Biosynthese der Labyrinthopeptine A1, A2 und A3, einer neuen Klasse von Lantibiotika aus *Actinomadura namibiensis*

vorgelegt von
Diplom-Biochemiker
Timo Schmiederer
aus Ludwigsburg

Von der Fakultät II – Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. rer. nat. Thomas Friedrich
Erster Berichter:	Prof. Dr. rer. nat. Roderich D. Süßmuth
Zweiter Berichter:	PD Dr. rer. nat. Ullrich Keller

Tag der wissenschaftlichen Aussprache: 18. April 2008

Berlin 2008

D 83

Zusammenfassung

Grundlage für die vorliegende Arbeit war der 1988 als *Actinomadura namibiensis* taxonomisch eingeordnete Mikroorganismus, der die drei peptidischen Sekundärmetabolite A1, A2 und A3 produziert. Die Struktur von A2 konnte mit Hilfe einer Röntgenkristallstrukturanalyse aufgeklärt werden, was die Einordnung in die Klasse der ribosomal synthetisierten Lantibiotika ermöglichte. Die im Folgenden als Labyrinthopeptin A2 bezeichnete Verbindung weist neben den typischerweise in Lantibiotika vorkommenden Lanthioninbrücken sowie einer Disulfidbrücke eine strukturelle Besonderheit auf, die bisher nicht literaturbekannt ist und als Labionin bezeichnet wurde. Dabei handelt es sich um eine Verbrückung, die zu quartär substituierten αC-Atomen führt und somit eine ungewöhnliche C-C-Bindungsknüpfung während der Biosynthese erfordert. In der vorliegenden Arbeit sollten diese Vorkenntnisse genutzt werden, um die Biosynthese von Labyrinthopeptin A2 (LabA2) und dabei insbesondere der Labionin-Verbrückung aufzuklären. Um dieses Ziel zu erreichen, sollte zunächst das entsprechende Biosynthese-Gencluster gesucht werden. Nachdem Bedingungen optimiert werden konnten, die die Isolierung hochmolekularer genomischer DNA aus *A. namibiensis* ermöglichten, wurde in Zusammenarbeit mit der Firma Agowa GmbH, Berlin, eine Cosmidbank erstellt. Die größte Herausforderung bestand nun darin, ausgehend von der 18 Aminosäuren lange Struktur von LabA2, von der nur zwölf als proteinogene Aminosäuren definiert waren, eine Sonde für ein Screening der Cosmidbank abzuleiten. Durch Kombination verschiedener PCR-Methoden gelang es schließlich, eine Sonde zu generieren, die im Cosmidbankscreening eingesetzt werden konnte. Nach der Sequenzierung eines positiv-getesteten Cosmids konnte mit Hilfe von verschiedenen Softwaretools das Biosynthese-Gencluster identifiziert und annotiert werden. Darin konnte u.a. ein Strukturgen identifiziert werden, das für die Präpropeptide der Labyrinthopeptine A1 und A3 kodiert. Basierend auf der hohen Homologie der Labyrinthopeptine untereinander konnte damit ein Strukturvorschlag für LabA1 und LabA3 erstellt werden. Die weitere Auswertung des Biosynthese-Genclusters zeigte neben den Strukturgenen noch zwei Gene für einen ABC-Transporter sowie ein Gen, das *C*-terminal eine Ser/Thr-Kinase- und *N*-terminal eine LanC-Domäne kodiert. Dabei handelt es sich vermutlich um das für die Labionin-Biosynthese zuständige Gen. Um diese Hypothese experimentell zu beweisen, wurden zwei unterschiedliche Strategien verfolgt. Die erste Strategie sah einen *in vitro*-Assay mit dem entsprechenden überexprimierten Protein und hypothetischen Substraten vor, während die zweite Strategie die heterologe Expression des gefundenen Biosynthese-Genclusters in *S. lividans* umfasste. In beiden Strategien konnten bedeutende Zwischenziele erreicht werden, die in zukünftigen Arbeiten die Bestätigung der Hypothese ermöglichen sollten.

für Diane, meine Familie und Sabine†

Danksagung

Ich danke Herrn Prof. Dr. Roderich Süßmuth für die ausgezeichneten Arbeitsbedingungen und das tolle Thema, das er mir zur Bearbeitung überlassen hat. Außerdem danke ich für die stetige Unterstützung und die freundschaftliche Atmosphäre.

Herzlichen Dank an Herrn PD Dr. Ullrich Keller für viele hilfreiche Gespräche und Diskussionen und für die Übernahme der zweiten Berichterstattung.

Vielen Dank an Herrn Prof. Dr. Thomas Friedrich für die Übernahme des Prüfungsvorsitzes.

Mein besonderer Dank gilt den Herren Dr. Berthold Fartmann, Dr. Stefan Pelzer und Dr. Tilmann Weber für wichtige Tipps und ehrliche Meinungen, wenn es mal wieder Probleme gab.

Herzlichen Dank auch an meine Kooperationspartner Dr. Marc Brönstrup, Dr. Klaus Hantke, Dr. Rainer Zocher und Dr. Peter Schmieder für die Zusammenarbeit.

Mein besonderer Dank geht an den gesamten Arbeitskreis Süßmuth für tolle Jahre und eine stets gute Arbeitsatmosphäre: Louise Aigrain, Dr. Christian Appelt, Mahsuni Atli, Diane Butz, Frank Dettner, Dr. Sven Feifel, Elvira Gottardi, Anne Hänchen, Soliman Helali, Maik Henkel, Dr. Simone Keller, Joanna Krawczyk, Wolfgang Müller, Jonny Nachtigall, Stefan Pohle, Marta Ramirez Hernández, Dr. Andreas Reicke, Pierre Loic Saaidi, Georg Sambeth, Heiko Schadt, Caroline Schlosser, Dr. Kathrin Schneider, Hanna von Suchodoletz, Dr. Suada Turkanović, Falko Wolter sowie meine Praktikanten Roland Kersten und Maria (Turbo) Woznik.

Dr. Kathrin Schneider und Falko Wolter danke ich für die schöne Zeit in Tübingen und Berlin sowie für ihre Freundschaft.

Ich danke dem VfB Stuttgart für die tolle Saison 2006/2007.

Unendlich dankbar bin ich meinen Eltern und meinem Bruder Dirk für die stetige Unterstützung und für ein immer offenes Ohr.

Diane danke ich für alles, was wir bis zum jetzigen Zeitpunkt zusammen erlebt haben und für die unendliche Unterstützung bei allem.

Inhaltsverzeichnis

1 Einleitung 1

1.1 Posttranslationale Modifikationen bei Lantibiotika 2

1.2 Aktivität von Lantibiotika 4

1.3 Biosynthese von Lantibiotika 6

1.3.1 Präpropeptide von Lantibiotika 7

1.3.2 Einführung der Lanthionin- und Methyllanthioninbrücken 8

1.3.3 Export der prozessierten Präpro- oder Propeptide und proteolytische Spaltung 12

1.4 Biosynthese-Gencluster von Lantibiotika 13

1.5 Einordnung der Lantibiotika in Gruppen 15

1.6 Die Entdeckungsgeschichte der Labyrinthopeptine 16

1.7 Die Labyrinthopeptine A1, A2 und A3 17

2 Zielsetzung 22

3 Materialien 23

3.1 Geräte 23

3.2 Chemikalien 24

3.3 Kits 26

3.3.1 DNA-Präparationen und Aufreinigung 26

3.3.2 Klonierung 26

3.3.3 Nichtradioaktive Markierung und Detektion von DNA-Sonden 26

3.3.4 Heterologe Expression 26

3.3.5 Mutagenese 26

3.3.6 Färben von Proteingelen 26

3.4 Puffer und andere Lösungen 27

3.4.1 SDS-Gelelektrophorese 27

3.4.2 Elektroporation 27

3.4.3 Protoplastentransformation 28

3.4.4 Agarosegelelektrophorese 28

3.4.5 *Southern-Blot* 28

3.4.6 Screening auf Hybridisierungsfilter der Cosmidbank 29

3.4.7 Isolierung genomischer DNA mit dem *NucleoSpin® Tissue-Kit* 29

3.4.8 Isolierung hochmolekularer genomischer DNA mittels Phenol/ Chloroform-Extraktion 30

3.4.9 SDS-PAGE nach Lämmli 30

3.4.10 SDS-PAGE nach Schägger und Jagow 30

3.4.11 Photometrische Proteinbestimmung nach Bradford 30

3.4.12 Gelfärbung mit Coomassie Brillant Blau 31

3.4.13 *Western-Blot* 31

3.4.14 Proteinreinigung mittels IMAC 31

3.4.15 Abspaltung des N-terminalen His-Tag mit der SUMO-Protease 32

3.4.16 Weitere Medienzusätze 32

3.5 Antibiotika 32

3.6 Enzyme 33

3.7 Bakterien 33

3.8 Nährmedien 34

3.9 Oligonukleotid-Primer 35

4 Methoden 39

4.1 Mikrobiologische Methoden 39
4.1.1 Lagerung und Kultivierung von *Actinomadura namibiensis* 39
4.1.2 Lagerung und Kultivierung von *Escherichia coli* 39
4.1.3 Lagerung und Kultivierung von *Streptomyces lividans* 40
4.1.4 Produktionskontrolle der Labyrinthopeptine A1/A2 aus Flüssigkulturen 40

4.2 Molekularbiologische Methoden 41
4.2.1 DNA-Transfer 41
4.2.2 Analyse, Isolierung und Reinigung von DNA 45
4.2.3 Polymerase-Kettenreaktion (PCR) 55
4.2.4 Klonierung 64

4.3 Bioinformatik 68
4.3.1 BLAST 68
4.3.2 Artemis 69
4.3.3 FramePlot 69

4.4 Proteinbiochemische Methoden 71
4.4.1 Photometrische Proteinbestimmung nach Bradford 71
4.4.2 Photometrische Proteinbestimmung durch Absorptionsmessung bei λ = 280 nm 72
4.4.3 Reduzierende diskontinuierliche SDS-PAGE (Disk-SDS-PAGE) nach Lämmli 73
4.4.4 Reduzierende diskontinuierliche SDS-PAGE (Disk-SDS-PAGE) nach Schägger und Jagow 75
4.4.5 Coomassie-Färbung von Proteingelen 76
4.4.6 Silberfärbung von Proteingelen 76
4.4.7 *Western-Blot* 77
4.4.8 Heterologe Expression mit dem *Champion pET SUMO-System* 78
4.4.9 Proteinreinigung mittels IMAC 80
4.4.10 Abspaltung des *N*-terminalen His-Tags mit Hilfe der SUMO-Protease und Aufreinigung der Produkte 81
4.4.11 GluC-Spaltung der Präpropeptide zur massenspektrometrischen Untersuchung und Charakterisierung 83
4.4.12 HPLC-ESI-MS und –MSMS von GluC-Verdauansätzen 84

5 Ergebnisse und Diskussion 85

5.1 Identifizierung des Biosynthese-Genclusters der Labyrinthopeptine 85
5.1.1 Isolierung genomischer DNA aus *Actinomadura namibiensis* 85
5.1.2 Herstellung von Cosmidbank und Hybridisierungsfilter 86
5.1.3 Erstellung eines *Codonusage* für die Gattung Actinomadura 88
5.1.4 Ableiten von DNA-Sonden und Screening von genomischen DNA-Verdaus sowie der Cosmidbank 91
5.1.5 Untersuchung der Integrität des Cosmids 1104 114

5.2 Auswertung und Annotation der Sequenzierung des Cosmids 1104 116
5.2.1 Auswertung des Sequenzabschnittes mit den Genen des Biosynthese-Genclusters für Labyrinthopeptin A2 117
5.2.2 Annotation der restlichen Gene von Cosmid 1104 140

5.3 Untersuchungen zur Biosynthese der Labyrinthopeptine 143
5.3.1 Generierung der Substrate für den *in vitro*-Assay 143
5.3.2 Expression des LabKC-Proteins für den *in vitro*-Assay 150
5.3.3 *In vitro*-Assay zur Untersuchung der katalytischen Aktivität des LabKC-Proteins 152

5.4 Heterologe Expression des Cosmids 1104 in *Streptomyces lividans* 153
5.4.1 Herstellung des Cosmids M1104 zur heterologen Expression des Labyrinthopeptin-Genclusters 154
5.4.2 Integration des Cosmids M1104 ins Genom von *S. lividans* und Untersuchungen zur heterologen Expression 159

6 Ausblick _____ *165*

7 DNA und Proteinsequenzen _____ *166*

7.1 Sequenz Cosmid1104 _____ **166**

8 Literaturverzeichnis _____ *186*

Abkürzungen

APS	Ammoniumperoxodisulfat
ASA	Aminosäurenanalytik
Da	Dalton
DAD	Dioden Array Detektor
DIG	Digoxigenin
Dha	Didehydroalanin
Dhb	Didehydrobutyrin
DNA	Desoxyribonucleinsäure
DMSO	Dimethylsulfoxid
EIC	*extracted ion chromatogram*
EMS	*enhanced MS*
EPI	*enhanced product ion scan*
ESI	Elektrospray-Ionisation
FT-ICR-MS	Fourier-Transform-Ionen-Zyklotron-Massenspektrometrie
g	Gramm
h	Stunde
HPLC	Hochleistungsflüssigkeitschromatographie
ICP-MS	*inductively-coupled-plasma MS*
IPTG	Isopropyl β-*D*-Thiogalaktopyranosid
kbp	Kilo-Basen-Paare
L	Liter
Lab	Labyrinthopeptin
Lan	Lanthionin
LC	Flüssigkeitschromatographie
m	Masse
m/z	Masse-zu-Ladung-Verhältnis
MeLan	Methyllanthionin

MeOH	Methanol
mg	Milligramm
min	Minuten
mL	Milliliter
mm	Millimeter
MS	Massenspektrometer / Massenspektrometerie
MSD	*membrane-spanning-domain*
NBD	*nucleotide-binding-domain*
nm	Nanometer
ORF	*open-reading frame*
PCR	Polymerase-Kettenreaktion (*polymerase chain reaction*)
PEG	Polyethylenglykol
RBS	Ribosomen-Bindungsstelle (*ribosome binding site*)
RP	Umkehrphase (*reverse phase*)
RT	Raumtemperatur
SDS	Natriumdodecylsulfat (*sodium dodecyl sulfate*)
SDS-PAGE	SDS-Polyacrylamid-Gelelektrophorese
SSP-PCR	*single-specific*-Primer-PCR
T	Temperatur
TEMED	*N,N,N´,N´*-Tetramethylethylendiamin
TIC	Totalionenchromatogramm
UV	ultraviolettes Licht
Vis	sichtbares Licht
X-Gal	5-Brom-4-chlor-3-indoxyl-β-*D*-galaktosid
λ	Wellenlänge

1 Einleitung

Viele Bakterien unterschiedlicher taxonomischer Zweige aus verschiedenen Habitaten produzieren ribosomal synthetisierte Peptidantibiotika, die posttranslational modifiziert werden und dadurch zu ihrer biologisch aktiven Form reifen. Aufgrund der großen Anzahl und Vielfalt der bisher gefundenen ribosomal synthetisierten Peptidantibiotika wurden diese zunächst gemäß ihrer Herkunft von Gram-positiven bzw. Gram-negativen Produzentenstämmen eingeteilt.[1] Ribosomal synthetisierte Peptidantibiotika aus Gram-positiven Bakterien werden demnach als Bacteriocine, die aus Gram-negativen Bakterien als Microcine bezeichnet. Die Bacteriocine werden wiederum in die drei Klassen I-III unterteilt. Unmodifizierte, thermosensible Proteine, z.B. Helveticin J, werden zu den Klasse III-Bacteriocinen zusammengefasst.[2] Die Klasse II-Bacteriocine sind thermostabile Peptide mit einer Länge von 37-58 Aminosäuren, die eine oder zwei Disulfidbrücken besitzen. Sie werden mit einer *N*-terminalen Leadersequenz als Präbacteriocine gebildet, wobei die Leadersequenz während des Exports aus der Zelle proteolytisch abgespalten wird und das reife Bacteriocin entsteht.[3] Die zu den Klasse I-Bacteriocinen zusammengefassten Peptide werden ebenfalls mit einer *N*-terminalen Leadersequenz zunächst als Präpropeptid synthetisiert und unterliegen anschließend mehr oder weniger starken posttranslationalen Modifikationen. Alle Mitglieder der Klasse I-Bacteriocine haben die Gemeinsamkeit, dass sie die ungewöhnlichen Strukturmotive Lanthionin (Lan) bzw. Methyllanthionin (MeLan) beinhalten.[1] Diese charakteristischen Strukturmerkmale führten schließlich auch zur Namensgebung dieser Klasse, Lantibiotika bzw. *lantibiotics* (aus dem *engl.: lanthionine-containing antibiotic peptides*).[4] Der Wortstamm geht auf die Entdeckungsgeschichte der Lanthionine (Lan) zurück, da nach der Behandlung von Wolle (*latein.: lana = Wolle*) mit Natriumcarbonat das Auftreten dieser ungewöhnlichen Aminosäuren-Seitenkettenverbrückung zum ersten Mal beobachtet werden konnte.[5] In den folgenden Abschnitten soll auf die Lantibiotika näher eingegangen werden.

1.1 Posttranslationale Modifikationen bei Lantibiotika

Bisher umfasst die Klasse der Lantibiotika über 50 verschiedene Peptide mit einer Länge zwischen 19 und 38 Aminosäuren.[6] Sie werden sowohl von GC-reichen als auch GC-armen Gram-positiven Bakterien gebildet, wobei die meisten Produzenten aus der Gattung Bacillus und Lactobacillus stammen. Das erste und bis heute am besten untersuchte Lantibiotikum ist Nisin (s. Abbildung 1 A)) aus *Lactobacillus lactis*, das seit über 40 Jahren in vielen Ländern zur Haltbarmachung von Lebensmitteln als Alternative zu chemischen Stoffen eingesetzt wird.[7] Ein anderer typischer Vertreter dieser Stoffklasse ist das ebenfalls gezeigte Cinnamycin (s. Abbildung 1 B)).

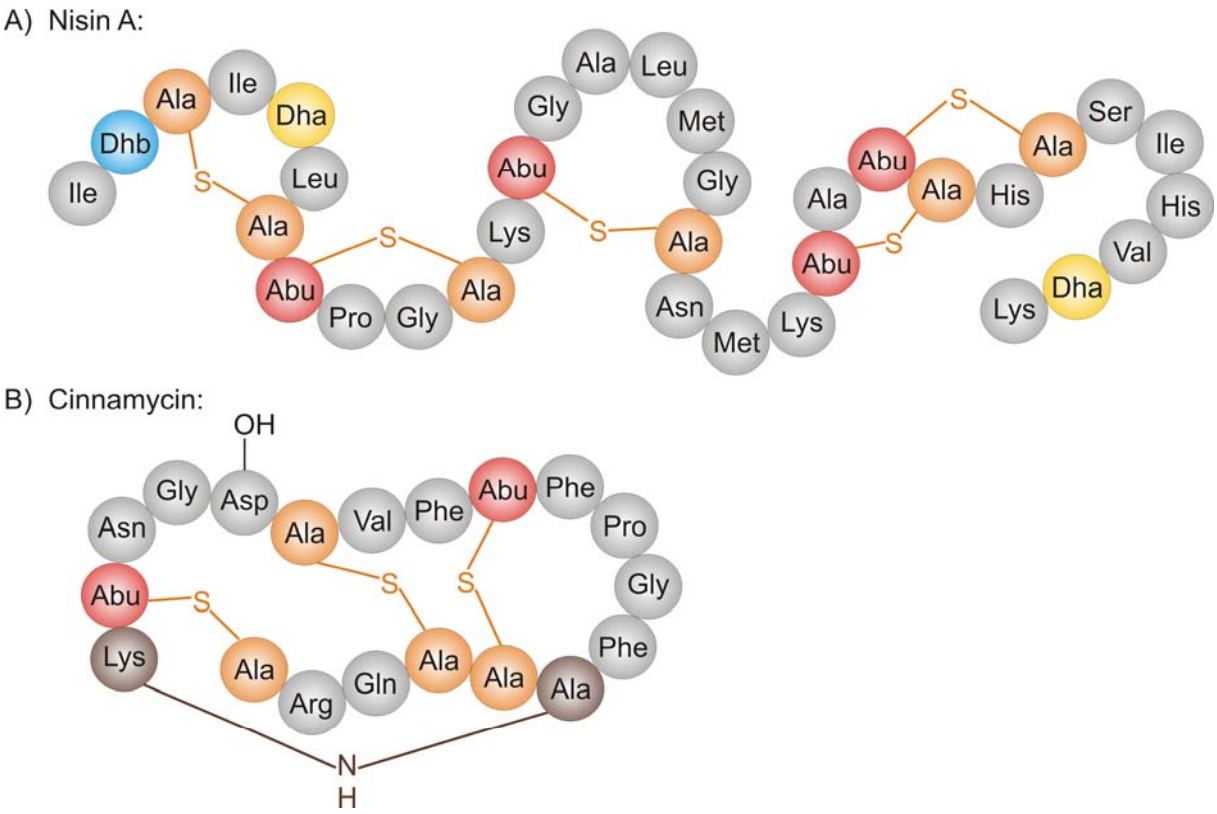


Abbildung 1: A) Kugelmodell des bisher am besten untersuchten Lantibiotikums Nisin. Es besteht aus einer fünffach verbrückten Sequenz aus 34 Aminosäuren (vier MeLan- und eine Lan-Brücke). B) Kugelmodell des Lantibiotikums Cinnamycin. Es besteht aus einer vierfach verbrückten Sequenz aus 19 Aminosäuren (zwei MeLan-, eine Lan- und eine Lysinoalaninbrücke). Dhb = 2,3-Didehydrobutyrin; Dha = 2,3-Didehydroalanin; Abu = 2-Aminobutyrat.

Wie diese beiden Beispiele exemplarisch zeigen, umfassen die Lantibiotika bzw. Klasse I-Bacteriocine ribosomal synthetisierte Peptide Gram-positiver Bakterien, die mindestens ein Lanthionin (Lan) oder Methyllanthionin (MeLan) als charakteristisches Strukturmotiv besitzen (s. Abbildung 2).

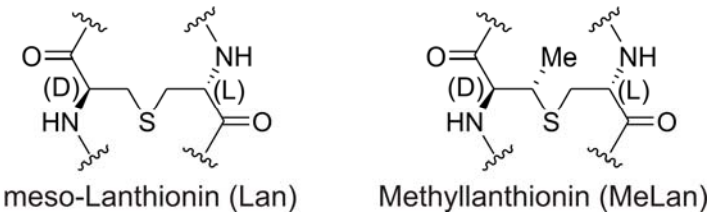


Abbildung 2: Darstellung der für Lantibiotika charakteristischen Strukturmotive Lanthionin (Lan) und Methyllanthionin (MeLan). Für die Stereochemie der bisher aufgeklärten Lan-Brücken wurde immer die meso-Form bestimmt.

Die Lan- bzw. MeLan-Strukturmotive werden posttranslational eingeführt, wobei es sich dabei formal um eine Thioetherbrücke zwischen zwei Ala-Resten (Lan) oder zwischen einem Ala und einer 2-Aminobuttersäure (MeLan) handelt. Bei allen bisher aufgeklärten und stereochemisch untersuchten Lan-Brücken konnte immer die meso-Form bestimmt werden (s. Abbildung 2). Neben diesen posttranslationalen Modifizierungen wurden in der Klasse der Lantibiotika eine Reihe weiterer Modifizierungen beobachtet (s. Abbildung 3).

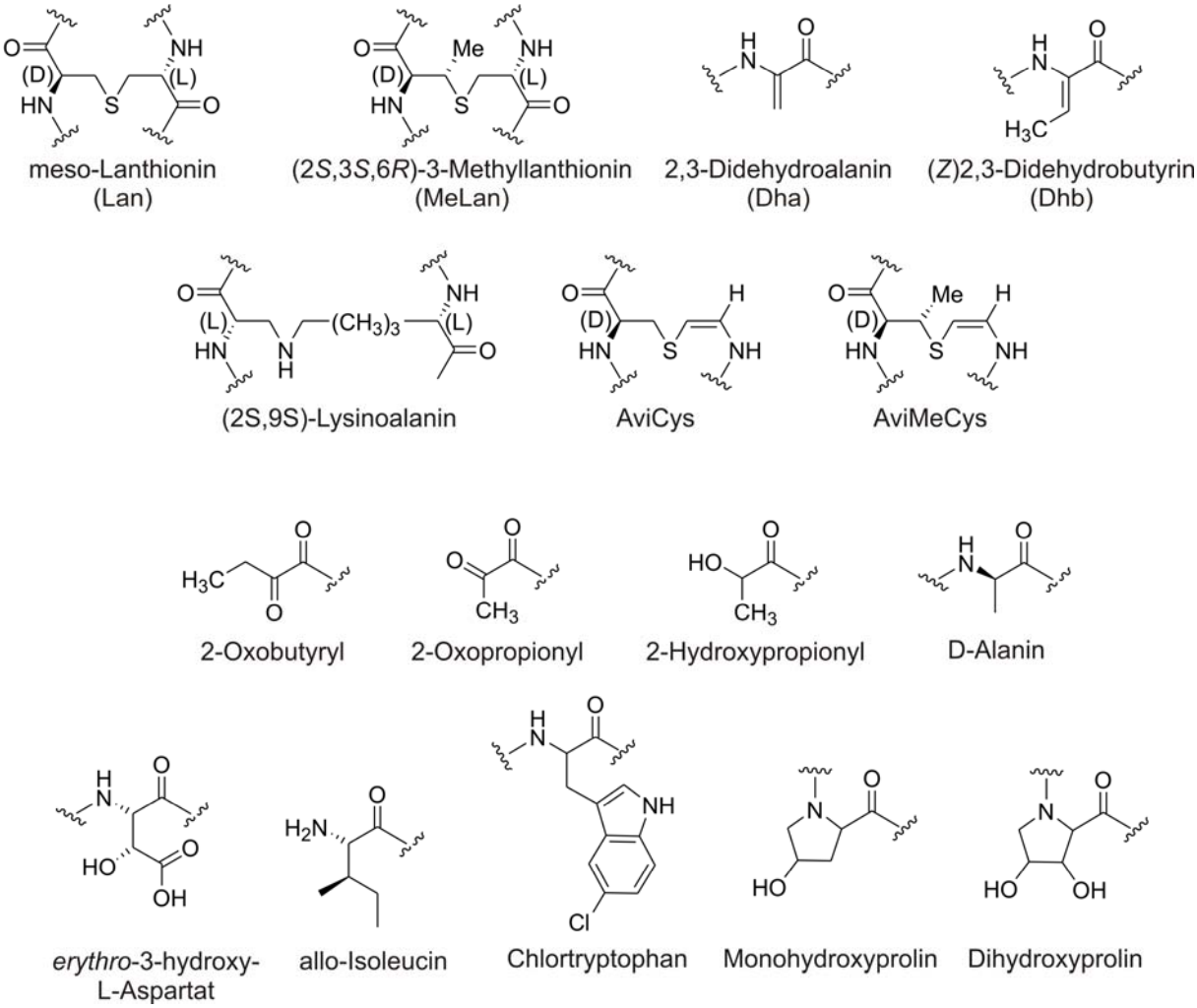


Abbildung 3: Übersicht über bisher bekannte posttranslationale Modifizierungen in Lantibiotika.

Dazu gehören z.B. die biosynthetischen Vorläufer der Lan- und MeLan-Strukturmotive, nämlich Didehydroalanin (Dha) und Didehydrobutyrin (Dhb). Liegen diese am *N*-Terminus vor, sind sie zu 2-Oxopropionyl- und 2-Oxobutyrylresten hydrolysiert,[8] wobei wie in Epilancin 15X der 2-Oxopropionylrest auch zu 2-Hydroxypropionyl weiter reduziert werden kann.[9] Außer Lan und MeLan gibt es mit Lysinoalanin, *S*-Aminovinyl-*D*-Cystein (AviCys) und *S*-Aminovinyl-3-Methyl-*D*-Cystein (AviMeCys) noch weitere Strukturmotive, die die Seitenketten zweier Aminosäuren verknüpfen. Die in AviCys und AviMeCys vorkommenden ursprünglichen Cys-Reste liegen dabei *C*-terminal. Die posttranslationale Einführung eines *D*-Ala ausgehend von *L*-Ser über ein Dha-Intermediat wurde für das Lantibiotikum Lacticin 3147 beschrieben.[10,11] Neben *erythro*-3-Hydroxy-*L*-Aspartat und *allo*-Isoleucin wurden kürzlich in Microbisporicin noch drei weitere für Lantibiotika bisher unbekannte posttranslationale Modifikationen identifiziert, nämlich Chlortryptophan, Monohydroxy- und Dihydroxy-Prolin.[12]

Der Gehalt an modifizierten Aminosäuren variiert je nach Lantibiotikum zwischen 24 % (Lactocin S) und 58 % in den kürzlich entdeckten Microbisporicinen.[12] Die mehr oder weniger vorhandenen posttranslationalen Modifikationen der Lantibiotika führen zu einer erhöhten Stabilität gegenüber hohen Temperaturen,[13] oxidierenden und sauren Bedingungen,[14] sowie zu einer Resistenz gegenüber Proteaseaktivitäten.[15]

1.2 Aktivität von Lantibiotika

Die meisten von Gram-positiven Bakterien gebildeten Lantibiotika zeigen eine antibiotische Aktivität gegen nahe verwandte Gram-positive Bakterien. Dazu gehören auch humanpathogene Erreger wie Methicillin-resistente *S. aureus* (MRSA) oder Vancomycin-resistente Enterokokken (VRE).[16] Trotzdem ist bisher nur Nisin (s. Abbildung 1A)) zur Haltbarmachung von Lebensmitteln in der Anwendung. Es befinden sich inzwischen jedoch weitere Lantibiotika in klinischen Studien, wie z.B. das Mutacin 1140, das eine antibiotische Aktivität gegen Karies-hervorrufende Streptokokken zeigt.[17] Bisher konnten verschiedene Wirkmechanismen für die antibiotische Aktivität von Lantibiotika identifiziert werden, wobei die Zellwand dabei einen zentralen Angriffsort darstellt.[18,23] So konnte z.B. für Nisin und andere strukturverwandte Lantibiotika eine Bindung an Lipid II und eine damit einhergehende Porenbildung beobachtet werden (s. Abbildung 4).[19] Für die Bindung an Lipid II ist die Ausbildung eines sog. Pyrophosphatkäfigs essentiell, der aus den ersten beiden *N*-terminalen Lan- und MeLan-Brücken in Nisin (s. Abbildung 1 A)) besteht. Dieser Pyrophosphatkäfig tritt über H-Brücken in Wechselwirkung mit dem Pyrophosphat von Lipid II (s. Abbildung 4 A)). Nach dieser Bindung von Nisin kommt es durch eine Konformationsänderung zur Einlagerung des *C*-terminalen Bereichs von Nisin in die

Lipiddoppelschicht. Durch die Ausbildung eines Komplexes aus acht Nisin- und vier Lipid II-Molekülen kommt es dann zur Porenbildung, was die Integrität der bakteriellen Zellwand zerstört (s. Abbildung 4 B)).[18] Viele strukturverwandte Lantibiotika zeigen einen zum Nisin sehr ähnlich aufgebauten *N*-Terminus, der den Pyrophosphatkäfig ausbilden kann und somit eine analoge Wirkungsweise vermuten lässt.[18]

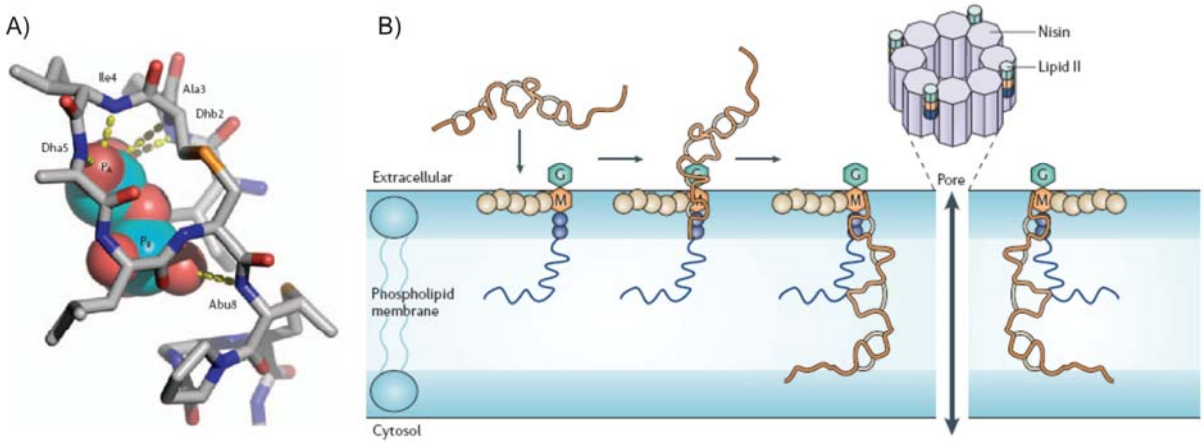


Abbildung 4: A) Hier sind die Wechselwirkungen zwischen dem durch die ersten beiden *N*-terminalen Lan- und MeLan-Brücken gebildeten Pyrophosphatkäfig von Nisin (s. Abbildung 1A)) und dem Pyrophosphat von Lipid II gezeigt. B) Nach der Bindung von Nisin an Lipid II kommt es zur Einlagerung des Nisins in die Cytoplasmamembran, wo sich ein Komplex aus acht Nisin- und vier Lipid II-Molekülen unter Porenbildung zusammenlagert.[18,19]

Neben der Ausbildung von Poren konnte zusätzlich eine Lipid II-abhängige antibiotische Aktivität gefunden werden, die auf einer Inhibition der Transglykosylierung und Transpeptidierung bei der Zellwandbiosynthese basiert.[18,20] Neben der Inhibition der Zellwandbiosynthese und Porenbildung in der Cytoplasmamembran wurden noch weitere Wirkmechanismen für Lantibiotika gefunden. So sind die Duramycine und das Cinnamycin potente Inhibitoren der Phospholipase A_2, indem sie deren natürliches Substrat Phosphatidylethanolamin binden.[21]

Die selektive Wirkung auf Gram-positive Bakterien wird der Größe der Lantibiotika zugeschrieben (1800-4600 Da), die eine Permeation durch die äußere Membran von Gram-negativen Bakterien verhindert, so dass diese ihre Targets nicht erreichen können.[20] Neben den bisher beschriebenen Lantibiotika mit antibakteriellem Wirkprofil wurden in letzter Zeit mit SapB und SapT auch Vertreter gefunden, die an der Induktion der Luftmycelbildung in Streptomyceten beteiligt sind.[22,41]

1.3 Biosynthese von Lantibiotika

Wie in Abschnitt 1.1 bereits dargestellt, unterliegen die ribosomal synthetisierten Lantibiotika vielfältigen posttranslationalen Modifikationen, bevor sie als prozessierte Verbindungen aus der Produzentenzelle freigesetzt werden. In Abbildung 5 ist der zeitliche Ablauf der Biosynthese eines Lantibiotikums schematisch dargestellt.

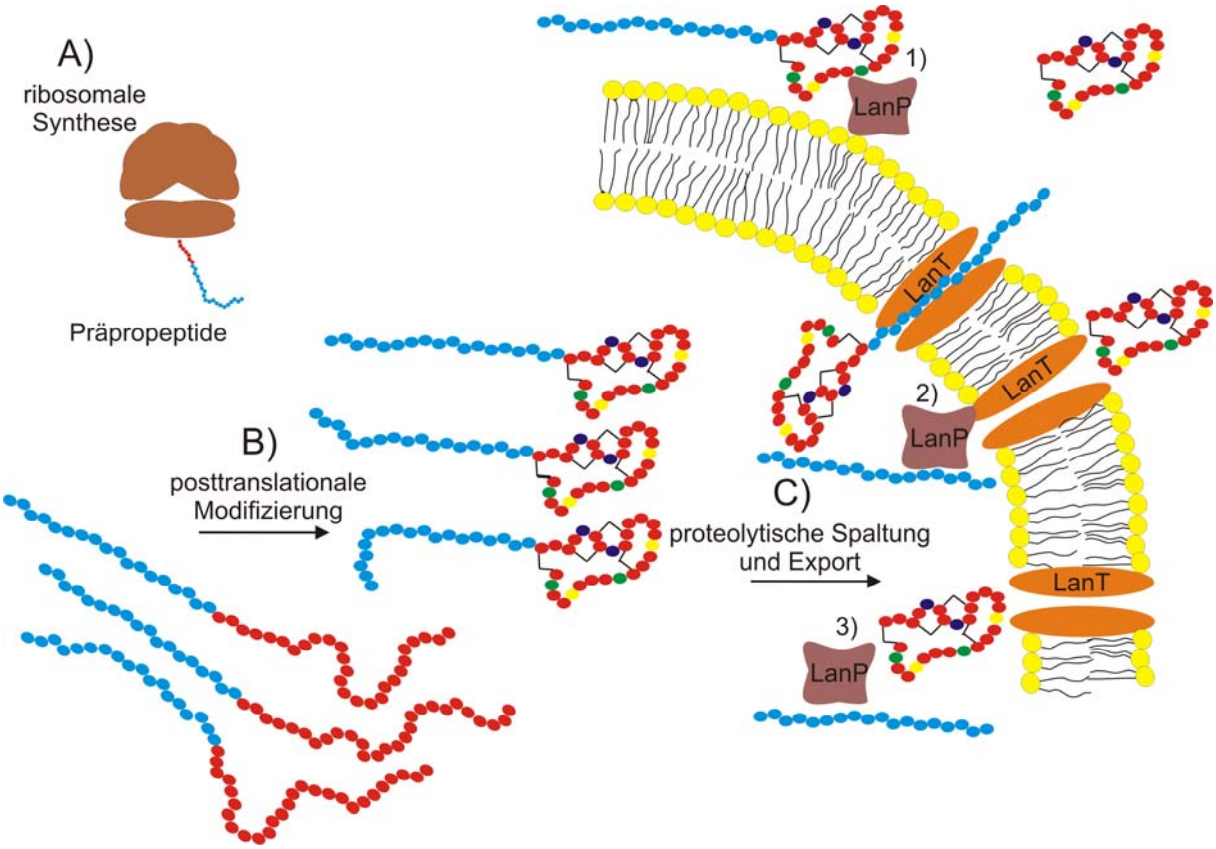


Abbildung 5: Schematischer Ablauf der Biosynthese eines Lantibiotikums. A) Bei der ribosomalen Synthese entsteht zunächst ein Präpropeptid mit einer *N*-terminalen Leadersequenz (blau) und dem *C*-terminalen strukturgebenden Propeptid (rot). B) Das Präpropeptid wird durch die entsprechenden Enzyme posttranslational modifiziert. C) Nach der Prozessierung erfolgt die proteolytische Abspaltung der *N*-terminalen Leadersequenz durch LanP unter Freisetzung des „reifen" Lantibiotikums. Die Proteasen LanP und die am Export beteiligten Transporter LanT unterscheiden sich von Fall zu Fall im Hinblick auf Lokalisation und Aufbau (1-3)).

Nach der ribosomalen Synthese des Präpropeptids (s. Abbildung 5 A)), bestehend aus *N*-terminaler Leadersequenz und *C*-terminalem Propeptid, erfolgen im Cytosol der Bakterienzelle die posttranslationalen Modifikationen (s. Abbildung 5 B)). Für die Einführung der charakteristischen Lan- bzw. MeLan-Verbrückungen sind entweder zwei Enzyme (LanB und LanC) oder ein Enzym (LanM) verantwortlich, worauf in Abschnitt 1.3.2 näher eingegangen werden soll. Neben diesen Lan- und MeLan-Verbrückungen können auf der Stufe des Präpropeptids auch andere, in Abschnitt 1.1 dargestellte Modifikationen in die Propeptide eingeführt werden. Nachdem das Propeptid fertig prozessiert wurde, erfolgt

schließlich die proteolytische Abspaltung der *N*-terminalen Leadersequenz und die Freisetzung aus der Produzentenzelle (s. Abbildung 5 C)). Für diese Schritte wurden drei unterschiedliche Wege beschrieben, die sich hauptsächlich in der Lokalisation der Protease (LanP) unterscheiden.[23] Bei dem in Abbildung 5 C-1)) dargestellten Weg liegt die Protease LanP teilweise in der Membran verankert[24] extrazellulär vor und spaltet nach dem Export die Leadersequenz unter Freisetzung des Lantibiotikums ab. Es wurden jedoch auch Lantibiotika-Exporttransporter identifiziert, die eine Protease-Domäne besitzen und die Leadersequenz während des Exports abspalten (s. Abbildung 5 C-2)).[23] In Abbildung 5 C-3) spaltet schließlich eine cytoplasmatisch lokalisierte Protease die Leadersequenz des prozessierten Lantibiotikums vor dem Export ab.[25] Alle drei Wege führen am Ende der Biosynthese zur Freisetzung des vollständig prozessierten und somit „reifen" Lantibiotikums aus der Zelle. In den folgenden Abschnitten soll jeweils kurz auf einige Aspekte der Biosynthese, wie Aufbau der Präpropeptide, Einführung der Lan- und MeLan-Verbrückungen sowie Export eingegangen werden.

1.3.1 Präpropeptide von Lantibiotika

Die Präpropeptide von Lantibiotika bestehen aus einem *C*-terminalen strukturgebenden Teil mit einer Länge von 20 bis 38 Aminosäuren, dem sog. Propeptid und einer *N*-terminalen Leadersequenz aus 23 bis 59 Aminosäuren. Während posttranslational unterschiedlichste Modifikationen (s. Abschnitt 1.1) in den Propeptidteil eingeführt werden können, bleibt der Teil der Leadersequenz unmodifiziert. Ser und Thr als Vorläufer für die Einführung von Lan und MeLan werden zwar auch in der Leadersequenz beobachtet, allerdings ist das Vorkommen von Cys auf das Propeptid begrenzt.[23]

Basierend auf Sequenzvergleichen verschiedener Leadersequenzen konnten zwei konservierte Sequenzmotive identifiziert werden, zum einen das „FNLD"-Motiv in Position -15 bis -20 mit einem Pro (-2) vor der Spaltstelle und zum anderen das „Doppelglycin"-Motiv (GG oder GA) mit mehreren Asp und Glu (s. Abbildung 6).[26] Verschiedene Experimente haben gezeigt, dass der Leadersequenz unterschiedliche Funktionen zugeordnet werden können. So konnte am Modellsystem Nisin (s. Abbildung 1 A)) der Beitrag der Leadersequenz zur Erkennung des Präpropeptids durch die posttranslational modifizierenden Enzyme nachgewiesen werden.[27,28,6] Daneben ist die Leadersequenz am Export des prozessierten Lantibiotikums aus der Zelle beteiligt,[6] sowie an der Inaktivierung des prozessierten Lantibiotikums.[24] Für Lantibiotikaproduzenten, bei denen die proteolytische Freisetzung erst extrazellulär stattfindet, ist dies eine sehr bedeutende Möglichkeit der Immunität bzw. des Schutzes gegen das eigene Produkt.

“FNLD”-Motiv:

```
Nisin A/Z                          MSTKDFNLDLVSVSKK--DSGASPR
Subtilin/Ericin S                  MSKFDDFDLDVVKVSKQ--DSKITPQ
Epidermin                        MEAVKEKNDLFNLDVKVNAKESNDSGAEPR
Gallidermin/Staphylococcin T     MEAVKEKNELFDLDVKVNAKESNDSGAEPR
Mutacin 1140/I                MSNTQLLEVLGTETFDVQEDLPAFDTTDTTIFASNDDPDTR
Pep5                             MKNNKNLFDLEIKKETSQNTD-ELEPQ
```

“Doppelglycin”-Motiv:

```
Lacticin 481                       MKEQNSPNLLQEVTESELDLILGA
Streptococcin A-FF22               MEKNNEVINSIQEVSLEELDQIIGA
Variacin                             MTNAFQALDEVTDAELDAILGG
Butyrivibriocin OR79                MNKELNALTNPIDEKELEQILGG
Mutacin II                        MNKLNSNAVVSLNEVSDSELDTILGG
Salivaricin A                     MKNSKDILNNAIEEVSEKELMEVAGG
Salivaricin A1                    MKNSKDILTNAIEEVSEKELMEVAGG
Sublancin 168                        MEKLFKEVKLEELENQKGS
```

Abbildung 6: Einteilung der Leadersequenzen in die zwei Gruppen „FNLD“-Motiv und „Doppelglycin“-Motiv.

1.3.2 Einführung der Lanthionin- und Methyllanthioninbrücken

Die Einteilung von bakteriellen Peptiden in die Gruppe der Lantibiotika basiert auf der Existenz von mindestens einer Lanthionin- oder Methyllanthioninbrücke.[1] Für die Einführung dieser charakteristischen Strukturmotive müssen in den Propeptiden die Aminosäuren Ser und Cys für Lan bzw. Thr und Cys für MeLan vorkommen. Der allgemeine Mechanismus für die Einführung dieser Seitenkettenverbrückungen ist in Abbildung 7 schematisch dargestellt.

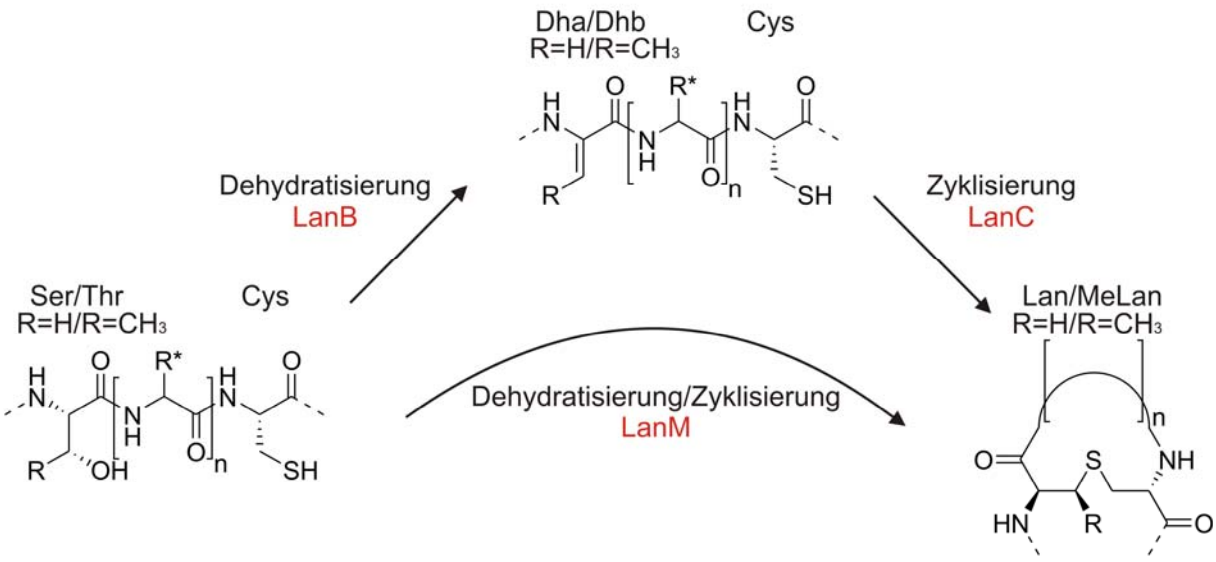


Abbildung 7: Zwei mögliche Wege zur posttranslationale Einführung einer Lan- bzw. MeLan-Brücke. Zunächst wird Ser bzw. Thr zu Dha bzw. Dhb dehydratisiert. Durch einen nukleophilen Angriff von Cys werden die Aminosäurenseitenketten schließlich zu Lan bzw. MeLan verbrückt.

Dieser Mechanismus sieht zunächst eine Dehydratisierung von Ser oder Thr zu Dha oder Dhb vor. Durch einen nukleophilen Angriff des als Thiolatanion aktivierten Cys auf das Michaelsystem von Dha oder Dhb erfolgt anschließend die Zyklisierung unter Ausbildung von Lan bzw. MeLan. Die bisherigen Untersuchungen zur Biosynthese von Lan und MeLan in Lantibiotika haben zwei mögliche Wege identifiziert, die sich in der Anzahl der beteiligten Proteine unterscheiden. So konnten die für die Modifikation notwendigen enzymatischen Aktivitäten im einen Fall auf zwei getrennten Proteinen (LanB/C) und im anderen Fall auf einem Protein (LanM) lokalisiert werden. Proteinalignments verschiedener LanB/C und LanM Sequenzen zeigten, dass LanC und der *C*-terminale Bereich von LanM eine 20-27%ige Homologie zueinander zeigen.[29] Die *N*-terminale Domäne von LanM zeigt hingegen keine Homologie zu LanB, weshalb auch eine evolutionäre Entstehung von *lanM* durch Genfusion von *lanB* und *lanC* ausgeschlossen werden kann.[29]

LanB-Proteine bestehen aus etwa 1000 Aminosäuren und zeigen untereinander eine geringe Sequenzidentität von nur 26-29 %, was mit der großen Substratdiversität in Verbindung gebracht wird.[23] Aufgrund einer schwachen Homologie zu IlvA, einer Threonin-Dehydratase von *E. coli*,[30] und aufgrund von Ergebnissen von LanB- und LanC-Geninaktivierungs-Mutanten[25] wird den LanB-Proteinen die Rolle der Dehydratase bei der Einführung der Lan- und MeLan-Brücken zugeschrieben. Die Ergebnisse aus Versuchen mit LanC-Geninaktivierungs-Mutanten zeigten auch, dass LanC Zyklaseaktivität besitzt und deshalb für den zweiten Schritt bei der Einführung von Lan oder MeLan verantwortlich ist. So konnten z.B. aus NisC-Geninaktivierungs-Mutanten des Nisin-Produzenten nichtzyklisierte Propeptide isoliert werden, die Dha- und Dhb-Reste an den entsprechenden Positionen der eigentlichen Lan- und MeLan-Brücken zeigten.[31] LanC-Proteine bestehen durchschnittlich aus ca. 400 Aminosäuren und zeigen untereinander eine Sequenzidentität von 24-32 %.[29] ICP-MS-(*inductively-coupled-plasma mass spectrometry*) Untersuchungen von LanC-Proteinisolaten zeigten einen stöchiometrischen Zink-Gehalt an,[32] was durch die Röntgenstrukturanalyse von NisC bestätigt werden konnte (s. Abbildung 8).[27] An der tetraedrischen Komplexierung des Zn^{2+} im aktiven Zentrum von NisC ist neben den Aminosäuren Cys^{284}, Cys^{330} und His^{331} (s. Abbildung 8 A)) auch ein Wassermolekül beteiligt. Neben diesen in LanC-Proteinen konservierten Aminosäureresten wird auch die Beteiligung der konservierten Aminosäuren His^{212} und Arg^{280} im Zuge eines Säure-Base-Mechanismus an der Zyklisierung angenommen (s. Abbildung 8 B)). Nach Bindung des dehydratisierten Substrats sieht der vorgeschlagene Mechanismus die Verdrängung des Wassers am Zn^{2+} durch die Seitengruppe des entsprechenden Cys unter gleichzeitiger Deprotonierung zum Thiolat vor. Dieses aktivierte Thiolatanion greift dann das β-Kohlenstoffatom von Dha oder Dhb unter Zyklisierung und Ausbildung eine Enolats an. Das Enolat wird dann protoniert, was zur *D*-Konfiguration am α-Kohlenstoffatom führt (s. Abbildung 8 C)).[27]

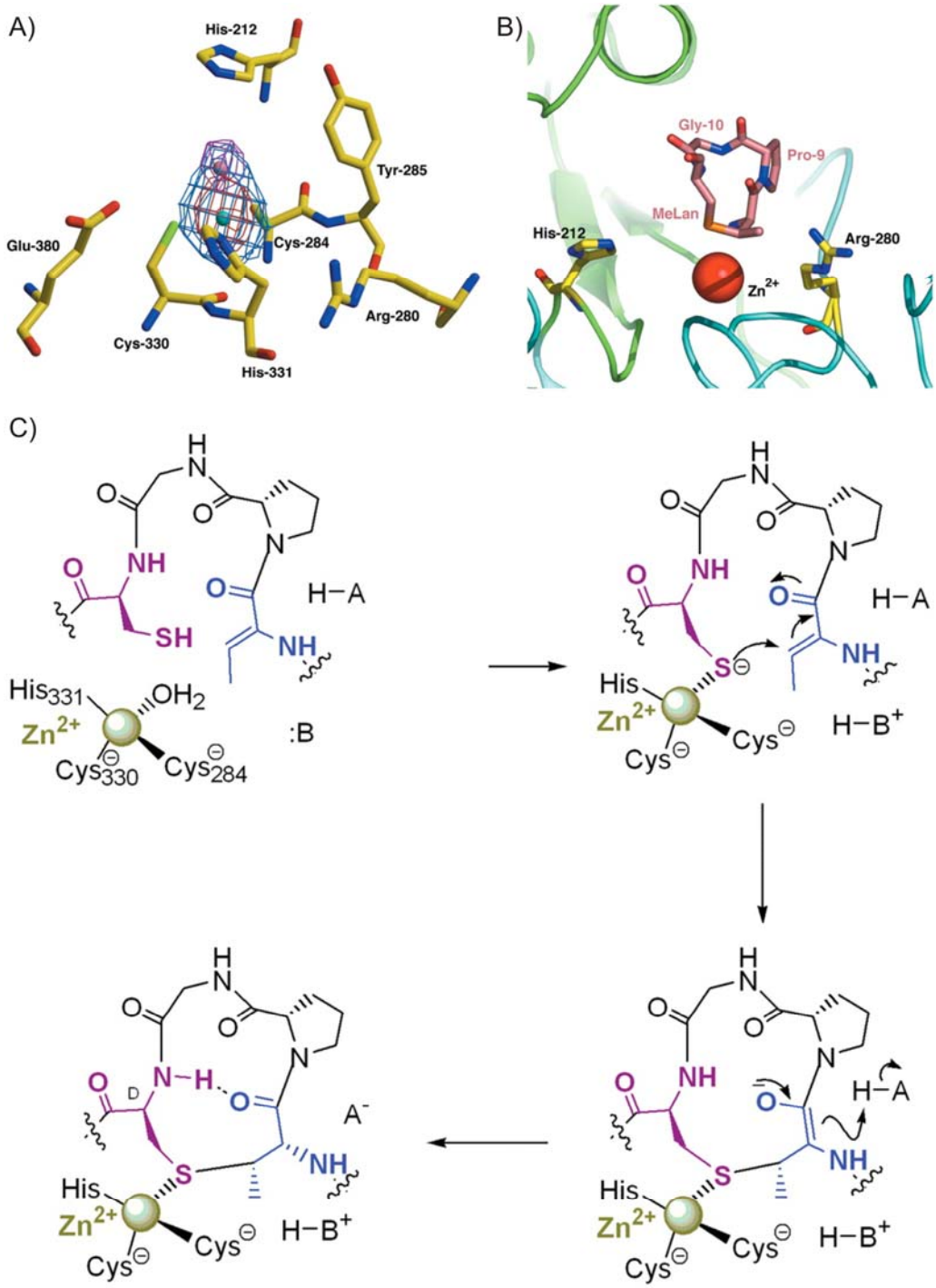


Abbildung 8: A) Gezeigt sind die an der Komplexierung von Zn^{2+} im aktiven Zentrum von NisC beteiligten Reste. Ein Wassermolekül (nicht dargestellt) und die Aminosäuren Cys^{284}, Cys^{330} und His^{331} koordinieren das Zn^{2+} tetraedrisch. B) Den beiden in LanC konservierten Aminosäuren His^{212} und Arg^{280} wird eine Beteiligung an der Zyklisierung durch einen Säure-Base-Mechanismus zugeschrieben. C) Hier ist der vorgeschlagene Mechanismus für die Zykllisierungsreaktion gezeigt.[27]

Wie bereits erwähnt wurde, zeigen Sequenzvergleiche zwischen LanC- und LanM nur eine relativ niedrige Sequenzidentität von 20-27 %. Es sind jedoch deutliche Homologien zu erkennen, vor allem was die konservierten Aminosäuren betrifft (AHG-, WCXG- und CHG-Motive), die für die Komplexierung von Zn^{2+} verantwortlich sind (s. Abbildung 9).

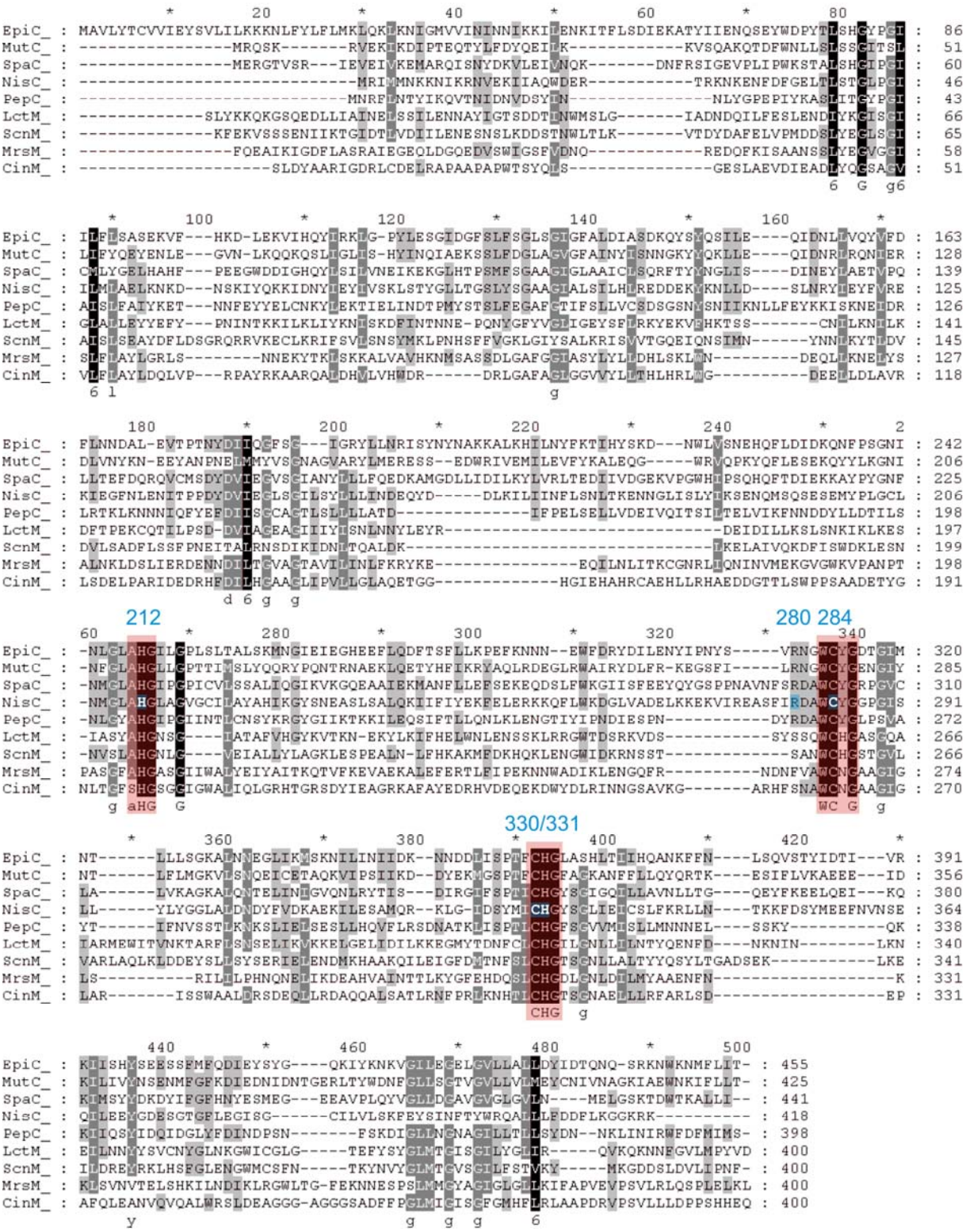


Abbildung 9: Proteinalignment von bekannten LanC-Sequenzen und den entsprechenden *C*-terminalen (400 Aminosäuren) Bereichen verschiedener LanM-Proteine. Die AHG-, WCXG- und CHG-Motive wurden rot, die für NisC bestimmten Reste des aktiven Zentrums blau markiert. Legende: EpiC = LanC für Epidermin und analog dazu: MutC (Mutacin I); SpaC (Subtilin); NisC (Nisin); PepC (Pep5). LactM = LanM von Lacticin481 und analog dazu: ScnM (StreptococcinA-FF22); MrsM (Mersacidin); CinM (Cinnamycin).

Der *N*-terminale Bereich der 900-1000 Aminosäuren großen LanM-Proteine zeigt aber, wie bereits erwähnt, keine Homologie zu LanB-Sequenzen. Trotzdem wurde dem *N*-terminalen Bereich von LanM die Dehydrataseaktivität analog zu LanB zugeschrieben, was durch *in vitro*-Studien am LctM, dem modifizierenden Enzym für Lacticin 481, bestätigt werden konnte.[90] Die *in vitro*-Studien an LctM haben auch gezeigt, dass dessen Aktivität die Anwesenheit von ATP und Mg^{2+} voraussetzt, wobei diese Beobachtung zunächst nicht näher gedeutet werden konnte.[90] Weitere Versuche mit LctM und verschiedenen Substratanaloga erbrachten dann, dass phosphorylierte Intermediate von Ser und Thr während der Dehydratation zu den entsprechenden Dha bzw. Dhb zu beobachten sind.[33]

Zusammenfassend kann gesagt werden, dass für die Einführung der für Lantibiotika charakteristischen Lan- bzw. MeLan-Brücken in das Präpropeptid zwei unterschiedliche Wege zur Verfügung stehen, wobei der Biosyntheseweg eines jeden Lantibiotikums genetisch durch die Enzymausstattung vorgeschrieben ist (s. Abschnitt 1.4). Beide Wege gehen von Dha- oder Dhb-Intermediaten aus, die dann in einer Zn-abhängigen Zyklisierung zu den Lan- oder MeLan-Resten verbrückt werden. Dafür ist LanC oder die *C*-terminale LanC-*like* Domäne von LanM verantwortlich. Die Dha- oder Dhb-Substrate für LanC werden von LanB auf einem bisher ungeklärten Mechanismus gebildet,[25] die Substrate für die *C*-terminale LanC-*like* Domäne von LanM werden dagegen von der *N*-terminalen Domäne von LanM nach einem aktivierenden Phosphorylierungsschritt selbst generiert.[33]

1.3.3 Export der prozessierten Präpro- oder Propeptide und proteolytische Spaltung

Mit Ausnahme des Cinnamycins werden die prozessierten Präpropeptide von Lantibiotika auf einem *sec*- und *TAT*- (twin-arginine) unabhängigen Weg aus der Produzentenzelle exportiert.[6] Für den Export sind Lantibiotika-Exporter (LanT) verantwortlich, die aufgrund einer ATP-Bindungsstelle in die Familie der ABC- (*ATP-binding-cassette*) Transporter eingeordnet werden können.[34] Die meisten bakteriellen ABC-Transporter bestehen aus einer intrazellulär lokalisierten *nukleotide binding domain* (NBD) und einer *membrane spanning domain* (MSD) mit üblicherweise sechs Transmembranhelices einer durchschnittlichen Länge von 20 Aminosäuren.[29] Diese beiden Domänen eines Transporters können entweder gemeinsam auf einem Protein oder separat auf zwei Proteinen liegen.[35] Die meisten LanT-Transporter von Lantibiotika tragen beide Domänen auf einem Protein, wobei MSD *N*-terminal und NBD *C*-terminal lokalisiert ist (s. Abbildung 10).[29] Während MSD für den Durchtritt der zu transportierenden Verbindung durch die Membran verantwortlich ist, wird von NBD die dafür notwendige Energie durch Hydrolyse von ATP bereitgestellt. In den NBDs wurden dafür fünf hochkonservierte Bereiche gefunden, die an der Bindung und Spaltung von ATP beteiligt sind und wie in Abbildung 10 gezeigt benannt werden.[34]

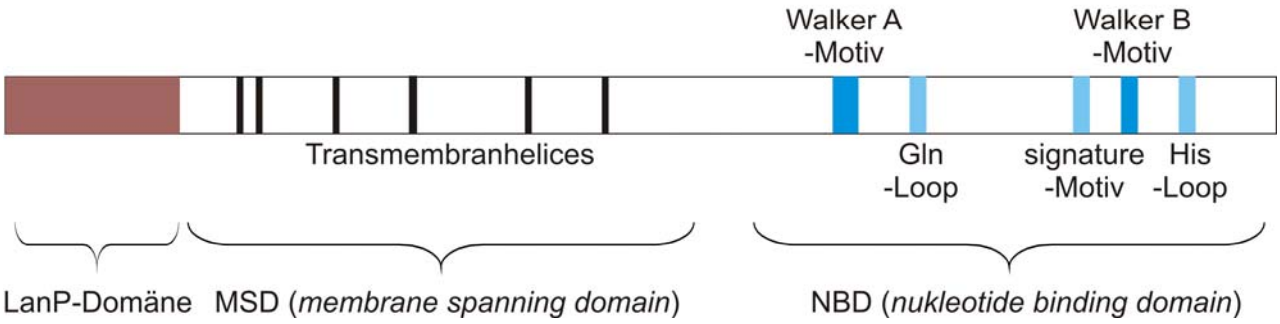


Abbildung 10: Domänenstruktur eines Lantibiotikatransporters (LanT). Die *N*-terminale, ca. 150 Aminosäuren lange Proteasedomäne LanP ist nicht für alle LanT Transporter zu beobachten. In die MSD sind die sechs Transmembranhelices durch schwarze Balken eingezeichnet. In der NBD wurden fünf hochkonservierte Bereiche identifiziert, die an der ATP-Bindung und -Spaltung beteiligt sind und hier blau markiert wurden.

Die Walker-Motive in den NBDs sind in allen ATP-bindenden Proteinen zu beobachten.[36] Im Gegensatz dazu sind die drei anderen, nämlich „Gln-Loop“, „signature-Motiv“ und „His-Loop“ spezifisch für ABC-Transporter.[37] Der in Abbildung 10 dargestellte schematische Aufbau eines LanT-Transporters zeigt außerdem eine *N*-terminale LanP-Domäne, die allerdings nicht für alle LanT-Proteine zu beobachten ist. Diese Proteasen, die eine Domäne des Exporttransporters darstellen, zählen aufgrund ihres katalytischen Cys zu der Klasse der Cysteinproteasen und spalten Leaderpeptide mit einem „Doppelglycin“-Motiv.[42] Proteasen, die durch ein eigenständiges Gen (*lanP*) kodiert werden, können in ihrer Größe relativ stark variieren (LasP 266 Aminosäuren; NisP 682 Aminosäuren). Sie zeigen jedoch alle Homologien zu Subtilisin-*like* Serinproteasen und spalten die Präpropeptide, die in ihrer Leadersequenz das „FNLD“-Motiv enthalten. Extrazellulär lokalisierte Proteasen (s. Abschnitt 1.3) besitzen zusätzlich *N*-terminal eine Signalsequenz für ihren Export und teilweise auch ein *C*-terminales LPXTGX-Konsensusmotiv für die anschließende Verankerung in der Membran.[23, 24]

1.4 Biosynthese-Gencluster von Lantibiotika

Die Gene, die an der Biosynthese von Antibiotika und anderen Sekundärmetaboliten beteiligt sind, liegen zumeist in sog. Biosynthese-Genclustern auf transposablen Elementen des Genoms oder Plasmiden der jeweiligen Produzenten vor.[38] Häufig werden daneben auch Gene gefunden, die an der Regulation der Expression beteiligt sind oder die dem Produzenten eine Resistenz gegen das eigene Produkt verleihen. Die Untersuchungen zur Biosynthese einiger Lantibiotika haben gezeigt, dass auch hier die Gene meist in einem Cluster vorliegen (s. Abbildung 11).[6,23]

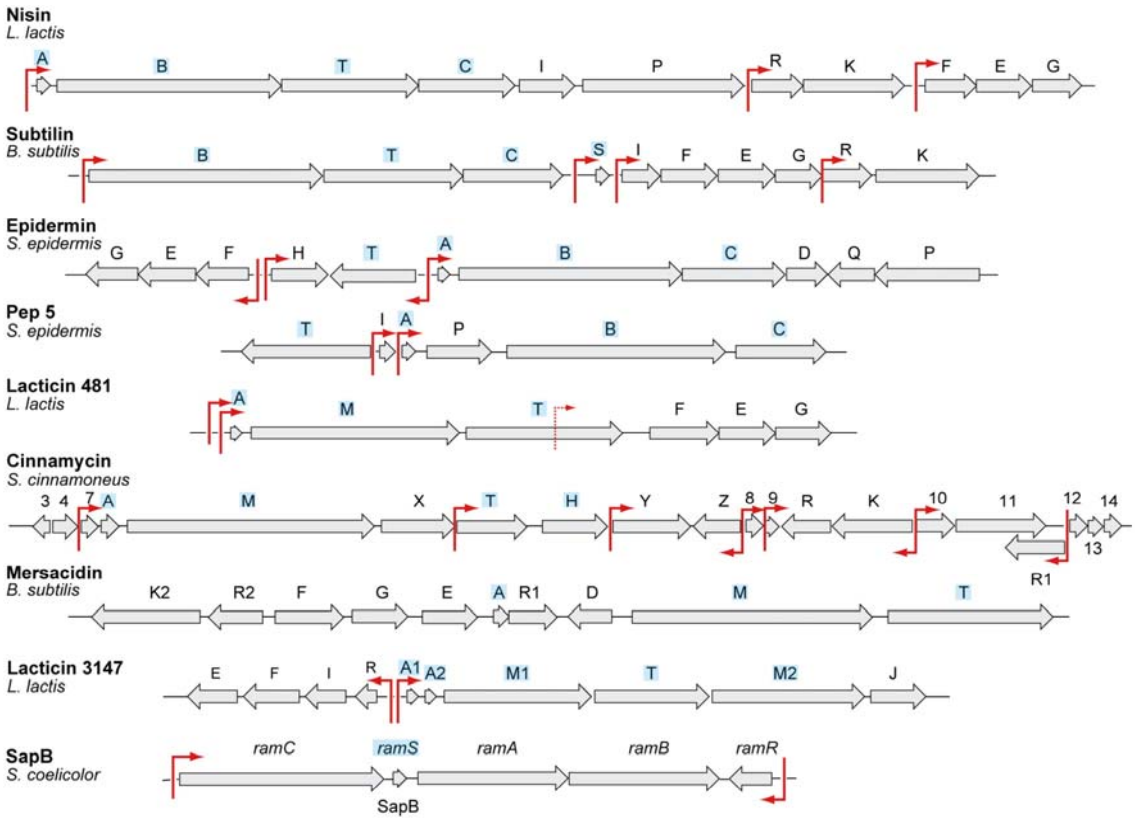


Abbildung 11: Übersicht verschiedener Biosynthese-Gencluster von Lantibiotika. Transkriptionseinheiten sind durch rote Pfeile gekennzeichnet.[6] Die in jedem Cluster vorhandenen Gene („Basisset“) wurden blau markiert.

Der Vergleich verschiedener Lantibiotika-Gencluster hat gezeigt, dass sich die meisten Cluster im Hinblick auf Orientierung, Anordnung und Ausstattung an Genen deutlich unterscheiden. Es konnte jedoch in allen Clustern ein „Basisset“ an Genen identifiziert werden, das in allen enthalten ist (*lanABCT* oder *lanAMT*). Dieses „Basisset“ besteht aus dem Gen für ein Präpropeptid (*lanA* bzw. aus historischen Gründen *spaS* bei Subtilin oder *ramS* bei SapB), teilweise werden sogar zwei gefunden, das im Verlauf der Biosynthese zum Lantibiotikum prozessiert wird. Daneben werden auch die Enzyme kodiert, die für die Einführung der Lan- und MeLan-Brücken verantwortlich sind. Wie in Abschnitt 1.3.2 beschrieben, handelt es sich dabei entweder um zwei Gene (*lanB* und *lanC*) oder nur um ein Gen (*lanM*). Zu dem „Basisset“ gehört auch das bisher in jedem Cluster gefundene Transportergen (*lanT*), das den in Abschnitt 1.3.3 beschriebenen Exporttransporter kodiert. Ausnahmen sind zum einen das Cinnymacin-Gencluster, bei dem die MSD und NBD als zwei separate Proteine kodiert werden (*cin*H und *cin*T)[39], und zum anderen das Epicidin 280-Gencluster, das kein *lanT* enthält.[40]

Das SapB-Gencluster passt von seiner Zusammensetzung her nicht zu den anderen genannten Clustern aus Abbildung 11, da die Gene des „Basissets“ hier nicht ohne weiteres zugeordnet werden können.[41]

Neben dem „Basisset“ an Genen sind in jedem Cluster auch immer die Gene zu finden, die für die Einführung weiterer posttranslationaler Modifikationen (s. Abschnitt 1.1) benötigt werden. Alle anderen Gene, die bisher in verschiedenen Lantibiotika-Genclustern identifiziert werden konnten, gehören nicht zum „Basisset“. So wurden in einzelnen Clustern z.B. Gene für Proteasen (*lanP*) gefunden, wenn diese nicht schon in *lanT* kodiert vorliegen oder Gene für Regulator- (*lanKR*) und Immunitätsproteine (*lanI* oder *lanEFG*). Teilweise sind diese Gene aber nicht im Cluster kodiert, was eine Vorhersage der erwarteten Cluster-zusammensetzung deutlich erschwert.

1.5 Einordnung der Lantibiotika in Gruppen

Aufgrund der stetig zunehmenden Zahl an gefundenen und strukturell charakterisierten Lantibiotika wurde eine Unterklassifizierung eingeführt.[42] Für die Einteilung in die drei Lantibiotika-Klassen wird hauptsächlich der Mechanismus der Einführung der charakteristischen Lan- und MeLan-Reste und das Vorhandensein einer antibiotischen Aktivität herangezogen. Zusätzlich wird auch die Struktur der Leadersequenz und der Aufbau der Operonstruktur des Biosynthese-Genclusters berücksichtigt.

Die Einführung der Lan- und MeLan-Brücken erfolgt bei den Klasse I-Lantibiotika ausschließlich durch LanB und LanC-Enzyme (s. Abschnitt 1.3.2). Die Leaderpeptide, die alle das „FNLD“-Motiv enthalten, werden hier vor oder nach dem Export von einer eigenständigen Serinprotease gespalten. Der Prototyp der Klasse I-Lantibiotika ist das Nisin (s. Abbildung 1 A)).

Bei den Klasse II-Lantibiotika sind an der Einführung der Lan- und MeLan-Brücken LanM-Enzyme beteiligt. Export und Abspaltung der Leadersequenzen mit dem „Doppelglycin“-Motiv wird hier durch ein LanT-Protein mit einer cytoplasmatischen Cysteinprotease-Domäne geleistet. Zu den Klasse II-Lantibiotika zählen neben den Duramycinen und dem Cinnamycin (s. Abbildung 1 B)) auch noch die zwei-Komponenten Lantibiotika.[6] Bei den zuletzt genannten handelt es sich immer um zwei Lantibiotika aus dem gleichen Produzenten, die zur vollen Entfaltung ihrer biologischen Aktivität synergistisch wirken.[43] In den entsprechenden Genclustern liegen deshalb immer zwei Gene für die Präpropeptide (*lanA1* und *lanA2*) vor. Im Fall von Lacticin 3147 A1 und A2 sind im Cluster zwar zwei *lanM*-Gene für die Modifikation, aber nur ein Transporter mit Proteasedomäne kodiert (s. Abbildung 11).

Zu den Klasse III-Lantibiotika werden die Verbindungen zusammengefasst, die in bisherigen Versuchen keine klare antibiotische Aktivität gezeigt haben. Dazu gehören nach aktuellem Kenntnisstand SapB, SapT und AmfS, von denen bisher nur SapT strukturell vollständig aufgeklärt wurde.[41,22,44] Die Zusammensetzung der Biosynthese-Gencluster weicht hier deutlich von denen aus Klasse I und Klasse II ab.

1.6 Die Entdeckungsgeschichte der Labyrinthopeptine

Im Zuge eines Screeningprogramms nach neuen Antibiotikaproduzenten wurde im Jahre 1988 der Stamm *Actinomadura namibiensis* aus der Wüste Namibias isoliert und taxonomisch eingeordnet.[45] Die Gattung Actinomadura gehört, wie die der Streptomyceten in die Ordnung der Actinomyceten, deren Mitglieder für ihr reichhaltiges Repertoire an Sekundärmetaboliten bekannt sind.[46,57] Der Stamm *A. namibiensis* zeigt eine lachsrosa Farbe und bildet ein Luftmycel mit spiralförmigen Sporenketten (s. Abbildung 12).

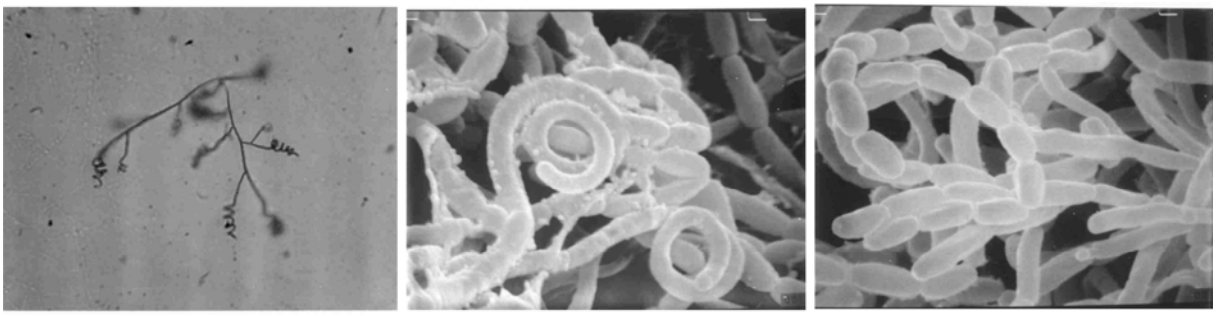


Abbildung 12: Licht- und elektronenmikroskopische Aufnahmen des Stammes *A. namibiensis*. Die Bilder zeigen die spiralförmigen Sporenketten des Luftmycels.[45]

Im Kulturüberstand von *A. namibiensis* konnten drei Sekundärmetabolite mit peptidischem Charakter beobachtet werden, die im Folgenden als A1, A2 und A3 bezeichnet wurden und in ein Screeningprogramm der Firma Sanofi-Aventis, Frankfurt, eingebracht wurden. Dort zeigten sie aber keine ansprechende Aktivität, da die gefundene antibakterielle und antivirale Wirkung erst bei hohen Konzentrationen zu beobachten war. Deshalb wurde die Strukturaufklärung von A1-A3 an den Arbeitskreis von Professor Süßmuth abgegeben, was außerdem mit dem Transfer des Stammes *A. namibiensis* in den Arbeitskreis verbunden war. Um strukturelle Informationen zu erhalten, bediente man sich verschiedener Methoden wie NMR, FTICR-MS und ASA sowie diversen Mikroreaktionen. Tabelle 1 zeigt die wichtigsten Daten, die zu den drei Verbindungen generiert werden konnten.

Tabelle 1: Zusammenstellung der analytischen Daten zu den Verbindungen A1-A3.

Methode	A1	A2	A3
FTICR-MS	2073,7624 amu ($C_{92}H_{119}N_{23}O_{25}S_4$)	1922,6930 amu ($C_{85}H_{110}N_{20}O_{24}S_4$)	2188,7943 amu
ASA	*L*-Ala(1), *L*-Thr(1), *L*-Asx(1), *L*-Cys(2), *L*-Phe(1), *L*-Glx(1), *L*-Trp(2), *L*-Gly(1), ***L*-Val(2), *L*-Pro(1)**	*L*-Ala(1), *L*-Thr(1), *L*-Asx(1), *L*-Cys(2), *L*-Phe(1), *L*-Glx(1), *L*-Trp(2), *L*-Gly(1), ***L*-Leu(2)**	-
Edman-Abbau	^{1}Xxx-^{2}Asn	^{1}Xxx-^{2}Asp	^{1}Asp-^{2}Xxx-^{3}Asn
DTT-Reduktion	eine Disulfidbrücke	eine Disulfidbrücke	-

Zu Beginn der Strukturaufklärung konzentrierte man sich zunächst auf die Verbindungen A1 und A2. Aufgrund von exakt bestimmten Molekülmassen aus hochauflösenden FTICR-MS-Spektren konnten von diesen beiden Verbindungen Summenformeln bestimmt werden. Eine ASA erbrachte für A1 und A2 die in Tabelle 1 aufgeführten Aminosäure-Zusammensetzungen. Hierbei fiel auf, dass bis auf Val und Pro in A1 und Leu in A2 die gleichen Aminosäuren im gleichen quantitativen Verhältnis gefunden wurden. Die Kalkulation der Molekülmassen und Summenformeln, basierend auf den Ergebnissen der ASA, zeigte durch Vergleich mit den experimentell bestimmten Molekülmassen, dass in beiden Verbindungen noch relativ große Strukturbereiche unbestimmt waren. Die Reduktion mit DTT zeigte in A1 und A2 die Existenz einer Disulfidbrücke an. Der Versuch einer *N*-terminalen Sequenzierung nach Edman[47] brach nach zwei bzw. drei Aminosäuren ab, wobei selbst in dieser kurzen Sequenz nicht alle Positionen (Xxx) zuzuordnen waren (s. Tabelle 1). Trotz großer Anstrengungen war es aufgrund von Sekundärstrukturbildung und der Größe des Moleküls nicht möglich, einen Strukturvorschlag aus 2D-NMR-Daten zu erstellen. Schließlich gelang es jedoch in Kooperation mit der Arbeitsgruppe von Professor Sheldrick (Universität Göttingen) eine Kristallstruktur von A2 zu erhalten, die deren Strukturaufklärung ermöglichte (s. Abschnitt 1.7).

1.7 Die Labyrinthopeptine A1, A2 und A3

Die Auswertung der Kristallstruktur von A2 ergab die in Abbildung 13 A) dargestellte, fünffach verbrückte Struktur (A-,B-,A`-,B`- und C-Ring), die aufgrund dieser starken Verbrückung der Seitenketten mit dem Namen Labyrinthopeptin A2 (LabA2) bezeichnet wurde. Zur Veranschaulichung zeigt Abbildung 13 verschiedene Darstellungsweisen der LabA2-Struktur.

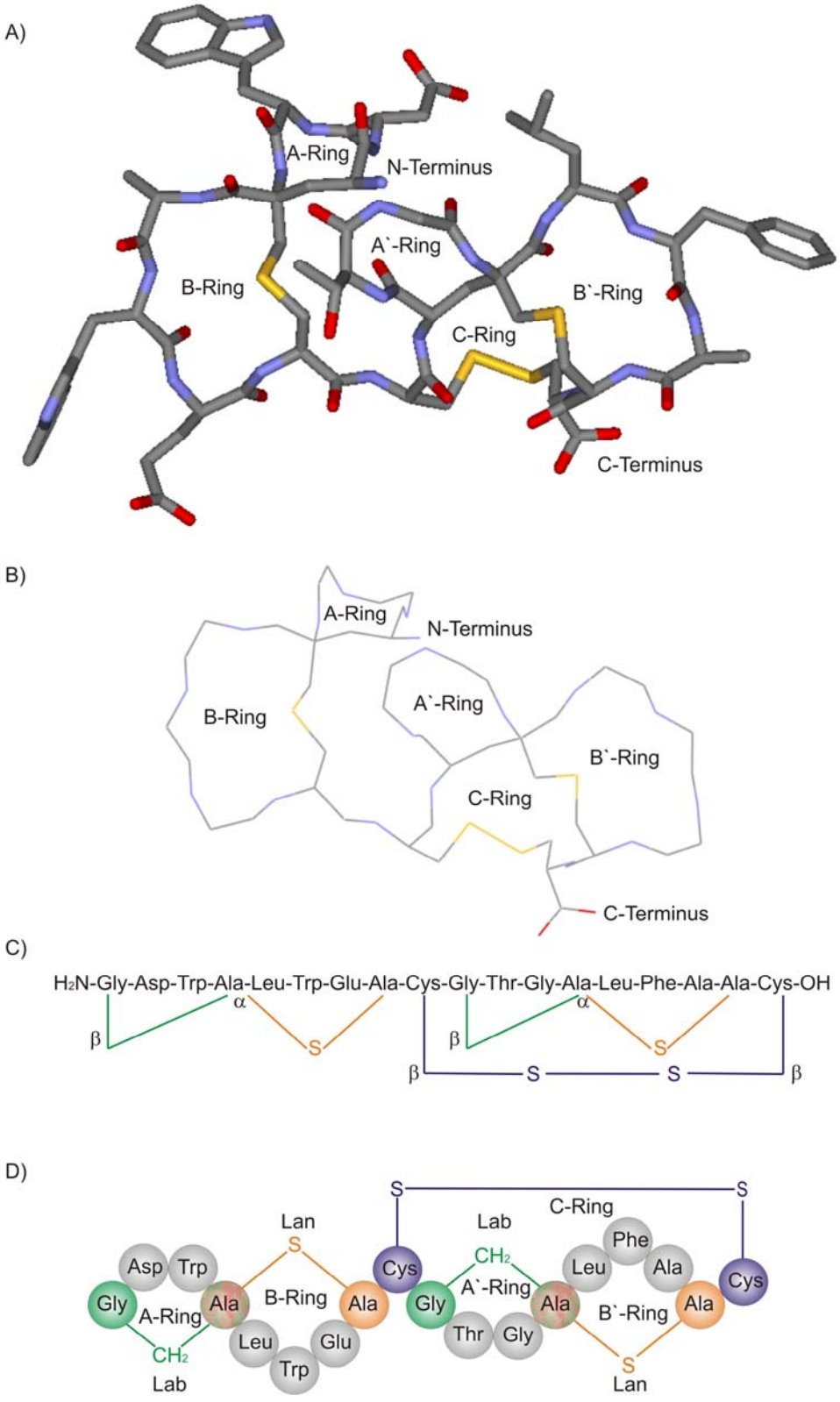


Abbildung 13: A) Stabmodell der Kristallstruktur von Labyrinthopeptin A2. C (grau); N (blau); O (rot); S (gelb) B) Dieses Modell zeigt den Verlauf des Rückgrats mit den Seitenkettenverbrückungen. C) und D) Schematische Darstellungen der LabA2-Struktur mit den fünf Ringen.

Die genaue Betrachtung der LabA2-Struktur zeigt, dass es sich um ein fünffach verbrücktes 18-er Peptid handelt, das aufgrund der Existenz zweier Lan-Brücken (B- und B`-Ring) zur Gruppe der Lantibiotika gehört (s. Abbildung 13 D)). Neben diesen zwei Lan-Brücken und einer Disulfidbrücke (C-Ring) konnten zwei weitere Verbrückungen (A- und A`-Ring) gefunden werden, die zuvor weder als posttranslationale Modifikation in Lantibiotika noch in anderen Sekundärmetaboliten beobachtet werden konnten. Dieses neue Strukturmotiv soll im Folgenden als Labionin (Lab) bezeichnet werden. In Abbildung 14 ist der Bereich des A`-Rings vergrößert dargestellt.

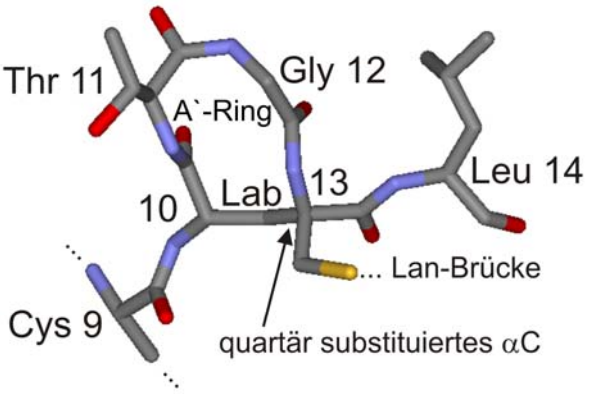


Abbildung 14: Ausschnitt aus der LabA2-Struktur. Der abgebildete A`-Ring entsteht formal aus einer Methylenbrücke zwischen zwei Gly-Resten (Labionin-Motiv oder Lab-Motiv). Da die Aminosäure 13 aber auch an der Bildung des B`-Rings (Lan-Brücke) beteiligt ist, entsteht hier ein quartär substituiertes αC-Atom.

Bei dem Labionin-Strukturmotiv handelt es sich formal um zwei Gly-Reste, die durch eine Methyleneinheit verbrückt werden. Da im A- und A`-Ring aber die jeweils *C*-terminal an der Bildung des Lab-Motivs involvierten Aminosäuren auch an der Ausbildung der nachfolgenden Lan-Brücke beteiligt sind, entstehen an diesen Positionen quartär substituierte αC-Atome (s. Abbildung 14). Dieses Strukturmotiv ist bisher nicht bekannt und sehr außergewöhnlich. Das 18-er Peptid kann formal in zwei 9-er Peptide zerlegt werden, die den gleichen Aufbau des Verbrückungsmusters zeigen und jeweils endständig ein Cys tragen, die in LabA2 die Disulfidbrücke bilden. Durch die Verbrückungen entstehen im 18-er Peptid kanonische Abfolgen von einem Labionin-Lanthionin-Motiv. Wie man in Abbildung 13 C) sehr gut erkennen kann, entsteht durch die Labionin-Verbrückung jeweils eine Dipeptidschleife und durch die Lanthionin-Verbrückung eine Tripeptidschleife. Die für den C-Ring verantwortliche Disulfidbrücke ist bei Lantibiotika bisher nur bei Sublancin beobachtet worden.[48] Die mehrfache Verbrückung führt schließlich zu der kompakten globulären Struktur von LabA2.

Bei genauer Betrachtung der einzelnen Peptidbindungen konnte festgestellt werden, dass in den A- bzw. A`-Ringen jeweils eine cis-Amidbindung vorkommt. Im A-Ring liegt diese Amidbindung zwischen Asp und Trp, im A`-Ring zwischen Thr und Gly, was in Abbildung 15 A) an den entsprechenden Ausschnitten der Kristallstruktur gezeigt wird. Durch diese Cis-Amidbindungen nehmen der A- und A`-Ring β-turn-ähnliche Strukturen an.

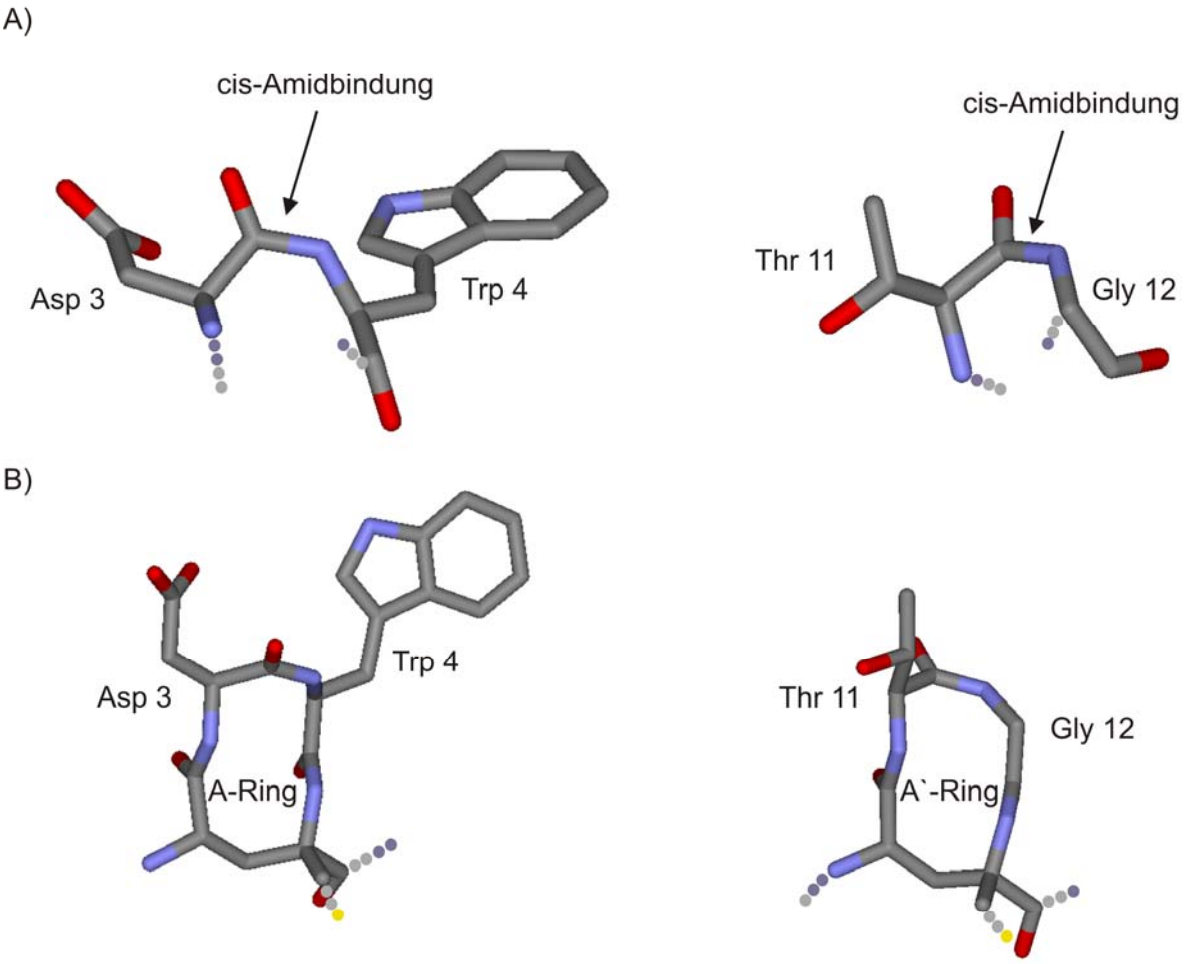


Abbildung 15: A) Ausschnitte aus den A- und A`-Ringen von LabA2, die die cis-Amidbindungen zeigen. B) Durch diese cis-Amidbindungen bilden die A- und A`-Ringe β-turn-ähnliche Strukturen.

Diese Kristallstruktur von LabA2 steht im Einklang mit den früher experimentell bestimmten Daten zu Molekülmasse, Summenformel und Aminosäure-Zusammensetzung (s. Tabelle 1). Die in LabA2 gefundenen Strukturmotive (Lab- und Lan-Brücken) erklären auch, warum die ASA nur einen Teil der Aminosäuren nachweisen konnte.

A1 zeigt bezüglich der Ergebnisse der ASA (s. Tabelle 1) eine große Übereinstimmung mit LabA2, so dass zumindest, was die nichtmodifizierten Aminosäuren betrifft, auf eine gewisse strukturelle Ähnlichkeit geschlossen werden kann. Aus diesem Grund erhielten A1 und A3 in Analogie zu Labyrinthopeptin A2 die Namen Labyrinthopeptin A1 und A3 (LabA1 und LabA3). Addiert man formal die Lab- und Lan-Verbrückungen aus LabA2 ($C_{18}H_{22}N_6O_6S_2$) zu

der über die ASA bestimmten Summenformel von LabA1, bleibt immer noch ein Rest übrig (C_4H_5NO), der die Existenz einer weiteren nicht proteinogenen Aminosäure andeutet. Die für LabA1 und LabA3 bestimmten Molekülmassen deuten darauf hin, dass LabA3 ein zusätzliches Asp im Vergleich zu LabA1 besitzt. Dieses könnte nach den Ergebnissen des Edman-Abbaus am *N*-Terminus sitzen (s. Tabelle 1). Bezüglich LabA1 und LabA3 existieren also nur strukturelle Vermutungen, die sich auf Teilergebnisse der Strukturaufklärung stützen.

2 Zielsetzung

Grundlage für die vorliegende Arbeit war der 1988 als *Actinomadura namibiensis* taxonomisch eingeordnete Mikroorganismus, der die drei peptidischen Sekundärmetabolite A1, A2 und A3 produziert. Die Struktur von A2 konnte mit Hilfe einer Röntgenkristallstrukturanalyse aufgeklärt werden, was die Einordnung in die Klasse der ribosomal synthetisierten Lantibiotika ermöglichte. Die im Folgenden als Labyrinthopeptin A2 bezeichnete Verbindung weist neben den typischerweise in Lantibiotika vorkommenden Lanthioninbrücken sowie einer Disulfidbrücke eine strukturelle Besonderheit auf, die bisher nicht literaturbekannt ist und als Labionin bezeichnet wurde. Dabei handelt es sich um eine Verbrückung, die zu quartär substituierten αC-Atomen führt und somit eine ungewöhnliche C-C-Bindungsknüpfung während der Biosynthese erfordert.

Ziel der vorliegenden Arbeit war es, ausgehend von den o.g. Vorkenntnissen die Biosynthese von Labyrinthopeptin A2 und dabei insbesondere der Labionin-Verbrückung aufzuklären. Bei Sekundärmetaboliten liegen typischerweise die für die Biosynthese zuständigen Gene zusammen mit den Genen für Regulation, Immunität und Transport in einem sog. Biosynthese-Gencluster vor. Bei den ribosomal synthetisierten Lantibiotika enthalten die entsprechenden Biosynthese-Gencluster immer ein „Basisset“ aus Strukturgen, Transportergen sowie Genen für die posttranslationalen Modifikationen, weshalb vermutet werden kann, dass das Gen bzw. die Gene für die Einführung der Labionin-Verbrückung in Labyrinthopeptin A2 auch in einem solchen Cluster vorliegt. Um das Ziel der Biosyntheseaufklärung von Labyrinthopeptin A2 bzw. der Labionin-Verbrückung zu erreichen, sollte deshalb zunächst das entsprechende Biosynthese-Gencluster mit Hilfe einer Cosmidbank gefunden und sequenziert werden. Die Auswertung der so generierten DNA-Sequenzdaten sollte die Identifizierung eines bzw. mehrerer möglicherweise für die Einführung der Labionin-Verbrückung zuständigen Gene ermöglichen. Eventuell befinden sich in diesem Biosynthese-Gencluster auch die Strukturgene für die Labyrinthopeptine A1 und A3, was wertvolle Informationen über die mittels NMR und MS schwer zugänglichen Strukturen liefern könnte.

3 Materialien

3.1 Geräte

Agarosegelkammern:	Mini-Sub Cell GT; Bio-Rad Laboratories GmbH, München
Blot-Apparatur für *Western-Blot*:	Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell; Bio-Rad Laboratories GmbH, München
Brutschrank:	BD 53; Binder, Tuttlingen
Eismaschine:	AF 80 AS; Scotsman Europe, Mailand, Italien
Elektroporator:	Gene Pulser; Bio-Rad Laboratories GmbH, München
French Press:	SIM-AMINCO; SLM Instruments, Inc. Rochester, New York, USA
Geldokumentation:	GelSystem Flexi 5020 S; biostep GmbH, Jahnsdorf
Heizblock:	Thermomixer comfort; Eppendorf AG, Hamburg
HPLC:	1100-Series; Agilent Technologies Deutschland GmbH, Waldbronn
Hybridisierungsofen:	Hybaid Maxi 14; Thermo Electron GmbH, Dreieich
Massenspektrometer:	QTRAP 2000; Applied Biosystems Applera Deutschland GmbH, Darmstadt QTOF-2; Waters GmbH, Eschborn
Thermocycler:	Primus 96 advanced® Gradient; PEQLAB Biotechnologie GMBH, Erlangen
pH-Meter:	pH/Ion 510; Eutech Instruments Pte Ltd, Singapur
Proteingelkammern:	Mini-Protean 3; Bio-Rad Laboratories GmbH, München
Rotoren:	Winkelrotor 220.87 V05; HERMLE GmbH, Wehingen Ausschwingrotor 221.08 V01; HERMLE GmbH, Wehingen (für Zentrifugationen bis 4500 rpm) Winkelrotor 220.78 V02; HERMLE GmbH, Wehingen (für Zentrifugationen über 4500 rpm)
Säulen:	Luna 3u C18(2) 100Å 50 x 1 mm; Phenomenex, Aschaffenburg *Jupiter*; 4u, 300Å 15 x 1 mm; Phenomenex, Aschaffenburg
Schlauchpumpe:	Reglo-Analog MS-2/6-160; ISMATEC Laboratoriumstechnik GmbH, Wertheim-Mondfeld
Schüttelinkubatoren:	MULTITRON II; Infors GmbH, Einsbach

Spannungsquellen:	PowerPac Basic; Bio-Rad Laboratories GmbH, München PowerPac HC; Bio-Rad Laboratories GmbH, München PowerPac 1000; Bio-Rad Laboratories GmbH, München
Sterilbank:	Holten Safe 2010; Thermo Electron, Dreieich
Taumeltisch:	Polymax 1040 Neigungswinkel 5°; Heidolph, Schwabach
Photometer:	Ultrospec 2100 pro Classic; Amersham Phamacia Biotech, Freiburg
Reagenzglasschüttler:	Reax top; Heidolph, Schwabach
Waagen:	EW 6000-1M; Kern & Sohn GmbH, Balingen XR 125SM; Precisa Instruments AG, Dietikon, Schweiz
Zentrifugen:	Mikroliterzentrifuge Z233MK-2; HERMLE GmbH, Wehingen Tischkühlzentrifuge Z383K; HERMLE GmbH, Wehingen

3.2 Chemikalien

Acetonitril (HPLC-MS grade):	Acros Organics, Geel, Belgien
Acrylamid/Bisacrylamid-Lösung (30 % / 0,8 %):	Carl Roth GmbH & Co, Karlsruhe
Agarose Broad Range ROTI® GAROSE für die Elektrophorese im gesamten Größenbereich:	Carl Roth GmbH & Co, Karlsruhe
Ameisensäure (HCOOH):	Carl Roth GmbH & Co, Karlsruhe
Ammoniumperoxodisulfat (APS):	Carl Roth GmbH & Co, Karlsruhe
Anti-polyHis-Antikörper (monoklonaler Antikörper aus der Maus):	Sigma-Aldrich Chemie GmbH, München
Anti-Maus-IgG-Antikörper aus der Ziege (mit alkalischer Phosphatase konjugiert):	Sigma-Aldrich Chemie GmbH, München
5-Brom-4-Chlorindolylphosphat (BCIP):	Carl Roth GmbH & Co, Karlsruhe
Bromphenolblau:	Carl Roth GmbH & Co, Karlsruhe
Corn Steep Powder:	Marcor, USA
Difco Bacto-Pepton:	Becton Dickinson GmbH, Heidelberg
Difco Casaminoacids:	Becton Dickinson GmbH, Heidelberg
Difco Hefe-Extrakt:	Becton Dickinson GmbH, Heidelberg
Dimethylformamid (DMF):	Acros Organics, Geel, Belgien
Dinatriumethylendiamintetraessigsäure (EDTA):	Carl Roth GmbH & Co, Karlsruhe
dNTP-Mix:	New England Biolabs GmbH, Frankfurt
Dithiothreitol (DTT):	Carl Roth GmbH & Co, Karlsruhe
Ethanol (mit Methylethylketon vergällt):	Carl Roth GmbH & Co, Karlsruhe
Ethidiumbromid:	Carl Roth GmbH & Co, Karlsruhe

Filterpapier (extra thick):	Bio-Rad Laboratories GmbH, München
Glukose (D+) wasserfrei:	Merck KGaA, Darmstadt
Glycerin:	Carl Roth GmbH & Co, Karlsruhe
Glycin (Gly):	Carl Roth GmbH & Co, Karlsruhe
Hefe-Extrakt:	Carl Roth GmbH & Co, Karlsruhe
H_2O_d:	hausinterne Wasserentsalzungsanlage
H_2O_{dd}:	membraPure, Bodenheim
H_2O (HPLC-MS grade):	Carl Roth GmbH & Co, Karlsruhe
Hybond-N Membran:	Amersham Phamacia Biotech, Freiburg
Hyperfilm MP Röntgenfilm:	Amersham Phamacia Biotech, Freiburg
Imidazol:	Carl Roth GmbH & Co, Karlsruhe
Isopropanol:	Carl Roth GmbH & Co, Karlsruhe
L-Arabinose:	Carl Roth GmbH & Co, Karlsruhe
Methanol (gradient grade):	Acros Organics, Geel, Belgien
Natriumdodecylsulfat (SDS):	Carl Roth GmbH & Co, Karlsruhe
Ni-NTA-Agarose:	Qiagen GmbH, Hilden
4-Nitroblautetrazoliumchlorid (NBT):	Carl Roth GmbH & Co, Karlsruhe
N,N,N´,N´-Tetramethylethylendiamin (TEMED):	Carl Roth GmbH & Co, Karlsruhe
Nutrient Broth:	Becton Dickinson GmbH, Heidelberg
Oxoid-Malz-Extrakt:	Carl Roth GmbH & Co, Karlsruhe
Polyethylenglykol (PEG) 1000:	Carl Roth GmbH & Co, Karlsruhe
vorgefärbte Proteinleiter, 10-180 kDa (#SM0671):	Fermentas GmbH, St. Leon-Rot
Proteinleiter, 10-200 kDa (#SM0661):	Fermentas GmbH, St. Leon-Rot
Protran® Nitrozellulose Transfer Membran BA85, 0,45 µm:	Schleicher & Schuell BioScience, Dassel
Rinderserumalbumin (BSA):	Carl Roth GmbH & Co, Karlsruhe
Roti®-Quant Reagenz:	Carl Roth GmbH & Co, Karlsruhe
Roti®-Phenol/Chloroform/Isoamylalkohol:	Carl Roth GmbH & Co, Karlsruhe
ROTIPURAN® Eisessig:	Carl Roth GmbH & Co, Karlsruhe
SERVAPOR® Dialysemembran (MWCO 12000 - 14000):	Serva, Feinbiochemica GmbH, Heidelberg
Spectra/Por® - MWCO 2000:	Carl Roth GmbH & Co, Karlsruhe
Saccharose:	Sigma-Aldrich Chemie GmbH, München
SYBRGreen:	Roche Diagnostics GmbH, Mannheim
Trichloressigsäure (p.A.) (TCA):	Carl Roth GmbH & Co, Karlsruhe
TriDye 100bp DNA Leiter:	New England Biolabs GmbH, Frankfurt
TriDye 2-log DNA Leiter:	New England Biolabs GmbH, Frankfurt
Tris-(hydroxymethyl)-aminomethan (Tris):	Carl Roth GmbH & Co, Karlsruhe

Triton X 100:	Carl Roth GmbH & Co, Karlsruhe
Trypton/Pepton aus Casein:	Carl Roth GmbH & Co, Karlsruhe
Tween® 20:	Carl Roth GmbH & Co, Karlsruhe

Alle hier nicht aufgeführten Chemikalien waren analysenreine Reagenzien der Firma Carl Roth GmbH & Co, Karlsruhe

3.3 Kits

3.3.1 DNA-Präparationen und Aufreinigung

peqGOLD Plasmid Miniprep-Kit I:	PEQLAB Biotechnologie GmbH, Erlangen
NucleoSpin® Tissue-Kit:	MACHEREY-NAGEL GmbH, Düren
NucleoBond® Bac 100-Kit:	MACHEREY-NAGEL GmbH, Düren
QIAEX® II Gel Extraction-Kit:	Qiagen GmbH, Hilden

3.3.2 Klonierung

A-Addition-Kit:	Qiagen GmbH, Hilden
pDive PCR Cloning-Kit:	Qiagen GmbH, Hilden
TOPO TA Cloning®-Kit:	Invitrogen GmbH, Karlsruhe

3.3.3 Nichtradioaktive Markierung und Detektion von DNA-Sonden

DIG-High Prime DNA labeling and detection Start-Kit I:	Roche Diagnostics GmbH, Mannheim
PCR DIG labeling-Mix:	Roche Diagnostics GmbH, Mannheim
DIG Nucleic Acid Detection-Kit:	Roche Diagnostics GmbH, Mannheim

3.3.4 Heterologe Expression

Champion pET SUMO Protein Expression-System:	Invitrogen GmbH, Karlsruhe

3.3.5 Mutagenese

QuickChange® II Site-Directed Mutagenesis-Kit:	Stratagene Europe, Amsterdam Zuidoost, Niederlande

3.3.6 Färben von Proteingelen

SilverQuest Silver Staining-Kit	Invitrogen, Karlsruhe

3.4 Puffer und andere Lösungen

3.4.1 SDS-Gelelektrophorese

10 x Elektrophoresepuffer:	25 mM Tris 1,9 M Glycin 1 % SDS pH ergibt sich zu 8,9
Trenngelpuffer:	1,5 M Tris/HCl; pH 8,8
Sammelgelpuffer:	0,25 M Tris/HCl; pH 6,8
APS-Stammlösung:	100 mg/mL
SDS-Stammlösung:	10 % SDS
Trenngel (10 %):	7,6 mL Acrylamid/Bisacrylamid-Lösung 9,0 mL H_2O_d 5,74 mL Trenngelpuffer 115 µl SDS-Stammlösung 57,5 µl TEMED 460 µl APS-Stammlösung
Sammelgel (5 %):	2,06 mL Acrylamid/Bisacrylamid-Lösung 3,06 mL H_2O_d 6,3 mL Sammmelgelpuffer 120 µl SDS-Stammlösung 30 µl TEMED 80 µl APS-Stammlösung
Probenauftragspuffer:	3,75 mL 1 M Tris/HCl; pH 6,8 1,2 g SDS 6 mL Glycerin 6 mg Bromphenolblau ad 13,5 mL H_2O_d
kurz vor Verwendung wird 1,5 mL 1M DTT zugesetzt	
Fixierlösung:	25 mL Isopropanol 10 mL Essigsäure ad 100 mL H_2O_d

3.4.2 Elektroporation

Mg^{2+}-Stammlösung:	1M $MgSO_4$ 1M $MgCl_2$

3.4.3 Protoplastentransformation

Sucrose-Lösung: 10,3 % (w/v) Saccharose

P (Protoplast) Puffer:
51,5 g Saccharose
0,125 g K_2SO_4
1,01 g $MgCl_2$*6H_2O
1 mL Spurenenlement-Lösung
ad 400 mL H_2O_d

in fünf Gefäße zu je 80 mL autoklavieren und kurz vor dem Gebrauch jeweils folgendes zusetzen:
1mL KH_2PO_4 (0,5 % (w/v))
10 mL $CaCl_2$ (3,68 % (w/v))

Lysozym-Lösung: 1 mg/mL Lysozym in P-Puffer

P-PEG-Puffer: 25 % (w/v) PEG1000 in P-Puffer

NB-Weichagar:
800 mg Nutrient Broth
500 mg Agar
ad 100 mL H_2O_d

autoklavieren und bis zum Gebrauch im 65°C-Schrank aufbewahren

3.4.4 Agarosegelelektrophorese

10x DNA-Probenauftragspuffer:
10 mM Tris
5 % Saccharose
0,02 % SDS
0,01 % Bromphenolblau

Ethidiumbromidfärbebad:
1 Spatelspitze Ethidiumbromid /
200mL H_2O_{dd}

TAE-Elektrophoresepuffer:
40 mM Tris/HCl
20 mM Natriumacetat
1 mM EDTA
pH 7,8

3.4.5 *Southern-Blot*

Southern I: 0,25 M HCl

Southern II:
0,5 M NaOH
1,5 M NaCl

Southern III:
1,0 M Tris
1,5 M NaCl
pH 7,5

20xSSC:
3,0 M NaCl
0,35 M Na-Citrat

PO_4-Lösung:
1,026 M NaH_2PO_4
0,974 M Na_2HPO_4
pH 7,0

Hybridisierungs-Lösung:	100 mL PO_4-Lösung 280 mL 10 % (w/v) SDS-Lösung 20 mL H_2O_d
Wasch-Lösung:	50 mL PO_4-Lösung 100 mL 10 % (w/v) SDS-Lösung 850 mL H_2O_d

3.4.6 Screening auf Hybridisierungsfilter der Cosmidbank

Church-Puffer:	500 mL NaH_2PO_4 500 mL 10 % (w/v) SDS-Lösung 2 mL EDTA (5,0 M) pH 7,2
2xSSC-Puffer:	2xSSC (0,3 M NaCl; 35 mM Na-Citrat) 0,1 % SDS
0,5xSSC-Puffer:	0,5xSSC (75 mM NaCl; 9 mM Na-Citrat) 0,1 % SDS
Maleinsäurepuffer:	0,1 M Maleinsäure 0,15M NaCl pH 7,5
Waschpuffer:	Maleinsäurepuffer 0,3 % (v/v) Tween 20
Blockpuffer:	90 mL Waschpuffer 10 mL 10XBlocking reagent
Antikörper-Lösung:	2 mL Blocking solution 18 mL Maleinsäurepuffer 4 µl Anti-Digoxigenin-AP-Antikörper
Detektions-Lösung:	0,1 M Tris 0,1 M NaCl pH 9,5
Färbe-Lösung:	10 mL Detektions-Lösung 200 µl NBT/BCIP-Stammlösung

3.4.7 Isolierung genomischer DNA mit dem *NucleoSpin® Tissue-Kit*

Wasch-Lösung:	10 mM Tris 1 mM EDTA pH 8,0 10 % (w/v) Saccharose
Lyse-Lösung:	20 mM Tris 2mM EDTA 1 % Triton X 100 pH 8,0 20 mg/mL Lysozym

3.4.8 Isolierung hochmolekularer genomischer DNA mittels Phenol/ Chloroform-Extraktion

Wasch-Lösung:
10 mM Tris
1 mM EDTA
pH 8,0
10 % (w/v) Saccharose

Lyse-Lösung-PC:
Wasch-Lösung
10 mg/mL Lysozym

TE-Puffer:
10 mM Tris
1 mM EDTA
pH 8,0

3.4.9 SDS-PAGE nach Lämmli

Trenngelpuffer: 1,5 M Tris/HCl; pH 8,8

Sammelgelpuffer: 0,25 M Tris/HCl; pH 6,8

10 x Elektrophoresepuffer:
25 mM Tris
1,9 M Glycin
1 % SDS
pH ergibt sich zu 8,9

Probenauftragspuffer:
3,75 mL 1 M Tris/HCl; pH 6,8
1,2 g SDS
6 mL Glycerin
6 mg Bromphenolblau
ad 13,5 mL H_2O_d
kurz vor Verwendung wird 1,5 mL 1M DTT zugesetzt

APS-Lösung: 100 mg/mL

3.4.10 SDS-PAGE nach Schägger und Jagow

Gelpuffer: 3 M Tris/HCl; pH 8,45

Anodenpuffer: 0,2 M Tris/HCl; pH 8,9

Kathodenpuffer: 0,1 M Tris/ 0,1 M Tricin; 0,3 % (w/v) SDS

APS-Lösung: 100 mg/mL

Probenauftragspuffer: Tricin Sample Buffer, Bio-Rad (161-0739)
kurz vor Verwendung wird 1/50 Volumen β-Mercaptoethanol (2 %; v/v) zugesetzt

3.4.11 Photometrische Proteinbestimmung nach Bradford

Färbe-Lösung:
2 Teile Roti®-Quant
5,5 Teile H_2O_{dd}

3.4.12 Gelfärbung mit Coomassie Brillant Blau

Färbe-Lösung:	0,85 g Coomassie Brillant Blau G 250 250 mL Ethanol 50 mL Eisessig ad 500 mL H_2O_{dd}
Entfärbe-Lösung:	100 mL Ethanol 100 mL Eisessig ad 1 L H_2O_d

3.4.13 *Western-Blot*

Blot-Puffer:	25 mM Tris 192 mM Glycin 20 % MeOH
Tris-Natriumchlorid-Tween-Puffer (TNT-Puffer):	20 mM Tris/HCl 500 mM NaCl 0,05 % Tween 20 pH 7,5
AP-Puffer:	0,1 M Tris/HCl 0,1 M NaCl 5 mM $MgCl_2$ pH 9,5
4-Nitroblautetrazolium-Lösung:	50 mg/mL in 70 % DMF
5-Brom-4-Chlorindolylphosphat-Lösung:	10 mg/mL in 100 % DMF
Färbe-Lösung:	135 µl 4-Nitroblautetrazolium-Lösung 300 µl 5-Brom-4-Chlorindolylphosphat-Lösung ad 15 mL AP-Puffer

3.4.14 Proteinreinigung mittels IMAC

Lysepuffer:	50 mM Natriumdihydrogenphosphat 300 mM NaCl 20 mM Imidazol pH 8,0
Waschpuffer:	50 mM Natriumdihydrogenphosphat 300 mM NaCl 35 mM Imidazol pH 8,0
Elutionspuffer 1:	50 mM Natriumdihydrogenphosphat 300 mM NaCl 130 mM Imidazol pH 8,0
Elutionspuffer 2:	50 mM Natriumdihydrogenphosphat 300 mM NaCl 250 mM Imidazol pH 8,0

Bradford-Lösung:
100 mg Coomassie Brillant Blau G 250
100 mL Phosphorsäure (85 %)
50 mL Ethanol
ad 1L H_2O_d

3.4.15 Abspaltung des N-terminalen His-Tag mit der SUMO-Protease

Dialysepuffer:
25 mM Tris
300 mM NaCl
20 % Glycerin
1 mM DTT

3.4.16 Weitere Medienzusätze

Spurenelement-Lösung:
200 mg $FeCl_3x6H_2O$
40 mg $ZnCl_2$
10 mg $Na_2B_4O_7x10H_2O$
10 mg $(NH_4)_6Mo_7O_{24}x4H_2O$
7,88 mg $CuCl_2$
10 mg $MnCl_2x4H_2O$
mit H_2O_d auf 1 L, autoklavieren

IPTG-Lösung (1000x): 11,9 mg/mL (50 mM)

X-Gal-Lösung (1000x): 80 mg/mL (200 mM)

3.5 Antibiotika

Antibiotikum	Abkürzung	Endkonz. [µg/mL]	Lösungsmittel	Hersteller
Ampicillin	Amp	100	H_2O	Carl Roth GmbH & Co, Karlsruhe
Apramycin	Apra	100	H_2O	Sigma-Aldrich Chemie GmbH, München
Chloramphenicol	Cm	34	100 % EtOH	Carl Roth GmbH & Co, Karlsruhe
Kanamycin	Kan	50	H_2O	Carl Roth GmbH & Co, Karlsruhe
Carbenicillin	Carb	50	H_2O	Carl Roth GmbH & Co, Karlsruhe

3.6 Enzyme

Lysozym	Serva, Feinbiochemica GmbH, Heidelberg
RNase A	Fermentas GmbH, St.Leon-Rot
Taq-Polymerase	Qiagen GmbH, Hilden
Pwo DNA-Polymerase	Roche Diagnostics GmbH, Mannheim
MasterTaq DNA-Polymerase	Eppendorf AG, Hamburg
Exonuklease I	Fermentas GmbH, St.Leon-Rot
T4-DNA-Ligase	New England Biolabs GmbH, Frankfurt
DNaseI	Sigma-Aldrich Chemie GmbH, München
Endoproteinase GluC	New England Biolabs GmbH, Frankfurt
Restriktionsendonukleasen	Fermentas GmbH, St.Leon-Rot New England Biolabs GmbH, Frankfurt

3.7 Bakterien

Stamm	Genotyp/Phänotyp	Referenz
E. coli DH5α	*supE44 ΔlacU169* (Φ80*lacZ*ΔM15) *hsdR17 recA1 endA1 gyrA thi-1 relA1*	Hanahan, 1983[49]
E. coli Mach1™-T1^R	F$^-$ Φ80*lacZ*ΔM15 Δ*lac*X74 *hsd*R($r_K^-m_K^+$) Δ*rec*A1398 *end*A1 *ton*A	Invitrogen
E. coli BL21 (DE3)	F$^-$ *ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3)	Invitrogen
E. coli ET12567	F-, *dam-13::Tn9*, *dcm-6*, *hsdM, hsdR, lacY1*	MacNeil, 1992[50]
E. coli BW25113	*lacI^q rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$	Datsenko, Wanner[94]
E. coli XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacIZΔM15* Tn*10* (Tet^r)]	Stratagene
Streptomyces lividans		Combinature Biopharm AG
Actinomadura namibiensis	Labyrinthopeptin-Produzent	Sanofi-Aventis

3.8 Nährmedien

Falls nicht gesondert angegeben, beziehen sich die folgenden Angaben auf 1 L H_2O_d. Zur Herstellung der entsprechenden Festmedien werden, wenn nicht gesondert vermerkt, 15 g/L Agar zugegeben.

TY-Medium	Trypton/Pepton aus Casein	8,0 g
	Hefe-Extrakt	5,0 g
	NaCl	5,0 g
LB-Medium	Trypton/Pepton aus Casein	10,0 g
	Hefe-Extrakt	5,0 g
	NaCl	10,0 g
SOB-Medium	Trypton/Pepton aus Casein	20,0 g
	Hefe-Extrakt	5,0 g
	NaCl	0,5 g
	KCl	186,4 mg
	pH 7,0	
	nach dem Autoklavieren folgende sterile Lösungen zugeben:	
	1M $MgCl_2$	5 mL
	1M $MgSO_4$	5 mL
SOC-Medium	SOB	1 L
	nach der Herstellung von SOB folgende sterile Lösung zugeben:	
	Glukose 1M	20,0 mL
R5-Medium	Saccharose	103,0 g
	Glukose (*D*+) wasserfrei	10,0 g
	Hefe-Extrakt	5,0 g
	$MgCl_2$	10,12 g
	K_2SO_4	0,25 g
	TES	5,73 g
	Difco Casaminoacids	0,1 g
	Spurenelement-Lösung	2 mL
	pH 7,2	
	nach dem Autoklavieren folgende sterile Lösungen zugeben:	
	KH_2PO_4 0,5 % (w/v)	10 mL
	$CaCl_2$ 1 M	20 mL
	L-Prolin 20 % (w/v)	15 mL
KM4-Medium	Hefe-Extrakt	4,0 g
	Malz-Extrakt	10,0 g
	Glukose (*D*+) wasserfrei	4,0 g
	pH 7,3 (Leitungswasser)	

Medium 5294	Stärke	10,0 g
	Hefe-Extrakt	2,0 g
	Glukose	10,0 g
	Glycerin 86 %	11,6 mL
	Corn Steep Powder	2,5 g
	Pepton	2,0 g
	NaCl	1,0 g
	pH 7,2 (Leitungswasser)	
YEME-Medium (Yeast extract-malt extract medium)	Hefe-Extrakt (Difco)	3,0 g
	Difco Bacto-Pepton	5,0 g
	Oxoid Malz-Extrakt	3,0 g
	Glukose (*D*+) wasserfrei	10,0 g
	Saccharose	340,0 g
	nach dem Autoklavieren folgende sterile Lösungen zugeben:	
	$MgCl_2$ (2,5 M)	2,0 mL
	Glycerin (20 %)	25,0 mL
R2YE-Medium	Saccharose	103,0 g
	K_2SO_4	0,25 g
	$MgCl_2*6H_2O$	10,12 g
	Glukose (*D*+) wasserfrei	10,0 g
	Difco Casaminoacids	0,1 g
	Agar	22,0 g
	ad 800mL H_2O_d	
	nach dem Autoklavieren folgende sterile Lösungen zugeben:	
	KH_2PO_4 0,5 % (w/V)	10,0 mL
	$CaCl_2*2H_2O$ 3,68 % (w/V)	80,0 mL
	L-Prolin 20 % (w/v)	15,0 mL
	TES-Puffer 5,73 % (w/v) pH 7,2	100,0 mL
	Spurenelement-Lösung	2,0 mL
	NaOH 1M	5,0 mL
	Hefe-Extrakt (Difco) 10 % (w/v)	50,0 mL

3.9 Oligonukleotid-Primer

Name	Sequenz (5´->3´)	Anmerkung
EV1	5´-CAGGAAACAGCTATGACCGAYTGGWSNYTNTGGG-3´	
EH1	5´-TGTAAAACGACGGCCAGTRCANGANGCRAANARRC-3´	Nur der rot markierte Bereich bindet an das Ursprungs-templat.
EH2	5´-TGTAAAACGACGGCCAGTRCARCTNGCRAANARRC-3´	
EH3	5´-TGTAAAACGACGGCCAGTRCARCANGCRAANARNG-3´	
EH4	5´-TGTAAAACGACGGCCAGTRCARCANGCRAANARRC-3´	
Φapra Fw	5´-AAAGTTCTGCTATGTGGCGCGGTATTATCCCGT GTTGACGGGTTCATGTGCAGCTCCATC-3´	Nur der rot markierte Bereich bindet an das Ursprungs-templat.
Φapra Rev	5´-CGAGACCCACGCTCACCGGCTCCAGATTTATCAG CAATAACAGGCTTCCCGGGTGTCTC-3´	

NVP1	5´-GAYTGGWSNYTNTGGGAG-3´	
NVP2	5´-TGGGAGTGCTGTAGCACGGGAA-3´	Degenerierte Primer für *nested*-PCR.
NHP1	5´-RCARCANGCRAANARRCTTCC-3´	
NHP2	5´-CTTCCCGTGCTACAGCACTCCC-3´	
rLAT3r	5´-CACGGTACCTAGACTAGTGACCAAGTGCGCCGGTC-3´	
Toyo Modul 5 5	5´-AAGCTAGCCCGGACACCCGT-3´	Unspezifische Primer.
16SrDNA ActinoFw	5´-GCTTAACACATGCAAGTCGAGCGGAAAGG-3´	Primer zur Kontrolle auf genomische DNA von *Actinomadura namibiensis*.
16SrDNA ActinoRev	5´-CGTTGCTGATCTGCGATTACTAGCGACTC-3´	
wABL1	5´-GGSTCSGGSAAGTCSACSC-3´	Degenerierte Primer zur Identifizierung von Lantibiotika-ABC-Transportern.
wABL3	5´-GGSTCSGGSAAGAGCACSC-3´	
wABL2	5´-GGSAGCGGSAAGTCSACSC-3´	
wABL4	5´-GGSAGCGGSAAGAGCACSC-3´	
wACinT	5´-GGSGCSGGSAAGACSACSAC-3´	
wB7B1	5´-CTGCCACTGSCCSCCSGASAG-3´	
wB7B2	5´-CTGCCACTGSCCSCCGCTSAG-3´	
wB7L1	5´-CTGGATCTGSCCSCCSGASAG-3´	
wB7L2	5´-CTGGATCTGSCCSCCGCTSAG-3´	
wB7CinT1	5´-CGCTGCATSCCSCCSGAG-3´	
wB7CinT2	5´-CGCTGCATSCCSCCGCTG-3´	
wBallesL	5´-SGTSGGRTCGTCSAGGAASAG-3´	
wBallesI	5´-SGTSGGRTCGTCSAGSAGSAG-3´	
wBCinT	5´-SGTSGGRTCGTCSAGGATSAG-3´	
V1	5´-GACTGGTCSCTSTGGGARTCSTGC-3´	Degenerierte Primer und Sonden aus dem 5´-Bereich der Primärsequenz von Labyrinthopeptin A2.
V2	5´-GACTGGWGCCTSTGGGARTCSTGC-3´	
V3	5´-GACTGGTCSCTSTGGGARWGCTGC-3´	
V4	5´-GACTGGWGCCTSTGGGARWGCTGC-3´	
H1	5´-ACSGGSTCSCTSTTCGCSTCSTGC-3´	Degenerierte Primer und Sonden aus dem 3´-Bereich der Primärsequenz von Labyrinthopeptin A2.
H2	5´-ACSGGSWGCCTSTTCGCSTCSTGC-3´	
H3	5´-ACSGGSTCSCTSTTCGCSWGCTGC-3´	
H4	5´-ACSGGSWGCCTSTTCGCSWGCTGC-3´	

A2-FwP1	5´-ATGGACCTCGCCACGGGCTC-3´	Primer zu Herstellung der homologen Sonde für das Filterscreening.
A2-M2 fw	5´-ATGGCGTCGATCCTGGAAC-3´	Primer zur heterologen Expression des LabA2-Präpropeptids.
A2-Stop rev	5´-TCAGCAGCAGGCGAACAGG-3´	
LabKC fw	5´-ATGGATCTGCGGTACCACGCCTATG-3´	Primer zur heterologen Expression des LabKC-Proteins.
LabKC-Stop rev	5´-TGCCATGGTGGATGTTCACCTCCTC-3´	
lKoFw1	5´-GTTCGTTCGACGGACCAATG-3´	Primer zur Untersuchung der Integrität des Cosmids 1104.
lKoRev1	5´-CCTGCTCGACGCAGTATTTG-3´	
lKoFw2	5´-CGCAGGACGAACGGTTTCAG-3´	
lKoRev2	5´-CCATGGGACTGTCACCTCTC-3´	
lKoFw3	5´-CATCCACCATGGCATCCATC-3´	
lKoRev3	5´-AGCAGCCGGGAGAGCAGCAG-3´	
lKoFw4	5´-ACTACCGGGCGATGTTCGAG-3´	
lKoRev4	5´-AGATGAAGCGGGCGATCGAG-3´	
lKoFw5	5´-AGGAGACGATCGCCTGGTTC-3´	
lKoRev5	5´-AACATGGCCAAGGCCCTGAC-3´	
lKoFw6	5´-AGGAGGTACCGCTGAACGAC-3´	
lKoRev6	5´-GTCCCGGACAGGTCATCAAG-3´	
A2 Ser2 ->Ala Fw	5´-GCGAGAACCGCTCCGACTGGGCCCTGTGGGAGTGCTGTAG-3´	Mutageneseprimer.
A2 Ser2 ->Ala Rev	5´-CTACAGCACTCCCACAGGGCCCAGTCGGAGCGGTTCTCGC-3´	
A2 Ser4 ->Ala Fw	5´-GGAGTGCTGTAGCACGGGAGCCCTGTTCGCCTGCTGCTG-3´	
A2 Ser4 ->Ala Rev	5´-CAGCAGCAGGCGAACAGGGCTCCCGTGCTACAGCACTCC-3´	
ΦintKon-trolle Fw	5´-GTCGCCGACATGACACAAG-3´	Kontrollprimer zur Untersuchung auf erfolgreiche Rekombination.
ΦintKon-trolle Rev	5´-TTCGCCCACGTGATCGAAG-3´	

attL	5´-GTCGACGATGTAGGTCACGG-3´	Primer zur Untersuchung der korrekten Integration von Cosmid M1104 in das Genom von *S. lividans*.
attR	5´-GTCGACATGCCCGCCGTGAC-3´	
cattL	5´-CGCGCTGAAGGTTCTTCGTG-3´	
cattR	5´-CCTATCCTGCCCGGCTGACG-3´	

Legende:
M = A oder C; R = A oder G; W = A oder T; S = C der G; Y = C oder T; K = G oder T; B = nicht A; D = nicht C; H = nicht G; V = nicht T; N = alle

4 Methoden

4.1 Mikrobiologische Methoden

4.1.1 Lagerung und Kultivierung von *Actinomadura namibiensis*

Der Stamm *Actinomadura namibiensis* HAG 010767 wurde von der Firma Sanofi-Aventis, Frankfurt, auf einer KM4-Platte erhalten. Von dieser Platte ausgehend wurden zunächst Kryokulturen für die dauerhafte Lagerung bei -80°C hergestellt. Dazu wurden zum einen Agarblöckchen der erhaltenen KM4-Platte mit 15 % Glycerin sowie zum anderen eine 72 h alte Flüssigkultur (Medium 5294) mit Kryo-Lösung (30 % Glycerin) im Verhältnis 1:1 versetzt. Mit diesen Kryokulturen (200 µl) konnten dann Vorkulturen (20 mL) in Erlenmeyerkolben (100 mL) mit drei Schikanen und einer Spirale beimpft werden. Hauptkulturen (100 mL) wurden mit der 72 h alten Vorkultur (2 %) beimpft und in Erlenmeyerkolben (500 mL) mit drei Schikanen und einer Spirale geschüttelt. Die Inkubation der Kulturen erfolgte bei 28°C und 120 rpm. Für Vorkulturen wurde unabhängig vom weiteren Verwendungszweck das Medium 5294 verwendet. Für die Isolierung genomischer DNA wurde CRM-Medium mit Glycin (0,5 %) für die Hauptkultur eingesetzt. Für die Produktion der Labyrinthopeptine A1-A3 erfolgte die Kultivierung analog zur Vorkultur in Medium 5294 für 168 h.

4.1.2 Lagerung und Kultivierung von *Escherichia coli*

Die Kultivierung von *E. coli*-Kulturen erfolgte grundsätzlich, wenn nicht gesondert angegeben, bei 37°C und 150 rpm in TY-Medium, wenn möglich unter Selektionsdruck eines oder mehrerer Antibiotika. Für Vorkulturen und Plasmidisolierungen im analytischen Maßstab wurden Kulturen (5 mL) mit Hilfe von sterilen Zahnstochern mit Einzelkolonien beimpft und für 16 h kultiviert. Größere Kulturvolumina wurden zu je 100 mL in Erlenmeyerkolben (500 mL) mit drei Schikanen geschüttelt. Die Hauptkulturen wurden mit einer 16 h alten Vorkultur (1 %) beimpft. Die Kontrolle der Wachstumsphase erfolgte photometrisch mit der Messung der optischen Dichte (OD) je nach Protokoll bei einer Wellenlänge von λ = 578 nm oder λ = 600 nm (OD_{578}/OD_{600}). Zur Stammhaltung wurden die Kulturen auf TY-Festmedium ausplattiert, nach 16 h Inkubation bei 37°C zum Schutz mit Parafilm verschlossen und dann bei 4°C im Kühlraum bis zu 2 Monate gelagert. Zur längerfristigen Lagerung wurden Flüssigkulturen mit Kryo-Lösung (30 % Glycerin) im Verhältnis 1:1 versetzt und bei -80°C eingefroren.

4.1.3 Lagerung und Kultivierung von *Streptomyces lividans*

Der Stamm *Streptomyces lividans* wurde von der Firma Combinature Biopharm AG, Berlin, auf Platte erhalten. Von dieser Platte ausgehend wurden zunächst Kryokulturen für die dauerhafte Lagerung bei -80°C hergestellt. Dazu wurden zum einen Agarblöckchen mit 15 % Glycerin sowie zum anderen eine 72 h alte Flüssigkultur mit Kryo-Lösung (30 % Glycerin) im Verhältnis 1:1 versetzt. Mit diesen Kryokulturen (200 µl) konnten dann Vorkulturen (20 mL) in Erlenmeyerkolben (100 mL) mit drei Schikanen und einer Spirale beimpft werden. Hauptkulturen (100 mL) wurden mit der 72 h alten Vorkultur (1 %) beimpft und in Erlenmeyerkolben (500 mL) mit drei Schikanen und einer Spirale geschüttelt. Die Inkubation der Kulturen erfolgte bei 28°C und 120 rpm. Für Vorkulturen wurden abhängig vom jeweiligen Verwendungszweck unterschiedliche Medien verwendet. So wurde für die Protoplastierung in YEME-Medium, für die Isolierung genomischer DNA in R5-Medium und für die heterologe Expression in R5-Medium und Medium 5294 kultiviert. Für die Protoplastierung und Isolierung genomischer DNA wurde YEME-Medium für die Hauptkulturen eingesetzt. Für die heterologe Expression erfolgte die Kultivierung analog zur Vorkultur in R5-Medium oder Medium 5294. Die Kultivierung der *S. lividans*-Klone (K1-4) erfolgte stets unter Selektion auf das integrierte Cosmid M1104 mit Kanamycin und Apramycin.

4.1.4 Produktionskontrolle der Labyrinthopeptine A1/A2 aus Flüssigkulturen

Zur Kontrolle der Produktion der Labyrinthopeptine aus Flüssigkulturen war keine Vorreinigung notwendig. Durch Zentrifugation (20 min; 21000 g) wurde ein klarer Kulturüberstand erhalten, der dann direkt zur Analyse mittels HPLC-ESI-MS eingesetzt werden konnte. Dabei wurde der in Tabelle 2 beschriebene Standardgradient mit einer Luna 3u C18(2) 100Å 50 x 1 mm Säule der Firma Phenomenex, Aschaffenburg verwendet. Als Laufmittel A wurde Wasser mit 0,1 % HCOOH und als Laufmittel B Acetonitril mit 0,1 % HCOOH verwendet.

Tabelle 2: Standardgradient zur Produktionskontrolle mittels HPLC-ESI-MS.

Zeit [min]	% Laufmittel A	% Laufmittel B	Flußrate [µl/min]
0,0	90	10	60
6,0	0	100	60
8,0	0	100	60
8,1	90	10	60
11,0	90	10	60

4.2 Molekularbiologische Methoden

4.2.1 DNA-Transfer

Man kann bei der Übertragung und Aufnahme fremder DNA in bakterielle Zellen prinzipiell zwischen Transduktion, Konjugation und Transformation unterscheiden. Während die Transduktion die Übertragung von genetischem Material von Viren auf Bakterien und die Konjugation den Austausch zwischen Bakterien über eine sich ausbildende Plasmabrücke beschreibt, versteht man unter der Transformation die Aufnahme freier DNA in ein Bakterium. Die Aufnahme freier DNA ist mit wenigen Ausnahmen nur nach einer Vorbehandlung der Empfängerzelle möglich. Eine bakterielle Zelle, die zur Transformation befähigt ist, wird auch als kompetente Zelle bezeichnet. Für *E. coli* gibt es mehrere Methoden der Kompetentmachung. Neben der klassischen $CaCl_2$-Methode werden auch elektrokompetente Zellen zur Transformation herangezogen. Bei Actinomyceten wird häufig eine durch Polyethylenglykol (PEG) induzierte DNA-Aufnahme von Protoplasten herangezogen, welche auch als Protoplastentransformation bezeichnet wird.

4.2.1.1 Herstellung chemisch kompetenter *E. coli* nach der $CaCl_2$-Methode

Die Herstellung chemisch kompetenter Zellen basiert auf der Behandlung von Bakterienzellen mit einer hohen Konzentration an divalenten Kationen bei Temperaturen nahe dem Gefrierpunkt. Am häufigsten werden dazu Ca^{2+}-Kationen herangezogen.[51] Die Hintergründe der dabei ablaufenden Vorgänge konnten jedoch bis heute nicht vollständig aufgeklärt werden.

Zur Herstellung kalziumkompetenter *E. coli* wurde in dieser Arbeit TY-Medium (100 mL) mit einer 16 h alten Vorkultur (1 %) des gewünschten Stammes beimpft und bis zu einer $OD_{578} \approx$ 0,5 – 0,8 kultiviert (37°C). Die Kultur wurde durch Zentrifugation (10 min; 4200 g) bei 4°C pelletiert und der Überstand quantitativ verworfen. Alle weiteren Schritte erfolgten auf Eis und mit gut vorgekühlten Lösungen. Das Pellet wurde durch vorsichtiges Triturieren in steriler $CaCl_2$-Lösung (5 mL; 100 mM) resuspendiert und erneut pelletiert. Nach erneuter Resuspension in steriler $CaCl_2$-Lösung (40 mL; 50 mM) schloss sich erneut eine Zentrifugation (10 min; 4200 g) bei 4°C an. Das Pellet wurde schließlich in steriler $CaCl_2$-Lösung (1,85 mL; 100 mM) aufgenommen und für 16 h auf Eis inkubiert, um so die Kompetenz der Zellen zu verbessern und die Transformationseffizienz zu erhöhen.

Die kompetenten *E. coli* wurden entweder direkt zur Transformation eingesetzt oder nach Zugabe von Glycerin-Lösung (60 %; ¼-Volumen der *E. coli*-Suspension) zur Lagerung bei -80°C eingefroren.

4.2.1.2 Transformation in chemisch kompetente *E. coli*

Die Transformation in chemisch kompetente *E. coli* wurde in dieser Arbeit zur Amplifikation und Expression unterschiedlichster Plasmide und Cosmide herangezogen. Es wurde dabei das Protokoll von Sambrook und Russell angewendet.[52]

Die zu transformierende Fremd-DNA und die kompetenten *E. coli*-Zellen wurden auf Eis aufgetaut bzw. gelagert. Einem Aliquot kompetenter Zellen (200 µl) wurde dann das jeweils zu transformierende Plasmid bzw. Cosmid (2 µl) oder der Ligationsansatz (5 µl) zupipettiert. Nach Inkubation auf Eis (30 min) schloss sich ein Hitzeschock (90 s) bei 42°C an. Der Transfer der Fremd-DNA findet bei 0°C statt und wird durch den Hitzeschock beendet. Nach Beendigung des Hitzeschocks wurde das Reaktionsgefäß sofort auf Eis überführt und TY-Medium (1 mL) zugegeben. Bei Transformationen, die Proteinexpressionen betrafen, wurde stattdessen SOC-Medium (250 µl) verwendet. Zur Entwicklung der eingebrachten Resistenz, welche eine Selektion auf positive Klone erlaubte, folgte eine Inkubation ohne Selektionsdruck (40 min) bei 37°C. Ansätze mit Cosmiden erforderten eine Inkubation bei 27°C, wobei dann die Inkubationsdauer ohne Selektionsdruck auf bis zu 3 h verlängert wurde. Nach einem Zentrifugationsschritt (1 min; 21000 g) wurde der Überstand nahezu quantitativ entfernt, das Pellet im restlichen Volumen (ca. 100 µl) resuspendiert und auf einer TY-Platte mit dem entsprechenden Antibiotikum mit einer sterilen Glaspipette (1 mL) ausplattiert. Die Platten wurden anschließend bei 37°C, bei Cosmidtransformationen bei 28°C im Brutschrank inkubiert (16 h), bis das Wachstum einzelner Kolonien beobachtet werden konnte.

4.2.1.3 Transformation in *E. coli* mittels Elektroporation

Eine weitere und im Vergleich zur Transformation in chemisch kompetente Zellen effizientere und schnellere Methode ist die Elektroporation, welche in gleichem Maße für prokaryotische und eukaryotische Zellen geeignet ist.[53,54] Durch die schnelle Entladung eines Kondensators wird ein kurzer elektrischer Puls hoher Spannung erzeugt, der zur Ausbildung kleiner Löcher in den Zellmembranen führt. Diese Poreninduktion ermöglicht neben dem Verlust der Semipermeabilität der Zellmembran und der damit einhergehenden Freisetzung intrazellulärer Bestandteile auch die Aufnahme freier DNA (Transformation). Es kann mit dieser Methode eine Effizienz von bis zu 80 % transformierter Zellen erhalten werden, wobei darauf geachtet werden muss, dass die Bakterien- und DNA-Lösungen möglichst salzarm eingesetzt werden, um einen Kurzschluss zu vermeiden. Die Elektroporation wurde in der vorliegenden Arbeit bei der Modifikation des Cosmids 1104 mit Hilfe des Redirect-Systems (s. Abschnitt 5.4) für die heterologe Expression in *S. lividans* angewendet.

SOB-Medium (9,9 mL) wurde mit einer Mg^{2+}-Lösung (100 µl) und den entsprechenden Antibiotika-Stammlösungen (10 µl) komplettiert und anschließend mit einer 16 h alten Vorkultur beimpft (1 %). Die Kulturen wurden bis zu einer $OD_{600} \approx 0{,}4$ kultiviert (30°C und 200 rpm). Durch Zentrifugation (10 min; 4200 g) bei 4°C wurden die Zellen vollständig pelletiert und anschließend zweimal mit eiskalter, steriler Glycerin-Lösung (10 mL/ 5 mL; 10 %) gewaschen. Nach dem letzten Abdekantieren wurde das Pellet in der verbleibenden Flüssigkeit (ca. 50 µl) resuspendiert und zur Elektroporation mit der DNA (1 µl) in vorgekühlten Küvetten gemischt. Für den Puls wurden folgende Bedingungen gewählt: 200 Ω, 25 µF und 2,5 kV. Nach dem Puls wurde sofort vorgekühltes LB-Medium (1 mL) in die Küvette gegeben. Zur Entwicklung der eingebrachten Resistenz folgte eine Inkubation ohne Selektionsdruck (3 h) bei 27°C. Nach einem Zentrifugationsschritt (1 min; 21000 g) wurde der Überstand nahezu quantitativ entfernt, das Pellet im restlichen Volumen (ca. 100 µl) resuspendiert und auf einer LB-Platte mit dem entsprechenden Antibiotikum ausplattiert. Die Platten wurden anschließend bei 27°C im Brutschrank inkubiert, bis das Wachstum einzelner Kolonien beobachtet werden konnte.

4.2.1.4 Protoplastentransformation in Actinomyceten

Für die Transformation von DNA in Actinomyceten konnte bisher keine vergleichsweise einfache Methode zur Herstellung chemisch kompetenter Zellen bei *E. coli* gefunden werden. Es wurde aber beobachtet, dass Protoplasten von Actinomyceten PEG-induziert Fremd-DNA mit hoher Ausbeute aufnehmen können.[55,56] Die Protoplastentransformation wurde in der vorliegenden Arbeit für die Transformation von Cosmiden in *Streptomyces lividans* für eine heterologe Expression verwendet. Die Herstellung der Protoplasten erfolgte nach einem Standardprotokoll, das für viele verschiedene Actinomyceten-Stämme Anwendung findet.[57]
Dazu wurde zunächst für jeden Ansatz eine Vorkultur mit YEME-Medium (25 mL) hergestellt und für 72 h geschüttelt (30°C, 120 rpm). Mit diesen Vorkulturen wurde je eine Hauptkultur (25 mL YEME-Medium) beimpft (250 µl) und 62 h kultiviert (30°C, 120 rpm). Durch Zentrifugation (10 min, 11500 g) wurden die Zellen pelletiert und anschließend zweimal mit Sucrose-Lösung (15 mL) gewaschen. Die Mycelpellets wurden dann in Lysozym-Lösung (4 mL) resuspendiert und für 30 min im Heizblock (30°C) unter ständigem Schütteln (750 rpm) inkubiert. Der Fortschritt der Protoplastierung wurde mikroskopisch kontrolliert. Nach dreimaligem Triturieren mit einer Glaspipette (5 mL) wurde für weitere 15 min inkubiert. Zur Entfernung von Mycelresten wurden die Protoplasten durch eine sterile, mit Watte gepackte Einwegspritze gedrückt und anschließend erneut pelletiert (7 min, 11500 g). Die Pellets wurden nach dem Abdekantieren zunächst möglichst ohne Blasenbildung im verbleibenden Flüssigkeitsrückstand resuspendiert und anschließend mit P-Puffer (1 mL) versetzt. Die

erhaltenen Protoplasten-Lösungen wurden dann auf Eis bei -80°C für die längere Lagerung langsam eingefroren.

Am Tag der Durchführung der Protoplastentransformation wurden die eingefrorenen Protoplasten schnell aufgetaut und anschließend direkt pelletiert (7 min, 4200 g). Die Pellets wurden nach dem Abdekantieren im verbleibenden Flüssigkeitsrückstand (ca. 50 µl) vorsichtig resuspendiert und dann mit der zu transformierenden Cosmid-DNA (5 µl) versetzt. Nach sofortiger Zugabe von sterilem P-PEG-Puffer (500 µl) wurde einmal trituriert und kurz (1 min) inkubiert. Nach der Zugabe von P-Puffer (5 mL) wurde zentrifugiert (5 min, 4200 g), das Pellet nach dem Abdekantieren zunächst im verbleibenden Flüssigkeitsrückstand und dann in P-Puffer (1 mL) resuspendiert. Die so erhaltenen Protoplastensuspensionen wurden dann zu gleichen Teilen (100 µl) auf zehn R2YE-Platten pro Ansatz ausplattiert. Die Inkubation der Platten erfolgte zunächst für 16 h bei 28 °C. Dann wurden die Platten zur Selektion auf positive Klone mit NB-Weichagar (2 mL) überschichtet, wobei die Antibiotika-Lösungen zehnfach konzentriert eingesetzt wurden. Die Platten wurden so lange bei 28°C kultiviert (ca. 120 h) bis einzelne Kolonien durch den NB-Weichagar durchgebrochen waren und für weitere Experimente überimpft werden konnten.

4.2.2 Analyse, Isolierung und Reinigung von DNA

4.2.2.1 Agarosegelelektrophorese

Die Gelelektrophorese ist eine sehr wichtige Methode zur Analyse von DNA. Als Trägermaterial wird vorwiegend Agarose, ein Polysaccharid, verwendet. Die Phosphatgruppen der DNA liegen innerhalb eines sehr großen pH-Bereichs negativ geladen vor, so dass DNA eine konstante Ladungsdichte, also ein konstantes Molekulargewichts-zu-Ladungs-Verhältnis besitzt. Deshalb ist die Wanderungsgeschwindigkeit im elektrischen Feld in freier Lösung für alle DNA-Moleküle gleich. In einer festen Gelmatrix hingegen treten Unterschiede in der Wanderungsgeschwindigkeit auf, abhängig von der Molekülgröße und der Form der DNA (superhelikal, relaxiert, doppelsträngig-linear oder einzelsträngig). Bei doppelsträngig-linearer DNA besteht über einen weiten Größenbereich eine empirisch beobachtete, lineare Abhängigkeit zwischen dem Logarithmus (log_{10}) der DNA-Fragmentlänge und der relativen Wanderungsdistanz im Gel. Superhelikale DNA wandert in der Regel schneller, relaxierte DNA dagegen wesentlich langsamer als doppelsträngig-lineare DNA. In der Praxis verwendet man Längenstandards zur Charakterisierung von Fragmentgrößen. Um Diffusionsverluste zu vermeiden, wird die DNA mit Hilfe von DNA-Probenauftragspuffer in die Geltaschen aufgetragen, die neben Saccharose zur Erhöhung der Dichte auch Farbstoffe wie beispielsweise Bromphenolblau enthalten. Diese wandern während der Elektrophorese aufgrund ihrer geringen Molekülgröße als erste in Richtung Anode und bieten somit einen Anhaltspunkt für die Wanderung der DNA. Zur Detektion der DNA wird in den meisten Fällen Ethidiumbromid, für spezielle Anwendungen auch SYBRGreen, verwendet. Diese Fluoreszenzfarbstoffe können aufgrund ihrer planaren Struktur in DNA interkalieren.

In der vorliegenden Arbeit wurden Gele unterschiedlicher Agarosekonzentrationen eingesetzt. Für Fragmente unter 1 kbp wurden 2%ige, für Fragmente über 1 kbp 1%ige und für Cosmidverdaus 0,8%ige Agarosegele verwendet, wobei TAE-Elektrophoresepuffer als Gel- und Laufpuffer diente. Je nach Trennproblem wurde die Elektrophorese bei verschiedenen Spannungen und unterschiedlich langen Laufzeiten durchgeführt, wobei die Kombinationen 80 V / 40 min, 70V / 60 min und 60 V / 90 min häufig verwendet wurden. Aus Laborsicherheitsgründen wurden alle mit dieser Arbeit verbundenen Agarosegele erst nach der Elektrophorese angefärbt und unter UV-Licht einer Geldokumentationsanlage analysiert bzw. fotografiert.

4.2.2.2 *Southern-Blot*

Mit der von E. Southern 1975 zum ersten Mal beschriebenen Methode[58] ist es möglich, gelelektrophoretisch getrennte DNA-Fragmente nach deren Transfer und Fixierung auf eine Membran spezifisch mit markierten Sonden nachzuweisen. Die Methode wurde in dieser Arbeit für die Identifizierung des Labyrinthopeptin-Genclusters und für den Nachweis der Integration des Cosmids M1104 in das Genom von *S. lividans* eingesetzt.

Die hochmolekularen genomischen DNA-Präparationen wurden dafür mit geeigneten Restriktionsenzymen verdaut und gelelektrophoretisch getrennt. Nach kurzer Färbung und Dokumentation wurde das Gel für den Transfer mittels Saugblot-Verfahren vorbereitet. Aufgrund der Herkunft der Proben und der daraus resultierenden Fragmentgrößen von teilweise weit über 5 kbp wurde das Gel für 20 min in Southern I auf dem Taumeltisch geschüttelt. Die Inkubation in HCl-Lösung führte zur Depurinierung und in Folge dessen zu Strangbrüchen, was einen effizienten Blotvorgang größerer DNA-Fragmente ermöglichte. Nach der Denaturierung der DNA-Fragmente in Southern II (30 min) wurde das Gel durch Inkubation (30 min) in Southern III wieder neutralisiert. Anschließend wurde der Blot wie in Abbildung 16 gezeigt aufgebaut und für 16 h geblottet.

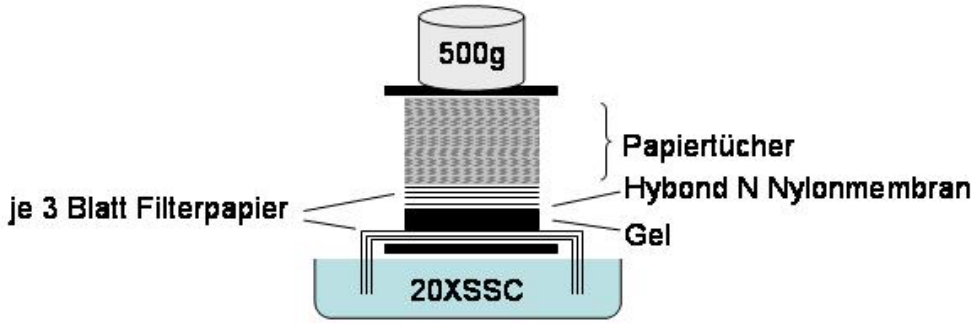


Abbildung 16: Aufbau der Apparatur zum Blotten von DNA aus einem Agarosegel nach dem Saugblot-Verfahren. Eine Schicht aus drei Streifen Filterpapier wurde auf einer Glasplatte so angebracht, dass sie auf beiden Seiten in das Vorratsgefäß des 20xSSC-Blotpuffers reichte. Das Gel wurde aufgelegt, so dass die Taschenöffnungen auf das untere Filterpapier zeigten und mit der Membran bedeckt. Auf diese Weise entsprach die Reihenfolge der Proben auf der Membran derjenigen auf dem Gel. Nach der Membran wurden wiederum drei Filterpapierstreifen in Gelgröße und ein großer Stapel Papiertücher aufgelegt. Zur gleichmäßigen Belastung wurde eine Glasplatte benutzt und beschwert. Beim Aufbau des Blots wurde darauf geachtet, dass keine Luftblasen und kein Bypass für den Blotpuffer eingebaut wurden.

Die kovalente Fixierung der DNA auf der Membran erfolgte durch eine kurze (2 min) Bestrahlung mit UV-Licht, wobei Thymidinreste der DNA und Aminogruppen der Nylonmembran kovalent verbunden wurden. Im Anschluss daran folgte die Hybridisierung mit Sonden und die Detektion dieser Hybridisierungsereignisse, was allerdings in der vorliegenden Arbeit auf zwei unterschiedliche Arten durchgeführt wurde. Für den Nachweis der Integration des Cosmids M1104 in *S. lividans* wurden DIG-markierte Sonden eingesetzt, deren Herstellung, Hybridisierung und Detektion im folgenden Abschnitt (s. Abschnitt 4.2.2.3)

beschrieben wird. Die Vorgehensweise bei der Identifizierung des Labyrinthopeptin-Genclusters soll dagegen hier weiter beschrieben werden. Als Sonden wurden kurze, degenerierte, [^{32}P]-markierte Oligonukleotide eingesetzt, die von der Firma Hartmann-Analytik, Braunschweig, bezogen wurden. Für die Hybridisierung wurde die Membran in eine Rollröhre überführt, mit Hybridisierungs-Lösung (50 mL) für 30 min bei 40°C prähybridisiert und anschließend mit den zuvor bei 90°C (10 min) denaturierten Sonden (in 25 mL Hybridisierungs-Lösung) hybridisiert. Die Hybridisierung erfolgte aufgrund der sehr kurzen Länge der Oligonukleotide ebenfalls bei 40°C im Rollofen. Nach 16 h wurden die Sonden entfernt und kurz mit Wasch-Lösung (25 mL) nachgespült. Anschließend wurde die Membran zweimal mit Wasch-Lösung (75 mL) bei 40°C im Rollofen gewaschen, bis die Wasch-Lösung keine Radioaktivität mehr zeigte. Die Membran wurde dann in Frischhaltefolie eingeschlagen, in einer Fotokassette fixiert und bei -80°C autoradiografiert. Die Inkubationsdauer wurde dem Alter bzw. der Aktivität der Sonde entsprechend gewählt (2 - 10 Tage).

4.2.2.3 Herstellung DIG-markierter Sonden und Nachweis der Integration des Cosmids M1104 in das Genom von *S. lividans*

Für die Detektion des in das Genom von *S. lividans* integrierten Cosmids M1104 mittels *Southern-Blot* mussten DIG-markierte Sonden hergestellt werden, was üblicherweise über PCR erfolgen kann. In diesem speziellen Fall sollte jedoch ein Gemisch kleiner, DIG-markierter Fragmente von Cosmid M1104 als Sonde verwendet werden, für deren Herstellung sich ein anderes Verfahren besser eignete. Dazu wurde das *DIG-High Prime DNA labeling and detection Start-KitI* der Firma Roche, Mannheim, verwendet, das auf einem *random primed oligo-labeling*-Mechanismus basiert. Voraussetzung war der Verdau des Cosmids M1104 mit MlyI, was Fragmente geeigneter Größe (800-1000 bp) ergab. Nach deren Denaturierung zu Einzelsträngen wurden DIG-markierte Nukleotide (DIG-11-dUTP), Klenow-Enzym und randomisierte kurze Oligonukleotide zugegeben. Dies führte zum Aufbau der gewünschten DIG-markierten Doppelstränge, die die gesamte Sequenz von Cosmid M1104 abdeckten und nach Denaturierung als Sonde eingesetzt werden konnten. Die Detektion erfolgte durch einen im Kit enthaltenen Antikörper und dem damit gekoppelten Farbassay. Die verwendeten Lösungen wurden dem Herstellerprotokoll entsprechend angesetzt. Puffer, die dort nicht vermerkt sind, entsprechen den in Abschnitt 3.4.6 angegebenen Lösungen. Im Folgenden soll die Vorgehensweise genau beschrieben werden. Aufgrund der bisher beobachteten Schwierigkeiten bei Verdauansätzen von Actinomadura-DNA wurde die Cosmid M1104-DNA für den MlyI-Verdau (ca. 8 µg) im Überschuss eingesetzt und der Verdau länger inkubiert (7 h). Nach Zugabe von H_2O_{dd} (4 µl) wurden die DNA-Fragmente bei 100°C denaturiert (10 min) und anschließend sofort auf Eis gestellt.

Nach der Zugabe von DIG-High Prime-Lösung (4 µl; Reaktionsmischung mit Klenow, Oligos und Nukleotiden) wurde der Ansatz bei 37°C inkubiert (16 h). Anschließend wurde der Gehalt an markierter DNA durch einen Dot-Blot ermittelt. Dazu wurde eine Verdünnungsreihe des Markierungsansatzes und einer markierten-DNA bestimmter Konzentration auf eine Blotmembran aufgetragen und wie der spätere Blot entwickelt. Der Vergleich zeigte dann eine Konzentration an markierter DNA im Markierungsansatz von ca. 400 ng/µl. Das so hergestellte DIG-markierte Sondengemisch wurde im Folgenden für den *Southern-Blot* eingesetzt.

Die genomische DNA der zu untersuchenden *S. lividans*-Klone, des *S. lividans*-Wildtyps (Negativkontrolle) und die DNA des Cosmids M1104 (Positivkontrolle) wurden zunächst mit BamHI verdaut und weiterhin dem in Abschnitt 4.2.2.2 beschriebenen Blotverfahren unterzogen. Nach der kovalenten Fixierung der geblotteten DNA auf der Membran wurde diese mit vorgewärmtem (65°C) *DIG Easy Hyb*-Puffer (30 ml) im Rollofen inkubiert (65°C, 30 min). Die DIG-markierte Sonde wurde denaturiert (100°C, 5 min), kurz auf Eis gestellt und dann auf eine Endkonzentration von 25 ng/ml in vorgewärmtem *DIG Easy Hyb*-Puffer (65°C, 30 ml) verdünnt. Die Prähybridisierungs-Lösung wurde dann gegen die Sonden-Lösung ausgetauscht und im Rollofen inkubiert (65°C, 16 h). Nach dem quantitativen Entfernen der Sonden-Lösung wurde zunächst zweimal mit 2xSSC-Puffer (50 mL) für 5 min bei RT gewaschen und anschließend zweimal mit 0,5xSSC-Puffer (50 mL) für 15 min bei 68°C. Vor der Inkubation der Membran mit dem Anti-Digoxigenin AP-Antikörper wurde erst für 5 min mit Waschpuffer (50 mL) vorinkubiert und dann durch Zugabe von Blockpuffer (100 mL) für 30 min bei RT zur Verminderung der unspezifischen Bindung geblockt. Die Inkubation des Antikörpers erfolgte dann mit der Antikörper-Lösung (30 mL) für 30 min bei RT. Nach der Inkubation des Antikörpers wurden die Reste der Lösung durch zweimaliges Waschen mit Waschpuffer (100 mL) für 15 min bei RT entfernt. Die Membran wurde dann durch Zugabe von Detektionspuffer (30 mL) kurz auf die für die Färbereaktion benötigten Bedingungen umgepuffert und der Färbevorgang durch Zugabe der Färbe-Lösung (30 mL) gestartet. Der Färbevorgang wurde bis zu einer ausreichenden Signalstärke unter Lichtausschluss durchgeführt und die Membran dann mehrfach mit H_2O_d gewaschen, ausgewertet und archiviert.

4.2.2.4 Screening auf Hybridisierungsfilter der Cosmidbank

Für die Identifizierung des Biosynthese-Genclusters wurden wie in Abschnitt 5.1.2 beschrieben Hybridisierungsfilter der Cosmidbank von der Firma RZPD, Berlin, erstellt. Um diese nach Gencluster-tragenden Cosmidklonen durchsuchen zu können, musste zunächst eine Sonde abgeleitet werden. Für die spätere Detektion und Auswertung des Screenings wurde diese Sonde mittels einer PCR mit Digoxigenin markiert (s. Abschnitt 4.2.3.4).
Für die Hybridisierung und Detektion wurde das *DIG Nucleic Acid Detection-Kit* der Firma Roche, Mannheim, unter Verwendung eines optimierten Protokolls der Firma Combinature Biopharm AG, Berlin, eingesetzt (Lösungen s. Abschnitt 3.4.6). Vor der Hybridisierung mit der Sonde wurde der Hybridisierungsfilter zunächst für 4 h bei 65°C in Church-Puffer (100 mL) vorhybridisiert. Die aus dem Gel gereinigte DNA-Sonde (40 µl) wurde für 5 min auf 95°C erhitzt, um die DNA-Stränge zu denaturieren und anschließend sofort bis zur Hybridisierung auf Eis gestellt. Für die Hybridisierung wurde der Church-Puffer der Vorhybridisierung entfernt und die Sonde in neuem Church-Puffer (50 mL) auf den Filter gegeben. Die Hybridisierung erfolgte für 16 h bei 65°C im Rollofen. Nach dem quantitativen Entfernen der Sonden-Lösung wurde zunächst zweimal mit 2xSSC-Puffer (50 mL) für 15 min bei RT gewaschen und anschließend zweimal mit 0,5xSSC-Puffer (50 mL) für 15 min bei 65°C. Vor der Inkubation des Filters mit dem Anti-Digoxigenin AP-Antikörper wurde der Filter erst für 5 min mit Waschpuffer vorinkubiert und dann durch Zugabe von Blockpuffer (100 mL) für 30 min bei RT zur Verminderung der unspezifischen Bindung geblockt. Die Inkubation des Antikörpers erfolgte dann mit der Antikörper-Lösung (20 mL) für 30 min bei RT. Nach der Inkubation des Antikörpers wurden die Reste der Lösung durch zweimaliges Waschen mit Waschpuffer (100 mL) für 15 min bei RT entfernt. Der Filter wurde dann durch Zugabe von Detektionspuffer (20 mL) kurz auf die für die Färbereaktion benötigten Bedingungen umgepuffert und der Färbevorgang durch Zugabe der Färbe-Lösung (10,2 mL) gestartet. Der Färbevorgang wurde bis zu einer ausreichenden Signalstärke durchgeführt und der Filter dann mehrfach mit H_2O_d gewaschen, ausgewertet und archiviert.

4.2.2.5 Sequenzierung

Die in dieser Arbeit zahlreich durchgeführten Standardsequenzierungen wurden der Firma GATC, Konstanz, in Auftrag gegeben. Die Sequenzierung des Cosmids 1104 erfolgte durch die Firma Agowa GmbH, Berlin. Zur Auswertung der Sequenzierergebnisse wurde der Clonemanager herangezogen, wobei bei schlechten Rohdaten die Sequenzen zuvor mittels BioEdit (*freeware*) kontrolliert wurden.

4.2.2.6 Plasmidisolierung aus *E. coli* mittels *peqGOLD Plasmid Miniprep-Kit I*

Zur Aufreinigung von bis zu 20 µg hochreiner Plasmid-DNA aus Flüssigkulturen wurde in dieser Arbeit das *peqGOLD Plasmid Miniprep-Kit I* der Firma PEQLAB, Erlangen, verwendet. Mit der so isolierten Plasmid-DNA wurde eine Vielzahl von Folgeexperimenten durchgeführt, unter anderem Restriktionsanalysen, Sequenzierungen, Mutagenesen und PCR.

Das diesem Kit zugrunde liegende Prinzip beruht auf der Arbeit zur alkalischen Lyse von Bakterienzellen von Birnboim und Doly.[59] Dabei werden die Bakterienzellen durch Zugabe von SDS und NaOH lysiert, wobei sich ein alkalischer pH-Wert einstellt. Die Zugabe von RNase A dient der Entfernung der RNA. Bei den vorliegenden alkalischen Bedingungen kommt es zur vollständigen Denaturierung der DNA, wobei sich die beiden Stränge eines zirkulären Plasmids aufgrund ihrer Topologie nicht vollständig voneinander trennen. Durch Zugabe einer sauren Natriumacetat-Lösung wird neutralisiert, wodurch es zur Renaturierung der DNA kommt. Während sich Plasmid-DNA vollständig renaturieren kann, kommt es bei genomischer DNA zur unspezifischen Verknüpfung mehrerer Stränge, was zur Ausbildung eines DNA-Knäuels und zur Präzipitation führt. Die genomische DNA kann dann durch Zentrifugation entfernt werden, genauso wie die aufgrund der hohen Konzentration an Natriumacetat ausgefallenen Protein-SDS-Komplexe. Die Reinigung der Plasmid-DNA erfolgt dann aus dem salzhaltigen Überstand durch reversible Adsorption der DNA an Silikasäulchen. Die Elution erfolgt nach diversen Waschschritten mit geringeren Konzentrationen an chaotropen Salzen mit Wasser oder TE-Puffer. Die praktische Vorgehensweise wird im Folgenden beschrieben.

Abhängig von der Kopienzahl pro Zelle wurden unterschiedliche Volumina einer 16 h alten Kultur pelletiert (5 mL bei high-copy (hc) Plasmiden/ 10 mL bei low-copy (lc) Plasmiden). Der Überstand wurde quantitativ entfernt und die Pellets in mit RNase A komplettierter Lösung I (hc: 200 µl; lc: 500 µl) resuspendiert. Nach der Zugabe von Lösung II (hc: 200 µl; lc 500 µl) wurde durch Invertieren (6x) gemischt und bei RT für 2 min inkubiert. Durch Zugabe von Lösung III (hc: 280 µl; lc: 700 µl) bildete sich ein weißer Niederschlag, dessen Bildung durch Invertieren (10x) unterstützt wurde. Der weiße Niederschlag wurde durch Zentrifugation (10 min, 16000 g) abgetrennt und der Überstand quantitativ auf *HiBind-Miniprep*-Zentrifugensäulen aufgetragen. Nach der Zentrifugation (1 min, 16000 g) folgten zwei Waschschritte (HB-Puffer (500 µl); DNA-Waschpuffer (600 µl)). Um das die folgenden Schritte störende Ethanol des DNA-Waschpuffers quantitativ zu entfernen, wurde erneut zentrifugiert (3 min, 16000 g). Die Elution erfolgte mit vorgewärmtem (65°C) sterilem H_2O_{dd} (40 µl) nach kurzer Inkubation (3 min) auf der Säule durch Zentrifugation (1 min, 16000 g).

4.2.2.7 Isolierung genomischer DNA mittels *NucleoSpin® Tissue-Kit*

Für die Isolierung genomischer DNA stehen verschiedene Methoden zur Verfügung, wobei die Auswahl unter anderem vom späteren Verwendungszweck abhängig gemacht werden sollte. Die mittels *NucleoSpin® Tissue-Kit* generierten genomischen DNA-Isolierungen haben den Nachteil, dass aufgrund der Verwendung der Säulchen auch kleine Bruchstücke entstehen und wurden deshalb in dieser Arbeit nur als Templat für die PCR eingesetzt. Die praktische Vorgehensweise wird im Folgenden beschrieben.

Abhängig von der Dichte der eingesetzten Kultur wurden unterschiedliche Volumina (10 -15 mL) durch Zentrifugation (10 min, 4200 g) pelletiert. Abweichend vom Herstellerprotokoll wurde ein Waschschritt (1 mL Wasch-Lösung) eingeschoben. Nach erneuter Zentrifugation (10 min, 4200 g) wurde der Überstand quantitativ entfernt und das Pellet in Lyse-Lösung (ca. 500 µl; je nach Pelletvolumen teilweise auch mehr) resuspendiert. Diese Lösung wurde bei 37°C inkubiert bis das Zellpellet deutlich kleiner wurde oder völlig gelöst war (30 - 60 min). Nach der Zugabe von 25 µl Proteinase K-Lösung wurde der Ansatz bei 56°C im Thermoblock geschüttelt (3 h). Zur Entfernung der RNA wurde RNase A (40 µl, 10 mg/mL) zugegeben und kurz (5 min) bei RT inkubiert. Nach der Zugabe von 200 µl B3-Puffer wurde die Lösung durch mehrfaches Invertieren gut gemischt und dann auf 70°C erhitzt (10 min). Während der Inkubation wurde der Ansatz immer wieder durch Invertieren gemischt. Nach kurzer Zentrifugation (5 min; 11500 g) wurde der Überstand direkt auf Ethanol (abs.) (210 µl) pipettiert und sofort invertiert. Die gesamte Probe wurde dann auf eine *NucleoSpin® Tissue*-Säule aufgetragen und zentrifugiert (1 min, 7700 g). Bei sehr zähflüssigen Lösungen musste die Zentrifugation entsprechend verlängert werden. Die Säule wurde zweimal gewaschen (BW-Puffer (500 µl); B5-Puffer (600 µl)). Um das die folgenden Schritte störende Ethanol des B5-Waschpuffers quantitativ zu entfernen, wurde erneut zentrifugiert (3 min 7700 g). Die Elution der genomischen DNA erfolgte mit vorgewärmtem (65°C) BE-Puffer (100 µl) nach kurzer Inkubation (3 min) auf der Säule durch Zentrifugation (1 min, 7700 g).

4.2.2.8 Isolierung hochmolekularer genomischer DNA mittels Phenol/ Chloroform-Extraktion

Manche Applikationen wie z.B. die Herstellung von Cosmidbanken oder *Southern-Blot*-Analysen von genomischen Restriktionsverdaus benötigen hochmolekulare genomische DNA als Ausgangsmaterial. Da sich die Isolierungen mittels Säulchen-Kits aufgrund der auftretenden Scherung der DNA-Moleküle als nicht geeignet erwiesen, wurde in der vorliegenden Arbeit eine andere Methode für die Isolierung hochmolekularer DNA herangezogen. Das verwendete Protokoll der Combinature Biopharm AG, Berlin, basiert auf

der Reinigung von DNA mit Hilfe der Phenol/Chloroform-Extraktion und wurde für die in dieser Arbeit verwendeten Stämme optimiert.

Für eine Isolierung wurde eine 48 h alte Kultur (100 mL) zunächst durch Zentrifugation (10 min, 2000 g) in einer vorgekühlten Zentrifuge (4°C) pelletiert. Danach wurde das Pellet mit Wasch-Lösung (20 mL) gewaschen und in Lyse-Lösung-PC (20 mL) aufgenommen. Diese Lösung wurde zu gleichen Teilen auf vier Zentrifugationsgefäße (15 mL) aufgeteilt und bei 37°C unter Schütteln (300 rpm) im Heizblock inkubiert (210 min). Nach der Zugabe einer SDS-Lösung (2 mL; 5 %) in jeden Ansatz wurden diese für 10 min auf 60°C erhitzt. Anschließend wurden die zähflüssigen Lösungen kurz abgekühlt, in ein Zentrifugationsgefäß (50 mL) vereinigt und mit einer Phenol/Chloroform/Isoamylalkohol-Mischung (8mL; 25:24:1) versetzt. Zur Durchmischung wurde mit einer umgedrehten Glaspipette (10 mL) vorsichtig trituriert (6x). Die Trennung der Phasen wurde durch Zentrifugation (15 min, 26000 g) unterstützt. Bei der Überführung der wässrigen Phase in ein neues Zentrifugengefäß war darauf zu achten, dass die DNA nicht geschert und keine Interphase mit überführt wurde. Die Phenol/Chloroform-Extraktion (8 mL) wurde solange wiederholt, bis die obere, wässrige Phase mit der DNA klar war. Um die hochmolekulare genomische DNA von Resten der organischen Phase zu befreien, wurde die DNA gefällt. Dazu wurde Natriumacetat-Lösung (1:10 Volumen, 3M, pH 8,0) und eiskaltes Isopropanol (1 Volumen) zugegeben. Die ausgefallene DNA wurde pelletiert (3 min, 4000 g) und der Überstand zweimal mit einer Ethanol-Lösung (8 mL, 70 %) gewaschen. Das Pellet wurde in TE-Puffer (10 mL) gelöst und damit ein RNase A-Verdau (90 min, 37°C, 100 µg/mL) durchgeführt. Die RNase A wurde wiederum durch eine Phenol/Chloroform-Extraktion entfernt. Das Pellet gereinigter, hochmolekularer DNA wurde dann in sterilem H_2O_{dd} (1 mL) gelöst, die Konzentration photometrisch bestimmt und so für weitere Versuche eingesetzt.

4.2.2.9 Cosmidisolierung aus *E. coli* mit dem *NucleoBond® Bac 100-Kit*

Cosmide können nicht analog zu Plasmiden nach alkalischer Lyse mittels Silikasäulchen isoliert und gereinigt werden, da diese Aufreinigung auf eine Größe unter 15 kbp begrenzt ist. Größere DNA-Moleküle, wie in diesem Fall Cosmide, binden nahezu irreversibel und lassen sich deshalb nur noch mit geringer Ausbeute eluieren. Außerdem sind Zentrifugationsschritte, wie sie in den meisten Protokollen von Plasmidreinigungs-Kits enthalten sind, aufgrund der auftretenden Scherkräfte für die Integrität der zu isolierenden Cosmide nicht geeignet. Aus diesem Grund wurden die Cosmide in dieser Arbeit mit dem *NucleoBond® Bac 100-Kit* isoliert, welches auf der reversiblen Bindung der DNA an eine Anionenaustauschersäule beruht. Die Elution der Lösungen erfolgt dabei durch Veränderung der Salzkonzentration und des pH-Wertes allein mittels Schwerkraft von der Säule.

Ausgehend von einem Pelletnassgewicht von 1,8 bis 2,0 g wurden die Isolierungen nach dem Protokoll des Herstellers durchgeführt.

4.2.2.10 Isolierung von DNA-Fragmenten aus Agarosegelen mittels *QIAEX® II Gel Extraction-Kit*

Nach PCR-Reaktionen oder Restriktionsverdaus werden die gewünschten PCR-Produkte bzw. Restriktionsfragmente häufig mittels Agarosegelelektrophorese analysiert und gereinigt. Um Schäden an der DNA zu vermeiden, wird hierbei das Agarosegel nur leicht mit Ethidiumbromid angefärbt und die Zeit der Exposition gegen UV-Licht möglichst kurz gehalten. Nach dem Ausschneiden der gewünschten DNA-Bande wird die DNA aus dem Gel extrahiert, wofür im Handel verschiedene Kits erhältlich sind. In der vorliegenden Arbeit wurde das *QIAEX II Gel Extraction-Kit* verwendet. Dazu muss zunächst das ausgeschnittene Agarosestück aufgelöst werden, was in Analogie zu der von Vogelstein und Gillespie[60] 1979 vorgeschlagenen Methode, Agarose mit Hilfe chaotroper Salze wie z.B. Natriumiodid zu lösen, verwirklicht wird. Dies geschieht durch das Aufbrechen der Wasserstoffbrückenbindungen zwischen den Zuckerresten der Agarose. In der damit hergestellten Gegenwart hoher Konzentrationen chaotroper Salze adsorbiert DNA an Silikagel. Man versucht, dieses Phänomen mit einer Modifikation der Struktur des Wassers zu erklären.[61] Im Fall des QIAEX II werden die Silikagelpartikel direkt in den Lösungen suspendiert und können durch Pelletieren und Abpipettieren leicht wieder abgetrennt werden. Verunreinigungen wie Agarose und Ethidiumbromid treten nicht in Wechselwirkung mit dem Silikagel und können durch Waschschritte leicht entfernt werden. Die abschließende Elution der DNA erfolgt bei niedrigen Salzkonzentrationen und eher basischem pH (Tris-Puffer oder Wasser), wobei Erwärmung des Elutionsmittels auf 60°C die Effizienz erhöht. Die genaue praktische Durchführung in der vorliegenden Arbeit erfolgte nach dem Herstellerprotokoll.

4.2.2.11 Photometrische Konzentrationsbestimmung

Eine einfache und schnelle Bestimmung der DNA-Konzentration erfolgt zumeist photometrisch bei einer Wellenlänge von λ = 260 nm (OD_{260}) in UV-durchlässigen Küvetten. Die Konzentration errechnet sich aus der Absorption bzw. optischen Dichte bei λ = 260 nm, was dem Absorptionsmaximum von Nukleinsäuren entspricht, dem Verdünnungsfaktor und einem spezifischen Koeffizienten, der sich für DNA, RNA und Oligonukleotide unterscheidet. Eine Lösung doppelsträngiger DNA mit einer Konzentration von 50 µg/mL hat eine OD $\approx$ 1,

was auch gleichzeitig das obere Ende des linearen Bereichs dieser Methode darstellt (OD = 0,1-1).

Die Messung der OD kann auch zur Reinheitskontrolle herangezogen werden. So liegt der Quotient OD_{260}/OD_{280} einer reinen DNA-Lösung bei 1,8. Da Proteine bei λ = 280 nm ein Absorptionsmaximum haben, wird dieser Quotient bei Proteinkontaminationen deutlich verringert. Neben diesem Quotient dient auch der OD_{260}/OD_{230}-Wert zur Charakterisierung der Reinheit der DNA-Lösung, da neben Peptidbindungen auch Tris, EDTA und andere Salze eine Absorption bei λ = 230 nm zeigen.

In der Praxis wird außerdem die OD_{320} bestimmt, die zur Hintergrundkompensation verwendet wird. Die Wellenlänge λ = 320 nm ist hierfür geeignet, da sie weit von den Absorptionsmaxima von DNA und Protein bei λ = 260 nm bzw. λ = 280 nm entfernt ist.

4.2.3 Polymerase-Kettenreaktion (PCR)

Die PCR ist eine repetitive Methode zur *in vitro*-Amplifikation beliebiger Nukleinsäureabschnitte. Sie wurde im Jahre 1983 von Kary B. Mullis entdeckt, 1986 veröffentlicht[62] und 1993 mit dem Nobelpreis für Chemie ausgezeichnet.

Prinzipiell umfasst ein Zyklus einer PCR drei Schritte, die bei unterschiedlichen Temperaturen stattfinden. Im ersten Schritt („Denaturierung“) erfolgt durch Temperaturerhöhung typischerweise auf 92-98°C eine Trennung der DNA-Doppelstränge. An die so erhaltenen Einzelstränge können sich im zweiten Schritt („Annealing“) die kurzen Oligonukleotidprimer anlagern, wenn der Reaktionsansatz auf eine durch die Primer (GC-Gehalt, Länge) vorgegebene Temperatur abgekühlt wird. Im dritten Schritt („Elongation“) wird die Temperatur je nach Aktivitätsoptimum der eingesetzten thermostabilen DNA-Polymerase auf 68-72°C erhöht.

Heutzutage gibt es abhängig von der Applikation verschiedenste Varianten bei der Durchführung einer PCR, sowohl was die PCR-Programme, die DNA-Polymerasen und weitere Additive, aber auch die Struktur der Template betrifft. In der vorliegenden Arbeit wurden neben der Standard-PCR auch eine Vielzahl abgewandelter PCR-Methoden angewendet, von denen in den folgenden Abschnitten einige kurz vorgestellt werden sollen.

4.2.3.1 Standard-PCR

Bei der Standard-PCR erfolgt die Amplifikation eines DNA-Fragments mit Hilfe von zwei möglichst spezifischen Primern. Als Templat kann hierbei genomische, Plasmid- oder Cosmid-DNA eingesetzt werden. Die Primer haben eine durchschnittliche Länge von 18 – 22 Nukleotiden und begrenzen das zu amplifizierende Fragment nach beiden Seiten. In dieser Arbeit wurde mit Hilfe dieser PCR-Methode eine Reihe von Genen und Genabschnitten amplifiziert, die dann zur Sequenzierung, heterologen Expression, Kontrolle von Klonierungsexperimenten, Überprüfung der Integrität von Cosmiden oder zum Screening der Cosmidbank als homologe oder heterologe Sonde verwendet wurden. Dabei kamen in Abhängigkeit der weiteren Verwendung zwei unerschiedliche Polymerasen zum Einsatz. Die Taq-Polymerase von Qiagen liefert doppelsträngige DNA mit einem A-Überhang und wurde für die meisten Anwendungen verwendet, da die Produkte leicht mittels TA-Klonierung (s. Abschnitt 4.2.4.4) weiter verarbeitet werden konnten. Für die Amplifikationen von Genen zur Überexpression wurde die Pwo SuperYield DNA-Polymerase von Roche eingesetzt, da diese aufgrund ihrer 3´->5´-Exonukleaseaktivität eine viel höhere Genauigkeit bzw. geringere Fehlerrate besitzt. Die jeweilige Zusammensetzung der Ansätze ist in Tabelle 3 aufgeführt.

Wenn nicht gesondert aufgeführt, wurde das in Tabelle 4 beschriebene Standard-PCR-Protokoll verwendet.

Tabelle 3: Zusammensetzung der Standard PCR-Ansätze. Die Additive für GC-reiche Lösungen wurden bei entsprechendem Templat zugegeben.

Reagenz	Taq-Polymerase von Qiagen		Pwo SuperYield DNA-Polymerase von Roche	
Puffer	5,0 µl	10 x CoralLoad PCR-Puffer	5,0 µl	10 x Pwo SuperYield PCR-Puffer
Additiv für GC-reiche Sequenzen	10,0 µl	Q-Solution	10,0 µl	5 x GC-rich Solution
Templat	1,0 µl	je nach Ansatz	1,0 µl	je nach Ansatz
dNTP`s	1,0 µl	dNTP-Mix von NEB	1,0 µl	PCR Grade Nucleotide Mix
Primer	je 1,0 µl	10 pmol/µl	je 1,0 µl	10 pmol/µl
Polymerase	0,25 µl	Taq-Polymerase	0,5 µl	Pwo SuperYield DNA-Polymerase
H_2O_{dd}	ad 50 µl	-	ad 50 µl	-

Tabelle 4: Standard-PCR-Protokoll.

Zyklenzahl	Zeit	Temperatur	Reaktion
1	3 min	94°C	initiale Denaturierung
30	45 s	94°C	Denaturierung
	60 s	je nach Primerpaar	Annealing
	1000 kbp/min	72°C	Elongation
1	7 min	72°C	finale Elongation
1	-	2,0°C/s	Abkühlen
1	∞	10°C	Lagerung

4.2.3.2 PCR mit elongierten Primern

In Abwandlung des Prinzips der Standard-PCR erfordern einige Applikationen die Verlängerung des Primerpaares über das spezifische 5´-Ende hinaus. In dieser Arbeit wurde diese Modifikation der Primer bei der Identifizierung des Genclusters und der Vorbereitung des Cosmids 1104 für die heterologe Expression eingesetzt.

Zur Identifizierung des Genclusters sollte eine PCR durchgeführt werden, bei der das erwartete Produkt mit einer Größe von 51 Nukleotiden für eine Detektion, Klonierung und Sequenzierung sehr schwer zugänglich gewesen wäre. Aus diesem Grund wurden die Primer (EV1, EH1-4 s. Abschnitt 3.9) jeweils am 5´-Ende um 18 Nukleotide einer unspezifischen Sequenz verlängert. Mit diesen Primern wurden Standard-PCR-Ansätze durchgeführt, wobei die Annealingtemperatur mit 50°C entsprechend niedrig gewählt wurde.

Zur Modifikation des Cosmids 1104 mittels λ-RED-System wurden Primer eingesetzt, die zusätzlich zu den spezifischen 19 bzw. 20 Nukleotiden noch 40 Nukleotide lange 5´-Überhänge besaßen. Diese unspezifischen Überhänge dienten dann zu einem späteren Zeitpunkt zur homologen Rekombination in das Cosmid. Die Amplifikation erfolgte in diesem Fall mit der Master Taq DNA-Polymerase der Firma Eppendorf, Hamburg, da alle Ansätze mit anderen Polymerasen in diesem Fall kein Produkt ergeben hatten. Der verwendete Ansatz ist in Tabelle 5 und das PCR-Programm in Tabelle 6 beschrieben.

Tabelle 5: Zusammensetzung der PCR-Ansätze mit dem Master Taq DNA-Polymerase Kit der Firma Eppendorf, Hamburg.

Reagenz	Master Taq DNA-Polymerase von Eppendorf	
Puffer	5,0 µl	10 x Taq Puffer mit Mg^{2+}
Additiv für GC-reiche Sequenzen	10,0 µl	Taq Master
Templat	2,0 µl	*Eco*RV-linearisierter pSET152
dNTP`s	1,5 µl	dNTP-Mix von NEB
Primer	je 0,5 µl	100 pmol/µl
Polymerase	0,20 µl	Taq DNA-Polymerase
H_2O_{dd}	ad 50 µl	-

Tabelle 6: PCR-Protokoll zur Amplifikation des DNA-Fragments zur Modifikation von Cosmid 1104.

Zyklenzahl	Zeit	Temperatur	Reaktion
1	3 min	94°C	initiale Denaturierung
	45 s	94°C	Denaturierung
10	45 s	56°C ± 2°C	Annealing
	5 min	72°C	Elongation
	45 s	94°C	Denaturierung
20	45 s	60°C ± 2°C	Annealing
	5 min	72°C	Elongation
1	7 min	72°C	finale Elongation
1	-	2,0°C/s	Abkühlen
1	∞	10°C	Lagerung

4.2.3.3 *Nested*-PCR und *single-specific*-Primer-PCR

Führt man zwei Standard-PCR-Reaktionen so durch, dass das Produkt der ersten PCR als Templat für die zweite PCR verwendet wird, und dass das Primerpaar für die zweite PCR innerhalb des Primerpaars für die erste PCR liegt („geschachtelte Primerpaarkombination"), kann man sehr spezifisch das gewünschte DNA-Fragment, auch von sehr wenig Templat ausgehend, amplifizieren. Diese in Abbildung 17 schematisch dargestellte *nested*-PCR wird hauptsächlich verwendet, wenn aufgrund zu vieler unspezifischer Banden oder zu geringer Ausbeute das gewünschte Produkt bei einer einfachen PCR nicht isoliert werden kann.

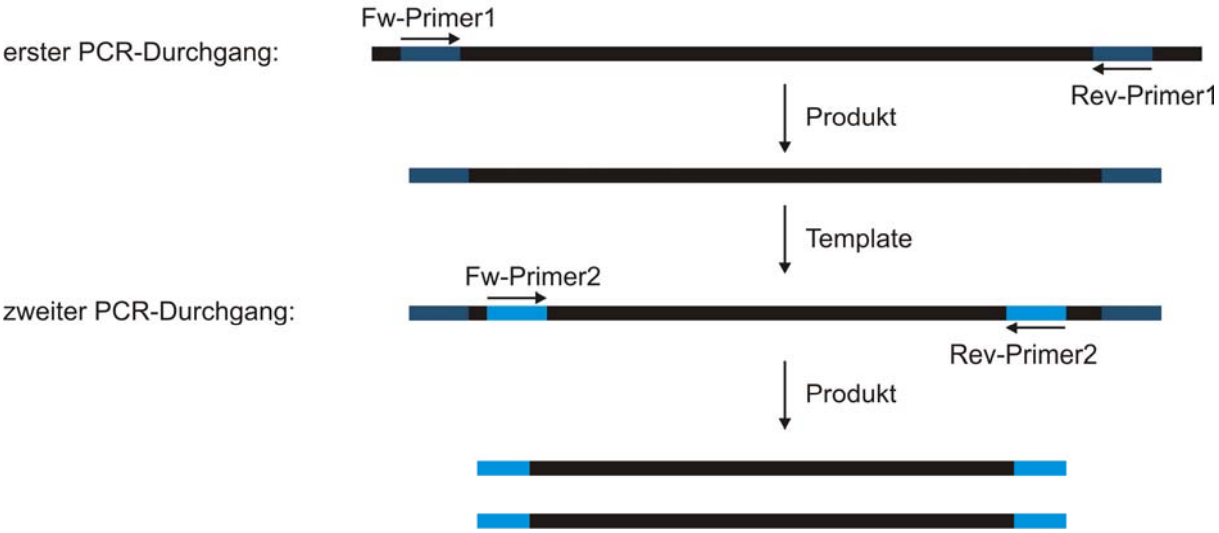


Abbildung 17: Schematische Darstellung der Standard-Anwendung einer *nested*-PCR. In der zweiten PCR-Reaktion wird das Primerpaar so gewählt, dass es innerhalb des Primerpaars der ersten PCR-Reaktion bindet. Der Ansatz der ersten Reaktion wird dann als Templat für die zweite Reaktion eingesetzt.

In der vorliegenden Arbeit wurde ein Spezialfall der *nested*-PCR verwendet, da in der ersten PCR neben einem spezifischen (allerdings stark degenerierten) auch ein unspezifischer Primer eingesetzt wurde (s. Abbildung 18). Eine solche PCR wird auch als *single-specific*-Primer-PCR (SSP-PCR) bezeichnet.

Abbildung 18: Schematische Darstellung einer *single-specific*-Primer-PCR. Durch Kombination eines spezifischen und eines unspezifischen Primers wird durch Wahl der geeigneten Annealingtemperatur eine randomisierte Bindung des unspezifischen Primers gefördert. Dadurch entstehen verschiedenste PCR-Produkte unterschiedlicher Länge.

Die dadurch bedingte geringe Spezifität erforderte die zweite PCR im Sinne einer *nested*-PCR. Insgesamt kann somit für diese Methode der Name SSP-*nested*-PCR verwendet werden.

Im Folgenden soll auf die Details der praktischen Durchführung eingegangen werden. Die Ansätze (s. Tabelle 7) des ersten PCR-Durchlaufs enthielten je einen spezifischen degenerierten (NHP1/NVP1 s. Abschnitt 3.9) und einen unspezifischen Primer (fwrLAT3r, fwToyo Modul 5 5 s. Abschnitt 3.9) und wurden mit dem in Tabelle 8 aufgeführten Protokoll durchgeführt.

Tabelle 7: Zusammensetzung der Ansätze für den ersten Durchlauf der SSP-*nested*-PCR mit der Taq-Polymerase von Qiagen.

Reagenz		Taq-Polymerase von Qiagen
Puffer	2,0 µl	10 x CoralLoad PCR-Puffer
Additiv für GC-reiche Sequenzen	4,0 µl	Q-Solution
Templat	0,4 µl	je nach Ansatz
dNTP`s	0,4 µl	dNTP-Mix von NEB
Primer	je 0,5 µl	spezifischer, degen. Primer: 100 pmol/µl unspezifischer Primer: 10 pmol/µl
Polymerase	0,20 µl	Taq-Polymerase
H_2O_{dd}	ad 20 µl	-

Tabelle 8: PCR-Protokoll für den ersten Durchlauf der SSP-*nested*-PCR.

Zyklenzahl	Zeit	Temperatur	Reaktion
1	3 min	95°C	initiale Denaturierung
	45 s	95°C	Denaturierung
10	45 s	38°C	Annealing
	3,5 min	72°C	Elongation
	45 s	94°C	Denaturierung
30	45 s	52°C	Annealing
	3,5 min	72°C	Elongation
1	5 min	72°C	finale Elongation
1	-	2,0°C/s	Abkühlen
1	∞	10°C	Lagerung

Da in die zweite PCR-Reaktion der Ansatz der ersten als Templat eingehen sollte, mussten zur Vermeidung von Kreuzreaktionen zunächst die Primer des ersten Ansatzes entfernt werden. Dazu wurde ein ExonukleaseI-Verdau durchgeführt, wobei die Exonuklease mit ihrer 3´->5´- Aktivität einzelsträngige DNA und damit die Primer abverdauen sollte. Dazu wurden jeweils 5 µl PCR-Produkt mit 0,5 µl ExonukleaseI (20 u/µl) versetzt und zunächst bei 37°C inkubiert (15 min), bevor das Enzym hitzeinaktiviert (80°C, 15 min) wurde. Die Ansätze der zweiten PCR besaßen die gleiche Zusammensetzung wie die der ersten PCR (s. Tabelle 7), allerdings mit dem jeweiligen PCR-Produkt der ersten Runde als Templat und einem anderen spezifischen Primer (NHP2 s. Abschnitt 3.9), der innerhalb der ersten PCR-Produkte liegen sollte. Als PCR-Programm wurde in diesem Fall ein Standard-Protokoll (s. Abschnitt 4.2.3.1) mit einer Annealingtemperatur von 56°C und einer Elongationsdauer von 5 min verwendet. Die Auswertung der Ansätze erfolgte mittels Agarosegelelektrophorese durch Vergleich mit diversen Kontrollen.

4.2.3.4 Herstellung DIG-markierter Sonden mittels PCR

Zum Screening der Cosmidbank wurde eine Vielzahl von homologen und heterologen Sonden hergestellt. Zur Signalgebung der gegen die Cosmidbankfilter hybridisierten Sonden wurde neben der radioaktiven Markierung mit [^{32}P] auch eine durch PCR eingeführte Markierung mit Digoxigenin eingesetzt. Die PCR-Ansätze hatten dazu die in Tabelle 9 angegebene Zusammensetzung. Als PCR-Programm wurde ein Standard-Protokoll herangezogen.

Tabelle 9: Zusammensetzung der PCR-Ansätze zur Herstellung Digoxigenin-markierter Sonden mit der Taq-Polymerase von Qiagen.

Reagenz	Taq-Polymerase von Qiagen	
Puffer	5,0 µl	10 x CoralLoad PCR-Puffer
Additiv für GC-reiche Sequenzen	10,0 µl	Q-Solution
Templat	1,0 µl	je nach Ansatz
dNTP`s	5 µl	DIG-dNTP-Mix (s. Abschnitt 3.3.3)
Primer	je 1,0 µl	10 pmol/µl
Polymerase	0,20 µl	Taq-Polymerase
H_2O_{dd}	ad 20 µl	-

4.2.3.5 Ortsgerichtete Mutagenese mit mutierten Primern

Eine weiteres Einsatzgebiet der PCR ist die ortsgerichtete Mutagenese zur Deletion, Einführung oder zum Austausch von Nukleotiden, was sich auf Proteinebene in unterschiedlicher Weise auswirken kann. Das *QuickChange® II Site-Directed Mutagenesis-Kit* der Firma Stratagene, Amsterdam, ermöglicht die Durchführung der ortsgerichteten Mutagenese in einem dreistufigen Prozess, ausgehend von einem doppelsträngigen, zirkulären Plasmid (s. Abbildung 19).

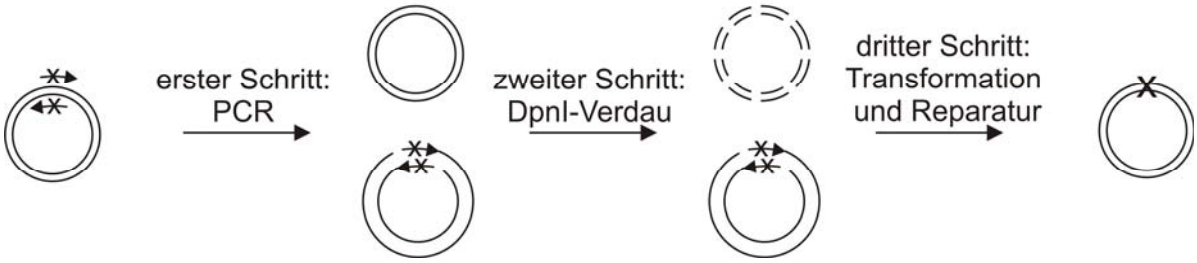


Abbildung 19: Schematische Darstellung des dreistufigen Prozesses zur ortsgerichteten Mutagenese mit dem *QuickChange® II Site-Directed Mutagenesis-Kit* der Firma Stratagene, Amsterdam.

Die für die PCR im ersten Schritt benötigten Primer werden an exakt dem gleichen Sequenzstück des Plasmids abgeleitet und tragen beide die einzuführende Mutation möglichst mittig. Nach der PCR mit dem mutierten Primerpaar liegen die Einzelstränge um den Primer versetzt vor, was nach der finalen Renaturierung zu einem „Knick“ in den linearen aber trotzdem pseudozirkulär-doppelsträngig vorliegenden Plasmiden führt. Nach dem Verdau mit dem Restriktionsenzym DpnI, das selektiv die unmutierten Templatplasmide

durch Erkennung ihrer Methylierung schneidet, werden die pseudozirkulären Plasmide transformiert, dort durch cytosolische Enzyme „repariert“ und anschließend amplifiziert. In dem so erhaltenen Plasmid ist jetzt die durch die Primer vorgegebene Sequenz enthalten.
Diese Methode wurde in der vorliegenden Arbeit für die Generierung von mutierten Präpropeptidsubstraten für den *in vitro*-Assay zur Untersuchung der LabKC-Aktivität eingesetzt. Bei der Ableitung der Primer wurden die gewünschten Mutationen durch gezieltes Austauschen einzelner Basen eingeführt. Die Primer wurden in die in Tabelle 10 aufgeführten Ansätze eingesetzt und die PCR mit dem in Tabelle 11 beschriebenen Protokoll durchgeführt.

Tabelle 10: Zusammensetzung der PCR-Ansätze für den ersten Schritt der ortsgerichteten Mutagenese mit der *PfuUltra* HF DNA-Polymerase von Stratagene**.**

Reagenz	*PfuUltra* HF DNA-Polymerase von Stratagene	
Puffer	5,0 µl	10 x Reaktions-Puffer
Templat	1,0 µl	LabA2-Expressionskonstrukt (s. Abschnitt 5.3.1)
dNTP`s	1,0 µl	dNTP-Mix
Primer	je 1,25 µl	10 pmol/µl
Polymerase	1,0 µl	Taq-Polymerase
H_2O_{dd}	ad 40,5 µl	-

Tabelle 11: PCR-Programm für die ortsgerichtete Mutagenese mit mutierten Primern.

Zyklenzahl	Zeit	Temperatur	Reaktion
1	30 s	95°C	initiale Denaturierung
	30 s	95°C	Denaturierung
12	60 s	55°C	Annealing
	6 min	68°C	Elongation
1	7 min	68°C	finale Elongation
1	-	1,5°C/s	Abkühlen
1	∞	10°C	Lagerung

Durch Zugabe von DpnI (1 µl) in jeden PCR-Ansatz wurden die Templatplasmide enzymatisch verdaut (1 h; 37°C). Anschließend wurde mit jedem Ansatz (1 µl) eine Transformation in *E. coli* XL1-Blue nach dem in Abschnitt 4.2.1.2 aufgeführten Protokoll durchgeführt.

4.2.3.6 Kolonie-PCR

Bei den bisher beschriebenen PCR-Ansätzen lag die eingesetzte Templat-DNA immer gereinigt vor. Es ist jedoch auch die Verwendung von DNA möglich, die direkt und ohne große Aufreinigung aus etwas Bakterienkolonie generiert wurde. Diese sog. Kolonie-PCR wurde auch in der vorliegenden Arbeit verwendet. Dazu wurde mit einem Zahnstocher etwas Kolonie in $H_2O_{dd,\ steril}$ (100 µl) aufgenommen und die Zellen mittels Erhitzen (10 min, 100°C) aufgeschlossen. Vor der Verwendung in der PCR wurde dieser Ansatz auf Eis abgekühlt und zentrifugiert (10 min, 16000 g). Statt wie sonst 1 µl mussten allerdings bei der Kolonie-PCR 10 bis 15 µl Templat-DNA in den in Tabelle 5 angegebenen Standardansatz eingesetzt werden.

4.2.4 Klonierung

Unter Klonierung versteht man allgemein die Einführung von DNA-Fragmenten in Vektoren, die die Amplifikation des eingeführten Fragments ermöglichen. Dieser Vorgang erfordert und bedingt eine Vielzahl zum Teil enzymatisch katalysierter Reaktionen, wobei im Folgenden nur die in dieser Arbeit angewendeten Methoden vorgestellt werden.

4.2.4.1 Restriktionsverdau

Restriktionsendonukleasen oder kurz Restriktionsenzyme erkennen je nach Spezifität vier bis acht normalerweise palindromische Basenpaare in einem DNA-Doppelstrang und schneiden diesen meist innerhalb dieser Erkennungssequenz (Typ II Restriktionsenzyme). Nach dem Restriktionsverdau bleiben an den 5´-Enden der DNA-Fragmente fast immer Phosphatreste zurück, die dann für eine anschließende Ligation wichtig werden können, während die 3´-Enden lediglich Hydroxygruppen tragen. Neben der Erkennungssequenz ist auch die Art der generierten Enden ein wichtiges Charakteristikum eines Restriktionsenzyms. Während manche Enzyme glatte Enden (*blunt ends*) produzieren, können andere 5´- bzw. 3´-Überhangenden (*sticky ends*) unterschiedlicher Länge generieren. Die käuflich erhältlichen Restriktionsenzyme liegen vorwiegend in einer Konzentration von 10 units/µl vor, wobei die unit definiert ist als die Enzymmenge, die 1 µg DNA in 1 h bei der angegebenen Temperatur in einem 50 µl Ansatz vollständig verdaut. Man setzt üblicherweise das Enzym in einem drei- bis fünffachen Überschuss in einer Reaktion ein. In der vorliegenden Arbeit wurde für analytische Zwecke meist folgender Ansatz verwendet und 1 h bei der vom Hersteller angegebenen Temperatur inkubiert:

jeweilige DNA-Lösung	5,0 bis 7,7 µl
10 x Restriktionspuffer	1,0 µl
Restriktionsenzym	0,3 µl
Optional 100 x BSA	0,1 µl
steriles H_2O_{dd}	ad 10,0 µl

Restriktionsverdaus wurden in der vorliegenden Arbeit neben der Analyse von Plasmiden und Cosmiden auch bei der Identifizierung des Labyrinthopeptin-Genclusters und zur Erzeugung von PCR-Templaten herangezogen.

4.2.4.2 Ligation

Unter Ligation versteht man das Verknüpfen von DNA-Strängen. Als Standardenzym findet hierbei die T4 DNA Ligase Verwendung, weil sie sowohl glatte als auch überhängende Enden doppelsträngiger DNA unter ATP-Verbrauch miteinander verbindet, indem sie Phosphodiesterbindungen zwischen den 3´-OH- und den 5´-Phosphatenden ausbildet.
Im Laufe dieser Arbeit wurden zwei verschiedene T4 DNA Ligase-Systeme eingesetzt, die T4 DNA Ligase und das *Quick Ligation-Kit* von New England Biolabs, Frankfurt. Das typische Verhältnis von Insert zu Vektor von 3:1 bis 5:1 wurde bei beiden meistens eingehalten, wobei die Unterschiede in der Gesamtgröße eines Ansatzes und in der Inkubationsdauer lagen. Während die Ansätze mit T4 DNA Ligase für 16 h bei 16°C inkubiert wurden, war die Inkubationszeit bei Verwendung des *Quick Ligation-Kits* mit 1 h deutlich kürzer, wobei in diesem Fall nach 30 min die Temperatur auf 10°C gesenkt wurde. Nach der Ligation wurde zumeist direkt transformiert (2 – 5 µl).

Tabelle 12: Ligationsansätze mit der T4 DNA Ligase und dem *Quick Ligation-Kit* von New England Biolabs, Frankfurt.

Reagenz		T4 DNA Ligase		*Quick Ligation-Kit*
Vektor	x µl	50 ng	x µl	50 ng
Insert	y µl	drei- bis fünffach molarer Überschuss	y µl	drei- bis fünffach molarer Überschuss
H_2O_{dd}	10-(x+y) µl		10-(x+y) µl	
Puffer	1,2 µl	10x T4 DNA Ligase Reaction Buffer	10 µl	2x Quick Ligation Reaction Buffer
Ligase	1,0 µl	T4 DNA Ligase	1,0 µl	Quick T4 DNA Ligase

4.2.4.3 *A-Addition*

Mit TA-Klonierung bezeichnet man eine häufig verwendete Methode, bei der PCR-Produkte mit A-Überhang leicht in einen bereits linearisierten Vektor mit T-Überhang kloniert werden können. Während DNA-Polymerasen ohne 3´->5´-Exonukleaseaktivität, z.B. die Taq-Polymerase, meist diesen unspezifischen A-Überhang synthetisieren, erzeugen DNA-Polymerasen mit 3´->5´-Exonukleaseaktivität ausschließlich glatte Enden. Um mit dabei erzeugten PCR-Produkten die Methode der TA-Klonierung auch anwenden zu können, bedient man sich der sog. *A-Addition*. Die Firma Qiagen, Hilden, nutzt hierzu nach eigenen Angaben eine „Art Klenow-Fragment“, welches in Anwesenheit von dATP die gewünschten

A-Überhänge generiert. Wie den Herstellerangaben zu entnehmen ist, sollen 0,15 bis 1,5 pmol PCR-Produkt in die Reaktion eingesetzt werden, wobei sich folgender Ansatz ergab, der 30 min bei 37°C inkubiert wurde:

PCR-Produkt	≤ 8 µl
5x *QIAGEN A-Addition Master Mix*	2 µl
$H_2O_{dd,\ steril}$	ad 10 µl

In der vorliegenden Arbeit wurde diese Reaktion zur Klonierung des *labKC*-Gens in den pET SUMO-Vektor für die heterologe Expression eingesetzt, da das mit der Pwo SuperYield DNA-Polymerase erzeugte *labKC*-Gen nicht direkt in den linearisierten pET SUMO-Vektor mit T-Überhang klonierbar war.

4.2.4.4 TA-Klonierung

Das bereits in Abschnitt 4.2.4.3 beschriebene Prinzip der TA-Klonierung wurde in der vorliegenden Arbeit zur Klonierung in den Klonierungsvektoren pDrive und den Expressionsvektor pET SUMO verwendet. Bei der Klonierung in pDrive wurde zur Berechnung der einzusetzenden DNA-Menge des Inserts Formel (1) herangezogen, wobei von 50 ng pDrive und in der Regel von einem fünffachen molaren Überschuss an Insert ausgegangen wurde:

$$\text{ng Insert} = \frac{50\ \text{ng} * \text{Insertgröße bp} * 5}{3851\ \text{bp}} \qquad (1)$$

Für die Klonierung in pET SUMO (5643 bp) wurde die Formel (1) entsprechend angepasst. Daraus ergaben sich die in Tabelle 13 verwendeten Ansätze.

Tabelle 13: Ligationsansätze zur TA-Klonierung mit pDrive und pET SUMO.

Reagenz	pDrive		pET SUMO	
Vektor	1,0 µl	50 ng	2,0 µl	50 ng
Insert	1,0-4,0 µl	fünffacher molarer Überschuss	1,0-6,0 µl	fünffacher molarer Überschuss
Puffer	5,0 µl	2x Ligations Master Mix	1,0 µl	10x Ligations-Puffer
Ligase	-	-	1,0 µl	T4 DNA Ligase
H_2O_{dd}	x µl	ad 10 µl	x µl	ad 10 µl

Nach Inkubation (16 h) bei langsam von 25°C auf 15°C abnehmender Temperatur und anschließender Transformation in $CaCl_2$-kompetente *E. coli* DH5α (pDrive) bzw. *E. coli* Mach1^{TM}-T1^{R} (pET SUMO) von Invitrogen wie in Abschnitt 4.2.1.2 beschrieben, wurden die Zellen auf TY-Platten mit entsprechender Selektion (pDrive: TY-Amp; pET SUMO: TY-Kan) ausplattiert. Durch die Kultivierung unter Selektionsdruck konnte die Aufnahme eines Plasmids in den gewachsenen Klonen sichergestellt werden, allerdings nicht die erfolgreiche TA-Klonierung eines PCR-Produkts und dessen Orientierung in dem Klonierungsvektor. Bei den Klonierungen in pDrive konnte zum Nachweis einer erfolgreichen Insertion die sog. Blau-Weiß-Selektion herangezogen werden, wozu den Platten die entsprechenden Mengen an IPTG (Isopropyl β-*D*-Thiogalaktopyranosid) und X-Gal (5-Brom-4-chlor-3-indoxyl-β-*D*-galaktosid) zugesetzt wurden. Das Prinzip dieser Selektion beruht darauf, dass die Klonierungsstelle in dem *lac*Z´-Gen liegt, welches für das *N*-terminale α-Fragment der β-Galaktosidase kodiert. Bei Plasmiden ohne Insert kommt es zu einer Wechselwirkung mit dem vom Bakterienstamm exprimierten ω-Fragment LacZΔM15, was zu einer Reaktivierung der β-Galaktosidaseaktivität führt. Der Medienzusatz von IPTG zur Induktion und X-Gal als Substrat führen dann zu einer enzymatischen Hydrolyse von X-Gal in Galaktose und 5-Brom-4-chlor-indoxyl, wobei letzteres vom Luftsauerstoff zum tiefblauen Farbstoff 5,5´-Dibrom-4,4´-dichlor-indigo oxidiert wird. Wurde das *lac*Z´-Gen allerdings durch Insertion eines PCR-Produktes inaktiviert, so bleiben die Klone weiß.[63] Einschränkend muss dazu jedoch gesagt werden, dass erst bei Inserts >1000 bp eine verlässliche Aussage über die Blau-Weiß-Selektion erhalten werden konnte. Zur Kontrolle auf Insertion des richtigen Fragments und dessen Orientierung wurden geeignete Restriktionsverdaus bzw. Sequenzierungen herangezogen.

4.3 Bioinformatik

Die Auswertung der generierten Sequenzdaten des Cosmids 1104 erfolgte unter Zuhilfenahme verschiedener Software-Tools. Da die vorliegende Arbeit im Rahmen einer Kooperation mit der Firma Sanofi-Aventis, Frankfurt, erstellt wurde und die Ergebnisse in zwei Patente eingehen werden, mussten bei der Annotation der Sequenzdaten einige Einschränkungen beachtet werden. Das Einstellen von Sequenzen zur Datenbank-Suche oder Nutzung anderer Software-Tools im Internet war nicht erlaubt und es mussten Wege gefunden werden, die benötigten Tools lokal zu installieren. Aufgrund dieser Problematik wurden die Programme *Artemis* (Sanger Institute, Hixton, UK) und BLAST (***B**asic **L**ocal **A**lignment **S**earch **T**ool*) sowie die dazu benötigten Datenbanken (nt = nicht-redundante Nukleotid-Datenbank; nr = nicht-redundante Protein-Datenbank) aus dem Internet heruntergeladen und lokal auf einem Rechner installiert. Dadurch konnten zumindest diese Software-Tools für die Annotation des Cosmids herangezogen werden.

4.3.1 BLAST

Blast steht für ***B**asic **L**ocal **A**lignment **S**earch **T**ool* und wird seit 1990 am National Center for Biotechnology Information NCBI, Bethesda, MD, USA entwickelt.[64] Dieses Programm vergleicht eingegebene Nukleotid- oder Proteinsequenzen (Query-Sequence) mit Nukleotid- oder Proteinsequenzdatenbanken und berechnet die statistische Signifikanz der Treffer, was ermöglicht, homologe Sequenzen zu extrahieren.

Das Blast-Programm besteht aus verschiedenen Unterprogrammen, wie z.B. BlastN, welches zum Vergleich von Nukleotidsequenzen mit einer Nukleotiddatenbank verwendet wird. Analog dazu entspricht BlastP dem Programm für Proteine. Mit Hilfe des Programms BlastX wird eine Nukleotidsequenz in die entsprechenden Proteine übersetzt und dann mit einer Proteindatenbank verglichen.

In einem ersten Schritt des heuristischen Blast-Algorithmus werden die eingegebene Sequenz und die Sequenzen der Datenbank in Fragmente („Words“ W; die Länge W beträgt zumeist 3 Aminosäuren) zerlegt und diese dann nach Homologien durchsucht. Anhand einer Scoring-Matrix, üblicherweise Blosum62, werden diese Homologien bewertet (T-Score). Liegt dieser T-Score über einem bestimmten Schwellenwert, so wird versucht, diesen homologen Bereich in beide Richtungen zu verlängern. Dadurch ensteht dann ein lokales Alignment, welches wiederum bewertet wird (S-Score) und über einem bestimmten Schwellenwert liegen muss, um als interessant abgespeichert zu werden. Neben dem Score, der direkt aus der Übereinstimmung der Sequenzen berechnet wird, werden weitere statistische Parameter angegeben. So gibt z.B. der E-value (*expectation-value*) die

Gesamtwahrscheinlichkeit an, dass der beobachtete Wert des Alignments von zwei zufällig nicht-homologen Sequenzen erreicht werden kann. Je kleiner dieser E-value ist, desto signifikanter ist der entsprechende Treffer.[65]

In dieser Arbeit wurde Blast als schnelle und einfache Methode verwendet, um die aus der Cosmidsequenzierung erhaltenen Sequenzdaten auszuwerten. So wurde mittels Nukleotid- und Proteinblasts versucht, die Funktionen der gefundenen Gene und der dadurch kodierten Proteine zuzuordnen. Die Blast-Analysen wurden aufgrund der patentrechtlichen Situation über das Kommandozeilen-Interface der Blast-Programme über eine lokale Installation gegen die nicht-redundante Nukleotiddatenbank und die nicht-redundante Proteindatenbank durchgeführt.

4.3.2 Artemis

Artemis kann zur Darstellung von DNA-Sequenzen und zu deren Annotation herangezogen werden.[66] Das Programm stellt außerdem eine Vielzahl anderer Software-Tools zur Verfügung, die direkt von der eingegebenen Sequenz ausgehend angewendet werden können. In der vorliegenden Arbeit wurde dieses Programm hauptsächlich zur Auffindung von ORFs und Startcodons unter Zuhilfenahme des FramePlot-Tools herangezogen. Dazu wurde das Programm lokal installiert.

4.3.3 FramePlot

Gene von Bakterien, die ein sehr GC-reiches Genom haben, besitzen einen unausgewogenen *Codonusage*. Das zeigt sich darin, dass die Codons dieser Gene einen extrem hohen GC-Gehalt von bis zu 92 % in der dritten Base haben. Dieser Sachverhalt wurde 1984 zum ersten Mal beschrieben und wird heutzutage vor allem in der Streptomyceten-Genetik angewendet, weil damit ermöglicht wird, kodierende Bereiche vorherzusagen.[67] Diese Vorgehensweise beruht auf der Tatsache, dass für alle 20 Aminosäuren Codons zur Verfügung stehen, die auf G oder C enden. Im Gegensatz dazu gibt es zwölf Aminosäuren, deren Codons an Position 2 ein A oder T haben, was zu einem durchschnittlichen GC-Gehalt von 40 bis 60 % an dieser Position führt. Weitere sieben, eher selten vorkommende Aminosäuren haben ein A oder T an erster Position ihrer Codons, was zu einem durchschnittlichen GC-Gehalt von 70 % an dieser Stelle führt. Die Konsequenz, die sich daraus für proteinkodierende, GC-reiche Sequenzen ergibt, ist, dass die Position drei eines Codons diesen hohen GC-Gehalt zeigen muss, um insgesamt auf den hohen GC-Gesamtgehalt zu kommen.

Bei *A. namibiensis* handelt es sich wie bei den Streptomyceten auch um ein Mitglied der Ordnung Actinomycetales mit einem sehr hohen GC-Gehalt. Der GC-Gehalt des Cosmids 1104 wurde auf 74,3 % bestimmt, weshalb dieses Software-Tool in der vorliegenden Arbeit dazu benutzt wurde, ORFs kodierender Bereiche zu finden und die Position des Startcodons zu lokalisieren.

4.4 Proteinbiochemische Methoden

4.4.1 Photometrische Proteinbestimmung nach Bradford

Bei der quantitativen Proteinbestimmung nach Bradford handelt es sich um die empfindlichste, auf einer Färbereaktion beruhende Methode, die allerdings für verschiedene Proteine sehr unterschiedliche Ergebnisse liefern kann.[68] Das Prinzip beruht darauf, dass der blaue Säurefarbstoff Coomassie Brillantblau G 250 in Gegenwart von Proteinen und in saurem Milieu sein Absorptionsmaximum von λ = 465 nach λ = 595 nm verschiebt. Der Grund hierfür ist wahrscheinlich die Stabilisierung des Farbstoffs in seiner unprotonierten, anionischen Sulfonatform durch Komplexbildung mit Proteinen. Der Farbstoff bindet dabei recht unspezifisch an kationische und nichtpolare, hydrophobe Seitenketten der Proteine, vor allem an Arg. Als Maß für die Proteinkonzentration kann deshalb die Absorption bei einer Wellenlänge zwischen λ = 570 nm und λ = 610 nm herangezogen werden. Als Referenz wird die Absorption einer Leerprobe des verwendeten Puffers ohne Protein gemessen und auf Null gesetzt. Um die Genauigkeit der Ergebnisse zu erhöhen, werden in der Praxis Mehrfachbestimmungen durchgeführt.

In der vorliegenden Arbeit wurde mit dem Roti®-Quant-Reagenz der Firma Carl Roth, Karlsruhe, gearbeitet. Nach Inkubation (10 min) bei 22°C wurde die Absorption dieser Ansätze (50 µl Proteinprobe + 200 µl Roti®-Quant-Lösung) bei λ = 610 nm gemessen. Um eine grobe Abschätzung der Absolutkonzentrationen vornehmen zu können, wurde zunächst nach Beschreibung des Herstellers eine Kalibriergerade mit Albumin erstellt. Aus der in Abbildung 20 dargestellten Kalibriergeraden durch den Nullpunkt ergab sich folgende Geradengleichung, die für die Bestimmung der Proteinkonzentration in dieser Arbeit herangezogen wurde:

$$\text{Absorption}_{610nm} = 0{,}0064 \times \text{Albuminkonzentration [µg/mL]}$$

$$\text{Albuminkonzentration [µg/mL]} = \text{Absorption}_{610nm} : 0{,}0064$$

Für die Messung der Proteinkonzentration der einzelnen Stufen der Proteinisolierung wurden die Lösungen zuvor so verdünnt, dass die Absorption im linearen Bereich der Kalibriergeraden lag.

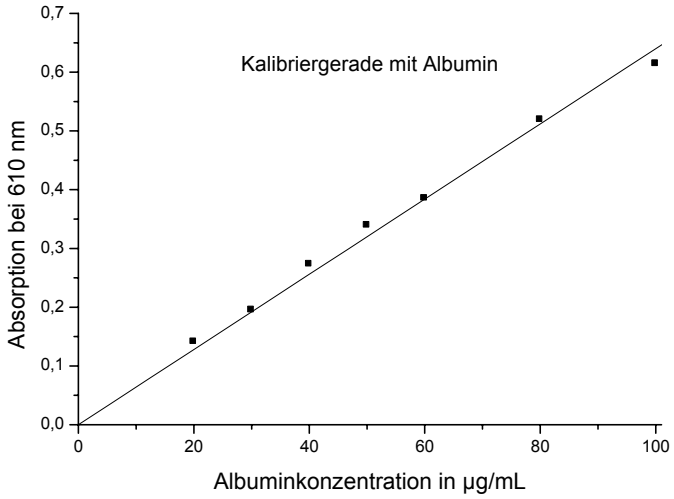


Abbildung 20: Kalibriergerade, die mit Hilfe einer Albumin-Konzentrationsreihe bestimmt wurde. Von jeder Verdünnung wurde eine Dreifachbestimmung durchgeführt und vom jeweiligen Mittelwert die Leerprobe abgezogen.

4.4.2 Photometrische Proteinbestimmung durch Absorptionsmessung bei λ = 280 nm

Die Konzentration einer reinen Lösung eines Proteins kann bei Kenntnis des Extinktionskoeffizienten von diesem Protein durch Messung der Absorption bei λ = 280 nm bestimmt werden. Der Zusammenhang wird durch das Lambert-Beersche Gesetz gegeben:

$$E_{280nm} = -\log\frac{I}{I_0} = c \cdot \varepsilon_{280} \cdot d$$

mit *I* (Intensität des transmittierten Lichts), I_0 (Intensität des eingestrahlten Lichts), *c* (Konzentration der reinen Protein-Lösung), ε_{280} (molarer Extinktionskoeffizient bei einer Wellenlänge von λ = 280 nm) und *d* (Weglänge des Lichtstrahls in der Probe in cm). Molare Extinktionskoeffizienten können im Internet aus einer gegebenen Aminosäuresequenz berechnet werden (**http://www.scripps.edu/~cdputnam/protcalc.html**). Aufgrund der patentrechtlichen Lage wurden die Berechnungen der Extinktionskoeffizienten mit randomisierten Sequenzen gleicher Zusammensetzung durchgeführt. Es ergaben sich dabei für LabA2 und die LabA2-His-Tag-Konstrukte folgende Extinktionskoeffizienten:

LabA2: $\varepsilon_{280nm} = 11380\frac{1}{M \cdot cm}$

LabA2-His-Tag-Konstrukte: $\varepsilon_{280nm} = 12660\frac{1}{M \cdot cm}$

Damit konnten die Konzentrationen nach folgender Formel berechnet werden:

$$E_{280nm} = c\left[\frac{mol}{l}\right] \cdot \varepsilon_{280nm}\left[\frac{1}{M \cdot cm}\right] \cdot d[cm]$$

$$c\left[\frac{mol}{l}\right] = \frac{E_{280nm}}{\varepsilon_{280}}$$

Für LabA2 folgte daraus: $c\left[\frac{mol}{l}\right] = \frac{E_{280nm}}{11380}$. Analog dazu ergab sich die Konzentration für die LabA2-His-Tag-Konstrukte nach folgender Formel: $c\left[\frac{mol}{l}\right] = \frac{E_{280nm}}{12660}$.

4.4.3 Reduzierende diskontinuierliche SDS-PAGE (Disk-SDS-PAGE) nach Lämmli

Die Gelelektrophorese ist nicht nur eine wichtige Methode zur Analyse von DNA, sondern auch von Proteinen. Allerdings gibt es einige grundlegende Unterschiede zwischen diesen beiden Gelelektrophoresearten, die auf der Verschiedenheit der beiden Stoffklassen beruhen. Für die meisten Proteine sind Agarosegele zu großporig, weshalb man hier typischerweise vertikale Polyacrylamidgele verwendet. Die radikalische Kopolymerisation zwischen Acrylamid und dem Quervernetzer *N,N´*-Methylenbisacrylamid wird hierbei durch den Radikalstarter Ammoniumperoxodisulfat (APS) und den Radikalüberträger Tetramethylethylendiamin (TEMED) gesteuert. Die in der vorliegenden Arbeit verwendete Elektrophoresemethode wird als diskontinuierliche SDS-Polyacrylamidgelelektrophorese (Disk-SDS-PAGE) bezeichnet und wurde mit den Puffern nach Lämmli angewendet, was im Folgenden kurz erläutert werden soll.[69] Da Proteine im Gegensatz zu DNA ganz unterschiedliche Nettoladungen aufweisen können, bedient man sich des anionischen Detergens SDS, welches sich über hydrophobe Wechselwirkungen an die Proteine anlagert (1,4 g SDS/g Protein in einer 1%igen SDS-Lösung) und so neben der Denaturierung auch für ein konstantes Masse-zu-Ladungs-Verhältnis sorgt. Dadurch ist es möglich, alle Proteine in Richtung Anode aufzutrennen, wobei die Trennung jetzt nur noch von der entsprechenden Größe des Proteins abhängt. Weiterhin charakteristisch ist, dass mit zwei verschiedenen Gelzonen gearbeitet wird, dem sog. Sammel- und Trenngel. Im Sammelgel gilt das Elektrophoreseprinzip der Isotachophorese, die sich durch ein diskontinuierliches Puffersystem aus Tris/Chlorid (pH 6,8) im Sammelgel und Tris/Glycin (pH 8,9) im Laufpuffer auszeichnet. Nach dem Auftragen der Proben auf das Sammelgel und dem Anlegen der Spannung trifft Glycin auf den pH des Sammelgels, welcher seinem isoelektrischen Punkt sehr nahe kommt. Dadurch hat Glycin eine niedrige elektrophoretische Mobilität und wird zum Folgeion. Im Gegensatz dazu haben die Chlorid-Ionen eine hohe elektrophoretische Mobilität und werden dementsprechend zum Leition. Die in der Probe enthaltenen Proteine liegen sowohl örtlich als auch von ihrer elektrophoretischen Mobilität her zwischen Leit- und Folgeion. Durch diese Gegebenheit bildet sich ein Feldstärkegradient über die nach ihrer

elektrophoretischen Mobilität sortierten Analyten als molekulare Stapel aus und alle Ionen laufen mit der gleichen Geschwindigkeit in Richtung Anode. Dieses System regelt sich selbst, da zu schnelle oder zu langsame Ionen durch die dort herrschende kleinere oder größere Feldstärke wieder an ihren Platz befördert und somit refokussiert werden. Treffen die Proteine schließlich auf das engmaschigere Trenngel, erfahren sie plötzlich einen großen Reibungswiderstand (*stacking*), der zu einer weiteren Aufkonzentrierung führt. Das Folgeion Glycin ist aufgrund seiner Größe davon nicht betroffen, überholt die Proteine und beendet damit das Elektrophoreseprinzip der Isotachophorese. Im Trenngel (pH 8,8) liegen die Glycin-Moleküle geladen vor und sorgen dadurch für eine konstante Feldstärke, so dass jetzt eine Auftrennung der Proteine nach dem Prinzip der Zonenelektrophorese erfolgen kann. Durch die Verwendung von SDS besteht über einen weiten Größenbereich eine lineare Beziehung zwischen dem Logarithmus (log_{10}) des Molekulargewichts und der relativen Wanderungsdistanz im Gel. In der Praxis verwendet man analog zur Gelelektrophorese von DNA Längenstandards oder Protein-Leitern zur Größenbestimmung. Als Ergänzung muss noch erwähnt werden, dass der Probenauftragspuffer neben dem bereits erwähnten SDS und den schon bei der DNA verwendeten Zusätzen Glycerin und Bromphenolblau auch das reduzierende Agens DTT enthält, um vorkommende Disulfidbrücken zu öffnen.

Für eine erfolgreiche Elektrophorese mussten die verwendeten Glasplatten zunächst mit Ethanol gründlichst gereinigt werden. Anschließend wurden die Gelkammern entsprechend der Herstellerangaben zusammengebaut und mit Hilfe eines Markers die gewünschte Höhe des Trenngels markiert (1 cm Sammelgel). Das Trenngel wurde gegossen und sofort vorsichtig mit Isopropanol überschichtet. Nach ausreichender Polymerisation wurde das Isopropanol zunächst quantitativ entfernt, was durch Abblasen mit Druckluft sichergestellt wurde. Anschließend wurde das Trenngel mit dem Sammelgel überschichtet. Während das Sammelgel polymerisierte, wurden die Proteinproben vorbereitet. Dazu wurden die Proteinproben mit dem entsprechenden Volumen (3/5 Probe + 2/5 Probenauftragspuffer) des zuvor mit einer DTT-Lösung (1M) komplettierten Probenauftragspuffers versetzt und auf 95°C erhitzt (5 min). Anschließend wurden die Proteinproben kurz auf Eis gestellt, auf das Gel aufgetragen und elektrophoretisch getrennt (15mA/Gel im Sammelgel und 20mA/Gel im Trenngel). Die Auswertung der Gele erfolgte entweder mittels Comassie- bzw. Silberfärbung oder *Western-Blot.*

Die gelbasierte Analyse der heterologen Expression des LabKC-Proteins mittels SDS-PAGE nach Lämmli und *Western-Blot* wurde mit den in Tabelle 14 aufgeführten Gelen durchgeführt.

Tabelle 14: Zusammensetzung der Trenn- und Sammelgele, die im Rahmen dieser Arbeit zur Untersuchung der Proteinexpressionen und Reinigungen herangezogen wurden.

Reagenz	10 % Trenngel (Angaben pro Gel)	5 % Sammelgel (Angaben pro Gel)
Acrylamid/Bisacrylamid-Lösung (37,5:1)	1,9 mL	515 µl
Trenngel- bzw Sammelgelpuffer	1,4 mL	1,58 mL
H_2O_{dd}	2,25 mL	750 µl
10 % SDS	28,8 µl	30 µl
TEMED	14,4 µl	7,5 µl
APS-Lösung	115 µl	20 µl

4.4.4 Reduzierende diskontinuierliche SDS-PAGE (Disk-SDS-PAGE) nach Schägger und Jagow

Eine Abwandlung der SDS-PAGE nach Lämmli für kleine Proteine und Peptide (1 bis 30 kDa) wurde im Jahre 1987 von Schägger und Jagow vorgestellt.[70] Die unbefriedigende Auflösung der Lämmli-Gele im unteren Größenbereich (< 10 kDa) wird darauf zurückgeführt, dass Proteine dieser Größe im Sammelgel nicht vom überschüssigen SDS getrennt werden, was dann zu einer Verbreiterung der Banden führt.[70] Um dieses Problem zu umgehen, wurde ein neues Puffersystem, basierend auf Tricin, eingeführt. Neben den Unterschieden im Puffersystem sind auch noch weitere Unterschiede im Vergleich zum System nach Lämmli zu erwähnen. So gibt es bei den Tricin-Gelen nur einen Gelpuffer für Trenn- und Sammelgel; dafür unterscheiden sich die Puffer an Anode und Kathode voneinander.

Dieses Gelsystem wurde in der vorliegenden Arbeit zur Analyse der heterologen Expression der Polypeptidsubstrate für den *in vitro*-Assay des LabKC-Proteins herangezogen. Dazu wurden Gele verwendet, die der in Tabelle 15 beschriebenen Zusammensetzung entsprachen. Die Durchführung erfolgte in Analogie zu der in Abschnitt 4.4.3 beschriebenen Methode nach Lämmli.

Tabelle 15: Zusammensetzung der Gele für die Analyse kleiner Proteine und Peptide nach der Methode von Schägger und Jagow.

Reagenz	16,5 % Trenngel (Angaben pro Gel)	5 % Sammelgel (Angaben pro Gel)
Acrylamid/Bisacrylamid-Lösung (37,5:1)	3,33 mL	330 µl
Puffer	1,125 mL	200 µl
H_2O_{dd}	1,41 mL	1,44 mL
10 % SDS	120 µl	20 µl
TEMED	15 µl	4 µl
APS-Lösung	30 µl	10 µl

4.4.5 Coomassie-Färbung von Proteingelen

Die gängigste Methode zur Färbung von Proteingelen ist wohl die mit Coomassie Brillantblau G 250, welches auch schon zur Konzentrationsbestimmung von Proteinen nach Bradford verwendet wurde. Diese Methode ist einfach in der Durchführung und hat je nach Protein eine Nachweisgrenze zwischen 100 ng und 1 µg. Das Gel wurde dazu nach der elektrophoretischen Auftrennung an einer Ecke zur Orientierung markiert und dann für mindestens 1 h in die Färbe-Lösung gelegt. Der Gehalt an Essigsäure (10 %) in der Färbe-Lösung führt gleichzeitig zur Fixierung der getrennten Proteinbanden. Für die Entfärbung der Gele wurden diese mehrfach in Entfärbe-Lösung inkubiert bis der gewünschte Kontrast eingestellt war.

4.4.6 Silberfärbung von Proteingelen

In der vorliegenden Arbeit wurden die Gele teilweise mittels Silberfärbung gefärbt, da diese Methode mit einer Nachweisgrenze im Subnanogrammbereich deutlich sensitiver ist. Das Prinzip der Silberfärbung beruht darauf, dass nach der Fixierung der Proteine im Gel ein Reduktionsschritt mit Natriumdithionit ($Na_2S_2O_4$) erfolgt, welcher die Disulfidbrücken der Proteine zu Thiolgruppen reduziert. Diese Thiolgruppen komplexieren neben den Seitenketten von Glu und Asp Silber-Ionen, die initiiert von den funktionellen Gruppen und den Peptidbindungen der Proteine zu Silber reduziert werden. Die Zugabe des starken Reduktionsmittels Natriumthiosulfat ($Na_2S_2O_3$) in alkalischem Milieu setzt diesen Prozess fort, zunächst vor allem in der Nähe der bereits gebildeten Silberkeime. Eine pH-Änderung stoppt diesen Prozess rechtzeitig, bevor sich überall auf dem Gel metallisches Silber

ablagert. Die Nachteile dieser Methode liegen darin, dass gleiche Mengen verschiedener Proteine unterschiedlich intensive Banden erzeugen, was eine Quantifizierung verhindert. Außerdem ist die Färbung nicht spezifisch für Proteine, denn auch Nukleinsäuren, Lipide und weitere Substanzklassen können das Gel anfärben.
In der vorliegenden Arbeit wurde das *SilverQuest Silver Staining-Kit* der Firma Invitrogen, Karlsruhe, nach dem dort beschriebenen Protokoll verwendet.

4.4.7 *Western-Blot*

Die *Western-Blot*-Methode erlaubt eine spezifische Identifizierung von Proteinbanden nach einer elektrophoretischen Trennung. Dazu werden die aufgetrennten Proteinbanden auf eine Membran transferiert. Der Transfer kann mittels Tank-Blot oder wie in dieser Arbeit durch Semi-Dry-Blot erfolgen. Die anschließend auf der Membran immobilisierten Proteine sind frei zugänglich und können durch spezifische Antikörper selektiv gebunden werden. Die Auswertung erfolgt über Enzyme, die kovalent an den sekundären, seltener an den primären Antikörper gekoppelt sind. Bei den Enzymen handelt es sich um Peroxidasen, β-Galaktosidasen oder wie in dieser Arbeit um alkalische Phosphatasen. Da es sich bei der farbgebenden Reaktion um eine enzymatische Umsetzung handelt, ist die erreichte Sensitivität im Vergleich zur Coomassie-Färbung um ein Vielfaches höher. Der *Western-Blot* wurde in dieser Arbeit zur spezifischen Detektion von mit His-Tag modifizierten, heterolog exprimierten Substrat- und LabKC-Proteinen herangezogen.
Die elektrophoretisch aufgetrennten Proteine wurden mittels Semi-Dry-Blotverfahren auf eine Nitrozellulose-Membran transferiert. Dazu wurden das Gel nach der Elektrophorese (analog zu Abschnitt 4.4.3 und 4.4.4) zunächst in Blot-Puffer (20 min) zusammen mit der Membran und dem Filterpapier getränkt und anschließend der Blot wie in Abbildung 21 gezeigt aufgebaut. Dabei wurde darauf geachtet, dass keine Luftblasen zwischen den Schichten eingeschlossen wurden.

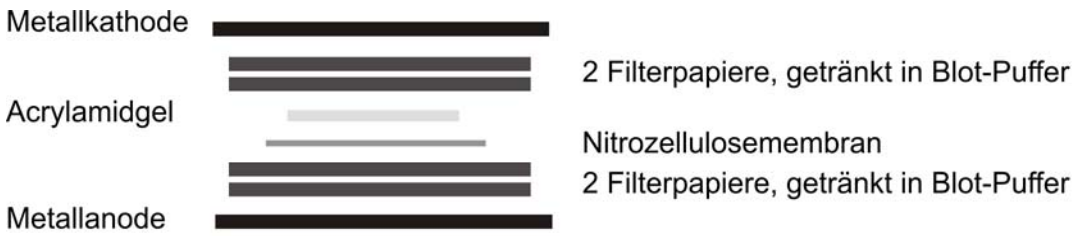


Abbildung 21: Schematischer Aufbau des Semi-Dry-Blots.

Die Blotbedingungen wurden nach der Faustregel (0,8 mA/cm^2) abgeleitet und ergaben sich so zu einer konstanten Stromstärke I = 50 mA (60 min). Nach Beendigung des Transfers wurde die Blotmembran für 90 min in mit BSA (0,8 g) komplettiertem TNT-Puffer (25 ml) zur Verhinderung von unspezifischer Bindung des primären Antikörpers gesättigt. Zur Inkubation des ersten Antikörpers wurde Anti-His-Antikörper (7 µl) zu mit 3 % BSA komplettiertem TNT-Puffer (30 ml) zugesetzt. Dieser Ansatz wurde für 16 h im 4°C-Kühlraum auf dem Taumeltisch inkubiert. Anschließend wurde die Lösung abgesaugt und die Membran für 3 x 5 min jeweils mit TNT-Puffer (30 ml) gewaschen. Der sekundäre Anti-Maus-IgG-Antikörper (7 µl) wurde in TNT-Puffer (30 ml) verdünnt und die Blotmembran darin für 90 min bei RT inkubiert. Nach Beendigung der Inkubation wurde die Lösung abgesaugt und 3 x mit TNT-Puffer (30 ml) für 5 min gewaschen. Anschließend wurde mit AP-Puffer (15 ml) äquilibriert und dieser schließlich durch Färbe-Lösung (15 ml) ersetzt. Der Fortschritt der Färbung wurde verfolgt und bei ausreichender Bandenintensität durch Zugabe von H_2O_d gestoppt. Nach mehrmaligem Waschen mit H_2O_d wurde die Blotmembran photographisch dokumentiert.

4.4.8 Heterologe Expression mit dem *Champion pET SUMO-System*

Der Begriff heterologe Expression beschreibt die Expression eines Gens in einem Organismus, der nicht dem Wildtyp-Organismus entspricht. Für viele Fragestellungen ist es wichtig, dass ein möglichst hohes Expressionsniveau eines fremden Gens erreicht werden kann. Dieses Problem hat im Laufe der Zeit zur Entwicklung von Expressionsvektoren geführt. Diese sind nicht nur dafür geeignet, das gewünschte Gen zu klonieren, sondern enthalten auch die notwendigen regulatorischen Sequenzen, die einen Einfluss des Experimentators auf die Expression dieses Gens ermöglichen.

In dieser Arbeit wurde aus Gründen, auf die in Abschnitt 5.3.1 näher eingegangen werden soll, mit dem *Champion pET SUMO Protein Expression-System* von Invitrogen, Karlsruhe gearbeitet.

Bei dem pET SUMO-Vektor handelt es sich wie bei den meisten Expressionsvektoren um ein low-copy-Plasmid, dessen Kopienzahl durch den vorhandenen pBR322-Origin und das im *ROP*-ORF kodierte Protein (*ROP* = *repressor of primer*) reguliert wird. Neben dem Kanamycin-Gen für die Selektion ist außerdem ein Lac-Repressor (*lacI*) kodiert. Das über TA-Klonierung eingebrachte heterolog zu exprimierende Protein steht unter der Kontrolle des T7-Promotors. Die üblicherweise starke Initiation am T7 RNA-Polymerase-Promotor bedingt, dass dieser Promotor nicht ganz abzuschalten ist (*leaky*). Aufgrund dessen ist stromabwärts des T7-Promotors noch ein zusätzlicher *lac*-Operator (*lac*O) eingefügt, um die Hintergrund-Expression weiter zu minimieren. Die für die Expression des gewünschten Proteins benötigte T7 RNA-Polymerase wird von dem verwendeten *E. coli* BL21 (DE3) zur Verfügung gestellt.

In diesem Stamm ist die Expression der T7 RNA-Polymerase unter der Kontrolle des IPTG-induzierbaren *lac*UV5-Promotors. Ein weiterer Vorteil von *E. coli* BL21 ist, dass dieser Stamm neben einem mutierten *lon*-Gen, was normalerweise für eine cytoplasmatische Protease kodiert, auch noch ein defektes *omf*T-Gen trägt. Dieses Gen kodiert eine Protease der äußeren Membran und spaltet unter anderem die T7 RNA-Polymerase. Diese komplexe Regulation der Expression führt dazu, dass die Zellen zunächst ohne detektierbaren Expressionshintergrund eines möglicherweise toxischen Produkts amplifiziert werden können, um dann durch Zugabe des Induktors große Mengen an heterologem Protein zu exprimieren. Die TA-Klonierung in den bereits als linearisierter Vektor gelieferten pET SUMO führt zu einem *N*-terminal, um 119 Aminosäuren verlängerten, Expressionsprodukt. Die zusätzlich angehängten Aminosäuren erhöhen nicht nur das Expressionslevel und die Löslichkeit des gebildeten Proteins, sondern erlauben auch durch den darin enthaltenen His_6-Tag eine einfache Aufreinigung mittels IMAC (immobilisierte Metallchelat-Affinitätschromatographie). Neben dem His-Tag ist zusätzlich eine Proteasespaltstelle (SUMO) vorhanden, die nach der Reinigung die Entfernung des *N*-terminalen His-Tags ermöglicht.

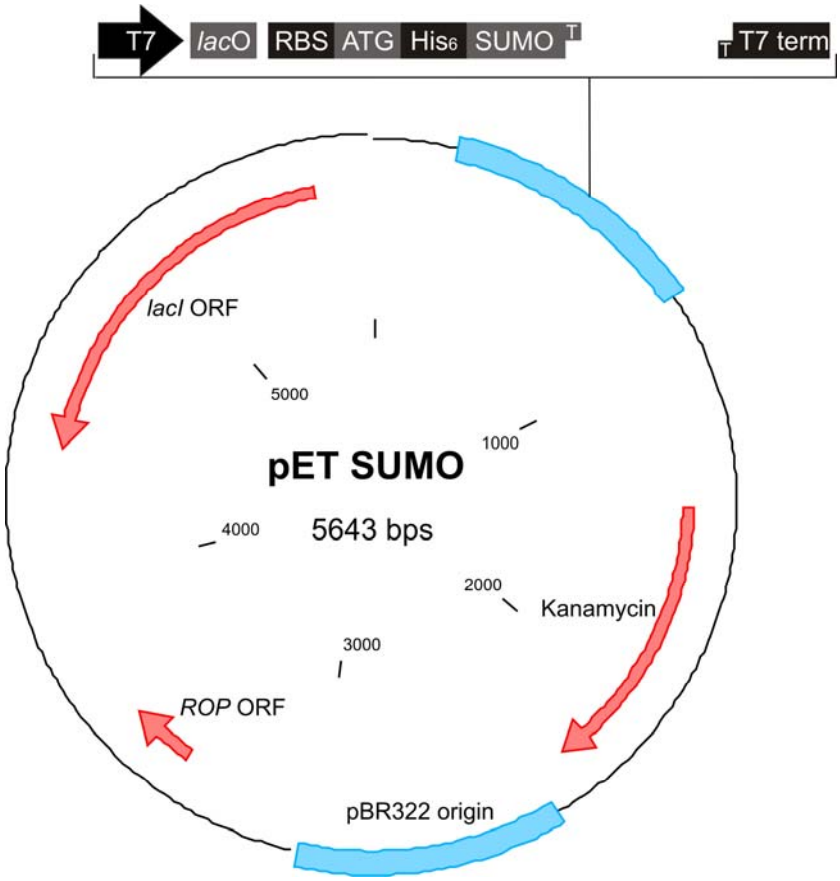


Abbildung 22: Vektorkarte des zur heterologen Proteinexpression herangezogenen Vektors pET SUMO von Invitrogen.

Das beschriebene Expressionssystem wurde in der vorliegenden Arbeit für die Bereitstellung des LabA2-Substrates und des Enzyms LabKC für den *in vitro*-Assay eingesetzt, wobei aufgrund der unterschiedlichen Problemstellungen die optimalen Expressionsbedingungen für die Substrate und LabKC unabhängig voneinander optimiert wurden.

Für die Substrate wurden zunächst Vorkulturen (40 mL TY-Kan; 1 % Glukose) mit Kryokulturen der die entsprechenden Expressionsplasmide tragenden Produzentenstämme beimpft (200 µl) und bei 37°C kultiviert (16 h; 150 rpm). Ausgehend von diesen Vorkulturen wurden dann die Hauptkulturen (100 mL TY-Kan; 1 % Glukose) beimpft (1 %) und bis zu einer $OD_{600} \approx 0{,}8$ kultiviert (37°C; 150 rpm). Die Induktion der Expression erfolgte dann durch Zugabe einer IPTG-Lösung (1 M) zu einer Endkonzentration von 1 mM. Die Expression der gewünschten Produkte (37°C; 4 h; 150 rpm) wurde durch Pelletieren (10 min, 4200 g) in einer auf 4°C vorgekühlten Zentrifuge beendet. Die Pellets wurden dann entweder direkt aufgeschlossen oder bis zur Weiterverarbeitung bei -80°C eingefroren.

Da das Enzym LabKC nach der Expression und Reinigung in einen *in vitro*-Assay eingesetzt werden sollte, wurden die Bedingungen für die Expression hier gesondert optimiert. Mehrfache Kultivierung unter verschiedenen Bedingungen erbrachten folgende optimale Parameter. Die Vorkultur (50 mL TY-Kan; 1,8 % Glukose) wurde mit einer Kryokultur (200 µl) des das entsprechende Expressionsplasmid tragenden Produzentenstammes beimpft und bei 28°C kultiviert (16 h; 150 rpm). Ausgehend von der Vorkultur wurden dann die Hauptkulturen (je 100 mL TY-Kan; 1,8 % Glukose) beimpft (1 %) und bis zu einer $OD_{600} \approx 0{,}6$ kultiviert (15°C; 150 rpm). Die Induktion der Expression erfolgte dann durch Zugabe einer IPTG-Lösung (1 M) zu einer Endkonzentration von 0,1 mM. Die Expression des LabKC-Proteins (15°C; 72 h; 150 rpm) wurde durch Pelletieren (10 min, 4200 g) in einer auf 4°C vorgekühlten Zentrifuge beendet. Das Pellet wurde danach sofort aufgeschlossen und weiter aufgearbeitet.

4.4.9 Proteinreinigung mittels IMAC

Die Methode der immobilisierten Metallchelat-Affinitätschromatographie (IMAC) ist heute für die Isolierung rekombinanter Proteine sehr weit verbreitet. Dieses Prinzip basiert darauf, dass an das zu exprimierende Peptid oder Protein *N*- oder *C*-terminal ein $His_{6\text{-}8}$-Tag angehängt wird, der eine schnelle Aufreinigung über Säulen mit multivalenten Kationen wie z.B. Ni^{2+}-Ionen ermöglicht.[71] Bei der in dieser Arbeit als feste Phase eingesetzten Ni-NTA-Agarose handelte es sich um eine Nitrilotriessigsäure-Gruppen tragende Agarosebasis, die Ni^{2+}-Ionen binden. An diesen Ni^{2+}-Ionen verbleiben noch weitere Stellen zur Ausbildung von koordinativen Bindungen mit His-Resten von Proteinen (s. Abbildung 23).

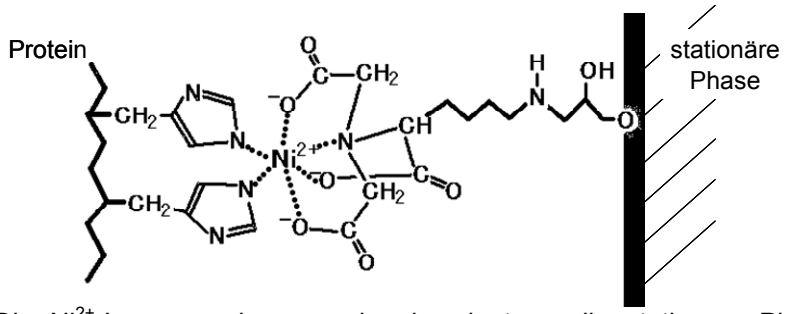


Abbildung 23: Die Ni^{2+}-Ionen werden von den kovalent an die stationäre Phase gebundenen Nitrilotriessigsäureresten komplexiert. Freie Bindungsstellen des Nickels sind in der Lage, koordinativ Poly-His-Reste zu binden.

Durch die reversible Bindung von Peptiden und Proteinen mit His-Tag kann eine erste Reinigung und Aufkonzentrierung aus einem komplexen Proteingemisch erreicht werden. Die Elution erfolgt durch Änderung des pHs oder mit steigenden Konzentrationen an freiem Imidazol.

Die Reinigung der heterolog exprimierten Substrate und des LabKC-Proteins erfolgte in der vorliegenden Arbeit mittels IMAC. Um proteolytischen Abbau zu minimieren, wurden alle Schritte möglichst schnell und unter ständiger Kühlung bei 4°C durchgeführt. Für die Verfolgung der Aufreinigung wurden auf jeder Stufe Proben für eine Proteinbestimmung, SDS-PAGE und *Western-Blot* genommen. Die Bakterien aus den generierten Pellets (s. Abschnitt 4.4.8) wurden nach der Resuspension in Lysepuffer zunächst in der French-Press aufgeschlossen. Nach Verdau mit einer Spatelspitze DNaseI (20 min auf Eis) wurden Zelltrümmer und ungelöste Proteine durch Zentrifugation (15 min, 25000 g) entfernt und der Überstand auf eine zuvor mit Lyse-Puffer äquilibrierte Ni-NTA-Säule (4 ml Bettvolumen) aufgetragen. Die Säule wurde danach mit Lyse- und Waschpuffer so lange gewaschen, bis die Elutionsfraktionen keine Färbung mehr im Bradford-Schnelltest zeigten (10 µl Eluat + 90 µl Bradford-Lösung). Die Elution erfolgte mit dem Elutionspuffer1 in 1 mL Fraktionen. Abschließend wurde mit Elutions-Lösung 2 nachgewaschen, wobei hier kaum noch Protein von der Säule eluierte. Die erhaltenen Fraktionen wurden mittels SDS-PAGE und *Western-Blot* untersucht.

4.4.10 Abspaltung des *N*-terminalen His-Tags mit Hilfe der SUMO-Protease und Aufreinigung der Produkte

Die Abspaltung des *N*-terminalen His-Tags von Peptiden und Proteinen, die mit dem pET SUMO-Vektor exprimiert wurden, erfolgt durch die SUMO-Protease. Dabei handelt es sich um eine Protease, die ihre Spaltstelle nicht an einer bestimmten Aminosäuresequenz erkennt, sondern aufgrund der Tertiärstruktur der SUMO-Domäne, die in den Konstrukten zwischen dem His-Tag und dem Peptid/Protein insertiert ist. Für die Spaltreaktion sind jedoch einige Bedingungen zu erfüllen. So darf z.B. der Imidazolgehalt der Lösung nicht

größer als 150 mM sein. Diese Tatsache ist aber auch noch aus einem anderen Grund wichtig. Die weitere Vorgehensweise nach der Spaltreaktion sieht eine zweite Ni-Säule zur Aufreinigung des gewünschten Peptids oder Proteins aus dem Spaltansatz vor. Um dabei die restlichen His-Tag tragenden Verunreinigungen an die Säule zu binden und somit zu entfernen muss die Imidazolkonzentration sehr gering sein.
In der vorliegenden Arbeit wurde zur Vorbereitung der SUMO-Spaltung und anschließenden Aufreinigung der Spaltprodukte ein Dialyseschritt eingeführt. Dabei wurden die Fraktionen der ersten Ni-Säule aufgrund von Ergebnissen der SDS-PAGE vereinigt und gegen einen Puffer dialysiert, der den richtigen Salzgehalt für die Proteasespaltung einstellte. Für die Dialyse der Substratkonstrukte (ca. 17,7 kDa) wurde in der vorliegenden Arbeit die Dialysemembran Spectra/Por® (MWCO 2000) verwendet. Für das LabKC-Konstrukt (108,4 kDa) erfolgte die Dialyse in einem Servapor®(MWCO 12000-14000)-Schlauch, der für den Gebrauch durch Aufkochen (15 min) in einer EDTA-Lösung (40 mM) vorbereitet wurde. Die Dialyse erfolgte bei 4°C unter leichtem Rühren, wobei der erste Liter Dialysepuffer nach 16 h gegen einen zweiten für 8 h ausgetauscht wurde. Die so erhaltenen Dialysate konnten somit direkt zur Spaltung in den in Tabelle 16 gezeigten Ansatz eingesetzt werden.

Tabelle 16: Zusammensetzung der Ansätze für die SUMO-Spaltung.

Volumen	Komponente	Bemerkung
100 µl	Dialysat	Laut Herstellervorschrift 20 µg Konstrukt pro Ansatz. Aufgrund der anschließenden Aufreinigung über eine Ni-Säule wurde hier das Konstrukt meist im Überschuss eingesetzt.
20 µl	10 x Puffer (- salt)	Die Zusammensetzung des Dialysepuffers hatte die optimale Konzentration an NaCl (150 mM) bereits vorgelegt, weshalb in diesem Fall nur noch der Puffer ohne Salz eingesetzt wurde.
10 µl	SUMO-Protease	
70 µl	H_2O_{dd}	

Die Ansätze der SUMO-Spaltung wurden bei 30°C inkubiert (16 h). Nach der Spaltung erfolgte die Reinigung des gewünschten Substrates bzw. des LabKC-Proteins durch eine zweite IMAC. Für diesen Schritt wurden wegen der geringeren Volumina und leichteren Handhabung Ni-NTA-Spin-Säulchen von Qiagen verwendet. Diese wurden vor der Auftragung der Spaltreaktion mit Lysepuffer äquilibriert (600 µl). Die Elution des Durchflusses mit dem gewünschten Produkt erfolgte durch Zentrifugation (3 min, 1000 g). Die hier beschriebenen Schritte wurden ebenfalls durch Gelelektrophorese und *Western-Blot* untersucht.

4.4.11 GluC-Spaltung der Präpropeptide zur massenspektrometrischen Untersuchung und Charakterisierung

Bei der Bestimmung von Aminosäuren-Sequenzen wurde der klassische Edman-Abbau inzwischen durch die Methode der Massenspektrometrie ergänzt bzw. abgelöst. Die Sequenzierung durch sog. MSMS-Spektren von Peptiden gehört heutzutage schon fast zur Routineanalytik. Für eine Sequenzanalyse müssen die Peptide oder Proteine jedoch zuvor durch einen proteolytischen Verdau in kleinere Peptide von einer Länge zwischen acht bis zwölf Aminosäuren gespalten werden. Die Sequenz dieser Peptide kann dann durch die für Peptide charakteristische Fragmentierung am Peptidrückgrat mittels MSMS-Experimenten bestimmt werden.[72] Die für diese Zwecke am häufigsten verwendete Endoproteinase ist Trypsin, da sie nach den basischen Aminosäuren Lys und Arg *C*-terminal schneidet und durch die natürliche Häufigkeit dieser beiden Aminosäuren durchschnittlich Nonapeptide erzeugt, die zudem in einem massenspektrometrisch außerordentlich gut erfassbaren Massenbereich ionisieren. Durch die positive Ladung am *C*-Terminus dieser Peptide kann die Sequenzierung dann meist anhand der y-Serie erfolgen.

Die massenspektrometrische Sequenzierung von Peptiden sollte in dieser Arbeit sowohl zur Charakterisierung und zum Nachweis der isolierten Präpropeptide als auch zur Auswertung der *in vitro*-Umsetzungen der Substrate mit dem LabKC-Protein eingesetzt werden. Bei der Betrachtung der Sequenz des LabA2-His-Tag-Konstrukts und LabA2 (s. Abbildung 24) wurde jedoch schnell deutlich, dass in diesem Fall Trypsin nicht als geeignete Endoproteinase in Frage kommt.

LabA2-Präpropeptid-His-Tag-Konstrukt:

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFF
KIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG
MASILELQNLDVEHARGENRSDWSLWECCSTGSLFACC

Trypsin-Verdau von LabA2:

MASILELQNLDVEHAR GENR SDWSLWECCSTGSLFACC

GluC-Verdau von LabA2:

MASILE LQNLDVE HARGE NRSDWSLWE CCSTGSLFACC

Abbildung 24: *in silico*-Verdaus des LabA2-His-Tag-Konstrukts mit besonderem Augenmerk auf das Präpropeptid (blau-rot). Ein Trypsin-Verdau würde das Präpropeptid nur in zwei sehr große und ein kleines Fragment spalten, wobei die großen Peptidfragmente massenspektrometrisch nur sehr schwer detektierbar gewesen wären. Das entstehende Fragmentmuster eines GluC-Verdaus war im Vergleich dazu deutlich besser geeignet.

Die einzige Endoproteinase, die ein geeignetes Fragmentmuster lieferte, war GluC. Die Substrate werden durch diese Serinprotease in fünf Peptide, die Substratkonstrukte in 17 Peptide gespalten.

Für die proteolytische Spaltung der Substrate und der Substrakonstrukte wurden zu den entsprechenden Lösungen GluC-Puffer (1 Volumen) und GluC (3 µl einer 1 µg/µl Lösung) zu einem Gesamtvolumen von 43 µl zusammengegeben. Die Verdauansätze wurden nach der Inkubation (mindestens 3 h) bei 25°C in Vials überführt und mittels HPLC-MS und -MSMS vermessen.

4.4.12 HPLC-ESI-MS und -MSMS von GluC-Verdauansätzen

Die Untersuchung der GluC-Verdaus der Substrate und Substratkonstrukte erfolgte mittels HPLC-ESI-MS und -MSMS. Dafür stand ein LC-MS-System bestehend aus einer Kapillar-HPLC der 1100-Serie von Agilent, Waldbronn, und einer QTRAP 2000 von Applied Biosystems, Darmstadt, zur Verfügung. Für die Chromatographie des Peptidgemischs wurde klassischerweise eine C18 *reversed-phase*-Säule (*Jupiter*; 4u, 300Å, 15 x 1 mm) der Firma Phenomenex, Aschaffenburg, verwendet. Als mobile Phase diente Wasser/0,1 % HCOOH (Laufmittel A) und Acetonitril/0,1 % HCOOH (Laufmittel B). Für die chromatographische Trennung der Peptide wurde der in Tabelle 17 beschriebene Gradient eingesetzt.

Tabelle 17: Gradient für die Auftrennung der Peptidgemische aus den GluC-Verdauansätzen.

Zeit [min]	Laufmittel A [%]	Laufmittel B [%]	Flussrate Kapillar-HPLC
0	95	5	50 µl/min
60	80	20	50 µl/min
61	0	100	50 µl/min
67	0	100	50 µl/min
68	95	5	50 µl/min
75	95	5	50 µl/min

Die Massendetektion für MS-Spektren erfolgte in einem Bereich von 300 bis 1400 *m/z* im positiven EPI-Scanmodus (*enhanced product ion*) unter Verwendung folgender Parameter: CUR 30; IS 5500; TEM 350; GS1 30; GS2 65; DP 40; EP 10; CE 10; Lit fill-time dynamic.
Für die Aufnahme von MSMS-Spektren im EPI-Modus (*enhanced product ion*) wurden folgende Parameter gewählt: CUR 30; IS 5500; TEM 350; GS1 30; GS2 65; DP 40; EP 10; CE 35; CES 10; Lit fill-time 400 ms. Für die Analysen wurden jeweils 2 µl der entsprechenden Proben injiziert.

5 Ergebnisse und Diskussion

5.1 Identifizierung des Biosynthese-Genclusters der Labyrinthopeptine

Die Grundlage der Arbeit war die Auffindung und Sequenzierung des Biosynthese-Genclusters der Labyrinthopeptine. Deshalb musste zu Beginn der Arbeit ein Weg zur Isolierung genomischer DNA gefunden werden, um daraus eine Cosmidbank erstellen und Sonden für ein Screening dieser Bank generieren zu können. Die Herstellung geeigneter spezifischer Sonden stellte eine große Herausforderung dar, da sich die Kenntnisse anfangs nur auf einen Teil der 18 Aminosäuren langen Primärsequenz von Labyrinthopeptin A2 beschränkten. Mit den generierten Sonden wurde dann die Cosmidbank untersucht und anschließend eines der positiv-getesteten Cosmide sequenziert.

5.1.1 Isolierung genomischer DNA aus *Actinomadura namibiensis*

Aufgrund der literaturbekannten Lysozymresistenz vieler Stämme der Gattung Actinomadura [73,74] musste zunächst ein Weg zur Isolierung genomischer DNA gefunden werden. Zur Kontrolle der Isolierungen wurde eine Standard-PCR (Annealingtemperatur: 57°C; Elongationsdauer: 2 min) mit dem spezifisch dafür abgeleiteten Primerpaar 16SrDNA ActinoFw - 16SrDNA ActinoRev durchgeführt. Die Sequenz der 16S rDNA, die auch zur taxonomischen Einordnung des Stammes grundlegend beigetragen hatte, wurde zur Ableitung der Primer herangezogen.[45] Durch Variation und Optimierung von Medien, Medienzusätzen (Glycin), Kultivierungs- und Aufschlussbedingungen war es schließlich möglich, genomische DNA zu isolieren (s. Abschnitt 4.2.2.7). Dabei stellten sich die in Abschnitt 4.1.1 beschriebenen Bedingungen mit einer Kultivierung der Vorkultur in Medium 5294 und der Hauptkultur in CRM-Medium mit einem Zusatz von 0,5 % Glycin als ideal heraus. Die Zugabe von Glycin hat einen verstärkten Einbau in die Zellwand zur Folge, was in diesem Fall durch eine Sensibilisierung gegenüber Lysozym beobachtet werden konnte und ist ein bekanntes Mittel zur Vereinfachung des Zellaufschlusses von vielen Gram-positiven Bakterien.[57]

Weitere Applikationen erforderten die Isolierung hochmolekularer genomischer DNA aus *Actinomdura namibiensis*, wozu ein Protokoll basierend auf einer Phenol/Chloroform-Extraktion (s. Abschnitt 4.2.2.8) herangezogen wurde. Das Gel in Abbildung 25 dokumentiert die erfolgreiche Isolierung hochmolekularer genomischer DNA.

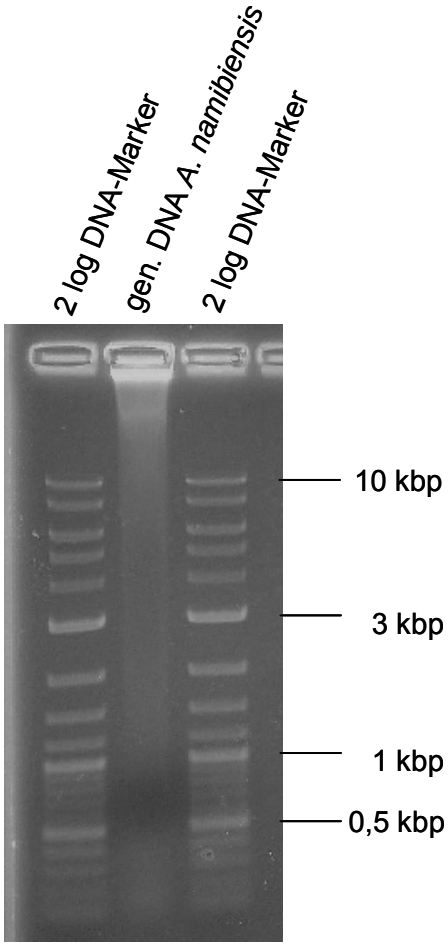


Abbildung 25: Gelelektrophorese der Isolierung genomischer DNA aus *A. namibiensis*. Die Bande der erfolgreich isolierten genomischen DNA hängt in der mittleren Tasche und konnte aufgrund der Größe nicht in das 1%ige Agarosegel einwandern.

5.1.2 Herstellung von Cosmidbank und Hybridisierungsfilter

Für die Identifizierung des Labyrinthopeptin-Genclusters wurde eine auf pWEB (EPICENTRE) basierende Cosmidbank in Zusammenarbeit mit der Firma Agowa GmbH, Berlin, erstellt. Dabei wurde die in Abbildung 26 schematisch dargestellte Vorgehensweise verfolgt.

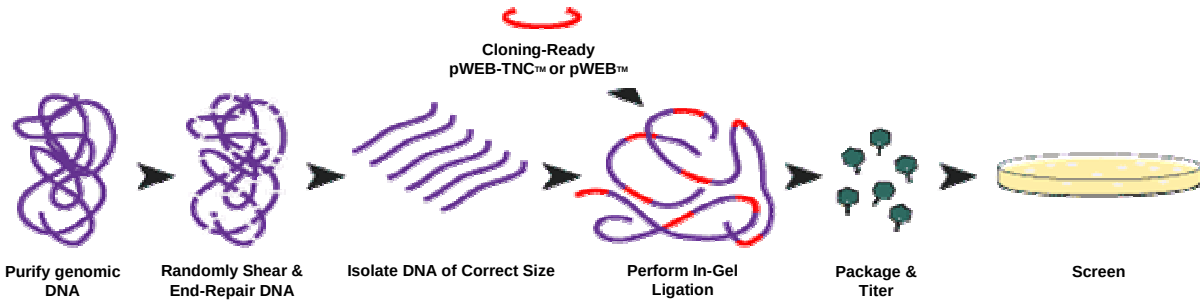


Abbildung 26: Übersicht über die Herstellung einer Cosmidbank mit dem pWEB-Vektor der Firma EPICENTRE, Madison (Abbildung aus Produktbeschreibung von **www.epibio.com**).

Die hochmolekulare genomische DNA wurde durch mehrfache Passage durch eine Spritze so geschert, dass hauptsächlich Fragmente mit einer durchschnittlichen Größe von 35 kb entstehen. Die gelgereinigten DNA-Fragmente dieser Größe wurden zur *blunt-end* Klonierung vorbereitet und mit dem Cosmidvektor pWEB ligiert. Nach erfolgreicher Verpackung in Phagen wurde in *E. coli* EPI100 transduziert und auf Selektionsmedium ausplattiert. Für eine Koloniehybridisierung wurden 2304 Klone (6*384) gepickt und zur internen Kontrolle jeweils doppelt auf einen Hybridisierungsfilter systematisch gespottet (s. Abbildung 27). Die Hybridisierungsfilter wurden von der Firma RZPD, Berlin, erstellt. Durch die Bearbeitung von insgesamt 2304 Klonen mit einer durchschnittlichen Insertgröße von 35 kb entspricht das bei einer angenommenen Genomgröße von 8 Mb einer Abdeckung von 10,0.

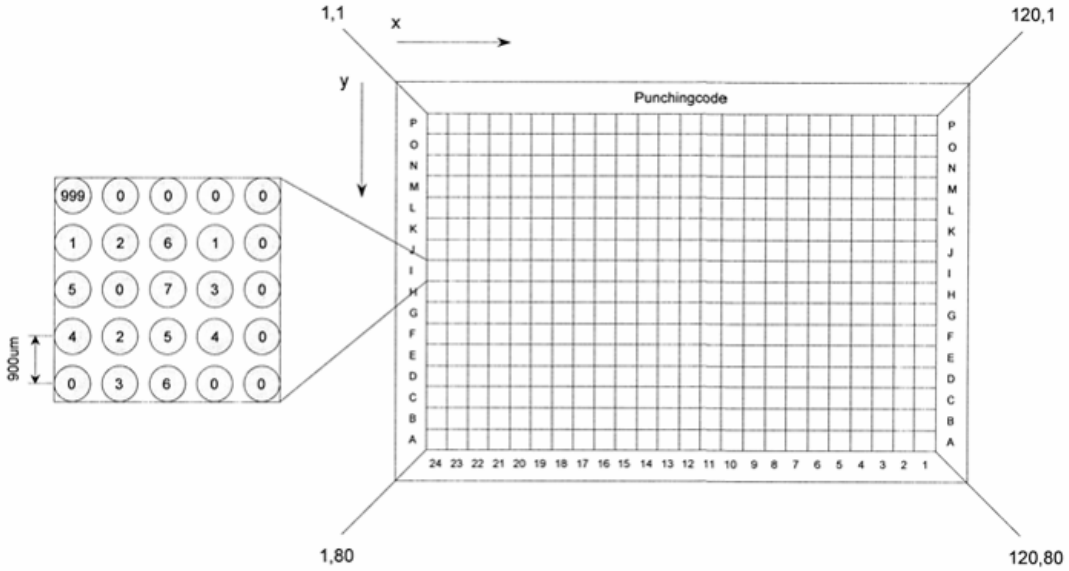


Abbildung 27: Belegungsplan eines Hybridisierungsfilters. Ein Filter hat 384 Quadranten (P24 bis A1), in die jeweils maximal 25 Spots gesetzt werden können. Das entspricht einer maximalen Beladung einer Filtereinheit mit 9600 Einzelspots. Zur Orientierung wird in die Mitte (7) jedes Quadranten ein Spot mit Tinte gesetzt. Mit einem 384-well-Stempel wird jeder Cosmidklon (aus den 384-well-Platten 1-6) jeweils zweimal gespottet. Durch die Lage und Anordnung der Signale kann dann der jeweilige Cosmidklon auf der Platte ermittelt werden. Die Positionen 0 sind jeweils nicht besetzt.

5.1.3 Erstellung eines *Codonusage* für die Gattung Actinomadura

Für die Ableitung spezifischer Primer und die Generierung homologer Sonden ausgehend von Aminosäuresequenzen müssen diese zunächst in den Nukleinsäurecode übersetzt werden. Für 20 proteinogene Aminosäuren stehen 61 kodierende Basentripletts zur Verfügung, was bedeutet, dass manche Aminosäuren durch mehr als ein Triplett kodiert werden. Z.B. werden Leu, Ser und Arg durch jeweils sechs unterschiedliche Tripletts (degeneriert), Trp und Met nur durch ein Triplett kodiert. Aus diesem Grund gibt es für Aminosäuren, die durch mehr als ein Triplett kodiert werden, eine stammspezifische Verteilung der Häufigkeit, mit der die unterschiedlichen Tripletts in Genen vorkommen (*Codonusage*) und für die Translation benutzt werden. Die Erstellung eines stammspezifischen *Codonusage* basiert auf bekannten Gensequenzen dieses Stammes und ist umso genauer, je mehr Sequenzinformation bekannt ist. Für *A. namibiensis* war zu Beginn des Projekts außer der Sequenz der 16S rDNA keine Sequenz bekannt, weshalb die Erstellung eines stammspezifischen *Codonusage* nicht möglich war. Deshalb wurde ein *Codonusage* auf Gattungsebene erstellt, wobei die in Tabelle 18 aufgeführten, im Januar 2006 verfügbaren Daten einbezogen wurden. Wie das 16S rDNA-Dendrogramm in Abbildung 28 zeigt, ist *A. hibisca* phylogenetisch *A. namibiensis* am nächsten und wurde deshalb bei der Erstellung des *Codonusage* zusätzlich gesondert betrachtet.

Tabelle 18: Actinomadura-Stämme, von denen *Codonusages* (Internetseite: **http://www.kazusa.or.jp**) verfügbar waren und die zur Erstellung eines gattungspezifischen *Codonusage* herangezogen wurden.

Stamm	Anzahl der Codons, die in die Statistik eingingen
Actinomadura verrucusospora	3381 Codons aus fünf Proteinen
Actinomadura verrucusospora R39	305 Codons
Actinomadura sp.	539 Codons
Actinomadura hibisca	2564 Codons aus elf Proteinen
Actinomadura madurae	3292 Codons aus fünf Proteinen

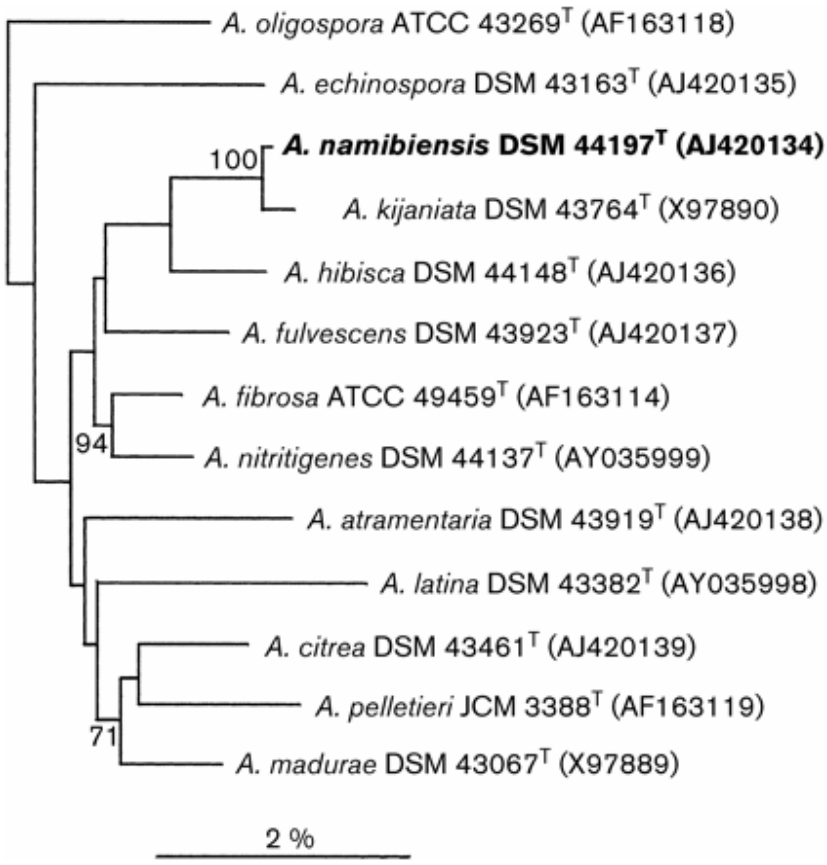


Abbildung 28: Das Dendrogramm der 16S rDNA verschiedener Actinomadura-Stämme zeigt, dass *A. hibisca* phylogenetisch *A. namibiensis* sehr nahe steht.[45]

Mit diesen *Codonusages* anderer Stämme der Gattung Actinomadura wurden dann die in Tabelle 19 dargestellten relativen Häufigkeiten der einzelnen Basentripletts berechnet, wobei sich die Prozentangaben immer auf die einzelnen Aminosäuren beziehen.

Tabelle 19: *Codonusage* der Gattung Actinomadura, die aus den *Codonusages* der in Tabelle 18 aufgeführten Stämme erstellt wurde. In der Spalte der relativen Häufigkeiten der Gattung Actinomadura ist mit Schrägstrich getrennt die Prozentzahl des Stammes *A. hibisca* angegeben. Zum Vergleich sind die Werte für Streptomyceten angegeben.[57] Die Häufigkeiten, bei denen die Unterschiede zwischen der Gattung Actinomadura und dem Stamm *A. hibisca* um mehr als 10 % differieren, wurden farblich hervorgehoben.

Amino-säure	Codon	relative Häufigkeiten Actinomadura	relative Häufigkeiten Streptomyces	Amino-säure	Codon	relative Häufigkeiten Actinomadura	relative Häufigkeiten Streptomyces
Phe	TTT	0,0/0,0	1,6	Ser	TCT	0,3/1,6	1,1
	TTC	100,0/100,0	98,4		TCC	36,9/20,2	39,1
Leu	TTA	0,0/0,0	0,4		TCA	1,9/0,0	2,4
	TTG	1,6/2,9	2,6		TCG	31,2/39,5	27,4
Tyr	TAT	2,2/0,0	5,0	Cys	TGT	1,1/0,0	12,0
	TAC	97,8/100,0	95,0		TGC	98,9/100,0	88,0
Stop	TAA	5,8/9,1	4,0	Stop	TGA	72,4/81,8	81,0
	TAG	21,8/9,1	15,0	Trp	TGG	100,0/100,0	100,0
Leu	CTT	0,6/1,3	2,0	Pro	CCT	0,8/0,6	2,7
	CTC	34,6/28,5	38,5		CCC	50,0/36,0	42,4
	CTA	0,2/0,4	0,3		CCA	0,2/0,0	1,6
	CTG	62,9/66,9	56,1		CCG	49,2/63,4	53,3
His	CAT	4,1/0,0	6,5	Arg	CGT	5,4/1,8	7,3
	CAC	95,9/100,0	93,5		CGC	47,7/57,9	45,0
Gln	CAA	4,1/4,0	6,1		CGA	0,7/1,4	3,7
	CAG	95,9/96,0	93,9		CGG	40,1/36,1	37,7
Ile	ATT	0,2/0,0	4,4	Thr	ACT	0,4/0,0	2,2
	ATC	99,6/100,0	92,0		ACC	65,4/61,9	65,2
	ATA	0,2/0,0	3,6		ACA	0,4/1,5	2,4
Met	ATG	100,0/100,0	100,0		ACG	33,7/36,6	30,2
Asn	AAT	2,2/2,7	4,6	Ser	AGT	0,8/0,8	2,9
	AAC	97,8/97,3	95,4		AGC	30,3/37,9	27,1
Lys	AAA	1,0/2,6	5,8	Arg	AGA	0,3/0,0	1,1
	AAG	99,0/97,4	94,2		AGG	5,8/2,8	5,2
Val	GTT	0,6/0,8	2,8	Ala	GCT	0,7/0,8	2,3
	GTC	57,4/49,1	56,0		GCC	54,4/48,9	58,7
	GTA	0,4/0,4	2,5		GCA	1,6/0,5	4,3
	GTG	41,5/49,6	38,8		GCG	43,4/49,7	34,7
Asp	GAT	3,5/4,6	4,7	Gly	GGT	7,9/2,6	8,7
	GAC	96,6/95,4	95,3		GGC	60,1/73,1	64,2
Glu	GAA	11,2/10,1	17,7		GGA	4,7/1,3	8,9
	GAG	88,8/89,9	82,3		GGG	27,4/23,1	18,2

5.1.4 Ableiten von DNA-Sonden und Screening von genomischen DNA-Verdaus sowie der Cosmidbank

5.1.4.1 Allgemeine Betrachtungen

Die Identifizierung von Genclustern, die für die Biosynthese bakterieller Sekundärmetabolite verantwortlich sind, verläuft meist nach demselben Schema. Ausgehend von der genomischen DNA des Produzenten wird eine Cosmidbank erstellt, die dann auf Hybridisierungsfilter gespottet ein Screening mit DNA-Sonden ermöglicht. Die dabei eingesetzten Sonden sollen möglichst aus einem >500 bp großen Fragment aus dem gesuchten Cluster bestehen (homologe Sonde). Die Generierung einer solchen Sonde erfolgt immer mittels PCR. Die dafür notwendigen Primer werden an hoch homologen Bereichen (5 bis 8 Aminosäuren) von Proteinalignments von Proteinen nahe verwandter Gencluster abgeleitet. Dieses Vorgehen basiert auf der Annahme, dass diese Proteine auch im gesuchten Gencluster vorkommen. Das Rückübersetzen der Aminosäuresequenzen in Nukleotide hat eine Degenerierung der Primer zur Folge. Gelingt es nicht, auf diese Weise eine homologe Sonde zu generieren, besteht noch die Möglichkeit der Herstellung einer heterologen Sonde. Dazu werden hoch homologe Bereiche aus strukturverwandten Clustern phylogenetisch nahe verwandter Stämme mittels PCR amplifiziert. Für verschiedene Substanzklassen wie z.B. nicht-ribosomale Peptide und Polyketide sind in der Literatur inzwischen Standardprimer zur Erzeugung homologer oder heterologer Sonden beschrieben.[75]

Bei dem dieser Arbeit zugrundeliegenden Labyrinthopeptin A2 konnte mit großer Sicherheit angenommen werden, dass es sich um ein ribosomal synthetisiertes Peptid handelte. Damit konnten die Standardprimer für NRPS und PKS nicht herangezogen werden. Es mussten also zunächst andere Angriffspunkte für die Generierung möglichst spezifischer homologer und heterologer Sonden gesucht werden. Alle bisher untersuchten Lantibiotika-Biosynthese-Gencluster zeigten als gemeinsame Schnittmenge die Gene für die modifizierenden Enzyme LanB/C oder M, für den ABC-Transporter LanT und für Enzyme, die weiterere posttranslationale Modifikationen (Dha, Dhb, AviCys usw.) einführen sowie die Strukturgene (*lanA*). Alle weiteren Gene wie z.B. das Proteasegen (*lanP*) oder Gene für die Regulation (*lanK/R*) oder Immunität (*lanI*) liegen nicht zwangsläufig im Cluster und wurden deshalb als nicht geeignet befunden (s. Abschnitt 1.4).

Proteinalignments (s. Abschnitt 1.3.2) der modifizierenden Enzyme LanB/C und LanM zeigten nur sehr geringe Sequenzhomologien (20-33 %), wobei diese auf der gesamten Sequenz verteilt zu finden waren und keine zusammenhängende Aminosäurensequenzen von mindestens sechs Aminosäuren für die Ableitung von Primern enthielten. Die Enzyme, die für die posttranslationale Einführung der Labionine verantwortlich sind, waren zu Beginn

dieser Arbeit gänzlich unbekannt. Die Leadersequenzen der Lantibiotika unterscheiden sich sehr im Hinblick auf Länge und Sequenz, was sie ebenfalls unbrauchbar machte.
Für die Generierung von Sonden zum Cosmidbankscreening konnten also nur der ABC-Transporter und die Primärsequenz des Labyrinthopeptins A2 selbst herangezogen werden.

5.1.4.2 Ableitung degenerierter Primer für die Identifizierung von ABC-Transportern

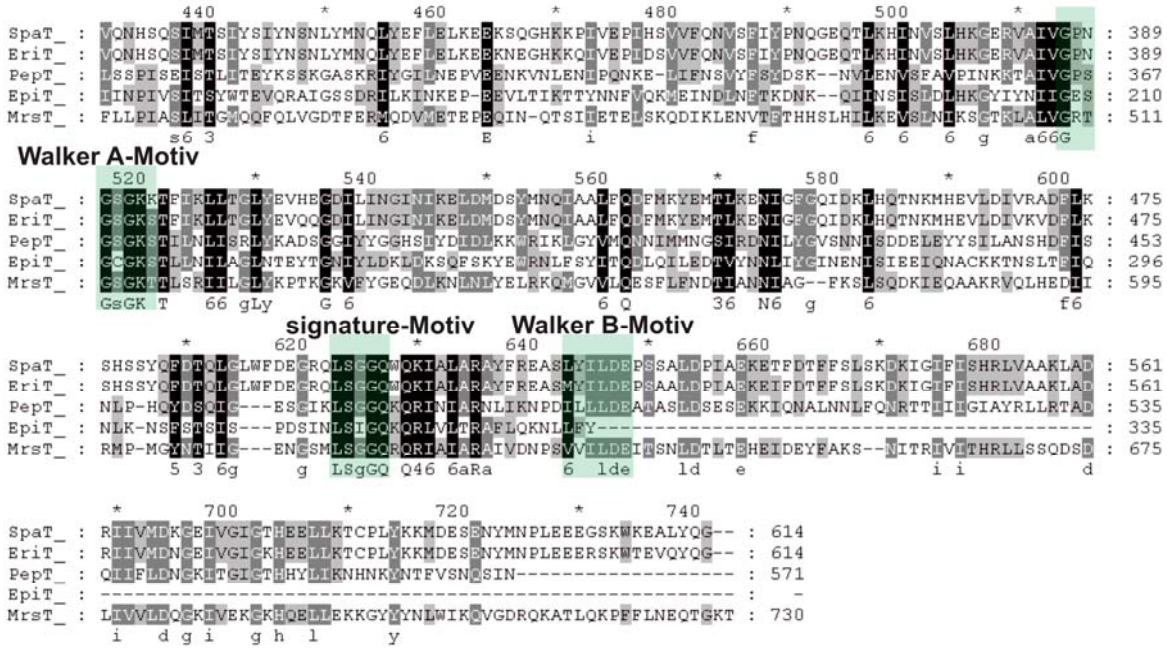
ABC-Transporter besitzen fünf hoch homologe Sequenzbereiche, Walker A- (GXXGXGKT/S; X = beliebig), Walker B- (hhhhDE; h = hydrophobe Aminosäure) und signature-Motive (LSGGQ), sowie die Gln- und His-Loops (s. Abschnitt 1.3.3). Da die Gln- und His-Loops aufgrund ihrer Kürze nicht für die Ableitung von Primer geeignet waren, blieben für den hier gewählten Ansatz nur die Walker- und signature-Motive. Diese Motive sind jedoch in jedem ABC-Transporter vorhanden, sodass damit gerechnet werden musste, dass bei einer PCR mit degenerierten Primern aus diesen drei Sequenzbereichen alle im Genom kodierten Transporter ansprechen. Um diesem Problem entgegenzuwirken, wurden Proteinalignments mit einer Vielzahl bekannter Lantibiotika-Transporter erstellt (s. Abbildung 29), wobei zur Verringerung der Komplexität die Lantibiotika-Transporter in zwei Hauptgruppen unterteilt wurden. Die Einteilung in Gruppen erfolgte aufgrund der Taxonomie der jeweiligen Lantibiotika-Produzenten. Mit Ausnahme von CinT, dem Transporter aus dem Cinnamycin-Cluster aus *Streptomyces cinnamoneus*, stammten alle Sequenzen aus Lantibiotika-Produzenten, die der Klasse der Bacilli angehören (s. Tabelle 20). Die Gruppeneinteilung erfolgte dann aufgrund der Unterordnungen in Bacillales und Lactobacillales.

Tabelle 20: Auflistung der in die Proteinalignments eingegangenen Transportersequenzen. Die Sequenzen wurden aus der Swissprot-Datenbank heruntergeladen (**www.expasy.ch**).

Abkürzung	Lantibiotikum	Spezies	Ordnung	Klasse
SpaT	Subtilin	*Bacillus subtilis*	Bacillales	Bacilli
EriT	Ericin	*Bacillus subtilis*	Bacillales	Bacilli
PepT	Pep5	*Staphylococcus epidermidis*	Bacillales	Bacilli
EpiT	Epidermin	*Staphylococcus epidermidis*	Bacillales	Bacilli
MrsT	Mersacidin	*Bacillus sp.* *(strain HIL-Y85/54728)*	Bacillales	Bacilli
NisT	Nisin	*Lactococcus lactis*	Lactobacillales	Bacilli
SrtT	Streptin	*Streptococcus pyogenes*	Lactobacillales	Bacilli

MutT	Mutacin III	*Streptococcus mutans*	Lactobacillales	Bacilli
SalT	Salivaricin A	*Streptococcus salivarius*	Lactobacillales	Bacilli
CylB	Cytolysin B	*Enterococcus faecalis*	Lactobacillales	Bacilli
ScnT	StreptococcinA-FF22	*Streptococcus pyogenes*	Lactobacillales	Bacilli
LasT	Lactocin S	*Lactobacillus sakei*	Lactobacillales	Bacilli
CinT	Cinnamycin	*Streptomyces cinnamoneus*	Actinomycetales	Actinobacteria

Proteinalignment Bacillales:



Proteinalignment Lactobacillales:

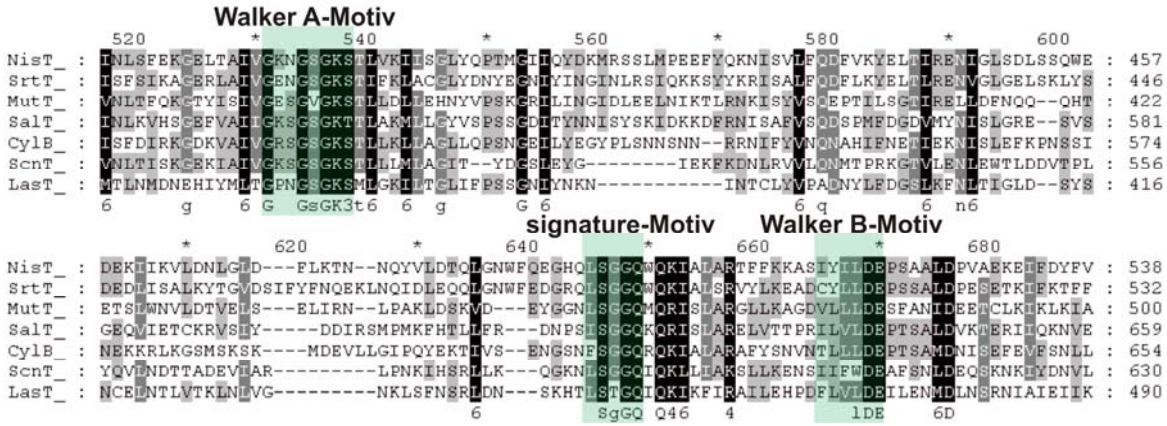


Abbildung 29: Darstellung der Proteinalignments, aufgrund derer die variablen Stellen der Walker-Motive durch konkrete Aminosäuren ersetzt werden sollten.

Anhand der abzulesenden Homologien aus den Proteinalignments wurde nun versucht, die variablen Stellen der Walker- und signature-Motive durch Aminosäuren zu ersetzen, die in Lantibiotika-Transportern gehäuft vorkommen. Dadurch sollte die Spezifität der Primer verbessert werden. Dies ist am Beispiel des Walker A-Motivs in Tabelle 21 gezeigt.

Tabelle 21: Vergleich der Ergebnisse der Proteinalignments in der Walker A-Region (GXXGXGKS/T). Dargestellt sind die Aminosäurereste aus Bacillales und Lactobacillales in der Walker A-Region und deren absolute Häufigkeit an den jeweiligen Positionen in Klammern. Außerdem ist zum Vergleich die entsprechende Sequenz von CinT abgebildet.

Walker A				G	X	X	G	X	G	K	S/T		
Bacillales:	$A_{(4)}$	$I_{(4)}$	$V_{(4)}$	$G_{(5)}$	$P_{(3)}$	$N_{(2)}$	$G_{(5)}$	$S_{(4)}$	$G_{(5)}$	$K_{(5)}$	$S_{(3)}$	$T_{(5)}$	$L_{(2)}$
	$N_{(1)}$	$L_{(1)}$	$I_{(1)}$		$E_{(1)}$	$S_{(2)}$		$C_{(1)}$			$T_{(1)}$		$F_{(2)}$
					$R_{(1)}$	$T_{(1)}$					$K_{(1)}$		$I_{(1)}$
Lactobacillales:	$A_{(5)}$	$I_{(6)}$	$V_{(5)}$	$G_{(7)}$	$K_{(3)}$	$S_{(4)}$	$G_{(7)}$	$S_{(6)}$	$G_{(7)}$	$K_{(7)}$	$S_{(6)}$	$T_{(6)}$	$L_{(6)}$
	$S_{(1)}$	$L_{(1)}$	$I_{(1)}$		$E_{(2)}$	$N_{(3)}$		$V_{(1)}$			$T_{(1)}$	$M_{(1)}$	$I_{(1)}$
	$M_{(1)}$		$T_{(1)}$		$R_{(1)}$								
					$P_{(1)}$								
CinT:	G	F	L	G	P	N	G	A	G	K	T	T	T

Die Sequenzhomologien im Walker A-Motiv selbst, vor allem im GXXG-Motiv waren zu gering, um dort eine akzeptable Sequenzvorhersage treffen zu können. Der Vergleich der Sequenzen von Bacillales und Lactobacillales zeigte jedoch eine gute Homologie der zwei Aminosäuren nach dem Walker A-Motiv, so dass von dieser Sequenz (GSGKSTL) ausgehend Primer abgeleitet wurden, was im Folgenden beschrieben wird.

Der Labyrinthopeptin A2-Produzent *A. namibiensis* ist ein Vertreter der Klasse Actinobacteria, von der nur eine Transportersequenz zur Verfügung stand (CinT). Um eine Aussage für mögliche homologe Sequenzen in dieser Klasse treffen zu können, wurde die CinT-Sequenz direkt mit den Sequenzen aus Bacillales und Lactobacillales verglichen. Wie man Tabelle 21 entnehmen kann, unterscheidet sich die Sequenz von CinT in dem Bereich hinter Walker A deutlich, so dass ausgehend von CinT ein zusätzlicher Primer abgeleitet werden musste. Die dabei gemachte Annahme war, dass die in diesem Bereich gezeigte Homologie der Lantibiotika-Transporter der Klasse Bacilli eventuell auch entsprechend bei der Klasse der Actinobacteria zu finden sein könnte. Ausgehend von den Aminosäuresequenzen GSGKSTL (Bacilli) und GAGKTTT (Actinobacteria) wurden mit dem *Codonusage* für Actinomadura (s. Abschnitt 5.1.3) degenerierte Primer abgeleitet.

Tabelle 22: Übersetzung der ausgewählten Bereiche des Walker A-Motivs auf Nukleotidebene mittels des *Codonusage* für die Gattung Actinomadura (prozentuale Häufigkeit der Codons in *A. hibisca*). Dabei wurde versucht, eine Abdeckung mit einer 95%igen Wahrscheinlichkeit bei den einzelnen Aminosäuren zu gewährleisten. Außerdem wurde darauf geachtet, dass die Primer nicht zu stark degeneriert waren. Die in die Primer eingegangenen Codons sind grau unterlegt. Die Codonverteilung bei Ser ist sehr ungünstig, weshalb je Ser-Rest zwei Primer abgeleitet wurden.

Bacilli	G	S	G	K	S	T	L
Codons	GGC(73.1)	TCG(39.5)	GGC(73.1)	AAG(97.4)	TCG(39.5)	ACC(61.9)	CTG(66.9)
	GGG(23.1)	AGC(37.9)	GGG(23.1)	AAA(2.6)	AGC(37.9)	ACG(36.6)	CTC(28.5)
	GGT(2.6)	TCC(20.2)	GGT(2.6)		TCC(20.2)	ACA(1.5)	TTG(2.9)
	GGA(1.3)	TCT(1.6)	GGA(1.3)		TCT(1.6)	ACT(0.0)	CTT(1.3)
		AGT(0.8)			AGT(0.8)		CTA(0.4)
		TCA(0.0)			TCA(0.0)		TTA(0.0)
Primer:							
wABL1(30.7)	GGS(96.2)	TCS(59.7)	GGS(96.2)	AAG(97.4)	TCS(59.7)	ACS(98.5)	CT(97.1)
wABL2(19.4)	GGS(96.2)	AGC(37.9)	GGS(96.2)	AAG(97.4)	TCS(59.7)	ACS(98.5)	CT(97.1)
wABL3(19.4)	GGS(96.2)	TCS(59.7)	GGS(96.2)	AAG(97.4)	AGC(37.9)	ACS(98.5)	CT(97.1)
wABL4(12.4)	GGS(96.2)	AGC(37.9)	GGS(96.2)	AAG(97.4)	AGC(37.9)	ACS(98.5)	CT(97.1)
CinT	G	A	G	K	T	T	T
Codons	GGC(73.1)	GCG(49.7)	GGC(73.1)	AAG(97.4)	ACC(61.9)	ACC(61.9)	ACC(61.9)
	GGG(23.1)	GCC(48.9)	GGG(23.1)	AAA(2.6)	ACG(36.6)	ACG(36.6)	ACG(36.6)
	GGT(2.6)	GCT(0.8)	GGT(2.6)		ACA(1.5)	ACA(1.5)	ACA(1.5)
	GGA(1.3)	GCA(0.5)	GGA(1.3)		ACT(0.0)	ACT(0.0)	ACT(0.0)
Primer:							
wACinT1(86.2)	GGS(96.2)	GCS(98.6)	GGS(96.2)	AAG(97.4)	ACS(98.5)	ACS(98.5)	AC(100.0)

Die Ableitung der degenerierten Primer ausgehend von den bestimmten Proteinsequenzen sollte eine maximale Abdeckung der Codons gewährleisten, ohne jedoch die Variation der Sequenz innerhalb eines Primers zu groß werden zu lassen. Dazu wurden zunächst alle Codons in die Betrachtung miteinbezogen und durch Kombination und Degeneration versucht, eine möglichst große Abdeckung zu erlangen (>95 %). Der Code für eine degenerierte Base entspricht dem in Abschnitt 3.9 gezeigten Standardcode. Bei schwer zusammenzufassenden Aminosäuren bzw. stark degenerierten Codontripletts, wie z.B. für Ser wurde aufgrund der zu großen Unterschiede der einzelnen Basentripletts die Verteilung auf zwei Primer bevorzugt. Es wurden auf diese Weise fünf Vorwärtsprimer abgeleitet, wobei für beide Klassen, Bacilli und Actinobacteria, jeweils eine Codonabdeckung von über 80 % erreicht wurde.

In analoger Weise zu den fünf Vorwärtsprimern aus der Walker A-Region wurden insgesamt neun Rückwärtsprimer aus den Regionen des signature- und Walker B-Motivs abgeleitet. Bei

der Ableitung aus dem signature-Motiv wurde hier zwischen Bacillales (B) und Lactobacillales (L) unterschieden und folglich Primer für jede Ordnung abgeleitet (wB7B1/2 und wB7L1/2). Die entsprechende CinT-Region wurde analog zu dem Walker A-Bereich gesondert betrachtet und daraus ein Primer abgeleitet (wB7CinT1/2). Bei der Ableitung der Rückwärtsprimer aus dem Walker B-Motiv konnte eine Konsensussequenz in den Sequenzen der Klasse Bacilli gefunden werden (wBallesI und wBallesL), wobei auch hier wieder die CinT-Sequenz getrennt betrachtet wurde. Als Gesamtergebnis wurden insgesamt fünf Vorwärtsprimer und neun Rückwärtsprimer abgeleitet, die in Tabelle 23 zusammengefasst sind.

Tabelle 23: Zusammenstellung aller abgeleiteter, degenerierter Primer für die Identifizierung von Lantibiotika-spezifischen ABC-Transportern. Die Sequenzen der Rückwärtsprimer wurden nach dem Ableiten aus den Motiven entsprechend übersetzt und invertiert.

Primer	Bezeichnung	zugrundeliegende Aminosäuresequenz	degenerierte Primersequenz
Vorwärts-primer	wABL1	GSGKSTL	5´-GGSTCSGGSAAGTCSACSC-3´
	wABL2		5´-GGSAGCGGSAAGTCSACSC-3´
	wABL3		5´-GGSTCSGGSAAGAGCACSC-3´
	wABL4		5´-GGSAGCGGSAAGAGCACSC-3´
	wACinT	GAGKTTT	5´-GGSGCSGGSAAGACSACSAC-3´
Rückwärts-primer	wB7B1	LSGGQWQ	5´-CTGCCACTGSCCSCCSGASAG-3´
	wB7B2		5´-CTGCCACTGSCCSCCGCTSAG-3´
	wB7L1	LSGGQIQ	5´-CTGGATCTGSCCSCCSGASAG-3´
	wB7L2		5´-CTGGATCTGSCCSCCGCTSAG-3´
	wB7CinT1	FSGGMQR	5´-CGCTGCATSCCSCCSGAG-3´
	wB7CinT2		5´-CGCTGCATSCCSCCGCTG-3´
	wBallesL	LLLDEPT	5´-SGTSGGRTCGTCSAGGAASAG-3´
	wBallesI	LILDEPT	5´-SGTSGGRTCGTCSAGSAGSAG-3´
	wBCinT	LFLDEPT	5´-SGTSGGRTCGTCSAGGATSAG-3´

5.1.4.3 Identifizierung von ABC-Transportern mit Hilfe der degenerierten Transporterprimer

Mit den in Abschnitt 5.1.4.2 abgeleiteten Primern konnten verschiedene Kombinationen von Vorwärts- und Rückwärtsprimern gebildet werden, die zur Standard-PCR (s. Abschnitt 4.2.3.1; Annealingtemperatur: 59°C; Elongation: 2 min) auf genomische DNA von *A. namibiensis* eingesetzt wurden. Die Auswertung erfolgte mittels Agarosegelelektrophorese in 2%igen Gelen. Einige Primerpaarkombinationen zeigten, im Vergleich zu den jeweiligen Kontrollen mit jeweils nur einem Primer, Banden auf der erwarteten Höhe bei 340-420 bp, was beispielhaft in Abbildung 30 gezeigt ist.

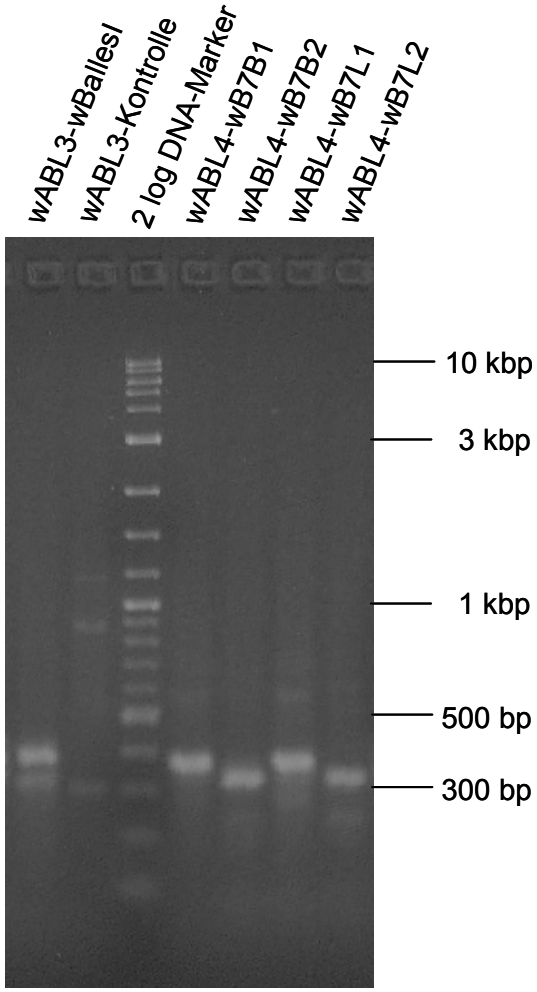


Abbildung 30: Auswertung der Standard-PCR-Reaktionen mit verschiedenen Primerkombinationen in einem 2%igen Agarosegel. Die Spur des Ansatzes wABL3-aBallesI zeigt eine Bande knapp unter 400bp, die in der Spur daneben (wABL3-Kontrolle) nicht zu beobachten ist.

Die als interessant erachteten Banden wurden aus dem Gel gereinigt und zur Sequenzierung in pDrive kloniert. Die Auswertung der Sequenzen erfolgte immer nach der gleichen Vorgehensweise. Zunächst wurde überprüft, ob die Sequenz überhaupt ein Produkt der PCR des eingesetzten Primerpaars war. War dies der Fall, wurde die Sequenz mittels BLAST

gegen die nr-Datenbank[76] untersucht, wofür die Bereiche vorher entfernt wurden, die durch die Primer vorgegeben waren.

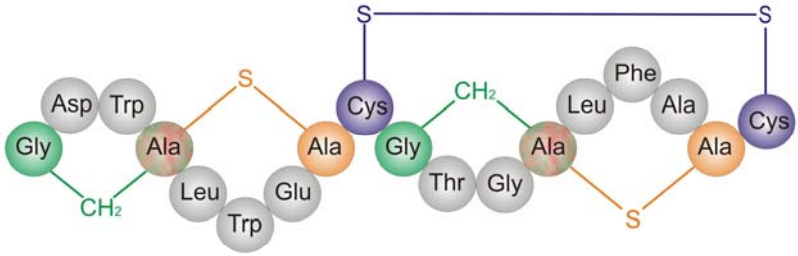
Auf diese Weise wurden insgesamt 17 unterschiedliche Sequenzen generiert, die alle im BLAST Hits zu ABC-Transportern unterschiedlicher Familien gezeigt haben. Auch nach eingehender Betrachtung der BLAST-Ergebnisse und nach Vergleich der Ergebnisse mit BLASTs von bekannten Lantibiotika-Transportern war es nicht möglich, einen oder mehrere Transporter als besonders geeignet für das Cosmidbankscreening zu identifizieren. Auch eine Hybridisierung einer oder mehrerer der gefundenen ABC-Sequenzen hatte wenig Aussicht auf Erfolg, da diese teilweise sehr große Sequenzhomologien untereinander zeigten. Aufgrund dieser Sequenzhomologien war zu befürchten, dass sehr viele ABC-Transporter aus dem Genom darauf ansprechen würden. Bei einer zehnfachen Abdeckung und der Annahme von mindestens 20 bis 25 ABC-Transportern im Gesamtgenom würde man zwischen 200 und 250 positive Cosmide erwarten, was die Identifizierung des Labyrinthopeptin-Genclusters nur geringfügig erleichtert hätte. Abschließend kann gesagt werden, dass die Clustersuche mittels ABC-Transportersonden nicht zum Erreichen des Ziels geführt hat.

5.1.4.4 Ableitung von Sonden aus der Primärsequenz

Wie in Abschnitt 5.1.4.1 bereits diskutiert, kam neben dem im Cluster vermuteten ABC-Transporter nur die Primärsequenz des Labyrinthopeptins A2 selbst für die Generierung von Sonden zum Cosmidbankscreening in Frage. Dazu wurden zunächst alle theoretisch möglichen Aminosäuresequenzen betrachtet, von denen ausgehend durch posttranslationale Modifikationen das fertige A2 gebildet werden könnte. Es wurden dabei an die Stellen, die jeweils den ersten Aminosäuren eines Labionin-Motivs entsprechen, die Aminosäuren Ser, Cys oder Ala als denkbare Vorläufer gesetzt. Bei der Orientierung der Ser- und Cys-Reste, die an der Lanthioninbildung beteiligt sind, wurden beide Varianten (Ser->Cys; Cys->Ser) berücksichtigt, so dass sich die in Abbildung 31 dargestellte Ausgangssituation ergab.

An den Positionen der Labionin-Motive, für die aufgrund der Neuartigkeit noch keinerlei Biosyntheseinformationen und somit auch nichts über die ursprünglich an dieser Stelle kodierten Aminosäuren bekannt war, hätten insgesamt zwölf Codontripletts berücksichtigt werden müssen (Ser: 6; Ala: 4; Cys: 2). Da das eine zu starke Degenerierung zur Folge gehabt hätte, wurden diese beiden Positionen nicht berücksichtigt und bei der Ableitung lediglich die in Abbildung 31 C) abgebildeten Oktapeptide herangezogen. Mit diesen beiden Oktapeptiden wurden dann unter Verwendung des *Codonusage* für die Gattung Actinomadura die in Tabelle 24 aufgelisteten Sonden abgeleitet. Dies geschah analog zur Ableitung der degenerierten Primer aus dem Walker A-Motiv (s. Abschnitt 5.1.4.2).

A)



B)

Ala
Cys - Asp - Trp - Cys - Leu - Trp - Glu - Cys - Cys - Ala - Thr - Gly - Cys - Leu - Phe - Ala - Cys - Cys
Ser Ser Cys Ser Ser Ser

C)

Asp - Trp - Cys - Leu - Trp - Glu - Cys - Cys Thr - Gly - Cys - Leu - Phe - Ala - Cys - Cys
Ser Ser Ser

Abbildung 31: A) Posttranslational modifiziertes A2. B) Theoretisch mögliche bzw. sinnvoll erscheinende Aminosäuresequenzen vor der posttranslationalen Modifikation. C) Zerlegung des Strukturpeptids in geeignete Bereiche (Oktapeptide) zur Ableitung von degenerierten Sonden und Primern.

Tabelle 24: Ableitung von Sonden aus dem *N*-terminalen und *C*-terminalen Oktapeptidbereich der Primärsequenz des Labyrinthopeptins A2.

N-terminale Sequenz	Asp	Trp	Ser Cys	Leu	Trp	Glu	Cys Ser	Cys
V1	5´-GACTGGTCSCTSTGGGARTCSTGC-3´							
V2	5´-GACTGGWGCCTSTGGGARTCSTGC-3´							
V3	5´-GACTGGTCSCTSTGGGARWGCTGC-3´							
V4	5´-GACTGGWGCCTSTGGGARWGCTGC-3´							
C-terminale Sequenz	Thr	Gly	Ser Cys	Leu	Phe	Ala	Cys Ser	Cys
H1	5´-ACSGGSTCSCTSTTCGCSTCSTGC-3´							
H2	5´-ACSGGSWGCCTSTTCGCSTCSTGC-3´							
H3	5´-ACSGGSTCSCTSTTCGCSWGCTGC-3´							
H4	5´-ACSGGSWGCCTSTTCGCSWGCTGC-3´							

Bei der Ableitung wurde berücksichtigt, dass eine Lanthioninbrücke immer aus einem Ser und einem Cys gebildet wird. Deshalb wurden zwei getrennte Sonden abgeleitet, einmal für Ser->Cys und einmal für Cys->Ser, so dass die Möglichkeiten Cys->Cys und Ser->Ser

ausgeschlossen waren. Da zusätzlich die sechs möglichen Codons für Ser auf zwei unterschiedliche Sonden aufgeteilt wurden, ergaben sich jeweils vier Oligonukleotide für die *N*-terminale und *C*-terminale Sequenz. Die erhaltenen Sonden hatten jeweils eine Länge von 24 Nukleotiden und bestanden aufgrund der Degenerierung aus 16 (V1-V4) bzw. 64 (H1-H4) unterschiedlichen Varianten. Daran kann auch abgeschätzt werden, dass bei einer Ableitung ausgehend von den Aminosäuren 2-18 von Labyrinthopeptin A2 die Sonden zu stark degeneriert gewesen wären (12288 Varianten ausgehend von DSS für die Ser/Cys/Ala-„Brücke"). Diese jeweils vier Oligonukleotide – für den *N*-terminalen und *C*-terminalen Sequenzbereich aus dem Strukturgen – wurden für zahlreiche Ansätze unterschiedlichster Methoden herangezogen, von denen eine im folgenden Abschnitt näher beschrieben werden soll.

5.1.4.5 Screening mit [^{32}P]-markierten degenerierten 24er Oligonukleotiden aus der Primärsequenz auf genomische Restriktionsverdaus

Mit den in Abschnitt 5.1.4.4 beschriebenen Oligonukleotiden V1-4 und H1-4 aus der Primärsequenz des Labyrinthopeptins A2 sollte versucht werden, das Strukturgen mittels *Southern-Blot* zu lokalisieren. Dabei wurde auf eine Methode zurückgegriffen, die bereits bei der Identifizierung des Strukturgens von Salivaricin A erfolgreich war.[77] Diese Methode sah die im Folgenden beschriebene Vorgehensweise vor. Zunächst wurde hochmolekulare genomische DNA mit Restriktionsenzymen gespalten und die entstandenen Fragmente auf einem Gel getrennt und anschließend auf eine Membran geblottet. Im Falle des Salivaricins wurde dann mit degenerierten, endständig [^{32}P]-markierten 17er Oligonukleotiden als Sonden hybridisiert, die zuvor aus der mittels Edman-Abbau bestimmten *N*-terminalen Aminosäuresequenz abgeleitet wurden. So konnte ein DNA-Fragment gefunden und sequenziert werden.

Analog zu diesem Vorgehen wurde in der vorliegenden Arbeit hochmolekulare genomische DNA von *A. namibiensis* mit den Restriktionsenzymen BamHI, PstI und SmaI verdaut und die entstandenen Fragmente anschließend auf einem 1%igen Agarosegel getrennt. Da kein radioaktiv markierter Größenmarker zur Verfügung stand, wurden die Gele mit einem anliegenden Lineal fotografiert, um so später eine Größenabschätzung der Banden auf den Radiofilmen vornehmen zu können. Die DNA-Fragmente wurden dann geblottet und gegen die endständig [^{32}P]-markierten 24-er Oligonukleotide wie in Abschnitt 3.4.5 beschrieben hybridisiert. Aufgrund der sehr kurzen Sonden war zunächst eine Optimierung der Hybridisierungstemperatur notwendig. Erst bei 40°C konnte eine Hybridisierung beobachtet werden, wobei die Radiofilme bei 42°C schon keine Banden mehr zeigten. Das Screening mit diesen Sonden, die stark degeneriert und sehr kurz waren, stellte eine große Herausforderung dar, da die erforderlichen niedrigen Temperaturen unspezifische

Bindungen ermöglichten. Der dadurch hervorgerufene, relativ starke und bandenreiche Hintergrund ist in Abbildung 32 deutlich zu sehen. Hier führte die unspezifische Bindung sogar zur Sichtbarmachung der unmarkierten Markerbanden. Um die eigentlichen „spezifischen“ Banden zu finden, wurden immer mehrere *Southern-Blots* mit jeweils unterschiedlichen Sonden parallel unter identischen Bedingungen und aus demselben Verdauansatz durchgeführt. Die entwickelten Filme der Blots wurden dann exakt, u.a. durch Übereinanderlegen miteinander verglichen (s. Abbildung 32). Gesucht wurden Banden, die auf Filmen unterschiedlicher V- und H-Sonden auf gleicher Höhe zu beobachten waren. Das DNA-Fragment mit dem Gen für die Primärsequenz müsste nämlich mit den Sonden für den *N*-terminalen und *C*-terminalen Bereich hybridisieren und somit ein Signal auf gleicher Höhe erzeugen. Wie Abbildung 32 zeigt, konnten auf den vier abgebildeten Filmen zwei interessante Banden in den BamHI-Spuren identifiziert werden. Die Größe der Banden wurde auf 0,9 und 1,6 kbp bestimmt.

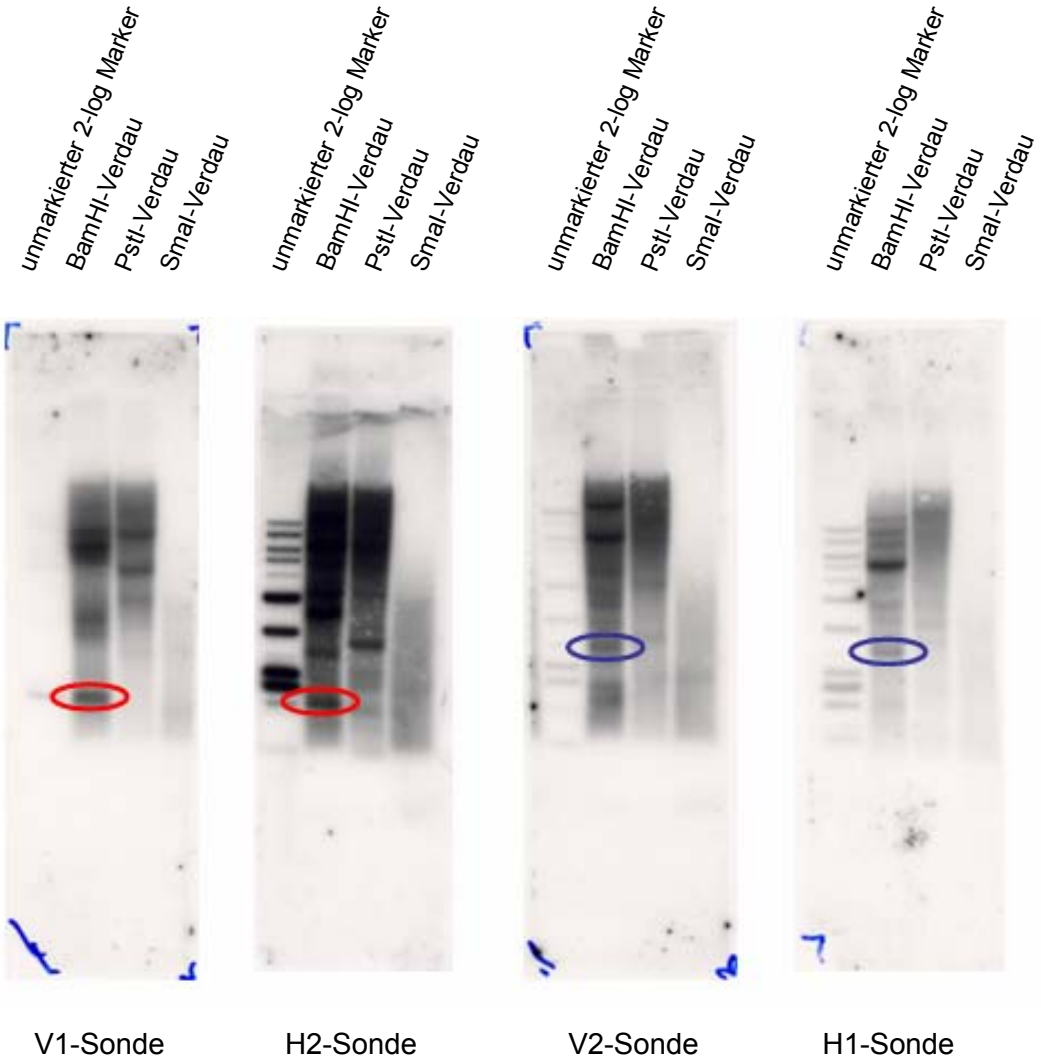


Abbildung 32: Auswertung der *Southern-Blots* mit den Sonden V1-2 und H1-2. Rot markiert ist eine Bande bei 0,9 kbp, die sowohl auf die V1- als auch auf die H2-Sonde ansprach. Die blaue Bande entspricht einer Fragmentgröße von 1,6 kbp und hybridisierte sowohl mit der V2- als auch mit der H1-Sonde.

Der bestimmte Größenbereich wurde aus einem nachgezogenen Gel ausgeschnitten und kloniert. Klone mit einem Insert richtiger Größe wurden verdaut, geblottet und mit den gleichen Sonden reanalysiert. Ein Klon von 20 untersuchten Klonen zeigte in der Reanalytik eine Bande, woraufhin dieser sequenziert wurde. Die Auswertung der erhaltenen Sequenz ergab jedoch, dass jeweils ein kleiner Sequenzbereich von 17 bzw. 18 Nukleotiden mit einer Sequenzhomologie von nur 65 % ausgereicht hatte, um die auf den Filmen beobachtete Bande zu erzeugen. Erschwerend kam hinzu, dass die Restriktionsverdaus eine schlechte Reproduzierbarkeit aufwiesen, was sich daran zeigte, dass die Bandenmuster verschiedener Blot-Ansätze mit der gleichen Sonde teilweise deutlich unterschiedlich aussahen. Probleme, wie nicht quantitativ verlaufende Restriktionsverdaus oder sehr geringe Ligationsausbeuten wurden während der praktischen Arbeit mit der genomischern DNA von *A. namibiensis* mehrfach beobachtet. Aus o.g. Gründen wurde diese Strategie nach einigen Versuchen verworfen.

5.1.4.6 PCR-Screening auf genomische DNA mit elongierten Primern der Primärsequenz

Aufgrund der geringen Erfolge mit den ersten beiden Strategien, Identifizierung von ABC-Transportern und *Southern-Blot* mit degenerierten Sonden aus der Primärsequenz, musste der Einsatz weiterer Methoden in Betracht gezogen werden. Eine bei der Strukturaufklärung von Ruminococcin A beschriebene Methode mit elongierten Primern erschien übertragbar auf das hier vorliegende Problem und wurde deshalb getestet.[78] Bei der Strukturaufklärung von Ruminoccocin A wurde eine PCR mit degenerierten Primern beschrieben, wobei die Ableitung des Vorwärtsprimers am *N*-terminalen Ende und die des Rückwärtsprimers am *C*-terminalen Ende der Primärsequenz des Lantibiotikums erfolgte. Das erwartete PCR-Produkt hatte jedoch nur eine Größe von 63 bp und wurde deshalb durch das Anhängen von 18 Nukleotiden an das jeweilige 5´-Ende der Primer auf 99 bp verlängert.

In Analogie zu dieser Methode wurden neue Primer aus der Primärsequenz des Labyrinthopeptins A2 abgeleitet, wobei diesmal unabhängig von ihrer relativen Häufigkeit alle Codons berücksichtigt wurden (s. Abbildung 33).

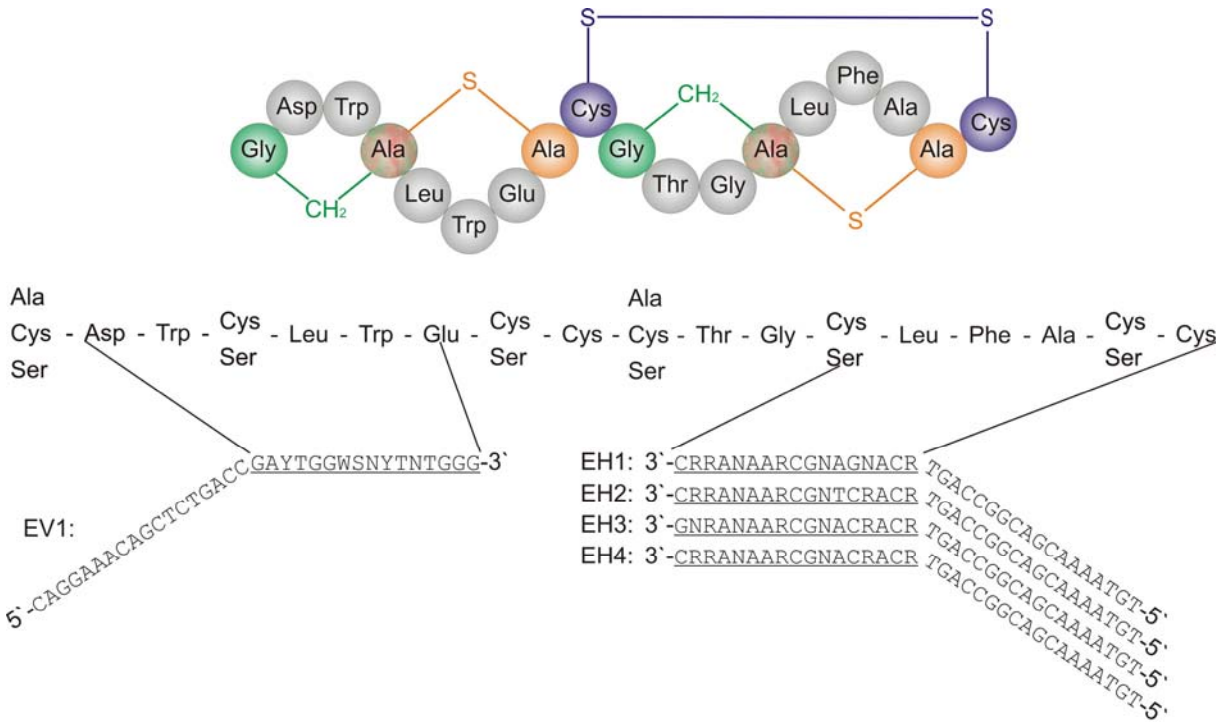


Abbildung 33: Gezeigt ist die Ableitung degenerierter Primer an der Primärsequenz und deren Elongation am 5´-Ende um jeweils 18 Nukleotide. Auf diese Weise wurden ein Vorwärtsprimer und vier Rückwärtsprimer abgeleitet, die bei erfolgreicher PCR ein Produkt von 87 bp Länge erzeugen sollten.

Im Gegensatz zu den in Abschnitt 5.1.4.4 abgeleiteten Sonden, bei denen die maximal mögliche Länge erreicht werden sollte, sollten hier nun die Primer möglichst randständig und von minimal möglicher Länge gewählt werden, denn nur so hätte das erwartete PCR-Produkt auch den maximalen Informationsgehalt. Der Vorwärtsprimer EV1 bestand aus 256 Varianten, begann mit der ersten Base des Codons von Asp und endete mit der ersten Base des Codons für Glu. Der schwierigste Bereich über das Ser oder Cys der Lanthioninbrücke wurde durch die Wahl eines WSN-Codons gelöst, womit alle acht Codons abgedeckt waren. Da der Bereich des Vorwärtsprimers außerdem zwei Trp-Reste enthielt, die nur durch ein Codontriplett kodiert werden, hielt sich die Degeneriertheit insgesamt in einem akzeptablen Bereich. Die Ableitung der vier Rückwärtsprimer erfolgte bezüglich der Labioninbrücke in Analogie zu den vier H-Sonden aus Abschnitt 5.1.4.4. Allerdings waren die neu ausgewählten Primer deutlich degenerierter als die früher verwendeten Sonden, da wie bereits erwähnt alle möglichen Codons berücksichtigt werden sollten und die vorkommenden Aminosäuren Ala und Leu durch vier bzw. sechs mögliche Codons kodiert werden. Bei der Inversion und Übersetzung der Primer musste darauf geachtet werden, dass die degenerierten Positionen richtig übersetzt werden. So wird z.B. aus Y (C oder T) R (G oder A) (s. Abschnitt 3.9). Trotz der Aufteilung auf vier verschiedene Primersets bestanden diese immer noch aus 512 (EH2 und EH4) bzw. 1024 (EH1 und EH3) unterschiedlichen Varianten. Mit den sich daraus ergebenden vier verschiedenen Primerpaarkombinationen und den entsprechenden Kontrollen wurde die PCR nach dem in Abschnitt 4.2.3.2 beschriebenen

Protokoll durchgeführt. Die erwartete Produktgröße der PCR-Ansätze wurde auf 87 bp berechnet. Auf einem ersten 2%igen Gel waren nur leichte Schatten zu sehen, weshalb die gleichen PCR-Ansätze noch einmal wiederholt wurden, diesmal jedoch mit je 1 µl des ersten Ansatzes als Templat. Die Ansätze wurden erneut auf einem 2%igen Agarosegel analysiert (s. Abbildung 34) und ausgewertet.

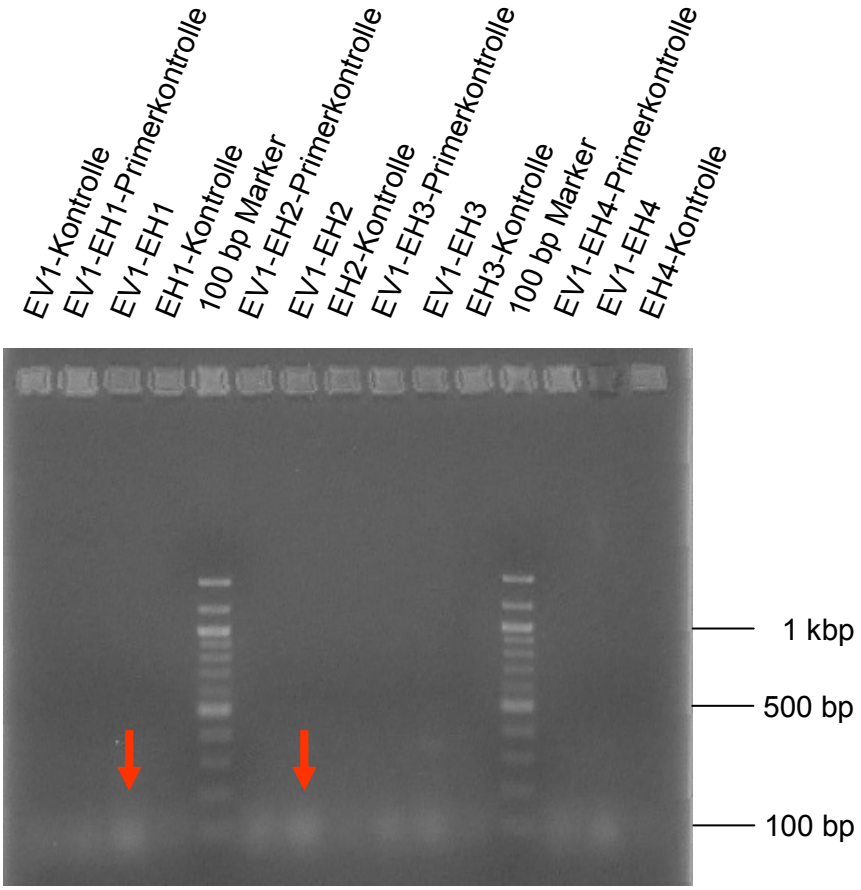


Abbildung 34: Auswertung der zweiten PCR-Reaktionen in einem 2%igen Agarosegel. Die mit einem roten Pfeil markierten Banden wurden weiter verwendet.

Um unterscheiden zu können, ob es sich bei den schwachen Banden knapp unter der 100 bp Markerbande nur um die Primer selbst oder auch um ein PCR-Produkt handelte, wurden diverse Kontrollen mit aufgetragen. Aufgrund der geringen Gelauflösung und der relativ kleinen Größenunterschiede zwischen Primern (36 Nukleotide) und PCR-Produkt (87 bp) wurde je Ansatz auch ein Ansatz ohne Polymerase angesetzt, so dass der DNA-Gehalt mit dem eigentlichen Ansatz identisch war. Außerdem wurde von jedem eingesetzten Primer auch ein Ansatz mit nur diesem einen Primer gemacht. Durch Vergleichen der Ansätze mit den jeweiligen Kontrollen konnten schwache Banden in den Ansätzen EV1-EH1, EV1-EH2 und EV1-EH4 beobachtet werden, die bei direkter Betrachtung auf dem UV-Tisch besser zu sehen waren. Aufgrund der deutlicheren Unterschiede zwischen den Ansätzen EV1-EH1 und EV2-EH2 mit ihren jeweiligen Kontrollen gegenüber anderen Ansätzen wurde mit diesen weitergearbeitet. Die PCR-Ansätze wurden direkt in pDrive kloniert und in *E. coli*

DH5α transformiert. Nach der Isolierung einiger Plasmide wurden jeweils RsaI-, PstI- und PstI/NcoI-Verdaus angesetzt und die Fragmente auf dem Gel analysiert. Man konnte in den Spuren der Verdaus zwar leichte Unterschiede in der Bandengröße ausmachen, aber keinesfalls absolute Größen bestimmen und damit auch nicht auf die Anwesenheit des erhofften Inserts schließen. Aus diesem Grund wurden mehrere Plasmide direkt zum Sequenzieren geschickt. Die Sequenzierung von insgesamt vier Plasmiden erbrachte zwei interessante Sequenzen von 87 bp Länge, wovon eine in Abbildung 35 genauer betrachtet werden soll.

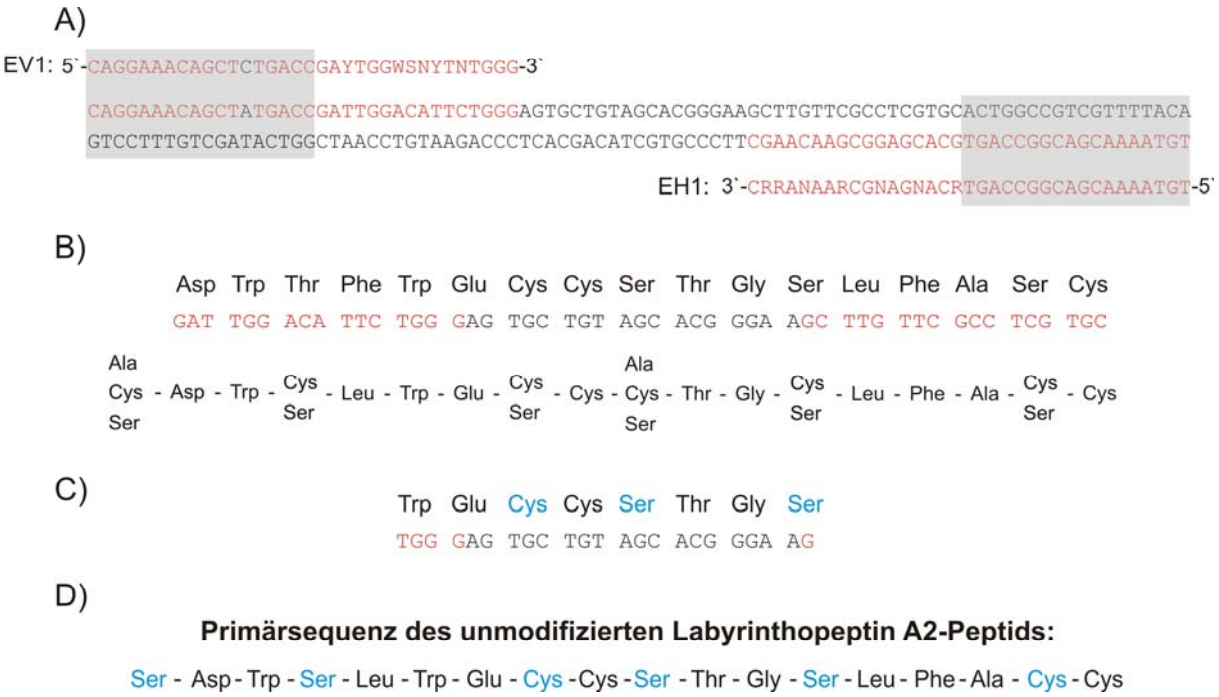


Abbildung 35: Auswertung der Sequenzierung. A) zeigt die Originalsequenz mit den darin liegenden Primern EV1 und EH1. Rot markiert sind die Basen, die mit den Vorgaben der degenerierten Primer übereinstimmen. Die Bereiche, die zur Produktverlängerung angehängt wurden, sind grau unterlegt. B) liefert die dazugehörige Übersetzung des interessanten Bereichs aus der Primärsequenz in die jeweiligen Aminosäuren. C) zeigt den nun bekannten, 23 Nukleotide langen Sequenzabschnitt des Strukturgens aus dem Cluster, der mittels der PCR mit elongierten Primern generiert werden konnte sowie die Aminosäureübersetzung. D) Mit Hilfe von im Text beschriebenen Analogieschlüssen konnte die komplette Aminosäure-Primärsequenz von Labyrinthopeptin A2 bestimmt werden.

Zunächst wurden die in der PCR eingesetzten Primer EV1/EH1 in der Sequenz gesucht und die Sequenz danach orientiert. Eine genauere Analyse beschränkte sich dann auf den Bereich, der nicht durch die degenerierten Primer vorgegeben war, denn die Bereiche des Primers mussten wegen dessen Degeneriertheit nicht zwingend der Sequenz auf dem Genom entsprechen. Das neu dazu gekommene Sequenzstück hatte die gesuchte Länge, so dass es nicht zu einem Frameshift im durch die Primer vorgegebenen Leseraster kam. Bei der Übersetzung der Basentripletts in die Aminosäuren konnte schnell festgestellt werden, dass es tatsächlich gelungen war, mit Hilfe dieser stark degenerierten und verlängerten Primer den Bereich der Aminosäuren 7-13 aus dem Strukturgen mittles PCR zu

amplifizieren und somit zu sequenzieren. Mit diesem kurzen Sequenzstück ließen sich jedoch noch weitere Erkenntnisse über die Primärsequenz gewinnen, die für das weitere Vorgehen wichtig waren. Anhand der Teilsequenz konnte nicht nur die Reihenfolge der Ser- und Cys-Reste als Lanthionin-Vorläufer bestimmt, sondern vor allem auch Ser als Vorläufer des bisher noch nicht beschriebenen Labionin-Motivs identifiziert werden. Somit war es nun möglich, die Primärsequenz auf Aminosäurestufe ohne variable Stellen wie in Abbildung 35 D) gezeigt, darzustellen. Hierbei wurde die Annahme gemacht, dass Ser auch der Rest ist, der in die Biosynthese des ersten Labionin-Motiv eingeht. Das zweite, hier nicht dargestellte Ergebnis der Sequenzierung eines PCR-Produkts aus dem Primerpaar EV1/EH2 zeigte abgesehen von kleinen Unterschieden im Bereich der Primer die gleiche Sequenz, was die Ergebnisse verifizierte.

5.1.4.7 Screening der Cosmidbank mit [^{32}P]-markierten degenerierten 24er Oligonukleotiden sowie einem nicht-degenerierten 23er Oligonukleotid aus der Primärsequenz

Aufgrund der nicht zufriedenstellenden Ergebnisse der in Abschnitt 5.1.4.5 beschriebenen Hybridisierung gegen Verdaus der genomischen DNA sollte jetzt eine Hybridisierung gegen die auf Hybridisierungsfilter gespottete Cosmidbank durchgeführt werden. Die Unterschiede zur Hybridisierung gegen die Restriktionsfragmente sind der Wegfall des nicht reproduzierbaren Restriktionsverdaus und die Hybridisierung gegen einen deutlich kleineren Anteil des Gesamtgenoms in Form der Cosmide. Dies ließ eine Verbesserung der Ergebnisse erwarten. Dieses Filterscreening wurde in Zusammenarbeit mit der Firma RZPD, Berlin, durchgeführt, wobei die Sonden ebenfalls endständig mit $^{[32]}$P-markiert eingesetzt wurden.

In einem ersten Screening wurden nur die in Abschnitt 5.1.4.4 beschriebenen degenerierten V- und H-Sonden, jeweils gepoolt, eingesetzt. Erst zu einem späteren Zeitpunkt, als die 23 Nukleotide lange Sequenz aus dem vorhergehenden Abschnitt 5.1.4.6 bestimmt werden konnte, wurde auch mit dieser Sonde (A2AS6-13) gegen einen Filter hybridisiert (s. Abbildung 36).

Auf den Filtern konnten 11 Cosmide identifiziert werden, die mehr oder weniger mit allen Sonden ein Signal zeigten. Aufgrund des Belegungsplans der Filter wurden die Cosmide aus der Cosmidbank aktiviert und mittels dem in Abschnitt 4.2.2.9 beschriebenen Protokoll isoliert. Die isolierten Cosmide wurden dann mit EcoRI verdaut und auf einem 0,8%igen Agarosegel analysiert.

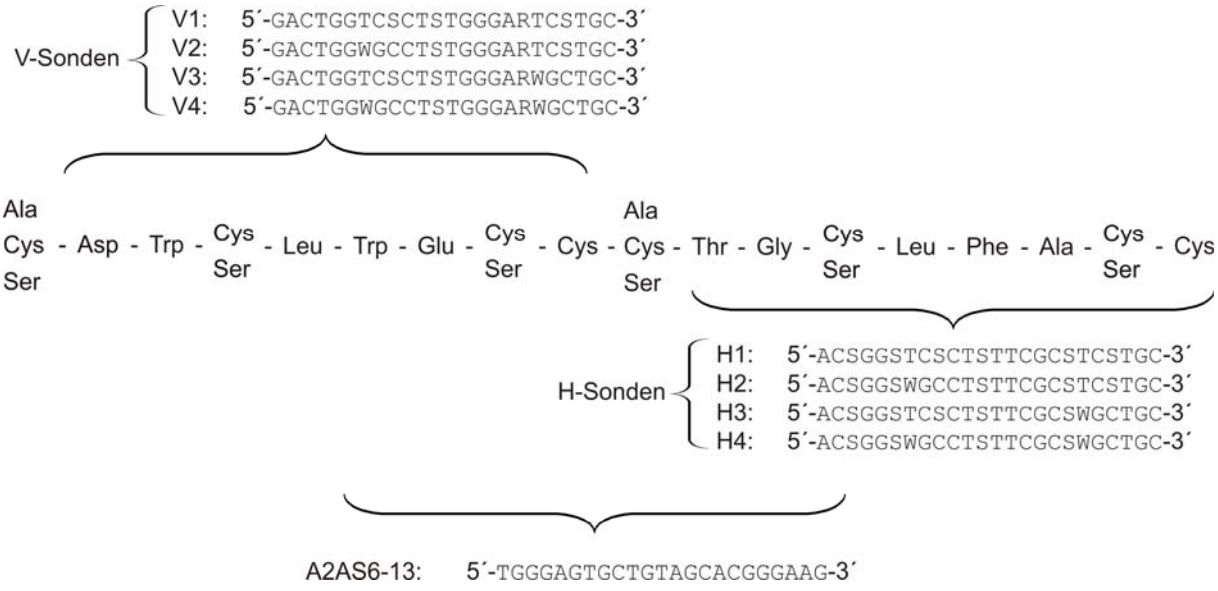


Abbildung 36: Übersicht über die zum Filterscreening eingesetzten Sonden.

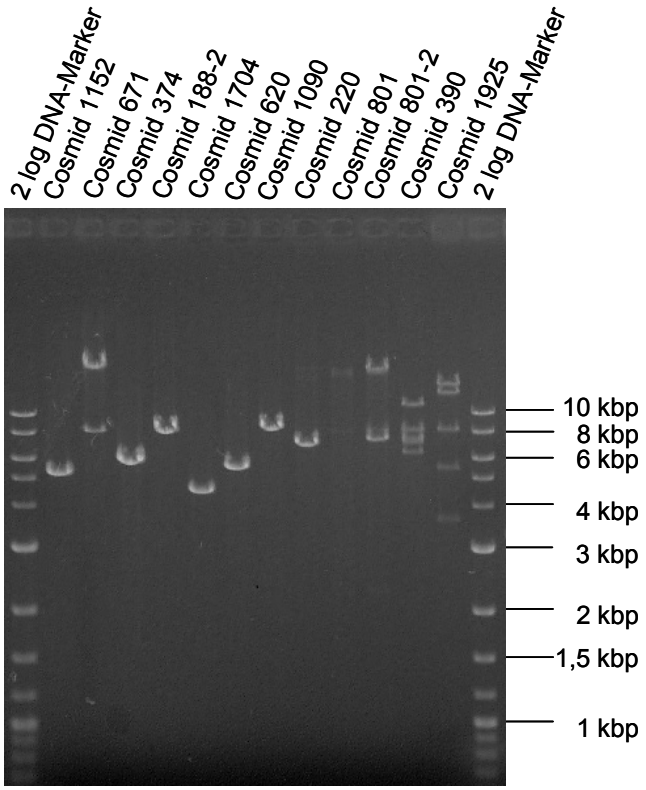


Abbildung 37: EcoRI-Verdau der isolierten Cosmide, die bei der Hybridisierung mit den V-, H- und A2AS6-13-Sonden ein positives Signal gezeigt hatten. EcoRI schneidet das Insert aus dem Vektor aus und erzeugt ein 8,1 kbp großes Fragment im Vektorrückgrat.

Die Auswertung des in Abbildung 37 gezeigten Gels führte zu der Vermutung, dass die Integrität der isolierten Cosmide nicht gegeben war. So zeigten nur vier der elf verschiedenen Cosmide (671, 801, 390 und 1925) überhaupt die notwendigerweise erwartete Vektorbande bei 8,1 kbp. Bei einer durchschnittlichen Insertgröße von 35 kbp

wurden bandenreiche Fragmentmuster erwartet, was bei keinem Cosmid der Fall war. Aufgrund dieser Ergebnisse wurden die Cosmidklone erneut aus der urspünglichsten Kopie der Bank reaktiviert, von Beginn an dann aber nur bei 27°C auf Platte kultiviert. Verdaus dieser nachgezogenen Ansätze zeigten keine Veränderung zum ersten Ansatz mit der Kultivierung bei 37°C und in Flüssigkulturen. Zur Überprüfung, ob nicht doch ein Teil des Clusters in den Cosmiden zu finden ist, wurden PCRs mit den Cosmiden als Templat angesetzt, wobei die elongierten Primer (EV1-EH1) aus Abschnitt 5.1.4.6 eingesetzt wurden. Auch der direkte Versuch einer Sequenzierung auf diesen Cosmiden mit der Sequenz der A2AS6-13-Sonde als Primer zeigte keinerlei Ergebnis. Folglich gelang es also nicht, Cosmide zu isolieren, die nachweislich ein Insert mit Bezug zu dem gesuchten Gencluster enthielten.

Ein möglicher Erklärungsversuch ist, dass die Cosmide gleich nach der Konjugation in *E. coli* den größten Teil des Inserts verloren hatten. Dadurch hatten sie im Vergleich zu den Klonen, die ein Insert und damit um bis zu 35 kbp mehr GC-reiche DNA beinhalteten, einen Wachstumsvorteil. Bei der Kultivierung für die Erstellung der Filter wuchsen diese dann schneller, was zu größeren Kulturen und mehr DNA-Material auf diesem Spot sorgte. Durch die Hybridisierung mit den sehr kurzen und teilweise degenerierten Sonden kam es dann dort durch unspezifische Bindung zu einem höheren Hintergrund, was zu den falsch-positiven Ergebnissen geführt haben könnte. Die Hybridisierung der 23 Nukleotide langen A2AS6-13-Sonde gegen diese defekten Cosmide zeigte, dass selbst nicht degenerierte Oligos dieser Länge durch unspezifische Bindung falsch-positive Cosmide anzeigen. Aufgrund der zu geringen Spezifität dieser kurzen Oligo-Sonden wurde dieser Ansatz zur Auffindung des Labyrinthopeptin-Biosysnthese-Genclusters verworfen.

5.1.4.8 Generierung einer größeren Sonde aus dem Bereich des Strukturgens mittels *single-specific-nested*-PCR

Die bisherigen Bemühungen, eine Sonde zu generieren, die die Lokalisierung des Biosynthese-Genclusters in der Cosmidbank ermöglicht, waren nicht erfolgreich. Die Identifizierung von ABC-Transportern war zu unselektiv und die Screeningansätze mit den kurzen degenerierten Sonden aus der Primärsequenz auf Verdaus genomischer DNA und die Cosmidbank zeigten keine reproduzierbaren oder falsch-positive Ergebnisse. Es war jedoch gelungen, mit Hilfe verlängerter degenerierter Primer ein 23 Nukleotide langes Stück aus dem Strukturgen zu sequenzieren (s. Abschnitt 5.1.4.6). Anhand dieses kurzen Sequenzteils konnten einige Informationen über die Sequenz des Strukturgens herausgefunden werden, wie z.B. die Orientierung der an den Lanthioninbrücken beteiligten Reste oder die Identität der Vorläufer des Labionin-Motivs. Ausgehend von diesen Fortschritten musste ein Weg gefunden werden, einen größeren Bereich mittels PCR zugänglich zu machen und so eine Sonde zu generieren.

Für diese PCR stand zum jetzigen Zeitpunkt nur ein spezifischer Primer (A2AS6-13) zur Verfügung. Für diesen speziellen Fall einer PCR gibt es in Lehrbüchern und Publikationen verschiedene Ansätze, die in der vorliegenden Arbeit getestet wurden. So wurde unter anderem die Etablierung einer inversen PCR versucht, was jedoch daran scheiterte, dass keine Bedingungen gefunden werden konnten, um ein recyklisiertes Templat zu erzeugen. Alle durchgeführten Positivkontrollen mit Primern aus der 16S rDNA waren stets negativ. Auch ein Vorgehen analog der Aufklärung des Strukturgens von Michiganin A blieb ohne Erfolg.[79] Dabei wurde versucht, die Fragmente eines genomischen DNA-Verdaus mit einem ebenfalls linearisierten Standard-Sequenziervektor, in diesem Falle pUC18, zu ligieren. Die anschließende PCR auf diese eventuell generierten hybriden Template mit verschiedenen Primern aus dem Strukturgen und aus der 16S rDNA als Kontrolle und einem Primer aus dem Vektor zeigten keine interessanten Banden. Dies war wenig verwunderlich, da die gemachten Erfahrungen bei der Handhabung der genomischen DNA von *A. namibiensis* im Laboralltag zeigten, dass schon „normale“ Ligationen nur mit sehr geringen Ausbeuten verliefen. Die Ligation, die bei dieser Strategie für die Generierung der Template sorgen sollte, könnte das Problem gewesen sein. Aus diesem Grund musste eine Methode gefunden werden, die keine molekularbiologische Vorbehandlung voraussetzte.

Deshalb wurde eine Kombination aus *single-specific*-Primer-PCR und *nested*-PCR (SSP-*nested*-PCR) gewählt, die in Abschnitt 4.2.3.3 ausführlich beschrieben wurde. Der unspezifische Primer für die erste PCR wurde dabei völlig willkürlich aus anderen Projekten und teilweise auch aus anderen Organismen gewählt, wie z.B. ein Primer für den Latrophilinrezeptor der Ratte. Als spezifische Primer für die erste PCR wurden degenerierte Primer aus dem Strukturgen (NVP1 oder NHP1) abgeleitet, wobei die in Abschnitt 5.1.4.6

gewonnenen Erkenntnisse berücksichtigt wurden. Die Enden der spezifischen Primer wurden so gewählt, dass mindestens die letzten fünf Nukleotide nicht mehr degeneriert waren. Für die zweite PCR wurden spezifische Primer aus dem mittleren, bekannten Sequenzbereich abgeleitet, die somit nicht degeneriert waren (NVP2 oder NHP2) (s. Abbildung 38).

Abbildung 38: Darstellung der spezifischen Primer für die SSP-*nested*-PCR. Die ersten Primer NVP1 und NHP1 sind stark degeneriert, so dass sie alle Codons berücksichtigen, wobei die 3´-Enden so gewählt wurden, dass diese im bekannten Sequenzbereich liegen und somit genau passen. Die Primer für die zweite PCR, NVP2 und NHP2, sind spezifisch und liegen auf einem vermeintlichen Produkt der ersten PCR.

In der ersten PCR entstanden aufgrund der unspezifischen Primer und der anfänglich sehr niedrigen Annealingtemperatur (10 Zyklen: 38°C; 30 Zyklen: 52°C) verschiedene PCR-Produkte, die durch Kombination zweier unspezifischer oder eines spezifischen und eines unspezifischen Primers zustande gekommen sein könnten. Aus diesem Gemisch unterschiedlichster PCR-Produkte sollte durch eine *nested*-PCR mit den zweiten spezifischen, nichtdegenerierten Primern (NVP2/NHP2) und den jeweiligen unspezifischen Primern nur die Template amplifiziert werden, die im ersten Durchgang aus dem spezifischen Primer entstanden waren. Durch die Verwendung der geschachtelten spezifischen Primer und die Abfolge an verschiedenen PCRs wurde erhofft, ausgehend vom bekannten Sequenzbereich noch zusätzliche Sequenzinformation zu erhalten.

Für die Auswertung der Gele war die Berücksichtigung einer Vielzahl von Kontrollen notwendig, da z.B. die unspezifischen Primer in beiden aufeinanderfolgenden PCR-Ansätzen enthalten waren und so die Bildung von PCR-Produkten aus zwei unspezifischen Primern möglich war. Die Auswertung der PCR-Ansätze erfolgte aus dem in Abbildung 39 gezeigten Agarose-Gel.

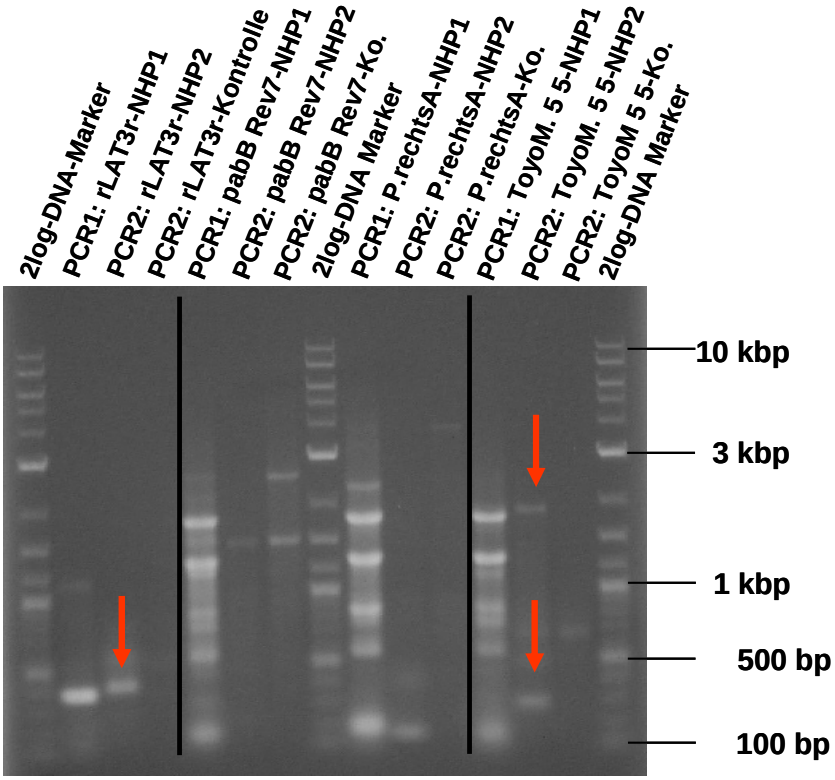


Abbildung 39: Auswertung der SSP-*nested*-PCR. Es wurden pro PCR-Ansatz zwei Kontrollen mit betrachtet, die erste PCR und die zweite PCR mit nur unspezifischem Primer. Es wurden Banden gesucht, die nur in den eigentlichen Ansätzen zu finden waren und nicht in den jeweiligen Kontrollen. Die drei mit Pfeilen markierten Banden wurden aus dem Gel eluiert, kloniert und sequenziert.

Es konnten insgesamt drei Banden gefunden werden, die nur im Ansatz und nicht in den Kontrollansätzen „erster PCR-Ansatz“ (SSP-PCR) und „zweiter PCR-Ansatz mit nur unspezifischem Primer“ zu beobachten waren. Diese drei Fragmente wurden aus dem Gel eluiert, kloniert und sequenziert. Die Auswertung der Sequenzierung ist am Beispiel der 0,4 kbp-Bande des rLat3r-NHP2-Ansatzes in Abbildung 40 gezeigt. Die Sequenz wurde zunächst auf das Vorkommen der eingesetzten Primer untersucht und dann so orientiert, dass die NHP2-Sequenz das 3´-Ende bildete. Dann wurde die Sequenz in dem vom NHP2-Primer vorgegebenen Leseraster in die entsprechende Aminosäuresequenz übersetzt. Die erhaltene Sequenz enthielt die aufgrund der bisherigen Ergebnisse vorhergesagte Primärsequenz (s. Abbildung 31D)). Die 0,3 kbp-Sequenz des Ansatzes Toyo-Modul 5 5-NHP2 zeigte bei der Auswertung die gleiche Sequenz, nur etwas kürzer, an. Das 1,8 kbp Fragment hingegen erbrachte keine Sequenz, die dem Cluster zugeordnet werden konnte. Man kann zusammenfassend sagen, dass es mittels dieser SSP-*nested*-PCR gelungen ist, eine 397 bp lange Sequenz aus dem Cluster zu bestimmen. Diese konnte im Folgenden als homologe Sonde zur Hybridisierung gegen die Filter der Cosmidbank eingesetzt werden.

Abbildung 40: Auswertung der Sequenzierung der 0,4 kbp-Bande der rLat3r-NHP2-PCR. Die Sequenz wurde so orientiert, dass der gefundene NHP2-Primer das 3´-Ende darstellte. Eine Übersetzung des davor liegenden Sequenzbereichs enthielt die erwartete kodierte Aminosäuresequenz (rot). Das war der Beweis, dass diese 0,4 kbp Bande aus dem Gencluster stammte.

5.1.4.9 Screening der Hybridisierungsfilter der Cosmidbank mit der homologen 0,4 kbp-Sonde

Mit der in Abschnitt 5.1.4.8 erhaltenen 397 bp langen Sequenz bestand nun endlich die Möglichkeit, die Cosmidbank auf den Hybridisierungsfiltern nach Biosynthese-Gencluster-haltigen Cosmiden zu untersuchen. Für die PCR wurde dazu ein neuer Vorwärtsprimer abgeleitet, so dass der Bereich des unspezifischen Primers aus der SSP-*nested*-PCR nicht enthalten war (A2-FwP1). Mit diesem Primer und NHP2 wurde dann die in Abschnitt 4.2.3.4 beschriebene PCR zur Generierung einer DIG-markierten Sonde (397 bp) durchgeführt. Die PCR-Bande wurde aus dem Gel gereinigt und nach dem Protokoll aus Abschnitt 4.2.2.4 gegen einen Filter hybridisiert (s. Abbildung 41). Die Auswertung des Filters erbrachte insgesamt zehn mehr oder weniger starke Signale. Diese Cosmide wurden dann mit dem *Bac-Kit* (s. Abschnitt 4.2.2.9) isoliert und der gleichen PCR unterzogen, mit der auch die Sonde generiert worden war. Von den zehn Cosmiden zeigten nur sechs die Bande bei 0,4 kbp, so dass in der Folge nur mit diesen weitergearbeitet wurde (s. Tabelle 25).

Von diesen sechs Cosmiden sollte eines sequenziert werden, um so die DNA-Sequenz des Biosynthese-Genclusters zu erhalten. Um eine Auswahl treffen zu können, wurden alle sechs Cosmide mit BamHI verdaut und nach einem 1%igen Agarosegel und anschließender Färbung mit SYBRGreen verglichen. Nach eingehender Betrachtung der Verdaus wurde Cosmid 1104 für die Sequenzierung ausgewählt (s. Abbildung 42).

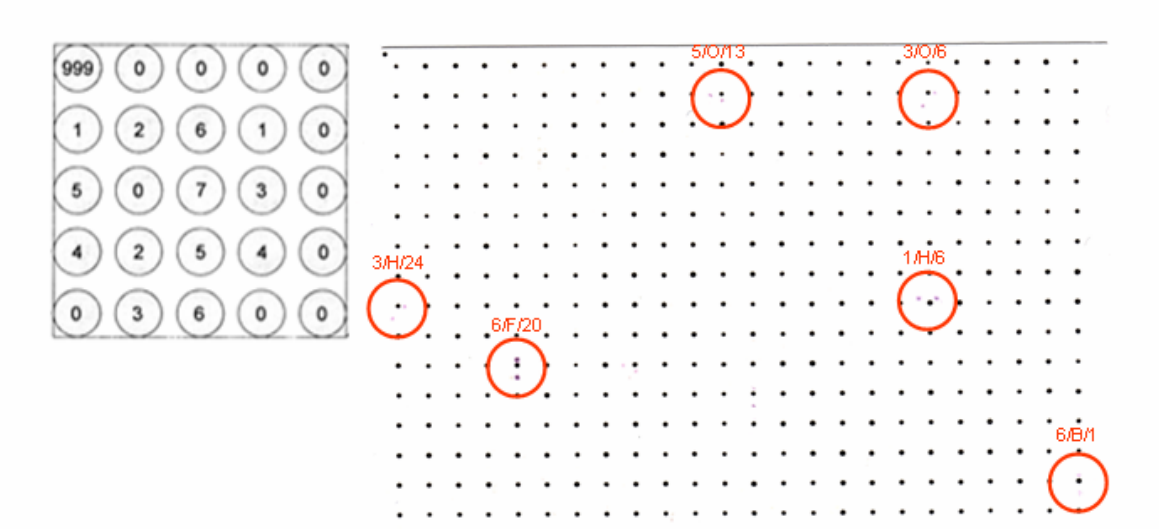


Abbildung 41: Auswertung des Filterscreenings mit der 397 bp großen homologen Sonde aus dem Labyrinthopeptin-Biosynthese-Gencluster. Die sechs rot markierten Cosmidklone zeigten sowohl auf dem Filter als auch bei einer anschließend durchgeführten Kontroll-PCR ein positives Ergebnis. Die Lage der positiven Spots auf dem Filter wurde mit dem Belegungsplan zunächst auf 384-well- und dann auf 96-well-Format umgerechnet. Die entsprechenden Klone wurden dann aus der Cosmidbank (96-well-Format) reaktiviert. (5/O/13 = 384-well-Platte 5; Koordinaten O/13; entspricht auf der 96-well-Platte 11 dem Klon H/17).

Tabelle 25: Auflistung der Cosmide, die sowohl auf dem Filter, als auch in der Kontroll-PCR (A2-FwP1-NHP2) ein positives Ergebnis gezeigt hatten.

Position auf dem Filter (Position/X/Y)	Position in der 384-well-Bank (Platte/X/Y)	Position in der 96-well-Bank (Platte/X/Y)	fortlaufende Nummer des Cosmids
5/O/13	5/O/13	17/H/7	1627
3/O/6	3/O/6	10/H/3	951
3/H/24	3/H/24	12/D/12	1104
1/H/6	1/H/6	4/D/3	327
6/F/20	6/F/20	24/C/10	2242
6/B/1	6/B/1	23/A/1	2113

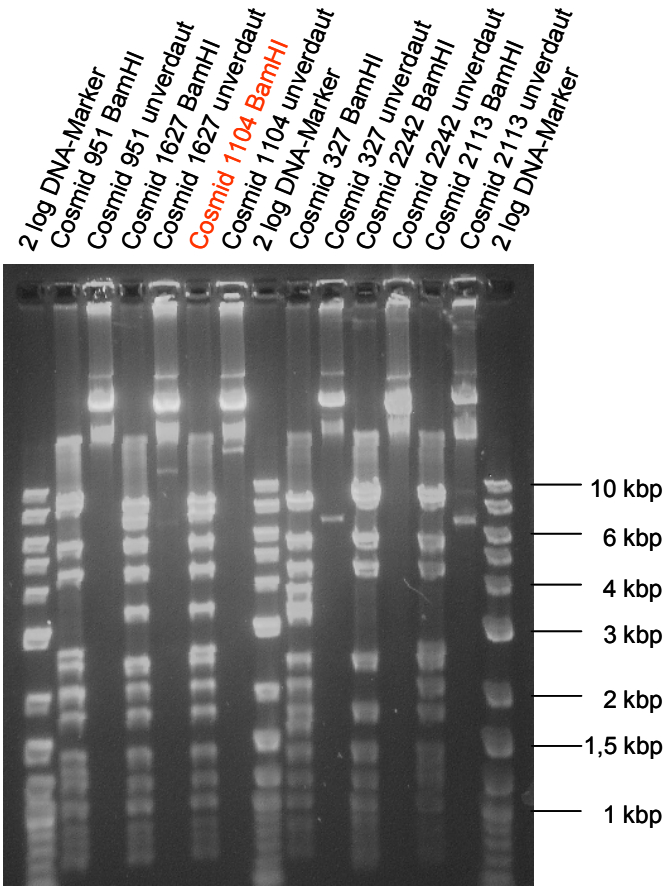


Abbildung 42: Vergleich der Labyrinthopeptin-Strukturgen-tragenden Cosmide. Es wurde nach dem Cosmid gesucht, welches das größte Insert enthält und damit die größte Wahrscheinlichkeit besitzt, das komplette Labyrinthopeptin-Biosynthese-Gencluster tragen. Der Vergleich der Fragmente ergab, dass die Cosmide 951, 1627, 1104, 2242 und 2113 sehr ähnliche Bandenmuster zeigten und nur Cosmid 327 deutliche Unterschiede aufwies. Aus den fünf relativ gleichen Cosmiden wurde schließlich Cosmid 1104 zur Sequenzierung ausgesucht.

5.1.5 Untersuchung der Integrität des Cosmids 1104

Die bisherigen Erfahrungen beim Arbeiten mit den Cosmiden hatten gezeigt, dass die Inserts der Cosmide nicht stabil sind (s. Abschnitt 5.1.4.7). Diese Beobachtung bestätigte sich bei der Auftragssequenzierung von Cosmid 1104. Dieses wurde für die Erstellung von dafür notwendigen Shotgun-Banken bei 37°C in *E. coli* angezogen und isoliert. Die Sequenzierung der Shotgun-Banken erbrachte Sequenzen, die nur einen kleinen Bereich der Gesamtsequenz darstellten. Um dieses Problem zu beheben, wurde das Cosmid 1104 ausgehend von klein gewachsenen Cosmidklonen aus der Originalbank bei lediglich 28°C angezogen, isoliert und für die Erstellung neuer Shotgun-Banken verwendet.

Vor der Auswertung und Annotation der schließlich erhaltenen DNA-Sequenzen sollte sichergestellt werden, dass diese vollständig den entsprechenden Bereich auf der genomischen DNA widerspiegeln und nicht Teile bei der Isolierung von Cosmid 1104 oder sogar schon bei der Erstellung der Cosmidbank verloren gegangen sind. Dazu wurde an der

erhaltenen Cosmidsequenz eine Reihe von Primerpaaren abgeleitet, die für PCR-Ansätze mit Cosmid und genomischer DNA als Templat verwendet wurden (s. Abbildung 43). Die Auswertung der analogen PCR-Ansätze zeigte für diese sechs Primerpaare Banden auf jeweils gleicher Höhe, was für die Primerpaare IKoFw5/Rev5 und IKoFw6/Rev6 in Abbildung 44 exemplarisch gezeigt ist. Aufgrund dieser Ergebnisse kann davon ausgegangen werden, dass zumindest dieser Teil auf Cosmid 1104 mit dem entsprechenden Bereich auf dem Genom übereinstimmt.

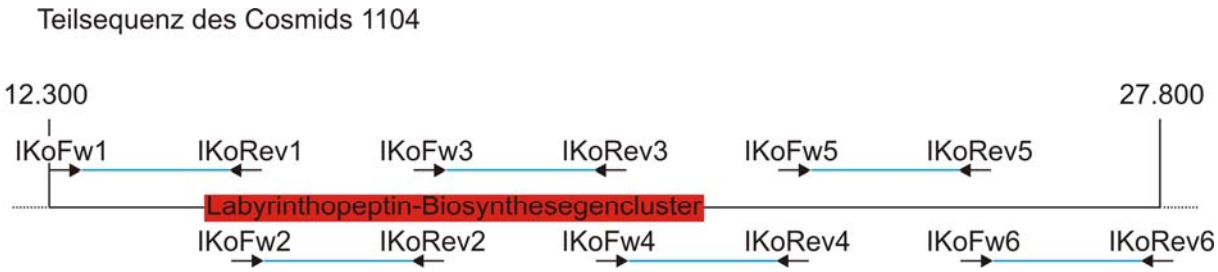


Abbildung 43: Übersicht über die PCR-Ansätze zur Kontrolle der Cosmidintegrität. In dem untersuchten Bereich befindet sich mittig das Biosynthese-Gencluster (s. Abschnitt 5.2.1).

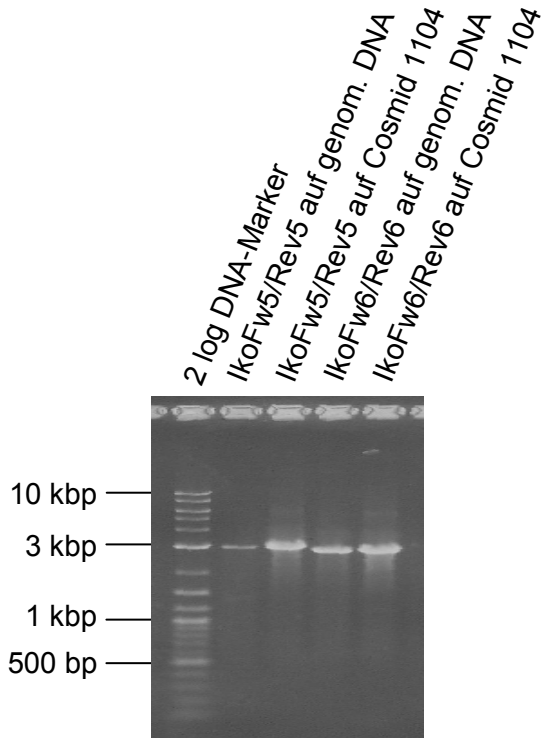


Abbildung 44: Auswertung der PCR-Ansätze zur Kontrolle der Integrität des Cosmids 1104 exemplarisch am Beispiel der Primerpaare IKoFw5/Rev5 und IKoFw6/Rev6. Man kann in den Ansätzen mit dem gleichen Primerpaar auf die unterschiedlichen Template jeweils eine PCR-Bande auf gleicher Höhe beobachten.

5.2 Auswertung und Annotation der Sequenzierung des Cosmids 1104

Wie in Abschnitt 5.1 ausführlich beschrieben, ist es im Zuge dieser Arbeit gelungen, das gesuchte Biosynthese-Gencluster nach Generierung einer ca. 400 bp großen, homologen Sonde mittels Filterscreening auf sechs Cosmiden zu lokalisieren. Die Analyse der isolierten Cosmide (s. Abschnitt 4.2.2.9) auf einem Agarosegel (s. Abbildung 42) erwies Cosmid 1104 als geeignet für die Sequenzierung, die von der Firma Agowa GmbH, Berlin, durchgeführt wurde. Zunächst erfolgte die Sequenzierung mittels einer Shotgun-Bank, wobei jedoch nur der Vektor und insgesamt 15 kbp Insert, aufgeteilt in drei Contigs, erhalten werden konnten. Aufgrund dieser Schwierigkeiten wurde auf Primerwalking umgestellt. An machen Stellen waren jedoch auch hierbei große Anstrengungen nötig, da das Primerwalking aufgrund ausgiebiger Sekundärstrukturbildung nur unter modifizierten Bedingungen möglich war. So konnte schließlich in mühsamer Kleinstarbeit die 36826 bp Insertsequenz von Cosmid 1104 erhalten werden. Die Sequenz wurde abschließend durch Gegenstrangsequenzierung verifiziert.

Allgemein ist die Auswertung genomischer Sequenzen aus Streptomyceten und anderen GC-reichen Mikroorganismen schwierig, da aufgrund des geringen AT-Gehalts häufig große *open-reading frames* (ORFs) in allen sechs möglichen Leserastern zu finden sind.[57] Ein ORF ist definiert als der Abschnitt eines Leserasters einer Sequenz, der durch zwei Stop-Codons begrenzt ist. Erschwerend kommt dann oftmals noch hinzu, dass die Ribosomen-Bindungsstelle (RBS) als Indikator eines Translationsstartpunktes nicht erkennbar oder vorhanden ist.[80]

Die Auswertung der 36826 bp langen Sequenz des Cosmids 1104 erfolgte durch Kombination verschiedener Programme, wobei vor allem die in Abschnitt 4.3 beschriebenen Programme Blast, Artemis und FramePlot verwendet wurden. Dabei wurde zunächst mit Hilfe des FramePlot-Programms nach ORFs gesucht, wodurch Leseraster und Orientierung der vermeintlichen Proteine bestimmt werden konnten. Durch Blasten dieser gefundenen Bereiche wurde im Folgenden versucht, den darin kodierten Proteinen eine Funktion durch Abgleich gegen Datenbanken zuzuordnen. Für die gefundenen Proteine, die dem Labyrinthopeptin-Gencluster zugeordnet werden konnten, wurde dann noch mittels exakter Analyse der Blast-Ergebnisse und des FramePlots versucht, den Startpunkt der Translation zu definieren.

5.2.1 Auswertung des Sequenzabschnittes mit den Genen des Biosynthese-Genclusters für Labyrinthopeptin A2

Nachdem die Sequenz des Cosmids 1104 vollständig und in guter Qualität zur Verfügung stand, war das primäre Ziel die Identifizierung von Protein-kodierenden Genen, die an der Biosynthese von Labyrinthopeptin A1, A2 und A3 beteiligt sind. Neben den Genen für die Biosynthese wurden außerdem Gene für die Regulation und die Immunität des Produzentenstammes erwartet, sowie die Strukturgene der Derivate A1 und A3. Dazu wurden zunächst die in direkter Umgebung des Strukturgens von Labyrinthopeptin A2 liegenden ORFs betrachtet (s. Abbildung 45).

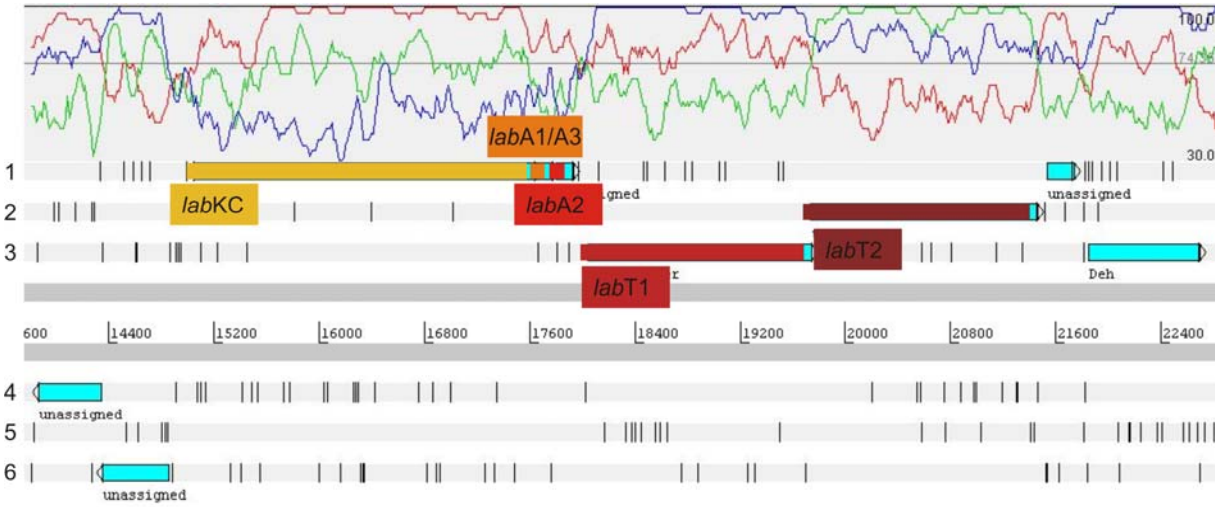


Abbildung 45: Artemis-Darstellung des Bereichs um das Strukturgen von Labyrinthopeptin A2 (LabA2). Gezeigt sind der FramePlot und die sechs möglichen Leseraster (1-6), wobei die kleinen vertikalen schwarzen Balken die Begrenzungen von ORFs anzeigen. Die aus dem FramePlot abgelesenen Gene sind in das jeweilige Leseraster eingetragen.

Der FramePlot zeigte stromaufwärts von *labA2* zunächst einen weiteren kleinen ORF (von *labA1/A3*) und dann einen großen (von *labKC*). Stromabwärts wurden zwei ORFs (von *labT1/T2*) gefunden, wobei sich diese über einen größeren Bereich überschneiden. In den folgenden Abschnitten soll auf die weitere Untersuchung dieser ORFs eingegangen werden.

5.2.1.1 Untersuchung der ORFs *labA2* und *labA1/3*

Wie im vorherigen Abschnitt gezeigt, wurde bei der Suche ausgehend von dem Strukturgen für Labyrinthopeptin A2 stromaufwärts ein weiterer kleiner ORF (111 bp) gefunden, bei dem es sich um das Strukturgen für Labyrinthopeptin A1 und/oder A3 handeln könnte. Deshalb wurden die bisherigen Kenntnisse (s. Abschnitt 1.7) zu den Labyrinthopeptinen A1 und A3 auf Übereinstimmung mit der Aminosäuresequenz des ORFs geprüft. Dabei zeigte sich, dass sowohl die für Labyrinthopeptin A1 durchgeführte Aminosäurenanalytik als auch die

mittels Edman-Abbau bestimmten *N*-terminalen Teilsequenzen von A1 und A3 in Einklang mit der hier kodierten Aminosäuresequenz waren. Daraus ergab sich die Hypothese, dass beide Labyrinthopeptine A1 und A3 durch dieses ORF kodiert werden und aus einem gemeinsamen Präpropeptid durch um eine Aminosäure versetzte Abspaltung der Leadersequenz entstehen. Allerdings ist das Vorkommen von zwei Proteasespaltstellen unwahrscheinlich, weshalb im einen Fall von einer fehlerhaften proteolytischen Spaltung ausgegangen werden muss. Dafür spricht auch die Beobachtung, dass bei längerer Kultivierung LabA3 in LabA1 umgewandelt wird, was durch eine nachträgliche Prozessierung durch eine Protease erklärt werden könnte. Bei genauerer Betrachtung der Propeptide von LabA1 und LabA3 waren deutliche Homologien zu der bekannten Sequenz des Propeptids von LabA2 zu erkennen (s. Abbildung 46 A)).

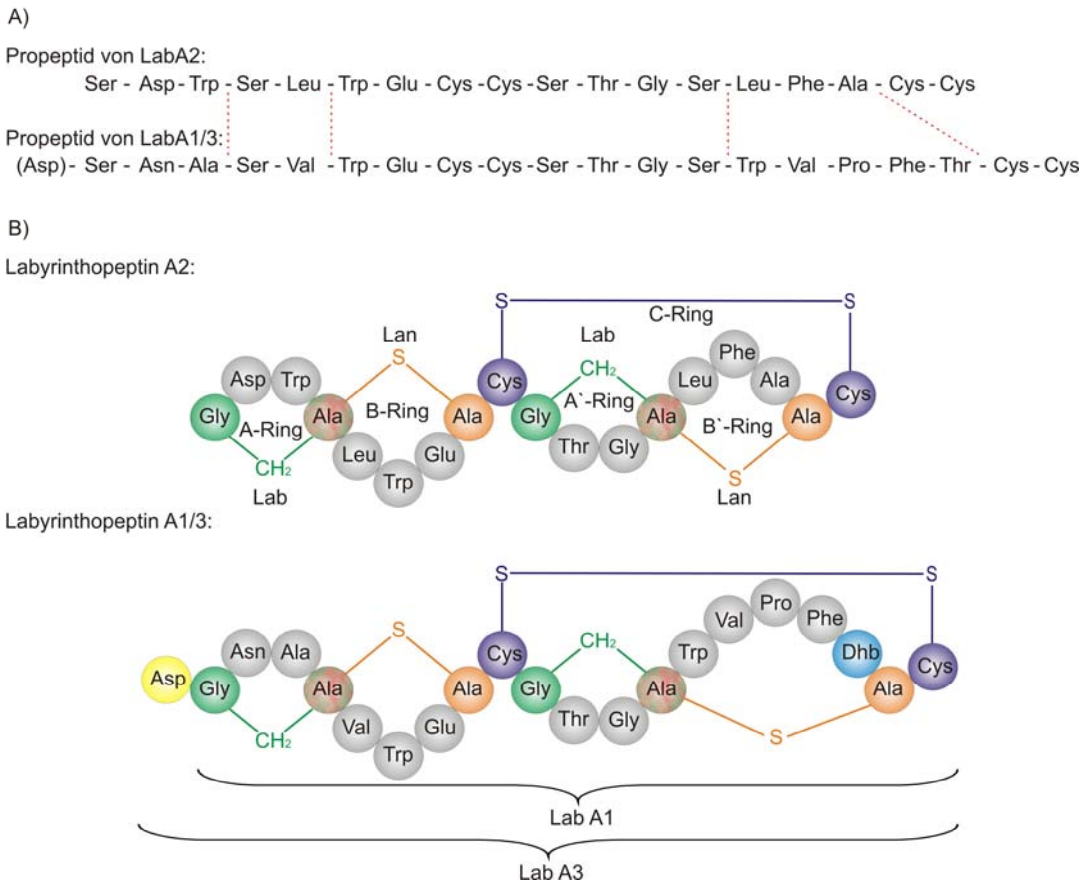


Abbildung 46: A) Gegenüberstellung der strukturgebenden Propeptide von LabA2 und LabA1/A3. B) Struktur von LabA2 in Gegenüberstellung mit den hypothetischen Strukturen von LabA1/A3.

Deshalb wurden die in Abbildung 46 B) dargestellten Strukturvorschläge für LabA1 und A3 gemacht, die von einer zu LabA2 analogen posttranslationalen Modifikation ausgingen und mit den Ergebnissen aus Molekülmassenanalysen und Summenformelbestimmungen übereinstimmten. So zeigten die Propeptide das gleiche SXXSXXXCCSXXS-Motiv (X = beliebige Aminosäure), was eine gleiche Verbrückung (Labionin-Lanthionin-Labionin) der Aminosäuren in diesem Bereich vermuten ließ. Aufgrund der Molekülmassen und der daraus

kalkulierten Summenformeln wurde erstens die Existenz einer zweiten Lanthioninbrücke vermutet, auch wenn dadurch im Gegensatz zu LabA2 eine Lanthioninbrücke über fünf statt über drei Aminosäuren vorliegt und zweitens von einer Wasserabspaltung an einem Thr ausgegangen werden muss. In NMR-Spektren konnte ein für Lantibiotika typischer Dhb-Rest in Nachbarschaft zu Phe lokalisiert werden, was diese Wasserabspaltung bestätigte. Drittens konnte durch eine Reduktion mit DTT die Existenz einer Disulfidbrücke nachgewiesen werden. Die vorgeschlagenen Strukturen stimmten somit mit den bisherigen analytischen Daten sehr gut überein, was in Tabelle 26 für LabA1 verdeutlicht wird.

Tabelle 26: Kalkulation der Molekülmassen und dazu gehörenden Summenformeln ausgehend von dem Propeptid von Labyrinthopeptin A1. Die kalkulierte Molekülmasse und Summenformel des „reifen" LabA1 stimmt sehr gut mit der gemessenen Molekülmasse und daraus bestimmten Summenformel (2073,7624; $C_{92}H_{119}N_{23}O_{25}S_4$) überein.

	Labyrinthopeptin A1 (kalkulierte Molekülmasse; Summenformel)
Propeptid	2165,8; $C_{92}H_{131}N_{23}O_{30}S_4$
+ Lab-Lan-Lab-Motiv (Ring A/B/A`)	2111,8; $C_{92}H_{125}N_{23}O_{27}S_4$
+ Lan-Motiv (Ring B`)	2093,8; $C_{92}H_{123}N_{23}O_{26}S_4$
+ Dhb	2075,8; $C_{92}H_{121}N_{23}O_{25}S_4$
+ Disulfid (Ring C)	2073,8; $C_{92}H_{119}N_{23}O_{25}S_4$

Der in Abbildung 47 dargestellte Vergleich der Strukturen zeigt, dass neben den an den Verbrückungen beteiligten Aminosäuren außerdem die in den Peptiden zentral angeordneten Aminosäuren in allen drei Verbindungen hochkonserviert sind (rot). So sind die Aminosäuren 6 (LabA1) bzw. 7 (LabA3) bis 13 identisch mit den entsprechenden Aminosäuren von LabA2. Die Aminosäure ^{5}Leu von LabA2 entspricht Val in LabA1 und LabA3 (grün), was nahezu keine Veränderung der Polarität bewirkt. Die einzigen deutlichen Unterschiede betreffen die zwei Aminosäuren des A-Rings und die Reste, die an der Ausbildung des B`-Rings beteiligt sind (blau).

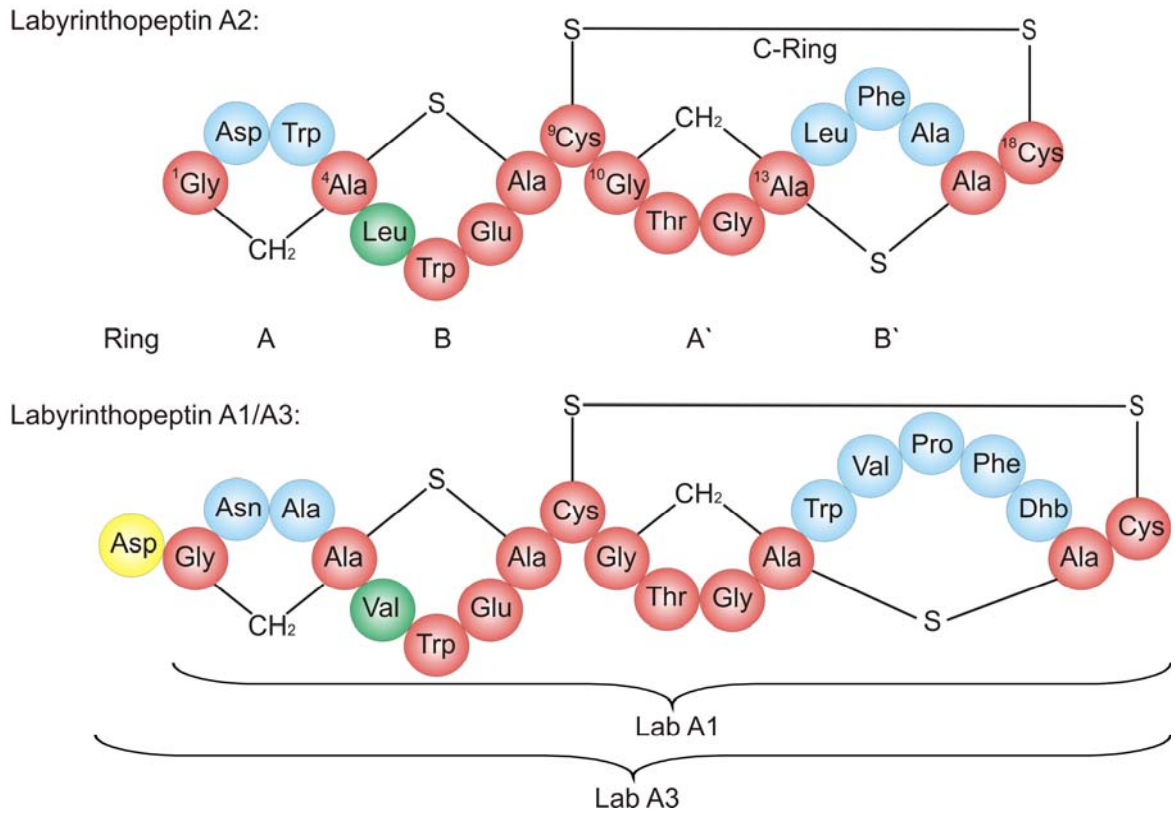


Abbildung 47: Vergleich der Struktur von Labyrinthopeptin A2 mit den hypothetischen Strukturen der Labyrinthopeptine A1 und A3. Rot markiert wurden identische Aminosäuren, blau und grün unterschiedliche Aminosäuren. Grün wurde für die Fälle gewählt, bei denen keine großen Polaritätsänderungen auftraten. Die zusätzliche Aminosäure Asp von LabA3 wurde gelb hervorgehoben.

Der A-Ring in LabA2 besteht aus der großen aromatischen Aminosäure Trp und der unter physiologischen Bedingungen negativ geladenen Aminosäure Asp. Diese Positionen sind in LabA1/A3 entsprechend durch Ala und ein zwar polares aber nicht geladenes Asn ersetzt. Zusammenfassend kann man sagen, dass der sperrige, geladene A-Ring in LabA2 durch einen kleinen, polaren A-Ring in LabA1/A3 ersetzt ist. Noch größer sind die strukturellen Unterschiede bezüglich des B`-Rings, an dem bei LabA2 drei unpolare Aminosäuren beteiligt sind, bei LabA1/A3 hingegen fünf unpolare Aminosäuren.

Stromaufwärts der Basentripletts für die Propeptide werden die Leadersequenzen kodiert, die zusammen die sog. Präpropeptide darstellen. Um den Translationsstartpunkt der Präpropeptide und damit deren Länge bestimmen zu können, wurden die beiden ORFs von LabA2 und LabA1/A3 miteinander verglichen. Eine Aussage aus den FramePlots war aufgrund der Kürze der ORFs nicht möglich. Der ORF von LabA1/A3 enthielt nur ein Startcodon und zeigte vor diesem (-13 bis -17) auch die typische Sequenz für eine RBS (GGAGG). Der ORF für LabA2 enthielt hingegen zwei mögliche Translationsstartpunkte und keine typische RBS in den Sequenzen davor.

Aufgrund der hohen Homologien in den Leadersequenzen von LabA2 und LabA1/A3 (s. Abbildung 48 A)) wird jedoch angenommen, dass das zweite ATG als Startcodon fungiert.

Somit ergab sich für LabA3 eine 19 und für LabA1/A2 eine 20 Aminosäuren lange Leadersequenz, die eine 60%ige Sequenzidentität zeigten. Der Vergleich mit einigen bekannten Leadersequenzen anderer Lantibiotika (s. Abbildung 48 B)) zeigte deutlich, dass die gefundenen Leadersequenzen für LabA1/A2/A3 nicht in die bisher definierten Klassen (FNLD- oder GG-Typ) eingegliedert werden konnten.[23] Außerdem unterschieden sich die gefundenen Leadersequenzen aufgrund ihrer extremen Kürze von nur 19 bzw. 20 Aminosäuren von den bisher gefundenen, die aus 23 bis 59 Aminosäuren bestehen. Diese Ergebnisse sprachen dafür, dass ein neuer Leadersequenztyp gefunden worden ist, der sich vor allem durch seine extrem kurzen Aminosäuresequenzen auszeichnet.

A)

Aminosäuresequenz des ORFs des Gens *labA1/A3*:

TSTMASILELQDLEVERASSAADSNASVWECCSTGSWVPFTCC

Aminosäuresequenz des ORFs des Gens *labA2*:

RPHTVPPMRGDSPMASILELQNLDVEHARGENRSDWSLWECCSTGSLFACC

B)

"FNLD"-Motiv:

```
Nisin A/Z                          MSTKDFNLDLVSVSKK--DSGASPR
Subtilin/Ericin S                  MSKFDDFDLDVVKVSKQ--DSKITPQ
Epidermin                        MEAVKEKNDLFNLDVKVNAKESNDSGAEPR
Gallidermin/Staphylococcin T     MEAVKEKNELFDLDVKVNAKESNDSGAEPR
Mutacin 1140/I                MSNTQLLEVLGTETFDVQEDLPAFDTTDTTIFASNDDPDTR
Pep5                               MKNNKNLFDLEIKKETSQNTD-ELEPQ
```

"Doppelglycin"-Motiv:

```
Lacticin 481                       MKEQNSPNLLQEVTESELDLILGA
Streptococcin A-FF22               MEKNNEVINSIQEVSLEELDQIIGA
Variacin                             MTNAFQALDEVTDAELDAILGG
Butyrivibriocin OR79                MNKELNALTNPIDEKELEQILGG
Mutacin II                        MNKLNSNAVVSLNEVSDSELDTILGG
Salivaricin A                     MKNSKDILNNAIEEVSEKELMEVAGG
Salivaricin A1                    MKNSKDILTNAIEEVSEKELMEVAGG
Sublancin 168                            MEKLFKEVKLEELENQKGS

Cinnamycin  MTASILQQSVVDADFRAALLENPAAFGASAAALPTPVEAQDQASLDFWTKDIAATEAFA
Mersacidin  MSQEAIIRSWKDPFSRENSTQNPAGN          PFSELKEAQMDKLVGAGDMEAA

Lactocin S  MTKEKKVLDELSLHASAKMGARDVESSMNAD

Labyrinthopeptin A1/3                    MASILELQDLEVERASSAA D
Labyrinthopeptin A2                      MASILELQNLDVEHARGENR
```

Abbildung 48: A) Vergleich der Aminosäuresequenzen der beiden ORFs für die Labyrinthopeptine. Den Propeptidsequenzen (rot) gehen die jeweiligen Leadersequenzen (blau) voraus. Der Beginn der Leadersequenz war im Falle von LabA1/A3 eindeutig, wohingegen im ORF von LabA2 zwei Met-Reste kodiert waren und somit zwei Startmöglichkeiten bestanden. Aufgrund der Sequenzhomologien wurde das zweite Met als Translationsstart vorgeschlagen. B) Vergleich der LabA1/A2/A3-Leadersequenzen mit anderen bekannten Leadersequenzen.

Die außerordentlich hohe Homologie der beiden Leadersequenzen deutete auch darauf hin, dass die Labyrinthopeptine durch die gleichen Enzyme posttranslational modifiziert werden müssen, die gewöhnlich ihre Substrate an der Leadersequenz erkennen. Um zu überprüfen, ob ähnliche Sequenzen schon einmal beschrieben worden sind, wurden die Präpropeptidsequenzen lokal geblastet. Die Blasts ergaben jedoch keinen einzigen Hit.
Zusammenfassend kann also gesagt werden, dass durch die Auffindung der Strukturgene Strukturen für LabA1/A3 vorhergesagt werden konnten, die zusammen mit der bereits bekannten Struktur von LabA2 einen neuen Strukturtyp von Lantibiotika bilden. Für das Hauptcharakteristikum dieses Strukturtyps, das Labionin-Lanthionin-Motiv, scheint das $SXXS(X)_{3\text{-}5}C$-Motiv in der Primärsequenz eine Grundvoraussetzung zu sein. Ein weiteres Charakteristikum sind die lediglich 19 bzw. 20 Aminosäuren langen Leadersequenzen, die sich außerdem deutlich von der bisherigen Klassifizierung von Leadersequenzgruppen bei Lantibiotika unterscheiden.

5.2.1.2 Untersuchung des ORFs von *labKC*

Vor den Strukturgenen *labA1/A3* und *labA2* wurde ein großer ORF gefunden, der ein Gen beinhaltete, das im Folgenden als *labKC* bezeichnet wird. Zunächst sollte versucht werden, diesem Gen bzw. dem dadurch kodierten Protein eine Funktion zuzuweisen. Dazu wurde mit der gesamten ORF-Sequenz ein lokaler BlastX durchgeführt. In Tabelle 27 sind die signifikantesten Ergebnisse aufgelistet.

Tabelle 27: Ergebnisse des BlastX mit dem ORF des Gens *labKC*. Aufgeführt sind die besten vier Ergebnisse.

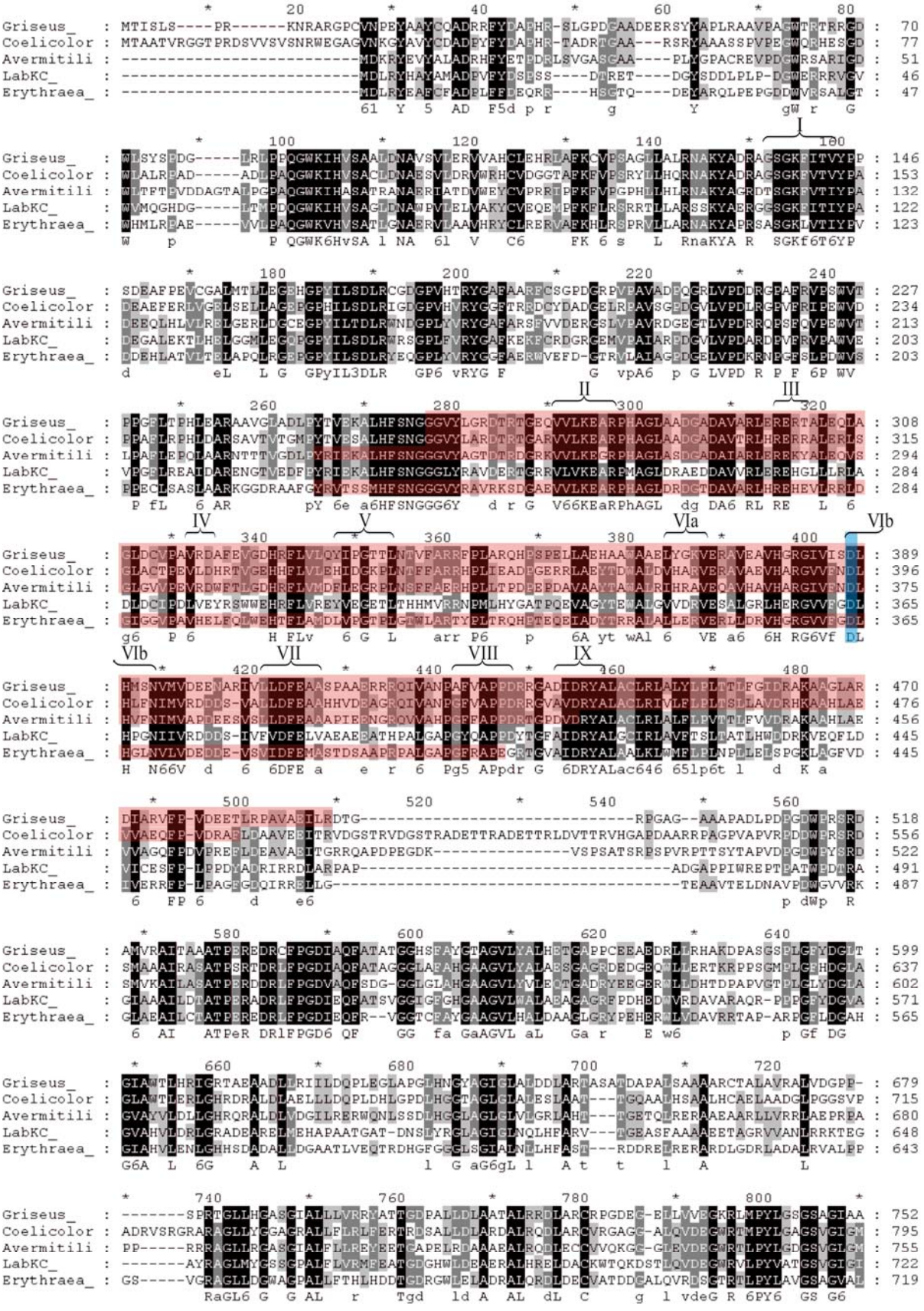
Referenz (Genname)	Beschreibung	Score (bits)	E-value
NP_630756.1 (*ram*C: SCO6681)	serine/threonine protein kinase [*S. coelicolor* A3(2)]	629	e-178
YP_001106424.1 (SACE_4230)	serine/threonine protein kinase [*Saccharop. erythraea* NRRL 2338]	619	e-175
NP_828679.1 (SAV7503)	membrane translocator [*S. avermitilis* MA-4680]	575	e-162
BAB62264.1 (AmfT)	membrane translocator [*Streptomyces griseus*]	561	e-158

Die Blastergebnisse zeigten signifikante Homologien sowohl zu Ser/Thr-Protein-Kinasen von *Streptomyces coelicolor* und *Sacharopolyspora erythraea* als auch zu Membrantranslokatoren von *Streptomyces avermitilis* und *Streptomyces griseus.* Alle weiteren Hits zeigten ebenfalls Homologien zu Ser/Thr-Kinasen an, jedoch mit deutlich geringerer Signifikanz (Score <350; E-value; >e-93). Anhand dieser Ergebnisse konnte das erste ATG im ORF als der wahrscheinliche Translationsstartpunkt von LabKC bestimmt werden, da die *N*-Termini der gefundenen homologen Proteine sehr gut zu diesem auch im FramePlot ablesbaren Beginn passten. Allerdings war im stromaufwärts liegenden Sequenzbereich keine RBS zu finden. Für alle weiteren Arbeiten wurde dieser Startpunkt und die daraus resultierende 862 Aminosäuren lange Proteinsequenz für das LabKC-Protein angenommen und verwendet.

Aufgrund der Tatsache, dass alle Hits außer den homologen Proteinen aus *S. avermitilis* und *S. griseus* Ser/Thr-Kinasen waren, sollte zunächst geprüft werden, ob es sich bei den beiden Membrantranslokatoren, die eine sehr hohe Homologie zu den anderen Ser/Thr-Kinasen aufweisen, nicht auch um Ser/Thr-Kinasen handeln könnte. Die Annotation ganzer Genome aufgrund von DNA-Sequenzdaten erlaubt es heute nicht mehr, jedes einzelne Gen und Protein zu charakterisieren und deren katalytische Aktivität zu bestimmen. Dies hat zur Folge, dass die Annotationen häufig automatisch aufgrund von reinen Datenbanksuchen erfolgen. Diese Automatisierung kann zu falschen bzw. uneindeutigen Annotationen führen. Um diesen Sachverhalt für die in Tabelle 27 aufgeführten Proteinsequenzen zu untersuchen, wurden diese den entsprechenden Datenbanken entnommen und mit dem LabKC-Protein verglichen (s. Abbildung 49).

Das Alignment zeigte für alle fünf Sequenzen eine sehr hohe Homologie, die sich über die gesamte Sequenz verteilte. Dies begründet noch einmal die Vermutung, dass es sich bei den beiden Proteinen aus *S. avermitilis* und *S. griseus* ebenfalls um Ser/Thr-Kinasen handeln könnte. Um diese Vermutung weiter zu verifizieren, sollten die Sequenzen auf konservierte Regionen von Ser/Thr-Kinasen mit Hilfe der **KinG**-Software untersucht werden.[81,82,83] Die Domänensuche erbrachte für alle vier Proteine jeweils eine Kinase-Domäne, die in Abbildung 49 rot markiert wurden. Die Lokalisation der Kinase-Domäne im LabKC-Protein konnte allerdings aus patentrechtlichen Gründen nicht direkt durch Eingabe in dieses Softwaretool bestätigt werden. Außerdem zeigte die **KinG**-Software für alle vier Proteine den gleichen katalytischen Asp-Rest an, der blau hervorgehoben wurde und auch im LabKC-Protein konserviert vorliegt. In Abbildung 49 wurden außerdem die für NP_630756.1 aus *S. coelicolor* publizierten, für Ser/Thr-Kinasen typischen Hanks–Motive[81] eingetragen und mit römischen Zahlen nummeriert.[83,84] In den konservierten Hanks-Motiven liegen hauptsächlich die Aminosäuren, die an der Ausbildung des katalytischen Zentrums beteiligt sind.[81,85] Bei NP_630756.1 handelte es sich um das Protein RamC aus *S.*

coelicolor,[41] welches bisher als einzigstes näher untersucht worden ist. Allerdings konnte weder dessen Aktivität nachgewiesen noch dessen Substrat benannt werden.



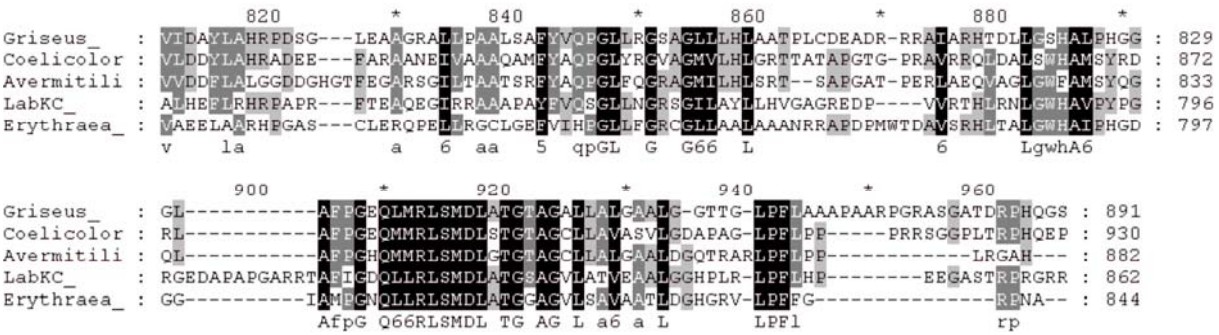
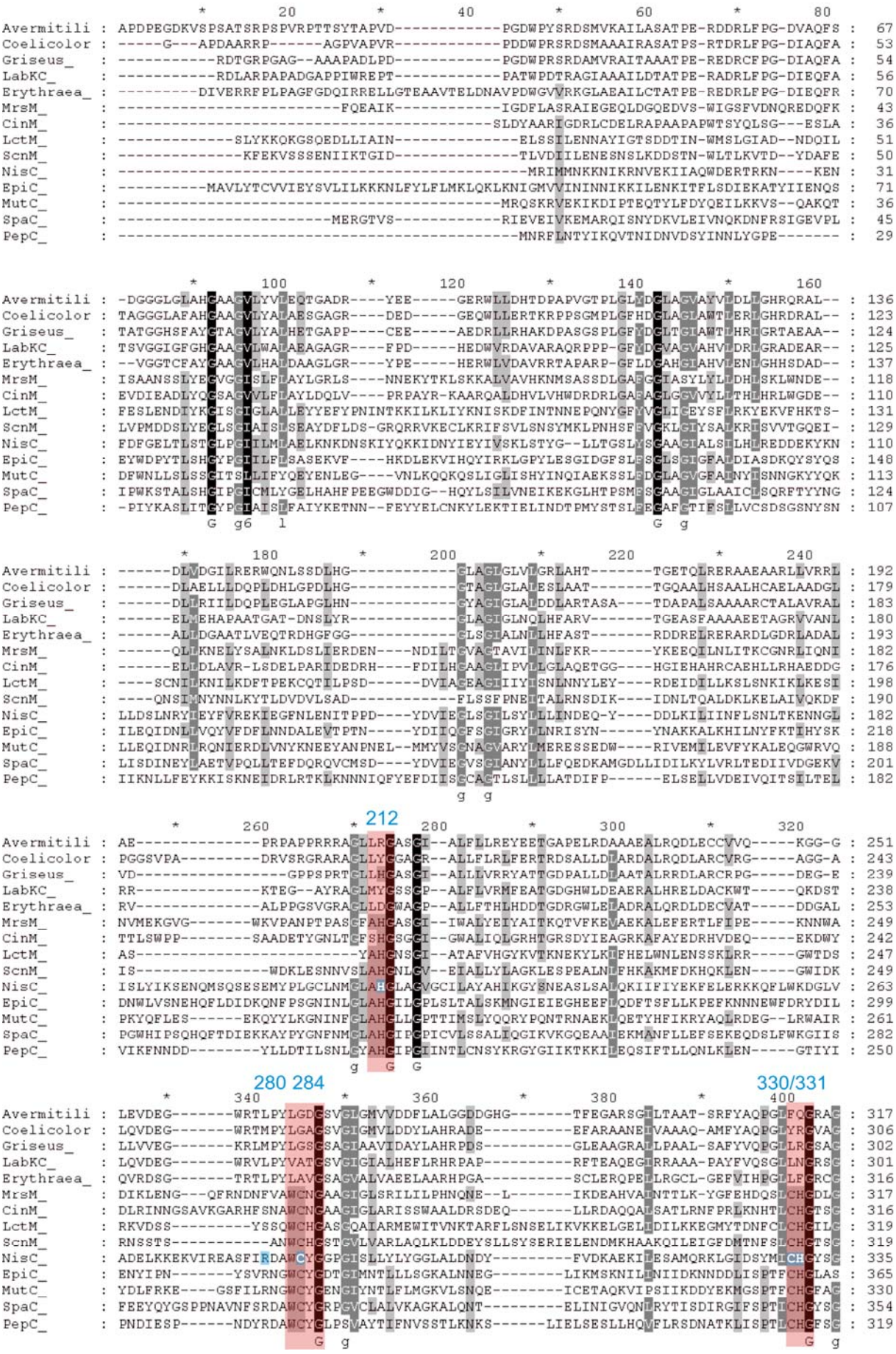


Abbildung 49: Sequenzalignment der Proteinsequenzen des LabKC-Proteins und der vier signifikantesten Hits des BlastX-Ergebnisses (s. Tabelle 27). Die Sequenzen aus der Datenbank wurden mit Hilfe des Softwaretools **KinG** (**K**inases **in** **G**enome; Internetseite: **http://hodgkin.mbu.iisc.ernet.in/~king/**) auf Ser/Thr-Kinase-Domänen untersucht und die gefundenen Domänen rot markiert. Der katalytische Asp-Rest war hoch konserviert und wurde deshalb blau markiert. Die Hanks-Motive[81] des Proteins RamC aus *S. coelicolor* wurden mit römischen Zahlen nummeriert.[84]

Anschließend sollte untersucht werden, ob neben der *N*-terminalen Ser/Thr-Kinase-Domäne noch weitere Domänen im LabKC-Protein vorkommen. Auch hier konnte diese Untersuchung aus Patentschutzgründen nur indirekt an den vier homologen Proteinen durchgeführt werden. Dazu wurden die Datenbankeinträge dieser vier signifikanten Hits in InterproScan und Swissprot betrachtet (s. Tabelle 27). Alle vier Proteine enthielten demnach *C*-terminal eine LanC-Domäne, was deshalb auch für das LabKC-Protein angenommen wurde. Das Vorkommen einer solchen Domäne war aufgrund der bisherigen Kenntnisse zur Biosynthese von Lantibiotika im Biosynthese-Gencluster erwartet worden. Wie in Abschnitt 1.3.2 beschrieben, enthalten LanC-Domänen typischerweise hoch konservierte Aminosäure-Motive (AHG; WCXG und CHG), die für die Koordination des katalytisch aktiven Zn-Restes verantwortlich sind. Deshalb wurde versucht, diese Aminosäure-Motive im *C*-terminalen Bereich der Ser/Thr-Kinasen zu finden. Dazu wurden verschiedene Alignments mit bekannten LanC- und LanM-Sequenzen herangezogen, die jedoch alle keine Hinweise auf das Vorhandensein dieser Aminosäure-Motive zeigten (s. Abbildung 50). Möglicherweise liegen hier also LanC-Domänen mit einem anderen Katalysemechanismus vor.


212
280 284
330/331

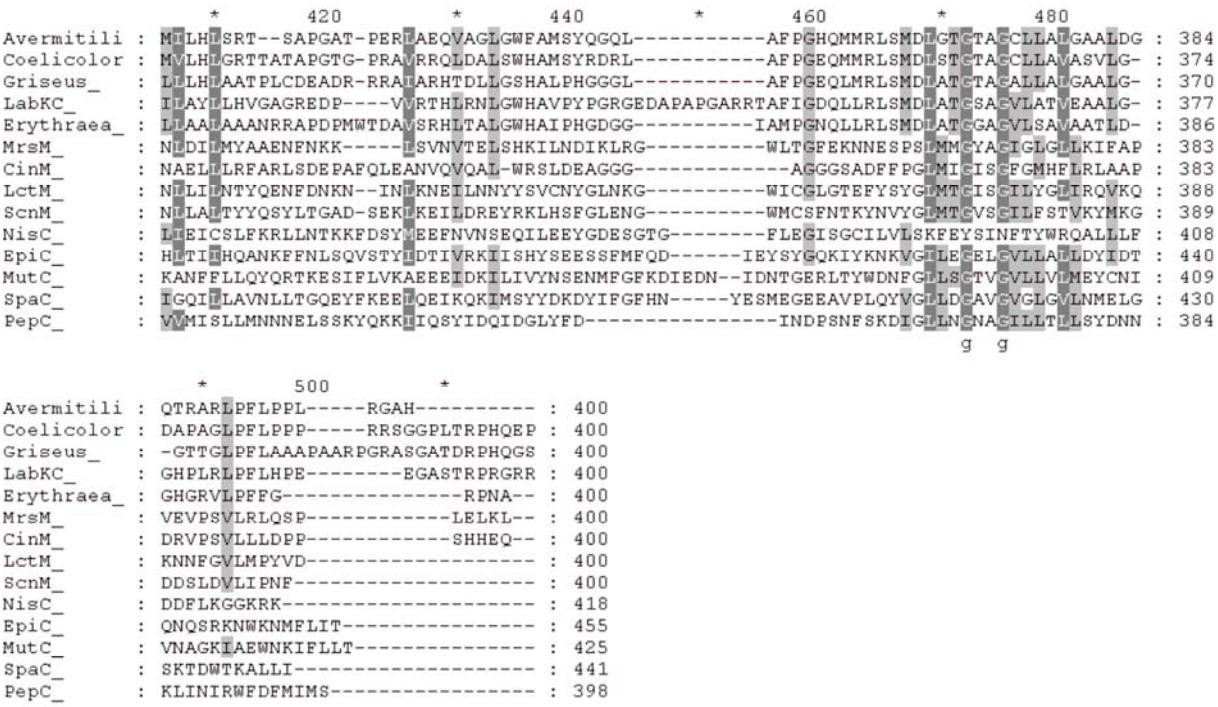


Abbildung 50: Proteinalignment von bekannten LanC-Sequenzen und von entsprechenden *N*-terminalen (400 Aminosäuren) Bereichen verschiedener LanM- und LanKC-Proteine. Die für NisC bestimmten Reste des aktiven Zentrums wurden blau markiert. Legende: EpiC = LanC für Epidermin und analog dazu: MutC (Mutacin I); SpaC (Subtilin); NisC (Nisin); PepC (Pep5); LactM = LanM von Lacticin481 und analog dazu: ScnM (StreptococcinA-FF22); MrsM (Mersacidin); CinM (Cinnamycin).

Zusammenfassend kann also gesagt werden, dass das LabKC-Protein eine Ser/Thr-Kinase-Domäne und eine LanC-Domäne besitzt, wobei letzterer die typischen Aminosäure-Motive für die Zn-Komplexierung fehlen. Die potentielle Rolle dieses Enzyms wird in der vorliegenden Arbeit zu einem späteren Zeitpunkt (s. Abschnitt 5.2.1.5) diskutiert. Außerdem wurden proteinbiochemische Untersuchungen durchgeführt, die zur Aufklärung der Funktion beitragen sollten (s. Abschnitt 5.3).

5.2.1.3 Untersuchung der ORFs von *labT1* und *labT2*

Die in Abbildung 45 markierten ORFs von *labT1* und *labT2* wurden zunächst mittels BlastX untersucht. Die Schwierigkeit dabei war die sehr große Überlappung der beiden ORFs, die das Ablesen der richtigen Gene und Translationsstartpunkte erschwerte. In Tabelle 28 und Tabelle 29 sind jeweils die vier Hits abgebildet, die den größten Score und den kleinsten E-value gezeigt haben.

Tabelle 28: BlastX-Ergebnisse gegen die ORFs für *labT1*.

Referenz	Beschreibung	Score (bits)	E-value
labT1			
NP_828677.1 (*amf*B: SAV7501)	ABC transporter ATP-binding membrane translocator, AmfB [*S. avermitilis* MA-4680]	362	4e-098
YP_001106224.1 (SACE_4028)	ABC transporter protein, ATP-binding component [*Saccharop. erythraea* NRRL 2338]	347	2e-093
BAA33538.2 (*amf*B)	membrane translocator [*Streptomyces griseus*]	335	6e-090
AAD33775.1 (*ram*A)	putative ATP binding membrane translocator [*Streptomyces lividans*]	323	3e-086

Tabelle 29: BlastX-Ergebnisse gegen die ORFs für *labT2*.

Referenz	Beschreibung	Score (bits)	E-value
labT2			
YP_001106427.1 (*ram*B: SACE_4233)	putative ABC transporter [*Saccharop. erythraea* NRRL 2338]	276	3e-072
YP_001106225.1 (SACE_4029)	putative ABC transporter ATP-binding protein [*Saccharop. erythraea* NRRL 2338]	266	4e-069
NP_828676.1 (*amf*A: SAV7500)	ABC transporter ATP-binding protein, AmfA [*S. avermitilis* MA-4680]	254	2e-065
AAD33776.1 (*ram*B)	putative ATP binding membrane translocator [*S. lividans*]	232	6e-059

Bei der Blast-Analyse des ORFs für *labT2* musste aufgrund der großen Überlappung mit dem ORF für *labT1* die Ausgangssequenz eingeschränkt werden. Die Blast-Ergebnisse zeigten eindeutig, dass es sich sowohl bei LabT1 (566 Aminosäuren) als auch bei LabT2 (571 Aminosäuren) um einen ABC-Transporter handeln musste. Deren Translationsstart wurde mit Hilfe der FramePlot-Analysen bestimmt. Die Ergebnisse der FramePlots passten auch immer sehr gut zu den gefundenen homologen Proteinen aus den Blasts, da diese jeweils sehr ähnliche Startpunkte und Sequenzlängen zeigten.

Untersuchungen zur Transmembran-Domänenstruktur waren nur aufgrund der Betrachtung der homologen Hits aus dem Proteinblast möglich und wurden für die vorliegende Arbeit mit dem TMHMM-Programm (http://www.cbs.dtu.dk/services/TMHMM-2.0/) durchgeführt. Abbildung 51 zeigt exemplarisch die graphische Auswertung der TMHMM-Analyse für AmfB (NP_828677.1) aus *S. avermitilis*, den Transporter mit der höchsten Homologie zu LabT1.[86]

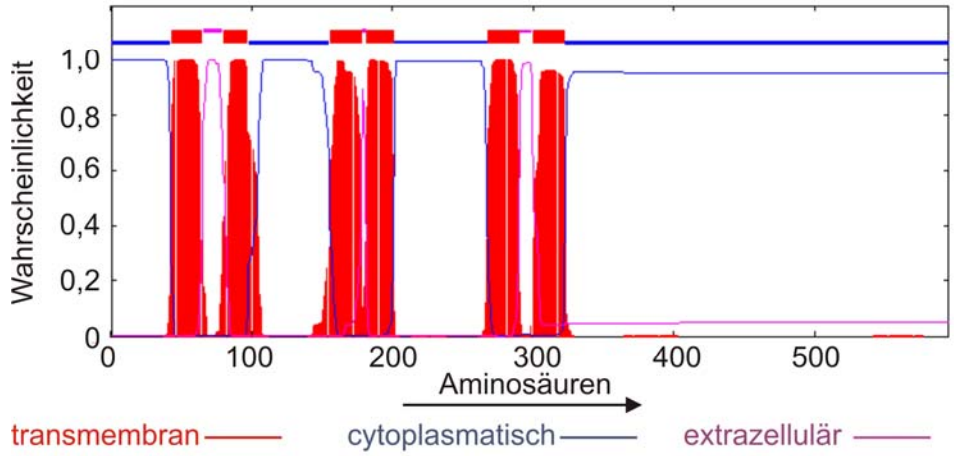


Abbildung 51: Graphische Auswertung der TMHMM-Analyse des ABC-Transporters AmfB (NP_828677.1) aus *S. avermitilis*, der Transporter mit der höchsten Homologie zu LabT1. Aufgetragen ist die Wahrscheinlichkeit der Lokalisation der Aminosäuren (transmembran rot; cytoplasmatisch blau; extrazellulär violett).

In der graphischen Auswertung der TMHMM-Analyse ist die Wahrscheinlichkeit der Lokalisation jeder einzelnen Aminosäure des zu untersuchenden Proteins aufgetragen. Die auf diese Weise erhaltenen Informationen über wahrscheinliche Transmembran-Domänen der homologen Proteine aus Tabelle 28 sind zusammen mit den Walker A-, B- und signature-Motiven in Abbildung 52 dargestellt.

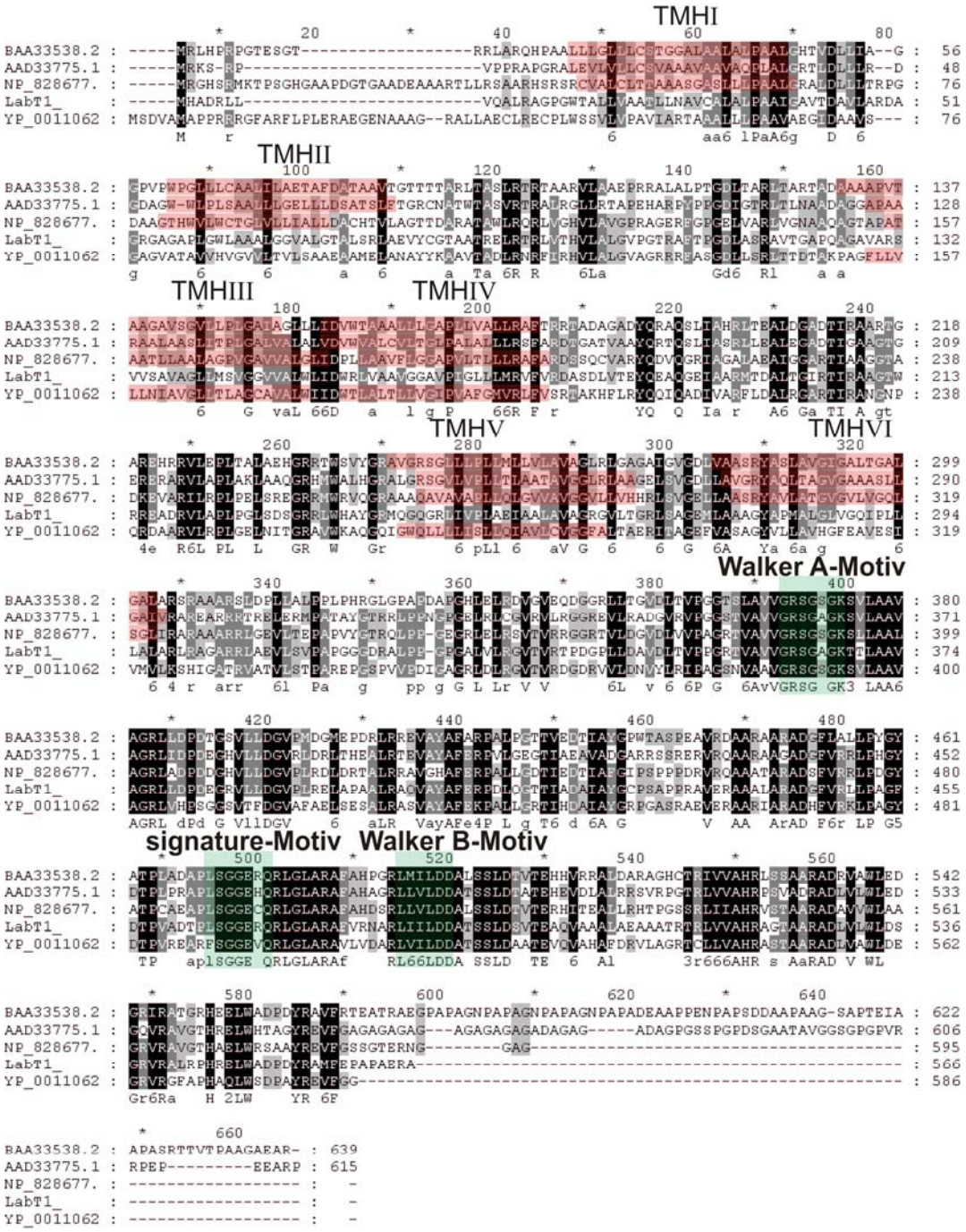


Abbildung 52: Ergebnisse der TMHMM-Analysen der vier zu LabT1 homologen ABC-Transporter in einem Proteinalignment mit LabT1. Die Walker A-, B- und signature-Motive wurden grün markiert.

Abbildung 52 zeigt, dass die zu LabT1 homologen Transporter mit Ausnahme des Transporters YP_001106224 aus *Saccharop. erythraea* in der TMHMM-Analyse die für Lantibiotika-Transporter typische Domänenstruktur mit sechs Transmembran-Domänen im *N*-terminalen Bereich (s. Abschnitt 1.3.3) aufweisen. Der Transporter YP_001106224 aus *Saccharop. erythraea* zeigte in der graphischen Auswertung (nicht abgebildet) noch drei weitere Transmembran-Domänen in den Bereichen von TMHI, II und VI an, wobei die

Ergebnisse hier nicht eindeutig waren und somit diese drei Transmembran-Domänen nicht in das Alignment eingetragen werden konnten.

Ein solcher ABC-Transporter wurde für alle bisher gefundenen Lantibiotika-Biosynthese-Gencluster beschrieben und gehört somit neben dem Enzym für die Einführung der Lanthioninbrücken und den Strukturgenen zum „Basisset“ eines Lantibiotika-Clusters. Diese Ergebnisse ließen vermuten, dass es sich bei LabT1 um diesen zum „Basisset“ gehörigen ABC-Transporter des Labyrinthopeptin-Clusters handelte. Die gleiche Vorgehensweise wurde auch für den zweiten ABC-Transporter angewendet. Hierbei waren jedoch die Ergebnisse der TMHMM-Analysen (s. Abbildung 53) nicht so eindeutig wie die für die LabT1-homologen ABC-Transporter. Jedoch konnte auch hier vermutet werden, dass es sich bei LabT2 ebenfalls um einen für Lantibiotika typischen Transporter handelte.

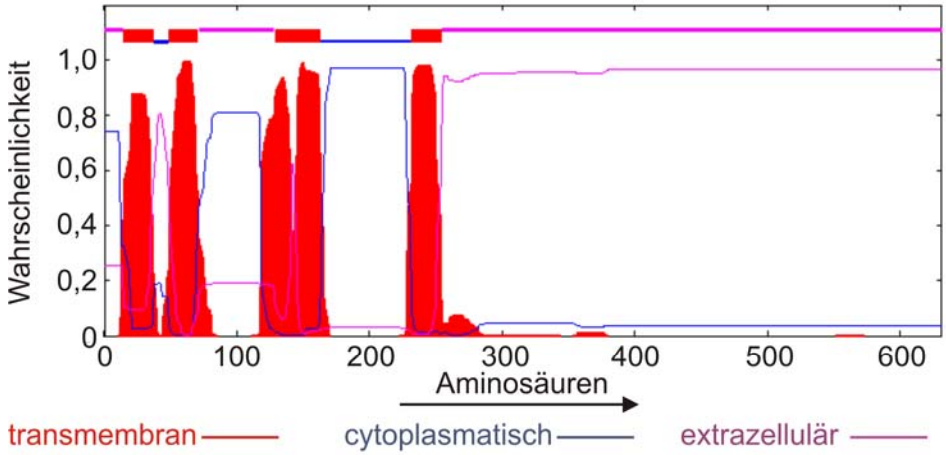


Abbildung 53: Graphische Auswertung der TMHMM-Analyse des AmfA-Transporters aus *S. avermitilis* (NP_828676.1), der Transporter mit der höchsten Homologie zu LabT2. Aufgetragen ist die Wahrscheinlichkeit der Lokalisation der Aminosäuren (transmembran rot; cytoplasmatisch blau; extrazellulär violett).

Abbildung 54 zeigt die mittels TMHMM bestimmten Transmembran-Domänen für die LabT2-homologen ABC-Transporter. Die Ergebnisse sind hier nicht eindeutig, trotzdem lassen sich fünf Bereiche finden, in denen sehr wahrscheinlich Transmembran-Domänen liegen. Alle graphischen Auswertungen zeigten, wie in Abbildung 53 am Beispiel des AmfA-Transporters aus *S. avermitilis* (NP_828676.1) dargestellt, auch noch schwach einen sechsten Bereich an. Ohne diese sechste Transmembran-Domäne würde die *C*-terminale ATP-Bindungsstelle laut TMHMM-Analyse extrazellulär liegen. Da dies sehr unwahrscheinlich ist, wird von der Existenz der sechsten Transmembran-Domäne ausgegangen.

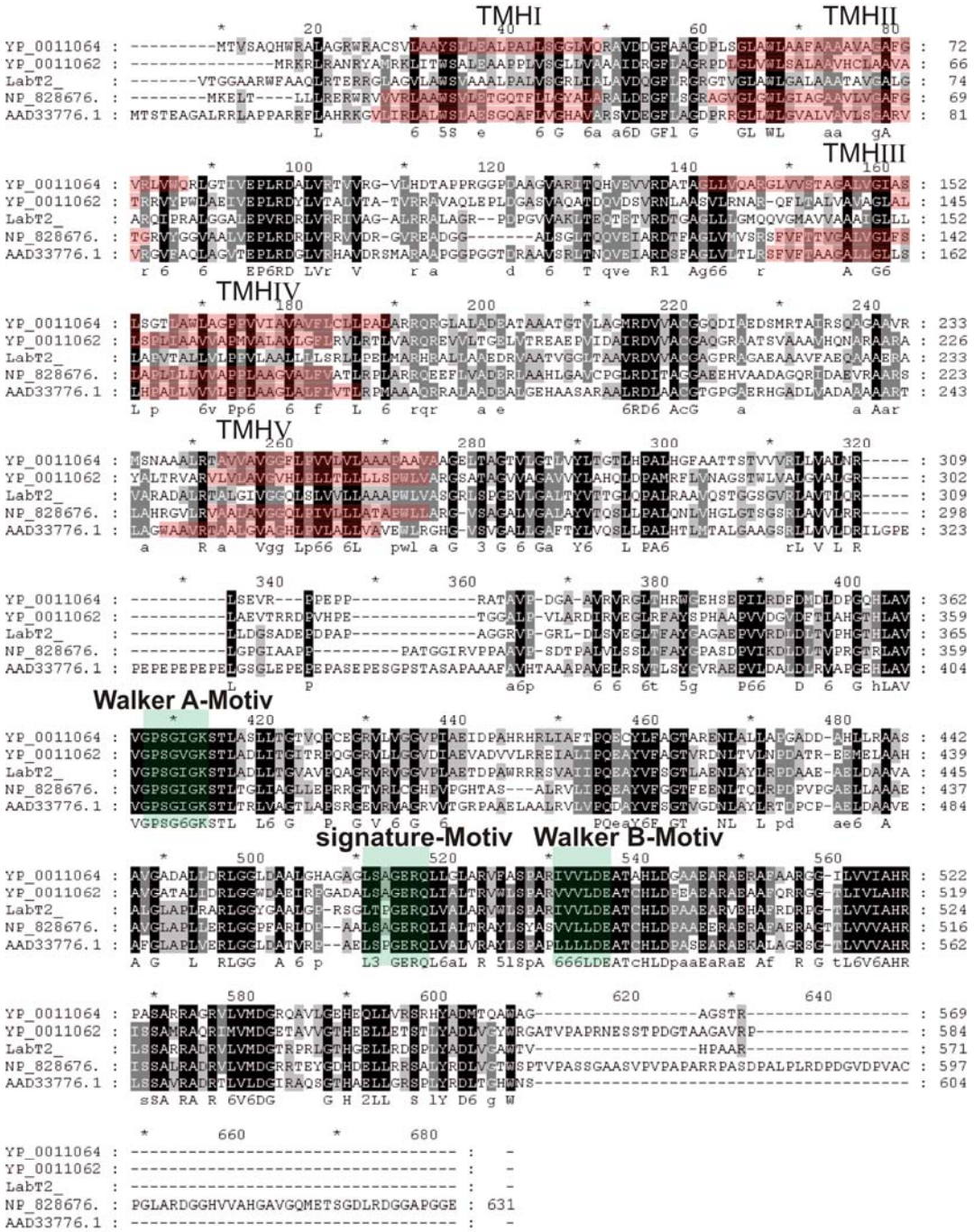


Abbildung 54: Ergebnisse der TMHMM-Analysen der vier zu LabT2 homologen ABC-Transporter in einem Proteinalignment mit LabT2. Die Walker A-, B- und signature-Motive wurden grün markiert.

Zu den Vorhersagen von Transmembran-Domänen basierend auf der Primärsequenz mit dem zurzeit wohl dafür besten Programm (TMHMM)[86] muss angemerkt werden, dass sie nur als Hinweise zu betrachten sind. Man kann mit den Ergebnissen die Existenz und Lage von Transmembranhelices abschätzen, aber keinesfalls sicher und genau vorhersagen. Für die LabT2-Transporter wurden aufgrund der relativ schlechten Ergebnisse noch andere Programme getestet, wobei jedoch kein zusätzlicher Erkenntnisgewinn erzielt werden konnte. Die mittels TMHMM gefundenen Transmembranhelices wurden zwar meistens

bestätigt, variierten von der Lage her aber immer um einige Aminosäuren. Auch Hinweise auf eine sechste Transmembran-Domäne wurden in den anderen Programmen gefunden.
Abschließend kann also zusammengefasst werden, dass zwei ABC-Transporter gefunden wurden, von denen zumindestens LabT1 dem für Lantibiotika typischen Transporter mit sechs *N*-terminalen Transmembranhelices und einer *C*-terminalen NBD-Domäne zu entsprechen scheint.

5.2.1.4 Untersuchung der umliegenden ORFs

Die Untersuchungen weiterer ORFs in direkter Umgebung der fünf bisher zugeordneten Genen und dadurch kodierten Proteinen (LabKC, LabA1, LabA2, LabT1 und LabT2) ergaben keine weiteren eindeutigen Ergebnisse. So konnte dem stromabwärts auf *labT2* folgenden kurzen ORF keinerlei Datenbankeintrag zugeordnet werden. Weiter stromabwärts zeigte ein ORF im BlastX eine hohe Homologie zu einer putativen Hydroxylase aus *Streptomyces ambofaciens*. Stromaufwärts zu *labKC* befand sich eine ca. 2500 bp lange Sequenz, in der es aufgrund uneindeutiger FramePlot-Analysen schwer war, ORFs zu definieren. Auch die Blast-Untersuchungen in diesem Bereich ließen keine klaren Zuordnungen erkennen. Die Zuordnung der restlichen auf der Cosmidsequenz kodierten Proteine erfolgt in Abschnitt 5.2.2.

5.2.1.5 Das Biosynthese-Gencluster der Labyrinthopeptine

Anhand der Untersuchungen der DNA-Sequenzen stromaufwärts und stromabwärts von der bekannten Ausgangssequenz des Strukturgens von LabA2 konnten insgesamt vier neue Gene eindeutig zugeordnet werden. Neben dem Strukturgen von LabA1/A3 wurde stromaufwärts ein Gen identifiziert, das *N*-terminal eine Ser/Thr-Kinase-Domäne und *C*-terminal eine LanC-Domäne kodiert. Stromabwärts wurden zwei Gene für ABC-Transporter identifiziert. Die Betrachtung weiterer ORFs erbrachte keine eindeutige Zuordnung der darin kodierten Proteine, weshalb deren Bezug zum Labyrinthopeptin-Gencluster nicht geklärt werden konnte.
Alle bisher identifizierten Lantibiotika-Gencluster beinhalten neben den Strukturgenen mindestens ein Gen für einen ABC-Transporter sowie Gene für die Enzyme, die für die posttranslationale Einführung der Lanthionin-Reste verantwortlich sind („Basisset“ s. Abschnitt 1.4). Wenn die Biosyntheseprodukte weitere posttranslationale Modifikationen aufweisen, beinhalten die Lantibiotika-Cluster außerdem immer die für diese Modifikationen notwendigen Enzyme. Basierend auf diesen Kenntnissen über die Grundstruktur eines Lantibiotika-Clusters sollen die bisher gefundenen und annotierten Gene des Labyrinthopeptin-Genclusters bzw. deren Produkte im Folgenden diskutiert werden.

Im Labyrinthopeptin-Gencluster wurden zwei Gene für ABC-Transporter gefunden. Normalerweise reicht für den Export der gereiften Struktur ein ABC-Transporter aus, allerdings wurden in einigen Clustern noch weitere Transporter beobachtet, die dem Produzenten Immunität gegenüber seinem Lantibiotikum verleihen. Diese Transporter bestehen aber immer aus mehreren Proteinen (LanFEG). Über die eigentliche Funktion der beiden Transporter LabT1 und LabT2 kann deshalb ohne weitere Versuche also nur spekuliert werden. Einige Lantibiotika-Exporter besitzen auch noch eine Protease-Domäne (s. Abbildung 10), was für LabT1 und LabT2 jedoch nicht beobachtet werden konnte.
Für die Komplettierung des „Basissets“ eines Clusters fehlten beim Labyrinthopeptin-Cluster nun noch die Gene für die modifizierenden Enzyme, die die Einführung der Lanthionin- und Labioninbrücken katalysieren. Für diese Funktion steht nach aktuellem Kenntnisstand nur ein Protein zur Verfügung, das LabKC. Die *C*-terminal liegende LanC-homologe Domäne weist trotz der fehlenden konservierten Reste für eine Zn-Bindung darauf hin, dass dieses Enzym sehr wahrscheinlich an der Zyklisierung zu Lanthioninen und damit an der Reifung des Lantibiotikums beteiligt ist. In den meisten Fällen erfordert die Zyklisierung eine Dehydratisierung, die entweder durch die Dehydratase LanB oder die *N*-terminale Domäne von LanM vorgenommen wird. Im Labyrinthopeptin-Gencluster fehlen allerdings homologe Domänen bzw. Enzyme. Als Vorstufe für die Zyklisierung zu Lanthioninen werden aber auch phosphorylierte Intermediate diskutiert,[87] die im Labyrinthopeptin-Gencluster durch die *N*-terminale Ser/Thr-Kinase-Domäne eingeführt werden könnten. Aufgrund der geringen Homologie von LabKC zu LanM-Enzymen (s. Abbildung 50 zeigt nur das Alignment der jeweils 400 *C*-terminalen Aminosäuren), die zwar auch eine *C*-terminale LanC-Domäne besitzen, aber *N*-terminal stark von der Dehydratase LanB abweichen, kann auch spekuliert werden, dass das LabKC-Enzym für die Einführung des Labionin-Motivs verantwortlich sein könnte. Bei der Einführung des Labionin-Motivs muss sehr wahrscheinlich eine Aktivierung des Ser-Restes für die Zyklisierung zum Labionin erfolgen, was wiederum die *N*-terminale Ser/Thr-Kinase-Domäne übernehmen könnte. Da bei der Zuständigkeit von LabKC für die Labionin-Motive das Enzym für die Einführung der Lanthionine fehlen würde, besteht theoretisch auch die Möglichkeit, dass das LabKC-Protein für die Einführung beider Verbrückungen verantwortlich ist. Der in Abbildung 55 vorgeschlagene Reaktionsmechanismus würde die gleichzeitige Einführung der Lanthionin- und Labionin-Motive, basierend auf den in LabKC vorhandenen Domänen, ermöglichen. Dieser hypothetische Reaktionsmechanismus sieht zunächst eine Aktivierung der Ser-Seitenketten durch Phosphorylierung durch die *N*-terminale Ser/Thr-Kinase-Domäne vor, die dann nach Abspaltung von Phosphat als Dha-Seitenketten vorliegen würden. Die LanC-Domäne von LabKC müsste dann die Cys-Seitenketten durch einen bisher unbekannten Mechanismus zu Thiolaten aktivieren, die die Michael-Systeme der Dha-Seitenketten nukleophil angreifen und

somit die Lanthionine ausbilden können. Dabei würden an den ehemaligen Ser-Seitenketten der Lanthioninbrücke Carbanionen entstehen, die in einem konzertierten Mechanismus wiederum als Nukleophil die Dha-Reste der Labionin-Vorstufen angreifen und unter Bildung der Methylenbrücke den Labioninring schließen könnten.

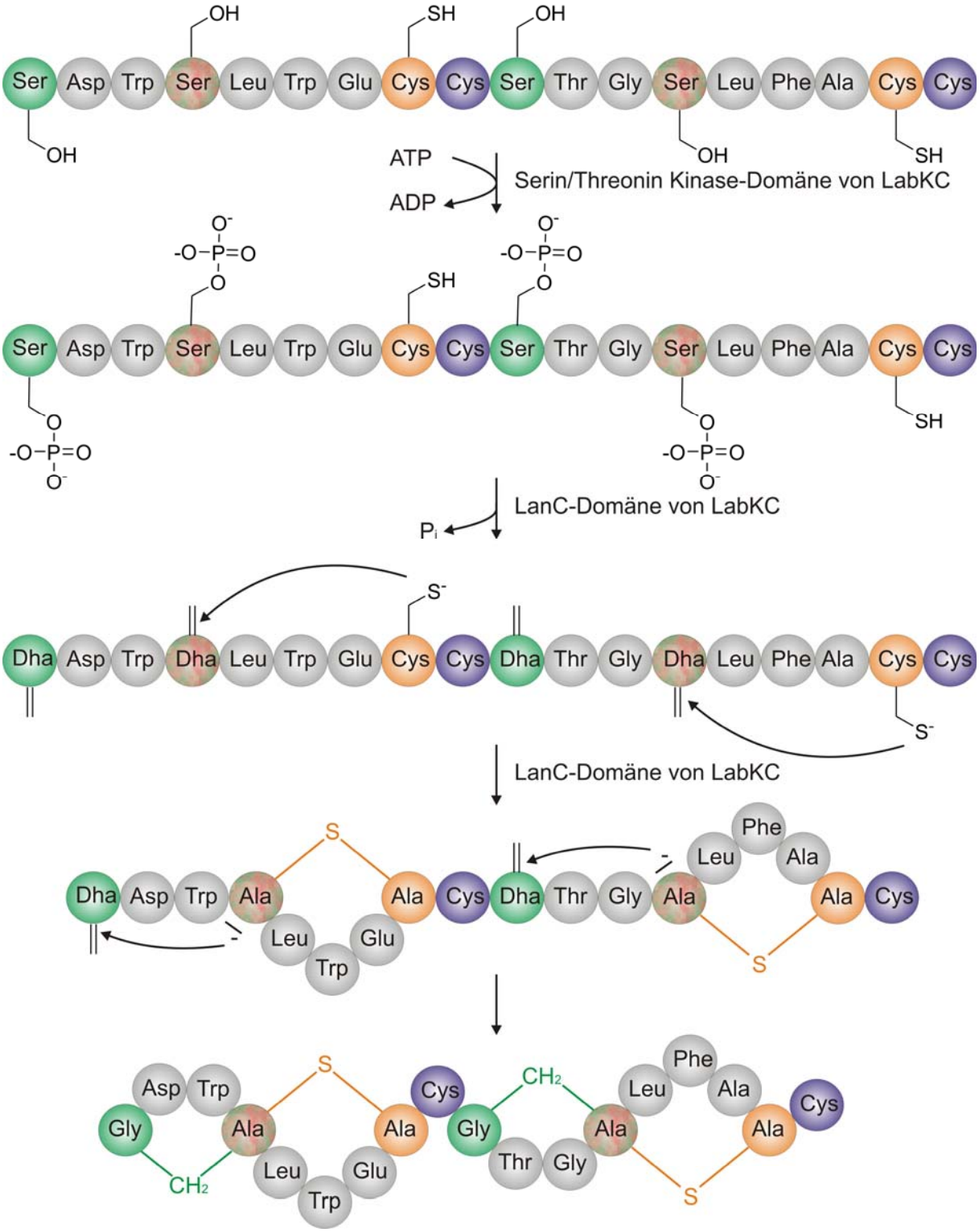


Abbildung 55: Vorschlag zum Reaktionsmechanismus der posttranslationalen Modifikationen des Labyrinthopeptins A2. Im Kugelmodell sind nur die an den Modifikationen beteiligten Aminosäureseitenketten abgebildet.

Diese Überlegungen basieren auf der Annahme, dass die gefundenen fünf Gene auch wirklich das gesamte Cluster repräsentieren. Eine Fortführung des Clusters stromauf- und abwärts ist sehr unwahrscheinlich, denn sonst hätte den angrenzenden ORFs die ein oder andere bekannte und erwartete Funktion zugeordnet werden können müssen.

Zusammfassend kann also gesagt werden, dass unter Annahme der Zuständigkeit von *labKC* für die Synthese der Lanthionin- und Labionin-Motive mit den fünf gefundenen Genen die Biosynthese der Labyrinthopeptine erklärt werden kann. Eine andere Möglichkeit wäre, dass es sich hierbei nur um einen Teil eines verstreut liegenden Clusters handelt, wobei andere Teile an anderen Stellen im Genom liegen. Diese Teile könnten beispielsweise Gene für Regulation, Immunität und für eine Protease zur Abspaltung der Leadersequenz enthalten.

Um diese Möglichkeiten zu überprüfen, müssen Geninaktivierungs-Mutanten des Wildtyps erzeugt oder eine heterologe Expression des Cosmids 1104 (s. Abschnitt 5.4) durchgeführt werden. Eine weitere mögliche Vorgehensweise ist die *in vitro*-Umsetzung des Substrates mit dem LabKC-Protein, um so dessen Aktivität bestimmen zu können (s. Abschnitt 5.3).

An dieser Stelle soll noch einer Beobachtung nachgegangen werden, die im Zuge der Sequenzvergleiche von LabKC, LabT1 und LabT2 gemacht wurde. Es war auffällig, dass die signifikanten Hits der Blasts aus immer den gleichen Stämmen kamen. Bei genauerer Betrachtung der fortlaufenden Nummern der Proteine wurde festgestellt, dass diese benachbart sein müssen. Aufgrund dessen wurden die direkten Umgebungen der vier Hits aus *S. coelicolor* (NP_630756.1), *S. avermitilis* (NP_828679.1), *Saccharop. erythraea* (YP_001106424.1) und *S. griseus* (BAB62264.1) zu LabKC in Sequenzdatenbanken betrachtet (s. Abbildung 56). Der Vergleich der Genanordnungen zeigte, dass sich in diesen vier Stämmen sehr ähnlich wie das Labyrinthopeptin-Cluster aufgebaute Cluster befinden. Die Abfolge und Identität der kodierten Proteine war dabei konserviert. Der einzige Unterschied zwischen den Clustern aus den Streptomyceten-Stämmen im Vergleich zu denen aus *A. namibiensis* und *Saccharopolyspora erythraea* ist das zusätzliche Gen *ramR*, das für einen Transkriptionsregulator kodiert, der für die Regulation der Expression des restlichen Clusters verantwortlich ist.[88] Den *ram*- bzw. *amf*-Clustern der Streptomyceten wird eine wichtige Rolle bei der Bildung eines Luftmycels zugeschrieben (*ram* = *r*apid *a*erial *m*ycelium formation).[41,44] Die meisten Arbeiten auf dem Gebiet der drei Streptomyceten-Cluster wurden bisher am Cluster aus *S. coelicolor* (*ramCSABR*) durchgeführt. Dabei wurde ein Lantibiotikum (SapB) als Produkt des *ramS*-Gens identifiziert. Die Proteine der Gene des restlichen *ram*-Operons sollen für die Biosynthese des SapB verantwortlich sein.[41] Die Kenntnisse zur Struktur von SapB sind zum aktuellen Zeitpunkt jedoch nicht vollständig. Auch die Beteiligung von RamC an der Biosynthese wird bisher nur aufgrund der LanC-Domäne vermutet.[41] Zu Produkten der *amf*-Cluster wurden bisher keine Daten veröffentlicht.

Die Existenz des Clusters in *Saccharopolyspora erythraea* ist nur aufgrund der Genomsequenzierung bekannt, wobei dieser Bereich dort als putatives Lantibiotikum annotiert wurde.[89]

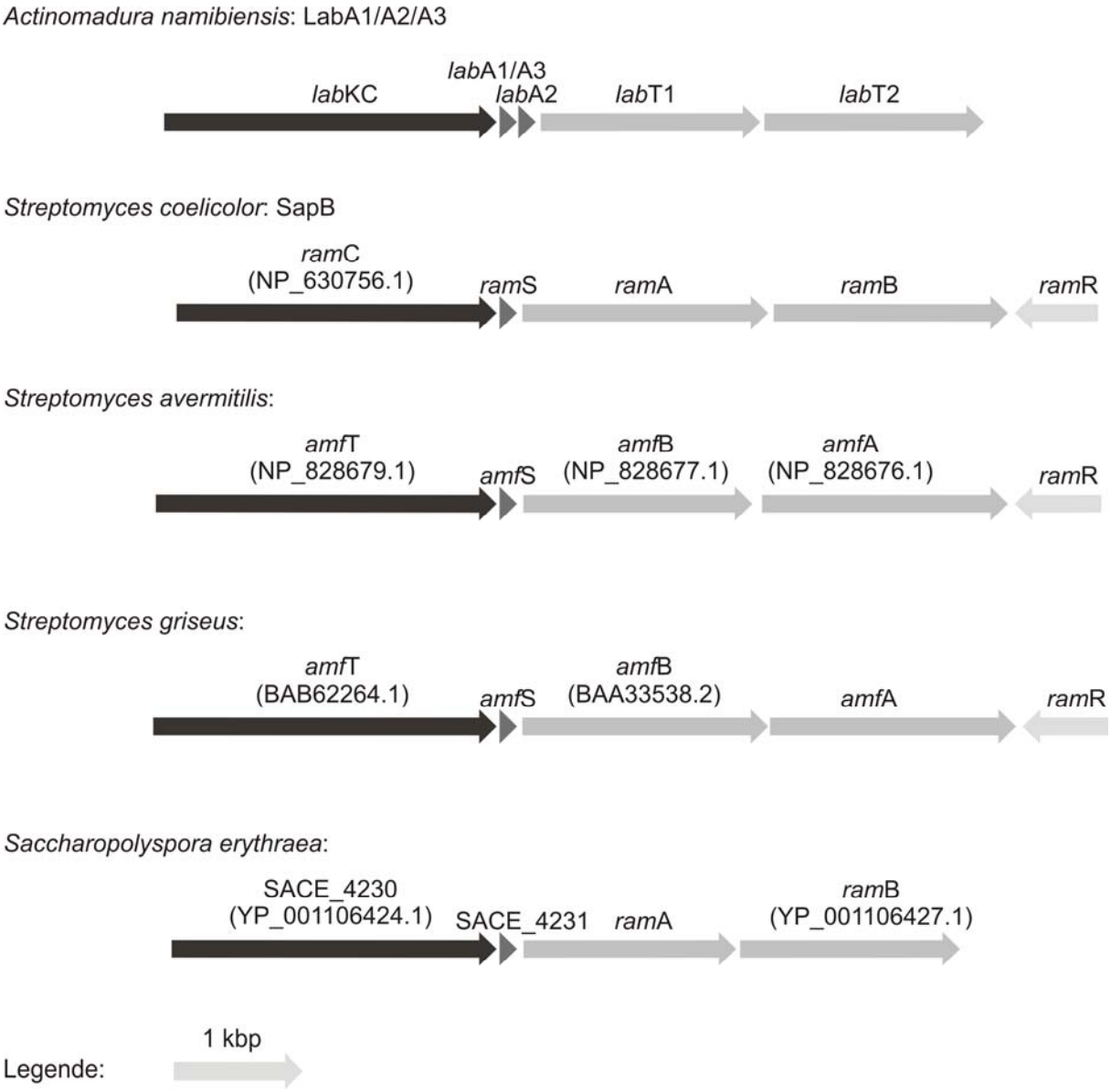


Abbildung 56: Vergleiche der Genombereiche um die zu *labKC*-homologen Gene aus *S. coelicolor*, *S. avermitilis*, *S. griseus* und *Saccharop. erythraea*. Die Transportergene, die ebenfalls unter den vier Treffern der Blasts mit *labT1* und *labT2* waren, wurden mit den entsprechenden Referenzen versehen.

Aufgrund der großen Homologie der entsprechenden Gene bzw. Proteine und der sehr ähnlichen Clusteranordnungen sollen im Folgenden einzelne Aspekte der bisherigen Ergebnisse noch einmal betrachtet werden.

Bisher war die Frage immer noch offen, ob die fünf Gene das gesamte Labyrinthopeptin-Cluster darstellen. Aufgrund der nahezu identischen Clusterstruktur von *Saccharopolyspora erythraea,* in dessen Nachbarschaft Proteine kodiert werden, deren Zugehörigkeit zum

Cluster aufgrund der annotierten Funktion zum jetzigen Zeitpunkt ausgeschlossen werden kann, ist es sehr wahrscheinlich, dass das Labyrinthopeptin-Gencluster auch nur aus den fünf gefundenen Genen besteht.

Ein zu dem Regulatorgen (*ramR*) der Streptomyceten-Cluster homologes Gen kann auch bei erneuter Betrachtung im Labyrinthopeptin-Gencluster nicht identifiziert werden. Das bedeutet, dass die Regulation hier über einen anderen Weg erfolgen muss.

Eine weitere wichtige Frage ist die der Rolle des LabKC-Proteins an der Biosynthese der Labyrinthopeptine. Aufgrund der genetischen Erkenntnisse kann nur spekuliert werden, ob es für die Einführung nur einer Modifikation (Labionin **oder** Lanthionin) verantwortlich ist oder ob doch die vollständige posttranslationale Modifikation (Labionin **und** Lanthionin) katalysiert wird. Für Analogieschlüsse müssten die Strukturen der Produkte der anderen Cluster bekannt sein. Bisher wurde jedoch nur an der Strukturaufklärung von SapB gearbeitet.[41] Trotzdem sollen die Primärsequenzen aller Präpropeptide verglichen werden, um zu untersuchen, ob diese eventuell auch die vorläufig bestimmte Konsensussequenz ($SXXS(X)_{3-5}C$-Motiv) für die Einführung von Labionin-Motiven beinhalten. Dazu wurden die in den Genomannotationen festgelegten Aminosäuren-Sequenzen herangezogen und mit den Präprosequenzen von LabA2 und LabA1/A3 verglichen (s. Abbildung 57).

Der Vergleich der Aminosäuren-Sequenzen zeigte, dass das $SXXS(X)_{3-5}C$-Motiv in jeder Sequenz zweimal zu finden ist. Dies lässt darauf schließen, dass das in LabA2 zum ersten Mal beobachtete Strukturmotiv des Labionins auch sehr wahrscheinlich noch in weiteren Lantibiotika vorkommt. Wenn man die in Abschnitt 1.3.1 gezeigten Propeptidsequenzen aufgeklärter Lantibiotikastrukturen auf das Strukturmotiv hin untersucht, ist es in seiner Gesamtheit dort so nie zu finden. Die außer den Labyrinthopeptinen einzige Verbindung aus Abbildung 57 A), zu der analytische Daten bekannt sind, ist das SapB aus *S. coelicolor*.[41] Da das SapB auch das $SXXS(X)_{3-5}C$-Motiv zeigte, wurde in Abbildung 57 B) eine hypothetische Struktur, basierend auf den Erkenntnissen dieser Arbeit, erstellt. Der Unterschied zur publizierten Struktur nach J. Willey et al. liegt darin, dass die zwei Dha-Reste durch zwei Labionine ersetzt sind. Die bisher publizierten Strukturdaten passen nach genauer Prüfung für beide Modelle, da es sich um isobare Unterschiede handelt. Das einzige Analyseergebnis, das für die Struktur nach Willey und gegen die hier vorgeschlagene Struktur sprechen würde, ist eine Reduktion der Dha-Reste mit einem Massenshift von 4 amu. Dies wurde in der Publikation zwar beschrieben, aber die entsprechenden experimentellen Daten nicht gezeigt.

A)

A. namibiensis: LabA2 MASILELQNLDVEHARGENR SDWSLWECCSTGSLFACC

A. namibiensis: LabA1/A3 MASILELQDLEVERASSAA D SNASVWECCSTGSWVPFTCC

S. coelicolor: SapB MNLFDLGSMETPKEEAMQDVE TGSRASLLLCDGSSLSITTCN

S. avermitilis: MALLDLGTMESDEHTGGGGASTVSLLSCVSAASVLLCL

S. griseus: MALLDLGAMDTPAEDSFGELRTGSQVSLLVCEYSSLSVVLCTP

Saccharop. erythraea: MEMVLELQELDAPNELAYGDPSHGGGSNLSLLASCANSTVSLLTCH

B)

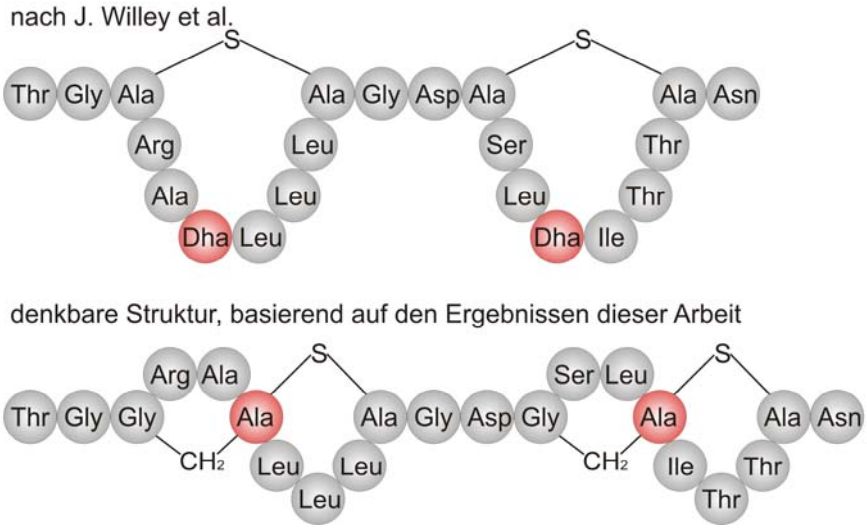


Abbildung 57: A) Darstellung der Aminosäuren-Sequenzen der Strukturgene aus den homologen Clustern. Die Trennung in Leadersequenz (blau) und Propeptid (schwarz/rot) erfolgte nur bei den Sequenzen von *A. namibiensis* und *S. coelicolor* aufgrund experimenteller Daten. Mögliche SXXS(X)$_{3-5}$C-Motiv sind rot hervorgehoben. B) Die bisherige Strukturaufklärung von SapB ist nicht eindeutig. Neben der von J. Willey et al. vorgeschlagene Struktur wäre auch eine Struktur mit Labionin-Motiven denkbar.[41] Der auf den Ergebnissen der vorliegenden Arbeit gründende Strukturvorschlag lässt sich mit den bisher publizierten Strukturdaten ebenfalls in Einklang bringen.

Abschließend kann man zusammenfassen, dass die in *A. namibiensis* gefundene Clusterstruktur so auch in anderen Stämmen zu beobachten ist und bisher für *S. coelicolor*[41] und *S. griseus*[44] beschrieben wurde. Auch die Betrachtung der in den homologen Clustern kodierten Präpropeptide lässt den Schluss zu, dass das Strukturmotiv der Labionin-Lanthionin-Verbrückung eine bisher unbekannte aber doch relativ weitläufig verbreitete posttranslationale Modifikation sein könnte. Da die Cluster auch nur die zu LabKC homologen Proteine besitzen, sind diese sehr wahrscheinlich für die Einführung der Lanthionin- und Labionin-Brücken verantwortlich.

5.2.2 Annotation der restlichen Gene von Cosmid 1104

Neben dem ca. 6,4 kbp umfassenden Bereich des Labyrinthopeptin-Genclusters wurden durch die Sequenzierung des Cosmids 1104 weitere ca. 30 kbp umfassende Sequenzdaten erhalten (s. Abbildung 58). Auch diese wurden annotiert und ergaben die in Tabelle 30 zusammengestellten Ergebnisse. Zur Annotation wurden die kodierenden ORFs mittels FramePlot bestimmt und dann einer Blastanalyse unterzogen.

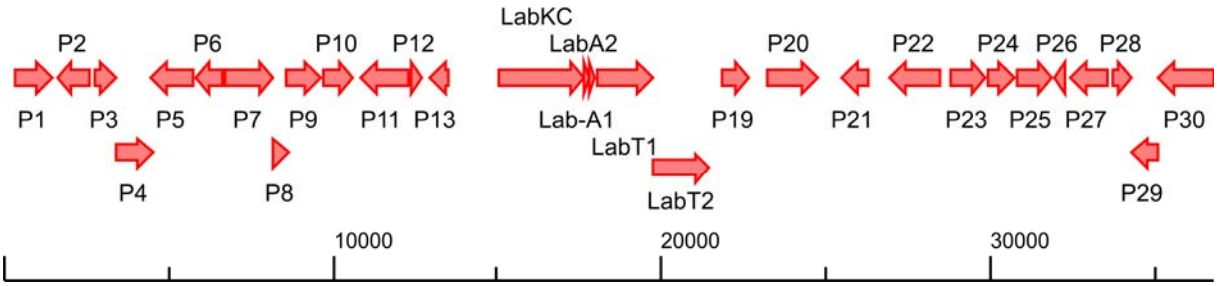


Abbildung 58: Gezeigt ist der ca. 36 kbp große sequenzierte Bereich mit den darauf kodierten Proteinen. (P = ORF in Tabelle 30)

Tabelle 30: Zusammenstellung der kodierten Proteine, die in den ORFs der 30 kbp gefunden wurden und nicht zum Cluster gehören. Es wurde jeweils der signifikanteste Hit des BlastP notiert und die Länge des Proteins angegeben.

ORF	Referenz	Beschreibung	Score (bits)	E-value	Protein (AS)
ORF1	ZP_01437506.1	Voltage-gated ion channel superfamily protein [*Fulvimarina pelagi* HTCC2506]	71.6	9e-011	388
ORF2	NP_927370.1	phosphoribulokinase [*Gloeobacter violaceus* PCC 7421]	283	1e-074	322
ORF3	YP_638226.1	two component transcriptional regulator, LuxR family [*Mycobacterium* sp. MCS]	218	3e-055	218
ORF4	YP_952197.1	GAF sensor signal transduction histidine kinase [*Mycobacterium vanbaalenii* PYR-1]	152	3e-035	379
ORF5	YP_001538428.1	Saccharopine dehydrogenase (NAD(+), *L*-glutamate-forming) [*Salinispora arenicola* CNS-205]	408	e-112	430

ORF6	BAA95688.1	LysR type transcriptional regulator CbbR [*Hydrogenophilus thermoluteolus*]	92.8	2e-017	295
ORF7	ABV69587.1	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit-1 [*Mycobacterium* sp. DSM 3803]	791	0.0	497
ORF8	ABV69590.1	ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit-2 [*Mycobacterium* sp. DSM 3803]	224	2e-057	139
ORF9	ABM17102.1	protein CbbX [*Sulfobacillus acidophilus*]	347	9e-094	346
ORF10	YP_288987.1	hypothetical protein Tfu_0926 [*Thermobifida fusca* YX]	327	6e-088	308
ORF11	YP_001109088.1	transcriptional regulator/aminotransferase [*Saccharopolyspora erythraea* NRRL 2338]	223	3e-056	480
ORF12	YP_756654.1	Cupin 2, conserved barrel domain protein [*Maricaulis maris* MCS10]	82.4	8e-015	117
ORF13		no hits found			182
ORF19	CAJ89591.1	putative hydroxylase [*Streptomyces ambofaciens* ATCC 23877]	362	1e-098	278
ORF20	NP_825262.1	hypothetical protein SAV_4085 [*Streptomyces avermitilis* MA-4680]	88.6	1e-015	503
ORF21		no hits found			264
ORF22	EDR11245.1	predicted protein [*Laccaria bicolor* S238N-H82]	53.5	4e-005	519

ORF23	CAJ88212.1	putative transcriptional regulator [*Streptomyces ambofaciens* ATCC 23877]	139	4e-031	343
ORF24	YP_482168.1	aminotransferase, class IV [*Frankia* sp. CcI3]	242	1e-062	254
ORF25	P19432	Glutamine synthetase 2 (Glutamine synthetase II) (Glutamate-ammonia ligase II) (GSII) [*Streptomyces viridochromogenes*]	542	e-152	372
ORF26	YP_001103878.1	hypothetical protein SACE_1634 [*Saccharopolyspora erythraea* NRRL 2338]	80.9	2e-014	104
ORF27	YP_704719.1	hypothetical protein RHA1_ro04776 [*Rhodococcus* sp. RHA1]	152	5e-035	379
ORF28		no hits found			196
ORF29	CAL53966.1	unnamed protein product [*Ostreococcus tauri*]	38.1	0.53	256
ORF30	YP_001104263.1	GTP pyrophosphokinase [*Saccharopolyspora erythraea* NRRL 2338]	399	e-109	560

Die Zuordnungen der in Tabelle 30 zusammengestellten Ergebnisse beruhen teilweise auf Hits mit sehr niedriger Signifikanz. Nur am Anfang des Cosmids 1104 wurde ein Sequenzabschnitt mit eindeutiger Zuordnung gefunden, welcher einen Teil des Calvin-Zyklus-Clusters enthält.

5.3 Untersuchungen zur Biosynthese der Labyrinthopeptine

Die manuelle und Blast-gestützte Auswertung der Sequenz von Cosmid 1104 (s. Abschnitt 5.2) hatte einen Bereich aus fünf Genen gezeigt, der dem Labyrinthopeptin-Gencluster zugeordnet werden konnte. Neben den beiden Strukturgenen, die für LabA1, A2 und A3 kodieren, wurden die zwei Gene *labT1* und *labT2* für ABC-Transporter gefunden. Dem fünften Gen *labKC* wurde aufgrund seiner Domänenstruktur die Einführung der Lanthionin- und Labioninbrücken zugeordnet. Um diese Annahme experimentell zu festigen, sollte die katalytische Aktivität des LabKC-Proteins untersucht werden. Eine klassische Herangehensweise in einem solchen Fall ist die Erzeugung einer Geninaktivierungs-Mutante, bei der das entsprechende Gen *in-frame* deletiert wird. Der Vergleich des Metabolitspektrums dieser Mutante mit dem des Wildtyps gibt Aufschluss über die Funktion des inaktivierten Gens bzw. des entsprechenden Enzyms. Leider ist es bis zum jetzigen Zeitpunkt nicht gelungen, DNA in den Stamm *Actinomadura namibiensis* zu transferieren, was essentiel ist für die Erzeugung einer Geninaktivierungs-Mutante. Deshalb wurde in der vorliegenden Arbeit die Untersuchung der katalytischen Aktivität des LabKC-Proteins auf alternativem Weg versucht. In einem *in vitro*-Assay sollte gezeigt werden, dass LabKC für die posttranslationalen Modifikationen von LabA1, A2 und A3 verantwortlich ist. Dazu sollte das heterolog exprimierte LabKC-Protein mit seinem hypothetischen Substrat, dem Präpropeptid von LabA2 sowie den vermuteten Cofaktoren inkubiert und anschließend das gebildete Produkt mittels HPLC-ESI-MS und -MSMS untersucht werden. Die dafür notwendigen Voraussetzungen waren die Verfügbarkeit von Enzym und Substrat, auf deren Generierung in den nächsten Abschnitten eingegangen werden soll.

5.3.1 Generierung der Substrate für den *in vitro*-Assay

Untersuchungen zur Funktion der Leadersequenzen von Lantibiotika haben gezeigt, dass diese unter anderem für die Substraterkennung der posttranslational modifizierenden Enzyme verantwortlich sind.[90] Diese Tatsache ist für die Planung des *in vitro*-Assays von großer Wichtigkeit, da die einzusetzenden Substrate die Leadersequenz tragen müssen, um von dem hypothetischen modifizierenden Enzym LabKC erkannt zu werden. Deshalb sollte das in Abbildung 48 A) gezeigte, 38 Aminosäuren lange Präpropeptid von Labyrinthopeptin A2 als Substrat verwendet werden. Ein Peptid dieser Größe ist normalerweise ohne große Probleme durch chemische Synthese an der festen Phase darzustellen. Für die Synthese des 38-er Peptids wurde die Firma EMC-Microcollections, Tübingen, beauftragt, die zunächst verschiedene Testansätze durchführte. Da allerdings aufgrund der schwierig zu synthetisierenden Sequenz kein Syntheseansatz zum gewünschten Ergebnis führte, wurde

der Auftrag wegen Undurchführbarkeit abgelehnt. Aufgrund dessen musste ein alternativer, biologischer Weg gefunden werden, um das Substrat für den Assay zu gewinnen. Um von diesem kurzen 38-er Peptid eine möglichst große Menge produzieren und aufreinigen zu können, wurde eine heterologe Expression in *E. coli* als Peptid mit His-Tag in Betracht gezogen. Als Problem wurde jedoch die damit verbundene Verlängerung des Substrates angesehen. Bei einem *N*-terminalen His-Tag werden normalerweise mindestens sechs His-Reste an die Leadersequenz angehängt, was bei einer Gesamtlänge von 20 Aminosäuren einer Verlängerung um ca. 30 % entspricht. Deshalb wurde nach einem Expressionssystem gesucht, bei dem der His-Tag nach der Reinigung wieder entfernt werden kann. Schließlich wurde das Champion pET SUMO-Expressionssystem der Firma Invitrogen, Karlsruhe, ausgewählt, welches in Abschnitt 4.4.8 näher beschrieben wurde. Dieses System ermöglicht es, das Produkt nach Abspaltung des *N*-terminalen His-Tags ohne die Verlängerung um auch nur eine Aminosäure zu generieren. Dabei kommt die SUMO-Protease zum Einsatz, die ihr Substrat nicht anhand einer sequenzspezifischen Aminosäuresequenz erkennt, sondern aufgrund der Tertiärstruktur der zwischen Tag und Präpropeptid insertierten SUMO-Domäne.

Für die heterologe Expression von LabA2 musste zunächst der gewünschte Genabschnitt aus dem *lab*A2-ORF mittels PCR amplifiziert und in den linearisierten pET SUMO-Vektor kloniert werden. Bei der Ableitung der Primer wurde so vorgegangen, dass der Vorwärtsprimer mit dem Startcodon beginnt und der Rückwärtsprimer das entsprechende Stopcodon enthält (s. Abbildung 59).

Abbildung 59: Ableitung der Primer für die heterologe Expression des LabA2-Präpropeptids.

Aufgrund der Größe des PCR-Produkts von nur 117 bp wurde hier mit einer Taq-Polymerase gearbeitet. Das Produkt wurde aus dem Gel gereinigt (s. Abschnitt 4.2.4.4), mittels TA-Klonierung in den pET SUMO-Vektor kloniert und in kompetente *E. coli* Mach1^{TM}-T1^{R} Zellen transformiert. Die Auswertung auf erfolgreiche Insertion in richtiger Orientierung erfolgte aufgrund der Kürze der Insertsequenz mittels Sequenzierung. Auf diese Weise konnte ein Expressionsplasmid für LabA2 erzeugt werden, das für die heterologe Expression in *E. coli* BL21 (DE3) transformiert wurde.

Neben dem Nachweis der Beteiligung des LabKC-Proteins an der Reifung der Labyrinthopeptine sollten mit Hilfe des *in vitro*-Assays auch mechanistische Untersuchungen zum Ablauf der posttranslationalen Modifikationen durchgeführt werden. Der in Abbildung 55 vorgeschlagene Mechanismus geht zunächst von phosphorylierten Ser-Intermediaten aus, die zu Dha-Resten dephosphoryliert werden. Nach der Aktivierung der Cys-Reste als Thiolat-Anionen erfolgt dann ein nukleophiler Angriff auf die dazu *N*-terminal liegenden Dha-Reste unter Bildung der Lanthioninbrücken. Wenn man von einem konzertierten Mechanismus ausgeht, könnten die dabei entstehenden Carbanionen durch einen erneuten nukleophilen Angriff auf die noch weiter *N*-terminal liegenden Dha-Reste zur Ausbildung der Labioninverbrückungen führen. Um diesen Vorgang genauer zu untersuchen, sollte der möglicherweise konzertierte Ablauf mit Hilfe eines mutierten Substrates unterbrochen werden. Ser->Ala-Mutationen an den zentralen Ser-Resten der Labionin-Lanthionin-Verbrückungen sollten es ermöglichen, Intermediate der Biosynthese abzufangen (s. Abbildung 60).

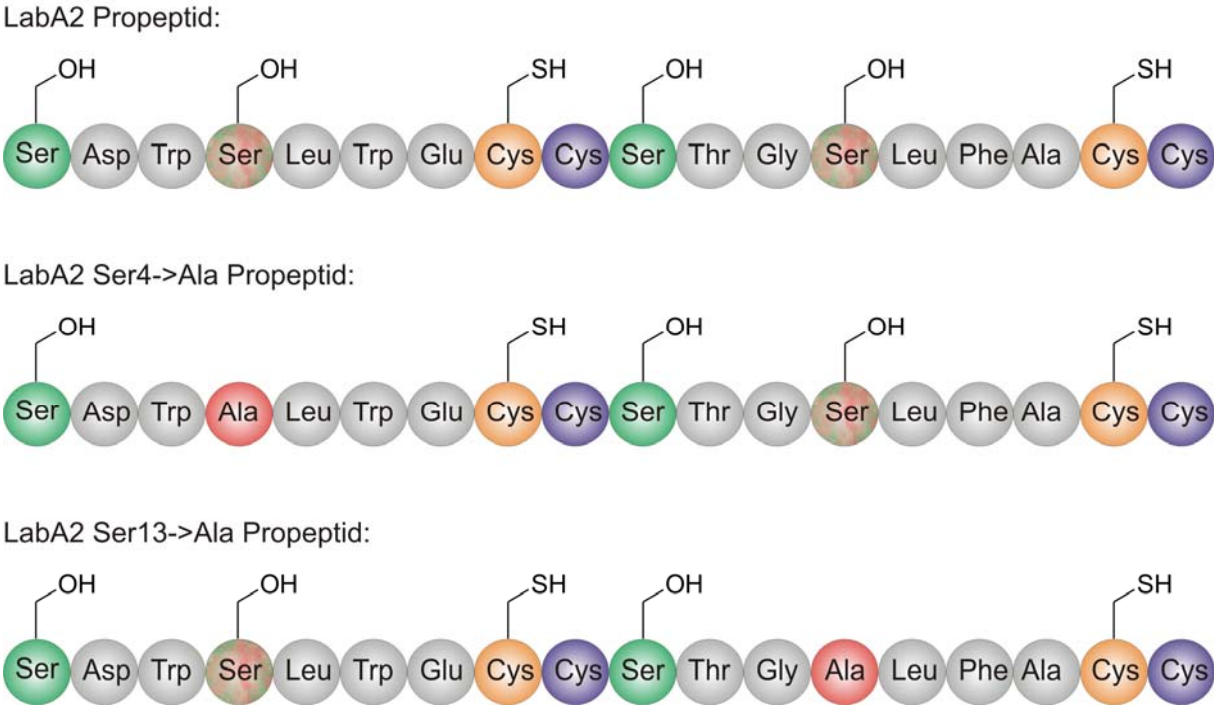


Abbildung 60: Eingeführte Mutationen im Präpropeptid für mechanistische Untersuchungen. Hier sind allerdings nur die entsprechenden Propeptide dargestellt.

Durch die Einführung von Ala an der Stelle der zentralen Ser-Reste (Position 4 bzw. 13) dürfte eine Ausbildung der Lanthioninbrücken an diesen Stellen nicht mehr möglich sein. Somit sollte bei einem konzertierten Mechanismus auch die Schließung der Labioninbrücke unterbunden sein, was dann zu Dha-Resten an den Positionen 1 (LabA2 Ser4->Ala) oder 10 (LabA2 Ser13->Ala) führen sollte. Verläuft die Labioninbildung allerdings nicht konzertiert,

sondern durch eine unabhängige Aktivierung des αC-Kohlenstoffes an den Ser-/Ala-Resten 4 oder 13, müsste stattdessen das teilweise zyklisierte Produkt nachzuweisen sein. Für erste mechanistische Untersuchungen zur Biosynthese sollten die beiden in Abbildung 60 gezeigten mutierten Substrate hergestellt werden, um dann nach ersten Hinweisen gezielt weitere mutierte Substrate generieren zu können.

Die Herstellung der mutierten Präpropeptidsubstrate erfolgte durch eine ortsgerichtete Mutagenese nach dem in Abschnitt 4.2.3.5 beschriebenen Protokoll, wobei die in Abbildung 61 dargestellten Primer für die Einführung der Mutationen herangezogen wurden.

Abbildung 61: Ausgehend von der Aminosäuresequenz von LabA2 (Leadersequenz blau, Strukturteil rot) sollten Primer abgeleitet werden, die an den Ser-Resten (grün) eine Ala-Mutation einführen. Da die beiden Aminosäuren zu weit auseinander lagen, um auf einen Schritt mutagenisiert zu werden, wurden zwei Primerpaare abgeleitet. Die abgeleiteten Primerpaare erlaubten für den Bedarfsfall die Einführung einer Doppelmutation durch eine zweite Mutageneserunde mit dem jeweils anderen Primerpaar.

Die Überprüfung der erhaltenen Expressionsplasmide konnte durch einen Verdau mit BanII erfolgen, da durch die Mutationen jeweils eine zusätzliche palindrome Sequenz (GRGCYC) eingeführt wurde. Diese ermöglichte die eindeutige Unterscheidung zwischen mutierter und nicht-mutierter Sequenz. Für beide Mutationen wurde je ein im Verdau positiv aufgefallenes Expressionsplasmid sequenziert, um Polymerasefehler bei der PCR ausschließen zu können. Die erhaltenen pET SUMO-Konstrukte (LabA2 Ser4->Ala und LabA2 Ser13->Ala) wurden für die anschließende Expression ebenfalls in *E. coli* BL21 (DE3) transformiert.

Die Expression der Substrat-His-Tag-Konstrukte erfolgte nach dem in Abschnitt 4.4.8 optimierten Protokoll. Dabei hatten sich die Zugabe von Glukose (1 %) zum Medium und eine Expression bei 37°C (4 h) als optimal herausgestellt. Mit den so erzeugten Pellets erfolgte eine Proteinreinigung über eine Ni-NTA-Agarose-Säule (s. Abschnitt 4.4.9). Die dabei erhaltenen Elutionsfraktionen wurden aufgrund der erwarteten Größe der Substrat-His-Tag-Konstrukte von nur 17,7 kDa mittels SDS-PAGE nach Schägger und Jagow (s. Abschnitt 4.4.4) analysiert. Für die weitere geplante Vorgehensweise zur Generierung der freien Präpropeptide mussten nun die entsprechenden Elutionsfraktionen vereinigt und aus den in Abschnitt 4.4.10 diskutierten Gründen dialysiert werden. Ausgehend von diesen Dialysaten wurden dann die Präpropeptide durch proteolytische Spaltung aus den Substrat-His-Tag-Konstrukten erhalten und anschließend über eine zweite Ni-Säule gereinigt (s. Abschnitt 4.4.10). Abbildung 62 zeigt den Verlauf der SUMO-Spaltung des nicht-mutierten Lab-A2-His-Tag-Konstruktes mit der Zeit.

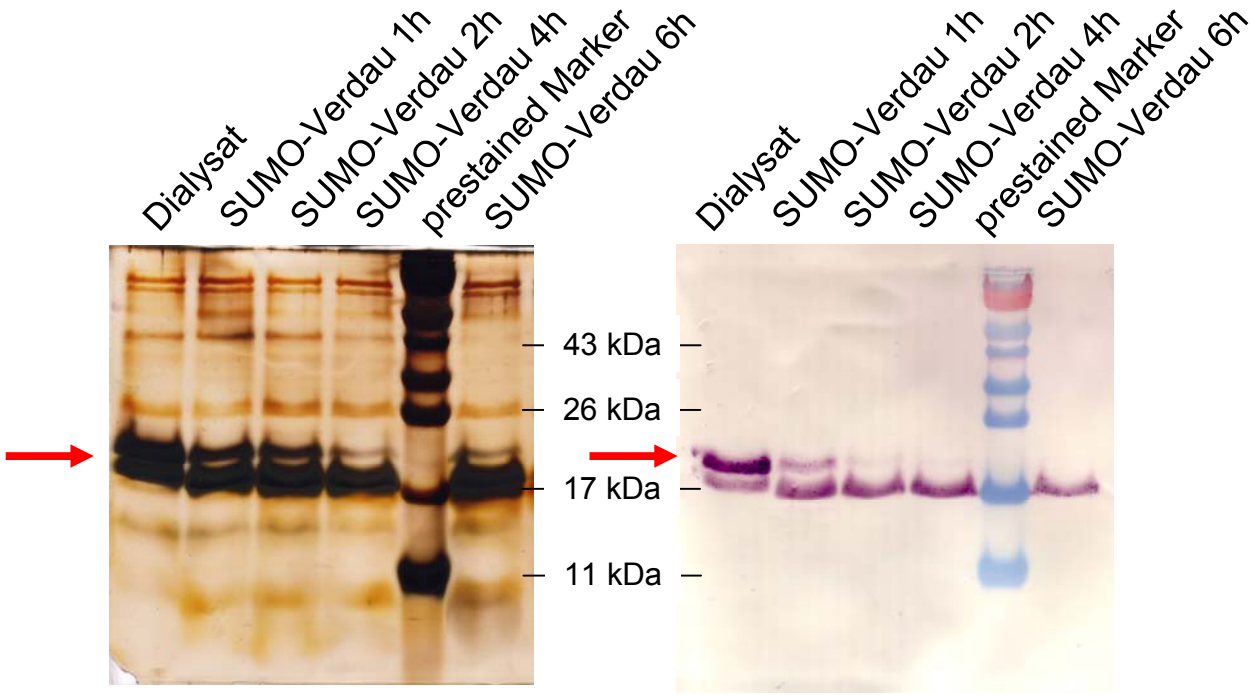


Abbildung 62: Zeitlicher Verlauf der SUMO-Spaltung des nicht-mutierten Lab-A2-His-Tag-Konstruktes im silbergefärbten Gel und im *Western-Blot* (16,5 % Trenngel und 5 % Sammelgel).

In den Dialysaten waren zwei Hauptbanden zu sehen, die im erwarteten Größenbereich des nicht-mutierten Lab-A2-His-Tag-Konstruktes um 17,7 kDa lagen. Man konnte beobachten, wie mit der Zeit die Intensität der oberen Bande deutlich abnahm, wobei die untere gleich intensiv blieb. Das deutete darauf hin, dass die untere der beiden Banden nur coaufgereinigt wurde. Eine Darstellung des gebildeten freien Präpropeptides von LabA2 war aufgrund deren geringer Größe von 4,3 kDa nicht gelelektrophoretisch zu verwirklichen.

Die Auswertung der Gele lief für die beiden mutierten Varianten analog. Diese Ergebnisse sowie die Bestätigung der richtigen DNA-Sequenzen der Expressionsplasmide mittels Sequenzierung ließen darauf schließen, dass die Generierung der Präpropeptide auf biologischem Wege funktioniert hat. Es sollten jedoch die Präpropeptide auch noch direkt nachgewiesen werden, was bei einer Länge von 38 Aminosäuren und einer Größe von 4,3 kDa wie bereits erwähnt gelelektrophoretisch nicht zu realisieren war und deshalb mittels HPLC-ESI-MS und -MSMS erfolgen sollte (s. Abschnitt 4.4.12). Für die massenspektrometrische Charakterisierung der Präpropeptide wurden diese zuvor noch durch einen proteolytischen GluC-Verdau in detektierbare Fragmente gespalten (s. Abschnitt 4.4.11). Durch Analyse der Verdauansätze mittels MS und MSMS konnten nahezu alle erwarteten Peptidfragmente gefunden und durch Sequenzierung bestätigt werden (s. Abbildung 63). Es gelang jedoch nicht, das *C*-terminale Fragment (CCSTGSLFACC) zu detektieren, wofür Oxidationen oder eine Polymerisation an den Cys-Seitenketten verantwortlich sein könnten.

LabA2-Präpropeptid-His-Tag-Konstrukt:

MGSSHHHHHHGSGLVPRGSASMSDSE VNQE AKPE VKPE VKPE THINLKVSDGSSE IFFKIKKTTPLRRLME AFAKRQGKE MDSLRFLYDGIRIQADQTPE DLDME DNDIIE AHRE QIGGMASILE LQNLDVE HARGE NRSDWSLWE CCSTGSLFACC

LabA2:

MASILE LQNLDVE HARGE NRSDWSLWE CCSTGSLFACC

LabA2 Ser2->Ala:

MASILE LQNLDVE HARGE NRSDWALWE CCSTGSLFACC

LabA2 Ser4->Ala:

MASILE LQNLDVE HARGE NRSDWSLWE CCSTGALFACC

Abbildung 63: Übersicht über die Fragmente aus den Verdauansätzen der Präpropeptide und Präpropeptid-His-Tag-Konstrukte. Alle grau unterlegten Fragmente konnten mittels HPLC-ESI-MS und -MSMS nachgewisen werden. Die Leadersequenzen sind blau, die Propeptide rot gekennzeichnet. Die grün hervorgehobenen Aminosäuren entsprechen den mutierten Stellen. Leider war es in keinem der Fälle möglich, das *C*-terminale Fragment (CCSTGSLFACC) zu detektieren.

Abbildung 64 zeigt an zwei Beispielen exemplarisch die Auswertung der MSMS-Spektren, wobei QIGGMASILE die SUMO-Spaltstelle enthält und MASILE dann den *N*-Terminus des Präpropeptids nach der Spaltung repräsentiert. Dieses Beispiel zeigte auch, dass die Abspaltung mit der SUMO-Protease erfolgreich war und dass wie erwartet keine zusätzlichen Aminosäuren am *N*-Terminus angehängt waren.

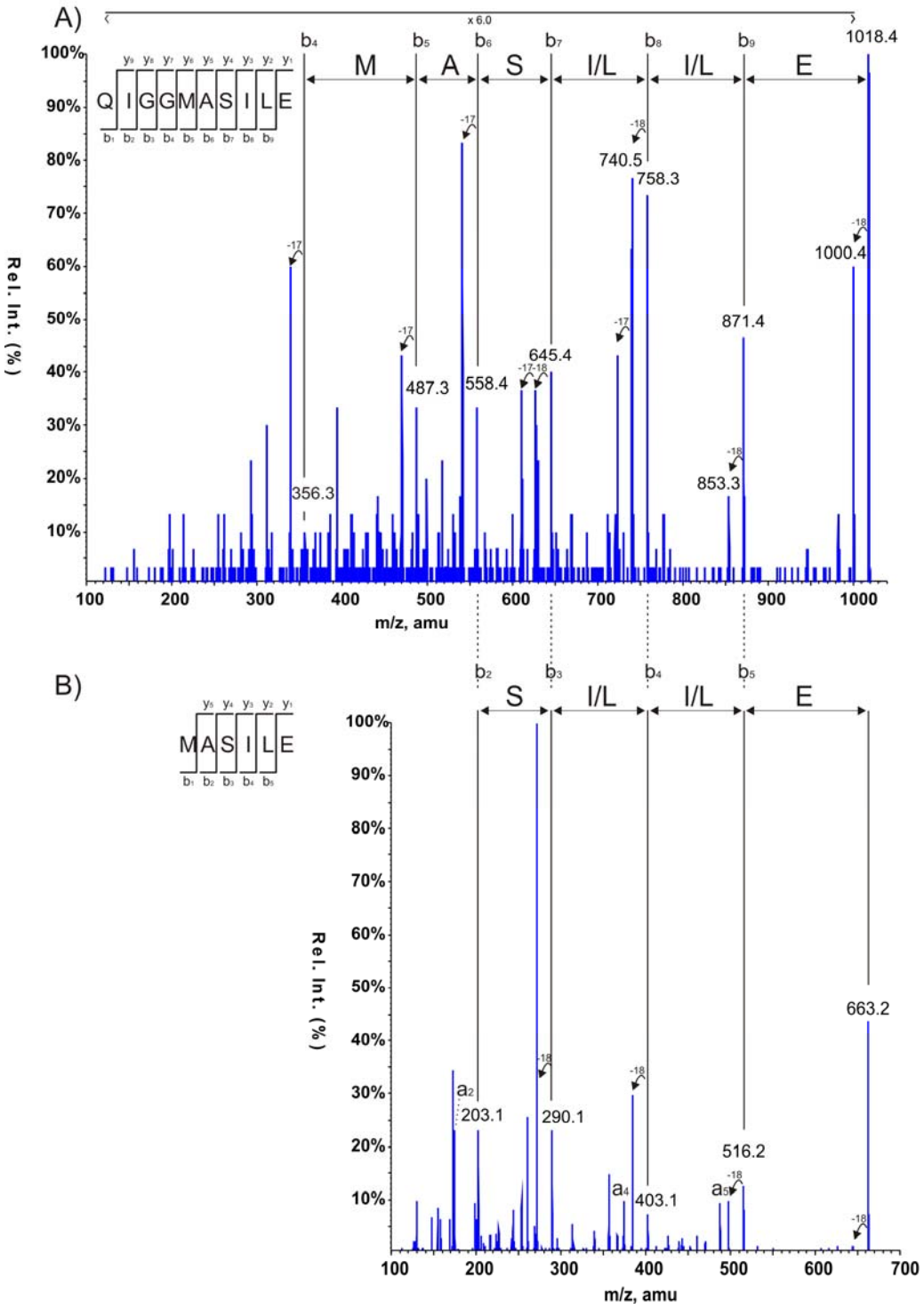


Abbildung 64: MSMS-Spektren zweier Peptide des LabA2-Präpropeptid-His-Tag-Konstrukts und des LabA2-Präpropeptids. Die Sequenzierung erfolgte anhand der zu beobachtenden b-Serie. A) zeigt das MSMS-Spektrum des Peptids QIGGMASILE über die SUMO-Spaltstelle des LabA2-His-Tag-Konstrukts. B) zeigt das MSMS-Spektrum des analogen Peptids MASILE des freien LabA2-Präpropeptids nach der SUMO-Spaltung.

Grundsätzlich muss angemerkt werden, dass die Konzentrationen auf Stufe der Substrat-His-Tag-Konstrukte ausreichend waren, um eine eindeutige massenspektrometrische Identifizierung der Peptide vornehmen zu können. Bei den freien Präpropeptiden waren die Konzentrationen allerdings deutlich geringer, was eine Folge der zusätzlich durchgeführten

Schritte war. Die Originalvorschrift von Invitrogen für die SUMO-Spaltung setzt für einen Ansatz lediglich 20µg Protein ein und erhält nach einer Inkubation (6h; 30°C) das gewünschte Produkt in einer Ausbeute von 95 %. Für die SUMO-Ansätze in dieser Arbeit wurden die Substratkonstrukte in einem großen Überschuß (100 µl einer 1 µg/µl Lösung) eingesetzt. Nach der anschließenden Aufreinigung über Ni-NTA-Spin-Säule konnten jedoch nur etwa 50 % wiedergefunden werden, was auf Verluste bei der Aufreinigung bzw. auf eine unvollständige SUMO-Spaltung zurückzuführen war. Wenn man jetzt noch berücksichtigte, dass es durch den für die Detektion unumgänglichen GluC-Verdau außerdem zu einer Verdünnung um den Faktor zwei kam, war der große Konzentrationsverlust von den Substrat-His-Tag-Konstrukte zu den Präpropeptiden nachvollziehbar, aber leider nicht zu vermeiden.

5.3.2 Expression des LabKC-Proteins für den *in vitro*-Assay

Für den *in vitro*-Assay zur Untersuchung der katalytischen Aktivität des Enzyms LabKC musste neben den Substraten natürlich auch das LabKC-Enzym selbst überexprimiert und gereinigt werden. Dafür wurde, wie auch schon zur Darstellung der Substrate, das Champion pET SUMO-System (s. Abschnitt 5.3.1) verwendet. Da die meisten Schritte nahezu identisch waren, sollen im Folgenden nur kurz die Unterschiede diskutiert werden.

Die PCR zur Generierung des *labKC*-Gens erfolgte nach dem in Abschnitt 4.2.3.1 beschriebenen Protokoll mit einer Polymerase mit 3´->5´-Exonukleaseaktivität, weshalb nach der Reinigung der Banden aus dem Gel (s. Abschnitt 4.2.2.10) zur Klonierung in den Vektor mit T-Überhang zunächst eine *A-Addition* erfolgen musste (s. Abschnitt 4.2.4.3). Die Primer wurden wie in Abbildung 65 dargestellt wiederum so abgeleitet, dass der Vorwärtsprimer den späteren *N*-Terminus darstellte und das Stopcodon im Rückwärtsprimer enthalten war.

Abbildung 65: Ableitung der Primer für die heterologe Expression des LabKC-Enzyms.

Zur Kontrolle der Orientierung wurden die erhaltenen Plasmide zunächst durch Restriktionsverdaus vorselektioniert. Um sicherzustellen, dass sich bei der PCR des ca. 3

kpb großen Produktes kein Fehler eingeschlichen hatte, wurde das Insert eines im Verdau positiv bewerteten Plasmids zunächst an den Ränder und dann komplett sequenziert. Die Sequenzierungen zeigten keinen Fehler an, weshalb Plasmid LabKC-Pl3 weiterverwendet und dafür in den Expressionsstamm *E. coli* BL21 (DE3) transformiert werden konnte.
Aufgrund der erwarteten Größe des exprimierten Proteinkonstrukts von 108,4 kDa und der zu befürchtenden Bildung von Inclusion-Bodies wurde hier eine sehr langsame (Kultivierung bei 15°C; 0,1 mM IPTG) dafür aber längere (72 h) Expression gewählt (s. Abschnitt 4.4.8). Testansätze bei höheren Temperaturen hatten gezeigt, dass bei einer Kultivierung über 27°C das exprimierte Protein nahzu ausschließlich in Inclusion-Bodies vorlag. Die durch die Kultivierung erhaltenen Zellen wurden aufgeschlossen und daraus das LabKC-His-Tag-Konstrukt über eine Ni-NTA-Agarose-Säule angereichert und gereinigt (s. Abschnitt 4.4.9). Die einzelnen Schritte wurden mittels Gelelektrophorese nach Lämmli (s. Abschnitt 4.4.3) analysiert und die interessanten Fraktionen zur Dialyse vereinigt. Das erhaltene Dialysat wurde zur Abspaltung des *N*-terminalen His-Tags und zur anschließenden Aufreinigung über eine Ni-NTA-Spin-Säule eingesetzt. Der Fortschritt der SUMO-Spaltung konnte durch die Bildung des gewünschten freien LabKC-Proteins mit der Zeit beobachtet werden (s. Abbildung 66).

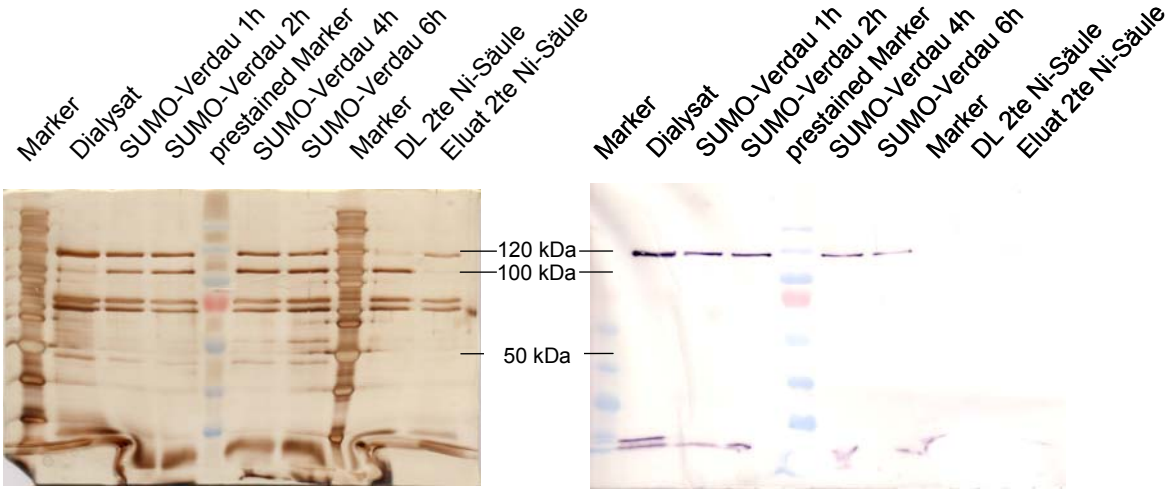


Abbildung 66: Zeitlicher Verlauf der SUMO-Spaltung im silbergefärbten Gel und im *Western-Blot* (10 % Trenngel und 5 % Sammelgel).

Die Bande des LabKC-His-Tag-Konstrukts (108,5 kDa) bei knapp unter 120 kDa zeigte sich nur in den induzierten Ansätzen und nahm ausgehend vom Dialysat mit fortlaufender Zeit der SUMO-Spaltung ab. Im Größenbereich knapp unter 100 kDa nahm die Intensität des Spaltproduktes LabKC (95 kDa) gleichermaßen zu. Der Durchlauf der Ni-NTA-Spin-Säule zeigte wie gewünscht nur noch das freie LabKC-Protein sowie eine Verunreinigung in Form einer Doppelbande bei 70 – 80 kDa, die auch nicht durch die zweite Ni-Säule entfernt werden konnte. Die Abnahme des LabKC-His-Tag-Konstrukts konnte auch eindrucksvoll im *Western-Blot* verfolgt werden.

Zusammenfassend kann man sagen, dass es gelungen ist, das LabKC-Protein in löslicher Form zu exprimieren. Für die im Proteinblast gefundenen homologen Proteine aus Streptomyceten (s. Abschnitt 5.2.1.5 und Tabelle 27) wurde bisher eine Lokalisation in der Membran diskutiert,[41,84] was aufgrund der Expression als lösliches Protein für LabKC nicht angenommen werden kann. Die nach der zweiten Ni-Säule noch zu beobachtenden Verunreinigungen, deren Gehalt aufgrund der nicht-quantitativen Silberfärbung nicht abgeschätzt werden konnte (s. Abschnitt 4.4.6), wurden für das geplante Vorhaben des *in vitro*-Assays nicht weiter berücksichtigt.

5.3.3 *In vitro*-Assay zur Untersuchung der katalytischen Aktivität des LabKC-Proteins

Wie in den Abschnitten 5.3.1 und 5.3.2 beschrieben, konnten die Präpropeptide von LabA2, LabA2 Ser4->Ala und LabA2 Ser13->Ala sowie das Enzym LabKC heterolog exprimiert und gereinigt werden. Mit diesen Voraussetzungen sollte nun ein *in vitro*-Assay zur Untersuchung der katalytischen Aktivität des LabKC-Proteins durchgeführt werden. Die Durchführung des Assays sah eine Mischung von Substrat- und Enzym-Lösung in einem geeigneten Puffer und die Zugabe von ATP und unterschiedlichen Kationen wie z.B. Zn vor. Nach einer bestimmten Inkubationszeit sollte der Ansatz dann durch GluC-Spaltung für eine Untersuchung mittels HPLC-ESI-MS vorbereitet und anschließend vermessen werden. Bevor jedoch die Reaktionsbedingungen optimiert wurden, sollte zunächst die Möglichkeit zur massenspektrometrischen Detektion der Präpropeptide geprüft werden. Wie in Abschnitt 5.3.1 bereits angedeutet war zwar die Darstellung der Präpropeptide gelungen, allerdings nur in relativ geringen Mengen. Da es bei der Durchführung des geplanten Assays durch Zugabe von Reaktionspuffer, Cosubstraten und Enzym noch zu einer zusätzlichen Verdünnung kommt, wurde zunächst in Vorversuchen untersucht, ob eine Detektion der interessierenden Fragmente mit dem zur Verfügung stehenden System überhaupt noch möglich ist. Dazu wurden Verdünnungen vermessen, die denen in den geplanten Assays entsprachen. Dabei stellte sich leider heraus, dass keine massenspektrometrische Detektion mehr möglich war. Bei der Detektion des *in vitro*-Assays käme außerdem noch ein unter Umständen nicht-quantitativer Umsatz der Präpropeptide dazu.

Zusammenfassend kann man sagen, dass es gelungen ist, die Präpropeptidsubstrate und das Enzym LabKC für einen *in vitro*-Assay zur Untersuchung der katalytischen Aktivität des Enzyms zu generieren. Aufgrund der Tasache, dass bei der Darstellung der Präpropeptidsubstrate die dafür notwendige enzymatisch katalysierte SUMO-Spaltung die Ausbeute deutlich verringert, war eine Detektion der Substrate unter *in vitro*-Assay-Bedingungen nicht mehr möglich. Das heißt, dass in Zukunft für den *in vitro*-Assay diverse Optimierungen oder gar ganz neue Wege der Substratgenerierung getestet werden müssen.

5.4 Heterologe Expression des Cosmids 1104 in *Streptomyces lividans*

Wie die vorherigen Abschnitte zeigten, ist es in dieser Arbeit gelungen, eine neues Lantibiotika-Biosynthese-Genclustern zu identifizieren. Dabei wurde das Enzym LabKC identifiziert, welches voraussichtlich maßgeblich an der posttranslationalen Modifikation und der damit einhergehenden Reifung der Labyrinthopeptine beteiligt ist. Es wird angenommen, dass es neben der Einführung der Lanthioninbrücken auch für die Zyklisierung zu den in den Labyrinthopeptinen zum ersten Mal beobachteten Labioninbrücken verantwortlich ist. Bisher gründen diese Annahmen jedoch nur auf Homologievergleichen mit anderen Lantibiotika-Genclustern (s. Abschnitt 5.2.1.5) und wurden noch nicht experimentell bewiesen. Zur Verifizierung der biosynthetischen Leistung der im Labyrinthopeptin-Gencluster kodierten Proteine (LabKC, LabT1 und LabT2) stehen verschiedene Methoden zur Verfügung. Die Erzeugung von Geninaktivierungs-Mutanten im Labyrinthopeptin Wildtyp-Produzenten *Actinomadura namibiensis* war bisher aufgrund der vergeblich gesuchten Möglichkeit zur Einführung von Fremd-DNA in den Wildtyp nicht möglich. Deshalb wurde, wie in Abschnitt 5.3 beschrieben, ein *in vitro*-Assay zur gezielten Untersuchung des Enzyms LabKC in Betracht gezogen. In der vorliegenden Arbeit konnten die Voraussetzungen für einen solchen Assay in Form der Bereitstellung des Enzyms LabKC und der vermuteten Präpropeptidsubstrate geschaffen werden. Allerdings muss zukünftig die Herstellung der Substrate optimiert werden, um Konzentrationen zu erhalten, mit denen eine Detektion des Assays möglich wird. Aufgrund dieser Schwierigkeiten wurde in der vorliegenden Arbeit eine dritte Methode herangezogen, die in diesem Abschnitt beschrieben werden soll. Analog zum Lantibiotikum Cinnamycin[39] sollte die Funktionalität und Vollständigkeit des gefundenen Biosynthese-Genclusters durch heterologe Expression bewiesen werden. Sollte es gelingen, die Biosynthese der Labyrinthopeptine heterolog in einem anderen, genetisch handhabbaren Stamm nachzustellen, könnten Experimente wie Geninaktivierungen oder kombinatorische Biosynthese, die bisher wegen der schweren Handhabung des Wildtyps nicht möglich gewesen waren, in diesem Stamm durchgeführt werden. Deshalb bot es sich an, dafür einen „Standardstamm“ wie z.B. *Streptomyces lividans* zu verwenden, da dessen Handhabung bereits optimiert und in zugänglichen Protokollen beschrieben wurde.

Für die heterologe Expression des Labyrinthopeptin-Genclusters gab es zunächst zwei denkbare Strategien, wobei beide auf einer ortsspezifischen Integration in die Bakteriophagen ΦC31 attachment-site (*attB*-site) des Genoms von *S. lividans* basierten.[91] Diese Integration ist sehr effizient und setzt die Anwesenheit des Integrasegens des Bakteriophagen ΦC31 und der *attP*-attachment-site auf dem zu integrierenden Cosmid oder Plasmid voraus. Einige Vektoren, die für die Herstellung von Cosmidbänken verwendet werden, z.B. pOJ436, besitzen diese Voraussetzungen (*int/attP*) für die Integration.[93,92] Bei der Herstellung der Cosmidbank für den Stamm *Actinomadura namibiensis* wurde zunächst

die Verwendung des Vektors pOJ436 versucht, allerdings gelang nur die Herstellung einer Bank basierend auf pWEB, der nicht die notwendigen Elemente (*int/attP*) für eine Integration besitzt. Deshalb ist mit dem Gencluster-tragenden Cosmid 1104 direkt keine ortsspezifische Integration möglich, was jedoch durch zwei verschiedene Strategien behoben werden könnte.

Die erste Strategie sah vor, ein durch Restriktionsverdau von Cosmid 1104 generiertes, Gencluster-tragendes Fragment in einen passenden Vektor wie z.B. pSET zu klonieren, der Integrasegen und *attP*-attachment-site enthält und deshalb in das Genom von *S. lividans* integriert werden kann.[93] Der Vorteil dieser Strategie ist, dass durch geschickten Restriktionsverdau das Cluster ohne große Reste des Cosmids 1104, die offensichtlich nicht mehr zum Cluster gehören, in das Genom integriert werden könnte. Die Schwierigkeit war in diesem Fall jedoch das Auffinden geeigneter Restriktionsenzyme bzw. -spaltstellen. Ein 14 kbp großes StuI-Fragment (von 11638 bis 25772 s. Abschnitt 7.1) war das „kleinste" Fragment, dass durch Spaltung generiert werden konnte und wurde deshalb verwendet. StuI erzeugt jedoch blunt-end Schnittstellen, weshalb ein Umweg über *A-Addition* (s. Abschnitt 4.2.4.3) und TA-Klonierung in einen Standardkloniervektor (pDrive oder pCR4-TOPO) (s. Abschnitt 4.2.4.4) gewählt wurde. Trotz vieler Versuche konnte mit dieser Strategie kein Ergebnis erzielt werden, was möglicherweise an der Größe des Fragments lag.

Aufgrund dessen wurde eine zweite Strategie verfolgt, die auf einer Modifikation des Vektorrückgrats von Cosmid 1104 (pWEB) basierte. Die Durchführung dieser Modifikation wird in Abschnitt 5.4.1 beschrieben.

5.4.1 Herstellung des Cosmids M1104 zur heterologen Expression des Labyrinthopeptin-Genclusters

Wie in Abschnitt 5.4 beschrieben, war für die ortsgerichtete Integration des Cosmids 1104 in das Genom von *S. lividans* die Einführung des ΦC31 Integrasegens (*int*) und der *attP*-attachment-site in das Cosmidrückgrat notwendig, welches dann als Cosmid M1104 bezeichnet wurde. Aus Selektionsgründen sollte außerdem eine Antibiotikaresistenz eingebracht werden.

Dazu wurde eine auf λ-RED-basierende Strategie herangezogen, die ursprünglich in *E. coli* entwickelt worden ist und später auf Streptomyceten erweitert wurde.[94,95] Dabei machte man sich die Beobachtung einer erhöhten Rekombinationsrate von linearer DNA bei Anwesenheit der Bakteriophagen λ-Proteine Redα (*exo*), Redβ (*bet*) und Redγ (*gam*) zunutze.[96] Die Standardanwendung dieses Systems ist die Ersetzung und damit Deletion von Genen durch Resistenzkassetten. Die Primer für die PCR zur Erzeugung der Resistenzkassetten tragen endständig mindestens 39 Nukleotide lange Überhange, die homolog zu den Rändern des

zu deletierenden Gens sind und deshalb die Rekombination des linearen PCR-Produkts ermöglichen.[94,95]

In dieser Arbeit war die Aufgabe jedoch nicht die Deletion eines Gens, sondern die Insertion der drei Sequenzabschnitte für ΦC31 Integrase (*int*), *attP*-attachment-site und Antibiotikaresistenz. In der Publikation zur heterologen Expression von Novobiocin und Clorobiocin wurde eine analoge Aufgabe schon einmal mit dem λ-RED-System gelöst, indem das Ampicillin-Resistenzgen (*bla*) im Vektorrückgrat von SuperCos1 durch das Integrasegen (*int*) und die *attP*-attachment-site ersetzt wurde.[97] Diese Vorgehensweise sollte in der vorliegenden Arbeit auf den Cosmidvektor pWEB übertragen werden, der auch das Ampicillin-Resistenzgen (*bla*) enthält. Ziel war es nun, mittels PCR ein lineares DNA-Stück zu amplifizieren, welches die drei einzubringenden Elemente enthält, aber keine Homologien zu pWEB zeigt. Dies ist wichtig, um ungewollte Rekombinationen zu vermeiden. Als Templat für die PCR wurde deshalb der Vektor pSET152 ausgewählt, der vom Apramycin-Resistenzgen (*apra*) mit Promotor bis zum Ende des *int*-Gens amplifiziert wurde. Als 39 Nukleotide lange Überhänge der Primer für die gewünschte Rekombination wurden der Anfang und das Ende des Ampicillin-Resistenzgens (*bla*) gewählt (s. Abbildung 67). Daraus ergaben sich die in Tabelle 31 angegebenen Primer. Das damit erzeugte PCR-Produkt sollte dann mit Hilfe des λ-RED-Systems in Cosmid 1104 rekombinieren und dadurch das Ampicillin-Resistenzgen (*bla*) durch das Apramycin-Resistenzgen (*apra*), die *attP*-attachment-site und das Integrasegen (*int*) ersetzen.

Tabelle 31: Primer für den Austausch des Ampicillin-Resistenzgens (*bla*) gegen das Apramycin- und Integrasegen sowie die *attP*-attachment-site aus pSET152 mit dem λ-RED-System.

Primer	Primerüberhang homolog zu *bla*	Primeranteil für pSET152
Φapra Fw	5´-AAAGTTCTGCTATGTGGCGCGGTATTATCCCTGGTTGACG	GGTTCATGTGCAGCTCCATC-3´
Φapra Rev	5´-CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAA	CAGGCTTCCCGGGTGTCTC-3´

Die PCR erfolgte unter den in Abschnitt 4.2.3.2 beschriebenen Bedingungen und wurde anschließend aus dem Gel eluiert (s. Abschnitt 4.2.2.10) und gereinigt.

Für die Rekombination mit dem λ-RED-System musste zunächst das Cosmid 1104 in *E. coli* BW25113/pIJ790 transformiert werden, was nach dem in Abschnitt 4.2.1.3 beschriebenen Protokoll mittels Elektroporation durchgeführt wurde. PIJ790 ist dabei die Bezeichnung für das Rekombinationsplasmid, auf dem die drei Proteine des λ-RED-Systems kodiert sind und das zur Selektion ein Chloramphenicol-Resistenzgen enthält. Der Elektroporationsansatz

wurde nach einer Präinkubationszeit ohne Selektion auf TY-Carb/Kan/Cm-Platten ausgestrichen, wobei Carbenicillin (als Ampicillinanalogon) und Kanamycin zur Selektion auf Cosmid 1104 und Chloramphenicol zur Selektion auf den Vektor pIJ790 verwendet wurden. Nach Inkubation (24 h) bei 28°C sollten die Cosmide der gewachsenen Klone aufgrund der mehrfach beobachteten Instabilität der Inserts zunächst auf Vollständigkeit untersucht werden. Dazu wurden ausgehend von Einzelkolonien von Platte je zwei Flüssigkulturen (TY-Carb/Kan/Cm) beimpft, wobei ein Ansatz bei 28°C und einer bei 37°C kultiviert wurde. Die 37°C-Kultur wurde zur Cosmidisolierung verwendet, da bei 37°C das pIJ790-Plasmid aufgrund eines temperatursensitiven Replikons verloren geht und damit die Cosmidisolierung erleichtert. Die isolierten Cosmide wurden verdaut (BamHI) und aufgrund des Bandenmusters analysiert. Der Verdau eines Klons zeigte ein intaktes Cosmid 1104 an, so dass mit der entsprechenden 28°C-Kultur eine Hauptkultur für den nächsten Schritt angeimpft werden konnte. Der nächste Schritt sah die Transformation und Rekombination des linearen, gelgereinigten PCR-Produkts nach dem in Abschnitt 4.2.1.3 beschriebenen Protokoll vor, wobei zur Expression der dafür notwendigen Red-Proteine zusätzlich *L*-Arabinose (100 µl; 1 M) zur Hauptkultur gegeben werden musste. Der Transfer des linearen PCR-Produkts erfolgte wiederum mittels Elektroporation und die anschließende Kultivierung (28°C) auf TY-Kan/ApraPlatten. Dieser Antibiotikamix selektierte folglich auf Klone, bei denen die Ampicillinresistenz (Carb) in Cosmid 1104 durch das lineare PCR-Produkt mit Kanamycinresistenz ersetzt worden ist und die somit Cosmid M1104 tragen. Da das Rekombinationsplasmid pIJ790 im Folgenden nicht mehr gebraucht wurde, wurde der Chloramphenicol-Selektionsdruck aufgehoben. Das Standardprotokoll sah an dieser Stelle zwar eine Kultivierung bei 37°C vor, um zusätzlich die Entfernung des pIJ790-Plasmids zu forcieren, allerdings wurde wegen der geringen Temperaturstabilität der Cosmide weiterhin bei 28°C kultiviert. Nach der Inkubation (24 h) war ein einzelner Klon auf der Platte gewachsen. Dieser wurde dann in Flüssigmedium (TY-Kan/Apra) kultiviert sowie zur Stammhaltung auf Platte (TY-Kan/Apra) ausgestrichen. Abbildung 67 zeigt noch einmal schematisch die einzelnen Schritte, die zu diesem gewünschten Klon bzw. Cosmid M1104 geführt hatten.

Die Bestätigung, dass dieser Klon das gewünschte und unbeschädigte Cosmid M1104 enthält, erfolgte auf zwei Wegen. Zum einen wurde mit den Kontrollprimern ΦintKontrolle Fw und ΦintKontrolle Rev (s. Abschnitt 3.9) eine Kolonie-PCR (s. Abschnitt 4.2.3.6) durchgeführt und zum anderen wurde das isolierte Cosmid mit BamHI verdaut, um das Fragmentmuster mit dem des unmodifizierten Cosmids 1104 vergleichen zu können. Die positiven Ergebnisse sind in Abbildung 68 dargestellt.

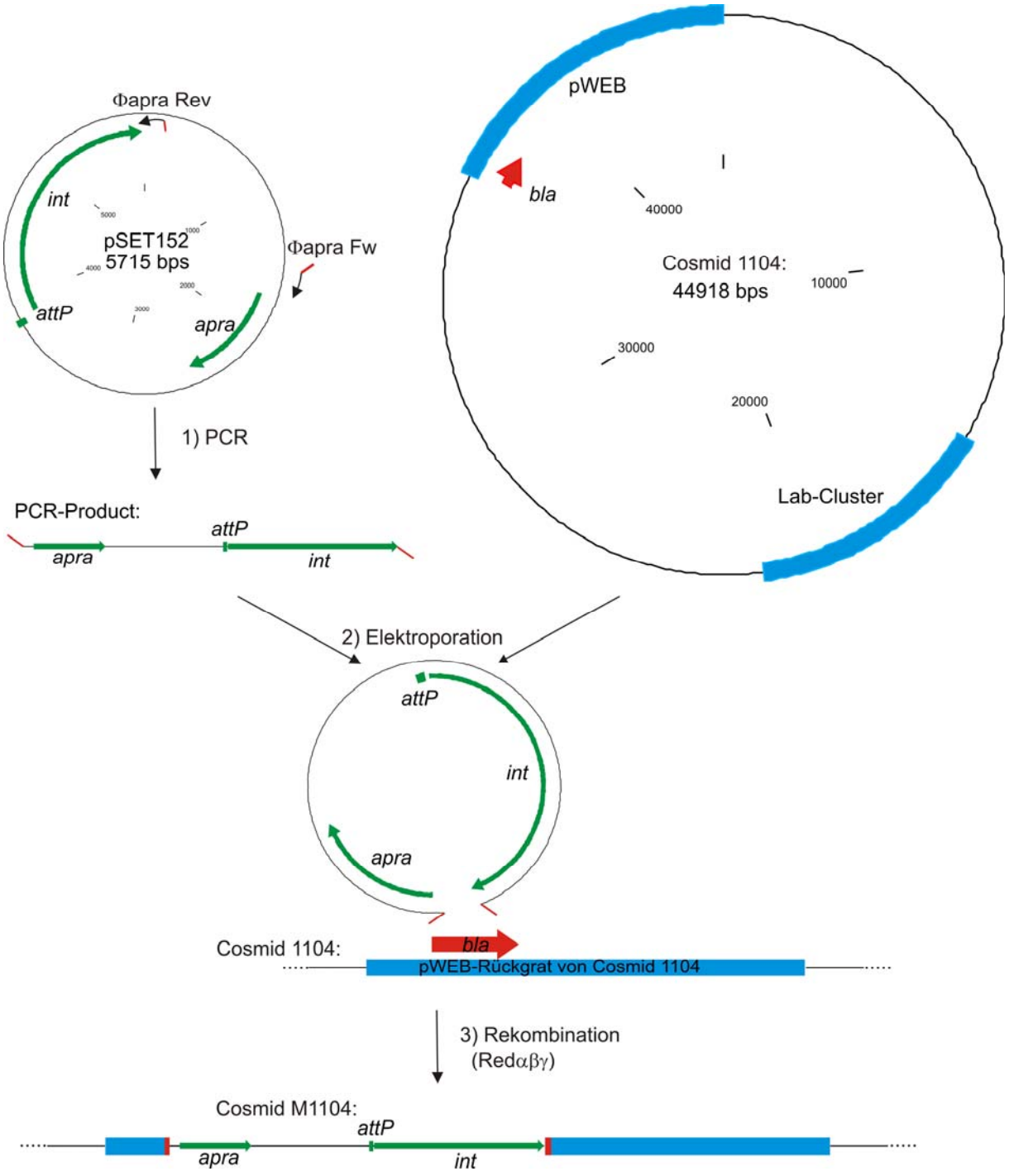


Abbildung 67: Schematische Übersicht über die in dieser Arbeit verwendete Strategie zur Modifizierung des pWEB-Rückgrats von Cosmid 1104 zur Generierung des Cosmids M1104 in *E. coli* als Wirtsorganismus.

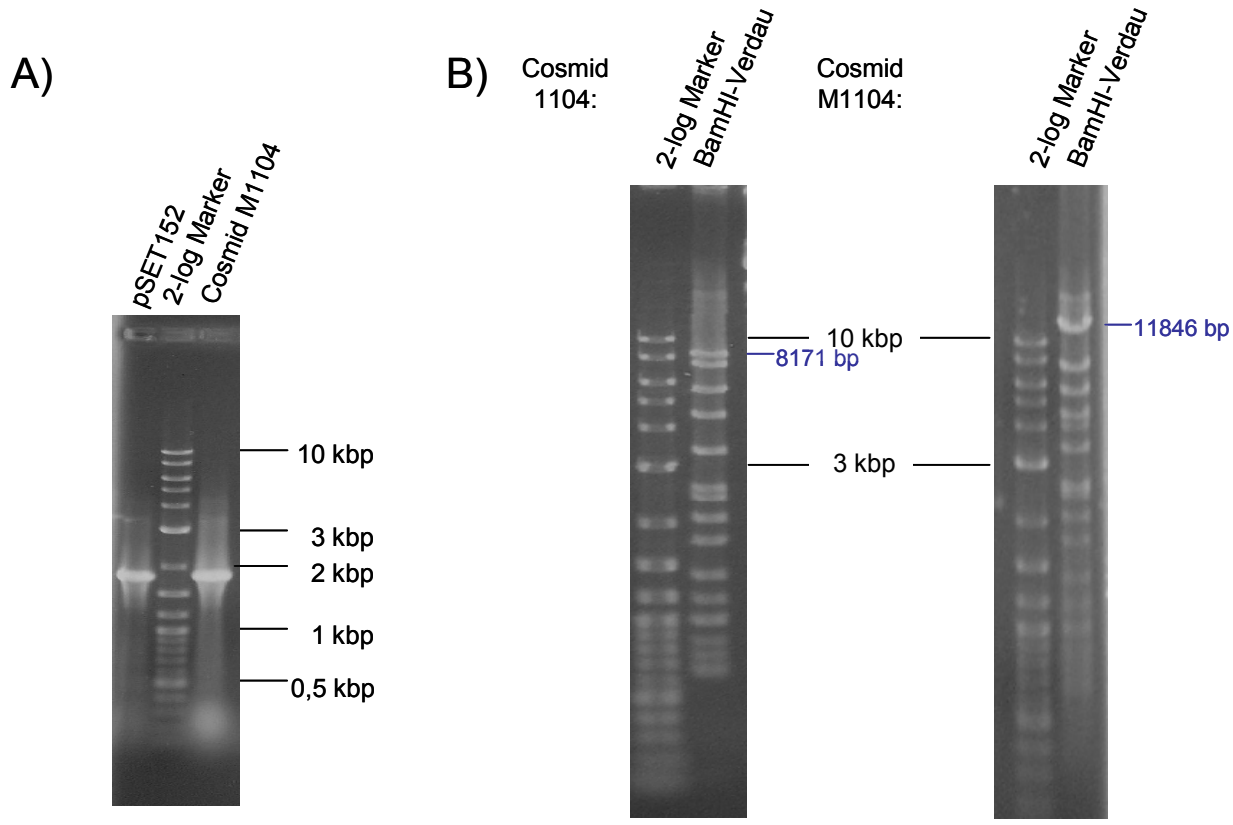


Abbildung 68: Untersuchung der Richtigkeit und Integrität des erzeugten Cosmids M1104. A) Ergebnisse der Kolonie-PCR mit den Kontrollprimern ΦintKontrolle Fw und ΦintKontrolle Rev. Der Ansatz mit Cosmid M1104 zeigte wie der Kontrollansatz mit pSET152 die erwartete Bande bei 1790 bp. B) BamHI Restriktionsverdaus von Cosmid 1104 und Cosmid M1104. Der Vergleich zeigt den erwarteten Unterschied bei Cosmid M1104, nämlich den Wegfall der in Cosmid 1104 vorkommenden Bande bei 8171 bp und das Erscheinen einer zusätzlichen Bande bei 11846 bp. Außerdem spricht der Bandenreichtum für die Unversehrtheit von Cosmid M1104.

Mit dem so erzeugten Cosmid M1104 sollte im Folgenden die Transformation in *S. lividans* vorgenommen werden, was allerdings das nicht-methylierte Vorliegen von Cosmid M1104 erforderte. Um dies zu erreichen, wurde es in $CaCl_2$-kompetente *E. coli* ET12567 (s. Abschnitt 4.2.1.1) transformiert (s. Abschnitt 4.2.1.2). Aufgrund der bekannten geringeren Transformationseffizienz in *E. coli* ET12567 wurden hier jedoch abweichend vom Standardprotokoll 10 µl Cosmid-DNA eingesetzt. Die Kultivierung des Transformationsansatzes erfolgte wie gewohnt bei 28°C, wobei neben den Selektionsmarkern Kan und Apra für Cosmid M1104 auch Chloramphenicol eingesetzt wurde, welches zur Selektion auf die *dam*$^-$-Mutation und damit auf die Nicht-Methylierung diente.[98] Nach 72 h konnte wiederum ein einzelner Klon auf den Platten beobachtet werden, der in Flüssigmedium (TY-Kan/Apra/Cm) kultiviert und auf Platte (TY-Kan/Apra/Cm) ausgestrichen wurde. Aus der Flüssigkultur konnte das nicht-methylierte Cosmid M1104 mittels dem in Abschnitt 4.2.2.9 beschriebenen Protokoll isoliert werden, wobei die Integrität über Kontrollverdau mit BamHI bestätigt wurde.

5.4.2 Integration des Cosmids M1104 ins Genom von *S. lividans* und Untersuchungen zur heterologen Expression

Zur heterologen Expression des Labyrinthopeptin-Genclusters in *S. lividans* wurde das Cosmid 1104 wie in Abschnitt 5.4.1 beschrieben modifiziert, um eine ortsspezifische Integration in das Genom über die *attP*-attachment-site zu ermöglichen. Zur Transformation des aus *E. coli* ET12567 isolierten und damit nicht-methylierten Cosmids M1104 in den Stamm *S. lividans* wurde das in Abschnitt 4.2.1.4 beschriebene Protokoll der Protoplastentransformation verwendet. Nach der Kultivierung der Transformationsansätze (ca. 120 h) bei 28°C konnten *S. lividans*-Klone beobachtet werden, die die Selektionsmarker-enthaltende Weichagarschicht durchbrochen hatten und somit höchstwahrscheinlich Cosmid M1104 ortsspezifisch ins Genom integriert hatten. Von einigen dieser Klone wurden zur Stammhaltung Verdünnungsausstriche erstellt. Die erfolgte Integration wurde im Folgenden durch PCR und *Southern-Blot* untersucht. Die Untersuchung mittels PCR erfolgte anhand von PCR-Reaktionen mit Primern aus der Umgebung der *attP*-site des Cosmids M1104 und der *attB*-site des *S. lividans*-Genoms (s. Abbildung 69).

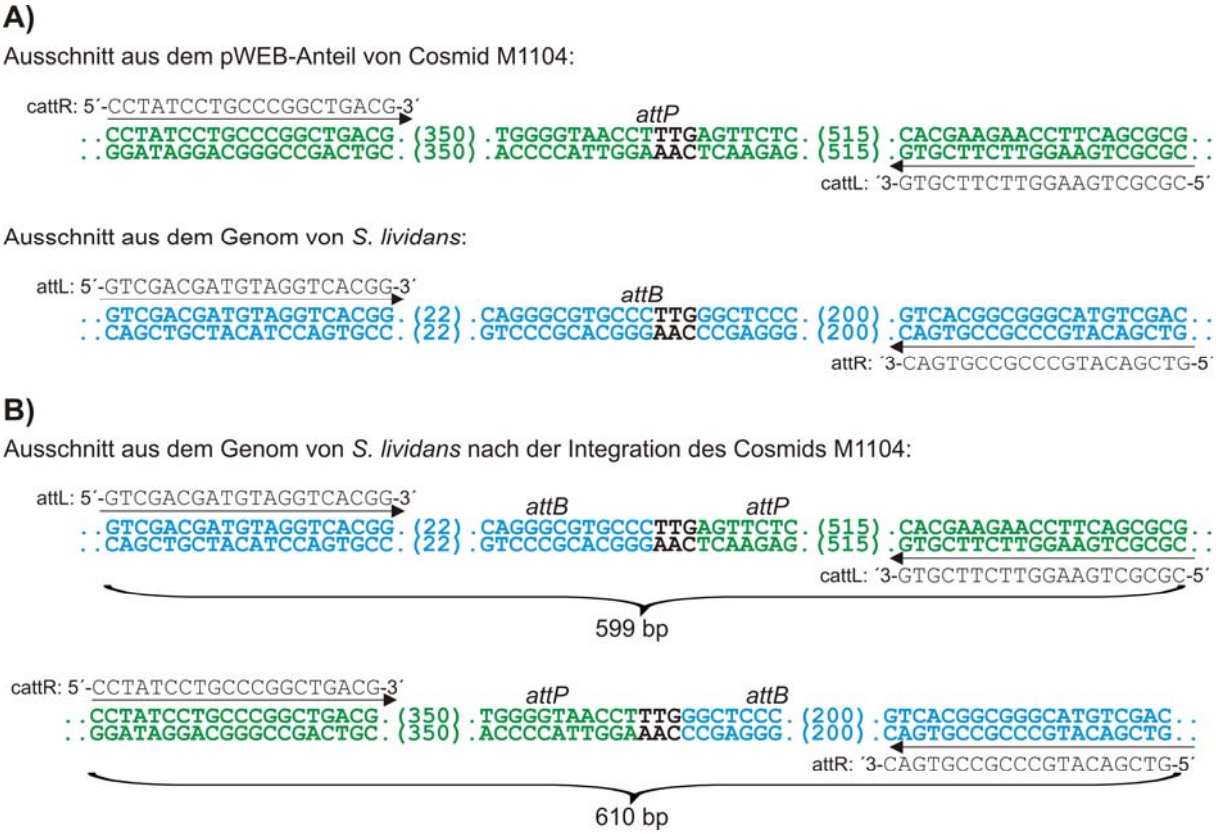


Abbildung 69: Schematische Darstellung der Kontroll-PCRs zur Untersuchung der korrekten Integration von Cosmid M1104 in das Genom von *S. lividans*. A) Die Primer wurden im Bereich der *attP*(Cosmid)- und *attB*(Genom)-site abgeleitet. B) Gezeigt sind die Übergänge von der Cosmid- in die Genom-Sequenz nach der Integration des Cosmids M1104. Bei richtiger Integration ergeben die Primerpaarkombinationen attL-cattL und attR-cattR PCR-Fragmente der Größen 599 bp und 610 bp.

Für die PCR-Ansätze (s. Abschnitt 4.2.3.1) mit den Primerpaaren attL-cattL und attR-cattR wurde genomische DNA aus vier *S. lividans*-Klonen (K1-4) sowie aus *S. lividans*-Wildtyp als Kontrolle isoliert (s. Abschnitt 4.2.2.7) und eingesetzt. Die Auswertung erfolgte mittels Agarosegelelektrophorese (s. Abschnitt 4.2.2.1) und ist in Abbildung 70 dargestellt.

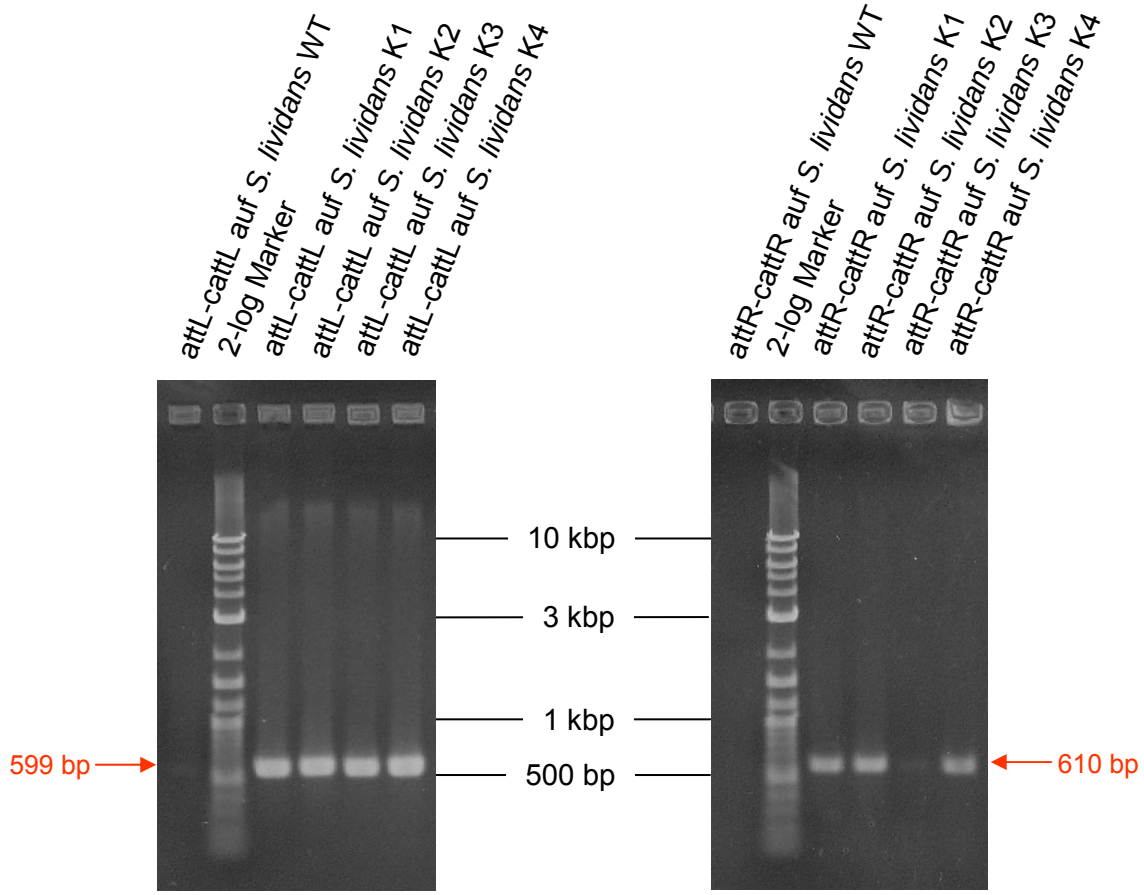


Abbildung 70: Auswertung der Gele der Kontroll-PCRs auf korrekte Integration des Cosmids M1104 in das Genom von *S. lividans.* Man sieht in den Ansätzen der Klone 1-4 die erwarteten PCR-Banden bei 599 bzw. 610 bp. Allerdings ist die Bande der attR-cattR-PCR bei Klon 3 verhältnismäßig schwach. Der als Negativkontrolle verwendete *S. lividans*-Wildtyp zeigte wie erwartet keine PCR-Produkte.

Die Auswertung der Gele zeigte vor allem bei den Klonen 1, 2 und 4 die erwarteten Banden, so dass davon ausgegangen werden konnte, dass in diesen Fällen die Integration des Cosmids M1104 in das Genom von *S. lividans* erfolgreich verlaufen war.

Zur Sicherheit und zur Überprüfung der Integrität des integrierten Cosmids M1104 wurde noch ein *Southern-Blot* wie in Abschnitt 4.2.2.3 beschrieben durchgeführt. Als Sonde für diesen *Southern-Blot* sollte ein Gemisch kleiner, DIG-markierter Fragmente von Cosmid M1104 verwendet werden. Dazu wurde das Cosmid M1104 mit MlyI in kleine Fragmente verdaut, die dann durch einen *random primed oligo-labeling*-Mechanismus mit DIG markiert wurden. Parallel dazu wurde genomische DNA der *S. lividans*-Klone K1-4, genomische DNA des *S. lividans*-Wildtyps als Negativkontrolle sowie DNA des Cosmids M1104 als

Positivkontrolle BamHI verdaut und gelelektrophoretisch aufgetrennt. Nach dem Blot erfolgte die Hybridisierung mit der DIG-markierten Sondenmischung, wobei bei den *S. lividans*-Klonen K1-4 mit Ausnahme der Fragmente, die den Übergang zwischen Genom und Cosmid enthalten, ein identisches Bandenmuster wie bei Cosmid M1104 erwartet wurde. Abbildung 71 zeigt den *Southern-Blot* exemplarisch für K1 und K2.

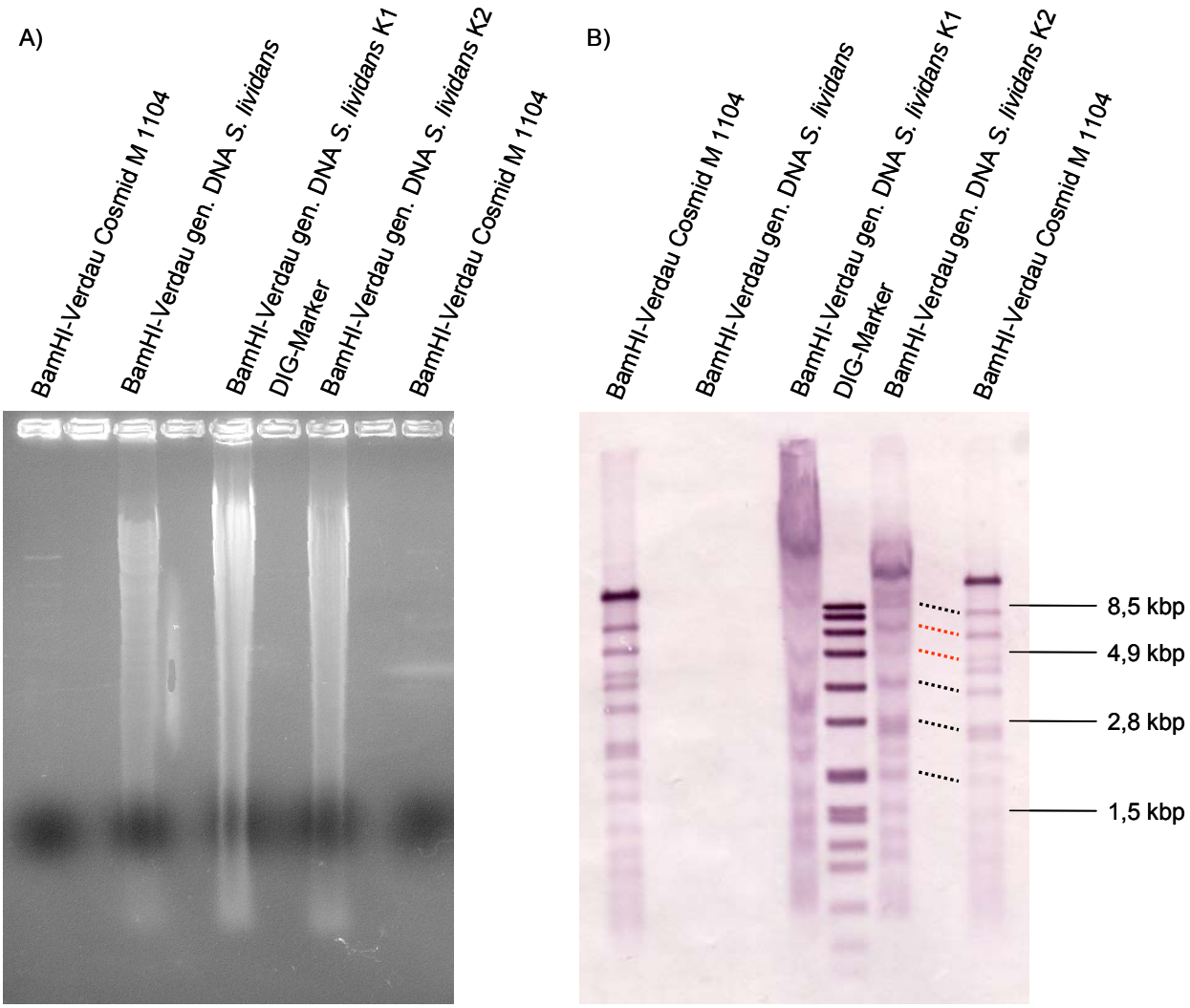


Abbildung 71: A) 1%iges Agarosegel der BamHI-Verdaue, welches für den *Southern-Blot* verwendet wurde. Man sieht deutlich das Vorhandensein von DNA, v.a. in den drei genomischen Ansätzen. B) Der *Southern-Blot* zeigte für *S. lividans* Wildtyp wie erwartet keine Bande und für die *S. lividans* Klone 1 und 2 ein sehr ähnliches Bandenmuster wie für das Cosmid M1104. Aufgrund einer leichten Verschiebung der Bandenmuster wurden schwarze Linien zur Orientierung eingezeichnet. Die roten Striche zeigen die Banden, die das Labyrinthopeptin-Biosynthese-Gencluster beinhalten.

Dieser *Southern-Blot* zeigte in der Spur des *S. lividans* Wildtyps wie erwartet keine Bande. Den Beweis, dass diese Tatsache nicht auf einer zu geringen DNA-Konzentration beruht, zeigt das Gel in Abbildung 71A). In den Spuren der *S. lividans* K1- und K2-Klone sieht man hingegen deutlich Banden, die ein leicht nach oben verschobenes, zum Cosmid M1104 sehr ähnliches Bandenmuster aufweisen. Dies deutet an, dass das Cosmid M1104 vollständig in

die Genome der *S. lividans* K1- und K2-Klone integriert wurde. Die leichte Verschiebung im Vergleich zum Cosmid wurde wahrscheinlich durch die hohe Viskosität der Proben aufgrund des hohen DNA-Gehalts hervorgerufen. Der Vergleich der Bandenmuster am Beispiel von *S. lividans* K1 und Cosmid M1104 zeigte, dass auch die Fragmente des Labyrinthopeptin-Biosynthese-Gencluster im genomischen Verdau identifiziert werden konnten (s. Abbildung 72).

Nachdem somit die Integration des vollständigen Cosmids M1104 in das Genom von *S. lividans* nachgewiesen war, sollte im Folgenden die heterologe Expression des nun im Genom enthaltenen Labyrinthopeptin-Genclusters anhand der Produktion der Labyrinthopeptine untersucht werden. Eine heterologe Expression ist im Allgemeinen nicht trivial und bedarf einiger Optimierung vor allem bezüglich Medienwahl und Kultivierungsbedingungen. In einem ersten Versuch wurde zunächst mit zwei verschiedenen Medien (5294 und R5) bei Standardkultivierungsbedingungen (28°C; 120 rpm) gearbeitet. Dazu wurden von den *S. lividans*-Klonen K1-4 zunächst Vorkulturen (20 ml) der entsprechenden Medien beimpft. Als Kontrollen wurden der *S. lividans*-Wildtyp (Negativkontrolle) und *Actinomadura namibiensis* (Positivkontrolle) verwendet. Nach 48 h wurden aus den Vorkulturen Hauptkulturen beimpft (1 %), die dann bei gleichen Bedingungen weiter kultiviert wurden. Nach 72, 144 und 192 h wurde aus jeder Hauptkultur eine Probe (500 µl) entnommen und wie in Abschnitt 4.1.4 beschrieben für die Analyse mittels HPLC-ESI-MS vorbereitet und vermessen. Die Auswertung der Analysen zeigte allerdings nur in den Proben der Positivkontrolle *Actinomadura namibiensis* eine Produktion der Labyrinthopeptine an, was in Abbildung 72 exemplarisch für den 72 h alten Kulturüberstand von *Actinomadura namibiensis* in Medium 5294 dargestellt ist.

In den *S. lividans*-Ansätzen konnte leider keine Masse beobachtet werden, die mit den Labyrinthopeptinen bzw. hypothetischen Biosyntheseintermediaten in Verbindung gebracht werden konnten. Dieses Ergebnis kann viele Ursachen haben, wobei im Folgenden einige Möglichkeiten diskutiert werden sollen. Eine mögliche Erklärung ist die Unvollständigkeit des eingebrachten Labyrinthopeptin-Biosynthese-Genclusters. Es konnten z.B. keine Gene für die Regulation identifiziert werden, was die nicht erfolgende Expression der Biosynthesegene erklären könnte. Desweiteren fiel bei der Betrachtung des Labyrinthopeptin-Genclusters (s. Abschnitt 5.2.1.5) auf, dass keine Protease zu finden ist, die die Abspaltung der Leaderpeptide vornimmt. Dieses Phänomen wurde auch schon in anderen bekannten Lantibiotika-Genclustern beobachtet, wobei dort die Wirkung zellulärer Proteasen vermutet wurde. Im hier vorliegenden Fall der heterologen Expression des Labyrinthopeptin-Genclusters könnte eine fehlende zelluläre Proteaseaktivität dazu führen, dass die gebildeten Labyrinthopeptine nicht von ihren Leaderpetiden befreit werden, was eine Detektion aus dem Kulturüberstand mit dem gewählten HPLC-ESI-MS-System aufgrund

der Größe erschwert. Außerdem könnte der Exportprozess davon negativ beeinflusst werden, was zu einer Akkumulation in der Zelle führen würde. Um dies auszuschließen, wurden die Mycelpellets nach der letzten Probennahme mit Methanol extrahiert (90 min) und die Extrakte ebenfalls vermessen. Allerdings konnte auch hier keine Labyrinthopeptinproduktion festgestellt werden.

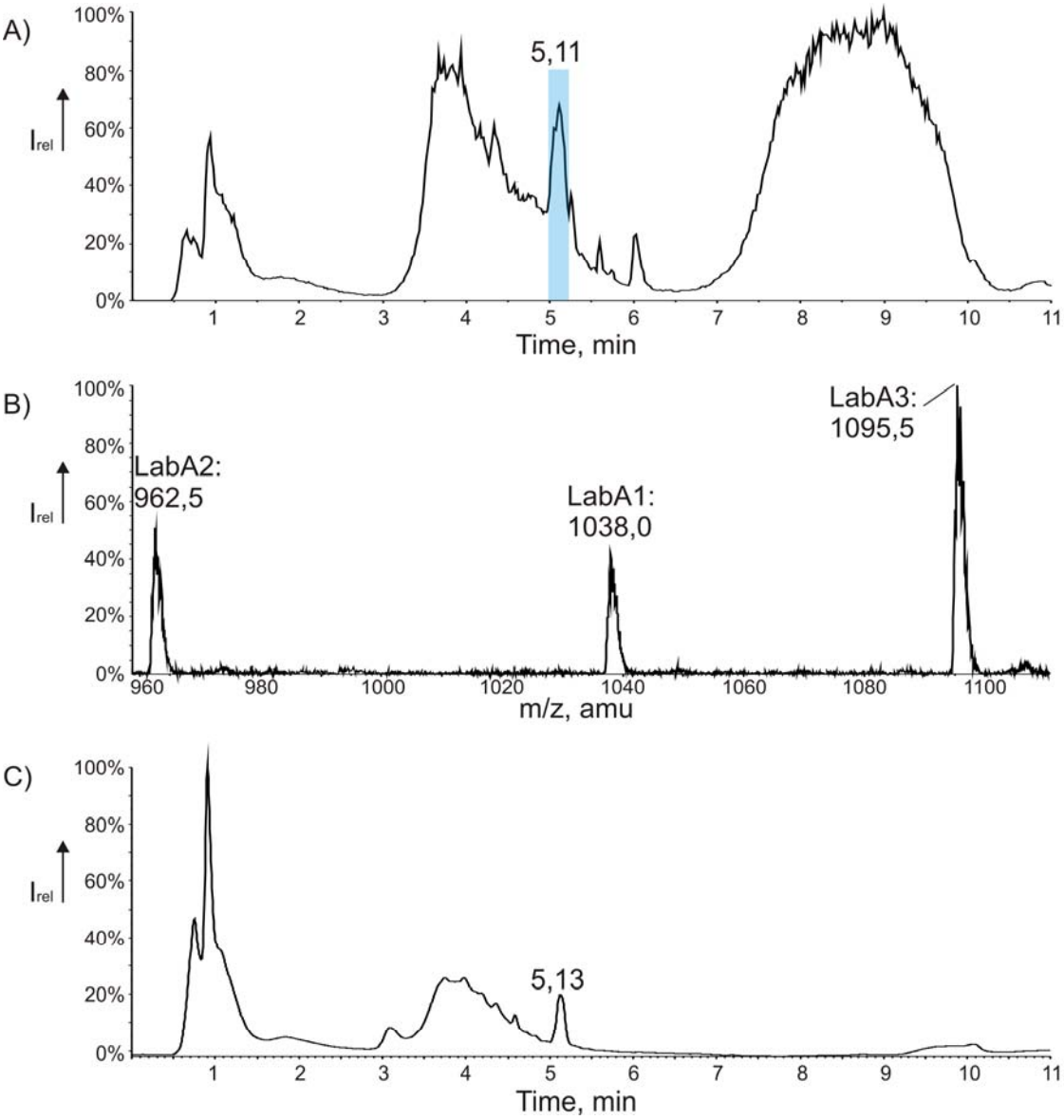


Abbildung 72: Beispiel der Auswertung von HPLC-ESI-MS-Daten zur Labyrinthopeptin-Produktion. Der Kulturüberstand von *Actinomadura namibiensis* (72 h Kultivierung; Medium 5294) zeigte A) im TIC (*total ion chromatogram*) bei 5,11 min einen Peak (blau), der aufsummiert B) im Spektrum die Molekülmassen der drei Labyrithopeptine zeigte (LabA1: $[M+2H]^{2+}$ = 1037,85; LabA2: $[M+2H]^{2+}$ = 962,35; LabA3: $[M+2H]^{2+}$ = 1095,35). C) Im TWC (*total wavelength chromatogram*) ist zur selben Retentionszeit ebenfalls ein deutlicher Peak zu beobachten.

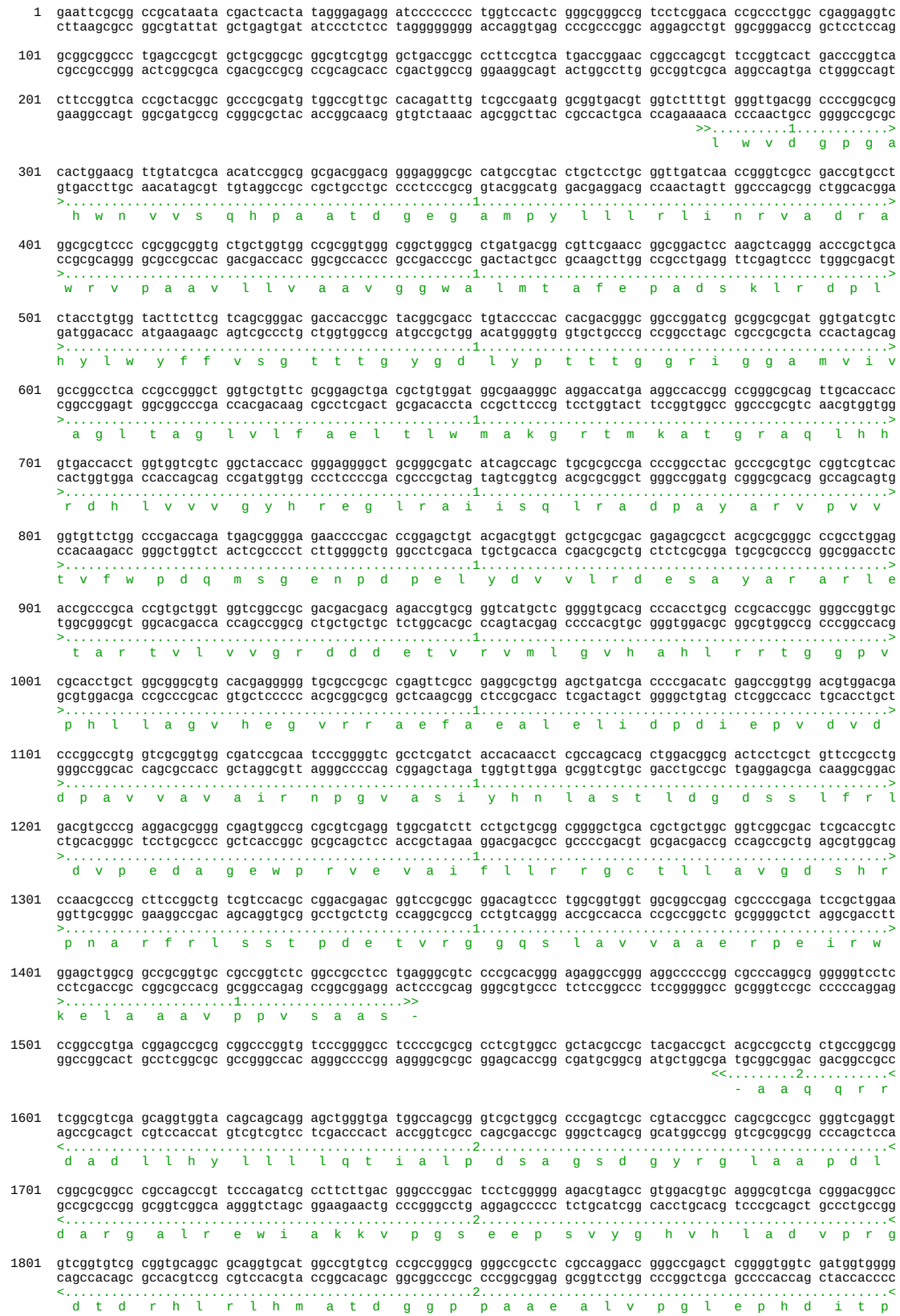
In zukünfigen Experimenten muss diesem Phänomen weiter nachgegangen werden, wobei Untersuchungen auf mRNA-Ebene geplant sind, die die Transkription der eingebrachten Gene überprüfen sollen.
Unter Umständen besteht allerdings auch die Möglichkeit, dass lediglich eine ungünstige Medienwahl bzw. Kultivierungsbedingungen zur Nicht-Produktion geführt haben, weshalb diese weiter optimiert werden müssen. Außerdem soll die Transformation und heterologe Expression in anderen Streptomyceten versucht werden, bevorzugt in den in Abbildung 56 beschriebenen Stämmen, die ähnlich aufgebaute Gencluster tragen und somit möglicherweise fehlende Funktionen kompensieren können.

6 Ausblick

In dieser Arbeit konnte ein das Labyrinthopeptin-Biosynthese-Gencluster tragendes Cosmid gefunden und sequenziert werden. Die Auswertung der Sequenz konnte fünf Gene zu diesem Biosynthese-Gencluster zuordnen, wobei es sich um zwei Strukturgene, zwei Transportergene und ein Gen für ein Protein mit einer *N*-terminalen Kinase-Domäne und einer *C*-terminalen LanC-Domäne handelte. Dieses Gen ist vermutlich für die Einführung der neuartigen Labionin-Verbrückungen zuständig, was mittels *in vitro*-Assay und heterologer Expression des gefundenen Genclusters nachgewiesen werden soll. Die in der vorliegenden Arbeit erzielten Ergebnisse bezüglich dieser zwei Strategien sollen in künftigen Arbeiten fortgeführt werden. Beim *in vitro*-Assay muss der Fokus auf der Optimierung der Substratbereitstellung liegen, um angemessene Konzentrationen zu erreichen. Die Integration des gefundenen Genclusters in das Genom von *S. lividans* konnte mittels PCR und *Southern-Blot* gezeigt werden, allerdings konnte bisher keine Labyrinthopeptin-produktion nachgewiesen werden. Da in Genomen von *S. coelicolor*, *S. griseus, S. avermitilis* und *Saccharopolyspora erythraea* ähnliche Clusteranordnungen gefunden wurden, ist momentan der vielversprechendste Ansatz, eine heterologe Expression in diesen Stämmen zu versuchen. Dabei könnten im Cluster fehlende Funktionen eventuell durch den Stamm komplementiert werden, so dass eine Labyrinthopeptinproduktion ermöglicht wird. Ziel künftiger Arbeiten muss auf jeden Fall sein, ein für kombinatorische Biosyntheseexperimente zugängliches System zu erhalten. Dabei kann es sich entweder um das bereits oben erwähnte heterologe System handeln oder es muss ein Weg gefunden werden, um den Wildtyp-Produzenten *A. namibiensis* genetisch zugänglich zu machen. An beiden Strategien wird momentan gearbeitet, wobei sich die bisherigen Arbeiten mit dem Wildtyp-Produzenten als sehr schwierig erwiesen haben. Kombinatorische Biosyntheseexperimente sind in diesem Zusammenhang deshalb so interessant, da die Firma Sanofi-Aventis, Frankfurt, in weiteren Aktivitätsuntersuchungen auf eine sehr interessante Wirkung der Labyrinthopeptine gestoßen ist, was weitreichende Struktur-Aktivitäts-Studien erfodert. Aufgrund der Komplexität der Labyrinthopeptin-Struktur ist eine chemisch-synthetische Herstellung von Derivaten äußerst schwierig und erfordert deshalb zumindest für pharmazeutische Anwendungen eine biotechnologische Herangehensweise.

7 DNA und Proteinsequenzen

7.1 Sequenz Cosmid1104



```
1901  cgcagcagca gctcggccga cagcggcgtg tccggcgggt cggcgcggcc cgcgaccggc gcgaaccgga cgacgatgtc ggcccagcgg cgctgcgggc
      gcgtcgtcgt cgagccggct gtcgccgcac aggccgccca gccgcgccgg gcgctggccg cgcttggcct gctgctacag ccgggtcgcc gcgacgcccg
      <.....................................................2.....................................................<
       r  l  l  l  e  a  s  l  p  t  d  p  p  d  a  r  g  a  v  p  a  f  r  v  v  i  d  a  w  r  r  q  p

2001  ggatgtaggc ggcgctctcg ggttcgcggc gggccagctc ggcgcggacc tgcccggggg tgtagccgcg gttggcgcag tcgcggcgga ccttccaggc
      cctacatccg ccgcgagagc ccaagcgccg cccggtcgag ccgcgcctgg acgggccccc acatcggcgc caaccgcgtc agcgccgcct ggaaggtccg
      <.....................................................2.....................................................<
      r  i  y  a  a  s  e  p  e  r  r  a  l  e  a  r  v  q  g  p  t  y  g  r  n  a  c  d  r  r  v  k  w  a

2101  gtggcgcagc ggctcgggcg ggtccaggta gacggtgatg tcgaagcagg cgcgcgccag ccgggtgtgc agcggcagca gcccctcgac gatcacgaag
      caccgcgtcg ccgagcccgc ccaggtccat ctgccactac agcttcgtcc gcgcgcggtc ggcccacacg tcgccgtcgt cggggagctg ctagtgcttc
      <.....................................................2.....................................................<
        h  r  l  p  e  p  p  d  l  y  v  t  i  d  f  c  a  r  a  l  r  t  h  l  p  l  l  g  e  v  i  v  f

2201  tcgcggggcc ggaccagctc gggccggacg agccggccgg tgccgtggtc gtagaccggt ttgaggatcg gctcgcccag cgacagcagt tgcaggtgct
      agcgccccgg cctggtcgag cccggcctgc tcggccggcc acggcaccag catctggcca aactcctagc cgagcgggtc gctgtcgtca acgtccacga
      <.....................................................2.....................................................<
       d  r  p  r  v  l  e  p  r  v  l  r  g  t  g  h  d  y  v  p  k  l  i  p  e  g  l  s  l  l  q  l  h

2301  gctccatgat gtcgacgtgg ttgcagtcgg ggtgcagcgc cgtgaacggc ctgcccgccc gttcccgccg gtcgaagcgg tggtagtcgt ccacgcagac
      cgaggtacta cagctgcacc aacgtcagcc ccacgtcgcg gcacttgccg gacgggcggg caagggcggc cagcttcgcc accatcagca ggtgcgtctg
      <.....................................................2.....................................................<
      q  e  m  i  d  v  h  n  c  d  p  h  l  a  t  f  p  r  g  a  r  e  r  r  d  f  r  h  y  d  d  v  c  v

2401  ggcggtcatc cggtccgccc ccagacaccg caccagcccc ctggtgagcg tcgtcttgcc cgccgcgctg tccccggcga tggcgagcat gacggggcgg
      ccgccagtag gccaggcggg ggtctgtggc gtggtcgggg gaccactcgc agcagaacgg gcggcgcgac aggggccgct accgctcgta ctgccccgcc
      <.....................................................2.....................................................<
        a  t  m  r  d  a  g  l  c  r  v  l  g  r  t  l  t  t  k  g  a  a  s  d  g  a  i  a  l  m  v  p  r

2501  cggcccgatc cgcgcgcccg gagcatgtgc acgatccgat ggggcacctt gcctcctagc gtgccgacgt cgtccggtgc ccggacggta gaacgggcgg
      gccgggctag gcgcgcgggc ctcgtacacg tgctaggcta ccccgtggaa cggaggatcg cacggctgca gcaggccacg ggcctgccat cttgcccgcc
      <.....................2.....................<<
       r  g  s  g  r  a  r  l  m  h  v  i  r  h  p  v

2601  cgggccgcgg cctcacccga acgggggaga gcacgggtga aacccgggag aagcctgacg ccacgggggc tactggcggg tagcaagggt tgttatcgtg
      gcccggcgcc ggagtgggct tgccccctct cgtgcccact ttgggccctc ttcggactgc ggtgcccccg atgaccgccc atcgttccca acaatagcac

2701  ccgggatggg cgtgcccgta cgcgtcgtcc tggtggacga ccacgagatg atcctggccg ggctgaaggc gatgctggcc gggttcgccg ggcgggtgcg
      ggccctaccc gcacgggcat gcgcagcagg accacctgct ggtgctctac taggaccggc ccgacttccg ctacgaccgg cccaagcggc ccgcccacgc
           >>.....................................................3.....................................................>
             m  g  v  p  v  r  v  v  l  v  d  d  h  e  m  i  l  a  g  l  k  a  m  l  a  g  f  a  g  r  v

2801  ggtggtgggg caggcgggga cgtccgccga cgccgcccgg ctcgtggcgg cgctgcggcc cgacgtggtg ctgctggacg tccggctggg gccggagagc
      ccaccacccc gtccgccccct gcaggcggct gcggcgggcc gagcaccgcc gcgacgccgg gctgcaccac gacgacctgc aggccgaccc cggcctctcg
      >.....................................................3.....................................................>
      r  v  v  g  q  a  g  t  s  a  d  a  a  r  l  v  a  a  l  r  p  d  v  v  l  l  d  v  r  l  g  p  e  s

2901  ggcctggacg tgtgccggac gctgaccggc ctggcccccg agtcgcgggt ggtgttcctg tcggtctacg acgacgagca gtacgtcttc gaggcgctgc
      ccggacctgc acacggcctg cgactggccg gaccgggggc tcagcgccca ccacaaggac agccagatgc tgctgctcgt catgcagaag ctccgcgacg
      >.....................................................3.....................................................>
        g  l  d  v  c  r  t  l  t  g  l  a  p  e  s  r  v  v  f  l  s  v  y  d  d  e  q  y  v  f  e  a  l

3001  gtgcgggcgc cggcggctac ctgctgaaac gcgtcgacgg accggagctg gtgcggcggc tggaggaggt cgcccgcggc gagaccgtcg tcgatcccac
      cacgcccgcg gccgccgatg gacgactttg cgcagctgcc tggcctcgac cacgccgccg acctcctcca gcgggcgccg ctctggcagc agctagggtg
      >.....................................................3.....................................................>
       r  a  g  a  g  g  y  l  l  k  r  v  d  g  p  e  l  v  r  r  l  e  e  v  a  r  g  e  t  v  v  d  p

3101  cctcgccggg cggatggcgg tgacggcggc gcggctgcgg cgcggcgagt tctggccggg cgccgaccgg ggcctcaccc agcgcgagag cgaggtgctg
      ggagcggccc gcctaccgcc actgccgccg cgccgacgcc gcgccgctca agaccggccc gcggctggcc ccggagtggg tcgcgctctc gctccacgac
      >.....................................................3.....................................................>
      t  l  a  g  r  m  a  v  t  a  a  r  l  r  r  g  e  f  w  p  g  a  d  r  g  l  t  q  r  e  s  e  v  l

3201  tcgctgctgg tcggcgggtt gtcgaacaag gcgatcgcgg cgcggctggt ggtcagcgag gagaccgtca agagccatct gcggaccctg taccggaagc
      agcgacgacc agccgcccaa cagcttgttc cgctagcgcc gcgccgacca ccagtcgctc ctctggcagt tctcggtaga cgcctgggac atggccttcg
      >.....................................................3.....................................................>
        s  l  l  v  g  g  l  s  n  k  a  i  a  a  r  l  v  v  s  e  e  t  v  k  s  h  l  r  t  l  y  r  k

3301  tggaggtggc cgaccggtcc gcggcggtcg cgctggcgct gcgggagggg ctgtaccggt gacggccgcg ctggacctgc tggcgggccg gggacgggac
      acctccaccg gctggccagg cgccgccagc gcgaccgcga cgccctcccc gacatggcca ctgccggcgc gacctggacg accgcccggc ccctgccctg
      >.....................................3.............................>>
       l  e  v  a  d  r  s  a  a  v  a  l  a  l  r  e  g  l  y  r  -
                                                        >>.....................4.....................>
                                                          v  t  a  a  l  d  l  l  a  g  r  g  r  d

3401  ctgctggtcc gggtcgtcgc ggtcgcgtgc tcggggcgga cgctgccgtc gatggccgag gagctggccg ggctggtggt gcggtcgggc gaggcgctcg
      gacgaccagg cccagcagcg ccagcgcacg agccccgcct gcgacggcag ctaccggctc ctcgaccggc ccgaccacca cgccagcccg ctccgcgagc
      >.....................................................4.....................................................>
        l  l  v  r  v  v  a  v  a  c  s  g  r  t  l  p  s  m  a  e  e  l  a  g  l  v  v  r  s  g  e  a  l

3501  tgttctgcga cgcgtacgtc ctggacgagt gcgggcgggt catggagggc gcgggccggg ccgtcccgct gggggagggg gacgtcgggt gggccgcggc
      acaagacgct gcgcatgcag gacctgctca cgcccgccca gtacctcccg cgcccggccc ggcagggcga ccccctcccc ctgcagccca cccggcgccg
      >.....................................................4.....................................................>
       v  f  c  d  a  y  v  l  d  e  c  g  r  v  m  e  g  a  g  r  a  v  p  l  g  e  g  d  v  g  w  a  a

3601  gcacgggcgg gcccggcggc accgcgacgg gcgcggcgtc gcggtgcccg tgcacagcgg gcccgcggtc atcggggtcc tcgacgtccg gtcggcgggc
      cgtgcccgcc cgggccgccg tggcgctgcc cgcgccgcag cgccacgggc acgtgtcgcc cgggcgccag tagccccagg agctgcaggc cagccgcccg
      >.....................................................4.....................................................>
      a  h  g  r  a  r  r  h  r  d  g  r  g  v  a  v  p  v  h  s  g  p  a  v  i  g  v  l  d  v  r  s  a  g

3701  ccctgccgcg agccggacgt ggcactggtc ggggcgctgg ccgagctgtt cgcgccggtg ctgtgctcgt gccgccgcct gcgcaccgcc cgggagcgcg
      gggacggcgc tcggcctgca ccgtgaccag ccccgcgacc ggctcgacaa gcgcggccac gacacgagca cggcggcgga cgcgtggcgg gccctcgcgc
      >.....................................................4.....................................................>
        p  c  r  e  p  d  v  a  l  v  g  a  l  a  e  l  f  a  p  v  l  c  s  c  r  r  l  r  t  a  r  e  r

3801  agttcgccgc cgagcgcttc gccgaacgcg ccgtccaggt ccaggaggcc gagcggtccc ggctcagccg cgagatccac gacgggatcg cccagcgcct
      tcaagcggcg gctcgcgaag cggcttgcgc ggcaggtcca ggtcctccgg ctcgccaggg ccgagtcggc gctctaggtg ctgccctagc gggtcgcgga
      >.....................................................4.....................................................>
       e  f  a  a  e  r  f  a  e  r  a  v  q  v  q  e  a  e  r  s  r  l  s  r  e  i  h  d  g  i  a  q  r

3901  ggccagcctg ggcttccacc tgtcggcggc gcaggcggcg ctggagggcc ccgggctgga cgggcccgcg ccgtccggcg cccgcgccca gatcgcccgc
      ccggtcggac ccgaaggtgg acagccgccg cgtccgccgc gacctcccgg ggcccgacct gcccgggcgc ggcaggccgc gggcgcgggt ctagcgggcg
      >.....................................................4.....................................................>
      l  a  s  l  g  f  h  l  s  a  a  q  a  a  l  e  g  p  g  l  d  g  p  a  p  s  g  a  r  a  q  i  a  r
```

```
4001  gcgcgcgagc tgtgcgagct ggccgccgcc gagacgcggg cggcgatcgg cgggctgcgg cccccggtcc tggacgacct ggggctgccc gccgcgctgg
      cgcgcgctcg acacgctcga ccggcggcgg ctctgcgccc gccgctagcc gcccgacgcc gggggccagg acctgctgga ccccgacggg cggcgcgacc
      >.....................................................4.....................................................>
        a  r  e  l  c  e  l  a  a  a  e  t  r  a  a  i  g  g  l  r  p  p  v  l  d  d  l  g  l  p  a  a  l

4101  cgacgctggc gcgggaggcg ggcgcgccca cctcgcgcgg ggcggtggac gtgaccgtca ccgtgaccgg ggagctggag gagccgctgc ccgaccacgt
      gctgcgaccg cgccctccgc ccgcgcgggt ggagcgcgcc ccgccacctg cactggcagt ggcactggcc cctcgacctc ctcggcgacg ggctggtgca
      >.....................................................4.....................................................>
       a  t  l  a  r  e  a  g  a  p  t  s  r  g  a  v  d  v  t  v  t  v  t  g  e  l  e  e  p  l  p  d  h

4201  gcagaccgcg ctgtaccgga tcgcgcagga ggcggtcggc aactgcctgc ggcacgccgc cgccacctgc gtcgacctgc tgctggagca cgaccgcgac
      cgtctggcgc gacatggcct agcgcgtcct ccgccagccg ttgacggacg ccgtgcggcg gcggtggacg cagctggacg acgacctcgt gctggcgctg
      >.....................................................4.....................................................>
      v  q  t  a  l  y  r  i  a  q  e  a  v  g  n  c  l  r  h  a  a  a  t  c  v  d  l  l  l  e  h  d  r  d

4301  cgcgtccggc tgaccgtcgc cgacgacggg gtggggttcg cgccgcggca gggccgccgg tccggggcct acgggctgcg cggcatggcc gaacgcgccg
      gcgcaggccg actggcagcg gctgctgccc caccccaagc gcggcgccgt cccggcggcc aggccccgga tgcccgacgc gccgtaccgg cttgcgcggc
      >.....................................................4.....................................................>
        r  v  r  l  t  v  a  d  d  g  v  g  f  a  p  r  q  g  r  r  s  g  a  y  g  l  r  g  m  a  e  r  a

4401  agctgctggg cggccggttc acggtcacca gccgccccgg cgccggcacc accgtcgagg cggtcgtccc gatcaggaat ggatcacccg gaaggtgatc
      tcgacgaccc gccggccaag tgccagtggt cggcggggcc gcggccgtgg tggcagctcc gccagcaggg ctagtcctta cctagtgggc cttccactag
      >.....................................................4.....................................................>>
       e  l  l  g  g  r  f  t  v  t  s  r  p  g  a  g  t  t  v  e  a  v  v  p  i  r  n  g  s  p  g  r  -
                                                                                                  <<.....5.....<
                                                                                                    -  s  h  i  v  r  f  t  i

4501  ccggccttct ggaggcggtc gatcagcgcg tcgcccatcg cgaccgccgt ggtgacctgg ccggagctcg gcggcaggtc gtcgaaggcc aggcacagcg
      ggccggaaga cctccgccag ctagtcgcgc agcgggtagc gctggcggca ccactggacc ggcctcgagc cgccgtccag cagcttccgg tccgtgtcgc
      <.....................................................5.....................................................<
       g  a  k  q  l  r  d  i  l  a  d  g  m  a  v  a  t  t  v  q  g  s  s  p  p  l  d  d  f  a  l  c  l

4601  ccgactcggc cagcatcttc gcggtctcgc cgtagccggg gtcgccgccc gcgacctcgg tgaccacgcg cctgccgccg ccctcgccga cgaacttcac
      ggctgagccg gtcgtagaag cgccagagcg gcatcggccc cagcggcggg cgctggagcc actggtgcgc ggacggcggc gggagcggct gcttgaagtg
      <.....................................................5.....................................................<
      a  s  e  a  l  m  k  a  t  e  g  y  g  p  d  g  g  a  v  e  t  v  v  r  r  g  g  g  e  g  v  f  k  v

4701  gctgaaccac gccttggcgc gcttctccgg cggcgggccc tcgcccgccc cgcgcacccg cagcagcagc gcgcgggtcg gcggcagctg cgccagcgcc
      cgacttggtg cggaaccgcg cgaagaggcc gccgcccggg agcgggcggg gcgcgtgggc gtcgtcgtcg cgcgcccagc cgccgtcgac gcggtcgcgg
      <.....................................................5.....................................................<
        s  f  w  a  k  a  r  k  e  p  p  p  g  e  g  a  g  r  v  r  l  l  l  a  r  t  p  p  l  q  a  l  a

4801  gtcacggccg tggcgccgac ggccagcccg accgccgacg gcagccgcct ggtcgccacg taggggggcgt aggagaagtc cgggccgtac cggtccagcg
      cagtgccggc accgcggctg ccggtcgggc tggcggctgc cgtcggcgga ccagcggtgc atcccccgca tcctcttcag gcccggcatg gccaggtcgc
      <.....................................................5.....................................................<
       t  v  a  t  a  g  v  a  l  g  v  a  s  p  l  r  r  t  a  v  y  p  a  y  s  f  d  p  g  y  r  d  l

4901  ccgccgccga ccgcagcacg atctgcgggt cgatcgtcgg catcggcagc gtccagccgt tcacccccga gcggggcacg ggccgccgcg agatccgcac
      ggcggcggct ggcgtcgtgc tagacgccca gctagcagcc gtagccgtcg caggtcggca agtgggggct cgccccgtgc ccggcggcgc tctaggcgtg
      <.....................................................5.....................................................<
      a  a  a  s  r  l  v  i  q  p  d  i  t  p  m  p  l  t  w  g  n  v  g  s  r  p  v  p  r  r  s  i  r  v

5001  ccgccgtccc tccggcttcg gcgcggcgcg gtggtactcg cgctcggagc ggaccatgcc ggggatgtcg gcgaagatcc cgacggccga gtgcagggtg
      ggcggcaggg aggccgaagc cgcgccgcgc caccatgagc gcgagcctcg cctggtacgg cccctacagc cgcttctagg gctgccggct cacgtcccac
      <.....................................................5.....................................................<
        r  r  g  e  p  k  p  a  a  r  h  y  e  r  e  s  r  v  m  g  p  i  d  a  f  i  g  v  a  s  h  l  t

5101  ccgccggaga agtcgcccga gccccgcacg aacccctcca cgtgcagcgg aacgcccgcc gggagccgct tgacggtgaa gtacgcgccc aggtcctgcg
      ggcggcctct tcagcgggct cggggcgtgc ttggggaggt gcacgtcgcc ttgcgggcgg ccctcggcga actgccactt catgcgcggg tccaggacgc
      <.....................................................5.....................................................<
       g  g  s  f  d  g  s  g  r  v  f  g  e  v  h  l  p  v  g  a  p  l  r  k  v  t  f  y  a  g  l  d  q

5201  ggatcgagtc gaacccgcag gcgtgcacga tccgcgcgcc cgtgcgcacg gcctcctcgt ggtggcgcac gtacatccgg ttcacgaacg tcggctcgcc
      cctagctcag cttgggcgtc cgcacgtgct aggcgcgcgg gcacgcgtgc cggaggagca ccaccgcgtg catgtaggcc aagtgcttgc agccgagcgg
      <.....................................................5.....................................................<
      p  i  s  d  f  g  c  a  h  v  i  r  a  g  t  r  v  a  e  e  h  h  r  v  y  m  r  n  v  f  t  p  e  g

5301  ggtcaggtcc acgtagtcgg tcccgttgcg ggcgcacgcg gcgaccagcg gctcgccgtg cctcacgtac gggccgacgg tggtgatgac gacgcgcgtc
      ccagtccagg tgcatcagcc agggcaacgc ccgcgtgcgc cgctggtcgc cgagcggcac ggagtgcatg cccggctgcc accactactg ctgcgcgcag
      <.....................................................5.....................................................<
        t  l  d  v  y  d  t  g  n  r  a  c  a  a  v  l  p  e  g  h  r  v  y  p  g  v  t  t  i  v  v  r  t

5401  cgcgcggcca gctcctcgac cgacgcggtg tcggcggtgt cggcgtgcag cagcggcagg tcggcgcagg ccgggtcgat cgcggccagc cggtcgcgga
      gcgcgccggt cgaggagctg gctgcgccac agccgccaca gccgcacgtc gtcgccgtcc agccgcgtcc ggcccagcta gcgccggtcg gccagcgcct
      <.....................................................5.....................................................<
       r  a  a  l  e  e  v  s  a  t  d  a  t  d  a  h  l  l  p  l  d  a  c  a  p  d  i  a  a  l  r  d  r

5501  ccgccgccag cctctcctgg ttgcgcccgg ccagcgccca cttggcgaag ggcggcgcgt gccgcgccag gtactcggcg gtcagcccgc cggtgaagcc
      ggcggcggtc ggagaggacc aacgcgggcc ggtcgcgggt gaaccgcttc ccgccgcgca cggcgcggtc catgagccgc cagtcgggcg gccacttcgg
      <.....................................................5.....................................................<
      v  a  a  l  r  e  q  n  r  g  a  l  a  w  k  a  f  p  p  a  h  r  a  l  y  e  a  t  l  g  g  t  f  g

5601  ggtggcgccg aagaggacga ggtcgtacgg gcgttcggcg gtcatcgtgg acccctttcg cgtggtcgcg gatgcgggcg ccctgccccc agggttttcc
      ccaccgcggc ttctcctgct ccagcatgcc cgcaagccgc cagtagcacc tggggaaagc gcaccagcgc ctacgcccgc gggacggggg tcccaaaagg
      <.....................................................5.....................................................<
        t  a  g  f  l  v  l  d  y  p  r  e  a  t  m  t  s  g  k  r  t  t  a  s  a  p  a  r  g  g  p  n  e

5701  gaatccgatt ctgtcgtgtc gtggccgctc cggcgcgctc cggtgcgcga tgtgaccgac cccacgcgcc cggcggggcg tcagcggggc cgggaccagc
      cttaggctaa gacagcacag caccggcgag gccgcgcgag gccacgcgct acactggctg gggtgcgcgg gccgccccgc agtcgccccg gccctggtcg
      <.................................5.................................<<
       s  d  s  e  t  t  d  h  g  s  r  r  a  g  t  r  s  t  v  s  g  v
                                                                                                  <<........6.........<
                                                                                                    -  r  p  r  s  w

5801  ccggggcctc caggagatga tccacgaaca gccgggtcgt cggggtcgga tcgggcccgc ccacggcgtg ccaggggcgg tgcagcggcg ttccgggcgc
      ggccccggag gtcctctact aggtgcttgt cggcccagca gccccagcct agcccgggcg ggtgccgcac ggtccccgcc acgtcgccgc aaggcccgcg
      <.....................................................6.....................................................<
      g  p  a  e  l  l  h  d  v  f  l  r  t  t  p  t  p  d  p  g  g  v  a  h  w  p  r  h  l  p  t  g  p  a

5901  gtcgatctcg accagggagc cgtccgccag gtggccgcgg accgcgtcac gggacagcag ggcgacgccc agccccgccg ccgcgcccgc gaccaccgcg
      cagctagagc tggtccctcg gcaggcggtc caccggcgcc tggcgcagtg ccctgtcgtc ccgctgcggg tcggggcggc ggcgcgggcg ctggtggcgc
      <.....................................................6.....................................................<
        d  i  e  v  l  s  g  d  a  l  h  g  r  v  a  d  r  s  l  l  a  v  g  l  g  a  a  a  g  a  v  v  a
```

```
6001  ccgttggagc cgaggaccag gcggggcggg gcgacgtcgc gttcggccag gtacgcctcg gccgtcgcgc gggtgcccga gcccggctcg cgcatcaccc
      ggcaacctcg gctcctggtc cgccccgccc cgctgcagcg caagccggtc catgcggagc cggcagcgcg cccacgggct cgggccgagc gcgtagtggg
      <.....................................................6.....................................................<
       g  n  s  g  l  v  l  r  p  p  a  v  d  r  e  a  l  y  a  e  a  t  a  r  t  g  s  g  p  e  r  m  v

6101  acggcgtcgt ggccgggtcg aacgccgccg ccggcgcggg cgcgcccacc acgaccaggt cgttgggccg ccgcgccagc acggtcgccg cgatgccggc
      tgccgcagca ccggcccagc ttgcggcggc ggccgcgccc gcgcgggtgg tgctggtcca gcaacccggc ggcgcggtcg tgccagcggc gctacggccg
      <.....................................................6.....................................................<
      w  p  t  t  a  p  d  f  a  a  a  p  a  p  a  g  v  v  v  l  d  n  p  r  r  a  l  v  t  a  a  i  g  a

6201  gggcggccgt ccggccagca ccagatcggc ccggtgggcg gccaggctgc tccacacctg gcggcgcggc cccacctcca gggccaggtc cacctccggc
      cccgccggca ggccggtcgt ggtctagccg ggccacccgc cggtccgacg aggtgtggac cgccgcgccg gggtggaggt cccggtccag gtggaggccg
      <.....................................................6.....................................................<
        p  p  r  g  a  l  v  l  d  a  r  h  a  a  l  s  s  w  v  q  r  r  p  g  v  e  l  a  l  d  v  e  p

6301  cagcgggcgc ggaacgaggc cagcagctcc ggcaggacct ggtcggcggc cgtggtcacc gccgccaggc gcagcgggcc gcgcgccggg tcggccgcgc
      gtcgcccgcg ccttgctccg gtcgtcgagg ccgtcctgga ccagccgccg gcaccagtgg cggcggtccg cgtcgcccgg cgcgcggccc agccggcgcg
      <.....................................................6.....................................................<
       w  r  a  r  f  s  a  l  l  e  p  l  v  q  d  a  a  t  t  v  a  a  l  r  l  p  g  r  a  p  d  a  a

6401  cccgcgccgc cgcgcgcccc tgctccagca ggccgaggac ccggcgggcg taggaggcgt agaccgtccc ggcgggggtg agcgccaggc cgcgcccctc
      gggcgcggcg gcgcgcgggg acgaggtcgt ccggctcctg ggccgcccgc atcctccgca tctggcaggg ccgcccccac tcgcggtccg gcgcggggag
      <.....................................................6.....................................................<
      g  r  a  a  a  r  g  q  e  l  l  g  l  v  r  r  a  y  s  a  y  v  t  g  a  p  t  l  a  l  g  r  g  e

6501  gcggcggacc agcgcgacgc cgaggtcccg ggccagcgcc gccagcgccg ccgacaccgc cgactcggtc acgtacagcc gccgcgccgc ctcccgcacc
      cgccgcctgg tcgcgctgcg gctccagggc ccggtcgcgg cggtcgcggc ggctgtggcg gctgagccag tgcatgtcgg cggcgcggcg gagggcgtgg
      <.....................................................6.....................................................<
        r  r  v  l  a  v  g  l  d  r  a  l  a  a  l  a  s  v  a  s  e  t  v  y  l  r  r  a  a  e  r  v

6601  gacccggcgt cggccaccgc cacgaacgtc cgcagccgcg cctcggtcac ggcgacctcc tggttcaagc cctcgctcag gtgtgggccg gaaacacttg
      ctgggccgca gccggtggcg gtgcttgcag gcgtcggcgc ggagccagtg ccgctggagg accaagttcg ggagcgagtc cacacccggc ctttgtgaac
      <.....................................6.....................................<<
       s  g  a  d  a  v  a  v  f  t  r  l  r  a  e  t  v  a  v  e  q  n  l

6701  atggacgatg cgagtatcgc cgggcgacgc tcgggacgtc cacgaacccc gacccgagcg gagggccatg agcgcgggca gatggtcggc gggtgtgatc
      tacctgctac gctcatagcg gcccgctgcg agccctgcag gtgcttgggg ctgggctcgc ctcccggtac tcgcgcccgt ctaccagccg cccacactag
            >>.....................................................7.....................................................>
              m  r  v  s  p  g  d  a  r  d  v  h  e  p  r  p  e  r  r  a  m  s  a  g  r  w  s  a  g  v  i

6801  ccgtacgcgg agatgggcta ctggcggccc gactaccggc cgaaggacgg cgacgtgctg gccgcgttcc ggatcacgcc gcagccgggc gtcccgccgg
      ggcatgcgcc tctacccgat gaccgccggg ctgatggccg gcttcctgcc gctgcacgac cggcgcaagg cctagtgcgg cgtcggcccg cagggcggcc
      >.....................................................7.....................................................>
        p  y  a  e  m  g  y  w  r  p  d  y  r  p  k  d  g  d  v  l  a  a  f  r  i  t  p  q  p  g  v  p  p

6901  aggaggcggg cgcggccgtc gccggggagt cgtccaccgc cacctggacg gtggtgtgga ccgaccggct cacctcccac gagaactacc aggccaagtg
      tcctccgccc gcgccggcag cggcccctca gcaggtggcg gtggacctgc caccacacct ggctggccga gtggagggtg ctcttgatgg tccggttcac
      >.....................................................7.....................................................>
       e  e  a  g  a  a  v  a  g  e  s  s  t  a  t  w  t  v  v  w  t  d  r  l  t  s  h  e  n  y  q  a  k

7001  ctaccgggtg gaggcggtgc cggggcgcga ggggcagttc atcgcctcca tcgcctacga cctggacctg ttcgaggagg gctcgatcgc caacctcacc
      gatggcccac ctccgccacg gccccgcgct ccccgtcaag tagcggaggt agcggatgct ggacctggac aagctcctcc cgagctagcg gttggagtgg
      >.....................................................7.....................................................>
      c  y  r  v  e  a  v  p  g  r  e  g  q  f  i  a  s  i  a  y  d  l  d  l  f  e  e  g  s  i  a  n  l  t

7101  tcctcgatca tcgggaacgt cttcgggttc aaggcgctgc gggcgctgcg gctggaggac atgcggatcc cgccgcacta cgtcaagacg ttccagggcc
      aggagctagt agcccttgca gaagcccaag ttccgcgacg cccgcgacgc cgacctcctg tacgcctagg gcggcgtgat gcagttctgc aaggtcccgg
      >.....................................................7.....................................................>
        s  s  i  i  g  n  v  f  g  f  k  a  l  r  a  l  r  l  e  d  m  r  i  p  p  h  y  v  k  t  f  q  g

7201  cggcgcacgg gatcgtgatg gagcgcgagt acctcggcaa gttcgggcgc ccgctgctgg gcgcgacggt caagccgaaa ctcgggctgt cggcgcgcaa
      gccgcgtgcc ctagcactac ctcgcgctca tggagccgtt caagcccgcg ggcgacgacc cgcgctgcca gttcggcttt gagcccgaca gccgcgcgtt
      >.....................................................7.....................................................>
       p  a  h  g  i  v  m  e  r  e  y  l  g  k  f  g  r  p  l  l  g  a  t  v  k  p  k  l  g  l  s  a  r

7301  ctacgggcgc gtcgtgtacg aggcgctgcg cggcggcctg gacttcacca aggacgacga gaacatcaac tcccagccgt tcatgcgctg gcgcgaccgg
      gatgcccgcg cagcacatgc tccgcgacgc gccgccggac ctgaagtggt tcctgctgct cttgtagttg agggtcggca agtacgcgac cgcgctggcc
      >.....................................................7.....................................................>
      n  y  g  r  v  v  y  e  a  l  r  g  g  l  d  f  t  k  d  d  e  n  i  n  s  q  p  f  m  r  w  r  d  r

7401  ttcctgtact gcatggaggg cgtcaaccgg gccgtcgccg cgaccggcga ggtcaagggg cactacctca acgtcaccgc cgccacgatg gaggacatgt
      aaggacatga cgtacctccc gcagttggcc cggcagcggc gctggccgct ccagttcccc gtgatggagt tgcagtggcg gcggtgctac ctcctgtaca
      >.....................................................7.....................................................>
        f  l  y  c  m  e  g  v  n  r  a  v  a  a  t  g  e  v  k  g  h  y  l  n  v  t  a  a  t  m  e  d  m

7501  acgagcgcgc cgagttcgcg cgcgacctcg gcagcgtcat cgtgatgatc gacctgaccg tcggctacac cgcgatccag tcgatggccc ggtgggcgcg
      tgctcgcgcg gctcaagcgc gcgctggagc cgtcgcagta gcactactag ctggactggc agccgatgtg gcgctaggtc agctaccggg ccacccgcgc
      >.....................................................7.....................................................>
       y  e  r  a  e  f  a  r  d  l  g  s  v  i  v  m  i  d  l  t  v  g  y  t  a  i  q  s  m  a  r  w  a

7601  ggccaacggg atgctgctgc acctgcaccg cgccgggcac tccacctaca cgcggcagaa gacccacggc gtcagcttcc gcgtgatcgc caagtggatg
      ccggttgccc tacgacgacg tggacgtggc gcggcccgtg aggtggatgt gcgccgtctt ctgggtgccg cagtcgaagg cgcactagcg gttcacctac
      >.....................................................7.....................................................>
      r  a  n  g  m  l  l  h  l  h  r  a  g  h  s  t  y  t  r  q  k  t  h  g  v  s  f  r  v  i  a  k  w  m

7701  cggctggccg gcgtcgacca catccacgcc gggacggtcg tcggcaagct ggagggcgac ccccacagcg tcgcgggctt ctacgacacc ctgcgcctcg
      gccgaccggc cgcagctggt gtaggtgcgg ccctgccagc agccgttcga cctcccgctg ggggtgtcgc agcgcccgaa gatgctgtgg gacgcggagc
      >.....................................................7.....................................................>
        r  l  a  g  v  d  h  i  h  a  g  t  v  v  g  k  l  e  g  d  p  h  s  v  a  g  f  y  d  t  l  r  l

7801  acaggatcgc cgcggatccg gtgaagggcc tgtacttcga ccaggactgg gcgtcgatgc cgggcacgat gccggtggcg tccggcggca tccacgccgg
      tgtcctagcg gcgcctaggc cacttcccgg acatgaagct ggtcctgacc cgcagctacg gcccgtgcta cggccaccgc aggccgccgt aggtgcggcc
      >.....................................................7.....................................................>
       d  r  i  a  a  d  p  v  k  g  l  y  f  d  q  d  w  a  s  m  p  g  t  m  p  v  a  s  g  g  i  h  a

7901  gcagatgcac cagctgctgc accacctggg cgaggacgtg gtcctccagt tcggcggcgg gacgatcggg cacccgatgg ggatcgccgc cggggccgcc
      cgtctacgtg gtcgacgacg tggtggaccc gctcctgcac caggaggtca agccgccgcc ctgctagccc gtgggctacc cctagcggcg gccccggcgg
      >.....................................................7.....................................................>
      g  q  m  h  q  l  l  h  h  l  g  e  d  v  v  l  q  f  g  g  g  t  i  g  h  p  m  g  i  a  a  g  a  a

8001  gccaaccgcg tcgccctcga agcgatgatc aaggcgcgca acgagggccg tgacctcctc gccgagggcc ccgacatcct gcgggccgcc gcccggcaca
      cggttggcgc agcgggagct tcgctactag ttccgcgcgt tgctcccggc actggaggag cggctcccgg ggctgtagga cgcccggcgg cgggccgtgt
      >.....................................................7.....................................................>
        a  n  r  v  a  l  e  a  m  i  k  a  r  n  e  g  r  d  l  l  a  e  g  p  d  i  l  r  a  a  a  r  h
```

```
8101  gccgcgagct ggaggtcgcg ctgtccacct ggggcgacgt gacgttcagc tacgagtcca ccgacacccc cgacgtcgtc gagacccctg tgagcctgtg
      cggcgctcga cctccagcgc gacaggtgga ccccgctgca ctgcaagtcg atgctcaggt ggctgtgggg gctgcagcag ctctggggac actcggacac
      >.....................................................7.....................................................>
       s r e  l e v a  l s t  w g d  v t f s  y e s  t d t  p d v v  e t p  v s l

8201  atgcgcgtga cccaagggac gttctcccac ctgcccgacc tcaccgacga cgagatcgcc gcgcaggtcg cctacgccct ggagcggggc tggccgtgct
      tacgcgcact gggttccctg caagagggtg gacgggctgg agtggctgct gctctagcgg cgcgtccagc ggatgcggga cctcgccccg accggcacga
      > 7
      -
      >>....................................................8.....................................................>
        m r v  t q g  t f s h  l p d  l t d  d e i a  a q v  a y a  l e r g  w p c

8301  cggtggagtt caccgacgac ccgcacccgc gcaacgcgta ctgggagatg tggggcctgc cgatgttcga cctggccgac cccgccgggg tgctgtacga
      gccacctcaa gtggctgctg ggcgtgggcg cgttgcgcat gaccctctac accccggacg gctacaagct ggaccggctg gggcggcccc acgacatgct
      >.....................................................8.....................................................>
       s v e  f t d d  p h p  r n a  y w e m  w g l  p m f  d l a d  p a g  v l y

8401  ggtcaacgag tgccgccgcg cctacccggg ccactacgtc cgggtgaacg cctacgacgc gggctacggg cggcagacga cggcgttgca gttcatcgtg
      ccagttgctc acggcggcgc ggatgggccc ggtgatgcag gcccacttgc ggatgctgcg cccgatgccc gccgtctgct gccgcaacgt caagtagcac
      >.....................................................8.....................................................>
      e v n e  c r r  a y p  g h y v  r v n  a y d  a g y g  r q t  t a l  q f i v

8501  cagcgcccgg ccgtcgaacc ggggttccgg ctggaccggg aggagacctc cgaccgccgc gtccgctacc ggatccaccc ctacgcgctt gaccggccgc
      gtcgcgggcc ggcagcttgg ccccaaggcc gacctggccc tcctctggag gctggcggcg caggcgatgg cctaggtggg gatgcgcgaa ctggccggcg
      >.....................................................8.....................................................>
        q r p  a v e  p g f r  l d r  e e t  s d r r  v r y  r i h  p y a l  d r p
                                                                                  >>....9.....>
                                                                                    l t g r

8601  agggcgagcg gtacgcatga ccccgcccgt caccgcacct tcgatcgcga cgttcttcgg tccggaccgg gcggcgccat gaccgggcgg cccgccttcg
      tcccgctcgc catgcgtact ggggcgggca gtggcgtgga agctagcgct gcaagaagcc aggcctggcc cgccgcggta ctggcccgcc gggcggaagc
      >.........8........>>
       q g e  r y a  -
      >.....................................................9.....................................................>
        r a s  g t h  d p a r  h r t  f d r  d v l r  s g p  g g a  m t g r  p a f

8701  gcatgcgttc ggcggacggc ccggcggacg gcccggcgga cgaggccccc gccgggccgc tgccccccga cgcccgcgtg gacctggccc gcgaacgcgc
      cgtacgcaag ccgcctgccg ggccgcctgc cgggccgcct gctccggggg cggcccggcg acgggggggct gcgggcgcac ctggaccggg cgcttgcgcg
      >.....................................................9.....................................................>
       g m r  s a d g  p a d  g p a  d e a p  a g p  l p p  d a r v  d l a  r e r

8801  cgagtcccag gccgacgagg tcctggactc cctggaccgc gacctggtcg gcctggcccc ggtgaaggcc cgcgtccgcg agatcgccgc gctgctgctg
      gctcagggtc cggctgctcc aggacctgag ggacctggcg ctggaccagc cggaccgggg ccacttccgg gcgcaggcgc tctagcggcg cgacgacgac
      >.....................................................9.....................................................>
      a e s q  a d e  v l d  s l d r  d l v  g l a  p v k a  r v r  e i a  a l l l

8901  gtggaccgcg tccgggcgcg cttcggcatc gactcgggcc gccccaacct gcacatgtgc ttcaccggcg gccccggcac gggcaagacg accgtcgcgg
      cacctggcgc aggcccgcgc gaagccgtag ctgagcccgg cggggttgga cgtgtacacg aagtggccgc cggggccgtg cccgttctgc tggcagcgcc
      >.....................................................9.....................................................>
        v d r  v r a  r f g i  d s g  r p n  l h m c  f t g  g p g  t g k t  t v a

9001  tccgcctcgc cgaactgctg caccgcctcg gctacgtccg caagggccac ctggtgccgg tcacccgcga cgacctggtc ggccagtacg tcgggcacac
      aggcggagcg gcttgacgac gtggcggagc cgatgcaggc gttcccggtg gaccacggcc agtgggcgct gctggaccag ccggtcatgc agcccgtgtg
      >.....................................................9.....................................................>
       v r l  a e l l  h r l  g y v  r k g h  l v p  v t r  d d l v  g q y  v g h

9101  cgcgcccaag accaaggagg tcctcaagcg cgccatgggc ggcgtcctgt tcatcgacga ggcctaccac ctgtaccggg ccgagaacga acgcgactac
      gcgcgggttc tggttcctcc aggagttcgc gcggtacccg ccgcaggaca agtagctgct ccggatggtg gacatggccc ggctcttgct tgcgctgatg
      >.....................................................9.....................................................>
      t a p k  t k e  v l k  r a m g  g v l  f i d  e a y h  l y r  a e n  e r d y

9201  ggccaggagg ccatcgagat cctcctccag gtcatggaga accagcgcga cgacctcgtc gtcgtcctgg cgggctacaa ggaccgcatg gacgcgttct
      ccggtcctcc ggtagctcta ggaggaggtc cagtacctct tggtcgcgct gctggagcag cagcaggacc gcccgatgtt cctggcgtac ctgcgcaaga
      >.....................................................9.....................................................>
        g q e  a i e  i l l q  v m e  n q r  d d l v  v v l  a g y  k d r m  d a f

9301  tcgcctccaa ccccggcatg agctcccgca tcgcccacca catcgacttc cccgactacg aacccgccga actggaggag atcggccgcc tgatgatggc
      agcggaggtt ggggccgtac tcgagggcgt agcgggtggt gtagctgaag gggctgatgc ttgggcggct tgacctcctc tagccggcgg actactaccg
      >.....................................................9.....................................................>
       f a s  n p g m  s s r  i a h  h i d f  p d y  e p a  e l e e  i g r  l m m

9401  ccgcgagggc taccggctcg cccccgagac cgttccagtc ttccgcgact acctgcggcg acgccgcgtc cggccgcggt tcgccaacgc ccgttcggtc
      ggcgctcccg atggccgagc gggggctctg gcaaggtcag aaggcgctga tggacgccgc tgcggcgcag gccggcgcca agcggttgcg ggcaagccag
      >.....................................................9.....................................................>
      a r e g  y r l  a p e  t v p v  f r d  y l r  r r r v  r p r  f a n  a r s v

9501  cgcaacgcga tcgagcgcgc ccgcctgcgc cacgccggcc ggctgctggc cgcccccggc gaggtggacc tggccgccct caccaccctg ctgcccgagg
      gcgttgcgct agctcgcgcg ggcggacgcg gtgcggccgg ccgacgaccg gcgggggccg ctccacctgg accggcggga gtggtgggac gacgggctcc
      >.....................................................9.....................................................>
        r n a  i e r  a r l r  h a g  r l l  a a p g  e v d  l a a  l t t l  l p e

9601  acttcctggc cagccgcgtc ttcgaatgag accgcgatca cacacggaca agggatgccc ccgcccatgt cctaaggttg aagggctcaa gacatgaatg
      tgaaggaccg gtcggcgcag aagcttactc tggcgctagt gtgtgcctgt tccctacggg ggcgggtaca ggattccaac ttcccgagtt ctgtacttac
      >.............9............>>
       d f l  a s r v  f e  -
                                                                                  >>10..>
                                                                                    m  n

9701  tgggggagac ggtgctgaac cctgaggatc tgtacgagct ggacgccgac ctgccggagc tgacgggccc ggtcctgctg cacagcatcg acgggttcgt
      acccccctctg ccacgacttg ggactcctag acatgctcga cctgcggctg gacggcctcg actgcccggg ccaggacgac gtgtcgtagc tgcccaagca
      >.....................................................10....................................................>
       v g e  t v l n  p e d  l y e  l d a d  l p e  l t g  p v l l  h s i  d g f

9801  ggacgccgga tcggtcgggc agctcgtgcg cgaacacctg ctggacgccc tcgaccaccg cgtcatcgcg cggttcgacg tggactccct gatcgactac
      cctgcggcct agccagcccg tcgagcacgc gcttgtggac gacctgcggg agctggtggc gcagtagcgc gccaagctgc acctgaggga ctagctgatg
      >.....................................................10....................................................>
      v d a g  s v g  q l v  r e h l  l d a  l d h  r v i a  r f d  v d s  l i d y

9901  cgggcccgcc gcccgacgat gatctacgac cgcgaccact gggcctccta cgacgccccc gagctggtcc tgcgcctgct gcacgacgag gtcggcacgc
      gcccgggcgg cgggctgcta ctagatgctg gcgctggtga cccggaggat gctgcggggg ctcgaccagg acgcggacga cgtgctgctc cagccgtgcg
      >.....................................................10....................................................>
        r a r  r p t  m i y d  r d h  w a s  y d a p  e l v  l r l  l h d e  v g t
```

```
10001  cgttcctgct gctgaccggg cccgagcccg accggatgtg ggagggcttc gtcgccgccg tccgctcgct ggtggaccgc ctgaacgtcc gcctcgtcgt
       gcaaggacga cgactggccc gggctcgggc tggcctacac cctcccgaag cagcggcggc aggcgagcga ccacctggcg gacttgcagg cggagcagca
       >.....................................................10.....................................................>
        p f l  l l t g  p e p  d r m  w e g f  v a a  v r s  l v d r  l n v  r l v

10101  cggattccac ggcatcccga tgggcgttcc gcacacccgc ccggtgggga tgacctcgca cgcgacccgg ccggagctga tcacgcgctc ctcgtggttc
       gcctaaggtg ccgtagggct acccgcaagg cgtgtgggcg ggccaccccc actggagcgt gcgctgggcc ggcctcgact agtgcgcgag gagcaccaag
       >.....................................................10.....................................................>
       v g f h  g i p  m g v  p h t r  p v g  m t s  h a t r  p e l  i t r  s s w f

10201  gacaaggtcc aggtgcccgg cagcgccgcg gccctgctgg agtaccggct gggcgagtcc ggccacgacg cgctcggctt cgccgtccac gtcccgcact
       ctgttccagg tccacgggcc gtcgcggcgc cgggacgacc tcatggccga cccgctcagg ccggtgctgc gcgagccgaa gcggcaggtg cagggcgtga
       >.....................................................10.....................................................>
         d k v  q v p  g s a a  a l l  e y r  l g e s  g h d  a l g  f a v h  v p h

10301  acctggccca gaccacctac ccgaccgccg ccgtcgccgc cctcgacgcg gtgatcgccg ccaccggcct ggtgctgccc tccgccgccc tccgcgaggc
       tggaccgggt ctggtggatg ggctggcggc ggcagcggcg ggagctgcgc cactagcggc ggtggccgga ccacgacggg aggcggcggg aggcgctccg
       >.....................................................10.....................................................>
        y l a  q t t y  p t a  a v a  a l d a  v i a  a t g  l v l p  s a a  l r e

10401  cgccctcacg accgacgccg agatcaccga gcaggtcagc ggcaacgagg aggtcgagaa ggtcgtgcgc gccctggagc agcagtacga cgcgttcgcg
       gcgggagtgc tggctgcggc tctagtggct cgtccagtcg ccgttgctcc tccagctctt ccagcacgcg cgggacctcg tcgtcatgct gcgcaagcgc
       >.....................................................10.....................................................>
       a a l t  t d a  e i t  e q v s  g n e  e v e  k v v r  a l e  q q y  d a f a

10501  ggcgccgccg accgcgagag cctgctcgcc gaggcccggg acatgccgac ggccgaggag ctggccgccc agttcgaacg cttcctggcc gaacaggacg
       ccgcggcggc tggcgctctc ggacgagcgg ctccgggccc tgtacggctg ccggctcctc gaccggcggg tcaagcttgc gaaggaccgg cttgtcctgc
       >.....................................................10.....................................................>
         g a a  d r e  s l l a  e a r  d m p  t a e e  l a a  q f e  r f l a  e q d

10601  gcggcgaccc ccagccatga tccgtttcgc gcacccaccc cggaaccggg tacgctttcc cggtcaccga ggcacgaacg tgcccacggg cgcgtagctc
       cgccgctggg ggtcggtact aggcaaagcg cgtgggtggg gccttggccc atgcgaaagg gccagtggct ccgtgcttgc acgggtgccc gcgcatcgag
       >........10........>>
        g g d  p q p  -

10701  aggggtagag cactcgcctt acaagcgagg ggtcgcaggt tcgaaacctg ccgcgcccac cagccataaa gaccccgttt ccgatcatgg aaacggggtc
       tccccatctc gtgagcggaa tgttcgctcc ccagcgtcca agctttggac ggcgcgggtg gtcggtattt ctggggcaaa ggctagtacc tttgccccag

10801  ttttcgcagc cagggcgatc ctccaaggct cttgtcatac cgccatggat cggggcgcgg gcgtgccctg gtcgcgaccg tcgccgtagg cgtgccatgc
       aaaagcgtcg gtcccgctag gaggttccga gaacagtatg gcggtaccta gccccgcgcc cgcacgggac cagcgctggc agcggcatcc gcacggtacg
                                     <<.............................11.............................<
                                       - v  a m s  r p a  p t g q  d r g  d g y  a h w a

10901  ggtggccagg cggcgcgccg cctcggtcag cgtgtcgggg gcggcgagga agtggatccg gatgaactcc ctgctgcctc ccgaggcgtc cagcccgctg
       ccaccggtcc gccgcgcggc ggagccagtc gcacagcccc cgccgctcct tcacctaggc ctacttgagg gacgacggag ggctccgcag gtcgggcgac
       <.....................................................11.....................................................<
         t a l  r r a  a e t l  t d p  a a l  f h i r  i f e  r s g  g s a d  l g s

11001  ccgggcagca cggcgacgcc gtggcggagg gctgtctgcg cgaaggagtc gccgtcgccg cgggggaggc gcacccacag ggtctgaccg ccggtgacgg
       ggcccgtcgt gccgctgcgg caccgcctcc cgacagacgc gcttcctcag cggcagcggc gccccctccg cgtgggtgtc ccagactggc ggccactgcc
       <.....................................................11.....................................................<
        g p l  v a v g  h r l  a t q  a f s d  g d g  r p l  r v w l  t q g  g t v

11101  gtggggcctg ccactgcggc agccgccgtg ccagttccgc gcgcaggtgg tcatggcggg cccggcggtc cgccgcgatg tcctggcaca gggtctccag
       caccccggac ggtgacgccg tcggcggcac ggtcaaggcg cgcgtccacc agtaccgccc gggccgccag gcggcgctac aggaccgtgt cccagaggtc
       <.....................................................11.....................................................<
       p p a q  w q p  l r r  a l e a  r l h  d h r  a r r d  a a i  d q c  l t e l

11201  gtggggcagc agctcggcgg cggcgagctg cgcggtgacg ttccccccga ggtcgtgaac cgctcggagc ctggcgaggc ggtcgatcac gggccgcggg
       caccccgtcg tcgagccgcc gccgctcgac gcgccactgc aaggggggct ccagcacttg gcgagcctcg gaccgctccg ccagctagtg cccggcgccc
       <.....................................................11.....................................................<
         h p l  l e a  a a l q  a t v  n g g  l d h v  a r l  r a l  r d i v  p r p

11301  gcccgtatcc agccgatgcg cagcccgccc cagacgctct tgctgagcga gcccacggag atgaccgttt cgccgaacgc agcgagtgga ggtggtgtct
       cgggcatagg tcggctacgc gtcgggcggg gtctgcgaga acgactcgct cgggtgcctc tactggcaaa gcggcttgcg tcgctcacct ccaccacaga
       <.....................................................11.....................................................<
        a r i  w g i r  l g g  w v s  k s l s  g v s  i v t  e g f a  a l p  p p t

11401  cctgcccggg aagcgccagg tgggcgagga tttcgtcgtc gatgagcggc acgctcgccg cggcggcggc ggcgacgagc cgtcggcgcg ccagcgacgg
       ggacgggccc ttcgcggtcc acccgctcct aaagcagcag ctactcgccg tgcgagcggc gccgccgccg ccgctgctcg gcagccgcgc ggtcgctgcc
       <.....................................................11.....................................................<
       e q g p  l a l  h a l  i e d d  i l p  v s a  a a a a  a v l  r r r  a l s p

11501  cagtaccgct ccggtcgggt tgtggtgggt ggacacgagg taggcgagtt ccaggcggcg gtcagaggtc cggaccctcg cttcgaagcc gtccaggccg
       gtcatggcga ggccagccca acaccaccca cctgtgctcc atccgctcaa ggtccgccgc cagtctccag gcctgggagc gaagcttcgg caggtccggc
       <.....................................................11.....................................................<
         l v a  g t p  n h h t  s v l  y a l  e l r r  d s t  r v r  a e f g  d l g

11601  acggggagcg cctccaggga cgtggtctgt tcgcggaagg cctcgagcgc gccggggtag gtgggcgcct cgacgagcac cgtgccgccg ggctcgacga
       tgccctcgc ggaggtccct gcaccagaca agcgccttcc ggagctcgcg cggccccatc cacccgcgga gctgctcgtg gcacggcggc ccgagctgct
       <.....................................................11.....................................................<
        v p l  a e l s  t t q  e r f  a e l a  g p y  t p a  e v l v  t g g  p e v

11701  aggcgcgggc gagcagggag agggcttgct ggccgccgtt ggtgaccagc acctgcccgg ggccggtcgg caccccacgc cccgtgtagt ggtcggcgac
       tccgcgcccg ctcgtccctc tcccgaacga ccggcggcaa ccactggtcg tggacgggcc ccggccagcc gtggggtgcg gggcacatca ccagccgctg
       <.....................................................11.....................................................<
       f a r a  l l s  l a q  q g g n  t v l  v q g  p g t p  v g r  g t y  h d a v

11801  cgcacggcgc agccgccagt ggcccgacgg gtggtagccg atgtcgctcg tgaccgtggc cagctggggc aaggcccgct cgtaggccgt cagcagcatc
       gcgtgccgcg tcggcggtca ccgggctgcc caccatcggc tacagcgagc actggcaccg gtcgaccccg ttccgggcga gcatccggca gtcgtcgtag
       <.....................................................11.....................................................<
         a r r  l r w  h g s p  h y g  i d s  t v t a  l q p  l a r  e y a t  l l m

11901  gccggcggcg cggacggggc ggcgcatgcc aggaggatca cgtcgtcgtc ccggggttcg agcaggtgca ggaagaccgg cgcgcccgtg ctctcccggg
       cggccgccgc gcctgccccg ccgcgtacgg tcctcctagt gcagcagcag ggccccaagc tcgtccacgt ccttctggcc gcgcgggcac gagagggccc
       <.....................................................11.....................................................<
        a p p  a s p a  a c a  l l i  v d d d  r p e  l l h  l f v p  a g t  s e r

12001  gtccggggcg cggtggcccg gccacgcggg tcccgcttcc ctggcgtcgg ctgatccggc cttccaggcg cagctggtcg tacgcggcca cgacggtccc
       caggccccgc gccaccgggc cggtgcgccc agggcgaagg gaccgcagcc gactaggccg gaaggtccgc gtcgaccagc atgcgccggt gctgccaggg
       <.....................................................11.....................................................<
       p g p r  p p g  a v r  t g s g  q r r  s i r  g e l r  l q d  y a a  v v t g
```

```
12101  gcgtcctacc gcgagcacgg acgcgagggc ccggtcgggt ggcagcgtgg tgccggacgg cagctcgccc tcgtcgacga ggcgcctcag ccgcgcggcg
       cgcaggatgg cgctcgtgcc tgcgctcccg ggccagccca ccgtcgcacc acggcctgcc gtcgagcggg agcagctgct ccgcggagtc ggcgcgccgc
       <.....................................................11.....................................................<
         r  g  v  a  l  v  s  a  l  a  r  d  p  p  l  t  t  g  s  p  l  e  g  e  d  v  l  r  r  l  r  a  a

12201  agcaggaggt acagcggtcc gcggccggcg gaccaccggc ccagacggtc ggccagctcg gccgggtcca tcgacatgat cgagttcttc cggacgcgat
       tcgtcctcca tgtcgccagg cgccggccgc ctggtggccg ggtctgccag ccggtcgagc cggcccaggt agctgtacta gctcaagaag gcctgcgcta
       <.....................................11.....................................<<
        l  l  l  y  l  p  g  r  g  a  s  w  r  g  l  r  d  a  l  e  a  p  d  m  s  m

12301  cgagatgatt ggttcgttcg acggaccaat catggcgcat tggctctgcc agccgtccac gatgacggcg aggatcgcga catgaccaca caacacgagg
       gctctactaa ccaagcaagc tgcctggtta gtaccgcgta accgagacgg tcggcaggtg ctactgccgc tcctagcgct gtactggtgt gttgtgctcc
                                                                                                 >>.......12........>
                                                                                                   m  t  t  q  h  e

12401  acctgcggca gctcgccgcg gtatcggtgc ccgaggccgt cgccgtgctc ccggggccct tccagcagcg cgatctggca atggtcaacg acaccaccgt
       tggacgccgt cgagcggcgc catagccacg ggctccggca gcggcacgag ggccccggga aggtcgtcgc gctagaccgt taccagttgc tgtggtggca
       >.....................................................12.....................................................>
        d  l  r  q  l  a  a  v  s  v  p  e  a  v  a  v  l  p  g  p  f  q  q  r  d  l  a  m  v  n  d  t  t

12501  cgtccgcctc gctcgtctgg aagacgagtt cccctggcac catcacgacg aggacgagct gttcctgtgc tgggacggca cctttgagat ccagttggag
       gcaggcggag cgagcagacc ttctgctcaa ggggaccgtg gtagtgctgc tcctgctcga caaggacacg accctgccgt ggaaactcta ggtcaacctc
       >.....................................................12.....................................................>
       v  v  r  l  a  r  l  e  d  e  f  p  w  h  h  h  d  e  d  e  l  f  l  c  w  d  g  t  f  e  i  q  l  e

12601  gggcgggacc cggtggtgct gaaggcgggc gaactcttcg tcgtccccag gggcgtacgg caccgccccg tcgctgagaa ggccgcccac atcctgctga
       cccgccctgg gccaccacga cttccgcccg cttgagaagc agcaggggtc cccgcatgcc gtggcggggc agcgactctt ccggcgggtg taggacgact
       >.....................................................12.....................................................>
         g  r  d  p  v  v  l  k  a  g  e  l  f  v  v  p  r  g  v  r  h  r  p  v  a  e  k  a  a  h  i  l  l

12701  tcgagcaccc ggacaccaag cagtacggaa gctgaccccg aggggcttca ccgcccgtgt ccggtccggg cactcgtcac cggacgtcgg cttgccgcat
       agctcgtggg cctgtggttc gtcatgcctt cgactggggc tccccgaagt ggcgggcaca ggccaggccc gtgagcagtg gcctgcagcc gaacggcgta
       >.................12.................>>
        i  e  h  p  d  t  k  q  y  g  s  -

12801  ggcccgggaa cgcggcaggc cgacgctccg tccttgccct ccgttcgact tcctgggccc ggagcactcc ggcatgtcgc cgcgtctgcg ctgtcagggg
       ccgggccctt gcgccgtccg gctgcgaggc aggaacggga ggcaagctga aggacccggg cctcgtgagg ccgtacagcg gcgcagacgc gacagtcccc

12901  agagtcgtcc ctcccccgga cgccgagagg tcaggacggg agggtggcgc gtagggctgt tatgcacttg aggacttcgg ggcggttgtc ggccgctttg
       tctcagcagg gagggggcct gcggctctcc agtcctgccc tcccaccgcg catcccgaca atacgtgaac tcctgaagcc ccgccaacag ccggcgaaac
                                <<.....................................13.....................................<
                                  -  s  p  l  t  a  r  l  a  t  i  c  k  l  v  e  p  r  n  d  a  a  k

13001  atccaggcgg gtaggcgttc caggaggctg cctgagcctg tccagcctcc tcgttcgcgt aggcgggcgg ccacgtaggc ttccagcagg tccaggtggc
       taggtccgcc catccgcaag gtcctccgac ggactcggac aggtcggagg agcaagcgca tccgcccgcc ggtgcatccg aaggtcgtcc aggtccaccg
       <.....................................................13.....................................................<
        i  w  a  p  l  r  e  l  l  s  g  s  g  t  w  g  g  r  e  r  l  r  a  a  v  y  a  e  l  l  d  l  h

13101  gttggttgta ggcccacagg acgtggccgc agcatgggcg ttgcagccac agcggcaggt cgaagtaggg gtcgttcggg ccggtgaggt gggggaccag
       caaccaacat ccgggtgtcc tgcaccggcg tcgtacccgc aacgtcggtg tcgccgtcca gcttcatccc cagcaagccc ggccactcca ccccctggtc
       <.....................................................13.....................................................<
       r  q  n  y  a  w  l  v  h  g  c  c  p  r  q  l  w  l  p  l  d  f  y  p  d  n  p  g  t  l  h  p  v  l

13201  caggggggcg ccgtgggtga agtaggggtc tccgtcgtcg ccgcgctgcc agcggcgtgg cgtccactcg gcggtgtggc cgcaccccgt gcagctcagc
       gtccccggc ggcacccact tcatccccag aggcagcagc ggcgcgacgg tcgccgcacc gcaggtgagc cgccacaccg gcgtggggca cgtcgagtcg
       <.....................................................13.....................................................<
         l  p  g  g  h  t  f  y  p  d  g  d  d  g  r  q  w  r  r  p  t  w  e  a  t  h  g  c  g  t  c  s  l

13301  cggtgcggcg ccgtcaggag gccgatcacg tagcgatggc cttcgcggtg gtcggggtga acgaccacga cggcccgtgc gtcgcagcgg gggcagcgga
       gccacgccgc ggcagtcctc cggctagtgc atcgctaccg gaagcgccac cagccccact tgctggtgct gccgggcacg cagcgtcgcc cccgtcgcct
       <.....................................................13.....................................................<
        r  h  p  a  t  l  l  g  i  v  y  r  h  g  e  r  h  d  p  h  v  v  v  v  a  r  a  d  c  r  p  c  r

13401  ccaggatctc gtcgccgaag ttccagagcc gccaccccgg gtcctggaag cgctcgtcgg tcatgcttga tctttccatg gcggccccgg acgcaggaca
       ggtcctagag cagcggcttc aaggtctcgg cggtggggcc caggaccttc gcgagcagcc agtacgaact agaaaggtac cgccggggcc tgcgtcctgt
       <.....................................13.....................................<<
       v  l  i  e  d  g  f  n  w  l  r  w  g  p  d  q  f  r  e  d  t  m  s  s  r  e  m

13501  atgcccgggt gggcaaggat tcctctcgga tggtcggcgc gtcgctcggt gagctgctgg cctggctggt cgatggggag cctcgggttc ggaggttggc
       tacgggccca cccgttccta aggagagcct accagccgcg cagcgagcca ctcgacgacc ggaccgacca gctacccctc ggagcccaag cctccaaccg

13601  ggccgcttat cttggggacc ggctcgtcca ggactgccgg gcggggcttg aggtctcgtc ggtcgtgatg ccgctggtcg aggctctgtc gcgggagacc
       ccggcgaata gaacccctgg ccgagcaggt cctgacggcc cgccccgaac tccagagcag ccagcactac ggcgaccagc tccgagacag cgccctctgg

13701  gatcccgtgg tgcaggagga gatcgctcac tcgctcgggt atctggtcga gtacggcacc gttccgcaga gtattgtccg gccgcttcgg gagtgtctgc
       ctagggcacc acgtcctcct ctagcgagtg agcgagccca tagaccagct catgccgtgg caaggcgtct cataacaggc cggcgaagcc ctcacagacg

13801  cgcggcttca cgccggggcc gccgagcatg tcaacgacgt ccttggggcg gtgcggtgac gctcagttcg gggacgggct ggcgagttcc tggagcaggt
       gcgccgaagt gcggccccgg cggctcgtac agttgctgca ggaaccccgc cacgccactg cgagtcaagc ccctgcccga ccgctcaagg acctcgtcca

13901  ccagcgagcg gacgggttcc gcgcgccgga cccttccggc ctcgaccagc accgcgaacg gcgtgcccgg cacccggtac tcgttgagca gggcctgttc
       ggtcgctcgc ctgcccaagg cgcgcggcct gggaaggccg gagctggtcg tggcgcttgc cgcacgggcc gtgggccatg agcaactcgt cccggacaag

14001  gccggtgtgc accgggatga cgttcccggc gcccatcggc agcgggtcgc ccgcgaagac ggcccgcatc gggacgacgt cgccggtggc ggccaggtcg
       cggccacacg tggccctact gcaagggccg cgggtagccg tcgcccagcg ggcgcttctg ccgggcgtag ccctgctgca gcggccaccg ccggtccagc

14101  agggccttct ccagtaccgc cgggcagcct ccgcagcccg agctgaggaa cagcagcagc gtcgggccgg acggcgcggg ggcgagccgg tccaggccgg
       tcccggaaga ggtcatggcg gcccgtcgga ggcgtcgggc tcgactcctt gtcgtcgtcg cagcccggcc tgccgcgccc ccgctcggcc aggtccggcc

14201  gggcgcggga cccggccggg aggcccagtt cgggaacgcg cgcccggagc gaccggacct gccggttcat cctgagcagg agcagggtga gcaccaggag
       cccgcgccct gggccggccc tccgggtcaa gcccttgcgc gcgggcctcg ctggcctgga cggccaagta ggactcgtcc tcgtcccact cgtggtcctc

14301  gacggcccag acgaggacgt cggtctcgac atgatccatg gcgccgcccc ctagttcgcc gggcgcggcg cggcgagtgc gggcagctcc cagaccaggg
       ctgccgggtc tgctcctgca gccagagctg tactaggtac cgcggcgggg gatcaagcgg cccgcgccgc gccgctcacg cccgtcgagg gtctggtccc

14401  cgccgagggt ggctcccgcc aggaccgcga ccaggaggtc cggcgcgtcc acggacgccg cgtgcgccgc gccgccggcg gccacggcgg cgacggccgc
       gcggctccca ccgagggcgg tcctggcgct ggtcctccag gccgcgcagg tgcctgcggc gcacgcggcg cggcggccgc cggtgccgcc gctgccggcg

14501  cagcgccacg gccctgagcg ggacccagcc ggtcatgggc gttcccgccg ccgagcaccc gcagggcacc gcggcctcgc tggtgcgggt gacgtgccag
       gtcgcggtgc cgggactcgc cctgggtcgg ccagtacccg caagggcggc ggctcgtggg cgtcccgtgg cgccggagcg accacgccca ctgcacggtc

14601  gtgtaggcgc cgtagcccgc catcagcgcg gccgcgcaga ccatcgcggc cgtgagggct ccgtcgtgcc cgcccagcag gccgaacgcg ccgagccccg
       cacatccgcg gcatcgggcg gtagtcgcgc cggcgcgtct ggtagcgccg gcactcccga ggcagcacgg gcgggtcgtc cggcttgcgc ggctcggggc
```

```
14701  ccggtacggc ttcgatgacc ggcaccagga ccgccagcgg gcggacgaag cgctccggca ccgtcgcgtg cgcgcgcagc gcgtcccgca ggacgaacgg
       ggccatgccg aagctactgg ccgtggtcct ggcggtcgcc cgcctgcttc gcgaggccgt ggcagcgcac gcgcgcgtcg cgcagggcgt cctgcttgcc

14801  tttcagcgcg tgccccgccg cgcccgagat cagcacgagc gtcaccagcg cggccgcgat cgcataaccc attcgtgttc acctccatga atcccctggt
       aaagtcgcgc acggggcggc gcgggctcta gtcgtgctcg cagtggtcgc gccggcgcta gcgtattggg taagcacaag tggaggtact taggggacca

14901  cggtgccgac tataaccgc ggccaagtga cggcgcggaa acccgtcgtg acgtatgcgc gatgaccagg ccgtttcggc gacggccgaa gtgactggtt
       gccacggctg atattgggcg ccggttcact gccgcgcctt tgggcagcac tgcatacgcg ctactggtcc ggcaaagccg ctgccggctt cactgaccaa

15001  cgcggggcgt ccggtggcac acgatggggg tgcgccgaaa gattggggta gtggatggat ctgcggtacc acgcctatgc gatggccgat cccgtctttt
       gcgccccgca ggccaccgtg tgctaccccc acgcggcttt ctaaccccat cacctaccta gacgccatgg tgcggatacg ctaccggcta gggcagaaaa
                                                                 >>.....................LabKC.....................>
                                                                   m  d  l  r  y  h  a  y  a  m  a  d  p  v  f

15101  atgactcgcc gtccagcgac acccgggaga ccgacggcta ctccgacgac ctccccctgc cggacggctg ggaacgtcgg cgcgtggggg tctgggtgat
       tactgagcgg caggtcgctg tgggccctct ggctgccgat gaggctgctg gaggggacg gcctgccgac ccttgcagcc gcgcaccccc agacccacta
       >.....................................................LabKC.....................................................>
        y  d  s  p  s  s  d  t  r  e  t  d  g  y  s  d  d  l  p  l  p  d  g  w  e  r  r  r  v  g  v  w  v

15201  gcagggccac gacgggttga ccatgcctga ccagggctgg aagatccatg tgtcggccgg tctggacaac gcctggcccg tcctcgaact ggtcgccaaa
       cgtcccggtg ctgcccaact ggtacggact ggtcccgacc ttctaggtac acagccggcc agacctgttg cggaccgggc aggagcttga ccagcggttt
       >.....................................................LabKC.....................................................>
       m  q  g  h  d  g  l  t  m  p  d  q  g  w  k  i  h  v  s  a  g  l  d  n  a  w  p  v  l  e  l  v  a  k

15301  tactgcgtcg agcaggaaat gcctttcaag ttcctgcgca gcagaaggac gttgctggcc cgcagcagca agtacgccga acgcggcggc agcggaaagt
       atgacgcagc tcgtcctta cggaaagttc aaggacgcgt cgtcttcctg caacgaccgg gcgtcgtcgt tcatgcggct tgcgccgccg tcgcctttca
       >.....................................................LabKC.....................................................>
         y  c  v  e  q  e  m  p  f  k  f  l  r  s  r  r  t  l  l  a  r  s  s  k  y  a  e  r  g  g  s  g  k

15401  tcatcacgat ctaccccgcc gacgaaggcg ctctggaaaa gaccctccat gaactcggtg gaatgctgga agggcagccc ggcccctata tcttgagcga
       agtagtgcta gatggggcgg ctgcttccgc gagacctttt ctgggaggta cttgagccac cttacgacct tcccgtcggg ccggggatat agaactcgct
       >.....................................................LabKC.....................................................>
        f  i  t  i  y  p  a  d  e  g  a  l  e  k  t  l  h  e  l  g  g  m  l  e  g  q  p  g  p  y  i  l  s

15501  tctgcgctgg cgttcggggc cgctgttcgt gcgctacggc gctttcaagg agaaattctg ccgggacggg cgcggcgaga tggtgcccgc gatcgcgcgc
       agacgcgacc gcaagccccg gcgacaagca cgcgatgccg cgaaagttcc tctttaagac ggccctgccc gcgccgctct accacgggcg ctagcgcgcg
       >.....................................................LabKC.....................................................>
       d  l  r  w  r  s  g  p  l  f  v  r  y  g  a  f  k  e  k  f  c  r  d  g  r  g  e  m  v  p  a  i  a  r

15601  ccggacggcg tgctggtgcc cgacgccagg gacccggtgt tccgggtgcc cgcgtgggtg gaggtgcccg gcttcctgcg ggaggccatc gacgcccggg
       ggcctgccgc acgaccacgg gctgcggtcc ctgggccaca aggcccacgg gcgcacccac ctccacgggc cgaaggacgc cctccggtag ctgcgggccc
       >.....................................................LabKC.....................................................>
         p  d  g  v  l  v  p  d  a  r  d  p  v  f  r  v  p  a  w  v  e  v  p  g  f  l  r  e  a  i  d  a  r

15701  agaacgggac cgtcgaggac ttcccctacc ggatcgagaa ggcgctgcac ttctccaacg gcggcggcct ctaccgcgcc gtggacgagc gcaccggccg
       tcttgccctg gcagctcctg aaggggatgg cctagctctt ccgcgacgtg aagaggttgc cgccgccgga gatggcgcgg cacctgctcg cgtggccggc
       >.....................................................LabKC.....................................................>
        e  n  g  t  v  e  d  f  p  y  r  i  e  k  a  l  h  f  s  n  g  g  g  l  y  r  a  v  d  e  r  t  g

15801  cagggtcctg gtgaaggagg cgcggccgat ggcgggcctg gaccgcgccg aggacgacgc cgtcgtccgg ctcgaacgcg agcacggcct gctgctccgc
       gtcccaggac cacttcctcc gcgccggcta ccgcccggac ctggcgcggc tcctgctgcg gcagcaggcc gagcttgcgc tcgtgccgga cgacgaggcg
       >.....................................................LabKC.....................................................>
       r  r  v  l  v  k  e  a  r  p  m  a  g  l  d  r  a  e  d  d  a  v  v  r  l  e  r  e  h  g  l  l  l  r

15901  ctcgccgacc tggactgcat ccccgacctg gtcgagtacc ggagctggtg ggagcaccgc ttcctcgtgc gggagtacgt cgagggcgag accctcaccc
       gagcggctgg acctgacgta ggggctggac cagctcatgg cctcgaccac cctcgtggcg aaggagcacg ccctcatgca gctcccgctc tgggagtggg
       >.....................................................LabKC.....................................................>
         l  a  d  l  d  c  i  p  d  l  v  e  y  r  s  w  w  e  h  r  f  l  v  r  e  y  v  e  g  e  t  l  t

16001  accacatggt gcgccgcaac ccgatgctgc actacggcgc gacgccgcag gaggtcgccg gctacaccga gtgggcgctc ggcgtcgtcg accgggtgga
       tggtgtacca cgcggcgttg ggctacgacg tgatgccgcg ctgcggcgtc ctccagcggc cgatgtggct cacccgcgag ccgcagcagc tggcccacct
       >.....................................................LabKC.....................................................>
        h  h  m  v  r  r  n  p  m  l  h  y  g  a  t  p  q  e  v  a  g  y  t  e  w  a  l  g  v  v  d  r  v

16101  gagcgcgctc ggccggctgc acgaacgcgg cgtggtcttc ggcgacctgc accccggcaa catcatcgtc cgggacgacg actccatcgt gttcgtcgac
       ctcgcgcgag ccggccgacg tgcttgcgcc gcaccagaag ccgctggacg tggggccgtt gtagtagcag gccctgctgc tgaggtagca caagcagctg
       >.....................................................LabKC.....................................................>
       e  s  a  l  g  r  l  h  e  r  g  v  v  f  g  d  l  h  p  g  n  i  i  v  r  d  d  d  s  i  v  f  v  d

16201  ttcgagctgg tcgccgaggc ggaggaggcg acgcacccgg cgctcggcgc gcccggctac caggccccgc ccgactacac cgggttcgcc atcgaccgct
       aagctcgacc agcggctccg cctcctccgc tgcgtgggcc gcgagccgcg cgggccgatg gtccggggcg ggctgatgtg gcccaagcgg tagctggcga
       >.....................................................LabKC.....................................................>
         f  e  l  v  a  e  a  e  e  a  t  h  p  a  l  g  a  p  g  y  q  a  p  p  d  y  t  g  f  a  i  d  r

16301  acgcgctcgg ctgcatcagg ctcgcggtct tcacctccct caccgcgacg ctgcactggg acgaccgcaa ggtcgagcag ttcctcgacg tgatctgcga
       tgcgcgagcc gacgtagtcc gagcgccaga agtggaggga gtggcgctgc gacgtgaccc tgctggcgtt ccagctcgtc aaggagctgc actagacgct
       >.....................................................LabKC.....................................................>
        y  a  l  g  c  i  r  l  a  v  f  t  s  l  t  a  t  l  h  w  d  d  r  k  v  e  q  f  l  d  v  i  c

16401  gtccttcccg ctgccgccgg actacgccga ccggatccgc cgcgacctcg cccgcccggc gccggcggac ggcgcgccgc cgatctggcg cgagccgacg
       caggaagggc gacggcggcc tgatgcggct ggcctaggcg gcgctggagc gggcgggccg cggccgcctg ccgcgcggcg gctagaccgc gctcggctgc
       >.....................................................LabKC.....................................................>
       e  s  f  p  l  p  p  d  y  a  d  r  i  r  r  d  l  a  r  p  a  p  a  d  g  a  p  p  i  w  r  e  p  t

16501  cccgccacct ggccggacac gcgggccggg atcgccgccg cgatcctcga cacggccacg cccgagcgcg ccgaccggct cttcccgggc gacatcgagc
       gggcggtgga ccggcctgtg cgcccggccc tagcggcggc gctaggagct gtgccggtgc gggctcgcgc ggctggccga gaagggcccg ctgtagctcg
       >.....................................................LabKC.....................................................>
         p  a  t  w  p  d  t  r  a  g  i  a  a  a  i  l  d  t  a  t  p  e  r  a  d  r  l  f  p  g  d  i  e

16601  agttcgcgac cagcgtcggt gggatcgggt tcggccacgg cgcggcgggc gtgctgtggg cgctggccga ggcgggcgcc ggccgcttcc ccgaccacga
       tcaagcgctg gtcgcagcca ccctagccca agccggtgcc gcgccgcccg cacgacaccc gcgaccggct ccgcccgcgg ccggcgaagg ggctggtgct
       >.....................................................LabKC.....................................................>
        q  f  a  t  s  v  g  g  i  g  f  g  h  g  a  a  g  v  l  w  a  l  a  e  a  g  a  g  r  f  p  d  h

16701  ggactgggtg cgggacgccg tcgccagggc gcaacggccg ccgcccggct tctacgacgg cgtcgcgggc gtcgcccacg tcctggaccg gctcggccgc
       cctgacccac gccctgcggc agcggtcccg cgttgccggc ggcgggccga agatgctgcc gcagcgcccg cagcgggtgc aggacctggc cgagccggcg
       >.....................................................LabKC.....................................................>
       e  d  w  v  r  d  a  v  a  r  a  q  r  p  p  p  g  f  y  d  g  v  a  g  v  a  h  v  l  d  r  l  g  r

16801  gccgacgagg cccgcgagct catggagcac gcgcccgccg cgacgggggc gaccgacaac agcctctacc gggggctggc gggcatcggc ctcaaccagc
       cggctgctcc gggcgctcga gtacctcgtg cgcgggcggc gctgcccccg ctggctgttg tcggagatgg cccccgaccg cccgtagccg gagttggtcg
       >.....................................................LabKC.....................................................>
         a  d  e  a  r  e  l  m  e  h  a  p  a  a  t  g  a  t  d  n  s  l  y  r  g  l  a  g  i  g  l  n  q
```

```
16901  tccacttcgc ccgcgtcacg ggcgaggcgt cgttcgccgc ggcggccgag gagaccgccg gccgggtggt cgcgaacctg cgccgcaaga cggagggcgc
       aggtgaagcg ggcgcagtgc ccgctccgca gcaagcggcg ccgccggctc ctctggcggc cggcccacca gcgcttggac gcggcgttct gcctcccgcg
       >.....................................................LabKC.....................................................>
        l  h  f  a  r  v  t  g  e  a  s  f  a  a  a  a  e  e  t  a  g  r  v  v  a  n  l  r  r  k  t  e  g

17001  ctaccgggcg ggcctgatgt acggctcctc cggcccggcc ctgttcctcg tccggatgtt cgaggcgacg ggcgacggcc actggctgga cgaggccgaa
       gatggcccgc ccggactaca tgccgaggag gccgggccgg gacaaggagc aggcctacaa gctccgctgc ccgctgccgg tgaccgacct gctccggctt
       >.....................................................LabKC.....................................................>
       a  y  r  a  g  l  m  y  g  s  s  g  p  a  l  f  l  v  r  m  f  e  a  t  g  d  g  h  w  l  d  e  a  e

17101  cgcgccctgc accgcgagct ggacgcctgc aagtggacgc agaaggacag cacgctccag gtcgacgagg gctggcgggt ccttccgtac gtcgccaccg
       gcgcgggacg tggcgctcga cctgcggacg ttcacctgcg tcttcctgtc gtgcgaggtc cagctgctcc cgaccgccca ggaaggcatg cagcggtggc
       >.....................................................LabKC.....................................................>
         r  a  l  h  r  e  l  d  a  c  k  w  t  q  k  d  s  t  l  q  v  d  e  g  w  r  v  l  p  y  v  a  t

17201  gcagcgtggg catcggcatc gcgctgcacg agttcctgcg gcaccgcccg gcgccgcgct tcaccgaggc gcaggagggg atccggaggg cggcggctcc
       cgtcgcaccc gtagccgtag cgcgacgtgc tcaaggacgc cgtggcgggc cgcggcgcga agtggctccg cgtcctcccc taggcctccc gccgccgagg
       >.....................................................LabKC.....................................................>
        g  s  v  g  i  g  i  a  l  h  e  f  l  r  h  r  p  a  p  r  f  t  e  a  q  e  g  i  r  r  a  a  a

17301  cgcgtacttc gtgcagagcg gcctgctcaa cggacggtcc ggcatcctcg cctacctgct gcacgtcggg gccggccggg aggacccggt cgtccggacg
       gcgcatgaag cacgtctcgc cggacgagtt gcctgccagg ccgtaggagc ggatggacga cgtgcagccc cggccggccc tcctgggcca gcaggcctgc
       >.....................................................LabKC.....................................................>
       p  a  y  f  v  q  s  g  l  l  n  g  r  s  g  i  l  a  y  l  l  h  v  g  a  g  r  e  d  p  v  v  r  t

17401  cacctgcgca acctcggctg gcacgccgtc ccgtaccccg gccggggcga ggacgccccg gcgcccggcg cgcggcggac cgcgttcatc ggggaccagc
       gtggacgcgt tggagccgac cgtgcggcag ggcatggggc cggccccgct cctgcggggc cgcgggccgc gcgccgcctg gcgcaagtag cccctggtcg
       >.....................................................LabKC.....................................................>
         h  l  r  n  l  g  w  h  a  v  p  y  p  g  r  g  e  d  a  p  a  p  g  a  r  r  t  a  f  i  g  d  q

17501  tgctgcgcct gtccatggac ctcgccacgg gctccgccgg ggtcctggcc acggtcgagg cggccctggg cggccacccc ctgagactcc ccttcctcca
       acgacgcgga caggtacctg gagcggtgcc cgaggcggcc ccaggaccgg tgccagctcc gccgggaccc gccggtgggg gactctgagg ggaaggaggt
       >.....................................................LabKC.....................................................>
        l  l  r  l  s  m  d  l  a  t  g  s  a  g  v  l  a  t  v  e  a  a  l  g  g  h  p  l  r  l  p  f  l

17601  tccggaggaa ggggcgagca cgcgaccccg ggggaggagg tgaacatcca ccatggcatc catccttgag ctccaggacc tggaggtcga gcgcgccagc
       aggcctcctt ccccgctcgt gcgctggggc cccctcctcc acttgtaggt ggtaccgtag gtaggaactc gaggtcctgg acctccagct cgcgcggtcg
       >...................LabKC...................>>
       h  p  e  e  g  a  s  t  r  p  r  g  r  r  -
                                                    >>...................LabA1/A3...................>
                                                      m  a  s  i  l  e  l  q  d  l  e  v  e  r  a  s

17701  tcggccgccg acagcaacgc cagcgtctgg gagtgctgca gcacgggcag ctgggttccc ttcacctgct gctgacgccc gcacaccgtt ccaccgatga
       agccggcggc tgtcgttgcg gtcgcagacc ctcacgacgt cgtgcccgtc gacccaaggg aagtggacga cgactgcggg cgtgtggcaa ggtggctact
       >.................................LabA1/A3.............................>>
         s  a  a  d  s  n  a  s  v  w  e  c  c  s  t  g  s  w  v  p  f  t  c  c  -

17801  gaggtgacag tcccatggcg tcgatcctgg aactccagaa cctggacgtc gagcacgccc gcggcgagaa ccgctccgac tggagcctgt gggagtgctg
       ctccactgtc agggtaccgc agctaggacc ttgaggtctt ggacctgcag ctcgtgcggg cgccgctctt ggcgaggctg acctcggaca ccctcacgac
                 >>.....................................LabA2.....................................>
                   m  a  s  i  l  e  l  q  n  l  d  v  e  h  a  r  g  e  n  r  s  d  w  s  l  w  e  c

17901  tagcacggga agcctgttcg cctgctgctg aacagcgctg accgaggccg ccgaagcggc acccatgtga aacgaccgcc cggggcggtg gggactccca
       atcgtgccct tcggacaagc ggacgacgac ttgtcgcgac tggctccggc ggcttcgccg tgggtacact ttgctggcgg gccccgccac ccctgagggt
       >............LabA2............>>
       c  s  t  g  s  l  f  a  c  c  -

18001  ccgccccggg cggaccacca tcagcgagga gcacgacagg aatgcacgcc gaccggctgc tcgtccaggc gctgcgcgcc ggccccgggt ggacggccct
       ggcggggccc gcctggtggt agtcgctcct cgtgctgtcc ttacgtgcgg ctggccgacg agcaggtccg cgacgcgcgg ccgggccca cctgccggga
                                              >>.........................LabT1.........................>
                                                m  h  a  d  r  l  l  v  q  a  l  r  a  g  p  g  w  t  a

18101  cctggtcgcc gcgacgctgc tgaacgcggt gtgcgcgctc gccctgcccg ccgcgatcgg cgcggtcacc gacgcggtgc tcgcccgcga cgcgggccgc
       ggaccagcgg cgctgcgacg acttgcgcca cacgcgcgag cgggacgggc ggcgctagcc gcgccagtgg ctgcgccacg agcgggcgct gcgcccggcg
       >.....................................................LabT1.....................................................>
       l  l  v  a  a  t  l  l  n  a  v  c  a  l  a  l  p  a  a  i  g  a  v  t  d  a  v  l  a  r  d  a  g  r

18201  ggcgccgggg ccccgctggg ctggctggcg gcggcgctcg gcggcgtggc gctgggcacg gcgctgtccc ggctcgccga ggtgtactgc ggcacggccg
       ccgcggcccc ggggcgaccc gaccgaccgc cgccgcgagc cgccgcaccg cgacccgtgc cgcgacaggg ccgagcggct ccacatgacg ccgtgccggc
       >.....................................................LabT1.....................................................>
         g  a  g  a  p  l  g  w  l  a  a  a  l  g  g  v  a  l  g  t  a  l  s  r  l  a  e  v  y  c  g  t  a

18301  cgaccaggga gctgcggacg cgcctggtca cgcacgtgct ggccctcggg gtgccgggga cgcgcgcgtt cacgcccggc gacctggcca gccgcgccgt
       gctggtccct cgacgcctgc gcggaccagt gcgtgcacga ccgggagccc cacggcccct gcgcgcgcaa gtgcgggccg ctggaccggt cggcgcggca
       >.....................................................LabT1.....................................................>
        a  t  r  e  l  r  t  r  l  v  t  h  v  l  a  l  g  v  p  g  t  r  a  f  t  p  g  d  l  a  s  r  a

18401  caccggcgcg ccgcaggccg gcgcggtcgc gcggtcggtg gtcagcgcgg tcgcggggct gctgatgtcg gtcggcggcg tggtcgcgct gtggctgatc
       gtggccgcgc ggcgtccggc cgcgccagcg cgccagccac cagtcgcgcc agcgccccga cgactacagc cagccgccgc accagcgcga caccgactag
       >.....................................................LabT1.....................................................>
       v  t  g  a  p  q  a  g  a  v  a  r  s  v  v  s  a  v  a  g  l  l  m  s  v  g  g  v  v  a  l  w  l  i

18501  gactggcggc tggtggcggc ggtgggcggc gcggtcccga tcgggctgct gctcatgcgc gtcttcgtcc gggacgcctc cgacctggtc accgagtacc
       ctgaccgccg accaccgccg ccacccgccg cgccagggct agcccgacga cgagtacgcg cagaagcagg ccctgcggag gctggaccag tggctcatgg
       >.....................................................LabT1.....................................................>
         d  w  r  l  v  a  a  v  g  g  a  v  p  i  g  l  l  l  m  r  v  f  v  r  d  a  s  d  l  v  t  e  y

18601  aggaggcgca gggcgagatc gccgcgcgga tgaccgacgc gctcaccggc atccgcacca tccgcgccgc cgggacctgg cggcgcgagg ccgaccgggt
       tcctccgcgt cccgctctag cggcgcgcct actggctgcg cgagtggccg taggcgtggt aggcgcggcg gccctggacc gccgcgctcc ggctggccca
       >.....................................................LabT1.....................................................>
        q  e  a  q  g  e  i  a  a  r  m  t  d  a  l  t  g  i  r  t  i  r  a  a  g  t  w  r  r  e  a  d  r

18701  cctggccccg ctgccgggcc tgtcggactc cgggcggcgg ctctggcacg cctacgggcg gatgcagggg cagggccgcc tgatcgtccc gctggccgag
       ggaccggggc gacggcccgg acagcctgag gcccgccgcc gagaccgtgc ggatgcccgc ctacgtcccc gtcccggcgg actagcaggg cgaccggctc
       >.....................................................LabT1.....................................................>
       v  l  a  p  l  p  g  l  s  d  s  g  r  r  l  w  h  a  y  g  r  m  q  g  q  g  r  l  i  v  p  l  a  e

18801  atcgcggcgc tggccgtcgc cgggcggggc gtcctgaccg gtcggctgtc ggcgggggag atgctcgcgg cggcgggcta cgccccgatg gcgctgggcc
       tagcgccgcg accggcagcg gcccgccccg caggactggc cagccgacag ccgccccctc tacgagcgcc gccgcccgat gcggggctac cgcgacccgg
       >.....................................................LabT1.....................................................>
         i  a  a  l  a  v  a  g  r  g  v  l  t  g  r  l  s  a  g  e  m  l  a  a  a  g  y  a  p  m  a  l  g
```

```
18901  tggtcggcca gatccccctg ctgctggccc tggcgcggct gcgggcgggg gcgcgccgcc tggccgaggt gctgtcggtc cccgcccccg gcggcggcga
       accagccggt ctagggggac gacgaccggg accgcgccga cgcccgcccc cgcgcggcgg accggctcca cgacagccag gggcgggggc cgccgccgct
       >.....................................................LabT1.....................................................>
        l v g  q i p l  l l a  l a r  l r a g  a r r  l a e  v l s v  p a p  g g g

19001  ccgcgcgctg ccgcccggcc ccggcgcgct ggtcctgcgc ggggtgaccg tgcggacgcc ggacgggccg ctgctcgacg cggtcgacct gaccgtcccg
       ggcgcgcgac ggcgggccgg ggccgcgcga ccaggacgcg ccccactggc acgcctgcgg cctgcccggc gacgagctgc gccagctgga ctggcagggc
       >.....................................................LabT1.....................................................>
       d r a l  p p g  p g a  l v l r  g v t  v r t  p d g p  l l d  a v d  l t v p

19101  cccggccgga cggtggcggt ggtcggccgg tccggcgcgg gcaagacgac gctcgccgcc gtggccgggc gcctcctcga ccccgacgag gggcgggtgc
       gggccggcct gccaccgcca ccagccggcc aggccgcgcc cgttctgctg cgagcggcgg caccggcccg cggaggagct ggggctgctc cccgcccacg
       >.....................................................LabT1.....................................................>
         p g r  t v a  v v g r  s g a  g k t  t l a a  v a g  r l l  d p d e  g r v

19201  tcctggacgg ggtcccgctg cgcgagctcg cgcccgccgc gctgcgcgcg caggtcgcct acgccttcga acgccccgac ctccagggga ccacgatcgc
       aggacctgcc ccagggcgac gcgctcgagc gcgggcggcg cgacgcgcgc gtccagcgga tgcggaagct tgcggggctg gaggtcccct ggtgctagcg
       >.....................................................LabT1.....................................................>
        l l d  g v p l  r e l  a p a  a l r a  q v a  y a f  e r p d  l q g  t t i

19301  cgacgccatc gcctacggct gcccgtccgc gccgccccgc gcggtcgaac gggcggccgc gctcgcgcgg gcggacggct tcgtccggct gctcccggcg
       gctgcggtag cggatgccga cgggcaggcg cggcggggcg cgccagcttg cccgccggcg cgagcgcgcc cgcctgccga agcaggccga cgagggccgc
       >.....................................................LabT1.....................................................>
       a d a i  a y g  c p s  a p p r  a v e  r a a  a l a r  a d g  f v r  l l p a

19401  ggcttcgaca cgcccgtcgc ggacacgccc ctgtccggcg gggaacggca gcggctcggg ctggcccggg ccttcgtccg gaacgcccgc ctgatcatcc
       ccgaagctgt gcgggcagcg cctgtgcggg gacaggccgc cccttgccgt cgccgagccc gaccgggccc ggaagcaggc cttgcgggcg gactagtagg
       >.....................................................LabT1.....................................................>
         g f d  t p v  a d t p  l s g  g e r  q r l g  l a r  a f v  r n a r  l i i

19501  tcgacgacgc gacctccagc ctcgactcgg tgaccgaggc gcaggtggcg gccgcgctcg cggaggccgc cgcgacccgc acccggctgg tggtcgcgca
       agctgctgcg ctggaggtcg gagctgagcc actggctccg cgtccaccgc cggcgcgagc gcctccggcg gcgctgggcg tgggccgacc accagcgcgt
       >.....................................................LabT1.....................................................>
        l d d  a t s s  l d s  v t e  a q v a  a a l  a e a  a a t r  t r l  v v a

19601  ccgcgccggg accgccgcgc gcgccgacct ggtcctgtgg ctggactccg gccgggtgcg cgcgctccgc ccgcaccgcg agctgtgggc cgatccggac
       ggcgcggccc tggcggcgcg cgcggctgga ccaggacacc gacctgaggc cggcccacgc gcgcgaggcg ggcgtggcgc tcgacacccg gctaggcctg
       >.....................................................LabT1.....................................................>
       h r a g  t a a  r a d  l v l w  l d s  g r v  r a l r  p h r  e l w  a d p d

19701  taccgggcga tgttcgagcc ggcgccggcg gagcgcgcgt gacggggggc gcggcgcgct ggttcgccgc ccagctgcgg accgaacggc gcggcctggc
       atggcccgct acaagctcgg ccgcggccgc ctcgcgcgca ctgccccccg cgccgcgcga ccaagcggcg ggtcgacgcc tggcttgccg cgccggaccg
       >..................LabT1.................>>
         y r a  m f e  p a p  a e r  a  -
                                    >>............................LabT2.............................>
                                      v t g  g a a r  w f a  a q l r  t e r  r g l

19801  cggggtgctg gcgtggtcgg tcgcggcggc gctcccggcg ctggtgtccg gacggctgat cgcgctcgcc gtcgaccagg gcttcctgcg cggacgcggc
       gccccacgac cgcaccagcc agcgccgccg cgagggccgc gaccacaggc ctgccgacta gcgcgagcgg cagctggtcc cgaaggacgc gcctgcgccg
       >.....................................................LabT2.....................................................>
       a g v l  a w s  v a a  a l p a  l v s  g r l  i a l a  v d q  g f l  r g r g

19901  acggtcggcc tggcgtggct gggggcgctc gcggcggcga cggcggtcgg cgcgctgggc gcgcggcaga tcccgcgcgc gctcggcggg gccctggagc
       tgccagccgg accgcaccga cccccgcgag cgccgccgct gccgccagcc gcgcgacccg cgcgccgtct agggcgcgcg cgagccgccc cgggacctcg
       >.....................................................LabT2.....................................................>
         t v g  l a w  l g a l  a a a  t a v  g a l g  a r q  i p r  a l g g  a l e

20001  cggtgcgcga ccggctcgtc cgccggatcg tggccggcgc gctgcgccgc gcgctggccg gccggcccga tccgggcgtg gtcgcgaagc tgaccgagca
       gccacgcgct ggccgagcag gcggcctagc accggccgcg cgacgcggcg cgcgaccggc cggccgggct aggcccgcac cagcgcttcg actggctcgt
       >.....................................................LabT2.....................................................>
        p v r  d r l v  r r i  v a g  a l r r  a l a  g r p  d p g v  v a k  l t e

20101  gaccgagacc gtccgggaca ccggcgcggg gctgctgctc gggatgcagc aggtcggcat ggcggtggtc gcggccgcga tcgggctgct gctgctcgcg
       ctggctctgg caggccctgt ggccgcgccc cgacgacgag ccctacgtcg tccagccgta ccgccaccag cgccggcgct agcccgacga cgacgagcgc
       >.....................................................LabT2.....................................................>
       q t e t  v r d  t g a  g l l l  g m q  q v g  m a v v  a a a  i g l  l l l a

20201  cccgtcacgg cgctgctggt cctcccgccg gtgctggcgg cgctgctgct gctctcccgg ctgctgccgg agctgatggc gcggcaccgg gccctgctgg
       gggcagtgcc gcgacgacca ggagggcggc cacgaccgcc gcgacgacga cgagagggcc gacgacggcc tcgactaccg cgccgtggcc cgggacgacc
       >.....................................................LabT2.....................................................>
         p v t  a l l  v l p p  v l a  a l l  l l s r  l l p  e l m  a r h r  a l l

20301  cggccgagga ccgcgtcgcg gcgaccgtcg gcggcctgac cgcggcggtg cgcgacgtgg tcgcgtgcgg ggccgggccc cgggccgggg ccgaggcggc
       gccggctcct ggcgcagcgc cgctggcagc cgccggactg gcgccgccac gcgctgcacc agcgcacgcc ccggcccggg gcccggcccc ggctccgccg
       >.....................................................LabT2.....................................................>
        a a e  d r v a  a t v  g g l  t a a v  r d v  v a c  g a g p  r a g  a e a

20401  ggcggtcttc gccgagcagg ccgcggccga acgcgccgtc gcccgcgccg acgcgctgcg caccgcgctc ggcatcgtcg gcggccagct gtcgctggtc
       ccgccagaag cggctcgtcc ggcgccggct tgcgcggcag cgggcgcggc tgcgcgacgc gtggcgcgag ccgtagcagc cgccggtcga cagcgaccag
       >.....................................................LabT2.....................................................>
       a a v f  a e q  a a a  e r a v  a r a  d a l  r t a l  g i v  g g q  l s l v

20501  gtcctgctgg ccgccgcgcc ctggctggtg gcctccgggc ggctcagccc cggcgaggtg ctcggcgcgc tcacctacgt gacgaccggc ctgcaaccgg
       caggacgacc ggcggcgcgg gaccgaccac cggaggcccg ccgagtcggg gccgctccac gagccgcgcg agtggatgca ctgctggccg gacgttggcc
       >.....................................................LabT2.....................................................>
         v l l  a a a  p w l v  a s g  r l s  p g e v  l g a  l t y  v t t g  l q p

20601  cgctgcgcgc ggccgtgcag agcaccgggg gttccggcgt gcggctggcg gtgaccctgc aacgcctcct ggacggctcc gcggacgaac ccgaccccgc
       gcgacgcgcg ccggcacgtc tcgtggcccc caaggccgca cgccgaccgc cactgggacg ttgcggagga cctgccgagg cgcctgcttg ggctggggcg
       >.....................................................LabT2.....................................................>
        a l r  a a v q  s t g  g s g  v r l a  v t l  q r l  l d g s  a d e  p d p

20701  gcccgccgga gggcgcgtcc cgggccgcct cgacctgtcg gtggaggggc tcacgttcgc ctacggcgcc ggtgccgaac cggtcgtgcg cgacctggac
       cgggcggcct cccgcgcagg gcccggcgga gctggacagc cacctccccg agtgcaagcg gatgccgcgg ccacggcttg gccagcacgc gctggacctg
       >.....................................................LabT2.....................................................>
       a p a g  g r v  p g r  l d l s  v e g  l t f  a y g a  g a e  p v v  r d l d

20801  ctgaccgtgc cgcacgggac gcacctggcc gtcgtcggcc ccagcgggat cggcaagtcg accctggccg acctgctcac cggggtcgcc gtcccgcagg
       gactggcacg gcgtgccctg cgtggaccgg cagcagccgg ggtcgcccta gccgttcagc tgggaccggc tggacgagtg gccccagcgg cagggcgtcc
       >.....................................................LabT2.....................................................>
         l t v  p h g  t h l a  v v g  p s g  i g k s  t l a  d l l  t g v a  v p q
```

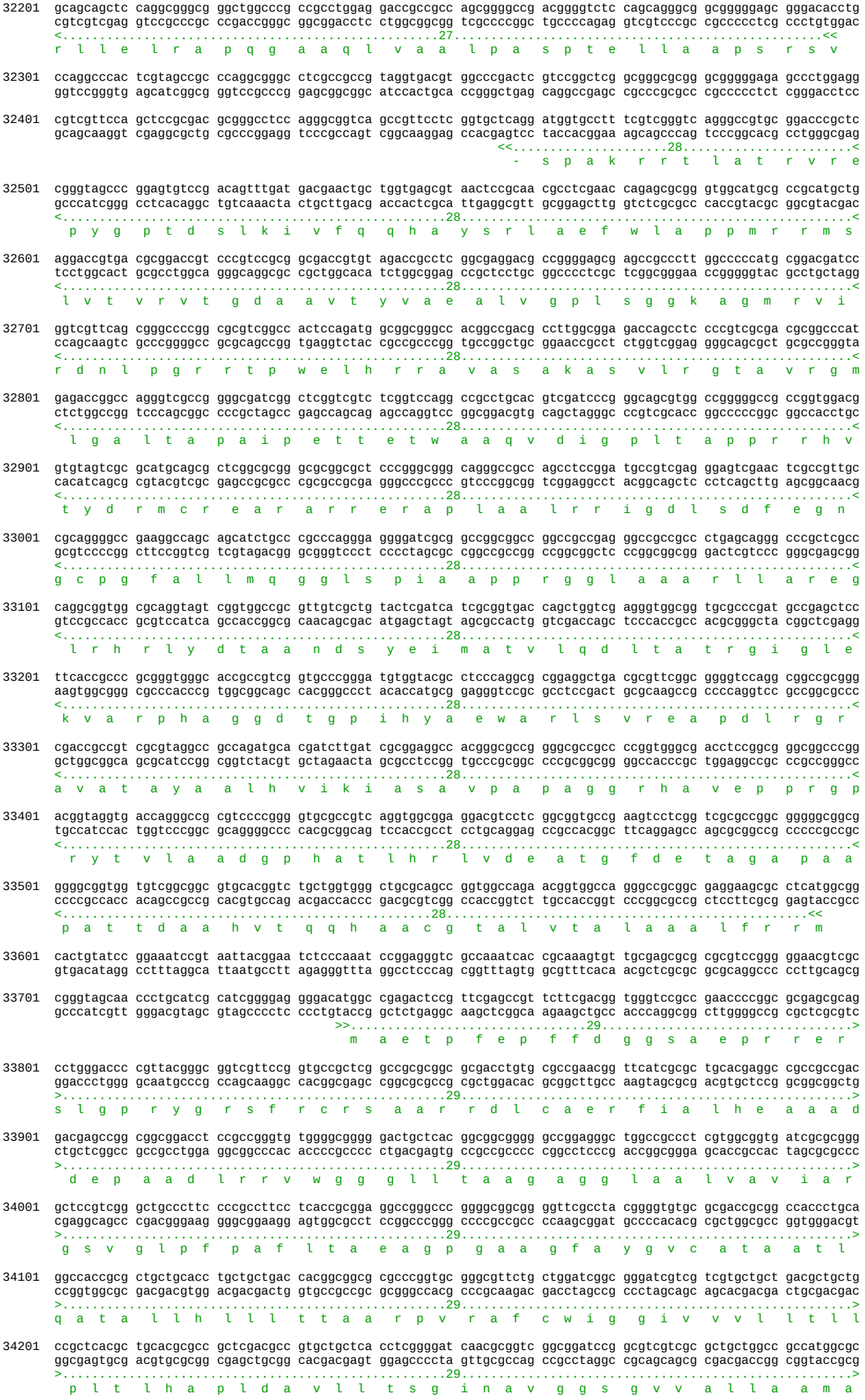
```
20901  cgggccgggt ccgcgtcggc ggcgtgccgc tcgccgagac cgatcccgcc tggaggcgcc ggtcggtcgc gatcatcccg caggaggcgt acgtcttcag
       gcccggccca ggcgcagccg ccgcacggcg agcggctctg gctagggcgg acctccgcgg ccagccagcg ctagtagggc gtcctccgca tgcagaagtc
       >.....................................................LabT2.....................................................>
        a  g  r  v  r  v  g  g  v  p  l  a  e  t  d  p  a  w  r  r  r  s  v  a  i  i  p  q  e  a  y  v  f

21001  cggcacgctc gccgagaacc tcgcctacct gcgcccggac gccgccgagg ccgagctcga cgccgccgtc gccgcgctgg gcctggcgcc gctgcgcgcc
       gccgtgcgag cggctcttgg agcggatgga cgcgggcctg cggcggctcc ggctcgagct gcggcggcag cggcgcgacc cggaccgcgg cgacgcgcgg
       >.....................................................LabT2.....................................................>
       s  g  t  l  a  e  n  l  a  y  l  r  p  d  a  a  e  a  e  l  d  a  a  v  a  a  l  g  l  a  p  l  r  a

21101  cggctcggcg ggtacggggc cgcgctcggg ccgcgttcgg gcctgacgcc gggggagcgg cagctcgtcg cgctcgcccg cgtctggctc agccccgcgc
       gccgagccgc ccatgccccg gcgcgagccc ggcgcaagcc cggactgcgg ccccctcgcc gtcgagcagc gcgagcgggc gcagaccgag tcggggcgcg
       >.....................................................LabT2.....................................................>
         r  l  g  g  y  g  a  a  l  g  p  r  s  g  l  t  p  g  e  r  q  l  v  a  l  a  r  v  w  l  s  p  a

21201  ggatcgtggt gctggacgag gcgacctgcc acctcgaccc cgccgccgag gcccgcgtgg agcacgcctt ccgggaccgg cccggcaccc tggtcgtcat
       cctagcacca cgacctgctc cgctggacgg tggagctggg gcggcggctc cgggcgcacc tcgtgcggaa ggccctggcc gggccgtggg accagcagta
       >.....................................................LabT2.....................................................>
        r  i  v  v  l  d  e  a  t  c  h  l  d  p  a  a  e  a  r  v  e  h  a  f  r  d  r  p  g  t  l  v  v

21301  cgcgcaccgg ctcagctcgg cgcgccgcgc cgaccgcgtc ctggtgatgg acgggacccg cccccggctg ggcacccacg gggaactcct ccgggactcc
       gcgcgtggcc gagtcgagcc gcgcggcgcg gctggcgcag gaccactacc tgccctgggc gggggccgac ccgtgggtgc cccttgagga ggccctgagg
       >.....................................................LabT2.....................................................>
       i  a  h  r  l  s  s  a  r  r  a  d  r  v  l  v  m  d  g  t  r  p  r  l  g  t  h  g  e  l  l  r  d  s

21401  ccgctctacg ccgacctggt cggggcctgg accgttcacc ccgcggcgcg gtagccgctc agccgggcgg ttcgagcccg tgctgccgca gccgcgcccg
       ggcgagatgc ggctggacca gccccggacc tggcaagtgg ggcgccgcgc catcggcgag tcggcccgcc aagctcgggc acgacggcgt cggcgcgggc
       >.........................LabT2.......................>>
         p  l  y  a  d  l  v  g  a  w  t  v  h  p  a  a  r  -

21501  ccgcaggtcg tcggcgctga tcatcaccgt cagcgtgctc gccaccagca ccagcagcgc gaccgccacg cgcgcccacc cttcggcggg cagcgccgcg
       ggcgtccagc agccgcgact agtagtggca gtcgcacgag cggtggtcgt ggtcgtcgcg ctggcggtgc gcgcgggtgg gaagccgccc gtcgcggcgc

21601  aacgccgcgc agaccggcgc catcagcccc aggtgccgga gcaggatccg cccgcgccat ccggcgtcgg tgaggtcgtg ccgcacccac tcccggttct
       ttgcggcgcg tctggccgcg gtagtcgggg tccacggcct cgtcctaggc gggcgcggta ggccgcagcc actccagcac ggcgtgggtg agggccaaga

21701  gcggcggcaa ccggaagccg agcgcgtaac gcacccttcc ccccactccc ggatccccgg gcagccgacc cataccgctc cccttccccg accggaggcc
       cgccgccgtt ggccttcggc tcgcgcattg cgtgggaagg ggggtgaggg cctaggggcc cgtcggctgg gtatggcgag gggaaggggc tggcctccgg

21801  cccgcgctta ggacatgaag cgtcctaacg cccggttacc ttccccgcat gacccacacc agcgacacct tcctcgtgac cggcgccacc gggaccgtcg
       gggcgcgaat cctgtacttc gcaggattgc gggccaatgg aaggggcgta ctgggtgtgg tcgctgtgga aggagcactg gccgcggtgg ccctggcagc
                                                  >>.......................20.......................>
                                                    m  t  h  t  s  d  t  f  l  v  t  g  a  t  g  t  v

21901  gccgccagat cgtccgtgag ctcctcgaac gcggccacgc cgtccgcgcg ctgacccgcg accccgcgaa cgcctccctt cccgggggcg tcgaggccgt
       cggcggtcta gcaggcactc gaggagcttg cgccggtgcg gcaggcgcgc gactgggcgc tggggcgctt gcggagggaa gggcccccgc agctccggca
       >.....................................................20.....................................................>
        g  r  q  i  v  r  e  l  l  e  r  g  h  a  v  r  a  l  t  r  d  p  a  n  a  s  l  p  g  g  v  e  a

22001  ccggggcgac ctgaccgacc ccgcgagcct ggcgcccgcc ctcgacggcg ccgccgggct ccacctgatc accttcggcg gcggctacga gccgctggag
       ggccccgctg gactggctgg ggcgctcgga ccgcgggcgg gagctgccgc ggcggcccga ggtggactag tggaagccgc cgccgatgct cggcgacctc
       >.....................................................20.....................................................>
       v  r  g  d  l  t  d  p  a  s  l  a  p  a  l  d  g  a  a  g  l  h  l  i  t  f  g  g  g  y  e  p  l  e

22101  accggcccgg agatcgtcga gctggcccgc ggggcgggcg tgcgccgggt caccgtcctc aacggcggcg gccccacccc gctccaggac gcggtccgcg
       tggccgggcc tctagcagct cgaccgggcg ccccgcccgc acgcggccca gtggcaggag ttgccgccgc cggggtgggg cgaggtcctg cgccaggcgc
       >.....................................................20.....................................................>
         t  g  p  e  i  v  e  l  a  r  g  a  g  v  r  r  v  t  v  l  n  g  g  g  p  t  p  l  q  d  a  v  r

22201  cgagcgacct ggagtggacc gtggtggcgc cggtcgagtt catggccaac gcgctggact gggccgagtc cgtccgcacc gaggacaccg tgcgcgcggc
       gctcgctgga cctcacctgg caccaccgcg gccagctcaa gtaccggttg cgcgacctga cccggctcag gcaggcgtgg ctcctgtggc acgcgcgccg
       >.....................................................20.....................................................>
        a  s  d  l  e  w  t  v  v  a  p  v  e  f  m  a  n  a  l  d  w  a  e  s  v  r  t  e  d  t  v  r  a

22301  gttcgtcgac cgccacagcg ccatggtcca cgaggccgac atcggcgcgg tcgccgcgac ggccctcacc gccccggcc acgcgggcga ggagtacgtg
       caagcagctg gcggtgtcgc ggtaccaggt gctccggctg tagccgcgcc agcggcgctg ccgggagtgg cgggggccgg tgcgcccgct cctcatgcac
       >.....................................................20.....................................................>
       a  f  v  d  r  h  s  a  m  v  h  e  a  d  i  g  a  v  a  a  t  a  l  t  a  p  g  h  a  g  e  e  y  v

22401  gtcaccggcc cgcaggccct gaccgtccgc gaccaggtcg tggccgtcgc cgccgcccgg ggccgcgaga tccgcctcgt cgagctgagc gaggacgagg
       cagtggccgg gcgtccggga ctggcaggcg ctggtccagc accggcagcg gcggcgggcc ccggcgctct aggcggagca gctcgactcg ctcctgctcc
       >.....................................................20.....................................................>
         v  t  g  p  q  a  l  t  v  r  d  q  v  v  a  v  a  a  a  r  g  r  e  i  r  l  v  e  l  s  e  d  e

22501  ccgtggcccg gtggcgcgcg gagggccacg acgaggagac gatcgcctgg ttcgtcgaga ccctcaagaa caccccgccg gtcggccgca cggtcgtcga
       ggcaccgggc caccgcgcgc ctcccggtgc tgctcctctg ctagcggacc aagcagctct gggagttctt gtggggcggc cagccggcgt gccagcagct
       >.....................................................20.....................................................>
        a  v  a  r  w  r  a  e  g  h  d  e  e  t  i  a  w  f  v  e  t  l  k  n  t  p  p  v  g  r  t  v  v

22601  caccgtcgaa cgcgtcaccg gccgcccggc ccgcaccttc gcccggtggg ccgccgagaa cgcctcggcg ttcaccgcct cctgaccagg cggtcagtcc
       gtggcagctt gcgcagtggc cggcgggccg ggcgtggaag cgggccaccc ggcggctctt gcggagccgc aagtggcgga ggactggtcc gccagtcagg
       >.....................................20.....................................>>
       d  t  v  e  r  v  t  g  r  p  a  r  t  f  a  r  w  a  a  e  n  a  s  a  f  t  a  s  -

22701  cgcgccgcgg cctcgatcgc ccgcttcatc tccaccgcgc cgcgttcgac gcggatccgg atctcggcgg cgagcgcgtg gggcagcacg tcggtggtcc
       gcgcggcgcc ggagctagcg ggcgaagtag aggtggcgcg gcgcaagctg cgcctaggcc tagagccgcc gctcgcgcac cccgtcgtgc agccaccagg

22801  agatcaggcg ccccgcccgc gcgccctcga ccaccgcgta ggccaggcgg tgcgcctcgt cgtcgacgtc gaggatcagt tcgcggatca cgcgcccgtc
       tctagtccgc ggggcgggcg cgcgggagct ggtggcgcat ccggtccgcc acgcggagca gcagctgcag ctcctagtca agcgcctagt gcgcgggcag

22901  gggaaaggcc aggaaccgct ggtcgccctc gatccgggtc ccggcgacac ggccgggcag cagccgctcg tgcacggcgc cgacgtcgcg cagcacgtcc
       ccctttccgg tccttggcga ccagcgggag ctaggcccag ggccgctgtg ccggcccgtc gtcggcgagc acgtgccgcg gctgcagcgc gtcgtgcagg

23001  cagacgtgcc cgggggacgc gtcgatcacg atcccgtggc ggatggaggc caagacgccg gggctcctct cgggcggaac ggatcaccgg cgcccaaccc
       gtctgcacgg gccccctgcg cagctagtgc tagggcaccg cctacctccg gttctgcggc cccgaggaga gcccgccttg cctagtggcc gcgggttggg

23101  agccgcaagt gtccggctcc ggaccggagg tggaagatca cccgaaacct gtggagaccg cccttggacg gcgccggagc acctaaagtg atcttgtctt
       tcggcgttca caggccgagg cctggcctcc accttctagt gggctttgga cacctctggc gggaacctgc cgcggcctcg tggatttcac tagaacagaa

23201  gtgaaccaag aggaagacgg tgcgcatgcg ttcatctctc gggctggcct gcgggctggc gatggccgcg gtcctgctgc ccgcccccgc gtacgccgcc
       cacttggttc tccttctgcc acgcgtacgc aagtagagag cccgaccgga cgcccgaccg ctaccggcgc caggacgacg ggcgggggcg catgcggcgg
                          >>.....................................21.....................................>
                            v  r  m  r  s  s  l  g  l  a  c  g  l  a  m  a  a  v  l  l  p  a  p  a  y  a  a
```

```
23301  gagccgtcgg cgcaggtcat cgaggccgcg acgggggtca agggggtcgt gcggccctcc ggcggcggcg cggaggaggt cggggcggtc acgacgcgtg
       ctcggcagcc gcgtccagta gctccggcgc tgcccccagt tccccccagca cgccgggagg ccgccgccgc gcctcctcca gccccgccag tgctgcgcac
       >.....................................................21.....................................................>
         e  p  s  a  q  v  i  e  a  a  t  g  v  k  g  v  v  r  p  s  g  g  g  a  e  e  v  g  a  v  t  t  r

23401  acaaggtgtg gggctcgacg acgacggtgc ggatcccccg ggacgcccgg cgcggggtgg ccgtcgccga cagcggttcg ggctcggtgg tgctggggct
       tgttccacac cccgagctgc tgctgccacg cctaggggggc cctgcgggcc gcgccccacc ggcagcggct gtcgccaagc ccgagccacc acgaccccga
       >.....................................................21.....................................................>
        d  k  v  w  g  s  t  t  t  v  r  i  p  r  d  a  r  r  g  v  a  v  a  d  s  g  s  g  s  v  v  l  g

23501  gccgggcggc ggccgtccca ccgtcagcgc gaaggggacg gtcgtctacc ccgggcccgg gttccggacg gcggtgcaga ccacgaccga cgggaccgtg
       cggcccgccg ccggcagggt ggcagtcgcg cttcccctgc cagcagatgg ggcccgggcc caaggcctgc cgccacgtct ggtgctggct gccctggcac
       >.....................................................21.....................................................>
       l  p  g  g  g  r  p  t  v  s  a  k  g  t  v  v  y  p  g  p  g  f  r  t  a  v  q  t  t  t  d  g  t  v

23601  cgaggcctgg tcgaggccaa ggggccgggg acgttccggt tcccgctgac ggtcccgcgg ggcgcccggc tcacgcggaa cgccgacggc agcctcgacg
       gctccggacc agctccggtt ccccggcccc tgcaaggcca agggcgactg ccagggcgcc ccgcgggccg agtgcgcctt gcggctgccg tcggagctgc
       >.....................................................21.....................................................>
         r  g  l  v  e  a  k  g  p  g  t  f  r  f  p  l  t  v  p  r  g  a  r  l  t  r  n  a  d  g  s  l  d

23701  tcatgcgggc agggacgccc gtcggtcaca tcgacgcgcc ctgggccagg agcgccgggg gcgggaaggt ccccgcctcg ttcggggtcg agggcgccac
       agtacgcccg tccctgcggg cagccagtgt agctgcgcgg gacccggtcc tcgcggcccc cgcccttcca ggggcggagc aagccccagc tcccgcggtg
       >.....................................................21.....................................................>
        v  m  r  a  g  t  p  v  g  h  i  d  a  p  w  a  r  s  a  g  g  g  k  v  p  a  s  f  g  v  e  g  a

23801  cgtcgtgctg aacctcggcc ggggcggcgc gtccccggtc gtggcgggcc tggcctacca cgcggctccc ggcaagggcg gctaccgggg cgcggcgacc
       gcagcacgac ttggagccgg ccccgccgcg caggggccag caccgcccgg accggatggt gcgccgaggg ccgttcccgc cgatggcccc gcgccgctgg
       >.....................................................21.....................................................>
       t  v  v  l  n  l  g  r  g  g  a  s  p  v  v  a  g  l  a  y  h  a  a  p  g  k  g  g  y  r  g  a  a  t

23901  atcgacgcca cgaagaggac caggacctgg accgggaagg agatgcggca ggcccagccg ctcgacgcgc ccgaggtcgc ccccggccgg accgaccgct
       tagctgcggt gcttctcctg gtcctggacc tggcccttcc tctacgccgt ccgggtcggc gagctgcgcg ggctccagcg ggggccggcc tggctggcga
       >.....................................................21.....................................................>
         i  d  a  t  k  r  t  r  t  w  t  g  k  e  m  r  q  a  q  p  l  d  a  p  e  v  a  p  g  r  t  d  r

24001  ccgggaccgg gcggacggcg cccgcgaccg ccgccgccac cgaccccgcc gtgggcgcga tgttcttcag cttcctgggc atcaagaagc gctgcacggc
       ggccctggcc cgcctgccgc gggcgctggc ggcggcggtg gctggggcgg cacccgcgct acaagaagtc gaaggacccg tagttcttcg cgacgtgccg
       >.....................................................21.....................................................>
        s  g  t  g  r  t  a  p  a  t  a  a  a  t  d  p  a  v  g  a  m  f  f  s  f  l  g  i  k  k  r  c  t

24101  cacgatcgtc caccgcaacc tgatcgcgac ggcggcccac tgcatccagc agggcggcgt catcttcaac cagtggacct tcgccccgca gtacgacaac
       gtgctagcag gtggcgttgg actagcgctg ccgccgggtg acgtaggtcg tcccgccgca gtagaagttg gtcacctgga agcggggcgt catgctgttg
       >.....................................................21.....................................................>
       a  t  i  v  h  r  n  l  i  a  t  a  a  h  c  i  q  q  g  g  v  i  f  n  q  w  t  f  a  p  q  y  d  n

24201  ggcgccaccc cctacggctc gtggaagatc aagacgatct actacgacca gcggtggggc aagctgaaca acaaccccga cttcgactac gcgttcatct
       ccgcggtggg ggatgccgag caccttctag ttctgctaga tgatgctggt cgccaccccg ttcgacttgt tgttggggct gaagctgatg cgcaagtaga
       >.....................................................21.....................................................>
         g  a  t  p  y  g  s  w  k  i  k  t  i  y  y  d  q  r  w  g  k  l  n  n  n  p  d  f  d  y  a  f  i

24301  ccctgcaacc gcggaacggc aagtacgtcg gggacgtggt gccgtggggc cgcgcggaca tcaccgacct gtcgtggcgc ggcgcctgct cgggcgacaa
       gggacgttgg cgccttgccg ttcatgcagc ccctgcacca cggcacccgg gcgcgcctgt agtggctgga cagcaccgcg ccgcggacga gcccgctgtt
       >.....................................................21.....................................................>
        s  l  q  p  r  n  g  k  y  v  g  d  v  v  p  w  g  r  a  d  i  t  d  l  s  w  r  g  a  c  s  g  d

24401  ctgcccggag atccccaccc acgtcaccgg ataccccaag aaccgcacca cccctctgga gtgcttcacc cggagccgcc tggtctacat ctccgagaac
       gacgggcctc taggggtggg tgcagtggcc tatggggttc ttggcgtggt ggggagacct cacgaagtgg gcctcggcgg accagatgta gaggctcttg
       >.....................................................21.....................................................>
       n  c  p  e  i  p  t  h  v  t  g  y  p  k  n  r  t  t  p  l  e  c  f  t  r  s  r  l  v  y  i  s  e  n

24501  ggtgagaagg ggagctacta ctacaccttc ccgtgcaagg gctacgccga gggcgtcagc ggcacggccc tggtcgggca gttccccgag cagggccacc
       ccactcttcc cctcgatgat gatgtggaag ggcacgttcc cgatgcggct cccgcagtcg ccgtgccggg accagcccgt caaggggctc gtcccggtgg
       >.....................................................21.....................................................>
         g  e  k  g  s  y  y  y  t  f  p  c  k  g  y  a  e  g  v  s  g  t  a  l  v  g  q  f  p  e  q  g  h

24601  ccggctacga ggtcatcttc ggcgtgctgg gcgggtacca gggaggcggc gacaccgacg acgtctctgta cgcctcgtac tgggtcgacg gcaaggtgct
       ggccgatgct ccagtagaag ccgcacgacc cgcccatggt ccctccgccg ctgtggctgc tgcagagcat gcggagcatg acccagctgc cgttccacga
       >.....................................................21.....................................................>
        p  g  y  e  v  i  f  g  v  l  g  g  y  q  g  g  g  d  t  d  d  v  s  y  a  s  y  w  v  d  g  k  v

24701  gttcgacgac gcaaagagca aggccgtcta ggtcctgggg gccggacgcc gagggtcccg cgagccgacg gctcgcggga ccctcgtccg tccacgggcc
       caagctgctg cgtttctcgt tccggcagat ccaggacccc cggcctgcgg ctcccagggc gctcggctgc cgagcgccct gggagcaggc aggtgcccgg
       >.............21.............>>
       l  f  d  d  a  k  s  k  a  v  -

24801  atggcaccat cgaccgacgg gaccgtgccg aggaggtacc gctgaacgac gcgcccaccg gacacagcgc cgggctccac gaactggcgg cgatcctcgc
       taccgtggta gctggctgcc ctggcacggc tcctccatgg cgacttgctg cgcgggtggc ctgtgtcgcg gcccgaggtg cttgaccgcc gctaggagcg

24901  cgacgccgcc cgccatctcg cctccgccga ccgcgccatc gaccgggccc tgcgcgagcg cacccgcgcc gtcggcctca tccgcccccg ggtccgcgcc
       gctgcggcgg gcggtagagc ggaggcggct ggcgcggtag ctggcccggg acgcgctcgc gtgggcgcgg cagccggagt aggcgggggc ccaggcgcgg

25001  gccgcccgcg cccgcgcccg cctccgcgcc gtcgaccggg cactcgaccg cgccccgggc ctcgtcccgg ccgggtccgg cgacccgccc ttcgacgccg
       cggcgggcgc gggcgcgggc ggaggcgcgg cagctggccc gtgagctggc gcggggcccg gagcagggcc ggcccaggcc gctgggcggg aagctgcggc

25101  ccctcctctc cttccgcgtc gacgccctcg cccgcgcgct ctgctccgtc ccctcctacg agctccccgc cgaccgttcg gagatccgcg gctggatccg
       gggaggagag gaaggcgcag ctgcgggagc gggcgcgcga gacgaggcag gggaggatgc tcgaggggcg gctggcaagc ctctaggcgc cgacctaggc

25201  cctgctcggc cgggcgctgg gcctggacat cggtctcgtc ccccgcgtcg gacccgacgg cgtccgcggc ctccagggcg atctcgccga ggcggtgcag
       ggacgagccg gcccgcgacc cggacctgta gccagagcag ggggcgcagc ctgggctgcc gcaggcgccg gaggtcccgc tagagcggct ccgccacgtc

25301  cacgcgcggg cccgcgtgga ggcgctgctg aacgcccggg acgagcccgg ccgcgaggac ggggagaacg acaccgccgg gtaggaggcg cgaacgtgac
       gtgcgcgccc gggcgcacct ccgcgacgac ttgcgggccc tgctcgggcc ggcgctcctg cccctcttgc tgtggcggcc catcctccgc gcttgcactg

25401  ccgccgttcg ccatgccggt ggccgagggg gcggcggccg gtaggccccc gccggttccc cgtcgggggc ctaccgcctg tccaggccct ggccgggcgc
       ggcggcaagc ggtacggcca ccggctcccc cgccgccggc catccggggg cggccaaggg gcagcccccg gatggcggac aggtccggga ccggcccgcg

25501  ggctcaggct aggtcagggc cttggccatg ttcgaggcga acttctccca tgtgctgtgg gcgccggtcc tgttcgccca gtccgggacc gacgggttgg
       ccgagtccga tccagtcccg gaaccggtac aagctccgct tgaagagggt acacgacacc cgcggccagg acaagcgggt caggccctgg ctgcccaacc
              <<.....................................22.....................................<
                 -  t  l  a  k  a  m  n  s  a  f  k  e  w  t  s  h  a  g  t  r  n  a  w  d  p  v  s  p  n

25601  ccacgtaggc gtagacgtgc aggcgcagga ccttcttctt gtcctggtac cagcggaagt agaccgtgga gcccttgccg tcgaagtgcc ccttctgcgc
       ggtgcatccg catctgcacg tccgcgtcct ggaagaagaa caggaccatg gtcgccttca tctggcacct cgggaacggc agcttcacgg ggaagacgcg
       <.....................................................22.....................................................<
       a  v  y  a  y  v  h  l  r  l  v  k  k  k  d  q  y  w  r  f  y  v  t  s  g  k  g  d  f  h  g  k  q  a
```

```
25701  gacgaagtac cagccgtact ttccccgggg gtagaacttg accggggcgt tgcacacctt cttgaaccag caggccttga gcggcagctt ctgccccctc
       ctgcttcatg gtcggcatga aaggggcccc catcttgaac tggccccgca acgtgtggaa gaacttggtc gtccggaact cgccgtcgaa gacgggggag
       <.....................................................22.....................................................<
         v  f  y  w  g  y  k  g  r  p  y  f  k  v  p  a  n  c  v  k  k  f  w  c  a  k  l  p  l  k  q  g  r

25801  ttggggtact tcttcggcgc gcccgaaacc gggaaggtgc agttgaagca gctcttcagc ttcttcatcg ccttggacgg gtgctttttg gcgaccgtgt
       aaccccatga agaagccgcg cgggctttgg cccttccacg tcaacttcgt cgagaagtcg aagaagtagc ggaacctgcc cacgaaaaac cgctggcaca
       <.....................................................22.....................................................<
        k  p  y  k  k  p  a  g  s  v  p  f  t  c  n  f  c  s  k  l  k  k  m  a  k  s  p  h  k  k  a  v  t

25901  acgtgatgca cctgaaggtg gaaggaaagc agtggcccgc cgtgaaggtg tagctgtaga gcttgcgcag cttcttcgcc ttgacttcgg tcgccacgga
       tgcactacgt ggacttccac cttcctttcg tcaccgggcg gcacttccac atcgacatct cgaacgcgtc gaagaagcgg aactgaagcc agcggtgcct
       <.....................................................22.....................................................<
       y  t  i  c  r  f  t  s  p  f  c  h  g  a  t  f  t  y  s  y  l  k  r  l  k  k  a  k  v  e  t  a  v  s

26001  agccggtgtc ggggcggcgg caggcgccgc ctgcgcgatc gcaggggtga ccggcttccg tgcctgggcg gcggcgcccg tgccgggggt ggcgagcacc
       tcggccacag ccccgccgcc gtccgcggcg gacgcgctag cgtccccact ggccgaaggc acggacccgc cgccgcgggc acggccccca ccgctcgtgg
       <.....................................................22.....................................................<
         a  p  t  p  a  a  a  p  a  a  q  a  i  a  p  t  v  p  k  r  a  q  a  a  a  g  t  g  p  t  a  l  v

26101  agcgcgccgg aggcgacggt cgccaccaag acctttctca aaaggctcat gcaaccttcc ttacgagggg agtcttcttt ccgggaggga gagcgtcggg
       tcgcgcggcc tccgctgcca gcggtggttc tggaaagagt tttccgagta cgttggaagg aatgctcccc tcagaagaaa ggccctccct ctcgcagccc
       <.....................................................22.....................................................<
        l  a  g  s  a  v  t  a  v  l  v  k  r  l  l  s  m  c  g  e  k  r  p  s  d  e  k  r  s  p  s  r  r

26201  accggaccgt tcgacgtgac ccgagcctcc ccgcacctgg ttgggaacag gttggaatcc ggttggagag ccgccggccg gtgggtcgga tacctcggga
       tggcctggca agctgcactg ggctcggagg ggcgtggacc aacccttgtc caaccttagg ccaacctctc ggcggccggc cacccagcct atggagccct
       <.....................................................22.....................................................<
       s  r  v  t  r  r  s  g  l  r  g  a  g  p  q  s  c  t  p  i  r  n  s  l  r  r  g  t  p  r  i  g  r  s

26301  catggggaaa acggcacggg ctgatcacgc ccattctcga tattccaggg ggaatcggtg ctgcccggat ccacggccgg cgcggccgcc gtgctgtgcg
       gtaccccttt tgccgtgccc gactagtgcg ggtaagagct ataaggtccc ccttagccac gacgggccta ggtgccggcc gcgccggcgg cacgacacgc
       <<< 22
         m

26401  ccgtcatggt gggacccgtg ctcacgacgg ggaacgccgc cgcccacacc cggcaggccg cccggaccgt ccagaccgcc caagccgtcc gggctccgta
       ggcagtacca ccctgggcac gagtgctgcc ccttgcggcg gcgggtgtgg gccgtccggc gggcctggca ggtctggcgg gttcggcagg cccgaggcat

26501  cgaccgggtg ctgtcgaacc cgtcgctcgc ccgcgccacc agggcgttcg tcagggcggc ggcgcggcgg cggcagcggg tgacgagggt cgtcacggag
       gctggcccac gacagcttgg gcagcgagcg ggcgcggtgg tcccgcaagc agtcccgccg ccgcgccgcc gccgtcgccc actgctccca gcagtgcctc

26601  cggtccggaa cgatccggct ggtcgtgcgc acccgccagc ggggccgcag acctacgagc cggccaagcg caaggcctac accgccgtga cgttcggtcg
       gccaggcctt gctaggccga ccagcacgcg tgggcggtcg ccccggcgtc tggatgctcg gccggttcgc gttccggatg tggcggcact gcaagccagc

26701  gcccaccagc cggctcgtcg gcaacgccac catcaggcag ctacccggca ccctggcgct gcccggcggc gtcccggtca agcgccgcgg cttccgggtc
       cgggtggtcg gccgagcagc cgttgcggtg gtagtccgtc gatgggccgt gggaccgcga cgggccgccg cagggccagt tcgcggcgcc gaaggcccag

26801  gcgtgcatcg gggtgggcgg cgcccccgac ggccggatcg acgaggcgat cgcccgcacg gtccgggcgg ccctgcgcgg ctgaccctgc gcggctgacc
       cgcacgtagc cccacccgcc gcgggggctg ccggcctagc tgctccgcta gcgggcgtgc caggcccgcc gggacgcgcc gactgggacg cgccgactgg

26901  ctgcgcggct gaccctcgcg gcacatcccg gccgggccgg tcagccggtc cggccgtcca gaggcgacca cccgccggtg cggcggaggg gagggaagcc
       gacgcgccga ctgggagcgc cgtgtagggc cggccccggcc agtcggccag gccggcaggt ctccgctggt gggcggccac gccgcctccc ctcccttcgg
                                          <<.............................23.............................<
                                            -  g  t  r  g  d  l  p  s  w  g  g  t  r  r  l  p  p  f  g

27001  cgtcacggac gtccagccgg tgcggggcgg ccactgcccc cggggcggct ccgggtggac gccgacgacg gccagcggcg gcaggtcgtc ccggtgccac
       gcagtgcctg caggtcggcc acgccccgcc ggtgacgggg gccccgccga ggcccacctg cggctgctgc cggtcgccgc cgtccagcag ggccacggtg
       <.....................................................23.....................................................<
         t  v  s  t  w  g  t  r  p  p  w  q  g  r  p  p  e  p  h  v  g  v  v  a  l  p  p  l  d  d  r  h  w

27101  agctcgcgcc actggtcgcg ggtcaccacc aggtggcgga gcccggacag gtcggcgatc gtggcgagat cgacggtcgc ggacgagacg tccagcgccc
       tcgagcgcgg tgaccagcgc ccagtggtgg tccaccgcct cgggcctgtc cagccgctag caccgctcta gctgccagcg cctgctctgc aggtcgcggg
       <.....................................................23.....................................................<
        l  e  r  w  q  d  r  t  v  v  l  h  r  l  g  s  l  d  a  i  t  a  l  d  v  t  a  s  s  v  d  l  a

27201  acaggcgggg aacggcgcgc agcggggcca ggtcgacggg ggtcccggac gcgaccgtca gcgaacgcag cccggcgtgg tcggccaacg gcgccagatc
       tgtccgcccc ttgccgcgcg tcgccccggt ccagctgccc ccagggcctg cgctggcagt cgcttgcgtc gggccgcacc agccggttgc cgcggtctag
       <.....................................................23.....................................................<
       w  l  r  p  v  a  r  l  p  a  l  d  v  p  t  g  s  a  v  t  l  s  r  l  g  a  h  d  a  l  p  a  l  d

27301  gaccgactcc agttcgagtt ccaggacctc gaccggcagg gcgcgcagcg gggccaggtc agggcggccg ggaccggaca ggaccgcctt gcgtacgttg
       ctggctgagg tcaagctcaa ggtcctggag ctggccgtcc cgcgcgtcgc cccggtccag tcccgccggc cctggcctgt cctggcggaa cgcatgcaac
       <.....................................................23.....................................................<
         v  s  e  l  e  l  e  l  v  e  v  p  l  a  r  l  p  a  l  d  p  r  g  p  g  s  l  v  a  k  r  v  n

27401  ggcagcgcct tgaggaagcc gaggtcctcc acatcggtca cccggaggtc ctggacgcgg ggacggacct ggaacgcgtc gatgctccgc gcgtcggtgt
       ccgtcgcgga actccttcgg ctccaggagg tgtagccagt gggcctccag gacctgcgcc cctgcctgga ccttgcgcag ctacgaggcg cgcagccaca
       <.....................................................23.....................................................<
        p  l  a  k  l  f  g  l  d  e  v  d  t  v  r  l  d  q  v  r  p  r  v  q  f  a  d  i  s  r  a  d  t

27501  acagggtgcg cgcgtcctcc accttcctct gggagggcag gtcgggctcg ctcctgaggt cgtcgccacg gcgcgtgtgg tccccccgct ccagcgcctc
       tgtcccacgc gcgcaggagg tggaaggaga ccctcccgtc cagcccgagc gaggactcca gcagcggtgc cgcgcacacc agggggggca ggtcgcggag
       <.....................................................23.....................................................<
       y  l  t  r  a  d  e  v  k  r  q  s  p  l  d  p  e  s  r  l  d  d  g  r  r  t  h  d  g  r  e  l  a  e

27601  cacgagccgc cgcaggaacg tggtgaccga gtcggcgacg tacaggggcg ctttccggaa gccgacgccg atcttgatga cctgtccggg acgcccgccg
       gtgctcggcg gcgtccttgc accactggct cagccgctgc atgtccccgc gaaaggcctt cggctgcggc tagaactact ggacaggccc tgcgggcggc
       <.....................................................23.....................................................<
         v  l  r  r  l  f  t  t  v  s  d  a  v  y  l  p  a  k  r  f  g  v  g  i  k  i  v  q  g  p  r  g  g

27701  gcggcggggt ccatgtccac ggcgagccag ttgccgccgg tgtcgtgcgc gaaccggatc cacccgggcc gcaggagcga gcgccgcacg gtatgcgcgg
       cgccgcccca ggtacaggtg ccgctcggtc aacggcggcc acagcacgcg cttggcctag gtgggcccgg cgtcctcgct cgcggcgtgc catacgcgcc
       <.....................................................23.....................................................<
        a  a  p  d  m  d  v  a  l  w  n  g  g  t  d  h  a  f  r  i  w  g  p  r  l  l  s  r  r  v  t  h  a

27801  gaacggcgtc gagcacgggc gcgcgttgcg gctcgtgctc ccattcccgg tcggcccact cgtcgtcctc ctcgcccacc tccccggcgc ggatccaggg
       cttgccgcag ctcgtgcccg cgcgcaacgc cgagcacgag ggtaagggcc agccgggtga gcagcaggag gagcgggtgg aggggccgcg cctaggtccc
       <.....................................................23.....................................................<
       p  v  a  d  l  v  p  a  r  q  p  e  h  e  w  e  r  d  a  w  e  d  d  e  e  g  v  e  g  a  r  i  w  p

27901  cttccggtcg aacagcgggg tgacgagggt gccgtcgccg tcggcgacgc ggtagagggc gcgcaggtcg gcgggcaggg cgacaccgag cctgtcctcc
       gaaggccagc ttgtcgcccc actgctccca cggcagcggc agccgctgcg ccatctcccg cgcgtccagc cgcccgtccc gctgtggctc ggacaggagg
       <.....................................................23.....................................................<
         k  r  d  f  l  p  t  v  l  t  g  d  g  d  a  v  r  y  l  a  r  l  d  a  p  l  a  v  g  l  r  d  e
```

```
28001  aggtcgtcca gctccgcgtc gtcggcgggc ggcggcagcc gttcctcgtg gccgaggatc tccgcgcgcc gccgcatgag ctcgcgcagg agccgcaccg
       tccagcaggt cgaggcgcag cagccgcccg ccgccgtcgg caaggagcac cggctcctag aggcgcgcgg cggcgtactc gagcgcgtcc tcggcgtggc
       <.....................................................23.....................................................<
        l  d  d   l  e  a  d   d  a  p   p  p  l   r  e  e  h   g  l  i   e  a  r   r  r  m  l   e  r  l   l  r  v

28101  cctcccgcgg atcgccggcc ggggcggcgg acgccggggc gtcctgctcg tcgccccggt cgggcggaac gaagtcctcg tcgcagacgt agagggtcgg
       ggagggcgcc tagcggccgg ccccgccgcc tgcggccccg caggacgagc agcggggcca gcccgccttg cttcaggagc agcgtctgca tctcccagcc
       <.....................................................23.....................................................<
       a  e  r  p   d  g  a   p  a  a   s  a  p  a   d  q  e   d  g  r   d  p  p  v   f  d  e   d  c  v   y  l  t  p

28201  ccgctcgtgc tcgacgggct cggcgggctc gccgacggtc gcctcgaacc ggcccgccgc ctccacgtcg agcaccaggt gcaccggcct gcggcccacg
       ggcgagcacg agctgcccga gccgcccgag cggctgccag cggagcttgg ccgggcggcg gaggtgcagc tcgtggtcca cgtggccgga cgccgggtgc
       <.....................................................23.....................................................<
         r  e  h   e  v  p   e  a  p  e   g  v  t   a  e  f   r  g  a  a   e  v  d   l  v  l   h  v  p  r   r  g  v

28301  agcctgacga cgtccagggc ccggtccgac aggtcgagcg cgacgcggcg gctctcgccg ccggcgagcg tgaaggaggg ttcgccgtga acggcgccgt
       tcggactgct gcaggtcccg ggccaggctg tccagctcgc gctgcgccgc cgagagcggc ggccgctcgc acttcctccc aagcggcact tgccgcggca
       <.....................................................23.....................................................<
        l  r  v   v  d  l  a   r  d  s   l  d  l   a  v  r  r   s  e  g   g  a  l   t  f  s  p   e  g  h   v  a  g

28401  cgtcgcggac gtgccagcgc aggcgcgcgc gttcccagcc ggggggtgcg gcggcggcga tccgcgccgc gagttcgtcc aggacggggg gagctgtcat
       gcagcgcctg cacggtcgcg tccgcgcgcg caagggtcgg ccccccacgc cgccgccgct aggcgcggcg ctcaagcagg tcctgccccc ctcgacagta
       <.....................................................23.....................................................<<
       d  d  r  v   h  w  r   l  r  a   r  e  w  g   p  p  a   a  a  a   i  r  a  a   l  e  d   l  v  p   p  a  t  m

28501  gctgccgcaa cctagcggac gactccgaca ccgcccgccc ggtgggcacg ggcatggtaa aggaattttg aacctcttcg gctgtcctgg aggaccaagg
       cgacggcgtt ggatcgcctg ctgaggctgt ggcgggcggg ccacccgtgc ccgtaccatt tccttaaaac ttggagaagc cgacaggacc tcctggttcc

28601  aaaaggggga acgggaaccc ggcgtgtcct ggtatccctt tctgactggt cggtcattgt ccgcgggaaa acgcgtgccc ccggcgaagg cggtcacgct
       ttttccccct tgcccttggg ccgcacagga ccatagggaa agactgacca gccagtaaca ggcgcccttt tgcgcacggg ggccgcttcc gccagtgcga

28701  cagcgatcaa ctcgcgcctt gacggacggc caatccgata tttagaaatg actatgtaat gcgctctgtc gagagtgacg agaactgctc ttggtgagct
       gtcgctagtt gagcgcggaa ctgcctgccg gttaggctat aaatctttac tgatacatta cgcgagacag ctctcactgc tcttgacgag aaccactcga

28801  ttcgttcacc ggcgcgaacg ctgtcggtcg gtgacgctcc gcgcggggga tcatgtgctc agcattcact tcacggcgga cgaccttgcc cggatccggg
       aagcaagtgg ccgcgcttgc gacagccagc cactgcgagg cgcgcccccc agtacacgag tcgtaagtga agtgccgcct gctggaacgg gcctaggccc
                                    >>.....................................24.....................................>
                                      v  t  l   r  a  g   d  h  v  l   s  i  h   f  t  a   d  d  l   a  r  i  r

28901  tgtcgggcgc ccgcccggca tgggagacgg tcctcagcgc gcggtcgctg cgcgaccgtg cgcgccaccc cgtgctgggg agctggcgtg cgcggacgct
       acagcccgcg ggcgggccgt accctctgcc aggagtcgcg cgccagcgac gcgctggcac gcgcggtggg gcacgacccc tcgaccgcac gcgcctgcga
       >.....................................................24.....................................................>
        v  s  g   a  r  p  a   w  e  t   v  l  s   a  r  s  l   r  d  r   a  r  h   p  v  l  g   s  w  r   a  r  t

29001  ggaggtgatg ccggccgcca gccggatgct gctcgcgctg tgcccgcccg ccggaagatt cgcggaattc ctcatatccg agaggggcgc gcgcgaatgg
       cctccactac ggccggcggt cggcctacga cgagcgcgac acgggcgggc ggccttctaa gcgccttaag gagtataggc tctccccgcg cgcgcttacc
       >.....................................................24.....................................................>
       l  e  v  m   p  a  a   s  r  m   l  l  a  l   c  p  p   a  g  r   f  a  e  f   l  i  s   e  r  g   a  r  e  w

29101  gaggggggatc tcgacgcgat cctggggacc ccgccggacc ggatcaggaa tgatcttgtc gggctcgcgt ggcggggccg cccgcccgaa tgggccgccg
       ctccccctag agctgcgcta ggaccctgg ggcggcctgg cctagtcctt actagaacag cccgagcgca ccgccccggc gggcgggctt acccggcggc
       >.....................................................24.....................................................>
         e  g  d   l  d  a   i  l  g  t   p  p  d   r  i  r   n  d  l  v   g  l  a   w  r  g   r  p  p  e   w  a  a

29201  ccctcgcggc cggggaggga aaggcgctgc acgccctcgg ccgggcgctg cgcgattacc accggatcgg aatagcgccc tattgggagg aaattcaggc
       gggagcgccg gcccctccct ttccgcgacg tgcgggagcc ggcccgcgac gcgctaatgg tggcctagcc ttatcgcggg ataaccctcc tttaagtccg
       >.....................................................24.....................................................>
        a  l  a   a  g  e  g   k  a  l   h  a  l   g  r  a  l   r  d  y   h  r  i   g  i  a  p   y  w  e   e  i  q

29301  gcgggtggaa gtcgaccgcg ctttccggct ccgcgtcctc gccgacggcg gggtggagca gctgctggcc acgctcgaca cggccgtccg gtgggactcg
       cgcccacctt cagctggcgc gaaaggccga ggcgcaggag cggctgccgc cccacctcgt cgacgaccgg tgcgagctgt gccggcaggc caccctgagc
       >.....................................................24.....................................................>
       a  r  v  e   v  d  r   a  f  r   l  r  v  l   a  d  g   g  v  e   q  l  l  a   t  l  d   t  a  v   r  w  d  s

29401  gggaccctgg agctgccgca ggccggtgag cggcgcgtgc ggctggacgg gcggggcgtc gtgctggccc cctcggtctt ctgcggccgc caggcggtga
       ccctgggacc tcgacggcgt ccggccactc gccgcgcacg ccgacctgcc cgccccgcag cacgaccggg ggagccagaa gacgccggcg gtccgccact
       >.....................................................24.....................................................>
         g  t  l   e  l  p   q  a  g  e   r  r  v   r  l  d   g  r  g  v   v  l  a   p  s  v   f  c  g   r  q  a  v

29501  tcagggaacg gtccgaccgg ccgccggtgc tgttcttccc ggtcgtgatg gggttcggca ggccgctcgc ggtgacgggc ggggacgacc gccggccgga
       agtcccttgc caggctggcc ggcggccacg acaagaaggg ccagcactac cccaagccgt ccggcgagcg ccactgcccg cccctgctgg cggccggcct
       >.....................................................24.....................................................>
        i  r  e   r  s  d  r   p  p  v   l  f  f   p  v  v  m   g  f  g   r  p  l   a  v  t  g   g  d  d   r  r  p

29601  cgccctgtcg gcgctcctgg ggcggacgcg cgccgccgcc ctgcaggagg cggccgacgg cacgaccacc acggagctgg cccgccggct ggaggtgggc
       gcgggacagc cgcgaggacc ccgcctgcgc gcggcggcgg gacgtcctcc gccggctgcc gtgctggtgg tgcctcgacc gggcggccga cctccacccg
       >.....................................................24.....................................................>
       d  a  l  s   a  l  l   g  r  t   r  a  a  a   l  q  e   a  a  d   g  t  t  t   t  e  l   a  r  r   l  e  v  g

29701  ctgtccacgg ccagcgagca cgtgaaggtg ctgcgcgagg ccgggctgat cgcgacgggc cggatggggc gcgaggcggt gcacacggtg acgccgctgg
       gacaggtgcc ggtcgctcgt gcacttccac gacgcgctcc ggcccgacta gcgctgcccg gcctaccccg cgctccgcca cgtgtgccac tgcggcgacc
       >.....................................................24.....................................................>
         l  s  t   a  s  e   h  v  k  v   l  r  e   a  g  l   i  a  t  g   r  m  g   r  e  a   v  h  t  v   t  p  l

29801  gggcggcgct gtcgcgcacc gggtgctgcg cgcggccgtc ggtccccgac cccggccggt gagccccgcg ggccgggcgc ccgcgcggac ctgccgggcg
       cccgccgcga cagcgcgtgg cccacgacgc gcgccggcag ccaggggctg gggccggcca ctcggggcgc ccggcccgcg ggcgcgcctg gacggcccgc
       >.............................24.............................>>
        g  a  a   l  s  r  t   g  c  c   a  r  p   s  v  p  d   p  g  r   -

29901  gggcgcgcgc ggcgcgcgcg gcgcgcgtca gactgtcggg catggcggaa ctggacgggg aaccggtcgg ggtcgagcag ctcgcggcgt tggcgctggt
       cccgcgcgcg ccgcgcgcgc cgcgcgcagt ctgacagccc gtaccgcctt gacctgcccc ttggccagcc ccagctcgtc gagcgccgca accgcgacca
                                          >>.....................25.....................>
                                            m  a  e   l  d  g   e  p  v   g  v  e  q   l  a  a   l  a  l

30001  caactacggg cacttcacca cgatgcgggt cgagggcggg ggcgtgcgcg ggctcgcgct ccatctggac cggctcacgc gcgactgccg ggcgctgttc
       gttgatgccc gtgaagtggt gctacgccca gctcccgccc ccgcacgcgc ccgagcgcga ggtagacctg gccgagtgcg cgctgacggc ccgcgacaag
       >.....................................................25.....................................................>
       v  n  y  g   h  f  t   t  m  r   v  e  g  g   g  v  r   g  l  a   l  h  l  d   r  l  t   r  d  c   r  a  l  f

30101  ggcgtcgagc cggacgtcga gcgggtgcgg cgcctcgcgc ggcgggccgc cgccggagcc gccgcgcccg tcgtcgtccg ggtcaccctc ttcgaccccg
       ccgcagctcg gcctgcagct cgcccacgcc gcggagcgcg ccgcccggcg gcggcctcgg cggcgcgggc agcagcaggc ccagtgggag aagctggggc
       >.....................................................25.....................................................>
         g  v  e   p  d  v   e  r  v  r   r  l  a   r  r  a   a  a  g  a   a  a  p   v  v  v   r  v  t  l   f  d  p
```

```
30201  gcctggacat cgggcggccc gtgagcaacg gcgagccccg cgtgctcgtc accgcgcggg ccgccgggga gcggccggtg ccggggatga gggtcgggac
       cggacctgta gcccgccggg cactcgttgc cgctcggggc gcacgagcag tggcgcgccc ggcggccccct cgccggccac ggcccctact cccagccctg
       >.....................................................25.....................................................>
        g  l  d  i  g  r  p  v  s  n  g  e  p  r  v  l  v  t  a  r  a  a  g  e  r  p  v  p  g  m  r  v  g

30301  cgtccggtac gcgcgggacc tgcccgcggt caagcacgtg gggctgttcg gggccgtgcg gcggcggagg gacgcgcagg ccgaggggtt cgacgacgcg
       gcaggccatg cgcgccctgg acgggcgcca gttcgtgcac cccgacaagc cccggcacgc cgccgcctcc ctgcgcgtcc ggctccccaa gctgctgcgc
       >.....................................................25.....................................................>
       t  v  r  y  a  r  d  l  p  a  v  k  h  v  g  l  f  g  a  v  r  r  r  r  d  a  q  a  e  g  f  d  d  a

30401  ctgttcgtgg acgccggcgc gtccgtcacc gaggggacca cctggaacgt cggcttcttc gacgggcggc gggtcgtgtg gcccgacgcc gaggtgctgc
       gacaagcacc tgcggccgcg caggcagtgg ctcccctggt ggaccttgca gccgaagaag ctgcccgccg cccagcacac cgggctgcgg ctccacgacg
       >.....................................................25.....................................................>
         l  f  v  d  a  g  a  s  v  t  e  g  t  t  w  n  v  g  f  f  d  g  r  r  v  v  w  p  d  a  e  v  l

30501  ccggggtgac gatgcggctg ctgcgggggа gcgtccggcc ggtgggggtc gccgacctcg ccgggatgga ggccgcgttc gccaccaacg cggccgtggg
       ggccccactg ctacgccgac gacgcccccт cgcaggccgg ccacccccag cggctggagc ggccctacct ccggcgcaag cggtggttgc gccggcaccc
       >.....................................................25.....................................................>
        p  g  v  t  m  r  l  l  r  g  s  v  r  p  v  g  v  a  d  l  a  g  m  e  a  a  f  a  t  n  a  a  v

30601  ggtgcgggcg atcaccgccg tggacgggcg ggcgttcccc gccgaccatc ccgtgctcga cgcgctgcgc cgggagtacg cggcggtccc gctcgacccg
       ccacgcccgc tagtggcggc acctgcccgc ccgcaagggg cggctggtag ggcacgagct gcgcgacgcg gccctcatgc gccgccaggg cgagctgggc
       >.....................................................25.....................................................>
       g  v  r  a  i  t  a  v  d  g  r  a  f  p  a  d  h  p  v  l  d  a  l  r  r  e  y  a  a  v  p  l  d  p

30701  ctctgacccc gccggaggcc cgaaaagtgg gagtgtgacc ggcttctcat cgacgttcgg cctggtcagg gttcgtttaa cgtgccgttc acaaatgggc
       gagactgggg cggcctccgg gcttttcacc ctcacactgg ccgaagagta gctgcaagcc ggaccagtcc caagcaaatt gcacggcaag tgtttacccg
       >.25>>
         l  -

                                                                                                  >>.......26........>
                                                                                                    v  p  f  t  n  g

30801  aacaggtccg aaatgccgta ttgccactct catcacggcc gagacgacgg accggccgcg ggaagactaa ggagtcgacg atgagctaca aggccgaata
       ttgtccaggc tttacggcat aacggtgaga gtagtgccgg ctctgctgcc tggccggcgc ccttctgatt cctcagctgc tactcgatgt tccggcttat
       >.....................................................26.....................................................>
        q  q  v  r  n  a  v  l  p  l  s  s  r  p  r  r  r  t  g  r  g  k  t  k  e  s  t  m  s  y  k  a  e

30901  catctgggtc gacggcaccg agcccaccgc gcggctccgc tccaagacca ggatccttgc cgacggcgcc gagctgcccg actggggctt cgacggctcc
       gtagacccag ctgccgtggc tcgggtggcg cgccgaggcg aggttctggt cctaggaacg gctgccgcgg ctcgacgggc tgaccccgaa gctgccgagg
       >.....................................................26.....................................................>
       y  i  w  v  d  g  t  e  p  t  a  r  l  r  s  k  t  r  i  l  a  d  g  a  e  l  p  d  w  g  f  d  g  s

31001  agcaccaacc aggcccccgg tgacaactcc gactgcgtgc tcaagccggt cttcgtcgta ccggacccgc tgcgcggcgg cgagaacaag ctggtgctgt
       tcgtggttgg tccgggggcc actgttgagg ctgacgcacg agttcggcca gaagcagcat ggcctgggcg acgcgccgcc gctcttgttc gaccacgaca
       >.....................................................26.....................................................>
         s  t  n  q  a  p  g  d  n  s  d  c  v  l  k  p  v  f  v  v  p  d  p  l  r  g  g  e  n  k  l  v  l

31101  gcgaggtgta cctggtggac ggcaacccgc accccaccaa caaccgcgcc aggctgcgcc cggtggccga gaagtacgcc gcccacgacg cctggttcgg
       cgctccacat ggaccacctg ccgttgggcg tggggtggtt gttggcgcgg tccgacgcgg gccaccggct cttcatgcgg cgggtgctgc ggaccaagcc
       >.....................................................26.....................................................>
        c  e  v  y  l  v  d  g  n  p  h  p  t  n  n  r  a  r  l  r  p  v  a  e  k  y  a  a  h  d  a  w  f

31201  catcgagcag gagtacacct tcttccaggc cggccggccg ctcggcttcc ccgagaccgg tttccccgcc ccgcagggcc cgtactactg cggcgtgggc
       gtagctcgtc ctcatgtgga agaaggtccg gccggccggc gagccgaagg ggctctggcc aaaggggcgg ggcgtcccgg gcatgatgac gccgcacccg
       >.....................................................26.....................................................>
       g  i  e  q  e  y  t  f  f  q  a  g  r  p  l  g  f  p  e  t  g  f  p  a  p  q  g  p  y  y  c  g  v  g

31301  gccgacgagg tcttcggccg cgagatcgtc gagcgccacc tggacgcctg cctggaggcc gggatcaagg tctccggcat caacgccgag gtcatgcccg
       cggctgctcc agaagccggc gctctagcag ctcgcggtgg acctgcggac ggacctccgg ccctagttcc agaggccgta gttgcggctc cagtacgggc
       >.....................................................26.....................................................>
         a  d  e  v  f  g  r  e  i  v  e  r  h  l  d  a  c  l  e  a  g  i  k  v  s  g  i  n  a  e  v  m  p

31401  gccagtggga gttccagatc ggcccggccg gcccgatcga ggtcgccgac cacctgtggg tcgcgcgctg gctgctctac cgcatcgccg aggagtacga
       cggtcaccct caaggtctag ccgggccggc cgggctagct ccagcggctg gtggacaccc agcgcgcgac cgacgagatg gcgtagcggc tcctcatgct
       >.....................................................26.....................................................>
        g  q  w  e  f  q  i  g  p  a  g  p  i  e  v  a  d  h  l  w  v  a  r  w  l  l  y  r  i  a  e  e  y

31501  cgtcgccgcc accctcgacc ccaagcccgc caagggcgac tggaacggcg ccggcgcgca caccaacttc tccaccaagg ccatgcgcga gggctacgac
       gcagcggcgg tgggagctgg ggttcgggcg gttcccgctg accttgccgc ggccgcgcgt gtggttgaag aggtggttcc ggtacgcgct cccgatgctg
       >.....................................................26.....................................................>
       d  v  a  a  t  l  d  p  k  p  a  k  g  d  w  n  g  a  g  a  h  t  n  f  s  t  k  a  m  r  e  g  y  d

31601  gccatcctca ccgcctgcga cgcgctcggc gccgagggca aggtcgagga gcacatcgcg ggctacggcg cgggcatcca ggagcgcctg accggccagc
       cggtaggagt ggcggacgct gcgcgagccg cggctcccgt tccagctcct cgtgtagcgc ccgatgccgc gcccgtaggt cctcgcggac tggccggtcg
       >.....................................................26.....................................................>
         a  i  l  t  a  c  d  a  l  g  a  e  g  k  v  e  e  h  i  a  g  y  g  a  g  i  q  e  r  l  t  g  q

31701  acgagaccgc cccgtggaac gtctacagct acggcgtctc cgaccgcggc gcgtcggtgc gcatcccgtg gcaggtccag gtggacaaga agggctacat
       tgctctggcg gggcaccttg cagatgtcga tgccgcagag gctggcgccg cgcagccacg cgtagggcac cgtccaggtc cacctgttct tcccgatgta
       >.....................................................26.....................................................>
        h  e  t  a  p  w  n  v  y  s  y  g  v  s  d  r  g  a  s  v  r  i  p  w  q  v  q  v  d  k  k  g  y

31801  cgaggaccgc cgcccgaacg ccaacatcga cccctacaac gtcacccgcc tgctggtgaa cacctgctgc gaggccctgg agaaggccgg ccaggtctga
       gctcctggcg gcgggcttgc ggttgtagct ggggatgttg cagtgggcgg acgaccactt gtggacgacg ctccgggacc tcttccggcc ggtccagact
       >.....................................................26.....................................................>>
       i  e  d  r  r  p  n  a  n  i  d  p  y  n  v  t  r  l  l  v  n  t  c  c  e  a  l  e  k  a  g  q  v  -

31901  tccggcccgc cgagggcatg cgccgcggcc cggtcccccg cccgggggcc gggccgcggc gcgttcccgg gcgccggcgc cctcaggagc cggtgcccgc
       aggccgggcg gctcccgtac gcggcgccgg gccaggggcc gggcccccgg cccggcgccg cgcaagggcc cgcggccgcg ggagtcctcg gccacgggcg
                                                                                                  <<......27........<
                                                                                                    -  s  g  t  g  a

32001  ggtgatgtcg aaggaggccg ccagcagggc caggacgggg aagacgacgc gcggcggcat ccggtaggtg gcccgctcgg actgctccac caggcccgcg
       ccactacagc ttcctccggc ggtcgtcccg gtcctgcccc ttctgctgcg cgccgccgta ggccatccac cgggcgagcc tgacgaggtg gtccgggcgc
       <.....................................................27.....................................................<
         t  i  d  f  s  a  a  l  l  a  l  v  p  f  v  v  r  p  p  m  r  y  t  a  r  e  s  q  e  v  l  g  a

32101  gccgacaggg ccttcagatg gtggtagagc tgcccggtcg agcccaggtt cagctcggcc tgcaggtccg ccaccgtgcg ggggcccacc agcagcaggc
       cggctgtccc ggaagtctac caccatctcg acgggccagc tcgggtccaa gtcgagccgg acgtccaggc ggtggcacgc ccccgggtgg tcgtcgtccg
       <.....................................................27.....................................................<
        a  s  l  a  k  l  h  h  y  l  q  g  t  s  g  l  n  l  e  a  q  l  d  a  v  t  r  p  g  v  l  l  l
```



```
34301  gcgccgcgcg ggacgggttc ccctaacggc ccacgctgat cgtcaggggc tcgggcgcgc gggccgtggt gacccggacc gggaccaggc ggcggcggac
       cgcggcgcgc cctgcccaag gggattgccg ggtgcgacta gcagtccccg agcccgcgcg cccggcacca ctgggcctgg ccctggtccg ccgccgcctg
       >...........29..........>>
        r  a  a  r  d  g  f  p  -
                        <<.....................................30.....................................<
                          -  r  g  v  s  i  t  l  p  e  p  a  r  a  t  t  v  r  v  p  v  l  r  r  r  v

34401  gtgcagctcc accgggtcgg cgggggcgaa gaccagcagg tcgcggtgca tcccggcgaa acccaggacc aggtgggccg cccaggtgcc gcggggcagc
       cacgtcgagg tggcccagcc gcccccgctt ctggtcgtcc agcgccacgt agggccgctt tgggtcctgg tccacccggc gggtccacgg cgccccgtcg
       <.....................................................30.....................................................<
         h  l  e  v  p  d  a  p  a  f  v  l  l  d  r  h  m  g  a  f  g  l  v  l  h  a  a  w  t  g  r  p  l

34501  gggtcgtcga cgaacgcgcg ggtcagcggg accgccgcgg tgtagcccag ccggtccggg cccgcctcca gcatcgcgat cacctcgtag gtcgcgcccc
       cccagcagct gcttgcgcgc ccagtcgccc tggcggcgcc acatcgggtc ggccaggccc gggcggaggt cgtagcgcta gtggagcatc cagcgcgggg
       <.....................................................30.....................................................<
        p  d  d  v  f  a  r  t  l  p  v  a  a  t  y  g  l  r  d  p  g  a  e  l  m  a  i  v  e  y  t  a  g

34601  cgcgctcgtg cagcaccagc gtcgccgaga ccggcgtctc ggcggcgggc acggcgaggt ggccgcgcac gaccagctcc tcgtcgtcgg cgttccaggc
       gcgcgagcac gtcgtggtcg cagcggctct ggccgcagag ccgccgcccg tgccgctcca ccggcgcgtg ctggtcgagg agcagcagcc gcaaggtccg
       <.....................................................30.....................................................<
       g  r  e  h  l  v  l  t  a  s  v  p  t  e  a  a  p  v  a  l  h  g  r  v  v  l  e  e  d  d  a  n  w  a

34701  gagcgcgtcg gccgtggtcg atccggcggg atggggcgcg ccgccgacgt ccagcgccag cccgccgccc cgcggggcga agtaggcgat caccatggtc
       ctcgcgcagc cggcaccagc taggccgccc taccccgcgc ggcggctgca ggtcgcggtc gggcggcggg gcgccccgct tcatccgcta gtggtaccag
       <.....................................................30.....................................................<
         l  a  d  a  t  t  s  g  a  p  h  p  a  g  g  v  d  l  a  l  g  g  g  r  p  a  f  y  a  i  v  m  t

34801  gagtccggca ggaaccggcg ccgcggcatg ctgtccaggc cggggtggta ggcgagcggc gcggtggcgg ggccggcgcg cagggacacg tcccacaggc
       ctcaggccgt ccttggccgc ggcgccgtac gacaggtccg gccccaccat ccgctcgccg cgccaccgcc ccggccgcgc gtccctgtgc agggtgtccg
       <.....................................................30.....................................................<
        s  d  p  l  f  r  r  r  p  m  s  d  l  g  p  h  y  a  l  p  a  t  a  p  g  a  r  l  s  v  d  w  l

34901  cccgggggag ggcgccggcc gcgacggacg cggcgaaacc ggcgtcccgg acctcggcgg gcagctcgtg gacgcggtcg ccgtcgcgtt cgcgcaggac
       gggccccctc ccgcggccgg cgctgcctgc gccgctttgg ccgcagggcc tggagccgcc cgtcgagcac ctgcgccagc ggcagcgcaa gcgcgtcctg
       <.....................................................30.....................................................<
       g  r  p  l  a  g  a  a  v  s  a  a  f  g  a  d  r  v  e  a  p  l  e  h  v  r  d  g  d  r  e  r  l  v

35001  cagcgcgacg tcggtggcgg gtccgggcag cgtgccccgc agccgcagcg tggccccgtc ccagcgcacc tcgtccagaa gcggatcgga catcaggccc
       gtcgcgctgc agccaccgcc caggcccgtc gcacggggcg tcggcgtcgc accggggcag ggtcgcgtgg agcaggtctt cgcctagcct gtagtccggg
       <.....................................................30.....................................................<<
         l  a  v  d  t  a  p  g  p  l  t  g  r  l  r  l  t  a  g  d  w  r  v  e  d  l  l  p  d  s  m

35101  ccctgccccg tcgtttccgc ggaccatcct gccgcacggg atcaggaacg cttgaagcgg ggcagcaggt cggccaggtc gagcgcgccc gagccgaccg
       gggacggggc agcaaaggcg cctggtagga cggcgtgccc tagtccttgc gaacttcgcc ccgtcgtcca gccggtccag ctcgcgcggg ctcggctggc
                                          <<.............................31.............................<
                                            -  s  r  k  f  r  p  l  l  d  a  l  d  l  a  g  s  g  v

35201  cggcgtacag ctggtcgacg tcggcgtagc ccagggcgcg ggccaccgcc agggagtcgc cgtgcgcctc ggccgacagc aggtcgaccc ggcccgccag
       gccgcatgtc gaccagctgc agccgcatcg ggtcccgcgc ccggtggcgg tccctcagcg gcacgcggag ccggctgtcg tccagctggg ccgggcggtc
       <.....................................................31.....................................................<
       a  a  y  l  q  d  v  d  a  y  g  l  a  r  a  v  a  l  s  d  g  h  a  e  a  s  l  l  d  v  r  g  a  l

35301  ggcgcgggcc aggcggccgc ggccggtctc ggccgcctcc tcggcgcggc gccgggccag ccaccgctgg acgtggacgc gcgcgggggc ggtcaccgcg
       ccgcgcccgg tccgccggcg ccggccagag ccggcggagg agccgcgccg cggcccggtc ggtggcgacc tgcacctgcg cgcgcccccg ccagtggcgc
       <.....................................................31.....................................................<
         a  r  a  l  r  g  r  g  t  e  a  a  e  e  a  r  r  r  a  l  w  r  q  v  h  v  r  a  p  a  t  v  a

35401  aacgacagcc actcctcgga ggggacgccc tcggggtcgg tgaggacctc caccacgttc ccgctgcgca gctccgccga cagcgggacc agccgcccgt
       ttgctgtcgg tgaggagcct cccctgcggg agccccagcc actcctggag gtggtgcaag ggcgacgcgt cgaggcggct gtcgccctgg tcggcgggca
       <.....................................................31.....................................................<
        f  s  l  w  e  e  s  p  v  g  e  p  d  t  l  v  e  v  v  n  g  s  r  l  e  a  s  l  p  v  l  r  g

35501  tgaccagggc gccgatgcag tggccgccgg tggcggcgcc cagcgcgtag gcgaagtcca gcgcggtcgc cccgtccggc agcgtgatca cctcaccggt
       actggtcccg cggctacgtc accggcggcc accgccgcgg gtcgcgcatc cgcttcaggt cgcgccagcg gggcaggccg tcgcactagt ggagtggcca
       <.....................................................31.....................................................<
       n  v  l  a  g  i  c  h  g  g  t  a  a  g  l  a  y  a  f  d  l  a  t  a  g  d  p  l  t  i  v  e  g  t

35601  cggggtgaac accgcgacgc tgccctccgc caggtcggcg cgcaggccgt ccaggaactc cgccgacggg acgcccgact gccaggccag caggcggctc
       gccccacttg tggcgctgcg acgggaggcg gtccagccgc gcgtccggca ggtccttgag gcggctgccc tgcgggctga cggtccggtc gtccgccgag
       <.....................................................31.....................................................<
         p  t  f  v  a  v  s  g  e  a  l  d  a  r  l  g  d  l  f  e  a  s  p  v  g  s  q  w  a  l  l  r  s

35701  agccacacca ggtcgcgccg cccggccacg gtgctggtcg cctcgtcggg cgcggaccgg atgtgcgcga cgatcccgta ctcggcgacc gggtgcatgt
       tcggtgtggt ccagcgcggc gggccggtgc cacgaccagc ggagcagccc gcgcctggcc tacacgcgct gctagggcat gagccgctgg cccacgtaca
       <.....................................................31.....................................................<
        l  w  v  l  d  r  r  g  a  v  t  s  t  a  e  d  p  a  s  r  i  h  a  v  i  g  y  e  a  v  p  h  m

35801  cccggctgcg gatgatgacc tcgatcgggt cgccgtccgg ggtgatcacg cgggtgtgca gcgaccggta catgttgtac ttcggcagcg cgatgaagtc
       gggccgacgc ctactactgg agctagccca gcggcaggcc ccactagtgc gcccacacgt cgctggccat gtacaacatg aagccgtcgc gctacttcag
       <.....................................................31.....................................................<
       d  r  s  r  i  i  v  e  i  p  d  g  d  p  t  i  v  r  t  h  l  s  r  y  m  n  y  k  p  l  a  i  f  d

35901  gcggacctgg ccggggatcg ggtgcatcgc cgcgtggacc gcgccgagcg cgacgtagca ctcctgctcg gcgccgtcca ccaccagcag cagccgcgcc
       cgcctggacc ggcccctagc ccacgtagcg gcgcacctgg cgcggctcgc gctgcatcgt gaggacgagc cgcggcaggt ggtggtcgtc gtcggcgcgg
       <.....................................................31.....................................................<
         r  v  q  g  p  i  p  h  m  a  a  h  v  a  g  l  a  v  y  c  e  q  e  a  g  d  v  v  l  l  l  r  a

36001  gcctcgcagg ggcgcagccc ggtcacgtcc ccgtccaggg cccggtgcac cgcgtacagg tgcgaggggc ggatctcgac gcgtccggtg acgccgtgct
       cggagcgtcc ccgcgtcggg ccagtgcagg ggcaggtccc gggccacgtg gcgcatgtcc acgctccccg cctagagctg cgcaggccac tgcggcacga
       <.....................................................31.....................................................<
        a  e  c  p  r  l  g  t  v  d  g  d  l  a  r  h  v  a  y  l  h  s  p  r  i  e  v  r  g  t  v  g  h

36101  cggccagcga ggcgcggacg agccgccgcg cgggggcgaa cgcggccggc gccccgcgca gcgccagggt cacggcccgc tcggtggcgg cgtgggcggc
       gccggtcgct ccgcgcctgc tcggcggcgc gcccccgctt gcgccggccg cggggcgcgt cgcggtccca gtgccgggcg agccaccgcc gcacccgccg
       <.....................................................31.....................................................<
       e  a  l  s  a  r  v  l  r  r  a  p  a  f  a  a  p  a  g  r  l  a  l  t  v  a  r  e  t  a  a  h  a  a

36201  cgggtcggcg gcggcgaacg ccaggtcgtc catctcccgc ttcagcacgt ggatgcccag ccgctcggcg aacggcacca gcagctcgtg cgacgccttg
       gcccagccgc cgccgcttgc ggtccagcag gtagagggcg aagtcgtgca cctacgggtc ggcgagccgc ttgccgtggt cgtcgagcac gctgcggaac
       <.....................................................31.....................................................<
         p  d  a  a  a  f  a  l  d  d  m  e  r  k  l  v  h  i  g  l  r  e  a  f  p  v  l  l  e  h  s  a  k
```

```
36301  gcgatgcgcg cacgcttgtg cggcggctgg aagcccaggg tgcgcaggtt gtgcagccgg tccgccagct tgatcaccag gacgcgcaga tcgtccgccg
       cgctacgcgc gtgcgaacac gccgccgacc ttcgggtccc acgcgtccaa cacgtcggcc aggcggtcga actagtggtc ctgcgcgtct agcaggcggc
       <.....................................................31.....................................................<
        a  i  r  a  r  k  h  p  p  q  f  g  l  t  r  l  n  h  l  r  d  a  l  k  i  v  l  v  r  l  d  d  a

36401  ccgccagcag gatcttgcgg aacgtctcgg cctcggcgcg ctcgccccac cgcttcccgt ccagcttggt gaccccgtcc accagcaccg cgacccgctc
       ggcggtcgtc ctagaacgcc ttgcagagcc ggagccgcgc gagcggggtg gcgaagggca ggtcgaacca ctggggcagg tggtcgtggc gctgggcgag
       <.....................................................31.....................................................<
       a  a  l  l  i  k  r  f  t  e  a  e  a  r  e  g  w  r  k  g  d  l  k  t  v  g  d  v  l  v  a  v  r  e

36501  gccgaagtcg gagcggacct cgccgagcgt gtaggggggtg tcctcgacgg tgtcgtgcag cagcgccgcc accagcgtgg tggtgtccat ccccatctcc
       cggcttcagc ctcgcctgga gcggctcgca catcccccac aggagctgcc acagcacgtc gtcgcggcgg tggtcgcacc accacaggta ggggtagagg
       <.....................................................31.....................................................<
         g  f  d  s  r  v  e  g  l  t  y  p  t  d  e  v  t  d  h  l  l  a  a  v  l  t  t  t  d  m  g  m  e

36601  gccaggatca tcgcgaccgc gagcggatgg gtgatgtagg gcgcgcccga cttgcggcgc tggtcccggt gcaggcgctc ggcgacctcg tagccgcgca
       cggtcctagt agcgctggcg ctcgcctacc cactacatcc cgcgcgggct gaacgccgcg accagggcca cgtccgcgag ccgctggagc atcggcgcgt
       <.....................................................31.....................................................<
        a  l  i  m  a  v  a  l  p  h  t  i  y  p  a  g  s  k  r  r  q  d  r  h  l  r  e  a  v  e  y  g  r

36701  gcagcgcggc ctcgtcggcg ccggggtagg cggcgcggtg cgcggccagc agcggcgcga cggcggggcg gaccgggtcg ggggggatc cgtcgtgact
       cgtcgcgccg gagcagccgc ggccccatcc gccgcgccac gcgccggtcg tcgccgcgct gccgccccgc ctggcccagc ccccccctag gcagcactga
       <.....................................................31.....................................................<
       l  l  a  a  e  d  a  g  p  y  a  a  r  h  a  a  l  l  p  a  v  a  p  r  v  p  d  p  p  s  g  d  h  s

36801  gggaaaaccc tggcggccgc gaattcttga agacgaaagg gcctcgtgat acgcctattt ttataggtta atgtcatgat aataatggtt tcttagacgt
       ccctttgggg accgccggcg cttaagaact tctgctttcc cggagcacta tgcggataaa aatatccaat tacagtacta ttattaccaa agaatctgca
       <..........31.........<<
         p  f  g  q  r  g  r  i

36901  caggtggcac ttttcgggga aatgtgcgcg gaacccctat ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagacaat aaccctgata
       gtccaccgtg aaaagcccct ttacacgcgc cttggggata aacaaataaa aagatttatg taagtttata cataggcgag tactctgtta ttgggactat

37001  aatgcttcaa taatattgaa aaaggaagag tatgagtatt caacatttcc gtgtcgccct tattcccttt tttgcggcat tttgccttcc tgtttttgct
       ttacgaagtt attataactt tttccttctc atactcataa gttgtaaagg cacagcggga ataagggaaa aaacgccgta aaacggaagg acaaaaacga

37101  cacccagaaa cgctggtgaa agtaaaagat gctgaagatc agttgggtgc acgagtgggt tacatcgaac tggatctcaa cagcggtaag atccttgaga
       gtgggtcttt gcgaccactt tcatttttcta cgacttctag tcaacccacg tgctcaccca atgtagcttg acctagagtt gtcgccattc taggaactct

37201  gttttcgccc cgaagaacgt tttccaatga tgagcacttt taaagttctg ctatgtggcg cggtattatc ccgtgttgac gccgggcaag agcaactcgg
       caaaagcggg gcttcttgca aaaggttact actcgtgaaa atttcaagac gataccacgc gccataatag ggcacaactg cggcccgttc tcgttgagcc

37301  tcgccgcata cactattctc agaatgactt ggttgagtac tcaccagtca cagaaaagca tcttacggat ggcatgacag taagagaatt atgcagtgct
       agcggcgtat gtgataagag tcttactgaa ccaactcatg agtggtcagt gtcttttcgt agaatgccta ccgtactgtc attctcttaa tacgtcacga

37401  gccataacca tgagtgataa cactgcggcc aacttacttc tgacaacgat cggaggaccg aaggagctaa ccgcttttt gcacaacatg ggggatcatg
       cggtattggt actcactatt gtgacgccgg ttgaatgaag actgttgcta gcctcctggc ttcctcgatt ggcgaaaaaa cgtgttgtac cccctagtac

37501  taactcgcct tgatcgttgg gaaccggagc tgaatgaagc cataccaaac gacgagcgtg acaccacgat gcctgcagca atggcaacaa cgttgcgcaa
       attgagcgga actagcaacc cttggcctcg acttacttcg gtatggtttg ctgctcgcac tgtggtgcta cggacgtcgt taccgttgtt gcaacgcgtt

37601  actattaact ggcgaactac ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa gttgcaggac cacttctgcg ctcggccctt
       tgataattga ccgcttgatg aatgagatcg aagggccgtt gttaattatc tgacctacct ccgcctattt caacgtcctg gtgaagacgc gagccgggaa

37701  ccggctggct ggtttattgc tgataaatct ggagccggtg agcgtgggtc tcgcggtatc attgcagcac tggggccaga tggtaagccc tcccgtatcg
       ggccgaccga ccaaataacg actatttaga cctcggccac tcgcacccag agcgccatag taacgtcgtg accccggtct accattcggg agggcatagc

37801  tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg agataggtgc ctcactgatt aagcattggt aactgtcaga
       atcaatagat gtgctgcccc tcagtccgtt gatacctact tgctttatct gtctagcgac tctatccacg gagtgactaa ttcgtaacca ttgacagtct

37901  ccaagtttac tcatatatac tttagattga tttaaaactt catttttaat ttaaaaggat ctaggtgaag atccttttg ataatctcat gaccaaaatc
       ggttcaaatg agtatatatg aaatctaact aaattttgaa gtaaaaatta aattttccta gatccacttc taggaaaaac tattagagta ctggttttag

38001  ccttaacgtg agttttcgtt ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc
       ggaattgcac tcaaaagcaa ggtgactcgc agtctggggc atcttttcta gtttcctaga agaactctag gaaaaaaaga cgcgcattag acgacgaacg

38101  aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc
       tttgtttttt tggtggcgat ggtcgccacc aaacaaacgg cctagttctc gatggttgag aaaaaggctt ccattgaccg aagtcgtctc gcgtctatgg

38201  aaatactgtc cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc gcctacatac ctcgctctgc taatcctgtt accagtggct
       tttatgacag gaagatcaca tcggcatcaa tccggtggtg aagttcttga gacatcgtgg cggatgtatg gagcgagacg attaggacaa tggtcaccga

38301  gctgccagtg gcgataagtc gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg aacggggggt tcgtgcacac
       cgacggtcac cgctattcag cacagaatgg cccaacctga gttctgctat caatggccta ttccgcgtcg ccagcccgac ttgcccccca agcacgtgtg

38401  agcccagctt ggagcgaacg acctacaccg aactgagata cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta
       tcgggtcgaa cctcgcttgc tggatgtggc ttgactctat ggatgtcgca ctcgatactc tttcgcggtg cgaagggctt ccctctttcc gcctgtccat

38501  tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg ccacctctga
       aggccattcg ccgtcccagc cttgtcctct cgcgtgctcc ctcgaaggtc ccccttt gcg gaccatagaa atatcaggac agcccaaagc ggtggagact

38601  cttgagcgtc gatttttgtg atgctcgtca ggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt
       gaactcgcag ctaaaaacac tacgagcagt ccccccgcct cggatacctt tttgcggtcg ttgcgccgga aaaatgccaa ggaccggaaa acgaccggaa

38701  ttgctggcct tttgctcaca tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc tcgccgcagc
       aacgaccgga aaacgagtgt acaagaaagg acgcaatagg ggactaagac acctattggc ataatggcgg aaactcactc gactatggcg agcggcgtcg

38801  cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aagagcgctg acttccgcgt ttccagactt tacgaaacac ggaaaccgaa gaccattcat
       gcttgctggc tcgcgtcgct cagtcactcg ctccttcgcc ttctcgcgac tgaaggcgca aaggtctgaa atgctttgtg cctttggctt ctggtaagta

38901  gttgttgctc aggtcgcaga cgttttgcag cagcagtcgc ttcacgttcg ctcgcgtatc ggtgattcat tctgctaacc agtaaggcaa ccccgccagc
       caacaacgag tccagcgtct gcaaaacgtc gtcgtcagcg aagtgcaagc gagcgcatag ccactaagta agacgattgg tcattccgtt ggggcggtcg

39001  ctagccgggt cctcaacgac aggagcacga tcatgcgcac ccgtcagatc cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg
       gatcggccca ggagttgctg tcctcgtgct agtacgcgtg ggcagtctag gtctgtacta ttctatgtaa ctactcaaac ctgtttggtg ttgatcttac

39101  cagtgaaaaa aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca ataaacaagt taacaacaac aattgcattc
       gtcacttttt ttacgaaata aacactttaa acactacgat aacgaaataa acattggtaa tattcgacgt tatttgttca attgttgttg ttaacgtaag

39201  attttatgtt tcaggttcag ggggaggtgt gggaggtttt ttaaagcaag taaaacctct acaaatgtgg tatggctgat tatgatctct agtcaaggca
       taaaatacaa agtccaagtc cccctccaca ccctccaaaa aatttcgttc attttggaga tgtttacacc ataccgacta atactagaga tcagttccgt

39301  ctatacatca aatattcctt attaacccct ttacaaatta aaaagctaaa ggtacacaat ttttgagcat agttattaat agcagacact ctatgcctgt
       gatatgtagt ttataaggaa taattgggga aatgtttaat ttttcgattt ccatgtgtta aaaactcgta tcaataatta tcgtctgtga gatacggaca
```

```
39401  gtggagtaag aaaaaacagt atgttatgat tataactgtt atgcctactt ataaaggtta cagaatattt ttccataatt ttcttgtata gcagtgcagc
       cacctcattc ttttttgtca tacaatacta atattgacaa tacggatgaa tatttccaat gtcttataaa aaggtattaa aagaacatat cgtcacgtcg

39501  tttttccttt gtggtgtaaa tagcaaagca agcaagagtt ctattactaa acacagcatg actcaaaaaa cttagcaatt ctgaaggaaa gtccttgggg
       aaaaaggaaa caccacattt atcgtttcgt tcgttctcaa gataatgatt tgtgtcgtac tgagtttttt gaatcgttaa gacttccttt caggaacccc

39601  tcttctacct ttctcttctt ttttggagga gtagaatgtt gagagtcagc agtagcctca tcatcactag atggcatttc ttctgagcaa aacaggtttt
       agaagatgga aagagaagaa aaaacctcct catcttacaa ctctcagtcg tcatcggagt agtagtgatc taccgtaaag aagactcgtt ttgtccaaaa

39701  cctcattaaa ggcattccac cactgctccc attcatcagt tccataggtt ggaatctaaa atacacaaac aattagaatc agtagtttaa cacattatac
       ggagtaattt ccgtaaggtg gtgacgaggg taagtagtca aggtatccaa ccttagattt tatgtgtttg ttaatcttag tcatcaaatt gtgtaatatg

39801  acttaaaaat tttatattta ccttagagct ttaaatctct gtaggtagtt tgtccaatta tgtcacacca cagaagtaag gttccttcac aaagatccgg
       tgaattttta aaatataaat ggaatctcga aatttagaga catccatcaa acaggttaat acagtgtggt gtcttcattc caaggaagtg tttctaggcc

39901  accaaagcgg ccatcgtgcc tccccactcc tgcagttcgg gggcatggat gcgcggatag ccgctgctgg tttcctggat gccgacggat ttgcactgcc
       tggtttcgcc ggtagcacgg aggggtgagg acgtcaagcc cccgtaccta cgcgcctatc ggcgacgacc aaaggaccta cggctgccta aacgtgacgg

40001  ggtagaactc cgcgaggtcg tccagcctca ggcagcagct gaaccaactc gcgaggggat cgagcccggt gggcgaagaa ctccagcatg agatccccgc
       ccatcttgag gcgctccagc aggtcggagt ccgtcgtcga cttggttgag cgctccccta gctcgggcca cccgcttctt gaggtcgtac tctaggggcg

40101  gctggaggat catccagccg gcgtcccgga aaacgattcc gaagcccaac ctttcataga aggcggcggt ggaatcgaaa tctcgtgatg gcaggttggg
       cgacctccta gtaggtcggc cgcagggcct tttgctaagg cttcgggttg gaaagtatct tccgccgcca ccttagcttt agagcactac cgtccaaccc

40201  cgtcgcttgg tcggtcattt cgaaccccag agtcccgctc agaagaactc gtcaagaagg cgatagaagg cgatgcgctg cgaatcggga gcggcgatac
       gcagcgaacc agccagtaaa gcttggggtc tcagggcgag tcttcttgag cagttcttcc gctatcttcc gctacgcgac gcttagccct cgccgctatg

40301  cgtaaagcac gaggaagcgg tcagcccatt cgccgccaag ctcttcagca atatcacggg tagccaacgc tatgtcctga tagcggtccg ccacacccag
       gcatttcgtg ctccttcgcc agtcgggtaa gcggcggttc gagaagtcgt tatagtgccc atcggttgcg atacaggact atcgccaggc ggtgtgggtc

40401  ccggccacag tcgatgaatc cagaaaagcg gccattttcc accatgatat tcggcaagca ggcatcgcca tgggtcacga cgagatcctc gccgtcgggc
       ggccggtgtc agctacttag gtcttttcgc cggtaaaagg tggtactata agccgttcgt ccgtagcggt acccagtgct gctctaggag cggcagcccg

40501  atgcgcgcct tgagcctggc gaacagttcg gctggcgcga gcccctgatg ctcttcgtcc agatcatcct gatcgacaag accggcttcc atccgagtac
       tacgcgcgga actcggaccg cttgtcaagc cgaccgcgct cggggactac gagaagcagg tctagtagga ctagctgttc tggccgaagg taggctcatg

40601  gtgctcgctc gatgcgatgt ttcgcttggt ggtcgaatgg gcaggtagcc ggatcaagcg tatgcagccg ccgcattgca tcagccatga tggatacttt
       cacgagcgag ctacgctaca aagcgaacca ccagcttacc cgtccatcgg cctagttcgc atacgtcggc ggcgtaacgt agtcggtact acctatgaaa

40701  ctcggcagga gcaaggtgag atgacaggag atcctgcccc ggcacttcgc ccaatagcag ccagtccctt cccgcttcag tgacaacgtc gagcacagct
       gagccgtcct cgttccactc tactgtcctc taggacgggg ccgtgaagcg ggttatcgtc ggtcagggaa gggcgaagtc actgttgcag ctcgtgtcga

40801  gcgcaaggaa cgcccgtcgt ggccagccac gatagccgcg ctgcctcgtc ctgcagttca ttcagggcac cggacaggtc ggtcttgaca aaaagaaccg
       cgcgttcctt gcgggcagca ccggtcggtg ctatcggcgc gacggagcag gacgtcaagt aagtcccgtg gcctgtccag ccagaactgt ttttcttggc

40901  ggcgcccctg cgctgacagc cggaacacgg cggcatcaga gcagccgatt gtctgttgtg cccagtcata gccgaatagc ctctccaccc aagcggccgg
       ccgcggggac gcgactgtcg gccttgtgcc gccgtagtct cgtcggctaa cagacaacac gggtcagtat cggcttatcg gagaggtggg ttcgccggcc

41001  agaacctgcg tgcaatccat cttgttcaat catgcgaaac gatcctcatc ctgtctcttg atcagatctt gatcccctgc gccatcagat ccttggcggc
       tcttggacgc acgttaggta gaacaagtta gtacgctttg ctaggagtag gacagagaac tagtctagaa ctaggggacg cggtagtcta ggaaccgccg

41101  aagaaagcca tccagtttac tttgcagggc ttcccaacct taccagaggg cgccccagct ggcaattccg gttcgcttgc tgtccataaa accgcccagt
       ttctttcggt aggtcaaatg aaacgtcccg aagggttgga atggtctccc gcggggtcga ccgttaaggc caagcgaacg acaggtattt tggcgggtca

41201  ctagctatcg ccatgtaagc ccactgcaag ctacctgctt tctctttgcg cttgcgtttt cccttgtcca gatagcccag tagctgacat tcatccgggg
       gatcgatagc ggtacattcg ggtgacgttc gatggacgaa agagaaacgc gaacgcaaaa gggaacaggt ctatcgggtc atcgactgta agtaggcccc

41301  tcagcaccgt ttctgcggac tggctttcta cgtgttccgc ttcctttagc agcccttgcg ccctgagtgc ttgcggcagc gtgaagcttt ttgcaaaagc
       agtcgtggca aagacgcctg accgaaagat gcacaaggcg aaggaaatcg tcgggaacgc gggactcacg aacgccgtcg cacttcgaaa aacgttttcg

41401  ctaggcctcc aaaaaagcct cctcactact tctggaatag ctcagaggcc gaggcggcct cggcctctgc ataaataaaa aaaattagtc agccatgggg
       gatccggagg ttttttcgga ggagtgatga agaccttatc gagtctccgg ctccgccgga gccggagacg tatttatttt ttttaatcag tcggtacccc

41501  cggagaatgg gcggaactgg gcggagttag gggcgggatg ggcggagtta ggggcgggac tatggttgct gactaattga gatgcatgct ttgcatactt
       gcctcttacc cgccttgacc cgcctcaatc cccgccctac ccgcctcaat ccccgccctg ataccaacga ctgattaact ctacgtacga aacgtatgaa

41601  ctgcctgctg gggagcctgg ggactttcca cacctggttg ctgactaatt gagatgcatg ctttgcatac ttctgcctgc tggggagcct ggggactttc
       gacggacgac ccctcggacc cctgaaaggt gtggaccaac gactgattaa ctctacgtac gaaacgtatg aagacggacg acccctcgga cccctgaaag

41701  cacaccctaa ctgacacaca ttccacagcc ggatctgcag gacccaacgc tgcccgagat gcgccgcgtg cggctgctgg agatggcgga cgcgatggat
       gtgtgggatt gactgtgtgt aaggtgtcgg cctagacgtc ctgggttgcg acgggctcta cgcggcgcac gccgacgacc tctaccgcct gcgctaccta

41801  atgttctgcc aagggttggt ttgcgcattc acagttctcc gcaagaattg attggctcca attcttggag tggtgaatcc gttagcgagg tgccgccggc
       tacaagacgg ttcccaacca aacgcgtaag tgtcaagagg cgttcttaac taaccgaggt taagaacctc accacttagg caatcgctcc acggcggccg

41901  ttccattcag gtcgaggtgg cccggctcca tgcaccgcga cgcaacgcgg ggaggcagac aaggtatagg gcggcgccta caatccatgc caacccgttc
       aaggtaagtc cagctccacc gggccgaggt acgtggcgct gcgttgcgcc cctccgtctg ttccatatcc cgccgcggat gttaggtacg gttgggcaag

42001  catgtgctcg ccgaggcggc ataaatcgcc gtgacgatca gcggtccaat gatcgaagtt aggctggtaa gagccgcgag cgatccttga agctgtccct
       gtacacgagc ggctccgccg tatttagcgg cactgctagt cgccaggtta ctagcttcaa tccgaccatt ctcggcgctc gctaggaact tcgacaggga

42101  gatggtcgtc atctacctgc ctggacagca tggcctgcaa cgcgggcatc ccgatgccgc cggaagcgag aagaatcata atggggaagg ccatccagcc
       ctaccagcag tagatggacg gacctgtcgt accggacgtt gcgcccgtag ggctacggcg gccttcgctc ttcttagtat taccccttcc ggtaggtcgg

42201  tcgcgtcgcg aacgccagca agacgtagcc cagcgcgtcg gccgccatgc cggcgataat ggcctgcttc tcgccgaaac gtttggtggc gggaccagtg
       agcgcagcgc ttgcggtcgt tctgcatcgg gtcgcgcagc cggcggtacg gccgctatta ccggacgaag agcggctttg caaaccaccg ccctggtcac

42301  acgaaggctt gagcgagggc gtgcaagatt ccgaataccg caagcgacag gccgatcatc gtcgcgctcc agcgaaagcg gtcctcgccg aaaatgaccc
       tgcttccgaa ctcgctcccg cacgttctaa ggcttatggc gttcgctgtc cggctagtag cagcgcgagg tcgctttcgc caggagcggc ttttactggg

42401  agagcgctgc cggcacctgt cctacgagtt gcatgataaa gaagacagtc ataagtgcgg cgacgatagt catgccccgc gcccaccgga aggagctgac
       tctcgcgacg gccgtggaca ggatgctcaa cgtactattt cttctgtcag tattcacgcc gctgctatca gtacggggcg cgggtggcct tcctcgactg

42501  tgggttgaag gctctcaagg gcatcggtcg acgctctccc ttatgcgact cctgcattag gaagcagccc agtagtaggt tgaggccgtt gagcaccgcc
       acccaacttc cgagagttcc cgtagccagc tgcgagaggg aatacgctga ggacgtaatc cttcgtcggg tcatcatcca actccggcaa ctcgtggcgg

42601  gccgcaagga atggtgcatg caaggagatg gcgcccaaca gtcccccggc cacggggcct gccaccatac ccacgccgaa acaagcgctc atgagcccga
       cggcgttcct taccacgtac gttcctctac cgcgggttgt caggggggcg gtgccccgga cggtggtatg ggtgcggctt tgttcgcgag tactcgggct

42701  agtggcgagc ccgatcttcc ccatcggtga tgtcggcgat ataggcgcca gcaaccgcac ctgtggcgcc ggtgatgccg gccacgatgc gtccggcgta
       tcaccgctcg ggctagaagg ggtagccact acagccgcta tatccgcggt cgttggcgtg gacaccgcgg ccactacggc cggtgctacg caggccgcat

42801  gaggatcttg gcagtcacag catgcgcata tccatgcttc gaccatgcgc tcacaaagta ggtgaatgcg caatgtagta cccacatcgt catcgctttc
       ctcctagaac cgtcagtgtc gtacgcgtat aggtacgaag ctggtacgcg agtgtttcat ccacttacgc gttacatcat gggtgtagca gtagcgaaag
```

```
42901  cactgctctc gcgaataaag atggaaaatc aatctcatgg taatagtcca tgaaaatcct tgtattcata aatcctccag gtagctatat gcaaattgaa
       gtgacgagag cgcttatttc tacctttag ttagagtacc attatcaggt acttttagga acataagtat ttaggaggtc catcgatata cgtttaactt

43001  acaaaagaga tggtgatctt tctaagagat gatggaatct cccttcagta tcccgatggt caatgcgctg gatatgggat agatgggaat atgctgattt
       tgttttctct accactagaa agattctcta ctaccttaga gggaagtcat agggctacca gttacgcgac ctatacccta tctaccctta tacgactaaa

43101  ttatgggaca gagttgcgaa ctgttcccaa ctaaaatcat tttgcacgat cagcgcacta cgaactttac ccacaaatag tcaggtaatg aatcctgata
       aataccctgt ctcaacgctt gacaagggtt gattttagta aaacgtgcta gtcgcgtgat gcttgaaatg ggtgtttatc agtccattac ttaggactat

43201  taaagacagg ttgataaatc agtcttctac gcgcatcgca cgcgcacacc gtagaaagtc tttcagttgt gagcctgggc aaaccgttaa ctttcggcgg
       atttctgtcc aactatttag tcagaagatg cgcgtagcgt gcgcgtgtgg catctttcag aaagtcaaca ctcggacccg tttggcaatt gaaagccgcc

43301  ctttgctgtg cgacaggctc acgtctaaaa ggaaataaat catgggtcat aaaattatca cgttgtccgg cgcggcgacg gatgttctgt atgcgctgtt
       gaaacgacac gctgtccgag tgcagatttt cctttattta gtacccagta ttttaatagt gcaacaggcc gcgccgctgc ctacaagaca tacgcgacaa

43401  tttccgtggc gcgttgctgt ctggtgatct gccttctaaa tctggcacag ccgaattgcg cgagcttggt tttgctgaaa ccagacacac agcaactgaa
       aaaggcaccg cgcaacgaca gaccactaga cggaagattt agaccgtgtc ggcttaacgc gctcgaacca aaacgacttt ggtctgtgtg tcgttgactt

43501  taccagaaag aaaatcactt tacctttctg acatcagaag ggcagaaatt tgccgttgaa cacctggtca atacgcgttt tggtgagcag caatattgcg
       atggtctttc ttttagtgaa atggaaagac tgtagtcttc ccgtctttaa acggcaactt gtggaccagt tatgcgcaaa accactcgtc gttataacgc

43601  cttcgatgac gcttggcgtt gagattgata cctctgctgc acaaaaggca atcgacgagc tggaccagcg cattcgtgac accgtctcct tcgaacttat
       gaagctactg cgaaccgcaa ctctaactat ggagacgacg tgttttccgt tagctgctcg acctggtcgc gtaagcactg tggcagagga agcttgaata

43701  tcgcaatgga gtgtcattca tcaaggacgc cgctatcgca aatggtgcta tccacgcagc ggcaatcgaa acacctcagc cggtgaccaa tatctacaac
       agcgttacct cacagtaagt agttcctgcg gcgatagcgt ttaccacgat aggtgcgtcg ccgttagctt tgtggagtcg gccactggtt atagatgttg

43801  atcagccttg gtatccagcg tgatgagcca gcgcagaaca aggtaaccgt cagtgccgat aagttcaaag ttaaacctgg tgttgatacc aacattgaaa
       tagtcggaac cataggtcgc actactcggt cgcgtcttgt tccattggca gtcacggcta ttcaagtttc aatttggacc acaactatgg ttgtaacttt

43901  cgttgatcga aaacgcgctg aaaaacgctg ctgaatgtgc ggcgctggat gtcacaaagc aaatggcagc agacaagaaa gcgatggatg aactggcttc
       gcaactagct tttgcgcgac tttttgcgac gacttacacg ccgcgaccta cagtgtttcg tttaccgtcg tctgttcttt cgctacctac ttgaccgaag

44001  ctatgtccgc acggccatca tgatggaatg tttccccggt ggtgttatct ggcagcagtg ccgtcgatag tatgcaattg ataattatta tcatttgcgg
       gatacaggcg tgccggtagt actaccttac aaaggggcca ccacaataga ccgtcgtcac ggcagctatc atacgttaac tattaataat agtaaacgcc

44101  gtcctttccg gcgatccgcc ttgttacggg gcggcgacct cgcgggtttt cgctatttat gaaaattttc cggtttaagg cgtttccgtt cttcttcgtc
       caggaaaggc cgctaggcgg aacaatgccc cgccgctgga gcgcccaaaa gcgataaata cttttaaaag gccaaattcc gcaaaggcaa gaagaagcag

44201  ataacttaat gtttttattt aaaataccct ctgaaaagaa aggaaacgac aggtgctgaa agcgagcttt ttggcctctg tcgtttcctt tctctgtttt
       tattgaatta caaaaataaa ttttatggga gacttttctt tcctttgctg tccacgactt tcgctcgaaa aaccggagac agcaaaggaa agagacaaaa

44301  tgtccgtgga atgaacaatg gaagtcaaca aaaagcagct ggctgacatt ttcggtgcga gtatccgtac cattcagaac tggcaggaac agggaatgcc
       acaggcacct tacttgttac cttcagttgt ttttcgtcga ccgactgtaa aagccacgct cataggcatg gtaagtcttg accgtccttg tcccttacgg

44401  cgttctgcga ggcggtggca agggtaatga ggtgctttat gactctgccg ccgtcataaa atggtatgcc gaaagggatg ctgaaattga gaacgaaaag
       gcaagacgct ccgccaccgt tcccattact ccacgaaata ctgagacggc ggcagtattt taccatacgg ctttccctac gactttaact cttgcttttc

44501  ctgcgccggg aggttgaaga actgcggcag gccagcgagg cagatccaca ggacgggtgt ggtcgccatg atcgcgtagt cgatagtggc tccaagtagc
       gacgcggccc tccaacttct tgacgccgtc cggtcgctcc gtctaggtgt cctgcccaca ccagcggtac tagcgcatca gctatcaccg aggttcatcg

44601  gaagcgagca ggactgggcg gcggccaaag cggtcggaca gtgctccgag aacgggtgcg catagaaatt gcatcaacgc atatagcgct agcagcacgc
       cttcgctcgt cctgacccgc cgccggtttc gccagcctgt cacgaggctc ttgcccacgc gtatctttaa cgtagttgcg tatatcgcga tcgtcgtgcg

44701  catagtgact ggcgatgctg tcggaatgga cgatatcccg caagaggccc ggcagtaccg gcataaccaa gcctatgcct acagcatcca gggtgacggt
       gtatcactga ccgctacgac agccttacct gctatagggc gttctccggg ccgtcatggc cgtattggtt cggatacgga tgtcgtaggt cccactgcca

44801  gccgaggatg acgatgagcg cattgttaga tttcatacac ggtgcctgac tgcgttagca atttaactgt gataaactac cgcattaaag cttatcgatg
       cggctcctac tgctactcgc gtaacaatct aaagtatgtg ccacggactg acgcaatcgt taaattgaca ctatttgatg gcgtaatttc gaatagctac

44901  ataagcggtc aaacatga
       tattcgccag tttgtact
```

8 Literaturverzeichnis

[1] Jack, R.W., Jung, G. (2000). Lantibiotics and microcins: polypeptides with unusual chemical diversity. Curr. Opin. Chem. Biol. *4*, 310-317.

[2] Joerger, M.C., Klaenhammer, T.R. (1986). Characterization and purification of helveticin J and evidence for a chromosomally determined bacteriocin produced by *Lactobacillus helveticus* 481. J. Bacteriol. *167*, 439-446.

[3] Drider, D., Fimland, G., Héchard, Y., McMullen, L.M., Prévost, H. (2006). The continuing story of class IIa bacteriocins. Microbiol. Mol. Biol. Rev. *70*, 564-582.

[4] Schnell, N., Entian, K.-D., Schneider, U., Götz, F., Zähner, H., Kellner, R., Jung, G. (1988). Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings. Nature, *333*, 276-278.

[5] Horn, M.J., Jones, D.B., Ringel, S.J. (1941). Isolation of a new sulfur-containing amino acid (lanthionine) from sodium carbonate-treated wool. J. Biol. Chem. *138*, 141-149.

[6] Willey, J.M., van der Donk, W.A. (2007). Lantibiotics: peptides of diverse structure and function. Annu. Rev. Microbiol. *61*, 477-501.

[7] Delves-Broughton, J., Blackburn, P., Evans, R.J., Hugenholtz, J. (1996). Applications of the bacteriocin, nisin. Antonie van Leeuwenhoek, *69*, 193-202.

[8] Kellner, R., Jung, G., Josten, M., Kaletta, C., Entian, K.-D., Sahl, H.-G. (1989). Pep5: Strukturaufklärung eines Lantibioticums. Angew. Chem. *101*, 618-621.

[9] Ekkelenkamp, M.B., Hanssen, M., Danny Hsu, S.-T., de Jong, A., Milatovic, D., Verhoef, J., van Nuland, N.A.J. (2005). Isolation and structural characterization of epilancin 15X, a novel lantibiotic from a clinical strain of *Staphylococcus epidermidis*. FEBS Letters, *579*, 1917-1922.

[10] Skaugen, M., Nissen-Meyer, J., Jung, G., Stevanovic, S., Sletten, K., Abildgaard, C.I.M., Nes, I.F., (1994). *In vivo* conversion of L-serine to D-alanine in a ribosomally synthesized polypeptide. J. Biol. Chem. *269*, 27183-27185.

[11] Cotter, P.D., O`Conner, P.M., Draper, L.A., Lawton, E.M., Deegan, L.H., Hill, C., Ross, R.P. (2005). Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. Proc. Natl. Acad. Sci. USA, *102*, 18584-18589.

[12] Castiglione F., Lazzarini, A., Carrano, L., Corti, E., Ciciliato, I., Gastaldo, L., Candiani, P., Losi, D., Marinelli, F., Selva, E., Parenti, F. (2008). Determining the structure and mode of action of microbisporicin, a potent lantibiotic active against multiresistant pathogens. Chem. Biol. *15*, 22-31.

[13] Hurst, A. (1981). Nisin. Adv. Appl. Microbiol. *27*, 85-123.

[14] Sahl, H.-G., Jack, R.W., Bierbaum, G. (1995). Biosynthesis and biological activities of lantibiotics with unique post-translational modifications. Eur. J. Biochem. *230*, 827-853.

[15] Bierbaum, G., Szekat, C., Josten, M., Heidrich, C., Kempter, C., Jung, G., Sahl, H.-G. (1996). Engineering of a novel thioether bridge and role of modified residues in the lantibiotic Pep5. Appl. Environ. Microbiol. *62*, 385-392.

[16] Cotter, P.D, Hill, C., Ross, R.P. (2005). Bacteriocins: developing innate immunity for food. Nat. Rev. Microbiol. *3*, 777-788.

[17] Hillmann, J.D. (2002). Genetically modified *Streptococcus mutans* for the prevention of dental caries. Antonie van Leeuwenhoek, *82*, 361-366.

[18] Breukink, E., de Kruijff, B. (2006). Lipid II as a target for antibiotics. Nat. Rev. Drug Discov. *5*,321-332.

[19] Hsu, S.-T.D., Breukink, E., Tischenko, E., Lutters, M.A.G., de Kruijff, B., Kaptein, R., Bonvin, A.M.J.J., van Nuland, N.A.J. (2004). The nisin-lipid II complex reveals a pyrophosphat cage that provides a blueprint for novel antibiotics. Nat. Struct. Mol. Biol. *11*, 963-567.

[20] Brötz, H., Sahl, H.-G. (2000). New insights into the mechanism of action of lantibiotics – diverse biological effects by binding to the same molecular target. J. Antimicrob. Chemother. *46*, 1-6.

[21] Machaidze, G., Seelig, J. (2003). Specific binding of cinnamycin (Ro 09-0198) to phosphatidylethanolamine. Comparison between micellar and membrane environments. Biochemistry, *42*, 12570-12576.

[22] Kodani, S., Lodato, M.A., Durrant, M.C., Picart, F., Willey, J.M. (2005). SapT, a lanthionine-containing peptide involved in aerial hyphae formation in the streptomycetes. Mol. Microbiol. *58*, 1368-1380.

[23] Chatterjee, C., Paul, M., Xie, L., van der Donk, W.A. (2005). Biosynthesis and mode of action of lantibiotics. Chem. Rev. *105*, 633-683.

[24] van der Meer, J.R., Polman, J., Beerthuyzen, M.M., Siezen, R.J., Kuipers, O.P., de Vos, W.M. (1993). Characterization of the *Lactococcus lactis* nisin A operon genes nisP, encoding a subtilisin-like serine protease involved in precursor processing, and nisR, encoding a regulatory protein involved in nisin biosynthesis. J. Bacteriol. *175*, 2578-2588.

[25] Meyer, C., Bierbaum, G., Heidrich, C., Reis, M., Süling, J., Iglesias-Wind, M.I., Kempter, C., Molitor, E., Sahl, H.-G. (1995). Nucleotide sequence of the lantibiotic Pep5 biosynthetic gene cluster and functional analysis of Pep5 and PepC. Eur. J. Biochem. *232*, 478-489.

[26] Nes, I.F., Tagg, J.R. (1996). Novel lantibiotics and their pre-peptides. Antonie van Leeuwenhoek, *69*, 89-97.

[27] Li, B., Yu, J.P.J., Brunzelle, J.S., Moll, G.N., van der Donk, W.A., Nair, S.K. (2006). Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. Science, *311*, 1464-1467.

[28] van der Meer, J.R., Rollema, H.S., Siezen, R.J., Beerthuyzen, M.M., Kuipers, O.P., de Vos, W.M. (1994). Influence of amino acid substitutions in the nisin leader peptide on biosynthesis and secretion of nisin by *Lactococcus lactis*. J. Biol. Chem. *269*, 3555-3562.

[29] Siezen, R.J., Kuipers, O.P., de Vos, W.M. (1996). Comparison of lantibiotic gene clusters and encoded proteins. Antonie van Leeuwenhoek, *69*, 171-184.

[30] Gutowski-Eckel, Z., Klein, C., Siegers, K., Bohm, K., Hammelmann, M., Entian, K.-D. (1994). Growth phase-dependent regulation and membrane localization of SpaB, a protein involved in biosynthesis of the lantibiotic subtilin. Appl. Environ. Microbiol. *60*, 1-11.

[31] Koponen, O., Tolonen, M., Qiao, M., Wahlström, G., Helin, J., Saris, P.E.J. (2002). NisB is required for the dehydration and NisC for the lanthionine formation in the posttranslational modification of nisin. Microbiology, *148*, 3561-3568.

[32] Okeley, N.M., Paul, M., Strasser, J.P., Blackburn, N., van der Donk, W.A. (2003). SpaC and NisC, the cyclases involved in subtilin and nisin biosynthesis, are zinc proteins. Biochemistry, *42*, 13613-13624.

[33] Chatterjee, C., Miller, L.M., Leung, Y.L., Xie, L., Yi, M., Kelleher, N.L., van der Donk, W.A. (2005). Lacticin 481 synthetase phosphorylates its substrate during lantibiotic production. J. Am. Chem. Soc. *127*, 15332-15333.

[34] Davidson, A.L., Chen, J. (2004). ATP-binding cassette transporters in bacteria. Annu. Rev. Biochem. *73*, 241-268.

[35] Krügel, H. (1997). How microorganisms produce and survive antibiotics. BIOspektrum, *Sonderausgabe 1997*, 38-41.

[36] Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. *1*, 945-951.

[37] Kerr, I.D. (2002). Structure and association of ATP-binding cassette transporter nucleotide-binding domains. Biochim. Biophys. Acta, 1-18.

[38] Martín, J.F., Liras, P. (1989). Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. Annu. Rev. Microbiol. *43*, 173-206.

[39] Widdick, D.A., Dodd, H.M., Barraille, P., White, J., Stein, T.H., Chater, K.F., Gasson, M.J., Bibb, M.J. (2003). Cloning and engineering of the cinnamycin biosynthetic gene cluster from *Streptomyces cinnamoneus cinnamoneus* DSM 40005. Proc. Natl. Acad. Sci. USA, *100*, 4316-4321.

[40] Heidrich, C., Pag, U., Josten, M., Metzger, J., Jack, R.W., Bierbaum, G., Jung, G., Sahl, H.-G. (1998). Isolation, characterization, and heterologous expression of the novel lantibiotic epicidin 280 and analysis of its biosynthetic gene cluster. Appl. Environ. Microbiol. *64*, 3140-3146.

[41] Kodani, S., Hudson, M.E., Durrant, M.C., Buttner, M.J., Nodwell, J.R., Willey, J.M. (2004). The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. Proc. Natl. Acad. Sci. USA, *101*, 11448-11453.

[42] Pag, U., Sahl, H.-G. (2002). Multiple activities in lantibiotics – models for the design of novel antibiotics? Curr. Pharm. Des. *8*, 815-833.

[43] Garneau, S., Martin, N.I., Vederas, J.C. (2002). Two-peptide bacteriocins produced by lactic acid bacteria. Biochimie, *84*, 577-592.

[44] Ueda, K., Oinuma, K.-I., Ikeda, G., Hosono, K., Ohnishi, Y., Horinouchi, S., Beppu, T. (2002). AmfS, an extracellular peptidic morphogen in *Streptomyces griseus.* J. Bacteriol. *184*, 1488-1492.

[45] Wink, J., Kroppenstedt, R.M., Seibert, G., Stackebrandt, E. (2003). *Actinomadura namibiensis* sp. nov. Int. J. Syst. Evol. Microbiol. *53*, 721-724.

[46] Lazzarini, A., Cavaletti, L., Toppo, G., Marinelli, F. (2000). Rare genera of actinomycetes as potential producers of new antibiotics. Antonie van Leeuwenhoek *78*, 399-405.

[47] Edman, P. (1949). A method for the determination of amino acid sequence in peptides. Arch Biochem*. 22*, 475.

[48] Paik, S.H., Chakicherla, A., Hansen, J.N. (1998). Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by *Bacillus subtilis* 168. J. Biol. Chem. *273*, 23134-23142.

[49] Hanahan, D. (1983). Studies on transformation of *Escherichia co*li with plasmids. J. Mol. Biol. *166*, 557-580.

[50] MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H., MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* required for avermectin biosynthesis utilizing a novel integration vector. Gene, *111*, 61-68.

[51] Dagert, M., Ehrlich, S.D. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene, *6*, 23-28.

[52] Sambrook, J., Russell, D.W. (2001). Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press.

[53] Dower, W.J., Miller, J.F., Ragsdale, W. (1988). High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res., *16*, 6127-6145.

[54] Neumann, E., Schäfer-Ridder, M., Wang, Y., Hofschneider, P.H. (1982). Gene transfer into mouse lyoma cells by electroporation in high electric fields. EMBO J. *1*, 841-845.

[55] Okanishi, M., Suzuki, K., Umezawa, H. (1974). Formation and reversion of Streptomycete protoplasts: cultural condition and morphological study. J. Gen. Microbiol. *80*, 389-400.

[56] Bibb, M.J., Ward, J.M., Hopwood, D.A. (1978). Transformation of plasmid DNA into Streptomyces at high frequency. Nature, *274*, 398-400.

[57] Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A. (2000). Practical streptomyces genetics, Crowes. Norwich, England.

[58] Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503-517.

[59] Birnboim, H.C., Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plamsid DNA. Nucleic Acids Res. *7*, 1513-1523.

[60] Vogelstein, B., Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA, *76*, 615-619.

[61] Hamaguchi, K., Geiduschek, P. (1962). Effect of electrolytes on the stability of the deoxyribonucleate helix. J. Am. Chem. Soc. *84*, 1329-1338.

[62] Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. *51*, 263-273.

[63] Ullmann, A., Perrin, D. (1970). Complementation in β-galactosidase, (The lactose operon) Beckwith, J.R., Zipser, D. (Hrsg.)) Cold Spring Harbor Laboratory Press, New York, 143-172.

[64] Altschul, A.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

[65] http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/glossary2.html

[66] Rutherford, K., Parkhill, J., Crook, J., Hornsell, T., Rice, P., Rajandream, M.-A., Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics, *16*, 944-945.

[67] Bibb, M.J., Findlay P.R., Johnson, M.W. (1984). The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene, *30*, 157-166.

[68] Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. *72*, 248-254.

[69] Lämmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*, 680-685.

[70] Schägger, H., Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. *166*, 368-379.

[71] Porat, J. (1992). Protein expression and purification. *3*, 263-281.

[72] Roeppstorff, P., Fohlmann, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. *11*, 601.

[73] Dairi, T., Hamano, Y., Furumai, T., Oki, T. (1999). Development of a self-cloning system for *Actinomadura verrucosospora* and identification of polyketide synthase genes essential for production of the angucyclic antibiotic pradimicin. Appl. Environ. Microbiol. *65*, 2703-2709.

[74] Dairi, T., Hamano, Y., Igarashi, Y., Furumai, T., Oki, T. (1997). Protoplasting and regeneration of strains belonging to the genus *Actinomadura*. Actinomycetologica, *11*, 1-5.

[75] Ayuso-Sacido, A., Genilloud, O. (2005). New PCR primers for the screening of NRPS and PKS-I systems in Actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. Microbiol Ecology, *49*, 10-24.

[76] *all non-redundant GenBank CDS translations* + RefSeq Proteins + PDB + SwissProt + PIR + PRF (von ww.ncbi.nlm.nih.gov/BLAST/blastcgihelp.shtml#protein_databases)

[77] Ross, K.F., Ronson, C.W., Tagg, J.R. (1993). Isolation and characterization of the lantibiotic Salivaricin A and its structural gene *salA* from *Streptococcus salivarius* 20P3. Appl. Environ. Microbiol. *59*, 2014-2021.

[78] Dabard, J., Bridonneau, C., Phillipe, C., Anglade, P., Molle, D., Nardi, M., Ladiré, M., Girardin, H., Marcille, F., Gomez, A., Fons, M. (2001). Ruminococcin A, a new lantibiotic produced by *Ruminococcus gnavus* strain isolated from human feces. Appl. Environ. Microbiol. *67*, 4111-4118.

[79] Holtsmark, I., Mantzilas, D., Eijsink, V.G.H., Brurberg M.B. (2006). Purification, characterization, and gene sequence of Michiganin A, an actagardine-like lantibiotic produced by the tomato pathogen *clavibacter michiganensis* subsp. *Michiganensis*. Appl. Environ. Microbiol. *72*, 5814-5821.

[80] Strohl, W.R. (1992). Compilation and analysis of DNA sequences associated with apparent streptomycete promoters. Nucleic Acids Res. *20*, 961-974.

[81] Hanks, S.K., Quinn, A.M., Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science, *241*, 42-52.

[82] Krupa, A., Srinivasan, N. (2005). Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. BMC Genomics, *6*, 129.

[83] Hudson, M.E., Nodwell, J.R. (2003). Dimerization of the RamC morphogenetic protein of *Streptomyces coelicolor*. Journal of Bacteriology, *186*, 1330-1336.

[84] Hudson, M.E., Zhang, D., Nodwell, J.R. (2002). Membrane association and kinase-like motifs of the RamC protein of *Streptomyces coelicolor*. J. Bacteriol. *184*, 4920-4924.

[85] Hanks, S.K., Hunter, T., (1995). The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J., *9*, 576-596.

[86] Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567-580.

[87] Chatterjee, C., Miller, L.M., Leung, Y.L., Xie, L., Yi, M., Kelleher, N.L., van der Donk W. (2005). Lacticin 481 synthetase phosphorylates its substrate during lantibiotic production. J. Am. Chem. Soc. *127*, 15332-15333.

[88] Yamazaki, H., Takano, Y., Ohnishi, Y., Horinouchi, S. (2003). *amf*R, an essential gene for aerial mycelium formation, is a member of the AdpA regulon in the A-factor regulatory cascade in *Streptomyces griseus*. Mole. Microbiol. *50*, 1173-1187.

[89] Oliynyk, M., Samborskyy, M., Lester, J.B.; Mironenko, T., Scott, N., Dickens, S., Haydock, S.F., Leadlay, P.F. (2007). Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. Nat. Biotechnol. *25*, 447-453.

[90] Xie, L., Miller, L.M., Chatterjee, C., Averin, O., Kelleher, N.L., van der Donk, W.A. (2004). Lacticin 481: in vitro reconstitution of lantibiotic synthetase activity. Science, *303*, 679-681.

[91] Kuhstoss, S., Rao, R.N. (1991). Analysis of the integration function of the streptomycete bacteriophage ΦC31. J. Mol. Biol. *222*, 897-908.

[92] Thorpe, H.M., Wilson, S.E., Smith, M.C.M. (2000). Control of directionality in the site-specific recombination system of the *Streptomyces* phage ΦC31. Mol. Microbiol. *38*, 232-241.

[93] Bierman, M., Logan, R., O`Brien, K., Seno. E.T., Nagaraja Rao, R., Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene, *116*, 43-49.

[94] Datsenko, K.A., Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA, *97*, 6640-6645.

[95] Gust, B., Challis, G.L., Fowler, K., Kieser, T., Chater, K.F. (2003). PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc. Natl. Acad. Sci. USA, *100*, 1541-1546.

[96] Murphy, K.C. (1998). Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. J. Bacteriol. *180*, 2063-2071.

[97] Eustáquio, A.S., Gust, B., Galm, U., Li, S.-M., Chater, K.F., Heide, L. (2005). Heterologous expression of Novobiocin and Clorobiocin biosynthetic gene clusters. Appl. Environ. Microbiol. *71*, 2452-2459.

[98] Gust, B., Kieser, T., Chater, K., *http://streptomyces.org.uk/redirect/protocol_V1.3.pdf*