

Identifizierung und Funktionsanalysen
von kleinen RNAs in
Corynebacterium glutamicum

DISSERTATION

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

vorgelegt von
Almut Mentz

aus Berlin

Dezember 2013

INHALTSVERZEICHNIS

INHALTSVERZEICHNIS	I
ABBILDUNGSVERZEICHNIS	VI
TABELLENVERZEICHNIS	VIII
ABKÜRZUNGSVERZEICHNIS.....	IX
I. Zusammenfassung.....	1
II. Einleitung	3
1 Corynebacterium glutamicum	3
1.1 Die Taxonomie und Eigenschaften der Corynebakterien.....	3
1.2 Biotechnologische Bedeutung von <i>Corynebacterium glutamicum</i>	4
1.3 Genomvergleich mit den verwandten Spezies <i>C. efficiens</i> und <i>C. diphtheriae</i>	5
2 RNA und die bakterielle Transkription.....	7
2.1 Transkriptionsinitiation.....	8
2.2 Transkriptionstermination.....	11
2.3 RNA-Metabolismus und RNA-Stabilität	12
2.4 RNA-Strukturen und Strukturvorhersagen	14
3 Regulatorische RNAs in Bakterien.....	16
3.1 <i>Cis</i> -regulatorische RNA-Motive von proteinkodierenden Genen.....	16
3.2 Nicht-kodierende kleine regulatorische RNAs in Bakterien	18
3.3 Die Rfam-Datenbank und Rfam-Vorhersagen für <i>C. glutamicum</i>	22
4 Methoden zur Identifizierung kleiner RNAs in Bakterien.....	23
4.1 Bioinformatische Suche nach kleinen RNAs	23
4.2 Microarray-basierte Identifizierung kleiner RNAs	25
4.3 Transkriptomanalyse über Hochdurchsatz-Sequenzierung (RNA-Seq).....	27
5 Identifizierung von Funktionen regulatorischer <i>trans</i>-kodierter RNAs	28
III. Material und Methoden	31
1 Bakterienstämme, Plasmide, Primer.....	31
2 Verwendete Chemikalien, Materialien, Geräte und Software.....	32
2.1 Chemikalien, Enzyme, Kits, Marker.....	32

2.2	Verbrauchsmaterialien	35
2.3	Geräte und Apparaturen	35
2.4	Software	36
3	Medien.....	38
3.1	Nährmedien.....	38
3.2	Zusätze zu den Nährmedien.....	39
3.3	Verwendete Puffer und Lösungen.....	40
3.3.1	Nukleinsäure-Isolierung, -Reinigung und –Amplifikation	40
3.3.2	Elektrophorese	40
3.3.3	DNA-Transfertechniken	41
3.3.4	Auftrennung der RNA im Polyacrylamidgel und Semi-Dry Elektroblot.....	41
4	Kultivierung, Ernten und Lagerung von Bakterien	44
4.1	Bestimmung des Bakterientiters	44
4.2	Anzucht der Bakterien für die Experimente zur Identifizierung kleiner RNAs	44
4.3	Anzucht der Bakterien zur Attenuator-Validierung (<i>aroF</i>)	44
4.4	Stochertest.....	45
5	Polymerase-Kettenreaktion.....	45
5.1	Primer-Design	45
5.2	PCR-Reaktionsansätze und –Programme	46
6	Klonierungsexperimente	48
6.1	DNA-Restriktionsspaltung.....	48
6.2	DNA-Ligation.....	48
7	DNA-Transfertechniken.....	49
7.1	Herstellung und Elektroporation elektrokompenter <i>C. glutamicum</i> -Zellen	49
7.2	Hitzeschock-Transformation von ultrakompetenten <i>E. coli</i> -Zellen	49
8	Allgemeine DNA-Arbeiten	50
8.1	Isolierung von Plasmid-DNA.....	50
8.2	Aufreinigung von PCR-Produkten.....	50
8.3	Agarose-Gelelektrophorese.....	50
8.4	Isolierung von DNA-Fragmenten aus einem Agarosegel	51
9	GeneSOEing	51
10	Allgemeine RNA-Arbeiten	53
10.1	Isolierung der Gesamt-RNA	53
10.2	Präzipitation kleiner RNAs	54
10.3	DNase-Verdau	54
10.4	Aufreinigung durch P:C:I-Extraktion	54

10.5	Aufreinigung über LiCl-Fällung	55
10.6	RNA-Quantifizierung	55
10.7	RNA-Qualitätskontrolle	56
11	Spezifische Einzelgen-Transkriptanalysen	56
11.1	RT-qPCR	56
11.2	Northern Blot-Analyse für kleine RNAs (bis 800 nt)	57
11.2.1	Erstellung Digoxigenin-markierter RNA-Sonden	57
11.2.2	Semidry-Elektroblot	58
12	Microarrayanalysen	59
12.1	<i>Tiling</i> -Microarrays zur intergenischen Transkriptanalyse in <i>C. glutamicum</i>	59
12.1.1	Berechnung des intergenischen <i>C. glutamicum tiling</i> -Microarrays	59
12.1.2	RNA-Probenvorbereitung und Hybridisierung	59
12.1.3	Normalisierung und Auswertung der Microarraydaten	60
12.2	<i>C. glutamicum</i> Zweifarben-cDNA-Microarrays	61
13	cDNA-Sequenzierung kleiner RNAs (RNA-Seq)	62
13.1	Erstellen der cDNA-Sequenzierungs Bibliotheken	62
13.2	<i>Read-mapping</i> und Datenvisualisierung	63
13.3	Identifizierung von Transkriptionstarts	63
13.4	Promotorsuche	64
13.5	Vorhersage proteinkodierender Transkripte	64
14	Methoden zur bioinformatischen Vorhersage regulatorischer RNAs	65
14.1	Erstellung eines multiplen Genom-Alignments mit Mauve	65
14.2	RNAz zur Vorhersage regulatorischer RNAs	66
IV.	Ergebnisse	67
1	Bioinformatische Vorhersage (intergenischer) kleiner RNAs in <i>C. glutamicum</i>	67
1.1	Vorhersage regulatorischer RNAs in <i>C. glutamicum</i> mit dem Programm RNAz	68
1.2	Suche nach Rho-unabhängigen Terminatoren und deren Korrelation mit RNAz-Vorhersagen	71
2	Experiment-spezifische <i>tiling</i>-Microarrayanalysen in den intergenischen Regionen in <i>C. glutamicum</i>	72
2.1	Etablierung einer Methode zur intergenischen Transkriptanalyse in <i>C. glutamicum</i> basierend auf <i>tiling</i> -Microarrays	73
2.2	Design eines intergenischen <i>tiling</i> -Microarrays für <i>C. glutamicum</i> über IRMA	74
2.3	Bestimmung der Schwellenwerte für transkriptionelle Aktivität mithilfe der Signale von Kontroll-Oligonukleotiden	76
2.4	Analyse der Signalverteilung zur quantitativen Einteilung der Ergebnisse	77

2.5	Erstellung einer Datenbank für regulatorische RNAs in <i>C. glutamicum</i>	79
3	Identifizierung kleiner RNAs mittels Hochdurchsatz-Sequenzierung in <i>C. glutamicum</i> ..	81
3.1	Erzeugung von zwei verschiedenen cDNA-Bibliotheken zur Hochdurchsatz-Sequenzierung kleiner RNAs und Kartierung der erhaltenen <i>reads</i> auf der <i>C. glutamicum</i> Genomsequenz.....	81
3.2	Filterungsprozesse zur Identifizierung von potenziellen kleinen RNA-Genen.....	83
3.3	Klassifizierung von kleinen RNA-Genen in Abhängigkeit von ihrer Position zu den benachbarten CDS.....	85
3.4	Suche nach putativen kleinen Proteinen	87
3.5	Betrachtung von <i>cis-antisense</i> RNA-Genen und ihre Verteilung auf funktionelle CDS-Klassen	88
4	Validierung, Charakterisierung und funktionelle Analysen kleiner RNAs in <i>C. glutamicum</i>	91
4.1	Analyse von <i>trans</i> -kodierten RNA-Genen in Bezug auf Sekundärstrukturkonservierung und ihre differenzielle Expression	91
4.2	Spezifische Transkriptanalyse am Beispiel von vier <i>trans</i> -kodierten kleinen RNAs.....	96
4.3	Funktionelle Analyse der <i>trans</i> -kodierten RNA <i>cgb_16115</i>	100
4.3.1	Zugänglichkeitsanalyse von <i>cgb_16115</i> und verschiedene bioinformatische Methoden zur Vorhersage von Ziel-mRNAs dieser <i>trans</i> -kodierten RNA.....	100
4.3.2	Gezielte Deletionen verschiedener Interaktionsregionen von <i>cgb_16115</i> über <i>GeneSOEing</i>	103
4.3.3	Vergleichende cDNA-Microarrayanalysen zwischen dem <i>C. glutamicum</i> -Wildtypstamm und den Mutanten <i>Δcgb_16115_Reg1</i> und <i>Δcgb_16115_Reg123</i>	105
4.3.4	Hypothesen zum Regulationsmodell der Interaktion <i>cgb_16115</i> mit <i>ppmA</i> (<i>cg3138</i>).....	110
4.4	Validierung eines neuen Attenuationsmechanismus in der Synthese aromatischer Aminosäuren	113
V.	Diskussion	116
1	Die Kombination verschiedener bioinformatischer und experimenteller Methoden führt zur umfassenden Identifizierung von kleinen RNA-Genen in <i>C. glutamicum</i>	116
2	RNA-Seq-Daten erlauben aufgrund ihrer Einzelnukleotid-Auflösung die präzise Bestimmung von Eigenschaften kleiner RNAs in <i>C. glutamicum</i>	119
3	Die Verknüpfung der Ergebnisse verschiedener Transkriptanalysemethoden eignet sich zur Bestimmung spezifischer Expressionsmuster <i>trans</i> -kodierter RNAs.....	122
4	<i>Cis-antisense</i> RNAs sind in <i>C. glutamicum</i> zahlreich vorhanden und bei proteinkodierenden Genen bestimmter funktioneller Klassen häufiger anzutreffen	125
5	Funktionelle Analysen der <i>trans</i> -kodierten RNA <i>cgb_16115</i> lassen auf eine Interaktion mit <i>ppmA</i> (<i>cg3138</i>) schließen und deuten auf weitere Regulationen im Zusammenhang mit Metall-Ionen-Transport hin	127

6	Transkriptionelle Attenuation ist ein Mechanismus zur Regulation verschiedener Aminosäuresynthese-Gene in <i>C. glutamicum</i> und wurde hier für das Gen der DAHP-Synthetase (<i>aroF</i>) validiert.....	133
VI.	Ausblick.....	136
VII.	Literaturverzeichnis.....	138
VIII.	Anhang	156
1	Geniom®one Microarrayplattform.....	156
2	pK18 <i>mobsacB</i> -Leervektor.....	158
3	CD (Ordnerstruktur).....	159

ABBILDUNGSVERZEICHNIS

Abbildung II-1 Rasterelektronenmikroskopische (REM)-Aufnahme von <i>C. glutamicum</i>	4
Abbildung II-2 Syntenie-Plot von <i>C. glutamicum</i> , <i>C. efficiens</i> und <i>C. diphtheriae</i>	7
Abbildung II-3 SigA-Konsensussequenz in <i>C. glutamicum</i>	10
Abbildung II-4 Regulatoren der Transkription in Bakterien.....	11
Abbildung II-5 Globale mRNA-Halbwertszeitanalysen über Rifampicin in <i>B. subtilis</i> und <i>C. glutamicum</i>	13
Abbildung II-6 RNA-Sekundärstrukturelemente.....	15
Abbildung II-7 Schematische Abbildung der 5'-UTR des Trp-Operons in <i>E. coli</i>	17
Abbildung II-8 Beispiele von hochkonservierten bakteriellen RNA-Haushaltsgenen.....	19
Abbildung II-9 <i>Cis-antisense</i> RNAs und <i>trans</i> -kodierte RNAs.....	20
Abbildung III-1 Schematische Darstellung zur Lage der Primer für GeneSOEing.....	52
Abbildung III-2 Vorbereitungsschritte zur Erstellung der cDNA-Bibliotheken aus den Transkripten der kleinen RNA-Fractionen mit dem Illumina® TruSeq™ Small RNA Kit.....	63
Abbildung IV-1 Multiples Genom-Alignment über <i>C. glutamicum</i> , <i>C. efficiens</i> , <i>C. diphtheriae</i>	68
Abbildung IV-2 RNAz-Vorhersage für <i>C. glutamicum</i> und stringente Reduktion.....	70
Abbildung IV-3 Arbeitsvorgang zur Detektion kleiner RNAs über intergenische <i>tiling</i> -Microarrays.....	74
Abbildung IV-4 Aufbau des intergenischen <i>C. glutamicum tiling</i> -Microarray (Beispiel-Region).....	75
Abbildung IV-5 Verteilung der Signale aller Negativ-Kontrollen.....	76
Abbildung IV-6 Überblick über die normalisierten Signale im Experiment ohne Stress.....	78
Abbildung IV-7 Datenbankstruktur.....	79
Abbildung IV-8 Schritte zur Erzeugung der RNA-Proben für die Erstellung von zwei cDNA-Bibliotheken.....	82
Abbildung IV-9 Definition von Start- und Stopp-Positionen potenzieller kleiner RNA-Gene.....	84
Abbildung IV-10 Klassifizierung der kleinen RNAs in Abhängigkeit von ihrer Position zu den benachbarten CDS.....	87
Abbildung IV-11 Vergleich der Längenverteilung für asRNAs, <i>trans</i> -kodierte RNAs und mRNA Leader.....	89
Abbildung IV-12 Funktionelle Klassikation von <i>C. glutamicum</i> Genen mit kodierten <i>cis-antisense</i> RNAs (asRNAs).....	90
Abbildung IV-13 RNA-Haushaltsgene in <i>C. glutamicum</i> im Längenvergleich über alle Methoden.....	92
Abbildung IV-14 Expressionsmuster über 32 <i>trans</i> -kodierte RNAs mit sehr deutlichem Transkriptsignal.....	95
Abbildung IV-15 Sequenzabdeckung, Validierung durch Northern Blot und Sekundärstrukturvorhersage von drei <i>trans</i> -kodierten kleinen RNAs.....	97
Abbildung IV-16 <i>Trans</i> -kodierte RNA <i>cgb_16115</i>	99
Abbildung IV-17 Zugänglichkeitsprofil von <i>cgb_16115</i> und drei häufig vorhergesagte Interaktionsregionen.....	102
Abbildung IV-18 Venn-Diagramm über die Ergebnisse bioinformatischen Zielsuche von <i>cgb_16115</i>	103
Abbildung IV-19 Überblick <i>cgb_16115</i> im Wildtypstamm im Vergleich zu zwei <i>erzeugten</i> Deletionskonstrukten.....	104
Abbildung IV-20 M/A-Diagramm für die Microarrayanalyse von <i>C. glutamicum Δcgb_16115-Reg123</i> im Vergleich zum Wildtypstamm nach hyperosmotischem Stress.....	106

Abbildung IV-21 Validierung der erhöhten relativen Expression für <i>cg3138</i> in Δ <i>cgb_16115-Reg123 bezogen auf den <i>C. glutamicum</i>-Wildtypstamm (WT) als Referenz jeweils nach NaCl-Stress (RT-qPCR-Messung mit Gesamt-RNA).</i>	108
Abbildung IV-22 Sekundärstruktur der 5'-UTR von <i>cg3138</i> und Visualisierung der Interaktion mit <i>cgb_16115</i>	111
Abbildung IV-23 Transkriptionelle Attenuation für das DAHP-Synthase kodierende Gen <i>aroF</i> (<i>cg1129</i>).	114
Abbildung IV-24 Potenzielle Terminator-/Antiterminatorstruktur im Attenuator stromaufwärts von <i>aroF</i> in <i>C. glutamicum</i>	115
Abbildung V-1 Netzwerk-Graph einer Untermenge von <i>trans</i> -kodierten RNAs, die in den durchgeführten Experimenten transkriptionell aktiv waren.	123
Abbildung VIII-1 pK18 <i>mobsacB</i> -Leervektor	158

TABELLENVERZEICHNIS

Tabelle II-1 Vergleich der Genomsequenzen von <i>C. glutamicum</i> , <i>C. efficiens</i> und <i>C. diphtheriae</i>	6
Tabelle II-2 Vorhersage kleiner RNAs über SIPHT in <i>C. glutamicum</i> ATCC 13032 (aus Livny <i>et al.</i> , 2008).....	25
Tabelle III-1 Verwendete Bakterienstämme	31
Tabelle III-2 Verwendete Plasmide	31
Tabelle III-3 Verwendete Primer	31
Tabelle III-4 Liste der verwendeten Chemikalien.....	32
Tabelle III-5 Liste der verwendeten Enzyme	34
Tabelle III-6 Liste der verwendeten Kits	34
Tabelle III-7 Liste der verwendeten Marker	35
Tabelle III-8 Verwendete Verbrauchsmaterialien.....	35
Tabelle III-9 Verwendete Geräte und Apparaturen.....	35
Tabelle III-10 Verwendete Software.....	36
Tabelle III-11 Verwendete Antibiotika	39
Tabelle III-12 Mix für einen Taq PCR-Ansatz (50µL)	46
Tabelle III-13 Programm für die Taq-PCR.....	47
Tabelle III-14 Mix für einen Phusion PCR-Ansatz (50µL)	47
Tabelle III-15 Programm für eine Phusion-PCR.....	47
Tabelle IV-1 Experiment-spezifische Standard-Schwellenwerte ab der transkriptionelle Aktivität angenommen wird.	77
Tabelle IV-2 Oligonukleotidsignale über alle Experimente (nach Standard-Schwellenwert).....	80
Tabelle IV-3 <i>Read</i> -Übersicht über die Sequenzierungs-Ergebnisse der zwei cDNA-Bibliotheken.....	83
Tabelle IV-4 Liste der detektierten und Rfam-vorhergesagten <i>cis</i> -regulatorische Motive in <i>C. glutamicum</i>	85
Tabelle IV-5 Putative Peptide	88
Tabelle IV-6 <i>Trans</i> -kodierte kleine RNA-Gene mit überlappender RNAz-Vorhersage ($p \geq 0.9$) und Vergleich mit NAPP-Daten für <i>C. glutamicum</i> (Ott <i>et al.</i> , 2012).	93
Tabelle IV-7 <i>Trans</i> -kodierte RNAs mit SigH-/ECF-ähnlichem Promotor	94
Tabelle IV-8 Ergebnisse von ABC-Typ Transportergenen im k-Means Gencluster C7 aus den vergleichenden Microarrayexperimenten der beiden <i>cgb_16115</i> Deletetionsmutanten vs. <i>C. glutamicum</i> - Wildtypstamm (Kultivierung ohne Stress) ^a	109

ABKÜRZUNGSVERZEICHNIS

% (v/v)	Volumenprozent
% (w/v)	Gewichtsprozent
°C	Grad Celsius
g	Mikrogramm
µL	Mikroliter
µM	Mikromolar
A	Adenin
AA	Acrylamid
Abb.	Abbildung
abh.	abhängig
AGE	Agarosegelelektrophorese
Amp	Ampicillin
APS	Ammoniumperoxodisulfat
AS	Aminosäure(n)
AsaC	<i>cis-antisense</i> RNA von <i>argC</i>
asRNA	<i>cis-antisense</i> RNA
ATCC	American Type Culture Collection
ATP	Adenosintriphosphat
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
BHI	<i>brain heart infusion</i>
bp	Basenpaar(e)
BPB	Bromphenolblau
Bsp.	Beispiel
bzw.	beziehungsweise
C	Cytosin
<i>C. diphtheriae</i>	<i>Corynebacterium diphtheriae</i>
<i>C. efficiens</i>	<i>Corynebacterium efficiens</i>
<i>C. glutamicum</i>	<i>Corynebacterium glutamicum</i>
ca.	circa
CaCl ₂	Calciumchlorid
CASO	Caseinpepton-Sojamehlpepton
cDNA	<i>complementary</i> DNA
CO ₂	Kohlenstoffdioxid
CP	<i>crossing point</i>
Da	Dalton, Masseinheit [g mol ⁻¹]
DEPC	Diethylpyrocarbonat
dH ₂ O	Destilliertes Wasser
DIG	Digoxigenin
DMF	Dimethylformamid
DMSO	Dimethylsulfoxid
dNTPs	Didesoxyribonukleosid-Triphosphate
ds	doppelsträngig
dsDNA	Doppelsträngige DNA
DTT	1,4-Dithiotreitol
<i>E. coli</i>	<i>Escherichia coli</i>

EC-Nummer	Enzyme Commission-Nummer
EDTA	Ethylendiamintetraessigsäure / Ethylendiamintetraacetat
engl.	englisch
EPS	Exopolysaccharid
<i>et al.</i>	<i>et alii</i> (und andere)
EtBr	Ethidiumbromid
EtOH	Ethanol
G	Guanidin
GC-Gehalt	Guanin-Cytosin-Gehalt
GeneSOEing	<i>Gene Splicing by overlap extension</i> (Methode zur Gendeletion)
ggf.	gegebenenfalls
GRAS	<i>generally regarded as safe</i>
griech.	Griechisch
Std.	Stunde(n)
H ₂ O	Wasser
H-Brücke	Wasserstoffbrückenbindung
HCl	Salzsäure
HGT	horizontaler Gentransfer
HPLC	<i>high performance liquid chromatography</i>
HTH	<i>helix-turn-helix</i>
i.d.R.	in der Regel
IGR(s)	intergenische Region(en)
IPA	Isopropanol bzw. 2-Propanol
IRMA	<i>Intergenic Regions by Microarray Analysis</i> (Java-Applet)
K ₂ HPO ₄	Dikaliumhydrogenphosphat
kb	<i>kilo bases</i> (Kilobasen; 1000 Nukleotide)
KCl	Kaliumchlorid
kDa	Kilo Dalton; Molmasse
KH ₂ PO ₄	Kaliumdihydrogenphosphat
K _m	Kanamycin
KOH	Kaliumhydroxid
LB	<i>lysogeny broth</i> , auch Luria-Bertani (Bertani, 2004)
LiCl	Lithiumchlorid
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
mA	Milli-Ampere (Einheit der elektrischen Stromstärke)
max.	maximal
MCS	multiple cloning site
MFE	minimale freie Energie
Mg	Milligramm
MgCl ₂	Magnesiumchlorid
MGA	Multiples Genom Alignment
MgSO ₄	Magnesiumsulfat
Min.	Minute(n)
mL	Milliliter
mM / M	Millimolar / Molar
MOPS	3-Morpholinopropansulfonsäure
mRNA	<i>messenger RNA</i>

MS	Massenspektrometrie
N ₂	Stickstoff
Na	Natrium
Na ₂ HPO ₄	Dinatriumphosphat
NaCl	Natriumchlorid
NaOH	Natronlauge
NH ₄	Ammonium(-Ion)
NH ₄ Cl	Ammoniumchlorid
nm	Nanometer
nt	Nukleotide
Nx	Nalidixinsäure
OD ₆₀₀	Optische Dichte bei einer Wellenlänge von 600 nm
ORF	<i>open reading frame</i> ; offenes Leseraster
<i>p.a.</i>	<i>per analysis</i> (zur Analyse)
PAA	Polyacrylamid
PAGE	Polyacrylamid-Gelelektrophorese
PCI	Phenol/Chloroform/Isoamylalkohol
PCR	<i>polymerase chain reaction</i> , Polymerase-Kettenreaktion
pH	pH-Wert
Phe/ F	Phenylalanin
PIPES	Piperazin-N,N'-bis(2-ethansulfonsäure)
r	resistent
RBS	Ribosomen-Bindestelle
RNA	Ribonukleinsäure
RNA-Seq	RNA-Hochdurchsatz-Sequenzierung
rpm	<i>rounds per minute</i> , Umdrehungen pro Minute
rRNA	ribosomale RNA
RT	Raumtemperatur
RT-qPCR	<i>reverse transcription-quantitative PCR</i>
SDS	<i>sodium dodecylsulfate</i> , Natriumdodecylsulfat
SAGE	<i>Serial Analysis of Gene Expression</i>
SDS-PAGE	Natriumdodecylsulfat-Polyacrylamidgelelektrophorese
Sek	Sekunde(n)
SLS / SDS	Natriumlaurylsulfat / Natriumdodecylsulfat
sRNA	small RNA, kleine RNA
T	Thymin
TA	TRIS-Acetat-EDTA
Tab.	Tabelle
TAE	TRIS-Acetat-EDTA(-Puffer)
<i>Taq</i>	<i>Thermus aquaticus</i>
TEMED	Tetramethylethylendiamin
TLS	Translationsstart
Tm	Schmelztemperatur (der Primer)
TRIS	Tris(hydroxymethyl)-aminomethan
Triton X-100	Octoxinol-9 (nicht-ionisches Tensid)
TY	<i>tryptone yeast extract</i>
Tyr/ T	Tyrosin

U	<i>unit</i> , Einheit
u. a.	unter anderem
ü/N	über Nacht
UTR	<i>untranslated region</i> (untranslatierter Bereich einer Kodiersequenz)
UV	Ultraviolett
V	Volt (Einheit der elektrischen Spannung)
vs.	<i>versus</i> , gegen
WT	Wildtyp
z.B.	zum Beispiel

I. Zusammenfassung

Traditionell stellen RNA-Moleküle einer jeden Zelle den Zwischenschritt vom DNA-kodierten Gen zum Protein als Boten-RNA (*messenger* RNA, mRNA) dar oder spielen als ribosomale oder *transfer* RNAs (rRNAs oder tRNAs) eine Rolle in der Proteinbiosynthese. Diese Anschauung hat sich innerhalb der letzten 10-15 Jahre stark gewandelt. Mittlerweile ist bekannt, dass es zusätzlich kleine regulatorische RNA-Moleküle gibt (*small* RNA; sRNA), die nicht für Proteine kodieren. Kleine RNAs regulieren die Genexpression für verschiedenste zelluläre Prozesse und wurden besonders in Anpassung an neue Umgebungs- und Stress-Bedingungen mit hoher Expression beobachtet. Auf Basis verschiedener Analysen werden Hunderte dieser kleinen RNA-Gene in Bakteriengenomen angenommen. *Corynebacterium glutamicum* stellt ein Gram-positives Bakterium von großer industrieller Bedeutung dar und gilt zudem als Modellorganismus für die nah verwandten und humanpathogenen Bakterien *C. diphtheriae* und *Mycobacterium tuberculosis*. Vor Beginn dieser Arbeit waren in *C. glutamicum* lediglich einige wenige hochkonservierte RNAs über Kovarianzmodelle vorhergesagt und in der Rfam-Datenbank hinterlegt (Gardner *et al.*, 2009). Die Hauptaufgabe der Dissertationsarbeit bestand darin, kleine regulatorische RNAs in diesem Gram-positiven Bakterium zu identifizieren und diese zu klassifizieren.

Zur Bestimmung kleiner RNA-Gene in *C. glutamicum* wurde in dieser Arbeit eine Hochdurchsatz-Sequenzierung (RNA-Seq) zur Nukleotid-genauen Auflösung kleiner RNAs mit dem Illumina GA IIx-System durchgeführt (Mentz *et al.*, 2013). Um ein möglichst umfangreiches Bild über kleine RNAs in *C. glutamicum* zu erhalten, wurde für das RNA-Seq-Experiment ein Gemisch von sechs Proben isolierter Gesamt-RNA aus Schüttelkolben-Kulturen im exponentiellen Stadium verwendet, wobei fünf dieser Kulturen zum Probenahme-Zeitpunkt physischen Stressen (Hitze, Kälte) oder chemischen Stressbedingungen (Diamid, H₂O₂, NaCl) ausgesetzt waren. Zusätzlich wurden zwei weitere große Analysen speziell in den intergenischen Regionen des *C. glutamicum*-Genoms durchgeführt. Hier sind zum einen Experiment-spezifische Transkriptanalysen mittels *tiling*-Microarrays zu nennen, wobei vergleichbare Proben wie im Sequenzierungsexperiment verwendet wurden. Zum anderen wurde mit dem bereits implementierten Programm RNAz (Washietl *et al.*, 2005) eine bioinformatische Vorhersage regulatorischer RNAs in *C. glutamicum* gemacht, die aus einer Sekundärstrukturkonservierung in den nah verwandten Bakterien *C. efficiens* und *C. diphtheriae* resultiert.

In Vorbereitung auf das RNA-Seq-Experiment wurden zwei cDNA-Bibliotheken kleiner RNAs erzeugt. Eine dieser Bibliotheken wurde zur Kartierung der Transkriptionsstarts (TSS) mittels enzymatischer Behandlung auf Primärtranskripte angereichert, während die andere unbehandelt blieb. Um die TSS

kleiner RNA-Gene zu bestimmen, wurden auf die erhaltenen RNA-Seq-*reads* aus cDNA-Bibliothek 1 verschiedene Filterschritte angewendet, und es wurde mithilfe einer *in silico*-Analyse nach zugehörigen Promotoren gesucht. Im Anschluss daran wurden für diese 5'-Enden von RNA-Genen mithilfe der cDNA-Bibliothek 2 die 3'-Enden ermittelt. Eine Klassifizierung der erhaltenen kleinen RNA-Gene über die Entfernung zu den benachbarten proteinkodierenden Genen ergab insgesamt 316 mRNA-Leader, 262 *trans*-kodierte RNAs und die große Anzahl von 543 *cis-antisense* RNAs (Mentz *et al.*, 2013). Ein Drittel der 262 *trans*-kodierten RNAs zeigen Übereinstimmungen mit den 599 RNAz-Vorhersagen und scheinen daher in *C. efficiens* und *C. diphtheriae* konserviert zu sein. Zu diesen zählen auch alle drei Haushalts-RNAs und die aus Actinomyceten bekannte 6C RNA. Im Rahmen der Microarrayanalysen erwies sich eine Kooperation mit der febit biomed GmbH (jetzt CBC, Heidelberg) als vielversprechend, wobei auf Basis der flexiblen genom@one Plattform spezielle *tiling*-Microarrays etabliert wurden, die alle intergenischen Bereiche eines Genoms abdecken. In diesem Zusammenhang wurde auch das Java-Applet IRMA (*Intergenic Regions by Microarray Analysis*) (Schroeder, 2007) zur Berechnung der intergenischen Oligonukleotidproben in Zusammenarbeit entwickelt. Eine Analyse der normalisierten Microarrayrohdaten hat zu einer quantitativen Einteilung der Signale geführt. Die Hälfte der *trans*-kodierten RNAs wurde mithilfe der Microarrayanalysen detektiert und diesen kann nun eine Experiment-spezifische transkriptionelle Aktivität zugewiesen werden.

Im Rahmen dieser Arbeit werden vier *trans*-kodierte RNAs über eine Northern Blot-Analyse validiert. Für eine dieser kleinen RNAs (*cgb_16115*) werden auf bioinformatischer und experimenteller Ebene Funktionsanalysen präsentiert, die auf eine mögliche Ziel-mRNA *cg3138* (*ppmA*) hinweisen. In vergleichenden cDNA-Microarrayexperimenten von Deletionsmutanten dieser kleinen RNA mit dem *C. glutamicum*-Wildtypstamm war zudem die Expression für Gene des Metall-Ionen-Transports bzw. speziell für Gene der Eisen-Homöostase auffällig verändert. Im letzten Kapitel werden für die peptidkodierende RNA *cgb_11283* stromaufwärts des DAHP-Synthase-Gens (*aroF*, *cg1129*) aus der Klasse „mRNA-Leader“ Ergebnisse vorgestellt, die einen neuen Mechanismus transkriptioneller Attenuation in der Synthese aromatischer Aminosäuren validieren.

II. Einleitung

1 *Corynebacterium glutamicum*

In der vorliegenden Dissertationsschrift über kleine RNAs in *Corynebacterium glutamicum* (ehemals *Brevibacterium flavum*) soll im ersten Kapitel die taxonomische Einordnung dieses Bakteriums und die Beschreibung der Eigenschaften von Corynebakterien als solchen erfolgen. Anschließend wird die industrielle und biotechnologische Bedeutung von *C. glutamicum* dargestellt. Den Abschluss dieses Kapitels bildet ein Genomvergleich zwischen *C. glutamicum* und den beiden nah verwandten Spezies *C. efficiens* und *C. diphtheriae*.

1.1 Die Taxonomie und Eigenschaften der Corynebakterien

Innerhalb der Klasse der Actinobakterien gehört die Gattung *Corynebacterium* zur Familie der *Corynebacteriaceae* und stellt Gram-positive Bakterien mit typischerweise hohem G+C-Gehalt > 50 Mol % dar (Stackebrandt *et al.*, 1997). Die Gattung *Corynebacterium* wurde zuerst 1896 von Lehmann und Neumann beschrieben (Lehmann und Neumann, 1896), wobei sich der Name aus der Eigenschaft ableitet, keulenförmige bzw. leicht gewölbte Bakterienformen anzunehmen (griechisch *coryne* = Keule; vgl. Abbildung II-1, nächste Seite). Die Vertreter der Gattung bilden 3-5 µm lange gerade oder leicht gekrümmte Stäbchen, die unbeweglich sind, keine Sporen bilden und aerob bis fakultativ anaerob leben (Barksdale, 1970). Eine weitere Eigenschaft dieser Bakterien ist die sogenannte Schnappteilung (*snapping division*). Oft bleiben Zellen nach der Teilung miteinander verbunden und erscheinen unter dem Mikroskop dadurch v-förmig (Letek *et al.*, 2008). Die Corynebakterien wurden aufgrund ihres Zellwandaufbaus außerdem der CMN-Gruppe zugewiesen (Barksdale, 1970). Diese Gruppe besteht aus den Gattungen *Corynebacterium*, *Mycobacterium* und *Nocardia* und weist als charakteristisches Merkmal eine zweite Mykolsäure-haltige Permeabilitäts-Barriere ähnlich der äußeren Membran von Gram-negativen Bakterien auf. Während berichtet wurde, dass diese äußere Barriere essentiell ist für die Virulenz von *Mycobacterium tuberculosis* (Dubnau *et al.*, 2000), konnte für *C. glutamicum* gezeigt werden, dass der Mykolsäure-Gehalt dieser Barriere die Sensitivität gegenüber Antibiotika sowie den Export von L-Lysin und von L-Glutamat beeinflusst (Gebhardt *et al.*, 2007). In den meisten der *Corynebacterium*-Spezies besteht die darüber liegende Schicht aus freien Polysacchariden, Glykolipiden und Proteinen, während darunter ein Arabinogalactan-Geflecht über kovalente Bindungen an eine Peptidoglycanstruktur gebunden ist, die den Übergang zur Plasmamembran darstellt (Eggeling *et al.*, 2008). Bisher wurden 88 verschiedene *Corynebacterium*-Spezies taxonomisch klassifiziert, die teilweise von medizinischer oder wirtschaftlicher Relevanz sind (Bernard, 2012). Einige Vertreter der Corynebakterien sind für Menschen oder Tiere pathogen, wobei *C. diphtheriae* (Cerdeño-Tárraga *et al.*, 2003) den prominentesten Vertreter darstellt und als Auslöser der Krankheit Diphtherie bekannt ist. Die

nicht-pathogenen Spezies wie zum Beispiel *C. glutamicum* (Kinoshita *et al.*, 1958) und *C. efficiens* (Fudou *et al.*, 2002) kommen im Bodenhabitat vor, während andere Vertreter Saprophyten darstellen oder auf der Haut bzw. den Schleimhäuten des Menschen zu finden sind.

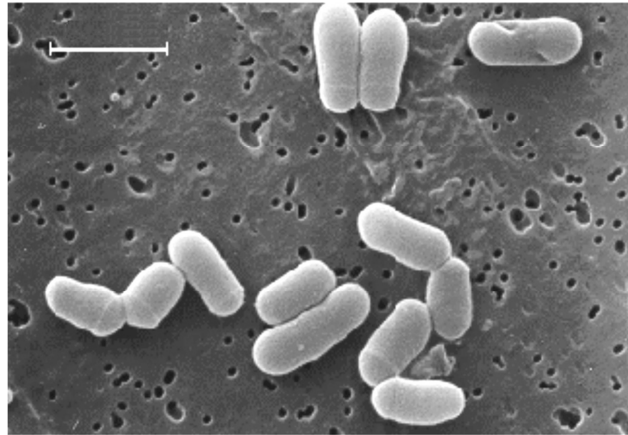


Abbildung II-1 Rasterelektronenmikroskopische (REM)-Aufnahme von *C. glutamicum*.

Die typische keulenförmige Morphologie und das v-förmige Abwinkeln der Zellen während der Zellteilung sind zu erkennen (Liebl, 2004).

1.2 Biotechnologische Bedeutung von *Corynebacterium glutamicum*

Zu den biotechnologischen Einsatzgebieten von Corynebakterien zählen überwiegend fermentative Produktionen von Aminosäuren (Leuchtenberger *et al.*, 2005). Auch bei der Herstellung und Umwandlung von Steroiden (Decréau *et al.*, 2003) und in der Käsereifung (Schröder *et al.*, 2011) sind Bakterien der Gattung *Corynebacterium* von Bedeutung. Neben *Escherichia coli* oder *Saccharomyces cerevisiae* ist das Gram-positive Bodenbakterium *C. glutamicum* heute einer der bedeutendsten industriell genutzten Organismen (Eggeling und Bott, 2005). Bei der gezielten Suche nach einem Mikroorganismus, der sich zur fermentativen Herstellung von L-Glutaminsäure eignet, wurde *C. glutamicum* im Jahr 1957 in Japan identifiziert (Kinoshita *et al.*, 1958). Hierbei wird die L-Glutaminsäure vom Bakterium als Folgeprodukt des Citratzyklus durch Transaminierung von 2-Oxoglutarinsäure produziert. Verschieden Produktionsstämme von *C. glutamicum* werden seither genutzt, um diese häufig als Geschmacksverstärker verwendete Aminosäure industriell in großen Mengen zu erzeugen und zwar mit einer weltweit jährlichen Produktion (2009) von circa 2,8 Mio. t (Ajinomoto, 2012). Als Produktionsstamm ist *C. glutamicum* hervorragend geeignet, da es aufgrund fehlender Pathogenität und Toxizität die Klassifikation GRAS (*generally regarded as safe*) besitzt und dazu verschiedene ökonomische Vorteile bietet. So wächst *C. glutamicum* vergleichsweise schnell in kostengünstigen Nährmedien, in denen landwirtschaftliche Nebenprodukte wie Melasse oder Stärke-Hydrolysat als Kohlenstoffquelle verwendet werden können und Ammoniak zumeist als Stickstoff-Quelle dient. Außerdem zeigt *C. glutamicum* hohe Wachstumsraten unter hohen Zuckerkonzentrationen (Ikeda und Takeno S., 2013). Gegenüber der chemischen Totalsynthese ausgehend von Acrylnitril werden bei der

fermentativen Gewinnung von L-Glutaminsäure vergleichsweise hohe Ausbeuten gemacht (Sano, 2009). Besonders vorteilhaft ist hierbei, dass im Gegensatz zur chemischen Synthese ist keine Racematen-Trennung notwendig ist. *C. glutamicum* hat sich auch für die Produktion einer Vielzahl anderer biologischer Verbindungen als geeignet erwiesen (Leuchtenberger *et al.*, 2005). Diese sind besonders Aminosäuren der Aspartat-Familie, wie die essentielle Aminosäure L-Lysin, welche mit einer jährlichen Produktion von 1,48 Mio. t hauptsächlich als Futtermittelzusatz große Verwendung findet (Ajinomoto, 2012). Zudem wird *C. glutamicum* für die Herstellung von Nukleotiden und Vitaminen wie z.B. Pantothensäure genutzt (Sahm und Eggeling, 1999).

1.3 Genomvergleich mit den verwandten Spezies *C. efficiens* und *C. diphtheriae*

Die vollständige Genomsequenz von *C. glutamicum* ATCC 13032 wurde im Jahr 2003 von zwei verschiedenen Arbeitsgruppen publiziert (Ikeda und Nakagawa, 2003; Kalinowski *et al.*, 2003), wobei eine Chromosomengröße von ~3,28 Mio. Basenpaaren (bp) ermittelt wurde. Mithilfe von rekombinanten DNA-Transfertechniken hat die Genomsequenz die Grundlage für die rationale Konstruktion verbesserter Produktionsstämme bereitet. In der aktuellen Genomannotation von *C. glutamicum* sind 3058 Gene kodiert, und es ergibt sich ein relativ hoher G+C-Gehalt von ~53 % (Kalinowski *et al.*, 2003). Grundsätzlich gibt es einige genomische Regionen in *C. glutamicum*, die einen davon unterschiedlichen G+C-Gehalt aufweisen und auf horizontalen Gentransfer hinweisen (HGC1- und LGC1-Region) oder auf die Integration von putativen Prophagen zurückzuführen sind (CGP1, CGP2, CGP3, CGP4) (Kalinowski *et al.*, 2003). Ungefähr gleichzeitig mit der Veröffentlichung des *C. glutamicum*-Genoms wurden die Sequenzen von *C. efficiens* (Nishio *et al.*, 2003) und *C. diphtheriae* (Cerdeño-Tárraga *et al.*, 2003) publiziert. Die Genome von *C. glutamicum* und *C. efficiens* zeigen ähnliche Genomgrößen (~3.2 Mbp), wobei *C. efficiens* jedoch noch zwei zusätzliche Plasmide enthält. Des Weiteren besitzen *C. glutamicum* und *C. efficiens* eine vergleichbare Anzahl von Genen (~3000) und auch die Aminosäuresequenzen der Proteine sind überwiegend gut konserviert. Nishio und Kollegen haben zudem den G+C-Gehalt und die Codon-Präferenz beider Spezies untersucht und festgestellt, dass sich die Genome in ihrem G+C-Gehalt um circa 10 Prozentpunkte unterscheiden, wobei das *C. glutamicum*-Genom mit 53 % den niedrigeren Gehalt aufweist (Tabelle II-1, nächste Seite). Außerdem wurde eine signifikante Verzerrung bei der Verwendung von Aminosäuren in Proteinen beobachtet. Die Austausche von Aminosäuren betreffen in *C. efficiens* häufig Lysin statt Arginin, Serin statt Alanin und Serin statt Threonin. Diese Substitutionen werden auch für die höhere Thermostabilität von *C. efficiens*-Proteinen verantwortlich gemacht, wodurch gutes Wachstum im Temperaturbereich über 40 °C möglich ist (ebd.).

Tabelle II-1 Vergleich der Genomsequenzen von *C. glutamicum*, *C. efficiens* und *C. diphtheriae* (modifiziert nach Eggeling und Bott, 2005).

Stamm	<i>C. glutamicum</i>	<i>C. efficiens</i>	<i>C. diphtheriae</i>
	ATCC 13032	YS-314	NCTC 13129
Größe [bp]	3.232.708	3.147.090	2.488.635
Durchschnittl. G+C-Gehalt [%]	53,8	63,4	53,5
ORF-Anzahl	3.002	2.950	2.320
Anzahl rRNA-Gene	6	5	5
Anzahl tRNA-Gene	60	56	54
Kodierregionen [%]	88,3	90,4	89,6

Das Genom des humanpathogenen Bakteriums *C. diphtheriae* zeigt mit nur ~2.4 Mbp Größe und circa 2300 Genen ein kleineres Genom als die beiden anderen genannten *Corynebacterium*-Spezies, aber im Vergleich zu *C. glutamicum* und *C. efficiens* sind eine ähnliche Anzahl von rRNA-Operons (5) und tRNA-Genen (54) annotiert (Cerdeño-Tárraga *et al.*, 2003). Auch die Anzahl von Kodierregionen und die mittlere ORF-Länge sind ähnlich groß (Tabelle II-1). Um Mutationen und Umordnungen von Genen in verschiedenen, aber verwandten Organismen zu vergleichen, werden häufig sogenannte Syntenie-Karten angefertigt, in denen die Genabschnitte aneinander ausgerichtet werden (Huynen und Bork, 1998). Der Begriff Syntenie entstammt dem Griechischen (*syn* = zusammen, *tainia* = Band) und bezeichnet Gene, die auf demselben Chromosom lokalisiert sind. Allerdings spricht man auch von Syntenie, wenn Gene in verschiedenen Organismen dieselbe Reihenfolge im Chromosom bzw. Genom aufweisen. Schon bald nachdem die ersten drei Corynebakterien Genome *C. glutamicum*, *C. efficiens*, *C. diphtheriae* sequenziert und annotiert waren, wurde die Konservierung der Genanordnung zwischen allen drei Spezies untersucht. Diese Analyse hat einen außerordentlich hohen Grad an Syntenie ergeben (Kalinowski *et al.*, 2003) (Abbildung II-2, nächste Seite). Eine mögliche Erklärung für diese ungewöhnliche Genomstabilität ist die Tatsache, dass Corynebakterien keine *recBCD*-Gene des rekombinatorischen Reparaturmechanismus besitzen (Nakamura *et al.*, 2003). Die Forscher vermuten, dass das Fehlen dieses Systems Gen-Umordnungen größeren Ausmaßes vorgebeugt hat.

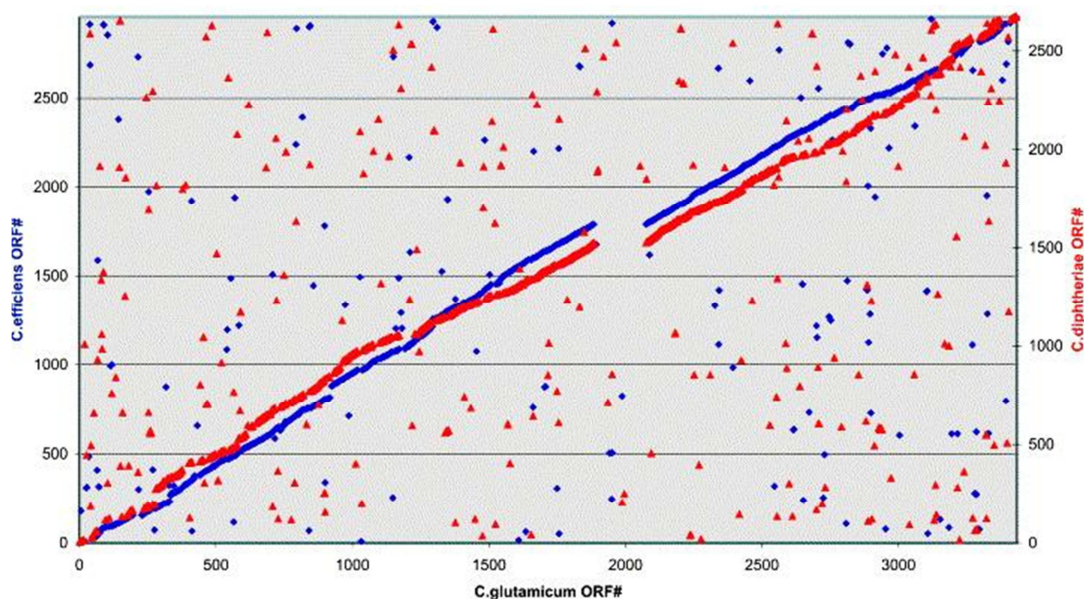


Abbildung II-2 Syntenie-Plot von *C. glutamicum*, *C. efficiens* und *C. diphtheriae*.

Die dargestellten Punkte ergeben sich aus den Ergebnissen des besten reziproken Treffers einer vergleichenden Aminosäure-Sequenzanalyse mit BLASTp (Altschul *et al.*, 1990) bezogen auf die mehr als 3000 ORFs der *C. glutamicum*-Sequenz. Treffer für *C. diphtheriae* sind als rote Dreiecke und Treffer für *C. efficiens* sind als blaue Punkte dargestellt (aus Kalinowski *et al.*, 2003).

2 RNA und die bakterielle Transkription

RNA ist die englische Abkürzung für *ribonucleic acid* und dient auch im deutschen Sprachgebrauch als Abkürzung für Ribonukleinsäure. RNA besteht aus Ribonukleotid-Einheiten, die wiederum Kondensationsprodukte aus einem Ribose-Zucker, einer Nukleinbase und einem Phosphat darstellen. Die Hypothese des zentralen Dogmas der Molekularbiologie (Crick, 1970) beschreibt RNA in lebenden Zellen als Vermittler der genetischen Information aus der Desoxyribonukleinsäure (DNA) während der Synthese von Proteinen. Auf der Suche nach dem Ursprung des Lebens stellt sich grundsätzlich die "Henne oder Ei-Frage": Um DNA herzustellen sind Enzyme notwendig, wobei die Information über den Aufbau der Enzyme in der DNA gespeichert ist - was war also zuerst da? Eine Antwort hierauf gibt die RNA-Welt-Hypothese (Gilbert, 1986). RNA kann sowohl wie DNA genetische Information speichern, aber auch wie Enzyme chemische Reaktionen katalysieren (Ribozyme; Kruger *et al.*, 1982). Die Hypothesen der chemischen Evolution erklären die Entstehung organischer Moleküle aus anorganischen Verbindungen (Miller, 1953; Miller und Urey, 1959). Da die abiotische Entstehung von RNA in der Ursuppe der Erde als möglich gilt, könnte diese den Übergang zum Aufkommen erster zellulärer Lebensformen darstellen (Gilbert, 1986). RNA ist der DNA sehr ähnlich, besitzt die fast gleichen vier Basen (Adenin, Cytosin und Guanin, nur Uracil statt Thymin) und trägt am C2-Atom des Ribose-Zuckers

zusätzlich ein Sauerstoffatom. Für eine Langzeitspeicherung von Informationen ist die DNA jedoch geeigneter als RNA, da durch die fehlende Hydroxylfunktion in der DNA eine Umesterung im Zucker-Phosphat-Gerüst verhindert wird (Chaudhuri, 2012). Des Weiteren geht der Baustein Cytosin relativ leicht eine Desaminierungsreaktion ein und wird zu Uracil. Die so entstandenen Fehler können in der DNA leicht identifiziert und korrigiert werden, während das Vorkommen von Uracil in der RNA normal ist und solche Basenaustausche nicht korrigiert werden können.

Um das Überleben einer Bakterienzelle und eine normale Zellfunktion zu ermöglichen, kodiert ein Bakteriengenom für viele tausend unterschiedliche Proteine. In Anpassung an besondere Umweltbedingungen werden verschiedene Proteine bzw. nicht alle Proteine zur gleichen Zeit in der gleichen Konzentration benötigt. Daher ist eine komplexe und vielschichtige Regulation der Genexpression nötig, um den Zellstoffwechsel und das Wachstum der Bakterien an die verschiedenen extrazelluläre Bedingungen anzugleichen (Wösten, 1998). Hierbei spielen Signaltransduktions-Mechanismen mittels Ein- oder Zwei-Komponentensysteme eine bedeutende Rolle, wobei letztere über eine Sensor-Histidinkinase Informationen ins Zellinnere auf Proteinregulatoren weiterleiten (Stock *et al.*, 2000). Im Rahmen von regulatorischen Prozessen kann die Initiation der Transkription entweder direkt oder auch posttranskriptionell beeinflusst werden. Dieses kann infolge des Mitwirkens von Proteinregulatoren, aber auch durch kleine RNA-Regulatoren geschehen. In den folgenden Kapiteln werden zunächst die Grundlagen der bakteriellen Transkription und deren Bestandteile für *C. glutamicum* erläutert, bevor auf posttranskriptionelle Prozesse im RNA-Metabolismus in Bakterien eingegangen wird.

2.1 Transkriptionsinitiation

Die unterste Ebene in der Genregulation ist die Initiation der Transkription zur Synthese von RNA, ausgehend von einer DNA-Vorlage durch das Schlüsselenzym RNA-Polymerase (RNAP). Die DNA-abhängige RNA-Polymerase (*core*-Enzym) benötigt zur Transkription ein zusätzliches Polypeptid, das Sigmafaktor (σ) genannt wird und an das RNAP *core*-Enzym binden kann. Das *core*-Enzym besteht aus insgesamt vier Untereinheiten $\alpha\alpha\beta\beta'$ und bildet zusammen mit dem Sigmafaktor das sogenannte RNAP-Holoenzym $\alpha\alpha\beta\beta'\sigma$ (Murakami und Darst, 2003). Die beiden α -Untereinheiten sorgen hierbei für die spezielle Struktur des Holoenzym und tragen durch eine Dimerbildung zur Stabilität bei. Die β -Untereinheit dient der Bindung an die DNA-Matrize und die β' -Untereinheit der Bindung an die Ribonukleosid-triphosphate (ebd.). In einer Zelle liegen dabei etwa 1500-2000 *core*-Enzyme vor, wovon jedoch der größte Teil bereits in die Transkription involviert ist (Jishage *et al.*, 1996). Die σ -Untereinheit erkennt und bindet direkt zwei konservierte Sequenzen der Promotor-Region, die als Konsensussequenz bekannt ist (Ghosh *et al.*, 2010). Die beiden Sequenzen liegen zum einen 35 und zum anderen 10

Basenpaare stromaufwärts vom Transkriptionsstart (TSS) entfernt, sind durch eine bestimmte Basenabfolge charakterisiert und meist durch einen 16-18 bp langen *spacer* getrennt (ebd.). Nach Bindung des RNAP-Holoenzym an die DNA bildet sich ein sogenannter „geschlossener Komplex“. Wird nun ein Teilstück der DNA entwunden, geht dieser in einen „offenen Komplex“ über und das RNAP-Holoenzym beginnt, einen kleinen bis zu 12 bp großen RNA-Abschnitt vom Transkriptionsstart aus (bezeichnet als +1) zu synthetisieren. Die Initiation der Transkription, die nur in Gegenwart von Mg^{2+} -Ionen abläuft, beginnt mit der Knüpfung der ersten Ribonukleotidbindung über eine Phosphodiesterbindung (ebd.). Die σ -Untereinheit löst sich im Anschluss vom RNAP-Holoenzym und das *core*-Enzym bewegt sich im Rahmen der Elongationsphase mitsamt zwei Elongationsfaktoren NusA und NusG entlang der fortlaufend entwundenen DNA und synthetisiert den neuen RNA-Strang bis es zur Transkriptionstermination kommt (Peters *et al.*, 2012).

Die Stärke eines Promotors bestimmt, mit welcher Frequenz das RNAP-Holoenzym die Transkription initiiert. Sie ist sowohl durch die Promotorsequenz als auch durch die Konformation der DNA bestimmt. Dabei existieren verschiedene Promotoren, die von verschiedenen Sigmafaktoren erkannt werden und in Bakterien die Anpassung an wechselnde Umweltbedingungen ermöglichen (Wösten, 1998). Es existieren mehrere, aber nur zwei große Familien von Sigmafaktoren: die σ^{70} - und die σ^{54} -Familie. Der Großteil der Sigmafaktoren gehört der σ^{70} -Familie an, die nach dem 70 kDa Hauptsigmafaktor aus *E. coli* benannt wurde (Silverstone *et al.*, 1972). Die σ^{54} -Familie ist nach dem 54 kDa großem Sigmafaktor aus *E. coli* benannt, welcher an der Stickstoffregulation beteiligt ist (Hirschman *et al.*, 1985; Hunt und Magasanik, 1985). Das Genom von *C. glutamicum* kodiert für insgesamt sieben Sigmafaktoren SigA, SigB, SigC, SigD, SigE, SigH und SigM (Kalinowski *et al.*, 2003), und diese lassen sich in essentielle, nicht essentielle und alternative Sigmafaktoren einteilen. Der Sigmafaktor SigA ist für die Expression der essentiellen konstitutiv exprimierten Gene (Haushaltsgene) verantwortlich und wird hauptsächlich während der exponentiellen Wachstumsphase benötigt (Pátek und Nešvera, 2011). Er besitzt Ähnlichkeit mit der σ^{70} -Familie aus *E. coli* und die zugehörige Konsensussequenz beruht auf 159 untersuchten mutmaßlichen SigA-abhängigen Promotoren (ebd.) (Abbildung II-3, nächste Seite). In *C. glutamicum* ist die SigA-abhängige Promotorsequenz insbesondere durch eine erweiterte (*extended*) -10-Region (GGTANNNT) gekennzeichnet. Eine solche Motiv-Erweiterung wurde auch für andere Gram-positive Bakterien beobachtet: In *Bacillus subtilis* ist zum Teil ein TGTG-Motiv an den Positionen -15 bis -18 bp stromaufwärts vom TSS konserviert beobachtet und als -16-Region bezeichnet worden (Voskuil *et al.*, 1995). Es wird vermutet, dass die -16-Region von Bedeutung ist für schwache Promotoren, aber eher ohne Bedeutung für starke Promotoren.

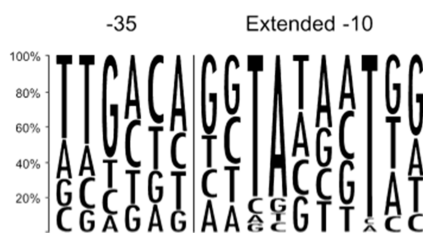


Abbildung II-3 SigA-Konsensussequenz in *C. glutamicum*.

Häufigkeits-Diagramm für die Nukleotid-Verteilung über 159 bekannte Promotoren (aus Pátek und Nešvera, 2011).

SigB ist der einzige charakterisierte nicht-essentielle σ^{70} -Faktor in *C. glutamicum* (Halgasova *et al.*, 2001) und seine Transkriptmenge steigt im Übergang von der exponentiellen in die stationäre Phase an (Larisch *et al.*, 2007). Gleichzeitig ist dieser Sigmafaktor für die Antwort auf Hitze-, Säure-, Kälte- und Ethanol-Stress verantwortlich und involviert in die stringente Antwort (Brockmann-Gretza und Kalinowski, 2006). Die Gruppe der *extracytoplasmic function* (ECF)-Sigmafaktoren wurde zuerst im Jahr 1994 beschrieben (Lonetto *et al.*, 1994). In *C. glutamicum* zählen SigC, SigD, SigE, SigG und SigH zu den ECF-Sigmafaktoren (Halgasova *et al.*, 2001). Die Regulation der Gene, die in der Reaktion der Zelle auf Oberflächen-Stress sowie Hitze-Stress benötigt werden, ist die Hauptaufgabe von SigE (Park *et al.*, 2008), wobei die Expression von SigE während des exponentiellen Wachstums eher gering ist. Es wird angenommen, dass die SigE-abhängige Promotorsequenz GGAATA (-35-Region) und GTTR (-10-Region) lautet, wobei R für A oder G stehen kann (Pátek und Nešvera, 2011). Eine Antwort auf Hitze- und oxidativen Stress liefert der Sigmafaktor SigM (Nakunst *et al.*, 2007). Aber auch SigH spielt eine wichtige Rolle bei der Antwort auf Hitze- und oxidativen Stress (Kim *et al.*, 2005; Ehira *et al.*, 2009). Zudem scheint es so, als wenn SigH eine zentrale Position im querregulierten Netzwerk der Sigmafaktoren einnimmt und dadurch eine abgestimmte Stressantwort in *C. glutamicum* ermöglicht. So scheint SigH auch in die Transkription von SigB und SigM involviert zu sein (Nakunst *et al.*, 2007). Es wird angenommen, dass SigH ein globaler Regulator im Netzwerk der Genregulation von *C. glutamicum* ist (Schröder und Tauch, 2010). Die Konsensussequenz von 45 bekannten SigH-abhängigen Promotoren lautet gGGAAta (-35-Region) und YGTTgaa (-10-Region) (Busche *et al.*, 2012). Dabei kann Y für ein C oder T stehen (Nukleotide in Großbuchstaben sind in mehr als 80 % der Fälle vertreten). Zusammenfassend lässt sich sagen, dass die alternativen Sigmafaktoren SigE, SigM und SigH relativ ähnliche Konsensussequenzen aufweisen. Die Sigmafaktoren SigC und SigD wurden bisher noch nicht im Detail charakterisiert. Analysen haben jedoch ergeben, dass SigD in die Transkriptionsregulation von Genen involviert ist, die bei niedriger Sauerstoffkonzentration erhöht sind (Ikeda *et al.*, 2009).

Zusätzlich zu Sigmafaktoren und der Promotorstärke können auch Proteinregulatoren an eine Operator-Region im 5'-Bereich von Genen binden und dadurch entweder die Transkription aktivierend oder reprimierend verändern. Das transkriptionelle Regulationsnetzwerk von *C. glutamicum* wurde in den

letzten Jahren intensiv untersucht (Schröder und Tauch, 2010). Aktuell sind in *C. glutamicum* 97 transkriptionelle Regulatorproteine bekannt, mit bisher 1443 identifizierten Interaktionen (CoryneRegNet Pauling *et al.*, 2012).

2.2 Transkriptionstermination

In Bakterien existieren grundsätzlich zwei unterschiedliche Mechanismen zum Beenden der Transkription. Bei der Rho-abhängigen Termination wird die *messenger* RNA (mRNA) erst nach Bindung eines Proteins von der DNA abgelöst. Dieses Rho-Faktor Protein ρ (Roberts, 1969) liegt in seiner aktiven Form als Homohexamer vor und bindet im Bereich der bereits synthetisierten mRNA aufwärts der Transkriptionsblase an eine ~80 nt lange unstrukturierte Cytosin-reiche Bindestelle mit dem Namen *rho utilization site* (*rut*). Danach bewegt sich das ρ -Protein als Translokase entlang der neu synthetisierten RNA in Richtung 3'-Ende der Transkriptionsblase entgegen (Banerjee *et al.*, 2006). Kommt es zur Pause der RNA-Polymerase, wird diese von ρ eingeholt und das Protein wirkt nun als ATP-abhängige Helikase (ebd.). Es löst die RNA vom Transkriptions-Elongationskomplex ab und beendet damit die RNA-Synthese (Richardson, 2003). Wie oben erwähnt, enthalten Bakterien zwei Elongationsfaktoren NusA und NusG, für die verschiedene Beobachtungen gemacht wurden. NusA und NusG binden mit ihrer N-terminalen Domäne an den Elongationskomplex und beschleunigen dadurch die Transkriptionsrate (Abbildung IV-3). Paradoxerweise scheint NusA, vor allem aber NusG, auch in die effiziente Rho-abhängige Termination involviert zu sein (Pasman und Hippel, 2000; Cardinale *et al.*, 2008). Es wurde gezeigt, dass NusG hierbei direkt mit dem ρ -Protein interagiert (Li *et al.*, 1992). Entdeckt wurde das NusG Protein ursprünglich im Rahmen der Antitermination des Phagen Lambda N-Proteins (Li *et al.*, 1992).

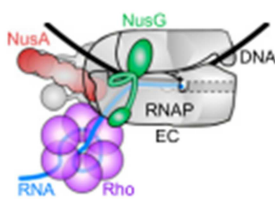


Abbildung II-4 Regulatoren der Transkription in Bakterien.

Der Elongationskomplex (EC) besteht aus RNAP, DNA Matrize und RNA Transkript. NusG steht in Kontakt mit RNAP über dessen N-terminale Domäne (NTD) und mit Rho über die C-terminale Domäne (CTD). Das Rho-Hexamer bindet nackte RNA. NusA bindet RNAP ebenfalls über dessen NTD (aus Peters *et al.*, 2012).

Die einfache oder Rho-unabhängige Termination funktioniert dagegen ohne Mitwirkung eines Proteins. Bei der Rho-unabhängigen Termination bildet die synthetisierte RNA in einer G+C-reichen Region am 3'-Ende eine Haarnadelschleife aus, gefolgt von einer Reihe aufeinanderfolgender Uridin-Reste (Brendel *et al.*, 1986). Es wird vermutet, dass die Haarnadelstruktur das RNA-DNA-Hybrid verkürzt und die instabile AU-Paarung den Komplex weiter destabilisiert. Darauf folgt die Transkript-Entlassung (Hippel und Yager, 1991).

2.3 RNA-Metabolismus und RNA-Stabilität

Bei Bakterien finden Transkription und Translation im Gegensatz zu Eukaryoten im gleichen Kompartiment statt. Die Ribosomen können sich daher bereits am wachsenden mRNA-Strang anlagern, so dass die Translation an der noch unfertigen mRNA beginnt. Dieser Vorgang wird als co-transkriptionelle Translation bezeichnet und ermöglicht den Zellen eine rasche Anpassung an veränderte Umgebungsbedingungen. Trotzdem kommt es nach der RNA-Polymerisation zum Teil noch zur Reifung der Transkripte über Prozessierung und/oder posttranskriptionelle RNA-Modifikationen. So sind zum Beispiel stabile rRNAs in der Zelle prozessierte Produkte eines Vorläufer-Primärtranskriptes, das die 16S-23S-5S rRNA als ein gesamtes prä-rRNA Molekül enthält (Deutscher, 2009). Modifikationen betreffen unter anderem die Methylierung von Cytosin zu 5-Methylcytidin (m^5C) in tRNAs (Edelheit *et al.*, 2013) oder die Adenin-zu-Inosin Umwandlung der Anticodon-Wobble-Base der tRNA^{Arg}, was als RNA-*editing* bezeichnet wird (Su und Randau, 2011). Im Anschluss an die Erzeugung durch die Transkription erfolgt nach einer gewissen Zeit auch der Abbau der Transkripte. Insgesamt lässt sich sagen, dass die verschiedenen Komponenten des RNA-Abbaus mit denen der RNA-Reifung überschneiden (Deutscher, 2006). Sie werden als Ribonukleasen (RNasen) bezeichnet und können in zwei Gruppen eingeteilt werden: Exoribonukleasen, die ein RNA-Nukleotid am 5'- oder 3'-Ende schneiden und Endoribonukleasen, die innere Phosphodiesterbindungen innerhalb der mRNA schneiden (Jain, 2002). Der Abbau von RNA ist allerdings ein hochorganisierter zellulärer Prozess, bei dem nicht nur Nukleasen, sondern auch verschiedene weitere Proteine beteiligt sind. Generell besteht ein grundlegender Unterschied zwischen dem Abbau von mRNA und von stabiler RNA: Stabile RNA bezeichnet meistens die Fraktionen von rRNA und tRNA, welche generell nur unter Stressbedingungen oder wenn RNA-Moleküle defekt sind abgebaut werden (Deutscher, 2006). Die durchschnittliche Halbwertszeit von mRNA in Bakterien ist im Vergleich dazu relativ kurz und schwankt zwischen 2 und 7 Minuten. Hiermit ist diese auch wesentlich kürzer als die Halbwertszeit der mRNA in eukaryotischen Zellen (Evguenieva-Hackenberg und Klug, 2011). Entsprechend ähnliche Beobachtungen wurden bei RNA-Stabilitätsanalysen für *C. glutamicum* gemacht, und zwar mit einer durchschnittlichen Halbwertszeit von ~5 Minuten (A. Mentz, nicht publiziert) (Abbildung II-5, nächste Seite). Die Stabilität der RNA hängt häufig von *cis*-agierenden Elementen ab, die die RNA durch stabile Stamm-Schleifen (*stem loops*) am 5'-Ende schützen (Emory *et al.*, 1992). Zusätzlich kann das Binden von Ribosomen oder das Binden von Proteinen die Abbaurate durch Schutz der mRNA vor Nuklease-Aktivität verändern (Jürgen *et al.*, 1998). Außerdem können endonukleolytische Schnittstellen über kleine regulatorische RNAs erzeugt, maskiert oder freigelegt werden (Evguenieva-Hackenberg und Klug, 2011). Im Bereich der 3'-untranslatierten Regionen (UTRs) von mRNAs sind ebenfalls oft *stem loops* vorhanden, die einerseits Transkriptionsterminatoren darstellen, aber im Fall polycistronischer RNA auch als Schutz gegen Exoribonukleasen dienen können (Klug *et al.*, 1987). Des Weiteren ist bekannt, dass die 3'-Enden bakterieller mRNAs durch das Enzym Poly(A)-Polymerase (PAP) polyadenyliert sein können, mit einer

Länge zwischen 15 nt und 60 nt Adenin-Basen (Sarkar, 1997). Die Eigenschaft Poly(A)-haltige 3'-Enden zu synthetisieren wurde ebenfalls für die 3'-5'-Exonuklease Polynukleotidphosphorylase (PNPase) in *E. coli* und auch für *S. coelicolor* berichtet (Mohanty und Kushner, 2000; Bralley *et al.*, 2006). Möglicherweise erleichtert die Polyadenylierung den Zugang für 3'-5'-Exonukleasen zum Transkript, denn in Anwesenheit von Transkripten mit Poly(A)-Schwänzen erhöht sich die Menge an Sequenzen, die für den Verdau zugänglich sind, stark (Rauhut und Klug, 1999). Im *C. glutamicum*-Genom sind diese beiden Gene ebenfalls vertreten, und zwar die PAP mit *cg3415* und die PNPase mit *cg2166*.

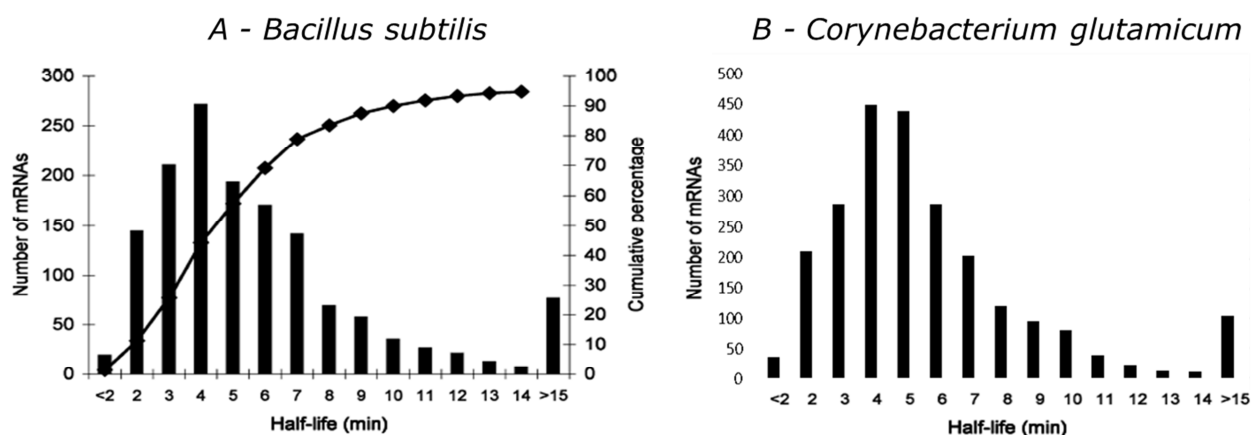


Abbildung II-5 Globale mRNA-Halbwertszeitanalysen über Rifampicin in *B. subtilis* und *C. glutamicum*.

A Berechnete Halbwertszeiten für *Bacillus subtilis* (aus Hambræus *et al.*, 2003), zusätzlich ist die prozentuale Anreicherung der Gene abgebildet. B *C. glutamicum* Halbwertszeiten für 2391 mRNAs über Microarray-Zeitreihenanalyse nach Zugabe von Rifampicin [100 µg/mL]. Die Berechnung dieser Werte erfolgte über mindestens vier aufeinanderfolgende Zeitpunkte der Probenahme nach 0, 2, 4, 6, 12, 16, 20 Minuten, Stringenz $R^2 \geq 0,9$ (A. Mentz, nicht publiziert).

Das Schlüsselenzym für den Start des RNA-Abbaus und den Zusammenbau des Degradosom-Enzymkomplexes ist die Endoribonuklease E (RNase E). Im Degradosom sind vier Proteine zu einem hochmolekularen Komplex organisiert, der aus der Endoribonuklease RNase E, einer RNA-Helikase RhlB, der Exoribonuklease Polynukleotidphosphorylase (PNPase) und einer Enolase besteht (Kushner, 2002). In Archaeen und auch in Eukaryoten übernimmt dagegen der sogenannte Exosom-Komplex diese Funktion (Lin-Chao *et al.*, 2007). Da bakterielle RNA-Primärtranskripte an ihrem 5'-Ende einen Triphosphat-Rest aufweisen, erleichtert die Pyrophosphohydrolase RppH in *E. coli* der RNase E das Schneiden von Primärtranskripten, indem sie am 5'-Ende die Pyrophosphate entfernt (Deana *et al.*, 2008). Infolgedessen erkennt die RNase E eine Schnittstelle am 5'-Ende der RNA und schneidet dort. Generell schneiden Endoribonukleasen meist in einzelsträngigen AU-reichen Regionen (Kaberdin, 2003), mit der Ausnahme von RNase III, die nur doppelsträngige RNA schneidet. Anschließend erfolgt der Abbau der Fragmente durch 3'-5'-Exoribonukleasen. Neben der PNPase gehören RNase II und RNase R zu den

wichtigsten 3'-5'-Exonukleasen in *E. coli* (Cheng und Deutscher, 2005). Zudem wurde in *Bacillus subtilis* eine Exoribonuklease J1 identifiziert, die Transkripte in 5'-3'-Richtung abbaut (Mathy *et al.*, 2007), ähnlich dem primären Abbaumechanismus eukaryotischer Transkripte.

Für kleine regulatorische RNAs mit kompakten und stabilen Sekundärstrukturen wurden in der Regel lange Halbwertszeiten zwischen 20-60 Min. beobachtet (Altuvia *et al.*, 1997; Barciszewski, 2003). Allerdings wurden im Gegensatz dazu nur sehr geringe Halbwertszeiten von ~1 Minute für *antisense* RNAs in Plasmiden beobachtet, die deren Replikation (CopA) und Segregation kontrollieren (Söderbom *et al.*, 1997).

2.4 RNA-Strukturen und Strukturvorhersagen

RNA-Struktur motive wurden in hoher Auflösung über Kernspinresonanzspektroskopie (NMR) und Röntgen-Kristallographiemethoden analysiert und haben einen Einblick in ihren charakteristischen Aufbau ermöglicht. Experimentell können Strukturen enzymatisch über partiellen Ribonuklease-Verdau (Waldminghaus *et al.*, 2007) oder auch chemischen Analyse (Waldminghaus *et al.*, 2009) bestimmt werden. Neben typischen Watson-Crick-Basenpaarungen existieren auch stabilisierende Nicht-Watson-Crick-Basenpaarungen sowie Base-Phosphat und Base-Ribose Wasserstoffbrücken (Shen *et al.*, 1995). Vollständig komplementäre RNA-Stränge können thermodynamisch stabile doppelsträngige antiparallele Helix-Stapel formen (Freier *et al.*, 1986). Viele Einzelstrang-RNAs bilden dagegen unvollständige Duplexe mit sich selbst, und bewirken Ausbuchtungen (*bulges*) sowie interne und Haarnadel/Stamm-Schleifen (*internal* und *hairpin stem loops*), die sich zu gut definierten Sekundärstrukturen falten (Tian *et al.*, 2004) (Abbildung II-6, nächste Seite). Einzelsträngige RNAs können über *loop-helix* oder *loop-loop* Interaktionen (*kissing-hairpins*) zur Bildung von Tertiärstrukturen wie Pseudoknoten führen (Dam *et al.*, 1992). Die häufigsten RNA *kissing-hairpins* in 16S und 23S rRNAs enthalten stabile dreidimensionale Tetranukleotid-Schleifen mit ungewöhnlichen Guanin-Adenin-Bindungen, und zeigen eine sogenannte GNR- Konsensussequenz (Woese *et al.*, 1990).

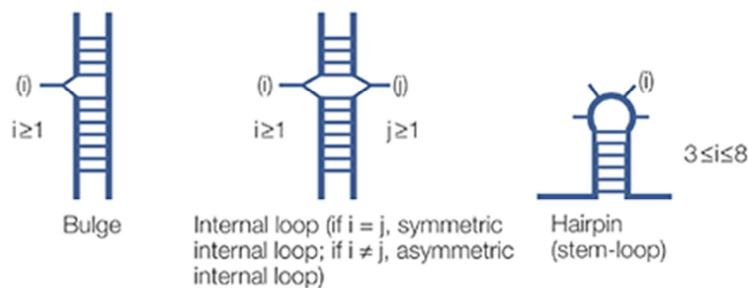


Abbildung II-6 RNA-Sekundärstrukturelemente.

Sekundärstrukturelemente sind definiert durch Muster gepaarter und ungepaarter Basen und bilden neben Helices typischerweise Ausbuchtungen (*bulges*), interne und Haarnadel-/Stamm-Schleifen (*internal* und *hairpin/stem loops*). *Bulges* besitzen ein oder mehrere ungepaarte Nukleotide in einem Strang. Interne Schleifen besitzen eine oder mehrere ungepaarte Nukleotide in beiden Strängen gegenüber. Symmetrische interne Schleifen liegen vor, wenn die Anzahl ungepaarter Nukleotide (angezeigt durch „i“ oder „j“) in beiden Strängen gleich ist, ansonsten liegt eine asymmetrische Schleife vor. Haarnadel- bzw. Stamm-Schleifen besitzen ungepaarte Nukleotide an der Spitze eines RNA-Doppelstranges (aus Tian *et al.*, 2004).

Mithilfe der publizierten Energiewerte für RNA-Strukturen wurden verschiedene Algorithmen zur Strukturvorhersage von RNAs und verschiedene Vorhersagemodelle entwickelt. Eine sehr ausführliche Liste der Programme ist unter dem englischen Wikipedia-Eintrag „*List of RNA structure prediction software*“ zu finden (en.wikipedia.org). Die ersten Algorithmen, die implementiert wurden, sind auch heute noch verbreitet, wie z.B. Mfold (Zuker und Stiegler, 1981) oder RNAfold (McCaskill, 1990). Hierbei wird die energetisch stabilste Sekundärstruktur für eine gegebene RNA-Sequenz berechnet bzw. die Konformation minimaler freier Energie (MFE) bestimmt. Ein weiterer Algorithmus, der auch in dieser Arbeit verwendet wurde ist RNAShapes (Steffen *et al.*, 2006). Die *shape*-Abstraktion (Form) behält die strukturellen Eigenschaften bei, vernachlässigt aber die Helix-Längen. Hierdurch wird die Anzahl suboptimaler Lösungen verringert, aber die signifikante Information bleibt erhalten. Außerdem repräsentieren *shapes* Strukturklassen, für die Wahrscheinlichkeiten berechnet werden können. Werden die Vertreter einer *shape*-Strukturklasse gruppiert, addieren sich ihre Wahrscheinlichkeiten. So kann es dazu kommen, dass *shapes* mit relativ schlechter Energie wahrscheinlicher sind als ein *shape* der MFE-Struktur (Giegerich *et al.*, 2004).

3 Regulatorische RNAs in Bakterien

Traditionell stellen die ribosomalen, *transfer* und *messenger* RNA-Moleküle (rRNAs, tRNAs, mRNAs) die Vermittler der genomischen Informationen auf dem Weg zur Proteinbiosynthese dar, wovon besonders rRNAs mit mehr als 95 % zu einem hohen Prozentsatz in den Zellen vorhanden sind (Peano *et al.*, 2013). Zusätzlich wurden innerhalb des letzten Jahrzehnts eine Vielzahl regulatorischer RNAs in allen drei Domänen des Lebens identifiziert. In Eukaryoten werden viele tausend regulatorische Transkripte in den intergenischen Regionen, Introns und auch in *antisense*-Richtung von proteinkodierenden Genen exprimiert (van Bakel *et al.*, 2010). Eine ähnlich pervasive Transkription existiert auch in Bakterien (Toledo-Arana *et al.*, 2009) und Archaeen (Jäger *et al.*, 2009). In Bakterien wird hierbei unterschieden zwischen *cis*-regulatorischen RNA-Motiven, die meistens in den 5'- oder 3'-untranslatierten Bereichen von proteinkodierenden Genen vorliegen, und eigenständigen regulatorischen RNA-Genen, die keine Proteine kodieren. Die Vertreter der zuletzt genannten Kategorie werden daher häufig auch mit dem Synonym nicht-kodierende bzw. *non-coding* RNAs (ncRNAs) bezeichnet. Andererseits spricht man auch von kleiner bzw. *small* RNA (sRNA), da diese Transkripte meist nur bis ~300 nt lang sind (Storz *et al.*, 2011). Vereinzelt wurde auch eine Bifunktionalität beobachtet und proteinkodierende RNAs können gleichzeitig regulatorische RNAs darstellen (Vanderpool *et al.*, 2011). Alle bisher identifizierten RNAs sind in der RNA-*families* Datenbank (Rfam) gesammelt (Gardner *et al.*, 2009). In *C. glutamicum* war bis zu den Ergebnissen dieser Arbeit (Mentz *et al.*, 2013) nur wenig über kleine RNAs und ihre potentiellen regulatorischen Funktionen bekannt und nur wenige regulatorische RNA-Gene waren über Kovarianzmodelle der Rfam-Datenbank vorhergesagt. Außer den rRNA-Genen und tRNA-Genen sind aktuell keinerlei Informationen über RNA-Spezies in der Genomannotation enthalten. Im Folgenden werden die Klassen der *cis*-regulatorischen RNAs und nicht-kodierenden kleinen RNAs werden einzeln vorgestellt und auch die Rfam-Datenbank soll näher erläutert werden.

3.1 *Cis*-regulatorische RNA-Motive von proteinkodierenden Genen

Eine fertige mRNA in Bakterien enthält vor und hinter der kodierenden Sequenz meistens noch UTRs. Die 5'-UTR liegt vor der Kodierregion und enthält einige Basen aufwärts des Start-Codons die Purinreiche ribosomale Bindestelle, auch als Shine-Dalgarno Sequenz bekannt, die den Einstiegspunkt der Ribosomen für die Translation darstellt (Shine und Dalgarno, 1974). Nur in speziellen Fällen liegen sogenannte *leaderless*-Transkripte vor, die keinerlei UTR besitzen und zuerst für den Bakteriophagen Lambda beschrieben wurden (Ptashne *et al.*, 1976). Über *in silico*-Analysen wurde besonders in Actinobakterien ein relativ hoher Prozentsatz mit bis zu 20 % für diese Art von Transkripten festgestellt (Zheng *et al.*, 2011). Da der Transkriptionsstart hier sogleich den Translationsstart darstellt, gibt es Vermutungen, dass bei *leaderless*-Transkripten eine Box stromabwärts des TLS für die Bindung an die

16S rRNA verantwortlich ist und so die Initiation der Translation unterstützt. Der experimentelle Nachweis hierfür ist bis heute allerdings nicht gelungen (Zheng *et al.*, 2011). Die 5'-UTRs sowie 3'-UTRs (stromabwärts des Translations-Stopp-Codons) können aus mehreren hundert Nukleotiden bestehen und eine komplizierte Sekundärstruktur aufweisen. Solche Sekundärstrukturen können regulatorische Aufgaben übernehmen, stehen aber zum Teil auch mit der Stabilität der Transkripte im Zusammenhang.

Ein weit verbreiteter Mechanismus in der Regulation der Genexpression in Bakterien und Archaeen ist die transkriptionelle Attenuation, die über einen strukturellen *switch* in der mRNA charakterisiert ist und über Ausbildung eines Rho-unabhängigen Terminators zur frühzeitigen Transkriptionstermination führen kann. Dieser Mechanismus kann von der Ribosomen-Aktivität, vom Binden bestimmter ungeladener tRNAs (T-Box Attenuation), vom Binden eines kleinen Metaboliten (Riboswitch) oder von Temperaturänderungen (RNA-Thermometer) (Naville und Gautheret, 2010) abhängig sein. Die Ribosomen-vermittelte Attenuation basiert auf der Tatsache, dass in Bakterien Transkription und Translation gleichzeitig stattfinden. In der Regel ist diese Form der Attenuation durch ein Leaderpeptid gekennzeichnet, das am 5'-Ende eines Operons mit großer Relevanz für einen bestimmten Aminosäure-Biosyntheseweg lokalisiert ist, wie im Fall des Tryptophan (Trp)-Operon in *E. coli* (Yanofsky, 1981). Eine Leaderpeptid-mRNA enthält verschiedene zueinander komplementäre Bereiche und eine Region aufeinanderfolgender gleicher Codons, an denen das Ribosom in Abhängigkeit von der Verfügbarkeit geladener tRNAs entweder schnell oder nur langsam vorankommt (Abbildung II-7). Pausiert das Ribosom an diesen Codons, kann es über Basenpaarkomplementaritäten zur Ausbildung eines Anti-Terminators in der mRNA kommen, und die Transkription läuft weiter. Teilweise bewirken Protein-RNA-Interaktionen die strukturelle Umwandlung in der Leader-mRNA, wie für das TRAP-Protein in *Bacillus subtilis* gezeigt wurde (Babitzke und Yanofsky, 1993).

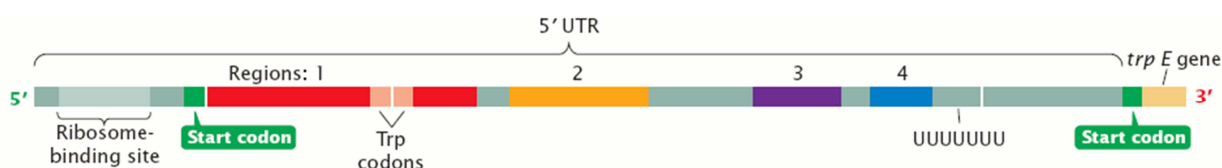


Abbildung II-7 Schematische Abbildung der 5'-UTR des Trp-Operons in *E. coli*.

Ist die Konzentration geladener $tRNA^{Trp}$ -Moleküle hoch, paaren Region 3 (lila) und 4 (blau), bilden eine Terminatorstruktur und beenden die Transkription (transkriptionelle Attenuation). In diesem Fall paaren ebenfalls die Regionen 1 (rot) und 2 (orange). Ist die $tRNA^{Trp}$ -Konzentration dagegen niedrig, bleibt das Ribosom an einer Reihe aufeinanderfolgender Trp-Codons in Region 1 (rot) hängen. Hierdurch kommt es zur Paarung komplementärer Basen von Region 2 (orange) und 3 (lila), die eine Anti-Terminatorstruktur ausbilden. Die Transkription des *trpE* Gens läuft in diesem Fall fort (modifiziert nach Ralston, 2008).

T-Box Attenuatorelemente beeinflussen meistens die Expression der zugehörigen Aminoacyl-tRNA-Synthetase-Operons und binden spezifisch an ungeladene tRNAs, wobei hohe Konzentrationen an ungeladenen tRNAs die Ausbildung einer Anti-Terminator Struktur begünstigen (Merino und Yanofsky, 2005). Transkriptionelle Attenuation kann auch durch Riboswitches verursacht werden. Riboswitches sind in den 5'-UTRs der mRNA von verschiedenen metabolischen Wegen lokalisiert und bestehen in ihrer komplexen Architektur aus zwei Funktionselementen, und zwar einem Aptamer und einer Expressionsplattform. Die Expressionsplattform ändert infolge der Aptamer-Bindung ihre Konformation und reguliert dadurch die Genexpression (Tucker und Breaker, 2005). Dabei kann das Aptamer je nach Spezifität verschiedene niedermolekulare Metaboliten binden. Flavin-Mononukleotid (FMN)-bindende Riboswitches sind in die Riboflavin-Biosynthese involviert (Gelfand *et al.*, 1999). Thiamin-Pyrophosphat (TPP)-bindende Riboswitches sind in die Thiamin-Biosynthese involviert bzw. regulieren den Thiamin-Transport (auch THI-Box genannt) (Miranda-Ríos *et al.*, 2001), aber auch Cobalamin (Vitamin B12), S-Adenosyl-Methionin (SAM), Lysin, Glycin, Purine und weitere kleine Liganden können von Riboswitches gebunden werden (Nudler und Mironov, 2004). Neben einer Transkriptionstermination wurden weitere Mechanismen beobachtet, die durch Riboswitches vermittelt werden können. Diese betreffen meistens Strukturänderungen, welche zur Blockade der ribosomalen Bindestelle führen und so die Translation inhibieren (Mironov *et al.*, 2002). Andererseits können Riboswitches aber auch Ribozym-Aktivität zeigen und zur katalytischen mRNA-Prozessierung führen. Dieses wurde für einen Riboswitch stromaufwärts des *glmS* Gens in Gram-positiven Bakterien gezeigt, der Glucosamin-6-Phosphat binden kann (Barrick *et al.*, 2004). Über alternative Sekundärstrukturen können auch RNA-Thermometer zum Abbruch der Transkription führen, wie es für Kälte-Schock Proteine in *E. coli* vermutet wird (Naville und Gautheret, 2010). Die meisten RNA-Thermometer regulieren jedoch die Translation von stromabwärts liegenden Operons und wirken auf die Weise, dass sie die Zugänglichkeit zur Shine-Dalgarno-Sequenz bei Temperaturerhöhung meistens freigeben, wie im Fall des FourU-Typs oder ROSE-Typs (Narberhaus *et al.*, 2006).

3.2 Nicht-kodierende kleine regulatorische RNAs in Bakterien

Wie schon erwähnt, sind regulatorische RNAs in Bakterien meist relativ klein, können aber auch bis zu 500 nt lang sein (Sorek und Cossart, 2010) wie für RNAlII in *Staphylococcus aureus*, der größten bekannten bakteriellen regulatorischen RNA mit 482 nt gezeigt wurde (Benito *et al.*, 2000). In einer Vielzahl von Bakterien wurden die meisten der vier folgenden RNA-Haushaltsgene als konserviert beobachtet. Die M1 RNA bildet zusammen mit dem C5-Protein die RNase P und ist in die Reifung der tRNA-Vorläufer Moleküle involviert (Evans *et al.*, 2006). Die 4.5S RNA (*ffs*) bildet zusammen mit dem Protein Ffh den sogenannten bakteriellen *signal recognition particle* (*bacterial_SRP*) und wird für den Prozess der Proteinsekretion benötigt (Poritz *et al.*, 1990). Die *transfer messenger* RNA (tmRNA, SsrA)

(Ray und Apirion, 1979; Muto *et al.*, 1998) hat ihre Aufgabe in der Proteinsynthese und befreit mithilfe des Proteins SmpB blockierte Ribosomen von unvollständigen Peptidketten. Für diese Aufgabe sind zwei Bestandteile der tmRNA von Bedeutung: Einerseits ahmt diese Strukturelemente von tRNAs nach und andererseits kodiert sie für ein kurzes Peptid-Tag, das die unvollständigen Peptidketten für den Abbau markiert (Keiler, 2008). Es fällt auf, dass diese Haushalts-RNAs Bestandteil von Ribonukleoprotein-Komplexen sind (Abbildung II-8). Die 6S RNA (SsrS) wurde als nicht ganz so stark konserviert beobachtet (Gardner *et al.*, 2009) und blockiert durch Nachahmen der Transkriptionsblase den σ^{70} -Faktor der RNA-Polymerase während der stationären Wachstumsphase (Wassarman, 2007).

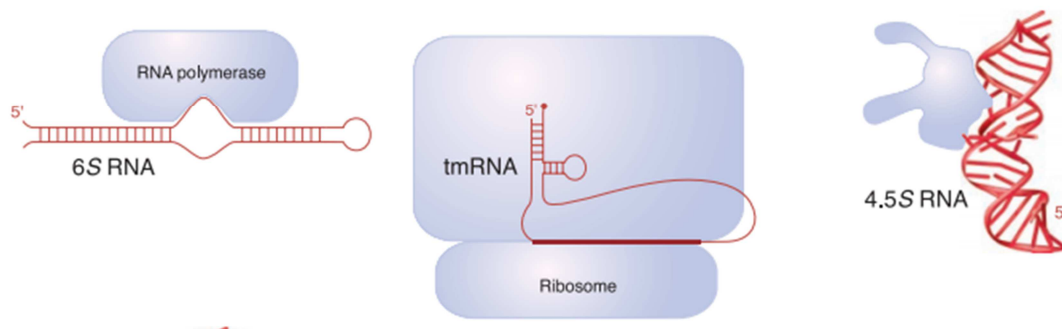


Abbildung II-8 Beispiele von hochkonservierten bakteriellen RNA-Haushaltsgenen.

Die 6S RNA und tmRNA ahmen beide Nukleinsäuren nach. Die tmRNA und 4.5S RNA sind Bestandteil von Ribonukleoprotein-Komplexen (aus Storz, 2002).

Insgesamt sind zwei Interaktionskategorien beschrieben: Regulatorische RNAs können entweder mit mRNAs oder auch mit Proteinen interagieren (Waters und Storz, 2009). In den hier aufgeführten Funktionen von RNA-Haushaltsgenen ist eine RNA-Protein Interaktion vertreten (6S RNA). Für die Mehrzahl der identifizierten und untersuchten kleinen regulatorischen RNAs in Bakterien wurden allerdings eher posttranskriptionelle Interaktionen mit mRNAs beobachtet (Storz *et al.*, 2011), wobei diese Interaktionen zum Teil durch Helferproteine wie dem Hfq-Protein in *E. coli* unterstützt werden (Valentin-Hansen *et al.*, 2004). Daneben wurden auch noch andere RNA-bindende Helferproteine in Bakterien beobachtet wie das RNA-Chaperon StpA (Waldsich *et al.*, 2002), das Histon-ähnliche Protein HU51 (Balandina *et al.*, 2002) und der transkriptionelle Regulator H-NS (Deighan *et al.*, 2000). Generell werden kleine RNA-Gene in zwei Typen unterschieden: *Trans*-kodierte RNA-Gene sind in Entfernung von ihrer Ziel-mRNA zwischen den annotierten proteinkodierenden Regionen lokalisiert. *Cis-antisense* RNA-Gene sind dagegen direkt *antisense* zu ihrer Ziel-mRNA kodiert (Abbildung II-9, nächste Seite). Diese Tatsache führt im Allgemeinen dazu, dass eine *cis-antisense* RNA (asRNA) vollständig komplementär mit ihrer Ziel-mRNA paart, wohingegen eine *trans*-kodierte RNA nur teilweise komplementär zu ihrer Ziel-mRNA ist (Waters und Storz, 2009) und über eine stabile, konservierte Sekundärstruktur an ihre Ziel-mRNA bindet. Des Weiteren besitzen kleine RNAs einen eigenen Promotor

und es findet häufig eine vom Rho-Faktor unabhängige Transkriptionstermination statt. Durch Bindung der *trans*-kodierte RNAs nahe der RBS wird an ihrer Ziel-mRNA primär eine Translationsinhibierung oder -aktivierung als rasche Antwort auf Stress, andere Umwelteinflüsse und verschiedene zelluläre Prozesse bewirkt. Bakterielle kleine RNAs wurden bisher schon im Zusammenhang mit Chromosomen Replikation und Zellteilung (DicF RNA Bouché und Bouché, 1989), mit der Zellteilung (6S RNA Wassarman und Storz, 2000), mit Virulenz und *quorum sensing* (Qrr RNA Lenz *et al.*, 2004), mit RNA-Prozessierung (RNaseP Kazantsev und Pace, 2006), mit Proteinstabilität (tmRNA Muto *et al.*, 1998), mit Plasmidreplikation (RepE RNA Chai und Winans, 2005), mit Kohlenstoffspeicherung (CsrBC Weilbacher *et al.*, 2003), mit Stress-Antwort (OxyS RNA Altuvia *et al.*, 1998) und mit Oligonukleotid-Transport (GcvB RNA Urbanowski *et al.*, 2000) beobachtet. Grundsätzlich kann eine mRNA das Ziel von mehreren regulatorischen RNAs darstellen, was zuerst für die Translation des RNA-Polymerase Sigmafaktors RpoS gezeigt wurde (Majdalani *et al.*, 2002). Andererseits kann eine kleine RNA auch mehrere Ziel-mRNAs regulieren, was zuerst für die kleine RNA DsrA in *E. coli* gezeigt wurde (Lease und Belfort, 2000). Heute ist bekannt, dass kleine RNAs eine Menge von Genen regulieren können, aus denen sich sogar regulatorische Netzwerke ergeben (Beisel und Storz, 2010; Storz *et al.*, 2011).

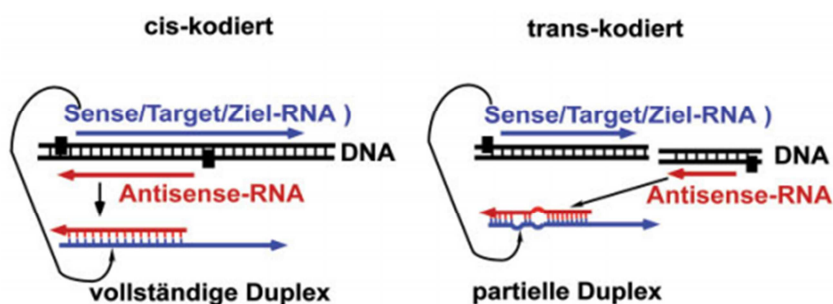


Abbildung II-9 *Cis-antisense* RNAs und *trans*-kodierte RNAs.

Cis-antisense kodierte RNAs (links) sind direkt gegenüber ihrer Ziel-mRNA lokalisiert und bilden mit dieser vollständige Duplexe. *Trans*-kodierte RNA (rechts) sind dagegen in Distanz zu ihrem Ziel lokalisiert und bilden mit diesem nur partielle Duplexe (modifiziert nach Brantl, 2008).

Die Anzahl von *cis-antisense*-Transkripten, die im Gegenstrang eine mRNA überlappen, variiert in den bisher untersuchten Bakterien zwischen 3-50 % (Lasa *et al.*, 2011). Weitergehende Analysen haben zur Entdeckung verschiedener neuartiger regulatorischer Funktionen von *cis-antisense* RNAs geführt. So kann eine überlappende asRNA auch hier entweder zur Blockade oder Öffnung der RBS der gegenüberliegenden mRNA führen. Des Weiteren können *cis-antisense* RNAs die Stabilität der Ziel-mRNA fördern (Opdyke *et al.*, 2004) oder ihren Abbau durch Ribonukleasen beschleunigen (Silvaggi *et al.*, 2005). Außerdem können durch Bindung einer asRNA Strukturänderungen in der gegenüberliegenden mRNA hervorgerufen werden und so zur Transkriptionstermination über transkriptionelle Attenuation

führen (Brantl und Wagner, 2002). Als transkriptionelle Interferenz wird ein Mechanismus bezeichnet, bei dem die simultane *in vivo*-Aktivität von zwei konvergierenden Promotoren im *sense*- und *antisense*-Strang die Aktivität des schwächeren Promotors durch die des Stärkeren in *cis* unterdrückt (Shearwin *et al.*, 2005).

Eine spezielle Gruppe von nicht-kodierenden RNAs basiert auf der Grundlage von DNA-Abschnitten, die als CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) bezeichnet werden und in ~45 % der bislang sequenzierten Bakterien-Genome und ~83 % der Archaeen-Genome detektiert wurden (CRISPRdB Stand September 2013, Grissa *et al.*, 2007). CRISPRs stellen kurze, sich wiederholende DNA-Abschnitte dar und sind von *spacern* ähnlicher Länge (bis zu 72 bp) getrennt. Dieser spiegel-symmetrische, allerdings nicht ganz palindromische, Aufbau hat zur Folge, dass die zugehörigen RNA-Transkripte eine stabile *hairpin*-ähnliche Sekundärstruktur aufweisen (Kunin *et al.*, 2007). Es konnte gezeigt werden, dass einige *spacer* Sequenzidentität mit Plasmiden und Phagen-DNA aufweisen (Bolotin *et al.*, 2005; Mojica *et al.*, 2005). Zusammen mit dem Cas-Proteinkomplex können CRISPRs eine Abwehr gegen exogene DNA bewirken (Barrangou *et al.*, 2007). Dieser Mechanismus weist Parallelen zur RNA-Interferenz (RNAi/*gene silencing*) (Fire *et al.*, 1998) in Eukaryoten auf. Es bestehen jedoch auch Unterschiede, und wie einige der Vorgänge im Detail ablaufen, ist noch unbekannt (Marraffini und Sontheimer, 2010).

In *C. glutamicum* ist bisher wenig über kleine regulatorische RNAs und ihre potentiellen regulatorischen Einflüsse bekannt. Es kann allerdings von der Genomsequenz abgeleitet werden, dass ähnlich wie in anderen Vertretern der Ordnung *Actinomycetales* (Vockenhuber *et al.*, 2011) kein Homolog des Hfq-Helferproteins existiert. Auch Homologe des RNA-Chaperons StpA (Waldsich *et al.*, 2002), des Histon-ähnliche Proteins HU51 (Balandina *et al.*, 2002) und der transkriptionelle Regulator H-NS (Deighan *et al.*, 2000) wurden nicht gefunden. In den Genomen der meisten anderen *Corynebacterium*-Spezies wurde zudem mindestens eine CRISPR-Region identifiziert (siehe CRISPRdB) (Grissa *et al.*, 2007), dies trifft allerdings nicht für *C. glutamicum* zu. Bis vor kurzem war die ArnA RNA die einzige experimentell beschriebene kleine RNA in *C. glutamicum*, die stromaufwärts und *antisense* zur mRNA des GntR-Regulators *cg1935* detektiert wurde (Zemanová *et al.*, 2008).

3.3 Die Rfam-Datenbank und Rfam-Vorhersagen für *C. glutamicum*

Die Rfam-Datenbank ist eine öffentlich zugängige Datenbank des Wellcome Trust Sanger Institutes (Cambridge, UK) und beinhaltet die Daten kleiner regulatorischer RNAs und anderer struktureller RNA-Elemente in allen Organismen. Andere Datenbanken dieser Art sind zum Beispiel fRNAdb (Kin *et al.*, 2007) und sRNAdb (Pischmarov *et al.*, 2012). In der Rfam-Datenbank sind jedoch nicht nur Sequenzinformationen gesammelt. Über bekannte homologe Sequenzen einer RNA-Spezies wird hier ein *Seed-Alignment* mittels der Software Infernal ("INFERence of RNA ALignment" Griffiths-Jones *et al.*, 2003) erzeugt, mithilfe dessen im Anschluss ein Kovarianzmodell erstellt wird. Dieses Kovarianzmodell stellt eine Art Sequenzprofil dar, in dem die Konsensussequenz und die konservierte RNA-Sekundärstruktur kombiniert sind. Hierüber können dann RNA-Homologe in anderen Organismen erfolgreicher als über ein Primärsequenz-Alignment identifiziert werden. Wenn die Infernal Software versucht, eine Eingabesequenz mit dem Kovarianzmodell zu vereinen, wird ein *bit score* erzeugt. Dieser Zahlenwert stellt den binären Logarithmus der Wahrscheinlichkeit der Eingabesequenz im Kovarianzmodell über das Nullmodell dar (Formel 1). In der Theorie bedeutet das, dass ein positiver *bit score* signifikant ist. In der Praxis gilt allerdings: Je besser das Modell passt, desto höher ist der *bit score*-Zahlenwert.

Formel 1
$$\text{Rfam } bit \text{ score} = \log_2 \left(\frac{P_{\text{cm}}}{P_{\text{null}}} \right)$$

Aktuell existieren für *C. glutamicum* 101 Rfam-vorhergesagte und als signifikant eingestufte RNA-Motive (*bit score* > 32), die als Tabelle öffentlich/online bereitgestellt werden. Die Mehrheit dieser Vorhersagen betrifft die annotierten sechs verschiedenen Kopien der 5S rRNA-Gene und 55 tRNA-Gene, sowie assoziierte rRNA-Strukturen der zwölf *archaeal small subunit ribosomal* RNAs innerhalb der 5S rRNA-Gene (*SSU_rRNA_archaea*; RF01959) und sechs Pseudoknoten *PK-G12rRNAs* (RF01118) in den verschiedenen 16S rRNA-Genkopien. Darüber hinaus werden sieben *cis*-regulatorische Riboswitch-Motive vor proteinkodierenden Genen vorhergesagt, wovon fünf einen Thiamin-Pyrophosphat-abhängigen Riboswitch darstellen (TPP-Riboswitch/THI-Element; RF00059). Zusätzlich sind ein S-Adenosylmethionin-abhängiger (SAM-IV; RF00634) und ein Flavin-Mononukleotid-abhängiger Riboswitch (FMN-Riboswitch; RF00050) vorhergesagt. Als weitere *cis*-regulatorische Motive sind drei *yybP-ykoY* (RF00080), zwei *ydaO-yuaA*-Motive (RF00379), das *mraW*-Motiv (RF01746), *msik* (RF01747), *cspA* (RF01766), und das *mini-ykkC*-Motiv (RF01068) und die Ribozyme Gruppe II katalytischer Introns (RF02004 und RF00029) zu erwähnen. Außerdem ist die aus Actinomyceten bekannte 6C RNA (RF01066) vorhergesagt (Weinberg *et al.*, 2010) sowie die drei oben erwähnten hochkonservierten RNA-Haushaltsgene 4.5S RNA (*Bacteria_small_SRP*; RF00169), tmRNA (RF00023) und M1 RNA (*RNaseP_bact_a*; RF00010). Das aus anderen Bakterien bekannte 6S RNA-Haushaltsgen ist nicht vorhergesagt. Auf die Genom-Positionen dieser Elemente in *C. glutamicum*, die fast alle im Rahmen dieser Arbeit validiert werden konnten, wird im Ergebnisteil eingegangen.

4 Methoden zur Identifizierung kleiner RNAs in Bakterien

Die ersten kleinen RNAs wurden schon vor einigen Jahrzehnten entdeckt, jedoch waren bis zum Jahr 2000 nur zehn solcher Gene in dem gut charakterisierten Enterobacterium *E. coli* bekannt (Wassarman *et al.*, 1999). Zu den ersten identifizierten RNAs zählen die RNA-Haushaltsgene 4.5S, 6S, tmRNA und M1 RNA, die früher über radioaktive Markierung der Gesamt-RNA und anschließende Gel-Isolierung identifiziert worden sind. Im Bereich experimenteller Strategien sind seit dem letzten Jahrzehnt systematische Analysen begonnen worden, wodurch viele hundert kleine RNAs in den untersuchten Bakterien detektiert wurden. In diesem Zusammenhang sind *shotgun*-Klonierungen kleiner RNAs (*RNomics*), SAGE (*Serial Analysis of Gene Expression*), aber vor allem genomweite Microarraystudien zu nennen (Vogel und Sharma, 2005). Seit wenigen Jahren leistet der technologische Fortschritt moderner Hochdurch-Sequenzierungen bzw. *next generation sequencing* (NGS)-Methoden einen bedeutenden Beitrag zur Identifizierung von kleinen RNAs (Sharma und Vogel, 2009; Wang *et al.*, 2009). In den Experimenten wurde dabei entweder die Fraktion kleiner RNAs aus Gesamt-RNA-Proben verwendet oder RNA, die mithilfe von Co-Immunopräzipitation über das Helfer-Protein Hfq gewonnen wurde (Zhang *et al.*, 2003; Sittka *et al.*, 2008). Während der letzten Jahre und besonders während der Arbeiten für diese Dissertation sind auch verschiedene Algorithmen und bioinformatische Tools entwickelt worden, die sich zur *in silico*-Vorhersage von kleinen RNAs eignen (Sridhar und Gunasekaran, 2013). Die einzelnen Methoden zur Identifizierung kleiner regulatorischen RNAs werden im Folgenden beschrieben.

4.1 Bioinformatische Suche nach kleinen RNAs

Die bioinformatische Suche nach kleinen RNAs ist im Vergleich zu proteinkodierenden Genen schwieriger, denn regulatorische kleine RNAs besitzen weder Start- noch Stopp-Codons noch eine ribosomale Bindestelle, die sich zur effektiven Identifizierung von Proteinen eignen. Die ersten Vorhersagen kleiner RNAs wurden über komparative Genomanalysen mit dem Programm QRNA in *E. coli* erzeugt (Rivas und Eddy, 2001) und basieren auf der Beobachtung, dass kleine RNAs häufig in den intergenischen Regionen nah verwandter Bakterien konserviert sind. In der Bioinformatik können mithilfe von Methoden der komparativen Genomik DNA-Sequenzen so angeordnet werden, dass es möglich ist, Regionen mit Ähnlichkeit zu identifizieren, die aus funktioneller, struktureller oder evolutionärer Verwandtschaft resultieren. Zur Erstellung dieser sogenannten Alignments wird häufig auf die Programmiermethode „*dynamic programming*“ zurückgegriffen, die sich zum Lösen von Optimierungsproblemen eignet (Nalbantoğlu, 2014) und von dem Mathematiker Richard Bellman eingeführt wurde (Bellmans Prinzip; Bellman, 1957). Grundsätzlich wird in bioinformatischen Alignments zwischen zwei Formen unterschieden. Bei einem globalen Alignment kommt es zur globalen Optimierung über die Gesamtlänge aller Sequenzen (Needleman-Wunsch Algorithmus; Needleman und Wunsch, 1970). Im Gegensatz dazu bilden lokale Alignments bestimmte Regionen innerhalb von langen

Sequenzen, die sonst eher voneinander unterschiedlich sind (Smith-Waterman Algorithmus; Smith und Waterman, 1981). Zusätzlich können paarweise Sequenz-Alignments (*pairwise sequence alignments*) berechnet werden, die das am besten übereinstimmende stückweise oder globale Alignment zweier Eingabesequenzen bestimmen. Eine Möglichkeit der Quantifizierung besteht in der Methode von *maximum unique matches* (MUM), wobei längere MUMs typischerweise für stärkere Ähnlichkeit und nähere Verwandtschaft von zwei Sequenzen stehen. Ein multiples Sequenz-Alignment stellt die Erweiterung eines paarweisen Alignments dar um mehrere Sequenzen gleichzeitig zu analysieren. Dieses kann mithilfe verschiedener bioinformatischer Tools erstellt werden wie z.B. ClustalW (Larkin *et al.*, 2007) oder Mauve (Darling *et al.*, 2004).

Auch die bereits beschriebenen Vorhersagen über Kovarianzmodelle über bekannte RNA-Sequenzen aus der Rfam-Datenbank beruhen auf komparativer Genomik. Diese Methode hat jedoch zur Bestimmung von nur wenigen RNA-Elementen außerhalb der Klasse der gut studierten γ -Proteobakterien geführt. Das Programm RNAz (Washietl *et al.*, 2005) implementiert dagegen einen anderen Ansatz, und kann dazu verwendet werden um in multiplen Genom-Alignments nah verwandter Bakterien nach konservierten Sekundärstrukturen mit thermodynamischer Stabilität zu suchen, die charakteristisch für kleine RNAs sind. Das Programm RNAz wurde im Rahmen dieser Arbeit zur Identifizierung kleiner RNAs in *C. glutamicum* angewendet und wird daher in Material und Methoden detaillierter besprochen. Zeitgleich mit der Durchführung dieser Arbeit wurden auch noch weitere bioinformatische Tools/Programme entwickelt. Grundsätzlich besteht auch die Möglichkeit der *ab initio*-Suche von RNAs in bakteriellen Genomen. So kann das Tool smyRNA (Salari *et al.*, 2009) kleine RNAs über die unterschiedliche Verteilung der Basenzusammensetzung in einer Genomsequenz erkennen. Hierbei zeigen Regionen, die für eine regulatorische RNA kodieren, einen höheren G+C-Gehalt als die Hintergrund-Genomsequenz. Dieses Tool wurde speziell für thermophile Bakterien entwickelt und erzielt dort auch gute Ergebnisse (Sridhar und Gunasekaran, 2013). Ein anderes Programm zur *ab initio*-Vorhersage genannt NAPP (*Nucleic Acids Phylogenetic Profiling*) (Ott *et al.*, 2012) bedient sich der Beobachtung, dass regulatorische kleine RNAs besonders häufig in intergenischen Regionen mit phylogenetischer Konservierung lokalisiert sind (Sridhar und Rafi, 2007). Die über NAPP erzeugten *contigs* ergeben zwar die Regionen, in denen vermutlich RNAs kodiert sind, sagen aber keine genauen Start- und Stopp-Positionen für diese RNAs vorher.¹ Neben den bereits vorgestellten Methoden gibt es weitere charakteristische Bestandteile von regulatorischen RNAs, die ebenfalls zur Entwicklung von Vorhersage-

¹ Die vorhergesagten Daten der NAPP-*Contigs* für *C. glutamicum* ATCC 13032 (569 intergenische Regionen) befinden sich im Anhang auf der Daten-CD im Ordner „Vorhersage“ in der Datei C_glutamicum_ATCC_13032_Bi-NAPP-Contigs.xlsx.

Tools eingesetzt wurden. Diese sind Transkriptionsmerkmale wie beispielsweise Promotoren, Transkriptionsfaktor-Bindestellen oder Rho-unabhängige Terminatoren. In diesem Zusammenhang ist der sRNAPredict Algorithmus (Livny *et al.*, 2005) von Bedeutung, auf dessen Basis im Laufe der Jahre immer wieder neue Programme aufgebaut wurden, wie sRNAPredict2 (Livny *et al.*, 2006) oder SIPHT (Livny *et al.*, 2008). SIPHT wendet für circa 1500 bakterielle Genome ein automatisiertes Verfahren an, welches die Konservierung der Primärsequenz mittels BLASTn (Altschul *et al.*, 1990) in intergenischen Regionen stromaufwärts von Rho-unabhängigen Terminatoren untersucht. Auf diesem Programm beruht auch eine bestehende Vorhersage für kleine RNAs in *C. glutamicum*, die im Rahmen der SIPHT-Publikation von Livny und Kollegen durchgeführt wurde (Livny *et al.*, 2008). Hierbei wurden die 4.5S RNA und tmRNA mit Haushaltsfunktion detektiert und weitere 15 kleine RNAs vorhergesagt, die bisher in keiner Datenbank gelistet sind (Tabelle II-2).

Tabelle II-2 Vorhersage kleiner RNAs über SIPHT in *C. glutamicum* ATCC 13032 (aus Livny *et al.*, 2008).

Name	Strang	Start	Stopp	Benachbarte Gene
Candidate_1_NC_006958	+	520496	520707	cg0577(+)/cg0578(+)
Candidate_2_NC_006958	-	2882538	2882238	cg3021(+)/cg3022(+)
Candidate_3_NC_006958	-	1091553	1091478	cg1178(-)/cg1179(+)
Candidate_4_NC_006958	-	857352	856963	cg0928(+)/trRNA(+)
Candidate_5_NC_006958	+	520496	520588	cg0577(+)/cg0578(+)
Candidate_6_NC_006958	-	1626891	1626720	cg1735(-)/cg1736(-)
Candidate_7_NC_006958	-	2110076	2110017	cg2221(-)/cg2222(-)
Candidate_8_NC_006958	-	2267029	2266865	cg2377(-)/cg2378(-)
Candidate_9_NC_006958	-	2644837	2644632	cg2780(-)/cg2781(-)
Candidate_10_NC_006958	+	1047865	1048031	cg1128(+)/cg1129(+)
Candidate_11_NC_006958	+	3122806	3123034	trRNA(-)/trRNA(-)
Candidate_12_NC_006958	+	3141674	3141947	cg3285(-)/cg3286(+)
Candidate_13_NC_006958	+	3141674	3141979	cg3285(-)/cg3286(+)
Candidate_14_NC_006958	+	509704	509926	cg0571(+)/cg0572(+)
Candidate_15_NC_006958	-	806978	806774	cg0877(+)/cg0878(-)

4.2 Microarray-basierte Identifizierung kleiner RNAs

Mithilfe der Microarraytechnologie ist die Grundlage für die simultane Expressionsanalyse einer Vielzahl von Genen auf Transkriptebene geschaffen worden (Fodor *et al.*, 1991). Diese Methode ist damit deutlich effizienter gegenüber Einzelgen-Transkriptanalysen über RT-qPCR oder Northern Blot. Das Prinzip von Microarrayanalysen ist es, eine Hybridisierung von fluoreszenzmarkierten Nukleinsäure-Sonden (meist cDNA) gegen immobilisierte Oligonukleotidproben durchzuführen und die Sequenz-spezifischen

Bindungen anschließend zu detektieren, wobei unspezifische Bindungen möglichst minimiert werden sollten. Dabei existieren zwei unterschiedliche Herstellungsarten von Microarrays. Eine Möglichkeit besteht in der *in situ*-Synthese von Oligonukleotiden auf einem Microarray. Diese Methode der „photolithographischen Festphasensynthese“ wurde von der Firma Affymetrix® entwickelt. Eine alternative Technik stellt die an der Stanford University entwickelte externe Synthese von Oligomeren dar, die anschließend auf der Microarrayoberfläche über mechanisches Kontakt-Verfahren oder piezoelektrisch immobilisiert werden (Sчена *et al.*, 1995). Die Immobilisierungsmatrix kann entweder aus Glas, Polymeren oder Nylon-/Nitrocellulosemembranen bestehen. *In situ* hergestellte Microarrays sind aufgrund des standardisierten Herstellungsverfahrens weniger anfällig gegenüber räumlichen Fehlern; zudem werden spezielle *mismatch*-Sonden zur Fehlerkorrektur eingesetzt (Gao *et al.*, 2004). Das experimentelle Design von Microarrays kann sehr unterschiedlich ausfallen. Als Sonde können entweder eine fluoreszenzmarkierte Probe (*one-channel* Microarrays) oder zwei Proben mit unterschiedlicher Fluoreszenzmarkierung (*two-channel* Microarrays) zum Einsatz kommen. Im letztgenannten Fall wird meistens gegen eine Referenz hybridisiert. Für einen Vergleich zwischen Probe und Referenz ist dann ein M/A-Diagramm interessant. Man betrachtet dabei logarithmische Intensitätsverhältnisse zwischen dem roten und grünen Kanal für den *Spot* i (R = Kanal 1, G = Kanal 2)

Formel 2
$$M_i = \log_2 \frac{R_i}{G_i}$$

sowie die mittlere *Spot*-Intensität (Maß für die Helligkeit des *Spots*)

Formel 3
$$A_i = \log_2 \sqrt{R_i G_i}$$

Je nach verwendeter Microarrayplattform werden die RNA-Proben nach der Isolierung entweder direkt präpariert und hybridisiert, oder es erfolgen noch Schritte, die eine reverse Transkription und Fluoreszenzmarkierung der Proben beinhalten. Im Anschluss an die Hybridisierungen wird der Microarray gescannt. Dabei kommt es zur Messung der Emissionssignale bei den exzitorischen Wellenlängen der Fluoreszenzfarbstoffe. Im Anschluss sollte immer eine Normalisierung der Rohdaten durchgeführt werden, da die absoluten Signalintensitäten durch eine Reihe von technischen Varianzen beeinflusst sein können (Quackenbush, 2001). Danach können statistische Analysen erfolgen oder Methoden der Mustererkennung angewendet werden, um komplexe Daten zu strukturieren und zu visualisieren. Beispiele hierfür sind hierarchische Cluster, *K-means*, *self-organizing-maps* (SOMs) und die *principle-component-analysis* (PCAs) (Allison *et al.*, 2006).

Die Anzahl an Oligonukleotidproben, die auf einem Array Platz finden, ist im Laufe der Jahre immer größer geworden und heute können Microarrays Millionen von Proben enthalten, die z.B. gleichmäßig über das gesamte Genom verteilt sind (*tiling*-Microarrays) (Mockler *et al.*, 2005). Solche *tiling*-Microarrays verschiedener Auflösung und Dichte sind auch bei Studien zur Suche nach kleinen RNAs in

anderen Bakterien zum Einsatz gekommen (Selinger *et al.*, 2000; Wassarman *et al.*, 2001; Tjaden *et al.*, 2002; Zhang *et al.*, 2003; Toledo-Arana *et al.*, 2009). Bei der Identifizierung kleiner RNAs können generell einige Schwierigkeiten auftreten, denn die geringe Größe, geringe Expression und stabile Sekundärstrukturen erschweren die cDNA-Synthese. Eine spezielle Detektionsmethode arbeitet mit Antikörpern, die spezifisch an RNA-DNA-Hybride binden und so cDNA- oder RNA-Markierungsprobleme umgehen kann (Zhang *et al.*, 2003). Heute werden im Rahmen von Transkriptomanalysen immer häufiger nur noch Hochdurchsatz-Sequenzierungen durchgeführt. Doch auch in den vorbereitenden Schritten zur Erstellung von Sequenzierung-Bibliotheken können Verzerrungen (Bias) auftreten, und quantitative RNA-Seq-Messungen sind noch nicht standardisiert. Die Anwendung der Microarraytechnologie kann daher auch noch heute sinnvoll sein, um Transkriptlevel zu bestimmen (Malone und Oliver, 2011).

4.3 Transkriptomanalyse über Hochdurchsatz-Sequenzierung (RNA-Seq)

Die Entwicklung von Hochdurchsatz-Sequenzierungstechniken hat einen fundamentalen Fortschritt auf dem Gebiet der Genom- und Transkriptom-Forschung mit sich gebracht. Die Einsatzgebiete dieser Hochdurchsatz-Sequenzierungstechniken reichen heute von *de novo* und Re-Sequenzierung von Genomen (Huang *et al.*, 2009; Brinkrolf *et al.*, 2013), Bisulfit-Sequenzierung (Cokus *et al.*, 2008), Chromatin-Immuno-Präzipitations-Sequenzierung (Chip-Seq) (Park, 2009) bis RNA-Sequenzierung (RNA-Seq) (Mortazavi *et al.*, 2008; Sittka *et al.*, 2008) und ermöglichen auch die Untersuchung genetischer Variationen innerhalb der einer Spezies (*single-nucleotide* Polymorphismen, SNPs) (Craig *et al.*, 2008). Die Durchführung von RNA-Seq-Experimenten ermöglicht nun seit ein paar Jahren den bisher detailliertesten Blick auf das Transkriptom einer Zelle (Güell *et al.*, 2011). Einen Vorläufer dieser RNA-Technik kann man in der 1995 entwickelten SAGE-Technologie (Velculescu *et al.*, 1995) und ihren Varianten sehen. SAGE-Experimente ermöglichen nicht nur wie Microarrays die Analyse bekannter Gene, sondern auch die Detektion unbekannter Gene über die Identifizierung von kurzen cDNA-Fragmenten, sogenannten *tags*. Der Vorteil von Transkriptomanalysen über die RNA-Seq-Technologie ist nun, dass eine genomweite Auflösung für einzelne Nukleotide erreicht wird, es gibt keine störenden Hintergrundsignale wie bei der Fluoreszenz-basierten Detektion, keine störenden Kreuzhybridisierungen, und insgesamt entsteht ein fast uneingeschränkter dynamischer Detektionsbereich (Wang *et al.*, 2009). Ein Nachteil dieser Methode ist jedoch, dass sie weiterhin vergleichsweise teuer ist. Im Rahmen bakterieller RNA-Seq-Studien wurden sowohl die Illumina (Chaudhuri *et al.*, 2011) und 454 (Sittka *et al.*, 2008) als auch die SOLiD (Passalacqua *et al.*, 2009) Sequenzierungsplattformen erfolgreich eingesetzt. Zusammengefasst kann man sagen, dass RNA-Seq zur umfangreichen Verbesserung von Genom-Annotationen führen kann, z.B. in Gesamt-Transkriptomanalysen (Liu und Camilli, 2011) und durch Detektion neuer bakterieller regulatorischer kleiner RNAs (Schlüter *et al.*, 2010; Mraheil *et al.*, 2011; Vockenhuber *et al.*, 2011; Wilms *et al.*, 2012). Darüber hinaus sind mittlerweile spezifische enzymatische

Methoden etabliert, die sich zur Bestimmung genomweiter Transkriptionsstarts (Sharma *et al.*, 2010) oder zur Detektion definierter Prozessierungsstellen innerhalb der Transkripte eignen (Underwood *et al.*, 2010).

Um cDNA-Bibliotheken für effiziente RNA-Sequenzierungen zu erstellen, muss eine Anzahl von technischen Schwierigkeiten überwunden werden. Zunächst sollte die Mehrheit der Gesamt-RNA (> 95 % Peano *et al.*, 2013) möglichst vollständig abgereichert werden. Dies kann entweder über Hybridisierung an rRNA-komplementäre Nukleotidsequenzen erfolgen, die ihrerseits an magnetische *beads* gebunden sind, oder auch über enzymatische Abreicherung prozessierter Transkripte (Sorek und Cossart, 2010). Daneben ist es wichtig, die Strang-Information beizubehalten, um später in den Daten zwischen *sense*- und *antisense*-Transkripten unterscheiden zu können. Dies kann über Ligation von Adaptern bekannter Sequenz an die RNA vor der cDNA-Synthese, PCR-Amplifikation und Cluster-Generierung erreicht werden (Sharma *et al.*, 2010). Nach der Sequenzierung beinhaltet die bioinformatische Auswertungs-Prozedur das Filtern von *reads* mit geringer Qualität und die Kartierung der erhaltenen *reads* auf das Referenz-Genom (Marguerat und Bähler, 2010). Heute steht besonders die Entwicklung von Programmen zur sinnvollen Verwaltung der generierten Sequenzierungs-Daten im Vordergrund, und es werden bioinformatische Tools zur Statistische Analyse der Sequenzierungs-Daten implementiert (Costa *et al.*, 2010).

5 Identifizierung von Funktionen regulatorischer *trans*-kodierter RNAs

Die Zuordnung der biologischen Funktion von kleinen RNAs gestaltet sich weitaus schwieriger als die reine Identifikation. Durch den Fortschritt neuer Sequenzierungsmethoden sind in den untersuchten Bakterien Hunderte kleine *trans*-kodierte RNAs bekannt, aber nur einigen davon konnte bisher eine konkrete Funktion zugeordnet werden (Vogel und Wagner, 2007; Prasse *et al.*, 2013). In der sRNATarBase-Datenbank (Cao *et al.*, 2010) werden alle experimentell verifizierten Interaktionen gespeichert. Als eine der ersten kleinen RNA-Funktionen wurde die von *MicF*, einer Porin-regulierenden kleinen RNA in *E. coli* aufgeklärt, die einen RNA-Duplex mit der 5'-UTRs der *ompF* mRNA formt und dadurch deren Translation inhibiert (Andersen *et al.*, 1989; Andersen und Delihias, 1990). Mit der Aufdeckung der ersten RNA-Regulationsnetzwerke in Enterobakterien der Gattungen *Escherichia* und *Salmonella* (Beisel und Storz, 2010; Storz *et al.*, 2011) gewinnt die Bestimmung der Funktionen kleiner RNAs nun immer mehr an Bedeutung. Bei der Bestimmung eines Ziels und der Funktion einer regulatorischen RNA können bioinformatische Programme hilfreich sein, sie können jedoch nur Hinweise zur Interaktion liefern, und experimentelle Analysen sind zur Bestätigung unbedingt notwendig. Im Bereich der Bioinformatik gibt es verschiedene Ansätze, die zur Identifizierung von einer oder mehreren

Ziel-mRNAs führen können (Backofen und Hess, 2010; Li *et al.*, 2012). Zum Teil werden dabei nur intermolekulare Interaktionen betrachtet, wobei die Struktur der interagierenden Moleküle vernachlässigt wird. Solche reinen Komplementaritätsanalysen können über Programme wie TargetRNA (Tjaden, 2008) durchgeführt werden. Hierbei wird ein Hybridisierungs-*score* analog zum Smith–Waterman Algorithmus für lokale Alignments zwischen den interagierenden RNA-Molekülen berechnet (Smith und Waterman, 1981). Allerdings mit der Ausnahme, dass anstatt des Homologiepotenzials das Basenpaarpotenzial bewertet wird. Zusätzlich sind aber auch intramolekulare Strukturen bzw. thermodynamische Berechnungen von großer Bedeutung für eine Interaktion, und können über die Ermittlung der Zugänglichkeit (*accessibility*) der interagierenden Moleküle berücksichtigt werden. Wird ein Zugänglichkeitsvergleich der Ziel-mRNA vor und nach der Bindung einer regulatorischen RNA durchgeführt, kann dadurch sogar der Effekt der Interaktion vorhergesagt werden. Standardmäßig wird die Zugänglichkeit darüber berechnet, wie viel an freier Energie benötigt wird, um einen Bereich von vier aufeinanderfolgenden Nukleotiden zu öffnen. Hierfür kommen Programme wie RNAup (Mückstein *et al.*, 2006), IntaRNA (Busch *et al.*, 2008) oder auch RNAPlex (Tafer und Hofacker, 2008) in Frage. In RNAup wird die gesamte freie Energie ΔG einer Interaktion berechnet über die Interaktionsenergie und unter Berücksichtigung der Energien, die zur Strukturöffnung der Interaktionspartner A und B benötigt werden (Formel 4).

Formel 4
$$\Delta G = \Delta G_{A \text{ Öffnung}} + \Delta G_{B \text{ Öffnung}} + \Delta G_{\text{Interaktion}}$$

In RNAPredator (Eggenhofer *et al.*, 2011) sind über RNAPlex die Zugänglichkeitsprofile von Genen für viele Genome schon vorberechnet und werden über ein Energiemodell mit den putativen intermolekularen Interaktionen kombiniert. Ist eine *trans*-kodierte RNA in mindestens drei verschiedenen Genomen konserviert, kann auch der CopraRNA (*Comparative Prediction Algorithm for sRNA Targets*)-Ansatz zur Suche nach einer Ziel-mRNA angewendet werden (Wright *et al.*, 2013). Dieses Tool integriert zusätzlich phylogenetische Informationen in genomweite IntaRNA-Vorhersagen (Busch *et al.*, 2008). An dieser Stelle sollen aber auch die Limitierungen der bioinformatischen Ansätze genannt sein, denn der Effekt von RNA-Helferproteinen wie Hfq (Valentin-Hansen *et al.*, 2004; Soper *et al.*, 2010) in Interaktionen kann natürlich zu Problemen in der Berechnung führen.

Experimentelle Ansätze zur Bestimmung von kleinen regulatorischen RNAs beginnen häufig mit der Erstellung einer Deletionsmutante einer kleinen RNA. Anhand von phänotypisch auffälligen Merkmalen in Wachstum, Virulenz oder Morphologie der Deletionsmutante können anschließend Rückschlüsse auf die putative Funktion einer kleinen RNA gemacht werden. Im zweiten Schritt kommen vergleichende Analysen auf Proteomebene zwischen dem Wildtypbakterium und der Deletionsmutante hinzu, die die globale Auswirkung einer RNA auf das Proteinmuster bzw. die Translation sichtbar machen können

(Rasmussen *et al.*, 2005). Über vergleichende Genexpressionsanalysen auf Transkriptebene (z.B. Microarrays) kann die Funktion einer kleinen RNA ebenfalls deutlich werden (Massé *et al.*, 2005), wobei zusätzlich auch der Einfluss einer kleinen RNA in Bezug auf die mRNA-Stabilitäten erkennbar wird (Papenfert *et al.*, 2006). Bei all diesen Experimenten sollte aber nicht vergessen werden, dass nicht nur direkte Interaktionen sichtbar werden, sondern auch indirekte Auswirkungen auf Transkriptionsregulatoren und deren Netzwerke möglich sind, die zur veränderten Expression von verschiedenen Genen führen können. Direkte Nachweise von bekannten RNA-mRNA Interaktionen sind bisher über ein 2-Plasmidexpressionssystem in *E. coli* gelungen (Urban und Vogel, 2007). Das Prinzip dieses Systems besteht darin, eine putative 5'-UTR-Zielsequenz als Plasmid-gebundene Translationsfusion vor die kodierende Sequenz von Grün-fluoreszierendem Protein (GFP) zu klonieren und in *E. coli* Zellen zu integrieren. Der Vergleich der Zellfluoreszenz vor und nach Integration eines zweiten Plasmids, das die kleine RNA trägt, kann dann den regulatorischen Effekt einer kleinen RNA auf die Translation direkt sichtbar und messbar machen. Außerdem konnte über *in vitro* toeprinting-Assays (*primer extension inhibition*) bestätigt werden, dass bestimmte kleine RNAs an der RBS mit dem Ribosom konkurrieren (Hartz *et al.*, 1988; Udekwu *et al.*, 2005; Gaubig *et al.*, 2011). Mithilfe von radioaktiv-markierten RNAs (Brantl und Wagner, 1994) konnten verschiedene Interaktionen verifiziert werden. Hierbei wurden auch Mobilitäts (*band-shift*)-Assays von *in vitro* synthetisierten kleinen RNAs und nicht-markierter Ziel-RNA durchgeführt und die veränderte Gel-Retardation im System überprüft (Wilms *et al.*, 2011).

Vor Beginn dieser Arbeit waren in *C. glutamicum* lediglich einige wenige hochkonservierte RNAs über Kovarianzmodelle vorhergesagt und in der Rfam-Datenbank hinterlegt (Gardner *et al.*, 2009). Die Hauptaufgabe dieser Dissertationsarbeit bestand darin, kleine regulatorische RNAs in diesem Gram-positiven Bakterium zu identifizieren und diese zu klassifizieren. Dieser Hauptteil der Dissertationsarbeit mit den Ergebnissen der cDNA-Sequenzierung kleiner RNAs wurde bereits publiziert (Mentz *et al.*, 2013). Hierbei wurde auch die bioinformatische Vorhersage kleiner RNAs aus dieser Arbeit für *C. glutamicum* veröffentlicht. Im Ergebnisteil dieser Arbeit werden zusätzlich *tiling*-Microarrayanalysen zwischen den annotierten proteinkodierenden Genen in *C. glutamicum* präsentiert, um die Experiment-spezifische transkriptionelle Aktivität der in der RNA-Seq erhaltenen kleinen RNAs zu analysieren. Das letzte Ergebniskapitel befasst sich mit Transkript- und Funktionsanalysen verschiedener und in dieser Arbeit detektierter kleiner RNAs.

III. Material und Methoden

1 Bakterienstämme, Plasmide, Primer

Tabelle III-1 Verwendete Bakterienstämme

Stamm	Relevante Eigenschaften	Referenz
<i>Corynebacterium glutamicum</i> ATCC 13032	Wildtyp, Nx ^r	ATCC
<i>C. glutamicum</i> <i>Δcgb_16115_Reg123</i>	Deletion in <i>cgb_16115</i> von Interaktionsregionen 1&2&3; Nx ^r (stromabwärts von <i>cg1611</i>)	(Neshat, 2010)
<i>C. glutamicum</i> <i>Δcgb_16115_Reg1</i>	Deletion in <i>cgb_16115</i> von Interaktionsregion 1; Nx ^r (stromabwärts von <i>cg1611</i>)	Diese Arbeit
<i>C. glutamicum</i> <i>pheA::IS6100</i>	Transposonmutante mit Insertion im <i>pheA</i> Gen (Phe-auxotroph)	(Mormann <i>et al.</i> , 2006)
<i>Escherichia coli</i> JM109	Klonierstamm	New England Biolabs

Tabelle III-2 Verwendete Plasmide

Plasmid	Relevante Eigenschaften	Referenz
pK18mobsacB	Mobilisierbarer <i>E. coli</i> Klonierungs- vektor pK18 <i>mob</i> , ermöglicht die Selektion auf Doppel- <i>crossover</i> in <i>C. glutamicum</i> , beinhaltet zusätzlich ein <i>sacB</i> -Derivat, <i>lacZα</i> , <i>mcs</i> , Km ^r	Schäfer <i>et al.</i> , 1994

Tabelle III-3 Verwendete Primer

Verwendete Primer (für Methode)	Sequenz ^a
<i>cgb_16115_d1-EcoRI</i> (GeneSOEing)	GGTGGTgaattcCAGATCACGTGCACCAAGTT
<i>cgb_16115_d2new-HindIII</i> (GeneSOEing)	CTTTCATGCGGTAGATGACAagcttCCAGCTCTTGTAGCTTAGGT
<i>cgb_16115_d3new-HindIII</i> (GeneSOEing)	aagcttGTCATCTACGGCATGGAAAG
<i>cgb_16115_d4-BamHI</i> (GeneSOEing)	GGTGGTggatccTGGTGGCGTTGAACTCATCG
5S RNA-Kontrolle- <i>sense</i> (NB*)	GTGTCGGTGGTGATAGTAGC
5S RNA-Kontrolle- <i>antisense</i> +T7 (NB*)	GAAATTAATACGACTCACTATAGGG-GTCGGCGGTAACCTACTCTC
6C_RNA- <i>sense</i> (NB*)	CAAGGCCCGATATACAGTG
6C_RNA- <i>antisense</i> +T7 (NB*)	GAAATTAATACGACTCACTATAGGG-CAAGGCCCGATATACAGTG
<i>cgb_00105-sense</i> (NB*)	GTGATTGCAGTGACTGATGT
<i>cgb_00105-antisense</i> +T7 (NB*)	GAAATTAATACGACTCACTATAGGG-TTTAAGCGGACGCACGGTCA
<i>cgb_20715-sense</i> (NB*)	TACAGGGGGTTTAGTTGTGG

<i>cgb_20715-antisense</i> +T7 (NB*)	<u>GAAATTAATACGACTCACTATAGGG</u> -TACAGGGGGTTTAGTTGTGG
<i>cgb_16115-sense</i> (NB*)	GCTGGTTTGCTGAAAAATGG
<i>cgb_16115-antisense</i> +T7 (NB*)	<u>GAAATTAATACGACTCACTATAGGG</u> -GGGGTGAGATGCTACAAAA
<i>aroR_sense</i> (NB*)	GCTGGTTTGCTGAAAAATGG
<i>aroR_antisense</i> +T7 (NB*)	<u>GAAATTAATACGACTCACTATAGGG</u> -GGGGTGAGATGCTACAAAA
<i>aroR_5'</i> (Amplifikat i) (RT-qPCR)	ACTATGTTTCGCAGCACGATT
<i>aroR_3'</i> (Amplifikat i) (RT-qPCR)	GCTCGACTCCGGTTCTGTTC
<i>aroG_5'</i> (Amplifikat ii) (RT-qPCR)	CGGCGTCAACCAGCAACAAG
<i>aroG_3'</i> (Amplifikat ii) (RT-qPCR)	GGTGCGAGGCTTCTCGAAGT
<i>aroR-aroG_5'</i> (Amplifikat iii) (RT-qPCR)	CGTCAGCTGGCCATTAGTTG
<i>aroR-aroG_3'</i> (Amplifikat iii) (RT-qPCR)	GCTAGGCAGCTCGTGGAAG
<i>cg3138-sense</i> (RT-qPCR)	CCACTGTCTGTGACGAAGAA
<i>cg3138-antisense</i> (RT-qPCR)	CTGTGCAGTATCGGCTACCT

^a Unterstrichene Nukleotide kennzeichnen die T7-RNA-Polymerase-Promotorsequenz. Kursivschrift gibt die *Annealing*-Regionen für GeneSOEing-Primer an, kleine Buchstaben kennzeichnen die Restriktionsschnittstellen

*NB = Northern Blot

2 Verwendete Chemikalien, Materialien, Geräte und Software

2.1 Chemikalien, Enzyme, Kits, Marker

Tabelle III-4 Liste der verwendeten Chemikalien

Substanz	Bezugsquelle
Aceton	Roth®
Acrylamid/N,N-methylbisacrylamid	Roth®
Agar	Invitrogen
Agarose	Peqlab
Ammoniumchlorid	Merck KGaA
Ammoniumperoxodisulfat	Sigma-Aldrich
Biotin	Sigma-Aldrich
Blocking-Reagenz (für Nukleinsäurehybridisierung)	Roche
Borsäure	Merck KGaA
Bromphenolblau	Serva Electrophoresis
Calciumchlorid	Merck KGaA
CASO-Bouillon	Merck KGaA
CDP-Star	Roche Applied Science
Chloroform	Sigma-Aldrich
D(+)-Glukosemonohydrat	Merck KGaA
DEPC-H ₂ O	Roth®

Desoxyribonukleotide	Fisher Scientific
Diethanolamin	Merck KGaA
Dikaliumhydrogenphosphat	Merck KGaA
Dimethylsulfoxid	Sigma-Aldrich
Dinatriumhydrogenphosphat	Merck KGaA
Dithiothreitol	Roth®
dNTPs	Fisher Scientific
<i>Dye-Reagent Concentrate</i>	Bio-Rad Laboratories
EDTA	Amersham
Essigsäure	Roth®
Ethanol	Roth®
Ethidiumbromid	Serva
Fluoreszenzfarbstoff Molecular probes FM 4-64	Life Technologies
FluoroLink Cy3 monofunctional dye 3-pack, Cy5 monofunctional dye 5-pack	Amersham
Formamid	Roth®
Glycerin 87%	Merck KGaA
Glycin	Merck KGaA
Harnstoff	Merck KGaA
Isopropanol	Roth®
Isopropylthiogalactosid	Roth®
Kaliumchlorid	Merck KGaA
Kaliumdihydrogenphosphat	Merck KGaA
LB-Medium	Invitrogen™
Magnesiumchlorid	Merck KGaA
Magnesiumsulfat	Merck KGaA
Maleinsäure	Merck KGaA
Methanol	Roth®
MOPS	Biomol GmbH
Natriumacetat	Merck KGaA
Natriumchlorid	Merck KGaA
Natriumcitrat	Merck KGaA
Natriumlaurylsulfat (SDS)	Serva
Natronlauge	Merck KGaA
N-Lauroylsarcosin	Sigma-Aldrich
Phenol/Chloroform/Isoamylalkohol	Roth®
PIPES	Roth®
Protocatechusäure	Sigma-Aldrich
Salzsäure	Roth®
Stickstoff	Linde AG
Sucrose	Merck KGaA

Tetramethylethylendiamin	Merck KGaA
Thiamin	Sigma-Aldrich
TRIS	MP Biomedicals
TRIS-HCl	Serva
Triton X-100	Serva
TRIzol-Reagenz	Invitrogen
Trypton	Oxoid
Tween® 20 (Polysorbat 20)	Roth®
X-Gal	Roth®
β-Mercaptoethanol	Roth®

Tabelle III-5 Liste der verwendeten Enzyme

Enzym	Hersteller
BIOTAQ™ DNA Polymerase	Bioline
DNase I (RNase-Free DNase Set)	QIAGEN
DNase I recombinant, RNase-free	Roche Applied Science
Lysozyme	Serva
Phusion™ High-Fidelity DNA Polymerase	Finnzymes
Restriktionsendonukleasen	Fisher Scientific
RiboLock™ RNase Inhibitor	Fisher Scientific
RNA 5' Polyphosphatase	EPICENTRE®
RNase A	Roche Applied Science
Shrimp Alkalische Phosphatase	Fisher Scientific
T4-DNA-Ligase	Fisher Scientific
T7-RNA-Polymerase	Roche Applied Science
Terminator™ 5'-Phosphate-Dependent Exonuclease	EPICENTRE®
ThermoScript™ Reverse Transcriptase	Invitrogen

Tabelle III-6 Liste der verwendeten Kits

Produktname	Zweck	Bezugsquelle
Agilent RNA 6000 Pico/Nano Kits	RNA-Qualitätsanalyse (Bioanalyzer)	Agilent Technologies
CloneJET™ PCR Cloning Kit	Klonierung von DNA-Fragmenten	Fisher Scientific
DIG RNA Labeling Mix	Markierung von RNA-Sonden	Roche
GeneJET™ Plasmid Miniprep Kit	Isolierung von Plasmid-DANN	Fisher Scientific
GenElute™ Bacterial Genomic DNA Kit	Isolierung von genomischer DNA	Sigma-Aldrich
High Sensitivity DNA Analysis Kits	DNA-Qualitätsanalyse (Bioanalyzer)	Agilent Technologies
Illumina® TruSeq™ Small RNA Kit	Erstellung von cDNA-Bibliotheken	Illumina
NucleoSpin® Extract II Kit	Aufreinigung von Nukleinsäuren	Macherey-Nagel
RiboZero (Gram-Positive Bacteria)	Abreicherung von rRNAs aus Gesamt-RNA	Epicentre®

RNeasy Mini Kit	Isolierung von Gesamt-RNA	Qiagen
SensiMix™ One-Step	RT-qPCR	Bioline

Tabelle III-7 Liste der verwendeten Marker

Marker	Hersteller
DNA-Marker 100bp-Leiter (P-805)	MBBL
DNA-Marker 100bp-Leiter, bis 5000 Bp (P-810)	MBBL
DNA-Marker 250bp-Leiter (P-705)	MBBL
GeneRuler™ Low Range DNA Ladder	Fisher Scientific
GeneRuler™ Ultra Low Range DNA Ladder	Fisher Scientific
RiboRuler™ Low Range RNA Ladder	Fisher Scientific

2.2 Verbrauchsmaterialien

Tabelle III-8 Verwendete Verbrauchsmaterialien

Name	Hersteller
Elektroporationsküvetten	Peqlab Biotechnologie GmbH
Glaswaren	Schott AG
Küvetten	Sarstedt
Nitrilhandschuhe	VWR International
Parafilm	Pechiney PlasticPackaging
Pipettenspitzen	Starlab GmbH
Plastikbedarf	Greiner, Biozym, Starlab
Reaktionsgefäße (1,5 und 2 ml)	Greiner
Ribotubes	MP Biomedicals
Sterilfilter	Milipore

2.3 Geräte und Apparaturen

Tabelle III-9 Verwendete Geräte und Apparaturen

Name	Hersteller
Agilent 2100 Bioanalyser	Agilent Technologies
Brutschränke	Memmert GmbH & Co. KG
Eismaschine	Scotsman Ice Systems
Eclipse 80i Mikroskop	Nikon Instruments
Elektrophoresekammern für Agarosegele	Eigenbau
Feinwaage	Sartorius AG
Gene Pulser / Elektroporationsgerät	Bio-Rad Laboratories

Illumina GA IIx Maschine	Illumina
Kühlschränke	Privileg / Siemens
Kühltischzentrifuge 5417 R	Eppendorf AG
Kühlzentrifuge 4K15 C	Sigma-Aldrich
LightCycler	Roche
Luftschüttler	Innova
Magnetrührer mit Heizplatte	Heidolph Elektro GmbH&Co. KG
Mikrowelle	cinex electronic GmbH
NanoDrop ND-1000 Photospektrometer	Peqlab Biotechnologie GmbH
PCR-Thermocycler	Eppendorf AG
PerfectBlue Semi-Dry Elektrobloetter	Peqlab Biotechnologie GmbH
pH-Meter	Knick Elektronische Messgeräte GmbH & Co
Photometer	Eppendorf AG
Pipetman	Eppendorf AG
Reinraumwerkbank HERAsafe®	Heraeus Holding GmbH
Reinstwasseranlage GenPure	TKA
Ribolyser	Peqlab Biotechnologie GmbH
Ribolyser FP120-HY-230	Hybaid
Schnellschüttler	New Brunswick Scientific
Spannungsgeber	Bio-Rad Laboratories
Thermomixer comfort	Eppendorf AG
Tischzentrifuge 5415 D	Eppendorf AG
UV-Transilluminator	Alpha Innotech Corporation
Vakuumzentrifuge Savant Speed-Vac SPD 111V	Thermo Fisher Scientific
Vortex-Genie	Scientific Industries, Inc.
Wasserbad	GFL – Gesellschaft für Labortechnik mbH

2.4 Software

Tabelle III-10 Verwendete Software

Name	Zweck	Hersteller
BLAST	Sequenzvergleiche (Protein- o. Nukleotidebene)	(Altschul <i>et al.</i> , 1990)
calculateExpressionValues	TSS-Bestimmung aus RNA-Seq Daten	Christian Rückert, CeBiTec, Bielefeld
Clone Manager	Berechnung von Primersequenzen	Scientific & Educational Software
ClustalW	Erstellen von Sequenz-Alignments	(Larkin <i>et al.</i> , 2007)
CopraRNA	Vorhersage von RNA-RNA Interaktionen	(Wright <i>et al.</i> , 2013)

DAVID-Tool	Funktionelle Genanalyse nach GO-Terminologie	(Dennis <i>et al.</i> , 2003)
EMMA2	Normalisierung und Auswertung von Microarrays	(Dondrup <i>et al.</i> , 2009)
Genesis 1.7.6.	Microarray-Datenanalyse	(Sturn <i>et al.</i> , 2002)
Improbizer	Motivsuche in Sequenzen	Jim Kent, UCSC University California (Ao <i>et al.</i> , 2004)
ImaGenge	Spoterkennung und Quantifizierung	Bio Discovery
IRMA (Java-Applet)	Berechnung intergenischer Oligonukleotidproben	(Schroeder, 2007)
Mauve 2.3.1	Erstellen Multipler Genom-Alignments	(Darling <i>et al.</i> , 2004)
NodeXL-Excel Template	Netzwerk Analyse-Tool	The Social-Media Research Foundation
Office 2010 Paket	Textverarbeitung, Tabellenkalkulation	Microsoft
ORFfinder	Detektion von offenen Leserastern	Uhmin, Osaka University
Origin 8.1	Tabellenkalkulation, Box plot-Diagramme	ADDITIVE
Overlap (Perl-Skript)	Berechnen übereinstimmender Sequenzbereiche	Thomas Kohl, CeBiTec Bielefeld
PseudoViewer3	RNA-Sekundärstrukturvisualisierung	(Byun und Han, 2009)
pknotsRG	Berechnung von RNA-Pseudoknoten	(Reeder <i>et al.</i> , 2007)
RBSfind	Detektion von ribosomalen Bindestellen	(Suzek <i>et al.</i> , 2001)
ReadXplorer	Visualisierung von RNA-Seq- <i>reads</i>	(Hilker <i>et al.</i> , <i>submittiert</i>)
RNAcofold	Interaktionsanalyse von RNA-Molekülen	(Lorenz <i>et al.</i> , 2011)
RNAfold	RNA-Sekundärstrukturanalyse	(Lorenz <i>et al.</i> , 2011)
RNAPredator	Vorhersage von RNA-RNA Interaktionen	(Eggenhofer <i>et al.</i> , 2011)
RNAshapes	RNA-Sekundärstrukturanalyse	(Steffen <i>et al.</i> , 2006)
RNAup	Interaktions-Analyse von RNA-Molekülen	(Lorenz <i>et al.</i> , 2011)
RNAz	Vorhersage regulatorischer RNAs	(Washietl <i>et al.</i> , 2005)
SARUMAN	<i>Read-mapping</i>	(Blom <i>et al.</i> , 2011)
TargetRNA	Vorhersage von RNA-RNA Interaktionen	(Tjaden, 2008)

3 Medien

3.1 Nährmedien

BHIS-Medium (Liebl *et al.*, 1989)

- 37 g/L *Brain Heart* Infusion
- 91 g/L Sorbitol

CASO-Bouillon

- 30,0 g/L CASO-Bouillon
- 17,0 g/L Caseinpepton
- 3,0 g/L Sojapepton
- 5,0 g/L NaCl
- 2,5 g/L K_2HPO_4
- 2,5 g/L D(+)-Glucose-Monohydrat

CGXII-Minimalmedium (Keilhauer *et al.*, 1993)

- 20,00 g/L $(NH_4)_2SO_4$
- 1,00 g/L K_2HPO_4
- 1,00 g/L KH_2PO_4
- 0,25 g/L $MgSO_4 \cdot 7 H_2O$
- 42,00 g/L MOPS

→ in 876 mL H_2O dest. lösen und autoklavieren, anschließend zugeben:

- 100 mL D(+)-Glukose, 50% (w/v)
- 20 mL Harnstoff, 25% (w/v)

unmittelbar vor Gebrauch hinzugeben:

- 1 mL Biotin (0,2 mg/mL)
- 1 mL Thiamin (0,42 mg/mL)
- 1 mL Protocatechusäure (30 mg/mL)
- 1 mL Spurenelemente

LB-Medium (*lysogeny broth*) (Bertani, 2004)

- 20 g/L LB-Medium (Lennox L *Broth Base*)
- 10 g/L Trypton
- 5 g/L Hefeextrakt
- 5 g/L NaCl

SOC-Medium (Hanahan, 1983)

- 20,0 g/L Trypton
- 5,0 g/L Hefeextrakt
- 4,0 g/L D(+)-Glucose-Monohydrat
- 0,6 g/L NaCl
- 0,2 g/L KCl
- 4,1 g/L MgCl₂
- 4,9 g/L MgSO₄

→ Medium wird ohne Mg²⁺ autoklaviert, anschließend Zugabe steriltfiltrierter Mg²⁺-Lösung

3.2 Zusätze zu den NährmedienFür Festmedien

- 15 g/L Agar

CGXII-Spurenelemente (Keilhauer *et al.*, 1993)

- 10,0 g/L FeSO₄·7 H₂O
- 10,0 g/L MnSO₄·7 H₂O
- 1,0 g/L ZnSO₄·7 H₂O
- 0,2 g/L CuSO₄
- 0,02 g/L NiCl₂·6 H₂O

Sucrose (für Festmedien)

- 100 g/L Sucrose

Stressoren für Wachstumsexperimente

- 1,50 M NaCl oder
- 0,33 M H₂O₂ oder
- 2,00 mM Diamid (N,N,N',N'-tetramethylazodicarboxamide)

Antibiotika für die selektive Anzucht

Tabelle III-11 Verwendete Antibiotika

Antibiotikum	Abkürzung	Lösungsmittel	Konzentration [mg/L]
Ampicillin	Amp	H ₂ O dest.	100
Kanamycin	Km	H ₂ O dest.	25 (<i>C. glutamicum</i>), 50 (<i>E. coli</i>)
Nalidixinsäure	Nx	0,1 N NaOH	50

3.3 Verwendete Puffer und Lösungen

3.3.1 Nukleinsäure-Isolierung, -Reinigung und –Amplifikation

dNTP-Mix

- 2,5 mM dATP
- 2,5 mM dCTP
- 2,5 mM dGTP
- 2,5 mM dTTP

RNase-Lösung

- 10 mg RNase A in 1 mL H₂O dest. lösen

Lysozym-Lösung (immer frisch ansetzen)

- 4 mg Lysozym in 200 µl Gram-Positive Lysis Solution (GenElute-Kit) lösen

Proteinase K-Lösung

- 20 mg Proteinase K in 1 mL H₂O dest. lösen und bei -20 °C lagern

Lösung zur Präzipitation kleiner RNAs

- 0,5 M NaAc
- 5,0 % PEG 8000

3.3.2 Elektrophorese

Agarose

- 0,8-3,0 % (w/v) Agarose in TA-Puffer (1-fach) lösen, unter Rühren aufkochen, bei 60 °C lagern

Bromphenolblau-Xylencyanol Ladepuffer (BPB)

- 1 mM EDTA
- 50,0 % (w/v) Glycerin
- 0,1 % (w/v) Bromphenolblau
- 0,1 % (w/v) Xylencyanol

→ mit HCl bis zum Umschlagpunkt (dunkelgrün) titrieren

Ethidiumbromid-Färbelösung

- 10 mg/mL EtBr

TA-Puffer (50 x)

- 2 M TRIS
- 500 mM Na-Acetat
- 50 mM EDTA

→ Einstellung des pH-Wertes auf 7,8 mittels konzentrierter Essigsäure

3.3.3 DNA-Transfertechniken

TG (TRIS-Glycerin)-Puffer (pH 7,5)

- 1 mM TRIS
- 10 % (w/v) Glycerin

TB-Puffer

- 10 mM PIPES
- 15 mM CaCl
- 250 mM KCl

→ in 800 mL H₂O lösen und auf pH=6,7 einstellen (mit KOH oder HCl)

3.3.4 Auftrennung der RNA im Polyacrylamidgel und Semi-Dry Elektrobplot

LiCl-Lösung

- 4 M LiCl in H₂O dest lösen

SDS-Lösung

- 10 % (w/v) SDS

APS-Lösung (10 %)

- 1 g Ammoniumperoxodisulfat in 10 mL H₂O dest lösen und im Kühlschrank lagern

EDTA-Lösung (0,5 M)

- 93,05 g ETDA in ca. 400 mL H₂O dest. aufnehmen,
- pH mittels NaOH auf 8,0 einstellen und auf 500 mL auffüllen

Blocking-Reagenz

- 20 g Blocking-Pulver
 - 200 mL Puffer I, 1-fach
- in H₂O dest. aufnehmen, in Mikrowelle lösen, dann autoklavieren, Lagerung bei 4 °C

SSC-Puffer (20-fach)

- 3 M Natriumcitrat·2 H₂O
- 3 M NaCl

Prähybridisierungslösung für DIG-markierte Sonden

- 187,50 µL H₂O dest.
 - 10,00 µL SDS, 10 %
 - 50,00 µL N-Lauroylsarcosin
 - 1,00 mL Blocking-Reagenz
 - 1,25 mL 20x SSC
 - 2,50 mL Formamid
 -
- Prähybridisierungslösung immer frisch ansetzen

PAA-Mix (200 mL) für 10 % PAA-Gele

- 66,7 mL 30 % AA
 - 20,0 mL 10 x TBE-Puffer
 - 84,0 g Harnstoff
- mit H₂O dest. auf 200 mL auffüllen und den Mix dunkel im Kühlschrank lagern
- zu 50 mL PAA-Mix (2 Gele) vor Gießen 50 µL TEMED (99 %) und 500 µL 10 % APS hinzu

TBE-Puffer (10-fach)

- 108,0 g/L TRIS
 - 55,0 g/L Borsäure in ca. 900 mL H₂O dest aufnehmen
 - 40,0 mL EDTA (0,5 M)
- auf 1 Liter auffüllen

Puffer I (10-fach)

- 1,0 M Maleinsäure
- 1,5 M NaCl

→ in H₂O dest aufnehmen, pH mittels NaOH auf 7,5 einstellen

Waschpuffer

- 3 mL Tween® 20 (Polysorbat 20) mit 1 Liter Puffer I (1-fach) auffüllen

Puffer II (1-fach)

- 25 mL Puffer I (1-fach)
- 25 mL Blocking-Reagenz

→ in 250 mL H₂O dest. aufnehmen; immer frisch ansetzen

Puffer III (1-fach)

- 9,6 mL Diethanolamin mit 1 Liter Puffer I (1-fach) auffüllen

Waschlösung I

- 12,50 mL 20x SSC
- 250,00 µL SDS, 10 %
- 37,25 mL H₂O dest.

Waschlösung II

- 2,50 mL 20x SSC
- 250,00 µL SDS, 10 %
- 47,25 mL H₂O dest.

Waschlösung III

- 1,25 mL 20x SSC
- 250,00 µL SDS, 10 %
- 48,25 mL H₂O dest.

4 Kultivierung, Ernten und Lagerung von Bakterien

4.1 Bestimmung des Bakterientiters

Durch Bestimmung der optischen Dichte der Flüssigkultur kann bei einer Wellenlänge von 600 nm in einem Spektralphotometer das Wachstum von Bakterien verfolgt werden. Als Referenz dient dabei das Nährmedium ohne Bakterien. Eine OD_{600} von 0,1 entspricht etwa $2 \cdot 10^7$ *E. coli*-Zellen bzw. $1 \cdot 10^7$ *C. glutamicum* Zellen. Hierbei ist zu beachten, dass mit dieser Methode der Gesamt-Bakterientiter aus lebenden und bereits abgestorbenen Zellen ermittelt wird.

4.2 Anzucht der Bakterien für die Experimente zur Identifizierung kleiner RNAs

C. glutamicum ATCC 13032 wurde in 25 mL Schüttelkolben-Kulturen in CGXII-Minimalmedium (Keilhauer *et al.*, 1993) angezogen. Die *C. glutamicum*-Zellen wurden in der exponentiellen Wachstumsphase bei einer optischen Dichte OD_{600} von 10 geerntet. Zusätzlich wurden zu diesem Zeitpunkt in fünf *C. glutamicum*-Kulturen die verschiedenen Stressbedingungen Hitzeschock (50 °C), Kälteschock (4 °C), hyperosmotischer Stress (1,5 M NaCl), H_2O_2 -Stress (0,33 M) und Diamid-Stress (2 mM N,N,N',N'-tetramethylazodicarboxamide) für jeweils 15 Minuten (Min) angewendet (Nakunst *et al.*, 2007) und im Anschluss daran geerntet. Zum Ernten wurden jeweils 1,5 mL Kultur schnellstmöglich in ein Eppendorf-Gefäß überführt und für 15 Sekunden (Sek) bei Raumtemperatur (RT) und 13.000 g (Erdbeschleunigung) zentrifugiert. Danach wurde der Überstand verworfen und die pelletierten Zellen sofort in flüssigem Stickstoff tiefgefroren. Die Proben wurden bei -80 °C bis zur weiteren Verwendung aufbewahrt. Von allen Kulturen wurde jeweils ein biologisches Replikat mitgeführt.

4.3 Anzucht der Bakterien zur Attenuator-Validierung (*aroF*)

Für die Kultivierungen wurde die Insertionsmutante *C. glutamicum* ATCC 13032 *pheA::IS6100* (Mormann *et al.*, 2006) verwendet, die auxotroph für die aromatische Aminosäure Phenylalanin ist. Zur Inokulation der Vorkulturen wurde die Cryo-Kultur auf Agar-Platten mit komplexem CASO-Medium und 50 mg/L Nalidixinsäure angezogen. Die Vorkulturen wurden in 10 mL desselben Mediums, allerdings ohne Agar als Flüssigkultur in 100 mL-Schüttelkolben für einen Tag bei 30 °C und 300 rpm in einem Schüttelinkubator inkubiert. Die Zellen von 1 mL Vorkultur wurden dann zentrifugiert und mit 1 mL gereinigtem H_2O gewaschen. Nach einem zweiten Zentrifugationsschritt und Verwerfen des Überstandes wurden die Zellen resuspendiert in 1 mL sterilem H_2O und für das Animpfen der Hauptkulturen mit einer $OD_{600} = 0,1$ verwendet. Die Hauptkulturen wurden in 20 mL CGXII-Minimalmedium (Keilhauer *et al.*, 1993) kultiviert und enthielten zusätzlich 0,1 mM oder 1 mM Alanyl-Phenylalanin Dipeptid (Ala-Phe) in 250 mL-Schüttelkolben. Für jede der zwei Dipeptidkonzentrationen

wurden drei biologische Replikate kultiviert. Nach 15 Std. Kultivierung wurden die Zellen geerntet zur Isolierung der Gesamt-RNA. Zum Ernten wurden jeweils 1,5 mL Kultur schnellstmöglich in ein Eppendorf-Gefäß überführt und für 15 Sek bei RT und bei 13.000 g zentrifugiert. Danach wurde der Überstand verworfen und die pelletierten Zellen sofort in flüssigem Stickstoff tiefgefroren. Diese Proben wurden bei -80 °C bis zur weiteren Verwendung aufbewahrt.

4.4 Stochertest

Etwas Zellmasse einer zu prüfenden Einzelkolonie wird für den Stochertest mit einem sterilen Zahnstocher aufgenommen und durch strichförmiges Ausstreichen auf verschiedene Festmedien-Platten überimpft. Nach Inkubation der Platten im Brutschrank über Nacht können diese ausgewertet werden.

5 Polymerase-Kettenreaktion

Die Polymerase-Kettenreaktion (*polymerase-chain-reaction*; PCR) bezeichnet eine Methode zur enzymatischen *in vitro* Amplifikation definierter DNA-Abschnitte, deren flankierende Sequenzen zur Konstruktion geeigneter Primer bekannt sein müssen (Saiki *et al.*, 1988). Ein PCR-Prozess besteht aus 20-50 Zyklen, die jeweils in drei Schritte aufgeteilt sind. Der erste Schritt eines PCR-Zyklus beginnt mit der Denaturierung der Matrizen-DNA (*template*) bei 95-98 °C, damit es im zweiten Schritt zur Bindung (*annealing*) der Primer an komplementäre einzelsträngige DNA-Bereiche über Wasserstoffbrückenbindungen kommen kann. Der Aufbau und die Länge von Primern bestimmen dabei die *annealing*-Temperatur, die etwa 5 °C unter der Primer-Schmelztemperatur T [°C] liegt und von Länge und G+C-Gehalt beeinflusst wird. Die 3'-OH-Enden der Primer dienen im dritten und letzten Schritt eines PCR-Zyklus als Ausgangspunkt für die Komplementierung der Einzelstränge zu Doppelsträngen über eine thermostabile DNA-Polymerase (Polymerisation). Voraussetzung für eine erfolgreiche Polymerisation sind neben einem geeigneten pH-Wert und Vorhandensein von Mg^{2+} -Ionen, eine optimale Polymerisations-Temperatur (72 °C), sowie das Vorhandensein von Desoxynukleosid-Triphosphaten (dNTPs).

5.1 Primer-Design

Die Wahl geeigneter Primer spielt bei der PCR eine wesentliche Rolle, weshalb einige Regeln beachtet werden sollten. Für eine ausreichende Spezifität der PCR sollte die Primer-Mindestlänge 20 Nukleotide betragen. Der G+C-Gehalt der Primer sollte bei 50 - 60 % liegen und die *annealing*-Temperaturen eines Primerpaares sollten möglichst nicht mehr als 2 °C voneinander abweichen. Angehängte Erweiterungen wie eine T7-Promotorsequenz, Restriktionsschnittstellen, oder chemische Modifikationen werden an das

5'-Ende eines Primers angehängt. Neben diesen Kriterien sind noch einige weitere Dinge zu beachten, z.B. die Vermeidung der Bildung von Haarnadelstrukturen, Dimerisierung etc., daher empfiehlt sich der Einsatz spezieller Computerprogramme wie beispielsweise der im Rahmen dieser Arbeit verwendete Clone Manager (Sci-Ed Software).

5.2 PCR-Reaktionsansätze und -Programme

In dieser Arbeit wurden sowohl eine Taq- als auch eine Phusion-DNA-Polymerase eingesetzt, die sich durch unterschiedliche Eigenschaften auszeichnen. Die Taq-Polymerase ist die thermostabile DNA-Polymerase des Bakteriums *Thermus aquaticus* (Taq). Phusion-Polymerasen kombinieren ein neues *Pyrococcus*-ähnliches Enzym mit einer prozessivitätssteigernden Domäne. Die Phusion-Polymerase verfügt darüberhinaus über eine 3'-5'-Exonuklease-Aktivität (*proof-reading*), die zu einer geringeren Fehlerrate bei der Amplifikation führt. Bei der Phusion handelt es sich außerdem um eine *hot start*-Polymerase, bei der die Polymerase bis zum ersten Denaturierungsschritt durch gebundene Antikörper blockiert wird. Unspezifische Amplifikationen, die durch falsches Binden von Primern während des Temperaturanstieges im ersten Zyklus entstehen können, werden somit verhindert. Ebenso werden Primerdegradierung und die Exonuklease-Aktivität der Phusion-Polymerase bei Raumtemperatur unterbunden. Im Gegensatz zur Phusion-Polymerase fügt die Taq-Polymerase unabhängig vom Aufbau der Matrizen-DNA ein Adenosin-Triphosphat als letztes Nukleotid an, sodass Amplifikate mit 3'-Überhängen (*sticky ends*) entstehen. Im Folgenden sind Pipettierschemata für beide Polymerasen dargestellt und die zugehörigen Standard-PCR-Programme.

PCR-Reaktionsansatz und -Programm für die taq-Polymerase

Tabelle III-12 Mix für einen Taq PCR-Ansatz (50µL)

5,0 µL 10× NH ₄ Reaktions-Puffer
2,0 µL dNTPs (2,5 mM)
0,5 µL Primer 1 (10 mM)
0,5 µL Primer 2 (10 mM)
1,5 µL MgCl ₂ (50mM)
1,0 µL <i>Template</i> -DNA
1,0 µL Taq-Polymerase

→ mit H₂O dest. auf 50 µL auffüllen

Tabelle III-13 Programm für die Taq-PCR

94 °C	2 Min.		Initiale Denaturierung
94 °C	30 Sek.	} 35 Zyklen	Denaturierung
~60 °C*	30 Sek.		<i>annealing</i>
72 °C**	30-90 Sek.		Elongation
72 °C	7 Min.		Finale Elongation
4 °C	∞		Ende

* abhängig von den verwendeten Primern

** abhängig von der Größe des zu amplifizierenden DNA-Fragments und der Reaktionsgeschwindigkeit der verwendeten Polymerase

PCR-Reaktionsansatz und -programm für die Phusion-Polymerase

Tabelle III-14 Mix für einen Phusion PCR-Ansatz (50µL)

5,0 µL 10× Phusion GC-Puffer
2,0 µL dNTPs (2,5mM)
0,5 µL Primer 1 (10 mM)
0,5 µL Primer 2 (10 mM)
1,5 µl DMSO
1,0 µL Template-DNA
1,0 µL Phusion-Polymerase

→ mit H₂O dest. auf 50 µL auffüllen

Tabelle III-15 Programm für eine Phusion-PCR

98 °C	1 Min.		Initiale Denaturierung
98 °C	5-10 Sek.	} 35 Zyklen	Denaturierung
~60 °C	40 Sek.*		<i>annealing</i>
72 °C	30 Sek.**		Elongation
72 °C	30-90 Sek.		Finale Elongation
4 °C	∞		Ende

* abhängig von den verwendeten Primern

** abhängig von der Größe des zu amplifizierenden DNA-Fragments und der Reaktionsgeschwindigkeit der verwendeten Polymerase

6 Klonierungsexperimente

6.1 DNA-Restriktionsspaltung

Die DNA-Restriktionsspaltung stellt eine Methode zur Analyse von DNA oder zur Vorbereitung einer anschließende Ligation dar, bei der das Pentose-Phosphat-Rückgrat der DNA an charakteristischen palindromischen Sequenzen mit Restriktionsendonukleasen vom Typ II hydrolysiert wird. Je nach verwendetem Enzym entstehen dabei entweder versetzt einzelsträngige Enden (*sticky ends*) oder glatte Enden (*blunt ends*). Es ist zu beachten, dass einige Restriktionsendonukleasen „Methylierungs-anfällig“ sind. Das bedeutet, dass diese Enzyme ihre Zielsequenzen nicht oder nur mit einer geringeren Effizienz erkennen, wenn eine Base darin im Rahmen einer zelltypischen Modifikation methyliert wurde. Vorbereitend für den Restriktionsansatz werden die Konzentrationen von Fragment und Vektor per Spektralphotometer bestimmt. Die Klonierung sollte mit 250 ng Vektor und 5-fach molarem Überschuss Fragment erfolgen. Die Restriktionsspaltung erfolgt analog zu den vom Hersteller empfohlenen Bedingungen (in der Regel für mindestens 1 Std. bei 37 °C im Brutschrank) unter Verwendung des mitgelieferten Puffers. Die Restriktionsenzyme sollten nach der Reaktion inaktiviert und anschließend mittels PCR-Aufreinigung (Macherey Nagel Kit) entfernt werden.

Der Spaltungsansatz lautet (20 µL):

- 2,0 µL Spaltungs-Puffer
- 0,5 µL je Restriktionsenzym
- 3-15 µL Vektor bzw. Fragment verwenden und mit H₂O dest. auf 20 µL auffüllen.

6.2 DNA-Ligation

Bei einer DNA-Ligation werden die endständigen 5'-Phosphatgruppen und 3'-Hydroxylgruppen des Pentose-Phosphat-Rückgrats der DNA mithilfe der ATP-abhängigen T4-DNA-Ligase unter Ausbildung von Phosphodiesterbindungen miteinander verknüpft. Auf diese Weise können durch Restriktionsenzyme modifizierte DNA-Fragmente miteinander kondensieren. Die Inkubation des Ansatzes erfolgt 1 Std. bei RT. Nachfolgend ist ein Ligationsansatz zur Integration eines DNA-Fragments in ein Plasmid aufgeführt (50 µL-Ansatz):

- 20,0 µL Fragment
- 20,0 µL Vektor
- 5,0 µL Ligase-Puffer
- 0,5 µL Ligase und mit H₂O dest. auf 50 µL auffüllen.

7 DNA-Transfertechniken

7.1 Herstellung und Elektroporation elektrokompetenter *C. glutamicum*-Zellen

Die Elektroporation ist eine verbreitete Methode, um freie DNA in Bakterienzellen zu transformieren. Dabei erfolgt die Aufnahme der DNA in vorbehandelte (kompetente) Zellen über einen kurzen elektrischen Puls von 10 kV/cm, wobei kurzzeitig Poren in der Cytoplasmamembran erzeugt werden und so Makromoleküle wie Fremd-DNA in die Zelle eindringen können. Hierbei ist es von Bedeutung, dass sowohl die Bakterien- als auch die DNA-Suspensionen deionisiert sind. Eine zu hohe Ionenkonzentration würde im Reaktionsansatz zu einem verstärkten Stromfluss und infolgedessen zu einer deutlichen Abnahme der Elektroporations-Effizienz führen. Anschließend muss die Membran wieder hergestellt werden, um das Überleben der Zellen zu sichern. Dabei müssen für jeden Zelltyp die richtigen Bedingungen wie Temperatur und Zusammensetzung gewählt werden.

Eine *C. glutamicum*-Kultur wird in CASO mit 10 % (w/v) Sorbitol bis zur früh-logarithmischen Phase angezogen und auf Eis schnell abgekühlt. Für das Waschen der Kultur wird zunächst TG-Puffer verwendet, dann 10 % Glycerin. Nach mehrfacher Wiederholung zur Erhöhung der Stringenz, werden die kompetenten Zellen in 150 µL-Aliquots aufgeteilt und können für mehrere Monate bei -80 °C gelagert werden (Tauch *et al.*, 2002). Für die Elektroporation wird das gewünschte Plasmid zu der aufgetauten Zellsuspension gegeben und folgende Pulsbedingungen eingestellt: 25 µF Kapazität; 200 Ω Parallelwiderstand; 2,5 kV Spannung (bei vorgekühlten 0,2 cm Küvetten). Im nächsten Schritt werden die Zellen sofort in die auf 46 °C vorgewärmten 4 mL BHIS-Medium überführt und dann im Wasserbad für 6 Min. bei 46 °C zur Inaktivierung des Restriktionsmodifikationssystems inkubiert. Anschließend erfolgt eine Regeneration der Zellen für 60 Min. im 30 °C-Roller. Nach Ablauf dieser Zeit werden die Zellen auf geeignetes Selektionsmedium ausplattiert (Tauch *et al.*, 2002).

7.2 Hitzeschock-Transformation von ultrakompetenten *E. coli*-Zellen

Eine weitere Methode zur DNA-Transformation ist der Hitzeschock (Inoue *et al.*, 1990). Die Bakterienzellen werden hierbei mit Calciumchlorid behandelt, was bewirkt, dass es zwischen der negativ geladenen DNA und der negativ geladenen Zellmembran weniger abstoßende Kräfte gibt (Palchevskiy und Finkel, 2006). Anschließend wird ein kurzer Hitzeschock durchgeführt und es entstehen Poren in der Membran, so dass DNA in die Zellen hinein gelangen kann. Für die Herstellung der ultrakompetenten *E. coli*-Zellen muss der gewünschte Stamm bei 19 °C bis zur früh-logarithmischen Phase (OD = 0,5) kultiviert werden. Nach Abkühlung der Zellen erfolgt einmaliges Waschen mit TB-Puffer. Die Zellen werden danach in TB-Puffer mit DMSO (Endkonzentration 7 %) resuspendiert, aliquotiert und bei -80 °C gelagert. Im Rahmen der Transformation mittels Hitzeschock werden die kompetenten Zellen auf Eis

aufgetaut und 1-25 µL Plasmid-DNA wird hinzugegeben. Nach 30 Min. Inkubation der Zellen auf Eis, erfolgt der Hitzeschock für ca. 45 Sek bei 42,5 °C. Die Zellen werden danach kurz auf Eis gestellt und anschließend werden zur Regeneration 0,5 mL SOC-Medium zugegeben und die Zellen 1 Std. bei 37 °C gelagert. Die Zellen werden auf einem geeigneten Selektionsmedium ausplattiert.

8 Allgemeine DNA-Arbeiten

8.1 Isolierung von Plasmid-DNA

Dieses Verfahren bietet die Möglichkeit der Isolierung besonders reiner Plasmid-DNA und basiert auf der alkalischen Lyse (Birnboim und Doly, 1979). Im Anschluss an die Zellyse erfolgt eine Neutralisierung, um passende Bindungen für die Bindung der Plasmid-DNA eine Quarzmembran-Säule zu schaffen. Vor der Beladung der Säule wird das neutralisierte Lysat zentrifugiert, um Zelltrümmer zu pelletieren. Der Überstand wird auf die Säule geladen und zur Entfernung von Proteinen, Polysacchariden, Metaboliten und kurzkettigen Oligonukleotiden gewaschen. Anschließend wird die DNA eluiert. Die genaue Durchführung der einzelnen Schritte ist der GeneJET™ Plasmid Miniprep Kit (Thermo Scientific, Waltham, U.S.A.) beiliegenden Beschreibung zu entnehmen.

8.2 Aufreinigung von PCR-Produkten

Nach einer durchgeführten PCR sollten überschüssige (Oligo-) Nukleotide, Primer und Salze im Amplifikat entfernt werden, da diese folgende Reaktionen behindern könnten. In dieser Arbeit wurde mit dem kommerziell erhältlichen NucleoSpin Extract II Kit (Macherey Nagel, Düren, Deutschland) gearbeitet. Dabei werden größere DNA-Moleküle (> 80 Nukleotide) an eine Säulenmatrix gebunden und mit einem Ethanol-haltigen Puffer gewaschen. Anschließend wird die DNA mit einem schwach basischen Puffer (pH 8) eluiert. Es wurde nach Herstellerangaben vorgegangen.

8.3 Agarose-Gelelektrophorese

Die Agarose-Gelelektrophorese ist ein Verfahren zur Auftrennung von DNA-Fragmenten. Die Trennung hierbei erfolgt im elektrischen Feld, wobei sich die DNA-Moleküle aufgrund ihrer negativen Nettoladung zur positiven Anode hin bewegen. Die Wanderungsgeschwindigkeit ist dabei abhängig von verschiedenen Faktoren, wie Molekulargewicht, Konzentration des Agarose-Gels, DNA-Konformation und der Stärke des angelegten elektrischen Feldes. Beim Vergleich mit einem Längenstandard von DNA-Fragmenten definierter Größe kann die Größe der unbekanntenen DNA-Fragmente ermittelt werden.

Je nach Experiment werden 0,8-3 %ige Agarose mit TA-Puffer angesetzt und aufgeköcht. Dann wird das flüssige Gel in eine Gelkammer gegossen und nach Erkalten in der Kammer mit TA-Puffer überschichtet. Vor dem Auftragen der DNA-Proben werden diese mit einem Glycerin-haltigen Ladepuffer (BPB) versetzt, um die Dichte der Proben zu erhöhen und damit das Einsinken der Proben in die Gel-Taschen zu ermöglichen. Die BPB-Banden dienen gleichzeitig als Marker für die zurückgelegte Laufstrecke während der Elektrophorese. Zur späteren Identifizierung der DNA-Molekülgrößen können verschiedene Marker dienen, die mit auf das Gel aufgetragen werden. Die Gelelektrophorese wird bei 80-100 V für ~30 Min. durchgeführt. Im Anschluss wird das Gel aus der Gelkammer genommen, wenige Minuten in EtBr-Lösung (1 µg/mL EtBr) inkubiert und danach kurz in einem H₂O-Bad geschwenkt, bevor es unter UV-Licht bei 320 nm betrachtet.

8.4 Isolierung von DNA-Fragmenten aus einem Agarosegel

Bei der Isolierung von DNA-Fragmenten aus Agarosegelen wird zum einen die Extraktion eines DNA-Fragments definierter Größe ermöglicht, aber gleichzeitig wird auch eine DNA-Reinigung durchgeführt. Man trägt dazu die DNA-haltige Probe in mehreren Spuren auf ein Agarosegel auf und trennt sie mittels Elektrophorese auf. Damit Mutationen durch EtBr oder UV-Licht vermieden werden, werden nur die außenliegenden Spuren des Gels herausgeschnitten und in EtBr-Lösung gefärbt. Die entsprechende Bande wird unter UV-Licht mit dem Skalpell möglichst genau ausgeschnitten (unbedingt UV-Geschichtsschutz verwenden!). Das mittlere Gelstück wird schließlich wieder zu den Randstücken gelegt, wobei die ausgeschnittenen Bereiche jetzt als Orientierung dienen um das gewünschte DNA-Fragment ohne UV-Detektion auszuschneiden. Die Aufreinigung aus dem Gel erfolgt mittels NucleoSpin Extract II Kit (Macherey Nagel, Düren, Deutschland) nach Herstellerangaben.

9 GeneSOEing

Bei der Methode *Gene Splicing by Overlap Extension (GeneSOEing)* werden *in vitro* DNA-Fragmente mittels PCR neu kombiniert. Eine mögliche Anwendung ist das Erzeugen von Deletions-Konstrukten, mithilfe derer definierte Mutanten generiert werden können. Mittels vier verschiedenen Primern d1-d4 werden zuerst zwei PCR-Produkte mit einer Größe von je ca. 500 bp erzeugt, die den zu deletierenden Bereich flankieren (Abbildung III-1, nächste Seite). Bei der Erstellung der Primer muss darauf geachtet werden, dass der Primer d2 am 5'-Ende einen Überhang trägt, welcher komplementär zur Primersequenz von d3 ist. In einer PCR-Reaktion werden die beiden Produkte dann im gleichen Verhältnis zur Erstellung des Fusionsprodukts eingesetzt. Mithilfe des Überhangs am zweiten Produkt und mithilfe der Primer d1 und d4 können die beiden PCR-Produkte fusioniert werden. Nach geeigneten Restriktionsreaktionen kann das Fusionsprodukt anschließend mittels Ligation in den pK18*mobsacB*-Vektor eingebracht werden.

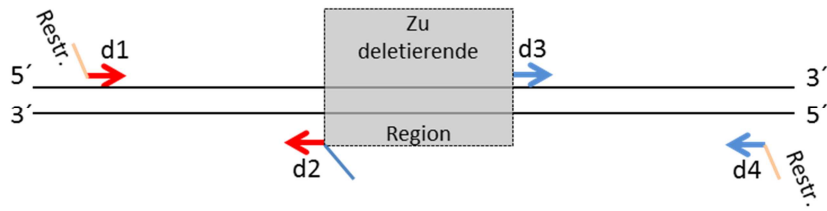


Abbildung III-1 Schematische Darstellung zur Lage der Primer für GeneSOEing.

An den Primern d1 und d4 (rote und blaue Pfeile) sind Schnittstellen für Restriktionsenzyme (Restr.) angehängt, um die Klonierung des Produkts zu ermöglichen (hellroter Bereich symbolisiert hier die Erkennungssequenzen). Der Primer d2 trägt einen komplementären Überhang (blau) zum Primer d3. Die zu deletierende Region ist durch ein gestricheltes Rechteck gekennzeichnet.

Um eine Deletion im Genom von *C. glutamicum* zu etablieren, wird das *GeneSOEing* Amplifikationsprodukt über das *pK18mobsacB*-System gegen den entsprechenden Abschnitt im Chromosom ausgetauscht. Hinsichtlich der Integration des Vektors können die Zellen aufgrund der auf dem Vektor kodierten Kanamycin-Resistenz selektiert werden. Außerdem führt der integrierte *pK18mobsacB* zu einer Sucrose-Sensitivität, welche auf die Expression des *sacB*-Gens (Levansucrase, ursprünglich aus *Bacillus subtilis*) zurückzuführen ist. Die Levansucrase wandelt Sucrose in das für *C. glutamicum* toxische Levan. Können die Zellen auf Sucrose wachsen, hat eine zweite homologe Rekombination stattgefunden, wodurch der Vektor verloren wurde (Schäfer *et al.*, 1994). Bei der Exzision des Vektors kommt es entweder zur homologen Rekombination über dieselbe Flanke wie bei der Integration, was bedeutet dass der Genotyp revertiert. Oder die Rekombination erfolgt über die zweite Flanke. In diesem Fall verbleibt das Deletions-Konstrukt im Chromosom, während das native Gen mit dem Vektor verloren geht. Mögliche Deletionsmutanten können in einer PCR-Reaktion auf den gewünschten Verlust hin untersucht werden. Mittels der außenliegenden Primer wird der DNA-Abschnitt, in dem die Deletion liegt, amplifiziert. Das entstandene Amplifikat ist im Fall der Mutante um die Größe der Deletion verkürzt. Der Wildtyp dient bei der Auftrennung über Agarosegel-Elektrophorese als Kontrolle. Für den Nachweis von eingebrachten Mutationen muss der amplifizierte Bereich anschließend sequenziert werden.

10 Allgemeine RNA-Arbeiten

Beim Arbeiten mit Ribonukleinsäuren (RNA) müssen die ubiquitär vorhandenen Ribonukleasen (RNasen) beachtet werden. Im Gegensatz zu DNasen sind RNasen temperaturstabil und daher auch noch nach Autoklavierung aktiv. RNasen zeigen allerdings eine Sensitivität gegenüber Diethylpyrocarbonat (DEPC). Daher sollte mit DEPC-behandeltem Wasser gearbeitet und es sollten Handschuhe getragen werden. Darüberhinaus sollten bei allen Pipettierschritten Filterspitzen und RNase-freie 1,5 mL Reaktionsgefäße verwendet werden. Wenn möglich sollte auf Eis gearbeitet werden. Außerdem wird empfohlen die Arbeitsflächen vor Beginn aller Arbeitsschritte mit Ambion® RNase-Zap® (Life Technologies GmbH, Darmstadt, Deutschland) zu reinigen. Optional kann in Lösung mit einem RNase-Inhibitor gearbeitet werden. RNase-Inhibitoren sind große (~49 kDa) Leucin-reiche Proteine, die mit RNasen stabile Komplexe formen und sie dadurch inaktivieren (Kobe und Deisenhofer, 1993). Der in dieser Arbeit verwendete RiboLock™ RNase Inhibitor (Fisher Scientific, Schwerte, Deutschland) bindet an die RNasen A, B, and C, allerdings nicht an RNasen I und H.

10.1 Isolierung der Gesamt-RNA

Für die Zweifarben-cDNA-Microarrayanalysen wurde zur Gesamt-RNA-Isolierung das RNeasy Mini Kit (Qiagen, Hilden, Deutschland) verwendet. Da hierbei besonders die Fraktion mit RNA-Spezies <200 nt abgereichert wird, wurde für alle Methoden zur Identifizierung kleiner RNAs eine andere Methode angewendet, die bekannt ist als Guanidiniumthiocyanat-Phenol-Chloroform-Extraktion (Chomczynski und Sacchi, 1987). Sie wurde in dieser Arbeit zum ersten Mal für Gesamt-RNA Isolierungen in *C. glutamicum* etabliert und wird daher ausführlich mit allen Arbeitsschritten beschrieben. Die bei -80 °C gelagerten Zellen werden auf Eis aufgetaut und mit jeweils 1 mL Trizol® (Invitrogen, Darmstadt, Deutschland) vermischt. Trizol® enthält das Thiocyanat-Salz des Guanidins. Dieses Chaotrop lysiert die Zellen und denaturiert gleichzeitig RNasen und andere Enzyme. Zusätzlich enthält das Reagenz Phenol, in dem sich die RNA löst. Der Zellaufschluss wird durch mechanischen Aufbruch unterstützt und erfolgt in sandgefüllten Ribotubes im Precellys Ribolyser. Durch Zugabe von Chloroform und anschließende Zentrifugation bilden sich drei Phasen aus: Ein wässriger Überstand, der die RNA enthält, und eine organische Phase, die Proteine und Lipide enthält. In der Interphase befindet sich die DNA. Die RNA kann aus dem Überstand mittels Isopropanol oder in Anwesenheit einwertiger Na-Ionen präzipitiert und anschließend in Wasser gelöst werden. Falls die erwartete RNA-Konzentration sehr niedrig ist, sollte Zugabe von Glycogen als Ko-Präzipitant (*carrier*) erfolgen. Es wird wie nachfolgend beschrieben vorgegangen.

- Zell-Pellet mit 1 mL Trizol® mischen und 20 Sekunden (Sek) bei Speed 6.5 bei RT ribolysieren
- Inkubation 1 Min. auf Eis inkubieren
- 20 Sek bei Speed 6.5 bei RT zentrifugieren

- 1 Min. auf Eis inkubieren
- 3 Min. bei 4 °C bei 18.000 g zentrifugieren
- jeweils 200 µL Chloroform hinzufügen und 30 Sek vortexen
- 5 Min. bei RT inkubieren
- 10 Min. bei RT bei 13.000 g zentrifugieren
- Je Ansatz 450 µL Isopropanol vorlegen und Überstand aus Zentrifugationsschritt hinzufügen
- 10 Min. auf Eis inkubieren
- 15 Min. bei 4 °C bei 16.000 g zentrifugieren und Überstand verwerfen
- RNA-Präzipitat in jeweils 500 µL 70 % Ethanol waschen und kurz vortexen
- 5 Min. bei 4 °C bei 7600 g zentrifugieren
- Überstand verwerfen und die Reaktionsgefäße im 37 °C Schrank lufttrocknen

Die gefällte RNA bildet ein Pellet, das nach dem Trocknen durchsichtig wird und in RNase-freiem DEPC-H₂O gelöst werden kann.

10.2 Präzipitation kleiner RNAs

Durchführung zur Fällung kleiner RNAs < 250 nt

- Gesamt-RNA mit 2.5 M NaAcetat und 25 % PEG 8000 mischen, so dass eine Endkonzentration von 0.5 M an NaAc und 5 % an PEG 8000 entsteht
- 30 Min. auf Eis inkubieren
- bei 4 °C und max. Geschwindigkeit zentrifugieren → kleine RNA-Fraktion im Überstand
- Pellet verwerfen (enthält die große RNA-Fraktion)

10.3 DNase-Verdau

Die RNA enthält nach der Isolierung meist noch DNA-Kontaminationen, die durch einen DNase-Verdau eliminiert werden müssen. Die Proben werden hierzu mit RNase-free DNase I (Roche, Mannheim, Deutschland) inkubiert.

- je 124 µL DEPC-H₂O hinzufügen und resuspendieren
- je 15 µL Puffer, 10,5 µL RNase-free DNase I und 0,5µL RNase Inhibitor hinzufügen
- 40 Min. bei 30 °C inkubieren

10.4 Aufreinigung durch P:C:I-Extraktion

Um die RNA von vorhandenen Enzyme wie DNasen aufzureinigen, wird eine Lösung aus Phenol, Chloroform und Isoamylalkohol im Verhältnis 25:24:1 (Roth GmbH, Karlsruhe, Deutschland) verwendet. Anschließend wird die RNA in Gegenwart monovalenter Kationen (Na⁺) und Alkohol präzipitiert.

- je DNase-Verdau Ansatz 150 µL P:C:I-Lösung hinzufügen und 30 Sek vortexen (Verhältnis 1:3)
- 15 Min. bei RT und bei 20.000 g zentrifugieren

- je 450 μL 0,3 M Natriumacetat (in Ethanol) werden in 1,5 mL Eppendorf-Reaktionsgefäßen vorgelegt und jeweils der Überstand des Zentrifugationsschrittes hinzugefügt
- 30 Min. bei $-20\text{ }^{\circ}\text{C}$ inkubieren
- 20 Min. bei $4\text{ }^{\circ}\text{C}$ bei 17.600 g zentrifugieren und den Überstand verwerfen
- je 500 μL 70 % eiskalten Ethanol hinzufügen und kurz vortexen
- 5 Min. bei $4\text{ }^{\circ}\text{C}$ bei 7600 g zentrifugieren
- Überstand verwerfen und die Reaktionsgefäße im $37\text{ }^{\circ}\text{C}$ Schrank lufttrocknen
- 50 μL bis 100 μL DEPC- H_2O hinzufügen und bei $-80\text{ }^{\circ}\text{C}$ zur weiteren Verwendung lagern

Die Abschnitte DNase-Verdau und P:C:I-Extraktion werden wiederholt bis die RNA DNA-frei ist. Um dies zu verifizieren, wird mit den Proben eine PCR mit Taq-Polymerase durchgeführt, wobei DNA-Kontaminationen über Amplifizierung sichtbar gemacht werden können. Neben den zu testenden Proben werden eine Positiv-Kontrolle mit *C. glutamicum*-DNA und dest. H_2O als Negativ-Kontrolle verwendet. In allen Ansätzen werden zwei genomspezifische Primerpaare verschiedener Größe benutzt. Anschließend werden die PCR-Proben auf ein 1 %iges Agarosegel aufgetrennt und ausgewertet.

10.5 Aufreinigung über LiCl-Fällung

Die LiCl-Fällung von RNA ist eine gebräuchliche Methode, um inhibitorische Moleküle aus präparierten RNA-Proben zu entfernen (Cathala *et al.*, 1983). Über diese Methode wird nur RNA effizient präzipitiert, da diese stärker hydratisiert ist als DNA. Die LiCl-Methode stellt eine besonders schnelle Methode dar, um RNA aus *in vitro* Transkriptionen zu reinigen.

- 2,5 μL 4 M LiCl zu 20 μL *in vitro*-Transkriptionsansatz hinzufügen
- 75 μL 99 % eiskaltes Ethanol hinzufügen
- 30 Min. bei $-80\text{ }^{\circ}\text{C}$ inkubieren
- 15 Min. bei $4\text{ }^{\circ}\text{C}$ für 13.000 g zentrifugieren
- Überstand verwerfen
- 500 μL 70 %iges Ethanol hinzufügen
- 5 Min. bei $4\text{ }^{\circ}\text{C}$ bei 7600 g zentrifugieren
- Überstand verwerfen
- Pellet für 20 Min. in Vakuumzentrifuge trocknen
- 50 μL DEPC- H_2O hinzufügen

10.6 RNA-Quantifizierung

Für die Quantifizierung der Gesamt-RNA wird ein ND-1000 NanoDrop-Spektrometer (Peqlab, Erlangen, Deutschland) verwendet. Die Oberflächenspannung der Flüssigkeit wird genutzt, um die aufgetragene Probe in Form einer Säule zu stabilisieren. Zwischen den Enden von zwei optischen Fasern wird die spektrometrische Messung durchgeführt. Neben dem Konzentrationswert werden das 260/230- und 260/280-Absorptionsverhältnis angezeigt. Eine Gesamt-RNA Probe von hoher Qualität weist ein

260/230-Verhältnis von $>1,8$ auf und ist damit relativ frei von organischen Komponenten. Zusätzlich sollte ein 260/280-Verhältnis von circa 2,0 gemessen werden, denn ab diesem Wert gilt die Nukleinsäureprobe als frei von Proteinen.

10.7 RNA-Qualitätskontrolle

Für die Qualitätskontrolle der Gesamt-RNA sowie der kleinen RNA-Fractionen wurde ein Agilent Bioanalyzer 2001 (Agilent Technologies, Waldbronn, Deutschland) verwendet. Das verwendete RNA 6000 Nano LabChip® Kit ist für RNA-Konzentrationsgehalte von 25-500 ng/ μ L geeignet. Es sind jedoch auch Kits für andere RNA-Konzentrationsgehalte und auch Kits für die Überprüfung von DNA-Proben erhältlich. Die Nukleinsäurefragmente werden während der Messung in gelgefüllten Kapillaren elektrophoretisch aufgetrennt. Neben einer Konzentrationsmessung, können die Integrität und die Größenverteilung der Gesamt-RNA Probe überprüft werden. Im Allgemeinen gilt Gesamt-RNA mit einem 23S:16S rRNA-Verhältnis größer als 1:1 und einer niedrigen Basislinie zwischen den rRNA-Fragmenten stringent genug für Anwendungen wie Northern Blot, cDNA-Synthese, Microarrays und RT-qPCR. Für die Durchführung wurde nach Herstellerangaben vorgegangen.

11 Spezifische Einzelgen-Transkriptanalysen

Im Fall von Einzelgen-Transkriptanalysen können qualitative und quantitative Aussagen über die Aktivität dieses Gens getroffen werden, wenn die Sequenzinformation des zu untersuchenden Transkripts bekannt ist. Zwei Analysemethoden dieser Art, der Northern Blot und die RT-qPCR, wurden in dieser Arbeit angewendet und werden in den nächsten Kapiteln beschrieben.

11.1 RT-qPCR

Die Abkürzung RT-qPCR (*real-time reverse transcription PCR*) bezeichnet die Kombination einer reversen Transkription (RT) und einer anschließenden quantitativen Polymerase-Kettenreaktion (qPCR, auch *real-time PCR* genannt). Hierbei werden dem Reaktionsgemisch Fluoreszenzfarbstoffe zugesetzt, die es ermöglichen, die Konzentration des PCR-Produktes in Echtzeit und direkt zu verfolgen. Die Zunahme der Fluoreszenz wird jeweils am Ende der Elongationsphase eines PCR-Zyklus gemessen (*real-time*), so dass die ansteigende Menge an PCR-Produkt am Bildschirm verfolgt werden kann (Wittwer *et al.*, 1997). Die Änderung der Konzentration des PCR-Produktes über die Zeit erlaubt dann Rückschlüsse auf die Ausgangskonzentration der betreffenden Nukleinsäure. Unterschiede in den Transkriptmengen werden durch Vergleiche der *crossing points* (CP-Wert) zwischen mindestens zwei biologischen Replikaten in jeweils zwei technischen Replikaten ermittelt. Die *crossing points* geben die

Schwelle an, bei der sich die Fluoreszenzsignale des PCR-Produkts statistisch signifikant von denen des Hintergrunds abheben und sind somit ein direktes Maß für die Menge an cDNA (und damit auch RNA) in der zu untersuchenden Probe. Aus mathematischer Sicht ist der CP als die x-Koordinate des Maximums der zweiten Ableitung der Amplifikationskurve definiert. Je niedriger der CP, d.h. je eher die Menge an gebildeter cDNA über das Hintergrundsignal hinausgeht, desto höher ist die RNA-Menge in der Ausgangsprobe.

Mit dem LightCycler der Firma Roche Diagnostics und dem „SensiMix™ SYBR One-Step Kit“ (Bioline, Luckenwalde, Deutschland) wurde die RT-qPCR in einem Reaktionsgefäß durchgeführt (1-Schritt-RT-qPCR). Dafür wurden 1 µL der RNA-Probe (300 ng/µL) in die Polycarbonat-Kapillaren vorgelegt, und anschließend wurden 19 µL des Master-Mixes zur RNA-Probe pipettiert. Nach Ablauf der reversen Transkription und der initialen Denaturierung, bei der zugleich die Aktivierung der Polymerase stattfindet, folgten 40 PCR-Zyklen (95 °C für 15 Sek, 58 °C für 20 Sek und 72 °C für 15 Sek). Die CP-Werte wurden mithilfe von CAMPER 1.2 (<http://camper.cebitec.uni-bielefeld.de>) ermittelt, das Effizienz-korrigierte *crossing points* berechnet (Tichopad *et al.*, 2003). Die ermittelten CP-Werte wurden dann in relative Expressionsstärken der Transkripte umgerechnet, welche jeweils die Änderung der Transkriptionsstärke eines Gens verglichen mit der eines Referenzstammes beschreibt.

11.2 Northern Blot-Analyse für kleine RNAs (bis 800 nt)

Der Northern Blot ist eine Methode zum spezifischen Nachweis und zur Bestimmung der Größe von RNA-Sequenzen. Die RNA-Proben werden zunächst mittels Gelelektrophorese nach ihrer Größe getrennt. Anschließend werden die getrennten RNA-Fragmente auf eine Membran übertragen (*blotting*) und mittels spezifisch markierter Sonden über komplementäre Basenpaarung markiert. Die Sondendetektion kann radioaktiv erfolgen oder wie in dieser Arbeit über Chemolumineszenz. Die Schritte zur Erstellung der RNA-Sonden und die verwendete *blotting*-Methode werden nachfolgend erläutert.

11.2.1 Erstellung Digoxigenin-markierter RNA-Sonden

Um RNA-Sonden herzustellen wurde zunächst genomische DNA aus *C. glutamicum* isoliert. Für die Gewinnung von Gesamt-DNA aus Gram-positiven Bakterien wurde das *Genelute Bacterial Genomic DNA Kit* (Sigma, Taufkirchen, Deutschland) verwendet. Bei dieser Methode spaltet das Enzym Lysozym die glykosidischen Bindungen des Peptidoglukangerüsts der Bakterienzellwand auf um die DNA freizustellen. Es wurde nach Herstellerangaben vorgegangen. Anschließend wurde die benötigte RNA-Sonde mittels isolierter DNA-Vorlage durch *in vitro*-Transkription über eine T7-RNA-Polymerase (Roche, Mannheim, Deutschland), einen DIG-RNA Labeling-Mix (Roche, Mannheim, Deutschland) mit genspezifischen Primern erzeugt. Der reverse Primer enthält jeweils die T7-Promotorsequenz

(GAAATTAATACGACTCACTATAGGG) am 5'-Ende der Sequenz. Zur Markierung der entstehenden RNA-Sonde wurde Digoxigenin (DIG) verwendet. DIG ist ein Steroid, also ein Naturstoff, der nur in Digitalis-Pflanzen vorkommt. Im NTP-Mix ist DIG kovalent über einen linearen Abstandshalter (*spacer*) an 35% der enthaltenen Uridintriphosphatmoleküle (UTP) gebunden und wird so von der T7-Polymerase bei der Transkription statistisch gesehen in jede 20.-25. Base in die entstehende RNA-Sonde eingebaut. Nach erfolgter *in vitro*-Transkription wird die Sonde über LiCl-Fällung gereinigt und bis zur späteren Verwendung bei -80 °C eingefroren.

11.2.2 Semidry-Elektroblot

Nach Vorbereiten der Gelkammer PerfectBlue Doppelsystem Twin M (Peqlab, Erlangen, Deutschland) wurden 10 µg isolierte Gesamt-RNA zur Größenauftrennung auf ein 10 % denaturierendes Polyacrylamidgel aufgetragen. Hierbei wurden für das Gel 16,7 mL 30 % Acrylamid, 5 mL 10x TBE und 21 g Harnstoff mit DEPC-behandeltem H₂O auf 50 mL aufgefüllt. Dann wurden 50 µL TEMED und 500 µL 10 % Ammoniumpersulphat (APS) hinzugefügt. Das Gel wurde anschließend sofort gegossen. Das polymerisierte Gel wurde in der vertikalen Gelkammer fixiert, und die Kammer wurde mit 0,5x TBE befüllt. Nach 30 Min. Vorlauf bei 300 V wurden die kurz erhitzten RNA-Proben und des RNA-Leiters Fermentas™ RiboRuler™ Low Range (Thermo Scientific, Dreieich, Deutschland) auf das Gel geladen und nachfolgend für 75 Min. bei 300 V aufgetrennt. Zur Überprüfung, ob die Auftrennung erfolgreich war, wurde das Gel in Ethidiumbromid gefärbt und gescannt. Der Transfer der separierten RNA auf eine Hybond-N₊2 Membran (GE Healthcare, München, Deutschland) erfolgte durch Semi-Dry Elektroblotting für 1 Std. bei 300 mA. Anschließend wurde die RNA durch Backen bei 120 °C für 0,5 Std. fixiert und der RNA-Leiter auf der Membran unter UV-Licht mit Bleistift nachgezeichnet. Die Membran wurde danach bei 37 °C für 1 Std. in Prähybridisierungslösung (ohne Sonde) inkubiert, um reaktive Membranbindestellen zu blockieren. Vor der Hybridisierung wurden 200 ng der Digoxigenin (DIG-) gelabelten RNA-Sonde in 3 mL Prähybridisierungslösung bei 95 °C für 10 Min. denaturiert. Anschließend wird die Sonde zu frischer Hybridisierungs-Lösung gegeben und zusammen mit der Membran bei 37 °C im Roller inkubiert. Um die Hybridisierungslösung und unspezifische gebundene RNA-Sonde zu entfernen, wurde die hybridisierte Membran am nächsten Tag zweimal für je 10 Min. mit Waschlösung I (2x SSC und 0,1 % SDS) bei RT gewaschen und dann dreimal für 15 Min. bei 68 °C mit Waschlösung II (0,1x SSC und 0,1 % SDS). Um die Hybridisierungssignale immunologisch zu detektieren wurde die Membran zunächst mit einem Anti-Digoxigenin-Antikörper, der mit Alkalischer Phosphatase konjugiert ist, behandelt (Anti-Digoxigenin-AP, *Fab fragments*, Roche, Mannheim, Deutschland). Anschließend wurde für den Chemilumineszenz-Scan CDP-Star (Roche, Mannheim, Deutschland) zugegeben, welches durch die Alkalische Phosphatase enzymatisch dephosphoryliert wird, was wiederum zu einer Lichtemission bei 466 nm führte. Die Signale wurden mittels eines Fusion FX7

Scanners (Peqlab, Erlangen, Deutschland) aufgenommen. Die Größe der detektierten Signale wurde mit dem RNA-Leiter verglichen, der auf der Membran markiert wurde.

12 Microarrayanalysen

In dieser Arbeit wurde einerseits eine neue *tiling*-Microarraymethode für intergenische Genombereiche entwickelt und andererseits schon länger etablierte vergleichende Zweifarben-cDNA-Microarrayanalysen (Hüser *et al.*, 2003) zwischen dem *C. glutamicum*-Wildtyp und den zwei Deletionsmutanten der kleinen RNA *cgb_16115* durchgeführt. Diese Methoden werden im Folgenden beschrieben, wobei besonders auf die neu etablierte Methode und Schritte zur Auswertung eingegangen wird.

12.1 *Tiling*-Microarrays zur intergenischen Transkriptanalyse in *C. glutamicum*

Die im Folgenden dargestellte Einfarben-Microarraymethode über die genom@one Plattform beruht auf Validierungsexperimenten, die mithilfe des Enterobakteriums *Escherichia coli* str. K-12 MG1655 durchgeführt wurden (siehe Anhang). Die Methode wurde daraufhin äquivalent auf *C. glutamicum* übertragen.

12.1.1 Berechnung des intergenischen *C. glutamicum tiling*-Microarrays

Die benötigten Daten zur Erstellung des intergenischen *C. glutamicum tiling*-Microarrays mittels IRMA (Schroeder, 2007) wie die Größe und die Sequenz der intergenischen Regionen, werden aus der GenBank-Datei (Benson *et al.*, 2013) von *C. glutamicum* ATCC 13032 (NC_006958.1) gewonnen. Insgesamt wurden 10.521 50-mer Oligonukleotidproben mit 15-mer Überlappungen in den Regionen zwischen allen bekannten Kodierregionen für einen Strang des *C. glutamicum* Genoms berechnet. Von jeder Probe wurde zusätzlich das reverse Komplement berechnet, so dass beide Stränge des Genoms berücksichtigt sind. Zusätzlich wurden als Kontrollen die Startsequenzen von 60 tRNAs für die Synthese von 50-mer Oligonukleotidproben berechnet, und zwar in *sense*-Richtung und dazu ebenfalls die reversen Komplemente (*antisense*-Richtung).

12.1.2 RNA-Probenvorbereitung und Hybridisierung

Die Qualität und Quantität der aus den sechs verschiedenen Experimenten aufgereinigten Gesamt-RNA wurde mit dem Agilent Bioanalyzer überprüft. Anschließend wurde die RNA mittels FlashPAGE™ Fractionator-System (Applied Biosystems, Foster City, U.S.A.) nach Herstellerangaben bei einer Laufzeit von 45 Minuten Größen-selektioniert, um RNA-Moleküle bis zu einer Größe von ~250 nt Länge zu

erhalten. Anschließend wurde die kleine RNA Fraktion ohne cDNA-Synthese direkt mit dem miRVana miRNA Labeling Kit (Applied Biosystems, Foster City, U.S.A.) nach Standard-Protokoll Cy3-markiert. Die markierten Proben wurden in einer Vakuumzentrifuge getrocknet und im miRNA Hybridization Buffer (febit biomed GmbH, jetzt CBC, Heidelberg, Deutschland) resuspendiert.

Auf den intergenischen *C. glutamicum*-Microarrays wurden spezifische Hybridisierungen mit je 3 µg der markierten kleinen RNA-Fractionen für 16 Stunden bei 42 °C in der Geniom® one-Maschine (febit biomed GmbH, jetzt CBC, Heidelberg, Deutschland) durchgeführt. Dabei wurden von jedem Experiment zwei biologische Replikate hybridisiert. Anschließend wurde nach der “miRNA Standard Waschprozedur“ gewaschen, worauf die Biotin-Detektion mittels Streptavidin-Phycoerythrin (SAPE) (Invitrogen, Carlsbad, U.S.A.) erfolgte. Vor Quantifizierung der Signale wurden biotinierte Anti-Streptavidin-Antikörper (Vector Laboratories, Burlingame, U.S.A.) zur Signal-Amplifikation eingesetzt und mit einer zweiten SAPE-Inkubation kombiniert. Zuletzt wurden die Signale mit einem geeigneten Absorptionsfilter für Cy3 (552 nm) erfasst, jeweils mit einer doppelt so hohen Belichtungszeit wie durch die automatische-Belichtungsfunktion der Geniom-Software vorgeschlagen wurde. Die emittierten Photonen der Fluoreszenzmessung wurden von einem *photo multiplier tube* (PMT) verstärkt. Die Signale, die empfangen werden können liegen im Intervall $[0, 65535 \triangleq 2^{16}-1]$ und werden gemessen als dimensionsloser Quotient in Graustufen. Zur Signalanalyse wurde der Geniom *feature extraction* Algorithmus eingesetzt. Danach wurden die Rohdaten als Textdatei exportiert.

12.1.3 Normalisierung und Auswertung der Microarraydaten

Die Rohdaten der durchgeführten Hybridisierungen wurden über die biologischen Replikate jedes Experimentes nach Quantil (Bolstad *et al.*, 2003) normalisiert. Diese Art der Normalisierung stellt eine Erweiterung der globalen Signalkorrektur durch Medianzentrierung dar (ebd.). Hier wird davon ausgegangen, dass alle Quantile (und nicht nur das 50 %-Quantil = Median) eines vergleichbaren Arrays identisch sind. Bei der Quantil-Normalisierung wird daher die Verteilung der Microarraysignale entsprechend ihrer Quantile aneinander angepasst. Der Algorithmus lautet wie folgt:

- Ordne die Signale eines jeden Microarrays der Größe nach.
- M_n sei der Mittelwert der Signale mit der n -t höchsten Signal. Ersetze den Messwert dieser Signale jeweils durch M_n
- Verfahre so für alle Positionen n .

Von der Geniom-Software als fehlerhaft markierte Signale wurden bei der Normalisierung verworfen. Anschließend wurde jeweils über den Array-Median reskaliert. Abschließend wurde das arithmetische Mittel zusammen mit der Standardabweichung über die Replikate berechnet.

Im experimentellen Design ist es nicht vorgesehen, die verschiedenen Experimente aufeinander zu beziehen und eine davon als Referenz zu betrachten. Den nächsten Schritt stellt also die Bestimmung eines spezifischen Signal-Schwellenwertes dar, der für transkriptionelle Aktivität in jedem einzelnen Experiment steht. Dieser Schwellenwert wurde über die Signale der Negativ-Kontrollen definiert, d.h. aus den Signalen der *antisense*-Sequenzen von tRNAs eines Microarrayexperimentes (Anzahl 120). Der maximale Wert aller Negativ-Kontrollen in einem Experiment stellt dabei die Grenze zum Hintergrundrauschen dar und gibt den Schwellenwert wieder, ab dem Transkription angenommen wird. Über die Analyse der Signalverteilung in den Experimenten wurde eine quantitative Einteilung der Signale vorgenommen. Bei Erreichen des doppelten Standard-Schwellenwertes ($\hat{=}$ den oberen ~25 % der Signale) wird ein „deutliches Transkriptsignal“ angenommen, und bei Erreichen des vierfachen Standard-Schwellenwertes ($\hat{=}$ den oberen ~10 % der Signale) wird ein „sehr deutliches Transkriptsignal“ angenommen.

12.2 *C. glutamicum* Zweifarben-cDNA-Microarrays

Die Vorbereitung der *C. glutamicum* Zweifarben-cDNA-Microarrays erfolgte jeweils mit 10 μ g Gesamt-RNA der isolierten Gesamt-RNA Proben (RNeasy Mini Kit, Qiagen, Hilden, Deutschland). Für die cDNA-Synthese wurde nach der Standardmethode (Hüser *et al.*, 2003) vorgegangen. Die Fluoreszenzmarkierung der Proben und der Referenz erfolgt mithilfe des Cy5- bzw. Cy3-Monoreactive Dye™ Pack (Amersham Biosciences) und in Farbumkehr für die biologischen Replikate. Während der Reaktion wird Erststrang-cDNA nachträglich über Kopplung von Carbocyanin-*N*-Hydroxysuccinimid (NHS)-Ester an Aminoallyl-substituierte (AA)-dUTPs fluoreszenzmarkiert. Ebenso nach der Standardmethode wurden die Hybridisierungen, das Waschen der Microarrays und die Analyse der *Spot*-Signale mit einem Laser-Fluoreszenzscanner (LS-Reloaded, Tecan) durchgeführt. Dabei werden Fluoreszenzbilder bei den jeweiligen exzitatorischen Wellenlängen (Cy5 = 663 nm, Cy3 = 552 nm) erstellt. Nach dem Scannen fanden die Lokalisierung der *Spots* und die Intensitätsbestimmung der Signale sowie der Hintergrundsignale statt. Dies wurde mithilfe der Computersoftware Image 6.0 (BioDiscovery) ausgeführt. Im Anschluss erfolgte die statistisch-mathematische Auswertung der Microarraydaten mithilfe des EMMA2-Programms (Dondrup *et al.*, 2009). Hierbei wurden die Daten nach LOWESS (Yang *et al.*, 2002) normalisiert, worauf ein Einstichproben-t-Test über die normalisierten Daten folgte. Alle Gene mit M-Werten $\geq 0,6$ oder $\leq -0,6$ bei signifikanten p-Werten $< 0,01$ (Irrtumswahrscheinlichkeiten $< 1\%$), wurden als signifikant verändert in einer Deletionsmutante gegenüber dem *C. glutamicum*-Wildtypstamm gewertet.

13 cDNA-Sequenzierung kleiner RNAs (RNA-Seq)

Das Experiment zur Sequenzierung kleiner RNAs umfasst sowohl die Vorbereitung der cDNA-Sequenzierungs-Bibliotheken im Labor als auch die bioinformatische Analyse und Auswertung der Sequenzierungsergebnisse. Die einzelnen Schritte werden nachfolgend erläutert.

13.1 Erstellen der cDNA-Sequenzierungs Bibliotheken

Für das Sequenzierungsexperiment wurden ~100 µg Gesamt-RNA eines RNA-Pools verwendet, der sich zu gleichen Teilen aus Trizol®-isolierter RNA der beschriebenen Kultivierungsexperimente zusammensetzt (16 µg Gesamt-RNA je Wachstumsbedingung). Zuerst wurde dieser RNA-Pool präzipitiert für kleine RNAs < 250 nt (mit 2.5 M Natriumacetat, 25 %; PEG 8000). Anschließend sind rRNAs mit dem Ribo-Zero Kit für Gram-positive Bakterien (Epicentre, Madison, U.S.A.) nach Herstellerangaben abgereichert worden. Dieser rRNA-abgereicherte Pool kleiner RNAs wurde dann aufgetrennt in zwei Ansätze, wobei ein Ansatz zur Vorbereitung einer cDNA-Bibliothek kleiner Primärtranskripte dient und der andere zur Erstellung einer cDNA-Bibliothek der Gesamt-Fraktion kleiner RNAs (dRNA-Seq Methode Sharma *et al.*, 2010). Zur Herstellung der cDNA-Bibliothek von Primärtranskripten wurde im ersten Schritt die Terminator™ 5'-Phosphate-Dependent Exonuclease (Epicentre, Madison, U.S.A.) nach Herstellerangaben eingesetzt. Diese 5'-3'-Exonuklease verdaut speziell bereits prozessierte oder degradierte RNAs, die an ihrem 5'-Ende einen Monophosphat-Rest aufweisen. Nicht-prozessierte primäre RNA-Transkripte weisen an ihren 5'-Ende einen Triphosphat-Rest auf und bleiben von diesem enzymatischen Verdau unbeeinflusst. Im zweiten Schritt wurde das Enzym RNA 5'-Polyphosphatase (Epicentre, Madison, U.S.A.) nach Herstellerangaben eingesetzt. Diese Mg²⁺-unabhängige Phosphohydrolase entfernt γ - und β -Phosphate der Primärtranskripte und bereitet sie damit für die Ligation der Sequenzierungsadapter vor.

Das Illumina® TruSeq™ Small RNA Kit (Illumina, San Diego, U.S.A.) setzt an dieser Stelle an und ligiert spezifische Adapter an die 5'-Monophosphat und 3'-Hydroxyl-Gruppen der kleinen RNAs, die später zur strangspezifischen Auswertung der Sequenzierungsdaten führen können (Abbildung III-2). Das Protokoll umfasst weiterhin die reverse Transkription der Adapter-ligierten RNA-Moleküle. Die Einzelstrang-cDNAs werden dann mittels eines Adapter-spezifischen Primerpaares PCR-amplifiziert. Eine abschließende Gelaufreinigung folgte in Vorbereitung auf die folgende Cluster-Generierung und Hochdurchsatz-Sequenzierung. Zur Erstellung beider cDNA-Bibliotheken wurde nach Herstellerangaben vorgegangen. Nach verschiedenen Schritten wurde jeweils die Größenverteilung der entstandenen RNA-Zwischenprodukte mit dem Agilent Bioanalyzer (Agilent Technologies, Waldbronn, Deutschland) getestet. Beide cDNA-Bibliotheken wurden anschließend mit dem Illumina GA IIx-System (Illumina, San Diego, U.S.A.) sequenziert, dabei wurden 35 bp-lange *single-reads* vom 5'-Ende aus gelesen.

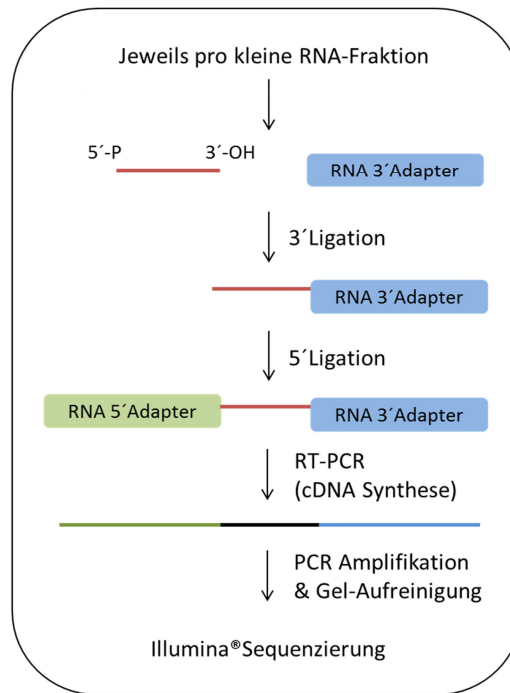


Abbildung III-2 Vorbereitungsschritte zur Erstellung der cDNA-Bibliotheken aus den Transkripten der kleinen RNA-Fractionen mit dem Illumina® TruSeq™ Small RNA Kit.

13.2 Read-mapping und Datenvisualisierung

Die Sequenzierungs-*reads* wurden mit dem SARUMAN-Algorithmus (Blom *et al.*, 2011) der *C. glutamicum* ATCC 13032 Genomsequenz (Kalinowski *et al.*, 2003) zugeordnet, wobei bis zu zwei Fehler pro Sequenzierungs-*read* erlaubt wurden. Das Programm *ReadXplorer* (Hilker *et al.*, Manuskript submitted) wurde verwendet, um die zugeordneten Sequenzierungsdaten in Bezug auf die *C. glutamicum* Referenz-Genomsequenz zu visualisieren.

13.3 Identifizierung von Transkriptionstarts

Um automatisch und systematisch Transkriptionsstarts zu bestimmen, wurden die Daten der Primärtranskript-Bibliothek 1 verwendet. Dabei wurden zuerst für jeden Strang und jede Position des Genoms die startenden *reads* (*read*-Stapel) analysiert (Programm zur Analyse entwickelt und implementiert von C. Rückert, CeBiTec, Universität Bielefeld). Als mögliche Transkriptionsstarts wurden alle Positionen pro Strang ausgewählt, die den folgenden Kriterien entsprachen: Für eine Position i überschreitet die Anzahl von *read*-Starts x_i auf diesem Strang an dieser Position einen Hintergrundschwellenwert T und das Verhältnis x_i/x_{i-1} an dieser Position muss über einem Schwellenwert R liegen. Nach manueller Analyse wurde T auf 19 gesetzt und R auf 5, da diese Parameter ein gutes Signal-Rausch Verhältnis ergeben haben.

13.4 Promotorsuche

Im Rahmen einer Promotoranalyse wurden die Sequenzen stromaufwärts der putativen Transkriptionsstarts untersucht. Das Tool Improbizer (Ao *et al.*, 2004; Tompa *et al.*, 2005) wurde hierfür jeweils mit 158 publizierten Sequenzen von SigA-Bindestellen (Pátek und Nešvera, 2011) und 45 SigH-Bindestellen (Busche *et al.*, 2012) trainiert. In beiden Trainingssets wurden die -10- und -35-Promotor-Regionen über den *expectation maximization* (EM)-Algorithmus korrekt bestimmt. Wie von Improbizer vorgeschlagen, wurden für die Bestimmung eines signifikanten Hintergrundwertes (*background significance score*) Kontrollläufe durchgeführt. Dieser Wert (arithmetisches Mittel) wurde als Schwellenwert in den Improbizeranalysen für die zu testenden Sequenzen verwendet. In den Improbizer-Läufen wurden die Aufwärtssequenzen möglicher Transkriptionsstarts mit Sequenzen des Trainingssets jeweils im Verhältnis 1:10 eingesetzt, was nahezu keine Auswirkung auf die Motivsuche und Wertung gezeigt hat. Lagen die Werte der -10 und -35 Regionen über den SigH-Schwellenwerten und zeigten einen *spacer*-Abstand von 16-20 nt, wurden die Testsequenzen als SigH-Promotor-indiziert eingestuft. Es wurde beobachtet, dass -35-Regionen von SigA-Promotoren vergleichsweise schlecht konserviert sind (Pátek und Nešvera, 2011). Daher galt ein SigA-Promotor schon als indiziert, wenn für die gut konservierte erweiterte -10-Region Werte größer oder gleich dem maximalen signifikanten Hintergrundwert erhalten wurden, unabhängig davon, welchen Wert die schlecht konservierte -35-Region betrug. Wenn für drei aufeinanderfolgende Transkriptionsstart-Positionen jeweils dieselbe Promotorsequenz identifiziert wurde (in 278 Fällen), ist die Genom-Position mit der höchsten Anzahl an *read*-Starts als Transkriptionsstart gewertet worden.

13.5 Vorhersage proteinkodierender Transkripte

In dieser Arbeit wurde die ORF (*open reading frame*) Suche mit einer minimalen Proteinelänge von 48 nt, die bereits vom Leaderpeptid des transkriptionellen Attenuator vor *ilvB* (Morbach *et al.*, 2000) bekannt ist, innerhalb der Sequenzen der erhaltenen kleinen RNA-Transkripte durchgeführt. Dabei wurde mit dem Tool ORFfinder (Uhmin, Osaka University, Japan, <http://www.gen-info.osaka-u.ac.jp>) gearbeitet unter Verwendung der folgenden Start-Codons: AUG, GUG, UUG und Stopp-Codons: UAA, UAG, UGA. Für einen vorhergesagten ORF wurde nach ribosomalen Bindestellen (RBS) mit dem RBSfinder (Suzek *et al.*, 2001) unter Verwendung der Standard RBS-Einstellungen (AGGAG) innerhalb eines Fensters von 15 bp gesucht. Nur Transkripte, die sowohl einen ORF und eine geeignete RBS besitzen wurden als möglicherweise proteinkodierend eingestuft. Im Fall von Transkripten, für die der Beginn des vorhergesagten ORF auf dieselbe Position des Transkriptionsstarts fiel (*leaderless transcripts*), wurde auf die RBS-Analyse verzichtet.

14 Methoden zur bioinformatischen Vorhersage regulatorischer RNAs

Für die Vorhersage regulatorischer RNAs wurde der RNAz-Algorithmus (Washietl *et al.*, 2005) angewendet. Die Voraussetzung hierfür ist ein multiples Genom-Alignment, das mithilfe der Mauve Software (*Multiples Alignment of Conserved Genomic Sequence with Rearrangements*) (Darling *et al.*, 2004) erstellt wurde. Hierfür wurden die GenBank-Sequenzdateien (Benson *et al.*, 2013) von *C. glutamicum* ATCC 13032 (NC_006958.1) und der verwandten Corynebakterien *C. diphtheriae* NCTC 13129 (NC_002935.2) sowie *C. efficiens* YS-314 (NC_004369.1) verwendet. Die Details der bioinformatischen Vorgehensweise werden in den nächsten Kapiteln erläutert.

14.1 Erstellung eines multiplen Genom-Alignments mit Mauve

Über Mechanismen wie Rekombination kann es zu Neuankordnungen im Genom kommen. So können orthologe Regionen in verwandten Genomen entweder invertiert oder an anderen Stellen wiederzufinden sein. Mauve (Darling *et al.*, 2004) identifiziert während des Alignierungsprozesses konservierte Segmente aus den eingegebenen Genomen, die frei von Genomumordnungen erscheinen. Solche Segmente werden als *locally collinear blocks* (LCBs) bezeichnet. Mauve verwendet hierzu eine *anchor*-Alignment Technik (Ma *et al.*, 2002). Entgegen anderen Methoden zur Generierung von Genom-Alignments, erlaubt Mauve eine Umordnung der Alignment-Anker (*anchors*) in jedem Genom. Um als *anchor* eines LCB in Betracht zu kommen, muss eine minimale Länge von Nukleotiden zwischen den Genomen übereinstimmen. Diese minimale Länge (*minimum match size*) ist benutzerdefinierbar. Standardmäßig berechnet Mauve einen Wert, der automatisch in Abhängigkeit von den zu alignierenden Genomen erzeugt wird. Durch Mauve zufällig identifizierte *anchor*, die aufgrund kleiner und nicht signifikanter Genomumordnungen gebildet werden, müssen über ein minimales Gewicht (*minimum weight*)-Kriterium herausgefiltert werden. Das Gewicht ist definiert über die Summe der Längen von Übereinstimmungen im LCB. Als Standardwert setzt Mauve das minimale Gewicht drei Mal so hoch an wie die *minimum match size*. Nachdem alle *anchor*-Paare bestimmt sind, die die einzelnen LCBs definieren, wendet Mauve auf jeden LCB den ClustalW-progressiven globalen Alignment-Algorithmus an (Larkin *et al.*, 2007). Bei der Erstellung des multiplen Genom-Alignments werden von Mauve fünf verschiedene Ausgabedateien erzeugt. Zwei dieser Dateien, die *.mauve*- und die *.alignment*-Datei enthalten das Alignment in unterschiedlichen Formaten. Die *.alignment* Datei beinhaltet das gesamte Genom-Alignment im *extended multi-FastA* (XMFA)-Format. Das XMFA-Dateiformat unterstützt das Speichern vieler kollinearer Sub-Alignments, die jeweils durch ein Komma-Zeichen getrennt sind. Jedes dieser Sub-Alignments besteht aus einer Sequenz im FastA-Format, wobei in der Überschrift Angaben zum Strang und zur Position im entsprechenden Genom verzeichnet sind. Das Format der *.mauve*-Datei stellt gegenüber dem XMFA-Format eine Speicherplatz sparende Variante dar, da hier nur die Koordinaten großer und exakt übereinstimmender Regionen notiert sind. Alle übrigen Dateien enthalten zusätzliche Informationen, die für die Erstellung des Alignments

benötigt wurden (phylogentischer Baum, Identitäts-Matrix, Positionen der in allen Genomen konservierten *backbone*-Regionen und Positionen der Insel-Regionen, in denen nur ein Subset der Genome eine eindeutig übereinstimmende Sequenz zeigt).

14.2 RNAz zur Vorhersage regulatorischer RNAs

Der RNAz-Algorithmus (Washietl *et al.*, 2005) kann zur Suche nach funktionellen RNA-Sekundärstrukturen aus multiplen Sequenz-Alignments verwendet werden. Dabei wird mit der Berechnung einer RNA Konsensus-Sekundärstruktur (*structure conservation index* (SCI) ein Maß für die Konservierung der Sekundärstruktur erzeugt. Der *z-score* stellt ein Maß für die thermodynamische Stabilität dieser Struktur dar. Diese Werte werden allerdings nicht für sich betrachtet, sondern im Algorithmus über einen Klassifikator kombiniert. Der Algorithmus basiert auf dem Klassifikator einer binären Support-Vektor-Maschine (SVM). Diese SVM wurde mithilfe von zwölf bekannten funktionellen RNA-Klassen aus der Rfam-Datenbank (Griffiths-Jones *et al.*, 2003) trainiert. Hierbei sind Test-Alignments generiert worden, die sowohl von Sequenzen einer RNA-Klasse als auch von Zufallssequenzen ohne native Sekundärstruktur erzeugt wurden. Die Kombination aus errechnetem *z-score*, SCI-Index und der *mean-pairwise identity* (MPI) des Alignments führte zur Klassifizierung. Die Vorhersagen über die zwölf getesteten RNA-Klassen mit einer MPI > 60 % bei einer *class-probability* $p > 0,5$ waren insgesamt sensitiver als bei $p > 0,9$. Wobei die Vorhersagen mit höherer Wahrscheinlichkeit ($p > 0,9$) dafür aber spezifischer waren (Washietl *et al.*, 2005).

In der durchgeführten RNAz-Suche wurde im Mauve erzeugten multiplen Genom-Alignment über *C. glutamicum*, *C. diphtheriae* und *C. efficiens* in Alignments mit einer Länge größer 200 bp in Fenstergrößen von 100 bis 200 bp, schrittweise um 40 bp versetzt, und in beiden Strängen (*class-probability* $p \geq 0,5$) gesucht. Dieses initiale Set von Vorhersagen wurde anschließend gefiltert nach Vorhersagen, die einer *mean-pairwise-identity* MPI > 60 % entsprachen. Außerdem wurden vorhergesagte rRNAs und tRNAs aus dem Set entfernt und alle übrigen überlappenden Vorhersagen zu einer neuen Region zusammengezogen, wobei deren äußerste 5'- und 3'-Enden als neue Start- und Stopp-Positionen definiert wurden.

IV. Ergebnisse

1 Bioinformatische Vorhersage (intergenischer) kleiner RNAs in *C. glutamicum*

Seit einigen Jahren nimmt das Wissen über die enorme Bedeutung von RNA-Regulationen durch kleine RNA-Moleküle (*small RNAs*, sRNAs) zu, die unter bestimmten Bedingungen die Genexpression für verschiedenste zelluläre Prozesse beeinflussen. In der Mehrzahl der publizierten Genome fehlen diese RNA-Elemente allerdings, da die meisten Programme zur automatischen Genannotation auf Basis von putativen ORFs oder konservierten Proteinsequenzen arbeiten und daneben nur die hochkonservierten rRNA- und tRNA-Gene detektieren, die eine elementare Rolle bei der Proteinbiosynthese in allen Organismen spielen. Kleine regulatorische RNAs liegen häufig in den intergenischen Bereichen zwischen den annotierten proteinkodierenden Genen. Sie wurden jedoch eher mit konservierten Sekundärstrukturen als konservierten Primärsequenzen beobachtet und das zumeist nur in nah verwandten Bakterien (Sridhar und Gunasekaran, 2013). Somit können spezifische Vergleiche und die *in silico*-Analyse von Genomsequenzen auch zur bioinformatischen Vorhersage von regulatorischen kleinen RNAs führen. Die heute bereits existierenden Vorhersagen von kleinen RNAs in *C. glutamicum* haben jeweils nur eine geringe Anzahl an vorhergesagten kleinen RNAs ergeben. In diesem Rahmen sind die Vorhersagen über Kovarianzmodelle der Rfam-Datenbank (Gardner *et al.*, 2009) zu nennen oder die zeitgleich mit dieser Arbeit entstandene SIPHT-Vorhersage für *C. glutamicum*, die auf konservierten intergenischen Primärsequenzen stromaufwärts von Rho-unabhängigen Terminatoren basiert (Livny *et al.*, 2008). Die vor kurzem publizierte NAPP-Pipeline (Ott *et al.*, 2012) liefert phylogenetisch konservierte Bereiche zwischen Kodierregionen, die möglicherweise kleine RNAs kodieren, wobei diese jedoch keine genauen Positionsangaben enthalten.² In dieser Arbeit wurde ein anderes Vorgehen auf Basis des bereits implementierten RNAz-Algorithmus (Washietl *et al.*, 2005) gewählt. Hierbei wird nach konservierten Sekundärstrukturen mit thermodynamischer Stabilität in multiplen Sequenz-Alignments nah verwandter Bakterien gesucht. Stabile Sekundärstrukturen stellen eine wichtige funktionelle Eigenschaft von kleinen RNAs dar und können daher effizient für Vorhersagen eingesetzt werden (Sridhar und Gunasekaran, 2013). Zusätzlich zum *C. glutamicum* Genom wurden die Genome der nah verwandten Corynebakterien *C. efficiens* und *C. diphtheriae* zur vergleichenden Analyse ausgewählt. Die zwei bodenbewohnenden Corynebakterien-Spezies *C. glutamicum* und *C. efficiens* sind beide natürliche Produzenten von L-Glutamat und zeigen viele übereinstimmende Merkmale wie die Genomgröße und die Anzahl von Genen. *C. diphtheriae* ist dagegen ein humanpathogenes Bakterium mit einer kleineren Genomgröße, es

² Die Vorhersage-Daten der NAPP-Contigs für *C. glutamicum* ATCC 13032 (569 intergenische Regionen) befinden sich im Anhang auf der Daten-CD im Ordner „Vorhersage“ in Datei C_glutamicum_ATCC_13032_Bi-NAPP-Contigs.xlsx.

wurde jedoch ein hoher Grad an Syntenie für alle drei Spezies beobachtet (Kalinowski *et al.*, 2003). Die über RNAz erhaltenen Kandidaten werden im Anschluss mit den bereits existierenden Vorhersagen kleiner regulatorischer RNAs und einer Vorhersage Rho-unabhängiger Terminatoren in *C. glutamicum* verglichen.

1.1 Vorhersage regulatorischer RNAs in *C. glutamicum* mit dem Programm RNAz

Um das Programm RNAz (Washietl *et al.*, 2005) zur Bestimmung kleiner RNAs anzuwenden, wurde im ersten Schritt ein multiples Genom-Alignment (MGA) über die Genome der verwandten *Corynebacterium*-Spezies *C. glutamicum* ATCC 13032, *C. diphtheriae* NCTC 13129 und *C. efficiens* YS-314 mit der Software Mauve (*Multiples Alignment of Conserved Genomic Sequence with Rearrangements*) (Darling *et al.*, 2004) über Standardeinstellungen erzeugt.³ Dieses Alignment wurde mit insgesamt 20 LCBs (*locally collinear blocks*) unterschiedlicher Größe berechnet, was bedeutet, dass diese frei von Genom-Umordnungen erscheinen (Abbildung IV-1). Diese LCBs ergeben sich aus insgesamt 83 Intervallen, wovon die letzten 63 Intervalle keine orthologen Regionen in einem der anderen Genome enthalten. Hierzu gehört auch Intervall 62, das ab Position 1868862 in *C. glutamicum* beginnt, und dem > 180 kbp großen Prophagen CGP3 (Kalinowski *et al.*, 2003) entspricht.



Abbildung IV-1 Multiples Genom-Alignment über *C. glutamicum*, *C. efficiens*, *C. diphtheriae*.

Das Alignment wurde mit der Software Mauve 2.3.1 (Darling *et al.*, 2004) berechnet und visualisiert mit der Mauve-GUI. Insgesamt wurden 20 unterschiedlich lange LCB Einheiten mit Konservierung in allen drei Genomen erhalten, in denen jeweils auch die Ähnlichkeiten abgebildet sind.

³ Die Datei Mauve.alignment befindet sich im Ordner „Vorhersagen“ auf der beiliegenden Daten-CD.

Von den 20 Intervallen, die in allen Genomen konserviert sind, stellen Intervall 4 (Länge 1.230.815 bp) und Intervall 15 (Länge 1.153.148 bp) die größten dar. Diese sind als solche gut in Abbildung IV-1 zu erkennen (LCBs-Farben Lila-Intervall 4 und Blau-Intervall 15). Das kleinste LCB betrifft Intervall 18 mit einer Größe von 205 bp. Dieses Intervall 18 sowie andere kleine Intervalle sind in Abbildung IV-1 nicht zu erkennen.

Mit Unterstützung von K. Gkogkoglou (Praktische Informatik, Universität Bielefeld) wurde das über Mauve generierte Genom-Alignment (nicht-Standard XMFA-Format) in das Standard XMFA-Format umgewandelt. Danach wurde das Alignment im Bereich zwischen allen Kodierregionen mit RNAz (Washietl *et al.*, 2005) nach konservierten Sekundärstrukturen untersucht, die in wenigstens zwei der Genome auftreten. Hierbei wurde mit RNAz-Standardinstellungen in Fenstergrößen zwischen 100 bp bis 200 bp, schrittweise um 40 bp versetzt und in beiden Strängen gesucht. Die Klassifikation in RNAz erfolgt basierend auf einer *support vector machine* (SVM), die *z-scores* als Maß für thermodynamische Stabilität und den *structure conservation index* (SCI) des Alignments ausgibt, wobei auch die *mean pairwise identity* (MPI) eine Rolle spielt. Es wurden 4046 Vorhersagen erhalten, die zur sensitiven Detektion regulatorischer RNAs reduziert wurden (Abbildung IV-2, nächste Seite). Um die Vorhersage Falsch-Positiven zu minimieren, wurden 3638 Vorhersagen selektiert, die einer *mean-pairwise-identity* M.P.I. > 60 % und einer *class-probability* $p \geq 0,5$ entsprechen (Washietl *et al.*, 2005). Anschließend wurden vorhergesagte rRNA- und tRNA-Regionen aus dem Set herausgefiltert, so dass noch 1777 Vorhersagen vorlagen. Zum Schluss wurden überlappende Vorhersagen zu einer neuen Region zusammengezogen und deren äußerste 5'- und 3'-Enden als neue Start- und Stopp-Positionen definiert. Hierbei wurde der maximale *p*-Wert innerhalb der jeweiligen Region notiert. Dieser Prozess lieferte ein Set von 599 Kandidaten, die teilweise recht große Regionen umfassen (Maximum 539 nt, Minimum 82 nt, arithmetisches Mittel 220 nt). Die Erhöhung der Stringenz auf $p \geq 0,9$ ergab noch 339 Kandidaten.⁴

Ein Vergleich der RNAz-Kandidaten mit Vorhersagen der Rfam-Datenbank (Gardner *et al.*, 2009) für *C. glutamicum* lieferte 13 überlappende Bereiche. So wurden alle drei RNA-Haushaltsgene (4.5S RNA, tmRNA, M1 RNA) in den entsprechenden Regionen übereinstimmend auch mittels RNAz detektiert. Außerdem wurden die Rfam-Vorhersagen der 6C RNA, vier TPP-Riboswitches, das *mraW*-Motiv, drei

⁴ Die gesamte RNAz-Vorhersage und die reduzierte endgültige Liste sind in der Datei RNAz-Vorhersage.xlsx im Ordner „Vorhersagen“ auf der beiliegenden Daten-CD zu finden.

Riboswitch-ähnliche *yybP-ykoY*-Motive und ein *ydaO-yuaA*-Motiv über die Analyse nach konservierten Sekundärstrukturen getroffen. Mehr als ein Drittel der RNAz-Vorhersagen (282) fallen in intergenische Bereiche mit phylogenetischer Konservierung, die über die NAPP-Pipeline bestimmt wurden. Von den 15 vorhergesagten Kandidaten kleiner RNAs über SIPHT (Livny *et al.*, 2008) konnten 11 Kandidaten ebenfalls über RNAz detektiert werden, jedoch nicht die folgenden:

Candidate_3_NC_006958, Candidate_4_NC_006958, Candidate_11_NC_006958,
Candidate_15_NC_006958.

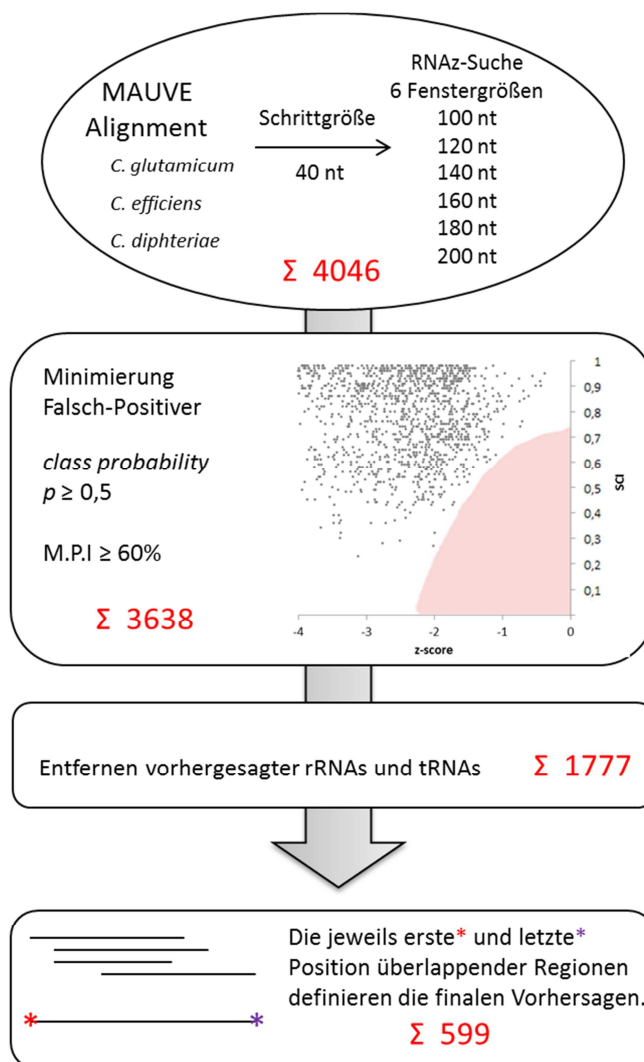


Abbildung IV-2 RNAz-Vorhersage für *C. glutamicum* und stringente Reduktion.

Nach Anwenden der Parameter für eine sensitive Detektion konservierter regulatorischer RNAs umfassen die Arbeitsschritte das Entfernen vorhergesagter rRNAs und tRNAs sowie das Zusammenfügen überlappender Vorhersagen. Diagramm zur Minimierung Falsch-Positiver: Die schwarzen Punkte zeigen die RNAz-Vorhersagen für *C. glutamicum* mit den erhaltenen *z-scores* und SCIs, die $p \geq 0,5$ entsprechen. Der rote Bereich im Diagramm steht für $p < 0,5$ (Washietl *et al.*, 2005).

1.2 Suche nach Rho-unabhängigen Terminatoren und deren Korrelation mit RNAz-Vorhersagen

Es ist bekannt, dass bei der Transkription kleiner RNAs häufig an deren 3'-Ende eine Sequenz kodiert ist, so dass im Transkript eine haarnadelförmige Sekundärstruktur entsteht. Die Struktur destabilisiert den Komplex aus RNA-Polymerase und DNA/RNA-Hybrid, so dass er zerfällt und damit eine weitere Elongation verhindert (Hippel und Yager, 1991). Diese vom Rho-Faktor unabhängige Transkriptionstermination ist charakterisiert durch einen G+C-reichen Stamm, innerhalb dessen die Nukleotide nicht paaren und eine Schleife (*stem loop*) formen. Hierauf folgen jeweils drei bis acht kodierte Uracil-Reste. Im Rahmen dieser Arbeit wurde eine Vorhersage von Rho-unabhängigen Terminatoren in *C. glutamicum* mit dem Programm TransTermHP (Kingsford *et al.*, 2007) erzeugt. Hierfür wurde nach G+C-reichen Stammlängen L zwischen $L \geq 4$ und $L \leq 23$ gesucht, wobei ein Konfidenzniveau von 0,75 gesetzt wurde. Weitere Einstellungen wurden nach TransTermHP-Standard gemacht. Dabei wurden insgesamt 2266 Rho-unabhängige Terminatoren über alle Genomregionen von *C. glutamicum* detektiert.⁵ Von dieser Gesamtzahl überlappen 221 Rho-unabhängige Terminatoren mit den über RNAz vorhergesagten konservierten kleinen RNAs und sind in der Datei RNAz-Vorhersage.xlsx gekennzeichnet. Zwei der drei vorhergesagten RNA Haushaltsgene sind Rho-unabhängig terminiert (4.5S RNA, tmRNA). Am Ende der M1 RNA konnte keine entsprechende Terminatorstruktur detektiert werden. Dies ist wahrscheinlich auch der Grund dafür, warum sie nicht über die SIPHT-Vorhersage (Livny *et al.*, 2008) detektiert wurde, denn dieses Tool sucht nach konservierten Primäsequenzen stromaufwärts von Rho-unabhängigen Terminatoren.

⁵ Das gesamte Ergebnis der TransTermHP-Vorhersage befindet sich in der Datei Rho-unabhängige-Terminatoren.xlsx auf der beiliegenden CD im Ordner „Vorhersagen“.

2 Experiment-spezifische *tiling*-Microarrayanalysen in den intergenischen Regionen in *C. glutamicum*

Mithilfe von Microarrayanalysen ist es möglich, eine Expressionsanalyse vieler Tausend Gene auf Transkriptebene simultan durchzuführen (Lipshutz *et al.*, 1999). Das Prinzip hierbei ist, eine Hybridisierung von fluoreszenzmarkierten Nukleinsäure-Sonden gegen immobilisierte Oligonukleotidproben durchzuführen. Unspezifischen Bindungen sollten anschließend über stringente Waschprozesse minimiert werden, um dann die Sequenz-spezifischen Bindungen über geeignete Absorptionsfilter zu detektieren. Die Anzahl an Oligonukleotidproben, die auf einem Array Platz finden, ist im Laufe der Jahre immer größer geworden. Heute können *tiling*-Microarrays mit ihren Proben ein Genom in verschiedener Auflösung und Dichte abdecken (Mockler *et al.*, 2005). Solche *tiling*-Microarrays sind auch bei Studien zur Identifizierung kleiner RNAs in anderen Bakterien zum Einsatz gekommen (Selinger *et al.*, 2000; Wassarman *et al.*, 2001; Tjaden *et al.*, 2002; Zhang *et al.*, 2003). Die nachfolgend dargestellte Einfarben-Microarrayanalyse in *C. glutamicum* beruht auf einer neuartigen Methode auf Basis einer flexiblen Plattform zur Oligonukleotid-Synthese und *tiling*-Microarrays, die speziell alle intergenischen Bereiche (zwischen den annotierten proteinkodierenden Genen) abdecken. In diesem Rahmen wurde außerdem in Zusammenarbeit mit S. Schroeder das Java-Applet IRMA (*Intergenic Regions by Microarray Analysis*) entwickelt (Schroeder, 2007), das zur Berechnung der intergenischen Oligonukleotidproben für die Microarrays dient. Die gesamte Methode und das in dieser Arbeit entwickelte Vorgehen zur Auswertung der Microarrays konnte mithilfe des Enterobakteriums *E. coli* K-12 MG1655 erfolgreich validiert werden (siehe Anhang). Die Vorgehensweise wurde anschließend auf *C. glutamicum* übertragen. Die folgenden Kapitel beschreiben einen Überblick über die Methode und die erhaltenen Ergebnisse der Experiment-spezifischen Hybridisierungen in *C. glutamicum*. Die präsentierten Ergebnisse stellen zugleich eine Art Validierung der Microarrayexperimente dar, da die gute Qualität und Reproduzierbarkeit der Ergebnisse deutlich werden. An dieser Stelle soll erwähnt werden, dass die Oligonukleotidsignale nicht dazu dienen sollen, kleine RNA-Kandidaten zu definieren. Die Ergebnisse sollen später helfen, die über Hochdurchsatz-Sequenzierung detektierten kleiner RNAs aus einem Probengemisch den einzelnen durchgeführten Experimenten und Stressen zuzuordnen. Darauf vorbereitend wird am Schluss dieses Kapitels eine Datenbank präsentiert, die neben den Experiment-spezifischen Oligonukleotidsignalen der Microarrays eine Verknüpfung zu den verschiedenen bioinformatischen Vorhersagen und allen intergenischen kleinen RNAs der cDNA-Sequenzierung beinhaltet, was eine Filterung nach all diesen Kriterien einzeln oder in Kombinationen ermöglicht.⁶

⁶ Die Microsoft Access-Datei Datenbank.accdb ist auf der beiliegenden CD zu finden.

2.1 Etablierung einer Methode zur intergenischen Transkriptanalyse in *C. glutamicum* basierend auf *tiling*-Microarrays

Kleine RNAs spielen oft eine besondere Rolle bei der Anpassung an veränderte Umgebungsbedingungen (Hoe *et al.*, 2013). Daher wurden bei der Probengewinnung nicht nur ungestresste *C. glutamicum*-Zellen in der exponentiellen Wachstumsphase geerntet, sondern zusätzlich auch fünf Kulturen, die zum Probenahme-Zeitpunkt Hitze- und Kälte-Stress, hyperosmotischem und oxidativem Stress ausgesetzt waren. Von allen sechs Kultivierungsproben wurde die Gesamt-RNA isoliert und für die Stress-spezifischen Microarrayexperimente wie folgt verfahren (Abbildung IV-3, nächste Seite): Die Gesamt-RNA Proben wurden nach Heidelberg zur febit biomed GmbH (jetzt CBC) verschickt und dort größenfraktioniert. Der Anteil aller kleinen RNAs bis ~250 nt Länge wurde ohne cDNA-Synthese direkt für den Hybridisierungsprozess mit dem genom@one System vorbereitet. Darauf folgten Experiment-spezifische Hybridisierungen auf speziellen intergenischen *tiling*-Microarrays, deren Design vorher über das Java-Applet IRMA (Schroeder, 2007) berechnet wurde. Diese Microarrays decken in beiden Strängen alle Bereiche zwischen den bekannten proteinkodierenden Regionen im *C. glutamicum*-Genom ab und enthalten zusätzlich tRNA-*sense*-Proben als Positiv-Kontrollen und deren reverse Komplemente (tRNA-*antisense*-Proben) als Negativ-Kontrollen, da im Gegenstrang von tRNAs keine Expression erwartet wird. Im Anschluss an die Experiment-spezifischen Hybridisierungen wurden die Emissionssignale der Microarrays erfasst und als Rohdaten von febit zur Verfügung gestellt. Nach Normalisierung und Median-Reskalierung der Daten wurden die spezifischen Signal-Schwellenwerte bestimmt, ab denen eine Expression in jedem einzelnen Experiment angenommen wird. Dieser Standard-Schwellenwert für transkriptionelle Aktivität wird über die Signale der Negativ-Kontrollen ermittelt, d.h. aus den Signalen der *antisense*-Sequenzen von tRNAs eines Microarrayexperimentes, wobei der maximale Wert dieser Negativ-Kontrollen die Grenze zum Hintergrundrauschen definiert.⁷

⁷ Die im Rahmen dieser Arbeit entwickelte Methode zur Auswertung der Microarrays wurde anschließend von S. Schroeder (Arbeitsgruppe BRF, CebiTec, Universität Bielefeld) in das Java-Applet IRMA integriert, so dass IRMA heute die komplette Methode zur Auswertung der Microarrayrohdaten zur Verfügung stellt.

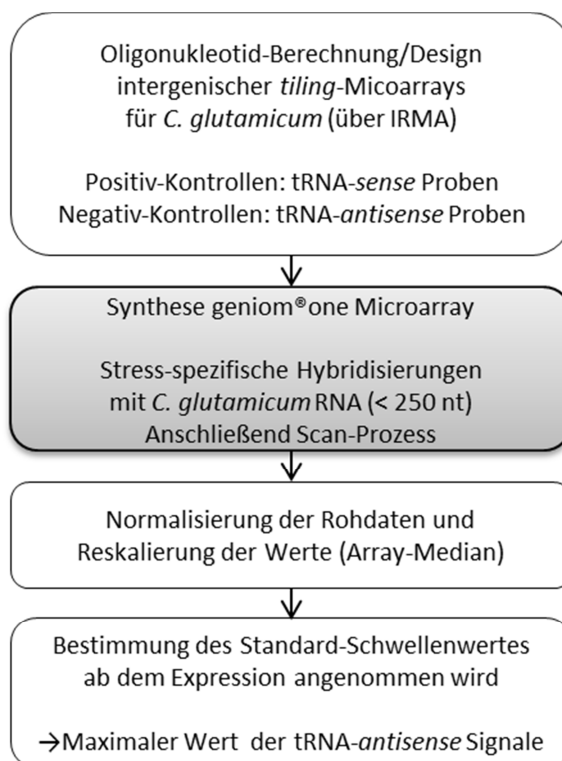


Abbildung IV-3 Arbeitsvorgang zur Detektion kleiner RNAs über intergenische *tiling*-Microarrays. Arbeitsschritte ohne Farbmarkierung wurden am CeBiTec Bielefeld durchgeführt und Grau-markierte Arbeitsschritte in Heidelberg (febit biomed GmbH, jetzt CBC). Alle Schritte der Auswertung wurden in dieser Arbeit entwickelt und von S. Schröder in das Java-Applet IRMA integriert.

2.2 Design eines intergenischen *tiling*-Microarrays für *C. glutamicum* über IRMA

Das Design der Oligonukleotidproben für die Synthese der genom[®]one Microarrays wurde mithilfe der Software IRMA erstellt. Hauptziel des Java-Applets ist die Implementierung eines Systems zur Analyse der intergenischen Regionen eines Genoms. IRMA berechnet Oligonukleotide mit gewünschter Länge und Überlappung aus den intergenischen Regionen eines Genoms in Anlehnung an *tiling*-Microarrays für Gesamt-Transkriptomanalysen. Die benötigten Daten hierfür, wie Größe und Sequenz der intergenischen Regionen, wird aus der Annotation der GenBank-Datei (Benson *et al.*, 2013) des zu untersuchenden Genoms gewonnen. Insgesamt wurden 10.521 50-mer Oligonukleotidproben mit 15-mer Überlappungen in den Regionen zwischen allen bekannten Kodierregionen für einen Strang des *C. glutamicum*-Genoms berechnet (Abbildung IV-4, nächste Seite). Von jeder Probe wurde zusätzlich das reverse Komplement ermittelt, so dass beide Stränge des Genoms berücksichtigt sind. Zusätzlich wurden als Positiv-Kontrollen die Startsequenzen 60 tRNAs für die Synthese von 50-mer Oligonukleotidproben berechnet, und zwar in *sense*-Richtung und dazu ebenfalls die reversen Komplemente als Negativ-Kontrollen (*antisense*-

Richtung). Hieraus ergeben sich Schmelztemperaturen der Oligonukleotide hauptsächlich im Intervall 64 bis 79 °C.⁸

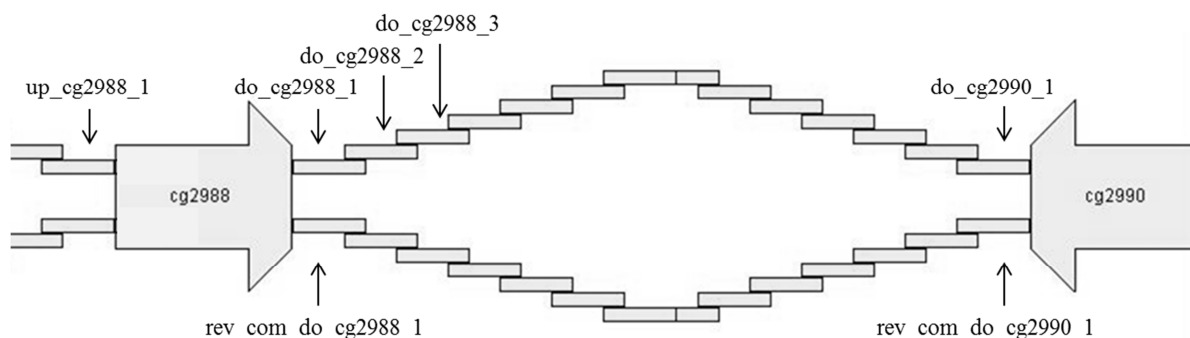


Abbildung IV-4 Aufbau des intergenischen *C. glutamicum* tiling-Microarray (Beispiel-Region).

Die Pfeile repräsentieren annotierte ORFs, dazwischen decken überlappende 50-mer Oligonukleotidproben alle intergenischen Bereiche des Genoms ab (15-mer Überlappung).

Die Identifikationsnummern der überlappenden Oligonukleotidproben folgen einer systematischen Benennung. Die ID „up_cg0002_x“ steht für eine Sequenz stromaufwärts von Gen *cg0002*, wogegen „do_cg0002_x“ für eine Sequenz stromabwärts von Gen *cg0002* steht. Das „x“ am Ende der ID gibt die Entfernung eines Oligonukleotids von der zugehörigen Kodierregion an. Ein hoher Wert „x“ findet sich nur zwischen Kodierregionen, die durch eine lange intergenische Region voneinander getrennt sind. Die IDs der revers-komplementären Sequenzen im Rückwärtsstrang heißen entsprechend, allerdings mit dem Namenszusatz „rev_com“. Da die Microarrays direkt mit markierter RNA hybridisiert wurden (ohne cDNA-Synthese), stellen die Signale der Vorwärtsstrang-Oligonukleotide Transkripte dar, die auf dem Rückwärtsstrang kodiert sind. Signale der Rückwärtsstrang-Oligonukleotide stellen Transkripte dar, die auf dem Vorwärtsstrang kodiert sind.

⁸ Die Datei IRMA-Oligonukleotide.xlsx im Ordner „Geniom-Microarrays“ auf der beiliegenden CD enthält alle 50-mer Sequenzen mit folgenden Zusatzinformationen: Identifikationsnummer (ID), Start- und Stopp-Positionen der 50-mer Sequenz, G+C-Gehalt, Schmelztemperatur, Name und Produkt und Orientierung der vorherigen und folgenden CDS.

2.3 Bestimmung der Schwellenwerte für transkriptionelle Aktivität mithilfe der Signale von Kontroll-Oligonukleotiden

Die erhaltenen Microarrayrohdaten wurden im ersten Schritt über die zwei Replikate jedes Experimentes nach Quantil (Bolstad *et al.*, 2003) normalisiert.⁹ Die für Microarrayexperimente verbreitete Lowess-Normalisierung (*locally-weighted scatterplot smoothing* (Cleveland und Devlin, 1988)) wurde für das Design der intergenischen *tiling*-Microarrays nicht angewendet, da nicht von einer Gleichverteilung der Signale über alle Intensitäten ausgegangen werden kann. Der Großteil der intergenischen Signale liegt an der Detektionsgrenze. Im Anschluss an die Normalisierung wurden die normalisierten Intensitäten im Intervall [0, 65535] reskaliert. Eine verbreitete Methode zur Reskalierung besteht darin, eine logarithmische Skalierung zu wählen. Hierdurch werden die vielen intergenischen Signale im niedrigen Intensitätsbereich allerdings sehr gestaucht. Daher wurden die Werte über Teilen des Gesamt-Median eines jeden Microarrays reskaliert, was zusätzlich einer Median-Zentrierung entspricht. Da der Array-Median in den Experimenten zwischen 500 und 1000 Einheiten liegt, ergibt sich nach der Reskalierung eine Signalspanne bis maximal ~ 130 ($65535/500$) Einheiten im Sättigungsbereich. Darauf folgend wurde die Verteilung der Signalstärken von insgesamt 720 Negativ-Kontrollen, die die *antisense*-Signale von tRNA-Transkripte darstellen, über alle Experimente als Box-Plot in einem Diagramm aufgetragen (Abbildung IV-5). Grundsätzlich eignen sich Box-Plots zur Visualisierung der Streuung von Datenmengen. Neben dem Median-Wert (0,5-Quantil), dem oberen Quartil und unteren Quartil (0,75- und 0,25-Quantil) der Datenverteilung werden auch die Extremwerte abgebildet. Im folgenden Box-Plot-Diagramm ist deutlich zu erkennen, dass die Signalwerte aller Negativ-Kontrollen sehr gering sind und auch die maximalen und minimalen Werte verteilt um ~ 1 liegen.

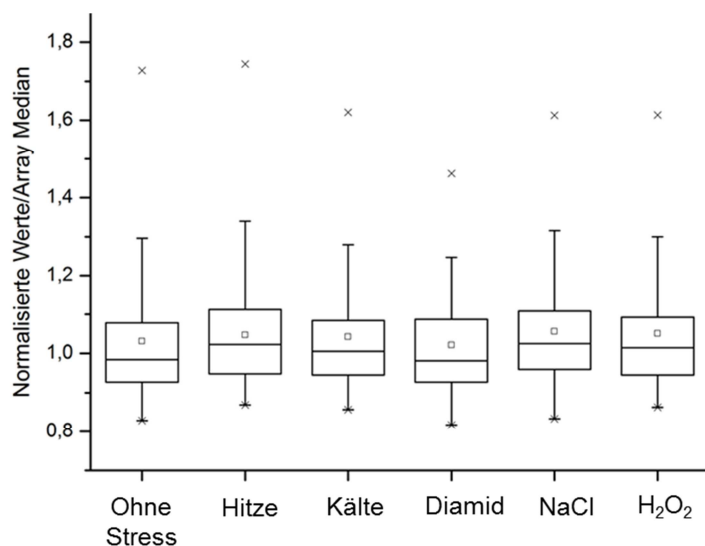


Abbildung IV-5 Verteilung der Signale aller Negativ-Kontrollen.

Überblick der 120 tRNA *antisense*-Signale je Experiment (Σ 720). Der Box-Bereich stellt das 0,25- bis 0,75-Quantil, das Kästchen in der Box stellt das arithmetische Mittel und die gerade Linie stellt den Median dar. Die *whisker* des Box-Plot-Diagramms repräsentieren Ausreißer. Maximum und Minimum der Datenverteilung werden als Kreuze visualisiert. Aus den maximalen Signalen der Negativ-Kontrollen wurden anschließend in jedem Experiment die spezifischen Standard-Schwellenwerte ermittelt, oberhalb derer ein Signal für transkriptionelle Aktivität angenommen wird.

⁹ Sowohl die Rohdaten der Oligonukleotidsignale als auch die Werte nach Normalisierung über alle Experimente befinden sich zusammengefasst in den Dateien Ergebnisse.xlsx und Ergebnisse_Kontrollen.xlsx im Ordner „Geniom-Microarrays“ auf der beiliegenden CD.

Im experimentellen Design ist es nicht vorgesehen, die verschiedenen Experimente aufeinander zu beziehen und eine davon als Referenz zu betrachten. Im Validierungsexperiment mit *E. coli* wurden gute Ergebnisse erzielt, wenn für die Signale oberhalb des maximalen Wertes aller tRNA-*antisense*-Proben transkriptionelle Aktivität angenommen wurde. In Anlehnung daran wurden auf diese Weise auch die Experiment-spezifischen Standard-Schwellenwerte im Microarrayexperiment mit *C. glutamicum* bestimmt (Tabelle IV-1).

Tabelle IV-1 Experiment-spezifische Standard-Schwellenwerte ab der transkriptionelle Aktivität angenommen wird.

Ohne Stress	Hitze-Stress	Kälte-Stress	Diamid-Stress	NaCl-Stress	H ₂ O ₂ -Stress
1,75	1,75	1,65	1,55	1,65	1,65

2.4 Analyse der Signalverteilung zur quantitativen Einteilung der Ergebnisse

Die Ergebnisse aller Microarrays aufzuführen wäre sehr umfangreich, daher werden hier zur Validierung nur die Signale im Experiment ohne Stress ausführlicher dargelegt. Die beschriebenen Beobachtungen sind aber mit den Ergebnissen der übrigen Experimente vergleichbar. In der folgenden Abbildung IV-6 (nächste Seite) ist die Verteilung der intergenischen Signale im Wachstumsexperiment ohne Stress für beide Replikate, allerdings ohne Kontrollen, dargestellt. Zur einfachen Übersicht werden die normalisierten Werte an dieser Stelle in der häufig verwendeten logarithmischen Skalierung eines *Scatter-Plots* visualisiert. Dabei ist zu erkennen, dass die Mehrheit der Werte im niedrigen Intensitätsbereichen liegt und dass sich die Werte mit guter Reproduzierbarkeit entlang der Winkelhalbierenden verteilen. Das stärkste Signal liegt im Bereich der Sättigung ($\log_2 64816,83 = 15,98$) und stammt von dem Oligonukleotid mit der ID *rev_com_up_cg0294_4* aus der genomischen Region, in der das Haushaltsgen 4.5S RNA kodiert ist. Die Signalwerte im blau-markierten Kasten entsprechen dem Wertebereich unterhalb des Standard-Schwellenwertes ab dem transkriptionelle Aktivität angenommen wird. Die Gesamtmenge der Signale oberhalb dieses Standard-Schwellenwertes beläuft sich auf 634. Die höchsten 10 % dieser Signale (63 von 634) liegen oberhalb eines Wertes, der dem vierfachen Standard-Schwellenwert entspricht (als grün-gestrichelte Linie dargestellt). Die Grenze zu den höchsten 25 % der Signalwerte (158 von 634) ist ebenfalls in Abbildung IV-6 markiert und ist als blau-gestrichelte Linie dargestellt. Dieser Anteil ergibt sich oberhalb von Signalen, die dem doppelten Standard-Schwellenwert entsprechen.

Nach Analyse der Signalwerte der übrigen Hybridisierungen wurde eine ähnliche Verteilung beobachtet wie beim Experiment ohne Stress, so dass definiert wurde: Bei Erreichen des doppelten Standard-Schwellenwertes für transkriptionelle Aktivität (obere ~25 % des Signalanteils) wird ein „deutliches Transkriptsignal“ angenommen. Bei Erreichen des vierfachen Standard-Schwellenwertes (obere ~10 % des Signalanteils) wird ein „sehr deutliches Transkriptsignal“ festgelegt. Die daraus resultierende quantitative Einteilung der Signale konnte in experimentellen Validierungen kleiner RNAs über Northern Blot-Analysen (Kapitel IV.4.1) größtenteils bestätigt werden.

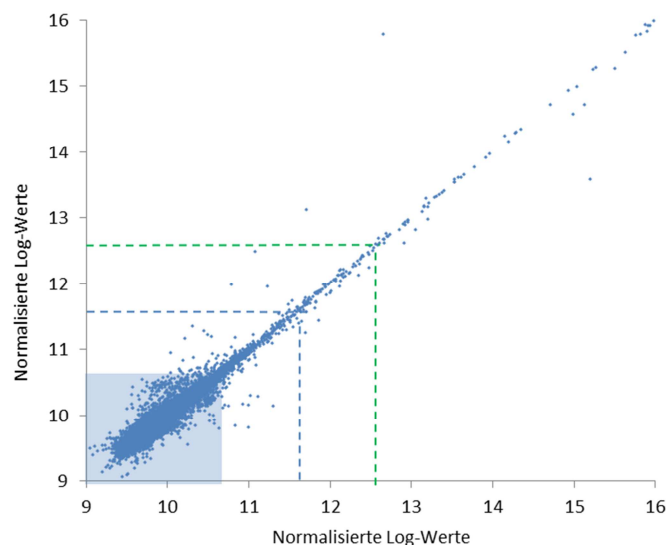


Abbildung IV-6 Überblick über die normalisierten Signale im Experiment ohne Stress.

Im Diagramm wurde eine logarithmische Skalierung der Werte gewählt. Der blau-markierte Kasten entspricht Signalen, die unterhalb des Standard-Schwellenwertes für transkriptionelle Aktivität liegen und das Hintergrundrauschen darstellen. Die gestrichelten Linien repräsentieren die Grenzen der oberen 25 % bzw. 10 % der Signale. Diese entsprechen dem doppelten (blau-gestrichelt) bzw. vierfachen Standard-Schwellenwert (grün-gestrichelt).

Um einen möglichst vollständigen Blick auf die Daten zu zeigen, werden nun auch die Ergebnisse der Positiv-Kontrollen (*sense*-Signale der tRNAs) im Experiment ohne Stress dargelegt. Die Werte der Positiv-Kontrollen liegen alle im Bereich sehr deutlicher Transkriptsignale, da jeweils der vierfache Standard-Schwellenwert ($4 \times 1,75$) überschritten wird. Allerdings ist eine Streuung der Werte zu beobachten, und zwar zwischen 8,05 für Glycin-tRNA (ID *rev_com_tRNA_3587_Start*) bis 74,51 für Cystein-tRNA (ID *tRNA_3564_Start*; Anticodon GCA). Gleichzeitig wird eine hohe Reproduzierbarkeit über die Replikate mit nur geringen Standardabweichungen (arithmetisches Mittel 0,77) deutlich. Es fällt auf, dass das geringste Signal von der Glycin-tRNA (Anticodon CCC) stammt, deren komplementäre Sequenz auf der DNA dem geringsten anteiligen Codongebrauch von 1 % in *C. glutamicum* zugeordnet werden kann (Liu *et al.*, 2010). Diese Auffälligkeit stellt jedoch eine Ausnahme dar, da sich aus dem Vergleich der Signalstärken der übrigen tRNAs und dem jeweils zugehörigen anteiligen Codongebrauch

keine Proportionalität ergeben hat. Im Rahmen dieser Analyse wurden jeweils die Signalwerte über tRNAs mit demselben Anticodon addiert (Daten nicht gezeigt).

2.5 Erstellung einer Datenbank für regulatorische RNAs in *C. glutamicum*

Im Laufe der Experimente und bioinformatischen Analysen wurden die produzierten Daten immer komplexer und zur besseren Übersicht in eine Datenbank eingepflegt (Abbildung IV-7). Diese Datenbank enthält die normalisierten und reskalierten Microarraysignale aller Experimente, und über flexible Filterprozesse können schnell vergleichende Abfragen unter Verwendung der Standard- oder alternativer Schwellenwerte in den spezifischen Microarrayexperimenten erstellt werden. Zusätzlich können verschiedene Informationen der zugehörigen Oligonukleotidproben abgerufen werden, sowie die Signal-Standardabweichung über die Replikate, Position im Genom, Namen und Orientierung und Funktionen der umliegenden Gene und einige weitere Details. Darüber hinaus wurde über die Positionen aller Oligonukleotide eine Verknüpfung zu den Vorhersagen der Rfam-Datenbank (Gardner *et al.*, 2009), der SIPHT-Analyse und der NAPP-Analyse erstellt. Zudem sind Verknüpfungen zur RNAz-Vorhersage regulatorischer RNAs und zur Liste von Rho-unabhängigen Terminatoren in *C. glutamicum* in der Datenbank enthalten. Außerdem sind die in dieser Arbeit aufgeführten Ergebnisse der cDNA-Sequenzierung kleiner RNAs integriert.

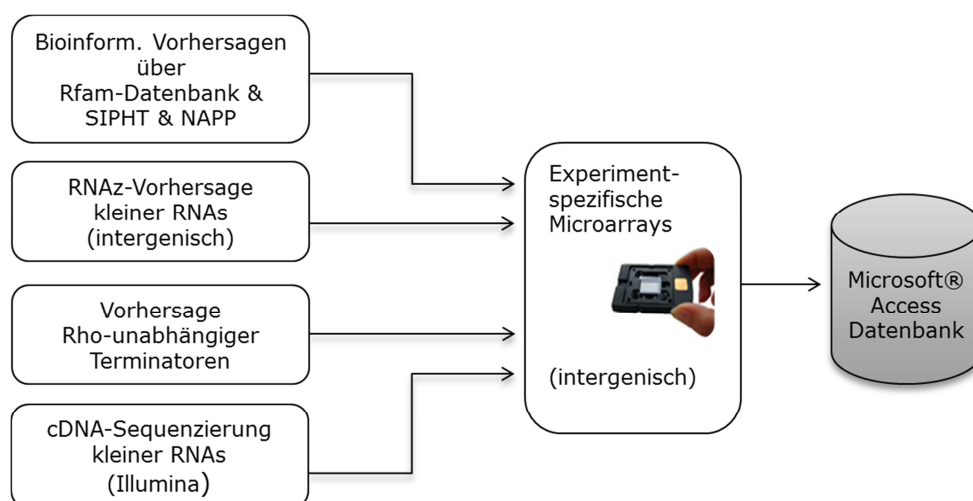


Abbildung IV-7 Datenbankstruktur.

In der Datenbank sind alle Signale der Oligonukleotidproben und weitere Details enthalten. Zusätzlich wurden die Signale mit den Informationen der Rfam-Datenbank (Gardner *et al.*, 2009) und den bioinformatischen Vorhersagen für kleine regulatorische RNAs und stromabwärts kodierten Rho-unabhängigen Terminatoren verknüpft. Außerdem können Vergleiche mit den Ergebnissen der cDNA-Sequenzierung kleiner RNAs vollzogen werden.

Über die sechs Microarrayexperimente zusammengefasst zeigen ca. 10 % (2143 von 21042) aller intergenischen Oligonukleotide ein Signal oberhalb des Standard-Schwellenwertes. Es ergibt sich, dass ~9 % (195 von 2143) dieser Signale mit RNAz-Kandidaten überlappen und ~15 % (317 von 2143) in intergenischen Bereichen mit phylogenetischer Konservierung nach NAPP-Analyse (Ott *et al.*, 2012) liegen. Insgesamt treffen ~21 % (442 von 2143) dieser Signale Regionen, in denen auch über die cDNA-Sequenzierung kleine RNAs detektiert wurden. Die folgende Datenbank-Abfrage stellt eine Gesamtübersicht über die Signale in den einzelnen Microarrayexperimenten dar (Tabelle IV-2).

Tabelle IV-2 Oligonukleotidsignale über alle Experimente (nach Standard-Schwellenwert)

	Ohne Stress	Hitze-Stress	Kälte-Stress	Diamid-Stress	NaCl-Stress	H ₂ O ₂ -Stress
Signale gesamt	634	842	1151	1225	1024	1483
Signale Experiment-spezifisch	93 (15 %)	47 (6 %)	97 (8 %)	139 (11 %)	59 (6 %)	256 (17 %)
Signale auch in anderen Experimenten	541 (85 %)	404 (48 %)	576 (50 %)	627 (51 %)	450 (44 %)	702 (47 %)
Signale auch im Experiment ohne Stress	-	391 (46%)	487 (42 %)	459 (37 %)	515 (50 %)	525 (35 %)

Der Tabelle ist zu entnehmen, dass die Anzahl von Signalen am geringsten im Experiment ohne Stress (634) und am größten nach H₂O₂-Stress (1483) ist. Insgesamt ist über die Experimente ein hoher Prozentsatz von Signalen vertreten (zwischen 35 % und 50 %), die auch ohne Stress detektiert wurden. Der Experiment-spezifische Signalanteil ist eher gering und liegt in den Experimenten zwischen 6 % und 17 %.

3 Identifizierung kleiner RNAs mittels Hochdurchsatz-Sequenzierung in

C. glutamicum

Die Hochdurchsatzsequenzierung von RNA-Proben (RNA-Seq) bzw. von den aus diesen Proben generierten cDNA-Bibliotheken stellt seit einiger Zeit eine sehr effiziente Methode zur Bestimmung kleiner RNA-Elemente in Bakterien und anderen Organismen dar (Güell *et al.*, 2011). Hierbei wird eine Nukleotid-genaue Auflösung der kleinen RNAs erreicht. Über spezielle Methoden in der Probenvorbereitung kann eine genaue Bestimmung der Transkriptionsstarts ermöglicht werden (Sharma *et al.*, 2010). In den folgenden Kapiteln werden die Ergebnisse der strangspezifischen cDNA-Sequenzierung zur Identifizierung kleiner RNAs in *C. glutamicum* präsentiert. Die erhaltenen kleinen RNAs wurden hierbei in die drei Klassen „mRNA-Leader“, „*trans*-kodierte RNAs“ und „*antisense*-Transkripte“ klassifiziert. Zusätzlich wurden über die Sequenzierungs-Daten kleine putative Proteine identifiziert. Im Rahmen dieses Kapitels werden außerdem die Klassen „mRNA-Leader“ und „*antisense*-Transkripte“ analysiert. Die Klasse „*trans*-kodierte RNAs“ wird erst im Kapitel 4 der Ergebnisse zur Validierung kleiner RNAs genauer dargelegt und dabei mit den vorangegangenen bioinformatischen Analysen und Experiment-spezifischen Microarrayanalysen verglichen.

3.1 Erzeugung von zwei verschiedenen cDNA-Bibliotheken zur Hochdurchsatz-Sequenzierung kleiner RNAs und Kartierung der erhaltenen *reads* auf der *C. glutamicum* Genomsequenz

Für die Hochdurchsatz-Sequenzierung kleiner RNAs in *C. glutamicum* wurden wie im Microarrayscreening die beschriebenen *C. glutamicum*-Kulturen verwendet, die während der exponentiellen Wachstumsphase geerntet und zusätzlich auch verschiedenen Stressbedingungen ausgesetzt waren. Im Gegensatz zu den Stress-spezifisch durchgeführten Microarrayexperimenten wurde nach Isolierung der Gesamt-RNA ein Gemisch aus gleichen Teilen jeder RNA-Probe gebildet (Pool). Aus diesem Gesamt-RNA-Pool wurden alle kleinen RNAs < 250 nt zunächst über Präzipitation fraktioniert (Abbildung IV-8, nächste Seite). Nach Abreicherung der ribosomalen RNAs über eine Hybridisierungs-Prozedur mit rRNA-bindenden magnetischen *beads*, wurde die kleine RNA-Fraktion in zwei gleichgroße Ansätze aufgeteilt zur Erstellung von zwei verschiedenen cDNA Bibliotheken mit dem Illumina® TruSeq™ Small RNA Kit. Dabei wurde nach dem dRNA-Seq-Ansatz (Sharma *et al.*, 2010) vorgegangen. Dies bedeutet konkret, dass ein Ansatz in Vorbereitung für cDNA-Bibliothek 1 enzymatisch auf Primärtranskripte angereichert wurde und zwar über die Behandlung mit einer 5'-Monophosphat-spezifischen Exonuklease, die speziell degradierte oder prozessierte Transkripte abbaut. Nicht-prozessierte primäre RNA-Transkripte weisen an ihren 5'-Ende einen Triphosphat-Rest auf und bleiben

von diesem enzymatischen Verdau unbeeinflusst. Die RNA für cDNA-Bibliothek 2 wurde dagegen nicht enzymatisch behandelt und enthält alle kleinen RNAs-Spezies.¹⁰

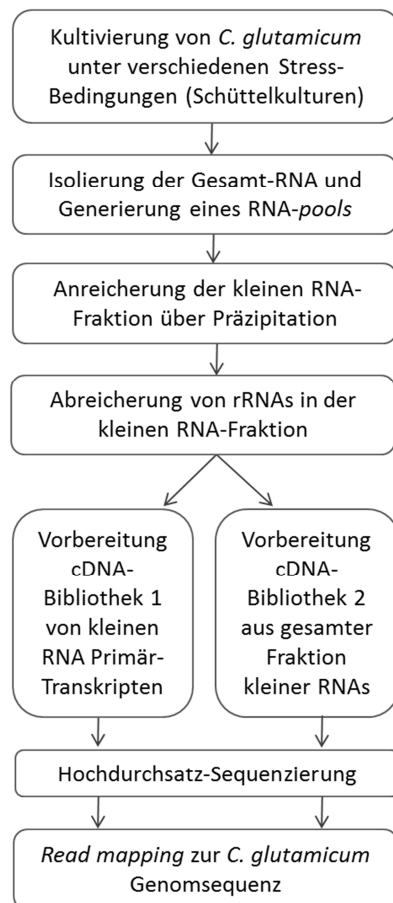


Abbildung IV-8 Schritte zur Erzeugung der RNA-Proben für die Erstellung von zwei cDNA-Bibliotheken.

Die Sequenzierung der beiden cDNA-Bibliotheken mittels dem Illumina GA IIx-System wurde mit einer Leselänge von 35 bp durchgeführt, die ausgehend vom 5'-Ende aus gelesen wurden.¹¹ Die erhaltenen *reads* wurden anschließend auf der Chromosomensequenz von *C. glutamicum* ATCC 13032 (Kalinowski *et al.*, 2003) kartiert (*read mapping*) und ergaben für die Primärtranskript-angereicherte Probe 7.869.859 eindeutig zuzuordnende *reads* (Bibliothek 1; Tabelle IV-3, nächste Seite) und 22.752.379 eindeutig zuzuordnende *reads* für die Gesamtfraction kleiner RNAs (Bibliothek 2; Tabelle IV-3). Für die meist reichlich abundanten *reads* im Bereich von rRNA-Genen konnte besonders in der enzymbehandelten cDNA-Bibliothek 1 ein hoher Grad an Abreicherung erreicht werden. Von der Gesamtzahl der *reads* in cDNA-Bibliothek 1 sind nur 3 % rRNA-Genen zuzuordnen. Außerdem wurde in Bibliothek 1 im Vergleich zur Bibliothek 2 ein kleinerer Anteil an *reads* in *sense*-Richtung von Kodierregionen (CDS) erhalten. Dies lässt vermuten, dass die Anzahl degradiertes oder bereits prozessierter größerer mRNA-Transkripte durch die Enzymbehandlung verringert werden konnte. Die restliche Anzahl von *reads* konnte entweder *cis-antisense* zu Kodierregionen zugeordnet werden oder zu Regionen, in denen es bisher keine Annotation gab. Die *cis-antisense reads* zeigen mit bis zu 3 % einen ähnlich hohen Anteil in beiden Bibliotheken. Interessanterweise gehört jeweils der größte Anteil von *reads* zu Regionen ohne bisherige Annotationen. In der Primärtranskript-angereicherten Bibliothek 1 machen diese sogar 59 % der

¹⁰ Ein Überblick über die Agilent Bioanalyzer Daten zu Beginn der Arbeiten (sRNA-Pool) und der beiden fertigen cDNA-Bibliotheken steht auf der Daten-CD zur Verfügung (Ordner „Sequenzierung“, Datei Bioanalyzer.pptx).

¹¹ Die Ergebnisse sind im Verzeichnis vol/rna-illumina/Run010 auf dem CeBiTec-Server gelistet als Probe S6 bzw. Name X1 (= nicht behandelt) und Probe S6 bzw. Name T1 (= Terminator Exonuklease behandelt).

Gesamt-*reads* aus, was das Vorhandensein einer großen Anzahl potenzieller kleiner RNAs in *C. glutamicum* vermuten lässt.¹²

Tabelle IV-3 *Read*-Übersicht über die Sequenzierungs-Ergebnisse der zwei cDNA-Bibliotheken

Kategorie	Primärtranskripte ^a kleiner RNAs (Bibliothek 1)		Gesamtfraction kleiner RNAs (Bibliothek 2)	
	<i>Reads</i>	[%]	<i>Reads</i>	[%]
rRNA-Gene	255.591	3,2	5.230.021	23,0
tRNA-Gene	1.050.962	13,4	2.130.360	9,3
mRNA-Gene (CDS, <i>sense</i>)	1.686.575	21,4	7.134.344	31,4
Übrige Verteilung von <i>reads</i>				
<i>cis-antisense</i> (CDS, <i>antisense</i>)	242.429	3,1	450.589	2,0
Regionen ohne Annotation	4.634.302	58,9	7.807.065	34,3
Gesamt	7.869.859		22.752.379	

^a Primärtranskripte kleiner RNAs wurden erhalten über eine 5'-Monophosphat-spezifische Exonuklease, die speziell Transkripte verdaut, die entweder prozessiert wurden oder gerade degradiert werden

3.2 Filterungsprozesse zur Identifizierung von potenziellen kleinen RNA-Genen

In den Sequenzierungs-Daten der Primärtranskript-angereicherten cDNA-Bibliothek 1 wird aufgrund des experimentellen Vorgehens jeweils an der 5'-Position eines *read*-Stapels ein TSS vermutet, wobei ein *read*-Stapel eine bestimmte Anzahl an einer Position startenden *reads* darstellt (Abbildung IV-9A, nächste Seite). Zur Berechnung der putativen TSS in der cDNA-Bibliothek 1 wurde nach manueller Analyse eine Mindestanzahl von 20 *read*-Starts an einer Position als sinnvoller Schwellenwert angenommen.¹³ Vor Beginn der weiteren Analysen wurden die Daten gefiltert, wobei alle Transkriptionsstarts im Bereich von tRNA- und rRNA-Genen und innerhalb annotierter proteinkodierender Sequenzen (CDS, *sense*-Richtung) entfernt wurden (Abbildung IV-9B). Dies führte zu einer Anzahl von 2899 Stapel (1304 Stapel *cis-antisense* zu CDS und 1595 Stapel in Regionen ohne Annotation). Ein weiterer Filterschritt bestand darin, die Sequenzen stromaufwärts der vermuteten TSS nach Promotoren zu analysieren (Abbildung IV-9C). Mit dem Tool Improbizer (Ao *et al.*, 2004; Tompa *et al.*, 2005) wurde in den Sequenzen nach Ähnlichkeiten zu den Konsensussequenzen des primären Sigmafaktor SigA (Pátek und Nešvera, 2011) oder des Stress-verwandten ECF-Familien-Sigmafaktors

¹² Alle detaillierten Daten der Verteilung der *reads*, auch umgerechnet in RPKM-Werte (*reads per kilo base per million*) sind auf der beiliegenden CD im Ordner „Sequenzierung“ in Datei Expression.xlsx zu finden.

¹³ Die Gesamttabelle der daraus ermittelten TSS in cDNA-Bibliothek 1 ist auf der CD im Ordner „Sequenzierung“ in der Datei Transkriptionsstarts-Rohdaten.xlsx zu entnehmen.

SigH (Busche *et al.*, 2012) gesucht, wobei SigH eine bedeutende Rolle unter oxidativem Stress (Kim *et al.*, 2005) und unter Hitzestress-Bedingungen spielt (Ehira *et al.*, 2009). Dabei wurden 1267 putative TSS gefunden, die stromaufwärts eine SigA-erkennbare Promotorsequenz tragen (531 *cis-antisense* zu CDS und 736 in Regionen ohne Annotation). Die Suche nach SigH-Promotersequenzen war stromaufwärts von 44 TSS erfolgreich (11 *cis-antisense* zu CDS und 33 in Regionen ohne Annotation). Für 14 Transkripte, wurden Promotoren beider Typen detektiert, was auf Transkriptvarianten mit unterschiedlichen 5'-Enden hindeutet. Um die 3'-Enden der potenziellen kleinen RNAs zu bestimmen wurden die *reads* beider cDNA-Bibliotheken analysiert. Jedes startende Transkript wurde bis zu einer Position verlängert, an der die Anzahl von *reads* unter einen gesetzten Wert von 10 fiel, und diese Position wurde als 3'-Ende definiert (Abbildung IV-9A).¹⁴

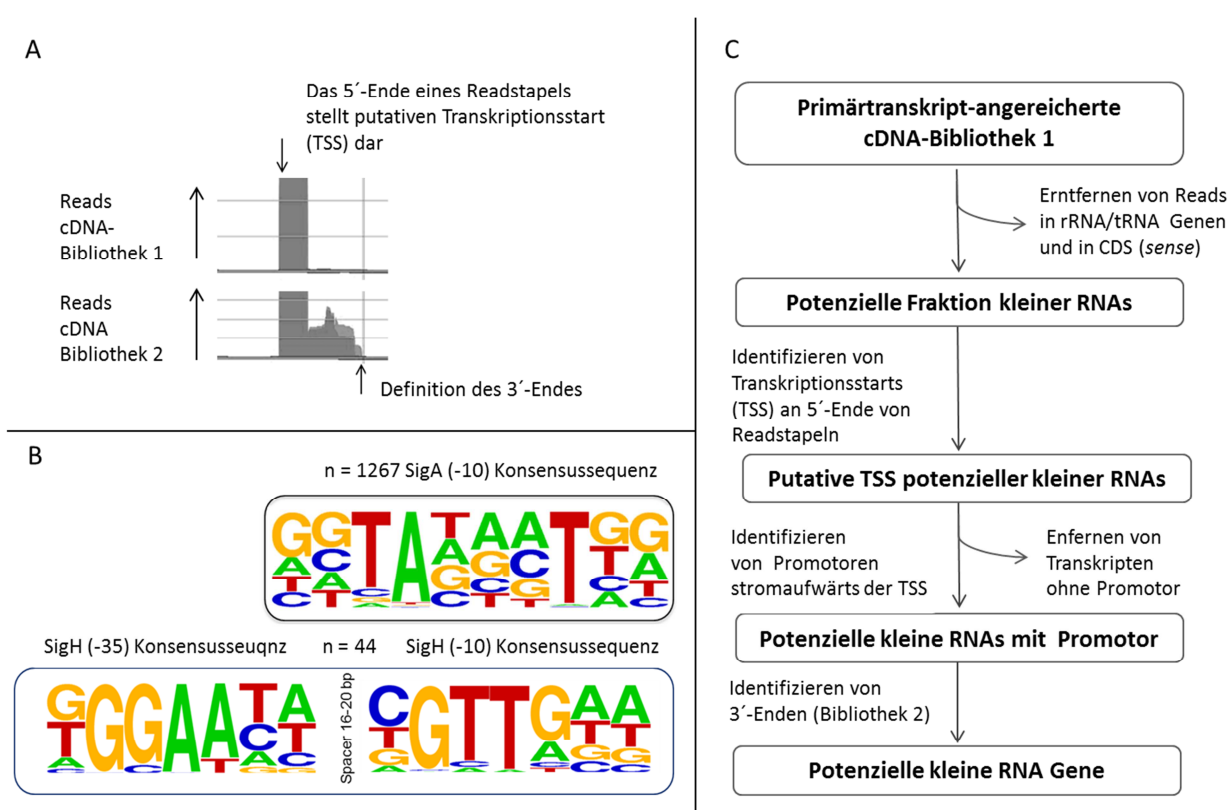


Abbildung IV-9 Definition von Start- und Stopp-Positionen potenzieller kleiner RNA-Gene.

A cDNA-Bibliothek 1 (Primärtranskripte) wurde zur Bestimmung der Transkriptionsstarts (TSS) an den 5'-Enden von *read*-Stapeln verwendet. **B** Workflow für die Charakterisierung potenzieller kleiner RNA-Gene mithilfe von bioinformatischer Promotoranalyse stromaufwärts der putativen TSS in cDNA-Bibliothek 1 (Primärtranskripte). Das 3'-Ende der potenziellen kleinen RNAs wurde mithilfe von cDNA-Bibliothek 2 bestimmt. **C** Weblogo (Crooks *et al.*, 2004) Präsentation der Konsensussequenzen der -35- und -10-Regionen von SigA- und SigH-abhängigen Promotoren, die mit dem Improbizer Tool (Ao *et al.*, 2004; Tompa *et al.*, 2005) detektiert wurden. Das prozentuale Auftreten eines Nukleotids an einer bestimmten Position ist durch die Größe des Nukleotidsymbols (A, C, G, T) repräsentiert.

¹⁴ Die Daten hierzu befinden sich im Ordner Sequenzierung in Datei Definition_3'-Enden.xlsx

Zusätzlich wurde um die ermittelten 3'-Enden herum innerhalb von 60 nt nach Rho-unabhängigen Terminatoren gesucht, da diese häufig die Transkripte kleiner bakterieller RNAs terminieren (Livny *et al.*, 2008). Hierfür wurden die bereits bestimmten Rho-unabhängige Terminatoren in *C. glutamicum* eingesetzt, und der Vergleich lieferte 69 potenziellen kleinen RNAs Rho-unabhängige Terminatoren in diesem Bereich (4 *cis-antisense* zu CDS und 65 in Regionen ohne Annotation). In 136 Fällen folgten putative Transkriptionsstarts dicht aufeinander, was auf multiple Promotoren hindeutet. Daher wurden alle angenommenen multiplen TSS innerhalb von 100 bp vom äußersten 5'-Ende aus und mit demselben 3'-Ende zu einer Region zusammengezogen und als solche annotiert.¹⁵

3.3 Klassifizierung von kleinen RNA-Genen in Abhängigkeit von ihrer Position zu den benachbarten CDS

Um die noch im Set enthaltenen putativen untranslatierten Regionen (UTRs) von mRNAs zu identifizieren, wurden alle potenziellen kleinen RNAs entsprechend ihrer Position und Richtung im Vergleich zu den umliegenden CDS gruppiert. Insgesamt wurden 298 Transkripte als "mRNA Leader" gekennzeichnet, auf deren TSS in weniger als 100 nt Distanz stromabwärts im gleichen Strang eine CDS folgt.¹⁶ Beim Vergleich mit der Rfam-Datenbank (Gardner *et al.*, 2009) konnten zusätzlich das gut konservierte *mraW*-Motiv (bei *cg2377*), das *cspA*-Motiv (bei *cg0215*) und neun vorhergesagte Riboswitches Genen zugeordnet werden, deren 5'-UTRs damit länger als 100 nt sind (Tabelle IV-4).

Tabelle IV-4 Liste der detektierten und Rfam-vorhergesagten *cis*-regulatorische Motive in *C. glutamicum*.

Rfam Nr.	Rfam Name	Gen	Rfam score	Rfam Start	Sequenzierung Start	Rfam Stopp	Sequenzierung Stopp
RF00050	FMN-Ribosw.	<i>cg0083</i>	111,76	66442	66438	66279	66406
RF01766	<i>cspA</i>	<i>cg0215</i>	60,19	186399	186327	186766	186500
RF00059	TPP-Ribosw.	<i>cg0825</i>	63,54	742654	742651	742547	742535
RF00059	TPP-Ribosw.	<i>cg1227</i>	51,03	1127774	1127765	1127883	1127860
RF00059	TPP-Ribosw.	<i>cg1476</i>	55,87	1373213	1373210	1373103	1373154
RF00634	SAM-IV-Ribosw.	<i>cg1478</i>	70,47	1374007	1374005	1374123	1374114
RF00059	TPP-Ribosw.	<i>cg1655</i>	52,1	1544490	1544485	1544383	1544419
RF00080	yybP-ykoY	<i>cg1660</i>	43,71	1550030	1550101	1550196	1550136
RF00080	yybP-ykoY	<i>cg2157</i>	49,13	2043157	2043151	2042989	2043038
RF00059	TPP-Ribosw.	<i>cg2236</i>	62,55	2120271	2120271	2120383	2120303
RF01746	<i>mraW</i>	<i>cg2377</i>	56,64	2267021	2266969	2266916	2266877

¹⁵ Eine Übersicht über alle potenziellen kleinen RNA Gene nach Filterung ist in der Datei *Potenzielle_kleine_RNA_Gene.xlsx* auf der CD im Ordner „Sequenzierung“ zu finden und stellt Informationen zur Sequenzierung (z.B. *read count*), alle zugehörigen Promotorsequenzen und Rho-unabhängigen Terminator-Positionen zur Verfügung.

¹⁶ Die Datei *Klasse_mRNA-Leader.xlsx* befindet sich auf der CD im Ordner Sequenzierung.

Hierbei konnten wir die Rfam-vorhergesagten Thiaminpyrophosphat (TPP)-abhängigen Riboswitches stromaufwärts von Genen des Thiamin-Biosyntheseweges *thiC* (*cg1476*), *thiM* (*cg1655*) und *thiE* (*cg2236*) validieren. Zwei TPP-Riboswitches wurden stromaufwärts von *cg0825* (putative β -Ketoacyl Carrier-Protein-Reduktase) und *cg1227* (putatives Membranprotein) gefunden. Ein Flavinmononukleotid (FMN)-abhängiger Riboswitch wurde bei der putativen Nikotinamid-Mononukleotidaufnahme-Permease (*cg0083*) identifiziert und ein S-Adenosylmethionin (SAM)-abhängiger Typ IV-Riboswitch stromaufwärts von *cg1478* (annotiert als hypothetisches Protein). Zusammen mit zwei Riboswitch-ähnlichen RNA-Motiven (beide *yybP-ykoY*) (Barrick *et al.*, 2004) und sieben weiteren Transkripten, die putative kleine Proteine kodieren, wurden insgesamt 316 Regionen der Klasse “mRNA Leader” zugeordnet.

Die Klasse “*antisense*-Transkripte” enthält 543 Regionen und besteht aus drei Untergruppen (i) *cis-antisense* RNAs (asRNA), die innerhalb einer gegenüberliegenden CDS starten, (ii) Transkripte, die in *antisense*-Richtung zu einer 5'-UTR liegen und innerhalb von 100 nt des 5'-Endes dieser CDS starten (as5'-UTR) und (iii) Transkripte, die in *antisense*-Richtung zu einer 3'-UTR liegen und innerhalb von 60 nt des 3'-Endes dieser CDS starten (as3'-UTR).¹⁷ Auf diese Weise wurden 464 Regionen als asRNAs, 63 als as5'-UTRs, und 16 als as3'-UTRs eingeteilt (Abbildung IV-10, nächste Seite). Aufgrund spezieller CDS-Anordnungen wurden 8 as5'-UTRs auch als as3'-UTRs und 48 mRNA Leader auch als as5'-UTRs gezählt. Alle übrigen 262 Regionen wurden als Klasse „*trans*-kodierte kleine RNAs“ definiert.¹⁸ Hierzu zählen auch die RNA-Haushaltsgene tmRNA (*cgb_09183*), M1 RNA (*cgb_24535*) und 4.5S RNA (*cgb_02933*). Zur Vorbereitung auf eine Aktualisierung der *C. glutamicum* Genomannotation wurden alle *antisense*-Transkripte und *trans*-kodierte RNAs mit neuen Locus-Tags versehen. In der neuen Nomenklatur werden die Locus-Tags jeder CDS mit einer nachfolgenden Null erweitert (z.B. alt: *cg0001*; neu: *cgb_00010*). Die letzte Ziffer der Nummer wird dazu benutzt um neue Gene zwischen die alten einzufügen.

¹⁷ Die Datei Klasse_antisense-Transkripte.xlsx befindet sich auf der CD im Ordner „Sequenzierung“.

¹⁸ Die Datei Klasse_trans-kodierte_RNAs.xlsx befindet sich auf der CD im Ordner „Sequenzierung“.

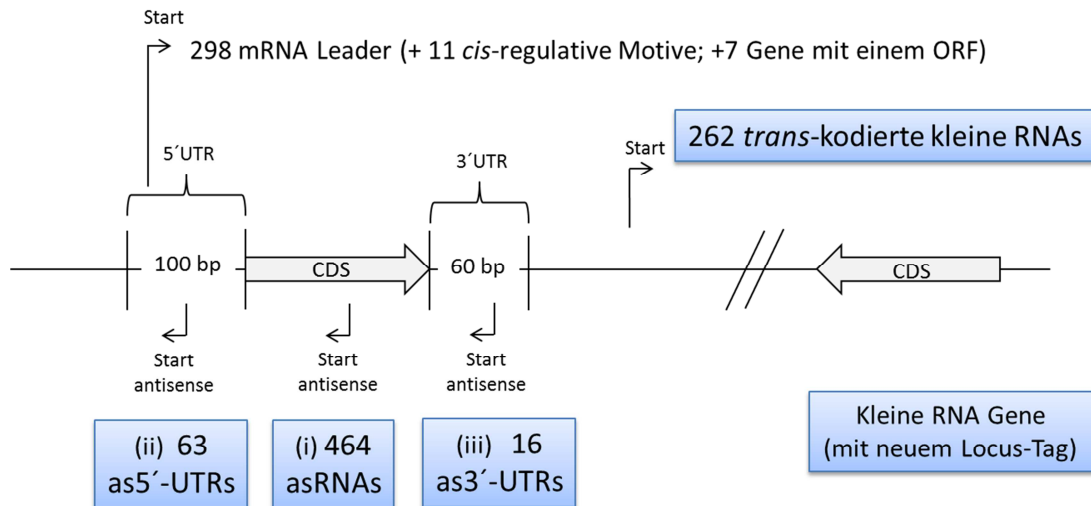


Abbildung IV-10 Klassifizierung der kleinen RNAs in Abhängigkeit von ihrer Position zu den benachbarten CDS. Alle neuen Gene der Klassen „*antisense*-Transkripte“ und „*trans*-kodierte RNAs“ werden einem neuen Locus-Tag zugeordnet.

3.4 Suche nach putativen kleinen Proteinen

Einige der detektierten kleinen RNAs könnten auch mRNAs repräsentieren, also kleine Peptide kodieren. Daher wurde im ersten Schritt das Tool „ORFfinder“ eingesetzt, um aus den Sequenzen aller kleinen RNA-Gene ORFs im multiplen FastA-Format zu extrahieren. Anschließend wurde nach ribosomalen Bindestellen (RBSs) stromaufwärts der extrahierten ORF Sequenzen mit dem RBSfinder (Suzek *et al.*, 2001) mit einer Fenstergröße von 15 bp und den Standard RBS Einstellungen gesucht. Dabei wurden nur ORFs mit einer minimalen Länge von 15 Aminosäureresten in Betracht gezogen, die entweder eine RBS besitzen oder mRNAs ohne Leader darstellen (*leaderless transcripts*). Insgesamt wurden acht kleine mRNAs erhalten (Tabelle IV-5, nächste Seite). Als nächstes wurde nach Konservierung dieser Peptidsequenzen in anderen bakteriellen Genomen mit dem tblastx-Algorithmus des NCBI Web Portals (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) gesucht. Dabei wurde eine Konservierung in sieben Fällen gefunden.¹⁹ Die ähnlichen kleinen Proteine (E-Werte $< 10^{-4}$) wurden zumeist in anderen *Corynebacterium*-Spezies, aber auch außerhalb von *Corynebakterien* detektiert. Am weitesten verbreitet scheint der gut konservierte tmRNA-Peptid-Tag zu sein (Peptidsequenz AEKSQRDYALAA in *C. glutamicum*). Dieser *C. glutamicum* tmRNA-Peptid-Tag korrespondiert gut mit den bekannten Sequenzen eines breiten phylogenetischen Spektrums (Zwieb *et al.*, 1999). Neben diesem, wurde nur das durch *cgb_08775* (*cg4014*) kodierte Peptid außerhalb der *Corynebacterium*-Spezies als konserviert gefunden und war auch als Candidate_15_NC_006958 über das SIPHT-Tool (Livny *et al.*, 2008) vorhergesagt. Das kleinste konservierte Peptid *cgb_14345* (*cg4016*) wurde mit 15 Aminosäuren Länge

¹⁹ Die Datei Liste_putative-Peptide.xlsx befindet sich auf der CD im Ordner „Sequenzierung“.

kodiert gefunden und ist bereits bekannt als Valin-reiches Leaderpeptid vor dem *ilvBNC*-Operon (Morbach *et al.*, 2000). Weitere Leaderpeptide möglicher Attenuatorstrukturen wurden bei *cgb_33575* (*cg4012*) vor *trpE* (*cg3359*) des Tryptophan-Operons identifiziert, bei *cgb_03035* (*cg4015*) vor dem *leuA* Gen (*cg0303*), das als Isopropylmalat-Synthase den ersten Schritt in der Leucin-Biosynthese darstellt, und vor *aroF* (*cg1129*), das als DAHP-Synthase den ersten Schritt in der Synthese aromatischer Aminosäuren darstellt. Die vermutete Funktionalität der genannten Peptide wird unterstützt durch das Auftreten von drei aufeinanderfolgenden Tryptophanresten im putativen Leaderpeptid stromaufwärts des Tryptophan-Operons, aufeinanderfolgenden Codons verzweigter Aminosäuren im vermuteten *leuA* Leaderpeptid (2x Isoleucin-, 3x Valin- und 2x Leucin-Codons) und den drei Aminosäureresten Phenylalanin-Tyrosin-Phenylalanin im Fall des *aroF* Leaderpeptids. Interessant ist, dass die mRNAs aller angenommenen Fälle transkriptioneller Attenuation *leaderless*-Transkripte sind. Eine Konservierung der Sekundärstruktur mittels RNAz wurde für alle putativen Leaderpeptid-mRNAs bis auf *cgb_33575* vor *trpE* und für *cgb_08775* erhalten.

Tabelle IV-5 Putative Peptide

Seq.- Start	Seq.- Stop	Strang	ORF Länge [nt]	ORF Start	ORF Stopp	RBS	Neuer Locus- Tag	Bisher Locus- Tag	Möglicher Attenuator für
268137	268081	-	57	268137	268081	<i>Leaderless</i>	<i>cgb_03035</i>	<i>cg4015</i>	<i>cg0303 (leuA)</i>
806926	806809	-	75	806893	806819	AGGAG	<i>cgb_08775</i>	<i>cg4014</i>	
848500	848922	+	39	848642	848680	-	<i>cgb_09185</i>	<i>tmRNA</i>	
1047891	1048044	+	78	1047891	1047968	<i>Leaderless</i>	<i>cgb_11283</i>	<i>cg4017</i>	<i>cg1129</i>
1339309	1339356	+	48	1339309	1339356	<i>Leaderless</i>	<i>cgb_14345</i>	<i>cg4016</i>	<i>cg1435 (ilvB)</i>
1367696	1368633	+	93	1367818	1367910	AGGAG	<i>cgb_14697</i>	<i>cg4019</i>	
2399351	2399146	-	108	2399280	2399173	AAGAG	<i>cgb_25195</i>	<i>cg4018</i>	
3206459	3206519	+	54	3206459	3206512	<i>Leaderless</i>	<i>cgb_33575</i>	<i>cg4012</i>	<i>cg3359 (trpE)</i>

3.5 Betrachtung von *cis-antisense* RNA-Genen und ihre Verteilung auf funktionelle CDS-Klassen

Mehr als die Hälfte aller kleinen RNA-Regionen (543 von 807) fallen in die Klasse “*antisense*-Transkripte” (Abbildung IV-10, CD-Datei Klasse_*antisense*-Transkripte.xlsx). Für die Unterklasse von asRNAs (464), die direkt gegenüber einer CDS lokalisiert sind, wurde eine mittlere Größe von 55 nt berechnet. Diese sehr kleine Größe insbesondere für asRNAs im Vergleich zu den Transkriptgrößen in den anderen Klassen wird in einem Box-Plot-Diagramm deutlich (Abbildung IV-11, nächste Seite). Für das Set der *trans*-kodierten RNAs wurde dagegen eine mittlere Größe von 90 nt erhalten und eine nicht symmetrische Verteilung in Richtung größerer Transkripte.

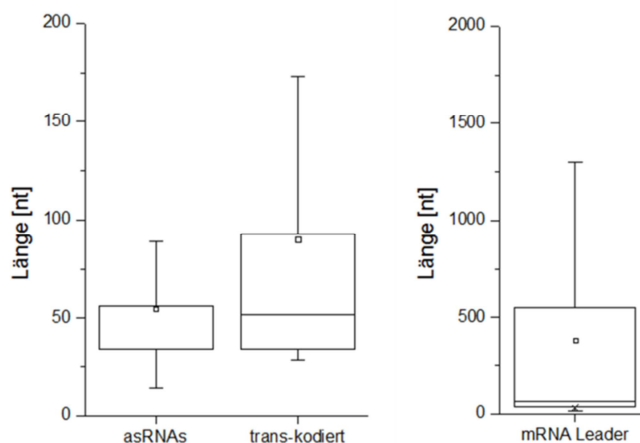


Abbildung IV-11 Vergleich der Längenverteilung für asRNAs, *trans*-kodierte RNAs und mRNA Leader.

Die Box Plots zeigen das arithmetische Mittel (kleines Quadrat) und den Median (gerade Linie) für die unterschiedlichen Klassen. Das untere Ende und das obere Ende der Boxen stellen jeweils das 0,25- und 0,75-Quantil dar und die *whisker* zeigen die Ausreißer der Verteilung.

Alle asRNAs verteilen sich auf 409 unterschiedliche CDS, und davon existieren 44 CDS, die mehr als ein *antisense*-Transkript besitzen. Es ist anzumerken, dass Transkription in *antisense*-Richtung nicht nur an den 5'-Enden der zugehörigen CDS zu beobachten ist, sondern oft auch am 3'-Ende oder in der Mitte der entsprechenden Kodierregion. Um die Funktionen von Genen mit kodierten asRNAs zu vergleichen, wurden diese klassifiziert nach dem eggNOG-Klassifikationssystem (*evolutionary genealogy of genes: Non-supervised Orthologous Groups*) (Jensen *et al.*, 2008). Im Fall von 264 von 464 asRNAs sind die entsprechenden CDS über die eggNOG-Klassifikation gut charakterisiert (Abbildung IV-12, nächste Seite). Für die übrigen 200 asRNAs wurden entweder keine Kategorie oder nur Kategorien mit schlechter Charakterisierung erhalten. Eine normale Verteilung der 464 asRNAs über alle aktuell annotierten CDS in *C. glutamicum* würde wohl darin resultieren, dass ~15 % der Mitglieder einer eggNOG-Klasse eine asRNA enthalten. Daher wurden eggNOG-Klassen, in denen mehr als 20 % oder weniger als 10 % der Mitglieder eine asRNA enthalten als über- bzw. unterrepräsentiert bezogen auf *antisense*-Transkription betrachtet. Der geringste Anteil von asRNAs wurde für Gene in den Klassen Zellteilung ("D", 0 %), Coenzym-Transport und Metabolismus ("H", 1 %), Transkription ("K", 7 %) und anorganischer Transport und Metabolismus ("P", 9 %) gefunden. Andererseits wurde ein höherer Anteil von asRNAs in der funktionellen Klasse Zellhüllen-Biogenese ("M", 35 %) beobachtet. Außerdem zeigte sich ein höherer Anteil von asRNAs für die Klasse von Proteinen, die in Sekretionsprozesse und Vesikeltransport involviert sind ("U", 32 %) (Abbildung IV-12). In der letztgenannten funktionellen Klasse sind die beiden bekannten Mechanismen der Proteinsekretion mit asRNAs repräsentiert und zwar *secY* (*cg0647*), das für eine Preprotein-Translokase-Untereinheit des Sec-Sekretionssystems kodiert, und *tatC* (*cg1684*), das für eine Twin-Arginin (Tat) Translokase-Untereinheit kodiert. Die *tatC antisense* RNA (*cgb_16835*) wird

offenbar von einem SigA-abhängigen Promoter reguliert, wobei die *secY* asRNA (*cgb_06475*) wahrscheinlich von einem SigH-kontrollierten Promotor transkribiert wird.

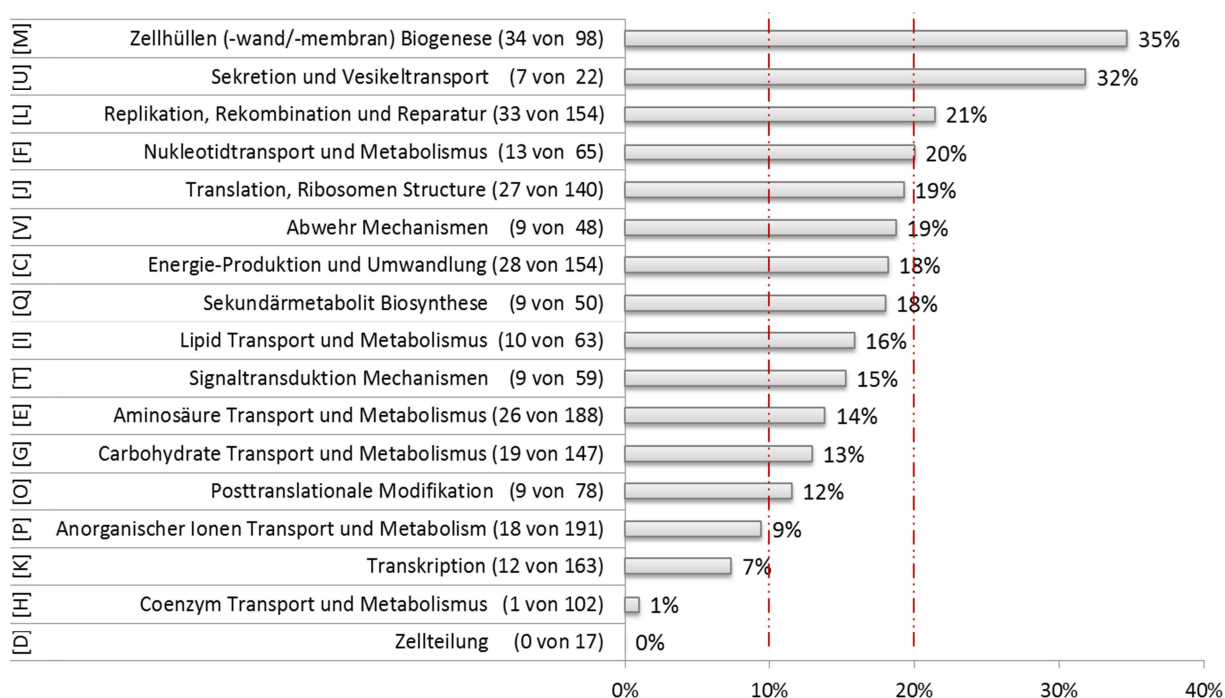


Abbildung IV-12 Funktionelle Klassifikation von *C. glutamicum* Genen mit kodierten *cis-antisense* RNAs (asRNAs). Für 264 von 464 asRNAs konnten die entsprechenden kodierenden Gene eggNOG Klassen (Jensen *et al.*, 2008) zugeordnet werden. Es werden die Häufigkeiten und Verteilung von asRNAs über alle eggNOG Klassen gezeigt. Klassen mit *cis-antisense* RNA Anteilen zwischen 10 % und 20 % stehen für eine normale Verteilung. Die roten gestrichelten Linien stellen die Grenzwerte dar, ab denen eggNOG-Klassen mit asRNAs unterrepräsentiert (< 10 %) oder überrepräsentiert (> 20 %) sind.

In diesem Zusammenhang wurde auch der Anteil von Genen mit asRNAs berechnet, die für eine Transmembranhelix oder für ein Signalpeptid zur Sekretion kodieren. Über eine bioinformatische Suche wurden 165 von 464 asRNAs (~35 %) für solche entsprechenden Gene detektiert. Statistisch gesehen ist dieser Anteil nicht signifikant unterschiedlich von einer Normalverteilung, da in *C. glutamicum* 974 Gene entweder für ein Signalpeptid zur Sekretion (304 Gene) oder für mindestens eine Transmembranhelix (670 Gene) kodieren, was ~ 32 % der aktuell annotierten Gene entspricht.

4 Validierung, Charakterisierung und funktionelle Analysen kleiner RNAs in *C. glutamicum*

Die Identifizierung von kleinen RNAs in *C. glutamicum* (Mentz *et al.*, 2013) stellt erst den Beginn der Analyse von RNA-Regulationen in diesem Bakterium dar. In den folgenden Jahren gilt es nun die Transkriptgrößen und Experiment-spezifische transkriptionelle Aktivität der neu definierten kleinen RNAs im Detail zu validieren, Sekundärstrukturen zu analysieren und natürlich die entsprechende(n) Funktion(en) zu bestimmen. Die über diese Arbeit erhaltenen Daten der bioinformatischen RNAz-Vorhersage und Experiment-spezifischen Microarraydetektion intergenischer kleiner RNAs liefern in diesem Zusammenhang eine großen Menge von Informationen und können einen Ausgangspunkt für funktionelle Analysen von *trans*-kodierten RNAs darstellen. Im Rahmen dieser Arbeit wurden die Transkripte von vier *trans*-kodierten RNAs mit einer hohen Anzahl erhaltener *reads* mittels Northern Blot-Analyse untersucht. Für eine dieser vier RNAs aus der Klasse „*trans*-kodierte RNAs“ (*cgb_16115*) werden anschließend Ergebnisse präsentiert, die eindeutige Hinweise auf eine mögliche Ziel-mRNA bzw. eine putative Funktion liefern. Zudem haben die Ergebnisse dieser Arbeit zur Identifizierung der kleinen RNA *cgb_15805* aus der Klasse „*cis*-*antisense* RNAs“ geführt, die gegenüber dem 5'-Ende von *argC* (*cg1580*) liegt und in Anlehnung daran *asaC* (*antisense argC*) benannt wurde. Analysen mit dieser kleinen RNA haben ergeben, dass ihr Wirkmechanismus transkriptionelle Interferenz ist und die Transkription von *argC* negativ beeinflusst (Petri *et al.*, Manuskript in Vorbereitung) (Daten nicht gezeigt). Im letzten Kapitel wird außerdem aus der Klasse „mRNA-Leader“ die peptidkodierende RNA *cgb_11283* stromaufwärts von *aroF* (*cg1129*) vorgestellt, über die ein neuer Mechanismus transkriptioneller Attenuation in der Synthese aromatischer Aminosäuren validiert werden konnte (Mentz *et al.*, Manuskript in Vorbereitung).

4.1 Analyse von *trans*-kodierten RNA-Genen in Bezug auf Sekundärstrukturkonservierung und ihre differenzielle Expression

Um die Klasse „*trans*-kodierte RNAs“ genauer zu charakterisieren, wurde das Set dieser 262 RNAs mit der bioinformatischen RNAz-Vorhersage und den Ergebnissen der intergenischen *tiling*-Microarrays verglichen. Insgesamt gibt es 47 *trans*-kodierte RNAs, die sowohl eine RNAz-Vorhersage besitzen als auch in den Microarrayexperimenten detektiert wurden. Dazu gehören alle drei RNA-Haushaltsgene tmRNA (*cgb_09183*), M1 RNA (*cgb_24535*) und 4.5S RNA (*cgb_02933*). Die Abbildung IV-13 (nächste Seite) kartiert einen Längenvergleich der RNA-Haushaltsgene nach allen in dieser Arbeit durchgeführten Analysen zur Identifizierung kleiner RNAs in *C. glutamicum* im Vergleich zu den Rfam-Vorhersagen (Gardner *et al.*, 2009). Es wird eine hohe Ähnlichkeit der Ergebnisse über die verschiedenen Bestimmungsmethoden deutlich, insbesondere bei Betrachtung der Ergebnisse der Microarrays (grün) und der cDNA-Sequenzierung (rot).

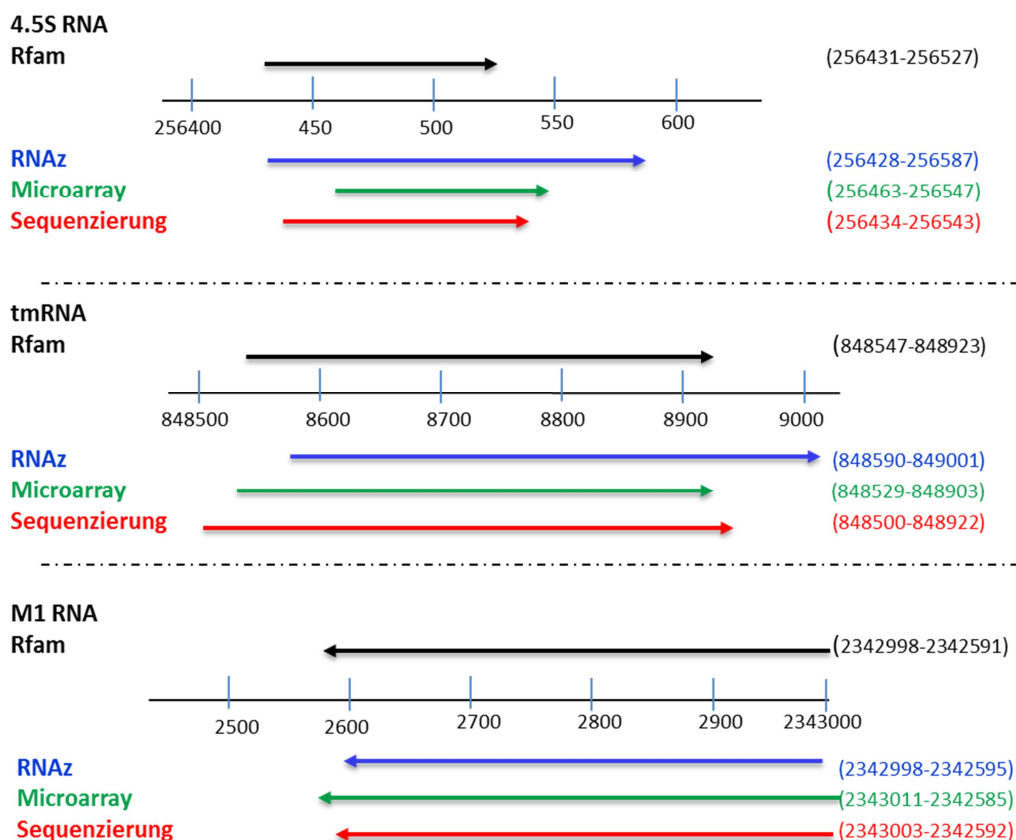


Abbildung IV-13 RNA-Haushaltsgene in *C. glutamicum* im Längenvergleich über alle Methoden.

Es sind alle in dieser Arbeit durchgeführten Methoden zur Detektion kleiner RNAs in *C. glutamicum* im Vergleich zur Rfam-Datenbank Vorhersage (Gardner *et al.*, 2009) gezeigt.

Der Gesamtvergleich der RNAz-Vorhersage mit den 262 *trans*-kodierte RNAs lieferte in der Summe 77 *trans*-kodierte RNAs (~ 30 %), die eine konservierte Sekundärstruktur aufzeigen (bei RNAz $p \geq 0.5$). Ein ähnlicher Wert von 83 Übereinstimmungen wurde über den Abgleich mit Bereichen phylogenetischer Konservierung über die NAPP-Pipeline (Ott *et al.*, 2012) bestimmt. Insgesamt 45 *trans*-kodierte RNAs wurden unter sehr stringenten Bedingungen (RNAz $p \geq 0.9$) erhalten, und davon sind 18 gleichzeitig auch in den NAPP-*contigs* vertreten (Tabelle IV-6, nächste Seite).

Tabelle IV-6 *Trans*-kodierte kleine RNA-Gene mit überlappender RNAz-Vorhersage ($p \geq 0.9$) und Vergleich mit NAPP-Daten für *C. glutamicum* (Ott *et al.*, 2012).

Neuer Locus Tag	Strang	Sequenzierung Start	Sequenzierung Stopp	Benachbarte Gene	RNAz Start	RNAz Stopp	RNAz max. <i>p</i> -score	NAPP
<i>cgb_00105</i>	-	10053	9921	<i>cg0010(-)/cg0012(-)</i>	10073	9921	0.95	✓
<i>cgb_00925</i>	+	74286	74320	<i>cg0092(+)/cg0095(+)</i>	74297	74476	0.99	
<i>cgb_03505</i>	-	307582	307548	<i>cg0350(-)/cg0352(-)</i>	307558	307474	0.95	✓
^a <i>cgb_03605</i>	+	314679	314787	<i>cg0360(-)/cg0362(+)</i>	314611	314792	0.99	✓
<i>cgb_03995</i>	-	346945	346882	<i>cg0399(-)/cg0400(-)</i>	346922	346590	0.96	✓
<i>cgb_05085</i>	+	452359	452408	<i>cg0508(-)/cg0510(+)</i>	452321	452622	0.99	✓
<i>cgb_05716</i>	+	509744	509990	<i>cg0571(+)/cg0572(+)</i>	509724	509981	0.99	✓
<i>cgb_05756</i>	+	512702	512814	<i>cg0575(+)/cg0576(+)</i>	512744	512906	0.99	✓
<i>cgb_08496</i>	+	782757	782836	<i>cg0849(+)/cg0850(+)</i>	782647	782889	0.99	
<i>cgb_08785</i>	-	807467	807331	<i>cg0878(-)/cg0879(+)</i>	807563	807292	0.92	✓
<i>cgb_09095</i>	-	842812	842715	<i>cg0909(-)/cg0910(-)</i>	842945	842791	0.97	
<i>cgb_09097</i>	-	842983	842911	<i>cg0909(-)/cg0910(-)</i>	842945	842791	0.97	
^b <i>cgb_09185</i>	+	848500	848922	<i>cg0918(+)/cg0919(+)</i>	848444	848993	0.90	✓
<i>cgb_09483</i>	+	878863	878996	<i>cg0948(-)/cg0949(+)</i>	878852	879125	0.99	✓
<i>cgb_13305</i>	-	1237440	1237333	<i>cg1330(+)/cg1332(-)</i>	1237507	1237208	0.90	
<i>cgb_14495</i>	+	1351975	1352041	<i>cg1449(+)/cg1451(+)</i>	1351833	1352246	0.99	
<i>cgb_17355</i>	-	1626662	1626596	<i>cg1735(-)/cg1736(-)</i>	1626885	1626583	0.95	✓
<i>cgb_17735</i>	+	1665705	1665791	<i>cg1773(-)/cg1774(+)</i>	1665657	1665835	0.99	
<i>cgb_17805</i>	-	1672717	1672673	<i>cg1780(+)/cg1781(-)</i>	1672744	1672565	0.92	
<i>cgb_18405</i>	-	1734383	1734304	<i>cg1840(-)/cg1841(-)</i>	1734440	1734264	0.99	✓
<i>cgb_18415</i>	-	1736390	1736347	<i>cg1841(-)/cg1842(+)</i>	1736478	1736333	0.96	
<i>cgb_21516</i>	-	2039580	2039466	<i>cg2151(-)/cg2152(-)</i>	2039656	2039493	0.98	
<i>cgb_21673</i>	-	2055867	2055764	<i>cg2167(-)/cg2168(-)</i>	2055929	2055750	0.99	✓
<i>cgb_22185</i>	-	2108839	2108800	<i>cg2218(-)/cg2221(-)</i>	2109027	2108748	0.98	
<i>cgb_22215</i>	-	2110108	2109924	<i>cg2221(-)/cg2222(-)</i>	2110102	2109913	0.99	✓
<i>cgb_22285</i>	-	2116294	2116236	<i>cg2228(-)/cg2229(-)</i>	2116297	2115998	0.96	
<i>cgb_22405</i>	-	2124418	2124384	<i>cg2240(+)/cg2241(-)</i>	2124504	2124285	0.92	
<i>cgb_23783</i>	-	2267593	2267546	<i>cg2378(-)/cg2380(-)</i>	2267720	2267551	0.93	✓
<i>cgb_24455</i>	-	2331195	2331116	<i>cg2445(-)/cg2446(-)</i>	2331257	2331109	0.99	✓
^c <i>cgb_24535</i>	-	2343003	2342592	<i>cg2453(-)/cg2455(-)</i>	2343050	2342650	0.95	
<i>cgb_24775</i>	-	2362678	2362640	<i>cg2477(-)/cg2478(-)</i>	2362704	2362548	0.94	✓
<i>cgb_25636</i>	+	2447380	2447441	<i>cg2563(+)/cg2564(-)</i>	2447231	2447490	0.94	
<i>cgb_25955</i>	-	2476453	2476419	<i>cg2595(-)/cg2597(-)</i>	2476484	2476295	0.92	✓
<i>cgb_26475</i>	-	2530082	2530004	<i>cg2647(-)/cg2648(+)</i>	2530150	2529972	0.91	
<i>cgb_28315</i>	+	2693243	2693292	<i>cg2831(-)/cg2833(+)</i>	2692958	2693349	0.99	
<i>cgb_28685</i>	-	2730160	2730126	<i>cg2868(+)/cg2869(-)</i>	2730432	2730065	0.97	
<i>cgb_29606</i>	-	2816731	2816535	<i>cg2960(+)/cg2962(-)</i>	2816711	2816532	0.99	
<i>cgb_30116</i>	-	2863994	2863960	<i>cg3011(-)/cg3012(-)</i>	2864045	2863846	0.95	
<i>cgb_30685</i>	-	2928726	2928634	<i>cg3068(-)/cg3069(-)</i>	2928804	2928585	0.99	
<i>cgb_31375</i>	-	2997519	2997485	<i>cg3137(-)/cg3138(+)</i>	2997679	2997500	0.94	
<i>cgb_31785</i>	-	3041252	3041154	<i>cg3178(-)/cg3179(-)</i>	3041427	3041124	0.97	
<i>cgb_33045</i>	-	3156331	3156287	<i>cg3304(-)/cg3306(-)</i>	3156621	3156254	0.91	
<i>cgb_33325</i>	+	3179928	3180066	<i>cg3332(-)/cg3334(+)</i>	3179728	3180003	0.99	
<i>cgb_34325</i>	-	3282122	3282086	<i>cg3432(-)/cg3434(+)</i>	3282124	3281877	0.95	

^a RF01066; 6C^b RF00023; tmRNA^c RF00010; RNaseP

Ein Microarraysignal nach Standard-Schwellenwert konnte für 130 *trans*-kodierte RNAs detektiert werden (~ 50 %).²⁰ Mithilfe der in dieser Arbeit erstellten Datenbank wurden 32 *trans*-kodierte RNAs erhalten, die mindestens in einem der Microarrayexperimente einen vierfachen Standard-Schwellenwert erreichen, also ein sehr deutliches Transkriptsignal ergeben haben, und sind in Form eines Signalmusters dargestellt (Abbildung IV-14, nächste Seite). Mit Ausnahme der M1 RNA zeigen die anderen RNA-Haushaltsgene in allen Experimenten sehr deutliche Transkriptsignale. Der Großteil der übrigen *trans*-kodierten RNAs wurde über die Experimente mit unterschiedlichen Signalen detektiert, was in der Abbildung über verschiedene Farben visualisiert wird.

Außerdem wurde verglichen, ob die sieben *trans*-kodierten RNAs, für die SigH-abhängige Promotoren detektiert wurden, auch über die Microarrays im Zusammenhang mit oxidativen Stressbedingungen (Diamid oder H₂O₂) oder bei Temperatur-Stress mit deutlichen oder sehr deutlichen Transkriptsignalen beobachtet wurden. Bis auf *cgb_11253* konnte dieser Zusammenhang für die übrigen sechs RNAs gezeigt werden (Tabelle IV-7). Wie schon oben beschrieben, ist *cgb_16115* zusätzlich auch nach hyperosmotischem Stress mittels NaCl mit einem sehr deutlichen Transkriptsignal vertreten.

Tabelle IV-7 *Trans*-kodierte RNAs mit SigH-/ECF-ähnlichem Promotor

<i>Trans</i> -kodierte RNA (SigH-/ECF-ähnlicher Promotor)	Mindestens deutliches Transkriptsignal/ doppelter Standard-Schwellenwert überschritten
<i>cgb_06075</i>	Hitze-, Kälte-, Diamid-, H ₂ O ₂ -Stress
<i>cgb_01363</i> und <i>cgb_31005</i>	Diamid-, Kälte-, H ₂ O ₂ -Stress
<i>cgb_16115</i>	Diamid-, NaCl-, H ₂ O ₂ -Stress
<i>cgb_20916</i>	Diamid-, H ₂ O ₂ -Stress
<i>cgb_31375</i>	Diamid-Stress
<i>cgb_11253</i>	-

²⁰ Alle Übereinstimmungen mit RNAz, NAPP und Microarrays wurden in der Datei *Trans-kodierte_RNAs.xlsx* entsprechend gekennzeichnet (CD-Ordner „Sequenzierung“).

Neuer Locus-Tag	Ohne Stress	Hitze	Kälte	Diamid	NaCl	H ₂ O ₂	RFAM Name
<i>cgb_00105</i>							
<i>cgb_01363</i>							
<i>cgb_02675</i>							
<i>cgb_02933</i>							4.5S RNA; Bacteria_small_SRP
<i>cgb_03605</i>							6C
<i>cgb_05835</i>							
<i>cgb_09185</i>							tmRNA
<i>cgb_09366</i>							
<i>cgb_09483</i>							
<i>cgb_11283</i>							
<i>cgb_12905</i>							
<i>cgb_14697</i>							
<i>cgb_15626</i>							
<i>cgb_16115</i>							
<i>cgb_17805</i>							
<i>cgb_18945</i>							
<i>cgb_19385</i>							
<i>cgb_20417</i>							
<i>cgb_20715</i>							
<i>cgb_21945</i>							
<i>cgb_22535</i>							
<i>cgb_23366</i>							
<i>cgb_23395</i>							
<i>cgb_23485</i>							
<i>cgb_23783</i>							
<i>cgb_24535</i>							M1 RNA
<i>cgb_24616</i>							
<i>cgb_24645</i>							
<i>cgb_24775</i>							
<i>cgb_28127</i>							
<i>cgb_30116</i>							
<i>cgb_31005</i>							

Abbildung IV-14 Expressionsmuster über 32 *trans*-kodierte RNAs mit sehr deutlichem Transkriptsignal. (weiß = kein Signal/ unterhalb des Standard-Schwellenwertes; gelb = Transkriptsignal/ Standard-Schwellenwert erreicht; rosa = deutliches Transkriptsignal/ doppelter Standard-Schwellenwert erreicht; orange = sehr deutliches Transkriptsignal/ vierfacher Standard-Schwellenwert erreicht).

4.2 Spezifische Transkriptanalyse am Beispiel von vier *trans*-kodierten kleinen RNAs

Zur Validierung der Sequenzierungsergebnisse wurden vier *trans*-kodierte RNAs ausgewählt und anschließend über Northern Blot-Experimente analysiert und auch mit den entsprechenden Stress-spezifischen Microarraysignalen verglichen (Abbildungen siehe nächste Seiten). Es sind die aus vielen *Actinomycetales* Gattungen bekannte 6C RNA (*cgb_03605*) (Weinberg *et al.*, 2010) (Abbildung IV-15a) und drei andere RNAs mit einer hoher Anzahl von *reads* (> 1000) an ihrem Transkriptionsstart. Hierfür wurde *cgb_00105* als Beispiel für eine kleine RNA mit Vorhersage über RNAz (und NAPP) ausgesucht (Abbildung IV-15b). Als Beispiele ohne Sekundärstrukturkonservierung wurden *cgb_20715* innerhalb und *cgb_16115* außerhalb der Prophagen-Region CGP3 ausgewählt (Abbildung IV-15c und Abbildung IV-16). Die zugehörigen RNA-Sekundärstrukturen ergaben über eine RNashapes-Analyse (Steffen *et al.*, 2006) unter MFE-Bedingungen jeweils hohe Faltungswahrscheinlichkeiten (*shape-probabilities*) von ~90 %. Um Stress-spezifische Signalmuster zu erhalten, wurden in den Northern Blot-Analysen jeweils gleiche Mengen isolierter Gesamt-RNA aus einem Experiment pro Spur aufgetragen. Hierbei korreliert jeweils die Größe der obersten Bande einer kleinen RNA im Northern Blot mit der Größe, die über Sequenzierung bestimmt wurde (~100 nt für 6C RNA, ~130 nt für *cgb_00105*, ~95 nt und ~70 nt für *cgb_20715*, ~140 nt und ~95 nt für *cgb_16115*).

Im Vergleich zur Rfam-Vorhersage (Gardner *et al.*, 2009) der 6C RNA mit Start-Position 314705 zeigt diese in der Sequenzierung am 5'-Ende eine 27 nt Erweiterung (Sequenzierung Start-Position 314679). Die erweiterte 5'-Region scheint zumindest in den Corynebakterien-Spezies *C. efficiens* und *C. diphtheriae* konserviert zu sein, da diese so auch über die bioinformatische RNAz-Analyse vorhergesagt war (Start der RNAz-Vorhersage Position 314679). Die 6C RNA ursprünglich nach ihren zwei Schleifen benannt, die in Actinobakterien beide häufig sechs aufeinanderfolgende Cytosin-Reste enthalten. In *C. glutamicum* besitzt die 6C RNA interessanterweise zwei Abschnitte mit jeweils acht aufeinanderfolgenden Cytosin-Resten (Abbildung IV-15a). Die Funktion der Cytosin-Homopolymere in der 6C RNA ist bislang nicht bekannt, daher bleibt die Relevanz dieses Unterschiedes unklar. Eine Analyse der Signalstärken im Northern Blot ergab für die 6C RNA keine Änderung in den verschiedenen Experimenten, aber jeweils eine sehr deutlich sichtbare Bande. Die entsprechenden Ergebnisse der intergenischen Microarrays sind in Abbildung IV-15a-c und Abbildung IV-16 unterhalb der Northern Blot-Bilder jeweils mit eingezeichnet. Die Northern Blot Banden der kleine RNA *cgb_20715* sind im Verhältnis zur Bande der 6C RNA zwar schwächer, aber dennoch deutlich und unterscheiden sich kaum in den Signalstärken untereinander (Abbildung IV-15c). Die quantitativen Aussagen wurden für die beiden genannten RNAs ähnlich über die Auswertung der spezifischen Microarrayexperimente erhalten.

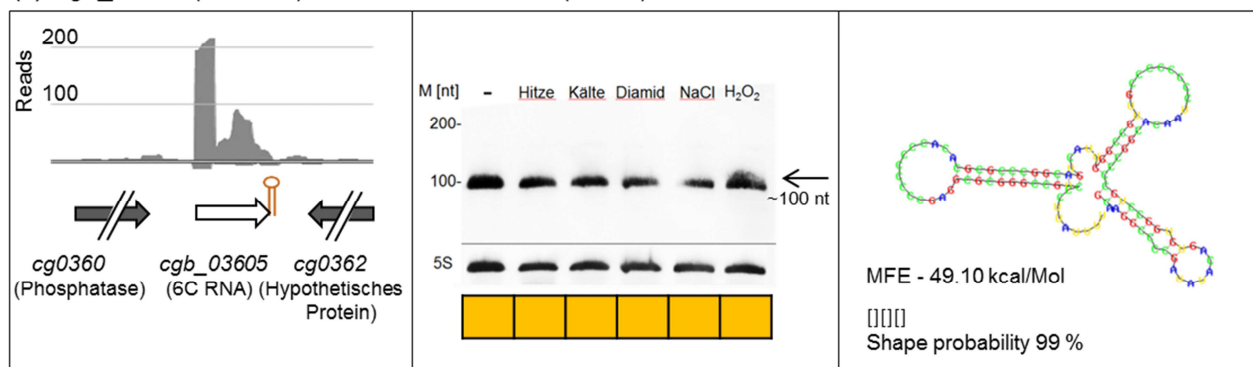
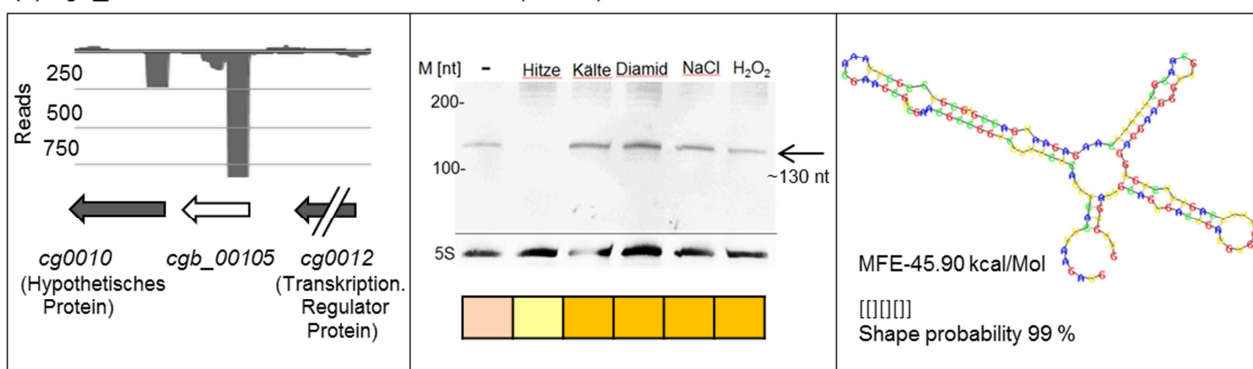
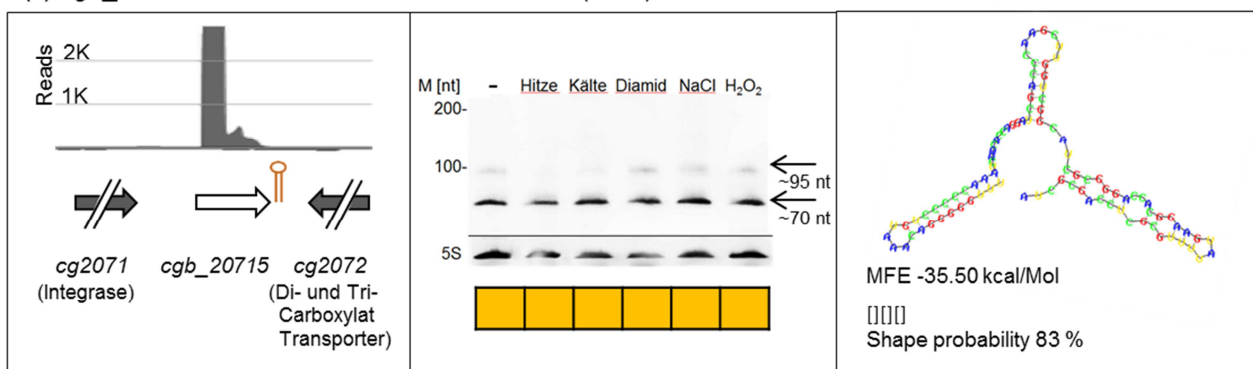
(a) *cgb_03605* (6C RNA) 314679 – 314787 (109 nt)(b) *cgb_00105* 10053 – 9921 (131 nt)(c) *cgb_20715* 1965663-1965755 (93 nt)

Abbildung IV-15 Sequenzabdeckung, Validierung durch Northern Blot und Sekundärstrukturvorhersage von drei *trans*-kodierte kleinen RNAs.

Validierung der kleinen RNAs (a) *cgb_03605* (6C RNA), (b) *cgb_00105* und (c) *cgb_20715*. **Linke Spalte:** Zeigt die Visualisierung und Sequenzabdeckung der *reads* (Hilker *et al.*, Manuskript submittiert) in cDNA-Bibliothek 1, wobei die x-Achse die Lokalisierung im Genom abbildet. Weiße Pfeile repräsentieren die kleinen RNA-Gene, graue Pfeile die Nachbargene, braune Haarnadelstrukturen die Rho-unabhängigen Terminatoren. **Mittlere Spalte:** Northern Blot-Ergebnisse unter allen Bedingungen bei denen die Zellen geerntet wurden. (-) $\hat{=}$ ohne Stress (OD₆₀₀ 10), Hitze-, Kälte-, Diamid-, Salz- und H₂O₂-Stress. Zusätzlich sind die Ergebnisse intergenischen Microarrayexperimente aufgeführt (gelb: Signal; rosa: deutliches Transkriptsignal; orange: sehr deutliches Transkriptsignal). **Rechte Spalte:** Sekundärstrukturen erzeugt über RNASHapes (Steffen *et al.*, 2006) bei MFE-Bedingungen (minimale freie Energie). Sequenz-Code: A-Blau; C-Grün; G-Rot; U-Gelb.

Ein vergleichbares Ergebnis mit den Microarrayexperimenten zeigt auch dass das Transkript von *cgb_00105* (Abbildung IV-15b), das bei Hitze-Stress im Northern Blot nicht detektiert werden konnte. Auch das entsprechende Microarrayexperiment liefert bei Hitze-Stress nur ein Signal im Bereich des Standard-Schwellenwertes, wogegen sonst mindestens deutliche Transkriptsignale erhalten werden. Es fällt auf, dass der Northern Blot von *cgb_20715* außerdem eine zweite, deutliche und kürzere Bande bei ~70 nt zeigt. Da in der entsprechenden Region kein zusätzlicher Promotor detektiert wurde, kann vermutet werden, dass das zweite Transkript dieser kleinen RNA über enzymatische Prozessierung generiert wird.

Auch die kleine RNA *cgb_16115* (Abbildung IV-16, nächste Seite) liegt in Form von mindestens zwei unterschiedlichen Transkriptgrößen vor und ergibt nach hyperosmotischem Stress mittels NaCl ein auffällig verändertes Bandenmuster. Anders als im Fall der kleinen RNA *cgb_20715*, wurde ein um 52 bp versetzter zweiter TSS für *cgb_16115* detektiert und zwar über die Suche nach SigH-abhängigen Promotoren. In der Tat zeigt die Promotorregion dieses zweiten TSS (-35: TGGATGA und -10: CGTTGTG) eine Ähnlichkeit zur SigH-Konsensussequenz (-35: gGGAAta und -10: YGTTgaa) (Busche *et al.*, 2012).²¹ Insgesamt besteht aber eine Ähnlichkeit zur Konsensussequenz von ECF-Sigmafaktoren (Pátek und Nešvera, 2011). Für SigH-abhängige Promotoren ist bekannt, dass sie Gene im Zusammenhang mit oxidativem Stress oder unter Hitze-Stress regulieren (Kim *et al.*, 2005; Ehira *et al.*, 2009). Unter diesen Stressbedingungen sind im Northern Blot von *cgb_16115* allerdings kaum stärkere Signale festzustellen. Zudem ist SigH unter normalen Wachstumsbedingungen ohne Stress durch den Anti-Sigmafaktor *rshA* (*cg0877*) inaktiviert (Busche *et al.*, 2012). Das kleinere Transkript der schwachen Bande knapp unter 100 nt kann wohl diesem zweiten TSS zugeordnet werden, wird aber auch ohne Stress detektiert. Unter hyperosmotischem Stress wird eine spezifische Prozessierung der Transkripte angenommen, die zu dem veränderten Bandenmuster im Northern Blot führt. Die Ergebnisse der Stress-spezifischen Microarrayexperimente sind an dieser Stelle nicht direkt vergleichbar, sie liefern ein sehr deutliches Transkriptsignal unter H₂O₂-Stress (9,2) und ergeben interessanterweise das mit Abstand höchste Signal unter NaCl-Stress (22,8). Vergleichende RT-qPCR Experimente waren wegen zu schwacher Signale nicht auszuwerten.

²¹ Dabei kann Y für ein C oder T stehen. Nukleotide in Großbuchstaben treten mit mehr als 80 % Einzelnukleotid-Wahrscheinlich auf.

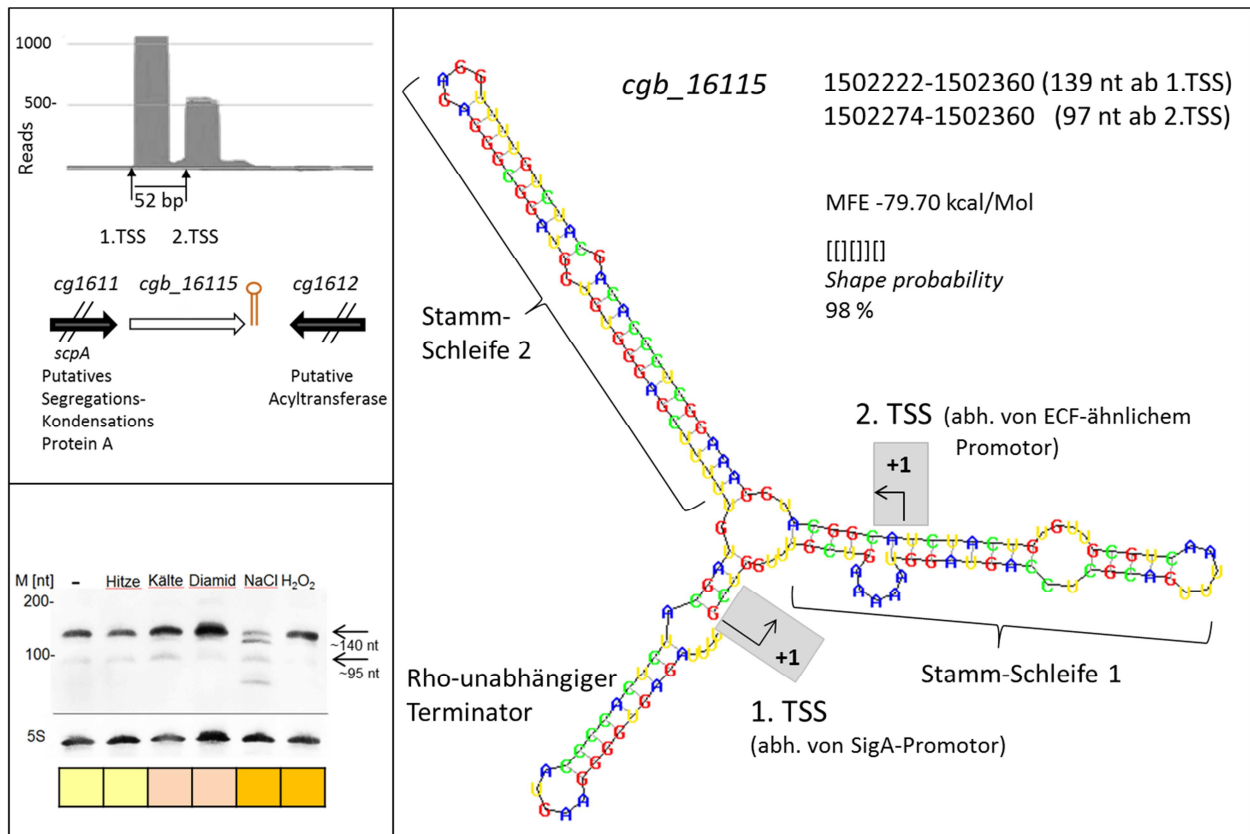


Abbildung IV-16 *Trans*-kodierte RNA *cgb_16115*.

Die Abbildung wurde in Analogie zu Abbildung IV-15a-c erstellt, zeigt hier aber eine vergrößerte Sekundärstruktur mit den Details zu den zwei Transkriptionsstarts (TSS) dieser kleine RNA. Der 1. TSS steht in Abhängigkeit von einem SigA-Promotor und der 2. TSS von einem ECF-ähnlichen (*extracytoplasmic function*) Promotor.

4.3 Funktionelle Analyse der *trans*-kodierten RNA *cgb_16115*

Die oben vorgestellte kleine RNA *cgb_16115* wurde im Rahmen dieser Arbeit genauer untersucht, da sie viele charakteristische Eigenschaften besitzt, die darauf hinweisen, dass es sich um eine regulatorische *trans*-kodierte RNA handelt. Außerdem stellt die Abhängigkeit von einem zweiten ECF-ähnlichen Sigmafaktor, der speziell zur Transkription der Stamm-Schleife 2 von *cgb_16115* führt, eine besondere und noch nicht beschriebene Art der Regulation einer kleinen RNA dar. Zusammenfassend sei hier noch einmal erwähnt, dass *cgb_16115* unter bestimmten Stress-Bedingungen vermutlich differenziell exprimiert ist und in der Geamtlänge (140 nt) eine stabile Sekundärstruktur mit zwei Stamm-Schleifen zeigt, gefolgt von einem Rho-unabhängigen Terminator. Direkt stromaufwärts der kleinen RNA liegt *scpA* (*cg1611*) und *scpB* (*cg1614*) in der näheren Umgebung, welche als Proteine beide in Wechselwirkung mit dem SMC-Protein (*structural maintenance of chromosome protein*) treten und damit eine wichtige Rolle bei der Kondensation und Segregation der Erbinformation spielen (Volkov *et al.*, 2003). Direkt stromabwärts auf die kleine RNA *cgb_16115* folgt in entgegengesetzter Orientierung das Gen *cg1612*. Zu diesem Gen sind bisher keine experimentellen Informationen vorhanden, aber anhand einer Datenbank von konservierten Domänen (Marchler-Bauer *et al.*, 2011) wird angenommen, dass es sich bei dem Produkt von *cg1612* um eine Acetyltransferase handelt. In den nächsten Kapiteln werden Ergebnisse zur funktionellen Analyse dieser kleinen RNA präsentiert, die sowohl über bioinformatische als auch mithilfe von experimentellen Methoden erhalten wurden.

4.3.1 Zugänglichkeitsanalyse von *cgb_16115* und verschiedene bioinformatische Methoden zur Vorhersage von Ziel-mRNAs dieser *trans*-kodierten RNA

Trans-kodierte kleine RNAs bewirken über ihre Bindung im Bereich der UTRs von Ziel-mRNAs häufig eine posttranskriptionelle Regulation. Hierüber kann entweder die Translation dieser Ziel-mRNAs reguliert und/oder deren Stabilität verändert werden (Storz *et al.*, 2011). Bei der Bestimmung der Funktion einer regulatorischen RNA können bioinformatische Programme hilfreich sein und können Hinweise auf eine putative Interaktion liefern. Verschiedene bioinformatische Programme erleichtern mittlerweile Vorhersagen von RNA-RNA-Interaktionen in Bakterien. Das Tool TargetRNA (Tjaden, 2008) dient zur genomweiten Suche von möglichen Zielen einer kleinen RNA und berechnet im Standardmodus (*individual basepair modus*) einen Hybridisierungs-score, bei dem das Basenpaarpotenzial für die interagierenden RNA-Moleküle bewertet wird. Die Tools RNAup (Mückstein *et al.*, 2006) betrachtet außerdem intramolekulare Strukturen sowie thermodynamische Eigenschaften und integrieren damit Zugänglichkeits- (*accessibility*) Berechnungen für zwei interagierende Sequenzen. In RNAPredator (Eggenhofer *et al.*, 2011) wurde diese Art von Zugänglichkeitsprofilen für diverse Genome vorberechnet. Daher kann auch mit diesem Programm sehr schnell nach möglichen Interaktionen einer kleinen RNA in einem Genom gesucht werden. Einen ähnlichen Ansatz stellt die Software CopraRNA

(Wright *et al.*, 2013) dar, der auch auf genomweiten distinkten Zugänglichkeitsanalysen basiert, aber zusätzlich phylogenetische Informationen aus anderen Genomen mit einbezieht, in denen die kleine RNA ebenfalls kodiert ist.

In der vorliegenden Arbeit wurden mit den Online-Tools TargetRNA und RNAPredator für die *trans*-kodierte RNA *cgb_16115* angepasste Suchen innerhalb einer angenommenen mittleren 5'-UTR-Länge von 60 bp für *C. glutamicum* und bis +20 bp aller annotierten Genstarts durchgeführt. Das Ergebnis wurde anschließend mit einer Analyse über das Tool CopraRNA verglichen. Da der Algorithmus von CopraRNA für die phylogenetischen Sequenzvergleiche einen größeren Suchbereich benötigt, wurde hier der Standardbereich von -200 bp bis +100 bp bezüglich aller annotierten Genstarts ausgewählt. Für die CopraRNA-Analyse musste auf die sehr ähnlichen Sequenzen der Stämme *C. glutamicum* R (Yukawa *et al.*, 2007) und *C. glutamicum* ATCC 13032 (Ikeda und Nakagawa, 2003) zurückgegriffen werden, weil über direkte BLASTn-Suchen (Altschul *et al.*, 1990) keine homologen Regionen von *cgb_16115* in anderen Bakterienspezies gefunden wurden. Insgesamt wurden über TargetRNA 21, über RNAPredator 40 und mit CopraRNA 155 Vorhersagen erhalten, wovon 35 Vorhersagen innerhalb des direkt vergleichbaren 5'-UTR-Suchbereichs bestimmt wurden (insgesamt wurden nur Kandidaten mit p-Werten < 0,05 gewertet, also < 5 % Irrtumswahrscheinlichkeit).²² In den Ergebnissen der beiden Methoden, die die Zugänglichkeit von *cgb_16115* berücksichtigen, werden drei distinkte Regionen auf der kleinen RNA für Interaktionen vorhergesagt (Abbildung IV-17A und B, nächste Seite), und zwar am Beginn der Stamm-Schleife 1 (Position 1-17), im Bereich zwischen Stamm-Schleife 1 und 2 bzw. nach dem 2. TSS (Position 52-62) und im Bereich der Stamm-Schleife 2 (Pos. 87-95). Über ein RNAup (Mückstein *et al.*, 2006) erzeugtes Zugänglichkeitsprofil von *cgb_16115* ist in den Bereichen mit einer Häufung von Interaktionsvorhersagen zu erkennen, dass die nötige Energie ΔG_i , um vier aufeinanderfolgende Nukleotide in der Struktur zu öffnen, jeweils sehr gering ist (Abbildung IV-17A). Die stärkste Häufung von Interaktionen wird für den Beginn der Stamm-Schleife 1 beobachtet, der Zwischenbereich ab dem 2.TSS wird am seltensten für Interaktionen vorhergesagt.

²² Alle detaillierten Ergebnisse sind der Datei Zielvorhersagen_cgb_16115.xlsx auf der Daten-CD zu entnehmen.

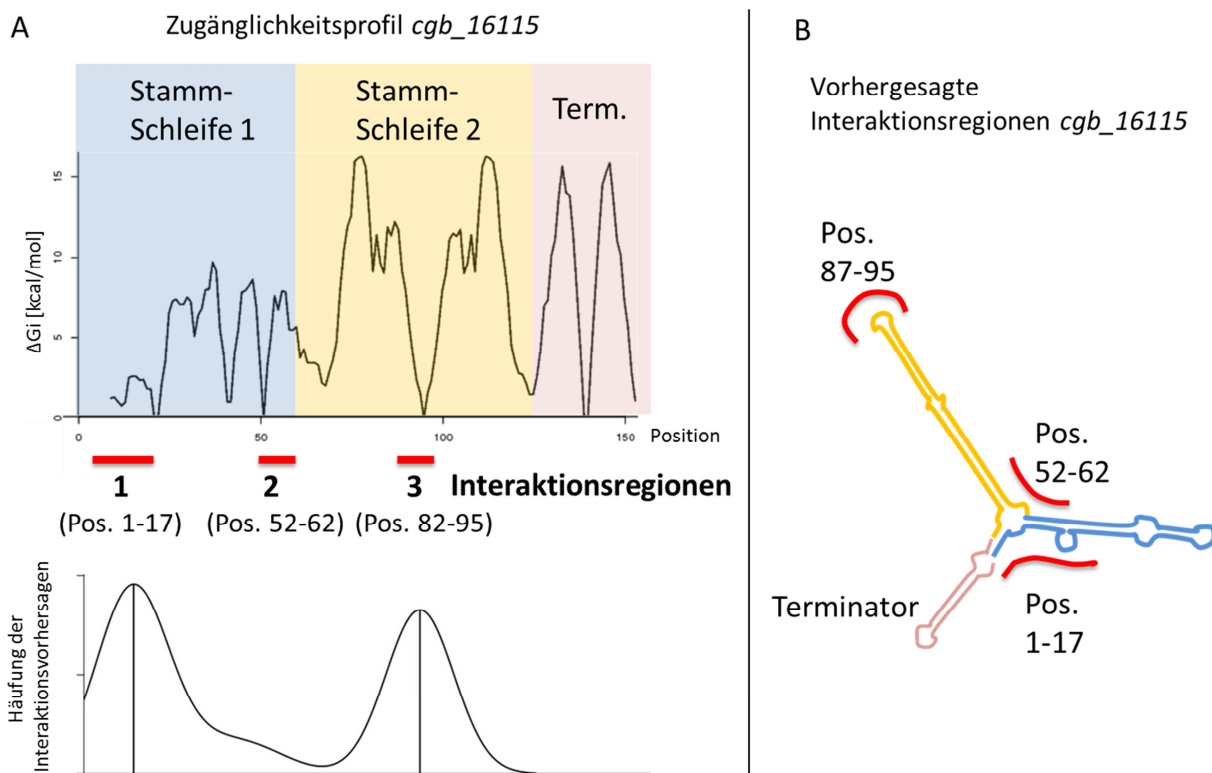


Abbildung IV-17 Zugänglichkeitsprofil von *cgb_16115* und drei häufig vorhergesagte Interaktionsregionen.

A Es existieren drei Regionen in denen sich Interaktionsvorhersagen über CopraRNA (Wright *et al.*, 2013) häufen. In diesen Bereichen ergibt das RNAup (Mückstein *et al.*, 2006) erstellte Zugänglichkeitsprofil, dass die nötige Energie ΔG_i , um vier aufeinanderfolgende Nukleotide in der Struktur zu öffnen, jeweils sehr gering ist. **B** Ermittelte Interaktionsregionen (rot) in der Sekundärstruktur-Darstellung von *cgb_16115* (verändert nach RNA-Shapes, Steffen *et al.*, 2006).

Ein Vergleich über alle Vorhersagen ist in einem Venn-Diagramm visualisiert (Abbildung IV-18, nächste Seite). Es existieren jeweils wenige Übereinstimmungen zwischen den erhaltenen Kandidaten, aber insgesamt stimmen nur zwei davon in allen drei Vorhersagen überein und betreffen die Region stromaufwärts von *cg3138* (*ppmA*), annotiert als Membran-Protease Untereinheit, Stomatin/Prohibitin Homolog) und *cg2618* (*vanK* annotiert als Vanillin-Transporter, Merkens *et al.*, 2005). Das Gen *cg3138* ist dabei als Ziel von Interaktionsregion 1 und *cg2618* als Ziel von Interaktionsregion 3 vorhergesagt. Das Gen *cg3138* wird dabei sogar über alle Methoden an der ersten Stelle der Vorhersagen berechnet. Dieses Gen ist als Membran-Protease-Untereinheit (Stomatin/Prohibitin-Homolog) annotiert und in einer Transkriptions-Einheit mit *cg3139* (hypothetisches Protein) und *cg3140* (*tag1*) kodiert (Pauling *et al.*, 2012). Über dieses Operon wurde bisher wenig publiziert, es ist aber bekannt, dass die Membran-Protease Untereinheit (*cg3138*) in einer Studie der Arbeitsgruppe von A. Poetsch als reichlich abundantes integrales Membranprotein detektiert wurde (fünfte Stelle in Bezug auf die relative Häufigkeit) (Fischer *et al.*, 2006). Außerdem wurde gezeigt, dass die Gene *cg3138-cg3139* vom dualen Transkriptionsregulator CysR (*cg0156*) des Schwefel-Metabolismus reprimiert werden (Rückert *et al.*, 2008).

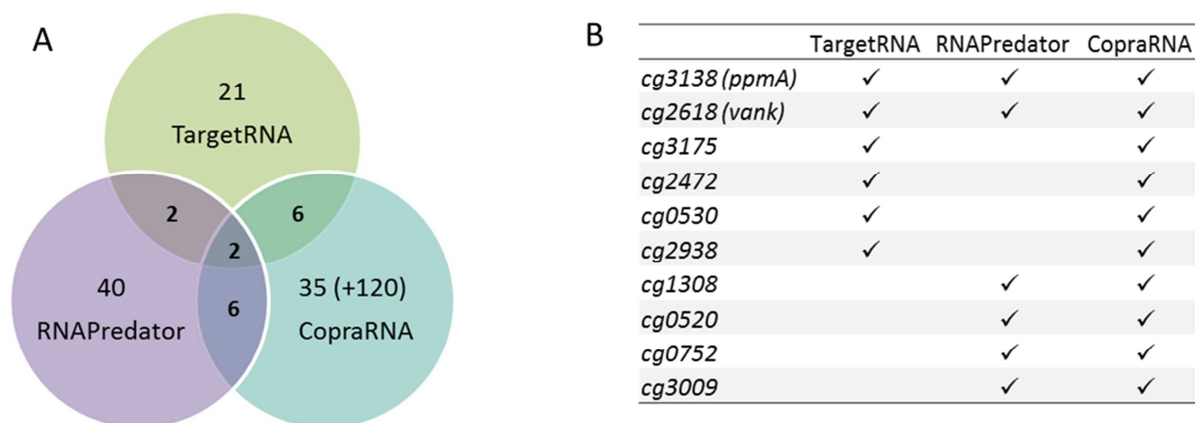


Abbildung IV-18 Venn-Diagramm über die Ergebnisse bioinformatischer Zielsuche von *cgb_16115*.

A In der direkt vergleichbaren Suchregion (-60bp bis +20 bp bezogen auf den TLS) wurden über CopraRNA 35 Vorhersagen erhalten, alle anderen 120 Kandidaten liegen außerhalb davon, aber innerhalb des Standard CopraRNA Suchbereichs. **B** Die Tabelle zeigt jeweils die Schnittmengen von vorhergesagten Ziel-mRNAs, die über die drei verschiedenen Programme für *cgb_16115* erhalten wurden.

Die verschiedenen Regionen mit Häufung von Interaktionsvorhersagen lassen vermuten, dass die kleine RNA *cgb_16115* auf mehrere Ziel-mRNAs in *C. glutamicum* regulatorisch wirken kann. Die Interaktionsregionen 2 und 3 sind dabei wahrscheinlich zusätzlich von einem ECF-ähnlichen-Promotor abhängig und besitzen damit wohl eine unterschiedliche Funktion als die Interaktionsregion 1. Die übereinstimmenden Vorhersagen der erhaltenen Ergebnisse deuten klar auf mindestens eine putative Interaktion mit der mRNA von *cg3138* und/oder *cg2618* hin. Es sind jedoch experimentelle Analysen zur Bestätigung und Validierung dieser bioinformatischen Zielvorhersagen notwendig.

4.3.2 Gezielte Deletionen verschiedener Interaktionsregionen von *cgb_16115* über *GeneSOEing*

Experimentell können Hinweise auf post-transkriptionelle Regulationen durch eine kleine RNA erhalten werden, wenn vergleichende Analysen auf Transkriptomebene zwischen dem Wildtypstamm und einer Mutante durchgeführt werden, in der die Expression der regulatorischen RNA verändert ist (Vogel und Wagner, 2007). Ein kleiner Teil der folgenden Ergebnisse und Methoden wurde in der Masterarbeit von A. Neshat beschrieben (Neshat, 2010), die unter meiner Anleitung erfolgt ist. Hierbei wurde in *C. glutamicum* ATCC 13032 nahe des 5'-Endes von *cgb_16115* mittels *GeneSOEing* eine 93 bp lange Deletion (Position 1502226-1502318) erzeugt, wobei alle drei Interaktionsregionen entfernt wurden, der Rho-unabhängige Terminator aber erhalten wurde.²³ Diese Deletionsmutante wird fortan als Δcgb_16115_Reg123 bezeichnet. Später wurde im *C. glutamicum*-Wildtypstamm eine zweite Deletion

²³ Die frühere Bezeichnung von Δcgb_16115_Reg123 nach einer alten Nomenklatur lautete Δscg_291 .

erzeugt und zwar von derselben Position an, aber nur innerhalb von Stamm-Schleife 1 bis Position 1502266 (41 bp). Im Detail wurde dabei die Interaktionsregion 1 mitsamt der ECF-Sigmafaktor-ähnlichen Promotorsequenz deletiert, womit die Interaktionsregionen 2 und 3 in dieser Mutante nun „speziell“ von einem SigA-abhängigen Promotor transkribiert werden. Diese Mutante wird fortan als Δcgb_16115_Reg1 bezeichnet. Aus praktikablen Gründen wurde beim Erzeugen dieser sehr kurzen Deletion zur Überprüfung von Mutanten eine Schnittstelle für die Restriktionsendonuklease *HindIII* kloniert. Ein Überblick über die Konstrukte der erzeugten Stämme ist in Abbildung IV-19 dargestellt.

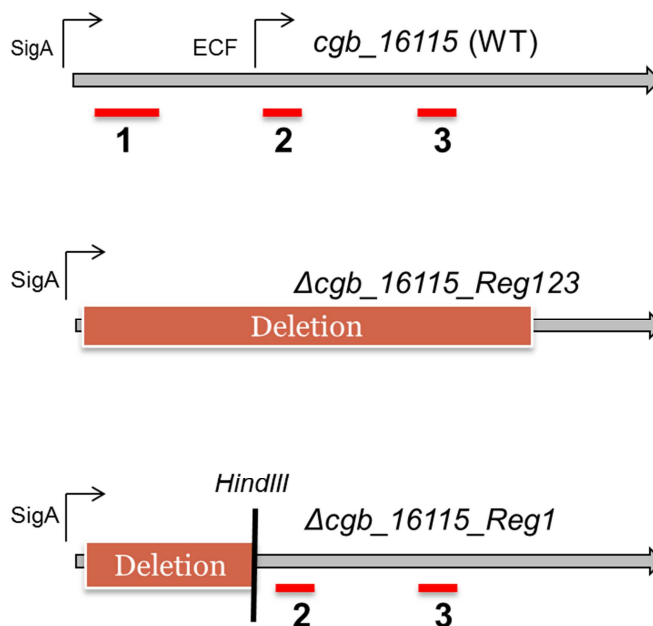


Abbildung IV-19 Überblick *cgb_16115* im Wildtypstamm im Vergleich zu zwei erzeugten Deletionskonstrukten.

Oben: *C. glutamicum*-Wildtyp (WT) von *cgb_16115* mit Interaktionsregionen 1-3 und Darstellung der zwei verschiedenen Transkriptionsstarts, die abhängig von einem SigA-Promotor bzw. ECF-(*extracytoplasmic function*) ähnlichen Promotor transkribiert werden. **Mitte:** Konstrukt der Gesamt-Deletion aller Interaktionsregionen (Rho-unabhängiger Terminator erhalten) **Unten:** Konstrukt mit Deletion von Interaktionsregion 1, wobei die Interaktionsregionen 2 und 3 nun speziell von dem SigA-abhängigen Promotor exprimiert werden. Zusätzlich wurde eine *HindIII*-Restriktionsschnittstelle kloniert.

Es soll erwähnt werden, dass Δcgb_16115_Reg123 kein verändertes Wachstumsverhalten in Minimalmedium gezeigt hat (Neshat, 2010), ebenso wurde für Δcgb_16115_Reg1 unter diesen Bedingungen keine Änderung in der Teilungsrate der Bakterien beobachtet. Auch ein mikroskopischer Vergleich aller Stämme ergab bei verschiedenen Bedingungen (H_2O und 1 M NaCl-Lösung) keine phänotypisch auffälligen Unterschiede. Hierbei wurden von den *C. glutamicum*-Zellen sowohl Hellfeld-

als auch Fluoreszenzaufnahmen mit dem Mikroskop Eclipse 80i (Nikon Instruments) gemacht. Für die Fluoreszenzaufnahmen wurden die Zellen mit einem Farbstoff markiert, der in die Zellmembran interkaliert (Molecular probes FM 4-64, Life Technologies, Darmstadt) (Vida und Emr, 1995).²⁴

4.3.3 Vergleichende cDNA-Microarrayanalysen zwischen dem *C. glutamicum*-Wildtypstamm und den Mutanten Δcgb_16115_Reg1 und Δcgb_16115_Reg123

Liegt ein Deletionsstamm einer bakteriellen *trans*-kodierten RNA vor, kann es aufgrund dieses fehlenden Regulationselementes bei entsprechenden Bedingungen zu Abweichungen in der relativen Expression der putativen Ziel-mRNAs gegenüber Wildtypstamm kommen (Massé *et al.*, 2005). Diese Abweichungen können primär auf einer veränderten Stabilität bzw. auf einem veränderten nukleolytischen Verdau beruhen, die entweder direkt auf die fehlende Interaktion zurückzuführen ist oder indirekt abhängig ist von der Anwesenheit translatierender Ribosomen auf der Ziel-mRNA. Aber auch sekundäre regulatorische Effekte der Deletion auf das Transkriptom werden sichtbar und können einen Hinweis auf den generellen funktionellen Zusammenhang einer kleinen RNA geben. Eine veränderte relative Expression kann global über vergleichende Microarraystudien detektiert oder für einzelne putative Ziel-mRNAs über RT-qPCR-Analysen validiert werden. Unter Verwendung isolierter Gesamt-RNA aus Kultivierungen mit den beschriebenen Mutanten Δcgb_16115_Reg123 und Δcgb_16115_Reg1 wurden jeweils gegen den *C. glutamicum*-Wildtypstamm ATCC 13032 als Referenz vergleichende cDNA-Expressionsanalysen über Zweifarben-cDNA-Microarrays nach der Standardmethode (Hüser *et al.*, 2003) durchgeführt. Nach LOWESS-Normalisierung (Yang *et al.*, 2002) der Hybridisierungsrohdaten und Einstichproben-t-Test über die biologischen Replikate wurden alle Gene mit M-Werten $\geq +0,6$ oder $\leq -0,6$ bei signifikanten p-Werten $< 0,01$ (Irrtumswahrscheinlichkeiten $< 1\%$) als signifikant erhöht oder erniedrigt in der Deletionsmutante gegenüber dem *C. glutamicum*-Wildtypstamm gewertet.²⁵ Im Folgenden soll nicht auf die Details der Microarrayergebnisse eingegangen werden, sondern die Tendenz wiedergegeben werden, die sich aus besonders auffälligen Ergebnissen ergeben haben.

Aufgrund der Northern Blot-Validierung und den intergenischen Microarrayanalysen ist eine differentielle Regulation durch die kleine RNA *cgb_16115* besonders unter hyperosmotischen Stressbedingungen zu vermuten. Daher wurde zunächst eine vergleichende cDNA-Microarrayanalyse mit Gesamt-RNA vom *C. glutamicum*-Wildtypstamm gegen die Deletionsmutante Δcgb_16115_Reg123

²⁴ Die Bilder der mikroskopischen Aufnahmen befinden sich auf der CD im Ordner „Analyse_cgb_16115“ im Unterordner „Mikroskopie“.

²⁵ Alle Ergebnisse zu den Zweifarben-cDNA-Microarrayexperimenten befinden sich auf der CD im Ordner „Analyse_cgb_16115“ im Unterordner „EMMA2-cDNA-Microarraydaten“.

(Deletion aller Interaktionsregionen) wie in den Validierungsexperimenten jeweils nach hyperosmotischem Stress mittels NaCl durchgeführt. Nachfolgend ist die Streuung der Microarraydaten in einem M/A-Diagramm visualisiert (Abbildung IV-20), wobei die M-Werte (y-Achse) das Verhältnis der Log_2 -Intensitäten und die A-Werte (x-Achse) die durchschnittliche Log_2 -Intensität für einen Punkt abbilden. Insgesamt ergibt sich in diesem Experiment eine hohe Anzahl von Genen mit signifikant differenziellen M-Werten (~ 300 Gene), wobei extreme Werte eher in negativer Richtung erhalten werden. Einzig das Gen *cg3138*, für das eine dreifache Interaktionsvorhersage besteht, weist mitsamt den übrigen Genen dieses Operons *cg3139* und *cg3140* M-Werte $> +1,5$ auf und zeigt damit eine besonders auffällig erhöhte relative Expression in der Deletionsmutante.

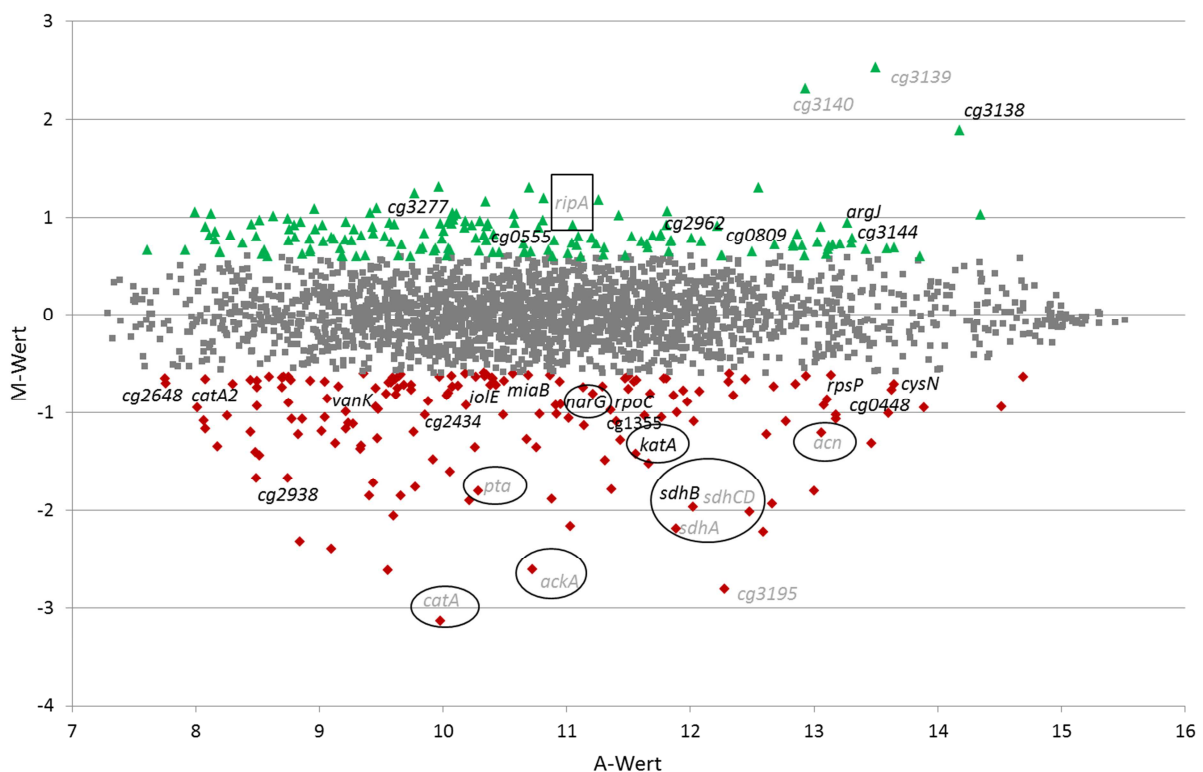


Abbildung IV-20 M/A-Diagramm für die Microarrayanalyse von *C. glutamicum* Δcgb_{16115} -Reg123 im Vergleich zum Wildtypstamm nach hyperosmotischem Stress.

Signale mit M-Werten $\geq +0,6$ oder $\leq -0,6$ gelten als signifikant erhöht oder erniedrigt gegenüber dem Wildtyp und sind farblich hervorgehoben (grün: stärkere Signale, rot: schwächere Signale, schwarz: keine Änderung). Alle Gennamen mit vorhergesagter Interaktion sind mit schwarzer Schrift bezeichnet, weitere Extremwerte wurden in grauer Schrift gekennzeichnet. Einkreiste Gene sind Teil des RipA-Regulons und das *ripA*-Gen ist von einem Viereck umrandet.

Es entsteht der Eindruck, dass sich die Zellen generell in einem Prozess größerer Umstrukturierung befinden, da 17 Transkriptionsregulatoren, in der Mehrzahl putative Regulatoren, signifikant differenzielle M-Werte zeigen.²⁶ Hierunter befindet sich mit einem leicht erhöhten M-Wert von +0,9 auch der AraC-Typ-Repressor von Eisenproteinen RipA (*cg1120*) (Wennerhold *et al.*, 2005). Insgesamt ergeben sich in diesem Experiment mit Δcgb_16115 -Reg123 nach hyperosmotischem Stress viele Ähnlichkeiten zur Regulation durch RipA (ebd.), da für fast alle Gene dieses Regulons signifikant negative M-Werte erhalten werden. Hierzu gehören die Gene des *sdhCDAB*-Operons, die Gene *narGHJI* (Nitrat Reduktase), *catA* (Catechol 1,2-Dioxygenase), *pta* (Phosphotransacetylase), *pcn* (Aconitase), *ackA* (Acetat Kinase) und das *kata* Katalase-Gen. Bis auf die Transkriptionseinheit *leuCD*, die in diesem Experiment keine differenzielle Expression zeigt, stimmt die Antwort zum RipA-Regulon überein (Wennerhold *et al.*, 2005). Neben *cg3138* werden in der Summe noch 21 weitere Gene detektiert, für die eine Interaktions-Vorhersage mit der *trans*-kodierte RNA besteht.²⁷ Hierunter befindet sich auch das Gen *vanK* (*cg2618*), für das ebenfalls eine dreifache Interaktions-Vorhersage besteht. Die Bezeichnungen dieser Gene wurden im M/A-Plot in schwarzer Schrift gekennzeichnet. Interessant ist, dass für einige Gene des Stimulons nach RipA-Deletion (Wennerhold *et al.*, 2005) ebenfalls Interaktionen mit *cgb_16115* vorhergesagt sind. Dies ist der Fall für die membrangebundene Succinat-Dehydrogenase Untereinheit *sdhB* (*cg0445*), *narG* (*cg1344*) und *kata* (*cg0310*). Zudem ist auffällig, dass eine Interaktion mit *catA2* (*cg1311*, Hydroxyquinol 1,2-Dioxygenase) vorhergesagt wurde (M-Wert -0,7) und *catA* (*cg2363*, Catechol 1,2-Dioxygenase) den signifikant geringsten M-Wert von -3,1 aufweist. Zusätzlich zeigen acht ABC-Typ putative Eisen-Siderophore Transporter-Untereinheiten leicht erhöhte M-Werte bei +0,7 (*cg0590*, *cg0591*, *cg0926*, *0927*, *cg0769*, *cg0771*, *cg0777*, *cg0778*). Aufgrund der vorliegenden Ergebnisse ist es möglich, dass nach hyperosmotischem Stress in Δcgb_16115 ein Eisenmangel entsteht. Es ist bekannt, dass RipA durch den DtxR-dualen Regulator (*cg2103*) bei Eisenüberschuss reprimiert wird (Wennerhold *et al.*, 2005; Brune *et al.*, 2006), wobei der M-Wert von *dtxR* im vorliegenden Experiment allerdings unauffällig bleibt.

Das Microarrayergebnis für *cg3138* konnte über RT-qPCR Experimente mit Gesamt-RNA Proben validiert werden (Abbildung IV-21, nächste Seite) und hat zu einer ~6-fach stärkeren relativen Expression in der Deletionsmutante Δcgb_16115 -Reg123 gegenüber dem *C. glutamicum*-Wildtypstamm als Referenz jeweils nach hyperosmotischem Stress mittels NaCl geführt.

²⁶ Alle Ergebnisse befinden sich im CD-Ordner „Analyse_cgb_16115“ im Unterordner „EMMA2-cDNA-Microarraydaten“.

²⁷ Eine Übersicht dazu befindet sich in Datei Microarray-trifft-Vorhersagen.xlsx, CD-Ordner „Analyse_cgb16115“.

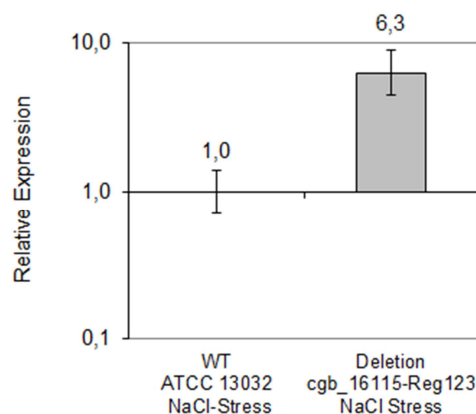


Abbildung IV-21 Validierung der erhöhten relativen Expression für *cg3138* in Δ *cgb_16115-Reg123 bezogen auf den *C. glutamicum*-Wildtypstamm (WT) als Referenz jeweils nach NaCl-Stress (RT-qPCR-Messung mit Gesamt-RNA).*

Zum Vergleich werden die Tendenzen der Hybridisierungs-Ergebnisse dargelegt, die mithilfe von isolierter Gesamt-RNA nach Zellernte bei exponentiellem Wachstum ohne Stress durchgeführt wurden, wobei nun die Hybridisierungs-Ergebnisse beider Deletionsmutanten jeweils gegen den *C. glutamicum*-Wildtypstamm als Referenz betrachtet werden. Im Gegensatz zum Experiment mit Δ *cgb_16115-Reg123 nach hyperosmotischem Stress zeigt sich im cDNA-Microarrayexperiment ohne Stress mit ~30 signifikant differenziellen Genen eine wesentlich schwächere Auswirkung auf das Transkriptom. Allerdings werden in der vergleichenden cDNA-Microarrayanalyse vom *C. glutamicum*-Wildtypstamm gegen Δ *cgb_16115-Reg1* über 300 Gene mit signifikant abweichenden M-Werten etwa gleichverteilt in positiver und negativer Richtung erhalten. Somit zeigt das Deletionskonstrukt, in dem die Interaktionsregionen 2 und 3 SigA-abhängig exprimiert sind, auch ohne Stress-Bedingungen einen enormen Einfluss auf das Transkriptom.²⁸ Für einen übersichtlichen Vergleich über die Gene mit besonders auffällig ähnlicher bzw. unterschiedlicher relativer Expression in beiden Experimenten wurden Gencluster über die signifikant differenziellen Gene in beiden Experimenten mit dem *k*-Means Algorithmus (Tavazoie *et al.*, 1999) über das Programm Genesis (Sturn *et al.*, 2002) berechnet. Anschließend wurden diese auf funktionelle Anreicherung nach Gen-Ontologie (GO) System (Ashburner *et al.*, 2000) über das David-Tool (Dennis *et al.*, 2003) untersucht.*

Nach der Clusteranalyse (*k*-Means 8) wurden die Gencluster (C1-C8) mit Größen zwischen 2 und 154 Genen erhalten.²⁹ Dabei ergab sich nur ein Gencluster mit tendenziell positiven M-Werten in beiden Experimenten (bis +1,9), und zwar das C6-Cluster (5 Gene). Hier ist interessanterweise das Gen *cg3138*

²⁸ Alle Ergebnisse befinden sich im CD-Ordner „Analyse_cgb_16115“ im Unterordner „EMMA2-cDNA-Microarraydaten“

²⁹ Die Datei *k*-Means-Gencluster.xls befindet sich im CD-Ordner „Analyse_cgb_16115“.

enthalten, für das die dreifache Interaktionsvorhersage besteht, mitsamt den anderen Genen der Transkriptionseinheit *cg3139* und *cg3140*. Dieses von der Tendenz her übereinstimmende Ergebnis in beiden Deletionsmutanten ist daher möglicherweise auf die Auswirkung der in beiden Mutanten gleichartigen Deletion von Interaktionsregion 1 zurückzuführen. In einer Analyse über alle *k*-Means-Cluster nach dem Gen-Ontologie (GO) System konnte nur für eines der acht Gencluster (C7) eine signifikante funktionelle Anreicherung detektiert werden (*score* > 1,3 bzw. *p*-Wert < 0,5). Dieses C7-Cluster stellt mit 154 Genen das größte *k*-Means-Cluster dar und zeigt als Einziges auffällig negative *M*-Werte nur im Experiment mit der Mutante Δ *cgb_16115_Reg1*. Nach der GO-Terminologie über 100 zugeordnete DAVID-IDs wurde eine signifikant funktionelle Anreicherung für zwei Annotations-Cluster erhalten: Mit dem höchsten *score* 5,5 (*p*-Wert < 0,00001) wurden die Gruppen Metall-Ionen-, Kationen-, Ionen-Transport in GO-Kategorie BP (biologische Prozesse) erhalten, gefolgt von den Gruppen di-, tri-valenter anorganischer Kationen-Transport, hochaffiner Eisen-Transport, Eisen-Ionen-Transport, Transmembran-Transport mit einem *score* 3,6 (*p*-Wert < 0,0005) in den GO-Kategorien BP oder MF (molekulare Funktion). Die folgende Tabelle IV-8 präsentiert einen Ausschnitt aus diesem C7-Cluster und zeigt beispielhaft die enthaltenen ABC-Typ Transportergene, wobei viele der Gene bzw. Operons im Zusammenhang mit dem Metall-Ionen Transport stehen. Insgesamt sind 30 der 154 Gene aus dem C7-Cluster (~20 %) im Modul Makroelement- und Metall-Homöostase (Pauling *et al.*, 2012) vertreten.

Tabelle IV-8 Ergebnisse von ABC-Typ Transportergenen im *k*-Means Gencluster C7 aus den vergleichenden Microarrayexperimenten der beiden *cgb_16115* Deletionsmutanten vs. *C. glutamicum*-Wildtypstamm (Kultivierung ohne Stress)^a

Locus-Tag	Gen-Name	Annotation	Δ <i>cgb_16115-Reg1</i>	Δ <i>cgb_16115-Reg123</i>
<i>cg0405</i>	-	ABC-type putative iron(III) dicitrate transporter, substrate-binding lipoprotein	-1,79	0,16
<i>cg0467</i>	-	ABC-type putative hemin transporter, substrate-binding lipoprotein	-1,27	0,35
<i>cg0469</i>	-	ABC-type putative hemin transporter, ATPase subunit	-1,26	0,37
<i>cg0506</i>	-	ABC-type putative spermidine/putrescine/iron(III) transporter, ATPase subunit	-1,33	0,11
<i>cg0507</i>	-	ABC-type putative spermidine/putrescine/iron(III) transporter, permease subunit	-1,42	0,41
<i>cg0508</i>	-	ABC-type putative spermidine/putrescine/iron(III) transporter,	-1,13	-0,09
<i>cg0589</i>	-	ABC-type putative iron-siderophore transporter, ATPase subunit	-0,78	0,21
<i>cg0590</i>	-	ABC-type putative iron-siderophore transporter, permease subunit	-0,72	-0,10
<i>cg0591</i>	-	ABC-type putative iron-siderophore transporter, permease subunit	-0,81	1,38
<i>cg0748</i>	-	ABC-type putative iron-siderophore transporter, substrate-binding lipoprotein	-0,87	0,56
<i>cg0832</i>	-	ABC-type putative sugar transporter, permease subunit	-0,81	0,09
<i>cg0924</i>	-	ABC-type putative iron-siderophore transporter, substrate-binding lipoprotein	-1,25	0,36
<i>cg0926</i>	-	ABC-type putative iron-siderophore transporter, permease subunit	-0,69	0,43
<i>cg0927</i>	-	ABC-type putative iron-siderophore transporter, permease subunit	-0,77	0,20
<i>cg0928</i>	-	ABC-type putative iron-siderophore transporter, ATPase subunit	-0,70	0,31
<i>cg1229</i>	-	ABC-type putative cobalt transporter, permease subunit	-1,84	-0,23
<i>cg1281</i>	-	ABC-type putative multidrug transporter, ATPase and permease subunit	-0,78	0,50
<i>cg1380</i>	<i>ssuA</i>	ABC-type aliphatic sulfonate transporter, substrate-binding lipoprotein	-1,53	-0,54
<i>cg2136</i>	<i>gluA</i>	ABC-type glutamate transporter, ATPase subunit (TC 3.A.1.3.9)	-0,79	0,16
<i>cg2137</i>	<i>gluB</i>	ABC-type glutamate transporter, substrate-binding lipoprotein (TC 3.A.1.3.9)	-0,62	0,33
<i>cg2234</i>	-	ABC-type putative iron(III) dicitrate transporter, substrate-binding lipoprotein	-1,41	0,03
<i>cg2911</i>	-	ABC-type putative Mn/Zn transporter, substrate-binding lipoprotein	-0,68	-0,15
<i>cg2912</i>	-	ABC-type putative Mn/Zn transporter, ATPase subunit	-0,83	0,35
<i>cg2939</i>	-	ABC-type putative dipeptide/oligopeptide transporter, ATPase subunit	-0,62	0,15

^a Die geschweiften Klammern fassen Gene in einem Operon zusammen

Neben *cg3138* wurden im Microarrayexperiment mit Δcgb_16115_Reg1 (ohne Stress) weitere 28 Gene mit signifikant differenziellen M-Werten ermittelt, für die auch eine Interaktions-Vorhersage mit der *trans*-kodierten RNA besteht. Hierunter befindet sich auch ein putative Transkriptionsregulator *cg2462* (*TetR*-Typ).³⁰ An dieser Stelle soll erwähnt werden, dass auch andere Transkriptionsregulatoren ohne Interaktionsvorhersage auffällige M-Werte zeigen: Der RipA-Repressor von Eisenproteinen A (*cg1120*; AraC-Typ) (Wennerhold *et al.*, 2005) und der GntR-Typ Regulator *cg1935*, der im Zusammenhang mit dem Glukonat-Stoffwechsel bekannt ist und die Glukose-Aufnahme reguliert (Frunzke *et al.*, 2008). Daneben sind neun putative Regulatoren differenziell reguliert, über die noch nichts oder nur sehr wenig bekannt ist.

4.3.4 Hypothesen zum Regulationsmodell der Interaktion *cgb_16115* mit *ppmA* (*cg3138*)

Aufgrund der übereinstimmenden Vorhersage über alle drei bioinformatischen Methoden und der Ergebnisse der vergleichenden cDNA-Microarrayanalysen wird besonders die Interaktion der *trans*-kodierten RNA *cgb_16115* mit *cg3138* als sehr wahrscheinlich angenommen. Daher wird nachfolgend versucht, ein putatives Regulationsmodell aufzustellen. Generell gilt: Wird im Zuge der Deletion einer kleinen RNA bei vergleichenden cDNA-Expressionsanalysen gegenüber dem Wildtypstamm eine erhöhte relative Expression der putativen Ziel-mRNA (wie im Fall von *cg3138*) festgestellt, so ist als Ursache eine höhere Stabilität dieser Ziel-mRNA aufgrund der fehlenden Regulation durch die kleine RNA möglich. In dem hier untersuchten System könnte die Bindung der kleinen RNA *cgb_16115* im Wildtypstamm somit eine direkte destabilisierende Wirkung auf die *cg3138*-mRNA haben und/oder die Instabilität wird aufgrund einer geringeren Frequenz von translatierenden Ribosomen auf dieser Ziel-mRNA hervorgerufen. Dies liegt daran, dass ein solch „nacktes“ Transkript ohne Ribosomen generell einem erleichterten nukleolytischen Verdau ausgesetzt ist. Somit ist auf Basis der experimentellen Daten eine Blockade der RBS von *cg3138* durch Bindung der kleinen RNA *cgb_16115* im Wildtypstamm denkbar.

Eine bioinformatische Sekundärstrukturanalyse der 5'UTR von *cg3138* soll helfen, die oben aufgestellten Vermutungen zu bestärken, jedoch ergibt sich ein unklares Bild. Die Sekundärstruktur der 5'-UTR von *cg3138* ergibt sich aus einer relativ kurzen 5'-UTR mit einer Länge von 41 bp (Möker *et al.*, 2004) und faltet sich mit hoher Wahrscheinlichkeit zu einer Haarnadelstruktur (*shape probability* 99 %, Abbildung IV-22A, nächste Seite). Es scheint so, als ob die RBS von *cg3138* schon im nativen Zustand auch ohne Interaktion von *cgb_16115* durch überwiegend komplementäre Adenin-Uracil Basenpaarungen blockiert ist. Kommt es nun zur Bindung mit der kleinen RNA *cgb_16115*, werden für diese bioinformatisch

³⁰ Eine Übersicht dazu befindet sich in Datei Microarray-trifft-Vorhersagen.xlsx, CD-Ordner „Analyse_cgb16115“.

Komplementaritäten mit der Anti-RBS vorhergesagt (Abbildung IV-22A). So könnte über die Interaktion der beiden RNA-Moleküle die RBS von *cg3138* möglicherweise für den Eintritt von Ribosomen freigelegt werden. Diese Hypothese spricht eher gegen die aus den Experimenten abgeleitete Annahme, dass die Bindung der kleinen RNA *cg16115* zu einer RBS-Blockade von *cg3138* führt. Es ist natürlich auch möglich, dass der Effekt der RBS-Blockade durch die Interaktion mit der RNA *cg16115* noch verstärkt wird. Für ein solches Regulationsmodell ergeben sich zusätzlich Hinweise über die Berechnung der Sekundärstruktur mit beiden interagierenden RNA-Molekülen über das Programm RNAcifold (Gruber *et al.*, 2008). Diese hat ergeben, dass während einer Interaktion bei MFE-Bedingungen weiterhin eine blockierte RBS von *cg3138* vorliegen könnte (Abbildung IV-22B), in diesem Fall aber hauptsächlich über komplementäre stabile Guanin-Cytosin-Paarungen. Wie sich schon über die bioinformatischen Zielvorhersagen und die cDNA-Microarrayanalysen ergeben hat, ist vermutlich Stamm-Schleife 1 von *cg16115* für die putative Interaktion mit der *cg3138*-mRNA verantwortlich und wurde so auch über RNAcifold (ebd.) erhalten.

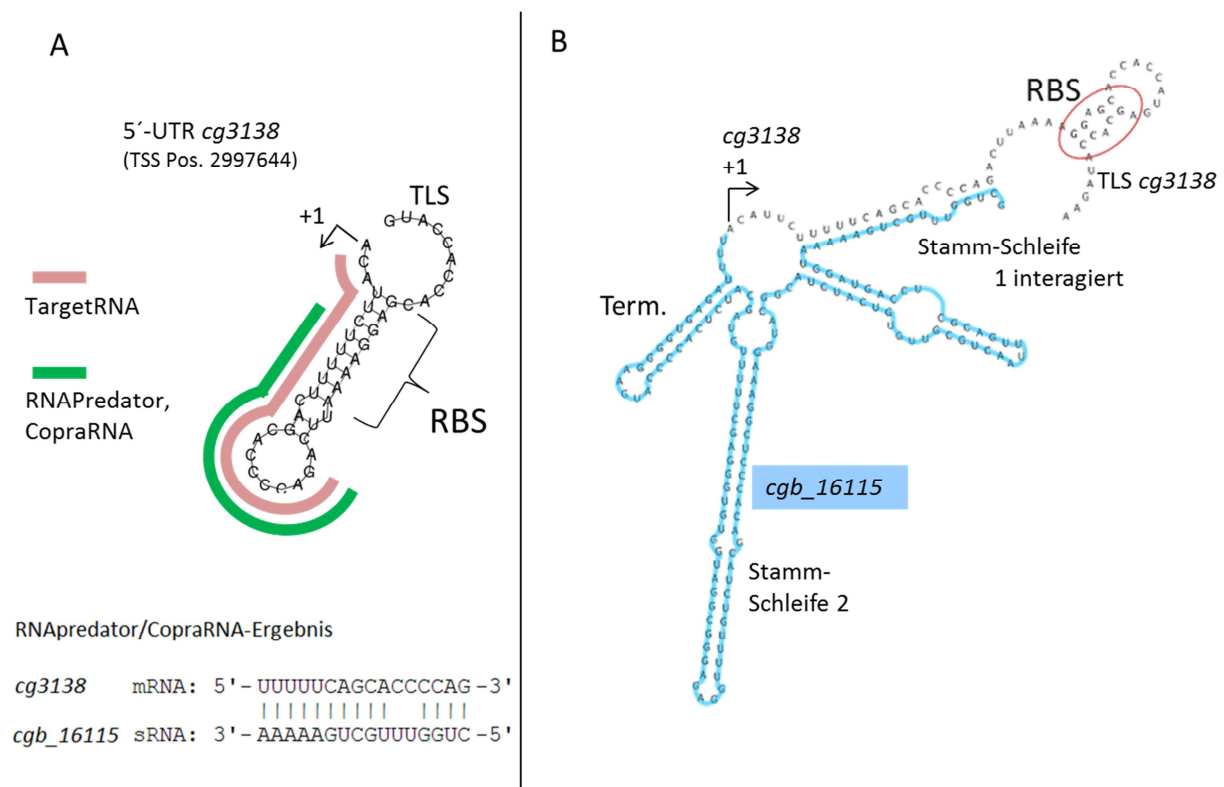


Abbildung IV-22 Sekundärstruktur der 5'-UTR von *cg3138* und Visualisierung der Interaktion mit *cg16115*. **A** Die Sekundärstruktur der 5'-UTR von *cg3138* berechnet über RNASHapes (Steffen *et al.*, 2006). In Rot bzw. Grün sind die entsprechenden Nukleotide markiert, für die eine Interaktion mit der *trans*-kodierte RNA *cg16115* über TargetRNA (Tjaden, 2008) bzw. RNAPredator (Eggenhofer *et al.*, 2011) und CopraRNA (Busch *et al.*, 2008) vorhergesagt ist. **B** Sekundärstruktur der Interaktion berechnet über RNAcifold (Gruber *et al.*, 2008) und visualisiert mit PseudoViewer3 (Byun und Han, 2009).

Die Überlegung, dass möglicherweise ein dreidimensionaler Pseudoknoten bei der Interaktion eine Rolle spielt, konnte über eine Analyse mit dem Programm pknotsRG (Reeder *et al.*, 2007) nicht bestätigt werden. Darüber hinaus kann auch eine weitere Hypothese aufgestellt werden, und zwar, dass die Expression von *cg3138* über einen komplexen regulatorischen Prozess erhöht wird. Hier ist z.B. ein Feedbackmechanismus denkbar, weil das RBS-freischaltende Regulationselement für *cg3138* mit der Deletion von *cgb_16115* fehlt.

Abschließend soll erwähnt werden, dass im Rahmen dieser Arbeit versucht wurde, die vorhergesagte Interaktion von *cgb_16115* mit *cg3138* über ein bereits etabliertes heterologes 2-Plasmid Co-Expressionssystem *in vivo* in *E. coli* mittels GFP-Translationsfusion (Urban und Vogel, 2007) nachzuweisen. Die erhaltenen Ergebnisse haben allerdings keine neuen Anhaltspunkte über den Wirkmechanismus ergeben. Es bleiben also die verschiedenen Hypothesen über das Regulationsmodell bestehen. Außerdem wurde eine vergleichende 2D-Gelelektrophorese mit den cytosolischen Proteinfractionen der erzeugten Deletionsmutanten Δcgb_16115_Reg1 sowie Δcgb_16115_Reg123 und dem *C. glutamicum*-Wildtypstamm durchgeführt. Auch hierbei haben sich keine neuen Hinweise auf die oben untersuchte bzw. andere putative Interaktionen ergeben. Die dazugehörigen Daten werden daher in der vorliegenden Arbeit auch nicht präsentiert.

4.4 Validierung eines neuen Attenuationsmechanismus in der Synthese aromatischer Aminosäuren

Als letztes Beispiel soll eine kleine RNA aus der Klasse „mRNA-Leader“ genauer untersucht und validiert werden. Unmittelbar stromaufwärts von dem Gen *aroF*, das als 3-Deoxy-d-Arabino-Heptulose-7-Phosphat-(DAHP)-Synthase annotiert ist, befindet sich ein peptidkodierendes Transkript *cgb_11283* RNA (*cg4017*) (Abbildung IV-23A, nächste Seite). AroF (*cg1129*) und das homologe AroG (*cg2391*) sind in *C. glutamicum* verantwortlich für den ersten Schritt in der Biosynthese aromatischer Aminosäuren (Ikeda, 2006) und synthetisieren Erythrose-4-Phosphat zu DAHP. Das im Sequenzierungsexperiment identifizierte Transkript *cgb_11283* zeigt mit dem Rho-unabhängigem Terminator eine Länge von 154 nt und wurde zuerst über eine Northern Blot-Analyse validiert (Abbildung IV-23b). Interessanterweise hat die Northern Blot-Analyse zusätzlich ein sehr viel längeres Transkript ergeben, vergleichbar mit der zu erwartenden Größe der *aroF* mRNA (~1.1 kb). Das vorhergesagte Peptid in *cgb_11283* ist 25 Aminosäuren lang und enthält ein Cluster von drei Codons für die aromatischen Aminosäuren Phenylalanin und Tyrosin (F-Y-F). Diese Tatsache und die Ergebnisse der nachfolgenden Analysen lassen auf einen Attenuationsmechanismus schließen, weshalb dieses Peptid AroR benannt wurde. Über Sequenzvergleiche wurde herausgefunden, dass die Peptidsequenz von AroR in verschiedenen *C. glutamicum* Isolaten und in den nah verwandten Spezies *C. efficiens* und *C. pekinense* gut konserviert ist (Abbildung IV-23c).

Um den Attenuator-Mechanismus von *aroR* als Leaderstruktur von *aroF* zu validieren, wurde eine *pheA*-Transposon-Mutante von *C. glutamicum* *pheA::IS6100* (Mormann *et al.*, 2006) verwendet. Das *pheA*-Gen von *C. glutamicum* kodiert für eine Prephenat-Dehydratase und ist als Enzym verantwortlich für den letzten Schritt in der Biosynthese von Phenylalanin. Die erzeugte *pheA*-Mutante ist daher auxotroph für Phenylalanin. Um die intrazelluläre Phenylalanin-Konzentration kontrollieren zu können, wurde bei Anzucht von *C. glutamicum* *pheA::IS6100* in CGXII-Minimalmedium (Keilhauer *et al.*, 1993) mit Alanyl-Phenylalanin (Ala-Phe) Dipeptid supplementiert. Für limitierende Bedingungen wurde eine Konzentration 0,1 mM Dipeptid gewählt und im Überschuss wurde 1 mM Dipeptid eingesetzt. Nach Isolierung der Gesamt-RNA aus beiden Kulturen, wurden RT-qPCR-Analysen mit drei verschiedenen Primerpaaren durchgeführt, um den Effekt von Phenylalanin auf die Transkription des *aroF*-Gens zu messen. Diese Primerpaare befinden sich (i) innerhalb der Leader-mRNA, die die Kodiersequenz von AroR umfassen (ii) innerhalb des Gens *aroF*, das die DAHP-Synthase kodiert und (iii) in einem Bereich, der die AroR- und AroF-kodierenden Regionen überlappt (Abbildung IV-23d).

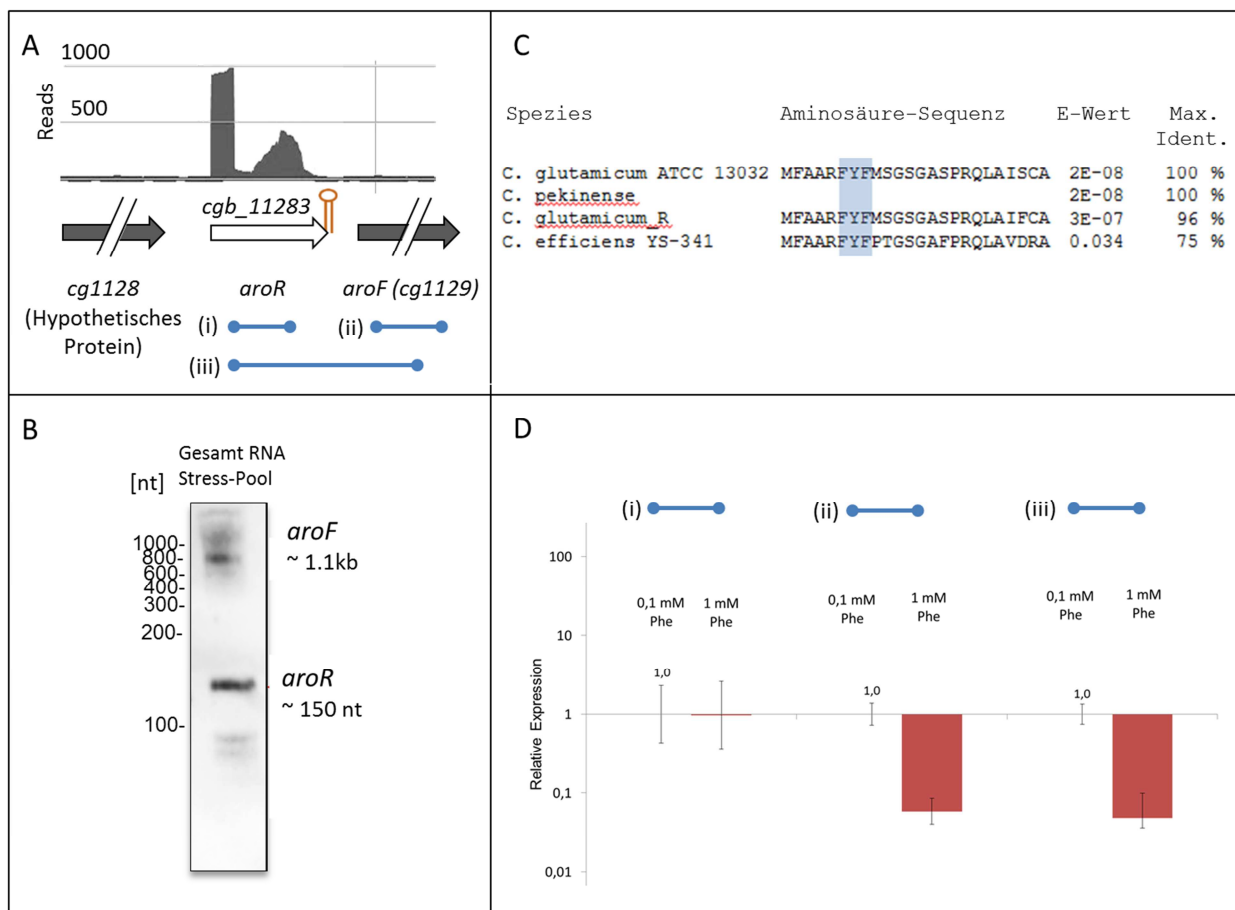


Abbildung IV-23 Transkriptionelle Attenuation für das DAHP-Synthase kodierende Gen *aroF* (*cg1129*).

A Die Visualisierung der *reads* (Hilker *et al.*, Manuskript submittiert) aus cDNA-Bibliothek 1 in der Region stromaufwärts von *aroF* zeigt das *aroR* Transkript mit Rho-unabhängigem Terminator. Die nummerierten blauen Linien repräsentieren die Regionen für drei verschiedene Primer-Amplifikate (i) in der *aroR* Region (ii) innerhalb der *aroF* kodierenden Region (DAHP-Synthase) und (iii) in der Region, die *aroR* und *aroF* überbrückt. **B** Northern Blot-Detektion der Leader-mRNA *aroR* und der langen *aroF* mRNA (~1.1 kb) über eine Digoxigenin-markierte RNA-Sonde für *aroR* (*cg11293*) in einem Gesamt-RNA-Pool von verschiedenen gestressten *C. glutamicum* ATCC 13032 Kulturen. **C** Resultat und Alignment einer BLASTn-Suche mit der AroR Peptidsequenz. **D** Validierung einer transkriptionellen Attenuation vor *aroF* über RT-qPCR-Analysen. Die Experimente wurden mit isolierter Gesamt-RNA einer Phe-auxotrophen Mutante *pheA::IS6100* von *C. glutamicum* ATCC 13032 durchgeführt. Für die Kultivierungen in CGXII-Minimalmedium wurde unter limitierenden Bedingungen mit 0,1 mM Ala-Phe Dipeptid supplementiert und mit 1 mM Ala-Phe im Überschuss.

Die RT-qPCR-Messungen haben gezeigt, dass die Attenuator-mRNA unter limitierenden Bedingungen dieselbe Konzentration aufweist wie bei Ala-Phe-Supplementation im Überschuss. Dagegen waren die relativen Mengen von *aroF*-Transkripten und von *aroR*- und *aroF*-überbrückenden Transkripten geringer als bei Supplementation im Überschuss, was auf eine Transkriptionstermination in der Attenuator-Region hinweist.

Eine bioinformatische Strukturanalyse hat potenzielle Nukleotidkomplementaritäten vorhergesagt, die eine Anti-Terminator Struktur begünstigen und die Bildung der Rho-unabhängigen Terminator-Stamm-Schleife verhindern könnten (Abbildung IV-24B). Die minimale freie Energie für die Attenuatorstruktur mit Rho-unabhängigem Terminator wurde mit 18 kcal/Mol berechnet und die Faltung des Anti-Terminators mit 22 kcal/Mol. Über Faltung der homologen Attenuator-Region in *C. efficiens* wurden sehr ähnliche Strukturen erhalten (Daten nicht gezeigt).

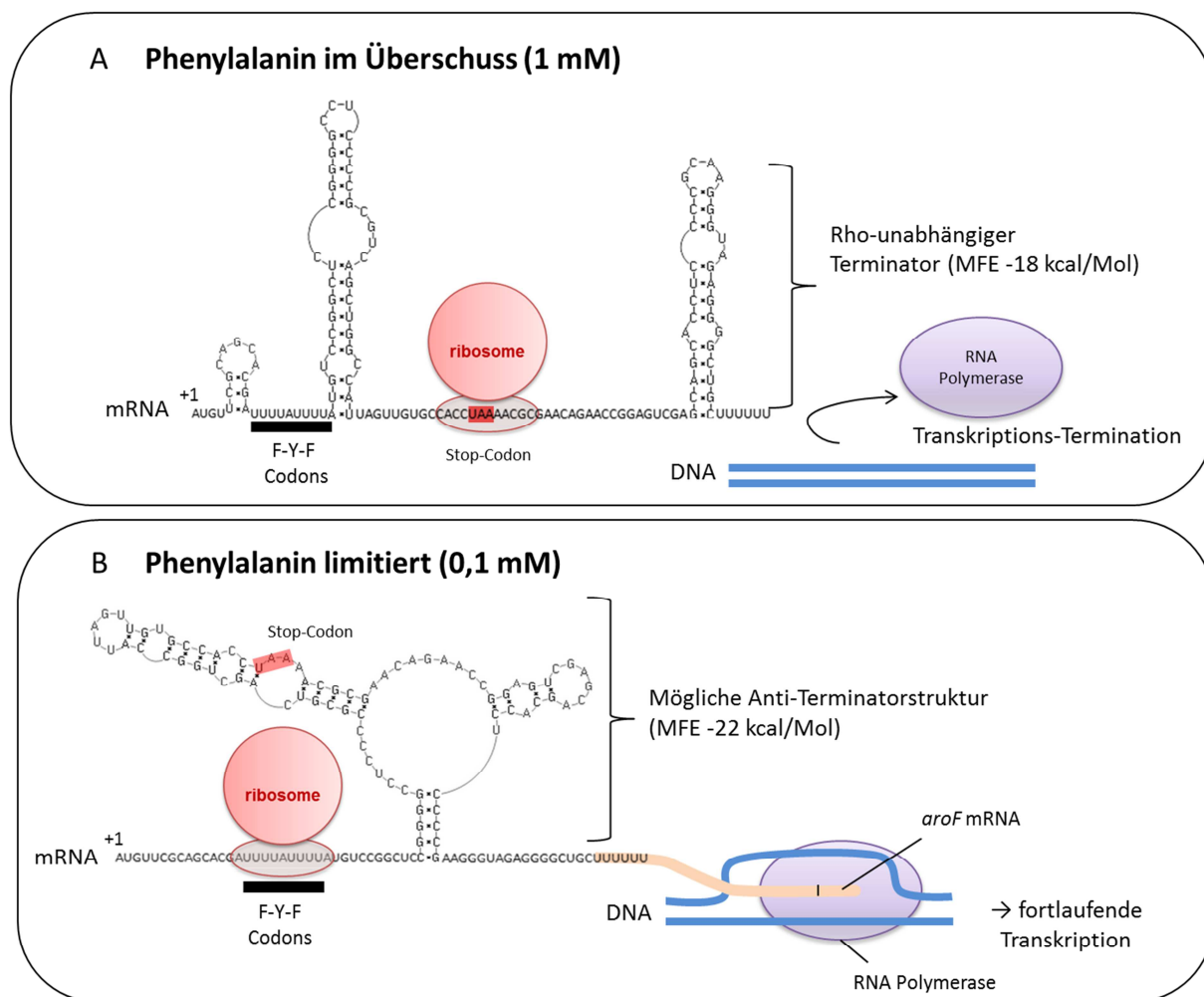


Abbildung IV-24 Potenzielle Terminator-/Antiterminatorstruktur im Attenuator stromaufwärts von *aroF* in *C. glutamicum*.

Die *aroR* Leader-mRNA ist in verschiedenen Sekundärstrukturen gezeigt. Nukleotide, die mit einem schwarzen Balken unterstrichen sind repräsentieren das Triple-Cluster der Codons Phenylalanin (F), Tyrosin (Y) und Phenylalanin (F). Nukleotide, die in Rot markiert sind, stellen das Stopp-Codon des AroR Leaderpeptids dar. **A** Überschuss von Phenylalanin führt zur Formation eines Rho-unabhängigen Terminators. **B** Mögliche Anti-Terminator-Struktur bei Phe-Limitierung. Beide möglichen Sekundärstrukturen wurden unter MFE-Bedingungen mit RNAShapes (Steffen *et al.*, 2006) berechnet.

V. Diskussion

1 Die Kombination verschiedener bioinformatischer und experimenteller Methoden führt zur umfassenden Identifizierung von kleinen RNA-Genen in *C. glutamicum*

Seit einigen Jahren nimmt das Wissen über die enorme Bedeutung von Regulationen durch kleine RNA-Moleküle in Bakterien zu. Das meiste Wissen über regulatorische kleine RNAs besteht für die Klasse von γ -*Proteobacteria* und hier besonders für *E. coli*- und *Salmonella*-Spezies. Mithilfe dieser Dissertationsarbeit wurden zum ersten Mal kleine RNA-Gene in dem industriell bedeutenden Bakterium *C. glutamicum* über eine Hochdurchsatz-Sequenzierung (RNA-Seq) identifiziert. Dieses Gram-positive Bakterium stellt gleichzeitig auch einen Modellorganismus für die nah verwandten Gattungen der *Corynebacterineae*, wie z.B. *Mycobacterium* oder *Nocardia* dar. Um ein möglichst umfangreiches Bild über die in *C. glutamicum* kodierten kleinen RNA-Gene zu erhalten, wurden in dieser Arbeit auch noch weitere Methoden angewendet. Hierbei sind eine bioinformatische *in silico*-Vorhersage regulatorischer RNAs und auch Transkriptanalysen über intergenische *tiling*-Microarrays zu nennen. Da die Techniken zur Hochdurchsatz-Sequenzierung von RNAs in Bakterien erst im Laufe dieser Dissertationsarbeit etabliert wurden (Sittka *et al.*, 2008; Sharma *et al.*, 2010) ist die RNA-Seq zeitlich betrachtet erst als letzte Methode zur Bestimmung kleiner RNAs in *C. glutamicum* angewendet worden (Mentz *et al.*, 2013). Zu Beginn dieser Arbeit hat die erfolgreiche Validierung der Methode von intergenischen *tiling*-Microarrays in *E. coli* (siehe Anhang) darauf schließen lassen, dass die entwickelte Prozedur zusammen mit dem Java-Applet IRMA (Schroeder, 2007) eine geeignete Methode ist, um kleine RNAs in Bakterien schnell und kosteneffizient zu detektieren. *Tiling*-Microarrays verschiedener Auflösung und Dichte sind auf anderen Microarrayplattformen bereits in den globalen Studien mit *E. coli* zur Identifizierung kleiner RNAs zum Einsatz gekommen (Selinger *et al.*, 2000; Wassarman *et al.*, 2001; Tjaden *et al.*, 2002; Zhang *et al.*, 2003) und werden bis heute auch in anderen Bakterien publiziert in *Caulobacter* (Landt *et al.*, 2008), in *Listeria* (Toledo-Arana *et al.*, 2009), in *Mycobacterium tuberculosis* (Miotto *et al.*, 2012), in *Streptococcus* (Kumar *et al.*, 2010; Patenge *et al.*, 2012) und in *Vibrionaceae* (Ahmad *et al.*, 2012). Der technologische Vorsprung von RNA-Seq zur exakten Kartierung kleiner RNAs wurde jedoch schnell deutlich (Schlüter *et al.*, 2010; Mraheil *et al.*, 2011; Vockenhuber *et al.*, 2011; Wilms *et al.*, 2012) und stellt heute die *state-of-the-art*-Technologie zur Identifizierung nach kleinen RNAs dar. Schon vor der Durchführung der RNA-Seq-Methode zur Identifizierung von kleiner RNAs in *C. glutamicum* haben also die Ergebnisse der intergenischen *tiling*-Microarrays auf eine rege Transkriptionsaktivität in ~10 % der intergenischen Regionen schließen lassen. Außerdem waren mittels einer RNAz-Vorhersage (Washietl *et al.*, 2005) zur Suche nach konservierten Sekundärstrukturen mit thermodynamischer Stabilität ~600 regulatorische RNA-Kandidaten in dieser Arbeit vorhergesagt, die auf Basis eines multiplem Genom-Alignments mit den verwandten *C. efficiens* und *C. diphtheriae* erhalten wurden.

Im Rahmen der experimentellen Analysen mittels der RNA-Seq- und Microarraytechnologien wurde jeweils Gesamt-RNA aus *C. glutamicum*-Zellen isoliert und weiterverwendet, die zum Zeitpunkt der Probenahme in der Überzahl Stress-Bedingungen ausgesetzt waren. Auf diese Weise sollten so viele kleine RNA-Transkripte wie möglich detektiert werden, da eine starke, differenzielle Expression regulatorischer RNAs besonders in Folge von Stresseinwirkung auf Bakterienzellen beobachtet worden ist (Hoe *et al.*, 2013). Obwohl für beide Analysemethoden jeweils Gesamt-RNA aus gleichartigen Kultivierungsexperimenten eingesetzt wurde, ergab sich über die vergleichbaren Signale in den intergenischen Regionen, wie beispielsweise für *trans*-kodierte RNAs, keine vollständige Übereinstimmung zwischen den beiden experimentellen Techniken. Dieses Ziel ist jedoch auch wenig realistisch aufgrund der vielen technischer Varianzen bei der Durchführung und wurde selbst zwischen vergleichbaren RNA-Seq Experimenten verschiedener Plattformen erreicht (Schlüter *et al.*, 2010).

In anderen Studien zeigte die (vergleichende) Gegenüberstellung von Microarraytechnologie und Hochdurchsatz-Sequenzierung jeweils eine höhere Sensitivität im RNA-Seq-Experiment und dies vor allem für schwach exprimierte Gene (Su *et al.*, 2011; Sîrbu *et al.*, 2012). Schon in Vorbereitung auf die Microarrayhybridisierungen kann die Effizienz der Fluoreszenzmarkierung durch die Rückfaltung kleiner RNAs zu stabilen Sekundärstrukturen negativ beeinflusst werden. Allerdings können sich durch diese stabilen Strukturen auch Verzerrungen in den RNA-Seq-Daten ergeben. So sind besonders die Adapterligation, die cDNA-Synthese und die Amplifikation zur cDNA-Bibliothek von elementarer Bedeutung. Kommt es zur Hinderung oder Bevorzugung dieser Reaktionen, kann dies die tatsächliche Höhe der Transkript-Level verfälschen. So weisen manche Transkripte eine extrem hohe Anzahl von *reads* auf und erscheinen besonders gut amplifiziert, andere sind dagegen nur schlecht zu detektieren (Raabe *et al.*, 2013). Im Rahmen der intergenischen Microarrayanalysen in dieser Arbeit wurde die markierte kleine RNA-Fraktion ohne cDNA-Synthese zwar direkt hybridisiert, allerdings ist hier zusätzlich mit nicht-spezifischen Kreuzhybridisierungen zu rechnen. Im Vergleich zu den sehr kurzen 25-mer Proben auf Affimetrix „GeneChip“-Microarrays und einen kritischen Punkt dieser Technologie in Bezug auf die Bindspezifität darstellen (Wu *et al.*, 2005), ermöglichen die verwendeten 50-mer Oligonukleotidproben wohl höhere Bindspezifitäten. Zusätzlich ist die Bindeeffizienz der kleinen RNA-Transkripte an die Oligonukleotidproben auf Microarrays aber auch abhängig von den Schmelztemperaturen der Proben, und diese variieren auf den verwendeten *tiling*-Microarrays. Während für die Signal-Variation unspezifischer Bindungen besonders die Schmelztemperaturen eine Rolle spielen, sind für spezifische Bindungen auch die intramolekularen Strukturen und thermodynamische Eigenschaften der Moleküle von Bedeutung, wie zumindest für NimbleGen Microarrays gezeigt wurde (Mueckstein *et al.*, 2010). Insgesamt konnten für die genom@one Plattform stabile Reaktionsbedingungen, eine hohe Genauigkeit und, wie auch in dieser Arbeit, eine hohe

Reproduzierbarkeit nachgewiesen werden (Baum *et al.*, 2003). Da die Reproduzierbarkeit und die Genauigkeit häufig ein großer Nachteil von Microarraystudien waren, wurde im Jahr 2006 das *MicroArray Quality Control* (MAQC) Konsortium (Shi und *et al.*, 2006, 2010) gegründet. Hier wurden Kriterien der Qualitäts-Kontrolle anhand der Analyse verschiedener Microarrayplattformen entwickelt, um eine Standardisierung von Microarraydaten zu ermöglichen.

Ein weiterer technischer Unterschied zwischen den experimentellen Ansätzen der RNA-Sequenzierung und der in dieser Arbeit gewählten Microarrayplattform ist folgender: Während im Sequenzierungsexperiment in *C. glutamicum* auch *cis-antisense* RNAs detektiert wurden, war dies mit den verwendeten intergenischen Microarrays nicht möglich. Die genom[®]one-Plattform bietet nur für eine kleinere Anzahl von Oligonukleotidproben Platz als andere hochauflösende Einfarben-Microarrayplattformen (Affymetrix „Gene Chip“, Illumina „Bead Chip“, Agilent „SurePrint-Arrays“). Verglichen mit Microarraystudien, in denen nur die Expression eines reduzierten Sets von hochkonservierten intergenischen Regionen analysiert wurde, wie beispielsweise in *Staphylococcus aureus* (Pichon und Felden, 2005), geben die Microarrayexperimente aus dieser Arbeit jedoch einen umfangreichen Überblick über die Transkriptionsaktivität in allen intergenischen Regionen und können damit viele Stamm-spezifische regulatorische RNA-Elemente detektieren. Insgesamt lässt sich zusammenfassen, dass die erhaltenen Transkriptlängen zumindest für starke Signale über beide experimentellen Analysemethoden gut vergleichbar sind, wie für die RNA-Haushaltsgene in *C. glutamicum* gezeigt wurde. Auch die Längen der bioinformatischen Vorhersagen waren an dieser Stelle gut vergleichbar.

Auf Grundlage verschiedener bioinformatischer Tools wurden immer wieder bioinformatische Vorhersagen kleiner RNAs in verschiedenen Bakterien unternommen, in *E. coli* beginnend im Jahr 2001 (Rivas und Eddy, 2001) und zuletzt in 2009 (Tran *et al.*, 2009). Innerhalb der letzten zehn Jahre sind dafür verschiedene bioinformatische Vorhersagemodelle implementiert worden (Sridhar und Gunasekaran, 2013). Die Übereinstimmung der aktuellen *in vivo*-Expression kleiner RNAs mit bioinformatischen Vorhersagen war in anderen Studien und anderen Bakterien dabei meist eher gering (Arnvig und Young, 2009; Perez *et al.*, 2009; Vockenhuber *et al.*, 2011). Auch in dieser Arbeit ist die Schnittmenge der mittels RNA-Seq detektierten *trans*-kodierte RNAs und gleichzeitigen RNAz-Vorhersagen nicht besonders hoch, beträgt aber immerhin 30 %. Eine Erklärung hierfür kann sein, dass es grundsätzlich unmöglich ist, in einem Experiment alle auslösenden Umweltfaktoren zu treffen, die die Expression regulatorischer RNAs bewirken. Andererseits existiert wohl auch eine größere Anzahl Stamm-spezifischer regulatorischer RNAs, die auf das Bewohnen einer ökologischen Nische eines

Bakterien-Stammes zurückzuführen ist bzw. dieses erst ermöglicht. In diesen Fällen sind bioinformatische Vorhersagen auf Basis komparativer Sequenzanalysen nicht möglich. Neben den rein vergleichenden Genomanalysemethoden existieren Vorhersage-Programme, die kleine RNAs auf Basis von Transkriptionsmerkmalen identifizieren. Hier stellen die sRNAPredict-Algorithmen (Livny *et al.*, 2005; Livny *et al.*, 2006) bekannte bioinformatische Tools dar, die kleine RNAs in konservierten Primärsequenzen stromaufwärts von Rho-unabhängigen Terminatoren vorhersagen. Die SIPHT-Analyse auf Basis des sRNAPredict Algorithmus (Livny *et al.*, 2008) ergab für *C. glutamicum* jedoch nur wenige kleine RNA-Kandidaten. Die erst vor kurzem entwickelte NAPP-Pipeline (Ott *et al.*, 2012) hat Hunderte phylogenetisch konservierter intergenischer Regionen in *C. glutamicum* ergeben, aber nur ungefähr ein Drittel der RNAz-Vorhersagen liegen auch in diesen mittels NAPP identifizierten Regionen. Im Abgleich mit den Sequenzierungsdaten *trans*-kodierter RNAs wurde allerdings eine vergleichbare Anzahl an Übereinstimmungen wie für die RNAz-Vorhersagen erhalten (ebenfalls ~30 %). Es kann also sinnvoll sein, für Vorhersagen kleiner RNAs in einem Bakterium verschiedene bioinformatische Methoden anzuwenden. Weiterhin sollte möglichst immer eine experimentelle Validierung zur Bestätigung der potentiellen kleinen RNAs vorgenommen werden und zur Ermittlung der exakten Start- und Stopp-Positionen.

2 RNA-Seq-Daten erlauben aufgrund ihrer Einzelnukleotid-Auflösung die präzise Bestimmung von Eigenschaften kleiner RNAs in *C. glutamicum*

Ein bedeutende Eigenschaft der RNA-Seq-Technik zur Analyse von Transkriptomen ist, dass sie sich zur Bestimmung von Transkriptionsstarts eignet (Sharma *et al.*, 2010). Die Anreicherung von Primärtranskripten wird mittels enzymatischer Behandlung der isolierten RNA-Proben erreicht. So führt die RNA-Seq-Technologie zur simultanen und exakten Bestimmung von Tausenden Transkriptionsstarts, die früher über 5'RACE (*rapid amplification of cDNA ends*) (Frohman, 1993) einzeln und arbeitsaufwendig bestimmt werden mussten. Der Ansatz nach Sharma *et al.* wurde auch in dieser Arbeit in *C. glutamicum* durchgeführt. Trotzdem ergeben sich in der Primärtranskript-angereicherten cDNA-Bibliothek noch eine Vielzahl von *reads* kleiner RNA-Transkripte, die wahrscheinlich über enzymatische Prozessierung entstanden bzw. aus Degradierungsprozessen hervorgegangen sind und nicht effizient genug über die enzymatische Exonuklease-Behandlung zur Anreicherung von Primärtranskripten abgebaut wurden. So werden beispielsweise nur prozessierte und degradierte Transkripte durch das verwendete Exonuklease-Enzym abgebaut, die eine 5'-Monophosphat-Gruppe aufweisen, Transkripte mit einer 5'-Diphosphat-Gruppe dagegen nicht. Aus diesen Gründen mussten in der vorliegenden Arbeit zur Bestimmung der putativen Transkriptionsstarts kleiner RNAs einige Filterschritte auf die erhaltenen *reads*

der Sequenzierungs-Daten angewendet werden. Hierbei wurde ein an die Größe des Datensets angepasster Anstieg von *read*-Starts von einer Position auf die nächste als Schwellenwert für einen putativen TSS eingesetzt. Im Anschluss daran wurde stromaufwärts aller putativen TSS über eine *in silico*-Analyse nach zugehörigen Promotoren gesucht. Hierbei wurde die bereits bekannte primäre SigA-Promotorsequenz verwendet (Pátek und Nešvera, 2011). Die erweiterte und besonders gut konservierte -10-Region dieser SigA-Konsensussequenz wurde gerade kürzlich in einer Gesamt-Transkriptom-Sequenzierung bei über 2000 Genen bestätigt (Pfeifer-Sancar *et al.*, 2013). Zusätzlich wurde nach Übereinstimmungen mit der SigH-Promotersequenz (Busche *et al.*, 2012) gesucht, denn von diesem ECF-Sigmafaktor liegen bisher die meisten Informationen über Promotorsequenzen vor. Da die Erkennungssequenz von SigH-Promotoren eine hohe Ähnlichkeit zu den der vier anderen ECF-Sigmafaktoren in *C. glutamicum* aufweist (Pátek und Nešvera, 2011), ist eine exakte Zuordnung der Promotor-Abhängigkeit nicht möglich. Es lässt sich feststellen, dass die Anzahl regulatorischer kleiner RNAs, die von einem ECF-ähnlichen Sigmafaktor abhängen sowohl für *cis-antisense* als auch für *trans*-kodierte RNAs sehr gering ist.

Zum Zeitpunkt der RNA-Seq-Auswertung kleiner RNAs in *C. glutamicum* lag noch keine Gesamt-Transkriptom-Sequenzierung mit definierten UTR-Längen der proteinkodierender Gene vor, diese wurde erst vor kurzem publiziert (Pfeifer-Sancar *et al.*, 2013), wobei allerdings speziell für RNAs > 200 nt untersucht wurden. Daher wurden die putativen kleiner RNAs für diese Arbeit über ihre relative Position zu den annotierten CDS in *trans*- und *cis-antisense* kodierte RNAs klassifiziert. Ein ähnliches Vorgehen wurde auch in in anderen Studien gewählt (DiChiara *et al.*, 2010; Schlüter *et al.*, 2010). Wie in dieser Arbeit wurden kleine RNAs bisher in *cis*- und *trans*-kodierte *antisense* RNAs eingeteilt bzw. klassifiziert, und die daraus abgeleitete Funktionalität der kleinen RNAs war zumeist eindeutig. Vor kurzem aber wurde zumindest in Archaeen das Paradigma einer strengen Grenze zwischen *cis*- und *trans*-kodierten kleinen RNAs überraschenderweise in Frage gestellt. Hier wurden in *Methanosarcina mazei* die ersten Funktionen einer kleinen RNA in Archaeen aufgeklärt (Prasse *et al.*, 2013). Diese Studie hat ergeben, dass *sRNA162* mit den 5'-Bereichen von zwei Ziel-mRNAs interagiert und dabei in einem Fall als *cis-antisense* RNA und im anderen Fall in *trans* auf eine entfernte mRNA wirkt.

Auch bei der Bestimmung der 3'-Enden mithilfe der Sequenzierungs-Daten für die cDNA-Bibliothek 2 wurde ein manueller Schwellenwert zur Detektion gesetzt. Nach einer anschließenden Längen-Analyse repräsentiert die Klasse der mRNA-Leader wie erwartet die längsten Transkripte. Dieses resultiert wohl auf der fortlaufenden Polymerisation bis in die zugehörige CDS beim Zeitpunkt eines Abbruchs der Transkription. Interessanterweise wurde in den vorliegenden Daten für *cis-antisense* RNAs (asRNAs), die

direkt gegenüber der CDS von mRNAs kodiert sind, eine besonders kurze Transkriptlänge festgestellt. Dieser Unterschied wurde in der Studie von Schlüter und Mitarbeitern nicht beobachtet, wobei die Sequenzierung hier auf dem 454-System (Roche, Basel, Schweiz) durchgeführt wurde (Schlüter *et al.*, 2010). Verglichen mit kleinen RNA-Genen in anderen Bakterien wurde in *C. glutamicum* eine kürzere Durchschnittslänge für kleine RNA-Gene erhalten, was ähnlich auch in *Streptomyces coelicolor* festgestellt wurde (Vockenhuber *et al.*, 2011) und im Gegensatz zum vorliegenden Sequenzierungsexperiment in dieser Arbeit (Illumina GA IIx-System) ebenfalls auf Basis des 454-Systems von Roche erhalten wurde. Die geringe Anzahl von Vorhersagen mit dem SIPHT-Tool (Livny *et al.*, 2008) hat dies schon vermuten lassen: Der Vergleich der in der Sequenzierung detektierten kleinen RNAs mit einer Vorhersage Rho (ρ)-unabhängiger Terminatoren ergeben, dass weniger als 25 % der *trans*-kodierten kleinen RNAs von einer solchen Terminatorstruktur gefolgt werden (Mentz *et al.*, 2013). Besonders die Anzahl der asRNAs, auf die ein Rho-unabhängiger Terminator folgt, ist marginal. Eine ähnliche Beobachtung wurde auch in einer Studie mit *Vibrio splendidus* erhalten (Toffano-Nioche *et al.*, 2012). Diese Tatsache passt zu der Beobachtung von Peters und Kollegen im Jahr 2009, die in einer genomweiten Analyse in *E. coli* ρ -abhängig terminierte Bereiche in mehreren hundert Fällen beobachtet haben, und zwar intergenisch am 3'-Ende von kleinen regulatorischen RNAs, für stabile tRNAs und intragenisch für kleine *antisense*-Transkripte (Peters *et al.*, 2009). Vor kurzem wurde ebenfalls in *E. coli* gezeigt, dass das ρ -Protein in der Tat die Hauptrolle in der umfassenden Unterdrückung von *antisense* Transkription spielt und dies in Abhängigkeit von seinem Cofaktor NusG (Peters *et al.*, 2012). Die Termination des kleinen RNA-Transkriptes *cgb_00105*, das keine eindeutige Terminatorstruktur zeigt und damit möglicherweise in Abhängigkeit vom ρ -Protein terminiert wird, konnte in dieser Arbeit über eine Northern Blot Analyse bestätigt werden.

Es wird deutlich, dass die erhaltenen *reads* nach einem durchgeführten RNA-Seq-Experiment eine sehr große und komplexe Menge von Informationen beinhalten. Dies beruht darauf, dass der Transkriptionsmechanismus in Bakterien wesentlich komplexer ist als lange Zeit angenommen wurde (Thomason und Storz, 2010). Daher sollten RNA-Seq-Rohdaten zur Detektion kleiner RNAs immer sorgfältig analysiert werden, um die relevanten Signale von möglichen Falsch-Positiven-Signalen abgrenzen zu können. Man muss allerdings davon ausgehen, dass aufgrund der angewendeten Filterschritte auch einige echte kleine RNAs mit untypischen bzw. unbekanntem Promoren verloren gehen.

3 Die Verknüpfung der Ergebnisse verschiedener Transkriptanalysemethoden eignet sich zur Bestimmung spezifischer Expressionsmuster *trans*-kodierter RNAs

Die mithilfe der neuen NGS-Sequenzierungstechnologien produzierten Daten weisen den bisher größten Informationsgehalt für die Transkriptom-Forschung überhaupt auf. Dem fundamentalen Aspekt des experimentellen Designs sowie der statistischen Analyse und Auswertung der produzierten RNA-Seq-Daten wurde jedoch einige Zeit lang wenig Aufmerksamkeit geschenkt, und besonders die Schritte zur quantitativen Auswertung sind bisher nicht standardisiert. Die entsprechenden Methoden werden aber zunehmend entwickelt (Chen *et al.*, 2011; Garber *et al.*, 2011). Ein statistisches Modell zur Abschätzung der Transkriptmengen wurde von Trapnell und Mitarbeitern aufgestellt und in dem Programm Cufflinks implementiert (Trapnell *et al.*, 2010). Aus der Microarraytechnologie ist ein Ansatz bekannt, der externe *spike-in*-Kontrollen zur Normalisierung von Transkriptmengen verwendet (Yang, 2006). Dieser Ansatz wurde auf RNA-Seq-Experimente übertragen und hat damit zur quantitativen Auswertung von Sequenzierungs-Daten sowohl in Eukaryoten (Mortazavi *et al.*, 2008) als auch in Prokaryoten (Wang *et al.*, 2011) geführt. Zur Analyse der differentiellen Expression kleiner RNAs ist es aus den oben genannten Gründen aber weiterhin sinnvoll, auch andere etablierte Methoden anzuwenden. In diesem Zusammenhang sind aber auch die vergleichsweise hohen Kosten als wesentliches Kriterium zu nennen, die bei der Durchführung von RNA-Seq-Experimenten immer noch entstehen.

Als eine gut etablierte Methode zur Transkriptanalyse ist die Northern Blot-Methode zu nennen, die für einige *trans*-kodierte RNAs auch in dieser Arbeit durchgeführt wurde und bisher zur stichprobenartigen Validierung in nahezu jeder Analyse zur Suche nach kleinen RNAs auch in anderen Bakterien durchgeführt wurde. Northern Blot-Analysen für alle neu detektierten kleinen RNAs durchzuführen wäre jedoch sehr zeitaufwendig. Wie auch in dieser Arbeit gezeigt, werden heute relativ kostengünstige Microarrayanalysen mit RNA-Seq-Experimenten kombiniert, um Experiment-spezifische Expressionsmuster der kleinen RNA-Transkripte zu erstellen (Schlüter *et al.*, 2010; Phillips *et al.*, 2013). In dieser Arbeit wurden kleine RNAs mittels Sequenzierung aus einem RNA-Gemisch detektiert, das aus Zellen verschiedener Kultivierungsexperimente stammte. Um die Experiment-spezifische transkriptionelle Aktivität speziell der intergenisch lokalisierten kleinen RNAs in größerem Umfang zu untersuchen, wurden die Sequenzierungsergebnisse mit den Experiment-spezifischen *tiling*-Microarrayanalysen verglichen. Wie im Ergebnisteil dieser Arbeit gezeigt, können über Datenbank-Abfragen nun speziell für die intergenisch detektierten kleinen RNAs Expressionsmuster aufgestellt werden. Um dies zu verdeutlichen, wurden an dieser Stelle 106 *trans*-kodierte kleine RNAs mit mindestens deutlichen Transkriptsignalen in mindestens einem Experiment in das Netzwerk-Tool

NodeXL (<http://nodexl.codeplex.com>) zur Erstellung eines Graphen importiert, was eine nützliche Methode zur Visualisierung größerer und vergleichbarer Datenmengen darstellt.³¹ Zur Erstellung der folgenden Abbildung V-1 wurde ein Algorithmus verwendet, der einen ungerichteten Graphen mit Mehrfachkanten visualisieren kann (Harel und Koren, 2000). Die Menge der Knoten ist definiert über die *trans*-kodierte RNA, wobei jeweils die Anzahl abgehender Kanten darüber Auskunft gibt, in welchen Experimenten die kleine RNA aufgetreten ist. Zur einfachen Darstellung wurden die Knoten je nach Anzahl der abgehenden Kanten farblich markiert. Im Netzwerk-Graph wird deutlich, dass in dieser Untermenge von 106 RNAs mit deutlichen Transkriptsignalen auffällig wenige spezifisch in einem Experiment auftreten (schwarze Knoten). Die Experimente mit Diamid- und Kälte-Stress zeigen jeweils mit einer Anzahl von 11 schwarzen Knoten die meisten *trans*-kodierte RNAs mit spezifischer Expression auf. Insgesamt sind 24 der kleinen RNAs in allen Experimenten vertreten (rote Knoten), zu denen auch alle RNA-Haushaltsgene gehören. Interessanterweise finden sich die meisten gemeinsamen Signale zwischen zwei Experimenten (jeweils grüne Knoten) mit 14 *trans*-kodierte RNAs in den Experimenten H₂O₂- und Diamid-Stress.

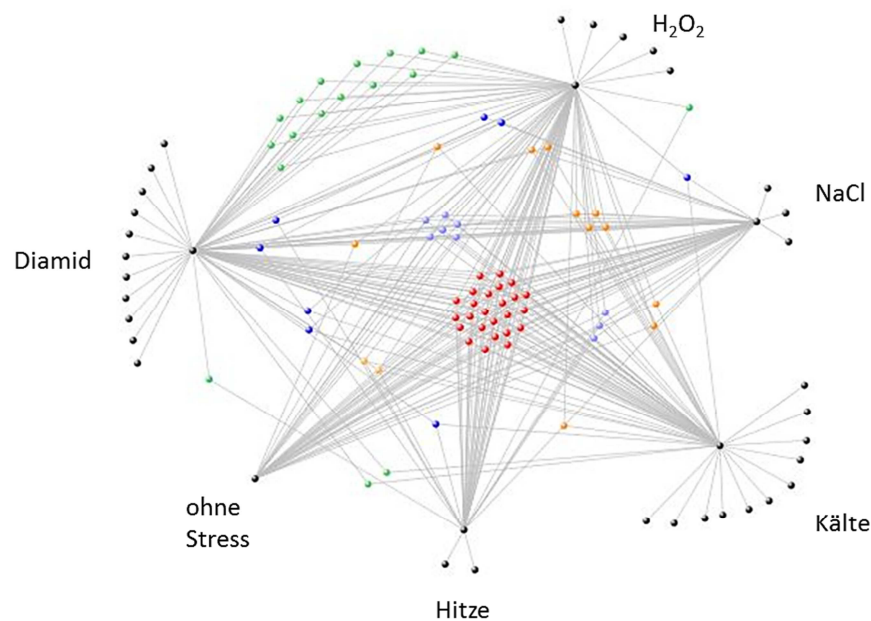


Abbildung V-1 Netzwerk-Graph einer Untermenge von *trans*-kodierten RNAs, die in den durchgeführten Experimenten transkriptionell aktiv waren.

Der Graph nach D. Harel und Y. Koren (Harel und Koren, 2000) zeigt Knoten, die stellvertretend für *trans*-kodierte RNAs stehen. Die abgehenden Kanten zwischen den Knoten repräsentieren dabei die Anzahl von verschiedenen Experimenten, in denen die kleine RNA mit mindestens einem deutlichen Transkriptsignal aufgetreten ist. Zur Visualisierung wurden die Knoten farblich markiert (schwarz: Signal spezifisch für ein Experiment; grün: zwei Experimente; blau: drei Experimente; gelb: vier Experimente; lila: fünf Experimente; rot: alle Experimente). Das Diagramm wurde mit NodeXL (nodexl.codeplex.com) erstellt.

³¹ In der Graphentheorie ist ein Graph $G(V, E)$ eine abstrakte Struktur, die eine Menge von V Objekten (Knoten oder Ecken) mit den zwischen diesen Objekten bestehenden Verbindungen E (Kanten) repräsentiert, wobei die Kanten gerichtet oder ungerichtet sein können.

Bakterienzellen sind in ihrer Umwelt häufig H_2O_2 oder anderen reaktiven Sauerstoff-Spezies ausgesetzt. Dieser oxidative Stress bedeutet für die Zellen, dass in den Proteinen Cysteine zu Disulfid-Brücken oxidiert werden, und dies führt zur Inaktivierung von Proteinen. Aber auch DNA und Lipide können geschädigt werden (Storz *et al.*, 1990). Daher gibt es in Bakterien ein komplexes System der zellulären Redox-Homöostase, das die Expression von Antioxidans-Genen wie Superoxid-Dismutasen (SOD), Peroxidasen und Katalasen reguliert (Aslund und Beckwith, 1999). Das in dieser Studie verwendete Diamid reagiert generell mit freien Thiolen in den Bakterienzellen. Die Produkte dieser chemischen Reaktion erzeugen nichtnatürliche Disulfid-Brücken sowie ein Hydrazin-Derivat im Cytoplasma und stellen somit Disulfid-Stress für die Zellen dar (Aslund und Beckwith, 1999). Da Thiol-oxidativer Stress als eine Unterkategorie von oxidativem Stress bezeichnet werden kann, sind die erhaltenen Gemeinsamkeiten für die Expression kleiner regulatorischer RNAs in den beiden Experimenten mit Diamid und H_2O_2 nachzuvollziehen. Unter oxidativen Stressbedingungen in *E. coli* spielt neben dem Redox-sensitiven Protein SoxR auch OxyR eine bedeutende Rolle (Aslund und Beckwith, 1999). OxyR stellt gleichzeitig einen Transkriptions-Aktivator aus der LysR-Familie von Regulatorproteinen dar, vermittelt als solcher die oxidative Stress-Antwort und induziert auch die kleine RNA OxyS (Altuvia *et al.*, 1998). Überraschenderweise wurde vor kurzem gezeigt, dass OxyR (*cg2109*) in *C. diphtheriae* und *C. glutamicum* als Repressor fungiert (Kim und Holmes, 2012). In dieser Arbeit konnte zudem kein OxyS-Homolog in *C. glutamicum* gefunden werden.

Das Homolog der aus Actinomyceten bekannten 6C RNA zeigte jeweils starke Bandensignale im Northern Blot und wurde mittels der *tiling*-Microarrays mit sehr deutlichen Transkriptionsignalen in allen Experimenten detektiert (abgebildet als roter Knoten in Abbildung V-1). Bis heute gibt es nur wenige Hinweise auf die Funktion der 6C RNA in *C. glutamicum*, wogegen in *Streptomyces coelicolor* eine gesteigerte Transkription der 6C RNA während der Sporulation ermittelt wurde (Swiercz *et al.*, 2008). Vor kurzem wurde allerdings gezeigt, dass die 6C RNA in *C. glutamicum* in das GlxR-regulatorische Netzwerk involviert ist (Jungwirth *et al.*, 2013). GlxR ist als globaler Regulator des Kohlenstoffquellen-Metabolismus und von Energie-Umwandlungsprozessen bekannt.

Aufgrund der Ergebnisse der Microarrayexperimente ist es nun möglich, die spezifische Expression für die Hälfte aller klassifizierten *trans*-kodierten RNAs zu überprüfen. Die Tatsache, dass in Vorbereitung auf die Hybridisierung der intergenischen *tiling*-Microarrays ohne reverse Transkription gearbeitet wurde, kann für die quantitative Auswertung der erhaltenen Signale vorteilhaft sein. Es soll an dieser Stelle aber erwähnt werden, dass sich quantitative Aussagen über die Microarraysignale als problematisch erweisen. Aus statistischer Sicht wurden nicht ausreichend viele Microarrayreplikate hybridisiert, bzw. es wurde keine Referenz-Probe oder externe *spike-in*-Kontrollen mitgeführt (Yang, 2006). So konnte auch keine statistische Analyse der erhaltenen Signale erfolgen. Beim Vergleich von Microarraysignalen einzelner *trans*-kodierter RNAs mit den Northern Blot-Signalestärken wurde unter den vergleichbaren

experimentellen Bedingungen meist eine ähnliche Tendenz erhalten. Insgesamt wird angenommen, dass sich die Ergebnisse der *tiling*-Microarrays aus dieser Arbeit zumindest zur ersten groben Einschätzung über quantitative Aussagen zur Experiment-spezifischen Expression kleiner RNAs eignen.

4 *Cis-antisense* RNAs sind in *C. glutamicum* zahlreich vorhanden und bei proteinkodierenden Genen bestimmter funktioneller Klassen häufiger anzutreffen

In den Daten der intergenischen *tiling*-Microarrays erschienen die vielen Signale, die gleichzeitig in beiden Strängen im Bereich der 5'-UTRs und 3'-UTRs von CDS auftraten zunächst noch rätselhaft und wurden zum Teil auf Kreuzhybridisierungen zurückgeführt. Solch ähnliche Beobachtungen wurden auch in anderen Studien zur Detektion kleiner RNAs mittels *tiling*-Microarrays auf anderen Plattformen gemacht, beginnend für *E. coli* (Wassarman *et al.*, 2001) bis zuletzt in *Listeria monocytogenes* (Toledo-Arana *et al.*, 2009). Eine extensive Transkription im *antisense* Strang proteinkodierender Gene konnte aber dann über RNA-Seq-Experimente in verschiedenen Bakterien bestätigt werden, und heute ist bekannt, dass *cis-antisense* RNAs einen grundsätzlicher Bestandteil der Transkription in Bakterienzellen darstellen (Lasa *et al.*, 2011). Auf Basis der Ergebnisse dieser Arbeit über kleine RNAs in *C. glutamicum* kann zusammengefasst werden, dass *cis-antisense* RNA-Gene auch in diesem Gram-positiven Actinobakterium zahlreich vorhanden sind und in ~15 % aller annotierten proteinkodierenden Gene zu finden sind (Mentz *et al.*, 2013). Dieser Wert gibt allerdings nur den Anteil von kleinen RNAs wieder, die direkt entgegengesetzt zu einer CDS starten und in dieser Arbeit sicher als *cis-antisense* identifiziert werden konnten (asRNAs). In vorhergegangenen Studien variierte der Anteil von Genen, die durch asRNA-Transkription gekennzeichnet sind, in verschiedenen bakteriellen Genomen recht stark und war am größten mit >46 % in *Helicobacter* (Sharma *et al.*, 2010). Die erste asRNA in *C. glutamicum* wurde stromaufwärts von *cg1935* detektiert (Zemanová *et al.*, 2008) und überlappt die mRNA des Regulatorproteins der GntR-Familie in *antisense*-Richtung. In dieser Arbeit wurden weitere asRNAs in *C. glutamicum* identifiziert, die gegenüber von Transkriptionsregulatoren lokalisiert sind, z.B. in *antisense*-Richtung von *cysR* (*cg0156*) und *sufR* (*cg1756*), die in die assimilatorische Sulfatreduktion involviert sind (Rückert *et al.*, 2008) und eine Rolle bei der Abwehr von Thiol-oxidativem Stress (Nakunst *et al.*, 2007) spielen. Außerdem wurden asRNAs am *acnR* Gen (*cg1738*) (Krug *et al.*, 2005) und anderen Mitgliedern der TetR-Familie (*cg2686*; *cg1308*) beobachtet. Ebenso wurden die folgenden Regulatorgene mit *antisense*-Transkription identifiziert: SugR (*cg2115*) als Regulator der Gene des PEP:Zucker-Phosphotransferasesystems (Engels und Wendisch, 2007; Gaigalat *et al.*, 2007), NdrR (*cg2112*) als Regulator der Desoxyribonukleotid-Reduktion (Jochmann *et al.*, 2009), PcaO (*cg2627*) als Aktivator der Gene des Ketoacid-Path-Metabolismus (Zhao *et al.*, 2010), FarR (*cg3202*) als transkriptioneller

Regulator der Gene des Stickstoff-Metabolismus (Hänssler *et al.*, 2007) und zwei Mitglieder der HTH_3-Familie (*cg1392*; *cg2040*).

Basierend auf dem eggNOG-Klassifikationssystem für orthologe Gene (Jensen *et al.*, 2008) hat die Analyse von asRNAs dieser Arbeit ergeben, dass diese Transkripte in *C. glutamicum* anscheinend häufiger in Genen auftreten, die eine Funktion in der Biogenese der Zellhülle und im Prozess der Proteinsekretion haben. Es existieren jedoch Hunderte von kurzen Transkripten in *antisense*-Richtung von proteinkodierenden Sequenzen in *C. glutamicum*, und ihre Funktionen müssen im Einzelnen noch aufgeklärt werden. Bisher konnte herausgefunden werden, dass asRNAs den Level von Transkriptionsregulatoren, metabolischen, toxischen und Virulenzproteinen beeinflussen sowie Transposasen reprimieren können (Thomason und Storz, 2010). Generell können *antisense*-Transkripte in Bakterien die Stabilität ihrer Ziel-mRNAs verändern, und zwar entweder über Erzeugung oder Blockierung von Ribonukleasen (Gerdes *et al.*, 1992; Opdyke *et al.*, 2004; Lee und Groisman, 2010). In diesem Zusammenhang konnten Lasa und Mitarbeiter für verschiedene Gram-positive Bakterien zeigen, dass überlappende *antisense*-Transkripte doppelsträngige Substrate für den Abbau mit RNaseIII erzeugen und dadurch viele kleine RNAs (~20 nt) generieren (Lasa *et al.*, 2011). Dieser Prozess stellt möglicherweise einen genomweiten posttranskriptionellen Mechanismus dar, um mRNA-Level anzupassen (ebd.). Andere Studien haben ergeben, dass asRNAs durch ihre Bindung zur Termination der Transkription führen können (Brantl und Wagner, 1994; Stork *et al.*, 2007) oder dass *antisense*-Transkription an der Ziel-mRNA einen veränderten Zugang zur ribosomalen Bindestelle bewirken und damit die Translation regulieren kann (Hernández *et al.*, 2006). Im Prozess der transkriptionellen Interferenz kommt es aufgrund der simultanen *in vivo*-Aktivität der Promotoren bzw. der konvergenten Transkription im *sense*- und *antisense*-Strang zur Konkurrenz der RNA-Polymerasen, wodurch die Aktivität des schwächeren Promotors durch die des Stärkeren in *cis* unterdrückt wird (Callen *et al.*, 2004, 2005). Auf Grundlage der vorliegenden Arbeit wurde eine *cis-antisense* RNA detektiert, für die ein Mechanismus transkriptioneller Interferenz auch in *C. glutamicum* gezeigt werden konnte, und zwar für *argC* (*cg1580*) (Petri *et al.*, Manuskript in Vorbereitung). Im Bereich nahe des 5'-Endes der *argC* mRNA liegt im *antisense* Strang die asRNA *cgb_15805*, die in Anlehnung daran *asaC* RNA benannt wurde (*antisense* zu *argC*) und ~90 nt lang ist (ebd.). Die in entgegengesetzter Richtung liegenden Promotoren von *argC* und von *asaC* führen zu Transkripten mit einer Überlappung von ungefähr 90 nt, was der erhaltenen Größe für die asRNA entspricht. Transkriptionelle Analysen nach der Zerstörung des *antisense*-Promotors und Überexpression von *AsaC* auf einem Vektor in *trans* haben ergeben, dass die Transkription von *argC* durch Transkription von *AsaC* beeinträchtigt wird und dass der Mechanismus dahinter transkriptionelle Interferenz ist (ebd.). Auch in anderen Organismen existieren Beispiele transkriptioneller Interferenz zur Regulation der Genexpression: In Bakteriophagen (Dodd und Egan), in Prokaryoten (Wang *et al.*, 1998), in Hefen (Hongay *et al.*, 2006) und Säugetieren (Martianov *et al.*, 2007).

5 Funktionelle Analysen der *trans*-kodierten RNA *cgb_16115* lassen auf eine Interaktion mit *ppmA* (*cg3138*) schließen und deuten auf weitere Regulationen im Zusammenhang mit Metall-Ionen-Transport hin

Der Untersuchung einiger *trans*-kodierter RNAs mittels Northern Blot-Analyse hat teilweise zwei oder mehr Signale ergeben. Im Fall von *cgb_20715* ist Reifung oder Degradierung über Endo- oder Exoribonukleasen wahrscheinlich, da kein zusätzlicher Promotor identifiziert wurde. Für *cgb_16115* wurde allerdings ein zweiter Promotor detektiert und führt speziell zur Expression von Stamm-Schleife 2 bzw. der Interaktionsregionen 2 und 3. Bisher ist bekannt, dass *trans*-kodierte RNAs wie DsrA in *E. coli* mit verschiedenen Ziel-mRNAs spezifisch durch unterschiedlichen Substrukturen bzw. Stamm-Schleifen interagieren können (Lease *et al.*, 1998). Die Expression einer einzelnen Substruktur einer kleinen RNA von einem zusätzlichen Promotor wurde aber noch nicht beobachtet. Darüberhinaus steht der zweite Promotor von *cgb_16115* nicht wie der erste in Abhängigkeit von dem primären Sigmafaktor SigA aus *C. glutamicum*, sondern besitzt Ähnlichkeiten mit einem SigH bzw. ECF-ähnlichen Promotor. Gene, die eine Abhängigkeit von SigH-Promotoren zeigen, wurden im Zusammenhang mit der Antwort auf oxidativen Stress oder Hitze-Stress beobachtet (Ehira *et al.*, 2009; Busche *et al.*, 2012). Da unter diesen genannten Bedingungen kein auffällig starkes Bandensignal im Northern Blot-Experiment beobachtet wurde, ist also eher eine Abhängigkeit von einem anderen ECF-Sigmafaktor zu vermuten, für welche recht ähnliche Erkennungssequenzen beobachtet worden sind (Pátek und Nešvera, 2011). Diese Annahme wird verstärkt durch die Tatsache, dass der hochkonservierte Bereich mit über 80 % Nukleotidwahrscheinlichkeit in der -35-Region mit „GGAT“ anders lautet als für SigH-abhängige Promotoren („GGAA“) und damit von der bekannten SigH-Promotor Konsensussequenz leicht abweicht. Erst kürzlich wurde von Busche und Mitarbeitern vermutet, dass gerade eine Änderung an dieser vierten Position zur spezifischen Erkennung eines anderen ECF-Sigmafaktors beiträgt. So wird an dieser Position ein „C“ für Promotoren vermutet, die in Abhängigkeit von SigE stehen (Tobias Busche, persönliche Mitteilung). Wie in dieser Arbeit durchgeführt, macht es für *cgb_16115* im Rahmen funktioneller Analysen Sinn, nicht die gesamte regulatorische RNA bzw. alle ihre Interaktionsregionen zu deletieren, sondern einzelne Interaktionsregionen zu deletieren, um so mehr Informationen über die Funktion(en) dieser speziellen kleinen RNA zu erhalten. In der erzeugten partiellen Deletionsmutante *cgb_16115_Reg1* wurde dabei der SigA-abhängige Promotor erhalten und führt nun speziell zur Expression der Interaktionsregionen 2 und 3.

Aktuell werden immer wieder neue bioinformatische Programme entwickelt, die sich zur Vorhersage von Ziel-mRNA regulatorischer RNAs eignen. Verlässliche *in silico*-Vorhersagen erweisen sich als kompliziert, denn im Gegensatz zu eukaryotischen miRNAs mit definierter Länge und vollkomplementären *seed*-Paarungen weisen bakterielle regulatorische RNAs eine extreme

Heterogenizität in ihrer Größe sowie Struktur auf und besitzen nur teilweise Komplementaritäten mit ihrer Ziel-mRNA (Papenfort *et al.*, 2010). Die aktuellen Vorhersage-Tools basieren entweder nur auf Basenpaarkomplementaritäten (Tjaden, 2008) oder zusätzlich thermodynamischen Modellen (Busch *et al.*, 2008; Eggenhofer *et al.*, 2011) und seit kurzer Zeit auch auf Kriterien phylogentischer Konservierung (Wright *et al.*, 2013). Trotz kontinuierlicher Verbesserung der Vorhersage-Algorithmen, wie beispielsweise die Berücksichtigung der Zugänglichkeit der Interaktionsregion, werden jeweils auch eine Anzahl falschpositiver Vorhersagen berechnet. So ist auch in dieser Arbeit die hohe Anzahl von Vorhersagen für *cgb_16115* zu erklären, die auf verschiedenen Algorithmen beruhen. Sicherlich findet nur ein Bruchteil der in dieser Arbeit erhaltenen Zielvorhersagen für *cgb_16115* auch *in vivo* statt. Möglicherweise gibt es *in vivo* auch Interaktionen, die über die bioinformatische Vorhersage gar nicht oder nur mit schlechten p-Werten berechnet wurden, z.B. weil für die entsprechende Interaktion noch ein unterstütztes Helferprotein wie Hfq (Valentin-Hansen *et al.*, 2004; Soper *et al.*, 2010), StpA (Waldsich *et al.*, 2002), HU51 (Balandina *et al.*, 2002) oder H-NS (Deighan *et al.*, 2000) benötigt wird. In dieser Arbeit wurde speziell nach putativen Interaktionsregionen in den 5'-Bereichen von mRNAs gesucht. Möglicherweise kommt es aber auch zu Interaktionen mit anderen Bereichen von mRNAs. Solche Mechanismen wurden allerdings für *trans*-kodierte RNAs noch nicht beschrieben. Insgesamt zeigt sich, dass ein größerer Teil bereits bekannter und experimentell bewiesener RNA:RNA-Interaktionen von den Programmen exakt berechnet wird (Busch *et al.*, 2008; Tjaden, 2008). Daher können sich über sorgfältiges Analysieren der bioinformatischen Vorhersagen Hinweise auf eine bisher unbekannte RNA:RNA-Interaktion ergeben wie beispielsweise in *Agrobacterium tumefaciens* gezeigt (Wilms *et al.*, 2011). Die Tatsache, dass in dieser Arbeit die zwei mRNAs *vanK* und *ppmA* mithilfe von allen drei Zielvorhersage-Programmen detektiert wurden, deutet in diesen Fällen besonders auf eine tatsächliche *in vivo*-Interaktion mit *cgb_16115* hin.

Die *trans*-kodierte RNA *cgb_16115* ist in einer genomischen Region lokalisiert, in der die umliegenden Gene für Proteine kodieren, die eine besondere Rolle bei der Zellteilung spielen. Das stromaufwärts liegende Gen *parA2* (*cg1610*) kodiert für ein Protein, welches in Abhängigkeit von ATP an der Verteilung der genetischen Information vor der Zellteilung beteiligt ist (Bignell und Thomas, 2001). Außerdem liegt *scpA* (*cg1611*) direkt stromaufwärts der kleinen RNA und *scpB* (*cg1614*) in der näheren Umgebung, welche als Proteine beide in Wechselwirkung *B. subtilis* mit dem SMC-Protein (*structural maintenance of chromosome protein*) treten und damit eine ebenso wichtige Rolle bei der Kondensation und Segregation der Erbinformation spielen (Volkov *et al.*, 2003). Direkt stromabwärts auf die kleine RNA *cgb_16115* folgt in entgegengesetzter Orientierung das Gen *cg1612*. Zu diesem Gen sind bisher keine experimentellen Informationen vorhanden, aber anhand einer Datenbank von konservierten Domänen (Marchler-Bauer *et al.*, 2011) wird angenommen, dass es sich bei dem Produkt von *cg1612* um eine Acetyltransferase handelt.

Vermutlich ist aber eine Funktion der kleinen RNA *cgb_16115* nicht direkt von ihrer Lokalisierung im Genom abzuleiten, denn die erzeugten Deletionsstämme zeigten keine Veränderung im Zellwachstum und auch keine sichtbaren Veränderungen der Zelllänge oder Zellteilung auf den Mikroskopiebildern, die in dieser Arbeit aufgenommen wurden. Trotzdem scheint die Auswirkung des Fehlens der kleinen RNA auf das Transkriptom in den meisten durchgeführten cDNA-Microarrayexperimenten enorm groß zu sein. Multiple Regulationen durch *trans*-kodierte RNAs wurden schon beschrieben. Für die kleine RNA GcbV in *Salmonella* wurde ein beeindruckend großes post-transkriptionelles Regulon aufgedeckt, das in die Regulation von ~1 % aller Salmonellen-Gene involviert ist (Sharma *et al.*, 2007; Sharma *et al.*, 2011). In den Ergebnissen der cDNA-Microarrayexperimente lassen sich jedoch primäre Effekte durch direkte Interaktionen mit der kleinen RNA *cgb_16115* nicht von sekundären Effekten unterscheiden, die sich beispielsweise durch Auswirkung auf Proteinregulatoren ergeben können. Deutlich wird jedoch, dass in allen präsentierten Microarrayexperimenten der Deletionsstämme im Vergleich zum *C. glutamicum*-Wildtyp das Gen *ppmA* (*cg3138*) mit erhöhten M-Werten aufgefallen ist, wobei jeweils die gesamte Transkriptionseinheit mit den Genen *cg3139* und *cg3140* auch signifikant verändert war. Hiermit wird also zusätzlich zu den bioinformatischen Zielvorhersagen ein weiterer Hinweis für eine Interaktion von *cgb_16115*, und wie vorhergesagt mit deren Stamm-Schleife 1, mit der *cg3138*-mRNA geliefert.

Genfunktionen können für *ppmA* anhand von Sequenzhomologien mithilfe der *Conserved Domain Database* (Marchler-Bauer *et al.*, 2011) vorhergesagt werden. So kann eine Homologie zwischen *ppmA* und einer Domäne Flotillin-ähnlicher Proteine identifiziert werden. Viele dieser Proteine sind als Mikrodomänen in der Zellmembran (*lipid rafts*) lokalisiert und können als Bestandteile von Multiproteinkomplexen an einer Vielzahl von physiologischen Prozessen beteiligt sein (Langhorst *et al.*, 2005). In einer Studie von Fischer und Mitarbeitern aus der Arbeitsgruppe von A. Poetsch wurde die Membran-Protease Untereinheit *ppmA* als reichlich abundantes integrales Membranprotein detektiert (fünfte Stelle in Bezug auf die relative Häufigkeit) (Fischer *et al.*, 2006). Zudem haben vorangegangene Microarrayanalysen ergeben, dass *cg3138* vom dualen Transkriptionsregulator CysR (*cg0156*) des Schwefel-Metabolismus reprimiert wird (Rückert *et al.*, 2008). Es ist außerdem über verschiedene Stimulons mit stark erhöhten M-Werten (bis +3,9) aufgefallen, und zwar (i) nach Deletion des Zur-Regulators in *C. glutamicum* (*cg2502*; Untergruppe Fur-Regulatoren), das den Schlüsselregulator der Zink-Homöostase darstellt (Schröder *et al.*, 2010), (ii) nach Deletion des dualen Transkriptionsregulators DtxR, der in das globale regulatorische Netzwerk des Eisen-Metabolismus involviert ist (Brune *et al.*, 2006), und (iii) nach Deletion des Zwei-Komponentensystems MtrAB (Möker *et al.*, 2004). Alle M-Werte dieser genannten regulatorischen Proteine sind in den Microarrayexperimenten allerdings unauffällig geblieben. Zu dem Gen *cg3139* und seiner Funktion konnten keine Informationen ermittelt werden. Das Gen *cg3140* (*tagA1*) kodiert laut CCD-Analyse für eine 3-Methyladenin-DNA-

Glykosylase I. Dieses Enzym ist an der Reparatur von DNA-Schäden durch alkylierende Verbindungen beteiligt, indem es mit hoher Effizienz 3-Methyladenin aus der DNA entfernt (Bjelland und Seeberg, 1996).

Im vergleichenden Microarrayexperiment zwischen dem *C. glutamicum*-Wildtypstamm und der erzeugten Deletionsmutante *cgb_16115_Reg123* (Verlust aller Interaktionsregionen) zeigte diese Transkriptionseinheit *cg3138-cg3139* nach hyperosmotischem Stress eine besonders auffällige differenzielle Expression. Bakterien und besonders Bodenbakterien sind häufig hyperosmotischen Stress-Situationen ausgesetzt. Man unterscheidet hier zwischen hyperosmotischem Stress, wenn aufgrund von hoher Osmolarität in der Umgebung ein Efflux von Wasser aus den Zellen heraus stattfindet, und hypoosmotischen Stress, wenn aufgrund von niedriger Osmolarität in der Umgebung Wasser in die Zellen hineinfließt. Um das dynamische Gleichgewicht wieder herzustellen, kommt es unter hyperosmotischen Bedingungen zur kontinuierlichen Aufnahme osmotisch aktiver gelöster Stoffe. Unter hypoosmotischen Bedingungen dagegen sind chemosensitive Kanäle geöffnet, und es werden kleine gelöste Stoffe aus den Zellen heraustransportiert. *C. glutamicum* ist mit mindestens zwei, MscL (Cg1001) und YggB (Cg1434), dieser chemosensitive Kanäle gegen hypoosmotischen Stress ausgestattet (Nottebrock *et al.*, 2003). Unter hyperosmotischen Stressbedingungen werden Glycin-Betain, Ectoin und Prolin mittels Aufnahme oder Biosynthese akkumuliert. Zu den osmosensitiven Transportern in *C. glutamicum* gehören BetP (Cg1016), EctP (Cg2539), LocP (Cg2563) und ProP (Cg3395), wovon BetP den aktivsten darstellt und Betain im Cotransport mit zwei Na⁺-Ionen aufnimmt (Krämer, 2009). Als Sensorsystem für die Antwort auf hyperosmotischen Stress wurde im Jahr 2004 von Möker und Mitarbeitern das Zweikomponentensystem MtrAB in *C. glutamicum* beschrieben (Möker *et al.*, 2004). Alle diese Komponenten der Osmoregulation zeigen im vergleichenden cDNA-Microarrayexperiment mit *Δcgb_16115_Reg123* nach hyperosmotischem Stress keine signifikant erhöhten M-Werte. So steht aus dieser Sicht die differenzielle Expression der *trans*-kodierte RNA *cgb_16115*, die besonders im Zusammenhang mit hyperosmotischen Stress beobachtet wurde, vermutlich nicht direkt im Zusammenhang mit der Osmoregulation in *C. glutamicum*. Interessanterweise ist aber gerade *betB* mit einem signifikant negativen M-Wert von -2 im cDNA-Microarrayexperiment mit *Δcgb_16115_Reg1* ohne Stress detektiert worden. Zudem ist *betP* in dem großen *k*-Means Cluster C7 enthalten, für das eine auffällige Anreicherung von Genen der funktionellen Klasse Metall-Ionen-Transport mithilfe des Gen-Ontologie (GO) Systems (Ashburner *et al.*, 2000) erhalten wurde.

Insgesamt kann vermutet werden, dass ein Großteil der zahlreich erhaltenen signifikant differenziellen Gene in den cDNA-Microarrayexperimenten nicht direkt auf eine Interaktion mit der kleinen RNA zurückzuführen ist. Der Effekt der putativen Interaktion mit der Membranprotease *ppmA* könnte möglicherweise aber verschiedene sekundäre Effekte auf das Transkriptom auslösen. Insgesamt zeigen

auch viele Proteinregulatoren eine auffällig veränderte relative Expression in den cDNA-Microarrayexperimenten, wobei die meisten davon putative Regulatoren darstellen, über die noch nicht sehr viel bekannt ist. Mit *ripA* (*cg1120*) wurde im vergleichenden Microarrayexperiment mit dem Deletionsstamm *cgb_16115_Reg123* nach NaCl-Stress jedoch ein bekanntes Regulatorgen getroffen, das für den AraC-Typ-Regulator zur Repression von Eisenproteinen kodiert (Wennerhold *et al.*, 2005). Fast alle Gene des RipA-Regulons zeigen deutlich negative M-Werte und erscheinen reprimiert. Von diesen Genen stimmen auch einige mit Interaktionsvorhersagen überein. Zusätzlich zeigen verschiedene Eisen-Transporter leicht erhöhte M-Werte. Aufgrund dieser Ergebnisse könnte vermutet werden, dass in der Deletionsmutante *cgb_16115_Reg123* nach hyperosmotischem Stress ein Eisenmangel auftritt. In dem ebenfalls Gram-positiven Bodenbakterium *B. subtilis* konnte gezeigt werden, dass erhöhte Salzkonzentrationen in der Umgebung zu Eisenmangel in den Zellen führen (Hoffmann *et al.*, 2002; Steil *et al.*, 2003). Eisen ist ein essentielles Element nicht nur für Bakterien, sondern alle Organismen und spielt als Cofaktor in Eisen-Schwefel-Proteinen in vielen biologischen Prozessen eine Rolle (Andrews *et al.*, 2003). In *E. coli* ist mit RhyB eine kleine RNA bekannt, die neben dem Fur-Regulator (*ferric uptake regulator*) einen Einfluss auf die Eisen-Homöostase zu haben scheint und zusätzlich selbst durch Fur reguliert wird (Massé *et al.*, 2005).

Die genaue Funktion der kleinen RNA *cgb_16115* in *C. glutamicum* bleibt unklar. Zusammengefasst deuten die komplexen Ergebnisse der vergleichenden cDNA-Microarrayanalysen im weiteren Sinn auf einen Zusammenhang mit der Regulation der Metall-Ionen-Homöostase hin und speziell der von membranständigen (Transport-) Proteinen. Zur putativen Interaktion von *cgb_16115* und der *ppmA*-mRNA wurden in dieser Arbeit verschiedene Hypothesen zum Regulationsmodell aufgestellt. Nach den bioinformatischen Vorhersagen ist eine Translationsaktivierung durch Bindung der kleinen RNA zu vermuten. Die MFE-Faltung der 5'-UTR von *ppmA* zeigt deren RBS über selbst-komplementäre Bindungen blockiert, und *cgb_16115* interagiert nach den Vorhersagen mit dem Bereich der Anti-RBS. Aus experimenteller Sicht ist im Wildtypstamm allerdings eine geringere Frequenz der Ribosomen auf der mRNA durch Bindung von *cgb_16115* anzunehmen, da die M-Werte von *ppmA* in den Deletionsmutanten jeweils erhöht sind. Wenn keine weiteren Transkriptionsfaktoren involviert sind, ist von einer höheren Stabilität der *ppmA*-mRNA ohne Bindung der kleinen RNA auszugehen. Dies würde dann auf eine erhöhte Translationsfrequenz und eine geringere Zugänglichkeit für Exo- und Endonukleasen in den Deletionsstämmen hinweisen.

Interaktionen kleiner RNA nahe der RBS einer mRNA wurden in der Literatur insgesamt häufiger beschrieben und bewirken meistens eine Blockierung der zugehörigen RBS und eine Reprimierung der Translation (Interaktionsbeispiele RyhB-*sodB* Tjaden *et al.*, 2006, SgrS-*ptsG* Kawamoto *et al.*, 2006, Spot42-*galK* Møller *et al.*, 2002 und *GcvB-dppA* Pulvermacher *et al.*, 2009). Die Bindung kleiner RNAs

in mehr als 100 nt Entfernung von der RBS hat in der Regel den Effekt einer thermodynamischen Umstrukturierung der gesamten 5'-UTR zur Folge und bewirkt über Öffnung der RBS eine Translationsaktivierende Wirkung (Interaktionsbeispiele DsrA-*rpoS* Majdalani *et al.*, 1998 und RprA-*rpoS* Majdalani *et al.*, 2002). Aufgrund der vorhergesagten Bindung von *cgb_16115* nahe der RBS von *ppmA* könnte man nach diesem Modell eher auf eine Blockierung dieser schließen, was auch zu den erhaltenen Ergebnissen der cDNA-Microarrayanalysen passt. Insgesamt wird deutlich, dass das Wissen über RNA-Regulationen in *C. glutamicum* im Allgemeinen noch sehr gering und nicht ausreichend ist, um die regulatorischen Vorgänge und die Bedeutung der Regulationen durch *cgb_16115* im Detail zu verstehen. Vermutlich spielen hierbei auch bisher unbekannte Komponenten eine Rolle wie z.B. RNA-Helferproteine.

6 Transkriptionelle Attenuation ist ein Mechanismus zur Regulation verschiedener Aminosäuresynthese-Gene in *C. glutamicum* und wurde hier für das Gen der DAHP-Synthetase (*aroF*) validiert

Ein verbreiteter Mechanismus zur Regulation der Genexpression ist die transkriptionelle Attenuation (Merino und Yanofsky, 2005). Hierbei erfährt das 5'-Ende einer mRNA eine Strukturänderung, die zur Bildung eines Rho-unabhängigen Terminators führt und dadurch ein verfrühtes Ende der Transkription bewirken kann (Naville und Gautheret, 2010). Der zu dieser Strukturänderung führende Mechanismus kann von der Ribosomenaktivität in einem kurzen Leaderpeptid abhängen (Yanofsky, 1981; Vitreschak *et al.*, 2004) oder auch von anderen regulatorischen Vorgängen in 5'-UTRs von Genen wie vom Binden kleiner Metabolit-Liganden an Aptamer-Sequenzen (Riboswitches) (Henkin, 2008) oder ungeladener tRNAs an T-Box-Sequenzen (Henkin, 1994; Vitreschak *et al.*, 2008). Auch Temperaturänderungen können eine Rolle bei der Attenuation spielen (RNA-Thermometer) (Naville und Gautheret, 2010). Vorhersagen indizieren, dass in Bakterien bis zu 10 % der Gene stromaufwärts Terminations-/Anti-Terminationsregulationen aufweisen (Henkin und Yanofsky, 2002). Es gibt Hinweise, dass das Histidine-Biosynthese-Operon *hisDCB-orf1-orf2-hisHA-impA-hisFI* in *C. glutamicum* AS019 über einen T-Box-vermittelten Attenuationsmechanismus reguliert wird (Jung *et al.*, 2010). Die Sequenzierungs-Ergebnisse dieser Arbeit haben in *C. glutamicum* ATCC 13032 zwar ein kurzes (Abbruch-) Transkript vor *hisD* (*cg2305*, Klasse mRNA-Leader) ergeben, aber aktuelle Ergebnisse von Kulis-Horn und Mitarbeiter haben einen solchen Mechanismus für diesen *C. glutamicum*-Stamm nicht zeigen können (Kulis-Horn *et al.*, 2014).

Als *cis*-regulatorische RNA-Motive sind transkriptionelle Attenuatoren generell auch in der Rfam-Databank (Gardner *et al.*, 2009) gelistet und können über diese mithilfe von Kovarianzmodellen vorhergesagt werden. Wie in der vorliegenden Arbeit schon dargelegt wurde, sind in *C. glutamicum* neun Riboswitches auf diese Weise bioinformatisch vorhergesagt worden, die auch alle in der Sequenzierung detektiert wurden und wahrscheinlich jeweils Polymerisationsabbruch-Transkripte darstellen. Mittels der Rfam-Datenbank wurden in *C. glutamicum* allerdings keine Mechanismen vorhergesagt, die eine Ribosomen-vermittelte Attenuation für Gene zur Biosynthese von Aminosäuren vorhersagen. Bisher wurde jedoch für drei Leaderpeptid-kodierende Transkriptionseinheiten eine Regulation durch transkriptionelle Attenuation in *C. glutamicum* vermutet und teilweise auch bewiesen. Diese betreffen das *trpE*-Gen (*cg3359*) des Trp-Operons, das *leuA*-Gene (*cg0303*) (Pátek *et al.*, 1994) und das *ilvB*-Gen (*cg1435*) des *ilvBNC*-Operons (Morbach *et al.*, 2000). Im RNA-Seq-Experiment zur Identifizierung kleiner RNAs konnte in allen drei Fällen eine kleine RNA detektiert werden. Die Bildung dieser Transkripte ist vermutlich darauf zurückzuführen, dass unter den gewählten Kultivierungs-Bedingungen und unter kurzzeitiger Stresseinwirkung auf die *C. glutamicum*-Zellen kein direkter Mangel an

aromatischen Aminosäuren entsteht. Aufgrund ausreichend geladener tRNA-Moleküle wird der Translationsprozess daher an den entsprechend aufeinanderfolgenden Codons in diesen Leaderpeptiden nicht verzögert, wodurch die Faltung der Terminatorstruktur begünstigt wird und es zur Transkriptionstermination kommt. Das historisch erste identifizierte Operon unter der Kontrolle transkriptioneller Attenuation ist das Trp-Operon in *E. coli*. Dieses enthält ein Leaderpeptid mit einer Reihe aufeinanderfolgender Trp-Codons (Hiraga und Yanofsky, 1973). In Analogie dazu zeigt die 5'-UTR von *trpE* (*cg3359*) in *C. glutamicum* ebenfalls einen kurzen ORF, der für ein Trp-haltiges Leaderpeptid kodiert und als TrpL (*Cg4012*, *Cbg_33575*) bezeichnet wird. Ergebnisse verschiedener Studien lassen vermuten, dass *trpL* die Expression des stromabwärts lokalisierten *trpE*-Gens in Abhängigkeit von der Verfügbarkeit geladener tRNA^{Trp} regliert (Sano und Matsui, 1987). In der Regulation des Trp-Operons im Gram-positiven Bakterium *Bacillus subtilis* gibt es dagegen deutliche Unterschiede, denn dort ist kein Leaderpeptid kodiert. Stattdessen existiert in *B. subtilis* ein Protein namens TRAP (*trp RNA-binding attenuation protein*), welches bei Trp-Mangel an Wiederholungen von GAG-Sequenzen bindet und so die Bildung einer Anti-Terminatorstruktur begünstigt (Babitzke *et al.*, 1994).

Eine weitere peptidkodierende kleine RNA wurde in dieser Arbeit stromaufwärts des *leuA*-Gens (*cg0303*) in *C. glutamicum* detektiert und wird als *leuL* (*cg4015*) bezeichnet. Bereits 1994 zeigten Pátek *et al.*, dass die spezifische Aktivität der durch *leuA* kodierten Isopropylmalat-Synthase mehr als zweifach erhöht war unter Leucin-Mangel (Pátek *et al.*, 1994). Zudem konnte dieser regulatorische Effekt von Leucin wurde nach Mutagenese Leaderpeptid aufgehoben werden (ebd.). Ein anderer untersuchter Attenuationsmechanismus in *C. glutamicum* wurde für das *ilvBNC*-Operons identifiziert (*cg1435*) und ist in die Biosynthese der verzweigten Aminosäuren involviert (Morbach *et al.*, 2000). Die Region stromaufwärts von *ilvB* kodiert für ein Leaderpeptid (*ilvL/cg4016*), das mehrere Codons für die verzweigten Aminosäuren Isoleucin, Leucin und Valin enthält. Unter Mangelbedingungen für eine oder alle der entsprechenden Aminosäuren konnte gezeigt werden, dass die Expression des Operons mehr als zweifach erhöht war (ebd.). Darüber hinaus vermuteten Morbach und Mitarbeiter ein *leaderless*-Transkript für dieses *ilvL*. In Übereinstimmung damit zeigen die RNA-Seq-Ergebnisse, dass der TSS von *ilvL* bzw. *ilvB* direkt am Beginn des Leaderpeptid-Start-Codons liegt. Nach Analyse der RNA-Seq-Daten liegt ein *leaderless*-Transkript ebenso für die *trpL*- und *leuL*-mRNAs vor. *Leaderless*-Transkripte wurden schon in verschiedenen Bakterien beobachtet. Die Bedeutung von *leaderless*-Transkripten und der genaue Mechanismus sind bislang nicht eindeutig geklärt (Zheng *et al.*, 2011). Über *in silico*-Analysen wurde besonders in Actinobakterien ein relativ hoher Prozentsatz von bis zu 20 % für diese Art von Transkripten festgestellt (ebd.). Erst ganz aktuell wurde in einer RNA-Sequenzierung des Gesamt-Transkriptoms von *C. glutamicum* gezeigt, dass *leaderless*-Transkripte in diesem Bakterium mit einem besonders hohen Prozentsatz von 33 % vertreten sind (Pfeifer-Sancar *et al.*, 2013).

In dieser Arbeit wurde zusätzlich noch ein weiteres *leaderless*-Transkript der peptidkodierenden kleinen RNA *cgb_11293* (*aroL*) detektiert und zwar am 5'-Ende des *aroF*-Gens (*cg1129*), dessen Genprodukt eine von zwei DAHP-Synthasen in *C. glutamicum* darstellt und den ersten Biosyntheseschritt des Shikimat-Biosyntheseweges aromatischer Aminosäuren katalysiert. Dieses Leadepeptid weist zudem eine Anzahl aufeinanderfolgender aromatischer Aminosäure-Codons auf (Mentz *et al.*, 2013) und konnte in einigen Corynebakterien-Spezies als konserviert beobachtet werden. Wie für die meisten der anderen detektierten Leaderpeptid-mRNAs besteht auch für *aroL* eine Sekundärstrukturkonservierung über die RNAz-Vorhersage regulatorischer RNAs, und es existiert eine Vorhersage mit dem SIPHT-Tool (*Candidate_10_NC_006958*). Die Funktion des putativen Attenuationsmechanismus für das *aroF*-Gen wurde über transkriptionelle Analysen untersucht, und diese weisen sehr darauf hin, dass es sich hier um einen solchen transkriptionellen Attenuationsmechanismus handelt (Mentz *et al.*, Manuskript in Vorbereitung). Insgesamt lässt sich zum aufgestellten Regulationsmodell der transkriptionellen Attenuation für *aroF* sagen, dass die vorhergesagten Anti-Terminator-Sequenzkomplementaritäten nicht so eindeutig sind wie für andere in der Literatur beschriebene Attenuatoren. Um die alternative Anti-Terminator-Faltung zu gewährleisten, wird *in vivo* möglicherweise unterstützend ein RNA-Helferprotein benötigt.

Bisher wurde eine transkriptionelle Attenuation in Abhängigkeit vom zellulären Gehalt geladener Phenylalanin- und Tyrosin-tRNAs für DAHP-Synthasegene überhaupt noch nicht gezeigt. Liu und Mitarbeiter haben im Jahr 2008 für die AroF DAHP-Synthase beobachtet, dass diese besonders sensitiv ist gegenüber einer Feedbackinhibierung durch Tyrosin und etwas weniger durch Phe und Trp (Liu *et al.*, 2008). Im Gegensatz dazu ist die AroG DAHP-Synthase (*cg2391*) nur gering sensitiv gegenüber Trp und nicht sensitiv gegenüber Tyr und Phe. Außerdem scheint AroG essenziell zu sein (ebd.). Im Gegensatz zu *C. glutamicum* gibt es in *E. coli* drei isofunktionelle DAHP-Synthasen, die jeweils in Abhängigkeit von der intrazellulären Konzentration der korrespondierenden Aminosäuren allosterisch inhibiert werden können (Panina *et al.*, 2001). In *E. coli* wird allerdings das *pheA* kodierte Phenylalanin-Biosynthese-Enzym Chorismat-Mutase/Prephenat-Dehydratase durch transkriptionelle Attenuation reguliert, in diesem Fall aber spezifisch in Abhängigkeit von der intrazellulären Konzentrationen von tRNA^{Phe} (Gavini und Davidson, 1991).

Die Analyse des Stammes *C. glutamicum* ATCC 21850, der Tryptophan überproduziert, führte im Jahr 1993 zur Identifizierung der dafür verantwortlichen Mutation in der Trp-Leaderpeptidsequenz (Heery und Dunican, 1993). Da heute verschiedene Stämme von *C. glutamicum* zur industriellen Produktion von Aminosäuren genutzt werden, könnten die erhaltenen Ergebnisse dieser Arbeit zur transkriptionellen Attenuation für die Entwicklung von Aminosäure-Produktionsstämmen relevant sein.

VI. Ausblick

Mithilfe der Ergebnisse dieser Arbeit wurde eine Vielzahl kleiner RNAs in *C. glutamicum* ATCC 13032 bestimmt, und so wurde zum ersten Mal die beachtliche genetische Ausstattung für potentielle RNA-Regulationen deutlich, die dieses Actinobakterium besitzt. Da zum Zeitpunkt der Auswertung der RNA-Seq kleiner RNAs noch keine Daten zur Gesamt-Transkriptom-Sequenzierung in *C. glutamicum* vorlagen, wurden einheitliche UTR-Längen für annotierte Gene definiert, um die erhaltenen kleinen RNAs zu klassifizieren. Erst vor kurzem wurde eine Gesamt-Transkriptom RNA-Seq-Analyse in *C. glutamicum* speziell für RNAs > 200 nt fertiggestellt (Pfeifer-Sancar *et al.*, 2013) und definiert nun die genaue Länge der UTRs von über 2000 proteinkodierenden Genen. Die erhaltenen 5'-UTR-Längen sind recht variabel, das arithmetische Mittel liegt bei ~60 nt. Allerdings gibt es auch ~500 5'-UTRs, die länger als 100 nt sind. Im Rahmen der neuen Genomannotation für *C. glutamicum* werden die Ergebnisse beider Studien (Pfeifer-Sancar *et al.*, 2013 und Mentz *et al.*, 2013) nun zusammengeführt und können einige Änderungen bezogen auf die Klassifizierung kleiner RNAs ergeben.

Die funktionelle Untersuchung der *trans*-kodierte RNA *cgb_16115* wurde in dieser Arbeit begonnen, und auf Basis bioinformatischer und experimenteller Analysen sind vor allem Hinweise einer Interaktion mit der *cg3138*-mRNA, einer putativen Membran-Protease, erhalten worden. Der direkte Nachweis dieser Interaktion ist allerdings noch nicht gelungen und das Modell zur Regulation ist weiterhin unklar. *In vitro* toeprinting-Assays (*primer extension inhibition*) könnten die Auswirkung der Deletion von *cgb_16115* auf die Bindung von Ribosomen an der *cg3138*-mRNA deutlich machen. Aufgrund einer *in silico*-Promotoranalyse wird vermutet, dass die kleine RNA in Abhängigkeit von einem SigA- und einem ECF-ähnlichen Promotor steht. Es entsteht die Frage, ob es über diese Sigmafaktoren hinaus noch Proteinregulatoren gibt, die diese kleine RNA regulieren. Um dies zu untersuchen, eignet sich die Methode der DNA-Affinitätschromatographie. Diese eignet sich zur Identifizierung von bindenden Regulatorproteinen an die DNA und wird damit als *regulator fishing* bezeichnet. Die Methode wurde zuerst für Zellkulturen durch Gishlain und Fish entwickelt (Gishlain und Fish, 1996) und außerdem erfolgreich in Corynebakterien angewendet (Brinkrolf *et al.*, 2008; Brune *et al.*, 2011).

Für den Nachweis einer Interaktion mit *cg3138* kann außerdem das Wissen über RNA-Helferproteine in *C. glutamicum* hilfreich sein, die möglicherweise eine Rolle bei der Vermittlung dieser und wahrscheinlich auch anderer regulatorischer RNA-RNA-Interaktionen spielen können. Mit der Identifizierung dieser Helferproteine wurde teilweise schon begonnen, und diese muss nun fortgeführt werden. Die Idee besteht darin, die RNA-bindenden Helferproteine zu „fischen“, in Anlehnung an das Fischen von Regulatorproteinen (ebd.). So sollen entweder *in vitro* amplifizierte regulatorische RNAs oder die *in vivo* RNA-Fraktion kleiner RNAs jeweils am 3'-Ende mit einem Biotin-Tag ligiert werden.

Eine andere Möglichkeit stellt die Ligation biotinierter RNA-Adapter an das 5'-Ende der kleiner RNAs dar. Anschließend sollen die biotinierten kleinen RNAs an magnetische Streptavidin-Beads und darüber an eine magnetische Säule gekoppelt werden. Spezifisch RNA-bindende Proteine aus einem aufgereinigten *C. glutamicum*-Proteinrohextrakt sollen dann von der Säule eluiert und im Anschluss mithilfe massenspektrometrischer Analysen identifiziert werden. Vor Beginn dieses Experimentes muss natürlich die ribosomale RNA abgereichert werden, um nicht die zahlreichen vorhandenen ribosomal bindenden Proteine zu detektieren.

In den vergleichenden cDNA-Microarrayexperimenten der beiden Deletionsmutanten von *cgb_16115* im Vergleich zum *C. glutamicum*-Wildtypstamm hat sich eine auffällige differenzielle Expression für Gene des Metall-Ionen-Transports ergeben bzw. speziell für Gene der Eisen-Homöostase. In diesem Zusammenhang könnten vergleichende Wachstumsanalysen und auch vergleichende transkriptionelle Expressionsanalysen mit den in dieser Arbeit erzeugten Deletionsstämmen unternommen werden, die auf unterschiedlicher Konzentration von Metall-Ionen im Kultivierungsmedium beruhen, z.B. durch Verwendung von CGXII-Minimalmedium ohne Eisen in den zugegebenen Spurenelementen. Zudem sollten die in dieser Arbeit schon begonnenen 2D-Gelelektrophoreseexperimente im Rahmen einer Funktionsanalyse von *cgb_16115* mit der Fraktion cytosolischer Proteine aus *C. glutamicum* nun für die Fraktion von Membranproteinen fortgeführt werden.

In Bezug auf Funktionsanalysen kleiner RNAs können generell auch Frag-Seq (*fragmentation sequencing*)-Experimente eine Rolle spielen (Underwood *et al.*, 2010). Die vorhergesagte Faltung der Sekundärstrukturen von RNA-Elementen könnten auf diese Weise verifiziert oder korrigiert werden, wodurch sich möglicherweise auch Schlüsse auf putative Interaktionen ergeben können.

Die Analyse der differenziellen Expression kleiner RNAs wurde mit den Ergebnissen der intergenischen *tiling*-Microarrays eine wichtige Informations-Grundlage geschaffen. Weiterhin wird die Durchführung von Northern Blot-Analysen noch mehr Informationen liefern um kleine RNAs im Detail zu analysieren. Für die Zukunft sollte aber auch über die Durchführung paralleler Multiplex-Hochdurchsatz-Sequenzierungen in *C. glutamicum* nachgedacht werden, wobei externe *spike-in*-Kontrollen zur quantitativen Auswertung und Erstellung von Expressionsmuster über die Transkripte verwendet werden könnten (Wang *et al.*, 2011).

VII. Literaturverzeichnis

- Ahmad R, Hansen GÅ, Hansen H, Hjerde E, Pedersen HL, Paulsen SM, Nyrud MLJ, Strauss A, Willassen N, Haugen P (2012):** Prediction, microarray and northern blot analyses identify new intergenic small RNAs in *Aliivibrio salmonicida*. *J. Mol. Microbiol. Biotechnol.* **22**: 352–360.
- Ajinomoto (2012):** Fact sheet: Food Products Business. <http://www1.aminovital.com/en/ir/pdf/FoodOct2012.pdf>.
- Allison DB, Cui X, Page GP, Sabripour M (2006):** Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.* **7**: 55–65.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990):** Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G (1997):** A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90**: 43–53.
- Altuvia S, Zhang A, Argaman L, Tiwari A, Storz G (1998):** The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.* **17**: 6069–6075.
- Andersen J, Delihans N (1990):** micF RNA binds to the 5' end of ompF mRNA and to a protein from *Escherichia coli*. *Biochemistry* **29**: 9249–9256.
- Andersen J, Forst SA, Zhao K, Inouye M, Delihans N (1989):** The function of micF RNA. micF RNA is a major factor in the thermal regulation of OmpF protein in *Escherichia coli*. *J. Biol. Chem.* **264**: 17961–17970.
- Andrews SC, Robinson AK, Rodríguez-Quinones F (2003):** Bacterial iron homeostasis. *FEMS Microbiol. Rev.* **27**: 215–237.
- Ao W, Gaudet J, Kent WJ, Muttumu S, Mango SE (2004):** Environmentally induced foregut remodeling by PHA-4/FoxA and DAF-12/NHR. *Science* **305**: 1743–1746.
- Arnvig KB, Young DB (2009):** Identification of small RNAs in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **73**: 397–408.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000):** Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**: 25–29.
- Aslund F, Beckwith J (1999):** Bridge over troubled waters: sensing stress by disulfide bond formation. *Cell* **96**: 751–753.
- Babitzke P, Stults JT, Shire SJ, Yanofsky C (1994):** TRAP, the *trp* RNA-binding attenuation protein of *Bacillus subtilis*, is a multisubunit complex that appears to recognize G/UAG repeats in the *trpEDCFBA* and *trpG* transcripts. *J. Biol. Chem.* **269**: 16597–16604.
- Babitzke P, Yanofsky C (1993):** Reconstitution of *Bacillus subtilis* *trp* attenuation *in vitro* with TRAP, the *trp* RNA-binding attenuation protein. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 133–137.
- Backofen R, Hess WR (2010):** Computational prediction of sRNAs and their targets in bacteria. *RNA Biol* **7**: 33–42.
- Balandina A, Kamashev D, Rouviere-Yaniv J (2002):** The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J. Biol. Chem.* **277**: 27622–27628.
- Banerjee S, Chalissery J, Bandey I, Sen R (2006):** Rho-dependent transcription termination: more questions than answers. *J. Microbiol.* **44**: 11–22.
- Barksdale L (1970):** *Corynebacterium diphtheriae* and its relatives. *Bacteriol Rev* **34**: 378–422.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007):** CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712.
- Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, Breaker RR (2004):** New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 6421–6426.

- Baum M, Bielau S, Rittner N, Schmid K, Eggelbusch K, Dahms M, Schlauersbach A, Tahedl H, Beier M, Guimil R, Scheffler M, Hermann C, Funk J, Wixmerten A, Rebscher H, Honig M, Andreae C, Buchner D, Moschel E, Glathe A, Jager E, Thom M, Greil A, Bestvater F, Obermeier F, Burgmaier J, Thome K, Weichert S, Hein S, Binnewies T, Foitzik V, Muller M, Stahler C, Stahler P (2003):** Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Res* **31**(23).
- Beisel CL, Storz G (2010):** Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev.* **34**: 866–882.
- Bellman R (1957):** Dynamic programming. Princeton University Press.
- Benito Y, Kolb FA, Romby P, Lina G, Etienne J, Vandenesch F (2000):** Probing the structure of RNAIII, the *Staphylococcus aureus* agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. *RNA* **6**: 668–679.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013):** GenBank. *Nucleic Acids Res.* **41**: D36–42.
- Bernard K (2012):** The genus corynebacterium and other medically relevant coryneform-like bacteria. *J. Clin. Microbiol.* **50**: 3152–3158.
- Bertani G (2004):** Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **186**: 595–600.
- Bignell C, Thomas CM (2001):** The bacterial ParA-ParB partitioning proteins. *J. Biotechnol.* **91**: 1–34.
- Birnboim HC, Doly J (1979):** A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**: 1513–1523.
- Bjelland S, Seeberg E (1996):** Different efficiencies of the Tag and AlkA DNA glycosylases from *Escherichia coli* in the removal of 3-methyladenine from single-stranded DNA. *FEBS Lett.* **397**: 127–129.
- Blomberg P, Wagner EG, Nordström K (1990):** Control of replication of plasmid R1: the duplex between the antisense RNA, CopA, and its target, CopT, is processed specifically *in vivo* and *in vitro* by RNase III. *EMBO J.* **9**: 2331–2340.
- Blom J, Jakobi T, Doppmeier D, Jaenicke S, Kalinowski J, Stoye J, Goesmann A (2011):** Exact and complete short-read alignment to microbial genomes using Graphics Processing Unit programming. *Bioinformatics* **27**: 1351–1358.
- Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005):** Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551–2561.
- Bolstad B, Irizarry R, Astrand M, Speed T (2003):** A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**: 185–193.
- Bouché F, Bouché JP (1989):** Genetic evidence that DicF, a second division inhibitor encoded by the *Escherichia coli* *dicB* operon, is probably RNA. *Mol. Microbiol.* **3**: 991–994.
- Bralley P, Gust B, Chang S, Chater KF, Jones GH (2006):** RNA 3'-tail synthesis in *Streptomyces*: *in vitro* and *in vivo* activities of RNase PH, the SCO3896 gene product and polynucleotide phosphorylase. *Microbiology* **152**: 627–636.
- Brantl S (2008).** *BIOspektrum*, 07.08., 14. Jahrgang.
- Brantl S, Wagner EG (1994):** Antisense RNA-mediated transcriptional attenuation occurs faster than stable antisense/target RNA pairing: an *in vitro* study of plasmid pIP501. *EMBO J.* **13**: 3599–3607.
- Brantl S, Wagner EGH (2002):** An antisense RNA-mediated transcriptional attenuation mechanism functions in *Escherichia coli*. *J. Bacteriol.* **184**: 2740–2747.
- Brendel V, Hamm GH, Trifonov EN (1986):** Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. *J. Biomol. Struct. Dyn.* **3**: 705–723.
- Brinkrolf K, Plöger S, Solle S, Brune I, Nentwich SS, Hüser AT, Kalinowski J, Pühler A, Tauch A (2008):** The LacI/GalR family transcriptional regulator UriR negatively controls uridine utilization of *Corynebacterium glutamicum* by binding to catabolite-responsive element (cre)-like sequences. *Microbiology* **154**: 1068–1081.

- Brinkrolf K, Rupp O, Laux H, Kollin F, Ernst W, Linke B, Kofler R, Romand S, Hesse F, Budach WE, Galosy S, Müller D, Noll T, Wienberg J, Jostock T, Leonard M, Grillari J, Tauch A, Goesmann A, Helk B, Mott JE, Pühler A, Borth N (2013):** Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* **31**: 694–695.
- Brockmann-Gretza O, Kalinowski J (2006):** Global gene expression during stringent response in *Corynebacterium glutamicum* in presence and absence of the *rel* gene encoding (p)ppGpp synthase. *BMC Genomics* **7**: 230.
- Brune I, Barzantny H, Klötzel M, Jones J, James G, Tauch A (2011):** Identification of McbR as transcription regulator of *aecD* and genes involved in methionine and cysteine biosynthesis in *Corynebacterium jeikeium* K411. *J. Biotechnol.* **151**: 22–29.
- Brune I, Werner H, Hüser AT, Kalinowski J, Pühler A, 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.
- Busch A, Richter AS, Backofen R (2008):** IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* **24**: 2849–2856.
- Busche T, Šilar R, Pičmanová M, Pátek M, Kalinowski J (2012):** Transcriptional regulation of the operon encoding stress-responsive ECF sigma factor SigH and its anti-sigma factor RshA, and control of its regulatory network in *Corynebacterium glutamicum*. *BMC Genomics* **13**: 445.
- Byun Y, Han K (2009):** PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics* **25**: 1435–1437.
- Callen BP, Shearwin KE, Egan JB (2004):** Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell* **14**: 647–656.
- Cao Y, Wu J, Liu Q, Zhao Y, Ying X, Cha L, Wang L, Li W (2010):** sRNATarBase: a comprehensive database of bacterial sRNA targets verified by experiments. *RNA* **16**: 2051–2057.
- Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E (2008):** Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**: 935–938.
- Cathala G, Savouret JF, Mendez B, West BL, Karin M, Martial JA, Baxter JD (1983):** A method for isolation of intact, translationally active ribonucleic acid. *DNA* **2**: 329–335.
- Cerdeño-Tárraga AM, Efstratiou A, Dover LG, Holden MTG, Pallen M, Bentley SD, Besra GS, Churcher C, James KD, Zoysa A de, Chillingworth T, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Quail MA, Rabinowitsch E, Rutherford KM, Thomson NR, Unwin L, Whitehead S, Barrell BG, Parkhill J (2003):** The complete genome sequence and analysis of *Corynebacterium diphtheriae* NCTC13129. *Nucleic Acids Res.* **31**: 6516–6523.
- Chai Y, Winans SC (2005):** A small antisense RNA downregulates expression of an essential replicase protein of an *Agrobacterium tumefaciens* Ti plasmid. *Mol. Microbiol.* **56**: 1574–1585.
- Chaudhuri K (2012):** Microbial Genetics. The Energy and Resources Institute, TERI.
- Chaudhuri RR, Yu L, Kanji A, Perkins TT, Gardner PP, Choudhary J, Maskell DJ, Grant AJ (2011):** Quantitative RNA-seq analysis of the *Campylobacter jejuni* transcriptome. *Microbiology* **157**: 2922–2932.
- Cheng Z, Deutscher MP (2005):** An important role for RNase R in mRNA decay. *Mol. Cell* **17**: 313–318.
- Chen Z, Liu J, Ng HKT, Nadarajah S, Kaufman HL, Yang JY, Deng Y (2011):** Statistical methods on detecting differentially expressed genes for RNA-seq data. *BMC Syst Biol* **5** Suppl 3: S1.
- Chomczynski P, Sacchi N (1987):** Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156–159.
- Cleveland WS, Devlin SJ (1988):** Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association* **Vol. 83**, No. 403: 596–610.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008):** Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215–219.

- Costa V, Angelini C, Feis I de, Ciccodicola A (2010):** Uncovering the complexity of transcriptomes with RNA-Seq. *J. Biomed. Biotechnol.* 2010: 853916.
- Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ (2008):** Identification of genetic variants using bar-coded multiplexed sequencing. *Nat. Methods* 5: 887–893.
- Crick F (1970):** Central dogma of molecular biology. *Nature* 227: 561–563.
- Crooks GE, Hon G, Chandonia J, Brenner SE (2004):** WebLogo: a sequence logo generator. *Genome Res.* 14: 1188–1190.
- Dam E, Pleij K, Draper D (1992):** Structural and functional aspects of RNA pseudoknots. *Biochemistry* 31: 11665–11676.
- Darling AC, Mau B, Blattner FR, Perna NT (2004):** Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research* 14: 1394–1403.
- Deana A, Celesnik H, Belasco JG (2008):** The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* 451: 355–358.
- Decréau RA, Marson CM, Smith KE, Behan JM (2003):** Production of malodorous steroids from androsta-5,16-dienes and androsta-4,16-dienes by *Corynebacteria* and other human axillary bacteria. *J. Steroid Biochem. Mol. Biol.* 87: 327–336.
- Deighan P, Free A, Dorman CJ (2000):** A role for the *Escherichia coli* H-NS-like protein StpA in OmpF porin expression through modulation of micF RNA stability. *Mol. Microbiol.* 38: 126–139.
- Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003):** DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4: P3.
- Deutscher MP (2006):** Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res.* 34: 659–666.
- Deutscher MP (2009):** Maturation and degradation of ribosomal RNA in bacteria. *Prog Mol Biol Transl Sci* 85: 369–391.
- DiChiara JM, Contreras-Martinez LM, Livny J, Smith D, McDonough KA, 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.
- Dodd IB, Egan JB (2002):** Action at a distance in CI repressor regulation of the bacteriophage 186 genetic switch. *Mol Microbiol.* 45(3):697-710.
- Dondrup M, Albaum SP, Griebel T, Henckel K, Jünemann S, Kahlke T, Kleindt CK, Küster H, Linke B, Mertens D, Mittard-Runte V, Neuweger H, Runte KJ, Tauch A, Tille F, Pühler A, Goesmann A (2009):** EMMA 2--a MAGe-compliant system for the collaborative analysis and integration of microarray data. *BMC Bioinformatics* 10: 50.
- Dubnau E, Chan J, Raynaud C, Mohan VP, Lanéelle MA, Yu K, Quémar A, Smith I, Daffé M (2000):** Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. *Mol. Microbiol.* 36: 630–637.
- Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R (2013):** Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.* 9: e1003602.
- Eggeling L, Bott M (2005):** Handbook of *Corynebacterium glutamicum*. CRC Press.
- Eggeling L, Gurdial S, Alderwick L (2008):** "Structure and synthesis of the cell wall" in *Corynebacteria*. Caister Academic Press, Norfolk, UK A. Burkovski, Ed.: 267–294.
- Eggenhofer F, Tafer H, Stadler PF, Hofacker IL (2011):** RNAPredator: fast accessibility-based prediction of sRNA targets. *Nucleic acids research* 39(Web Server issue):W149-54.
- Ehira S, Teramoto H, Inui M, Yukawa H (2009):** Regulation of *Corynebacterium glutamicum* Heat Shock Response by the Extracytoplasmic-Function Sigma Factor SigH and Transcriptional Regulators HspR and HrcA. *Journal of bacteriology* 191: 2964–2972.

- Emory SA, Bouvet P, Belasco JG (1992):** A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev.* **6**: 135–148.
- Engels V, Wendisch VF (2007):** The DeoR-type regulator SugR represses expression of *ptsG* in *Corynebacterium glutamicum*. *J. Bacteriol.* **189**: 2955–2966.
- Evans D, Marquez SM, Pace NR (2006):** RNase P: interface of the RNA and protein worlds. *Trends Biochem. Sci.* **31**: 333–341.
- Evguenieva-Hackenberg E, Klug G (2011):** New aspects of RNA processing in prokaryotes. *Curr. Opin. Microbiol.* **14**: 587–592.
- Fire A, Xu S, Montgomery M, Kostas S, Driver S, Mello C (1998):** Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.
- Fischer F, Wolters D, Rögner M, Poetsch A (2006):** Toward the complete membrane proteome: high coverage of integral membrane proteins through transmembrane peptide detection. *Mol. Cell Proteomics* **5**: 444–453.
- Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D (1991):** Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**: 767–773.
- Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, Turner DH (1986):** Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 9373–9377.
- Frohman MA (1993):** Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Meth. Enzymol.* **218**: 340–356.
- Frunzke J, Engels V, Hasenbein S, Gätgens C, Bott M (2008):** Co-ordinated regulation of gluconate catabolism and glucose uptake in *Corynebacterium glutamicum* by two functionally equivalent transcriptional regulators, GntR1 and GntR2. *Mol. Microbiol.* **67**: 305–322.
- Fudou R, Jojima Y, Seto A, Yamada K, Kimura E, Nakamatsu T, Hiraishi A, Yamanaka S (2002):** *Corynebacterium efficiens* sp. nov., a glutamic-acid-producing species from soil and vegetables. *Int. J. Syst. Evol. Microbiol.* **52**: 1127–1131.
- Gaigalat L, Schlüter J, Hartmann M, Mormann S, Tauch A, Pühler A, Kalinowski J (2007):** The DeoR-type transcriptional regulator SugR acts as a repressor for genes encoding the phosphoenolpyruvate:sugar phosphotransferase system (PTS) in *Corynebacterium glutamicum*. *BMC Mol. Biol.* **8**: 104.
- Gao X, Gulari E, Zhou X (2004):** *In situ* synthesis of oligonucleotide microarrays. *Biopolymers* **73**: 579–596.
- Garber M, Grabherr MG, Guttman M, Trapnell C (2011):** Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8**: 469–477.
- Gaubig LC, Waldminghaus T, Narberhaus F (2011):** Multiple layers of control govern expression of the *Escherichia coli* *ibpAB* heat-shock operon. *Microbiology* **157**: 66–76.
- Gavini N, Davidson BE (1991):** Regulation of *pheA* expression by the *pheR* product in *Escherichia coli* is mediated through attenuation of transcription. *J. Biol. Chem.* **266**: 7750–7753.
- Gebhardt H, Meniche X, Tropis M, Krämer R, Daffé M, Morbach S (2007):** The key role of the mycolic acid content in the functionality of the cell wall permeability barrier in *Corynebacterineae*. *Microbiology* **153**: 1424–1434.
- Gelfand MS, Mironov AA, Jomantas J, Kozlov YI, Perumov DA (1999):** A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.* **15**: 439–442.
- Gerdes K, Nielsen A, Thorsted P, Wagner EG (1992):** Mechanism of killer gene activation. Antisense RNA-dependent RNase III cleavage ensures rapid turn-over of the stable *hok*, *srnB* and *pndA* effector messenger RNAs. *J. Mol. Biol.* **226**: 637–649.
- Ghislain JJ, Fish EN (1996):** Application of genomic DNA affinity chromatography identifies multiple interferon- α -regulated Stat2 complexes. *J. Biol. Chem.* **271**: 12408–12413.
- Ghosh T, Bose D, Zhang X (2010):** Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiol. Rev.* **34**: 611–627.
- Giegerich R, Voß B, Rehmsmeier M (2004):** Abstract shapes of RNA. *Nucleic acids research* **32**: 4843–4851.
- Gilbert W (1986):** Origin of life: The RNA world. *Nature*: 618.

- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003):** Rfam: an RNA family database. *Nucleic Acids Res.* **31**: 439–441.
- Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A (2009):** Rfam: updates to the RNA families database. *Nucleic Acids Res* **37**(Database issue):D136-40
- Grissa I, Vergnaud G, Pourcel C (2007):** The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.
- Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008):** The Vienna RNA websuite. *Nucleic Acids Res.* **36**: W70-4.
- Güell M, Yus E, Lluch-Senar M, Serrano L (2011):** Bacterial transcriptomics: what is beyond the RNA horizon? *Nat. Rev. Microbiol.* **9**: 658–669.
- Halgasova N, Bukovska G, Timko J, Kormanec J (2001):** Cloning and transcriptional characterization of two sigma factor genes, sigA and sigB, from *Brevibacterium flavum*. *Curr. Microbiol.* **43**: 249–254.
- Hambraeus G, Wachenfeldt C von, Hederstedt L (2003):** Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol. Genet. Genomics* **269**: 706–714.
- Hanahan D (1983):** Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**: 557–580.
- Hänssler E, Müller T, Jessberger N, Völzke A, Plassmeier J, Kalinowski J, Krämer R, Burkovski A (2007):** FarR, a putative regulator of amino acid metabolism in *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* **76**: 625–632.
- Harel D, Koren Y (2000):** A fast multi-scale method for drawing large graphs, Proceedings of the working conference on Advanced visual interfaces. ACM, New York, NY, U.S.A., 282–285.
- Hartz D, McPheeters DS, Traut R, Gold L (1988):** Extension inhibition analysis of translation initiation complexes. *Meth. Enzymol.* **164**: 419–425.
- Heery DM, Dunican LK (1993):** Cloning of the trp gene cluster from a tryptophan-hyperproducing strain of *Corynebacterium glutamicum*: identification of a mutation in the trp leader sequence. *Appl. Environ. Microbiol.* **59**: 791–799.
- Henkin TM (1994):** tRNA-directed transcription antitermination. *Mol. Microbiol.* **13**: 381–387.
- Henkin TM (2008):** Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev.* **22**: 3383–3390.
- Henkin TM, Yanofsky C (2002):** Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. *Bioessays* **24**: 700–707.
- Hernández JA, Muro-Pastor AM, Flores E, Bes MT, Peleato ML, Fillat MF (2006):** Identification of a *furA* cis-antisense RNA in the cyanobacterium *Anabaena* sp. PCC 7120. *J. Mol. Biol.* **355**: 325–334.
- Hilker R, Stademann KB, Doppmeier D, Kalinowski J, Stoye J, Straube J., Winnebold J, Goesmann A (Manuskript submitted):** ReadXplorer - Visualization and Analysis of Mapped Sequences.
- Hippel PH von, Yager TD (1991):** Transcript elongation and termination are competitive kinetic processes. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 2307–2311.
- Hiraga S, Yanofsky C (1973):** Inhibition of the progress of transcription on the tryptophan operon of *Escherichia coli*. *J. Mol. Biol.* **79**: 339–349.
- Hirschman J, Wong PK, Sei K, Keener J, Kustu S (1985):** Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription in vitro: evidence that the *ntrA* product is a sigma factor. *Proc. Natl. Acad. Sci. U.S.A.* **82**: 7525–7529.
- Hoe C, Raabe CA, Rozhdestvensky TS, Tang T (2013):** Bacterial sRNAs: regulation in stress. *Int. J. Med. Microbiol.* **303**: 217–229.
- Hoffmann T, Schütz A, Brosius M, Völker A, Völker U, Bremer E (2002):** High-salinity-induced iron limitation in *Bacillus subtilis*. *J. Bacteriol.* **184**: 718–727.
- Hongay CF, Grisafi PL, Galitski T, Fink GR (2006):** Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* **127**: 735–745.

- Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009):** High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19:** 1068–1076.
- Hunt TP, Magasanik B (1985):** Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnG*, and *glnL*. *Proc. Natl. Acad. Sci. U.S.A.* **82:** 8453–8457.
- Hüser AT, Becker A, Brune I, Dondrup M, Kalinowski J, Plassmeier J, Pühler A, Wieggräbe I, Tauch A (2003):** Development of a *Corynebacterium glutamicum* DNA microarray and validation by genome-wide expression profiling during growth with propionate as carbon source. *J. Biotechnol.* **106:** 269–286.
- Huynen MA, Bork P (1998):** Measuring genome evolution. *Proc. Natl. Acad. Sci. U.S.A.* **95:** 5849–5856.
- Ikeda M (2006):** Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. *Appl. Microbiol. Biotechnol.* **69:** 615–626.
- Ikeda M, Baba M, Tsukamoto N, Komatsu T, Mitsunashi S, Takeno S (2009):** Elucidation of genes relevant to the microaerobic growth of *Corynebacterium glutamicum*. *Biosci. Biotechnol. Biochem.* **73:** 2806–2808.
- Ikeda M, Nakagawa S (2003):** The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl. Microbiol. Biotechnol.* **62:** 99–109.
- Ikeda M, Takeno S. (2013):** Amino Acid Production by *Corynebacterium glutamicum*: Biology and Biotechnology. *Microbiology Monographs:* 107–147.
- Inoue H, Nojima H, Okayama H (1990):** High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96:** 23–28.
- Jäger D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA (2009):** Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. U.S.A.* **106:** 21878–21882.
- Jain C (2002):** Degradation of mRNA in *Escherichia coli*. *IUBMB Life* **54:** 315–321.
- Barciszewski J (Hrsg.) (2003):** Non-Coding RNAs: Molecular Biology and Molecular Medicine. Noncoding RNAs Encoded by Bacterial Chromosomes (Wagner, R.; Vogel, J.). Springer.
- Jensen LJ, Julien P, Kuhn M, Mering C von, Muller J, Doerks T, Bork P (2008):** eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* **36:** D250–4.
- Jishage M, Iwata A, Ueda S, Ishihama A (1996):** Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J. Bacteriol.* **178:** 5447–5451.
- Jochmann N, Kurze A, Czaja LF, Brinkrolf K, Brune I, Hüser AT, Hansmeier N, Pühler A, Borovok I, Tauch A (2009):** Genetic makeup of the *Corynebacterium glutamicum* LexA regulon deduced from comparative transcriptomics and in vitro DNA band shift assays. *Microbiology* **155:** 1459–1477.
- Jung S, Chun J, Yim S, Lee S, Cheon C, Song E, Lee M (2010):** Transcriptional regulation of histidine biosynthesis genes in *Corynebacterium glutamicum*. *Can. J. Microbiol.* **56:** 178–187.
- Jungwirth B, Sala C, Kohl TA, Uplekar S, Baumbach J, Cole ST, Pühler A, Tauch A (2013):** High-resolution detection of DNA binding sites of the global transcriptional regulator GlxR in *Corynebacterium glutamicum*. *Microbiology* **159:** 12–22.
- Jürgen B, Schweder T, Hecker M (1998):** The stability of mRNA from the *gsiB* gene of *Bacillus subtilis* is dependent on the presence of a strong ribosome binding site. *Mol. Gen. Genet.* **258:** 538–545.
- Kaberdin VR (2003):** Probing the substrate specificity of *Escherichia coli* RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res.* **31:** 4710–4716.
- Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Krämer R, Linke B, McHardy AC, Meyer F, Möckel B, Pfefferle W, Pühler A, Rey DA, Rückert C, Rupp O, Sahm H, Wendisch VF, Wieggräbe I, 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:** 5–25.
- Kawamoto H, Koide Y, Morita T, Aiba H (2006):** Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol. Microbiol.* **61:** 1013–1022.

- Kazantsev AV, Pace NR (2006):** Bacterial RNase P: a new view of an ancient enzyme. *Nat. Rev. Microbiol.* **4**: 729–740.
- Keiler KC (2008):** Biology of *trans*-translation. *Annu. Rev. Microbiol.* **62**: 133–151.
- Keilhauer C, Eggeling L, Sahn H (1993):** Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon. *J Bacteriol* **175**: 5595–5603.
- Kim J, Holmes RK (2012):** Characterization of OxyR as a negative transcriptional regulator that represses catalase production in *Corynebacterium diphtheriae*. *PLoS ONE* **7**: e31709.
- Kim T, Kim H, Park J, Kim Y, Kim P, Lee H (2005):** Functional analysis of sigH expression in *Corynebacterium glutamicum*. *Biochem. Biophys. Res. Commun.* **331**: 1542–1547.
- Kingsford C, Ayanbule K, Salzberg S (2007):** Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* **8**.
- Kinoshita S, Takayama S, Akita S (1958):** Taxonomical study of glutamic acid accumulating bacteria, *Micrococcus glutamicus* nov. sp. *Bull. Agr. Chem. Soc. Jpn.*: 176–185.
- Kim T, Yamada K, Terai G, Okida H, Yoshinari Y, Ono Y, Kojima A, Kimura Y, Komori T, Asai K (2007):** iRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res.* **35**: D145-8.
- Klug G, Adams CW, Belasco J, Doerge B, Cohen SN (1987):** Biological consequences of segmental alterations in mRNA stability: effects of deletion of the intercistronic hairpin loop region of the *Rhodobacter capsulatus puf* operon. *EMBO J.* **6**: 3515–3520.
- Kobe B, Deisenhofer J (1993):** Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**: 751–756.
- Krämer R (2009):** Osmosensing and osmosignaling in *Corynebacterium glutamicum*. *Amino Acids* **37**: 487–497.
- Krug A, Wendisch VF, Bott M (2005):** Identification of AcnR, a TetR-type repressor of the aconitase gene *acn* in *Corynebacterium glutamicum*. *J. Biol. Chem.* **280**: 585–595.
- Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982):** Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**: 147–157.
- Kulis-Horn RK, Persicke M, Kalinowski J (2014):** Histidine biosynthesis, its regulation and biotechnological application in *Corynebacterium glutamicum*. *Microb Biotechnol* **7**: 5–25.
- Kumar R, Shah P, Swiatlo E, Burgess SC, Lawrence ML, Nanduri B (2010):** Identification of novel non-coding small RNAs from *Streptococcus pneumoniae* TIGR4 using high-resolution genome tiling arrays. *BMC Genomics* **11**: 350.
- Kunin V, Sorek R, Hugenholtz P (2007):** Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**: R61.
- Kushner SR (2002):** mRNA decay in *Escherichia coli* comes of age. *J. Bacteriol.* **184**: 4658-65; discussion 4657.
- Landt SG, Abeliuk E, McGrath PT, Lesley JA, McAdams HH, Shapiro L (2008):** Small non-coding RNAs in *Caulobacter crescentus*. *Mol. Microbiol.* **68**: 600–614.
- Langhorst MF, Reuter A, Stuermer CAO (2005):** Scaffolding microdomains and beyond: the function of reggie/flotillin proteins. *Cell. Mol. Life Sci.* **62**: 2228–2240.
- Larisch C, Nakunst D, Hüser AT, Tauch A, Kalinowski J (2007):** The alternative sigma factor SigB of *Corynebacterium glutamicum* modulates global gene expression during transition from exponential growth to stationary phase. *BMC Genomics* **8**: 4.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007):** Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.
- Lasa I, Toledo-Arana A, Dobin A, Villanueva M, los Mozos IR de, Vergara-Irigaray M, Segura V, Fagegaltier D, Penades JR, Valle J, Solano C, Gingeras TR (2011):** Genome-wide antisense transcription drives mRNA processing in bacteria. *Proceedings of the National Academy of Sciences* **108**: 20172–20177.

- Lease RA, Belfort M (2000):** Riboregulation by DsrA RNA: *trans*-actions for global economy. *Mol. Microbiol.* **38**: 667–672.
- Lease RA, Cusick ME, Belfort M (1998):** Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 12456–12461.
- Lee E, Groisman EA (2010):** An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol. Microbiol.* **76**: 1020–1033.
- Lehmann KB, Neumann R (1896):** Atlas und Grundriss der Bakteriologie und Lehrbuch der speciellen bakteriologischen Diagnostik 1st edition, Munich: Lehmann.
- Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL (2004):** The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118**: 69–82.
- Letek M, Fiuza M, Ordóñez E, Villadangos AF, Ramos A, Mateos LM, Gil JA (2008):** Cell growth and cell division in the rod-shaped actinomycete *Corynebacterium glutamicum*. *Antonie Van Leeuwenhoek* **94**: 99–109.
- Leuchtenberger W, Huthmacher K, Drauz K (2005):** Biotechnological production of amino acids and derivatives: current status and prospects. *Appl. Microbiol. Biotechnol.* **69**: 1–8.
- Liebl W (2004):** In: Eggeling L. & Bott M. (Hrsg.) Handbook of *Corynebacterium glutamicum*. CRC Press:2:9-34.
- Liebl W, Bayerl A, Schein B, Stillner U, Schleifer KH (1989):** High efficiency electroporation of intact *Corynebacterium glutamicum* cells. *FEMS Microbiol. Lett.* **53**: 299–303.
- Li J, Horwitz R, McCracken S, Greenblatt J (1992):** NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. *J. Biol. Chem.* **267**: 6012–6019.
- Lin-Chao S, Chiou N, Schuster G (2007):** The PNPase, exosome and RNA helicases as the building components of evolutionarily-conserved RNA degradation machines. *J. Biomed. Sci.* **14**: 523–532.
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999):** High density synthetic oligonucleotide arrays. *Nat. Genet.* **21**: 20–24.
- Liu G, Wu J, Yang H, Bao Q (2010):** Codon Usage Patterns in *Corynebacterium glutamicum*: Mutational Bias, Natural Selection and Amino Acid Conservation. *Comp. Funct. Genomics* 2010: 343569.
- Liu JM, Camilli A (2011):** Discovery of Bacterial sRNAs by High-Throughput Sequencing. *Methods in molecular biology* (Clifton, N.J.) **733**: 63–79.
- Liu Y, Li P, Zhao K, Wang B, Jiang C, Drake HL, Liu S (2008):** *Corynebacterium glutamicum* contains 3-deoxy-D-arabino-heptulosonate 7-phosphate synthases that display novel biochemical features. *Appl. Environ. Microbiol.* **74**: 5497–5503.
- Livny J, Brencic A, Lory S, Waldor MK (2006):** Identification of 17 *Pseudomonas aeruginosa* sRNAs and prediction of sRNA-encoding genes in 10 diverse pathogens using the bioinformatic tool sRNAPredict2. *Nucleic Acids Res.* **34**: 3484–3493.
- Livny J, Fogel MA, Davis BM, Waldor MK (2005):** sRNAPredict: an integrative computational approach to identify sRNAs in bacterial genomes. *Nucleic Acids Res.* **33**: 4096–4105.
- Livny J, Teonadi H, Livny M, Waldor M (2008):** High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. *PLoS ONE* **3**(11).
- Li W, Ying X, Lu Q, Chen L (2012):** Predicting sRNAs and their targets in bacteria. *Genomics Proteomics Bioinformatics* **10**: 276–284.
- Lonetto MA, Brown KL, Rudd KE, Buttner MJ (1994):** Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 7573–7577.
- Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011):** ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**: 26.
- Ma B, Tromp J, Li M (2002):** PatternHunter: faster and more sensitive homology search. *Bioinformatics* **18**: 440–445.

- Majdalani N, Cuning C, Sledjeski D, Elliott T, Gottesman S (1998):** DsrA RNA regulates translation of RpoS message by an anti-*antisense* mechanism, independent of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 12462–12467.
- Majdalani N, Hernandez D, Gottesman S (2002):** Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol. Microbiol.* **46**: 813–826.
- Malone JH, Oliver B (2011):** Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* **9**: 34.
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011):** CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**: D225-9.
- Marguerat S, Bähler J (2010):** RNA-seq: from technology to biology. *Cell. Mol. Life Sci.* **67**: 569–579.
- Marraffini LA, Sontheimer EJ (2010):** CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**: 181–190.
- Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007):** Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**: 666–670.
- Massé E, Vanderpool CK, Gottesman S (2005):** Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J. Bacteriol.* **187**: 6962–6971.
- Mathy N, Bénard L, Pellegrini O, Daou R, Wen T, Condon C (2007):** 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* **129**: 681–692.
- McCaskill JS (1990):** The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**: 1105–1119.
- Mentz A, Neshat A, Pfeifer-Sancar K, Pühler A, Rückert C, Kalinowski J (2013):** Comprehensive discovery and characterization of small RNAs in *Corynebacterium glutamicum* ATCC 13032. *BMC Genomics* **14**: 714.
- Mentz A, Neshat A, Rückert C, Kalinowski J. (Manuskript in Vorbereitung):** Transcriptional attenuation in *Corynebacterium glutamicum* and the experimental validation of a new attenuator of the aromatic amino acid synthesis.
- Merino E, Yanofsky C (2005):** Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* **21**: 260–264.
- Merkens H, Beckers G, Wirtz A, Burkovski A (2005):** Vanillate metabolism in *Corynebacterium glutamicum*. *Curr. Microbiol.* **51**: 59–65.
- Miller SL (1953):** A production of amino acids under possible primitive earth conditions. *Science* **117**: 528–529.
- Miller SL, Urey HC (1959):** Organic compound synthesis on the primitive earth. *Science* **130**: 245–251.
- Miotto P, Forti F, Ambrosi A, Pellin D, Veiga DF, Balazsi G, Gennaro ML, Di Serio C, Ghisotti D, Cirillo DM (2012):** Genome-wide discovery of small RNAs in *Mycobacterium tuberculosis*. *PLoS ONE* **7**: e51950.
- Miranda-Ríos J, Navarro M, Soberón M (2001):** A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 9736–9741.
- Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E (2002):** Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**: 747–756.
- Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR (2005):** Applications of DNA tiling arrays for whole-genome analysis. *Genomics* **85**: 1–15.
- Mohanty BK, Kushner SR (2000):** Polynucleotide phosphorylase functions both as a 3' right-arrow 5' exonuclease and a poly(A) polymerase in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 11966–11971.
- Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E (2005):** Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**: 174–182.

- Möker N, Brocker M, Schaffer S, Krämer R, Morbach S, Bott M (2004):** Deletion of the genes encoding the MtrA-MtrB two-component system of *Corynebacterium glutamicum* has a strong influence on cell morphology, antibiotics susceptibility and expression of genes involved in osmoprotection. *Mol. Microbiol.* **54:** 420–438.
- Møller T, Franch T, Udesen C, Gerdes K, Valentin-Hansen P (2002):** Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes Dev.* **16:** 1696–1706.
- Morbach S, Junger C, Sahm H, Eggeling L (2000):** Attenuation control of *ilvBNC* in *Corynebacterium glutamicum*: evidence of leader peptide formation without the presence of a ribosome binding site. *J. Biosci. Bioeng.* **90:** 501–507.
- Mormann S, Lömker A, Rückert C, Gaigalat L, Tauch A, Pühler A, Kalinowski J (2006):** Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. *BMC Genomics* **7:** 205.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008):** Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5:** 621–628.
- Mraheil MA, Billion A, Mohamed W, Mukherjee K, Kuenne C, Pischmarov J, Krawitz C, Retey J, Hartsch T, Chakraborty T, Hain T (2011):** The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages. *Nucleic acids research* **39:** 4235–4248.
- Mückstein U, Tafer H, Hackermüller J, Bernhart SH, Stadler PF, Hofacker IL (2006):** Thermodynamics of RNA-RNA binding. *Bioinformatics* **22:** 1177–1182.
- Mueckstein U, Leparac GG, Posekany A, Hofacker I, Kreil DP (2010):** Hybridization thermodynamics of NimbleGen microarrays. *BMC Bioinformatics* **11:** 35.
- Murakami KS, Darst SA (2003):** Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* **13:** 31–39.
- Muto A, Ushida C, Himeno H (1998):** A bacterial RNA that functions as both a tRNA and an mRNA. *Trends Biochem. Sci.* **23:** 25–29.
- Nakamura Y, Nishio Y, Ieko K, Gojobori T (2003):** The genome stability in *Corynebacterium* species due to lack of the recombinational repair system. *Gene* **317:** 149–155.
- Nakunst D, Larisch C, Hüser AT, Tauch A, Pühler A, Kalinowski J (2007):** The extracytoplasmic function-type sigma factor SigM of *Corynebacterium glutamicum* ATCC 13032 is involved in transcription of disulfide stress-related genes. *J. Bacteriol.* **189:** 4696–4707.
- Nalbantoğlu OU (2014):** Dynamic programming. *Methods Mol. Biol.* **1079:** 3–27.
- Narberhaus F, Waldminghaus T, Chowdhury S (2006):** RNA thermometers. *FEMS Microbiol. Rev.* **30:** 3–16.
- Naville M, Gautheret D (2010):** Transcription attenuation in bacteria: theme and variations. *Brief Funct Genomics* **9:** 178–189.
- Needleman SB, Wunsch CD (1970):** A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48:** 443–453.
- Neshat A (2010):** Molekulargenetische und bioinformatische Untersuchung kleiner RNAs in *Corynebacterium glutamicum*, Universität Bielefeld, CeBiTec, Masterarbeit.
- Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ieko K, Gojobori T (2003):** Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Res.* **13:** 1572–1579.
- Nottebrock D, Meyer U, Krämer R, Morbach S (2003):** Molecular and biochemical characterization of mechanosensitive channels in *Corynebacterium glutamicum*. *FEMS Microbiol. Lett.* **218:** 305–309.
- Nudler E, Mironov AS (2004):** The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29:** 11–17.
- Opdyke JA, Kang J, Storz G (2004):** GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J. Bacteriol.* **186:** 6698–6705.
- Ott A, Idali A, Marchais A, Gautheret D (2012):** NAPP: the Nucleic Acid Phylogenetic Profile Database. *Nucleic Acids Res.* **40:** D205–9.
- Palchevskiy V, Finkel SE (2006):** *Escherichia coli* competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. *J. Bacteriol.* **188:** 3902–3910.

- Panina EM, Vitreschak AG, Mironov AA, Gelfand MS (2001):** Regulation of aromatic amino acid biosynthesis in gamma-proteobacteria. *J. Mol. Microbiol. Biotechnol.* **3**: 529–543.
- Papenfort K, Bouvier M, Mika F, Sharma CM, Vogel J (2010):** Evidence for an autonomous 5' target recognition domain in an Hfq-associated small RNA. *Proc. Natl. Acad. Sci. U.S.A.* **107**: 20435–20440.
- Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JCD, Vogel J (2006):** SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global *omp* mRNA decay. *Mol. Microbiol.* **62**: 1674–1688.
- Park PJ (2009):** ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**: 669–680.
- Park S, Youn J, Kim Y, Lee S, Kim Y, Lee H (2008):** *Corynebacterium glutamicum* sigmaE is involved in responses to cell surface stresses and its activity is controlled by the anti-sigma factor CseE. *Microbiology* **154**: 915–923.
- Pasman Z, Hippel PH von (2000):** Regulation of rho-dependent transcription termination by NusG is specific to the *Escherichia coli* elongation complex. *Biochemistry* **39**: 5573–5585.
- Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH (2009):** Structure and complexity of a bacterial transcriptome. *J. Bacteriol.* **191**: 3203–3211.
- Pátek M, Krumbach K, Eggeling L, Sahm H (1994):** Leucine synthesis in *Corynebacterium glutamicum*: enzyme activities, structure of *leuA*, and effect of *leuA* inactivation on lysine synthesis. *Appl. Environ. Microbiol.* **60**: 133–140.
- Pátek M, Nešvera J (2011):** Sigma factors and promoters in *Corynebacterium glutamicum*. *J. Biotechnol.* **154**: 101–113.
- Patenge N, Billion A, Raasch P, Normann J, Wisniewska-Kucper A, Retey J, Boisguérin V, Hartsch T, Hain T, Kreikemeyer B (2012):** Identification of novel growth phase- and media-dependent small non-coding RNAs in *Streptococcus pyogenes* M49 using intergenic tiling arrays. *BMC Genomics* **13**: 550.
- Pauling J, Röttger R, Tauch A, Azevedo V, Baumbach J (2012):** CoryneRegNet 6.0--Updated database content, new analysis methods and novel features focusing on community demands. *Nucleic Acids Res.* **40**: D610–4.
- Peano C, Pietrelli A, Consolandi C, Rossi E, Petiti L, Tagliabue L, Bellis G de, Landini P (2013):** An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microb Inform Exp* **3**: 1.
- Perez N, Treviño J, Liu Z, Ho SCM, Babitzke P, Sumby P, Aziz RK (2009):** A Genome-Wide Analysis of Small Regulatory RNAs in the Human Pathogen Group A *Streptococcus*. *PLoS ONE* **4**: e7668.
- Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R (2012):** Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev.* **26**: 2621–2633.
- Peters JM, Mooney RA, Kuan PF, Rowland JL, Keles S, Landick R (2009):** Rho directs widespread termination of intragenic and stable RNA transcription. *Proc. Natl. Acad. Sci. U.S.A.* **106**: 15406–15411.
- Petri K, Pfeifer-Sancar K, Mentz A, Rückert C, Persicke M, Kalinowski J:** (Manuskript in Vorbereitung) Basal transcription of the arginine biosynthesis operon of *Corynebacterium glutamicum* is controlled by transcriptional interference from an antisense promoter in *argC*.
- Pfeifer-Sancar K, Mentz A, Rückert C, Kalinowski J (2013):** Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics* **14**: 888.
- Phillips P, Progulské-Fox A, Grieshaber S, Grieshaber N (2013):** Expression of *Porphyromonas gingivalis* small RNA in response to hemin availability identified using microarray and RNA-seq analysis. *FEMS Microbiol. Lett.*
- Pichon C, Felden B (2005):** Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc Natl Acad Sci U.S.A.* **102**: 14249–14254.
- Pischmarov J, Kuenne C, Billion A, Hemberger J, Cemič F, Chakraborty T, Hain T (2012):** sRNAdb: a small non-coding RNA database for gram-positive bacteria. *BMC Genomics* **13**: 384.
- Poritz MA, Bernstein HD, Strub K, Zopf D, Wilhelm H, Walter P (1990):** An *E. coli* ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. *Science* **250**: 1111–1117.
- Prasse D, Ehlers C, Backofen R, Schmitz RA (2013):** Regulatory RNAs in archaea: first target identification in *Methanoarchaea*. *Biochem. Soc. Trans.* **41**: 344–349.

- Ptashne M, Backman K, Humayun MZ, Jeffrey A, Maurer R, Meyer B, Sauer RT (1976):** Autoregulation and function of a repressor in bacteriophage *lambda*. *Science* **194**: 156–161.
- Pulvermacher SC, Stauffer LT, Stauffer GV (2009):** Role of the *Escherichia coli* Hfq protein in GcvB regulation of oppA and dppA mRNAs. *Microbiology* **155**: 115–123.
- Quackenbush J (2001):** Computational analysis of microarray data. *Nat. Rev. Genet.* **2**: 418–427.
- Raabe CA, Tang T, Brosius J, Rozhdestvensky TS (2013):** Biases in small RNA deep sequencing data. *Nucleic Acids Res.* [Epub ahead of print]
- Ralston A (2008):** Simultaneous gene transcription and translation in bacteria. *Nature Education* **1**(1).
- Rasmussen AA, Eriksen M, Gilany K, Udesen C, Franch T, Petersen C, Valentin-Hansen P (2005):** Regulation of ompA mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. *Mol. Microbiol.* **58**: 1421–1429.
- Rauhut R, Klug G (1999):** mRNA degradation in bacteria. *FEMS Microbiol. Rev.* **23**: 353–370.
- Ray BK, Apirion D (1979):** Characterization of 10S RNA: a new stable rna molecule from *Escherichia coli*. *Mol. Gen. Genet.* **174**: 25–32.
- Reeder J, Steffen P, Giegerich R (2007):** pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic Acids Res.* **35**: W320–4.
- Richardson JP (2003):** Loading Rho to terminate transcription. *Cell* **114**: 157–159.
- Rivas E, Eddy SR (2001):** Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**: 8.
- Roberts JW (1969):** Termination factor for RNA synthesis. *Nature* **224**: 1168–1174.
- Rückert C, Milse J, Albersmeier A, Koch DJ, Pühler A, Kalinowski J (2008):** The dual transcriptional regulator CysR in *Corynebacterium glutamicum* ATCC 13032 controls a subset of genes of the McbR regulon in response to the availability of sulphide acceptor molecules. *BMC Genomics* **9**: 483.
- Sahm H, Eggeling L (1999):** D-Pantothenate synthesis in *Corynebacterium glutamicum* and use of panBC and genes encoding L-valine synthesis for D-pantothenate overproduction. *Appl. Environ. Microbiol.* **65**: 1973–1979.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988):** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
- Salari R, Aksay C, Karakoc E, Unrau PJ, Hajirasouliha I, Sahinalp SC (2009):** smyRNA: a novel Ab initio ncRNA gene finder. *PLoS ONE* **4**: e5433.
- Sano C (2009):** History of glutamate production. *Am. J. Clin. Nutr.* **90**: 728S–732S.
- Sano K, Matsui K (1987):** Structure and function of the trp operon control regions of *Brevibacterium lactofermentum*, a glutamic-acid-producing bacterium. *Gene* **53**: 191–200.
- Sarkar N (1997):** Polyadenylation of mRNA in prokaryotes. *Annu. Rev. Biochem.* **66**: 173–197.
- Schäfer A, Tauch A, Jäger W, Kalinowski J, Thierbach G, 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**: 69–73.
- Schena M, Shalon D, Davis RW, Brown PO (1995):** Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Schlüter JP, Reinkensmeier J, Daschkey S, Hackenberg EE, Janssen S, Janicke S, Becker J, Giegerich R, Becker A (2010):** A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genomics* **11**: 245.
- Schröder J, Jochmann N, Rodionov DA, Tauch A (2010):** The Zur regulon of *Corynebacterium glutamicum* ATCC 13032. *BMC Genomics* **11**: 12.
- Schröder J, Maus I, Trost E, 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.

- Schröder J, Tauch A (2010):** Transcriptional regulation of gene expression in *Corynebacterium glutamicum*: the role of global, master and local regulators in the modular and hierarchical gene regulatory network. *FEMS Microbiol. Rev.* **34**: 685–737.
- Schroeder S (2007):** IRMA Berechnung und Visualisierung von Oligonukleotiden aus Intergenischen Regionen zur Identifikation nichtcodierender RNAs und Transkriptions-Starts durch Microarray Analysen, Universität Bielefeld, Technische Fakultät, Bioinformatics Resource Facility, CeBiTec, Bachelorarbeit.
- Selinger D, Cheung K, Mei R, Johansson E, Richmond C, Blattner F, Lockhart D, Church G (2000):** RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat Biotechnol* **18**: 1262–1268.
- Sharma CM, Darfeuille F, Plantinga TH, Vogel J (2007):** A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.* **21**: 2804–2817.
- Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiß S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF, Vogel J (2010):** The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**: 250–255.
- Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf H, Hinton JCD, Vogel J (2011):** Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol. Microbiol.* **81**: 1144–1165.
- Sharma CM, Vogel J (2009):** Experimental approaches for the discovery and characterization of regulatory small RNA. *Curr. Opin. Microbiol.* **12**: 536–546.
- Shearwin KE, Callen BP, Egan JB (2005):** Transcriptional interference--a crash course. *Trends Genet.* **21**: 339–345.
- Shen LX, Cai Z, Tinoco I (1995):** RNA structure at high resolution. *FASEB J.* **9**: 1023–1033.
- Shi L, et al. (2006):** The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.* **24**: 1151–1161.
- Shi L, et al. (2010):** The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.* **28**: 827–838.
- Shine J, Dalgarno L (1974):** The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U.S.A.* **71**: 1342–1346.
- Silvaggi JM, Perkins JB, Losick R (2005):** Small untranslated RNA antitoxin in *Bacillus subtilis*. *J. Bacteriol.* **187**: 6641–6650.
- Silverstone AE, Goman M, Scaife JG (1972):** ALT: a new factor involved in the synthesis of RNA by *Escherichia coli*. *Mol. Gen. Genet.* **118**: 223–234.
- Sîrbu A, Kerr G, Crane M, Ruskin HJ (2012):** RNA-Seq vs dual- and single-channel microarray data: sensitivity analysis for differential expression and clustering. *PLoS ONE* **7**: e50986.
- Sittka A, Lucchini S, Papenfort K, Sharma C, Rolle K, Binnewies T, Hinton J, Vogel J (2008):** Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* **4**.
- Smith TF, Waterman MS (1981):** Identification of common molecular subsequences. *J. Mol. Biol.* **147**: 195–197.
- Söderbom F, Binnie U, Masters M, Wagner EG (1997):** Regulation of plasmid R1 replication: PcnB and RNase E expedite the decay of the antisense RNA, CopA. *Mol. Microbiol.* **26**: 493–504.
- Soper T, Mandin P, Majdalani N, Gottesman S, Woodson SA (2010):** Positive regulation by small RNAs and the role of Hfq. *Proceedings of the National Academy of Sciences of the U.S.A.* **107**: 9602–9607.
- Sorek R, Cossart P (2010):** Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* **11**: 9–16.
- Sridhar J, Gunasekaran P (2013):** Computational small RNA prediction in bacteria. *Bioinform Biol Insights* **7**: 83–95.
- Sridhar J, Rafi ZA (2007):** Small RNA identification in *Enterobacteriaceae* using synteny and genomic backbone retention. *OMICS* **11**: 74–99.

- Stackebrandt E, Rainey FA, Ward-Rainey NL (1997):** Proposal for a new hierarchal classification system, *Actinobacteria* classis nov. *Int. J. Syst. Bacteriol.*: 479–491.
- Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R (2006):** RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics* **22**: 500–503.
- Steil L, Hoffmann T, Budde I, Völker U, Bremer E (2003):** Genome-wide transcriptional profiling analysis of adaptation of *Bacillus subtilis* to high salinity. *J. Bacteriol.* **185**: 6358–6370.
- Stock AM, Robinson VL, Goudreau PN (2000):** Two-component signal transduction. *Annu. Rev. Biochem.* **69**: 183–215.
- Stork M, Di Lorenzo M, Welch TJ, Crosa JH (2007):** Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *J. Bacteriol.* **189**: 3479–3488.
- Storz G (2002):** An expanding universe of noncoding RNAs. *Science* **296**: 1260–1263.
- Storz G, Tartaglia LA, Farr SB, Ames BN (1990):** Bacterial defenses against oxidative stress. *Trends Genet.* **6**: 363–368.
- Storz G, Vogel J, Wassarman KM (2011):** Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell* **43**: 880–891.
- Sturn A, Quackenbush J, Trajanoski Z (2002):** Genesis: cluster analysis of microarray data. *Bioinformatics* **18**: 207–208.
- Su AAH, Randau L (2011):** A-to-I and C-to-U editing within transfer RNAs. *Biochemistry Mosc.* **76**: 932–937.
- Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL (2001):** A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* **17**: 1123–1130.
- Su Z, Li Z, Chen T, Li Q, Fang H, Ding D, Ge W, Ning B, Hong H, Perkins RG, Tong W, Shi L (2011):** Comparing next-generation sequencing and microarray technologies in a toxicological study of the effects of aristolochic acid on rat kidneys. *Chem. Res. Toxicol.* **24**: 1486–1493.
- Swiercz JP, Hindra, Bobek J, Haiser HJ, Di Berardo C, Tjaden B, Elliot MA (2008):** Small non-coding RNAs in *Streptomyces coelicolor*. *Nucleic Acids Res.* **36**: 7240–7251.
- Tafer H, Hofacker IL (2008):** RNAplex: a fast tool for RNA-RNA interaction search. *Bioinformatics* **24**: 2657–2663.
- Tauch A, Kirchner O, Löffler B, Götter S, Pühler A, Kalinowski J (2002):** Efficient electrotransformation of *Corynebacterium diphtheriae* with a mini-replicon derived from the *Corynebacterium glutamicum* plasmid pGA1. *Curr. Microbiol.* **45**: 362–367.
- Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999):** Systematic determination of genetic network architecture. *Nat. Genet.* **22**: 281–285.
- Thomason MK, Storz G (2010):** Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annu. Rev. Genet.* **44**: 167–188.
- Tian B, Bevilacqua PC, Diegelman-Parente A, Mathews MB (2004):** The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.* **5**: 1013–1023.
- Tichopad A, Dilger M, Schwarz G, Pfaffl MW (2003):** Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res.* **31**: e122.
- Tjaden B (2008):** TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res* **36**: W109–113.
- Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G (2006):** Target prediction for small, noncoding RNAs in bacteria. *Nucleic acids research* **34**: 2791–2802.
- Tjaden B, Saxena R, Stolyar S, Haynor D, Kolker E, Rosenow C (2002):** Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res* **30**: 3732–3738.
- Toffano-Nioche C, Nguyen AN, Kuchly C, Ott A, Gautheret D, Bouloc P, Jacq A (2012):** Transcriptomic profiling of the oyster pathogen *Vibrio splendidus* opens a window on the evolutionary dynamics of the small RNA repertoire in the *Vibrio* genus. *RNA* **18**: 2201–2219.

- Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori M, Soubigou G, Regnault B, Coppee J, Lecuit M, Johansson J, Cossart P (2009):** The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**: 950–956.
- Tompa M, Li N, Bailey TL, Church GM, Moor B de, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Régnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z (2005):** Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.* **23**: 137–144.
- Tran TT, Zhou F, Marshburn S, Stead M, Kushner SR, Xu Y (2009):** *De novo* computational prediction of non-coding RNA genes in prokaryotic genomes. *Bioinformatics* **25**: 2897–2905.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010):** Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**: 511–515.
- Tucker BJ, Breaker RR (2005):** Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* **15**: 342–348.
- Udekwu KI, Darfeuille F, Vogel J, Reimegård J, Holmqvist E, Wagner EGH (2005):** Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev.* **19**: 2355–2366.
- Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, Lowe TM, Salama SR, Haussler D (2010):** FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7**: 995–1001.
- Urban JH, Vogel J (2007):** Translational control and target recognition by *Escherichia coli* small RNAs in vivo. *Nucleic acids research* **35**: 1018–1037.
- Urbanowski ML, Stauffer LT, Stauffer GV (2000):** The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Mol. Microbiol.* **37**: 856–868.
- Valentin-Hansen P, Eriksen M, Udesen C (2004):** The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol* **51**: 1525–1533.
- van Bakel H, Nislow C, Blencowe BJ, Hughes TR (2010):** Most "dark matter" transcripts are associated with known genes. *PLoS Biol.* **8**(5)
- Vanderpool CK, Balasubramanian D, Lloyd CR (2011):** Dual-function RNA regulators in bacteria. *Biochimie.* **93**(11):1943-9.
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995):** Serial analysis of gene expression. *Science* **270**: 484–487.
- Vida TA, Emr SD (1995):** A new vital stain for visualizing vacuolar membrane dynamics and endocytosis in yeast. *J. Cell Biol.* **128**: 779–792.
- Vitreschak AG, Lyubetskaya EV, Shirshin MA, Gelfand MS, Lyubetsky VA (2004):** Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiol. Lett.* **234**: 357–370.
- Vitreschak AG, Mironov AA, Lyubetsky VA, Gelfand MS (2008):** Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA* **14**: 717–735.
- Vockenhuber M, Sharma CM, Statt MG, Schmidt D, Xu Z, Dietrich S, Liesegang H, Mathews DH, Suess B (2011):** Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. *RNA Biol* **8**: 468–477.
- Vogel J, Sharma C (2005):** How to find small non-coding RNAs in bacteria. *Biol Chem* **386**: 1219–1238.
- Vogel J, Wagner EGH (2007):** Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol.* **10**: 262–270.
- Volkov A, Mascarenhas J, Andrei-Selmer C, Ulrich HD, Graumann PL (2003):** A prokaryotic condensin/cohesin-like complex can actively compact chromosomes from a single position on the nucleoid and binds to DNA as a ring-like structure. *Mol. Cell. Biol.* **23**: 5638–5650.

- Voskuil MI, Voepel K, Chambliss GH (1995):** The -16 region, a vital sequence for the utilization of a promoter in *Bacillus subtilis* and *Escherichia coli*. *Mol. Microbiol.* **17**: 271–279.
- Waldminghaus T, Gaubig LC, Klinkert B, Narberhaus F (2009):** The *Escherichia coli* *ibpA* thermometer is comprised of stable and unstable structural elements. *RNA Biol* **6**: 455–463.
- Waldminghaus T, Heidrich N, Brantl S, Narberhaus F (2007):** FourU: a novel type of RNA thermometer in *Salmonella*. *Mol. Microbiol.* **65**: 413–424.
- Waldsich C, Grossberger R, Schroeder R (2002):** RNA chaperone StpA loosens interactions of the tertiary structure in the *td* group I intron in vivo. *Genes Dev.* **16**: 2300–2312.
- Wang L, Si Y, Dedow LK, Shao Y, Liu P, Brutnell TP (2011):** A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PLoS ONE* **6**: e26426.
- Wang P, Yang J, Ishihama A, Pittard AJ (1998):** Demonstration that the TyrR protein and RNA polymerase complex formed at the divergent P3 promoter inhibits binding of RNA polymerase to the major promoter, P1, of the *aroP* gene of *Escherichia coli*. *J. Bacteriol.* **180**: 5466–5472.
- Wang Z, Gerstein M, Snyder M (2009):** RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**: 57–63.
- Washietl S, Hofacker IL, Stadler PF (2005):** Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 2454–2459.
- Wassarman KM (2007):** 6S RNA: a small RNA regulator of transcription. *Curr. Opin. Microbiol.* **10**: 164–168.
- Wassarman K, Repoila F, Rosenow C, Storz G, Gottesman S (2001):** Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* **15**: 1637–1651.
- Wassarman K, Storz G (2000):** 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101**: 613–623.
- Wassarman K, Zhang A, Storz G (1999):** Small RNAs in *Escherichia coli*. *Trends Microbiol* **7**: 37–45.
- Waters L, Storz G (2009):** Regulatory RNAs in bacteria. *Cell* **136**: 615–628.
- Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, Baker CS, Georgellis D, Babitzke P, Romeo T (2003):** A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol. Microbiol.* **48**: 657–670.
- Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR (2010):** Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome biology* **11**.
- Wennerhold J, Krug A, Bott M (2005):** The AraC-type regulator RipA represses aconitase and other iron proteins from *Corynebacterium* under iron limitation and is itself repressed by DtxR. *J. Biol. Chem.* **280**: 40500–40508.
- Wilms I, Overlöper A, Nowrousian M, Sharma CM, Narberhaus F (2012):** Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium tumefaciens*. *RNA biology* **9**.
- Wilms I, Voss B, Hess WR, Leichert LI, Narberhaus F (2011):** Small RNA-mediated control of the *Agrobacterium tumefaciens* GABA binding protein. *Mol. Microbiol.* **80**: 492–506.
- Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP (1997):** Continuous fluorescence monitoring of rapid cycle DNA amplification. *BioTechniques* **22**: 130-1, 134-8.
- Woese CR, Winker S, Gutell RR (1990):** Architecture of ribosomal RNA: constraints on the sequence of "tetraloops". *Proc. Natl. Acad. Sci. U.S.A.* **87**: 8467–8471.
- Wösten MM (1998):** Eubacterial sigma-factors. *FEMS Microbiol. Rev.* **22**: 127–150.
- Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J (2013):** Comparative genomics boosts target prediction for bacterial small RNAs. *Proc. Natl. Acad. Sci. U.S.A.* **110**: E3487-96.
- Wu C, Carta R, Zhang L (2005):** Sequence dependence of cross-hybridization on short oligo microarrays. *Nucleic Acids Res.* **33**: e84.
- Yang IV (2006):** Use of external controls in microarray experiments. *Meth. Enzymol.* **411**: 50–63.
- Yang Y, Dudoit S, Luu P, Lin D, Peng V, Ngai J, Speed T (2002):** Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* **30**(4): e15.
- Yanofsky C (1981):** Attenuation in the control of expression of bacterial operons. *Nature* **289**: 751–758.

- Yukawa H, Omumasaba CA, Nonaka H, Kós P, Okai N, Suzuki N, Suda M, Tsuge Y, Watanabe J, Ikeda Y, Vertès AA, Inui M (2007):** Comparative analysis of the *Corynebacterium glutamicum* group and complete genome sequence of strain R. *Microbiology* **153**: 1042–1058.
- Zemanová M, Kaderábková P, Pátek M, Knoppová M, Silar R, Nesvera J (2008):** Chromosomally encoded small antisense RNA in *Corynebacterium glutamicum*. *FEMS Microbiol. Lett.* **279**: 195–201.
- Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S (2003):** Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.* **50**: 1111–1124.
- Zhao K, Huang Y, Chen X, Wang N, Liu S (2010):** PcaO positively regulates pcaHG of the beta-ketoadipate pathway in *Corynebacterium glutamicum*. *J. Bacteriol.* **192**: 1565–1572.
- Zheng X, Hu G, She Z, Zhu H (2011):** Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes. *BMC Genomics* **12**: 361.
- Zuker M, Stiegler P (1981):** Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**: 133–148.
- Zwieb C, Wower I, Wower J (1999):** Comparative sequence analysis of tmRNA. *Nucleic acids research* **27**: 2063–2071.

VIII. Anhang

1 Geniom@one Microarrayplattform

Die Microarrayplattform geniom@one (febit biomed GmbH, jetzt CBC Comprehensive Biomarker Center, Heidelberg, Deutschland) stellt ein Einfarben-Microarray System dar, das alle Prozessschritte von der Microarraysynthese über die Hybridisierung und Detektion (Baum *et al.*, 2003) integriert. Zur Herstellung der Microarrays dient ein mikrofluidischer Reaktionsträger, der acht Einheiten mit je 6776 Oligonukleotidproben enthält, wodurch insgesamt bis zu 54.000 Proben realisiert werden können. Sowohl die Länge als auch die Anzahl der Oligonukleotidproben können variiert werden und ermöglichen damit eine flexible *in situ*-Synthese der Microarrays über einen fotolithografischen Prozess.

Im Rahmen dieser Arbeit und auf Basis dieser Plattform wurde zusammen mit dem Java-Applet IRMA³² zur Berechnung von intergenischen Oligonukleotidproben eine Methode zur schnellen und relativ kostengünstigen Detektion kleiner bakterieller RNAs entwickelt. Die Validierungsexperimente wurden mithilfe des Enterobakteriums *Escherichia coli* K-12 MG1655 durchgeführt, da zum Zeitpunkt der Methodvalidierung (2010) bereits 60 kleine RNAs für dieses Genom annotiert waren. Anschließend sollte die erfolgreiche Methodvalidierung publiziert werden. Allerdings wurde die „febit biomed GmbH“ Ende des Jahres 2011 umstrukturiert, bekam einen neuen Namen und wurde zur „Comprehensive Biomarker Center GmbH“, kurz CBC. Die Services des Unternehmens wurden grundlegend umgestellt und fokussieren sich nun ausschließlich auf den Bereich Biomarker-Analytik in der personalisierten Medizin. In diesem Zuge wurden die Technik und Herstellung der geniom@one Plattform und der Support für bereits verkaufte Geräte dieser Serie eingestellt. Die zu diesem Zeitpunkt bereits existierenden Daten für *C. glutamicum* beinhalten jedoch viele Informationen, die nicht verloren gehen sollten und werden daher in dieser Arbeit präsentiert.

Im Zuge der Methodvalidierung wurden für das Design des *E. coli* Microarrays über IRMA jeweils 15.409 50-mer Oligonukleotidproben mit 15-mer Überlappung für alle intergenischen Regionen für beide Stränge des Genoms berechnet. Zusätzlich wurden jeweils die Startsequenzen der tRNAs in *sense*-Richtung als Positiv-Kontrollen und deren reverse Komplemente (*antisense*-Richtung) als Negativ-Kontrollen berechnet. In Vorbereitung auf die Hybridisierungen wurden *E. coli* Zellen in LB-Medium kultiviert und bei drei verschiedenen Bedingungen geerntet: (i) Bei exponentiellem Wachstum ohne Stress (ii) in dieser Wachstumsphase nach Hitzeschock und (iii) in der Stationärphase. Nach Isolierung der Gesamt-RNA und Selektion der kleinen RNA-Fraktionen (< 250 nt) wurde diese nach dem Standard-

³² IRMA (*Intergenic Regions by Microarray Analysis*) in Zusammenarbeit mit S. Schroeder Schroeder (2007)

Verfahren Cy3-markiert und mit einem biologischen Replikat Experiment-spezifisch hybridisiert (Baum *et al.*, 2003). Die während dieser Arbeit entwickelte Methode zur Auswertung der Daten ergab folgende Ergebnisse: Die Normalisierung der Rohdaten über Quantil (Bolstad *et al.*, 2003) führte im Vergleich zur Normalisierung über Medianzentrierung zur besseren Reproduzierbarkeit über die Replikate. Der Schwellenwert, ab dem transkriptionelle Aktivität angenommen wird, wurde definiert als der maximale Wert aller tRNA-*antisense* Signale bzw. Negativ-Kontrollen in einem Experiment. Auf diese Weise konnten 50 (83 %) der 60 annotierten kleinen RNAs in *E. coli* über die Experimente detektiert werden, als Beispiel ist das Signal der 6S RNA im Experiment Stationär Phase gezeigt (Abbildung Anhang-1). Darüber hinaus konnte eine Korrelation zwischen der Länge der annotierten kleinen RNAs und der Anzahl der Signale über dem Schwellenwert für transkriptionelle Aktivität in den entsprechenden Genomregionen nachgewiesen werden (Abbildung Anhang-2).

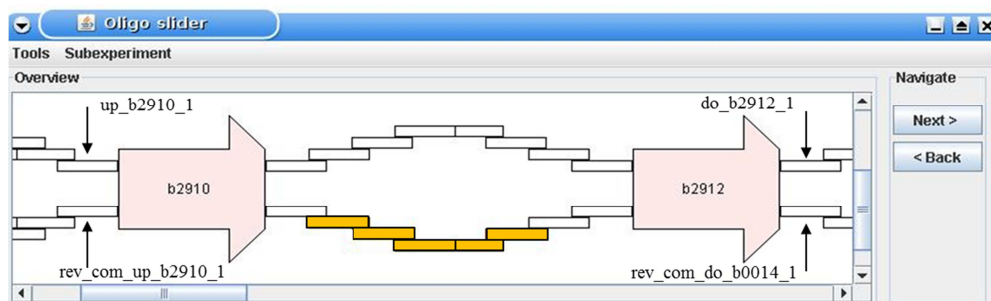


Abbildung Anhang-1 Signal der 6S RNA in *E. coli* K12 MG1655 (Stationär Phase Experiment).

Der Screenshot zeigt die graphische Benutzeroberfläche von IRMA, in der die genomische Region und die intergenischen Oligonukleotide zwischen zwei ORFs (*b2910* und *b2912*, dargestellt als Pfeile) abgebildet sind. Zusätzlich sind in Orange die erhaltenen Signale der 6S RNA gezeigt, die oberhalb des Schwellenwertes für transkriptionelle Aktivität liegen.

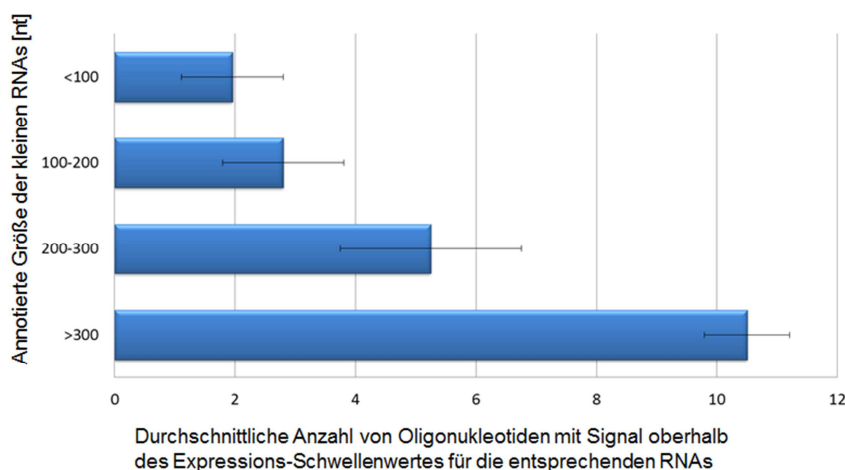


Abbildung Anhang-2 Korrelation der Signallänge mit den Transkriptgrößen annotierter kleiner RNAs.

Alle 50 detektierten annotierten kleinen RNAs wurden anhand ihrer Länge in Klassen sortiert und gegen die Durchschnittsanzahl von Oligonukleotiden mit Transkriptsignal in diesen Regionen aufgetragen.

2 pK18*mobsacB*-Leervektor

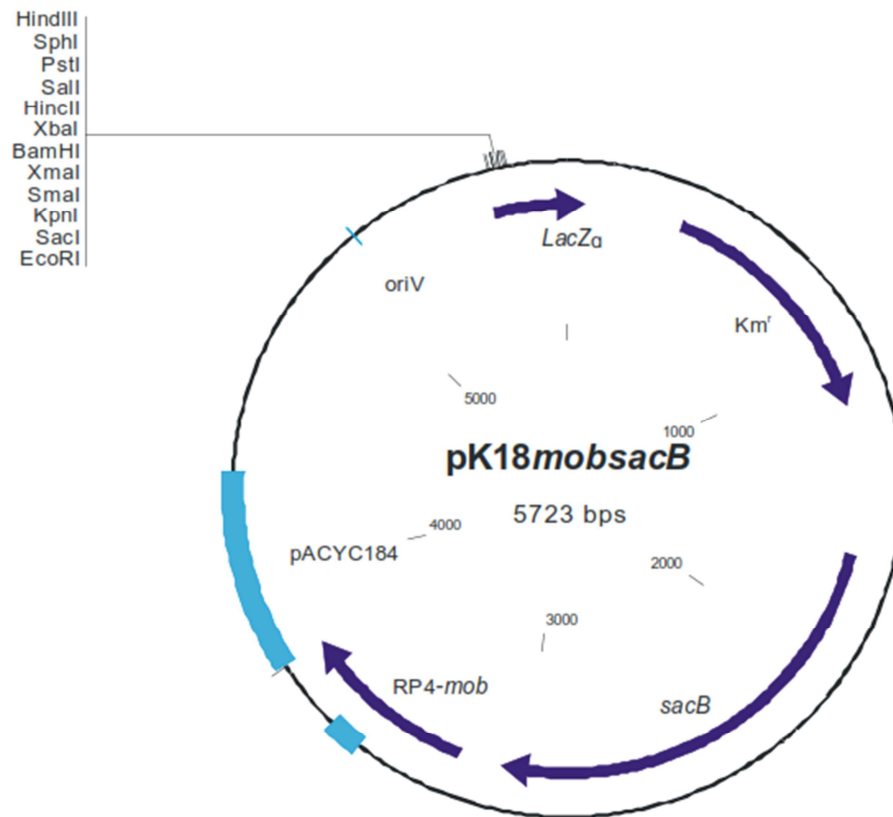


Abbildung VIII-1 pK18*mobsacB*-Leervektor

3 CD (Ordnerstruktur)

- Datenbank.accdb
- Ordner „geniom-Microarrays“
 - Ergebnisse.xlsx
 - Ergebnisse_Kontrollen.xlsx
 - IRMA Oligonukleotide.xlsx
 - IRMA.jar
- Ordner „Sequenzierung“
 - Bioanalyzer.pptx
 - Definition_3´Enden.xlsx
 - Expression.xlsx
 - Transkriptionsstarts-Rohdaten.xlsx
 - Potenzielle_kleine_RNA_Gene.xlsx
 - Liste_putative_Peptide.xlsx
 - Klasse_antisense-Transkripte.xlsx
 - Klasse_mRNA-Leader.xlsx
 - Klasse_trans-kodierte_RNAs.xlsx
- Ordner „Vorhersagen“
 - RNAz-Vorhersage.xlsx
 - Rho-unabhängige Terminatoren.xlsx
 - Mauve.txt
 - Mauve.alignment
 - Mauve.backbone
 - Mauve.guide_tree
 - Mauve.islands
- Ordner „Analyse_cgb_16115“
 - Zielvorhersagen_cgb_16115.xlsx
 - k-Means-Cluster.xlsx
 - Microarray-trifft-Vorhersagen.xlsx
 - Unterordner „EMMA2-cDNA-Microarraydaten“
 - Unterordner „Mikroskopie“

gratiarum actio

Im Rahmen meiner Danksagung möchte ich bei meinem Mann beginnen und ihm ganz besonders danken. Während der letzten Jahre hat er mich in der Betreuung unserer Kinder und auch sonst sehr unterstützt und mir damit erst das Durchführen und Schreiben dieser Dissertation ermöglicht.

Ein ganz großer Dank gilt meinem Betreuer apl. Prof. Dr. Jörn Kalinowski für seine hervorragende wissenschaftliche Betreuung. Er hat mir mit dieser Dissertationsarbeit ein umfangreiches, aber sehr spannendes Thema anvertraut und meine Arbeit immer und in jeder Hinsicht unterstützt.

Natürlich gilt auch meinem Erstgutachter Prof. Dr. Alfred Pühler großer Dank für die wissenschaftliche Betreuung im Hintergrund und wichtige und hilfreiche Anregungen beim Schreiben der Publikationen.

Außerdem möchte ich mich auch bei Prof. Dr. Anke Becker bedanken, dass sie bei der Bewertung dieser Arbeit als Zweitgutachterin zur Verfügung steht.

Daneben möchte ich mich auch bei Prof. Dr. R. Giegerich von der Technischen Fakultät bedanken, der mich während dieser Arbeit mit bioinformatischem Fokus bei aufkommenden Fragen oder Problemen immer gut beraten und unterstützt hat.

Ein spezieller Dank gilt meinem Arbeits-Kollegen Armin Neshat, der diese Arbeit Korrektur gelesen hat. Und ich möchte mich bei meinen Kolleginnen Katharina Pfeifer-Sancar, Kathrin Petri und Johanna Milse bedanken. Sie alle haben mir ein besonders angenehmes und schönes Arbeiten im gemeinsamen Labor ermöglicht. Einige der genannten KollegInnen gehören dem „RNA-Seq-Team“ an, zu dem auch Christian Rückert zählt. Zusammen haben wir mit der Etablierung dieser Methode am CeBiTec einen großen wissenschaftlichen Fortschritt geleistet, und ich möchte mich für die tolle Zusammenarbeit bedanken. Natürlich gilt mein Dank auch allen anderen Mitarbeitern der Arbeitsgruppe von J. Kalinowski, die z.B. immer wenn es mal nötig war ein Ethidium-Bromid Gel für mich gefärbt haben (während meiner Schwangerschaften!) und auch sonst wirklich sehr nette Kollegen sind!

Schließlich möchte an dieser Stelle natürlich auch meinen Eltern großer Dank ausgesprochen werden. Denn sie haben mir auf dem Weg bis heute Vieles ermöglicht und mich während meiner Berufsausbildung in jeder Hinsicht unterstützt und somit natürlich auch zum Entstehen dieser Arbeit beigetragen.

Erklärung

Mit der vorliegenden Dissertation beantrage ich beim Promotionsausschuss der Fakultät für Biologie der Universität Bielefeld die Eröffnung des Promotionsverfahrens nach § 9 der Promotionsordnung in der Fassung vom 10. Mai 2006.

Die Anfertigung dieser Dissertation mit dem Titel „Identifizierung und Funktionsanalysen von kleinen RNAs in *Corynebacterium glutamicum*“ erfolgte unter Betreuung von Prof. Dr. Alfred Pühler in der AG „Genomforschung industrieller Mikroorganismen“ am *Center for Biotechnology – CeBiTec* an der Fakultät für Biologie der Universität Bielefeld.

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Dissertation, die anderen Quellen im Wortlaut oder dem Sinn nach entnommen wurden, sind durch Angabe 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 hiermit erstmalig um den Doktorgrad der Naturwissenschaften der Universität Bielefeld.

Bielefeld, den 28.12.2013

Almut Mentz