

Identifizierung von diversitätsgenerierenden Retroelementen (DGRs) in Cyanobakterien und strukturelle Charakterisierung der Templat RNA eines DGRs aus dem Cyanobakterium *Nostoc sp. PCC7120*

Vom Fachbereich Biologie der TU Kaiserslautern zur Verleihung des akademischen Grades „Doktor der Naturwissenschaften“
genehmigte Dissertation

Vorgelegt von:

Philipp Möller

Tag der wissenschaftlichen Aussprache,

19. Juni 2015

Berichterstatter:

Dr. Nora Zingler

Prof. Dr. Matthias Hahn

D386

Inhaltsverzeichnis

INHALTSVERZEICHNIS	I
ABBILDUNGSVERZEICHNIS	IV
TABELLENVERZEICHNIS	V
ABKÜRZUNGSVERZEICHNIS	VI
ZUSAMMENFASSUNG	VIII
SUMMARY	IX
1 EINLEITUNG	1
1.1 MOBILE ELEMENTE	1
1.1.1 <i>Selbstpleißende Gruppe II Introns</i>	2
1.1.2 <i>Diversitätsgenerierende Retroelemente</i>	5
1.1.2.1 Der Mechanismus von diversitätsgenerierenden Retroelementen.....	8
1.1.2.2 Verbreitung von diversitätsgenerierenden Retroelementen	10
1.1.3 <i>Zielsetzung</i>	11
2 MATERIAL UND METHODEN	12
2.1 MATERIAL.....	12
2.1.1 <i>Chemikalien</i>	12
2.1.2 <i>Enzyme</i>	12
2.1.3 <i>Geräte</i>	12
2.1.4 <i>Kit-Systeme</i>	13
2.1.5 <i>Bakterienstämme</i>	13
2.1.6 <i>Medien</i>	14
2.1.7 <i>Lösungen und Puffer</i>	15
2.2 METHODEN.....	17
2.2.1 <i>Arbeiten mit Bakterien</i>	17
2.2.1.1 Kultivierung von <i>Escherichia coli</i>	17
2.2.1.2 Kultivierung von Cyanobakterien	17
2.2.2 <i>Arbeiten mit DNA</i>	18
2.2.2.1 Klonierungen	18
2.2.2.1.1 Präparative PCR	18
2.2.2.1.2 Agarose-Gelelektrophorese von DNA	19
2.2.2.1.3 Aufreinigung von DNA aus PCR-Produkten und Agarosegelen.....	19
2.2.2.1.4 Restriktionsverdau von DNA.....	19
2.2.2.1.5 Ligation von Insert und Vektor	20
2.2.2.1.6 Transformation in <i>Escherichia coli</i>	20
2.2.2.1.7 Kolonie-PCR	20
2.2.2.1.8 Plasmidisolierung.....	21
2.2.2.1.9 Sequenzierungen	21

Inhaltsverzeichnis | II

2.2.2.2	Cyanobakterielles Screening auf die DGR-kodierte RT	21
2.2.2.3	Inverse PCR.....	22
2.2.2.3.1	Phenol-Chloroform-Extraktion von Nukleinsäuren.....	22
2.2.2.3.2	Restriktionsverdau der Inversen PCR.....	23
2.2.2.3.3	Intramolekulare Ligation	23
2.2.2.3.4	Touchdown-PCR	23
2.2.2.4	Restriktionsenzym-basierter Variabilitäts-Assay	24
2.2.3	<i>Arbeiten mit RNA</i>	25
2.2.3.1	Northern-Blot-Experimente.....	25
2.2.3.1.1	RNA-Isolation	25
2.2.3.1.2	Denaturierende Formaldehyd-Agarose-Gelelektrophorese von RNA	25
2.2.3.1.3	Transfer von RNA auf eine Nylonmembran	26
2.2.3.1.4	Radioaktive Markierung von DNA-Sonden	26
2.2.3.1.5	Hybridisierung von geblotteter RNA.....	26
2.2.3.2	RT-PCR-Analysen	27
2.2.3.3	RACE-Experimente.....	28
2.2.3.4	SHAPE-Experimente	30
2.2.3.4.1	In vitro Transkription	30
2.2.3.4.2	Denaturierende Polyacrylamidgelelektrophorese.....	31
2.2.3.4.3	Modifizierung durch NMIA	31
2.2.3.4.4	Primer-Extension	32
2.2.4	<i>Arbeiten mit Computerprogrammen</i>	33
2.2.4.1	Quantifizierung der Bandenintensitäten in RT-PCR-Analysen	33
2.2.4.2	Verarbeitung der SHAPE-Daten.....	33
2.2.4.3	Bioinformatische Sekundärstrukturanalysen	34
2.2.4.4	Erstellung des phylogenetischen Baums der DGR-kodierten RTs.....	34
2.2.4.5	Identifizierung von Sequenzhomologien und Erstellung von Alignments	35
2.2.4.6	Erstellung der Vektorkarten	35
2.2.4.7	Identifizierung von invertierten Repeats	35
2.2.4.8	Taxonomie von Cyanobakterien.....	35
3	ERGEBNISSE	36
3.1	EIN VERFAHREN FÜR DIE IDENTIFIZIERUNG VON DGRS IN CYANOBAKTERIEN	36
3.1.1	<i>Etablierung eines PCR-Screenings auf die DGR-kodierte RT</i>	36
3.1.2	<i>Alignments der identifizierten RT-Sequenzen</i>	40
3.1.3	<i>DGR-kodierte RTs aus Cyanobakterien bilden einen Subtree im phylogenetischen Baum</i>	42
3.1.4	<i>Geographische Verteilung der DGR-assoziierten Cyanobakterienstämme</i>	44
3.1.5	<i>Entdeckung eines neuen DGRs in Anabaena flos-aquae</i>	45
3.1.5.1	<i>Anabaena flos-aquae</i> besitzt eine durch Kältestress induzierte RT.....	50
3.2	AKTIVITÄTSANALYSEN.....	53
3.2.1	<i>Das native DGR aus Nostoc sp. PCC7120 zeigt Aktivität</i>	53
3.2.2	<i>Das native DGR aus Anabaena flos-aquae zeigt keine Aktivität</i>	55

3.2.3	<i>Das native DGR aus Treponema denticola zeigt Variabilität</i>	56
3.2.4	<i>Heterolog exprimierte DGRs zeigen keine Aktivität</i>	58
3.3	STRUKTURANALYSE DER TEMPLAT RNA AUS <i>NOSTOC SP. PCC7120</i>	60
3.3.1	<i>Identifizierung des Haupttranskripts der Templat RNA</i>	60
3.3.2	<i>Bioinformatische Sekundärstrukturberechnung der Templat RNA</i>	68
3.3.3	<i>Experimentelle Daten unterstützen die Sekundärstrukturberechnung der Templat RNA</i>	72
3.3.4	<i>Berechnung einer Konsensus-Struktur für mehrere Templat RNAs</i>	78
4	DISKUSSION	83
4.1	DAS SCREENING DER CYANOBAKTERIENSAMMLUNG IDENTIFIZIERT 34 DGRS	83
4.1.1	<i>DGRs sind in der Stammsammlung weiter verbreitet als in öffentlichen Sequenzdatenbanken</i>	86
4.1.2	<i>Die Verbreitung von DGRs: Weltweit und bevorzugt über horizontalen Gentransfer?</i>	89
4.1.3	<i>Das neu entdeckte DGR aus Anabaena flos-aquae könnte durch den Wirt reguliert sein</i>	91
4.2	AKTIVITÄTSANALYSEN BAKTERIELLER DGRS	93
4.2.1	<i>Analysen cyanobakterieller DGRs zeigen Aktivität in Nostoc sp. PCC7120</i>	94
4.2.2	<i>Das DGR aus Treponema denticola ist über ein Jahr nicht aktiv</i>	101
4.2.3	<i>Heterolog exprimierte DGRs zeigen keine Aktivität</i>	102
4.2.4	<i>Regulation der DGR-Aktivität durch den Wirt</i>	103
4.3	STRUKTURAUFKLÄRUNG DER TEMPLAT RNA AUS <i>NOSTOC SP. PCC7120</i>	104
4.3.1	<i>Für den Templat Repeat existieren unterschiedlich lange Transkripte</i>	104
4.3.2	<i>Analyse der identifizierten Sekundärstruktur der Templat RNA</i>	106
4.3.3	<i>Die Templat RNA könnte konservierte Strukturelemente besitzen</i>	111
4.4	AUSBLICK	116
5	LITERATURVERZEICHNIS	117
	ANHANG	131
	TABELLE A1. LISTE DER CYANOBAKTERIENSAMMLUNG	131
	TABELLE A2: PATIENTENPROBEN AUS SUBGINGIVALEM PLAQUE	136
	TABELLE A3. SHAPE-DATEN VON DER TEMPLAT RNA	138
	TABELLE A4. VERWENDETE OLIGONUKLEOTIDE	148
	ABBILDUNG A1: VEKTORKARTEN	152
	ABBILDUNG A2: PHYLOGENETISCHER BAUM DER DGR-KODIERTEN RTS	154
	ABBILDUNG A3: <i>IN VITRO</i> TRANSKRIPTION VIER MÖGLICHER TEMPLAT RNAs AUS <i>NOSTOC SP. PCC7120</i>	157
	ABBILDUNG A4: SHAPE-EXPERIMENT ZUR STRUKTURAUFKLÄRUNG DER TEMPLAT RNA	158
	ABBILDUNG A5: KONTROLLE DER KONSENSUS-STRUKTUR VON DER TEMPLAT RNA	160
	DANKSAGUNG	162
	LEBENSLAUF	163
	EIDESSTÄTLICHE VERSICHERUNG	164

Abbildungsverzeichnis

ABBILDUNG 1: KONSERVIERTE SEKUNDÄRSTRUKTUR VON GRUPPE IIA INTRONS.....	2
ABBILDUNG 2: SPLEIßEN UND REVERSES SPLEIßEN VON GRUPPE II INTRONS.	3
ABBILDUNG 3: RETROHOMING VON GRUPPE II INTRONS.	4
ABBILDUNG 4: AUFBAU VON DGRS.....	5
ABBILDUNG 5: SCHEMATISCHE DARSTELLUNG DES DGR-MECHANISMUS.....	8
ABBILDUNG 6: AUSSCHNITT AUS DEM CYANOBAKTERIELLEN SCREENING AUF DIE DGR-KODIERTE RT.....	37
ABBILDUNG 7: SEQUENZ-ALIGNMENTS DER CYANOBAKTERIELLEN RT-FRAGMENTE.	41
ABBILDUNG 8: SUBTREE DER CYANOBAKTERIELLEN DGR-KODIERTEN RTs IM PHYLOGENETISCHEN BAUM.	43
ABBILDUNG 9: SCHEMATISCHE DARSTELLUNG DER INVERSEN PCR.	45
ABBILDUNG 10: INVERSE PCR VON <i>ANABAENA FLOS-AQUAE</i>	46
ABBILDUNG 11: SCHEMATISCHE DARSTELLUNG DES DGRs AUS <i>ANABAENA FLOS-AQUAE</i>	47
ABBILDUNG 12: ALIGNMENT DER DGR-KODIERTEN RT-DOMÄNEN.	47
ABBILDUNG 13: ALIGNMENT VON AVD-SEQUENZEN.	48
ABBILDUNG 14: VARIABLE REGION DES DGRs AUS <i>ANABAENA FLOS-AQUAE</i>	48
ABBILDUNG 15: DETEKTION DER CYANOBAKTERIELLEN DGR-TRANSKRIPTE.	50
ABBILDUNG 16: TRANSKRIPTREGULATION DER RT AUS <i>ANABAENA FLOS-AQUAE</i>	51
ABBILDUNG 17: AUSSCHNITT DES RESTRIKTIONSENZYM-BASIERTEN VARIABILITÄTS-ASSAYS.....	54
ABBILDUNG 18: VARIABLE REGION DES DGRs AUS <i>NOSTOC SP. PCC7120</i>	54
ABBILDUNG 19: NACHWEIS DER VARIABLEN REGION AUS <i>TREPONEMA DENTICOLA</i>	56
ABBILDUNG 20: VARIABLE REGION DES DGRs AUS <i>TREPONEMA DENTICOLA</i>	58
ABBILDUNG 21: STABILITÄTSTEST HETEROLOG EXPRIMIERTER DGRs.	59
ABBILDUNG 22: KONTROLLE DER TEX-AKTIVITÄT.....	61
ABBILDUNG 23: IDENTIFIZIERUNG DER TRANSKRIPTENDEN DER TEMPLAT RNA ÜBER NORTHERN-BLOT-EXPERIMENTE.	62
ABBILDUNG 24: NORTHERN-BLOT-EXPERIMENT ZUR GRÖßENBESTIMMUNG DER TEMPLAT RNA.	64
ABBILDUNG 25: IDENTIFIZIERUNG DER TRANSKRIPTENDEN DER TEMPLAT RNA ÜBER RT-PCR.	65
ABBILDUNG 26: DETEKTION DES TR4-TRANSKRIPTS IN TEX-BEHANDELTER GESAMT-RNA.	67
ABBILDUNG 27: VERGLEICH DER VIER POTENTIELLEN STRUKTUREN FÜR DIE TEMPLAT RNA.	69
ABBILDUNG 28. SEKUNDÄRSTRUKTURBERECHNUNG DES HAUPTTRANSKRIPTS DER TEMPLAT RNA.	71
ABBILDUNG 29: AUSSCHNITT EINES SHAPE-EXPERIMENTS DER TEMPLAT RNA.....	73
ABBILDUNG 30: SEKUNDÄRSTRUKTURBERECHNUNG DER TEMPLAT RNA MIT INTEGRIERTEN SHAPE-REAKTIVITÄTEN.....	75
ABBILDUNG 31: ALIGNMENT DER TEMPLAT RNAs FÜR DIE BERECHNUNG DER KONSENSUS-STRUKTUR.....	80
ABBILDUNG 32: KONSENSUS-STRUKTUR DER TEMPLAT RNAs.	81
ABBILDUNG 33: ÜBERSICHT DES CYANOBAKTERIELLEN SCREENINGS.	83
ABBILDUNG 34: MODELL DES MUTAGENEN HOMINGS.	96
ABBILDUNG 35: INTRAMOLEKULARER TEMPLAT-WECHSEL VON REVERSEN TRANSKRIPTASEN.	99
ABBILDUNG 36: POSITION DER DIREKTEN REPEATS IN DER TEMPLAT RNA.....	99

Tabellenverzeichnis

TABELLE 1: EIGENSCHAFTEN DER BESCHRIEBENEN DGRS	7
TABELLE 2. VERWENDETE GERÄTE.....	12
TABELLE 3. VERWENDETE KIT-SYSTEME.....	13
TABELLE 4. VERWENDETE BAKTERIENSTÄMME.....	13
TABELLE 5. VERWENDETE MEDIEN	14
TABELLE 6. SPURENELEMENTLÖSUNGEN DER MEDIEN (100 ML).....	15
TABELLE 7: VERWENDETE RESTRIKTIONSENZYME INNERHALB DES REBVA.....	24
TABELLE 8: CYANOBAKTERIEN MIT EINEM PCR-PRODUKT DER KORREKTEN LÄNGE FÜR DIE RT	39
TABELLE 9: GEOGRAPHISCHE VERTEILUNG DER DGR-ASSOZIIERTEN CYANOBAKTERIENSTÄMME	44
TABELLE 10: PROBEN VON FÜNF PATIENTEN MIT AMPLIFIZIERBARER VR	57
TABELLE 11: UNTERSCHIEDE ZWISCHEN DEN BERECHNETEN SEKUNDÄRSTRUKTUREN	77
TABELLE 12: VERWENDETE TR-SEQUENZEN FÜR DIE SEKUNDÄRSTRUKTURBERECHNUNG	79
TABELLE 13: KULTIVIERUNGSBEDINGUNGEN DREIER STAMMSAMMLUNGEN.....	87
TABELLE 14: SEQUENZIDENTITÄTEN DER POTENTIELLEN TEMPLAT RNAS.....	111
TABELLE 15: POSITION DES TR ZU DEN DGR-KOMPONENTEN IN DEN ANALYSIERTEN ORGANISMEN	113

Abkürzungsverzeichnis

<i>avd</i>	: accessory variability determinant
BLAST	: Basic Local Alignment Search Tool
Bp	: Basenpaar
cDNA	: komplementäre DNS
cpm	: Counts per minute
Da	: Dalton (atomare Masseneinheit)
DEPC	: Diethyl pyrocarbonate
DiGReF	: Diversity-generating retroelement finder
DGR	: Diversitätsgenerierendes Retroelement
DMSO	: Dimethylsulfoxid
DTT	: Dithiothreitol
FGE	: Formylglycin-generierendes Enzym
gDNA	: genomische DNA
HIV	: Humanes Immundefizienzvirus
HRDC	: Helikase und RNaseD C-terminale (Domäne)
IEP	: Intron encoded protein
IMH	: Initiator des mutagenen Homings
IPTG	: Isopropyl- β -D-thiogalactopyranosid
IVT	: <i>In vitro</i> Transkription
LTR	: Long terminal Repeat
<i>mtd</i>	: major tropism determinant
N	: Jedes Nukleotid
NMIA	: N-methylisotoic anhydride
ORF	: Offener Leserahmen
PAGE	: Polyacrylamidgelelektrophorese
PCR	: Polymerase Kettenreaktion
R	: Purin
REBVA	: Restriktionsenzym basierter Variabilitäts-Assay
RACE	: rapid amplification of cDNA-ends
ROS	: Reaktive Sauerstoffspezies
rRNA	: ribosomale RNA
RT	: Reverse Transkriptase
SHAPE	: Selective 2'-hydroxyl acylation analyzed by primer extension
snRNP	: small nuclear ribonucleic particles
ss	: Einzelsträngig

STOPP	: Transkriptionsstopp
TEX	: Terminator 5'-Phosphate-Dependent Exonuclease
TPRT	: Target primed reverse transcription
TR	: Templat Repeat
TSS	: Transkriptionsstart
VR	: Variable Region
X	: Unbekannte Aminosäure
Y	: Pyrimidin

Zusammenfassung

Die kürzlich entdeckte Klasse der diversitätsgenerierenden Retroelemente (DGRs) kann ihrem Wirt einen selektiven Vorteil über eine beschleunigte Proteinevolution verschaffen. Dazu bedient sich das DGR eines nicht proliferativen „copy and replace“ Mechanismus, der kodierende Sequenzinformation zielgerichtet von einem Templat Repeat über ein RNA-Intermediat zu einem spezifischen Gen transferiert. Die Sequenz wird während dem Prozess hypermutiert, was vermutlich durch eine Fehleranfälligkeit der DGR-kodierten reversen Transkriptase (RT) geschieht. Dabei kann die mutierte Sequenz eine höhere Diversität erreichen, als es für die Antikörper und T-Zell-Rezeptoren des Immunsystems von Vertebraten beobachtet wurde.

In dieser Arbeit wurde die Verteilung von DGRs in einer Stammsammlung von Cyanobakterien untersucht. Dafür wurde ein Screening mit degenerierten Primern auf die DGR-kodierte RT durchgeführt. Es konnten ca. 30 % (34) der analysierbaren Cyanobakterienstämme positiv auf Präsenz eines DGRs getestet werden. Dazu gehört ein DGR aus *Anabaena flos-aquae*, von dem auch die Sequenz ermittelt werden konnte. Dieses neu entdeckte DGR wurde zusammen mit zwei weiteren DGRs aus *Nostoc sp. PCC7120* und *Treponema denticola* auf Aktivität untersucht, wobei die letzten beiden Elemente eine DGR-vermittelte Variation gezeigt haben. Das demonstriert die Funktionsfähigkeit der Elemente, gibt aber zugleich einen Hinweis auf eine starke Regulation, da die beobachtete Frequenz der Diversifizierung sehr gering war. Eine Regulation wäre vorteilhaft für den Wirt, da vermutlich ein Großteil der Mutationen die Funktion der variablen Proteine beeinträchtigt.

Von dem funktionsfähigen DGR aus *Nostoc sp. PCC7120* wurde anschließend die Struktur des RNA-Intermediats bioinformatisch und experimentell aufgeklärt. Dabei handelt es sich um die erste aufgeklärte Struktur von RNA-Intermediaten aus DGRs. Basierend auf den Daten konnte eine Konsensus-Struktur für 13 Sequenzen aus Cyanobakterien, grünen Schwefelbakterien, Purpurbakterien, *Treponema denticola* und dem *Bordetella*-Phagen berechnet werden, in der vier Haarnadelstrukturen konserviert zu sein scheinen. Diese Strukturelemente könnten auf eine konservierte Funktion des RNA-Intermediats hinweisen und eine hochaffine Bindestelle für die DGR-kodierte RT bereitstellen bzw. für eine katalytische Aktivität als Endonuklease benötigt werden.

Damit liefert diese Arbeit einen wichtigen Beitrag für die experimentelle Identifizierung von DGRs, sowie deren Verteilung und Regulation in Bakterien. Desweiteren bietet die Arbeit einen Hinweis darauf, dass es sich bei dem RNA-Intermediat nicht nur um eine mobile Komponente handelt, sondern weitere Funktionen hinzukommen könnten.

Summary

The recently discovered class of diversity-generating retroelements (DGRs) is able to provide a selective advantage to their host via an accelerated protein evolution. Therefore the DGR uses a non proliferative „copy and replace“ mechanism that transfers coding sequence information from a template repeat over an RNA intermediate to a specific gene in a site-directed manner. The sequence gets hypermutated during the process which probably happens due to an error-proneness of the DGR-encoded reverse transcriptase (RT). The mutated sequence can achieve a higher diversity than the antibody and T cell receptors of the vertebrate immune system.

In the present work the distribution of DGRs in a cyanobacterial culture collection was investigated. Therefore a screening was performed that uses degenerate primers to amplify a region of the DGR-encoded RT. The screening showed that about 30 % (34) of the analyzable strains is associated with DGRs. From this strains the sequence of a theoretical functional DGR could be determined in the species *Anabaena flos-aquae*. This newly discovered DGR was analyzed for activity together with two DGRs from *Nostoc sp. PCC7120* and *Treponema denticola* at which the latter two showed DGR mediated variation. This shows the functional capability of these elements but indicates a strong regulation since the observed frequency of diversification was very low. A regulation would be beneficial for the host since a large part of the variable proteins is probably inactive after the hypermutation.

The structure of the RNA intermediate from the functional DGR in *Nostoc sp. PCC7120* was identified using a bioinformatical and experimental approach. This is the first identified structure of a RNA intermediate from DGRs. Based on this data a consensus structure could be calculated that uses 13 sequences from cyanobacteria, green sulfur bacteria, purple bacteria, *Treponema denticola* and the *Bordetella* phage. In this consensus structure there seem to be four conserved hairpin loops that indicate a conserved function for the RNA intermediate. A potential function could include a high affinity binding site for the DGR-encoded RT or a catalytic activity as an endonuclease.

Thereby the present work provides an important contribution to the experimental identification of DGRs and their distribution and regulation in bacteria. Furthermore the results indicate that the RNA intermediate is not solely a mobile component but might have other functions important for the DGR mechanism.

1 Einleitung

1.1 Mobile Elemente

Die ersten mobilen Elemente wurden von Barbara McClintock beschrieben (McClintock 1950). Bei mobilen Elementen handelt es sich um DNA-Sequenzen, die ihre Position innerhalb des Genoms einer Zelle wechseln können (Transposition). Mobile Elemente die mit Hilfe eines RNA-Intermediats transponieren, werden als Retrotransposons bezeichnet. Retrotransposons benötigen eine reverse Transkriptase, um das RNA-Intermediat in komplementäre DNA (cDNA) umzuschreiben. Die cDNA kann anschließend an einer anderen Stelle im Genom integriert werden. Dadurch kommt es zu einer Verdopplung des Elements (zusammengefasst in Han 2010).

Retrotransposons sind weit verbreitet in Pflanzen und Tieren und können hohe Kopienzahlen erreichen. So besteht das menschliche Genom zu über 40 % aus Retrotransposons (Lander et al. 2001). Dabei können Retrotranspositionen einen schädlichen Einfluss auf ihren Wirt haben. Finden die Retrotranspositionen in proteinkodierenden oder regulatorischen Regionen statt, kann sich dies negativ auf die Funktion der betroffenen Proteine auswirken und Krankheiten hervorrufen (zusammengefasst in Hancks & Kazazian 2012).

Dabei wird ein Großteil der Krankheiten durch non-LTR Retrotransposons hervorgerufen. Die non-LTR Retrotransposons haben ihre Bezeichnung daher, dass sie im Gegensatz zu den LTR Retrotransposons nicht über lange terminale Repeats verfügen. Bei der Retrotransposition integrieren sie über einen Prozess in das Chromosom, der als „target primed reverse transkription“ (TPRT) bezeichnet wird (Luan et al. 1993). Dabei wird die an der Stelle der Integration vorliegende DNA als Primer für die reverse Transkription verwendet.

Die in den mitochondrialen und chloroplastidären Genomen von Pilzen und Pflanzen, sowie den Genomen einiger Bakterien vorkommenden Gruppe II Introns, besitzen einen ähnlichen Mechanismus der reversen Transkription. Das ist einer der Gründe, weshalb Gruppe II Introns als potentielle Vorfahren der non-LTR Retrotransposons angesehen werden (Keller & Michel 1985; Ferat & Michel 1993; Zimmerly et al. 1995).

1.1.1 Selbstspießende Gruppe II Introns

Gruppe II Introns sind Retroelemente deren RNA-Intermediat katalytische Aktivität besitzt. Dadurch können sie sich aus der prä-mRNA ausschneiden und die flankierenden Exon-Sequenzen kovalent verbinden (Selbstspießen). Entscheidend für die katalytische Aktivität ist die Ausbildung einer konservierten RNA-Struktur. Anhand dieser Struktur lassen sich Gruppe II Introns in die Klassen A-C einteilen (Lambowitz & Zimmerly 2004). In Abbildung 1 ist die konservierte Sekundärstruktur der Gruppe IIA Introns gezeigt.

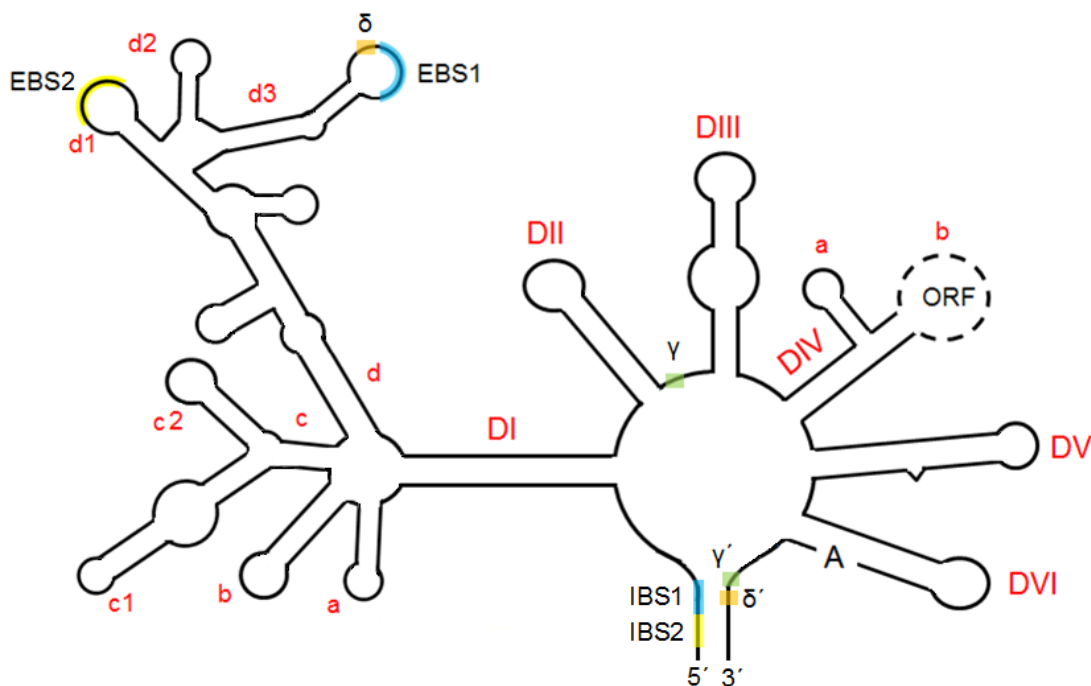


Abbildung 1: Konservierte Sekundärstruktur von Gruppe IIA Introns. Sechs doppelhelikale Domänen (DI-DVI) gehen von einem zentralen Rad aus. Dabei signalisieren kleine Buchstaben die Subdomänen. Subdomäne IVb kodiert für das multifunktionale Protein, welches Subdomäne IVa bindet und damit die Tertiärstruktur stabilisiert. Die Exon-bindenden Sequenzen EBS, δ und γ , binden die Intron-bindenden Sequenzen IBS, γ' und δ' , wodurch die Exon-Sequenzen für den nukleophilen Angriff des Adenosins in DVI positioniert werden. (Modifiziert nach Lambowitz & Zimmerly 2004)

Die Sekundärstruktur besteht aus sechs Domänen (DI-DVI), die von einem zentralen Rad ausstrahlen (Michel & Ferat 1995). Damit die native Tertiärstruktur ausgebildet werden kann, muss diese stabilisiert werden. Das geschieht *in vivo* über eine Bindung an ein multifunktionales Protein, welches von dem Intron kodiert wird (IEP). Das IEP besitzt eine Reverse Transkriptase-Domäne, eine Maturase-Domäne, eine DNA-Bindedomäne und eine Endonuklease-Domäne. Über die Maturase-Domäne und Teile der Reverse Transkriptase-Domäne bindet das IEP an die RNA, was preferentiell *in cis* geschieht (Moran et al. 1994; Cui et al. 2004; Dai et al. 2008). *In vitro* wird die RNA-Struktur durch hohe Temperaturen und Salzkonzentrationen stabilisiert, welche nicht den physiologischen Bedingungen entsprechen (Peebles et al. 1986; Schmelzer & Schweyen 1986; Van der Veen et al. 1986).

Nachdem die native Tertiärstruktur ausgebildet wurde übernehmen die Domänen spezifische Funktionen (zusammengefasst in Fedorova & Zingler 2007). Dabei fungiert DI als Gerüst für den Aufbau der katalytisch aktiven Tertiärstruktur indem es mit anderen Domänen interagiert. DII und DIII werden ebenfalls für die Ausbildung der Tertiärstruktur benötigt und beschleunigen das Spleißen (Pyle 2010; Toor et al. 2010). Domäne IVb kodiert für das IEP, welches eine hochaffine Bindungsstelle in DIVa bindet und dadurch die aktive Tertiärstruktur stabilisiert (Watanabe & Lambowitz 2004). DV konnte anhand einer Kristallstruktur als aktives Zentrum identifiziert werden (Toor et al. 2008). Ein konserviertes, nicht basengepaartes Adenosin in DVI wird in der Spleiß-Reaktion benötigt (Abbildung 2). Damit die Spleiß-Reaktion stattfinden kann müssen die Exons zu dem katalytischen Zentrum und dem Adenosin positioniert werden. Dies geschieht über Interaktionen zwischen Intron- und Exon-bindenden Sequenzen (Robart & Zimmerly 2005).

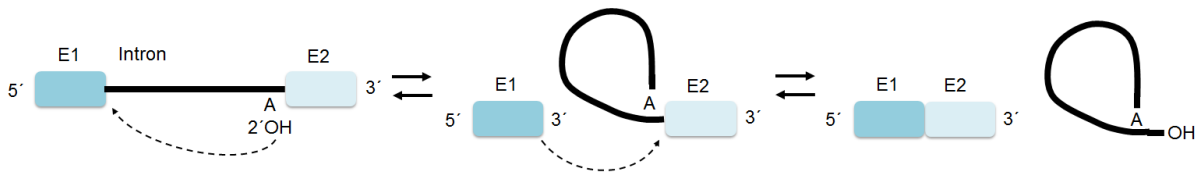


Abbildung 2: Spleißen und reverse Spleißen von Gruppe II Introns. In einer S_N2 -Reaktion attackiert das 2'-OH eines konservierten Adenosins nukleophil die 5'-Spleißstelle. Dabei entsteht ein Intron-Lariat/3'-Exon-Intermediat. Anschließend agiert das 3'-OH des 5'-Exons als Nukleophil und attackiert die 3'-Spleißstelle. Dadurch kommt es zu einer Ligation der Exons und dem Ausschneiden des Intron-Lariats (Modifiziert nach Lambowitz & Zimmerly 2011).

In Abbildung 2 ist der *in vivo* vorherrschende Spleiß-Mechanismus von Gruppe II Introns abgebildet. Dabei initiiert das nicht basengepaarte Adenosin in DVI zwei aufeinanderfolgende Umesterungsreaktionen (Cech 1993). Zunächst agiert die 2'-Hydroxylgruppe des Adenosins als Nukleophil und attackiert das Phosphat am 5'Ende des Introns. Dabei kommt es zu einer ungewöhnlichen 2'-5'-Phosphodiesterbindung, die zur Ausbildung einer Lassoform (Lariat) des Introns und dem Freisetzen des 5'-Exons führt. Im zweiten Schritt der Reaktion attackiert die freie 3'-Hydroxylgruppe des 5'-Exons nukleophil die Spleißstelle am 3'-Exon. Dadurch kommt es zur Verknüpfung beider Exons und dem Freisetzen des Lariats (Peebles et al. 1986). Der Spleißvorgang lässt sich umkehren, was als reverse Spleißen bezeichnet wird.

Das reverse Spleißen hat eine Bedeutung im Retrohoming (Abbildung 3). Dafür muss jedoch zunächst die Zielsequenz erkannt werden. Dies geschieht für das Gruppe IIC Intron aus *Bacillus halodurans in vitro* über eine Haarnadelschleifenstruktur ähnlich einem Rho-unabhängigen Transkriptionsterminator, in dessen Nähe das Intron *in vivo* inseriert (Robart, Seo & Zimmerly 2007). Anschließend interagiert die DNA-Bindedomäne des IEP mit der Zielsequenz und fördert die lokale Trennung der doppelsträngigen DNA. An den entstehenden einzelsträngigen Bereich bindet die Intron-RNA über Interaktionen zwischen den Intron- und

Exon-bindenden Sequenzen (Singh & Lambowitz 2001). Daraufhin kann die RNA in den Sense-Strang der DNA inserieren. Dies geschieht unabhängig von der RecA-vermittelten homologen Rekombination über reverses Spleißen, wodurch die RNA linear zwischen den beiden Exons vorliegt (Cousineau et al. 1998). Die Endonuklease-Domäne des IEP schneidet den gegenüberliegenden Strang, wodurch die entstehende 3'-Hydroxylgruppe als Primer für die reverse Transkription der inserierten RNA verwendet werden kann. Alternativ können Gruppe II Introns ohne Endonuklease-Domäne den entstehenden Strang während der DNA-Replikation als Primer verwenden (Zhong & Lambowitz 2003). Da die Zielsequenz als Primer für die reverse Transkription dient, handelt es sich um einen TPRT-Mechanismus (Zimmerly et al. 1995). Die cDNA-Integration wird über Wirts-Enzyme vermittelt. Der RNA-Strang wird durch die RNase H entfernt, der Zweitstrang von der Polymerase III synthetisiert und die DNA-Stränge durch die DNA-Ligase verknüpft (Smith et al. 2005).

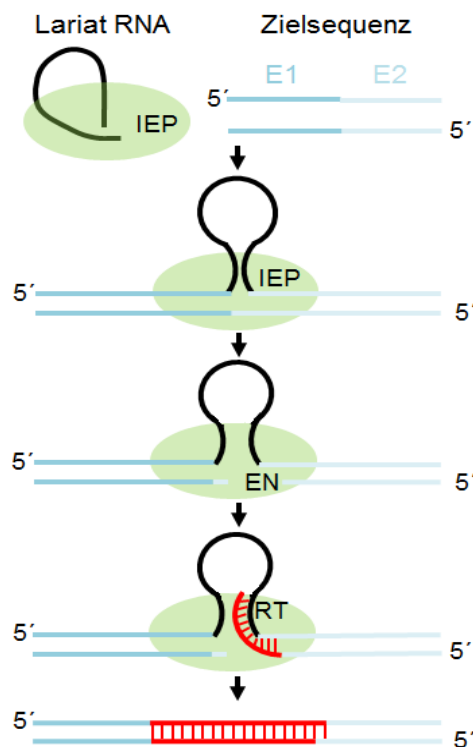


Abbildung 3: Retrohomie von Gruppe II Introns. Das Intron-Lariat ist an das Intron-kodierte Protein (IEP) gebunden, welches die Zielsequenz erkennt. Daraufhin kann die RNA revers in den Sense-Strang der DNA spließen. Der gegenüberliegende Strang wird durch die Endonuklease-Domäne (EN) geschnitten und das Intron durch die Reverse Transkriptase-Domäne (RT) revers transkribiert. Die Insertion wird durch Replikations- und Reparaturmechanismen fertiggestellt.

Gruppe II Introns sind evolutionär mit den diversitätsgenerierenden Retroelementen verwandt, welche ihrem Wirt einen selektiven Vorteil vermitteln können (Liu et al. 2002; Doulatov et al. 2004).

1.1.2 Diversitätsgenerierende Retroelemente

Das erste diversitätsgenerierende Retroelement (DGR) wurde im *Bordetella*-Phagen beschrieben (Liu et al. 2002). Dieser Phage infiziert Bakterien der Spezies *Bordetella*, die ihre Oberflächenproteine innerhalb des Infektionszyklus verändern können. Darunter befindet sich auch der Phagenrezeptor, weshalb der Phage nur Bakterien einer bestimmten Phase im Infektionszyklus infizieren kann. Durch das in Abbildung 4 dargestellte DGR kann der Phage den Rezeptor wechseln und Bakterien anderer Phasen infizieren. Die Anpassung des Phagen an die dynamische Oberflächenstruktur des Wirts wird über eine beschleunigte Proteinevolution des Liganden vermittelt. Neben dem *Bordetella*-Phagen wurden weitere DGRs, unter anderem in *Nostoc sp.* PCC7120 und *Treponema denticola*, identifiziert (Doulatov et al. 2004).

Nostoc sp. PCC7120 ist ein cyanobakterieller Modellorganismus für Untersuchungen an Heterocysten (Mitschke et al. 2011). In dem Genom befinden sich zwei DGRs, von denen eines (RT: *alr5014*) aufgrund unspezifischer Mutationen als inaktiv gewertet wird. In dieser Arbeit wurde das andere DGR (RT: *alr3497*) untersucht (Abbildung 4).

Treponema denticola ist ein humanpathogener Spirochät, der mit Parodontitis assoziiert wird (Loesche & Grossman 2001). Für das DGR aus *Treponema denticola* wurden sieben potentielle Zielgene beschrieben (Le Coq & Ghosh 2011). In dieser Arbeit wurde das zur RT nächstgelegene Zielgen (TDE2269) analysiert (Abbildung 4).

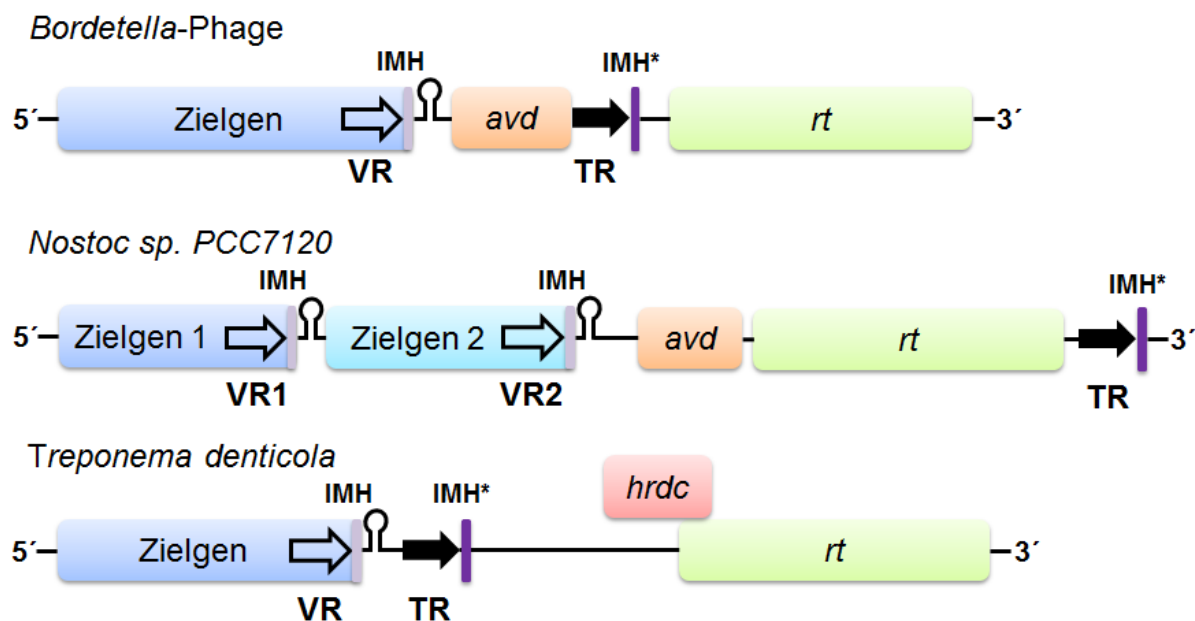


Abbildung 4: Aufbau von DGRs. Die DGRs besitzen eine reverse Transkriptase (RT), sowie ein oder mehrere Zielgene mit variabler Region (VR) am 3'-Ende. Die VR ist fast identisch mit dem Templat Repeat (TR). VR und TR besitzen am 3'-Ende eine Region, die sich in einigen Positionen voneinander unterscheidet und als IMH („initiation of mutagenic homing“) bezeichnet wird. Wenn das DGR nicht für die *Avd* („accessory variability determinant“) kodiert, überlappt der RT-ORF mit einer Helikase und RNaseD C-terminalen (HRDC) Domäne. Die Ausbildung einer Haarnadel/Kreuz-Struktur downstream zur VR ist eingetragen.

Die Zielgene der DGRs kodieren für die variablen Proteine. Eine Kenntnis über die Funktion des variablen Proteins kann entscheidend für das Verständnis der biologischen Funktion des DGRs sein. Das Zielgen des *Bordetella*-Phagen *mtd* („major tropism determinant“) kodiert für ein Schwanzfaserprotein das in die Rezeptorerkennung involviert ist und erklärt weshalb der Phage den Rezeptor wechseln kann (Liu et al. 2002). Das Zielgen von *Treponema denticola* (TDE2269) kodiert für ein variables Protein (TvpA) das an der äußeren Bakterienoberfläche lokalisiert ist und dadurch mit der Umwelt interagieren kann (Le Coq & Ghosh 2011). Mtd, TvpA und die variablen Proteine aus *Nostoc sp.* PCC7120 werden als Formylglycinegenerierende Enzyme (FGEs) annotiert. Diese Annotation wurde für ca. 27 % der variablen Proteine beobachtet (Schillinger & Zingler 2012).

Die Komponente der DGRs, welche hypermutiert wird, ist die variable Region (VR) am 3´-Ende des Zielgens. Die VR besitzt mehrere variable Positionen, die mit jedem der vier Nukleotide besetzt werden können. Daher ergibt sich eine theoretisch maximale Sequenzdiversität von 4^X . Für die Berechnung der theoretisch maximalen Aminosäurediversität ist die Lage der variablen Positionen innerhalb der Kodons entscheidend. Da diese vorrangig an den ersten beiden Positionen eines Kodons lokalisiert sind, ist die Chance auf einen Aminosäureaustausch massiv erhöht (Liu et al. 2002; Medhekar & Miller 2007). Die variablen Positionen entsprechen Adeninen in dem Templat Repeat (TR). Dieser ist, bis auf die Adenine, fast identisch mit der VR.

Am 3´-Ende von VR und TR ist ein nicht variabler Bereich vorzufinden, der als IMH („initiation of mutagenic homing“) bezeichnet wird (Doulatov et al. 2004). Die IMH des *Bordetella*-Phagen besteht aus einer 14 bp langen GC-reichen Region, gefolgt von einer 21 bp langen Region mit fünf Unterschieden zwischen VR (IMH) und TR (IMH*). Der Aufbau der IMH ist zwischen DGRs nicht konserviert. Dennoch ist eine Region vorzufinden, die keine Adenine enthält (G/C/T-Region). Darauf folgt eine Region die in der Sequenz über einige Abweichungen zwischen VR und TR verfügt. In dieser Arbeit wurde angenommen, dass diese Regionen der IMH des *Bordetella*-Phagen entsprechen.

Desweiteren gehört eine Haarnadelstruktur oder eine kreuzförmige DNA-Struktur downstream der VR zum DGR. Die Sequenz der Schleife ist in Phagen-assoziierten DGRs konserviert, zeigt aber Abweichungen in anderen DGRs. Die Ausbildung der Struktur wurde *in vitro* durch einen Struktur-spezifischen Nuklease-Assay bestätigt und wird *in vivo* für die Generierung der Variabilität benötigt (Guo et al. 2011; Arambula et al. 2013).

Eine weitere Komponente der DGRs ist die reverse Transkriptase (RT). Diese verfügt über sechs gemeinsame Domänen mit anderen RTs (Simon & Zimmerly 2008). In Domäne fünf befindet sich bei allen RTs das YxDD-Motiv, welches für die katalytische Aktivität notwendig ist. In Domäne vier konnte dagegen ein Motiv identifiziert werden das diagnostisch für DGR-kodierte RTs ist. Dabei handelt es sich um das SQ-Motiv mit der Konsensus-Sequenz (L/I/V)GxxxSQ (Medhekar & Miller 2007; Schillinger et al. 2012).

Die Avd („accessory variability determinant“) ist nicht immer Teil des DGRs (s. *Treponema denticola*). In solchen Fällen überlappt der RT-ORF mit einer Helikase und RNaseD C-terminalen (HRDC) Domäne (Medhekar & Miller 2007). Sowohl für die HRDC-Domäne als auch für die Avd wird eine Interaktion mit Nukleinsäuren postuliert (Liu et al. 1999; Alayyoubi et al. 2013). Daher könnte die HRDC-Domäne die Funktion der Avd übernehmen, wenn diese nicht vom DGR kodiert wird. Ein Merkmal der Avd ist das GGW-Motiv mit der Konsensus-Sequenz [L/I/V]G[K/R]xxGGWxK am C-Terminus (Schillinger 2013). Die Bedeutung des Motivs ist jedoch noch nicht geklärt.

Die wichtigsten Eigenschaften der DGRs aus dem *Bordetella*-Phagen, *Nostoc sp.* PCC7120 und *Treponema denticola* sind in Tabelle 1 zusammengefasst.

Tabelle 1: Eigenschaften der beschriebenen DGRs

	<i>Bordetella</i> - Phage	<i>Nostoc sp.</i> PCC7120	<i>Treponema</i> <i>denticola</i>
Größe	2828 bp	3483 bp	2879 bp
SQ-Motiv (RT)	IGSLTSQ	IGNLTSQ	
GGW-Motiv (Avd)	VGRILGSWIA	IGTELGGWIK	-----
Zielgen	Homologie zur FGE-Familie		
Länge der VR	134	137	136
IMH	(G/C)14, 21 bp Sequenz mit 5 Abweichungen	(G/C/T)29, 11 bp Sequenz mit 2 Abweichungen	(G/C/T)27 bp Sequenz mit 2 Abweichungen
Variable Positionen	23	31	40
Nukleotiddiversität	$7,0 \times 10^{13}$	$4,6 \times 10^{18}$	$1,2 \times 10^{24}$
Aminosäurediversität	$9,2 \times 10^{12}$	$5,0 \times 10^{17}$	$6,0 \times 10^{20}$
Haarnadel/Kreuz- Struktur	Stamm: 8 bp Schleife: 5'GAAA'3	Stamm: 7 bp Schleife: 5'TTT'3	Stamm: 4 bp Schleife: 5'AGG'3

1.1.2.1 Der Mechanismus von diversitätsgenerierenden Retroelementen

Das DGR kann einen definierten Genabschnitt hypermutieren und seinem Wirtorganismus dadurch einen selektiven Vorteil vermitteln. Der postulierte Mechanismus des DGRs beinhaltet einen zielgerichteten Sequenztransfer von dem TR über ein RNA-Intermediat zur VR, der von Adenin-spezifischen Mutationen begleitet wird (Abbildung 5). Dabei wird das Retrohoming der mutierten Sequenz als mutagenes Homing bezeichnet.

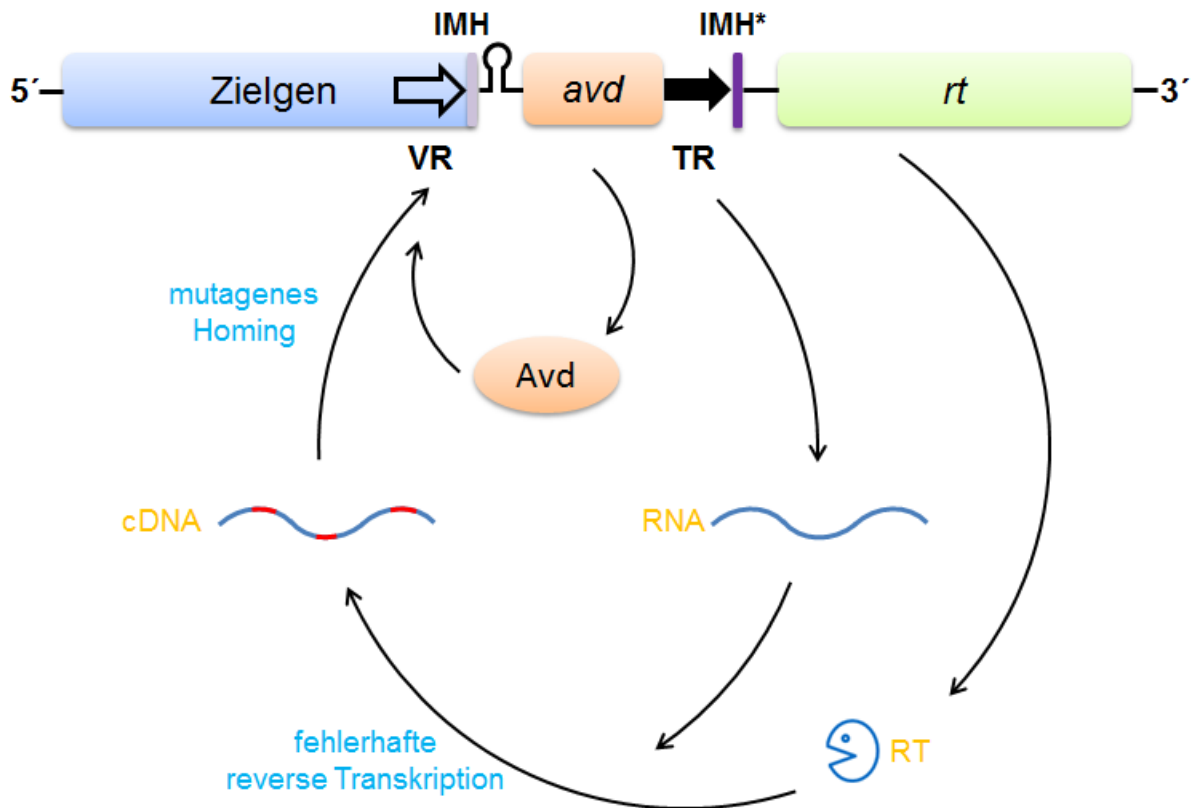


Abbildung 5: Schematische Darstellung des DGR-Mechanismus. Das DGR kodiert für eine fehleranfällige reverse Transkriptase (RT), welche die RNA des Templat Repeats (TR) in cDNA umschreibt und dabei mutiert. Die cDNA ersetzt in einem Prozess, der als mutagenes Homing bezeichnet wird, die variable Region (VR) am 3'-Ende des Zielgens. Das mutagenes Homing könnte durch die Avd koordinierend unterstützt und die Haarnadel/Kreuz-Struktur downstream der VR für die Erkennung der Zielsequenz benötigt werden. Der Diversifizierungszyklus verläuft aufgrund der IMH/IMH*-Regionen gerichtet von dem TR zur VR.

Der Mechanismus des DGRs wurde anhand mehrerer Experimente postuliert. Dabei wurde der Sequenztransfer vom TR zur VR über eine stille Mutation im TR nachgewiesen, welche nach der Diversifizierung ebenfalls in der VR beobachtet werden konnte (Liu et al. 2002). Damit der Sequenztransfer gerichtet vom TR zur VR verläuft, wird die IMH benötigt. Dies wurde durch Experimente aufgeklärt, bei denen die IMH mit der IMH* vertauscht wurde (Doulatov et al. 2004).

Die Verwendung eines RNA-Intermediats für den Sequenztransfer konnte mit einem Versuch bestätigt werden, bei dem ein Gruppe I Intron mit flankierenden Exons in den TR eingebracht wurde (Guo et al. 2008). Nach dem Sequenztransfer konnten die ligierten Exons, mit den charakteristischen Mutationen an den variablen Positionen, in der VR identifiziert werden.

Das RNA-Intermediat (Templat RNA) wurde auf Mutationen untersucht die spezifisch Adenine betreffen. Dabei konnten keine Mutationen beobachtet werden (Guo et al. 2008). Demnach könnten die beobachteten Mutationen in der VR durch die DGR-kodierte RT entstanden sein, wenn diese eine fehlerhafte Adenin-Erkennung aufweist. Das SQ-Motiv der RT könnte für die Adenin-spezifische Mutagenese verantwortlich sein, da das korrespondierende Motiv (GQxxxSP) in der RT des humanen Immundefizienz-Virus (HIV), an der Koordinierung von Templat und dNTPs im aktiven Zentrum beteiligt ist (Huang et al. 1998). Die HIV-RT ist ebenfalls ein Beispiel für ungenau arbeitende RTs, fördert die genetische Variation jedoch nicht über eine basenspezifische Ungenauigkeit (Preston, Poesz & Loeb 1988).

Das mutagene Homing besitzt Parallelen zu dem Retrohoming der Gruppe II Introns (Kap. 1.1.1). Die Erkennung der Zielsequenz könnte analog zu dem Gruppe IIC Intron aus *Bacillus halodurans* über die Haarnadel/Kreuz-Struktur downstream der VR stattfinden (Guo et al. 2011). Auch scheint es trotz der Homologie zwischen VR und TR keine RecA-vermittelte homologe Rekombination zu geben, da ein Sequenztransfer sowohl im Wildtyp, als auch in einem RecA-Knockout-Stamm beobachtet wurde (Guo et al. 2008). Als Primer für die cDNA-Synthese wurde die Verwendung des entstehenden Strangs während der DNA-Replikation ausgeschlossen. Stattdessen wird ein Endonuklease-abhängiger TPRT-Mechanismus postuliert, für den die Endonuklease noch identifiziert werden muss. Ein möglicher Kandidat wäre eine katalytisch aktive Templat RNA (Guo et al. 2011). Die cDNA-Integration am 5'-Ende der VR erfordert einen homologen Bereich zwischen VR und cDNA und könnte über einen Templat-Wechsel der DGR-kodierten RT erfolgen (Guo et al. 2008). Die entstehenden Unterschiede an den variablen Positionen könnten während der DNA-Replikation behoben werden (Doulatov et al. 2004).

Die enorme Aminosäurediversität der variablen Proteine (10^{20} in *Treponema denticola*) muss von der Proteinstruktur toleriert werden, während die variable Region an der Oberfläche präsentiert wird. Dies geschieht durch die Ausbildung einer CLec-Faltung innerhalb der C-terminalen Domäne des variablen Proteins (McMahon et al. 2005; Medhekar & Miller 2007). Die CLec-Faltung ist damit vergleichbar mit der Immunglobulin-Faltung des Immunsystems, welche die hypervariablen Regionen der Antikörper und T-Zell-Rezeptoren toleriert ($\sim 10^{14}$ - 10^{16} Sequenzen) (Chothia & Lesk 1987; Davis & Bjorkman 1988).

Die Avd ist notwendig für die Generierung der Diversität (Doulatov et al. 2004). Über die genaue Funktion der Avd ist aber nichts bekannt. Aufgrund eines nukleinsäurebindenden Charakters und einer spezifischen Bindung an die RT werden jedoch zwei Funktionen postuliert. Zum einen könnte Avd die katalytische Aktivität der RT verstärken. Dies konnte jedoch nicht zufriedenstellend überprüft werden, da die Primärsequenz der Templat RNA, welche das Substrat der RT darstellt, nicht bekannt ist. Desweiteren ist eine Funktion von Avd vorstellbar, welche das mutagene Homing koordinierend unterstützt (Alayyoubi et al. 2013).

Die gesammelten Daten der DGRs lassen auf einen konservierten Mechanismus schließen (Medhekar & Miller 2007). Ein Verständnis von diesem Mechanismus würde es erlauben DGRs in Verfahren einzusetzen, bei denen Proteine diversifiziert werden. Die DNA-Sequenz des zu testenden Proteins kann dabei in den TR eingebracht werden. Der TR toleriert Sequenzen bis zu 200 bp und transferiert diese mit den charakteristischen Mutationen zur VR (Guo et al. 2008). Diese Verwendung von DGRs wurde bereits erfolgreich für die Diversifizierung von Phagen-Display Bibliotheken angewendet (Overstreet et al. 2012). Dadurch könnte die Entwicklung neuer biologischer Arzneistoffe beschleunigt werden.

1.1.2.2 Verbreitung von diversitätsgenerierenden Retroelementen

Der potentielle Nutzen einer beschleunigten Proteinevolution durch das DGR könnte auch in anderen Organismen Verwendung finden. Um die Verbreitung von DGRs zu untersuchen, wurde zunächst von Hand und später durch das bioinformatische Programm DiGReF („Diversity Gerating Retroelement Finder“) nach weiteren Vertretern gesucht (Doulatov et al. 2004; Schillinger et al. 2012). Durch DiGReF konnten 155 DGRs in den öffentlichen Sequenzdatenbanken identifiziert werden. Diese wurden am häufigsten in den Genomen von Bakterien, aber auch in Phagen und Plasmiden vorgefunden. Da über 6000 Genome von Prokaryoten und Phagen auf die Präsenz von DGRs untersucht wurden, erscheint die Anzahl der identifizierten DGRs (< 3 %) gering (Schillinger et al. 2012). Eine mögliche Ursache stellen die konstanten Kultivierungsbedingungen der sequenzierten Stämme dar. Diesen fehlt infolge ein Anpassungsdruck auf sich ändernde Umweltbedingungen. Dadurch wird das DGR nicht benötigt und könnte als genomischer Ballast entfernt werden. Etwas Vergleichbares konnte für Virulenzfaktoren in kultivierten Bakterien der Gattung *Shigella* beobachtet werden (Chosa et al. 1989). Daher wurde die Suche nach DGRs auf Metagenomdatenbanken ausgeweitet, wobei Hinweise auf eine stärkere Verbreitung unter natürlichen Bedingungen gefunden wurden (Schillinger & Zingler 2012; Ye 2014).

1.1.3 Zielsetzung

Ein bioinformatischer Ansatz zur Identifizierung von DGRs in öffentlichen Sequenzdatenbanken konnte zeigen, dass nur 6,5 % (9) der analysierten Cyanobakterien mit DGRs assoziiert sind (Schillinger et al. 2012). Im Hinblick auf den potentiellen Nutzen einer beschleunigten Proteinevolution durch das DGR scheint dieser Prozentsatz sehr gering. Eine Hypothese verbindet das geringe Vorkommen mit den konstanten Kultivierungsbedingungen der sequenzierten Stämme (Schillinger & Zingler 2012). Unter diesen Bedingungen werden DGRs nicht benötigt und könnten als genomischer Ballast entfernt werden.

Durch eine Kooperation mit der Abteilung Pflanzenökologie & Systematik (TU Kaiserslautern) ergab sich die Möglichkeit die oben genannte Hypothese experimentell zu überprüfen und neue DGRs zu identifizieren. Dafür sollte ein PCR-Screening etabliert werden, welches DGRs in kultivierten Cyanobakterien systematisch und effizient erfassen kann. Um die komplette Sequenz der identifizierten DGRs zu erhalten, sollte zudem die inverse PCR als nachgeschaltete Methode getestet werden.

Ein weiteres Ziel dieser Arbeit war es einen Eindruck zu bekommen, wie Häufig eine Diversifizierung der VR in bakteriellen DGRs auftritt. Dafür sollte die Aktivität der identifizierten DGRs über einen längeren Zeitraum beobachtet werden. Eine beobachtete Aktivität würde zudem zeigen, dass auch die einzelnen Komponenten funktionsfähig sind, was für weitere Untersuchungen an diesen Komponenten sinnvoll wäre.

Eine dieser Untersuchungen beschäftigte sich mit der Templat RNA, welche bereits als mobile Komponente für den Sequenztransfer identifiziert wurde und für die eine weitere Funktion als Endonuklease postuliert wurde (Guo et al. 2008; Guo et al. 2011). Diese Funktion würde durch die Struktur vermittelt werden, die bislang jedoch für keine Templat RNA aufgeklärt wurde. Daher sollte die Templat RNA-Struktur eines aktiven DGRs als weiteres Ziel dieser Arbeit bestimmt werden. Dafür mussten zunächst die Transkriptenden der Templat RNA identifiziert werden. Anschließend bot sich für die Strukturaufklärung eine Kombination aus bioinformatischer Berechnung und experimentell erlangten Daten an. Ausgehend von der Struktur könnten anschließende Analysen Strukturelemente aufzeigen, die für eine spezifische Bindung zur DGR-kodierten RT benötigt werden und/oder der Templat RNA eine katalytische Aktivität verleihen.

2 Material und Methoden

2.1 Material

2.1.1 Chemikalien

Chemikalien wurden im molekularbiologischen Grad von folgenden Firmen erworben: Sigma-Aldrich (Steinheim), AppliChem (Darmstadt), New England Biolabs (NEB) (Frankfurt a. M.), Serva (Heidelberg), Roth (Karlsruhe), Fluka (Deisenhofen), Merck (Darmstadt), ChemSolute (Renningen), Roche (Mannheim) und Invitrogen (Darmstadt). [³²P]-markierte Radionuklide wurden von PerkinElmer (Baesweiler) und Hartmann Analytik (Braunschweig) bezogen.

2.1.2 Enzyme

DNA und RNA modifizierende Enzyme wurden von folgenden Herstellern bezogen und nach deren Angaben verwendet: New England Biolabs (NEB) (Frankfurt a. M.), Roche (Mannheim), Bioline (Luckenwalde), Epicentre (USA), Fisher Scientific (Schwerte) und Invitrogen (Darmstadt).

2.1.3 Geräte

Eine Auswahl der verwendeten Geräte ist in Tabelle 2 zusammengestellt.

Tabelle 2. Verwendete Geräte

Gerät	Hersteller
Cyclone Phosphor Imager	PerkinElmer Inc, Rodgau
Gel iX Imager Geldokumentationssystem	Intas Science Imaging Instruments GmbH, Göttingen
Innova Incubation Shaker 44/44R	New Brunswick/Eppendorf AG, Hamburg
Kühlmikrozentrifuge 5415R	Eppendorf AG, Hamburg
MyCycler Thermocycler	Bio-Rad Laboratories GmbH, München
Qubit 1.0 Fluorometer	Invitrogen/Life Technologies GmbH, Darmstadt
Vortex-Schüttler VF2	IKA-Werke GmbH & Co. KG, Staufen

2.1.4 Kit-Systeme

Die Kit-Systeme sind in Tabelle 3 aufgelistet und wurden nach Herstellerangaben verwendet.

Tabelle 3. Verwendete Kit-Systeme

Kit-System	Hersteller
High Pure RNA Isolation Kit for total RNA	Roche Diagnostics GmbH, Mannheim
HiYield® Plasmid Mini Kit	Süd-Laborbedarf GmbH, Gauting
NEBlot® Kit	New England Biolabs GmbH, Frankfurt a. M.
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel, Düren
QIAquick Nucleotide Removal Kit	Qiagen, Hilden

2.1.5 Bakterienstämme

Die in der Arbeit verwendeten Bakterienstämme sind in Tabelle 4 beschrieben

Tabelle 4. Verwendete Bakterienstämme

Stamm	Eigenschaften	Referenz
<i>Nostoc sp.</i> PCC7120	Wildtyp	Sammlung von Algenkulturen, Göttingen (SAG 25.82)
<i>Anabaena</i> <i>flos-aquae</i>	Wildtyp	Pflanzenökologie & Systematik, TU Kaiserslautern (BB 97.35)
DH5α T1R	F' φ80lacZΔM15 Δ(lacZYA-argF)U169 recA1 endA1 hsdR17(rk-,mk+) phoA supE44 λ- thi-1 gyrA96 relA1 tonA	Invitrogen/ Life Technologies GmbH, Darmstadt
BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHI ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	New England Biolabs GmbH, Frankfurt
T7 Express	fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC- mrr)114::IS10	New England Biolabs GmbH, Frankfurt

Für sämtliche Klonierungen wurde der *Escherichia coli* (*E. coli*)-Stamm DH5 α verwendet. Die Stabilitätstests wurden zudem mit den *E. coli*-Stämmen BL21(DE3) und T7 Express durchgeführt.

2.1.6 Medien

Die in der Arbeit verwendeten Medien sind in Tabelle 5 aufgeführt.

Tabelle 5. Verwendete Medien

LB Medium	Z-Medium	BG11-Medium
10 g Trypton	21.0 mg Na ₂ CO ₃	1.5 g NaNO ₃
5 g Hefeextrakt	6.42 mg Fe(III)-EDTA	40 mg K ₂ HPO ₄ ·3H ₂ O
10 g NaCl	25.0 mg MgSO ₄ ·7H ₂ O	75 mg MgSO ₄ ·7H ₂ O
Ad. 1l dH ₂ O	31.0 mg K ₂ HPO ₄	36 mg CaCl ₂ ·2H ₂ O
	59.0 mg Ca(NO ₃) ₂ ·4H ₂ O	6 mg Citronensäure
	467.0 mg NaNO ₃	1 mg Na ₂ EDTA
	0.08 ml/l Spurenelementlösung ¹	6 mg Ammoniumeisen(III)-citrat
	Ad. 1l dH ₂ O	20 mg Na ₂ CO ₃
		1ml/l Spurenelementlösung ²
		Ad. 1l dH ₂ O

Die Medien wurden anschließend autoklaviert. Um LB-Agar zu erzeugen, wurden vor dem autoklavieren 15 g Agar/L zu dem LB-Medium hinzugefügt. Das LB-Medium/Agar wurde für *E. coli*, das BG11-Medium für *Nostoc sp.* PCC7120 und das Z-Medium für *Anabaena flos-aquae* verwendet. Für die Medien der Cyanobakterien wurden separat Spurenelementlösungen angesetzt (Tabelle 6).

Tabelle 6. Spurenelementlösungen der Medien (100 ml)

Komponente	H. Gaffron ¹ (Z-Medium)	Kuhl & Lorenzen ² (BG11-Medium)
H ₃ BO ₃	310.0 mg	6.1 mg
MnSO ₄ ·H ₂ O	16.9 mg	16.9 mg
Na ₂ WO ₄ ·2H ₂ O	3.3 mg	
(NH ₄) ₆ Mo ₇ O ₂₄ ·4H ₂ O	8.8 mg	1.25 mg
KBr	11.9 mg	
KJ	8.3 mg	
ZnSO ₄ ·7H ₂ O	28.7 mg	28.7 mg
Cd(NO ₃) ₂ ·4H ₂ O	15.4 mg	
Co(NO ₃) ₂ ·6H ₂ O	14.6 mg	
CuSO ₄ ·5H ₂ O	12.5 mg	0.25 mg
NiSO ₄ (NH ₄) ₂ SO ₄ ·6H ₂ O	19.8 mg	
Cr(NO ₃) ₃ ·7H ₂ O	3.7 mg	
VO ₂ SO ₄ ·2H ₂ O	2.0 mg	
Al ₂ (SO ₄) ₃ K ₂ SO ₄ ·24H ₂ O	47.4 mg	

2.1.7 Lösungen und Puffer

Die verwendeten Lösungen und Puffer sind nachfolgend genannt und wurden, wenn möglich, nach dem Ansetzen autoklaviert bzw. sterilfiltriert (Porengröße 0,2 µm).

<u>DNA-Ladepuffer (6x)</u>	<u>PAGE-Elutionspuffer</u>	<u>TAE-Puffer (50x)</u>
0.15 % Farbstoff	0.3 M NaCl	2 M Tris-Acetat
50 mM Tris, pH 7.6	10 mM MOPS, pH 6.0	0.1 M Na ₂ EDTA, pH 8.2
60 % Glycerol	1 mM Na ₂ EDTA, pH 8.0	
10 mM EDTA		<u>TBE-Puffer (10x)</u>
	<u>PAGE-Ladepuffer (2x)</u>	0.89 M Tris
<u>Faltungsmix (3x)</u>	40 µl Saccharose (50 %)	0.89 M Borsäure
333 mM HEPES, pH 8.0	10 µl TBE (10x)	20 mM Na ₂ EDTA, pH 8.3
50 mM MgCl ₂	10 µl XC (2 %)	
333 mM NaCl	40 µl DEPC-H ₂ O	<u>TE-Puffer (0.5x)</u>
	400 µl Formamid	5 mM Tris-HCl
<u>KCM-Puffer (5x)</u>	55 µl EDTA (0,5 M)	0.5 mM EDTA, pH 8.0
10 ml KCl (1 M)	5 min auf 65 °C erhitzen	
3 ml CaCl ₂ (1 M)	bei 4 °C lagern	<u>TSB-Puffer</u>
5 ml MgCl ₂ (1 M)		5 ml LB (2x)
Ad. 20 ml dH ₂ O	<u>PBS-Puffer</u>	1 g PEG 3350
	137 mM NaCl	1 ml MgCl ₂ (100 mM)
<u>Lysepuffer</u>	6.5 mM Na ₂ HPO ₄	1ml MgSO ₄ (100 mM)
4 M Guanidinthiocyanat	1.5 mM K ₂ HPO ₄	0,5 ml Dimethylsulfoxid
20 mM Na-Acetat, pH 5,2	2.7 mM KCl	Ad. 10 ml H ₂ O
0,1 mM DTT	pH 7.4	
0,5 % N-Lauryl-Sarcosin	<u>RNA-Ladepuffer</u>	<u>Transkriptionspuffer (10x)</u>
bei 65 °C lösen	15 µl MOPS (10x)	0.15 M MgCl ₂
vor gebrauch 10 µl/ml	25 µl Formaldehyd (37 %)	0.4 M Tris-HCl
β-Mercaptoethanol zugeben	50 µl Formamid	0.02 M Spermidin
	10 µl Glycerin/ Bromphenolblau	0.05 M DTT
<u>MOPS-Puffer (10x)</u>	5 µl Ethidiumbromid	Ad. 10 ml DEPC-H ₂ O
200 mM MOPS		<u>Saure Stopp-Lösung</u>
50 mM Na-Acetat	<u>SSC-Puffer (20x)</u>	85 % Formamid
10 mM Na ₂ EDTA	3 M NaCl	0,5x TBE
pH 6.8 (NaOH)		50 mM EDTA, pH 8.0
Add. 0.1 % DEPC	0.3 M Natriumcitrat, pH 7.0	Bromphenolblau
37 °C Übernacht inkubieren		

2.2 Methoden

2.2.1 Arbeiten mit Bakterien

2.2.1.1 Kultivierung von *Escherichia coli*

Bakterien wurden in Luria-Bertani (LB)-Medium bzw. LB-Agarplatten bei 37 °C kultiviert (Bertani 1951). Um selektive Wachstumsbedingungen für transformierte Bakterien zu schaffen, wurden Medium und Platten mit Ampicillin (Endkonzentration: 100 µg/ml) oder Kanamycin (Endkonzentration: 50 µg/ml) versetzt. Flüssigkulturen wurden ÜN bei 37 °C und 180 rpm auf einem Schüttler inkubiert. Kulturen auf Nährboden wurden ebenfalls ÜN bei 37 °C inkubiert. Um eine Lagerung der Bakterien über längere Zeit zu ermöglichen, wurden Glycerolstocks hergestellt. Dafür wurden 500 µl einer flüssigen ÜN-Kultur mit 500 µl sterilem Glycerol (15 %) vermischt. Die Glycerolstocks wurden bei -80 °C gelagert.

Stabilitätstest zur Untersuchung von heterolog exprimierten DGRs:

Transformanten, welche mit dem REBVA (Kap. 2.2.2.4) analysiert werden sollten, wurden zuvor einem Stabilitätstest unterzogen. Bei diesem Test wurden die Bakterien eine Woche über in Kultur gehalten. Für eine bessere Fitness der Bakterien, wurden diese maximal bis zu einer $OD_{600} \sim 0,8$ wachsen gelassen und so der Übergang in die stationäre Phase verhindert. Die Bakterien wurden morgens und abends in frisches LB-Medium überimpft ($\sim 1/1000$) und währenddessen Proben entnommen. Tagsüber (~ 8 h) wurden die Bakterien bei 37 °C und nachts (~ 16 h) bei 30 °C wachsen gelassen. Nach der Plasmidisolierung wurden die Proben auf einem Agarosegel auf ihre Stabilität untersucht.

2.2.1.2 Kultivierung von Cyanobakterien

Cyanobakterienstämme wurden in Erlenmeyerkolben bei 22 °C und einer Bestrahlungsstärke von 90-100 µE/m²*s auf einem Schüttler bei 120 rpm kultiviert. Dabei wurde eine OD_{750} von 0,5 bis 0,8 eingestellt. Der Stamm *Nostoc sp.* PCC7120 wurde mit BG11-Medium (Stanier et al. 1971) und der Stamm *Anabaena flos-aquae* mit Z-Medium (Staub 1961) kultiviert.

Stressbedingungen für Transkriptlevel-Analysen:

Für die Generierung der Stressbedingungen wurden 50 ml Cyanobakterien ($OD_{750} \sim 0.6$) für 3 Stunden bzw. 24 Stunden einer der folgenden Bedingungen ausgesetzt: Referenz (22 °C), Kältestress (4 °C), Hitzestress (37 °C), Stress durch reaktive Sauerstoffspezies (ROS) (5-10 mM H₂O₂), Dunkelstress und Salzstress (50-250 mM NaCl).

2.2.2 Arbeiten mit DNA

2.2.2.1 Klonierungen

Für die Herstellung von Plasmiden wurden die DNA-Fragmente von Interesse über die Polymerasekettenreaktion (polymerase chain reaction, PCR) amplifiziert und auf einem Agarosegel analysiert (Mullis et al. 1986). Die amplifizierten DNA-Fragmente wurden in Vektoren ligiert und anschließend in den Klonierungsstamm DH5 α transformiert. Positive Transformanten wurden über Kolonie-PCR identifiziert und anschließend sequenziert.

2.2.2.1.1 Präparative PCR

Für die präparative PCR wurde mit der Phusion DNA-Polymerase (Fisher Scientific, Schwerte) nach Herstellerangaben gearbeitet, da diese über eine Korrekturlesefunktion verfügt. Der Reaktionsansatz der PCR (20 μ l) setzt sich aus den Primern (je 250 nM), dem Templat (z.B. gDNA), dNTPs (je 250 μ M), der Phusion DNA-Polymerase (0,5 U) und 1x Phusion-Puffer (GC) zusammen. Der Reaktionsansatz wurde im MyCycler (Bio-Rad, München) mit einem angepassten Thermoprofil inkubiert.

Thermoprofil für die präparative PCR

Anfängliche Denaturierung der DNA	98 °C	5 min
Exponentielle Amplifikation (25 Zyklen):		
Denaturierung der DNA	98 °C	30 sec
Annealing der Primer	50-60 °C	25 sec
Primer-Elongation mit der Phusion	72 °C	15-30 sec/kb
Finale DNA-Elongation	72 °C	10 min

Für schwer zu klonierende Konstrukte wurden Primer verwendet die am 5'-Ende über Erkennungssequenzen für Restriktionsenzyme verfügen. Das geschnittene PCR-Produkt konnte über die Überhänge gerichtet in den Vektor ligiert werden, was die Effizienz der Ligation erhöht (Kap. 2.2.2.1.5).

2.2.2.1.2 Agarose-Gelelektrophorese von DNA

Zur Auftrennung der PCR-Produkte wurden Agarosegele verwendet. Diese eignen sich durch ihre Porengröße für die Auftrennung von Nukleinsäuren in einem Bereich von 0,5 bis 25 kb. Das verwendete Standard-Gel setzt sich aus 1 % Agarose, gelöst in 1x TAE-Puffer und einer Ethidiumbromid-Konzentration von 0.3 µg/ml zusammen. Dabei wird das Ethidiumbromid erst nach dem Aufkochen hinzugegeben. Die Proben wurden mit DNA-Ladepuffer versetzt und die Größe der Fragmente durch Vergleich mit einem parallel gelaufenen Marker bestimmt. Als Marker wurden entweder 0.5 µg pro Spur des „DNA Ladder 1 kb“ oder des „DNA Ladder 100 bp“ (NEB, Frankfurt) verwendet. Die Elektrophorese wurde bei einer konstanten Spannung von 100 V in 1x TAE-Laufpuffer durchgeführt. Analysiert wurden die Gele in einem Geldokumentationssystem (Intas, Göttingen).

2.2.2.1.3 Aufreinigung von DNA aus PCR-Produkten und Agarosegelen

Die Aufreinigung der DNA wurde mit dem „NucleoSpin Gel and PCR Clean-up Kit“ (Macherey-Nagel, Düren) nach Herstellerangaben durchgeführt. Dabei wurde je nach Zustand der analysierten PCR-Produkte auf dem Agarosegel in PCR-Aufreinigung (eine diskrete Bande) und Gelextraktion (mehrere PCR-Produkte) unterschieden. Anschließend wurde die Konzentration der DNA fluorometrisch mit dem Qubit 1.0 (Invitrogen, Darmstadt) bestimmt.

2.2.2.1.4 Restriktionsverdau von DNA

Damit das amplifizierte und aufgereinigte DNA-Fragment (Insert) in einen Vektor ligiert werden kann, muss dieser linearisiert vorliegen und über geeignete Enden verfügen. Wurde das Insert mit der Phusion DNA-Polymerase (Fisher Scientific, Schwerte) erzeugt, liegt dieses mit glatten Enden vor. Im Normalfall wurde daher nur der Vektor (pBSK) mit einem Restriktionsenzym geschnitten, das glatte Enden produziert (EcoRV). Wurden in der PCR Primer mit zusätzlich angefügten Erkennungssequenzen für Restriktionsenzyme verwendet, kann das PCR-Produkt einem doppelten Restriktionsverdau unterzogen werden und die Ligation gerichtet stattfinden. Die verwendeten Restriktionsenzyme wurden nach Herstellerangaben verwendet. Dabei wurde darauf geachtet, dass nicht mehr als fünffach überverdaut wurde. Das bedeutet 1 µg DNA wurde maximal mit 5 U eines NEB-Enzyms für eine Stunde bei der geeigneten Inkubationstemperatur mit dem empfohlenen Puffer inkubiert. Bei einem doppelten Restriktionsverdau muss zudem darauf geachtet werden, dass beide Enzyme in dem verwendeten Puffer arbeiten können. Der Restriktionsverdau wurde anschließend auf einem Agarosegel gegen die unverdaute Probe aufgetragen, analysiert und im geeigneten Fall aus dem Gel extrahiert (Kap. 2.2.2.1.2).

2.2.2.1.5 Ligation von Insert und Vektor

Für die Ligation des geschnittenen Vektors mit dem Insert, wurde die T4 DNA-Ligase (NEB, Frankfurt) nach Herstellerangaben verwendet. Für einen 20 µl Ligationsansatz wurden ~ 100 ng DNA, bestehend aus Vektor und Insert, mit 400 U der T4 DNA-Ligase und 1x Ligase-Puffer zusammengegeben. Dabei wurde ein molares Verhältnis von 1 zu 3 (Vektor: Insert) eingestellt. Anschließend wurde der Ligationsansatz ÜN bei 16 °C inkubiert.

2.2.2.1.6 Transformation in *Escherichia coli*

Damit *E. coli* Fremd-DNA aufnimmt, wurden die Zellen KCM-Kompetent gemacht. Dafür wurde 1 ml ÜN-Kultur in 100 ml frisches LB-Medium überimpft und bis zu einer $OD_{600} = 0.5$ wachsen gelassen. Danach wurden die Zellen zentrifugiert, das Pellet in 7.5 ml vorgekühltem TBS-Puffer resuspendiert und auf Eis für eine Stunde inkubiert. Anschließend wurden die Zellen aliquotiert (100 µl) und nach einem Gefrierschritt in flüssigem Stickstoff bei -80 °C gelagert.

Für die KCM-Transformation wurden die KCM-Kompetenten Zellen auf Eis aufgetaut und 100 µl Transformationslösung, bestehend aus 10 ng DNA bzw. dem Ligationsansatz (Kap. 2.2.2.1.5) in 1x KCM-Puffer, hinzugegeben. Die Bakterien mit der Transformationslösung wurden für 20 Minuten auf Eis und anschließend 10 Minuten bei Raumtemperatur inkubiert. Danach wurde 1 ml LB-Medium hinzugefügt und die Bakterien für eine Stunde bei 37 °C und 180 rpm wachsen gelassen. In dieser Zeit konnten die Bakterien ihre Antibiotikaresistenz ausbilden. Anschließend wurde die Lösung auf Agarplatten mit entsprechendem Antibiotikum ausplattiert (100 µl) und ÜN bei 37 °C wachsen gelassen.

2.2.2.1.7 Kolonie-PCR

Da eine Ligation mit glatten Enden nicht gerichtet verläuft und die Möglichkeit einer Religation des Vektors hoch ist, wurden die Transformanten mit einer Kolonie-PCR auf das Insert getestet. Dabei wurden die Bakterienkolonien direkt zum Reaktionsansatz gegeben und parallel auf einer nummerierten Agarplatte ausgestrichen. Für diese analytische PCR wurde die MyTaq DNA-Polymerase (Bioline, Luckenwalde) nach Herstellerangaben verwendet. Die vektorspezifischen-Primer TS3 und TS43 wurden für Inserts unter 1000 bp eingesetzt. Die Komponenten des PCR-Ansatzes (20 µl) setzen sich aus dem Templat (Kolonie), TS3 und TS43 (je 250 nM), der myTaq DNA-Polymerase (0,04 U) und 1x MyTaq-Puffer (mit 1 mM dNTPs) zusammen. Der PCR-Ansatz wurde im MyCycler (Bio-Rad, München) mit dem folgenden Thermoprofil inkubiert. Positive Transformanten konnten gelelektrophoretisch identifiziert und der Agarplatte zugeordnet werden (Kap. 2.2.2.1.2).

Thermoprofil der Kolonie-PCR

Bakterielle Lyse und anfängliche Denaturierung der DNA	95 °C	8 min
Exponentielle Amplifikation (25 Zyklen):		
Denaturierung der DNA	95 °C	15 sec
Annealing der Primer	50-60 °C	15 sec
Primer-Elongation mit der myTaq	72 °C	10 sec/kb
Finale DNA-Elongation	72 °C	5 min

2.2.2.1.8 Plasmidisolierung

Die in der Kolonie-PCR identifizierten positiven Transformanten wurden in frisches LB-Medium überimpft und ÜN bei 37 °C wachsen gelassen. Aus der ÜN-Kultur wurden die Plasmide mit dem „HiYield® Plasmid Mini Kit“ (Süd-Laborbedarf GmbH, Gauting) isoliert.

2.2.2.1.9 Sequenzierungen

Für die Sequenzierung von isolierten Plasmiden oder PCR-Produkten, wurden 1 µg DNA in 10 µl dH₂O gelöst und mit 2 µl einer 10 mM Primerlösung zu der Firma LGC Genomics GmbH (Berlin) geschickt. Standard-Sequenzierungen wurden mit den vektorspezifischen-Primern TS3 (T7-Promotorsequenz) und TS43 (T3-Promotorsequenz) durchgeführt.

2.2.2.2 Cyanobakterielles Screening auf die DGR-kodierte RT

Für die PCR des Screenings wurden ganze Zellen der Cyanobakterien als Templat eingesetzt. Diese wurden von Prof. Büdel (Pflanzenökologie & Systematik, TU Kaiserslautern) zur Verfügung gestellt und befanden sich zunächst in den entsprechenden Wachstumsmedien. Daher wurden die Cyanobakterien zentrifugiert (8000 rpm, 5 min) und in einer geeigneten Menge dH₂O resuspendiert. Für die Amplifikation der RT wurde ein PCR-Ansatz (20 µl) bestehend aus der Cyanobakteriensuspension (3 µl), den Primern NZ112 und NZ114 (je 5 µM), der MyTaq DNA-Polymerase (0,04 U) (Bioline, Luckenwalde) und 1x MyTaq-Puffer (mit 1 mM dNTPs) zusammengestellt. Kontrollen auf DNA-Kontamination (Wasserkontrolle) und einen erfolgreichen Aufschluss der Zellen (16S rDNA-Kontrolle) wurden parallel durchgeführt. Die PCR auf die 16S rDNA weicht von der PCR auf die RT durch die Verwendung anderer Primer (je 500 nM von PM24 und PM57), sowie einer höheren Annealingtemperatur (55 °C) ab. Die Reaktionslösungen wurden im MyCycler (Bio-Rad, München) mit Gradientenfunktion nach dem folgenden Thermoprofil inkubiert und die PCR-Produkte auf einem Agarosegel analysiert (Kap. 2.2.2.1.2), sowie teilweise sequenziert (Kap. 2.2.2.1.9).

Thermoprofil des cyanobakteriellen Screenings für die Amplifikation der RT

Bakterielle Lyse und anfängliche Denaturierung der DNA	95 °C	5 min
Exponentielle Amplifikation (25 Zyklen):		
Denaturierung der DNA	95 °C	15 sec
Annealing der Primer	42 °C	15 sec
Primer-Elongation mit der myTaq	72 °C	10 sec
Finale DNA-Elongation	72 °C	5 min

2.2.2.3 Inverse PCR

Die Sequenz des kompletten DGR-Elements aus dem Cyanobakterium *Anabaena flos-aquae* wurde mittels inverser PCR identifiziert (Ochman, Gerber & Hartl 1988). Dafür wurde zunächst genomische DNA (gDNA) isoliert, mit einem Restriktionsenzym geschnitten und intramolekular ligiert. Die entstehende zirkuläre DNA wurde über Touchdown-PCR amplifiziert.

2.2.2.3.1 Phenol-Chloroform-Extraktion von Nukleinsäuren

Für die Phenol-Chloroform-Extraktion wurden 50 ml einer cyanobakteriellen Kultur ($OD_{750} \sim 0,5$) für 10 Minuten zentrifugiert (9.500 rpm, 4 °C), das Zellpellet in 10 ml vorgekühltem PBS-Puffer resuspendiert und wieder für 10 Minuten gewaschen (9.500 rpm, 4 °C). Das gewaschene Zellpellet wurde in 1 ml kaltem PBS-Puffer gelöst, 2 ml Lysepuffer und 0,4 g Glasperlen (0,25 mm Durchmesser) hinzugegeben und der Aufschluss mittels Vortex-Schüttler (IKA-Werke, Staufen) begünstigt. Nach dem Zellaufschluss erfolgte die Zugabe des gleichen Volumens Phenol-Chloroform-Isoamylalkohol (25:24:1) pH 8,0. Durch nachfolgendes vortexen entstehen eine wässrige und eine organische Phenol-Phase. In der wässrigen Phase sind die polaren Nukleinsäuren gelöst, wobei sich die Proteine zwischen den Phasen (Interphase) ansammeln. Die wässrige Phase wurde abgenommen und die Phenol-Chloroform-Extraktion sofort wiederholt, bis keine Interphase mehr zu erkennen war. Für eine reine RNA-Extraktion wurde saures Phenol verwendet (Kap. 2.2.3.1.1).

An die Extraktion wurde im Anschluss eine Ethanol-Präzipitation angefügt. Dazu wurde die Nukleinsäurelösung mit 0.1 Vol. 3 M Natriumacetatlösung (pH 5,5) und ~ 3 Vol. 100 % Ethanol versetzt, invertiert und bei -80 °C für 1 Stunde gefällt. Anschließend wurden die Nukleinsäuren pelletiert (30 Minuten, 12.300 rpm, 4 °C), mit 500 µl 70 % Ethanol gewaschen (20 Minuten, 12.300 rpm, 4 °C) und an der Luft getrocknet. Das Pellet wurde in dH₂O bzw. DEPC-H₂O resuspendiert und die Konzentration fluorometrisch mit dem Qubit 1.0 (Invitrogen, Darmstadt) bestimmt. Die Lagerung der Nukleinsäuren erfolgte bei -80 °C.

2.2.2.3.2 Restriktionsverdau der Inversen PCR

Für die inverse PCR wurde 1 µg der gDNA aus *Anabaena flos-aquae* mit 50 U des EcoRI-HF Restriktionsenzym (NEB) in 100 µl 1x NEBuffer4 bei 37 °C für eine Stunde inkubiert. Anschließend wurde das Enzym Hitzeinaktiviert (65 °C, 20 Minuten).

2.2.2.3.3 Intramolekulare Ligation

Die linearisierte DNA des Restriktionsverdaus (Kap. 2.2.2.3.2), wurde mit 2000 U der T4 DNA-Ligase (NEB, Frankfurt) in 2 ml 1x Ligase-Puffer ligiert. Dabei wird durch die starke Verdünnung eine intramolekulare Ligation favorisiert. Der Ligationsansatz wurde ÜN bei 16 °C inkubiert, die DNA mittels Phenol-Chloroform-Extraktion aufgereinigt und in 160 µl dH₂O resuspendiert (Kap. 2.2.2.3.1).

2.2.2.3.4 Touchdown-PCR

In den vorhergehenden Schritten wurde zirkuläre DNA aus gDNA von *Anabaena flos-aquae* hergestellt. Darunter befindet sich auch der DNA-Abschnitt der sequenzierten RT (Kap. 2.2.2.2) mit unbekanntem, flankierendem Bereich. Dieser Bereich wurde durch das verwendete Restriktionsenzym (EcoRI) vorgegeben. Um ihn amplifizieren zu können, wurden außwärts orientierte Primer für die bekannte RT-Sequenz entworfen. Der Reaktionsansatz der PCR (20 µl) besteht aus den Primern PM21 und PM22 (je 250 nM), zirkulärer DNA (~ 30 ng), DMSO (3 % finale Konzentration), dNTPs (je 250 µM), der Phusion DNA-Polymerase (0,5 U) (Fisher Scientific, Schwerte) und 1x Phusion-Puffer (GC). Der Reaktionsansatz wurde im MyCycler (Bio-Rad, München) mit dem Thermoprofil der Touchdown-PCR inkubiert. Diese verwendet eine hohe anfängliche Annealingtemperatur, die in den folgenden Amplifikationszyklen abnimmt. Dadurch wird die Amplifikation unspezifischer Sequenzen verringert. Das erhaltene PCR-Produkt wurde anschließend kloniert und sequenziert (Kap. 2.2.2.1).

Thermoprofil der Touchdown-PCR

Anfängliche Denaturierung der DNA	98 °C	5 min
Exponentielle Amplifikation (30 Zyklen):		
Die Annealingtemperatur der ersten 20 Zyklen verringert sich um 0.5 °C/Zyklus		
Denaturierung der DNA	98 °C	30 sec
Annealing der Primer	60 °C (bis 50 °C)	25 sec
Primer-Elongation mit der Phusion	72 °C	3 min
Finale DNA-Elongation	72 °C	10 min

2.2.2.4 Restriktionsenzym-basierter Variabilitäts-Assay

Der Restriktionsenzym-basierte Variabilitäts-Assay (REBVA) wurde zur Bestimmung der Aktivität von diversitätsgenerierenden Retroelementen (DGRs) verwendet (Liu et al. 2002). Für den REBVA wird ein potentiell mutierter Sequenzbereich alternierend restriktionsverdaut (Kap. 2.2.2.1.4) und amplifiziert (Kap. 2.2.2.1.1). Die Amplifikation ist nur bei einer mutierten Erkennungssequenz möglich, die das Restriktionsenzym nicht schneiden konnte. Durch DGR-Aktivität können nur bestimmte variable Positionen in einer variablen Region (VR) mutiert werden, die mit Adeninen in einem Templat Repeat (TR) korrelieren. Die verwendeten Restriktionsenzyme, deren Erkennungssequenzen und die analysierten variablen Positionen, sind für die untersuchten DGRs in Tabelle 7 zusammengefasst.

Tabelle 7: Verwendete Restriktionsenzyme innerhalb des REBVA

DGR	Restriktionsenzym	Erkennungssequenz	Variable Positionen
<i>Nostoc sp.</i> PCC7120	SacI	5'...GAGCT↓C...3' 3'...C↑TCGAG...5'	VR 5'...GAGCTC...3' TR 5'...GAACAA...3'
<i>Anabaena flos-aquae</i>	AclI	5'...AA↓CGTT...3' 3'...TTGC↑AA...5'	VR 5'...AACGTT...3' TR 5'...AACAAAT...3'
<i>Treponema denticola</i>	PvuII	5'...CAG↓CTG...3' 3'...GTC↑GAC...5'	VR 5'...CAGCTG...3' TR 5'...CAGCTG...3'

Für die Amplifikation wurden zwei ineinander verschachtelte Primerpaare verwendet (nested PCR). Für *Nostoc sp.* PCC7120 wurden die Primerpaare PM73/TS69 und PM125/PM82, für *Anabaena flos-aquae* die Primerpaare PM58/PM59 und PM42/PM49 und für *Treponema denticola* die Primerpaare PM5/PM25 und PM55/PM56 verwendet.

Für jeden REBVA wurden schätzungsweise 10^7 Zellen eingesetzt. Die zugrundeliegende Berechnung ist am Beispiel des Cyanobakteriums *Nostoc sp.* PCC7120 dargestellt:

$$1 \times 10^{-7} \text{ g} / (7200000 \text{ bp} \times 650 \text{ Da}) = 0,021 \text{ fmol}$$

$$0,021 \text{ fmol} \times 6 \times 10^{23} \text{ (Avogadro-Konstante)} = 1,3 \times 10^7 \text{ Moleküle}$$

Dabei beziehen sich die 10^{-7} g auf die eingesetzte Menge genomischer DNA, die 7200000 bp beziehen sich auf die Genomgröße von *Nostoc sp.* PCC7120 (Kaneko et al. 2001) und die 650 Da beziehen sich auf das durchschnittliche Molekulargewicht eines DNA-Basenpaares. Für *Anabaena flos-aquae* wurde mit der Genomgröße des verwandten Cyanobakteriums *Anabaena variabilis* (7,1 Mbp) (Thiel et al. 2014) und für *Escherichia coli* K-12 (heterologe Expression) wurde mit einer Genomgröße von 4,6 Mbp gerechnet (Blattner et al. 1997).

2.2.3 Arbeiten mit RNA

Um eine Degradation der RNA durch RNasen zu vermeiden, wurden Handschuhe getragen und auf Eis gearbeitet. Zudem wurden hitzestabile Materialien 6 Stunden bei 220 °C gebacken und hitzelabile Materialien für 3 Stunden in 3 % (v/v) H₂O₂ inkubiert. Lösungen wurden ÜN mit 0,1 % (v/v) Diethylpyrocarbonat (DEPC) bei 37 °C behandelt. Um DEPC zu inaktivieren, wurde die Lösung im Anschluss autoklaviert.

2.2.3.1 Northern-Blot-Experimente

Diese Methode dient zum Nachweis von spezifischen Transkripten in einer RNA-Probe. Zunächst wurde Gesamt-RNA isoliert, gelelektrophoretisch aufgetrennt und auf eine Nylonmembran transferiert. Im Anschluss wurde die RNA-Spezies von Interesse über Hybridisierung mit einer komplementären Gensonde spezifisch detektiert.

2.2.3.1.1 RNA-Isolation

Die RNA-Isolation wurde nach der bereits beschriebenen Methode der Phenol-Chloroform-Extraktion durchgeführt (Kap. 2.2.2.3.1). Mit dieser Methode kann sowohl DNA als auch RNA isoliert werden. Für eine reine RNA-Extraktion wurde Phenol-Chloroform-Isoamylalkohol (25:24:1) pH 5.0 verwendet. DNA wird bei niedrigen pH-Werten selektiv in der organischen Phenol-Phase und Interphase zurückgehalten, während RNA in der wässrigen Phase vorzufinden ist. 20 µg der in DEPC-H₂O gelösten RNA wurden anschließend mit dem „High Pure RNA Isolation Kit for total RNA“ (Roche, Mannheim) weiter aufgereinigt. In dem Kit ist ein Restriktionsverdau mit DNase I inbegriffen, um isolierte DNA zu entfernen. Die RNA wurde anschließend in 50 µl DEPC-H₂O eluiert.

2.2.3.1.2 Denaturierende Formaldehyd-Agarose-Gelelektrophorese von RNA

Formaldehyd-Agarosegele wurden für die Überprüfung der Integrität von RNA-Präparationen und für Northern-Blot-Experimente verwendet. Formaldehyd wird zur Denaturierung der RNA eingesetzt und verhindert die Ausbildung von Sekundärstrukturen. Für die Herstellung von 100 ml Gellösung werden 1-2 g Agarose in 94 ml 1x MOPS-Puffer aufgekocht und anschließend mit 6 ml 37 % Formaldehyd versetzt. Nach Erhärtung des Gels, können die RNA-Proben aufgetragen werden. Diese wurden zuvor mit 3/4 des Volumens an RNA-Ladepuffer versehen, für 5 Minuten bei 65 °C denaturiert und anschließend für 5 Minuten bei 4 °C abgekühlt. Die Elektrophorese wurde in 1x MOPS-Laufpuffer bei einer konstanten Spannung von 70 V durchgeführt.

2.2.3.1.3 Transfer von RNA auf eine Nylonmembran

Die Gesamt-RNA wurde auf einem denaturierenden Formaldehyd-Agarosegel aufgetrennt. Um die aufgetrennte RNA auf eine Nylonmembran (Hybond N+, GE Healthcare Life Science GmbH, Freiburg) zu transferieren, wurde das Gel zunächst zweimal für 10 Minuten in DEPC-H₂O gewaschen und anschließend für 10 Minuten in 10x SSC-Puffer äquilibriert. Der Transfer der RNA erfolgte ÜN durch Kapillarkräfte. Dafür wurde ein ca. 10 cm hoher Stapel Saugpapier mit zwei Lagen 3MM Whatman-Papier überschichtet und zunächst die Membran und dann das Gel von oben aufgelegt. Auf das Gel folgten wieder zwei Lagen und eine Brücke aus 3MM Whatman-Papier. Die Brücke wurde in Behälter mit 10x SSC-Puffer gelegt und zur Verstärkung der kapillaren Sogwirkung mit einem Gewicht von ca. 1 kg beschwert. Nach dem Transfer wurde die RNA durch UV-Licht bei 120 mJ/cm² (UV-Crosslinker, Amersham Biosciences) kovalent an die Membran gebunden.

2.2.3.1.4 Radioaktive Markierung von DNA-Sonden

Die geblottete RNA wurde mit DNA-Sonden hybridisiert, die entweder am 5'-Ende oder innerhalb der Sequenz radioaktiv markiert waren. Um die Sequenz eines PCR-Produktes zu markieren, wurde das „NEBlot Kit“ (NEB, Frankfurt) mit [α -³²P]-dCTP verwendet. Für die Markierung des 5'-Endes von Oligonukleotiden wurde das Oligonukleotid (0,67 μ M) zusammen mit [γ -³²P]-ATP (0,15 mCi) und der T4 Polynukleotidkinase (15 U) (NEB, Frankfurt) in 1x Kinase-Puffer für eine Stunde bei 37 °C inkubiert. Nach der Herstellung der Sonden, wurden diese durch das „Nukleotide Removal Kit“ (Qiagen, Hilden) aufgereinigt und die Radioaktivität mittels Szintillationszähler (Packard BioScience Company) gemessen.

2.2.3.1.5 Hybridisierung von geblotteter RNA

Für die Hybridisierung von geblotteter RNA mit einer radioaktiv markierten Sonde wurden Puffer der ULTRAhyb-Reihe (Invitrogen, Darmstadt) verwendet. Markierte Oligonukleotide wurden mit dem ULTRAhyb-Puffer (Oligo) und markierte PCR-Produkte mit dem ULTRAhyb-Puffer (Standard) behandelt. In beiden Fällen wurde die Membran für eine Stunde prähybridisiert um unspezifische Bindestellen zu besetzen und anschließend ÜN hybridisiert. Dabei wurde eine Sondenkonzentration von mindestens 1x10⁶ cpm/ml Hybridisierungspuffer eingestellt. Die Prähybridisierungs- und Hybridisierungstemperatur für 5'-markierte Sonden lag bei 43 °C und die für sequenzmarkierte Sonden bei 50 °C. Anschließend wurde die Membran nach Herstellerangaben des Hybridisierungspuffers gewaschen und in Klarsichtfolie fixiert. Für die Detektion wurde eine Phosphorimagerplatte über einen Zeitraum von 10 Minuten bis zu 16 Stunden mit der Membran exponiert und anschließend im Phosphorimager (Packard BioScience Company) ausgelesen.

2.2.3.2 RT-PCR-Analysen

Die RT-PCR setzt sich aus zwei Schritten zusammen. Im ersten Schritt wird die als Ausgangsmaterial dienende RNA in cDNA umgeschrieben und im zweiten Schritt als Templat in eine PCR eingesetzt. Die verwendeten Primer für die jeweilige Applikation sind in Tabelle A4 zusammengefasst. Der Reaktionsansatz (20 µl) für die cDNA-Synthese setzt sich aus einem Primer (250 nM), der RNA (500 ng), einem RNase Inhibitor (20 U), dNTPs (je 250 µM), der AMV-RT (10 U) und 1x RT-Puffer zusammen. Sowohl der RNase Inhibitor, als auch die AMV-RT wurden von NEB (Frankfurt) bezogen. Dabei wurden zunächst Primer und RNA preannealed (5 min bei 95 °C, gefolgt von 5 min bei 4 C) und die anderen Komponenten anschließend hinzugegeben. Die Reaktion wurde für 30 Minuten bei 55 °C durchgeführt und das Enzym anschließend Hitzeinaktiviert (5 min bei 90 °C). Im Anschluss wurde die zur cDNA komplementäre RNA durch die RNase H (NEB, Frankfurt) abgebaut. Dafür wurden die Proben mit 2 U der RNase H für 20 Minuten bei 37 °C inkubiert und das Enzym anschließend Hitzeinaktiviert (20 min bei 65 °C).

Die cDNA wurde im Anschluss als Templat für die PCR eingesetzt. Je nach Applikation wurden unterschiedliche PCR-Bedingungen gewählt. Für die Transkriptanalyse und die Primerbasierte Verifizierung des Transkriptionsstarts, wurde die myTaq DNA-Polymerase (Bioline, Luckenwalde) nach Herstellerangaben verwendet. Der PCR-Ansatz (20 µl) setzt sich aus der cDNA (~ 50 ng), den Primern (je 250 nM), der myTaq DNA-Polymerase (0,04 U) und 1x myTaq-Puffer (mit 1 mM dNTPs) zusammen. Inkubiert wurde der Ansatz im MyCycler (Bio-Rad, München) mit dem folgenden Thermoprofil.

Thermoprofil der PCR auf cDNA (Transkriptanalyse und Transkriptionsstart)

Denaturierung der cDNA	95 °C	5 min
Exponentielle Amplifikation (25 Zyklen):		
Denaturierung der DNA	95 °C	15 sec
Annealing der Primer	50 °C	15 sec
Primer-Elongation mit der myTaq	72 °C	10 sec/kb
Finale DNA-Elongation	72 °C	5 min

Die RT-PCR ist auch Bestandteil der RACE-Experimente (Kap. 2.2.3.3). Diese wurden mit der Phusion DNA-Polymerase (Fisher Scientific, Schwerte) nach Herstellerangaben durchgeführt. Für die 5'-RACE wurde der Primer PM137 (cDNA-Synthese) und das Primerpaar PM141/PM142 (PCR) und für die 3'-RACE der Primer PM103 (cDNA-Synthese) und das Primerpaar PM121/PM104 (PCR) verwendet. Der PCR-Ansatz (20 µl) besteht aus der cDNA (~ 100 ng), den Primern (je 250 nM), dNTPs (je 250 µM), der Phusion DNA-Polymerase (0,5 U)

und 1x Phusion-Puffer (GC). Der Reaktionsansatz wurde im MyCycler (Bio-Rad, München) mit dem folgenden Thermoprofil inkubiert.

<u>Thermoprofil der PCR auf cDNA (RACE)</u>		
Denaturierung der cDNA	98 °C	1 min
Exponentielle Amplifikation (30 Zyklen):		
Denaturierung der DNA	98 °C	10 sec
Annealing der Primer (5'-RACE/3'-RACE)	55/60 °C	30 sec
Primer-Elongation mit der Phusion (5'-RACE/3'-RACE)	72 °C	10/40 sec
Finale DNA-Elongation	72 °C	5 min

Bei einer RNA-Extraktion können auch geringe Mengen DNA isoliert werden. Daher muss sichergestellt werden, dass ein PCR-Produkt von der cDNA und nicht der DNA stammt. Dazu wurde eine Kontrollreaktion bei der cDNA-Synthese durchgeführt, in der das Enzym mit DEPC-H₂O ausgetauscht wurde (noRT-Kontrolle). Ein PCR-Produkt in der noRT-Kontrolle deutet auf eine DNA-Kontamination in der RNA-Präparation hin. In diesem Fall wurde ein weiterer Restriktionsverdau mit der DNaseI durchgeführt (Kap. 2.2.3.1.1).

2.2.3.3 RACE-Experimente

Die „rapid amplification of cDNA-ends“ (RACE) ist eine Methode zur Bestimmung von Transkriptenden. Dafür wurde Gesamt-RNA isoliert (Kap. 2.2.3.1.1), mit der T4 RNA-Ligase ein RNA-Adapter an das 5'-Ende der RNA-Moleküle ligiert und ein künstlicher Poly(A)-Schwanz über eine Poly(A)-Polymerase angefügt. Um prozessierte RNA vor der Modifizierung zu entfernen, wurde die RNA mit der Terminator 5'-Phosphate-Dependent Exonuclease (TEX) behandelt. Die TEX ist eine prozessive Exonuklease, die RNA mit einem 5'-Monophosphat als Substrat verwendet. RNA mit einem 5'-Triphosphat (Primärtranskript) stellt kein Substrat für die TEX dar. Anschließend wurden die Primärtranskripte mit der Tobacco Acid Pyrophosphatase (TAP) behandelt. Die TAP hydrolysiert die terminale Pyrophosphatbindung, wodurch bakterielle Primärtranskripte nur noch über ein 5'-Monophosphat verfügen. Das 5'-Monophosphat wurde benötigt, um den RNA-Adapter an das 5'-Ende ligieren zu können. Die modifizierten Sequenzen wurden in einer cDNA-Synthese für die Generierung von Primerbindestellen benutzt, die in der darauf folgenden PCR (RT-PCR) benötigt wurden (Kap. 2.2.3.2). Die Sequenzierung des PCR-Produkts erlaubt anschließend eine Analyse der Transkriptenden.

TEX-Reaktion auf Gesamt-RNA (20 µl)

Gesamt-RNA	10 µg
RNase Inhibitor (Epicentre, USA)	20 U
TEX (Epicentre, USA)	1 U
Reaktionspuffer A	1x konzentriert

Inkubation für 1 Stunde bei 30 °C

Stoppen der Reaktion durch Phenol-Chloroform-Extraktion (Kap. 2.2.2.3.1)

TAP-Reaktion (10 µl)

TEX-behandelte RNA	~ 10 µg
RNase Inhibitor (Epicentre, USA)	20 U
TAP (Epicentre, USA)	5 U
TAP-Puffer	1x konzentriert

Inkubation für 30 Minuten bei 37 °C

RNA-Adapter Ligation (20 µl)

TAP-behandelte RNA	~ 10 µg
RNase Inhibitor (Epicentre, USA)	20 U
RNA-Adapter (PM140)	5 µM
ATP	1 mM
T4 RNA-Ligase (NEB, Frankfurt)	400 U
Ligase-Puffer	1x konzentriert
NEBuffer 3 (TAP-Stopp über pH-Shift)	1x konzentriert

Inkubation für 30 Minuten bei 37 °C

Verwendung für RT-PCR (5'-RACE)

Polyadenylierung von Gesamt-RNA (20 µl)

TEX-behandelte RNA	~ 10 µg
RNase Inhibitor (Epicentre, USA)	20 U
ATP	1 mM
Poly(A)-Polymerase (Epicentre, USA)	4 U
Reaktionspuffer	1x konzentriert

Inkubation für 30 Minuten bei 37 °C

Stoppen der Reaktion durch Phenol-Chloroform-Extraktion (Kap. 2.2.2.3.1)

Verwendung für RT-PCR (3'-RACE)

2.2.3.4 SHAPE-Experimente

„Selective 2'-hydroxyl acylation analyzed by primer extension“ (SHAPE)-Experimente erlauben die Strukturanalyse von RNA in Einzel-Nukleotid-Auflösung (Merino et al. 2005). In dem Versuch wird die RNA mit einem Elektrophil behandelt, welches selektiv mit der 2'-Hydroxylgruppe der Ribose reagiert, wenn die Nukleotide in ihrer Konformation flexibel sind. Es bildet sich ein 2'-O-Addukt, welches über Primer-Extension mit radioaktiv markierten Primern nachgewiesen werden kann. Für die Primer-Extension wird eine RT verwendet, die bei Erreichen der modifizierten Position vom Templat abdissoziiert. Die radioaktiv markierte cDNA kann auf einem denaturierenden Polyacrylamidgel aufgetrennt und analysiert werden. Die gemessene Bandenintensität korreliert mit der Reaktivität des modifizierten Nukleotids auf das Elektrophil und stellt ein Maß für die Einbindung des Nukleotids in Basenpaarungen oder andere Interaktionen dar. SHAPE wurde für die Aufklärung der Sekundärstruktur der *in vitro* transkribierten Templat RNA aus *Nostoc sp.* PCC7120 mit dem Elektrophil NMIA (N-methylisotoic anhydride) verwendet.

2.2.3.4.1 *In vitro* Transkription

Unter der *in vitro* Transkription (IVT) wird die zellfreie Synthese von RNA über ein DNA-Templat mit Hilfe einer RNA-Polymerase verstanden. Bei dem verwendeten DNA-Templat handelt es sich um den pKH2 (Abbildung A1). Der pKH2 beinhaltet die TR-Sequenz aus *Nostoc sp.* PCC7120 unter Kontrolle des T7-Promotors. Downstream der TR-Sequenz wurde das Plasmid mit dem Restriktionsenzym HindIII geschnitten (Kap. 2.2.2.1.4), um eine „run off transcription“ mit der T7-Polymerase zu ermöglichen. Die „run off transcription“ limitiert die Transkription auf die Sequenz von Interesse, da die Polymerase am Ende des geschnittenen DNA-Strangs abdissoziiert.

Für die IVT wurde ein Reaktionsansatz (500 µl) bestehend aus dem linearisierten pKH2 (10 µg), einer in der Arbeitsgruppe aufgereinigten T7-RNA-Polymerase (20 µl), 1x Transkriptionspuffer, DTT (10 mM), RNase Inhibitor (500 U) (NEB, Frankfurt) und NTPs hergestellt. Die Konzentration aller NTPs wurde auf 16 mM eingestellt, wobei die Konzentration der einzelnen NTPs entsprechend ihrem Anteil im Transkript berechnet wurde. Der Reaktionsansatz wurde zwei Stunden bei 37 °C inkubiert. Dabei bildet sich ein Magnesium-Pyrophosphat-Komplex, der nach der Inkubationszeit abzentrifugiert wurde. Im Überstand ist die synthetisierte RNA. Diese wurde präzipitiert (Kap. 2.2.2.3.1) und anschließend über eine denaturierende Polyacrylamidgelelektrophorese aufgereinigt (Kap. 2.2.3.4.2).

2.2.3.4.2 Denaturierende Polyacrylamidgelelektrophorese

Die Polyacrylamidgelelektrophorese (PAGE) eignet sich zur Auftrennung von Nukleinsäuren bis zu 1 kb. Polyacrylamidgele entstehen durch radikalische Polymerisation der Stoffe Acrylamid und Bisacrylamid. Dabei ist die für die Auftrennung entscheidende Porengröße abhängig von der Totalacrylamidkonzentration. Bei denaturierenden Polyacrylamidgelen wird zusätzlich Harnstoff hinzugegeben, was die Ausbildung von Sekundärstrukturen innerhalb der aufzutrennenden Nukleinsäuren reduziert.

Für die Herstellung einer denaturierenden Polyacrylamidgellösung werden Acrylamid und Bisacrylamid (29:1) mit Harnstoff (8 M) in 1x TBE-Puffer gelöst und steril filtriert (Porengröße 0,2 µm). Anschließend wird die Gellösung mit dem Radikalstarter Ammoniumperoxodisulfat (0,05 %) und dem Polymerisierungskatalysator Tetramethylethyldiamin (0,05 %) versetzt und das Gel gegossen. Nach ca. 45 Minuten ist das Gel auspolymerisiert und wird in 1x TBE-Puffer für eine Stunde bei ~ 1000 V vorlaufen gelassen, um die Salzfront zu entfernen. Anschließend wird das Gel mit Proben beladen, die mit PAGE-Ladepuffer versetzt und zuvor für 3 Minuten bei 95 °C denaturiert wurden.

Die Aufreinigung von der IVT (Kap. 2.2.3.4.1) und den Primern der SHAPE-Experimente (Kap. 2.2.3.4.4) wurde mit denaturierenden Polyacrylamidgelen durchgeführt, die eine Polyacrylamidkonzentration von 5 % hatten. Anschließend wurde das Polyacrylamidgel mit UV-Licht bestrahlt und der „Schatten“ der Nukleinsäuren ausgeschnitten. Der „Schatten“ entsteht, da Nukleinsäuren UV-Licht absorbieren, die Umgebung im Gel jedoch nicht. Die ausgeschnittene Bande wurde zerkleinert, mit 3 ml Elutionspuffer gemischt, in flüssigem Stickstoff gefroren und für vier Stunden bei 4 °C inkubiert. Anschließend wurden die Gelreste durch eine Filtration (Porengröße 0,2 µm) abgetrennt und die Nukleinsäuren präzipitiert (Kap. 2.2.2.3.1).

Denaturierende Polyacrylamidgele wurden zudem zur Auftrennung der radioaktiv markierten cDNA in den SHAPE-Experimenten verwendet. Dabei wurde eine Polyacrylamidkonzentration von 12 % bzw. 15 % eingestellt. Nach der Gelelektrophorese wurde das Polyacrylamidgel im Vakuum bei 80 °C für ca. 1 Stunde getrocknet und eine Phosphorimagerplatte ÜN aufgelegt. Die Phosphorimagerplatte wurde anschließend im Phosphorimager (Packard BioScience Company) ausgelesen.

2.2.3.4.3 Modifizierung durch NMIA

Für die *in vitro* transkribierte RNA (Kap. 2.2.3.4.1) wurden Bedingungen eingestellt, welche die Ausbildung der Tertiärstruktur erlauben. Anschließend wurden die flexiblen Nukleotide der gefalteten RNA mittels NMIA (N-methylisotoic anhydride) modifiziert.

Für die Faltung wurde angenommen, dass die stabilste Struktur gleichzeitig die native ist. Die RNA-Moleküle (2 pmol) wurden zunächst denaturiert (5 min bei 65 °C) und danach abgekühlt (5 min bei 4 °C). Anschließend wurden die Faltungsbedingungen durch Zugabe von 1x Faltungsmix eingestellt und die RNA für 20 Minuten bei 25 °C inkubiert. Zu der gefalteten RNA wurde in DMSO gelöstes NMIA (8 mM) gegeben und der Ansatz für fünf Hydrolyse-Halbwertzeiten inkubiert. Die Halbwertzeit ist Temperaturabhängig und lässt sich über folgende Formel berechnen (Wilkinson, Merino & Weeks 2006):

$$\text{Halbwertzeit (Minuten)} = 360 * \exp [-0,102 * \text{Temperatur (}^\circ\text{C)}]$$

Das entspricht einer Inkubationszeit von 140 Minuten bei einer Modifizierungstemperatur von 25 °C. Neben der NMIA-Reaktion wurde parallel eine Kontrolle ohne NMIA durchgeführt. Anschließend wurde die RNA präzipitiert (Kap. 2.2.2.3.1) und in 10 µl 0.5x TE-Puffer resuspendiert.

2.2.3.4.4 *Primer-Extension*

Die Primer-Extension der modifizierten RNA (Kap. 2.2.3.4.3) wurde mit der Superscript III-RT (Invitrogen, Darmstadt) und vier Primern (PM135-137, PM151) durchgeführt, welche ~ 80 nt voneinander entfernt auf der RNA binden. Die Primer wurden zuvor aufgereinigt (Kap. 2.2.3.4.2) und radioaktiv markiert (Kap. 2.2.3.1.4).

Zunächst wurde die modifizierte RNA (~ 2 pmol) mit einem Primer (~ 3x10⁶ cpm) annealed. Dafür wurden beide denaturiert (5 min bei 65 °C) und über einen Zwischenschritt (5 min bei 35 °C) auf 4 °C abgekühlt. Anschließend wurden DTT (5 mM) und dNTPs (je 500 µM) hinzugefügt und die Pufferbedingungen mit dem Superscript III-Puffer (FS) eingestellt. Der Ansatz wurde für 2 Minuten bei 52 °C inkubiert, die Superscript III-RT (100 U) hinzugegeben und der fertige Reaktionsansatz (20 µl) für weitere 30 Minuten bei 52 °C inkubiert. Parallel wurden Sequenzierungsreaktionen nach der Standardmethode mit Dideoxynukleotiden angesetzt (Sanger, Nicklen & Coulson 1977). Dabei wurde das Abdissoziieren der Polymerase von der RNA (5 pmol) durch Zugabe eines ddNTPs (250 µM) forciert. Die Sequenzierungsreaktionen waren für eine Zuordnung der NMIA-Banden zu der RNA-Sequenz erforderlich.

Nach der Primer-Extension wurde die cDNA mit NaOH (200 mM) behandelt und der Ansatz für 5 Minuten auf 95 °C erhitzt. Dabei wurde die zur cDNA komplementäre RNA abgebaut. Der pH-Wert wurde durch Zugabe von 1,4 Vol. Stopp-Mix (4:25 (v/v) Mischung aus 1 M Tris-HCl, pH 8.0 und saurer Stopp-Lösung) neutralisiert. Der Stopp-Mix diente zudem als Ladepuffer für die Polyacrylamidgelelektrophorese (Kap. 2.2.3.4.2).

2.2.4 Arbeiten mit Computerprogrammen

2.2.4.1 Quantifizierung der Bandenintensitäten in RT-PCR-Analysen

Die Gelbilder der RT-PCR-Analysen wurden in einem Bildokumentationssystem (Intas, Göttingen) mit einer Belichtung aufgenommen, die eine Quantifizierung der Bandenintensitäten erlaubt. Die Quantifizierung wurde mit dem Programm ImageJ Version 1.47 als Dreifachbestimmung durchgeführt (Schneider, Rasband & Eliceiri 2012). Dabei wurden die Werte der RT-Banden ins Verhältnis zur zugehörigen Positivkontrolle gesetzt, um unterschiedliche Bindungsaffinitäten der Primer berücksichtigen zu können.

2.2.4.2 Verarbeitung der SHAPE-Daten

Für die Auswertung eines Versuchs wurden mehrere Phosphorimagerplatten verwendet. Dabei wurden die Bandenintensitäten mit dem Programm ImageQuant TL Version 7.0 (GE Healthcare Life Science GmbH, Freiburg) quantifiziert, was eine Subtraktion des Hintergrundsignals mit einschließt. Um beurteilen zu können inwiefern die Werte unterschiedlicher Phosphorimagerplatten miteinander korrelieren, wurde eine Spearman-Rangsummen-Korrelation für überlappende Werte durchgeführt. Zudem wurden die Werte gegeneinander geplottet und visuell analysiert (R-Development-Core-Team 2008). Dabei konnte in allen Fällen ein hochsignifikanter linearer Zusammenhang festgestellt werden ($R^2 > 0,97$, $p < 1,6^{-6}$, $n: 8-27$), was einen Vergleich der Werte ermöglicht.

Die Banden der NMIA-Spuren wurden über die Sequenzierungsreaktionen der RNA-Sequenz zugeordnet. Dabei wurde berücksichtigt, dass die cDNA der Sequenzierung um ein Nukleotid länger ist, als die korrespondierende NMIA-cDNA. Die Werte der NMIA(-)-Banden wurden von denen der zugehörigen NMIA(+)-Banden abgezogen und hochreaktive Ausreißer entfernt. Diese wurden über Boxplot-Analysen identifiziert, wenn die Reaktivität höher war als das 1,5-Fache des Interquartilsabstands. Die maximale Anzahl der Ausreißer wurde für die ausgelesenen Phosphorimagerplatten (30-40 Werte) auf 5 % begrenzt. Anschließend wurden die Werte (SHAPE-Reaktivitäten) auf einen normalisierten Maßstab gebracht, indem diese durch den Durchschnitt der obersten 10 % der reaktiven Nukleotide geteilt wurden. Durch diese Berechnung wurden alle Reaktivitäten auf einen Bereich von 0 bis ~2 gebracht (durchführung in Deigan et al. 2009). Die vorher entfernten Ausreißer wurden anschließend ebenfalls normalisiert und alle Werte, auch die nicht analysierbaren Nukleotide (Wert: -999), in RNAstructure integriert.

2.2.4.3 Bioinformatische Sekundärstrukturanalysen

Die bioinformatische Sekundärstrukturanalyse der Templat RNA aus *Nostoc sp.* PCC7120 wurde mit den Programmen RNAstructure Version 5.6 (Reuter & Mathews 2010) und Mfold Version 3.1 (Zucker 2003) unter den vorgegebenen Einstellungen durchgeführt. Beide Programme bedienen sich der freien Energieminimierung mit dem Ansatz der dynamischen Programmierung (Zucker 1989), basierend auf Energieparametern des nearest-neighbor-Modells (Tinoco, Uhlenbeck & Levine 1971; Delisi & Crothers 1971). Die RNA wurde mit RNAstructure zudem unter Verwendung der SHAPE-Daten (Kap. 2.2.4.2) gefaltet. Dabei wurden die SHAPE-Daten mit einem „Slope-Wert“ von 1,8 kcal/ mol und einem „intercept-Wert“ von -0,6 kcal/ mol als Einschränkungen der Pseudoenergie integriert (Deigan et al. 2009).

Konsensus-Strukturen für RNAs wurden mit dem Programm LocARNA der Version 1.7.16 berechnet. Dabei wurde zunächst ein multiples Alignment der RNA-Moleküle über RIBOSUM-like similarity scoring durchgeführt und die Konsensus-Struktur über das Vienna RNA Packet der Version 2.1.7 berechnet (Smith et al. 2010; Will et al. 2007). Die zufälligen Sequenzen für die Kontrollberechnung wurden mit dem Programm Random DNA Generator erzeugt <<http://www.faculty.ucr.edu/~mmaduro/random.htm>> (Morris Maduro Labor, Universität von Kalifornien, Riverside). Die erzeugten Sequenzen besitzen eine Länge von 537 nt (Länge der VR aus *Nostoc sp.* PCC7120 (137 nt) und weitere 400 nt) und einen GC-Gehalt von 50 % (entspricht der Templat RNA aus *Nostoc sp.* PCC7120 (TR4)).

Die snRNP („small nuclear ribonucleic particles“) Bindemotive in der Templat RNA wurden über das Programm RNA Analyzer identifiziert (Bengert & Dandekar 2003). Um Homologien zu funktionellen RNA-Motiven zu identifizieren, wurde zudem das Programm RegRNA 2.0 verwendet (Chang et al. 2013).

2.2.4.4 Erstellung des phylogenetischen Baums der DGR-kodierten RTs

Das multiple Alignment der Proteinsequenzen von den DGR-kodierten RTs wurde mit ClustalW2 erstellt (Larkin et al. 2007). Die Entwicklungsgeschichte wurde abgeleitet über die Maximum-Likelihood Methode (Whelan & Goldman 2001). Der Baum wurde mit IQ-TREE erstellt und die Integrität über 2000 Bootstrap-Replikate getestet (Minh, Nguyen & von Haeseler 2013).

2.2.4.5 Identifizierung von Sequenzhomologien und Erstellung von Alignments

Die Suche nach Sequenzhomologien wurde mit dem Programm BLAST durchgeführt (Altschul et al. 1990). Um ein Alignment der homologen Sequenzen zu erstellen, wurde das Programm Lasergene (Megalalign) Version 7.1.0 (DNASTAR Inc., Madison, USA) mit der ClustalW Methode verwendet. Mit diesem Programm konnten auch die Sequenzidentitäten festgestellt werden. Sequenzähnlichkeiten wurden mit Hilfe des SIAS-Servers über die Identitäts-Methode PID_1 berechnet <<http://imed.med.ucm.es/Tools/sias.html>> (Pedro Reche Labor, Universität von Complutense, Madrid).

2.2.4.6 Erstellung der Vektorkarten

Die Vektorkarten wurden mit dem Programm Lasergene (Seqbuilder) Version 7.1.0 (DNASTAR Inc., Madison, USA) erstellt.

2.2.4.7 Identifizierung von invertierten Repeats

Für die Identifizierung invertierter Sequenzwiederholungen downstream der variablen Region der untersuchten DGRs, wurden 350 Nukleotide unter Verwendung des EMBOSS 6.3.1 Palindrom-Programms analysiert (Rice, Longden & Bleasby 2000). Dabei wurden folgende

Parameter eingestellt:

- Minimale Länge des Palindroms: 4
- Maximale Länge des Palindroms: 100
- Maximale Länge der Region zwischen den Repeats: 30
- Anzahl der erlaubten Fehler: 0

2.2.4.8 Taxonomie von Cyanobakterien

Die Cyanobakterien wurden nach der Klassifizierung von Prof. Büdel (Pflanzenökologie & Systematik, TU Kaiserslautern) und Juniorprofessor Dr. Kauff (Molekulare Phylogenetik niederer Pflanzen, TU Kaiserslautern) in sechs Ordnungen eingeteilt (Büdel & Kauff 2012). Bei den sechs Ordnungen handelt es sich um die Chroococcales, Nostocales, Oscillatoriales, Pleurocapsales, Pseudanabaenales und die Synechococcales. Eine siebte Ordnung, welche von den Gloeobacterales repräsentiert wird, kommt in der Stammsammlung nicht vor. Den Chroococciopsidaceae kann nach dieser Klassifizierung gegenwärtig keine Ordnung zugewiesen werden.

3 Ergebnisse

3.1 Ein Verfahren für die Identifizierung von DGRs in Cyanobakterien

Diversitätsgenerierende Retroelemente (DGRs) können ihrem Wirt einen selektiven Vorteil über eine zielgerichtete Proteinevolution vermitteln. Trotz dieses potentiellen Nutzens wurden in einer bioinformatischen Analyse öffentlicher Sequenzdatenbanken nur 2,3 % (9) der analysierten Cyanobakterien mit DGRs assoziiert vorgefunden (Schillinger et al. 2012). Einer Hypothese nach könnte das geringe Vorkommen der DGRs mit den konstanten Kultivierungsbedingungen der sequenzierten Stämme zusammenhängen (Schillinger & Zingler 2012). Unter diesen Bedingungen gibt es keinen Anpassungsdruck auf sich ändernde Umweltbedingungen. In der Folge könnten die DGRs als genomischer Ballast entfernt worden sein. Der Zweck des folgenden Versuchs war die experimentelle Überprüfung der oben genannten Hypothese, sowie die Identifizierung neuer DGRs in Cyanobakterien.

3.1.1 Etablierung eines PCR-Screenings auf die DGR-kodierte RT

Eine Strategie für die Identifizierung von Retroelementen in Organismen stellt die Amplifizierung der RT über degenerierte Primer dar (Koniczny et al. 1991; Flavell, Smith & Kumar 1992; Friesen, Brandes & Heslop-Harrison 2001; Ahmed et al. 2011).

Bei dem zu untersuchenden Organismus handelt es sich in dieser Arbeit um Cyanobakterien. Die verwendeten Cyanobakterien wurden von Prof. Burkhard Büdel (Pflanzenökologie & Systematik, TU Kaiserslautern) zur Verfügung gestellt. Es handelt sich um eine Sammlung von 169 Isolaten aus den verschiedensten ökologischen Habitaten (Tabelle A1). Die Stammmummern der Sammlung bekommen im Folgenden den Zusatz BB (Burkhard Büdel).

Zu Beginn der Arbeit waren bereits einige Cyanobakterien mit DGRs bekannt (Doulatov et al. 2004). Die RTs dieser Cyanobakterien wurden für ein Aminosäure-Alignment verwendet, bei dem zwei hoch konservierte Bereiche festgestellt wurden. Dabei handelt es sich um einen Bereich in der katalytischen Domäne fünf, welcher das YxDD-Motiv beinhaltet (YVDDFALF), sowie einen Bereich in Domäne drei (QCDIXKYFP) (Abbildung 12, S. 47). Für diese Bereiche wurden von Dr. Nora Zingler degenerierte Primer erstellt, die ein PCR-Produkt von ca. 320 bp ergeben sollten.

Die Primer wurden in Vorversuchen von Lars Fölsch getestet und das Screening anschließend von Christine Pakebusch optimiert und durchgeführt. Für die Optimierung wurde versucht die zeitaufwändige DNA-Isolierung vor der PCR zu umgehen. Dabei wurde ein Protokoll entwickelt mit dem ganze Cyanobakterienzellen in die PCR eingesetzt werden können. In der PCR wurde parallel zur RT auch die 16S rDNA amplifiziert. Der Nachweis der 16S rDNA dient als Kontrolle für die Zugänglichkeit der Probe für die PCR. Um falsche Ergebnisse aufgrund einer fehlerhaft durchgeführten PCR zu vermeiden, wurden PCR-unzugängliche Cyanobakterien mindestens ein weiteres Mal getestet.

Ein Ausschnitt aus dem Screening ist in Abbildung 6 dargestellt. Dabei zeigen drei der getesteten Stämme keine Bande für die 16S rDNA und sind damit nicht für die PCR zugänglich. Drei der getesteten Stämme zeigen eine Bande für die RT. Diese liegt bei *Anabaena flos-aquae* (BB97.35) und *Tolypothrix distorta* (BB97.80) bei den theoretisch berechneten 320 bp. Ein größeres PCR-Produkt konnte für *Aphanizomenon flos-aquae* (BB97.85) beobachtet werden.

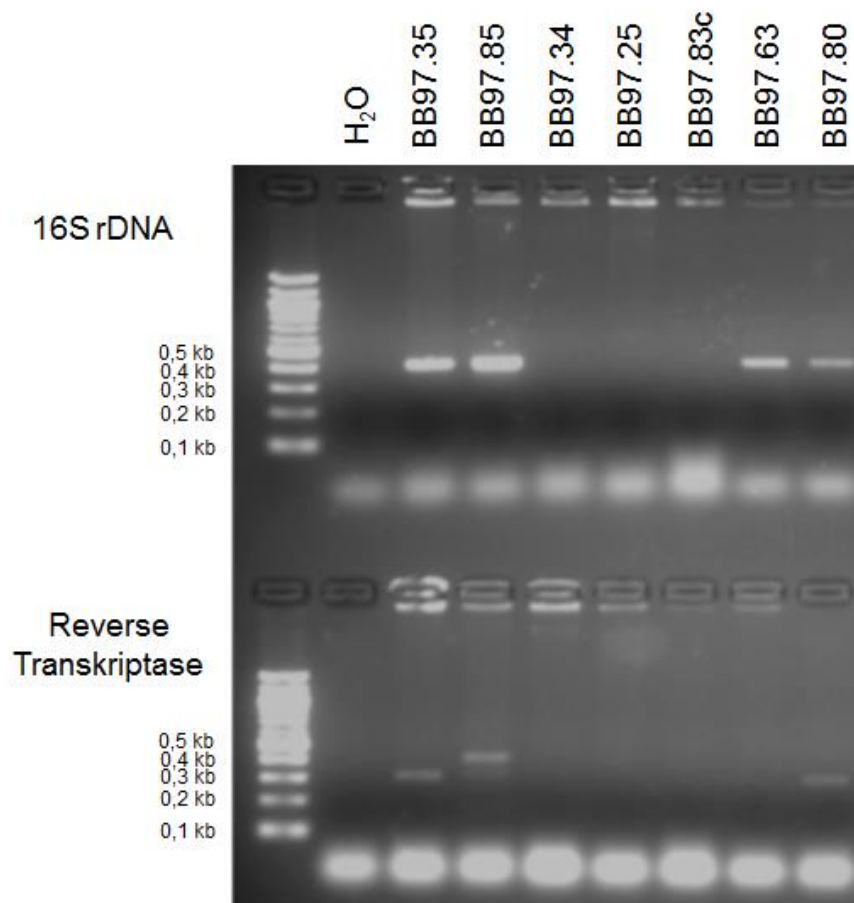


Abbildung 6: Ausschnitt aus dem cyanobakteriellen Screening auf die DGR-kodierte RT. Verschiedene Stämme der Cyanobakteriensammlung von Prof. Büdel (BB) wurden mittels PCR auf Vorhandensein eines RT-Leserahmens untersucht. Von jedem Stamm wurde parallel ein Teil der 16S rDNA (405 bp) amplifiziert, um eine Zugänglichkeit der Probe für die PCR zu testen. *Anabaena flos-aquae* (BB97.35) und *Tolypothrix distorta* (BB97.80) weisen jeweils eine Bande bei ca. 320 bp auf. Eine größere Bande von ca. 400 bp konnte für *Aphanizomenon flos-aquae* (BB97.85) beobachtet werden.

In dem kompletten Screening wurde für 15 Cyanobakterien eine Bande beobachtet, die größer war als die berechneten 320 bp (Tabelle A1). Da die Größe von einer zusätzlichen RT-Domäne resultieren könnte, wurden vier dieser Banden sequenziert. Dabei konnte in keinem Fall eine Homologie zu einer bekannten RT-Sequenz festgestellt werden.

Banden der theoretisch korrekten Länge konnten für 34 Cyanobakterien beobachtet werden (Tabelle 8). Von diesen Banden wurden 18 sequenziert, wobei in allen Fällen eine Homologie zu RT-Sequenzen festgestellt werden konnte. Dabei wurden teilweise Banden sequenziert, die von verschiedenen Umweltisolaten der gleichen Cyanobakterienart stammen (*Tolypothrix distorta*: BB97.17, BB97.18 und BB97.80 und *Scytonema myochrous*: BB97.22 und BB97.84). Es konnte beobachtet werden dass sich die Nukleotidsequenzen der drei Isolate von *Tolypothrix distorta* exakt gleichen, während sich die Nukleotidsequenzen der beiden Isolate von *Scytonema myochrous* in einem Nukleotid an Position 111 unterscheiden (Thymin (BB97.22) bzw. Guanin (BB97.84)).

Zu den 18 RT-Sequenzen gehören auch zwei Mischsequenzen (BB97.81 und BB97.24), die durch mehrere RTs innerhalb des Organismus zu erklären sind.

Tabelle 8: Cyanobakterien mit einem PCR-Produkt der korrekten Länge für die RT

Cyanobakterium	BB	Ordnung	Anmerkung
<i>Anabaena flos-aquae</i>	97.35	<i>Nostocales</i>	RT-Sequenz
<i>Anabaena planctonica</i>	97.102	<i>Nostocales</i>	RT-Sequenz
<i>Aphanizomenon flos-aquae</i>	97.111b	<i>Nostocales</i>	
<i>Chroococciopsis</i> sp.	01.01	ungewisse Ordnung	
<i>Chroococcus</i> sp.	97.96	<i>Chroococcales</i>	RT-Sequenz
<i>Cylindrospermum</i> sp.	97.12	<i>Nostocales</i>	
<i>Dichothrix</i> sp.	97.87	<i>Nostocales</i>	RT-Sequenz
<i>Fischerella muscicola</i>	98.1	<i>Nostocales</i>	
<i>Microcoleus chthonoplastes</i>	02.2	<i>Oscillatoriales</i>	
<i>Microcoleus chthonoplastes</i>	92.3	<i>Oscillatoriales</i>	RT-Sequenz
<i>Microcystis aeruginosa</i>	97.43	<i>Chroococcales</i>	
Nicht näher bestimmt	97.108	--	RT-Sequenz
<i>Nostoc commune vaucher</i>	97.40	<i>Nostocales</i>	RT-Sequenz
<i>Nostoc</i> sp.	97.53	<i>Nostocales</i>	
<i>Nostoc</i> sp.	97.88	<i>Nostocales</i>	
<i>Nostochopsis lobatus</i>	92.1	<i>Nostocales</i>	
<i>Oscillatoria limosa</i>	97.105	<i>Oscillatoriales</i>	
<i>Oscillatoria rubescens</i>	97.107	<i>Oscillatoriales</i>	
<i>Phormidium favosum</i>	97.124	<i>Oscillatoriales</i>	RT-Sequenz
<i>Rivularia</i> sp.	97.24	<i>Nostocales</i>	Mischsequenz (RTs)
<i>Scytonema capitatum</i>	97.62	<i>Nostocales</i>	
<i>Scytonema javanicum</i>	97.58	<i>Nostocales</i>	
<i>Scytonema myochrous</i>	97.84	<i>Nostocales</i>	RT-Sequenz
<i>Scytonema myochrous.</i>	97.22	<i>Nostocales</i>	RT-Sequenz
<i>Scytonema</i> sp.	97.127b	<i>Nostocales</i>	
<i>Scytonema stuposum</i>	97.78	<i>Nostocales</i>	
<i>Stigonema mamillosum</i>	97.104b	<i>Nostocales</i>	RT-Sequenz
<i>Stigonema panniforme</i>	97.27	<i>Nostocales</i>	
<i>Tolypothrix byssoidea</i>	97.10	<i>Nostocales</i>	RT-Sequenz
<i>Tolypothrix byssoidea</i>	97.81	<i>Nostocales</i>	Mischsequenz (RTs)
<i>Tolypothrix distorta</i>	97.17	<i>Nostocales</i>	RT-Sequenz
<i>Tolypothrix distorta</i>	97.18	<i>Nostocales</i>	RT-Sequenz
<i>Tolypothrix distorta</i>	97.80	<i>Nostocales</i>	RT-Sequenz
<i>Tolypothrix</i> sp.	97.26	<i>Nostocales</i>	RT-Sequenz

3.1.2 Alignments der identifizierten RT-Sequenzen

Im Screening auf die DGR-kodierte RT wurden Banden sequenziert die eine Größe von ca. 320 bp aufweisen. Die Sequenzierung ergab in allen Fällen eine Homologie zu bekannten RT-Sequenzen. Anhand dieser Sequenzen wurden ein Nukleotid- und Aminosäure-Alignment erstellt (Abbildung 7). In das Alignment wurden keine Mischsequenzen aufgenommen, da für die Nukleotide keine exakte Zuordnung zu einer RT möglich war. Zudem wurde bei mehreren sequenzierten Banden unterschiedlicher Isolate einer Art nur eine RT-Sequenz verwendet, wenn diese sich von den anderen nicht deutlich unterscheidet. Das betrifft *Tolypothrix distorta* (BB97.17 im Alignment) und *Scytonema myochrous* (BB97.22 im Alignment). Die meisten Banden des Screenings wurden nur von einer Seite sequenziert. Da die ersten Nukleotide einer Sequenzierung nicht verlässlich sind, wurden diese vor dem Alignment entfernt, was durch Boxen im Alignment gekennzeichnet ist. In den Alignments ist das (L/I/V)GxxxSQ-Motiv (SQ-Motiv) farblich hervorgehoben. Es ist charakteristisch für DGR-kodierte RTs und in allen Sequenzen des Alignments vorhanden. Auch die Abweichung (SQ → NQ) im Falle von *Stigonema mamillosum* ist für das Motiv bereits beschrieben (Schillinger et al. 2012). Somit kann eine 320 bp Bande im cyanobakteriellen Screening als ein verlässlicher Indikator für DGRs angesehen werden.

In den Alignments wurden die Sequenzidentitäten und anhand der physikalisch-chemischen Eigenschaften der Aminosäuren, auch die Sequenzähnlichkeiten berechnet.

Für das Nukleotid-Alignment wurde eine Sequenzidentität von 60 % (*Anabaena flos-aquae* vs. *Nostoc commune vaucher*) bis 97 % (*Tolypothrix distorta* vs. *Tolypothrix sp.*) festgestellt. Dabei treten Abweichungen von der Konsensus-Sequenz vor allem an dem dritten Nukleotid eines Kodons auf (Wobble-Position), wodurch Aminosäureaustausche minimiert werden. Dies ist unter anderem an der farblich unterlegten Sequenz (entspricht dem SQ-Motiv im Aminosäure-Alignment) zu beobachten. Dabei beschreibt ein „+“ unter der Konsensus-Sequenz das erste Nukleotid eines Kodons.

Im Aminosäure-Alignment wurde eine Sequenzidentität von 64 % (*Anabaena flos-aquae* vs. *Nostoc commune vaucher*) bis 97 % (*Tolypothrix distorta* vs. *Tolypothrix sp.*) festgestellt. Zudem wurde eine Sequenzähnlichkeit von 76 % (*Anabaena-flos-aquae* vs. *Chroococcus-sp.*) bis 97 % (*Tolypothrix distorta* vs. *Tolypothrix sp.*) berechnet. Im Beispiel von *Anabaena flos-aquae* und *Nostoc commune vaucher* wird eine Sequenzähnlichkeit von 79 % berechnet, obwohl nur eine Sequenzidentität von 64 % vorliegt.

Nukleotid-Alignment



Aminosäure-Alignment



Abbildung 7: Sequenz-Alignments der cyanobakteriellen RT-Fragmente. Nukleotid- und Aminosäuresequenzen wurden mit ClustalW aligned. Dabei sind Aminosäuren und Nukleotide nur eingezeichnet, wenn sich diese von der Konsensus-Sequenz unterscheiden. Die ersten sequenzierten Nukleotide einer Sequenzierreaktion sind nicht verlässlich und wurden entfernt (Boxen). In der Konsensus-Sequenz des Alignments ist das charakteristische (L/I/V)GxxxSQ-Motiv farblich hervorgehoben. Alle Sequenzen im Alignment verfügen über dieses Motiv. Die Aminosäuren wurden nach ihren physikalisch-chemischen Eigenschaften gefärbt. Einteilung der Aminosäuregruppen: aromatisch = F; Y; W (Lila); hydrophob = A; M; C; G (hellgrün); aliphatisch (und hydrophob) = V; I; L (dunkelgrün); positiv geladen = K; R; H (blau); negativ geladen = D; E (rot); polar = T; Q; S; N (orange) und klein = P (schwarz).

3.1.3 DGR-kodierte RTs aus Cyanobakterien bilden einen Subtree im phylogenetischen Baum

Schillinger und Kollegen haben für 155 DGR-kodierte RT-Sequenzen einen phylogenetischen Baum auf Aminosäureebene erstellt (Schillinger et al. 2012). Die RT-Sequenzen wurden in dieser Arbeit verwendet um einen weiteren phylogenetischen Baum auf Aminosäureebene zu generieren, bei dem die 13 RT-Sequenzen des Screenings hinzugefügt wurden (Abbildung A2). Die Erstellung des Baums wurde von Damir Baranasic (Genetik, TU Kaiserslautern) durchgeführt. Dabei wurde die RT des Gruppe II Introns aus *Bacillus halodurans* als Wurzel verwendet. Um die Ergebnisse zu validieren, wurde ein Bootstrapping mit 2000 Replikaten durchgeführt. Für ein verlässliches Bootstrapping wurden die Sequenzlängen aus der Datenbank und dem Screening angepasst.

In Abbildung 8 ist ein Subtree des phylogenetischen Baums dargestellt. Darin gruppieren sich die cyanobakteriellen DGR-kodierten RT-Sequenzen der Datenbank, mit denen des Screenings. In diesem Subtree sind ebenso die Sequenzen von zwei grünen Schwefelbakterien (*Chlorobium phaeobacteroides* und *Pelodictyon phaeoclathratiforme*), einem Schwefelpurpurbakterium (*Thiorhodovibrio* sp.) und einem Nichtschwefelpurpurbakterium (*Rhodomicrobium vannielii*) vorzufinden. Die Cyanobakterienstämme *Cyanothece* sp. CCY0110 und *Nodularia spumigena* CCY9414 verfügen über zwei DGRs, deren RT-Sequenzen voneinander abweichen. Die RT-Sequenzen aus *Cyanothece* sp. CCY0110 besitzen eine Sequenzidentität von 68 %, während die RT-Sequenzen aus *Nodularia spumigena* CCY9414 eine Sequenzidentität von 72 % aufzeigen.

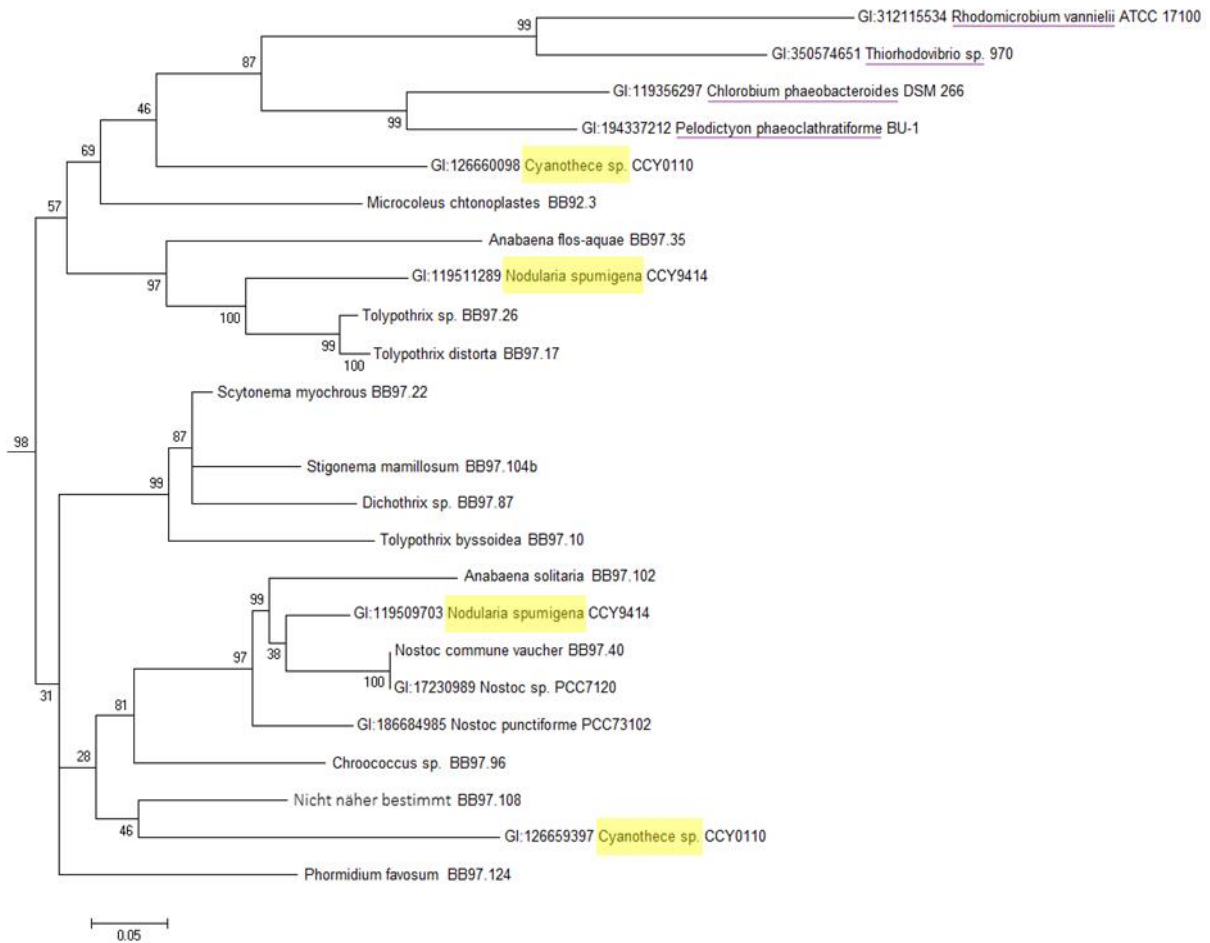


Abbildung 8: Subtree der cyanobakteriellen DGR-kodierten RTs im phylogenetischen Baum. Abgebildet ist ein Ausschnitt des phylogenetischen Baums der DGR-kodierten RTs (s. Abbildung A2). Sämtliche cyanobakterielle Sequenzen (Datenbank und Screening) formen zusammen eine eigenständige Gruppe, zu der auch grüne Schwefelbakterien und Purpurbakterien gehören (lila unterstrichen). Die Stämme *Nodularia spumigena* CCY9414 und *Cyanothece sp.* CCY0110 (gelb hervorgehoben) besitzen jeweils zwei DGRs. Die Bootstrap-Werte für 2000 Replikate sind angegeben.

3.1.4 Geographische Verteilung der DGR-assozierten Cyanobakterienstämme

In den vorangehenden Kapiteln konnte gezeigt werden, dass eine 320 bp Bande im Screening der Cyanobakterien als ein verlässlicher Indikator für DGRs angesehen werden kann. Um einen Zusammenhang zwischen dem Isolationsort der Cyanobakterien und dem Vorkommen der DGRs zu untersuchen, wurde die geographische Verteilung der DGR-assozierten Cyanobakterienstämme analysiert (Tabelle 9).

Tabelle 9: Geographische Verteilung der DGR-assozierten Cyanobakterienstämme

Isolationsort	Isolate		DGRs	
	[%]	[Anzahl]	[%]	[Anzahl]
Europa (Schweiz)	49,7 (35,5)	84 (60)	58,8 (50,0)	20 (17)
Afrika (Namibia)	13,0 (6,5)	22 (11)	8,8 (2,9)	3 (1)
Nordamerika (Kanada)	11,2 (6,5)	19 (11)	2,9 (2,9)	1 (1)
Indien/Nepal	8,3	14	11,8	4
Zentral- & Südamerika	3,0	5	0,0	0
Antarktika	0,6	1	0,0	0
nicht angegeben	14,2	24	17,6	6

Anhand von Tabelle 9 ist zu erkennen, dass die meisten Stämme in Europa isoliert wurden (49,7 %), mit einem Fokus in der Schweiz (35,5 %). Daher ist es nicht verwunderlich, dass auch die meisten DGRs in diesen Stämmen vorgefunden wurden (58,8 % bzw. 50 %). Hinzu kommen Stämme aus Afrika (13 %), die 8,8 % der DGR-assozierten Stämme ausmachen, Nordamerika (11,2 %), die 2,9 % der DGR-assozierten Stämme ausmachen und Indien/Nepal (8,3 %), welche 11,8 % der DGR-assozierten Stämme ausmachen. Die wenigsten Stämme wurden in Zentral- und Südamerika (5) bzw. Antarktika (1) isoliert. In diesen Stämmen wurden keine DGRs identifiziert, was höchstwahrscheinlich mit der geringen Anzahl der isolierten Stämme zusammenhängt.

3.1.5 Entdeckung eines neuen DGRs in *Anabaena flos-aquae*

In dem cyanobakteriellen Screening auf die DGR-kodierte RT konnten von den 169 Isolaten der Stammsammlung 34 Stämme mit einer Bande der korrekten Länge identifiziert werden. Beispielhafte Sequenzierungen haben gezeigt, dass es sich um RT-Sequenzen mit dem für DGRs charakteristischen SQ-Motiv handelt. Diese Sequenzen beanspruchen, zusammen mit cyanobakteriellen DGR-kodierten RT-Sequenzen der Datenbank, eine eigene Gruppe im phylogenetischen Baum. Für weitere Experimente wird die komplette Sequenz der identifizierten DGRs benötigt. Diese kann durch Genomsequenzierungen erlangt werden. Eine kostengünstige Alternative stellt die inverse PCR dar. Mit der inversen PCR kann ausgehend von einem bekannten Sequenzbereich, in diesem Fall dem RT-Fragment, die umliegende Sequenz identifiziert werden. Hierzu wird genomische DNA isoliert, mit einem Restriktionsenzym geschnitten und unter hoher Verdünnung intramolekular ligiert. Die bekannte RT-Sequenz sollte nun Teil eines zirkulären DNA-Moleküls sein, auf dem sich ebenfalls die umgebende Sequenz befindet. Die umgebende Sequenz ist dabei durch die Erkennungssequenz des verwendeten Restriktionsenzym definiert. Die umgebende, unbekannte Sequenz kann durch auswärts gerichtete Primer amplifiziert werden, deren Richtung nicht aufeinander zu, sondern in den unbekanntem Bereich zielt. Die Abfolge der Reaktionsschritte für die inverse PCR ist schematisch in Abbildung 9 dargestellt.

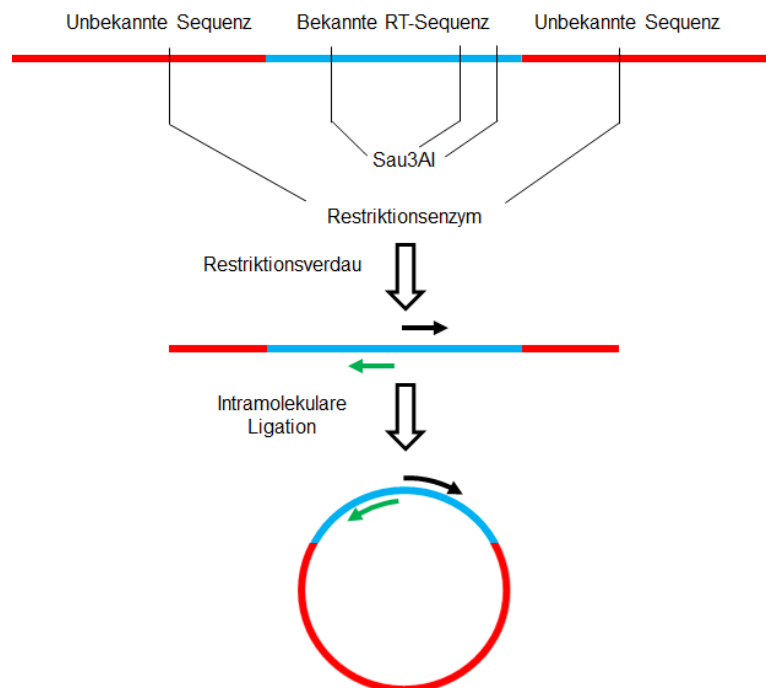


Abbildung 9: Schematische Darstellung der inversen PCR. Nach der Isolation genomischer DNA wird diese mit einem Restriktionsenzym geschnitten. Die linearisierten DNA-Fragmente werden intramolekular ligiert und die entstehende zirkuläre DNA als Template für eine PCR verwendet. Dafür werden Primer benutzt, deren Richtung in die unbekannte Sequenz zielt (schwarzer und grüner Pfeil). Die Erkennungssequenz von Sau3AI kommt in der bekannten RT-Sequenz von *Anabaena flos-aquae* dreimal vor. Dieses Restriktionsenzym wurde in der inversen PCR als Positivkontrolle verwendet.

Die inverse PCR wurde auf der RT-Sequenz von *Anabaena flos-aquae* durchgeführt (Abbildung 10). Um zu testen ob intramolekular ligierte DNA entstanden ist, wurde eine Positivkontrolle mitgeführt. Dabei wurde das Restriktionsenzym *Sau3AI* verwendet, dessen 4 bp-Erkennungssequenz dreimal in der bekannten RT-Sequenz vorkommt. Dadurch können unter anderen zwei PCR-Produkte bekannter Größe (216 bp und 291 bp) entstehen. Während das 216 bp große PCR-Produkt nicht beobachtet werden konnte, sind dafür neben der 291 bp Bande noch weitere, größere Banden zu erkennen. Eine Sequenzierung dieser Banden ergab, dass es sich um Sequenzwiederholungen der 216 bp und 291 bp Sequenzen handelt. Diese Sequenzwiederholungen können entstehen, wenn die DNA-Synthese auf zirkulärer DNA kontinuiert, ähnlich der bakteriellen „rolling circle replication“ (zusammengefasst in Khan 1997). Ausgehend von DNA, die zuvor einem Restriktionsverdau mit *Bam*HI, *Hind*III, *Pvu*II oder dem Doppel-Restriktionsverdau *Bam*HI/*Bgl*II unterzogen wurde, konnte kein deutliches PCR-Produkt amplifiziert werden. Dagegen konnte mit DNA, welche mit *Eco*RI geschnitten wurde, ein PCR-Produkt auf Höhe von ca. 3 kb amplifiziert werden.

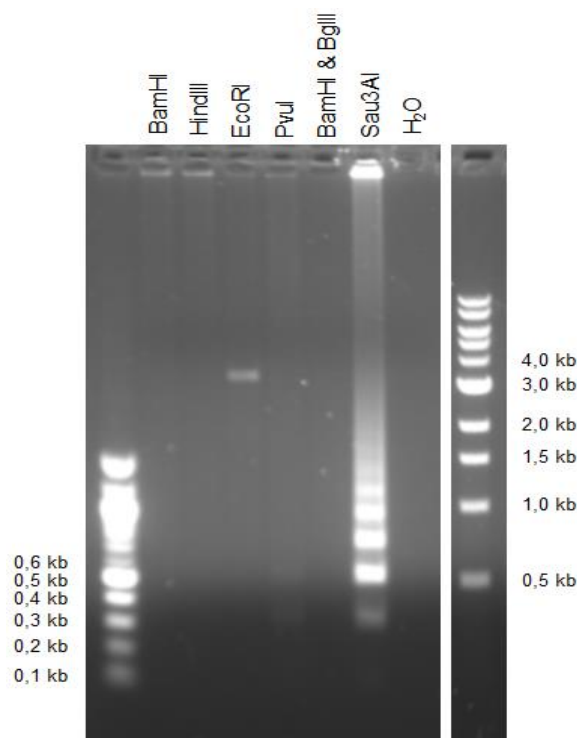


Abbildung 10: Inverse PCR von *Anabaena flos-aquae*. Die inverse PCR auf die RT-Sequenz von *Anabaena flos-aquae* wurde wie in Abbildung 9 beschrieben durchgeführt. Die verwendeten Restriktionsenzyme sind eingezeichnet und die resultierenden PCR-Produkte abgebildet. Dabei ist für den Restriktionsverdau mit *Eco*RI eine Bande auf Höhe von ca. 3 kb zu beobachten. Als Positivkontrolle diente der Restriktionsverdau mit *Sau3AI*, da dieses über drei Erkennungssequenzen in der bekannten Sequenz verfügt. Die resultierenden PCR-Produkte sollten unter anderen eine Größe von 216 bp bzw. 291 bp besitzen. Es konnten jedoch neben der 291 bp Bande noch weitere Banden beobachtet werden.

Das PCR-Produkt der mittels *Eco*RI restriktionsverdauten DNA wurde kloniert und anschließend sequenziert. Dabei konnte die Sequenz eines ca. 2900 bp großen DGRs identifiziert werden. Dieses setzt sich aus einem Zielgen mit VR am 3'-Ende, *Avd*, *RT* und *TR*

zusammen (Abbildung 11). Da das komplette Genom nicht sequenziert wurde, können weitere Zielgene vorliegen.

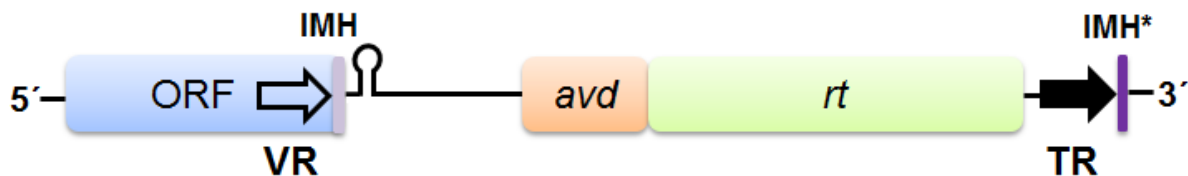


Abbildung 11: Schematische Darstellung des DGRs aus *Anabaena flos-aquae*. Das abgebildete DGR wurde über inverse PCR in dem Cyanobakterium *Anabaena flos-aquae* identifiziert. Es besitzt ein Zielgen mit VR am 3'-Ende, sowie einen *avd*- und *rt*-ORF. Der TR ist downstream zum *rt*-ORF lokalisiert. Die IMH und IMH* Region, sowie eine potentielle Haarnadel/Kreuz-Struktur downstream der VR sind eingezeichnet.

Die RT des DGRs aus *Anabaena flos-aquae* ist 349 Aminosäuren lang und verfügt über die sechs Domänen, die für DGR-kodierte RTs beschrieben wurden (Simon & Zimmerly 2008) (Abbildung 12). Die wichtigsten Merkmale innerhalb der sechs Domänen stellen das für DGRs charakteristische SQ-Motiv in Domäne vier und das YxDD-Motiv in der katalytischen Domäne fünf dar. Der Bereich um das YxDD-Motiv sowie ein konservierter Bereich in Domäne drei, dienen als Bindestelle der degenerierten Primer des cyanobakteriellen Screenings.

Auf die komplette Aminosäuresequenz der RT bezogen, zeigt die RT aus *Anabaena flos-aquae* eine Sequenzidentität von 65 % (Sequenzähnlichkeit = 80 %) zu der RT aus *Nostoc sp.* PCC7120, aber nur eine Sequenzidentität von 35 % (Sequenzähnlichkeit = 49 %) zu der RT des *Bordetella*-Phagen. Wie bei anderen DGR-kodierten RTs, verfügt die RT aus *Anabaena flos-aquae* weder über eine Homing Endonuklease-Domäne, wie sie bei den Gruppe II Introns vorkommt, noch über eine RNase H-Domäne, wie sie z.B. für die HIV-RT beobachtet wurde (Johnson et al. 1986).

	0	1	2	2a	3
<i>Anabaena flos-aquae</i>			RIISAAPYRDRVV	DSYANR	YVIQCDIKKYFFPSIDHHILKKIL
<i>Nostoc sp.</i> PCC7120			RLISAAPYRDRVV	DSYANR	YVIQCDIRKYFFPSIDHIILKELI
<i>Bordetella</i> -Phage			RLISALEFKDRLV	YTYACR	HFLKSDFSKFFPSIDRAALYAMI

	4	5	6	7
<i>Anabaena flos-aquae</i>	RRGLEIGNLTSQFLSNVYLN	KYVRYVDDFAIFADDQ	LHPVKQTQL	VNFLGFRI
<i>Nostoc sp.</i> PCC7120	RKGLEIGNLTSQFFANIYLN	KYVRYVDDFALESDDR	IHPIKSQL	ATFLGFRI
<i>Bordetella</i> -Phage	-VGIFIGSLTSQLFANVYGG	HWARYMDDIVVLGDDP	LKISHWQV	INFLGYRI

Abbildung 12: Alignment der DGR-kodierten RT-Domänen. Aminosäure-Alignment der RT-Domänen 0-7 für die DGR-kodierten RTs von *Anabaena flos-aquae*, *Nostoc sp.* PCC7120 und dem *Bordetella*-Phagen. DGR-kodierte RTs verfügen über keine Sequenz, die in Domäne 0-1 mit anderen RT-Sequenzen aligned. Das SQ-Motiv in Domäne vier ist für die drei RT-Sequenzen hervorgehoben (gelb). Für die cyanobakteriellen RT-Sequenzen sind zudem die Bindestellen der degenerierten Primer des Screenings hervorgehoben (rot). Diese liegen auf dem YxDD-Motiv in Domäne fünf und einem konservierten Bereich in Domäne drei. Die Einteilung der Aminosäuren nach den physikalisch-chemischen Eigenschaften (farbliche Kodierung) entspricht der aus Abbildung 7 (S.41).

Unmittelbar upstream zur RT ist die Avd lokalisiert. Bei der Avd handelt es sich um ein 115 Aminosäuren großes Protein mit der Sequenz IGNELGGWIK am C-Terminus, was dem von Schillinger beschriebenen GW-Motiv entspricht (Schillinger 2013) (Abbildung 13). Die Aminosäuresequenz der Avd aus *Anabaena flos-aquae* besitzt eine Sequenzidentität von 63 % (Sequenzähnlichkeit = 79 %) zu der Avd aus *Nostoc sp.* PCC7120, aber nur eine Sequenzidentität von 23 % (Sequenzähnlichkeit = 33 %) zu der Avd des *Bordetella*-Phagen.

Anabaena flos-aquae	-----MSDLPIVQKTYDLIKWYVPIIDKLPKNHKFGLGDRIVSGLYDLLEGFILARYARE	55
Nostoc sp. PCC7120	-----MKELSVIQKTYDCIKWYVPIIERLPKIHKFTLGDRIINQLYDLLEGLIKAKYAKN	55
Bordetella-Phage	MEPIEEATKCYDQMLIVERYERVISYLYPIAQSIPRKHGVAR-EMFLKCLLQVELFIVAGKS-N	63
Anabaena flos-aquae	KLGQLELLNAKIDILRHQTRLLDFDGL---FDERRYEYVGQLINDIGNELGGWIKQQRQVRHS	115
Nostoc sp. PCC7120	KLPQLESLSQLDILRYQTRMLDFDK---MSIEREYEVIKLIDEIGTELGGWIKNQRNREK	114
Bordetella-Phage	QVSKLYAADAGLAMLRFWLRFVLAGIQKPHAMTPHQVETAQVLIAEVGRILGSWIARVNRKQAGK	128

Abbildung 13: Alignment von Avd-Sequenzen. Aminosäure-Alignment der Avd-Sequenzen von *Anabaena flos-aquae*, *Nostoc sp.* PCC7120 und dem *Bordetella*-Phagen. Das GW-Motiv am C-terminalen Ende ist hervorgehoben. Die Einteilung der Aminosäuren nach den physikalisch-chemischen Eigenschaften (farbliche Kodierung) entspricht der aus Abbildung 7 (S.41).

Das variable Protein ist 252 Aminosäuren lang und besitzt eine konservierte Proteindomäne, die es in die Superfamilie der Formylglycin-generierenden Enzyme (FGEs) einordnet. Am 3'-Ende des offenen Leserahmens befindet sich die 131 bp lange VR, die sich von dem downstream zur RT gelegenen TR fast ausschließlich an Positionen unterscheidet, an denen der TR über Adenine verfügt (Abbildung 14).

TR	AA AA AA	A AAA	A	AA AA A	65
VR	GGCTGCTGCGCGGTGGTTCTTGGGACAACAACCCTAGACTTTGCCGTTCTGCGTATCGTACCTTCT				65
TR	A AA	A A AA A AAA AA AA AA	A A	C	131
VR	ACTACCCAGATTT	CAGGGACAGCAACGT	TGGGTTTCGGTTAGTAGTTT	CCGGCGCAAGGACTCTT	131

Abbildung 14: Variable Region des DGRs aus *Anabaena flos-aquae*. Die variable Region (VR) am 3'-Ende des Zielgens des DGRs aus *Anabaena flos-aquae* ist gezeigt. Bei dem fast identischen Templat Repeat (TR) wurden übereinstimmende Nukleotide, bis auf die Adenine, ausgeblendet. Adenine in dem TR können zu variablen Positionen in der VR führen (rot). Die Initiierungssequenz des mutagenen Homings (IMH, blau) besitzt keine variablen Positionen. Daher wurden die Adenine in dem korrespondierenden Bereich des TR (IMH*) ausgeblendet. Zwischen IMH und IMH* ist eine Abweichung zu beobachten (unterstrichen).

Der variable Bereich der VR ist 114 bp lang und beinhaltet 35 potentiell variable Positionen, die durch Adenine in dem TR vorgegeben werden. Davon zeigen 15 Positionen eine Abweichung zum TR. Das Zielgen kann eine maximale Diversität von $(4^{35}) 1,2 \times 10^{21}$ Nukleotidsequenzen generieren, was $1,6 \times 10^{18}$ Proteinsequenzen entspricht.

Am 3'-Ende der VR ist eine Sequenz lokalisiert, die aufgrund der Ähnlichkeit zum DGR des *Bordetella*-Phagen als Initiierungssequenz des mutagenen Homings (IMH) definiert wurde. Die IMH besteht aus einer (GC)₇ Region, gefolgt von einer 10 bp Sequenz mit einem Adenin-unspezifischen Unterschied zu der entsprechenden Sequenz des TR (IMH*). In der IMH* kommen drei Adenine vor, die in der IMH nicht mutiert vorliegen und im Hinblick auf die IMH des *Bordetella*-Phagen auch nicht als variabel angesehen werden.

68 bp downstream zur VR ist eine Sequenz lokalisiert, welche die Ausbildung einer Haarnadel/Kreuz-Struktur erlaubt. Die ausgebildete Struktur würde aus einer 6 bp langen, GC-reichen Doppelhelix und einer 3 nt langen Schleife (rot) bestehen.

5'-GTCGCCTCTGGCGAC-3'
3'-CAGCGGAGACCGCTG-5'

Die Ausbildung einer solchen Struktur scheint für DGRs konserviert zu sein und wird mit der Erkennung der Zielsequenz im mutagenen Homing in Verbindung gebracht (Guo et al. 2011).

Das DGR aus *Anabaena flos-aquae* sollte theoretisch funktionsfähig sein, da keine inaktivierenden Mutationen beobachtet werden konnten. Für weitere Experimente wurde das DGR in einen Standardvektor (pBSK) kloniert (Abbildung A1). Mit der inversen PCR wurden keine weiteren DGRs identifiziert.

3.1.5.1 *Anabaena flos-aquae* besitzt eine durch Kältestress induzierte RT

Das DGR aus *Anabaena flos-aquae* scheint auf DNA-Ebene funktionsfähig zu sein (Kap. 3.1.5). Daher wurde neben *Nostoc sp.* PCC7120, welches ebenfalls ein theoretisch funktionsfähiges DGR besitzt (Abbildung 4), auch *Anabaena flos-aquae* im Labor kultiviert. Um zu prüfen, ob die DGRs unter den Kultivierungsbedingungen transkribiert werden, wurde eine Reverse-Transkriptase-PCR (RT-PCR) auf die Transkripte von TR, RT und Avd durchgeführt. Hierfür wurde Gesamt-RNA aus Cyanobakterien isoliert, die entweder Standardbedingungen (22 °C) oder einem zusätzlichen Kältestress (drei Stunden bei 4 °C) ausgesetzt waren. Die isolierte Gesamt-RNA wurde durch eine reverse Transkriptase in cDNA umgeschrieben und als Templat in eine PCR eingesetzt. Die Analysen wurden mit Unterstützung von Thomas Kafka im Triplikат durchgeführt und sind in Abbildung 15 zusammengestellt.

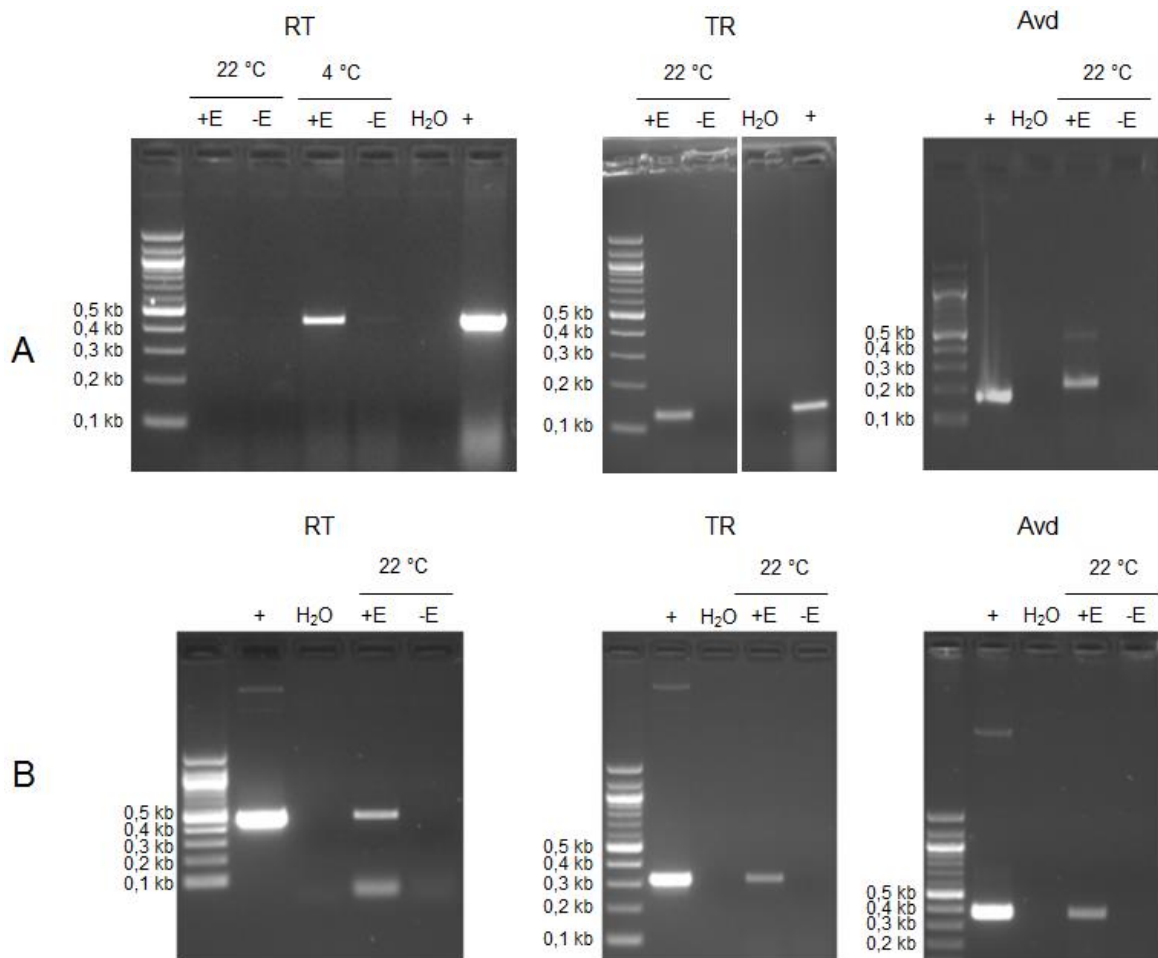


Abbildung 15: Detektion der cyanobakteriellen DGR-Transkripte. Die Transkripte der RT, TR und Avd der DGRs aus (A): *Anabaena flos-aquae* und (B): *Nostoc sp.* PCC7120 wurden mittels Reverse-Transkriptase-PCR amplifiziert. Die Reaktion wurde mit (+E) und ohne (-E) Enzym (reverse Transkriptase) durchgeführt, um auf Präsenz von gDNA zu testen. Als Positivkontrolle (+) dienten die Plasmide der klonierten DGRs (pPM6 und pPM7). Die RNA wurde aus Cyanobakterien isoliert, die entweder bei 22 °C oder zusätzlich für drei Stunden bei 4 °C kultiviert wurden. Alle Transkripte konnten nachgewiesen werden, wobei die RT aus *Anabaena flos-aquae* nur in Kulturen detektiert wurde, die zuvor 4 °C ausgesetzt waren. Die beobachtete Größe der PCR-Produkte stimmt mit der theoretischen überein (Ausnahme: *Anabaena flos-aquae*, Avd): *Anabaena flos-aquae* RT (436 bp), TR (130 bp) und Avd (182 bp), sowie *Nostoc sp.* PCC7120 RT (488 bp), TR (320 bp) und Avd (342 bp).

Mit der RT-PCR konnten die Transkripte der RT, TR und Avd aus *Nostoc sp.* PCC7120 Kulturen nachgewiesen werden, die bei 22 °C kultiviert wurden. Bei *Anabaena flos-aquae* konnten bei dieser Kultivierungstemperatur nur die Transkripte von TR und Avd detektiert werden. Dabei ist zu beobachten, dass die Avd-Bande etwas höher läuft als die zugehörige Positivkontrolle, was durch einen Probenüberschuss und/oder Salzeffekte zu erklären sein könnte. Das Transkript der RT ist dagegen nur in kältegestressten *Anabaena flos-aquae* Kulturen nachweisbar. Um die Transkriptregulation der RT näher zu untersuchen, wurde die RT-PCR mit *Anabaena flos-aquae* Kulturen wiederholt, die für drei Stunden verschiedenen Stressbedingungen ausgesetzt waren (Abbildung 16). Das Experiment wurde im Duplikat durchgeführt und Salzstress (55 mM NaCl), oxidativer Stress (10 mM H₂O₂), Hitzestress (37°C), Dunkelstress und Kältestress (4 °C) getestet.

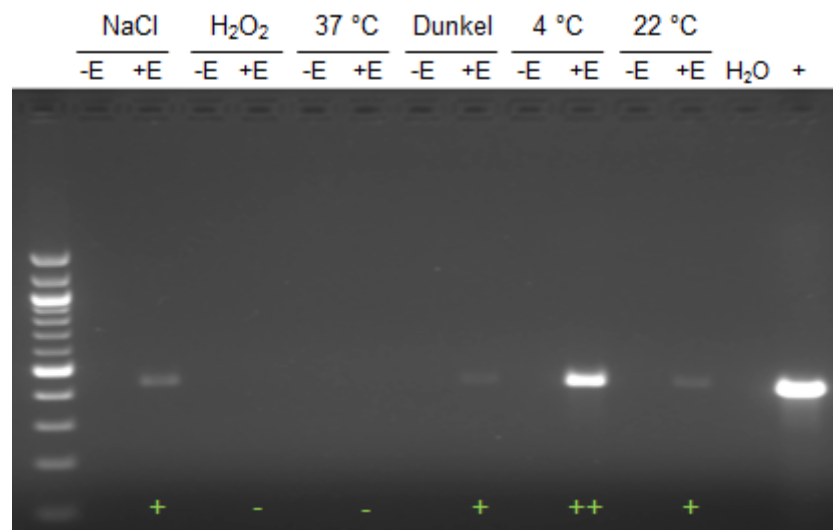


Abbildung 16: Transkriptregulation der RT aus *Anabaena flos-aquae*. *Anabaena flos-aquae* Kulturen wurden für drei Stunden verschiedenen Stressbedingungen ausgesetzt und die RT anschließend über Reverse-Transkriptase-PCR amplifiziert. Dabei wurde die Anzahl der Amplifikationszyklen erhöht (von 25 auf 30), um schwache Transkripte nachweisen zu können. Die Reaktion wurde mit (+E) und ohne (-E) Enzym (reverse Transkriptase) durchgeführt, um auf Präsenz von gDNA zu testen. Als Positivkontrolle (+) diente der pPM6. Stress durch H₂O₂ (10 mM) und Hitze (37 °C) führte zu keinem PCR-Produkt (-). Unter Dunkelstress, Salzstress (55 mM NaCl) und den Standardbedingungen (22 °C) konnte eine schwache Bande detektiert werden (+). Die Bande für die RT ist nach Kältestress (4°C) am intensivsten (++).

Die RT-PCR ist eine Form der Endpunkt-PCR. Diese eignet sich nicht für eine genaue Quantifizierung, denn gleiche DNA-Mengen zu Reaktionsbeginn können durch Variationen in der Reaktionskinetik zu unterschiedlichen DNA-Mengen in der gemessenen Plateauphase führen. Um die Bandenintensitäten in dem Experiment dennoch vergleichen zu können, wurde eine Ordinalskala mit den Variablen - (keine Bande), + (schwache Bande) und ++ (starke Bande) verwendet.

Nach oxidativem Stress und Hitzestress konnte kein Transkript von der RT nachgewiesen werden. Das könnte mit einem Absterben der Zellen zusammenhängen, welches beobachtet und durch zu intensive Stressbedingungen ausgelöst wurde. Für die Standardbedingungen, Salz- und Dunkelstress konnte ein schwaches Transkript nachgewiesen werden. Die stärkste Transkription wurde nach Kältestress detektiert, wie anhand der intensiven Bande zu erkennen ist.

Es konnten Bedingungen gefunden werden, unter denen die Templat RNA, die RT und die Avd in den kultivierten Cyanobakterien *Anabaena flos-aquae* und *Nostoc sp. PCC7120* transkribiert vorliegen. Als nächstes wurden die DGRs auf ihre Fähigkeit zur Hypermutation der VR untersucht.

3.2 Aktivitätsanalysen

Die DGRs aus den Cyanobakterien *Anabaena flos-aquae* und *Nostoc sp.* PCC7120, sowie aus dem humanpathogenen Bakterium *Treponema denticola*, wurden auf Aktivität untersucht. Unter Aktivität wird die Fähigkeit des DGRs zur Erzeugung von Variabilität in der VR verstanden. Um diese nachzuweisen, wurde mit Hilfe von Christine Förster-Schorr ein Restriktionsenzym-basierter Variabilitäts-Assay (REBVA) durchgeführt. Dieser wurde bereits zur Untersuchung der VR aus dem *Bordetella*-Phagen eingesetzt (Liu et al. 2002). Im REBVA wird die VR durch ein Restriktionsenzym geschnitten, wenn die Erkennungssequenz nicht mutiert vorliegt. Eine nachfolgende PCR amplifiziert nur die ungeschnittene VR, in der sich demnach Mutationen befinden.

Bei der Durchführung des REBVA wurden zwei generelle Probleme festgestellt. Zum einen ist die Effizienz des Restriktionsverdau nicht absolut, sodass alternierende Schritte aus Restriktionsverdau und PCR solange wiederholt wurden, bis keine restriktionsverdauten Banden mehr zu beobachten waren. Wurde das ungeschnittene PCR-Produkt sequenziert, obwohl restriktionsverdauten Banden zu beobachten waren, ergab die Sequenzierung in allen Fällen eine unmutierte VR-Sequenz. Das zweite Problem ist eine ausbleibende Amplifikation bei einer wiederholten PCR auf dasselbe PCR-Produkt unter Verwendung derselben Primer. Das Problem konnte durch eine nested PCR verbessert werden. Dabei wird eine unspezifische Primer-Bindung durch Verwendung eines weiteren, verschachtelten Primerpaares reduziert. Die verwendeten Primer, die Restriktionsenzyme mit den Erkennungssequenzen und die analysierten variablen Positionen sind in Kap. 2.2.2.4 aufgeführt.

Die DGRs wurden sowohl nativ, als auch heterolog exprimiert untersucht. Untersuchungen am nativen DGR bieten den Vorteil, dass alle benötigten Wirtsfaktoren vorhanden sind.

3.2.1 Das native DGR aus *Nostoc sp.* PCC7120 zeigt Aktivität

Bei der Untersuchung der Aktivität des nativen DGRs aus *Nostoc sp.* PCC7120 wurden Cyanobakterien verwendet, die seit zwei Jahren im Labor der Molekularen Genetik (TU Kaiserslautern) in Kultur gehalten wurden. Im REBVA wurde nur die VR von Zielgen2 untersucht, da für die VR von Zielgen1 kein geeignetes Restriktionsenzym gefunden wurde. Die Durchführung des REBVA gestaltete sich als schwierig und führte nur in einem von vielen Versuchen zu einem PCR-Produkt, das von dem Restriktionsenzym nicht weiter geschnitten wurde (Abbildung 17).

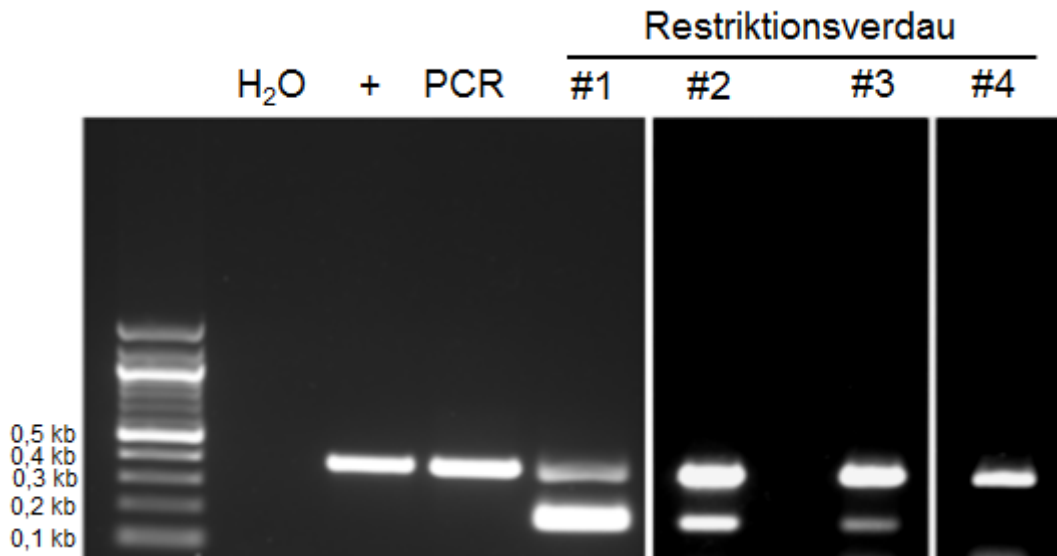


Abbildung 17: Ausschnitt des Restriktionsenzym-basierten Variabilitäts-Assays. Die variable Region am 3'-Ende von Zielgen2 des DGRs aus *Nostoc sp.* PCC7120 wurde abwechselnd amplifiziert (382 bp) und restriktionsverdaut (SacI), wodurch zwei Fragmente entstehen (190 bp und 192 bp). Abgebildet sind die erste PCR und vier Restriktionsverdaue. Die relative Menge der geschnittenen Fragmente nimmt über die einzelnen Schritte hinweg ab, bis diese nicht mehr zu beobachten sind (Restriktionsverdau #4). Bei der Positivkontrolle zur PCR (+), wurde das klonierte DGR (pPM7) als Templat verwendet.

Das PCR-Produkt, welches im vierten Restriktionsverdau keine geschnittenen Fragmente aufwies, wurde sequenziert. Dabei wurden zwei unterschiedliche Sequenzen beobachtet (Abbildung 18).

TR		AA AA AA	AAAA	AA AA AA AA A	69	
VR	CTGCGGGGCGGCTCCTG	GAGCTCCTC	TCTCTCTG	TAGTCTGCCGTTCCGCGTCCC	GGCAACAACGATAGG	69
R1		A A T	TT		69	
R2	-----				10	
TR	A	A AA A	AA AA AA A	T C	137	
VR	GCGGGGCGCGCT	TACCGCTACT	ACGCTG	TTGGTTTTTCGTGTTGTCTGCGCGTT	CGGGAGGACTTTTCA	137
R1					137	
R2	-----				70	

Abbildung 18: Variable Region des DGRs aus *Nostoc sp.* PCC7120. Die variable Region (VR) von Zielgen2 des DGRs aus *Nostoc sp.* PCC7120 ist gezeigt. In dem fast identischen Templat Repeat (TR) wurden übereinstimmende Nukleotide, bis auf Adenine, ausgeblendet. Adenine in dem TR können zu variablen Positionen in der VR führen (rot). Die Initiierungssequenz des mutagenen Homings (IMH, blau) verfügt über drei Adenine in dem korrespondierenden TR. Diese Positionen werden nicht als variabel angesehen, weshalb die Adenine in dem TR ausgeblendet sind. In der IMH sind zwei Adenin-unspezifische Abweichungen zu beobachten (unterstrichen). Die durch den REBVA erhaltenen Sequenzen R1 und R2 besitzen eine mutierte SacI-Erkennungssequenz (gelb hervorgehoben). **(R1):** Der Bereich der Mutationen umfasst 15 bp und beinhaltet fünf Mutationen an variablen Positionen. **(R2):** Die Erkennungssequenz wurde zusammen mit einem Bereich der VR deletiert (Minus).

Die VR-Sequenz, die in Abbildung 18 als R1 bezeichnet wird, scheint durch DGR-Aktivität verändert worden zu sein. Innerhalb eines 15 bp langen Bereichs, der die Erkennungssequenz des Restriktionsenzym mit einschließt, können fünf Adenin-spezifische Mutationen beobachtet werden. In der zweiten VR-Sequenz (R2) wurde die Erkennungssequenz des Restriktionsenzym durch eine großflächige Deletion von 67 bp entfernt.

3.2.2 Das native DGR aus *Anabaena flos-aquae* zeigt keine Aktivität

Die Untersuchung der Aktivität des nativen DGRs aus *Anabaena flos-aquae* wurde mit Cyanobakterien durchgeführt, die seit zwei Jahren im Labor der Molekularen Genetik (TU Kaiserslautern) kultiviert wurden. Diese wurden für den REBVA unter Standardbedingungen (22 °C) angezogen und für einen parallelen Versuch zusätzlich einem Kältestress (4 °C, 3 Stunden) ausgesetzt. Der Kältestress wurde durchgeführt um die Transkription der DGR-kodierten RT zu induzieren (Kap. 3.1.5.1).

Der REBVA für die unter Standardbedingungen angezogenen Zellen gestaltete sich als schwierig, da eine wiederholte Amplifikation der VR nicht möglich war. Dieses Problem konnte nicht durch eingerückte Primer behoben werden. Eine Möglichkeit mit der dieses Problem behoben werden konnte, lag in der Verwendung einer Polymerase ohne Korrekturlesefunktion (myTaq). Mit der myTaq konnte der REBVA durchgeführt werden, bis keine geschnittenen Fragmente mehr zu beobachten waren. Die Sequenzierung des PCR-Produkts ergab jedoch, dass die Erkennungssequenz des Restriktionsenzym durch eine Punktmutation verändert wurde (Daten nicht gezeigt). Eine DGR-Aktivität konnte nicht beobachtet werden.

Bei der kältegestressten *Anabaena flos-aquae* Kultur bestand ebenfalls das Problem, dass eine wiederholte Amplifikation der VR, auch mit eingerückten Primern, nicht möglich war. Anders als bei den unter Standardbedingungen angezogenen Zellen, wurde nicht die myTaq DNA-Polymerase verwendet, da die fehlende Korrekturlesefunktion vermutlich zu der Punktmutation im Selektionsbereich geführt hat. Der REBVA wurde mit der Phusion DNA-Polymerase (mit Korrekturlesefunktion) mehrmals wiederholt, konnte aber nicht zuende geführt werden.

Aufgrund der geringen Anzahl erfolgreich durchgeführter REBVA-Versuche konnte in den kultivierten Cyanobakterien nur ein Ereignis von DGR-Aktivität beobachtet werden (Abbildung 18). Bei der Durchführung gestaltete sich die Amplifikation des PCR-Produkts nach dem Restriktionsverdau als schwierig. Dies könnte daran liegen, dass die amplifizierte VR-Sequenz unmutiert vorlag und nach dem Restriktionsverdau nicht genügend Moleküle für die PCR zur Verfügung standen. Sollte die Hypermuation der VR durch einen unbekanntem äußeren Reiz ausgelöst werden, dann wären die konstanten Kultivierungsbedingungen eine schlechte Voraussetzung für die Analyse der Aktivität. Um weitere DGR-vermittelte Unterschiede in der VR feststellen zu können, wurde daher ein DGR untersucht, dessen variables Protein einem natürlichen Anpassungsdruck unterworfen ist.

3.2.3 Das native DGR aus *Treponema denticola* zeigt Variabilität

Im Jahr 2011 veröffentlichte Dr. Sigrun Eick, Leiterin des Labors Orale Mikrobiologie an der Klinik für Parodontologie (Bern), einen Artikel über den Vergleich zweier Methoden zur Diagnose von chronischer Parodontitis (Eick et al. 2011). Zu Beginn der Studie (t1), sowie 3, 6 und 12 Monate nach Ansetzen einer Behandlung (t2-4), wurden jeweils drei Proben (A-C) des subgingivalen Plaques aus 25 Patienten entnommen und mittels DNA-STRIP-Technologie und Real-Time-PCR untersucht. Die isolierte bakterielle DNA der Plaques von 53 Proben aus 15 Patienten der Studie wurde unserer Arbeitsgruppe zur Verfügung gestellt (Tabelle A2). Die Proben waren pseudonymisiert, ein Votum der Ethikkommission des Universitätsklinikums Jena zur bakterienspezifischen PCR-Analyse der Proben lag vor. In allen 53 Proben wurde *Treponema denticola* über die oben genannten Methoden nachgewiesen. *Treponema denticola* besitzt ein DGR, welches in Abbildung 4 skizziert ist. In dieser Arbeit wurden die Proben verwendet, um die VR von *Treponema denticola* zu untersuchen. Diese wurde mit Hilfe von Christine Förster-Schorr über PCR amplifiziert (Abbildung 19).

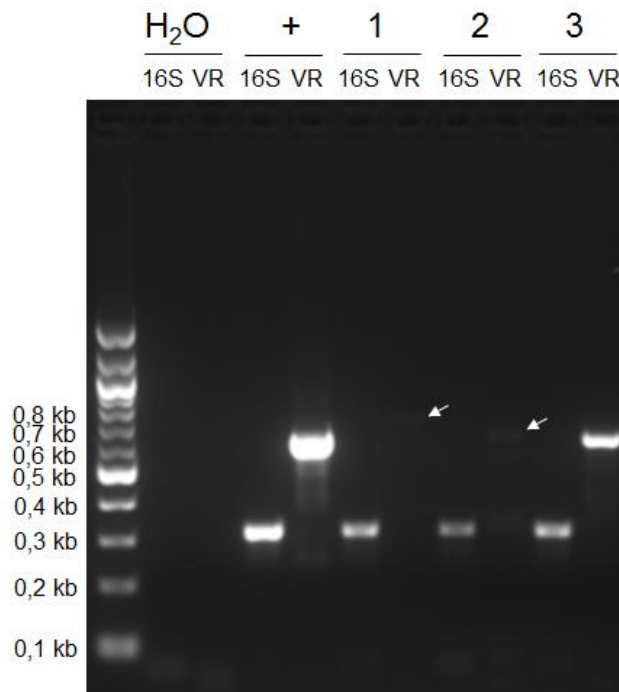


Abbildung 19: Nachweis der variablen Region aus *Treponema denticola*. Die variable Region (VR) (622 bp) wurde aus Patientenproben (Spur 1-3) amplifiziert. Um die Integrität der genomischen DNA zu überprüfen, wurde zusätzlich die 16S rDNA (316 bp) amplifiziert. Die schwache Bande in Spur 2 und die Bande in Spur 3 laufen auf Höhe der VR, während die schwache Bande in Spur 1 größer ist (~ 800 bp). Als Positivkontrolle (+) diente genomische DNA von *Treponema denticola* (Deutsche Sammlung von Mikroorganismen und Zellkulturen, #14222)

In Abbildung 19 ist zu sehen, dass die PCR-Produkte für die VR in Spur 2 und 3, auf Höhe der Positivkontrolle (622 bp) laufen. Eine anschließende Sequenzierung der PCR-Produkte bestätigte für Spur 2 (Patient 2, Probe A, t1) und 3 (Patient 5, Probe A, t3), dass es sich um

die VR aus *Treponema denticola* handelt. Das PCR-Produkt in Spur 1 ist größer (~ 800 bp) als die Positivkontrolle. Es wurde sequenziert, wobei keine Homologie zu der VR aus *Treponema denticola* beobachtet werden konnte. Ein Nachweis der VR konnte für 41 der untersuchten 53 Proben nicht erbracht werden. Da die PCR sensibel genug ist, um einzelne DNA-Moleküle nachzuweisen, könnte die DNA im Laufe der Zeit (mind. 8 Monate seit Entnahme) oder beim Transport Schaden genommen haben. Die 12 Proben, in denen die VR nachgewiesen wurde, gehören zu fünf Patienten und sind in Tabelle 10 zusammengefasst.

Tabelle 10: Proben von fünf Patienten mit amplifizierbarer VR

Patient	Probe	Zeitpunkt	Unterschied zur Datenbank
1	C	t1	Nein
2	A	t1, t2	Nein
	B	t1, t2, t4	
3	A	t1	Nein
	B	t3	
4	C	t1	Nein
5	A	t3	Ja
	B	t1	
	C	t1	

Die sequenzierten *Treponema*-Sequenzen von den Patienten 1-4 wiesen keine Variabilität auf, sodass alle Proben zu VR-Sequenzen führten, die identisch mit der Datenbank sind.

Die VRs der *Treponema*-Sequenzen von den drei Proben aus Patient 5 (VR-5) sind untereinander identisch, unterscheiden sich aber an 29 Positionen von der VR-Sequenz der Datenbank (VR-D) (Abbildung 20). Um die Unterschiede näher zu bestimmen, wurde auch der TR von Patient 5 sequenziert (TR-5). Dabei korrespondieren 25 variable Positionen (rot) in der VR-5 mit Adeninen in dem TR-5. Diese Adenin-spezifischen Mutationen sprechen für DGR-Aktivität. Ebenfalls für eine DGR-Aktivität spricht Position 32 (grün). Dabei handelt es sich nicht um eine Mutation in der VR-5, sondern um eine Mutation in dem TR-5 (Cytosin zu Thymin), der zur VR-5 transferiert wurde. Die anderen drei Unterschiede zur VR-D sind Adenin-unspezifisch (unterstrichen). Diese könnten durch einen fehlerhaften Einbau der DGR-kodierten RT entstanden sein. Die Fehlerrate von RTs liegt bei etwa einem falschen Einbau pro 1700 bis 30.000 Nukleotiden (Roberts, Bebenek & Kunkel 1988). Bei drei Mutationen innerhalb der 136 bp langen VR, könnte neben einer fehlerhaften Erkennung von Adeninen, auch eine generell erhöhte Fehlerrate für die DGR-kodierte RT vermutet werden.

TR-5		A	AA	A	AAA	T	A	AA	AA	AA	AA	AA	A	A	68
VR-D	CCGCGTCAGGCTCT	GGCCG	TGTTT	TACGCGG	GGCAGCTGG	GCCGG	CAGCGCG	GACTACTGCG	CTGTA						68
VR-5	<u>A</u>	G	AA	GC	T		A	<u>TAT</u> TA	AT	A	A				68
TR-5	AAA	AA	AA	AA	A	A	AA	A	AA	AA	AA		T	T	136
VR-D	GGCGAACGGGT	CAACAT	CAGTCCT	GGCGT	CAGGT	GCAGCG	ATCTTGGCT	TCCGCC	TGGCTT	GCCGGCC					136
VR-5	TCT	A	G	TA		A	TA	A	<u>G</u>						136

Abbildung 20: Variable Region des DGRs aus *Treponema denticola*. Die variable Region des DGRs von *Treponema denticola* aus der Datenbank (VR-D) und der Probe aus Patient 5 (VR-5) ist gezeigt. Bei dem fast identischen Templat Repeat aus Patient 5 (TR-5) wurden zur VR-D übereinstimmende Nukleotide (bis auf Adenine) ausgeblendet. Adenine im TR-5 können zu variablen Positionen in der variablen Region führen (rot). Adenin-unspezifische Abweichungen sind unterstrichen. **(VR-5):** In der VR-5 sind die Unterschiede zur VR-D abgebildet. Es liegen mindestens 25 Adenin-spezifische, sowie drei unspezifische Mutationen vor, von denen eine in der Initiierungssequenz des mutagenen Homings (IMH, blau) vorkommt. Der Unterschied zur VR-D an Position 32 (grün) ist durch eine transferierte Mutation in dem TR-5 entstanden.

3.2.4 Heterolog exprimierte DGRs zeigen keine Aktivität

In dem kultivierten Cyanobakterium *Nostoc sp.* PCC7120 und dem natürlich vorkommenden Humanpathogen *Treponema denticola* konnte eine durch das DGR diversifizierte VR-Sequenz beobachtet werden. In einem weiteren Experiment sollte getestet werden, ob DGRs auf einen anderen Wirt transferierbar sind und inwiefern eine Diversifizierung des Zielgens im heterologen System möglich ist.

Dafür wurden die DGRs aus *Anabaena flos-aquae*, *Nostoc sp.* PCC7120 und *Treponema denticola* in den Vektor pBluescript kloniert (Abbildung A1). Dort stehen die Elemente unter Kontrolle des T7-Promotors. Anschließend wurden die Plasmide in einen *E. coli*-Expressionsstamm transformiert, der die T7-Polymerase unter IPTG-Zugabe exprimiert. Dadurch ist es möglich die Expression des DGRs durch IPTG-Zugabe zu induzieren. Für die Expression wurden die Stämme BL21(DE3) und T7 Express verwendet. Die transformierten Zellen wurden sowohl ohne, als auch mit IPTG (1mM) bis zu einer Woche in Kultur gehalten und dabei täglich Proben entnommen. Aus den Proben wurden Plasmide isoliert und restriktionsverdaut. Das entstehende Bandenmuster wurde mit dem theoretischen verglichen. Dadurch konnte die Stabilität der Plasmide in den Zellen beobachtet und dem DGR Zeit zur Diversifizierung des Zielgens gegeben werden. Ein Auszug aus diesem Stabilitätstest ist für die klonierten DGRs aus *Anabaena flos-aquae* und *Nostoc sp.* PCC7120 in Abbildung 21 dargestellt.

In Abbildung 21 ist zu sehen, dass die restriktionsverdauten Plasmide mit dem DGR aus *Anabaena flos-aquae* am zweiten Tag nach der Transformation in T7 Express Zellen, ein von der Kontrolle abweichendes Bandenmuster aufweisen. Diese Fragmentierung kann sowohl unter induzierten, als auch uninduzierten Bedingungen beobachtet werden und tritt ebenfalls in BL21(DE3) Zellen und dem Klonierungsstamm DH5α auf (Daten nicht gezeigt).

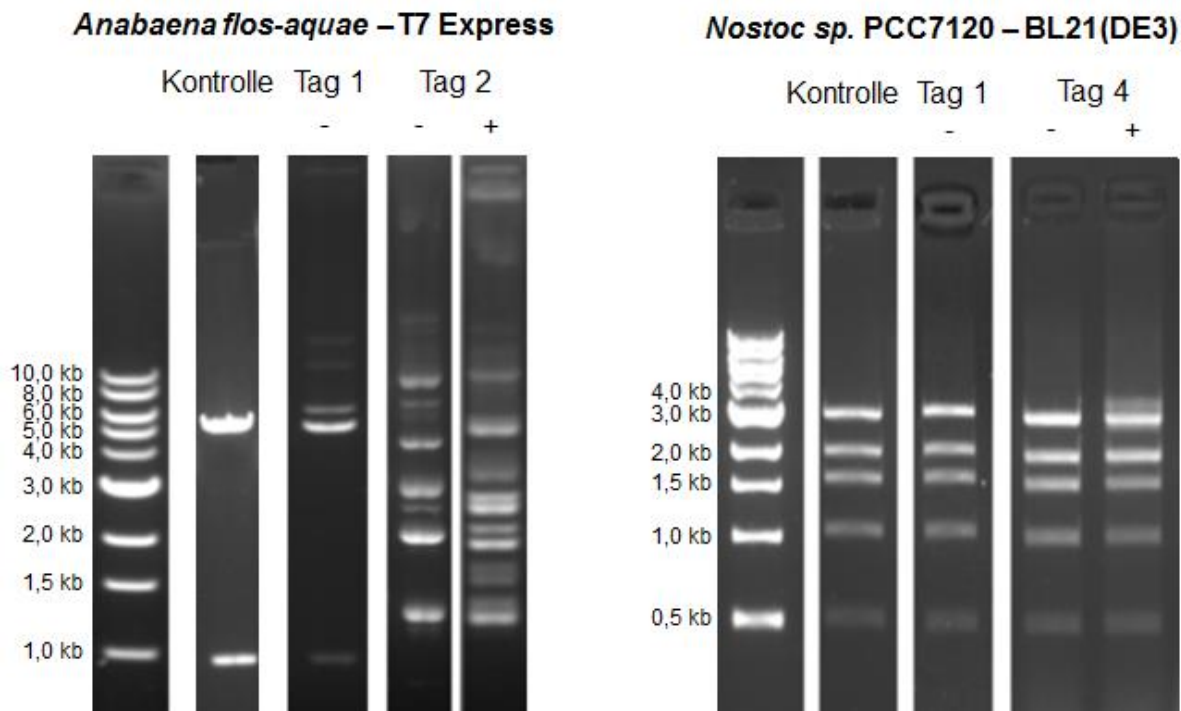


Abbildung 21: Stabilitätstest heterolog exprimierter DGRs. Die klonierten DGRs wurden in Expressionsstämme transformiert und ohne (-) bzw. mit (+) 1 mM IPTG mehrere Tage in Kultur gehalten. Der Restriktionsverdau (PvuI) isolierter Plasmide ist gezeigt. Das Bandenmuster kann mit einer Kontrolle verglichen werden. Das DGR aus *Anabaena flos-aquae* zeigt am ersten Tag nach der Transformation das korrekte Bandenmuster (5311 bp und 1045 bp), fragmentiert aber am zweiten Tag mit und ohne IPTG. Das DGR aus *Nostoc sp. PCC7120* zeigt auch am vierten Tag nach der Transformation das korrekte Bandenmuster (512 bp, 1045 bp, 1576 bp, 1974 bp und 2831 bp).

Das klonierte DGR aus *Nostoc sp. PCC7120* ist in den verwendeten BL21(DE3) Zellen auch am vierten Tag nach der Transformation stabil. Um eventuelle Veränderungen der DNA-Sequenz zu analysieren, die keine Auswirkungen auf das Bandenmuster haben, wurde die Plasmidisolierung vom letzten Tag des Stabilitätstests (Tag 4) sequenziert. Die erhaltenen Sequenzen liegen ohne Mutationen vor und entsprechen denen der Datenbank.

Das klonierte DGR aus *Treponema denticola* ist ebenfalls über vier Tage stabil in BL21(DE3) Zellen (Daten nicht gezeigt). Die Plasmide wurden mit dem Restriktionsenzym PvuI geschnitten. Die dadurch entstehenden Restriktionsfragmente (3858 bp, 1893 bp und 1354 bp) konnten für die Plasmidisolierungen beobachtet werden. Eine Sequenzierung der Plasmidisolierung vom letzten Tag des Stabilitätstests wurde nicht durchgeführt.

Der Stabilitätstest wurde für die getesteten DGRs dreimal wiederholt und die Plasmide anschließend im REBVA verwendet (Kap. 3.2). Dabei wurden stabile und fragmentierte Plasmide eingesetzt, solange die VR zu amplifizieren war. Sequenzierte PCR-Produkte des REBVA wiesen in allen VRs eine Veränderung der Restriktionsenzymkennungssequenz auf (Daten nicht gezeigt). Dabei handelt es sich um einzelne Punktmutationen, die nicht mit einer DGR-Aktivität zu assoziieren sind. Wahrscheinlicher ist ein fehlerhafter Einbau durch die Polymerase während der PCR (Fehlerrate: $4,4 \times 10^{-7}$).

3.3 Strukturanalyse der Templat RNA aus *Nostoc sp.* PCC7120

DGRs verwenden einen gerichteten Sequenztransfer und eine wahrscheinlich fehleranfällige RT, um einen definierten Genbereich zu hypermutieren. Bei dem mobilen Element, welches den Sequenztransfer vermittelt, handelt es sich um die Templat RNA. Für die Templat RNA wurde eine mögliche Endonuklease-Funktion im mutagenen Homing postuliert (Guo et al. 2011). Auch wäre, analog zu den Gruppe II Introns, eine hochaffine Bindung zu der RT denkbar, welche eine definierte Substraterkennung ermöglicht. In beiden Fällen ist die ausgebildete Struktur der Templat RNA entscheidend für die Funktion. Die Struktur der Templat RNA wurde jedoch bislang für kein bekanntes DGR aufgeklärt.

Daher wurde das DGR aus dem Cyanobakterium *Nostoc sp.* PCC7120 verwendet, um die Struktur der Templat RNA aufzuklären. Für dieses DGR konnte Aktivität nachgewiesen werden (Kap. 3.2.1), was die Funktionsfähigkeit der Templat RNA bestätigt. Damit die Struktur aufgeklärt werden kann, muss zunächst die Primärsequenz bekannt sein, aus der sich die Sekundärstruktur über intramolekulare Basenpaarungen bildet.

3.3.1 Identifizierung des Haupttranskripts der Templat RNA

Die Primärstruktur der RNA beginnt am Transkriptionsstart (TSS) und endet am Transkriptionsstopp (STOPP). Um TSS und STOPP der Templat RNA aus *Nostoc sp.* PCC7120 zu identifizieren, wurden RACE (rapid amplification of cDNA-ends)-Experimente im Triplikat durchgeführt. Bei RACE-Experimenten wird eine bekannte Sequenz an die RNA-Enden angefügt (5'-Ende: RNA-Adapter; 3'-Ende: Poly(A)-Schwanz), die nach einer cDNA-Synthese als Primerbindestelle fungieren kann. Da beide Enden in unterschiedlichen Experimenten analysiert wurden, konnte der jeweils entgegengesetzte Primer basierend auf der bekannten TR-Sequenz generiert werden. Durch eine PCR können somit auch die unbekanntes Transkriptenden amplifiziert und anschließend sequenziert werden.

Die RNA der RACE-Experimente wurde mit dem Enzym „Terminator 5'-Phosphate-Dependent Exonuclease“ (TEX) vorbehandelt. Dieses Enzym baut RNA mit einem 5'-Monophosphat ab. Dies betrifft die rRNA und andere prozessierte Transkripte. Da bakterielle Primärtranskripte über ein 5'-Triphosphat verfügen, stellen sie kein Substrat für die TEX dar und können über diese Methode angereichert werden. Getestet wurde die Aktivität der TEX an der rRNA aus *Nostoc sp.* PCC7120, die in der Zelle prozessiert wird und dadurch ein angreifbares 5'-Monophosphat besitzt (Abbildung 22).

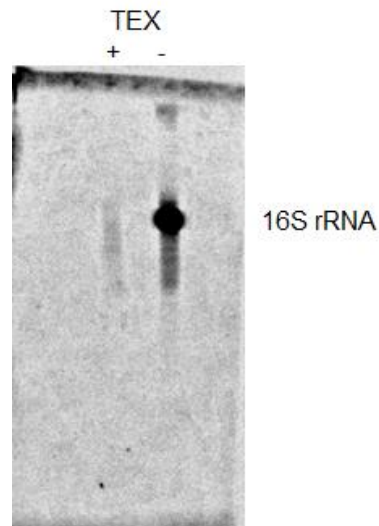


Abbildung 22: Kontrolle der TEX-Aktivität. Gesamt-RNA aus *Nostoc sp.* PCC7120 wurde einer TEX (Terminator 5'-Phosphate-Dependent Exonuclease)-Behandlung unterzogen. Diese RNA wurde in Northern-Blot-Experimenten verwendet, um die 16S rRNA nachzuweisen. Der einzige Unterschied zwischen TEX + und TEX - ist die Zugabe des Enzyms. Die prozessierte rRNA stellt ein Substrat für die TEX dar und wird abgebaut.

In *Nostoc sp.* PCC7120 wurde bereits ein genomweites TSS-Mapping durchgeführt (Mitschke et al. 2011). Unter den identifizierten TSS befindet sich auch ein potentieller TSS für den TR. Die Position des TSS im Genom liegt innerhalb des RT-Gens (*alr3497*). Mit der von Thomas Kafka durchgeführten 5'-RACE konnte dieser TSS ebenfalls identifiziert werden (TSS1) (Abbildung 23A). Zusätzlich wurde ein weiterer TSS gefunden (TSS2), dessen Bande in der PCR intensiver ist, als die des TSS1. Das spricht für einen höheren Anteil dieser RNA-Spezies. In der 3'-RACE wurden ebenfalls zwei STOPPs identifiziert (Abbildung 23A). Bei der Untersuchung der STOPPs wurde ein Bereich sequenziert, der sich aus dem 3'-Ende der Templat RNA und dem künstlich angefügten Poly(A)-Schwanz zusammensetzt. Für STOPP2 sind die ersten drei Adenine des Poly(A)-Schwanzes jedoch auch an der entsprechenden Position der DNA-Sequenz vorzufinden und könnten damit sowohl zu dem Poly(A)-Schwanz, als auch zu der Templat RNA gehören. Für weitere Experimente wurden die Adenine zu der Templat RNA hinzugezählt. Die identifizierten Transkriptenden befinden sich im Chromosom von *Nostoc sp.* PCC7120 an Position 4212285 (TSS1), 4212361 (TSS2), 4212612 (STOPP1) und 4212695 (STOPP2).

Ausgehend von den vier potentiellen Transkriptenden, ergeben sich die vier möglichen Transkripte TR1-TR4 (Abbildung 23B). Die Sequenzen dieser Transkripte wurden für eine *in vitro* Transkription (IVT) kloniert. Dabei wurde die Klonierung von TR1 und TR3 erst später von Kathrin Haag übernommen. Anschließend wurden sämtliche *in vitro* Transkripte von ihr auf durchgehende Transkription untersucht (Abbildung A3). Bei der IVT entstehen RNA-Moleküle, die um 5-7 Nukleotide länger sind als das zu untersuchende Transkript. Die zusätzlichen Nukleotide stammen von der Erkennungssequenz des verwendeten Restriktionsenzym bei der „run-off“ Transkription (Kap. 2.2.3.4.1).

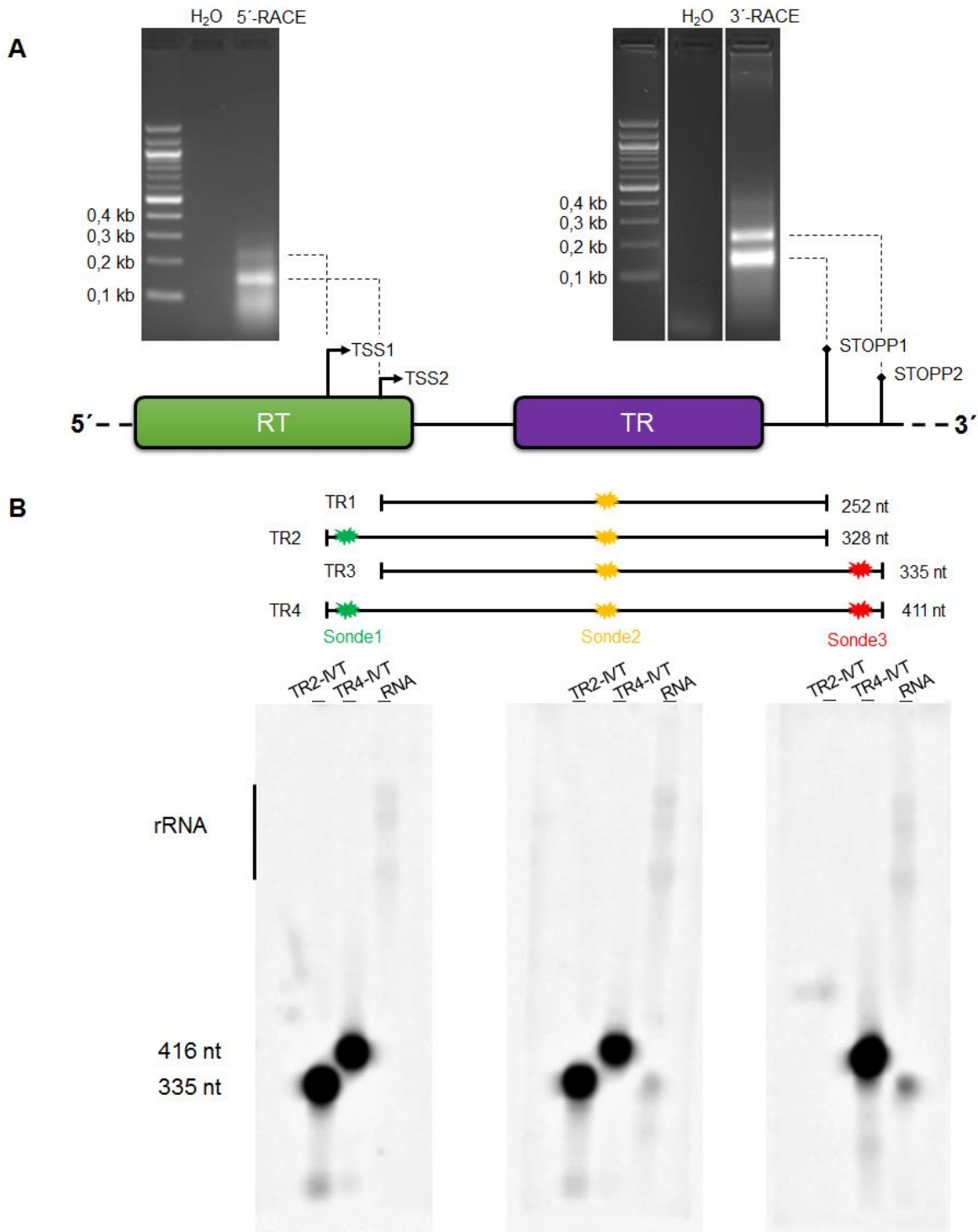


Abbildung 23: Identifizierung der Transkriptenden der Templat RNA über Northern-Blot-Experimente. (A): Eine PCR auf die modifizierte cDNA des TR (s. Text) führt zur Identifizierung zweier Transkriptionsstarts (TSS) bei der 5'-RACE (links) und zweier Transkriptionsstopps (STOPP) bei der 3'-RACE (rechts). **(B):** Durch die Identifizierung der zwei TSS und STOPPs ergeben sich vier potentielle Transkripte (TR1-4). Um zu testen, welches Transkript *in vivo* verwendet wird, wurden Sonden für Northern-Blot-Experimente erstellt. Die Hybridisierungsorte der Sonden auf den Transkripten sind markiert. In den Northern-Blot-Experimenten wurde Gesamt-RNA aus *Nostoc sp.* PCC7120, sowie die *in vitro* Transkripte TR2 und TR4, mit den vorgestellten Sonden hybridisiert. Das *in vivo* Transkript läuft unterhalb der TR4-IVT (416 nt) auf Höhe der TR2-IVT (335 nt), wird aber nicht von Sonde1 gebunden. Zusammen mit der intensiven Bande des TSS2 bei der PCR, spricht dies für eine Templat RNA in Form der TR3 (335 nt). Mit Sonde2 kann ein weiteres, schwaches Signal unterhalb des potentiellen TR3-Signals detektiert werden.

Die *in vitro* Transkripte wurden als Hybridisierungskontrolle und Größenstandards in Northern-Blot-Experimenten eingesetzt. In den Northern-Blot-Experimenten wurde die *in vivo* vorkommende Templat RNA über verschiedene Oligonukleotid-Sonden in Gesamt-RNA nachgewiesen (Abbildung 23B). Die Northern-Blot-Experimente wurden mit unabhängigen RNA-Präparationen im Triplikate durchgeführt. Dabei wurden sämtliche nicht-radioaktive Schritte von Thomas Kafka übernommen. Die verschiedenen Sonden wurden mit gleicher Konzentration an Radioaktivität eingesetzt, um gleiche Mengen RNA zu hybridisieren. Zu den Sonden gehört eine Positivkontrolle, welche alle der möglichen vier Transkripte binden kann (Sonde2). Diese Sonde detektiert ein Transkript in der Gesamt-RNA, welches ungefähr auf Höhe der TR2-IVT läuft (335 nt). Die STOPP2-spezifische Sonde (Sonde3), kann ebenfalls ein Transkript auf dieser Höhe detektieren. Dagegen kann durch die TSS1-spezifische Sonde (Sonde1) kein Transkript in der Gesamt-RNA nachgewiesen werden. Diese Ergebnisse sprechen für eine Templat RNA in Form der TR3, die bei TSS2 beginnt und bei STOPP2 endet (335 nt). Mit Sonde2 kann zudem ein schwaches Signal unterhalb des potentiellen TR3-Signals detektiert werden, bei dem es sich um die TR1-Variante handeln könnte (252 nt). Da die TR1-IVT und die TR3-IVT zu diesem Zeitpunkt noch nicht verfügbar waren, konnte ein direkter Vergleich nicht erfolgen.

Die RNA in Abbildung 23B ist nicht sonderlich stark aufgetrennt. Daher ist eine genaue Größenbestimmung der Templat RNA nicht möglich. Um die abgeschätzte Größe zu bestätigen, wurden weitere Northern-Blot-Experimente mit stärker aufgetrennter RNA durchgeführt (Abbildung 24). Die dafür verwendete Sonde4 wurde über „random primed labeling“ generiert und kann alle vier TR-Transkripte binden. Um die Spezifität der Sonde zu testen, wurde diese mit Gesamt-RNA aus *Anabaena flos-aquae* hybridisiert (Abbildung 24A). Die Sequenzen der Templat RNA aus *Anabaena flos-aquae* und *Nostoc sp.* PCC7120 weichen voneinander ab. Daher sollte in *Anabaena flos-aquae* kein Signal detektiert werden. Ein Signal würde für eine Kreuzhybridisierung an ein anderes, eventuell abundantes Transkript sprechen und könnte auch in dem verwandten Cyanobakterium *Nostoc sp.* PCC7120 zur Detektion geführt haben. Es konnte jedoch keine Kreuzhybridisierung beobachtet werden. Im selben Northern-Blot wird der potentielle TR3 in *Nostoc sp.* PCC7120 nachgewiesen. Die Größe des Signals ordnet sich durch Vergleich mit einem ssRNA Größenmarker auf 300-500 nt ein. Dieser wurde verwendet, um einen alternativen Größenvergleich zu ermöglichen. Denn trotz des denaturierenden Gels, könnten sich stabile Sekundärstrukturen in der Templat RNA/IVT ausbilden und zu einem veränderten Laufverhalten führen. Der potentielle TR3 (335 nt) wird auch in Abbildung 24B detektiert und läuft gut erkennbar auf Höhe der TR2-IVT (335 nt). Darunter kann erneut das schwächere Signal detektiert werden, bei dem es sich um die TR1-Variante handeln könnte. Auch für diesen Versuch standen die TR1-IVT und die TR3-IVT nicht zur Verfügung und konnten daher nicht direkt verglichen werden.

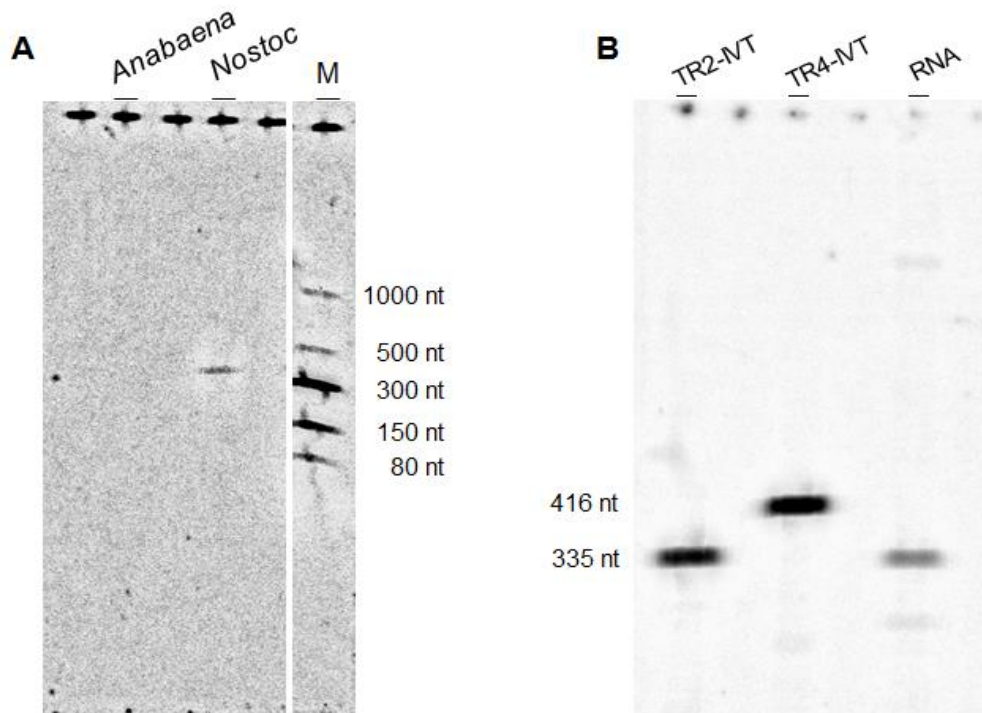


Abbildung 24: Northern-Blot-Experiment zur Größenbestimmung der Templat RNA. (A): In *Nostoc sp.* PCC7120 kann ein Signal für die Templat RNA detektiert werden, dessen Größe im Vergleich mit einem ssRNA Größenmarker (M), auf 300-500 nt einzuordnen ist. Zusätzlich wurde die gleiche Menge RNA aus *Anabaena flos-aquae* aufgetragen, um eine Kreuzhybridisierung der Sonde zu untersuchen **(B):** Die Templat RNA aus *Nostoc sp.* PCC7120 läuft auf Höhe der TR2-IVT (335 nt). Darunter befindet sich ein schwächeres Signal, bei dem es sich um die TR1-Variante handeln könnte.

Die aufgestellte Hypothese, das Haupttranskript des TR würde von TSS2 bis STOPP2 verlaufen (TR3), wurde mittels RT-PCR überprüft. Die RT-PCR liefert im Gegensatz zu den Northern-Blot-Experimenten einen indirekten Nachweis, da nicht die RNA, sondern die umgeschriebene cDNA detektiert wird. Durch den Amplifikationsschritt besitzt die RT-PCR aber eine höhere Sensitivität. Die RT-PCR wurde von Katrin Haag mit sechs unabhängigen RNA-Präparationen durchgeführt und ist in Abbildung 25 veranschaulicht.

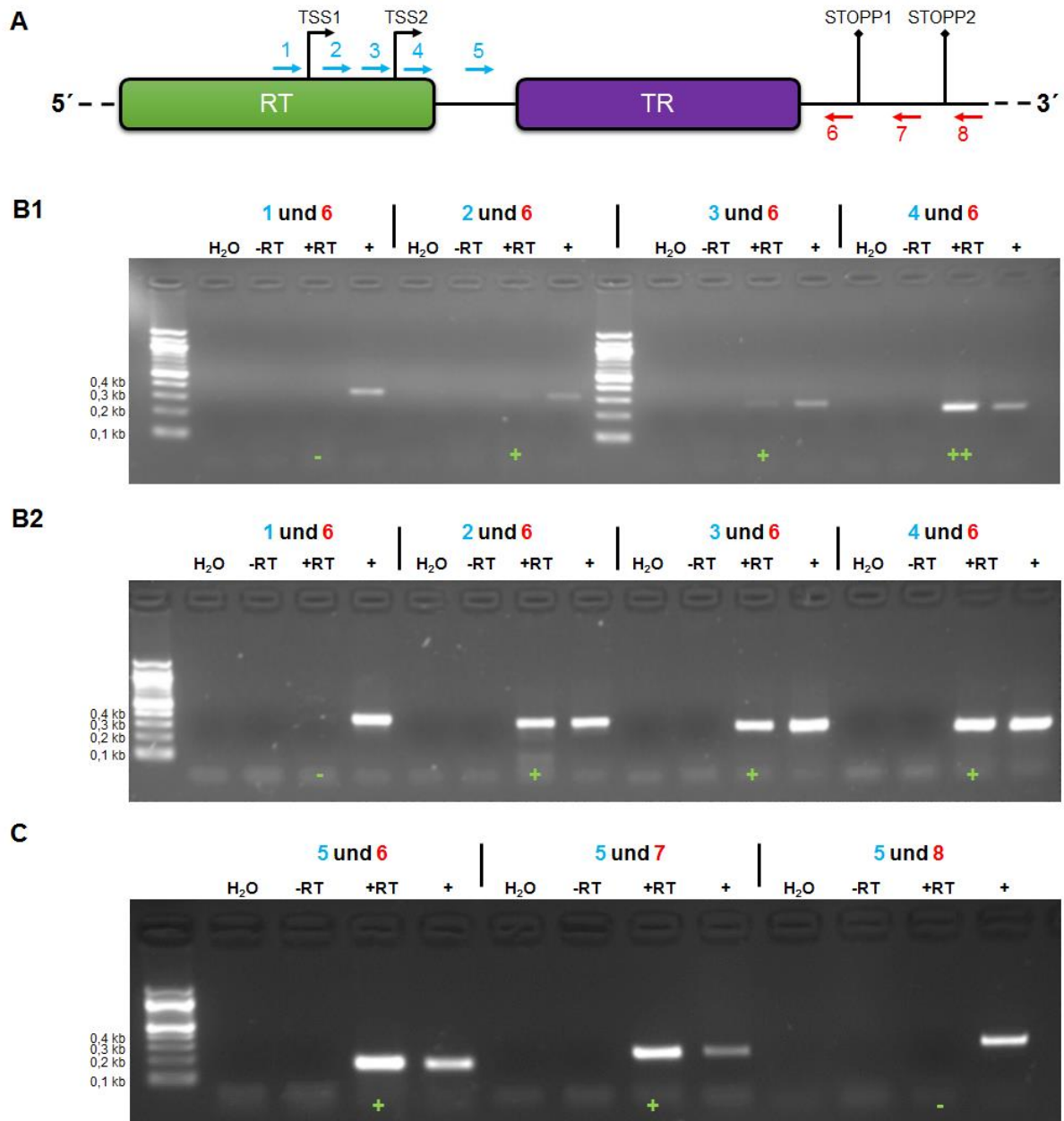


Abbildung 25: Identifizierung der Transkriptenden der Templat RNA über RT-PCR. Die Templat RNA aus *Nostoc sp.* PCC7120 wurde mittels RT-PCR aus Gesamt-RNA amplifiziert. **(A)**: Übersicht der verwendeten Primer. Primer 1-5 (blau) spezifizieren den Transkriptionsstart (TSS) und binden auf dem Minus-Einzelstrangabschnitt. Die Primer 6-8 (rot) binden den Plus-Einzelstrangabschnitt und dienen der Untersuchung des Transkriptionsstopps (STOPP). **B** und **C** zeigen die PCR-Produkte der RT-PCR. Als gDNA-Kontrolle wurde eine Reaktion ohne RT-Enzym verwendet (-RT). Für die Positivkontrolle (+) wurde das Plasmid pPM7 eingesetzt. Die relativen Bandenintensitäten der RT-Reaktion (+RT) sind angegeben: - (keine Bande), + (schwache Bande), ++ (starke Bande). **(B1)**: Untersuchung des TSS. Es kann kein Transkript beobachtet werden, das upstream von TSS1 beginnt (1 und 6). Für die Verwendung von TSS1 sprechen die Banden von 2 und 6 bzw. 3 und 6 (TSS1-Banden). Die Intensitäten der TSS1-Banden sind schwächer als die der TSS2-Bande (4 und 6). **(B2)**: Teilweise konnten ähnliche Bandenintensitäten für TSS1 und TSS2 beobachtet werden. **(C)**: Untersuchung des STOPP. Die Banden, welche für die Verwendung von STOPP1 (5 und 6) bzw. STOPP2 (5 und 7) sprechen, sind ähnlich intensiv. Es wird kein Transkript detektiert, das downstream von TSS2 endet (5 und 8).

Die Bandenintensitäten in der RT-PCR wurden gemessen und der Wert der RT-Reaktion auf den korrespondierenden Wert der Positivkontrolle bezogen. Dadurch wurden unterschiedliche Bindungsaffinitäten zwischen verschiedenen Primerpaaren ausgeglichen. Die resultierenden Intensitäten eines Versuchs wurden miteinander verglichen und sind als Ordinal-Variablen dargestellt: - (keine Bande), + (schwache Bande) und ++ (starke Bande). Der Rang einer Variablen im Kontext mit den Primern, erlaubt Rückschlüsse auf den verwendeten TSS bzw. STOPP der Templat RNA.

Das Primerpaar 1/6 wurde benutzt, um einen TSS upstream von TSS1 analysieren zu können. Da kein PCR-Produkt für die Primerkombination amplifiziert wurde, ist dies nicht der Fall (Abbildung 25B1 und B2). Die Primerpaare 2/6 und 3/6 ergeben bei Verwendung von TSS1 ein PCR-Produkt (TSS1-Banden). Ein PCR-Produkt mit Primerpaar 4/6, bei gleichzeitigem Fehlen der TSS1-Banden, spricht für die Verwendung des TSS2 (TSS2-Bande). In allen analysierten RNA-Präparationen konnten die TSS1-Banden detektiert werden. In der Hälfte der RNA-Präparationen ist die Intensität der TSS1-Banden nur schwach, während die Intensität der TSS2-Bande stark ist (Abbildung 25B1). Das deutet auf eine Verwendung von TSS2 hin. Die andere Hälfte der RNA-Präparationen führt zu ähnlichen Bandenintensitäten für den TSS1 und TSS2 (Abbildung 25B2). Das spricht für eine Verwendung des TSS1.

Der Transkriptionsstopp wurde in Abbildung 25C analysiert. Ein PCR-Produkt mit Primerpaar 5/7 spricht für die Verwendung von STOPP2 (STOPP2-Bande). Bei Verwendung von STOPP1 sollte keine STOPP2-Bande, dafür aber die STOPP1-Bande (Primerpaar 5/6) amplifiziert werden. In den Versuchen konnte sowohl die STOPP1- als auch die STOPP2-Bande amplifiziert werden. Um zu testen, ob der STOPP downstream von STOPP2 lokalisiert ist, wurde Primerpaar 5/8 verwendet. Für dieses Primerpaar ist in Abbildung 25C kein Transkript zu beobachten. In vier RNA-Präparationen konnte allerdings eine schwache Bande mit Primerpaar 5/8 amplifiziert werden (Daten nicht gezeigt). Dies könnte durch einen undichten Transkriptionsterminator zu erklären sein, wie es bereits für den Hepatitis B Virus beschrieben wurde (Guo et al. 1991). Die Analyse des Transkriptionsstopps spricht für eine Verwendung von STOPP2.

Während die Ergebnisse der RT-PCR eindeutig für eine Verwendung von STOPP2 sprechen, konnten beide TSS nachgewiesen werden. In den vorigen Northern-Blot-Experimenten konnte der TSS1 nicht detektiert werden. Daher wurden die RNA-Präparationen, die innerhalb der RT-PCR für den TSS1 gesprochen haben, für einen weiteren Northern-Blot verwendet. Auch in diesem Fall konnte mit der TSS1-spezifischen Sonde kein Transkript detektiert werden. Eine Erhöhung der Expositionszeit führte auch zu einer Verstärkung unspezifischer Signale (rRNA), wodurch weitere mögliche Banden überdeckt wurden (Daten nicht gezeigt). Um dies zu umgehen, wurden die RNA-Präparationen mit TEX behandelt. Durch TEX-behandelte RNA

kann zudem die Fragestellung beantwortet werden, ob es sich bei TSS2 um einen realen TSS handelt, oder um eine prozessierte RNA. Für letzteres würde sprechen, dass Mitschke und Kollegen den TSS2 nicht beobachten konnten (Mitschke et al. 2011). In Abbildung 26 ist der Northern-Blot mit TEX-behandelter RNA zu sehen. Dabei konnte mit der TSS1-spezifischen Sonde1 ein schwaches Signal auf Höhe der TR4-IVT (416 nt) detektiert werden. Dieses Signal ist viel schwächer als das durch Sonde2 und Sonde3 detektierte Signal, welches dem TR3 (335 nt) entspricht. Das deutliche Signal für den TR3 in TEX-behandelter RNA zeigt, dass es sich bei TSS2 um einen richtigen TSS handelt. Das Signal von Sonde1 spricht dafür, dass TSS1 ebenfalls ein richtiger TSS ist. Das schwache Signal lässt jedoch auf einen TSS schließen, der unter den verwendeten Kultivierungsbedingungen viel seltener verwendet wird als der TSS2.

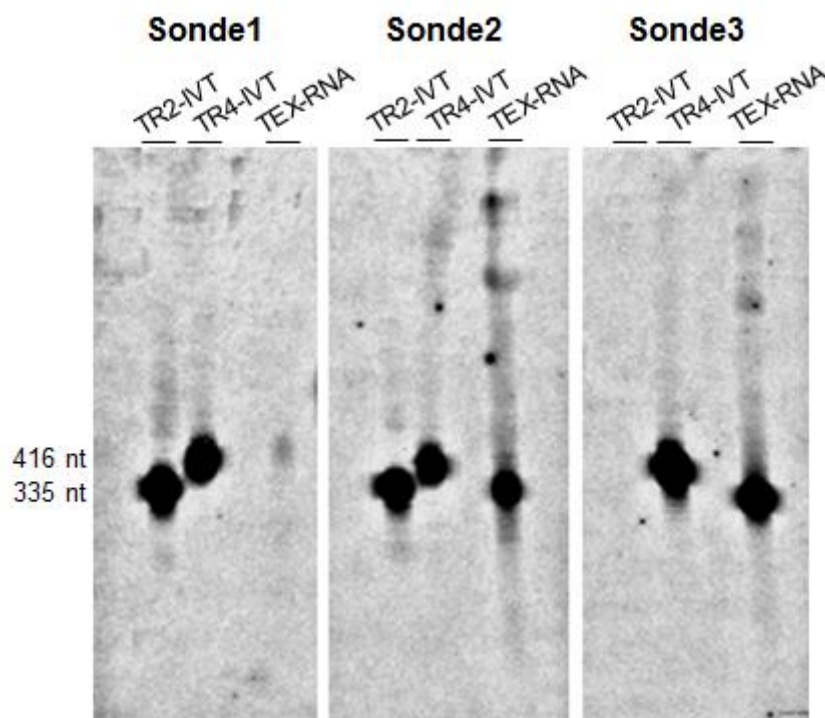


Abbildung 26: Detektion des TR4-Transkripts in TEX-behandelter Gesamt-RNA. Zur Abschätzung der Größe wurden die TR2- und TR4-IVT (335 nt bzw. 416 nt) zusammen mit der TEX-behandelten Gesamt-RNA aufgetragen. Das TR4-Transkript beginnt bei TSS1 und endet bei STOPP2, wodurch es eine Länge von 411 nt besitzt. Die für den TSS1 spezifische Sonde1 kann dieses seltene Transkript in TEX-behandelter RNA detektieren. Das TR3-Transkript, welches bei TSS2 beginnt und bei STOPP2 endet, wird auch in TEX-behandelter RNA nachgewiesen.

Zusammengenommen ergibt sich aus den Ergebnissen der Northern-Blot-Experimente und RT-PCR-Analysen ein Auftreten verschiedener TR-Transkripte. Für die verwendeten Kultivierungsbedingungen konnte aber ein eindeutiges Haupttranskript identifiziert werden, das sich von TSS2 bis STOPP2 erstreckt und 335 nt lang ist (TR3). Dieses Transkript wurde im Folgenden für Strukturanalysen verwendet.

3.3.2 Bioinformatische Sekundärstrukturberechnung der Templat RNA

Im letzten Abschnitt konnten vier mögliche Transkripte für die Templat RNA aus *Nostoc sp.* PCC7120 identifiziert werden. Von diesen Transkripten wird unter den verwendeten Kultivierungsbedingungen ein Haupttranskript *in vivo* gebildet. Da die Primärstruktur dieser Transkripte nun bekannt ist, kann eine bioinformatische Sekundärstrukturberechnung erfolgen. Es gibt verschiedene Programme, die sich unterschiedlicher Algorithmen zur Berechnung der Sekundärstruktur bedienen. Zu den populärsten Programmen gehören Mfold (Zucker 2003) und RNAstructure (Reuter & Mathews 2010). Diese Programme berechnen die Veränderung der Gibbsschen freien Energie (ΔG) in theoretisch möglichen Strukturmodellen. Der ΔG -Wert beschreibt den Netto-Energieaustausch zwischen einem System mit der Umwelt und beruht auf Arbeiten des Physikers Josiah Willard Gibbs (Gibbs 1873). Je geringer der ΔG -Wert ist, desto wahrscheinlicher ist die Bildung der entsprechenden Struktur unter thermodynamischen Gesichtspunkten. Neben der Struktur mit der niedrigsten freien Energie (optimale Struktur) werden auch Strukturen mit geringer freier Energie (suboptimale Strukturen) berechnet. Für die suboptimalen Strukturen existieren schätzungsweise $1,8^n$ (n = Anzahl der Nukleotide in der Sequenz) Strukturmodelle (Zuker & Sankoff 1984). Daher wurde bei der Strukturberechnung, welche unter voreingestellten Parametern durchgeführt wurde, nur die Struktur mit der am niedrigsten berechneten freien Energie analysiert.

Zunächst wurde das Programm Mfold für die *in silico* Modellierung der Sekundärstrukturen von den vier möglichen Templat RNAs verwendet (Abbildung 27).

Im Vergleich der vier Strukturen ist gut zu erkennen, dass sich der Kernbereich der Templat RNA (TR1: TSS2 bis STOPP1) unabhängig von der zusätzlichen Sequenzinformation faltet, die durch den upstream gelegenen TSS1 oder den downstream gelegenen STOPP2 entsteht. Dadurch bleibt die Struktur des Kernbereichs erhalten.

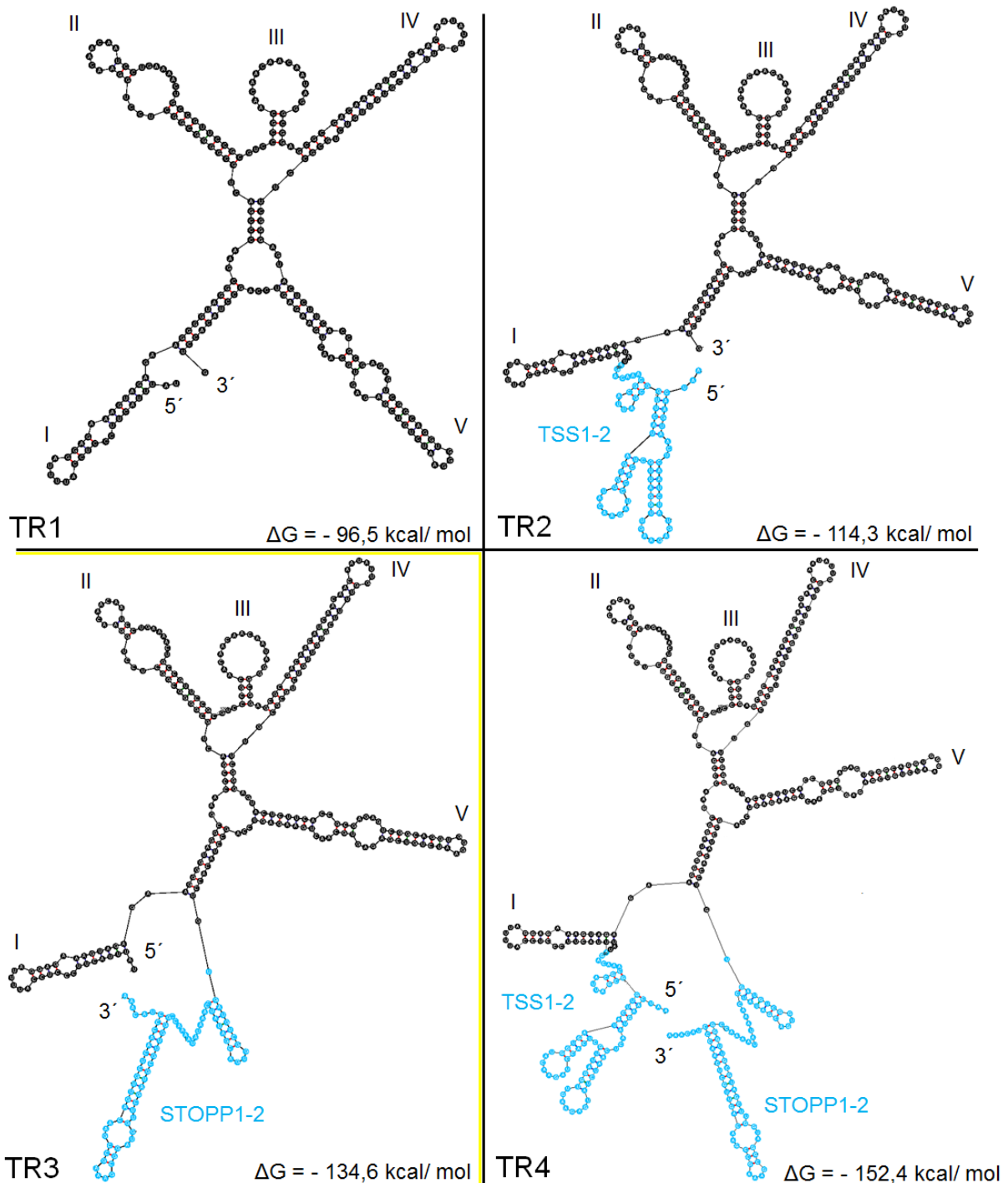


Abbildung 27: Vergleich der vier potentiellen Strukturen für die Templat RNA. Die Sekundärstrukturberechnung erfolgte mit dem Programm Mfold unter den vorgegebenen Einstellungen. Dabei ist jeweils die Struktur mit der am niedrigsten berechneten freien Energie (ΔG) abgebildet. Die vier Varianten der Templat RNA (TR1-4) unterscheiden sich in dem verwendeten Transkriptionsstart (TSS) und Transkriptionsstopp (STOPP). Die Sequenzbereiche TSS1-TSS2 bzw. STOPP1-STOPP2 sind blau markiert. Die Struktur des Haupttranskripts (TR3) ist gelb umrandet. Für eine einfache Zuordnung wurden größere Haarnadelstrukturen mit römischen Zahlen versehen. Es ist zu erkennen, dass sich der Kernbereich der Templat RNA (schwarz) unabhängig von dem zusätzlichen Sequenzbereich faltet, der durch den früheren TSS bzw. späteren STOPP entsteht.

Nachdem die Struktur für die vier möglichen Templat RNAs berechnet wurde, wurde die Strukturberechnung des Haupttranskripts (TR3) mit dem Programm RNAstructure überprüft (Abbildung 28).

Die mit RNAstructure berechnete Sekundärstruktur gleicht vom Aufbau der Struktur, die mit Mfold berechnet wurde. Die wenigen Unterschiede zwischen den Strukturen sind in Tabelle 11 (S. 77) aufgeführt. Die gemeinsame Struktur besteht aus sieben Haarnadelstrukturen (Haarnadelstrukturen VI-VII entsprechen STOPP1-STOPP2 der Mfold-Struktur). Der Bereich mit Sequenzhomologie zur VR erstreckt sich von Nukleotidposition 53-190 und ist markiert (gelb). Dieser Bereich schließt die Haarnadelstrukturen II bis IV mit ein. Beginn und Ende des Bereichs bestehen jeweils aus stabilen helikalen Strukturen, von denen besonders der 5'-Bereich durch einen hohen G/C-Anteil hervorsteht (Position 60). Innerhalb des Bereichs befinden sich mehrere Schleifen (Positionen: 75, 120 und 150) sowie eine Ausbuchtung (Position 85), die durch einen hohen Anteil an Adeninen gekennzeichnet sind. Direkt downstream (3') zu dem Bereich schließt sich eine interne Schleife an (Position 191), die zu Haarnadelstruktur V gehört. Die apikalen Schleifen von Haarnadelstruktur V, VI und VII besitzen jeweils die Nukleotidsequenz 5'GNRA'3 (N = jedes Nukleotid, R = Purin). Dabei lautet die genaue Nukleotidsequenz: 5'GGGA'3 (Haarnadelstruktur V), 5'GUGA'3 (Haarnadelstruktur VI) und 5'GAAA'3 (Haarnadelstruktur VII).

Bei diesen GNRA-Tetraloops handelt es sich um RNA-Struktur motive, die spezifische Bindestellen für RNA-Rezeptoren und Proteine bereitstellen können und dadurch eine Bedeutung bei der Bildung von RNA-Tertiärstrukturen und Ribonukleoproteinen besitzen (Jaeger, Michel & Westhof 1994; Wool, Gluck & Endo 1992). Bei dem RNA-Rezeptor handelt es sich im einfachsten Fall um einen helikalen Bereich. Auf Grundlage einer komparativen Sequenzanalyse wurde vermutet, dass die beiden Nukleotide am 3'-Ende des Tetraloops (RA) mit zwei aufeinanderfolgenden Purinen der „minor groove“ einer Helix interagieren können (Michel & Westhof 1990). Eine solche Interaktion ist in einem Hammerhead-Ribozym beobachtet worden, bei dem ein 5'GAAA'3 Tetraloop mit einem tandem C:G-Basenpaar in der „minor groove“ der Helix interagiert (Pley, Flaherty & MCKay 1994). GNRA-Tetraloops sind weit verbreitet und machen ca. ein Drittel der Tetraloops in rRNA und die Hälfte der Tetraloops einiger katalytischer RNAs aus (Klosterman et al. 2004; Fiore & Nesbitt 2013).

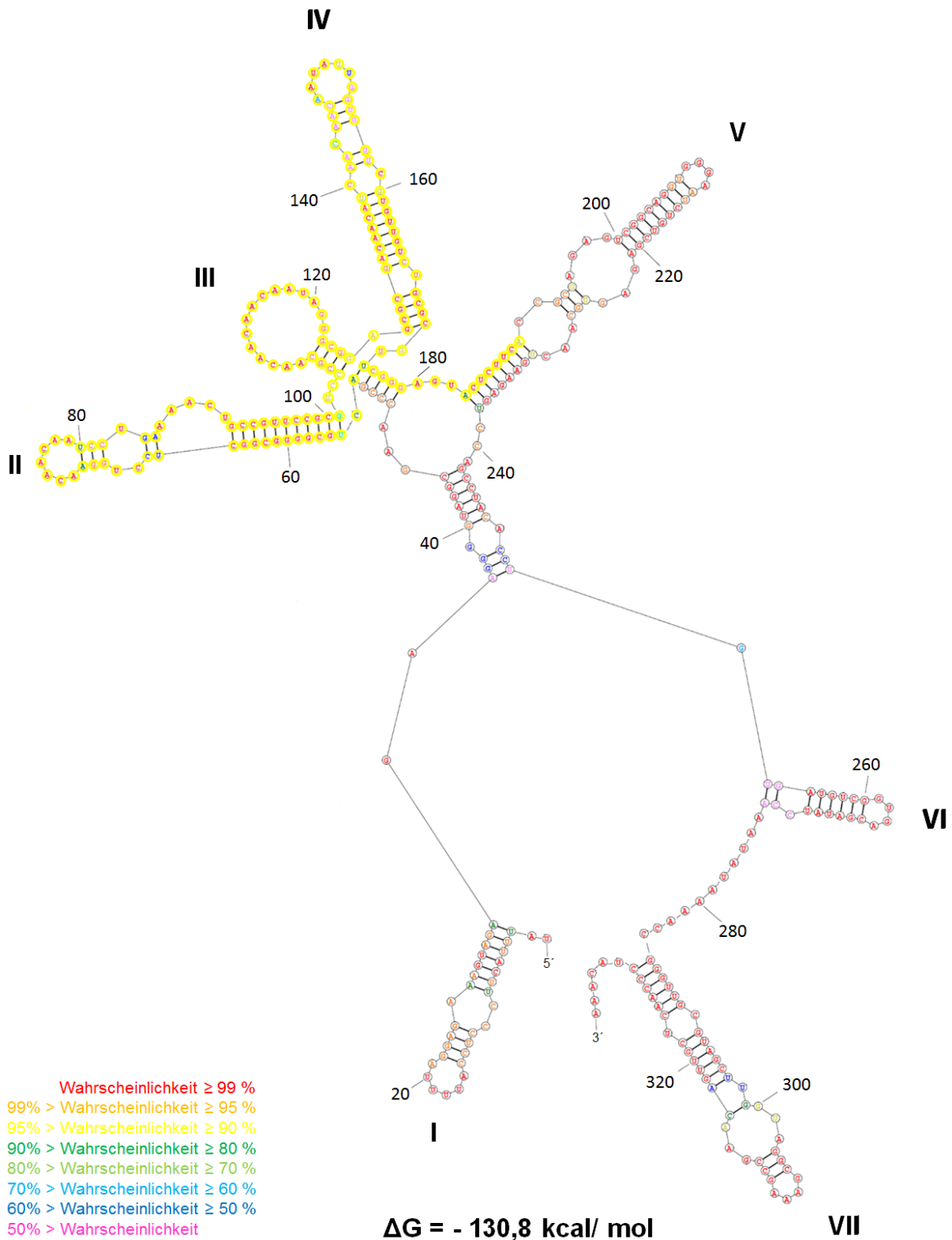


Abbildung 28. Sekundärstrukturberechnung des Haupttranskripts der Templat RNA. Die Sequenz der Templat RNA wurde mit RNAstructure unter den vorgegebenen Einstellungen gefaltet. Die Struktur mit der am niedrigsten berechneten freien Energie ($\Delta G = -130,8 \text{ kcal/mol}$) ist abgebildet. Die Wahrscheinlichkeit für eine korrekte Darstellung der einzelnen Basen ist mit einem Farbcode angegeben. Der zur VR homologe Teil verläuft von Position 53-190 und ist markiert (gelb). Größere Haarnadelstrukturen sind mit römischen Zahlen versehen, um eine einfache Zuordnung zu erlauben.

3.3.3 Experimentelle Daten unterstützen die Sekundärstrukturberechnung der Templat RNA

Die Sekundärstruktur der Templat RNA aus *Nostoc sp.* PCC7120 wurde im letzten Abschnitt bioinformatisch berechnet. Um die erhaltenen Daten experimentell zu validieren, wurden SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension)-Experimente durchgeführt. In SHAPE-Experimenten werden bestimmte Positionen der gefalteten RNA durch eine Chemikalie modifiziert und die Modifikationen über Primer-Extension nachgewiesen (Merino et al. 2005). Als Chemikalie kann NMIA (N-methylisotoic anhydride) verwendet werden. NMIA modifiziert flexible Nukleotide an der 2'-Hydroxylgruppe der Ribose und bildet ein stabiles 2'-O-Addukt, weshalb die verwendete RT während der Primer-Extension von der RNA abdissoziiert. Dadurch gibt es keine Unterscheidung zwischen verschiedenen Basen, sodass alle Positionen der RNA in einem Versuch überprüft werden können. Interaktionen, bei denen die Flexibilität der Nukleotide beschränkt wird (Basenpaarungen, Protein/RNA- und Tertiärstruktur-Wechselwirkungen), führen zu einer verminderten NMIA-Reaktivität. Die NMIA-Reaktivität kann durch eine Quantifizierung der Bandenintensitäten aus der Primer-Extension bestimmt werden. Aus den NMIA-Reaktivitäten werden SHAPE-Reaktivitäten berechnet, die anschließend in RNAstructure integriert werden können (Deigan et al. 2009).

Die TR3-IVT, welche dem Haupttranskript der Templat RNA entspricht, wurde für SHAPE-Experimente mit NMIA verwendet. Die IVT wurde vor der Modifizierung mit NMIA gefaltet. Da für die gefaltete RNA keine Möglichkeit bestand die Funktionalität zu überprüfen, wurde nach der stabilsten Konformation gesucht. Es wurde davon ausgegangen, dass es sich dabei um die native Konformation handelt. Um die stabilste Konformation zu identifizieren, wurde zunächst ein Vorversuch für die SHAPE-Experimente durchgeführt, bei dem verschiedene Faltungsbedingungen (Temperatur: 25 °C und 37 °C; MgCl₂-Konzentration: 0, 1, 6, 25, 50 und 100 mM) getestet wurden (Daten nicht gezeigt). Die Proben mit der unterschiedlich gefalteten RNA wurden mit NMIA modifiziert und für die Primer-Extension verwendet. Anschließend wurden die Bandenmuster der Proben miteinander verglichen. Bei beiden Temperaturen konnte für MgCl₂-Konzentrationen ≥ 6 mM ein gleichbleibendes Bandenmuster beobachtet werden, welches wahrscheinlich der stabilsten Konformation entspricht. Um sicher zu gehen das diese Konformation ausgebildet wird, wurde die MgCl₂-Konzentration in den SHAPE-Experimenten auf 16,7 mM erhöht (bei 25 °C für 20 Minuten) (s. Kap. 2.2.3.4.3).

Ein Ausschnitt aus einem typischen SHAPE-Experiment ist in Abbildung 29 dargestellt.

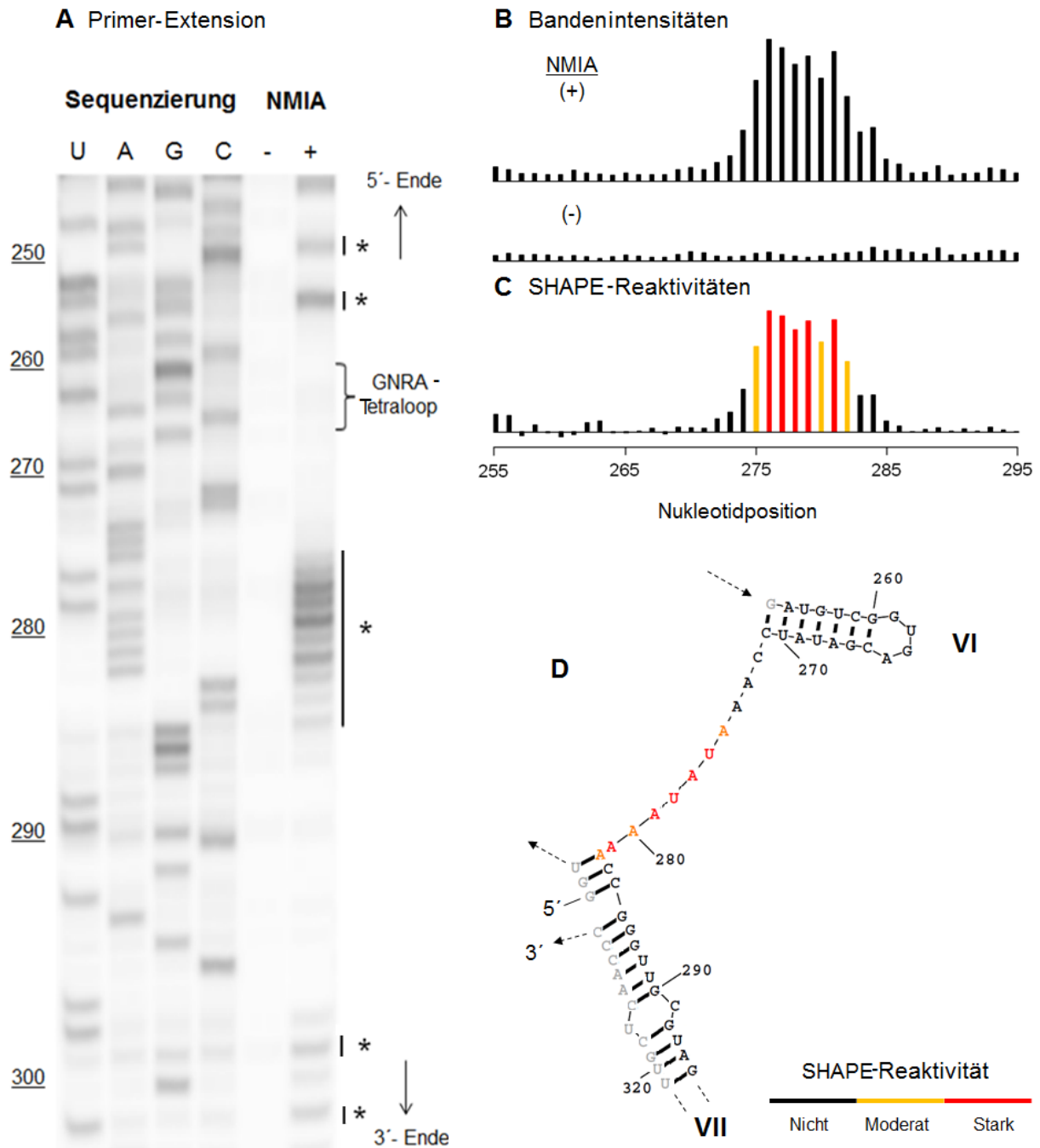


Abbildung 29: Ausschnitt eines SHAPE-Experiments der Templat RNA. (A): Primer-Extension. Die Sequenzierungsreaktionen (U, A, G, C) wurden mit den Dideoxynukleotiden (ddATP, ddTTP, ddCTP, ddGTP) für einen Kettenabbruch durchgeführt. Diese Banden laufen 1 Nukleotid höher als die der korrespondierenden NMIA-Spuren. Dabei unterscheidet sich NMIA(-) von (+) nur durch das Fehlen der Chemikalie. Die Position eines vorhergesagten GNRA-Tetraloops, sowie von flexiblen Nukleotiden (*) ist gekennzeichnet. **(B):** Bandenintensitäten. Zur Demonstration wurden die Bandenintensitäten der Nukleotide 255-295 für die NMIA(+) und (-)-Spuren aus **A** aufgetragen. **(C):** SHAPE-Reaktivitäten. Die Intensitäten aus **B** wurden genutzt um die SHAPE-Reaktivitäten zu berechnen. Die Reaktivitäten der Nukleotide sind mit einem Farbschema versehen: Nicht reaktiv (schwarz), moderat reaktiv (orange) und stark reaktiv (rot). **(D):** Integrierung der SHAPE-Reaktivitäten in RNAstructure. Die Daten aus **C** wurden als Einschränkungen der Pseudoenergie in RNAstructure integriert und die Sekundärstruktur berechnet. Der Strukturbereich von Domäne VI und VII ist gezeigt. Die Farbmarkierung entspricht der aus **C**. Nicht analysierte Nukleotide sind grau gekennzeichnet. Der GNRA-Tetraloop (261-264 nt) weist keine flexiblen Nukleotide auf. Dagegen ist der einzelsträngige Bereich um Position 280 flexibel.

An dem beispielhaften Polyacrylamidgel der Primer-Extension in Abbildung 29A ist zu erkennen, wie flexible Nukleotide in der RNA durch NMIA modifiziert wurden (NMIA+). Durch die Modifizierung bildet sich ein 2'-O-Addukt. Dadurch dissoziiert die RT ein Nukleotid vor der modifizierten Stelle vom Templat ab und es entstehen Abbruchprodukte, die als Banden in der NMIA(+)-Spur zu erkennen sind. Um zu prüfen, ob die Banden durch die Modifizierung oder eine Degradation der RNA entstanden sind, wurde eine Kontrolle ohne NMIA mitgeführt (NMIA -). Damit die modifizierten Nukleotide zu der RNA-Sequenz zugeordnet werden können, wurden parallel Sequenzierungsreaktionen durchgeführt (U, A, G, C).

In Abbildung 29B wurden die Bandenintensitäten für die Nukleotide 255-295 der NMIA(+) und (-)-Spuren quantifiziert und gegen die Nukleotidposition aufgetragen. Aus diesen Werten wurden die SHAPE-Reaktivitäten berechnet (Abbildung 29C), welche in RNAstructure integriert wurden. Ein Ausschnitt der Sekundärstruktur, die mit diesen SHAPE-Reaktivitäten berechnet wurde, ist in Abbildung 29D abgebildet. Die berechnete Struktur zeigt einen einzelsträngigen Bereich um Position 280. In diesem Bereich wurden flexible Positionen für viele, aber nicht alle Nukleotide mittels SHAPE bestimmt. Nukleotid 282 besitzt eine moderate Reaktivität, liegt in der Sekundärstruktur allerdings gepaart vor. Dagegen zeigt der berechnete GNRA-Tetraloop (Nukleotide 261-264) keine SHAPE-Reaktivität.

Für einen kompletten SHAPE-Versuch wurde mit vier Primern gearbeitet (Abbildung A4). Zudem wurden Gele unterschiedlicher Polyacrylamidkonzentrationen verwendet, um die Auftrennung der cDNA in bestimmten Bereichen zu verbessern. Der SHAPE-Versuch wurde dreimal mit zwei unabhängigen IVTs durchgeführt. Dabei konnten 46 der 340 Nukleotide der TR3-IVT mittels SHAPE nicht analysiert werden. Dazu gehören Nukleotide die nahe der vollständig revers transkribierten cDNA oder der Primerbindestelle lokalisiert sind. Auch betrifft es Bereiche komprimierter Banden, die aufgrund stabiler helikaler Regionen innerhalb der Sekundärstruktur auf dem Gel zusammen wandern. Dies wurde im Bereich der Nukleotide 97-106 und 127-132 festgestellt. Für die Nukleotide 65, 71 und 72 wurden zwischen den drei SHAPE-Versuchen stark voneinander abweichende NMIA-Reaktivitäten beobachtet. An diesen Nukleotiden ist auch die NMIA(-)-Bande sehr intensiv. Das kann durch stabile Sekundärstrukturen zu erklären sein, an denen die RT abdissoziiert. Die 294 analysierbaren Nukleotide wurden entsprechend ihrer SHAPE-Reaktivitäten in drei Gruppen unterteilt (Tabelle A3). Dabei wurden die Nukleotide in schwarz (nicht reaktiv), orange (moderat reaktiv) und rot (stark reaktiv) annotiert. Anschließend wurden die Daten als Einschränkungen der Pseudoenergie in RNAstructure integriert und die Struktur berechnet (Abbildung 30).

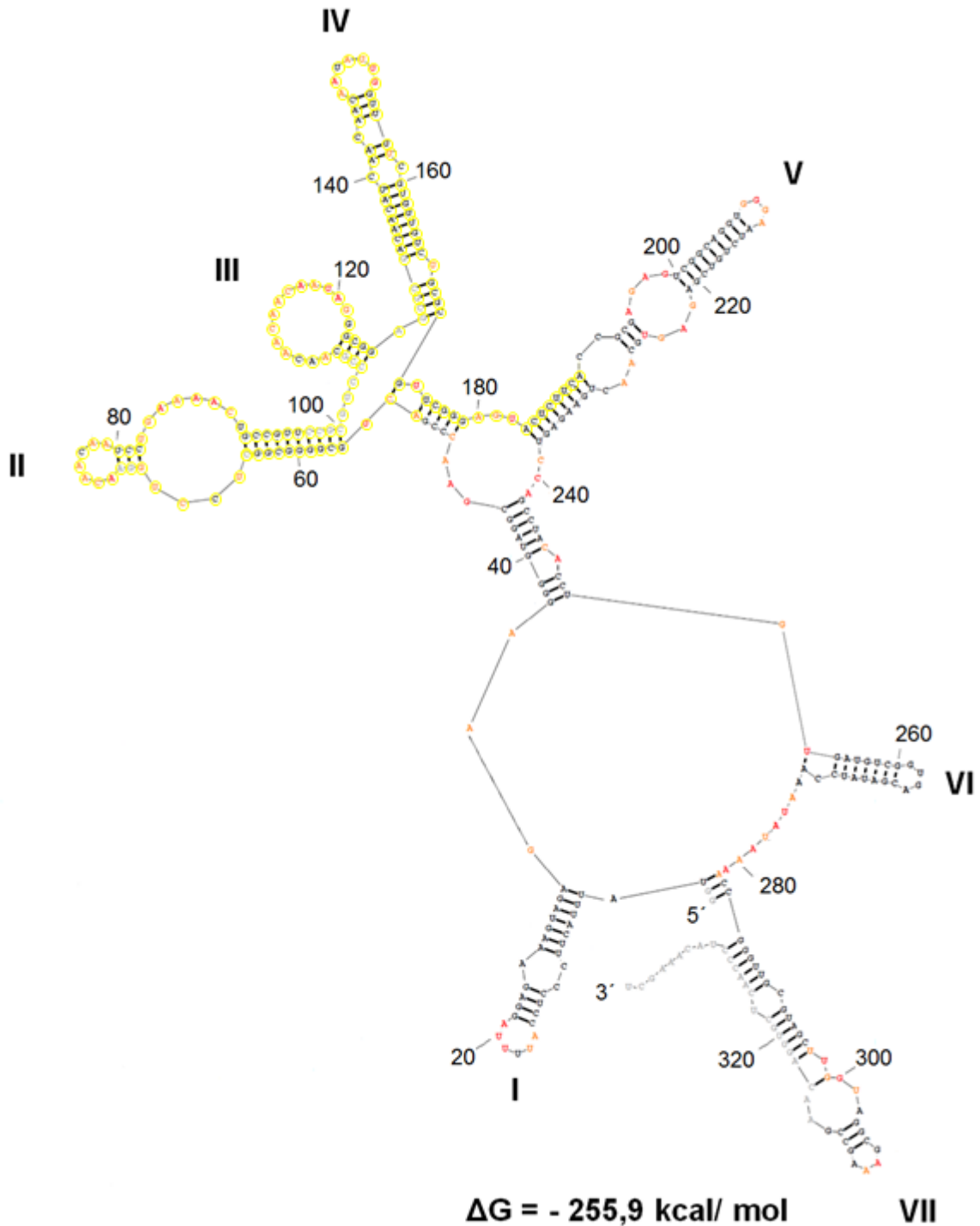


Abbildung 30: Sekundärstrukturberechnung der Templat RNA mit integrierten SHAPE-Reaktivitäten. Die Sequenz der TR3-IVT wurde mit RNAstructure unter Hinzufügen der SHAPE-Reaktivitäten als Einschränkungen der Pseudoenergie gefaltet. Die berechnete Struktur besitzt eine freie Energie (ΔG) von $-255,9 \text{ kcal/ mol}$. Die ersten beiden Nukleotide am 5'-Ende, sowie die letzten drei Nukleotide am 3'-Ende sind Teil der IVT und gehören nicht zur Templat RNA. Daher wurden sie von der Nummerierung ausgeschlossen. Der zur VR homologe Teil verläuft von Position 53-190 und ist markiert (gelb). Größere Haarnadelstrukturen sind mit römischen Zahlen versehen um eine einfache Zuordnung zu erlauben. Die Annotation für die SHAPE-Reaktivitäten ist: Nicht gemessen (grau), keine Reaktivität (schwarz), moderate Reaktivität (orange) und starke Reaktivität (rot).

In Abbildung 30 passen die berechneten SHAPE-Reaktivitäten von 39 Nukleotiden (~ 13 %) nicht mit der berechneten Sekundärstruktur zusammen. Dabei gibt es zwei Möglichkeiten:

In 29 Fällen zeigt die Sekundärstruktur ein ungepaartes Nukleotid, obwohl dieses nach den SHAPE-Daten nicht flexibel sein sollte. Das kann durch intramolekulare Wechselwirkungen in der Tertiärstruktur begründen sein, die in einem SHAPE-Experiment ebenfalls erfasst werden. Auch können die Reaktivitäten nur knapp unter dem festgelegten Wert für moderat flexible Nukleotide ($\geq 0,4$) liegen. So zeigen fünf nicht flexible Nukleotide Reaktivitäten $\geq 0,37$. Dabei ist zu beachten, dass der durchschnittliche Wert nicht reaktiver Nukleotide bei 0,09 liegt.

Desweiteren kann in zehn Fällen ein gepaartes Nukleotid beobachtet werden, das nach den SHAPE-Daten flexibel sein sollte. Dabei handelt es sich in acht Fällen um moderat reaktive und in zwei Fällen um stark reaktive Nukleotide.

Bei den moderat reaktiven Nukleotiden ist Position 53 (Adenosin) hervorzuheben. Der Basenpartner 175 (Uridin) wird als nicht flexibel eingestuft, verfügt aber über einen Reaktivitätswert von 0,399. Das forcierte Öffnen dieser Bindung führt zu einer marginalen Erhöhung der minimalen freien Energie um 0,1 (auf -255,8 kcal/ mol). Die anderen moderat reaktiven Nukleotide, die als gepaart berechnet wurden, sind an Positionen lokalisiert die direkt an einzelsträngige Bereiche, wie Schleifen oder Ausbuchtungen, grenzen. Diese Zuordnung moderat reaktiver Nukleotide wurde bereits beobachtet (Wilkinson, Merino & Weeks 2006).

Dagegen liegen stark reaktive Nukleotide typischerweise nur einzelsträngig vor. Das ist für zwei Nukleotide jedoch nicht der Fall. Diese beiden Nukleotide sind an den Basenpaarungen G-U an Position 195/225 (Reaktivität 195: 0,32), sowie U-A an Position 253/273 (Reaktivität 273: 0,13) involviert. Das forcierte Öffnen der Bindung des Wobble-Basenpaars 195/225 erhöht die minimale Energie um 1,7 (auf -254,2 kcal/ mol), während das Öffnen von Basenpaar 253/273 die minimale Energie um 0,6 (auf -255,3 kcal/ mol) erhöht.

Die minimalen freien Energien der Sekundärstrukturen, die mit bzw. ohne experimentelle Daten berechnet wurden, weichen stark voneinander ab. So zeigen die Sekundärstrukturen ohne integrierte SHAPE-Reaktivitäten Werte von -134,6 kcal/ mol und -130,8 kcal/ mol, während die Struktur mit integrierten SHAPE-Reaktivitäten einen Wert von -255,9 kcal/ mol besitzt. Das kommt daher, dass die über Einschränkungen der Pseudoenergie integrierten SHAPE-Reaktivitäten über einen zusätzlichen nearest-neighbor-Parameter in den Algorithmus der dynamischen Programmierung von RNAstructure eingebunden werden (Deigan et al. 2009). Die Strukturen weisen dagegen im Vergleich kaum Unterschiede auf. Mit Zusatz von experimentellen Daten sind die ersten drei Nukleotide am 5'-Ende in Basenpaarungen involviert (nicht nummeriert). Das ist ein Unterschied zu den Strukturen, welche ohne experimentelle Daten berechnet wurden. Der Unterschied kommt durch die zusätzlichen Nukleotide zustande, durch die sich die IVT von der Templat RNA unterscheidet.

Abweichungen, die nicht durch diesen Sequenzunterschied berechnet wurden, sind in Tabelle 11 aufgelistet.

Tabelle 11: Unterschiede zwischen den berechneten Sekundärstrukturen

Nukleotid	Mfold	RNAstructure(-)	RNAstructure(SHAPE)
36	bp [251]	bp [251]: < 50 %	ss
37	bp [250]	bp [250]: 50-60 %	bp [251]
38	bp [249]	bp [249]: 50-60 %	bp [250]
39	ss	ss: 50-60 %	bp [249]
54	ss	ss: 80-95 %	bp [173]
55	ss	bp [101]: 70-80 %	ss
66	ss	bp [85]: 50-60 %	ss
67	ss	bp [84]: 50-60 %	ss
253	ss	bp [273]: < 50 %	bp [273]
254	bp [271]	bp [272]: < 50 %	bp [271]
272	ss	bp [254]: < 50 %	ss

Die Sekundärstrukturen von Mfold und RNAstructure ohne SHAPE-Daten (-) bzw. mit SHAPE-Daten (SHAPE) wurden miteinander verglichen und die Unterschiede aufgelistet. Dabei ist die Nukleotidposition angegeben. Die Nukleotide liegen in den Strukturen entweder einzelsträngig (ss) oder in Basenpaarung (bp) mit einem weiteren Nukleotid vor (eckige Klammern). Für RNAstructure(-) sind zusätzlich die berechneten Wahrscheinlichkeiten aufgeführt.

Bei den ersten fünf Unterschieden sind die Strukturen von Mfold und RNAstructure(-) konform. RNAstructure(SHAPE) gibt Nukleotid 36, für das eine moderate Reaktivität berechnet wurde, als einzelsträngig aus. Dadurch verschiebt sich der Bindungspartner in dem Bereich (Nukleotide 36-39) um ein Nukleotid. Nukleotid 54 besitzt eine moderate SHAPE-Reaktivität. Dennoch berechnet RNAstructure(SHAPE) eine Basenpaarung mit Nukleotid 173. Dies steht im Gegensatz zu Mfold und RNAstructure(-), welche Nukleotid 54 als einzelsträngig berechnen.

Bei den weiteren Unterschieden kommt es zu Abweichungen zwischen den berechneten Strukturen von Mfold und RNAstructure(-). Der Zustand der betroffenen Nukleotide kann durch die hinzugefügten SHAPE-Reaktivitäten genauer bestimmt werden. Nukleotide 55, 66 und 67 scheinen einzelsträngig vorzuliegen, da für die Nukleotide bzw. den potentiellen Bindungspartner eine Flexibilität über SHAPE berechnet wurde. Nukleotid 253 wurde bereits erwähnt, da es eine starke SHAPE-Reaktivität besitzt, in RNAstructure(SHAPE) allerdings als gepaart berechnet wurde. RNAstructure(-) berechnet auch eine Basenpaarung, diese jedoch

mit geringer Wahrscheinlichkeit (< 50 %). Mfold berechnet die Basenpaarung dagegen nicht. Es ist denkbar, dass Nukleotid 253 keine stabile Bindung zu seinem Basenpartner hat und dadurch sowohl der gepaarte als auch der ungepaarte Zustand in der Struktur vorkommt. Die letzten Abweichungen betreffen Nukleotide 254 und 272, welche in RNAstructure(-) mit geringer Wahrscheinlichkeit (< 50 %) als gepaart berechnet wurden. Mfold und RNAstructure(SHAPE) berechnen die Nukleotide 254 und 271 als Bindungspartner, wobei Nukleotid 272 einzelsträngig ist.

Aus den Unterschieden zwischen den Strukturen in Tabelle 11 lässt sich der Grad der Übereinstimmung berechnen. Die Strukturen aus RNAstructure(-) und RNAstructure(SHAPE) besitzen zehn Abweichungen, wodurch eine Übereinstimmung von 97 % erreicht wird. Durch jeweils sechs Abweichungen konnte eine Übereinstimmung von 98 % für die Strukturen von Mfold und RNAstructure(-) bzw. Mfold und RNAstructure(SHAPE) erreicht werden.

3.3.4 Berechnung einer Konsensus-Struktur für mehrere Templat RNAs

Da die bioinformatisch berechnete Sekundärstruktur der Templat RNA zu 98 % mit der SHAPE-optimierten Struktur übereinstimmt (Tabelle 11), bietet sich eine Berechnung weiterer Templat RNA-Sekundärstrukturen an.

Die Transkriptenden anderer Templat RNAs sind aber nicht bekannt. Daher wurde der Versuch unternommen weitere Transkriptenden über ein Alignment (ClustalW) mit der Templat RNA aus *Nostoc sp.* PCC7120 zu identifizieren. Dabei wurde die DNA-Vorlage der größten Variante der Templat RNA (TR4) verwendet. Diese besteht aus dem zur VR homologen Bereich und zusätzlichen 129 Nukleotiden upstream und 145 Nukleotiden downstream. Die größte Variante der Templat RNA wurde ausgewählt, da der verwendete Transkriptionsstart bzw. Transkriptionsstopp keinen Einfluss auf den Kernbereich der berechneten Sekundärstruktur hat (Abbildung 27). Für das Alignment wurden neben der Sequenz aus *Nostoc sp.* PCC7120 noch 12 weitere TR-Sequenzen verwendet, die sich aus dem zur VR homologen Bereich und jeweils 200 nt upstream und downstream zusammensetzen.

Für die Auswahl der TR-Sequenzen wurden sämtliche sequenzierte DGRs aus dem cyanobakteriellen Subtree der DGR-kodierten RTs verwendet (Abbildung 8). Dazu gehören sieben TR-Sequenzen aus Cyanobakterien (*Nostoc sp.* PCC7120, *Nostoc punctiforme*, *Anabaena flos-aquae*, *Nodularia spumigena* und *Cyanothece sp.* CCY0110), sowie zwei TR-Sequenzen aus grünen Schwefelbakterien (*Chlorobium phaeobacteroides* und *Pelodictyon phaeoclathratiforme*) und zwei TR-Sequenzen aus Purpurbakterien (*Thiorhodovibrio sp.* 970 (Schwefelpurpurbakterium) und *Rhodomicrobium vannielii* (Nichtschwefelpurpurbakterium)).

Um auch TR-Sequenzen außerhalb des cyanobakteriellen Subtrees untersuchen zu können, wurde ein Spirochät (*Treponema denticola*) und ein Phage (*Bordetella*-Phage) hinzugenommen. Die verwendeten TR-Sequenzen sind in Tabelle 12 zusammengefasst.

Tabelle 12: Verwendete TR-Sequenzen für die Sekundärstrukturberechnung

Organismus	GI-Nummer der RT	Bezeichnung
<i>Nostoc sp. PCC7120</i>	17230989	<i>Nostoc</i>
<i>Nostoc punctiforme</i>	186684985	<i>Nostoc punc.</i>
<i>Anabaena flos-aquae</i>		<i>Anabaena</i>
<i>Nodularia spumigena</i>	119509703	<i>Nodularia A</i>
	119511289	<i>Nodularia B</i>
<i>Cyanothece sp. CCY0110</i>	126659397	<i>Cyanothece A</i>
	126660098	<i>Cyanothece B</i>
<i>Chlorobium phaeobacteroides</i>	119356297	<i>Chlorobium</i>
<i>Pelodictyon phaeoclathratiforme</i>	194337212	<i>Pelodictyon</i>
<i>Thiorhodovibrio sp. 970</i>	350574651	<i>Thiorhodo.</i>
<i>Rhodomicrobium vannielii</i>	312115534	<i>Rhodomicrob.</i>
<i>Treponema denticola</i>	42527768	<i>Treponema</i>
<i>Bordetella</i> -Phage	41179367	Phage

Ausgehend von dem Alignment wurde der Sequenzbereich, welcher die DNA-Vorlage für die Templat RNA in *Nostoc sp. PCC7120* darstellt, extrahiert. Die 13 extrahierten TR-Sequenzen wurden anschließend in RNA umgeschrieben und für die Berechnung einer Konsensus-Struktur mit dem Programm LocARNA verwendet (Smith et al. 2010; Will et al. 2007; Will et al. 2012). LocARNA führt einen Vergleich der RNA-Moleküle durch, der nicht nur die Sequenzen, sondern auch die Strukturen berücksichtigt. Ausgehend von den vielen potentiellen Strukturen der einzelnen RNA-Moleküle, wird dadurch die Struktur identifiziert, welche die beste Übereinstimmung für alle analysierten RNA-Moleküle aufweist. Die Datenausgabe besteht neben der Konsensus-Struktur auch aus dem multiplen Alignment, welches für die Berechnung verwendet wurde (Abbildung 31, Abbildung 32).

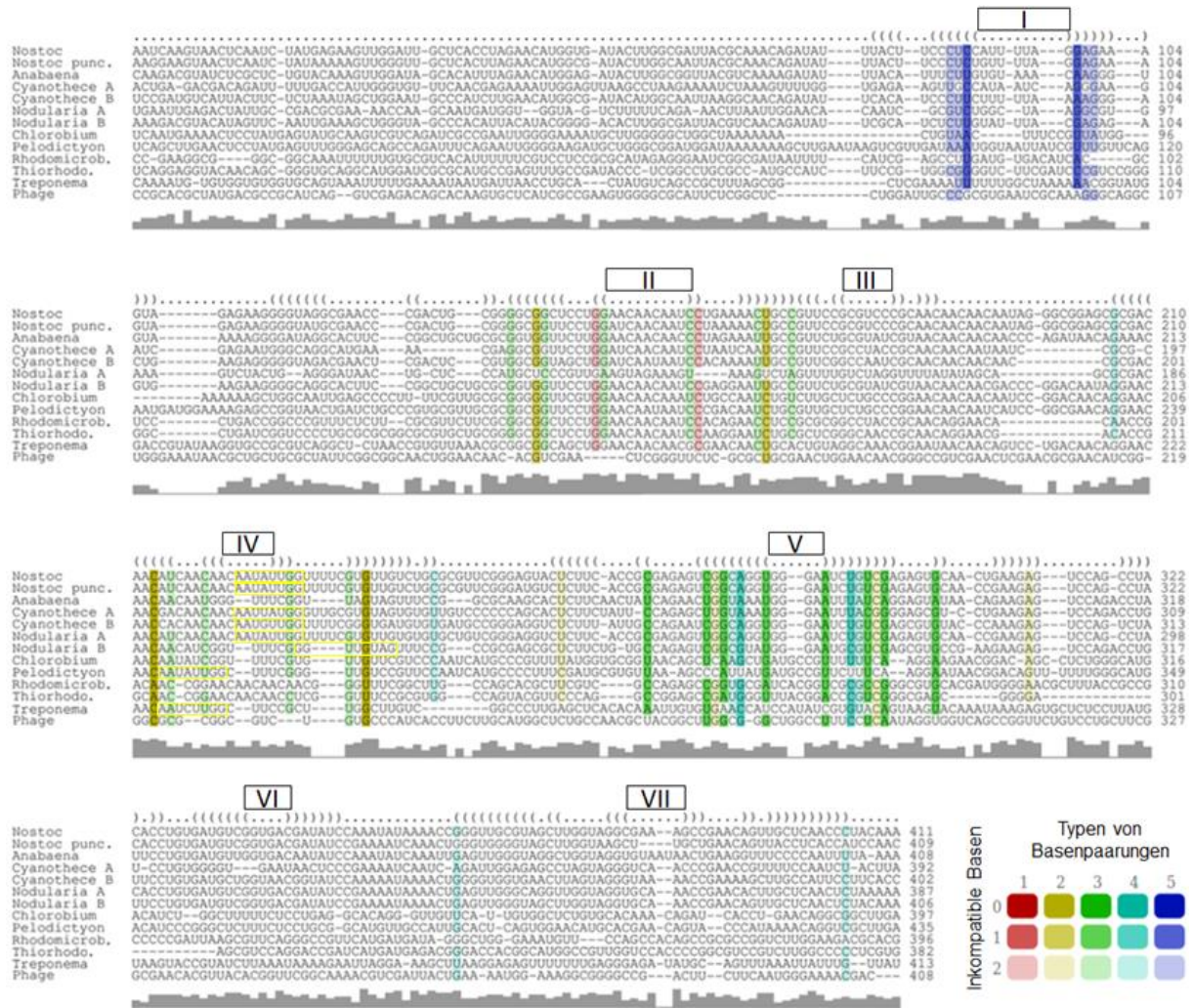


Abbildung 31: Alignment der Templat RNAs für die Berechnung der Konsensus-Struktur. Die TR-Sequenzen aus Tabelle 12 wurden für die Berechnung einer Konsensus-Struktur verwendet. Das multiple Alignment ist abgebildet. In dem Alignment sind kompatible Basenpaare gefärbt. Die Farbe gibt Aufschluss über die Anzahl unterschiedlicher Basenpaartypen (C-G, G-C, A-U, U-A, G-U oder U-G) und ist ein Maß der Sequenzkonservierung. Die Farbsättigung nimmt mit steigender Anzahl an inkompatiblen Basen ab und stellt ein Maß der Strukturkonservierung für das Basenpaar dar. Die apikalen Schleifen von sieben Haarnadelstrukturen sind gekennzeichnet. Acht der Templat RNAs besitzen in Haarnadelstruktur IV ein snRNP-Bindemotiv (gelb markiert).

In dem multiplen Alignment aus Abbildung 31 sind sieben potentielle Haarnadelstrukturen hervorgehoben. Es ist zu erkennen, dass die Haarnadelstrukturen I, II, IV und V recht konserviert sind. Dabei weist die Sequenz des *Bordetella*-Phagen die geringste Konservierung auf. Die apikalen Schleifen der Haarnadelstrukturen präsentieren einen Uracil-Stretch (I), veränderliche Adenine (entsprechen variablen Positionen in der VR) (II), snRNP (small nuclear ribonucleic particles)-Bindemotive (IV) und einen GNRA-Tetraloop mit der Sequenz 5'GGA3' (nur Cyanobakterien) (V). Die Haarnadelstrukturen III, VI und VII sind nur in Cyanobakterien konserviert (Ausnahme: Nodularia A (III) und Cyanobacterium A (VI)). Dabei präsentiert Haarnadelstruktur VI einen GNRA-Tetraloop mit der Sequenz 5'GUGA3'.

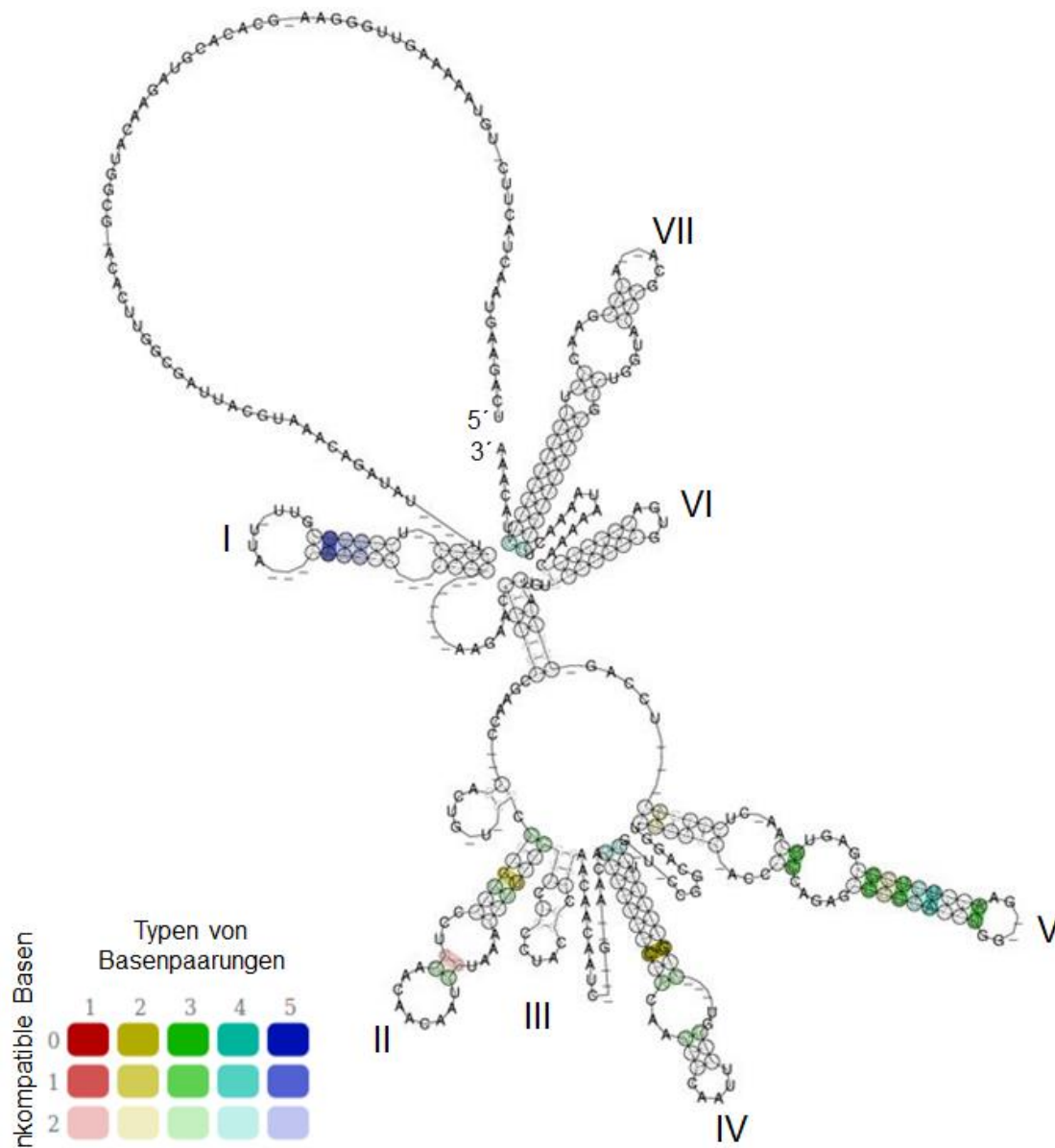


Abbildung 32: Konsensus-Struktur der Templat RNAs. Basierend auf dem Alignment der Templat RNAs wurde die Konsensus-Struktur berechnet. Dabei entspricht die farbliche Markierung der des Alignments und gibt Aufschluss über Sequenz- und Strukturkonservierung (Abbildung 31). Die gekennzeichneten Haarnadelstrukturen (I-VII) können dem Alignment zugeordnet werden.

Die berechnete Konsensus-Struktur in Abbildung 32 weist mit den sieben Haarnadelstrukturen (I-VII) eine hohe Ähnlichkeit zu dem Haupttranskript der Templat RNA aus *Nostoc* PCC7120 auf (Abbildung 30). Das Haupttranskript beinhaltet aber nicht den einzelsträngigen Bereich vor Haarnadelschleife I (TSS1-TSS2). Dieser Bereich ist für die untersuchten Templat RNAs nicht konserviert genug, um ein Konsensus-Strukturelement berechnen zu können. Die Sequenz von STOPP1 bis STOPP2 (VI-VII) zeigt zwar eine Struktur, diese ist jedoch nur für cyanobakterielle Templat RNAs konserviert (Abbildung 31). Dagegen zeigt der Kernbereich der Konsensus-Struktur, mit den Haarnadelstrukturen I, II, IV und V (Ausnahme: III), für alle analysierten Templat RNAs konservierte Strukturelemente.

Um abschätzen zu können, wie aussagekräftig die Ergebnisse der berechneten Konsensus-Struktur sind, wurde eine Kontrolle durchgeführt. Als Kontrolle wurde die Templat RNA aus *Nostoc sp.* PCC7120 (TR4) mit 12 zufällig generierten Sequenzen aligned (ClustalW) und der Sequenzbereich der Templat RNA für die Berechnung einer Konsensus-Struktur mit LocARNA verwendet (Abbildung A5). Die zufällig generierten Sequenzen wurden auf eine Länge von 537 nt eingestellt. Die Länge setzt sich, analog zu den untersuchten TR-Sequenzen, aus dem zur VR homologen Bereich des DGRs aus *Nostoc sp.* PCC7120 (137 nt) und einem umgebenden Bereich von 400 nt zusammen.

Die Kontrolle wurde dreimal durchgeführt, wobei keine konservierten Sequenzen oder Strukturelemente beobachtet werden konnten. Das spricht dafür, dass die beobachteten Strukturelemente in Abbildung 32 nicht durch zufällige Sequenzen entstanden sind, sondern funktionell wichtige und daher konservierte Strukturelemente repräsentieren.

4 Diskussion

4.1 Das Screening der Cyanobakteriensammlung identifiziert 34 DGRs

In der vorliegenden Arbeit wurden offene Fragen zu diversitätsgenerierenden Retroelementen (DGRs) bearbeitet. Eine der offenen Fragen beschäftigt sich mit der Verteilung von DGRs in den Wirtsorganismen. Nach einer bioinformatischen Analyse von Schillinger und Kollegen sind DGRs in Prokaryoten zwar weit verbreitet, konnten aber nur in ~ 3 % der sequenzierten Genome identifiziert werden (Schillinger et al. 2012).

Um die von Schillinger und Kollegen beobachtete geringe Inzidenz experimentell zu prüfen, wurde das Vorkommen von DGRs in 169 kultivierten Cyanobakterien aus der Stammsammlung von Prof. Burkhard Büdel (Pflanzenökologie & Systematik, TU Kaiserslautern) untersucht (Tabelle A1). Dafür wurde ein Screening auf die DGR-kodierte RT durchgeführt (Kap. 3.1.1), von dem eine Übersicht in Abbildung 33 dargestellt ist.

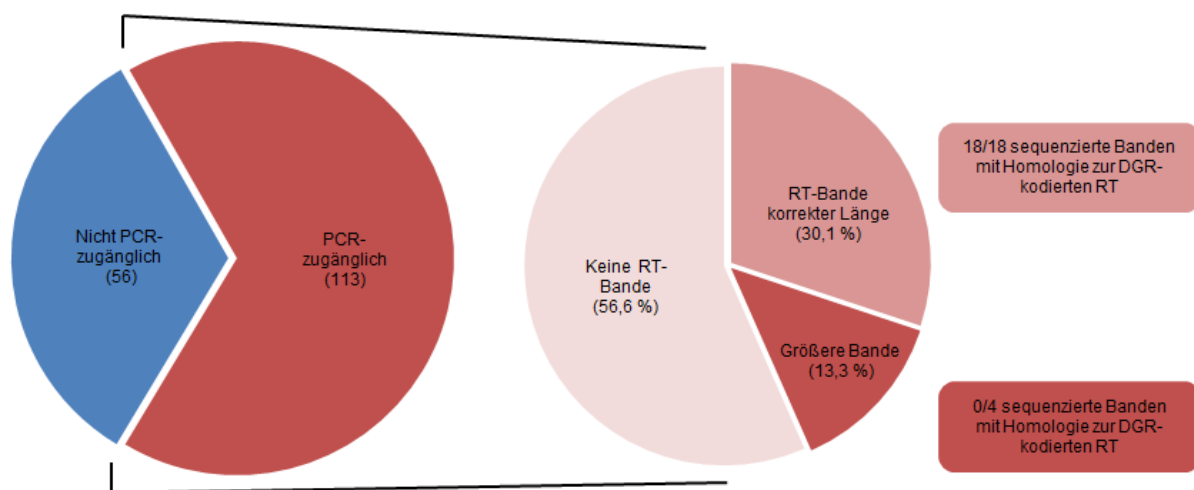


Abbildung 33: Übersicht des cyanobakteriellen Screenings. Von den 169 Isolaten der cyanobakteriellen Stammsammlung konnte nur bei ca. zwei Dritteln die 16S rDNA amplifiziert werden. Diese wurden als PCR-zugänglich bewertet. In ~ 30 % (34) der PCR-zugänglichen Isolate konnte eine Bande für das RT-Fragment amplifiziert werden, die der theoretisch korrekten Größe entspricht. Die Sequenzierung einiger PCR-Produkte konnte zeigen, dass es sich um DGR-kodierte RT-Sequenzen handelt. Größere PCR-Produkte besitzen keine Homologie zu RT-Sequenzen.

Anhand der Übersicht ist zu sehen, dass etwa ein Drittel der Proben für die PCR nicht zugänglich waren. Dies wurde über eine parallele PCR auf die 16S rDNA bestimmt (Abbildung 6, S. 37) und kann mit der Verwendung ganzer Cyanobakterienzellen in der PCR zusammenhängen. Cyanobakterien können eine widerstandsfähige Zellwand und PCR-inhibierende Substanzen aufweisen, welche die PCR stören (Hoiczky & Hansel 2000; Koskeniemi et al. 2007). In dem Screening ging es jedoch primär darum, schnell und

spezifisch den Großteil der DGRs zu identifizieren. Durch das Verwenden ganzer Zellen in der PCR wurde die zeitaufwändige DNA-Isolation umgangen und das Screening konnte zeitsparend durchgeführt werden. Zudem kommt es bei der DNA-Isolation auch immer zu einem Verlust von DNA. Daher kann das für diese Arbeit optimierte PCR-Protokoll in Applikationen Verwendung finden, bei denen nur eine begrenzte Probenmenge zur Verfügung steht.

Zugänglich für die PCR waren 113 Cyanobakterienstämme. In ca. der Hälfte (56.6 %) dieser Stämme konnte die 16S rDNA amplifiziert werden, aber keine RT-Sequenz. Das könnte bedeuten, dass im Genom der Zellen kein DGR vorkommt. Alternativ könnte die RT-Sequenz so weit von dem Konsensus abweichen, dass die degenerierten Primer nicht mehr effizient binden können. Die degenerierten Primer wurden auf Grundlage eines Alignments von sieben bekannten cyanobakteriellen DGR-kodierten RT-Sequenzen erstellt und binden diese an konservierten Bereichen von Domäne drei (QCDIXKYFP) und Domäne fünf (YVDDFALF). In diesen Bereichen kommen unbestimmte bzw. sechsfach degenerierte Aminosäuren nur einmal vor (X bzw. L (Leucin)). Diese Aminosäuren wurden vermieden, um die Degeneriertheit der Primer gering zu halten, wodurch sich die Spezifität erhöht. Um die Degeneriertheit weiter zu senken, wurden bei der Erstellung der Primer Kodons verwendet, die bevorzugt von den Cyanobakterien im Alignment verwendet wurden. Ein Beispiel ist die Aminosäure Valin (V) mit dem degenerierten Kodon „GUN“. Da die Mehrzahl der Cyanobakterien im Alignment an der Wobble-Position dieses Kodons ein Pyrimidin besitzt, wurde für den Primer das degenerierte Kodon „GUY“ verwendet. Die degenerierten Primer NZ112 und NZ114 sind 256-fach und 128-fach degeneriert (Tabelle A4), was einen Kompromiss zwischen einer spezifischen und einer möglichst umfangreichen Erfassung der RT-Sequenzen darstellt. Sind die Primer nicht degeneriert genug, können RT-Sequenzen im Screening übersehen werden, wenn diese andere Aminosäuren oder einen anderen Kodongebrauch verwenden, als die sieben DGR-kodierten RT-Sequenzen des Alignments.

Wenn die Primer nicht spezifisch genug sind, können diese nicht binden bzw. es kommt zu einer Fehlbindung. Fehlbindungen wurden für vier sequenzierte PCR-Produkte beobachtet, die größer waren als das theoretisch amplifizierte Produkt (320 bp). Hierbei konnte keine Homologie zu RT-Sequenzen festgestellt werden. PCR-Produkte mit einer Größe über 320 bp wurden in 13,3 % (15) der analysierbaren Cyanobakterien beobachtet. Es wird davon ausgegangen, dass diese Banden durch Fehlbindungen entstanden sind und keine RT-Sequenzen repräsentieren.

30,1 % (34) der analysierbaren Cyanobakterien ergaben ein PCR-Produkt mit einer Größe von ca. 320 bp (Tabelle 8, S. 39). 18 dieser PCR-Produkte wurden sequenziert und eine Homologie zu RT-Sequenzen festgestellt. Dies ist ein Hinweis auf das Vorhandensein eines DGRs in dem

Genom der betreffenden Cyanobakterien (Mischsequenzen von RTs können durch mehrere DGRs entstehen). Es könnten jedoch auch andere bakterielle RT-Fragmente, wie z.B. von der Reverse Transkriptase-Domäne der Gruppe II Introns, amplifiziert worden sein. Dies ist ebenfalls denkbar, da besonders das YxDD-Motiv in der katalytischen Domäne fünf in allen RTs hoch konserviert ist und einen Teil der Bindestelle von Primer NZ114 darstellt (Simon & Zimmerly 2008). Die Primer wurden so erstellt, dass sich das SQ-Motiv, welches für DGR-kodierte RTs diagnostisch ist (Medhekar & Miller 2007; Schillinger et al. 2012), auf dem Amplifikat befindet. Das SQ-Motiv kann in dem Aminosäure-Alignment der RT-Sequenzen für jede Sequenz nachgewiesen werden (Abbildung 7, S. 41). Demnach ist eine 320 bp Bande für die RT ein verlässliches Zeichen für die Identifizierung von DGRs. Das bedeutet, dass alle 34 Cyanobakterienstämme mit dieser Bande ein DGR besitzen.

Eine Gefahr für Stammsammlungen besteht in der Kontamination von Stämmen durch andere Stämme, während der Überimpfung in frische Medien. In diesem Fall würden identische Nukleotidsequenzen für die RT-Fragmente zu erwarten sein. Die Nukleotidsequenzen von unterschiedlichen Arten zeigen aber Abweichungen zueinander (3-40 %), was gegen eine Kontamination spricht (Kap. 3.1.2). Eine frühe Kontamination, welche zusammen mit der natürlichen Mutationsrate zu den Sequenzunterschieden geführt hat, ist ebenfalls unwahrscheinlich. In diesem Fall sollten auch verschiedene Umweltisolate einer Art die beobachteten Sequenzunterschiede vorweisen. Die drei Umweltisolate von *Tolypothrix distorta* (BB97.17 (1963 isoliert), BB97.18 (1963 isoliert) und BB97.80 (1970 isoliert)) weisen aber identische Nukleotidsequenzen für die RT-Fragmente auf (Kap. 3.1.1). Daher scheint es in der Stammsammlung keine dahingehende Kontamination gegeben zu haben.

Mit dem Screening auf die DGR-kodierte RT konnten schnell und spezifisch DGRs in über 30 % der analysierbaren Cyanobakterien identifiziert werden. Dabei wurden aufgrund des etablierten PCR-Protokolls, bei dem ganze Zellen eingesetzt wurden, nur geringe Probenmengen benötigt. Diese Methode könnte auch in zukünftigen Studien mit anderen Organismen Verwendung finden, bei denen keine Zeit für die Etablierung der Anzuchtbedingungen, sowie die Anzucht der Organismen zur Verfügung steht. Dabei ist jedoch ein Kenntnis über DGR-kodierte RT-Sequenzen aus einigen dieser Organismen notwendig, um die degenerierten Primer erstellen zu können. Zudem ist es sinnvoll das diagnostische SQ-Motiv in den amplifizierten Bereich zu integrieren, um die RT-Sequenzen eindeutig den DGRs zuordnen zu können.

4.1.1 DGRs sind in der Stammsammlung weiter verbreitet als in öffentlichen Sequenzdatenbanken

Bei der bioinformatischen Analyse von Schillinger und Kollegen wurden 6000 Genome aus öffentlichen Sequenzdatenbanken auf das Vorhandensein von DGRs untersucht und dabei 155 DGRs identifiziert (Schillinger et al. 2012). Dabei stammen 2,3 % der untersuchten Genome von Cyanobakterien, welche 5,8 % der identifizierten DGRs stellen. Dieser Analyse nach sind nur etwa 6,5 % der Cyanobakterien mit DGRs assoziiert.

Um die Ergebnisse der bioinformatischen Analyse und des cyanobakteriellen Screenings vergleichen zu können, muss zunächst die Auswertung des Screenings bewertet werden. Als DGR-assoziiert werden ~ 30 % (34) der Cyanobakterienstämme angesehen. In der Berechnung wurden Stämme, die nicht für die PCR zugänglich waren (keine Bande bei der 16S rDNA-PCR), ausgeschlossen. Da aber eine Korrelation zwischen der 16S rDNA-PCR und einer DGR-Präsenz sehr unwahrscheinlich ist, sollte es gerechtfertigt sein von 30 % DGR-assoziierten Cyanobakterienstämmen zu sprechen (werden die PCR-unzugänglichen Stämme einbezogen, sind 20,1 % der Cyanobakterienstämme mit DGRs assoziiert).

Bei den 113 PCR-zugänglichen Cyanobakterien handelt es sich nicht um 113 verschiedene Arten. Denn die Stammsammlung enthält zum einen gleiche Isolate einer Art und zum anderen nicht näher bestimmte Gattungen (Tabelle A1). Eine exakte Aussage über die Verteilung von DGRs in verschiedenen Cyanobakterienarten ist demnach nicht möglich. In Tabelle 8 ist aber zu sehen, dass 24 der 34 DGR-assoziierten Cyanobakterien definitiv zu verschiedenen Arten gehören. Demnach würde der Prozentsatz DGR-assoziiierter Cyanobakterienarten mindestens bei 21 % liegen. Da bei der bioinformatischen Analyse von Schillinger und Kollegen unterschiedliche Genome analysiert wurden, sollten für einen Vergleich mit dem cyanobakteriellen Screening ebenfalls verschiedene Cyanobakterienarten herangezogen werden. Selbst nach dieser Berechnung wären DGRs in der Stammsammlung mehr als dreifach stärker repräsentiert als in den öffentlichen Sequenzdatenbanken.

Da das geringe Vorkommen der DGRs in den öffentlichen Sequenzdatenbanken mit den konstanten Kultivierungsbedingungen der sequenzierten Stämme in Verbindung gebracht wurde (Schillinger & Zingler 2012), wurden die Kultivierungsbedingungen verschiedener Stammsammlungen miteinander verglichen (Tabelle 13). Dabei sind die Kultivierungsbedingungen der Stammsammlungen von Prof. Büdel (BB), der Roscoff Culture Collection (RCC) und der Sammlung für Algenkulturen in Göttingen (SAG) aufgeführt.

Tabelle 13: Kultivierungsbedingungen dreier Stammsammlungen

Stammsammlung	Kultivierungsbedingungen	Referenz
BB	Flüssigkultur ohne Luftdurchmischung bei 17 °C und 14 Stunden Licht/Tag (~ 60 $\mu\text{E}/\text{m}^2\cdot\text{s}$)	Hans Reichenberger
RCC	Flüssigkultur ohne Luftdurchmischung bei 21 °C und 16 Stunden Licht/Tag (~ 15 $\mu\text{E}/\text{m}^2\cdot\text{s}$)	Dr. Frederic Partensky
SAG	Schrägagarkultur bei 10 °C und 16 Stunden Licht/Tag (~ 50 $\mu\text{E}/\text{m}^2\cdot\text{s}$)	Dr. Maike Lorenz

Die Kultivierungsbedingungen der drei Stammsammlungen sind ausgelegt um ein Überleben der Zellen zu sichern, ohne das Wachstum zu fördern. Dabei unterscheiden sich die Kultivierungsbedingungen nicht stark voneinander. Sollten die konstanten Kultivierungsbedingungen zu einem Verlust der DGRs geführt haben, dann wäre entscheidend wie lange die Cyanobakterien vor der Sequenzierung in Kultur waren. Ein Beispiel ist das Cyanobakterium *Nostoc sp.* PCC7120, welches vor 1982 isoliert und dessen Sequenz 2001 publiziert wurde (Kaneko et al. 2001). Nach den ~ 20 Jahren, die das Cyanobakterium in Kultur war, ist das DGR noch immer im Genom vorhanden.

In der Stammsammlung von Prof. Büdel wurden über 75 % der PCR-zugänglichen Stämme mit bekanntem Isolationsdatum zwischen 1950 und 1979 isoliert (Tabelle A1). Beispiele hierfür sind die Umweltisolate von *Tolypothrix distorta* (BB97.17, BB97.18 und BB97.80), welche über DGR-kodierte RT-Sequenzen verfügen. Da die Stämme über einen langen Zeitraum kultiviert wurden ohne das DGR (die RT-Sequenz) zu verlieren und sich die Nukleotidsequenzen der RTs exakt gleichen (Kap. 3.1.1), spricht dies für stabil im Genom verankerte DGRs. Es wurden jedoch auch DGR-assoziierte Stämme gefunden, bei denen andere Isolate derselben Art keine Bande für die RT hervorgebracht haben (obwohl die 16S rDNA-PCR funktioniert hat) (Kap. 4.1.2). Diese Beispiele sprechen gegen stabil im Genom verankerte DGRs.

Wenn die bioinformatisch beobachtete geringe Inzidenz der DGRs nicht allein durch die Kultivierungsbedingungen zu erklären ist, könnte es auch möglich sein, dass nicht alle DGRs über das Programm (DiGreF) gefunden wurden. Die Parameter des Programms identifizieren ein DGR nur, wenn ein VR/TR-Paar vorliegt das höchstens 5 kb von der RT entfernt ist und mindestens 50 Nukleotide lang ist. Das VR/TR-Paar muss dabei, bis auf die variablen Positionen in der VR, eine identische Sequenz haben. Dementsprechend können Mutationen, die durch die natürliche Mutationsrate oder während der reversen Transkription entstanden sind, zu einem Entfernen des vermeintlichen DGRs aus der DiGreF-Liste führen. Desweiteren muss der TR mindestens über zehn Adenine verfügen und die VR mindestens sieben Adenin-spezifische Austausch vorzuweisen haben (Schillinger et al. 2012). Diese Parameter wurden

etabliert um inaktive DGRs oder deren Überreste aus der Suche auszuschließen. Es ist demnach möglich, dass im Screening genau solche Überreste identifiziert wurden, was aufgrund der mangelnden Sequenzinformation nicht nachvollziehbar ist. Gegen diese Vermutung spricht, dass die RT-Sequenzen in dem Aminosäure-Alignment recht konserviert vorliegen (Abbildung 7, S. 41). Aminosäuren, welche vom Konsensus abweichen, verfügen oft dennoch über dieselben physikalisch-chemischen Eigenschaften. Dadurch ergibt sich für die RT-Sequenzen eine Sequenzähnlichkeit von 76-97 % (Kap. 3.1.2). Auf DNA-Ebene ist ein geringerer Grad an Konservierung zu beobachten (60-97 %), wobei die meisten vom Konsensus abweichenden Positionen an den Wobble-Positionen der Kodons vorkommen. Das und die Tatsache, dass keine nonsense-Mutationen beobachtet werden konnten, sind starke Indizien für einen selektiven Druck auf die Funktionalität der untersuchten RTs und würden dafür sprechen, dass die identifizierten DGR-kodierten RT-Sequenzen zu funktionsfähigen DGRs gehören. Dies wurde exemplarisch für eine RT-Sequenz überprüft, indem die umgebende Sequenz identifiziert wurde (Kap. 4.1.3).

Der Unterschied zwischen den bioinformatisch und experimentell erlangten Ergebnissen in der Verteilung von DGRs könnte auch durch einen Bias in den Stichproben entstanden sein. Von den 34 DGR-assoziierten Stämmen in Tabelle 8 (S. 39) gehören 74 % zu der Ordnung *Nostocales*. Da die *Nostocales* 46 % der 113 analysierbaren Stämme ausmachen (Tabelle A1), ist diese Ordnung bei den DGR-assoziierten Stämmen überrepräsentiert. Unter der Annahme, dass kein verstärktes Vorkommen von DGRs in einer bestimmten cyanobakteriellen Ordnung vorliegt, wurde die Wahrscheinlichkeit der beobachteten Verteilung von Christine Pakebusch über eine Monte-Carlo-Simulation geprüft (Daten nicht gezeigt). Zunächst wurde der Chi-Quadrat-Wert berechnet um messen zu können, wie die beobachtete Verteilung der cyanobakteriellen Ordnungen in den DGR-assoziierten Stämmen von der Verteilung in den PCR-zugänglichen Stämmen abweicht. Dann wurden 10.000 Zufallsexperimente durchgeführt, bei denen jeweils 34 zufällige Stämme von den 113 PCR-zugänglichen Stämmen gezogen wurden. Für jedes Zufallsexperiment wurde ebenfalls der Chi-Quadrat-Wert berechnet. Der Anteil der Zufallsexperimente mit Chi-Quadrat-Werten größer oder gleich dem Wert der ursprünglich beobachteten Chi-Quadrat-Statistik entspricht dem p-Wert. Dieser liegt für die durchgeführte Simulation bei 0,07. Damit kann kein statistisch signifikanter Bias von DGRs zu einer bestimmten cyanobakteriellen Ordnung nachgewiesen werden.

Der potentielle Nutzen von DGRs für ihren Wirt macht eine starke Verbreitung wahrscheinlich. Das in dieser Arbeit durchgeführte Screening zeigt, dass DGRs in Cyanobakterien stärker verbreitet sind, als die bioinformatische Analyse von Schillinger und Kollegen vermuten lies (s.o.). Wodurch es zu dieser Diskrepanz kommt, ist mit den vorliegenden Daten nicht eindeutig zu ermitteln. Die in der bioinformatischen Analyse beobachtete geringe Inzidenz wurde zunächst mit den konstanten Kultivierungsbedingungen der sequenzierten Stämme in

Verbindung gebracht, da vermutet wird, dass diese zu einem Verlust von Virulenzfaktoren in Bakterien der Gattung *Shigella* geführt haben (Chosa et al. 1989; Schillinger & Zingler 2012). Da die Cyanobakterien im Screening aber ebenfalls über einen längeren Zeitraum kultiviert wurden, scheinen die konstanten Kultivierungsbedingungen nicht der alleinige Grund für einen Verlust der DGRs zu sein. Durch ein weiteres Screening, bei dem nicht kultivierte Cyanobakterien verwendet werden, könnte das Vorkommen von DGRs ohne den Aspekt der konstanten Kultivierungsbedingungen untersucht werden. Ein solches Screening sollte mit dem in dieser Arbeit etablierten Protokoll möglich sein (Kap. 4.1).

4.1.2 Die Verbreitung von DGRs: Weltweit und bevorzugt über horizontalen Gentransfer?

Cyanobakterien kommen in fast allen Habitaten, wie z.B. Wüsten, heißen Quellen, dem arktischen Eis und stark salzhaltigen Seen, vor (zusammengefasst in Wynn-Williams 2000; Oren 2000; Vincent 2000). Eine Möglichkeit für die Anpassung an solch extreme Bedingungen stellt die Symbiose mit einem Pilz (Flechte) dar. Dabei versorgen die Cyanobakterien den Pilz mit Kohlenstoff- und/oder Stickstoffverbindungen, welche sie über Photosynthese bzw. Stickstofffixierung gewinnen. Dafür liefert der Pilz Wasser und Mineralstoffe und schützt die Cyanobakterien durch das Pilzgeflecht vor Austrocknung, intensivem Licht und Frassfeinden (zusammengefasst in Adams 2000). Die extremen Bedingungen und die Notwendigkeit für Interaktionen mit anderen Organismen könnten Gründe sein, weshalb Cyanobakterien DGRs besitzen. Um dies näher zu untersuchen, wurden die Isolationsorte der DGR-assoziierten Stämme aus der Cyanobakteriensammlung von Prof. Büdel analysiert (Kap. 3.1.4). Die Analyse der geographischen Verteilung DGR-assoziiierter Stämme zeigt, dass DGRs nicht nur auf einem Kontinent, sondern wahrscheinlich weltweit vorkommen. Dabei kann kein Zusammenhang zwischen einer bestimmten Region (Klima) und dem Vorkommen von DGRs festgestellt werden (Tabelle 9).

Anschließend wurde das Auftreten von DGRs innerhalb der einzelnen Cyanobakterienarten untersucht, wobei eine ungleichmäßige Verteilung vorgefunden wurde. Das bedeutet, dass von DGR-assoziierten Stämmen auch Vertreter derselben Art gefunden wurden, die keine Bande für die RT aufweisen, obwohl sie für die PCR zugänglich sind (Tabelle A1). Beispiele hierfür sind u.a. in *Phormidium favosum*, *Scytonema capitatum*, *Scytonema stuposum*, *Stigonema mammilosum* und *Nostoc commune vaucher* zu finden. Da in solchen Fällen die PCR wiederholt wurde, sind Fehler bei der Durchführung der PCR zwar nicht auszuschließen, aber unwahrscheinlich. Mutationen in den Primerbindestellen sind ebenfalls unwahrscheinlich, da die natürliche Mutationsrate gering ist (keine Abweichung zwischen den Nukleotid-

sequenzen dreier *Tolypothrix distorta*-Isolate von 1963 bzw. 1970) (Kap. 3.1.1, Tabelle A1). Eine mögliche Erklärung wäre eine fehlerhafte taxonomische Bestimmung der Art, die sich für Cyanobakterien als besonders schwierig erweist (zusammengefasst in Rikkinen 2013). In diesem Fall könnten DGRs eventuell als genetische Marker eingesetzt werden, um eine genauere Bestimmung verwandter Arten zu ermöglichen.

Wenn sich die ungleichmäßige Verteilung der DGRs in einer Cyanobakterienart aber durch den Verlust von DGRs in einigen Stämmen ergeben hat, dann spricht das für Elemente, die nicht besonders stabil im Genom verankert sind. In diesem Fall würden DGRs für die Verbreitung besonders auf den horizontalen Gentransfer angewiesen sein. Ein horizontaler Gentransfer über Phagen, Transposons und Plasmide wurde für einige DGRs bereits vermutet (Schillinger et al. 2012; Arambula et al. 2013).

Da außer dem sequenzierten Bereich der RT keine Information über die genetische Umgebung der DGRs bekannt ist, lässt sich keine Aussage über eventuell auftretende genomische Inseln machen, welche Hinweise auf einen speziesunspezifischen horizontalen Gentransfer geben könnten. Ein in die Übertragung der DGRs involvierter Phage sollte sich in dem cyanobakteriellen Subtree des phylogenetischen Baums der DGR-kodierten RTs eingliedern (Abbildung 8). Dies kann nicht beobachtet werden, könnte aber damit zusammenhängen, dass der entsprechende Phage noch nicht sequenziert wurde.

Der phylogenetische Baum wurde aus Sequenzen konstruiert, die eine Länge von etwa 80 Aminosäuren aufweisen. Dabei sind neben den Sequenzen des cyanobakteriellen Screenings noch weitere Sequenzen vorhanden. Diese wurden aus dem phylogenetischen Baum von Schillinger und Kollegen übernommen und nach einem Alignment angepasst (Schillinger et al. 2012). Die Anordnung der RTs in beiden Bäumen ähnelt sich stark. Deshalb kann der in dieser Arbeit konstruierte Baum, trotz der geringen Sequenzlänge, als zuverlässig angesehen werden.

Die Verwandtschaft der cyanobakteriellen DGRs äußert sich an der Gruppierung der cyanobakteriellen RT-Sequenzen aus der Datenbank und dem Screening in einem Subtree (Abbildung 8). In diesem cyanobakteriellen Subtree befinden sich zudem zwei grüne Schwefelbakterien (*Chlorobium phaeobacteroides* und *Pelodictyon phaeoclathratiforme* (GI: 194337212)), ein Schwefelpurpurbakterium (*Thiorhodovibrio* sp.) und ein Nichtschwefelpurpurbakterium (*Rhodomicrobium vannieli*).

Im phylogenetischen Baum der DGR-kodierten RTs sind noch weitere Schwefelpurpurbakterien (*Thiorhodospira sibirica* und *Thiorhodococcus drewsii*), Nichtschwefelpurpurbakterien (*Rhodospirillum centenum* und *Rhodobacter sphaeroides*) und grüne Schwefelbakterien (*Pelodictyon phaeoclathratiforme* (GI: 194337359) und *Chlorobium chloromati*) aufgeführt, die jedoch über den kompletten Baum verteilt sind (Abbildung A2). Es gibt noch weitere sequenzierte Stämme von Purpurbakterien und grünen Schwefelbakterien welche in

der Genom-Liste von NCBI (National Center for Biotechnology Information) einzusehen sind, aber nicht mit DGRs assoziiert sind. Diese Beobachtungen lassen einen horizontalen Gentransfer der entsprechenden DGRs vermuten.

Sollte der DGR-Transfer zwischen Cyanobakterien und grünen Schwefelbakterien/Purpurbakterien vertikal verlaufen sein, dann gibt es DGRs seit ~ 3,5 Mrd. Jahren. Denn die anoxygenen phototrophen Organismen (grüne Schwefelbakterien, Purpurbakterien) werden als Vorfahren der oxygenen phototrophen Organismen (Cyanobakterien) angesehen, welche seit ~ 3,5 Mrd. Jahren existieren (Taylor & Taylor 1993; Xiong, Inoue & Bauer 1998).

Die Verteilung der DGRs in der cyanobakteriellen Stammsammlung und die Anordnung der Organismen im phylogenetischen Baum der DGR-kodierten RTs lassen darauf schließen, dass DGRs weltweit verbreitet sind und die Übertragung oft über horizontalen Gentransfer erfolgt. Eine Sequenzierung der genetischen Umgebung von den DGRs könnte dabei helfen Vektoren für den Gentransfer zu identifizieren.

4.1.3 Das neu entdeckte DGR aus *Anabaena flos-aquae* könnte durch den Wirt reguliert sein

Durch das cyanobakterielle Screening konnte eine Vielzahl an DGR-kodierten RT-Sequenzen identifiziert werden, für die vermutet wird, dass sie zu funktionsfähigen DGRs gehören (Kap. 4.1.1). Um diese Vermutung zu überprüfen, wurde die umgebende Sequenz des RT-Fragments aus dem Cyanobakterium *Anabaena flos-aquae* exemplarisch aufgeklärt. Dafür wurde die inverse PCR als Methode verwendet (Abbildung 9, Abbildung 10). Diese Methode sollte als kostengünstige Alternative zu den angebotenen Genomsequenzierungen getestet werden. Durch die inverse PCR konnte ein neues DGR in *Anabaena flos-aquae* identifiziert werden, welches ca. 2900 bp groß ist und aus einem Zielgen mit der VR am 3'-Ende, einer Avd, einer RT und dem TR besteht (Abbildung 11). Diese Komponenten weisen charakteristische Eigenschaften von DGRs auf. Dazu gehören das GGW-Motiv am C-terminalen Ende der Avd (Abbildung 13), die typischen Domänen DGR-kodierter RTs mit dem SQ-Motiv in Domäne IV (Abbildung 12) und die Adenin-spezifischen Unterschiede zwischen VR und TR (Abbildung 14). Ein Vergleich der Komponenten mit den entsprechenden Komponenten aus dem Cyanobakterium *Nostoc sp.* PCC7120 und dem *Bordetella*-Phagen, zeigt eine jeweils höhere Sequenzähnlichkeit zu dem verwandten Cyanobakterium. Dies bestätigt die Ergebnisse des phylogenetischen Baums, welcher die RT-Sequenzen der Cyanobakterien in einem Subtree gruppiert (Abbildung 8).

Das Zielgen besitzt, wie ca. 27 % der DGR-Zielgene, eine Homologie zu den Formylglycin-generierenden Enzymen (FGEs) (Schillinger & Zingler 2012). FGE-Proteine katalysieren die Umwandlung von Cystein zu Formylglycin im aktiven Zentrum von Sulfatasen (Dierks et al. 2005). Daher wäre eine katalytische Aktivität auch für einige variable Proteine der DGRs denkbar (Le Coq & Ghosh 2011). Im Kontext der DGR-Funktion ist es aber wahrscheinlicher, dass eine Gemeinsamkeit zwischen den FGE-Proteinen und den variablen DGR-Proteinen in der Erkennung von Proteinen liegt (Schillinger & Zingler 2012). Eine strukturelle Gemeinsamkeit besteht in der Ausbildung einer CLec-Faltung des FGE-Typs (Le Coq & Ghosh 2011). Die CLec-Faltung ist in DGRs konserviert und dient dazu, die enorme Sequenzdiversität der variablen Proteine zu tolerieren und diese an der Oberfläche zu präsentieren (Medhekar & Miller 2007; McMahon et al. 2005). Das variable Protein aus *Anabaena flos-aquae* kann dabei theoretisch bis zu $1,6 \times 10^{18}$ Proteinsequenzen generieren (Kap. 3.1.5). Damit übertrifft es die Diversität des Mtd-Proteins ($9,2 \times 10^{12}$ Proteinsequenzen) aus dem Modell-DGR des *Bordetella*-Phagen um fünf Zehnerpotenzen (Tabelle 1).

Die hohe Proteindiversität entsteht durch die Anzahl und Anordnung der variablen Positionen in der VR. Dabei sind 26 der 35 variablen Positionen in NNY-Kodons involviert (N = variable Position, Y = C, T). Aus den NNC- und NNT-Kodons können zum einen keine Stoppkodons gebildet werden (UAA, UAG und UGA) und zum anderen kodieren sie jeweils für 15 verschiedene Aminosäuren. Diese häufig anzufindende Anordnung von variablen Positionen an den ersten beiden Positionen eines Kodons wurde bereits beschrieben und zeigt die Optimierung der Sequenzdiversifizierung durch das Element (Medhekar & Miller 2007). In der VR von *Anabaena flos-aquae* wurden nur 15 der 35 variablen Positionen mit anderen Nukleotiden substituiert. Das bedeutet aber nicht, dass das Potential zur Diversifizierung nicht ausgeschöpft wurde. Es ist wahrscheinlicher, dass die momentane VR den Zustand repräsentiert, der am besten an die Funktion angepasst ist und sich durchgesetzt hat. Über die Funktion des variablen Proteins ist aber nichts bekannt.

Um zu testen, ob eine Hypermutation des variablen Proteins unter den Kultivierungsbedingungen im Labor überhaupt möglich ist, wurden die Transkripte der RT, der Avd und des TR über RT-PCR nachgewiesen (Abbildung 15, Abbildung 16). Die RT-PCR wurde ebenfalls für das DGR aus kultivierten *Nostoc sp.* PCC7120 durchgeführt, dessen Aufbau in Abbildung 4 einzusehen ist. Die meisten Transkripte konnten problemlos unter den Standardbedingungen nachgewiesen werden. Eine Ausnahme ist die RT aus *Anabaena flos-aquae*, welche erst mit erhöhter Anzahl an PCR-Amplifikationszyklen (30) als schwache Bande zu detektieren war. Anhand der RT-PCR-Analysen ist aber auch zu erkennen, dass die Transkriptmenge der RT aus *Anabaena flos-aquae* nach einem dreistündigen Kältestress bei 4 °C erhöht vorliegt. Dabei könnte der Kältestress direkt mit der Funktion des variablen

Proteins zusammenhängen. Eine andere Möglichkeit wäre eine Erhöhung der Transkriptmenge aufgrund einer generellen Stressantwort. So wurde die Primärantwort auf Kältestress in dem Cyanobakterium *Synechocystis* mit Microarray-Analysen untersucht und dabei über 100 Kältestress-induzierte Gene identifiziert (Suzuki et al. 2001). Auf jeden Fall scheint die Aktivität des DGRs aus *Anabaena flos-aquae* über die Transkriptmenge der RT regulierbar zu sein. Dadurch könnte eine Hypermutation des variablen Proteins auf Situationen begrenzt werden, in denen eine Anpassung an neue Umweltbedingungen sinnvoll ist. Das bringt einen Vorteil für den Wirt, da Mutationen auch zur Inaktivierung funktioneller Proteine führen können (Kap. 4.2).

Die Frage, ob die RT-Fragmente des cyanobakteriellen Screenings zu kompletten DGRs gehören, oder es sich dabei nur um Überreste von DGRs handelt, konnte für das RT-Fragment aus *Anabaena flos-aquae* eindeutig geklärt werden. Das identifizierte DGR besitzt alle benötigten Komponenten, wobei keine offensichtlich inaktivierenden Mutationen beobachtet werden konnten. Da auch der variable Bereich der VR keine Adenin-unspezifischen Mutationen zeigt (Abbildung 14), sollte das DGR vor nicht zu langer Zeit aktiv gewesen sein. Ansonsten wären durch die natürliche Mutationsrate zufällige Mutationen in der VR entstanden. Da der Cyanobakterienstamm (BB97.35) 1965 isoliert wurde, ist das DGR seit ca. 50 Jahren im Genom der kultivierten *Anabaena flos-aquae* vorhanden (Tabelle A1). Diese Beobachtungen sprechen für ein stabiles und funktionsfähiges DGR.

Um die Sequenz des DGRs aus *Anabaena flos-aquae* zu identifizieren, wurde eine inverse PCR auf das RT-Fragment durchgeführt. Die inverse PCR ist zwar kostengünstig, die Länge der erhaltenen Sequenzinformation ist aber zufallsbedingt und hängt von den verwendeten Restriktionsenzymen ab. Bezogen auf das neu identifizierte DGR bedeutet dies, dass theoretisch weitere Zielgene existieren könnten, welche außerhalb des identifizierten Bereichs lokalisiert sind. Eine Analyse der um das DGR liegenden Sequenz ist ebenfalls nicht möglich, könnte aber Hinweise auf einen horizontalen Gentransfer geben, der eine Möglichkeit für die Übertragung von DGRs darstellt (Kap. 4.1.2). Diese Probleme würden bei einer Genomsequenzierung nicht auftreten.

4.2 Aktivitätsanalysen bakterieller DGRs

Unter der DGR-Aktivität wird die Fähigkeit des Elements verstanden, die VR des Zielgens zu hypermutieren. Um dies für die untersuchten Elemente bestätigen zu können und einen Eindruck darüber zu gewinnen, wie häufig eine Diversifizierung der VR auftritt, wurde die VR zu unterschiedlichen Zeitpunkten analysiert.

4.2.1 Analysen cyanobakterieller DGRs zeigen Aktivität in *Nostoc sp.*

PCC7120

Die DGRs aus *Anabaena flos-aquae* (Abbildung 11) und *Nostoc sp.* PCC7120 (Abbildung 4) sind auf DNA-Ebene theoretisch funktionsfähig. Das bedeutet, dass alle benötigten Komponenten vorliegen, diese keine (offensichtlich) inaktivierenden Mutationen beinhalten und im variablen Bereich der VR keine Adenin-unspezifischen Mutationen vorzufinden sind, welche durch die natürliche Mutationsrate in inaktiven Elementen auftreten würden (Kap. 3.1.5 und Doulatov et al. 2004).

Für beide DGRs wurden zudem Bedingungen gefunden, unter denen die Transkripte der Komponenten RT, Avd und TR nachweislich vorliegen (Kap. 3.1.5.1). Die Aktivitätsanalysen wurden mit Cyanobakterien durchgeführt, die diesen Bedingungen ausgesetzt waren (Standardbedingungen für *Nostoc sp.* PCC7120 (22 °C) und ein dreistündiger Kältestress für *Anabaena flos-aquae* (4 °C)). Dabei wurde die isolierte gDNA von ca. 10^7 Zellen mit dem REBVA analysiert (Kap. 2.2.2.4).

Aktivitätsanalyse des DGRs aus Anabaena flos-aquae

Für die Untersuchung der VR von dem DGR aus *Anabaena flos-aquae* wurde AclI als Restriktionsenzym verwendet. Dieses lässt die Analyse von vier variablen Positionen zu (Kap. 2.2.2.4). Der REBVA, welcher zur Selektionierung mutierter VR-Sequenzen dient, wurde mehrmals begonnen, konnte aber nicht komplett durchgeführt werden (Kap. 3.2.2). Die Schwierigkeit bestand darin, das PCR-Produkt der VR im PCR/Restriktionsverdau-Zyklus weiterhin amplifizieren zu können. Die Durchführung der ersten Zyklen konnte durch die Verwendung neuer, eingerückter Primer (nested PCR) verbessert werden. Nach einigen Zyklen brachte die nested PCR aber auch keine PCR-Produkte mehr hervor. Die Sequenzierung von PCR-Produkten aus mehreren, unvollständig durchgeführten REBVAs hat für die VR der kältegestressten *Anabaena flos-aquae* keine Diversifizierung ergeben (Kap. 3.2.2). Dementsprechend funktioniert entweder die Anreicherung mutierter Sequenzen durch den REBVA nicht, oder es liegen keine mutierten Sequenzen vor.

Für *Anabaena flos-aquae* wurden neben dem Kältestress auch die Standardbedingungen analysiert. Unter Standardbedingungen liegen nur sehr wenige RT-Transkripte vor (Kap. 3.1.5.1). Daher wurden die Standardbedingungen analysiert, um zu testen, ob die geringe Anzahl von RT-Molekülen für eine Diversifizierung ausreicht. Auch für die Standardbedingungen konnte der REBVA nicht komplett durchgeführt werden. Um zu prüfen, ob der REBVA nicht funktioniert weil keine mutierten VR-Sequenzen vorliegen, wurde die verwendete Polymerase gewechselt. Anstelle der Phusion DNA-Polymerase (mit

Korrekturlesefunktion) wurde die MyTaq DNA-Polymerase (ohne Korrekturlesefunktion) verwendet. Durch die fehlende Korrekturlesefunktion der Polymerase können Mutationen in der Erkennungssequenz des Restriktionsenzym entstehen, welche in einem funktionierenden REBVA selektioniert werden sollten. Mit der MyTaq DNA-Polymerase konnte der REBVA komplett durchgeführt werden und das sequenzierte PCR-Produkt wies eine Punktmutation in der Erkennungssequenz des Restriktionsenzym auf (jedoch keine DGR-vermittelte Hypermuation). Demnach eignet sich der REBVA für die Anreicherung mutierter Sequenzen.

Da der REBVA für die Kulturen, welche Standardbedingungen bzw. Kältestress ausgesetzt waren, nicht komplett durchgeführt werden konnte, spricht dies gegen eine Variabilität der VR und gegen eine Aktivität des DGRs. Die Zeitspanne von der Sequenzierung des DGRs bis zur Durchführung des REBVA beträgt in etwa zwei Jahre. Das in dieser Zeit keine detektierbare Hypermuation der VR aufgetreten ist, könnte mit der geringen Anzahl transkribierter RT-Moleküle unter den Standardbedingungen zusammenhängen. Demnach könnte die schwache Transkription eine effektive Regulation des DGRs durch den Wirt darstellen (Kap. 4.2.4). Nach dem dreistündigen Kältestress liegt die RT aber deutlich detektierbar vor. Das der zugehörige REBVA ebenfalls nicht komplett durchgeführt werden konnte (keine mutierten Sequenzen vorlagen), könnte entweder bedeuten das die drei Stunden nicht ausreichen um eine Diversifizierung auszulösen oder das weitere Mechanismen für eine Regulation existieren.

Aktivitätsanalyse des DGRs aus Nostoc sp. PCC7120

Für die Untersuchung der VR2 von dem analysierten DGR aus *Nostoc sp. PCC7120* (RT: *alr3497*) wurde SacI als Restriktionsenzym verwendet. Dieses lässt die Analyse von vier variablen Positionen zu (Kap. 2.2.2.4). Die Untersuchung mit dem REBVA wurde mehrmals begonnen, konnte aber nur einmal komplett durchgeführt werden (Kap. 3.2.1). Die Sequenzierung des PCR-Produkts erbrachte die VR-Sequenzen R1 und R2 (Abbildung 18).

Die VR-Sequenz R1 besitzt fünf Abweichungen zu der VR-Sequenz, die vor der Kultivierung von *Nostoc sp. PCC7120* im Labor (Molekulare Genetik, TU Kaiserslautern) festgestellt wurde. Alle fünf Abweichungen sind an variablen Positionen (Adeninen an den korrespondierenden Positionen im TR) lokalisiert, was eine Diversifizierung durch das DGR wahrscheinlich macht. Die Abweichungen betreffen einen 15 bp langen Bereich, der die Erkennungssequenz des Restriktionsenzym mit einschließt und verändert. Daher konnte die Sequenz im REBVA angereichert werden. Ein Modell des mutagenen Homings, welches die DGR-charakteristischen Mutationen in der VR erklärt, ist in Abbildung 34 gezeigt.

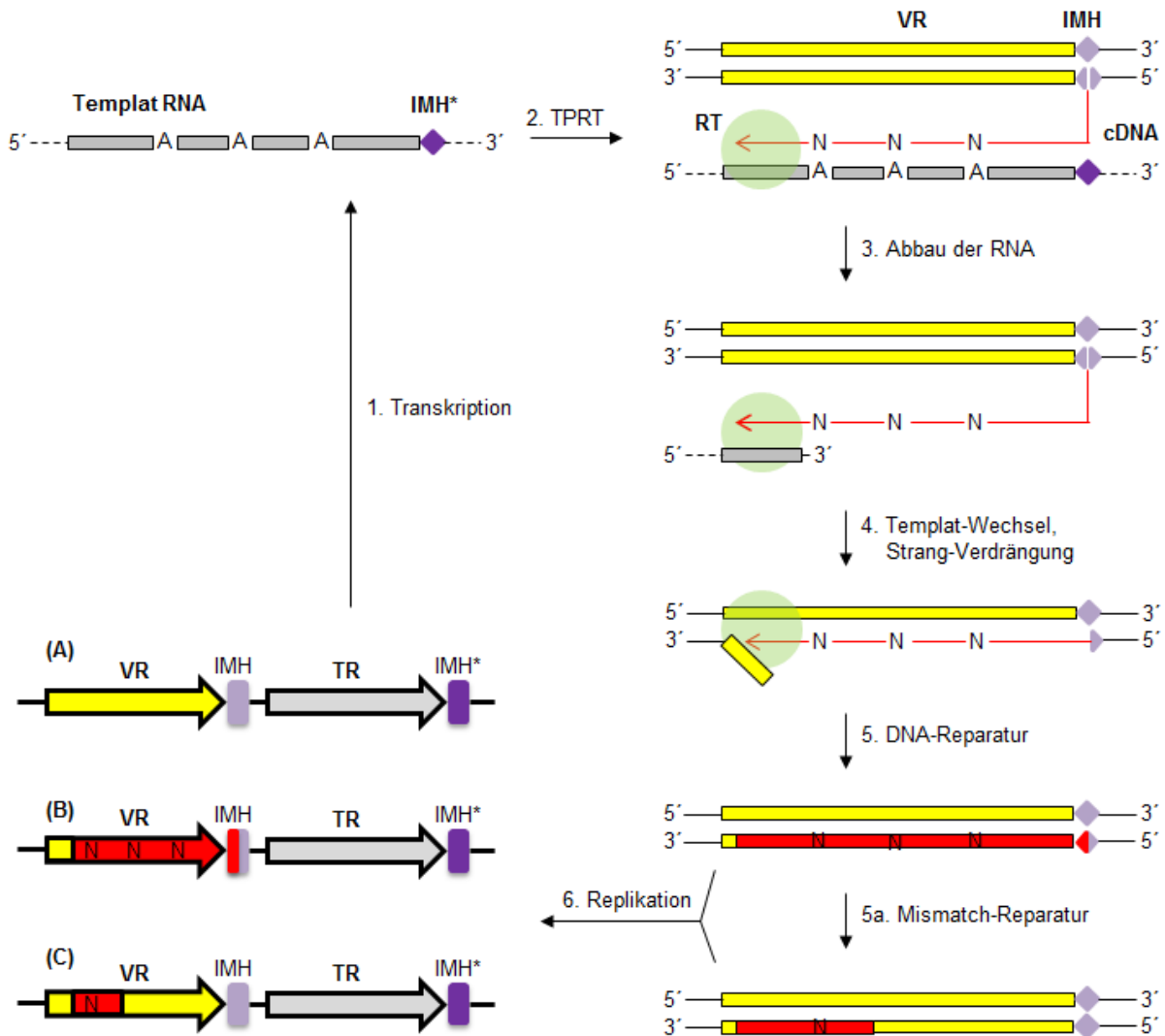


Abbildung 34: Modell des mutagenen Homings. Im mutagenen Homing wird die VR des Zielgens durch eine veränderte VR ersetzt. Dafür wird der TR transkribiert (1) und die entstehende Templat RNA von der DGR-kodierten RT revers transkribiert, wobei es vermutlich zu einer fehlerhaften Erkennung der Adenine kommt. Die Initiierung der cDNA-Synthese am 3'-Ende der VR (IMH) könnte über einen TPRT-Mechanismus stattfinden, der durch eine bislang nicht identifizierte Endonuklease induziert wird (2). Der Abbau der Templat RNA könnte durch eine zelluläre RNase H erfolgen (3). Durch einen sequenzhomologen Bereich zwischen Templat RNA und VR könnte es zu einem Templat-Wechsel der DGR-kodierten RT zu dem Plusstrang der VR kommen (4). Die Befestigung des 5'-Endes der cDNA könnte über DNA-Reparaturmechanismen erfolgen (5). Die resultierenden Fehlpaarungen zwischen den DNA-Strängen der VR können über die DNA-Replikation behoben werden (6). Dabei entstehen eine unveränderte und eine mutierte Version der VR (A, B). Durch Mismatch-Reparatur oder einen analogen Prozess (5a) vor der Replikation könnten kleinere Diversitätsbereiche in der VR entstehen (C). Da die Sequenz bei dem „copy and replace“ Mechanismus des DGRs wiederhergestellt wird, sind iterative Zyklen an Diversifizierung möglich. N: Jedes der vier Nucleotide (modifiziert nach Guo et al. 2008).

Abbildung 34 skizziert ein Modell für das mutagenes Homing, welches durch die Wiederherstellung der Zielsequenz iterative Zyklen an Sequenzdiversifizierung erlaubt. Diese Möglichkeit zur ständigen „Neuanpassung“ des variablen Proteins ist vorteilhaft für den Wirt. Als mobiles Element dient das Transkript des TR (Templat RNA). Da die DGR-charakteristischen Mutationen nicht in der Templat RNA vorkommen, scheinen diese durch eine fehlerhafte Adenin-Erkennung der DGR-kodierten RT bei der cDNA-Erststrangsynthese zu entstehen (Kap. 1.1.2.1). Die cDNA-Synthese benötigt jedoch einen Primer, für dessen Bereitstellung mehrere Mechanismen postuliert wurden (Medhekar & Miller 2007):

1. Ein bislang nicht identifizierter Primer könnte die cDNA-Synthese unabhängig von der VR-Sequenz initiieren (kein TPRT-Mechanismus).
2. Ein Okazaki-Fragment innerhalb der Replikationsgabel könnte als Primer verwendet werden. Dies wurde für Gruppe II Introns ohne Endonuklease-Domäne beobachtet (Martínez-Abarca et al. 2004). Dabei kommt es zu einem strangspezifischen Bias, der für das DGR des *Bordetella*-Phagen aber nicht festgestellt wurde (Guo et al. 2011).
3. Ein bislang nicht identifizierter Faktor/Endonuklease könnte einen Einzelstrangbruch/Doppelstrangbruch in der VR bewirken, wonach die entstehende 3'-Hydroxylgruppe des Minusstrangs als Primer verwendet werden kann. Zu den Retroelementen die eine Endonuklease im Retrohoming verwenden, gehören auch die Gruppe II Introns und non-LTR Retrotransposons (Kap. 1.1.1; Gasior et al. 2006; Christensen & Eickbush 2005).

In Abbildung 34 ist aufgrund der engen evolutionären Verwandtschaft zwischen DGRs und Gruppe II Introns ein Endonuklease-induzierter TPRT-Mechanismus für das Retrohoming postuliert (Doulatov et al. 2004). Ein möglicher Kandidat für die Endonuklease wäre eine katalytisch aktive Templat RNA. Diese könnte zusammen mit der DGR-kodierten RT und der Avd (nukleinsäurebindend, spezifische Bindung zur DGR-kodierten RT) einen Homing-Komplex bilden, der analog zu den Gruppe IIC Introns, die Haarnadel/Kreuz-Struktur downstream der VR erkennt (Alayyoubi et al. 2013; Guo et al. 2011; Robart, Seo & Zimmerly 2007). Anschließend erfolgt die Initiierung der cDNA-Synthese, welche im *Bordetella*-Phagen innerhalb der GC-reichen Region der IMH stattzufinden scheint. Dies zeigen Versuche, bei denen Marker in den TR eingebracht wurden, die nach einem Sequenztransfer nur dann in der VR auftauchen, wenn sie zwischen der 3'- und 5'-cDNA-Integrationsstelle lokalisiert sind (Marker-Coconversion) (Guo et al. 2008). Die Ergebnisse der Marker-Coconversion stimmen mit der Beobachtung überein, dass Adenine in der upstream gelegenen Sequenz der IMH nicht variabel sind.

Die cDNA-Integration von Retroelementen kann über homologe Rekombination stattfinden (Ke & Voytas 1997). Die zentrale Reaktion bei der Rekombination ist der Austausch zwischen zwei homologen DNA-Molekülen, welcher durch RecA-Proteine katalysiert wird (Chen, Yang & Pavletich 2008). Aufgrund der hohen TR/VR-Homologie ist die homologe Rekombination für DGRs ebenfalls denkbar. Da ein DGR-Sequenztransfer aber in RecA-Knockout-Mutanten nachgewiesen wurde, ist das mutagene Homing RecA-unabhängig (Guo et al. 2008).

Ein anderes Modell für die Integration des 5'-Endes der cDNA wurde für das non-LTR Retroelement R2 aus *Bombyx mori* beschrieben, bei dem die cDNA-Synthese ebenfalls über einen Endonuklease-induzierten TPRT-Mechanismus stattfindet. In dem Modell kann die R2-RT, nachdem die RNA revers transkribiert wurde, in den geschnittenen Plusstrang der DNA springen (Templat-Wechsel) und dort die Zweitstrangsynthese unter Verdrängung des RNA-

Templats durchführen (Kurzynska-Kokorniak et al. 2007). Dem Modell liegt zugrunde, dass die R2-RT in der Lage ist vom Ende des RNA-Templats zu doppelsträngigen DNA-Templaten zu wechseln, was durch Mikrohomologien (bis 5 bp) gefördert wird (Bibillo & Eickbush 2002; Bibillo & Eickbush 2004; Kurzynska-Kokorniak et al. 2007).

Auch bei DGRs wird ein sequenzhomologer Bereich (4-12 bp) zwischen VR und TR für die cDNA-Integration am 5'-Ende der VR benötigt. Zudem konnte die Marker-Coconversion zeigen, dass kein exakt definierter Integrationsbereich existiert (Guo et al. 2008). Es ist demnach denkbar, dass die DGR-kodierte RT von der Templat RNA zum Plusstrang der VR springen kann. Um dies zu testen, könnte die Elongation eines markierten Primers auf einem Donor-Templat in Anwesenheit eines Akzeptor-Templats untersucht werden. Dies setzt jedoch eine aufgereinigte, enzymatisch aktive DGR-kodierte RT voraus.

Eine Zweitstrangsynthese durch die DGR-kodierte RT ist dagegen unwahrscheinlich, da es durch die vermutete Adenin-spezifische Fehleranfälligkeit auch zu einer Variabilität der Thyminen kommen müsste (Thyminen im TR werden zu Adeninen im Erstrang und würden zu variablen Positionen im Zweitstrang führen). Wahrscheinlicher ist das cDNA-Erststrang und DNA-Plusstrang am sequenzhomologen Bereich miteinander annealen. Dafür müsste die Templat RNA jedoch abgebaut sein. Retrovirale RTs bauen das RNA-Templat während der reversen Transkription mit einer RNase H-Domäne ab, die daher wichtig für den Templat-Wechsel ist (Svarovskaia et al. 2000; Wisniewski et al. 2000). DGR-kodierte RTs besitzen eine solche RNase H-Domäne nicht. Diese Aufgabe könnte daher von einer zellulären RNase H übernommen werden.

Nach dem Annealen von cDNA-Erststrang und DNA-Plusstrang, könnte die Integration des 5'-Endes der cDNA in den Minusstrang über DNA-Reparaturmechanismen erfolgen. Zwischen den DNA-Strängen entstehen dann aufgrund der Mutationen Fehlpaarungen im Bereich der VR. Diese könnten während der DNA-Replikation behoben werden, wobei eine unveränderte und eine mutierte Version der VR entstehen (Abbildung 34, VRs A und B).

Die in *Nostoc sp.* PCC7120 beobachtete VR-Sequenz R1 weist nur einen kleinen mutierten Bereich am 5'-Ende der VR auf (Abbildung 18). Ein solches Mosaikmuster, bei dem die VR aus mutierten und unmutierten Sequenzen zusammengesetzt ist, wurde bereits beobachtet (Doulatov et al. 2004). Eine Erklärung wäre, dass eine Mismatch-Reparatur oder ein analoger Prozess fehlgepaarte Bereiche vor der Replikation korrigiert (Abbildung 34, VR C).

Die VR-Sequenz R2 besitzt eine Deletion von 67 bp, welche die Erkennungssequenz des im REBVA verwendeten Restriktionsenzym beinhaltet (Abbildung 18). Bei der Deletion handelt es sich um einen Bereich zwischen zwei direkten Repeats (5'-GCGGGGCG-3, Nukleotide 3-10 und 70-77 in der VR) inklusive einem Repeat. Da ein Templat-Wechsel der DGR-kodierten

RT in das mutagene Homing involviert sein könnte und ein intramolekularer Templat-Wechsel zu einer Deletion führen kann, wurde dies in Betracht gezogen (Abbildung 35).

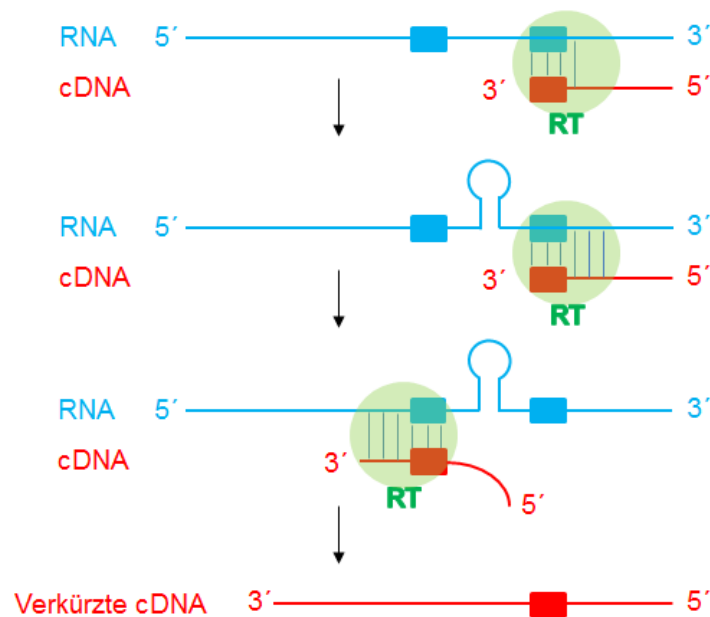


Abbildung 35: Intramolekularer Templat-Wechsel von reversen Transkriptasen. Eine RNA mit zwei direkten Repeats (Boxen) besitzt das Potential für einen Templat-Wechsel der RT, der von einem Repeat zum anderen erfolgt. Die Wahrscheinlichkeit für einen Templat-Wechsel wird durch die Ausbildung einer stabilen Haarnadelstruktur zwischen den Repeats erhöht. In dem entstehenden cDNA-Molekül ist einer der Repeats und die Sequenz zwischen den Repeats deletiert (modifiziert nach Cocquet et al. 2006).

Die Wahrscheinlichkeit für einen intramolekularen Templat-Wechsel wird durch direkte Repeats und die Ausbildung einer stabilen Haarnadelstruktur zwischen den Repeats erhöht (zusammengefasst in Cocquet et al. 2006). In dieser Arbeit wurde eine Sekundärstruktur der Templat RNA bestimmt, die eine stabile Haarnadelstruktur zwischen den Repeats zeigt (Kap. 3.3.3). In der Templat RNA kommt ein „variables“ Adenin in einem der Repeats vor (Guanin in der VR), weshalb an dieser Stelle ein Mismatch existiert (Abbildung 36).

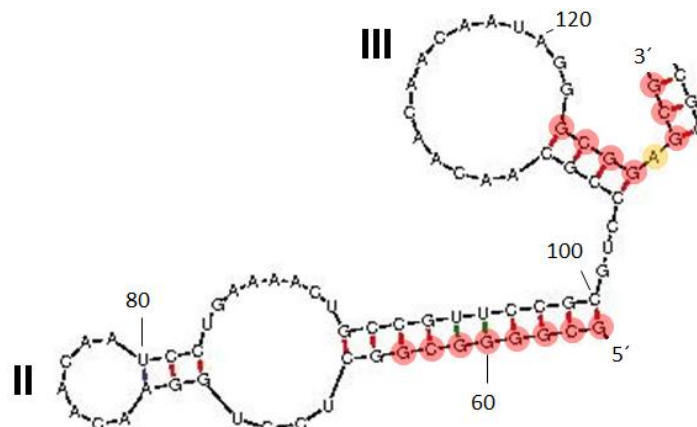


Abbildung 36: Position der direkten Repeats in der Templat RNA. Ein Ausschnitt der identifizierten Templat RNA-Sekundärstruktur ist gezeigt (s. Abbildung 30). Zwischen den direkten Repeats (rot) gibt es einen Mismatch (Adenin, Orange) der in der VR nicht existiert, da das „variable“ Adenin dort mit einem Guanin substituiert wurde. Zwischen den (unvollständigen) Repeats bildet sich eine stabile Haarnadelstruktur aus (II).

Der Mismatch zwischen den Repeats in der Templat RNA und die Tatsache, dass keine der 12 verbleibenden variablen Positionen in der R2-Sequenz Variabilität aufweist (Abbildung 18, verglichen mit der VR-Sequenz vor der Kultivierung von *Nostoc sp.* PCC7120), spricht gegen DGR-Aktivität. Da fast jedes DGR-Retrohoming (> 90 %) von Mutationen an den variablen Positionen begleitet wird, sollten diese auch in der R2-Sequenz zu finden sein (Guo et al. 2008). Die Deletion könnte alternativ durch einen intramolekularen Templat-Wechsel während der PCR im REBVA entstanden sein. Dem Hersteller (Fisher Scientific, Schwerte) der verwendeten Phusion DNA-Polymerase ist nicht bekannt, ob diese einen Templat-Wechsel durchführen kann (persönliche Mitteilung Dr. Isabel Molina), an anderer Stelle wurde es aber bereits vermutet (Gao et al. 2012).

Im Laufe von zwei Jahren konnte in dem DGR aus *Nostoc sp.* PCC7120 nur eine VR-Sequenz (R1) beobachtet werden, die sich durch Adenin-spezifische Mutationen von der VR-Sequenz unterscheidet, die vor der Kultivierung des Cyanobakteriums im Labor (Molekulare Genetik, TU Kaiserslautern) festgestellt wurde. Die Zellen von *Nostoc sp.* PCC7120 wurden vor der zweijährigen Kultivierung durch eine Verdünnungsreihe vereinzelt. Daher sollte die VR-Sequenz R1 ein Aktivitätsereignis des DGRs repräsentieren, das innerhalb dieser zwei Jahre aufgetreten ist. Die zu Beginn identifizierte VR-Sequenz entspricht auch der 2001 publizierten Sequenz (Kaneko et al. 2001), weshalb eine Diversifizierung nicht sehr frequent aufzutreten scheint. Das würde auch erklären, weshalb in dem DGR aus *Anabaena flos-quaе* innerhalb von zwei Jahren keine Aktivität festgestellt werden konnte (s.o.). Eine Erklärung für die seltene Diversifizierung könnten die konstanten Kultivierungsbedingungen darstellen. Dadurch fehlt ein Anpassungsdruck auf sich ändernde Umweltbedingungen, die eine Diversifizierung notwendig machen. Hinzu kommen fehlende Interaktionen mit anderen Mikroorganismen bzw. Symbiosen mit Pilzen und Farnen (z.B. *Anabaena flos-aquae* und der Algenfarn *Azolla*) (Cole et al. 2014; Stewart, Rowell & Rai 1983; Peters & Mayne 1974). Symbiosen können eine Veränderung der transkriptionalen Organisation bewirken, die neben einer zellulären Differenzierung (vegetative Zellen zu stickstofffixierenden Heterocysten) auch zu einer „Aktivierung“ des DGRs führen könnte (Mitschke et al. 2011; Meeks & Elhai 2002). Diese Form der Regulation könnte die Diversifizierung mit dem Bedarf koppeln (Kap. 4.2.4).

4.2.2 Das DGR aus *Treponema denticola* ist über ein Jahr nicht aktiv

Die Hypermutation der VR könnte im natürlichen Lebensraum des Wirts frequenter auftreten als in kultivierten Zellen (Kap. 4.2.1). Um dies zu untersuchen, wurde das DGR aus *Treponema denticola* zu verschiedenen Zeitpunkten in Patientenproben analysiert (Abbildung 4, Kap. 3.2.3). Bei den Patientenproben handelt es sich um subgingivalen Plaque, einen Biofilm aus über hundert Bakterienarten. In dem Biofilm ist *Treponema denticola* ein später Ansiedler, der sich über Rezeptor-Ligand-Interaktionen mit anderen Bakterien verbindet. An den Interaktionen könnte auch das an der Zelloberfläche lokalisierte *Treponema variable* Protein A (TvpA) beteiligt sein. Da die Zusammensetzung der Bakterienarten im Biofilm variieren kann, ist ein Anpassungsdruck auf TvpA wahrscheinlich (zusammengefasst in Le Coq & Ghosh 2011). Aus diesem Grund wurde die VR von Zielgen TDE2269 untersucht, welches für TvpA kodiert. Zu dem DGR aus *Treponema denticola* gehören sieben potentiell variable Proteine, die eine Sequenzidentität von 25-67 % besitzen (Le Coq & Ghosh 2011). Bei der Analyse der sequenzierten VRs wurde daher überprüft, ob diese zu TDE2269 gehören. Dies war immer der Fall, was die Spezifität der Primer demonstriert. Die analysierten VR-Sequenzen wurden nicht über den REBVA angereichert, da eine wiederholte Amplifizierung der VR schwierig und die Probenmenge gering war. Der REBVA sollte für diese Analyse aber nicht nötig sein, da die natürliche Selektion das am besten angepasste variable Protein hervorbringt.

Die VR-Sequenzen von *Treponema denticola* konnten aus Proben von fünf Patienten zu unterschiedlichen Zeitpunkten sequenziert werden (Tabelle 10). Der Abstand zwischen der Probenentnahme beträgt dabei maximal ein Jahr. Zwischen keiner der Proben konnte DGR-Aktivität festgestellt werden. Die einzige VR-Sequenz die sich von der 2004 publizierte Sequenz an variablen Positionen unterscheidet existiert seit Beginn der Studie und zeigt auch ein halbes Jahr nach Beginn der Studie keine Veränderung (Patient 5 in Abbildung 20, Seshadri et al. 2004). Die Sequenz weist Variabilität über den kompletten Bereich der VR auf. Das spricht für eine komplexe Selektion auf die mutierte Sequenz durch den Anpassungsdruck der Proteinevolution (Doulatov et al. 2004). In der Sequenz sind zudem mehrere unspezifische Mutationen zu beobachten. Diese könnten durch eine generell erhöhte Fehleranfälligkeit der DGR-kodierten RT entstanden sein, was eine Identifizierung von DGRs durch die Parameter der bioinformatischen Analysen erschweren würde (vgl. Kap. 4.1.1).

Die Aktivität des DGRs aus *Treponema denticola* scheint auch im natürlichen Lebensraum des Wirts stark reguliert zu sein. Bei der diversifizierten Sequenz könnte das DGR durch einen neuen Interaktionspartner bei der Infektion des Patienten „aktiviert“ worden sein. Wenn sich die Zusammensetzung der Bakterien im Biofilm nicht weiter verändert hat, wäre eine erneute Anpassung nicht notwendig und würde das Ausbleiben weiterer Aktivitätsereignisse erklären.

4.2.3 Heterolog exprimierte DGRs zeigen keine Aktivität

Die Beobachtungen betreffend der Aktivität von DGRs wurden bislang nur im ursprünglichen Wirtssystem gemacht (u.a. Liu et al. 2002; Minot et al. 2013; Arambula et al. 2013). Eine heterologe Expression ist für DGRs nicht beschrieben. Da DGRs vermutlich den horizontalen Gentransfer zur Verbreitung nutzen (Schillinger et al. 2012), sollten benötigte Wirtsfaktoren artunspezifisch sein. Das macht eine DGR-Aktivität im heterologen Organismus wahrscheinlich.

Um DGR-Aktivität im heterologen Organismus zu untersuchen, wurden die DGRs aus *Anabaena flos-aquae*, *Nostoc sp. PCC7120* und *Treponema denticola* in Standardvektoren kloniert und in *E. coli* transformiert (Kap. 3.2.4). Die transformierten Zellen wurden bis zu einer Woche in Kultur gehalten und die Expression des DGRs induziert. Diese Zeitspanne sollte ausreichen, damit die benötigten Prozesse (Transkription, Translation und mutagenes Homing) ablaufen können. Während die Plasmide mit den DGRs aus *Nostoc sp. PCC7120* und *Treponema denticola* über den Zeitraum stabil vorlagen, wurde für das Plasmid mit dem DGR aus *Anabaena flos-aquae* eine Fragmentierung beobachtet, die auch uninduziert auftrat (Abbildung 21). Dies könnte ein Hinweis darauf sein, dass die Polymerase aus *E. coli* den DGR-Promotor erkennen kann, die Genprodukte aber toxisch für die Zelle sind. Eine fehlende Spezifität der DGR-kodierten RT oder des mutagenen Homings könnte zur Hypermutation zufälliger Genbereiche führen und damit dem heterologen Wirt schaden.

Um zu untersuchen, ob die Fragmentierung mit einer DGR-Aktivität in Verbindung steht, wurden neben intakten auch fragmentierte Plasmide mit dem REBVA analysiert. Dabei konnte für kein Element Aktivität festgestellt werden. Da aber auch in den ursprünglichen Wirtsorganismen keine bzw. kaum DGR-Aktivität festgestellt wurde (Kap. 4.2.1-4.2.2), lässt sich nicht mit Sicherheit sagen, dass DGRs in heterologen Organismen inaktiv sind. Zu der geringen Frequenz des mutagenen Homings können im heterologen System noch ein anderer Kodongebrauch und ungeeignete Chaperone hinzukommen, welche für die native Faltung von Proteinen und RNA benötigt werden.

Da die Frequenz des mutagenen Homings für das DGR des *Bordetella*-Phagen abschätzbar ist (~ 1 % der neu gebildeten Phagen), wäre es für einen Transfer in *E. coli* gut geeignet und könnte in einem zukünftigen Versuch getestet werden (Medhekar & Miller 2007). Die Kultivierung und genetische Manipulation von *E. coli* ist im Vergleich zum *Bordetella*-Phagen einfach und kostengünstig, was eine Diversifizierung heterologer Sequenzen, die im *Bordetella*-Phagen bereits erfolgreich durchgeführt wurde, einfacher gestalten sollte (Guo et al. 2008; Overstreet et al. 2012). DGRs könnten dann im Protein-Engineering eingesetzt werden, bei dem Proteine mittels zufallsbasierter Mutagenese verändert und optimiert werden.

Ein Beispiel stellt die Tre-Rekombinase dar, bei der die Erkennungssequenz der Cre-Rekombinase verändert wurde, um spezifisch die provirale DNA des HI-Virus aus der DNA menschlicher Zellen herauszuschneiden (Sarkar et al. 2007). Der Tre-Rekombinase könnte daher in Zukunft eine starke medizinische Bedeutung zukommen.

4.2.4 Regulation der DGR-Aktivität durch den Wirt

In dieser Arbeit konnte nur in *Nostoc sp.* PCC7120 DGR-Aktivität festgestellt werden (Kap. 3.2.1). Das Cyanobakterium wurde für ~ 2 Jahre kultiviert, wobei nur ein Aktivitätsereignis beobachtet werden konnte (Abbildung 18). Der verwendete REBVA analysiert vier variable Positionen in der VR hingehend auf Mutationen. Da fast jedes DGR-Retrohoming (> 90 %) von Mutationen an den variablen Positionen begleitet wird, sollte ein Großteil der Aktivitätsereignisse durch den REBVA identifiziert werden (Guo et al. 2008).

Für das DGR aus *Treponema denticola* konnte im natürlichen Lebensraum des Wirts über ein Jahr keine Aktivität festgestellt werden (Kap. 3.2.3). Eine VR-Sequenz mit DGR-charakteristischen Unterschieden zur VR der Datenbank konnte aber beobachtet werden und lässt ein aktives Element vermuten (Abbildung 20).

Die Beobachtungen sprechen für eine geringe Frequenz des mutagenen Homings. Dies wurde bereits für DGRs aus Viren des menschlichen Darms und für ein DGR aus dem humanpathogenen Bakterium *Legionella pneumophila* (Legionärskrankheit) beobachtet (Minot et al. 2013; Arambula et al. 2013). Arambula und Kollegen konnten die Frequenz des mutagenen Homings in *Legionella pneumophila* durch eine exogene Expression der DGR-Komponenten Avd, TR und RT erhöhen. Das spricht für eine Regulation des DGRs über die Transkriptmenge der Komponenten. Eine Regulation des DGRs wäre für den Wirt vorteilhaft, da durch die Hypermutation vermutlich auch ein Großteil funktionsunfähiger Proteine entsteht. Eine Regulation könnte das DGR dann „aktivieren“, wenn ein entsprechender Anpassungsdruck vorhanden ist. Dadurch würde das Risiko funktionsunfähige Proteine zu erhalten auf Situationen begrenzt werden, in denen eine Proteinanpassung notwendig ist.

Die Ausbildung der CLec-Faltung innerhalb der C-terminalen Domäne von variablen Proteinen reduziert das Risiko funktionsunfähige Proteine zu erhalten, da die grundlegende Proteinstruktur von dem variablen Bereich nicht beeinflusst wird (McMahon et al. 2005; Medhekar & Miller 2007).

Der *Bordetella*-Phage kann zudem die Affinität des variablen Proteins (Mtd) zum Bakterienrezeptor durch Avidität verstärken, wodurch auch schlecht angepasste Mtd-Proteine zu einer Infektion führen können (McMahon et al. 2005; Miller et al. 2008; Dai et al. 2010).

4.3 Strukturaufklärung der Templat RNA aus *Nostoc sp.* PCC7120

Ein weiterer Aspekt dieser Arbeit beschäftigt sich mit der Templat RNA von DGRs. Dabei handelt es sich um das Transkript des TR, welches bereits als mobiles Element für den Sequenztransfer identifiziert wurde (Guo et al. 2008). Analog zu den verwandten Gruppe II Introns wäre die Ausbildung von Strukturelementen denkbar, welche eine hochaffine Bindestelle für die DGR-kodierte RT stellen und/oder der Templat RNA katalytische Aktivität verleihen könnten (Kap. 1.1.1; Doulatov et al. 2004). Durch die katalytische Aktivität könnte die Templat RNA eine zusätzliche Funktion als Endonuklease bekommen, welche für den vorgeschlagenen TPRT-Mechanismus des mutagenen Homings benötigt wird, aber noch nicht identifiziert werden konnte (Kap. 4.2.1; Guo et al. 2011). Welche Struktur die Templat RNA ausbildet, wurde bislang ebenfalls nicht identifiziert. Daher wurde in dieser Arbeit die erste Strukturaufklärung für Templat RNAs durchgeführt. Dafür wurde die Templat RNA aus *Nostoc sp.* PCC7120 ausgewählt und bioinformatisch und experimentell untersucht. Dieses DGR wurde ausgewählt, da es Aktivität gezeigt hat und die einzelnen Komponenten (z.B. Templat RNA) daher funktionsfähig sein sollten (Kap. 3.2.1).

4.3.1 Für den Templat Repeat existieren unterschiedlich lange Transkripte

Damit die Strukturaufklärung der Templat RNA aus *Nostoc sp.* PCC7120 durchgeführt werden konnte, mussten zunächst Transkriptionsstart (TSS) und Transkriptionsstopp (STOPP) identifiziert werden. Dafür wurden RACE-, Northern-Blot- und RT-PCR-Analysen als unabhängige Methoden verwendet (Kap. 3.3.1).

Die RACE-Experimente ergaben jeweils zwei TSS für die Templat RNA (Abbildung 23A). Während TSS1 bereits innerhalb eines genomweiten TSS-Mappings identifiziert wurde, konnte der dazu downstream gelegene TSS2 in dem Mapping nicht beobachtet werden (Mitschke et al. 2011). Sowohl in dem genomweiten TSS-Mapping, als auch in den RACE-Experimenten, wurde mit der „Terminator 5'-Phosphate-Dependent Exonuclease“ (TEX) gearbeitet. Dieses Enzym baut prozessierte RNA (5'-Monophosphat) ab, während bakterielle Primärtranskripte (5'-Triphosphat) kein Substrat darstellen. Dadurch sollten beide Analysen selektiv für Primärtranskripte sein. Da TSS2 von Mitschke und Kollegen nicht beobachtet wurde, könnte es sich alternativ um das Transkriptende einer prozessierten Templat RNA handeln, wenn die in dieser Arbeit verwendete TEX keine ausreichend enzymatische Aktivität besessen hat. Um die Aktivität der TEX zu überprüfen, wurde ein Northern-Blot-Experiment durchgeführt, in dem die hoch abundand vorliegende, prozessierte 16S rRNA erfolgreich abgebaut werden konnte (Abbildung 22). Das mit TSS2 beginnende Transkript wurde in einem

Northern-Blot-Experiment dagegen nicht von der TEX abgebaut (Abbildung 26). Das ist ein Hinweis darauf, dass TSS2 ein Transkriptionsstart ist und kein prozessiertes Transkriptende repräsentiert. TSS2 konnte auch durch weitere Northern-Blot- und RT-PCR-Experimente als Beginn eines TR-Haupttranskripts identifiziert werden (Abbildung 23B, Abbildung 25B1).

Der von Mitschke und Kollegen identifizierte TSS1 konnte dagegen nur schwer in einem Northern-Blot-Experiment identifiziert werden (Abbildung 26). Demnach könnte es zu einem Wechsel des TSS gekommen sein.

Die Verwendung verschiedener TSS wurde in Bakterien bereits beobachtet und mit einer Regulation in Verbindung gebracht, die es ermöglicht alternative Sekundärstrukturen auszubilden (Filiatrault et al. 2011). Da für die Struktur der Templat RNA Funktionen postuliert wurden (Kap. 4.3), die entscheidend für die Aktivität des Elements sind, könnte ein Wechsel des TSS eine Möglichkeit für den Wirt darstellen, das DGR zu regulieren (Kap. 4.2.4).

Der TSS-Wechsel könnte dabei durch die unterschiedlichen Kultivierungsbedingungen der Labore begünstigt worden sein. Ein Versuch die Kultivierungsbedingungen des Labors „Genetik & experimentelle Bioinformatik“ (Uni Freiburg), welches das genomweite TSS-Mapping durchgeführt hat, zu übernehmen, führte aber nicht zu einem Wechsel des TSS (Daten nicht gezeigt). Da die Ursache für einen möglichen TSS-Wechsel nicht bekannt ist, müssten die Zellen beider Labore in einem parallelen Versuch miteinander verglichen werden. Dafür würde sich ein Northern-Blot-Experiment mit TSS1-spezifischer Sonde anbieten (vgl. Abbildung 23B). Würde ein TSS-Wechsel beobachtet werden, könnten die Auswirkungen der veränderten Sequenz/Struktur auf die Aktivität des DGRs über Aktivitätsanalysen untersucht werden (vgl. Kap. 4.2). Da der Versuch jedoch nicht durchgeführt wurde, bleibt eine Regulation der DGR-Aktivität über den TSS in *Nostoc sp.* PCC7120 spekulativ.

Neben den TSS konnten durch die RACE-Experimente auch zwei STOPPs identifiziert werden (Abbildung 23A). Dabei wird STOPP2 für das Haupttranskript der Templat RNA verwendet. Dies konnte durch Northern-Blot- und RT-PCR-Experimente bestätigt werden (Abbildung 23B, Abbildung 25C). Es konnte aber auch ein schwaches Transkript beobachtet werden, bei dem es sich um eine Variante handeln könnte, die mit STOPP1 endet (Abbildung 24B). Demnach könnte eine mögliche Regulation des DGRs auch über das Transkriptende der Templat RNA stattfinden. Dabei könnte ein Wechsel im verwendeten Transkriptionsterminator über Antiterminationsfaktoren gesteuert werden, welche in Phagen und Bakterien vorkommen (zusammengefasst in Santangelo & Artsimovitch 2011).

Die verschiedenen Varianten der Templat RNA reichen von 252 nt (TSS2 bis STOPP1) bis zu 411 nt (TSS1 bis STOPP2). Da keine der beobachteten Varianten größer ist als 1000 nt, ist ein Kotranskript mit der RT unwahrscheinlich (offener Leserahmen der RT: 1060 nt). Dieser experimentelle Hinweis bestätigt eine bioinformatische Beobachtung von Schillinger und

Kollegen, die ein DGR identifizieren konnten, bei dem RT und TR auf unterschiedlichen Strängen lokalisiert sind (Schillinger et al. 2012). Ein gemeinsames Transkript für RT und TR wurde zunächst angenommen, da beide Komponenten oft in direkter Nähe zueinander vorgefunden wurden und auch die RT der verwandten Gruppe II Introns ihre eigene RNA als Templat verwendet (Kap. 1.1.1; Doulatov et al. 2004).

Die Analyse der Sekundärstruktur wurde mit der Variante der Templat RNA durchgeführt, die in dieser Arbeit am eindeutigsten nachgewiesen werden konnte. Dabei handelt es sich um das Transkript, welches bei TSS2 beginnt und bei STOPP2 endet (TR3). Dieses Transkript besitzt eine theoretische Größe von 335 nt, die auch mit der abgeschätzten Größe im Northern-Blot übereinstimmt (Abbildung 24).

4.3.2 Analyse der identifizierten Sekundärstruktur der Templat RNA

RNAs können über Wasserstoffbrücken zwischen komplementären Abschnitten Sekundärstrukturen ausbilden. Die dabei entstehenden Strukturelemente (z.B. Haarnadelschleifen) können wiederum untereinander interagieren und komplexe Tertiärstrukturen formen. Die Tertiärstruktur kann der RNA eine katalytische Aktivität verleihen bzw. eine spezifische Bindestelle für Proteine bereitstellen. Dies ist am Beispiel der Gruppe II Introns zu erkennen, welche sowohl spezifisch das Intron-kodierte Protein binden, als auch das Selbstspießen katalysieren können (Kap. 1.1.1). Für das RNA-Intermediat (Templat RNA) der DGRs, welche eng mit den Gruppe II Introns verwandt sind, wären ähnliche Funktionen denkbar (Doulatov et al. 2004; Guo et al. 2011). Daher sollte die Struktur des TR-Haupttranskripts, welches in Kap. 3.3.1 identifiziert wurde, aufgeklärt werden. Die Tertiärstruktur des Gruppe IIC Introns aus *Oceanobacillus iheyensis* wurde erfolgreich über Kristallstrukturanalysen aufgeklärt (Toor et al. 2008). Eine weitere Methode zur Analyse der RNA-Tertiärstruktur ist die Kernspinresonanzspektroskopie (zusammengefasst in Furtig et al. 2003). Diese Methoden benötigen allerdings entsprechende Gerätschaften und Fachkenntnisse, die für diese Arbeit nicht zur Verfügung standen.

Der Faltungsprozess der RNA verläuft hierarchisch (Tinoco & Bustamante 1999). Das bedeutet, dass die Sekundärstruktur ohne Kenntnis der Tertiärinteraktionen berechnet werden kann, da die lokalen Interaktionen vorher entstehen und energetisch stärker sind als Tertiärinteraktionen. Die Sekundärstruktur lässt sich bioinformatisch berechnen, wodurch sich eine Aufklärung einfacher gestaltet als bei der Tertiärstruktur. Zudem erlaubt eine Kenntnis der Sekundärstruktur teilweise Aussagen über die Funktion der RNA (z.B. über Struktur motive

oder Protein-zugängliche Regionen (Li 2010)). Daher wurde die Sekundärstruktur der Templat RNA bioinformatisch analysiert (Kap. 3.3.2).

Die akkuratete Methode der bioinformatischen Strukturberechnung ist die komparative Analyse verwandter Nukleotidsequenzen in einem Alignment (Pace, Thomas & Woese 1999). Der Methode liegt die Annahme zugrunde das Mutationen, welche eine Basenpaarung in einem funktionellen Stamm zerstören (z.B. G:C zu G:A), einen schädlichen Einfluss auf die RNA haben. Dieser schädliche Einfluss kann durch eine zweite, kompensierende Mutation auf der anderen Seite des Stamms beseitigt werden (z.B. G:A zu U:A). Wird ein solcher Polymorphismus zwischen den analysierten Sequenzen beobachtet, wird von Kovarianz gesprochen. Die Kovarianz unterstützt die Existenz der Basenpaarung in der Struktur, da eine kompensierende Mutation während der Evolution nicht auftreten würde, wenn die Basenpaarung für die Struktur oder Funktion der RNA nicht erforderlich wäre. Die Genauigkeit der Faltungsvorhersage wurde für die 30S und 50S ribosomalen Untereinheiten bestimmt, von denen Kristallstrukturen mit hoher Auflösung vorliegen (Ban et al. 2000; Wimberly et al. 2000). Dabei lagen über 97 % der Basenpaare in der Sekundärstruktur, welche durch die komparative Methode vorhergesagt wurde, auch in der Kristallstruktur vor (Gutell, Lee & Cannone 2002). Eine Voraussetzung für die komparative Methode ist jedoch, dass die analysierten RNA-Moleküle über konservierte, funktionelle Sekundärstrukturelemente verfügen. Da dies für die Templat RNA nicht bekannt ist, wurde zunächst auf die komparative Analyse verzichtet.

Die akkuratete Methode für die Faltungsvorhersage einzelner Sequenzen erfolgt über eine Minimierung der freien Energie in den berechneten Strukturen. Dabei wird den Basen, je nach Kontext, ein bestimmter Energiewert zugeordnet (komplementäre Basen bekommen z.B. relativ niedrige Energiewerte). Das populärste Programm, welches diese Methode über eine dynamische Programmierung mit nearest-neighbor Parametern verwirklicht, heisst Mfold (Zucker 1989; Zucker 2003). Die Genauigkeit der Faltungsvorhersage von Mfold wurde in einer Studie getestet, welche die berechneten Strukturen von tRNAs (569 Strukturen) und 16S rRNAs (496 Strukturen) mit den jeweiligen komparativen Strukturmodellen verglichen hat (Doshi et al. 2004; Cannone et al. 2002). Die durchschnittliche Genauigkeit liegt dabei für die 16S rRNA (~ 1500 nt) bei 41 % und für die kleinere tRNA (~ 75 nt) bei 69 %. Aufgrund der hohen Genauigkeit der Faltungsvorhersage kleinerer RNAs (Kontaktdistanz zwischen den Basenpaaren unter 100 Nukleotiden), wurde Mfold auch für die Sekundärstrukturberechnung der Templat RNA verwendet. Dabei wurden die vier möglichen RNA-Sequenzen, die sich aus den identifizierten Transkriptenden ergeben, gefaltet (Abbildung 27). Es ist zu erkennen, dass die Kernregion der Templat RNA (TSS2 bis STOPP1) aus fünf Haarnadelstrukturen besteht und sich hinzukommende Sequenzen (TSS1-TSS2 und STOPP1-STOPP2) unabhängig von der Kernregion falten. In der Berechnung sind jedoch keine Tertiärinteraktionen berücksichtigt.

Tertiärinteraktionen können z.B. durch Interaktionen von GNRA-Tetraloops mit helikalen Bereichen in der RNA entstehen (Kap. 3.3.2). Für die Sequenz STOPP1-STOPP2 wurden zwei GNRA-Tetraloops vorhergesagt. Die für die biologische Funktion entscheidende Tertiärstruktur könnte sich also abhängig von dem verwendeten STOPP verändern. Demnach könnte eine mögliche Regulation der DGR-Aktivität über die Länge der Templat RNA erfolgen (Kap. 4.2.4).

Eine Limitierung von Mfold stellt die Berechnung von Pseudoknoten dar. Pseudoknoten sind Strukturelemente, die minimal aus zwei Haarnadelstrukturen bestehen, von denen die Schleife eines Stamms gleichzeitig Teil des zweiten Stamms ist. Die Berechnung gestaltet sich als schwierig, da sich die Positionen in der Sequenz überlappen können und Mfold dann lediglich den stabileren der beiden Stämme im Pseudoknoten vorhersagt. Pseudoknoten können aber wichtig für biologische Prozesse sein. So besitzt die menschliche Telomerase einen Pseudoknoten der essentiell für die Aktivität des Enzyms ist und das Gruppe I Intron aus *Azoarcus* sp. BH72 einen Pseudoknoten der entscheidend für den Aufbau der nativen Struktur ist (Chen & Greiger 2005; Adams et al. 2004). Es gibt Programme für die Sekundärstrukturvorhersage, welche nicht nur eine hohe Genauigkeit aufweisen, sondern auch Pseudoknoten berechnen können. Zu den Programmen gehören das zu Mfold verwandte RNAstructure (Probknot) (Reuter & Mathews 2010; Bellaousov & Mathews 2010), ILM (Ruan, Stormo & Zhang 2004), HotKnots (Ren et al. 2005) und pknotsRG (Reeder & Giegerich 2004). Von diesen Programmen besitzt RNAstructure (Probknot) die genaueste Strukturvorhersage, weshalb es für die bioinformatische Überprüfung der Mfold-Struktur und für die Vorhersage von Pseudoknoten in der Templat RNA verwendet wurde (Bellaousov & Mathews 2010). Dabei zeigt die berechnete Sekundärstruktur des TR-Haupttranskripts kaum Unterschiede zu der Mfold-Struktur (Abbildung 28; Tabelle 11). In der Struktur konnten keine Pseudoknoten identifiziert werden. Da die Vorhersage von Pseudoknoten mit RNAstructure (Probknot) und den oben genannten Programmen aber unzuverlässig ist, d.h. nur ein geringer Anteil der Pseudoknoten in bekannten Strukturen korrekt identifiziert wird, könnten dennoch Pseudoknoten in der Templat RNA vorkommen (Bellaousov & Mathews 2010).

Bioinformatische Strukturvorhersagen sind meist nicht hundertprozentig korrekt. Das liegt zum einen an der hohen Anzahl möglicher Strukturmodelle ($\sim 1,8^n$ Strukturmodelle für eine Sequenz mit n Nukleotiden) (Zuker & Sankoff 1984). Die Strukturmodelle können ähnliche Energiewerte besitzen, sodass geringe Störungen der thermodynamischen Parameter zu einem Wechsel der Struktur mit der niedrigsten freien Energie führen können (Layton & Bundschuh 2005). Zum anderen können die Programme bei der Strukturvorhersage keine stabilisierenden Tertiärinteraktionen berechnen. Daher bietet sich eine Überprüfung bzw. Optimierung der bioinformatisch berechneten Struktur durch experimentelle Daten an.

Bioinformatische Strukturvorhersagen können durch SHAPE-Experimente verbessert werden (Deigan et al. 2009; Washietl et al. 2012; Sükosd et al. 2013; Merino et al. 2005). SHAPE basiert auf der Entdeckung, dass die Nukleophilie der 2'-Hydroxylgruppe der Ribose durch die angrenzende 3'-Phosphodiestergruppe beeinflusst wird (Chamberlin, Merino & Weeks 2002). Dabei können flexible (einzelsträngige) Nukleotide Konformationen eingehen, welche die 2'-Hydroxylgruppe nukleophiler macht, als bei eingeschränkten (z.B. basengepaarten) Nukleotiden. Ein Elektrophil wie NMIA kann dadurch mit flexiblen Nukleotiden ein stabiles 2'-O-Addukt bilden, welches in einer anschließenden Primer-Extension erfasst wird. Dadurch erlaubt SHAPE eine Analyse der Flexibilität sämtlicher Nukleotide der analysierten RNA und berücksichtigt Tertiärinteraktionen. Das ist ein Vorteil gegenüber Methoden mit basenmodifizierenden Chemikalien, welche nur spezifische Nukleotide analysieren können. Ein Beispiel ist Dimethylsulfat, welches Adenosin (N1) und Cytosin (N3) methyliert (zusammengefasst in Tijerina, Mohr & Russel 2007). Aus dem Grund wurde die *in vitro* transkribierte Templat RNA mittels SHAPE analysiert.

SHAPE-Analysen können bei einer unsachgemäßen Durchführung zu ungenauen Sekundärstrukturvorhersagen führen (Kladwang et al. 2011(a); Kladwang et al. 2011(b)). Daher wurden generelle Richtlinien für eine erfolgreiche Durchführung der SHAPE-Experimente aufgestellt, die sich u.a. mit dem richtigen Umgang der verwendeten Chemikalie (NMIA) beschäftigen (zusammengefasst in Leonard et al. 2013). Die in dieser Arbeit beschriebenen SHAPE-Experimente wurden unter Beachtung dieser Richtlinien durchgeführt, wobei eine Herausforderung in der Bestimmung der korrekten Faltungsbedingungen lag. Die Faltungsbedingungen können entscheidend für die Aktivität der RNA sein. So führt eine Faltungstemperatur von 25 °C zu einem inaktiv gefalteten Gruppe I Intron (*Tetrahymena thermophila*), während bei 50 °C die aktive Form begünstigt wird (Tijerina, Mohr & Russel 2007). Um die Faltungsbedingungen der Templat RNA zu prüfen, bietet sich ein *in vitro* Funktionalitätsassay an, der Diversifizierungen in der VR detektiert. Der Assay ließ sich aber nicht verwirklichen, da neben den aufgereinigten DGR-Komponenten (RT, Avd, Templat RNA und Plasmid mit Zielgen) weitere Faktoren (z.B. DNA-Reparaturmechanismen) berücksichtigt werden müssen. Daher wurden stattdessen die Faltungsbedingungen identifiziert, welche die stabilste und damit wahrscheinlich native Konformation begünstigen (Kap. 3.3.3).

Bei der Auswertung der SHAPE-Experimente wurde zunächst die radioaktiv markierte cDNA, die bei der Primer-Extension entstanden ist, in einem Phosphorimager detektiert und die Bandenintensität über eine Software quantifiziert. Die Größe einer Phosphorimagerplatte reichte dabei aus, um etwa 30-40 Nukleotide abzudecken. Ein komplettes Polyacrylamidgel, welches in Abbildung A4 dargestellt ist, konnte nicht abgedeckt werden. Daher wurden für einen Versuch mehrere Phosphorimagerplatten eingesetzt, bei denen die Expositionsdauer

variierte. Für die überlappenden Bereiche zwischen den Phosphorimagerplatten wurden die Bandenintensitäten miteinander verglichen. Dabei konnte ein hochsignifikanter linearer Zusammenhang festgestellt werden (Kap. 2.2.4.2). Demnach ist die automatische Erfassung und Quantifizierung der Banden durch die Software verlässlich und die Verwendung mehrerer Phosphorimagerplatten akzeptabel. Aus den Bandenintensitäten wurden die SHAPE-Reaktivitäten berechnet (Kap. 2.2.4.2). Die SHAPE-Analyse wurde im Triplikate durchgeführt und die SHAPE-Reaktivitäten der drei Experimente miteinander verglichen (Tabelle A3). Dabei ist zu sehen, dass mit drei Ausnahmen (Erklärung in Kap. 3.3.3) ähnliche SHAPE-Reaktivitäten für die einzelnen Nukleotide berechnet wurden. Demnach lag die RNA in derselben Konformation vor und die Nukleotide wurden zuverlässig modifiziert und analysiert. Eine einfachere Handhabung der SHAPE-Experimente kann durch fluoreszenzmarkierte Primer (geringeres Risikopotential) und die Verwendung einer Kapillarelektrophorese (Auslesen bis 400 nt) erreicht werden (Wilkinson et al. 2008). Für dieses „high-throughput“ SHAPE wurden datenverarbeitende Programme wie SHAPEFinder und FAST entwickelt, die eine zeitsparende Auswertung erlauben (Vasa et al. 2008; Pang et al. 2011).

Die SHAPE-Reaktivitäten wurden als „soft constraints“ in RNAstructure integriert und die Struktur des TR-Haupttranskripts berechnet. Zwischen den bioinformatisch berechneten Sekundärstrukturen und der SHAPE-optimierten Struktur konnten kaum Unterschiede beobachtet werden (97-98 % Übereinstimmung, s. Tabelle 11). Das ist ein Hinweis auf die Identifizierung der korrekten Konformation und zeigt, dass Mfold und RNAstructure eine genaue Strukturvorhersage für die Templat RNA machen konnten. Das könnte bedeuten, dass die Struktur der Templat RNA sehr stabil ist. Die Struktur besteht aus sieben Haarnadelstrukturen (I-VII) (Abbildung 30). Die Haarnadelstrukturen V-VII weisen jeweils einen GNRA-Tetraloop auf. Diese Tetraloops können nicht nur für den Aufbau der Tertiärstruktur entscheidend sein, sondern auch in Interaktion mit Proteinen stehen (Kap. 3.3.2). Mögliche Kandidaten für eine spezifische Bindung an die Templat RNA wären die DGR-kodierte RT (analog zu den verwandten Gruppe II Introns, Kap. 1.1.1) und die Avd. Für die Avd wurde eine nukleinsäurebindende Funktion postuliert (Alayyoubi et al. 2013). Alayyoubi und Kollegen konnten eine spezifische Bindung zu der Templat RNA aus dem *Bordetella*-Phagen aber nicht überprüfen, da die Transkriptenden nicht bekannt waren. Da sowohl die Avd, als auch die DGR-kodierte RT aus *Nostoc sp.* PCC7120 in der Molekularen Genetik (TU Kaiserslautern) aufgereinigt werden konnten, bieten sich nun Footprinting-Analysen an (z.B. SHAPE), welche eine spezifische Bindung zu der Templat RNA testen können.

4.3.3 Die Templat RNA könnte konservierte Strukturelemente besitzen

Die identifizierten Strukturen der Templat RNA aus *Nostoc sp.* PCC7120 stimmen im bioinformatischen (Mfold, RNAstructure) und experimentell erweiterten (SHAPE) Ansatz zu 97-98 % überein (Tabelle 11). Demnach lässt sich die Templat RNA bioinformatisch gut berechnen, was eine Berechnung weiterer Templat RNA-Strukturen nahelegt (Kap. 3.3.4). Dafür wurden neben der Sequenz aus *Nostoc sp.* PCC7120 noch 12 weitere Sequenzen untersucht (Tabelle 12). Hierbei stellte sich das Problem, dass die Transkriptenden nur für die Templat RNA aus *Nostoc sp.* PCC7120 bekannt sind (Kap. 3.3.1). Unter der Annahme, dass die verschiedenen Templat RNAs ähnlich lang sind und die Transkriptenden einen ähnlichen Abstand zum TR besitzen, wurden die 13 TRs mit umgebender Sequenz (200 bp up- und downstream) für ein Alignment verwendet. Anschließend wurde der Sequenzbereich ausgewählt, der mit der Sequenz aus *Nostoc sp.* PCC7120 aligned, die der Templat RNA entspricht. Dabei wurde die größte Variante der Templat RNA (411 nt) ausgewählt, da der verwendete TSS bzw. STOPP keine Auswirkung auf den Kernbereich der bioinformatisch berechneten Templat RNA besitzt (Abbildung 27). Diese Methode zum Erlangen der Templat RNAs beruht auf oben genannter Annahme und ersetzt keine exakte Identifizierung der Transkriptenden, weshalb die Ergebnisse dieses Versuchsteils kritisch zu betrachten sind. Die 13 Templat RNAs wurden für Strukturberechnungen herangezogen. Dabei wurden Hinweise auf konservierte Strukturelemente gefunden (Daten nicht gezeigt). Um diese besser herausstellen zu können, wurde die Struktur mit der besten Übereinstimmung für die 13 Sequenzen berechnet (Konsensus-Struktur). Für deren Berechnung gibt es verschiedene Programme (zusammengefasst in Will et al. 2012). Die Programme EvoFold (Pedersen et al. 2006) und RNAz (Washietl, Hofacker & Stadler 2005; Gruber et al. 2010) verwenden sequenzbasierte Alignments. Dadurch können strukturelle RNAs mit Sequenzidentitäten unter 60 % nur schlecht aligned werden (Gardner, Wilm & Washietl 2005). Diese Programme eignen sich daher nicht für eine Analyse der Templat RNA, welche zwischen nicht verwandten Arten eine kaum konservierte Primärstruktur aufweist (Sequenzidentitäten um 20 %) (Tabelle 14).

Tabelle 14: Sequenzidentitäten der potentiellen Templat RNAs

Organismen	Sequenzidentität
Cyanobakterien	52-92 %
Grüne Schwefelbakterien	68 %
Purpurbakterien	59 %
<i>Bordetella</i> -Phage vs. <i>Treponema denticola</i>	29 %
Cyanobakterien vs. grüne Schwefelbakterien (Purpurbakterien)	26-40 % (32-43 %)
Cyanobakterien vs. <i>Bordetella</i> -Phage (<i>Treponema denticola</i>)	22-26 % (29-39 %)

Die Sekundärstruktur der Templat RNA könnte dagegen evolutionär konserviert sein, da potentielle Funktionen über die Struktur vermittelt werden. Daher wurde ein Programm verwendet, das auch strukturelle Ähnlichkeiten berücksichtigt (Sankoff-Algorithmus; Sankoff 1985). Zu diesen Programmen gehören FoldAlign (Havgaard et al. 2005) und LocARNA (Will et al. 2007; Smith et al. 2010). FoldAlign berechnet ein loop-basiertes Energiemodell für die RNA, welches durch sequenzbasierte Heuristik kompensiert werden muss. Dagegen verwendet LocARNA ein Energiemodell, welches die Wahrscheinlichkeit von Basenpaarungen in den einzelnen Sequenzen einbezieht (McCaskill 1990). Dadurch wird der Rechenaufwand reduziert und unwahrscheinliche Basenpaarungen können ignoriert werden, ohne die Genauigkeit des Alignments zu kompromittieren. LocARNA wurde bereits verwendet um die Rfam-Datenbank (sequenzbasierte Klassifizierung nicht kodierender RNAs) und ein Set vorhergesagter struktureller RNAs aus *Ciona intestinalis* (Schlauchseescheide) strukturbasiert zu überprüfen (Will et al. 2007; Griffiths-Jones et al. 2005; Missal, Rose & Stadler 2005). Dabei wurden bestehende Klassen (z.B. t-RNAs) erfolgreich wiedererlangt und neue Klassen identifiziert, von denen bislang kein Vertreter experimentell charakterisiert wurde.

Die Berechnung der Konsensus-Struktur für die Templat RNA wurde mit LocARNA durchgeführt. Die Konsensus-Struktur zeigt dabei eine gewisse Ähnlichkeit zu der SHAPE-optimierten Struktur aus *Nostoc sp.* PCC7120 (Hinweis auf die Korrektheit der Konsensus-Struktur) (Abbildung 30, Abbildung 32). Die Ähnlichkeit bezieht sich dabei nicht auf die exakte Stamm- oder Schleifenlänge, sondern auf den Aufbau, der jeweils aus sieben Haarnadelstrukturen (I-VII) besteht. Die Haarnadelstrukturen I, II, IV und V sind in allen analysierten Templat RNAs recht konserviert und präsentieren an den apikalen Schleifen einen Uracil-Stretch (I), veränderliche Adenine (entsprechen variablen Positionen in der VR) (II), snRNP (small nuclear ribonucleic particles)-Bindemotive (IV) und einen GNRA-Tetraloop mit der Sequenz 5´GGGA´3 (nur Cyanobakterien) (V). Die Haarnadelstrukturen VI und VII sind dagegen nur in den cyanobakteriellen Templat RNAs konserviert. Demnach könnten diese Strukturelemente für die grundlegende Funktion der Templat RNA unerheblich sein bzw. den cyanobakteriellen Templat RNAs eine zusätzliche Funktion verleihen.

Für den Bereich TSS1-TSS2 lässt sich dagegen kein gemeinsames Strukturelement berechnen. Das ist ein Hinweis darauf, dass TSS1 nicht der Transkriptionsstart der Templat RNA ist. Dies ist zudem konsistent mit der Beobachtung, dass das TR-Haupttranskript aus *Nostoc sp.* PCC7120 mit TSS2 beginnt (Kap. 3.3.1).

Um zu prüfen, ob die Strukturelemente der Konsensus-Struktur zufällig bei der Berechnung entstanden sind, wurde eine Kontrollberechnung mit der Templat RNA aus *Nostoc sp.* PCC7120 und 12 zufällig generierten Sequenzen durchgeführt (Abbildung A5). Dabei konnten keine zufällig ausgebildeten Strukturelemente beobachtet werden.

Eine weitere Fehlerquelle wäre, wenn ein konservierter Genabschnitt (z.B. der DGR-kodierten RT) Teil sämtlicher Templat RNAs ist. Dann würde die Struktur dieses konservierten Bereichs berechnet und fälschlicherweise der Templat RNA zugeordnet werden. Daher wurde die Position des TR zu den DGR-Komponenten (RT, Avd und Zielgen) geprüft (Tabelle 15).

Tabelle 15: Position des TR zu den DGR-Komponenten in den analysierten Organismen

Organismus	Position des TR up- (↑) bzw. downstream (↓) zur:		
	RT	Avd	Zielgen
<i>Nostoc sp. PCC7120</i>	↓ (20 nt)	↓ (1096 nt)	↓ (5569 nt)
<i>Nostoc punctiforme</i>	↓ (20 nt)	↓ (1099 nt)	↓ (5526 nt)
<i>Anabaena flos-aquae</i>	↓ (30 nt)	↓ (1078 nt)	↓ (2009 nt)
<i>Nodularia spumigena</i> (A)	↓ (393 nt)	↓ (1473 nt)	↓ (1928 nt)
<i>Nodularia spumigena</i> (B)	↓ (18 nt)	↓ (1073 nt)	↓ (1588 nt)
<i>Cyanothece sp. CCY0110</i> (A)	↓ (69 nt)	↓ (1146 nt)	↓ (1721 nt)
<i>Cyanothece sp. CCY0110</i> (B)	↓ (22 nt)	↓ (1123 nt)	↓ (1626 nt)
<i>Chlorobium phaeobacteroides</i>	↑ (197 nt)	↓ (34 nt)	↓ (412 nt)
<i>Pelodictyon phaeoclathratiforme</i>	↑ (205 nt)	↓ (62 nt)	↓ (443 nt)
<i>Thiorhodovibrio sp. 970</i>	Teil der RT	↓ (1064 nt)	↓ (1508 nt)
<i>Rhodomicrobium vannielii</i>	↓ (1673 nt)	↓ (3492 nt)	↓ (180 nt)
<i>Treponema denticola</i>	↑ (571 nt)	---	↓ (194 nt)
<i>Bordetella</i> -Phage	↑ (144 nt)	↓ (1 nt)	↓ (417 nt)

Anhand der großen Entfernung (> 400 nt) des TR zur Avd und dem Zielgen, können diese Komponenten als „Kontamination“ der Berechnung ausgeschlossen werden. Der TR liegt dagegen bei sechs cyanobakteriellen Sequenzen (Sequenzidentität: 55-92 %) relativ nahe (18-69 nt) downstream zur DGR-kodierten RT vor. Demnach bestehen diese Sequenzen zum Teil aus kodierenden und untranslatierten Bereichen der DGR-kodierten RT. Es ist denkbar, dass sich diese Bereiche zu Strukturelementen falten, die für die Funktion der DGR-kodierten RT entscheidend sind und daher konserviert vorliegen. Bei der Darstellung der Konsensus-Struktur könnte eine Vorhersage dieser Strukturelemente dann auch den restlichen sieben Sequenzen „aufgezwungen“ werden. Etwas Derartiges trifft für die Haarnadelstrukturen VI und VII zu. Im Alignment (Abbildung 31) ist zu erkennen, dass diese Strukturen nur in den cyanobakteriellen Sequenzen konserviert sind, was bei der Konsensus-Struktur (Abbildung 32) aber zu einer einheitlichen Darstellung führt. Für die Haarnadelstrukturen I, II, IV und V trifft dies aber nicht zu, da eine Vorhersage dieser Strukturen von allen Sequenzen im Alignment relativ gut unterstützt wird.

Die Konservierung in den Haarnadelstrukturen I, II, IV und V ist beachtlich, da die Templat RNAs teilweise nur Sequenzidentitäten um 20 % aufweisen (Tabelle 14). Das lässt auf einen evolutionären Druck zur Ausbildung dieser Strukturelemente schließen, die daher entscheidend für die Funktion der Templat RNA sein könnten. Dabei könnte es sich um eine hochaffine Bindestelle zur DGR-kodierten RT (analog zu den verwandten Gruppe II Introns, Kap. 1.1.1) bzw. eine katalytische Funktion als Endonuklease handeln (Guo et al. 2011).

Um die Strukturelemente näher zu untersuchen, würden sich Mutationsexperimente anbieten, welche die Ausbildung eines Stamms unterbinden bzw. die Sequenz einer Schleife (z.B. Uracil-Stretch) verändern. Auf diese Weise könnten auch die GNRA-Tetraloops der cyanobakteriellen Templat RNAs (Haarnadelstrukturen V und VI) auf ihre Bedeutung analysiert werden. Voraussetzung für dieses Experiment ist ein Test (z.B. PCR-basierter Homing-Assay), der die Auswirkungen der Mutationen auf die Funktionalität des DGRs analysieren kann. Für den Homing-Assay wird ein Plasmid (kodiert für Avd, modifizierten TR und RT) in einen Wirt eingebracht, bei dem diese Region deletiert wurde. Homing-Ereignisse werden analysiert, indem die VR mit VR-spezifischem Primer und TR-spezifischem Primer (Sequenz-Tag in modifiziertem TR) amplifiziert wird (Guo et al. 2008; Arambula et al. 2013). Eine Kenntnis der funktionalen Strukturelemente in der Templat RNA könnte ein vielversprechendes Anwendungsgebiet von DGRs verbessern. Dabei geht es um die Diversifizierung heterologer Proteine, welche durch die Toleranz des TR für heterologe Sequenzen (~ 200 bp) limitiert wird (Overstreet et al. 2012; Guo et al. 2008). Diese Toleranz könnte erhöht werden, wenn die korrekte Faltung der Templat RNA berücksichtigt wird.

Die Templat RNAs wurden mit den Programmen RegRNA 2.0 (Chang et al. 2013) und RNA Analyzer (Bengert & Dandekar 2003) auf Homologien zu funktionellen RNA-Motiven getestet (Daten nicht gezeigt). Dabei konnte kein Motiv identifiziert werden, welches in den Kontext der Templat RNA passen würde. So wurden in acht von 13 Templat RNAs snRNP-Bindemotive in der Schleife von Haarnadelstruktur IV identifiziert (Abbildung 31). SnRNP-Bindemotive sind konservierte Sequenzen in snRNA-Molekülen, die an der Bindung von Sm-Proteinen beteiligt sind und zur Bildung der snRNPs beitragen. SnRNPs kommen in Eukaryoten vor und sind als essentielle Komponenten des Spleißosoms an der Prozessierung von RNA beteiligt (zusammengefasst in Urlaub et al. 2001). In Bakterien kommen Sm-like Proteine vor, welche ebenfalls SnRNP-Bindemotive erkennen und eine Ringstruktur (Hexamer) um die RNA bilden können. Dabei wird postuliert, dass diese Proteine an intermolekularen Nukleinsäureinteraktionen beteiligt sind (zusammengefasst in Møller et al. 2002).

Sm/Sm-like Proteine erinnern an die Avd, für die eine nukleinsäurebindende Funktion postuliert und die Ausbildung einer Faß-Struktur (Pentamer) mit zentraler Pore nachgewiesen wurde (Alayyoubi et al. 2013). Sm/Sm-like Proteine verfügen über ein gemeinsames Sequenz-

motiv (Sm-Motiv), das zuständig für die Bindung der RNA ist (Branlant et al. 1982; Urlaub et al. 2001) und von Alayyoubi und Kollegen nicht für die Avd beschrieben wurde. Daher ist eine Bindung der Avd an die snRNP-Bindemotive der Templat RNA unwahrscheinlich, was aber über Footprinting-Analysen überprüft werden könnte.

Zusammenfassend kann gesagt werden, dass es möglich war eine Konsensus-Struktur für 13 Sequenzen zu berechnen, die den Templat RNAs von DGRs aus verschiedenen Organismen entsprechen sollten. Die vorhergesagte Konsensus-Struktur weist dabei vier relativ konservierte Haarnadelstrukturen auf. Die Ausbildung dieser Haarnadelstrukturen wird nicht nur von den cyanobakteriellen Sequenzen, sondern auch von Sequenzen aus grünen Schwefelbakterien, Purpurbakterien, *Treponema denticola* und dem *Bordetella*-Phagen unterstützt und könnte mit einer konservierten Funktion der Templat RNA zusammenhängen. Hierbei ist allerdings zu beachten, dass lediglich eine Methode (Algorithmus) verwendet wurde, um die Sequenzen zu analysieren. Zudem reicht die Analyse von 13 Sequenzen nicht aus, um eine valide Aussage über die generelle Struktur der Templat RNA treffen zu können. In der durchgeführten Analyse wurden die Sequenzen per Hand extrahiert. Eine bioinformatische Analyse mit automatisierter Extraktion der Sequenzen würde es erlauben, eine Berechnung der Konsensus-Struktur aller bekannten Templat RNAs (momentan 155) durchzuführen. Zudem könnten experimentelle Strukturaufklärungen anderer Templat RNAs dabei helfen, die hier getroffenen Vermutungen zu prüfen. Sollte sich der in dieser Arbeit gefundene Hinweis auf konservierte Haarnadelstrukturen verstärken, bieten sich Mutationsexperimente an, welche die Bedeutung der Strukturelemente für die Funktion der Templat RNA weiter analysieren können. Die Strukturelemente könnten z.B. hochaffine Bindestellen für die DGR-kodierte RT oder die Avd bereitstellen, was durch Footprinting-Analysen überprüft werden kann.

4.4 Ausblick

In dieser Arbeit wurde das Vorkommen von DGRs in einer cyanobakteriellen Stammsammlung untersucht. Dabei wurden ~ 30 % (34) der Stämme mit DGRs assoziiert vorgefunden. Diese Stämme wurden auf unterschiedlichen Kontinenten isoliert, was auf eine weltweite Verbreitung der DGRs hindeutet. Weiterführende Experimente könnten das Vorkommen von DGRs in unkultivierten Cyanobakterien analysieren, deren Probenentnahme weitere Regionen abdeckt. Ein solches Experiment ist mit dem in dieser Arbeit etablierten Protokoll möglich, da nur eine geringe Probenmenge benötigt wird. Das Experiment würde keine Kultivierung der Cyanobakterien beinhalten, welche zu einem DGR-Verlust führen könnte.

Ein weiterer Aspekt dieser Arbeit beschäftigt sich mit der Aktivität bakterieller DGRs. Dabei wurden drei theoretisch funktionsfähige DGRs auf Hypermutationen in der VR untersucht. Es konnte aber nur ein Aktivitätsereignis beobachtet werden, das in dem Cyanobakterium *Nostoc sp.* PCC7120 stattgefunden hat. Wie bereits in vorhergehenden Untersuchungen vermutet wurde, könnte die beobachtete geringe Frequenz der Diversifizierung mit einer Regulation des DGRs durch den Wirt zusammenhängen. Daher wäre ein heterologes System sinnvoll, bei dem die Diversifizierung induziert werden kann und eine einfache und kostengünstige Kultivierung bzw. genetische Modifizierung möglich ist. Eine Bemühung in dieser Arbeit war es, die drei analysierten DGRs heterolog in *E. coli* zu untersuchen. Dabei konnte aber keine Diversifizierung beobachtet werden. Ein zukünftiges Experiment könnte das Modell-DGR aus dem *Bordetella*-Phagen verwenden, das eine häufigere Diversifizierung aufzuweisen scheint. Ein heterologes System könnte im Bereich des Protein-Engineerings Verwendung finden und u.a. dafür eingesetzt werden Enzyme zu modifizieren. Dabei kann z.B. die Erkennungssequenz verändert oder die katalytische Aktivität bzw. Resistenz gegenüber pH-Wertveränderungen optimiert werden.

In dieser Arbeit wurde zudem die Struktur der Templat RNA aus *Nostoc sp.* PCC7120 untersucht, da diese entscheidende Funktionen für den DGR-Mechanismus besitzen könnte. Zu den postulierten Funktionen gehört neben einer katalytischen Funktion als Endonuklease auch die Bereitstellung einer Bindestelle für die DGR-kodierte RT. Die Sekundärstruktur konnte bioinformatisch und experimentell aufgeklärt werden, wobei Footprinting-Experimente noch ausstehen, welche eine spezifische Bindung der DGR-kodierten RT (und Avd), sowie deren Lokalisierung auf der Templat RNA untersuchen können. Die Strukturinformation könnte auch im oben genannten Protein-Engineering Verwendung finden, da heterologe Sequenzen bislang nur bedingt vom TR toleriert werden. Wenn die Ausbildung der korrekten Struktur bei der Insertion der heterologen Sequenz in den TR berücksichtigt wird, sollte sich die Toleranz gegenüber der Sequenz erhöhen.

5 Literaturverzeichnis

Adams, DG 2000, 'Symbiotic interactions', in BA Whitton, M Potts (eds.), *The Ecology of Cyanobacteria*, Kluwer Academic Publishers, Dordrecht.

Adams, PL, Stahley, R, Gill, ML, Kosek, AB, Wang, J & Strobel, SA 2004, 'Crystal structure of a group I intron splicing intermediate.', *RNA*, vol 10, pp. 1867-1887.

Ahmed, S, Shafiuddin, MD, Azam, MS, Islam, MS, Ghosh, A & Khan, H 2011, 'Identification and characterization of jute LTR retrotransposons: Their abundance, heterogeneity and transcriptional activity.', *Mobile Genetic Elements*, vol 1, no. 1, pp. 18-28.

Alayyoubi, M, Guo, H, Dey, S, Golnazarian, T, Brooks, AG, Rong, A, Miller, JF & Ghosh, P 2013, 'Structure of the Essential Diversity-Generating Retroelement Protein bAvd and Its Functionally Important Interaction with Reverse Transcriptase.', *Structure*, vol 21, pp. 1-11.

Altschul, SF, Gish, W, Miller, W, Myers, EW & Lipman, DJ 1990, 'Basic local alignment search tool.', *J Mol. Biol.*, vol 215, no. 3, pp. 403-410, <<http://blast.ncbi.nlm.nih.gov/Blast.cgi>>.

Arambula, D, Wong, W, Medhekar, BA, Guo, H, Gingery, M, Czornyj, E, Liu, M, Dey, S, Ghosh, P & Miller, JF 2013, 'Surface display of a massively variable lipoprotein by a Legionella diversity-generating retroelement.', *PNAS Early Edition*, pp. 1-6.

Ashimoto, A, Chen, C, Bakker, I & Slots, J 1996, 'Polymerase chain reaction detection of 8 putative periodontal pathogens in subgingival plaque of gingivitis and advanced periodontitis lesions.', *Oral Microbiol Immunol*, vol 11, pp. 266-273.

Baba, T, Ara, T, Hasegawa, M, Takai, YOY, Baba, M, Datsenko, KA, Tomita, M, Wanner, BL & Mori, H 2006, 'Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.', *Mol Syst Biol.*, vol 2, pp. 1-11.

Ban, N, Nissen, P, Hansen, J, Moore, PB & Steitz, TA 2000, 'The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution.', *Science*, vol 289, pp. 905-920.

Bartrons, M, Catalan, J & Casamayor, EO 2012, 'High Bacterial Diversity in Epilithic Biofilms of Oligotrophic Mountain Lakes.', *Microb. Ecol.*, vol 64, pp. 860-869.

Belhocine, K, Plante, I & Cousineau, B 2004, 'Conjugation mediates transfer of the LI.LtrB group II intron between different bacterial species.', *Mol. Microbiol.*, vol 51, pp. 1459-1469.

Bellaousov, S & Mathews, DH 2010, 'ProbKnot: Fast prediction of RNA secondary structure including pseudoknots.', *RNA*, vol 16, pp. 1870-1880.

Bengert, P & Dandekar, T 2003, 'A software tool-box for analysis of regulatory RNA elements.', *Nucl. Acids. Res.*, vol 31, pp. 3441-3445.

Benson, DA, Karsch-Mizrachi, I, Lipman, DJ, Ostell, J & Sayers, EW 2009, 'GenBank.', *Nucleic Acids Res.*, vol 37, pp. D26-31.

Bertani, G 1951, 'Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*.' *J. Bacteriol.*, vol 62, no. 3, pp. 293-300.

Bibillo, A & Eickbush, TH 2002, 'The Reverse Transcriptase of the R2 Non-LTR Retrotransposon: Continuous Synthesis of cDNA on Non-continuous RNA Templates.', *J. Mol. Biol.*, vol 316, pp. 459-473.

Bibillo, A & Eickbush, TH 2004, 'End-to-End Template Jumping by the Reverse Transcriptase Encoded by the R2 Retrotransposon.', *J. Biol. Chem.*, vol 279, pp. 14945-14953.

Blattner, FR, Plunkett, G, Bloch, CA, Perna, NT, Burland, V, Riley, M, Collado-Vides, J, Glasner, JD, Rode, CK, Mayhew, GF, Gregor, J, Davis, NW, Kirkpatrick, HA, Goeden, MA, Rose, DJ, Nau, B & Shao, Y 1997, 'The Complete Genome Sequence of *Escherichia coli* K-12.', *Science*, vol 277, no. 5331, pp. 1453-1462.

Branlant, C, Krol, A, Ebel, JP, Lazar, E, Haendler, B & Jacob, M 1982, 'U2 RNA shares a structural domain with U1, U4, and U5 RNAs.', *EMBO J.*, vol 1, pp. 1259-1265.

Brenner, DJ, Steigerwalt, AG & McDade, JE 1979, 'Classification of the Legionnaires' Disease Bacterium: *Legionella pneumophila*, genus novum, species nova, of the Family Legionellaceae, familia nova.', *Ann Intern Med.*, vol 90, no. 4, pp. 656-658.

Büdel, B & Kauff, F 2012, 'Blue-green Algae, Myxomycetes, Myxomycete-like organisms, Phytoparasitic protists, Heterotrophic Heterokontobionta and Fungi p.p.', in W Frey (ed.), *Syllabus of Plant Families - A. Engler's Syllabus der Pflanzenfamilien*, 13th edn, Gebr. Borntraeger, Stuttgart.

Cannone, JJ, Subramanian, S, Schnare, MN, Collett, JR, D'Souza, LM, Du, Y, Feng, B, Lin, N, Madabusi, LV, Muller, KM, Pande, N, Shang, Z, Yu, N & Gutell, RR 2002, 'The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs.', *BMC Bioinformatics*, vol 3, <<http://www.rna.icmb.utexas.edu/>>.

- Cech, TR 1993, 'The RNP World', in *The RNA World*, 3rd edn, Cold Spring Harbor Laboratory Press, New York.
- Chamberlin, SI, Merino, EJ & Weeks, KM 2002, 'Catalysis of amide synthesis by RNA phosphodiester and hydroxyl groups.', *PNAS*, vol 99, pp. 14688-14693.
- Chang, TH, Huang, HY, Hsu, JB, Wenig, SL, Horng, JT & Huang, HD 2013, 'An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs.', *BMC bioinformatics*, vol 14, <<http://regrna2.mbc.nctu.edu.tw/>>.
- Chen, JL & Greiger, CW 2005, 'Functional analysis of the pseudoknot structure in human telomerase RNA.', *PNAS*, vol 102, no. 23, pp. 8080-8085.
- Chen, M, Quinnell, RG & Larkum, AWD 2002, 'The major light-harvesting pigment protein of *Acaryochloris marina*.', *FEBS Letters*, vol 514, pp. 149-152.
- Chen, Z, Yang, H & Pavletich, NP 2008, 'Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures.', *Nature*, vol 453, pp. 489-494.
- Chosa, H, Makino, S, Sasakawa, C, Okada, N, Yamada, M, Komatsu, K, Suk, JS & Yoshikawa, M 1989, 'Loss of virulence in *Shigella* strains preserved in culture collections due to molecular alteration of the invasion plasmid.', *Microb. Pathog.*, vol 6, pp. 337-342.
- Chothia, C & Lesk, AM 1987, 'Canonical structures for the hypervariable regions of immunoglobulins.', *J. Mol. Biol.*, vol 196, pp. 901-917.
- Christensen, SM & Eickbush, TH 2005, 'R2 Target-Primed Reverse Transcription: Ordered Cleavage and Polymerization Steps by Protein Subunits Asymmetrically Bound to the Target DNA.', *Mol. Cell. Biol.*, vol 25, no. 15, pp. 6617-6628.
- Cocquet, J, Chong, A, Zhang, G & Veitia, RA 2006, 'Reverse transcriptase template switching and false alternative transcripts.', *Genomics*, vol 88, pp. 127-131.
- Cole, JK, Hutchison, JR, Renslow, RS, Kim, YM, Chrisler, WB, Engelmann, HE, Dohnalkova, AC, Hu, D, Metz, TO, Fredrickson, JK & Lindemann, SR 2014, 'Phototrophic biofilm assembly in microbial-mat-derived unicyanobacterial consortia: model systems for the study of autotroph-heterotroph interactions.', *Front Microbiol*, vol 5, no. 109, p. PMC3985010.
- Cousineau, B, Smith, D, Lawrence-Cavanagh, S, Mueller, JE, Yang, J, Mills, D, Manias, D, Dunny, G, Lambowitz, AM & Belfort, M 1998, 'Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. ', *Cell*, vol 94, pp. 451-462.

Cui, X, Matsuura, M, Wang, Q, Ma, H & Lambowitz, AM 2004, 'A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing.', *J. Mol. Biol.*, vol 340, no. 2, pp. 211-231.

Dai, L, Chai, D, Gu, SQ, Gabel, J, Noskov, SY, Blocker, FJ, Lambowitz, AM & Zimmerly, S 2008, 'A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase.', *Mol. Cell*, vol 30, pp. 472-485.

Dai, W, Hodes, A, Hui, WH, Gingery, M, Miller, JF & Zhou, HZ 2010, 'Three-dimensional structure of tropism-switching Boredetella bacteriophage.', *PNAS* 107, 2010, pp. 4347-4352.

Davis, MM & Bjorkman, PJ 1988, 'T-cell antigen receptor genes and T-cell recognition.', *Nature*, vol 334, pp. 395-402.

Deigan, KE, Li, TW, Mathews, DH & Weeks, KM 2009, 'Accurate SHAPE-directed RNA structure determination.', *PNAS*, vol 106, pp. 97-102.

Delisi, C & Crothers, DM 1971, 'Prediction of RNA secondary structure.', *PNAS*, vol 68, pp. 2682-2685.

Dierks, T, Dickmanns, A, Preusser-Kunze, A, Schmidt, B, Mariappan, M, von Figura, K, Ficner, R & Rudolph, MG 2005, 'Molecular Basis for Multiple Sulfatase Deficiency and Mechanism for Formylglycine Generation of the Human Formylglycine-Generating Enzyme.', *Cell*, vol 121, no. 4, pp. 541-552.

Doshi, KJ, Cannone, JJ, Cobaugh, CW & Gutell, RR 2004, 'Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction.', *BMC Bioinformatics*, vol 5, p. 105.

Doulatov, S, Hodes, A, Dai, L, Mandhana, N, Liu, M, Deora, R, Simons, RW, Zimmerly, S & Miller, JF 2004, 'Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements', *Nature*, vol 431, pp. 476-481.

Eick, S, Straube, A, Guentsch, A, Pfister, W & Jentsch, H 2011, 'Comparison of real-time polymerase chain reaction and DNA-strip technology in microbiological evaluation of periodontitis treatment.', *Diagnostic Microbiology and Infectious Disease*, vol 69, pp. 12-20.

Fedorova, O & Zingler, N 2007, 'Group II introns: structure, folding and splicing mechanism.', *Biological chemistry*, vol 388, no. 7, pp. 665-678.

Ferat, JL & Michel, F 1993, 'Group II self-splicing introns in bacteria.', *Nature*, vol 364, pp. 358-361.

- Filiatrault, MJ, Stodghill, PV, Myers, CR, Bronstein, PA, Butcher, BG, Lam, H, Grills, G, Schweitzer, P, Wang, W, Schneider, DJ & Cartinhour, SW 2011, 'Genome-Wide Identification of Transcriptional Start Sites in the Plant Pathogen *Pseudomonas syringae* pv. Tomato str. DC3000.', *PLoS one*, vol 6, no. 12, p. e29335.
- Fiore, JL & Nesbitt, DJ 2013, 'An RNA folding motif :GNRA tetraloop-receptor interactions.', *Quarterly Reviews of Biophysics*, vol 46, no. 3, pp. 223-264.
- Flavell, AJ, Smith, DB & Kumar, A 1992, 'Extreme heterogeneity of Ty1-copia group retrotransposons in plants.', *Mol. Gen. Genet.*, vol 231, pp. 233-242.
- Friesen, N, Brandes, A & Heslop-Harrison, JS 2001, 'Diversity, Origin, and Distribution of Retrotransposons (gypsy and copia) in Conifers.', *Mol. Biol. Evol.*, vol 18, no. 7, pp. 1176-1188.
- Furtig, B, Richter, C, Wohnert, J & Schwalbe, H 2003, 'NMR spectroscopy of RNA.', *ChemBiochem.*, vol 4, no. 10, pp. 936-962.
- Gao, R, Zhao, AH, Du, Y, Ho, WT, Fu, X & Zhao, ZJ 2012, 'PCR artifacts can explain the reported biallelic JAK2 mutations.', *Blood Cancer Journal*, vol 2, p. e56.
- Gardner, PP, Wilm, A & Washietl, S 2005, 'A benchmark of multiple sequence alignment programs upon structural RNAs.', *Nucleic Acids Res.*, vol 33, pp. 2433-2439.
- Gasior, SL, Wakeman, TP, Xu, B & Deininger, PL 2006, 'The human LINE-1 retrotransposon creates DNA double-strand breaks.', *J. Mol. Biol.*, vol 357, no. 5, pp. 1383-1393.
- Gibbs, JW 1873, 'A Method of Geometrical Representation of the Thermodynamic Properties of Substances by Means of Surfaces.', *Transactions of the Connecticut Academy*, vol 2, pp. 382-404.
- Griffiths-Jones, S, Moxon, S, Marshall, M, Khanna, A & Eddy, SR 2005, 'Rfam: Annotating non-coding RNAs in complete genomes.', *Nucleic Acids Res.*, vol 33, pp. D121-D124.
- Gruber, AR, Findeiss, S, Washietl, S, Hofacker, IL & Stadler, PF 2010, 'RNAZ 2.0: Improved noncoding RNA detection.', *Pac. Symp. Biocomput.*, vol 15, pp. 69-79.
- Gumbs, OH, Padgett, RA & Dayie, KT 2006, 'Fluorescence and solution NMR study of the active site of a 160-kDa group II intron ribozyme.', *RNA*, vol 12, pp. 1693-1707.
- Guo, H, Tse, LV, Barbalat, R, Sivaamnuaiaphorn, S, Xu, M, Doulatov, S & Miller, JF 2008, 'Diversity-Generating Retroelement Homing Regenerates Target Sequences for Repeated Rounds of Codon Rewriting and Protein Diversification.', *Mol. Cell*, vol 31, pp. 813-823.

Guo, H, Tse, LV, Nieh, AW, Czornyj, E, Williams, S, Oukil, S, Liu, VB & Miller, JF 2011, 'Target Site Recognition by a Diversity-Generating Retroelement.', *PLoS Genetics*, vol 7, p. e1002414.

Guo, WT, Wang, J, Tam, G, Yen, TS & Ou, JS 1991, 'Leaky transcription termination produces larger and smaller than genome size hepatitis B virus X gene transcripts.', *Virology*, vol 181, pp. 630-636.

Gutell, RR, Lee, JC & Cannone, JJ 2002, 'The accuracy of ribosomal RNA comparative structure models.', *Current Opinion in Structural Biology*, vol 12, no. 3, pp. 301-310.

Han, JS 2010, 'Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions.', *Mobile DNA*, vol 1, no. 15, pp. 1-12.

Hancks, DC & Kazazian, HH 2012, 'Active Human Retrotransposons: Variation and Disease.', *Curr. Opin. Genet. Dev.*, vol 22, no. 3, pp. 191-203.

Havgaard, JH, Lyngso, RB, Stormo, GD & Gorodkin, J 2005, 'Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%.', *Bioinformatics*, vol 21, pp. 1815-1824.

Hofacker, IL, Fekete, M & Stadler, PF 2002, 'Secondary Structure Prediction for Aligned RNA Sequences.', *Journal of Molecular Biology*, vol 319, no. 5, pp. 1059-1066.

Hoiczyk, E & Hansel, A 2000, 'Cyanobacterial Cell Walls: News from an Unusual Prokaryotic Envelope.', *Journal of Bacteriology*, vol 182, no. 5, pp. 1191-1199.

Huang, H, Chopra, R, Verdine, GL & Harrison, SC 1998, 'Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. ', *Science*, vol 282, pp. 1669-1675.

Jaeger, L, Michel, F & Westhof, E 1994, 'Involvement of a GNRA Tetraloop in Long-range RNA Tertiary Interactions.', *J. Mol. Biol.*, vol 236, pp. 1271-1276.

Johnson, MS, McClure, MA, Feng, DF, Gray, J & Doolittle, RF 1986, 'Computer analysis of retroviral pol genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes.', *PNAS*, vol 83, pp. 7648-7652.

Joo, J, Gunny, M, Cases, M, Hudson, P, Albert, R & Harvill, E 2006, 'Bacteriophage-mediated competition in *Bordetella* bacteria.', *Proceedings of the Royal Society B: Biological Sciences*, vol 273, no. 1595, pp. 1843-1848.

- Kaneko, T, Nakamura, Y, Wolk, CP, Kuritz, T, Sasamoto, S, Watanabe, A, Iriguchi, M, Ishikawa, A, Kawashima, K, Kimura, T, Kishida, Y, Kohara, M, Matsumoto, M, Matsuno, A, (etc.) & Tabata, S 2001, 'Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120.', *DNA Research*, vol 8, pp. 205-213.
- Keller, M & Michel, F 1985, 'The introns of the *Euglena gracilis* chloroplast gene which codes for the 32-kDa protein of photosystem II.', *FEBS Lett.*, vol 179, pp. 69-73.
- Ke, N & Voytas, DF 1997, 'High Frequency cDNA Recombination of the *Saccharomyces* Retrotransposon Ty5: The LTR Mediates Formation of Tandem Elements.', *Genetics*, vol 147, pp. 545-556.
- Khan, SA 1997, 'Rolling-circle replication of bacterial plasmids.', *Microbiol. Mol. Biol. Rev.*, vol 61, no. 4, pp. 442-455.
- Kladwang, W, VanLang, CC, Cordero, P & Das, R 2011(a), 'A two-dimensional mutate-and-map strategy for non-coding RNA structure.', *Nature Chemistry*, vol 3, pp. 954-962.
- Kladwang, W, VanLang, CC, Cordero, P & Das, R 2011(b), 'Understanding the errors of SHAPE-directed RNA structure modeling.', *Biochemistry*, vol 50, no. 37, pp. 8049-8056.
- Klosterman, PS, Hendrix, DK, Tamura, M, Holbrook, SR & Brenner, SE 2004, 'Threedimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns.', *Nucleic Acids Res.*, vol 32, pp. 2342-2352.
- Konieczny, A, Voytas, D, Cummings, M & Ausubel, FM 1991, 'Konieczny A, VoytaA superfamily of *Arabidopsis thaliana* retrotransposons.', *Genetics*, vol 127, no. 4, pp. 801-809.
- Koskenniemi, K, Lyra, C, Rajaniemi-Wacklin, P, Jokela, J & Sivonen, K 2007, 'Quantitative Real-Time PCR Detection of Toxic *Nodularia* Cyanobacteria in the Baltic Sea.', *Applied and Environmental Microbiology*, vol 73, no. 7, pp. 2173-2179.
- Kurzynska-Kokorniak, A, Jamburuthugoda, VK, Bibillo, A & Eickbush, TH 2007, 'DNA-directed DNA Polymerase and Strand Displacement Activity of the Reverse Transcriptase Encoded by the R2 Retrotransposon.', *J. Mol. Biol.*, vol 374, pp. 322-333.
- Lambowitz, AM & Zimmerly, S 2004, 'Mobile group II introns.', *Annual review of genetics*, vol 38, pp. 1-35.
- Lambowitz, AM & Zimmerly, S 2011, 'Group II introns: mobile ribozymes that invade DNA.', *Cold Spring Harbor perspectives in biology*, vol 3, no. 8, p. a003616.

Lander, ES, Linton, LM, Birren, B, Nusbaum, C, Zody, MC, Baldwin, J, Devon, K, Dewar, K, Doyle, M, Aristides, P & Michael, JM 2001, 'Initial sequencing and analysis of the human genome.', *Nature*, vol 409, pp. 860-921.

Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, Gibson, TJ & Higgins, DG 2007, 'Clustal W and Clustal X version 2.0.', *Bioinformatics*, vol 23, no. 21, pp. 2947-2948.

Layton, DM & Bundschuh, R 2005, 'A statistical analysis of RNA folding algorithms through thermodynamic parameter perturbation.', *Nucleic Acids Res.*, vol 33, no. 2, pp. 519-524.

Le Coq, J & Ghosh, P 2011, 'Conservation of the C-type lectin fold for massive sequence variation in a *Treponema* diversity-generating retroelement.', *PNAS*, vol 108, pp. 14649-14653.

Leonard, CW, Hajdin, CE, Karabiber, F, Mathews, DH, Favorov, OV, Dokholyan, NV & Weeks, KM 2013, 'Principles for Understanding the Accuracy of SHAPE-Directed RNA Structure Modeling.', *Biochemistry*, vol 52, no. 4, pp. 588-595.

Li, X,Q,LDAM 2010, 'Predicting in vivo binding sites of RNA-binding proteins using mRNA secondary structure.', *RNA*, vol 16, pp. 1096-1107.

Liu, M, Deora, R, Doulatov, SR, Gingery, M, Eiserling, FA, Preston, A, Maskell, DJ, Simons, RW, Cotter, PA, Parkhill, J & Miller, JF 2002, 'Reverse Transcriptase-Mediated Tropism Switching in *Bordetella* Bacteriophage.', *Science*, vol 295, pp. 2091-2094.

Liu, Z, Macias, MJ, Bottomley, MJ, Stier, G, Linge, JP, Nilges, M, Bork, P & Sattler, M 1999, 'The three-dimensional structure of the HRDC domain and implications for the Werner and Bloom syndrome proteins.', *Structure*, vol 7, no. 12, pp. 1557-1566.

Loesche, WJ & Grossman, NS 2001, 'Periodontal disease as a specific, albeit chronic, infection: Diagnosis and treatment.', *Clin. Microbiol. Rev.*, vol 14, no. 4, pp. 727-752.

Luan, DD, Korman, MH, Jakubczak, JL & Eickbush, TH 1993, 'Reverse transcription of R2BmRNAs primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.', *Cell*, vol 72, pp. 595-605.

Manthey, JA 2005, 'mFold, Delta G, and Melting Temperature: What Does it Mean?', *Integrated DNA Technologies*, vol Bioinformatics Group.

Martínez-Abarca, F, Barrientos-Durán, A, Fernández-López, M & Toro, N 2004, 'The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the

nascent lagging strand at DNA replication forks for priming.', *Nucleic Acids Res.*, vol 32, no. 9, pp. 2880-2888.

McCaskill, JS 1990, 'The equilibrium partition function and base pair binding probabilities for RNA secondary structure.', *Biopolymers*, vol 29, pp. 1105-1119.

McClintock, B 1950, 'The Origin and Behavior of Mutable Loci in Maize.', *PNAS*, vol 36, pp. 344-355.

McMahon, SA, Miller, JL, Lawton, JA, Kerkow, DE, Hodes, A, Marti-Renom, MA, Doulatov, S, Narayanan, E, Sali, A, Miller, JF & Ghosh, P 2005, 'The C-type lectin fold as an evolutionary solution for massive sequence variation.', *Nat. Struct. Mol. Biol.*, vol 12, pp. 886-892.

Medhekar, B & Miller, JF 2007, 'Diversity-generating retroelements.', *Current Opinion in Microbiology*, vol 10, pp. 388-395.

Meeks, JC & Elhai, J 2002, 'Regulation of Cellular Differentiation in Filamentous Cyanobacteria in Free-Living and Plant-Associated Symbiotic Growth States.', *Microbiol. Mol. Biol. Rev.*, vol 66, no. 1, pp. 94-121.

Merino, EJ, Wilkinson, KA, Coughlan, JL & Weeks, KM 2005, 'RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE).', *J. Am. Chem. Soc.*, vol 127, pp. 4223-4231.

Michel, F & Ferat, JL 1995, 'Structure and activities of group II introns.', *Annu. Rev. Biochem.*, vol 64, pp. 435-461.

Michel, F & Westhof, E 1990, 'Modeling of the 3-dimensional architecture of group-I catalytic introns based on comparative sequence analysis.', *Journal of Molecular Biology*, vol 216, pp. 585-610.

Miller, JL, Le Coq, J, Hodes, A, Barbalat, R, Miller, JF & Ghosh, P 2008, 'Selective Ligand Recognition by a Diversity-Generating Retroelement Variable Protein.', *PLoS Biology*, vol 6, pp. 1195-1207.

Minh, BQ, Nguyen, MA & von Haeseler, A 2013, 'Ultrafast Approximation for Phylogenetic Bootstrap.', *Mol. Biol. Evol.*, vol 30, no. 5, pp. 1188-1195.

Minot, S, Bryson, A, Chehoud, C, Wu, GD, Lewis, JD & Bushman, FD 2013, 'Rapid evolution of the human gut virome.', *PNAS*, vol 110, no. 30, pp. 12450-12455.

Missal, K, Rose, D & Stadler, PF 2005, 'Non-coding RNAs in *Ciona intestinalis*.', *Bioinformatics*, vol 21, pp. i77-i78.

Mitschke, J, Vioque, A, Haas, F, Hess, WR & Muro-Pastor, AM 2011, 'Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120.', *PNAS*, vol 108, pp. 20130-20135.

Møller, T, Franch, T, Højrup, P, Keene, DR, Bächinger, HP, Brennan, RG & Valentin-Hansen, P 2002, 'Hfq: A Bacterial Sm-like Protein that Mediates RNA-RNA Interaction.', *Mol. Cell*, vol 9, no. 1, pp. 23-30.

Moran, JV, Mecklenburg, KL, Sass, P, Belcher, SM, Mahnke, D, Lewin, A & Perlman, P 1994, 'Splicing defective mutants of the COXI gene of yeast mitochondrial DNA: Initial definition of the maturase domain of the group II intron AI2.', *Nucleic Acids Res.*, vol 22, pp. 2057-2064.

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.*, vol 51, pp. 263-273.

Murina, VN & Nikulin, AD 2011, 'RNA-Binding Sm-Like Proteins of Bacteria and Archaea. Similarity and Difference in Structure and Function.', *Biochemistry*, vol 76, no. 13, pp. 1434-1449.

Murzin, AG, Brenner, SE, Hubbard, T & Chothia, C 1995, 'SCOP: A structural classification of proteins database for the investigation of sequences and structures.', *J. Mol. Biol.*, vol 247, pp. 536-540.

Ochman, H, Gerber, AS & Hartl, DL 1988, 'Genetic applications of an inverse polymerase chain reaction.', *Genetics*, vol 120, no. 3, pp. 621-623.

Oren, A 2000, 'Salts and brines ', in BA Whitton, M Potts (eds.), *The Ecology of Cyanobacteria*, Kluwer Academic Publishers, Dordrecht.

Overstreet, CM, Yuan, TZ, Levin, AM, Kong, C, Coroneus, JG & Weiss, GA 2012, 'Self-made phage libraries with heterologous inserts in the Mtd of *Bordetella bronchiseptica*.', *PEDS*, pp. 1-7.

Pace, NR, Thomas, BC & Woese, CR 1999, 'Probing RNA structure, function, and history by comparative analysis.', in JF Atkins, RF Gesteland, TR Cech (eds.), *The RNA world*, 2nd edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

- Pang, PS, Elazar, M, Pham, EA & Glenn, JS 2011, 'Simplified RNA secondary structure mapping by automation of SHAPE data analysis.', *Nucleic Acids Res.*, vol 39, no. 22, p. e151.
- Pedersen, JS, Bejerano, G, Siepel, A, Rosenbloom, K, Lindblad-Toh, K, Lander, ES, Kent, J, Miller, W & Haussler, D 2006, 'Identification and classification of conserved RNA secondary structures in the human genome.', *PLoS Comput. Biol.*, vol 2, p. e33.
- Peebles, CL, Perlman, PS, Mecklenburg, KL, Petrillo, ML, Tabor, JH, A, JK & Cheng, HL 1986, 'A self-splicing RNA excises an intron lariat.', *Cell*, vol 44, no. 2, pp. 213-223.
- Peters, GA & Mayne, BC 1974, 'The Azolla, Anabaena Azollae Relationship I. INITIAL CHARACTERIZATION OF THE ASSOCIATION.', *Plant Physiol.*, vol 53, pp. 813-819.
- Pley, HW, Flaherty, KM & MCKay, DB 1994, 'Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix.', *Nature*, vol 372, pp. 111-113.
- Preston, BD, Poiesz, BJ & Loeb, LA 1988, 'Fidelity of HIV-1 reverse transcriptase.', *Science*, vol 242, pp. 1168-1171.
- Pyle, AM 2010, 'The tertiary structure of group II introns: implications for biological function and evolution.', *Crit. Rev. Biochem. Mol. Biol.*, vol 45, pp. 215-232.
- Qin, PZ & Pyle, AM 1997, 'Stopped-Flow Fluorescence Spectroscopy of a Group II Intron Ribozyme Reveals that Domain 1 Is an Independent Folding Unit with a Requirement for Specific Mg²⁺ Ions in the Tertiary Structure.', *Biochemistry*, vol 36, no. 16, pp. 4718-4730.
- R-Development-Core-Team 2008, 'R: A language and environment for statistical computing.', <<http://www.R-project.org>>.
- Reche, P 2008, *SIAS (sequences identities and similarities)*, <<http://imed.med.ucm.es/Tools/sias.html>>.
- Reeder, J & Giegerich, R 2004, 'Design, implementation, and evaluation of a practical pseudoknot folding algorithm based on thermodynamics.', *BMC Bioinformatics*, vol 5, no. 104.
- Ren, J, Rastegari, B, Condon, A & Hoos, HH 2005, 'HotKnots: Heuristic prediction of RNA secondary structures including pseudoknots.', *RNA*, vol 11, pp. 1494-1504.
- Reuter, JS & Mathews, DH 2010, 'RNAstructure: software for RNA secondary structure prediction and analysis.', *BMC Bioinformatics*, vol 11, p. 129.

Rice, P, Longden, I & Bleasby, A 2000, 'EMBOSS: The European Molecular Biology Open Software Suite.', *Trends in Genetics*, vol 16, no. 6, pp. 276-277, <<http://mobyli.pasteur.fr/cgi-bin/portal.py#forms:palindrome>>.

Richards, J, Luciano, DJ & Belasco, JG 2012, 'Influence of translation on RppH-dependent mRNA degradation in Escherichia coli.', *Mol. Microbiol.*, vol 86, no. 5, pp. 1063-1072.

Rikkinen, J 2013, 'Molecular studies on cyanobacterial diversity in lichen symbioses.', *MycKeys*, vol 6, pp. 3-32.

Robart, AR, Seo, W & Zimmerly, S 2007, 'Insertion of group II intron retroelements after intrinsic transcriptional terminators.', *PNAS*, vol 104, pp. 6620-6625.

Robart, AR & Zimmerly, S 2005, 'Group II intron retroelements: function and diversity. ', *Cytogenetic and genome research*, vol 110, pp. 589-597.

Roberts, JD, Bebenek, K & Kunkel, TA 1988, 'The accuracy of reverse transcriptase from HIV-1.', *Science*, vol 242, pp. 1171-1173.

Ruan, J, Stormo, GD & Zhang, W 2004, 'An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots.', *Bioinformatics*, vol 20, pp. 58-66.

Sambrook, J & Russell, DW 2001, *Molecular Cloning*, 3rd edn, Cold Spring harbor Laboratory Press, New York.

Sanger, F, Nicklen, S & Coulson, AR 1977, 'DNA sequencing with chain-terminating inhibitors.', *PNAS*, vol 74, pp. 5463-5467.

Sankoff, D 1985, 'Simultaneous solution of the RNA folding, alignment and protosequence problems.', *SIAM J. Appl. Math.*, vol 45, pp. 810-825.

Santangelo, TJ & Artsimovitch, I 2011, 'Termination and antitermination: RNA polymerase runs a stop sign.', *Nature Reviews Microbiology*, vol 9, pp. 319-329.

Sarkar, I, Hauber, I, Hauber, J & Buchholz, F 2007, 'HIV-1 Proviral DNA Excision Using an Evolved Recombinase.', *Science*, vol 316, pp. 1912-1915.

Sayak, G, Dey, SK, Priyanka, D, Protip, B, Paushali, R & Abhijit, D 2010, 'Catalytic RNA world relics in Dicer RNAs.', *International Journal of Genetics*, vol 2, no. 1, pp. 8-17.

Sayers, EW, Barrett, T, Benson, DA, Bryant, SH, Canese, K, Chetvernin, V, Church, DM, DiCuccio, M, Edgar, R, Federhen, S, Feolo, M, Geer, LY, Helmberg, W, Kapustin, Y,

- Landsman, D, Lipman, DJ, Madden, TL, Maglott, D, Miller, V & Ye, J 2009, 'Database resources of the National Center for Biotechnology Information.', *Nucleic Acids Res.*, vol 37, no. 9, pp. D5-15.
- Schillinger, T 2013, *Diversitätsgenerierende Retroelemente – Identifikation, Klassifizierung, Phylogenie und in vitro Funktionsanalysen.*, unv. Dissertation, TU Kaiserslautern.
- Schillinger, T, Lisfi, M, Chi, J, Cullum, J & Zingler, N 2012, 'Analysis of a comprehensive dataset of diversity generating retroelements generated by the program DiGReF.', *BMC Genomics*, vol 13, p. 430.
- Schillinger, T & Zingler, N 2012, 'The low incidence of diversity-generating retroelements in sequenced genomes.', *Mobile Genetic Elements*, vol 2, no. 6, pp. 1-5.
- Schmelzer, C & Schweyen, RJ 1986, 'Self-Splicing of Group II Introns In Vitro: Mapping of the Branch Point and Mutational Inhibition of Lariat Formation.', *Cell*, vol 46, pp. 557-565.
- Schneider, CA, Rasband, WS & Eliceiri, KW 2012, 'NIH Image to ImageJ: 25 years of image analysis.', *Nat Methods*, vol 9, no. 7, pp. 671-675.
- Seshadri, R, Myers, GSA, Tettelin, H, Eisen, JA, Heidelberg, JF, Dodson, RJ, Davidsen, TMA, DeBoy, RT, Fouts, DE, Haft, DH, Selengut, J, Ren, Q, Brinkac, LM, Madupu, R, Kolonay, J, Durkin, SA, Daugherty, SC, Shetty, J, (etc.) & Paulsen, IT 2004, 'Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes.', *PNAS*, vol 101, pp. 5646-5651.
- Shearman, C, Godon, JJ & Gasson, M 1996, 'Splicing of a group II intron in a functional transfer gene of *Lactococcus lactis*.', *Mol. Microbiol.*, vol 21, pp. 45-53.
- Simon, D & Zimmerly, S 2008, 'A diversity of uncharacterized reverse transcriptases in bacteria.', *Nucleic Acids Res.*, vol 36, pp. 7219-7229.
- Singh, NN & Lambowitz, AM 2001, 'Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference.', *J. Mol. Biol.*, vol 309, pp. 361-386.
- Smith, C, Heyne, S, Richter, AS, Will, S & Backofen, R 2010, 'Freiburg RNA Tools: a web server integrating IntaRNA, ExpaRNA and LocARNA.', *Nucleic Acids Res.*, vol 38, pp. W373-377.

Smith, D, Zhong, J, Matsuura, M, Lambowitz, AM & Belfort, M 2005, 'Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming.', *Genes Dev*, vol 19, pp. 2477-2487.

Stanier, RY, Kunisawa, R, Mandel, M & Cohen-Bazire, G 1971, 'Purification and properties of unicellular blue-green algae (order Chroococcales).', *Bacteriol. Rev.*, vol 35, pp. 171-205.

Staub, R 1961, 'Ernährungsphysiologisch-autökologische Untersuchungen an der planktischen Blaualge *Oscillatoria rubescens* DC.', *Schweiz Zeit f Hydrologie*, vol 23, no. 1, pp. 82-198.

Stewart, WD, Rowell, P & Rai, AN 1983, 'Cyanobacteria-eukaryotic plant symbioses.', *Ann. Microbiol. (Paris)*, vol 134, no. 1, pp. 205-228.

Sükösd, Z, Swenson, MS, Kjems, J & Heitsch, CE 2013, 'Evaluating the accuracy of SHAPE-directed RNA secondary structure predictions.', *Nucleic Acids Res.*, vol 41, no. 5, pp. 2807-2816.

Suzuki, I, Kanesaki, Y, Mikami, K, Kanehisa, M & Murata, N 2001, 'Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*.', *Molecular Microbiology*, vol 40, no. 1, pp. 235-244.

Svarovskaia, ES, Delviks, KA, Hwang, CK & Pathak, VK 2000, 'Structural Determinants of Murine Leukemia Virus Reverse Transcriptase That Affect the Frequency of Template Switching.', *J. Virol.*, vol 74, no. 15, pp. 7171-7178.

Taylor, TN & Taylor, EL 1993, *The Biology and Evolution of Fossil Plants.*, Prentice Hall, New Jersey.

Thiel, T, Pratte, BS, Zhong, J, Goodwin, L, Copeland, A, Lucas, S, Han, C, Pitluck, S, Land, ML, Kyripides, NC & Woyke, T 2014, 'Complete genome sequence of *Anabaena variabilis* ATCC 29413.', *SIGS*, vol 9, no. 3, pp. 562-573.

Tijerina, P, Mohr, S & Russel, R 2007, 'DMS Footprinting of Structured RNAs and RNA-Protein Complexes.', *Nat. Protoc.*, vol 2, no. 10, pp. 2608-2623.

Tinoco, IJR & Bustamante, C 1999, 'How RNA folds.', *J. Mol. Biol.*, vol 293, pp. 271-281.

Tinoco, I, Uhlenbeck, OC & Levine, MD 1971, 'Estimation of secondary structure in ribonucleic acids.', *Nature*, vol 230, pp. 362-367.

Toor, N, Keating, KS, Fedorova, O, Rajashankar, K, Wang, J & Pyle, AM 2010, 'Tertiary architecture of the *Oceanobacillus iheyensis* group II intron.', *RNA*, vol 16, pp. 57-69.

Toor, N, Keating, KS, Taylor, SD & Pyle, AM 2008, 'Crystal structure of a self-spliced group II intron.', *Science*, vol 320, pp. 77-82.

Urlaub, H, Raker, VA, Kostka, S & Lührmann, R 2001, 'Sm protein–Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure.', *EMBO J*, vol 20, no. 1, pp. 187-196.

Van der Veen, R, Arnberg, AC, Van der Horst, G, Bonen, L, Tabak, HF & Grivell, LA 1986, 'Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro.', *Cell*, vol 44, pp. 225-234.

Vasa, SM, Guex, N, Wilkinson, KA, Weeks, KM & Giddings, MC 2008, 'ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis.', *RNA*, vol 14, pp. 1979-1990.

Vincent, WF 2000, 'Cyanobacterial dominance in the polar regions.', in BA Whitton, M Potts (eds.), *The Ecology of Cyanobacteria*, Kluwer Academic Publishers, Dordrecht.

Washietl, S, Hofacker, IL & Stadler, PF 2005, 'Fast and reliable prediction of noncoding RNAs.', *PNAS*, vol 102, pp. 2454-2459.

Washietl, S, Hofacker, IL, Stadler, PF & Kellis, M 2012, 'RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction.', *Nucleic Acids Res.*, vol 40, no. 10, pp. 4261-4272.

Watanabe, K & Lambowitz, AM 2004, 'High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA.', *RNA*, vol 10, pp. 1433-1443.

Whelan, S & Goldman, N 2001, 'A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach.', *Mol. Biol. Evol.*, vol 18, no. 5, pp. 691-699.

Wilkinson, KA, Gorelick, RJ, Vasa, SM, Guex, N, Rein, A, Mathews, DH, Giddings, MC & Weeks, KM 2008, 'High-Throughput SHAPE Analysis Reveals Structures in HIV-1 Genomic RNA Strongly Conserved across Distinct Biological States.', *PLoS Biology*, vol 6, no. 4, p. e96.

Wilkinson, KA, Merino, EJ & Weeks, KM 2006, 'Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution.', *Nature protocols*, vol 1, no. 3, pp. 1610-1616.

Will, S, Joshi, T, Hofacker, IL, Stadler, PF & Backofen, R 2012, 'LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs.', *RNA*, vol 18, no. 5, pp. 900-914.

Will, S, Reiche, K, Hofacker, IL, Stadler, PF & Backofen, R 2007, 'Inferring non-coding RNA families and classes by means of genome-scale structure-based clustering.', *PLoS Computational Biology*, vol 3, no. 4, p. e65.

Wimberly, BT, Brodersen, DE, Clemons, WMJR, Morgan-Warren, RJ, Carter, AP, Vornrhein, C, Hartsch, T & Ramakrishnan, V 2000, 'Structure of the 30S ribosomal subunit.', *Nature*, vol 407, pp. 327-339.

Wisniewski, M, Balakrishnan, M, Palaniappan, C, Fay, PJ & Bambara, RA 2000, 'Unique progressive cleavage mechanism of HIV reverse transcriptase RNase H.', *PNAS*, vol 97, no. 22, pp. 11978-11983.

Wool, IG, Gluck, A & Endo, Y 1992, 'Ribotoxin recognition of ribosomal RNA and a proposal for the mechanism of translocation.', *Trends. Biochem. Sci.*, vol 17, pp. 266-269.

Wynn-Williams, DD 2000, 'Cyanobacteria in deserts – life at the limit.', in BA Whitton, M Potts (eds.), *The Ecology of Cyanobacteria*, Kluwer Academic Publishers, Dordrecht.

Xiong, J, Inoue, K & Bauer, CE 1998, 'Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*.', *PNAS*, vol 95, no. 25, pp. 14851-14856.

Ye, Y 2014, 'Identification of Diversity-Generating Retroelements in Human Microbiomes.', *Int. J. Mol. Sci.*, vol 15, pp. 14234-14246.

Zhang, L & Doudna, JA 2002, 'Structural Insights into Group II Intron Catalysis and Branch-Site Selection.', *Science*, vol 295, no. 5562, pp. 2084-2088.

Zhong, J & Lambowitz, AM 2003, 'Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription.', *EMBO J.*, vol 22, no. 17, pp. 4555-4565.

Zimmerly, S, Guo, H, Perlman, PS & Lambowitz, AM 1995, 'Group II intron mobility occurs by target DNA-primed reverse transcription.', *Cell*, vol 82, pp. 545-554.

Zucker, M 2003, 'Mfold web server for nucleic acid folding and hybridization prediction.', *Nucleic Acids Res.*, vol 31, no. 13, pp. 3406-3415, <<http://mfold.rna.albany.edu>>.

Zuker, M 1989, 'On finding all suboptimal foldings of an RNA Molecule.', *Science*, vol 244, pp. 48-52.

Zuker, M & Sankoff, D 1984, 'RNA Secondary Structures and their Prediction.', *B. Math. Biol.*, vol 46, pp. 591-621.

Anhang

Tabelle A1. Liste der Cyanobakteriensammlung

Name	Ordnung	BB	PCR	Herkunft	Datum
<i>Anabaena cylindrica</i>	<i>Nostocales</i>	97.11	-	Kanada	1966
<i>Anabaena flos-aquae</i>	<i>Nostocales</i>	97.35	+	Schweiz	1965
<i>Anabaena planctonica</i>	<i>Nostocales</i>	97.102	+	Schweiz	1979
<i>Anabaena planctonica</i>	<i>Nostocales</i>	97.34	0	N.S.	N.S.
<i>Aphanizomenon flos-aquae</i>	<i>Nostocales</i>	97.111b	+	Schweiz	1990
<i>Aphanizomenon flos-aquae</i>	<i>Nostocales</i>	97.85	++	Schweiz	1974
<i>Aphanizomenon flos-aquae</i>	<i>Nostocales</i>	97.25	0	Schweiz	1963
<i>Aphanocapsa grevillei</i>	<i>Chroococcales</i>	97.29	-	Schweiz	1964
<i>Chamaesiphon polonicus</i>	<i>Synechococcales</i>	97.37	0	Schweiz	1965
<i>Chlorogloea fritschii</i>	<i>Chroococcales</i>	97.1	0	Indien	N.S.
<i>Chroococcales sp.</i>	--	89.2	-	Spanien	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	01.01	+	N.S.	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	79.1	-	Österreich	1979
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	04.1	0	Antarktika	2003
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	81.1	-	Österreich	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	82.1	-	Mexiko	1982
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	82.1a	-	Mexiko	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	82.1b	-	N.S.	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	82.3	-	Mexiko	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	84.1	-	Hawaii	1981
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	95.2	-	Südafrika	1994
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	95.6	-	Südafrika	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	96.1a	-	N.S.	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	96.1b	-	N.S.	N.S.
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	90.5	0	Südafrika	1990
<i>Chroococcidiopsis sp.</i>	ungewisse Ordnung	90.1	0	Südafrika	1990
<i>Chroococcidopsis sp.</i>	ungewisse Ordnung	79.2	0	Österreich	N.S.
<i>Chroococcidiopsis thermalis</i>	ungewisse Ordnung	82.2	++	Kuba	N.S.
<i>Chroococcus minor</i>	<i>Chroococcales</i>	97.45	0	Schweiz	1965
<i>Chroococcus sp.</i>	<i>Chroococcales</i>	97.96	+	Slowakei	1977
<i>Chroococcus sp.</i>	<i>Chroococcales</i>	97.97	-	Slowakei	1977
<i>Cyanosarcina sp.</i>	<i>Chroococcales</i>	80.2	-	Spanien	1980

Name	Ordnung	BB	PCR	Herkunft	Datum
<i>Cylindrospermum sp.</i>	Nostocales	97.12	+	Kamerun	1976
<i>Dermocapsa</i>	--	96.19	-	Südafrika	N.S.
<i>Dichothrix orsiniana</i>	Nostocales	97.98b	-	Slowakei	1976
<i>Dichothrix sp.</i>	Nostocales	97.87	+	N.S.	N.S.
<i>Entophysalidaceae</i>	Chroococcales	81.2	-	Tschechien	N.S.
<i>Fischerella ambigua</i>	Nostocales	97.28b	0	Schweiz	1965
<i>Fischerella muscicola</i>	Nostocales	98.1	+	N.S.	N.S.
<i>Gloeocapsa c.f. montana</i>	Chroococcales	80.3	-	Spanien	1980
<i>Gloeocapsa sp.</i>	Chroococcales	80.5	-	Spanien	N.S.
<i>Gloeocapsa sp.</i>	Chroococcales	81.3	0	Tschechien	N.S.
<i>Gloeothece fusco-lutea</i>	Chroococcales	97.36	-	Schweiz	1965
<i>Gloeothece sp.</i>	Chroococcales	89.15	-	U.S.A.	1988
<i>Gloeothece sp.</i>	Chroococcales	89.5	0	U.S.A.	1989
<i>Gloeothece sp.</i>	Chroococcales	89.6	0	U.S.A.	1988
<i>Gloeotrichia echinulata</i>	Nostocales	97.72	0	N.S.	N.S.
<i>Lyngbya putealis</i>	Oscillatoriales	97.125	-	Schweiz	1996
<i>Lyngbya sp.</i>	Oscillatoriales	97.64	++	Nepal	1967
<i>Lyngbya sp.</i>	Oscillatoriales	97.65b	++	Nepal	1967
<i>Lyngbya sp.</i>	Oscillatoriales	97.54	0	Schweiz	1965
<i>Microcoleus chthonoplastes</i>	Oscillatoriales	02.16	0	Namibia	N.S.
<i>Microcoleus chthonoplastes</i>	Oscillatoriales	02.17	0	Namibia	N.S.
<i>Microcoleus chthonoplastes</i>	Oscillatoriales	02.2	+	Namibia	N.S.
<i>Microcoleus chthonoplastes</i>	Oscillatoriales	92.3	+	Deutschland	1984
<i>Microcystis aeruginosa</i>	Chroococcales	97.43	+	Polen	1965
<i>Microcystis aeruginosa</i>	Chroococcales	97.30	0	Kanada	1958
<i>Microcystis aeruginosa</i>	Chroococcales	97.31	0	Kanada	1955
<i>Microcystis aeruginosa</i>	Chroococcales	97.14	-	Kanada	N.S.
<i>Microcystis aeruginosa</i>	Chroococcales	97.15	-	Kanada	N.S.
<i>Microcystis aeruginosa</i>	Chroococcales	97.16	-	Kanada	N.S.
<i>Microcystis aeruginosa</i>	Chroococcales	97.44	++	Polen	1965
<i>Microcystis aeruginosa</i>	Chroococcales	97.68	-	Kanada	1954
<i>Microcystis aeruginosa</i>	Chroococcales	97.69	-	Kanada	1955
<i>Microcystis aeruginosa</i>	Chroococcales	97.70	-	Kanada	1955
<i>Microcystis aeruginosa</i>	Chroococcales	97.71	-	N.S.	1956
<i>Microcystis aeruginosa</i>	Chroococcales	97.83c	0	Schweiz	1971
<i>Microcystis sp.</i>	Chroococcales	97.114	-	N.S.	N.S.

Name	Ordnung	BB	PCR	Herkunft	Datum
<i>Microcystis sp.</i>	<i>Chroococcales</i>	97.115	-	N.S.	N.S.
<i>Microcystis sp.</i>	<i>Chroococcales</i>	97.73	-	N.S.	1955
<i>Microcystis sp.</i>	<i>Chroococcales</i>	97.39	0	Kanada	1958
<i>Myxosarcina sp.</i>	<i>Pleurocapsales</i>	97.116	-	Schweiz	1991
N.S.	--	97.108	+	N.S.	N.S.
<i>Nostoc commune Vaucher</i>	<i>Nostocales</i>	97.40	+	Schweiz	1965
<i>Nostoc commune Vaucher</i>	<i>Nostocales</i>	97.122	-	Schweiz	1996
<i>Nostoc commune Vaucher</i>	<i>Nostocales</i>	97.41	-	Schweiz	1965
<i>Nostoc commune Vaucher</i>	<i>Nostocales</i>	97.42	-	Schweiz	1965
<i>Nostoc microscopicum</i>	<i>Nostocales</i>	97.32	-	Schweiz	1965
<i>Nostoc muscorum</i>	<i>Nostocales</i>	90.3	0	PCC7906	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	95.9	0	Deutschland	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	97.9	-	Schweiz	1958
<i>Nostoc sp.</i>	<i>Nostocales</i>	94.3	-	Mexiko	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	95.10	-	Deutschland	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	95.11	-	Deutschland	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	97.128	0	N.S.	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	90.10	0	N.S.	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	94.2	0	Südafrika	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	97.53	+	Schweiz	1966
<i>Nostoc sp.</i>	<i>Nostocales</i>	97.88	+	Schweiz	1976
<i>Nostoc sp.</i>	<i>Nostocales</i>	94.4	-	Südafrika	N.S.
<i>Nostoc sp.</i>	<i>Nostocales</i>	95.1	-	Südafrika	1994
<i>Nostoc sp.</i>	<i>Nostocales</i>	95.4	-	Südafrika	1994
<i>Nostoc sp.</i>	<i>Nostocales</i>	97.66	++	Nepal	1967
<i>Nostoc sp.</i>	<i>Nostocales</i>	94.1	0	Deutschland	1994
<i>Nostoc sphaericum Vaucher</i>	<i>Nostocales</i>	97.51	0	Schweiz	1965
<i>Nostoc sphaericum Vaucher</i>	<i>Nostocales</i>	97.55	0	Schweiz	1965
<i>Nostochopsis lobatus</i>	<i>Nostocales</i>	92.1	+	Südafrika	1990
<i>Oscillatoria amoena</i>	<i>Oscillatoriales</i>	97.119	-	Schweiz	1996
<i>Oscillatoria limosa</i>	<i>Oscillatoriales</i>	97.120	0	Schweiz	1996
<i>Oscillatoria tenuis</i>	<i>Oscillatoriales</i>	97.121	++	Schweiz	1996
<i>Oscillatoria cortiana</i>	<i>Oscillatoriales</i>	97.92b	-	Slowakei	1976
<i>Oscillatoria formosa</i>	<i>Oscillatoriales</i>	97.123	++	Schweiz	1996
<i>Oscillatoria limosa</i>	<i>Oscillatoriales</i>	97.105	+	Schweiz	1983
<i>Oscillatoria limosa</i>	<i>Oscillatoriales</i>	97.52	++	Schweiz	1966

Name	Ordnung	BB	PCR	Herkunft	Datum
<i>Oscillatoria rubescens</i>	Oscillatoriales	97.107	+	N.S.	1979
<i>Oscillatoria rubescens</i>	Oscillatoriales	97.129	0	N.S.	N.S.
<i>Oscillatoria rubescens</i>	Oscillatoriales	97.94	++	Schweiz	1976
<i>Oscillatoria rubescens</i>	Oscillatoriales	97.5	0	Schweiz	1959
<i>Oscillatoria tortuosa</i>	Oscillatoriales	02.8	++	Namibia	N.S.
<i>Petalonema alatum</i>	Nostocales	97.79d	0	Schweiz	1969
<i>Petalonema alatum</i>	Nostocales	97.82	0	Schweiz	1969
<i>Phormidium autumnale</i>	Oscillatoriales	97.19	-	Schweiz	1964
<i>Phormidium autumnale</i>	Oscillatoriales	97.20	-	Schweiz	1964
<i>Phormidium autumnale</i>	Oscillatoriales	97.21	-	N.S.	N.S.
<i>Phormidium favosum</i>	Oscillatoriales	97.124	+	Schweiz	1996
<i>Phormidium favosum</i>	Oscillatoriales	97.46	-	Schweiz	1965
<i>Phormidium frigidum</i>	Oscillatoriales	97.47	0	Schweiz	1965
<i>Phormidium sp.</i>	Oscillatoriales	02.4	0	Namibia	N.S.
<i>Phormidium sp.</i>	Oscillatoriales	02.9	0	Namibia	N.S.
<i>Phormidium sp.</i>	Oscillatoriales	97.49	++	Schweiz	1965
<i>Phormidium sp.</i>	Oscillatoriales	97.50	-	Schweiz	1965
<i>Phormidium sp.</i>	Oscillatoriales	97.75	-	Nepal	1967
<i>Plectonema boryanum</i>	Oscillatoriales	97.56	0	Schweiz	1967
<i>Plectonema sp.</i>	Oscillatoriales	97.67	0	Nepal	1967
<i>Pleurocapsa minor</i>	Pleurocapsales	89.12	0	Namibia	1988
<i>Pleurocapsa sp.</i>	Pleurocapsales	97.117	0	Schweiz	1991
<i>Pseudanabaena sp.</i>	Pseudanabaenales	02.18	++	Namibia	N.S.
<i>Pseudanabaena catenata</i>	Pseudanabaenales	02.13	-	Namibia	N.S.
<i>Pseudanabaena catenata</i>	Pseudanabaenales	97.2	-	Schweiz	1957
<i>Rivularia biasolettiana</i>	Nostocales	97.101	++	Schweden	1978
<i>Rivularia sp.</i>	Nostocales	97.24	+	Schweiz	1964
<i>Scytonema capitatum</i>	Nostocales	97.62	+	Nepal	1967
<i>Scytonema capitatum</i>	Nostocales	97.59b	-	Nepal	1967
<i>Scytonema capitatum</i>	Nostocales	97.60	-	Nepal	1967
<i>Scytonema capitatum</i>	Nostocales	97.91	-	Schweiz	1976
<i>Scytonema capitatum</i>	Nostocales	97.61	0	Nepal	1967
<i>Scytonema javanicum</i>	Nostocales	96.12	0	Costa Rica	N.S.
<i>Scytonema javanicum</i>	Nostocales	96.13	0	Costa Rica	N.S.
<i>Scytonema javanicum</i>	Nostocales	96.14	0	Costa Rica	N.S.
<i>Scytonema javanicum</i>	Nostocales	97.58	+	Nepal	1968

Name	Ordnung	BB	PCR	Herkunft	Datum
<i>Scytonema myochrous</i>	<i>Nostocales</i>	97.84	+	Schweiz	1973
<i>Scytonema myochrous</i>	<i>Nostocales</i>	97.106a	0	Schweiz	1979
<i>Scytonema myochrous</i>	<i>Nostocales</i>	97.77d	0	Schweiz	1969
<i>Scytonema myochrous.</i>	<i>Nostocales</i>	97.22	+	Schweiz	1964
<i>Scytonema ocellatum</i>	<i>Nostocales</i>	02.1	-	Namibia	N.S.
<i>Scytonema sp.</i>	<i>Nostocales</i>	97.127b	+	N.S.	N.S.
<i>Scytonema sp.</i>	<i>Nostocales</i>	97.63	-	Nepal	1967
<i>Scytonema sp.</i>	<i>Nostocales</i>	97.90c	0	Schweiz	1976
<i>Scytonema stuposum</i>	<i>Nostocales</i>	97.78	+	Nepal	1967
<i>Scytonema stuposum</i>	<i>Nostocales</i>	02.11	-	Namibia	N.S.
<i>Stigonema mamillosum</i>	<i>Nostocales</i>	97.104b	+	Schweiz	1979
<i>Stigonema mammosum</i>	<i>Nostocales</i>	96.17	-	Brasilien	N.S.
<i>Stigonema panniforme</i>	<i>Nostocales</i>	97.27	+	Schweiz	1964
<i>Synechococcus elongatus</i>	<i>Synechococcales</i>	98.10	0	N.S.	N.S.
<i>Synechococcus elongatus</i>	<i>Synechococcales</i>	97.3	0	England	1940
<i>Synechocystis sp.</i>	<i>Synechococcales</i>	98.3	-	SAG 51.71	N.S.
<i>Synechocystis sp.</i>	<i>Synechococcales</i>	98.4	0	SAG 51.79	N.S.
<i>Tolypothrix sp.</i>	<i>Nostocales</i>	97.126	++	N.S.	N.S.
<i>Tolypothrix sp.</i>	<i>Nostocales</i>	97.100b	0	Österreich	1977
<i>Tolypothrix byssoidea</i>	<i>Nostocales</i>	97.10	+	Kanada	N.S.
<i>Tolypothrix byssoidea</i>	<i>Nostocales</i>	97.81	+	Nepal	1967
<i>Tolypothrix distorta</i>	<i>Nostocales</i>	97.57	0	Österreich	1968
<i>Tolypothrix distorta</i>	<i>Nostocales</i>	97.18	+	Schweiz	1963
<i>Tolypothrix distorta</i>	<i>Nostocales</i>	97.17	+	Schweiz	1963
<i>Tolypothrix distorta</i>	<i>Nostocales</i>	97.80	+	Schweiz	1970
<i>Tolypothrix elenkinii</i>	<i>Nostocales</i>	97.23	-	Schweiz	1964
<i>Tolypothrix elenkinii</i>	<i>Nostocales</i>	97.86	-	Schweiz	1975
<i>Tolypothrix sp.</i>	<i>Nostocales</i>	97.26	+	Schweiz	1964
<i>Xenococcus sp.</i>	<i>Pleurocapsales</i>	97.118	0	Schweiz	1991

Die 169 Cyanobakterien aus der Stammsammlung sind zusammen mit Stammnummer (BB), Herkunftsland und Zeitpunkt der Aufnahme angegeben. Dabei sind nicht immer alle Daten spezifiziert (N.S.). Das Screening unterteilt die Cyanobakterien in PCR-unzugänglich (0), PCR-zugänglich ohne RT-Bande (-), PCR-zugänglich mit RT-Bande der theoretisch korrekten Größe (+) und PCR-zugänglich mit zu großer RT-Bande (++) . Sequenzierte Banden sind farblich hervorgehoben: Homologie zu RTs (grün) und keine Homologie zu RTs (rot).

Tabelle A2: Patientenproben aus subgingivalem Plaque

Patient	Probe	Zeitpunkt
1	C	t1
2	A	t1, t2
	B	t1, t2 t4
3	C	t1, t4
	A	t1, t4
4	B	t1, t3
	C	t1
5	A	t3
	B	t1
6	C	t1
	B	t4
7	C	t4
	A	t1
8	C	t1, t2, t3, t4
	A	t1
9	C	t1
	A	t1, t3
10	C	t2
	C	t1, t4
11	A	t1, t2
	B	t1, t3
12	C	t1, t4
	A	t3
	B	t2
13	C	t1, t2
	A	t1, t2, t3, t4
14	B	t1, t3
	A	t1, t3
15	C	t1, t2
	A	t1, t4
	C	t1

Die Proben stammen aus der Studie von Dr. Sigrun Eick (Eick et al. 2011). Den Patienten wurden jeweils drei Proben (A-C) des subgingivalen Plaque zu Beginn (t1), sowie 3, 6 und 12 Monate nach Ansetzen einer Behandlung (t2-4) entnommen. Die Proben waren pseudonymisiert, ein Ethikvotum der Universität Jena lag vor.

Tabelle A3. SHAPE-Daten von der Templat RNA

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
3	0,0636	0,0467	0,1256	0,0786	0,03
4	0,0089	0,0247	-0,0222	0,0038	0,02
5	-0,0246	0,0067	-0,0174	-0,0118	0,01
6	-0,0178	0,0073	-0,0172	-0,0092	0,01
7	-0,0442	-0,0015	-0,0380	-0,0279	0,02
8	-0,0301	0,0052	-0,0104	-0,0118	0,01
9	-0,0207	-0,0015	-0,0001	-0,0074	0,01
10	-0,0082	0,0094	0,0044	0,0019	0,01
11	0,0488	0,0872	0,0572	0,0644	0,02
12	0,0158	0,0392	0,0101	0,0217	0,01
13	0,0365	0,0745	0,0172	0,0427	0,02
14	-0,0035	-0,0025	-0,0340	-0,0134	0,01
15	-0,0101	-0,0027	-0,0302	-0,0143	0,01
16	-0,0011	-0,0028	-0,0033	-0,0024	0,00
17	0,0944	0,0879	0,1252	0,1025	0,02
18	0,3768	0,3761	0,5079	0,4203	0,06
19	0,6463	0,6130	0,5394	0,5996	0,04
20	0,1742	0,2781	0,1702	0,2075	0,05
21	0,8838	0,8491	1,6932	1,1421	0,39
22	1,7399	1,6939	1,8248	1,7529	0,05
23	1,4590	1,5743	2,1262	1,7198	0,29
24	0,0306	0,1066	0,0706	0,0693	0,03
25	-0,0013	0,0108	-0,0141	-0,0015	0,01
26	-0,0131	-0,0009	-0,0076	-0,0072	0,00
27	-0,0174	-0,0013	0,0003	-0,0061	0,01
28	0,0139	0,0306	0,0257	0,0234	0,01
29	0,0334	0,0229	0,0247	0,0270	0,00
30	0,0172	-0,0042	0,0347	0,0159	0,02
31	0,0246	0,0377	-0,0076	0,0182	0,02
32	0,0219	0,0394	0,0163	0,0259	0,01
33	0,0535	0,1084	-0,0094	0,0508	0,05
34	0,0527	0,1024	-0,0122	0,0476	0,05
35	0,1340	0,1887	0,0937	0,1388	0,04
36	0,2995	0,5324	0,4319	0,4213	0,10

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
37	0,7598	0,7991	0,4253	0,6614	0,17
38	0,3516	0,5384	0,6056	0,4986	0,11
39	0,2775	0,4336	0,1843	0,2985	0,10
40	-0,0122	0,1001	-0,0397	0,0161	0,06
41	-0,0369	0,0116	-0,0401	-0,0218	0,02
42	-0,0002	0,0349	-0,0249	0,0032	0,02
43	0,0507	0,1252	0,0440	0,0733	0,04
44	0,1498	0,1903	0,0914	0,1438	0,04
45	0,2738	0,3390	0,2729	0,2952	0,03
46	0,3990	0,3310	0,1138	0,2813	0,12
47	0,2905	0,2699	0,2902	0,2835	0,01
48	0,8676	0,8038	0,9681	0,8798	0,07
49	1,3599	1,9059	1,6382	1,6347	0,22
50	0,7953	0,5957	0,6908	0,6939	0,08
51	0,9486	1,0496	0,5483	0,8488	0,22
52	0,1110	0,1138	0,0028	0,0759	0,05
53	-0,0017	0,1055	-0,0325	0,0238	0,06
54	0,2001	0,2883	0,1936	0,2273	0,04
55	0,4507	0,4057	0,4263	0,4275	0,02
56	0,6923	0,4871	0,5388	0,5728	0,09
57	1,3123	1,2967	1,7767	1,4619	0,22
58	0,2882	0,2062	0,3684	0,2876	0,07
59	0,0421	0,1079	0,0663	0,0721	0,03
60	-0,0379	0,0305	0,0379	0,0102	0,03
61	-0,1080	0,0174	0,0549	-0,0119	0,07
62	-0,1770	0,0596	0,0710	-0,0155	0,11
63	-0,1091	0,0528	0,0677	0,0038	0,08
64	-0,3716	0,0327	0,0712	-0,0892	0,20
65	-0,0696	0,0746	0,0717	0,0256	0,07
66	-0,3022	-0,0114	0,1148	-0,0663	0,17
67	-0,7689	0,5084	-0,5004	-0,2537	0,55
68	0,3447	0,4429	0,6229	0,4702	0,12
69	0,2293	0,3638	0,3510	0,3147	0,06
70	0,2448	0,4360	0,5714	0,4174	0,13
71	0,5158	0,8793	1,0270	0,8074	0,21

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
72	0,5873	0,8004	0,8775	0,7551	0,12
73	0,4352	0,6369	-0,0813	0,3303	0,30
74	0,9663	1,6215	-0,1261	0,8205	0,72
75	1,0337	1,1207	2,0000	1,3848	0,44
76	0,3425	0,4284	0,4342	0,4017	0,04
77	0,6160	0,7616	0,9807	0,7861	0,15
78	0,7527	0,5378	0,5938	0,6281	0,09
79	0,3974	0,3609	0,3512	0,3698	0,02
80	0,3720	0,4180	0,5748	0,4549	0,09
81	0,7066	0,7498	0,5919	0,6827	0,07
82	0,2154	0,1809	0,3167	0,2377	0,06
83	0,0674	0,0827	0,1930	0,1143	0,06
84	0,0599	0,0949	0,1577	0,1041	0,04
85	0,1858	0,5125	0,5358	0,4114	0,16
86	0,3904	0,6336	0,8099	0,6113	0,17
87	0,5503	1,0993	1,0193	0,8896	0,24
88	0,6002	0,9007	0,8900	0,7970	0,14
89	1,3973	0,8017	0,7575	0,9855	0,29
90	1,3438	0,8271	0,5666	0,9125	0,32
91	0,4825	0,4152	0,3457	0,4145	0,06
92	0,3389	0,4183	0,4036	0,3869	0,03
93	0,1339	0,2501	0,2029	0,1956	0,05
94	0,0907	0,1622	0,1390	0,1306	0,03
95	0,0659	0,1328	0,1229	0,1072	0,03
96	0,0004	0,0117	0,0348	0,0156	0,01
97	0,0489	0,0912	0,1191	0,0864	0,03
98	0,0546	0,1963	0,1867	0,1459	0,06
99	0,0000	0,0000	0,0000	0,0000	0,00
100	0,0000	0,0000	0,0000	0,0000	0,00
101	0,0000	0,0000	0,0000	0,0000	0,00
102	0,0000	0,0000	0,0000	0,0000	0,00
103	0,0000	0,0000	0,0000	0,0000	0,00
104	0,0000	0,0000	0,0000	0,0000	0,00
105	0,0000	0,0000	0,0000	0,0000	0,00
106	0,0000	0,0000	0,0000	0,0000	0,00

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
107	0,0000	0,0000	0,0000	0,0000	0,00
108	0,0000	0,0000	0,0000	0,0000	0,00
109	0,1149	0,1376	0,1936	0,1487	0,03
110	0,3503	0,6461	0,4499	0,4821	0,12
111	0,2124	0,3188	0,2240	0,2517	0,05
112	0,1405	0,3464	0,2432	0,2434	0,08
113	0,6536	1,0735	0,6539	0,7937	0,20
114	0,4183	0,5973	0,4208	0,4788	0,08
115	0,3590	0,6345	0,5354	0,5097	0,11
116	0,4953	1,1527	0,6824	0,7768	0,28
117	0,7210	0,8994	0,8544	0,8249	0,08
118	0,5200	0,6974	0,5339	0,5838	0,08
119	0,8536	1,2881	1,0459	1,0626	0,18
120	0,5188	0,8385	0,6925	0,6833	0,13
121	1,2064	3,8444	1,6342	2,2283	1,16
122	1,5312	1,3353	1,0998	1,3221	0,18
123	0,6715	0,5487	0,6360	0,6187	0,05
124	0,3402	0,2984	0,5251	0,3879	0,10
125	-0,0016	0,0814	0,1977	0,0925	0,08
126	-0,0023	0,0447	0,1810	0,0745	0,08
127	-0,0110	0,0082	0,2481	0,0817	0,12
128	-0,0303	0,0050	0,3578	0,1108	0,18
129	0,0000	0,0000	0,0000	0,0000	0,01
130	0,0000	0,0000	0,0000	0,0000	0,00
131	0,0000	0,0000	0,0000	0,0000	0,00
132	0,0000	0,0000	0,0000	0,0000	0,00
133	0,0000	0,0000	0,0000	0,0000	0,00
134	0,0000	0,0000	0,0000	0,0000	0,00
135	0,0024	-0,0057	0,0510	0,0159	0,03
136	-0,0296	-0,0432	-0,0068	-0,0265	0,02
137	-0,0076	-0,0245	0,0787	0,0156	0,05
138	0,0069	0,0057	0,0990	0,0372	0,04
139	-0,0158	-0,0605	0,1625	0,0287	0,10
140	0,0339	0,0040	0,2309	0,0896	0,10
141	0,0594	0,0613	0,2224	0,1144	0,08

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
142	0,0687	0,0957	0,2456	0,1366	0,08
143	0,1321	0,1794	0,3103	0,2073	0,08
144	0,0677	0,0863	0,2542	0,1361	0,08
145	0,3065	0,3644	0,4580	0,3763	0,06
146	0,1284	0,1546	0,2532	0,1788	0,05
147	0,1367	0,1642	0,2465	0,1825	0,05
148	0,0915	0,0822	0,2197	0,1311	0,06
149	0,3506	0,4283	0,5096	0,4295	0,06
150	0,8337	0,8742	0,8622	0,8567	0,02
151	0,3183	0,3341	0,5314	0,3946	0,10
152	0,7486	0,6349	0,6568	0,6801	0,05
153	0,7196	0,5978	0,7904	0,7026	0,08
154	1,4741	1,4921	1,3769	1,4477	0,05
155	1,3934	2,0510	1,3777	1,6074	0,31
156	0,2446	0,1864	0,1471	0,1927	0,04
157	0,0850	0,1059	0,1538	0,1149	0,03
158	0,0598	0,0699	0,1141	0,0813	0,02
159	0,2067	0,2004	0,2325	0,2132	0,01
160	0,5486	0,4205	0,5687	0,5126	0,07
161	0,3123	0,2351	0,3210	0,2895	0,04
162	0,0860	0,1287	0,0833	0,0993	0,02
163	0,0135	0,0202	0,0581	0,0306	0,02
164	0,0166	-0,0147	0,0303	0,0107	0,02
165	0,0283	0,0087	0,0807	0,0392	0,03
166	0,0259	0,0152	0,0533	0,0315	0,02
167	0,0929	0,0817	0,1697	0,1148	0,04
168	0,0622	0,0555	0,1097	0,0758	0,02
169	0,4580	0,2841	0,4029	0,3817	0,07
170	0,7217	0,8936	0,8520	0,8224	0,07
171	0,0530	0,1170	0,1197	0,0966	0,03
172	0,0329	0,0179	0,0054	0,0187	0,01
173	0,0804	0,0692	0,0459	0,0652	0,01
174	0,1592	0,1490	0,1695	0,1593	0,01
175	0,3696	0,4032	0,4234	0,3987	0,02
176	1,1489	1,2821	1,1113	1,1807	0,07

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
177	0,2668	0,3212	0,3051	0,2977	0,02
178	0,0688	0,0657	0,0610	0,0652	0,00
179	0,0545	0,0705	0,0494	0,0581	0,01
180	0,0603	0,0418	0,0963	0,0661	0,02
181	0,3720	0,2476	0,2730	0,2976	0,05
182	0,4152	0,3850	0,4054	0,4019	0,01
183	0,4837	0,4784	0,5056	0,4892	0,01
184	0,9466	1,5128	0,8188	1,0927	0,30
185	0,0743	0,1036	0,0291	0,0690	0,03
186	0,0053	-0,0173	-0,0409	-0,0177	0,02
187	-0,0010	-0,0170	-0,0164	-0,0114	0,01
188	-0,0100	-0,0304	-0,0312	-0,0239	0,01
189	-0,0192	-0,0191	-0,0210	-0,0197	0,00
190	-0,0262	-0,0036	-0,0110	-0,0136	0,01
191	-0,0096	0,0360	0,0430	0,0232	0,02
192	0,0384	0,0718	0,0349	0,0484	0,02
193	0,1314	0,1109	0,1216	0,1213	0,01
194	0,2494	0,2891	0,2402	0,2595	0,02
195	0,1128	0,0674	0,0936	0,0913	0,02
196	0,1156	0,1233	0,0660	0,1017	0,03
197	0,3369	0,4049	0,2188	0,3202	0,08
198	1,3863	1,0796	1,0759	1,1806	0,15
199	0,4415	0,7994	0,4680	0,5696	0,16
200	1,5446	0,6943	0,5138	0,9176	0,45
201	0,6075	1,0903	1,7682	1,1553	0,48
202	-0,0210	0,0703	0,0359	0,0284	0,04
203	-0,0258	-0,0275	0,0490	-0,0014	0,04
204	-0,0041	-0,0058	0,0126	0,0009	0,01
205	-0,0011	-0,0029	0,0180	0,0047	0,01
206	-0,0013	-0,0015	-0,0055	-0,0028	0,00
207	-0,0020	-0,0012	-0,0027	-0,0020	0,00
208	0,0047	0,0011	-0,0167	-0,0036	0,01
209	0,0414	0,0366	0,0560	0,0446	0,01
210	0,1313	0,1622	0,0318	0,1085	0,06
211	0,5919	0,5827	0,5576	0,5774	0,01

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
212	0,8617	0,9204	1,8678	1,2166	0,46
213	0,4784	0,4435	0,4486	0,4568	0,02
214	0,6121	0,5664	0,5425	0,5737	0,03
215	0,1351	0,1145	0,0921	0,1139	0,02
216	0,0818	0,0809	0,1008	0,0879	0,01
217	-0,0019	0,0014	0,0135	0,0043	0,01
218	-0,0029	-0,0053	0,0263	0,0060	0,01
219	-0,0119	0,0153	0,0336	0,0123	0,02
220	0,0063	0,0046	-0,0153	-0,0015	0,01
221	-0,0019	0,0048	-0,0149	-0,0040	0,01
222	0,0217	0,0241	0,0372	0,0277	0,01
223	0,3348	0,2881	0,2579	0,2936	0,03
224	0,4092	0,4513	0,4384	0,4330	0,02
225	1,2040	1,4858	1,5508	1,4135	0,15
226	0,6672	0,6074	0,6437	0,6394	0,02
227	1,3671	1,4199	1,4657	1,4176	0,04
228	0,1203	0,0905	0,1178	0,1095	0,01
229	0,1295	0,1671	0,0778	0,1248	0,04
230	0,7960	0,7395	0,7142	0,7499	0,03
231	0,5125	0,4596	0,4932	0,4884	0,02
232	0,2076	0,1933	0,1693	0,1901	0,02
233	0,2178	0,2059	0,2206	0,2148	0,01
234	0,0363	0,0286	0,0156	0,0268	0,01
235	0,0457	0,0736	0,0582	0,0592	0,01
236	0,0464	0,0640	0,0090	0,0398	0,02
237	0,0119	0,0202	-0,0313	0,0002	0,02
238	0,0607	0,0642	0,2222	0,1157	0,08
239	0,0621	0,1206	0,1309	0,1045	0,03
240	0,2813	0,3102	0,4343	0,3419	0,07
241	0,3469	0,3605	0,7494	0,4856	0,19
242	0,9832	0,8671	0,8010	0,8838	0,08
243	1,0920	1,3626	1,4036	1,2861	0,14
244	0,2761	0,2148	0,1621	0,2177	0,05
245	0,0363	0,0445	0,0167	0,0325	0,01
246	0,0005	0,0403	-0,1094	-0,0229	0,06

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
247	0,1045	0,1281	0,0621	0,0983	0,03
248	0,2332	0,1621	0,3158	0,2370	0,06
249	0,6212	0,4091	0,3302	0,4535	0,12
250	1,2752	1,0542	1,2556	1,1950	0,10
251	0,0898	0,0518	0,0004	0,0473	0,04
252	0,0119	0,0358	-0,2739	-0,0754	0,14
253	0,0808	0,1337	-0,1463	0,0227	0,12
254	0,3568	0,5211	0,3665	0,4148	0,08
255	2,1203	1,7996	2,1631	2,0277	0,16
256	0,4420	0,3905	0,3628	0,3984	0,03
257	0,0762	0,1452	0,1019	0,1078	0,03
258	0,0183	0,1367	0,0514	0,0688	0,05
259	-0,0486	-0,0337	-0,0831	-0,0551	0,02
260	0,0644	0,0580	0,0703	0,0642	0,01
261	-0,0234	-0,0108	-0,0820	-0,0387	0,03
262	-0,0132	-0,0420	-0,0430	-0,0327	0,01
263	0,0164	-0,0247	-0,1003	-0,0362	0,05
264	0,1279	0,0765	0,0877	0,0974	0,02
265	0,1417	0,0885	0,0629	0,0977	0,03
266	0,1862	-0,0097	0,1181	0,0982	0,08
267	0,0385	-0,0076	-0,0855	-0,0182	0,05
268	0,0845	0,0012	-0,0419	0,0146	0,05
269	0,1362	0,0263	0,0349	0,0658	0,05
270	0,0711	-0,0200	-0,0202	0,0103	0,04
271	0,0315	0,0386	0,0137	0,0279	0,01
272	0,0210	0,0338	0,0347	0,0298	0,01
273	0,0241	0,0272	0,0321	0,0278	0,00
274	0,0936	0,1097	0,0913	0,0982	0,01
275	0,1361	0,1645	0,0854	0,1287	0,03
276	0,2988	0,3573	0,1633	0,2731	0,08
277	0,6870	0,7189	0,3883	0,5981	0,15
278	0,9400	1,0232	0,7029	0,8887	0,14
279	0,8540	0,9768	0,8784	0,9031	0,05
280	0,6347	0,8612	0,7263	0,7407	0,09
281	1,0600	0,9349	1,1216	1,0389	0,08

Nukleotid	SHAPE 1	SHAPE 2	SHAPE 3	Mittelwert	STABWN
282	0,9362	0,7589	0,7673	0,8208	0,08
283	0,8696	0,9452	0,8686	0,8945	0,04
284	0,8306	0,5929	0,4437	0,6224	0,16
285	0,4849	0,3062	0,2530	0,3481	0,10
286	0,3887	0,3101	0,2351	0,3113	0,06
287	0,0776	0,0888	0,0153	0,0606	0,03
288	0,0578	0,0476	0,0321	0,0458	0,01
289	0,0095	-0,0034	-0,0649	-0,0196	0,03
290	0,0318	0,0214	0,0223	0,0252	0,00
291	0,0108	0,0335	-0,0376	0,0022	0,03
292	0,0324	0,0047	-0,0279	0,0030	0,02
293	0,0514	0,0237	0,0157	0,0303	0,02
294	0,0290	0,0057	-0,0643	-0,0099	0,04
295	0,1006	0,0369	-0,0679	0,0232	0,07
296	0,0776	0,0167	-0,0201	0,0247	0,04
297	0,0629	-0,0041	-0,0589	-0,0001	0,05
298	0,0766	0,0332	-0,0130	0,0322	0,04
299	0,5115	0,3831	0,3662	0,4203	0,06
300	1,2737	1,3370	1,2344	1,2817	0,04
301	0,5839	0,4295	0,4679	0,4938	0,07
302	0,9947	2,1851	0,9800	1,3866	0,56
303	0,4716	0,4364	0,3698	0,4259	0,04
304	0,0522	0,0859	0,0468	0,0617	0,02
305	-0,0024	-0,0153	0,0100	-0,0026	0,01
306	0,0523	-0,0122	-0,0218	0,0061	0,03
307	0,1643	0,1359	0,0964	0,1322	0,03
308	0,0059	0,0307	0,0017	0,0128	0,01
309	1,2274	1,8249	1,6658	1,5727	0,25
310	0,3714	0,5683	0,6725	0,5374	0,12
311	0,2643	0,4242	0,3422	0,3436	0,07
312	-0,1724	-0,0841	0,0890	-0,0558	0,11
313	-0,2571	-0,0626	0,0645	-0,0851	0,13
314	-0,3335	0,0099	-0,2766	-0,2001	0,15
315	-0,2322	0,0894	0,0875	-0,0184	0,15

Die aus den gemessenen NMIA-Reaktivitäten berechneten SHAPE-Daten sind für drei unabhängige Versuche angegeben. Die Aufteilung der Reaktivitäten entspricht der von RNAstructure: Nicht reaktiv ($< 0,4$ = schwarz), moderat reaktiv ($0,4-0,85$ = gelb) und stark reaktiv ($> 0,85$ = rot). Die Standardabweichung (STABWN) aus den drei Versuchen ist angegeben und bei einem Wert $\geq 0,3$ rot markiert. Das betrifft einige der stark reaktiven Nukleotide und die Nukleotide 67 und 73-74. Nicht analysierbare Nukleotide sind blau markiert (1-2, 67, 73-74, 99-108, 129-134 und 316-340). Diese wurden mit dem Wert „-999“ belegt und zusammen mit den anderen Mittelwerten in RNAstructure integriert.

Tabelle A4. Verwendete Oligonukleotide

Verwendung		Sequenz (5' -> 3')	Bezeichnung
Cyanobakterien-Screening	DGR RT	CARTGYGAYATYARRAARTATTTYCC AATARNGCRAAGTCATCRACRTARC	NZ112 NZ114
	16S rDNA	GAGCTACGCCTAGTATCCATCG GGAAGGCTCTTGGGTTGTAAACC	PM24 PM57
Inverse PCR	<i>Anabaena</i> DGR	GGTTTACCCATTGGGAATCTCACC CAGAGTATCTTGACATTTAAGTTTACGC	PM21 PM22
		cDNA-Synthese	TAGACTTAGAAATTAATACGACTCACTATAGGCGCGCCACCGGTG-D(T)18
3'-RACE	RT-PCR	TAGACTTAGAAATTAATACGACTCACTATAGGCGCGCCACCG CGAGAGTCGGCAGGTGGGAATCTGTGCGAG	PM104 PM121
	cDNA-Synthese	GGAACGGCAGTTTTTCAGGATTG	PM137
5'-RACE	RT-PCR	GCTATTCTCGAGGAGCGGCCGCCTGCAGGAAA GGTCTTGAGCTCCGGCAGTTTTTCAGGATTGTTGTTCCAGG	PM141 PM142
	Primer A	GCTTTGTAGGGTTGAGCAACTG	PM151
SHAPE (Primer-Extension)	Primer B	AGCTTCAGGTGTAGGCTGGAC	PM135
	Primer C	CGAACGCGCAGACAACACG	PM136
	Primer D	GGAACGGCAGTTTTTCAGGATTG	PM137
Vektor-Sequenzierung	T7-Promotor	TAATACGACTCACTATAGGG	TS3
	T3-Promotor	ATTAACCCTCACTAAAGGGA	TS43
Aktivitätsmessung (VR)	<i>Anabaena</i> (1)	GAACGACGACACCTTTTCAC CGTTGCTGAATTGAGGTATG	PM58 PM59
		<i>Anabaena</i> (2)	CTATCACAGGTAAGTTAGC

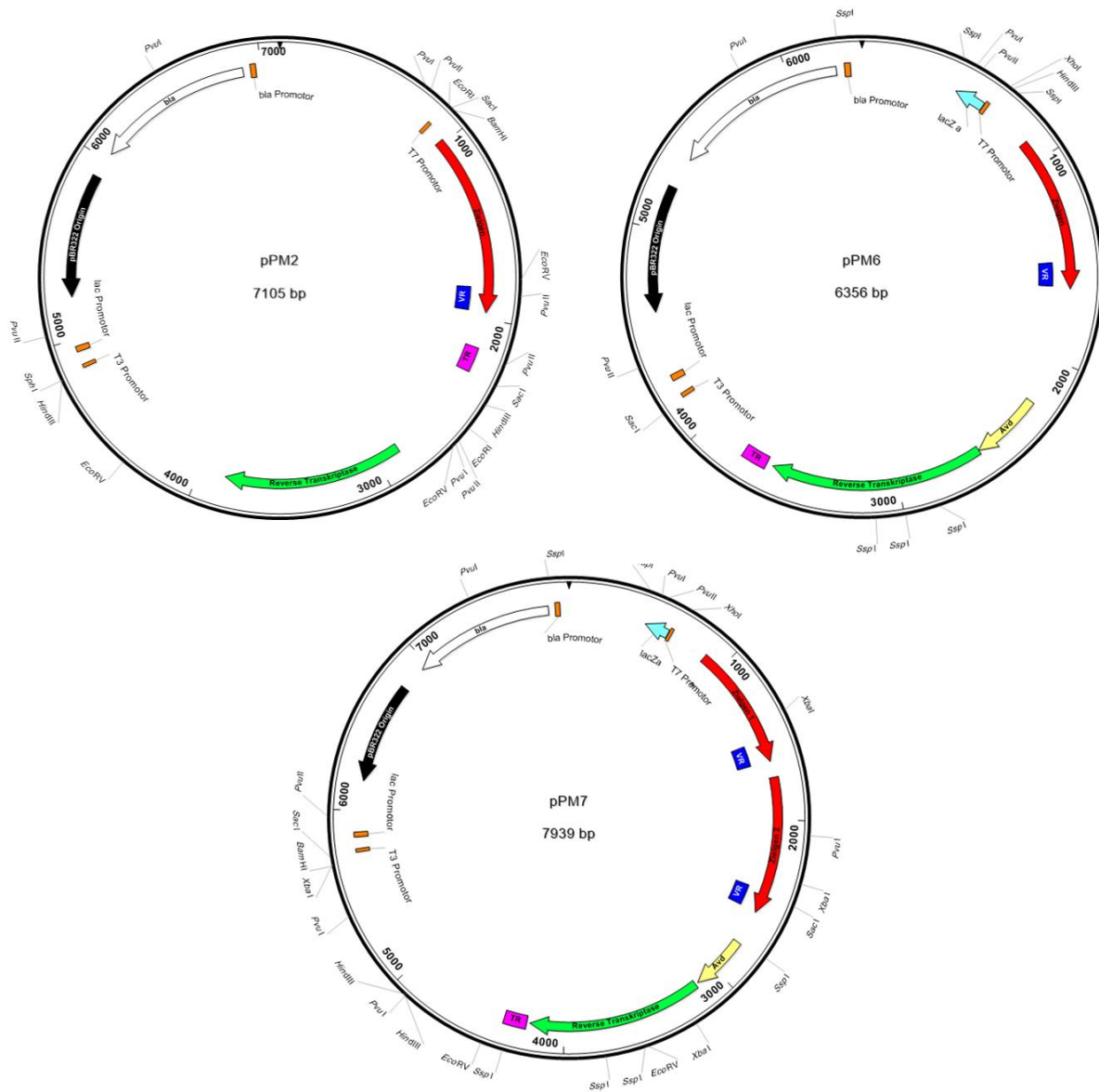
Verwendung	Sequenz (5' -> 3')	Bezeichnung	
Aktivitätsmessung (VR)	<i>Nostoc</i> (1)	CTCTTACTCTCTGCGAC	PM49
		CACCAAAAACAGAGGAAGCC	PM73
		TGGTGGTGGTGCTCGAGTTTTTCTCTGTTTCTTTGATTTTTAATCC	TS69
	<i>Nostoc</i> (2)	GTAGGGAGCTTTGGCGTAGC	PM125
		GCTCCACTCCGTTACGCTAAG	PM82
	<i>Treponema</i> (1)	ATGGTGCTGGGATTGGC	PM5
		GATACGGTACTTACATAAGGAG	PM25
	<i>Treponema</i> (2)	GTGCCGATATCCATACC	PM55
		CCTTATACGGTCCATACC	PM56
	<i>Treponema</i> 16S rDNA (Ashimoto et al. 1996)	TAATACCGAATGTGCTCATTTACAT	PM63
TCAAAGAAGCATTCCCTCTTCTTCTTA		PM64	
Erzeugung von Plasmiden	pKH1	GCTACGGAATTCTAATACGACTCACTATAGGTATTTACTTCCCTCC	PM147
		TGCACGAAGCTTCAGGTGTAGGCTGGACTC	PM123
	pKH2	GCTACGGAATTCTAATACGACTCACTATAGGTATTTACTTCCCTCC	PM147
		CGTAGCAAGCTTTGTAGGGTTGAGCAACTG	PM124
	pPM2	CGCACCGGATCCGCAGGGTTGAATTTATTCGGCTTTGC	SF5
		CGCGCGGCATGCGGCATTTTAAAAGGCCGTGTTTAATTTAACC	SF6
	pPM6	GCATCTCTCGAGCAGAAGTACAAGCTTTTGCTAAGG	PM41
		GTCACTGAGCTCGTCAGAGACTGGGAATGAGACG	PM46
	pPM7	ATCGCTCTCGAGGATGATGAGGAAAAGGCAATTGCACAAG	PM60
		CGTATCGGATCCCCAATCCTGAAATTTGTCGAGCCAATGG	PM61
pPM10	GCTACGGAATTCTAATACGACTCACTATAGGAATCAAGTAACTCAATC	PM122	

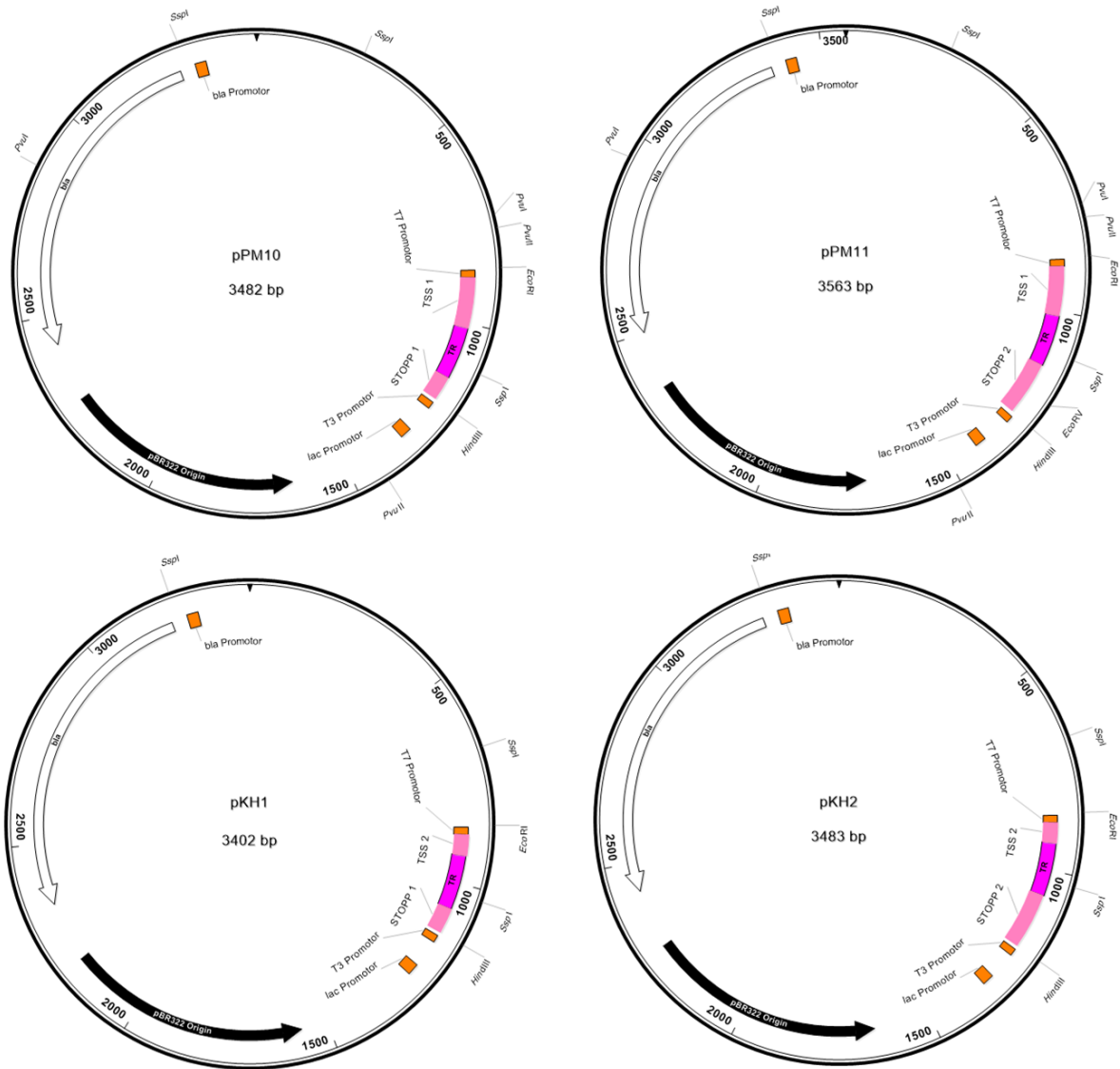
Verwendung	Sequenz (5' -> 3')	Bezeichnung
	TGCACGAAGCTTCAGGTGTAGGCTGGACTC	PM123
pPM11	GCTACGGAATTCTAATACGACTCACTATAGGAATCAAGTAACTCAATC	PM122
	CGTAGCAAGCTTTGTAGGGTTGAGCAACTG	PM124
	CACCATGCACTATGCAACATC	PM75
RT (<i>Nostoc</i>)	CATCGGCTAATAATTCTCGGTC	PM76
	GGTGTAGGCTGGACTCTTCAG	PM77
TR (<i>Nostoc</i>)	GTA ACTCAATCTATGAGAAGTTGG	PM109
	AAGAAGGAGATATACATATGAAAGAATTATCGGTCATCCAAAAG	TS68
Avd (<i>Nostoc</i>)	TGGTGGTGGTGCTCGAGTTTTTCTCTGTTTCTTTGATTTTTAATCC	TS69
RT-PCR	GGTTTACCCATTGGGAATCTCACC	PM21
RT (<i>Anabaena</i>)	CGAGATACGTCTTGATC	PM47
	GGTGGTTCTTGGAACAACAACC	PM67
TR (<i>Anabaena</i>)	CTGATAGTTGAAGAGTGCTTGC	PM68
	CAACCGGGTTTGATGACG	PM48
Avd (<i>Anabaena</i>)	GGTATGTGCCTATTTTAGACAAG	PM107
	CTGTTTGCGTAATCGCCAAGTATCACCATGTTCTAGGT	PM146
Northern-Blot-Experimente	GGTCTTGAGCTCCGGCAGTTTTTCAGGATTGTTGTTCCAGG	PM142
	GTTCCGGCTTTCGCCTACCAAGCTACGCAACCCGGTTTTATATTTGG	PM139
	GCTACGGAATTCTAATACGACTCACTATAGGAATCAAGTAACTCAATC	PM122
Sonde 4	CGTAGCAAGCTTTGTAGGGTTGAGCAACTG	PM124
Primer 1	CAGACTATGCTCAAGGTAACTTG	PM108

Verwendung		Sequenz (5' -> 3')	Bezeichnung
Verifizierung der Transkriptlänge des TR (RT-PCR)	Primer 2	GTA ACTCAATCTATGAGAAGTTGG	PM109
	Primer 3	GATACTTGGCGATTACGCAAAC	PM144
	Primer 4	CTTCCCTCCATTTTAGGAGAAAG	PM145
	Primer 5	CTGGAACAACAATCCTGAAAAC	PM69
	Primer 6	GGTGTAGGCTGGACTCTTCAG	PM77
	Primer 7	GTTCGGCTTTCGCCTACCAAGCTACGCAACCCGGTTTTATATTTGG	PM139
	Primer 8	CCTGAAATGGTCAAGACAGAG	PM79

Das Kürzel in der Bezeichnung gibt die Person an, die das Oligonukleotid entworfen hat: PM (Philipp Möller), TS (Dr. Thomas Schillinger), NZ (Dr. Nora Zingler) und SF (Dr. Sabine Filker). Die Oligonukleotide wurden von der Firma Eurofins MWG Operon (Ebersberg) bestellt.

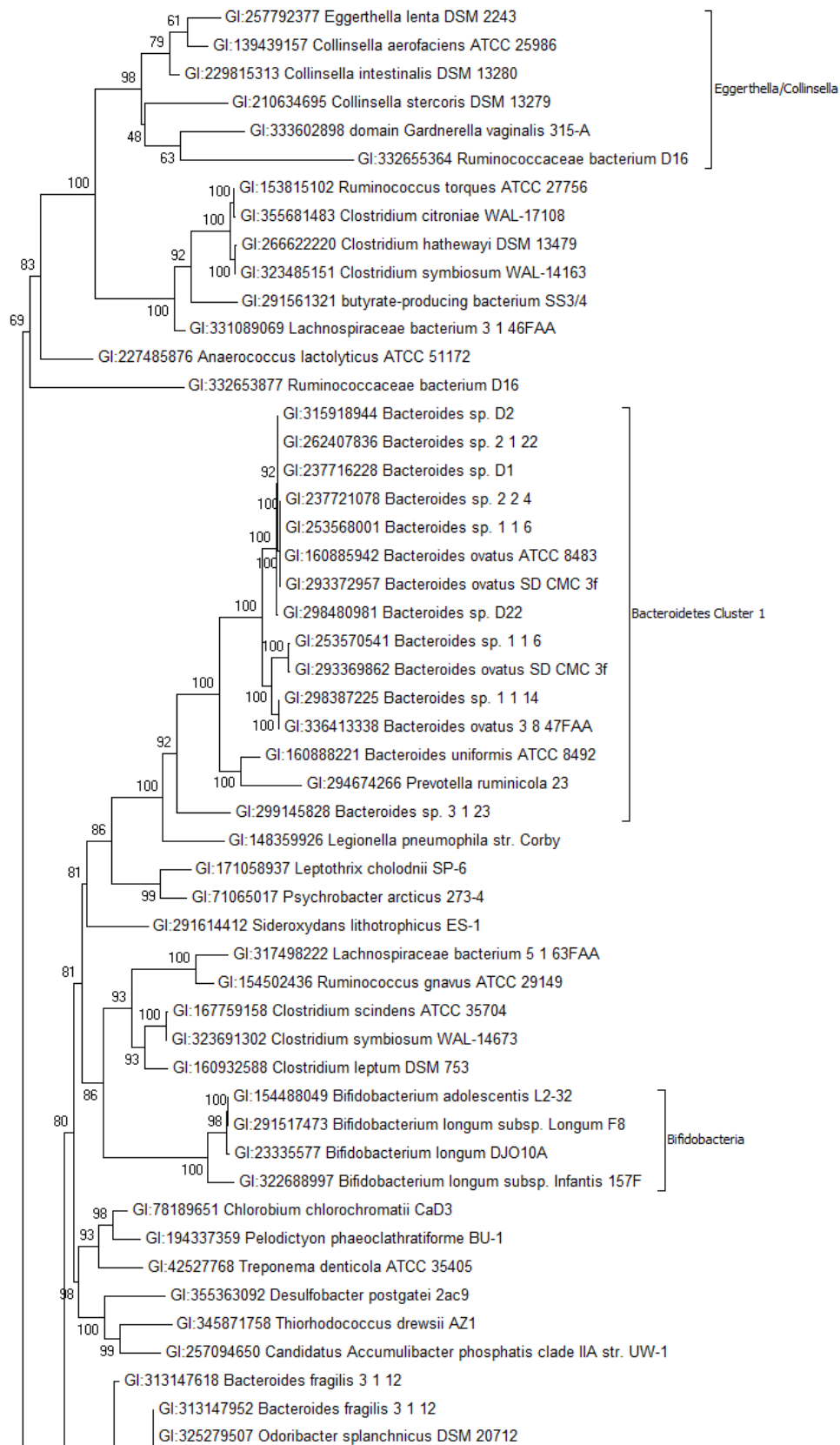
Abbildung A1: Vektorkarten

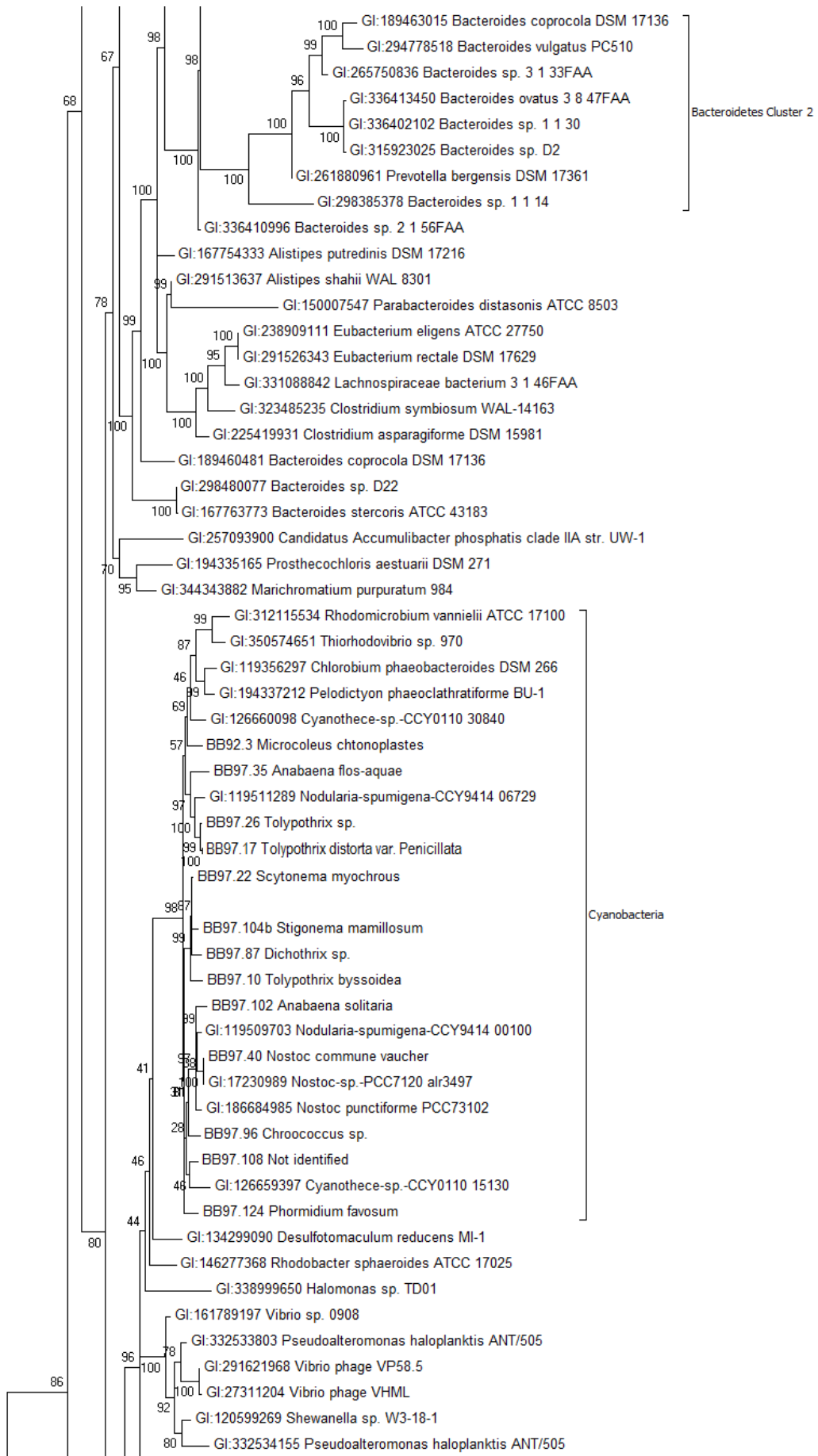


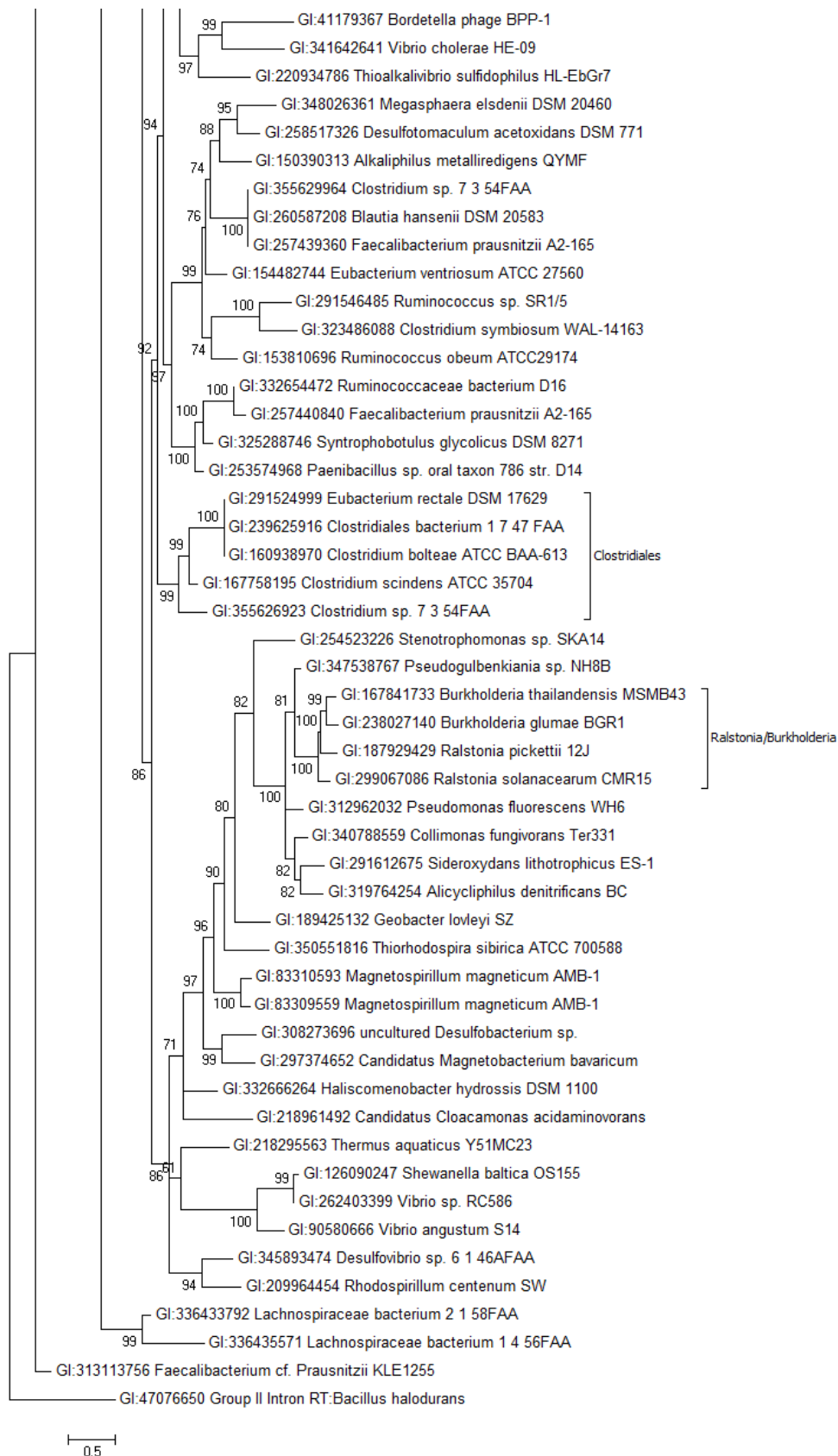


Die in der Arbeit verwendeten Plasmide sind abgebildet. Die Plasmide pPM2 (*Treponema denticola*), pPM6 (*Anabaena flos-aquae*) und pPM7 (*Nostoc sp.* PCC7120) tragen die kompletten DGRs der jeweiligen Organismen. Die Plasmide pPM10 (TSS1-STOPP1), pPM11 (TSS1-STOPP2), sowie pKH1 (TSS2-STOPP1) und pKH2 (TSS2-STOPP2) wurden für die IVT der Templat RNA aus *Nostoc sp.* PCC7120 verwendet. Als Vektor diente der pBluescript (pPM2, pPM6 und pPM7) bzw. der pBluescript Δ T7 (pPM10, pPM11, pKH1 und pKH2). Die genomische DNA von *Treponema denticola*, die für die Erzeugung des pPM2 verwendet wurde, stammt von der Deutschen Sammlung von Mikroorganismen und Zellkulturen (#14222). Farbgebung: Zielgen (rot), VR (blau), TR (lila), Avd (gelb), RT (grün), Resistenzgen (weiss), Replikationsursprung (schwarz) und Promotoren (orange).

Abbildung A2: Phylogenetischer Baum der DGR-kodierten RTs

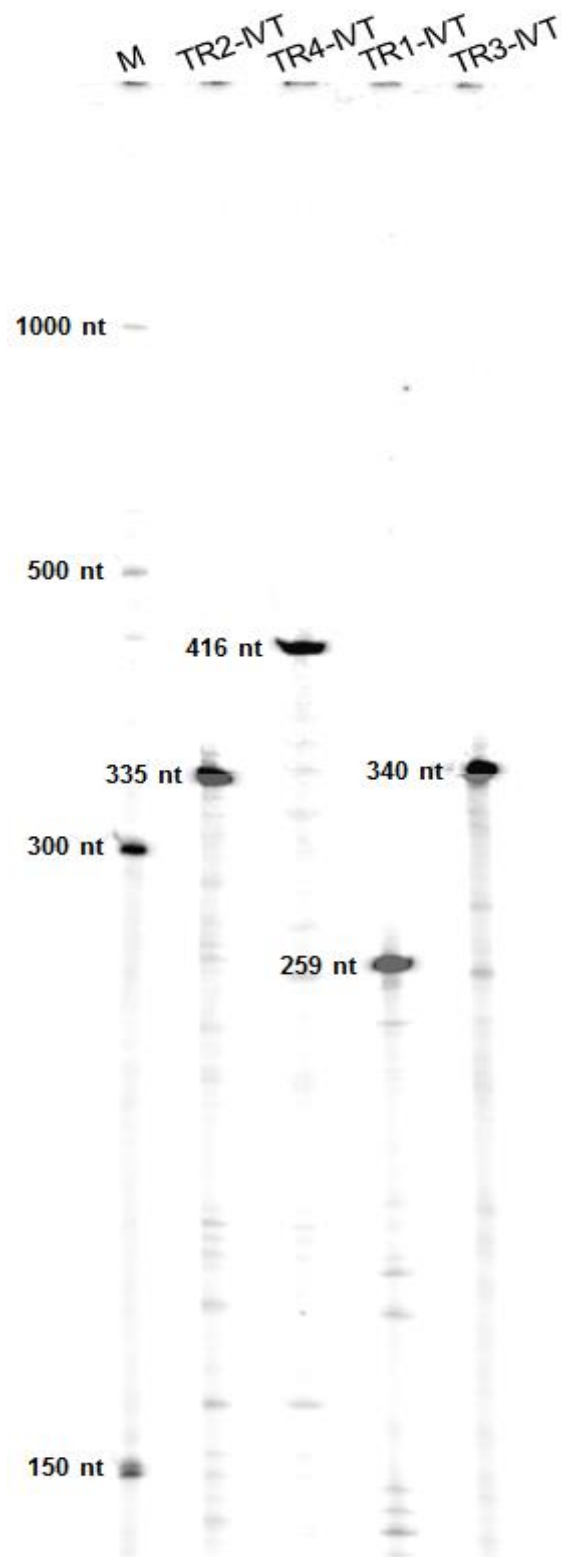






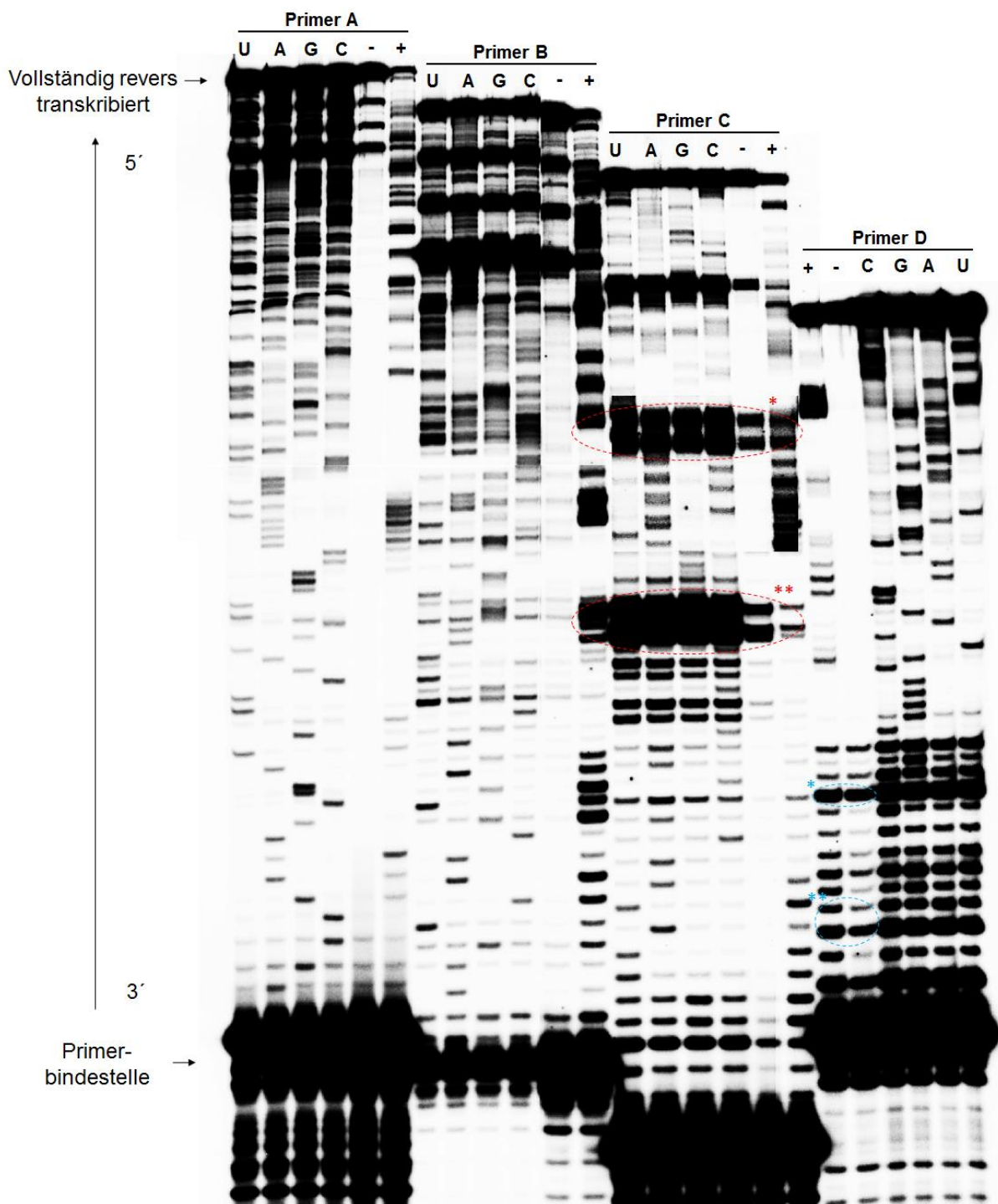
Die Proteinsequenzen der RTs wurden mit ClustalW2 aligned und ein phylogenetischer Baum über Maximum-Likelihood generiert. Die Bootstrap-Werte von 2000 Replikaten sind angezeigt. Verwandte Organismen, die über RT-Sequenzen gruppiert werden, sind markiert. Als Wurzel des Baums wurde die RT von *Bacillus halodurans* (GI 47076650) verwendet.

Abbildung A3: *In vitro* Transkription vier möglicher Templat RNAs aus *Nostoc sp.* PCC7120



Es wurden vier mögliche Transkripte für die Templat RNA identifiziert (TR1-4). Diese wurden *in vitro* transkribiert und zusammen mit einem ssRNA Größenmarker (M) radioaktiv markiert. In der Abbildung sind die *in vitro* Transkripte auf einem Polyacrylamidgel aufgetrennt. Es ist zu erkennen, dass die RNA ohne größere Abbrüche bis zum Ende transkribiert wurde.

Abbildung A4: SHAPE-Experiment zur Strukturaufklärung der Templat RNA

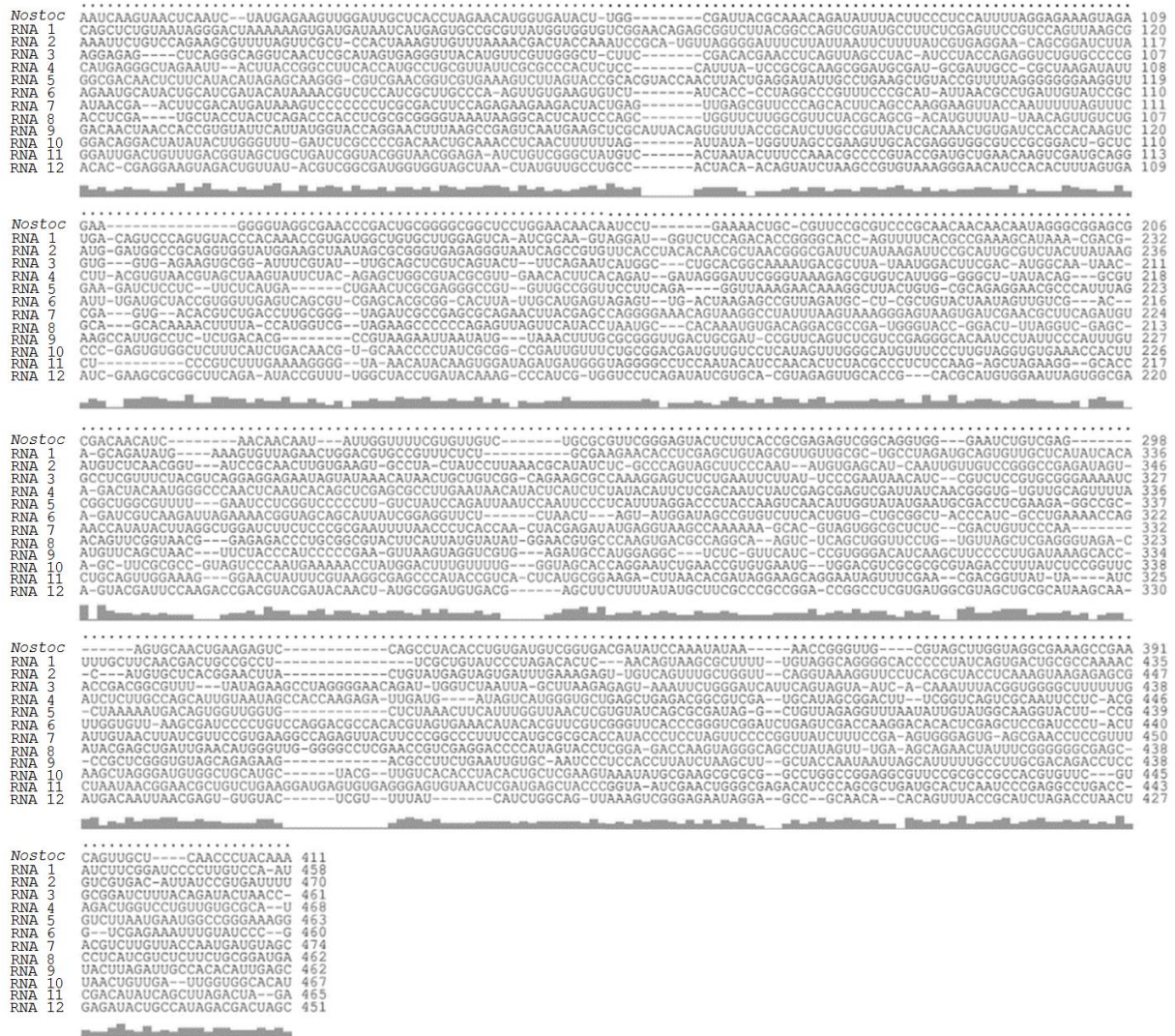


Die TR3-IVT (340 nt) wurde in einem SHAPE-Experiment untersucht. Dabei bindet Primer A am 3'-Ende und Primer B-D folgen im Abstand von ca. 80 nt. Die Sequenzierungsreaktionen (U, A, G, C) wurden mit den Dideoxynukleotiden (ddATP, ddTTP, ddCTP, ddGTP) für einen Kettenabbruch durchgeführt. NMIA(-) unterscheidet sich von NMIA(+) nur durch das Fehlen der Chemikalie. Die Chemikalie modifiziert flexible Positionen in der RNA, wodurch die RT bei der Primer-Extension vom Templat abdissoziiert. Die cDNA wurde auf einem 15 % Polyacrylamidgel aufgetrennt und analysiert. Nicht analysierbare Nukleotide befinden sich

direkt an der Primerbindestelle und der vollständig revers transkribierten cDNA. Auch konnten Nukleotide in komprimierten Banden (Primer C, rot markiert, 97-106 (*) und 127-132 (**)) nicht ausgelesen werden. Nicht gewertet wurden Nukleotide mit abweichenden NMIA-Reaktivitäten zwischen den Experimenten (Primer D, blau markiert, 65 (*) und 71-72 (**)).

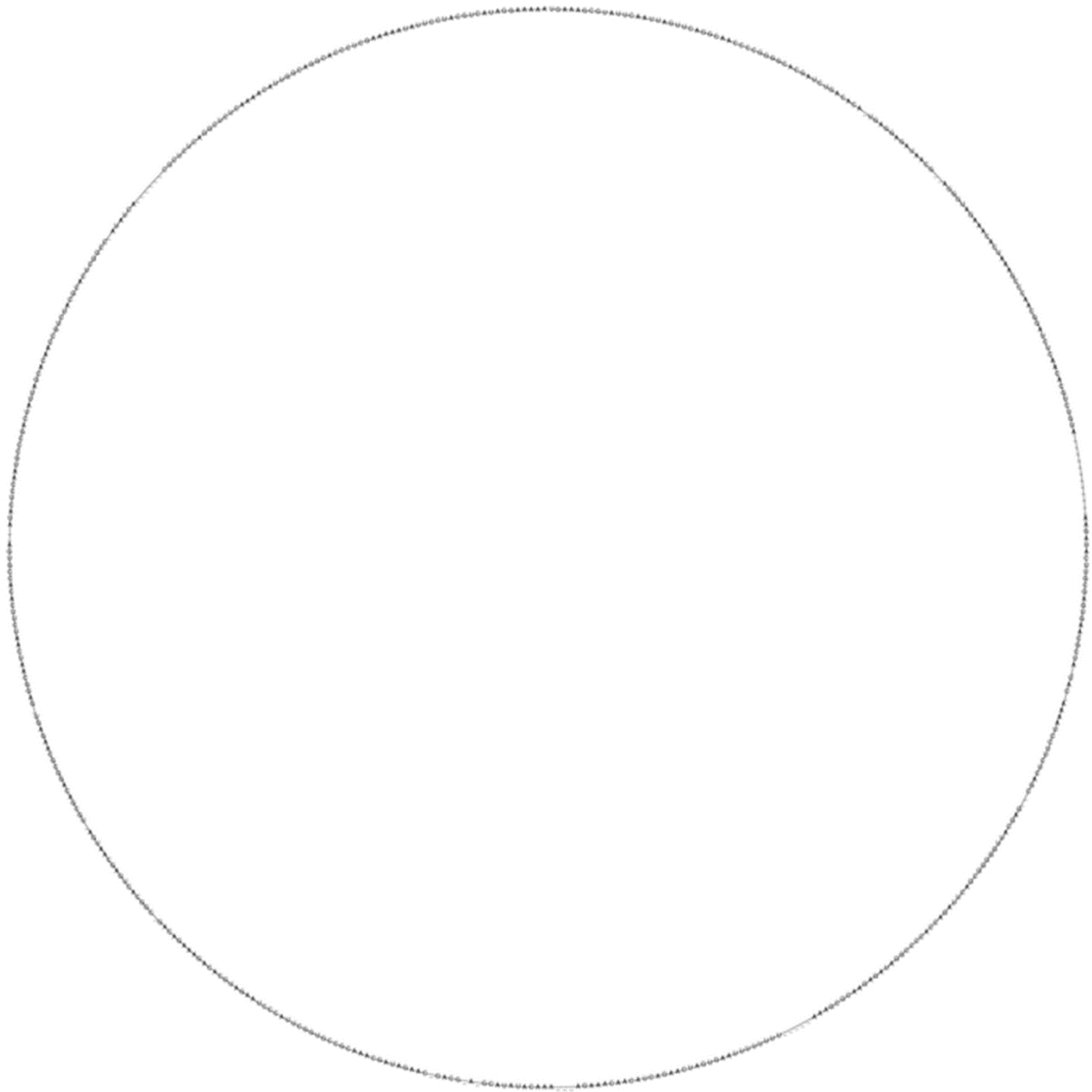
Abbildung A5: Kontrolle der Konsensus-Struktur von der Templat RNA

A: Kontroll-Alignment



Das multiple Alignment wurde mit LocARNA erstellt und besteht aus der Templat RNA aus *Nostoc* sp. PCC7120 (Transkriptionsstart1 bis Transkriptionsstopp2) und 12 zufällig generierten Sequenzen. Die zufällig generierten Sequenzen (GC-Gehalt: 50 %, Länge: 537 nt) wurden bereits zuvor mit der Templat RNA aus *Nostoc* sp. PCC7120 aligned (ClustalW), um den Sequenzbereich der Templat RNA für die LocARNA-Berechnung verwenden zu können. LocARNA führt einen sequenz- und strukturbasierten Vergleich der RNA-Moleküle durch. Die Datenausgabe besteht aus einer Konsensus-Struktur (Abbildung A5, B) und dem abgebildeten multiplen Alignment. Das Alignment zeigt keine konservierten Bereiche.

B: Kontroll-Konsensus-Struktur



Die Konsensus-Struktur von 12 zufällig generierten Sequenzen und der Templat RNA aus *Nostoc sp.* PCC7120 wurde unter Hinzunahme der Information aus einem multiplen Alignment mittels LocARNA erstellt (Abbildung A5, A). In der Konsensus-Struktur kommen keine konservierten Strukturelemente vor. Demnach scheinen die konservierten Strukturelemente der verschiedenen Templat RNAs (Abbildung 32) nicht durch zufällige Sequenzen entstanden zu sein, sondern repräsentieren funktionell wichtige, konservierte Strukturelemente.

Danksagung

Zunächst möchte ich mich bei Dr. Nora Zingler für die Aufnahme in die Arbeitsgruppe und die Betreuung meiner Arbeit bedanken.

Zudem soll an dieser Stelle Prof. Dr Matthias Hahn für die Übernahme der Zweitkorrektur, sowie Prof. Dr. John A Cullum für die Übernahme des Vorsitzes der Promotionskommission gedankt werden.

Ein besonderer Dank geht auch an die Arbeitsgruppe der Pflanzenökologie & Systematik der TU Kaiserslautern (Prof. Dr. Burkhard Büdel) für die Bereitstellung der Stammsammlung von Cyanobakterien, sowie die Hilfe im Umgang mit derselbigen.

In diesem Kontext würde ich mich auch gerne bei der Arbeitsgruppe Pflanzenphysiologie der TU Kaiserslautern (Prof. Dr. Ekkehard Neuhaus) für die Nutzung der Kniese-Apparatur bedanken, die für die Aufzucht der Cyanobakterien hilfreich war.

Die Erstellung des phylogenetischen Baums der DGR-kodierten RTs wurde von Damir Baranasic aus der Arbeitsgruppe Genetik der TU Kaiserslautern (Prof. Dr. John A Cullum) übernommen und ihm soll an dieser Stelle recht herzlich dafür gedankt werden.

Zu guter Letzt möchte ich mich bei den früheren Mitgliedern der Abteilung Molekulare Genetik der TU Kaiserslautern (Dr. Nora Zingler) bedanken. Meine Büronachbarn Lisa Jakobi, Maïke Gieseke und Dr. Thomas Schillinger standen mir mit Rat und Tat zur Seite. Insbesondere möchte ich mich hier bei Dr. Thomas Schillinger bedanken, der Teile dieser Arbeit gelesen und mir anschließend ein nützliches Feedback gegeben hat. Auch gilt Christine Förster-Schorr ein besonderer Dank für ihre helfende Hand in unzähligen Versuchen.

Zudem würde ich mich gerne dafür Bedanken, die Bachelorarbeiten von Christine Pakebusch und Katrin Haag, sowie die Diplomarbeit von Thomas Kafka betreut haben zu dürfen. Die Zusammenarbeit war freundlich und konstruktiv. Euch gilt an dieser Stelle ein kräftiges Dankeschön.

Lebenslauf

Name

Philipp Möller

Schullaufbahn

1990 – 1994

Grundschule Bugenhagenschule, Hamburg

1995 - 2003

Heinrich-Hertz Schule, Hamburg

Abschluss: Allgemeine Hochschulreife

Studium

2004 – 2009

Studium der Biologie an der Technischen Universität
Kaiserslautern

Dezember 2009

Diplomarbeit in der Abteilung Pflanzenphysiologie unter der
Leitung von Prof. Dr. Neuhaus

Titel: Molekulare Charakterisierung tonoplastidärer Proteine

Promotion

Seit Januar 2010

Promotion an der Technischen Universität Kaiserslautern in der
Abteilung Molekulare Genetik unter der Leitung von Jun. Prof.
Dr. Nora Zingler

Thema: Identifizierung von diversitätsgenerierenden
Retroelementen (DGRs) in Cyanobakterien und strukturelle
Charakterisierung der Templat RNA eines DGRs aus dem
Cyanobakterium *Nostoc sp.* PCC7120

Eidesstattliche Versicherung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbständig und ohne unerlaubte Hilfe verfasst und keine anderen Quellen und Hilfsmittel als die angegebenen verwendet habe.

Ich versichere, dass ich weder an der Technischen Universität Kaiserslautern noch anderweitig versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

Kaiserslautern, 19. Juni 2015

Philipp Möller