

Transkriptomsequenzierung bei Corynebakterien zur Charakterisierung von Mutanten in der RNA-Synthese und im RNA-Abbau

Dissertation zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)
an der Fakultät für Biologie der Universität Bielefeld

Verfasst und vorgelegt von
Manuel Wittchen
Geboren in Herdecke

Bielefeld, Mai 2018

Gedruckt auf alterungsbeständigem Papier ISO 9706.

Inhaltsverzeichnis

Abkürzungsverzeichnis	9
1 Zusammenfassung	11
2 Theoretischer Hintergrund	13
2.1 Taxonomie der Corynebakterien	13
2.1.1 Medizinische Relevanz von <i>Corynebacterium diphtheriae</i>	13
2.1.2 <i>Corynebacterium glutamicum</i> als industriell genutztes Bakterium	15
2.2 Transkriptionsinitiation in Bakterien	16
2.3 Rho-abhängige Transkriptionstermination	18
2.3.1 Prinzip der Rho-abhängigen Transkriptionstermination	19
2.3.2 Struktur des Rho-Faktors	21
2.3.3 Verbreitung des Rho-Faktors und weitere Funktionen	21
2.4 Prokaryotische Transkript-Degradation	22
2.4.1 Transkript-Degradation in gramnegativen und grampositiven Bakterien	22
2.4.2 Regulation der Transkript-Degradation	26
2.5 RNA-Sequenzierung zur Transkriptomanalyse	27
2.6 Ziele der Arbeit	30
2.7 Bisherige Publikationen	31
3 Material und Methoden	33
3.1 Bakterienstämme, Plasmide und Primer	33
3.2 Verwendete Chemikalien, Materialien, Geräte und Software	36
3.3 Medien und Lösungen	40
3.4 Kultivierung und Lagerung von Bakterien	41
3.5 Bestimmung des Bakterientiters	42
3.6 Allgemeine DNA-Arbeiten	42
3.6.1 Polymerase-Kettenreaktion (PCR)	42
3.6.2 Isolierung von DNA	44
3.6.3 Agarose-Gelelektrophorese	44

3.6.4	DNA-Aufreinigung und -Quantifizierung	44
3.6.5	Klonierung mittels Gibson-Assembly	45
3.6.6	Herstellung von kompetenten Zellen	45
3.6.7	Transformation von DNA	47
3.7	Bestimmung der Transformationsfrequenz	48
3.7.1	Durchführen von Gendelektionen	48
3.7.2	Plasmidbasierte Genexpression	49
3.7.3	Insertion von Genen mittels pPRIM2-System	49
3.8	Allgemeine Protein-Arbeiten	50
3.8.1	Proteinaufreinigung mit Affinitätschromatographie	50
3.8.2	Proteinquantifizierung mittels Bradford-Test	52
3.8.3	Proteinidentifikation mit SDS-PAGE	53
3.8.4	Proteinidentifikation mit MALDI-ToF/ToF-MS	55
3.9	Allgemeine RNA-Arbeiten	58
3.9.1	Isolierung von RNA	58
3.9.2	Quantifizierung von RNA	58
3.10	Quantitative Reverse-Transkriptase-PCR	59
3.11	Northern Blot	60
3.12	RNA-Sequenzierung	61
3.12.1	Sequenzierung von Vollängentranskripten	61
3.12.2	Sequenzierung von ribosomaler RNA	62
3.12.3	Sequenzierung der 5'-Enden von Primärtranskripten	63
3.12.4	Sequenzierung von RNase-Schnittstellen mittels <i>RNase-Assay</i>	64
3.13	Bioinformatische Analysen der RNA-Sequenzierungsdaten	66
3.13.1	Qualitätskontrolle und Read-Mapping	66
3.13.2	Bestimmung differenziell transkribierter Gene	68
3.13.3	Identifizierung von Transkriptionsstarts	69
3.13.4	Identifizierung von neuen Transkripten	69
3.13.5	Identifizierung von Operonstrukturen	70
3.13.6	Identifizierung von Sequenzmotiven	70
3.13.7	Identifizierung von verlängerten Transkripten	71
3.13.8	Identifizierung von RNase-Schnittstellen im <i>RNase-Assay</i>	71

4	Ergebnisse	73
4.1	Analyse der Transkriptionsinitiation in <i>Corynebacterium diphtheriae</i>	73
4.1.1	Probengenerierung und Sequenzierung der <i>C. diphtheriae</i> cDNA-Bibliotheken	74
4.1.2	Identifizierung von Transkriptionsstarts	75
4.1.3	Charakterisierung des σ^A Promotormotivs anhand der identifizierten Transkriptionsstarts	76
4.1.4	Die Analyse der 5'-Untranslatierten Regionen ergibt eine für Corynebakterien charakteristische Längenverteilung	77
4.1.5	Verbesserung der Gen-Annotation durch Identifizierung von neuen Transkripten und Korrektur von Translationsstarts	80
4.1.6	Identifizierung von Operon-Strukturen durch die Kombination von Gesamttranskriptom- und TSS-Daten	82
4.1.7	Das Diphtherietoxin kodierende Gen <i>tox</i> besitzt zwei bisher unbekannte <i>antisense</i> Transkripte	85
4.1.8	Vergleichende Analyse der DtxR-regulierten Gene	87
4.2	Rho-abhängige Transkriptionstermination in <i>Corynebacterium glutamicum</i> . .	91
4.2.1	Vergleich des Rho-Proteins aus <i>C. glutamicum</i> mit den Homologen der Modellorganismen	91
4.2.2	Die Deletion des Rho-Faktor Gens führt zu verlangsamtem Wachstum .	92
4.2.3	Die Deletion von <i>rho</i> führt zu einer deutlichen Veränderung des Transkriptoms	94
4.2.4	Der Rho-Faktor beeinflusst die Transkription potentieller Fremd-DNA	98
4.2.5	Hinweise auf verlängerte Transkripte in der $\Delta\rho$ -Mutante	101
4.2.6	Identifizierung von Rho-abhängig terminierten Transkripten	103
4.2.7	Längenanalyse der Rho-abhängig terminierten Transkripte	106
4.2.8	Analyse von <i>Rho utilization sites</i>	107
4.2.9	Die Termination von <i>antisense</i> RNA ist Rho-abhängig	111
4.2.10	Rho reguliert die Termination bestimmter Riboswitches	112
4.2.11	Die Transkription eines Typ IVb Pilus-Gen-Clusters in <i>C. glutamicum</i> ist Rho-abhängig	115
4.2.12	Rho aktiviert die Transkription von Kompetenzgenen	119
4.2.13	Überexpression des Typ IVb Pilus-Gen-Clusters führt zu höherer Transformationsfrequenz	120

4.3	RNA-Degradation durch die RNasen E/G und J in <i>C. glutamicum</i>	122
4.3.1	Vergleich von RNase E/G und RNase J mit Homologen der Modellorganismen	122
4.3.2	Deletion der für die RNasen E/G und J kodierenden Gene und Analyse des Wachstums	124
4.3.3	Die Deletionen der für die RNasen E/G und J kodierenden Gene haben einen starken Einfluss auf das Transkriptom	126
4.3.4	Prozessierung des 5S rRNA Vorläufers durch RNase E/G	131
4.3.5	Entwicklung einer RNA-Sequenzierungsmethode zur Erfassung von RNase-Schnittstellen	136
4.3.6	Endoribonukleolytische mRNA-Degradation durch die RNase E/G	139
4.3.7	Die Hauptfunktion der bifunktionalen RNase J ist der exoribonukleolytische Abbau ausgehend vom 5'-Ende	142
4.3.8	Regulation von Riboswitches und RNA-Motiven durch die RNasen E/G und J	148
5	Diskussion	151
5.1	Primärtranskriptsequenzierung als geeignete Methode zur Analyse der Transkriptionsinitiation	151
5.2	Analyse der Transkriptionsinitiation in <i>Corynebacterium diphtheriae</i>	152
5.2.1	Analyse der genomweit identifizierten Transkriptionsstarts	152
5.2.2	Nachweis neuer potenziell regulatoriver RNAs	153
5.2.3	Charakterisierung des σ^A Promotormotivs und von regulatorischen Elementen in 5'-UTRs	154
5.2.4	Identifizierung von Operon-Strukturen	156
5.2.5	Analyse des DtxR-Regulons in <i>C. diphtheriae</i>	157
5.2.6	Analyse der Phagen-Insel mit Fokus auf dem Diphtherietoxin-kodierenden Gen <i>tox</i>	158
5.3	Rho-abhängige Transkripttermination in <i>C. glutamicum</i>	160
5.3.1	Vergleich der Struktur des Rho-Faktors aus <i>C. glutamicum</i>	160
5.3.2	Rho reprimiert die Transkription von potenzieller Fremd-DNA	161
5.3.3	Die Deletion von <i>rho</i> führt zu verlängerten Transkripten	162
5.3.4	Identifikation der in der Δrho -Mutante verlängerten Transkripte	163
5.3.5	Rho terminiert <i>antisense</i> RNAs	163
5.3.6	Rho ist an der Regulation einiger Riboswitches beteiligt	165
5.3.7	Motivanalyse von möglichen <i>rut sites</i> von mRNAs	165

5.3.8	Rho reguliert ein Typ IVb Pilussystem	167
5.4	RNA-Degradation durch die RNasen E/G und J in <i>C. glutamicum</i>	171
5.4.1	<i>C. glutamicum</i> besitzt eine von den Modellorganismen abweichende RNase-Ausstattung	171
5.4.2	Die RNasen E/G und J besitzen zusätzliche N-terminale Bereiche . . .	172
5.4.3	Die RNasen E/G und J nicht essenziell	173
5.4.4	Einfluss der RNasen E/G und J das Transkriptom in <i>C. glutamicum</i> . .	174
5.4.5	Die Entwicklung des <i>RNase-Assays</i> ermöglicht die Identifizierung von RNase-Schnittstellen	176
5.4.6	Cg.RNase E/G ist an der Initiation der RNA-Degradation beteiligt . . .	178
5.4.7	Cg.RNase E/G schneidet RNA an einem konservierten Motiv	180
5.4.8	RNase J besitzt hauptsächlich Exoribonuklease-Aktivität	181
5.4.9	Prozessierung und Degradation von Riboswitches und anderen regula- torischen Elementen durch die RNasen E/G und J	184
6	Ausblick	187
	Literaturverzeichnis	189
	Tabellenverzeichnis	221
	Abbildungsverzeichnis	223
	Anhang	227
A.1	Ergänzende Abbildungen	227
A.2	Ergänzende Tabellen	236
A.3	Große ergänzende Tabellen	236
A.4	Detaillierte Protokolle der <i>RNase-Assays</i>	237
A.5	Quellcode der im Rahmen dieser Arbeit entwickelten Programme	237
	Danksagung	239
	Erklärung	241

Abkürzungsverzeichnis

5'-UTR	<i>5'-untranslated Region</i> (5'-untranslatierte Region)
% (v/v)	Volumenprozent
% (w/v)	massebezogenes Volumenprozent
μ_{\max}	maximale spezifische Wachstumsgeschwindigkeit
Amp^R	Ampicillin-Resistenz
AS	Aminosäure
asRNA	<i>antisense RNA</i> (komplementäre RNA auf Gegenstrang)
ATCC	<i>American Type Culture Collection</i>
bp	Basenpaare
cDNA	<i>complementary DNA</i> (komplementäre DNA)
CDS	<i>Coding Sequence</i> (Kodiersequenz)
Cm^R	Chloramphenicol-Resistenz
DNA	<i>Deoxyribonucleic Acid</i> (Desoxyribonukleinsäure)
IPTG	Isopropyl- β -D-thiogalactopyranosid
Km^R	Kanamycin-Resistenz
LFC	\log_2 (FoldChange) (Logarithmus zur Basis 2 des Änderungsverhältnis)
mRNA	<i>messenger RNA</i> (Boten-RNA)
NCTC	<i>National Collection of Type Cultures</i>
ncRNA	<i>noncoding RNA</i> (nichtkodierende RNA)
nt	Nukleotide
Nx^R	Nalidixinsäure-Resistenz
ORF	<i>Open Reading Frame</i> (offener Leserahmen)
RNA	<i>Ribonucleic Acid</i> (Ribonukleinsäure)
rpm	<i>Revolutions per minute</i> (Umdrehungen pro Minute)
rRNA	<i>ribosomal RNA</i> (ribosomale RNA)
TLS	<i>Translation Start Site</i> (Translationsstart)
TSS	<i>Transcription Start Site</i> (Transkriptionsstart)

1 Zusammenfassung

Der Prozess der Transkription umfasst das Umschreiben von DNA in RNA und ist somit Teil der Genexpression. Das Transkriptom ist die Gesamtheit aller RNA in einer Zelle zu einem bestimmten Zeitpunkt. Die Transkription besteht im Wesentlichen aus drei Unterprozessen: der Transkriptionsinitiation, der -termination und der Transkript-Degradation. An allen drei Unterprozessen sind eine Vielzahl von Enzymen und Proteinen beteiligt, die diese Prozesse katalysieren und regulieren.

Die in den letzten Jahren entwickelte Technologie der Transkriptom- bzw. RNA-Sequenzierung ermöglicht die Hochdurchsatz-Sequenzierung aller Transkripte und somit ganzer Transkriptome mit Einzelnukleotidauflösung. Darüber hinaus bildet sie einen extrem hohen dynamischen Bereich für die relative Quantifizierung von Transkripten ab. Neben der klassischen RNA-Sequenzierung von Vollängentranskripten wurden weitere Methoden entwickelt, um beispielsweise Primärtranskripte spezifisch anzureichern und somit Transkriptionsstarts genomweit zu identifizieren.

Im Rahmen dieser Arbeit wurden die Transkriptome in den medizinisch und industriell relevanten Corynebakterien *Corynebacterium diphtheriae* und *C. glutamicum* hinsichtlich der RNA-Synthese und des RNA-Abbaus untersucht. Im Speziellen wurden die oben genannten Prozesse der Transkriptionsinitiation und -termination sowie der Transkript-Degradation analysiert.

Obwohl das Genom von *C. diphtheriae* seit etwa 15 Jahren bekannt ist, ist nur wenig über die transkriptionelle Organisation dieses wichtigen Humanpathogens bekannt. Mittels Primär- und Vollängentranskript-Sequenzierung wurden in dieser Arbeit eine umfassende Analyse der genetischen Elemente, wie des σ^A Promotormotivs, 5'-UTRs sowie Operonstrukturen durchgeführt. Von 1.656 identifizierten Transkriptionsstarts (TSS) wurden 1.202 TSS bekannten und 454 TSS neuen Transkripten, wie *antisense* RNAs, zugeordnet. Insgesamt wurde ca. 87 % des *C. diphtheriae* Genoms als aktiv transkribiert nachgewiesen. Darüber hinaus wurde die Gen-Annotation von 120 Genen anhand der identifizierten TSS verbessert. Über die Vollängentranskriptdaten wurde außerdem das DtxR-Regulon, welches unter anderem das Diphtheritoxin-kodierende *tox* Gen umfasst, charakterisiert. Für das *tox* Gen wurden

1 Zusammenfassung

bisher unbekannte *antisense* RNAs und benachbarte intergenische Transkripte nachgewiesen, die möglicherweise regulatorische Funktionen besitzen.

Neben der Transkriptionsinitiation und daraus abgeleiteter Informationen in *C. diphtheriae* wurde in dieser Arbeit in dem industriell genutzten Aminosäureproduzenten *C. glutamicum* die Rho-abhängige Transkriptionstermination untersucht. Mittels RNA-Sequenzierung einer $\Delta\rho$ -Mutante wurde gezeigt, dass der Rho-Faktor überwiegend *antisense* RNAs terminiert und DNA aus Fremdquellen, wie Prophagen-Gene und Transposons, transkriptionell stilllegt. Des Weiteren wurde die Funktion und genetische Struktur eines durch den Rho-Faktor regulierten DNA-Aufnahmesystems in *C. glutamicum* nachgewiesen. Anhand von 3'-verlängerten Transkripten wurden 930 Rho-abhängig terminierte Transkripte im *C. glutamicum* Genom identifiziert. Die Analyse der Rho-abhängig terminierten Transkripte ergab eine von *Escherichia coli* Rho abweichende Rho-Erkennungsstelle (*rut site*), sodass die für *E. coli* Rho bekannten C-reichen *rut sites* kein Erkennungsmerkmal für *C. glutamicum* Rho darstellen.

Von der Zelle nicht mehr benötigte Transkripte werden über die Transkript-Degradation abgebaut. Die Vielzahl an der Transkript-Degradation beteiligter RNasen und anderer Proteine erlaubt eine sehr feingliedrige Regulation und damit strenge Prozesskontrolle. Im Vergleich zu den Modellorganismen *E. coli* und *Bacillus subtilis*, dessen wichtigste RNasen die RNasen E bzw. J und Y darstellen, besitzt *C. glutamicum* ein abweichendes RNase-Repertoire bestehend aus RNase E/G und J. Im Rahmen dieser Arbeit wurden die Einflüsse der RNasen E/G und J auf das *C. glutamicum* Transkriptom analysiert. So wirken sich die Deletionen der für RNase E/G und RNase J kodierenden Gene besonders stark auf das Transkriptom aus. Über eine in dieser Arbeit entwickelte RNA-Sequenzierungsmethode zur transkriptomweiten Identifizierung von RNase-Schnittstellen, wurden 2.528 RNase E/G und 5.850 RNase J Schnittstellen im Transkriptom von *C. glutamicum* identifiziert. Die Verteilung der Schnittstellen ergab für beide RNasen charakteristische Muster. So wurde über eine graduelle Abnahme der Schnittstellenverteilung, eine 5'-3'-Exoribonuklease-Aktivität für RNase J gezeigt. Für die Endoribonuklease RNase E/G wurde eine Schnitt-Präferenz im vorderen Transkriptbereich und das Motiv AAA↓CUU nachgewiesen.

Die hier vorgestellten Ergebnisse erweitern das Wissen über die Transkriptionsinitiation in *C. diphtheriae* und legen den Grundstein für darauf aufbauende Transkriptomanalysen. Darüber hinaus erlauben die hier präsentierten Ergebnisse tiefe Einblicke in die Transkriptionstermination und Transkript-Degradation in *C. glutamicum* und ermöglichen ein besseres Verständnis dieser stark regulierten Transkriptionsprozesse.

2 Theoretischer Hintergrund

2.1 Taxonomie der Corynebakterien

Die Familie Corynebacteriaceae (Lehmann und Neumann, 1896; Stackebrandt *et al.*, 1997; Zhi *et al.*, 2009) gehört zum Phylum Actinobacteria, Klasse Actinobacteria und Ordnung der Corynebacteriales (Goodfellow, 2012; Goodfellow und Jones, 2012; Stackebrandt *et al.*, 1997). Die Familie umfasst das Typ-Genus *Corynebacterium* (Lehmann und Neumann, 1896; Bernard *et al.*, 2010) sowie das monospezifische Genus *Turicella* (Funke *et al.*, 1994).

Das Genus *Corynebacterium* wurde erstmals von Lehmann und Neumann 1896 in *Bergey's Manual of Systematic Bacteriology* mit der Beschreibung des „diphtheroiden Bazillus“ (später als *Corynebacterium diphtheriae* bezeichnet) erwähnt. Die Namensgebung erfolgte aufgrund der keulenförmigen Morphologie der Zellen (griechisch coryne = Keule). Die Zellteilung erfolgt über *snapping division*, mit einem ungleichmäßigen Aufreißen der Zellwand, das zu V-förmig angeordneten Zellen führt (Abbildung 2.1) (Eggeling und Bott, 2005). Das Genus umfasst grampositive, 3 – 5 μm lange, nicht-motile, nicht-sporulierende und fakultativ anaerob bis aerob lebende Bakterien mit hohem G+C-Gehalt des Genoms (Goodfellow und Jones, 2012). Das Genus beinhaltet nach aktuellem Stand 90 Spezies (Tauch und Sandbote, 2014). Der komplexe Aufbau der Zellmembran besteht aus Peptidoglycan verbunden an eine Schicht Arabinogalactan welche mit Mycolsäuren und Glycolipiden versehen ist (Eggeling *et al.*, 2008). Corynebakterien gehören daher zu den mycolsäurehaltigen Actinomyceten und sind eng mit den Mycobakterien verwandt (Nolden *et al.*, 2000).

Wichtige Vertreter der Corynebakterien sind zum einen das Pathogen *Corynebacterium diphtheriae* sowie das biotechnologisch relevante Bakterium *Corynebacterium glutamicum*, auf die im Folgenden gesondert eingegangen wird.

2.1.1 Medizinische Relevanz von *Corynebacterium diphtheriae*

C. diphtheriae wurde 1884 von Edwin Klebs und Friedrich Loeffler entdeckt und ist daher auch als Klebs-Loeffler-Bazillus bekannt. *C. diphtheriae* ist der Erreger der Atemwegserkrankung Diphtherie, welche sich durch starke toxin-induzierte Entzündung im oberen Bereich der Atemwege und eine graue Schicht aus totem Gewebe (Pseudomembran) äußert (Hoskisson, 2018).

2 Theoretischer Hintergrund

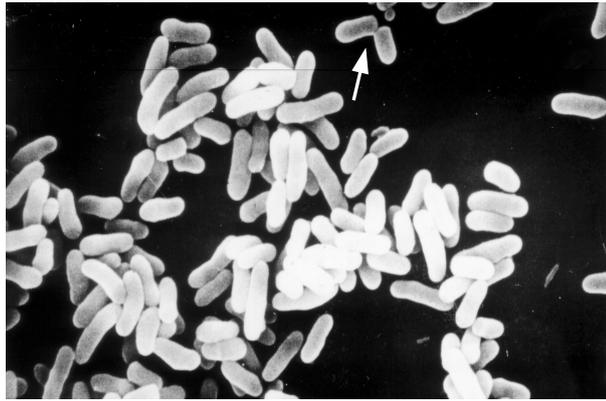


Abbildung 2.1: *Corynebacterium glutamicum* Zellen aufgenommen mit einem Rasterelektronenmikroskop. Aufgrund der *snapping division* bei der Zellteilung ordnen sich Zellen nach der Teilung V-förmig an (Pfeil). Bildquelle: Jörn Kalinowski (modifiziert). Abgedruckt mit freundlicher Genehmigung von Jörn Kalinowski.

Das Diphtherietoxin (DT) ist ein Exotoxin welches an den Vorläufer des Heparin-bindenden epidermalen Wachstumsfaktors von epidermalen Zellen bindet. Anschließend gelangt DT über Endozytose in die Zelle und modifiziert den Elongationsfaktor eEF-2 durch anhängen eines Ribosylrests. Diese Modifikation führt zu einer Inhibition der Proteinbiosynthese und somit zum Zelltod (Holmes, 2000). Neben den Toxin-produzierenden *C. diphtheriae* Stämmen, treten auch Erkrankungen durch toxin-freie *C. diphtheriae* Stämme auf. Diese können zu Endokarditis, septischer Arthritis, Osteomyelitis und kutanen Infektionen führen (Sangal und Hoskisson, 2014).

Durch die Einführung eines Impfstoffs und globaler Immunisierungsinitiativen konnten die Infektionsrate auf weltweit 5.000 Infektionen pro Jahr reduziert werden. Dennoch ist Diphtherie in Ländern mit schlechter Routineimpfung ein erhebliches Gesundheitsproblem (Hoskisson, 2018). Große Epidemien traten unter anderem in den 1990er Jahren in ehemaligen Staaten der Sowjetunion auf.

Das Genom des in weiten Teilen für diesen Ausbruch verantwortlichen Stamms *C. diphtheriae* NCTC 13129 wurde 2003 sequenziert (Cerdeño-Tárraga *et al.*, 2003). Das Genom von *C. diphtheriae* NCTC 13129 umfasst 2,5 Mbp mit einem G+C-Gehalt von 53,5 % und enthält 2.320 proteinkodierende Sequenzen (CDS) inklusive des DT kodierenden Gens *tox* (Cerdeño-Tárraga *et al.*, 2003).

Die Produktion von DT ist Eisen-abhängig und wird durch den Diphtherietoxin-Regulator (DtxR) transkriptionell reguliert (Boyd *et al.*, 1990; Schmitt und Holmes, 1991). Bei hoher Eisenkonzentration bindet DtxR Eisenionen und bildet ein Dimer, welches sich an eine palindromische Sequenz im Genom *upstream* des *tox* Promotors anlagert und so die Transkription

von *tox* inhibiert (Tao *et al.*, 1994). Das Gen *tox* stammt von einer Infektion des chromosomal integrierten Corynephagen, das für DtxR kodierende Gen *dtxR* befindet sich im bakteriellen Genom. Somit befindet sich die Genexpression von *tox* unter bakterieller Kontrolle (Holmes, 2000). Weiter konnte gezeigt werden, dass DtxR ein dualer Regulator der Eisenhomöostase ist (Brune *et al.*, 2006) und in vielen Bakterien Gene der Eisenaufnahme und -speicherung reguliert (Schmitt, 1997; Qian *et al.*, 2002; Kunkle und Schmitt, 2003, 2005; Trost *et al.*, 2012).

Aufgrund der großen medizinischen Relevanz wurde das Diphtherietoxin umfassend bezüglich seiner Struktur (Kantardjieff *et al.*, 1989; Choe *et al.*, 1992; Louie *et al.*, 1997; Li *et al.*, 2014) und seinem Wirkmechanismus charakterisiert (Hoch *et al.*, 1985; Ladokhin, 2013). Zur transkriptionellen Organisation von *C. diphtheriae* existieren allerdings nur wenige Studien (Leong und Murphy, 1985; Schmitt, 1997; Oram *et al.*, 2006; Burgos und Schmitt, 2012). Promotoren und Operonstrukturen sind für den überwiegenden Teil der Gene unbekannt.

2.1.2 *Corynebacterium glutamicum* als industriell genutztes Bakterium

Vertreter der Gattung *Corynebacterium* zeichnen sich durch eine Vielzahl an StoffwechsellLeistungen aus und werden für die fermentative Produktion von Aminosäuren als Futter-, Nahrungs- und Arzneimittelzusätze eingesetzt (Schaffer *et al.*, 2001). Auch in der Herstellung und Konversion von Steroiden (Constantinides, 1980) und dem Abbau halogenierter Kohlenwasserstoffe (Yokota *et al.*, 1987) werden Corynebakterien eingesetzt. Darüber hinaus wurden Corynebakterien in der Käsereifung nachgewiesen (Schröder *et al.*, 2011).

Corynebacterium glutamicum ist vermutlich der bekannteste Vertreter der nicht-pathogenen Corynebakterien und wird aufgrund seines Status als GRAS-Organismus (*Generally Regarded As Safe*) häufig für die fermentative Produktion von Chemikalien eingesetzt.

C. glutamicum, damals noch als *Micrococcus glutamicus* bezeichnet, wurde 1957 aufgrund seiner Fähigkeit zur natürlichen Sekretion von Aminosäuren bei der Suche nach einem Aminosäure-produzierenden Mikroorganismus identifiziert (Kinoshita *et al.*, 2004). Die Aminosäuresekretion von *C. glutamicum* basiert auf einem *Overflow*-Metabolismus, sowie begrenzten katabolischen Stoffwechselwege zum Recycling überflüssiger Aminosäuren (Hillemann *et al.*, 1993). Um die industrielle Produktion von Aminosäuren weiter zu steigern, wurde in die Entwicklung und Optimierung von gentechnischen Methoden für *C. glutamicum* investiert. Darüber hinaus wurde die Grundlagenforschung als auch gezieltes *Metabolic Engineering* intensiviert (Bardonnet und Blanco, 1991; Cadenas *et al.*, 1991; Eikmanns *et al.*, 1991; Schäfer *et al.*, 1994; Tauch *et al.*, 1995; Kirchner und Tauch, 2003). Aktuell beträgt die weltweite

2 Theoretischer Hintergrund

fermentative Produktion von Aminosäuren, wie Glutamat, mehrere Millionen Tonnen pro Jahr (Hirasawa und Shimizu, 2016).

Im Jahr 2003 wurde das Genom von *C. glutamicum* ATCC 13032 zeitgleich von zwei Arbeitsgruppen publiziert (Ikeda und Nakagawa, 2003; Kalinowski *et al.*, 2003). Das *C. glutamicum* Genom umfasst 3,3 Mbp mit einem G+C-Gehalt von 53,8 % und enthält knapp 3.000 CDS (Kalinowski *et al.*, 2003). Das sequenzierte Genom bildet die Grundlage für die Entwicklung von Produktionsstämmen und etablierte *C. glutamicum* als Modellorganismus der Actinomycetales, insbesondere für die pathogenen Organismen, wie *C. diphtheriae* und *Mycobacterium tuberculosis*. Darüber hinaus ermöglicht die Kenntnis der Genomsequenz die globale Analyse der transkriptionellen Landschaft und erlaubt so detaillierte Erkenntnisse zum Transkriptom (Pfeifer-Sancar *et al.*, 2013).

2.2 Transkriptionsinitiation in Bakterien

Die Anpassung an sich ändernde Umweltbedingungen ist eine wesentliche Voraussetzung für Lebewesen im Allgemeinen und für Bakterien im Speziellen. Da sich Umweltbedingungen schnell verändern können, müssen der Metabolismus angepasst und entsprechende Proteine gebildet werden. Dieser Prozess geschieht unter anderem auf der Ebene der Genexpression (Browning und Busby, 2016).

Die Regulation der Genexpression kann an jedem Einzelschritt der Genexpression von der Transkription bis zur Translation erfolgen. Eine posttranslationale Modifizierung der Proteinausstattung ist zusätzlich möglich. Die Transkription oder genauer gesagt die Transkriptionsinitiation ist aber in Bakterien der am häufigsten und am stärksten regulierte Schritt (Browning und Busby, 2004).

Die Transkription umfasst drei Phasen: Initiation, Elongation und Termination (Helmann, 2009). Die Transkriptioninitiation beginnt mit der Bindung der RNA-Polymerase II (RNAP) an die Promotorregion auf der DNA. Die RNAP besteht aus zwei α -Untereinheiten und jeweils einer β -, β' - und ω -Untereinheit die zusammen das katalytische Grundgerüst der RNAP (*RNAP Core Enzyme*) bilden. Die Untereinheiten β und β' sind für die Bindung an die DNA notwendig und bilden auch das aktive Zentrum, in dem Nukleotide an den 3'-Hydroxylrest des naszierenden RNA-Strangs angehängen werden. Die beiden α -Untereinheiten sind für die Assemblierung sowie die Stabilität der RNAP verantwortlich und unterstützen die Bindung an die Promotorregion durch die Erkennung des *upstream* befindlichen UP-Elements. Die ω -Untereinheit stabilisiert ebenfalls die RNAP (Vassylyev *et al.*, 2002). Sigmafaktoren (σ) ermöglichen die spezifische Bindung der RNAP an Promotorregionen (Abbildung 2.2). Sind Sigmafaktoren an das RNAP *Core Enzyme* gebunden, wird der Komplex als RNAP-Holoenzym

2.2 Transkriptionsinitiation in Bakterien

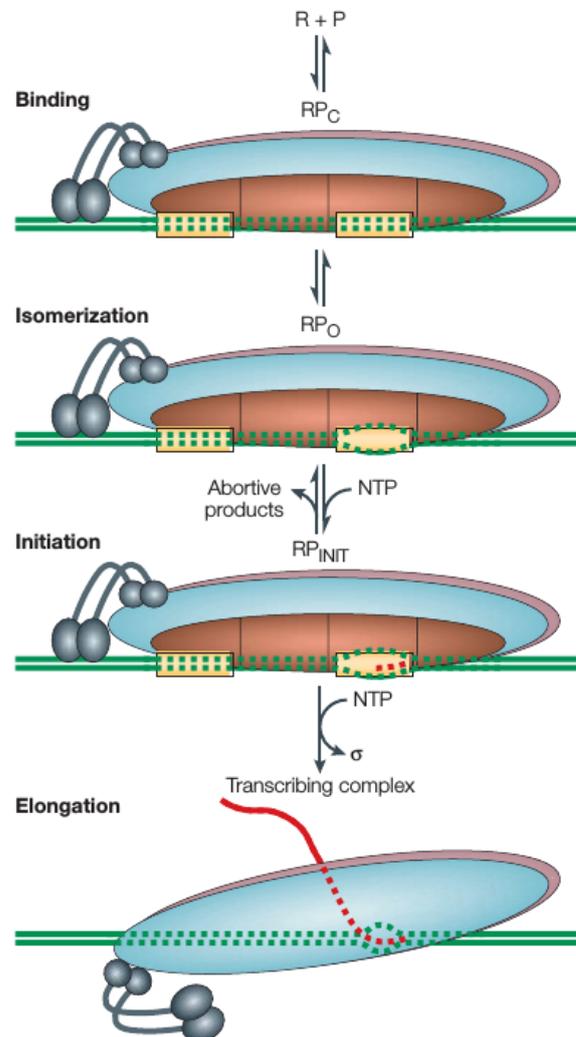


Abbildung 2.2: Mechanismus der Transkriptionsinitiation an bakteriellen Promotoren. Die RNA-Polymerase (R) interagiert mit dem Promotor auf der DNA (P) und bildet einen geschlossenen Komplex (RP_C). Gestrichelte Linien zeigen die Promotorregionen die von dem RNA-Polymerase-Holoenzym gebunden werden. Der DNA-Doppelstrang am Transkriptionsstart wird aufgewunden (dargestellt durch die „Blase“ in der DNA die von der RNA-Polymerase gebunden wird), um den offenen Komplex (RP_O) zu formen. Der Initiationskomplex (RP_{INIT}) beginnt mit der Synthese von RNA (gestrichelte rote Linie) anhand der DNA-Vorlage über die Ausbildung einer Phosphodiester-Bindung zwischen dem ersten und dem nachfolgenden Nucleosidtriphosphat (NTP). Die Elongation ist die finale Stufe in der der gebildete RNA-Strang (rote Linie) verlängert wird. In dieser Stufe ist der Sigmafaktor (braunes Oval) nicht mehr an die RNA-Polymerase gebunden. Bildquelle: Browning und Busby (2004). Abgedruckt mit freundlicher Genehmigung von Springer Nature.

2 Theoretischer Hintergrund

bezeichnet (Borukhov und Nudler, 2008). Verschiedene Sigmafaktoren ermöglichen die Erkennung von unterschiedlichen Promotorregionen und bilden so eine grundlegende Regulation der Transkription. Der Hauptsigmafaktor σ^{70} aus *E. coli* besitzt eine Größe von ca. 70 kDa und ist für die Transkription von Haushaltsgenen (*housekeeping genes*) verantwortlich und ist in allen Bakterien konserviert (Silverstone *et al.*, 1972; Browning und Busby, 2004). In Corynebakterien wird der σ^{70} als σ^A bezeichnet (Pátek *et al.*, 2003; Pátek und Nešvera, 2011). Auf der DNA erkennt σ^{70} Promotorkernmotive, die sich ca. 10 bzw. 35 bp *upstream* vom Transkriptionsstart befinden. Diese Motive werden daher auch -10 bzw. -35 Motiv genannt (Helmann, 2009). Nach der Erkennung des Promotors durch das RNAP-Holoenzym werden zunächst kurze Fragmente von unter zehn Nukleotiden gebildet, bis der Übergang zum Elongationskomplex (EC) vollzogen ist (Abbildung 2.2) (Hsu, 2009). Im EC ist der Sigmafaktor von der RNAP dissoziiert und die nächste Phase der Transkription, die Elongation, beginnt. Die Elongation erfolgt über tausende Nukleotide bis die Transkripttermination einsetzt und das Transkript von der RNAP freigesetzt wird.

Neben Sigmafaktoren existieren unterschiedlichste Transkriptionsregulatoren mit vielfältigen Wirkmechanismen (Übersichtsartikel von Browning und Busby (2016)). Beispielhaft wurde oben bereits der Diphtherietoxin-Repressor (DtxR) beschrieben.

2.3 Rho-abhängige Transkriptionstermination

Die Termination der Transkription ist ein wichtiger Prozess der Transkription. Durch die Transkripttermination werden RNA-Polymerasen frei, die so wieder für die Expression anderer Gene zur Verfügung stehen. Auch verhindert die Transkriptionstermination die Genexpression von *downstream* befindlichen Genen und dadurch auftretende Interferenzen mit *antisense* Transkripten und ist für die Erzeugung von definierten 3'-Enden für präzise RNA-Strukturen, wie zum Beispiel nichtkodierenden RNAs (*non-coding RNAs*, ncRNAs), wichtig. Des Weiteren verringert die Termination der Transkription Kollisionen zwischen Replikationskomplexen (Ray-Soni *et al.*, 2016).

Grundsätzlich wird zwischen zwei Arten der Transkriptionstermination unterschieden: (a) intrinsische, Faktor-unabhängige und (b) Faktor-abhängige Termination. Die intrinsische Transkriptionstermination benötigt nur die bei der Transkription vorhandenen Komponenten RNAP, RNA und DNA. Die Faktor-abhängige Termination benötigt hingegen weitere Faktoren, wie den Rho- oder Mfd-Faktor (Ray-Soni *et al.*, 2016). Der Mfd-Faktor spielt für die allgemeine Transkription eine unwichtige Rolle, da er nur bei der Auflösung und Reparatur von beschädigten Elongationskomplexen relevant ist (Roberts und Park, 2004) und daher hier nicht weiter behandelt wird.

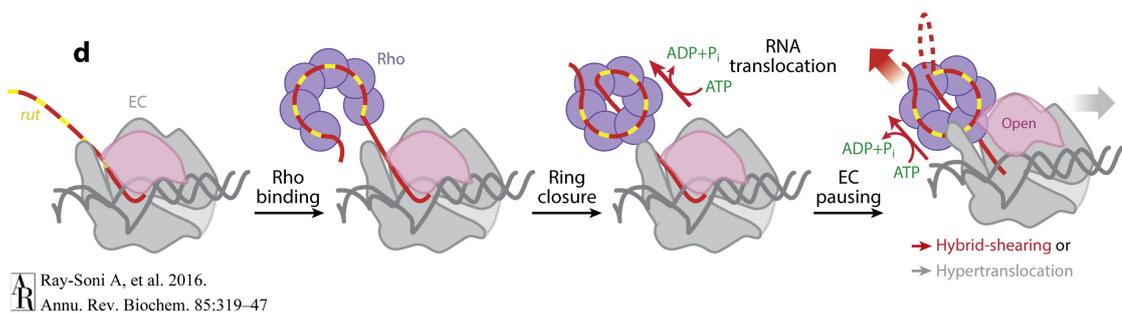
2.3 Rho-abhängige Transkriptionstermination

Der Fokus dieser Arbeit liegt auf der Rho-abhängigen Termination. Daher wird die intrinsische Transkriptionstermination im Folgenden nur kurz umrissen. Der Übersichtsartikel von Ray-Soni *et al.* (2016) bietet eine detaillierte Beschreibung des Terminationsmechanismus über Haarnadelterminatoren. Die intrinsische Transkriptionstermination erfolgt durch Ausbildung einer Sekundärstruktur in der RNA. Die Sekundärstruktur erinnert an die Form einer Haarnadel (*hairpin*) und wird daher auch Haarnadel-Terminator genannt. Der Haarnadel-Terminator besteht aus zwei kanonischen Elementen: einer GC-reichen Sequenz, die die Haarnadel bildet und einem *downstream* liegenden U-reichen Abschnitt. Während der Transkription faltet sich die RNA, die die Haarnadelstruktur kodiert, zu einem Haarnadelterminator innerhalb der RNAP. Dieser Haarnadelterminator führt zu einem instabilen RNA-DNA-Hybrid, der zur Dissoziation der RNAP und somit zur Transkriptionstermination führt (Ray-Soni *et al.*, 2016). Aufgrund der einfachen Sequenz- und Struktureigenschaften lassen sich Haarnadelterminatoren über die Genomsequenz bioinformatisch vorhersagen (d'Aubenton Carafa *et al.*, 1990).

2.3.1 Prinzip der Rho-abhängigen Transkriptionstermination

Die Rho-abhängige Transkriptionstermination wird durch die ATP-abhängige RNA-DNA-Helicase Rho (auch Rho-Faktor oder kurz Rho genannt) vermittelt (Ray-Soni *et al.*, 2016). Die zelluläre Funktion und Struktur des Rho-Faktors wurde hauptsächlich in *E. coli* analysiert (Cardinale *et al.*, 2008; Peters *et al.*, 2012; Leela *et al.*, 2013). Rho bildet eine offene Ringstruktur aus sechs Monomeren. In der aktiven Form schließt sich der Ring und bildet eine zentrale Pore (Skordalakes und Berger, 2003). Der Rho-Faktor erkennt und bindet an unspezifische Signale auf dem naszierenden Transkript (Lowery-Goldhammer und Richardson, 1974). Nach erfolgter Bindung über primäre Bindestellen wird die RNA durch die zentrale Pore geführt und die Translocase-Aktivität durch den Ringschluss aktiviert (Skordalakes und Berger, 2003) (Abbildung 2.3). Anschließend bewegt sich der Rho-Faktor unter ATP-Verbrauch entlang der RNA zum Elongationskomplex (EC). Sobald der Rho-Faktor die RNA vollständig durch die Pore gezogen hat, pausiert der EC und die Dissoziation des EC wird eingeleitet (Park und Roberts, 2006). Entweder wird der RNA-DNA-Hybrid getrennt, indem der Rho-Faktor die RNA aus dem EC herauszieht (*hybrid-shearing*) oder die RNAP durch Rho auf der DNA verschoben wird, sodass die RNA von der RNAP gelöst wird (*hypertranslocation*) (Park und Roberts, 2006). Es ist allerdings noch ungeklärt, ob der Rho-Faktor freie RNA-Enden erkennt, bindet und sich dann in Richtung RNAP bewegt oder ob der Rho-Faktor bereits an der RNAP gebunden ist und während der Transkription die RNA direkt erkennt (Ray-Soni *et al.*, 2016).

2 Theoretischer Hintergrund



Ray-Soni A, et al. 2016.
Annu. Rev. Biochem. 85:319–47

Abbildung 2.3: Mechanismus der Rho-abhängigen Transkriptionstermination. Der Rho-Faktor erkennt *rut sites* (gelb) auf der naszierenden RNA (rote Linie) und bindet an diese in der offenen Ringform. Durch schließen des Hexamer-Rings wird die RNA durch die zentrale Pore gezogen und Rho wandert unter ATP-Verbrauch in Richtung des Elongationskomplexes (EC) bestehend aus RNAP (grau und rosa), RNA und DNA (graue Linien). Bei Kontakt mit dem EC wird die RNA aus dem *exit channel* der RNAP über *hybrid-shearing* oder *hybertranslocation* (rote und graue Pfeile) entfernt. In dieser Abbildung bindet der Rho-Faktor an das freie Ende der RNA. Möglicherweise bindet der Rho-Faktor aber schon während der Transkription an die RNAP und interagiert direkt mit dem Transkript (nicht dargestellt). Bildquelle: Ray-Soni *et al.* (2016) (modifiziert). Abgedruckt mit freundlicher Genehmigung von *Annual Reviews*.

Der Rho-Faktor bindet bevorzugt RNA-Sequenzen mit hohem C- und geringem G-Gehalt die eine geringe Sekundärstruktur aufweisen. Die höchste Affinität zeigt Rho für Poly(C)-RNA (Lowery-Goldhammer und Richardson, 1974; Alifano *et al.*, 1991). Diese Bereiche werden auch CG-Blase (*CG bubble*) genannt und bilden die als *Rho utilization sites (rut sites)* bezeichneten Erkennungsstellen auf der RNA von denen aus die Transkriptionstermination eingeleitet wird. In *E. coli* sind die CG-Blasen ca. 80 – 90 nt lang (Chen und Richardson, 1987; McSwiggen *et al.*, 1988; Zalatan und Platt, 1992). Die Länge ist ausreichend damit jede der sechs Rho-Untereinheiten an die *rut site* binden kann (Koslover *et al.*, 2012). Allerdings ist unklar, ob sich die *rut site* über alle Untereinheiten erstrecken muss, um die Translokation zu aktivieren oder nur einige Untereinheiten dafür mit der *rut site* in Kontakt treten müssen (Canals *et al.*, 2010). Die Termination der Transkription erfolgt meist 10 – 20 nt, nur selten mehr als 100 nt *downstream* von der *rut site* (Thomsen und Berger, 2009). Sekundärstrukturen innerhalb von *rut sites* verhindern eine Bindung des Rho-Faktors. Sobald die Translokase-Aktivität des Rho-Faktors aktiviert wurde, kann der Rho-Faktor auch komplexe Sekundärstrukturen auflösen. Es wird davon ausgegangen, dass ein einzelnes Konsensusmotiv für *rut sites* nicht existiert, da Rho von diversen Poly(C)-RNAs aktiviert werden kann (Hart und Roberts, 1991; Ray-Soni *et al.*, 2016). Eine bioinformatische Vorhersage von *rut sites* über die Genomsequenz ist daher vermutlich nicht möglich (Ray-Soni *et al.*, 2016).

2.3.2 Struktur des Rho-Faktors

Basierend auf der Strukturanalyse von *E. coli* Rho (Ec.Rho) wird der Rho-Faktor in eine N-terminale (NTD) und eine C-terminale Domäne (CTD) unterteilt (Boudvillain *et al.*, 2013; Ray-Soni *et al.*, 2016). Die CTD beinhaltet die Bindestellen zur Bindung der Rho-Untereinheiten zur Bildung des aktiven Rho-Hexamers. Außerdem enthält die CTD die ATP-abhängige Motorfunktion zur Translokation des Rho-Hexamers entlang einer RNA, sowie die sekundären RNA-Bindestellen (SBS), bezeichnet als Q- und R-Loop (Skordalakes und Berger, 2003). Die NTD von Ec.Rho besitzt eine primäre Bindestelle (PBS) zur Bindung des Rho-Faktors an RNA. Die PBS ermöglicht die Erkennung von *rut sites* auf der RNA, anhand deren die Transkripttermination eingeleitet wird (Alifano *et al.*, 1991; Koslover *et al.*, 2012). Die PBS bildet eine Spalte, die gerade groß genug ist, um ein 5'-YC-Dimer (Y und C stehen für einen Pyrimidin- und einen Cytosin-Rest) unterzubringen. Dies erklärt die Präferenz von Ec.Rho für einzelsträngige C-reiche *rut sites* (Skordalakes und Berger, 2003; Boudvillain *et al.*, 2013).

2.3.3 Verbreitung des Rho-Faktors und weitere Funktionen

Der Rho-Faktor ist ein Terminationsfaktor, der in Bakterien weit verbreitet ist, nicht aber in Eukaryoten vorkommt und daher ein interessantes Ziel für die Entwicklung von antibakteriellen Substanzen darstellt (D'Heygère *et al.*, 2013). Der Rho-Faktor kann spezifisch über Bicyclomycin (BCM) inhibiert werden (Tran *et al.*, 2011). Allerdings ist die Wirkung von BCM in grampositiven Bakterien sehr begrenzt, da bisher nur *Micrococcus luteus* als einziges grampositives Bakterium mit BCM-Sensitivität bekannt ist (Nowatzke *et al.*, 1997; Williams und Durham, 1988). Auch besitzen circa 8 % der bekannten bakteriellen Spezies kein Rho-Homolog, sodass diese ebenfalls nicht auf BCM sensitiv sein dürften. Die Spezies ohne Rho-Homolog gehören zu den Cyanobakterien, Tenericutes und Firmicutes (D'Heygère *et al.*, 2013).

Neben der Funktion als Transkriptionsterminator von mRNAs wurden für den Rho-Faktor weitere Funktionen in der Zelle nachgewiesen. So werden auch *antisense* RNAs spezifisch von Rho terminiert (Peters *et al.*, 2012). Darüber hinaus ist der Rho-Faktor besonders an der Stilllegung (*silencing*) der Transkription von fremder DNA, wie Prophagen-DNA oder horizontal transferierter DNA, beteiligt, indem die Transkripte dieser Gene vorzeitig terminiert werden (Cardinale *et al.*, 2008). Weiter konnte gezeigt werden, dass durch eine Integration von Phagen-DNA ins bakterielle Genom intrinsische Terminatoren zerstört werden können. Die Transkriptionstermination für diese Transkripte wird in solchen Fällen durch den Rho-Faktor übernommen (Peters *et al.*, 2009).

2.4 Prokaryotische Transkript-Degradation

Die Transkript-Degradation dient der Anpassung der Genexpression an wechselnde Umweltbedingungen und ist der wichtigste Mechanismus der posttranskriptionellen Kontrolle der Genexpression (Arraiano *et al.*, 2010; Laalami *et al.*, 2014; Durand und Condon, 2018). Weiter ermöglicht die Transkript-Degradation die Steuerung der Proteinmenge (im Fall von mRNA) und das Recycling von Ribonukleotiden zur Synthese neuer RNA (Laalami *et al.*, 2014). Aus diesen Gründen ist eine präzise Kontrolle der Transkript-Degradation notwendig (Laalami *et al.*, 2014). Die Regulation der Transkript-Degradation erfolgt über die Kontrolle der initialen Degradationsschritte durch eine RNase. Der initiale Schnitt einer RNase ist somit der geschwindigkeitsbestimmende Schritt. Zwischenprodukte der Degradation werden meist nicht beobachtet (Laalami *et al.*, 2014).

Die Transkript-Stabilität bezeichnet die Periode zwischen Synthese (Transkription) und Degradation des Transkripts. Die Transkript-Stabilität hängt von verschiedenen Faktoren, wie Translationseffizienz, RNA-Sequenz, Sekundärstrukturen und der Interaktion des Transkripts mit Proteinen und anderen RNAs, sowie der subzellulären Lokalisation ab (Petersen, 1992; Dreyfus, 2009; Laalami *et al.*, 2014). In Prokaryoten beträgt die Transkript-Stabilität oft wenige Minuten und korreliert nicht mit der Verdopplungszeit des jeweiligen Bakteriums (Laalami *et al.*, 2014).

RNasen haben einen großen Einfluss auf die Transkript-Stabilität, da sie die Degradation von Transkripten vermitteln bzw. katalysieren. Allerdings bedeuten nicht alle Schnitte von RNasen die Degradation eines Transkripts, sondern können auch prozessierende Funktionen haben. Als Beispiel ist hier die Prozessierung der 16S, 23S und 5S rRNA aus einem rRNA-Transkript über verschiedene Zwischenstufen genannt. An der Prozessierung von rRNA sind viele RNasen beteiligt, die auch an der Transkript-Degradation von anderen Transkripten, wie mRNA, involviert sind (Apirion und Lassar, 1978; Misra und Apirion, 1979; Mathy *et al.*, 2007). Ob ein Transkript durch Schnitte von einer RNase degradiert oder prozessiert wird, ist nicht immer direkt ersichtlich, da RNasen zum Teil an beiden Prozessen beteiligt sind (Arraiano *et al.*, 2010; Mackie, 2013).

2.4.1 Transkript-Degradation in gramnegativen und grampositiven Bakterien

Der Abbau von RNA durch RNasen erfolgt auf unterschiedliche Weisen. Generell wird zwischen den RNase-Ausstattungen von gramnegativen und grampositiven Bakterien unterschieden (Laalami *et al.*, 2014). Der Modellorganismus für die RNA-Degradation in gramnegativen Bakterien ist *Escherichia coli*. Für die RNA-Degradation in grampositiven ist der Modellorga-

2.4 Prokaryotische Transkript-Degradation

nismus *Bacillus subtilis* (Arraiano *et al.*, 2010; Silva *et al.*, 2011). Beide Organismen besitzen eine unterschiedliche RNase-Ausstattung von jeweils mehr als 20 RNasen, von denen circa die Hälfte für die Prozessierung von stabiler RNA (rRNA und tRNA) zuständig ist (Durand und Condon, 2018). In *E. coli* ist RNase E die zentrale RNase in der RNA-Degradation (Stead *et al.*, 2011), in *B. subtilis* sind es RNase Y und RNase III sowie die RNasen J1 und J2 (Durand *et al.*, 2012a).

In *E. coli* erfolgt die RNA-Degradation über RNase E in einem Membran-gebundenen Multienzymkomplex, dem Degradosom. Dieser Multienzymkomplex besteht aus einem RNase E-Tetramer sowie der RNA-Helicase RhlB, der 3'-5'-Exoribonuklease PNPase und dem glykolytischen Enzym Enolase, die jeweils an die CTD einer RNase E-Untereinheit gebunden sind (Callaghan *et al.*, 2004). Das Degradosom ermöglicht den Abbau von stark gefalteter RNA durch das Auflösen von Sekundärstrukturen durch die Helicase. Nach dem initialen Schnitt von RNase E wird das geschnittene Transkript schnell von der PNPase exoribonukleolytisch abgebaut. Die genaue Funktion der Enolase ist nicht geklärt, es wird aber vermutet, dass über die Enolase der Stoffwechselstatus der Zelle durch das Degradosom erkannt werden kann, um die RNA-Degradation entsprechend zu regulieren (Bandyra *et al.*, 2013; Ait-Bara und Carpousis, 2015). *E. coli* besitzt neben RNase E das Paralog RNase G. RNase G besitzt keine CTD zur Ausbildung eines Degradosoms und besteht im wesentlichen aus der katalytischen Domäne von RNase E (Arraiano *et al.*, 2010). Aufgrund der Ähnlichkeit der beiden RNasen werden sie in der RNase E/G-Familie zusammengefasst (Laalami *et al.*, 2014).

RNase Y in *B. subtilis* übernimmt funktional eine ähnliche Aufgabe in der RNA-Degradation wie RNase E in *E. coli* (Lehnik-Habrink *et al.*, 2011), besitzt aber auf Sequenzebene große Unterschiede (Shahbadian *et al.*, 2009). RNase Y besitzt keine strenge Präferenz für 5'-Monophosphat-Transkriptenden, sondern kann auch bestimmte Transkripte mit 5'-Triphosphaten spalten (Marincola und Wolz, 2017). Die Intermembrandomäne ermöglicht eine Verankerung von RNase Y in der Zellmembran. Weitere Domänen ermöglichen die Bindung von PNPase, Enolase und Phosphofruktokinase sowie der RNA-Helicase CshA (Commichau *et al.*, 2009; Lehnik-Habrink *et al.*, 2010; Newman *et al.*, 2012; Salvo *et al.*, 2016; Cascante-Esteva *et al.*, 2016). Aufgrund der geringen Co-Lokalisation und der schwachen Bindung der einzelnen Enzyme an RNase Y wird teilweise anstatt von einem Degradosom von einem transienten Degradosom-ähnlichen Netzwerk (*Degradosome-like network*) gesprochen (Durand und Condon, 2018). Die RNase III ist ebenfalls in der RNA-Degradation von *B. subtilis* beteiligt und schneidet in doppelsträngigen Bereichen von RNA (Durand *et al.*, 2012a,b). RNase III ist auch in *E. coli* vorhanden, spielt dort aber eine weniger wichtige Rolle in der Degradation des Hauptteils der RNA (Durand und Condon, 2018).

2 Theoretischer Hintergrund

Die RNasen J1 und dessen Paralog J2 wurden zuerst in *B. subtilis* identifiziert und sind bifunktionale Enzyme mit Endo- als auch 5'-3'-Exoribonuklease-Aktivität (Even *et al.*, 2005; Mathy *et al.*, 2010). Es ist allerdings umstritten, ob RNase J *in vivo* Endoribonuklease-Aktivität besitzt, da diese nicht eindeutig nachgewiesen werden konnte (Durand und Condon, 2018). Bisher wurde die Endoribonuklease-Aktivität nur *in vitro* nachgewiesen (Dorléans *et al.*, 2011; Newman *et al.*, 2011). Aufgrund der Enzymstruktur können *B. subtilis* RNase J1 bzw. J2 nur Transkripte mit 5'-Monophosphat- oder 5'-Hydroxylrest degradieren (Dorléans *et al.*, 2011; Mathy *et al.*, 2007; Li de la Sierra-Gallay *et al.*, 2008). RNase J besteht aus drei Domänen: einer (Metallo-Hydrolase-) β -Lactamase-Domäne zwischen die eine β -CASP-Domäne eingeschoben ist, sowie einer C-terminalen Domäne (Li de la Sierra-Gallay *et al.*, 2008; Newman *et al.*, 2011). β -Lactamasen bilden eine Superfamilie aus hydrolytischen Enzymen mit einer großen Zahl an biologischen Funktionen (Aravind, 1999; Daiyasu *et al.*, 2001). β -Lactamasen mit einer eingeschobenen β -CASP-Domäne werden der β -CASP-Familie zugeordnet, dessen Mitglieder in der DNA-Reparatur und RNA-Prozessierung/Degradation involviert sind. Die Bezeichnung CASP steht für die vier Vertreter CPSE, Artemis, SNM1 und PSO2 dieser Familie (Callebaut *et al.*, 2002; Dominski *et al.*, 2013). Das aktive Zentrum von RNase J befindet sich in der Spalte zwischen der β -Lactamase- und der β -CASP-Domäne. Die genaue Funktion der CTD ist nicht bekannt, führt aber bei Deletion der CTD zu einer reduzierten Enzymaktivität (Li de la Sierra-Gallay *et al.*, 2008). In *B. subtilis* bilden die beiden RNasen J1 und J2 ein Heterodimer, können aber auch als Monomere vorliegen. Das Monomer von RNase J2 ist um zwei Größenordnungen weniger aktiv als das Monomer aus J1 (Mathy *et al.*, 2010). Die RNasen J1 und J2 sind nicht am Degradosom-ähnlichen Netzwerk direkt beteiligt. Daher ist der RNA-Abbau in *B. subtilis* vermutlich komplexer als in *E. coli* (Lehnik-Habrink *et al.*, 2010, 2011).

Die RNasen E/G sind nicht auf gramnegative Bakterien beschränkt. So besitzen viele grampositive Actinobakterien ein Homolog von RNase E/G und RNase J, aber kein RNase Y-Homolog (Zeller *et al.*, 2007; Csanadi *et al.*, 2009; Taverniti *et al.*, 2011). Umgekehrt sind auch gramnegative Bakterien mit RNase Y-Homolog, wie beispielsweise *Helicobacter pylori*, bekannt (Redko *et al.*, 2013). In den gramnegativen Chlostridia wurden auch Kombinationen aus RNase E/G und RNase Y beobachtet (Ait-Bara und Carpousis, 2015).

Transkript-Degradation in *E. coli*

Trotz der unterschiedlichen RNase-Ausstattung von *E. coli* und *B. subtilis* erfolgt die Transkript-Degradation in beiden Organismen über vergleichbare Abläufe. In *E. coli* ist RNase E die wichtigste RNase für die Transkript-Degradation. RNase E ist in >95 % der γ -Proteobacteria vorhan-

2.4 Prokaryotische Transkript-Degradation

den (Durand und Condon, 2018) und hat eine Präferenz für 5'-Monophosphat-Transkriptenden (Mackie, 1998). AU-reiche einzelsträngige Regionen sind häufige Schnittstellen für diese endoribonukleolytische RNase (Mackie, 1991; Ehretsmann *et al.*, 1992; Kaberdin, 2003). Strukturell wird RNase E in eine N-terminale (NTD) und eine C-terminale Domäne (CTD) unterteilt. Die NTD beinhaltet das katalytische Zentrum, während die CTD hauptsächlich Bindestellen für die Enzyme des Degradosoms bereitstellt (Schubert *et al.*, 2004; Callaghan *et al.*, 2005).

Primärtranskripte besitzen ein 5'-Triphosphat, welche für die RNA-Degradation durch RNase E nicht zugänglich sind (Mackie, 1991; Ehretsmann *et al.*, 1992). Um die RNA-Degradation eines Primärtranskripts einzuleiten, wandelt die RNA-Pyrophosphohydrolase RppH das 5'-Triphosphat über Abspaltung eines Pyrophosphatrests zu einem 5'-Monophosphat um (Deana *et al.*, 2008; Richards *et al.*, 2011). Das monophosphorylierte Transkript wird im Folgenden von RNase E durch den initialen Schnitt für die Degradation „markiert“. Die nach dem Schnitt von RNase E entstandenen Fragmente werden über prozessive 3'-5'-Exoribonukleasen, wie PNPase und RNase II, in kleinere Fragmente bzw. in einzelne Nukleotide abgebaut. Neben dem initialen Schnitt von RNase E sind weitere sekundäre Schnitte möglich (Arraiano *et al.*, 2010; Laalami *et al.*, 2014).

Neben der RNase E besitzt *E. coli* das Paralog RNase G. RNase E und G schneiden in AU-reichen Regionen und präferieren monophosphorylierte 5'-Enden (Arraiano *et al.*, 2010). Beide RNasen sind in ähnliche Degradations- und Prozessierungsschritte involviert (Jiang *et al.*, 2000; Tock *et al.*, 2000). Die genauen Schnittstellen von RNase E und G sind aber nicht streng konserviert (Tock *et al.*, 2000). RNase G besteht im Wesentlichen aus der NTD von RNase E. Die CTD von RNase E fehlt RNase G, sodass RNase G kein Degradosom bilden kann (Jiang *et al.*, 2000; Tock *et al.*, 2000). Gewöhnlich besitzen RNase E-Homologe in anderen Spezies hohe Ähnlichkeit mit der katalytischen Domäne von RNase E und RNase G und werden daher gesammelt als RNase E/G bezeichnet (Condon und Putzer, 2002).

Transkript-Degradation in *B. subtilis*

In *B. subtilis* erfolgt der initiale Schnitt entweder durch RNase Y oder durch die Doppelstrang-sensitive RNase III (Abbildung 2.4A) (Durand *et al.*, 2012a). Durch den Schnitt wird ein 5'-Monophosphat-Ende am 3'-Fragment erzeugt, welches nun von RNase J exoribonukleolytisch vom 5'-Ende Richtung 3'-Ende verdaut wird. Das 5'-Fragment wird von den 3'-5'-Exoribonukleasen PNPase oder RNase R vom 3'-Ende verdaut. Alternativ kann auch wie bei *E. coli* das 5'-Triphosphat eines Transkripts in *B. subtilis* durch die RppH in ein 5'-Monophosphat umgewandelt werden, was RNase J direkt den Abbau vom ausgehend vom 5'-Ende ermöglicht

2 Theoretischer Hintergrund

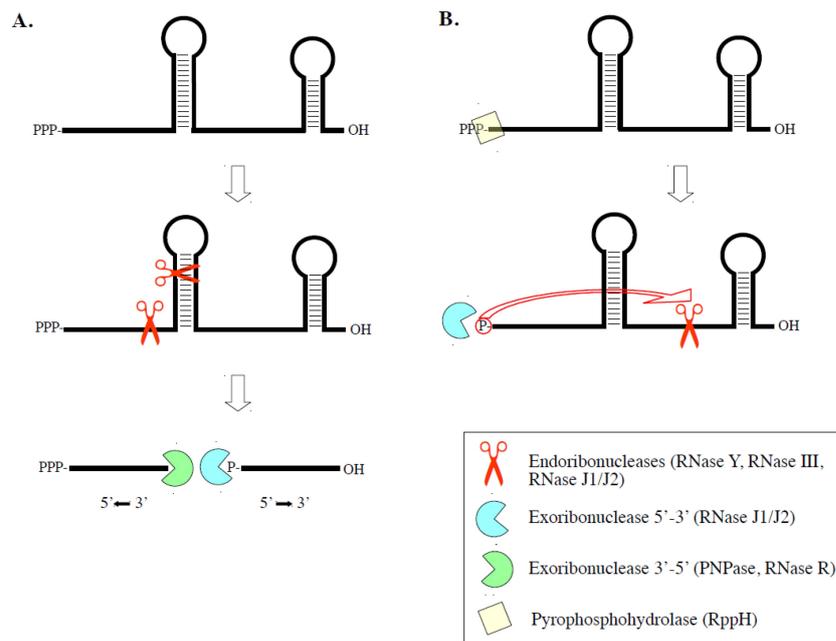


Abbildung 2.4: RNA-Degradation in grampositiven Bakterien am Beispiel von *B. subtilis*. (A) Primäre Transkripte sind am 5'-Ende durch ein Triphosphat geschützt. Die RNA-Degradation erfolgt über einen Schnitt einer Endoribonuklease (RNase Y oder RNase III). Dieser Schnitt ist der geschwindigkeitslimitierende Schritt der RNA-Degradation. Das durch den Schnitt entstandene 3'-Fragment besitzt ein 5'-Monophosphat welches von einer 5'-3'-Exoribonuklease (RNase J) angegriffen werden kann. Das 3'-Ende des 5'-Fragments kann durch eine 3'-5'-Exoribonuklease (PNPase oder RNase R) abgebaut werden. (B) In einem alternativen Abbau-Weg wird das 5'-Triphosphat der RNA von RppH in ein 5'-Monophosphat umgewandelt. Nach der Entfernung des 5'-Triphosphats kann das Transkript direkt von einer 5'-3'-Exoribonuklease abgebaut werden. In Fällen in denen RNase Y sensitiv auf den 5'-Phosphat-Status reagiert, kann RNase Y die Degradation des Transkripts durch einen initialen endoribonukleolytischen Schnitt einleiten. Bildquelle: Durand und Condon (2018). ©2018 American Society for Microbiology. Used with permission. No further reproduction or distribution is permitted without the prior written permission of American Society for Microbiology.

(Abbildung 2.4B). In manchen Fällen benötigt RNase Y ebenfalls ein 5'-Monophosphat um das Transkript zu degradieren (roter Pfeil in Abbildung 2.4B) (Laalami *et al.*, 2014; Durand und Condon, 2018). Alternativ kann RNase J1 auch als *sliding exoribonuclease* arbeiten, indem sie über das 5'-Triphosphat „rutscht“ und kurz hinter dem 5'-Transkriptende schneidet (Li de la Sierra-Gallay *et al.*, 2008).

2.4.2 Regulation der Transkript-Degradation

Die Transkript-Degradation bildet einen sehr effizienten Weg zur Kontrolle der Genexpression (Laalami *et al.*, 2014). Wie oben erwähnt, hängt die RNA-Stabilität und somit auch die

2.5 RNA-Sequenzierung zur Transkriptomanalyse

RNA-Degradation von vielen Faktoren ab. So verhindern translatierende Ribosomen oft die Bindung einer RNase an das Transkript und schützen es so vor dem Abbau (Arraiano *et al.*, 2010). Auch die Menge an RNasen in der Zelle beeinflusst die RNA-Stabilität. Das Transkript von RNase E in *E. coli* wird von RNase E selbst degradiert, um die Menge an aktiver RNase E in der Zelle zu regulieren und hält die RNase-Konzentration so in einem engen Rahmen konstant (Autoregulation) (Schuck *et al.*, 2009; Jain und Belasco, 1995). Eine vergleichbare Autoregulation wurde auch für PNPase (Carzaniga *et al.*, 2009; Jarrige *et al.*, 2001) und RNase III (Bardwell *et al.*, 1989; Matsunaga *et al.*, 1996) gezeigt.

Die komplementäre Bindung von kleinen regulativen RNAs (*small RNAs*, sRNAs) an eine Ziel-RNA erzeugt lokale Doppelstränge die eine potenzielle Schnittstelle für RNase III bilden. Alternativ können sRNAs eine Einzelstrang-sensitive RNase zu einer Schnittstelle „führen“, indem zuvor doppelsträngige Bereiche durch die benachbarte Bindung der sRNA aufgebrochen werden. Die RNase-Schnittstellen können dabei sowohl in der Nähe oder auch weit entfernt von der sRNA-Bindestelle liegen (Pfeiffer *et al.*, 2009; Prévost *et al.*, 2011). Das RNA-Bindeprotein Hfq ist in gramnegativen Bakterien für die Funktion und Stabilität der sRNAs von großer Bedeutung (Vogel und Luisi, 2011).

Riboswitches sind regulative Elemente in RNAs die über Sekundärstrukturen Metabolite binden und so die Konzentration eines Metaboliten sensieren (Breaker, 2011; Serganov und Nudler, 2013). Manche Riboswitches regulieren die Translationsinitiation der eigenen RNA durch die Verdeckung der RBS, sodass Ribosomen nicht binden können und die RNA so für RNasen angreifbar wird. Ein Beispiel ist hier der *glmS* Riboswitch in *B. subtilis*. Das Gen *glmS* kodiert für eine Glucosamin-6-Phosphat-(GlcN6P)-Synthase. Bei Überschuss an GlcN6P bindet der *glmS* Riboswitch an GlcN6P und induziert eine autokatalytische Spaltung im eigenen Transkript. Das dabei entstandene 5'-Hydroxyl wird von RNase J erkannt und das Transkript degradiert (Winkler *et al.*, 2004; Collins *et al.*, 2007).

2.5 RNA-Sequenzierung zur Transkriptomanalyse

Die Methoden zur Transkript- und Transkriptomanalyse sind vielfältig. Northern Blots (Alwine *et al.*, 1977), qRT-PCR (quantitative Reverse Transkriptase PCR) (Kubista *et al.*, 2006), 5'/3'-RACE (*Rapid Amplification of cDNA Ends*) (Frohman, 1994) und Microarrays (Taub *et al.*, 1983) sind für die Analyse von Transkripten und deren Abundanzen geeignet. Der Nachteil dieser Methoden besteht meist in der geringen Zahl parallel untersuchbarer Transkripte, was die Analyse ganzer Transkriptome sehr erschwert. Microarrays sind für die Hochdurchsatz-Analyse von Transkripten ausgelegt, bieten aber wenig Information über die jeweiligen Transkriptsequenzen (Chen *et al.*, 2006a).

2 Theoretischer Hintergrund

Die Methode der RNA-Sequenzierung (oder auch Transkript- oder cDNA-Sequenzierung) löst einen Großteil der Nachteile der erwähnten Methoden. Die RNA-Sequenzierung ermöglicht die Erfassung von Transkripten im Hochdurchsatz mit Einzelnukleotid-Auflösung. Parallel werden die Transkriptabundanzen über einen hohen dynamischen Bereich ermittelt. Auch die Identifizierung von neuen bzw. bisher unbekanntem Transkripten ist mit RNA-Sequenzierung möglich. Diese Vorteile bilden eine ideale Grundlage für die Analyse ganzer Transkriptome und finden daher bei Eukaryoten als auch Prokaryoten breite Anwendung (Mortazavi *et al.*, 2008; Lister *et al.*, 2008; Nagalakshmi *et al.*, 2008; Perkins *et al.*, 2009).

Mittels RNA-Sequenzierung wurde die unerwartet hohe Komplexität von bakteriellen Transkriptomen aufgedeckt und besonders viele neue kleine und *antisense* RNAs identifiziert (Wang *et al.*, 2009; Dornenburg *et al.*, 2010). Darüber hinaus können die bei der Sequenzierung von Transkriptomen gewonnenen Daten für die Korrektur und Verbesserung der Genomannotation eingesetzt werden (Denoëud *et al.*, 2008; Raghavan *et al.*, 2011). Neben der Analyse von Transkriptabundanzen über die Sequenzierung von Vollängentranskripten, welche für die Untersuchung der Genexpression genutzt werden (Mortazavi *et al.*, 2008; Wang *et al.*, 2009), wurden alternative Protokolle entwickelt, die auf spezifische Eigenschaften von Transkripten abzielen. So existieren unter anderem Methoden zur spezifischen Anreicherung von Primärtranskripten zur Analyse von Transkriptionsstarts (Sharma *et al.*, 2010; Pfeifer-Sancar *et al.*, 2013) oder der Charakterisierung von Sekundärstrukturen (Kertesz *et al.*, 2010). Die in dieser Arbeit verwendeten Methoden der RNA-Sequenzierung sind im Methodenteil in Abschnitt 3.12 ab Seite 61 umfassend beschrieben.

Aufgrund der enormen Vorteile ist die RNA-Sequenzierung die ideale Methode zur Analyse der Transkriptome von Corynebakterien hinsichtlich Transkriptionsinitiation, -termination und Transkript-Degradation. Die „klassischen“ Methoden zur Transkriptanalyse, wie Northern Blots und qRT-PCR sind allerdings nicht durch die Entwicklung der RNA-Sequenzierung obsolet geworden, sondern bieten vielmehr eine Möglichkeit die mittels RNA-Sequenzierung erhaltenen Ergebnisse zu validieren.

Aufbau und Funktion von Typ IV Pilussystemen

In den Daten der *C. glutamicum* $\Delta\rho$ -Mutante sind Gene eines Typ IV Pilussystems stark differenziell transkribiert. Daher wird hier vorgreifend auf die Struktur und Funktion von diesen Pilussystemen eingegangen.

Typ IV Pili sind komplexe, in gramnegativen wie auch in grampositiven Bakterien weit verbreitete, Pilusstrukturen (Tomich *et al.*, 2007). Generell wird zwischen Typ IVa und Typ IVb Pili anhand der Länge und Sequenz der *leader*-Peptide unterschieden. Typ IVa Pili sind

2.5 RNA-Sequenzierung zur Transkriptomanalyse

eine weitgehend homogene Gruppe mit vergleichsweise kurzen *leader*-Peptidsequenzen der Haupt-Pilinproteine von 6 bis 7 Aminosäuren. Typ IVb Pili hingegen bilden eine heterogene Gruppe und besitzen *leader*-Peptidsequenzen in den Haupt-Pilinproteinen von 15 bis 30 Aminosäuren (Berry und Pelicic, 2015). Die Unterscheidung anhand der *leader*-Peptidsequenzen ist allerdings ungenau, da Ausnahmen existieren und keine exakte Grenze definiert ist (Giltner *et al.*, 2012). Es wird daher empfohlen eine Einteilung anhand der charakteristischen Assemblierungssysteme vorzunehmen (Ayers *et al.*, 2010).

Der Typ IVb Pilus in *Aggregatibacter actinomycetemcomitans* vermittelt die feste Adhäsion des Bakteriums an Oberflächen und eukaryotische Zellen und wird daher auch Tad-Pilus (*Tight adherence*-Pilus) genannt (Tomich *et al.*, 2007). Neben der Bezeichnung Tad-Pilus wird auch oft Flp-Pilus (*fimbrial low molecular protein*-Pilus) verwendet (Giltner *et al.*, 2012; Berry und Pelicic, 2015). Der Tad-Pilus ist eine Untergruppe der Typ IVb Pili für den das Bakterium *A. actinomycetemcomitans* einen Modellorganismus darstellt (Tomich *et al.*, 2007). Die Namensgebung der für den Tad-Pilus kodierenden Gene ist nicht einheitlich. So werden die für den Tad-Pilus kodierenden Gen-Cluster und entsprechenden Proteine in manchen Bakterien als *tad*-, *comP*-, *pil*- oder *flp*-Cluster bezeichnet (Tomich *et al.*, 2007; Giltner *et al.*, 2012; Melville und Craig, 2013).

Auf Basis von vergleichenden Genomanalysen sowie Analysen der *leader*-Sequenzen der Prepilin-Proteine, wird der Typ IV Pilus in *C. glutamicum* der Gruppe der Typ IVb Pilussysteme zugeordnet (Imam *et al.*, 2011). In dieser Arbeit wird das den Typ IVb Pilus kodierende Gencluster entsprechend als *tad*-Cluster bezeichnet.

Typ IV Pili ähneln Typ II und Typ IV Sekretionssystemen sowie den Kompetenzpseudopili und den in Archaea vorkommenden Flagellen (Tomich *et al.*, 2007; Berry und Pelicic, 2015). Der grundsätzliche Aufbau von Typ IV Pilussystemen ist in Abbildung 2.5 für gramnegative wie auch für grampositive Bakterien dargestellt. Um die Ähnlichkeit zwischen Typ IV Pilus und Typ II Sekretionssystemen zu verdeutlichen, ist ein Typ II Sekretionssystem ebenfalls in der Abbildung gezeigt. Der Typ IV Pilus besteht aus Pili-Untereinheiten und ragt in gramnegativen Bakterien von der inneren über die äußere Membran weit aus der Zelle hinaus. In grampositiven Bakterien ragt der Pilus von der inneren Membran durch die Peptidoglykanschicht und ebenfalls über die Zelle hinaus. In der inneren Membran von gramnegativen als auch grampositiven befindet sich eine ATPase mit Ankerproteinen, die zusammen eine Assemblierungs-Plattform bilden, an der der Pilus aufgebaut und verankert wird. Vor dem Einbau in den Pilus werden von den Prepilin-Untereinheiten die *leader*-Peptide durch eine membrangebundene Prepilin-Peptidase entfernt. In gramnegativen ragt der Pilus nicht direkt durch das Periplasma, sondern wird von dem Hüllprotein Secretin umhüllt. Das in Abbildung 2.5 für grampositive Bakterien gezeigte Hüllprotein, welches den Pilus von der

2 Theoretischer Hintergrund

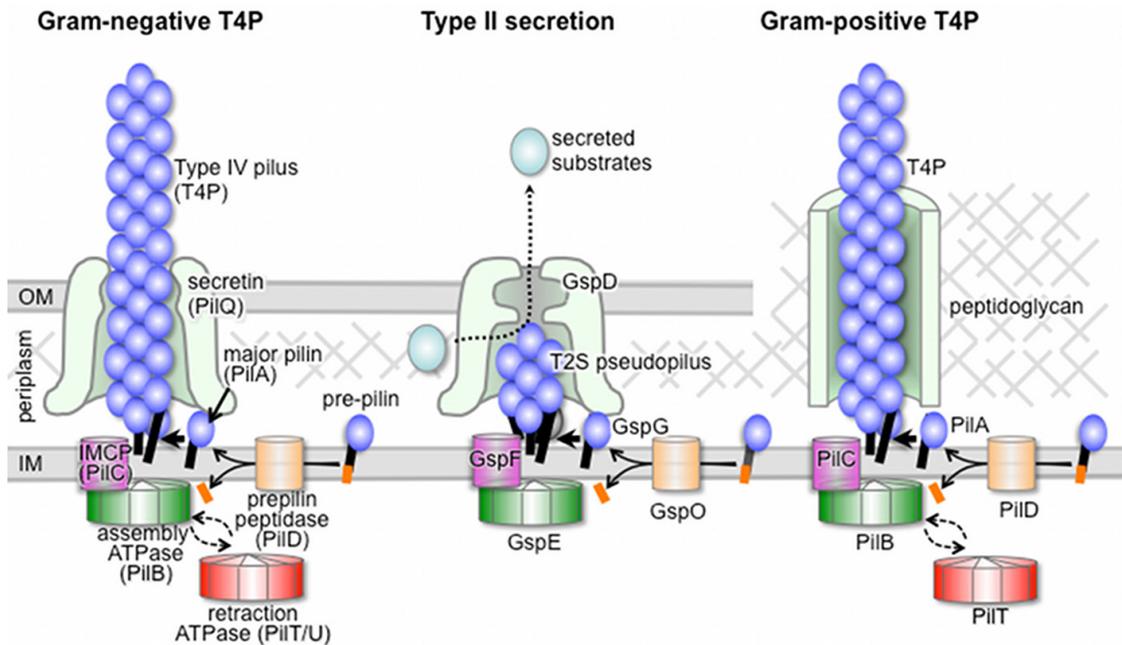


Abbildung 2.5: Darstellung der Hauptkomponenten des Typ IV Pilussystems und des Typ II Sekretionsystems in gramnegativen und des Typ IV Pilussystems in grampositiven Bakterien. Die Komponenten sind nach den gramnegativen Typ IV Pilussystemen benannt. Gezeigt sind das *inner membrane core protein* ICMP, die für den Aufbau bzw. Abbau nötigen ATPasen, Pili-Untereinheiten und das Sekretin in gramnegativen bzw. ein hypothetisches Hüllprotein in grampositiven Bakterien. IM, *inner membrane*; OM, *outer membrane*. Bildquelle: Melville und Craig (2013). Abgedruckt mit freundlicher Genehmigung von *Microbiology and Molecular Biology Reviews*.

Peptidoglykanschicht trennt, ist hypothetisch. Den Abbau bzw. den Rückzug des Pilus in die Zelle übernimmt eine weitere ATPase (Tomich *et al.*, 2007; Melville und Craig, 2013; Berry und Pelicic, 2015; Piepenbrink und Sundberg, 2016).

Die Funktionen von Typ IV Pili sind vielfältig (Giltner *et al.*, 2012). Sie vermitteln die Adhäsion an Oberflächen bzw. eukaryotischen Zellen (Tomich *et al.*, 2007; Motherway *et al.*, 2011), ermöglichen die Aufnahme von DNA (Chen *et al.*, 2006b), dienen der Fortbewegung (Piepenbrink und Sundberg, 2016) oder fungieren als Nanokabel für den Elektronentransport zwischen Zellen (Reguera *et al.*, 2005).

2.6 Ziele der Arbeit

Das Transkriptom umfasst alle RNA-Transkripte einer Zelle zu einem gegebenen Zeitpunkt. Mit Hilfe der neuen Technologie der RNA-Sequenzierung ist es möglich, das Transkriptom

in seiner Gesamtheit zu analysieren und ermöglicht völlig neue Einblicke in die bakterielle Transkription.

Ziel dieser Arbeit war es unter Anwendung der RNA-Sequenzierung neue Erkenntnisse in den medizinisch bzw. biotechnologisch relevanten Organismen *C. diphtheriae* und *C. glutamicum* hinsichtlich der RNA-Synthese und des RNA-Abbaus zu erlangen. Dazu wurden drei Ziele formuliert:

Das Ziel bezüglich der RNA-Synthese bestand in der Analyse der Transkriptionsinitiation in *C. diphtheriae* anhand von Transkriptionsstartpunkten. Daneben wurde außerdem das Regulon des Transkriptionsregulators DtxR charakterisiert. Das zweite Ziel war Charakterisierung der Transkriptionstermination anhand der Identifizierung und Analyse von Rho-abhängig terminierten Transkripten in *C. glutamicum*. Das dritte Ziel umfasste die Charakterisierung der in der Transkript-Degradation beteiligten RNasen E/G und J und somit einen Teil des RNA-Abbaus in *C. glutamicum*.

Da alle drei Ziele grundlegend mit der Transkription verknüpft sind, wurde für die Beantwortung der Ziele auf die Methode der RNA-Sequenzierung gesetzt. Falls für die Beantwortung spezifischer Fragestellungen die Entwicklung und Erprobung neuer Analysewerkzeuge und Sequenzierprotokolle erforderlich war, gehörte dies ebenfalls zu der Zielsetzung dieser Arbeit.

2.7 Bisherige Publikationen

Teile dieser Arbeit wurden bereits in folgender Publikation veröffentlicht:

Wittchen, Manuel; Busche, Tobias; Gaspar, Andrew H.; Lee, Ju Huck; Ton-That, Hung; Kalinowski, Jörn and Tauch, Andreas. *Transcriptome sequencing of the human pathogen Corynebacterium diphtheriae NCTC 13129 provides detailed insights into its transcriptional landscape and into DtxR-mediated transcriptional regulation*. BMC Genomics 2018 **19**:82. <https://doi.org/10.1186/s12864-018-4481-8>

3 Material und Methoden

Im Folgenden werden die verwendeten Materialien und Methoden aufgeführt, die im Rahmen dieser Arbeit benötigt und verwendet wurden. Bei Lösungen sind im Allgemeinen alle Mengen- oder Konzentrationsangaben auf einen Liter bezogen. Wenn nicht anders angegeben, wurde als Lösungsmittel reines Wasser, welches aus einer MilliQ-Anlage von Millipore (Merck) gewonnen wurde, verwendet.

3.1 Bakterienstämme, Plasmide und Primer

Die Plasmidkarten befinden sich im Anhang.

Tabelle 3.1: Liste der verwendeten Stämme von *Corynebacterium diphtheriae* und *Corynebacterium glutamicum*. Alle Stämme und Mutanten von *C. glutamicum* basieren auf dem Wildtyp ATCC 13032. Abkürzungen: Bs., *Bacillus subtilis*.

<i>Corynebacterium diphtheriae</i>		
Stamm	Relevante Eigenschaft	Referenz
NCTC 13129	Wildtyp	NCTC, Cerdeño-Tárraga <i>et al.</i> (2003)
$\Delta dtxR$	$\Delta DIP1414 (\Delta dtxR)$	Andrew H. Gaspar
<i>Corynebacterium glutamicum</i>		
Stamm	Relevante Eigenschaft	Referenz
ATCC 13032	Wildtyp; Nx ^R	ATCC, Kinoshita <i>et al.</i> (1958)
Δrho	$\Delta cg1354 (\Delta rho)$	Diese Arbeit
Δrho pEC-rho	$\Delta cg1354$ pEC-XC99E:: <i>Cg.rho</i> ; Cm ^R	Diese Arbeit
pEC-XC99E	Leervektor; Cm ^R	Diese Arbeit
pRIM2	Genomische Integration des Plasmids pRIM2; Km ^R	Diese Arbeit
Δrho pRIM2: <i>rho</i>	$\Delta cg1354$ pRIM2:: $\Delta cg1354$; Km ^R	Diese Arbeit
Δrho pRIM2	$\Delta cg1354$ pRIM2; Km ^R	Diese Arbeit
Δrne	$\Delta cg2597 (\Delta rne)$	Diese Arbeit
Δrnj	$\Delta cg2160 (\Delta rnj)$	Diese Arbeit
Δrnj pEC-rnjA	$\Delta cg2160$ pEC-XC99E:: <i>Bs.rnjA</i> ; Cm ^R	Diese Arbeit
Δrnj pEC-rnjB	$\Delta cg2160$ pEC-XC99E:: <i>Bs.rnjB</i> ; Cm ^R	Diese Arbeit
<i>tadOE</i>	pEC-XC99E:: <i>cg0362-cg0369</i> (<i>tad</i> -Cluster); Cm ^R	Diese Arbeit
<i>tadOE</i> pZMP	pEC-XC99E:: <i>cg0362-cg0369</i> ; Cm ^R und pZMP; Km ^R	Diese Arbeit

3 Material und Methoden

Tabelle 3.2: Liste der verwendeten Stämme von *E. coli*.

Stamm	Relevante Eigenschaft	Referenz
S17-1 DH5αMCR	Konjugationsstamm <i>F-endA1 supE44 thi-l λ- re-ca1 deoR Δ(lacZYA - argF) U169 φ80dlacZλM15 mcrA Δ(mrr-hsdRMS-mcrBC)</i>	ATCC 47055 Grant <i>et al.</i> (1990)
ER2566	Chromosomal kodierte, mit IPTG induzierbare T7 RNA-Polymerase, reduzierte Protease-Ausstattung,	New England Biolabs
S17-1 pK18 <i>mobsacB</i> :: Δ <i>dtxR</i>	Deletionsplasmid für DIP1414 / <i>dtxR</i>	Andrew H. Gaspar, Wittchen <i>et al.</i> (2018)
DH5αMCR pK18 <i>mobsacB</i> :: Δ <i>rho</i>	Deletionsplasmid für <i>cg1354 / rho</i>	Diese Arbeit
DH5αMCR pK18 <i>mobsacB</i> :: Δ <i>rne</i>	Deletionsplasmid für <i>cg2597 / rne</i>	Diese Arbeit
DH5αMCR pK18 <i>mobsacB</i> :: Δ <i>rnj</i>	Deletionsplasmid für <i>cg2160 / rnj</i>	Diese Arbeit
DH5αMCR pRIM2:: <i>rho</i>	Plasmid für Geninsertion von <i>C. glutamicum rho</i> inkl. nativem Promotor und nativem Haarnadel-Terminator des Operons	Diese Arbeit
DH5αMCR pEC-XC99E:: <i>tadOE</i>	Expressionsplasmid für <i>tad</i> -Cluster (<i>cg0362 - cg0369</i>)	Diese Arbeit
DH5αMCR pEC-XC99E::Bs. <i>rnjA</i>	Expressionsplasmid für <i>B. subtilis</i> RNase J1	Diese Arbeit
DH5αMCR pEC-XC99E::Bs. <i>rnjB</i>	Expressionsplasmid für <i>B. subtilis</i> RNase J2	Diese Arbeit
ER2566 pTXB1:: <i>rne</i>	Expressionsplasmid für Proteinaufreinigung von <i>C. glutamicum</i> RNase E/G, Amp ^R	Diese Arbeit
ER2566 pTXB1:: <i>rnj</i>	Expressionsplasmid für Proteinaufreinigung von <i>C. glutamicum</i> RNase J, Amp ^R	Diese Arbeit

Tabelle 3.3: Für Klonierungen, Northern Blot und qRT-PCR verwendete Primer. Diese Liste befindet sich auch auf der beiliegenden Daten-CD (Tabelle S1). Das Zeichen \leftrightarrow zeigt einen Zeilenumbruch an. LF, Linke Flanke; RF, Rechte Flanke; Reskl., Restriktionsklonierung; GA, Gibson Assembly; Plasmidamp., Plasmidamplifikation; Del., Deletion; Überexpr., Überexpression.

Verwendungszweck	Name	Sequenz (5' - 3')
Primer für Deletionen (pK18<i>mobsacB</i>-System)		
Del. von <i>dtxR</i> in <i>C. diphtheriae</i> LF (Reskl.)	<i>dtxR</i> -A	CGGGATCCGGAAGAATCTGGCATTGAAGA
Del. von <i>dtxR</i> LF (Reskl.)	<i>dtxR</i> -B	CCCATCCACTAAACTTAAACAATCGACTAAGTCCTTCATTGTA
Del. von <i>dtxR</i> RF (Reskl.)	<i>dtxR</i> -C	TGTTTAAGTTTAGTGGATGGGGGCCACATCACGTTGAGC
Del. von <i>dtxR</i> RF (Reskl.)	<i>dtxR</i> -D	CGGGATCCCGCAGACTCGTCGCCGC
Del. von <i>rho</i> (GA) LF (Reskl.)	<i>rho</i> -d1	GGTGGTCTCGAATTCACGCAACGACAGTGC AAGC
Del. von <i>rho</i> (GA) LF (Reskl.)	<i>rho</i> -d2	AGATTTCCTTTTCATTGCAGTTTC
Del. von <i>rho</i> (GA) RF (Reskl.)	<i>rho</i> -d3	GAACTGCAATGAAAGGAAATCTACTCGTTGGGCCGTGCGTAC
Del. von <i>rho</i> (GA) RF (Reskl.)	<i>rho</i> -d4	GGTGGTCTCGGATCGCCATTAGGAGTAATCCTCCTC
Test-Primer Del. von <i>rho</i>	MW95	TGGGTGCCACGTTCAATTAC
Test-Primer Del. von <i>rho</i>	MW96	CGAGTAGCGCTTACCTACCC
Plasmidamp. für Del.en (GA)	MW66	GGTACCCGGGGATCCTCTAG
Plasmidamp. für Del.en (GA)	MW67	GAGCTCGAATTCGTAATCATG
Del. von <i>rne</i> (GA) LF	W281	TTTAAGAAGGAGATATACATATGCCAAATAACAAGGCAGT
Del. von <i>rne</i> (GA) LF	W282	AGTGCATCTCCCGTGTGCAGTATCGTGGCCCTCGTCGGGTAG
Del. von <i>rne</i> (GA) RF	W187	ACCATCGGGTGTCTCCGACGCCAGCCCTGCTCAATGCGAC
Del. von <i>rne</i> (GA) RF	W277	GCTTCTACGTGTTCCCGTTCCGCAACCTACTGACTTCGAGGGCCG
Test-Primer Del. von <i>rne</i>	MW23	TCACCTGGGTGGAACCTTG
Test-Primer Del. von <i>rne</i>	MW24	GACAGCGCACTAATTACAG
Del. von <i>rnj</i> (GA) LF	<i>rnj</i> -d1	ACTCTAGAAATCGCGGCCGGAAGGTTAC
Del. von <i>rnj</i> (GA) LF	<i>rnj</i> -d2	AGGTGAAACGGGATGCCAAG
Del. von <i>rnj</i> (GA) RF	<i>rnj</i> -d3	CTTGGCATCCGTTTCACCTCAGATCACTGTCCGTGATGG
Del. von <i>rnj</i> (GA) RF	<i>rnj</i> -d4	CTTCTAGATTACAGGGACTCGCGTGAAAG
Test-Primer Del. von <i>rnj</i>	MW85	CCTCATATGGTCCGGTAAAG
Test-Primer Del. von <i>rnj</i>	MW86	GGGACTTGCCTGGTATTC
Primer für Überexpressionen (pEC-XC99E-System)		

3.1 Bakterienstämme, Plasmide und Primer

Tabelle 3.3: Fortgesetzt

Verwendungszweck	Name	Sequenz (5' - 3')
Plasmidamplifikation von pEC-XC99E (GA)	pEC-for	TTTGGCCCGACATCATAACG
Plasmidamplifikation von pEC-XC99E (GA)	pEC-rev	GGCCGATTTGTCTACTCAG
Überexpr. von <i>rho</i> (GA)	rho-ov-1	AACCCCTATAAAAGGAGGACAACCGTGACAACCACAGACAACACC
Überexpr. von <i>rho</i> (GA)	rho-ov-2	CCTGGCAGTCCCTACTCTCGCATGTTAGGAGTAATCCTCCTTT
Überexpr. von <i>Bs.rnjA</i> (GA)	W111	CACAGGAAGGAGAATAGACAATGAAATTTGTAAAAATGA
Überexpr. von <i>Bs.rnjA</i> (GA)	W112	CCACACTACCATCGGCGCTATTAAACCTCCATAATGATCG
Überexpr. von <i>Bs.rnjB</i> (GA)	W113	CACAGGAAGGAGAATAGACAATGAAAAAGAAAAATACAGAAAAAC
Überexpr. von <i>Bs.rnjB</i> (GA)	W114	CCACACTACCATCGGCGCTATTACTTCCATAATAATTG
Überexpr. von <i>tadOE</i> (cg0362 – cg0369) (GA)	W243	CACAGGAAGGAGAATAGACAATGAAAGCACGTAAGCCTCAG
Überexpr. von <i>tadOE</i> (cg0362 – cg0369) (GA)	W242	CCACACTACCATCGGCGCTATCATAAAGGCTCTGCTTTTG
Primer für Insertion (PRIM2-System)		
Insertion von <i>rho</i> in PRIM2 Restkl.	W316	AGCTGGGTACCCGGGGATCCTCTAGGCAGCTTTTAGCGAGGGGAATGAGA
Insertion von <i>rho</i> in PRIM2 Restkl.	W314	GCCTCGAGGTGCGACTCTAGCGCCCCACTCCCTGTTTCAAGGGGTGTGG← GGCGTTGAAATTAGGAGTAATCCTCCTCT
Primer für Proteinaufreinigung (IMPACT-System)		
Amplifikation von <i>rnj</i> (GA)	MW31	AGAAGGAGATATACATATGA
Amplifikation von <i>rnj</i> (GA)	MW32	TCGACTCGCGAGCTAGCCAT
Amplifikation von pTXB1 mit Überhängen zu <i>rnj</i> (GA)	MW33	CGCCTTACGCGAGTCCCTG
Amplifikation von pTXB1 mit Überhängen zu <i>rnj</i> (GA)	MW34	CGATTTGGGGAATCATTCAAT
Amplifikation pTXB1 (GA)	MW21	ATGGCTAGCTCGCGAGTCGAC
Amplifikation pTXB1 (GA)	MW22	TCATATGTATATCTCCTTCTTAAAG
Aufreinigung von <i>rne</i> (GA)	W274	CTTTAAGAAGGAGATATACATATGAGTGCCAAATAACAAGGCAGTAG
Aufreinigung von <i>rne</i> (GA)	W275	CGCCGTCGACTCGCGAGCTAGCCATGTATCGTGGGCCCTCGTCGGGTAG
Primer für Northern Blot		
Sonde 6C RNA	N009	CAAGGCCCGATATACAGTG
Sonde 6C RNA	N010	TAATACGACTCACTATAGGGCTAAGGGCACTAGTTTAGAA
Sonde 5S rRNA	N021	TGTGTCGGTGGTAGTAGTAG
Sonde 5S rRNA	N022	TAATACGACTCACTATAGGGGTCGGGGTAACCTACTCTC
Primer für qRT-PCR		
cg1478	q097	ATGGGAGCCAGTTACTAACC
cg1478	q098	CGTACTTCGGACCAATACTGAC
cg1478 Übergang	q127	AAAAACGGCTAACCCAAAAG
cg1478 Übergang	q128	CACGCATGGCTGGATCAAC
cg1478 Verlängerung	q099	AATCGCCAAAGCAGTAGTAG
cg1478 Verlängerung	q100	CCAACGTAGAGATCGTGGATTCTG
cg0369	q079	TGGCCAGGGAACAAGCACAG
cg0369	q080	GTGGGCTTCTGTCCACGAACG
cg0369 Übergang	q129	TCCAACGCCAACGGATCAAT
cg0369 Übergang	q130	TTGGAGGACGAGGTTTTTGA
cg0369 Verlängerung	q131	GCCGAATCAGGACTCAGCTC
cg0369 Verlängerung	q132	TTCCGGGTCATGTTTATTGC
6C RNA	q173	CCCCGATATACAGTGTGGC
6C RNA	q174	CCCCGCCCTCGGGGGGGGTG
6C RNA Übergang	q177	CGTTGCCGGGTTACTGACG
6C RNA Übergang	q178	TTCACTTTCGAATGCAGAGG
6C RNA Verlängerung	q179	TTATTTTCTAACTAGTGCCCTT
6C RNA Verlängerung	q180	TGTGGATAACTCACCTCTCAAC
ascg0066	q149	GGTAAATCTTGCAAAAAGTAG
ascg0066	q150	TAGTCTTCCGCATGTACAG
ascg0066 Übergang	q151	GGCGATGTCAGGTTTTCTCAC
ascg0066 Übergang	q152	ACACCTCCGTGCAGTACATC
ascg0066 Verlängerung	q153	GGAGACGCCGAAAATACGTG
ascg0066 Verlängerung	q154	GGTGCTCACGGGAGTTTCAA
ascg0511	q155	CACCGATGACTACGGCACTG
ascg0511	q156	ACGACCACGTCGGTACTAT
ascg0511 Übergang	q157	ACAGGAGGCTGTTTTTCATT
ascg0511 Übergang	q158	GTCACTCGGTGATGACGCC
ascg0511 Verlängerung	q159	ACCACCAAGCGCTGATATCC
ascg0511 Verlängerung	q160	GATCTCTGGCATGACATCAA
ascg0700	q093	GATGTCGCTGCGGCACGTCGTGATTAC
ascg0700	q094	CACAGCATCTGCGCCACAGGCAATAG
ascg0700 Übergang	q163	AGGTGATTGCTGCGTATGAG
ascg0700 Übergang	q164	ATCTCACGAACATTTTGTGG
ascg0700 Verlängerung	q095	AGCAGCACGGCGATCCATTGTTTTTTAG
ascg0700 Verlängerung	q096	GTCCGCTGCAGTGTGATGGT
cg0362	q067	CAGTCAACCACTGTGCTTATC
cg0362	q068	TAGACAACACCACGACCTCATC
cg0365	q073	GCTGACCCAGTGGATGTCTC
cg0365	q074	GCACTACTCCAATCGATAG

3 Material und Methoden

Tabelle 3.4: Liste der verwendeten Plasmide.

Plasmid	Beschreibung	Referenz
pK18 <i>mobsacB</i>	Plasmid für die Gendeletion / Genaustausch in <i>C. diphtheriae</i> und <i>C. glutamicum</i> , Km ^R	Schäfer <i>et al.</i> (1994)
pK18 <i>mobsacB</i> :: Δ <i>dtxR</i>	Deletionsplasmid für DIP1414 / <i>dtxR</i>	Andrew H. Gaspar, Wittchen <i>et al.</i> (2018)
pK18 <i>mobsacB</i> :: Δ <i>rho</i>	Deletionsplasmid für <i>cg1354</i> / <i>rho</i>	Diese Arbeit
pK18 <i>mobsacB</i> :: Δ <i>rne</i>	Deletionsplasmid für <i>cg2597</i> / <i>rne</i>	Diese Arbeit
pK18 <i>mobsacB</i> :: Δ <i>rnj</i>	Deletionsplasmid für <i>cg2160</i> / <i>rnj</i>	Diese Arbeit
pRIM2	Plasmid für Geninsertion, Km ^R	Vašicová <i>et al.</i> (1998)
pRIM2:: <i>rho</i>	Plasmid für Geninsertion von <i>C. glutamicum rho</i> inkl. nativem Promotor und nativem Haarnadel-Terminator des Operons	Diese Arbeit
pEC-XC99E	Shuttle-Expressionsplasmid, Cm ^R	Kirchner und Tauch (2003)
pEC-XC99E:: <i>tadOE</i>	Expressionsplasmid für <i>tad</i> -Cluster (<i>cg0362</i> - <i>cg0369</i>)	Diese Arbeit
pEC-XC99E:: <i>rho</i>	Expressionsplasmid für <i>rho</i>	Diese Arbeit
pEC-XC99E::Bs. <i>rnjA</i>	Expressionsplasmid für <i>B. subtilis</i> RNase J1	Diese Arbeit
pEC-XC99E::Bs. <i>rnjB</i>	Expressionsplasmid für <i>B. subtilis</i> RNase J2	Diese Arbeit
pZMP	Shuttle-Expressionsplasmid, Km ^R	Walter <i>et al.</i> (2016)
pTXB1	Expressionsplasmid für Proteinaufreinigung über das IMPACT-System, C-terminale Fusion, T7-Promotor, Amp ^R	New England Biolabs
pTXB1:: <i>rne</i>	Expressionsplasmid für Proteinaufreinigung von <i>C. glutamicum</i> RNase E/G, Amp ^R	Diese Arbeit
pTXB1:: <i>rnj</i>	Expressionsplasmid für Proteinaufreinigung von <i>C. glutamicum</i> RNase J, Amp ^R	Diese Arbeit

3.2 Verwendete Chemikalien, Materialien, Geräte und Software

Im Folgenden sind alle verwendeten Chemikalien, Materialien und Kits sowie deren Bezugsquellen als auch Geräte und Software aufgeführt.

Tabelle 3.5: Verwendete Chemikalien und deren Bezugsquelle.

Substanz	Bezugsquelle
(d)ATP, (d)GTP, (d)TTP, (d)CTP, UTP	Thermo Fisher
Acrylamid / Bisacrylamid-Lösung	Roth
Agar	Invitrogen
Agarose	Peqlab
Ammoniumsulfat	Merck
Antibiotika	Sigma
APS	Sigma

3.2 Verwendete Chemikalien, Materialien, Geräte und Software

Tabelle 3.5: Fortgesetzt

Substanz	Bezugsquelle
Biotin	Sigma Aldrich
Blocking-Pulver	Roche
Blocking-Reagenz	Roche
Brain-Heart-Infusion (BHI)	Roth
Bromphenolblau (BPB)	Serva
BSA	Applichem
Calciumchlorid	Merck
CASO-Bouillon	Roth
Chitinkügelchen	NEB
Coomassie Brilliant Blue R250/G250	Serva
D(+)-Glucose Monohydrat	VWR
DEPC-Wasser	eigene Herstellung / Roth
Dikaliumhydrogenphosphat	Merck
Dimethylsulfoxid (DMSO)	Fluka
Dinatriumhydrogenphosphat	Merck
Dithiothreitol (DTT)	Sigma
DTT	Applichem
Eisessig	Roth
Ethanol	Roth
Ethidiumbromid	Serva
Ethylendiamintetraessigsäure (EDTA)	Amersham
Glucose	Merck
Glycerol	Roth
Glykogen RNase-frei	Invitrogen
Isopropyl- β -D-Thiogalactopyranosid (IPTG)	Thermo Fisher
Kaliumdihydrogenphosphat	Merck
LB Agar	Gibco BRL
LB Broth Base	Invitrogen
Magnesiumchlorid	Merck
Magnesiumsulfat Heptahydrat	Merck
MOPS	Serva
Natriumacetat	Merck
Natriumchlorid	Roth
Natriumdodecylphosphat (SDS)	Serva
Natriumhydroxid	Merck
Natronlauge	Merck
PCR-Puffer für KOD-Polymerase	Merck Millipore
PCR-Puffer für Taq-, Phusion-Polymerase	Thermo Fisher
Phenol-Chloroform-Isoamylalkohol (PCI)	Roth
Restriktionspuffer	Thermo Fisher
RNase Zap	Sigma
Sorbitol	Sigma
Stickstoff (flüssig)	Linde
Sucrose	Merck
T4-DNA-Ligasepuffer	Thermo Fisher
TEMED	Sigma
Tris-HCl	Serva
Triton X-100	Merck

3 Material und Methoden

Tabelle 3.6: Verwendete Materialien und deren Bezugsquelle.

Material	Bezugsquelle
1,5 mL-Reaktionsgefäße	Starlab
2 mL-Schraubgefäße	Starlab
Elektroporationsküvetten	Peqlab
Filterspitzen, verschiedene Größen	Biozym
Glasperlen (0,1 mm Durchmesser)	Retsch
Glaswaren	Schott
Halb-Mikroküvetten	Sarstedt
Hybond-N Membran	GE Healthcare
Nitrilhandschuhe	VWR
Parafilm	American National Can
PCR-Reaktionsgefäße	VWR
PE-Röhrchen, mit / ohne Deckel, verschiedene Größen	Starlab, Greiner, Sarstedt
Petrischalen	Sarstedt
Pipettenspitzen, verschiedene Größen	Starlab
Proteinausschlussäulen (10 kDa)	Millipore
Proteinofilter Columns (35 mL)	Macherey-Nagel
Sterilfilter	Sartorius, Whatman

Tabelle 3.7: Verwendete Kits und deren Bezugsquelle.

Kit	Bezugsquelle
Agilent High Sensitivity DNA Kit	Agilent Technologies
Agilent RNA 6000 Nano Kit	Agilent Technologies
Agilent RNA 6000 Pico Kit	Agilent Technologies
IMPACT-CN Kit	New England BioLabs
NucleoSpin Extract II	Macherey-Nagel
Qiagen QIAquick Gel Extraction Kit	Qiagen
Quant-iT Pico Green	Invitrogen
Ribo-Zero rRNA Removal Kit	Epicenter
RNase-Free DNase Set	Qiagen
Roti Nanoquant Kit	Carl Roth
RNeasy Mini Kit	Qiagen
ThermoScript RT-PCR System	Invitrogen
TruSeq stranded RNA Sample Preparation Kit	Illumina

Tabelle 3.8: Verwendete Geräte für die generelle Laborarbeit und deren Hersteller. Spezielle Geräte sind in den jeweiligen Abschnitten genannt.

Gerät	Hersteller
Biflex III-MALDI-ToF-MS	Bruker
Brutschränke	Memmert
Elektrophoresekammer	BioRAD
Feinwaagen	Sartorius
French Press	Thermo Fisher Scientific

3.2 Verwendete Chemikalien, Materialien, Geräte und Software

Tabelle 3.8: Fortgesetzt

Gerät	Hersteller
Gefrierschränke (-80 °C)	Sanyo
Gene Pulser, Pulse Control	BioRad
Hybridisierungssofen (Roller)	Bachofer
Kühlschrank	Privileg/Siemens
Kühltischzentrifuge 5417R	Eppendorf
Kühlzentrifuge groß (4K 15)	Sigma
LightCycler	Roche
Luftschüttler	New Brunswick, Gerhardt
Luminescent Image Analyzer LAS-3000	Fujifilm Europe
Magnetrührer mit Heizplatte	IKA Labortechnik
Mastercycler Pro S Thermocycler	Eppendorf
Multifuge 3L-R	Heraeus
Nanodrop 2000 Spectrometer	Peqlab
pH-Meter	Knick
Photometer	Eppendorf
Pipetten	Eppendorf
Reinstwasseranlage	Millipore
Ribolyser	Hybaid
Spannungsgeber	BioRAD
Speed-Vac SPD 111V	Thermo Savant
Tisch-Zentrifugen	Eppendorf
Typhoon 8600-Scanner	Amersham Biosciences
UV-Transilluminator	UVP
Vortex-Genie 2	Scientific Industries
Wasserbäder	GFL

Tabelle 3.9: Verwendete Software. Weitere Software ist in den jeweiligen Abschnitten genannt.

Software	Version	Hersteller / Bezugsquelle
CloneManager	9.3	Scientific Education Software
GenDB	2.4	Meyer <i>et al.</i> (2003), https://gendb.computational.bio.uni-giessen.de
Inkscape	0.92.3	Inkscape Community, https://inkscape.org/
LibreOffice Calc	5.4.6.2	The Document Foundation, https://www.libreoffice.org/
LightCycler Software	1.1.0	Roche
Python	2.7	Python Software Foundation https://www.python.org/
R	3.2.5	R Core Team https://www.r-project.org/
TeX Live	2015	TeX Users Group https://www.tug.org/texlive/

3 Material und Methoden

3.3 Medien und Lösungen

Die folgende Tabelle 3.10 enthält die verwendeten Medien. Um aus den aufgeführten Medien Festmedien für die Plattenkultivierung herzustellen, werden 15 g L^{-1} Agar hinzugefügt. Die Zusätze zu den Medien sind in Tabelle 3.11 aufgelistet.

Tabelle 3.10: Verwendete Medien zur Kultivierung von Bakterien.

CG XII nach Keilhauer <i>et al.</i> (1993)	
In 940 mL Wasser lösen	
MOPS	42 g L^{-1}
Ammoniumsulfat	20 g L^{-1}
Dikaliumhydrogenphosphat	1 g L^{-1}
Kaliumdihydrogenphosphat	1 g L^{-1}
Den pH-Wert mit 10 M Kaliumhydroxid-Lösung auf 7,0 einstellen und autoklavieren.	
Nach autoklavieren zugeben:	
Harnstoff-Lösung (500 g L^{-1})	10 mL L^{-1}
Glucose-Monohydrat-Lösung (440 g L^{-1})	50 mL L^{-1}
Magnesiumsulfat-Heptahydrat-Lösung (250 g L^{-1})	1 mL L^{-1}
Biotin-Stammlösung	1 mL L^{-1}
Protocatechusäure-Stammlösung	1 mL L^{-1}
CG XII Spurenelement-Stammlösung	1 mL L^{-1}
(Brain) Heart Infusion (B/HI) Medium nach Liebl <i>et al.</i> (1989)	
BHI Broth Base	20 g L^{-1}
Wahlweise mit 90 g L^{-1} Sorbitol (BHIS)	
In Wasser lösen und autoklavieren.	
Caseinpepton (CASO) Medium	
CASO-Broth Base	30 g L^{-1}
Glucose-Monohydrat-Lösung (440 g L^{-1})	50 mL L^{-1}
Wahlweise mit Sorbitol (CASO+S)	90 g L^{-1}
In Wasser lösen und autoklavieren.	
SOC Medium nach Hanahan (1983)	
Trypton	20 g L^{-1}
Hefeextrakt	5 g L^{-1}
Glucose-Monohydrat	4 g L^{-1}
Natriumchlorid	580 mg L^{-1}
Kaliumchlorid	180 mg L^{-1}
Magnesiumchlorid-Hexahydrat	2 g L^{-1}
Magnesiumsulfat-Heptahydrat	$2,4 \text{ g L}^{-1}$
In Wasser lösen und autoklavieren.	
LB (Luria Bertani) nach Sambrook <i>et al.</i> (2001)	
LB-Broth Base	20 g L^{-1}
In Wasser lösen und autoklavieren.	
SOB-Medium (Super Optimal Broth) nach Hanahan (1983)	
Hefeextrakt	5 g L^{-1}

3.4 Kultivierung und Lagerung von Bakterien

Tabelle 3.10: Fortgesetzt

Trypton	20 g L ⁻¹
Natriumchlorid	10 mM
Kaliumchlorid	2,5 mM
Magnesiumchlorid	10 mM
Magnesiumsulfat	10 mM
Glucose (für SOC-Medium)	20 mM
In Wasser lösen und autoklavieren.	

Tabelle 3.11: Medien-Zusätze. Wenn nicht anders angegeben, werden die Medienzusätze in Wasser gelöst und sterilfiltriert.

Antibiotika	
Kanamycin (für <i>E. coli</i>)	50 mg L ⁻¹
Kanamycin (für <i>C. glutamicum</i>)	25 mg L ⁻¹
Biotin-Stammlösung	
Biotin	200 mg L ⁻¹
In Wasser mit etwas Natriumhydroxid lösen und sterilfiltrieren	
CG XII-Spurenelement-Stammlösung	
Eisensulfat-Heptahydrat	10 g L ⁻¹
Mangansulfat-Heptahydrat	10 g L ⁻¹
Zinksulfat-Heptahydrat	1 g L ⁻¹
Kupfersulfat	0,2 g L ⁻¹
Nickeldichlorid-Hexahydrat	0,02 g L ⁻¹
In 0,1 M Salzsäure lösen und autoklavieren.	
Protokatechusäure-Stammlösung	
Protokatechusäure	30 g L ⁻¹
In Ethanol lösen und sterilfiltrieren.	

3.4 Kultivierung und Lagerung von Bakterien

Die Anzucht von Bakterien auf Festmedien erfolgt durch Ausstrich einer Einzelkolonie bzw. aus der Glycerinkultur der Stammsammlung. Die optimale Temperatur zur Kultivierung von *E. coli* im Brutschrank beträgt 37 °C, die für *C. glutamicum* beträgt 30 °C. Die Kultivierung von *E. coli* und *C. glutamicum* in Flüssigmedium erfolgt bei 37 °C bzw. 30 °C im Luftschüttler bei 150 rpm bzw. 300 rpm. Als Kultivierungsgefäße werden mit einem Deckel verschließbare Erlenmeyerkolben verwendet, die einen Luftaustausch mit der Umgebung erlauben. Alternativ wurden auch Kultivierungen in Bioreaktoren durchgeführt. Bioreaktoren erlauben zusätzlich

3 Material und Methoden

die Kontrolle und Regelung von Parametern wie Sauerstoffsättigung und des pH-Werts. In dieser Arbeit wurden die Bioreaktorsysteme von Satorius verwendet. Typische Kultivierungsparameter für *C. glutamicum* sind 30 °C, pH 7,00 bei einer Rührerdrehzahl von anfangs 120 rpm. Das Volumen der Kultur im Bioreaktor beträgt 0,8 L. Die Bioreaktor-Kontrolleinheit wurde so eingestellt, dass der pH-Wert über die Zugabe von Ammoniak-Lösung auf 7,00 konstant gehalten wurde. Außerdem wurde die Sauerstoffsättigung von 30 % eingestellt, die über die Erhöhung des Gasflusses (Luft) und der Rührerdrehzahl geregelt wurde.

Festmedienkulturen können im Kühlschrank bei 4 °C für einige Wochen gelagert werden. Durch regelmäßiges Überimpfen von Einzelkolonien auf frische Festmedien, kann die Lagerzeit erhöht werden. Um die Platten vor Austrocknung zu bewahren, werden sie mit Parafilm versiegelt. Für eine Lagerung über Monate und Jahre eignen sich Glycerinkulturen. Dazu wird eine Kultur, die sich gerade in der exponentiellen Wachstumsphase befindet, mit Glycerin im Verhältnis 40:60 vermischt und direkt in flüssigem Stickstoff schock gefroren. Eine Lagerung bei -80 °C ist so über Jahre möglich.

3.5 Bestimmung des Bakterientiters

Der Bakterientiter gibt die Zahl der (lebenden) Zellen pro Milliliter Kultur an. Eine einfache und schnelle Methode zur Bestimmung des Bakterientiters ist die Bestimmung der optischen Dichte bei einer Wellenlänge von 600 nm (OD_{600}) mit einem Photometer. Eine OD_{600} von 1 entspricht $2 \cdot 10^8$ Zellen pro Milliliter bei *E. coli*. Bei *C. glutamicum* entspricht eine OD_{600} von $1 \cdot 10^8$ Zellen / mL. Die Bestimmung des Bakterientiters über die optische Dichte ist nicht exakt, da idealisiert von einer konstanten Zellgröße ausgegangen wird. Außerdem kann nicht zwischen lebenden und bereits abgestorbenen, aber noch ganzen, Zellen unterschieden werden. Eine genauere, aber wesentlich zeitaufwändigere Bestimmung des Bakterientiters ist über die Zellzählung mittels Thomakammer möglich (Sambrook *et al.*, 2001). Aus zeitlichen Gründen wurde in dieser Arbeit auf die Zellzählung verzichtet und der photometrischen Bestimmung der optischen Dichte der Vorzug gegeben.

3.6 Allgemeine DNA-Arbeiten

3.6.1 Polymerase-Kettenreaktion (PCR)

Die Polymerase-Kettenreaktion (*Polymerase Chain Reaction*, PCR) ist eine Methode zur Vervielfältigung von DNA anhand einer Vorlage (*template*) (Sambrook *et al.*, 2001). Im ersten Schritt wird die DNA-Vorlage (z. B. das bakterielle Chromosom) mittels Hitze in Einzelstränge

3.6 Allgemeine DNA-Arbeiten

denaturiert. Anschließend lagern sich spezifisch Primer (Oligonukleotide) über komplementäre Basenpaarung an die DNA-Einzelstränge an. Eine DNA-Polymerase beginnt hinter den Primern den DNA-Einzelstrang mit dem komplementären Einbau von Nukleotiden. Diese drei Schritte werden in mehreren Zyklen wiederholt, so dass das DNA-Fragment zwischen den beiden Primern exponentiell vervielfältigt wird.

Die für die PCR benötigten Primer werden mit Hilfe von speziellen Programmen entworfen. Dabei sind die Schmelztemperatur und der GC-Gehalt die wichtigsten zu beachtenden Parameter und sollten für beide Primer nach Möglichkeit ähnlich sein.

Im Rahmen dieser Arbeit wurden alle PCRs für Klonierungen von DNA-Fragmenten mit der „KOD Hot Start“ DNA-Polymerase (Novagen / Toyobo) durchgeführt, da diese eine gute Fehlerkorrektur (*proof reading*) besitzt und auch bei anspruchsvollen DNA-Vorlagen gute Ergebnisse liefert.

Für einfache PCRs die nur zum Längennachweis in der Agarose-Gelelektrophorese dienen, wurde die günstigere DNA-Polymerase „BioTAQ“ (Bioline) verwendet. Diese besitzt nur eingeschränkte Fehlerkorrekturfähigkeiten.

Abhängig von der zu amplifizierenden DNA werden unterschiedliche PCR-Protokolle durchgeführt, sodass hier nicht die Reaktionsbedingungen für jede PCR aufgeführt werden können. Die grundlegenden PCR-Reaktionsbedingungen für beide DNA-Polymerasen sind in Tabellen 3.13 und 3.14 aufgeführt.

Tabelle 3.13: Pipettierschema für PCR-Reaktionen. Für die Amplifikation von chromosomaler DNA wurden 100 ng DNA, für die Amplifikation von Plasmid-DNA maximal 10 ng DNA als Vorlage eingesetzt.

KOD-PCR	Volumen [μL]	BioTAQ-PCR	Volumen [μL]
„KOD Hot Start Buffer“ (10-fach)	5	„NH ₄ -Reaction-Buffer“ (10-fach)	5
Magnesiumsulfat (25 mM)	3	Magnesiumchlorid (50 mM)	2,5
dNTP-Mix (2,5 mM)	5	dNTP-Mix (2,5 mM)	5
KOD Hot Start Polymerase (1 U/ μL)	1	BioTaq-Polymerase (1 U/ μL)	0,5
Primer 1 (10 μM)	1,5	Primer 1 (10 μM)	1,5
Primer 2 (10 μM)	1,5	Primer 2 (10 μM)	1,5
DNA-Vorlage	1	DNA-Vorlage	1
Wasser	32	Wasser	33
Gesamtvolumen	50	Gesamtvolumen	50

3 Material und Methoden

Tabelle 3.14: PCR-Programme für die „KOD Hot Start“ und BioTAQ DNA-Polymerasen. Abhängig von den verwendeten Primern und der Länge des zu amplifizierenden DNA-Fragments müssen die Annealing-Temperatur sowie die Elongationszeit angepasst werden.

#	Schritt	Temperatur [°C]	Zeit
1	Initiale Denaturierung	95	2 min
2	Denaturierung	95	20 s
3	Annealing	55	10 s
4	Elongation (10 – 25 s/kbp)	70	X
	20 – 40 Zyklen der Schritte 2 bis 4		
5	Finale Elongation	70	5 min

3.6.2 Isolierung von DNA

Zur Isolierung von chromosomaler oder plasmidärer DNA aus Zellen wurde das Kit „GeneJET Plasmid Midiprep Kit“ (Thermo Fisher Scientific) nach Anleitung verwendet. Für die DNA-Isolierung aus *C. glutamicum* wurde der Zellaufschluss angepasst, da die Zellen nur schwer chemisch aufgeschlossen werden können. *C. glutamicum*-Zellen wurden von einer frischen Festmedium-Petrischale in Resuspensionspuffer des Kits resuspendiert. Anschließend wurden die Zellen mittels Glaskügelchen (0,1 mm Durchmesser) in der Kugelmühle Ribolyzer (Peqlab) 3x für 30 s bei 6.500 rpm aufgeschlossen und die Glaskügelchen danach durch Zentrifugation (5 min, 13.000 x g) pelletiert. Der Überstand konnte im Anschluss wieder nach der Anleitung des Kits weiterverarbeitet werden. Die anschließende DNA-Konzentrationsbestimmung erfolgte mittels NanoDrop (Peqlab).

3.6.3 Agarose-Gelelektrophorese

Die Auftrennung von DNA anhand ihrer Größe erfolgte über Agarose-Gelelektrophorese mit einer Agarosekonzentration von 0,7 bis 2,0 %, abhängig von der zu trennenden DNA. Die Auftrennung im Gel erfolgte für ca. 30 min bei 80 V

3.6.4 DNA-Aufreinigung und -Quantifizierung

DNA aus PCR-Ansätzen und aus Agarosegelen ausgeschnittene DNA-Fragmente wurden mit Hilfe des Kits „NucleoSpin Gel and PCR Clean-up“ (Macherey-Nagel) nach Anleitung aufgereinigt. Die anschließende DNA-Konzentrationsbestimmung erfolgte mittels NanoDrop (Peqlab).

3.6.5 Klonierung mittels Gibson-Assembly

DNA-Fragmente wurden mittels Gibson-Assembly (Gibson *et al.*, 2009) kloniert. Die Methode basiert auf dem isothermalen Zusammenbau von DNA-Fragmenten mit komplementären Enden. In einem Ansatz befinden sich die T5-Exonuklease, eine DNA-Polymerase, sowie eine DNA-Ligase. Die T5-Exonuklease verdaut die 5'-Enden der DNA-Fragmente, sodass die komplementären Enden an einander binden können. Diese komplementären Enden dienen der DNA-Polymerase als Primer um die, durch die Exonuklease entstandenen, Lücken aufzufüllen. Die Einzelstrangbrüche (*nicks*) werden durch die DNA-Ligase aufgefüllt.

Das Gibson-Assembly eignet sich für die Klonierung von mehreren DNA-Fragmenten in Plasmide. Ein typisches Protokoll zum klonieren von einem DNA-Fragment in ein Plasmid ist im Folgenden aufgeführt: Es wurden zwei Primer P1 und P2 entworfen, die an die Enden des zu klonierenden Fragments binden. Zwei weitere Primer P3 und P4 wurden entworfen, die an der Stelle im Plasmid binden, an der das DNA-Fragment eingefügt werden soll. Diese Primer müssen in entgegengesetzte Richtungen zeigen. An die Primer P1 und P2 wurden Überhänge angefügt, die komplementär zum Einfügeort auf dem Plasmid sind. Es wurde analog für Primer P3 und P4 vorgegangen. Mit Hilfe von PCR wurden das DNA-Fragment mit den Primern P1 und P2 und das Plasmid mit den Primern P3 und P4 amplifiziert. Falls Fehlbanden neben dem gewünschten PCR-Produkt erkennbar waren, wurden die jeweiligen Produkte aus dem Agarosegel ausgeschnitten und aufgereinigt. Das PCR-Produkt des DNA-Fragments wurde mit dem des amplifizierten Plasmids in einem Molverhältnis von 1:1 oder 2:1 gemischt und mit Wasser auf ein Gesamtvolumen von 5 μL aufgefüllt. Zu den gemischten PCR-Produkten wurden 15 μL Gibson-Assembly Master Mix hinzugefügt (enthält Exonuklease, Polymerase und Ligase, sowie dNTPs und Kofaktoren). Der Reaktionsansatz wurde für 60 min bei 50 °C inkubiert und anschließend in kompetente Zellen transformiert.

3.6.6 Herstellung von kompetenten Zellen

Kompetente Zellen werden für die Transformation von Plasmiden benötigt. Für *E. coli* und *C. glutamicum* wurden jeweils unterschiedliche Protokolle verwendet. Die benötigten Lösungen und Puffer sind in Tabelle 3.15 aufgeführt.

3 Material und Methoden

Tabelle 3.15: Puffer und Lösungen für die Herstellung von kompetenten Zellen.

TB-Puffer	
PIPES	10 mM
Calciumchlorid	15 mM
Kaliumchlorid	250 mM
Auf ca. 800 mL auffüllen, pH auf 6,7 einstellen (mit KOH oder HCl)	
Manganchlorid	55 mM
In Wasser lösen, sterilfiltrieren und bei 4 °C lagern.	
TG-Puffer	
Tris-HCl	1 mM
Glycerin (auf pH 7,5 einstellen)	10 %
Glycerin 10 %ig	
Glycerin (auf pH 7,0 einstellen)	10 %

Herstellung von ultrakompetenten *E. coli*-Zellen für Hitzeschocktransformation

Herstellung der ultrakompetenten Ecolis nach einer modifizierten Calciumchlorid-Methode zur Hitzeschocktransformation nach Inoue *et al.* (1990).

Als Vorkultur wurden 10 mL LB-Medium in einem 100 mL-Schüttelkolben mit 50 μ L einer Glycerinkultur des gewünschten Stammes beimpft und bei 180 rpm und 37 °C über Nacht inkubiert. Für die Hauptkultur wurden 250 mL SOB-Medium in einem 1 L-Kolben mit 2,5 mL der Vorkultur beimpft und bei 180 rpm und 19 °C bis OD600 = 0,5 inkubiert (Dauer 24 - 36 h). Anschließend wurde der Kolben 10 min im Eiswasser abgekühlt. Da alle weiteren Schritte bei 4 °C durchgeführt wurden, wurden die Lösungen und die Zentrifuge ebenfalls vorgekühlt. Die Hauptkultur wurde in fünf 50 mL-Zentrifugengefäßen umgefüllt und bei 5.000 g für 5 min pelletiert. Der Überstand wurde verworfen. Das Zellpellet wurde mit je 20 mL eiskaltem TB-Puffer resuspendiert und 10 min auf Eis inkubiert. Anschließend wurden die Zellen erneut bei 5.000 g für 5 min zentrifugiert und der Überstand vorsichtig abgegossen. Das Zellpellet wurde in 6 mL eiskaltem TB-Puffer resuspendiert mit 420 μ L DMSO (Dimethylsulfoxid) gemischt und sofort gevortext. Die nun kompetenten Zellen werden in 150 μ L Aliquots abgefüllt, in flüssigem Stickstoff schockgefroren und bei -80 °C gelagert.

Herstellung von elektrokompetenten *C. glutamicum*-Zellen

Elektrokompetente *C. glutamicum*-Zellen wurden mit einem modifizierten Protokoll nach Kirchner und Tauch (2003) erstellt.

Es wurde eine *C. glutamicum* Glycerinkultur frisch auf einer CASO-Platte mit Nalidixin ausgestrichen und über Nacht bei 30 °C inkubiert. Als Vorkultur wurde eine Einzelkolonie in 50 ml CASO+S-Medium (CASO mit 91 g/L Sorbitol) überimpft und über Nacht bei 30 °C bei 300 rpm inkubiert. Die Hauptkultur wurde mit einer OD600 0,5 in 2x 250 ml CASO+S-Medium (auf 30 °C vorgewärmt) angesetzt und bei 30 °C bis zu einer OD600 von ca. 2,0 bis 2,5 angezogen (Dauer ca. 2 - 2,5 h). Anschließend wurde der Kolben ca. 15 min auf Eis gekühlt. Da alle weiteren Schritte bei 4 °C durchgeführt wurden, wurden die Lösungen und die Zentrifuge ebenfalls vorgekühlt. Es wurden 500 ml Kultur in Zentrifugenbecher transferiert und diese für 8 min bei 4.500 rpm zentrifugiert. Der Überstand wurde verworfen. Das Pellet wurde in 100 ml eiskaltem TG-Puffer resuspendiert und 7 min bei 4.500 rpm zentrifugiert. Der Überstand wurde verworfen. Das Pellet wurde anschließend in 5 ml kaltem 10 %igem Glycerin resuspendiert und 7 min bei 4.500 rpm zentrifugiert. Die nun kompetenten Zellen wurden in Aliquots von 150 µL in vorgekühlte Eppis überführt und bei -80 °C eingefroren.

3.6.7 Transformation von DNA

Die Transformation von DNA in Zellen erfolgt über Hitzeschock oder Elektroporation. Für die Hitzeschocktransformation von *E. coli* wurde auf das Protokoll von Sambrook *et al.* (2001) zurückgegriffen: Es wurden 100 µL kompetente *E. coli*-Zellen mit maximal 20 µL der zu transformierenden DNA vermischt und für 2 min auf Eis gestellt. Anschließend wurde der Ansatz bei 42 °C für 30 s im Wasserbad erhitzt und direkt im Anschluss für 1 min auf Eis gestellt. Für die Regeneration der transformierten Zellen wurden diese in 1 mL SOC-Medium resuspendiert und für 45 bis 60 min bei 37 °C im Wasserbad inkubiert. Im Anschluss wurden die transformierten Zellen auf Selektionsmedien ausgestrichen.

Für *C. glutamicum*-Zellen eignet sich die Hitzeschocktransformation nicht. Daher wurde hier die Transformation über Elektroporation nach Tauch *et al.* (2002) durchgeführt. Dazu wurden die zu transformierenden elektrokompententen Zellen zunächst auf Eis aufgetaut und anschließend mit 0,1 bis 10,0 µg DNA in einer Elektroküvette vermischt. Die Elektroporation wurde bei einer Kapazität von 25 µF, einem Parallelwiderstand von 200 Ω und einer Spannung von 2,5 kV in einem Elektroporator von BioRad durchgeführt. Anschließend wurden die transformierten Zellen in auf 46 °C vorgewärmtes 4 mL BHIS-Medium für 6 min bei 46 °C inkubiert und danach für 50 min in einem Rotationsschüttler bei 30 °C regeneriert. Die Zellen wurden im Anschluss auf Selektionsmedien ausgestrichen.

3.7 Bestimmung der Transformationsfrequenz

Die Transformationsfrequenz ist der Anteil an Zellen, die die zu transformierende DNA aufgenommen haben (Transformanten), bezogen auf die Gesamtzahl der eingesetzten lebendigen Zellen. Zur Bestimmung der Transformationsfrequenz von *C. glutamicum*-Zellen wurden, wie oben beschrieben, kompetente Zellen hergestellt und diese am selben Tag transformiert. Die Transformation erfolgte mittels Elektroporation, da mittels Hitzeschocktransformation keine *C. glutamicum*-Zellen transformiert werden konnten. Transformiert wurde jeweils 100 ng des Plasmids pZMP in einem Volumen von 1 μ L Wasser. Die Elektroporation wurde wie oben bereits beschrieben durchgeführt. Allerdings wurde vor der Regeneration bei 30 °C ein Aliquot abgenommen, um die Gesamtzahl an lebendigen Zellen (Z_0) nach Poration zu messen. Die Lebendzellzahl Z_0 wurde über eine Verdünnungsreihe der Zellen, die auf CASO-Agar-Platten ohne Antibiotikum ausgestrichen wurden, ermittelt. Nach der Regeneration wurde die Lebendzellzahl Z_R über eine Verdünnungsreihe sowie die Anzahl an Zellen mit aufgenommenem Plasmid (Transformanten) nach der Regeneration T_R bestimmt. Die Transformanten wurden dazu auf Selektionsplatten ausgestrichen. Die Bestimmung der Lebendzellzahl vor und nach Regeneration war notwendig, um auf das dabei stattfindende Wachstum korrigieren zu können. Andernfalls würde für Zellen mit höherer Wachstumsgeschwindigkeit, im Vergleich zu Zellen mit geringerer Wachstumsgeschwindigkeit, fälschlicherweise eine erhöhte Transformationsfrequenz gemessen werden, obwohl nicht mehr Zellen vorhanden waren. Der Korrekturfaktor A setzt sich aus dem Quotienten von Z_R/Z_0 zusammen. Die Transformationsfrequenz F errechnet sich somit über $F = Z_0/(T_R \cdot A)$. Die Transformationsfrequenzen wurden für jede Mutante bzw. den Wildtyp in mindestens vier biologischen Replikaten bestimmt.

3.7.1 Durchführen von Gendeletionen

Gendeletionen in *C. glutamicum* wurden mit dem pK18*mobsacB*-System (Schäfer *et al.*, 1994) durchgeführt. Die Deletion von Genen erfolgt bei diesem System über zwei homologe Rekombinationsereignisse. Für die Deletion eines Gens werden zwei, zum Genom homologe, DNA-Abschnitte (Flanken) benötigt, die in das Plasmid pK18*mobsacB* kloniert werden. Dieses Plasmid lässt sich in *E. coli*, nicht aber in *C. glutamicum* replizieren. Über die plasmidcodierte Kanamycin-Resistenz und eine Levan-Sucrase (*sacB*), die bei Gegenwart von Sucrose die Zellwandsynthese stört und zum Zelltod führt, kann auf die beiden Rekombinationsereignisse selektioniert werden.

Die Flanken mit einer Länge von ca. 500 bp wurden über spezifisch entworfene Primer amplifiziert und mittels Gibson-Assembly in das Plasmid pK18*mobsacB* kloniert. Anschließend wurde das Plasmid in *C. glutamicum* transformiert und die Zellen auf Vollmediumplatten

3.7 Bestimmung der Transformationsfrequenz

mit Kanamycin ausplattiert. Die erste Selektion auf Integration des Plasmids in das zelluläre Genom erfolgte über die Kanamycin-Resistenz. Zellen die das Plasmid integriert hatten, konnten auf diesen Platten wachsen. Anschließend wurden die Kolonien mit einem Zahnstocher gepickt und mit Wasser 1:10 und 1:100 verdünnt. Diese Verdünnungen wurden auf eine Vollmediumplatte mit 10 % Sucrose ausplattiert. Danach wurden die Zellen erneut mit einem Zahnstocher gepickt und auf eine Vollmediumplatte mit Kanamycin und auf eine Vollmediumplatte mit 10 % Sucrose überstocht. Dabei wurden gerasterte Platten verwendet, um die Positionen der Zellen auf beiden Platten festhalten zu können. Zellen, die nicht auf der letzten Kanamycin-Platte, aber dafür auf der Sucrose-Platte wuchsen, hatten das Plasmid erfolgreich desintegriert und wahrscheinlich auch die Deletion übernommen. Alternativ könnte auch der ursprüngliche Wildtyp wieder entstanden sein. Mittels PCR wurde zwischen Wildtyp und korrekter Deletion unterscheiden, indem zwei Primer eingesetzt wurden, die außerhalb des Deletionsbereichs binden. Zellen mit korrekter Deletion zeigen in dieser PCR ein im Vergleich zum Wildtyp verkürztes PCR-Produkt. Klone mit korrektem Genotyp wurden anschließend auf Vollmediumplatten mit Nalidixinsäure ausplattiert.

3.7.2 Plasmidbasierte Genexpression

Für die plasmidbasierte Genexpression wurde das Plasmid pEC-XC99E verwendet. Es lässt sich in *E. coli* als auch in *C. glutamicum* replizieren (*Shuttle-Vector*). Das Plasmid pEC-XC99E basiert auf dem *medium copy number* Plasmid pGA1 (Nešvera *et al.*, 1997; Tauch *et al.*, 2002). Es besitzt den starken Promotor des *trc*-Gens aus *E. coli* sowie einen *downstream* befindlichen *lac*-Operator, welches eine Induktion der *downstream* klonierten Gene mittels IPTG erlaubt. Der dazu benötigte *lac*-Repressor ist dazu auf dem Plasmid kodiert. Darüber hinaus vermittelt das Plasmid eine Chloramphenicol-Resistenz über das kodierte *cat*₁-Gen (Kirchner und Tauch, 2003).

Die plasmidbasierte Genexpression mittels pEC-XC99E wurde in dieser Arbeit für die Komplementation von Deletionsmutanten verwendet. Die Klonierungen der zu überexprimierenden Gene erfolgte über Gibson-Assembly. Da das Plasmid keine ribosomale Bindestelle besitzt, wurde die Konsensussequenz mit einem Abstand von 7 bp zur CDS wie in Pfeifer-Sancar *et al.* (2013) beschrieben, eingesetzt.

3.7.3 Insertion von Genen mittels pRIM2-System

Das Plasmid pRIM2 ist von Vašicová *et al.* (1998) als Testsystem zur Bestimmung der Promotorstärke entworfen worden. Das Plasmid ist in *C. glutamicum* nicht replikativ und integriert über homologe Rekombination in das *C. glutamicum* Genom *downstream* des Gens *ppc*, welches

3 Material und Methoden

für eine Phosphoenolpyruvate-Carboxylase kodiert. Die genomische Integration des Plasmids hat keinen Effekt auf das Zellwachstum (Vašicová *et al.*, 1998).

Für die Komplementation der *C. glutamicum* $\Delta\rho$ -Mutante wurde das pRIM2-System gewählt, da eine Komplementation über induzierbare *medium-copy* Plasmide nicht erfolgreich war. In das Plasmid pRIM2 wurde daher das *rho*-Gen aus *C. glutamicum* mit nativem Promotor kloniert. Da *rho* an erster Stelle in einem Operon mit dem Gen *prfA* liegt, wurde zusätzlich der Haarnadel-Terminator von *prfA* hinter das *rho*-Gen in pRIM2 kloniert, um nach der genomischen Integration des Plasmids die Expression der umliegenden Gene im Genom nicht durch den natürlichen Promotor von *rho* zu beeinflussen.

3.8 Allgemeine Protein-Arbeiten

3.8.1 Proteinaufreinigung mit Affinitätschromatographie

Die Aufreinigung von Proteinen wurde mit Hilfe des IMPACT-CN Kits (*Intein Mediated Purification with an Affinity Chitin-binding Tag*) von *New England Biolabs* durchgeführt. Über die Fusion des Zielproteins mit einem Protein-splice-Element (Intein) und einer Chitinbindedomäne wird die Aufreinigung über Affinitätschromatographie ermöglicht. Das Intein erlaubt die vollständige Abspaltung der Chitinbindedomäne bei Zugabe von Thiolen (z. B. DTT), so dass das ursprüngliche Protein Marker-frei eluiert werden kann. Für die C- bzw. N-terminale Klonierung des Fusionsproteins werden die dem Kit beiliegenden Plasmide pTXB1 und pTYB21 benötigt.

Plasmidpräparation

Das Zielprotein wird in einen der beiden IMPACT-Vektoren kloniert. Für die Plasmidamplifikation eignet sich der *E. coli*-Stamm JM109. Für die Expression des Zielproteins ist der *E. coli*-Stamm ER2566 geeignet. Dieser Stamm besitzt das Gen für die T7 RNA-Polymerase, das durch den *lacI*-Repressor reprimiert ist. Durch Zugabe von IPTG wird die Expression der T7 RNA-Polymerase und damit auch des Zielproteins eingeleitet. Dem genannten Stamm fehlen außerdem die Proteasen *Ion* und *ompT*.

Tabelle 3.17: Lösungen für das IMPACT-CN Kit zur Proteinaufreinigung

EDTA-Stammlösung	
EDTA	1 M
In Wasser lösen	
PMSF-Stammlösung	

Tabelle 3.17: Fortgesetzt

PMSF (Protease-Inhibitor) In DMSO lösen	100 mM
TCEP-Stammlösung	
TCEP (Stabilisator) In Wasser lösen	1 M
Column Buffer	
Na ₂ HPO ₄ x H ₂ O	20 mM
NaCl	500 mM
EDTA	1 mM
Alle Substanzen in Wasser lösen Mit Phosphorsäure tropfenweise auf pH 8 einstellen.	
Lysis Buffer	
Triton X-100	0,1 %
PMSF	20 μM
TCEP	0,1 - 1 mM
Alle Substanzen in Column Buffer lösen	
Cleavage Buffer	
Na ₂ HPO ₄ x H ₂ O	20 mM
NaCl	500 mM
EDTA	1 mM
DTT	50 mM
Alle Substanzen in bidest Wasser lösen Mit Phosphorsäure tropfenweise auf pH 7 - 9 einstellen	
Protein-Wash Buffer	
Na ₂ HPO ₄ x H ₂ O	20 mM
NaCl	500 mM
EDTA	1 mM
In Wasser lösen	

Zellanzucht

Für eine Vorkultur wurde etwas Zellmaterial in 10 mL LB-Medium (mit 200 μg mL⁻¹ Ampicillin) resuspendiert und bei 37 °C inkubiert. Für die Hauptkultur wurden 250 mL LB mit einer OD600 = 0,1 angeimpft und bei 37 °C bis OD600 = 0,5 - 0,8 inkubiert. Anschließend wurden 50 μL für die SDS-PAGE abgenommen und eingefroren (1. Probe). Die Proteinexpression wurde durch Zugabe von IPTG in einer Konzentration von 0,3 - 0,5 mM induziert. Die Expression erfolgte über Nacht bei 15 - 20 °C und 180 rpm. Es wurden wieder 50 μL für die SDS-PAGE abgenommen und eingefroren (2. Probe).

3 Material und Methoden

Zellernte- und Aufschluss

Für die Zellernte wurde die gesamte Zellkultur in großen Zentrifugengefäßen bei 4.500 rpm, 10 - 20 min bei 4 °C pelletiert. Anschließend wurde das Pellet in 20 - 30 mL Lysis Buffer resuspendiert. Der Zellaufschluss erfolgte mittels zwei Durchläufen in einer French Press. Die Zelltrümmer wurden bei 4.500 rpm für 30 - 60 min bei 4 °C abzentrifugiert und die Überstände auf Eis gekühlt. Für die SDS-PAGE wurden 50 µL Überstand abgenommen und eingefroren (3. Probe). Auch eine Spatelspitze vom Pellet wurde für die SDS-PAGE eingefroren (4. Probe).

Vorbereiten, Beladen und Waschen der Säule

Der unterer Teil der Säule wurde mit Parafilm verschlossen und 5 mL Wasser in Säule pipettiert. Die Füllhöhe wurde markiert und das Wasser wieder entfernt. Anschließend wurde die Säule bis zur Markierung mit Chitin Beads (NEB) befüllt und mit einer Filterscheibe von oben leicht komprimiert. Das Säulenmaterial wurde mit 10 Volumen Column Buffer equilibriert. Die Säulen sind nun vorbereitet. Zum Beladen der Säule wurde der klare Überstand auf die Säule gegeben, so dass die Durchflussrate etwa 0,5 - 1 mL min⁻¹ betrug. Anschließend wurden 50 µL vom Durchlauf für die SDS-PAGE eingefroren (5. Probe). Die Säulen wurden anschließend mit 10 Volumen Column Buffer gewaschen.

Induktion der Spaltung und Elution

Zur Induktion der Spaltung des Inteins von dem Fusionsprotein wurde die Säule mit 3 Volumen Cleavage Buffer gewaschen. Anschließend wurden die Säulen für 16 – 40 h bei 4 bis 23 °C inkubiert. Das Zielprotein wurde mit 2 Volumen Column Buffer eluiert und das Eluat aufgefangen. Es wurden 50 µL für die SDS-PAGE abgenommen und eingefroren (6. Probe). Das Eluat wurde mit Hilfe eines geeigneten Proteinfilters (Porengröße kleiner als halbe Größe des Protein) durch Zentrifugation aufkonzentriert. Die aufkonzentrierte Proteinlösung wurde anschließend 2 – 3 mal mit 4 mL Protein-Wash Buffer gewaschen. Zur Proteinquantifizierung wurde eine Bradford-Bestimmung durchgeführt. Zur qualitativen Überprüfung der Aufreinigung wurde eine SDS-PAGE mit den aufgeführten Proben durchgeführt. Die Lagerung des aufgereinigten Proteins erfolgte im Kühlschrank bzw. Eisfach (evtl. mit Glycerin).

3.8.2 Proteinquantifizierung mittels Bradford-Test

Für die Quantifizierung von Proteinen wird häufig der Bradford-Test verwendet. Der Farbstoff Coomassie-Brillant-Blau G250 kann in saurer Lösung mit kationischen oder unpolaren

Seitenketten von Proteinen Komplexe bilden. Diese haben ein Absorptionsmaximum bei 590 nm, wobei der ungebundene Farbstoff ein Absorptionsmaximum bei 450 nm besitzt. Durch die Messung des Verhältnisses von gebundenem zu ungebundenem Farbstoff kann die Proteinkonzentration bestimmt werden. In dieser Arbeit wurde das Roti-Nanoquant-Kit von Carl Roth nach Herstellerangaben zur Bestimmung der Proteinkonzentration verwendet.

3.8.3 Proteinidentifikation mit SDS-PAGE

Die SDS-PAGE (Sodiumdodecylsulfat-Polyacrylamidgelelektrophorese) ist ein Verfahren zur Trennung von Proteinen nach ihrer Größe. Die Trennung erfolgt durch anlegen eines elektrischen Feldes an ein Polyacrylamidgel. Durch Anlagerung von SDS an die Proteine, werden diese linearisiert und erhalten eine gleichmäßige negative Ladungsverteilung an der Oberfläche. Die Eigenladung von Proteinen wird durch SDS maskiert. Die Wanderungsgeschwindigkeit im Gel ist somit nur noch von der Proteingröße (Molekulargewicht) abhängig. Die Wanderungsstrecke im Gel korreliert linear zum dekadischen Logarithmus des Molekulargewichts.

Tabelle 3.18: Lösungen für die SDS-PAGE

SDS-Lösung	
SDS	0,5 % (w/v)
PBJR-Probenpuffer (6 fach)	
TRIS-HCl (pH 6,8)	3,5 mL
SDS-Lösung	3,5 mL
Glycerin (87 % (w/v))	3,5 mL
DTT	930,0 mg
Bromphenolblau	1,2 mg
Mit Wasser auf 10 mL auffüllen.	
Sammelgel-Puffer	
TRIS	500 mM
SDS	0,4 % (w/v)
Den pH-Wert auf 6,8 einstellen	
Trenngel-Puffer	
TRIS	1,5 M
SDS	0,4 % (w/v)
Den pH-Wert auf 8,8 einstellen	
SDS-Laufpuffer (10 fach)	
TRIS	250 mM
Glycin	1,9 M
SDS	1 % (w/v)
Den pH-Wert auf 8,2 bis 8,4 einstellen.	
Coomassie-Proteinfärbelösung	

3 Material und Methoden

Tabelle 3.18: Fortgesetzt

Ethanol	425 mL
Methanol	50 mL
Coomassie Brillant Blau G-250	2,0 g L ⁻¹
Coomassie Brillant Blau R-250	0,5 g L ⁻¹
Wasser	425 mL
Schnell-Entfärber	
Essigsäure	100 mL
Ethanol	450 mL
Wasser	450 mL
Langsam-Entfärber	
Essigsäure	7 % (w/v)
In Wasser lösen.	

Herstellung von diskontinuierlichen SDS-Polyacrylamidgelen

Für die SDS-PAGE wurde das Elektrophorese-System Mini-PROTEAN Tetra Cell von Bio-Rad verwendet. Die benötigten Materialien (Glasplatten, Spacer, Kämme) werden mit Ethanol und Wasser gereinigt und anschließend getrocknet. Die Glasplatten werden den Anweisungen der Anleitung entsprechend zusammgebaut. Die Substanzen für das Trenngel werden in einem Reaktionsgefäß gemischt und die Polymerisation durch Zugabe von TEMED gestartet (Tabelle 3.19). Das Trenngel wird mit einer Pipette zwischen die Glasplatten gegossen und mit Wasser überschichtet. Die Polymerisation ist nach 45 min abgeschlossen. Anschließend wird das Wasser abgegossen und jegliche Flüssigkeitsreste mit einem Papiertuch entfernt. Das Sammelgel wird in analog zum Trenngel angesetzt und mit einer Pipette über das Trenngel geschichtet. Der Kamm wird luftblasenfrei eingesetzt. Nach erfolgter Polymerisation wird das Gel in die Elektrophoresekammer gestellt. Die Kammer wird mit SDS-Laufpuffer gefüllt. Das Gel wird ebenfalls mit SDS-Laufpuffer überschichtet. Anschließend wird der Kamm gezogen und die Taschen mit Laufpuffer gespült.

Tabelle 3.19: Zusammensetzung des Trenn- und Sammelgels der SDS-PAGE

Substanz	Trenngel	Sammelgel
Acrylamid/N,N-methylbisacrylamid	2,50 mL	330 μ L
TRIS-HCl	1,20 mL 1,88 M, pH 8,8	400 μ L 0,625 M, pH 6,8
Wasser	1,10 mL	870 μ L
SDS (0,5 %)	1,20 mL	400 μ L
APS (10 %)	0,03 mL	10 μ L
TEMED (10 %)	0,05 mL	20 μ L

Probenvorbereitung, Gellauf und Färbung mit Coomassie Brilliant Blue

Es wurden eine definierte Menge Protein (z. B. 8 μ g) mit PBJR-Puffer (6-fach) versetzt und die Proben 5 bis 10 min aufgekocht. Anschließend wurden die Proben kurz zentrifugiert und bis zur weiteren Verwendung auf Eis gelagert. Die Proben wurden in die gespülten Geltaschen geladen und der Vorlauf für ca. 30 min bei 15 mA gestartet. Die Elektrophorese fand bei 20 bis 30 mA für 1 - 1,5 h statt, bis die Bromphenolblau-Bande den unteren Gelrand erreichte. Anschließend wurde das Polyacrylamidgel über Nacht mit Coomassie-Lösung gefärbt. Danach wird das Gel für 30 min mit Schnell-Entfärber inkubiert und im Langsam-Entfärber so lange entfärbt, bis die Hintergrundfärbung gering geworden ist. Das Gel wurde anschließend in Klarsichtfolie eingeschweißt und eingescannt.

3.8.4 Proteinidentifikation mit MALDI-ToF/ToF-MS

Für die Identifikation von Proteinen ist eine Größenbestimmung mittels SDS-PAGE nicht ausreichend. Mittels MALDI-MS (*Matrix Assisted Laser Desorption/Ionisation-Mass spectrometry*) kann eine Identifikation auf der Ebene der Aminosäuresequenz durchgeführt werden. Dazu werden die zu untersuchenden Proteine zu Peptiden verdaut und in eine Matrix eingebettet. Die Ionisierung erfolgt durch Beschuss mit einem Laser. Die Matrix-Moleküle werden dabei ionisiert und geben die Ladung über Stoßionisation an die Peptide weiter. Diese können nun im elektrischen Feld über ihr Masse-Ladungsverhältnis (m/z) im ToF-Massenanalysator getrennt werden.

Die Identifikation des Proteins über seine Peptide erfolgt durch Vergleich mit einer Datenbank. Die Datenbank enthält die *in silico* translatierten und verdauten Genprodukte (Proteine bzw. dessen Peptidfragmente) und die entsprechenden (*Peptide Mass Fingerprints*, PMF). Über den Vergleich des PMFs von *in silico* und realen Peptidfragmenten kann das untersuchte Protein eindeutig identifiziert werden. Für den Datenvergleich wird der MOWSE-Algorithmus (*MOlecular Weight SEarch*) (Pappin *et al.*, 1993) verwendet, der den Grad der Übereinstim-

3 Material und Methoden

mung über einen *Mascot-Score* (Perkins *et al.*, 1999) ausdrückt. Im Allgemeinen gilt: je höher der *Mascot-Score*, desto höher die Wahrscheinlichkeit einer korrekten Identifikation.

Das hier verwendete Gerät „ultrafleXtreme“ von Bruker Daltonics besitzt eine MALDI-Ionenquelle. Der Massenanalysator besteht aus einem ToF (*Time of Flight*) mit Reflektor. Als Massendetektor kommt der von Bruker entwickelte FlashDetektor zum Einsatz. Für die Identifikation wurde die Software *BioTools* von Bruker Daltonics verwendet.

Tabelle 3.20: Lösungen für die Proteinidentifikation mit MALDI-ToF-ToF-MS

Matrix-Lösung	
Acetonitril	90 % (w/v)
Trifluoressigsäure	0,1 % (w/v)
Wasser	10 % (w/v)
Ammoniumhydrogenphosphat (NH ₄ H ₂ PO ₄)	1 mM
Matrix	
α-cyano-4-Hydroxymzimtsäure (HCCA)	0,7 mg
Matrix-Lösung	1 mL
TA30-Lösung	
Trifluoressigsäure	0,1 % (w/v)
Acetonitril	33 % (w/v)
Wasser	66,9 % (w/v)
MALDI-Standard	
Angiotensin II	30 μL
Substance P	30 μL
ACTH-Clip 1-17	120 μL
ACTH-Clip 18-39	200 μL
TA30-Lösung	480 μL
TDL-1-Lösung	
Trifluoressigsäure	0,1 % (w/v)
Acetonitril	60 % (w/v)
Wasser	39,9 % (w/v)
TFA-Lösung	
Trifluoressigsäure	0,1 % (w/v)
Acetonitril	50 % (w/v)
Waschlösung	
Ammoniumhydrogenphosphat	0,1 M
Acetonitril	30 % (w/v)
Wasser	70 % (w/v)

Probenvorbereitung und tryptischer Verdau

Als Proben können aus einer SDS-PAGE ausgeschnittene Banden benutzt werden, wenn das Gel nicht mit Silber gefärbt wurde. Die Probenvorbereitung findet unter einem Abzug statt. Da die in Handschuhen vorhandenen Weichmacher im MALDI stören, muss ohne diese gearbeitet werden. Alle Plastikgefäße müssen vor Benutzung mit TDL-1-Lösung von Weichmachern gereinigt werden.

Das aus der SDS-PAGE ausgeschnittene Gelstück wird in 200 μL Waschlösung für 10 min geschüttelt. Der Waschschrift wird ein weiteres Mal wiederholt und das Gelstück anschließend im Rotations-Vakuum-Konzentrator eingetrocknet. Das aktivierte Trypsin wird in einem Verhältnis von 1:14 in 10 mM Ammoniumhydrogencarbonat gelöst und davon 15 μL auf das getrocknete Gelstück gegeben. Der Verdau findet für 30 min bei Raumtemperatur statt. Nach erneuter Trocknung in der Rotations-Vakuum-Konzentrator wird das verdaute Protein in 15 μL TFA-Lösung rehydriert. Die Probe kann nun im MALDI untersucht werden, alternativ ist eine Lagerung bei $-20\text{ }^{\circ}\text{C}$ möglich.

Probenvermessung im MALDI-ToF/ToF-MS

Die Proben werden auf das sogenannte Target (384-Well Mikrotiterplatte Anchor Chip 800/384, Bruker Daltonics) gegeben und im MALDI vermessen. Zuvor muss das Target mit Wasser gewaschen und einem fusselreien Tuch abgetrocknet werden. Nachfolgend wird das Target mit Aceton gereinigt und die Acetonrückstände mit Wasser entfernt. Das Target wird erneut mit einem fusselreien Tuch gründlich getrocknet.

Von den Standards werden jeweils 0,5 μL , von den Proben jeweils 1 μL , auf einen Spot des Targets gegeben. Um die Oberfläche des Targets nicht zu beschädigen, darf die Pipette das Target nicht berühren. Kurz vor dem Eintrocknen der Proben wird pro Spot 1 μL Matrix hinzugegeben. Das Target ist nun für die Messung im MALDI bereit. Vor jeder Messung muss die Quelle im MALDI gereinigt werden.

Die Software BioTools führt nach der erfolgten Messung eine automatische Identifikation der Proben mittels einer für den entsprechenden Organismus spezifische Datenbank durch und gibt die entsprechenden Mascot-Scores aus.

3.9 Allgemeine RNA-Arbeiten

3.9.1 Isolierung von RNA

Grundsätzlich wurde die RNA aus einem Zellpellet, gewonnen aus 1 mL Flüssigkultur, isoliert. Der Zellaufschluss wurde mittels Kugelmühle (Peqlab) bei 6.500 rpm für 30 s (3 Wiederholungen) durchgeführt. Für den Zellaufschluss wurden „FastPrep Lysis Beads (Lysing Matrix B)“ (MPbio) eingesetzt. Abhängig vom verwendeten Protokoll der RNA-Isolation geschah der Zellaufschluss in unterschiedlichen Lösungen.

Für die Isolierung von RNA aus Zellen wurden zwei unterschiedliche Protokolle durchgeführt. So wurde die RNA für qRT-PCR oder RNA-Sequenzierung mittels dem Kit „NucleoSpin RNA Plus“ (Macherey-Nagel) entsprechend den Herstellerangaben isoliert. Diese Methode basiert auf einer Säulenaufreinigung bei der kleiner RNA-Moleküle unter 200 nt nicht mehr quantitativ erhalten bleiben und teilweise verloren gehen. Diese kurzen RNA-Moleküle gehen auch während der Herstellung von cDNA-Bibliotheken für RNA-Sequenzierung nach „TrueSeq Stranded mRNA library“-Protokoll (Illumina) verloren. Daher ist eine Aufreinigung mit dem Kit für den Zweck der RNA-Sequenzierung ausreichend. Auch für die im Rahmen dieser Arbeit durchgeführten qRT-PCR-Reaktionen war die mit dem Kit aufgereinigte RNA ausreichend.

Für Northern Blots wurde die RNA mit TRIzol-Reagenz (Ambion) isoliert. Diese Methode erlaubt auch die quantitative Isolierung von kurzen RNA-Molekülen, die für den Northern Blot als Qualitätsmarker benötigt werden.

Nach erfolgter RNA-Isolierung wurde die RNA auf verbliebene DNA-Kontaminationen mittels PCR überprüft. Konnte DNA in der RNA-Probe nachgewiesen werden, wurde ein DNA-Verdau mit DNase (Qiagen) durchgeführt und ggf. wiederholt, bis keine DNA per PCR mehr nachweisbar war. Abschließend wurde die RNA quantifiziert und die RNA-Qualität bestimmt.

3.9.2 Quantifizierung von RNA

Die Konzentration der isolierten RNA wurde mittels NanoDrop (Peqlab) überprüft. Mit Hilfe des NanoDrops können Verunreinigungen mit Phenol oder Guanidinisothiocyanat anhand des Absorptionsverhältnis A_{260}/A_{230} ermittelt werden. RNA ohne Verunreinigungen haben ein Verhältnis von 2,0 bis 2,2. Im Rahmen dieser Arbeit wurde nur RNA für weitere Analysen eingesetzt, dessen Absorptionsverhältnis $>2,0$ betrug.

Die Analyse der Größenverteilung der isolierten RNA wurde mittels „Agilent RNA Pico 6000 Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) überprüft. Die RNA wird über Kapillargelelektrophorese der Größe nach getrennt und anschließend über ein

Fluoreszenzsignal quantifiziert. Neben der RNA-Konzentration gibt der „Agilent 2100 Bioanalyzer“ einen Qualitätsparameter der RNA aus (*RNA Integrity Number*, RIN). Dieser wird unter anderem anhand des Verhältnisses aus 23S zu 16S rRNA ermittelt. RIN-Werte von 8 bis 10 entsprechen hoher bis sehr hoher RNA-Qualität. Der „Agilent 2100 Bioanalyzer“ kann RNA im Picomol-Bereich quantifizieren.

3.10 Quantitative Reverse-Transkriptase-PCR

Die quantitative Reverse-Transkriptase-PCR (qRT-PCR) ist eine Kombination aus einer Reverse-Transkriptase- mit einer Polymerase-Kettenreaktion zur quantitativen Messung von RNA. Dazu wird die RNA mittels Reverse-Transkriptase und einem Primer in cDNA umgeschrieben und anschließend mit Hilfe eines weiteren Primers amplifiziert. Die dabei entstehende doppelsträngige cDNA wird von einem interkalierenden Fluoreszenzfarbstoff gebunden, so dass die cDNA-Menge (und darüber die RNA-Menge) über das Fluoreszenzsignal in Echtzeit verfolgt werden kann. Dazu wird der C_q-Wert gemessen, der den Zeitpunkt markiert ab dem das Fluoreszenzsignal über das Hintergrundrauschen tritt. Je niedriger der C_q-Wert, desto mehr RNA ist in der Probe vorhanden.

In dieser Arbeit wurde das „SensiFAST SYBR No-ROX One-Step Kit“ (Bioline) verwendet, welches alle notwendigen Enzyme und Chemikalien vorgemischt beinhaltet und somit in einem Reaktionsgefäß (One Step) direkt verwendet werden kann. Zur Vermessung der Proben wurde der Lightcycler 96 von Roche verwendet. Zur Erstellung des qRT-PCR-Protokolls sowie der anschließenden Datenauswertung wurde die zugehörige Geräte-Software eingesetzt.

Für nicht-strangspezifische qRT-PCRs wurden beide Primer im Mastermix vorgelegt und die quantitative PCR direkt nach der reversen Transkription durchgeführt. Im Falle von strangspezifischen qRT-PCRs wurde zunächst der für die Ziel-RNA komplementäre Primer mit der Gesamt-RNA und dem Mastermix gemischt, um in einer reversen Transkription zur Ziel-RNA komplementäre einzelsträngige cDNA zu erhalten. Für die nachfolgende Zweitstrangsynthese und cDNA-Amplifikation wurde anschließend der zweite Primer hinzugefügt und nur noch die quantitative PCR-Reaktion durchgeführt.

Für eine qRT-PCR wurden alle Proben auf 20 ng Gesamt-RNA eingestellt und davon 1 μ L mit 19 μ L Mastermix (18 μ L bei strangspezifischer qRT-PCR) in einer 96-Well-Platte gemischt. Jede zu untersuchende Bedingung wurde dabei in drei biologischen und zwei technischen Replikaten vermessen. Die reverse Transkription erfolgte bei 60 °C für 20 min, mit anschließend 40 Zyklen PCR-Reaktion (95 °C für 15 s, 55 °C für 20 s und 72 °C für 15 s). Als Negativkontrollen wurde 1 μ L RNA- und DNA-freies Wasser anstelle von Gesamt-RNA in zwei technischen Replikaten eingesetzt. Die für die qRT-PCR entworfenen Primer wurden vor dem ersten Einsatz

3 Material und Methoden

mit genomischer DNA auf Funktionalität mit einer PCR überprüft. Wurde in dieser PCR nur ein Produkt erwarteter Länge (ca. 150 – 200 bp) gebildet, sind die Primer für die qRT-PCR tauglich.

3.11 Northern Blot

Der Northern Blot erlaubt den spezifischen Nachweis von RNA mittels Hybridisierung einer Sonde nach Längenauftrennung im Gel. Im Rahmen dieser Arbeit erfolgte der Nachweis der RNA mittels mit Digoxigenin markierten RNA-Sonden basierend auf den Angaben von Roche.

Herstellung Digoxigenin-markierter RNA-Sonden

Für den Northern Blot werden mit Digoxigenin markierte RNA-Sonden benötigt, die spezifisch an die jeweilige Ziel-RNA binden. Dazu wurde zunächst eine PCR mit genomischer DNA als template durchgeführt, um die Sonden-DNA zu amplifizieren und über den *antisense* Primer einen T7-RNA-Polymerase-Promotor in die Amplifikate einzubringen. Nach der Aufreinigung der Sonden-DNA wurde diese über reverse Transkription spezifisch mit T7-RNA-Polymerase in RNA umgeschrieben. Die Markierung der Sonden erfolgte dabei durch den Einbau von Digoxigenin-markierter dNTPs. Anschließend wurden die RNA-Sonden mittels Lithiumchlorid-Fällung gereinigt und in Prähybridisierungslösung aufgenommen.

Blotting und Detektion

Die Gesamt-RNA-Proben wurden für den Northern Blot zunächst in einem denaturierenden Urea-Polyacrylamidgel (12,5 %) nach ihrer Größe getrennt. Anschließend wurde das Gel in Ethidiumbromid gefärbt, in Klarsichtfolie gelegt und unter UV-Licht fotografiert. Die Übertragung der RNA auf eine Nitrocellulosemembran erfolgte anhand eines Semidry-Elektroblots. Hierzu wurde das Gel zwischen Whatman-Papiere und die Membran gelegt (Blotting-Sandwich) und durch anlegen von elektrischer Spannung auf die Membran übertragen. Anschließend wurde die Membran umgedreht und unter UV-Licht die Markerbanden mit Bleistift markiert. Die Membran wurde danach wieder umgedreht und dort die RNA-Banden mit einem Bleistift eingezeichnet. Vor der Hybridisierung der auf der Membran immobilisierten RNA mit den RNA-Sonden wurde eine Prähybridisierung durchgeführt, um unspezifische Bindungen zwischen Sonde und Membran auszuschließen. Die anschließende Hybridisierung mit dem RNA-Sonden erfolgte über Nacht bei Raumtemperatur.

Der Nachweis der Sonden-Hybridisierung erfolgte durch Inkubation von Anti-Digoxigenin-Antikörpern, die die alkalische Phosphatase am Fc-Teil tragen. Das künstliche Substrat CDP-

Star wird von der alkalischen Phosphatase, unter Lichtemission dephosphoryliert und kann mit einem Chemilumineszenz-Scanner aufgenommen werden.

3.12 RNA-Sequenzierung

3.12.1 Sequenzierung von Vollängentranskripten

Für die Sequenzierung von Vollängentranskripten wurden 2 μg Gesamt-RNA aus den jeweiligen RNA-Isolaten eingesetzt. Für die Sequenzierung von mRNA wurde die Ribosomale RNA mittels „Ribo-Zero rRNA Removal Kit (Bacteria)“ (Epicentre) nach Herstellerangaben abgereinigt. Anschließend wurde die mRNA mittels „RNA MinElute“-Säulen (Qiagen) aufgereinigt und die RNA-Qualität mittels „Agilent RNA Pico 6000 Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) überprüft. Die mRNA-Fragmentierung, reverse Transkription zu cDNA, Adenylierung von 3'-Enden, Adapterligation und PCR-Amplifikation wurden nach Anleitung des „TrueSeq Stranded mRNA library“ Protokolls (Illumina) durchgeführt. Vor der *paired-end* Sequenzierung der cDNA auf einer Illumina MiSeq bzw. HiSeq Maschine wurde die cDNA-Qualität mit dem „Agilent High Sensitivity DNA Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) überprüft.

Prinzip der Illumina RNA-Sequenzierung

Das Prinzip der RNA-Sequenzierung von Vollängentranskripten nach Illumina basiert auf drei Schritten: RNA-Fragmentierung, reverse Transkription der RNA in eine cDNA-Bibliothek und der cDNA-Sequenzierung durch Synthese (*sequencing by synthesis*) (Illumina, 2017).

Die aus den Zellen isolierte RNA wird zunächst von stabilen RNAs, wie rRNA, befreit, da diese meist nicht die zu untersuchenden Transkripte darstellen, aber gleichzeitig ca. 95 % der Gesamt-RNA einer Zelle ausmachen (Peano *et al.*, 2013). Die Abreicherung der rRNA basiert auf der komplementären Bindung von zur rRNA komplementären Sequenzen, die auf magnetischen Kügelchen (*beads*) immobilisiert sind. Anschließend wird die verbliebene RNA chemisch fragmentiert, da kürzere Fragmente geringere Sekundärstrukturen aufweisen und so später eine gleichmäßigere Read-Abdeckung in der Sequenzierung gewährleistet wird. Die reverse Transkription zu cDNA erfolgt über degenerierte Hexamer-Primer (*random hexamer primers*). Bei strangspezifischen cDNA-Bibliotheken wird für die Zweitstrangsynthese dUTP anstelle von dTTP eingesetzt. Das Vorhandensein von dUTP im Gegenstrang dämpft (*quenching*) die Amplifikation des Gegenstrangs in der späteren PCR, da die Polymerase nicht über dieses Nukleotid hinaus amplifizieren kann. Über das Anbringen eines zusätzlichen Adenosinrests an die 3'-Enden der *blunt end* und doppelsträngigen cDNA werden Chimären

3 Material und Methoden

bei der Adapter-Ligation verhindert, da die Adapter ein überhängenden Thyminrest aufweisen. Die Adapter enthalten Sequenzen für die spätere Bindung der cDNA an die *flow-cell*. Außerdem enthalten die Adapter Index-Sequenzen, sodass verschiedene cDNA-Bibliotheken parallel auf einer Sequenziermaschine sequenziert werden können (*multiplexing*). Im letzten Schritt der Vorbereitung der cDNA-Bibliothek wird eine PCR-Amplifikation durchgeführt, die DNA-Fragmente, die die Adaptersequenzen an beiden Enden besitzen, spezifisch vervielfältigt. Durch die PCR wird auch die für die Sequenzierung nötige Menge an DNA bereitgestellt (Illumina, 2017).

Für die Sequenzierung mittels Synthese (*sequencing by synthesis*) werden die cDNA-Fragmente der cDNA-Bibliothek an über die Adaptersequenzen an auf der *flow-cell* immobilisierte komplementäre DNA-Sequenzen hybridisiert. Über sogenannte *bridge amplification* werden lokal Gruppen (*cluster*) aus identischen cDNA-Fragmenten generiert, um das Signal in der anschließenden Polymerase-Reaktion zu verstärken. Die eigentliche Sequenzierung erfolgt durch Inkorporation von Fluoreszenz-markierten Nukleotiden, die bei Inkorporation ein für jede Base spezifisches Farbsignal abgeben. Diese Farbsignale werden über eine Kamera in ein digitales Bild der *flow-cell* umgewandelt, wobei jede Base (A, C, T oder G) ein spezifisches Farbsignal erzeugt. Über Zyklen aus Zugabe einzelner Nukleotide und anschließendem Spülen kann die Sequenz der cDNA-Fragmente über die zeitliche Aufnahme der Farbsignale ermittelt werden. Die Anzahl der Zyklen entspricht der Read-Länge (Illumina, 2017).

Nach der erfolgten Sequenzierung liegen die Sequenzen der cDNA-Fragmente in Form von Reads vor. Allerdings umfassen die Sequenzen der Reads nur die äußeren Enden der cDNA-Fragmente, da die Polymerase-Reaktion nicht über die volle Zahl an nötigen Zyklen qualitativ konstant bleibt. Aktuell sind Readlängen von 75 nt bei cDNA möglich. Je nach Sequenzierungsmethode werden *single-end* oder *paired-end* Reads erzeugt. *Single-end* Reads stammen nur von einem Ende des cDNA-Fragments, während *paired-end* Reads von beiden Enden des cDNA-Fragments stammen (Illumina, 2017). Über die Kenntnis beider Enden der cDNA-Fragmente, können die weitergehende Analysen durchgeführt werden, da sie mehr Informationen enthalten, als *single-end* Reads. Von dieser Eigenschaft von *paired-end* Reads wurde in dieser Arbeit mehrfach Gebrauch gemacht.

Die erhaltenen Reads werden gewöhnlich im Anschluss gegen eine genomische Referenzsequenz gemappt bzw. aligniert, um darauf aufbauende Analysen durchzuführen.

3.12.2 Sequenzierung von ribosomaler RNA

Die Sequenzierung von ribosomaler RNA wurde analog zur oben beschriebenen Sequenzierung von Vollängentranskripten durchgeführt. Allerdings wurde auf eine vorherige Abreicherung

der rRNA mittels „Ribo-Zero rRNA Removal Kit (Bacteria)“ (Epicentre) verzichtet. Aufgrund der hohen Abundanz von rRNA in bakteriellen Zellen, wurde wesentlich weniger RNA für die Erstellung der cDNA-Bibliotheken eingesetzt. Eine Menge von 300 ng RNA war ausreichend.

3.12.3 Sequenzierung der 5'-Enden von Primärtranskripten

Für die Sequenzierung der 5'-Enden von Primärtranskripten nach Pfeifer-Sancar *et al.* (2013) wurden 2x 5 µg Gesamt-RNA aus den jeweiligen RNA-Isolation eingesetzt.

Wie bei der Sequenzierung von Vollängentranskripten wird die rRNA mittels „Ribo-Zero rRNA Removal Kit (Bacteria)“ (Epicentre) aus der Gesamt-RNA abgereinigt. Anschließend wird die mRNA über Nacht mit Ethanol, Glycogen und Natriumacetat bei -80 °C präzipitiert. Am Folgetag wird die RNA auf erfolgreiche Abreicherung der rRNA mit „Agilent RNA Pico 6000 Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) kontrolliert, anschließend chemisch fragmentiert, über „RNA MinElute“-Säulen (Qiagen) gereinigt und nicht-primäre Transkripte (Transkripte mit 5'-Monophosphat oder 5'-Diphosphat) über die Terminator-Exonuklease (TEN) (Epicentre) abgebaut. Anschließend wird die TEN durch Phenol-Chloroform-Isoamylalkohol-(PCI)-Extraktion entfernt und die RNA mit „RNA MinElute“-Säulen (Qiagen) aufgereinigt. Die Triphosphate an den 5'-Enden von Primärtranskripten werden mittels RNA Pyrophosphohydrolase (RPP) (Epicentre) zu Monophosphaten umgewandelt. Es erfolgt erneut eine Ethanol-Präzipitation der RNA und im Anschluss die RNA-Adapter-Ligation mittels T4 Ligase (NEB). Überschüssiger Adapter wird durch Aufreinigung mit „RNA MinElute“-Säulen (Qiagen) entfernt. Für die reverse Transkription durch den Schlaufen-Adapter (*loop adapter*) muss dieser zunächst die korrekte Struktur annehmen. Dazu wird der Schlaufen-Adapter in einem Thermocycler zunächst aufgeköcht und anschließend mit einer Geschwindigkeit von 1 °C pro 10 s auf 25 °C abgekühlt. Die reverse Transkription über den Schlaufen-Adapter wird von der Superscript II RT (Thermo Fischer) katalysiert und die RNA danach aus dem RNA-DNA-Hybrid über Verdau mit RNase H (Thermo Fischer) entfernt. Anschließend erfolgt die Zweitstrangsynthese und Amplifikation mittels PCR. Dazu werden Primer eingesetzt, die für die Illumina Sequenzierung kompatible Enden aufweisen und auf den ligierten RNA- und Schlaufen-Adapter binden. Die in dieser Arbeit verwendeten Primer sind in Tabelle 3.21 aufgeführt. Die PCR erfolgt über die Phusion Polymerase (Thermo Fischer). Nach der PCR werden die PCR-Produkte mit „Agencourt AMPure XP Beads“ aus dem „SMARTer Stranded RNA Kit“ (Beckman Coulter) aufgereinigt. Anschließend wird die Qualität der erstellten cDNA-Bibliothek mit dem „Agilent High Sensitivity DNA Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) überprüft. Um verbliebene Primer oder Adapter zu entfernen werden die PCR-Produkte mit Größen zwischen 150 – 1.000 bp nach

3 Material und Methoden

der Aufreinigung präparativ aus einem Agarosegel mit „Ultra Low Range Agarose“ (Bio-Rad) ausgeschnitten und die cDNA-Bibliothek anschließend mit dem Illumina MiSeq oder HiSeq 1500 sequenziert.

Das vollständige Protokoll ist in Pfeifer-Sancar *et al.* (2013) detailliert beschrieben. Im Rahmen dieser Arbeit wurden kleine Anpassungen am Protokoll vorgenommen: Nicht-primäre Transkripte wurden mittels Terminator-Exonuklease (Epicentre) bei 30 °C für 60 min und anschließend 42 °C für 30 min verdaut. Die Adapterligation am 5'-Ende wurde für 60 min bei 30 °C mit 1 μ L 60 μ M Adapter durchgeführt. Nach der Amplifikation der cDNA wurden die beiden cDNA-Bibliotheken mittels Gelelektrophorese aufgereinigt und nach Größe selektioniert. Es wurden nur Fragmente mit einer Größe zwischen 100 und 1.000 bp aus dem Gel ausgeschnitten. Die untere Grenze von 100 bp wurde gewählt, um Adapterdimere abzureinigen. Da zwei Adapter, mit einer Gesamtlänge von 66 nt, für die Erstellung der cDNA-Bibliotheken benutzt wurden, wurden nur Transkripte mit einer Länge kleiner als 44 nt nicht sequenziert. Die Sequenzierung der 5'-Enden von Primärtranskripten erfolgt im *single read* Modus mit einer Readlänge von 75 nt. Die verwendeten Oligonukleotide sind in Tabelle 3.21 aufgeführt.

Tabelle 3.21: Oligonukleotide für die RNA-Sequenzierung. Das Zeichen \leftrightarrow zeigt einen Zeilenumbruch an.

Oligonukleotid	Sequenz (5' - 3')
RNA-Adapter (aus RNA)	CCCUACACGACGCUCUCCGAUCGAG
Schlaufen-Adapter	AGATCGGAAGAGAGACGTGTGCTCTTCCGATCTNNNNNNN
Amplifikations-Primer 1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC \leftrightarrow GCTCTCCGATCGAG
Amplifikations-Primer 2	CAAGCAGAAGACGGCATACGAGATGCCTAAGTACTGGAGTTCAG \leftrightarrow ACGTGTGCTCTTCCGATCT

3.12.4 Sequenzierung von RNase-Schnittstellen mittels *RNase-Assay*

Im Rahmen dieser Arbeit wurde eine Methode zur Identifizierung von RNase-Schnittstellen mittels RNA-Sequenzierung entwickelt. Die Methode wurde mit *RNase-Assay* benannt und wird in *in vivo* und *in vitro RNase-Assay* unterschieden. Beide Typen der *RNase-Assays* basieren auf der Markierung von RNase-Schnittstellen durch Ligation eines RNA-Adapters. Für die Identifizierung der Schnittstellen wird eine behandelte und eine unbehandelte Probe (Kontrolle) benötigt. Bei *in vivo RNase-Assay* ist die behandelte Probe die RNA des Wildtyps, da dieser alle RNasen besitzt und entsprechend geschnittene Transkripte aufweist. Die Kontrolle ist die RNA einer RNase-defizienten Mutante (Deletions- oder *knock down*-Mutante), da dieser RNA

die Schnittstellen der zu untersuchenden RNase fehlen. Bei *in vitro RNase-Assay* wird nur die RNA einer RNase-Mutante eingesetzt, diese aber einmal mit und einmal ohne (Kontrolle) der zu untersuchenden RNase *in vitro* behandelt. Dazu muss zum einen die zu untersuchende RNase in dem Organismus deletierbar sein als auch aufgereinigt werden können. Über den Vergleich der behandelten Probe mit der Kontrolle werden die RNase-Schnittstellen ermittelt. Die dazu nötigen bioinformatischen Prozesse werden weiter unten in einem separaten Abschnitt aufgeführt.

Im Folgenden wird der Ablauf der *RNase-Assays* beschrieben. Auf der dieser Arbeit beiliegenden Daten-CD befinden sich die Protokolle mit detaillierten Angaben zur Durchführung. Hintergründe zur Entwicklung, ein grafisches Ablaufschema, den erzielten Ergebnissen und der Ergebnisdiskussion werden in den nachfolgenden Kapiteln beschrieben.

Der *in vivo RNase-Assay*

Für den *in vivo RNase-Assay* wird die RNA des Wildtyps (als Kontrolle), sowie die RNA einer RNase-defizienten Mutante benötigt. Beide RNA-Proben sind separat zu behandeln und dürfen nicht vermischt werden. Der *in vivo RNase-Assay* beginnt mit der Abreicherung der rRNA aus der isolierten Gesamt-RNA mittels „Ribo-Zero rRNA Removal Kit (Bacteria)“ (Epicentre). Falls Schnittstellen innerhalb der rRNA untersucht werden sollen, wird die rRNA-Abreicherung übersprungen. Nach erfolgter rRNA-Abreicherung wird die verbliebene RNA über Nacht mit Ethanol präzipitiert, am nächsten Tag gereinigt und die rRNA-Abreicherung mit „Agilent RNA Pico 6000 Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) überprüft. Anschließend erfolgt der wichtigste Schritt: das Markieren der RNase-Schnittstellen durch Ligation eines RNA-Adapters. Dazu werden etwaige Sekundärstrukturen zunächst über Hitze denaturiert und die Probe direkt auf Eis gestellt. Anschließend wird die Ligation mit dem RNA-Adapter (Tabelle 3.21) durchgeführt. Anschließend wird überschüssiger RNA-Adapter über „RNA MinElute“-Säulen (Qiagen) entfernt, die RNA chemisch fragmentiert und erneut über „RNA MinElute“-Säulen (Qiagen) aufgereinigt.

Die folgenden Schritte sind mit der Methode zur Sequenzierung von Primärtranskripten (Pfeifer-Sancar *et al.*, 2013) ab dem Schritt zur Vorbereitung des Schlaufen-Adapters identisch.

Der *in vitro RNase-Assay*

Für den *in vitro RNase-Assay* wird die RNA einer RNase-defizienten Mutante benötigt, da das Einbringen der RNase-Schnittstellen *in vitro* mit aufgereinigter RNase erfolgt. Bei der aufgereinigten RNase muss es sich um die selbe RNase handeln, dessen Gen in der Mutante deletiert wurde. Die RNA der RNase-defizienten Mutante wird in einer Probe mit der aufgerei-

3 Material und Methoden

nigten RNase behandelt und einmal unbehandelt (als Kontrolle) weiterverarbeitet. Im ersten Schritt des *in vitro* RNase-Assays wird die rRNA mit „Ribo-Zero rRNA Removal Kit (Bacteria)“ (Epicentre) abgereinigt, mit Ethanol präzipitiert und die Abreicherung mittels „Agilent RNA Pico 6000 Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) kontrolliert. Falls RNase-Schnittstellen in der rRNA untersucht werden sollen, wird die rRNA-Abreicherung übersprungen.

Anschließend wird die RNA der RNase-defizienten Mutante mit der jeweiligen (aufgereinigten) RNase behandelt. Die Inkubationsdauer ist abhängig von der RNase-Aktivität und sollte vorher ermittelt werden. Als Kontrolle dient RNA die nicht mit der RNase behandelt wurde. Um die Vergleichbarkeit zwischen behandelter und unbehandelter Probe zu erhöhen, sollte die unbehandelte RNA allerdings mit möglichst ähnlichen Bedingungen (Puffer, Temperatur, Dauer) inkubiert werden. Anschließend wird die RNase der behandelten Probe mittels PCI-Reagenz (Carl Roth) entfernt und die RNA mit Ethanol präzipitiert. Für die Vergleichbarkeit wird die unbehandelte Probe ebenfalls mit PCI-Reagenz behandelt und mit Ethanol präzipitiert. Falls möglich kann der Effekt der RNase-Behandlung bereits im „Agilent RNA Pico 6000 Kit“ und dem „Agilent 2100 Bioanalyzer“ (Agilent Technologies) beobachtet werden.

Im Folgenden werden in beiden Proben die Schnittstellen über Ligation eines RNA-Adapters (Tabelle 3.21) markiert. Alle Schritte nach der Adapter-Ligation sind mit denen des *in vivo* RNase-Assays identisch.

3.13 Bioinformatische Analysen der RNA-Sequenzierungsdaten

3.13.1 Qualitätskontrolle und Read-Mapping

Nach Erhalt der Daten aus der RNA-Sequenzierung wurden die Daten auf Qualität überprüft, bevor sie weiterverarbeitet wurden. Qualitätsparameter sind vor allem die mittlere Qualität der Basen (Phred-Scores) über alle Reads, der zu erwartende GC-Gehalt der Reads und die mittlere Basenhäufigkeit für jede Read-Base. Diese Qualitätsparameter wurden mit dem Programm *FastQC*¹ (Version 0.11.5) berechnet. Generell wurden folgende Werte für die Parameter als Mindestmaß für die Sequenzierungsdaten festgelegt:

- Mittlere Basenqualität Q (Phred-Score) von mindestens 30 (entspricht einem Fehler auf 1.000 bp)
- Erwarteter GC-Gehalt entsprechend dem GC-Gehalt des sequenzierten Organismus (um 53 % für RNA-Sequenzierungen von *Corynebacterium glutamicum*)

¹Simon Andrews, *Bioinformatics Group, Babraham Institute*, Babraham, Großbritannien (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)

3.13 Bioinformatische Analysen der RNA-Sequenzierungsdaten

- Mittlere Basenhäufigkeit entsprechend dem GC-/AT-Gehalt des sequenzierten Organismus

Von allen im Rahmen dieser Arbeit durchgeführten RNA-Sequenzierungen erfüllten die Daten die jeweiligen Qualitätsparameter. Da die Basenqualität technisch bedingt zum 3'-Ende des Reads abnimmt, wurden die Reads im nächsten Schritt hinsichtlich der Qualität getrimmt, um möglichst nur Reads mit hoher Basenqualität zu mappen. Das trimmen wurde mit dem Programm *trimmomatic* (Version 0.36) (Bolger *et al.*, 2014) in Standardeinstellung für *paired end* Reads durchgeführt. Für die Sequenzierung von Primärtranskripten wurde nur vom 3'-Ende getrimmt, um die Information am 5'-Ende zu erhalten. In beiden Fällen wurden Reads, die nach dem trimmen kürzer als 39 nt waren, verworfen.

Das Programm *MultiQC* (Ewels *et al.*, 2016) wurde genutzt, um die Ausgaben der Programme *FastQC*, *bowtie2* und *trimmomatic* für jede verarbeitete cDNA-Bibliothek übersichtlich und vergleichend darzustellen.

Für die korrekte Darstellung der *paired end* Reads in *ReadXplorer* (Hilker *et al.*, 2014) wurden die Reads vom Gegenstrang (Datei `R1.fastq`) vor dem *mapping* mit dem Programm *FASTQ/A Reverse-Complement* aus der Sammlung *FASTX Toolkit*² (Version 0.0.13.2) revers komplementiert. Für das anschließende *mapping*, also der Lokalisation der Reads auf dem Genom mittels Übereinstimmung (*alignment*), wurden, je nach Anforderung, die Programme *bowtie2* (Version 2.3.0) (Langmead und Salzberg, 2012) oder *segemehl* (Version 0.2.0) (Hoffmann *et al.*, 2009) verwendet. Für generelles *mapping* von Vollängentranskript- sowie Primärtranskript-Reads wurde *bowtie2* in Standardeinstellungen mit Vorwärts-Orientierung für Folge- und Gegenstrang-Reads eingesetzt, da dieser in der Laufzeit einerseits sowie in der Qualität des *mappings* andererseits sehr gute Eigenschaften aufweist.

Für die Versuche zur Identifizierung von RNase-Schnittstellen mittels *RNase-Assay* wurde eine höhere Präzision der Read-Lokalisation benötigt und daher *segemehl* eingesetzt. Dieses Programm bietet die Angabe eines festen Schwellenwerts für die Übereinstimmung zwischen Reads und Referenzgenom, unabhängig von der Readlänge. Jeder Read musste mindestens 95 % Übereinstimmung zum Referenzgenom aufweisen. Außerdem wurden nur Reads mit eindeutiger Übereinstimmung zugelassen. Reads mit mehrdeutigen Übereinstimmungen (*multiple mapped reads*) im Genom wurden verworfen.

²Hannon Lab, Cold Spring Harbor Laboratory, New York, USA (http://hannonlab.cshl.edu/fastx_toolkit/index.html)

3.13.2 Bestimmung differenziell transkribierter Gene

Da die Anzahl der Reads eines Gens mit der Transkriptionsstärke des Gens korreliert (Mor-tazavi *et al.*, 2008), wird zur Bestimmung der differenziell transkribierten Gene aus RNA-Sequenzierungen die Anzahl der Reads pro Gen und Bedingung benötigt. Gewöhnlich sind dies bei prokaryotischen Datensätzen die Reads die innerhalb einer Proteinkodierregion (CDS) liegen. Wenn nicht anders angegeben, wurden drei biologische Replikate pro Bedingung verwendet und davon die Reads pro Gen bestimmt. Die Reads wurden entweder mit *ReadXplorer* (Hilker *et al.*, 2014) als Einzel-Reads (*single reads*) oder mit *featureCounts* (Liao *et al.*, 2013) als gepaarte Reads (*paired reads*) gezählt.

Anschließend werden mittels statistischer Verfahren die Reads normalisiert und geprüft, welche Gene in den zu untersuchenden Bedingungen differenziell transkribiert sind. In dieser Arbeit wurde dazu die Software *DESeq2* (Love *et al.*, 2014) verwendet. Als Ergebnis gibt *DESeq2* für jedes Gen die relative Transkriptmenge (in normalisierten Reads) sowie die Änderung der Transkriptmenge (*foldChange*) bezogen auf die Referenz als Logarithmus zur Basis 2 ($\log_2\text{foldChange}$ LFC), aus. Als Referenz wird meist die unbehandelte Bedingung bzw. bei Mutanten der Wildtyp herangezogen.

Da für jedes Gen einzeln geprüft wird, ob es differenziell transkribiert ist und Organismen mehrere Tausend Gene besitzen, reicht die einfache Prüfung auf statistische Signifikanz mittels p-Wert nicht aus. Würde zum Beispiel ein p-Wert von 0,05 als Signifikanzgrenze angenommen, würden bei einem Organismus mit 3000 Genen 150 Gene (5 %) fälschlicherweise als differenziell transkribiert bestimmt werden (Alphafehler-Kumulierung bzw. *multiple testing problem*). Um diese *False Discovery Rate* (FDR) zu korrigieren, werden von *DESeq2* zusätzliche statistische Tests durchgeführt. Der FDR-korrigierte bzw. angepasste p-Wert (*adjusted p-value*) gibt somit die statistische Signifikanz des $\log_2\text{foldChange}$ an (Love *et al.*, 2014). Im Rahmen dieser Arbeit wurde ein angepasster p-Wert von 0,05 als statistische Signifikanzgrenze für differenziell transkribierte Gene angenommen. Darüber hinaus wurde zusätzlich nach dem $\log_2\text{foldChange}$ (LFC) gefiltert, so dass nur Gene mit einem angepassten p-Wert kleiner als 0,05 und einem LFC von $\geq +1,0$ bzw. $\leq -1,0$ als differenziell transkribiert gelten. Falls andere, als die oben genannten, Kriterien angesetzt wurden, wurde dies im Text deutlich gekennzeichnet.

Neben den Kriterien zur Bestimmung von differenziell transkribierten Genen wurden außerdem die Replikate der untersuchten Bedingungen auf ihre Ähnlichkeit bzw. Ausreißer mittels hierarchischem Clustering oder linearer Regression überprüft.

3.13.3 Identifizierung von Transkriptionsstarts

Für die automatische Detektion von Transkriptionsstarts (*transcription start site*, TSS) aus den Primärtranskript-Datensätzen wurde die „TSS detection“ der Software *ReadXplorer* (Hilker *et al.*, 2014) (Version 2.2.3) eingesetzt. Nach anfänglichen empirischen Tests mit verschiedenen Parametereinstellungen und darauf aufbauenden Inspektionen der TSS-Daten wurden, ausgehend von der durch die Software angebotenen automatischen Parameterabschätzung, die Parameter für die Detektion von putativen TSS mit *ReadXplorer* festgelegt. Für den Primärtranskript-Datensatz aus *Corynebacterium diphtheriae* gilt: ein putativer TSS wurde detektiert, wenn an einer Position im Genom ein minimaler Anstieg der Coverage von 100 % ausgehend von mindestens 28 Readstarts vorhanden war. Für die TSS-Detektion aus den Primärtranskript-Datensatz aus *Corynebacterium glutamicum Δrho* gelten identische Bedingungen, wobei hier mindestens neun Readstarts vorhanden sein mussten.

3.13.4 Identifizierung von neuen Transkripten

Mit Hilfe der TSS-Daten aus den Primärtranskript-Datensätzen von *C. diphtheriae* ist es möglich neue, bisher nicht annotierte, Transkripte zu identifizieren. Ein TSS wurde einem annotierten Gen zugeordnet, wenn der TSS maximal 500 bp *upstream* vom Startcodon des jeweiligen Gens entfernt liegt. Alle TSS die keinem bekannten Gen zugeordnet werden konnten, wurden neuen Transkripten zugeordnet. Um die Klasse dieser TSS weiter zu charakterisieren, wurden die TSS in drei Gruppen unterteilt: (a) intergenische TSS; TSS die zwischen zwei annotierten Genen liegen, (b) intragenische TSS; TSS die innerhalb von annotierten Genen bzw. Kodierregionen liegen und (c) *antisense* TSS; TSS die auf dem Gegenstrang zu einem Gen bzw. Kodierregion liegen.

Die Transkriptlänge von neuen Transkripten, die für putative Proteine kodieren, wurde über die Read-Abdeckung aus den Vollängentranskript-Datensätzen des Wildtyps abgeschätzt. Diese mit Reads abgedeckten Regionen wurden anschließend nach offenen Leserahmen (*open reading frames*, ORFs) mit dem Programm UGENE (Okonechnikov *et al.*, 2012) mit AUG, GUG, UUG und UTG als möglichen Startcodons abgesucht. Zusätzlich musste das Stopcodon ebenfalls in der abgedeckten Region liegen. Die vorhergesagten ORFs wurden mittels NCBI BLAST (Altschul *et al.*, 1997, 2005) nach Proteinhomologen untersucht. Konnte kein ORF oder Proteinhomolog gefunden werden, wurde die Sequenz der abgedeckten Region mit dem Programm *Infernal* (Nawrocki und Eddy, 2013) auf potenzielle Homologe von nicht-kodierende RNAs (ncRNAs) und RNA-Motive in der RFAM-Datenbank (Nawrocki *et al.*, 2015) untersucht. Konnte ein putatives Protein- oder ncRNA-Homolog für ein neues Transkript

3 Material und Methoden

vorhergesagt werden, wurde das neue Transkript mit einem eigenen Locus-Namen der Form DIPXXXX (mit XXXX als eindeutige Nummer) versehen.

3.13.5 Identifizierung von Operonstrukturen

Zwei oder mehr Gene sind transkriptionell verbunden und Teil eines Operons, wenn sie von einem gemeinsamen Promotor transkribiert werden. Die Detektion von Operonstrukturen in *C. diphtheriae* wurde mit ReadXplorer (Hilker *et al.*, 2014) durchgeführt. Dazu wurden die *paired-end* Reads (mit einer minimalen Mapping-Qualität ≥ 30) im Vollängentranskript-Datensatz des Wildtyps gezählt, die zwei Gene bzw. Kodierregionen überspannen. Wenn mehr als fünf überspannende *paired-end* Reads gefunden wurden, wurden diese Gene einem Primäroperon zugeordnet. Der Prozess wurde für die nachfolgenden Gene wiederholt bis dem Operon keine weiteren Gene zugeordnet werden konnten. Um die Zahl der *paired-end* Reads in Regionen mit schwacher Read-Abdeckung zu erhöhen, wurden die drei Replikate der Vollängentranskript-Datensätze des Wildtyps zu einem kombiniert. ReadXplorer kann nur bereits annotierte Gene zu Operons zuordnen. Gene die durch neue, noch nicht annotierte, Transkripte kodiert werden (zum Beispiel tmRNA) wurden daher manuell auf überlappende *paired-end* Reads geprüft und gegebenenfalls neuen oder bestehenden Operons zugeordnet.

Die Liste der Primäroperons wurde in zwei Gruppen unterteilt: Operons denen ein TSS zum ersten Gen zugeordnet werden konnte (experimentell validiertes Operon) und Operons denen kein TSS zum ersten Gen zugeordnet werden konnte (vorhergesagtes Operon).

Darüber hinaus wurden Primäroperons in Sub-Operons unterteilt, wenn ein TSS für ein hinteres Gen (ab dem 2. Gen im Primäroperon) detektiert wurde. Dieser TSS zeigt einen weiteren Promotor an, der zu einem, im Vergleich zum Primäroperon, verkürzten Transkript führt.

Alle Gene die nicht mittels überlappender *paired-end* Reads einem Operons zugeordnet werden konnten, wurden als monocistronische Transkripte klassifiziert.

3.13.6 Identifizierung von Sequenzmotiven

Um konservierte Sequenzmotive, wie zum Beispiel Promotoren oder RNase-Schnittstellen, im Genom zu identifizieren, wurde das Programm *Improbizer* (Ao *et al.*, 2004) verwendet. Die damit vorhergesagten Motive wurden anschließend mit dem Programm *WebLogo 3* (Crooks, 2004) visualisiert. Abhängig von der angenommenen Position des Motivs wurden unterschiedliche DNA-Sequenzen für die Identifizierung verwendet.

Für die Identifizierung von sogenannten CG-Blasen bzw. *rut sites* (*Rho utilization sites*), die eine Rho-abhängige Transkripttermination andeuten können, wurde der Gehalt an C bzw. G

3.13 Bioinformatische Analysen der RNA-Sequenzierungsdaten

mit Hilfe der Software *EMBOSS FREAK* (Rice *et al.*, 2000) in Standardeinstellungen (Fensterbreite 30 nt, Schrittbreite 1 nt) bestimmt.

3.13.7 Identifizierung von verlängerten Transkripten

Im Vergleich zum *C. glutamicum* Wildtyp besitzt die $\Delta\rho$ -Mutante verlängerte Transkripte. Für eine Analyse dieser verlängerten Transkripte wurde eine genomweite Identifizierung der verlängerten Transkripte aus den RNA-Sequenzierungsdaten durchgeführt. Dazu wurde die, auf die Read-Gesamtmenge normalisierte, Read-Abdeckung der Replikate des Wildtyps mit der der $\Delta\rho$ -Mutante verglichen. Überschreitet das Verhältnis der Read-Abdeckung der $\Delta\rho$ -Mutante im Vergleich zum Wildtyp in einem *Sliding Window*-Bereich L am Ende eines transkribierten Bereichs einen festgelegten Faktor F , liegt ein verlängertes Transkript in der $\Delta\rho$ -Mutante vor.

Die Breite des *Sliding Window*-Bereichs L und die minimale Höhe des Verhältnis F der Read-Abdeckung wurden anhand einer Liste mit manuell klassifizierten verlängerten Transkripten auf optimale Werte getestet. Die Werte, mit denen die höchste Sensitivität erreicht werden konnte, wurden für die genomweite Identifizierung der verlängerten Transkripte übernommen. Die optimalen Werte sind ein Faktor F von 5 für das Verhältnis der Read-Abdeckung sowie ein Wert von 75 bp für die Breite des *Sliding Window* L .

Die so identifizierten verlängerten Transkripte wurden anschließend anhand des *upstream* befindlichen Features in die Kategorien (a) CDS, (b) intragenisches Transkript, (c) *antisense* RNA oder (d) nicht-kodierende RNA (ncRNA) eingeteilt.

Das Programm zur genomweiten automatischen Identifizierung von verlängerten Transkripten wurde vom Autor dieser Dissertation konzipiert. Die Programmierung wurde von David Brandt in der Sprache *Python* (Version 2.7) durchgeführt.

3.13.8 Identifizierung von RNase-Schnittstellen im *RNase-Assay*

Im Folgenden wird der bioinformatische Auswertung der im *RNase-Assay* erhaltenen Reads für die Identifizierung von RNase-Schnittstellen beschrieben.

Im Idealfall sollte in der Probe des Wildtyps (*in vivo RNase-Assay*) bzw. in der mit RNase behandelten Probe (*in vitro RNase-Assay*) ein Schnittstellen-Signal vorliegen, während es in der jeweiligen Kontrolle nicht vorliegen sollte. Das Signal sollte demnach ein Read-Stapel, der vergleichbar mit der eines Read-Stapels der einen TSS in den Primärtranskriptdatensätzen anzeigt. In der Kontrolle hingegen sollte an der selben Position kein Read-Stapel vorhanden sein. Dennoch kann die Identifikation von Read-Stapeln in den Sequenzdaten der *RNase-Assays* nicht über eine TSS-Detektion mit *ReadXplorer* (Hilker *et al.*, 2014) ablaufen, da die

3 Material und Methoden

Analyse der Signale nicht innerhalb eines Datensatzes, sondern über den Vergleich von zwei Datensätzen erfolgt.

Nach erfolgter Sequenzierung werden die aus den *RNase-Assays* erhaltenen Reads je Probe bzw. Bedingung wie im Abschnitt „Qualitätskontrolle und Read-Mapping“ beschrieben auf Qualität überprüft, vom 3'-Ende aus getrimmt und anschließend gegen das *C. glutamicum* ATCC 13032 Genom gemappt. Allerdings wird für das Mapping nicht *bowtie2* (Langmead und Salzberg, 2012), sondern *segemehl* (Version 0.2.0) (Hoffmann *et al.*, 2009) verwendet, da diese Software eine Angabe der Anzahl an erlaubten Fehlern in Prozent bei unterschiedlichen Read-Längen erlaubt. Im Gegensatz dazu erlaubt *bowtie2* dies nur bei Reads mit identischer Länge. Die Ausgabe von *multi reads* wurde beim Mapping mit *segemehl* deaktiviert. Anschließend werden die Reads künstlich auf eine Länge von 1 bp gekürzt, sodass nur noch die äußerste Base am 5'-Ende erhalten bleibt, und die Read-Anzahl an jeder Position im Genom strangspezifisch mit *featureCounts* (Liao *et al.*, 2013). Für das 3,3 Mbp große *C. glutamicum* Genom wurden demnach die Reads an 6,6 Millionen Positionen ($2 \cdot 3,3$ Mbp) gezählt. Die Software *featureCounts* bietet einen Parameter (`-read2pos 5`), der die künstliche Verkürzung der Reads direkt vor dem Zählen durchführt. Um die Datenmenge zu verringern, werden die Positionen, an denen unter beiden Bedingungen (behandelt und Kontrolle) keine Reads vorkommen (Read-Anzahl = 0) aus dem Datensatz entfernt.

Anschließend werden die Reads mittels *DESeq2* (Love *et al.*, 2014) normalisiert und auf ihren differentiellen Status für jede Genomposition überprüft. Für den *in vivo RNase-Assay* werden die Reads der RNase-Mutante gegen die des Wildtyps (Kontrolle) verglichen. Im *in vitro RNase-Assay* werden die Reads der mit RNase behandelten RNA der RNase-Mutante gegen die Reads der unbehandelten RNA verglichen. Positionen für die eine mindestens dreifache Änderung (*foldChange*) in der Read-Anzahl und einen angepassten p-Wert von $< 0,05$ besitzen, werden als mögliche RNase-Schnittstelle identifiziert. Anschließend wird aus der erhaltenen Liste an RNase-Schnittstellen die die Positionen entfernt, die einem Transkriptionsstart entsprechen, da hier keine RNase-Schnittstelle vorliegen kann. Daraus resultiert die finale Liste an identifizierten RNase-Schnittstellen, die mittels *RNase-Assay* identifiziert wurden.

4 Ergebnisse

Neue Entwicklungen im Bereich der RNA-Sequenzierung ermöglichen eine Analyse von Transkriptomen in bisher nicht gekanntem Detailgrad und Umfang. Im Rahmen dieser Arbeit wurden mit Hilfe der RNA-Sequenzierung die Transkriptome von *Corynebacterium diphtheriae* und *Corynebacterium glutamicum* bezüglich der RNA-Synthese und des RNA-Abbaus untersucht. Die dabei verfolgten Ziele waren die Analyse der Transkriptionsinitiation über die Identifizierung der 5'-Enden von Primärtranskripten des Humanpathogens *C. diphtheriae* sowie die Charakterisierung der Rho-abhängigen Transkriptionstermination und der Transkript-Degradation im biotechnologisch relevanten Modellorganismus *C. glutamicum*. Entsprechend dieser drei formulierten Ziele werden die jeweiligen Ergebnisse in drei Abschnitten präsentiert. Der erste Abschnitt behandelt die Ergebnisse der Primärtranskript-Analysen und der Charakterisierung des DtxR-Regulons in *C. diphtheriae*. Anschließend werden die Ergebnisse von Rho-abhängig terminierten Transkripten in *C. glutamicum* beschrieben. Der letzte Abschnitt behandelt die Charakterisierung zweier an der RNA-Degradation beteiligter RNasen.

4.1 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

In diesem Abschnitt werden die Resultate der Analyse der Transkriptionsinitiation in *C. diphtheriae* vorgestellt. Die Transkriptionsinitiation wurde anhand der spezifischen Sequenzierung von Primärtranskriptenden zur Identifikation von Transkriptionsstarts (TSS) eingesetzt. Aufbauend auf den TSS-Daten konnten so Promotormotive, 5'-UTRs und weitere genetische Elemente für *C. diphtheriae* analysiert und charakterisiert werden. Darüber hinaus wurde das Regulon des Diphtherietoxin-Repressors DtxR über differenzielle Transkriptionsanalysen untersucht.

Die in diesem Abschnitt gezeigten Ergebnisse und Diagramme sind in die Publikation von Wittchen *et al.* (2018)¹ übernommen worden.

¹Die Publikation kann über folgenden Link bezogen werden: <https://doi.org/10.1186/s12864-018-4481-8>

4 Ergebnisse

4.1.1 Probengenerierung und Sequenzierung der *C. diphtheriae* cDNA-Bibliotheken

Für die Analyse der Transkriptionsinitiation in *C. diphtheriae* NCTC 13129 mittels RNA-Sequenzierung wurde die RNA aus dem in HIB-Medium wachsenden Wildtyp in der exponentiellen Wachstumsphase isoliert und entsprechend dem Protokoll zur Sequenzierung von Primärtranskripten (Pfeifer-Sancar *et al.*, 2013) in eine Primärtranskript-cDNA-Bibliothek umgewandelt. Die Bibliothek wurde auf einem Illumina MiSeq mit einer Read-Länge von 75 nt im *single-end* Modus sequenziert und anschließend mit *bowtie2* (Langmead und Salzberg, 2012) gegen das *C. diphtheriae* NCTC 13129 Referenzgenom (Cerdeño-Tárraga *et al.*, 2003) gemappt. Über 97 % der ca. 2,8 Mio sequenzierten Reads konnten auf das Genom gemappt werden (Tabelle 4.1).

Tabelle 4.1: Sequenzierte und gemappte Reads der cDNA-Bibliotheken des *C. diphtheriae* Wildtyps und der $\Delta dtxR$ -Mutante.

cDNA-Bibliothek	Read-Typ	Gesamtzahl der Reads	Anzahl der gemappten Reads	Gemappte Reads in %
Primärtranskripte	<i>single-end</i>	2.812.117	2.742.095	97,51
Wildtyp Replik 1	<i>paired-end</i>	3.402.352	3.367.307	98,97
Wildtyp Replik 2	<i>paired-end</i>	2.009.475	1.960.443	97,56
Wildtyp Replik 3	<i>paired-end</i>	2.534.579	2.434.970	96,07
$\Delta dtxR$ -Mutante Replik 1	<i>paired-end</i>	2.945.471	2.920.139	99,14
$\Delta dtxR$ -Mutante Replik 2	<i>paired-end</i>	2.550.287	2.458.986	99,42
$\Delta dtxR$ -Mutante Replik 3	<i>paired-end</i>	2.231.733	2.221.243	98,53

Die Deletion des für den Diphtherietoxin-Regulator (DtxR) kodierenden Gens *dtxR* erfolgte über das *pK18mobsacB*-System (Schäfer *et al.*, 1994) und wurde von Andrew H. Gaspar (*Department of Molecular Biology and Biophysics, University of Connecticut Health Center, Farmington, CT, USA.*) durchgeführt. Die damit erzeugte *C. diphtheriae* $\Delta dtxR$ -Mutante wurde mittels Sequenzierung auf Korrektheit überprüft. Anschließend wurde die $\Delta dtxR$ -Mutante sowie der Wildtyp in Triplikaten in HIB-Medium bis zur exponentiellen Wachstumsphase angezogen, die Zellen geerntet und anschließend die RNA isoliert. Die isolierte RNA wurde über das *TrueSeq stranded mRNA* Protokoll (Illumina) zu strangspezifischen Vollängen-cDNA-Bibliotheken umgewandelt und auf einem Illumina MiSeq im *paired-end* Modus sequenziert. Die erhaltenen Reads mit einer Länge von 75 nt wurden mit *bowtie2* (Langmead und Salzberg, 2012) gegen das *C. diphtheriae* Referenzgenom gemappt. Wie auch bei der Primärtranskript-Bibliothek wurden über 96 % der sequenzierten Reads gemappt (Tabelle 4.1).

4.1.2 Identifizierung von Transkriptionsstarts

Die Analyse der Primärtranskript-cDNA-Bibliothek zur Detektion von Transkriptionsstarts (*Transcription Start Site*, TSS) wurde mit dem Programm *ReadXplorer* (Hilker *et al.*, 2014) durchgeführt. Über die automatische TSS-Detektion wurden 3.987 putative TSS detektiert. Nach manueller Inspektion der automatisch detektierten TSS wurden 2.310 falsch-positive TSS sowie 21 TSS, die rRNA und tRNA zugeordnet wurden, entfernt. Somit wurden 1.656 manuell inspizierte korrekte TSS identifiziert. Diese TSS wurden in zwei Hauptkategorien unterteilt: TSS die annotierten Genen zugeordnet werden können und TSS die neuen, bisher unbekannt, TSS zugeordnet wurden. Die zu bekannten Genen zugeordneten TSS wurden weiter unterteilt, je nachdem ob die Gene einen oder mehrere TSS aufwiesen. So wurden 874 Gene mit nur einem einzigen TSS (Einzel-TSS) und 137 Gene mit mehreren TSS (insgesamt 328 TSS) identifiziert (Abbildung 4.1). TSS die zu neuen Transkripten und somit zu neuen Genen gehören, wurden in drei Unterkategorien unterteilt: (a) intergenische TSS bei denen sich das neue Transkript zwischen zwei bekannten Genen befindet, (b) intragenische TSS bei denen sich das neue Transkript innerhalb der annotierten Kodierregion eines bekannten Gens befindet und (c) *antisense* TSS bei denen sich das neue Transkript auf dem Gegenstrang zu einem bereits bekannten Gen befindet.

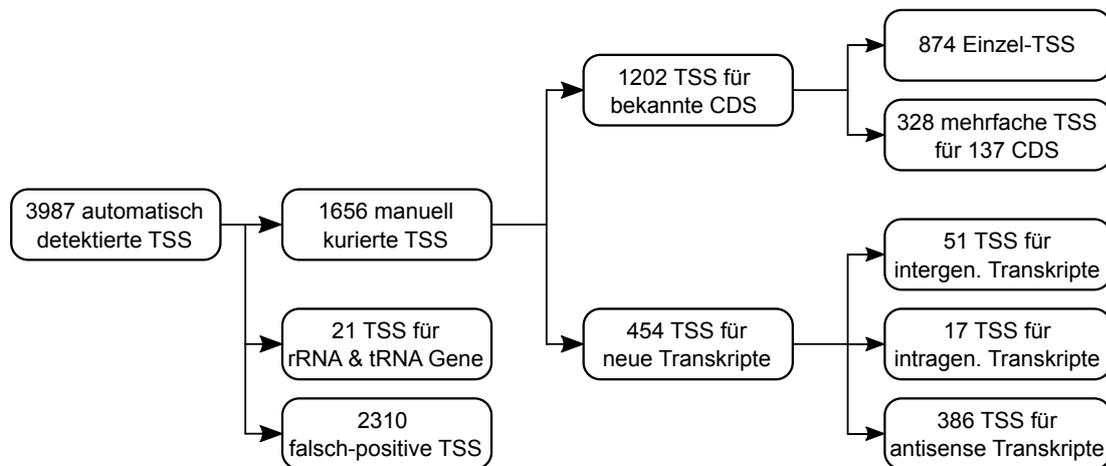


Abbildung 4.1: Klassifizierung der Transkriptionsstarts (TSS) in *C. diphtheriae* anhand von bereits bekannten (annotierten) Genen und neuen, bisher unbekannt, Transkripten.

Insgesamt wurden 1.202 TSS proteinkodierenden Genen zugeordnet. Darüberhinaus wurden 454 TSS neuen Transkripten zugeordnet. Von diesen TSS wurden 51 TSS als intergenisch, 17 TSS als intragenisch und 386 TSS als *antisense* TSS kategorisiert (Abbildung 4.1). Die

4 Ergebnisse

vollständige Liste der identifizierten TSS befindet sich im Anhang (Tabelle S2, Tabelle S3 & Tabelle S5).

Für die 4.5S RNA sowie die 6C RNA detektiert, obwohl beide RNAs in den Vollängen-cDNA-Bibliotheken eine deutlich erkennbare Read-Abdeckung aufwiesen. Dennoch wurde für den überwiegenden Teil der annotierten Gene ein TSS identifiziert.

4.1.3 Charakterisierung des σ^A Promotormotivs anhand der identifizierten Transkriptionsstarts

Die identifizierten TSS die bekannten Genen zugeordnet wurden, wurden für die Analyse von σ^A Promotorregionen eingesetzt. Für die Identifizierung wurde die DNA-Sequenz 40 bp *upstream* von der TSS des jeweiligen Gens eingesetzt. Dazu wurden die Sequenzen mit *Improbizer* (Ao *et al.*, 2004) auf ein konserviertes -10 Motiv untersucht. Da das -35 Motiv von σ^A Promotoren häufiger auftaucht, wenn das -10 Motiv nur geringe Ähnlichkeit zum Konsensus des -10 Motivs aufweist (Albersmeier *et al.*, 2017), wurden für die Identifizierung des -35 Motivs nur DNA-Sequenzen 23 bp *upstream* analysiert, die eine nicht optimale -10 Region aufwiesen.

Das -10 Motiv TAGaaT wurde *upstream* von 1.190 (98,9 %) TSS identifiziert (Abbildung 4.2). Das schwächer konservierte -35 Motiv ttgcaa wurde 16–20 bp *upstream* des (schwach konservierten) -10 Motivs bei 1.031 TSS gefunden. Für die Distanz zwischen dem -10 Motiv und der TSS wurde ein Wert von mindestens 6 bp und maximal 9 bp (Mittelwert 6,9 bp) gemessen. Die Position des TSS selbst bestand in 91 % der Fälle aus einem Purin (A oder G).



Abbildung 4.2: Promotormotiv des Sigmafaktors σ^A in *C. diphtheriae*. Die Größe der Nukleotide entspricht der relativen Abundanz des jeweiligen Nukleotids an dieser Position. Das -10 Motiv wurde *upstream* von 1.190 TSS und das -35 Motiv *upstream* von 1.031 TSS identifiziert. Die Daten wurden mit dem Programm *Weblogo* (Crooks, 2004) visualisiert.

4.1.4 Die Analyse der 5'-Untranslatierten Regionen ergibt eine für Corynebakterien charakteristische Längenverteilung

Anhand der Analyse der Regionen zwischen den TSS und dem Translationsstart (TLS) wurden die 5'-Untranslatierten Regionen (5'-UTRs) von mRNA in *C. diphtheriae* charakterisiert. Für die Analyse wurden die 1.202 TSS, die proteinkodierenden Genen zugeordnet wurden, sowie die 29 intergenischen TSS, die putativen proteinkodierenden Genen zugeordnet wurden, verwendet.

Die gemessene Länge von 5'-UTRs variierte zwischen 0 nt bis zu 463 nt (Abbildung 4.3A). Die 5'-UTR von 463 nt wurde für das neue Gen DIP1924A gemessen, welches erst im Rahmen dieser Arbeit identifiziert wurde. Das Gen kodiert für ein hypothetisches Protein. Transkripte dessen 5'-UTR eine Länge von 0 – 5 nt aufwies, sind zu kurz um eine ribosomale Bindestelle zu beinhalten und wurden daher als *leader*-freie (*leaderless*) Transkripte klassifiziert. Insgesamt wurden 452 (20 %) von 2.265 Genen als *leader*-freie Transkripte klassifiziert. Des Weiteren wurde die Anzahl an Startcodons für *leader*-freie und *leader*-tragende Transkripte untersucht. Für *leader*-freie als auch -tragende Transkripte wurde AUG als das häufigste und GUG als zweithäufigstes Startcodon identifiziert. Ungefähr 80 % der *leader*-freien und 62 % der *leader*-tragenden Transkripte besitzen AUG als Startcodon. Das Startcodon UUG wurde nur in circa 8 % der *leader*-tragenden Transkripte identifiziert. Wie in Abbildung 4.3A ersichtlich, beträgt die Länge der 5'-UTRs für eine kleine Fraktion an Transkripten 26 – 49 nt. Diese 5'-UTRs sind lang genug um eine ribosomale Bindestelle im nötigen Abstand zum Startcodon beinhalten, aber vermutlich zu kurz um *cis*-regulatorische Elemente zu umfassen.

Vorhersage von Riboswitches, RNA-Motiven und *leader* Sequenzen

Für 264 (21 %) Gene wurde eine Länge der 5'-UTR von >100 nt gemessen. Diese langen 5'-UTRs könnten *cis*-regulatorische Elemente besitzen, die die Transkription oder Translation der mRNA durch bestimmte Sequenzmotive oder Sekundärstrukturen beeinflussen. Um diese regulatorischen Elemente zu identifizieren wurde die Genomsequenz von *C. diphtheriae* mit dem Programm *Infernal* (Nawrocki und Eddy, 2013) auf bekannte Elemente in der Rfam-Datenbank (Nawrocki *et al.*, 2015) analysiert. Sieben regulatorische Elemente innerhalb von 5'-UTRs wurden für *C. diphtheriae* vorhergesagt, von denen 5 regulatorische Elemente im Primärtranskriptdatensatz als transkribiert identifiziert wurden (Tabelle 4.2). Neben einem Cobalamin- und zwei TPP-Riboswitches wurden die für Actinobakterien typischen *mraW* und *msiK* RNA-Motive mit den Primärtranskriptdaten als transkribiert bestätigt.

Desweiteren wurde das *C. glutamicum* Transkriptom nach in 5'-UTRs kodierten *leader*-Peptiden durchsucht. Diese Peptide befinden sich meist *upstream* von Operons die Gene

4 Ergebnisse

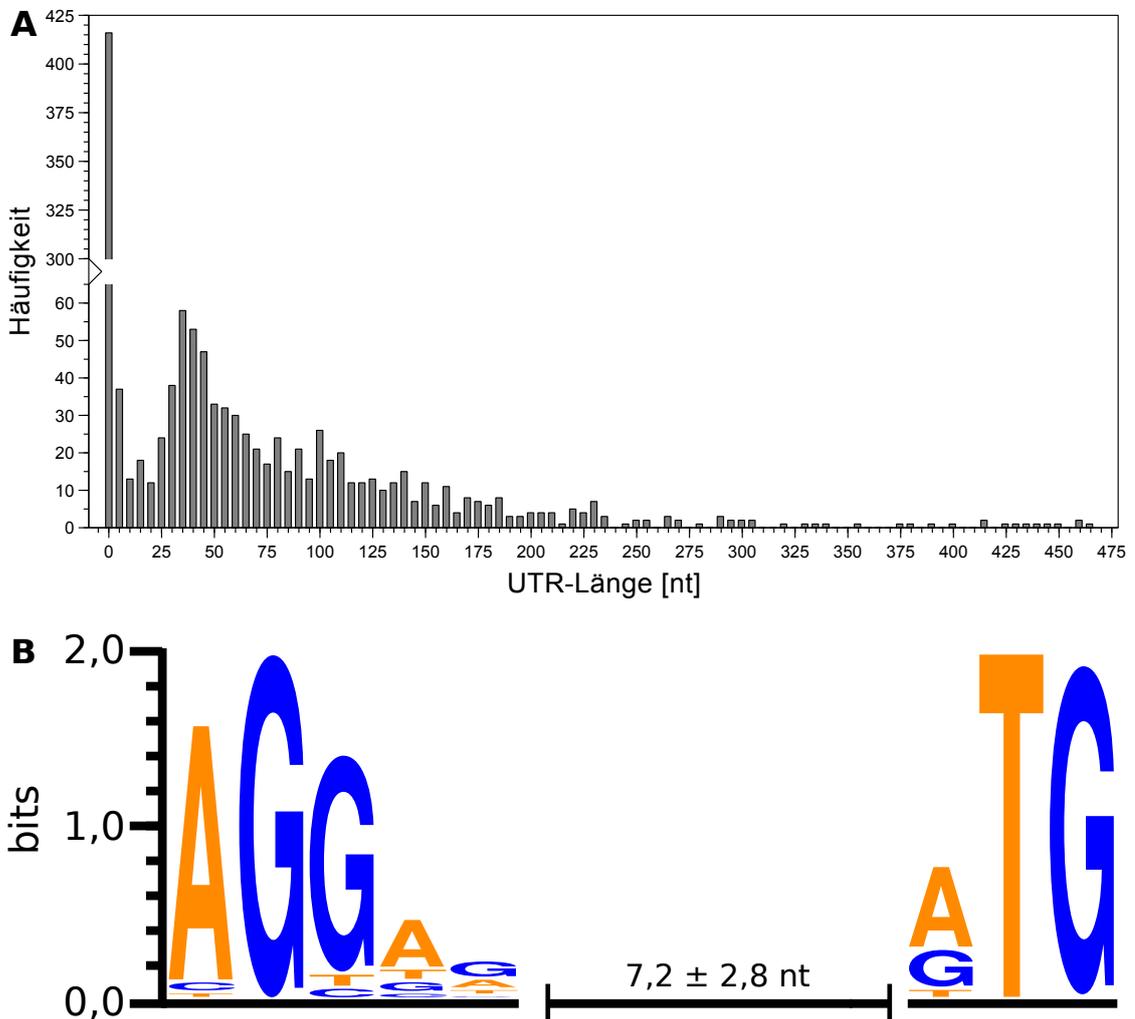


Abbildung 4.3: Histogramm der 5'-UTRs und Motiv der ribosomalen Bindestelle in *C. diphtheriae*. (A) Für die Analyse der 5'-UTRs wurden 1.232 TSS von bekannten und neuen für proteinkodierende Transkripten verwendet. Die Verteilung der 5'-UTRs rangiert zwischen 0 nt (*leader*-freie Transkripte) bis zu einem Maximum von 463 nt. Transkripte mit einer 5'-UTR-Länge kürzer als 5 nt machen etwa ein Drittel aller proteinkodierenden Transkripte aus. Die Balken repräsentieren Inkremente von fünf Nukleotiden (1 – 5 nt, 6 – 10 nt, usw.). Nur der erste Balken repräsentiert die UTR-Längen von 0 nt. (B) Darstellung des Motivs der ribosomalen Bindestelle (RBS) und dem mittleren Abstand zum Startcodon. Alle 5'-UTRs mit mehr als 5 nt Länge wurden für die Analyse des RBS-Motivs verwendet. Die Ordinaten-Achse gibt den Informationsgehalt der Sequenz an der jeweiligen Position in Bits an. Das Diagramm wurde mit *Weblogo* (Crooks, 2004) erstellt.

4.1 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

für die Aminosäurebiosynthese kodieren. Die Translation dieser Peptide führt zu einer veränderten Sekundärstruktur der Attenuator-RNA in der 5'-UTR des Gens in Abhängigkeit zur Verfügbarkeit bestimmter intrezellulärer Aminosäuren (Kolter und Yanofsky, 1982). In *C. diphtheriae* wurden drei putative *leader*-Peptidgene identifiziert: das *trp*-, das *ilvB*- und das *leu-leader*-Peptid. Das *trp-leader*-Peptid (MTNMNAHNWWWRA*) befindet sich *upstream* des ersten Gens des Tryptophan-Biosynthese-Operons im Genombereich 2.456.505 – 2.456.545 bp. Das *ilvB-leader*-Peptid (MNIIRLVVITTRRLP*) befindet sich *upstream* des Gens *ilvB* im Genombereich 1.081.747 – 1.081.794 bp. Das Genprodukt von *ilvB* ist in der Biosynthese von Isoleucin (I) und Valin (V) beteiligt. Das *leu-leader*-Peptid (MNRANLLLLRRGGSQA*) befindet sich im Genomabschnitt 230.506 – 230.455 bp *upstream* des Gens *leuA*, dessen Genprodukt den ersten Schritt in der Leucin-Biosynthese (L) katalysiert. Aufgrund von geringer Transkription in den jeweiligen Bereichen, konnte nur für das Gen *ilvB* ein TSS in den Primärtranskriptdaten identifiziert werden.

Tabelle 4.2: Vorhergesagte *cis*-regulatorische Elemente in 5'-UTRs von *C. diphtheriae*. Die Vorhersage wurde mit *Infernal* (Nawrocki und Eddy, 2013) und der Rfam-Datenbank (Nawrocki *et al.*, 2015) durchgeführt und anschließend mit den Primärtranskriptdaten auf tatsächlich Transkription untersucht. Die Tabelle ist nach *Bit Score* (Spalte „Score“) absteigend sortiert. Abkürzungen: RS, Riboswitch; n.a., nicht anwendbar.

Vorhersage mit Infernal und RFAM						Daten der RNA-Sequenzierung			
Name	RFAM ID	Start	Stop	Score	Strang	Status	Start	Stop	Gen
Cobalamin-RS	RF00174	1.066.117	1.066.317	90,8	+	beobachtet	1.066.117	1.066.317	DIP1084
Cobalamin-RS	RF00174	862.214	862.414	81,8	+	nicht beobachtet ^a	n.a.	n.a.	n.a.
<i>cspA</i> mRNA 5'-UTR	RF01766	285.506	285.923	64,8	+	beobachtet	285.506	285.923	DIP0320/ <i>cspA</i>
<i>mraW</i> RNA-Motiv	RF01746	1.640.095	1.639.987	70,2	-	beobachtet ^b	1.640.095	1.639.987	DIP1606/ <i>mraW</i>
TPP-RS	RF00059	924.778	924.890	60,8	+	beobachtet ^c	924.778	924.890	DIP0953
TPP-RS	RF00059	922.838	922.969	60,0	+	beobachtet ^c	922.838	922.969	DIP0951
<i>msiK</i> RNA-Motiv	RF01747	509.334	509.277	48,2	-	beobachtet	509.334	509.277	DIP0539
TPP-RS	RF00059	27.456	27.566	57,6	+	nicht beobachtet ^a	n.a.	n.a.	n.a.

^a Kein TSS detektiert und unzureichende Read-Abdeckung in diesem Bereich.

^b Das *mraW* RNA-Motiv befindet sich *upstream* vom TSS von DIP1606 / *mraW*.

^c Letzte Base des TPP-Riboswitch befindet sich innerhalb der CDS.

Analyse der ribosomalen Bindestelle

Anhand der mit den Primärtranskriptdaten erhaltenen 5'-UTR-Sequenzen wurde nach ribosomalen Bindestellen (RBS) innerhalb der 5'-UTRs gesucht. Dazu wurden 779 5'-UTRs mit einer Länge größer als 5 nt mit *Improbizer* (Ao *et al.*, 2004) auf konservierten Motiven analysiert. Als RBS wurde das Motiv AGGag in circa 87 % der analysierten 5'-UTRs identifiziert. Der mittlere Abstand zwischen TSS und Translationsstart (TLS) beträgt 7,2 nt ($\pm 2,8$ nt). Das für *C. diphtheriae* identifizierte RBS-Motiv ist in der Sequenz identisch und der Abstand zwischen TSS und TLS ähnlich zu der für *C. glutamicum* identifizierten RBS (Pfeifer-Sancar *et al.*, 2013). Dies war zu erwarten, da das RBS-bindende 3'-Ende der 16S rRNA in beiden Organismen identisch ist.

4.1.5 Verbesserung der Gen-Annotation durch Identifizierung von neuen Transkripten und Korrektur von Translationsstarts

Mit Hilfe der für mRNA identifizierten Transkriptionsstarts (TSS) ist es möglich annotierte proteinkodierende Sequenzen (CDS) im *C. diphtheriae* Genom zu verifizieren, zu korrigieren und zu re-annotieren. Des Weiteren können mit den TSS-Daten neue Transkripte detektiert und annotiert werden. Translationsstarts (TLS) von CDS wurden korrigiert, wenn sich der TSS *downstream* des TLS befindet: (a) Befindet sich der TSS auf einem potenziellen Startcodon (ATG oder GTG) welches *in-frame* mit der annotierten CDS liegt, wurde der TLS auf die Position des TSS verschoben. Da TSS und TLS auf der selben Position liegen, ist das neu annotierte Transkript *leader-frei* (*leaderless*). (b) Befindet sich der TSS nicht an einem potenziellen Startcodon, wurde der TLS *downstream* auf das nächste *in-frame* Startcodon verschoben. Dies führt zu einer verkürzten CDS mit einer 5'-UTR mit einer Länge > 0 nt.

Durch Anwendung dieser beiden Regeln wurden 120 TLS von CDS im *C. diphtheriae* Genom korrigiert. Von diesen 120 für CDS kodierenden mRNAs wurden 104 als *leader-freie* und 16 als *leader-tragende* (5'-UTR > 5 nt) Transkripte kategorisiert. Die korrigierten TLS wurden für die Analyse der 5'-UTR-Längenverteilung sowie bei der Motivanalyse der ribosomalen Bindestelle berücksichtigt. Eine Auflistung der korrigierten TLS befindet sich im Anhang (Tabelle S4).

Über die Analyse der intergenischen, intragenischen und *antisense* TSS ist es möglich neue Transkripte in *C. diphtheriae* zu identifizieren. Wie oben aufgeführt, wurden 454 TSS als zu möglichen neuen Transkripten zugehörig kategorisiert (Abbildung 4.1). Intergenische TSS weisen auf neue, bisher nicht annotierte proteinkodierende Gene oder nicht-kodierende *non-coding* RNAs (ncRNAs) hin. Insgesamt wurden 51 intergenische und 17 intragenische

4.1 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

TSS identifiziert. Da intragenische TSS zu verkürzten Proteinen führen, ist eine Aussage über die Funktion der intragenischen TSS schwierig.

Um neue proteinkodierende Transkripte zu identifizieren wurden die Sequenzen *downstream* von intergenischen TSS mit dem Programm *UGENE* (Okonechnikov *et al.*, 2012) nach offenen Leserahmen *open reading frames* (ORFs) durchsucht. Wurde ein potenzielle ORF gefunden, wurde dessen Aminosäuresequenz mit *BLASTp* (Altschul *et al.*, 1997, 2005) auf mögliche Proteinhomologe analysiert. Wurde kein ORF oder Proteinhomolog gefunden, wurde die Sequenz *downstream* des TSS auf mögliche ncRNAs und RNA-Motive mit *Infernal* (Nawrocki und Eddy, 2013) in der RFAM-Datenbank (Nawrocki *et al.*, 2015) durchsucht. Für 29 der 51 intergenischen TSS wurde ein potenzieller ORF gefunden und ein neuer Lokusname vergeben. Außerdem wurden die zwei ncRNAs tmRNA und RNase P M1 RNA als neue Transkripte identifiziert. Einige der neu identifizierten ORFs kodieren für Proteine mit metabolischer Funktion. So wurden eine Format-C-Acetyltransferase, ein Ammoniumtransporter, Magnesium-Chalatasen und eine Glycinderhydrogenase identifiziert. Außerdem wurde ein für einen putativen Transkriptionsregulator der HTH-Gruppe (*helix-turn-helix*) kodierendes Gen identifiziert. Dieses Gen befindet sich zwischen den Genen DIP1816 und DIP1817 und wurde daher mit dem Lokusnamen DIP1817A versehen (Abbildung 4.4). Für ungefähr 40 % der neu identifizierten ORFs wurde ein hypothetisches Protein oder Proteine mit unbekannter Funktion vorhergesagt.

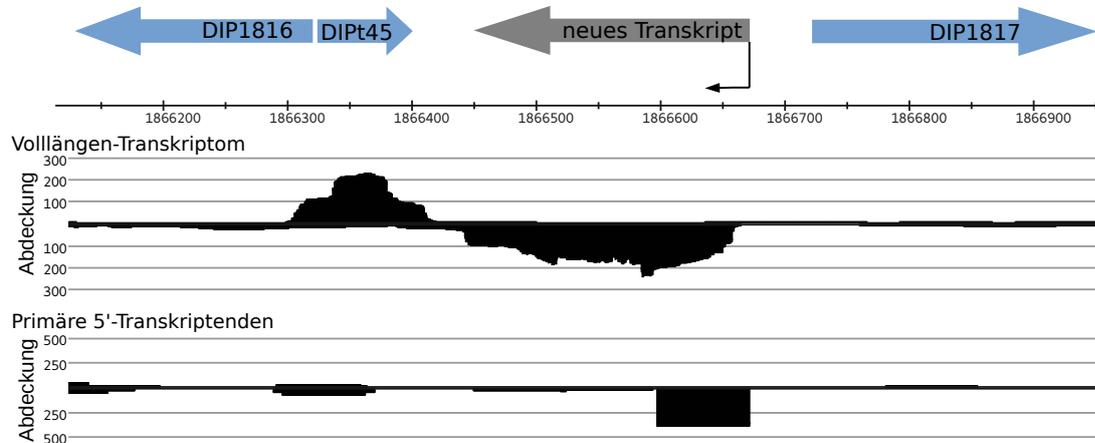


Abbildung 4.4: Darstellung eines intergenischen TSS mit neu identifiziertem Transkript. Für den intergenischen TSS an Position 1.866.670 bp im *C. diphtheriae* Genom wurde eine entsprechende Read-Abdeckung in den Daten der Vollängen-cDNA-Bibliothek des Wildtyps gefunden. Dem identifizierten ORF, welcher für einen putativen Transkriptionsregulator der HTH-Gruppe (*helix-turn-helix*) wurde der Lokusname DIP1817A gegeben.

4.1.6 Identifizierung von Operon-Strukturen durch die Kombination von Gesamttranskriptom- und TSS-Daten

Durch die Kombination der Daten aus der Primärtranskript- und den Vollängen-cDNA-Bibliotheken ist es möglich Operonstrukturen in *C. diphtheriae* zu identifizieren. Ein Operon ist ein polycistronisches Transkript welches aus mindestens zwei Genen besteht, die einen gemeinsamen Promotor besitzen. Für die Operon-Detektion wurde *ReadXplorer* (Hilker *et al.*, 2014) eingesetzt. Existieren mindestens fünf Reads die zwei benachbarte Gene auf dem selben Strang verbinden, so wurden diese Gene einem primären Operon zugeordnet. Weitere Gene wurden dem Operon hinzugefügt, wenn diese ebenfalls über mindestens fünf Reads mit den Genen, die sich bereits im Operon befinden, verbunden sind. Weiter wurden die identifizierten primären Operons als „experimentell validiert“ klassifiziert, wenn ein Transkriptionsstart (TSS) für das Operon identifiziert wurde. Wurde kein TSS für das jeweilige primäre Operon identifiziert, so wurde es als „validiert durch *paired-end* Reads“ klassifiziert. Wurde ein TSS innerhalb eines primären Operons detektiert, so entsteht ein verkürztes Transkript, welches weniger Gene als das primäre Operon trägt. Diese Operons wurden als sekundäre Operons bezeichnet. Gene die keinem Operon zugeordnet wurden, wurden der Gruppe der monocistronischen Gene zugeordnet.

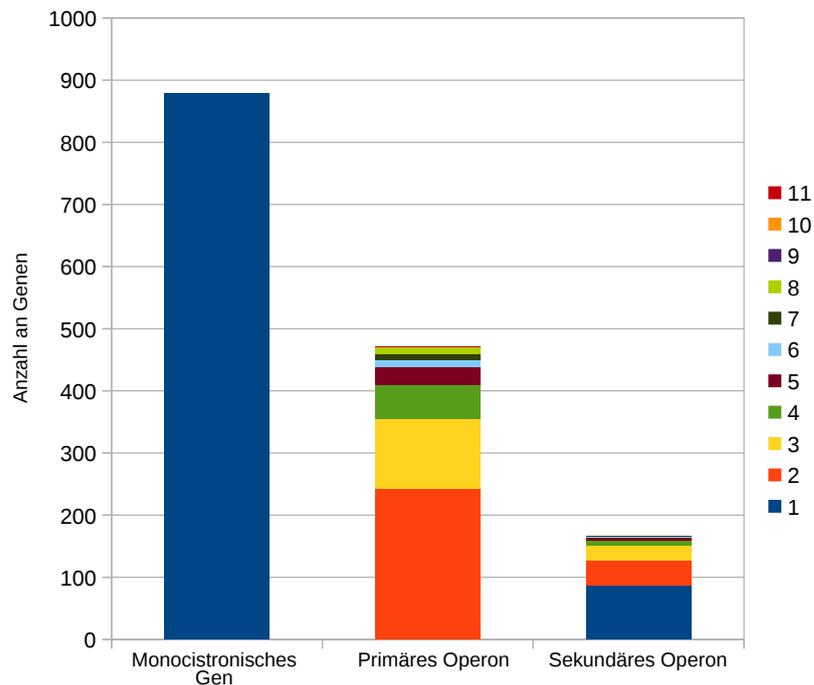


Abbildung 4.5: Anzahl an Monocistrons, primär und sekundär Operons in *C. diphtheriae*. Die Zahl an Genen innerhalb von primären bzw. sekundären Operons ist farb-kodiert.

4.1 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

Unter den hier gewählten Bedingungen wurden 471 primäre und 167 sekundäre Operons in *C. diphtheriae* identifiziert (Abbildung 4.5). Die 471 primären Operons umfassen 1.417 Gene. Von den primären Operons wurden 337 (72 %) Operons als „experimentell validiert“ kategorisiert, da für diese Operons ein TSS nachgewiesen wurde. Für 134 primäre Operons wurde kein TSS nachgewiesen. Diese wurden als „experimentell validiert durch *paired-end* Reads“ kategorisiert. Die beiden ncRNAs tmRNA und RNase P M1 RNA werden in Operons mit proteinkodierenden Genen transkribiert. Die tmRNA ist teil eines Operons mit *smfB* (DIP0750), welches für das SsrA-Bindeprotein kodiert. Die M1 RNA befindet sich in einem primären Operon bestehend aus DIP1679, DIP1678, M1 RNA und DIP1677. Dieses primäre Operon ist unterteilt in die zwei sekundären Operons bestehend aus DIP1678 bis DIP1677 und M1 RNA bis DIP1677.

Das längste identifizierte primäre Operon umfasst elf Gene die für verschiedene ribosomale Proteine der 30S und 50S Untereinheiten kodieren. Weiter wurden zehn Operons, die jeweils acht Gene umfassen, identifiziert. Diese Gene kodieren für Proteine mit unterschiedlichen Funktionen, wie der Replikation oder dem Kohlenstoffmetabolismus (Tabelle 4.3). Eine Liste der identifizierten primären und sekundären Operons befindet sich im Anhang (Tabelle S6).

Für das *C. diphtheriae* Genom wurden 878 monocistronische Gene (38 % der vorhergesagten Gene) identifiziert. Von diesen wurde für 550 Genen (63 %) ein TSS zugeordnet (Abbildung 4.5). Bei Betrachtung der Anzahl an Genen in primären Operons sowie der Zahl an Monocistrons mit einem TSS, so wurden fast 87 % aller annotierten Gene im Rahmen dieser Arbeit als aktiv transkribiert detektiert.

Tabelle 4.3: Längste primäre Operons in *C. diphtheriae*. Den Genen wurden ihre Funktionen anhand der *eggNOG* Datenbank (Version 4.5) (Huerta-Cepas *et al.*, 2016) zugeordnet. Die Annotation wurde aus der Datenbank übernommen.

Genbereich	Genzahl	Strang	TSS	Gennamen und Funktionsklassifizierung
DIP0472 – DIP0482	11	+	detektiert	<i>rpsJ, rplC, rplD, rplW, rplB, rpsS, rplV, rpsC, rplB, rpmC, rpsQ</i> . Translation, ribosomal structure and biogenesis (ribosomal proteins)
DIP0203 – DIP0209	8	+	nicht detektiert	DIP0203, DIP0204, DIP0205, DIP0206, DIP0207, DIP0208, DIP0208A, DIP0209. Function unknown, phage proteins
DIP0407 – DIP0414	8	+	detektiert	<i>hemE, hemG, hemL</i> , DIP0410, DIP0411, DIP0412, DIP0413, DIP0414. Coenzyme metabolism; Carbohydrate metabolism and transport; Post-translational modification, protein turnover, and chaperones

4 Ergebnisse

Tabelle 4.3: Fortgesetzt

Genbereich	Genzahl	Strang	TSS	Gennamen und Funktionsklassifizierung
DIP0438 – DIP0445	8	+	detektiert	DIP0438, DIP0439, DIP0440, DIP0441, DIP0442, DIP0443, DIP0444, DIP0445. Inorganic ion transport and metabolism; Function unknown
DIP0719 – DIP0726	8	+	detektiert	DIP0719, DIP0720, DIP0721, DIP0722, DIP0723, DIP0724, DIP0725, <i>uvrD</i> . Function unknown; Replication, recombination and repair; Inorganic ion transport and metabolism
DIP0969 – DIP0976	8	+	detektiert	DIP0969, DIP0970, DIP0971, DIP0972, <i>fdxA</i> , DIP0974, DIP0975, DIP0976. Signal transduction mechanisms; Amino Acid metabolism and transport; Function unknown; Energy production and conversion
DIP1248 – DIP1241	8	-	detektiert	DIP1248, DIP1247, DIP1246, DIP1245, DIP1244, DIP1243, <i>tatA</i> , <i>tatC</i> . Intracellular trafficking, secretion, and vesicular transport; Transcription; Amino Acid metabolism and transport; Post-translational modification, protein turnover, and chaperones; Function unknown
DIP1603 – DIP1596	8	-	nicht detektiert	<i>murE</i> , <i>murE</i> , <i>mraY</i> , <i>murD</i> , DIP1599, <i>murG</i> , <i>murC</i> , DIP1596. Cell wall, membrane, envelop biogenesis
DIP1720 – DIP1713	8	-	nicht detektiert	<i>dnaJ2</i> , DIP1719, DIP1718, DIP1717, DIP1716, DIP1715, <i>era</i> , <i>recO</i> . Replication, recombination and repair; Cell wall, membrane, envelop biogenesis; Coenzyme metabolism; Inorganic ion transport and metabolism; Function unknown; Signal transduction mechanisms; Post-translational modification, protein turnover, and chaperones
DIP1779 – DIP1772	8	-	nicht detektiert	<i>obgE</i> , DIP1778, <i>proB</i> , <i>proA</i> , <i>nadD</i> , DIP1774, DIP1773, DIP1772. Function unknown; Carbohydrate metabolism and transport; Coenzyme metabolism; Amino Acid metabolism and transport; Energy production and conversion
DIP1857 – DIP1850	8	-	detektiert	<i>clpS</i> , DIP1856, DIP1855, DIP1854, <i>murI</i> , DIP1852, <i>rph</i> , DIP1850. Nucleotide metabolism and transport; Translation, ribosomal structure and biogenesis; Function unknown; Cell wall, membrane, envelop biogenesis; Amino Acid metabolism and transport; Secondary metabolites biosynthesis, transport, and catabolism; Post-translational modification, protein turnover, and chaperones

4.1.7 Das Diphtherietoxin kodierende Gen *tox* besitzt zwei bisher unbekannte *antisense* Transkripte

C. diphtheriae NCTC 13129 ist ein *tox*⁺ Stamm, da es einen chromosomal integrierten Corynephagen mit dem Diphtherietoxin (DT) kodierenden Gen *tox* besitzt. Das DT ist eins der stärksten bakteriellen Toxine und essenziell für die Pathogenität von *C. diphtheriae* (Freeman, 1951; Baron, 1996). Im Rahmen dieser Arbeit wurden die Daten der Primärtranskript- als auch der Vollängen-cDNA-Bibliotheken genutzt, um die Phagen-Insel (*phage island*) im Allgemeinen und das *tox* Gens im Speziellen hinsichtlich ihrer transkriptionellen Strukturen zu beleuchten.

Die Phagen-Insel befindet sich im Genom zwischen den tRNA_{Arg} Genen DIPt10 und DIPt11 und enthält 44 Gene (von DIP0180 bis DIP0222 / *tox*). Nur sechs TSS wurden Genen des Corynephagen zugeordnet. Während das Gen DIP0180 *leader*-frei ist, besitzen die anderen fünf Gene 5'-UTRs mit Längen zwischen 28 nt bis 413 nt. Darüber hinaus wurden sechs TSS, die *antisense* Transkripten zugeordnet wurden, identifiziert. In der Phagen-Insel wurden sieben primäre (34 Gene) sowie ein sekundäres (ein Gen) Operon identifiziert. Die restlichen zehn Gene wurden der Klasse der Monocistrons zugeordnet (Abbildung 4.6A).

Für die Gene im mittleren Bereich der Phagen-Insel wurde eine, im Vergleich zu den Genen in den äußeren Bereichen, geringere Transkriptabundanz für den Wildtyp als auch für die Δ *dtxR*-Mutante gemessen. Bezogen auf den Wildtyp wurde für 20 Gene (45 %) des Corynephagen eine differenzielle Transkription in der Δ *dtxR*-Mutante gemessen. Diese 20 differenziell transkribierten Gene sind über den Corynephagen verteilt und sind nicht auf einen bestimmten Bereich des Phagen beschränkt. Obwohl nur für das *tox* Gen im Corynephagen eine Bindestelle des Diphtherietoxin-Repressors (DtxR) bekannt ist, sind in der Δ *dtxR*-Mutante 19 weitere Gene ohne diese Bindestelle stärker transkribiert als im Wildtyp. Unter diesen Genen sind Gene die für die Repression des Phagen, Capsid-Zusammensetzung involviert sind sowie für eine putative Phagen-Integrase und einen putativen Transkriptionsregulator kodieren.

Für das *tox* Gen wurden zwei TSS (genannt TSS 1 und TSS 2) *upstream* der CDS mit einer 5'-UTR von 43 nt bzw. 38 nt identifiziert (Abbildung 4.6B). Die DtxR-Bindestelle befindet sich 32 nt *upstream* des Startcodons und überlappt somit das -10 Motiv von beiden TSS. Für beide TSS wurde ein σ^A Promotormotiv detektiert. Wie erwartet, wurde für *tox* in der Δ *dtxR*-Mutante eine >4-fache Verstärkung der Transkription im Vergleich zum Wildtyp gemessen.

In frühen Studien zur Transkription von *tox* wurden zwei überlappende Promotoren gefunden, dessen TSS 38 bzw. 43 nt vom Startcodon entfernt liegen (Kaczorek *et al.*, 1985; Boyd und Murphy, 1988). Anhand von Mutagenese-Experimenten konnte gezeigt werden, dass das

4 Ergebnisse

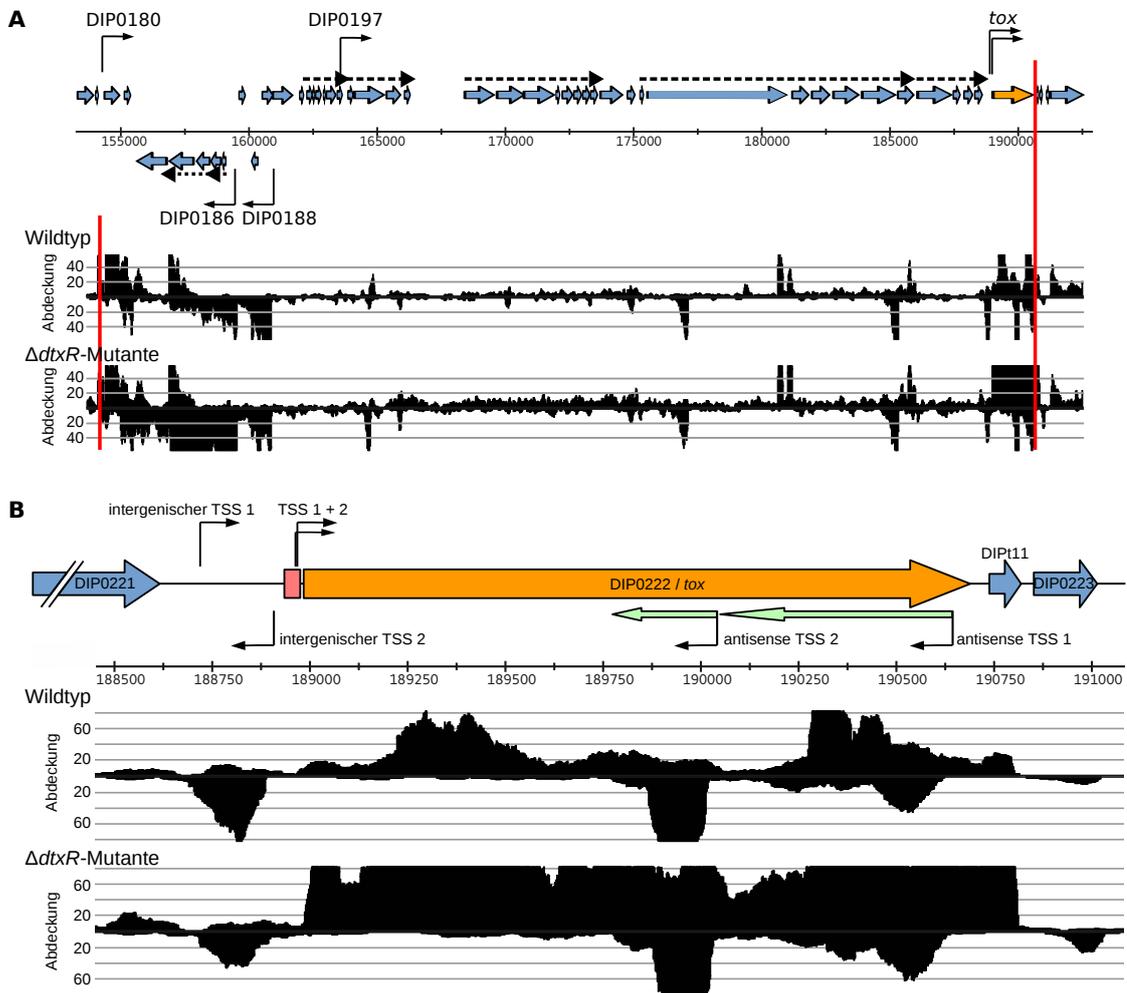


Abbildung 4.6: Transkriptionsprofil der Phagen-Insel (*phage island*) und dem *tox* Gen in *C. diphtheriae*. (A) Dargestellt ist das Transkriptionsprofil der Phagen-Insel (umrahmt von zwei roten Linien) für jeweils ein Replikat des Wildtyps und der $\Delta dtxR$ -Mutante. Identifizierte TSS sind mit einem schwarzen Pfeil und primäre Operons mit einem gestrichelten Pfeil angezeigt. Das *tox* Gen ist mit einem gelben Pfeil dargestellt. Intergenische und intragenische TSS wurden aus Gründen der Übersichtlichkeit weggelassen. (B) Detailansicht des monocistronischen *tox* Gens. Das Gen besitzt zwei TSS (TSS 1 bzw. 2) mit 5'UTR-Längen von 43 und 38 nt. Die DtxR-Bindestelle ist als roter Kasten eingezeichnet und überlappt die -10 Motive der beiden TSS. Zwei TSS für *antisense* RNAs (grüne Pfeile) sowie zwei intergenische TSS sind ebenfalls eingezeichnet. Bei der Abbildung handelt es sich um ein modifiziertes Bildschirmfoto von *ReadXplorer* (Hilker *et al.*, 2014) und zeigt daher nur die nicht normalisierte Read-Abdeckung.

4.1 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

näher zum Startcodon gelegene -10 Motiv das stärkere von beiden ist (Boyd und Murphy, 1988). Dieses Ergebnis konnte in den hier gezeigten Primärtranskriptdaten bestätigt werden, da die Zahl an Reads am TSS, der sich näher zum Startcodon befindet (TSS 2), ~5-fach größer ist, als am weiter entfernten TSS (Daten nicht gezeigt).

Neben den TSS des *tox* Gens wurden in den Primär- und Vollängentranskript-Daten zwei putative *antisense* RNAs (asRNAs) inklusive TSS sowie zwei intergenische TSS *upstream* von *tox* identifiziert (Abbildung 4.6B). Diese *antisense* und intergenischen Transkripte wurden bisher in keiner, dem Autor dieser Arbeit bekannten, Publikation beschrieben. Die beiden asRNAs beginnen in der Nähe des 3'-Endes der CDS asRNA TSS 1) bzw. in der Mitte der CDS (asRNA TSS 2) von *tox*. Die Längen der asRNAs wurden anhand der Read-Abdeckung in der Vollängen-cDNA-Bibliothek bestimmt und betragen ca. 580 nt für asRNA 1 und 260 nt für asRNA 2. *Antisense* RNAs können sowohl stabilisierende als auch destabilisierende Effekte auf das *sense*-Transkript haben (Thomason und Storz, 2010). Die Funktion der hier gezeigten asRNAs ist allerdings unbekannt.

Die beiden intergenischen TSS (intergenischer TSS 1 bzw. 2) deuten auf das Vorhandensein von neuen Transkripten *upstream* des *tox* Gens hin (Abbildung 4.6B). Die beiden Transkripte befinden sich auf gegenüberliegenden Strängen. Die Read-Abdeckung des Transkripts auf dem Gegenstrang ist deutlich erkennbar, während die Read-Abdeckung des Transkripts auf dem Vorwärtsstrang nur gering ist. Sowohl eine Suche nach möglichen kodierten Proteinen mittels *BLAST* (Altschul *et al.*, 1997), als auch nach ncRNAs in der RFAM-Datenbank (Nawrocki *et al.*, 2015) ergaben keine Hinweise auf mögliche Funktionen.

4.1.8 Vergleichende Analyse der DtxR-regulierten Gene

Der Diphtherietoxin-Repressor (DtxR) ist der Transkriptionsregulator der Eisen-Homöostase und des, für das Diphtherietoxin kodierende, Gen *tox* in *C. diphtheriae* und daher essenziell für die Pathogenizität des Bakteriums (Baron, 1996). Der Eisen-bindende Transkriptionsregulator DtxR erkennt und bindet das DtxR-Motiv auf der DNA bei Eisenüberschuss und reguliert damit die Expression von Genen die in den Eisenstoffwechsel involviert sind (Trost *et al.*, 2012; Merchant und Spatafora, 2014). Für *C. glutamicum* wurde experimentell gezeigt, dass DtxR ein dualer Regulator ist, der bei Eisenüberschuss Gene für die Eisenaufnahme reprimiert, aber Gene für die Eisenspeicherung aktiviert (Brune *et al.*, 2006). Um die Eigenschaften des dualen Regulators DtxR in *C. diphtheriae* zu analysieren, wurde die genomweite Transkription einer Δ *dtxR*-Mutante mit der des Wildtyps verglichen. Die Probengenerierung und Erstellung der Vollängen-cDNA-Bibliotheken ist im ersten Abschnitt 4.1.1 auf Seite 74 beschrieben. Die gemappten *paired-end* Reads wurden mit *ReadXplorer* (Hilker *et al.*, 2014) gezählt und mit

4.1 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

Gene in *C. diphtheriae* durch direkte oder indirekte Effekte differenziell transkribiert wurden. Das Gen DIP2330, welches für ein putatives sekretiertes Protein kodiert, war in der $\Delta dtxR$ -Mutante am stärksten differenziell transkribiert (LFC 6,28). Von den 40 Genen mit bekannter oder vorhergesagter DtxR-Bindestelle waren 25 (63 %) differenziell transkribiert. Abhängig von dem Status der differenziellen Transkription wurden die Gene mit DtxR-Bindestelle durch DtxR entweder aktiviert oder reprimiert (Tabelle 4.4). Der Großteil der 25 differenziell transkribierten Gene zeigte eine erhöhte Transkriptabundanz in der $\Delta dtxR$ -Mutante und wird daher im Wildtyp bei Eisenüberschuss von DtxR reprimiert. Diese Klasse von Genen kodiert für Hemin-Rezeptoren, Eisentransportern und Siderophoren. Die Gene *ftn*, *sdhB*, *narK* und *ycdA* sind in der $\Delta dtxR$ -Mutante schwächer transkribiert. Unter Eisenüberschuss werden diese Gene im Wildtyp vermutlich von DtxR aktiviert. Das Gen *ftn*, welches für ein Eisenspeicherprotein kodiert, war in der $\Delta dtxR$ -Mutante am schwächsten differenziell transkribiert und unterstreicht die mögliche duale Funktion von DtxR in *C. diphtheriae*.

Die Positionen der DtxR-Bindestellen und der TSS stimmen mit den regulatorischen Modellen überein. Die DtxR-Bindestellen von allen durch DtxR aktivierten Genen befinden sich mindestens 37 bp *upstream* von der jeweiligen TSS. Im Gegensatz dazu überlappen alle DtxR-Bindestellen von durch DtxR reprimierten Genen mit dem -10 Motiv des entsprechenden σ^A Promotors oder befinden sich *downstream* der jeweiligen TSS. Der Mechanismus der Repression durch DtxR funktioniert wahrscheinlich über das simple Verdecken des Promotors, so dass die RNA-Polymerase nicht an diesen binden kann. Alternativ bindet DtxR *downstream* des TSS und blockiert die RNA-Polymerase, so dass diese vorzeitig anhält (Rodionov, 2007). Die Bindung von DtxR *upstream* des TSS scheint einen aktivierenden Effekt auf die Transkription zu haben, wobei unklar ist, wie dieser Mechanismus der Aktivierung funktioniert. Dennoch wurde für einige Gene mit bekannter bzw. vorhergesagter DtxR-Bindestelle kein TSS identifiziert (Tabelle 4.4). Dies ist vermutlich auf geringe Transkription in diesen Bereichen zurückzuführen.

4 Ergebnisse

Tabelle 4.4: Gene mit DtxR-Bindestelle und Status der differentiellen Transkription sowie Promotormotiv in der *C. diphtheriae* Δ dtxR-Mutante. Basierend auf dem Status der differentiellen Transkription wurden die Gene in drei Kategorien eingeteilt: reprimiert, aktiviert oder nicht differentiell transkribiert. Gene für die ein TSS identifiziert, aber kein -10 Motiv detektiert wurde, sind anhand eines Startcodons mit gepunkteten Linie dargestellt. Mit einem Stern markierte Gene besitzen eine experimentell bestätigte DtxR Regulation. Expression, normalisierte Transkriptabundanz in Reads; LFC, \log_2 FoldChange; Start, Startcodon.

Status	Gen	Expression	LFC	angepasster p-Wert	-35	σ^A -Promoterregion	-10	Start
Reprimiert	DIP2330	15933,8	6,28	2,7E-159	CCAGCT ... N ₁₈	...TAGGGT...	N ₃₈₆	...TTG
	DIP0523 / <i>cirA</i> ^{a*}	1603,1	5,21	9,0E-101GTG
	DIP1520 / <i>chtA</i> [*]	9247,5	4,74	3,0E-053	TTAGGT ... N ₁₉	...TATGAT...	N ₅₀	...TTG
	DIP1519 / <i>chtB</i> [*]	3776,1	4,53	4,7E-026	CTGGTT ... N ₁₈	...TAAGGT...	N ₁₃	...GTG
	DIP0625 / <i>htaA</i> [*]	5566,6	4,28	9,1E-023	CTGGAC ... N ₁₈	...TAGCCT...	N ₁₅₄	...GTG
	DIP0629 / <i>htaB</i> [*]	2522,4	3,81	2,9E-042	TTGAAA ... N ₁₇	...TAAGGT...	N ₄₁	...ATG
	DIP0586 / <i>ciuE</i> ^{b*}	1184,3	3,79	3,6E-031	GTG
	DIP2161 / <i>sidA</i> ^{a*}	308,4	3,08	4,0E-035	ATG
	DIP1669 / <i>hmuO</i> [*]	478,0	2,99	2,4E-039	ATGTGA ... N ₁₇	...TAICTT...	N ₅₄	...ATG
	DIP1510 / <i>xerC2</i>	832,7	2,81	2,6E-040	TTGCGT ... N ₁₇	...TAATGT...	N ₂₁₇	...GTG
	DIP0624 / <i>htaC</i> [*]	334,6	2,62	7,7E-004TAGGCT...	N ₅₃	...ATG
	DIP0922 / <i>ripA</i> [*]	1450,5	2,58	5,1E-004	TTGCCT ... N ₁₇	...TAACCT...	N ₈₉	...GTG
	DIP0108 / <i>irp6A</i> [*]	16792,9	2,56	6,6E-009	TTCTTT ... N ₁₇	...TATAGT...	N ₄₂	...GTG
	DIP2356 / <i>irp4</i> [*]	208,5	2,18	5,6E-008	ATACGT ... N ₁₈	...TAACCT...	N ₅₁	...ATG
	DIP0222 / <i>tox</i> ^{c*}	1435,8	2,07	5,3E-018TAGGAT...	N ₃₈	...GTG
	DIP1062 / <i>frgD</i> ^{c*}	11524,2	2,02	9,2E-008	CTAACC ... N ₁₉	...TAACCT...	N ₇	...ATG
	DIP0582 / <i>ciuA</i> [*]	4077,3	1,82	3,8E-008	TTCCCC ... N ₁₈	...TACCTT...	N ₈	...ATG
	DIP0415 / <i>glyR</i> ^b	230,8	1,78	5,8E-012	ATG
	DIP2162 / <i>ddpA</i> ^{a*}	93,8	1,65	2,2E-006	ATG
	DIP2114 / <i>adhA</i> ^{a*}	1561,2	1,52	1,6E-005	ATG
DIP1061 / <i>frgA</i> [*]	1257,1	1,25	5,3E-004	TTACCC ... N ₁₉	...TAGAAT...	N ₆	...ATG	
Nicht differentiell	DIP1189 / <i>bioD1</i> ^a	105,7	0,89	2,3E-002	GTG
	DIP0894 / <i>irp5</i> ^{d*}	290,3	0,89	3,3E-002	TTCACC ... N ₁₈	...TACACT...	N ₈	...ATG
	DIP0169 / <i>iutA</i>	939,4	0,85	1,5E-002	CTAGCT ... N ₁₉	...TAAGAT...	N ₄₂	...GTG
	DIP0124 / <i>piuB</i> ^{a*}	58,4	0,83	2,4E-002	ATG
	DIP1296 / <i>sufR</i> [*]	772,4	0,66	1,1E-001	TTGGCG ... N ₁₈	...TAGTGT...	N ₁₀₄	...TTG
	DIP2303 / <i>dps</i>	211,2	0,65	7,4E-003	TTGCAC ... N ₁₆	...CACACT...	N ₇	...ATG
	DIP0579	269,8	0,56	7,0E-002	TTGCCT ... N ₁₈	...TAGGGT...	N ₇	...GTG
	DIP2219 ^a	906,5	0,52	3,4E-002	TTG
	DIP0540 / <i>secY</i> [*]	4874,4	0,43	4,9E-002	CTAGTG ... N ₁₉	...TAGGGT...	N ₁₁₉	...GTG
	DIP0699 / <i>secA1</i>	8693,0	0,10	6,5E-001	TTGCTG ... N ₂₀	...TAGGCT...	N ₁₁₂	...GTG
	DIP0437 / <i>rplL</i> ^b	24043,2	-0,12	6,3E-001	ATG
	DIP1714 / <i>era</i> ^d	2424,0	-0,15	4,9E-001TAACAT...	N ₄₈	...GTG
	DIP1626 / <i>qcrC</i> ^a	6020,6	-0,39	6,7E-002	ATG
DIP1473 / <i>yppH</i> ^d	557,8	-0,61	5,6E-002	TCCGCT ... N ₁₈	...TATCCT...	N ₄₈	...ATG	
DIP2202 / <i>beta</i> A	788,3	-0,81	2,9E-005	GTGGCA ... N ₁₈	...CACACT...	N ₁₂₅	...GTG	
Aktiviert	DIP1391 / <i>ycdA</i> ^c	2322,6	-1,01	1,2E-007	TTGCTG ... N ₁₉	...TATATT...	N ₁₁₄	...TTG
	DIP0501 / <i>narK</i>	7249,6	-2,17	1,0E-006	TTAATT ... N ₁₆	...TAGCTT...	N ₉₄	...ATG
	DIP0370 / <i>sdhB</i>	5816,8	-3,51	1,1E-031	GTGGTC ... N ₁₉	...TAAAAT...	N ₂₁	...ATG
	DIP1866 / <i>ftn</i>	1551,5	-5,97	1,9E-045	CTGACA ... N ₁₇	...TAGCAT...	N ₇	...ATG

^a Kein TSS identifiziert, daher kein Promotormotiv detektierbar.

^b Das Gen ist Teil eines primären Operons und besitzt keinen eigenen TSS.

^c Von mehreren identifizierten TSS wird nur der TSS dargestellt, der sich am nächsten an der DtxR-Bindestelle befindet. Für *tox* wurde das -10 Motiv von dem TSS dargestellt, der sich am nächsten zum Startcodon befindet (TSS 2).

^d Das Gen ist das erste in einem sekundären Operon und besitzt einen eigenen TSS.

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

Der vorherige Abschnitt behandelte Analyse der Transkriptionsinitiation über die transkriptomweite Charakterisierung von Transkriptionsstartpunkten in *C. diphtheriae* sowie der differentiellen Transkription in der *C. diphtheriae* Δ *dtxR*-Mutante. Der folgende Abschnitt richtet den Blick auf die 3'-Enden der RNA und die Analyse der Transkriptionstermination durch den Rho-Faktor im Modellorganismus *Corynebacterium glutamicum*.

4.2.1 Vergleich des Rho-Proteins aus *C. glutamicum* mit den Homologen der Modellorganismen

Im Folgenden wird das Rho-Protein aus *C. glutamicum* mit den Homologen aus den Modellorganismen *Escherichia coli* und *Bacillus subtilis*, sowie dem nahen Verwandten *Mycobacterium tuberculosis* verglichen. Während die Rho-Proteine in den klassischen Modellorganismen *E. coli* (Ec.Rho) und *B. subtilis* (Bs.Rho) eine Länge von 419 bzw. 427 AS (Aminosäuren) aufweisen, ist das Rho-Protein aus *C. glutamicum* (Cg.Rho) mit 762 AS deutlich länger (Abbildung 4.8A). Das Rho-Protein aus *M. tuberculosis* (Mt.Rho) ist mit einer Länge von 602 AS ebenfalls länger als die Homologe aus *E. coli* und *B. subtilis*, aber insgesamt 20 % kürzer als das Homolog in *C. glutamicum*. Wie von D'Heygère *et al.* (2015) festgestellt, besitzt Mt.Rho im Vergleich zu Ec.Rho eine Insertion von 144 AS in der N-terminalen Domäne (NTD) (Abbildung 4.8A). Anhand von Vergleichen der Aminosäuresequenzen der Rho-Homologe untereinander konnte für Cg.Rho ebenfalls eine Insertion in der NTD mit einer Länge von ca. 250 AS im Vergleich zu Ec.Rho beobachtet werden. Die Funktion der NTD-Insertion von Cg.Rho ist unbekannt.

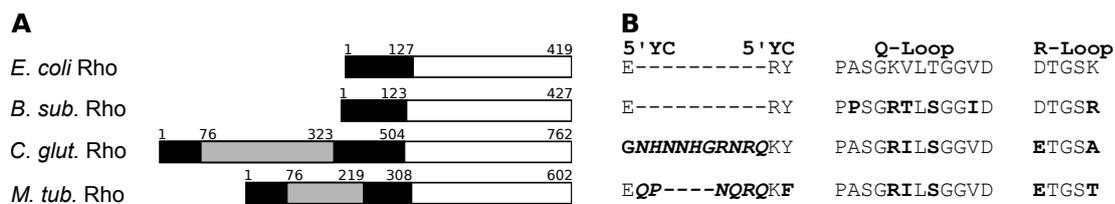


Abbildung 4.8: Vergleich der Rho-Homologe der Modellorganismen *E. coli* und *B. subtilis* mit *C. glutamicum* und *M. tuberculosis*, sowie einigen der charakteristischen Motive. (A) Längenvergleich von Rho mit den Hauptdomänen. Die N-terminale Domäne ist in schwarz, die C-terminale Domäne in weiß dargestellt. *C. glutamicum* und *M. tuberculosis* besitzen eine Insertion in der N-terminalen Domäne, dargestellt in grau. (B) Vergleich einiger für Rho charakteristischer Motive wie dem 5'-YC-Dimer-Motiv sowie dem Q- und dem R-Loop. Aminosäurereste sind in fett gedruckt, wenn sie sich von den Aminosäureresten in *E. coli* Rho unterscheiden. Insertionen sind kursiv und fett gedruckt.

4 Ergebnisse

Des Weiteren besitzt Cg.Rho eine weitere Insertion von 10 AS zwischen den beiden 5'-YC-Dimer-Motiven (Abbildung 4.8B). Mt.Rho besitzt ebenfalls eine Insertion mit einer Länge von 6 AS zwischen diesen beiden Motiven (D'Heygère *et al.*, 2015). In Ec.Rho ermöglichen die beiden 5'-YC-Dimer-Motiv die Erkennung eines 5'-YC-Dimers (Y = C oder T) innerhalb von *rut sites* auf der RNA, was die hohe Präferenz für C-reiche *rut sites* von Ec.Rho erklärt (Bogden *et al.*, 1999).

In der C-terminalen Domäne (Abbildung 4.8A) befinden sich unter anderem zwei weitere wichtige Motive für die RNA-Bindung. Diese Motive werden als Q- und R-Loop bezeichnet (Abbildung 4.8B). Das Q-Loop-Motiv von Cg.Rho besitzt im Vergleich eine höhere Übereinstimmung zu Bs.Rho als zu Ec.Rho. Das R-Loop-Motiv von Cg.Rho besitzt ebenfalls keine vollständige Identität zu Ec.Rho oder Bs.Rho, ist aber bis auf eine Aminosäure mit dem von Mt.Rho identisch.

Aufgrund der großen Insertion in der NTD ist die globale Sequenzidentität und -ähnlichkeit von Cg.Rho zu den Homologen der klassischen Modellorganismen *E. coli* und *B. subtilis* wesentlich geringer, als zu dem Homolog in *M. tuberculosis* (Tabelle 4.5).

Tabelle 4.5: Globales Alignment der Rho-Homologe der Modellorganismen *E. coli* und *B. subtilis* mit denen von *C. glutamicum* und *M. tuberculosis*. Das Alignment wurde mittels Needleman-Wunsch-Algorithmus (Needleman und Wunsch, 1970) durchgeführt. Angegeben ist die Sequenzidentität und -ähnlichkeit.

Homolog	Ec.Rho	Bs.Rho	Cg.Rho	Mt.Rho
Ec.Rho	100 / 100	55 / 73	27 / 35	34 / 46
Bs.Rho	–	100 / 100	24 / 34	31 / 45
Cg.Rho	–	–	100 / 100	47 / 55
Mt.Rho	–	–	–	100 / 100

4.2.2 Die Deletion des Rho-Faktor Gens führt zu verlangsamten Wachstum

Um den Einfluss des Rho-Proteins in *C. glutamicum* zu analysieren, wurde das 2289 bp lange *rho*-Gen (*cg1354*) mittels des pK18*mobsacB*-Systems (Schäfer *et al.*, 1994) deletiert. Da *rho* das erste Gen im Operon *rho-prfA-cg1356* ist, wurden die 5'-UTR sowie die hinteren 550 Basen der *rho*-CDS beibehalten, um die Transkription der beiden *downstream* gelegenen Gene möglichst unbeeinflusst zu lassen.

Als erste Charakterisierung der $\Delta\rho$ -Mutante wurde das Wachstum im Vergleich zum Wildtyp in CGXII-Minimalmedium untersucht. Wie in Abbildung 4.9 zu erkennen, ist die Wachstumsgeschwindigkeit der Deletionsmutante im Vergleich zum Wildtyp verringert. Die maximale spezifische Wachstumsgeschwindigkeit μ_{\max} der $\Delta\rho$ -Mutante beträgt $0,127 \text{ h}^{-1}$,

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

während der Wildtyp mit einer μ_{\max} von $0,194 \text{ h}^{-1}$ in der exponentiellen Phase wächst. Die Verringerung der Wachstumsgeschwindigkeit deutet auf eine wichtige Funktion von Rho in der Zelle hin.

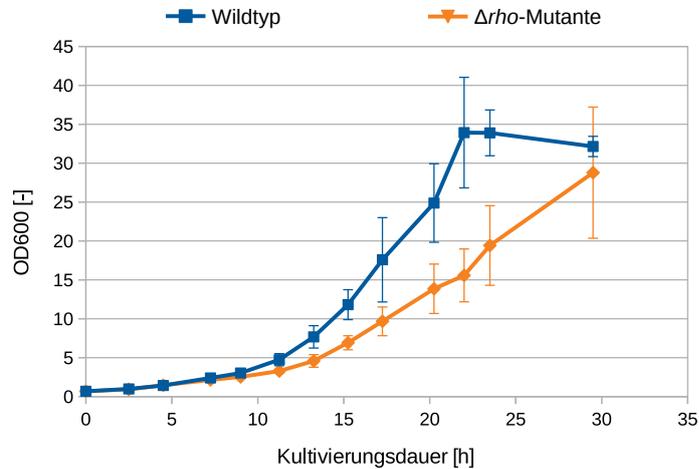


Abbildung 4.9: Wachstumskurve des *C. glutamicum* Wildtyps sowie der $\Delta\rho$ -Mutante in Minimalmedium mit Glucose. Dargestellt sind die Mittelwerte von drei biologischen Replikaten. Die Fehlerbalken stellen die Standardabweichung dar.

Nach der Deletion des *rho*-Gens wurde eine Komplementation der Deletionsmutante durchgeführt. Dazu wurde die CDS des *rho*-Gens in den Vektor pEC-XC99E kloniert und in die $\Delta\rho$ -Mutante transformiert. Dieser Vektor besitzt neben einem Gen für eine Chloramphenicol-Resistenz, einen *lac*-Operator *downstream* des starken P_{trc} -Promotors, sodass die Expression des klonierten *rho*-Gens mit IPTG (Isopropyl- β -D-Thiogalactopyranosid) induzierbar ist.

Das Wachstumsdefizit der $\Delta\rho$ -Mutante gegenüber dem Wildtyp konnte nicht über die plasmidbasierte Expression von *rho* auf dem Plasmid pEC-XC99E komplementiert werden. Durch das plasmidkodierte *rho*-Gen wurde das Wachstumsdefizit verstärkt (Abbildung A.4A im Anhang). Wahrscheinlich ist die Expression von *rho* in einem *high copy* Plasmid zu stark.

Um das Problem der plasmidbasierten Komplementation zu umgehen, wurde eine chromosomale Integration des *rho*-Gens inklusive nativem Promotor und 5'-UTR mit dem pRIM2-System (Vašicová *et al.*, 1998) durchgeführt. Das Plasmid pRIM2 integriert *downstream* von Gen *ppc*, welches für eine Phosphoenolpyruvate-Carboxylase kodiert. Die Integration des Leervektors pRIM2 in das Chromosom hat keinen Einfluss auf das Wachstum (Vašicová *et al.*, 1998). Durch die chromosomale Integration von pRIM2::*rho* konnte das Wachstumsdefizit der $\Delta\rho$ -Mutante vollständig komplementiert werden (Abbildung A.4B im Anhang).

4.2.3 Die Deletion von *rho* führt zu einer deutlichen Veränderung des Transkriptoms

Das verlangsamte Wachstum der Δrho -Mutante deutet auf die Wichtigkeit des Rho-Faktors für die Zelle hin. Als Transkriptionsterminator ist der Rho-Faktor direkt in die Transkription involviert. Aus diesem Grund wurde das Transkriptom der Δrho -Mutante mittels RNA-Sequenzierung untersucht.

Für die RNA-Sequenzierung wurden der Wildtyp sowie die Δrho -Mutante in CGXII-Minimalmedium in Schüttelkolben in jeweils drei Replikaten kultiviert. Die Zellernte erfolgte in der frühen exponentiellen Wachstumsphase. Die aus den Zellen isolierte DNA-freie RNA wurde nach Entfernen der rRNA entsprechend dem Protokoll *TruSeq stranded mRNA* (Illumina) in strangspezifische Vollängen-cDNA-Bibliotheken umgewandelt und anschließend auf einem MiSeq (Illumina) im *paired-end* Modus sequenziert. Die Read-Länge betrug 71 bp. Die sequenzierten Reads wurden anschließend gegen das *C. glutamicum* ATCC 13032 Genom (Kalinowski *et al.*, 2003) gemappt. In Tabelle 4.6 sind die sequenzierten und gemappten Reads aufgeführt. Der Anteil der gemappten Reads beträgt für alle cDNA-Bibliotheken mehr als 98 %.

Tabelle 4.6: Sequenzierte und gemappte Reads der cDNA-Bibliotheken von der Δrho -Mutante und dem Wildtyp.

cDNA-Bibliothek	Sequenzierte Reads	Gemappte Reads	Gemappte Reads in %
Wildtyp Replikat 1	10.728.816	10.578.613	98,60
Wildtyp Replikat 2	5.375.148	5.328.384	99,13
Wildtyp Replikat 3	7.405.849	7.381.410	99,67
Δrho -Mutante Replikat 1	8.112.497	8.083.292	99,64
Δrho -Mutante Replikat 2	7.143.761	7.121.615	99,69
Δrho -Mutante Replikat 3	11.302.008	11.103.093	98,24

Die Normalisierung der cDNA-Bibliotheken und die Identifizierung der differenziell transkribierten Gene der Δrho -Mutante im Vergleich zum Wildtyp wurde mit der Software *DESeq2* (Love *et al.*, 2014) in Standardeinstellungen durchgeführt. Außerdem wurden die Replikate mittels hierarchischem Clustering auf Ähnlichkeit überprüft, um mögliche Ausreißer zu identifizieren (Abbildung A.3 im Anhang). Gene die eine Änderung in der Transkription ($\log_2\text{FoldChange}$, LFC) $\geq +1,0$ bzw. $\leq -1,0$ und einen angepassten p-Wert $< 0,05$ besitzen, wurden als differenziell transkribiert klassifiziert.

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

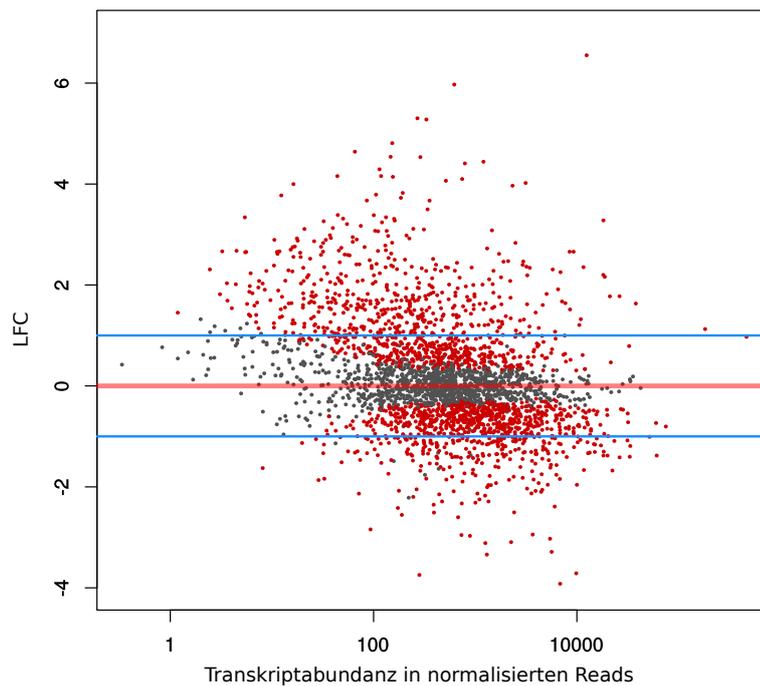


Abbildung 4.10: M/A-Plot der $\Delta\rho$ -Mutante im Vergleich zum Wildtyp. Aufgetragen sind der \log_2 FoldChange (LFC) gegen die Transkriptabundanz in (mit DESeq2 normalisierten) Reads. Jedes Gen ist durch einen Punkt dargestellt. Gene die die statistischen Bedingungen erfüllen (angepasster p-Wert < 0,05) sind rot markiert. Die blauen Linien markieren einen LFC von +1,0 bzw. -1,0. Gene die rot markiert sind und über der oberen bzw. unterhalb der unteren blauen Linie liegen wurden als differenziell transkribiert gewertet.

In der $\Delta\rho$ -Mutante sind 488 Gene stärker und 267 Gene schwächer transkribiert als im Wildtyp. Insgesamt ist somit ein Viertel des *C. glutamicum* Genoms (755 Gene) in der $\Delta\rho$ -Mutante differenziell transkribiert (Abbildung 4.10). Dies ist eine vergleichsweise sehr große Zahl an differenziell transkribierten Genen und deutet auf einen großen Einfluss des Rho-Faktors auf das Transkriptom von *C. glutamicum* hin. Das Gen *cg1478*, welches für ein hypothetisches Protein unbekannter Funktion kodiert, ist mit einem LFC von 6,55 am stärksten differenziell transkribiert. Das am schwächsten transkribierte Gen *cg3195*, kodierend für eine putative Flavin-Monooxygenase, ist in der $\Delta\rho$ -Mutante im Vergleich zum Wildtyp mit einem LFC von -3,92 am schwächsten transkribiert.

Eine Auswahl der am stärksten und schwächsten differenziell transkribierten Gene ist in Tabelle 4.7 aufgeführt. Die vollständige Liste aller differenziell transkribierter Gene befindet sich im Anhang (Tabelle S8).

Aufgrund der hohen Zahl an differenziell transkribierten Genen wurde eine Analyse auf Basis der COG-Klassen durchgeführt. Die COG-Klassen (*Clusters of Orthologous Groups*) klas-

4 Ergebnisse

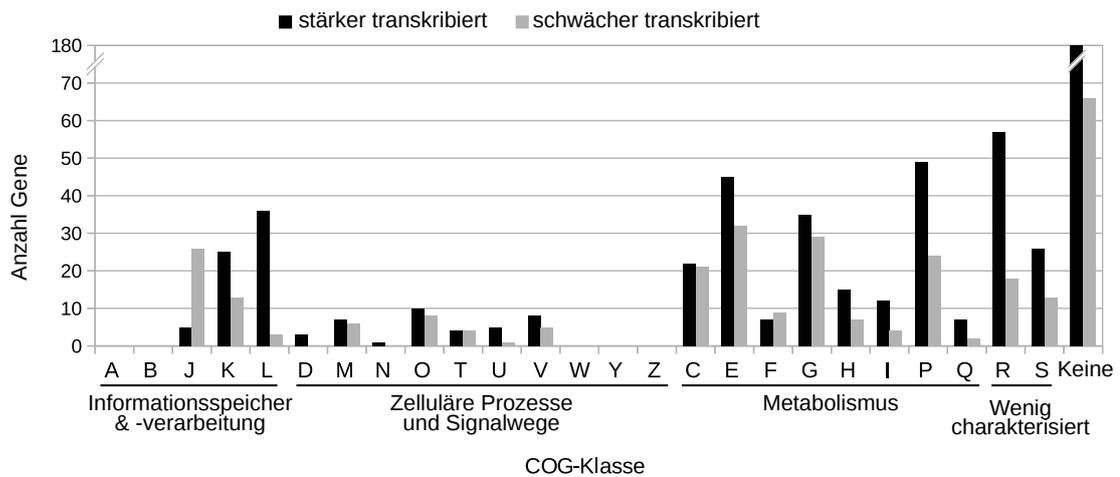


Abbildung 4.11: Klassifizierung der in der $\Delta\rho$ -Mutante differenziell transkribierten Gene nach COG-Klasse. Eine detaillierte Aufschlüsselung der COG-Klassen (Tatusov *et al.*, 1997) mit Beschreibung befindet sich in Tabelle A.1 auf Seite 236 im Anhang. Die Gruppe „Keine“ enthält alle Gene die keiner COG-Klasse zugeordnet sind. Die y-Achse ist zwischen den Werten 70 und 180 unterbrochen dargestellt.

sifizieren Gene anhand ihrer vorhergesagten Funktion in verschiedene Stoffwechselgruppen (Tatusov *et al.*, 1997). Eine Beschreibung der einzelnen COG-Klassen befindet sich in Tabelle A.1 im Anhang auf Seite 236. Wie in Abbildung 4.11 zu sehen, gehören besonders viele der differenziell transkribierten Gene den COG-Klassen „Informationsspeicher und -verarbeitung“ (J, K und L) sowie „Metabolismus“ (E, G und P) an. Hier ist auffällig, dass der überwiegende Teil der Gene, die in der Replikation und DNA-Reparatur (COG-Klasse L) involviert sind, in der $\Delta\rho$ -Mutante stärker transkribiert sind. Auch Gene, die in der Transkription (COG-Klasse K) involviert sind, sind überwiegend stärker transkribiert. Gleichzeitig sind Gene der Translation und Ribosomenstruktur (COG-Klasse J) überwiegend schwächer transkribiert. Die Deletion des für den Transkriptionsterminator Rho kodierenden Gens *rho* hat somit großen Einfluss auf essenzielle zelluläre Funktionen.

Tabelle 4.7: Liste der am stärksten bzw. schwächsten differenziell transkribierten Gene in der $\Delta\rho$ -Mutante im Vergleich zum Wildtyp (Ausschnitt). Gezeigt sind jeweils die 30 am stärksten bzw. schwächsten transkribierten Gene. Die vollständige Liste aller differenziell transkribierten Gene befindet sich im Anhang (Tabelle S8). LFC, \log_2 FoldChange.

Gen	Name	Genprodukt ¹	LFC
<i>cg1478</i>		Hypothetical protein	6,55
<i>cg0362</i>		Conserved hypothetical protein	5,97
<i>cg0363</i>		Putative secretion ATPase	5,30
<i>cg0428</i>	<i>ISCg17a</i>	Transposase, putative pseudogene	5,28

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

Tabelle 4.7: Fortgesetzt

Gen	Name	Genprodukt ¹	LFC
cg0369		Conserved putative secreted protein	4,81
cg0051		Putative transcriptional regulator, AraC-family	4,64
cg2048		Hypothetical protein	4,54
cg0427	<i>ISCg17a</i>	Transposase, putative pseudogene	4,53
cg0365		Putative membrane protein	4,44
cg1471		Hypothetical protein	4,41
cg0364		Putative membrane protein	4,29
cg0366		Conserved putative membrane protein	4,16
cg0426	<i>ISCg17a</i>	Transposase, putative pseudogene	4,16
cg0368		Conserved putative secreted protein	4,14
cg1118		Pyrimidine reductase, riboflavin biosynthesis	4,10
cg2443		Permease of the major facilitator superfamily - putative pseudogene	4,06
cg3374		Putative NADH-dependent flavin oxidoreductase	4,02
cg3290		Putative oxidoreductase	4,00
cg0083		Putative nicotinamide mononucleotide uptake permease	3,97
cg2244		Lyase class I family	3,82
cg2941		Putative threonine efflux transporter	3,79
cg1898		Hypothetical protein	3,77
cg0485		Hypothetical protein	3,73
cg1526		Putative multidrug efflux permease	3,67
cg0777		ABC-type putative iron-siderophore transporter	3,67
cg2577		Putative membrane protein Com family	3,50
cg3111		Putative secreted protein	3,39
cg2726		Putative membrane protein	3,38
cg1199		Hypothetical protein	3,34
cg0037	<i>ISCg22a</i>	Transposase, putative pseudogene	3,34
cg0834	<i>tusE</i>	ABC-type trehalose transporter	-2,13
cg2468		ABC-type transporter, permease subunit	-2,13
cg3156		Putative secreted protein	-2,13
cg3212		Putative carboxymuconolactone decarboxylase subunit	-2,15
cg2096		Putative membrane protein	-2,17
cg1373		Putative glyoxalase	-2,20
cg3048	<i>pta</i>	Phosphate acetyltransferase	-2,25
cg3047	<i>ackA</i>	Acetate kinase	-2,29
cg2610		ABC-type putative dipeptide/oligopeptide transporter	-2,29
cg2118		Transcriptional regulator protein, DeoR-family	-2,31
cg1595	<i>uspA2</i>	Universal stress protein UspA	-2,33
cg0759	<i>prpD2</i>	2-Methylcitrate dehydratase (propionate catabolism)	-2,35
cg0563	<i>rplK</i>	50S ribosomal protein L11	-2,39
cg2937		ABC-type putative dipeptide/oligopeptide transporter	-2,42
cg2470		ABC-type transporter	-2,50
cg1665		Putative secreted protein	-2,51
cg0762	<i>prpC2</i>	2-Methylcitrate synthase (propionate catabolism)	-2,56
cg0286		Conserved putative membrane protein	-2,60
cg0760	<i>prpB2</i>	Methylisocitrate lyase (propionate catabolism)	-2,84
cg2560	<i>aceA</i>	Isocitrate lyase	-2,94
cg2836	<i>sucD</i>	Succinate-CoA ligase (ADP-forming), alpha subunit	-2,95
cg2837	<i>sucC</i>	Succinate-CoA ligase (ADP-forming), beta subunit	-2,97

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

4 Ergebnisse

Tabelle 4.7: Fortgesetzt

Gen	Name	Genprodukt ¹	LFC
<i>cg2181</i>		ABC-type putative dipeptide/oligopeptide transporter, substrate-binding	-3,03
<i>cg2184</i>		ABC-type putative dipeptide/oligopeptide transporter, ATPase subunit	-3,09
<i>cg2183</i>		ABC-type putative dipeptide/oligopeptide transporter, permease subunit	-3,11
<i>cg3107</i>	<i>adhA</i>	Alcohol dehydrogenase	-3,29
<i>cg2182</i>		ABC-type putative dipeptide/oligopeptide transporter, permease subunit	-3,34
<i>cg3096</i>	<i>ald</i>	acetaldehyde dehydrogenase (acetylating)	-3,71
<i>cg1612</i>		Putative acetyltransferase	-3,74
<i>cg3195</i>		Putative flavin-containing monooxygenase	-3,92

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

Die Gene *cg0362* bis *cg0369* gehören zu den am stärksten differenziell transkribierten Genen in der $\Delta\rho$ -Mutante (LFC von 4,14 bis 5,97; Tabelle 4.7). Sie bilden ein Gen-Cluster das für einen, innerhalb der Actinobakterien hoch konservierten, putativen Typ IVb Pilus kodiert. Die mittlere Expressionsstärke der einzelnen Gene im Wildtyp reicht von circa 10 (*cg0364*) bis etwa 100 (*cg0365*) normalisierten Reads. Im Gegensatz dazu beträgt die mittlere Expressionsstärke in der $\Delta\rho$ -Mutante 230 (*cg0364*) bis 2.700 (*cg0365*) normalisierte Reads. Bisher sind in der Literatur keine Bedingungen beschrieben worden in denen dieses Gen-Cluster verstärkt transkribiert wird. Die $\Delta\rho$ -Mutante stellt somit die bisher einzige bekannte Bedingung dar, die zu einer verstärkten Transkription dieses Typ VI Pilus-Gen-Clusters führt. Aus diesem Grund befasst sich das Kapitel 4.2.11 ab Seite 115 vertiefend mit den Eigenschaften dieses Gen-Clusters.

Anhand der hohen Zahl an differenziell transkribierten Genen in der $\Delta\rho$ -Mutante, kann davon ausgegangen werden, dass der Rho-Faktor einen großen Einfluss auf das Transkriptom der Zelle hat. Da für *C. glutamicum* die Rho-abhängig terminierten Gene unbekannt sind, ist es schwierig, allein über die differenziell transkribierten Gene zwischen direkten (primären) und indirekten (sekundären) Effekten der *rho* Deletion zu unterscheiden.

4.2.4 Der Rho-Faktor beeinflusst die Transkription potentieller Fremd-DNA

C. glutamicum besitzt drei Prophagen-Regionen im Genom. Diese Regionen, genannt CGP1, CGP2 und CGP3, sind Überreste früherer Phageninfektionen (Kalinowski, 2005). CGP3 ist mit 192 Genen die mit Abstand größte Prophagen-Region, gefolgt von CGP1 mit 22 Genen und CGP2 mit 6 Genen. Unter den differenziell transkribierten Genen in der $\Delta\rho$ -Mutante waren auffällig viele Prophagen-Gene. So waren ca. 40 - 50 % der Gene von CGP3 und CGP2 in der $\Delta\rho$ -Mutante stärker transkribiert als im Wildtyp (Abbildung 4.12). Das Einzige in der $\Delta\rho$ -Mutante schwächer transkribierte Prophagen-Gen war *cg1905*, welches für ein hypothetisches Protein unbekannter Funktion kodiert. Die Prophagen-Regionen wurden somit

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

im Wildtyp unterschiedlich durch den Rho-Faktor direkt oder indirekt transkriptionell negativ beeinflusst bzw. „stillgelegt“. Während ein großer Anteil an Genen von CGP2 und CGP3 betroffen war, waren von CGP1 weniger als 10 % der Gene differenziell transkribiert.

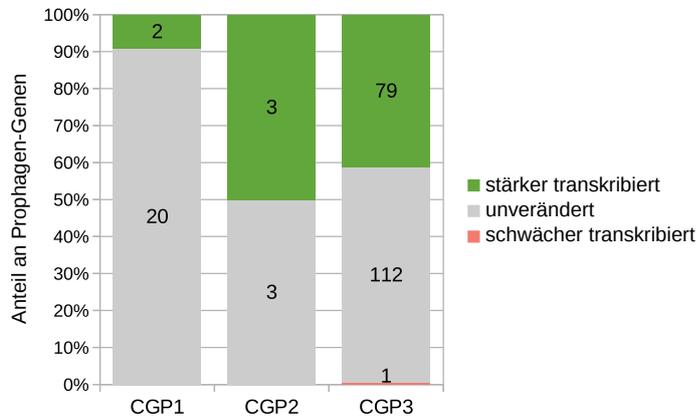


Abbildung 4.12: Differenziell transkribierte Gene der Prophagen-Regionen CGP1, CGP2 und CGP3 in der $\Delta\rho$ -Mutante. Die absoluten Genzahlen sind in den jeweiligen Balken dargestellt. Keine der Gene in den Prophagen-Regionen CGP1 und CGP2 sind in der $\Delta\rho$ -Mutante schwächer transkribiert als im Wildtyp. Von CGP3 ist das Gen *cg1905*, welches ein hypothetisches Protein kodiert, schwächer transkribiert als im Wildtyp.

In der $\Delta\rho$ -Mutante wurden im Vergleich zum Wildtyp verlängerte Transkripte gefunden, die offensichtlich durch das Fehlen des Transkriptionsterminators Rho nicht mehr korrekt terminiert werden können. Verlängerte Transkripte sind demnach ein direkter Effekt der Deletion von *rho*. Die folgenden Kapitel gehen detailliert auf diese verlängerten Transkripte ein. Trotz des hohen Anteils differenziell transkribierter Prophagen-Gene in der $\Delta\rho$ -Mutante konnte nur für vier Prophagen-Gene ein verlängertes Transkript festgestellt werden: *cg1524* (CGP1; LFC 1,05) und *cg1966* (CGP3; LFC 0,08), *cg2046* (CGP3; LFC 0,22) und *cg2057* (CGP3; LFC 1,57). Sie kodieren für hypothetische Proteine unbekannter Funktion. Nur zwei der vier Gene mit verlängertem Transkript sind differenziell transkribiert (*cg1524* und *cg2057*). Die Deletion des für den Terminationsfaktor Rho kodierenden Gens führt demnach nicht zu einer verstärkten Transkription Rho-abhängig terminierter Gene. Indirekte bzw. sekundäre Effekte, wie Transkriptionsaktivierung, sind somit für die verstärkte Transkription von Prophagen-Genen verantwortlich.

Die Liste der differenziell transkribierten Gene enthält viele Gene die zu Insertions-Elementen (IS-Elemente) gehören. Insgesamt besitzt *C. glutamicum* neun IS-Elementtypen die 49 Gene umfassen, von denen in der $\Delta\rho$ -Mutante 31 Gene (fast 66 %) verstärkt transkribiert sind (Abbildung 4.13). Auffällig ist, dass kein IS-Element schwächer transkribiert wird. Der Rho-Faktor ist demnach direkt oder indirekt an der transkriptionellen Regulation von IS-Elementen

4 Ergebnisse

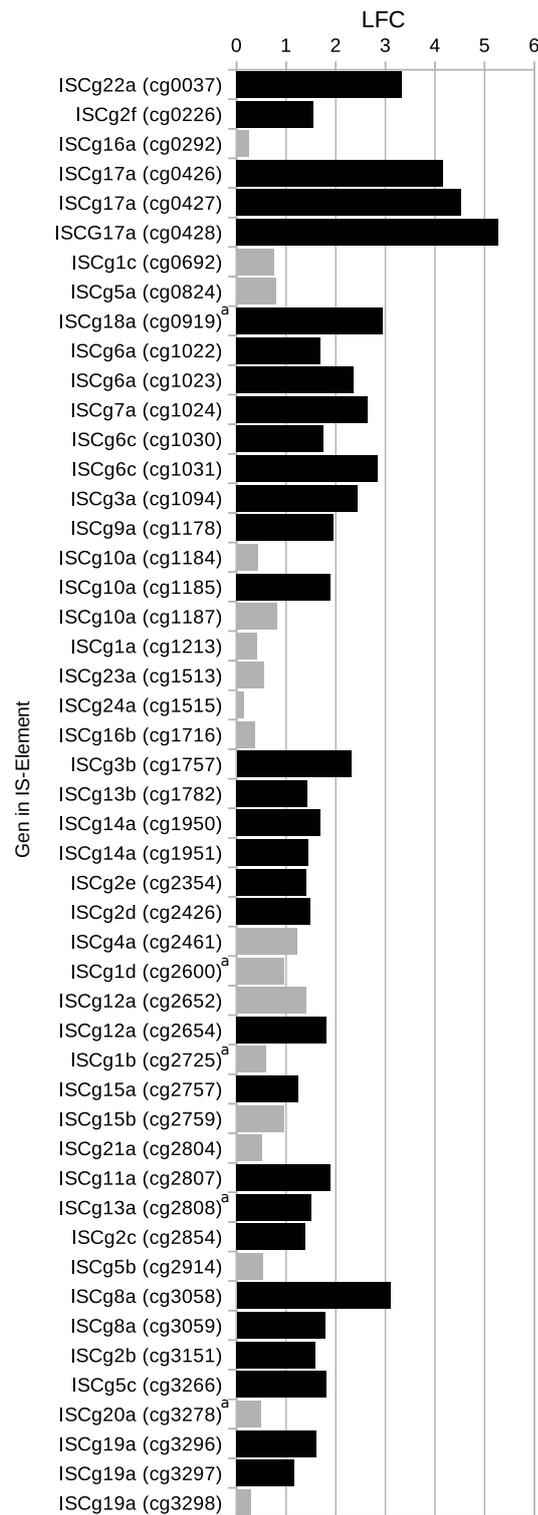


Abbildung 4.13: LFC-Werte (\log_2 (FoldChange)-Werte) von IS-Elementen der Δrho -Mutante im Vergleich zum Wildtyp. Gene mit schwarzem Balken sind differenziell transkribiert, Gene mit grauem Balken sind nicht differenziell transkribiert. ^averlängertes Transkript in der Δrho -Mutante.

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

beteiligt. Wie auch bei den Prophagen-Genen beobachtet, sind auch nur wenige Transkripte von IS-Elementen verlängert. IS-Elemente enthalten Gene, die möglicherweise ursprünglich aus fremden Quellen stammen. Für den Rho-Faktor aus *E. coli* ist bekannt, dass dieser die Transkription von Fremd-DNA, wie zum Beispiel Prophagen-Gene, inhibiert (Cardinale *et al.*, 2008; Menouni *et al.*, 2013). Die in dieser Arbeit vorgestellten Daten zeigen eine vergleichbare Funktion von Rho in *C. glutamicum*.

Verstärkt transkribierte IS-Elemente können zu einer Transposition des jeweiligen IS-Elements im Genom führen. Um auszuschließen, dass *nicht* die hohe Anzahl an verstärkt transkribierten IS-Elementen in der $\Delta\rho$ -Mutante, sondern die Deletion des für den Rho-Faktor kodierenden Gens selbst, für den transkriptionellen Phänotyp (differenziell transkribierte Gene & verlängerte Transkripte) verantwortlich ist, wurde geprüft, ob diese IS-Elemente transponiert sind. Dazu wurden die Distanzen der beiden Read-Partner von allen *paired-end* Reads gemessen, von denen ein Read-Partner innerhalb eines IS-Elements liegt. Insgesamt konnten für den Wildtyp 28 Reads und für die $\Delta\rho$ -Mutante 117 Reads gefunden werden, die auf ein möglicherweise transponiertes IS-Element hindeuten können (Abbildung A.5 im Anhang). Beide Werte sind im Vergleich zu allen *paired-end* Reads, die innerhalb von IS-Elementen liegen, verschwindend gering (Wildtyp 6.648 Reads, $\Delta\rho$ -Mutante 16.697 Reads). Der Anteil an eventuell transponierten IS-Elementen ist daher vernachlässigbar klein. Möglicherweise transponierte IS-Elemente sind demnach *nicht* für den Phänotyp der $\Delta\rho$ -Mutante verantwortlich.

4.2.5 Hinweise auf verlängerte Transkripte in der $\Delta\rho$ -Mutante

Die Hauptfunktion des Rho-Faktors ist die Termination von Transkripten (Hart und Roberts, 1991). Die Deletion des für Rho kodierenden Gens *rho* sollte sich daher phänotypisch auf die Länge von Rho-abhängig terminierten Transkripten auswirken. Bei Betrachtung der visualisierten RNA-Sequenzierungsdaten des *C. glutamicum* Wildtyps sowie der $\Delta\rho$ -Mutante mit der Software *ReadXplorer* fielen Bereiche *downstream* von einigen Genen auf, deren Read-Abdeckung in der $\Delta\rho$ -Mutante wesentlich später auf den Wert der Hintergrund-Read-Abdeckung abnehmen, als es im Wildtyp der Fall ist. Einer dieser Bereiche befindet sich *downstream* vom Gen *cg0700* und ist in Abbildung 4.14 exemplarisch dargestellt. Wie in der Abbildung zu sehen, sind neben proteinkodierenden Genen auch andere Transkriptklassen, wie *antisense* RNAs, von dieser „Verlängerung“ der Read-Abdeckung betroffen.

Abhängig von der Anzahl an Rho-abhängig terminierten Genen im Genom müsste die Gesamt-Read-Abdeckung des Genoms in der $\Delta\rho$ -Mutante größer sein als im Wildtyp, da die RNA-Polymerase kein Terminationssignal erhält und daher die Transkription erst verspätet

4 Ergebnisse

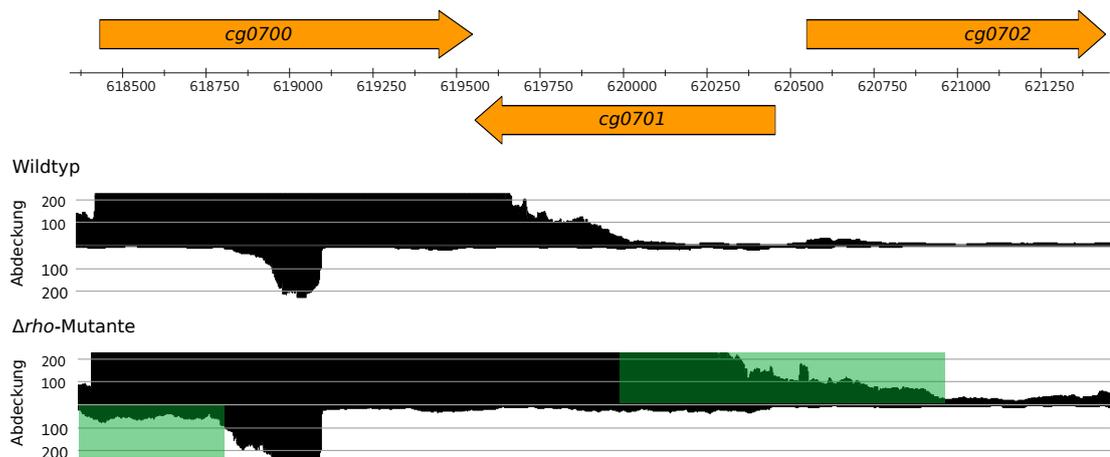


Abbildung 4.14: Beispiel verlängerter Transkripte in der $\Delta\rho$ -Mutante. Gezeigt ist der genomische Bereich der Gene *cg0700* bis *cg0702* (oberes Drittel), sowie die jeweilige Read-Abdeckung des *C. glutamicum* Wildtyps (Mitte) und der $\Delta\rho$ -Mutante (unteres Drittel), jeweils von einem Replikat. Die grünen Kästen markieren die ungefähren Bereiche der in der $\Delta\rho$ -Mutante verlängerten Transkripte des Gens *cg0700* sowie einer *antisense* RNA komplementär zu *cg0700*.

stoppt. Um dies zu überprüfen, wurde die Read-Abdeckung beider Stränge der drei sequenzierten Replikate des Wildtyps und der $\Delta\rho$ -Mutante des gesamten Genoms bestimmt und verglichen. Die Gesamt-Read-Abdeckung der $\Delta\rho$ -Mutante beträgt 93,39 % ($\pm 0,24$ %), die des Wildtyps 78,15 % ($\pm 4,61$ %) (Abbildung 4.15A). Der Unterschied in der Gesamt-Read-Abdeckung zwischen Wildtyp und $\Delta\rho$ -Mutante ist signifikant ($p < 0,05$) und beträgt etwa 15 %. Dies entspricht einer Read-Abdeckung von ca. 250 kbp pro Strang für das 3,28 Mbp große *C. glutamicum*-Genom.

Antisense RNAs, die komplementär zu proteinkodierenden Genen liegen, werden durch den Rho-Faktor terminiert (Peters *et al.*, 2012). Würden *antisense* RNAs durch einen Rho-unabhängigen intrinsischen Haarnadel-Terminator terminiert werden, würde dieser auch zu einer (vorzeitigen) Transkriptionstermination in der proteinkodierenden Region auf dem gegenüberliegenden Strang führen, was eine verkürzte mRNA zur Folge hätte. Aus diesem Grund wurde geprüft, ob die $\Delta\rho$ -Mutante verlängerte *antisense* RNAs besitzt. Dazu wurde die Read-Abdeckung auf dem gegenüberliegenden Strang von proteinkodierenden Regionen (CDS) zwischen $\Delta\rho$ -Mutante und Wildtyp verglichen. Die Read-Abdeckung von *antisense* RNAs beträgt im Wildtyp 1,87 % ($\pm 0,19$ %). Im Gegensatz dazu beträgt die Read-Abdeckung auf dem Gegenstrang zu CDS in der $\Delta\rho$ -Mutante 6,45 % ($\pm 0,55$ %) (Abbildung 4.15B). Der Unterschied der Read-Abdeckung von *antisense* RNAs zwischen Wildtyp und $\Delta\rho$ -Mutante ist signifikant ($p < 0,01$). Dies ist ein deutliches Indiz, dass *antisense* RNAs auch in *C. glutamicum* Rho-abhängig terminiert werden.

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

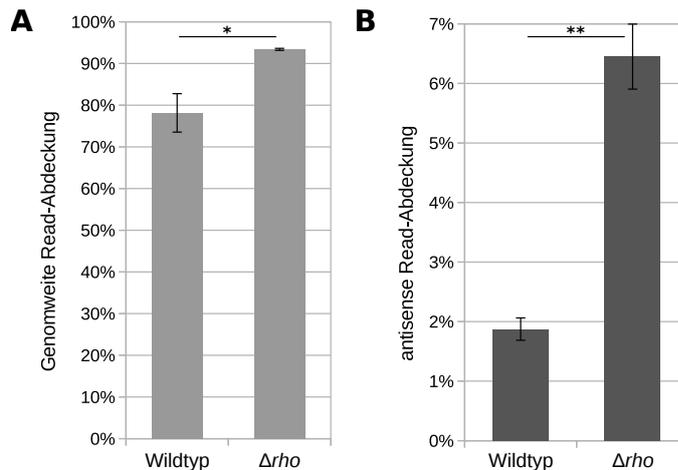


Abbildung 4.15: Vergleich der (A) genomweiten und (B) *antisense* Read-Abdeckung des *C. glutamicum* Wildtyps mit der $\Delta\rho$ -Mutante. Es wurde die normalisierte Read-Abdeckung von jeweils drei Replikaten verglichen. Die Unterschiede der genomweiten und der *antisense* Read-Abdeckung zwischen Wildtyp und Mutante sind signifikant (* p-Wert < 0,05; ** p-Wert < 0,01). Als Signifikanztest wurde der Students t-Test (beidseitig) durchgeführt.

Da in der $\Delta\rho$ -Mutante 489 Gene im Vergleich zum Wildtyp verstärkt transkribiert sind und die mittlere Genlänge in *C. glutamicum* etwa 800 bp beträgt, könnte die in der $\Delta\rho$ -Mutante beobachtete Gesamt-Read-Abdeckung allein von der Zahl an differenziell transkribierten Genen hervorgerufen sein ($489 \cdot 800 \text{ bp} \approx 391 \text{ kbp}$). Um zu zeigen, dass die erhöhte Gesamt-Read-Abdeckung tatsächlich von verlängerten Transkripten und nicht von verstärkt transkribierten Transkripten hervorgerufen wird, wurde eine bioinformatische Methode zur genomweiten Identifizierung von verlängerten Transkripten aus RNA-Sequenzierungsdaten entwickelt.

4.2.6 Identifizierung von Rho-abhängig terminierten Transkripten

Wie im vorherigen Abschnitt beschrieben, wurden in der $\Delta\rho$ -Mutante im Vergleich zum Wildtyp verlängerte Transkripte von Genen gefunden. Durch die Deletion des für den Transkriptionsterminator kodierenden Gens *rho*, findet keine korrekte Termination von Rho-abhängig terminierten Transkripten statt, so dass die RNA-Polymerase kein Terminationssignal erhält und die Transkription über die native Länge des Transkripts fortsetzt. Dies führt zu verlängerten Transkripten in der $\Delta\rho$ -Mutante. Verlängerte Transkripte sind somit ein direkter (primärer) Effekt der Deletion von *rho*. Da für Rho keine Binde- und Erkennungsstellen bekannt sind, ist eine Vorhersage von Rho-abhängig terminierten Genen *a priori* nicht möglich.

4 Ergebnisse

(Ray-Soni *et al.*, 2016). Durch die Deletion des für Rho kodierenden Gens bietet sich hier nun die Möglichkeit, in Kombination mit RNA-Sequenzierung, Rho-abhängig terminierte Gene anhand ihrer Transkriptverlängerung genomweit in *C. glutamicum* zu identifizieren.

Verlängerte Transkripte zeigen sich in der $\Delta\rho$ -Mutante durch eine im Vergleich zum Wildtyp verspätete Abnahme der Read-Abdeckung (vgl. Abbildung 4.14 auf Seite 102). Die Identifizierung von verlängerten Transkripten kann somit über den genomweiten Vergleich der Read-Abdeckung zwischen Wildtyp und $\Delta\rho$ -Mutante durchgeführt werden. Ein verlängertes Transkript wurde identifiziert, wenn die mittlere, auf die cDNA-Bibliotheksgröße normalisierte, Read-Abdeckung in der $\Delta\rho$ -Mutante innerhalb eines Bereichs L mindestens um den Faktor F höher ist, als im Wildtyp. Zur Erhöhung der Sensitivität in schwach transkribierten Bereichen wurde die Read-Abdeckung der Replikate zusammengefasst. Über eine Auswahl an manuell identifizierten, eindeutig verlängerten Transkripten (*true positives*), wurden die Parameter L und F getestet und soweit optimiert, bis alle manuell identifizierte verlängerte Transkripte identifiziert wurden. Die ermittelten optimalen Werte der Parameter sind 5 für den Faktor F und 75 bp für den Bereich L . Das Programm zur genomweiten automatischen Identifizierung von verlängerten Transkripten wurde vom Autor dieser Dissertation konzipiert und von David Brandt (CeBiTec, Bielefeld) in der Programmiersprache *Python* Version 2.7 (*Python Software Foundation*) implementiert.

Tabelle 4.8: Transkripte für die eine Rho-abhängige Termination durch Transkriptverlängerung als auch eine Rho-unabhängige Termination über einen Haarnadel-Terminator von Pfeifer-Sancar *et al.* (2013) identifiziert wurde. Diese Transkripte wurden manuell auf eine Verlängerung der Read-Abdeckung untersucht, um den tatsächlichen Terminator zu bestimmen. Die Transkripte, die trotz vorhergesagtem Haarnadel-Terminator eine offensichtliche Transkriptverlängerung aufwiesen, wurden den Rho-abhängig terminierten Transkripten zugeordnet. Der manuell bestimmte Terminator ist in der letzten Spalte mit Begründung angegeben.

Gen	Name	Vorhergesagte Haarnadel-Struktur	Vorhergesagte Verlängerung	Korrekt	Terminator (Begründung)
cg0310	<i>katA</i>	<i>loop</i>	ja	Haarnadel	(Verlängerung nicht erkannt)
cg0463		<i>stem 1</i>	ja	Haarnadel	(Verlängerung nicht erkannt)
cg0634	<i>rplO</i>	<i>loop</i>	ja	Haarnadel	(Verlängerung nicht erkannt)
cg0808	<i>wbpC</i>	<i>stem 1</i>	ja	Rho	(Verlängerung erkannt)
cg0858		<i>T-Tail</i>	ja	Rho	(Verlängerung erkannt)
cg1164	<i>ispH</i>	<i>loop</i>	ja	Haarnadel	(Verlängerung nicht erkannt)
cg1214		<i>loop</i>	ja	Rho	(Verlängerung erkannt)
cg1311		<i>stem 1</i>	ja	Haarnadel	(Verlängerung nicht erkannt)
cg1479	<i>glgP1</i>	<i>stem 1</i>	ja	Haarnadel	(Verlängerung nicht erkannt)
cg1714		<i>stem 1</i>	ja	Rho	(Verlängerung erkannt)
cg1966		<i>stem 1</i>	ja	Rho	(Verlängerung erkannt)
cg2325		<i>stem 1</i>	ja	Haarnadel	(Verlängerung nicht erkannt)
cg2880		<i>stem 1</i>	ja	Rho	(Verlängerung erkannt)

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

Mit Hilfe des Programms zur automatischen Identifizierung von verlängerten Transkripten wurden insgesamt 930 verlängerte Transkripte identifiziert. Davon wurden 626 (67 %) *anti-sense* RNAs, 252 (27 %) proteinkodierenden Genen (CDS), 23 (2 %) intergenischen, 18 (2 %) intragenischen Transkripten und 11 (1 %) nichtkodierenden RNAs (ncRNAs) zugeordnet. Zur weiteren Validierung der vorhergesagten verlängerten Transkripte wurden diese mit den für *C. glutamicum* bereits bekannten Rho-unabhängig terminierten proteinkodierenden Genen (Pfeifer-Sancar *et al.*, 2013) verglichen. Für 95 % (239) der hier identifizierten verlängerten proteinkodierenden Transkripte wurden keine Strukturen für Rho-unabhängige Terminatoren beschrieben. Für die 13 verbliebenen verlängerten Transkripte wurde ein kompletter oder teilweiser Rho-unabhängiger Haarnadel-Terminator vorhergesagt (Pfeifer-Sancar *et al.*, 2013). Diese Transkripte wurden manuell auf eine tatsächliche Transkriptverlängerung untersucht (Tabelle 4.8). Nur sieben der als verlängert vorhergesagten Transkripte wiesen tatsächlich keine sichtbare Verlängerung der Read-Abdeckung auf. Die Rate an falsch-positiv identifizierten verlängerten Transkripten ist demnach mit 2,8 % (7 von 252) gering. Die entwickelte Methode zur Vorhersage von eindeutig verlängerten Transkripten ist somit verlässlich. Es verbleiben 245 proteinkodierende verlängerte Transkripte, also eine Gesamtzahl von 923 eindeutig verlängerten Transkripten (Abbildung 4.16). Die vollständige Liste der als verlängert identifizierten Transkripte befindet sich im Anhang (Tabelle S9 & Tabelle S10). Wie oben bereits erwähnt, sind diese verlängerten Transkripte ein direkter Effekt der Deletion des für den Rho-Faktor kodierenden Gens. Somit entspricht die Liste der verlängerten Transkripte den Rho-abhängig terminierten Transkripten bzw. Genen. Dies ist die erste umfassende Liste von Rho-abhängig terminierten Genen in *C. glutamicum*.

Eine Auswahl der mittels RNA-Sequenzierung identifizierten verlängerten Transkripte wurde mit qRT-PCR validiert. Dazu wurden drei Primerpaare für jedes zu validierende Transkript erstellt: das erste Primerpaar bindet auf dem annotierten Bereich des Gens, das zweite Primerpaar bindet in dem Bereich, in dem die Read-Abdeckung im Wildtyp abnimmt, in der $\Delta\rho$ -Mutante aber noch konstant ist („Übergang“) und das dritte Primerpaar bindet im Bereich hinter dem zweiten Primerpaar in dem die Read-Abdeckung im Wildtyp auf Hintergrundniveau gesunken, in der $\Delta\rho$ -Mutante aber noch deutlich erkennbar ist („Verlängerung“). Abbildung A.6 zeigt beispielhaft die Primerpaare für das Gen *cg1478*. Da viele der identifizierten verlängerten Transkripte in Bereichen lagen, die eine hohe Read-Abdeckung auf beiden Strängen aufwiesen, musste eine strangspezifische qRT-PCR durchgeführt werden, um das Transkript auf einem Strang quantifizieren zu können. Für alle untersuchten Transkripte konnten in den Bereichen „Übergang“ (zweites Primerpaar) und/oder „Verlängerung“ (drittes Primerpaar) eine in der $\Delta\rho$ -Mutante deutlich erhöhte Transkriptmenge im Vergleich zum Wildtyp mittels qRT-PCR gemessen werden (Abbildung A.7 im Anhang). Alle

4 Ergebnisse

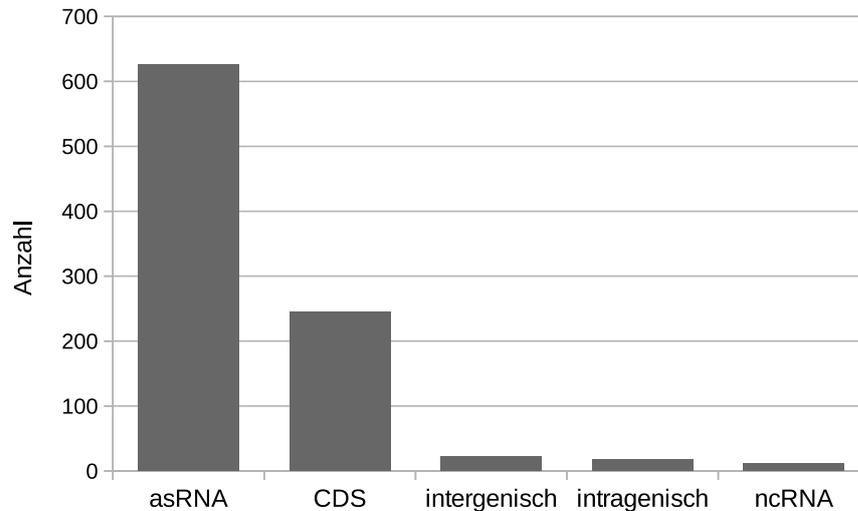


Abbildung 4.16: In der $\Delta\rho$ -Mutante als verlängert identifizierte Transkripte unterteilt nach Transkriptklasse. asRNA, antisense RNA; CDS, proteinkodierende Transkripte (mRNA); intergenisch, intergenisches Transkript; intragenisch, intragenisches Transkript; ncRNA, nichtkodierende RNA.

getesteten Transkripte sind tatsächlich in der $\Delta\rho$ -Mutante verlängert. Die Daten aus der RNA-Sequenzierung sind somit valide.

4.2.7 Längenanalyse der Rho-abhängig terminierten Transkripte

Ausgehend von der genomweiten Identifizierung von in der $\Delta\rho$ -Mutante verlängerten und somit Rho-abhängig terminierten Genen können weitere Analysen durchgeführt werden. Im Folgenden wird auf die gemessenen Längen der verlängerten proteinkodierenden Transkripte eingegangen. So wurden Verlängerungen relativ zur ursprünglichen Transkriptlänge im Wildtyp von ca. 25 bis ca. 6225 nt (2 – 2500 %, Mittelwert ca. 900 nt) ermittelt (Abbildung 4.17). Da die Read-Abdeckung eines verlängerten Transkripts zum 3'-Ende graduell abnimmt (vgl. Abbildung 4.14), wird vermutet, dass die RNA-Polymerase die naszierende RNA spontan entlässt und die Transkription abbricht. Somit existieren unterschiedlich lange Spezies von verlängerten Transkripten des selben Gens in der Zelle. Die Längenbestimmung eines verlängerten Transkripts bezieht sich daher auf die Summe aller Spezies dieses Transkripts. Aufgrund der Fragmentierung der RNA während der Erstellung der cDNA-Bibliothek können Einzeltranskripte nicht mehr direkt aufgelöst werden.

Eine Verlängerung eines Transkripts von einigen hundert Nukleotiden könnte aufgrund der hohen Kodierdichte von bakteriellen Genomen zur Transkription von *downstream* Genen führen. Es würde eine Art „künstliches“ Operon entstehen, dessen Transkript neben den UTR-

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

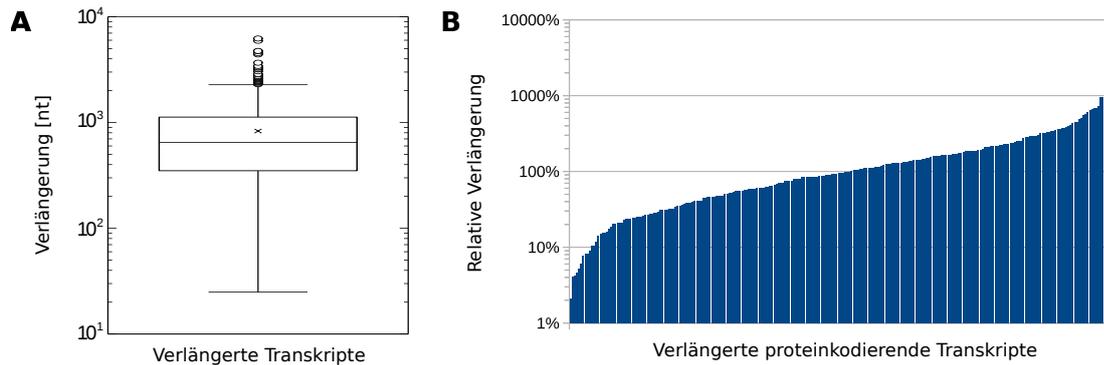


Abbildung 4.17: Längenverteilungen der verlängerten Transkripte in der $\Delta\rho$ -Mutante. Der Boxplot in (A) zeigt die Längen der Transkript-Verlängerungen aller verlängerten Transkripte (asRNA, proteinkodierend, ncRNA, inter- und intragenisch). In (B) ist die relative Länge der Verlängerung für proteinkodierenden Transkripte, bezogen auf die Gen-Länge im Wildtyp, dargestellt. Eine relative Verlängerung von 100 % bedeutet eine Verdoppelung der Transkriptlänge in der $\Delta\rho$ -Mutante im Vergleich zur Transkriptlänge im Wildtyp. Aus Gründen der Übersichtlichkeit wurden die Gennamen auf der Abszissenachse weggelassen.

und CDS-Bereichen auch den kompletten Promotorbereich des nachfolgenden Gens enthält. Im Transkriptom der $\Delta\rho$ -Mutante konnten diese „künstlichen“ Operons nicht nachgewiesen werden. Der größte Teil von extrem langen Transkriptverlängerungen überspannt keine *downstream* Gene auf dem selben, sondern auf dem Gegenstrang und bildet somit sehr lange *antisense* RNAs zu den Genen auf dem gegenüberliegenden Strang. Es ist unklar, ob diese langen Transkripte *in vivo* tatsächlich RNA-Duplexe mit *sense* Transkripten bilden. Es wäre denkbar, dass diese langen Transkripte komplexe interne Sekundärstrukturen bilden, die eine komplementäre Bindung von *sense* Transkripten verhindern.

Generell hat eine Verlängerung des Transkripts nur geringen Einfluss auf die Transkriptionsstärke eines Gens. Von den 245 proteinkodierenden Genen mit verlängertem Transkript sind nur 43 Gene (18 %) in der $\Delta\rho$ -Mutante stärker und fünf (2 %) Gene schwächer transkribiert als im Wildtyp (Abbildung 4.20, linker Balken).

4.2.8 Analyse von *Rho utilization sites*

Rho terminiert die Transkription von bestimmten RNAs. Da Rho nicht alle RNAs in der Zelle terminiert, müssen Erkennungsstellen für Rho-terminierte RNAs, so genannte *Rho utilization sites* (*rut sites*) existieren. Für den Rho-Faktor in *E. coli* konnte eine Präferenz von C-reichen und G-armen RNA-Sequenzen beobachtet werden. Diese Sequenzen werden als CG-Blase (*CG-bubble*) bezeichnet und als Indikatoren für Rho-abhängige Termination herangezogen (Alifano *et al.*, 1991). Die Transkripttermination findet meist 10 bis 20 nt, maximal 100 nt,

4 Ergebnisse

downstream der *rut site* statt (Alifano *et al.*, 1991; Koslover *et al.*, 2012; Ray-Soni *et al.*, 2016). Anhand von Tertiärstrukturanalysen wird davon ausgegangen, dass eine *rut site* von 70 bis 90 nt Länge benötigt wird, damit alle Untereinheiten des Rho-Hexamers mit der CG-Blase in Kontakt kommen können (McSwiggen *et al.*, 1988). Diese Annahme ist jedoch umstritten, da nicht klar ist, ob alle Untereinheiten gleichzeitig oder konsekutiv an die CG-Blasen binden (Canals *et al.*, 2010).

Der Rho-Faktor in *C. glutamicum* ist über 50 % größer als der Rho-Faktor in *E. coli* und weist deutliche Unterschiede in den Motiven für die RNA-Bindung auf (vgl. Abbildung 4.8 und Tabelle 4.5 auf Seite 91). Es ist daher davon auszugehen, dass sich die *rut site* in Zusammensetzung und Länge von *C. glutamicum* Rho ebenfalls deutlich von der von *E. coli* Rho unterscheidet.

In dieser Arbeit wurde erstmalig eine genomweite Identifizierung der Rho-abhängig terminierten Transkripte in *C. glutamicum* durchgeführt (siehe Abschnitt „Identifizierung von Rho-abhängig terminierten Transkripten“ auf Seite 103). Anhand der dabei erhaltenen umfangreichen Liste mit 245 Rho-abhängigen proteinkodierenden Transkripten bietet sich hier die Möglichkeit, *rut sites* für den Rho-Faktor in *C. glutamicum* zu analysieren. Dazu wurde zunächst der Gehalt an C und G der Rho-abhängigen Transkripte mit der Software *EMBOSS FREAK* (Rice *et al.*, 2000) in Standardeinstellungen bestimmt und anschließend gegen die Transkriptsequenz grafisch dargestellt. Abbildung 4.18 zeigt beispielhaft die C- bzw. G-Anteile für die Gene *cg1478*, *rluC1*, *ribM* und *thiF*. Zur besseren Veranschaulichung sind CG-Blasen (höherer C- als G-Anteil) rot und GC-Blasen (höherer G- als C-Anteil) blau dargestellt. Beim manuellen Sichten der 245 Diagramme fiel auf, dass die Länge und Höhe der CG-Blasen stark variieren. So haben manche Rho-abhängig terminierte Transkripte CG-Blasen eine Länge von wenigen Nukleotiden (vgl. *rluC1* in Abbildung 4.18). Andere Transkripte hingegen besitzen CG-Blasen von >300 nt Länge und nur kurze GC-Blasen (vgl. *thiF* in Abbildung 4.18). Auch die relativen Anteile von C bzw. G variieren zwischen den Rho-abhängig terminierten Transkripten und über deren Transkriptsequenz enorm. So beträgt der Anteil an C in einer CG-Blase im Transkript von *rluC1* maximal 35 %, während er im Transkript von *thiF* auf fast 60 % ansteigt (Abbildung 4.18).

In einer quantitativen Analyse wurden die Längen und Höhen aller CG-Blasen von Rho-abhängig terminierten Transkripten ermittelt. Außer für das Transkript von *cg4014* konnten für alle untersuchten Transkripte mindestens eine CG-Blase mit einer Länge von mindestens 10 nt identifiziert werden. Die längste CG-Blase beträgt 478 nt und befindet sich im Transkript von *cg3046*. Im Mittel sind die CG-Blasen 35 nt lang. Die CG-Blasenhöhe in *C. glutamicum* für Rho-abhängig terminierte Transkripte reicht von mindestens 0,07 bis maximal 0,63 und beträgt im Mittel 0,21. Eine Blasenhöhe von 0,21 bedeutet 21 % mehr C als G an dieser Stelle.

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

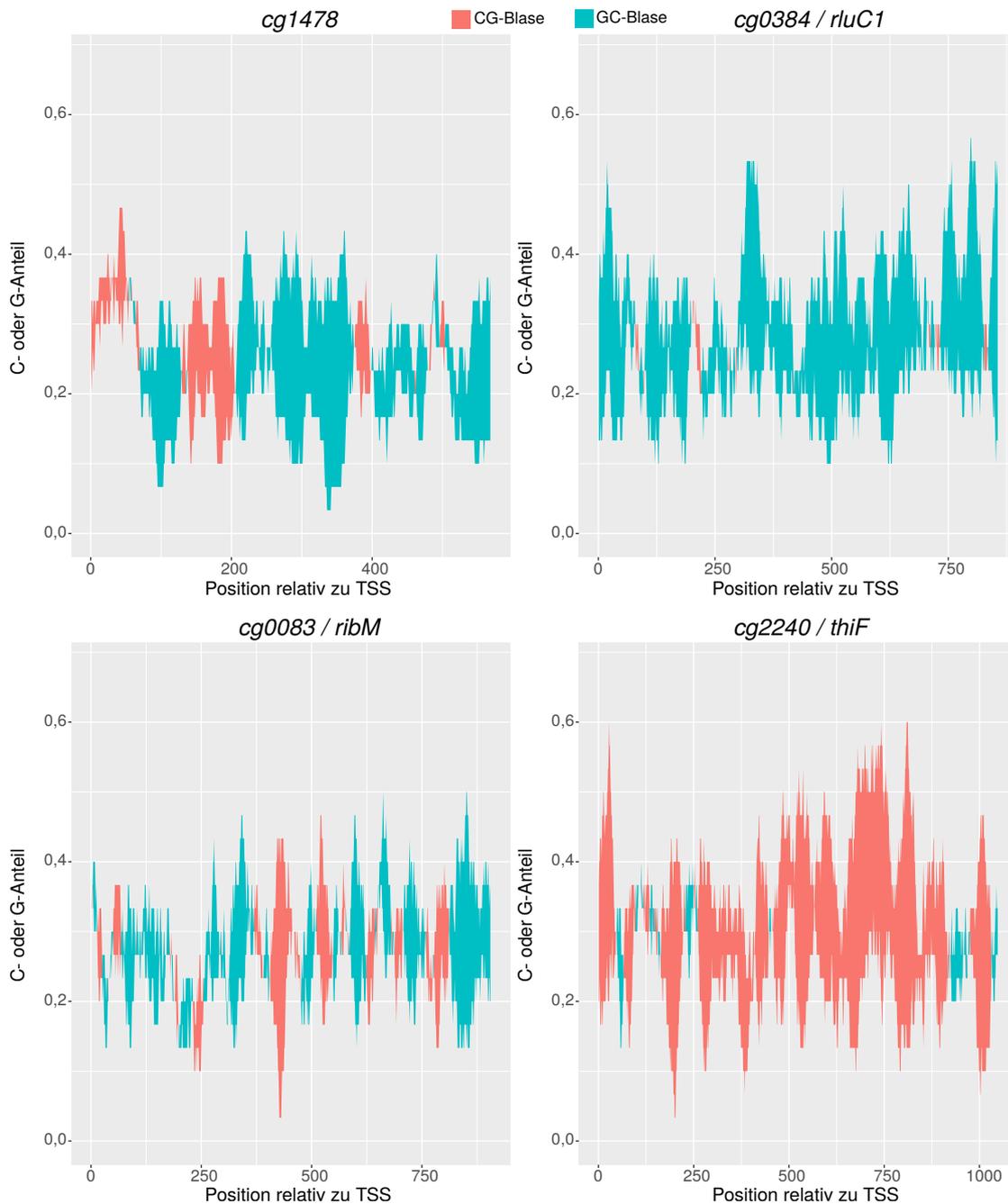


Abbildung 4.18: CG-Blasen einiger Rho-abhängig terminierter Transkripte. Dargestellt sind die Anteile der Basen C bzw. G über die Transkriptposition relativ zum Transkriptionsstart (TSS) beispielhaft für *cg1478* (links oben), *rluC1* (rechts oben), *ribM* (links unten) und *thiF* (rechts unten). Bereiche in denen der Anteil an Cytosin (C) den von Guanin (G) übersteigt, werden als CG-Blase bezeichnet (rote Bereiche). Analoges gilt für GC-Blasen (blaue Bereiche). Die Basenanteile wurden mit der Software *EMBOSS Freak* (Rice *et al.*, 2000) berechnet.

4 Ergebnisse

Die Parameter Länge und Höhe der CG-Blasen von Rho-abhängig terminierten Transkripten besitzen eine hohe Variation. Es ist daher nicht klar, welche CG-Blasen als *rut sites* agieren und somit tatsächlich von Rho erkannt werden.

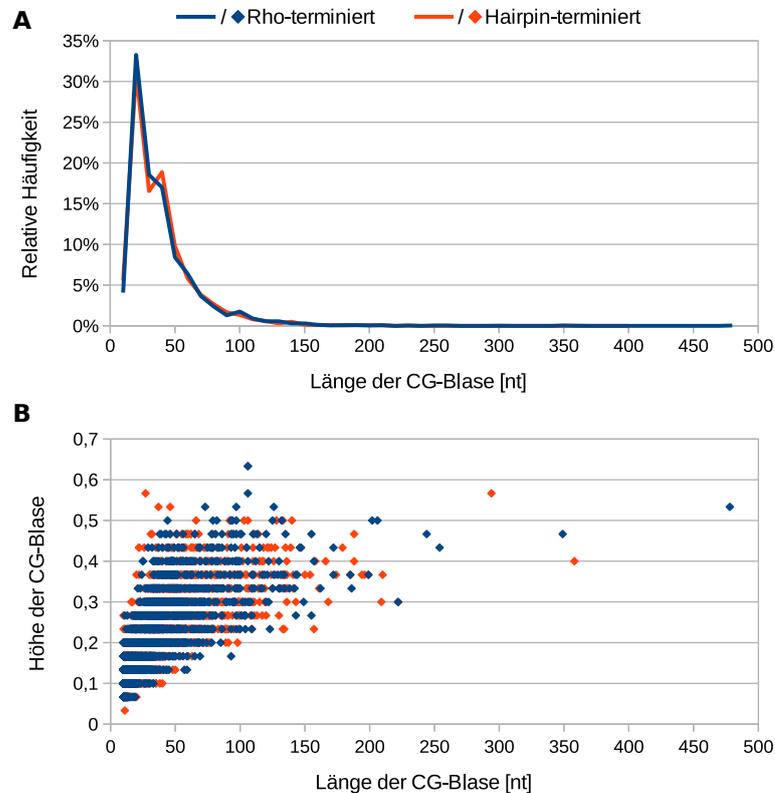


Abbildung 4.19: Längen- und Höhenvergleich der CG-Blasen von Rho- und Haarnadel-terminierten Transkripten. (A) Histogramm der relativen Längenverteilung der CG-Blasen von Rho- und Haarnadel-terminierten Transkripte. (B) Gegenüberstellung der CG-Blasenhöhe gegen die CG-Blasenlänge von Rho- und Haarnadel-terminierten Transkripten.

Für *C. glutamicum* sind auch die Haarnadel-terminierten Transkripte bekannt (Pfeifer-Sancar *et al.*, 2013). Dies erlaubt eine Kontrolle der für Rho-abhängig terminierte Transkripte bestimmten CG-Blasen. Es wurden die CG-Blasen sowie deren Längen und Höhen für die 320 bekannten Haarnadel-terminierten Transkripte bestimmt. Die Haarnadel-terminierten Transkripte besitzen ebenfalls CG-Blasen mit vergleichbaren Eigenschaften wie die Rho-abhängig terminierten Transkripte. So beträgt die mittlere CG-Blasenlänge ebenfalls 35 nt. Die Häufigkeitsverteilung der CG-Blasenlänge von Haarnadel-terminierten und Rho-terminierten Transkripten ist nahezu identisch (Abbildung 4.19A). Auch die Verteilung von CG-Blasenlänge zu -Blasenlänge ist für beide Terminationsarten nicht unterscheidbar (Abbildung 4.19B). Weder die Länge noch die Höhe von CG-Blasen, und daher vermutlich auch das bloße Vor-

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

handensein von CG-Blasen ist demnach kein Indikator für eine *rut site*. Eine Rho-abhängige Termination ist somit nicht über die CG-Blasen in *C. glutamicum* vorhersagbar. Auch eine Sequenzanalyse mit der Software *Improbizer* (Ao *et al.*, 2004) ergab keine für Rho-terminierte Transkripte spezifischen Resultate (Daten nicht gezeigt). Es wird daher vermutet, dass es weitere Faktoren geben muss, die eine Erkennung und Termination durch den Rho-Faktor vermitteln. Diese Faktoren sind allerdings unbekannt.

4.2.9 Die Termination von *antisense* RNA ist Rho-abhängig

Den mit Abstand größten Anteil an verlängerten Transkripten in der $\Delta\rho$ -Mutante machen die *antisense* RNAs aus. So sind von 923 verlängerten Transkripten 626 (67 %) verlängerte *antisense* RNAs (Abbildung 4.16). *Antisense* RNAs (asRNAs) befinden sich auf dem Gegenstrang zu einem anderen Gen und sind verglichen mit mRNA kurz. Von 626 identifizierten verlängerten asRNAs liegen 618 auf dem Gegenstrang zu proteinkodierenden Sequenzen (CDS), zwei auf dem Gegenstrang von nichtkodierenden RNAs (*scg0220* und *scg1630*) und sechs auf dem Gegenstrang der sechs für die rRNA kodierenden rDNA-Operons (*rrnA* bis *rrnF*).

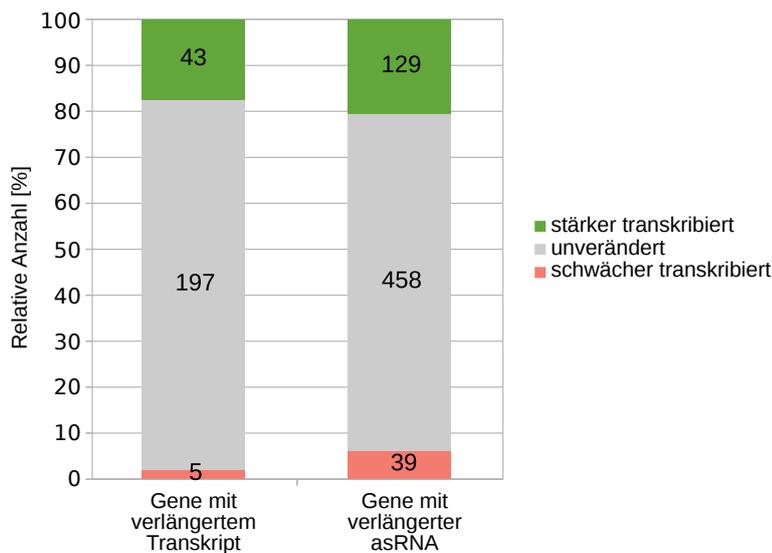


Abbildung 4.20: Einfluss der Transkriptverlängerung auf die differenzielle Transkription. Dargestellt sind proteinkodierende Gene mit Transkriptverlängerung (links) und proteinkodierende Gene mit verlängerter *antisense* RNA (asRNA) auf dem Gegenstrang (rechts) unterteilt nach ihrem differenziellen Transkriptionsstatus (stärker [grün], unverändert [grau] oder schwächer [rot] transkribiert als im Wildtyp). Die absoluten Genzahlen sind in den jeweiligen Balken angegeben.

4 Ergebnisse

Antisense RNAs sind teilweise oder vollständig komplementär zum Transkript des Gegenstranges und können mit ihm einen RNA-Duplex bilden. Die Bindung von asRNA kann einen stabilisierenden oder destabilisierenden Effekt auf das gebundene Transkript haben (Thomason und Storz, 2010). In der $\Delta\rho$ -Mutante sind asRNAs stark verlängert, was längere komplementäre Sequenzen und wahrscheinlich verstärkte Bindungen zwischen asRNA und *sense* Transkript zur Folge hat. Der Effekt einer vermuteten verstärkten Bindung auf proteinkodierende Gene ist anhand ihrer differentiellen Transkription abschätzbar. So sind 129 Gene mit asRNAs in der $\Delta\rho$ -Mutante stärker und 39 Gene mit asRNAs schwächer transkribiert. Die meisten Gene (450) sind in ihrer Transkription unbeeinflusst (Abbildung 4.20, rechter Balken). Die meisten asRNAs scheinen daher, zumindest unter den hier vorliegenden Bedingungen, keine regulatorische Funktion auf das *sense* Transkript zu haben.

Für die circa 21 % (129) der Gene mit asRNA konnte eine verstärkte Transkription festgestellt werden (Abbildung 4.20, rechter Balken). Möglicherweise hat die Bindung der verlängerten asRNA einen stabilisierenden Effekt auf das *sense* Transkript. Es kann hier aber nicht unterschieden werden, ob es sich um einen direkten oder indirekten Effekt der Deletion von *rho* handelt. Möglicherweise ist die Transkription des Gens durch einen Transkriptionsaktivator unabhängig von der asRNA verstärkt worden. Dann wäre der Effekt der gebundenen asRNA zu vernachlässigen. Bei einem direkten Effekt der komplementären Bindung der asRNA, könnte der stabilisierende Effekt dadurch entstanden sein, dass die verlängerte asRNA eine RNase-Schnittstelle überdeckt und somit nicht mehr von einer Einzelstrang-spezifischen RNase abgebaut werden kann.

Über die Funktion der zu den rDNA-Operons komplementären asRNAs sowie zu den beiden asRNAs, die sich auf dem Gegenstrang der nichtkodierenden RNAs (ncRNAs) *scg0220* und *scg1630* befinden, ist nichts bekannt.

4.2.10 Rho reguliert die Termination bestimmter Riboswitches

Für den FMN-Riboswitch in der 5'-UTR von *ribM* (*cgR_0083*) in *C. glutamicum* R wurde eine Rho-abhängige Regulation nachgewiesen (Takemoto *et al.*, 2015). Das Homolog von *ribM* in *C. glutamicum* ATCC 13032 ist *cg0083*, welches auch den FMN-Riboswitch in der 5'-UTR besitzt (Pfeifer-Sancar *et al.*, 2013). Aus diesem Grund wurde überprüft, ob der FMN-Riboswitch in *C. glutamicum* ATCC 13032 ebenfalls Rho-abhängig reguliert wird.

Das Gen *ribM* kodiert für einen Riboflavin-Transporter. Unter normalen Kulturbedingungen liegt genug Flavinmononukleotid (FMN) vor, sodass nur der FMN-Riboswitch in der 5'-UTR, nicht aber die *downstream* befindliche CDS, von *ribM* transkribiert wird. In der *C. glutamicum* R $\Delta\rho$ -Mutante wurde das Transkript nicht vorzeitig terminiert, sondern in voller Länge

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

(5'-UTR und CDS) transkribiert und zeigt somit eine Rho-abhängige Termination des FMN-Riboswitches an (Takemoto *et al.*, 2015). Zusätzlich konnten Takemoto *et al.* (2015) zeigen, dass der FMN-Riboswitch in *ribM* auch durch die RNase E/G reguliert wird. Die genaue Position der RNase E/G-Schnittstelle konnten sie allerdings nicht angeben (Takemoto *et al.*, 2015). Im Rahmen dieser Arbeit konnte die genaue Position der RNase E/G-Schnittstelle in *cg0083* für den Stamm ATCC 13032 identifiziert werden. Das Kapitel 4.3.8 ab Seite 148 befasst sich mit den Schnittstellen der RNasen E/G und J in Riboswitches.

Tabelle 4.9: Rho-abhängig regulierte Riboswitches und RNA-Motive. Für jeden Riboswitch sowie RNA-Motiv bzw. RNA *leader* wurde geprüft, ob eine Verlängerung des Transkripts im Bereich des Riboswitches vorliegt und ob das jeweilige Gen in der $\Delta\rho$ -Mutante differenziell transkribiert ist. Die angegebenen Transkriptionsstarts (TSS) wurden mit Hilfe einer für die $\Delta\rho$ -Mutante angefertigten Primärtranskript-cDNA-Bibliothek identifiziert bzw. mit den bereits für *C. glutamicum* bekannten TSS aus Pfeifer-Sancar *et al.* (2013) abgeglichen. LFC, \log_2 FoldChange.

Name	Gen	TSS	Rho-abhängige Verlängerung	Differenziell in $\Delta\rho$ -Mutante
FMN-Riboswitch	<i>cg0083</i>	beobachtet ^a	ja	ja (LFC +4,0)
<i>cspA</i> mRNA 5'-UTR	<i>cg0215 / cspA</i>	beobachtet ^a	nein	ja (LFC -1,1)
TPP-Riboswitch	<i>cg0825</i>	beobachtet ^a	nein	nein
<i>ydaO-yuaA leader</i>	<i>cg0936 / rpf1</i>	beobachtet ^a	nein	nein
TPP-Riboswitch	<i>cg1227</i>	beobachtet ^a	nein	nein
mini- <i>ykkC</i> RNA-Motiv	<i>cg1231 / chaA</i>	beobachtet ^b (Pos. 1.131.045 bp)	nein	nein
TPP-Riboswitch	<i>cg1476 / thiC</i>	beobachtet ^a	nein	nein
SAM-IV-Riboswitch	<i>cg1478</i>	beobachtet ^a	ja	ja (LFC +6,6)
TPP-Riboswitch	<i>cg1655 / thiM</i>	beobachtet ^a	nein	ja (LFC +1,3)
<i>yybP-ykoY leader</i>	<i>cg1660</i>	nicht beobachtet ^c	nein	nein
<i>yybP-ykoY leader</i>	<i>cg2157 / terC</i>	beobachtet ^a	nein	nein
TPP-Riboswitch	<i>cg2236 / thiE</i>	beobachtet ^a	nein	nein
<i>mraW</i> RNA-Motiv	<i>cg2377 / mraW</i>	beobachtet ^a	nein	nein
<i>ydaO-yuaA leader</i>	<i>cg2402</i>	beobachtet ^a	nein	nein
<i>msiK</i> RNA-Motiv	<i>cg2708 / msiK1</i>	beobachtet ^a	nein	ja (LFC -1,4)
<i>yybP-ykoY leader</i>	<i>cg2785</i>	nicht beobachtet ^c	nein	nein

^a Bekannter TSS aus Pfeifer-Sancar *et al.* (2013).

^b Neuer TSS; nur in $\Delta\rho$ -Mutante detektiert.

^c Kein TSS beobachtet, weder von Pfeifer-Sancar *et al.* (2013) noch in $\Delta\rho$ -Mutante.

Die für den FMN-Riboswitch in *ribM* von Takemoto *et al.* (2015) gezeigte Regulation durch Rho und RNase E/G gab Anlass zu einer Überprüfung aller in *C. glutamicum* ATCC 13032 bekannten Riboswitches und RNA-Motive sowie RNA *leader* auf eine mögliche Regulation durch Rho. Auch mögliche Regulationen durch die beiden RNasen E/G und J hinsichtlich der Riboswitches wurden überprüft. Das Kapitel 4.3.8 auf Seite 148 beschäftigt sich mit den

4 Ergebnisse

RNase-Schnittstellen in Riboswitches im Detail. Neben der Verlängerung des *ribM* Transkripts in der $\Delta\rho$ -Mutante, konnte auch eine im Vergleich zum Wildtyp 16-fache (LFC +4,0) Transkriptionsverstärkung beobachtet werden (Tabelle 4.9). Die Read-Abdeckung von *ribM* ist im Wildtyp im Bereich des FMN-Riboswitches am höchsten und fällt weiter *downstream* in Richtung CDS stark ab (Abbildung 4.21). In der $\Delta\rho$ -Mutante ist die Read-Abdeckung über das gesamte Gen nahezu konstant hoch und geht über den annotierten Bereich der CDS hinaus. Diese Beobachtung bestätigt die Resultate von Takemoto *et al.* (2015). Davon ausgehend, dass eine Transkriptverlängerung und verstärkte Transkription in der $\Delta\rho$ -Mutante Indikatoren für eine Regulation des Riboswitches durch Rho darstellen, wurden weitere Riboswitches in *C. glutamicum* dahingehend überprüft.

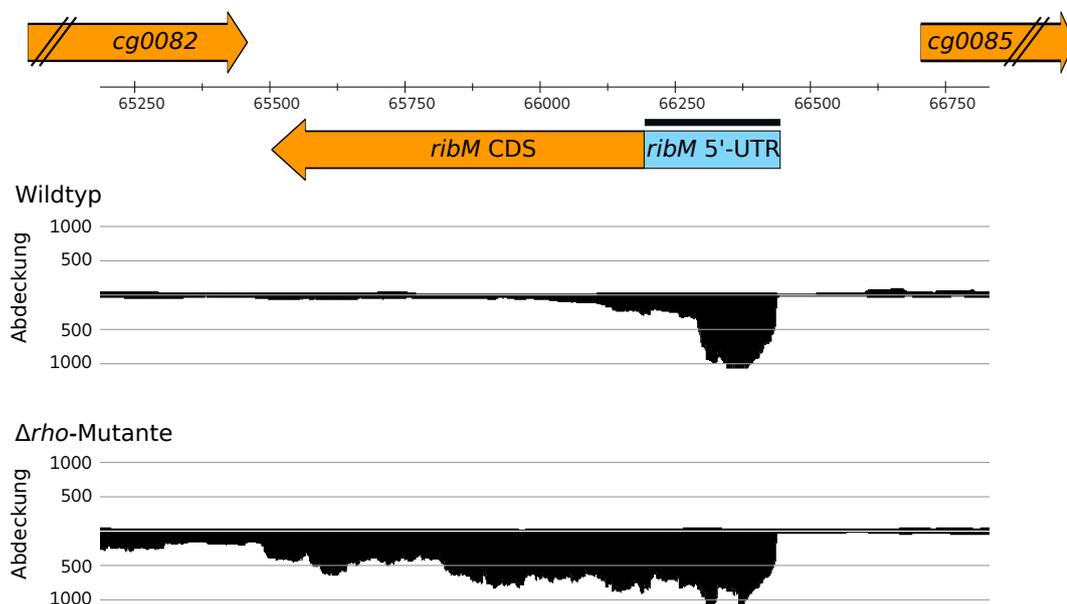


Abbildung 4.21: Read-Abdeckung im Bereich des FMN-Riboswitches von *ribM* im Wildtyp und der $\Delta\rho$ -Mutante. Der FMN-Riboswitch (schwarzer Balken) befindet sich in der 5'-UTR von *ribM* (blauer Kasten).

Der SAM-IV-Riboswitch im Gen *cg1478* ist sehr wahrscheinlich ebenfalls durch Rho reguliert, da das Gen zum Einen ein in der $\Delta\rho$ -Mutante verlängertes Transkript aufweist und zum Anderen das am stärksten differenziell transkribierte Gen in der $\Delta\rho$ -Mutante ist (LFC +6,6; entspricht einer >90-fachen Transkriptionsverstärkung).

Der TPP-Riboswitch in *thiM* ist möglicherweise ebenfalls durch den Rho-Faktor reguliert. Diese Regulation ist vermutlich indirekt, da das *thiM*-Transkript nicht verlängert, sondern nur verstärkt differenziell transkribiert ist (LFC +1,3).

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

In der Δrho -Mutante sind *msiK1*, das das *msiK* RNA-Motiv besitzt und *cspA*, welches den *cspA* leader trägt, differenziell transkribiert. Für keines dieser beiden Gene wurde eine Verlängerung am 3'-Ende festgestellt. Die Gene *msiK1* und *cspA* sind beide in der Δrho -Mutante schwächer transkribiert als im Wildtyp. Aufgrund des negativen LFC-Werts von *msiK1* und *cspA* ist eine Regulation durch Rho daher unwahrscheinlich.

Neben möglichen Regulationen von Riboswitches durch Rho, konnte ein neuer Transkriptionsstart (*Transcription Start Site*, TSS) an Genomposition 1.313.045 bp für das mini-*ykkC* RNA-Motiv in *cg1231* mit Hilfe der Primärtranskript-Sequenzierung der Δrho -Mutante identifiziert werden. Interessanterweise ist dieser TSS nicht unter der Vielzahl an von Pfeifer-Sancar *et al.* (2013) untersuchten Stressbedingungen (Hoher Sauerstoffpartialdruck, Wasserstoffperoxid-, Diamid-, Ethanol-, Salz-, Hitze- und Kältestress) beobachtet worden und nur in der Δrho -Mutante identifizierbar (Tabelle 4.9). Die Deletion von *rho* ruft somit ein von diesen Stressen unterschiedliches Transkriptionsprofil hervor.

4.2.11 Die Transkription eines Typ IVb Pilus-Gen-Clusters in *C. glutamicum* ist Rho-abhängig

Die sieben Gene *cg0362* bis *cg0369* *downstream* der 6C RNA gehören mit zu den in der Δrho -Mutante am stärksten differenziell transkribierten Genen (Tabelle 4.7). Sie kodieren für einen putativen Typ IV Pilus, im speziellen einen Typ IVb Pilus. Typ IV Pili sind weit verbreitet und haben unterschiedliche Funktionen. So nutzen ihn *Mycobacterium tuberculosis* und *Bifidobacterium breve* zur Adhäsion an eukaryotische Zellen (Ramsugit und Pillay, 2015; Motherway *et al.*, 2011), *Micrococcus luteus* als DNA-Aufnahmesystem (Angelov *et al.*, 2015) und *Clostridium perfringens* (Varga *et al.*, 2006) für die Zellmobilität. Da der Modellorganismus für diesen Pilustyp *Aggregatibacter actinomycetemcomitans* den Pilus für die Oberflächenadhäsion nutzt, wird der für den Pilus kodierende Gen-Cluster häufig als *tight adherence*-Cluster (*tad*-Cluster) bezeichnet. Im Folgenden werden die Gene *cg0362* bis *cg0369* und *cg1830* in *C. glutamicum* ebenfalls als *tad*-Cluster bezeichnet und nach den Homologen in *A. actinomycetemcomitans* benannt. Das Gen *tadV* (*cg1830*), welches für eine Prepilinpeptidase kodiert, befindet sich weit von den anderen *tad*-Genen entfernt. Dieses Gen gehört funktional ebenfalls zum Typ IVb Pilus, wird in dieser Arbeit aber aufgrund der entfernten Lage separat zum *tad*-Cluster betrachtet.

Das *tad*-Cluster ist in vielen Actinobakterien konserviert und wurde bioinformatisch unter anderem bereits für *C. glutamicum* und *C. diphtheriae* vorhergesagt (Tomich *et al.*, 2007; Imam *et al.*, 2011). Im *C. glutamicum* Wildtyp unter Standard-Laborbedingungen ist das *tad*-Cluster nicht bzw. nur sehr schwach transkribiert. Obwohl in der Studie von Pfeifer-

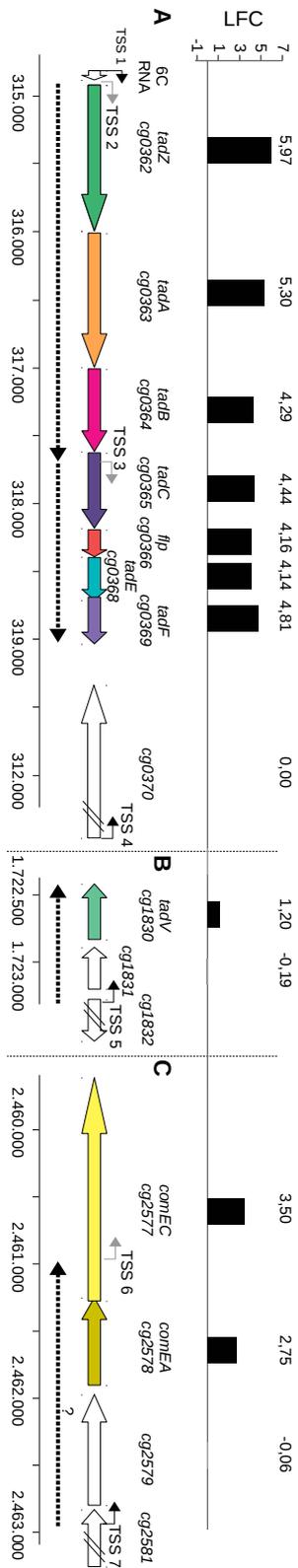


Abbildung 4.22: Grafische Darstellung des *tad*-Clusters und der *com*-Gene in *C. glutamicum* und deren differenzielle Transkription in der *Arho*-Mutante. Die Gene des Clusters wurden nach den entsprechenden Homologen in *A. actinomycetemcomitans* benannt. Die Homologe der *com*-Gene sind nach der Annotation der *com*-Gene in *B. subtilis* benannt worden. Alle Gene des *tad*-Clusters, sowie die beiden Kompetenzgene *comEA* und *comEC*, sind in der *Arho*-Mutante im Vergleich zum Wildtyp verstärkt transkribiert (schwarze Balken). (A) Der für einen Typ IVb Pilus kodierende *tad*-Cluster besteht aus den Genen *tadZABC-flp-tadEF*. Der Gen-Cluster wird *upstream* von der 6C RNA und *downstream* von dem, für eine putative RNA-Helicase kodierenden, Gen *cg0370* flankiert. Die 6C RNA wie auch *cg0370* werden jeweils monocistronisch transkribiert. Anhand der Primärtranskript-Daten der *Arho*-Mutante konnten die neuen Transkriptionsstartpunkte (graue Pfeile) TSS 2 (Position 314.802 bp), TSS 3 (Position 317.994 bp) und TSS 6 (Position 2.460.908 bp) identifiziert werden. Die Gene *tadZAB* sowie *tadC-flp-tadEF* werden jeweils in einem Operon transkribiert (gestrichelte Pfeile). (B) Das Gen der ebenfalls für den Typ IVb Pilus benötigten Prepilinpeptidase *tadV* befindet sich an einer weit entfernten genomischen Position und wird mit *cg1831* in einem Operon transkribiert. Das Gen *cg1831* kodiert für einen Transkriptionsregulator der AtrSR-Familie. (C) Die *com*-Gene, *comEA* und *comEC*, kodieren für einen putativen DNA-Rezeptor und eine putative Transmembranpore. Für *comEA* konnte in der *Arho*-Mutante ein neuer TSS (TSS 6) identifiziert werden. Das Gen *cg2579*, kodierend für ein Protein unbekannter Funktion, wird mit *comEA* in einem Operon transkribiert. Aufgrund der in diesem Bereich stark schwankenden Read-Abdeckung ist nicht eindeutig feststellbar, ob *comEC* ebenfalls Teil des Operons ist (gestrichelter Pfeil mit Fragezeichen). Die Farbkodierung der *tad*-Gene entspricht der von Abbildung 3 in Tomich *et al.* (2007). LFC, log₂FoldChange.

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

Sancar *et al.* (2013) neun unterschiedliche Stressfaktoren gepoolt wurden, um umfassend Transkriptionsstarts (TSS) in *C. glutamicum* zu identifizieren, konnte kein TSS für Gene des *tad*-Clusters gefunden werden. Bisher ist die, in dieser Arbeit beschriebene, $\Delta\rho$ -Mutante die einzige bekannte Bedingung, unter der die Transkription des *tad*-Cluster differenziell ist. So werden die Gene des *tad*-Clusters in der $\Delta\rho$ -Mutante 18 bis 63-fach (LFC 4,14 bzw. LFC 5,97) stärker als im Wildtyp transkribiert (Abbildung 4.22A). Auch das Gen der Prepilinpeptidase *tadV* (*cg1830*) ist in der $\Delta\rho$ -Mutante stärker transkribiert als im Wildtyp (Abbildung 4.22B).

Da die Read-Abdeckung vom Bereich der 6C RNA, ohne auf den Hintergrundwert abzusinken, in die Read-Abdeckung des ersten Gens des *tad*-Clusters (*tadZ*) übergeht, konnte nicht anhand der Read-Abdeckung unterschieden werden, ob es sich bei der 6C RNA in der $\Delta\rho$ -Mutante um ein verlängertes Transkript handelt (welches das ganze *tad*-Cluster umfasst) oder ob das *tad*-Cluster unabhängig der Verlängerung der 6C RNA differenziell transkribiert wurde. Wäre Ersteres der Fall, wäre die differenzielle Transkription des *tad*-Clusters in der $\Delta\rho$ -Mutante nicht auf klassische Transkriptionsregulation (z. B. Aktivierung oder Derepression) zurückzuführen. Somit wäre es nicht korrekt, von einer differenziellen Transkription des *tad*-Clusters in der $\Delta\rho$ -Mutante zu sprechen. Wäre im anderen Fall, das *tad*-Cluster unabhängig von der Verlängerung der 6C RNA differenziell transkribiert, so müsste ein TSS vor *tadZ* (*cg0362*) detektierbar sein. Um festzustellen, ob die 6C RNA in der $\Delta\rho$ -Mutante tatsächlich als verlängertes Transkript, welches das *tad*-Cluster umfasst, vorliegt, wurden die *paired-end* Reads, von denen einer der beiden Read-Partner in der 6C RNA liegt, aus allen sequenzierten Replikaten extrahiert und mit denen des Wildtyps verglichen. Wie in Abbildung 4.23C und F zu sehen, sind nur in der $\Delta\rho$ -Mutante Reads innerhalb der 6C RNA detektiert worden, dessen Read-Partner sich außerhalb der 6C RNA befinden. Im Wildtyp gibt es hingegen nur Reads, die innerhalb der 6C RNA liegen (Abbildung 4.23A). Dies beweist, dass die 6C RNA in der $\Delta\rho$ -Mutante als verlängertes Transkript vorliegt. Über diese Analyse lässt sich die Mindestlänge der 6C RNA-Verlängerung in der $\Delta\rho$ -Mutante abschätzen. Sie beträgt ca. 480 nt und ist fast fünfmal so lang wie im Wildtyp. Diese Transkriptlänge reicht nicht aus um das *tad*-Cluster von ca. 4,5 kbp Länge zu enthalten.

Mittels Northern Blot konnte im Gegensatz zur RNA-Sequenzierung und der qRT-PCR (vgl. Abschnitt 4.2.6 auf Seite 103) keine Verlängerung der 6C RNA in der $\Delta\rho$ -Mutante im Vergleich zum Wildtyp nachgewiesen werden (Abbildung 4.23G). Es wird vermutet, dass die Menge an verlängerter 6C RNA nicht ausreicht, um sie im Northern Blot nachzuweisen.

Für eine weitere Verifikation wurde aus der RNA der $\Delta\rho$ -Mutante, die für die Vollängen-transkript-Bibliothek verwendet wurde, zusätzlich eine Primärtranskript-Bibliothek erstellt und sequenziert, um die TSS des *tad*-Clusters identifizieren zu können. Für die Gene *tadZ*

4 Ergebnisse

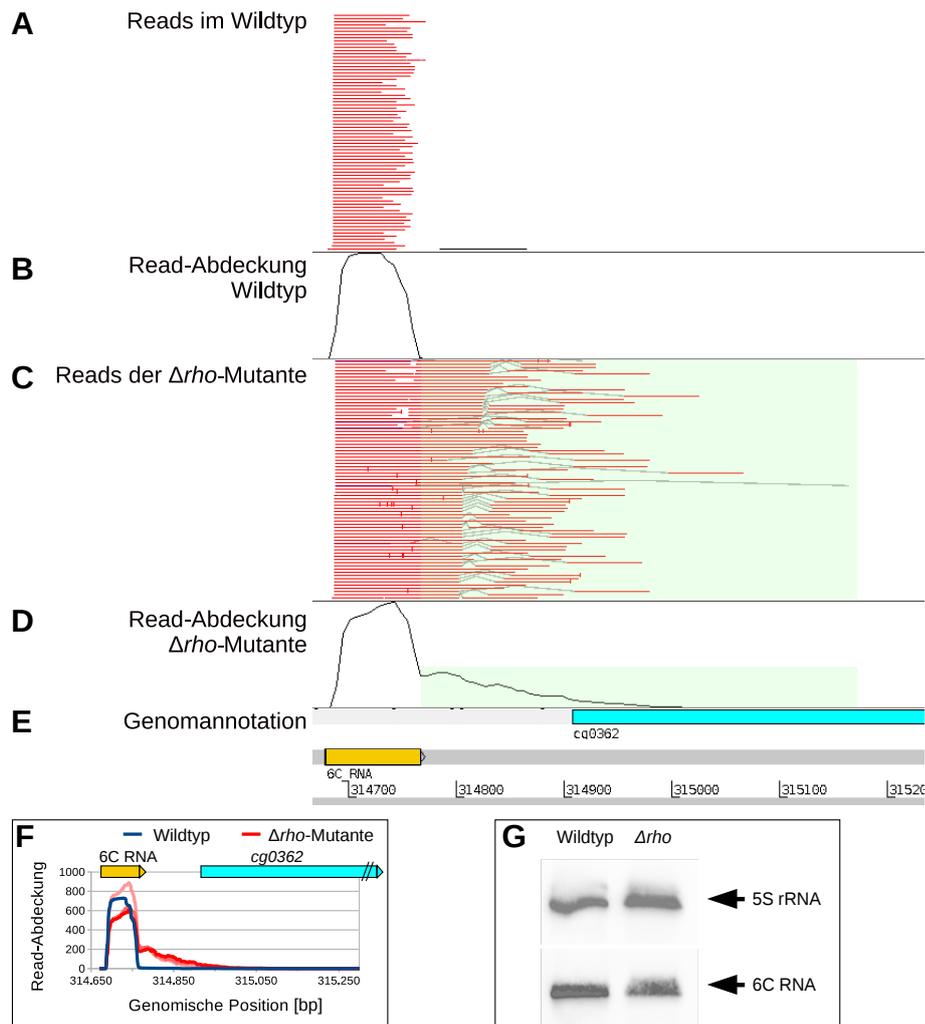


Abbildung 4.23: Verlängerung des 6C RNA-Transkripts in der $\Delta\rho$ -Mutante. Gezeigt ist die (E) Genomannotation im Bereich der 6C RNA und dem nachfolgenden Gen *cg0362*. In (A) und (C) sind nur die Reads (rote Linien) die innerhalb der 6C RNA-Region liegen bzw. von dieser Region ausgehen und deren *downstream* Read-Partner (verbunden über einen grauen Strich) für den Wildtyp und die $\Delta\rho$ -Mutante dargestellt. In (B) bzw. (D) ist die Read-Abdeckung, die sich aus den Reads aus (A) bzw. (C) ergibt, gezeigt. Die grünen Kästen in (C) und (D) markieren die Reads bzw. Read-Abdeckung, die von dem verlängerten Transkript der 6C RNA stammen. Reads die *nicht* innerhalb der 6C RNA-Region liegen bzw. aus dieser entsprungen sind, wurden in dieser Darstellung weggelassen (etwa Reads innerhalb von *cg0362*). Das Schema in Abbildung A.8 im Anhang auf Seite 233 gibt eine vereinfachte Übersicht über die hier dargestellten Reads. Während (A) bis (E) die Daten von jeweils nur einem Replikat des Wildtyps bzw. der $\Delta\rho$ -Mutante zeigen, ist in (F) die Read-Abdeckung aller drei sequenzierten Replikate für den Wildtyp (blau) und die $\Delta\rho$ -Mutante (rot) gezeigt. (G) Northern Blot der 6C RNA in der $\Delta\rho$ -Mutante und dem Wildtyp. Die Verlängerung der 6C RNA in der $\Delta\rho$ -Mutante ist im Northern Blot nicht ersichtlich. Die Verlängerung konnte aber mittels RNA-Sequenzierung (vgl. A bis F) und qRT-PCR nachgewiesen werden (vgl. Abbildung A.7). Die 5S rRNA dient als Positivkontrolle für den Northern Blot.

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

(*cg0362*) und *tadC* (*cg0365*) konnten jeweils ein TSS an Position 314.802 bp bzw. 317.994 bp identifiziert werden (Abbildung 4.22A). Das *tad*-Cluster wird in zwei Operons bestehend aus *tadZAB* und *tadC-flp-tadEF* transkribiert. Die Transkription des *tad*-Clusters ist somit unabhängig von der Transkriptverlängerung der 6C RNA.

Der TSS von *tadC* befindet sich 101 bp *downstream* von dessen Translationsstart (Abbildung 4.22A). Das nächste *in-frame* Startcodon befindet sich an Position 317.994 bp (Codon CTG). Daraus ergibt sich eine 5'-UTR von 25 nt und ein um 42 Aminosäuren verkürztes TadC-Protein.

Die Read-Abdeckung in der Δ *rho*-Mutante *downstream* vom letzten Gen des *tad*-Clusters (*tadF*) erreicht erst ca. 1.300 bp hinter dem Stopcodon von *tadF* den Wert des Hintergrundrauschens. Dies entspräche einer Transkriptverlängerung von 50 %. Es wird daher angenommen, dass *tadF* Rho-abhängig terminiert wird.

4.2.12 Rho aktiviert die Transkription von Kompetenzgenen

In der Δ *rho*-Mutante sind die beiden Gene *cg2577* und *cg2578* mit LFC-Werten von 2,75 bzw. 3,50 deutlich differenziell transkribiert. Die Gene sind als putative Pore zur DNA-Aufnahme der *Com*-Familie (*cg2577* / *comEC*) und als putativer DNA-Rezeptor (*cg2578* / *comEA*) annotiert und somit vermutlich in der Aufnahme von DNA (sogenannte „Kompetenz“) involviert. DNA-Aufnahmesysteme sind sehr komplex und bestehen neben den beiden genannten Proteinen auch aus einem Pilus, der die DNA bindet und vermutlich zur Pore leitet (Laurenceau *et al.*, 2013; Johnston *et al.*, 2014). Da das für einen putativen Typ IVb Pilus kodierende *tad*-Cluster in der Δ *rho*-Mutante ebenfalls stark differenziell transkribiert ist, könnte es sich hier um den für die DNA-Aufnahme benötigten Pilus handeln.

Für beide *com*-Gene war bisher kein Transkriptionsstartpunkt (TSS) in *C. glutamicum* bekannt. Anhand der in dieser Arbeit angefertigten Primärtranskript-Bibliothek der Δ *rho*-Mutante konnte für das Gen *comEC* ein TSS an Position 2.460.908 bp identifiziert werden (Abbildung 4.22C auf Seite 116). Der TSS besitzt ein für einen σ^A Promotor charakteristisches -10 Motiv (Sequenz TACTCT). Allerdings befindet sich dieser TSS 342 bp *downstream* des annotierten Translationsstarts (Position 2.461.250 bp). Für das nächstmögliche *in-frame* Startcodon (GTG) an Position 2.460.878 bp konnte keine RBS detektiert werden. Die 5'-UTR zum Startcodon GTG würde 25 nt betragen und ist somit zu lang für ein *leader*-freies (*leaderless*) Transkript. Das Gen *comEA* wird mit dem *upstream* liegenden Gen *cg2579*, welches für ein Protein unbekannter Funktion kodiert, in einem Operon transkribiert, da nur für *cg2579* ein TSS bekannt ist (Abbildung 4.22C).

4.2.13 Überexpression des Typ IVb Pilus-Gen-Clusters führt zu höherer Transformationsfrequenz

In der Δrho -Mutante sind neben dem für den Typ IVb Pilus kodierenden *tad*-Cluster (*cg0362* bis *cg0369*) auch Gene, die für einen putativen DNA-Rezeptor (*cg2578*) und eine putative DNA-Pore zur DNA-Aufnahme (*cg2577*) kodieren, stark differenziell transkribiert. Somit wären alle oder zumindest ein Großteil der theoretisch benötigten Komponenten für ein DNA-Aufnahmesystem zur Ausbildung von Kompetenz in der Δrho -Mutante exprimiert. Nur wenige Bakterien sind als natürlich kompetent bekannt (Johnston *et al.*, 2014). Für *C. glutamicum* ist natürliche Kompetenz nicht beschrieben. Aufgrund der in der Δrho -Mutante stark differenziell transkribierten Gene, liegt der Schluss nahe, dass die Δrho -Mutante besser zu transformieren sein könnte als der Wildtyp. Es wurde daher die Transformationsfrequenzen des Wildtyps und der Δrho -Mutante bestimmt. Außerdem wurde der Stamm *tadOE* (basierend auf dem Wildtyp) konstruiert. Der Stamm trägt das *tad*-Cluster auf dem induzierbaren Plasmid pEC-XC99E mit P_{trc} -Promotor anstelle des nativen Promotors und der nativen 5'-UTR von *cg0362*. Auch von diesem Stamm wurde die Transformationsfrequenz bestimmt. Darüber hinaus wurde mittels qRT-PCR die relative Transkriptmenge der Gene *cg0362* und *cg0365* im Wildtyp und im *tadOE*-Stamm bestimmt.

Das Wachstum des *tadOE*-Stamms bricht nach Induktion mit IPTG ein, während das Wachstum des uninduzierten *tadOE*-Stamms dem des Wildtyps ähnelt (Abbildung 4.24A). Durch die Induktion steigt die Transkriptmenge um den Faktor 100 an (Abbildung 4.24B). Die starke Überexpression des *tad*-Clusters hat vermutlich negative Auswirkungen auf die Membranintegrität der Zelle hat und somit das Zellwachstum stört. Aus diesem Grund wurde die Transformationsfrequenz nur für den uninduzierten *tadOE* Stamm bestimmt.

Die Transformationsfrequenz der Δrho -Mutante ist etwa 50 % höher als die des Wildtyps (Abbildung 4.24C). Die Transformationsfrequenz des uninduzierten *tadOE* Stamms übertrifft die der Δrho -Mutante und ist ca. 225 % höher als die des Wildtyps. Es ist sehr wahrscheinlich, dass die Funktion des vom *tad*-Cluster kodierten Typ IVb Pilus in *C. glutamicum* die DNA-Aufnahme der Zelle ist. Allerdings wächst der *tadOE* Stamm schlechter als der Wildtyp. Die Induktion der Expression des *tad*-Clusters führt zu einem Wachstumseinbruch (Abbildung 4.24A). Aufgrund des schlechteren Wachstums des induzierten *tadOE* Stamms wurden die Transformationsfrequenzen für den nicht induzierten *tadOE* Stamm bestimmt.

Mit einem Rasterkraftmikroskop wurde untersucht, ob Pilus-Strukturen auf der Zelloberfläche der Δrho -Mutante erkennbar sind. Dabei konnte kein Unterschied in der Zelloberfläche der Δrho -Mutante von der des Wildtyps beobachtet werden (Abbildung A.9). Der ebenfalls als DNA-Aufnahmesystem charakterisierte Typ IVb Pilus in *Micrococcus luteus* konnte, vermutlich

4.2 Rho-abhängige Transkriptionstermination in *Corynebacterium glutamicum*

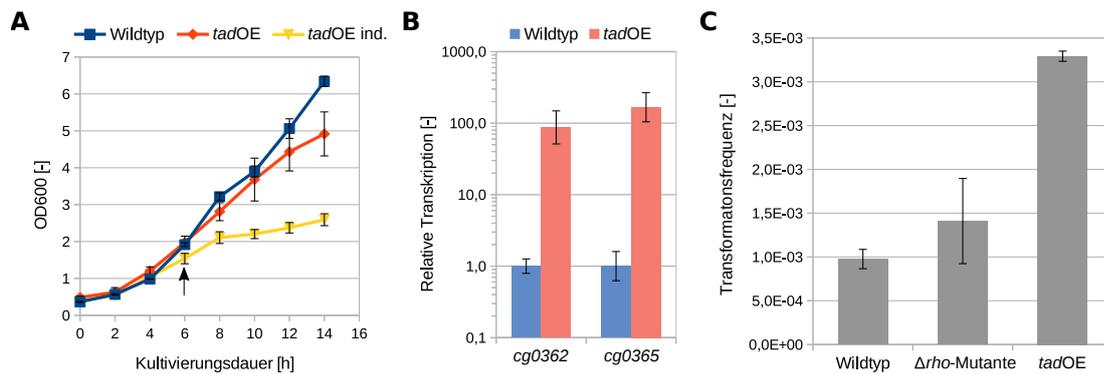


Abbildung 4.24: Charakterisierung des *tad*-Clusters in *C. glutamicum*. (A) Wachstumsvergleich des *tadOE* Stamms uninduziert (orange) und induziert (gelb) im Vergleich zum Wildtyp mit Leerplasmid (blau). Induziert wurde nach 6 Stunden mit 0,4 mM IPTG (Pfeil). (B) Relative Transkriptmenge der zwei *tad*-Cluster-Gene *cg0362* und *cg0365* nach Induktion ermittelt mit qRT-PCR. (C) Transformationsfrequenzen des uninduzierten *tadOE* Stamms im Vergleich zur $\Delta\rho$ -Mutante und dem Wildtyp.

aufgrund der geringen Größe, mikroskopisch ebenfalls nicht nachgewiesen werden (Angelov *et al.*, 2015). Es wird daher vermutet, dass der Typ IVb Pilus in *C. glutamicum* ebenfalls sehr kurz, aber dennoch funktional ist.

Über DNA-Aufnahmesysteme kann potenziell Fremd-DNA in die Zelle gelangen. Da dieses DNA-Aufnahmesystem in der $\Delta\rho$ -Mutante stark differenziell transkribiert wird, zeigt dies wiederum, dass der Rho-Faktor neben der Transkripttermination auch eine Art *Silencer* von Fremd-DNA ist. Dies passt sehr gut zu den weiter oben beschriebenen anderen, ebenfalls in der $\Delta\rho$ -Mutante differenziell transkribierten Genen aus potentiell zellfremden Quellen, wie IS-Elementen und Prophagen-Genen

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Die beiden vorherigen Kapitel des Ergebnisteils behandelten die Analyse der RNA ausgehend von den Transkriptenden. So wurden im ersten Kapitel mit Hilfe der 5'-Transkriptenden unter anderem Promotoren und 5'-UTRs in *C. diphtheriae* charakterisiert. Im zweiten Kapitel wurde auf die Transkripttermination von Rho-abhängig terminierten Transkripten in *C. glutamicum* eingegangen und diese untersucht. Es wurden die Prozesse der Transkription von der Bildung bis zur Termination behandelt. Das folgende Kapitel beschäftigt sich mit der Prozessierung bzw. dem Abbau von Transkripten durch RNasen. Hierzu wurde der Einfluss der RNasen E/G und J auf das Transkriptom von *C. glutamicum* analysiert und mit einer eigens entwickelten Methode die RNase-Schnittstellen identifiziert.

4.3.1 Vergleich von RNase E/G und RNase J mit Homologen der Modellorganismen

Die zu untersuchenden RNasen E/G und J aus *C. glutamicum* wurden bioinformatisch mit den Homologen der Modellorganismen *E. coli* bzw. *B. subtilis* sowie den Actinobakterien *S. coelicolor* und *M. tuberculosis* auf Ebene der Aminosäuresequenzen verglichen.

Die RNasen E/G besitzt einen zusätzlichen N-terminale Bereiche

Die RNase E/G aus *C. glutamicum* (Cg.RNase E/G) ist ein 1.021 Aminosäuren langes Protein. Obwohl die Längen von *E. coli* RNase E (Ec.RNase E) mit 1.061 Aminosäuren und Cg.RNase E/G mit 1.021 Aminosäuren ähnlich sind, unterscheiden sich beide Proteine in der Lokalisation der katalytischen Domäne. Vor der katalytischen Domäne von Cg.RNase E/G befindet sich ein zusätzlicher N-terminaler Bereich, den weder Ec.RNase E noch Ec.RNase G besitzen (Abbildung 4.25A). Cg.RNase E/G besitzt dafür einen im Vergleich zu Ec.RNase E kürzeren C-terminalen Bereich. Die Funktion des zusätzlichen N-terminalen Bereichs ist unbekannt, da keine bekannten Domänen für diesen Bereich identifiziert wurden. Der zusätzliche N-terminale Bereich ist auch in den Homologen der Actinobakterien *S. coelicolor* (Sc.RNase ES) und *M. tuberculosis* (Mt.RNase E/G) vorhanden und dort noch länger als in *C. glutamicum*. Aufgrund des zusätzlichen N-terminalen Bereichs, ist die Ähnlichkeit von Cg.RNase E/G zu Ec.RNase G und Ec.RNase E mit 31 bis 33 % gering. Die höchste Ähnlichkeit besitzt Cg.RNase E/G zu Mt.RNase E/G mit 47 % (Tabelle 4.10).

Für RNase E und G aus *E. coli* wurden für die Funktion wichtige Motive (hier als M1 bis M5 bezeichnet) in der katalytischen S1-Domäne identifiziert (Schubert *et al.*, 2004; Callaghan

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

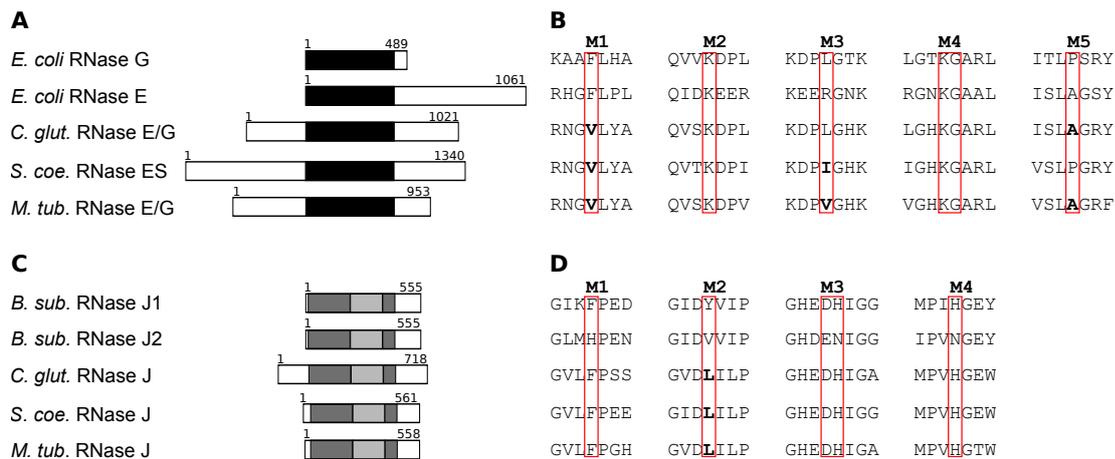


Abbildung 4.25: Vergleich der RNase E bzw. G und J-Homologe der Modellorganismen mit *C. glutamicum*, *S. coelicolor* und *M. tuberculosis* sowie deren charakteristischen Motiven. Längenvergleich von RNase E bzw. G (A) und J-Homologen (C); Vergleich der charakteristischen Motive für RNase E bzw. G (B) und J (D). Wichtige Domänen sind in schwarz (katalytische S1-Domäne), grau (Metallo-β-Lactamase-Domäne) und hellgrau (β-CASP-Domäne) eingezeichnet. Wichtige Aminosäure-Reste in den charakteristischen Motiven (M1 - M5 bzw. M1 bis M4) sind in rot markiert und in fett gedruckt, wenn sie von den RNase G bzw. RNase J1 Homologen der Modellorganismen abweichen.

et al., 2005). Die Motive M2, M3 und M4 von Cg.RNase E/G sind identisch zu Ec.RNase G (Abbildung 4.25B). Anstelle eines Phenylalanins (F) im Motiv M1 besitzt Cg.RNase E/G ein Valin (V). Dieser Austausch befindet sich ebenfalls in den RNase E-Homologen von *S. coelicolor* und *M. tuberculosis*. Da sowohl Phenylalanin als auch Valin unpolare Aminosäuren sind, wurde davon ausgegangen, dass dieser Aminosäureaustausch keinen wesentlichen Effekt auf die Eigenschaften der RNasen der gezeigten Actinobakterien hat. Motiv M5 von Cg.RNase E/G ist identisch zu dem aus Ec.RNase E. Cg.RNase E/G besitzt somit Motive die sowohl identisch zu Ec.RNase G (Motiv M1) als auch zu Ec.RNase E (Motiv M5) sind.

Vergleich der RNase J mit bekannten Homologen

C. glutamicum besitzt im Gegensatz zu *B. subtilis*, in dem RNase J zuerst beschrieben wurde, nur ein RNase J-Homolog. Im Allgemeinen wird RNase J in drei Domänen unterteilt: einer β-Lactamase-, einer β-CASP-Domäne und einer C-terminalen Domäne (Li de la Sierra-Gallay et al., 2008). Cg.RNase J besitzt einen im Vergleich zu *B. subtilis* RNase J1 bzw. J2 (Bs.RNase J1 bzw. J2) um 136 Aminosäuren verlängerten N-terminalen Bereich (Abbildung 4.25C). Cg.RNase J ist somit ca. 22 % länger als die Homologe in den nahen Verwandten *M. tuberculosis* und *S. coelicolor*. Für den in *C. glutamicum* verlängerten N-terminalen Bereich

4 Ergebnisse

Tabelle 4.10: Globales Alignment der RNase E bzw. E/G und RNase J-Homologe der Modellorganismen mit denen von *C. glutamicum*, *S. coelicolor* und *M. tuberculosis*. Das Alignment wurde mittels Needleman-Wunsch-Algorithmus (Needleman und Wunsch, 1970) durchgeführt. Angegeben ist die Sequenzidentität und -ähnlichkeit in Prozent. Ec, *E. coli*; Cg, *C. glutamicum*; Sc, *S. coelicolor*; Mt, *M. tuberculosis*; Bs, *B. subtilis*.

RNase E/G					
Homolog	Ec. RNase G	Ec. RNase E	Cg. RNase E/G	Sc. RNase ES	Mt. RNase E
Ec.RNase G	100 / 100	16 / 24	25 / 31	13 / 20	18 / 27
Ec.RNase E	–	100 / 100	25 / 33	17 / 25	15 / 25
Cg.RNase E/G	–	–	100 / 100	36 / 44	39 / 47
Sc.RNase ES	–	–	–	100 / 100	32 / 41
Mt.RNase E	–	–	–	–	100 / 100
RNase J					
Homolog	Bs. RNase J1	Bs. RNase J2	Cg. RNase J	Sc. RNase J	Mt. RNase J
Bs.RNase J1	100 / 100	49 / 70	28 / 44	37 / 57	35 / 54
Bs.RNase J2	–	100 / 100	26 / 47	38 / 59	34 / 58
Cg.RNase J	–	–	100 / 100	41 / 55	44 / 57
Sc.RNase J	–	–	–	100 / 100	55 / 71
Mt.RNase J	–	–	–	–	100 / 100

von RNase J wurden keine bekannten Motive oder Domänen identifiziert. Eine Aussage über die Funktion des zusätzlichen N-terminalen Bereichs ist somit nicht möglich.

Die größte Ähnlichkeit von 57 % besitzt Cg.RNase J mit dem Homolog in *M. tuberculosis*. Die geringste Ähnlichkeit weist Cg.RNase J zu Bs.RNase J2 auf (Tabelle 4.10).

Beim Vergleich der für RNase J charakteristischen Motive zeigte sich, dass Cg.RNase J eine höhere Identität zu Bs.RNase J1 als zu Bs.RNase J2 besitzt. Nur das Motiv M2 von Cg.RNase J besitzt eine von Bs.RNase J1 unterschiedliche Aminosäure. In dem Motiv M2 ist die polare Aminosäure Tyrosin (Y) gegen die unpolare Aminosäure Leucin (L) ausgetauscht (Abbildung 4.25D). Dieser Austausch wurde auch in den RNase J-Homologen von *M. tuberculosis* und *S. coelicolor* beobachtet und könnte die Aktivität oder Spezifität dieser RNase J-Homologe beeinflussen.

4.3.2 Deletion der für die RNasen E/G und J kodierenden Gene und Analyse des Wachstums

Um Rückschlüsse auf den Einfluss der RNasen E/G und J in *C. glutamicum* zu erhalten, wurden die beiden entsprechenden Gene *cg2597 / rne* und *cg2160 / rnj* mittels des pK18*mobsacB*-Systems (Schäfer *et al.*, 1994) deletiert. Das für RNase E/G kodierende Gen *rne* befindet sich als erstes Gen in einem Operon mit den Genen *rplU* und *rpmA* (Pfeifer-Sancar *et al.*,

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

2013), welche für die ribosomalen Proteine L21 bzw. L27 kodieren. Um die Transkription dieser wichtigen Gene nicht zu beeinflussen, wurde ein Abschnitt innerhalb der CDS von *rne* deletiert. Auch das für die RNase J kodierende Gen *rnj* befindet sich in einem Operon mit den Genen *dapA*, *thyX* und *dapB* (Pfeifer-Sancar *et al.*, 2013). Während die Gene *dapA* und *dapB* in der Lysin-Biosynthese involviert sind, ist *thyX* für die Umwandlung von Uridin zu Thymin zuständig. Um auch hier die Transkription der restlichen Gene im Operon nicht durch die Deletion von *rnj* zu beeinflussen, wurde nur ein Teil innerhalb der CDS von *rnj* deletiert.

Nach erfolgter Deletion der RNasen E/G bzw. J wurde das Wachstum der Deletionsmutanten analysiert. Dazu wurde die Δrne -Mutante sowie der Wildtyp in Erlenmeyerkolben mit CGXII-Minimalmedium in drei biologischen Replikaten kultiviert und die optische Dichte (OD600) alle zwei Stunden gemessen. Wie in Abbildung 4.26A zu sehen, ist das Wachstum der Δrne -Mutante identisch zu dem des Wildtyps. Die Deletion der RNase E/G hat somit keinen Einfluss auf das Zellwachstum von *C. glutamicum*.

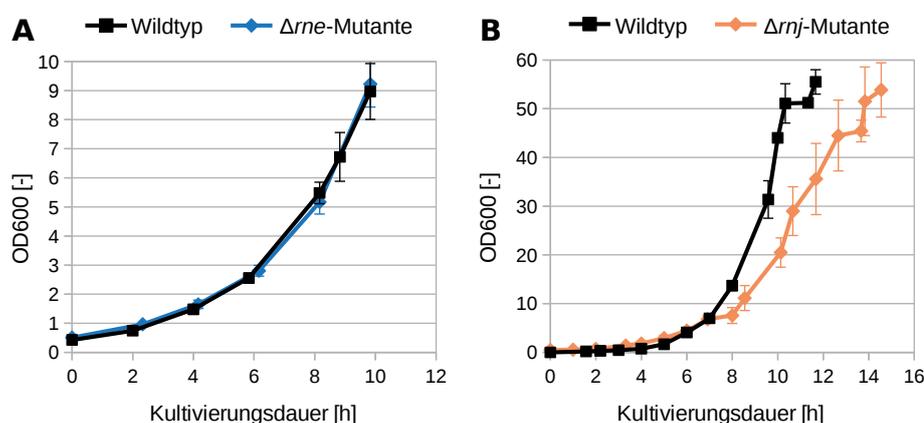


Abbildung 4.26: Vergleich des Wachstums der (A) Δrne -Mutante und (B) Δrnj -Mutante mit dem Wildtyp. Die Kultivierung der Δrne -Mutante wurde in Erlenmeyerkolben mit CGXII-Minimalmedium durchgeführt. Die Δrnj -Mutante wurde in Komplexmedium in einem Fermenter kultiviert. Dargestellt sind die Mittelwerte von drei biologischen Replikaten. Der Anstieg der optischen Dichte bei 12 h bzw. 14 h ist auf eine stark erhöhte Rührerdrehzahl zurückzuführen. Die Fehlerbalken entsprechen der Standardabweichung.

Die Deletion des für die RNase J kodierenden Gens *rnj* hat dagegen einen starken Einfluss auf das Wachstum. Für die Δrnj -Mutante wurde in CASO-Komplexmedium im Fermenter ein im Vergleich zum Wildtyp deutlich verlangsamtes Wachstum gemessen (Abbildung 4.26B). Die Δrnj -Mutante wurde in einem Fermenter mit Regelung des pH-Werts und der Sauerstoffsättigung kultiviert, da das Wachstum in Minimalmedium unter weniger geregelten Bedingungen im Erlenmeyerkolben noch stärker verlangsamt ist. Die maximale spezifische Wachstums-

4 Ergebnisse

geschwindigkeit μ_{\max} der Δrnj -Mutante betrug $0,349 \text{ h}^{-1}$, während die des Wildtyps unter identischen Bedingungen $0,614 \text{ h}^{-1}$ betrug.

Die Δrnj -Mutante wurde mit den Genen für RNase J1 (Gen *rnjA*) und J2 (Gen *rnjB*) aus dem Modellorganismus *B. subtilis* komplementiert. Dazu wurde das jeweilige Gen aus *B. subtilis* in das Plasmid pEC-XC99E kloniert und in die *C. glutamicum* Δrnj -Mutante transformiert. Obwohl die Ähnlichkeit von *C. glutamicum* RNase J zu *B. subtilis* RNase J1 höher ist als zu RNase J2 (siehe Abbildung 4.25), wurde das Wachstumsdefizit nur mit RNase J2 vollständig komplementiert (Abbildung A.10 im Anhang auf Seite 234).

Neben der Erstellung der RNase-Einzeldelentionsmutanten wurde ebenfalls versucht eine Doppeldelentionsmutante $\Delta rne \Delta rnj$ zu generieren. Dazu wurde das Deletionsplasmid pK18*mobsacB::* Δrne in kompetente Zellen der Δrnj -Mutante transformiert. Auch nach mehrmaligen Versuchen wurde das Gen *rne* nicht in der Δrnj -Mutante deletiert. Es wurde daher davon ausgegangen, dass zwar die Einzeldelentionen von RNase E/G und J für *C. glutamicum* kompensierbar sind, eine Doppeldelention beider RNasen allerdings letal ist.

4.3.3 Die Deletionen der für die RNasen E/G und J kodierenden Gene haben einen starken Einfluss auf das Transkriptom

Probengenerierung und RNA-Sequenzierung der Δrne -Mutante und Δrnj -Mutante

Die Deletionen der für die RNasen E/G und J kodierenden Gene sollten sich auf das Transkriptom der Zelle auswirken. Um eine umfassende Analyse des Transkriptoms durchführen zu können, wurde auf die Methode der RNA-Sequenzierung zurückgegriffen.

Für die Probengewinnung wurden die Δrne -Mutante mit dem Wildtyp in jeweils drei biologischen Replikaten in Minimalmedium mit Glucose in Schüttelkolben kultiviert und die Zellen in der exponentiellen Wachstumsphase (OD600 ca. 8) geerntet (Abbildung 4.26A). Die aus den Zellen isolierte RNA wurde, nach Abreicherung der rRNA, über das Protokoll *TruSeq stranded mRNA* (Illumina) in strangspezifische Vollängen-cDNA-Bibliotheken umgewandelt und diese anschließend auf dem Illumina MiSeq sequenziert. Die erhaltenen *paired-end* Reads hatten eine Read-Länge von 75 bp und wurden mit dem Programm *bowtie2* (Langmead und Salzberg, 2012) gegen das *C. glutamicum* ATCC 13032 Genom (Kalinowski *et al.*, 2003) gemappt (Tabelle 4.11).

Für die RNA-Sequenzierung der Δrnj -Mutante sowie des Wildtyps wurden diese in je drei biologischen Replikaten in Komplexmedium (CASO) mit Glucose in einem Fermenter kultiviert (Abbildung 4.26B). Die Zellernte erfolgte in der exponentiellen Wachstumsphase. Um den Wildtyp und die Δrnj -Mutante zu einer identischen Wachstumsphase zu ernten, wurde die langsamer wachsende Δrnj -Mutante nicht wie der Wildtyp nach 7 h, sondern

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Tabelle 4.11: Sequenzierte und gemappte Reads der cDNA-Bibliotheken von der Δrne -Mutante und der Δrnj -Mutante und den dazugehörigen Wildtypen.

RNase E/G Deletionsmutante			
cDNA-Bibliothek	Sequenzierte Reads	Gemappte Reads	Gemappte Reads in %
Wildtyp Replik 1	4.335.740	4.317.530	99,58
Wildtyp Replik 2	3.404.644	3.395.451	99,73
Wildtyp Replik 3	3.414.875	3.487.224	99,59
Δrne -Mutante Replik 1	3.358.088	3.339.619	99,45
Δrne -Mutante Replik 2	4.158.551	4.109.064	98,81
Δrne -Mutante Replik 3	3.501.580	3.465.164	98,96
RNase J Deletionsmutante			
cDNA-Bibliothek	Sequenzierte Reads	Gemappte Reads	Gemappte Reads in %
Wildtyp Replik 1	13.889.743	13.267.483	95,52
Wildtyp Replik 2	17.823.532	15.203.473	85,30
Wildtyp Replik 3	10.352.101	8.938.004	86,34
Δrnj -Mutante Replik 1	4.237.144	4.202.823	99,19
Δrnj -Mutante Replik 2	4.183.619	4.163.119	99,51
Δrnj -Mutante Replik 3	3.659.388	3.617.671	98,86

erst nach 8 h geerntet. Zu diesen Zeitpunkten wiesen Wildtyp und Deletionsmutante eine OD600 von ca. 9 auf. Anschließend wurde die RNA isoliert, die rRNA abgereichert und Vollängen-cDNA-Bibliotheken mit dem Protokoll *TruSeq stranded mRNA* (Illumina) erstellt. Aus organisatorischen Gründen wurden die cDNA-Bibliotheken der Δrnj -Mutante auf einem Illumina MiSeq und die cDNA-Bibliotheken des Wildtyps auf einem Illumina HiSeq 1500 sequenziert. Da der HiSeq im Vergleich zum MiSeq eine größere Kapazität besitzt, wurden die cDNA-Bibliotheken des Wildtyps tiefer sequenziert. Die nach der Sequenzierung erhaltenen *paired-end* Reads wurden mit *bowtie2* (Langmead und Salzberg, 2012) gegen das *C. glutamicum* ATCC 13032 Genom (Kalinowski *et al.*, 2003) gemappt (Tabelle 4.11).

Analyse der differentiellen Transkription der Δrne -Mutante

Die sequenzierten Reads der cDNA-Bibliotheken der Δrne -Mutante sowie die der entsprechenden Wildtypen wurden mittels *DESeq2* (Love *et al.*, 2014) in Standardeinstellungen normalisiert und auf differentiell transkribierte Gene analysiert. Außerdem wurden die Replikate mittels hierarchischem *Clustering* auf Ähnlichkeit überprüft, um mögliche Ausreißer zu identifizieren (Abbildung A.3 im Anhang). Gene die eine Änderung in der Transkription ($\log_2\text{FoldChange}$, LFC) $\geq +1,0$ bzw. $\leq -1,0$ und einen angepassten p-Wert $< 0,05$ besitzen, wurden als differentiell transkribiert klassifiziert.

4 Ergebnisse

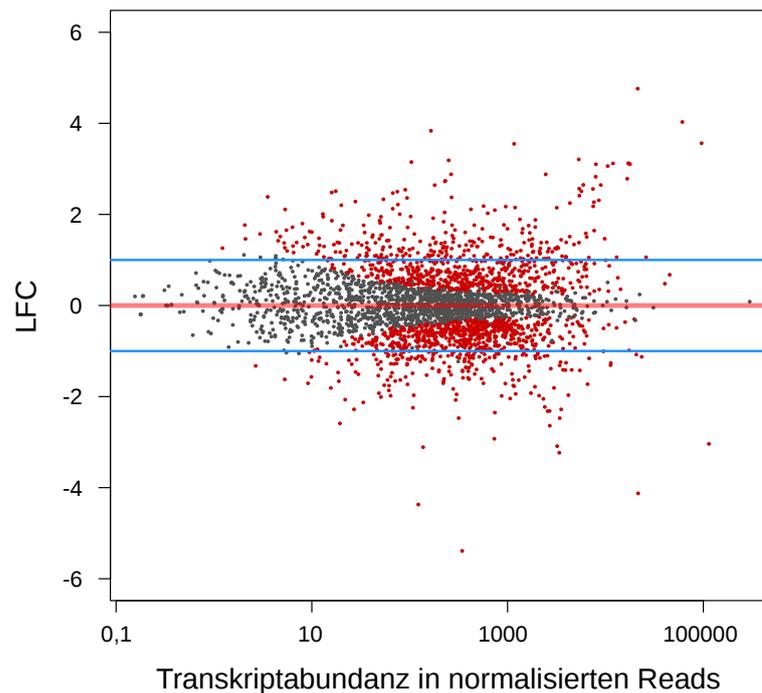


Abbildung 4.27: M/A-Plot der *C. glutamicum* Δrne -Mutante. Aufgetragen sind der \log_2 FoldChange (LFC) gegen die Transkriptabundanz in normalisierten Reads. Jedes Gen ist durch einen Punkt dargestellt. Gene die die statistischen Bedingungen erfüllen (angepasster p-Wert < 0,05) sind rot markiert. Die blauen Linien markieren einen LFC von +1,5 bzw. -1,5. Gene die rot markiert sind und über der oberen bzw. unterhalb der unteren blauen Linie liegen wurden als differenziell transkribiert gewertet.

In der Δrne -Mutante wurden insgesamt 176 Gene als differenziell transkribiert klassifiziert. Von diesen waren 104 Gene stärker und 72 Gene schwächer transkribiert als im Wildtyp (Abbildung 4.27). Im Vergleich zu den anderen im Rahmen dieser Arbeit untersuchten Deletionsmutanten ist die Zahl der differenziell transkribierten Gene gering. Eine Auswahl der in der Δrne -Mutante am stärksten und am schwächsten differenziell transkribierten Gene ist in Tabelle 4.12 aufgeführt. Eine vollständige Liste der differenziell transkribierten Gene befindet sich im Anhang (Tabelle S11).

Tabelle 4.12: Liste der am stärksten bzw. schwächsten differenziell transkribierten Gene in der Δrne -Mutante im Vergleich zum Wildtyp (Ausschnitt). Die vollständige Liste aller differenziell transkribierten Gene befindet sich im Anhang. LFC, \log_2 FoldChange.

Gen	Name	Genprodukt ¹	LFC
cg2253	<i>rpsP</i>	30S ribosomal protein S16	4,76
cg2222	<i>rpsB</i>	30S ribosomal protein S2	4,03
cg2557		Putative secondary Na ⁺ /bile acid symporter	3,84

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Tabelle 4.12: Fortgesetzt

Gen	Name	Genprodukt ¹	LFC
<i>cg1531</i>	<i>rpsA</i>	30S ribosomal protein S1	3,56
<i>cg0436</i>		Hypothetical protein	3,55
<i>cg0632</i>	<i>rpmD</i>	50S ribosomal protein L30	3,21
<i>cg1380</i>	<i>ssuA</i>	ABC-type aliphatic sulfonate transporter, substrate-binding lipoprotein	3,19
<i>cg1377</i>	<i>ssuC</i>	ABC-type aliphatic sulfonate transporter, permease subunit	3,15
<i>cg0631</i>	<i>rpsE</i>	30S ribosomal protein S5	3,12
<i>cg0936</i>	<i>rpf1</i>	RPF-protein precursor	3,12
<i>cg0629</i>	<i>rplF</i>	50S ribosomal protein L6	3,12
<i>cg0628</i>	<i>rpsH</i>	30S ribosomal protein S8	3,10
<i>cg0630</i>	<i>rplR</i>	50S ribosomal protein L18	3,10
<i>cg0634</i>	<i>rplO</i>	50S ribosomal protein L15	3,06
<i>cg1147</i>	<i>ssuI</i>	NAD(P)H-dependent FMN reductase	2,88
<i>cg2789</i>	<i>mrX2</i>	mycoredoxin 2	2,88
<i>cg0656</i>	<i>rplQ</i>	50S ribosomal protein L17	2,83
<i>cg0655</i>	<i>rpoA</i>	DNA-directed RNA polymerase, alpha subunit	2,78
<i>cg1376</i>	<i>ssuD1</i>	FMNH ₂ -dependent aliphatic sulfonate monooxygenase	2,74
<i>cg1118</i>		Pyrimidine reductase, riboflavin biosynthesis	2,72
<i>cg0653</i>	<i>rpsK</i>	30S ribosomal protein S11	2,65
<i>cg0654</i>	<i>rpsD</i>	30S ribosomal protein S4	2,65
<i>cg1156</i>	<i>ssuD2</i>	FMNH ₂ -dependent aliphatic sulfonate monooxygenase	2,64
<i>cg0563</i>	<i>rplK</i>	50S ribosomal protein L11	2,57
<i>cg0652</i>	<i>rpsM</i>	30S ribosomal protein S13	2,56
<i>cg1151</i>	<i>seuA</i>	FMNH ₂ -dependent monooxygenase, involved in sulfonate ester degradation	2,54
<i>cg0611</i>		Putative secreted protein	2,51
<i>cg0564</i>	<i>rplA</i>	50S ribosomal protein L1	2,51
<i>cg1379</i>	<i>ssuB</i>	ABC-type aliphatic sulfonate transporter, ATPase subunit	2,50
<i>cg0073</i>		RHOD domain-containing protein	2,48
<i>cg1227</i>		Putative membrane protein	-1,97
<i>cg1374</i>		Conserved hypothetical protein	-1,97
<i>cg1765</i>	<i>sufR</i>	Transcriptional repressor of suf operon	-1,98
<i>cg1785</i>	<i>amt</i>	Putative secondary ammonium transporter, Amt-family	-1,98
<i>cg1585</i>	<i>argR</i>	Transcriptional repressor of arginine biosynthesis, ArgR-family	-1,98
<i>cg1296</i>		Conserved hypothetical protein	-2,01
<i>cg1665</i>		Putative secreted protein	-2,04
<i>cg0825</i>		Putative short chain dehydrogenase	-2,05
<i>cg1061</i>	<i>urtA</i>	ABC-type putative branched-chain amino acid transporter	-2,07
<i>cg2962</i>		Uncharacterized enzyme, biosynthesis of extracellular polysaccharides	-2,07
<i>cg1783</i>	<i>soxA'</i>	Putative oxidase, pseudogene (N-terminal fragment)	-2,13
<i>cg1584</i>	<i>argF</i>	Ornithine carbamoyltransferase	-2,23
<i>cg0175</i>		Putative secreted protein	-2,25
<i>cg0160</i>		Hypothetical protein	-2,28
<i>cg1583</i>	<i>argD</i>	Acetylornithine aminotransferase, AT class II	-2,28
<i>cg1580</i>	<i>argC</i>	N-acetyl-gamma-glutamyl-phosphate reductase	-2,32
<i>cg1582</i>	<i>argB</i>	Acetylglutamate kinase	-2,32
<i>cg2236</i>	<i>thiE</i>	Thiamine-phosphate diphosphorylase	-2,35
<i>cg3374</i>		Putative NADH-dependent flavin oxidoreductase	-2,47
<i>cg1581</i>	<i>argJ</i>	Glutamate N-acetyltransferase	-2,47
<i>cg1784</i>	<i>ocd</i>	Putative ornithine cyclodeaminase	-2,59

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

4 Ergebnisse

Tabelle 4.12: Fortgesetzt

Gen	Name	Genprodukt ¹	LFC
<i>cg1930</i>		Putative secreted hydrolase	-2,64
<i>cg0120</i>		Putative hydrolase	-2,93
<i>cg4019</i>		Hypothetical protein	-3,04
<i>cg1478</i>		Hypothetical protein	-3,09
<i>cg0111</i>		Hypothetical protein	-3,11
<i>cg3372</i>		Conserved hypothetical protein	-3,23
<i>cg1476</i>	<i>thiC</i>	Thiamine biosynthesis protein ThiC	-4,13
<i>cg0701</i>		Putative secondary drug/metabolite transporter, DMT superfamily	-4,37
<i>cg0314</i>	<i>brnF</i>	Secondary branched-chain amino acid efflux transporter, large subunit	-5,39

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

Unter den stärker transkribierten Genen befanden sich auffällig viele die für ribosomale Proteine kodieren. Dies spiegelt sich auch in der Unterteilung der differenziell transkribierten Gene in COG-Klassen (Tatusov *et al.*, 1997) wider. Die meisten stärker transkribierten Gene (34) wurden der COG-Klasse für Translation (Klasse J) zugeordnet (Abbildung 4.28). Möglicherweise ist die Funktion von RNase E/G in *C. glutamicum* die Prozessierung der Transkripte dieser Gene, so dass die Transkriptabundanz dieser Gene in der Δrne -Mutante erhöht ist. Es ist interessant, dass ein in der Transkription involvierte RNase spezifisch die Transkription von in der Translation involvierten Genen regulieren könnte. Anhand der differenziellen Transkriptionsanalyse kann allerdings nicht zwischen einem direkten oder indirekten Effekt der RNase-Deletion auf die Transkriptabundanz der für die ribosomalen Proteine kodierenden Gene unterschieden werden. Weitere, der als stärker transkribiert gemessenen Gene, sind am anorganischen Ionentransport und -stoffwechsel (COG-Klasse P) beteiligt. Unter den schwächer transkribierten Genen der Δrne -Mutante wurden vor allem Gene die im Aminosäure-Stoffwechsel und -transport (COG-Klasse E) involviert sind beobachtet. So war der überwiegende Teil der in der Argininbiosynthese involvierten Gene schwächer transkribiert, obwohl der transkriptionale Repressor *argR* ebenfalls schwächer transkribiert war (Tabelle 4.12). Würde RNase E/G die Transkripte von *argR* und den anderen *arg*-Genen im Wildtyp abbauen, so müssten diese Transkripte in der Δrne -Mutante stärker und nicht schwächer transkribiert sein als im Wildtyp. Daher handelt es sich in diesem Fall vermutlich um einen indirekten Effekt der die Transkriptabundanz dieser Gene verringert. Eine genauere Unterscheidung zwischen direkten und indirekten Effekten der Deletion von *rne* ist allerdings nicht möglich.

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

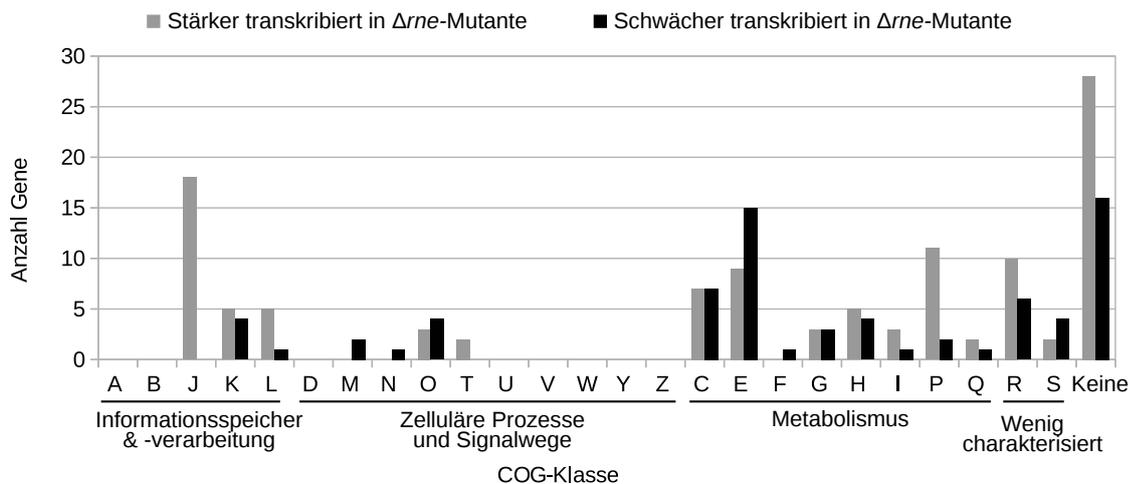


Abbildung 4.28: Klassifizierung der in der Δrne -Mutante differenziell transkribierten Gene nach COG-Klasse. Eine detaillierte Aufschlüsselung der COG-Klassen mit Beschreibung befindet sich in Tabelle A.1 auf Seite 236 im Anhang. Die Gruppe „Keine“ enthält alle Gene die keiner COG-Klasse zugeordnet sind.

4.3.4 Prozessierung des 5S rRNA Vorläufers durch RNase E/G

Eine für *C. glutamicum* RNase E/G bekannte Funktion ist die Prozessierung der 5S rRNA *in vivo* (Maeda und Wachi, 2012b). Im Rahmen dieser Arbeit wurde daher die ribosomale RNA mittels RNA-Sequenzierung auf charakteristische Unterschiede zwischen Δrne -Mutante und Wildtyp untersucht. Die oben genannten cDNA-Bibliotheken sind für diesen Zweck ungeeignet, da die rRNA vor Erstellung der Bibliotheken qualitativ abgereichert wurde. Es wurden daher jeweils eine Vollängen-cDNA-Bibliothek ohne vorherige rRNA-Abreicherung der RNA vom Wildtyp und der Δrne -Mutante erstellt. Die bei der Sequenzierung auf einem Illumina MiSeq erhaltenen *paired-end* Reads wurden wie die Reads der vorherigen Sequenzierungen gegen das *C. glutamicum* ATCC 13032 Genom gemappt. Die RNA-Sequenzierung des Wildtyps ergab 199.853 *paired-end* Reads. Für die Sequenzierung der Δrne -Mutante wurden 332.028 *paired-end* Reads sequenziert und jeweils zu 99,83 % auf das *C. glutamicum* ATCC 13032 Genom gemappt. Anschließend wurden die rRNA-kodierenden rDNA-Operons auf Unterschiede hinsichtlich der Read-Abdeckung zwischen Wildtyp und Mutante untersucht. Wie in Abbildung 4.29A beispielhaft für das rDNA-Operon Kopie A gezeigt, ist die Read-Abdeckung *upstream* und *downstream* der 5S rRNA in der Δrne -Mutante deutlich erkennbar. Im Wildtyp hingegen ist die Read-Abdeckung nur auf den annotierten Bereich der 5S rRNA beschränkt. Die von Maeda und Wachi (2012b) mit der Methode Primer-Extension identifizierte Bereich der RNase E/G-Schnittstelle befindet sich wenige Basen *upstream* der 5S rRNA und liegt in etwa

4 Ergebnisse

an der selben Position wie die erste Base mit hoher Read-Abdeckung im Wildtyp der 5S rRNA (Pfeil in Abbildung 4.29A). Dies bedeutet, dass die 5S rRNA in der Δrne -Mutante nicht wie im Wildtyp prozessiert wurde und eine längere Vorläufer-rRNA vorliegt. Somit konnte die Prozessierung des 5S rRNA-Vorläufers mittels RNase E/G auch über RNA-Sequenzierung gezeigt werden. Mittels Northern Blot wurden neben der 5S rRNA drei Vorläufer-rRNAs mit Längen von ca. 150 nt, ca. 200 nt und ca. 290 nt identifiziert (Abbildung 4.29B). Die fertige 5S rRNA wurde ebenfalls nachgewiesen. Es muss demnach eine weitere RNase geben, die die Prozessierung zur endgültigen 5S rRNA übernehmen kann.

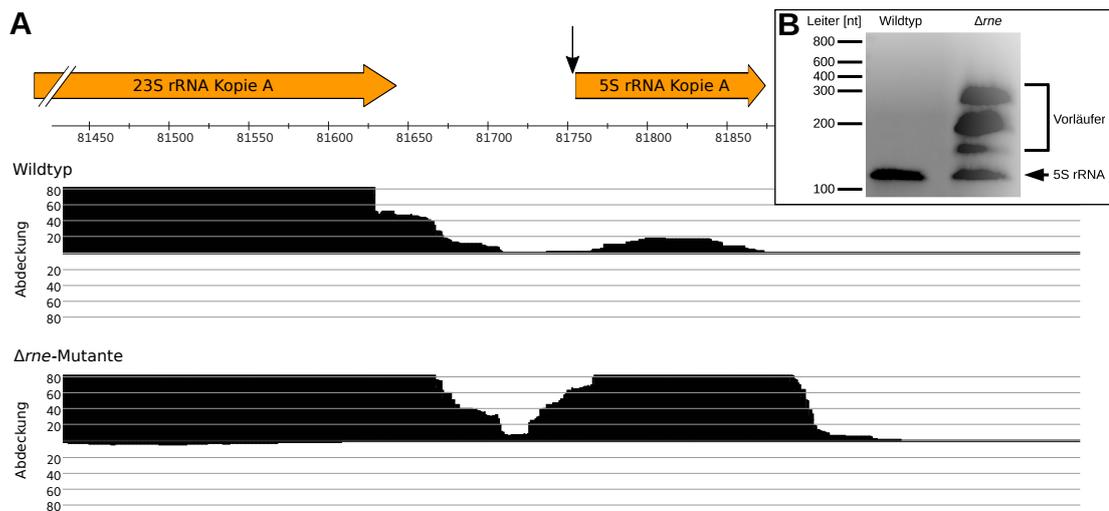


Abbildung 4.29: Unprozessierte Vorläufer-Transkripte der 5S rRNA konnten sowohl in den (A) RNA-Sequenzierungsdaten der nicht rRNA-abgereicherten cDNA-Bibliotheken, als auch mittels (B) Northern Blot für die Δrne -Mutante nachgewiesen werden. *C. glutamicum* besitzt insgesamt sechs rDNA-Operons (Kopien A bis F) mit identischen 5S rRNA-Sequenzen. Abbildungsteil (A) zeigt die Read-Abdeckung beispielhaft für die 5S rRNA des rDNA-Operons Kopie A. Der Pfeil markiert die von Maeda und Wachi (2012b) beschriebene RNase E/G Schnittstelle *upstream* der 5S rRNA.

Mittels Northern Blot der 5S rRNA wurde das Ergebnis der RNA-Sequenzierung bestätigt. Aufgrund der RNA-Fragmentierung zur Sequenzierung ist eine Unterscheidung von Transkriptspezies mit ähnlicher Länge nicht über die erstellte Vollängen-cDNA-Bibliothek möglich, um so Rückschlüsse auf RNase-Schnittstellen in Transkripten ziehen zu können. Im Rahmen dieser Arbeit wurde daher eine Methode zur transkriptomweiten Identifizierung von RNase-Schnittstellen mittels RNA-Sequenzierung entwickelt.

Analyse der differentiellen Transkription der Δrnj -Mutante

Die sequenzierten Reads der Vollängen-cDNA-Bibliotheken der Δrnj -Mutante sowie die der jeweiligen Wildtypen wurden mit *DESeq2* (Love *et al.*, 2014) in Standardeinstellungen normalisiert und die differentiell transkribierten Gene identifiziert. Die drei biologischen Replikate wurden mittels hierarchischem Clustering auf Ähnlichkeit überprüft, um mögliche Ausreißer zu identifizieren (Abbildung A.3 im Anhang).

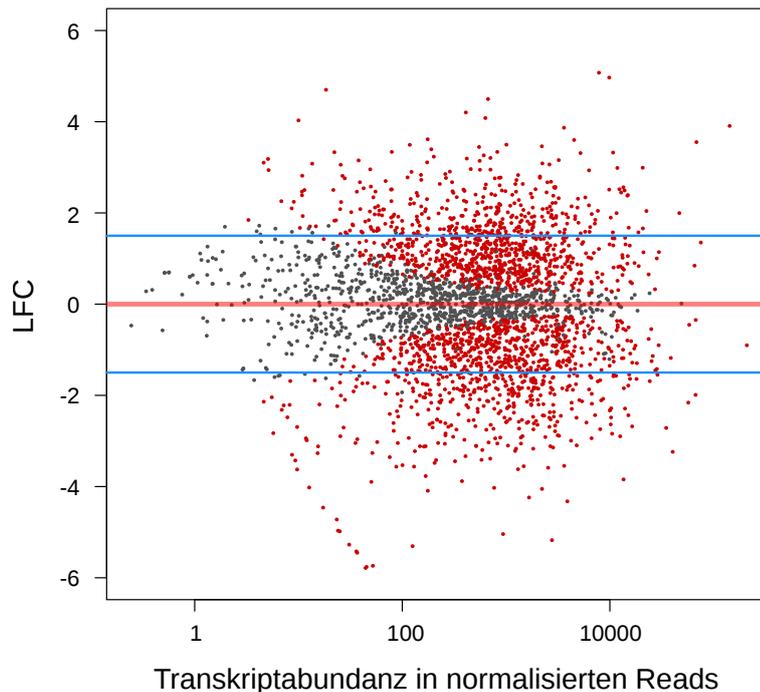


Abbildung 4.30: M/A-Plot der *C. glutamicum* Δrnj -Mutante. Aufgetragen sind der \log_2 FoldChange (LFC) gegen die Transkriptabundanz in normalisierten Reads. Jedes Gen ist durch einen Punkt dargestellt. Gene die die statistischen Bedingungen erfüllen (angepasster p-Wert < 0,05) sind rot markiert. Die blauen Linien markieren einen LFC von +1,5 bzw. -1,5. Gene die rot markiert sind und über der oberen bzw. unterhalb der unteren blauen Linie liegen wurden als differentiell transkribiert gewertet.

Wie bei den anderen RNA-Sequenzierungen wurden Gene die eine Änderung in der Transkription (\log_2 FoldChange, LFC) $\geq +1,0$ bzw. $\leq -1,0$ und einen angepassten p-Wert < 0,05 besitzen, als differentiell transkribiert klassifiziert. Insgesamt wurden so 1114 Gene (37 % des Genoms) als in der Δrnj -Mutante differentiell transkribiert klassifiziert. Um die sehr hohe Zahl an differentiell transkribierten Genen zu verringern, wurde ein strengerer LFC-Grenzwert $\geq +1,5$ bzw. $\leq -1,5$ gewählt. Mit dem strengerem Grenzwert wurden insgesamt 723 Gene als differentiell transkribiert klassifiziert, von denen 358 Gene in der Δrnj -Mutante stärker und 365 Gene schwächer transkribiert sind als im Wildtyp (Abbildung 4.30). Tabelle 4.13 enthält

4 Ergebnisse

einen Ausschnitt der am stärksten und schwächsten differenziell transkribierten Gene. Eine vollständige Liste befindet sich im Anhang (Tabelle S13).

Um die Übersicht über die differenziell transkribierten Gene zu erhöhen, wurden diese nach ihren COG-Klassen (Tatusov *et al.*, 1997) unterteilt. Auffällig ist, dass viele der differenziell transkribierten Gene den Hauptkategorien „Informationsspeicher und -verarbeitung“ sowie „Metabolismus“ zugeordnet wurden (Abbildung 4.31). Vergleichsweise wenig differenziell transkribierte Gene wurden der Kategorie „Zelluläre Prozesse und Signalwege“ zugeordnet. Wie bei der Δrne -Mutante beinhaltet die COG-Klasse „Translation“ (Klasse J) in der Δrnj -Mutante mehr stärker, als schwächer transkribierte Gene. Allerdings sind in der Δrnj -Mutante nur wenige Gene die für ribosomale Proteine kodieren betroffen. Der überwiegende Teil der differenziell transkribierten Gene dieser COG-Klasse kodieren für tRNA-Methylasen oder Translationsfaktoren, wie den *Translation Initiation Factor 2*. Eine weitere Klasse mit mehr stärker als schwächer transkribierten Genen ist die des Coenzym-Metabolismus (Klasse H). Nur die COG-Klasse G („Kohlenhydrat-Metabolismus“) besitzt ein Verhältnis von mehr schwächer als stärker transkribierten Genen.

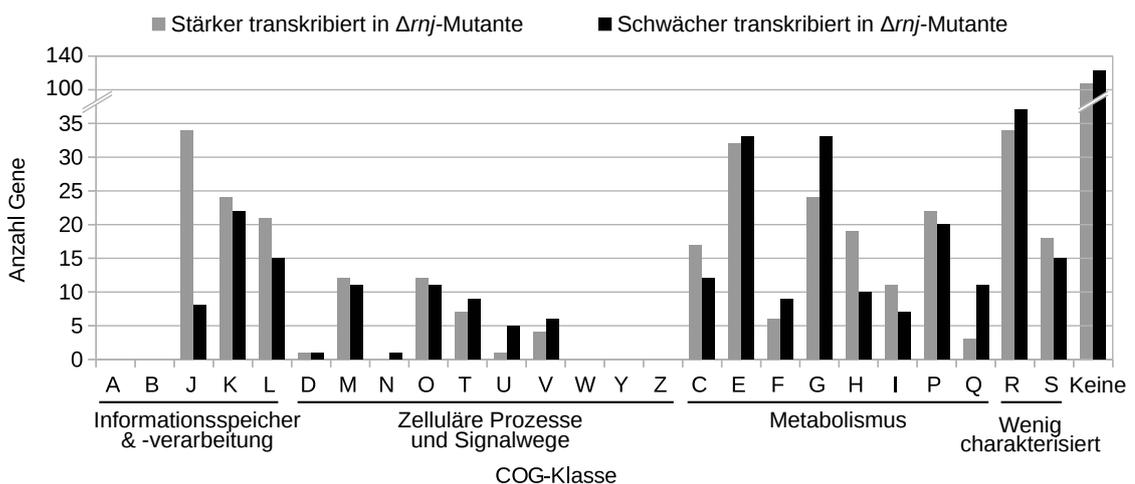


Abbildung 4.31: Klassifizierung der in der Δrnj -Mutante differenziell transkribierten Gene nach COG-Klasse. Eine detaillierte Aufschlüsselung der COG-Klassen mit Beschreibung befindet sich in Tabelle A.1 auf Seite 236 im Anhang. Die Gruppe „Keine“ enthält alle Gene die keiner COG-Klasse zugeordnet sind. Die y-Achse ist zwischen den Werten 35 und 100 unterbrochen dargestellt.

Als das am stärksten differenziell transkribierte Gen in der Δrnj -Mutante mit einem LFC-Wert von +5,07 wurde *rnpA* identifiziert, welches für den proteinogenen Anteil von RNase P kodiert (Tabelle 4.13). Die Endoribonuclease RNase P ist ein Ribozym, welches aus der katalytisch aktiven M1 RNA und einer proteinogenen Untereinheit, die die M1 RNA stabilisiert,

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

besteht (Guerrier-Takada *et al.*, 1983). Im Gegensatz zu *rnpA* wurde für das Gen der M1 RNA in der Δrnj -Mutante keine differenzielle Transkription.

Anhand der Analyse der differenziell transkribierten Gene ist es schwierig festzustellen, welches Gen durch einen direkten Effekt der Deletion des RNase J kodierenden Gens differenziell transkribiert wurde. Da RNasen RNA durch einbringen von Schnitten prozessieren bzw. degradieren, ist die Identifizierung von RNase-Schnittstellen eine Möglichkeit direkte Effekte auf differenzielle Gentranskription zu erfassen. Für die transkriptomweite Identifizierung von RNase-Schnittstellen mittels RNA-Sequenzierung wurde im Rahmen dieser Arbeit eine eigene Methode entwickelt.

Tabelle 4.13: Liste der am stärksten bzw. schwächsten differenziell transkribierten Gene in der Δrnj -Mutante im Vergleich zum Wildtyp (Ausschnitt). Die vollständige Liste aller differenziell transkribierten Gene befindet sich im Anhang (Tabelle S13). LFC, \log_2 FoldChange.

Gen	Name	Genprodukt ¹	LFC
cg3431	<i>rnpA</i>	Ribonuclease P	5,07
cg1369	<i>atpC</i>	ATP synthase F1, epsilon chain	4,97
cg1954		Hypothetical protein	4,70
cg0568		Putative secondary aromatic amino acid permease, HAAAP family	4,50
cg1173		Conserved hypothetical protein	4,20
cg0281		tRNA-specific adenosine deaminase	4,08
cg1939		Hypothetical protein	4,03
cg0655	<i>rpoA</i>	DNA-directed RNA polymerase, alpha subunit	3,91
cg2946		Conserved hypothetical protein	3,87
cg0693	<i>groEL</i>	60kDa chaperonin, putative pseudogene (C-terminal fragment)	3,61
cg3099	<i>grpE</i>	Chaperone GrpE, heat shock protein	3,60
cg0577	<i>rpoC</i>	DNA-directed RNA polymerase, beta' chain	3,55
cg3425		Conserved hypothetical protein	3,50
cg3412	<i>azlD</i>	Putative permease, branched-chain amino acid permease	3,49
cg3097	<i>hspR</i>	Putative transcriptional regulator, MerR-family	3,46
cg3281		Putative Cu ²⁺ transporting P-type ATPase	3,45
cg0338		Conserved hypothetical protein	3,40
cg2719		Putative enterochelin esterase	3,34
cg3289		Putative Thiol-disulfide isomerase or thioredoxin	3,34
cg1542		Putative membrane protein	3,33
cg0604	<i>rpsQ</i>	30S ribosomal protein S17	3,32
cg1253	<i>dapC</i>	Succinyldiaminopimelate aminotransferase, AT class I	3,31
cg0566	<i>gabT</i>	Putative 4-aminobutyrate aminotransferase, AT class II	3,26
cg2525		Putative membrane protein	3,23
cg2173	<i>dinF</i>	Putative MATE multidrug efflux protein	3,21
cg0374		Conserved hypothetical protein	3,19
cg1468		Conserved hypothetical protein	3,18
cg1338	<i>thrB</i>	Homoserine kinase	3,16
cg4012	<i>trpL</i>	Trp operon leader peptide	3,15
cg0567	<i>gabD2</i>	Putative succinate-semialdehyde dehydrogenase (NAD(P)(⁺))	3,14
cg3142		Putative membrane protein	-3,22
cg2648		Putative transcriptional regulator, ArsR-family	-3,22

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

4 Ergebnisse

Tabelle 4.13: Fortgesetzt

Gen	Name	Genprodukt ¹	LFC
cg2833	<i>cysK</i>	Cysteine synthase	-3,24
cg0332		Putative secreted protein	-3,26
cg1922		Hypothetical protein	-3,30
cg1413	<i>rbsB</i>	ABC-type ribose transporter, substrate-binding lipoprotein (TC 3.A.1.2.1)	-3,33
cg0128		Putative secreted protein	-3,35
cg1663		Conserved hypothetical protein	-3,39
cg1411	<i>rbsA</i>	ABC-type ribose transporter, ATPase subunit (TC 3.A.1.2.1)	-3,41
cg2931	<i>nanA</i>	N-acetylneuraminase lyase	-3,42
cg2252		Hypothetical protein	-3,45
cg0979		Putative transcriptional regulator, PadR-family	-3,48
cg0317		Putative transcriptional regulator, ArsR-family	-3,53
cg1044		Conserved hypothetical protein	-3,55
cg1139		Allophanate hydrolase subunit 2	-3,55
cg2601		Pirin-related protein, putative pseudogen (C-terminal fragment)	-3,56
cg3378		Conserved hypothetical protein	-3,56
cg2937		ABC-type putative dipeptide/oligopeptide transporter	-3,59
cg0768		ABC-type putative iron-siderophore transporter, ATPase subunit	-3,62
cg3082		Putative transcriptional regulator, ArsR-family	-3,77
cg2782	<i>ftn</i>	Ferritin-like protein	-3,84
cg1150		Putative NADPH-dependent FMN reductase	-3,88
cg1187	<i>ISCg10a</i>	Transposase, putative pseudogene	-3,90
cg1484		Putative secreted protein	-4,03
cg2880		Putative hydrolase, HIT-family	-4,05
cg1846		Putative transcriptional regulator, TetR-family	-4,09
cg1140		Allophanate hydrolase subunit 1	-4,24
cg1142		Putative Mn ²⁺ transporter	-4,32
cg0549		Putative pterin-4-alpha-carbinolamine dehydratase	-5,04
cg3212		Putative carboxymuconolactone decarboxylase subunit	-5,18

¹Annotation entnommen aus GenDB-Projekt „C.glut._ZFG-2.2“

4.3.5 Entwicklung einer RNA-Sequenzierungsmethode zur Erfassung von RNase-Schnittstellen

RNasen schneiden Transkripte bzw. RNA an Schnittstellen innerhalb der RNA-Sequenz. Abhängig von der Schnittstellen-Position innerhalb der Sequenz und der Häufigkeit der Schnitte durch die RNase im Transkript wird die RNA-Stabilität beeinflusst. Häufig führt das Schneiden eines Transkripts durch eine RNase zur Degradation des jeweiligen Transkripts (Arraiano *et al.*, 2010; Laalami *et al.*, 2014). Alternativ kann ein Transkript durch das Schneiden durch eine RNase Prozessiert werden, um die Funktionalität des Transkripts zu ermöglichen (Arraiano *et al.*, 2010; Durand und Condon, 2018). RNase-Schnittstellen zur Degradation oder Prozessierung von Transkripten sind somit ein direkter Effekt der RNase-Einwirkung.

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Mit den Protokollen zur Erzeugung von Vollängen-cDNA- (Illumina) oder Primärtranskript-Bibliotheken (Pfeifer-Sancar *et al.*, 2013) können RNase-Schnittstellen nicht erfasst werden. Daher wurde in dieser Arbeit eine eigene RNA-Sequenzierungsmethode zur Identifizierung von RNase-Schnittstellen entwickelt, um die direkten Effekte von RNasen auf das Transkriptom zu charakterisieren. Die entwickelte Methode wird im Folgenden als *RNase-Assay* bezeichnet.

Das generelle Prinzip des entwickelten *RNase-Assays* wird im Folgenden beschrieben und in Abbildung 4.32 grafisch dargestellt.

(1.) RNasen schneiden RNA an unterschiedlichen Positionen im Transkript, sodass (2.) unterschiedliche Fragmente entstehen. (3.) Nach erfolgter RNA-Isolierung wird an das 5'-Ende dieser Fragmente ein RNA-Adapter (schwarzer gestrichelter Kasten) ligiert, der die Schnittstelle markiert. (4.) Über einen Schlaufen-Adapter (grauer gestrichelter Kasten mit Pfeil) wird das RNA-Fragment in cDNA revers-transkribiert (Erststrangsynthese). (5.) Es entsteht ein RNA-cDNA-Hybrid. Anschließend wird die RNA mittels RNase H verdaut und (6.) die Zweitstrangsynthese sowie die cDNA-Amplifikation über PCR durchgeführt. Die dabei eingesetzten Primer (schwarze Striche) enthalten Überhänge, die eine Sequenzierung auf Illumina Sequenziermaschinen ermöglichen. (7.) Die erste sequenzierte Base entspricht der Position einer RNase-Schnittstelle (farbiger Balken im Diagramm). Um die Schnittstellen der untersuchten RNase zu identifizieren, werden die Schnittstellen-Positionen von einem RNase-defizienten Stamm mit denen des Wildtyps verglichen und entsprechend gefiltert. Dazu werden die Reads zunächst auf die erste Base am 5'-Ende reduziert und anschließend die Read-Anzahl an jeder Genomposition des Vorwärts- und Gegenstrangs bestimmt. (8.) Mit *DESeq2* (Love *et al.*, 2014) werden die Read-Anzahlen der Proben normalisiert und anschließend an jeder Position auf differenzielle Read-Anzahlen geprüft. Die Positionen, an denen ein differenzieller Unterschied vorliegt wurde, wurde als mögliche RNase-Schnittstelle betrachtet. Nach der Auswertung mit *DESeq2* verbleiben nur Positionen die den Schnittstellen der untersuchten RNase zugehörig sind. Eine detaillierte Beschreibung der Probenvorbereitung, Erstellung der cDNA-Bibliothek und bioinformatischen Auswertung befindet sich im Methodenteil.

Der *in vitro RNase-Assay* funktioniert auf ähnliche Weise wie der oben beschriebene *in vivo RNase-Assay*. Zur Untersuchung einer RNase wird diese aufgereinigt und die RNA des Deletionsstamms dieser RNase mit der aufgereinigten RNase behandelt. Der RNA des für die aufgereinigte RNase defizienten Stamms fehlen die Schnittstellen der zu untersuchenden RNase. Als Kontrolle wird die unbehandelte RNA der Deletionsmutante verwendet. Beide Ansätze („behandelt“ / „unbehandelt“) werden anschließend wie in Abbildung 4.32 ab Schritt 3. gezeigt verarbeitet. Ein detailliertes Protokoll für den *in vivo* bzw. *in vitro RNase-Assay* in Form einer Schritt-für-Schritt-Anleitung befindet sich auf der beiliegenden Daten-CD.

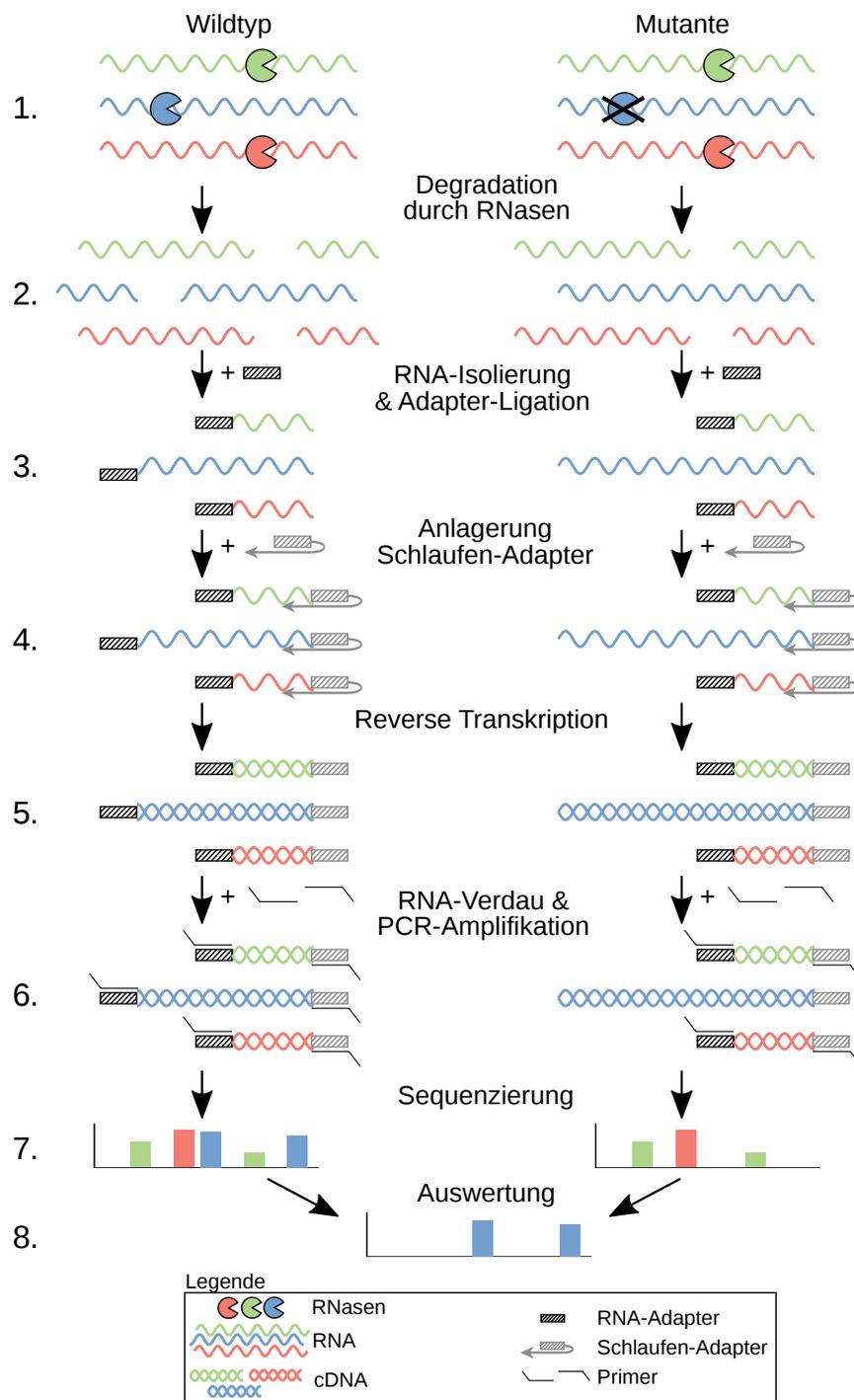


Abbildung 4.32: Schematischer Ablauf des *RNase-Assays* zur Identifizierung von RNase-Schnittstellen. Der *RNase-Assay* kann *in vivo* oder *in vitro* durchgeführt werden. Gezeigt ist der Ablauf des *in vivo* Protokolls mit dem Wildtyp (linke Spalte) und einem RNase-defizienten Stamm (rechte Spalte, durchkreuzte RNase). Eine detaillierte Beschreibung des Ablaufs befindet sich im Text.

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Der hier entwickelte *RNase-Assay* wurde für die Charakterisierung der Schnittstellen der RNase E/G und RNase J eingesetzt. Die dabei erzielten Ergebnisse sind in den nachfolgenden Kapiteln aufgeführt.

4.3.6 Endoribonukleolytische mRNA-Degradation durch die RNase E/G

Der im Rahmen dieser Arbeit entwickelte *in vivo RNase-Assay* wurde für die Identifizierung der Schnittstellen von RNase E/G in *C. glutamicum* eingesetzt. Dazu wurde die RNA der Δrne -Mutante sowie die RNA des Wildtyps jeweils in zwei biologischen Replikaten verwendet und nach der oben beschriebenen Methode zu cDNA-Bibliotheken verarbeitet und auf einem Illumina MiSeq sequenziert. Die erhaltenen *single-end* Reads wurden gegen das *C. glutamicum* ATCC 13032 (Kalinowski *et al.*, 2003) mit *segemehl* (Hoffmann *et al.*, 2009) mit einer festgelegten Genauigkeit von 95 % gemappt (Tabelle 4.14). Anschließend wurden die RNase E/G Schnittstellen über Vergleich der Signalhöhen an jeder Genomposition auf beiden Strängen zwischen Wildtyp und Δrne -Mutante mittels *DESeq2* (Love *et al.*, 2014) identifiziert. Alle Genompositionen die im Wildtyp mindestens dreimal so viele Reads wie in der Δrne -Mutante, einen angepassten p-Wert $\leq 0,05$ besaßen und nicht mit einem in *C. glutamicum* bekannten Transkriptionsstart (*Transcription Start Site*, TSS) übereinstimmen, wurden als RNase E/G Schnittstelle identifiziert.

Tabelle 4.14: Sequenzierte und gemappte Reads der cDNA-Bibliotheken für die Identifizierung von RNase E/G Schnittstellen mittels *RNase-Assay*.

<i>RNase-Assay</i> cDNA-Bibliothek	Sequenzierte Reads	Gemappte Reads	Gemappte Reads in %
Wildtyp Replikat 1	1.612.474	1.535.321	95,22
Wildtyp Replikat 2	2.815.527	2.659.348	94,45
Δrne -Mutante Replikat 1	2.770.855	2.556.231	92,25
Δrne -Mutante Replikat 2	3.635.820	3.313.664	91,14

Es wurden 2.528 RNase E/G Schnittstellen im Transkriptom von *C. glutamicum* identifiziert. Von diesen wurden 567 Schnittstellen 5'-UTRs und 1.607 proteinkodierenden Sequenzen (CDS) zugeordnet (Abbildung 4.33). Die restlichen 354 Schnittstellen wurden *antisense* RNAs und intergenischen Transkripten zugeordnet. Insgesamt wurden 210 5'-UTRs und 518 CDS identifiziert, die mindestens eine RNase E/G Schnittstelle enthalten. Die vollständige Liste aller RNase E/G Schnittstellen befindet sich im Anhang (Tabelle S12).

Obwohl für RNase E/G gezeigt werden konnte, dass diese an der Prozessierung von Vorläufern der 5S rRNA beteiligt ist (vgl. Abschnitt 4.3.4), wurde im *RNase-Assay* nur eine RNase E/G Schnittstelle innerhalb der 23S rRNA identifiziert. Es wird vermutet, dass die

4 Ergebnisse

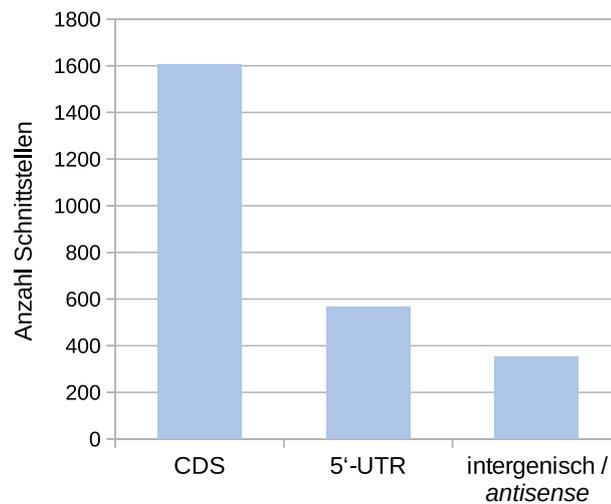


Abbildung 4.33: Anzahl der RNase E/G Schnittstellen nach Transkript bzw. -abschnitt. Schnittstellen, die im intergenischen Bereich bzw. *antisense* zu annotierten Transkripten liegen, wurden unter der Kategorie „intergenisch / *antisense*“ zusammengefasst.

rRNA-Abreicherung während der Erstellung der cDNA-Bibliotheken dazu führt, dass keine bzw. nur wenige RNase-Schnittstellen gefunden werden können.

Der überwiegende Teil an mRNAs wird von RNase E/G wenige Male geschnitten (Abbildung 4.34A). Nur für wenige mRNAs wurden mehr als zehn Schnitte beobachtet. Dies deutet auf eine endoribonukleolytische Funktionalität von RNase E/G hin und entspricht den Erwartungen aus den Modellen von RNase E aus *E. coli*. RNase E initiiert dort die Transkriptdegradation durch einige wenige Schnitte. Den eigentlichen Abbau übernehmen prozessive Exoribonukleasen (Arraiano *et al.*, 2010; Laalami *et al.*, 2014). Für *C. glutamicum* RNase E/G wurden allerdings auch Transkripte mit über 80 Schnitten gefunden. Diese Transkripte wurden durch RNase E/G in sehr kurze Fragmente gespalten und somit nahezu vollständig abgebaut. Für das Transkript von *cg1476 / thiC* wurden die meisten Schnittstellen (174 Schnitte) identifiziert. Bei einer Transkriptlänge von knapp 1.900 nt wurde das *thiC*-Transkript in ca. 10 nt kurze Fragmente geschnitten und so direkt von RNase E/G degradiert.

Da RNase E/G hauptsächlich Endoribonuklease-Aktivität besitzt, wurden die Sequenzen um die identifizierten Schnittstellen auf konservierte Motive mit *Improbizer* (Ao *et al.*, 2004) analysiert. Das von *Improbizer* gefundene Motiv mit der Sequenz AAA↓CUU wurde von *MEME* (Bailey und Elkan, 1994) (Version 4.8.1) bestätigt (p-Wert 4,2e-185) (Abbildung 4.34B). Wie auch *E. coli* RNase E, hat *C. glutamicum* RNase E/G ebenfalls eine Präferenz für AU-reiche Sequenzen, mit dem Unterschied, dass in der Sequenz ein konserviertes Cytosin liegt. Diese

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

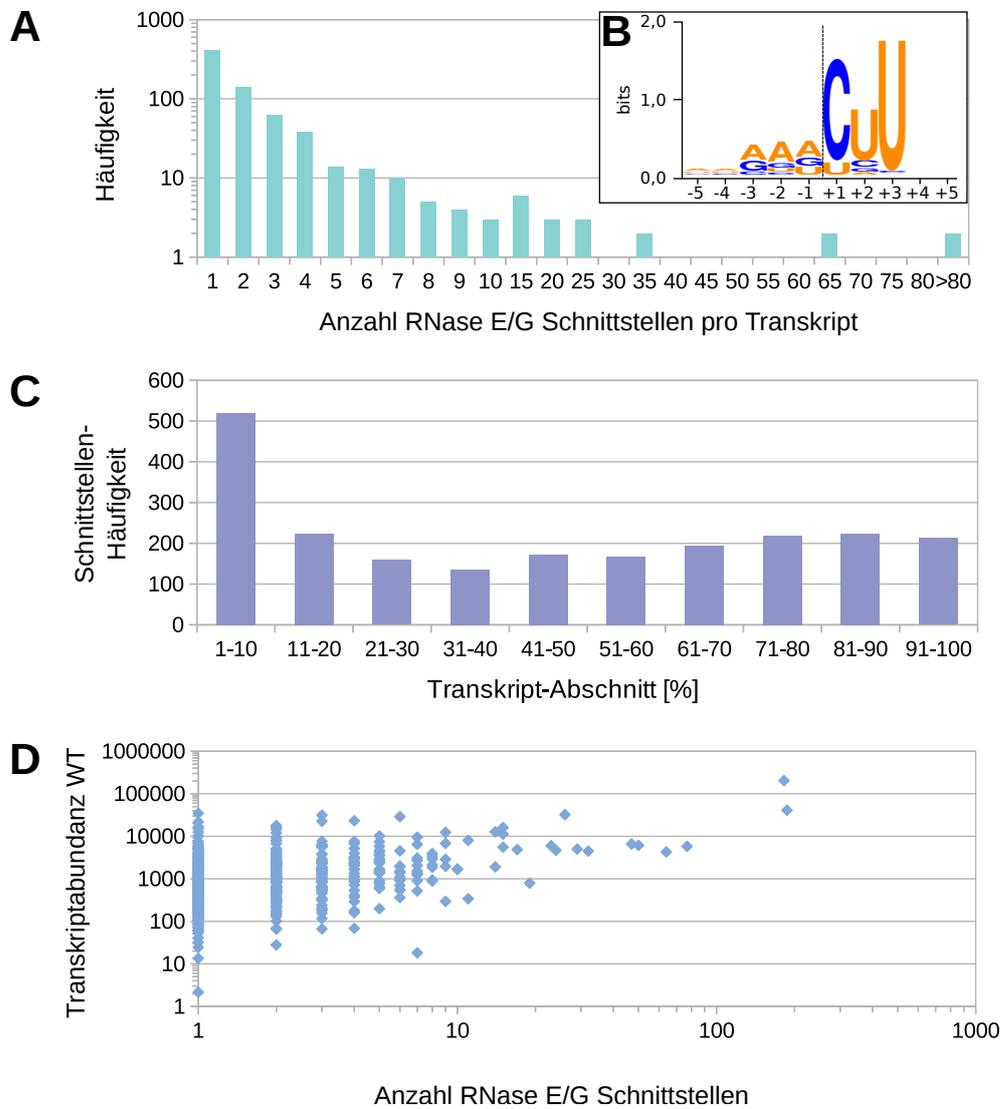


Abbildung 4.34: Analyse der RNase E/G Schnittstellen: (A) Histogramm der RNase E/G Schnittstellen pro Transkript. Transkripte mit mehr als 80 Schnittstellen wurden unter der Kategorie „>80“ zusammengefasst. (B) Motiv der RNase E/G Schnittstellen. Die gestrichelte Linie zeigt die Schnittstelle an. (C) Häufigkeit der RNase E/G Schnittstellen abhängig von der relativen Position im Transkript. Um die unterschiedlich langen Transkripte vergleichen zu können, wurde jedes Transkript in zehn Abschnitte unterteilt und die Schnittstellen in den jeweiligen Bereichen gezählt. (D) Anzahl der RNase E/G Schnittstellen pro Transkript in Abhängigkeit zur Transkriptabundanz (gemessen in mit *DESeq2* normalisierten Reads) des Transkripts im Wildtyp (WT).

deutliche Abweichung vom *E. coli* RNase E Motiv könnte in der deutlich unterschiedlichen Struktur von RNase E/G begründet sein.

Für einen Vergleich der Schnittstellenpositionen von unterschiedlich langen Transkripten wurden alle Transkripte mit identifizierten Schnittstellen längennormalisiert und in zehn Abschnitte unterteilt. Die meisten Schnittstellen innerhalb der Transkripte wurden in den ersten zehn Prozent eines Transkripts identifiziert (Abbildung 4.34C). In den nachfolgenden Abschnitten sinkt die Schnittstellen-Häufigkeit stark ab und steigt erst wieder zum 3'-Ende leicht an.

Es ist denkbar, dass Transkripte mit hoher Zahl an Schnittstellen nur eine geringe Transkriptabundanz aufweisen, da diese Transkripte schnell degradiert werden müssten. Um dies zu überprüfen, wurde die Zahl an identifizierten Schnittstellen pro Transkript gegen die Transkriptabundanz im Wildtyp (da dieser RNase E/G besitzt) aufgetragen (Abbildung 4.34D). Für Transkripte mit nur einer Schnittstelle wurden über fünf Größenordnungen schwankende Transkriptabundanzen gemessen. Ähnliches wurde für Transkripte mit einigen wenigen Schnittstellen beobachtet. Nur für Transkripte mit sehr hoher Zahl an Schnittstellen wurde eine geringe Varianz in der Transkriptabundanz gemessen. Demnach hat die Anzahl an Schnittstellen keinen direkten Effekt auf die Transkriptabundanz. Transkripte von Genen die besonders viele RNase E/G Schnittstellen besitzen (z. B. *thiC*), wurden in der Δrne -Mutante besonders stark differenziell transkribiert. Für diese Gene ist ein direkter Effekt auf die Transkription durch die Deletion der RNase E/G wahrscheinlich.

Der *in vitro* RNase-Assay konnte mit RNase E/G nicht durchgeführt werden, da die RNase nicht aufgereinigt werden konnte. Obwohl das Plasmidkonstrukt zur Aufreinigung von RNase E/G über das IMPACT-CN System (New England Biolabs) in *E. coli* stabil repliziert wurde, konnte unter verschiedenen Bedingungen kein RNase-Protein mittels SDS-PAGE nachgewiesen werden (Daten nicht gezeigt). Im Gegensatz dazu wurde RNase J über das IMPACT-CN System in ausreichender Menge aufgereinigt und im *in vitro* RNase-Assay eingesetzt. Der Grund warum RNase E/G nicht in *E. coli* aufgereinigt werden konnte ist unklar.

4.3.7 Die Hauptfunktion der bifunktionalen RNase J ist der exoribonukleolytische Abbau ausgehend vom 5'-Ende

Der in dieser Arbeit entwickelte RNase-Assay wurde auch für die Identifizierung der RNase J Schnittstellen in *C. glutamicum* eingesetzt. Für RNase J wurde der *in vivo* RNase-Assay, als auch der *in vitro* RNase-Assay durchgeführt. Für den *in vitro* RNase-Assay wurde RNase J als aufgereinigtes Protein benötigt.

Aufreinigung von RNase J

Die Aufreinigung von RNase J erfolgte mittels IMPACT-CN System (New England Biolabs), welches die Marker-freie Aufreinigung von Proteinen ermöglicht. Dazu wurde das für RNase J kodierende Gen *rnj* aus *C. glutamicum* mittels Gibson-Assembly in das Plasmid pTXB1 kloniert. Die Primer wurden dabei so gewählt, dass das Stop-Codon von *rnj* entfernt wird, um eine translationale Fusion am C-Terminus des RNase-Proteins mit dem Intein und der Chitinbinde-domäne zu erreichen. Das entstandene Konstrukt pTXB1::*rnj* wurde zur Aufreinigung in den Expressionsstamm *E. coli* ER2566 transformiert und anschließend nach Anleitung des Herstellers über Affinitätschromatographie aufgereinigt. Die Reinheit der aufgereinigten RNase J wurde mittels SDS-PAGE überprüft und die Proteinkonzentration über den Bradford-Test mit Roti-Nanoquant (Carl Roth) nach Herstellerangaben bestimmt. Die RNase J wurde in einer Konzentration von 278 $\mu\text{g mL}^{-1}$ aufgereinigt. Das aufgereinigte Protein wurde einerseits über die korrekte Größe von 77,6 kDa in der SDS-PAGE (Abbildung A.11), als auch dem mittels MALDI-MS (*Matrix Assisted Laser Desorption/Ionisation-Mass spectrometry*) bestimmten Mascott-Wert (Perkins *et al.*, 1999) von 291 eindeutig als *C. glutamicum* RNase J identifiziert.

Aktivitätstest der aufgereinigten RNase J

Bevor die aufgereinigte RNase J für den *in vitro* RNase-Assay verwendet wurde, wurde sie auf Aktivität überprüft. Dazu wurde RNase J mit der RNA der Δrnj -Mutante bei 30 °C inkubiert und die Abnahme der *bulk* RNA über die Zeit im Bioanalyzer 2100 (Agilent) mit dem „RNA Pico 6000 Kit“ (Agilent) gemessen. Die RNA wird dabei mit einem Fluoreszenzfarbstoff markiert und über Kapillarelektrophorese der Größe nach aufgetrennt. Die Fläche unter der Kurve des Chromatogramms entspricht der RNA-Menge der Probe. Da nur *bulk* RNA gemessen werden kann, ist eine Aussage über konkrete Schnittstellen oder bestimmte Transkriptspezies nicht möglich. Es kann nur der allgemeine Verdau durch die RNase anhand der Abnahme der charakteristischen *Peak*-Höhen abgeschätzt werden. Bereits nach 10 min Inkubationszeit mit RNase J wurde eine deutliche Abnahme der *Peak*-Höhen beobachtet (Abbildung A.12). Aus diesem Grund wurde für alle *in vitro* RNase-Assays eine Inkubationszeit von 10 min gewählt.

Durchführung des *in vivo* und *in vitro* RNase-Assays mit RNase J

Die aufgereinigte RNase J wurde anschließend im *in vitro* RNase-Assay eingesetzt. Dazu wurden 50 pmol RNase J mit 2,5 μg RNA aus der Δrnj -Mutante für 10 min bei 30 °C inkubiert, anschließend die RNase mit Phenol-Chloroform-Isoamylalkohol-Reagenz (PCI-Reagenz) gefällt, die RNA aufgereinigt und die rRNA abgereichert. Als Kontrolle wurde die

4 Ergebnisse

selbe Menge RNA der Δrnj -Mutante ohne RNase-Behandlung mit PCI-Reagenz inkubiert und anschließend aufgereinigt. Die RNA von beiden Proben („Behandelt“ bzw. „Unbehandelt“) wurde anschließend im *RNase-Assays*, wie oben in Abschnitt 4.3.5 beschrieben, zu cDNA-Bibliotheken weiterverarbeitet und auf einem Illumina MiSeq sequenziert (Tabelle 4.15).

Tabelle 4.15: Sequenzierte und gemappte Reads der cDNA-Bibliotheken für die Identifizierung von RNase J Schnittstellen mittels *RNase-Assay*.

cDNA-Bibliothek	Sequenzierte Reads	Gemappte Reads	Gemappte Reads in %
<i>in vitro RNase-Assay</i>			
RNA Δrnj -Mutante, mit RNase J behandelt	1.679.993	1.568.278	93,35
RNA Δrnj -Mutante, unbehandelt	848.158	805.728	95,00
<i>in vivo RNase-Assay</i>			
Δrnj -Mutante	1.508.661	1.349.201	89,43
Wildtyp	1.926.961	699.199	36,29

Für RNase J wurde außerdem auch der *in vivo RNase-Assay* mit der RNA des *C. glutamicum* Wildtyps und der *C. glutamicum* Δrnj -Mutante durchgeführt. Für den *in vivo RNase-Assay* wurden analog zum *in vivo RNase-Assay* von RNase E/G jeweils 5 μ g RNA eingesetzt, die ribosomale RNA abgereichert und die zu cDNA-Bibliotheken verarbeitet und auf einem Illumina MiSeq sequenziert (Tabelle 4.15).

Die Kriterien zur Identifizierung einer RNase J Schnittstelle waren zum einen eine mindestens 3-fache Änderung in der Read-Abundanz an einer Genomposition im Vergleich zur Kontrolle (Probe „Unbehandelt“ bzw. RNA der Δrnj -Mutante). Zum anderen wurden die im *in vitro RNase-Assay* erhaltenen Schnittstellen-Positionen mit denen des *in vivo RNase-Assays* verglichen und nur in beiden *RNase-Assays* identische Positionen als tatsächliche Schnittstellen für RNase J gewertet. Insgesamt wurden 5.850 RNase J Schnittstellen im *C. glutamicum* Transkriptom identifiziert. Von diesen Schnittstellen wurden 4.705 proteinkodierenden Sequenzen (CDS), 770 5'-UTRs und 365 *antisense* RNAs und intergenischen Transkripten zugeordnet (Abbildung 4.35). Insgesamt wurden 1.029 CDS und 276 5'-UTRs identifiziert, die mindestens eine RNase J Schnittstelle besitzen. Zehn RNase J Schnittstellen wurden innerhalb von rRNA-Regionen identifiziert. So wurden drei Schnittstellen *upstream* von der 16S rRNA und sieben Schnittstellen innerhalb der 23S rRNA identifiziert. Die vollständige Liste aller identifizierten RNase J Schnittstellen befindet sich im Anhang (Tabelle S14).

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

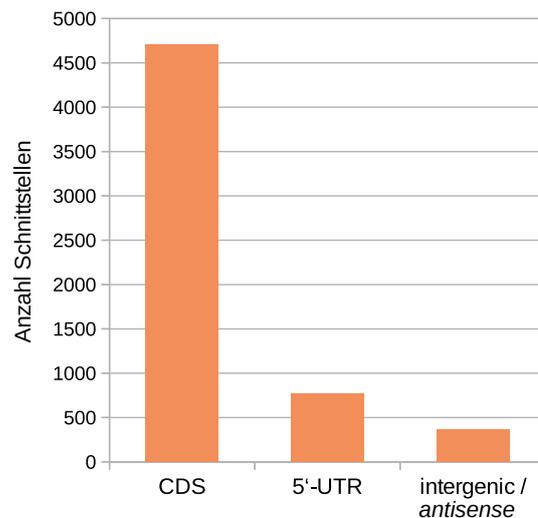


Abbildung 4.35: Anzahl der RNase J Schnittstellen nach Transkript bzw. -abschnitt. Schnittstellen, die im intergenischen Bereich bzw. *antisense* zu annotierten Transkripten liegen, wurden unter der Kategorie „intergenisch / *antisense*“ zusammengefasst. CDS, proteinkodierende Sequenz (*coding sequence*); 5'-UTR, 5' untranslatierte Region.

Obwohl die meisten Transkripte nur wenige Male von RNase J geschnitten wurden, nimmt die Schnittstellenzahl pro Transkript für RNase J über die Transkriptlänge weniger stark ab (Abbildung 4.36A), als bei RNase E/G (vgl. Abbildung 4.34A auf Seite 141). Darüber hinaus wurden mehr Transkripte mit großer Zahl an RNase J Schnittstellen identifiziert, als für RNase E/G. Für das Gen *cg0587 / tuf*, welches für den Elongationsfaktor Tu kodiert, wurden 159 Schnittstellen identifiziert. Es ist damit das Gen mit der höchsten Zahl an identifizierten RNase J Schnittstellen.

Um die relative Position der Schnittstellen innerhalb verschiedener Transkripte vergleichen zu können, wurden alle Transkripte (bestehend aus 5'-UTR mit CDS) auf Länge normalisiert, in zehn Abschnitte unterteilt und anschließend die Schnittstellenhäufigkeiten bestimmt (Abbildung 4.36B). Die höchste Schnittstellenhäufigkeit wurde in den ersten 10 % eines Transkripts gemessen. Über alle folgenden Transkriptabschnitte sinkt die Schnittstellenhäufigkeit langsam ab. Dies deutet darauf hin, dass RNase J aus *C. glutamicum* eine prozessive 5'-3'-Exoribonuklease-Aktivität besitzt.

4 Ergebnisse

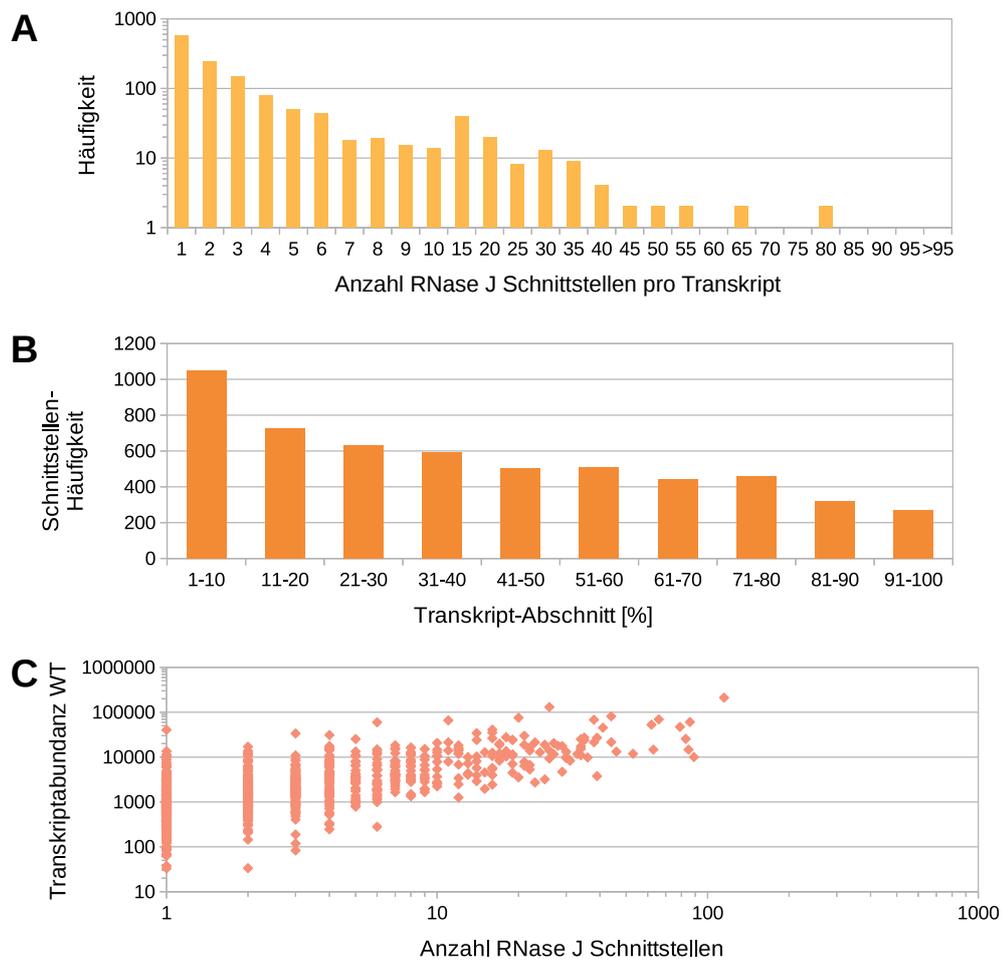


Abbildung 4.36: Analyse der RNase E/G Schnittstellen: (A) Histogramm der RNase J Schnittstellen pro Transkript. Transkripte mit mehr als 95 Schnittstellen wurden unter der Kategorie „>95“ zusammengefasst. (B) Häufigkeit der RNase J Schnittstellen abhängig von der relativen Position im Transkript. Um die unterschiedlich langen Transkripte vergleichen zu können, wurde jedes Transkript in zehn Abschnitte unterteilt und die Schnittstellen in den jeweiligen Bereichen gezählt. (C) Anzahl der RNase E/G Schnittstellen pro Transkript in Abhängigkeit zur Transkriptabundanz (gemessen in mit *DESeq2* normalisierten Reads) des Transkripts im Wildtyp (WT).

Für *B. subtilis* RNase J1 ist bekannt, dass das Enzym sowohl 5'-3-Exoribonuklease- als auch Endoribonuklease-Aktivität besitzt (Even *et al.*, 2005; Mathy *et al.*, 2007). Für *C. glutamicum* wäre es denkbar, dass Transkripte, für die nur eine RNase J Schnittstelle identifiziert wurde, endoribonukleolytisch gespalten wurden. Insgesamt wurden für 412 Transkripte nur eine RNase J-Schnittstelle nachgewiesen. Diese putativen endoribonukleolytischen Schnittstellen wurden auf konservierte Motive mit Improbizer (Ao *et al.*, 2004) überprüft. Im Gegensatz zu RNase E/G wurde für die endoribonukleolytischen RNase J Schnittstellen kein konserviertes

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Motiv vorhergesagt (Daten nicht gezeigt). Möglicherweise befinden sich in den Transkripten mit nur einer RNase J Schnittstelle mehr Schnittstellen, die nicht detektiert wurden, so dass diese Transkripte nicht endo-, sondern exoribonukleolytisch abgebaut wurden. Aufgrund der dabei entstehenden Zahl an sehr kleinen Fragmenten ist es schwierig alle Schnittstellen einer prozessiven Exoribonuklease zu identifizieren.

Wie auch bei der Schnittstellenanalyse von RNase E/G, korreliert die Anzahl der Schnittstellen pro Transkript nicht mit der Transkriptabundanz im Wildtyp (Abbildung 4.36C). Es kann somit nicht pauschal anhand der Zahl der RNase J Schnittstellen unterschieden werden, ob es sich bei einem in der Δrnj -Mutante differenziell transkribierten Gen um einen direkten oder indirekten Effekt der RNase-Deletion handelt.

Cg.RNase J schneidet Transkripte der RNase P und des *trpL*-leader-Peptids

In der *C. glutamicum* Δrnj -Mutante ist das Gen *rnpA*, welches den proteinogenen Teil des RNase P Ribozyms kodiert (Guerrier-Takada *et al.*, 1983), das am stärksten differenziell transkribierte Gen (LFC +5,07). Das Gen *rnpB* kodiert die M1 RNA, die die RNA-Komponente von RNase P darstellt (Guerrier-Takada *et al.*, 1983). Im Gegensatz zu *rnpA* ist *rnpB* in der Δrnj -Mutante schwächer transkribiert (LFC -1,42). Die M1 RNA wird in einem primären Operon aus den Genen *cg2457*, *cg2456*, *cg2455* und M1 RNA sowie monocistronisch in einem sekundären Operon transkribiert (Pfeifer-Sancar *et al.*, 2013). Für die M1 RNA wurden im RNase-Assay sechs RNase J Schnittstellen identifiziert. Zwei Schnittstellen befinden sich 1 bzw. 2 nt *upstream* des TSS von M1 RNA und müssen somit aus dem Transkript des primären Operons stammen. Zwei weitere Schnittstellen befinden sich 1 bzw. 2 nt *downstream* des M1 RNA TSS. Diese Schnittstellen können sowohl aus dem Transkript des primären als auch dem des sekundären Operon entstammen. Die verbleibenden zwei Schnittstellen befinden sich 62 bzw. 75 nt *downstream* des TSS.

Für das Gen *rnpA* wurden keine RNase J Schnittstellen identifiziert. Allerdings befindet sich *rnpA* in einem Operon aus den Genen *rpmH*, *rnpA*, *cg3430* und *cg3429*. Innerhalb von *rpmH* wurden neun RNase J Schnittstellen identifiziert, die über das gesamte Transkript verteilt liegen und somit auf Degradation des *rpmH*-Transkripts durch RNase J hindeuten.

Eine bekannte Degradation durch RNase J in *B. subtilis* findet in der 5'-UTR des *trp*-Operons statt (Deikus *et al.*, 2008). Die Gene des *trp*-Operons in *C. glutamicum* sind bis auf das Gen des leader-Peptids *trpL* nicht differenziell in der Δrnj -Mutante transkribiert. In der 5'-UTR des ersten Gens des *trp*-Operons befindet sich das Gen für das leader-Peptid TrpL (Neshat *et al.*, 2014). Dieses Gen ist mit einem LFC von +3,15 in der Δrnj -Mutante deutlich verstärkt transkribiert. Mit dem in dieser Arbeit entwickelten RNase-Assays zur Identifizierung von

4 Ergebnisse

RNase-Schnittstellen wurde eine RNase J Schnittstelle in *trpL* 11 nt *downstream* des TSS nachgewiesen, welche eine Interaktion zwischen *leader*-Peptid und RNase J hindeutet.

4.3.8 Regulation von Riboswitches und RNA-Motiven durch die RNasen E/G und J

In Kapitel 4.2.10 (Seite 4.2.10) wurde die Regulation des FMN-Riboswitches in *cg0083* / *ribM* durch den Rho-Faktor beschrieben und somit die Ergebnisse von Takemoto *et al.* (2015) bestätigt. Weiterhin hatten Takemoto und Kollegen den Einfluss der RNase E/G auf den FMN-Riboswitch gezeigt. Allerdings hatten sie keine exakte Schnittstelle angegeben. Riboswitches und vermutlich auch andere regulatorisch wirkende Sequenzen in Transkripten sind somit Ziele von RNasen, die diese kodierten Regulatoren regulieren.

Tabelle 4.16: Riboswitches und RNA-Motive mit RNase E/G und RNase J Schnittstellen. Für jeden Riboswitch sowie RNA-Motiv bzw. RNA *leader* wurde geprüft, ob RNase-Schnittstellen innerhalb und in der Nähe (max. 25 nt *downstream* des Startcodons) des jeweiligen Riboswitches vorliegen und ob das jeweilige Gen in der Δrne -Mutante bzw. Δrnj -Mutante differenziell transkribiert ist. RNase-Schnittstellen innerhalb eines Riboswitches, Motivs oder RNA *leaders* stehen vor der Klammer, Schnittstellen die sich in der Nähe befinden stehen in der Klammer. Die Namen wurden aus der RFAM-Datenbank (Nawrocki *et al.*, 2015) übernommen. LFC, \log_2 FoldChange.

Name	Gen	RNase E/G Schnittstellen	Differenziell in Δrne -Mutante	RNase J Schnittstellen	Differenziell in Δrnj -Mutante
FMN-Riboswitch	<i>cg0083</i>	0 (1)	nein	–	nein
<i>cspA</i> mRNA 5'-UTR	<i>cg0215</i> / <i>cspA</i>	–	nein	–	ja (LFC +2,0)
TPP-Riboswitch	<i>cg0825</i>	1 (0)	ja (LFC –2,1)	–	nein
<i>ydaO-yuaA leader</i>	<i>cg0936</i> / <i>rpf1</i>	–	ja (LFC +3,1)	–	nein
TPP-Riboswitch	<i>cg1227</i>	4 (0)	ja (LFC –2,0)	–	nein
mini- <i>ykkC</i> RNA-Motiv	<i>cg1231</i> / <i>chaA</i>	–	nein	–	ja (LFC +2,7)
TPP-Riboswitch	<i>cg1476</i> / <i>thiC</i>	13 (2)	ja (LFC –4,1)	–	ja (LFC –1,6)
SAM-IV-Riboswitch	<i>cg1478</i>	–	ja (LFC –3,1)	–	ja (LFC –2,0)
TPP-Riboswitch	<i>cg1655</i> / <i>thiM</i>	–	ja (LFC –1,4)	–	nein
<i>yybP-ykoY leader</i>	<i>cg1660</i>	–	ja (LFC +2,0)	–	nein
<i>yybP-ykoY leader</i>	<i>cg2157</i> / <i>terC</i>	–	nein	–	nein
TPP-Riboswitch	<i>cg2236</i> / <i>thiE</i>	–	ja (LFC –2,4)	–	nein
<i>mraW</i> RNA-Motiv	<i>cg2377</i> / <i>mraW</i>	2 (0)	nein	–	nein
<i>ydaO-yuaA leader</i>	<i>cg2402</i>	2 (0)	ja (LFC +1,3)	10 (0)	nein
<i>msiK</i> RNA-Motiv	<i>cg2708</i> / <i>msiK1</i>	–	nein	4 (3)	nein
<i>yybP-ykoY leader</i>	<i>cg2785</i>	–	nein	–	nein

Im Rahmen dieser Arbeit wurde ein *RNase-Assay* entwickelt, um RNase-Schnittstellen transkriptomweit identifizieren zu können. Im Folgenden wird daher der Fokus auf die in *C. glutamicum* bekannten Riboswitches, RNA-Motive und *leader*-Sequenzen gelegt und auf RNase E/G und RNase J Schnittstellen untersucht.

4.3 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Für drei TPP-Riboswitches (*cg0825*, *cg1227* und *cg1476 / thiC*) wurden mittels *RNase-Assay* RNase E/G Schnittstellen identifiziert (Tabelle 4.16). Desweiteren wurden Schnittstellen im *mraW* RNA-Motiv von *cg2377 / mraW*, sowie im *ydaO-yuaA leader* von *cg2402* identifiziert. Innerhalb der Sequenz des FMN-Riboswitches selbst wurde keine RNase E/G Schnittstelle identifiziert. Allerdings wurde 6 nt *downstream* des Startcodons eine RNase E/G Schnittstelle an Genomposition 66.192 bp identifiziert (Abbildung 4.37A). Bei dieser Schnittstelle handelt es sich mit aller Wahrscheinlichkeit um die von Takemoto *et al.* (2015) beschriebene Schnittstelle. Die exakte Position der Schnittstelle konnte nun erstmalig nachgewiesen werden. Da sich die Schnittstelle *downstream* des Startcodons befindet, wird durch RNase E/G neben dem Riboswitch die ribosomale Bindestelle als auch das Startcodon aus dem Transkript entfernt. Trotz der Entfernung des Riboswitches aus dem Transkript kann daher keine unregulierte Translation der CDS erfolgen, da die ribosomale Bindestelle ebenfalls aus dem Transkript entfernt wurde. Durch die Beseitigung des Startcodons wird auch eine mögliche *leader*-freie (*leaderless*) Translation unterbunden. Der Zweck der RNase E/G Schnittstelle ist demnach eindeutig der RNA-Degradation zuzuordnen.

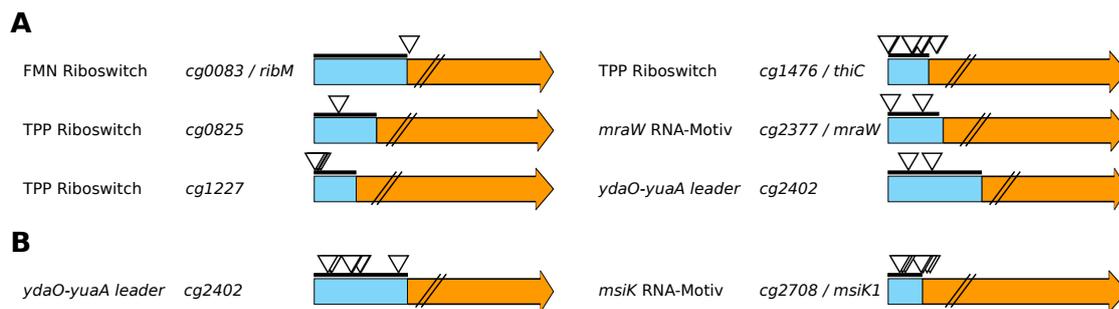


Abbildung 4.37: Identifizierte Schnittstellen der (A) RNase E/G und (B) RNase J in Riboswitches, RNA-Motiven und *leader*-Sequenzen. Die 5'-UTR ist in blau, der Bereich des Riboswitches, RNA-Motivs bzw. *leader*-Sequenz mit einem schwarzen Balken und die proteinkodierende Sequenz (CDS) in orange dargestellt. Identifizierte RNase-Schnittstellen sind mit einem auf der Spitze stehenden Dreieck angezeigt. Im Gegensatz zu den 5'-UTRs sind die Längen der CDS nicht maßstabsgetreu dargestellt.

Alle Gene für die mindestens eine Schnittstelle innerhalb der jeweiligen Riboswitches, RNA-Motive oder *leader*-Sequenzen identifiziert wurde, wurden in der Δrne -Mutante schwächer transkribiert detektiert (siehe LFC-Werte in Tabelle 4.16). Einzige Ausnahme ist der *ydaO-yuaA leader* von *cg2402*. Das Gen *cg2402* wurde in der Δrne -Mutante stärker transkribiert, als im Wildtyp. Das bedeutet, dass wenn RNase E/G durch die Deletion des entsprechenden Gens nicht mehr in der Zelle vorliegt, die Transkripte nicht mehr an den identifizierten Positionen geschnitten werden. Dennoch ist die Transkriptabundanz in der Δrne -Mutante im Vergleich

4 Ergebnisse

zum Wildtyp geringer. Folglich scheinen die durch RNase E/G eingebrachten Schnitte die Stabilität dieser Transkripte zu erhöhen.

RNase J Schnittstellen wurden nur im *ydaO-yuaA leader* von *cg2402* und im *msiK* RNA-Motiv von *cg2708 / msiK* gefunden (Tabelle 4.16 und Abbildung 4.37B). Obwohl in den regulatorischen Sequenzen dieser Transkripte Schnittstellen identifiziert wurden, sind sie in der Δ *rnj*-Mutante nicht differenziell transkribiert. Vermutlich ist RNase J nicht spezifisch an der Initiation der RNA-Degradation dieser Transkripte beteiligt. Regulatorische Elemente in 5'-UTRs, wie Riboswitches und *leader*-Sequenzen, weisen häufig Sekundärstrukturen auf (Serganov und Nudler, 2013). Es wird daher davon ausgegangen, dass die Hauptfunktion von RNase J in *C. glutamicum* die Degradation von Transkripten mit geringer Sekundärstruktur darstellt.

5 Diskussion

In dieser Arbeit wurden die beiden bakteriellen Organismen *Corynebacterium diphtheriae* und *Corynebacterium glutamicum* hinsichtlich der Prozesse der RNA-Synthese und des RNA-Abbaus mittels RNA-Sequenzierung untersucht. Konkret wurde die Transkriptionsinitiation über die 5'-Enden von Primärtranskripten im Humanpathogen *C. diphtheriae* und die Rho-abhängige Transkriptionstermination sowie die RNA-Degradation im biotechnologisch relevanten Modellorganismus *C. glutamicum* analysiert. In den folgenden drei Unterkapiteln werden die dabei erzielten Ergebnisse diskutiert.

5.1 Primärtranskriptsequenzierung als geeignete Methode zur Analyse der Transkriptionsinitiation

Um sowohl qualitative wie auch quantitative Analysen der transkriptionellen Landschaft von *C. diphtheriae* durchführen zu können, bedarf es einer Technologie mit Einzelnukleotidauflösung und hoher Genauigkeit über einen großen dynamischen Bereich. Die relativ neue Technologie der Hochdurchsatz-cDNA-Sequenzierung bzw. RNA-Sequenzierung (RNA-Seq) erfüllt alle Erfordernisse für qualitative als auch für quantitative Analysen (Wang *et al.*, 2009; Lu *et al.*, 2010; van Vliet, 2010).

Mit der Methode der Primärtranskriptsequenzierung (Pfeifer-Sancar *et al.*, 2013) wurden im Genom von *C. diphtheriae* 3.987 putative TSS durch die Software *ReadXplorer* (Hilker *et al.*, 2014) automatisch detektiert. Nach manueller Prüfung der automatisch detektierten TSS wurden 1.656 TSS für bekannte CDS und neue Transkripte identifiziert. Eine manuelle Prüfung der automatisch detektierten TSS ist nötig, da die Qualität der softwarebasierten TSS-Detektion von den Eigenschaften (z. B. Anteil an Sekundärstrukturen) der eingesetzten RNA abhängt.

Abhängig von den Eigenschaften der eingesetzten RNA werden nicht alle monophosphorylierten Transkripte von der Terminator-Exonuklease (TEN) abgebaut. Häufig ist die Read-Abdeckung in stark transkribierten Regionen auch in den Primärtranskriptdatensätzen erhöht (Tobias Busche, persönliche Mitteilung). Es wird daher vermutet, dass sehr abundante Transkripte nicht restlos abgebaut und somit ebenfalls mit einem RNA-Adapter ligiert werden.

Dies kann zu einem falsch-positiven Signal im Primärtranskript-Datensatz führen. Weiter ist bekannt, dass Transkripte mit starken Sekundärstrukturen von Exoribonukleasen nicht oder nur schlecht abgebaut werden (Spickler und Mackie, 2000). Es ist davon auszugehen, dass auch die TEN von starken Sekundärstrukturen behindert wird. Aus diesem Grund können Sekundärstrukturen innerhalb von nur teilweise durch die TEN abgebauten Transkripten ebenfalls zu falsch-positiven Signalen im Primärtranskriptdatensatz führen. Eine manuelle Prüfung der automatisch detektierten TSS ist somit häufig notwendig Pfeifer-Sancar *et al.* (2013); Irla *et al.* (2015); Wolf *et al.* (2017).

Ähnlich zu den Primärtranskriptanalysen in *C. glutamicum* (Pfeifer-Sancar *et al.*, 2013), *Bacillus methanolicus* (Irla *et al.*, 2015) und *Actinoplanes* (Wolf *et al.*, 2017) wurde in *C. diphtheriae* ein automatisch detektierter TSS als falsch-positiv klassifiziert, wenn keine klare Akkumulation von Readstarts an der betreffenden genomischen Position erkennbar war. Auch wurden automatisch detektierte TSS die innerhalb eines ungleichmäßigen Readstart-Gradienten lagen, ebenfalls als falsch-positive TSS klassifiziert, was besonders häufig innerhalb stark transkribierten Regionen der Fall war. Diese Beobachtung passt zu der oben genannten Theorie, dass stark abundante Transkripte nicht vollständig von der TEN abgebaut werden.

5.2 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

Obwohl das Genom von *Corynebacterium diphtheriae* NCTC 13129 schon im Jahr 2003 publiziert wurde (Cerdeño-Tárraga *et al.*, 2003) existieren nur einige Arbeiten über dessen transkriptionelle Organisation, Promotormotive und nichtkodierenden RNAs (ncRNAs) (Leong und Murphy, 1985; Schmitt, 1997; Oram *et al.*, 2006; Burgos und Schmitt, 2012).

Diese Arbeit beinhaltet die erste umfassende Analyse des Vollängentranskriptom sowie der Transkriptionsstarts (TSS) in *C. diphtheriae* NCTC 13129 mittels RNA-Sequenzierung. Für die meisten Gene wurden die TSS, Promotormotive und ribosomalen Bindestellen sowie 5'-UTRs identifiziert und analysiert. Darüber hinaus wurden anhand der experimentell identifizierten TSS 120 Translationsstarts (TLS) von proteinkodierenden Sequenzen (CDS) korrigiert und mehr als 400 putative neue Transkripte vorhergesagt.

5.2.1 Analyse der genomweit identifizierten Transkriptionsstarts

Insgesamt wurden 1.656 TSS in *C. diphtheriae* NCTC 13129 identifiziert. Von diesen wurden 1.202 TSS bekannten Genen und 454 TSS neuen Transkripten zugeordnet. Abhängig von der Position der TSS zu bekannten Genen wurden diese als intergenische (51 TSS), intragenische

5.2 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

(17 TSS) oder *antisense* TSS (386 TSS) klassifiziert. Diese TSS sind ein starker Indikator für neue, noch unbekannte Transkripte.

Intragenische TSS befinden sich innerhalb einer CDS und führen demnach zu verkürzten Transkripten. Es ist unklar, ob diese Transkripte für alternative verkürzte Proteine, neue proteinkodierende Regionen oder nichtkodierende RNA kodieren. Intragenische Transkripte scheinen nicht selten vorzukommen, so wurden sie bereits in verschiedenen Bakterien (Sharma *et al.*, 2010; Nicolas *et al.*, 2012; Wang *et al.*, 2013; Pfeifer-Sancar *et al.*, 2013; Irla *et al.*, 2015) identifiziert. Auch in Viren (Löchelt *et al.*, 1993) und im Menschen (Schoenfeld *et al.*, 1998) wurden verkürzte Genprodukte durch intragenische TSS nachgewiesen. Intragenische TSS könnten so die Zahl an möglichen kodierten Proteinen erhöhen während die Genomgröße (in Basenpaaren) konstant bleibt (Denoeud *et al.*, 2007).

In seltenen Fällen wurde für im Vollängendatensatz deutlich transkribierte Transkripte kein TSS im Primärtranskriptdatensatz identifiziert. Dies war beispielsweise bei den nichtkodierenden RNAs 4.5S RNA und 6C RNA der Fall. Beide RNAs sind unter 150 nt lang und besitzen starke Sekundärstrukturen mit schlecht zugänglichen 3'-Enden (Rosenblad *et al.*, 2009; Mentz *et al.*, 2013). Die während der Erstellung der Primärtranskriptbibliothek verwendete T4 RNA Ligase ist bei doppelsträngigen Transkriptenden weniger effizient, als bei einzelsträngigen Transkriptenden (Zhelkovsky und McReynolds, 2012). Möglicherweise verhindern die starken Sekundärstrukturen der 4.5S bzw. 6C RNA eine effiziente Ligation des RNA-Adapters, sodass letztlich kein TSS-Signal detektiert werden kann.

5.2.2 Nachweis neuer potenziell regulativer RNAs

Durch die Methode der RNA-Sequenzierung sind in bakteriellen Transkriptomen unerwartet viele neue RNAs identifiziert worden (Thomason und Storz, 2010; Georg und Hess, 2011). So wurden in gramnegativen und grampositiven Bakterien für 2 bis 30 % der annotierten Gene asRNAs nachgewiesen (Thomason *et al.*, 2015). In *C. diphtheriae* besitzen circa 15 % aller Gene eine asRNA. Die Zahl an detektierten asRNAs hängt stark von den untersuchten Bedingungen, der Erstellung der cDNA-Bibliothek und den für die Detektion gewählten Parametern ab (Thomason *et al.*, 2015). So wurden in unterschiedlichen Studien allein in *E. coli* zwischen hunderten und tausenden asRNAs detektiert (Dornenburg *et al.*, 2010; Shinhara *et al.*, 2011; Raghavan *et al.*, 2012; Li *et al.*, 2013). Die tatsächliche Anzahl an asRNAs in *C. diphtheriae* dürfte daher auch von den hier detektierten asRNAs abweichen und sollte daher als untere Grenze verstanden werden. Für die spezifische Analyse von kurzen RNAs, zu denen die meisten asRNAs gehören, sind besondere Protokolle zu verwenden, die zu einer besseren Repräsentation der asRNAs in der cDNA-Bibliothek führen (Wang *et al.*, 2009).

Transkripte die *antisense* zu einem Gen liegen, sind in den überlappenden Bereichen komplementär zueinander. Die komplementäre Bindung einer *antisense* RNA (asRNA) an ein *sense* Transkript kann unterschiedliche Auswirkungen auf die Stabilität des *sense* Transkripts haben (Thomason und Storz, 2010). So konnte für einige asRNAs gezeigt werden, dass diese die Stabilität des jeweiligen *sense* Transkripts erhöhen oder verringern (Saramago *et al.*, 2014). Eine Vorhersage der Funktion oder Auswirkung einer asRNA auf das *sense* Transkript ist nicht trivial und erfordert meist spezifische Experimente für einzelne asRNAs (Thomason und Storz, 2010). Die Funktionen der für *C. diphtheriae* über *antisense* TSS identifizierten asRNAs sind daher nicht bekannt. Es wird davon ausgegangen, dass die asRNAs zu den nichtkodierenden RNAs gehören.

Alternativ können Sequenzen in der Codon-Folge eines Transkripts vorhanden sein, die auf dem Gegenstrang einen Promotor bilden und so zu einem transkriptionellen „Hintergrundrauschen“ führen (Struhl, 2007). Die so gebildeten asRNAs sind in dem Fall ein Nebenprodukt und haben möglicherweise keine Funktion in der Zelle. Bei Bakterien mit hohem A+T-Gehalt treten -10 Motive häufiger auf als bei Bakterien mit geringem A+T-Gehalt (Lloréns-Rico *et al.*, 2016). Dies wird beispielsweise für *Staphylococcus aureus* vermutet, da dessen Genom einen vergleichsweise hohen A+T-Gehalt von 67,2 % aufweist (Mäder *et al.*, 2016).

In Bakterien mit geringerem A+T-Gehalt, wie *C. diphtheriae*, könnte das Auftreten von asRNAs ebenfalls auf transkriptionelles Hintergrundrauschen zurückzuführen sein. Für diese Bakterien ist aber nicht der A+T-Gehalt, sondern das Vorhandensein von „zufälligen“ Bindestellen von alternativen Sigmafaktoren ausschlaggebend. In *B. subtilis* wurde eine große Zahl an asRNAs nachgewiesen, die keinen Einfluss auf die *fitness* der Zelle haben (Nicolas *et al.*, 2012). *S. aureus* besitzt nur drei alternative Sigmafaktoren (Gertz *et al.*, 2000; Morikawa *et al.*, 2003; Miller *et al.*, 2012), während *B. subtilis* 15 alternative Sigmafaktoren besitzt (Helmann, 1995; Eichenberger *et al.*, 2004; Wang *et al.*, 2006; Helmann, 2002; Lane und Darst, 2006). Das Genom von *C. diphtheriae* besitzt einen geringen AT-Gehalt von 46,5 % und kodiert acht alternative Sigmafaktoren (Pátek und Nešvera, 2011). Es ist daher wahrscheinlich, dass die asRNAs ebenfalls durch Bindestellen der alternativen Sigmafaktoren entstehen.

5.2.3 Charakterisierung des σ^A Promotormotivs und von regulatorischen Elementen in 5'-UTRs

Analyse des σ^A Promotormotivs

Anhand der in *C. diphtheriae* identifizierten TSS wurde das σ^A Promotormotiv analysiert. Das -10 Motiv TAGaaT wurde für 1.190 (98,9 %) TSS detektiert. Das -10 Motiv sowie die Distanz zum TSS in *C. diphtheriae* stimmen mit dem in *C. glutamicum* identifizierten Motiv

5.2 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

sowie dem Abstand überein (Pátek *et al.*, 2003; Pátek und Nešvera, 2011; Pfeifer-Sancar *et al.*, 2013). Das schwach konservierte –35 Motiv (ttgcaa) wurde in *C. diphtheriae* nur bei Promotoren mit schwach konserviertem –10 Motiv detektiert. Diese Beobachtung trifft ebenfalls auf das –35 Motiv in *C. glutamicum* zu (Albersmeier *et al.*, 2017). Die hier präsentierten umfassenden TSS-Daten bilden den Grundstein für darauf aufbauende detaillierte Analysen der Promotormotive in *C. diphtheriae*, wie sie schon für *C. glutamicum* durchgeführt wurden (Albersmeier *et al.*, 2017).

Analyse der 5'-UTRs

Die in *C. diphtheriae* gemessenen 5'-UTR-Längen reichen von 0 nt bis 463 nt. Insgesamt wurden ca. 20 % als *leader*-freie Transkripte klassifiziert. *Leader*-freie Transkripte wurden unter anderem bereits in *H. pylori* (Sharma *et al.*, 2010), *S. coelicolor* (Passalacqua *et al.*, 2009) und *Pseudomonas putida* (Lawrence und Roth, 1996) nachgewiesen. Während der Anteil an *leader*-freien Transkripten in den meisten Bakterien gering ist, besitzen Actinobakterien im Allgemeinen einen erhöhten Anteil an *leader*-freien Transkripten (Zheng *et al.*, 2011). Im Speziellen weisen die Corynebakterien *C. diphtheriae* (Wittchen *et al.*, 2018), *C. glutamicum* (Pfeifer-Sancar *et al.*, 2013) und *C. glyciniphilum* (Al-Dilaimi, nicht publiziert) Anteile von 20 % bis 67 % *leader*-freie Transkripten auf.

Der Mechanismus der Translation von *leader*-freien Transkripten war lange unbekannt und ist auch heute nicht vollständig geklärt. In *E. coli* konnte gezeigt werden, dass das Startcodon AUG mit der Initiator-tRNA eine stabile Bindung mit dem Ribosom eingehen kann (Zheng *et al.*, 2011; Brock *et al.*, 2008). Vermutlich funktioniert die Translations-Initiation von *leader*-freien Transkripten anders, da *leader*-freie Transkripte eine Präferenz zu monomeren 70S Ribosomen gegenüber der 30S Untereinheit aufweisen (Moll *et al.*, 2004). Im Rahmen dieser Arbeit wurde GUG als zweithäufigstes Startcodon für *leader*-freie Transkripte in *C. diphtheriae* identifiziert. Es ist unklar, wie die Translations-Initiation mit diesem Startcodon abläuft. Da *leader*-freie Transkripte in einer Vielzahl unterschiedlicher Taxa gefunden wurden und heterologe *leader*-freie Transkripte aus Bakterien in Archaea und Eukaryoten translatiert werden konnten (Zheng *et al.*, 2011), wird vermutet, dass *leader*-freie Transkripte eine evolutionär ursprüngliche Form der Translation darstellen.

Transkripte mit einer 5'-UTR größer als 5 nt wurden hinsichtlich eines RBS-Motivs untersucht. Das identifizierte Motiv AGGag wurde in 87 % der analysierten Sequenzen gefunden. Der Abstand zum Startcodon beträgt im Durchschnitt 7,2 nt ($\pm 2,8$ nt). In *C. glutamicum* wurde ebenfalls das Motiv AGGag in einem vergleichbaren Abstand von 7,7 nt ($\pm 2,7$ nt) nachgewiesen (Pfeifer-Sancar *et al.*, 2013). Die beiden Spezies *C. diphtheriae* und *C. glutami-*

cum sind eng miteinander verwandt und besitzen an der 16S rRNA ein identisches 3'-Ende. Somit wurde ein vergleichbares RBS-Motiv erwartet.

Neben der RBS können 5'-UTRs regulatorische Elemente, wie Riboswitches, RNA-Motive oder *leader* Sequenzen enthalten. In den 5'-UTRs von *C. diphtheriae* wurden im Rahmen dieser Arbeit fünf Riboswitches, zwei RNA-Motive und ein RNA-Thermometer identifiziert. Im Vergleich mit *C. glutamicum* besitzt *C. diphtheriae* weniger regulatorische Elemente innerhalb von 5'-UTRs. Der Flavinmononukleotid-(FMN)- und der S-Adenosyl-Methionin-(SAM-IV)-Riboswitch sowie alle in *C. glutamicum* detektierten *leader* Sequenzen fehlen in *C. diphtheriae*. *C. glutamicum* ist ein Bodenbakterium welches sich wechselnden Umweltbedingungen und Nahrungsquellen anpassen muss (Kalinowski, 2005). *C. diphtheriae* als Humanpathogen hingegen ist hoch spezialisiert auf eine enge biologische Nische (Baron, 1996; Holmes, 2000). Es benötigt somit nur eine kleinere Genausstattung und somit ein kleineres Genom als *C. glutamicum*. Das Genom von *C. diphtheriae* umfasst ca. 2,4 Mbp (Cerdeño-Tárraga *et al.*, 2003), das von *C. glutamicum* etwa 3,3 Mbp (Kalinowski *et al.*, 2003). Das Genom von *C. diphtheriae* ist somit knapp ein Drittel kleiner, als das von *C. glutamicum*.

5.2.4 Identifizierung von Operon-Strukturen

Ein Operon ist ein Transkript, welches aus mehreren Genen besteht, die sich einen gemeinsamen Promotor teilen (Jacob und Monod, 1961). Insgesamt wurden für *C. diphtheriae* 471 primäre und 167 sekundäre Operons detektiert. Somit sind in *C. diphtheriae* 1.417 Gene (63 %) in Operons lokalisiert. Unter Berücksichtigung aller identifizierten TSS, primären Operons und Monocistrons wurden unter den vorliegenden Bedingungen (Komplexmedium, exponentielle Wachstumsphase) ca. 87 % der annotierten Gene als aktiv transkribiert detektiert. Die übrigen Gene wurden nicht detektiert, da sie unter den gewählten Bedingungen entweder zu gering für die Operon-Detektion oder gar nicht transkribiert vorliegen.

Für *C. glutamicum* wurden ebenfalls Operonstrukturen detektiert. *C. glutamicum* besitzt 616 primäre Operons bestehend aus circa 2.000 Genen (Pfeifer-Sancar *et al.*, 2013). Im Vergleich zu *C. glutamicum* besitzt *C. diphtheriae* kürzere und weniger Operons. Dies ist darauf zurückzuführen, dass *C. diphtheriae* um circa ein Drittel kürzeres Genom aufweist als *C. glutamicum* und somit die Zahl an Operons sowie deren Länge entsprechend geringer ausfallen. Ein Großteil der Operonstrukturen in *C. diphtheriae* ist ähnlich (syntän) zu denen in *C. glutamicum*.

Die experimentell detektierten Operons wurden mit den *in silico* vorhergesagten Operons aus der DOOR-Datenbank (*Database for prokaryotic Operons*) (Mao *et al.*, 2013) verglichen. Ein Großteil (89 %) der in DOOR vorhergesagten Operons, stimmt mit den über

5.2 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

RNA-Sequenzierung nachgewiesenen primären Operons überein. Die fehlenden 11 % wurden aufgrund der geringen Read-Abdeckung in Bereichen mit geringer Transkription nicht evaluiert. Die in dieser Arbeit detektierten Operonstrukturen bestätigen somit den überwiegenden Teil der *in silico* Vorhersagen von DOOR. Ein besonderer Vorteil der experimentellen Operon-Detektion mittels Primärtranskriptdaten ist die zusätzliche Detailanalyse die zum Auffinden von *in silico* nicht vorhersagbaren sekundären Operons genutzt wurde. So wurden über die hier präsentierten Primärtranskriptdaten 167 sekundäre Operons detektiert, die insgesamt 316 Gene umfassen.

Sekundäre Operons wurden bereits über die Analyse von Primärtranskripten in verschiedenen Bakterien, wie *Helicobacter pylori* (Sharma *et al.*, 2010), *B. methanolicus* (Irla *et al.*, 2015) und *Actinoplanes* (Wolf *et al.*, 2017) nachgewiesen. Die Unterteilung von primären Operons in kürzere sekundäre Operons ermöglichen eine feinere Regulation der Genexpression (Sorek und Cossart, 2010; Qiu *et al.*, 2010). Darüber hinaus sind manche Operonstrukturen nicht statisch, sondern dynamisch. So wurde in *Mycoplasma pneumoniae* gezeigt, dass ein Gen unter einer Bedingung in einem Operon, unter einer anderen Bedingung allerdings als Monocistron transkribiert wurde (Koide *et al.*, 2009; Güell *et al.*, 2009). Demnach sind Operonstrukturen flexibel und werden von den vorherrschenden Bedingungen bestimmt (Sorek und Cossart, 2010).

5.2.5 Analyse des DtxR-Regulons in *C. diphtheriae*

Insgesamt wurden in der $\Delta dtxR$ -Mutante 337 Gene als differenziell transkribiert identifiziert. Allerdings ist nur für 40 Gene eine DtxR-Bindestelle bekannt oder vorhergesagt worden (Kunkle und Schmitt, 2003). Diese Differenz zwischen Anzahl an Genen mit DtxR-Bindestelle und differenziell transkribierten Genen ist durch indirekte Effekte der *dtxR*-Deletion zu erklären. Direkte Effekte der Deletion von *dtxR* sind die differenzielle Transkription der Gene mit DtxR-Bindestelle. Diese Gene werden im Wildtyp unter Verfügbarkeit von Eisen durch DtxR reprimiert. Bei den im Wildtyp reprimierten Genen handelt es sich überwiegend um Gene der Eisenfreisetzung. Unter den in der $\Delta dtxR$ -Mutante aktivierten Genen befinden sich Gene, dessen Produkte in der Speicherung von Eisen (z. B. Ferritin) involviert sind. Die Transkriptionsaktivierung und -reprimierung durch DtxR wurde bereits in *C. glutamicum* beobachtet (Brune *et al.*, 2006), der Mechanismus ist allerdings nicht aufgeklärt. Die Ergebnisse der differenziellen Transkriptionsanalyse in *C. diphtheriae* bestätigen die Eigenschaft von DtxR als dualen Transkriptionsregulator.

Die DtxR-Bindestelle befindet sich bei allen in *C. diphtheriae* durch DtxR reprimierten Genen im Promotor oder *downstream* der TSS. Die Repression erfolgt daher wahrscheinlich über die

einfache Blockade des Promotors, sodass die RNA-Polymerase (RNAP) sterisch gehindert wird und nicht an die DNA binden kann oder über eine Hinderung von bereits an den Promotor gebundener RNAP (Browning und Busby, 2016).

Der Mechanismus der Transkriptionsaktivierung durch DtxR ist nicht bekannt. Die Abstände zwischen DtxR-Bindestelle und TSS betragen in *C. diphtheriae* 37 bp (*ftn*) bis 189 bp (*narK*). Die Transkriptionsaktivierung bei geringem Abstand von DtxR-Bindestelle zum Promotor (Gen *ftn*) ist als Klasse II Aktivierung denkbar. Klasse II Aktivatoren binden in der Nähe des -35 Promotormotivs und aktivieren die Transkription über Bindung der α -Untereinheiten der RNAP (Busby und Savery, 2007).

Der Mechanismus der Transkriptionsaktivierung über Distanzen von bis zu 189 bp (Gen *narK*) kann nicht über eine Klasse II Aktivierung erfolgen, da die Abstände zwischen DtxR-Bindestelle und Promotor zu groß sind. Bei σ^{54} Promotoren ist zwar eine Aktivierung über größere Distanzen beschrieben worden (Busby und Savery, 2007), allerdings besitzt *C. diphtheriae* keinen σ^{54} (Pátek und Nešvera, 2011). Auch ist für die Aktivierung von σ^{54} Promotoren ein Verstärkerprotein mit ATP-Motorfunktion nötig um die RNA-Polymerase fortzubewegen (Busby und Savery, 2007). DtxR besitzt allerdings keine ATP-Motorfunktion (Pohl *et al.*, 1998). Außerdem wurde in dieser Arbeit für die in *C. diphtheriae* durch DtxR aktivierten Gene eindeutig ein Promotormotiv des *house keeping* Sigmafaktors σ^A identifiziert. Die Transkriptionsaktivierung dieser Gene muss demnach über einen anderen, noch unbekanntem, Mechanismus ablaufen.

5.2.6 Analyse der Phagen-Insel mit Fokus auf dem Diphtherietoxin-kodierenden Gen *tox*

Die transkriptomweiten Datensätze über Primär- und Vollängentranskripte bieten die Möglichkeit die Phagen-Insel (*phage island*) sowie das Gen *tox*, welches für das Diphtherietoxin kodiert, näher zu betrachten. Das *tox* Gen ist ein wichtiger Pathogenitätsfaktor von *C. diphtheriae* (Baron, 1996; Freeman, 1951).

Das Gen *tox* ist Teil des Corynephagen (Cerdeño-Tárraga *et al.*, 2003), besitzt aber als einziges Gen der Phagen-Insel eine DtxR-Bindestelle (Tao *et al.*, 1994). Die Transkription von *tox* ist somit unter bakterieller Kontrolle, da das DtxR kodierende Gen *dtxR* im bakteriellen Genom kodiert ist (Cerdeño-Tárraga *et al.*, 2003; Oram *et al.*, 2006). Die Transkription von *tox* wird unter hohen Eisenkonzentrationen von DtxR reprimiert. Eisen, in Form von Fe^{2+} , ist ein Spurenelement und daher auch für *C. diphtheriae* essenziell. Eine Strategie von eukaryotischen Wirtsorganismen zur Abwehr von mikrobiellen Infektionen, wie durch *C. diphtheriae*, ist die Verringerung der extrazellulären Eisenkonzentration (Andrews *et al.*, 2003; Kehl-Fie

5.2 Analyse der Transkriptionsinitiation in *Corynebacterium diphtheriae*

und Skaar, 2010). Die Repression von *tox* durch DtxR kann somit als Gegenmaßnahme von *C. diphtheriae* gegen die Eisenverknappung durch den Wirt angesehen werden. Liegt kein Eisenmangel vor, wird auch kein Diphtherietoxin benötigt. Somit kann die Regulation der Expression von *tox* aus evolutionärer Sicht einen Vorteil darstellen. Corynebakterien, die eine DtxR-Bindestelle vor dem *tox* Gen besitzen, könnten ihrem Wirt (hier *C. diphtheriae*) so einen Vorteil gegenüber Wirten mit unregulierter *tox* Expression aufweisen (Merchant und Spatafora, 2014). Eine vergleichbare Regulation wurde für Shiga-Toxin-produzierende pathogene *E. coli* gezeigt. Das Shiga-Toxin-Gen ist ebenfalls Phagen-kodiert, die Expression ist allerdings über den Eisensensitiven Fur-Regulator, dessen Gen *fur* im bakteriellen Chromosom vorliegt, reguliert (Brüssow *et al.*, 2004).

Wie oben bereits erwähnt ist das *tox* Gen und dessen Genprodukt, das Diphtherietoxin (DT), für die Pathogenität von *C. diphtheriae* wichtig. Aus diesem Grund wurden schon relativ früh Studien zur Struktur und einzelner Domänen von DT durchgeführt (Kantardjieff *et al.*, 1989; Choe *et al.*, 1992; Louie *et al.*, 1997; Li *et al.*, 2014). Nur wenige Studien haben die Transkription des Gens analysiert (Leong und Murphy, 1985; Boyd und Murphy, 1988). Im Rahmen dieser Arbeit wurde daher eine umfassende Analyse der transkriptionellen Landschaft des *tox* Gens durchgeführt.

Für *tox* wurden die zwei bisher unbekannte asRNAs identifiziert. *Upstream* des *tox* Gens wurden zwei sich gegenüberliegende intergenische Transkripte identifiziert. Diese Transkripte wurden bisher nicht in der Literatur erwähnt. Über die Funktion der asRNAs zum *tox* Gen kann nur spekuliert werden, da asRNAs positive als auch negative Auswirkungen auf das *sense* Transkript haben können (Saramago *et al.*, 2014; Thomason und Storz, 2010). Das intergenische Transkript auf dem Gegenstrang ist stärker transkribiert als das intergenische Transkript auf dem Leitstrang. Es wird daher vermutet, dass das Transkript auf dem Leitstrang die Transkriptmenge über komplementäre Bindung des Transkripts auf dem Gegenstrang reguliert. Das *sense* Transkript könnte somit eine regulatorische asRNA für das Transkript auf dem Gegenstrang darstellen. Die beiden intergenischen Transkripte besitzen keine ORFs oder Homologien zu bekannten ncRNAs. Der TSS des intergenischen Transkripts auf dem Gegenstrang (inragenischer TSS 2 in Abbildung 4.6B) befindet sich 55 nt *upstream* vom TSS 1 von *tox*, welches auf dem Vorwärtsstrang kodiert ist. Aufgrund des geringen Abstands überlagern sich die -35 Motive beider Promotoren, sodass es zu sterischen Hinderungen zwischen RNA-Polymerasen auf beiden Strängen kommen könnte. Möglicherweise bildet das intergenische Transkript mit seinem Promotor ein zusätzliches regulatorisches Element der Expression von *tox*. Die neuen intergenischen Transkripte, als auch die *tox* asRNAs sind besonders für die medizinische Forschung relevant und sollten in zukünftigen Studien untersucht werden.

5.3 Rho-abhängige Transkripttermination in *C. glutamicum*

In *C. glutamicum* wurde die Funktion des Terminationsfaktors Rho über differenzielle Transkription zwischen einer $\Delta\rho$ -Mutante und dem *C. glutamicum* Wildtyp analysiert. Darüber hinaus wurden die Rho-abhängig terminierten Transkripte über die Transkriptverlängerung am 3'-Ende in der $\Delta\rho$ -Mutante identifiziert und charakterisiert. Außerdem wurde das in der $\Delta\rho$ -Mutante aktiv transkribierte *tad*-Cluster, welches für einen putativen Typ IV Pilus kodiert, genauer untersucht.

5.3.1 Vergleich der Struktur des Rho-Faktors aus *C. glutamicum*

Die grundlegende Struktur von *C. glutamicum* Rho wurde im Rahmen dieser Arbeit mit den Homologen aus *E. coli*, *B. subtilis* und *M. tuberculosis* verglichen. *C. glutamicum* Rho (Cg.Rho) besitzt wie Ec.Rho ebenfalls eine NTD sowie eine CTD, allerdings befindet sich in der NTD eine Insertion von 144 Aminosäuren, die weder in *E. coli* noch in *B. subtilis* vorkommt. Aus diesem Grund ist das Rho-Protein in *C. glutamicum* circa 45 % größer als in *E. coli* oder *B. subtilis* (Abbildung 4.8A). Das Rho-Protein in *M. tuberculosis* besitzt ebenfalls eine Insertion in der NTD und ist daher auch größer als das *E. coli* bzw. *B. subtilis* Homolog (D'Heygère *et al.*, 2015). Allerdings ist die Insertion in *M. tuberculosis* kürzer als in *C. glutamicum*. Diese Insertion in der NTD ist nicht auf Cg.Rho und Mt.Rho beschränkt, sondern ist eine typische Eigenschaft der Rho-Faktoren in Actinobakterien als auch der Bacteroidetes (D'Heygère *et al.*, 2013). Generell schwankt die Größe des Rho-Proteins stark von Phylum zu Phylum (D'Heygère *et al.*, 2013). Das größte bisher bekannte Rho-Protein mit 865 Aminosäuren wurde in *Thermaerobacter marianensis* (Firmicutes) gefunden (D'Heygère *et al.*, 2013). Eine Korrelation zwischen der Größe des Rho-Proteins und einer spezifischen Eigenschaft des entsprechenden Genoms (wie Anzahl an Genen, G+C-Gehalt oder Genomgröße) konnte nicht gefunden werden (D'Heygère *et al.*, 2013). Die kleinsten bisher gefundenen Rho-Proteine bestehen aus circa 300 Aminosäuren, denen wichtige Domänen fehlen. Bakterien in denen dieses verkürzte Rho-Protein gefunden wurde, besitzen ein weiteres Gen welches für Rho mit allen Domänen kodiert (D'Heygère *et al.*, 2013).

Für Mt.Rho konnte gezeigt werden, dass die NTD-Insertion die Affinität zu *rut sites* auf der RNA im Vergleich zu Ec.Rho erhöht und gleichzeitig die Motoraktivität von Mt.Rho verringert (D'Heygère *et al.*, 2015). Die genaue Funktion der NTD-Insertion von Cg.Rho ist unbekannt. Aufgrund der hohen Ähnlichkeit zu Mt.Rho wird vermutet, dass sie eine vergleichbare Funktion wie die von Mt.Rho besitzt. Ein verändertes Rho-Protein in *M. tuberculosis*, in dessen Gensequenz ein Threonin durch ein Alanin im R-Loop (Position 501 in der Proteinsequenz) ersetzt wurde, konnte besser kurze RNA-Fragmente terminieren und

5.3 Rho-abhängige Transkripttermination in *C. glutamicum*

wies eine geringere Aktivität als Mt.Rho aus dem Wildtyp auf (D'Heygère *et al.*, 2015). Der Rho-Faktor in *C. glutamicum* besitzt nativ ein Alanin anstelle eines Threonin im R-Loop (aufgrund der N-terminalen Insertion Position 655 in der Proteinsequenz) (Abbildung 4.8B). Die Erkenntnisse über das modifizierte Mt.Rho könnten daher auf Cg.Rho übertragbar sein.

Das für die Erkennung der C-reichen RNA-Abschnitte wichtige 5'-YC-Dimer-Motiv in Mt.Rho besitzt sechs weitere Aminosäurereste als Ec.Rho, was sich auf die Erkennung von *rut sites* auswirken könnte (D'Heygère *et al.*, 2015). Verglichen mit Ec.Rho besitzt Cg.Rho im 5'-YC-Dimer-Motiv eine Insertion aus zehn Aminosäuren (Abbildung 4.8B). Die Insertion im 5'-YC-Dimer in Cg.Rho ist somit nochmals größer als in Mt.Rho, was sich möglicherweise stärker auf die Erkennung von *rut sites* auswirken könnte.

5.3.2 Rho reprimiert die Transkription von potenzieller Fremd-DNA

Der Rho-Faktor ist bisher hauptsächlich in *E. coli* untersucht worden (Cardinale *et al.*, 2008; Peters *et al.*, 2012; Leela *et al.*, 2013; Ray-Soni *et al.*, 2016). Aufgrund der großen strukturellen Unterschiede sind Studien zu Rho-Proteinen aus anderen Bakterien wichtig, um Unterschiede und Gemeinsamkeiten ausfindig zu machen.

Alle Cyanobakterien und Tenericutes besitzen kein Rho-Homolog. Auch einige Vertreter der Klassen Clostridia, Bacilli und Negativicutes innerhalb der Abteilung der Firmicutes besitzen kein Rho-Homolog (D'Heygère *et al.*, 2013). Für *E. coli* als auch für *M. tuberculosis* ist der Rho-Faktor essenziell (Das *et al.*, 1978; Botella *et al.*, 2017). *B. subtilis* und *S. aureus* besitzen einen Rho-Faktor, dieser ist jedoch verzichtbar (Quirk *et al.*, 1993; Washburn *et al.*, 2001). Wie auch in *E. coli* ist der Rho-Faktor in *B. subtilis* und *S. aureus* für Transkriptionstermination Rho-abhängig terminierter Transkripte involviert (de Hoon *et al.*, 2005; Mäder *et al.*, 2016). In *C. glutamicum* ist eine Deletion des für Rho kodierenden Gens *rho* möglich, wie in dieser Arbeit gezeigt wurde. Über die Deletion von *rho* können die Auswirkungen des Rho-Faktors auf das Transkriptom und möglicherweise direkt durch den Rho-Faktor beeinflusste Gene analysiert werden. Somit ist *C. glutamicum* ein idealer Organismus um Rho-Proteine mit großer NTD-Insertion untersuchen zu können.

In der *C. glutamicum* Δrho -Mutante sind auffällig viele Gene der genomisch integrierten Phagen CGP2 und CGP3 (jeweils 40 – 50 % der Gene) (Abbildung 4.12) sowie der überwiegende Teil (66 %) der IS-Elemente (Abbildung 4.13) verstärkt transkribiert. Diese Gene stammen aus potenziellen Fremdquellen bzw. ermöglichen die Aufnahme von potenziell fremder DNA. In *E. coli* konnte ebenfalls gezeigt werden, dass bei Inhibierung des Rho-Faktors mit Bicyclomycin (BCM) Prophagen-Gene verstärkt transkribiert werden (Cardinale *et al.*, 2008; Menouni *et al.*, 2013). Auch in *M. tuberculosis* wurden bei einem *knock down* von

rho Gene aus Fremdquellen verstärkt transkribiert (Botella *et al.*, 2017). Trotz sehr unterschiedlicher Strukturen des Rho-Faktors in den verschiedenen Phyla besitzt der Rho-Faktor neben der allgemeinen Transkripttermination offensichtlich eine wichtige Funktion bei der transkriptionellen Stilllegung (*silencing*) von Fremd-DNA (Ray-Soni *et al.*, 2016).

5.3.3 Die Deletion von *rho* führt zu verlängerten Transkripten

Der Rho-Faktor kann über BCM in *E. coli* spezifisch inhibiert werden (Kohn und Widger, 2005). BCM bindet in *Ec.Rho* an die ATPase-Tasche und blockiert die Bindung eines nukleophilen Wassermoleküls, welches für die Hydrolyse von ATP benötigt wird (Kohn und Widger, 2005). Über die spezifische Inhibierung wurden mehr als 1.250 putative Rho-Bindestellen in *E. coli* bestimmt (Peters *et al.*, 2012). Allerdings ist nicht jedes Rho-Homolog mit BCM inhibierbar. So ist BCM in grampositiven Bakterien kaum effektiv. Von allen getesteten grampositiven Bakterien zeigt nur *Micrococcus luteus* eine BCM-Sensitivität (Williams und Durham, 1988; Nowatzke *et al.*, 1997).

Obwohl der Rho-Faktor für *C. glutamicum* nicht essenziell ist, ist die $\Delta\rho$ -Mutante stark im Wachstum verlangsamt (Abbildung 4.9). Über die differenzielle Transkriptionsanalyse der $\Delta\rho$ -Mutante mit dem Wildtyp, wurden der Einfluss der Deletion von *rho* auf das Transkriptom in *C. glutamicum* untersucht. Insgesamt sind in der $\Delta\rho$ -Mutante 488 Gene stärker und 267 Gene schwächer transkribiert als im Wildtyp (Abbildung 4.10). Besonders viele Gene die in der Replikation und DNA-Reparatur (COG-Klasse L) sowie in der Transkription involviert sind (COG-Klasse K), sind in der $\Delta\rho$ -Mutante verstärkt transkribiert (Abbildung 4.11). Im Gegensatz dazu sind auffällig viele Gene die in der Translation und Ribosomenstruktur involviert sind, schwächer transkribiert (COG-Klasse J). Da besonders die Expression von ribosomalen Proteinen an das Zellwachstum gekoppelt sind (Jörn Kalinowski, persönliche Mitteilung), ist die verringerte Transkription der Gene der COG-Klasse J wahrscheinlich nicht auf die Deletion von *rho* zurückzuführen. Wie in *E. coli*, *S. aureus* und *M. tuberculosis* (Cardinale *et al.*, 2008; Menouni *et al.*, 2013; Botella *et al.*, 2017), sind in der *C. glutamicum* $\Delta\rho$ -Mutante viele Gene die aus potenziellen Fremdquellen stammen verstärkt transkribiert. Anhand der Analyse der differenziell transkribierten Gene ist nicht erkennbar, ob diese Gene direkt oder indirekt von der Deletion von *rho* beeinflusst wurden.

Der Rho-Faktor vermittelt die Termination von Rho-abhängig terminierten Transkripten (Alifano *et al.*, 1991). Die Deletion des Rho kodierenden Gens *rho* hat zur Folge, dass Rho-abhängig terminierte Gene in der $\Delta\rho$ -Mutante nicht mehr korrekt terminiert werden. Die RNA-Polymerase fährt mit der Transkription am eigentlichen Genende fort, was zu am 3'-Ende

5.3 Rho-abhängige Transkripttermination in *C. glutamicum*

verlängerten Transkripten führt. Diese verlängerten Transkripte sind ein direkter Effekt der Deletion des *rho* Gens und damit Rho-abhängig terminiert.

5.3.4 Identifikation der in der $\Delta\rho$ -Mutante verlängerten Transkripte

Mit Hilfe eines eigens für die Identifizierung von verlängerten Transkripten entwickelten Programms wurden insgesamt 923 verlängerte Transkripte im *C. glutamicum* Transkriptom identifiziert. Die Parameter zur Identifikation wurden konservativ gewählt, um die Fehlerrate von als falsch-positiv identifizierten verlängerten Transkripten in Grenzfällen zu reduzieren. Die 923 als verlängert identifizierten Transkripte bilden somit die untere Grenze der tatsächlich unter diesen Bedingungen vorliegenden verlängerten Transkripten.

Grenzfälle für die Identifizierung von verlängerten Transkripten sind genomische Regionen mit eng benachbarten oder überlappenden Genen. Der Abstand der Gene ist in diesen Fällen zu gering für eine auf der Read-Abdeckung basierenden Identifikation von verlängerten Transkripten. In diesen Bereichen kann nicht unterschieden werden, ob die Reads vom *upstream* oder vom *downstream* Gen (bzw. dessen Transkript) stammen. Dieser Umstand geht auf die Fragmentierung der RNA vor der Sequenzierung zurück. Die Fragmentierung ist ein notwendiger Schritt bei der Illumina-Sequenzierung (Illumina, 2017).

Eine Alternative zur RNA-Sequenzierung mit Illumina ist die jüngst entwickelte Technik der direkten RNA-Sequenzierung (*Direct RNA Sequencing*) mit Nanoporen (Oxford Nanopore). Diese Sequenzieretechnik erlaubt die direkte Sequenzierung von Transkripten in ihrer originalen Länge ohne Fragmentierung oder PCR-Amplifikation (Jenjaroenpun *et al.*, 2018). Die Technik bietet allerdings eine im Vergleich zur Illumina RNA-Sequenzierung wesentlich geringere Read-Abdeckung und somit geringeren dynamischen Bereich über das gesamte Transkriptom (Christian Rückert, persönliche Mitteilung). Neben anderen Nachteilen der Nanopore-Sequenzierung, können gegenwärtig nur mRNAs mit Poly(A)-Schwanz sequenziert werden (Oxford Nanopore, 2017). Dieser Umstand schließt die Sequenzierung von (unmodifizierten) prokaryotischen Transkriptomen aus, da ein Poly(A)-Schwanz in Prokaryoten die Markierung für die Transkript-Degradation darstellt (Arraiano *et al.*, 2010).

Sollte die direkte RNA-Sequenzierung von prokaryotischer RNA in Zukunft möglich sein, wäre diese Technik geeignet, um verlängerte Transkripte in der *C. glutamicum* $\Delta\rho$ -Mutante auch für die oben genannten Grenzfälle eindeutig nachweisen können.

5.3.5 Rho terminiert *antisense* RNAs

Von den als verlängert identifizierten Transkripten machen *antisense* RNAs (asRNAs) mit 67 % den überwiegenden Anteil aus. Nur 27 % der verlängerten Transkripte sind proteinkodierende

5 Diskussion

Transkripte. Demnach ist eine der Hauptaufgaben von Rho in *C. glutamicum* die Termination von asRNAs. Diese Funktion des Rho-Faktors ist nicht auf Corynebakterien beschränkt, da die Termination von asRNAs auch in *E. coli* (Peters *et al.*, 2012) und *S. aureus* (Mäder *et al.*, 2016) gezeigt wurde.

Antisense RNAs können die Transkriptabundanz und -stabilität des *sense* Transkripts positiv als auch negativ beeinflussen (Thomason und Storz, 2010; Saramago *et al.*, 2014). Die Deletion des für den Rho-Faktor kodierenden Gens *rho* führte in *C. glutamicum* zu einer Verlängerung der asRNA-Transkripte. Allerdings sind nur circa ein Viertel der Transkripte, die asRNAs auf dem Gegenstrang besitzen, differenziell transkribiert (Abbildung 4.20). Es ist daher wahrscheinlich, dass ein Großteil der asRNAs in *C. glutamicum* keine regulative Funktion für das *sense* Transkript besitzen. Möglicherweise entstehen diese funktionslosen asRNAs aufgrund von Promotor-ähnlichen Sequenzen die durch die Codonfolge der CDS auf dem Leitstrang hervorgerufen werden und bilden eine Art transkriptionelles Hintergrundrauschen (Struhl, 2007). Wenn eine Veränderung der Codonfolge im Gen und somit der Proteinsequenz, zu einem Fitness-Verlust führt, könnte das Beibehalten eines asRNA-Promotors eine energetisch akzeptable Einschränkung für die Zelle darstellen. In so einem Fall wäre die Funktion des Transkriptionsterminators Rho das Stilllegen von Promotoren funktionsloser asRNAs. Diese Funktion des Rho-Faktors wird im Pathogen *S. aureus* vermutet, da in diesem Bakterium nur sehr wenige asRNAs identifiziert wurden (Mäder *et al.*, 2016). Es wäre denkbar, dass Rho in *C. glutamicum* eine identische Funktion aufweist.

In *C. glutamicum* wurden bisher zwei asRNAs charakterisiert. Die erste in *C. glutamicum* entdeckte asRNA ArnA befindet sich auf den Gegenstrang von *cg1935 / gntR2*, welches für einen Transkriptionsregulator der GntR-Familie kodiert (Zemanová *et al.*, 2008). Die Funktion von ArnA ist unbekannt, es wird vermutet, dass sie die Expression von *gntR2* reguliert, da ArnA unter Hitzestress verstärkt exprimiert wird (Zemanová *et al.*, 2008). In der $\Delta\rho$ -Mutante und im Wildtyp wurde ArnA nicht nachgewiesen, da die Read-Abdeckung auf dem Gegenstrang zu *gntR2* dem Hintergrundrauschen entspricht. Die zweite in *C. glutamicum* beschriebene asRNA ist AsgR (Maeda *et al.*, 2018). AsgR befindet sich auf dem Gegenstrang eines intrinsisch terminierten Gens (Pfeifer-Sancar *et al.*, 2013). In der $\Delta\rho$ -Mutante wurde kein verlängertes Transkript von AsgR nachgewiesen, was die Rho-unabhängige Termination von AsgR bestätigt.

Für die in der $\Delta\rho$ -Mutante identifizierten asRNAs wurden Transkriptlängen bis über 4.000 nt nachgewiesen (Abbildung 4.17A). Es ist unklar, ob diese langen asRNAs *in vivo* tatsächlich RNA-Duplexe mit den jeweiligen *sense* Transkripten bilden können. Es wäre denkbar, dass diese langen Transkripte komplexe interne Sekundärstrukturen bilden, die eine komplementäre Bindung von *sense* Transkripten verhindern. Translatierende Ribosomen können Sekundärstrukturen in der mRNA auflösen (Xie und Chen, 2017) und sollten daher ebenfalls

5.3 Rho-abhängige Transkripttermination in *C. glutamicum*

mögliche RNA-Duplexe aus *sense* mRNA und asRNA aufbrechen können, vorausgesetzt, die Duplexe blockieren nicht die RBS.

Auch für proteinkodierende Transkripte wurden Verlängerungen am 3'-Ende identifiziert, die ebenfalls über mehrere tausend Nukleotide lang sind (Abbildung 4.17B). Die verlängerten mRNAs umfassen keine *downstream* befindlichen Gene, sodass keine neuen Operons in der $\Delta\rho$ -Mutante entstehen (Daten nicht gezeigt).

Eine Möglichkeit zum Nachweis der RNA-Duplexe funktioniert über den Einbau von Substanzen wie Thionukleosiden in Transkripte, die unter UV-Bestrahlung RNA-Duplexe quervernetzen (Harris und Christian, 2009). Auf diese Weise könnte der Umfang der RNA-Duplex-Bildung durch verlängerte Transkripte in der $\Delta\rho$ -Mutante in zukünftigen Untersuchungen bestimmt werden.

5.3.6 Rho ist an der Regulation einiger Riboswitches beteiligt

Für den Flavinmononukleotid-(FMN)-Riboswitch in *ribM* wurde eine Rho-abhängige Termination in *C. glutamicum* R nachgewiesen (Takemoto *et al.*, 2015). Auch in der in dieser Arbeit erstellten *C. glutamicum* ATCC 13032 $\Delta\rho$ -Mutante wurde eine Rho-abhängige Termination des FMN-Riboswitches in *cg0083 / ribM* beobachtet. Unter normalen Kulturbedingungen liegt ausreichend FMN vor, sodass nur der FMN-Riboswitch in der 5'-UTR von *ribM*, nicht aber die CDS, transkribiert wird (Abbildung 4.21). Im Gegensatz dazu findet in der $\Delta\rho$ -Mutante eine deutliche Transkription der 5'-UTR und der CDS statt.

Das Transkript von *cg1478*, welches den SAM-IV-Riboswitch enthält, ist in der $\Delta\rho$ -Mutante ebenfalls verlängert (Tabelle 4.9). Außerdem ist *cg1478* das in der $\Delta\rho$ -Mutante am stärksten differenziell transkribierte Gen (LFC +6,6). Somit ist es sehr wahrscheinlich, dass neben dem FMN-Riboswitch auch der SAM-IV-Riboswitch durch Rho terminiert und somit reguliert wird.

Für *E. coli* wurde gezeigt, dass der TPP-Riboswitch in *thiM* ebenfalls durch Rho terminiert und so die Transkription von *thiM* reguliert wird (Bastet *et al.*, 2017). In der *C. glutamicum* $\Delta\rho$ -Mutante ist *thiM* zwar stärker transkribiert, eine Transkriptverlängerung wurde nicht beobachtet. Allerdings wurde auch kein Haarnadel-Terminator für dieses Gen nachgewiesen (Pfeifer-Sancar *et al.*, 2013). Für den TPP-Riboswitch in *thiM* ist daher eher von einer indirekten Regulation durch Rho auszugehen.

5.3.7 Motivanalyse von möglichen *rut sites* von mRNAs

Für *E. coli* konnte gezeigt werden, dass der Rho-Faktor C-reiche und G-arme Sequenzen, sogenannte CG-Blasen (*CG-Bubbles*), auf der RNA erkennt und anschließend die Transkriptionstermination einleitet (Alifano *et al.*, 1991; Walmacq *et al.*, 2004). Aus diesem Grund

5 Diskussion

gelten CG-Blasen als Indikatoren für Rho-abhängige Terminatorstrukturen und werden daher auch *Rho utilisation sites* (*rut sites*) genannt (Menouni *et al.*, 2013; Peters *et al.*, 2012). Eine distinkte Sequenzabfolge mit einem Erkennungsmotiv ist nicht bekannt. Daher sind bioinformatische Vorhersagen Rho-abhängiger Terminatoren im Gegensatz zu Rho-unabhängigen (intrinsischen) Terminatoren nicht möglich (Hart und Roberts, 1991; Ray-Soni *et al.*, 2016).

Es stellt sich die Frage, wie der Rho-Faktor in Bakterien mit hohem G+C-Gehalt zwischen zu terminierenden und nicht zu terminierenden Transkripten unterscheidet, da CG-Blasen häufiger auftreten sollten, als in Bakterien mit geringem G+C-Gehalt. Die Abteilung der Actinobakterien umfasst viele Bakterien die einen hohen bis sehr hohen G+C-Gehalt besitzen (Ventura *et al.*, 2007). Daneben besitzen viele Actinobakterien eine NTD-Insertion im Rho-Faktor (D'Heygère *et al.*, 2013), für die zumindest in *M. tuberculosis* eine RNA-bindende Eigenschaft nachgewiesen wurde (Mitra *et al.*, 2014; D'Heygère *et al.*, 2015). Zumindest für Actinobakterien besteht eine gewisse Korrelation zwischen G+C-Gehalt und NTD-Insertion im Rho-Protein.

Für *C. glutamicum* wurden in dieser Arbeit die Rho-abhängig terminierten Transkripte genomweit identifiziert. Obwohl einige Studien zur genomweiten Identifikation von Rho-abhängig terminierten Transkripten durchgeführt wurden (Peters *et al.*, 2012; Botella *et al.*, 2017; Mäder *et al.*, 2016), wurde auf dessen Basis bisher keine umfassende Analyse der CG-Blasen bzw. *rut sites* durchgeführt. Aufgrund von Strukturanalysen des Rho-Hexamers in *E. coli* wird von einer *rut site* mit einer Länge von ca. 80 bis 90 nt ausgegangen (Chen und Richardson, 1987; Zalatan und Platt, 1992; McSwiggen *et al.*, 1988). Diese Länge wird benötigt, damit jedes Rho-Monomer gleichzeitig mit einem Teil der *rut site* in Kontakt treten kann (Koslover *et al.*, 2012). Der Erkennungsmechanismus des Rho-Hexamers ist allerdings nicht geklärt. So könnte es ebenso wahrscheinlich sein, dass die einzelnen Rho-Monomere erst nacheinander mit der *rut site* in Kontakt treten. In diesem Fall wäre eine wesentlich kürzere *rut site*, die von Monomer zu Monomer wandert, für eine Induktion der Termination ausreichend (Canals *et al.*, 2010). Darüber hinaus könnte der Erkennungsmechanismus aufgrund der in vielen Rho-Homologen identifizierten NTD-Insertion und Verlängerungen im 5'-YC-Dimer-Motiv von dem in *E. coli* vermuteten Mechanismus unterschiedlich ablaufen. Für den Rho-Faktor aus *C. glutamicum* wurde in dieser Arbeit eine Analyse der verlängerten proteinkodierenden Transkripte hinsichtlich möglicher CG-Blasen bzw. *rut sites* durchgeführt. In 244 von 245 der Rho-abhängig terminierten mRNAs wurden CG-Blasen mit einer Länge von mindestens 10 nt bis maximal 478 nt gefunden. Im Durchschnitt haben die CG-Blasen eine Länge von circa 35 nt (Abbildung 4.19). Da der Rho-Faktor in *C. glutamicum* aufgrund der NTD-Insertion wesentlich größer ist als in *E. coli*, würde nach der oben genannten Annahme eine CG-Blase von über 90 nt benötigt werden, um eine Rho-abhängige Termination einzuleiten.

5.3 Rho-abhängige Transkripttermination in *C. glutamicum*

Nur 66 (27 %) der verlängerten proteinkodierenden Transkripte besitzen eine CG-Blase die mindestens 90 nt lang ist. Die Längen der CG-Blasen ist somit, zumindest für *C. glutamicum*, kein eindeutiges Indiz für eine Rho-abhängige Termination.

Interessanterweise besitzen die intrinsisch terminierten Transkripte ebenfalls CG-Blasen mit einer zu den Rho-abhängig terminierten Transkripten identischen Längenverteilung (Abbildung 4.19). Im Mittel sind die CG-Blasen der intrinsisch terminierten wie die Rho-abhängig terminierten Transkripte 35 nt lang und sind somit nicht von den CG-Blasen der Rho-abhängig terminierten Transkripte zu unterscheiden. In *C. glutamicum* können CG-Blasen daher keine *rut site* darstellen. Die Rho-abhängige Termination muss, zumindest in *C. glutamicum*, möglicherweise auch in allen Rho-Homologen mit vergleichbarer NTD-Insertion und/oder verlängertem 5'-YC-Dimer-Motiv, anders ablaufen als es für *E. coli* beschrieben wird.

In *E. coli* konnte gezeigt werden, dass die Rho-abhängige Termination von den Antiterminatorproteinen NusA und NusG beeinflusst wird (Peters *et al.*, 2012). Auch die Packungsstruktur der chromosomalen DNA durch das Histon-ähnliche Nukleoid-strukturierende Protein H-NS beeinflusst die Termination durch den Rho-Faktor in *E. coli* (Kotlajich *et al.*, 2015). Der Mechanismus der Rho-abhängigen Transkriptionstermination ist daher *in vivo* wesentlich komplexer, als über das bloße Vorhandensein von CG-Blasen erkennbar ist. *C. glutamicum* besitzt Homologe von NusA und NusG (Charlebois und Doolittle, 2004) mit zu den *E. coli* Proteinen ähnlicher Domänenstruktur. Ein H-NS-Homolog existiert allerdings nicht in *C. glutamicum* (Kalinowski *et al.*, 2003). Es ist daher wahrscheinlich, dass auch NusA oder NusG die Termination durch den Rho-Faktor in *C. glutamicum* beeinflussen. Dies könnte in weiteren Studien analysiert werden.

5.3.8 Rho reguliert ein Typ IVb Pilussystem

In der *C. glutamicum* $\Delta\rho$ -Mutante gehören die Gene *cg0362* bis *cg0369* zu den am stärksten differenziell transkribierten Genen. Diese Gene kodieren für ein putatives Typ IVb Pilussystem und wurden daher auch zusammenfassend als *tad*-Cluster bezeichnet (Imam *et al.*, 2011; Tomich *et al.*, 2007).

Funktion des Typ IVb Pilussystems in *C. glutamicum*

Bifidobacterium breve exprimiert den Typ IVb Pilus nur, wenn es in Kontakt mit eukaryotischen Zellen tritt, da es Teil der natürlichen Darmflora von Säugetieren ist und den Pilus zur Adhäsion an den Darmzellen verwendet (Motherway *et al.*, 2011). Neben *B. breve* wurde der Typ IV Pilus auch in *Micrococcus luteus* funktional beschrieben (Angelov *et al.*, 2015). Obwohl *M. luteus* wie *B. breve* zu den Actinobakterien zählt, nutzt dieses Bakterium den Typ IV Pilus

als DNA-Aufnahmesystem, was die natürliche Kompetenz von *M. luteus* erklärt (Angelov *et al.*, 2015). Die Fähigkeit von *C. glutamicum* zur natürlichen Kompetenz ist nicht beschrieben (Johnston *et al.*, 2014).

Im Rahmen dieser Arbeit konnte gezeigt werden, dass der Typ IVb Pilus von *C. glutamicum* ebenfalls die DNA-Aufnahme verbessert (Abbildung 4.24C). Eine Nutzung des Typ IVb Pilus durch *C. glutamicum* zur Adhäsion an Oberflächen ist für ein Bodenbakterium eher unwahrscheinlich, kann aber nicht ausgeschlossen werden. Obwohl die Ergebnisse der Charakterisierung des Typ IV Pilus in *C. glutamicum* und *M. luteus* für die DNA-Aufnahme sprechen, sind die für Kompetenzpilissysteme in grampositiven Bakterien bekannten Proteine ComEC, ComEA und EndA für die DNA-Aufnahme nicht im *tad*-Cluster, sondern an einem genomisch entfernten Bereich kodiert. Das Protein ComEC bildet eine Pore für die DNA-Aufnahme in der Zellmembran durch die einzelsträngige DNA mit Hilfe des DNA-Rezeptors ComEA hindurch geleitet wird. Die einzelsträngige DNA wird dabei durch die Nuklease EndA aus doppelsträngiger DNA erzeugt. Der Kompetenzpilus fungiert in diesem System als eine Art Antenne an die DNA bindet (Johnston *et al.*, 2014). Die DNA-Aufnahme kann unspezifisch oder über die Erkennung von DNA-Aufnahmesequenzen, wie bei *Neisseria gonorrhoeae*, erfolgen (Aas *et al.*, 2002). Nach erfolgter Bindung wird der Pilus in die Zelle zurückgezogen, um die DNA in Richtung Pore zu leiten (Provvedi und Dubnau, 1999; Chen und Gotschlich, 2001; Seitz *et al.*, 2014). In der *C. glutamicum* $\Delta\rho$ -Mutante sind neben dem *tad*-Cluster auch die Gene *comEC* und *comEA* verstärkt differenziell transkribiert (Abbildung 4.22). Ein *endA* Homolog konnte in *C. glutamicum* nicht nachgewiesen werden (Daten nicht gezeigt). Die im Vergleich zum Wildtyp erhöhte Transformierbarkeit der $\Delta\rho$ -Mutante lässt sich vermutlich auf die Expression des *tad*-Clusters und/oder der beiden Gene *comEC* und *comEA* zurückzuführen.

Genetische Organisation und mögliche Regulation des Typ IVb Pilussystems in *C. glutamicum*

Das in *C. glutamicum* für den Typ IVb Pilus kodierende *tad*-Cluster besteht aus sieben Genen (plus der abseits kodierten Prepillinpeptidase TadV) und ist damit wesentlich kleiner als das in *A. actinomycetemcomitans* (14 Gene; Tomich *et al.* (2007)). In Actinobakterien befindet sich die das Gen der Prepillinpeptidase häufig abseits des eigentlichen *tad*-Clusters (Tomich *et al.*, 2007). Da grampositive Bakterien keine äußere Membran besitzen, ist der Aufbau des Typ IV Pilus vermutlich weniger komplex und erfordert daher weniger Proteine und somit weniger Gene.

Im *C. glutamicum* Wildtyp ist unter Standard-Kultivierungsbedingungen das für einen putativen Typ IVb Pilus kodierende *tad*-Cluster nicht bzw. nur sehr gering transkribiert. Auch

5.3 Rho-abhängige Transkripttermination in *C. glutamicum*

unter Einfluss von unterschiedlichen Stressbedingungen ist keine erhöhte Transkription beobachtet worden (Pfeifer-Sancar *et al.*, 2013). Nur in der *C. glutamicum* $\Delta\rho$ -Mutante ist das *tad*-Cluster verstärkt differenziell transkribiert. Anhand der Vollängen- sowie der Primärtranskriptdaten der $\Delta\rho$ -Mutante konnte in dieser Arbeit die genetische Organisation des *tad*-Clusters in *C. glutamicum* erstmalig beschrieben werden. Demnach ist das *tad*-Cluster in zwei Operons mit jeweils eigenem TSS unterteilt: Operon 1 von *cg0362* bis *cg0364* und Operon 2 von *cg0365* bis *cg0369*. *B. breve* besitzt ebenfalls ein *tad*-Cluster mit identischer Genzahl und ähnlicher Operonstruktur (Motherway *et al.*, 2011). Dieses Bakterium exprimiert den Typ IVb Pilus nur, wenn es in Kontakt mit eukaryotischen Zellen tritt, um an diese zu binden (Motherway *et al.*, 2011). In vielen Spezies (*Myxococcus xanthus*, *Dichelobacter nodosus*, *Geobacter sulfurreducens*, *Kingella kingae*, *Xanthomonas axonopodis* und *Pseudomonas aeruginosa*) wird das Typ IV Pilussystem über einen σ^{54} Promotor reguliert (Giltner *et al.*, 2012). Eine weitere Regulation in *P. aeruginosa* findet über das Zweikomponentensystem PilR/PilS statt. PilS ist eine membrangebundene Histidinkinase die durch ein unbekanntes Signal autophosphoryliert wird. Anschließend phosphoryliert PilS PilR welcher die Pilin-Expression verstärkt (Giltner *et al.*, 2012).

C. glutamicum besitzt weder ein PilR/PilS-System, noch den Sigmafaktor σ^{54} . Die Regulation des *tad*-Clusters muss daher über einen anderen Mechanismus ablaufen. Die Bedingungen, unter denen *C. glutamicum* das *tad*-Cluster nativ exprimiert, sind nicht bekannt. Unter den von Pfeifer-Sancar *et al.* (2013) getesteten Stressbedingungen (Hoher Sauerstoffpartialdruck, Wasserstoffperoxid-, Diamid-, Ethanol-, Salz-, Hitze- und Kältestress) findet keine verstärkte Expression statt. In der $\Delta\rho$ -Mutante ist das *tad*-Cluster stark differenziell transkribiert. Die Regulation des *tad*-Clusters muss daher direkt oder indirekt über den Rho-Faktor erfolgen. Das *tad*-Cluster in *A. actinomycetemcomitans* besitzt drei Terminatoren T1 bis T3. Terminator T2 reagiert auf Bicyclomycin (BCM), was auf eine Rho-abhängige Termination hindeutet (Kram *et al.*, 2008), da BCM den Rho-Faktor spezifisch inhibiert. Diese Beobachtung unterstützt somit die Hypothese, dass das *tad*-Cluster in *C. glutamicum* durch den Rho-Faktor reguliert wird.

In *C. glutamicum* befindet sich direkt *upstream* des *tad*-Clusters die nichtkodierende RNA „6C RNA“. Die 6C RNA ist in Actinomyceten weit verbreitet (Weinberg *et al.*, 2010) und ist in *C. glutamicum* in der SOS-Antwort involviert (Pahlke *et al.*, 2016). In *M. tuberculosis* und *M. smegmatis* wird die 6C RNA als B11 RNA bezeichnet und befindet sich ebenfalls *upstream* eines *tad*-Clusters (DiChiara *et al.*, 2010; Arnvig und Young, 2009). Obwohl für die 6C RNA bzw. B11 RNA ein Rho-unabhängiger intrinsischer Terminator vorhergesagt wurde (Mentz *et al.*, 2013; DiChiara *et al.*, 2010; Arnvig und Young, 2009), konnte in der *C. glutamicum* $\Delta\rho$ -Mutante gezeigt werden, dass die 6C RNA um den Faktor fünf verlängert

5 Diskussion

vorliegt und somit Rho-abhängig terminiert sein muss. Dieses Ergebnis wird auch durch die Primärtranskriptdaten der $\Delta\rho$ -Mutante unterstützt (Abbildung 4.23). In einem Microarray-Experiment wurde nachgewiesen, dass unter anderem die Gene des *tad*-Clusters in einer 6C RNA-Deletionsmutante verstärkt exprimiert werden (Pahlke, 2014). Somit sind die 6C RNA-Deletionsmutante sowie die $\Delta\rho$ -Mutante die einzigen bekannten Bedingungen, in der das *tad*-Cluster in *C. glutamicum* verstärkt transkribiert wird. Die Transkriptionsregulation des *tad*-Clusters könnte daher über die Rho-abhängig terminierte 6C RNA erfolgen. Allerdings konnte bei einem *knock down* des Rho-Faktors in *M. tuberculosis* gezeigt werden, dass die Transkription des *tad*-Clusters leicht verringert ist (Botella *et al.*, 2017). Die mögliche Regulation des *tad*-Clusters durch den Rho-Faktor ist somit nicht einheitlich innerhalb der Corynebacterineae.

In *Legionella pneumophila* wurde kürzlich nachgewiesen, dass die Expression eines Kompetenzpilus-Clusters über eine ncRNA reguliert wird (Attaiech *et al.*, 2016). Im Folgenden wird Regulation vorgestellt und Parallelen zu *C. glutamicum* diskutiert.

Die ncRNA „RocR“ reprimiert zusammen mit dem RNA-Chaperon RocC die Expression eines Kompetenzpilus-Clusters in *L. pneumophila* (Attaiech *et al.*, 2016). RocR besteht aus drei *stem loops* und ist der 6C RNA aus *C. glutamicum* strukturell ähnlich. Der Repressionsmechanismus des Kompetenzpilus-Clusters durch RocR/RocC erfolgt über die Bindung von RocR an RocC, wobei ein *stem loop* von RocR linearisiert und frei exponiert wird. Über eine komplementäre Bindung des exponierten Teils von RocR an das 5'-Ende der mRNA von *comEA*, *comEC*, *comM* und *comF* wird die ribosomale Bindestelle blockiert. Durch die fehlende Translation durch Ribosomen ist die jeweilige mRNA vor RNasen ungeschützt und wird schnell degradiert. Erst durch verringerte Expression von RocR wird das Kompetenzpilus-Cluster exprimiert und die Zellen kompetent (Attaiech *et al.*, 2016). Aufgrund der großen Ähnlichkeit zwischen den beiden ncRNAs RocR und 6C RNA sowie der verstärkten Expression des *tad*-Clusters bei Deletion von *rho* bzw. der 6C RNA könnte die Regulation des *tad*-Clusters in *C. glutamicum* ähnlich wie die Regulation des Kompetenzpilus in *L. pneumophila* funktionieren. *Downstream* des *tad*-Clusters in *C. glutamicum*, wie auch in *M. tuberculosis*, befindet sich eine putative RNA-Helicase (in *C. glutamicum* kodiert durch das Gen *cg0370*). Es wäre denkbar, dass diese RNA-Helicase die 6C RNA bindet, einen Teil der 6C RNA-Sequenz frei exponiert und so eine Bindung zur mRNA von *cg0362* oder auch *comEA* bzw. *comEC* ermöglicht. Da die 6C RNA konstitutiv exprimiert wird, würde dies die stets geringe Transkriptabundanz des *tad*-Clusters in *C. glutamicum* erklären. Somit wäre die verstärkte Transkription des *tad*-Clusters in der $\Delta\rho$ -Mutante über die stark verlängerte 6C RNA, die nicht mehr von der RNA-Helicase gebunden werden kann, zu erklären. Um diese Hypothese zu überprüfen wurde eine *C. glutamicum* Deletionsmutante des Gens *cg0370* erstellt und die Transkriptabundanz des *tad*-Clusters im Vergleich zum Wildtyp gemessen. Die mittels qRT-PCR gemessene Transkriptabundanz des

tad-Clusters in der RNA-Helicase Deletionsmutante war identisch zu der des Wildtyps (Daten nicht gezeigt). Die RNA-Helicase Cg0370 ist somit nicht über den vermuteten Mechanismus an der Regulation des *tad*-Clusters beteiligt. Möglicherweise übernimmt eine andere RNA-Helicase die Funktion von RocC in *C. glutamicum*.

5.4 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Der Abbau und die Stabilität von RNA sind essenzielle Funktionen zur Aufrechterhaltung der zellulären Prozesse und erfolgen über unterschiedliche RNasen (Arraiano *et al.*, 2010; Laalami *et al.*, 2014). Aus dem Grund ist eine Charakterisierung und Analyse der RNasen für das Verständnis des RNA-Abbaus essenziell.

In dieser Arbeit wurden die RNasen E/G und J in *C. glutamicum* untersucht und charakterisiert. Dazu wurden die RNasen mit entsprechenden Homologen aus anderen Organismen verglichen. Da RNasen auf Transkriptebene wirken, wurden die Transkriptome von den jeweiligen Deletionsmutanten der RNasen untersucht. Darüber hinaus wurde eine Methode zur Identifizierung von RNase-Schnittstellen mit RNA-Sequenzierung entwickelt und zur Charakterisierung der Schnittstellen von RNase E/G und J angewendet.

5.4.1 *C. glutamicum* besitzt eine von den Modellorganismen abweichende RNase-Ausstattung

C. glutamicum besitzt die RNasen E/G und J, aber kein RNase Y-Homolog. Somit ist die Enzyenausstattung von *C. glutamicum* an zentralen RNasen weder mit der Enzyenausstattung der Modellorganismen *E. coli* und *B. subtilis* identisch. *C. glutamicum* besitzt, wie *E. coli* (Arraiano *et al.*, 2010), die 3'-5'-Exoribonukleasen PNPase und RNase R. *B. subtilis* besitzt ebenfalls je ein PNPase- und RNase R-Homolog (Condon, 2007). Somit ist die Enzyenausstattung an 3'-5'-Exoribonukleasen in *C. glutamicum* der Enzyenausstattung in *E. coli* ähnlicher als der von *B. subtilis*. Ein RppH-Homolog zur Abspaltung von Triphosphaten an den 5'-Enden von Primärtranskripten ist in *C. glutamicum*, wie auch in *B. subtilis* (Durand und Condon, 2018) und *E. coli* (Deana *et al.*, 2008), vorhanden. Da Ec.RNase E und Bs.RNase J zentrale Funktionen in der RNA-Degradation der jeweiligen Modellorganismen besitzen, wird vermutet, dass die Homologe RNase E/G und J in *C. glutamicum* ebenfalls zentrale Rollen einnehmen. Die Initiation der RNA-Degradation in *C. glutamicum* würde demnach zum Teil aus RNase E/G, wie bei *E. coli* und zum Teil aus RNase J, wie bei *B. subtilis*, erfolgen. Der anschließende Abbau würde wie bei *B. subtilis* vom 5'- als auch vom 3'-Ende durch RNase J, PNPase und RNase R erfolgen.

Da *C. glutamicum*, wie *E. coli*, ein Homolog der Oligo-Ribonuklease besitzt, erfolgt der Verdau der Fragmente zusätzlich über diese RNase. Allerdings sind die 3'-5'-Exoribonukleasen nicht Gegenstand dieser Arbeit gewesen. Um den tatsächlichen Einfluss dieser Exoribonukleasen auf die RNA-Degradation zu ermitteln, müssen weitere Studien erfolgen.

5.4.2 Die RNasen E/G und J besitzen zusätzliche N-terminale Bereiche

Da Cg.RNase E/G zwar eine vergleichbare Länge wie *E. coli* RNase E aufweist, aber eine konditionale RNase E Mutante in *E. coli* nicht komplementieren konnte, wird vermutet, dass Cg.RNase E/G weniger der Funktion von *E. coli* RNase E (Ec.RNase E), sondern mehr der Funktion von RNase G (Ec.RNase G) ähnelt (Maeda und Wachi, 2012b). Im Gegensatz zu Ec.RNase E fehlt Ec.RNase G die Transmembrandomäne, um sich in der Zellmembran zu verankern, sowie die *scaffolding* Domäne, um ein Degradosom mit PNPase, der RNA-Helicase Rhlb und der Enolase auszubilden (McDowall *et al.*, 1993). Cg.RNase E/G besitzt im Vergleich zu Ec.RNase E und G eine zusätzliche N-terminale Erweiterung, dessen Funktion unbekannt ist. Auch Mt.RNase E/G besitzt eine im Vergleich zu Ec.RNase E und G N-terminale Erweiterung und im Vergleich zu allen RNase E/G-Homologen (außer Ec.RNase G) die kürzeste CTD (Abbildung 4.25C). Für Mt.RNase E/G wurde gezeigt, dass diese eine putative Acetyltransferase sowie die Polyphosphat/ATP-NAD-Kinase Ppnk binden kann (Kovacs *et al.*, 2005). Aufgrund der kurzen CTD ist davon auszugehen, dass diese Enzyme über die NTD mit Mt.RNase E/G interagieren und so eine Art Degradosom bilden. Es ist allerdings unklar, welche Funktionen diese gebundenen Enzyme in der RNA-Degradation übernehmen könnten. Aufgrund der ähnlichen NTD-Länge von Cg.RNase E/G zu Mt.RNase E/G ist eine prinzipielle Interaktion zwischen Cg.RNase E/G und anderen Enzymen denkbar. Ob sich Cg.RNase E/G tatsächlich ein Degradosom formen kann, muss in weiteren Arbeiten gezeigt werden.

B. subtilis besitzt zwei RNase J-Homologe, RNase J1 und J2, die hauptsächlich als Heterodimer aktiv sind (Mathy *et al.*, 2010). *C. glutamicum* besitzt nur ein RNase J-Homolog, sodass diese RNase nur als Monomer oder Homodimer aktiv sein kann. Cg.RNase J besitzt einen um 136 Aminosäuren verlängerten N-terminalen Bereich im Vergleich zu Bs.RNase J1 bzw. J2 und ist ca. 22 % länger als die RNase J-Homologe in *S. coelicolor* und *B. subtilis* (Abbildung 4.25C). Die N-terminale Verlängerung scheint für Corynebakterien charakteristisch zu sein, da sie in nah verwandten Actinobakterien wie *S. coelicolor* (Pei *et al.*, 2015), *M. tuberculosis* (Sun *et al.*, 2011) und *M. smegmatis* (Taverniti *et al.*, 2011) nicht vorkommt. Die Funktion der N-terminalen Verlängerung ist allerdings unbekannt.

5.4.3 Die RNasen E/G und J nicht essenziell

RNase E ist für *E. coli* essenziell und eine Deletion des entsprechenden Gens daher letal (Apirion und Lassar, 1978), kann aber durch die Überexpression des für RNase G kodierenden Gens *rng* komplementiert werden (Lee *et al.*, 2002). Neue Untersuchungen haben ergeben, dass die Deletion von RNase E die Expression des Gens der Phosphoenolpyruvat-Synthetase (*ppsA*) verringert und somit die Glyconeogenese verhindert. Bei paralleler Überexpression von *ppsA* und Glycerol als Kohlenstoffquelle ist die Deletion von RNase E in *E. coli* möglich (Tamura *et al.*, 2016). In *M. smegmatis* ist das RNase E-Homolog essenziell (Taverniti *et al.*, 2011).

Im Gegensatz zu *E. coli* und *M. smegmatis* ist die Deletion des für die RNase E/G kodierenden Gens *rne* in *C. glutamicum* direkt möglich und hat keinen Einfluss auf das Wachstum in Minimalmedium mit Glukose (Abbildung 4.26A). Die Δrne -Mutante wächst auch in Komplexmedium mit Glukose wie der Wildtyp (Maeda und Wachi, 2012b). Vermutlich wird die Funktionalität von RNase E/G in *C. glutamicum* durch eine andere RNase kompensiert. Liegt Acetat als einzige Kohlenstoffquelle vor, ist die Wachstumsgeschwindigkeit der Δrne -Mutante sogar höher als die des Wildtyps, da die für die Isocitrat-Lyase kodierende mRNA von *aceA* in der Mutante verlangsamt abgebaut wird (Maeda und Wachi, 2012a). Die *E. coli* *rng::cat* Deletionsmutante kann durch das für Cg.RNase E/G kodierende Gen *rne* komplementiert werden (Maeda *et al.*, 2009). Allerdings komplementiert *rne* nicht die konditionale *E. coli* RNase E Mutante *rne-1* (Maeda *et al.*, 2009). *S. coelicolor* besitzt ebenfalls ein RNase E-Homolog, genannt RNase ES. RNase ES ist in *S. coelicolor* nicht essenziell. Das Gen für RNase ES komplementiert *E. coli* *rng::cat* als auch *rne-1* Mutanten (Lee und Cohen, 2003). Obwohl RNase ES wie Cg.RNase E/G eine N-terminale Verlängerung aufweisen und somit strukturell ähnlicher zu Cg.RNase E/G ist, als zu *E. coli* RNase E (Tabelle 4.10A), wurde über Komplementationsexperimente gezeigt, dass beide RNasen aus *C. glutamicum* und *S. coelicolor* funktional verschieden sind.

Die wichtigsten RNasen in der RNA-Degradation in *B. subtilis* sind die RNase Y und die RNase J1 sowie die doppelstrang-sensitive RNase III (Durand *et al.*, 2012a). Die RNasen Y und J1 sind nicht essenziell für *B. subtilis* (Even *et al.*, 2005; Britton *et al.*, 2007; Figaro *et al.*, 2013). RNase III degradiert die RNA eines Toxin-Antitoxinsystems in *B. subtilis*. Wird dieses System deletiert, ist auch RNase III in *B. subtilis* nicht mehr essenziell (Durand *et al.*, 2012b,a). Wie in dieser Arbeit gezeigt wurde, konnten für die für RNase E/G und RNase J kodierenden Gene Deletionsmutanten in *C. glutamicum* erstellt werden. Auch RNase III ist in *C. glutamicum* nicht essenziell (Maeda *et al.*, 2015). Obwohl die RNasen E, Y, J und III in den jeweiligen Modellorganismen als Haupt-RNasen wichtige Funktionen in der RNA-

Degradation und -Prozessierung übernehmen, sind diese RNasen für die Zelle verzichtbar. Die RNA-Degradation und -Prozessierung muss demnach, zumindest für die hier erwähnten Prokaryoten *E. coli*, *B. subtilis* und *C. glutamicum*, redundant organisiert sein, sodass der Verlust einer RNase durch eine andere innerhalb eines gewissen Rahmens kompensierbar ist.

5.4.4 Einfluss der RNasen E/G und J das Transkriptom in *C. glutamicum*

Um den Einfluss der RNasen E/G und J in *C. glutamicum* auf das Transkriptom zu untersuchen, wurden jeweils Einzeldelensionsmutanten der entsprechenden RNase-Gene erstellt und die RNA der Mutanten mittels Vollängentranskript-Sequenzierung auf differenziell transkribierte Gene analysiert.

Einfluss auf das Transkriptom durch RNase E/G

RNase E ist für den Abbau des Großteils der mRNA verantwortlich (Apirion und Lassar, 1978; Arraiano *et al.*, 1988; Taraseviciene *et al.*, 1991). Neben der Degradation von mRNA ist RNase E auch in der Prozessierung von 5S und 16S rRNA-Vorläufern (Apirion und Lassar, 1978; Misra und Apirion, 1979), tRNAs (Ow und Kushner, 2002) und der tmRNA (Lin-Chao *et al.*, 1999) involviert. Kürzlich wurde gezeigt, dass RNase E auch die Reifung von crRNAs in dem CRISPR-Cas System vom Subtyp III-Bv übernimmt (Behler *et al.*, 2018). Auch in dem Transkript, welches RNase E kodiert wurde eine RNase E Schnittstelle nachgewiesen. Über diese Schnittstelle erfolgt eine Autoregulation, sodass die RNase E Konzentration in der Zelle konstant gehalten wird (Mudd und Higgins, 1993; Jain und Belasco, 1995).

Die RNase G ist weniger gut untersucht und weniger über Degradations- und Prozessierungsmechanismen dieser RNase bekannt (Arraiano *et al.*, 2010). Wie RNase E ist RNase G auch an der Degradation eines Großteils der mRNA beteiligt (Carpousis *et al.*, 2009) und wird für die Reifung der 16S rRNA benötigt (Li *et al.*, 1999; Wachi *et al.*, 1999).

In der Δrne -Mutante wurden viele Gene die für ribosomale Proteine kodieren, stärker transkribiert, als im Wildtyp (Abbildung 4.27). Die Expression der Gene die für ribosomale Proteine kodieren und der rRNA-Operons ist gekoppelt, um stets eine ausreichende Menge an aktiven Ribosomen gewährleisten zu können (Gourse *et al.*, 1985). Obwohl über die differenzielle Transkriptionsanalyse keine RNase-Schnittstellen bestimmt werden können, liegt daher eine Regulation der für die ribosomalen Proteine kodierenden mRNA durch RNase E/G nahe. Wie für *E. coli* RNase E ist *M. smegmatis* RNase E ebenfalls in der Reifung von 16S, 23S und 5S rRNA beteiligt (Taverniti *et al.*, 2011). Dies unterstützt die Hypothese, dass Cg.RNase E/G über die Reifung von rRNA die Expression der ribosomalen Proteine beeinflusst und die Gene daher in der Δrne -Mutante differenziell transkribiert vorliegen.

5.4 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Die 5S rRNA Prozessierung durch Cg.RNase E/G wurde von Maeda und Wachi (2012b) gezeigt. Allerdings scheint Cg.RNase E/G keinen Einfluss auf die 16S und 23S rRNA zu haben (Maeda und Wachi, 2012b). Im Rahmen dieser Arbeit konnte ebenfalls in nicht rRNA-abgereicherten Vollängentranskript-Daten gezeigt werden, dass die Abundanz der 5S RNA in der Δrne -Mutante im Vergleich zum Wildtyp höher ist (Abbildung 4.29A). Mittels Northern Blot wurden neben der gereiften 5S rRNA (120 nt) drei 5S rRNA-Vorläufer mit Längen von ca. 150 nt, ca. 200 nt und ca. 290 nt nachgewiesen (Abbildung 4.29B). In *E. coli* sind neben den RNasen E und G weitere, noch unbekannte, RNasen beteiligt (Hayes und Vasseur, 1976; Kaczanowska und Rydén-Aulin, 2007). Es wird daher vermutet, dass die Prozessierung der Vorläufertranskripte zu gereifter 5S rRNA in *C. glutamicum* redundant organisiert ist und in der Δrne -Mutante von weiteren RNasen übernommen wird.

Einfluss von RNase J auf das Transkriptom

Die beiden RNase J Paraloge in *B. subtilis* sind als Exoribonukleasen an der Degradation eines Großteils der mRNA beteiligt (Mäder *et al.*, 2008; Durand *et al.*, 2012a). RNase J verhält sich somit ähnlich wie andere Exoribonukleasen (z. B. PNPase) (Wang und Bechhofer, 1996; Oussenko *et al.*, 2005). Während die PNPase die RNA vom 3'-Ende aus abbaut, findet der Abbau durch RNase J vom 5'-Ende aus statt. Wie RNase E und G in *E. coli* ist auch Bs.RNase J an der Prozessierung von 16S rRNA-Vorläufern beteiligt (Mathy *et al.*, 2007; Britton *et al.*, 2007). In *M. smegmatis* ist RNase J sogar an der Prozessierung der 16S, 23S und 5S rRNA beteiligt (Taverniti *et al.*, 2011). Obwohl Cg.RNase E/G in *C. glutamicum* an der Prozessierung der 5S rRNA-Vorläufer beteiligt ist, wurde für Cg.RNase J keine differenzielle Expression der rRNA in den nicht rRNA-abgereicherten Vollängentranskript-Daten der Δrnj -Mutante nachgewiesen. Möglicherweise wird die Prozessierung durch RNase J in *C. glutamicum* durch andere RNasen, wie z. B. RNase E/G kompensiert, sodass der Effekt in der Δrnj -Mutante nicht erkennbar ist. Über die Deletion weiterer RNase-Gene in der Δrnj -Mutante könnte der Effekt sichtbar werden.

In *B. subtilis* beeinflusst RNase J die Transkriptabundanz von etwa 21 % der mRNAs (Durand *et al.*, 2012a). Mit 723 differenziell transkribierten Genen in der Δrnj -Mutante, ist ein großer Teil des *C. glutamicum* Genoms (knapp 24 %) differenziell transkribiert. Demnach ist von einer wichtigen Funktion von RNase J in *C. glutamicum* auszugehen.

Die Deletion des für RNase J1 kodierenden Gens *rnjA* in *B. subtilis* beeinflusst die Transkriptmenge von RNase J2 und umgekehrt, um equimolare Mengen an Monomeren für die Bildung des Heterotetramers bereitzustellen (Durand *et al.*, 2012a; Jamalli *et al.*, 2014). Eine Kreuzregulation zwischen den Genen für RNase E/G und J wurde in der Δrne -Mutante bzw.

Δrnj -Mutante nicht beobachtet. Somit werden die Transkripte dieser RNasen wahrscheinlich nicht durch die jeweils andere RNase Prozessiert bzw. degradiert.

Vermutlich wird aber das Transkript der proteinogenen Untereinheit der RNase P von RNase J spezifisch degradiert oder prozessiert, da das entsprechende Gen *rnpA* in der Δrnj -Mutante am stärksten transkribiert vorliegt (LFC +5,07). Die katalytische M1 RNA der RNase P ist hingegen in der Δrnj -Mutante nicht differenziell transkribiert. RNase P prozessiert tRNA-Vorläufer durch Spaltung der *leader*-Sequenz am 5'-Ende (Guerrier-Takada *et al.*, 1983). Aufgrund der hohen Substratspezifität der RNase P ist eine Übernahme der RNase J Funktion in der Δrnj -Mutante unwahrscheinlich. Vermutlich wird das Transkript von *rnpA* nicht aber die M1 RNA durch RNase J im Wildtyp degradiert, so dass die Transkriptabundanz von *rnpA* in der Δrnj -Mutante erhöht ist.

5.4.5 Die Entwicklung des *RNase-Assays* ermöglicht die Identifizierung von RNase-Schnittstellen

Wie oben beschrieben ist die Unterscheidung von direkten und indirekten Effekten der Deletionen der für die RNasen E/G und J kodierenden Gene anhand von differenziellen Transkriptionsanalysen schwierig und nur für einige wenige Gene bzw. Transkripte möglich. Aus diesem Grund wurde in dieser Arbeit eine Methode mit Namen *RNase-Assay* entwickelt, um mittels Transkriptomsequenzierung spezifisch die RNase-Schnittstellen transkriptomweit zu identifizieren.

Der Nachweis von RNase-Schnittstellen mittels *RNase-Assay* erfolgt über den Vergleich der Signale einer RNA-Probe mit den jeweiligen RNase-Schnittstellen (Positivkontrolle) mit denen einer RNA-Probe ohne die jeweiligen RNase-Schnittstellen (Negativkontrolle). Im Fall des *in vivo RNase-Assays* ist die Positivkontrolle der Wildtyp und die Negativkontrolle eine Mutante in der das jeweilige RNase-Gen deletiert wurde. Im Fall des *in vitro RNase-Assays* wird die RNA einer RNase-Mutante mit der zu untersuchenden RNase behandelt (Positivkontrolle). Als Negativkontrolle dient die unbehandelte RNA.

Der *RNase-Assay* lässt sich auf alle RNasen anwenden, die nach dem Schneiden des Transkripts ein 5'-Monophosphat am 3'-Fragment hinterlassen. Die RNasen E, G, J und Y sowie die Doppelstränge degradierende RNase III hinterlassen nach dem Schneiden ein Monophosphat am 5'-Ende des 3'-Fragments (Arraiano *et al.*, 2010) und sind daher prinzipiell für den *RNase-Assay* geeignet. Exoribonukleasen die, wie PNPase, Transkripte vom 3'- zum 5'-Ende abbauen, können mit dem *RNase-Assay* nicht untersucht werden, da die dabei entstehenden 3'-Fragmente nur aus einzelnen Nukleotiden bestehen und für eine Adapter-Ligation zu kurz sind.

5.4 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Die RNA-5'-Pyrophosphohydrolase (RppH) konvertiert 5'-Triphosphate von Primärtranskripten durch Abspaltung eines Pyrophosphatrests zu 5'-Monophosphaten, um die Erkennung von 5'-Monophosphat-bevorzugenden RNasen (wie z. B. RNase E) zu ermöglichen und die Transkriptdegradation einzuleiten (Deana *et al.*, 2008). An die durch die RppH gebildeten 5'-Monophosphatenden wird ebenfalls der RNA-Adapter von der T4 RNA-Ligase ligiert. Dies führt zu falsch-positiven Signalen die TSS aber keine RNase-Schnittstellen anzeigen. Um die TSS von möglichen RNase-Schnittstellen unterscheiden zu können, sollten die TSS vor Anwendung des *RNase-Assays* für den jeweiligen Organismus bekannt sein.

Neben der hier entwickelten Methode zur Identifizierung von RNase-Schnittstellen mit Transkriptomsequenzierung existieren weitere Methoden mit vergleichbarer Anwendung. Die Methoden TIER-seq (*Transiently Inactivating an Endoribonuclease followed by RNA-seq*) (Chao *et al.*, 2017) und EMOTE (*Exact Mapping Of Transcriptome Ends*) (Linder *et al.*, 2014) sind hier als Beispiele genannt.

Die Methode von Chao *et al.* (2017) ist eine Weiterentwicklung der von Sharma *et al.* (2010) entwickelten Methode *dRNA-Seq* zur Sequenzierung von Primärtranskripten und daher zu dem hier vorgestellten *RNase-Assay* vergleichbar. TIER-seq basiert wie der *RNase-Assay* auf der Ligation eines RNA-Adapters an das 5'-Monophosphat des 3'-Fragments nach erfolgtem Schnitt einer RNase. Der wesentliche Unterschied besteht in der reversen Transkription. Bei TIER-seq werden die isolierten Transkripte mittels Poly(A)-Polymerase bei 37 °C mit einem Poly(A)-Schwanz versehen über den die reverse Transkription mit einem Oligo(dT)-Primer stattfindet. Beim *RNase-Assay* hingegen findet die reverse Transkription über einen Schlaufen-Adapter mit degenerierter Sequenz statt, wobei die Struktur des Schlaufen-Adapters nur eine Bindung am 3'-Ende eines Transkripts zulässt und damit zufällige Bindungen innerhalb des Transkripts verhindert (Pfeifer-Sancar *et al.*, 2013). Bakterielle Transkripte besitzen häufig einen Rho-unabhängigen Haarnadel-Terminator der die Bindung der Poly(A)-Polymerase behindern kann und so die Polyadenylierung der jeweiligen Transkripte verhindert (Yehudai-Resheff und Schuster, 2000; Raabe *et al.*, 2011). Dies hat zur Folge, dass nicht alle Transkripte ausreichend polyadenyliert und somit nicht sequenziert werden. Im Gegensatz dazu werden beim *RNase-Assay* vor der reversen Transkription mittels Schlaufen-Adapter die Sekundärstrukturen bei 70 °C aufgeschmolzen, direkt im Anschluss der Schlaufen-Adapter an die 3'-Transkriptenden hybridisiert und anschließend revers transkribiert. Die reverse Transkription ist so vermutlich weniger für Sekundärstrukturen am 3'-Transkriptende anfällig. Außerdem kann über den Schlaufenadapter direkt revers transkribiert werden, sodass im Vergleich zur Polyadenylierung ein Zwischenschritt gespart wird (Pfeifer-Sancar *et al.*, 2013).

Die Methode EMOTE ist ebenfalls auf eine Identifizierung von RNase-Schnittstellen über die direkte Ligation eines RNA-Adapters an das freie 5'-Monophosphat entwickelt worden (Linder

et al., 2014). Diese Methode setzt degenerierte Hexamere (*random hexamers*) als Primer für die reverse Transkription ein. Die Hexamere können ebenfalls an vorher denaturierte RNA binden und sind somit auch vermutlich weniger für Sekundärstrukturen anfällig, als die Behandlung mit Poly(A)-Polymerase. Allerdings binden die degenerierten Hexamere auch *upstream* des 3'-Transkriptendes, sodass vermutlich nicht alle 3'-Fragmente mit ausreichender Länge für die Sequenzierung revers transkribiert werden können, was letztlich zu einer geringeren Anzahl an identifizierten RNase-Schnittstellen führen sollte.

Der *RNase-Assay* könnte aufgrund des Schlaufen-Adapters im Vergleich zu den beiden anderen vorgestellten Methoden bessere Resultate bei der reversen Transkription erzielen. Die Bindung des Schlaufen-Adapters an RNA erfolgt über eine degenerierte Heptamersequenz. Über die Schlaufen-Struktur bindet der Schlaufenadapter allerdings nicht zufällig, sondern kann nur am 3'-Ende eines Transkripts binden. Dennoch kann nicht ausgeschlossen werden, dass die Bindung über die degenerierte Heptamersequenz fehlerfrei erfolgt. So wurde für degenerierte Hexamere bereits gezeigt, dass diese die Nukleotidkomposition der sequenzierten Reads verändern können (Hansen *et al.*, 2010).

Die Methoden *RNase-Assay*, TIER-seq und EMOTE wurden für den Nachweis von RNasen, die nach dem Schneiden ein 5'-Monophosphat am 3'-Fragment hinterlassen, entwickelt. Daneben existieren weitere RNasen die nach dem Schneiden ein 5'-Hydroxyl am 5'-Fragment und ein zyklisches 2'-3'-Phosphat am 3'-Fragment hinterlassen. Zu diesen RNasen zählen unter anderem die RNase MazF aus dem Toxin-Antitoxin-System MazE/MazF (Aizenman *et al.*, 1996) und die unspezifische RNase T2 (Luhtala und Parker, 2010). Für den Nachweis dieser RNase-Schnittstellen mittels Transkriptomsequenzierung existieren ebenfalls Methoden, wie beispielsweise MORE RNA-Seq (*Mapping by Overexpression of an RNase in Escherichia coli*) (Schifano *et al.*, 2014).

5.4.6 Cg.RNase E/G ist an der Initiation der RNA-Degradation beteiligt

Der in dieser Arbeit entwickelte *RNase-Assay* wurde zur transkriptomweiten Identifizierung der RNase E/G Schnittstellen in *C. glutamicum* angewendet. Konkret wurde der *in vivo RNase-Assay* für die Schnittstellen-Identifizierung verwendet. Da RNase E/G nicht aufgereinigt werden konnte, musste auf die Durchführung des *in vitro RNase-Assays* verzichtet werden.

Mit Hilfe des *in vivo RNase-Assays* konnten für RNase E/G 2.528 Schnittstellen im Transkriptom von *C. glutamicum* identifiziert werden. In *E. coli* bzw. *B. subtilis* wird die RNA-Degradation eines Transkripts durch einen bzw. wenige initiale Schnitte von RNase E bzw. Y eingeleitet (Laalami *et al.*, 2014). Die Degradation des Transkripts in kleinere Fragmente bis zu den einzelnen Nukleotiden erfolgt anschließend durch prozessive Exoribonukleasen, wie PNPase,

5.4 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Oligo-Ribonuklease oder RNase J (Andrade *et al.*, 2009; Arraiano *et al.*, 2010). Die meisten Transkripte werden von Cg.RNase E/G nur wenige Male geschnitten (Abbildung 4.34A). Dies deutet auf eine Initiation der Transkript-Degradation durch Cg.RNase E/G hin. Somit entspricht die Initiation der RNA-Degradation in *C. glutamicum* den Modellen von *E. coli* bzw. *B. subtilis*.

Die Verteilung der RNase E/G Schnittstellen über die Transkriptlänge in *C. glutamicum* ist nicht gleichmäßig. Der am häufigsten geschnittene Bereich liegt im vorderen Abschnitt eines Transkripts (Abbildung 4.34C). Im nachfolgenden Transkriptabschnitt fällt die Schnittstellenhäufigkeit rapide ab und steigt erst in den hinteren Abschnitten wieder leicht an. In *C. glutamicum* sind ca. 33 % der proteinkodierenden Transkripte *leader*-frei (Pfeifer-Sancar *et al.*, 2013). Ein Schnitt durch RNase E/G im vorderen Transkriptabschnitt führt bei *leader*-freien Transkripten zum Verlust des Startcodons und verhindert so die Translation des Transkripts. Das Transkript kann ohne den Schutz von translatierenden Ribosomen leicht von Exoribonukleasen degradiert werden (Joyce und Dreyfus, 1998). Neben der Schnittstellenhäufigkeit pro Transkript, ist dies ebenfalls ein Indikator, dass Cg.RNase E/G an der Initiation der Transkript-Degradation beteiligt ist.

In *C. glutamicum* wurden neben *leader*-freien Transkripten auch Transkripte mit 5'-UTRs von mehreren hundert Nukleotiden identifiziert, die möglicherweise regulatorische Funktionen besitzen (Pfeifer-Sancar *et al.*, 2013). Schnitte durch RNase E/G im vorderen Transkriptabschnitt bei *leader*-haltigen Transkripten könnten daher auch prozessierende Wirkung auf das Transkript haben, da die RBS nach dem Schnitt durch eine RNase weiter vorhanden bleibt und Ribosomen das Transkript nach wie vor translatieren können. Für die 5'-UTR von *colA* in *Clostridium perfringens* wurde gezeigt, dass der Schnitt von RNase Y erst die RBS durch eine sRNA-induzierte Konformationsänderung der 5'-UTR freilegt, gleichzeitig eine Haarnadel-Struktur am 5'-Ende erzeugt, die das Transkript vor 5'-3'-Exoribonukleasen schützt und somit die Translation der *colA* mRNA ermöglicht (Obana *et al.*, 2010). In *C. glutamicum* wäre eine ähnliche Regulation für Transkripte mit langen 5'-UTRs prinzipiell denkbar, da RNase Y und RNase E und somit vermutlich auch Cg.RNase E/G prinzipiell ähnliche Eigenschaften aufweisen (Shahbadian *et al.*, 2009; Laalami und Putzer, 2011). Der Mechanismus der RNA-Prozessierung von *colA* ist sehr komplex, da neben der prozessierenden RNase auch eine sRNA und die Sekundärstruktur der *colA*-mRNA eine Rolle spielen. Dieser Mechanismus ist daher nicht einfach über das Vorhandensein und die Position von RNase-Schnittstellen nachvollziehbar, sondern erfordert weitere Analysen.

In *Salmonella enterica* konnte gezeigt werden, dass RNase E Haarnadel-Terminatoren am 3'-Ende von Transkripten abspaltet (Chao *et al.*, 2017). Das Entfernen von Sekundärstrukturen am 3'-Ende führt einerseits zu einer vereinfachten Degradation des 5'-Fragments durch

5 Diskussion

3'-5'-Exoribonukleasen (Arraiano *et al.*, 2010). Andererseits konnte in *S. enterica* gezeigt werden, dass die abgespaltenen 3'-Fragmente von dem regulatorischen RNA-Chaperon Hfq gebunden werden und so an der Biogenese von sRNAs teilnehmen (Chao *et al.*, 2017). In *C. glutamicum* könnte RNase E/G prinzipiell zur Biogenese von sRNAs durch Abspaltung des Haarnadel-Terminators beitragen. Allerdings ist in *C. glutamicum* kein zu Hfq vergleichbares Protein bekannt (Daten nicht gezeigt). Auch die geringe Häufigkeit von RNase E/G Schnittstellen am 3'-Transkriptende spricht gegen eine mit *S. enterica* vergleichbare sRNA-Biogenese in *C. glutamicum*.

5.4.7 Cg.RNase E/G schneidet RNA an einem konservierten Motiv

E. coli RNase E schneidet in AU-reichen Sequenzen (Mackie, 1991; Ehretsmann *et al.*, 1992; Kaberdin, 2003). Auch das Motiv von RNase E aus *S. enterica* (RN↓WUU, mit R für G/A und W für A/U) ist dem von *E. coli* RNase E sehr ähnlich (Chao *et al.*, 2017). Mit dem RNase-Assay wurde für *C. glutamicum* RNase E/G ebenfalls das AU-reiche Motiv AAA↓CUU identifiziert (Abbildung 4.34B). Allerdings befindet sich zentral in dem Motiv ein konserviertes C, was einen deutlichen Unterschied zu den Motiven von RNase E aus *E. coli* und *S. enterica* darstellt. Für die RNase E-Homologe aus *Aquifex aeolicus* und *M. tuberculosis* wurden ebenfalls zu *E. coli* RNase E abweichende Eigenschaften nachgewiesen (Kaberdin und Bizebard, 2005; Zeller *et al.*, 2007). Für *E. coli* RNase E wurde außerdem gezeigt, dass die Schnitteffizienz stark schwanken kann und kontextabhängig ist (Tock *et al.*, 2000). Auch für RNase Y wurde in *S. aureus* ein Motiv mit einem konservierten G zwischen einer AU-reichen Sequenz nachgewiesen (Khemic *et al.*, 2015, 2016; Marincola und Wolz, 2017). Somit sind die Eigenschaften der jeweiligen RNase E-Homologe stark vom jeweiligen Organismus abhängig. Das Motiv von Cg.RNase E/G könnte auf die, im Vergleich zu *E. coli* RNase E, unterschiedliche Proteinstruktur, insbesondere die N-terminale Verlängerung, zurückzuführen sein (Abbildung 4.25A).

Zusammenhang zwischen RNase E/G Schnittstellen und Status der differentiellen Transkription in der Δrne -Mutante

Der RNase-Assay wurde entwickelt, um zwischen direkten und indirekten Effekten in der differentiellen Transkriptionsanalyse der RNase-defizienten Mutanten unterscheiden zu können. So sollten Rückschlüsse über die identifizierten RNase-Schnittstellen auf die Transkripte gezogen werden, die direkt von der jeweiligen zu untersuchenden RNase Prozessiert bzw. degradiert werden. Transkripte für die besonders viele RNase E/G Schnittstellen identifiziert wurden, wie z. B. *thiC* mit 174 Schnitten, sind auch in der Δrne -Mutante besonders stark differentiell transkribiert (LFC von *thiC* -4,13). Für diese Transkripte besteht eine Korrelation

5.4 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

zwischen der Anzahl an RNase E/G Schnittstellen und dem Status der differentiellen Transkription. Allerdings wurden auch sehr viele Transkripte mit nur einer RNase E/G Schnittstelle identifiziert, deren Transkriptabundanzen über vier Größenordnungen variieren (Abbildung 4.34D). Generell besteht demnach keine Korrelation zwischen Anzahl der RNase E/G Schnittstellen und dem Status der differentiellen Transkription. Der Effekt eines Schnitts durch RNase E/G auf ein bestimmtes Transkript ist daher stark kontextabhängig. So kann ein Schnitt beispielsweise die Sekundärstruktur eines Transkripts so verändern, dass die RBS für Ribosomen zugänglicher wird und das Transkript infolgedessen erst durch die Prozessierung seinen Zweck erfüllen kann (Obana *et al.*, 2010). Ein Schnitt in der CDS hingegen führt, wie am Beispiel von *ribM* gezeigt, zu verringerter Translation und somit zu erhöhter Anfälligkeit durch RNasen (Takemoto *et al.*, 2015). Neben der Position und Anzahl an RNase-Schnittstellen im Transkript beeinflussen auch andere Faktoren die Transkriptstabilität. Diese Faktoren sind vielfältig und umfassen unter anderem Sekundärstrukturen (Del Campo *et al.*, 2015), Helferproteine (wie RNA-Chaperone und -Helicasen) (Durand *et al.*, 2015), regulative sRNAs (Laalami *et al.*, 2014) und im Fall von Riboswitches niedermolekulare Substanzen (Serganov und Nudler, 2013; Saramago *et al.*, 2014).

Um die Regulation durch RNasen und weiteren die RNA-Degradation beeinflussenden Faktoren eingehender zu untersuchen, könnte nach bisher unbekanntem Funktionshomologen zu bekannten RNA-Chaperonen in *C. glutamicum* gesucht werden. Auch die Funktionsanalyse von möglichen regulativen *antisense* und *small* RNAs dürfte sich als vielversprechend erweisen. Darüber hinaus könnten über die Messung der RNA-Halbwertszeiten in Kombination mit den hier vorgestellten RNase-Schnittstellen weitere Einblicke in die unterschiedlichen Prozessierungs- bzw. Degradierungsmechanismen ermöglicht werden.

5.4.8 RNase J besitzt hauptsächlich Exoribonuklease-Aktivität

RNase J ist ein bifunktionales Enzym mit Endo- und 5'-3'-Exoribonuklease-Aktivität mit Präferenz zu Monophosphat- oder Hydroxyl-Rest am 5'-Transkriptende (Even *et al.*, 2005; Mathy *et al.*, 2007). Die bifunktionale Eigenschaft ist charakteristisch für β -CASP-Nukleasen (Dominski *et al.*, 2013). Die Endoribonuklease-Aktivität ist bei 5'-Triphosphaten und die Exoribonuklease-Aktivität bei 5'-Monophosphaten bevorzugt (Li de la Sierra-Gallay *et al.*, 2008; Dorléans *et al.*, 2011; Taverniti *et al.*, 2011). Die RNasen J1 und J2 im Modellorganismus *B. subtilis* sind die bisher am besten untersuchten RNasen dieses Typs in Bakterien. (Li de la Sierra-Gallay *et al.*, 2008; Mathy *et al.*, 2010). Die RNasen J1 und J2 sind als Heterodimer oder Heterotetramer aktiv, wobei die Rolle von J2 noch nicht voll verstanden ist (Mathy *et al.*, 2010; Commichau *et al.*, 2009). Allerdings wurde die Endoribonuklease-Aktivität von

5 Diskussion

RNase J in *B. subtilis* nur *in vitro* nachgewiesen (Mathy *et al.*, 2007, 2010). Ob RNase J *in vivo* ebenfalls Endoribonuklease-Aktivität aufweist, wird aktuell diskutiert (Durand und Condon, 2018). Zum einen gibt es keine eindeutigen Daten zu endoribonukleolytischen Schnitten *in vivo* (Durand und Condon, 2018). Zum anderen würde der Wechsel zwischen endoribonukleolytischem und exonukleolytischem Modus größere strukturelle Änderungen im RNase J-Dimer erfordern (Dorléans *et al.*, 2011; Newman *et al.*, 2011).

Das RNase J1-Homolog in *S. aureus* ist in der Lage sowohl Transkripte mit 5'-Tri- als auch 5'-Monophosphat oder 5'-Hydroxyl zu degradieren. Auch wurde von den selben Autoren gezeigt, dass *S. aureus* RNase J1 sowohl Endo- als auch 5'-3'-Exoribonuklease-Aktivität aufweist (Hausmann *et al.*, 2017). Wie *B. subtilis* RNase J1 und J2 besitzt auch *S. aureus* RNase J1 keine N-terminale Erweiterung (Daten nicht gezeigt), wie sie bei *C. glutamicum* RNase J nachgewiesen wurde (Abbildung 4.25C). Für *M. smegmatis* RNase J konnte *in vitro* eine Endo- und 5'-3'-Exoribonuklease-Aktivität mit Präferenz zu Transkripten mit 5'-Monophosphat nachgewiesen werden (Taverniti *et al.*, 2011).

Insgesamt wurden mittels kombiniertem *in vivo* / *in vitro* RNase-Assay über 5.000 RNase J Schnittstellen im Transkriptom von *C. glutamicum* identifiziert (Abbildung 4.35). Die Schnittstellenhäufigkeit über alle Transkripte mit RNase J Schnittstelle ist wie bei Cg.RNase E/G am 5'-Ende am höchsten, nimmt über die weiteren Transkriptabschnitte aber nur langsam ab (Abbildung 4.36B). Dieses Muster entspricht einer RNase mit überwiegender 5'-3'-Exoribonuklease-Aktivität. Da die Transkripte vom 5'-Ende aus prozessiv abgebaut werden, liegen von einer Transkriptspezies viele um wenige Nukleotide verkürzte Fragmente in der Zelle vor. Aus diesem Grund sinkt die Schnittstellenhäufigkeit nur langsam vom 5'- zum 3'-Transkriptende über alle Transkripte ab.

Mittels RNase-Assay wurden auch Transkripte in *C. glutamicum* identifiziert, die nur eine RNase J Schnittstelle aufwiesen. Diese Schnittstellen weisen auf eine Endoribonuklease-Aktivität von Cg.RNase J hin. Eine RNase J Schnittstelle wurde identifiziert, wenn an der selben Position im *in vivo* als auch im *in vitro* RNase-Assay ein charakteristisches Signal vorlag. Somit wurden die möglichen Schnittstellen im *in vivo* RNase-Assay mit denen des *in vitro* RNase-Assays (und umgekehrt) validiert. Dies ist ein deutlicher Hinweis, dass Cg.RNase J auch *in vivo* Endoribonuklease-Aktivität aufweist. Allerdings konnte für diese Schnittstellen kein Motiv identifiziert werden (Daten nicht gezeigt). Es wäre denkbar, dass ein Teil der Endoribonuklease-Schnittstellen tatsächlich Exoribonuklease-Schnittstellen sind, die Signale der benachbarten Schnittstellen aber nicht deutlich genug waren und somit ausgefiltert wurden. Eine Möglichkeit zur Überprüfung dieser vermeintlichen Endoribonuklease-Schnitte könnte über Zeitreihen des RNase-Assays erfolgen.

5.4 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Nichtsdestotrotz deuten die Ergebnisse auf eine Cg.RNase J mit 5'-3'-Exo- als auch Endoribonuklease-Aktivität *in vitro* und *in vivo* hin. Diese Beobachtung widerspricht nicht den oben genannten Vorbehalten von Durand und Condon (2018), ob die RNase J-Homologe in *B. subtilis* eine Endoribonuklease-Aktivität *in vivo* zeigen. Vielmehr zeigen die Ergebnisse für Cg.RNase J auf, dass die RNase J-Homologe in anderen Bakterien verschiedene Eigenschaften besitzen. Die Studie zu RNase J1 aus *S. aureus* bestätigt diese Vermutung indirekt (Hausmann *et al.*, 2017).

Zusammenhang zwischen RNase J Schnitten und Status der differentiellen Transkription in der Δrnj -Mutante

In dieser Arbeit wurde RNase J aufgereinigt, sodass für diese RNase neben dem *in vivo* auch der *in vitro* RNase-Assay durchgeführt werden konnte. So konnten die mittels *in vivo* RNase-Assay identifizierten RNase J Schnittstellen über den *in vitro* RNase-Assay validiert werden.

Wie bei der Analyse der RNase E/G Schnittstellen konnte auch zwischen der Anzahl an RNase J Schnittstellen pro Transkript und der Transkriptabundanz keine Korrelation festgestellt werden (Abbildung 4.34D). So variiert die Transkriptabundanz der Transkripte mit nur einer Schnittstelle über drei Größenordnungen. Demnach müssen weitere Faktoren, wie beispielsweise Helferproteine oder Sekundärstrukturen, auf die Transkriptstabilität einwirken.

Für RNase J sind einige Zieltranskripte bekannt. So werden die 5'-UTRs der Transkripte von *thrS*, *thrZ* und des *trp*-Operons von RNase J in *B. subtilis* geschnitten (Even *et al.*, 2005; Deikus *et al.*, 2008). Auch die *small cytoplasmic* RNA (scRNA) in *B. subtilis* sowie die M1 RNA des RNase P Ribozyms in *S. aureus* werden von RNase J prozessiert (Yao *et al.*, 2007; Linder *et al.*, 2014). Außerdem ist RNase J in *B. subtilis* und *M. smegmatis* an der Prozessierung der rRNA beteiligt (Mathy *et al.*, 2007; Taverniti *et al.*, 2011).

In den Transkripten von *trhS* und *thrZ*, sowie der scRNA (auch 4.5S RNA genannt) wurden keine RNase J Schnittstellen in *C. glutamicum* identifiziert (Daten nicht gezeigt). Auch wurden keine Schnittstellen in den rRNA-Transkripten nachgewiesen. Aufgrund der vorherigen rRNA-Abreicherung in der Erstellung der cDNA-Bibliotheken könnten rRNA-Vorläufer abgereichert worden sein, sodass keine Schnitte in der rRNA identifiziert werden konnten. In *C. glutamicum* wurde eine RNase J Schnittstelle in der 5'-UTR des *trp*-Operons identifiziert. Innerhalb der 5'-UTR ist das *leader*-Peptid TrpL kodiert (Neshat *et al.*, 2014). Da *trpL* in der Δrnj -Mutante verstärkt transkribiert wird, kann davon ausgegangen werden, dass Cg.RNase J durch den Schnitt an der Degradation des *trp*-Operon-Transkripts beteiligt ist.

5 Diskussion

Für die M1 RNA in *S. aureus* wurde gezeigt, dass RNase J die Prozessierung des 5'-Endes durchführt (Linder *et al.*, 2014). Aufgrund der Positionen der beiden RNase J Schnittstellen *downstream* des M1 RNA TSS wird von einer vergleichbaren Funktion in *C. glutamicum* ausgegangen. RNase J ist somit wahrscheinlich an der Reifung der M1 RNA beteiligt, indem RNase J die M1 RNA aus dem Operon-Transkript trennt.

Das Gen, das die proteinogene Untereinheit der RNase P (*rnpA*) in *C. glutamicum* kodiert, befindet sich in einem Operon aus *rpmH*, *rnpA*, *cg3430* und *cg3429*. Die RNase J schneidet hier häufig innerhalb von *rpmH*. Das Operon-Transkript wird demnach bis einschließlich *rpmH* degradiert, sodass das 5'-Ende von *rnpA* frei liegt. Demnach wird das Transkript bis zu *rnpA* exonukleolytisch durch RNase J abgebaut. Möglicherweise stoppt RNase J danach, um aufgrund von Sekundärstrukturen, die das 3'-Fragment mit *rnpA* stabilisieren. Anschließend übernehmen vermutlich andere RNasen die Prozessierung oder Degradation des Transkripts bestehend aus *rnpA*, *cg3430* und *cg3429*, was die starke differenzielle Transkription von *rnpA* erklären könnte. Der Grund für die verringerte Transkription von *rnpB* in der Δrnj -Mutante kann mit den vorliegenden Daten nicht erklärt werden. Hier wären weiterführende Studien zur Klärung der dahinter liegenden Mechanismen nötig.

5.4.9 Prozessierung und Degradation von Riboswitches und anderen regulatorischen Elementen durch die RNasen E/G und J

RNase Y ist an der Degradation von etwa 30 % der regulatorischen Elemente, wie Riboswitches, in *B. subtilis* beteiligt (Laalami *et al.*, 2013). Auch in *C. glutamicum* besitzen etwa 40 % der regulatorischen Elemente, wie Riboswitches und RNA-Motive, RNase E/G Schnittstellen (Abbildung 4.37A). Auch für RNase J wurden Schnittstellen in einem RNA-Motiv und einer *leader* Sequenz nachgewiesen (Abbildung 4.37B). Demnach werden auch in *C. glutamicum* diese regulatorischen Elemente durch die beiden wichtigsten RNasen degradiert.

C. glutamicum RNase E/G ist an der Degradation des FMN-Riboswitches in *cg0083* / *ribM* beteiligt (Takemoto *et al.*, 2015). Allerdings konnte von Takemoto *et al.* (2015) keine exakte Position der Schnittstelle identifiziert werden. Mit Hilfe des RNase-Assays konnte die exakte Position der RNase E/G Schnittstelle im FMN-Riboswitch detektiert werden. Die Schnittstelle befindet sich nicht innerhalb des Riboswitches, sondern 6 nt *downstream* des Startcodons. Durch den von RNase E/G durchgeführten Schnitt wird somit die RBS und das Startcodon aus dem *ribM*-Transkript entfernt und so die Translation der proteinkodierenden Sequenz verhindert. Auch eine Translation wie bei *leader*-freien Transkripten ist nicht mehr möglich, da auch das Startcodon durch RNase E/G entfernt wird. Die Funktion der RNase E/G Schnittstelle ist demnach eindeutig die Degradation des *ribM*-Transkripts. Ferner konnten Takemoto *et al.*

5.4 RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

(2015) zeigen, dass auch der Rho-Faktor an der Regulation des FMN-Riboswitches in *ribM* beteiligt ist. Die Beteiligung des Rho-Faktors an der Riboswitch Regulation wurde, wie die RNase E/G Schnittstelle, ebenfalls in dieser Arbeit bestätigt.

RNase J Schnittstellen wurden in den Sequenzen des *ydaO-yuaA-leader* von *cg2402* (zehn Schnittstellen) sowie dem RNA-Motiv von *msiK1* (sieben Schnittstellen) identifiziert (Tabelle 4.16 und Abbildung 4.37B). Obwohl in den regulatorischen Sequenzen dieser Transkripte Schnittstellen identifiziert wurden, sind sie in der Δrnj -Mutante nicht differenziell transkribiert. Vermutlich ist RNase J nicht spezifisch an der Initiation der RNA-Degradation dieser Transkripte beteiligt, sondern nimmt erst nach initialem Schnitt einer anderen (endoribonukleolytischen) RNase an der RNA-Degradation des Transkripts teil. Im Fall von *cg2402* könnte der initiale Schnitt durch RNase E/G stattfinden, da für dieses Transkript Schnittstellen dieser RNase identifiziert wurden. Die hier aufgeführten regulatorischen Elemente in 5'-UTRs besitzen teilweise starke Sekundärstrukturen. In *B. subtilis* wurde gezeigt, dass RNase J nicht in Bereichen mit Sekundärstrukturen schneidet (Mathy *et al.*, 2007). Da RNase J nur wenige dieser Transkripte mit regulatorischen Elementen (und somit Sekundärstrukturen) in *C. glutamicum* schneidet, wird davon ausgegangen, dass andere RNasen die Degradation dieser Transkripte übernehmen.

6 Ausblick

Im Rahmen dieser Arbeit wurden die Transkriptome von *C. diphtheriae* und *C. glutamicum* bezüglich der RNA-Synthese und des RNA-Abbaus untersucht. Im Detail wurden umfassende Analysen der Transkriptionsinitiation und -termination sowie der Transkript-Degradation durchgeführt. Die Anwendung von modernen RNA-Sequenzierungstechniken und neuen Auswertemethoden ermöglichte detaillierte Analysen der zugrundeliegenden Transkriptionsprozesse und beleuchtete komplexe transkriptionelle Strukturen, Regulationen und Wechselwirkungen.

Die hier für *C. diphtheriae* präsentierten Charakterisierungen der transkriptionellen Strukturen, wie Transkriptionsstarts, 5'-UTRs und Operons, erweitern das Verständnis der Transkriptionsregulation und bilden eine Basis für weitere Transkriptomanalysen zu diesem wichtigen Pathogen. Basierend auf den hier vorgestellten Daten sind nun Transkriptom-Studien zu Wirt-Pathogen-Interaktionen möglich. Auch die dadurch erzielte Verbesserung der Genomannotation ist ein wichtiger Schritt in diese Richtung.

Besonders die identifizierten asRNAs im Diphtherietoxin-kodierenden Gen *tox* und die benachbarten intragenischen Transkripte sind interessante Elemente mit möglichen regulativen Funktionen. Im Hinblick auf die weltweite Verbreitung von *C. diphtheriae* ist Diphtherie als Krankheit immer noch relevant. Daher ist das Verständnis von Pathogenitätsfaktoren und dessen Regulation wichtig und sollte in zukünftigen Arbeiten weiter untersucht werden.

Die für *C. glutamicum* vorhandenen Primärtranskriptdaten sind bereits in weitere Studien eingeflossen, um beispielsweise die Produktion von Metaboliten zu optimieren (Taniguchi und Wendisch, 2015; Zhang *et al.*, 2015). Für die hier präsentierten Primärtranskriptdaten in *C. diphtheriae* ist eine ähnliche Entwicklung im Bereich der medizinischen Forschung zu erwarten.

Die in dieser Arbeit vorgestellte transkriptomweite Analyse der in *C. glutamicum* Rho-abhängig terminierten Transkripte vervollständigt zusammen mit den bereits bekannten Haarnadel-Terminatoren das Bild der Transkriptionstermination in *C. glutamicum*. Weiter konnte gezeigt werden, dass der Rho-Faktor besonders Gene aus Fremd-Quellen transkriptionell stilllegt. In zukünftigen Studien sollte der hier zugrunde liegende Mechanismus der Erkennung von Fremd-DNA durch Rho analysiert und aufgeklärt werden. Da bisher die meis-

ten Studien zu Rho in *E. coli* durchgeführt wurden, könnte der Rho-Faktor aus *C. glutamicum* ein Modell für Rho-Homologe mit NTD-Insertion darstellen.

In dieser Arbeit wurde durch die erste transkriptomweite Identifizierung von RNase-Schnittstellen in *C. glutamicum* zum einen das Bild der Transkript-Degradation in *C. glutamicum* erweitert und zum anderen der Grundstein für weitere Transkriptomanalysen gelegt. Vermutlich sind in *C. glutamicum* weitere RNasen, wie zum Beispiel die doppelstrang-sensitive RNase III, ebenfalls für die RNA-Prozessierung und -Degradation relevant. Diese RNasen könnten nun mittels *RNase-Assay* eingehend untersucht werden. Sekundärstrukturen beeinflussen die Transkript-Degradation und können mit speziellen RNA-Sequenzierungstechniken, wie *Parallel Analysis of RNA Structure* (PARS) (Kertesz *et al.*, 2010), für das Transkriptom identifiziert werden. Die Kombination aus mittels *RNase-Assay* identifizierten RNase-Schnittstellen und mit PARS detektierten RNA-Sekundärstrukturen ist ein vielversprechender Ansatz um weitere Einblicke in die RNA-Degradation zu erhalten.

Auch die Funktionsanalyse von möglichen regulativen *antisense* und *small* RNAs dürfte sich als vielversprechend erweisen. Darüber hinaus sollte die Messung der RNA-Halbwertszeiten in Kombination mit den hier vorgestellten RNase-Schnittstellen weitere Einblicke in die unterschiedlichen Prozessierungs- bzw. Degradierungsmechanismen erbringen.

Insgesamt bilden die hier gezeigten Analysen ein weiteres Puzzlestück zur Vervollständigung des transkriptionellen Atlas der beiden Corynebakterien *C. diphtheriae* und *C. glutamicum*. Da *C. glutamicum* ein Modellorganismus für die Actinomycetales darstellt, sind die für die Transkriptionstermination und Transkript-Degradation erhaltenen Ergebnisse und Schlussfolgerungen auf andere Corynebakterien, wie *C. diphtheriae*, übertragbar und somit relevant für zukünftige Arbeiten in der Grundlagen- oder anwendungsbezogenen Forschung.

Literaturverzeichnis

- Aas, F. E., Løvold, C. und Koomey, M. (2002) An inhibitor of DNA binding and uptake events dictates the proficiency of genetic transformation in *Neisseria gonorrhoeae*: mechanism of action and links to Type IV pilus expression. *Molecular microbiology*, 46:1441–1450.
- Ait-Bara, S. und Carpousis, A. J. (2015) RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs. *Mol Microbiol*, 97(6):1021–1135.
- Aizenman, E., Engelberg-Kulka, H. und Glaser, G. (1996) An *Escherichia coli* chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proceedings of the National Academy of Sciences of the United States of America*, 93:6059–6063.
- Albersmeier, A., Pfeifer-Sancar, K., Rückert, C. und Kalinowski, J. (2017) Genome-wide determination of transcription start sites reveals new insights into promoter structures in the actinomycete *Corynebacterium glutamicum*. *Journal of biotechnology*, 257:99 – 109.
- Alifano, P., Rivellini, F., Limauro, D., Bruni, C. B. und Carlomagno, M. S. (1991) A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell*, 64(3):553–563.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. und Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25:3389–3402.
- Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A. und Yu, Y.-K. (2005) Protein database searches using compositionally adjusted substitution matrices. *The FEBS journal*, 272:5101–5109.
- Alwine, J. C., Kemp, D. J. und Stark, G. R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. U.S.A.*, 74:5350–5354.

- Andrade, J. M., Pobre, V., Silva, I. J., Domingues, S. und Arraiano, C. M. (2009) The role of 3'-5' exoribonucleases in RNA degradation. *Prog Mol Biol Transl Sci*, 85:187–229.
- Andrews, S. C., Robinson, A. K. und Rodríguez-Quiñones, F. (2003) Bacterial iron homeostasis. *FEMS microbiology reviews*, 27:215–237.
- Angelov, A., Bergen, P., Nadler, F., Hornburg, P., Lichev, A., Übelacker, M., Pachl, F., Kuster, B. und Liebl, W. (2015) Novel Flp pilus biogenesis-dependent natural transformation. *Front Microbiol*, 6:84.
- Ao, W., Gaudet, J., Kent, W. J., Muttumu, S. und Mango, S. E. (2004) Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. *Science*, 305(5691):1743–1746.
- Apirion, D. und Lassar, A. B. (1978) A conditional lethal mutant of *Escherichia coli* which affects the processing of ribosomal RNA. *The Journal of biological chemistry*, 253:1738–1742.
- Aravind, L. (1999) An evolutionary classification of the metallo-beta-lactamase fold proteins. *In silico biology*, 1:69–91.
- Arnvig, K. B. und Young, D. B. (2009) Identification of small RNAs in *Mycobacterium tuberculosis*. *Molecular microbiology*, 73:397–408.
- Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., Moreira, R. N., Pobre, V., Reis, F. P., Saramago, M., Silva, I. J. und Viegas, S. C. (2010) The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev*, 34(5):883–923.
- Arraiano, C. M., Yancey, S. D. und Kushner, S. R. (1988) Stabilization of discrete mRNA breakdown products in *ams pnp rnb* multiple mutants of *Escherichia coli* K-12. *Journal of bacteriology*, 170:4625–4633.
- Attaiech, L., Boughammoura, A., Brochier-Armanet, C., Allatif, O., Peillard-Fiorente, F., Edwards, R. A., Omar, A. R., MacMillan, A. M., Glover, M. und Charpentier, X. (2016) Silencing of natural transformation by an RNA chaperone and a multitarget small RNA. *Proceedings of the National Academy of Sciences of the United States of America*, 113:8813–8818.
- Ayers, M., Howell, P. L. und Burrows, L. L. (2010) Architecture of the type II secretion and type IV pilus machineries. *Future microbiology*, 5:1203–1218.
- Bailey, T. L. und Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, 2:28–36.

- Bandyra, K. J., Bouvier, M., Carpousis, A. J. und Luisi, B. F. (2013) The social fabric of the RNA degradosome. *Biochim Biophys Acta*, 1829(6-7):514–522.
- Bardonnnet, N. und Blanco, C. (1991) Improved vectors for transcriptional signal screening in corynebacteria. *FEMS microbiology letters*, 68:97–102.
- Bardwell, J. C., Régnier, P., Chen, S. M., Nakamura, Y., Grunberg-Manago, M. und Court, D. L. (1989) Autoregulation of RNase III operon by mRNA processing. *The EMBO journal*, 8:3401–3407.
- Baron, S. (1996) *Medical Microbiology*. University of Texas Medical Branch at Galveston, Galveston (TX), 4th Auflage.
- Bastet, L., Chauvier, A., Singh, N., Lussier, A., Lamontagne, A.-M., Prévost, K., Massé, E., Wade, J. T. und Lafontaine, D. A. (2017) Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic acids research*, 45:7474–7486.
- Behler, J., Sharma, K., Reimann, V., Wilde, A., Urlaub, H. und Hess, W. R. (2018) The host-encoded RNase E endonuclease as the crRNA maturation enzyme in a CRISPR–Cas subtype III-Bv system. *Nature Microbiology*, Seite 1.
- Bernard, K. A., Wiebe, D., Burdz, T., Reimer, A., Ng, B., Singh, C., Schindle, S. und Pacheco, A. L. (2010) Assignment of *Brevibacterium stationis* (ZoBell and Upham 1944) Breed 1953 to the genus *Corynebacterium*, as *Corynebacterium stationis* comb. nov., and emended description of the genus *Corynebacterium* to include isolates that can alkalinize citrate. *International journal of systematic and evolutionary microbiology*, 60:874–879.
- Berry, J.-L. und Pelicic, V. (2015) Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. *FEMS Microbiology Reviews*, 39(1):1–21.
- Bogden, C. E., Fass, D., Bergman, N., Nichols, M. D. und Berger, J. M. (1999) The structural basis for terminator recognition by the Rho transcription termination factor. *Molecular cell*, 3(4):487–493.
- Bolger, A. M., Lohse, M. und Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.
- Borukhov, S. und Nudler, E. (2008) RNA polymerase: the vehicle of transcription. *Trends in microbiology*, 16:126–134.

- Botella, L., Vaubourgeix, J., Livny, J. und Schnappinger, D. (2017) Depleting Mycobacterium tuberculosis of the transcription termination factor Rho causes pervasive transcription and rapid death. *Nature communications*, 8:14731.
- Boudvillain, M., Figueroa-Bossi, N. und Bossi, L. (2013) Terminator still moving forward: expanding roles for Rho factor. *Curr Opin Microbiol*, 16(2):118–124.
- Boyd, J. und Murphy, J. R. (1988) Analysis of the diphtheria tox promoter by site-directed mutagenesis. *J Bacteriol*, 170:5949–5952.
- Boyd, J., Oza, M. N. und Murphy, J. R. (1990) Molecular cloning and DNA sequence analysis of a diphtheria tox iron-dependent regulatory element (dtxR) from *Corynebacterium diphtheriae*. *Proc Natl Acad Sci U S A*, 87:5968–5972.
- Breaker, R. R. (2011) Prospects for riboswitch discovery and analysis. *Molecular cell*, 43:867–879.
- Britton, R. A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W. C., Mathy, N., Tobin, C., Daou, R., Szyk, J. und Condon, C. (2007) Maturation of the 5' end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol*, 63(1):127–138.
- Brock, J. E., Pourshahian, S., Giliberti, J., Limbach, P. A. und Janssen, G. R. (2008) Ribosomes bind leaderless mRNA in *Escherichia coli* through recognition of their 5'-terminal AUG. *RNA (New York, N.Y.)*, 14:2159–2169.
- Browning, D. F. und Busby, S. J. (2004) The regulation of bacterial transcription initiation. *Nature reviews. Microbiology*, 2:57–65.
- Browning, D. F. und Busby, S. J. (2016) Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, 14(10):638–650.
- Brune, I., Werner, H., Hüser, A. T., Kalinowski, J., Pühler, A. und Tauch, A. (2006) The DtxR protein acting as dual transcriptional regulator directs a global regulatory network involved in iron metabolism of *Corynebacterium glutamicum*. *BMC Genomics*, 7:21.
- Brüssow, H., Canchaya, C. und Hardt, W.-D. (2004) Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and molecular biology reviews : MMBR*, 68:560–602, table of contents.
- Burgos, J. M. und Schmitt, M. P. (2012) The ChrA response regulator in *Corynebacterium diphtheriae* controls hemin-regulated gene expression through binding to the hmuO and hrtAB promoter regions. *J Bacteriol*, 194(7):1717–1729.

- Busby, S. J. W. und Savery, N. J. (2007) Transcription activation at bacterial promoters. *eLS*.
- Cadenas, R. F., Martín, J. F. und Gil, J. A. (1991) Construction and characterization of promoter-probe vectors for *Corynebacteria* using the kanamycin-resistance reporter gene. *Gene*, 98:117–121.
- Callaghan, A. J., Aurikko, J. P., Ilag, L. L., Günter Grossmann, J., Chandran, V., Kühnel, K., Poljak, L., Carpousis, A. J., Robinson, C. V., Symmons, M. F. und Luisi, B. F. (2004) Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. *Journal of molecular biology*, 340:965–979.
- Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G. und Luisi, B. F. (2005) Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature*, 437:1187–1191.
- Callebaut, I., Moshous, D., Mornon, J.-P. und de Villartay, J.-P. (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res*, 30(16):3592–3601.
- Canals, A., Usón, I. und Coll, M. (2010) The structure of RNA-free Rho termination factor indicates a dynamic mechanism of transcript capture. *Journal of molecular biology*, 400(1):16–23.
- Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E. und Nudler, E. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science*, 320(5878):935–938.
- Carpousis, A. J., Luisi, B. F. und McDowall, K. J. (2009) Endonucleolytic initiation of mRNA decay in *Escherichia coli*. *Progress in molecular biology and translational science*, 85:91–135.
- Carzaniga, T., Briani, F., Zangrossi, S., Merlino, G., Marchi, P. und Dehò, G. (2009) Autogenous regulation of *Escherichia coli* polynucleotide phosphorylase expression revisited. *Journal of bacteriology*, 191:1738–1748.
- Cascante-Esteva, N., Gunka, K. und Stülke, J. (2016) Localization of Components of the RNA-Degrading Machine in *Bacillus subtilis*. *Frontiers in Microbiology*, 07.
- Cerdeño-Tárraga, A. M., Efstratiou, A., Dover, L. G., Holden, M. T. G., Pallen, M., Bentley, S. D., Besra, G. S., Churcher, C., James, K. D., De Zoysa, A., Chillingworth, T., Cronin, A., Dowd, L., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Moule, S., Quail, M. A., Rabinowitsch, E., Rutherford, K. M., Thomson, N. R., Unwin, L., Whitehead, S., Barrell, B. G. und Parkhill,

- J. (2003) The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res*, 31(22):6516–6523.
- Chao, Y., Li, L., Girodat, D., Förstner, K. U., Said, N., Corcoran, C., Šmiga, M., Papenfort, K., Reinhardt, R., Wieden, H.-J., Luisi, B. F. und Vogel, J. (2017) In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Molecular cell*, 65:39–51.
- Charlebois, R. L. und Doolittle, W. F. (2004) Computing prokaryotic gene ubiquity: rescuing the core from extinction. *Genome research*, 14:2469–2477.
- Chen, Y. A., Chou, C.-C., Lu, X., Slate, E. H., Peck, K., Xu, W., Voit, E. O. und Almeida, J. S. (2006a) A multivariate prediction model for microarray cross-hybridization. *BMC bioinformatics*, 7:101.
- Chen, I. und Gotschlich, E. C. (2001) ComE, a competence protein from *Neisseria gonorrhoeae* with DNA-binding activity. *Journal of bacteriology*, 183:3160–3168.
- Chen, I., Provvedi, R. und Dubnau, D. (2006b) A macromolecular complex formed by a pilin-like protein in competent *Bacillus subtilis*. *The Journal of biological chemistry*, 281:21720–21727.
- Chen, C. Y. und Richardson, J. P. (1987) Sequence elements essential for rho-dependent transcription termination at lambda tR1. *The Journal of biological chemistry*, 262:11292–11299.
- Choe, S., Bennett, M. J., Fujii, G., Curmi, P. M., Kantardjieff, K. A., Collier, R. J. und Eisenberg, D. (1992) The crystal structure of diphtheria toxin. *Nature*, 357:216–222.
- Collins, J. A., Irnov, I., Baker, S. und Winkler, W. C. (2007) Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev*, 21(24):3356–3368.
- Commichau, F. M., Rothe, F. M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., Hammer, E., Völker, U. und Stülke, J. (2009) Novel activities of glycolytic enzymes in *Bacillus subtilis*: interactions with essential proteins involved in mRNA processing. *Molecular & cellular proteomics : MCP*, 8:1350–1360.
- Condon, C. (2007) Maturation and degradation of RNA in bacteria. *Curr Opin Microbiol*, 10(3):271–278.
- Condon, C. und Putzer, H. (2002) The phylogenetic distribution of bacterial ribonucleases. *Nucleic acids research*, 30:5339–5346.

- Constantinides, A. (1980) Steroid transformation at high substrate concentrations using immobilized *Corynebacterium simplex* cells. *Biotechnology and bioengineering*, 22:119–136.
- Crooks, G. E. (2004) WebLogo: A Sequence Logo Generator. *Genome Res*, 14(6):1188–1190.
- Csanadi, A., Faludi, I. und Miczak, A. (2009) MSMEG_4626 ribonuclease from *Mycobacterium smegmatis*. *Molecular biology reports*, 36:2341–2344.
- Daiyasu, H., Osaka, K., Ishino, Y. und Toh, H. (2001) Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS letters*, 503:1–6.
- Das, A., Merrill, C. und Adhya, S. (1978) Interaction of RNA polymerase and rho in transcription termination: coupled ATPase. *Proceedings of the National Academy of Sciences of the United States of America*, 75:4828–4832.
- d'Aubenton Carafa, Y., Brody, E. und Thermes, C. (1990) Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *Journal of molecular biology*, 216:835–858.
- de Hoon, M. J. L., Makita, Y., Nakai, K. und Miyano, S. (2005) Prediction of transcriptional terminators in *Bacillus subtilis* and related species. *PLoS computational biology*, 1:e25.
- Deana, A., Celesnik, H. und Belasco, J. G. (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature*, 451(7176):355–358.
- Deikus, G., Condon, C. und Bechhofer, D. H. (2008) Role of *Bacillus subtilis* RNase J1 endonuclease and 5'-exonuclease activities in *trp* leader RNA turnover. *J Biol Chem*, 283(25):17158–17167.
- Del Campo, C., Bartholomäus, A., Fedyunin, I. und Ignatova, Z. (2015) Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet*, 11(10):e1005613.
- Denoëud, F., Aury, J.-M., Da Silva, C., Noel, B., Rogier, O., Delledonne, M., Morgante, M., Valle, G., Wincker, P., Scarpelli, C., Jaillon, O. und Artiguenave, F. (2008) Annotating genomes with massive-scale RNA sequencing. *Genome Biol*, 9:R175.
- Denoëud, F., Kapranov, P., Ucla, C., Frankish, A., Castelo, R., Drenkow, J., Lagarde, J., Alioto, T., Manzano, C., Chrast, J. *et al.* (2007) Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome research*, 17(6):746–759.

- D'Heygère, F., Rabhi, M. und Boudvillain, M. (2013) Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology (Reading, England)*, 159:1423–1436.
- D'Heygère, F., Schwartz, A., Coste, F., Castaing, B. und Boudvillain, M. (2015) ATP-dependent motor activity of the transcription termination factor Rho from *Mycobacterium tuberculosis*. *Nucleic acids research*, 43:6099–6111.
- DiChiara, J. M., Contreras-Martinez, L. M., Livny, J., Smith, D., McDonough, K. A. und Belfort, M. (2010) Multiple small RNAs identified in *Mycobacterium bovis* BCG are also expressed in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Nucleic acids research*, 38:4067–4078.
- Dominski, Z., Carpousis, A. J. und Clouet-d'Orval, B. (2013) Emergence of the β -CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochim Biophys Acta*, 1829(6-7):532–551.
- Dorléans, A., Li de la Sierra-Gallay, I., Piton, J., Zig, L., Gilet, L., Putzer, H. und Condon, C. (2011) Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J. *Structure*, 19(9):1252–1261.
- Dornenburg, J. E., Devita, A. M., Palumbo, M. J. und Wade, J. T. (2010) Widespread antisense transcription in *Escherichia coli*. *mBio*, 1.
- Dreyfus, M. (2009) Killer and protective ribosomes. *Progress in molecular biology and translational science*, 85:423–466.
- Durand, S., Braun, F., Lioliou, E., Romilly, C., Helfer, A.-C., Kuhn, L., Quittot, N., Nicolas, P., Romby, P. und Condon, C. (2015) A nitric oxide regulated small RNA controls expression of genes involved in redox homeostasis in *Bacillus subtilis*. *PLoS genetics*, 11:e1004957.
- Durand, S. und Condon, C. (2018) RNases and Helicases in Gram-Positive Bacteria. *Microbiology spectrum*, 6.
- Durand, S., Gilet, L., Bessières, P., Nicolas, P. und Condon, C. (2012a) Three essential ribonucleases-RNase Y, J1, and III-control the abundance of a majority of *Bacillus subtilis* mRNAs. *PLoS Genet*, 8(3):e1002520.
- Durand, S., Gilet, L. und Condon, C. (2012b) The essential function of *B. subtilis* RNase III is to silence foreign toxin genes. *PLoS genetics*, 8:e1003181.

- Eggeling, L., Besra, G. S. und Alderwick, L. (2008) *Corynebacteria: Genomics and Molecular Biology*, Kapitel Structure and Synthesis of the Cell Wall in Corynebacteria, Seiten 267–294. Caister Academic Press.
- Eggeling, L. und Bott, M. (2005) *Handbook of Corynebacterium glutamicum*. CRC press.
- Ehretsmann, C. P., Carpousis, A. J. und Krisch, H. M. (1992) Specificity of Escherichia coli endoribonuclease RNase E: in vivo and in vitro analysis of mutants in a bacteriophage T4 mRNA processing site. *Genes & development*, 6:149–159.
- Eichenberger, P., Fujita, M., Jensen, S. T., Conlon, E. M., Rudner, D. Z., Wang, S. T., Ferguson, C., Haga, K., Sato, T., Liu, J. S. und Losick, R. (2004) The program of gene transcription for a single differentiating cell type during sporulation in Bacillus subtilis. *PLoS biology*, 2:e328.
- Eikmanns, B. J., Kleinertz, E., Liebl, W. und Sahm, H. (1991) A family of Corynebacterium glutamicum/Escherichia coli shuttle vectors for cloning, controlled gene expression, and promoter probing. *Gene*, 102(1):93–98.
- Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Bréchemmier-Baey, D. und Putzer, H. (2005) Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E. *Nucleic Acids Res*, 33(7):2141–2152.
- Ewels, P., Magnusson, M., Lundin, S. und Käller, M. (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, 32:3047–3048.
- Figaro, S., Durand, S., Gilet, L., Cayet, N., Sachse, M. und Condon, C. (2013) Bacillus subtilis mutants with knockouts of the genes encoding ribonucleases RNase Y and RNase J1 are viable, with major defects in cell morphology, sporulation, and competence. *J Bacteriol*, 195(10):2340–2348.
- Freeman, V. J. (1951) Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae. *Journal of bacteriology*, 61(6):675.
- Frohman, M. A. (1994) On beyond classic RACE (rapid amplification of cDNA ends). *PCR Methods Appl*, 4:S40–S58.
- Funke, G., Stubbs, S., Altwegg, M., Carlotti, A. und Collins, M. D. (1994) Turicella otitidis gen. nov., sp. nov., a coryneform bacterium isolated from patients with otitis media. *International journal of systematic bacteriology*, 44:270–273.

- Georg, J. und Hess, W. R. (2011) cis-antisense RNA, another level of gene regulation in bacteria. *Microbiology and molecular biology reviews : MMBR*, 75:286–300.
- Gertz, S., Engelmann, S., Schmid, R., Ziebandt, A. K., Tischer, K., Scharf, C., Hacker, J. und Hecker, M. (2000) Characterization of the sigma(B) regulon in *Staphylococcus aureus*. *Journal of bacteriology*, 182:6983–6991.
- Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. und Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, 6:343–345.
- Giltner, C. L., Nguyen, Y. und Burrows, L. L. (2012) Type IV Pilin Proteins: Versatile Molecular Modules. *Microbiology and Molecular Biology Reviews*, 76(4):740–772.
- Goodfellow, M. (2012) *Bergey's Manual of Systematic Bacteriology*, Band 5, Kapitel Phylum XXVI. Actinobacteria phyl. nov., Seiten 33–34. Springer New-York.
- Goodfellow, M. und Jones, A. L. (2012) *Bergey's Manual of Systematic Bacteriology*, Band 5, Kapitel Order V. Corynebacteriales ord. nov., Seiten 235–243. Springer New-York.
- Gourse, R. L., Takebe, Y., Sharrock, R. A. und Nomura, M. (1985) Feedback regulation of rRNA and tRNA synthesis and accumulation of free ribosomes after conditional expression of rRNA genes. *Proc Natl Acad Sci U S A*, 82(4):1069–1073.
- Grant, S. G., Jessee, J., Bloom, F. R. und Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proceedings of the National Academy of Sciences of the United States of America*, 87:4645–4649.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. und Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 Pt 2):849–857.
- Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A.-C., Bork, P. und Serrano, L. (2009) Transcriptome complexity in a genome-reduced bacterium. *Science (New York, N.Y.)*, 326:1268–1271.
- Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology*, 166:557–580.
- Hansen, K. D., Brenner, S. E. und Dudoit, S. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res*, 38(12):e131.

- Harris, M. E. und Christian, E. L. (2009) RNA crosslinking methods. *Methods in enzymology*, 468:127–146.
- Hart, C. M. und Roberts, J. W. (1991) Rho-dependent transcription termination. Characterization of the requirement for cytidine in the nascent transcript. *J Biol Chem*, 266(35):24140–24148.
- Hausmann, S., Guimarães, V. A., Garcin, D., Baumann, N., Linder, P. und Redder, P. (2017) Both exo- and endo-nucleolytic activities of RNase J1 from *Staphylococcus aureus* are manganese dependent and active on triphosphorylated 5'-ends. *RNA biology*, Seite 0.
- Hayes, F. und Vasseur, M. (1976) Processing of the 17-S *Escherichia coli* precursor RNA in the 27-S pre-ribosomal particle. *European journal of biochemistry*, 61:433–442.
- Helmann, J. D. (1995) Compilation and analysis of *Bacillus subtilis* sigma A-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic acids research*, 23:2351–2360.
- Helmann, J. D. (2002) The extracytoplasmic function (ECF) sigma factors. *Advances in microbial physiology*, 46:47–110.
- Helmann, J. D. (2009) RNA polymerase: a nexus of gene regulation. *Methods (San Diego, Calif.)*, 47:1–5.
- Hilker, R., Stadermann, K. B., Doppmeier, D., Kalinowski, J., Stoye, J., Straube, J., Winnebold, J. und Goesmann, A. (2014) ReadXplorer—visualization and analysis of mapped sequences. *Bioinformatics*, 30(16):2247–2254.
- Hillemann, D., Dammann, T., Hillemann, A. und Wohlleben, W. (1993) Genetic and biochemical characterization of the two glutamine synthetases GSI and GSII of the phosphinothricyl-alanyl-alanine producer, *Streptomyces viridochromogenes* Tü494. *Journal of general microbiology*, 139:1773–1783.
- Hirasawa, T. und Shimizu, H. (2016) Recent advances in amino acid production by microbial cells. *Current opinion in biotechnology*, 42:133–146.
- Hoch, D. H., Romero-Mira, M., Ehrlich, B. E., Finkelstein, A., DasGupta, B. R. und Simpson, L. L. (1985) Channels formed by botulinum, tetanus, and diphtheria toxins in planar lipid bilayers: relevance to translocation of proteins across membranes. *Proc Natl Acad Sci U S A*, 82:1692–1696.

- Hoffmann, S., Otto, C., Kurtz, S., Sharma, C. M., Khaitovich, P., Vogel, J., Stadler, P. F. und Hackermüller, J. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS computational biology*, 5(9):e1000502.
- Holmes, R. K. (2000) Biology and molecular epidemiology of diphtheria toxin and the tox gene. *The Journal of infectious diseases*, 181 Suppl 1:S156–S167.
- Hoskisson, P. A. (2018) Microbe Profile: *Corynebacterium diphtheriae* - an old foe always ready to seize opportunity. *Microbiology (Reading, England)*.
- Hsu, L. M. (2009) Monitoring abortive initiation. *Methods (San Diego, Calif.)*, 47:25–36.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C. und Bork, P. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic acids research*, 44:D286–D293.
- Ikeda, M. und Nakagawa, S. (2003) The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Applied microbiology and biotechnology*, 62(2-3):99–109.
- Illumina (2017) *An Introduction to Next-Generation Sequencing Technology*. Illumina Inc.
- Imam, S., Chen, Z., Roos, D. S. und Pohlschröder, M. (2011) Identification of surprisingly diverse type IV pili, across a broad range of gram-positive bacteria. *PLoS One*, 6(12):e28919.
- Inoue, H., Nojima, H. und Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96:23–28.
- Irla, M., Neshat, A., Brautaset, T., Rückert, C., Kalinowski, J. und Wendisch, V. F. (2015) Transcriptome analysis of thermophilic methylotrophic *Bacillus methanolicus* MGA3 using RNA-sequencing provides detailed insights into its previously uncharted transcriptional landscape. *BMC Genomics*, 16:73.
- Jacob, F. und Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, 3:318–356.
- Jain, C. und Belasco, J. G. (1995) Autoregulation of RNase E synthesis in *Escherichia coli*. *Nucleic acids symposium series*, Seiten 85–88.
- Jamalli, A., Hébert, A., Zig, L. und Putzer, H. (2014) Control of expression of the RNases J1 and J2 in *Bacillus subtilis*. *J Bacteriol*, 196(2):318–324.

- Jarrige, A. C., Mathy, N. und Portier, C. (2001) PNPase autocontrols its expression by degrading a double-stranded structure in the pnp mRNA leader. *The EMBO journal*, 20:6845–6855.
- Jenjaroenpun, P., Wongsurawat, T., Pereira, R., Patumcharoenpol, P., Ussery, D. W., Nielsen, J. und Nookaew, I. (2018) Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D. *Nucleic Acids Research*, 46(7):e38–e38.
- Jiang, X., Diwa, A. und Belasco, J. G. (2000) Regions of RNase E important for 5'-end-dependent RNA cleavage and autoregulated synthesis. *Journal of bacteriology*, 182:2468–2475.
- Johnston, C., Martin, B., Fichant, G., Polard, P. und Claverys, J.-P. (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature reviews. Microbiology*, 12:181–196.
- Joyce, S. A. und Dreyfus, M. (1998) In the absence of translation, RNase E can bypass 5' mRNA stabilizers in *Escherichia coli*. *Journal of molecular biology*, 282:241–254.
- Kaberdin, V. R. (2003) Probing the substrate specificity of *Escherichia coli* RNase E using a novel oligonucleotide-based assay. *Nucleic acids research*, 31:4710–4716.
- Kaberdin, V. R. und Bizebard, T. (2005) Characterization of *Aquifex aeolicus* RNase E/G. *Biochemical and biophysical research communications*, 327:382–392.
- Kaczanowska, M. und Rydén-Aulin, M. (2007) Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol Mol Biol Rev*, 71(3):477–494.
- Kaczorek, M., Zettlmeissl, G., Delpyroux, F. und Streeck, R. (1985) Diphtheria toxin promoter function in *Corynebacterium diphtheriae* and *Escherichia coli*. *Nucleic acids research*, 13(9):3147–3159.
- Kalinowski, J. (2005) *Handbook of Corynebacterium glutamicum*, Kapitel 3, Seiten 37–57s. CRC Press.
- Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Krämer, R., Linke, B., McHardy, A. C., Meyer, F., Möckel, B., Pfefferle, W., Pühler, A., Rey, D. A., Rückert, C., Rupp, O., Sahm, H., Wendisch, V. F., Wiegräbe, I. und Tauch, A. (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol*, 104(1-3):5–25.

- Kantardjieff, K., Collier, R. und Eisenberg, D. (1989) X-ray grade crystals of the enzymatic fragment of diphtheria toxin. *Journal of Biological Chemistry*, 264(18):10402–10404.
- Kehl-Fie, T. E. und Skaar, E. P. (2010) Nutritional immunity beyond iron: a role for manganese and zinc. *Current opinion in chemical biology*, 14:218–224.
- Keilhauer, C., Eggeling, L. und Sahm, H. (1993) Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. *Journal of Bacteriology*, 175(17):5595–5603.
- Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y. und Segal, E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, 467(7311):103–107.
- Khemici, V., Prados, J., Linder, P. und Redder, P. (2015) Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. *PLoS genetics*, 11:e1005577.
- Khemici, V., Prados, J., Linder, P. und Redder, P. (2016) Correction: Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. *PLoS genetics*, 12:e1006320.
- Kinoshita, S., Nakayama, K. und Akita, S. (1958) Taxonomical study of glutamic acid accumulating bacteria, *Micrococcus glutamicus* nov. sp. *Journal of the Agricultural Chemical Society of Japan*, 22(3):176–185.
- Kinoshita, S., Udaka, S. und Shimono, M. (2004) Studies on the amino acid fermentation. Part 1. Production of L-glutamic acid by various microorganisms. *The Journal of general and applied microbiology*, 50:331–343.
- Kirchner, O. und Tauch, A. (2003) Tools for genetic engineering in the amino acid-producing bacterium *Corynebacterium glutamicum*. *J Biotechnol*, 104(1-3):287–299.
- Kohn, H. und Widger, W. (2005) The molecular basis for the mode of action of bicyclomycin. *Current drug targets. Infectious disorders*, 5:273–295.
- Koide, T., Reiss, D. J., Bare, J. C., Pang, W. L., Facciotti, M. T., Schmid, A. K., Pan, M., Marzolf, B., Van, P. T., Lo, F.-Y., Pratap, A., Deutsch, E. W., Peterson, A., Martin, D. und Baliga, N. S. (2009) Prevalence of transcription promoters within archaeal operons and coding sequences. *Molecular systems biology*, 5:285.

- Kolter, R. und Yanofsky, C. (1982) Attenuation in amino acid biosynthetic operons. *Annual review of genetics*, 16(1):113–134.
- Koslover, D. J., Fazal, F. M., Mooney, R. A., Landick, R. und Block, S. M. (2012) Binding and translocation of termination factor rho studied at the single-molecule level. *Journal of molecular biology*, 423:664–676.
- Kotlajich, M. V., Hron, D. R., Boudreau, B. A., Sun, Z., Lyubchenko, Y. L. und Landick, R. (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife*, 4.
- Kovacs, L., Csanadi, A., Megyeri, K., Kaberdin, V. R. und Miczak, A. (2005) Mycobacterial RNase E-associated proteins. *Microbiology and immunology*, 49:1003–1007.
- Kram, K. E., Hovel-Miner, G. A., Tomich, M. und Figurski, D. H. (2008) Transcriptional regulation of the tad locus in *Aggregatibacter actinomycetemcomitans*: a termination cascade. *J Bacteriol*, 190(11):3859–3868.
- Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strömbom, L., Ståhlberg, A. und Zoric, N. (2006) The real-time polymerase chain reaction. *Mol Aspects Med*, 27:95–125.
- Kunkle, C. A. und Schmitt, M. P. (2003) Analysis of the *Corynebacterium diphtheriae* DtxR regulon: identification of a putative siderophore synthesis and transport system that is similar to the *Yersinia* high-pathogenicity island-encoded yersiniabactin synthesis and uptake system. *Journal of bacteriology*, 185(23):6826–6840.
- Kunkle, C. A. und Schmitt, M. P. (2005) Analysis of a DtxR-regulated iron transport and siderophore biosynthesis gene cluster in *Corynebacterium diphtheriae*. *J Bacteriol*, 187:422–433.
- Laalami, S., Bessières, P, Rocca, A., Zig, L., Nicolas, P und Putzer, H. (2013) *Bacillus subtilis* RNase Y activity in vivo analysed by tiling microarrays. *PLoS One*, 8(1):e54062.
- Laalami, S. und Putzer, H. (2011) mRNA degradation and maturation in prokaryotes: the global players. *Biomolecular concepts*, 2:491–506.
- Laalami, S., Zig, L. und Putzer, H. (2014) Initiation of mRNA decay in bacteria. *Cellular and Molecular Life Sciences*, 71(10):1799–1828.
- Ladokhin, A. S. (2013) pH-triggered conformational switching along the membrane insertion pathway of the diphtheria toxin T-domain. *Toxins*, 5:1362–1380.

- Lane, W. J. und Darst, S. A. (2006) The structural basis for promoter -35 element recognition by the group IV sigma factors. *PLoS biology*, 4:e269.
- Langmead, B. und Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9:357–359.
- Laurenceau, R., Péhau-Arnaudet, G., Bacconnais, S., Gault, J., Malosse, C., Dujeancourt, A., Campo, N., Chamot-Rooke, J., Le Cam, E., Claverys, J.-P. und Fronzes, R. (2013) A type IV pilus mediates DNA binding during natural transformation in *Streptococcus pneumoniae*. *PLoS pathogens*, 9:e1003473.
- Lawrence, J. G. und Roth, J. R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*, 143:1843–1860.
- Lee, K., Bernstein, J. A. und Cohen, S. N. (2002) RNase G complementation of *rne* null mutation identifies functional interrelationships with RNase E in *Escherichia coli*. *Mol Microbiol*, 43(6):1445–1456.
- Lee, K. und Cohen, S. N. (2003) A *Streptomyces coelicolor* functional orthologue of *Escherichia coli* RNase E shows shuffling of catalytic and PNPase-binding domains. *Molecular microbiology*, 48:349–360.
- Leela, J. K., Syeda, A. H., Anupama, K. und Gowrishankar, J. (2013) Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 110:258–263.
- Lehmann, K. B. und Neumann, R. O. (1896) *Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik*. München, J.F. Lehmann.
- Lehnik-Habrink, M., Newman, J., Rothe, F. M., Solovyova, A. S., Rodrigues, C., Herzberg, C., Commichau, F. M., Lewis, R. J. und Stülke, J. (2011) RNase Y in *Bacillus subtilis*: a Natively disordered protein that is the functional equivalent of RNase E from *Escherichia coli*. *Journal of bacteriology*, 193:5431–5441.
- Lehnik-Habrink, M., Pförtner, H., Rempeters, L., Pietack, N., Herzberg, C. und Stülke, J. (2010) The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex. *Mol Microbiol*, 77(4):958–971.
- Leong, D. und Murphy, J. R. (1985) Characterization of the diphtheria tox transcript in *Corynebacterium diphtheriae* and *Escherichia coli*. *Journal of bacteriology*, 163:1114–1119.

- Li, S., Dong, X. und Su, Z. (2013) Directional RNA-seq reveals highly complex condition-dependent transcriptomes in *E. coli* K12 through accurate full-length transcripts assembling. *BMC genomics*, 14:520.
- Li, Z., Pandit, S. und Deutscher, M. P. (1999) RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. *The EMBO journal*, 18:2878–2885.
- Li, J., Rodnin, M. V., Ladokhin, A. S. und Gross, M. L. (2014) Hydrogen–deuterium exchange and mass spectrometry reveal the pH-dependent conformational changes of diphtheria toxin T domain. *Biochemistry*, 53(43):6849–6856.
- Li de la Sierra-Gallay, I., Zig, L., Jamalli, A. und Putzer, H. (2008) Structural insights into the dual activity of RNase J. *Nat Struct Mol Biol*, 15(2):206–212.
- Liao, Y., Smyth, G. K. und Shi, W. (2013) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930.
- Liebl, W., Bayerl, A., Schein, B., Stillner, U. und Schleifer, K. (1989) High efficiency electroporation of intact *Corynebacterium glutamicum* cells. *FEMS microbiology letters*, 65(3):299–303.
- Lin-Chao, S., Wei, C. L. und Lin, Y. T. (1999) RNase E is required for the maturation of *ssrA* RNA and normal *ssrA* RNA peptide-tagging activity. *Proceedings of the National Academy of Sciences of the United States of America*, 96:12406–12411.
- Linder, P., Lemeille, S. und Redder, P. (2014) Transcriptome-wide analyses of 5'-ends in RNase J mutants of a gram-positive pathogen reveal a role in RNA maturation, regulation and degradation. *PLoS Genet*, 10(2):e1004207.
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H. und Ecker, J. R. (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 133:523–536.
- Lloréns-Rico, V., Cano, J., Kamminga, T., Gil, R., Latorre, A., Chen, W.-H., Bork, P., Glass, J. I., Serrano, L. und Lluch-Senar, M. (2016) Bacterial antisense RNAs are mainly the product of transcriptional noise. *Science advances*, 2:e1501363. Original DateCompleted: 20160314.
- Löchelt, M., Muranyi, W. und Flügel, R. (1993) Human foamy virus genome possesses an internal, Bel-1-dependent and functional promoter. *Proceedings of the National Academy of Sciences*, 90(15):7317–7321.
- Louie, G. V., Yang, W., Bowman, M. E. und Choe, S. (1997) Crystal structure of the complex of diphtheria toxin with an extracellular fragment of its receptor. *Molecular cell*, 1(1):67–78.

- Love, M. I., Huber, W. und Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12):550.
- Lowery-Goldhammer, C. und Richardson, J. P. (1974) An RNA-dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. *Proceedings of the National Academy of Sciences of the United States of America*, 71:2003–2007.
- Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., Feng, Q., Zhao, Y., Guo, Y., Li, W., Huang, X. und Han, B. (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome research*, 20:1238–1249.
- Luhtala, N. und Parker, R. (2010) T2 Family ribonucleases: ancient enzymes with diverse roles. *Trends in biochemical sciences*, 35:253–259.
- Mackie, G. A. (1991) Specific endonucleolytic cleavage of the mRNA for ribosomal protein S20 of *Escherichia coli* requires the product of the *ams* gene in vivo and in vitro. *Journal of bacteriology*, 173:2488–2497.
- Mackie, G. A. (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature*, 395:720–723.
- Mackie, G. A. (2013) RNase E: at the interface of bacterial RNA processing and decay. *Nat Rev Microbiol*, 11(1):45–57.
- Maeda, T., Sakai, T. und Wachi, M. (2009) The *Corynebacterium glutamicum* NCgl2281 gene encoding an RNase E/G family endoribonuclease can complement the *Escherichia coli* *rng::cat* mutation but not the *rne-1* mutation. *Biosci Biotechnol Biochem*, 73(10):2281–2286.
- Maeda, T., Tanaka, Y. und Inui, M. (2018) Glutamine-rich toxic proteins GrtA, GrtB, and GrtC together with the antisense RNA AsgR constitute a toxin-antitoxin-like system in *Corynebacterium glutamicum*. *Molecular microbiology*.
- Maeda, T., Tanaka, Y., Takemoto, N., Hamamoto, N. und Inui, M. (2015) RNase III mediated cleavage of the coding region of *mraZ* mRNA is required for efficient cell division in *Corynebacterium glutamicum*. *Mol Microbiol*.
- Maeda, T. und Wachi, M. (2012a) 3' Untranslated region-dependent degradation of the *aceA* mRNA, encoding the glyoxylate cycle enzyme isocitrate lyase, by RNase E/G in *Corynebacterium glutamicum*. *Appl Environ Microbiol*, 78(24):8753–8761.

- Maeda, T. und Wachi, M. (2012b) *Corynebacterium glutamicum* RNase E/G-type endoribonuclease encoded by NCgl2281 is involved in the 5' maturation of 5S rRNA. *Arch Microbiol*, 194(2):65–73.
- Mao, X., Ma, Q., Zhou, C., Chen, X., Zhang, H., Yang, J., Mao, F., Lai, W. und Xu, Y. (2013) DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic acids research*, 42(D1):D654–D659.
- Marincola, G. und Wolz, C. (2017) Downstream element determines RNase Y cleavage of the saePQRS operon in *Staphylococcus aureus*. *Nucleic acids research*, 45:5980–5994.
- Mathy, N., Bénard, L., Pellegrini, O., Daou, R., Wen, T. und Condon, C. (2007) 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell*, 129(4):681–692.
- Mathy, N., Hébert, A., Mervelet, P., Bénard, L., Dorléans, A., Li de la Sierra-Gallay, I., Noirot, P., Putzer, H. und Condon, C. (2010) *Bacillus subtilis* ribonucleases J1 and J2 form a complex with altered enzyme behaviour. *Mol Microbiol*, 75(2):489–498.
- Matsunaga, J., Simons, E. L. und Simons, R. W. (1996) RNase III autoregulation: structure and function of rncO, the posttranscriptional operator". *RNA (New York, N.Y.)*, 2:1228–1240.
- McDowall, K. J., Hernandez, R. G., Lin-Chao, S. und Cohen, S. N. (1993) The ams-1 and rne-3071 temperature-sensitive mutations in the ams gene are in close proximity to each other and cause substitutions within a domain that resembles a product of the *Escherichia coli* mre locus. *Journal of bacteriology*, 175:4245–4249.
- McSwiggen, J. A., Bear, D. G. und von Hippel, P. H. (1988) Interactions of *Escherichia coli* transcription termination factor rho with RNA. I. Binding stoichiometries and free energies. *Journal of molecular biology*, 199:609–622.
- Melville, S. und Craig, L. (2013) Type IV Pili in Gram-Positive Bacteria. *Microbiology and Molecular Biology Reviews*, 77(3):323–341.
- Menouni, R., Champ, S., Espinosa, L., Boudvillain, M. und Ansaldi, M. (2013) Transcription termination controls prophage maintenance in *Escherichia coli* genomes. *Proc Natl Acad Sci U S A*, 110(35):14414–14419.
- Mentz, A., Neshat, A., Pfeifer-Sancar, K., Pühler, A., Rückert, C. und Kalinowski, J. (2013) Comprehensive discovery and characterization of small RNAs in *Corynebacterium glutamicum* ATCC 13032. *BMC Genomics*, 14(1):714.

- Merchant, A. T. und Spatafora, G. A. (2014) A role for the DtxR family of metalloregulators in gram-positive pathogenesis. *Molecular oral microbiology*, 29(1):1–10.
- Meyer, F., Goesmann, A., McHardy, A. C., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O., Giegerich, R. *et al.* (2003) GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic acids research*, 31(8):2187–2195.
- Miller, H. K., Carroll, R. K., Burda, W. N., Krute, C. N., Davenport, J. E. und Shaw, L. N. (2012) The extracytoplasmic function sigma factor σ^S protects against both intracellular and extracytoplasmic stresses in *Staphylococcus aureus*. *Journal of bacteriology*, 194:4342–4354.
- Misra, T. K. und Apirion, D. (1979) RNase E, an RNA processing enzyme from *Escherichia coli*. *The Journal of biological chemistry*, 254:11154–11159.
- Mitra, A., Misquitta, R. und Nagaraja, V. (2014) *Mycobacterium tuberculosis* Rho is an NT-Pase with distinct kinetic properties and a novel RNA-binding subdomain. *PLoS One*, 9(9):e107474.
- Moll, I., Hirokawa, G., Kiel, M. C., Kaji, A. und Bläsi, U. (2004) Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res*, 32(11):3354–3363.
- Morikawa, K., Inose, Y., Okamura, H., Maruyama, A., Hayashi, H., Takeyasu, K. und Ohta, T. (2003) A new staphylococcal sigma factor in the conserved gene cassette: functional significance and implication for the evolutionary processes. *Genes to cells : devoted to molecular & cellular mechanisms*, 8:699–712.
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. und Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*, 5(7):621–628.
- Motherway, M. O., Zomer, A., Leahy, S. C., Reunanen, J., Bottacini, F., Claesson, M. J., O'Brien, F., Flynn, K., Casey, P. G., Munoz, J. A. M., Kearney, B., Houston, A. M., O'Mahony, C., Higgins, D. G., Shanahan, F., Palva, A., de Vos, W. M., Fitzgerald, G. F., Ventura, M., O'Toole, P. W. und van Sinderen, D. (2011) Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proceedings of the National Academy of Sciences of the United States of America*, 108:11217–11222.
- Mudd, E. A. und Higgins, C. F. (1993) *Escherichia coli* endoribonuclease RNase E: autoregulation of expression and site-specific cleavage of mRNA. *Molecular microbiology*, 9:557–568.

- Mäder, U., Nicolas, P., Depke, M., Pané-Farré, J., Debarbouille, M., van der Kooi-Pol, M. M., Guérin, C., Dérozier, S., Hiron, A., Jarmer, H., Leduc, A., Michalik, S., Reilman, E., Schaffer, M., Schmidt, F., Bessières, P., Noirot, P., Hecker, M., Msadek, T., Völker, U. und van Dijl, J. M. (2016) Staphylococcus aureus Transcriptome Architecture: From Laboratory to Infection-Mimicking Conditions. *PLoS genetics*, 12:e1005962.
- Mäder, U., Zig, L., Kretschmer, J., Homuth, G. und Putzer, H. (2008) mRNA processing by RNases J1 and J2 affects *Bacillus subtilis* gene expression on a global scale. *Mol Microbiol*, 70(1):183–196.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. und Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (New York, N.Y.)*, 320:1344–1349.
- Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., Floden, E. W., Gardner, P. P., Jones, T. A., Tate, J. und Finn, R. D. (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res*, 43:D130–D137.
- Nawrocki, E. P. und Eddy, S. R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 29:2933–2935.
- Needleman, S. B. und Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48:443–453.
- Neshat, A., Mentz, A., Rückert, C. und Kalinowski, J. (2014) Transcriptome sequencing revealed the transcriptional organization at ribosome-mediated attenuation sites in *Corynebacterium glutamicum* and identified a novel attenuator involved in aromatic amino acid biosynthesis. *J Biotechnol*.
- Nešvera, J., Pátek, M., Hochmannová, J., Abrhánová, Z., Becvárová, V., Jelínková, M. und Vohradský, J. (1997) Plasmid pGA1 from *Corynebacterium glutamicum* codes for a gene product that positively influences plasmid copy number. *J Bacteriol*, 179(5):1525–1532.
- Newman, J. A., Hewitt, L., Rodrigues, C., Solovyova, A., Harwood, C. R. und Lewis, R. J. (2011) Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure*, 19(9):1241–1251.
- Newman, J. A., Hewitt, L., Rodrigues, C., Solovyova, A. S., Harwood, C. R. und Lewis, R. J. (2012) Dissection of the network of interactions that links RNA processing with glycolysis in the *Bacillus subtilis* degradosome. *J Mol Biol*, 416(1):121–136.

- Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., Bidnenko, E., Marchadier, E., Hoebeke, M., Aymerich, S., Becher, D., Bisicchia, P., Botella, E., Delumeau, O., Doherty, G., Denham, E. L., Fogg, M. J., Fromion, V., Goelzer, A., Hansen, A., Härtig, E., Harwood, C. R., Homuth, G., Jarmer, H., Jules, M., Klipp, E., Le Chat, L., Lecointe, F., Lewis, P., Liebermeister, W., March, A., Mars, R. A. T., Nannapaneni, P., Noone, D., Pohl, S., Rinn, B., Rügheimer, F., Sappa, P. K., Samson, F., Schaffer, M., Schwikowski, B., Steil, L., Stülke, J., Wiegert, T., Devine, K. M., Wilkinson, A. J., van Dijl, J. M., Hecker, M., Völker, U., Bessières, P. und Noirot, P. (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science (New York, N.Y.)*, 335:1103–1106.
- Nolden, L., Beckers, G., Möckel, B., Pfefferle, W., Nampoothiri, K. M., Krämera, R. und Burkovskia, A. (2000) Urease of *Corynebacterium glutamicum*: organization of corresponding genes and investigation of activity. *FEMS microbiology letters*, 189:305–310.
- Nowatzke, W. L., Keller, E., Koch, G. und Richardson, J. P. (1997) Transcription termination factor Rho is essential for *Micrococcus luteus*. *Journal of bacteriology*, 179:5238–5240.
- Obana, N., Shirahama, Y., Abe, K. und Nakamura, K. (2010) Stabilization of *Clostridium perfringens* collagenase mRNA by VR-RNA-dependent cleavage in 5' leader sequence. *Molecular microbiology*, 77:1416–1428.
- Okonechnikov, K., Golosova, O. und Fursov, M. (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 28(8):1166–1167.
- Oram, D. M., Jacobson, A. D. und Holmes, R. K. (2006) Transcription of the contiguous sigB, dtxR, and galE genes in *Corynebacterium diphtheriae*: evidence for multiple transcripts and regulation by environmental factors. *J Bacteriol*, 188:2959–2973.
- Oussenko, I. A., Abe, T., Ujiiie, H., Muto, A. und Bechhofer, D. H. (2005) Participation of 3'-to-5' exoribonucleases in the turnover of *Bacillus subtilis* mRNA. *Journal of bacteriology*, 187:2758–2767.
- Ow, M. C. und Kushner, S. R. (2002) Initiation of tRNA maturation by RNase E is essential for cell viability in *E. coli*. *Genes & development*, 16:1102–1115.
- Oxford Nanopore (2017) *Direct RNA Sequencing Protocol for the MinION*. Oxford Nanopore.
- Pahlke, J. (2014) *The 6C RNA of Corynebacterium glutamicum*. Doktorarbeit, Heinrich Heine University Düsseldorf.

- Pahlke, J., Dostálová, H., Holátko, J., Degner, U., Bott, M., Pátek, M. und Polen, T. (2016) The small 6C RNA of *Corynebacterium glutamicum* is involved in the SOS response. *RNA Biol*, Seite 0.
- Pappin, D. J., Hojrup, P. und Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Current biology : CB*, 3:327–332.
- Park, J.-S. und Roberts, J. W. (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proceedings of the National Academy of Sciences of the United States of America*, 103:4870–4875.
- Passalacqua, K. D., Varadarajan, A., Ondov, B. D., Okou, D. T., Zwick, M. E. und Bergman, N. H. (2009) Structure and complexity of a bacterial transcriptome. *Journal of bacteriology*, 191:3203–3211.
- Peano, C., Pietrelli, A., Consolandi, C., Rossi, E., Petiti, L., Tagliabue, L., De Bellis, G. und Landini, P. (2013) An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microbial informatics and experimentation*, 3:1.
- Pei, X.-Y., Bralley, P., Jones, G. H. und Luisi, B. F. (2015) Linkage of catalysis and 5' end recognition in ribonuclease RNase J. *Nucleic Acids Res.*
- Perkins, T. T., Kingsley, R. A., Fookes, M. C., Gardner, P. P., James, K. D., Yu, L., Assefa, S. A., He, M., Croucher, N. J., Pickard, D. J., Maskell, D. J., Parkhill, J., Choudhary, J., Thomson, N. R. und Dougan, G. (2009) A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella typhi*. *PLoS genetics*, 5:e1000569.
- Perkins, D. N., Pappin, D. J., Creasy, D. M. und Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20:3551–3567.
- Peters, J. M., Mooney, R. A., Grass, J. A., Jessen, E. D., Tran, F. und Landick, R. (2012) Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev*, 26(23):2621–2633.
- Peters, J. M., Mooney, R. A., Kuan, P. F., Rowland, J. L., Keles, S. und Landick, R. (2009) Rho directs widespread termination of intragenic and stable RNA transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 106:15406–15411.
- Petersen, C. (1992) Control of functional mRNA stability in bacteria: multiple mechanisms of nucleolytic and non-nucleolytic inactivation. *Molecular microbiology*, 6:277–282.

- Pfeifer-Sancar, K., Mentz, A., Rückert, C. und Kalinowski, J. (2013) Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics*, 14(1):888.
- Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C. D. und Vogel, J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nature structural & molecular biology*, 16:840–846.
- Piepenbrink, K. H. und Sundberg, E. J. (2016) Motility and adhesion through type IV pili in Gram-positive bacteria. *Biochemical Society transactions*, 44:1659–1666.
- Pohl, E., Holmes, R. K. und Hol, W. G. (1998) Motion of the DNA-binding domain with respect to the core of the diphtheria toxin repressor (DtxR) revealed in the crystal structures of apo- and holo-DtxR. *The Journal of biological chemistry*, 273:22420–22427.
- Provvedi, R. und Dubnau, D. (1999) ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Molecular microbiology*, 31:271–280.
- Prévost, K., Desnoyers, G., Jacques, J.-F., Lavoie, F. und Massé, E. (2011) Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes & development*, 25:385–396.
- Pátek, M. und Nešvera, J. (2011) Sigma factors and promoters in *Corynebacterium glutamicum*. *Journal of biotechnology*, 154:101–113.
- Pátek, M., Nešvera, J., Guyonvarch, A., Reyes, O. und Leblon, G. (2003) Promoters of *Corynebacterium glutamicum*. *Journal of biotechnology*, 104:311–323.
- Qian, Y., Lee, J. H. und Holmes, R. K. (2002) Identification of a DtxR-regulated operon that is essential for siderophore-dependent iron uptake in *Corynebacterium diphtheriae*. *Journal of bacteriology*, 184(17):4846–4856.
- Qiu, Y., Cho, B.-K., Park, Y. S., Lovley, D., Palsson, B. Ø. und Zengler, K. (2010) Structural and operational complexity of the *Geobacter sulfurreducens* genome. *Genome research*, 20:1304–1311.
- Quirk, P. G., Dunkley, E. A., Lee, P. und Krulwich, T. A. (1993) Identification of a putative *Bacillus subtilis* rho gene. *Journal of bacteriology*, 175:647–654.
- Raabe, C. A., Hoe, C. H., Randau, G., Brosius, J., Tang, T. H. und Rozhdestvensky, T. S. (2011) The rocks and shallows of deep RNA sequencing: Examples in the *Vibrio cholerae* RNome. *RNA (New York, N.Y.)*, 17:1357–1366.

- Raghavan, R., Groisman, E. A. und Ochman, H. (2011) Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res*, 21:1487–1497.
- Raghavan, R., Sloan, D. B. und Ochman, H. (2012) Antisense transcription is pervasive but rarely conserved in enteric bacteria. *mBio*, 3.
- Ramsugit, S. und Pillay, M. (2015) Pili of *Mycobacterium tuberculosis*: current knowledge and future prospects. *Archives of microbiology*, 197:737–744.
- Ray-Soni, A., Bellecourt, M. J. und Landick, R. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem*, 2(85):319–47.
- Redko, Y., Aubert, S., Stachowicz, A., Lenormand, P., Namane, A., Darfeuille, F., Thibonnier, M. und De Reuse, H. (2013) A minimal bacterial RNase J-based degradosome is associated with translating ribosomes. *Nucleic Acids Res*, 41(1):288–301.
- Reguera, G., McCarthy, K. D., Mehta, T., Nicoll, J. S., Tuominen, M. T. und Lovley, D. R. (2005) Extracellular electron transfer via microbial nanowires. *Nature*, 435:1098–1101.
- Rice, P., Longden, I. und Bleasby, A. (2000) EMBL: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG*, 16:276–277.
- Richards, J., Liu, Q., Pellegrini, O., Celesnik, H., Yao, S., Bechhofer, D. H., Condon, C. und Belasco, J. G. (2011) An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in *Bacillus subtilis*. *Molecular cell*, 43:940–949.
- Roberts, J. und Park, J.-S. (2004) Mfd, the bacterial transcription repair coupling factor: translocation, repair and termination. *Current opinion in microbiology*, 7:120–125.
- Rodionov, D. A. (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chemical reviews*, 107(8):3467–3497.
- Rosenblad, M. A., Larsen, N., Samuelsson, T. und Zwieb, C. (2009) Kinship in the SRP RNA family. *RNA biology*, 6:508–516.
- Salvo, E., Alabi, S., Liu, B., Schlessinger, A. und Bechhofer, D. H. (2016) Interaction of *Bacillus subtilis* Polynucleotide Phosphorylase and RNase Y: STRUCTURAL MAPPING AND EFFECT ON mRNA TURNOVER. *The Journal of biological chemistry*, 291:6655–6663.
- Sambrook, J., Fritsch, E. F., Maniatis, T. et al. (2001) *Molecular cloning: a laboratory manual*. Ed. 2. Cold spring harbor laboratory press.

- Sangal, V. und Hoskisson, P. A. (2014) *Corynebacterium diphtheriae and Related Toxigenic Species: Genomics, Pathogenicity and Applications*, Kapitel Corynephages: Infections of the Infectors, Seiten 67–81. Springer Netherlands.
- Saramago, M., Bárria, C., Dos Santos, R. F., Silva, I. J., Pobre, V., Domingues, S., Andrade, J. M., Viegas, S. C. und Arraiano, C. M. (2014) The role of RNases in the regulation of small RNAs. *Current opinion in microbiology*, 18:105–115.
- Schaffer, S., Weil, B., Nguyen, V. D., Dongmann, G., Günther, K., Nickolaus, M., Hermann, T. und Bott, M. (2001) A high-resolution reference map for cytoplasmic and membrane-associated proteins of *Corynebacterium glutamicum*. *Electrophoresis*, 22:4404–4422.
- Schifano, J. M., Vvedenskaya, I. O., Knoblauch, J. G., Ouyang, M., Nickels, B. E. und Woychik, N. A. (2014) An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-mt3. *Nat Commun*, 5:3538.
- Schmitt, M. P. (1997) Transcription of the *Corynebacterium diphtheriae* hmuO gene is regulated by iron and heme. *Infect Immun*, 65(11):4634–4641.
- Schmitt, M. P. und Holmes, R. K. (1991) Iron-dependent regulation of diphtheria toxin and siderophore expression by the cloned *Corynebacterium diphtheriae* repressor gene dtxR in *C. diphtheriae* C7 strains. *Infection and immunity*, 59(6):1899–1904.
- Schoenfeld, A., Davidowitz, E. J. und Burk, R. D. (1998) A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor. *Proceedings of the National Academy of Sciences*, 95(15):8817–8822.
- Schröder, J., Maus, I., Trost, E. und Tauch, A. (2011) Complete genome sequence of *Corynebacterium variabile* DSM 44702 isolated from the surface of smear-ripened cheeses and insights into cheese ripening and flavor generation. *BMC genomics*, 12:545.
- Schubert, M., Edge, R. E., Lario, P., Cook, M. A., Strynadka, N. C. J., Mackie, G. A. und McIntosh, L. P. (2004) Structural characterization of the RNase E S1 domain and identification of its oligonucleotide-binding and dimerization interfaces. *Journal of molecular biology*, 341:37–54.
- Schuck, A., Diwa, A. und Belasco, J. G. (2009) RNase E autoregulates its synthesis in *Escherichia coli* by binding directly to a stem-loop in the rne 5' untranslated region. *Molecular microbiology*, 72:470–478.

- Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G. und Pühler, A. (1994) Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene*, 145(1):69–73.
- Seitz, P., Pezeshgi Modarres, H., Borgeaud, S., Bulushev, R. D., Steinbock, L. J., Radenovic, A., Dal Peraro, M. und Blokesch, M. (2014) ComEA is essential for the transfer of external DNA into the periplasm in naturally transformable *Vibrio cholerae* cells. *PLoS genetics*, 10:e1004066.
- Serganov, A. und Nudler, E. (2013) A decade of riboswitches. *Cell*, 152(1-2):17–24.
- Shahbadian, K., Jamalli, A., Zig, L. und Putzer, H. (2009) RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *The EMBO journal*, 28:3523–3533.
- Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P. F. und Vogel, J. (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, 464(7286):250–255.
- Shinhara, A., Matsui, M., Hiraoka, K., Nomura, W., Hirano, R., Nakahigashi, K., Tomita, M., Mori, H. und Kanai, A. (2011) Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. *BMC genomics*, 12:428.
- Silva, I. J., Saramago, M., Dressaire, C., Domingues, S., Viegas, S. C. und Arraiano, C. M. (2011) Importance and key events of prokaryotic RNA decay: the ultimate fate of an RNA molecule. *Wiley Interdiscip Rev RNA*, 2(6):818–836.
- Silverstone, A. E., Goman, M. und Scaife, J. G. (1972) ALT: a new factor involved in the synthesis of RNA by *Escherichia coli*. *Molecular & general genetics : MGG*, 118:223–234.
- Skordalakes, E. und Berger, J. M. (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell*, 114(1):135–146.
- Sorek, R. und Cossart, P. (2010) Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nature reviews. Genetics*, 11:9–16.
- Spickler, C. und Mackie, G. A. (2000) Action of RNase II and polynucleotide phosphorylase against RNAs containing stem-loops of defined structure. *Journal of bacteriology*, 182:2422–2427.

- Stackebrandt, E., Rainey, F. A. und Ward-Rainey, N. L. (1997) Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol*, 47:479–491.
- Stead, M. B., Marshburn, S., Mohanty, B. K., Mitra, J., Pena Castillo, L., Ray, D., van Bakel, H., Hughes, T. R. und Kushner, S. R. (2011) Analysis of *Escherichia coli* RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic acids research*, 39:3188–3203.
- Struhl, K. (2007) Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nature structural & molecular biology*, 14:103–105.
- Sun, L., Zhang, L., Zhang, H. und He, Z.-G. (2011) Characterization of a bifunctional β -lactamase/ribonuclease and its interaction with a chaperone-like protein in the pathogen *Mycobacterium tuberculosis* H37Rv. *Biochemistry. Biokhimiia*, 76:350–358.
- Takemoto, N., Tanaka, Y. und Inui, M. (2015) Rho and RNase play a central role in FMN riboswitch regulation in *Corynebacterium glutamicum*. *Nucleic Acids Res*, 43(1):520–529.
- Tamura, M., Honda, N., Fujimoto, H., Cohen, S. N. und Kato, A. (2016) PpsA-mediated alternative pathway to complement RNase E essentiality in *Escherichia coli*. *Arch Microbiol*.
- Taniguchi, H. und Wendisch, V. F. (2015) Exploring the role of sigma factor gene expression on production by *Corynebacterium glutamicum*: sigma factor H and FMN as example. *Front Microbiol*, 6:740.
- Tao, X., Schiering, N., Zeng, H. Y., Ringe, D. und Murphy, J. R. (1994) Iron, DtxR, and the regulation of diphtheria toxin expression. *Mol Microbiol*, 14:191–197.
- Taraseviciene, L., Miczak, A. und Apirion, D. (1991) The gene specifying RNase E (*rne*) and a gene affecting mRNA stability (*ams*) are the same gene. *Molecular microbiology*, 5:851–855.
- Tatusov, R. L., Koonin, E. V. und Lipman, D. J. (1997) A genomic perspective on protein families. *Science (New York, N.Y.)*, 278:631–637.
- Taub, F., DeLeo, J. M. und Thompson, E. B. (1983) Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated RNAs. *DNA*, 2(4):309–327.
- Tauch, A., Kassing, F., Kalinowski, J. und Pühler, A. (1995) The *Corynebacterium xerosis* composite transposon Tn5432 consists of two identical insertion sequences, designated IS1249, flanking the erythromycin resistance gene *ermCX*. *Plasmid*, 34:119–131.

- Tauch, A., Kirchner, O., Löffler, B., Götter, S., Pühler, A. und Kalinowski, J. (2002) Efficient electrotransformation of *Corynebacterium diphtheriae* with a mini-replicon derived from the *Corynebacterium glutamicum* plasmid pGA1. *Current microbiology*, 45:362–367.
- Tauch, A. und Sandbote, J. (2014) The Family Corynebacteriaceae. In *The Prokaryotes*, Seiten 239–277. Springer Berlin Heidelberg.
- Taverniti, V., Forti, F., Ghisotti, D. und Putzer, H. (2011) Mycobacterium smegmatis RNase J is a 5'-3' exo-/endoribonuclease and both RNase J and RNase E are involved in ribosomal RNA maturation. *Mol Microbiol*, 82(5):1260–1276.
- Thomason, M. K., Bischler, T., Eisenbart, S. K., Förstner, K. U., Zhang, A., Herbig, A., Nieselt, K., Sharma, C. M. und Storz, G. (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J Bacteriol*, 197(1):18–28.
- Thomason, M. K. und Storz, G. (2010) Bacterial antisense RNAs: how many are there, and what are they doing? *Annu Rev Genet*, 44:167–188.
- Thomsen, N. D. und Berger, J. M. (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell*, 139:523–534.
- Tock, M. R., Walsh, A. P., Carroll, G. und McDowall, K. J. (2000) The CafA protein required for the 5'-maturation of 16 S rRNA is a 5'-end-dependent ribonuclease that has context-dependent broad sequence specificity. *The Journal of biological chemistry*, 275:8726–8732.
- Tomich, M., Planet, P. J. und Figurski, D. H. (2007) The tad locus: postcards from the widespread colonization island. *Nat Rev Microbiol*, 5(5):363–375.
- Tran, L., van Baarsel, J. A., Washburn, R. S., Gottesman, M. E. und Miller, J. H. (2011) Single-gene deletion mutants of *Escherichia coli* with altered sensitivity to bicyclomycin, an inhibitor of transcription termination factor Rho. *Journal of bacteriology*, 193:2229–2235.
- Trost, E., Blom, J., Soares, S. d. C., Huang, I.-H., Al-Dilaimi, A., Schröder, J., Jaenicke, S., Dorella, F. A., Rocha, F. S., Miyoshi, A., Azevedo, V., Schneider, M. P., Silva, A., Camello, T. C., Sabbadini, P. S., Santos, C. S., Santos, L. S., Hirata, R., Jr, Mattos-Guaraldi, A. L., Efstratiou, A., Schmitt, M. P., Ton-That, H. und Tauch, A. (2012) Pangenomic study of *Corynebacterium diphtheriae* that provides insights into the genomic diversity of pathogenic isolates from cases of classical diphtheria, endocarditis, and pneumonia. *J Bacteriol*, 194(12):3199–3215.

- van Vliet, A. H. M. (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS microbiology letters*, 302:1–7.
- Varga, J. J., Nguyen, V., O'Brien, D. K., Rodgers, K., Walker, R. A. und Melville, S. B. (2006) Type IV pili-dependent gliding motility in the Gram-positive pathogen *Clostridium perfringens* and other Clostridia. *Molecular microbiology*, 62:680–694.
- Vassylyev, D. G., Sekine, S.-i., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. und Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*, 417:712–719.
- Vašicová, P., Abrhamova, Z., Nešvera, J., Pátek, M., Sahm, H. und Eikmanns, B. (1998) Integrative and autonomously replicating vectors for analysis of promoters in *Corynebacterium glutamicum*. *Biotechnology Techniques*, 12(10):743–746.
- Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F. und van Sinderen, D. (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and molecular biology reviews : MMBR*, 71:495–548.
- Vogel, J. und Luisi, B. F. (2011) Hfq and its constellation of RNA. *Nature reviews. Microbiology*, 9:578–589.
- Wachi, M., Umitsuki, G., Shimizu, M., Takada, A. und Nagai, K. (1999) Escherichia coli cafA gene encodes a novel RNase, designated as RNase G, involved in processing of the 5' end of 16S rRNA. *Biochemical and biophysical research communications*, 259:483–488.
- Walmacq, C., Rahmouni, A. R. und Boudvillain, M. (2004) Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. *Journal of molecular biology*, 342:403–420.
- Walter, F., Grenz, S., Ortseifen, V., Persicke, M. und Kalinowski, J. (2016) *Corynebacterium glutamicum* ggtB encodes a functional γ -glutamyl transpeptidase with γ -glutamyl dipeptide synthetic and hydrolytic activity. *Journal of biotechnology*, 232:99–109.
- Wang, J., Ai, X., Mei, H., Fu, Y., Chen, B., Yu, Z. und He, J. (2013) High-throughput identification of promoters and screening of highly active promoter-5'-UTR DNA region with different characteristics from *Bacillus thuringiensis*. *PLoS one*, 8:e62960.
- Wang, W. und Bechhofer, D. H. (1996) Properties of a *Bacillus subtilis* polynucleotide phosphorylase deletion strain. *Journal of bacteriology*, 178:2375–2382.

- Wang, Z., Gerstein, M. und Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10:57–63.
- Wang, S. T., Setlow, B., Conlon, E. M., Lyon, J. L., Imamura, D., Sato, T., Setlow, P., Losick, R. und Eichenberger, P. (2006) The forespore line of gene expression in *Bacillus subtilis*. *Journal of molecular biology*, 358:16–37.
- Washburn, R. S., Marra, A., Bryant, A. P., Rosenberg, M. und Gentry, D. R. (2001) rho is not essential for viability or virulence in *Staphylococcus aureus*. *Antimicrobial agents and chemotherapy*, 45:1099–1103.
- Weinberg, Z., Wang, J. X., Bogue, J., Yang, J., Corbino, K., Moy, R. H. und Breaker, R. R. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome biology*, 11(3):R31.
- Williams, R. M. und Durham, C. A. (1988) Bicyclomycin: synthetic, mechanistic, and biological studies. *Chemical Reviews*, 88(3):511–540.
- Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. und Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428:281–286.
- Wittchen, M., Busche, T., Gaspar, A. H., Lee, J. H., Ton-That, H., Kalinowski, J. und Tauch, A. (2018) Transcriptome sequencing of the human pathogen *Corynebacterium diphtheriae* NCTC 13129 provides detailed insights into its transcriptional landscape and into DtxR-mediated transcriptional regulation. *BMC Genomics*, 19:82.
- Wolf, T., Schneiker-Bekel, S., Neshat, A., Ortseifen, V., Wibberg, D., Zemke, T., Pühler, A. und Kalinowski, J. (2017) Genome improvement of the acarbose producer *Actinoplanes* sp. SE50/110 and annotation refinement based on RNA-seq analysis. *Journal of biotechnology*, 251:112–123.
- Xie, P. und Chen, H. (2017) Mechanism of ribosome translation through mRNA secondary structures. *International journal of biological sciences*, 13:712–722.
- Yao, S., Blaustein, J. B. und Bechhofer, D. H. (2007) Processing of *Bacillus subtilis* small cytoplasmic RNA: evidence for an additional endonuclease cleavage site. *Nucleic acids research*, 35:4464–4473.
- Yehudai-Resheff, S. und Schuster, G. (2000) Characterization of the *E. coli* poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic acids research*, 28:1139–1144.

- Yokota, T., Omori, T. und Kodama, T. (1987) Purification and properties of haloalkane dehalogenase from *Corynebacterium* sp. strain m15-3. *Journal of bacteriology*, 169:4049–4054.
- Zalatan, F. und Platt, T. (1992) Effects of decreased cytosine content on rho interaction with the rho-dependent terminator trp t' in *Escherichia coli*. *The Journal of biological chemistry*, 267:19082–19088.
- Zeller, M.-E., Csanadi, A., Miczak, A., Rose, T., Bizebard, T. und Kaberdin, V. R. (2007) Quaternary structure and biochemical properties of mycobacterial RNase E/G. *Biochem J*, 403(1):207–215.
- Zemanová, M., Kaderábková, P., Pátek, M., Knoppová, M., Silar, R. und Nesvera, J. (2008) Chromosomally encoded small antisense RNA in *Corynebacterium glutamicum*. *FEMS microbiology letters*, 279:195–201.
- Zhang, B., Zhou, N., Liu, Y.-M., Liu, C., Lou, C.-B., Jiang, C.-Y. und Liu, S.-J. (2015) Ribosome binding site libraries and pathway modules for shikimic acid synthesis with *Corynebacterium glutamicum*. *Microbial cell factories*, 14:71.
- Zhelkovsky, A. M. und McReynolds, L. A. (2012) Structure-function analysis of *Methanobacterium thermoautotrophicum* RNA ligase - engineering a thermostable ATP independent enzyme. *BMC molecular biology*, 13:24.
- Zheng, X., Hu, G.-Q., She, Z.-S. und Zhu, H. (2011) Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. *BMC genomics*, 12:361.
- Zhi, X.-Y., Li, W.-J. und Stackebrandt, E. (2009) An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *International journal of systematic and evolutionary microbiology*, 59:589–608.

Tabellenverzeichnis

3.1	Liste der verwendeten Stämme von <i>Corynebacterium diphtheriae</i> und <i>Corynebacterium glutamicum</i>	33
3.2	Liste der verwendeten Stämme von <i>E. coli</i>	34
3.3	Verwendete Primer.	34
3.4	Liste der verwendeten Plasmide.	36
3.5	Verwendete Chemikalien und deren Bezugsquelle.	36
3.6	Verwendete Materialien und deren Bezugsquelle.	38
3.7	Verwendete Kits und deren Bezugsquelle.	38
3.8	Verwendete Geräte für die generelle Laborarbeit und deren Hersteller.	38
3.9	Verwendete Software. Weitere Software ist in den jeweiligen Abschnitten genannt.	39
3.10	Verwendete Medien zur Kultivierung von Bakterien.	40
3.11	Medien-Zusätze	41
3.13	Pipettierschema für PCR-Reaktionen	43
3.14	PCR-Programm	44
3.15	Puffer und Lösungen für die Herstellung von kompetenten Zellen.	46
3.17	Lösungen für das IMPACT-CN Kit zur Proteinaufreinigung	50
3.18	Lösungen für die SDS-PAGE	53
3.19	Zusammensetzung des Trenn- und Sammelgels der SDS-PAGE	55
3.20	Lösungen für die Proteinidentifikation mit MALDI-ToF-ToF-MS	56
3.21	Oligonukleotide für die RNA-Sequenzierung.	64
4.1	Sequenzierte und gemappte Reads der cDNA-Bibliotheken des <i>C. diphtheriae</i> Wildtyps und der $\Delta dtxR$ -Mutante.	74
4.2	Vorhergesagte <i>cis</i> -regulatorische Elemente in 5'-UTRs von <i>C. diphtheriae</i>	79
4.3	Längste primäre Operons in <i>C. diphtheriae</i>	83
4.4	Gene mit DtxR-Bindestelle und Status der differentiellen Transkription und Promotormotiv in der <i>C. diphtheriae</i> $\Delta dtxR$ -Mutante.	90
4.5	Globales Alignment der Rho-Homologe	92

Tabellenverzeichnis

4.6	Sequenzierte und gemappte Reads der cDNA-Bibliotheken von der Δrho -Mutante und dem Wildtyp.	94
4.7	Differenziell transkribierte Gene in der Δrho -Mutante (Ausschnitt)	96
4.8	Transkripte für die sowohl ein Rho-abhängiger als auch Rho-unabhängiger Terminator identifiziert wurde.	104
4.9	Rho-abhängig regulierte Riboswitches und RNA-Motive.	113
4.10	Globales Alignment der RNase E bzw. G und RNase J-Homologe	124
4.11	Sequenzierte und gemappte Reads der cDNA-Bibliotheken von der Δrne -Mutante und der Δrnj -Mutante und den dazugehörigen Wildtypen.	127
4.12	Differenziell transkribierte Gene in der Δrne -Mutante (Ausschnitt)	128
4.13	Differenziell transkribierte Gene in der Δrnj -Mutante (Ausschnitt)	135
4.14	Sequenzierte und gemappte Reads der cDNA-Bibliotheken für die Identifizierung von RNase E/G Schnittstellen mittels <i>RNase-Assay</i>	139
4.15	Sequenzierte und gemappte Reads der cDNA-Bibliotheken für die Identifizierung von RNase J Schnittstellen mittels <i>RNase-Assay</i>	144
4.16	Riboswitches und RNA-Motive mit RNase E/G und RNase J Schnittstellen. . .	148
A.1	Aufschlüsselung der COG-Klassenbezeichnungen und deren Beschreibung nach Tatusov <i>et al.</i> (1997).	236
A.2	Liste der in dieser Arbeit entwickelten Programme.	238

Abbildungsverzeichnis

2.1	Rasterelektronenmikroskopische Aufnahme von <i>C. glutamicum</i> Zellen.	14
2.2	Mechanismus der Transkriptionsinitiation an bakteriellen Promotoren.	17
2.3	Mechanismus der Rho-abhängigen Transkriptionstermination.	20
2.4	RNA-Degradation in grampositiven Bakterien am Beispiel von <i>B. subtilis</i>	26
2.5	Darstellung der Hauptkomponenten des Typ IV Pilusystems.	30
4.1	Klassifizierung der Transkriptionsstarts in <i>C. diphtheriae</i>	75
4.2	Promotormotiv des Sigmafaktors σ^A in <i>C. diphtheriae</i>	76
4.3	Histogramm der 5'-UTRs und Motiv der ribosomalen Bindestelle.	78
4.4	Darstellung eines intergenischen TSS mit neu identifizierten Transkript in <i>C. diphtheriae</i>	81
4.5	Anzahl an Monocistrons, primär und sekundär Operons in <i>C. diphtheriae</i>	82
4.6	Transkriptions-Profil der Phagen-Insel (<i>phage island</i>) und dem <i>tox</i> Gen in <i>C. diphtheriae</i>	86
4.7	M/A-Plot der differenziell transkribierten Gene in der <i>C. diphtheriae</i> $\Delta dtxR$ -Mutante.	88
4.8	Vergleich der Rho-Homologe verschiedener Modellorganismen	91
4.9	Wachstumskurven der <i>C. glutamicum</i> Δrho -Mutante.	93
4.10	M/A-Plot der Δrho -Mutante im Vergleich zum Wildtyp.	95
4.11	Klassifizierung der in der Δrho -Mutante differenziell transkribierten Gene nach COG-Klasse	96
4.12	Differenziell transkribierte Gene der Prophagen-Regionen in der Δrho -Mutante.	99
4.13	LFC-Werte von Genen in IS-Elementen in der Δrho -Mutante	100
4.14	Beispiel eines verlängerten Transkripts in der Δrho -Mutante.	102
4.15	Vergleich der genomweiten und <i>antisense</i> Read-Abdeckung vom <i>C. glutamicum</i> Wildtyp mit der Δrho -Mutante.	103
4.16	In der Δrho -Mutante als verlängert identifizierte Transkripte unterteilt nach Transkriptklasse	106
4.17	Längenverteilung der verlängerten Transkripte in der Δrho -Mutante.	107

Abbildungsverzeichnis

4.18	CG-Blasen einiger Rho-abhängig terminierter Transkripte.	109
4.19	Längen- und Höhenvergleich der CG-Blasen von Rho- und Haarnadel-terminierten Transkripten.	110
4.20	Einfluss der Transkriptverlängerung auf die differenzielle Transkription.	111
4.21	Read-Abdeckung im Bereich des FMN-Riboswitches von <i>ribM</i> im Wildtyp und der Δrho -Mutante.	114
4.22	Grafische Darstellung des <i>tad</i> -Clusters und der <i>com</i> -Gene in <i>C. glutamicum</i> mit LFC-Werten in der Δrho -Mutante.	116
4.23	Verlängerung des 6C RNA-Transkripts in der Δrho -Mutante.	118
4.24	Charakterisierung des <i>tad</i> -Clusters in <i>C. glutamicum</i>	121
4.25	Vergleich der RNase E bzw. G und J-Homologe der Modellorganismen	123
4.26	Vergleich des Wachstums der Δrne -Mutante und Δrnj -Mutante mit dem Wildtyp.	125
4.27	M/A-Plot der <i>C. glutamicum</i> Δrne -Mutante.	128
4.28	Klassifizierung der in der Δrne -Mutante differenziell transkribierten Gene nach COG-Klasse.	131
4.29	Unprozessierte Vorläufer-Transkripte der 5S rRNA in der Δrne -Mutante identi- fiziert in den (A) RNA-Sequenzierungsdaten und mittels (B) Northern Blot.	132
4.30	M/A-Plot der <i>C. glutamicum</i> Δrnj -Mutante.	133
4.31	Klassifizierung der in der Δrnj -Mutante differenziell transkribierten Gene nach COG-Klasse.	134
4.32	Schematischer Ablauf des <i>RNase-Assays</i> zur Identifizierung von RNase-Schnitt- stellen.	138
4.33	Anzahl der RNase E/G Schnittstellen nach Transkript bzw. -abschnitt.	140
4.34	Analyse der RNase E/G Schnittstellen.	141
4.35	Anzahl der RNase J Schnittstellen nach Transkript bzw. -abschnitt.	145
4.36	Analyse der RNase J Schnittstellen.	146
4.37	Identifizierte Schnittstellen der RNase E/G und J in Riboswitches, RNA-Motiven und <i>leader</i> -Sequenzen.	149
A.1	Plasmidkarten zur Komplementation von <i>rho</i> und zur Aufreinigung von RNase E/G und J.	227
A.2	Plasmidkarten zur Deletion und Überexpression.	228
A.3	Distanzen der sequenzierten Proben für die verschiedenen Deletionsmutanten.	229
A.4	Komplementation des Wachstumsdefizits der Δrho -Mutante.	230
A.5	Anzahl diskordanter Reads von IS-Elementen im Wildtyp und der Δrho -Mutante.	231

A.6	Beispiel der Bereiche der Primerpaare für die Validierung von verlängerten Transkripten in der Δrho -Mutante mittels qRT-PCR	232
A.7	Validierung verlängerter Transkripte in der Δrho -Mutante mittels qRT-PCR . .	232
A.8	Schema der Reads in der 6C RNA-Region von Wildtyp und Δrho -Mutante. . .	233
A.9	Rasterkraftmikroskopie-Aufnahme (<i>Atomic Force Microscopy</i> , AFM) der Δrho -Mutante und des Wildtyps zur Identifizierung von Typ IVb Pilus-Strukturen.) Aufnahme des Wildtyps und der Δrho -Mutante	233
A.10	Komplementation der Δrnj -Mutante.	234
A.11	SDS-PAGE der Elution der Aufreinigung von RNase J (Cg2160).	235
A.12	Aktivitätsprüfung der aufgereinigten RNase J.	235

Anhang

A.1 Ergänzende Abbildungen

Die folgenden Abbildungen sind aufgrund ihrer Größe oder ihres ergänzenden Charakters hier aufgeführt.

Allgemeine Abbildungen

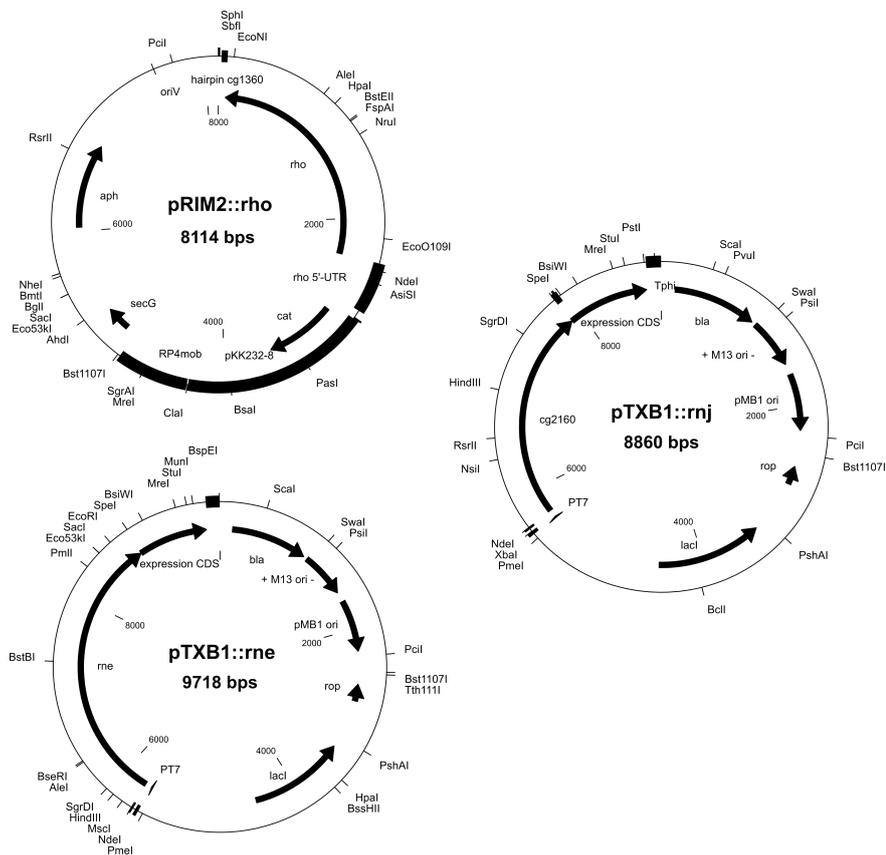


Abbildung A.1: Plasmidkarten zur Komplementation von *rho* und zur Aufreinigung von RNase E/G und J.

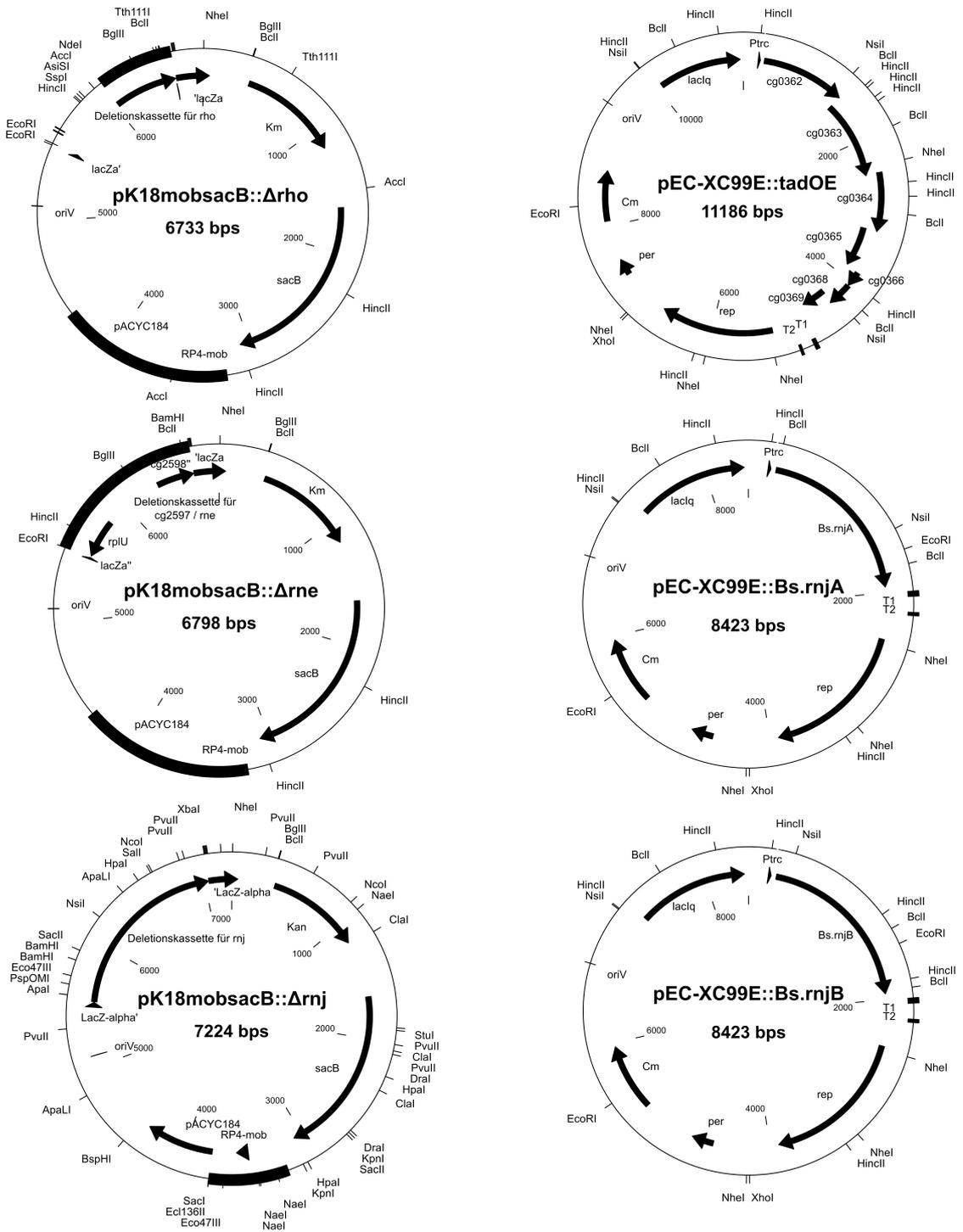


Abbildung A.2: Plasmidkarten zur Deletion der Gene *C. glutamicum* Gene *rho*, *rne* und *rnj* (links) und Überexpression des *tad*-Clusters sowie der Gene *rnjA* und *rnjB* aus *B. subtilis* (rechts).

A.1 Ergänzende Abbildungen

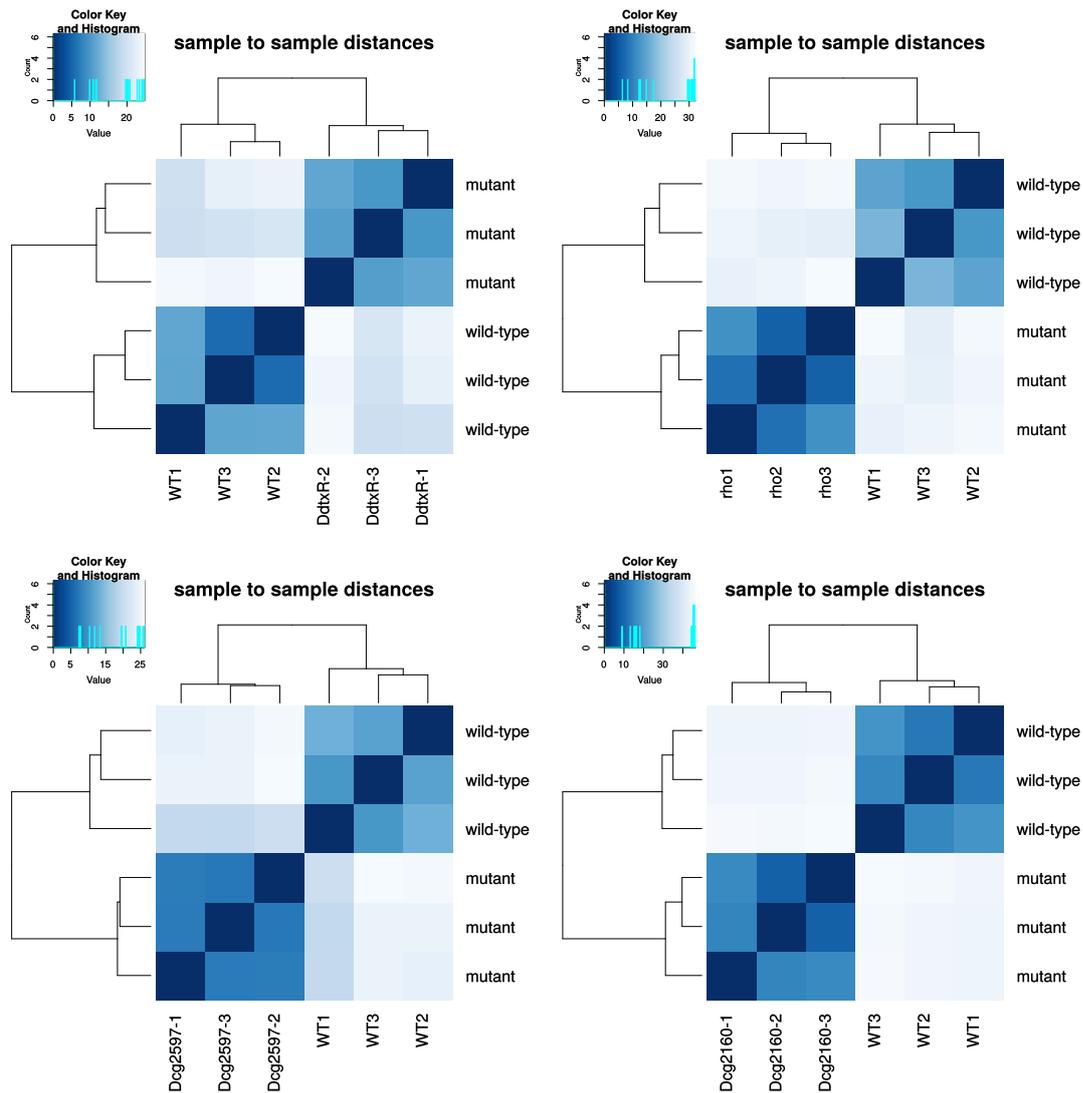


Abbildung A.3: Distanzen der sequenzierten Proben für die *C. diphtheriae* $\Delta dtxR$ -Mutante (links oben), *C. glutamicum* Δrho -Mutante (rechts oben), Δrne -Mutante ($\Delta cg2597$, links unten) und Δrnj -Mutante ($\Delta cg2160$, rechts unten). Die Probedistanzen wurden aus den mit *DESeq2* (Love *et al.*, 2014) normalisierten Reads über eine *rlog*-Transformation berechnet. Alle Replikate haben eine hohe Ähnlichkeit zueinander, so dass alle Replikate für die Analysen verwendet werden konnten.

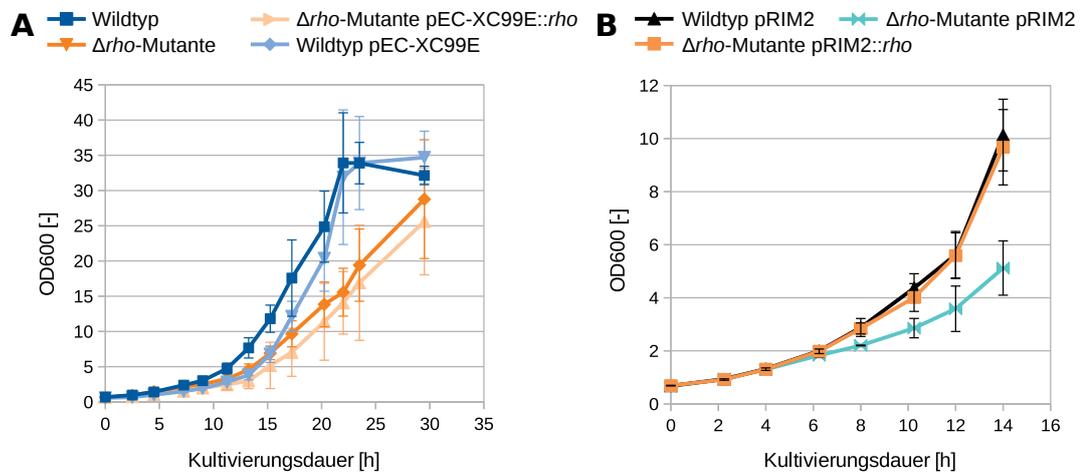
Rho-abhängige Transkriptionstermination in *C. glutamicum*

Abbildung A.4: Komplementation des Wachstumsdefizits der $\Delta\rho$ -Mutante mittels (A) plasmidbasiertem System (basierend auf dem Plasmid pEC-XC99E) und (B) chromosomaler Integration basierend auf pIM2. Gezeigt ist der Mittelwert von jeweils drei Replikaten. Die Fehlerbalken sind die Standardabweichung.

A.1 Ergänzende Abbildungen

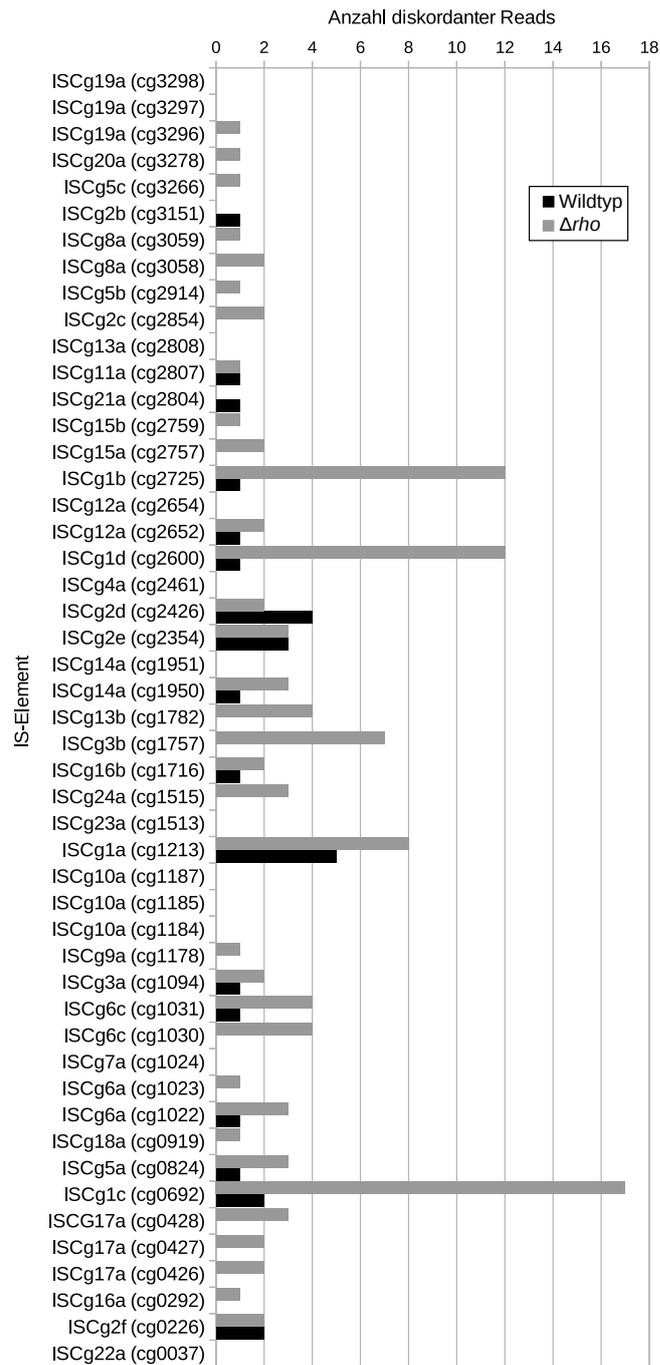


Abbildung A.5: Anzahl diskordanter Reads von IS-Elementen im Wildtyp (schwarz) und der $\Delta\rho$ -Mutante (grau). Beträgt die Distanz zweier *read pairs* weit mehr als die mittlere Fragmentlänge (hier >1000 bp), handelt es sich um ein diskordantes Read-Paar. Dieses Read-Paar umspannt einen Sequenzbereich, der nicht in der Genom-Referenz vorkommt und daher auf eine Genom-Neuanordnung hinweist. Somit sind diskordante Reads, von denen der eine Read im IS-Element und der andere Read >1000 bp entfernt liegt, ein Hinweis auf gesprungene IS-Elemente.

Anhang

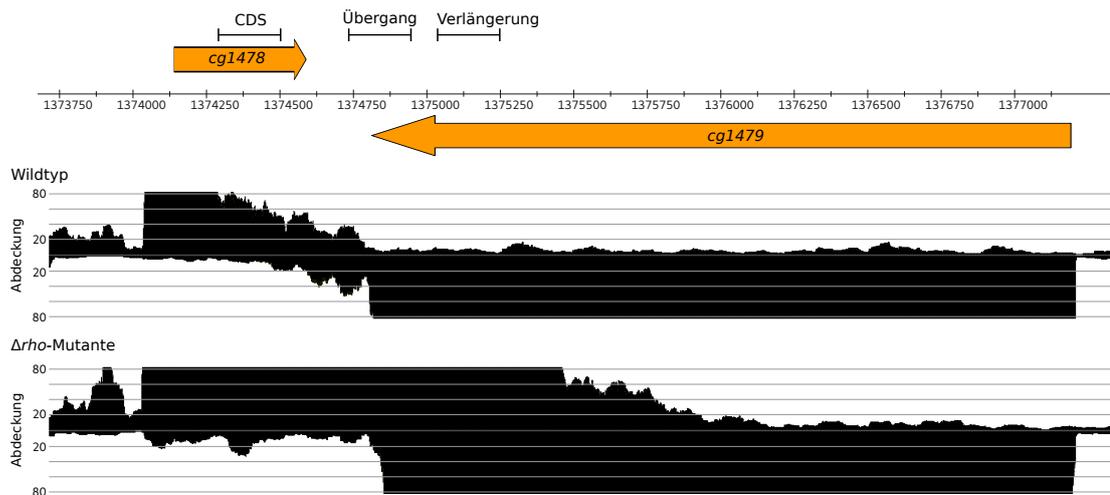


Abbildung A.6: Beispiel der Bereiche der Primerpaare für die Validierung von verlängerten Transkripten in der $\Delta\rho$ -Mutante mittels qRT-PCR gezeigt am Gen *cg1478*. Für jedes zu prüfende Transkript wurden drei Primerpaare erstellt: „CDS“, Primerpaar in der CDS bzw. dem annotierten Bereich; „Übergang“, Primerpaar welches den Bereich abnehmender Read-Abdeckung im Wildtyp überspannt und „Verlängerung“, Primerpaar welches den Bereich überspannt, der in der $\Delta\rho$ -Mutante noch deutliche Read-Abdeckung zeigt, im Wildtyp aber schon auf den Hintergrundwert abgesunken ist.

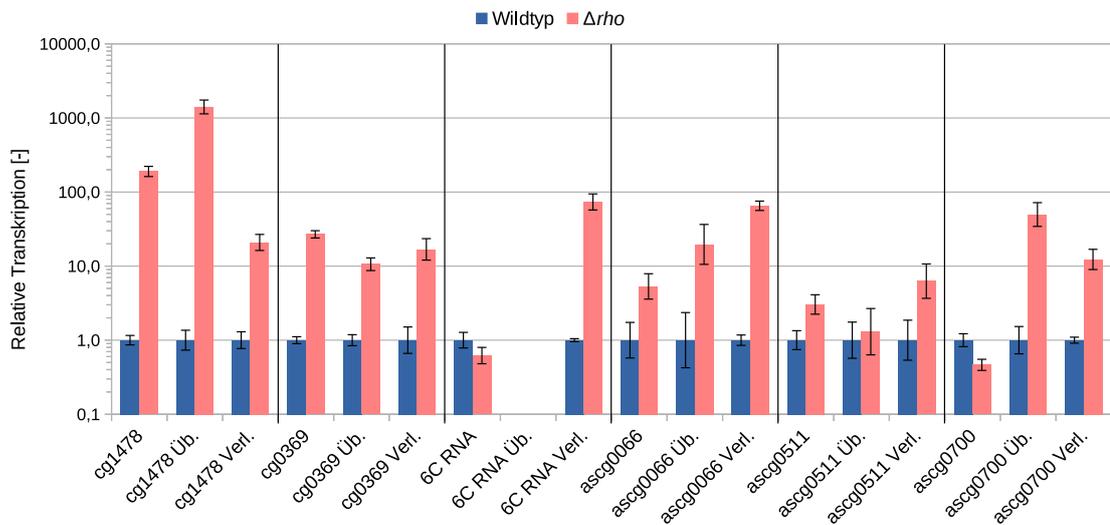


Abbildung A.7: Validierung einiger verlängerter Transkripte in der $\Delta\rho$ -Mutante mittels qRT-PCR. Es wurden insgesamt sechs verlängerte Transkripte der Transkriptklassen CDS (*cg1478* und *cg0369*), ncRNA (6C RNA) und antisense RNA (*ascg0700*, *ascg0066* und *ascg0511*) überprüft. Die relative Transkriptmenge ist für jedes Primerpaar auf die Transkriptmenge im Wildtyp normalisiert. Nur für den Bereich „Übergang“ der 6C RNA konnte kein für die qRT-PCR geeignetes Primerpaar entworfen werden. Es wurden jeweils drei biologische und zwei technische Replikate pro untersuchtem Transkript eingesetzt. Üb., Übergang; Verl., Verlängerung

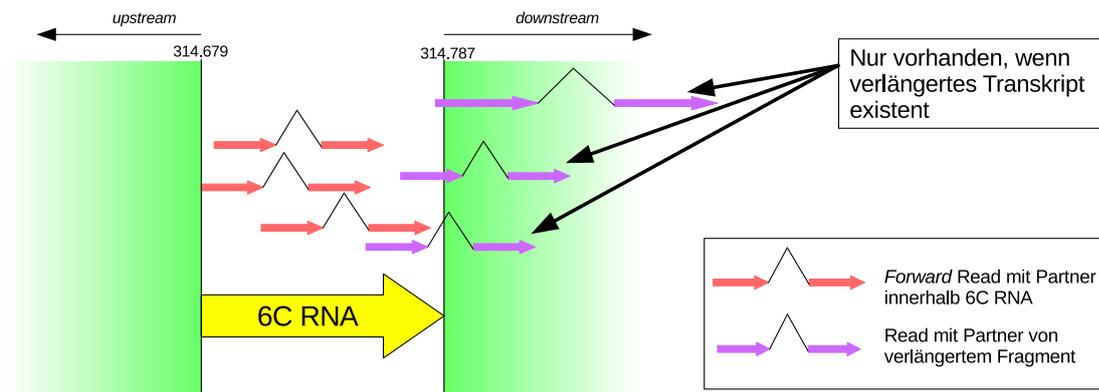


Abbildung A.8: Schema der Reads in der 6C RNA-Region von Wildtyp und $\Delta\rho$ -Mutante. Um zu zeigen, dass in der $\Delta\rho$ -Mutante im Vergleich zum Wildtyp verlängerte Transkripte der 6C RNA existieren, wurden die *paired-end* Reads, die innerhalb der 6C RNA-Region liegen, aus den RNA-Sequenzierungsdaten vom Wildtyp und der $\Delta\rho$ -Mutante extrahiert. Im Wildtyp existieren nur *paired-end* Reads, die vollständig in der 6C RNA-Region liegen (rote Pfeile). In der $\Delta\rho$ -Mutante existieren neben diesen Reads auch *paired-end* Reads, die in der 6C RNA-Region beginnen, dessen Read-Partner aber *downstream* der 6C RNA-Region liegt (violette Pfeile). Diese *paired-end* Reads beweisen die (Rho-abhängige) Verlängerung der 6C RNA in *C. glutamicum*. Zur Veranschaulichung zeigen die Read-Partner von Illumina *paired-end* Reads hier in die selbe Richtung, sind aber eigentlich gegenläufig orientiert.

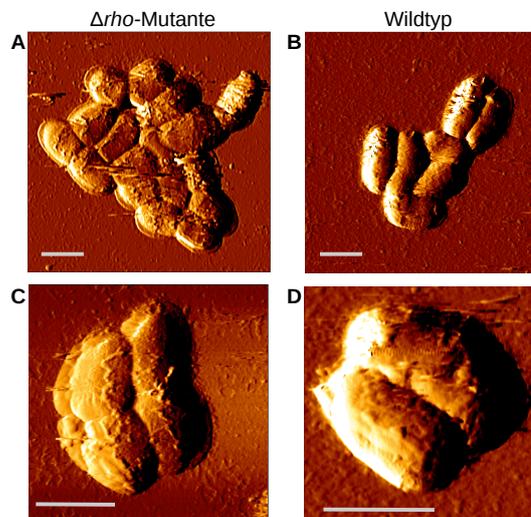


Abbildung A.9: Rasterkraftmikroskopie-Aufnahme (*Atomic Force Microscopy*, AFM) der $\Delta\rho$ -Mutante und des Wildtyps zur Identifizierung von Typ IVb Pilus-Strukturen. Dargestellt sind einige Zellen der $\Delta\rho$ -Mutante in einer (A) Übersichts- und in einer (C) Detailaufnahme. Zum Vergleich sind Zellen des Wildtyps in einer (B) Übersichts- und einer (D) Detailaufnahme gezeigt. Der Maßstab beträgt 1 μm (grauer Balken). Die Aufnahmen wurden von Dr. Nicole Hansmeier (Abteilung Mikrobiologie, Universität Osnabrück) durchgeführt.

RNA-Degradation durch die RNasen E/G und J in *Corynebacterium glutamicum*

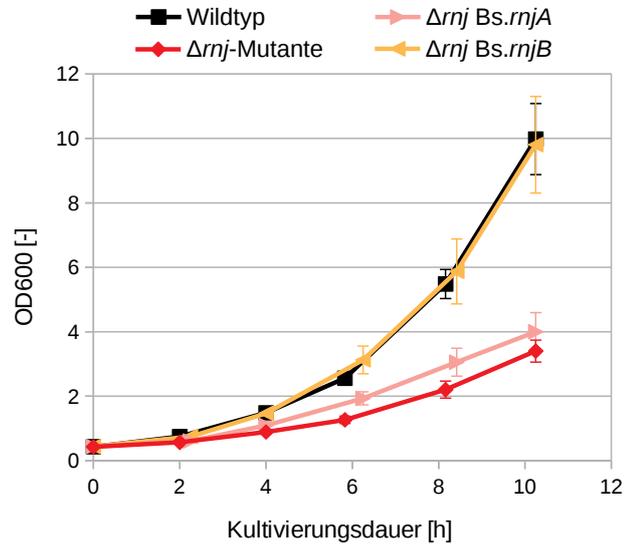


Abbildung A.10: Komplementation der *C. glutamicum* Δrnj -Mutante mit den Genen von RNase J1 (*Bs.rnjA*) und RNase J2 (*Bs.rnjB*) aus *Bacillus subtilis*. Die Kultivierungen wurden in CGXII-Minimalmedium durchgeführt. Dargestellt sind die Mittelwerte von drei biologischen Replikaten. Die Fehlerbalken sind die Standardabweichung.

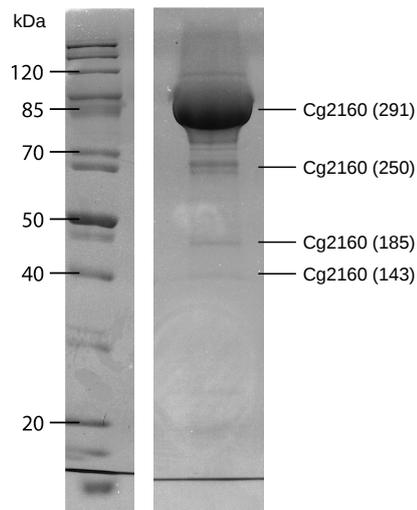


Abbildung A.11: SDS-PAGE der Elution der Aufreinigung von RNase J (Cg2160). In der linken Spalte ist ein Proteinstandard mit unterschiedlichen Größen als Marker aufgetragen. Die Beschriftung neben den einzelnen Banden zeigt das mittels MALDI-MS am wahrscheinlichsten identifizierte Protein an. In Klammern ist der Mascott-Wert angegeben. RNase J besitzt ein Molekulargewicht von 77,6 kDa. Die Bande mit der höchsten Intensität befindet sich ebenfalls auf der Höhe von ca. 75 kDa. Kleinere, vermutlich inaktive, Fragmente von RNase J wurden ebenfalls eluiert.

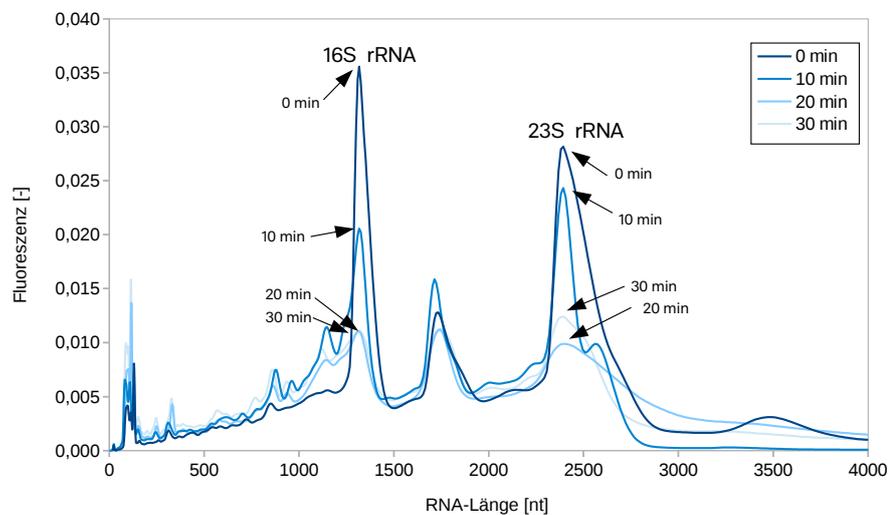


Abbildung A.12: Aktivitätsprüfung der RNase J mit dem Bioanalyzer 2100 (Agilent). Die RNA der Δrnj -Mutante wurde über eine Zeitreihe mit 50 pmol RNase J bei 30 °C inkubiert und die Abnahme der charakteristischen *Peak*-Höhen verfolgt. Als Negativkontrolle wurde die RNA für 30 min bei 30 °C ohne Enzym inkubiert.

A.2 Ergänzende Tabellen

Die folgenden Tabellen sind aufgrund ihres ergänzenden Charakters hier aufgeführt.

Tabelle A.1: Aufschlüsselung der COG-Klassenbezeichnungen und deren Beschreibung nach Tatusov *et al.* (1997).

COG-Klasse	Beschreibung auf englisch
<i>Information Storage and Processing</i>	
A	RNA processing and modification
B	Chromatin Structure and dynamics
J	Translation, ribosomal structure and biogenesis
K	Transcription
L	Replication, recombination and repair
<i>Cellular Processes and Signaling</i>	
D	Cell cycle control, cell division, chromosome partitioning
M	Cell wall, membrane, envelop biogenesis
N	Cell motility
O	Post-translational modification, protein turnover, and chaperones
T	Signal transduction mechanisms
U	Intracellular trafficking, secretion, and vesicular transport
V	Defense mechanisms
W	Extracellular structures
Y	Nuclear structure
Z	Cytoskeleton
<i>Metabolism</i>	
C	Energy production and conversion
E	Amino Acid metabolism and transport
F	Nucleotide metabolism and transport
G	Carbohydrate metabolism and transport
H	Coenzyme metabolism
I	Lipid transport and metabolism
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport, and catabolism
<i>Poorly Characterized</i>	
R	General Functional Prediction only
S	Function Unknown

A.3 Große ergänzende Tabellen

Die folgenden Tabellen können aufgrund ihrer Größe nicht sinnvoll in dieser Arbeit dargestellt werden und sind daher auf der beigefügten Daten-CD als Dokumente im OpenDocument-Format '.ods' hinterlegt. Zum Öffnen dieser Dokumente wird *Microsoft Excel*, *LibreOffice Calc* oder eine vergleichbare Software benötigt.

Allgemeine Tabellen

Tabelle S1 Liste der verwendeten Primer und deren Sequenzen.ods

Analyse der Transkriptionstartpunkte von *C. diphtheriae*

Tabelle S2 Liste der zu CDS zugehörigen TSS.ods

Tabelle S3 Liste der intragenischen und antisense TSS.ods

Tabelle S4 Liste der TLS-korrigierten CDS.ods

Tabelle S5 Liste der intergenischen TSS.ods

Tabelle S6 Liste der Operons und Monocistrons.ods

Tabelle S7 Liste der in der $\Delta dtxR$ -Mutante differentiell transkribierten Gene.ods

Rho-abhängige Transkriptionstermination in *C. glutamicum*

Tabelle S8 Liste der in der Δrho -Mutante differentiell transkribierten Gene.ods

Tabelle S9 Liste der verlängerten proteinkodierenden Transkripte.ods

Tabelle S10 Liste der verlängerten antisense intragenischen und anderen nichtkodierenden Transkripten.ods

RNA-Degradation durch die RNasen E/G und J in *C. glutamicum*

Tabelle S11 Liste der in der Δrne -Mutante differentiell transkribierten Gene.ods

Tabelle S12 Liste der identifizierten RNase E/G Schnittstellen.ods

Tabelle S13 Liste der in der Δrnj -Mutante differentiell transkribierten Gene.ods

Tabelle S14 Liste der identifizierten RNase J Schnittstellen.ods

A.4 Detaillierte Protokolle der *RNase-Assays*

In dieser Arbeit wurden der *in vivo* und der *in vitro RNase-Assay* zur Identifizierung von RNase-Schnittstellen mittels RNA-Sequenzierung entwickelt. Die detaillierten Protokolle befinden sich auf der beiliegenden Daten-CD.

A.5 Quellcode der im Rahmen dieser Arbeit entwickelten Programme

Im Rahmen dieser Arbeit wurden diverse Programme entwickelt, um spezifische Aufgaben und Fragestellungen zu behandeln. Die Programme wurden zum Teil vom Autor dieser

Anhang

Arbeit selbst, zum Teil von David Brandt (CeBiTec) programmiert. In jedem Fall ging der Entwurf sowie die Konzipierung auf die spezifische Aufgabe bzw. Fragestellung für jedes dieser Programme vom Autor dieser Arbeit aus. Die Programme befinden sich auf der beiliegenden Daten-CD.

Tabelle A.2: Liste der in dieser Arbeit entwickelten Programme. Die Programme wurden vom Autor dieser Arbeit entworfen bzw. konzipiert und selbst (Autor MW) oder von David Brandt (Autor DB) programmiert.

Programm	Autor	Funktion
create_base_SAF_for_featureCounts_v2.sh	MW	Benennung jeder Genomposition mit einer ID, um die Reads an jeder Position zählen zu können
FREAKmerge.sh	MW	Ausgaben von Jemboss Freak für C- und G-Anteile kombinieren
plot_all_freaks.R	MW	Generierung der CG-Blasen-Diagramme für alle Transkripte basierend auf <i>EMBOSS FREAK</i> Ausgaben
annotate_positions_v20161203.py	DB	Zuordnung von RNase-Schnittstellen zu einem Feature (Gen, UTR, etc)
find_longer_transcripts.py	DB	Identifikation von verlängerten Transkripten
generate_exons_v2.py	DB	Aufteilung aller Transkripte in jeweils 10 künstliche Exons zur Kategorisierung von RNase-Schnittstellen

Danksagung

Mein Dank gilt Herrn apl. Prof. Dr. Jörn Kalinowski für die Vergabe des sehr interessanten und spannenden Promotionsthemas. Gleichzeitig bedanke ich mich für die hervorragende wissenschaftliche Betreuung und Unterstützung während meiner Zeit in der AG *Mikrobielle Genomik und Biotechnologie*. Die vielen Diskussionen über fachliche aber auch weniger fachliche Themen waren stets sehr motivierend und haben des Öfteren zu kreativen Problemlösungen beigetragen.

Bei Herrn Dr. Stefan Albaum von der *Bioinformatics Ressource Facility* bedanke ich mich ebenfalls für die Betreuung und Unterstützung während meiner Promotion.

Weiter gilt mein Dank natürlich auch den Postdocs, Doktoranden, Technischen Assistenten und Studenten der Coryne-Gruppe. Mit euch ging die Laborarbeit immer gut von der Hand. Das lag natürlich auch an den vielen unterhaltenden Gesprächsthemen und Freizeitaktivitäten. Ein weiterer Dank geht an meine Labor- und Bürokollegen Arwa Al-Dilaimi, David Brandt, Tobias Busche, Julian Droste, Nils Lübke, Armin Neshat, Vera Ortseifen, Marcus Persicke, Lena Schaffert, Frederik Walter und Timo Wolf.

Für die immer freundliche und hilfreiche Unterstützung möchte ich mich auch bei Anika, Katharina und Yvonne vom Sequenzierteam bedanken.

Zuletzt gilt mein Dank meiner Familie, die mir während meines Studiums und meiner Promotion immer unter die Arme gegriffen hat.

Ein ganz besonderer Dank richtet sich an meine Frau Karina. Ihre liebevolle Unterstützung und Ermutigungen haben mich immer weiter angetrieben und motiviert.

Danke!

Erklärung

Hiermit erkläre ich, dass ich die hier vorliegende Dissertation mit dem Titel „Transkriptomsequenzierung bei Corynebakterien zur Charakterisierung von Mutanten in der RNA-Synthese und im RNA-Abbau“ eigenständig und ohne unerlaubte Hilfe angefertigt habe. Ich versichere, dass ich keine anderen als die angegebenen Quellen sowie Hilfsmittel verwendet habe. Die Stellen der Arbeit, die anderen Quellen im Wortlaut oder dem Sinn nach entnommen wurden, sind durch Angaben der Herkunft kenntlich gemacht. Dies gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen sowie für Quellen aus dem Internet.

Weiterhin erkläre ich, dass die vorliegende Dissertation weder vollständig noch in Auszügen einer anderen Fakultät mit dem Ziel vorgelegt worden ist, einen akademischen Titel zu erwerben. Ich bewerbe mich hier erstmalig um den Doktorgrad der Naturwissenschaften.

Bielefeld, den 25. Mai 2018

Manuel Wittchen