißbereiche ohne eine Einführung von Linkeraminosäuren fusioniert. Interessanterweise war das entstandene *cis*-Intein spleißaktiv und zeigte den typischen HINT-Faltungstyp [137]. Dies unterstützt die Annahme einer möglichen Abstammung der gespaltenen Inteine von ehemaligen Mini-Inteinen durch chromosomale Umorganisation (s. oben).



**Abbildung 19 Intein-Kristallstrukturen.** A) Das Maxi-Intein *Sce* VMA Intein aus *Saccharomyces cerevisiae* (pdb-Code 1JVA) [138]; B) Das natürliche Mini-Intein *Mxe* GyrA aus *Mycobacterium xenopi* (pdb-Code 1AM2) [139]; C) Das eigentlich natürlich gespalte *Ssp* DnaE Intein aus *Synechocystis* species PCC6803, dessen Spleißbereiche fusioniert vorliegen (pdb-Code 1ZDE) [137]; In allen drei Strukturen sind die spleißaktiven Bereiche in grau, sowie die N- und C-Exteinreste (soweit vorhanden) in violett, bzw. in orange in der Stäbchendarstellung gezeigt.

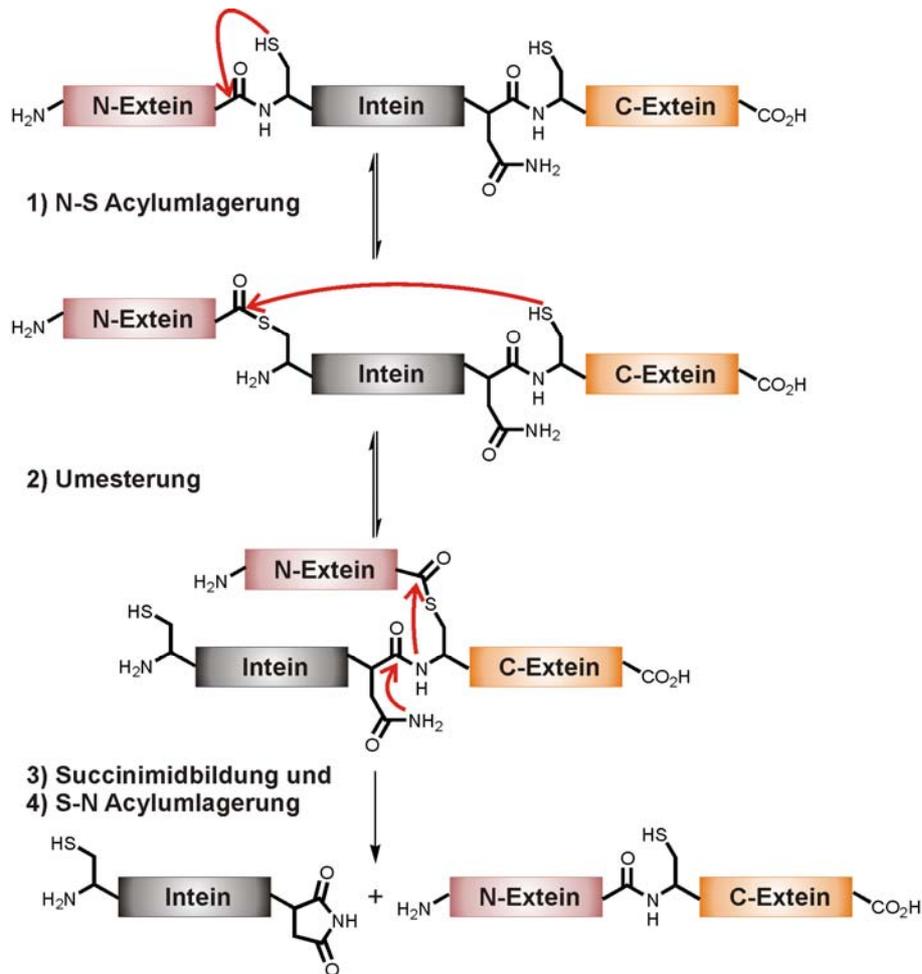
Inteine sind im Laufe des letzten Jahrzehnts zu einem interessanten Werkzeug für den segmentellen Aufbau und/oder die selektive Modifikation von Proteinen geworden. Zwei wichtige Eigenschaften waren dafür von entscheidender Bedeutung. Erstens besitzen Inteine die Fähigkeit auch mit nicht-nativen Exteinresten die Spleißreaktion eingehen zu können. Ein Intein ist demnach ein genetisches Element, welches leicht in verschiedene Zielproteine eingebaut werden kann. Eine gewisse Abhängigkeit der Spleißaktivität vom heterologen Kontext besteht jedoch, die in Kapitel 1.2.2 diskutiert wird. Die zweite erstaunliche Eigenschaft ist die Möglichkeit der Konstruktion von künstlich gespaltenen Inteinen ausgehend von natürlichen Maxi- oder Mini-Inteinen. Sie haben die Einsatzbreite der gespaltenen Inteine in Feldern wie der Herstellung zyklischer Peptide, der segmentellen Isotopenmarkierung von Proteinen oder der Protein-Semisynthese entscheidend erhöht. Auf weitere Anwendungsmöglichkeiten gespalteener Inteine sowie entscheidende Aspekte dieser

Proteinklasse im Allgemeinen wird in den Kapiteln 1.2.3 und 1.2.4 detaillierter eingegangen. Im nächsten Kapitel soll zunächst der Mechanismus der Protein-Spleißreaktion vorgestellt werden.

### 1.2.1 Der Mechanismus der Protein-Spleißreaktion

Ein bemerkenswerter Aspekt der Protein-Spleißreaktion ist, dass in deren Verlauf weder Kofaktoren noch zusätzliche Energie in Form von ATP benötigt werden. Der Vorgang wird demnach allein durch die Primärsequenz des Inteins und seiner drei-dimensionalen Struktur zusammen mit der ersten Aminosäure des Exteins an der C-terminalen Spleißstelle bestimmt. In den 90er Jahren gelang es hauptsächlich durch Etablierung eines *in vitro* Spleißsystems und durch Mutagenesestudien essentieller Reste, den Mechanismus der Protein-Spleißreaktion aufzuklären [140-143]. Abbildung 20 zeigt die bis heute allgemein akzeptierten Reaktionsschritte zur Bildung des Spleißproduktes am Beispiel eines Inteins mit zwei Cysteinresten an den nukleophilen Positionen [122,123].

Der erste Schritt ist die Bildung eines linearen Thioesterintermediats durch einen nukleophilen Angriff der Aminosäure am N-Terminus des Inteins (Position 1, Motiv A) auf die vorangehende Peptidbindung. Zur Katalyse dieses energetisch sehr ungünstigen Reaktionsschritts sind die Aminosäuren im Sequenzmotiv B von enormer Wichtigkeit. Nach Aufklärung der Kristallstruktur des *Mxe* GyrA Inteins [139] wurde deutlich, dass sich ein essentielles Histidin als Bestandteil einer konservierten TXXH Sequenz innerhalb von Motiv B in geeigneter Position befindet, um den Amidstickstoff der zu spaltenden Peptidbindung zu protonieren. Zusätzlich liegt nach NMR-Untersuchungen diese Peptidbindung in einer verzerrten *trans*- oder sogar in der *cis*-Konfiguration vor [144]. Beide Faktoren erleichtern den nukleophilen Angriff der Seitenkette der Aminosäure an Position 1. Das gebildete tetraedrische Intermediat, ein Oxythiazolidin, wird wahrscheinlich durch ein von den weiteren Resten des Motivs B (TXX) aufgebautes „Oxyanionenloch“ stabilisiert [139]. Eine weitere Bestätigung für diesen Reaktionsablauf liefert der Nachweis eines Thiazolidin-Nebenproduktes bei der Reaktion des an Position 11 gespaltenen *Ssp* DnaB-Inteins mit einem unnatürlichen Exteinrest an der Position -1 [145].



**Abbildung 20 Mechanismus des Proteinspleißens.** Vier hoch-koordinierte nukleophile Substitutionen führen zur Spaltung der Peptidbindungen an den N- und C-terminalen Spleißstellen, sowie zur kovalenten Verknüpfung der beiden Exteine. Analog kann diese Reaktion auch mit Serin an Position 1 des Inteins, Serin oder Threonin an Pos. +1 oder Glutamin anstelle von Asparagin an der letzten Position des Inteins ablaufen (für Details s. Text).

Der zweite Reaktionsschritt im Laufe der Proteinspleißreaktion ist eine intramolekulare Transthoesterifizierung, in deren Verlauf das N-Extein von der Seitenkette von Cys1 auf die Seitenkette der +1 Aminosäure des C-Exteins übertragen wird (s. Schritt 2 in Abbildung 20). Für diese Gleichgewichtsreaktion wird aufgrund von Kristallstrukturdaten mit und ohne Exteinreste eine, wenn auch nur geringfügige, Umorientierung der katalytischen Zentren postuliert [146,147].

Für den letzten katalysierten Reaktionsschritt, die Asparaginzyklisierung, sind die Aminosäuren in den Sequenzmotiven F und G im C-terminalen Spleißbereich verantwortlich. Hier greift die Amidgruppe der Asparaginseitenkette am C-Terminus des Inteins nukleophil die Peptidbindung zwischen dem Intein und dem C-Extein an. Das dabei auftretende tetraedrische Intermediat reagiert in einer Weise weiter, die zur Bindungsspaltung im Peptidrückgrat und zur Freisetzung des nun als Succinimid vorliegenden Inteins führt. Kristallstrukturanalysen und Mutationsstudien unter anderem mit Hilfe des künstlich

erzeugten *Ssp* DnaB Mini-Intein identifizierten ein Ladungsverschiebungssystem, das wahrscheinlich durch Deprotonierung die Nukleophilie der Asparaginseitenkette erhöht und zusätzlich ein „Oxyanionenloch“ bildet, um das tetraedrische Intermediat zu stabilisieren. Eine verdrehte Peptidbindung zwischen Intein und Extein erleichtert zusätzlich den nukleophilen Angriff des Asparagins. Daneben wurde eine Stabilisierung der Abgangsgruppe, also der N-terminalen Aminogruppe des C-Exteins, postuliert [146]. Keine Aminosäure in den Sequenzmotiven F und G ist zwischen allen Inteinen zu 100% konserviert, sodass angenommen werden kann, dass für jedes Intein der Ablauf des Ladungsverschiebungssystem zwar mechanistisch sehr ähnlich ist, aber strukturell je nach Anordnung der Aminosäuren unterschiedlich sein kann.

Nach der Asparaginzyklisierung liefert eine spontane S-N Acylwanderung die native Peptidbindung zwischen den beiden Exteinsequenzen und verschiebt die davor ablaufenden Gleichgewichtsreaktionen auf die Seite der Produkte (s. Schritt 4 in Abbildung 20). Die vier Schritte des Spleißmechanismus müssen exakt aufeinander abgestimmt sein. Eine Störung in der zeitlichen Abfolge der Reaktionen kann Nebenreaktionen zur Folge haben. Eine Zyklisierung des Asparagins vor der Bildung des verzweigten Intermediates verhindert eine Ligation der Exteine und bedingt nur die Freisetzung des C-Exteins, die so genannte C-terminale Spaltung. Diese kann, analog zu Hydrolysereaktionen der Thioesterintermediate an der N-terminalen Spaltstelle (N-terminale Hydrolyse), auch durch Mutation von essentiellen Aminosäureresten oder durch nicht-natürliche Exteinsequenzen verursacht werden.

### 1.2.2 Abhängigkeit der Protein-Spleißreaktion von heterologen Exteinsequenzen

In verschiedenen Arbeiten konnte gezeigt werden, dass die Proteinspleißreaktion auch in einem vollständig unnatürlichen Exteinkontext stattfinden kann und dass die Spleißaktivität hauptsächlich auf das sich herauschneidende Intein zurückgeht [140,148]. Dennoch läuft die Spleißreaktion nicht spurlos ab, da die essentielle Aminosäure an Position +1 des C-Exteins im Spleißprodukt verbleibt. Zusätzlich konnte in diversen Mutagenesestudien gezeigt werden, dass die Aminosäuren in direkter Nähe zum Intein (-3, -2, -1, +2, +3) das Proteinspleißen in Bezug auf die Ausbeute, den Anteil der Nebenreaktion und die Reaktionsgeschwindigkeit beeinflussen können [149-152]. In den meisten Fällen unterstützt die Verwendung von drei bis fünf nativen Aminosäureresten an beiden Spleißstellen die Aktivität des Inteins [153].

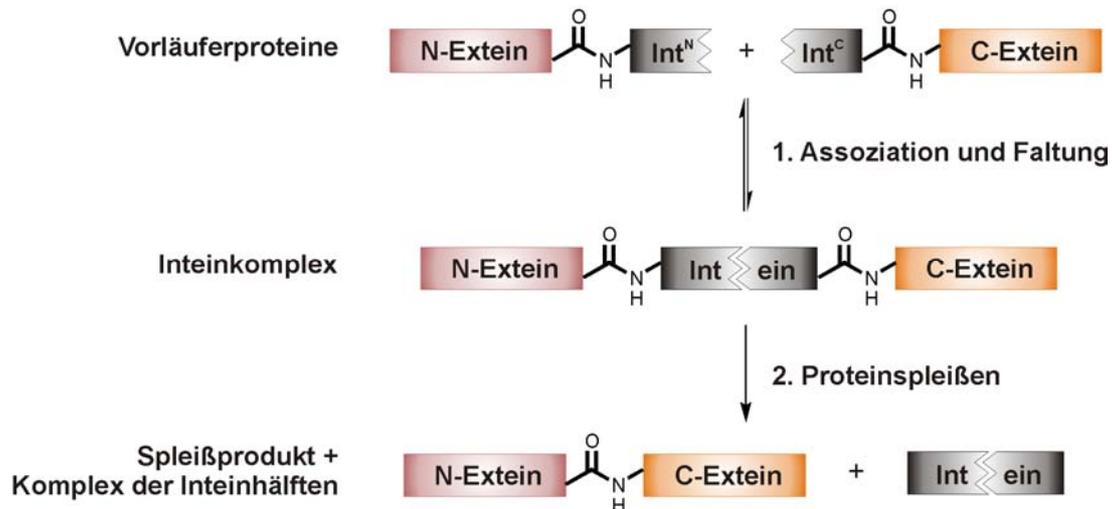
Andere Aminosäuren, die sich in der Primärsequenz in größerer Entfernung zum Intein befinden, scheinen nicht am Spleißprozess beteiligt zu sein. Kristallstrukturen von nicht gespleißten Vorläuferproteinen geben bisher keine definitiven Hinweise in welcher Art und Weise die flankierenden Aminosäuren das Intein beeinflussen. Zwar konnte z.B. das *Ssp* DnaE Intein zusammen mit acht N-terminalen und drei C-terminalen Exteinresten kristallisiert werden (s. Abbildung 19C) [137]. Jedoch ließen sich keine definitiven Voraussagen ableiten, wie der Exteinkontext das Intein beeinflusst. Die evolutionäre Anpassung eines Intein-Exteinpaars und die große Varianz der natürlichen Exteinreste an allen Positionen machen es unwahrscheinlich, dass in nächster Zeit generelle *ab initio* Aussagen zur Spleißaktivität der Inteine in einem nicht-natürlichen Exteinkontext getroffen werden können.

### 1.2.3 Gespaltene Inteine

Unter den natürlich vorkommenden Inteinen ist die Klasse der gespaltenen Inteine die kleinste. Mit Ausnahme des *Nanoarchaeum equitans* (*Neq*) B-Typ DNA Polymerase Inteins [154] gehören alle bisher gefundenen und charakterisierten gespaltenen Inteine zu einem Allel, welches in die  $\alpha$  Untereinheit der DNA Polymerase III (DnaE) in diversen Cyanobakterien inseriert ist [155-157]. Sequenzanalysen mit metagenomischen Daten von Tiefseeorganismen deuten jedoch darauf hin, dass sich die Klasse der natürlich gespaltenen Inteine in Zukunft erweitern wird [131]. Mechanistisch verläuft das Protein *trans*-Spleißen (PTS) analog zu dem in Abbildung 20 dargestellten Proteinspleißen der Maxi- und Mini-Inteine. Zur Ausbildung eines aktiven Inteinkomplexes müssen jedoch die beiden auf separaten Proteinen vorliegenden Inteinhälften zuerst assoziieren und die korrekte Faltung annehmen (s. Abbildung 21).

Ungefähr zeitgleich mit der Entdeckung des ersten natürlich gespaltenen Inteins, dem DnaE Intein aus *Synechocystis species* PCC6803 (*Ssp*) [155], wurden ausgehend von Maxi- und Mini-Inteinen die ersten künstlich gespaltenen Inteine erzeugt [135,136,149,158]. Je nach Intein und ausgesuchter Spaltstelle variiert die Aktivität dieser artifiziell gespaltenen Ligationssysteme stark. Die ersten an der Position der Endonuklease gespaltenen Inteine waren zwar spleißaktiv, benötigten aber einen Denaturierungsschritt und anschließend eine gemeinsame Rückfaltung beider Inteinhälften, um einen spleißkompetenten Komplex *in vitro* auszubilden [149,158]. Problematisch war vor allem, dass die individuell exprimierten Inteinhälften zur Aggregation oder Fehlfaltung neigten. Das Fehlen einer natürlichen Affinität

der getrennten N- und C-terminalen Spleißbereiche im *Sce* VMA Intein konnte zur Entwicklung eines konditionellen Spleißsystems ausgenutzt werden [159-161]. Hier ermöglichte erst eine durch fusionierte Affinitätsdomänen induzierte räumliche Nähe der Inteinhälften die Bildung eines aktiven Inteinkomplexes. Die Spleißaktivität konnte damit durch Zugabe eines kleinen Liganden, welcher entweder die Komplexbildung der fusionierten Affinitätsdomänen auslöst oder verhindert, ein- oder ausgeschaltet werden [160,162].



**Abbildung 21 Schematische Darstellung des Protein *trans*-Spleißens.** Nach einer Assoziation der beiden Inteinhälften führt der Inteinkomplex analog zu Abbildung 20 die Proteinspleißreaktion durch.

In den letzten Jahren wurden weitere entweder natürliche oder durch Deletion der Homing Endonuklease erzeugte Mini-Inteine an der Position gespalten, an der in homologen Sequenzen eine Endonuklease insertiert vorliegt. Interessanterweise zeigten einige von ihnen im Gegensatz zu den oben genannten Beispielen eine Spleißaktivität auch unter komplett nativen Bedingungen. So erlaubte die natürliche Affinität der Inteinhälften sowohl des *Ssp* DnaB Inteins [163,164], als auch des *Mxe* GyrA Inteins [165], selektive Ligationsreaktionen ohne Denaturierungs- und Rückfaltungsschritte. Zusätzlich zur Position der Endonuklease kann die Inteindomäne auch an anderen Positionen gespalten werden, von denen einige ebenfalls aktive Protein *trans*-Spleißsysteme hervorbrachten [166-168]. Erwähnenswert sind vor allem zwei künstlich gesplattene Inteinvarianten, die jeweils zu sehr kurzen Inteinhälften führten. So lässt sich das *Ssp* DnaB Intein ohne den Verlust der Spleißaktivität in einen 11 Aminosäuren großen N-terminalen und einen 144 AS großen C-terminalen Teil zerlegen [166,169,170]. Die kürzeste noch aktive Spaltstelle vom C-Terminus eines Inteins aus wurde in dem *Ssp* GyrB Intein (Untereinheit B der DNA-Gyrase aus *Synechocystis* sp. PCC6803) gefunden [171]. Hier reichen 6 Aminosäuren aus, um das ~ 150 AS große N-terminale Fragment zu komplementieren und die Protein *trans*-Spleißreaktion durchzuführen. Auf die

Wichtigkeit dieser beiden Systeme für die Proteinsemissynthese wird noch detailliert in Kapitel 1.2.4 eingegangen. Ein gemeinsames Charakteristikum aller bisher untersuchten gespaltenen Inteinen ist, dass ihre einzelnen Hälften alleine keine erkennbare Reaktivität zeigen. Darüber hinaus eignen sie sich zusätzlich für *in vivo* Anwendungen, da aufgrund ihres peptidischen Charakters keine Wechselwirkungen mit anderen Zellbestandteilen, mit Ausnahme von Proteasen und der Möglichkeit zur Disulfidbildung, zu erwarten sind [170,172]. Erst durch das Vorhandensein der komplementären Hälfte kann die PTS-Reaktion stattfinden. Kreuzreaktivitäten zwischen verschiedenen gespaltenen Inteinen wurden bisher nur innerhalb von Inteinallelen beobachtet [157] - dies macht das Protein *trans*-Spleißen zu einer höchst chemoselektiven Ligationsreaktion.

Unterschiede zwischen gespaltenen Inteinen sind jedoch sowohl bei der Betrachtung der Geschwindigkeit und Art der Assoziation der Inteinhälften als auch bei der Gesamtreaktionsgeschwindigkeit vorhanden. Sowohl bei den natürlich gespaltenen DnaE Inteinallelen, als auch bei dem an der Position der Endonuklease gespaltenem *Ssp* DnaB Intein scheint die Assoziation der Hälften sehr schnell und über elektrostatische Wechselwirkungen abzulaufen [163,173-175]. Für das *Ssp* DnaE Intein wurde ein Wert für die Komplexbildung nahe des Diffusionslimits nach der Einstein-Smoluchowski Gleichung bestimmt [174]. Weiterführende Studien grenzten die elektrostatischen Wechselwirkungen auf lokale Ladungcluster auf gegenüberliegende Stellen in den langen  $\beta$ -Faltblättern ein [157,176]. Im Gegensatz dazu ist die Assoziation des künstlich an Position 11 gespaltenen *Ssp* DnaB Inteins zum einen um ca. sechs Größenordnungen langsamer und zum anderen sind hier hauptsächlich hydrophobe Wechselwirkungen für die Komplexbildung verantwortlich [145]. Die große Stabilität eines einmal gebildeten Inteinkomplexes wird durch geringe  $K_D$ -Werte im mikro- bis nanomolaren Bereich und kleine Geschwindigkeitskonstanten der Dissoziation dokumentiert [145,174].

Im Gegensatz zu den Assoziationsraten weichen die oftmals mit einem Geschwindigkeitsgesetz erster Ordnung analysierten Gesamtreaktionsraten zwischen den gespaltenen Inteinen weniger voneinander ab (s. Tabelle 1). Die Halbwertszeiten der Protein *trans*-Spleißreaktionen bewegen sich im Bereich von Minuten bis Stunden und sind vergleichbar mit den bisher bestimmten Kinetiken für *cis*-spleißende Inteine. Jedoch existiert für letztere nur ein begrenzter Datensatz, da es oft nicht möglich ist ein noch nicht gespleißtes Vorläuferprotein zu isolieren.

**Tabelle 1** Geschwindigkeitskonstanten erster Ordnung für natürlich und künstlich gespaltene Inteine.

Intein	T /°C	k /s <sup>-1</sup>	t <sub>1/2</sub> /min	Literatur
<i>Ssp</i> DnaE	23	(6,6 ± 1,3) × 10 <sup>-5</sup>	175	[173]
<i>Ssp</i> DnaE	23	(3,3 ± 0,4) × 10 <sup>-4</sup>	35	[177]
<i>Sce</i> VMA	25	(2,0 ± 0,3) × 10 <sup>-3</sup>	6	[163]
<i>Ssp</i> DnaB (Endo)	25	(9,9 ± 0,8) × 10 <sup>-4</sup>	12	[163]
<i>Mtu</i> RecA	30	n.b.	60 - 120	[178]
<i>Ssp</i> DnaB (11)	25	(4,1 ± 0,2) × 10 <sup>-5</sup>	280	[169]
<i>Ssp</i> GyrB (6)	25	(6,9 ± 2,2) × 10 <sup>-5</sup>	167	[171]
GBD Pol-1 (cis)	37	n.b.	30	[141]
<i>Pab</i> Pol II (cis)	70	1,5 × 10 <sup>-5</sup>	722	[179]

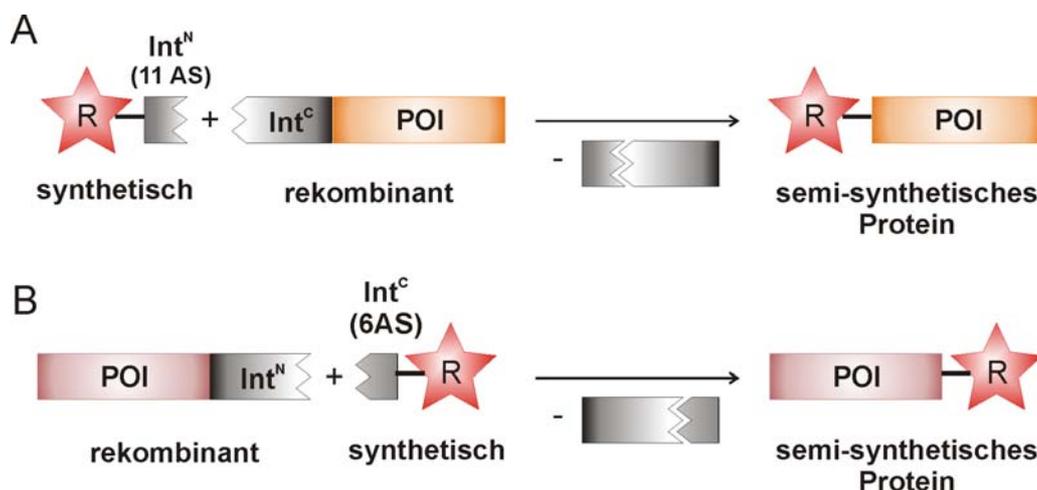
(n.b. = nicht berichtet; *Ssp* DnaB (Endo) = an der Position der Endonuklease gespaltenes *Ssp* DnaB Mini-Intein; *Ssp* DnaB (11) = 11 AS vom N-Terminus gespaltenes *Ssp* DnaB Mini-Intein; *Ssp* GyrB (6) = 6 AS vom C-Terminus gespaltenes *Ssp* GyrB Mini-Intein).

Bislang sind wenige strukturelle Informationen über gespaltene Inteinhälften ohne das Partnerprotein vorhanden. Dementsprechend ist nicht geklärt, ob eine Vorfaltung der isolierten Inteinhälften vorliegt und ob diese die Bildung des Komplexes beschleunigt oder verlangsamt. Ferner sind mögliche und postulierte Faltungsintermediate auf dem Weg zu einem spleißaktiven Komplex experimentell schwer zugänglich und konnten bisher nicht zweifelsfrei nachgewiesen werden [174]. Auch ohne ein exaktes Verständnis des Protein *trans*-Spleißens auf struktureller Ebene sind gespaltene Inteine zu einem sehr wichtigen Werkzeug in der gezielten Modifikation von Peptiden und Proteinen geworden. Ihre breiten Anwendungsmöglichkeiten sind Gegenstand des nächsten Kapitels.

#### 1.2.4 Biochemische Anwendungen gespaltener Inteine

Gespaltene Inteine können wie oben beschrieben zur Ligation von zwei Peptid- oder Proteinfragmenten eingesetzt werden. Eine getrennte Synthese und Präparation der beiden Inteinhälften erlaubt damit den segmentellen Aufbau eines Proteins aus mehreren Fragmenten, die unterschiedliche Modifikationen enthalten können. Bei der Proteinsemisynthese wird eine Inteinhälfte zusammen mit dem Zielprotein durch Expression in *E. coli* rekombinant hergestellt. Die andere Hälfte ist durch Festphasenpeptidsynthese zugänglich und erlaubt den Einbau diverser chemischer Gruppen und Funktionalitäten in die betreffende Exteinsequenz. In der anschließenden PTS-Reaktion wird dieses modifizierte Extein auf das Zielprotein übertragen. Durch die Größenlimitation der

Festphasenpeptidsynthese (40 - 50 AS-Bausteine) waren die ersten Reaktionen mit dem künstlich gespaltenen *Mtu* RecA (38 AS großes C-Intein [178]) und dem natürlich gespaltenen *Ssp* DnaE Intein (36 AS großes C-Intein [180]) auf ein kleines, synthetisches C-Extein beschränkt. Jedoch gelang vor allem mit dem *Ssp* DnaE Intein eindrucksvoll der Aufbau von semi-synthetischen Proteinen [172,181], von denen hier exemplarisch die C-terminale Modifikation eines Prionenproteins mit zwei Palmitinsäureresten, einem Mimikri eines Glykosylphosphatidylinositol (GPI) Ankers erwähnt werden soll [182]. Die Entdeckungen von spleißaktiven gespaltenen Inteinen mit einem sehr kurzen Inteinfragment (s. oben) ermöglichten den Zugang zur N-terminalen Modifikation von Proteinen [169,170] und zur vereinfachten Darstellung der C-terminalen Hälfte [171]. Diese Ansätze basieren sowohl auf dem *Ssp* DnaB als auch dem *Ssp* GyrB Intein und sind schematisch in Abbildung 22 dargestellt.



**Abbildung 22 Proteinsemisynthese mit Hilfe künstlich gespalteener Inteine.** A) Schematische Darstellung der N-terminalen Modifikation von Proteinen mit dem *Ssp* DnaB Intein (gespalten an Position 11) [169]; B) Schematische Darstellung der C-terminalen Modifikation von Proteinen mit dem *Ssp* GyrB Intein (der C-terminale Teil des aktiven Inteins ist nur 6 AS groß) [171]; POI = Protein von Interesse; R = gewünschte chemische Modifikation (Fluorophore, Biotin, reaktive Gruppe, etc.).

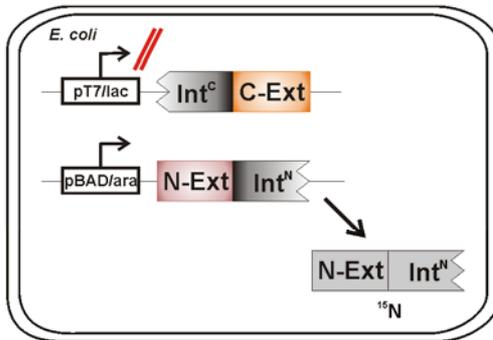
Einige Aminosäureseitenketten lassen sich selektiv in Gegenwart der anderen funktionellen Gruppen von Proteinen mit Markierungsreagenzien adressieren [183]. Lysin, Tyrosin und vor allem Cystein besitzen eine einzigartige Reaktivität. Sind jedoch mehrere Aminosäurereste eines Typs in einem Protein vorhanden, so kann keine regioselektive und stöchiometrische Markierung mehr erfolgen. Im Falle der Cysteinmodifikationen wurden in den letzten Jahren Ansätze entwickelt, die mittels gespalteener Inteine dieses Problem umgehen konnten [164,165,184]. Das zu markierende Protein wird so geteilt, dass nur noch ein Cysteinrest in einem Fragment vorhanden ist. Nach der rekombinanten Expression und Reinigung der Fusionsproteine wird nur eine Hälfte mit dem Markierungsreagenz behandelt.

In der anschließenden PTS-Reaktion wird somit ein modifiziertes Segment mit einem unmodifiziertem Segment verknüpft [184] - dieses Reaktionsschema ist somit analog zu einer regioselektiven Adressierung einer Seitenkette in Gegenwart von weiteren identischen Funktionalitäten.

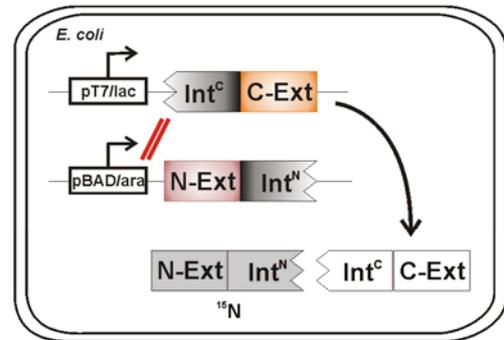
Die Verwendung von gespaltenen Inteinen zur segmentellen Isotopenanreicherung in Proteinen für NMR-Untersuchungen war eine der ersten Anwendungen für diese Proteinklasse überhaupt [185]. Die Strukturaufklärung von Proteinen in Lösung mittels NMR kann aufgrund ihrer Größe auf schwerwiegende Probleme treffen. Zum einen erzeugen langsam rotierende Moleküle, wie große Proteine oder Proteinkomplexe, Signale mit einer großen Linienbreite. Zum anderen kann bei einer zu ausgeprägten Überlagerung der Resonanzen keine exakte Zuordnung der Signale mehr erfolgen. Letzteres Problem, das durch die schiere Zahl an Resonanzen oder durch kaum disperse Spektren verursacht wird, kann mit Hilfe von gespaltenen Inteinen gelöst werden. Die getrennte Expression von Proteinfragmenten ermöglicht die Herstellung eines Segments, welches aufgrund der Anzucht in speziellem Nährmedium mit NMR-aktiven Isotopen meistens aber  $^{15}\text{N}$  und  $^{13}\text{C}$  angereichert ist. Das komplementäre Fragment kann in herkömmlichem Medium exprimiert werden und ist folglich nicht markiert. Nach der PTS-Reaktion ist im ligierten Protein nur ein Teil „NMR-sichtbar“ und damit die Wahrscheinlichkeit einer Signalüberlagerung deutlich reduziert. Yagi *et al.* nutzten dieses Prinzip elegant zur Untersuchung von konformationellen Änderungen der 50 kDa großen  $\beta$ -Untereinheit der  $\text{H}^+$ -ATPase [186]. Mit Hilfe des PI-*PfuI* Inteins aus *Pyrococcus furiosus* gelang die Zuordnung von ca. 90 % der NMR-Signale (Amid-NH,  $\text{C}_\alpha$ ,  $\text{C}_\beta$ , CO von 473 Aminosäuren) und darauf aufbauend die Bestimmung einer offenen und einer geschlossenen Form der  $\beta$ -Untereinheit der  $\text{H}^+$ -ATPase in Abhängigkeit von ADP und ATP [186]. Zwei weitere Ansätze erhöhen die Einsetzbarkeit von gespaltenen Inteinen für die segmentelle Isotopenmarkierung von Proteinen. Zum einen ist es möglich, mit Hilfe zweier orthogonaler Inteine durch sequentielles PTS auch zentrale Bereiche eines Proteins selektiv zu markieren (s. Abbildung 23B) [185,187]. Zum anderen bietet ein von Iwai und Kollegen entwickeltes duales Expressionssystem die Möglichkeit zur segmentellen Isotopenmarkierung *in vivo* [188,189]. Dabei kann mit zwei unabhängig induzierbaren Promotoren die Expression der Inteinhälften in *E. coli* gezielt gesteuert werden, sodass in Kombination mit einem Mediumswechsel das segmentell markierte Spleißprodukt in einem Expressionsansatz entsteht (s. Abbildung 23A). Eine zeit- und arbeitsaufwendige Reinigung der individuellen Hälften entfällt somit [188,189].

A

Medium 1 (z.B. M9 mit  $^{15}\text{N}$ )  
+ Arabinose



Medium 2 (z.B. LB)  
+ IPTG

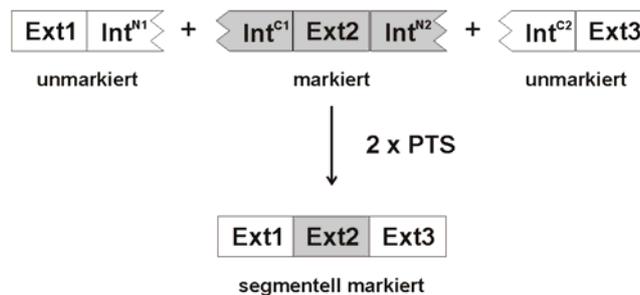


Mediumswechsel

*In vivo* PTS  
Proteinreinigung

Segmentelle  
Isotopenmarkierung  
 $^{15}\text{N}$

B

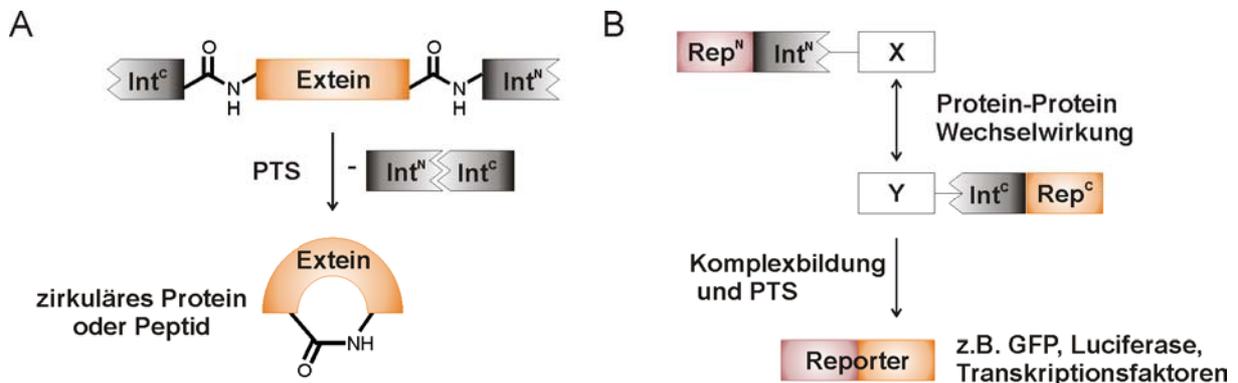


**Abbildung 23 Anwendungen von gespaltenen Inteinen für die segmentelle Isotopenmarkierung von Proteinen für NMR-Untersuchungen.** A) Von Iwai und Kollegen entwickeltes duales Expressionssystem zur segmentellen Isotopenmarkierung von Proteinen *in vivo* [188,189]. Über die pT7/lac und pBAD/ara Promoter/Operator-Sequenzen gelingt die selektive Induktion der Expression von jeweils nur einem Fusionsprotein. In Kombination mit einem Mediumswechsel wird nur eine Hälfte mit NMR-aktiven Kernen ( $^{15}\text{N}$ ,  $^{13}\text{C}$ ) angereichert. Nach PTS in der Zelle kann das segmentell isotopenangereicherte Protein in einem Schritt erhalten werden; B) Schematische Darstellung der Isotopenmarkierung eines internen Proteinfragments mit zwei orthogonalen Inteinen [185,187].

Weitere sehr interessante Anwendungsmöglichkeiten gespalteener Inteine sollen im Folgenden nur kurz angesprochen werden. Durch eine Änderung der Intein-Extein Anordnung auf genetischer Ebene gelingt mit Hilfe von  $\text{Int}^{\text{C}}$ -Extein- $\text{Int}^{\text{N}}$  Konstrukten die Synthese von zyklischen Peptiden und Proteinen (s. Abbildung 24A). Durch eine Zyklisierung des Peptidrückgrates entstehen proteolyse- und thermostabile Proteine [180,190-192]. Weiterhin konnten mit der SICLOPPS Technologie (*split intein circular ligation of peptides and proteins*) zyklische Peptidbibliotheken unterschiedlicher Größe (4 bis 9 AS) erzeugt werden [193,194]. Interessanterweise schien das in diesen Studien verwendete *Ssp DnaE* Intein bei den Zyklisierungsreaktionen toleranter im Vergleich zum normalen PTS in Bezug auf seine

Exteinsequenzen zu sein. Eine interessante Anwendung zur Detektion von Proteaseaktivitäten in Zellkultur und in lebenden Mäusen gelang 2007 mit einer zyklischen Luciferase. Dieses durch PTS mit dem *Ssp* DnaE Intein hergestellte Protein war in seiner zyklischen Form inaktiv. Jedoch enthielt der unnatürliche Verbindungsbereich zwischen den beiden Subdomänen (s. ANL Superfamilie in Kapitel 1.1.4) eine Erkennungssequenz für die Protease Caspase-3. Eine proteolytische Spaltung erzeugte eine aktive Luciferase und ermöglichte damit das Auslesen der Proteaseaktivität durch Biolumineszenz (s. Abbildung 8C) [195].

Die Gruppe um Umezawa entwickelte eine Vielzahl von Systemen zur Detektion von Protein-Protein Wechselwirkungen basierend auf gespaltenen Inteinen. Das Prinzip ist hierbei, dass die Komplexbildung zweier Proteine, die an sie fusionierten Inteinhälften in räumliche Nähe bringt. Nun verbindet das gespaltene Intein durch PTS die zwei Hälften eines Reporterproteins und aktiviert dieses dadurch (s. Abbildung 24B). Als Reporter wurden sowohl das grün fluoreszierende Protein (GFP), verschiedene Luciferasen, als auch Transkriptionsfaktoren eingesetzt [159,196,197]. Erstaunlicherweise konnte in diesen Fällen die intrinsische Affinität der Inteinhälften des *Ssp* DnaE Inteins noch signifikant durch die zusätzliche Protein-Protein Wechselwirkung verbessert werden. Die Methode erlaubte auch den Nachweis der Freisetzung von Proteinen aus Zellkompartimenten wie dem Mitochondrium in das Zytosol während der Apoptose [198].



**Abbildung 24 Herstellung von zyklischen Proteinen oder Peptiden und Detektion von Protein-Protein Wechselwirkungen mit Hilfe gespaltenener Inteine.** A) Durch eine  $\text{Int}^{\text{C}}$ -Extein- $\text{Int}^{\text{N}}$  Anordnung auf genetischer Ebene gelingt mit einer intramolekularen Protein „trans“-Spleißreaktion die Synthese zirkulärer Proteine und Peptide - als Nebenreaktionen sind Oligomerisierungen denkbar [180,190]; B) Die alleine inaktiven Teile eines Reportersystems ( $\text{Rep}^{\text{N}}$  und  $\text{Rep}^{\text{C}}$ ) sind mit den gespaltenen Inteinhälften, sowie den zu untersuchenden Proteinen X und Y fusioniert. Eine Protein-Protein Wechselwirkung zwischen den Proteinen X und Y verstärkt die Affinität der Inteinhälften und führt nach Komplexbildung und PTS zur vermehrten Bildung des Reporters (GFP, Luciferase, Transkriptionsfaktoren, etc.) [159,196,197].

In transgenen Pflanzen eingebrachte Resistenzgene können eine enorme Verbesserung im Vergleich mit Wildpflanzen darstellen. Um das Auskreuzen des funktionellen Transgens in Wildpflanzen zu erschweren, ist es möglich, dieses mit Hilfe gespaltenener Inteine in zwei

Hälften zu teilen. Nun wäre für das Auskreuzen die sehr unwahrscheinliche, simultane Übertragung der genetischen Informationen beider Genhälften notwendig [199-201]. Die Herstellung zytotoxischer Proteine wird durch den Einsatz gespaltener Inteine stark verbessert oder überhaupt erst zugänglich. Die nicht toxischen Proteinhälften können im Standardverfahren durch heterologe Genexpression erhalten werden. Durch eine *in vitro* PTS Reaktion kann das für lebende Zellen toxische Protein hergestellt werden [202].

Im Vergleich zu den gespaltenen Inteinen finden mutierte *cis*-Inteine bereits eine noch häufigere Anwendung in biochemischen, kommerziell erhältlichen Applikationen. Sie sind ein wichtiges Hilfsmittel in der Proteinreinigung und der Synthese von C-terminalen Proteinthioestern und Proteinen mit N-terminalem Cysteinen - enorm wichtige Bausteine für chemische Ligationsreaktionen, wie der Expressed Protein Ligation (EPL). Auch zu diesen Themen existieren eine Vielzahl hervorragender Übersichtsartikel, auf die hier jedoch nur hingewiesen werden kann [113,114,203,204].

Aufgrund der oben beschriebenen Anwendungen spielen gespaltene Inteine eine wichtige Rolle im Bereich der Chemischen Biologie. Ob sie sich auch in Zukunft zur Lösung von wichtigen biologischen und medizinischen Fragestellungen durchsetzen werden, hängt jedoch von den folgenden Herausforderungen ab: So wäre z.B. eine generelle Verbesserung der Reaktionsgeschwindigkeit der PTS-Reaktion ebenso wünschenswert wie stabilere Ausbeuten unter diversen Reaktionsbedingungen wie unterschiedliche Temperaturen, pH-Werte und Lösungsmittel. Eine größere Unabhängigkeit von den flankierenden Exteinsequenzen könnte die Einsatzbreite der gespaltene Inteine zusätzlich erhöhen.

### **1.3 Aufgabenstellung**

Dem detaillierten Wissen über einzelne NRPS-Domänen steht eine zwar wachsende, jedoch noch geringe Zahl an Veröffentlichungen zur Domänen-Domänen Interaktion in diesen Multi-Domänensystemen gegenüber. Zum besseren Verständnis der konformationellen Änderungen während des Katalysezyklus sollte die Wechselwirkung zwischen der A- und der *in cis* vorliegenden PCP-Domäne aus dem Gramacidin S Synthetase I Initiationsmodul mit biochemischen Methoden charakterisiert werden. Die weiteren Projekte dieser Doktorarbeit zielten darauf hin, die generelle Einsetzbarkeit von gespaltenen Inteinen für die selektive Modifikation von Proteinen zu verbessern. Dazu sollte der kritische Punkt der Exteinabhängigkeit durch die Entwicklung eines schnellen und unkomplizierten Systems zum Testen der Spleißaktivität eines gespaltenen Inteins in einem Zielprotein vereinfacht werden. Zusätzlich sollte das bisher nicht *in vitro* charakterisierte DnaE Intein aus *Nostoc punctiforme* biochemisch und mittels NMR-Spektroskopie untersucht werden.

---

## 2 Material

### 2.1 Geräte und Apparaturen

Tabelle 2 Liste der verwendeten Geräte.

Gerät	Hersteller mit Typenbezeichnung
Analysenwaagen	<i>Mettler</i> PM400 <i>Sartorius</i> Basic
Autoklav	<i>Thermo Fischer Scientific</i> Varioklav
Chromatographiesysteme	<i>GE Healthcare</i> Äkta prime plus und Äkta Purifier 10
Gefriertrocknungsanlage	<i>Christ</i> LT-105
Geldokumentationssystem	<i>Canon</i> PowerShot G5 <i>H. Saur</i> Transilluminator IL-200-M
Elektroporationssystem	<i>Bio-Rad</i> Gene Pulser und Pulse-Controller
Festphasenpeptidsynthesizer	<i>Multisyntech</i> Syro XP
French Press	<i>Avestin</i> EmulsiFlex <sup>®</sup> -C5 High pressure homogenizer
Heizblock	<i>Eppendorf</i> ThermoStat plus
HPLC-Anlage (analytisch)	<i>Hewlett Packard</i> Series 1100 Säule: Eclipse XDB-C18 3,5 µM
HPLC-Anlage (präparativ)	<i>Varian</i> ProStar Serie Säule: Microsorb 300-5 C18
Inkubationsschrank	<i>Memmert</i>
Inkubationsschüttler	<i>Infors</i> HT Aerotron <i>New Brunswick Scientific</i> I26, TC-7
MALDI-TOF	<i>Bruker Daltonics</i> Autoflex II
NMR-Geräte	<i>Bruker</i> Avance Spektrometer mit entweder 600, 700, 800, 900 oder 950 MHz und kryogenen Probenköpfen
PCR-Gerät	<i>Biometra</i> Tpersonal Kombi
pH-Meter	<i>HANNA instruments</i> pH211
Photometer	<i>Amersham Biosciences</i> Ultrospec 3100 <i>pro</i>
Pipetten	<i>Eppendorf</i> Reference
Reinstwasser Anlage	<i>Millipore</i> Milli-Q
SDS-PAGE Apparatur	<i>BioRad</i> Mini-PROTEAN 3 Cell <i>BioRad</i> PowerPac Basic Netzgerät
Ultraschallgerät	<i>Bandelin</i> Sonorex Digitec
Vakuumpumpe	<i>KNF Neuberger</i> Laboport SR
Vortexer	<i>LMS</i> VTX 300
Zentrifugen	<i>Eppendorf</i> Centrifuge 5424 <i>Eppendorf</i> Centrifuge 5702 R <i>Heraeus</i> Biofuge pico <i>Sorvall</i> Evolution

## 2.2 Chemikalien, Enzyme und Verbrauchsmaterialien

Hersteller	Produkt(e)
<i>Acros Organics</i> (New Jersey, USA)	Imidazol, N-Methylmorpholin
<i>Anaspec</i> (San Jose, USA)	Tetramethylrhodamin-5-Maleinimid
<i>AppliChem GmbH</i> (Darmstadt)	Glycin, IPTG, Tris, Trypton, HCl, Ampicillin, Kanamycin, Isopropanol, Fe(III)Cl <sub>3</sub> , ATP, AMP, MgCl <sub>2</sub> , Thiamin-Hydrochlorid, CoA
<i>Biosolve</i> (Valkenswaard, Niederlande)	DMF
<i>Bruker Daltonics</i> (Bremen)	Peptide Calibration Standard I
<i>Cambridge Isotope Labs</i> (Andover, USA)	<sup>15</sup> N-Ammoniumchlorid
<i>Deutsche Hefewerke</i> (Nürnberg)	Hefeextrakt
<i>Eurogentech</i> (Seraing, Belgien)	Agarose, Elektroporationsküvetten
<i>Fermentas</i> (St. Leon-Rot)	CIAP, Restriktionsendonukleasen und -puffer, Protein- und DNA-Größenstandards, T4-DNA-Ligase, DTT
<i>Finnzymes</i> (Espoo, Finnland)	Phusion™ DNS Polymerase
<i>Fluka</i> (Neu-Ulm)	TCEP, Mercaptoethanol, L-Phe
<i>GE Healthcare</i> (München)	ECL Westernblotting System, Ni-Sepharose 6 fastflow, Gel Filtration Calibration Kit, „HiTrap-Q-HP“ Anionenaustauscher-Säule
<i>IBA</i> (Göttingen)	Streptactin-Sepharose, Desthiobiotin
<i>Invitrogen</i> (Karlsruhe)	Texas-Red C5 Bromacetamid
<i>Jena Bioscience</i> (Jena)	ApCpp
<i>KNF optichem</i> (Lohmar)	Ethanol, Acetonitril
<i>Macherey &amp; Nagel</i> (Düren)	porablot NCP 0,45 µm Nitrocellulosemembran
<i>Merck KGaA</i> (Darmstadt)	KOD Hot Start DNA Polymerase, MESNA, SDS, Tween-20, 5-IAF
<i>Millipore</i> (Molsheim)	Dialysemembran (Porendurchmesser 0,025 µm)
<i>MultiSynTech</i> (Witten)	HBTU
<i>New England Biolabs</i> (Schwalbach)	Desoxyribonucleotide (dATP, dTTP, dGTP, dCTP), Restriktionsendonukleasen, DNA-Größenstandard
<i>Nova Biochem</i> (Darmstadt)	Fmoc-geschützte Aminosäuren, Harze
<i>Novagen</i> (Darmstadt)	KOD-Hot Start Polymerase-PCR Kit
<i>OPERON Biotechnologies</i> (Köln)	Oligonukleotide
<i>Promega</i> (Madison, USA)	Modifiziertes Trypsin (sequencing grade), SoftLink™ Soft Release Avidin Resin
<i>Qiagen</i> (Hilden)	QIAquick® Gel Extraction Kit, QIAquick® Spin Säule, Ni <sup>2+</sup> -NTA-Superflow
<i>Roth</i> (Karlsruhe)	Acrylamid-Lsg. für SDS-PAGE, APS, BSA, DMSO, Ethidiumbromid, β-Mercaptoethanol, Bromphenolblau, TEMED, Visking-Dialyseschlauch (MWCO 15.000), Harnstoff, Agar, Coomassie Brilliant Blue R 250, L-Arabinose, Glukose, L-Pro
<i>Sigma-Aldrich</i> (München)	5(6)-Carboxyfluorescein, TFA, TIS, DIPEA, Biotin, DHB, CHCA

## 2.3 Vektoren

### pET16b

Der Vektor pET-16b (5711 bp, *Novagen*) verfügt über eine *multiple cloning site* (MCS), die eine Zielgensequenz am 5'-Ende mit den Codons einer Faktor Xa Proteaseschnittstelle sowie einer Sequenz eines H<sub>6</sub>-Tags verbindet. Durch Klonierung der unterschiedlichen Zielgensequenzen über die Schnittstellen *NcoI* und *HindIII* wurden diese Eigenschaften jedoch aus dem Vektor entfernt. Der T7 Promotor reguliert die Expression des Zielgens und wird durch den *lac* Repressor (*lacI*) reprimiert, bis IPTG zur Induktion der Expression des Zielgens zugegeben wird. Die kodierte  $\beta$ -Lactamase verleiht Ampicillinresistenz und die Vervielfältigung geschieht über den ColE1 Replikationsursprung aus pBR322.

### pRSFDuet

Der pRSFDuet Vektor (3829 bp) der Firma *Novagen* wurde zur Koexpression von zwei Zielgensequenzen entwickelt. Er enthält zwei MCS, die jeweils am 5'-Ende eine T7/*lac* Promotor/Operator Sequenz, sowie eine Ribosomenbindestelle besitzen. Außerdem enthält der Vektor das Kanamycin-Resistenzgen, den *lac*-Repressor (*lacI*) und den vom RSF1030 abgeleiteten RSF Replikationsursprung. In dieser Arbeit wurde keine duale Koexpression durchgeführt, sondern der Vektor wurde nur zur Expression eines Zielgens verwendet. Dieses wurde in die erste MCS kloniert.

### pET42b

Der pET42b-Vektor (5930 bp) der Firma *Novagen* wurde zur Expression einer Zielgensequenz mit einem N-terminalen GST-Fusionsprotein und einem C-terminalen H<sub>6</sub>-Tag entwickelt. Über die Klonierung mit der *NdeI* Schnittstelle am 5'-Ende wurde jedoch die komplette GST-kodierende Sequenz entfernt. Mittels der T7/*lac* Promotor/Operator Sequenzen und des *lac* Repressors (*lacI*) kann die Expression des Zielgens durch IPTG induziert werden. Der pET42b-Vektor enthält zusätzlich ein Gen zur Vermittlung der Kanamycin-Resistenz und den ColE1 Replikationsursprung aus pBR322.

### **pET28a**

Der pET28a-Vektor (5369 bp) der Firma *Novagen* verfügt über eine N-terminale H<sub>6</sub>-Tag-Sequenz, eine anschließende Thrombin Protease-Schnittstelle und einen T7-Tag, sowie einen C-terminalen H<sub>6</sub>-Tag. In dieser Arbeit wurden durch Verwendung der *NcoI* Schnittstelle alle N-terminalen Eigenschaften des Vektors entfernt und meistens durch eine StrepTag II (ST)-Sequenz ersetzt. Außerdem enthält der Vektor das Kanamycin-Resistenzgen, einen T7-Promotor, den *lac*-Repressor (*lacI*) und den Replikationsursprung ColE1 aus pBR322.

### **pSU38**

Der pSU38-Vektor wurde ausgehend von dem Plasmid pACYC184 [205] von Bartolomé *et al.* konstruiert [206]. Unter anderem enthält dieser Vektor eine Kanamycin-Resistenz und ein *lacZα*-Reporterogen. Der Vektor liegt in mittlerer Kopienzahl (20-30 Kopien) in der Zelle vor und der Promoter ist unter nicht-induzierten Bedingungen vollständig blockiert.

### **pBluescript**

Der pBluescript II KS(-) Vektor (2962 bp) ist ein weit verbreiteter Klonierungsvektor. Der Vektor besitzt sowohl einen *lacI*-Promotor, als auch einen T3-Promotor am 5'-Ende der MCS. Am 3'-Ende der MCS liegt noch die Sequenz für einen T7-Promotor um eine Expression in zwei Richtungen zu ermöglichen. Der Vektor verleiht Resistenz gegen Ampicillin durch die Expression des  $\beta$ -Lactamase-Gens und repliziert sich in *E. coli* durch den ColE1 Replikationsursprung, der aus dem pBR322 Plasmid stammt.

### **pBAD**

Die pBAD Vektoren der Firma *Invitrogen* erlauben die Expression eines Proteins von Interesse durch die Zugabe von L-Arabinose. Dazu wird eine *araBAD* Promotorsequenz am 5'-Ende des Zielgens, sowie die Sequenz des *araC*-Repressors verwendet. Zusätzlich zur Bindung von L-Arabinose an den AraC-Repressor wird für eine maximale Expression der Komplex aus cAMP und dem cAMP Aktivatorprotein (CAP) benötigt. Über diese Katabolitrepression kann damit ebenfalls durch Zugabe von Glukose eine unerwünschte Transkription der Zielgensequenz verringert werden. Der Vektor verleiht Resistenz gegen Ampicillin durch die Expression des  $\beta$ -Lactamase-Gens (*bla*) und repliziert sich in *E. coli* durch den ColE1 Replikationsursprung, der aus dem pBR322 Plasmid stammt.

## 2.4 Mikroorganismen

### 2.4.1 *E. coli* Top10

Elektrokompetente Zellen des Stammes *E. coli* Top10 wurden zu Klonierungszwecken verwendet.

Genotyp: F' {*lacI*<sup>q</sup> *Tn10*(Tet<sup>R</sup>)} *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX74*  
*deoR recA1 araD139* Δ(*ara-leu*)7697 *galU galK rpsL*(StrR)*endA1 nupG*

### 2.4.2 *E. coli* BL21 Gold (DE3)

Dem Expressionsstamm *E. coli* BL21 Gold (DE3) fehlen zur Erhöhung der Proteinstabilität die Gene für die Proteasen *lon* und *ompT*. Er enthält ein Gen für die T7 RNA Polymerase, das im Chromosom nach *lacZ* und dem Promotor *lacV5* auf einem λ-Prophagen insertiert vorliegt. Die Polymerase ist nach Induktion mit IPTG für die effiziente Expression von unter Kontrolle des T7 Promotors stehenden Zielgenen essentiell. Der Stamm enthält eine Resistenz gegen Tetrazyklin. Das Gen der Endonuclease I (*endA*) ist inaktiviert.

Genotyp: F<sup>-</sup> *ompT hsdS<sub>B</sub>* (r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>) *dcm*<sup>+</sup> Tet<sup>r</sup> *gal*λ(DE3) *endA* Hte

## 2.5 Medien

Als Vollmedium zur Anzucht von *E. coli* BL21 Gold (DE3)- und *E. coli* Top10-Kulturen wurde autoklaviertes LB-Medium verwendet. Zur Herstellung von Agarplatten wurde das Medium vor dem Autoklavieren mit 1,5 % (w/v) Agar Nr. 1 versetzt.

**LB-Medium:**                    10 g/L Bactotrypton  
                                         5 g/L Hefeextrakt  
                                         5 g/L NaCl

Für die <sup>15</sup>N-Isotopenmarkierung der Proteine für NMR-Untersuchungen wurde M9 Minimal Medium zur Anzucht der *E. coli* BL21 Gold (DE3)-Zellen verwendet.

<b>M9 Minimal Medium:</b>	45 mM Na <sub>2</sub> HPO <sub>4</sub>
	25 mM KH <sub>2</sub> PO <sub>4</sub>
	8,5 mM NaCl
	0,1 mM CaCl <sub>2</sub>
	1 mM MgSO <sub>4</sub>
	0,03 mg/mL Thiamin (steril filtriert)
	0,1 % NH <sub>4</sub> Cl oder 0,1 % <sup>15</sup> NH <sub>4</sub> Cl (steril filtriert)
	0,2 % Glukose (steril filtriert)
	22 nM FeCl <sub>3</sub> (steril filtriert)

Hitzeempfindliche Substanzen, wie z.B Antibiotika, wurden dem Medium als sterilfiltrierte Lösungen erst nach dem Autoklavieren in folgenden Konzentrationen zugegeben:

Ampicillin: 100 µg/mL

Kanamycin: 50 µg/mL

oder eine Mischung aus Amp<sup>100</sup> und Kan<sup>50</sup>.

## 2.6 Puffer und Lösungen

Ni <sup>2+</sup> -NTA-Puffer A	50 mM Tris/HCl pH 8,0
	300 mM NaCl

Ni <sup>2+</sup> -NTA-Puffer B	50 mM Tris/HCl pH 8,0
	300 mM NaCl
	250 mM Imidazol

Ni <sup>2+</sup> -NTA-Puffer A (Harnstoff)	50 mM Tris/HCl pH 8,0
	300 mM NaCl
	8 M Harnstoff

Ni <sup>2+</sup> -NTA-Puffer B (Harnstoff)	50 mM Tris/HCl pH 8,0
	300 mM NaCl
	250 mM Imidazol
	8 M Harnstoff

## 2 Material

---

Spleißpuffer	50 mM Tris/HCl pH 7,0 300 mM NaCl 1 mM EDTA 2 mM DTT 10 % Glycerin
Strep-Tag Puffer W	100 mM Tris/HCl pH 8,0 150 mM NaCl 1 mM EDTA
Strep-Tag Puffer E	100 mM Tris/HCl pH 8,0 150 mM NaCl 1 mM EDTA 2,5 mM Desthiobiotin
4x SDS Probenpuffer	50 mM Tris/HCl pH 6,8 8 % (w/v) SDS 40 % (v/v) Glycerin 20 % (v/v) $\beta$ -Mercaptoethanol 0,2 % (w/v) Bromphenolblau
4x nativer Probenpuffer	50 mM Tris/HCl pH 6,8 40 % (v/v) Glycerin 20 % (v/v) $\beta$ -Mercaptoethanol 0,2 % (w/v) Bromphenolblau
SDS-Gel Laufpuffer	50 mM Tris 250 mM Glycin 0,1 % (w/v) SDS
Coomassie-Färbelösung	40 % (v/v) Ethanol 10 % (v/v) Essigsäure 2,5 g/L Coomassie Brilliant Blau R250 in ddH <sub>2</sub> O

## 2 Material

---

Coomassie-Entfärbelösung	40 % (v/v) Ethanol 10 % (v/v) Essigsäure in ddH <sub>2</sub> O
NMR-Puffer	25 mM Tris pH 7,0 50 mM NaCl
Waschlösung	200 mM NH <sub>4</sub> HCO <sub>3</sub> 50 % (v/v) Acetonitril
Trypsinlösung	50 mM NH <sub>4</sub> HCO <sub>3</sub> pH 8,1 10 % (v/v) Acetonitril 0,02 µg/µL Trypsin
Diffusionslösung	10 % (v/v) Acetonitril 1 % Trifluoressigsäure
Assay-Puffer	50 mM HEPES pH 8,0 100 mM NaCl 1 mM EDTA 10 mM MgCl <sub>2</sub> 2 mM DTT (optional)

---

## 3 Methoden

### 3.1 Konstruktion der Plasmide

Alle die in dieser Arbeit angewendeten molekularbiologischen Techniken wurden, wenn nicht ausdrücklich anders beschrieben, nach Standardvorschriften durchgeführt [207]. Diese Methoden umfassen: die Plasmidpräparation aus *E. coli* [208], die DNA-Amplifikation mittels Polymerasekettenreaktion (PCR) [209], die Punktmutations-PCR nach dem *QuickChange SiteDirected Mutagenesis*-Protokoll der Firma *Stratagene* [210], die DNA-Fragment- und Plasmidreinigung mit Hilfe der *QIAquick-spin-Säulen* der Firma *Qiagen*, die Dephosphorylierung mittels alkalischer Phosphatase, die Ligation von DNA-Fragmenten durch die T4-DNA-Ligase, sowie die abschließende Sequenzierung des hergestellten Plasmidkonstrukts bei der Firma *GATC Biotech AG* (Konstanz). Eine detaillierte Beschreibung dieser Methoden, sowie der Transformation von kompetenten *E. coli*-Zellen mit Plasmid-DNA befindet sich außerdem in folgender Diplomarbeit [211].

#### *Klonierung der Konstrukte mit den NRPS-Fragmenten aus der Gramicidin S Biosynthese*

Die beiden Plasmide zur Expression des GrsA A-PCP-H<sub>6</sub> Fragments (pJZ06) und seiner cystein-freie Variante GrsA A-PCP(Δ-4Cys)-H<sub>6</sub> (pJZ13) basierten auf dem pET16b-Vektor und wurden bereits in früheren Arbeiten hergestellt [211]. Ausgehend von pJZ13 wurden nun mittels Punktmutagenese Cysteine wieder in das Konstrukt eingeführt. Durch Verwendung der Oligonukleotide oJZ20 (5'-GTAGCCACCA AAGAATGCAT AGGTCATTCA GTTC) und oJZ21 (5'-GGAACTGAAT GACCTATGCA TTCTTTGGTG GCTAC) konnte z.B. die Mutation T341C eingebracht werden. In den Sequenzen der Oligonukleotide sind die zu mutierenden Basenpaare unterstrichen. In einer zweiten Punktmutagenese mit oJZ18 (5'-CCTTCTTATT TTATCCAGCT GTGCAAAATG CCGCTTAC) und oJZ19 (5'-GTAAGCGGCA TTTTGACACAG CTGGATAAAA TAAGAAGG) wurde pJZ67 zur Expression von A-PCP(Δ-4Cys, T341C, D508C) - Protein **11** erhalten. Zur Expression des kompletten Elongationsmoduls GrsB1 (ST-GrsB1 C-A-PCP-H<sub>6</sub>) wurde das auf dem pET28a basierende Plasmid pJZ56b verwendet. Mit den Oligonukleotiden oJZ56 (5'-ATAGCTAGCA GTACATTTAA AAAAGAACAT G) und oJZ57 (5'-TATGGATCCG CTTTCTTTCT TTGCCCCGTT T) wurde das 3,1 kbp große Insert von der genomischen DNA des Gramicidin S Produzenten *Bacillus brevis* ATC9999 amplifiziert und mit den

Restriktionsenzymen *NheI* und *BamHI* geschnitten. Die Ligation erfolgte mit einem identisch geschnittenen, modifizierten pET28a-Vektorfragment, das in 5'-Richtung zur *NheI*-Schnittstelle die Sequenz eines StrepTag II und in 3'-Richtung zur *BamHI*-Schnittstelle die Sequenz eines H<sub>6</sub>-Tags besaß.

#### *Klonierung der Expressionsplasmide des natürlich gespaltenen Npu DnaE Inteins*

Die Expressionsplasmide mit den N- und C-terminalen Hälften des *Npu DnaE* Inteins und Sequenzen von Hitzeschockproteinen als Exteine (pSTDuet51F, pSTBad1a) wurden freundlicherweise von Dr. Sylvain Tourel und Prof. Dr. Horst Kessler (Technische Universität München) zur Verfügung gestellt. Das N-Extein konnte aus dem pSTDuet51F-Vektor über die *NcoI* und *BamHI* Schnittstellen entfernt und durch ein ST-gpD kodierendes Fragment ersetzt werden. Dieses wurde durch PCR von pSB34 [164] mit den Oligonukleotiden oJZ35 (5'-ATACCATGGC CAGTTGGAGC CACCCGCAGT TCGAAAAAGC GAGCAAAGAA ACCTTTACCC) und (5'-ATAGGATCCA ACGATGCTGA TTGCC) amplifiziert und die Ligation ergab das Expressionsplasmid pAU08 für ST-gpD-*Npu DnaE*<sup>N</sup> (1) [212]. Ausgehend von pAU08 wurde das Plasmid für die C1A Mutation durch Punktmutagenese mit den Oligonukleotiden oJZ81 (5'-ATCAGCATCG TTGGATCTGC TTTAAGCTAT GAAAC) und oJZ82 (5'-GTTTCATAGC TTAAAGCAGA TCCAACGATG CTGAT) kloniert (pJZ62 - 1<sub>C1A</sub>). Das Expressionsplasmid für das Fusionskonstrukt ST-eGFP-*Npu DnaE*<sup>N</sup> (pVS07) wurde von Vivien Schütz zur Verfügung gestellt [213]. Für die C-terminalen Inteininfusionskonstrukte wurde das C-Extein aus pSTBad1a über die Schnittstellen *KpnI* und *HindIII* entfernt und in einem ersten Schritt durch ein verbessertes grün fluoreszierendes Protein (eGFP)-H<sub>6</sub> Fragment ersetzt (s. [212]). Der resultierende Vektor pAU07 konnte mit *KpnI* und *BglII* geschnitten werden. Die Ligation mit einem identisch geschnittenen durch Amplifikation von dem Templat pSB34 [164] mit den Oligonukleotiden (5'-ATAGGTACCG CGAGCAAAGA AACCTTTAC) und (5'-TATAGATCTG GATCCAACGA TGCTGATTGC CG) erhaltenem PCR-Produkt ergab pJZ60 (*Npu DnaE*<sup>C</sup>-gpD-H<sub>6</sub>). Aufgrund schlechter Expression wurde das komplette Insert über die *NdeI* und *HindIII* Schnittstellen in ein pET42b-Vektorgrundgerüst kloniert. Damit war pJZ68 das Expressionsplasmid für *Npu DnaE*<sup>C</sup>-gpD-H<sub>6</sub> (2). Die N36A Mutante wurde ausgehend von pJZ68 mit den Oligonukleotiden oJZ83 (5'-GGCTTCATAG CTTCTGCATG CTTCAATGGT ACCGCG) und oJZ84 (5'-CGCGGTACCA TTGAAGCATG CAGAAGCTAT GAAGCC) mittels Punktmutagenese-PCR erstellt (pJZ69 - 2<sub>N36A</sub>). Analog wurde zusätzlich in pJZ69 die C+1A Mutation mit oJZ95 (5'-CTTCATAGCT TCTGCAGCCT TCAATGGTAC CGCG) und

oJZ96 (5'-CGCGGTACCA TTGAAGGCTG CAGAAGCTAT GAAG) eingebracht (pJZ82 - 2<sub>N36A, C+1A</sub>). Um eine native Reinigung der C-terminalen Intein-fusionsproteine zu ermöglichen wurden Plasmide mit einer N-terminalen StrepTag II Sequenz benötigt. Plasmid pJZ57b kodiert für das Protein ST-*Npu* DnaE<sup>C</sup>-gpD-H<sub>6</sub> (**3**). Das Insert wurde mittels PCR von pJZ60 als Templat und mit den Oligonukleotiden (5'-ATACCATGGC CAGTTGGAGC CACCCGCAGT TCGAAAAAGC TAGCATGATC AAAATAGCCA CAC) und (5'-TATAAGCTTA GTGATGGTGA TGG) amplifiziert, anschließend mit *Nco*I und *Hind*III geschnitten und in einen identisch geschnittenen pET16b-Vektor ligiert. Aus pJZ57b konnte über *Kpn*I und *Hind*III das C-Extein entfernt werden und durch ein Thioredoxin (Trx)-H<sub>6</sub> Fragment (erhalten aus einem identischen Restriktionsverdau von pVS01 [214]) ersetzt werden (pJZ70 - ST-*Npu* DnaE<sup>C</sup>-Trx-H<sub>6</sub> - **4**). Die detaillierte Herstellung der Plasmide zur Expression der in Tabelle 5 erwähnten zusätzlichen Fusionsproteine ist in folgender Veröffentlichung nachzulesen [214].

#### *Herstellung der Konstrukte für NMR-Untersuchungen an gespaltenen Inteinen*

Neben den im vorangegangenen Kapitel beschriebenen Intein-gpD Konstrukten wurden weitere Proteine per NMR-Spektroskopie untersucht. Als Referenz für die Signale des gpD-Proteins diente Protein **5** (ST-gpD-H<sub>6</sub>). Das Insert wurde über eine PCR mit pSB34 [164] und den Oligonukleotiden (5'-ATACCATGGC CAGTTGGAGC CACCCGCAGT TCGAAAAAGC GAGCAAAGAA ACCTTTA) und (5'-CGCAAGCTTA GTGATGGTGA TGGTGATGAG ATCTGGATCC AACGATGCTG ATTGCCGTTTC) amplifiziert und mittels *Nco*I und *Hind*III geschnitten. Die Ligation mit einem identisch verdauten pET28a-Vektor ergab pJZ49. Die C-terminale Hälfte des an der Endonukleaseposition künstlich gespaltenen *Ssp* DnaB Inteins wurde mit Hilfe von Protein **6** ST-gpD-DnaB<sup>C</sup>(107-154)-H<sub>6</sub> NMR-spektroskopisch untersucht. Der zugehörige Expressionsvektor pJZ35 basierte ebenfalls auf einem pET28a Grundgerüst, welches über einen *Nco*I und *Hind*III Verdau erzeugt wurde. Für das Insert diente erneut pSB34 als Templat und die Oligonukleotide waren oJZ35 (5'-ATACCATGGC CAGTTGGAGC CACCCGCAGT TCGAAAAAGC GAGCAAAGAA ACCTTTACCC) und oJZ36 (5'-ATAAAGCTTA CCGCAGGATC TAATACTGTT ATGG). Für die Klonierung des gpD-Fusionsproteins mit der C-terminalen Hälfte das an Position 11 gespaltenen *Ssp* DnaB Intein, DnaB<sup>C</sup>(12-154)-gpD-H<sub>6</sub> (**7**), wurden die Plasmide pCL30 (bereitgestellt von Dr. Christina Ludwig) und pTS49 (bereitgestellt von Tim Sonntag) benötigt. Für das Vektorfragment wurde pCL30 (DnaB<sup>C</sup>(12-154)-A-PCP-H<sub>6</sub> in pSU38) mit *Bam*HI geschnitten und dephosphoryliert. Die Ligation erfolgte mit dem über PCR

hergestellten und ebenfalls mit *Bam*HI verdauten Insert (pTS49 als Templat und den Oligonukleotiden oJZ26 (5'-ATATCTAGAA AGGAGATATA CCATGGGATC CGCGAGCAAA GAAACCTTTA CC) und oJZ34 (5'-ATAACTAGTG GATCCAACGA TGCTGATTGC C)) und ergab pJZ29. Zusätzlich wurde für eine vereinfachte Reinigung mit besseren Ausbeuten im Vergleich mit Protein 1 (ST-gpD-*Npu* DnaE<sup>N</sup>) die StrepTag II Sequenz durch einen H<sub>6</sub>-Tag ersetzt. pAU08 diente sowohl als Vektorgrundgerüst (Restriktionsverdau mit *Nco*I und *Bgl*I), als auch als Templat in einer PCR mit den Oligonukleotiden oJZ97 (5'-ATACCATGGG CCATCACCAT CACCATCACA CTAGTAAAGA AACCTTTACC CATTAC) und oJZ34 (5'-ATAACTAGTG GATCCAACGA TGCTGATTGC C). Das Insert wurde ebenfalls mit *Nco*I und *Bgl*I geschnitten und die Ligation mit dem Vektorfragment ergab pJZ83 (H<sub>6</sub>-gpD-*Npu* DnaE<sup>N</sup> - 8). Zur Klonierung des Expressionsvektors für das Protein H<sub>6</sub>-gpD-*Npu* DnaE<sup>N</sup><sub>C1A</sub> (8<sub>C1A</sub>) wurden sowohl pJZ62 (ST-gpD-*Npu* DnaE<sup>N</sup><sub>C1A</sub>), als auch pJZ83 mit *Nco*I und *Bgl*I geschnitten. Der Austausch der ST-gpD Sequenz in pJZ62 durch die H<sub>6</sub>-gpD Sequenz ergab pJZ84.

#### *Klonierung der Intein-kassetten für das SPLICEFINDER-System*

Alle Inteinintegrationskassetten basierten auf dem Plasmid pTS13, welches über eine Fusions-PCR (gespaltenes *Sce* VMA Intein von dem HIS3-Gen unterbrochen) in einem pBluescript-Vektor generiert wurde. pTS13 wurde von Tim Sonntag zur Verfügung gestellt. In den folgenden Schritten wurden die 5'- und 3'-Bereiche des *Sce* VMA Inteins mittels RF-PCR [215] durch die Intein-fusionsgene der unterschiedlichen gespaltenen Inteine ausgetauscht. Für das künstlich an der Position der Endonuklease gesplante *Ssp* DnaB Intein wurde die C-terminale Sequenz von dem Templat pAU04 (*Ssp* DnaB<sup>C</sup>(107-154)-GrsA A<sup>Sub B</sup>-PCP-H<sub>6</sub> [212]) mit den Oligonukleotiden oTS158 (5'-TACTAACGCCG CCATCCAGTT TAAACTCAGC CATACTTTTC ATACTCC) und oTS159 (5'-CCCTCACTAA AGGGAACAAA AGCTGGCGGA TACATATTTG AATGTAT) amplifiziert. Das PCR-Produkt enthielt den araBAD Promotorbereich aus dem kommerziell erhältlichen pBAD Expressionssystem (*Invitrogen*) und wurde als Insert in einer RF-PCR mit pTS13 als Templat verwendet. Das Produkt wurde pCasDnaB1 genannt und diente als Templat in einer zweiten RF-PCR zum Austausch der 5'-kodierenden Sequenzen. Hier wurde das Insert über eine PCR von pJZ28 (H<sub>6</sub>-GrsA A<sup>Sub A</sup>-DnaB<sup>N</sup>(1-104)) und den Oligonukleotiden oTS156 (5'-AATACGACTC ACTATAGGGC GAATTGCGAG ATCTCGATCC CGCGAAAT) und oTS157 (5'-ATCTGGCGCG CCTTAATTAA CCCGGGGCGC TGGCAAGTGT AGCGGTC) generiert. Die vollständige *Ssp* DnaB Intein-kassette war pCasDnaB2. In zwei

weiteren RF-PCR Schritten wurde die *Npu* DnaE Intein-Kassette hergestellt. Zuerst erfolgte der Austausch der Gensequenz für den N-terminalen Bereich mit dem Fragment ST-gpD-*Npu* DnaE<sup>N</sup>, welches über eine PCR von pAU08 mit den Oligonukleotiden oJZ62 (5'-AATTGTGAGC GGATAACAAT TCCCCTGTAG AAATAATTTT G) und oJZ63 (5'-TGGTGGTGGT GCTCGAGTGC GGCCGCAAGC TTAATT) amplifiziert wurde. Anschließend wurde der 3'-Bereich durch ein *Npu* DnaE<sup>C</sup>-eGPF-H<sub>6</sub>-Fragment ersetzt, welches über eine PCR von pAU07 (*Npu* DnaE<sup>C</sup>-eGPF-H<sub>6</sub>) mit den Oligonukleotiden oJZ64 (5'-CGCTTTTTAT CGCAACTCTC TACTG) und oJZ65 (5'-AAGCTTAGTG ATGGTGATGG TGATG) generiert wurde. Das endgültige Plasmid wurde pCasDnaE2 genannt. Analog zu den beiden anderen Intein-Kassetten wurde auch die *Mxe* GyrA Kassette über zwei aufeinander folgende RF-PCRs ausgehend von pCasDnaE2 hergestellt. Der 5'-kodierende Bereich wurde von pAI13 (MBP-*Mxe* GyrA<sup>N</sup>(1-119)-FKBP, zur Verfügung gestellt von Dr. Thomas Kurpiers) mit den Oligonukleotiden oJZ91 (5'-ACTTTAATAA GGAGATATAC CATGGAAATC GAAGAAGG) und oJZ92 (5'-GTGCTCGAGT GCGGCCGCAA GCTTAATAAC TAGTTTCCAG TTTTAG) amplifiziert. Das Zwischenkonstrukt pCasGyrA1 diente als Templat in der zweiten RF-PCR. Hier wurde der 3'-kodierende Bereich durch das Insert gpD-*Mxe* GyrA<sup>C</sup>(120-199)-H<sub>6</sub>, hergestellt aus einer PCR von pDK01 (zur Verfügung gestellt von Dania Kendziora) und den Oligonukleotiden oJZ93 (5'-ACTTTAAGAA GGAGATATAC ATATGGCGAG CAAAGAAACC TTTAC) und oJZ94 (5'-AAGCTTAGTG ATGGTGATGG TGATG), ausgetauscht.

Das Helferplasmid pRSFara mit den kodierenden Bereichen für die regulatorischen Proteine LacI und AraC wurde wie folgt hergestellt: Ausgehend von dem kommerziell erhältlichen RSFDuet Vektor (*Novagen*, Kanamycin-Resistenz, *lacI*) wurden beide multiple Klonierungsstellen durch die Sequenz des *araC*-Gens und den korrespondierenden Promotorregionen mittels RF-PCR ausgetauscht. Das Insert für diesen Klonierungsschritt wurde über eine PCR von einem kommerziell erhältlichen pBAD-Vektor mit den Oligonukleotiden oJZ69 (5'-CGCTCTCCCT TATGCGACTC CTGCAGGAGA AACAGTAGAG AGTTGCG) und oJZ70 (5'-TTCGCAGCAG CGGTTTCTTT ACCAGGCCTG CCACCATACC CACGCCG) erzeugt.

#### *Herstellung des Beispielgens ST-gpD-Trx-H<sub>6</sub> und der Integrationsplasmide*

Zur Integration der *Ssp* DnaB- und der *Npu* DnaE-Intein-Kassetten in das Beispielgen ST-gpD-Trx-H<sub>6</sub> wurde als Templat das Plasmid pSE01 verwendet. Dieses konnte über eine Drei-Fragmentligation hergestellt werden. Ein *Nco*I, *Eco*RI Verdau des Plasmids pJZ42 (ST-gpD-

*Ssp* DnaB<sup>N</sup>(1-104)) führte zu dem ST-gpD Fragment. Zusätzlich wurde die Sequenz Trx-H<sub>6</sub> über eine PCR von pCL20 [169] mit den Oligonukleotiden oJZ85 (5'-ATAGAATTCG GCGGAGGCGG AGGAGGATCT GATAAAAT) und oJZ55 (5'-TATCTCGAGT TAGTGATGGT GATGGTGATG AG) und einem anschließenden Verdau mit *Eco*RI und *Xho*I erhalten. oJZ85 baut zusätzlich sechs Glycin-Codons in den Linkerbereich zwischen beiden Domänen ein. Die Ligation beider Fragmente mit einem *Nco*I, *Xho*I geschnittenen pET16b-Vektorfragment ergab pSE01. Um größere Expressionsausbeuten für die NMR-Untersuchungen zu erhalten, wurde das gesamte Insertfragment mittels eines *Nco*I, *Xho*I Restriktionsverdau in einen pET28a Vektor transferiert (pSE02, Protein 9).

Die über RF-PCR erhaltenen Integrationsplasmide der *Ssp* DnaB- und der *Npu* DnaE-Inteinkassetten in das Beispielgen ST-gpD-Trx-H<sub>6</sub> sind in Tabelle 3, die dazu verwendeten Oligonukleotide sind in Tabelle 10 im Anhang zusammengefasst.

**Tabelle 3** Generierte Integrationsplasmide der *Ssp* DnaB- und *Npu* DnaE-Inteinkassetten in das Beispielgen ST-gpD-Trx-H<sub>6</sub>.

Kombination	Oligonukleotide	AS an der N-term. Spleißstelle	AS an der C-term. Spleißstelle	Plasmidname
K1, <i>Ssp</i> DnaB	oSE03, oSE01	GGG	SGGG	pTS216
K2, <i>Ssp</i> DnaB	oSE03, oSE02	GGG	SIEQ	pTS217
K3, <i>Ssp</i> DnaB	oSE04, oSE01	ESG	SGGG	pTS218
K4, <i>Ssp</i> DnaB	oSE04, oSE02	ESG	SIEQ	pTS219
K5, <i>Npu</i> DnaE	oSE05, oSE07	GGG	CGGG	pTS210
K6, <i>Npu</i> DnaE	oSE05, oSE08	GGG	CFNK	pTS211
K7, <i>Npu</i> DnaE	oSE06, oSE07	AEY	CGGG	pTS212
K8, <i>Npu</i> DnaE	oSE06, oSE08	AEY	CFNK	pTS213

#### *Plasmide für die separate Expression der Npu DnaE-Inteinfusionsproteine*

Zur Durchführung von *in vitro* Spleißreaktionen mit gereinigten Proteinen wurden die individuellen *Npu* DnaE-Inteinfusionskonstrukte aus den Integrationsplasmiden mittels Restriktionsverdau entfernt und in Expressionsvektoren kloniert. Das Vektorfragment der 5'-Inteinfusionsgene wurde durch einen *Nco*I, *Hind*III-Verdau eines RSFDuet Vektors (*Novagen*) erhalten und mit den *Nco*I, *Hind*III-Fragmenten entweder von pSE07 (ST-gpD-GGG-*Npu* DnaE<sup>N</sup>) oder von pSE10 (ST-gpD-AEY-*Npu* DnaE<sup>N</sup>) ligiert. Die entstandenen Plasmide wurden pSE13, respektive pSE14 genannt. Die 3'-Fusionsgene wurden in

kommerziell erhältliche pBAD-Vektoren transferiert. Das Vektorgrundgerüst wurde durch einen Restriktionsverdau mit *NdeI* und *BglII* erhalten. Das *Npu* DnaE<sup>C</sup>-CGGG-Trx-H<sub>6</sub> Insert kam von einem *NdeI* und *BglII*-Verdau von pSE07 und die Ligation ergab pSE11. Das andere Plasmid, welches für *Npu* DnaE<sup>C</sup>-CFNK-Trx-H<sub>6</sub> kodiert, wurde pSE12 genannt und resultierte aus einem analogen Verdau von pSE10 mit anschließender Ligation in das pBAD-Vektorfragment.

#### *Klonierung von NRPS-Inteinfusionskonstrukten*

Die Klonierung des Expressionsplasmids pJZ16 für das Protein **10**, ST-*Ssp* DnaB<sup>C</sup>(12-154)-A-PCP-H<sub>6</sub>, gelang bereits in früheren Arbeiten [211]. Die Herstellung der Vektoren zur Expression der Proteine **12**, pAU06 - ST-GrsA A<sub>1-421</sub>-*Ssp* DnaB<sup>N</sup>(1-104), und **13**, pAU04 - *Ssp* DnaB<sup>C</sup>(107-154)-S-GrsA A<sub>423-540</sub>-PCP-H<sub>6</sub> ist in der Bachelorarbeit von Annika Urbanek nachzulesen [212].

#### *Konstrukte zur Expression von Ssp DnaB-Inteinfusionskonstrukten für Gelfiltrationsuntersuchungen*

Die Plasmide zur Expression von MBP-*Ssp* DnaB<sup>N</sup>-FKBP-H<sub>6</sub> und MBP-*Ssp* DnaB<sup>C</sup>-H<sub>6</sub> heißen pSB13, respektive pTK56 und wurden von Dr. Steffen Brenzel und Dr. Thomas Kurpiers zur Verfügung gestellt [163,164].

## **3.2 Expression und Proteinreinigung**

### **3.2.1 Heterologe Expression in *E. coli* Zellen**

Zur Herstellung der gewünschten Proteine wurde der *E. coli* Expressionsstamm BL21 Gold (DE3) mit den in dieser Arbeit hergestellten Plasmiden transformiert. Die Expression des gewünschten Proteins wurde über eine Testexpression im 4 mL Maßstab überprüft, dazu wurde mit Hilfe der SDS-Polyacrylamidgelelektrophorese (SDS-PAGE) der gesamte Proteininhalt der Zellen vor und nach der Induktion analysiert.

Zur präparativen Herstellung größerer Mengen an Protein wurden 300 - 800 mL LB-Medium im Verhältnis 1:100 mit einer Übernachtskultur angeimpft und bei 37°C und 250 rpm herangezogen. Durch Zugabe entsprechender Antibiotika (50 µg/mL Kanamycin, 100 µg/mL Ampicillin oder einer Kombination aus beiden) wurde ein Selektionsdruck aufgebaut. Bei

einer  $OD_{600}$  von 0,5 bis 0,7 wurde die Überexpression durch Zugabe von IPTG in einer Endkonzentration von 0,4 mM oder durch Zugabe von L-Arabinose (0,02 - 0,2 %) induziert. Nach der Induktion wurden die Zellen für weitere 3 bis 4 h bei 28°C inkubiert. Für den Einbau der NMR-aktiven  $^{15}\text{N}$ -Isotope wurde bei einer  $OD_{600}$  von 0,5 bis 0,7 ein Mediumswechsel zu M9 Minimal Medium mit  $^{15}\text{N}$ -Ammoniumchlorid als einziger Stickstoffquelle durchgeführt. Dazu wurden die Zellen pelletiert (4.000 rpm, 10 min, RT) und das LB-Medium entfernt. Anschließend wurden die Zellen in Minimalmedium resuspendiert. Vor der Induktion wurden die Zellen für 15 min bei der Induktionstemperatur (meistens 25°C) zur Anpassung an die neuen Mediumsbedingungen geschüttelt. Erfolgte die Induktion mit L-Arabinose, so wurde auf komplett auf Glukose im Minimalmedium verzichtet. Die Expression in isotope-markiertem Medium wurde meistens über Nacht durchgeführt. Nach der Expression des gewünschten Proteins wurden die Zellen geerntet (7.000 rpm, 15 min, 4°C) und in einem möglichst kleinen Volumen eines für die Reinigung geeigneten Puffers resuspendiert. Bis zum endgültigen Zellaufschluss wurden die Zellen bei -20°C gelagert.

#### 3.2.2 Zellaufschluss

Der Zellaufschluss dient zur Freisetzung der exprimierten Proteine aus den *E. coli* Zellen. Dazu wurde ein mechanischer Aufschluss mittels einer French Press Zelle gewählt. In dieser Arbeit wurde ein EmulsiFlex<sup>®</sup>-C5 High pressure homogenizer der Firma *Avestin* verwendet. Das Gerät wurde vor der Benutzung mit Eis gekühlt, um die aufgeschlossenen Proteine zu schonen. Die resuspendierten Zellen der Expression wurden mit hohem Druck durch eine kleine Öffnung gepresst, wobei ein Druckunterschied von 6,9 MPa erreicht wurde. Die dabei resultierenden Scherkräfte zerstörten die Zellwände und setzten die im Cytosol vorliegenden Proteine frei. Um einen möglichst kompletten Aufschluss zu erreichen, wurde die Zellsuspension zweimal durch das Gerät geführt. Zur Trennung des Zelllysats von den Zelltrümmern und den unlöslichen Proteinen wurde die Suspension 30 min bei 17.000 rpm und 4°C zentrifugiert. Der klare Überstand wurde direkt für die jeweilige Affinitätschromatographie eingesetzt.

### 3.2.3 Proteinreinigung

Die Reinigung der Proteine erfolgte in dieser Arbeit mittels N- und/oder C-terminal fusionierter Affinitätstags. Verwendet wurden hierfür sowohl die Hexahistidinsequenz (H<sub>6</sub>) als auch der Strep-Tag II (ST). Die exakten Schritte richteten sich nach den Angaben der Hersteller des Säulenmaterials. Für die Affinitätschromatographie an immobilisierten Nickelionen zur Reinigung von Proteinen mit einem H<sub>6</sub>-Tag wurde das Material der Firma *Qiagen* verwendet (Ni<sup>2+</sup>-NTA Superflow). Proteine mit Strep-Tag II konnten mit dem Strep-Tactin Sepharose Material der Firma *IBA* von den anderen Komponenten des Zellextraktes gereinigt werden. Detailliert sind die Reinigungsbedingungen ebenfalls in früheren Arbeiten zusammengefasst und dort nachzulesen [211]. Wurde ein Protein über beide Affinitätstags gereinigt, so erfolgte nach der Elution von der ersten Chromatographiesäule ein Dialyseschritt in den Auftragspuffer des zweiten Chromatographieschrittes (Ni<sup>2+</sup>-NTA-Puffer A oder Strep-Tag Puffer W). Die Reinigung über eine mittels PTS eingebrachte Biotinfunktionalität ist in Kapitel 3.9.1 beschrieben.

## 3.3 Biochemische Untersuchungen von NRPS-Proteinen

### 3.3.1 Partielle tryptische Verdauung von NRPS-Fragmenten

Um den Einfluss der nicht-hydrolysierbaren Sulfamoylinhibitoren auf die Konformation der GrsA A-PCP Konstrukte zu testen wurden partielle Proteolyse-Experimente durchgeführt. Die apo- und holo-Proteine wurden in Konzentrationen von 6 bis 12 µM in Assay-Puffer mit den jeweiligen Substraten (ATP, L-Phe, ApCpp, AMP - 1mM) und Inhibitoren (100 µM) für 10 Minuten bei 37°C vorinkubiert. Der Verdauung wurde durch Zugabe einer Trypsinlösung (0,08 µg/µL, modified Trypsin, *Promega*) in einem Protease/Protein Verhältnis von 1:250 (w/w) gestartet. Zu unterschiedlichen Zeitpunkten wurden Aliquots entnommen und die Reaktion durch Zugabe von 4x SDS Probenpuffer gestoppt. Um einen nachweisbaren Verdauung in Gegenwart des kognaten Inhibitors (**Inh1**) zu erzielen, musste die Proteasemenge auf 1:25 (w/w) erhöht werden. In einem Kontrollexperiment mit Sfp in An- und Abwesenheit von **Inh1** konnte nachgewiesen werden, dass Trypsin selbst nicht von dem Inhibitor beeinflusst wurde.

### 3.3.2 Native PAGE- und Gelfiltrations-Assays

Zur Analyse der Größe und Form der unterschiedlich behandelten GrsA A-PCP Proteine wurden sowohl native PAGE-, als auch Gelfiltrationsassays verwendet. Die diskontinuierliche native Gelelektrophorese wurde analog zu dem Standard SDS-PAGE Protokoll nur ohne SDS durchgeführt [216]. Es wurden ein 5 % Sammelgel und ein 8 % Trenngel benutzt. Vor Zugabe des 4x nativen Probenpuffers wurden die Proteine mit und ohne Substrate und Inhibitoren für 15 min bei 37°C inkubiert. Die Proben wurden nicht aufgekocht.

Die Gelfiltrationsexperimente wurden mit einer Superdex 200 10/300 GL Säule der Firma *GE Healthcare* an einem Äkta Purifier 10 der gleichen Firma durchgeführt. Die Säule wurde vor den Läufen mit Assay-Puffer und 2 mM DTT äquilibriert. Nach einem Vorinkubationsschritt mit oder ohne Inhibitoren für 10 min bei 37°C wurden 200 µL einer 20 µM Proteinlösung auf die Säule aufgegeben. Bei einer Flussrate von 0,5 mL/min wurde die Absorption bei 280 nm aufgezeichnet. Die Säule wurde mit Hilfe des Gel Filtration Calibration Kit - Low Molecular Weight der Firma *GE Healthcare* kalibriert.

### 3.3.3 Chemisches Markieren des Ppant-Arms

Die Zugänglichkeit des Ppant-Arms zum Lösungsmittel im cystein-freien holo-A-PCP( $\Delta$ -4Cys) wurde durch eine chemische Markierungsreaktion mit Texas-Red C<sub>5</sub> Bromacetamid getestet. Dazu wurde eine 7,5 µM Proteinlösung in Assay-Puffer (pH 7,0 mit 2 mM TCEP) mit entweder 1 mM der Substrate (ATP, ApCpp, AMP und/oder L-Phe) oder 100 µM der Inhibitoren versetzt. Die Mischung wurde für 10 min bei 37°C und anschließend für 10 min bei 25°C inkubiert. Die Reaktion wurde durch die Zugabe von 8 eq. (im Vergleich zum Protein) von Texas-Red C<sub>5</sub> Bromacetamid (*Invitrogen*, 1 mM Stocklösung in DMSO) gestartet und bei 25°C durchgeführt. Zu den Zeitpunkten 2½, 5, 10, 20, 40 und 60 min wurden Aliquots aus der Reaktionsmischung entnommen und die Markierung durch die Zugabe von 4x SDS Probenpuffer gestoppt. Die erhaltenen Proben wurden auf einem SDS-PAGE Gel analysiert und der Einbau des Texas-Red Fluorophors konnte unter UV-Beleuchtung des Gels nachgewiesen werden. Eine densitometrische Auswertung der Markierung gelang mit dem Programm Scion Image (<http://www.scioncorp.com>). Die Werte wurden anhand der Intensitäten der Coomassie-Färbung korrigiert und auf den Wert des holo-Proteins ohne Inhibitor nach einer Stunde normiert. Alle Messreihen waren das Ergebnis von mindestens drei unabhängigen Experimenten.

### 3.4 Kinetische Analyse der Protein *trans*-Spleißreaktion des Npu DnaE Inteins

#### 3.4.1 Durchführung der PTS-Reaktion

Die gereinigten *Npu* DnaE Inteininfusionsproteine wurden in äquimolaren Verhältnissen mit Konzentrationen von 10 - 20  $\mu\text{M}$  in Spleißpuffer gemischt und bei Temperaturen in einem Bereich von 6 bis  $37^\circ\text{C}$  inkubiert. Jedem Reaktionsansatz wurden zusätzlich vor der Reaktion 2 mM DTT aus einer frisch angelegten Stammlösung zugesetzt. Zu unterschiedlichen Zeitpunkten wurden Aliquots entnommen und durch Zugabe von 4x SDS-Probenpuffer wurde die PTS-Reaktion gestoppt. Die Nullstundenproben wurden erhalten, in dem zuerst ein Konstrukt mit 4x SDS-Probenpuffer versetzt wurde und anschließend das Partnerprotein zugegeben wurde. Die kinetische Analyse erfolgte anhand der Bandenintensitäten in einem Coomassie gefärbten SDS-PAGE Gel. Diese wurden mit Hilfe des Programms Scion Image (<http://www.scioncorp.com>) ermittelt und auf die Proteingröße normiert. Alle Messreihen waren das Ergebnis von mindestens drei unabhängigen Experimenten.

#### 3.4.2 Auswertung der kinetischen Daten

Protein *trans*-Spleißreaktionen lassen sich formal in zwei unterschiedliche Reaktionsschritte unterteilen. Nach der reversiblen Assoziation der Inteininfusionsproteine,  $\text{P}^{\text{N}}$  und  $\text{P}^{\text{C}}$ , zum spleißaktiven Komplex K, reagiert dieser irreversibel zu dem Spleißprodukt SP und den herausgeschnittenen Inteinhälften  $\text{I}^{\text{N}}$  und  $\text{I}^{\text{C}}$  weiter. Der zweite Schritt umfasst hier alle Reaktionen des Spleißmechanismus, sowie mögliche konformationelle Änderungen der Inteinstruktur. Das Reaktionsschema mit den zugehörigen Geschwindigkeitskonstanten ist in Abbildung 25 gezeigt.

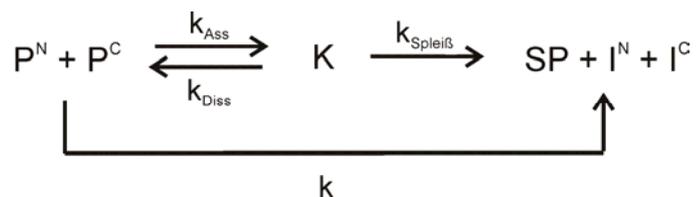


Abbildung 25 Reaktionsschema der PTS-Reaktion mit den zugehörigen Geschwindigkeitskonstanten.

In der Arbeit von Shi und Muir [174] wurde festgestellt, dass für das natürlich gespaltene *Ssp* DnaE Intein die Inteinfragmente  $\text{P}^{\text{N}}$  und  $\text{P}^{\text{C}}$  nahe dem Diffusionslimit assoziieren und dass

die Dissoziation des Komplexes K nur sehr langsam stattfindet ( $k_{\text{Ass}} > k_{\text{Diss}}$ ). Daraus resultiert eine sehr geringe Dissoziationskonstante von 36 nM für das *Ssp* DnaE Intein. In der Gesamtreaktion ist damit die unimolekulare Spleißreaktion definitiv der geschwindigkeitsbestimmende Schritt ( $k_{\text{Ass}} \gg k_{\text{Spleiß}}$ ) [174,175]. Dieser Sachverhalt wurde für eine vereinfachte kinetische Auswertung auch für das mit dem *Ssp* verwandte *Npu* DnaE Intein angenommen. Damit ergibt sich eine Betrachtung, in der das Inteininfusionsprotein im Unterschuss, z.B.  $P^N$ , aufgrund der großen Geschwindigkeitskonstante  $k_{\text{Ass}}$  zu Beginn der Reaktion vollständig im Komplex K vorliegt. Die Spleißreaktion lässt sich nun mit einem Geschwindigkeitsgesetz erster Ordnung bezüglich des Komplexes K und damit auch des Inteininfusionsproteins im Unterschuss beschreiben (s. Gleichung 1).

$$v = -\frac{d[K]}{dt} = k_{\text{Spleiß}} * [K] = \frac{d[\text{SP}]}{dt} \quad \text{oder} \quad v = -\frac{d[P^N]}{dt} = k_{\text{Spleiß}} * [P^N] = \frac{d[\text{SP}]}{dt} \quad (\text{Gl. 1})$$

Dabei ist  $v$  die Reaktionsgeschwindigkeit,  $[K]$  die Konzentration des Komplexes,  $-d[K]/dt$  die Abnahme des Komplexes über die Zeit,  $-d[P^N]/dt$  die Abnahme des Inteininfusionsproteins im Unterschuss und  $d[\text{SP}]/dt$  die Spleißproduktbildung über die Zeit. Die Integrationen ergeben folgende Geschwindigkeitsgesetze:

$$[K]_t = [K]_0 * e^{-k_{\text{Spleiß}} * t} \quad \text{oder} \quad [P^N]_t = [P^N]_0 * e^{-k_{\text{Spleiß}} * t} \quad (\text{Gl. 2})$$

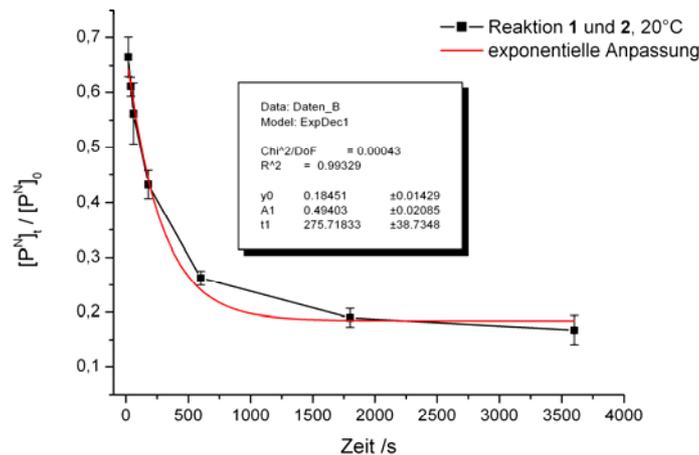
Dabei ist  $[K]_t$  die Konzentration des Komplexes zum Zeitpunkt  $t$ ,  $[K]_0$  die Anfangskonzentration des Komplexes. Analog sind die Symbole  $[P^N]_t$  und  $[P^N]_0$  für die Konzentrationen des Inteininfusionsproteins im Unterschuss zu betrachten. Die Bestimmung des Parameters  $k_{\text{Spleiß}}$ , der aufgrund des geschwindigkeitsbestimmenden Charakters der Spleißreaktion mit dem Parameter der Gesamtreaktion  $k$  gleichgesetzt wurde, erfolgte durch eine Kurvenanpassung einer exponentiellen Abnahme an die experimentellen Daten mit Hilfe des Programms Origin 6.1G (<http://www.originlab.com>). Hierzu wurde die Konzentration des Inteininfusionsproteins im Unterschuss zu den verschiedenen Zeitpunkten wie folgt berechnet. Herangezogen wurde nicht nur der Intensitätswert der Eduktbande, sondern zusätzlich auch noch die Intensität der Spleißproduktbande und einer möglichen Hydrolysebande um eine genauere Auswertung zu ermöglichen. Zu jedem Zeitpunkt  $t$  der Reaktion gilt:

$$[P^N]_t = [P^N]_0 - ([\text{SP}]_t + [\text{H}]_t) \quad (\text{Gl. 3})$$

Dabei ist  $[H]_t$  die Konzentration des zu  $P^N$ -gehörigen Hydrolyseproduktes zu dem Zeitpunkt  $t$ . In den meisten Fällen trat diese Nebenreaktion in einem vernachlässigbaren Ausmaß aus und wurde dann nicht berücksichtigt. Durch Umformen ergibt sich die auf die Anfangskonzentration normierte Konzentration des Inteininfusionsproteins im Unterschuss damit zu Gleichung 4.

$$\frac{[P^N]_t}{[P^N]_0} = \frac{[P^N]_t}{[P^N]_t + [SP]_t + [H]_t} = e^{-k_{\text{Spleiß}} \cdot t} \quad (\text{Gl. 4})$$

Die so erhaltenen, über mehrere, unabhängige Experimente gemittelten Werte wurden einer Kurvenanpassung einer exponentiellen Abnahme erster Ordnung mit dem Programm Origin 6.1G unterzogen. Ein Beispiel für die Reaktion zwischen den Proteinen 1 und 2 bei 20°C, bei der die Werte für Protein 1 aufgetragen wurden, ist in Abbildung 26 dargestellt.



**Abbildung 26** Beispiel für eine Kurvenanpassung an die Daten der Spleißreaktion der Proteine 1 und 2 bei 20°C durch exponentielle Abnahme erster Ordnung mittels des Programms Origin 6.1G.

Der erhaltene Parameter  $1/t_1 = k_{\text{Spleiß}}$  wurde als Geschwindigkeitskonstante erster Ordnung der Gesamtreaktion interpretiert. Für die monomolekulare Reaktion konnte er zusätzlich in die Halbwertszeit  $t_{1/2}$  umgerechnet werden:

$$t_{1/2} = \frac{\ln 2}{k_{\text{Spleiß}}} \quad (\text{Gl. 5})$$

## 3.5 Die SPLICEFINDER-Methode

### 3.5.1 Testexpressionen

Um zu testen, ob eine bestimmte Integrationsposition in einem Protein von Interesse spleißaktiv ist, wurden bei der SPLICEFINDER-Methode Testexpression mit anschließender Western Blot Analyse durchgeführt. Selektive Induktionen der N- und C-terminalen Inteininfusionsproteine benötigen hier die beiden regulatorischen Proteine, LacI und AraC. Deshalb wurden zuerst hitzekompetente *E. coli* BL21 Gold (DE3) Zellen mit dem Helferplasmid pRSFara (s. oben) erzeugt. Diese Zellen wurden dann mit den erzeugten Integrationsplasmiden (s. Tabelle 3) transformiert und die Selektion erfolgte auf Agarplatten mit Ampicillin und Kanamycin. Die Testexpressionen erfolgten in 4 mL LB-Medium mit 100 mg/L Ampicillin und 50 mg/L Kanamycin. Aus einer Übernachtskultur wurde eine Übertagkultur mit einer OD<sub>600</sub> von 0,2 angeimpft, bis auf eine OD<sub>600</sub> von 0,5 - 0,7 geschüttelt und dann mit 0,2 % L-Arabinose für zwei Stunden induziert. Nach einem Waschschrift (erste Zentrifugation bei 5.000 rpm für 1 min und Resuspendieren der Zellen in frischem LB-Medium mit Antibiotika und einem weiteren Zentrifugationsschritt) wurden die *E. coli* Zellen in frischem LB-Medium mit Antibiotika resuspendiert. Für die *Npu* DnaE Intein-kassetten wurden die Zellen nun für 15 min bei 37°C inkubiert, bevor die zweite Induktion mit 1,2 mM IPTG für 4,5 h gestartet wurde. Bei den *Ssp* DnaB Intein-kassetten wurde die Induktionstemperatur auf 25°C gesenkt. Zu den angezeigten Zeitpunkten wurden Proben für die SDS-PAGE und/oder Western Blot Analyse entnommen.

### 3.5.2 Western Blot Analyse

In allen Western Blots in dieser Arbeit wurde das ECL Western Blotting Kit (*GE Healthcare*) nach den Vorgaben des Herstellers verwendet. Die benutzten Verdünnungen für die Erstantikörper waren 1:2.000 (Anti-H<sub>6</sub>, *Qiagen*), 1:2.000 (Anti-ST, *IBA*) und 1:10.000 (Anti-MBP, *Fermentas*). Für alle Erstantikörper wurde der gleiche Sekundärantikörper (HRP-Anti Maus Konjugat, *GE Healthcare*) eingesetzt.

### 3.5.3 Präparation segmentell isotopenmarkierter ST-gpD-Trx-H<sub>6</sub> Proteine

#### <sup>15</sup>N(ST-gpD-Trx-H<sub>6</sub>)

Für die Expression des komplett markierten Proteins <sup>15</sup>N(ST-gpD-Trx-H<sub>6</sub>) wurden *E. coli* BL21 Gold (DE3) Zellen mit dem Plasmid pSE02 transformiert. Die Expression erfolgte im 300 mL Maßstab in M9 Minimal Medium mit <sup>15</sup>N-Ammoniumchlorid als einziger Stickstoffquelle, wie in Kapitel 3.2.1 beschrieben. Eine Ni<sup>2+</sup>-NTA Affinitätschromatographie war ausreichend, um das Protein in für NMR-Messungen akzeptabler Reinheit zu erhalten. Die Elutionsfraktionen wurden vereinigt und extensiv gegen NMR-Puffer dialysiert (2 - 3 Tage mit 4 - 5maligem Pufferwechsel). Anschließend wurde die Proteinlösung mit VivaSpin Säulen MWCO 5.000 der Firma *Sartorius* auf eine Konzentration von 500 µM gebracht. Die Konzentration wurde hierbei photometrisch bestimmt (s. Kapitel 3.7.2).

#### ST-gpD-<sup>15</sup>N(Trx-H<sub>6</sub>)

Die segmentelle Markierung des C-terminalen Fragmentes Trx-H<sub>6</sub> wurde durch *in vitro* Spleißen mit dem *Npu* DnaE Intein und gereinigten Proteinen erreicht. Die Expressionen des Proteins ST-gpD-GGG-Int<sup>N</sup> (pSE13) und des Proteins Int<sup>C</sup>-CFNK-Trx-H<sub>6</sub> (pSE12) erfolgten separat, wobei die C-terminale Hälfte in 600 mL M9 Minimal Medium mit <sup>15</sup>N-Ammoniumchlorid als einziger Stickstoffquelle exprimiert wurde. Zur Reinigung der individuellen Hälften kamen Standardmethoden zum Einsatz. Beide gereinigten Proteine wurden gegen Spleißpuffer dialysiert und in einem Verhältnis von 1,5:1 (N- zu C-terminaler Hälfte) bei 4°C über Nacht inkubiert. Die Konzentration der C-terminalen Komponente betrug hierbei 20 µM. Der Überschuss an N-terminalem Konstrukt sollte einen kompletten Umsatz der markierten C-terminalen Hälfte erzwingen. Nach der PTS-Reaktion wurde die Reaktionsmischung gegen Ni<sup>2+</sup>-NTA-Puffer A dialysiert und anschließend über eine Ni<sup>2+</sup>-NTA Säule gereinigt. Neben dem Spleißprodukt eluierte zusätzlich das C-terminale Spaltprodukt, welches mit einer weiteren StrepTag II-Reinigung abgetrennt werden konnte. In einem letzten Schritt erfolgte die Dialyse gegen NMR-Puffer und das Ankonzentrieren mittels VivaSpin Säulen MWCO 5.000 der Firma *Sartorius* auf ca. 200 µM.

#### <sup>15</sup>N(ST-gpD)-Trx-H<sub>6</sub>:

Das N-terminale ST-gpD Segment des Beispielproteins konnte durch *in vivo* PTS mit dem an der Position der Endonuklease gespaltenen *Ssp* DnaB Intein erreicht werden. Dafür wurden *E. coli* BL21 Gold (DE3) Zellen mit dem Integrationsplasmid der Kombination K3, pTS218,

und dem Helferplasmid pRSFara transformiert. Die Kombination K3 erzeugt die Aminosäuresequenz ESGSGGG in der Linkerregion des Spleißproduktes. Die Zellen wurden in 300 mL LB-Medium bei 37°C zu einer optischen Dichte von  $OD_{600} \sim 0,6$  wachsen gelassen und anschließend wurde die Expression des C-terminalen Hälfte durch die Zugabe von 0,2% L-Arabinose für 2h bei 37°C induziert. Die Zellen wurden durch Zentrifugation bei 4.000 rpm für 10 min pelletiert und zur Entfernung von überschüssiger L-Arabinose mit 100 mL unmarkiertem M9 Minimalmedium gewaschen. Nach einem weiteren Zentrifugationsschritt (4.000 rpm für 10 min) wurde das Zellpellet in 300 mL M9 Minimal Medium mit  $^{15}\text{N}$ -Ammoniumchlorid als einziger Stickstoffquelle resuspendiert und für 15 min bei 25°C unter Schütteln inkubiert. Die Induktion der N-terminalen Hälfte erfolgte durch die Zugabe von 1,2 mM IPTG für weitere 4h bei 25°C. Da unter diesen Expressionsbedingungen die C-terminale Hälfte durch PTS komplett verbraucht war, konnte das Spleißprodukt über den  $H_6$ -Affinitätstag in einem einzigen chromatographischen Reinigungsschritt in für NMR-Messungen akzeptabler Reinheit erhalten werden. Spuren des C-terminalen Spaltproduktes fanden sich in den Elutionsfraktionen, welche extensiv gegen NMR-Puffer dialysiert wurden. Über VivaSpin Säulen MWCO 5.000 der Firma *Sartorius* konnte die Konzentration des Spleißproduktes auf  $\sim 500 \mu\text{M}$  erhöht werden.

#### **3.5.4 Bestimmung der Markierungseffizienz von segmentell isotopenmarkierten Proteinen**

Zur Bestimmung der Markierungseffizienz von segmentell isotopenmarkierten Proteinen wurde in dieser Arbeit ein System entwickelt, welches auf einem tryptischen Verdau der Spleißproduktbande in einem SDS-PAGE Gel und anschließender Analyse der Peptidmischung mittels MALDI-TOF MS beruht. Die theoretischen Grundlagen der Methode, sowie einige Beispiele sind ausführlich in Kapitel 4.3.3 im Ergebnisteil präsentiert. Im kleinen Maßstab (8 mL) genauso wie in den präparativen Ansätzen erfolgte die Anzucht der Bakterienkulturen nach den Standardvorschriften (s. Kapitel 3.2.1). Ebenfalls wurden die Reinigungsprozeduren der Spleißprodukte wie beschrieben durchgeführt (s. Kapitel 3.2.2 und 3.2.3). Die Durchführung des tryptischen Verdau und die massenspektrometrische Analyse sind in Kapitel 3.7.3 dargestellt.

## 3.6 Modifikationen von NRPS-Proteinen mit Fluorophoren

### 3.6.1 Semi-synthetisches PTS mit dem an Position 11 gespaltenen Ssp DnaB Intein

Zur Einbringung unnatürlicher Funktionalitäten an den N-Terminus von Proteinen kann ein semisynthetischer Ansatz mit dem an Position 11 gespaltenen *Ssp* DnaB Intein verwendet werden. Dabei wird die C-terminale Komponente über Festphasenpeptidsynthese (SPPS) und das N-terminale Inteinprotein über rekombinante Genexpression in *E. coli* hergestellt (s. Abbildung 22). Die schematische Darstellung der Reaktion zur Generierung eines am N-Terminus mit Fluorescein modifiziertem A-PCP Protein ist in Abbildung 55 zu sehen. Zur präparativen Herstellung wurden 1 mL einer 5 - 10  $\mu$ M Lösung des Proteins **10** (ST-*Ssp* DnaB<sup>C</sup>(12-154)-A-PCP-H<sub>6</sub>) in Spleißpuffer mit 2 mM DTT (200 mM Stammlösung) versetzt. Nach Zugabe von 10 Äquivalenten **Pep2** (Biotin- $\beta$ AKKK(Fluorescein)ESG-*Ssp* DnaB<sup>N</sup>(1-11) - 1 mM Stammlösung in Spleißpuffer) wurde die Mischung bei 25°C unter Lichtausschluss für 30 h inkubiert. Zur Entfernung des überschüssigen Peptids wurde die Reaktionsmischung exzessiv gegen den Äquilibrierungspuffer (50 mM Tris - pH 8,0, 100 mM NaCl, 2 mM DTT, 1 mM EDTA) der monomeren Avidin-Säule dialysiert. Die Reinigung des biotinylierten Spleißproduktes **10mod** mit Hilfe des SoftLink<sup>TM</sup> Soft Release Avidin Resin (*Promega*) wurde nach Angaben des Herstellers durchgeführt. Nach der Äquilibrierung der Säule wurde die Reaktionsmischung mehrmals aufgetragen und die Säule mit fünf Säulenvolumina Äquilibrierungspuffer gewaschen. Die Elution erfolgte mit 5 mM Biotin in Äquilibrierungspuffer.

### 3.6.2 Modifikation der Cysteinmutanten von GrsA A-PCP

Die mittels Punktmutagenese neu in das Protein A-PCP( $\Delta$ -4Cys) eingebrachten Cysteine konnten selektiv mit 5-Iodacetamidfluorescein markiert werden. Dazu wurde eine 5 - 20  $\mu$ M Lösung des cysteinhaltigen A-PCP Konstrukts in Assay-Puffer mit 1 mM TCEP (100 mM Stammlösung) versetzt. Der pH-Wert der Lösung wurde überprüft und wenn nötig auf 7,0 eingestellt. Anschließend wurde das Protein für 5 min bei 25°C zur Reduktion der Cysteine inkubiert. Die eigentliche Reaktion fand durch Zugabe von 100 - 500  $\mu$ M 5-Iodacetamidfluorescein (10 mM Stammlösung in DMSO oder DMF) bei 25°C für 30 min statt. Nach der Reaktion wurde die Mischung mit 5 Äquivalenten DTT in Bezug auf das Iodacetamidreagenz

gequencht und alle niedermolekularen Substanzen durch Dialyse gegen Assay-Puffer entfernt. Die Markierungseffizienz wurde, wie in früheren Arbeiten, photometrisch bestimmt [211].

### 3.6.3 Sfp-katalysierte Modifikation der PCP-Domäne

Zur Markierung der PCP-Domäne mit einem synthetischen Fluorophor wurde die enzymatische Modifikation mittels der 4'-Phosphopantetheinyltransferase Sfp eingesetzt. Sfp überträgt nicht nur den Phosphopantethein-Rest von CoA auf das invariante Serin der PCP-Domänen, sondern akzeptiert ebenfalls über die Thiolfunktionalität modifizierte CoA-Analoga. Tetramethylrhodamin-CoA wurde ausgehend von Trilithium-CoA und Tetramethylrhodamin-5-maleinimid nach bekannten Vorschriften hergestellt [80,211]. Zur präparativen Herstellung von A-PCP(TAMRA) wurde eine 12,5  $\mu\text{M}$  Lösung von apo-A-PCP in Assay-Puffer zusätzlich mit 10 mM  $\text{MgCl}_2$  und 2 mM DTT, sowie 5 Äquivalenten Tetramethylrhodamin-CoA versetzt. Die Reaktion wurde durch die Zugabe von 0,02 Äquivalenten der *Bacillus subtilis* Phosphopantetheinyltransferase Sfp gestartet und bei 4°C über Nacht durchgeführt. Überschüssiges Fluorophor wurde durch Dialyse gegen Assay-Puffer entfernt.

## 3.7 Allgemeine Techniken

### 3.7.1 Präparative Herstellung von holo-A-PCP

Zur präparativen Herstellung von holo-A-PCP wurde eine 5 - 20  $\mu\text{M}$  Lösung von A-PCP in Assay-Puffer zusätzlich mit 10 mM  $\text{MgCl}_2$  und 2 mM DTT, sowie 40 Äquivalenten CoA versetzt. Die Reaktion wurde durch die Zugabe von 0,02 Äquivalenten der *Bacillus subtilis* Phosphopantetheinyltransferase Sfp gestartet und bei 4°C über Nacht durchgeführt. Überschüssiges CoA wurde durch Dialyse gegen Assay-Puffer entfernt. Eine analoge Prozedur wurde für das cystein-freie Protein A-PCP( $\Delta$ -4Cys) angewendet.

### 3.7.2 Photometrische Bestimmung der Proteinkonzentration

Die Konzentrationsbestimmung eines gereinigten Proteins in Lösung erfolgte photometrisch bei einer Wellenlänge von  $\lambda = 280 \text{ nm}$  [217]. Der Extinktionskoeffizient bei

dieser Wellenlänge wurde mit Hilfe des Programms Protean für das jeweilige Protein berechnet. Die photometrischen Messungen wurden mit einem *Pharmacia Biotech* Ultrospec 3100 Photometer in Quarzküvetten durchgeführt.

### **3.7.3 Massenspektrometrische Analyse von SDS-PAGE Gel-Proteinbanden**

Zur massenspektrometrischen Charakterisierung einzelner Proteine wurden diese nach gelelektrophoretischer Trennung auf einem SDS-PAGE Gel mit der Protease Trypsin verdaut, die die Polypeptidkette nach den basischen Aminosäuren Arginin sowie Lysin hydrolytisch schneidet. Die Massen der auf diese Weise erhaltenen kurzen Proteinfragmente konnten anschließend mittels MALDI-TOF massenspektrometrisch bestimmt werden, wodurch auf die erwartete Zusammensetzung des eingesetzten Proteins zurück geschlossen werden konnte.

Zum In-Gel Verdau einzelner Proteinbanden wurden diese nach SDS-PAGE und Anfärbung mit Coomassie Brilliant Blau knapp aus dem SDS-Gel per Skalpell ausgeschnitten und mit 300 µl Waschlösung durch Schütteln bei 37 °C entfärbt. War ein Waschschrift nicht ausreichend, so wurde die Prozedur ein weiteres Mal wiederholt. Nach Entfernung der Waschlösung wurden die Gelstücke im Vakuum getrocknet und anschließend mit 8 - 12 µl Trypsinlösung je nach Größe der Gelstücke versetzt, bis diese bei RT vollständig in das Gelstück eingezogen war. Nach Verdau über Nacht bei 37 °C wurden 15 µl Diffusionslösung hinzugefügt und das Gelstück 45 min bei Raumtemperatur im Ultraschallbad inkubiert. Zur anschließenden Analyse mittels MALDI-TOF MS wurden 0,8 µl der Peptidlösung auf der Targetplatte mit 0,8 µl Matrixlösung (2,5-Dihydroxybenzoesäure (DHB) oder  $\alpha$ -Cyano-4-hydroxymizsäure (CHCA) gesättigt in 30 % Acetonitril mit 0,1 % TFA) gemischt und bei Raumtemperatur getrocknet. Zur Kalibration des Massenspektrometers (AutoFlex II, *Bruker Daltonics*) diente der Peptide Calibration Standard I der Firma *Bruker Daltonics*.

### **3.7.4 Festphasenpeptidsynthese**

Die Festphasenpeptidsynthese wurde in dieser Arbeit entweder am Peptidsynthesizer der Firma *Multisyntech* (Syro XP) oder manuell in einer Glasfritte durchgeführt. Die Reaktionsbedingungen waren, solange im Folgenden nicht explizit daraufhin gewiesen wird, identisch. Auf die theoretischen Grundlagen der Fluorenyl-9-methyloxycarbonyl (Fmoc)-

Festphasenpeptidsynthese soll an dieser Stelle nicht eingegangen werden und stattdessen auf frühere Arbeiten verwiesen werden [211].

Die Ansatzgröße betrug 0,1 mmol und als Trägermaterial diente mit Alanin vorbeladenes Wang-Harz (*Novabiochem*). Die Kupplung einer Aminosäure erfolgte standardmäßig unter folgenden Bedingungen. Nach dem Quellen des Harzes in DMF einmalig für 10 min wurde durch zweimalige Inkubation mit einer 20%igen Piperidinlösung in DMF für 10 min die Fmoc-Gruppe der N-terminalen Aminosäure am Harz abgespalten. Nach dreimaligem Waschen mit DMF erfolgte die Kupplungsreaktion durch Zugabe von 5 Äquivalenten (eq.) des folgenden Aminosäurebausteins, 4,7 eq. HBTU und 6 eq. N-Methylmorpholin (NMM) bei manueller Kupplung oder N,N-Diisopropylethylamin (DIPEA) bei Kupplungen am Peptidsynthesizer in DMF. Die Reaktionszeiten betragen bei einer manuellen Synthese ca. 2 h und am Peptidsynthesizer 1h. Abschließend wurde das Harz wieder in drei Schritten mit DMF gewaschen, so dass ein neuer Zyklus zur Aminosäurekupplung beginnen konnte. Die Aminosäuren Ile<sup>2</sup>, Leu<sup>7</sup>, Ile<sup>8</sup> und Leu<sup>10</sup> der Int<sup>N</sup>-Sequenz wurden doppelt ohne zwischenzeitliche Abspaltung der Fmoc-Gruppe gekuppelt. Zum Einbau von 5,6-Carboxyfluorescein wurden ebenfalls zwei Inkubationsschritte durchgeführt, wobei der erste über Nacht und der zweite 8 h lang war. Die Aminosäure Cystein stellte eine Ausnahme dar und wurde nicht unter den Standard-Bedingungen eingeführt. Stattdessen wurde in diesem Fall ein Pentafluorophenyl-aktivierter Cystein-Baustein benutzt (Fmoc-Cys(Trt)-OPfp), um eine Racemisierung des Cysteins zu verhindern. Zur Kupplung von Biotin wurden analog zu den normalen Aminosäurebausteinen 5 eq. Biotin, 4,7 eq. HBTU und 6 eq. NMM in 1,5 mL DMF verwendet. Zur besseren Solubilisierung des Biotins wurden zusätzlich 4 mL NMP zugegeben. Die selektive Entschützung einer Lysinseitenkette zur späteren Modifikation mit 5,6-Carboxyfluorescein wurde mittels der säurelabilen Methyl-triphenylmethyl (Mtt)-Schutzgruppe erreicht. Dazu wurde das Harz mit Dichlormethan (DCM) gewaschen, mit einer Mischung aus TFA/DCM/TIS (1/94/5) dreimal für 30 min inkubiert und anschließend erneut mit DCM gewaschen.

Die Abspaltung des Peptids vom Harz mit gleichzeitiger Entfernung aller Schutzgruppen erfolgte mit „Reagent K“, einem Abspaltmix aus 82,5% TFA, 5% Phenol, 5% H<sub>2</sub>O, 5% Thioanisol und 2,5% Ethandithiol, unter Standardbedingungen. Die Reinigung von **Pep1** konnte über präparative HPLC-Läufe nach bereits beschriebenen Bedingungen [211] erfolgen. **Pep2** konnte aufgrund von Löslichkeitsproblemen nur im Basischen über eine Anionenaustauschsäule gereinigt werden. Dazu wurde das lyophilisierte Crude-Peptid in Anionenaustauschpuffer A (20 mM Tris, 2 mM DTT, pH 8,0) gelöst. Der pH-Wert musste

nachträglich mit NaOH eingestellt werden. Nach Abtrennung der unlöslichen Bestandteile durch Zentrifugation wurde der Überstand auf eine „HiTrap-Q-HP“ Anionenaustauscher-Säule der Firma *GE Healthcare* aufgetragen. Die Elution des Peptids von der Säule erfolgte mit einem linearen Gradienten von 0 nach 100 % Anionenaustauschpuffer B (20 mM Tris, 2 mM DTT, 500 mM NaCl, pH 8,0) mit einer Flussrate von 0,5 mL/min über 30 min. Die gewünschten Fraktionen wurden vereinigt und nach Dialyse gegen ddH<sub>2</sub>O mit 2 mM Mercaptoethanol konnte das gereinigte Peptid lyophilisiert werden.

### 3.7.5 NMR-Spektroskopie

Die kernmagnetische Resonanz beruht auf der Wechselwirkung zwischen dem magnetischen Moment des Atomkerns und magnetischen Feldern. Das magnetische Moment des Kerns ist mit dem Kernspin  $S$  verknüpft. In der hochauflösenden NMR-Spektroskopie verwendet man hauptsächlich Kerne mit einem Kernspin  $S = \frac{1}{2}$ . Für die NMR-spektroskopische Untersuchung von Proteinen sind hier  $^1\text{H}$ ,  $^{13}\text{C}$  und  $^{15}\text{N}$  zu nennen. Kerne, wie  $^{12}\text{C}$ ,  $^{16}\text{O}$  und  $^{18}\text{O}$ , haben einen Kernspin  $S$  von null und sind damit NMR-inaktiv.  $^{14}\text{N}$  hat einen Kernspin von größer  $\frac{1}{2}$  und besitzt damit ein Quadrupolmoment, das zu sehr breiten NMR-Signalen führt. In einem von außen angelegten Magnetfeld  $B_0$  orientieren sich Spin- $\frac{1}{2}$ -Kerne entweder parallel oder antiparallel zu diesem Feld. Der parallel zum Feld orientierte  $\alpha$ -Zustand ist energetisch günstiger als der antiparallele  $\beta$ -Zustand. Aufgrund der Boltzmann-Verteilung ist der  $\alpha$ -Zustand je nach Größe des Energieunterschieds stärker besetzt als der  $\beta$ -Zustand. Daraus resultiert eine makroskopische Magnetisierung der Probe in Richtung des  $B_0$ -Feldes. Die Energiedifferenz der beiden Zustände  $\Delta E$  ist das Produkt aus dem Planckschen Wirkungsquantum  $h$  und der Larmor-Frequenz des Kerns  $\nu_0$ . Letztere wiederum kann als die Frequenz mit der der Kern um den  $B_0$ -Vektor präzessiert angesehen werden und ist proportional zum angelegten Feld  $B_0$  und dem gyromagnetischen Verhältnis  $\gamma$  des Kerns.

$$\Delta E = h \nu_0 = \frac{h \gamma B_0}{2 \pi} \quad (\text{Gl. 6})$$

Das gyromagnetische Verhältnis  $\gamma$  ist eine Eigenschaft jedes einzelnen Atomkerns. Für  $^1\text{H}$  ist dieser Wert, z.B., viermal größer als für  $^{13}\text{C}$  und zehnmal größer als für  $^{15}\text{N}$ . Daraus resultieren eine größere Larmorfrequenz  $\nu_0$ , eine größere Energiedifferenz  $\Delta E$  und damit auch eine größere makroskopische Magnetisierung für Protonen im Vergleich zu  $^{13}\text{C}$  und  $^{15}\text{N}$ . Sie sind damit am einfachsten in NMR-Experimenten zu detektieren. Das Magnetfeld eines NMR-Spektrometers wird nicht in Tesla, sondern in der theoretischen Larmorfrequenz eines

$^1\text{H}$ -Kerns angegeben. So ergibt ein  $B_0$ -Feld von 7,05 T eine Resonanzfrequenz von 300,1 MHz. Die eigentliche Resonanzfrequenz eines Protons in einer chemischen Verbindung wird etwas von diesem theoretischen Wert abweichen, da es anhand der unterschiedlichen chemischen Umgebung auch ein anderes effektives Magnetfeld  $B_{\text{eff}}$  spürt.

$$B_{\text{eff}} = B_0 (1 - \sigma) \quad (\text{Gl. 7})$$

Hier ist  $\sigma$  die Abschirmungskonstante, die ausdrückt wie stark die Elektronen den Kern von dem äußeren Magnetfeld abschirmen. Diese auch chemische Verschiebung genannte Eigenschaft ist im Vergleich mit dem  $B_0$ -Feld sehr klein und wird in ppm (engl. *parts per million*) angegeben. In einem Molekül besitzt jeder Atomkern seine eigene charakteristische chemische Verschiebung, solange er nicht mit anderen Kernen magnetisch äquivalent ist. In dieser Arbeit wurde sich auf die NMR-spektroskopische Untersuchung der Stickstoffkerne und der Amidprotonen in der Peptidbindung von Proteinen konzentriert. Die chemische Verschiebung dieser Kerne lässt Rückschlüsse auf den Faltungszustand des Proteins zu. In unstrukturierten Peptiden und Proteinen variieren die chemischen Verschiebungen nicht sehr stark, man spricht von einer geringen Dispersität. So besitzen zwar die Aminosäureseitenketten und die Nachbaraminosäuren einen Einfluss auf die chemische Verschiebung des Amidprotons in einer Peptidbindung, allerdings sind alle Resonanzen im schmalen Bereich von 8 bis 9 ppm zu finden. Vergleichbares gilt für die  $^{15}\text{N}$ -Kerne der Peptidbindung, so dass in einem als „random coil“ vorliegendem Protein alle Amid-Resonanzen in einen relativ kleinen Ausschnitt des Spektrums zu finden sein sollten. In gefalteten Proteinen treten Abweichungen der chemischen Verschiebungen für einzelne Aminosäuren von den „random coil“-Werten auf. Die spezifische Anordnung des Proteinerückgrats und der Seitenketten führt dazu, dass für einzelne Kerne ein unterschiedliches effektives magnetisches Feld vorliegt. Diese anisotropischen Effekte gehen hauptsächlich auf aromatische Seitenketten (Ringströme), aber auch auf die Positionierung der Peptidbindung in Sekundärstrukturelementen, wie  $\alpha$ -Helices oder  $\beta$ -Faltblättern, zurück. Die NMR-Signale eines gefalteten Proteins müssen nicht zwangsweise dispers sein. Treten jedoch in dem NMR-Spektrum eines Proteins Signale auf, die stark von den „random coil“-Werten abweichen, so ist der Rückschluss möglich, dass es sich hierbei um ein gefaltetes Protein handelt.

Neben der chemischen Verschiebung ist das Phänomen der Kopplung zwischen Atomkernen von enormer Wichtigkeit für die NMR-Spektroskopie. So beruhen die in dieser Arbeit durchgeführten  $^{15}\text{N}$ ,  $^1\text{H}$ -Korrelationsspektren (TROSY - transverse relaxation optimized spectroscopy [218]) auf der über eine Bindung vermittelten Kopplung zwischen

dem  $^{15}\text{N}$ -Kern und seinem Proton in der Peptidbindung. Das Proton ist hier entweder im  $\alpha$ - oder im  $\beta$ -Zustand und sorgt damit für zwei unterschiedliche effektive magnetische Felder, denen der  $^{15}\text{N}$ -Kern ausgesetzt ist. Umgekehrt gilt der analoge Fall. Die zugehörige Kopplungskonstante ist  $J$ , in dem Fall der Peptidbindung  $^1J_{\text{NH}} \approx 90$  Hz, und erlaubt es durch Einstrahlung von Radiofrequenzpulsen in die Probe Magnetisierung von den Amid-Protonen auf die Stickstoffkerne und wieder zurück zu übertragen - diese damit miteinander zu korrelieren. Ein zugehöriges 2D-Spektrum wird üblicherweise als ein Contour-Plot mit einer  $^1\text{H}$ - und einer  $^{15}\text{N}$ -Frequenzachse dargestellt. Ein Signal bei 128 ppm in der  $^{15}\text{N}$ - und bei 9,8 ppm in der  $^1\text{H}$ -Achse bedeutet, dass ein Proton mit einem Stickstoffatom bei diesen Resonanzfrequenzen direkt über eine Bindung verbunden ist. Eine exakte Zuordnung zu welcher Aminosäure dieses Amidpaar gehört kann zwar nicht getroffen werden, aber aufgrund der chemischen Verschiebung ist die Aussage möglich, dass sich diese Peptidbindung in einem gefalteten Bereich des Proteins befindet. Auf eine detaillierte Beschreibung der verwendeten NMR-Experimente und Pulssequenzen wird an dieser Stelle verzichtet und auf allgemeine Lehrbücher der NMR-Spektroskopie verwiesen.

Auch anhand der gemessenen Linienbreiten der NMR-Signale der Inteinhälften wurden in dieser Arbeit Rückschlüsse auf den Faltungszustand dieser Proteinfragmente gezogen (s. Kapitel 4.2.2). Die Linienverbreiterung spielt zusätzlich in den Spektren der NRPS-Multidomänenproteine mit sehr großem Molekulargewicht eine entscheidende Rolle (s. Kapitel 4.4.2). An dieser Stelle soll kurz auf die wichtigsten Parameter und deren Einfluss auf die Linienbreite von NMR-Signalen eingegangen werden. Einen detaillierteren Überblick findet man in den Lehrbüchern der NMR-Spektroskopie. Im Laufe eines NMR-Experimentes wird die, durch das angelegte äußere Magnetfeld  $B_0$  entstandene makroskopische Magnetisierung, durch Radiofrequenzpulse, die senkrecht zu dem statischen Feld eingestrahlt werden, beeinflusst. Man bezeichnet diese magnetischen Felder mit  $B_1$ . Sie können sowohl eine Besetzungsinversion der Zustände (mehr Kerne im  $\beta$ -Zustand, als im  $\alpha$ -Zustand), als auch eine Organisation der einzelnen magnetischen Kernmomente senkrecht zum  $B_0$ -Feld, die so genannte Kohärenz, erzeugen. Diese Kohärenz der magnetischen Momente dreht sich in der x-y Ebene, die senkrecht zum  $B_0$ -Feld liegt, und kann damit durch Empfängerspulen detektiert werden. Sie erzeugt den aufgenommenen FID (engl. *free induction decay*), bei dem nach der Pulssequenz des NMR-Experiments, die Magnetisierung der Probe wieder in den Ausgangszustand zurückgeht. Dabei gleicht sich die Besetzung der Zustände wieder an das Ausgangsniveau an ( $T_1$  - longitudinale Relaxation) und außerdem geht die Kohärenz der magnetischen Momente verloren ( $T_2$  - transversale Relaxation). Aus dem FID wird durch

Fourier-Transformation das eigentliche Spektrum generiert. Bei langen transversalen Relaxationszeiten ist eine lange Aufnahme eines FID möglich und dies führt zu sehr kleinen Linienbreiten im NMR-Spektrum. Umgekehrt wird durch einen schnellen Verlust der Kohärenz nur ein zeitlich kurzer FID aufgenommen und die Fourier-Transformation erzeugt ein Spektrum mit großen Linienbreiten. Neben technischen Problemen, wie ein inhomogenes Magnetfeld, können die intrinsischen Eigenschaften der Probe zu einer schnellen transversalen Relaxation führen. So weisen langsam rotierende Moleküle, wie große Proteine, aufgrund von Dipol-Dipol Kopplungen und der chemischen Verschiebungsanisotropie (CSA) sehr kurze  $T_2$ -Zeiten auf. Bei der CSA besitzt ein Kern je nach Orientierung des Moleküls zum äußeren Magnetfeld unterschiedliche chemische Verschiebungswerte. Durch langsame Rotation des Moleküls mitteln sich diese während der Aufnahme des FID nicht heraus und sorgen damit für einen schnellen Abbau der Kohärenz. Neben experimentellen Methoden, wie der TROSY-Spektroskopie, führt eine Deuterierung des Proteins aufgrund des geringeren magnetischen Moments von  $^2\text{H}$  im Vergleich mit  $^1\text{H}$  zu besseren Relaxationseigenschaften der Probe. Bei thermostabilen Proteinen ist zusätzlich eine Aufnahme des Spektrums bei höheren Temperaturen vorzuziehen. Neben der langsamen Rotation kann noch eine weitere Eigenschaft eines Proteins zu großen Linienbreiten im NMR-Spektrum führen. Befindet sich ein Atomkern aufgrund einer konformationellen Änderung des Moleküls in zwei oder mehr unterschiedlichen Umgebungen, die zu unterschiedlichen chemischen Verschiebungen führen, so kann dies je nach Geschwindigkeit der Umwandlung in Linienverbreiterung resultieren. Bei einem im Vergleich mit der NMR-Zeitskala (Millisekunden) schnellen Austausch mitteln sich die Signale der einzelnen Konformation zu einem Signal. Eine langsame Umwandlung hat getrennte Signale zur Folge und im Zwischenbereich werden verbreiterte Linien detektiert.

Die eigentlichen NMR-Messungen wurden an der Goethe-Universität Frankfurt am Main im Arbeitskreis Prof. V. Dötsch von Alena Busche und Dr. Frank Löhr durchgeführt. Alle Proteine wurden nach der Affinitätschromatographischen Reinigung extensiv gegen NMR-Puffer ohne Glycerin dialysiert. Für die NRPS-Proteine wurde zusätzlich 10 mM  $\text{MgCl}_2$  zugegeben. Direkt vor der Messung wurde die Probe mit 10%  $\text{D}_2\text{O}$ , mit 2,2-Dimethyl-2-silapentan-5 sulfonsäure (DSS) zur Referenz und mit dem Complete Proteaseinhibitor der Firma *Roche*, sowie 2 mM DTT versetzt. Die Messungen erfolgten bei 20°C.

---

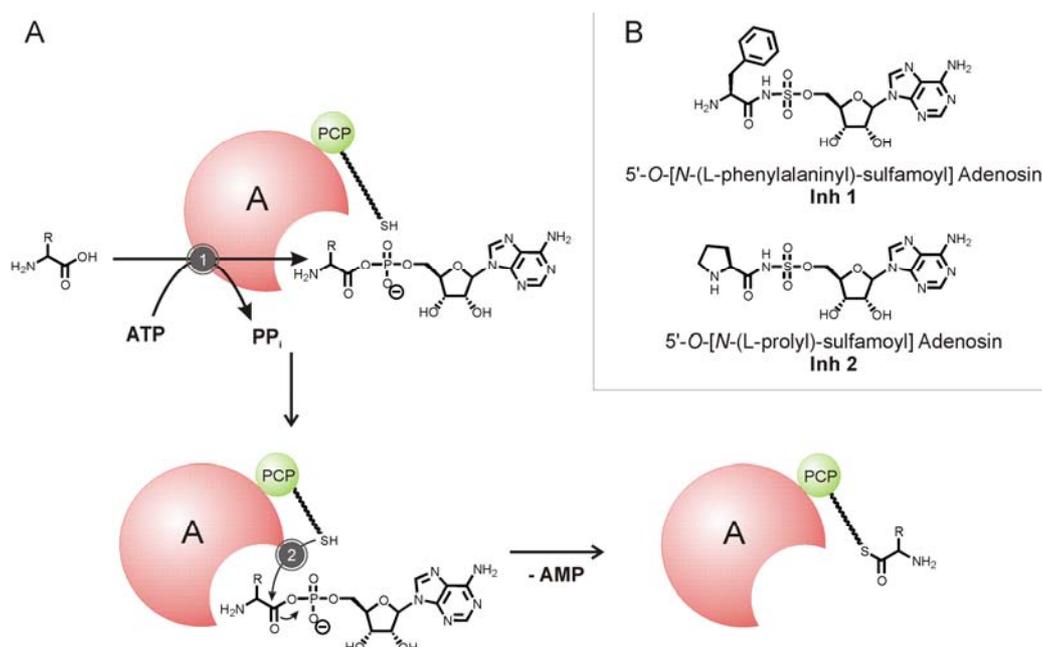
## 4 Ergebnisse

Das Vorhaben die Domäneninteraktionen innerhalb der NRPS Enzymkomplexe im Detail sowohl zu untersuchen, als auch zu verstehen führte im Laufe dieser Doktorarbeit zu verschiedenen Projekten. Neben der biochemischen Charakterisierung der Wechselwirkungen von Adenylierungs- und Peptidyl Carrier Protein-Domänen in einem NRPS Initiationsmodul (s. Kapitel 4.1) wurden auch Untersuchungen an gespaltenen Inteinen durchgeführt. So wurde in Zusammenarbeit mit Vivien Schütz das natürlich gespaltene DnaE Intein aus *Nostoc punctiforme* biochemisch charakterisiert und strukturell mittels NMR-Spektroskopie analysiert (s. Kapitel 4.2). Gemeinsam mit Tim Sonntag und Simone Eppmann gelang die Entwicklung eines Systems zur schnellen und einfachen Aktivitätsbestimmung von gespaltenen Inteinen in Zielproteinen (s. Kapitel 4.3). Diese Inteinstudien sollten in Zukunft die generellen Einsatzmöglichkeiten von gespaltenen Inteinen, unter anderem auch bei dem gezielten Einbau von biophysikalischen Sonden in NRPS-Multidomänenproteine, verbessern. Spektroskopischen Untersuchungen wie die NMR- oder die Fluoreszenzspektroskopie, die diese biophysikalischen Sonden benötigen, werden für ein Verständnis der NRPS Domäneninteraktion mit atomarer Auflösung notwendig sein. Kapitel 4.4 stellt die in dieser Arbeit durchgeführten Vorarbeiten für spektroskopische Untersuchungen von NRPS-Proteinen vor.

### 4.1 Wechselwirkungen zwischen NRPS A- und PCP-Domänen

Wie in der Einleitung beschrieben steht dem tiefen Verständnis der individuellen NRPS-Domänen noch ein relativ geringes Wissen über das Zusammenspiel der Domänen während der Naturstoffsynthese gegenüber. Für die PCP-Domänen, den zentralen Ankerpunkten der Peptidintermediate, werden anhand der Kristallstruktur eines kompletten Terminationsmoduls [94] Verschiebungen relativ zu den anderen Domänen postuliert [107] (s. Kapitel 1.1.6 für Details). Die Schwäche der kristallographischen Studien besteht jedoch darin, dass sie nur eine Momentaufnahme auf einen einzigen konformationellen Zustand während des komplizierten Katalysezyklus ermöglichen. Darüberhinaus gelingt die Kristallisation der NRPS-Proteine oft nur mit einer inaktiven apo-Form der PCP-Domäne. Die Dynamik in den NRPS Biosynthesemaschinen und deren Steuerung lassen sich mittels dieser Technik nicht entschlüsseln. Der in dieser Arbeit verfolgte Ansatz basierte im

Gegensatz dazu auf der biochemischen Untersuchung von katalytisch aktiven Proteinen in Lösung. Als Modellprotein wurde ein Didomänenkonstrukt aus dem Initiationsmodul GrsA der Gramicidin S Biosynthese, bestehend aus einer Adenylierungs- und einer Peptidyl Carrier Protein-Domäne, verwendet (s. Kapitel 1.1.3). Die C-terminale Epimerisierungsdomäne von GrsA wurde durch einen His<sub>6</sub>-Tag ersetzt [211]. Mit der holo-Form dieses Konstruktes lassen sich zwei Reaktionsschritte der NRPS Katalyse untersuchen - die von der A-Domäne katalysierte Bildung des Aminoacyladenylats ausgehend von ATP und einer Aminosäure (s. Schritt 1 in Abbildung 27A), sowie die Übertragung der aktivierten Aminosäure auf den 4'-Phosphopantethein-Arm der PCP-Domäne (s. Schritt 2 in Abbildung 27A). Der zweite Reaktionsschritt benötigt eine korrekte Interaktion zwischen der A- und der PCP-Domäne. Diese Wechselwirkung ist in dieser Form auch in Elongations- und Terminationsmodulen zu finden (s. 1.1.2).



**Abbildung 27 Schematische Darstellung der Reaktionsschritte des in dieser Arbeit verwendeten GrsA A-PCP Didomänenkonstruktes.** A) Holo-A-PCP ist in der Lage eine Aminosäure (bevorzugt Phenylalanin) unter ATP-Verbrauch zum Aminoacyladenylat zu aktivieren (Schritt 1); bei der kovalenten Beladung des Ppant-Arms der holo-PCP Domäne müssen die A- und die PCP produktiv interagieren (Schritt 2), es bildet sich der A-PCP-Thioester; B) Chemische Struktur der in dieser Arbeit verwendeten Sulfamoyl-Inhibitoren; die nicht-hydrolysierbaren Analoga der Acyladenylate binden mit hoher Affinität an die korrespondierenden A-Domänen [219].

Mit den natürlichen Substraten ATP und L-Phe lässt sich eine Fixierung der produktiven Interaktion der beiden Domänen im holo-Protein nicht erreichen. Wie in Abbildung 27 gezeigt, führt die Reaktionssequenz zu einem L-Phe-PCP Thioester, welcher nach der biosynthetischen Logik nicht mehr mit der A-Domäne interagieren sollte. Im kompletten

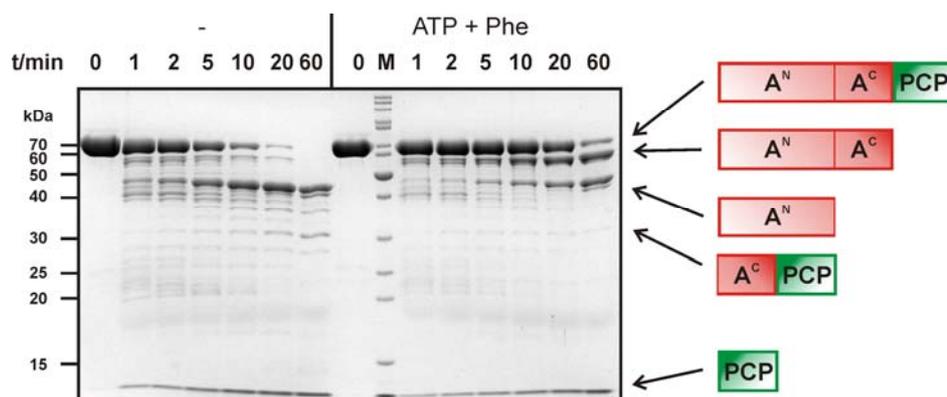
GrsA Protein müsste die beladene PCP-Domäne für den nächsten Reaktionsschritt mit der Epimerisierungs-Domäne wechselwirken. Jedoch bieten strukturelle Analoga der Aminoacyladenylate, die nicht hydrolysiert werden können, theoretisch die Möglichkeit das A-PCP Didomänenkonstrukt vor der Stufe der Thioesterbildung zu blockieren. Frühere Studien zeigten, dass 5'-O-[N-(L-phenylalaninyl)-sulfamoyl] Adenosin (**Inh 1**) (s. Abbildung 27B) die GrsA A-Domäne im nanomolaren Bereich inhibiert [219]. Die starke Bindung des Inhibitors an das Protein induziert mit hoher Wahrscheinlichkeit die Thioesterkonformation in der A-Domäne (für Details s. Kapitel 1.3). Die Sulfamoyladenosen-Inhibitoren (zur Verfügung gestellt von Dr. Martin Hahn und Prof. Mohamed Marahiel, Philipps-Universität Marburg) wurden in dieser Arbeit zur Untersuchung der Domäneninteraktion in dem GrsA A-PCP Didomänenkonstrukt verwendet.

#### **4.1.1 Der Einfluss der Sulfamoyl-Inhibitoren auf die Konformation von GrsA A-PCP im Vergleich mit den Substraten ATP und L-Phe**

Vor weiterführenden biochemischen Untersuchungen der Domäneninteraktion zwischen der A- und der PCP-Domäne musste zuerst sichergestellt werden, dass die Sulfamoylinhibitoren einen vergleichbaren Einfluss auf die Konformation der A-Domäne besitzen wie die natürlichen Substrate ATP und die Aminosäure. Die bisher einzigen veröffentlichten Studien zur Bestimmung von konformationellen Änderungen in NRPS Proteinen basierten auf partiellen Proteolyseexperimenten [220,221], einer allgemeinen Technik zur Untersuchung des strukturellen Aufbaus von Proteinen [222]. Unter der Annahme, dass die ersten Wechselwirkungen zwischen der Protease und dem Proteinsubstrat an flexiblen Linkern zwischen kompakten Domänen oder an hoch mobilen und gut zugänglichen Bereichen stattfinden, konnten mit der nicht vollständigen Proteolyse auch konformationelle Änderungen von Proteinen bei der Bindung von Liganden nachgewiesen werden [223,224]. Eine Voraussetzung für diese Methode ist jedoch, dass die unterschiedlichen Konformationen der Proteine auch unterschiedlich von der Protease geschnitten werden und daraus eine Veränderung des Verdaumusters oder der Kinetik der Proteolyse resultiert. Dieckmann *et al.* [220,221] benutzten die partielle Proteolyse zur Untersuchung der apo-Form der Tyrocidin A Synthetase I (TycA). Dieses NRPS Protein besitzt eine identische Domänenanordnung wie GrsA (A<sub>Phe</sub>-PCP-E). Die Inkubation von apo-TycA mit substöchiometrischen Mengen an Proteasen, wie Trypsin und Proteinase K, bestätigte den allgemeinen Domänenaufbau und die Unterteilung der A-Domäne in zwei

Subdomänen (s. Kapitel 1.1.4) [221]. Bei Inkubation des Proteins mit den Substraten ATP und L-Phe veränderte sich das Fragmentierungsmuster nicht, jedoch erwies sich eine identifizierte Proteaseschnittstelle zwischen den Subdomänen der A-Domäne als schlechter zugänglich [221] und damit konnte die partielle Proteolyse als Methode zur biochemischen Charakterisierung von konformationellen Änderungen in NRPS A-Domänen etabliert werden. Spätere Kristallstrukturanalysen an Mitgliedern der ANL-Superfamilie erklärten die veränderte Kinetik der Proteolyse mit dem Übergang der A-Domäne von der offenen in eine geschlossene Konformation bei der Bindung der Substrate (s. Kapitel 1.1.4 und [53]). In Bezug auf diese Vorarbeiten sollte in der vorliegenden Arbeit der Einfluss der Sulfamoylinhibitoren auf die Konformationen des GrsA A-PCP Didomänenkonstruktes im Vergleich zu den natürlichen Substraten mittels partieller Proteolyse untersucht werden.

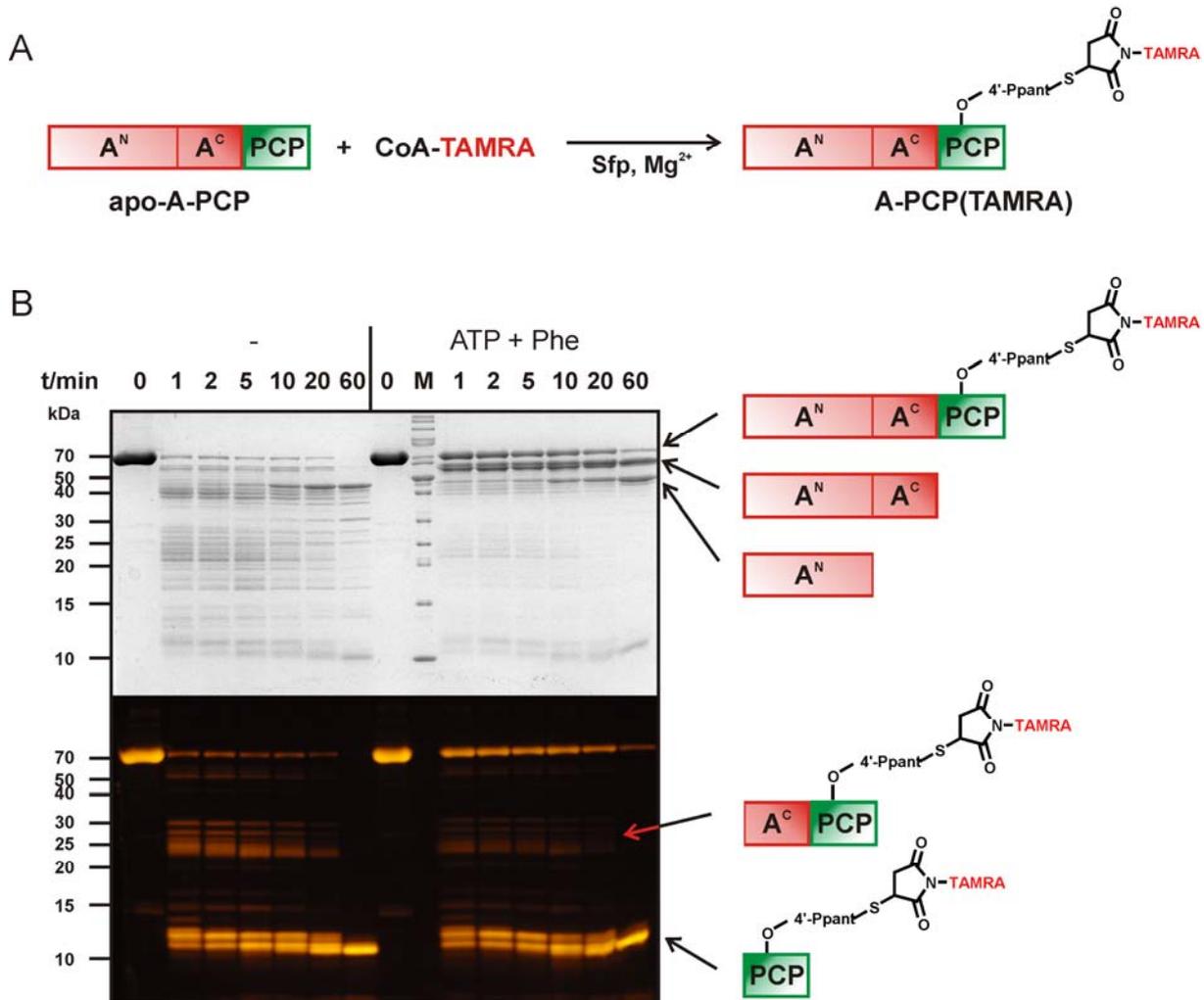
Für apo-GrsA A-PCP wurde in guter Übereinstimmung mit den Arbeiten von Dieckmann *et al.* eine veränderte Kinetik der Proteolyse bei Inkubation des Proteins mit ATP und L-Phe nachgewiesen. Wie in Abbildung 28 zu erkennen ist, befindet sich die bevorzugte Proteaseschnittstelle in Abwesenheit der Substrate zwischen den Subdomänen der A-Domäne. In der Präsenz von ATP und L-Phe (jeweils 2 mM) zeigt das A-PCP Protein, wahrscheinlich aufgrund einer kompakteren, schlechter zugänglichen Konformation, eine größere Resistenz gegenüber der Protease Trypsin.



**Abbildung 28** Partieller tryptischer Verdau von apo GrsA A-PCP in Lösung. SDS-PAGE Analyse der Verdau in Abwesenheit (links) und in Gegenwart (rechts) der Substrate ATP und L-Phe (2 mM); das Protease zu Protein Verhältnis betrug in beiden Fällen 1:250 (w/w).

Die Identifikation der durch die Proteolyse entstandenen Proteinfragmente gelang im Experiment von Abbildung 28, sowie in allen weiteren Experimenten durch massenspektrometrische Analyse (tryptischer in-Gelverdau mit anschließender MALDI-TOF

MS) und mittels Vergleich mit einem fluorophor-markierten A-PCP Protein (für Details s. Abbildung 29 und Abbildung 71A im Anhang).



**Abbildung 29 Herstellung und tryptischer Verdau in Lösung von A-PCP(TAMRA).** A) Die kovalente Verknüpfung eines Fluorophors an die PCP-Domäne des Didomänenkonstruktes A-PCP gelang durch eine Sfp-katalysierte Reaktion mit CoA-TAMRA; B) SDS-PAGE Analyse der Verdau von A-PCP(TAMRA) in Abwesenheit (links) und in Gegenwart (rechts) der Substrate ATP und L-Phe (2 mM); unten ist das Gel unter UV-Bestrahlung zu sehen, oben dasselbe Gel in der Coomassie-Färbung; das Protease zu Protein Verhältnis betrug in beiden Fällen 1:250 (w/w); durch den Vergleich des Gels mit der UV-Bestrahlung und mit der Coomassie-Färbung lassen sich die Proteinfragmente identifizieren, die die PCP-Domäne enthalten - zusammen mit einer Größenabschätzung und dem tryptischen in-Gel Verdau konnte eine exakte Zuordnung der erhaltenen Proteinfragmente erfolgen.

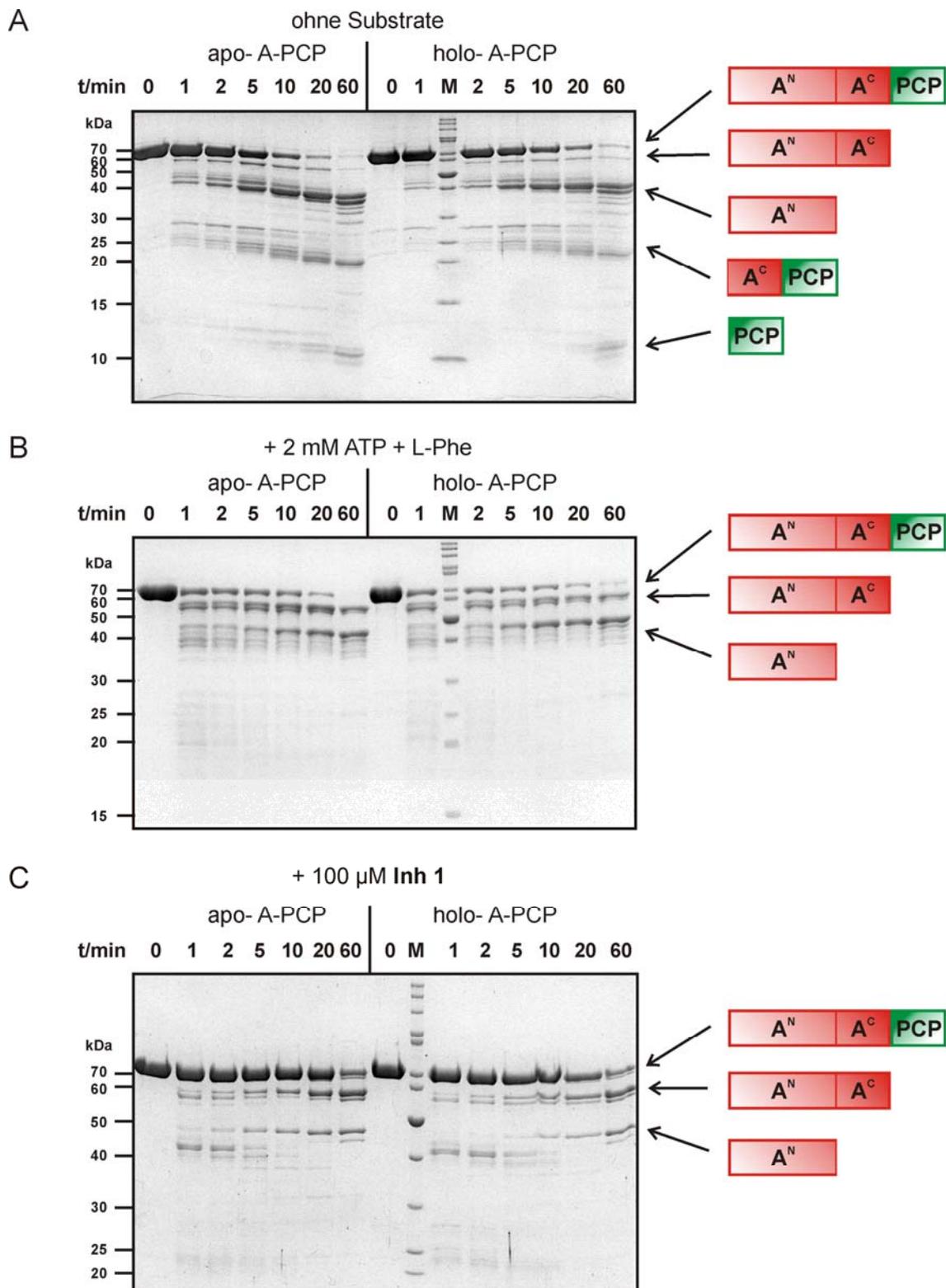
In Kontrollreaktionen, bei denen das Protein mit Pyrophosphat, AMP und ATP alleine inkubiert wurden, zeigte sich kein Einfluss dieser Moleküle auf die Kinetik des tryptischen Verdau (Daten nicht gezeigt). Die Inkubation mit L-Phe allein verlangsamte die Proteolyse zwischen den beiden Subdomänen der A-Domäne, jedoch nicht so stark wie in Kombination mit ATP. Der inhibierende Effekt von L-Phe konnte durch die zusätzliche Zugabe von AMP und Adenosin-5'-[( $\alpha,\beta$ )-methylen]-triphosphat (ApCpp) leicht verstärkt werden (Daten nicht

gezeigt). Diese zu früheren Arbeiten [221] vergleichbaren Ergebnisse bestätigten die Einsetzbarkeit der partiellen proteolytischen Verdau zur Untersuchung von konformationellen Änderungen in A-Domänen.

Der Vergleich des Einflusses von 5'-O-[N-(L-phenylalaninyl)-sulfamoyl] Adenosin (**Inh 1**) mit dem natürlichen Substrate ATP und L-Phe auf sowohl das apo-, als auch auf das über eine *in vitro* Modifikation mit Sfp und CoA hergestellte holo-A-PCP Protein ist in Abbildung 30 gezeigt. Auch die Zugabe von **Inh 1** (s. Abbildung 30 C) verschlechterte die Zugänglichkeit der Schnittstelle zwischen den A<sup>N</sup> und A<sup>C</sup>-Subdomänen, analog zur Inkubation mit ATP und L-Phe (s. Abbildung 30 B). Um in der untersuchten Zeitspanne von einer Stunde überhaupt einen messbaren Verdau des mit **Inh 1** inkubierten Proteins zu erhalten, musste sogar die Proteasemenge im Vergleich zu den ATP und L-Phe-Experimenten vervierfacht werden. Diese signifikant größere Proteolyseresistenz kann mit der sehr starken Bindung des Inhibitors an die A-Domäne ( $K_i = 61 \text{ nM}$  [219]) und dadurch einer effektiveren Fixierung des Proteins in einer kompakten Konformation erklärt werden. Die Inkubation mit einem nicht kognaten Inhibitor der GrsA A-Domäne, 5'-O-[N-(L-prolyl)-sulfamoyl] Adenosin (**Inh 2**), ergab keine Veränderung der Proteolysekinetik im Vergleich zum Protein ohne Substrate oder Inhibitoren (Daten nicht gezeigt). Interessanterweise zeigte auch die post-translationale Modifikation der PCP-Domäne mit dem Ppant-Arm keinen gravierenden Einfluss. Im Vergleich mit dem apo-Protein lieferte das holo-Protein ohne Substrate (s. Abbildung 30 A), mit ATP und L-Phe (s. Abbildung 30 B) und mit **Inh 1** (s. Abbildung 30 C) ähnliche Proteolysemuster und Kinetiken. Vor allem bei der Inkubation mit ATP und L-Phe sollten unterschiedliche Zustände der PCP-Domäne zu erwarten sein, da im Fall des holo-Proteins eine L-Phe-Ppant Thioester vorliegt. Jedoch konnte keine Veränderung im partiellen Verdau festgestellt werden.

Zusammenfassend lassen sich zwei Aussagen von den oben beschriebenen Beobachtungen bezüglich der partiellen Proteolyse ableiten. Erstens scheint **Inh 1**, im Rahmen der Auflösung dieses Assays, die gleiche Konformation in der A-Domäne auszulösen wie die Substrate ATP und L-Phe. **Inh 1** kann damit als analog zu dem Phenylalaninacyladenylat angesehen werden. Zweitens sind die partiellen Proteolyseexperimente nicht in der Lage mögliche Unterschiede in der Position der PCP-Domäne relativ zur A-Domäne aufzulösen. Unter Umständen ist der Effekt potentieller Domänen-Wechselwirkungen auf die Zugänglichkeit der Proteaseschnittstellen zu klein. Eine kinetische Analyse der Schnittstelle zwischen den Domänen wird dadurch erschwert, dass ein

entstehendes PCP-Fragment entweder aus einem intakten Protein oder aus einem schon vorher geschnittenen A<sup>C</sup>-PCP-Fragment hervorgehen kann.



**Abbildung 30** Partiieller tryptischer Verdau von apo- und holo-GrA A-PCP unter verschiedenen Bedingungen. Die Verdau für das apo-Protein sind links, die für das holo-Protein sind rechts dargestellt; A) Reaktionen ohne Substrate (Protease zu Protein 1:250 (w/w)); B) Vorinkubation mit 2 mM ATP und L-Phe (Protease zu Protein 1:100 (w/w)); C) Vorinkubation mit 100  $\mu$ M Inh 1 (Protease zu Protein 1:25 (w/w)).

### 4.1.2 Globale Konformationsanalyse von A-PCP Didomänenproteinen

Sowohl die in Kapitel 4.1.1 dargestellten Daten der partiellen Proteolyse, als auch der für A-Domänen vorgeschlagene Katalysezyklus weisen daraufhin, dass das untersuchte NRPS A-PCP Fragment eine kompaktere Konformation durch die Bindung der Substrate oder der Sulfamoylinhibitoren einnimmt. Zur Bestimmung der globalen Konformation von Proteinen können sowohl die Gelfiltrationschromatographie, als auch die native Polyacrylamid Gelelektrophorese (native PAGE) eingesetzt werden. Erstere kann über Bestimmung des Stokes Radius, zweitere über eine Änderung in der elektrophoretischen Mobilität verschiedene konformationelle Zustände eines Proteins detektieren, wenn diese die Form und Größe des Proteins beeinflussen. In dieser Arbeit wurde mittels dieser Techniken die apo- und die holo-Form von GrsA A-PCP in Abwesenheit und Gegenwart von Substraten und Inhibitoren untersucht.

In den Gelfiltrationsexperimenten wurde das Elutionsvolumen der A-PCP Proteine ohne Substrate und Inhibitoren auf einer Superdex 200 10/300 GL Säule (GE Healthcare) durch die post-translationale Modifikation des Ppant-Arms nicht beeinflusst. So eluierten apo- und holo-A-PCP innerhalb der Fehlergrenzen mit identischen Retentionszeiten (s. Tabelle 4) (t-Test der Mittelwerte der Elutionsvolumina: die Nullhypothese apo = holo kann mit einem Signifikanzniveau von  $\alpha = 0,05$  nicht verworfen werden).

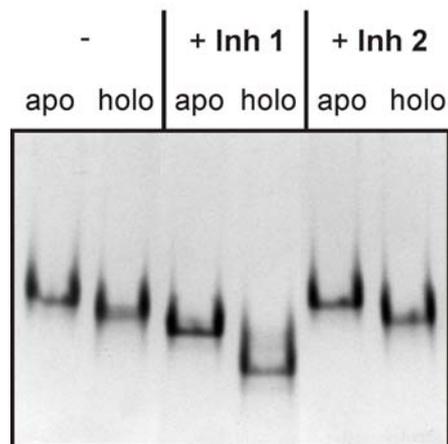
**Tabelle 4 Ergebnisse der Gelfiltrationsanalyse von unterschiedlich inkubierten GrsA A-PCP Konstrukten.**

Protein	MW <sub>theor.</sub> /kDa	Elutionsvolumen /mL	MW <sub>berechnet</sub> * /kDa
apo-A-PCP	71,1	14,560 ± 0,028	74,2 ± 1,1
holo-A-PCP	71,5	14,567 ± 0,023	73,9 ± 0,9
apo-A-PCP + <b>Inh 1</b>	71,6	14,732 ± 0,019	67,8 ± 0,7
holo-A-PCP + <b>Inh 1</b>	72,0	14,776 ± 0,017	66,3 ± 0,6
apo-A-PCP + <b>Inh 2</b>	71,6	14,567 ± 0,013	73,9 ± 0,5
holo-A-PCP + <b>Inh 2</b>	71,9	14,551 ± 0,015	74,5 ± 0,5

\* Die Säule wurde mit Hilfe des „Gel Filtration Calibration Kit - Low Molecular Weight“ von GE Healthcare kalibriert; in der zweiten Spalte sind die theoretischen Molekulargewichte der Komplexe aufgelistet.

In der nativen PAGE zeigte das holo-Protein im Vergleich mit dem apo-Protein eine etwas größere elektrophoretische Mobilität (s. Abbildung 31 linke Spalte). Dies könnte das Resultat von unterschiedlichen Konformationen der beiden Proteine sein. Jedoch erhält das A-PCP

Protein durch die Addition des Ppant-Arms eine zusätzliche negative Ladung (die berechnete Ladung des apo-Proteins unter den Bedingungen der nativen PAGE ist  $\sim -20$ ). Berücksichtigt man dies zusammen mit den Daten der Gelfiltrationsläufe so ist es wahrscheinlicher, dass die Anordnung der Domänen in beiden Proteinen (apo und holo) vergleichbar ist. Die Inkubation der Proteine mit dem Inhibitor der A-Domäne, **Inh 1**, veränderte sowohl die Retentionszeiten der Gelfiltrationsläufe, als auch das Verhalten in der nativen PAGE. In Übereinstimmung mit der Annahme, dass eine Bindung des Inhibitors eine kompaktere Konformation induziert, eluierten sowohl das apo-, als auch das holo-A-PCP Protein mit signifikant größeren Retentionszeiten in der Gelfiltrationschromatographie (s. Tabelle 4). In der nativen PAGE konnte in beiden Fällen eine größere elektrophoretische Mobilität im Vergleich zu den Proteinen ohne Inhibitoren festgestellt werden (vergleiche in Abbildung 31 die linke und mittlere Spalte).

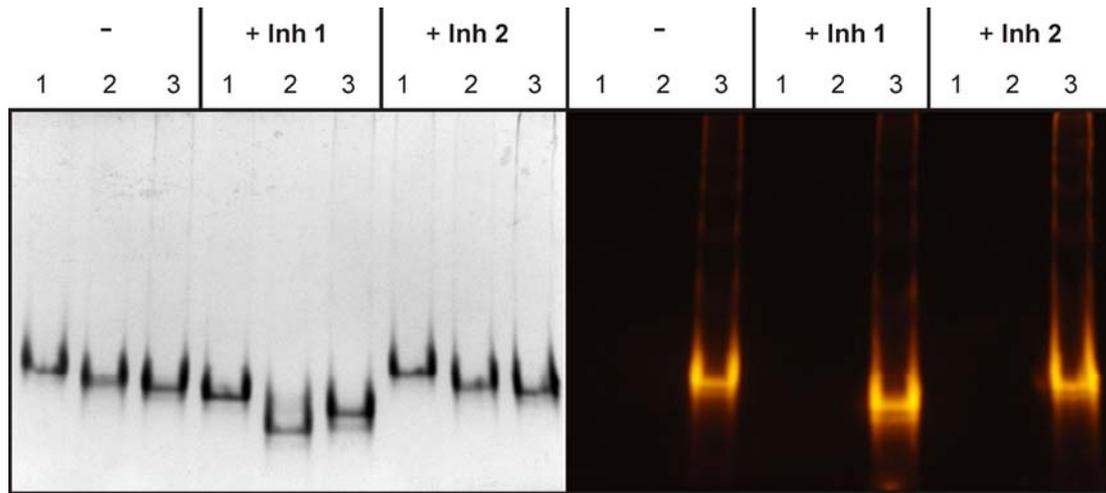


**Abbildung 31 Native PAGE der GrsA A-PCP Proteine in Abwesenheit und Gegenwart von Substraten.** Sowohl die apo-, als auch die holo-Form wurden ohne oder mit den Inhibitoren (jeweils 100  $\mu\text{M}$ ) inkubiert und anschließend der nativen PAGE unterzogen; nur die kognate Verbindung, 5'-O-[N-(L-phenylalaninyl)-sulfamoyl] Adenosin (**Inh 1**), und nicht 5'-O-[N-(L-prolyl)-sulfamoyl] Adenosin (**Inh 2**) verändert die elektrophoretische Mobilität der A-PCP Proteine.

Die Inkubation mit **Inh 2** zeigte als Kontrollexperiment, dass dieser Inhibitor keinen Einfluss auf die Konformation der A-PCP Proteine besitzt (s. Tabelle 4 und Abbildung 31 rechte Spalte). Auch die Inkubation mit ATP, L-Phe, AMP, ApCpp und Kombinationen der Adenosinverbindungen mit L-Phe ergaben keine veränderten elektrophoretischen Mobilitäten in der nativen PAGE (Daten nicht gezeigt). Nur die Bindung von **Inh 1** an das apo- und das holo-Protein war stark genug, um den Elektrophoreseprozess zu überstehen. Die kompaktere Konformation, induziert durch **Inh 1**, resultiert wahrscheinlich zu einem großen Teil aus dem Schließen der beiden Subdomänen  $A^N$  und  $A^C$  bei Bindung des Inhibitors. Der Übergang der A-Domäne von einer offenen in eine geschlossene Konformation sollte die effektive Größe des Proteins verkleinern (s. Abbildung 9). Ein detaillierter Vergleich zwischen dem apo- und

dem holo-Protein in diesen zwei Assays deuten jedoch daraufhin, dass sich zusätzlich die relative Anordnung der A- und der PCP-Domäne verändert. In der native PAGE führte die Inkubation des holo-Proteins mit **Inh 1** zu einer größeren relativen Steigerung der elektrophoretischen Mobilität im Vergleich zum apo-Protein (vergleiche Abbildung 31). Dieser Trend konnte in den Gelfiltrationsexperimenten bestätigt werden. Hier war das Elutionsvolumen des holo-Proteins mit **Inh 1** statistisch signifikant größer als das Elutionsvolumen des apo-Proteins mit **Inh 1** (t-Test der Mittelwerte der Elutionsvolumina: die Nullhypothese  $\text{apo} \geq \text{holo}$  wird mit einem Signifikanzniveau von  $\alpha = 0,05$  zugunsten der einseitigen Alternativhypothese  $\text{apo} < \text{holo}$  verworfen). Für einen Vergleich der berechneten Elutionsdifferenzen (apo-A-PCP  $0,172 \pm 0,034$  mL und holo-A-PCP  $0,209 \pm 0,029$  mL) konnte allerdings keine statistische Signifikanz erreicht werden. Die Daten erlauben jedoch trotzdem den Schluss, dass die Bindung von **Inh 1** an holo-A-PCP eine zusätzliche konformationelle Änderung zu einer kompakteren Konformation auslöst. Die Bindung der holo-PCP-Domäne an die A-Domäne erscheint als die nahe liegende Erklärung für diese Verkleinerung. In Übereinstimmung mit der Logik der NRPS Biosynthese ist nicht die apo-Form, sondern nur die Ppant-PCP-Domäne ein Substrat für die A-Domäne. Dies sind die ersten biochemischen Indizien, dass die Wechselwirkungen zwischen einer A- und einer sich *in cis* befindlichen PCP-Domäne von dem post-translationalen Zustand der letzteren abhängig sind.

In den vorangegangenen Experimenten wurde durch die Bindung des Inhibitors an die A-Domäne die Situation vor der Thioesterbildung simuliert. Eine Untersuchung des nächsten Reaktionsschrittes, die mit einem Thioester beladene PCP-Domäne, mittels nativer PAGE war jedoch erschwert. Die Inkubation des holo-Proteins mit ATP und L-Phe resultierte in keiner veränderten elektrophoretischen Mobilität des Proteins, wahrscheinlich weil die Stabilität des Thioesters unter den Assaybedingungen nicht gewährleistet war (z.B.  $\beta$ -Mercaptoethanol im Probenpuffer). Eine stabile, wenn auch nicht natürliche Modifikation des Ppant-Arms gelang durch Sfp katalysierte Umsetzung von apo-A-PCP mit CoA-TAMRA (s. Abbildung 29A). Das mit dem sperrigen Fluorophor Tetramethylrhodamin modifizierte Protein zeigte in den nativen PAGE Untersuchungen eine analoge Steigerung der elektrophoretischen Mobilität bei Zugabe des **Inh 1** wie das apo-Protein (s. Abbildung 32). Auch wenn diese nicht natürliche Situation keine eindeutigen Aussagen zulässt, kann man dieses Ergebnis als Hinweis interpretieren, dass eine beladene (modifizierte) PCP-Domäne bei Bindung des **Inh 1** im Vergleich zum holo-Protein nicht die gleiche Affinität zur A-Domäne besitzt.



**Abbildung 32 Native PAGE von A-PCP-TAMRA im Vergleich mit dem apo- und dem holo-Protein.** In den Spuren 1 ist das apo-Protein, in den Spuren 2 das holo-Protein und in den Spuren 3 das mit Tetramethylrhodamin beladene A-PCP Protein aufgetragen; A-PCP-TAMRA zeigt keine so ausgeprägte Veränderung der elektrophoretischen Mobilität bei Zugabe von **Inh 1** wie das holo-Protein; links ist das Gel in der Coomassie-Färbung, rechts dasselbe Gel unter UV-Bestrahlung zu sehen.

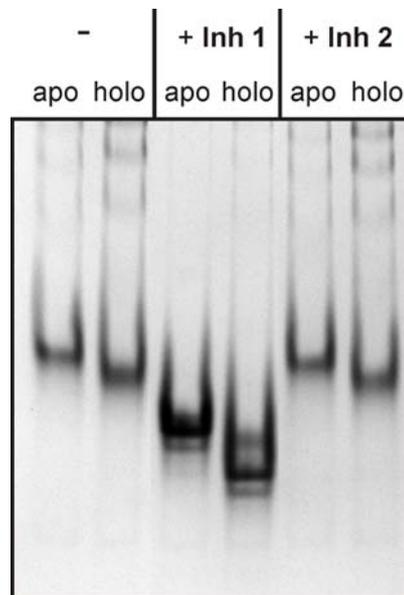
Zusammenfassend ermöglichen die in der vorliegenden Arbeit benutzten Gelfiltrations- und native PAGE-Assays die Untersuchung der globalen Konformationen von NRPS A-PCP Proteinen. Dies wurde auch durch die Analyse eines A-PCP Fragments aus einem Prolin aktivierenden Modul (Tyrocidin Synthetase II Modul 1) bestätigt (s. Abbildung 72A im Anhang). Hier führte nur die Inkubation mit 5'-*O*-[*N*-(L-prolyl)-sulfamoyl] Adenosin (**Inh 2**) und nicht die Inkubation mit **Inh 1** zu einer veränderten elektrophoretischen Mobilität des Proteins.

#### 4.1.3 Untersuchung der Position des Ppant-Arms in A-PCP Proteinen

Die Gelfiltrations- und nativen PAGE-Assays lassen die Interpretation zu, dass bei Bindung von **Inh 1** an holo-GrsA A-PCP das Protein eine hoch kompakte Konformation einnimmt, in der sich die PCP-Domäne in engem Kontakt zur A-Domäne befindet. Verdeutlicht man sich den NRPS Katalysezyklus (s. Abbildung 27), so müsste in dieser Konformation der Ppant-Arm der PCP-Domäne in die aktive Tasche der A-Domäne zeigen. Im Vergleich mit dem Aminoacyladenylat ist jedoch mit **Inh 1** keine Weiterreaktion möglich und das Protein wird in dieser Position fixiert. Dieses Modell postuliert damit eine Abschirmung des Ppant-Arms in dieser Konformation und kann mittels Zugänglichkeitsstudien des Ppant-Arms überprüft werden. In nachfolgenden Experimenten sollten thiol-modifizierende Reagenzien die Sulfhydrylgruppe des Phosphopantetheins

markieren und im Falle der hoch kompakten Konformation von holo-A-PCP eine langsamere Kinetik zeigen.

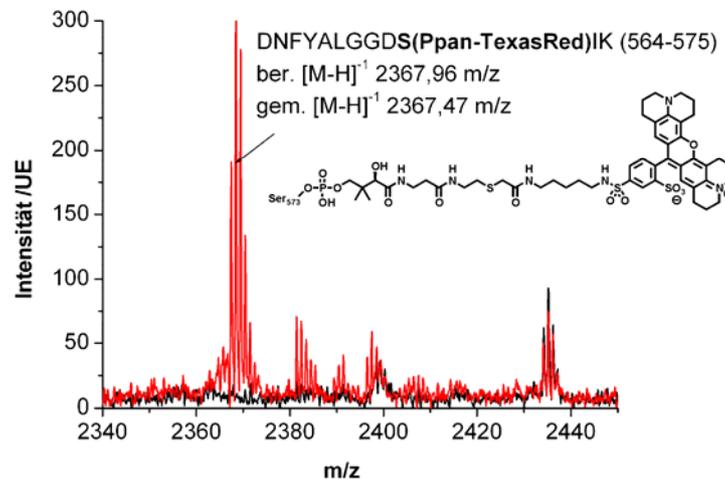
Um Kreuzreaktionen mit anderen Thiofunktionalitäten zu vermeiden, wurde in diesen Experimenten eine cystein-freie Mutante von GrsA A-PCP verwendet. Das Protein mit den Mutationen C60F, C331A, C376S und C473A (A-PCP( $\Delta$ -4Cys)) wurde bereits in einer früheren Arbeit hergestellt und biochemisch anhand des ATP/PP<sub>i</sub>-Austauschs charakterisiert [211]. Die vier eingeführten Mutationen in der A-Domäne hatten nur einen geringen Einfluss auf die Aktivität. So vergrößerte sich der K<sub>M</sub>-Wert in Bezug auf L-Phe um den Faktor 2,5 während k<sub>cat</sub> um den Faktor 1,5 verringert wurde [211]. Da sich A-PCP( $\Delta$ -4Cys) zusätzlich in dem nativen PAGE-Assay vergleichbar wie das Wildtyp-Protein verhielt (s. Abbildung 33), konnte es für die folgenden Untersuchungen eingesetzt werden.



**Abbildung 33 Native PAGE der GrsA A-PCP( $\Delta$ -4Cys) Proteine in Abwesenheit und Gegenwart von Substraten.** Sowohl die apo-, als auch die holo-Form wurden ohne oder mit den Inhibitoren (jeweils 100  $\mu$ M) inkubiert und anschließend der native PAGE unterzogen.

Die einzige Thiofunktionalität in holo-A-PCP( $\Delta$ -4Cys) befindet sich am Ppant-Arm. Bei einer Inkubation dieses Proteins mit Thiol-modifizierenden Reagenzien und der Bestimmung der zeitabhängigen Markierung können damit Aussagen über die Zugänglichkeit des Ppant-Arms zum Lösungsmittel getroffen werden. Der Vorteil dieses Systems ist, dass eine Vorinkubation des Proteins mit Substraten und Inhibitoren leicht durchführbar ist. Die Behandlung von holo-A-PCP( $\Delta$ -4Cys) mit Fluorescein-Maleinimid oder mit Fluorescein-5-Iodacetamid führte zu einer schnellen, vollständigen Markierung des Ppant-Arms, ungeachtet ob der **Inh 1** zugegen war oder nicht (Daten nicht gezeigt). Erst der Einsatz des weniger reaktiven, sterisch etwas anspruchsvolleren TexasRed-Bromacetamids ermöglichte die

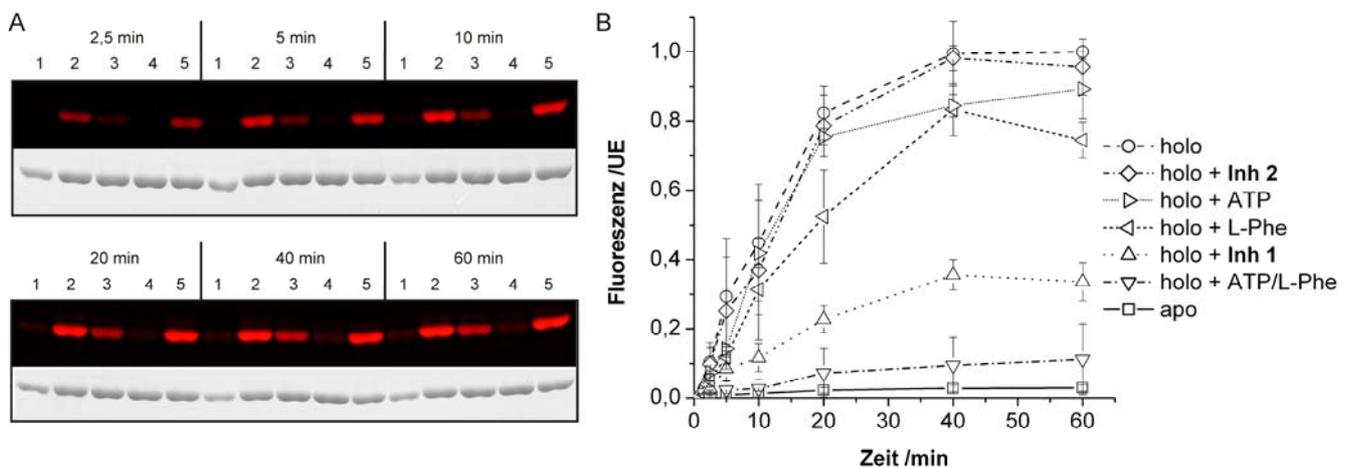
Bestimmung von unterschiedlichen Markierungskinetiken. Dass die Modifikation des Proteins mit dem Fluoreszenzfarbstoff tatsächlich am Ppant-Arm stattfand, konnte durch einen tryptischen in-Gelverdau und der Detektion der Masse des berechneten Fragments mittels MALDI-TOF MS nachgewiesen werden (s. Abbildung 34).



**Abbildung 34** Markierung des Ppant-Arms von holo-A-PCP( $\Delta$ -4Cys) mit TexasRed-Bromacetamid. Ausschnitt aus den MALDI-TOF MS Spektren der tryptischen in-Gelverdau vor (schwarz) und nach der Modifikationsreaktion mit TexasRed-Bromacetamid (rot); das modifizierte Ppant-TexasRed Fragment konnte nur nach der Reaktion nachgewiesen werden.

Der Umfang der Modifikation wurde anhand einer densitometrischen Auswertung des Fluoreszenzsignals im SDS-PAGE Gel (normiert auf den Gesamtproteingehalt der Bande in der Coomassie-Färbung) bestimmt. In Abbildung 35A ist beispielhaft ein SDS-PAGE Gel der Modifikationsreaktion von A-PCP( $\Delta$ -4Cys) mit TexasRed-Bromacetamid unter verschiedenen Inkubationsbedingungen gezeigt. Die detaillierte Auswertung und die graphische Auftragung des Markierungsgrads über die Zeit ist in Abbildung 35B zu sehen. Für holo-A-PCP( $\Delta$ -4Cys) ohne Liganden ergab sich die schnellste Markierungsreaktion. Der Wert dieser Reaktion nach einer Stunde wurde als relative Referenz verwendet. Die Inkubation des apo-A-PCP( $\Delta$ -4Cys) Proteins unter identischen Bedingungen zeigte wie erwartet nur eine minimale Hintergrundreaktivität, da keine Thiofunktionalität in diesem Protein vorhanden war (s. Spur 1 in Abbildung 35 A). Im Rückschluss bedeutete dies, dass der Einbau des Fluorophors in holo-A-PCP( $\Delta$ -4Cys) fast vollständig auf eine Markierung des Ppant-Arms zurückgeht. Wurde holo-A-PCP( $\Delta$ -4Cys) vor der Markierungsreaktion mit den Substraten ATP und L-Phe inkubiert, konnte nur ein sehr geringer Einbau des Fluorophors nachgewiesen werden. Unter diesen Bedingungen blockierte die Bildung des L-Phe-Ppant Thioesters die Sulfhydrylgruppe des Ppant-Arms und verhinderte die Markierung (s. Spur 4 in Abbildung 35 A). ATP und L-Phe zeigten alleine nur eine geringe Reduktion der Markierungsgeschwindigkeit, wobei die

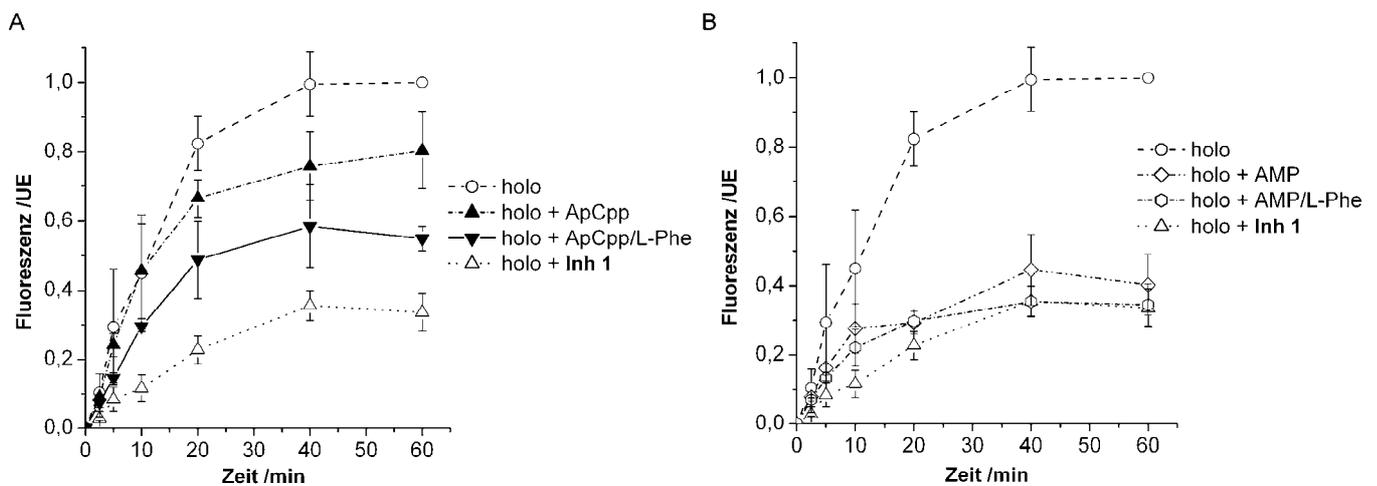
Aminosäure, in Übereinstimmung mit den Daten der partiellen Proteolyse, einen etwas stärkeren Effekt ausübte. In einer weiteren Kontrollreaktion führte die Zugabe des nicht kognaten Prolininhibitors **Inh 2** zu einem vergleichbaren Markierungsverlauf wie die Inkubation ohne Liganden (s. Spuren 2 und 5 in Abbildung 35 A). Im Gegensatz dazu verminderte die Präsenz des **Inh 1** die Zugänglichkeit der Ppant-Gruppe signifikant (s. Spur 3 in Abbildung 35 A). Bei Bindung von **Inh 1** an holo-A-PCP( $\Delta$ -4Cys) nimmt das Protein - in Übereinstimmung mit dem oben beschriebenen Modell - eine Konformation ein, in der der Ppant-Arm der PCP-Domäne effektiv vom Lösungsmittel abgeschirmt ist.



**Abbildung 35 Chemische Markierung von apo- und holo-A-PCP( $\Delta$ -4Cys) mit TexasRed-Bromacetamid.** A) Repräsentatives SDS-PAGE Gel der Markierungsreaktion zu verschiedenen Zeitpunkten unter UV-Bestrahlung (oben) und in der Coomassie-Färbung (unten); Spur 1 = apo-A-PCP( $\Delta$ -4Cys), Spur 2 = holo-A-PCP( $\Delta$ -4Cys), Spur 3 = holo-A-PCP( $\Delta$ -4Cys) + 100  $\mu$ M **Inh 1**, Spur 4 = holo-A-PCP( $\Delta$ -4Cys) + 7,5 mM ATP/L-Phe, Spur 5 = holo-A-PCP( $\Delta$ -4Cys) + 100  $\mu$ M **Inh 2**; B) Densitometrische Auswertung des Fluoreszenzsignals im SDS-PAGE Gel nach der Normierung auf den Gesamtproteingehalt der Bande in der Coomassie-Färbung.

In weiteren Experimenten dieser Arbeit wurde der chemische Markierungsassay eingesetzt, um den Einfluss von ATP-Analoga auf die Position des Ppant-Arms zu untersuchen. Nach der biosynthetischen Logik sollte das nicht hydrolysierbare Adenosin-5'-[( $\alpha,\beta$ )-methylen]-triphosphat (ApCpp) zusammen mit L-Phe den Zustand in der A-Domäne vor der Adenylierungsreaktion simulieren. Die Methylenverbrückung zwischen der  $\alpha$ - und der  $\beta$ -Phosphatgruppe erlaubt keine Abspaltung der Diphosphateinheit und damit keine Bildung des Aminoacyladenylats. Übereinstimmend damit wurde eine humane Acyl-CoA Synthetase ACSM2A, ein Mitglied der ANL-Superfamilie, im Komplex mit ApCpp in der Adenylierungskonformation kristallisiert [58]. Die Autoren weisen daraufhin, dass die  $\beta$ - und  $\gamma$ -Phosphatgruppen sowohl in ApCpp, als auch in ATP eine konformationelle Änderung des ANL-Proteins in die Thioesterkonformation verhindern. In Abbildung 36A ist der Einfluss von ApCpp auf die Zugänglichkeit des Ppant-Arms gezeigt. In einer Konzentration von 1 mM

hat es nur eine geringe Verminderung der Markierungsgeschwindigkeit des Ppant-Arms, vergleichbar mit dem Effekt von L-Phe, zur Folge. Die gemeinsame Inkubation mit L-Phe verstärkt diesen Einfluss etwas, jedoch ist die Ppant-Thiolgruppe nicht so effektiv abgeschirmt, wie im Fall der Inkubation des Proteins mit 100  $\mu\text{M}$  **Inh 1**. Dies steht im Einklang mit den Ergebnissen der partiellen Proteolyse, in der diese Ligandenkombination (ApCpp/L-Phe) einen geringeren Effekt im Vergleich zu sowohl ATP und L-Phe, als auch zu **Inh 1** zeigte. Unter der Annahme, dass eine Abschirmung des Ppant-Arms mit einer Positionierung des Ppant-Arms im aktiven Zentrum der A-Domäne gleichzusetzen ist, ist folgende Interpretation möglich. Nur in der Thioesterkonformation der A-Domäne (induziert durch **Inh 1**) und nicht in der Adenylierungskonformation (ausgelöst durch ApCpp/L-Phe) kann die holo-PCP-Domäne mit hoher Affinität an die A-Domäne binden und den Ppant-Arm im aktiven Zentrum positionieren. Kristallographische Studien unterstützen diese Hypothese, da nur in der Thioesterkonformation der A-Domäne ein Ppant-Bindungstunnel vorhanden ist (s. Kapitel 1.1.4).



**Abbildung 36 Chemische Markierung von holo-A-PCP( $\Delta$ -4Cys) in Gegenwart von ATP-Analoga.** A) Vor der Reaktion mit TexasRed-Bromacetamid (TR-BA) wurde das Protein mit 1 mM Adenosin-5'-[( $\alpha,\beta$ )-methylen]-triphosphat (ApCpp) alleine und mit der Kombination 1 mM ApCpp und L-Phe inkubiert; B) Inkubation mit 1 mM AMP und jeweils 1 mM AMP und L-Phe; zum Vergleich ist noch einmal die Reaktion in Gegenwart von 100  $\mu\text{M}$  **Inh 1** gezeigt.

Welche Konformation die Bindung des Reaktionsprodukts AMP in der A-Domäne auslöst, ist bisher in der Literatur nicht abschließend geklärt. Die bereits oben erwähnten Kristallisationsstudien an der humanen Acyl-CoA Synthetase ACSM2A [58] zeigten das Protein im Komplex mit AMP in der Thioesterkonformation, jedoch wurden mit diesem Liganden andere ANL-Mitglieder auch in der Adenylierungskonformation kristallisiert, z.B. [56]. Frühere Studien bestimmten eine ca. 20fach größere Affinität von AMP im Vergleich zu

ATP ( $K_D$ -Werte: 0,064 mM (AMP) und 1,3 mM (ATP) [221]. Bei den in dieser Arbeit durchgeführten chemischen Markierungsassays zeigte AMP (Inkubation mit einer Konzentration von 1 mM) sowohl alleine, als auch zusammen mit L-Phe eine gravierende Reduktion der Markierungsgeschwindigkeit, vergleichbar mit dem Einfluss von 100  $\mu$ M **Inh 1** (s. Abbildung 36 B). Dies spricht dafür, dass AMP die Thioesterkonformation in der A-Domäne auslöst und damit die Bindung von holo-PCP ermöglicht. Im Gegensatz dazu zeigten jedoch weder AMP alleine noch die Kombination aus AMP/L-Phe in der partiellen Proteolyse in Kapitel 4.1.1. einen größeren, nennenswerten Effekt, so dass anhand der hier präsentierten Ergebnisse keine definitive Aussage zu dem Bindungsmodus dieses Liganden getroffen werden kann.

Zwei Einwände können gegen die bis hierhin präsentierten Daten der chemischen Markierungsexperimente vorgebracht werden. Zum einen ist bekannt, dass aromatische Fluorophore, wie das im Vergleich zu TexasRed kleinere Fluorescein, eine natürliche Affinität zur ATP Bindetasche besitzen [225]. Fände die Modifikation des Ppant-Thiols im aktiven Zentrum statt, könnte sich der Einfluss des **Inh 1** auf eine kompetitive Bindung an das Protein und eine Verdrängung des Markierungsreagenz beschränken. Aussagen über konformationelle Änderungen wären damit erschwert. Zur Kontrolle wurde holo-A-PCP( $\Delta$ -4Cys) in Abwesenheit von Liganden, aber in Gegenwart unterschiedlicher Mengen von freiem TexasRed Farbstoff mit TexasRed-Bromacetamid umgesetzt. In dem oben beschriebenen Szenario würde die Zugabe des freien Farbstoffs zu einer Verringerung der Markierungsgeschwindigkeit führen, da er mit dem Markierungsreagenz um die Bindestelle am Protein konkurrieren würde. Jedoch hatte die Zugabe von bis zu 16  $\mu$ M TexasRed (in Bezug auf TexasRed-Bromacetamid) keinen Einfluss auf die Alkylierungsreaktion (s. Abbildung 37). Damit konnte diese alternative Erklärung ausgeschlossen und dieser Einwand widerlegt werden.

Als zweiten Einwand könnte eine alternative Deutung der Ergebnisse der chemischen Markierungsexperimente vorgebracht werden. Es lässt sich nicht vollständig ausschließen, dass die Unterschiede in der Markierungseffizienz nicht durch große konformationelle Änderungen, wie die Bindung der holo-PCP-Domäne an die A-Domäne, ausgelöst werden. Auch das Öffnen und Schließen von Proteinkanälen zum aktiven Zentrum würde die gesammelten Daten erklären. Betrachtet man jedoch den sterischen Anspruch des Markierungsreagenz und zusätzlich die Ergebnisse der nativen PAGE- und Gelfiltrationsassays so erscheint eine große Konformationsänderung mit einer Bewegung der PCP-Domäne wahrscheinlicher.



**Abbildung 37 Kontrolle der chemischen Markierungsreaktion.** A) Chemische Strukturen von TexasRed-Bromacetamid (TR-BA, links) und dem freien TexasRed Farbstoff (TR, rechts); B) Markierung von holo-A-PCP( $\Delta$ -4Cys) mit TR-BA in Gegenwart von unterschiedlichen Konzentrationen an freiem TR Farbstoff; gezeigt sind die Proben zu verschiedenen Zeitpunkten auf einem SDS-PAGE Gel unter der UV-Beleuchtung; Spur 1 = 1 mM TR; Spur 2 = 60  $\mu$ M TR-BA; Spur 3 = 60  $\mu$ M TR-BA + 60  $\mu$ M TR (1:1); Spur 4 = 60  $\mu$ M TR-BA + 600  $\mu$ M TR (1:10); Spur 5 = 60  $\mu$ M TR-BA + 1 mM TR (1:16).

Die in diesem Kapitel präsentierten Daten führen zu dem Schluss, dass sich der Ppant-Arm in holo-A-PCP in mindestens zwei unterschiedlichen Positionen während des Katalysezyklus befindet, die sich stark in ihrer Zugänglichkeit zum Lösungsmittel unterscheiden. Zusätzlich ist der Übergang von der zugänglichen in die abgeschirmte Position - wahrscheinlich zeigt hier der Ppant-Arm in das aktive Zentrum der A-Domäne - von dem Zustand der A-Domäne abhängig. Eine korrekte Wechselwirkung zwischen der A- und der PCP-Domäne benötigt dabei zum einen die Thioesterkonformation der A-Domäne (induziert durch **Inh 1**) und zum anderen die unbeladene holo-Form der PCP-Domäne. Befindet sich entweder die A-Domäne in der Adenylierungskonformation (induziert durch ApCpp und L-Phe) oder hat die PCP-Domäne nicht den richtigen post-translationalen Zustand (apo-Form oder mit TAMRA-beladener Ppant-Arm) so scheint nicht die richtige Kontaktfläche zwischen den beiden Domänen vorhanden zu sein.

Es bleibt anzumerken, dass die Auflösung der verwendeten biochemischen Charakterisierungen nicht ausreicht, um zu definitiven Aussagen über die hier untersuchte Domäneninteraktion zu gelangen. In wie weit die an freistehenden PCP-Domänen gefundenen Konformationen [87] auch in dem Didomänenprotein eine Rolle spielen kann nicht geklärt werden. Mit den hier präsentierten Assays ist weder eine strukturelle Beschreibung des Proteins mit atomarer Auflösung, noch eine Charakterisierung der kinetischen Parameter der konformationellen Änderungen möglich. Höher auflösende Techniken, wie die Einzelmolekül-Fluoreszenz- oder die NMR-Spektroskopie, müssen in Zukunft zur Aufklärung eingesetzt werden. Wie bereits in der Einleitung beschrieben, setzen diese Techniken jedoch

den definierten Einbau biophysikalischer Sonden in das zu untersuchende Protein voraus. Bevor auf Vorarbeiten zur spektroskopischen Untersuchung von Multi-Domänen NRPS-Proteinen, nämlich der Herstellung selektiv modifizierter Proteine, eingegangen wird (Kapitel 4.4), sollen zuerst die im Laufe dieser Arbeit durchgeführten Studien zu gespaltenen Inteinen (Kapitel 4.2 und 4.3) vorgestellt werden. Denn diese besitzen das Potential einen entscheidenden Beitrag zur Herstellung der modifizierten Proteine zu leisten (siehe vor allem Kapitel 1.2.4 in der Einleitung).

## 4.2 Biochemische und strukturelle Untersuchungen zu gespaltenen Inteinen

Die Protein *trans*-Spleißen (PTS)-Reaktion mit Hilfe von gespaltenen Inteinen ist bereits ein wichtiges Werkzeug in zahlreichen biochemischen Anwendungen. Unter anderem wird diese Ligrationsreaktion bei der Protein-Semisynthese, bei der selektiven Modifikation von Proteinen und bei der segmentellen Isotopenmarkierung benutzt (s. Kapitel 1.2.4). Der erfolgreiche Einsatz von gespaltenen Inteinen in diesen Applikationen setzt jedoch vier wichtige Eigenschaften der Inteinsysteme voraus. (1) Eine hohe Ausbeute der PTS-Reaktion mit möglichst geringer Nebenproduktbildung ist essentiell. Für *in vitro* Reaktionen wurden in einigen Fällen bereits Ausbeuten von 70 - 100% beschrieben [145,171]. Jedoch scheinen Umsätze zwischen 30% und 60% bei gespaltenen Inteinen die Regel zu sein [163,165,177,187]. (2) Eine hohe Reaktionsgeschwindigkeit ist für alle auf gespaltenen Inteinen basierenden Anwendungen, besonders für *in vivo* Applikationen, von Vorteil. Die Halbwertszeiten der Reaktionen liegen bisher im Bereich von einigen Minuten bis Stunden (s. Tabelle 1). (3) Ein effektives Inteinsystem sollte eine hohe Toleranz gegenüber seinen Substraten, also den N- und C-Exteinen, besitzen. Bei eingeschränkter Toleranz wäre es wichtig, schnell entscheiden zu können, ob das gespaltene Intein in einem bestimmten Exteinkontext aktiv ist. (4) Für *in vitro* Anwendungen ist die Löslichkeit der Intein-Extein-Fusionsproteine natürlich von großer Wichtigkeit. Generell ist zu erwarten, dass in diesem Punkt natürlich gespaltene Inteine einen Vorteil gegenüber den künstlich gespaltenen besitzen.

Aktuell ist noch kein Inteinsystem bekannt, dass in allen oben angesprochenen Punkten herausragende Eigenschaften besitzt. Die Suche nach besseren gespaltenen Inteinen und die Optimierung bereits vorhandener Systeme sind deshalb notwendig, um die Einsatzmöglichkeiten des Protein *trans*-Spleißens weiter zu verbessern.

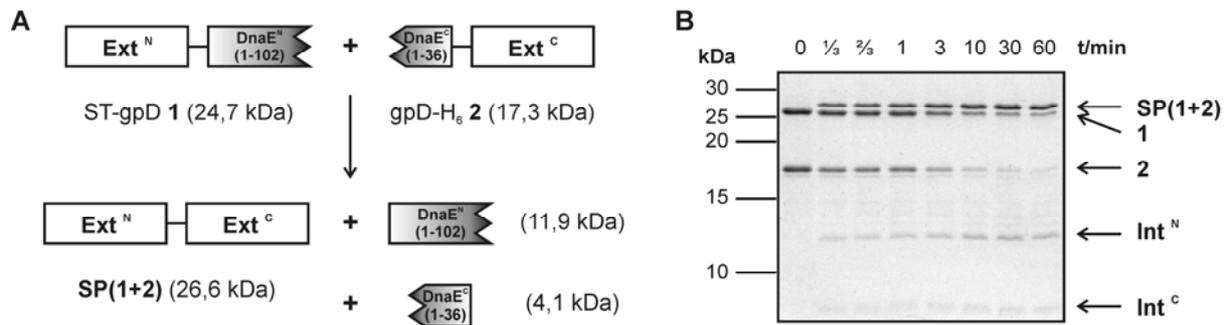
### 4.2.1 Biochemische Charakterisierung des natürlich gespaltenen *Npu* DnaE Inteins

Unter den natürlich vorkommenden DnaE-Inteinallelen aus diversen Cyanobakterien ist das DnaE Intein aus *Synechocystis* species PCC6803 (*Ssp*) das am besten charakterisierte. Die beiden Hälften des *Ssp* DnaE Inteins assoziieren sehr schnell nahe dem Diffusionslimit der Einstein-Smoluchowski Gleichung und zeigen zudem eine sehr geringe

Dissoziationskonstante, die im nanomolaren Bereich liegt [174]. In einem nicht-natürlichen Exteinkontext wurden für dieses Intein Ausbeuten von 50 - 80% und je nach Studie Halbwertszeiten der Spleißreaktion von 35 - 175 min bestimmt [151,173,177]. Eine Veröffentlichung von Iwai *et al.* [151] im Jahre 2006 lieferte jedoch Hinweise für die Annahme, dass das DnaE Intein aus *Nostoc punctiforme* (*Npu*) im Vergleich zum *Ssp* DnaE Intein in diesen Punkten überlegene Eigenschaften besitzt. Diese Arbeit belegte die *in vivo* Spleißaktivität des *Npu* DnaE Inteinallels durch Coexpression in *E. coli*. Die Spleißreaktion mit nicht-natürlichen Exteinresten verlief mit einer Ausbeute von über 98% deutlich effizienter als für das *Ssp* DnaE Intein. Zusätzlich zeigte die N-terminale *Npu* DnaE Inteinhälfte bei Kreuz-Spleißreaktionen mit der C-terminalen Hälfte des *Ssp* DnaE Inteins eine größere Toleranz gegenüber der +2 Position im Vergleich zur entsprechenden Hälfte des *Ssp* DnaE Inteins [151]. Die Frage war nun, ob diese vorteilhaften *in vivo* Spleißereigenschaften des *Npu* DnaE Inteins auch beim *in vitro* Spleißen mit gereinigten Proteinen auftreten würden. Dazu wurde das *Npu* DnaE Intein in Zusammenarbeit mit Vivien Schütz biochemisch charakterisiert [213].

### **Kinetische Charakterisierung des *Npu* DnaE Inteins**

Aufgrund ihrer hervorragenden Löslichkeitseigenschaften wurden das Kopffprotein des Bakteriophagen  $\lambda$  (gpD), Thioredoxin (Trx) und das verbesserte grün-fluoreszierende Protein (eGFP) als Modell-Exteine gewählt. Zur Reinigung der Proteine wurde bei den N-terminalen Inteininfusionsproteinen eine StrepTag II (ST)- und bei den C-terminalen eine Hexahistidin-Sequenz (H<sub>6</sub>) eingesetzt. An der N-terminalen Spleißstelle waren die nicht-nativen Aminosäuren GS (-2, -1 Position) vorhanden. In früheren Arbeiten hatte diese Substitution der natürlichen EY-Resten keinen negativen Einfluss auf die Aktivität sowohl des *Ssp*, als auch des *Npu* DnaE Inteins gezeigt [151,188]. Die C-terminale Spleißstelle enthielt die native Sequenz CFN. Alle Inteininfusionsproteine konnten in löslicher Form durch Expression in *E. coli* hergestellt werden. Allerdings neigten die C-terminalen Inteinstrukturen zur Degradation während der Zelllyse [212,213]. Der Aufschluss der Zellen in Gegenwart von 8M Harnstoff konnte den Abbau der C-terminalen Proteine vermindern und erlaubte die Isolierung der Inteininfusionsproteine nach Rückfaltung in Spleißpuffer mit akzeptabler Reinheit. Zum Testen der Spleißaktivität wurden die N- und C-terminalen Inteinhälften in äquimolaren Konzentrationen von 10 - 20  $\mu$ M umgesetzt. Am Beispiel der Proteine ST-gpD-Int<sup>N</sup> (**1**) und Int<sup>C</sup>-gpD-H<sub>6</sub> (**2**) ist die PTS-Reaktion des *Npu* DnaE Inteins in Abbildung 38 gezeigt.



**Abbildung 38 Protein *trans*-Spleißen mit dem *Npu* DnaE Intein.** A) Schematische Darstellung der PTS-Reaktion mit dem Kopfprotein des Bakteriophagen  $\lambda$  (gpD) als N- und C-Extein; B) SDS-PAGE Gel der Reaktion bei 20°C zwischen 1 und 2 zu verschiedenen Zeitpunkten; das Spleißprodukt SP(1+2), sowie die Inteinhälfte Int<sup>N</sup> und Int<sup>C</sup> sind zugeordnet.

Die drei im Laufe der Reaktion entstandenen Proteinbanden konnten mittels tryptischem Verdau der ausgeschnittenen Gelbanden und anschließender MALDI-TOF MS Analyse eindeutig dem Spleißprodukt ST-gpD-gpD-H<sub>6</sub> (SP1+2) und den beiden Inteinhälften Int<sup>N</sup> und Int<sup>C</sup> zugeordnet werden (Daten nicht gezeigt). Zusätzlich konnten bei den weiteren Exteinkombinationen die jeweiligen Affinitätstags und damit die Identität aller Proteine durch Western Blot Analyse bestätigt werden [213]. Die Bildung des Spleißproduktes fand im beschriebenen Beispiel ebenso wie bei den weiteren Kombinationen mit einer sehr hohen Geschwindigkeit statt. Unter der Annahme einer Reaktionskinetik erster Ordnung, wobei die Komplexbildung nicht den geschwindigkeitsbestimmenden Schritt der Reaktion darstellt [174], wurde die Geschwindigkeitskonstante der Reaktion über eine densitometrische Analyse der Proteinbandenintensität bestimmt (s. Kapitel 3.6). Für die Reaktion der Proteine 1 und 2 wurde bei 20°C ein Wert von  $k_{\text{Spleißen}, 20^\circ\text{C}} = 3,6 \pm 0,5 \times 10^{-3} \text{ s}^{-1}$  ermittelt. Eine Temperaturerhöhung auf 37°C führte zu einer Vergrößerung der Reaktionsgeschwindigkeit auf  $k_{\text{Spleißen}, 37^\circ\text{C}} = 9,0 \pm 0,4 \times 10^{-3} \text{ s}^{-1}$ .

In Zusammenarbeit mit Vivien Schütz wurden die Reaktionskonstanten der PTS-Reaktion mit weiteren Exteinkombinationen bei verschiedenen Temperaturen bestimmt (s. Tabelle 5). Im Temperaturbereich von 6 - 37°C zeigte das *Npu* DnaE Intein eine robuste Spleißaktivität. Die Reaktionsrate war nur in geringem Maße von den zu ligierenden Exteinsequenzen abhängig. Im Vergleich mit der Reaktion zwischen 1 und 2 führte die Reaktion mit Thioredoxin als C-Extein zu einer etwas höheren Reaktionsgeschwindigkeit mit  $k_{\text{Spleißen}, 37^\circ\text{C}} = 1,1 \pm 0,2 \times 10^{-2} \text{ s}^{-1}$ . Die daraus resultierende Halbwertszeit der Reaktion von 63 s stellt den bisher kleinsten Wert dar, der bisher für eine Protein *trans*-Spleißreaktion mit gereinigten Proteinen ermittelt wurde. Die Reaktion des *Npu* DnaE Inteins ist damit im Vergleich zu dem bisherigen Spitzenreiter, dem *Sce* VMA Intein [163], etwa sechsmal schneller. Die Reaktionskonstanten erster Ordnung des *Ssp* DnaE Inteins mit den Exteinen gpD und

Thioredoxin waren etwa zwei Zehnerpotenzen kleiner als die entsprechenden Werte für das *Npu* DnaE Intein (s. Tabelle 5). Die in dieser Arbeit ebenfalls ermittelten Werte für das *Ssp* DnaE Intein befanden sich im Bereich der bisher veröffentlichten Daten [173,177]. Damit erscheint es als sehr unwahrscheinlich, dass die hohen Raten für das *Npu* DnaE Intein auf spezielle Laborbedingungen zurückzuführen sind.

**Tabelle 5** Ermittelte Reaktionskonstanten erster Ordnung und prozentuale Spleißproduktbildung der untersuchten *Npu* und *Ssp* DnaE Inteinfusionsproteine bei verschiedenen Temperaturen.

N-terminales Konstrukt	C-terminales Konstrukt	T /°C	Spleißproduktbildung /%	$k_{\text{Spleißen}} / \text{s}^{-1}$
<b>1</b>	<b>2</b>	20	80 - 85	$(3.6 \pm 0.5) \times 10^{-3}$
		37	80 - 90	$(9.0 \pm 0.4) \times 10^{-3}$
<b>1</b>	Int <sup>C</sup> -Trx-H <sub>6</sub>	6	85 - 95	$(8.5 \pm 0.4) \times 10^{-4}$
		12	75 - 80	$(2.2 \pm 0.5) \times 10^{-3}$
		25	75 - 85	$(3.5 \pm 0.2) \times 10^{-3}$
		37	75 - 80	$(1.1 \pm 0.2) \times 10^{-2}$
<b>1</b>	Int <sup>C</sup> -eGFP-H <sub>6</sub>	6	80 - 90	$(1.6 \pm 0.1) \times 10^{-3}$
		25	75 - 80	$(4.3 \pm 0.7) \times 10^{-3}$
ST-eGFP-Int <sup>N</sup>	Int <sup>C</sup> -Trx-H <sub>6</sub>	25	50 - 55	$(4.7 \pm 1.2) \times 10^{-3}$
ST-gpD-Int <sup>N</sup> <sub><i>Ssp</i></sub>	Int <sup>C</sup> <sub><i>Ssp</i></sub> -Trx-H <sub>6</sub>	25	40 - 50	$(8.0 \pm 1.4) \times 10^{-5}$
		37	20 - 25	$(1.6 \pm 0.3) \times 10^{-4}$

Um darüberhinaus ausschließen zu können, dass die hohen Reaktionsgeschwindigkeiten auf die Reinigungsprozedur der C-terminalen Hälfte zurückgehen, wurde ein Reinigungsprotokoll unter vollständig nativen Bedingungen entwickelt. Die Fusionierung eines N-terminalen StrepTag II auf genetischer Ebene erlaubte die affinitätschromatographische Reinigung des Proteins ST-Int<sup>C</sup>-gpD-H<sub>6</sub> (**3**) in zwei Schritten ohne den Einsatz von 8M Harnstoff. Das nativ gereinigte Protein verhielt sich in PTS-Reaktionen mit der N-terminalen Hälfte ST-gpD-Int<sup>N</sup> **1** identisch zu dem unter denaturierenden Bedingungen isolierten Protein **3** (s. Abbildung 39). Im Vergleich mit Protein **2** (ohne den N-terminalen StrepTag II) zeigten beide Präparationen jedoch eine um den Faktor 1,5 verminderte Geschwindigkeitskonstante (s. Tabelle 12A im Anhang). Vergleichbare Ergebnisse konnten zudem mit Trx als C-Extein erhalten werden. Die Reinigungsbedingungen hatten demnach keinen messbaren Einfluss auf die PTS-

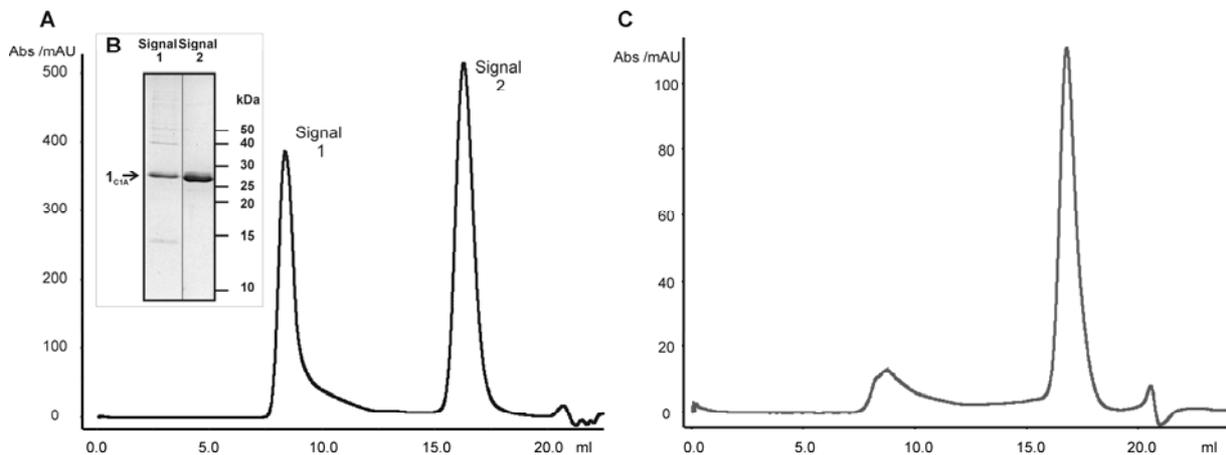


### Untersuchung der Bildung löslicher Aggregate von Intefusionsproteinen

Das natürlich gespaltene *Npu* DnaE Intein zeigt in den hier durchgeführten Studien nicht nur eine sehr hohe Reaktionsgeschwindigkeit, sondern auch gute Spleißausbeuten von etwa 85% (s. Tabelle 5). Allerdings waren die Ausbeuten der PTS-Reaktionen sowohl von den individuellen Proteinpräparationen (Daten nicht gezeigt), als auch von der Natur der Exteine abhängig. So konnte z.B. für eGFP als N-Extein nur eine maximale Spleißausbeute von 50% erreicht werden (s. Tabelle 5). Zur Erklärung unvollständiger Reaktionsumsätze kommen für gespaltene Inteine unterschiedliche Möglichkeiten in Frage (s. auch [163,165,169,177,187]). Eine oxidative Inaktivierung der katalytischen Cysteinreste ist genauso denkbar wie potentielle Fehlfaltungen bei der Assoziation der Inteinhälften. Gespaltene Inteine neigen insgesamt zur Bildung von so genannten „inclusion bodies“ bei der Überexpression in *E. coli* und tragen oft zur schlechten Löslichkeit der Fusionsproteine bei [158]. Vor diesem Hintergrund wäre das Auftreten von löslichen Aggregaten der individuellen Inteinhälften denkbar, die keine produktive Komplexbildung mit einer komplementären Inteinhälfte zulassen würden. Dieser Erklärungsansatz wurde im Laufe dieser Doktorarbeit in Zusammenarbeit mit Simone Eppmann mittels Gelfiltrationsexperimenten an den Proteinen ST-gpD-Int<sup>N</sup><sub>C1A</sub> (**1**<sub>C1A</sub>) und Int<sup>C</sup><sub>N36A</sub>-gpD-H<sub>6</sub> (**2**<sub>N36A</sub>) überprüft. Durch Mutation der katalytischen Reste, der ersten und letzten Aminosäure des Inteins, konnte bei diesen Proteinen nur die Komplexbildung der Inteinhälften und keine Spleiß- oder Spaltungsreaktionen stattfinden.

Das Vorhandensein von löslichen Aggregaten der C-terminalen *Npu* DnaE Inteinhälften wurde bereits anhand der Spleißaktivität des nativ und denaturierend gereinigten Proteins **3** widerlegt (s. oben). Gelfiltrationsexperimente mit Int<sup>C</sup><sub>N36A</sub>-gpD-H<sub>6</sub> (**2**<sub>N36A</sub>) unterstützten diese Daten, da neben einem Signal bei ungefähr der erwarteten Größe des monomeren Intefusionsproteins nur ein kleiner Anteil an Komplexen mit höherem Molekulargewicht auftrat (s. Abbildung 40C). Im Unterschied dazu ergab ein Gelfiltrationslauf der N-terminalen Hälfte (**1**<sub>C1A</sub>) zwei deutlich voneinander getrennte Signale mit vergleichbarer Intensität (s. Abbildung 40A). Zu beiden Zeitpunkten eluierte das Protein **1**<sub>C1A</sub> (s. Abbildung 40B). Im Vergleich mit einem Größenstandard (Gel Filtration Calibration Kit - Low Molecular Weight von GE Healthcare) konnte für die Retentionszeit des Signals 2 ein Molekulargewicht von 29,8 kDa, in Übereinstimmung mit dem theoretischen Wert für das Monomer von 24,7 kDa, berechnet werden. Signal 1 zeigte eine komplexe Spezies mit einem Elutionsvolumen knapp unterhalb des Ausschlussvolumens der Gelfiltrationssäule an, so dass hier auf Aggregate mit einem sehr großen Molekulargewicht (~ 1,3 MDa) geschlossen werden kann. Die zusätzlich

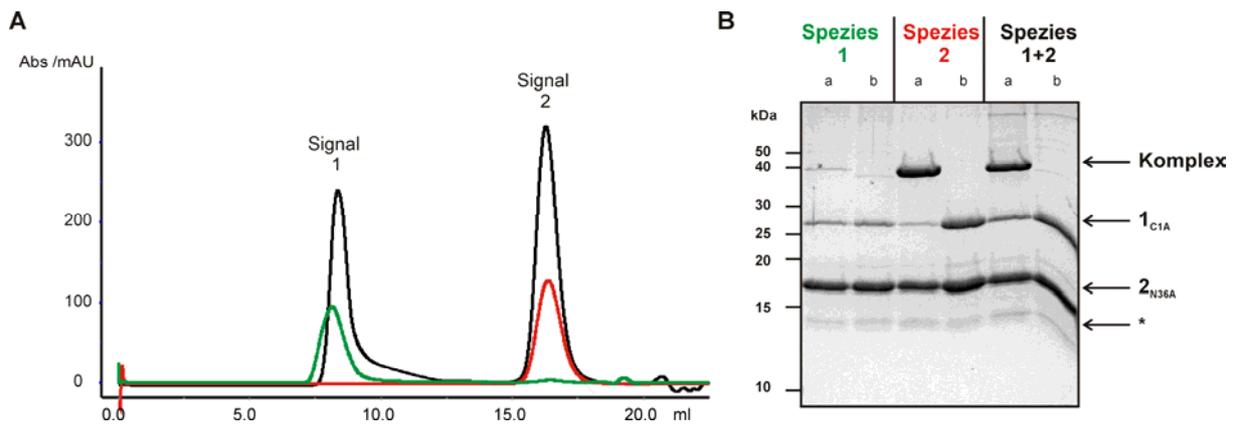
neben  $1_{C1A}$  auftretenden Banden in der SDS-PAGE Analyse von Signal 1 (s. Abbildung 40B) wurden mittels tryptischen Verdau und anschließender MALDI-TOF MS analysiert. Jedoch konnte nach einer Datenbanksuche den Banden kein *E. coli* Protein eindeutig zugeordnet werden (Daten nicht gezeigt). Aus diesem Grund konnte die Zusammensetzung der löslichen Aggregate in dieser Arbeit letztlich nicht geklärt werden.



**Abbildung 40 Gelfiltrationsläufe von *Npu* DnaE Inteininfusionsproteinen.** Gezeigt sind repräsentative Chromatogramme von ST-gpD-Int<sup>N</sup><sub>C1A</sub> ( $1_{C1A}$ ) (A) und Int<sup>C</sup><sub>N36A</sub>-gpD-H<sub>6</sub> ( $2_{N36A}$ ) (C); B) SDS-PAGE Analyse jeweils einer Fraktion von Signal 1 und 2; in beiden Signalen ist das Protein ST-gpD-Int<sup>N</sup><sub>C1A</sub> ( $1_{C1A}$ ) vorhanden.

Die in Abbildung 40A präsentierte Präparation von ST-gpD-Int<sup>N</sup><sub>C1A</sub> ( $1_{C1A}$ ) wurde bei einer Induktionstemperatur von 28°C exprimiert. Eine Expression bei 20°C führte zu einem im Vergleich mit Signal 2 stärkeren Signal 1, also der vermehrten Bildung von löslichen Aggregaten (s. Abbildung 73A im Anhang). Bei einer Erhöhung der Induktionstemperatur auf 34°C befand sich nach Zelllyse ein größerer Anteil des Proteins in der unlöslichen Fraktion (Daten nicht gezeigt) und das Aggregat/Monomer Verhältnis in der löslichen Fraktion war identisch zu der Expression bei 28°C (s. Abbildung 73A im Anhang). Das Auftreten von Aggregaten der N-terminalen Fusionsproteine lässt sich demnach nicht durch eine Änderung der Expressionsbedingungen umgehen.

Das stark unterschiedliche Laufverhalten der monomeren und aggregierten Spezies in den Gelfiltrationsexperimenten ermöglichte deren präparative Trennung (s. Abbildung 40A). Bei einer Inkubation über 24 Stunden bei 25°C konnte keine Umwandlung der Spezies von Signal 2 in die aggregierte Form festgestellt werden (s. Abbildung 41A, rotes Chromatogramm). Das Gelfiltrationschromatogramm des identisch behandelten, gereinigten Proteins des Signals 1 zeigte nur Spuren der monomeren Spezies (s. Abbildung 41A, grünes Chromatogramm), so dass sich innerhalb dieses Zeitraumes beide Spezies stabil verhielten und keine oder nur eine sehr geringe Umwandlung festzustellen war.



**Abbildung 41 Trennung von verschiedenen Spezies eines N-terminalen *Npu* DnaE Inteininfusionsproteins mittels Gelfiltration.** A) Nach einer präparativen Trennung der Proteine der Signale 1 und 2 (ST-gpD-Int<sup>N</sup><sub>C1A</sub> (**1**<sub>C1A</sub>), schwarzes Chromatogramm) wurden die Fraktionen separat jeweils für 24h bei 25°C inkubiert und anschließend erneut mittels Gelfiltration analysiert - Präparation des Signals 1 (grünes Chromatogramm), Präparation des Signals 2 (rotes Chromatogramm); innerhalb dieses Zeitraums ist keine oder nur eine sehr geringe Umwandlung der Spezies ineinander festzustellen; B) Fähigkeit zur Komplexbildung der unterschiedlichen ST-gpD-Int<sup>N</sup><sub>C1A</sub> (**1**<sub>C1A</sub>) Präparationen; sowohl die präparativ getrennten Signale 1 (Spezies 1) und 2 (Spezies 2), als auch die Mischung vor der Trennung (Spezies 1+2) wurden mit einem Überschuss der C-terminalen Hälfte Int<sup>C</sup><sub>N36A</sub>-gpD-H<sub>6</sub> (**2**<sub>N36A</sub>) inkubiert und mittels SDS-PAGE analysiert; gezeigt sind die nicht hitzedenaturierten (a) und die hitzedenaturierten (b) SDS-Proben (\* = Verunreinigung).

Die Fähigkeit zur Komplexbildung der präparativ gereinigten Spezies von ST-gpD-Int<sup>N</sup><sub>C1A</sub> (**1**<sub>C1A</sub>) wurde mit einem Überschuss an Int<sup>C</sup><sub>N36A</sub>-gpD-H<sub>6</sub> (**2**<sub>N36A</sub>) untersucht und ist in Abbildung 41B gezeigt. Das Auftreten einer neuen Bande bei den nicht hitzedenaturierten SDS-PAGE Proben (s. Proben a) in Abbildung 41B) auf der Höhe von ~ 40 kDa zeigt die Bildung des heterodimeren Komplexes an (theoretisches Molekulargewicht des Komplexes 41,9 kDa). Tryptische Verdauung und MALDI-TOF MS Analysen bestätigten das Vorhandensein sowohl des N- als auch des C-terminalen Inteininfusionsproteins in dieser Bande (Daten nicht gezeigt). Beim Aufkochen der Proben (s. Proben b) in Abbildung 41B) verschwand dieser nicht-kovalente Komplex wieder. Die ST-gpD-Int<sup>N</sup><sub>C1A</sub>-Proteine der Spezies 1 mit einem sehr geringen Elutionsvolumen in der Gelfiltration bildeten zu einem marginalen Anteil den Komplex mit der C-terminalen Hälfte. Jedoch ist die monomere Spezies (die gereinigten Proteine des Signals 2) in der Lage den heterodimeren Komplex fast vollständig zu bilden. Es ist also möglich, die nach der Expression in *E. coli* vorliegende Mischung aus inaktiven, löslichen Aggregaten der N-terminalen Hälfte **1**<sub>C1A</sub> und der reaktiven, monomeren Spezies **1**<sub>C1A</sub> mit Hilfe der Gelfiltration zu trennen und so Reaktanden zu erhalten, die einen nahezu vollständigen Reaktionsumsatz ermöglichen könnten.

In Übereinstimmung mit der nur sehr geringen PTS-Ausbeute von ~ 50% zeigte ST-eGFP-Int<sup>N</sup> (s. Tabelle 5) einen hohen Anteil an löslichen Aggregaten bei der Gelfiltrationsanalyse (s. Abbildung 74A im Anhang). Durch einen einfachen

Denaturierungsschritt (30 min bei 37°C in Gegenwart von 8M Harnstoff) und anschließender Rückfaltung in Spleißpuffer konnte der Anteil der monomeren Spezies erhöht werden (s. Abbildung 74A im Anhang). Jedoch sind weitere Versuche notwendig, um hier ein effektives Protokoll der Umwandlung zu etablieren.

Ob die in dieser Arbeit nachgewiesenen löslichen Aggregate der Inteinproteine auch in anderen gespaltenen Inteinsystemen eine Erklärung für die unvollständigen Spleißausbeuten sind, wurde am Beispiel des künstlich gespaltenen *Ssp* DnaB Inteins überprüft. Dieses an der Position der Endonuklease gespalte Intein zeigte bei Reaktionen von äquimolaren Mengen der N- und C-terminalen Fusionsproteinen eine maximale Spleißausbeute von 60% [163,164]. Gelfiltrationsanalysen der Proteine MBP-*Ssp* DnaB<sup>N</sup>-FKBP-H<sub>6</sub> und MBP-*Ssp* DnaB<sup>C</sup>-H<sub>6</sub> zeigten jedoch nur einen sehr geringen Anteil löslicher Aggregate für die N-terminale Hälfte. Die C-terminale Hälfte war frei von Aggregaten (s. Abbildung 75A im Anhang). Allerdings bleibt anzumerken, dass bei Inkubationen mit einer Inteinhälfte im Überschuss das Partnerprotein im Falle der C-terminalen Hälfte komplett und im Falle der N-terminalen Hälfte fast vollständig verbraucht wird (Daten nicht gezeigt).

Zusammenfassend lässt sich festhalten, dass wahrscheinlich das Auftreten von löslichen Aggregaten zum einen eine für jede gespalte Inteinhälfte individuelle Eigenschaft ist und zum anderen auch von der Natur des fusionierten Exteins abhängt. Im Falle des *Npu* DnaE Inteins wandeln sich die inaktiven löslichen Aggregate der N-terminalen Hälfte, wenn überhaupt, nur sehr langsam in die monomere, aktive Form um. Obwohl sicherlich noch andere Faktoren wie z.B. die Oxidation der katalytischen Reste eine wichtige Rolle spielen, bieten die hier präsentierten Daten doch eine mögliche Erklärung für Spleißreaktionen mit unvollständigem Umsatz. Eine Abtrennung der inaktiven Aggregate konnte leicht durch einen zusätzlichen Reinigungsschritt mittels Gelfiltration erfolgen.

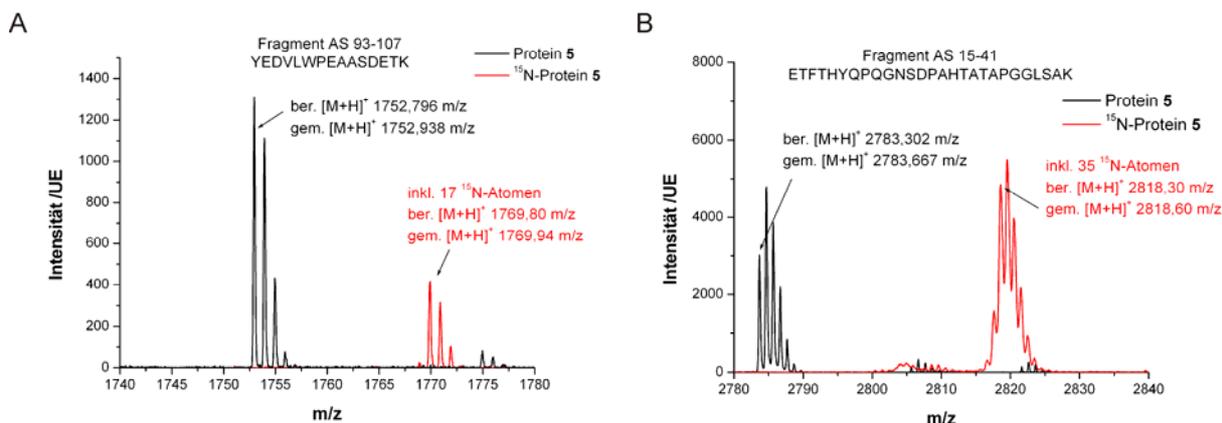
### 4.2.2 NMR-Untersuchungen von gespaltenen Inteinen

Die bisher ermittelten strukturellen Daten von Inteinen basieren alle auf Kristall- oder NMR-Strukturanalysen von *cis*-Inteinen (s. z.B. [133,146,226] und Kapitel 1.2 der Einleitung). Zur Untersuchung von natürlich gespaltenen Inteinen wurden diese bisher immer auf genetischer Ebene fusioniert, so dass auch in diesen Fällen ein artifizielles *cis*-Intein analysiert wurde [137,168]. Dementsprechend liegen im Moment keine Aussagen über die strukturellen Eigenschaften der individuellen Hälften gespalteener Inteine ohne das korrespondierende Partnerprotein vor. Die Bestimmung einer möglichen Vorfaltung der Inteinhälften und ein detailliertes Verständnis der Fragmentassoziation könnten jedoch unter anderem zur Erklärung der unterschiedlichen Reaktionsgeschwindigkeiten der gespaltenen Inteine entscheidend beitragen. Weiterhin wäre ein generelles Verständnis der verschiedenen Zwischenstufen der PTS-Reaktion auf atomarer Ebene wünschenswert. So erschwert z.B. das fehlende Wissen über den Einfluss der Exteinaminosäuren auf den Spleißprozess die Vorhersage, ob ein gespaltenes Intein in einem bestimmten Exteinkontext aktiv ist oder nicht. Eine gezielte rationale Optimierung von gespaltenen Inteinen in Bezug auf Reaktionsgeschwindigkeit, Exteinkompatibilität und Spleißausbeute ist anhand der aktuellen Datenlage nicht möglich. Um eine mögliche Grundlage für eine solche Herangehensweise zu schaffen wurden in Zusammenarbeit mit Alena Busche und Dr. Frank Löhr (AK Prof. V. Dötsch, Goethe-Universität Frankfurt am Main), sowie Simone Eppmann NMR-Untersuchungen an gespaltenen Inteinen in Lösung durchgeführt.

#### Strukturelle Untersuchungen von gespaltenen Inteinen ohne das Partnerprotein

Für die NMR-Spektroskopie werden Lösungen des zu untersuchenden Proteins in hoher Konzentration benötigt. Vor allem für künstlich gespaltene Inteine sind Werte über 300  $\mu\text{M}$  nur mit effektiven Schlepperproteinen zu erreichen. In dieser Arbeit wurde das Kopfprotein des Bakteriophagen  $\lambda$  (gpD) zur Verbesserung der Löslichkeit eingesetzt [227]. Neben seiner geringen Größe (110 AS, 11,6 kDa) ist gpD zusätzlich frei von Cysteinresten und NMR-spektroskopisch bereits ausführlich untersucht worden [228]. Alle in dieser Arbeit untersuchten gpD-Inteinfusionsproteine sind in Tabelle 6 dargestellt. Die Anreicherung des NMR-aktiven Isotops  $^{15}\text{N}$  in den Proteinen gelang durch Expression in M9 Minimalmedium mit  $^{15}\text{N}$ -Ammoniumchlorid als einziger Stickstoffquelle (s. Kapitel 3.2.1). Tryptische Verdauung der SDS-PAGE Gelbanden mit anschließender massenspektrometrischer Analyse bestätigten

den Einbau der schwereren Isotope im Vergleich zu einer in LB-Medium exprimierten Probe (s. Abbildung 42).



**Abbildung 42**  $^{15}\text{N}$ -Isotopenmarkierung von ST-gpD-H<sub>6</sub> (5). A) und B) Ausschnitte aus den MALDI-TOF MS Analysen der tryptischen in-Gelverdaue der gereinigten Proteine, die entweder in LB-Medium (schwarz) oder in M9 Minimal Medium mit  $^{15}\text{N}$ -Ammoniumchlorid als einziger Stickstoffquelle (rot) exprimiert wurden; da der Einbau der schwereren Isotope mit der Anzahl der Stickstoffatome in den untersuchten Fragmenten korrespondiert, kann auf eine vollständige Markierung geschlossen werden.

Die gpD-Inteinfusionsproteine (Konzentrationen zwischen 200 und 700  $\mu\text{M}$  - Puffer: 25 M Tris, 50 mM NaCl, 2 mM DTT, pH 7,0) wurden bei 20°C mittels  $^{15}\text{N}$ ,  $^1\text{H}$  Korrelationsspektroskopie untersucht ( $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY - transverse relaxation-optimised spectroscopy [218]). Eine Zuordnung der gemessenen Resonanzen wurde im Rahmen dieser Doktorarbeit nicht durchgeführt, jedoch gelang die Identifikation der gpD-Signale in den Spektren der Fusionsproteine durch Vergleich mit dem Protein ST-gpD-H<sub>6</sub> (5). Über die Linienbreite und chemische Verschiebung der übrigen, zusätzlichen Signale konnten Hinweise auf die strukturellen Eigenschaften der untersuchten individuellen Inteinhälften gezogen werden (s. Abbildung 43 und Abbildung 44, sowie Tabelle 6).

**Tabelle 6** Mittels NMR-Spektroskopie untersuchte Inteinfusionsproteine.

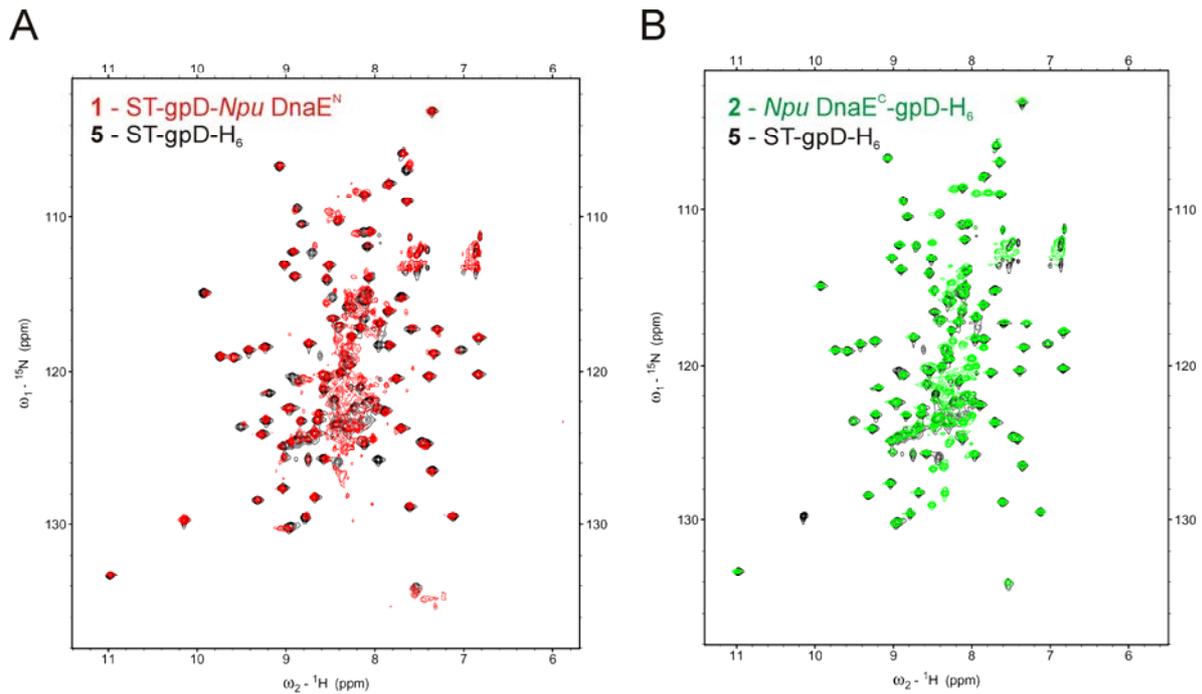
Fusionsprotein	Sequenz	Die NMR-Signale der Inteinhälften sind		Wahrscheinliche Struktur der Inteinhälfte
		scharf und dispers	linienverbreitert	
gpD- <i>Npu</i> DnaE <sup>N</sup> (1)	MASWSHPQFEKAS-gpD-GS- <i>Npu</i> DnaE <sup>N</sup> (1-102)	nein	ja	MG
<i>Npu</i> DnaE <sup>C</sup> -gpD (2)	<i>Npu</i> DnaE <sup>C</sup> (1-36)-CFNGTAS-gpD-GSRSHHHHHH	nein	nein	RC
gpD-DnaB <sup>C</sup> (Endo) (6)	MASWSHPQFEKAS-gpD-GSSS- <i>Ssp</i> DnaB <sup>C</sup> (107-154)-SIRSCGKLAALAEHHHHHHH	nein	teilweise	MG oder RC
DnaB <sup>C</sup> (11)-gpD (7)	MGTS- <i>Ssp</i> DnaB <sup>C</sup> (12-154)-SIEGSAS-gpD-GSRSHHHHHH	nein	ja	MG

Verwendete Abkürzungen: DnaB<sup>C</sup>(Endo) = C-terminales Inteinfragment, des an der Position der Endonuklease gespaltenen *Ssp* DnaB Inteins; DnaB<sup>C</sup>(11) = C-terminales Inteinfragment, des an der Position 11 gespaltenem *Ssp* DnaB Inteins; MG = „molten globule“; RC = „random coil“, für Details s. Text.

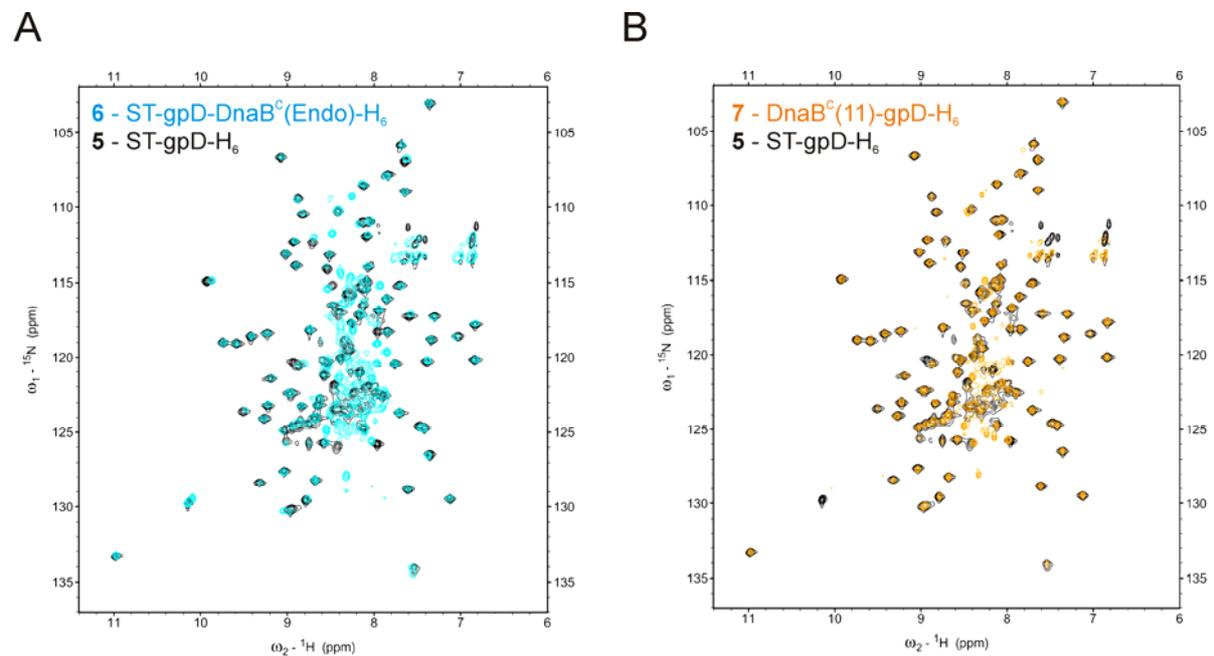
Scharfe und disperse Signale im  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY Spektrum würden auf eine Faltung des untersuchten Proteinteils hindeuten. Jedoch zeigten sowohl die C-terminalen Hälften des künstlich gespaltenen *Ssp* DnaB Inteins (an Position 11 und an der Position der Endonuklease gespalten), als auch die N- und C-terminale Hälften des natürlich gespaltenen *Npu* DnaE Inteins keine definierte Vorfaltung in Abwesenheit des Partnerproteins. Im Laufe dieser Doktorarbeit gelang es nicht, die N-terminale Hälfte des an der Position der Endonuklease gespaltenen *Ssp* DnaB Inteins (AS 1-104) im notwendigen Konzentrationsbereich zu erhalten. Auch durch die Fusion mit drei gpD-Domänen (ST-gpD-*Ssp* DnaB<sup>N</sup>(1-104)-gpD-gpD-H<sub>6</sub>) konnten nur sehr geringe Mengen an löslichem Protein erhalten werden (Daten nicht gezeigt). Es wird angenommen, dass auch diese Inteinhälfte in keiner definierten Faltung vorliegt und damit eine verstärkte Tendenz zur Bildung von unlöslichen Aggregaten besitzt.

Ungefaltete Proteine oder Proteinfragmente werden in der Literatur (für Reviews s. unter anderem [229-232] in weitere, oft ineinander übergehende Gruppen unterteilt. In den als „molten globule“ vorliegenden Proteinen [233-235] sind die Sekundärstrukturelemente bereits vorhanden, jedoch bilden die hydrophoben Seitenketten im Inneren des Proteins einen dynamischen, vom Lösungsmittel geschützten Kern („geschmolzener Tropfen“). Die NMR-Signale der Rückgrataminosäuren solcher Proteine sind wenig dispers [230] und besitzen im Vergleich mit vollständig gefalteten Proteinen ähnliche oder größere Linienbreiten [236-239]. Vollständig denaturierte Proteine (*engl.* „random coil“) ergeben NMR-Spektren mit scharfen, nicht dispersen Signalen.

Wie in Abbildung 43A und Abbildung 44B ersichtlich, deuten die geringe Signalintensität und -dispersität der Spektren der *Npu* DnaE<sup>N</sup>- und der *Ssp* DnaB<sup>C</sup>(12-154)-Inteinhälften darauf hin, dass sie als Ensemble von Konformationen im dynamischen Gleichgewicht vorliegen und damit eher einer „molten globule“ Struktur zugeordnet werden können. Aufgrund der geringen Größe von nur 36 Aminosäuren und den scharfen, aber wenig dispersen Signalen im  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY Spektrum nimmt die C-terminale Hälfte des *Npu* DnaE Inteins wahrscheinlich einen komplett ungefalteten Zustand („random coil“) ein (s. Abbildung 43B). Das Spektrum des Proteins ST-gpD-*Ssp* DnaB<sup>C</sup>(Endo)-H<sub>6</sub> (**6**) enthält sowohl verbreiterte, als auch scharfe, nicht disperse Signale (s. Abbildung 44A). Letztere könnten von dem ungefalteten, relativ großen C-Extein (13 AS ohne den H<sub>6</sub>-Tag, s. Tabelle 6) stammen. Damit wären die verbreiterten Signale einer als „molten globule“ vorliegenden Inteinhälfte zuzuordnen. Eine Linienverbreiterung der Inteinsignale aufgrund einer Oligomerisierung der Proteine bei hohen Konzentrationen kann ausgeschlossen werden, da die Signale der gpD-Proteine weiterhin scharf und dispers sind.



**Abbildung 43** NMR-Untersuchungen der Inteinhälften des *Npu* DnaE Inteins.  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektren der N-terminalen (ST-gpD-*Npu* DnaE<sup>N</sup> (1), (A), rot), sowie der C-terminalen Inteinhälfte (*Npu* DnaE<sup>C</sup>-gpD-H<sub>6</sub> (2), (B), grün); zum Vergleich ist jeweils das  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektrum des Referenzproteins ST-gpD-H<sub>6</sub> (5) in schwarz gezeigt.

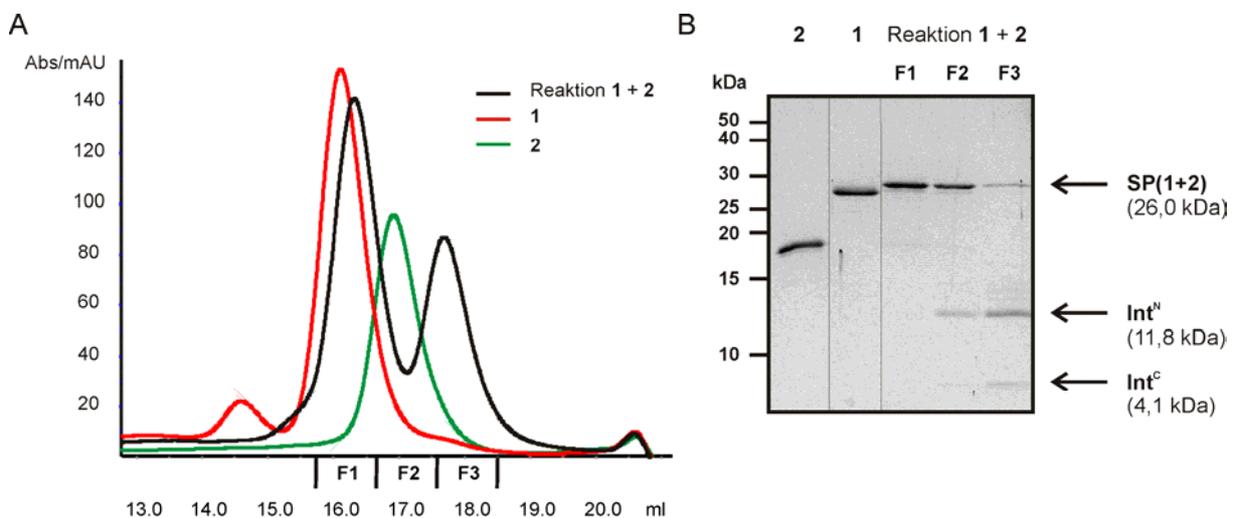


**Abbildung 44** NMR-Untersuchungen der C-terminalen Inteinhälften des *Ssp* DnaB Inteins.  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektren des an der Position der Endonuklease (ST-gpD-DnaB<sup>C</sup>(Endo)-H<sub>6</sub> (6), (A), cyan), sowie des an Position 11 gespaltenen Inteins (DnaB<sup>C</sup>(11)-gpD-H<sub>6</sub> (7) (B), orange); zum Vergleich ist jeweils das  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektrum des Referenzproteins ST-gpD-H<sub>6</sub> (5) in schwarz gezeigt.

### Charakterisierungen von Inteincomplexen

Zur Katalyse der PTS-Reaktion müssen die zwei ungefalteten Inteinhälften assoziieren und den aktiven Inteincomplex ausbilden. Die Faltung einer Inteinhälfte lässt sich mit der hier

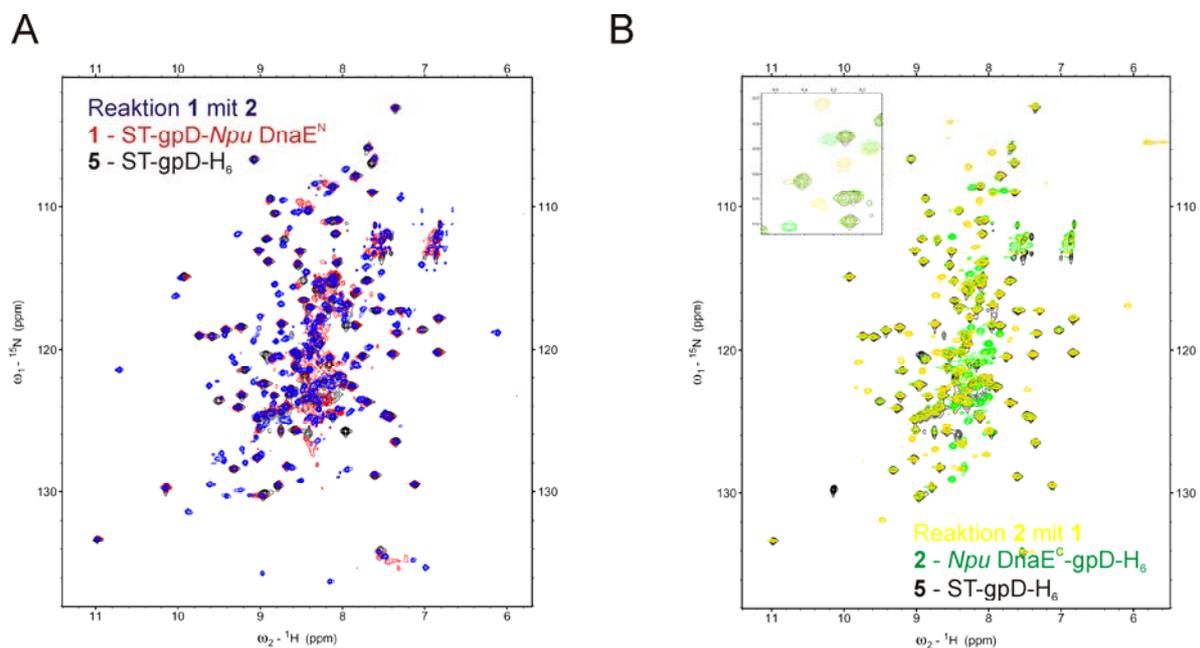
präsentierten Methode sehr gut untersuchen, indem zu der  $^{15}\text{N}$ -markierten Hälfte ein unmarkiertes Partnerprotein hinzugegeben wird. In den entstehenden Komplexen wäre dann nur ein Teil des Inteins NMR-sichtbar. Bei Inkubation der *Npu* DnaE Inteinhälften (Proteine **1** und **2**) kann die PTS-Reaktion stattfinden und es entstehen neben dem Spleißprodukt noch die gespleißten Inteinhälften (s. Abbildung 38). Die sehr schnelle Kinetik dieser Reaktion erlaubte jedoch keine Analyse von Reaktionsintermediaten. Bereits mit der Aufnahme des ersten  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY Spektrums etwa 1h nach dem Mischen der beiden Proteine, war der Endpunkt der Reaktion erreicht. Bis 16h später aufgenommene Spektren waren mit dem ersten Spektrum identisch (Daten nicht gezeigt). In Abbildung 46A ist das  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY Spektrum von  $^{15}\text{N}$ -ST-gpD-Int<sup>N</sup> (**1**) mit dem unmarkiertem Protein Int<sup>C</sup>-gpD-H<sub>6</sub> (**2**) nach der vollständigen Reaktion gezeigt. Das Verschwinden der ursprünglichen Inteinsignale und das Entstehen von neuen dispersen Signalen impliziert, dass die Int<sup>N</sup>-Hälfte nun in einem definierten, gefalteten Zustand vorliegt. Ein analoges Bild ergibt sich bei der Reaktion des mit  $^{15}\text{N}$  markierten C-terminalen Fusionsprotein **2** mit unmarkiertem Protein **1** (s. Abbildung 46B). Die daraus abgeleitete Annahme, dass die beiden Inteinhälften nach der Spleißreaktion in einem gefalteten Komplex vorliegen, wurde durch Gelfiltrationsexperimente weiter bestätigt. Hier eluieren Int<sup>N</sup> und Int<sup>C</sup> trotz unterschiedlicher Größe (11,9 kDa, bzw. 4,1 kDa) gemeinsam (s. Abbildung 45).



**Abbildung 45 Die Hälften des gespaltenen *Npu* DnaE Inteins liegen nach Spleißen in einem stabilen Komplex vor.** A) Gelfiltrationsläufe des Proteins ST-gpD-*Npu* DnaE<sup>N</sup> (**1**) (gelbes Chromatogramm), des Proteins *Npu* DnaE<sup>C</sup>-gpD-H<sub>6</sub> (**2**) (rotes Chromatogramm) und der PTS-Spleißreaktion von **1** und **2** (schwarzes Chromatogramm); B) Mittels SDS-PAGE Analyse der Fraktionen konnte den beiden Signalen im Gelfiltrationschromatogramm nach Spleißen das Spleißprodukt ST-gpD-gpD-H<sub>6</sub> **SP(1+2)**, sowie der Komplex der Inteinhälften zugeordnet werden, da beide Hälften **Int<sup>N</sup>** und **Int<sup>C</sup>** gleichzeitig eluieren.

Eine Zuordnung der Signale des Inteincomplexes mit Hilfe von  $^{15}\text{N}$ ,  $^{13}\text{C}$ -markierten Proben wurde im Rahmen dieser Arbeit nicht durchgeführt. Im Vergleich mit den von der

Gruppe um Iwai veröffentlichen NMR-Daten des fusionierten *cis*-Inteins [168,240] konnten nur die stark dispersen Signale eindeutig zugeordnet werden (s. Abbildung 76A im Anhang). Die Zuordnung von 8 der 134 Amid-Signale lässt jedoch den Schluss zu, dass der hier untersuchte Inteinkomplex strukturell nicht völlig von der veröffentlichten Struktur des fusionierten *cis*-Inteins [168] abweicht. Unterschiede in den chemischen Verschiebungen der Signale sind durch die verschiedenen Pufferbedingungen (pH 6,0 im Vergleich zu pH 7,0), sowie die leicht abweichende Primärstruktur der untersuchten Proteine (von Heinämäki *et al.* untersuchtes Protein: GG-A-Int<sup>N</sup>(2-102)-Int<sup>C</sup>(2-36)) zu erklären.



**Abbildung 46 NMR-Untersuchungen der PTS-Reaktion des *Npu DnaE* Inteins.**  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektren der N-terminalen (ST-gpD-*Npu DnaE*<sup>N</sup> (1), (A), blau), sowie der C-terminalen Inteinhälfte (*Npu DnaE*<sup>C</sup>-gpD-H<sub>6</sub> (2), (B), gelb) nach der vollständigen PTS-Reaktion mit dem unmarkierten Partnerprotein 2, respektive 1; zum Vergleich ist jeweils das  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektrum des Referenzproteins ST-gpD-H<sub>6</sub> (5) in schwarz, sowie die Spektren der N-terminalen (ST-gpD-*Npu DnaE*<sup>N</sup> (1), (A), rot), sowie der C-terminalen Inteinhälfte (*Npu DnaE*<sup>C</sup>-gpD-H<sub>6</sub> (2), (B), grün) alleine gezeigt; das Entstehen von neuen, dispersen Signalen in beiden Spektren deutet auf die Bildung eines gefalteten Komplexes nach Spleißen hin.

Neben der Untersuchung des Inteinkomplexes nach der Spleißreaktion ermöglicht das hier präsentierte *Npu DnaE* System durch Mutationen der katalytischen Reste eine Betrachtung der verschiedenen Zwischenstufen der PTS-Reaktion. Die Reaktion von Int<sup>C</sup><sub>N36A</sub>-gpD-H<sub>6</sub> (2<sub>N36A</sub>) mit Protein 1 (ST-gpD-Int<sup>N</sup>) kann nur bis zur Stufe des verzweigten Intermediats ablaufen. Die abschließende Asparaginzyklisierung wird durch die N36A Mutation verhindert. Das  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY Spektrum von 2<sub>N36A</sub> ist identisch mit dem Spektrum des nativen Proteins 2, so dass auch für die mutierte C-terminale *Npu DnaE* Inteinhälfte eine „random coil“ Konformation angenommen werden kann. Bei Zugabe eines Überschusses an unmarkiertem

N-terminalen Protein **1** verschwinden die Inteinsignale und es entstehen wenige neue Signale mit einer sehr geringen Intensität (s. Abbildung 77A im Anhang). Diese Linienverbreiterung deutet auf einen konformationellen Austausch hin und könnte mit der Umwandlung der während der Spleißreaktion auftretenden Komplexe (Anlagerungskomplex, erster Thioesterkomplex und verzweigtes Intermediat) ineinander erklärt werden.

Weiterhin wurde im Rahmen dieser Doktorarbeit versucht den Anlagerungskomplex der beiden *Npu* DnaE Inteinhälften mittels NMR zu verfolgen. Dazu wurden die beiden Proteine  $H_6\text{-gpD-Int}^N_{C1A}$  (**8**<sub>C1A</sub>) und  $\text{Int}^C_{N36A,C+1A}\text{-gpD-H}_6$  (**2**<sub>N36A,C+1A</sub>) verwendet. In der N-terminalen Hälfte **8**<sub>C1A</sub> wurde zur vereinfachten Reinigung im Vergleich zu Protein **1**<sub>C1A</sub> der StrepTag II gegen eine Hexahistidinsequenz ausgetauscht. Ansonsten sind diese Proteine identisch und zeigen auch vergleichbare <sup>15</sup>N, <sup>1</sup>H-TROSY Spektren. Die Komplexbildung zwischen den Proteinen **8**<sub>C1A</sub> und **2**<sub>N36A,C+1A</sub> kann durch Gelfiltrationsexperimente und SDS-PAGE Analyse, der Komplex ist ohne Aufkochen der Proben stabil, nachgewiesen werden (Daten nicht gezeigt). Bei Zugabe von unmarkiertem C-terminalen Protein **2**<sub>N36A,C+1A</sub> zu <sup>15</sup>N-markiertem Protein **8**<sub>C1A</sub> verschwinden die Signale der mutierten N-terminalen Inteinhälfte. Das Auftreten neuer jedoch im Vergleich zum Komplex nach Spleißen schwächerer Signale deutet auf eine Faltung der Inteinhälfte auch im Anlagerungskomplex hin (s. Abbildung 78A im Anhang). Die größere Linienbreite in dem Komplex **8**<sub>C1A</sub> und **2**<sub>N36A,C+1A</sub> kann entweder durch ein dynamisches Verhalten in diesem ersten Anlagerungskomplex oder durch die schnellere Spin-Relaxation in dem nun größeren Protein (41,9 kDa) erklärt werden. Da auch die Signale des gpD-Exteins eine größere Linienbreite zeigen, ist letztere Erklärung wahrscheinlicher.

Nicht nur für das *Npu* DnaE Intein, sondern auch im Falle des *Ssp* DnaB Inteins, welches an Position 11 gespalten ist, war die Untersuchung der PTS-Reaktion mittels NMR-Spektroskopie möglich. Das Protein *Ssp* DnaB<sup>C</sup>(11)-gpD-H<sub>6</sub> (**7**) konnte mit dem über Festphasenpeptidsynthese hergestellten Peptid Fluorescein-KKESG-CISGDSLISLA (**Pep1**, die Inteinsequenz ist unterstrichen) umgesetzt werden. Jedoch erschwerte die Präzipitation des Proteins bei der Messung definitive Aussagen. Die vorläufigen Daten deuten darauf hin, dass sich nach dem Spleißen auch hier die Inteinhälften in einem gefalteten Komplex befinden (Daten nicht gezeigt).

Obwohl für eine exakte Charakterisierung der ungefalteten Zustände der individuellen Inteinhälften weitere Experimente, wie CD-Spektroskopie oder ANS-Fluoreszenzspektroskopie, notwendig sind, lässt sich doch zusammenfassend festhalten, dass in den hier untersuchten sowohl natürlich als auch künstlich gespaltenen Inteinhälften keine

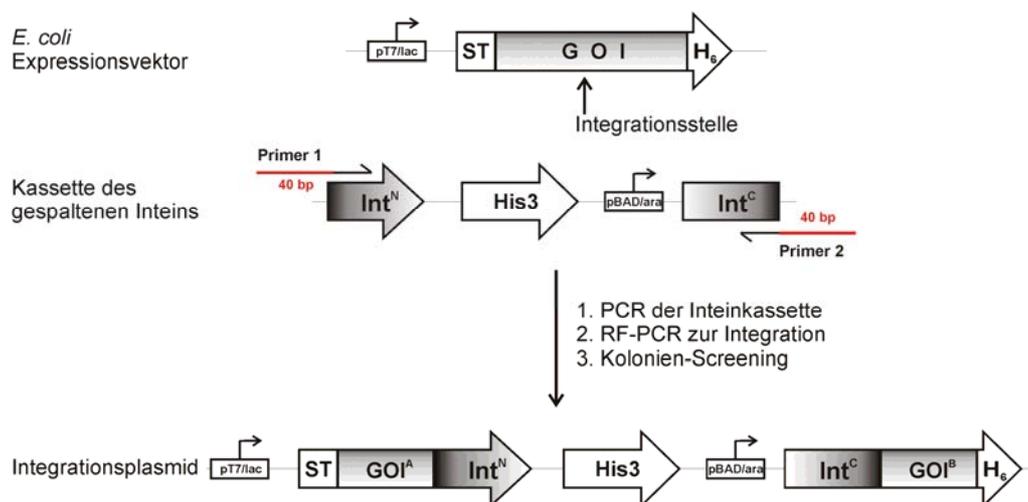
definierte Vorfaltung vorhanden ist. Das Fehlen einer stabilen Konformation ohne das Partnerprotein erklärt möglicherweise auch die Tendenz der individuellen Inteinhälften zur Aggregatbildung (s. Kapitel 4.2.1). Für das *Npu* DnaE Intein konnte weiterhin zum ersten Mal experimentell nachgewiesen werden, dass sich die Inteinhälften bei der Assoziation falten und nach dem Spleißen in einem stabilen Komplex vorliegen. Zu weiterführenden Aussagen über den Faltungsprozess sind jedoch zusätzliche Experimente zumindest aber die Zuordnung der NMR-Signale, notwendig. Diese Untersuchungen konnten aus Zeitgründen im Rahmen dieser Arbeit jedoch nicht mehr durchgeführt werden.

### **4.3 SPLICEFINDER - Schnelle Klonierung und Aktivitätsbestimmung von gespaltenen Inteinen in Zielproteinen**

Für alle biochemischen Anwendungen von gespaltenen Inteinen (s. Kapitel 1.2.4) ist die Spleißaktivität des Inteins an einer Position von Interesse in dem Zielprotein eine notwendige, wenn auch triviale, Voraussetzung. Wie in Kapitel 1.2.2 beschrieben ist ein Intein keine vollkommen autarke Proteindomäne, die unabhängig von seinen zu ligierenden Substraten, den Exteinen, die Spleißreaktion katalysiert. Die +1 Position des C-Exteins ist durch seine nukleophile Seitenkette an der Reaktion beteiligt und findet sich anschließend ebenfalls im Spleißprodukt wieder. Zusätzlich konnten frühere Studien zeigen, dass direkt an die Spleißstelle angrenzenden Aminosäuren (-2, -1, +2, +3 Positionen) die Spleißreaktion signifikant beeinflussen bzw. sogar vollkommen verhindern können [134,151,152,241]. Aktuell sind keine aussagekräftigen Daten vorhanden, die diese Exteinabhängigkeit rational erklären könnten und definitive Vorhersagen über die Spleißaktivität eines bestimmten Inteins an einer definierten Position in einem Zielprotein ermöglichen würden. Für den Anwender von gespaltenen Inteinen bedeutet dies, dass die Position der Inteininsertion und die flankierenden Aminosäuren an der Spleißstelle zum einen an die Sequenz des Zielproteins (möglichst geringfügige Abweichung des Spleißproduktes von der Wildtyp-Sequenz) und zum anderen an die nativen Exteinreste des Inteins (Erhalten der Spleißaktivität) angepasst werden müssen. Zur Identifikation des besten Kompromisses ist das Testen einer Vielzahl an unterschiedlichen Positionen und Kombinationen von flankierenden Aminosäuren oft ein notwendiger Prozess. In Zusammenarbeit mit Tim Sonntag und Simone Eppmann wurde das SPLICEFINDER-System entwickelt, welches die bisher oftmals langwierigen Klonierungsschritte zur Herstellung von verschiedenen Intein-Exteinfusionskonstrukten dramatisch verkürzt und einen schnellen und parallelen Aktivitätstest von unterschiedlichen Insertionspositionen gespalteener Inteine in Zielproteinen erlaubt. Basierend auf dem von Iwaï und Mitarbeitern entwickelten dualen Expressionssystem in *E. coli* [188,189] ist eine mögliche Anwendung des SPLICEFINDER-Systems die segmentelle *in vivo* Isotopenmarkierung von Proteinen für NMR-Untersuchungen. Eine in dieser Doktorarbeit entwickelte Methode zur Überprüfung der Markierungseffizienz erleichtert zusätzlich für diese Applikation die schnelle Bestimmung der optimalen Expressionsbedingung.

### 4.3.1 Prinzip des SPLICEFINDER-Systems

Im Gegensatz zu den klassischen Klonierungsmethoden zur Herstellung von zwei Exteintein Fusionskonstrukten basiert das hier vorgestellte System auf der Integration einer gespaltenen Inteinkassette in das Zielprotein in einem Schritt (s. schematische Darstellung des SPLICEFINDER-Systems in Abbildung 47). Eine vergleichbare Kassettenstrategie wurde erstmals von Tim Sonntag für das gesplattene VMA-Intein aus *S. cerevisiae* entwickelt [242]. Zur späteren gezielten Induktion der N- und C-terminalen Hälften wurden die gut beschriebenen IPTG- und Arabinose-Expressionssysteme verwendet [188,189]. Das zu untersuchende Gen von Interesse kann in einen kommerziell erhältlichen *E. coli* Expressionsvektor mit den pT7/lac Promotor/Operator Sequenzen kloniert werden (in dieser Arbeit wurde der pET16b-Vektor der Firma Novagen benutzt). Die pBAD/ara Kontrollelemente wurden in die drei hergestellten Inteinkassetten, basierend auf dem natürlich gespaltenen *Npu* DnaE Intein, sowie den beiden künstlich gespaltenen *Ssp* DnaB und *Mxe* GyrA Inteinen, integriert. Die Inteinkassetten (s. Tabelle 7) bestehen aus dem N-terminalen Inteinteil, einem HIS3 Markergen und der C-terminalen Inteinhälfte unter der Kontrolle der pBAD/ara Promotor/Operator Sequenzen. Alle drei benutzten gespaltenen Inteine sind aktiv unter nativen Bedingungen und besitzen unterschiedliche Nukleophile an der Position +1 [163,165].



**Abbildung 47 Schematische Darstellung des SPLICEFINDER-Systems.** Die Integration der gespaltenen Inteinkassetten in den Zielvektor ST-GOI-H<sub>6</sub> kann über eine „restriction-free“ (RF)-PCR und anschließendem Kolonien-Screening erfolgen. Die Aminosäuren an der Spleißstelle können über die Primer der ersten PCR frei gewählt werden und in dem entstandenen Integrationsplasmid sind die 5'- und 3'-Inteinfusionsgene unter der Kontrolle der pT7/lac und pBAD/ara Promotor/Operator Sequenzen. Ein einfacher Aktivitätstest der Insertionsposition mit den jeweiligen flankierenden Aminosäuren ist über eine Testexpression mit anschließender Western Blot Analyse möglich (GOI = engl. gene of interest).

Die Integration der Inteinkassette in das Gen von Interesse (am besten fusioniert mit unterschiedlichen 5'- und 3'-Affinitätstags) benötigt drei Klonierungsoperationen. Die erste PCR verlängert die Inteinkassette auf beiden Seiten um je 40 zur Integrationsposition in dem Gen von Interesse homologe Basenpaare. Durch unterschiedliche Oligonukleotide kann in diesem Schritt eine Vielzahl an verschiedenen Positionen, sowie flankierenden Aminosäuren an der Spleißstelle parallel generiert werden. Der nächste Schritt ist eine „restriction-free“ PCR mit der Inteinkassette als Insert und dem *E. coli* Expressionsvektor als Templat [215,243]. Nach *DpnI*-Verdau, Transformation in *E. coli* und einem Kolonien Screening wird das gewünschte Integrationsplasmid erhalten. Zusammen mit einem Helferplasmid, welches für die regulatorischen Proteine LacI und AraC kodiert, kann die Expression des N-terminalen Intein-Exteinfusionskonstrukt mit IPTG und die des C-terminalen durch Arabinose induziert werden. Zum Testen der PTS-Aktivität ist eine Expression im kleinen Maßstab (3 - 4 mL) mit anschließender Western Blot Analyse gegen die Affinitätstags des Spleißproduktes ausreichend.

**Tabelle 7 Übersicht über die in dieser Arbeit hergestellten gespaltenen Inteinkassetten.**

Kassette	Int <sup>N</sup> -Sequenz	Int <sup>C</sup> -Sequenz	+1 Nukleophil	Literatur
<i>Ssp</i> DnaB	<i>Ssp</i> DnaB(1-104)	DnaB(107-154)	Serin	[163]
<i>Npu</i> DnaE	<i>Npu</i> DnaE(1-102)	DnaE(103-138)	Cystein	[151]
<i>Mxe</i> GyrA	<i>Mxe</i> GyrA(1-119)-FKBP	gpD-GyrA(120-199)	Threonin	[165]

Abkürzungen: FKBP = FK506 bindendes Protein; gpD = Kopfprotein D des Bakteriophagen  $\lambda$ .

### 4.3.2 Das Modellprotein gpD-Trx

Die erste Machbarkeitsstudie der SPLICEFINDER-Methode wurde an einem Protein aus zwei gut löslichen und gefalteten Domänen, dem gpD-Protein und Thioredoxin (Trx), fusioniert über sechs Glycinreste durchgeführt. Am Beispiel dieses Proteins konnte gleichzeitig die segmentelle Isotopenmarkierung für NMR-Untersuchungen getestet werden. Eine Dreifragmentligation ergab den für das Protein ST-gpD-Trx-H<sub>6</sub> (9) kodierenden, pET16b-Ausgangsvektor (pT7/lac Promotor/Operator Sequenz, Ampicillinresistenz, für Details s. Kapitel 3.1) Die Integration sowohl der *Ssp* DnaB, als auch der *Npu* DnaE Inteinkassette in den Glycinlinkerbereich von ST-gpD-Trx-H<sub>6</sub> wurde mit jeweils vier verschiedenen Kombinationen an flankierenden Aminosäuren realisiert. Die Primer für die

Amplifikation der Inteinkassetten wurden so gewählt, dass in einer Kombination das mögliche Spleißprodukt nur eine Insertion des +1 Nukleophils im Vergleich zum „Wildtyp“ Protein 9 enthalten würde (Kombinationen K1 und K5). In den anderen Fällen wurden jeweils die drei Aminosäuren N- und/oder C-terminal an die natürlichen Exteinsequenzen der Inteine angepasst. Die in dieser Arbeit hergestellten Plasmide, sowie der prozentuale Erfolg der RF-PCR Klonierung sind zusammenfassend in Tabelle 8 dargestellt.

**Tabelle 8 Übersicht über die in dieser Arbeit hergestellten Integrationsplasmide mit dem Modellgen ST-gpD-Trx-H<sub>6</sub>.**

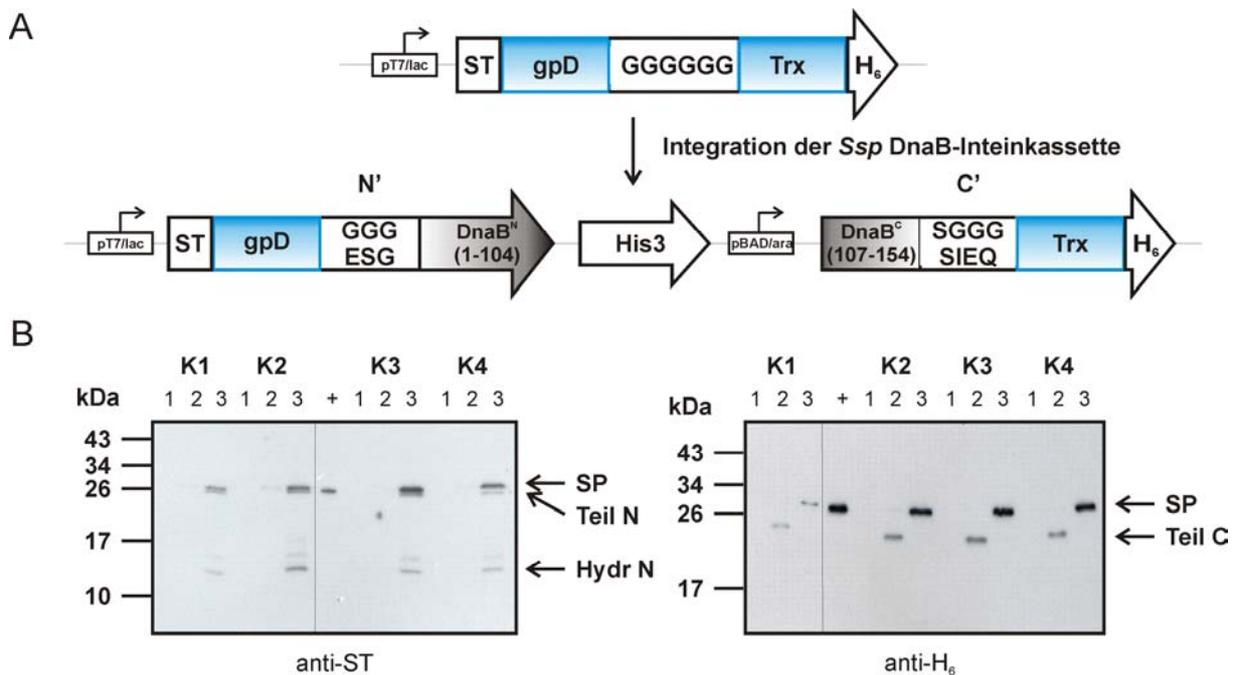
Kombination	Verwendete Inteinkassette	N-terminale AS	C-terminale AS	Untersuchte Kolonien	Positive Kolonien	Prozentualer Erfolg /%
ST-gpD-Trx-H <sub>6</sub>		GGG	GGG			
K1	<i>Ssp</i> DnaB	GGG	SGGG	3	1	33,3
K2	<i>Ssp</i> DnaB	GGG	SIEQ	3	3	100
K3	<i>Ssp</i> DnaB	ESG	SGGG	11	1	9,1
K4	<i>Ssp</i> DnaB	ESG	SIEQ	3	2	66,7
K5	<i>Npu</i> DnaE	GGG	CGGG	3	3	100
K6	<i>Npu</i> DnaE	GGG	CFNK	3	2	66,7
K7	<i>Npu</i> DnaE	AEY	CGGG	3	3	100
K8	<i>Npu</i> DnaE	AEY	CFNK	3	3	100
				<b>Σ 32</b>	<b>Σ 18</b>	<b>56,3</b>

Die Erfolgsquote der auf RF-PCR basierenden Klonierung ist zusammen mit den unterschiedlichen Aminosäurekombinationen an der Spleißstelle und den verwendeten Inteinkassetten dargestellt.

Ein Gesamtprozentsatz von über 50% positiver Kolonien deutet darauf hin, dass die RF-PCR auch unter den hier gewählten Bedingungen (die Inteinkassette ist ~ 2,5 kbp groß und es wird nur eine geringe Anzahl an bp im Templat ausgetauscht) eine stabile Klonierungsmethode darstellt [215,243]. Alternativ erlauben die Inteinkassetten auch eine Integration in einen Hefeshuttle-Vektor via homologe Rekombination in Hefe [244] - auf die in dieser Arbeit jedoch nicht detailliert eingegangen werden soll.

Zur Überprüfung, ob die gewählten Extein-Intein Kombinationen spleißaktiv sind, wurden *E. coli*-Zellen mit den jeweiligen Integrationsplasmiden (pET16b, Ampicillin-Resistenz) und einem Helferplasmid kotransformiert. Dieses Helferplasmid basierte auf einem kommerziell erhältlichen RSFDuet-Vektor der Firma Novagen (Kanamycin-Resistenz) und enthielt die für die regulatorischen Proteine LacI und AraC kodierenden Bereiche. Über die Selektion mit Ampicillin und Kanamycin konnte sichergestellt werden, dass beide Plasmide in den *E. coli*-Zellen vorhanden waren. Mit allen acht Integrationskombinationen wurden Testexpressionen

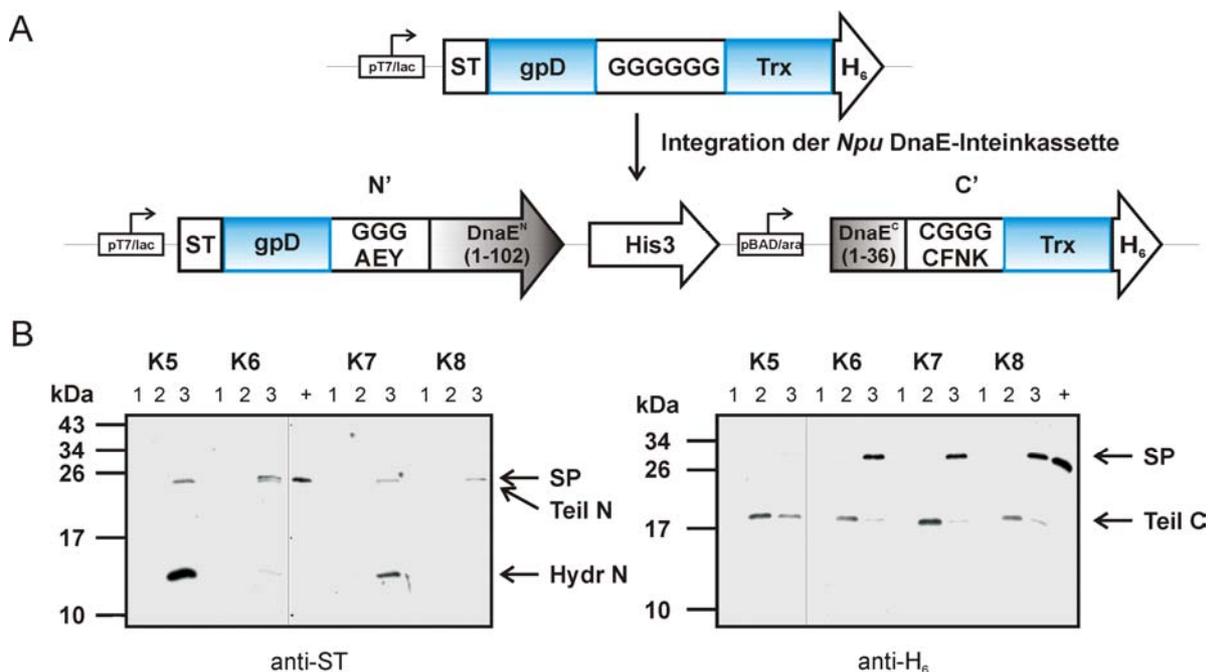
im 4 mL Maßstab durchgeführt. Dafür wurden die Kulturen bei einer OD von  $\sim 0,6$  zuerst mit 0,2% Arabinose für 2h bei 37°C induziert. Nach einem Waschschrift zur Entfernung von möglicherweise überschüssiger Arabinose wurden die *Ssp* DnaB Kombinationen bei 25°C und die *Npu* DnaE Kombinationen bei 37°C mit 1,2 mM IPTG für zusätzliche 4,5h inkubiert. Die Ergebnisse der Western Blot Analysen sind zusammen mit den schematischen Darstellungen der Integrationen in Abbildung 48 und in Abbildung 49 gezeigt. Eine selektive Induktion des C-terminalen Fusionsproteins mit Arabinose und eine anschließende Produktion der N-terminalen Hälfte durch die Zugabe von IPTG fanden in allen acht Kombinationen statt (s. Spuren 2 und 3 in Abbildung 48 und in Abbildung 49). Für die vier Kombinationen des *Ssp* DnaB Inteins konnte in allen Fällen die Bildung des Spleißproduktes ST-gpD-Trx-H<sub>6</sub> in beiden Western Blots nachgewiesen werden. Ebenfalls war das Auftreten von N-terminalen Hydrolysebanden im anti-ST Western Blot unabhängig von den flankierenden Aminosäuren an der Spleißstelle. Das *Ssp* DnaB Intein zeigt somit eine hohe Toleranz für die Insertion in glycin-reiche Linkerregionen.



**Abbildung 48 Integration der *Ssp* DnaB Inteinkassette in das Modellgen ST-gpD-Trx-H<sub>6</sub> (Kombinationen 1 - 4).** A) Schematische Darstellung der Integration mit den Aminosäure-Kombinationen an der N-terminalen (N') und der C-terminalen (C') Spleißstelle; B) Western Blot Analyse gegen den StrepTag II (links) und den H<sub>6</sub>-Tag (rechts) der Expressionen im kleinen Maßstab; Spur 1 = Probe vor Induktion; Spur 2 = Probe nach Arabinose-Induktion; Spur 3 = Probe nach anschließender IPTG-Induktion; Spur + = Positivkontrolle; die theoretischen Molekulargewichte der Proteine sind wie folgt: Spleißprodukt (SP)/Positivkontrolle = 26,4 - 26,8 kDa; Teil N = 24,9 - 25,1 kDa; Teil C = 18,8 - 19,0 kDa; N-terminale Hydrolyse (Hydr N) = 13,2 - 13,3 kDa.

Im Falle des *Npu* DnaE Inteins war eine größere Abhängigkeit von den fusionierten Exteinsequenzen zu beobachten. Bei Kombination 5, mit drei Glycinresten an beiden Seiten

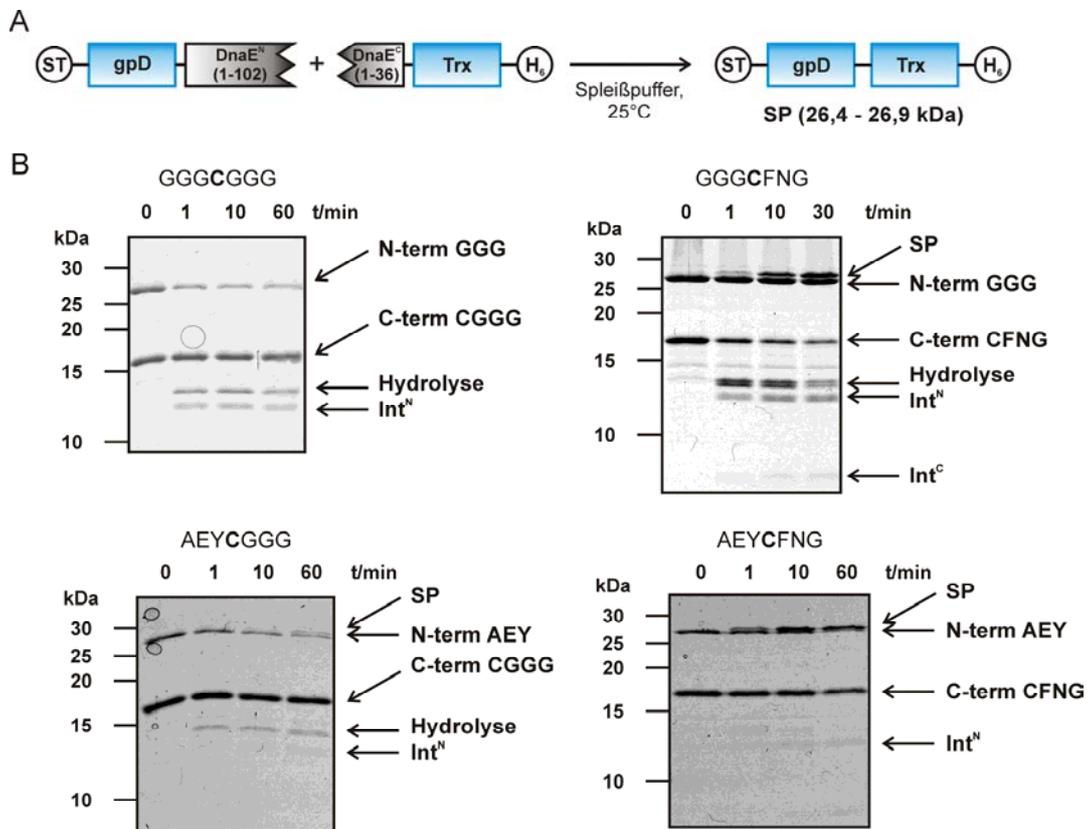
der Spleißstelle, konnte keine Spleißproduktbildung, jedoch das vermehrte Auftreten einer N-terminalen Hydrolysebande beobachtet werden. Diese Exteinminosäuren ermöglichen also die Komplexbildung, sowie den ersten N-S Acylshift zur Bildung eines Thioesters, blockieren die Spleißreaktion aber in einem der folgenden Reaktionsschritte. Interessanterweise führt das Anpassen der N-terminalen Aminosäuren an die nativen Exteinsequenzen (AEY) zwar immer noch zur N-terminalen Hydrolyse, jedoch konnte in dieser Kombination 7 auch Spleißprodukt (AEYCGGG) in beiden Western Blots nachgewiesen werden. Die Kombinationen 6 und 8 mit angepassten C-terminalen Aminosäuren zeigten Spleißaktivität ohne die Bildung von N-terminalen Hydrolyseprodukten. Frühere Studien hatten postuliert, dass zur Aktivität der DnaE Inteinallele die Peptidsequenz CFN an der C-terminalen Spleißstelle essentiell sei [245]. Die hier gefundenen, teilweise widersprechenden Ergebnisse (s. Kombination 7) bestätigen erneut, dass *ab initio* Aussagen zur Spleißfähigkeit eines Inteins in einem bestimmten Kontext nur schwer möglich sind.



**Abbildung 49 Integration der *Npu* DnaE Inteinkassette in das Modellgen ST-gpD-Trx-H<sub>6</sub> (Kombinationen 5 - 8).** A) Schematische Darstellung der Integration mit den Aminosäure-Kombinationen an der N-terminalen (N') und der C-terminalen (C') Spleißstelle; B) Western Blot Analyse gegen den StrepTag II (links) und den H<sub>6</sub>-Tag (rechts) der Expressionen im kleinen Maßstab; Spur 1 = Probe vor Induktion; Spur 2 = Probe nach Arabinose-Induktion; Spur 3 = Probe nach anschließender IPTG-Induktion; Spur + = Positivkontrolle; die theoretischen Molekulargewichte der Proteine sind wie folgt: Spleißprodukt (SP)/Positivkontrolle = 26,4 - 26,9 kDa; Teil N = 25,0 - 25,2 kDa; Teil C = 17,4 - 17,7 kDa; N-terminale Hydrolyse (Hydr N) = 13,2 - 13,4 kDa.

Der Frage, ob die Ergebnisse der PTS-Aktivitätstest *in vivo* auch auf gereinigte Proteine *in vitro* übertragbar sind, wurde am Beispiel des *Npu* DnaE Inteins nachgegangen. Die Inkubationen der individuellen N- und C-terminalen Inteininfusionsproteine bestätigten

qualitativ die mittels Western Blot Analyse erhaltenen Resultate (s. Abbildung 50). So war z.B. die Kombination der Proteine ST-gpD-GGG-Int<sup>N</sup> und Int<sup>C</sup>-CGGG-Trx-H<sub>6</sub> spleißinaktiv und zeigte in Übereinstimmung mit den *in vivo* Ergebnissen nur N-terminale Hydrolyse.



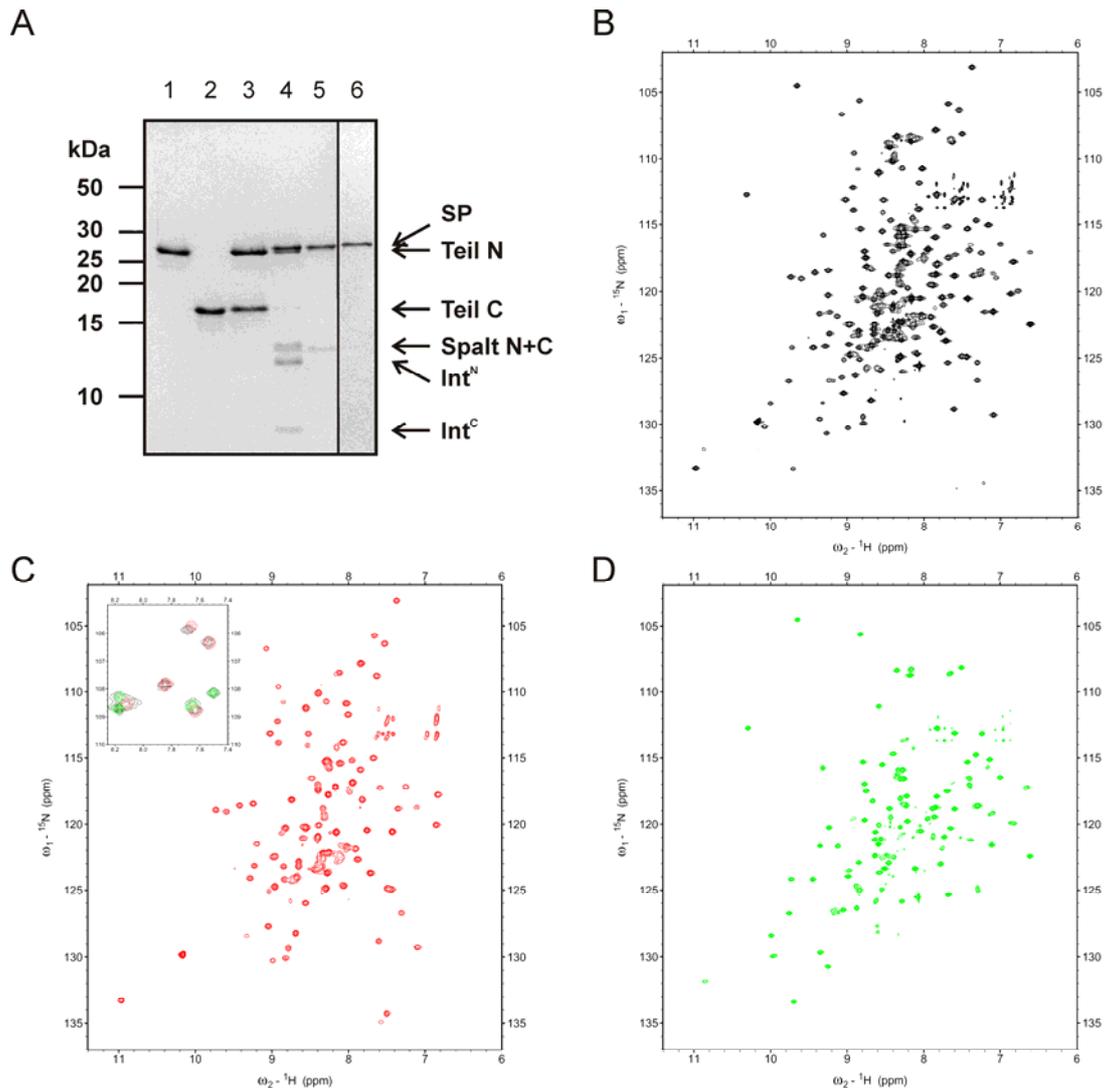
**Abbildung 50** *In vitro* PTS-Reaktionen zur Bildung des Modellproteins ST-gpD-Trx-H<sub>6</sub>. A) Schematische Darstellung der Reaktion; B) SDS-PAGE Gele der vier verschiedenen Kombinationen an N- und C-terminalen Inteininfusionsproteinen; die Spleißprodukte (SP), sowie die jeweiligen Protein- und Inteinhälften sind beschriftet. Die *in vitro* PTS-Reaktionen bestätigen die, über *in vivo* Expressionen im kleinen Maßstab und anschließender Western Blot Analyse erhaltenen, Ergebnisse. Die Kombination GGG-CGCG ist spleißinaktiv (oben, links). GGG-CFNK (oben, rechts), AEY-CGCG (unten, links) und AEY-CFNK (unten, rechts) zeigen Spleißproduktbildung und sind damit aktiv.

Schließlich wurde das Beispielprotein **9** im großen Maßstab segmentell <sup>15</sup>N-isotopenmarkiert und mittels NMR-Spektroskopie untersucht. Zur Isotopenanreicherung des C-terminalen Trx-Teils (ST-gpD-<sup>15</sup>N(Trx-H<sub>6</sub>)) kam das *in vitro* PTS mit dem *Npu* DnaE Intein und gereinigten Proteinen zum Einsatz. Die C-terminale Hälfte Int<sup>C</sup>-CFNK-Trx-H<sub>6</sub> wurde in M9 Minimal Medium mit <sup>15</sup>N-Ammoniumchlorid als einziger Stickstoffquelle separat von dem N-terminalen Fusionsprotein ST-gpD-GGG-Int<sup>N</sup> (Expression in LB-Medium) synthetisiert und gereinigt. Die PTS-Reaktion der beiden Proteine, sowie die anschließenden Schritte zur Reinigung des Spleißproduktes über die N- und C-terminalen Affinitätstags sind in Abbildung 51A gezeigt. Das Verhältnis der beiden Proteine wurde so gewählt, dass die N-terminale, unmarkierte Komponente im Überschuss (1,5 Äquivalente)

vorlag, um einen vollständigen Umsatz des markierten C-terminalen Proteins zu gewährleisten (s. Spuren 3 und 4 in Abbildung 51A). Das nach der ersten Ni<sup>2+</sup>-NTA Affinitätschromatographie ko-eluierte C-terminale Spaltungsprodukt (Spalt C) wurde über eine weitere StrepTag II Affinitätschromatographie abgetrennt (s. Spuren 5 und 6 in Abbildung 51A). Eine detaillierte Analyse (s. Kapitel 4.3.3) ergab für das Spleißprodukt eine Anreicherung der <sup>15</sup>N-Isotope im N-terminalen Teil von unter 1% und im C-terminalen Teil von 95% (s. Abbildung 80A im Anhang).

Die Isotopenanreicherung im N-terminalen gpD-Teil (<sup>15</sup>N(ST-gpD)-Trx-H<sub>6</sub>) gelang über eine *in vivo* Koexpression und PTS mit dem *Ssp* DnaB Intein (Kombination 3 - Spleißprodukt ESGSGGG) in *E. coli*. Nach Induktion der C-terminalen Hälfte in LB-Medium und anschließendem Mediumswechsel zu M9 Minimal Medium mit <sup>15</sup>N-Ammoniumchlorid als einziger Stickstoffquelle erfolgte die Synthese des N-terminalen Fusionsproteins durch Induktion mit IPTG (für Details s. Kapitel 3.5.3). Die Reinigung des entstandenen Spleißproduktes gelang in einem Schritt durch Ni<sup>2+</sup>-NTA Affinitätschromatographie, da die C-terminale Inteinhälfte bei den gewählten Expressionsbedingungen komplett durch die PTS-Reaktion verbraucht wurde. Der Anteil der C-terminalen Spaltungsreaktion war gering, da nur Spuren des Nebenproduktes Trx-H<sub>6</sub> mitgereinigt wurden (Daten nicht gezeigt). Eine detaillierte Analyse (s. Kapitel 4.3.3) ergab für das Spleißprodukt eine Anreicherung der <sup>15</sup>N-Isotope im N-terminalen Teil von 83% und im C-terminalen Teil von unter 1% (s. Abbildung 79A im Anhang).

Die beiden isotopenmarkierten Spleißprodukte wurden analog zu dem komplett <sup>15</sup>N-markierten Kontrollprotein (ST-gpD-GGGGGG-Trx-H<sub>6</sub>) gegen NMR-Puffer (25 mM Tris, 50 mM NaCl, 2 mM DTT, pH 7,0) dialysiert und ankonzentriert. Die von Alena Busche und Dr. Frank Löhr aufgenommenen <sup>15</sup>N, <sup>1</sup>H-TROSY Spektren aller drei Proteine sind in Abbildung 51B-D dargestellt. Vor allem in der Vergrößerung von Abbildung 51C ist schön zu erkennen, dass eine segmentelle Isotopenmarkierung die Komplexität der Spektren reduziert und durch eine geringere Signalüberlappung eine leichtere Zuordnung der Signale ermöglichen würde.



**Abbildung 51 Segmentelle Isotopenmarkierung des Beispielproteins ST-gpD-Trx-H<sub>6</sub>.** A) *In vitro* PTS mit dem *Npu* DnaE Intein und präparative Reinigung des segmentell markierten Spleißprodukts ST-gpD-<sup>15</sup>N(Trx-H<sub>6</sub>); gezeigt ist das SDS-PAGE Gel in der Coomassie-Färbung; Spur 1 = gereinigtes N-terminales Konstrukt (ST-gpD-GGG-Int<sup>N</sup>); Spur 2 = gereinigtes C-terminales Konstrukt <sup>15</sup>N(Int<sup>C</sup>-CFNK-Trx-H<sub>6</sub>); Spur 3 = Nullstundenprobe der PTS-Reaktion; Spur 4 = PTS-Reaktion nach 16h; Spur 5 = Elutionsfraktionen der Ni<sup>2+</sup>-NTA Affinitätschromatographie; Spur 6 = Elutionsfraktionen der StrepTag II Affinitätschromatographie; die theoretischen Molekulargewichte der Proteine sind wie folgt: **SP** (26.7 kDa); **Teil N** (25.0 kDa); **Teil C** (17.6 kDa); **Spalt N** (13.2 kDa); **Spalt C** (13.5 kDa); **Int<sup>N</sup>** (11.9 kDa); **Int<sup>C</sup>** (4.1 kDa); B)-D) <sup>15</sup>N, <sup>1</sup>H-TROSY Spektren von dem komplett markierten Protein <sup>15</sup>N(ST-gpD-Trx-H<sub>6</sub>) (schwarz, B), sowie von den segmentell markierten Proteinen <sup>15</sup>N(ST-gpD)-Trx-H<sub>6</sub>) (rot, C) und ST-gpD-<sup>15</sup>N(Trx-H<sub>6</sub>) (grün, D); die Vergrößerung in C) zeigt eine Überlagerung der drei Spektren und demonstriert die Reduktion der Signalüberlappung durch segmentelle Isotopenmarkierung.

### 4.3.3 Überprüfung der segmentellen Isotopenmarkierung bei *in vivo* Expressionen

Das von Iwai und Mitarbeitern entwickelte und in dieser Arbeit verwendete *in vivo* Expressionssystem [188,189] ermöglicht eine deutlich vereinfachte Darstellung von segmentell markierten Spleißprodukten im Vergleich zu einer *in vitro* PTS-Reaktion. Dem

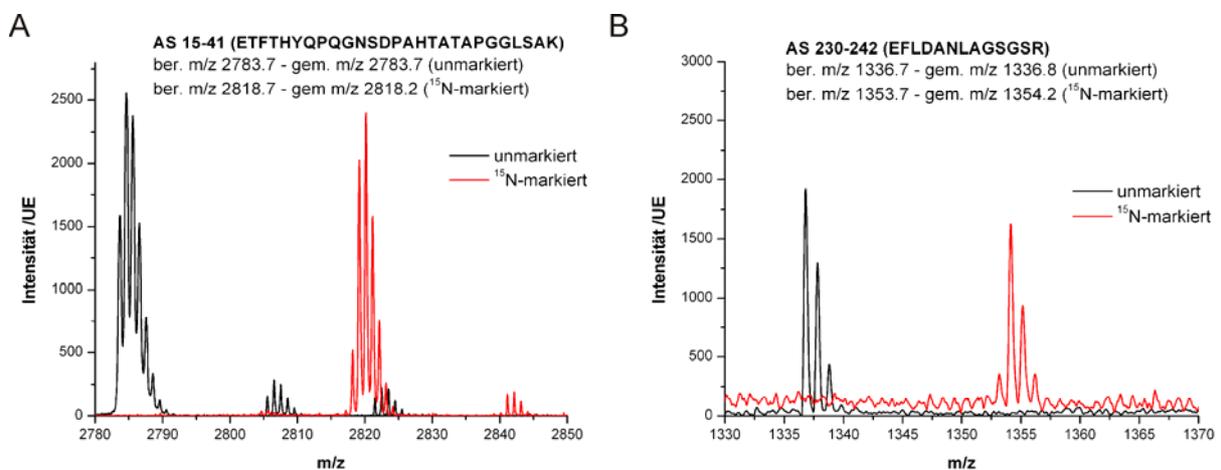
Vorteil der Ein-Schritt-Synthese und Reinigung steht jedoch der Nachteil des so genannten „isotope scrambling“ gegenüber. Hierbei soll angedeutet werden, dass die NMR-aktiven Isotope nicht nur in dem gewünschten Segment angereichert sind, sondern vielmehr über das gesamte Protein „verrührt“ vorliegen. Die Voraussetzung für eine erfolgreiche segmentelle Isotopenmarkierung ist eine hoch selektive Synthese der Proteinhälften in den unterschiedlichen Medien. Eine effektive Kontrolle der Genexpression durch die Promotor/Operator-Sequenzen und den jeweiligen Repressorproteinen kann von den Expressionsbedingungen, z.B. der Induktionstemperatur und -zeit, sowie der Konzentration des Induktionsmittels, abhängig sein. Oftmals müssen hier, auch im Hinblick auf die Ausbeute an segmentell markiertem Spleißprodukt, eine Reihe an verschiedenen Parametern getestet werden, um die optimale Expressionsbedingung zu ermitteln. Bisherige Studien bestimmten die Markierungseffizienz entweder über die Aufnahme eines NMR-Spektrums [185,188,189] oder durch die Massenanalyse des kompletten Proteins [187,189]. Die erste Methode ermöglicht eine exakte Bestimmung, erfordert jedoch einen sehr hohen experimentellen Aufwand. Die zweite Methode kann die Massendifferenz durch den Einbau der schwereren Isotope im Vergleich zu einer unmarkierten Probe nachweisen, liefert aber keine Hinweise darauf, in welchem Teil des Proteins die NMR-aktiven Kerne angereichert sind. Zur vereinfachten Analyse von verschiedenen Expressionsbedingungen wurde im Laufe dieser Doktorarbeit eine Strategie entwickelt, die auf Expressionen im kleinen Maßstab und tryptischem in-Gelverdau der Spleißproduktbande mit anschließender MALDI-TOF MS beruht. Sie ermöglicht eine exakte Bestimmung der Markierungseffizienz mit geringem Arbeitsaufwand. Das Prinzip basiert darauf, dass durch Zerlegung des Spleißproduktes in kleinere tryptische Peptide der Einbau der schwereren NMR-aktiven Isotope nicht für das gesamte Protein, sondern für die individuellen Peptide nachgewiesen werden kann. Üblicherweise können, abhängig von der Ionisierungswahrscheinlichkeit, nicht alle theoretischen tryptischen Peptide in den MALDI-TOF MS Spektren detektiert werden. Jedoch ist der Nachweis eines einzelnen Peptidfragments pro Proteinsegment ausreichend um die Markierungseffizienz zu bestimmen. Zum Vergleich mit den segmentell markierten Spleißprodukten werden zwei Referenzproteine, ein komplett Unmarkiertes und ein vollständig  $^{15}\text{N}$ -Markiertes, benötigt. Die Markierungseffizienz in einem segmentell markierten Spleißprodukt wird über das mittlere Molekulargewicht eines tryptischen Peptidfragments ( $MG_{\text{Peptid}}$ ) bestimmt. Dies berechnet sich aus der Summe des Produktes aus Intensität ( $I_i$ ) und korrespondierender Masse ( $m_i$ ) aller Basislinien-getrennten Isotopensignale geteilt durch die Gesamtintensität des Peptidsignals (s. Gleichung 8).

$$MG_{\text{Peptid}} = \frac{\sum_i \text{Zahl der Isotopensignale } I_i * m_i}{\sum_i \text{Zahl der Isotopensignale } I_i} \quad (\text{Gl. 8})$$

Der prozentuale Isotopeneinbau für das zu dem Peptid gehörende Proteinsegment berechnet sich anschließend über Gleichung 9. Dabei sind  $MG_{\text{markiert}}$  und  $MG_{\text{unmarkiert}}$  die mittleren Molekulargewichte der tryptischen Peptidfragmente der  $^{15}\text{N}$ -markierten und der unmarkierten Referenzproben.

$$\% \text{ Markierung} = \frac{MG_{\text{Peptid}} - MG_{\text{unmarkiert}}}{MG_{\text{markiert}} - MG_{\text{unmarkiert}}} * 100\% \quad (\text{Gl. 9})$$

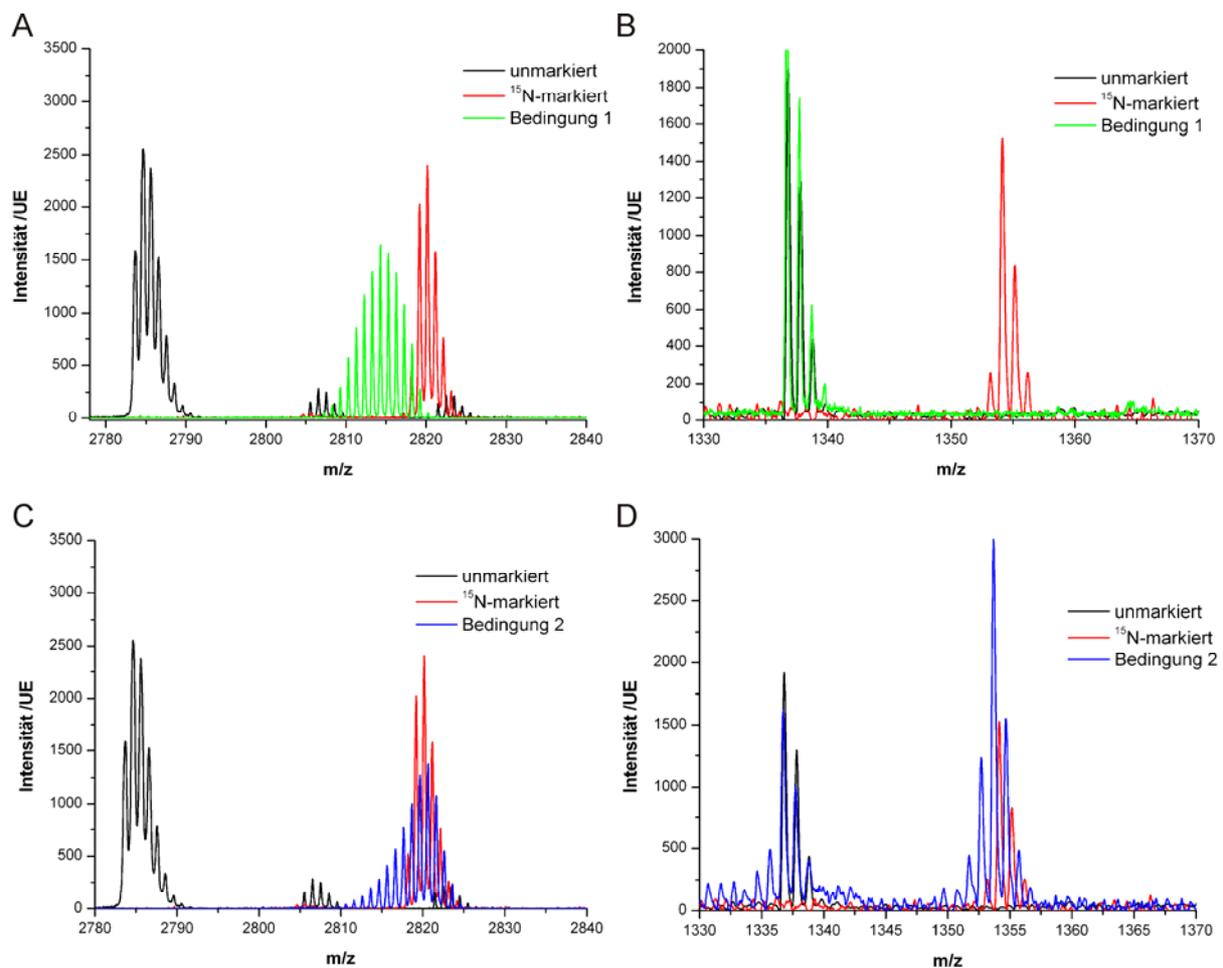
An zwei Expressionsbedingungen des *in vivo* Spleißen von Kombination 3 (*Ssp* DnaB Intein - Spleißprodukt ST-gpD-ESGSGGG-Trx-H<sub>6</sub>, s. Kapitel 4.3.2) soll nun die Methode exemplarisch vorgestellt werden. Nach Koexpression der beiden Hälften, Spleißproduktbildung *in vivo*,  $\text{Ni}^{2+}$ -NTA Chromatographie und anschließender SDS-PAGE konnte die Spleißproduktbande tryptisch in-Gel verdaut werden. Der zusätzliche Reinigungsschritt war nur notwendig, da das Spleißprodukt und die N-terminale Hälfte ähnliche elektrophoretische Beweglichkeiten im SDS-PAGE Gel aufwiesen und nicht eindeutig voneinander getrennt werden konnten. In den MS Spektren der Referenzproteine konnte erwartungsgemäß eine Verschiebung der Peptidsignale abhängig von der Anzahl der Stickstoffatome festgestellt werden (s. Abbildung 52). Für das N-terminale gpD Segment wurde das tryptische Peptidfragment AS 15-41 mit 35 N-Atomen zur Analyse herangezogen. Das C-terminale Trx Segment wurde mit den Aminosäuren 230-242 (17 N-Atome) untersucht (s. Abbildung 52).



**Abbildung 52 MALDI-TOF MS Spektren der Referenzproteine von ST-gpD-Trx-H<sub>6</sub>.** Die Spektren der tryptischen in-Gelverdaue des unmarkierten (schwarz) und des komplett in  $^{15}\text{N}$ -haltigem Medium synthetisierten

Proteins (rot) sind dargestellt; die Ausschnitte zeigen ein N-terminales gpD-Fragment (A) und ein C-terminales Trx-Fragment (B).

Die untersuchten Expressionsbedingungen waren wie folgt: Bei Bedingung 1 wurden die mit dem Plasmid von Kombination 3 und dem Helferplasmid transformierten *E. coli* BL21 Gold (DE3) Zellen in LB-Medium bis zu einer OD von 0,6 bei 37°C inkubiert. Dann erfolgte die Induktion der C-terminalen Hälfte mit 0,2% Arabinose für 2h bei weiterhin 37°C in LB-Medium. Nach einem Waschschrift wurde die Synthese der N-terminalen Hälfte mit 1,2 mM IPTG bei 25°C in <sup>15</sup>N-haltigem Minimalmedium für 4h induziert. Bei der zweiten Expressionsbedingung wurde im Vergleich zu Bedingung 1 nur die Dauer der zweiten Induktion von 4h auf 16h verlängert. Die MALDI-TOF MS Spektren der beiden Proben sind in Abbildung 53 dargestellt. Eine detaillierte Auswertung der MS Spektren von Bedingung 1 ergab, dass das Spleißprodukt eine Isotopenanreicherung von 80% im N-terminalen gpD-Teil (s. Abbildung 53A) und von unter 1% im C-terminalen Trx-Teil (s. Abbildung 53B) aufwies. Für <sup>15</sup>N-<sup>1</sup>H Korrelations-Untersuchungen des gpD-Teils sind dies damit optimale Expressionsbedingungen, da das Thioredoxin vollkommen „NMR-unsichtbar“ ist. Es bleibt jedoch anzumerken, dass in 3D-NMR Techniken, die die Verknüpfung dreier isotopenangereicherter Kerne benötigen, die Signalintensität um 49% (0,8<sup>3</sup>) im Vergleich mit einer vollständig markierten Probe reduziert wäre. Die unvollständige Markierung im N-terminalen Teil resultiert wahrscheinlich aus dem Vorhandensein einer großen Anzahl an unmarkierten Vorläuferaminosäuren, die sich während den Inkubationszeiten im LB-Medium gebildet haben. In Übereinstimmung damit erhöhte sich die Isotopenanreicherung im N-terminalen Teil bei einer Verlängerung der Induktionszeit im <sup>15</sup>N-haltigen Minimalmedium auf 97% (Bedingung 2, s. Abbildung 53C). Im Vergleich mit Bedingung 1 konnte jedoch bei dieser IPTG-Induktion kein vollständiges Ausschalten der Synthese des C-terminalen Teils gewährleistet werden. Das MALDI-TOF MS Spektrum des tryptischen in-Gelverdaus des Spleißproduktes (s. Abbildung 53D) zeigt zwei Fraktionen des Trx-Peptids. Die Synthese einer Fraktion (~ 39%) erfolgte wie gewünscht bei der ersten Induktion in LB-Medium, die Zweite (~ 61%) wurde in <sup>15</sup>N-haltigen Minimalmedium synthetisiert. Eine längere Inkubation in Minimalmedium führt wahrscheinlich durch niedrige Glukose- und hohe 3'-5'-zyklisches AMP (cAMP)-Konzentrationen zu einer Verminderung der Katabolitrepression [246,247] und damit zur unerwünschten Synthese des C-terminalen Inteinproteins.



**Abbildung 53** Untersuchung von verschiedenen Expressionsbedingungen zur segmentellen Isotopenmarkierung von ST-gpD-Trx-H<sub>6</sub> *in vivo*. Dargestellt sind Ausschnitte aus den MALDI-TOF MS Spektren der tryptischen in-Gelverdaue des gebildeten Spleißprodukts von Bedingung 1 (4h IPTG-Induktion, grünes Spektrum, A (N-terminales Fragment) und B (C-terminales Fragment)), sowie von Bedingung 2 (16h IPTG-Induktion, blaues Spektrum, C (N-terminales Fragment) und D (C-terminales Fragment)); zum Vergleich sind jeweils die Spektren des unmarkierten (schwarz) und des vollständig <sup>15</sup>N-markierten (rot) Referenzproteins gezeigt.

Die hier präsentierte Methode erlaubt somit die exakte Analyse einer Expressionsbedingung mit geringem Aufwand und die Möglichkeit des schnellen Testens einer Vielzahl an unterschiedlichen Expressionsbedingungen.

Zusammenfassend lässt sich sagen, dass das SPLICEFINDER-System ein schnelles, paralleles Testen von verschiedenen Insertionspositionen von gespaltenen Inteinen in Zielproteinen auf PTS-Aktivität ermöglicht. Für das natürlich gespaltene *Npu* DnaE Intein und das künstlich gespaltene *Ssp* DnaB Intein konnte die Methode eindeutig experimentell bestätigt werden. Für das künstlich gespaltene *Mxe* GyrA Intein wurden keine Integrationen in Zielproteine, sondern nur die Aktivität der Inteinkassette selbst überprüft (s. Abbildung 81A im Anhang). Damit stehen drei verschiedene Inteinkassetten mit unterschiedlichen +1 Nukleophilen (Cystein, Serin und Threonin) zur Verfügung. Die Aussagen der

Aktivitätsüberprüfung mittels Expressionen im kleinen Maßstab und Western Blot Analyse sollten sich auch auf *in vitro* PTS-Anwendungen übertragen lassen, solange keine Unlöslichkeitsprobleme der Inteinfusionproteine auftreten. Die Machbarkeit der segmentellen Isotopenmarkierung von Proteinen für NMR-Untersuchungen konnte am Beispielprotein ST-gpD-Trx-H<sub>6</sub> (9) für das SPLICEFINDER-System nachgewiesen werden. Ebenfalls wurde eine vereinfachte Methode zur Bestimmung der Markierungseffizienz bei *in vivo* Koexpression der Inteinhälften präsentiert. Weiterhin wurde die SPLICEFINDER-Methode an zwei Beispielproteinen, einem kompletten NRPS-Modul (GrsB1 C-A-PCP, Integration der *Ssp* DnaB Intein-kassette zwischen die A- und die PCP-Domäne) und einer Uroporphyrinogen III Methyltransferase (CobA, Integration der *Npu* DnaE Intein-kassette innerhalb der Proteindomäne), erfolgreich angewendet (Zettler, J. *et al.* unpublizierte Daten).

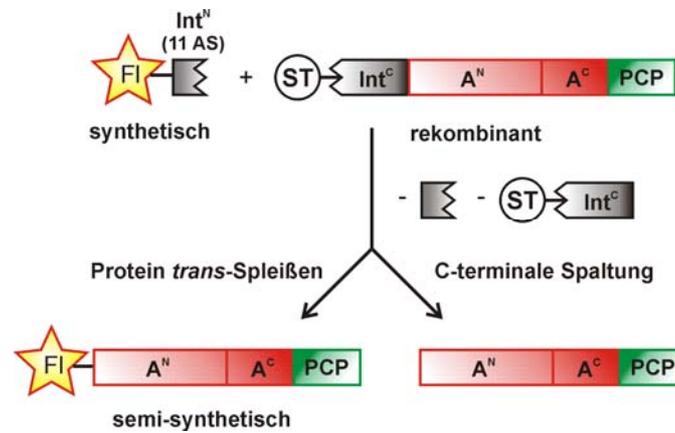
## **4.4 Vorarbeiten zu spektroskopischen Untersuchungen von NRPS-Proteinen**

Wie in Kapitel 4.1 beschrieben sind zum detaillierten Verständnis der konformationellen Änderungen in den NRPS-Multidomänenproteinen die verwendeten biochemischen Untersuchungen nicht ausreichend. Spektroskopische Techniken, wie z. B. die Einzelmolekülfluoreszenz- und die NMR-Spektroskopie könnten eine strukturelle Beschreibung des Proteins mit atomarer Auflösung und eine exakte Charakterisierung der Dynamik des Proteins erlauben. Die Fluoreszenz-Resonanz-Energie-Transfer (FRET) Spektroskopie ermöglicht, z.B., die Abstandsbestimmung zwischen einem Donor- und einem Akzeptorfluorophor innerhalb eines Proteins und damit auf Einzelmolekülebene (smFRET) den Nachweis und die Bestimmung von verschiedenen Konformationen und den kinetischen Austausch zwischen diesen Konformationen. Im Laufe dieser Arbeit wurden Methoden zum selektiven Einbau von Fluorophoren in NRPS-Proteine entwickelt, darunter auch eine vereinfachte, generelle Darstellung von semi-synthetischen Proteinen (s. Kapitel 4.4.1). Die NMR-Spektroskopie von großen Multi-Domänen-Proteinen sieht sich zwei Problemen gegenüber. Zum einen führen große Proteine mit einer langen Korrelationszeit zu einer Verbreiterung der NMR-Signale. Zum anderen führt allein die große Anzahl an NMR-aktiven Kernen zu einer Signalüberlappung und zu einer erschwerten Zuordnung der Resonanzen. Das zweite Problem kann mit Hilfe gespaltener Inteine zur segmentellen Isotopenmarkierung umgangen werden. Kapitel 4.4.2 präsentiert diese Vorarbeiten zusammen mit NMR-spektroskopischen Untersuchungen an vollständig markierten NRPS-Proteinen.

### **4.4.1 Fluoreszenzspektroskopie**

Für alle fluoreszenzspektroskopischen Techniken ist das selektive Einbringen von Fluorophoren in die zu untersuchenden Proteine von großer Wichtigkeit. Die Proteinsemisynthese mit dem an Position 11 künstlich gespaltenen *Ssp* DnaB Intein erlaubt die gezielte N-terminale Modifikation von Zielproteinen mit unnatürlichen Gruppen, wie z.B. Fluorophore [169,170]. In der dieser Arbeit vorangegangenen Diplomarbeit wurde dieses Prinzip erfolgreich auf das in dieser Arbeit verwendete NRPS-Didomänenkonstrukt GrsA A-PCP übertragen [211]. Die PTS-Reaktion zwischen dem über Festphasenpeptidsynthese hergestellten Fluoresceinpeptid (Fluorescein-SEFSG-*Ssp* DnaB<sup>N</sup>(1-11)) und dem rekombinant

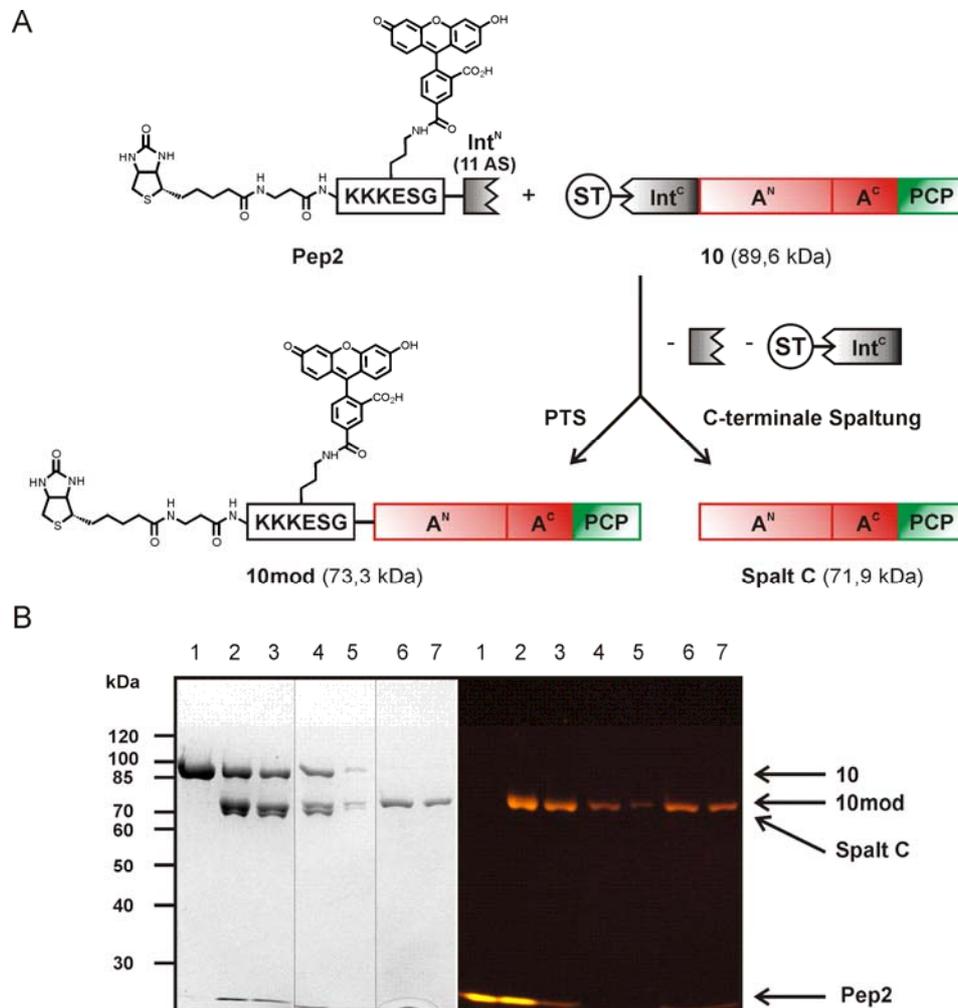
hergestellten Protein StrepTag II-*Ssp* DnaB<sup>C</sup>(12-154)-A-PCP-H<sub>6</sub> ergab das semisynthetische Spleißprodukt Fluorescein-A-PCP-H<sub>6</sub> (s. Abbildung 54).



**Abbildung 54** Schematische Darstellung des in früheren Arbeiten entwickelten Systems zur N-terminalen Modifikation von GrsA A-PCP [211]. Mittels des künstlich an Position 11 gespaltenen *Ssp* DnaB Inteins gelingt die selektive Einbringung eines Fluorophors (FI) an den N-Terminus des NRPS Didomänenproteins GrsA A-PCP; durch Auftragen der Mischung nach der PTS-Reaktion auf eine Streptactin-Säule konnten sowohl nicht-reagiertes Ausgangsmaterial, als auch die C-terminale Inteinhälfte abgetrennt werden.

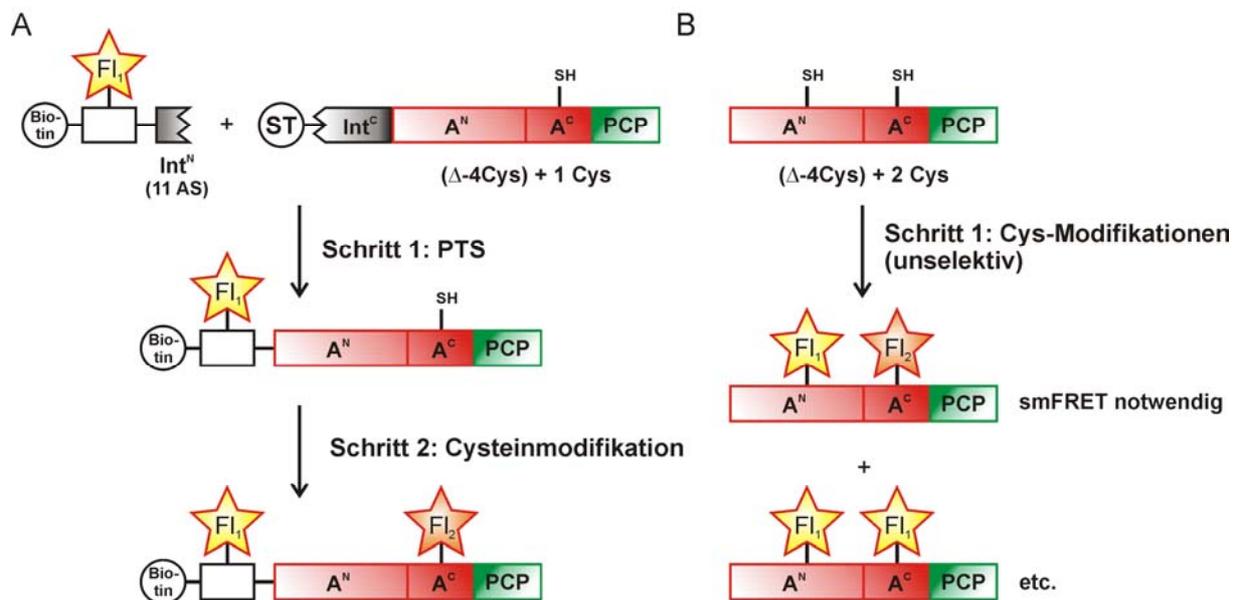
Diese Arbeit zeigte nicht nur die Synthese eines selektiv fluorophormarkierten NRPS-Proteins, sondern erlaubte mittels des N-terminalen StrepTag II am C-terminalen Konstrukt die Abtrennung des nicht-umgesetzten Edukts, sowie des nach der PTS-Reaktion herausgeschnittenen Int<sup>C</sup>-Fragmentes [211]. Jedoch war keine Trennung von Spleißprodukt und C-terminalem Spaltprodukt in diesem System möglich. Die Lösung dieses Problems gelang über den Einbau einer zusätzlichen unnatürlichen Gruppe in das N-Extein des synthetischen Peptids. Zusätzlich zu der Fluorescein-Gruppe wird ein Biotinrest durch PTS kovalent mit dem N-Terminus von A-PCP verknüpft. Nach einem Dialyseschritt zur Abtrennung des kleinen synthetischen Peptids ist das Spleißprodukt das einzige Molekül in der Mischung mit einer Biotin-Gruppe und kann über eine monomere Avidin-Säule gereinigt werden. Die Reaktion ist schematisch in Abbildung 55A dargestellt. Abbildung 55B zeigt die SDS-PAGE Analyse der Reaktion und der Reinigung. Die Synthese von **Pep2** (Biotin-βAKKK(Fluorescein)ESG-*Ssp* DnaB<sup>N</sup>(1-11)) gelang standardmäßig über Festphasenpeptidsynthese. Protein **10** (ST-*Ssp* DnaB Int<sup>C</sup>(12-154)-A-PCP-H<sub>6</sub>) wurde durch rekombinante Expression in *E. coli* hergestellt. Nach Inkubation für 30h bei 25°C (mit einem Überschuss an **Pep2**) konnte ein 70%iger Umsatz des Eduktproteins festgestellt werden. Davon waren ungefähr zwei Drittel durch PTS, das andere Drittel durch die C-terminale Spaltungsreaktion verbraucht worden (s. Abbildung 55B, Spuren 1 und 2). Durch Affinitätschromatographie gegen die Biotin-Gruppe konnte das Spleißprodukt von allen weiteren Komponenten, vor allem von dem C-terminalen Spaltprodukt (**Spalt C**), getrennt

werden (s. Abbildung 55B, Spuren 4 bis 7). Damit konnte gezeigt werden, dass die hier beschriebene Technik die Synthese und Reinigung von N-terminal modifizierten semisynthetischen Proteinen über positive Affinitätschromatographie möglich macht.



**Abbildung 55** Synthese und Reinigung des mit Fluorescein und Biotin modifiziertem GrsA A-PCP Proteins. A) Schematische Darstellung der Reaktion zwischen **Pep2** und Protein **10** (ST-*Ssp* DnaB Int<sup>C</sup>(12-154)-A-PCP-H<sub>6</sub>); B) SDS-PAGE Analyse der Synthese und Reinigung; links ist das Gel in der Coomassie-Färbung, rechts unter UV-Bestrahlung gezeigt; Spur 1 = Reaktion 0h; Spur 2 = Reaktion 30h; Spur 3 = Dialyse; Spuren 4-7 = Durchfluss, Waschen und zwei Elutionsfraktionen der Reinigung über eine monomere Avidinsäule.

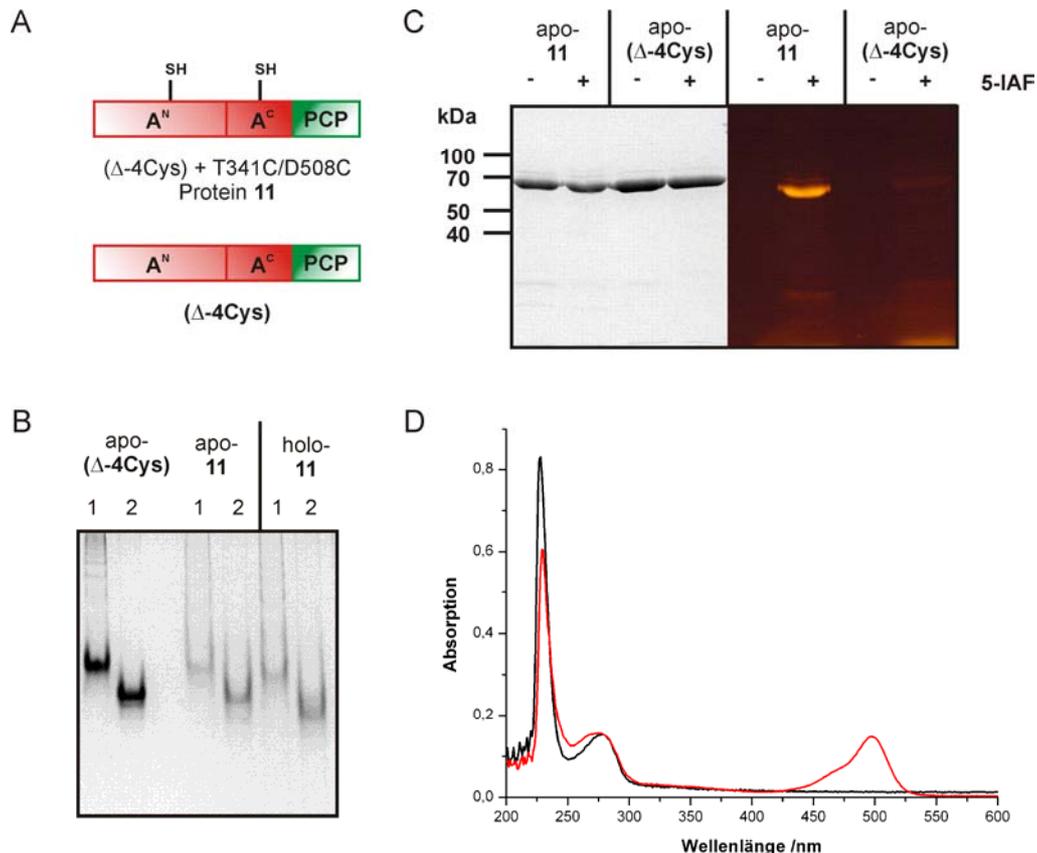
Zusätzlich zu der Fluorophormarkierung von GrsA A-PCP über die PTS-Reaktion ist für die FRET-Spektroskopie ein zweiter Fluorophor an einer definierten Position notwendig. Diese Einbringung könnte z.B. über eine Cysteinmodifikation erfolgen. Diese Strategie zum selektiven Einbau zweier Fluorophore in das GrsA A-PCP Protein ist schematisch in Abbildung 56A dargestellt. Die Vorarbeiten auf diesem Gebiet beschränken sich jedoch auf die Klonierung eines Konstruktes analog zu Protein **10** mit einem cystein-freien NRPS-Teil (C60F, C331A, C376S und C473A - A-PCP( $\Delta$ -4Cys)), sowie der Herstellung von Mutanten mit einem zusätzlich eingeführten nicht-nativen Cysteinrest.



**Abbildung 56 Strategien zur A) selektiven und B) unselektiven Einbringung zweier Fluorophore in GrsA A-PCP für die FRET-Spektroskopie.** A) Über PTS mit dem an Position 11 künstlich gespaltenen *Ssp* DnaB Intein kann in einem ersten Schritt ein Fluorophor ( $Fl_1$ ) selektiv an den N-Terminus von GrsA A-PCP eingebracht werden; die Reinigung des Spleißproduktes ist über die zusätzlichen Biotingruppe möglich; durch Mutation der nativen vier Cysteinreste und dem Einbau eines weiteren Cysteinrestes besitzt dieser nun die einzige Thiolfunktionalität im Spleißprodukt und kann im zweiten Schritt selektiv mit einem zweiten Fluorophor ( $Fl_2$ ) modifiziert werden; B) Zwei nicht-native Cysteinreste in GrsA A-PCP( $\Delta-4Cys$ ) erlauben den unselektiven Einbau zweier Fluorophore mittels Cysteinmodifikation; da nach der Reaktion ein Gemisch aus unterschiedlich modifizierten Proteine vorliegt, muss mit dieser Strategie auf die Einzelmolekülspektroskopie (smFRET) zurückgegriffen werden.

Die in Abbildung 56A vorgestellte Strategie ist in der Hinsicht begrenzt, dass ein Fluorophor am N-Terminus des Proteins positioniert werden muss. Zwar lässt sich die Position des zweiten Fluorophors über die Herstellung unterschiedlicher Mutanten variabel gestalten, jedoch würden alle Abstandsmessungen über die FRET-Spektroskopie relativ auf einen Punkt bezogen werden. Eine größere Freiheit erlaubt die in Abbildung 56B dargestellte Strategie. Hier werden zwei Cysteinreste an freiwählbaren Punkten in GrsA A-PCP über Punktmutagenese eingebracht. Die Inkubation mit einer Mischung von thiol-modifizierenden Donor- und Akzeptorfarbstoffen führt zu einer statistischen Markierung der Reste. Die relevanten Kombinationen von Donor und Akzeptor könnten nach Immobilisierung an einer festen Oberfläche mittels Einzelmolekülspektroskopie selektiv und getrennt von den Donor-Donor und Akzeptor-Akzeptor Kombinationen untersucht werden. Frühere Arbeiten zeigten, dass die vier Wildtyp-Cysteine in der A-Domäne von GrsA A-PCP sich nur schlecht für eine Fluorophormarkierung eignen [211]. Die Konjugation mit Fluorescein-Maleinimid war im Mittel nur zwischen 10 - 40% möglich [211]. Ausgehend von dem cystein-freien Konstrukt A-PCP( $\Delta-4Cys$ ) wurden unterschiedliche Cystein-Einzelmutanten erstellt (N152C, N260C, T341C, H481C, D508C, S585C). Angeleitet von der Kristallstruktur der GrsA A-Domäne

[57] und den NMR-Untersuchungen an TycC3 PCP [87] wurde versucht diese an solvenszugängliche Positionen in sowohl beide Subdomänen der A-Domäne, als auch in die PCP-Domäne einzubringen. Auch mit diesen neu eingeführten Cysteinresten konnte keine vollständige Markierung erreicht werden, jedoch ließ sich die Markierungseffizienz deutlich auf 60-70% steigern (Daten nicht gezeigt). Weiterhin wurde beispielhaft für eine Doppelmutante die Kombination T341C, D508C in A-PCP( $\Delta$ -4Cys) erzeugt.



**Abbildung 57 Die Doppelmutante A-PCP( $\Delta$ -4Cys, T341C, D508C)-H<sub>6</sub> (Protein 11).** A) Schematische Darstellung der in dieser Abbildung verwendeten Proteine; B) Native PAGE Analyse von apo- und holo-Protein **11** in Abwesenheit (Spuren 1) und in Gegenwart von 100  $\mu$ M **Inh 1** (Spuren 2); zum Vergleich ist die apo-Form des cystein-freien Proteins gezeigt; C) SDS-PAGE Gel der Konjugationsansätze von den apo-Proteinen **11** und ( $\Delta$ -4Cys) mit 5-Iodacetamid-fluorescein (5-IAF), links ist das Gel in der Coomassiefärbung, rechts unter UV-Beleuchtung gezeigt; D) UV/Vis-Absorptionsspektren von apo-**11** vor (schwarzes Spektrum) und nach der Reaktion mit 5-IAF (rotes Spektrum), die Spektren sind auf die Absorption von 280 nm normiert.

Dieses Protein verhielt sich in den nativen PAGE Assays vergleichbar zu dem Wildtyp-Protein (s. Abbildung 57B) und konnte selektiv an den Cystein-Resten mit 5-Iodacetamid-Fluorescein modifiziert werden (vergleiche in Abbildung 57C die Reaktion mit dem cystein-freien Protein unter identischen Reaktionsbedingungen). Die Auswertung der UV/Vis-Spektroskopie ergab ein Verhältnis von Protein zu Fluorophor von 1:1,4 (s. Abbildung 57D), was den Rückschluss nahe legt, dass eine Modifikation an beiden Cystein-Resten stattgefunden haben muss. Auch das modifizierte Protein veränderte seine elektrophoretische

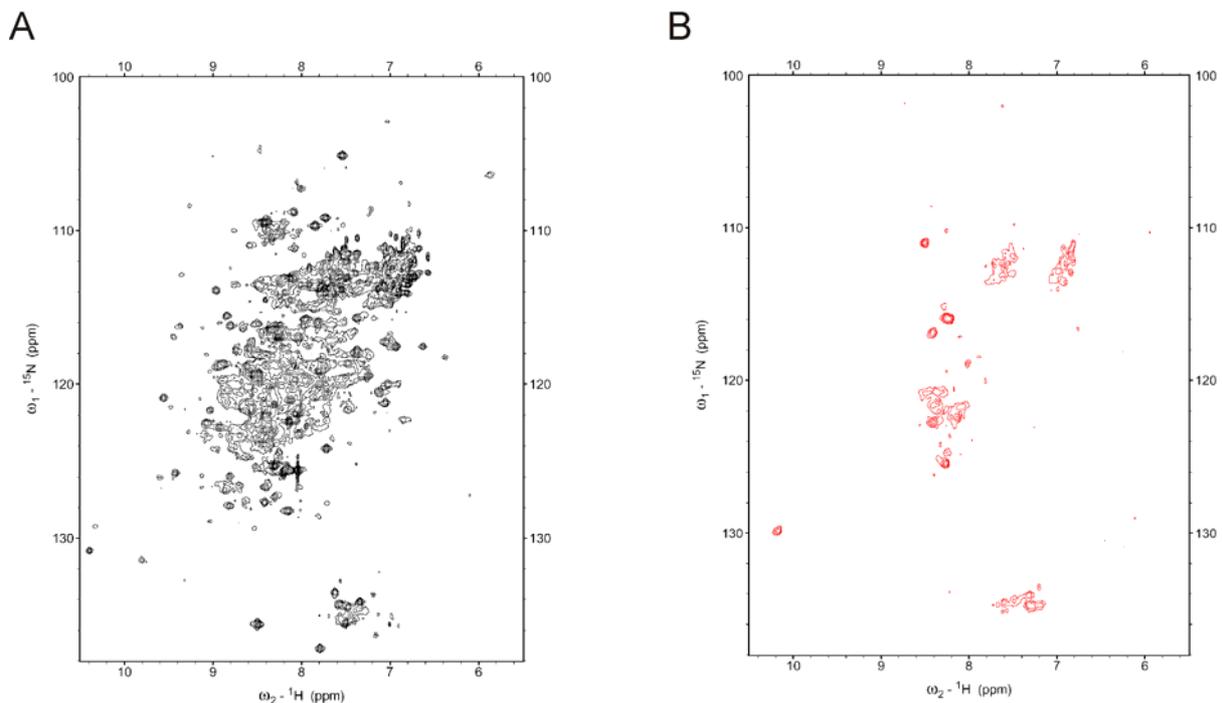
Mobilität in den nativen PAGE Assays bei Zugabe des kognaten Inhibitors (**Inh1**) (Daten nicht gezeigt) und scheint damit ebenfalls eine Konformationsänderung bei der Komplexbildung zu vollziehen. Alles in allem ist diese Doppelmutante geeignet um nach der Strategie in Abbildung 56B per smFRET-Spektroskopie analysiert zu werden. Die in dieser Arbeit durchgeführten Vorarbeiten werden in der Arbeitsgruppe Mootz in Kollaboration mit der Gruppe von Prof. H. Yang (Princeton University, Department of Chemistry, USA) weitergeführt.

#### 4.4.2 NMR-Spektroskopie

Die NMR-Spektroskopie in Lösung hat bereits bedeutende Erfolge in der Aufklärung der strukturellen Eigenschaften von NRPS-Proteinen erzielt [87,93]. Kürzlich wurde das dynamische Verhalten eines PCP-TE Didomänenfragments aus EntF mittels NMR aufgeklärt [93]. Die Zuordnung von 93% der Rückgratresonanzen (313 von 337 Resten) gelang in dem 37 kDa Didomänenprotein durch eine große Zahl an unterschiedlich isotoopenmarkierter Proben. In teilweise deuterierten Proben war die Spinrelaxation noch ausreichend langsam um 3D NMR-Spektren mit einer geringen Linienbreite und damit einem guten Signal zu Rausch Verhältnis zu ermöglichen. Vor allem diese Arbeit [93] verdeutlicht das Potential der NMR-Spektroskopie, die es erlaubt NRPS-Proteine in Lösung mit atomarer Auflösung zu untersuchen.

Um zu ermitteln, in wie weit auch größere NRPS Multidomänen-Proteine einer NMR-Untersuchung in Lösung zugänglich sind, wurden ein A-PCP Didomänenfragment und ein komplettes Elongationsmodul vollständig  $^{15}\text{N}$ -markiert und von Alena Busche und Dr. Frank Löhr (AK Prof. V. Dötsch, Goethe-Universität Frankfurt am Main) mittels NMR spektroskopisch vermessen. Die  $^{15}\text{N}$ - $^1\text{H}$  TROSY Spektren von GrsA A-PCP- $\text{H}_6$  (71,1 kDa) und ST-GrsB1 C-A-PCP- $\text{H}_6$  (124,2 kDa) sind in Abbildung 58 zu sehen. Im Spektrum des kompletten NRPS-Moduls (s. Abbildung 58B) sind die Resonanzen wahrscheinlich aufgrund der großen Korrelationszeit und einer daraus resultierenden schnellen Relaxation sehr verbreitert. Das aufgenommene Spektrum spricht dagegen, dass ein kleinerer Teil, z.B. die PCP-Domäne, flexibel ist und unabhängig vom Rest des Proteins rotieren kann. Es ist anzunehmen, dass auch eine Deuterierung des Proteins nicht zu einer dramatischen Verbesserung des Spektrums führt und damit mit den aktuellen NMR-Techniken keine Untersuchung eines kompletten NRPS-Moduls in Lösung möglich ist. Auch bei dem NMR Spektrum des A-PCP Didomänenproteins (s. Abbildung 58A) ist die Linienverbreiterung

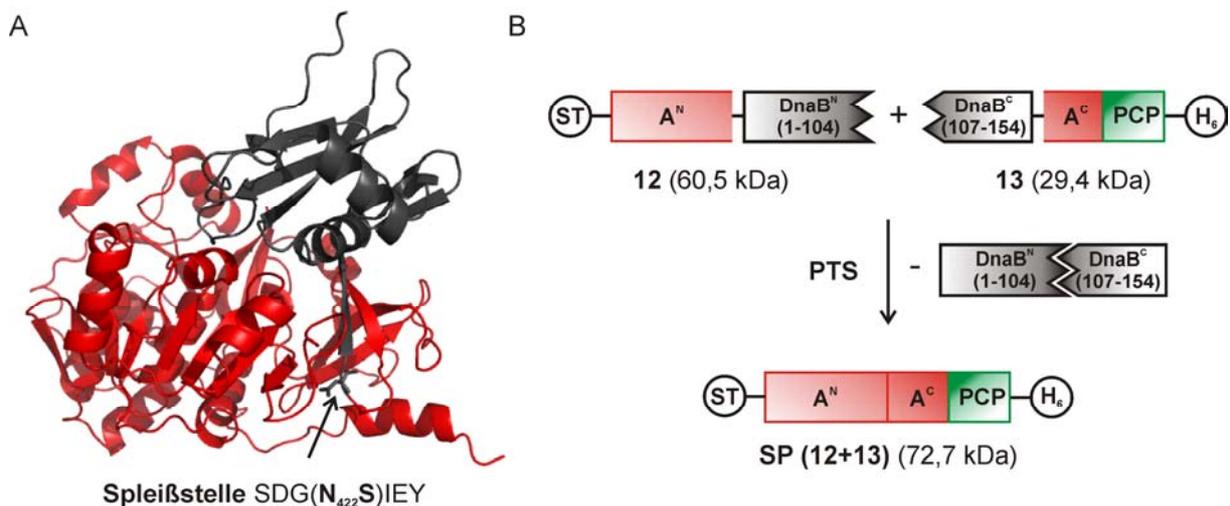
vorhanden, wenn auch nicht so stark ausgeprägt wie im Falle des größeren GrsB1 C-A-PCPs. Jedoch erlaubt die Signalüberlappung definitiv keine Zuordnung der Resonanzen. Eine Deuterierung des GrsA A-PCP Proteins dürfte in Kombination mit einer selektiven Markierung in  $^{15}\text{N}$  und  $^{13}\text{C}$  die Qualität des NMR Spektrums erhöhen und vielleicht in Zukunft eine strukturelle Untersuchung ermöglichen. Als Vorarbeit in Richtung der segmentellen Isotopenmarkierung des GrsA A-PCP NRPS-Proteins wurde im Laufe dieser Arbeit eine aktive Insertionsposition für das an der Position der Endonuklease gespaltene *Ssp* DnaB Intein in Zusammenarbeit mit Annika Urbanek ermittelt [212].



**Abbildung 58** NMR Untersuchungen an vollständig  $^{15}\text{N}$ -markierten Multidomänen NRPS-Proteinen.  $^{15}\text{N}$ - $^1\text{H}$  TROSY Spektren von GrsA A-PCP- $\text{H}_6$  (A) und ST-GrsB1 C-A-PCP- $\text{H}_6$  (B).

Nach den folgenden Kriterien wurde die Intein-Insertionsstelle  $\text{N}_{422}$  in dem GrsA A-PCP Protein ausgewählt. Eine Trennung des Proteins zwischen den beiden Subdomänen der A-Domäne erschien sinnvoll, da damit eine simultane Untersuchung der beiden wahrscheinlich mobilen Bereiche, der  $\text{A}^{\text{C}}$ -Subdomäne und der PCP-Domäne, gewährleistet werden könnte. Eine Betrachtung der Primärsequenz des NRPS Proteins zeigte, dass die Region um  $\text{N}_{422}$  große Homologie mit den natürlichen Exteinsequenzen des *Ssp* DnaB Inteins aufwies. Nimmt man  $\text{N}_{422}$  als das +1 Nukleophil der Spleißreaktion, so stimmen die -1, die +2 und +3 Positionen überein. Nach der Kristallstruktur der GrsA A-Domäne [57] befand sich dieser Rest zusätzlich in einer nicht-konservierten Loop-Region (s. Abbildung 59A). Die Klonierung des N-terminalen (ST-GrsA  $\text{A}_{1-421}$ -*Ssp* DnaB $^{\text{N}}$ (1-104), Protein **12**) und des C-terminalen

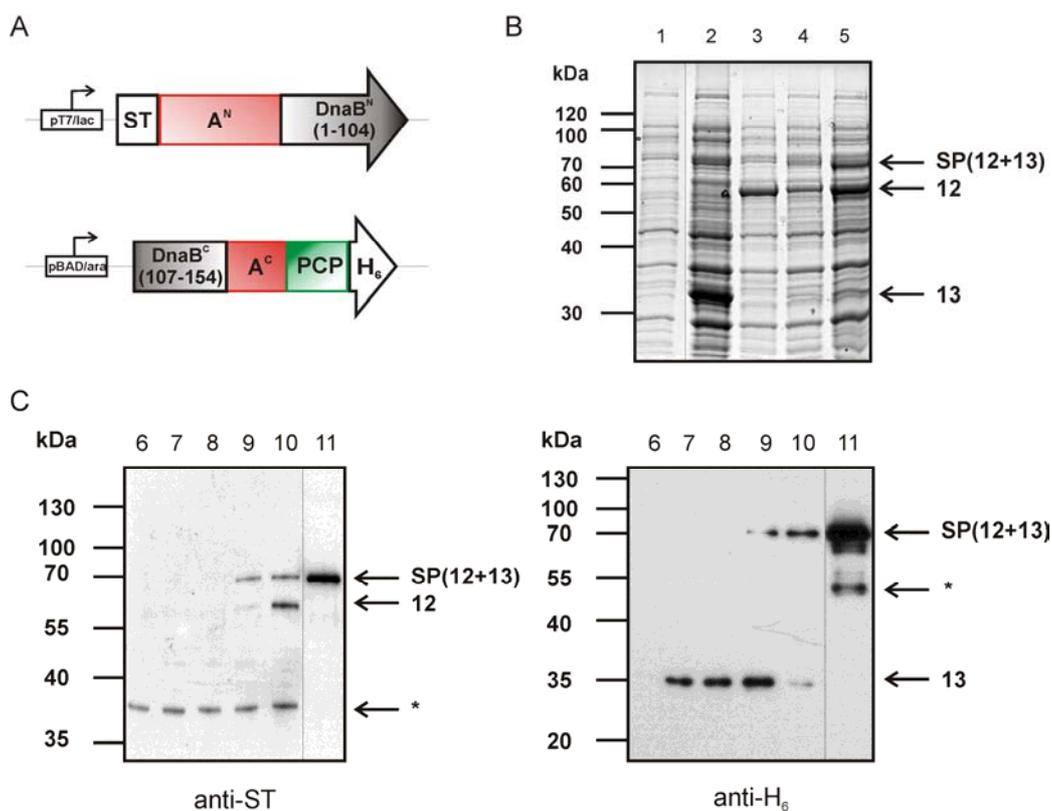
Konstruktes (*Ssp* DnaB<sup>C</sup>(107-154)-S-GrsA A<sub>423-540</sub>-PCP-H<sub>6</sub>, Protein **13**) erfolgte noch nicht mit dem SPLICEFINDER-System, sondern klassisch über Fusions-PCR und Standardtechniken [212]. Das Schema der PTS-Reaktion von Protein **12** mit Protein **13** ist in Abbildung 59B dargestellt. Das resultierende potentielle Spleißprodukt **SP(12+13)** enthält im Vergleich zum Wildtyp-Protein eine N<sub>422</sub>S Mutation und ist zur vereinfachten Reinigung das einzige Protein mit sowohl einem ST-, als auch einem H<sub>6</sub>-Affinitätstag. Ein Testen der *in vitro* PTS-Aktivität mit gereinigten Proteinen war aufgrund der Unlöslichkeit des N-terminalen Proteins **12** nicht durchführbar. Auch bei einem Versuch das Spleißen im Zellextrakt durchzuführen, konnte keine Spleißproduktbildung beobachtet werden (Daten nicht gezeigt).



**Abbildung 59 Insertion des künstlich gespaltenen *Ssp* DnaB Inteins in GrsA A-PCP an Position 422.** A) Kristallstruktur der GrsA A-Domäne (Conti, EMBO, 1997) (pdb-Code 1AMU); die Aminosäurereste 1-421 sind in rot, die Reste 422-530 sind in schwarz dargestellt; B) Schematische Darstellung der PTS-Reaktion zwischen den Proteinen **12** (ST-GrsA A<sub>1-421</sub>-*Ssp* DnaB<sup>N</sup>(1-104)) und **13** (*Ssp* DnaB<sup>C</sup>(107-154)-S-GrsA A<sub>423-540</sub>-PCP-H<sub>6</sub>).

Die letzte Möglichkeit die Unlöslichkeit der N-terminalen Hälfte zu umgehen bestand in der *in vivo* Koexpression beider Proteine in einer *E. coli* Zelle. Die beiden Vektoren der Inteininfusionskonstrukte waren analog zu dem von Iwai und Mitarbeitern entwickelten Expressionssystem so gewählt, dass sie unterschiedliche Antibiotikaresistenzen aufwiesen und die Proteine orthogonal induzierbar waren (s. Abbildung 60A). Durch die Zugabe von Arabinose konnte die Expression des C-terminalen Konstruktes (vergl. Spuren 1 und 2 in Abbildung 60B) und mit IPTG die des N-terminalen Konstruktes induziert werden (vergl. Spuren 1 und 3 in Abbildung 60B). Die Koexpression beider Proteine führte bei 25°C und 22°C, nicht jedoch bei 37°C zur Bildung des Spleißproduktes (ST-GrsA A<sub>N422S</sub>-PCP-H<sub>6</sub>, **SP(12+13)**) (s. Spuren 4 und 5 in Abbildung 60B). 22°C wurde in weiteren Versuchen als Induktionstemperatur verwendet, da hier zwar ein geringeres Zellwachstum, jedoch ebenfalls

eine größere Spleißproduktbildung im Vergleich mit 25°C festzustellen war [212]. Das Spleißprodukt **SP(12+13)** konnte mittels tryptischen in-Gelverdau und anschließender MALDI-TOF MS eindeutig nachgewiesen werden [212]. Interessanterweise war das an Position 11 gesplattene *Ssp* DnaB Intein an der identischen Insertionsposition N<sub>422</sub>S bei den *in vivo* Koexpressionen nicht aktiv [212]. In diesem System hatte die Unlöslichkeit der C-terminalen Hälfte das Testen der *in vitro* Spleißaktivität verhindert (Daten nicht gezeigt). Vielleicht ist bei dem an Position 11 gesplatteten *Ssp* DnaB Inteins die Assoziation der beiden Inteinhälften nicht schnell genug um die Bildung von unlöslichen Aggregaten des C-terminalen Proteins in der Zelle zu verhindern.



**Abbildung 60 Vorversuche zur segmentellen Isotopenmarkierung von ST-GrsA A-PCP-H<sub>6</sub>.** A) Schematische Darstellung der beiden verwendeten Expressionsplasmide; das N-terminale Konstrukt ist unter Kontrolle der pT7/lac Promotor/Operator Sequenzen in einem Vektor mit Kanamycin-Resistenz und RSF Replikationsursprung, das C-terminale Konstrukt wird durch die pBAD/ara Promotor/Operator Sequenzen kontrolliert und befindet sich in einem Plasmid mit Ampicillin-Resistenz und dem pBR322 Replikationsursprung; B) SDS-PAGE Gel der Testexpressionen des dualen Expressionssystems; Spur 1 = Probe vor Induktion; Spur 2 = Probe nach Arabinose-Induktion, 4h, 37°C; Spur 3 = Probe nach IPTG-Induktion, 4h, 37°C; Spuren 4 und 5 = Proben nach simultaner IPTG und Arabinose-Induktion für 4h und über Nacht, 22°C; C) Western Blot Analyse gegen den StrepTag II (links) und den H<sub>6</sub>-Tag (rechts) der sukzessiven Induktionen mit Mediumswechsel; Spur 6 = Probe vor Induktion; Spur 7 = Probe nach Arabinose-Induktion in M9 Minimalmedium, 2h, 37°C; Spur 8 = nach Waschen und Aufnahme der Zellen in LB-Medium; Spur 9 und 10 = anschließende IPTG-Induktion für 2h und 16h in LB-Medium, 22°C; Spur 11 = gereinigtes Spleißprodukt **SP(12+13)**; die theoretischen Molekulargewichte der Proteine sind wie folgt: **SP(12+13)** = 72,7 kDa; **12** = 60,5 kDa; **13** = 29,4 kDa; Proteinverunreinigungen \* sind markiert.

Eine segmentelle Isotopenmarkierung des A<sup>C</sup>-PCP-H<sub>6</sub> Fragmentes mittels des an der Endonuklease gespaltenen *Ssp* DnaB Inteins wurde über sukzessive Proteininduktionen mit einem Mediumswechsel simuliert (s. Western Blot Analysen in Abbildung 60C). Die selektive Expression der C-terminalen Hälfte (Protein **13**) konnte in unmarkiertem M9 Minimalmedium durch die Zugabe von 0,2% Arabinose bei 37°C herbeigeführt werden (vergl. Spuren 6 und 7 in Abbildung 60C). Nach einem Waschschrift (s. Spur 8 in Abbildung 60C) und einer Aufnahme der Zellen in LB-Medium wurde die Temperatur auf 22°C gesenkt und durch 0,8 mM IPTG die Genexpression des N-terminalen Konstruktes gestartet. Bereits nach 2h konnte die Bildung des Spleißproduktes **SP(12+13)** detektiert werden (s. Spur 9 in Abbildung 60C). Weitere 14h später war das C-terminale Protein **13** durch PTS fast vollständig verbraucht (s. Spur 10 in Abbildung 60C). Das erzeugte Spleißprodukt ließ sich über Ni<sup>2+</sup>-Affinitätschromatographie reinigen und zeigte in biochemischen Assays zur Aktivität der A-Domäne ein vergleichbares Verhalten wie das Wildtyp-Protein [212].

In diesem Kapitel konnte damit die erfolgreiche Synthese des GrsA A-PCP Didomänenproteins aus zwei Segmenten über PTS mit dem an der Endonuklease gespaltenen *Ssp* DnaB Intein vorgestellt werden. Das System erzeugt ein aktives Spleißprodukt mit einer N<sub>422</sub>S Mutation und ist für eine segmentelle Isotopenmarkierung durch PTS *in vivo* geeignet. Sollte diese Arbeit in Zukunft fortgesetzt werden, so ist vor allem eine Verbesserung der Ausbeute (0,7 mg Spleißprodukt pro L Expressionskultur) anzustreben.

---

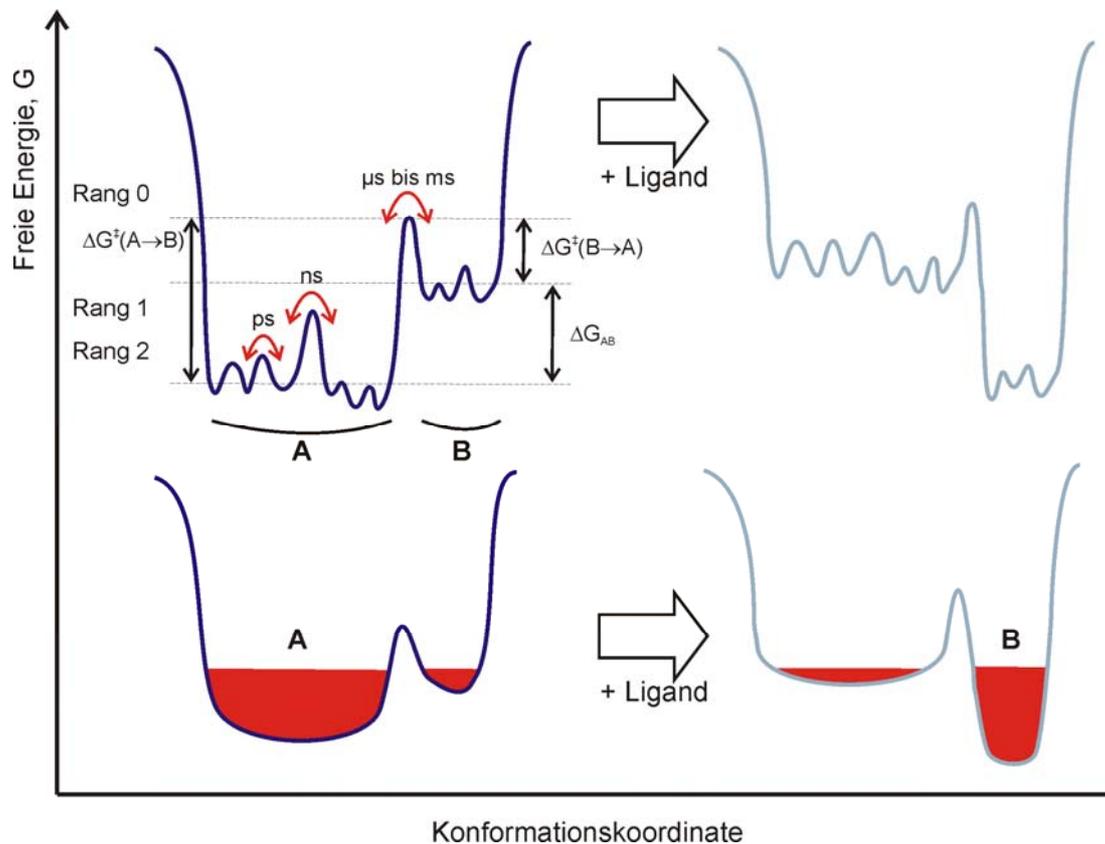
## 5 Diskussion und Ausblick

Weil Proteine in allen Prozessen des Lebens eine essentielle Rolle spielen, ist das Verständnis ihrer Funktion und Wirkungsweise von fundamentaler Bedeutung. Interessanterweise sind die von Enzymen katalysierten Reaktionen genauso wie das Leben selbst inhärent dynamisch. Die Bindung und Positionierung von Substraten und reaktiven Gruppen, sowie die Bindungsknüpfung und Brechung, verlangen zwangsläufig eine strukturelle Änderung der Protein-Ligand- oder der Protein-Protein-Komplexe. Hochauflösende Techniken, wie die Röntgenkristallographie oder die NMR-Spektroskopie haben es erlaubt Proteinstrukturen mit atomarer Auflösung zu erhalten, jedoch implizieren diese Analysen ein Bild von einem Protein als statischem Objekt. Die Interpretation, dass eine definiert gefaltete Konformation des Proteins ausreichend ist, um die Funktion des Proteins zu verstehen wird in den letzten Jahrzehnten von zwei unterschiedlichen Seiten in Frage gestellt. Zum einen scheint nicht nur der energetische Grundzustand des Proteins, sondern eine Vielzahl an ineinander übergehenden Konformationen für die katalytische Wirkungsweise essentiell zu sein. Zum anderen existiert eine Klasse von Proteinen, die intrinsisch ungefaltet vorliegt. Beispiele aus beiden Gruppen wurden in dieser Doktorarbeit untersucht. So sind nicht-ribosomale Peptidsynthetasen ein Paradebeispiel für dynamische Proteine, die große konformationelle Änderungen während des Katalysezyklus eingehen. Außerdem liegen die individuellen Hälften von gespaltenen Inteinen in ihrem nativen Zustand ohne das Partnerprotein ungefaltet vor.

### **5.1 Dynamische Eigenschaften von Proteinen und Protein-Ligand Komplexen**

Zur Beschreibung der strukturellen und dynamischen Eigenschaften eines Proteins wird in der aktuellen Forschung das Konzept der Energielandschaften (*engl.* energy landscapes) herangezogen. Es wurde zuerst für Faltungsstudien, d.h. den Übergang eines Proteins aus einer ungefalteten in eine definierte gefaltete Konformation, entwickelt [248-251]. Die Energielandschaft eines Proteins ist der Raum aller möglichen Konformationen dieses Proteins mit ihrer zugehörigen freien Energie ( $\Delta G$ ). Die Struktur des Proteins kann durch die Koordinaten aller N Atome (die Lösungsumgebung eingeschlossen) beschrieben werden und ein konformationeller Zustand ist ein Punkt in diesem  $3N-3$  dimensionalen Raum. Die

Anzahl aller möglicher Konformationen ist immens groß und sie bilden zusammen die Energielandschaft des Proteins. Eine reale Veränderung in der Proteinstruktur ist mit der Bewegung von einem Punkt zum anderen im Konformationsraum der Energielandschaft gleichzusetzen [249,252,253]. Die exakte Form einer Energielandschaft ist verständlicherweise von einer Vielzahl an Parametern wie Temperatur, Druck und Zusammensetzung des Lösungsmittels abhängig.



**Abbildung 61 Ein-dimensionaler Schnitt durch die Energielandschaft eines Proteins zeigt das dynamische Verhalten eines Proteins alleine und bei Bindung eines Liganden.** Gezeigt ist eine Energielandschaft eines hypothetischen Proteins mit zwei unterschiedlichen Rang 0 Konformationen; A und B. Ein Zustand ist durch ein Minimum, ein Übergangszustand durch ein Maximum in der Energielandschaft repräsentiert. Die Besetzung der Rang 0 Zustände A und B ist thermodynamisch anhand der Boltzmann-Verteilung durch die mittlere Energiedifferenz  $\Delta G_{AB}$  festgelegt und schematisch im unteren Teil der Abbildung dargestellt. Die Höhe der Barriere  $\Delta G^\ddagger$  bestimmt die Umwandlungsrate  $k$  zwischen diesen Zuständen. Rang 1 und Rang 2 Übergänge sind kleinere konformationelle Änderungen innerhalb der Rang 0 Zustände. Bei Bindung eines Liganden verändert sich die Energielandschaft des Proteins. Dabei kann der Gleichgewichtszustand in Richtung einer ehemals energetisch angeregten Konformation verschoben werden (reproduziert nach [254,255]).

Der in Abbildung 61 oben links dargestellte eindimensionale Schnitt durch die multidimensionale Energielandschaft soll das dynamische Verhalten eines Proteins vermitteln, welches im Gegensatz zu den traditionellen statischen Strukturen des thermischen Grundzustandes oder des über die Zeit gemittelten Ensembles steht. Das globale Minimum der Energielandschaft wird dem energetischen Grundzustand zugeordnet. Für viele Proteine

existieren weiterhin mehrere lokale Minima, die konformationelle Unterzustände repräsentieren. Die Besetzung der unterschiedlichen Konformationen in einem identischen Proteinensemble ist von den Energiedifferenzen der Zustände ( $\Delta G$ ) in der Energielandschaft abhängig (s. Abbildung 61). Die Kinetik der Umwandlung der Strukturen ineinander wird hingegen von der Höhe der Barrieren ( $\Delta G^\ddagger$ ) diktiert (s. z.B. [256,257] und Abbildung 61). Die konformationellen Änderungen werden anhand der Reaktionsrate in unterschiedliche Ränge (*engl. tiers*) klassifiziert. Die Rang 0 Übergänge sind durch große Energiebarrieren im Bereich von mehreren  $kT$  (dem Produkt der Boltzmann-Konstante und der absoluten Temperatur) getrennt und spielen sich im Bereich von  $\mu s$  bis  $ms$  ab. Üblicherweise sind das große konzertierte und kollektive Regionen- oder Domänenbewegungen. Innerhalb dieser Rang 0 Zustände sind die Proteine ebenfalls nicht statisch sondern können Rang 1 (z.B. kleinere Loop-Bewegungen) und Rang 2 (z.B. Seitenkettenrotationen) Übergänge eingehen (s. Abbildung 61) [252,255].

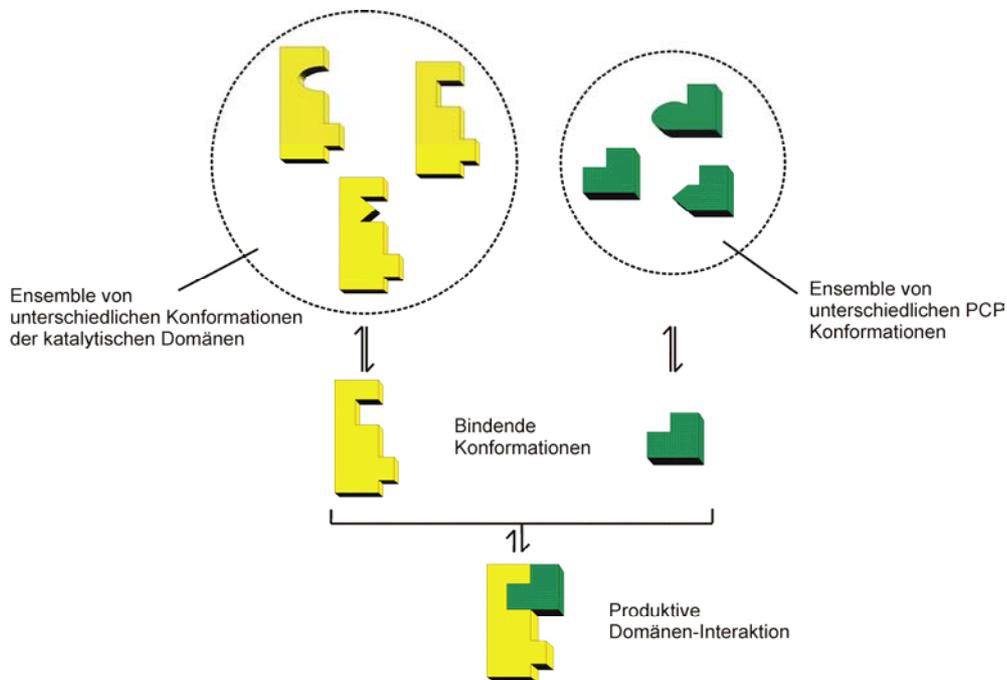
In dieser Arbeit standen die Protein-Protein Wechselwirkungen innerhalb der NRPS-Biosynthese auch im Zusammenspiel mit den niedermolekularen Substraten und Inhibitoren im Fokus der Untersuchung. Zur Erklärung der Bindung zwischen einem Liganden und einem Protein gingen die frühesten Modelle auch von einem statischen Bild der Assoziation aus. So postulierte Emil Fischer 1894 sein berühmtes Schlüssel-Schloß-Prinzip, in dem das Substrat und der Ligand perfekt ohne konformationelle Änderung ineinander passen würden. Die Entdeckung von allosterischen und ko-operativen Effekten bei der Aktivität von Enzymen führte zu dem so genannten „induced fit“ Mechanismus, der eine gewisse konformationelle Plastizität an den Ligandenbindungsstellen und unterschiedliche Konformationen postuliert [258]. Die neuere Forschung beschreibt die Bindung eines Liganden an ein Protein oder eine analog zu betrachtende Protein-Protein Wechselwirkung mit Hilfe der Energielandschaften eher wie folgt: Unter den sich ineinander umwandelnden Konformationen des freien Proteins, wird eine Struktur durch die Bindung des Liganden energetisch stabilisiert und damit am meisten besetzt (s. Abbildung 61). Anders ausgedrückt induziert ein Ligand nicht eine neue Konformation im Protein, sondern selektiert eine bereits vorhandene Konformation bevorzugt. In der neu erzeugten Energielandschaft des Komplexes ist der neue Grundzustand in einer Vielzahl der Fälle ein ehemals angeregter Zustand [255,259,260] (s. Abbildung 61). Das aus Kristallstrukturuntersuchungen erhaltene Bild eines Proteins in einer definierten, gefalteten Konformation ist somit nicht ausreichend um die vielen unterschiedlichen Wechselwirkungen eines Proteins mit Liganden und anderen Proteinen zu erklären. Für ein

tieferes Verständnis der enzymatischen Katalyse ist eine strukturelle Charakterisierung des Proteins über die Zeit und damit seiner dynamischen Eigenschaften notwendig [254,255].

### 5.1.1 Konformationelle Änderungen und Domäneninteraktionen in NRPS Proteinen

NRPS-Domänen sind nach dem aktuellen Forschungsstand (zusammengefasst in Kapitel 1.1) ein Paradebeispiel für dynamische Proteine, die in einem Gleichgewicht zwischen verschiedenen Konformationen vorliegen. So wurden für eine isolierte PCP-Domäne mehrere sich ineinander umwandelnde Strukturen mittels NMR-Spektroskopie gefunden [87]. Interessanterweise verändert die post-translationale Modifikation mit dem Ppant-Arm die Energielandschaft der PCP-Domäne derart, dass in der apo- und in der holo-Form der dynamische Austausch zwischen unterschiedlichen Konformationen vonstatten geht. Es existiert jeweils eine Struktur, die nur von dem apo-Protein und eine Struktur, die nur von dem holo-Protein eingenommen wird [87]. In sehr guter Analogie mit dem oben beschriebenen Modell der Protein-Protein Wechselwirkung verschiebt die Gegenwart eines externen, interagierenden Proteins, der PPTase Sfp, das Gleichgewicht von apo-PCP in Richtung einer bestimmten Konformation, dem A-Zustand. Dieser Zustand wird bevorzugt von Sfp gebunden, damit stabilisiert und thermodynamisch begünstigt. Das holo-Protein nimmt den A-Zustand aufgrund der veränderten Energielandschaft im Vergleich mit apo-PCP kaum ein. Deshalb ist auch keine Wechselwirkung zwischen Sfp und dem holo-Protein festzustellen [87]. Nicht nur PCP-Domänen, sondern auch andere NRPS-Proteine unterliegen einem konformationellen Austausch. Die TEII-Domäne aus *Bacillus subtilis* existiert ebenfalls in unterschiedlichen Konformationen, aber nur eine wird bei der Bindung an falsch beladenes holo-PCP stabilisiert [91]. Auch für die TEI-Domäne aus der Enterobactin-Biosynthese wurden Proteindynamiken mittels NMR-Spektroskopie nachgewiesen [93]. Hier kontrolliert das Öffnen und Schließen einer Deckelregion den Zugang zum aktiven Zentrum dieser Domäne. Dass auch für *in cis* vorliegende NRPS Didomänenproteine analoge dynamische Prozesse eine Rolle spielen, konnte zum ersten Mal in der herausragenden Arbeit von Frueh *et al.* nachgewiesen werden [93]. Das untersuchte apo-PCP-TE Fragment liegt in zwei sich langsam ineinander umwandelnden Konformationen vor. Eine Struktur zeigt beide Domänen in engem Kontakt; in der Anderen hat sich die PCP-Domäne räumlich von der TE-Domäne entfernt. Letztere Konformation wird durch die Titration mit Sfp stabilisiert und ist in Gegenwart der Ppant-Transferase der energetische Grundzustand [93].

Die innerhalb der letzten Jahre hauptsächlich über NMR-Untersuchungen erhaltenen Informationen über die dynamischen Eigenschaften von NRPS-Domänen erlaubten die Aufstellung eines dynamischen Modells der Protein-Protein Interaktion in diesen Systemen [261]. Hier befinden sich sowohl die Carrier-Proteindomäne, als auch die interagierenden katalytischen Domänen in einem konformationellen Austausch zwischen unterschiedlichen Strukturen. In nur jeweils einer Konformation präsentieren beide Domänen die Kontaktflächen, die über die richtige Form und Ladungsverteilung komplementär zueinander passen (s. Abbildung 62). Diese Auswahl aus dem dynamischen Ensemble wird durch die Bindung der Domänen stabilisiert und führt zu einer produktiven Protein-Protein Wechselwirkung.



**Abbildung 62 Schematische Darstellung des dynamischen Modells der Protein-Protein Interaktion in NRPS Enzymen.** Eine produktive Domänen-Interaktion zwischen der PCP-Domäne und einer katalytischen Domäne entsteht aus der Auswahl von zueinander komplementären Konformationen jeweils aus einem Ensemble von unterschiedlichen Strukturen im dynamischen Gleichgewicht. Die Bindung der Domänen stabilisiert die passende Konformation und verschiebt das Gleichgewicht in diese Richtung (reproduziert nach [261]).

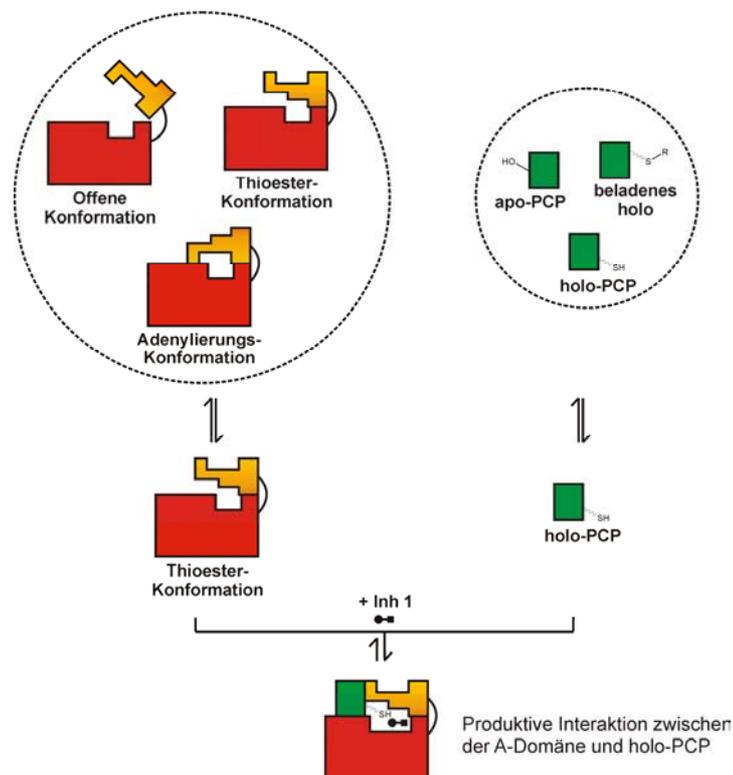
Wie schon in der Einleitung beschrieben können auch NRPS A-Domänen drei unterschiedliche Konformationen während der Katalyse einnehmen (s. Kapitel 1.1.4). Einer offenen Konformation stehen zwei geschlossene Strukturen gegenüber, die sich durch eine Drehung der kleineren Subdomäne voneinander unterscheiden. Yonus *et al.* postulierten, dass keine großen Energiedifferenzen zwischen den unterschiedlichen Zustände vorhanden sind, da alle geladenen Aminosäuren während des Katalysezyklus ihre Partneramino-säuren

austauschen und dadurch keine energetisch ungünstige Situation eintritt [55]. Die in dieser Arbeit durchgeführten partiellen Proteolyse-Experimente (s. Kapitel 4.1.1) eines A-PCP Didomänenkonstrukts aus einem NRPS Initiationsmodul in Abwesenheit von Substraten oder Inhibitoren lassen sich mit dem Bild einer intrinsisch flexiblen A-Domäne erklären. Hier wechselt das Protein zwischen unterschiedlichen Konformationen hin und her und ist damit gut zugänglich für die Protease Trypsin. Bei Bindung der Substrate oder des **Inh 1** verändert sich die Energielandschaft der A-Domäne und eine der geschlossenen Konformationen ist bevorzugt. Die große Resistenz des gesamten Proteins bei Inkubation mit **Inh 1** deutet auf eine effektive Fixierung der A-Domäne in einer definierten Struktur, der Thioesterkonformation, hin.

Die in dieser Doktorarbeit erhaltenen Ergebnisse analysieren die Wechselwirkungen zwischen der A- und der PCP-Domäne. Sie lassen sich vollständig in Einklang mit dem in Abbildung 62 präsentierten Modell der NRPS Domäneninteraktion bringen. Durch Untersuchung des GrsA A-PCP Didomänenkonstrukts mittels biochemischer Assays konnte aufbauend auf früheren Untersuchungen folgendes Modell für diese Interaktion aufgestellt werden (s. Abbildung 63). Sowohl die A-, als auch die PCP-Domäne liegen in einem Ensemble an unterschiedlichen Konformationen vor. Die Zugabe von Substraten und Inhibitoren und der post-translationale Zustand der PCP-Domäne beeinflussen diese konformationellen Gleichgewichte. Für eine produktive Interaktion zwischen einer A- und einer PCP-Domäne müssen zwei Voraussetzungen erfüllt sein. Zum einen muss sich die A-Domäne in der Thioesterkonformation befinden. Zum anderen muss die PCP-Domäne in der freien holo-Form vorliegen. Diese Bedingungen sind in dem Komplex zwischen holo-A-PCP und **Inh 1** gegeben. Der erhaltene Komplex nimmt in den Gelfiltrations- und nativen PAGE-Assays (s. Kapitel 4.1.2) die kompakteste Konformation ein, was mit einer korrekten Wechselwirkung der Domänen erklärt werden kann. Die chemischen Markierungs-Experimente (s. Kapitel 4.1.3) bestätigen die Annahme, dass in dieser Situation der PCP-Ppant-Arm vom Lösungsmittel abgeschirmt ist und sich im aktiven Zentrum der A-Domäne befindet. In allen anderen Fällen ist wahrscheinlich eine dissoziierte Form der beiden Domänen der energetische Grundzustand. Liegt in dem A-PCP Didomänenprotein die PCP-Domäne entweder in der apo- oder in einer modifizierten holo-Form vor, so kann sich die produktive Domäneninteraktion nicht ausbilden. Entsprechend ist die durch den **Inh 1** stabilisierte Thioesterkonformation der A-Domäne für die korrekte Wechselwirkung notwendig. Durch Kristallstrukturanalysen konnte gezeigt werden, dass sich nur in dieser Konformation ein Ppant-Bindungstunnel zum aktiven Zentrum bildet (z.B. [53]). In der

Adenylierungskonformation der A-Domäne, die durch das ATP-Analogon ApCpp und L-Phe stabilisiert wird, ist der Ppant-Arm nicht gut vom Lösungsmittel geschützt und damit kann hier wahrscheinlich auch die holo-PCP-Domäne nicht korrekt an die A-Domäne binden. Alternativ könnten bei der Möglichkeit einer richtigen Bindung der beiden Domänen größere on/off Raten der holo-PCP-Domäne ebenfalls dieses Verhalten erklären.

Das in dieser Arbeit entwickelte und in Abbildung 63 dargestellte Modell erweitert das dynamische Modell der Protein-Protein Interaktionen in NRPS Enzymen auf die Wechselwirkung einer A- mit einer PCP-Domäne. Dabei wurde erstmalig ein Teilschritt des Katalysezyklus zum Einbau eines peptidischen Bausteins in das nicht-ribosomale Produkt untersucht und der Einfluss von Substraten und Inhibitoren berücksichtigt.



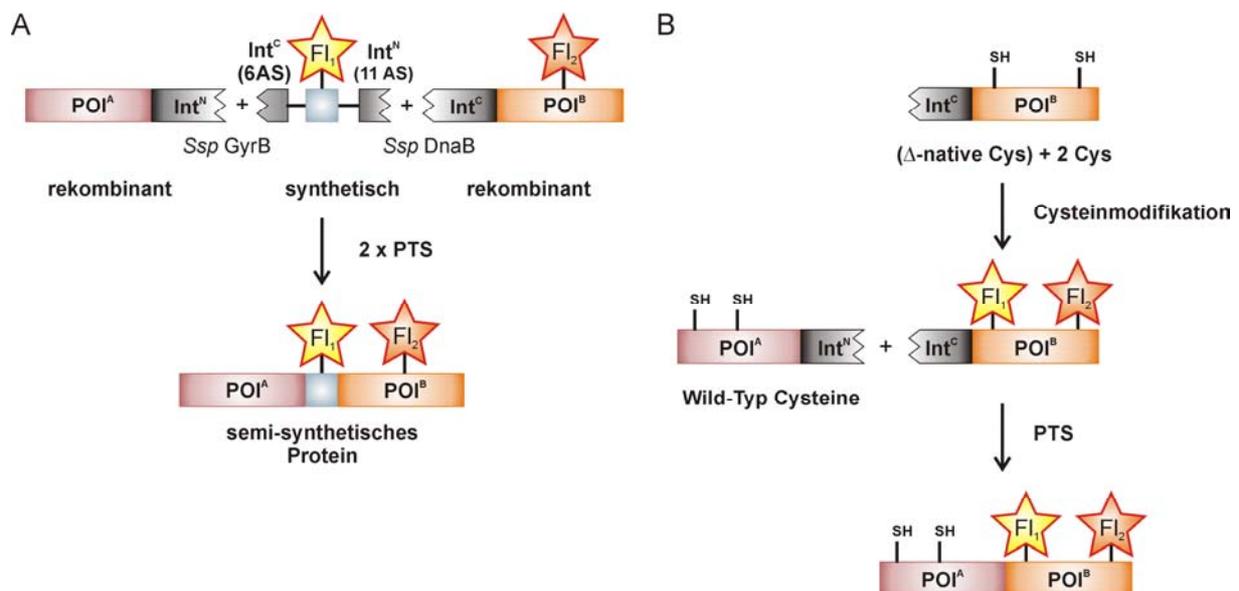
**Abbildung 63 In dieser Arbeit aufgestelltes Modell zur Interaktion von A- und PCP-Domänen in NRPS-Systemen.** NRPS A-Domänen sind dynamische Proteine und existieren in einem Gleichgewicht aus drei unterschiedlichen Konformationen. Dieses Gleichgewicht wird stark von der Abwesenheit und Gegenwart von Substraten und Inhibitoren beeinflusst; so stabilisiert **Inh 1** die Thioesterkonformation und die Kombination aus dem ATP Analogon ApCpp und L-Phe die Adenylierungskonformation. Die PCP-Domäne liegt in unterschiedlichen post-translationalen Modifikationen vor (apo, holo, beladenes holo). Eine produktive Interaktion der beiden Domänen ist nur in der Thioesterkonformation der A-Domäne und mit der holo-Form der PCP-Domäne möglich. Diese Interaktion wird in der Reaktionssequenz durch das Aminoacyladenylat oder artifizell durch den **Inh 1** stabilisiert. Unterschiedliche konformationelle Zustände der PCP-Domäne konnten mit den in dieser Arbeit verwendeten Methoden nicht aufgelöst werden. Dieses Modell erweitert das in Abbildung 62 dargestellte Schema zur NRPS Domäneninteraktion um einen Teilschritt des Katalysezyklus und berücksichtigt zum ersten Mal den Einfluss von Substraten und Inhibitoren.

Die in dieser Arbeit durchgeführten Untersuchungen zeigen, dass die Wechselwirkungen der PCP-Domäne mit anderen NRPS Domänen von ihrem post-translationalen Zustand abhängig sind. So reagiert apo-PCP anders mit einer inhibitor-gebundenen A-Domäne als holo-PCP. Zwei Szenarien könnten zur Erklärung herangezogen werden. Zum einen könnte für die Bindung der PCP-Domäne der Ppant-Arm die kritische Komponente sein, indem dessen Wechselwirkung mit der A-Domäne hauptsächlich den produktiven Kontakt der beiden Domänen stabilisiert. Zum anderen wäre es möglich, dass eine apo-PCP-Domäne nicht die richtige Konformation zur Bindung an die A-Domäne einnehmen kann und dies nur der holo-PCP-Domäne möglich ist. In beiden Fällen beeinflusst der post-translationalen Zustand der PCP-Domäne deren Affinität zu den interagierenden Domänen. Analysiert man die früheren Arbeiten an NRPS Systemen, so kann man postulieren, dass dies auch für die weiteren Schritte im Katalysezyklus der Fall und allgemein gültig ist. So wird nach der Thioesterbildung wahrscheinlich eine mit der Aminosäure beladene PCP-Domäne den Kontakt zur A-Domäne wieder verlieren. In GrsA und in zu GrsA analogen Proteinen, wie TycA, befindet sich das beladene PCP wahrscheinlich viel häufiger an der Epimerisierungs-Domäne als an der A-Domäne des Initiationsmoduls. Dies wird durch die schnelle permanente Umwandlung vom L- zum D-Phe-Ppant-Thioester und zurück dokumentiert [41]. In Elongationsmodulen besitzt die nach der produktiven Interaktion mit der A-Domäne beladene PCP-Domäne eine Affinität zur Akzeptorposition der C-Domäne des eigenen Moduls [44]. Linne *et al.* zeigten elegant, dass dieser Mechanismus eine verfrühte Initiation der NRP Biosynthese an internen Modulen verhindert [44]. Es sind jedoch zweifelsfrei noch weiterführende Experimente notwendig, um den exakten Einfluss der post-translationalen Modifikation auf die Struktur und die Affinität der PCP-Domäne zu bestimmen.

### 5.1.2 Zukünftige Arbeiten an NRPS-Proteinen

Die in NRPS Proteinen während der Katalyse auftretenden, hier postulierten Rang 0 Übergänge (s. oben und Abbildung 61) machen die Multidomänenproteine zu viel versprechenden Kandidaten für die Einzelmolekülfluoreszenzspektroskopie (smFRET-Spektroskopie). Auf dieser Ebene lassen sich sowohl die thermodynamischen Eigenschaften der Zustandsverteilungen als auch die kinetische Umwandlung der verschiedenen Zustände ineinander beobachten. Für den selektiven Einbau zweier unterschiedlicher Fluorophore an definierten Positionen in holo-GrsA A-PCP wurden in dieser Arbeit bereits Vorarbeiten geleistet (s. Kapitel 4.4.1). Diese Vorarbeiten sind erfolgsversprechend und sollten weiter

verfolgt werden. Eine große Zahl an alternativen Ansätzen aus dem Werkzeugkasten der Chemischen Biologie könnten weiterhin zum selektiven Einbringen zweier Fluorophore in NRPS-Proteine angewendet werden. Zwar existieren einige nicht auf Inteinen basierende Techniken, jedoch sollen als Ausblick im Folgenden zwei neue Strategien vorgestellt werden, die sich gespaltenen Inteinen bedienen. Das gezielte Einbringen eines Fluorophors in die Mitte eines Proteins könnte semi-synthetisch über die Verwendung der künstlich gespaltenen *Ssp* GyrB und *Ssp* DnaB Inteine erfolgen. Bei dieser Drei-Fragmentligation würden die N- und C-terminalen Fragmente durch rekombinante Genexpression in *E. coli* und das mittlere Segment mit der Modifikation über Festphasenpeptidsynthese hergestellt werden (s. Abbildung 64A). Die anzunehmende Orthogonalität der Inteine würde eine Ein-Topfsynthese des vollständigen Proteins erlauben. Ein zweites Fluorophor könnte unter anderem, wie in Abbildung 64A dargestellt, durch eine Cysteinmodifikation in Segment 3 eingebaut werden.



**Abbildung 64** Mögliche, neue intein-basierte Strategien zur Einbringung zweier Fluorophore in Zielproteine. A) Drei-Fragment-Ligation basierend auf den künstlich gespaltenen *Ssp* GyrB und DnaB Inteinen zur selektiven Modifikation eines inneren Segment des Proteins; der Einbau des zweiten Fluorophors könnte über chemische Markierung eines Cysteinrestes erfolgen; B) Das unselektive Einbringen zweier Fluorophore an zwei unnatürliche Cysteine in einem größeren Zielprotein kann durch eine Inteinstrategie vereinfacht werden; wird das Protein segmentell über PTS aufgebaut, so müssen nur in einem Segment die natürlichen Cysteine entfernt werden.

Die in Kapitel 4.4.1 vorgestellte Methode zum unselektiven Einbringen zweier Fluorophore mittels zweier unnatürlicher Cysteine trifft bei der Untersuchung größerer NRPS-Proteine, wie komplette Module oder multimodularer Proteine, auf folgendes Problem. Die notwendigen Mutationen der natürlichen Cysteinreste sind zum einen zeit- und arbeitsintensiv und zum anderen kann ein Aktivitätsverlust durch diese Mutationen nicht ausgeschlossen werden. Eine intein-vermittelte Ligation des zu untersuchenden Proteins

reduziert dieses Problem auf ein kleineres Fragment des Proteins. Wie in Abbildung 64B dargestellt, müssen nur in Segment 2 die natürlichen Cysteine entfernt und zwei neue Cysteine in solvenz-exponierte Positionen eingebracht werden. In Segment bleibt die Wildtyp-Sequenz erhalten. Analog zur Cys-Tag Strategie [164,165] können für diesen Ansatz sowohl das *Ssp* DnaB, als auch das *Mxe* GyrA Intein verwendet werden. Für beide Inteine ist es mit dem SPLICEFINDER-System möglich, schnell und unkompliziert eine aktive Insertionsposition in dem zu untersuchenden Protein zu finden.

Zwar erlaubt die smFRET-Spektroskopie die direkte Untersuchung der Dynamik der Katalyse bei Zugabe von Substraten, jedoch vereinfacht das Arretieren zu verschiedenen Zeitpunkten im Laufe des Katalysezyklus die Analyse der konformationellen Änderungen. Die Fixierung der produktiven Interaktion einer A- und einer PCP-Domäne gelang in dieser Arbeit durch die Zugabe des nicht hydrolysierbaren Amino-Acyladenylatanalogon (**Inh 1**). Der Zustand vor der Adenylierungsreaktion ist durch die Addition des nicht-hydrolysierbaren ApCpp und der Aminosäure zu erreichen. Weitere Zwischenstufen der Katalyse lassen sich möglicherweise durch nicht hydrolysierbare Amino-Ppant Analoga stabilisieren und bevorzugt untersuchen [262]. Mit Hilfe einer Kombination von spektroskopischen Methoden, wie smFRET, und Techniken zur selektiven Modifikation von Proteinen sollte es in Zukunft möglich sein, die Dynamiken und zugrunde liegenden Prinzipien der NRP Biosynthese zu entschlüsseln und besser zu verstehen.

## **5.2 Neue pharmazeutische Wirkstoffe durch Manipulation der NRPS-Biosynthesewege**

Wie in der Einleitung beschrieben werden die Herausforderungen an den Antibiotika-Markt und die zugehörige Forschung in den nächsten Jahren wachsen. Eine immer größere Zahl an bakteriellen Infektionen kann nicht mehr mit Standardtherapien behandelt werden. Resistenzen gegen mehrere Antibiotika breiten sich vor allem in den Intensivstationen der Krankenhäuser vermehrt aus und verursachen zusätzlich hohe Kosten im Gesundheitssystem. Das Problem wird durch eine unsachgemäße Anwendung von Antibiotika, weltweites Reisen, mangelnde Hygiene und einen wachsenden Anteil älterer und immunsupprimierter Menschen noch verstärkt. Neben effizienteren Maßnahmen zur Infektionskontrolle und Förderinitiativen für den sachgemäßen Umgang von Antibiotika werden neue antibakterielle Wirkstoffe dringend benötigt [263]. Naturstoffe, wie die nicht-ribosomalen Peptide, sind immer noch eine wichtige Quelle für pharmazeutisch interessante Moleküle. Interessanterweise sind weniger als 1 % der bekannten organischen Verbindungen Naturstoffe; 99 % sind Synthetika. Jedoch basierten im Zeitraum von 1984 bis 2004 75 % der eingereichten neuen Wirkstoffe gegen antibakterielle Erkrankungen auf Naturstoff-Leitstrukturen [263]. Das Screening von Naturstoffbibliotheken auf pharmakologische Aktivität ist meist weit erfolgreicher als das Screening von synthetisch erzeugten Bibliotheken [264]. Es ist nicht überraschend, dass die hauptsächlich von Pilzen und Bakterien hergestellten Naturstoffe besonders gute Leitstrukturen für Antibiotika darstellen. In den Jahrmillionen der Evolution entwickelten diese Produzentenstämme ein effektives, chemisches Angriffssystem gegen bakterielle Konkurrenten, um sich einen Vorteil im Kampf um Ressourcen zu verschaffen [263]. Jedoch besitzen diese evolutionär getesteten Verbindungen nicht zwangsläufig die herausragenden Eigenschaften von Medikamenten. Sie können deswegen oft durch strukturelle Veränderungen in Bezug auf Verfügbarkeit und Distribution im Patienten, sowie Toxizität und Stabilität verbessert werden. Des Weiteren erlaubt eine Modifikation des ursprünglichen Naturstoffs oft die für die pharmazeutische Industrie so wichtige Patentierbarkeit. Aus all diesen Gründen ist die Modifikation von Naturstoffen ein wichtiger Zweig bei der Entwicklung von antibakteriellen Medikamenten.

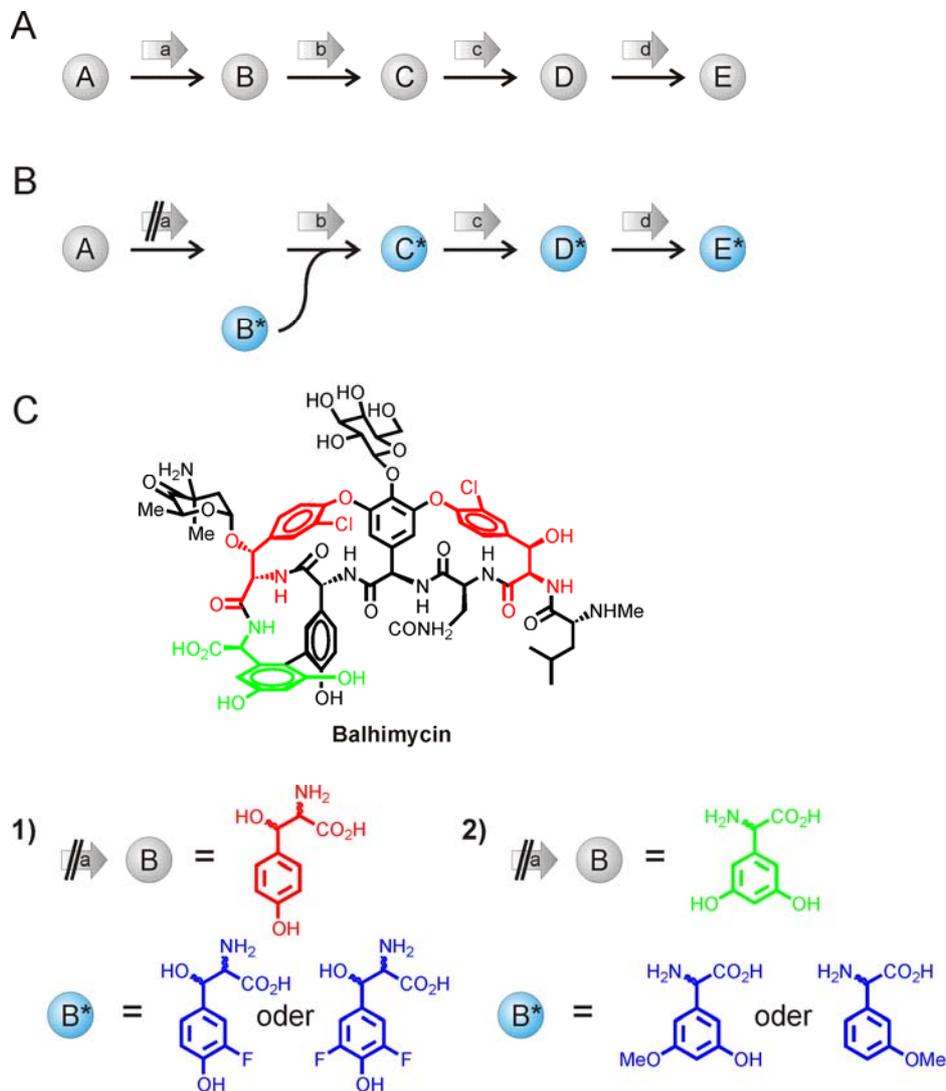
Prinzipiell sind drei verschiedene konzeptionelle Zugänge zu veränderten Naturstoffbibliotheken denkbar. Die chemische Totalsynthese besitzt das größte Potential zum Einbau gezielter Modifikationen in den Naturstoff. Die Aufklärung einer Vielzahl an Struktur-Wirkungs-Beziehungen von Naturstoffen konnte mit komplett synthetischen

Molekülen erreicht werden. Die chemische Komplexität einer Naturstoffstruktur verlangt jedoch oft eine lange, mehrstufige Synthese und erschwert damit die Produktion im größeren Maßstab. Bei der Semisynthese oder der Derivatisierung eines Naturstoffes greift man auf die biologisch hergestellte Substanz oder deren Vorläufer zurück und modifiziert ihn anschließend mit chemischen Methoden. Obwohl auch hier bereits eine Vielzahl an Erfolgen verzeichnet werden konnte, so sind die Modifikationen natürlich auf die vorhandenen funktionellen Gruppen des Naturstoffs begrenzt. Dies limitiert die Variationsmöglichkeiten bei dem Aufbau der Substanzbibliotheken. Sowohl die chemische Totalsynthese, als auch die Semisynthese sind enorm wichtige Standbeine der Wirkstoffforschung, jedoch soll in Folge in dieser Diskussion eine alternative Strategie, die Generierung neuer Verbindungen durch Manipulation der Biosynthesewege (*engl.* metabolic engineering) am Beispiel von NRPS-Reaktionen im Fokus stehen.

Bei dem „metabolic engineering“ verändert man gezielt auf DNA-Ebene die für die Naturstoffsynthese verantwortlichen Gene eines Produzentenstammes. Die veränderte enzymatische Zusammensetzung im Organismus führt zu der Produktion einer neuen Variante des ursprünglichen Naturstoffes. Liegt nach der zeit- und kostenintensiven Herstellung eines Mutantenstammes dieser erstmals vor, so kann relativ leicht durch Fermentation eine große Menge des gewünschten Biosyntheseprodukts erhalten werden. Von entscheidender Bedeutung ist die Auswahl des Produzentenstammes. Eine genetische Manipulation ist oft in heterologen Systemen, wie in *E. coli*, am leichtesten durchführbar. Jedoch existieren hier oft die Biosynthesewege für Vorläufermoleküle nicht. Diese sind zwar in den eigentlichen Produzenten des Naturstoffs vorhanden. Für diese liegen aber nicht immer effiziente Systeme zur Aufzucht und genetischen Manipulation vor. An drei Beispielen zur Manipulation von NRPS-Biosynthesewegen soll das noch junge Forschungsfeld des „metabolic engineering“ mit seinen Möglichkeiten und Limitationen im Folgenden vorgestellt werden.

Die natürliche Biosynthese eines Naturstoffs geht von einem Vorläufermolekül aus und dieses wird sukzessive durch enzym-katalysierte Reaktionen über verschiedene Zwischenstufen aufgebaut und verändert (s. schematische Darstellung in Abbildung 65A). Die so genannte Mutasyntese greift in den natürlichen Ablauf durch Deletion eines Gens im Stadium der Vorläuferbiosynthese ein. Diese Blockmutation verhindert die endogene Zulieferung des benötigten Bausteins B und stoppt damit die Produktion des Naturstoffs E. Durch Zugabe von synthetisch veränderten Bausteinen B\* zu diesen Blockmutanten entstehen neue Naturstoffderivate E\*, solange die unnatürlichen Zwischenprodukte von dem weiteren

enzymatischen Biosyntheseweg akzeptiert und prozessiert werden (s. schematische Darstellung in Abbildung 65B).

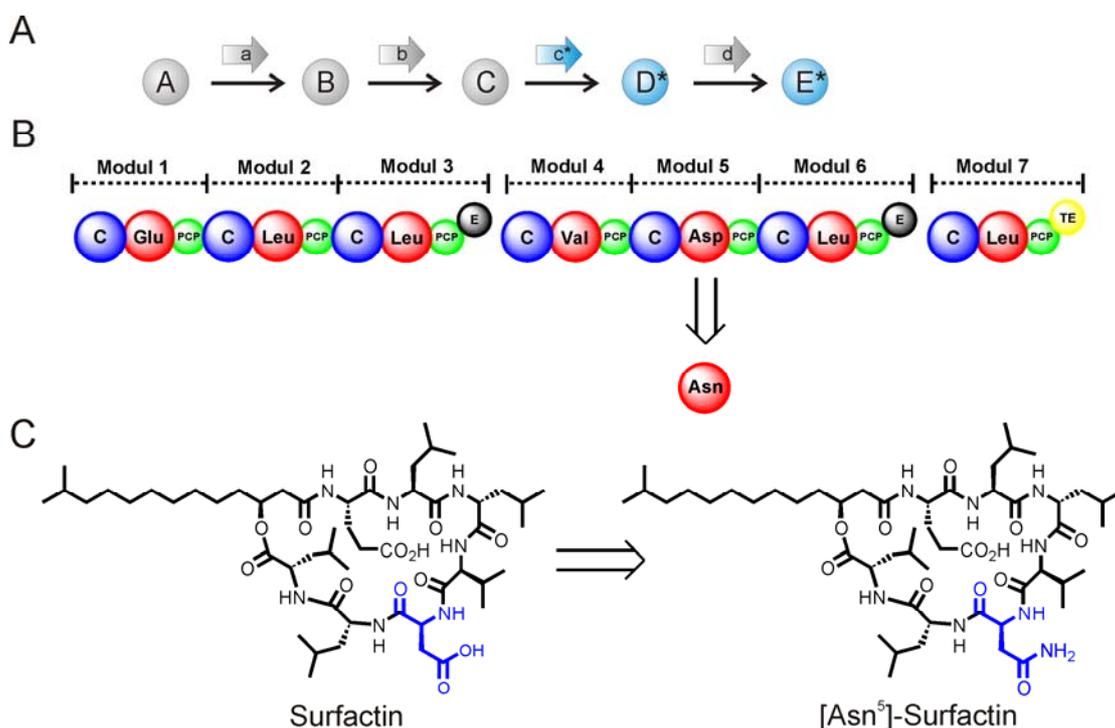


**Abbildung 65** Schematische Darstellung der Mutasynthese zur biosynthetischen Herstellung von neuen Naturstoffen und das Beispielmolekül Balhimycin. A) Schematische Darstellung eines natürlichen Biosynthesewegs ausgehend von dem Vorläufermolekül A hin zu dem Naturstoff E, die Pfeile stellen enzymatische Transformationen dar; B) Schema der Mutasynthese: Durch Ausschalten eines Gens am Anfang der Biosynthese (a) wird die Produktion des Intermediats B verhindert, die Zugabe des synthetischen Bausteins B\* kann zur Bildung eines modifizierten Naturstoffs E\* führen; C) Struktur von Balhimycin und Mutasynthese von Derivaten dieses Naturstoffs in *Amycolatopsis balhimycina*, 1) Bei Deletion der Biosynthese von  $\beta$ -Hydroxytyrosin (rot) können die in blau dargestellten fluorinierten Bausteine erfolgreich eingebaut werden [265], 2) Ebenso gelingt die Supplementierung einer in der Synthese von 3,5-Dihydroxyphenylglycin (grün) blockierten Mutante mit den in blau gezeigten Derivaten [266].

Erfolgreich konnten Süssmuth, Wohlleben und Mitarbeiter zwei Mutasynthesen am Beispiel des mit Vancomycin verwandten Antibiotikums Balhimycin durchführen [265,266]. Balhimycin unterscheidet sich von dem wichtigen Notfallantibiotikum Vancomycin nur durch ein unterschiedliches Glykosylierungsmuster. Die Autoren stellten ausgehend vom Wildtyp-Stamm *Amycolatopsis balhimycina* eine Deletionsmutante  $\Delta bhp$  für die Synthese von  $\beta$ -

Hydroxytyrosin her [265]. Dieser Baustein findet sich später zweimal in den 3-Chlor- $\beta$ -Hydroxytyrosin-Teilstrukturen des trizyklischen Aglycons wieder (s. rote Strukturen in Abbildung 65C). Die Deletions-Nullmutante  $\Delta bhp$  ist nicht mehr zur Balhimycin-Produktion fähig [265]. Durch Zufütterung fluorierter  $\beta$ -Hydroxytyrosine konnten neue antibiotische Verbindungen isoliert werden. Zusätzliche Diversität entstand durch Verwendung einer in der Biosynthese von 3,5-Dihydroxyphenylglycin blockierten Mutante [266]. Diese Komponente ist nach erfolgreicher Synthese im AB-Ring des Antibiotikums zu finden (s. grüne Struktur in Abbildung 65C). Die  $\Delta dpgA$  Blockmutante von *A. balhimycina* konnte bei Zugabe unterschiedlicher synthetischer Vorläufermoleküle weitere veränderte Naturstoffprodukte herstellen [266]. Es bleibt anzumerken, dass die Mutasynthese nur in, für den Produzenten nicht essentielle, Biosynthesewege eingreifen kann und das System stark von der Substratspezifität der weiteren enzymatischen Reaktionen abhängig ist.

Eine direkte Veränderung des Proteintemplats zur Generierung neuer Peptidantibiotika gelang Marahiel und Mitarbeitern 2002 in dem Surfactin-Biosyntheseweg [267]. Sie nutzten geschickt den in früheren Arbeiten bestimmten Selektivitätscode der A-Domänen [61] aus und veränderten durch eine Punktmutation (His322  $\rightarrow$  Glu) die A-Domäne des fünften NRPS-Moduls. Die so erzeugte Mutante aktivierte anstatt Aspartat nun bevorzugt Asparagin. Zusätzlich akzeptierten die folgenden Module das veränderte Substrat und der mutierte *Bacillus subtilis* Stamm produzierte das vorausgesagte [Asn<sup>5</sup>]-Surfactin (s. Abbildung 66) [267].

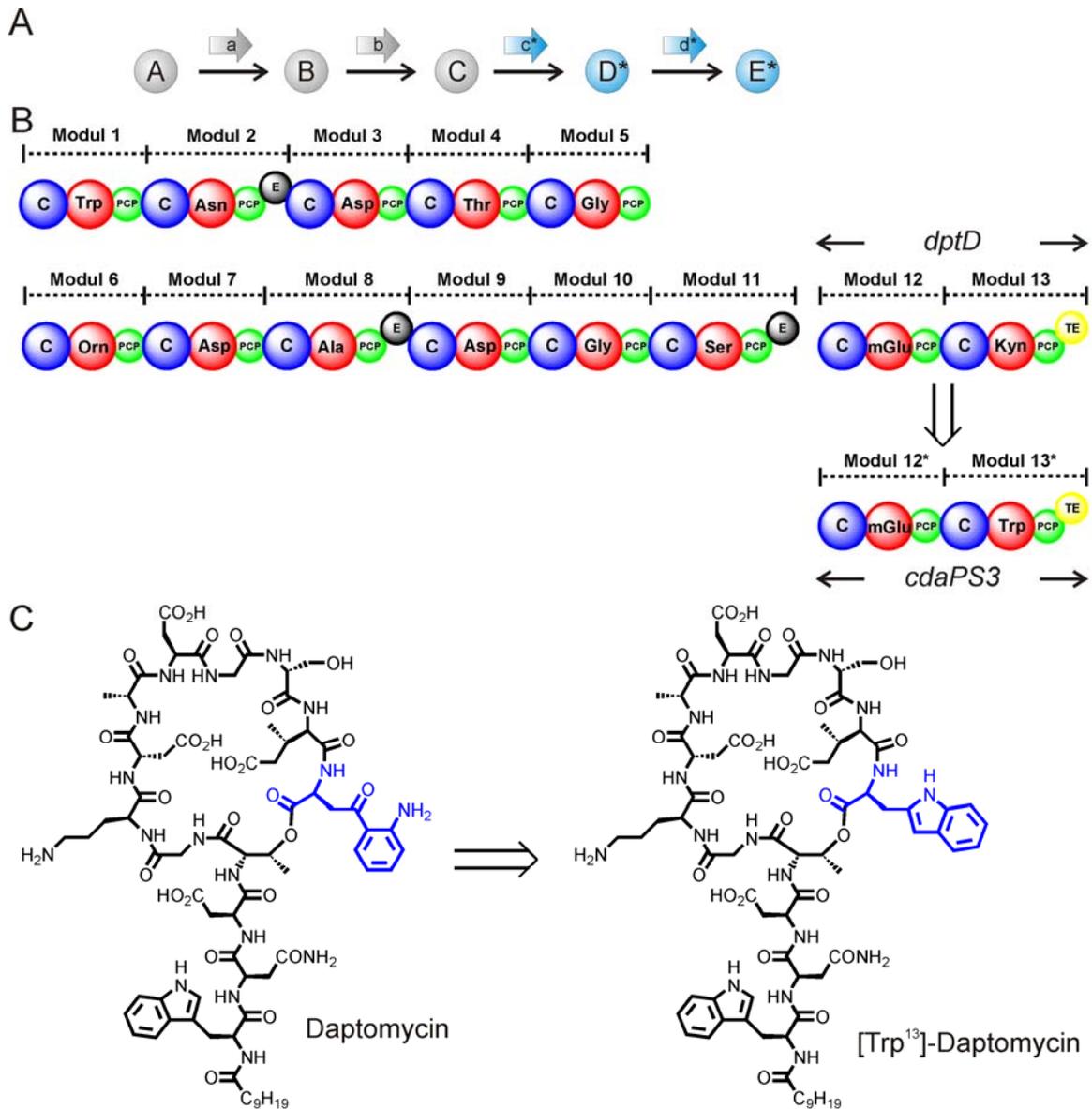


**Abbildung 66** Mutation des Surfactin-Biosyntheseweges in *Bacillus subtilis* [267]. A) Schematische Darstellung: Eine Änderung des enzymatischen Transformationsschritts (c\*) auf genetischer Ebene kann bei

Substrattoleranz der Folgeschritte einen modifizierten Naturstoff E\* erzeugen; B) Der NRPS-Biosynthesecluster zur Synthese von Surfactin, Eppelmann *et al.* veränderten durch Punktmutation die Substratspezifität der A-Domäne des fünften Moduls von Asp zu Asn; C) Chemische Strukturen von Surfactin und des erzeugten [Asn<sup>5</sup>]-Surfactinderivats.

Mit der Arbeit von Eppelmann *et al.* konnte gezeigt werden, dass die Herstellung neuer Antibiotika über eine Veränderung der A-Domänenselektivität durchführbar ist. Allerdings sind die Variationen im Substratspektrum der A-Domäne wahrscheinlich auf kleinere Änderungen in der Form und/oder der Polarität der Aminosäure beschränkt [267]. Größere Eingriffe in die NRPS-Synthesegene wurden bereits früher in der Arbeitsgruppe Marahiel durch Modulfusion und -deletion am Surfactin-Biosynthesecluster durchgeführt [268,269]. Wissenschaftler der Firma *Cubist Pharmaceuticals* manipulierten ebenfalls erfolgreich den Biosyntheseweg des zyklischen anionischen Lipopeptids Daptomycin durch Genaustausch (für Übersichtsartikel s. [270,271]). Daptomycin ist aus dreizehn Aminosäuren aufgebaut, von denen zehn an der Bildung eines über eine Esterbindung geschlossenen Rings beteiligt sind. Es wird zur Behandlung von Haut- und Hautstrukturinfektionen, die von gram-positiven Pathogenen wie *Staphylococcus aureus* ausgelöst werden, eingesetzt. Aufgrund des erschwerten, synthetischen Zugangs zu Vorläufer-Aminosäurebausteinen (3-Methyl-Glutaminsäure) ist ein vollsynthetischer Zugang zu Daptomycin und analogen Verbindungen nicht leicht durchführbar [272]. Eine Alternative ist die Manipulation des NRPS-Biosyntheseweges in dem Wildtyp-Produzenten *Streptomyces roseosporus* durch Multidomänen-, Modul- oder Multimodulaustausch (s. Abbildung 67A+B). Als ein Beispiel der Arbeit von Baltz und Mitarbeitern sei der Austausch der letzten zwei Module der Daptomycin-Biosynthese (kodiert von *dptD*) gegen das vergleichbare NRPS-Gen *cdaPS3* der Biosynthese des Lipopeptids CDA aus *Streptomyces coelicolor* erwähnt (s. Abbildung 67B) [273]. In dem *dptD*-defizienten *S. roseosporus* Stamm konnte durch Integration des *cdaPS3*-Gens die Produktion des vorhergesagten [Trp<sup>13</sup>]-Daptomycin wieder komplementiert werden [273] (s. Abbildung 67C). Die Verwandtschaft der NRPS-Proteine erlaubte funktionale Protein-Protein-Wechselwirkungen und eine geringe Substratspezifität der CDA-NRPS Enzyme ermöglichte die Prozessierung der unnatürlichen Substrate. Allerdings war die Ausbeute des biosynthetisch hergestellten neuen Naturstoffs um ca. 50 % im Vergleich mit der Wildtyp-Produktion reduziert. Zusammen mit ausgetauschten Modulen an den Positionen 8 und 11 von Daptomycin konnte eine große Zahl an Varianten des Daptomycin-Grundgerüsts erzeugt werden, von denen einige eine vergleichbare antibiotische Aktivität wie der eigentliche Naturstoff aufwiesen [271]. Die reduzierten Ausbeuten führten die Autoren

auf die nicht-natürliche und damit vermindert aktive Kommunikation zwischen den NRPS-Untereinheiten zurück.



**Abbildung 67 Mutation des Biosyntheseweges von Daptomycin in *Streptomyces roseosporus*.** A) Schematische Darstellung: Im Vergleich zu Abbildung 66A wird nicht nur ein Teilschritt, sondern gleich mehrere Schritte genetisch verändert; B) Der NRPS-Biosynthesekluster zur Synthese von Daptomycin, Miao *et al.* tauschten die beiden letzten Module durch das analoge Gen des CDA-Clusters aus *Streptomyces coelicolor* aus [273]; C) Chemische Strukturen von Daptomycin und seinem Derivat [Trp<sup>13</sup>]-Daptomycin.

Gerade diese Ergebnisse verdeutlichen die Wichtigkeit der in dieser Arbeit durchgeführten Studien zum besseren Verständnis der NRPS-Enzyme. Sind die essentiellen Wechselwirkungen der einzelnen Domänen auf molekularer Ebene im Detail verstanden, so sollte auch eine effizientere Veränderung der Biosynthesewege möglich sein. Aktuell sind zur Herstellung eines bestimmten Wirkstoffs die Manipulation der NRPS-Biosynthesewege und die Fermentation des Naturstoffs eine sinnvolle Alternative zu total- und semisynthetischen

Ansätzen. Durch ein besseres Verständnis des Enzymtemplats wird die Biosynthese sicherlich auch in Zukunft einen wichtigen Stellenwert in der pharmazeutischen Forschung einnehmen. Ob eine wünschenswerte, kombinatorische Peptidsynthese mit NRPS-Enzymen zur Erzeugung sehr großer Substanzbibliotheken durchführbar ist, kann mit dem heutigen Wissensstand nicht vorhergesagt werden.

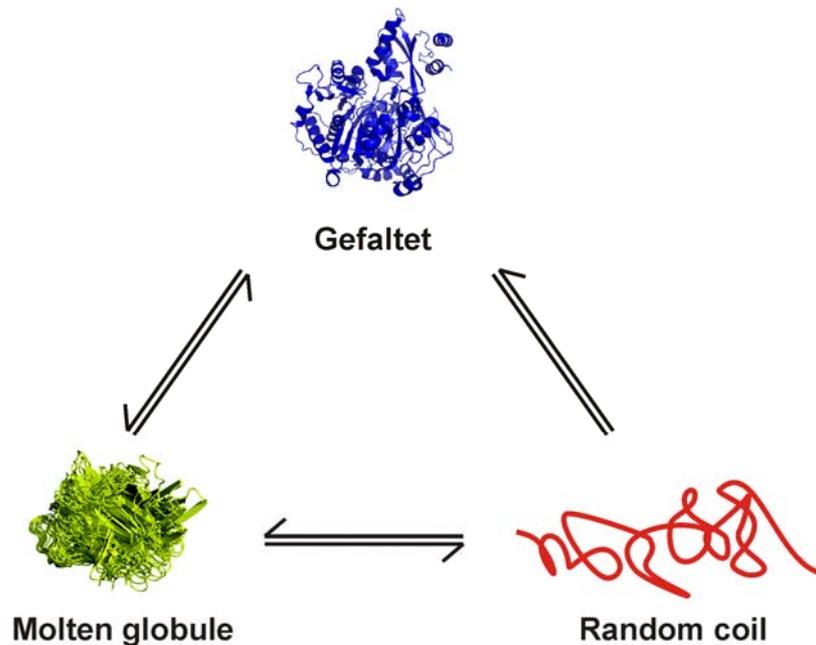
### **5.3 Gespaltene Inteine zur selektiven Modifikation von Proteinen**

Gespaltene Inteine sind ein wichtiges Werkzeug bei der Synthese von selektiv veränderten Proteinen (s. Kapitel 1.2.4). Diese Modifikationen sind für die exakte biophysikalische Charakterisierung und damit für das Verständnis der Proteine oft zwingend erforderlich. Allerdings werden gespaltene Inteine momentan nur in einer begrenzten Anzahl von biochemischen Laboratorien eingesetzt. Andere Ligationsreaktionen, wie die native chemische Ligation (NCL), finden in der wissenschaftlichen Gemeinschaft eine größere Verbreitung. Die Limitationen jedes gespaltenen Intein-Systems müssen separat betrachtet werden, jedoch sind hauptsächlich die Ausbeute und Geschwindigkeit der Reaktion, die Löslichkeit der Intein-Extein-Fusionsproteine und die Toleranz des Inteins gegenüber den Exteinsequenzen die entscheidenden Parameter. In der vorliegenden Arbeit wurden mit dem natürlich gespaltenen *Npu* DnaE Intein ein Intein biochemisch charakterisiert, welches in den Punkten Ausbeute und vor allem Geschwindigkeit der PTS-Reaktion eine deutliche Verbesserung zu den bisherigen Systemen darstellt. Die schnelle Kinetik, auch im Vergleich mit dem prominenteren *Ssp* DnaE Allel, wird in Kapitel 5.3.2 diskutiert. Auch die in dieser Arbeit vorgenommene Entwicklung des SPLICEFINDER-Systems zur schnellen Klonierung und Überprüfung von Intein-Insertionsstellen in Zielproteinen sollte die Einsatzbreite der gespaltenen Inteine weiter erhöhen. Die Thematik der Spleißfähigkeit in Abhängigkeit der fusionierten Exteinsequenzen ist Diskussionsgegenstand von Kapitel 5.3.3. NMR-Untersuchungen an gespaltenen Inteinhälften ohne das korrespondierende Partnerprotein ergaben, dass die individuellen Teile keine definierte Faltung aufweisen. Eine Diskussion über die Eigenschaften nativ ungefalteter Proteine wird im folgenden Kapitel 5.3.1 präsentiert.

### 5.3.1 Gespaltene Inteinhälften als Beispiele für nativ unstrukturierte Proteine

Vor allem im letzten Jahrzehnt wurde eine Vielzahl an Proteinen charakterisiert, die in ihrem nativen Zustand unter physiologischen Bedingungen keine definierte drei-dimensionale tertiäre und/oder sekundäre Struktur aufweisen [230,232]. Entgegen des lange akzeptierten Paradigmas, dass die Funktion eines Proteins an eine definierte Struktur gebunden ist, spielt die Klasse der intrinsisch ungefalteten Proteine (*engl.* intrinsically disordered proteins oder natively unfolded proteins) eine wichtige biologische Rolle vor allem in regulatorischen Signalnetzwerken. Diese Proteine komplementieren in ihrer Funktion oft die nativ gefalteten Proteine und sind bezogen auf die schiere Zahl keine Seltenheit. Nativ ungefaltete Proteine unterscheiden sich von gefalteten Proteinen anhand ihrer Aminosäurezusammensetzung, da sie eine hohe Nettoladung bei neutralem pH und einen geringen Anteil an hydrophoben Aminosäuren besitzen [274,275]. Diese Tatsache ermöglichte eine computergestützte Suche nach intrinsisch ungeordneten Proteinen in den vorhandenen Proteindatenbanken (s. unter anderem das Programm FoldIndex <http://biportal.weizmann.ac.il/fldbin/findex> [276]). Basierend auf diesen Algorithmen wird postuliert, dass in Säugetieren ~ 50% der Proteine lange (> 30 AS) ungefaltete Regionen aufweisen und dass ca. 25% komplett ungeordnet vorliegen [232]. Ungefaltete Proteine lassen sich weiterhin in die so genannten „Molten globule“ (MG) und „Random coil“ (RC) Proteine unterteilen. MG-Proteine [233-235] besitzen zwar Sekundärstrukturelemente und eine globuläre Struktur analog zu gefalteten Proteinen, jedoch befinden sich die hydrophoben Seitenketten im inneren Kern des Proteins in einem dynamischen, unstrukturierten Zustand. Dahingegen liegt ein RC-Protein in einem vollständig entfalteten und gestreckten Zustand ohne Sekundärstrukturelemente vor.

Experimentell lassen sich ungefaltete Proteine über eine Kombination von NMR-, CD- und 1-Anilin-8-naphthalen-sulfonat (ANS) Fluoreszenz-Spektroskopie, sowie über die Zugänglichkeit zu Proteasen und den Stokes Radius nachweisen und unterscheiden (für eine detaillierte Zusammenfassung s. [230] oder [277]). Mit der Bestimmung von aktiven MG- und RC-Proteinen unter physiologischen Bedingungen wurden somit zusätzlich zu definiert gefalteten Konformationen zwei weitere strukturell mögliche Zustände für funktionelle Proteine gefunden. Interessanterweise ist die Proteinfunktion nicht nur von einem der drei Zustände, sondern auch von den Übergängen zwischen den Zuständen abhängig. Z.B. gehen viele intrinsisch ungefaltete Proteine bei Bindung an ein Zielprotein in eine definierte Faltung über [230]. Dieses von Dunker *et al.* zum ersten Mal präsentierte neue Modell einer Kombination von Proteinstruktur und Proteinfunktion ist in Abbildung 68 gezeigt.



**Abbildung 68** Von Dunker *et al.* aufgestelltes Modell zur Beziehung von Proteinstruktur und -funktion. Die Proteinfunktion kann sowohl durch eine der drei strukturellen Formen (gefaltet - molten globule - random coil) als auch durch die Übergänge zwischen diesen Formen entstehen [230]. Dieses Modell widerspricht dem Paradigma, dass die Funktion eines Proteins an eine definierte, gefaltete Struktur gebunden ist. Das Bild des molten globule Proteins wurde aus [232] entnommen.

Die in dieser Arbeit durchgeführten NMR-Untersuchungen an individuellen Inteinhälften sind die ersten experimentellen Indizien dafür, dass auch gespaltene Inteine der Klasse der intrinsisch ungefalteten Proteine zuzuordnen sind (s. Kapitel 4.2.2). Anhand der Linienbreite und der Dispersität der Amidsignale in den  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektren konnte z.B. der N-terminalen Hälfte des *Npu* DnaE Inteins eine MG-Struktur zugeordnet werden. Ebenfalls gelang der Nachweis, dass die Inteinhälften bei Bindung an das korrespondierende Partnerprotein einen strukturellen Übergang von einem ungeordneten in einen gefalteten Zustand vollziehen. Interessanterweise scheint dieses strukturelle und dynamische Verhalten sowohl bei den natürlich, als auch bei den künstlich gespaltenen, aktiven Inteinen aufzutreten.

Ungefaltete Proteine werden in der Zelle im Vergleich zu den stabileren, gefalteten Proteinen schneller degradiert. Diesem Nachteil steht eine Reihe von bekannten Vorteilen [229,230] gegenüber, deren Relevanz für gespaltene Inteine in der Folge diskutiert werden soll. Für die Proteine, die in Prozesse wie die Zellregulation und Signalkaskaden verwickelt sind, erleichtert das Fehlen einer definierten Vorzugskonformation die Bindung an unterschiedliche Ziele. Post-translationale Modifikationen, wie Phosphorylierungen und Glykosylierungen, erlauben zusätzlich eine präzise Kontrolle und eine einfache Regulation des Bindeverhaltens. Dieses „one to many signalling“ Prinzip [278,279] wurde für gespaltene Inteine jedoch bisher nicht beobachtet, da diese nur mit ihrer komplementären Hälfte reagieren können. Genau diese Spezifität und Selektivität ist wahrscheinlich eng mit dem

gekoppelten Ereignis aus Bindung und Faltung verknüpft [229,280]. Die thermodynamischen Konsequenzen dieses Prozesses sind, dass zusätzlich freie Energie für den Faltungsübergang aufgebracht werden muss und damit die Komplexbildung unwahrscheinlicher wird. Nur eine maximale Komplementarität der beiden Hälften erlaubt eine energetisch günstige Assoziation und damit die hohe Spezifität. Die dafür benötigten Interaktionsflächen können im Gegensatz zu gefalteten, globulären Proteinen für ungefaltete Proteine signifikant größer sein. Im Einklang damit sind die beiden Hälften der sowohl natürlich gespaltene DnaE, als auch der künstlich an der Position der Endonuklease gespaltenen Inteine stark ineinander verzahnt und bilden viele attraktive Wechselwirkungen aus [137,146,168]. Eine weitere Eigenschaft der ungefalteten Proteine, die Tendenz zur Aggregation und Missfaltung, wird von den gespaltenen Inteinen geteilt. Die in früheren Arbeiten erzeugten künstlich gespaltenen Inteine konnten bei rekombinanter Genexpression in *E. coli* nach Zellaufschluss nur in der unlöslichen Fraktion erhalten werden [149,158]. In dieser Doktorarbeit wurden zum ersten Mal lösliche Aggregate eines natürlich gespaltenen Inteins nachgewiesen (s. Kapitel 4.2.1). Diese hochmolekularen Verbindungen weisen eine große Stabilität und nur eine minimale Freisetzung der Monomere auf. PTS-Ausbeuten unter 100% lassen sich teilweise durch dieses Aggregationsverhalten erklären. Eine Verbesserung der Ausbeute ist aber durch Isolierung der monomeren Spezies über Gelfiltrationsläufe möglich (s. Kapitel 4.2.1).

Die intrinsische Mobilität der beiden ungefalteten Inteinhälften könnte sich auch positiv auf die Geschwindigkeit der Assoziation auswirken. Gerade bei niedrigen Konzentrationen der Partnerproteine wäre eine schnelle Komplexbildung vorteilhaft um z.B. der proteolytischen Degradation zuvor zukommen. Shoemaker *et al.* konnten durch theoretische Berechnungen zeigen, dass ein unstrukturiertes Protein im Vergleich mit einem gefalteten Protein mit begrenzter konformationeller Flexibilität einen größeren Einfangradius in Bezug auf eine spezifische Bindestelle besitzt [281]. Schon in relativ großer Entfernung können sich aufgrund der großen Fluktuationen im ungefalteten Protein schwach bindende Kontakte aufbauen. Diese Kontakte vereinfachen die Ausbildung von Strukturen bei der konzertierten Bindung und Faltung [281]. Die Assoziationsgeschwindigkeit von komplexen Biomolekülen, wie Proteinen, ist theoretisch durch die Einstein-Smoluchowski nach oben begrenzt. Ist eine bestimmte, relative Orientierung der beiden Komplexpartner für eine produktive Interaktion notwendig, so kann dies die Assoziationsrate um drei bis sechs Zehnerpotenzen verlangsamen [282-284]. Dass das natürlich gespaltene *Ssp* DnaE Intein trotzdem mit Werten nahe dem Einstein- Smoluchowski Limit assoziiert [174], könnte auf eine Kombination aus

elektrostatischen Wechselwirkungen [157,174] und der konformationellen Flexibilität der ungefalteten Inteinhälften zurückzuführen sein.

Die exakten Assoziations- und Faltungsschritte bei der Bildung eines aktiven Inteincomplexes sind noch vollständig ungeklärt. Weitere Untersuchungen an dem in dieser Arbeit vorgestellten *Npu* DnaE Inteinssystem mittels NMR-Spektroskopie könnten diese Fragen beantworten. Die Vorarbeiten hatten gezeigt, dass durch Mutation der katalytischen Reste zumindest der erste Intein-Anlagerungskomplex für eine Analyse zugänglich ist. Im Gegensatz zu der Assoziation eines ungefalteten an ein strukturiertes Protein (s. z.B. [285] und [286] für Review) wäre hier die konzertierte Faltung und Assoziation zweier ungefalteter Proteinsegmente zu beobachten.

### 5.3.2 Die Geschwindigkeit der PTS-Reaktion

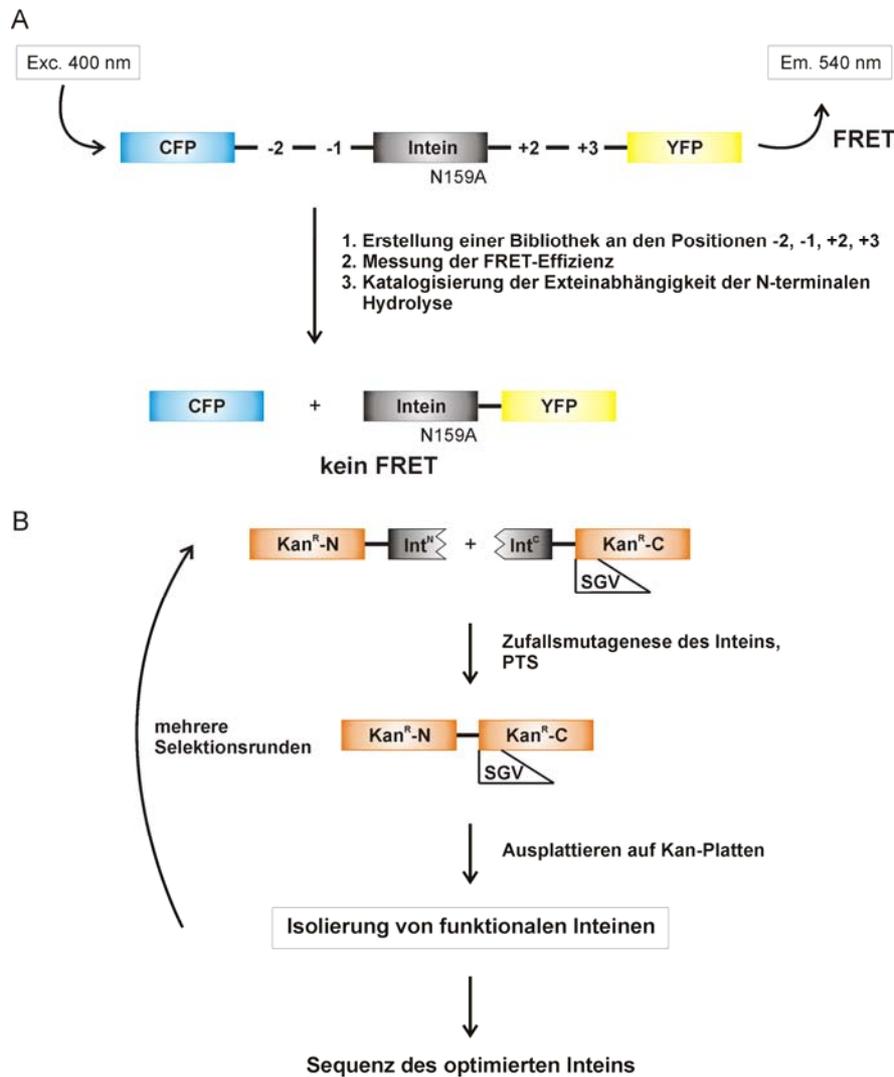
Für eine Vielzahl der in Kapitel 1.2.4 vorgestellten Anwendungen von gespaltenen Inteinen ist ein schneller Umsatz der Inteinfusionsproteine durch PTS von großem Vorteil. Das in dieser Arbeit zum ersten Mal biochemisch mit gereinigten Proteinen charakterisierte DnaE Intein aus *Nostoc punctiforme* (*Npu* DnaE) vermittelt die PTS-Reaktion um den Faktor 6 schneller als das bisher schnellste Intein, das *Sce* VMA Intein. Im Vergleich zu dem homologen Inteinallel aus *Synechocystis* species PCC6803 (*Ssp*) ist die unimolekulare Reaktionskonstante sogar zwischen 33- und 170-mal größer (s. 1.2.3 in der Einleitung). Die Gründe für diese herausragenden kinetischen Eigenschaften vor allem im Vergleich zum *Ssp* DnaE Intein wurden im Laufe dieser Arbeit nicht experimentell untersucht. Interessanterweise besitzen die beiden hoch homologen Inteinallele eine große Sequenzidentität von 67% im N-terminalen und von 53% im C-terminalen Teil [151]. Zusätzlich sind alle bisher bekannten für den Inteinmechanismus wichtigen Reste in beiden Inteinen identisch. Der einzige bei Betrachtung der Primärsequenz auffallende Unterschied ist eine 21 AS große Erweiterung des N-terminalen Inteinteils im *Ssp* DnaE. Anhand der Kristallstrukturen [137] lässt sich postulieren, dass diese Reste nicht an der eigentlichen Inteinfaltung beteiligt sind. Ghosh *et al.* konnten zeigen, dass diese N-terminale Erweiterung ohne Beeinträchtigung der Spleißfähigkeit entfernt werden kann [287]. Allerdings wurde hier nicht explizit der Einfluss der Deletionsmutation auf die Reaktionsgeschwindigkeit bestimmt. Hinweise darauf, dass die zusätzlichen Aminosäuren nicht entscheidend für den kinetischen Ablauf sind, kommen von der Untersuchung eines anderen Mitglieds des DnaE Allels. Das DnaE Intein aus dem Cyanobakterium *Nostoc* species PCC7120 (*Nsp*) ist ebenfalls hoch homolog zu den beiden

anderen Vertretern und besitzt mit 102 Aminosäuren einen exakt so großen N-terminalen Inteinanteil, wie das *Npu* DnaE Intein. Diese Gemeinsamkeit alleine ist jedoch nicht ausreichend um dem *Nsp* Intein ebenfalls herausragende Eigenschaften zu verleihen. Dieses Intein wurde zwar ebenfalls nicht detailliert kinetisch charakterisiert, aber es zeigt mit nur ~15% Spleißausbeute deutlich schlechtere Charakteristika im Vergleich zum *Npu* DnaE Intein [157]. Damit können die unterschiedlichen Reaktionsgeschwindigkeiten zwischen *Npu* und *Ssp* DnaE Intein wahrscheinlich nicht auf die N-terminale Erweiterung des *Ssp* DnaE Inteins zurückgeführt werden.

Der Mechanismus der PTS-Reaktion erfordert die Koordination von vier einzelnen Teilschritten. Für das *Ssp* DnaE Intein laufen sowohl die Komplexbildung [174], als auch der N-S-Acyltransfer und die Asparaginzyklisierung [173] mit deutlich höheren Geschwindigkeiten im Vergleich zur Gesamtreaktion ab. Damit ist für dieses Intein mit nativen Exteinresten an den Spleißstellen wahrscheinlich die Umesterung der die Gesamtreaktion verlangsamende, geschwindigkeitsbestimmende Schritt. Für die Übertragung des N-Exteins auf die Seitenkette der +1 Aminosäure wird eine konformationelle Änderung der Inteinstruktur postuliert, die die reaktiven Gruppen in räumliche Nähe bringt [137]. Die Hypothese, dass eine Feinabstimmung der dynamischen Eigenschaften der katalytischen Reste die Reaktionsgeschwindigkeit der PTS-Reaktion erhöhen kann, wird durch neueste Evolutionsstudien zur Verbesserung der Spleißereigenschaften von Inteinen bestätigt. Hiraga *et al.* benutzten ein *in vitro* Phagedisplay System zur Optimierung des *cis*-spleißenden *Mycobacterium tuberculosis* RecA Mini-Inteins (*Mtu* RecA) [288]. Den Autoren gelang es durch fehlerbehaftete PCR Inteinvarianten zu erzeugen, die im Vergleich zum Wildtyp verbesserte Spleißkinetiken besaßen. Interessanterweise befanden sich die mutierten Reste nicht in der Nähe der katalytischen Zentren, sondern an der Peripherie der Inteinstruktur. Anhand von NMR-Untersuchungen konnte weiterhin gezeigt werden, dass die Mutationen über einen Welleneffekt die dynamischen Eigenschaften der Reste im aktiven Zentrum beeinflussen [288]. Eine Feinabstimmung der gesamten Inteinstruktur und -dynamik könnte damit auch als Erklärung für die verbesserte Spleißkinetik des *Npu* DnaE Inteins im Vergleich mit dem *Ssp* DnaE herangezogen werden. Das *Npu* scheint die Koordination der einzelnen Teilschritte effektiver zu katalysieren als das *Ssp*. Auch ohne ein detailliertes Verständnis der Reaktivitätsunterschiede bietet sich das *Npu* DnaE Intein aufgrund seiner hervorragenden Eigenschaften als das Intein der Wahl für zukünftige biochemische Anwendungen an.

### 5.3.3 Exteinabhängigkeit

Ein großer Nachteil aller auf gespaltenen Inteinen basierenden Techniken ist die Tatsache, dass die Aktivität jedes Inteins zu einem gewissen Grad abhängig von seinen Substraten, also seinen Exteinsequenzen, ist (s. Kapitel 1.2.2 in der Einleitung). Zwar geht die Spleißfähigkeit vollständig auf die Inteinstruktur zurück, jedoch sind sowohl *cis*- als auch gesplattene Inteine keine vollständig transportierbaren genetischen Elemente. Im günstigsten Fall kann ein heterologer Exteinkontext im Vergleich mit der Wildtypsituation vergleichbare oder bessere Charakteristika in Bezug auf Ausbeute und Spleißgeschwindigkeit auslösen. Im anderen Extrem führt die Integration eines gespaltenen Inteins in ein fremdes Zielprotein zur vermehrten Nebenproduktbildung und/oder einer kompletten Inaktivierung der PTS-Reaktion. Neben dem essentiellen Nukleophil an der Position +1 konnten auch den Aminosäureresten direkt an den Spleißstellen (-2, -1, +2, +3 Position) ein modulierender Einfluss auf die Spleißaktivität nachgewiesen werden [134,151,241]. Aktuell lassen sich auch anhand der vorhandenen Kristallstrukturen von Inteinen keine definitiven Rückschlüsse über die Wirkungsweise der Exteinsequenzen ableiten. Die lange Ko-evolution jedes Intein-Exteinpaars und die große Varianz der natürlichen Exteinreste machen theoretische *ab initio* Vorhersagen zur Spleißaktivität eher unwahrscheinlich. Ein erstes Beispiel für eine experimentelle Katalogisierung der Exteinabhängigkeit für das fusionierte *Ssp* DnaE Intein lieferten 2009 Amitai *et al.* [152]. Ihr Ansatz erlaubte ein schnelles Testen einer Vielzahl von Exteinresten im Hochdurchsatz basierend auf einem FRET-System. Die Autoren fusionierten auf genetischer Ebene das cyan-fluoreszierende Protein (CFP), die beiden Hälften des *Ssp* DnaE Inteins mit einer Mutation der C-terminalen Aminosäure (N159A) und das gelb-fluoreszierende Protein (YFP) miteinander. Durch die Mutation des Inteins war die Bildung des Spleißprodukts aus CFP and YFP nicht möglich. Damit konnte einzig die Abspaltung des N-Exteins (CFP) von dem Intein katalysiert und durch einen Rückgang der FRET-Effizienz nachgewiesen werden (s. Abbildung 69A). Eine vollständige Mutation der -2, -1, +2 und +3 Positionen demonstrierte erneut den großen Einfluss der Exteinreste auf die Aktivität des Inteins [152]. Amitai *et al.* identifizierten Kombinationen von Aminosäuren im N- und im C-Extein, die im Vergleich zu der Wildtyp-Sequenz EY-CFN die durch DTT-induzierte N-terminale Hydrolyse entweder um den Faktor 4 beschleunigten oder 1000fach verlangsamten [152]. Interessanterweise konnte keine definitive Korrelation zwischen verbesserter N-terminaler Hydrolyse und verbesserter Spleißfähigkeit nachgewiesen werden.

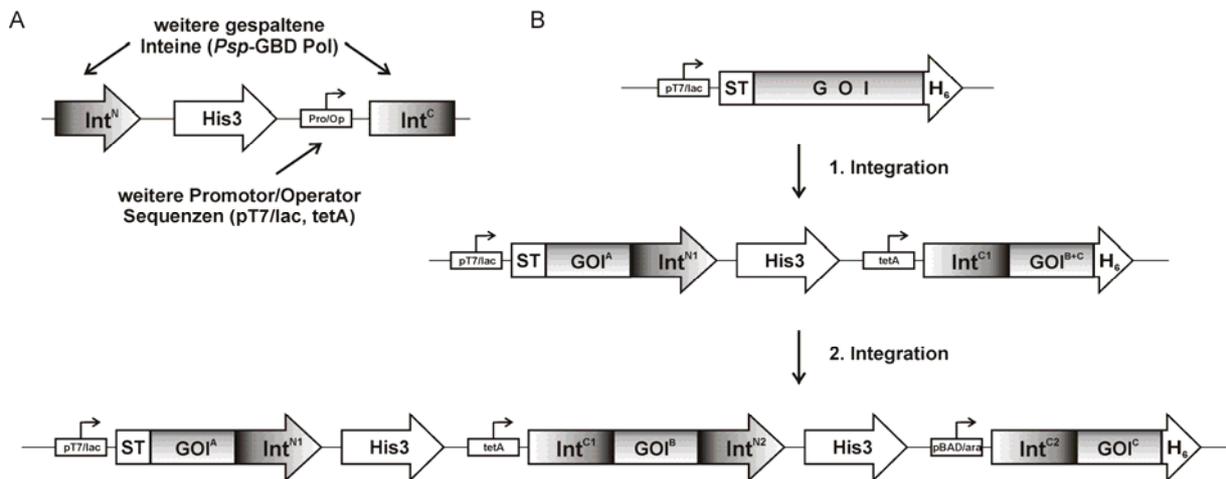


**Abbildung 69 Aktuelle Arbeiten zur Exteinabhängigkeit von Inteinen.** A) Katalogisierung der Abhängigkeit der N-terminalen Hydrolyse des fusionierten *Ssp* DnaE Inteins von den -2, -1, +2, +3 Exteinpositionen; das System basiert auf einem Hochdurchsatz-Screening der FRET-Effizienz [152]; B) Proteinevolution des *Npu* DnaE Int<sup>N</sup> und des *Ssp* DnaE Int<sup>C</sup> Systems mittels *in vivo* Selektion durch Ligation der Aminoglykosidtransferase KanR; das Prinzip erlaubt die Anpassung des Inteins an definierte Exteinsequenzen [245].

Auch wenn in der Arbeit von Amitai *et al.* [152] nicht direkt die Spleißfähigkeit des *Ssp* DnaE Inteins getestet wurde, so kann doch potentiell mit dieser oder einer verwandten Methode die Abhängigkeit eines Inteins von seinen flankierenden Exteinminosäuren katalogisiert werden. Es bleibt anzumerken, dass zum einen ein großer experimenteller Aufwand betrieben und zum anderen diese Katalogisierung für jedes Intein separat durchgeführt werden muss. Zusätzlich sind in einer konkreten Anwendung, also der Integration eines gespaltenen Inteins in ein Zielprotein, noch andere Parameter, wie der Faltungszustand der Exteinhälften, zu beachten [289]. Dieses gezielte Testen der Spleißfähigkeit eines Inteins an einer bestimmten Position in einem Protein von Interesse wird mit dem in dieser Arbeit entwickeltem SPLICEFINDER-System dramatisch vereinfacht (s.

Kapitel 4.3). Die Zahl der Klonierungsschritte wird durch die Integration einer Intein-Kassette in das Zielgen mittels RF-PCR deutlich verringert. Auch hier können durch unterschiedliche Primerwahl eine größere Anzahl an flankierenden Aminosäuren sowie unterschiedliche Integrationspositionen parallel getestet werden. Eine detaillierte Katalogisierung der Abhängigkeit der Inteine von ihren flankierenden Aminosäuren wie von Amitai *et al.* beschrieben [152] würde jedoch auch für die SPLICEFINDER-Methode von Vorteil sein. Durch eine geschickte Vorauswahl ließe sich damit die Zahl der zu untersuchenden Kombinationen reduzieren.

In zukünftigen Arbeiten könnte die SPLICEFINDER-Methode relativ schnell zu einem System der zweiten Generation (SPLICEFINDER 2.0) weiterentwickelt werden (s. Abbildung 70A).



**Abbildung 70 Ausbau des SPLICEFINDER-Systems zu SF 2.0.** A) Weitere Möglichkeiten zum Austausch der gespaltenen Inteine und/oder der Promotor/Operator Sequenzen; B) Schematische Darstellung der zweifachen Integration zweier orthogonaler Intein-Kassetten in einen Zielvektor ST-GOI- $H_6$ ; in einem solchen System wäre unter anderem die segmentelle Isotopenmarkierung eines inneren Fragments möglich.

Zum einen könnten weitere Intein-Kassetten, wie die des künstlich gespaltenen *Psp*-GBD Pol-Inteins, die Einsetzbarkeit erhöhen. Mit diesem gespaltenen Intein gelang die Ligation zweier Segmente eines integralen Membranproteins [184] durch Spleißen in 6M Harnstoff und damit eröffnet sich der Zugang zur selektiven Modifikation dieser so wichtigen Proteinklasse. Durch zusätzliche Intein-Kassetten kann somit für jede spezielle Anwendung das am besten geeignete Intein ausgesucht werden. Zum anderen bräuchte der Austausch der Promotor/Operator Sequenz innerhalb der Intein-Kassette weitere Vorteile. Die Integration einer T7/lac Kasette in ein Zielgen unter pBAD/ara Kontrolle würde die Genexpression des C-terminalen Fragments mit IPTG induzierbar machen. Dies ist vor allem für eine segmentelle  $^{13}C$ -Isotopenmarkierung eines C-terminalen Fragments wichtig, da somit der

Einsatz von sehr teurer  $^{13}\text{C}$ -Arabinose umgangen werden kann. Der Einbau eines mit Anhydrotetrazyklin induzierbaren Promotors [290] in die Inteinkassette ermöglicht eine Ligation des zu untersuchenden Proteins aus drei selektiv induzierbaren Fragmenten (s. Abbildung 70B). Dazu müssen sukzessiv zwei orthogonale Inteinkassetten in das Zielgen eingebracht werden. Zweifaches PTS würde dann das vollständige Protein ergeben. Mit diesem System kann durch gezielte Wahl der Expressionsmedien eine definierte, unterschiedliche Isotopenanreicherung in den drei Segmenten erreicht werden. Allerdings wäre die Konstruktion eines weiteren Helferplasmids mit dem zusätzlichen Gen für das regulatorische Protein TetR [290] notwendig.

Oft ermöglichen die Exteinabhängigkeiten der gespaltenen Inteine keine spurlose Ligation eines Zielproteins, d.h., das Spleißprodukt unterscheidet sich in der Primärsequenz in einer oder in mehreren Positionen von dem Wildtyp-Protein. In den meisten Fällen werden solche Mutationen, gerade in flexiblen Schleifenregionen, ohne Probleme von dem Zielprotein toleriert. Führt jedoch diese Änderung in der Primärsequenz zu einer Inaktivierung des Proteins oder wird besonders viel Wert auf die identische Aminosäurezusammensetzung gelegt, so kann eine von Lockless *et al.* kürzlich beschriebene Methode dieses Problem lösen (s. Abbildung 69B) [245]. Sie entwickelten ein Proteinevolutionssystem zur Anpassung eines gespaltenen Inteins an eine bestimmte Exteinsequenz. Die Selektion erfolgte *in vivo* über das Zusammenspleißen der Aminoglykosidtransferase (kodiert durch das Kan<sup>R</sup>-Gen). Die einzelnen Inteininfusionsproteine konnten den *E. coli*-Zellen keine Resistenz gegenüber dem Antibiotikum Kanamycin verleihen. Nur bei erfolgreicher Spleißproduktbildung waren die Zellen in Kanamycin-haltigem Medium lebensfähig [245]. Die Autoren verwendeten eine Kombination aus dem *Npu* DnaE Int<sup>N</sup> und dem *Ssp* DnaE Int<sup>C</sup> und evolvierten dieses gespalte Intein auf die Exteinsequenz S<sub>+1</sub>GV anstelle des nativen CFN. Nach drei Runden Selektion wurde ein mutiertes Intein isoliert (E15D, L25I, L92M im N-terminalen Teil und D23Y im C-terminalen Teil), welches spezifisch verbessert für die oben genannte C-Exteinsequenz war. Zusätzlich konnte gezeigt werden, dass mit dieser Inteinvariante die spurlose Ligation des proto-onkogenischen Adaptor-Proteins Crk-II mit SGV an der C-terminalen Spleißstelle möglich war [245]. Abschließend lässt sich sagen, dass diese Evolutionsmethode eine elegante Lösung des Problems der Exteinabhängigkeit von gespaltenen Inteinen ist, jedoch wird wahrscheinlich der große experimentelle Aufwand einen routinemäßigen Einsatz dieser Technik verhindern.

---

## Literatur

- [1] Nathan C (2004) Antibiotics at the crossroads, *Nature* **431**, 899-902.
- [2] Davies J (1994) Inactivation of antibiotics and the dissemination of resistance genes, *Science* **264**, 375-82.
- [3] Fischbach MA & Walsh CT (2009) Antibiotics for emerging pathogens, *Science* **325**, 1089-93.
- [4] Klevens RM et al. (2007) Invasive methicillin-resistant *Staphylococcus aureus* infections in the United States, *Jama* **298**, 1763-71.
- [5] Falagas ME, Bliziotis IA, Kasiakou SK, Samonis G, Athanassopoulou P & Michalopoulos A (2005) Outcome of infections due to pandrug-resistant (PDR) Gram-negative bacteria, *BMC Infect Dis* **5**, 24.
- [6] Dorman SE & Chaisson RE (2007) From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis, *Nat Med* **13**, 295-8.
- [7] Byford MF, Baldwin JE, Shiao CY & Schofield CJ (1997) The Mechanism of ACV Synthetase, *Chem Rev* **97**, 2631-2650.
- [8] Mootz HD & Marahiel MA (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: Complete nucleotide sequence and biochemical characterization of functional internal adenylation domains, *Journal of Bacteriology* **179**, 6843-6850.
- [9] Kratzschmar J, Krause M & Marahiel MA (1989) Gramicidin-S Biosynthesis Operon Containing the Structural Genes *Grsa* and *Grsb* Has an Open Reading Frame Encoding a Protein Homologous to Fatty-Acid Thioesterases, *Journal of Bacteriology* **171**, 5422-5429.
- [10] Krause M & Marahiel MA (1988) Organization of the Biosynthesis Genes for the Peptide Antibiotic Gramicidin-S, *Journal of Bacteriology* **170**, 4669-4674.
- [11] Tally FP & DeBruin MF (2000) Development of daptomycin for gram-positive infections, *J Antimicrob Chemother* **46**, 523-6.
- [12] Kohli RM & Walsh CT (2003) Enzymology of acyl chain macrocyclization in natural product biosynthesis, *Chem Commun (Camb)*, 297-307.
- [13] Grunewald J & Marahiel MA (2006) Chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides, *Microbiology and Molecular Biology Reviews* **70**, 121-+.
- [14] Velkov T & Lawen A (2003) Non-ribosomal peptide synthetases as technological platforms for the synthesis of highly modified peptide bioeffectors--Cyclosporin synthetase as a complex example, *Biotechnol Annu Rev* **9**, 151-97.
- [15] Peypoux F, Bonmatin JM & Wallach J (1999) Recent trends in the biochemistry of surfactin, *Appl Microbiol Biotechnol* **51**, 553-63.
- [16] Raymond KN, Dertz EA & Kim SS (2003) Enterobactin: an archetype for microbial iron transport, *Proc Natl Acad Sci U S A* **100**, 3584-8.
- [17] May JJ, Wendrich TM & Marahiel MA (2001) The *dhb* operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin, *Journal of Biological Chemistry* **276**, 7209-7217.
- [18] Barry SM & Challis GL (2009) Recent advances in siderophore biosynthesis, *Curr Opin Chem Biol* **13**, 205-15.
- [19] Crosa JH & Walsh CT (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria, *Microbiol Mol Biol Rev* **66**, 223-49.
- [20] Xu Y, Sugar IP & Krishna NR (1995) A variable target intensity-restrained global optimization (VARTIGO) procedure for determining three-dimensional structures of polypeptides from NOESY data: application to gramicidin-S, *J Biomol NMR* **5**, 37-48.
- [21] Ando S, Nishikawa H, Takiguchi H, Lee S & Sugihara G (1993) Antimicrobial specificity and hemolytic activity of cyclized basic amphiphilic beta-structural model peptides and their interactions with phospholipid bilayers, *Biochim Biophys Acta* **1147**, 42-9.
- [22] Kondejewski LH, Jelokhani-Niaraki M, Farmer SW, Lix B, Kay CM, Sykes BD, Hancock RE & Hodges RS (1999) Dissociation of antimicrobial and hemolytic activities in cyclic peptide diastereomers by systematic alterations in amphipathicity, *J Biol Chem* **274**, 13181-92.
- [23] Tishchenko GN, Andrianov VI, Vainstein BK, Woolfson MM & Dodson E (1997) Channels in the gramicidin S-with-urea structure and their possible relation to transmembrane ion transport, *Acta Crystallogr D Biol Crystallogr* **53**, 151-9.
- [24] Prenner EJ, Lewis RN, Kondejewski LH, Hodges RS & McElhaney RN (1999) Differential scanning calorimetric study of the effect of the antimicrobial peptide gramicidin S on the thermotropic phase behavior of phosphatidylcholine, phosphatidylethanolamine and phosphatidylglycerol lipid bilayer membranes, *Biochim Biophys Acta* **1417**, 211-23.
- [25] Salgado J, Grage SL, Kondejewski LH, Hodges RS, McElhaney RN & Ulrich AS (2001) Membrane-bound structure and alignment of the antimicrobial beta-sheet peptide gramicidin S derived from angular and distance constraints by solid state 19F-NMR, *J Biomol NMR* **21**, 191-208.
- [26] Lee DL, Powers JP, Pfliegerl K, Vasil ML, Hancock RE & Hodges RS (2004) Effects of single D-amino acid substitutions on disruption of beta-sheet structure and hydrophobicity in cyclic 14-residue antimicrobial peptide analogs related to gramicidin S, *J Pept Res* **63**, 69-84.
- [27] Prenner EJ, Lewis RN & McElhaney RN (1999) The interaction of the antimicrobial peptide gramicidin S with lipid bilayer model and biological membranes, *Biochim Biophys Acta* **1462**, 201-21.
- [28] Mogi T, Ui H, Shiomi K, Omura S & Kita K (2008) Gramicidin S identified as a potent inhibitor for cytochrome bd-type quinol oxidase, *FEBS Lett* **582**, 2299-302.

- [29] Kawai M, Yamamura H, Tanaka R, Umemoto H, Ohmizo C, Higuchi S & Katsu T (2005) Proline residue-modified polycationic analogs of gramicidin S with high antibacterial activity against both Gram-positive and Gram-negative bacteria and low hemolytic activity, *J Pept Res* **65**, 98-104.
- [30] Prenner EJ, Kiricsi M, Jelokhani-Niaraki M, Lewis RN, Hodges RS & McElhane RN (2005) Structure-activity relationships of diastereomeric lysine ring size analogs of the antimicrobial peptide gramicidin S: mechanism of action and discrimination between bacterial and animal cell membranes, *J Biol Chem* **280**, 2002-11.
- [31] Prenner EJ, Lewis RN, Jelokhani-Niaraki M, Hodges RS & McElhane RN (2001) Cholesterol attenuates the interaction of the antimicrobial peptide gramicidin S with phospholipid bilayer membranes, *Biochim Biophys Acta* **1510**, 83-92.
- [32] Mootz HD, Schwarzer D & Marahiel MA (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases, *ChemBiochem* **3**, 490-504.
- [33] Pohlmann V & Marahiel MA (2008) Delta-amino group hydroxylation of L-ornithine during coelichelin biosynthesis, *Org Biomol Chem* **6**, 1843-8.
- [34] Stachelhaus T, Mootz HD, Bergendahl V & Marahiel MA (1998) Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain, *J Biol Chem* **273**, 22773-81.
- [35] Bergendahl V, Linne U & Marahiel MA (2002) Mutational analysis of the C-domain in nonribosomal peptide synthesis, *Eur J Biochem* **269**, 620-9.
- [36] Keating TA, Marshall CG, Walsh CT & Keating AE (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains, *Nat Struct Biol* **9**, 522-6.
- [37] Borchert S, Stachelhaus T & Marahiel MA (1994) Induction of Surfactin Production in *Bacillus-Subtilis* by Gsp, a Gene Located Upstream of the Gramicidin-S Operon in *Bacillus-Brevis*, *Journal of Bacteriology* **176**, 2458-2462.
- [38] Hoyer KM, Mahler C & Marahiel MA (2007) The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization, *Chemistry & Biology* **14**, 13-22.
- [39] Gevers W, Kleinkauf H & Lipmann F (1968) The activation of amino acids for biosynthesis of gramicidin S, *Proc Natl Acad Sci U S A* **60**, 269-76.
- [40] Stachelhaus T & Marahiel MA (1995) Modular Structure of Peptide Synthetases Revealed by Dissection of the Multifunctional Enzyme Grsa, *Journal of Biological Chemistry* **270**, 6163-6169.
- [41] Stachelhaus T & Walsh CT (2000) Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase, *Biochemistry* **39**, 5775-87.
- [42] Ehmann DE, Trauger JW, Stachelhaus T & Walsh CT (2000) Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases, *Chem Biol* **7**, 765-72.
- [43] Belshaw PJ, Walsh CT & Stachelhaus T (1999) Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis, *Science* **284**, 486-9.
- [44] Linne U & Marahiel MA (2000) Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization, *Biochemistry* **39**, 10439-47.
- [45] Shaw-Reid CA, Kelleher NL, Losey HC, Gehring AM, Berg C & Walsh CT (1999) Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of *E. coli* EntF catalyzes both elongation and cyclolactonization, *Chem Biol* **6**, 385-400.
- [46] Bu X, Wu X, Ng NL, Mak CK, Qin C & Guo Z (2004) Synthesis of gramicidin S and its analogues via an on-resin macrocyclization assisted by a predisposed conformation of the linear precursors, *J Org Chem* **69**, 2681-5.
- [47] Wadhvani P, Afonin S, Ieronimo M, Buerck J & Ulrich AS (2006) Optimized protocol for synthesis of cyclic gramicidin S: starting amino acid is key to high yield, *J Org Chem* **71**, 55-61.
- [48] Schwarzer D & Marahiel MA (2001) Multimodular biocatalysts for natural product assembly, *Naturwissenschaften* **88**, 93-101.
- [49] Nolan EM & Walsh CT (2009) How nature morphs peptide scaffolds into antibiotics, *ChemBiochem* **10**, 34-53.
- [50] Donadio S, Monciardini P & Sosio M (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics, *Nat Prod Rep* **24**, 1073-109.
- [51] Cane DE & Walsh CT (1999) The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases, *Chem Biol* **6**, R319-25.
- [52] Marahiel MA, Stachelhaus T & Mootz HD (1997) Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis, *Chem Rev* **97**, 2651-2674.
- [53] Gulick AM (2009) Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase, *ACS Chem Biol* **4**, 811-27.
- [54] Wood KV, Lam YA & McElroy WD (1989) Introduction to beetle luciferases and their applications, *J Biolumin Chemilumin* **4**, 289-301.
- [55] Yonus H, Neumann P, Zimmermann S, May JJ, Marahiel MA & Stubbs MT (2008) Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains, *J Biol Chem* **283**, 32484-91.
- [56] Conti E, Franks NP & Brick P (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes, *Structure* **4**, 287-98.
- [57] Conti E, Stachelhaus T, Marahiel MA & Brick P (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S, *Embo J* **16**, 4174-83.
- [58] Kochan G, Pilka ES, von Delft F, Oppermann U & Yue WW (2009) Structural snapshots for the conformation-dependent catalysis by human medium-chain acyl-coenzyme A synthetase ACSM2A, *J Mol Biol* **388**, 997-1008.
- [59] Branchini BR, Murtiashaw MH, Magyar RA & Anderson SM (2000) The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase, *Biochemistry* **39**, 5433-40.
- [60] Horswill AR & Escalante-Semerena JC (2002) Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: residue Lys592 is required for propionyl-AMP synthesis, *Biochemistry* **41**, 2379-87.

- [61] Stachelhaus T, Mootz HD & Marahiel MA (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases, *Chem Biol* **6**, 493-505.
- [62] Saraste M, Sibbald PR & Wittinghofer A (1990) The P-loop--a common motif in ATP- and GTP-binding proteins, *Trends Biochem Sci* **15**, 430-4.
- [63] Osman KT, Du L, He Y & Luo Y (2009) Crystal structure of *Bacillus cereus* D-alanyl carrier protein ligase (DltA) in complex with ATP, *J Mol Biol* **388**, 345-55.
- [64] Cisar JS, Ferreras JA, Soni RK, Quadri LE & Tan DS (2007) Exploiting ligand conformation in selective inhibition of non-ribosomal peptide synthetase amino acid adenylation with designed macrocyclic small molecules, *J Am Chem Soc* **129**, 7752-3.
- [65] Reshetnikova L, Moor N, Lavrik O & Vassilyev DG (1999) Crystal structures of phenylalanyl-tRNA synthetase complexed with phenylalanine and a phenylalanyl-adenylate analogue, *J Mol Biol* **287**, 555-68.
- [66] Branchini BR, Southworth TL, Murtiashaw MH, Wilkinson SR, Khattak NF, Rosenberg JC & Zimmer M (2005) Mutagenesis evidence that the partial reactions of firefly bioluminescence are catalyzed by different conformations of the luciferase C-terminal domain, *Biochemistry* **44**, 1385-93.
- [67] Reger AS, Carney JM & Gulick AM (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase, *Biochemistry* **46**, 6536-46.
- [68] Drake EJ, Nicolai DA & Gulick AM (2006) Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain, *Chem Biol* **13**, 409-19.
- [69] Wu J, Xu D, Lu X, Wang C, Guo H & Dunaway-Mariano D (2006) Contributions of long-range electrostatic interactions to 4-chlorobenzoyl-CoA dehalogenase catalysis: a combined theoretical and experimental study, *Biochemistry* **45**, 102-12.
- [70] Luo LS, Kohli RM, Onishi M, Linne U, Marahiel MA & Walsh CT (2002) Timing of epimerization and condensation reactions in nonribosomal peptide assembly lines: Kinetic analysis of phenylalanine activating elongation modules of tyrocidine synthetase B, *Biochemistry* **41**, 9184-9196.
- [71] Reger AS, Wu R, Dunaway-Mariano D & Gulick AM (2008) Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase, *Biochemistry* **47**, 8016-25.
- [72] Wu R, Reger AS, Lu X, Gulick AM & Dunaway-Mariano D (2009) The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase, *Biochemistry* **48**, 4115-25.
- [73] Lambalot RH et al. (1996) A new enzyme superfamily - The phosphopantetheinyl transferases, *Chemistry & Biology* **3**, 923-936.
- [74] Walsh CT, Gehring AM, Weinreb PH, Quadri LE & Flugel RS (1997) Post-translational modification of polyketide and nonribosomal peptide synthetases, *Curr Opin Chem Biol* **1**, 309-15.
- [75] Quadri LE, Weinreb PH, Lei M, Nakano MM, Zuber P & Walsh CT (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases, *Biochemistry* **37**, 1585-95.
- [76] Mootz HD, Finking R & Marahiel MA (2001) 4'-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*, *Journal of Biological Chemistry* **276**, 37289-37298.
- [77] La Clair JJ, Foley TL, Schegg TR, Regan CM & Burkart MD (2004) Manipulation of carrier proteins in antibiotic biosynthesis, *Chem Biol* **11**, 195-201.
- [78] George N, Pick H, Vogel H, Johnsson N & Johnsson K (2004) Specific labeling of cell surface proteins with chemically diverse compounds, *J Am Chem Soc* **126**, 8896-7.
- [79] Yin J, Liu F, Li X & Walsh CT (2004) Labeling proteins with small molecules by site-specific posttranslational modification, *J Am Chem Soc* **126**, 7754-5.
- [80] Yin J et al. (2005) Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase, *Proc Natl Acad Sci U S A* **102**, 15815-20.
- [81] Stein DB, Linne U, Hahn M & Marahiel MA (2006) Impact of epimerization domains on the intermodular transfer of enzyme-bound intermediates in nonribosomal peptide synthesis, *Chembiochem* **7**, 1807-14.
- [82] Mootz HD, Schwarzer D & Marahiel MA (2000) Construction of hybrid peptide synthetases by module and domain fusions, *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5848-5853.
- [83] Weissman KJ & Muller R (2008) Protein-protein interactions in multienzyme megasynthetases, *Chembiochem* **9**, 826-848.
- [84] Spivey HO & Ovadi J (1999) Substrate channeling, *Methods* **19**, 306-21.
- [85] Castonguay R, He W, Chen AY, Khosla C & Cane DE (2007) Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase, *J Am Chem Soc* **129**, 13758-69.
- [86] Weber T, Baumgartner R, Renner C, Marahiel MA & Holak TA (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases, *Structure* **8**, 407-18.
- [87] Koglin A et al. (2006) Conformational switches modulate protein interactions in peptide antibiotic synthetases, *Science* **312**, 273-6.
- [88] Findlow SC, Winsor C, Simpson TJ, Crosby J & Crump MP (2003) Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*, *Biochemistry* **42**, 8423-33.
- [89] Xu GY, Tam A, Lin L, Hixon J, Fritz CC & Powers R (2001) Solution structure of *B. subtilis* acyl carrier protein, *Structure* **9**, 277-87.
- [90] Kim Y & Prestegard JH (1989) A dynamic model for the structure of acyl carrier protein in solution, *Biochemistry* **28**, 8792-7.
- [91] Koglin A et al. (2008) Structural basis for the selectivity of the external thioesterase of the surfactin synthetase, *Nature* **454**, 907-11.
- [92] Byers DM & Gong H (2007) Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family, *Biochem Cell Biol* **85**, 649-62.

- [93] Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT & Wagner G (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase, *Nature* **454**, 903-6.
- [94] Tanovic A, Samel SA, Essen LO & Marahiel MA (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase, *Science* **321**, 659-63.
- [95] Lai JR, Fischbach MA, Liu DR & Walsh CT (2006) Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection, *J Am Chem Soc* **128**, 11002-3.
- [96] Lai JR, Fischbach MA, Liu DR & Walsh CT (2006) A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection, *Proc Natl Acad Sci U S A* **103**, 5314-9.
- [97] Zhou Z, Lai JR & Walsh CT (2006) Interdomain communication between the thiolation and thioesterase domains of EntF explored by combinatorial mutagenesis and selection, *Chem Biol* **13**, 869-79.
- [98] Zhou Z, Lai JR & Walsh CT (2007) Directed evolution of aryl carrier proteins in the enterobactin synthetase, *Proc Natl Acad Sci U S A* **104**, 11621-6.
- [99] Mofid MR, Finking R & Marahiel MA (2002) Recognition of hybrid peptidyl carrier proteins/Acyl carrier proteins in nonribosomal peptide synthetase modules by the 4'-phosphopantetheinyl transferases AcpS and Sfp, *Journal of Biological Chemistry* **277**, 17023-17031.
- [100] Finking R, Mofid MR & Marahiel MA (2004) Mutational analysis of peptidyl carrier protein and acyl carrier protein synthase unveils residues involved in protein-protein recognition, *Biochemistry* **43**, 8946-56.
- [101] Marshall CG, Burkart MD, Meray RK & Walsh CT (2002) Carrier protein recognition in siderophore-producing nonribosomal peptide synthetases, *Biochemistry* **41**, 8429-37.
- [102] Lai JR, Koglin A & Walsh CT (2006) Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis, *Biochemistry* **45**, 14869-79.
- [103] Samel SA, Schoenafinger G, Knappe TA, Marahiel MA & Essen LO (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase, *Structure* **15**, 781-792.
- [104] Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT & Stubbs MT (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE, *Structure* **10**, 301-10.
- [105] Kleinkauf H & von Dohren H (1987) Biosynthesis of peptide antibiotics, *Annu Rev Microbiol* **41**, 259-89.
- [106] Kleinkauf H & von Dohren H (1990) Nonribosomal biosynthesis of peptide antibiotics, *Eur J Biochem* **192**, 1-15.
- [107] Weissman KJ & Muller R (2008) Crystal structure of a molecular assembly line, *Angew Chem Int Ed Engl* **47**, 8344-6.
- [108] Maier T, Jenni S & Ban N (2006) Architecture of mammalian fatty acid synthase at 4.5 angstrom resolution, *Science* **311**, 1258-1262.
- [109] Jenni S, Leibundgut M, Maier T & Ban N (2006) Architecture of a fungal fatty acid synthase at 5 angstrom resolution, *Science* **311**, 1263-1267.
- [110] Smith S (2006) Structural biology. Architectural options for a fatty acid synthase, *Science* **311**, 1251-2.
- [111] Wang L & Schultz PG (2004) Expanding the genetic code, *Angew Chem Int Ed Engl* **44**, 34-66.
- [112] Dawson PE & Kent SB (2000) Synthesis of native proteins by chemical ligation, *Annu Rev Biochem* **69**, 923-60.
- [113] Muir TW (2003) Semisynthesis of proteins by expressed protein ligation, *Annu Rev Biochem* **72**, 249-89.
- [114] David R, Richter MP & Beck-Sickinger AG (2004) Expressed protein ligation. Method and applications, *Eur J Biochem* **271**, 663-77.
- [115] Saxon E, Armstrong JI & Bertozzi CR (2000) A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds, *Org Lett* **2**, 2141-3.
- [116] Schellenberger V, Gorner A, Konnecke A & Jakubke HD (1991) Protease-catalyzed peptide synthesis: prevention of side reactions in kinetically controlled reactions, *Pept Res* **4**, 265-9.
- [117] Schellenberger V, Schellenberger U, Jakubke HD, Zapevalova NP & Mitin YV (1991) Protease-catalyzed peptide synthesis using inverse substrates: The synthesis of Pro-Xaa-bonds by trypsin, *Biotechnol Bioeng* **38**, 319-21.
- [118] Mazmanian SK, Liu G, Ton-That H & Schneewind O (1999) Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall, *Science* **285**, 760-3.
- [119] Mao H, Hart SA, Schink A & Pollok BA (2004) Sortase-mediated protein ligation: a new method for protein engineering, *J Am Chem Soc* **126**, 2670-1.
- [120] Hackenberger CP & Schwarzer D (2008) Chemoselective ligation and modification strategies for peptides and proteins, *Angew Chem Int Ed Engl* **47**, 10030-74.
- [121] Haugen P, Simon DM & Bhattacharya D (2005) The natural history of group I introns, *Trends Genet* **21**, 111-9.
- [122] Noren CJ, Wang J & Perler FB (2000) Dissecting the Chemistry of Protein Splicing and Its Applications, *Angew Chem Int Ed Engl* **39**, 450-466.
- [123] Paulus H (2000) Protein splicing and related forms of protein autoprocessing, *Annu Rev Biochem* **69**, 447-96.
- [124] Perler FB (2002) InBase: the Intein Database, *Nucleic Acids Res* **30**, 383-4.
- [125] Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L & Hilario E (2002) Inteins: structure, function, and evolution, *Annu Rev Microbiol* **56**, 263-87.
- [126] Perler FB et al. (1994) Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature, *Nucleic Acids Res* **22**, 1125-7.
- [127] Liu XQ (2000) Protein-splicing intein: Genetic mobility, origin, and evolution, *Annu Rev Genet* **34**, 61-76.
- [128] Gimble FS & Thorner J (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*, *Nature* **357**, 301-6.
- [129] Pietrokovski S (2001) Intein spread and extinction in evolution, *Trends Genet* **17**, 465-72.
- [130] Belfort M & Roberts RJ (1997) Homing endonucleases: keeping the house in order, *Nucleic Acids Res* **25**, 3379-88.
- [131] Dassa B, London N, Stoddard BL, Schueler-Furman O & Pietrokovski S (2009) Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family, *Nucleic Acids Res* **37**, 2560-73.
- [132] Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA & Leahy DJ (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins, *Cell* **91**, 85-97.

- [133] Duan X, Gimble FS & Quioco FA (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity, *Cell* **89**, 555-64.
- [134] Chong S & Xu MQ (1997) Protein splicing of the *Saccharomyces cerevisiae* VMA intein without the endonuclease motifs, *J Biol Chem* **272**, 15587-90.
- [135] Shingledecker K, Jiang SQ & Paulus H (1998) Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: design of a minimal intein and of a trans-splicing system involving two intein fragments, *Gene* **207**, 187-95.
- [136] Wu H, Xu MQ & Liu XQ (1998) Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein, *Biochim Biophys Acta* **1387**, 422-32.
- [137] Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ & Rao Z (2005) Crystal structures of an intein from the split dnaE gene of *Synechocystis* sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing, *J Mol Biol* **353**, 1093-105.
- [138] Mizutani R, Nogami S, Kawasaki M, Ohya Y, Anraku Y & Satow Y (2002) Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides, *J Mol Biol* **316**, 919-29.
- [139] Klabunde T, Sharma S, Telenti A, Jacobs WR, Jr. & Sacchettini JC (1998) Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing, *Nat Struct Biol* **5**, 31-6.
- [140] Xu MQ, Southworth MW, Mersha FB, Hornstra LJ & Perler FB (1993) In vitro protein splicing of purified precursor and the identification of a branched intermediate, *Cell* **75**, 1371-7.
- [141] Xu MQ & Perler FB (1996) The mechanism of protein splicing and its modulation by mutation, *Embo J* **15**, 5146-53.
- [142] Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M & Stevens TH (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase, *Science* **250**, 651-7.
- [143] Hirata R, Ohsumi Y, Nakano A, Kawasaki H, Suzuki K & Anraku Y (1990) Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*, *J Biol Chem* **265**, 6726-33.
- [144] Romanelli A, Shekhtman A, Cowburn D & Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction, *Proc Natl Acad Sci U S A* **101**, 6397-402.
- [145] Ludwig C, Schwarzer D & Mootz HD (2008) Interaction studies and alanine scanning analysis of a semi-synthetic split intein reveal thiazoline ring formation from an intermediate of the protein splicing reaction, *J Biol Chem* **283**, 25264-72.
- [146] Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, Lesage G & Rao Z (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing, *J Biol Chem* **278**, 39133-42.
- [147] Poland BW, Xu MQ & Quioco FA (2000) Structural insights into the protein splicing mechanism of PI-SceI, *J Biol Chem* **275**, 16408-13.
- [148] Davis EO, Jenner PJ, Brooks PC, Colston MJ & Sedgwick SG (1992) Protein splicing in the maturation of *M. tuberculosis* recA protein: a mechanism for tolerating a novel class of intervening sequence, *Cell* **71**, 201-10.
- [149] Southworth MW, Adam E, Panne D, Byer R, Kautz R & Perler FB (1998) Control of protein splicing by intein fragment reassembly, *Embo J* **17**, 918-26.
- [150] Lew BM & Paulus H (2002) An in vivo screening system against protein splicing useful for the isolation of non-splicing mutants or inhibitors of the RecA intein of *Mycobacterium tuberculosis*, *Gene* **282**, 169-77.
- [151] Iwai H, Zuger S, Jin J & Tam PH (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*, *FEBS Lett* **580**, 1853-8.
- [152] Amitai G, Callahan BP, Stanger MJ, Belfort G & Belfort M (2009) Modulation of intein activity by its neighboring extein substrates, *Proc Natl Acad Sci U S A* **106**, 11005-10.
- [153] Mootz HD (2009) Split inteins as versatile tools for protein semisynthesis, *Chembiochem* **10**, 2579-89.
- [154] Choi JJ, Nam KH, Min B, Kim SJ, Soll D & Kwon ST (2006) Protein trans-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*, *J Mol Biol* **356**, 1093-106.
- [155] Wu H, Hu Z & Liu XQ (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803, *Proc Natl Acad Sci U S A* **95**, 9226-31.
- [156] Caspi J, Amitai G, Belenkiy O & Pietrokovski S (2003) Distribution of split DnaE inteins in cyanobacteria, *Mol Microbiol* **50**, 1569-77.
- [157] Dassa B, Amitai G, Caspi J, Schueler-Furman O & Pietrokovski S (2007) Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations, *Biochemistry* **46**, 322-30.
- [158] Mills KV, Lew BM, Jiang S & Paulus H (1998) Protein splicing in trans by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein, *Proc Natl Acad Sci U S A* **95**, 3543-8.
- [159] Ozawa T, Nogami S, Sato M, Ohya Y & Umezawa Y (2000) A fluorescent indicator for detecting protein-protein interactions in vivo based on protein splicing, *Anal Chem* **72**, 5151-7.
- [160] Mootz HD & Muir TW (2002) Protein splicing triggered by a small molecule, *J Am Chem Soc* **124**, 9044-5.
- [161] Mootz HD, Blum ES, Tyszkiewicz AB & Muir TW (2003) Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo, *J Am Chem Soc* **125**, 10561-9.
- [162] Brenzel S & Mootz HD (2005) Design of an intein that can be inhibited with a small molecule ligand, *J Am Chem Soc* **127**, 4176-7.
- [163] Brenzel S, Kurpiers T & Mootz HD (2006) Engineering artificially split inteins for applications in protein chemistry: biochemical characterization of the split Ssp DnaB intein and comparison to the split Sce VMA intein, *Biochemistry* **45**, 1571-8.

- [164] Kurpiers T & Mootz HD (2007) Regioselective cysteine bioconjugation by appending a labeled cystein tag to a protein by using protein splicing in trans, *Angew Chem Int Ed Engl* **46**, 5234-7.
- [165] Kurpiers T & Mootz HD (2008) Site-specific chemical modification of proteins with a prelabelled cysteine tag using the artificially split Mxe GyrA intein, *Chembiochem* **9**, 2317-25.
- [166] Sun W, Yang J & Liu XQ (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing, *J Biol Chem* **279**, 35281-6.
- [167] Aranko AS, Zuger S, Buchinger E & Iwai H (2009) In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins, *PLoS One* **4**, e5185.
- [168] Oeemig JS, Aranko AS, Djupsjobacka J, Heinamaki K & Iwai H (2009) 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 Lett* **583**, 1451-6.
- [169] Ludwig C, Pfeiff M, Linne U & Mootz HD (2006) Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing, *Angew Chem Int Ed Engl* **45**, 5218-21.
- [170] Ando T, Tsukiji S, Tanaka T & Nagamune T (2007) Construction of a small-molecule-integrated semisynthetic split intein for in vivo protein ligation, *Chem Commun (Camb)*, 4995-7.
- [171] Appleby JH, Zhou K, Volkmann G & Liu XQ (2009) Novel split intein for trans-splicing synthetic peptide onto C terminus of protein, *J Biol Chem* **284**, 6194-9.
- [172] Giriat I & Muir TW (2003) Protein semi-synthesis in living cells, *J Am Chem Soc* **125**, 7180-1.
- [173] Martin DD, Xu MQ & Evans TC, Jr. (2001) Characterization of a naturally occurring trans-splicing intein from *Synechocystis* sp. PCC6803, *Biochemistry* **40**, 1393-402.
- [174] Shi J & Muir TW (2005) Development of a tandem protein trans-splicing system based on native and engineered split inteins, *J Am Chem Soc* **127**, 6198-206.
- [175] Brenzel S (2009) Künstlich gespaltene Inteine für die Protein-Semisynthese - Charakterisierung des *Ssp* DnaB Inteins und Modifikation des Ionenkanals *OmpF* mit Hilfe des *Psp*-GBD Pol Inteins, Doktorarbeit, *TU Dortmund*.
- [176] Borg N (2007) Entwicklung von induzierbarem Protein *trans*-Spleißen basierend auf dem *Ssp* DnaB Intein, Bachelorarbeit, *TU Dortmund*.
- [177] Nichols NM, Benner JS, Martin DD & Evans TC, Jr. (2003) Zinc ion effects on individual *Ssp* DnaE intein splicing steps: regulating pathway progression, *Biochemistry* **42**, 5301-11.
- [178] Lew BM, Mills KV & Paulus H (1999) Characteristics of protein splicing in trans mediated by a semisynthetic split intein, *Biopolymers* **51**, 355-62.
- [179] Mills KV, Manning JS, Garcia AM & Wuerdeman LA (2004) Protein splicing of a *Pyrococcus abyssi* intein with a C-terminal glutamine, *J Biol Chem* **279**, 20685-91.
- [180] Evans TC, Jr. et al. (2000) Protein trans-splicing and cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* species PCC6803, *J Biol Chem* **275**, 9091-4.
- [181] Kwon Y, Coleman MA & Camarero JA (2006) Selective immobilization of proteins onto solid supports through split-intein-mediated protein trans-splicing, *Angew Chem Int Ed Engl* **45**, 1726-9.
- [182] Olschewski D et al. (2007) Semisynthetic murine prion protein equipped with a GPI anchor mimic incorporates into cellular membranes, *Chem Biol* **14**, 994-1006.
- [183] Hermanson GT (2008) Bioconjugate techniques, *Elsevier*.
- [184] Brenzel S, Cebi M, Reiss P, Koert U & Mootz HD (2009) Expanding the scope of protein trans-splicing to fragment ligation of an integral membrane protein: towards modulation of porin-based ion channels by chemical modification, *Chembiochem* **10**, 983-6.
- [185] Otomo T, Ito N, Kyogoku Y & Yamazaki T (1999) NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation, *Biochemistry* **38**, 16040-4.
- [186] Yagi H, Tsujimoto T, Yamazaki T, Yoshida M & Akutsu H (2004) Conformational change of H<sup>+</sup>-ATPase beta monomer revealed on segmental isotope labeling NMR spectroscopy, *J Am Chem Soc* **126**, 16632-8.
- [187] Busche AE, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dotsch V & Iwai H (2009) Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein, *Angew Chem Int Ed Engl* **48**, 6128-31.
- [188] Zuger S & Iwai H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies, *Nat Biotechnol* **23**, 736-40.
- [189] Muona M, Aranko AS & Iwai H (2008) Segmental isotopic labelling of a multidomain protein by protein ligation by protein trans-splicing, *Chembiochem* **9**, 2958-61.
- [190] Scott CP, Abel-Santos E, Wall M, Wahn DC & Benkovic SJ (1999) Production of cyclic peptides and proteins in vivo, *Proc Natl Acad Sci U S A* **96**, 13638-43.
- [191] Iwai H, Lingel A & Pluckthun A (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-Pf1 intein from *Pyrococcus furiosus*, *J Biol Chem* **276**, 16548-54.
- [192] Volkmann G, Murphy PW, Rowland EE, Cronan JE, Jr., Liu XQ, Blouin C & Byers DM (2010) Intein-mediated cyclization of bacterial acyl carrier protein stabilizes its folded conformation but does not abolish function, *J Biol Chem* **285**, 8605-14.
- [193] Scott CP, Abel-Santos E, Jones AD & Benkovic SJ (2001) Structural requirements for the biosynthesis of backbone cyclic peptide libraries, *Chem Biol* **8**, 801-15.
- [194] Kinsella TM et al. (2002) Retrovirally delivered random cyclic Peptide libraries yield inhibitors of interleukin-4 signaling in human B cells, *J Biol Chem* **277**, 37512-8.
- [195] Kanno A, Yamanaka Y, Hirano H, Umezawa Y & Ozawa T (2007) Cyclic luciferase for real-time sensing of caspase-3 activities in living mammals, *Angew Chem Int Ed Engl* **46**, 7595-9.
- [196] Ozawa T, Kaihara A, Sato M, Tachihara K & Umezawa Y (2001) Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing, *Anal Chem* **73**, 2516-21.

- [197] Ozawa T, Takeuchi TM, Kaihara A, Sato M & Umezawa Y (2001) Protein splicing-based reconstitution of split green fluorescent protein for monitoring protein-protein interactions in bacteria: improved sensitivity and reduced screening time, *Anal Chem* **73**, 5866-74.
- [198] Kanno A, Ozawa T & Umezawa Y (2006) Genetically encoded optical probe for detecting release of proteins from mitochondria toward cytosol in living cells and mammals, *Anal Chem* **78**, 8076-81.
- [199] Daniell H (2002) Molecular strategies for gene containment in transgenic crops, *Nat Biotechnol* **20**, 581-6.
- [200] Yang J, Fox GC, Jr. & Henry-Smith TV (2003) Intein-mediated assembly of a functional beta-glucuronidase in transgenic plants, *Proc Natl Acad Sci U S A* **100**, 3513-8.
- [201] Dun BQ et al. (2007) Reconstitution of glyphosate resistance from a split 5-enolpyruvyl shikimate-3-phosphate synthase gene in Escherichia coli and transgenic tobacco, *Appl Environ Microbiol* **73**, 7997-8000.
- [202] Evans TC, Jr., Benner J & Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element, *Protein Sci* **7**, 2256-64.
- [203] Schwarzer D & Cole PA (2005) Protein semisynthesis and expressed protein ligation: chasing a protein's tail, *Curr Opin Chem Biol* **9**, 561-9.
- [204] Flavell RR & Muir TW (2009) Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology, *Acc Chem Res* **42**, 107-16.
- [205] Rose RE (1988) The nucleotide sequence of pACYC184, *Nucleic Acids Res* **16**, 355.
- [206] Bartolomé B, Jubete, Y., Martinez, E., de la Cruz, F. (1991) Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives, *Gene* **102**, 75-78.
- [207] Sambrook J, Fritsch, E. F., Maniatis, T. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [208] Birnboim HC, Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucleic Acids Res* **7**, 1513-1523.
- [209] Mullis KB & Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction, *Methods Enzymol* **155**, 335-50.
- [210] Braman J, Papworth C & Greener A (1996) Site-directed mutagenesis using double-stranded plasmid DNA templates, *Methods Mol Biol* **57**, 31-44.
- [211] Zettler J (2006) Präparation und spektroskopische Untersuchungen an fluorophormarkierten Peptidsynthetasenfragmenten, *Philipps-Universität Marburg*.
- [212] Urbanek A (2007) Untersuchungen für die segmentelle Isotopenmarkierung einer nichtribosomalen Peptidsynthetase durch Protein-Spleißen *in trans*, Bachelorarbeit, *TU Dortmund*.
- [213] Schütz V (2008) Biochemische Charakterisierung des Npu DnaE Inteins. Masterarbeit, *TU Dortmund*.
- [214] Zettler J, Schutz V & Mootz HD (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction, *FEBS Lett* **583**, 909-14.
- [215] van den Ent F & Lowe J (2006) RF cloning: a restriction-free method for inserting target genes into plasmids, *J Biochem Biophys Methods* **67**, 67-74.
- [216] Laemmli UL (1970) Cleavage of structural proteins during the assembly of the head of the bacteriophage T4, *Nature* **227**, 680-685.
- [217] Gill SC & von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data, *Anal Biochem* **182**, 319-26.
- [218] Pervushin K, Riek R, Wider G & Wuthrich K (1997) 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, *Proc Natl Acad Sci U S A* **94**, 12366-71.
- [219] Finking R, Neumuller A, Solsbacher J, Konz D, Kretzschmar G, Schweitzer M, Krumm T & Marahiel MA (2003) Aminoacyl adenylate substrate analogues for the inhibition of adenylation domains of nonribosomal peptide synthetases, *ChemBiochem* **4**, 903-906.
- [220] Dieckmann R, Pavela-Vrancic M, Pfeifer E, von Dohren H & Kleinkauf H (1997) The adenylation domain of tyrocidine synthetase 1--structural and functional role of the interdomain linker region and the (S/T)GT(T/S)GXPKG core sequence, *Eur J Biochem* **247**, 1074-82.
- [221] Dieckmann R, Pavela-Vrancic M, von Dohren H & Kleinkauf H (1999) Probing the domain structure and ligand-induced conformational changes by limited proteolysis of tyrocidine synthetase 1, *J Mol Biol* **288**, 129-40.
- [222] Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M & Zamboni M (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin, *Biochemistry* **25**, 1847-51.
- [223] Arnone MI, Birolo L, Giamberini M, Cubellis MV, Nitti G, Sanna G & Marino G (1992) Limited proteolysis as a probe of conformational changes in aspartate aminotransferase from Sulfolobus solfataricus, *Eur J Biochem* **204**, 1183-9.
- [224] Cui J & Somerville RL (1993) The TyrR protein of Escherichia coli, analysis by limited proteolysis of domain structure and ligand-mediated conformational changes, *J Biol Chem* **268**, 5040-7.
- [225] Pavela-Vrancic M, Pfeifer E, Schroder W, von Dohren H & Kleinkauf H (1994) Identification of the ATP binding site in tyrocidine synthetase 1 by selective modification with fluorescein 5'-isothiocyanate, *J Biol Chem* **269**, 14962-6.
- [226] Johnson MA, Southworth MW, Herrmann T, Brace L, Perler FB & Wuthrich K (2007) NMR structure of a K1bA intein precursor from Methanococcus jannaschii, *Protein Sci* **16**, 1316-28.
- [227] Forrer P & Jaussi R (1998) High-level expression of soluble heterologous proteins in the cytoplasm of Escherichia coli by fusion to the bacteriophage lambda head protein D, *Gene* **224**, 45-52.
- [228] Iwai H, Forrer P, Pluckthun A & Guntert P (2005) NMR solution structure of the monomeric form of the bacteriophage lambda capsid stabilizing protein gpD, *J Biomol NMR* **31**, 351-6.
- [229] Wright PE & Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, *J Mol Biol* **293**, 321-31.
- [230] Dunker AK et al. (2001) Intrinsically disordered protein, *J Mol Graph Model* **19**, 26-59.

- [231] Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci* **11**, 739-56.
- [232] Uversky VN (2010) The mysterious unfoldome: structureless, underappreciated, yet vital part of any given proteome, *J Biomed Biotechnol*.
- [233] Dolgikh DA, Gilmanshin RI, Brazhnikov EV, Bychkova VE, Semisotnov GV, Venyaminov S & Ptitsyn OB (1981) Alpha-Lactalbumin: compact state with fluctuating tertiary structure?, *FEBS Lett* **136**, 311-5.
- [234] Ptitsyn OB & Uversky VN (1994) The molten globule is a third thermodynamical state of protein molecules, *FEBS Lett* **341**, 15-8.
- [235] Ohgushi M & Wada A (1983) 'Molten-globule state': a compact form of globular proteins with mobile side-chains, *FEBS Lett* **164**, 21-4.
- [236] Baum J, Dobson CM, Evans PA & Hanley C (1989) Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin, *Biochemistry* **28**, 7-13.
- [237] Eliezer D, Yao J, Dyson HJ & Wright PE (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding, *Nat Struct Biol* **5**, 148-55.
- [238] Alexandrescu AT, Evans PA, Pitkeathly M, Baum J & Dobson CM (1993) Structure and dynamics of the acid-denatured molten globule state of alpha-lactalbumin: a two-dimensional NMR study, *Biochemistry* **32**, 1707-18.
- [239] Vamvaca K, Vogeli B, Kast P, Pervushin K & Hilvert D (2004) An enzymatic molten globule: efficient coupling of folding and catalysis, *Proc Natl Acad Sci U S A* **101**, 12860-4.
- [240] Heinamaki K, Oeemig JS, Djupsjobacka J & Iwai H (2009) NMR resonance assignment of DnaE intein from *Nostoc punctiforme*, *Biomol NMR Assign* **3**, 41-3.
- [241] Southworth MW, Amaya K, Evans TC, Xu MQ & Perler FB (1999) Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein, *Biotechniques* **27**, 110-4, 116, 118-20.
- [242] Sonntag T (2006) Entwicklung einer *cis* und *trans* Inteinkassette, *Philipps-Universität Marburg*.
- [243] Geiser M, Cebe R, Drewello D & Schmitz R (2001) Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase, *Biotechniques* **31**, 88-90, 92.
- [244] Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P & Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications, *Yeast* **14**, 115-32.
- [245] Lockless SW & Muir TW (2009) Traceless protein splicing utilizing evolved split inteins, *Proc Natl Acad Sci U S A* **106**, 10999-1004.
- [246] Miyada CG, Stoltzfus L & Wilcox G (1984) Regulation of the *araC* gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on *araBAD* expression, *Proc Natl Acad Sci U S A* **81**, 4120-4.
- [247] Guzman LM, Belin D, Carson MJ & Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter, *J Bacteriol* **177**, 4121-30.
- [248] Bryngelson JD, Onuchic JN, Socci ND & Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis, *Proteins* **21**, 167-95.
- [249] Frauenfelder H, Sligar SG & Wolynes PG (1991) The energy landscapes and motions of proteins, *Science* **254**, 1598-603.
- [250] Leopold PE, Montal M & Onuchic JN (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship, *Proc Natl Acad Sci U S A* **89**, 8721-5.
- [251] Lazaridis T & Karplus M (1997) "New view" of protein folding reconciled with the old through multiple unfolding simulations, *Science* **278**, 1928-31.
- [252] Ansari A, Berendzen J, Bowne SF, Frauenfelder H, Iben IE, Sauke TB, Shyamsunder E & Young RD (1985) Protein states and proteinquakes, *Proc Natl Acad Sci U S A* **82**, 5000-4.
- [253] Frauenfelder H et al. (2009) A unified model of protein dynamics, *Proc Natl Acad Sci U S A* **106**, 5129-34.
- [254] Vendruscolo M & Dobson CM (2006) Structural biology. Dynamic visions of enzymatic reactions, *Science* **313**, 1586-7.
- [255] Henzler-Wildman K & Kern D (2007) Dynamic personalities of proteins, *Nature* **450**, 964-72.
- [256] Dill KA & Chan HS (1997) From Levinthal to pathways to funnels, *Nat Struct Biol* **4**, 10-9.
- [257] Hilser VJ, Garcia-Moreno EB, Oas TG, Kapp G & Whitten ST (2006) A statistical thermodynamic model of the protein ensemble, *Chem Rev* **106**, 1545-58.
- [258] Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis, *Proc Natl Acad Sci U S A* **44**, 98-104.
- [259] Tsai CJ, Kumar S, Ma B & Nussinov R (1999) Folding funnels, binding funnels, and protein function, *Protein Sci* **8**, 1181-90.
- [260] Boehr DD, Dyson HJ & Wright PE (2006) An NMR perspective on enzyme dynamics, *Chem Rev* **106**, 3055-79.
- [261] Kapur S & Khosla C (2008) Biochemistry: Fit for an enzyme, *Nature* **454**, 832-3.
- [262] Liu Y & Bruner SD (2007) Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases, *Chembiochem* **8**, 617-621.
- [263] von Nussbaum F, Brands M, Hinzen B, Weigand S & Habich D (2006) Antibacterial natural products in medicinal chemistry--exodus or revival?, *Angew Chem Int Ed Engl* **45**, 5072-129.
- [264] Floss HG (2006) Combinatorial biosynthesis--potential and problems, *J Biotechnol* **124**, 242-57.
- [265] Weist S et al. (2002) Fluorobalhimycin--a new chapter in glycopeptide antibiotic research, *Angew Chem Int Ed Engl* **41**, 3383-5.
- [266] Weist S, Kittel C, Bischoff D, Bister B, Pfeifer V, Nicholson GJ, Wohlleben W & Süssmuth RD (2004) Mutasynthesis of glycopeptide antibiotics: variations of vancomycin's AB-ring amino acid 3,5-dihydroxyphenylglycine, *J Am Chem Soc* **126**, 5942-3.
- [267] Eppelmann K, Stachelhaus T & Marahiel MA (2002) Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics, *Biochemistry* **41**, 9718-26.

- [268] Stachelhaus T, Schneider A & Marahiel MA (1995) Rational Design of Peptide Antibiotics by Targeted Replacement of Bacterial and Fungal Domains, *Science* **269**, 69-72.
- [269] Mootz HD, Kessler N, Linne U, Eppelmann K, Schwarzer D & Marahiel MA (2002) Decreasing the ring size of a cyclic nonribosomal peptide antibiotic by in-frame module deletion in the biosynthetic genes, *J Am Chem Soc* **124**, 10980-1.
- [270] Baltz RH (2008) Biosynthesis and genetic engineering of lipopeptide antibiotics related to daptomycin, *Curr Top Med Chem* **8**, 618-38.
- [271] Baltz RH (2009) Daptomycin: mechanisms of action and resistance, and biosynthetic engineering, *Curr Opin Chem Biol* **13**, 144-51.
- [272] Nguyen KT, Ritz D, Gu JQ, Alexander D, Chu M, Miao V, Brian P & Baltz RH (2006) Combinatorial biosynthesis of novel antibiotics related to daptomycin, *Proc Natl Acad Sci U S A* **103**, 17462-7.
- [273] Miao V et al. (2006) Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics, *Chem Biol* **13**, 269-76.
- [274] Garner E, Cannon P, Romero P, Obradovic Z & Dunker AK (1998) Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization, *Genome Inform Ser Workshop Genome Inform* **9**, 201-213.
- [275] Uversky VN, Gillespie JR & Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions?, *Proteins* **41**, 415-27.
- [276] Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I & Sussman JL (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded, *Bioinformatics* **21**, 3435-3438.
- [277] Uversky VN (2002) What does it mean to be natively unfolded?, *Eur J Biochem* **269**, 2-12.
- [278] Romero P, Obradovic Z & Dunker AK (2004) Natively disordered proteins: functions and predictions, *Appl Bioinformatics* **3**, 105-13.
- [279] Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guillot S & Dunker AK (1998) Thousands of proteins likely to have long disordered regions, *Pac Symp Biocomput*, 437-48.
- [280] Spolar RS & Record MT, Jr. (1994) Coupling of local folding to site-specific binding of proteins to DNA, *Science* **263**, 777-84.
- [281] Shoemaker BA, Portman JJ & Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, *Proc Natl Acad Sci U S A* **97**, 8868-73.
- [282] Gabdouliline RR & Wade RC (2002) Biomolecular diffusional association, *Curr Opin Struct Biol* **12**, 204-13.
- [283] Selzer T & Schreiber G (2001) New insights into the mechanism of protein-protein association, *Proteins* **45**, 190-8.
- [284] Goh CS, Milburn D & Gerstein M (2004) Conformational changes associated with protein-protein interactions, *Curr Opin Struct Biol* **14**, 104-9.
- [285] Sugase K, Dyson HJ & Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature* **447**, 1021-5.
- [286] Dyson HJ & Wright PE (2002) Coupling of folding and binding for unstructured proteins, *Curr Opin Struct Biol* **12**, 54-60.
- [287] Ghosh I, Sun L & Xu MQ (2001) Zinc inhibition of protein trans-splicing and identification of regions essential for splicing and association of a split intein\*, *J Biol Chem* **276**, 24051-8.
- [288] Hiraga K et al. (2009) Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center, *J Mol Biol* **393**, 1106-17.
- [289] Gangopadhyay JP, Jiang SQ, van Berkel P & Paulus H (2003) In vitro splicing of erythropoietin by the *Mycobacterium tuberculosis* RecA intein without substituting amino acids at the splice junctions, *Biochim Biophys Acta* **1619**, 193-200.
- [290] Lutz R & Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I<sub>1</sub>-I<sub>2</sub> regulatory elements, *Nucleic Acids Research* **25**, 1203-1210.

## Anhang

Tabelle 9A Übersicht der Expressionsplasmide.

Name / Nr.	Protein	Masse	Plasmid	Vektor
GrsA A-PCP	GrsA A-PCP-H <sub>6</sub>	71125,3	pJZ06	pET16b
GrsA A-PCP(Δ-4Cys)	GrsA A-PCP(Δ-4Cys)-H <sub>6</sub>	71089,1	pJZ13	pET16b
GrsB1	ST-GrsB1 C-A-PCP-H <sub>6</sub>	124181,9	pJZ56b	pET28a
<b>1</b>	ST-gpD- <i>Npu</i> DnaE <sup>N</sup>	24721,8	pAU08	pRSFDuet
<b>1<sub>C1A</sub></b>	ST-gpD- <i>Npu</i> DnaE <sup>N</sup> <sub>C1A</sub>	24689,7	pJZ62	pRSFDuet
<b>2</b>	<i>Npu</i> DnaE <sup>C</sup> -gpD-H <sub>6</sub>	17250,3	pJZ68	pET42b
<b>2<sub>N36A</sub></b>	<i>Npu</i> DnaE <sup>C</sup> <sub>N36A</sub> -gpD-H <sub>6</sub>	17207,2	pJZ69	pET42b
<b>2<sub>N36A, C+1A</sub></b>	<i>Npu</i> DnaE <sup>C</sup> <sub>N36A, C+1A</sub> -gpD-H <sub>6</sub>	17175,2	pJZ82	pET42b
<b>3</b>	ST- <i>Npu</i> DnaE <sup>C</sup> -gpD-H <sub>6</sub>	18737,9	pJZ57b	pET16b
<b>4</b>	ST- <i>Npu</i> DnaE <sup>C</sup> -Trx-H <sub>6</sub>	19353,0	pJZ70	pET16b
<b>5</b>	ST-gpD-H <sub>6</sub>	13949,4	pJZ49	pET28a
<b>6</b>	ST-gpD- <i>Ssp</i> DnaB <sup>C</sup> (107-154)-H <sub>6</sub>	20518,8	pJZ35	pET28a
<b>7</b>	<i>Ssp</i> DnaB <sup>C</sup> (12-154)-gpD-H <sub>6</sub>	29746,6	pJZ29	pSU38
<b>8</b>	H <sub>6</sub> -gpD- <i>Npu</i> DnaE <sup>N</sup>	24433,4	pJZ83	pRSFDuet
<b>8<sub>C1A</sub></b>	H <sub>6</sub> -gpD- <i>Npu</i> DnaE <sup>N</sup> <sub>C1A</sub>	24401,3	pJZ84	pRSFDuet
<b>9</b>	ST-gpD-Trx-H <sub>6</sub>	26386,6	pSE02	pET28a
<b>10</b>	ST- <i>Ssp</i> DnaB <sup>C</sup> (12-154)-GrsA A-PCP-H <sub>6</sub>	89620,4	pJZ16	pSU38
<b>11</b>	GrsA A-PCP(Δ-4Cys, T341C, D508C)-H <sub>6</sub>	71079,2	pJZ67	pET16b
<b>12</b>	ST-GrsA A <sub>1-421</sub> - <i>Ssp</i> DnaB <sup>N</sup> (1-104)	60529,5	pAU06	pRSFDuet
<b>13</b>	<i>Ssp</i> DnaB <sup>C</sup> (107-154)-S-GrsA <sub>423-540</sub> -PCP-H <sub>6</sub>	29385,4	pAU04	pBAD

**Tabelle 10A Zur Integration der *Ssp* DnaB- und *Npu* DnaE-Inteinkassetten in das Beispielgen ST-gpD-Trx-H<sub>6</sub> verwendeten Oligonukleotide.**

Oligo-nukleotid	Sequenz
oSE01	5'-AATCATCAGT CAGATGAATA ATTTTATCAG ATCCTCCTCC AGAGTTATGG ACAATGATG
oSE02	5'-ATCAAAAAGAA TCATCAGTCA GATGAATAAT TTTATCAGAC TGTTCAATAG AGTTATGGAC
oSE03	5'-TGCCGGAACG GCAATCAGCA TCGTTGAATT CGGCGGAGGC TGCATCAGTG GAGATAGTT
oSE04	5'-GCGTTTGCCG GAACGGCAAT CAGCATCGTT GAATTCGAAA GCGGATGCAT CAGTGGAGAT
oSE05	5'-TGCCGGAACG GCAATCAGCA TCGTTGAATT CGGCGGAGGC TGTTTAAGCT ATGAAACGG
oSE06	5'-GACCGCGTTT GCCGGAACGG CAATCAGCAT CGTTGAATTC GCGGAGTACT GTTTAAGCTA TGAAAC
oSE07	5'-AATCATCAGT CAGATGAATA ATTTTATCAG ATCCTCCTCC ACAATTAGAA GCTATGAAGC
oSE08	5'-ATCAAAAAGAA TCATCAGTCA GATGAATAATT TTATCAGATT TATTGAAACA ATTAGAAGC

**Tabelle 11A In dieser Arbeit über Festphasenpeptidsynthese hergestellte und verwendete Peptide.**

Peptid	Sequenz	Masse /Da
<b>Pep1</b>	Fluorescein-KKESG <u>CISGDSLISLA</u>	1964,9
<b>Pep2</b>	Biotin-βAKKK(Fluorescein)ESG <u>CISGDSLISLA</u>	2390,1

Die *Ssp* DnaB Int<sup>N</sup>(1-11) Sequenz ist unterstrichen.

## A

```

10    20    30    40    50    60    70    80    90    100   110   120
MVNSSKSLI HAQNKNGTHE EEQYLFVAVNN TKAEYPRDKT IHQLFEEQVSKRPNNVAIVC ENEQLTYHEL NVKANQLARI FIEKGIGKDT LVGIMMEKSI DLFIGILAVL KAGGAYVPID
130   140   150   160   170   180   190   200   210   220   230   240
IEYFKERIYQ ILDDSQARM LTKQHLVHLI HNIQFNGQVE IFEEDTIKIR EGTNLHVPSK STDLAYVIYT SGTGNPKGT MLEHKGISNL KVFENSLNV TEKDRIGQFA SISFDASVWE
250   260   270   280   290   300   310   320   330   340   350   360
MFMAALTGAS LYIILKDTIN DFKFEQYIN QKEITVITLP PTYVHLDPE RILSIQTLIT AGSATSPSLV NKWKEKVYI NAYGPTETI CATTWATKE TIGHSVPIGA FIQNTQIYIV
370   380   390   400   410   420   430   440   450   460   470   480
DENLQLKSVG EAGELCIGGE GLARGYKRP ELTSQKFDVN PFVPEKLYK TGDQARWLSG GNIYELGRID NQVKIRGHRV ELEEVEISILL KHYISETAV SVHKDHQEQF YLCAYFVSEK
490   500   510   520   530   540   550   560   570   580   590   600
HIPLEQLRQF SSEELPTYMI PSYFIQLDKM PLTNSGKIDR KQLPEPDLTF GMRVDYEAPR NEIEETLVTI WQVVLGIEKI GIKDNFYALG GDSIKAIQVA ARLHSYQLKL ETKDLLKYPT
610   620
IDQLVHYIKD SKRERSRSHH HHHH

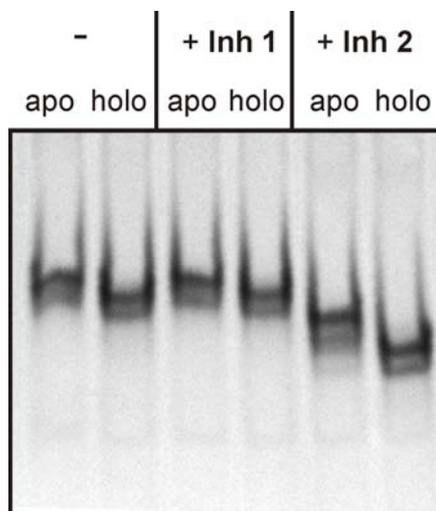
```

## B

Fragment	Peptides detected	m/z calc.	m/z meas.
A-PCP-H <sub>6</sub> (Complete protein)	AS 7-15 SILIHQNK	1023.60	1023.46
	AS 16-32 NGTHEEEQYLFVAVNNTK	1993.93	1993.88
	AS 40-51 TIHQLFEEQVSK	1458.76	1458.67
	AS 52-73 RPNNVAIVC (PAM) ENEQLTYHELNVK	2654.33	2653.93
	AS 112-125 AGGAYVPIDIEYPK	1492.77	1492.69
	AS 128-138 IQYILDDSQAR	1321.68	1321.60
	AS 181-198 STDLAYVIYTSGTTGNPK	1887.93	1887.84
	AS 265-272 FEQYINQK	1069.53	1069.42
	AS 273-291 EITVITLPPTYVHLDPER	2192.20	2192.15
	AS 292-312 ILSIQTLITAGSATSPSLVSK	2114.21	2114.05
	AS 340-367 ETIGHSVPIGAPIQNTQIYIVDENLQLK	3090.65	3090.01
	AS 368-384 SVGEAGELC (PAM) IGGEGLAR	1688.83	1688.77
	AS 397-407 FVDNPFVPGEK	1248.63	1248.50
	AS 417-428 WLSDGNIEYLGR	1422.70	1422.65
	AS 481-488 HIPLEQLR	1005.58	1005.51
	AS 489-509 QFSSEELPTYMIPSYFIQLDK	2536.23	2535.89
	AS 522-533 QLPEPDLTFGMR	1403.70	1403.63
AS 534-540 VDYEAPR	849.41	849.36	
AS 564-575 DNFYALGGDSIK	1299.62	1299.49	
AS 598-609 YPTIDQLVHYIK	1489.91	1489.72	
A	AS 16-32 NGTHEEEQYLFVAVNNTK	1993.93	1993.95
	AS 40-51 TIHQLFEEQVSK	1458.76	1458.77
	AS 52-73 RPNNVAIVC (PAM) ENEQLTYHELNVK	2654.33	2654.02
	AS 112-125 AGGAYVPIDIEYPK	1492.77	1492.77
	AS 128-138 IQYILDDSQAR	1321.68	1321.72
	AS 181-198 STDLAYVIYTSGTTGNPK	1887.93	1887.94
	AS 265-272 FEQYINQK	1069.53	1069.51
	AS 273-291 EITVITLPPTYVHLDPER	2192.20	2192.19
	AS 397-407 FVDNPFVPGEK	1248.63	1248.60
	AS 417-428 WLSDGNIEYLGR	1422.70	1422.74
AS 481-488 HIPLEQLR	1005.58	1005.61	
AS 489-509 QFSSEELPTYMIPSYFIQLDK	2536.23	2536.03	
AS 522-533 QLPEPDLTFGMR	1403.70	1403.73	
A <sup>N</sup>	AS 16-32 NGTHEEEQYLFVAVNNTK	1993.93	1993.87
	AS 40-51 TIHQLFEEQVSK	1458.76	1458.69
	AS 112-125 AGGAYVPIDIEYPK	1492.77	1492.69
	AS 128-138 IQYILDDSQAR	1321.68	1321.64
	AS 181-198 STDLAYVIYTSGTTGNPK	1887.93	1887.86
	AS 265-272 FEQYINQK	1069.53	1069.43
	AS 273-291 EITVITLPPTYVHLDPER	2192.20	2192.09
	AS 340-367 ETIGHSVPIGAPIQNTQIYIVDENLQLK	3090.65	3090.99
AS 397-407 FVDNPFVPGEK	1248.63	1248.53	
AS 417-428 WLSDGNIEYLGR	1422.70	1422.66	
PCP	AS 534-540 VDYEAPR	849.41	849.38
	AS 564-575 DNFYALGGDSIK	1299.62	1299.58
	AS 583-589 LHSYQLK	888.49	888.43

**Abbildung 71A Massenspektrometrische Analyse der bei der partiellen Proteolyse von apo-A-PCP entstandenen Proteinfragmente.** A) Aminosäuresequenz des in der vorliegenden Arbeit verwendeten GrsA A-PCP Proteins; dabei sind die N-terminale Subdomäne der A-Domäne A<sup>N</sup> in rot, die C-terminale Subdomäne A<sup>C</sup> in schwarz, der Linkerbereich in blau, die PCP-Domäne in grün und der zusätzliche His<sub>6</sub>-Tag in grau dargestellt; die Bestimmung des Linkerbereichs erfolgte durch einen Vergleich mit der Kristallstruktur des kompletten

Terminationsmoduls SrfA-C [94]; B) Nach der SDS-PAGE wurden die wichtigsten entstandenen Proteinfragmentbänder ausgeschnitten und tryptisch im Gel verdaut; die anschließend in der MALDI-TOF MS Analyse gefundenen Peptidfragmente sind tabellarisch dargestellt; die Ergebnisse bestätigen den in A) gezeigten Domänenbau und ermöglichten die Bestimmung der Proteinfragmente.

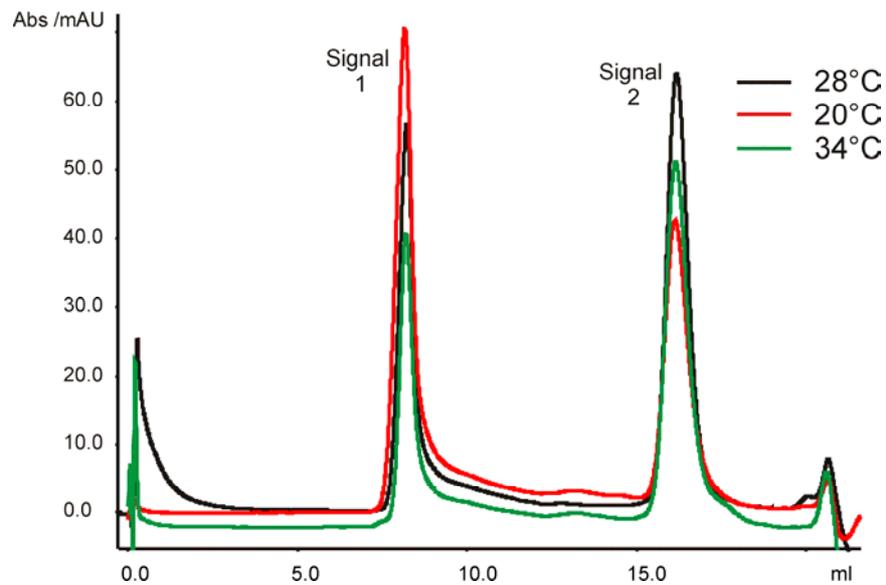


**Abbildung 72A Native PAGE der TycB1 A-PCP Proteine in Abwesenheit und Gegenwart von Inhibitoren.** Sowohl die apo-, als auch die holo-Form wurden ohne oder mit den Inhibitoren (jeweils 100  $\mu$ M) inkubiert und anschließend der native PAGE unterzogen; nur die kognate Verbindung, 5'-O-[N-(L-prolyl)-sulfamoyl] Adenosin (**Inh 2**), verändert die elektrophoretische Mobilität der A-PCP Proteine.

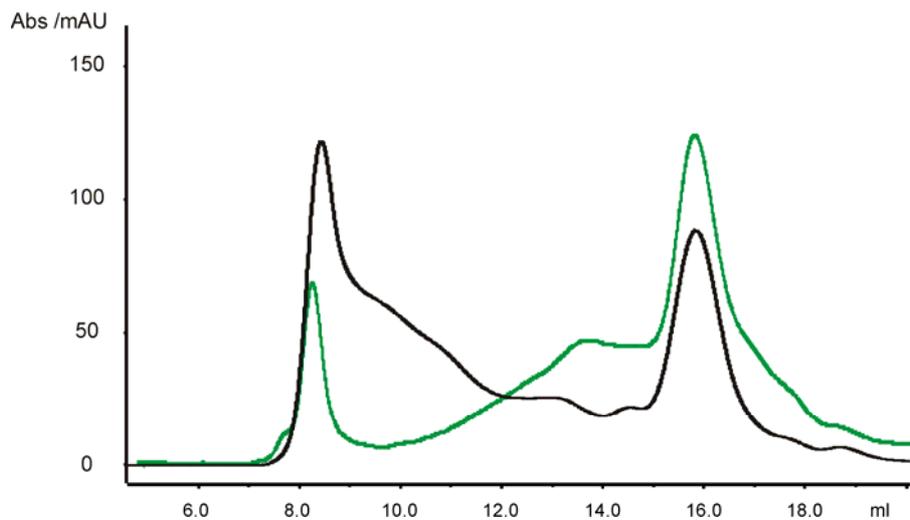
**Tabelle 12A Reaktionsraten und prozentuale Spleißproduktbildung der *Npu* DnaE Konstrukte.**

N-terminales Konstrukt	C-terminales Konstrukt	T /°C	Spleißproduktbildung /%	$k_{\text{Spleißen}} / \text{s}^{-1}$
1 (N)*	3 (N)*	20	75 - 85	$(2.4 \pm 0.1) \times 10^{-3}$
		37	85 - 95	$(6.2 \pm 0.8) \times 10^{-3}$
	3 (D)*	20	75 - 85	$(2.4 \pm 0.1) \times 10^{-3}$
		37	85 - 95	$(6.6 \pm 0.3) \times 10^{-3}$
	2 (D)*	20	80 - 85	$(3.6 \pm 0.5) \times 10^{-3}$
		37	80 - 90	$(9.0 \pm 0.4) \times 10^{-3}$
1 (N)*	4 (N)*	37	85 - 90	$(7.1 \pm 0.9) \times 10^{-3}$
	4 (D)*	37	85 - 90	$(6.6 \pm 1.3) \times 10^{-3}$
	Int <sup>C</sup> -Trx-H <sub>6</sub> (D)*	37	75 - 80	$(1.1 \pm 0.2) \times 10^{-2}$

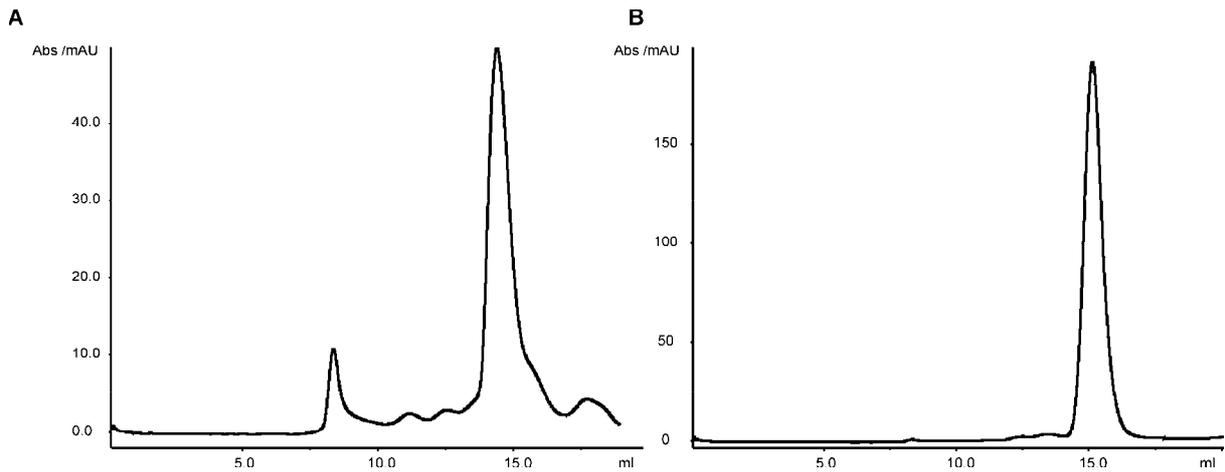
\* (N) bedeutet native Reinigung und (D) bedeutet denaturierende Reinigung des Proteins; Proteinomenklatur: 1 = ST-gpD-Int<sup>N</sup>; 2 = Int<sup>C</sup>-gpD-H<sub>6</sub>; 3 = ST-Int<sup>C</sup>-gpD-H<sub>6</sub>; 4 = ST-Int<sup>C</sup>-Trx-H<sub>6</sub>; sowohl die Reaktionsraten, als auch die Ausbeuten der PTS-Reaktionen sind nicht von den Reinigungsbedingungen der C-terminalen Inteinhälften abhängig.



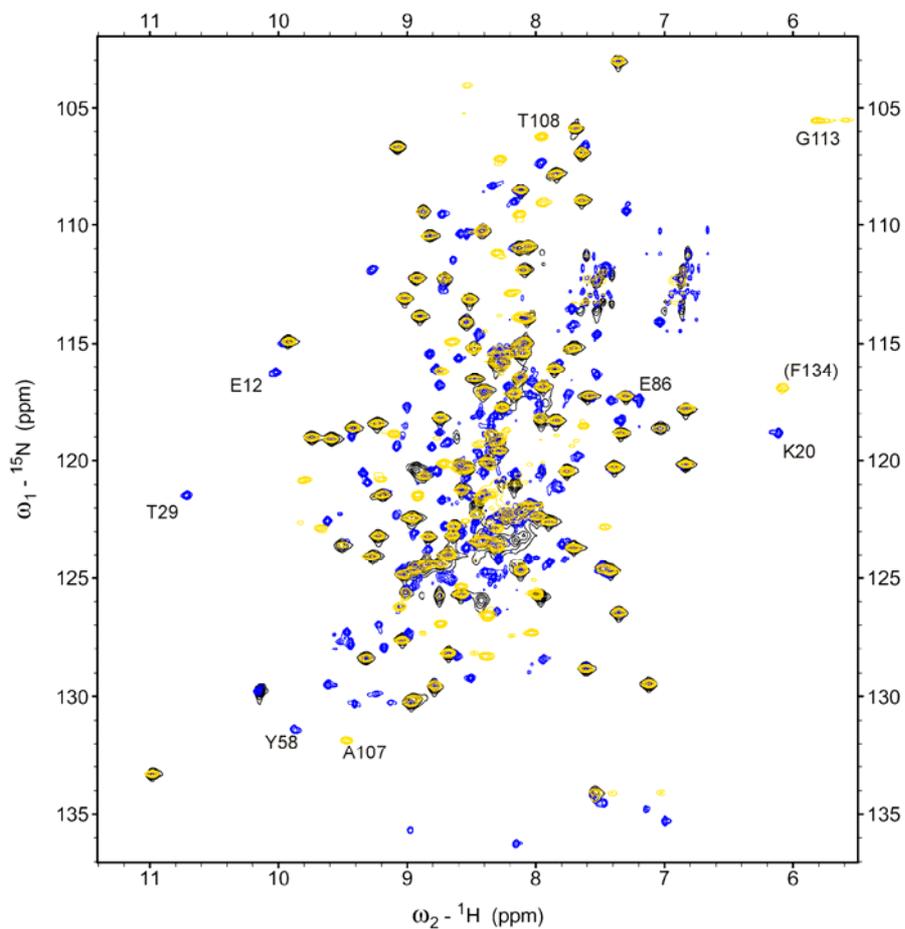
**Abbildung 73A Einfluss der Induktionstemperatur auf das Verhältnis von löslichen Aggregaten zur monomeren Spezies im Falle der N-terminalen Hälfte ST-gpD-Int<sup>N</sup><sub>CIA</sub> (I<sub>CIA</sub>).** Die Gel-filtrationschromatogramme der Expression bei 28°C (schwarz), bei 20°C (rot) und bei 34°C (grün) sind gezeigt.



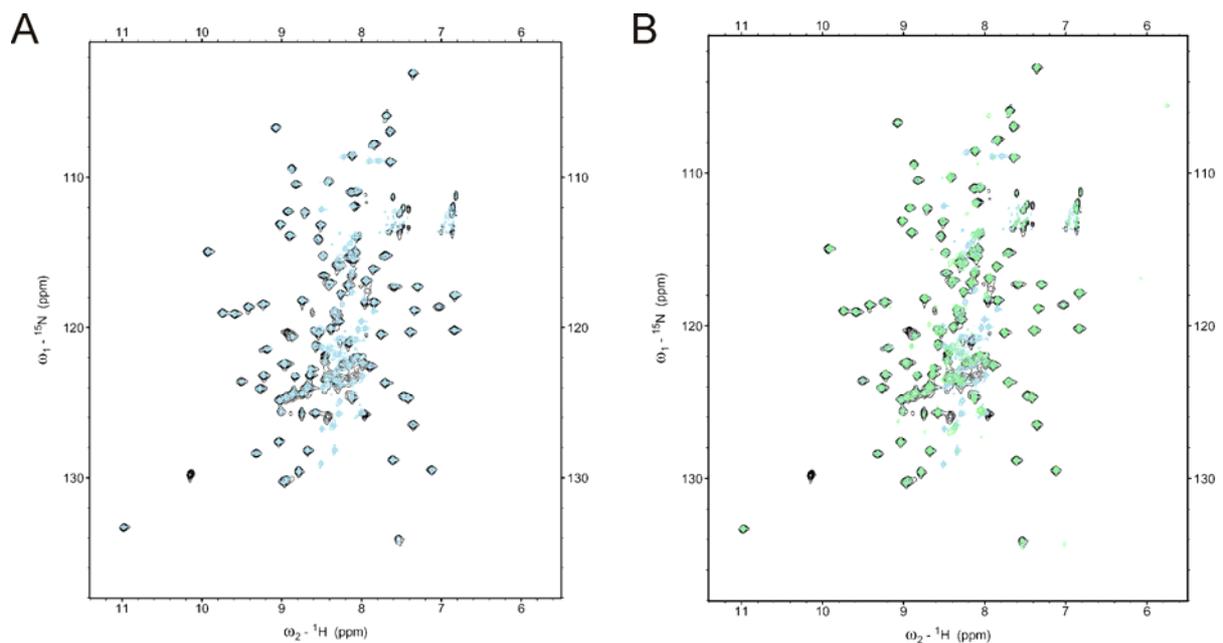
**Abbildung 74A Gel-filtrationsläufe des N-terminalen *Npu* DnaE Intein-fusionsproteins ST-eGFP-Int<sup>N</sup>.** Gezeigt sind die Chromatogramme des löslichen, gereinigten Proteins nach der Expression (schwarz) und des Proteins nach Denaturierung (30 min bei 37°C in Gegenwart von 8M Harnstoff) und anschließender Rückfaltung in Spleißpuffer (grün); diese Experimente zeigen, dass eine Umwandlung der löslichen Aggregate in die monomere, aktive Spezies prinzipiell möglich ist.



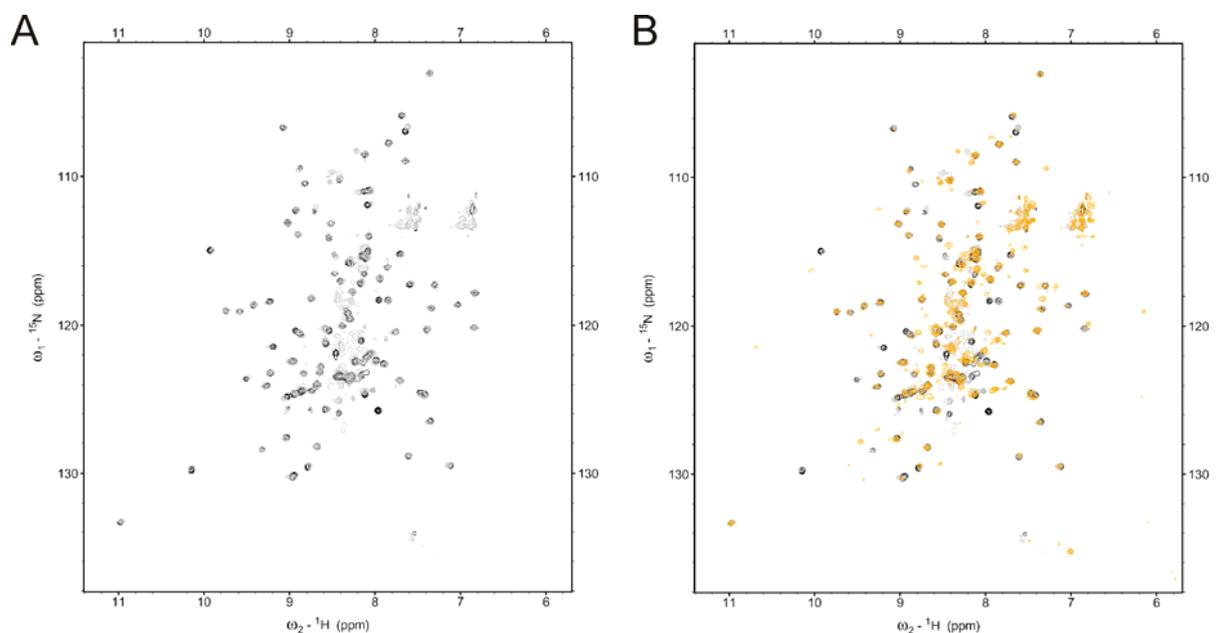
**Abbildung 75A** Gelfiltrationschromatogramme von N- und C-terminalen Fusionsproteinen des *Ssp* DnaB Inteins [163]. A) MBP-*Ssp* DnaB<sup>N</sup>-FKBP-H<sub>6</sub>; B) MBP-*Ssp* DnaB<sup>C</sup>-H<sub>6</sub>; nur die N-terminale Hälfte zeigt einen geringen Anteil an löslichen Aggregaten.



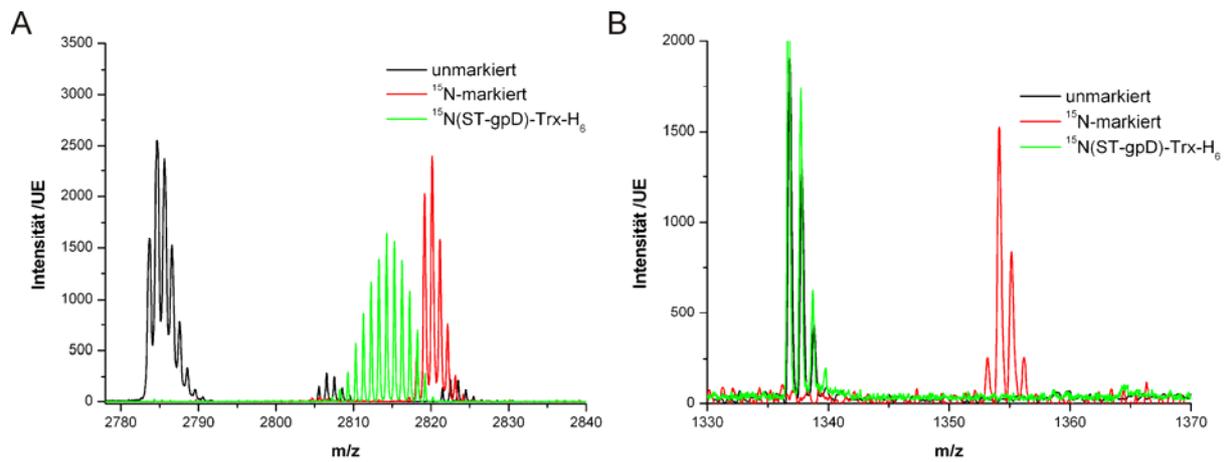
**Abbildung 76A** Zuordnung der Inteinsignale im Komplex nach Spleißen durch Vergleich mit literaturbekanntem Daten [240]. Gezeigt sind die  ${}^{15}\text{N}$ ,  ${}^1\text{H}$ -TROSY-Spektren der N-terminalen (ST-gpD-*Npu* DnaE<sup>N</sup> (1), (A), blau), sowie der C-terminalen Inteinhälfte (*Npu* DnaE<sup>C</sup>-gpD-H<sub>6</sub> (2), (B), gelb) nach der vollständigen PTS-Reaktion mit dem unmarkierten Partnerprotein 2, respektive 1; zum Vergleich ist das  ${}^{15}\text{N}$ ,  ${}^1\text{H}$ -TROSY-Spektrum des Referenzproteins ST-gpD-H<sub>6</sub> (5) in schwarz dargestellt; nur ein Bruchteil der Amid-Signale (8 von 134) konnte einwandfrei zugeordnet werden; die Nummerierung der Aminosäuren geht von einem fusionierten *cis*-Intein inklusive Startmethionin der C-terminalen Hälfte aus.



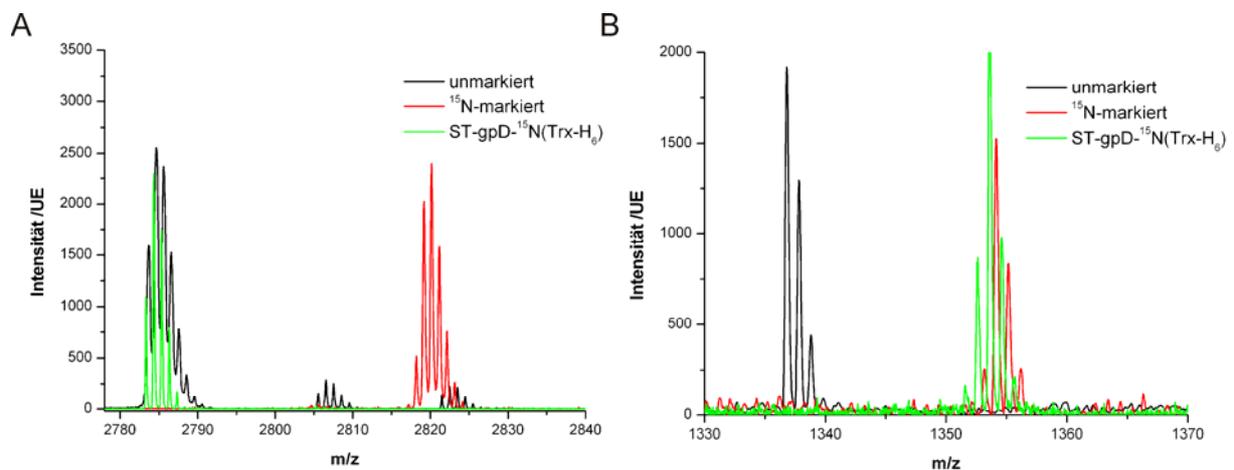
**Abbildung 77A NMR-Untersuchungen einer mutierten C-terminalen Inteinhälfte des *Npu* DnaE Inteins.**  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektren der C-terminalen Inteinhälfte ( $\text{Int}_{\text{N36A}}^{\text{C}}$ -gpD- $\text{H}_6$  ( $2_{\text{N36A}}$ ) vor (A, hellblaues Spektrum) und nach der Inkubation mit der unmarkierten N-terminalen Inteinhälfte (ST-gpD- $\text{Int}^{\text{N}}$  (**1**)) (B, hellgrünes Spektrum); zum Vergleich ist jeweils das  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektrum des Referenzproteins ST-gpD- $\text{H}_6$  (**5**) in schwarz gezeigt; mit diesen Proteinen ist die Reaktionsabfolge des PTS bis zur Bildung des Thioesterkomplexes möglich - die Asparaginyklisierung ist blockiert.



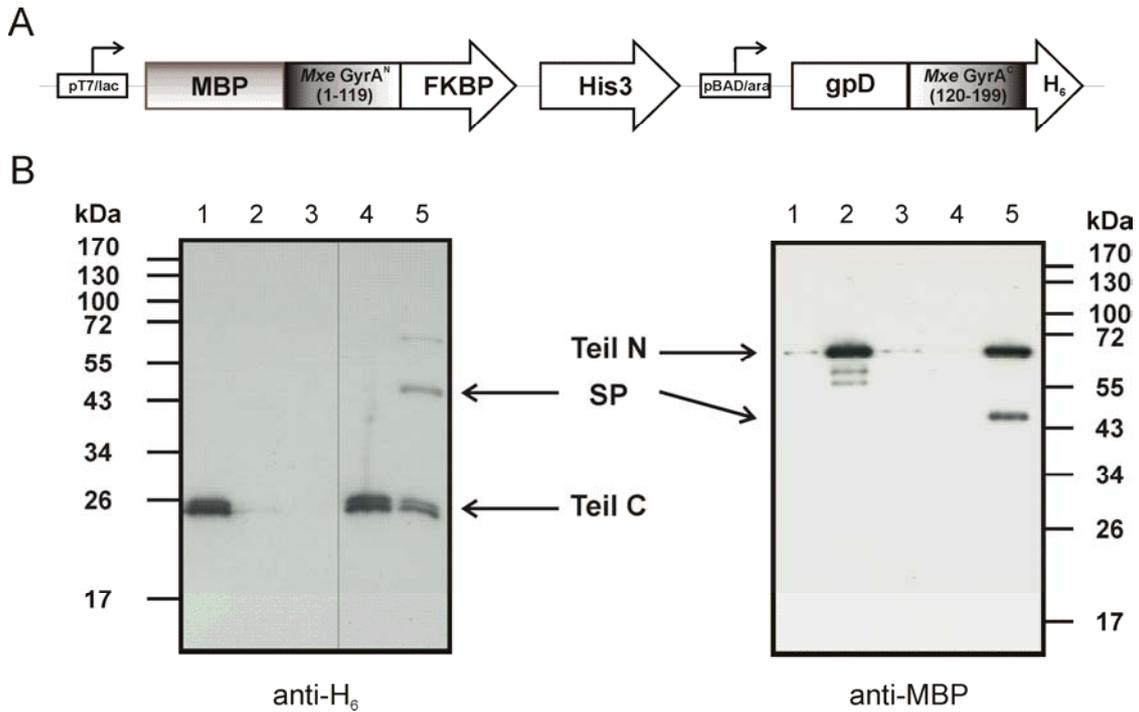
**Abbildung 78A NMR-Untersuchungen einer mutierten N-terminalen Inteinhälfte des *Npu* DnaE Inteins.**  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektren der N-terminalen Inteinhälfte ( $\text{H}_6$ -gpD- $\text{Int}_{\text{C1A}}^{\text{N}}$  (**8** $_{\text{C1A}}$ )) vor (A, graues Spektrum) und nach der Inkubation mit der unmarkierten C-terminalen Inteinhälfte ( $\text{Int}_{\text{N36A,C+1A}}^{\text{C}}$ -gpD- $\text{H}_6$  ( $2_{\text{N36A,C+1A}}$ )) (B, oranges Spektrum); zum Vergleich ist jeweils das  $^{15}\text{N}$ ,  $^1\text{H}$ -TROSY-Spektrum des Referenzproteins ST-gpD- $\text{H}_6$  (**5**) in schwarz gezeigt; mit diesen Proteinen ist die Reaktionsabfolge des PTS nur bis zur Bildung des ersten Anlagerungskomplexes möglich - alle weiteren Schritte sind blockiert.



**Abbildung 79A Überprüfung der  $^{15}\text{N}$ -Isotopenanreicherung des über *in vivo* PTS mit dem *Ssp* DnaB Intein hergestellte Spleißprodukts  $^{15}\text{N}(\text{ST-gpD})\text{-Trx-H}_6$ .** Gezeigt sind Ausschnitte aus den MALDI-TOF MS Spektrum des tryptischen in-Gelverdaus des Proteins (grünes Spektrum, A) Fragment 15-41 im N-terminalen Teil, B) Fragment 230-242 im C-terminalen Teil); zum Vergleich sind jeweils die Spektren des unmarkierten (schwarz) und des vollständig  $^{15}\text{N}$ -markierten (rot) Referenzproteins gezeigt; die Anreicherung der  $^{15}\text{N}$ -Isotope betrug 83% im N-terminalen Teil und unter 1% im C-terminalen Teil.



**Abbildung 80A Überprüfung der  $^{15}\text{N}$ -Isotopenanreicherung des über *in vitro* PTS mit dem *Npu* DnaE Intein hergestellte Spleißprodukts  $\text{ST-gpD-}^{15}\text{N}(\text{Trx-H}_6)$ .** Gezeigt sind Ausschnitte aus den MALDI-TOF MS Spektrum des tryptischen in-Gelverdaus des Proteins (grünes Spektrum, A) Fragment 15-41 im N-terminalen Teil, B) Fragment 230-242 im C-terminalen Teil); zum Vergleich sind jeweils die Spektren des unmarkierten (schwarz) und des vollständig  $^{15}\text{N}$ -markierten (rot) Referenzproteins gezeigt; die Anreicherung der  $^{15}\text{N}$ -Isotope betrug unter 1% im N-terminalen Teil und 95% im C-terminalen Teil.



**Abbildung 81A Die *Mxe GyrA* Inteinkassette.** A) Schematischer Aufbau der Kassette; B) Western Blot Analyse von Testexpressionen der Kassette zusammen mit dem Helferplasmid (LacI, AraC); Spur 1 = Arabinose Induktion, 37°C, 4h; Spur 2 = IPTG Induktion, 37°C, 4h; Spur 3 = vor Induktion; Spur 4 = Arabinose Induktion, 37°C, 2h; Spur 5 = Probe aus 4 und anschließende IPTG Induktion bei 25°C, 4h; die theoretischen Molekulargewichte der Proteine sind wie folgt: **SP** (44.7 kDa); **Teil N** (69.3 kDa); **Teil C** (22.2 kDa).

---

## Danksagung

“Bücher wie diese sind eigentlich gar nicht zum Lesen da. Sie werden in erster Linie geschrieben.“

T. Pratchett, *Einfach göttlich*, S. 183

Mir ist durchaus bewusst, dass das oben stehende Zitat sehr wahrscheinlich auch auf diese Doktorarbeit zutrifft. Jedoch findet in den meisten Fällen zumindest die Danksagung die ungeteilte Aufmerksamkeit der Leser. Nun, hier ist sie!

An erster Stelle gilt es meinem Doktorvater Prof. Dr. Henning Mootz zu danken. Er ermöglichte mir die Durchführung dieser Doktorarbeit in seinem Arbeitskreis. Neben seinem Interesse an dieser Arbeit, danke ich ihm für die gewährte Unterstützung in der gesamten Zeit. Meine wissenschaftliche Denk- und Arbeitsweise sind durch die Zusammenarbeit stark von ihm beeinflusst worden.

Herrn Prof. Martin Engelhard danke ich recht herzlich für die freundliche Übernahme des Zweitgutachtens.

Ein großer Dank gebührt zusätzlich all denjenigen, die sowohl praktisch, als auch theoretisch, an dieser Arbeit mitgewirkt haben. Diese Zusammenarbeiten haben mir in der Zeit der Doktorarbeit immer am meisten Spaß bereitet. Simone Eppmann war eine sehr große Hilfe bei Klonierungen, Proteinreinigungen und Assays. Viele Gele und Blots dieser Arbeit stammen von ihr. Die Idee der SPLICEFINDER-Methode entstand bei einer Heimfahrt nach Frankfurt mit Tim Sonntag und ich möchte mich bei ihm für die spannende und abwechslungsreiche Zusammenarbeit an diesem Projekt bedanken. Mit Vivien Schütz gelang die sehr arbeitsintensive, kinetische Charakterisierung des *Npu* DnaE Inteins. Ein weiterer großer Dank gebührt den Bachelor- und Masterstudenten, die ich während meiner Zeit betreuen durfte. Ihr großer Einsatz hat diese Arbeit entscheidend vorangebracht, auch wenn aus Platzgründen nicht alle Ergebnisse aufgenommen werden konnten. In chronologischer Reihenfolge sind Nadine Borg, Jens Binschik, Vivien Schütz, Annika Urbanek, Julian Matern und Michael Albers zu nennen. Allen anderen ehemaligen und aktuellen Mitgliedern des AK Mootz danke ich natürlich für eine angenehme und lustige Arbeitsatmosphäre. Helena Janzen sei für ihre Hilfe bei den statistischen Auswertungen der Gelfiltrationsläufe, sowie Alena Busche und Dr. Frank Löhr für die Aufnahme der NMR-Spektren gedankt.

Für das kritische Korrekturlesen dieser Arbeit danke ich Dr. Steffen Brenzel, Tim Sonntag, Jonas Alfermann und Michael Stockmar.

Bei Prof. Dr. Volker Dötsch bedanke ich mich für die Gelegenheit eine Woche in seiner Arbeitsgruppe verbringen zu können. Dr. Frank Löhr hat mir während dieses Aufenthalts versucht die Geheimnisse der NMR-Spektroskopie näher zu bringen.

Mein persönlicher und größter Dank geht an meine Familie. Die Unterstützung meiner Eltern und meines Bruders Arnd hat mich während der gesamten Zeit des Studiums und der Promotion immer begleitet und gestärkt. Meiner Familie, insbesondere meiner Mutter, ist diese Arbeit gewidmet.

---

## Eidesstattliche Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel

*„Wechselwirkungen zwischen Adenylierungs- und Peptidyl Carrier Protein-Domänen in nicht-ribosomalen Peptidsynthetasen sowie biochemische und strukturelle Untersuchungen zu gespaltenen Inteinen“*

selbständig, ohne unerlaubte Hilfe angefertigt und mich keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

---

(Ort, Datum)

---

(Unterschrift mit Vor- und Zuname)

# Lebenslauf

## PERSÖNLICHE DATEN

---

Name	Joachim Zettler
Geburtsdatum	19.12.1979
Geburtsort	Offenbach am Main
Familienstand	ledig
Staatsangehörigkeit	deutsch

## SCHULE

---

1986 - `90	Trinkbornschule, Ober-Roden (Grundschule)
1991 - `96	Gymnasialer Zweig der Heinrich-Mann-Schule, Dietzenbach
1997 - `99	Oberstufe mit Mathematik- und Physik-LK
Juni 1999	Abitur mit Ø 1,2

## ZIVILDIENTST

---

12-monatiger Zivildienst beim Arbeiter-Samariter-Bund Offenbach (Juli 1999 - Juli 2000)

## STUDIUM

---

2000- `02	Grundstudium der Chemie an der TU Darmstadt Vordiplom in Chemie Preis der Dr. Anton-Keller-Stiftung für Vordiplom
Okt. 2002 – Juni 2003	ERASMUS-Auslandsstudium an der University of Bristol mit Schwerpunkten in Biochemie und Chemie
2003 – `06	Studium der Chemie an der Philipps-Universität Marburg mit Abschluss als Diplomchemiker

## INDUSTRIEPRAKTIKA

---

6-wöchiges Praktikum bei der Firma Biotest in Dreieich im Bereich  
Forschung und Entwicklung (Aug. - Sep. 2003)  
7-wöchiges Praktikum bei der Firma Sanofi-Aventis (Frankfurt-Höchst) im  
Bereich Discovery Analytics (Feb. - Apr. 2005)

## PROMOTION

---

ab Okt. 2006 bis voraus. Jun. 2010	Promotion an der Technischen Universität Dortmund in der Fakultät Chemie im Bereich Chemische Biologie (Doktorvater Prof. Dr. H. Mootz)
------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------