

Entwicklung der Transkriptomsequenzierung und Anwendung zur Analyse des Transkriptoms von *Corynebacterium glutamicum*

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

Verfasst und vorgelegt von
Katharina Pfeifer-Sancar
Geboren in Tomsk

Rheda-Wiedenbrück, April 2014

Inhaltsverzeichnis

I	ZUSAMMENFASSUNG DER ARBEIT	1
II	EINLEITUNG.....	4
1	Das Bakterium <i>Corynebacterium glutamicum</i>.....	4
1.1	Taxonomie und Charakteristika der Gattung <i>Corynebacterium</i>	4
1.2	Entdeckung und wirtschaftliche Bedeutung von <i>Corynebacterium glutamicum</i>	6
1.3	Annotation sowie aktuelle Version des Genoms von <i>Corynebacterium glutamicum</i>	8
2	Die Transkriptionsmechanismen in Prokaryoten.....	10
2.1	Struktur und Aufbau der DNA-abhängigen RNA-Polymerase	10
2.2	Transkriptionsinitiation, -elongation und deren Regulation.....	13
2.3	Transkriptionstermination	15
3	Methoden der Transkriptomanalyse.....	17
3.1	Prinzipien der Transkriptom-Hochdurchsatz-Sequenzierung	17
3.2	Genomweite Transkriptomanalyse mittels RNAseq.....	21
4	Zielsetzung der Arbeit	24
III	MATERIALIEN UND METHODEN.....	26
1	Verwendete Materialien und Geräte.....	26
1.1	Bakterienstamm	26
1.2	Nährmedien	26
1.3	Oligonukleotide	28
1.4	Labor-Verbrauchsmaterialien, Enzyme und Kits	28
1.5	Chemikalien.....	30
1.6	Geräte	31
1.7	Software.....	32
1.8	Puffer und Lösungen	33
2	Mikrobiologische Methoden.....	35
2.1	Kultivierung von <i>C. glutamicum</i>	35
2.2	Lagerung von <i>C. glutamicum</i>	35
2.3	Bestimmung des Bakterientiters.....	35
2.4	Anzucht von <i>C. glutamicum</i> für die Transkriptom-Sequenzierung.....	36
3	Molekularbiologische Methoden	37
3.1	Isolierung von Gesamt-DNA aus <i>C. glutamicum</i>	37

3.2	Primer-Design.....	37
3.3	Polymerase-Kettenreaktion (PCR).....	38
3.4	Agarose-Gelelektrophorese.....	39
3.5	RNA-Aufreinigung mit Phenol-Chloroform-Isoamylalkohol (PCI).....	39
3.6	Kit-basierte Gesamt-RNA-Isolierung aus <i>C. glutamicum</i>	40
3.7	Zwei-Phasen-Extraktion von Gesamt-RNA aus <i>C. glutamicum</i>	41
3.8	Überprüfung der RNA auf DNA-Verunreinigung mittels PCR.....	42
3.9	Qualitäts- und Quantitätskontrolle von RNA.....	43
3.10	Abreicherung von rRNA.....	44
3.11	RNA-Fragmentierung.....	45
3.12	Anreicherung von nativen Transkriptfragmenten.....	46
3.13	Prozessierung von 5'-Tri- und Diphosphatenden zu 5'-Monophosphatenden.....	46
3.14	Phosphorylierung der 5'-Transkriptenden.....	46
3.15	RNA-Adapter-Ligation.....	46
3.16	Reverse Transkription.....	47
3.17	Amplifikation der cDNA-Bibliotheken nach der reversen Transkription.....	48
3.18	Aufreinigung von cDNA-Bibliotheken.....	48
3.19	Quantifizierungs- und Qualitätskontrolle von cDNA-Bibliotheken.....	49
3.20	Sequenzierung von cDNA-Bibliotheken.....	49
4	Bioinformatische Analysen.....	51
4.1	<i>Read</i> -Kartierung und Visualisierung der RNAseq-Daten.....	51
4.2	Detektion von Transkriptionsstarts.....	52
4.3	Suche von Promotor- und Ribosomenbindestellen-Motiven.....	52
4.4	Detektion neuer Transkripte.....	53
4.5	Identifizierung von offenen Leserastern und die Suche nach homologen Proteinen.....	54
4.6	Detektion und Definition von Operon-Strukturen.....	54
4.7	Bestimmung und Definition von Transkriptenden.....	55
IV	ERGEBNISSE.....	57
1	Die Entwicklung zweier RNAseq-Vorbereitungsprotokolle: Sequenzierung nativer 5'-Transkriptenden und Vollängentranskripte.....	57
2	Die Generierung der Transkriptom-Daten und die Kartierung der erhaltenen <i>Reads</i> auf das Genom von <i>Corynebacterium glutamicum</i>.....	62
3	Die Auswertung des Datensatzes der nativen 5'-Transkriptenden.....	63
3.1	Detektion und Klassifizierung von Transkriptionsstarts.....	63
3.2	Reannotation von proteinkodierenden Genen.....	66
3.3	Umfassende Motivsuche von σ^A -abhängigen Promotoren.....	67
3.4	Definition und Charakterisierung von 5'-UTRs.....	69
3.5	Analyse und Motivsuche von Ribosomenbindestellen.....	75

4	Die Auswertung des Datensatzes der Vollängentranskripte	78
4.1	Detektion und Beschreibung von Operon-Strukturen.....	78
4.1.1	Analyse und Motivsuche von internen Ribosomenbindestellen	82
4.2	Detektion und Klassifizierung neuer Transkripte.....	84
4.3	Detektion von Transkriptenden und Vergleich mit vorhergesagten ρ -unabhängigen Terminatoren.....	85
V	DISKUSSION.....	89
1	Die zwei etablierten RNAseq-Protokolle ermöglichen eine genomweite Analyse der transkriptionellen Organisation mit Einzelnukleotid-Auflösung	89
2	Der Datensatz der nativen 5'-Transkriptenden liefert eine enorme Anzahl an Transkriptionsstarts und ermöglicht die Identifizierung zugehöriger σ^A -abhängiger Promotoren in <i>C. glutamicum</i>	91
3	Die Transkriptomsequenzierung zeigt eine hohe Abundanz von <i>leaderless</i> mRNAs in <i>C. glutamicum</i>	93
4	Die Charakterisierung von 5'-untranslatierten Regionen gibt Hinweise auf neue regulatorische Elemente	95
5	Die umfassende Analyse der Transkriptomsequenzierung liefert das Motiv „AGGag“ als Ribosomenbindestelle für <i>C. glutamicum</i>	99
6	Der Datensatz der Vollängentranskripte klärt die Operon-Strukturen für <i>C. glutamicum</i> auf.....	101
7	Die Detektion neuer, unbekannter Transkripte liefert eine erhebliche Anzahl von <i>antisense</i> RNAs.....	103
8	Die Auswertung des Vollängentranskript-Datensatzes ermöglicht eine genaue Untersuchung der Transkriptenden	105
VI	AUSBLICK.....	108
VII	LITERATUR.....	112
VIII	ANHANG.....	129
1	Abkürzungsverzeichnis.....	129
2	Ergänzende Abbildungen	131
3	Ergänzende Tabellen	134
4	Abbildungsverzeichnis.....	135
5	Tabellenverzeichnis.....	136

I Zusammenfassung der Arbeit

Das Transkriptom beschreibt die Gesamtheit aller RNA-Moleküle einer Zelle zu einem gegebenen Zeitpunkt, welche die mRNA, rRNA, tRNA und nicht-kodierender RNA umfassen. Anders als das Genom, welches statisch in einer Zelle vorliegt, kann das Transkriptom als Anpassung an Umgebungsbedingungen variieren. Die Genomsequenz und die entsprechende Annotation von *Corynebacterium glutamicum* liegen seit nunmehr einem Jahrzehnt vor und stellen somit die Basis für die Aufklärung der transkriptionellen Organisation dieses industriell relevanten Modellorganismus dar. Die Transkriptomanalyse mittels Hochdurchsatz-Sequenzierung (RNAseq) findet seit Kurzem Anwendung und erlaubt gleichzeitig eine äußerst sensitive, qualitative Detektion von Transkripten mit Einzelnukleotid-Auflösung.

Infolgedessen sollten in dieser Arbeit RNAseq-Methoden für *C. glutamicum* etabliert werden, die eine genomweite, vollständige Erfassung des Transkriptoms unter gewählten Bedingungen erlaubt und dadurch die Validierung sowie Verbesserung der Genomannotation als auch die Identifizierung wichtiger Elemente der Transkription und Translation ermöglicht. Somit wurden im Rahmen dieser Arbeit zwei RNAseq-Protokolle entwickelt: das Volllängentranskript-Protokoll, welches Transkripte in ihrer vollen Länge abdeckt und das Protokoll zur Sequenzierung nativer 5'-Transkriptenden. Bei der Applikation des Volllängentranskript-Protokolls wurde die RNA aus *C. glutamicum*-Zellen isoliert, die unter verschiedenen Bedingungen kultiviert wurden (in CGXII-Minimalmedium und LB-Komplexmedium ohne Stressoren, in CGXII-Minimalmedium gestresst mit 10 % Ethanol, 10 % NaCl, 90 % gelöstem Sauerstoff, 1 % Wasserstoffperoxid, 2 mM Diamid, Hitzeschock bei 50 °C und Kälteschock bei 4 °C für je 15 min. vor Zellernte nach Erreichen der exponentiellen Wachstumsphase), da die Identifizierung von möglichst vielen Transkripten im Vordergrund stand. Für die Verwendung des Protokolls für native 5'-Transkriptenden wurde RNA aus *C. glutamicum*-Zellen isoliert, die in CGXII-Minimalmedium und LB-Komplexmedium kultiviert wurden, um die Aktivität des Sigmafaktors σ^A zu gewährleisten und die Sequenzierung von σ^A -entstammenden Transkriptionsstarts (TSSs) zu ermöglichen. Mit der Entwicklung und Anwendung beider RNAseq-Protokolle war es nun erstmals möglich, genomweit die vollständige transkriptionelle Organisation von *C. glutamicum* zu beschreiben.

Die Analyse der erlangten Sequenzierdaten ermöglichte die Identifizierung von 2.591 TSSs, die die Basis für weitere Transkriptom-Untersuchungen bildete. Eine nähere Betrachtung der gefundenen TSSs zeigte, dass unter den gewählten Bedingungen *C. glutamicum* für 30 % der Gene mehr als einen TSS verwenden kann. Zusätzlich konnte vor mehr als 2.500 TSSs das -10-Promotormotiv „TAnnnT“ und vor etwas mehr als 700 TSSs das -35-Motiv „ttgna“ gefunden werden sowie für 7 % der proteinkodierenden Gene in *C. glutamicum* eine Reannotation des Translationsstarts vorgenommen werden. Es zeigte sich, dass etwa

30 % der gefundenen Transkripte keine 5'-untranslatierte Region (5'-UTR) und etwa 25 % der Transkripte 5'-UTRs länger als 100 nt besitzen, was Hinweise auf bisher unbekannte regulatorische Elemente gibt. Die Analyse und Motivsuche von Ribosomenbindestellen (RBSs) innerhalb der detektierten 5'-UTRs ergab für *C. glutamicum* das Motiv „AGGag“, welches das perfekte Revers-Komplement zum 3'-Ende der 16S rRNA darstellt. Weiterhin lieferte die Daten-Analyse Operon-Strukturen für *C. glutamicum*, die unter den gewählten Bedingungen etwa 30 % der Gene als Monocistrons und etwa 70 % der Gene als Primär-Operons mit bis zu 16 Genen klassifizieren, in denen zahlreiche (über 500) alternative Sub-Operons liegen. Außerdem konnte eine Vielfalt an *antisense* RNAs (insgesamt 700), intragenischen (insgesamt 186) und einigen intergenischen Transkripten (insgesamt 30) gefunden und näher charakterisiert werden. Des Weiteren konnte gezeigt werden, dass der Großteil der Transkripte (über 50 %) in *C. glutamicum* unter den untersuchten Bedingungen ein graduelles, schrittweise auslaufendes 3'-Ende zeigt, was zum Einen auf eine kontinuierliche Neusynthese der RNA und zum Anderen auf einen schnellen Abbau der Transkripte durch 3'-5'-Exoribonukleasen hinweist. Für 23,2 % der Transkripte wurde ein deutliches Transkriptende im Bereich der vorhergesagten, ρ -unabhängigen Terminatoren bestimmt, wobei überraschenderweise die Mehrzahl der Transkripte im Stamm 1 (72 %) und nur einige in der Schleife (18,6 %), im Stamm 2 (7,5 %) oder in der Poly(T)-Sequenz (1,9 %) enden. Die Untersuchung der Transkriptenden liefert somit ein Indiz dafür, dass die Prozesse der ρ -unabhängige Termination in *C. glutamicum* noch nicht vollständig verstanden sind.

Mithilfe der entwickelten und auch *C. glutamicum* angewendeten RNAseq-Protokolle konnte eine umfassende Transkriptom-Karte erstellt werden, die eine nahezu vollständige Definition von genetischen Elementen wie Promotoren, Gen-Starts und -Stopps, 5'-UTRs, RBS, Transkript-Starts und -Stopps, repräsentiert. Diese Karte liefert eine ideale Basis für weitere Analysen der Transkriptionsregulation, insbesondere auf RNA-Ebene. Ferner eignen sich diese vorgestellten RNAseq-Protokolle auch in der Anwendung für andere Organismen und besitzen auch das Potential für quantitative Analysen.

II Einleitung

1 Das Bakterium *Corynebacterium glutamicum*

Die vorliegende Arbeit beschäftigt sich mit der genomweiten Transkriptomanalyse von *Corynebacterium glutamicum*. In den folgenden Abschnitten soll dieses Bakterium, seine Taxonomie, die industrielle Relevanz und die Genomannotation näher beschrieben werden.

1.1 Taxonomie und Charakteristika der Gattung *Corynebacterium*

Corynebacterium glutamicum gehört der Klasse *Actinobacteria* an, die eine Gruppe von Gram-positiven Bakterien mit einem hohen G+C-Gehalt der DNA (in der Regel > 50 %) darstellen und sich phylogenetisch von der Gruppe von Gram-positiven Bakterien mit einem niedrigen G+C-Gehalt der DNA unterscheiden [Stackebrandt & Woese, 1981; Stackebrandt *et al.*, 1997]. Die Mitglieder dieser Klasse besitzen eine 16S rDNA/rRNA-Sequenzidentität von > 80 % und weisen spezifische 16S rDNA/rRNA-Sequenzmotive auf [Stackebrandt *et al.*, 1997].

Eine weitere wichtige taxonomische Einteilung von *C. glutamicum* findet auf der Ebene der Unterordnung der *Corynebacterineae* statt, die die Familien *Corynebacteriaceae*, *Mycobacteriaceae*, *Nocardiaceae*, *Dietziaceae* und *Tsukamurellaceae* umfasst [Stackebrandt *et al.*, 1997]. Die meisten Vertreter dieser Gruppe zeichnen sich durch einen besonderen Aufbau der Zellwand aus. Die Besonderheit der Zellwand ist durch die eingelagerten langkettigen, verzweigten Fettsäuren (Mykolsäuren) gegeben, die über Polysaccharide (Arabinogalactan) an das Peptidoglykan (quervernetzt mit *meso*-Diaminopimelinsäure) der Plasmamembran gebunden vorliegen [Lechevalier & Lechevalier, 1970]. Eine weitere Besonderheit des Zellwandaufbaus dieser Unterordnung ist eine zusätzliche Schicht aus Lipiden, einfachen Mykolsäuren und Proteinen, die auf den verzweigten Mykolsäuren aufliegt und die Außenschicht der Zellwand bildet (Abbildung II-1 a). Die Bildung einer solchen äußeren Schicht als zusätzliche Permeabilitätsbarriere ist sonst für Gram-negative Bakterien charakteristisch, die sich aber im detaillierten Aufbau unterscheidet [Puech *et al.*, 2001]. Dieser Zellwandaufbau sorgt bei den Bakterien für eine besondere Resistenz gegenüber Umwelteinflüssen, Säuren sowie Antibiotika und ist bei den pathogenen Vertretern mit für deren Pathogenität verantwortlich [Liu *et al.*, 1996; Ramulu *et al.*, 2006]. Aufgrund der ersten chemotaxonomischen Studien bezüglich der Peptidoglykanstruktur sowie Lipid- und Mykolsäurezusammensetzung der Zellwand für die Gattungen *Corynebacterium*, *Mycobacterium* und *Nocardia* werden heute noch alle Gattungen mit diesem Zellwand-Profil in die sogenannte CMN-Gruppe eingeteilt [Barksdale, 1970; Lechevalier & Lechevalier, 1970; Bansal-Mutalik & Nikaido, 2011].

Die Gattung *Corynebacterium* wird erstmals im Jahre 1896 von Lehmann und Neumann mit den pathogenen Vertretern *Corynebacterium diphtheriae* (ehemals Diphtheriebacillus), *Bacillus pseudotuberculosis ovis* (heute *Corynebacterium pseudotuberculosis* [Breed *et al.*, 1948; Baird & Fontaine, 2007]) und *Bacillus pseudotuberculosis murium* (heute *Corynebacterium kutscheri* [Migula, 1900; Bergey *et al.*, 1925]) eingeführt, die von nicht-motilen, nicht-sporulierenden Bakterien mit einer Stäbchenform und keulig angeschwollenen Enden gebildet wird [Lehmann & Neumann, 1896]. Zwei Jahre vor der Einführung der Gattung *Corynebacterium* wurde der erste Vertreter als Diphtheriebacillus (später als *Corynebacterium diphtheriae* benannt [Lehmann & Neumann, 1896]) und die diphtherieähnlichen Bakterien von Dr. Theodor Escherich als Pseudo-Diphtheriebacilli im Rahmen der Beschreibung der epidemischen Diphtherie charakterisiert [Escherich, 1894]. Die keulenförmige Zellstruktur (griech.: *coryne* = Keule) und das V-förmige Abwinkeln (*snapping division*) der Tochterzellen während der Zellteilung (Abbildung II-1 b), welches durch die asymmetrische Trennung der neu gebildeten, mehrschichtigen Zellwand entsteht, zählt als ein weiteres wichtiges Charakteristikum der Gattung *Corynebacterium* [Thanky *et al.*, 2007; Letek *et al.*, 2008].

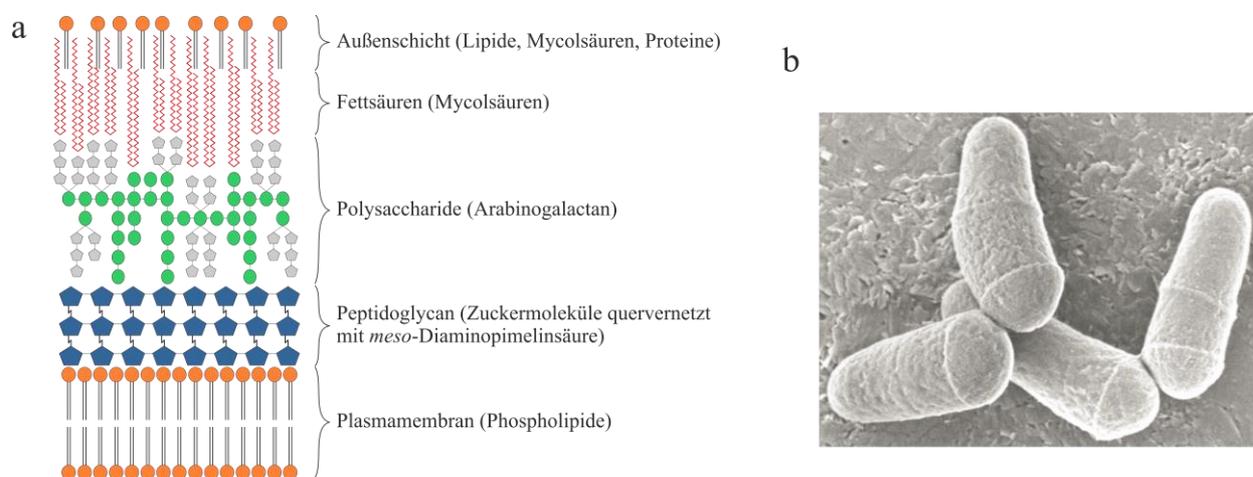


Abbildung II-1: Morphologie und Zellwandaufbau von *C. glutamicum*

a. Schematische Darstellung der mehrschichtigen Zellwand-Struktur von *C. glutamicum* (nach [Lechevalier & Lechevalier, 1970; Puech *et al.*, 2001]).

b. Elektronenmikroskopische Aufnahme von *C. glutamicum*. Die Keulenform und die V-förmige Ausbildung der *snapping division* bei der Zellteilung sind deutlich zu erkennen (aus [Wachi, 2013]).

Aufgrund der 16S rDNA/rRNA-Sequenzidentität lassen sich die Spezies der Gattung *Corynebacterium* als monophyletische Gruppe beschreiben [Pascual *et al.*, 1995; Ruimy *et al.*, 1995]. Nichtsdestotrotz zeigt die Gattung *Corynebacterium* eine phylogenetische Tiefe, zu der derzeit 94 Spezies gehören (*List of*

Prokaryotic Names with Standing in Nomenclature; Stand April 2014; [Euzéby, 1997; Parte, 2014]), die auch Vertreter mit fehlenden Mykolsäuren in ihrer Zellwand besitzt (wie *C. amycolatum* [Collins *et al.*, 1988], *C. kroppenstedtii* [Collins *et al.*, 1998], und *C. atypicum* [Hall *et al.*, 2003]), was der Gattung eine gewisse Heterogenität verleiht. Für 57 dieser zahlreichen *Corynebacterium*-Spezies liegen Genomsequenzen vor, die industriell und medizinisch interessant sind, wie *C. glutamicum* als Aminosäureproduzent [Mitsuhashi, 2014], *C. diphtheriae* als Erreger der Diphtherie [Escherich, 1894; Lehmann & Neumann, 1896] und *C. jeikeium* als Erreger der Endokarditis [Jackman *et al.*, 1987].

Eine Besonderheit des Genus *Corynebacterium* ist die starke Syntenie der Genome der verschiedenen Spezies, die im Zuge der Evolution nur eine geringe Reorganisation erfahren haben. Eine mögliche Erklärung für die geringe Genom-Reorganisation innerhalb der Gattung *Corynebacterium* liefert das fehlende Rekombinations- und Reparatursystem (kodiert von *recBCD* in *E. coli*), welches für chromosomale Inversionen verantwortlich ist [Nakamura *et al.*, 2003; Ventura *et al.*, 2007]. Die Entstehung der Artenvielfalt innerhalb dieser Gattung liegt im Verlust und Gewinn von Genen sowie in Einzelnukleotid-Substitutionen begründet [Ventura *et al.*, 2007]. Die Artenvielfalt der Gattung *Corynebacterium* zeigt sich in der Besiedelung verschiedenster Lebensräume. Die apathogenen Spezies *C. glutamicum*, *C. efficiens*, *C. variabile*, *C. casei* oder *C. halotolerans* konnten im Salzwasser, Abwasser, Boden, auf Gemüse oder gereiftem Käse [Kinoshita *et al.*, 1958; Brennan *et al.*, 2001; Fudou *et al.*, 2002; Chen *et al.*, 2004] gefunden werden. Andere Spezies wie *C. ciconiae* oder *C. caspium*, wurden von/aus Tieren isoliert [Collins *et al.*, 2004; Fernández-Garayzábal *et al.*, 2004] und humanpathogene Arten wie *C. urealyticum*, *C. diphtheriae* oder *C. jeikeium*, kommen auf bestimmten Partien der menschlichen Haut vor, die durch Eindringen in andere Bereiche zahlreiche Infektionen auslösen können [Escherich, 1894; Lehmann & Neumann, 1896; Jackman *et al.*, 1987; Soriano & Tauch, 2008].

Die Spezies *Corynebacterium glutamicum* lässt sich also aufgrund der 16S rDNA/rRNA-Sequenzanalysen, des G+C-Gehalts der DNA, des Zellwandaufbaus und ihrer Morphologie in die folgende hierarchische Systematik einteilen [Stackebrandt *et al.*, 1997]:

Klasse *Actinobacteria* – Unterklasse *Actinobacteridae* – Ordnung *Actinomycetales* – Unterordnung *Corynebacterineae* – Familie *Corynebacteriaceae* – Gattung *Corynebacterium*.

1.2 Entdeckung und wirtschaftliche Bedeutung von *Corynebacterium glutamicum*

Im Jahr 1957 gelang es Dr. Shukuo Kinoshita und seinen Kollegen das Bakterium *Corynebacterium glutamicum* im Rahmen eines Screenings nach Mikroorganismen, die Glutaminsäure extrazellulär anreichern, zu isolieren. Sie benannten dieses Isolat zunächst als *Micrococcus glutamicus* Nr. 534, das später aufgrund von taxonomischen Studien den heutigen Namen erhielt [Kinoshita *et al.*, 1958]. Das Interesse an dieser Aminosäure nahm im Jahr 1909 seinen Anfang als Prof. Dr. Kikunae Ikeda Studien

mit dem Ziel aufnahm die Ernährung und die damals kurze Lebenserwartung von Japanern zu verbessern [Ikeda, 2002; Kinoshita, 2005]. Dabei gelang es ihm, während seiner Untersuchungen am Seetang, die geschmacksverstärkende Wirkung von Mononatriumglutamat (Natriumsalz der Glutaminsäure) zu beschreiben und die davon ausgehende fünfte Geschmacksqualität (damals die sechste, da scharf ebenfalls als Geschmacksqualität zählte) „umami“ als fleischig, herzhaft einzuführen [Ikeda, 2002]. Von diesem Zeitpunkt an war das Interesse - besonders in Japan - sehr groß, Glutaminsäure im großen Maßstab zu gewinnen. So startete man zunächst die Gewinnung durch chemischen Abbau von Sojabohnen und Weizenproteinen, gefolgt von der fermentativen Produktion von α -Ketoglutarsäure, einer Vorstufe von Glutaminsäure, die durch Transformation zu Glutaminsäure umgewandelt werden musste [Udaka, 2008]. Mit der Isolierung von *C. glutamicum* fand dann die direkte fermentative Produktion von Glutaminsäure ihren Anfang und eröffnete ferner die Möglichkeit zur Produktion weiterer Aminosäuren, wie Lysin, Threonin, Asparaginsäure, Valin, Alanin, Tryptophan, usw., die heute Verwendung als Futtermittelzusatz, Süßstoff sowie als Komponenten von Medikamenten und Kosmetika finden [Hermann, 2003; Kinoshita, 2005; Udaka, 2008].

Seither haben die apathogenen *Corynebacterium* Spezies (wie z.B. *C. glutamicum*, *C. ammoniagenes* und *C. efficiens*), aber auch andere Mikroorganismen wie *Escherichia coli* oder *Saccharomyces cerevisiae*, aufgrund ihres breiten Spektrums an Stoffwechselleistungen und der Möglichkeit zur fermentativen Gewinnung verschiedenster Produkte eine erhebliche wirtschaftliche Relevanz erlangt [Eggeling & Bott, 2005]. Zu den Bereichen, in denen *Corynebakterien* Spezies Einsatz finden, zählen neben der Aminosäureproduktion u.a. die Herstellung und Konversion von Steroiden [Constantinides, 1980; Decréau *et al.*, 2003], die Käseerzeugung [Schröder *et al.*, 2011], die Herstellung von Nukleotiden und Vitaminen [Sahm & Eggeling, 1999] sowie die Bioremediation [Calabrese *et al.*, 1993; Mateos *et al.*, 2010]. Nach wie vor eignet sich insbesondere *C. glutamicum* hervorragend als Aminosäure-Produktionsstamm, da dieser wichtige ökonomische Vorteile mit sich bringt. Aufgrund der fehlenden Pathogenität und Toxizität besitzt diese Spezies die Klassifikation *GRAS* (*generally regarded as safe*). Dadurch wird Vertrieb und die Zulassung der von *C. glutamicum* gebildeten Produkte in der Lebensmittelindustrie erheblich erleichtert. Weiterhin ist *C. glutamicum* in der Lage mit hohen Wachstumsraten in kostengünstigen Nährmedien zu wachsen, verschiedenste Kohlenstoff- und Energiequellen zu verwerten und ist im Recycling überflüssiger Aminosäuren begrenzt [Wohlleben *et al.*, 1993; Gopinath *et al.*, 2012; Ikeda & Takeno, 2013]. Diese Vorteile sowie der stetige Bedarf führten zur kommerziellen Produktion von Aminosäuren u.a. durch *Ajinomoto Co., Inc*, *Evonik Industries AG* und *Archer Daniels Midland Co.*

Die industrielle Relevanz von *C. glutamicum* sowie die Etablierung neuer molekularbiologischer Methoden und der Fortschritt biotechnologischer Herstellungsverfahren trieben die Entwicklung und Optimierung gentechnischer Methoden für *C. glutamicum* voran, sodass die Konstruktion von zunehmend effizienterer Produktionsstämme ermöglicht wurde [Wohlleben *et al.*, 1993]. Mittlerweile ist die

Genomsequenz von *C. glutamicum* vollständig entschlüsselt [Ikeda & Nakagawa, 2003; Kalinowski *et al.*, 2003], viele hilfreiche bioinformatische Software und Datenbanken sind entwickelt [Meyer *et al.*, 2003; Dondrup, Albaum, *et al.*, 2009; Dondrup, Hüser, *et al.*, 2009; Schneider *et al.*, 2010; Blom *et al.*, 2011; Pauling *et al.*, 2012; Kessler *et al.*, 2013] und auch neue, verbesserte Hochdurchsatz-Methoden sind verfügbar [Baldrian & López-Mondéjar, 2014], die eine optimale Grundlage für weitere Analysen u.a. im Bereich der Physiologie, Genetik und transkriptionellen Regulationsmechanismen bieten [Eggeling & Bott, 2005; Burkovski, 2008; Inui & Yukawa, 2013]. Diese Methoden und Analysen ermöglichen nun die Gewinnung genomweiter Erkenntnisse für *C. glutamicum* sowie das Verständnis komplexer Vorgänge in der Zelle, um z.B. verbesserte Produktionsstämme zu erstellen.

1.3 Annotation sowie aktuelle Version des Genoms von *Corynebacterium glutamicum*

In einem gemeinsamen Projekt der *Degussa AG* (heute *Evonik Industries AG*) und dem Lehrstuhl für Genetik der Universität Bielefeld wurde im Jahr 2003 das Genom von *C. glutamicum* (Typstamm ATCC 13032) mithilfe von 116 BAC- (*Bacterial Artificial Chromosome*) und Cosmid-Klonen [Tauch *et al.*, 2002] und der *Shotgun*-Sequenzierung entschlüsselt und erreichte eine 5 bis 9fache Sequenzabdeckung (*Coverage*) [Kalinowski *et al.*, 2003]. Die Anordnung der erhaltenen Nukleotidsequenzen (*Assembly*) der BAC- und Cosmid-Klone lieferte eine Gesamt-Genomsequenz von 3.282.708 bp mit einem G+C-Gehalt von 53,8 %. Nach der darauffolgenden automatischen (mittels *CRITICA* [Badger & Olsen, 1999] und *GLIMMER* [Delcher *et al.*, 1999]) und manuellen Genannotation (Vergleich über *TBLASTN* [Altschul *et al.*, 1990] mit anderen *Actinomycetales*-Genomsequenzen) innerhalb der *GenDB*-Datenbank [Meyer *et al.*, 2003] konnten 3.002 proteinkodierende Gene identifiziert werden, von denen etwa 83 % eine Ähnlichkeit zu bekannten Genen aus anderen Organismen aufwiesen [Kalinowski *et al.*, 2003]. Weiterhin besitzt das *C. glutamicum*-Genom einige genomische Regionen, die einen für *C. glutamicum* untypischen G+C-Gehalt besitzen und somit auf horizontalen Gentransfer hinweisen: Die HGC1- und LGC1-Region, die einen hohen (bis zu 66 %) bzw. einen niedrigen (41 bis 49 %) G+C-Gehalt aufweisen sowie drei Bakteriophagen (CGP1, CGP2 und CGP3), die ebenfalls einen niedrigen G+C-Gehalt aufzeigen [Kalinowski *et al.*, 2003]. Die Genomsequenz von *C. glutamicum* wurde außerdem zeitgleich von einem weiteren Forscherteam der Universität Kitasato (Tokyo, Japan) in Zusammenarbeit mit *Kyowa Hakko Kirin Co., Ltd.* veröffentlicht, die im Vergleich zur Bielefelder Genomsequenz nur wenige Abweichungen enthält [Ikeda & Nakagawa, 2003]. Die Entschlüsselung der Genomsequenzen von *C. glutamicum* lieferte wertvolle Informationen für die weitere Grundlagenforschung und die industrielle Anwendung [Ikeda & Nakagawa, 2003; Kalinowski *et al.*, 2003].

Seit der Veröffentlichung der Genome von *C. glutamicum* in 2003 [Ikeda & Nakagawa, 2003; Kalinowski *et al.*, 2003] konnte das Wissen über viele Genfunktionen und –annotationen deutlich erweitert werden,

was sich in über 800 Veröffentlichungen niederschlägt (*PubMed*, Stand April 2014). Somit wurde basierend auf der entschlüsselten Genomsequenz u.a. die transkriptionelle Regulation in *C. glutamicum* extensiv untersucht [Brinkrolf *et al.*, 2007] und die Ergebnisse in der *CoryneRegNet*-Datenbank, hinterlegt. Diese Analyse offenbarte für *C. glutamicum* ein komplexes regulatorisches Netzwerk mit derzeit 98 Transkriptionsregulatoren und 1.441 regulatorischen Interaktionen (*CoryneRegNet*, Stand April 2014 [Pauling *et al.*, 2012]). Außerdem konnte festgestellt werden, dass *C. glutamicum* insgesamt sieben Sigmafaktoren besitzt, die die Transkription global regulieren, indem sie als Teil des RNA-Polymerase-Holoenzym spezifische Promotoren erkennen und binden [Pátek & Nešvera, 2011]. In den letzten Jahrzehnten konnten Promotoren von etwa 200 Genen identifiziert werden, wobei für die Mehrzahl der Gene die Promotoren unbekannt bleiben.

Diese neuerhaltenen Daten wurden innerhalb der *GenDB*-Datenbank gesammelt und charakterisiert, sodass das aktuelle Genom von *C. glutamicum* einige reannotierte Genstarts und einige neue Gene basierend auf verbesserten Genvorhersagen und –annotationen enthält und aktuell 3.039 proteinkodierende Gene umfasst. Weiterhin konnten mit der kürzlich durchgeführten Sequenzierung von kleinen RNAs für *C. glutamicum* [Mentz *et al.*, 2013] u.a. vier nicht-kodierende RNAs identifiziert werden, die zuvor innerhalb der *Rfam*-Datenbank an ähnlichen Positionen vorhergesagt wurden [Gardner *et al.*, 2011]: Die 4.5S RNA, 6C RNA, tmRNA, und M1 RNA (s. Tabelle II-1). Somit umfasst das aktualisierte Genom von *C. glutamicum* derzeit 3.043 Gene, welches als Grundlage für die vorliegende Arbeit diente.

Tabelle II-1: Vier mittels RNAseq verifizierten [Mentz *et al.*, 2013; Mentz, 2013] und *Rfam* vorhergesagten [Gardner *et al.*, 2011] sRNAs für *C. glutamicum*

Name	ID	Rfam-Vorhersage			Strang	RNAseq-Detektion	
		Start	Stopp	Bit Score		Start	Stopp
4.5S RNA	RF00169	256.431	256.527	64,76	+	256.434	256.543
6C RNA	RF01066	314.705	314.782	32,73	+	314.679	314.765
tmRNA	RF00023	848.547	848.923	134,11	+	848.547	848.943
M1 RNA	RF00010	2.342.998	2.342.591	256,02	-	2.343.003	2.342.586

2 Die Transkriptionsmechanismen in Prokaryoten

Das prokaryotische Genom enthält in der Regel mehrere Tausend Gene, die für Proteine und für regulatorische RNA kodieren. Das Genom liegt statisch vor und bleibt bei der Zellteilung über die Replikation unverändert (abgesehen von Mutationen) erhalten. Die Transkription beschreibt die Synthese der RNA anhand der DNA, wobei nur benötigte DNA-Abschnitte abgelesen werden und das Transkriptom (Gesamtheit aller RNA-Spezies einer Zelle zu einem gegebenen Zeitpunkt) bilden. Darauf folgt die Translation der synthetisierten RNA in Proteine oder die direkte Funktion als regulatorische RNA. Somit stellt die RNA ein Bindeglied zwischen dem Genom und der Umsetzung der genetischen Information dar. Da das Transkriptom sowie das Proteom (Gesamtheit aller Proteine einer Zelle zu einem gegebenen Zeitpunkt) im Gegensatz zum Genom als Anpassung an die Umgebungsbedingungen variieren, werden die Prozesse der Transkription und Translation reguliert, um die Synthese von Proteinen und regulatorischen RNAs in benötigten Mengen zu gewährleisten. Die Transkription ist dabei der erste Kontrollpunkt und kann direkt während der Initiation durch vielerlei Faktoren beeinflusst werden.

Der Transkriptionsprozess an sich wird in die drei Schritte Transkriptionsinitiation, Transkriptionselongation und Transkriptionstermination eingeteilt und wird in den folgenden Abschnitten näher beschrieben.

2.1 Struktur und Aufbau der DNA-abhängigen RNA-Polymerase

Die Information für die Synthese der RNA liefert der nicht-codogene Strang der DNA (3'-5') und wird an spezifischen DNA-Sequenzen (Promotorregionen) eingeleitet, die sich einige Nukleotide stromaufwärts des Transkriptionsstarts (+1) befinden. Während der RNA-Synthese wird ein dem nicht-codogenen Strang komplementärer RNA-Strang durch Verknüpfung der Nukleosidtriphosphate (NTPs; Adenosintriphosphat, Guanosintriphosphat, Cytidintriphosphat und Uridintriphosphat) aufgebaut. Das Enzym, welches für die Transkription zuständig ist, ist die DNA-abhängige RNA-Polymerase, die in Bakterien als Kernenzym und Holoenzym vorkommt und die Synthese aller RNA-Spezies (mRNA, rRNA, tRNA und sRNA) durchführt.

Das Kernenzym der RNA-Polymerase, welches die Transkription, jedoch aber nicht die gezielte Transkriptionsinitiation am Promotor durchführen kann, setzt sich aus fünf Untereinheiten zusammen: einem α -Dimer, der β -, β' - und ω -Untereinheit [Vassylyev *et al.*, 2002]. Diese sind in Eukaryoten, Archaeen und Prokaryoten in ihrer Struktur, Faltung, Funktion und Position innerhalb des Kernenzyms konserviert [Ebright, 2000]. Das aktive Zentrum der RNA-Polymerase wird durch die β - und β' -Untereinheit gebildet, die beide in der Lage sind die DNA und die synthetisierte RNA während der Transkription zu binden [Ebright, 2000; Browning & Busby, 2004]. Jede α -Untereinheit besteht aus zwei

unabhängig gefalteten Domänen, einer N-terminalen Domäne (α NTD) und einer C-terminalen Domäne (α CTD), die über einige Aminosäuren, den sogenannten Linker, verbunden sind [Busby & Ebright, 1994; Browning & Busby, 2004]. Das Dimer α_2 ist für den Zusammenbau der Untereinheiten β und β' verantwortlich, wobei die dimerisierten α NTD-Module direkt mit der β - und β' -Untereinheit interagieren und die dimerisierten α CTD-Module an bestimmten Promotoren an der DNA (dem sogenannten UP-Element stromaufwärts der -35-Promotorregion) binden und mit Transkriptionsregulatoren in Wechselwirkung treten können [Busby & Ebright, 1994; Ebright, 2000]. Die kleinste Untereinheit der RNA-Polymerase (ω) fungiert als Chaperon exklusiv für die β' -Untereinheit und rekrutiert diese für den Zusammenbau mit den $\alpha_2\beta$ -Untereinheiten [Mukherjee *et al.*, 1999; Mathew & Chatterji, 2006]. Das Chaperon GroEL ist nicht Teil der RNA-Polymerase, kann aber mit dieser assoziieren und ist für die korrekte Faltung der übrigen Untereinheiten verantwortlich [Houry *et al.*, 1999; Mukherjee *et al.*, 1999]. Somit folgt die Zusammenlagerung der einzelnen Untereinheiten der RNA-Polymerase keinem zufälligen, sondern einem systematischen Prozess: zunächst lagern sich die α -Untereinheiten zu α_2 zusammen; darauffolgend lagert sich die β -Untereinheit an zu $\alpha_2\beta$; parallel dazu bindet und faltet die ω -Untereinheit die β' -Untereinheit, wobei sich dann $\alpha_2\beta$ und $\beta'\omega$ zum Kernenzym ($\alpha_2\beta\beta'\omega$) zusammensetzen [Ishihama, 1980; Mathew & Chatterji, 2006]. Das Kernenzym besitzt eine V-förmige bzw. scherenförmige Struktur, die durch die großen β - und β' -Untereinheiten gegeben wird und eine feste Bindung der DNA gewährleistet [Vassylyev *et al.*, 2002; Browning & Busby, 2004]. Neben der korrekten Struktur der RNA-Polymerase werden für die Transkription Magnesiumionen benötigt, die vermutlich für die Bindung und Faltung der DNA an der RNA-Polymerase notwendig sind [Vassylyev *et al.*, 2002]. Das Holoenzym der RNA-Polymerase enthält zusätzlich zum Kernenzym eine σ -Untereinheit, die der RNA-Polymerase die Erkennung der Promotoren und die Bindung an die Promotorregion der DNA ermöglicht sowie die Entspiralisierung und Öffnung des DNA-Doppelstrangs einleitet [Vassylyev *et al.*, 2002; Browning & Busby, 2004]. Die σ -Untereinheit selbst kann strukturell und funktionell in bis zu vier Domänen (1 - 4) eingeteilt werden, die jeweils einzelne Abschnitte der Promotorregionen erkennen und binden (Abbildung II-2 a) [Browning & Busby, 2004; Pátek & Nešvera, 2011]. Die meisten Bakterien besitzen diverse σ -Faktoren, die unterschiedliche Promotoren erkennen und so die Transkription bestimmter Gene einleiten [Ishihama, 1988, 2000; Paget & Helmann, 2003]. Grundsätzlich werden bakterielle Sigmafaktoren strukturell und funktionsgemäß in die Familie der σ^{70} - oder σ^{54} -Untereinheiten eingeteilt [Pátek & Nešvera, 2011]. Die Benennung der beiden Sigmafaktor-Familien erfolgte nach dem Hauptsigmafaktor σ^{70} aus *E. coli*, der etwa 70 kDa groß ist und die Transkription von Haushaltsgenen initiiert [Silverstone *et al.*, 1972], und dem alternativen Sigmafaktor σ^{54} aus *E. coli* mit einer Größe von etwa 54 kDa, der die Transkription von Genen einleitet, dessen Produkte an der Stickstoffregulation beteiligt sind [Hirschman *et al.*, 1985]. Sigmafaktoren der σ^{70} -Familie wurden bisher in allen bekannten Bakterien gefunden, die konservierte -10- und -35-Promotormotive erkennen und binden, wohingegen Sigmafaktoren der σ^{54} -Familie recht selten zu finden sind [Pátek & Nešvera, 2011].

Tabelle II-2: Untereinheiten der RNA-Polymerase in *C. glutamicum* und dessen Funktion

Untereinheit	Gen	Molekulargewicht	Funktion
α	<i>cg0655 (rpoA)</i>	36,7 kDa	Zusammenbau und Interaktion der β - und β' -Untereinheit; Wechselwirkung mit Promotoren und Transkriptionsregulatoren
β	<i>cg0576 (rpoB)</i>	128,8 kDa	aktives Zentrum der RNA-Polymerase; Bindung der DNA und RNA
β'	<i>cg0577 (rpoC)</i>	147,3 kDa	aktives Zentrum der RNA-Polymerase; Bindung der DNA und RNA
ω	<i>cg1809</i>	10,3 kDa	Faltung der β' -Untereinheit; Zusammenbau von $\alpha_2\beta$ mit $\beta'\omega$
σ^A^*	<i>cg2092 (sigA)</i>	54,3 kDa	Erkennung und Bindung von Promotorregionen

* *C. glutamicum* besitzt sieben verschiedene σ -Faktoren, die unterschiedliche Promotorsequenzen erkennen und binden. Hier ist beispielhaft der primäre, essentielle σ -Faktor aufgeführt, der für die Transkription der Haushaltsgene initiiert.

Die RNA-Polymerase aus *C. glutamicum* setzt sich ebenfalls aus den Untereinheiten $\alpha_2\beta\beta'\omega$ zusammen (Tabelle II-1), die durch die Gene *cg0655 (rpoA)*, *cg0576 (rpoB)*, *cg0577 (rpoC)* und *cg1809* kodiert werden [Kalinowski *et al.*, 2003]. Des Weiteren besitzt *C. glutamicum* sieben verschiedene σ -Faktoren, die alle der σ^{70} -Familie zugeordnet sind: den essentiellen, primären σ^A -Faktor, der für die Transkription der Haushaltsgene während der exponentiellen Wachstumsphase zuständig ist und Promotoren mit den Konsensussequenzen TTGNCA (-35-Region) und GNTANANTNG (erweiterte -10-Region) erkennt [Pátek & Nešvera, 2011]; den primär-ähnlichen, jedoch nicht essentiellen σ^B -Faktor, der vermutlich dieselben Promotoren wie σ^A erkennt und die Transkription beim Übergang von der exponentiellen in die stationäre Wachstumsphase und in Antwort auf viele Stressbedingungen, wie Säure-, Salz-, Ethanol-, Kälte- und Hitzestress, einleitet [Larisch *et al.*, 2007; Pátek & Nešvera, 2011]; und fünf σ -Faktoren der ECF-Familie (σ^C , σ^D , σ^E , σ^H , σ^M), deren eigene Expression in Antwort auf verschiedene Stressbedingungen erfolgt (wie Oberflächen-, Hitze-, und oxidativer Stress) und deren Promotor-Erkennungssequenzen sich deutlich von denen der σ^A/σ^B -Faktoren unterscheiden [Pátek & Nešvera, 2011]. Die zugehörigen Gene sind dabei in spezifische Funktionen und Stress-Antworten involviert [Pátek & Nešvera, 2011].

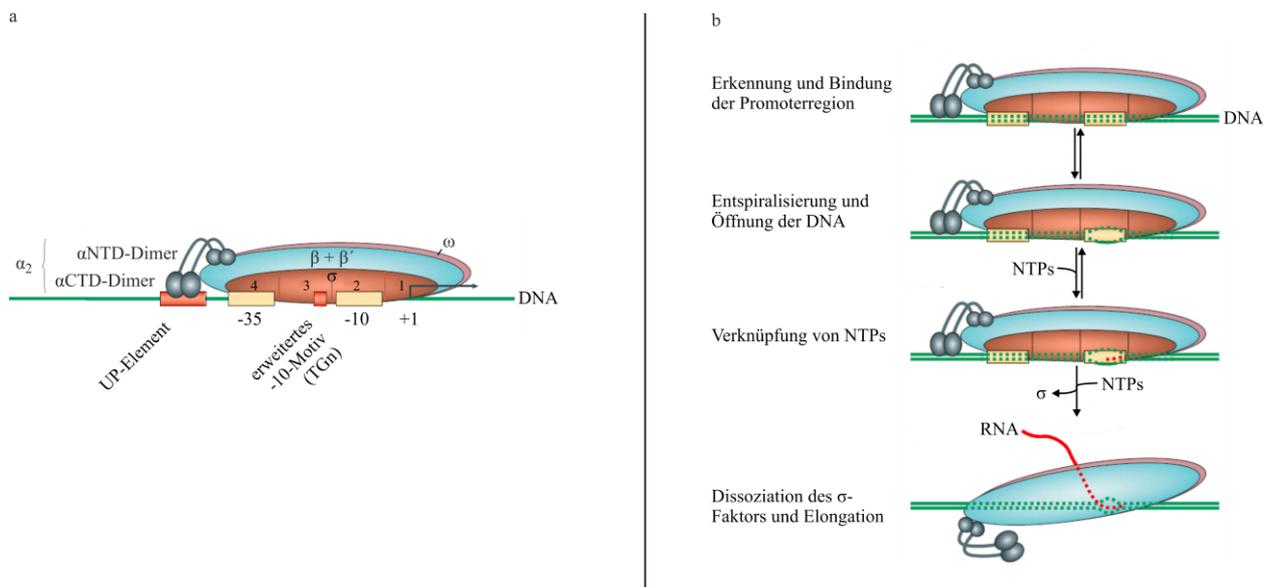


Abbildung II-2: Schematische Darstellung des Aufbaus der prokaryotischen RNA-Polymerase mit einer σ^{70} -Untereinheit, die Interaktion mit Promotorregionen und der Mechanismus der Transkriptionsinitiation (modifiziert nach [Browning & Busby, 2004])

Die einzelnen Untereinheiten des RNA-Polymerase-Holoenzym sind farblich dargestellt. Der DNA-Strang ist in grün und die NTPs bzw. die synthetisierte RNA in rot gezeigt.

a. Die α -Untereinheiten (grau), die sich jeweils aus den Domänen α NTD und α CTD zusammensetzen, interagieren mit dem UP-Element des Promotors (größere rote Box) und der β - bzw. β' -Untereinheit (blau). Die ω -Untereinheit (magenta) faltet die β' -Untereinheit und rekrutiert diese zum Zusammenbau zum Holoenzym. Die σ -Untereinheit ist mit den möglichen vier Domänen (1 - 4) gezeigt (braun), wobei Domäne 2 mit der -10-Region (gelbe Box), Domäne 3 mit dem erweiterten -10-Motiv und Domäne 4 mit der -35-Region (gelbe Box) des Promotors interagiert. Der Transkriptionsstartpunkt ist mit +1 markiert.

b. Die Transkriptionsinitiation wird mit der Erkennung und Bindung der Promotorregion durch das RNA-Polymerase-Holoenzym eingeleitet. Der gebundene DNA-Abschnitt ist durch grüne, gepunktete Linien dargestellt. Darauf folgend wird der DNA-Doppelstrang in der Nähe des Transkriptionsstarts geöffnet und die RNA-Polymerase beginnt mit der Verknüpfung der NTPs (rot, gepunktete Linie) anhand des nicht-codogenen Strangs der DNA. Anschließend dissoziiert die σ -Untereinheit und das RNA-Polymerase-Kernenzym setzt mit der Elongation des RNA-Strangs fort.

2.2 Transkriptionsinitiation, -elongation und deren Regulation

Für die Transkriptionsinitiation sind ein zusammengelagertes Holoenzym und die direkte Interaktion mit der DNA innerhalb der Promotorregion notwendig. Sobald das RNA-Polymerase-Holoenzym an die Promotorregion der DNA bindet, wird ein sogenannter offener Komplex gebildet, bei dem die DNA um den Transkriptionsstart herum für einige Basen entspiralisiert (etwa von -10 bis +2) und geöffnet wird [Browning & Busby, 2004]. Diese Strangtrennung macht den Matrizen-DNA-Strang (nicht-codogener Strang) für das aktive Zentrum der RNA-Polymerase zugänglich, sodass anschließend die RNA-Synthese starten kann (Abbildung II-2 b) [Vassylyev *et al.*, 2002; Browning & Busby, 2004].

Ein entscheidender Punkt in der Transkriptionsregulation ist die begrenzte Menge der RNA-Polymerase innerhalb der Zelle. Viele aktive RNA-Polymerasen werden für die Transkription der rRNA- und tRNA-

Gene gebraucht, die oft in vielfachen Kopien vorkommen, sodass die Anzahl an RNA-Polymerasen für die Transkription von Tausenden bakteriellen Genen limitiert ist [Ishihama, 2000] und Promotoren um die RNA-Polymerasen konkurrieren [Browning & Busby, 2004]. Um die Transkription der benötigten Gene zu gewährleisten, wurden Mechanismen entwickelt, die eine sinnvolle Verteilung der RNA-Polymerase erlauben [Browning & Busby, 2004]. So ergeben unterschiedliche Promotorsequenzen eine unterschiedlich gewichtete Verteilung der RNA-Polymerase auf die Transkriptionseinheiten, wobei Promotorsequenzen, die der Konsensussequenz sehr ähneln, eine effiziente Transkription erlauben [Busby & Ebricht, 1994; Browning & Busby, 2004]. Neben den Promotoren erlauben auch die σ -Faktoren eine Modulation der Transkription. Da Bakterien meistens neben einem Haushaltssigmafaktor weitere alternative σ -Faktoren besitzen, die jeweils in Antwort auf unterschiedliche Umgebungsbedingungen akkumulieren, somit den Haushaltssigmafaktor verdrängen und die RNA-Polymerase zu spezifischen Genen mit den zugehörigen Promotoren führen [Maeda *et al.*, 2000]. Zusätzlich werden σ -Faktoren oft durch sogenannte Anti- σ -Faktoren kontrolliert, die sich durch physiologische Signale anhäufen, an die σ -Faktoren binden und die Zusammenlagerung mit der RNA-Polymerase verhindern [Brown & Hughes, 1995]. Des Weiteren kodieren bakterielle Genome für Transkriptionsregulatoren, die die Transkription positiv oder negativ beeinflussen können [Babu & Teichmann, 2003]. Diese haben meistens spezifische Erkennungs- und Bindesequenzen innerhalb von Promotorregionen und zielen somit auf bestimmte Gene ab. Die Aktivität der Transkriptionsregulatoren wird durch umgebungsbedingte Signale, wie die steigende oder fallende Konzentration von Liganden oder durch bestimmte Modifizierungen, bzw. durch die eigene Konzentration bestimmt [Babu & Teichmann, 2003; Browning & Busby, 2004]. Grundsätzlich wird zwischen Transkriptionsaktivatoren und Transkriptionsrepressoren unterschieden, die die Bindung der RNA-Polymerase an die Promotorregion verstärken oder verhindern [Browning & Busby, 2004]. Außerdem geht man davon aus, dass die Faltung der gepackten, spiralisierten DNA (*supercoiled DNA*) einen Effekt auf die Verteilung der DNA-Polymerase auf die Promotoren haben muss [Browning & Busby, 2004]. Ein Beispiel hierfür ist das Protein H-NS aus *E. coli*, welches an der Kondensation der DNA beteiligt ist und die Transkription durch Bildung komplexer Nukleoprotein-Strukturen verhindern kann [Petersen *et al.*, 2002; Oshima *et al.*, 2006]. Diese vielfältigen Regulationsmechanismen erlauben die Transkription bestimmter Gene zu einem bestimmten Zeitpunkt und in Antwort auf Stressbedingungen und Nährstoffverfügbarkeit, und sind somit für die Dynamik des Transkriptomts zuständig.

Die Transkriptionselongation wird eingeleitet, wenn die RNA-Polymerase eine Konformationsänderung erfährt. Dabei dissoziiert die σ -Untereinheit vom Holoenzym, woraufhin das Kernenzym vom Promotor bewegt wird und durch die V-förmige Struktur des Kernenzyms fest an den Matrizen-DNA-Strang bindet [Uptain *et al.*, 1997]. Der Prozess der Elongation ist effizient und läuft mit hoher Rate von > 20 nt pro sek. ab [Epshtein & Nudler, 2003]. Die Elongationsrate ist von der Dichte der Holoenzyme am Promotor abhängig, welche vermutlich durch gegenseitiges Voranschieben der RNA-Polymerasen (mit der

Richtung der Elongation) verstärkt wird [Epshtein & Nudler, 2003]. Da das Holoenzym der RNA-Polymerase nur in Promotorregionen initiieren kann, ist für eine vollständige Transkription wichtig, dass die RNA-Polymerase einen stabilen Komplex bildet und extrem resistent gegen Dissoziation und Denaturierung ist [Uptain *et al.*, 1997]. Somit wird die Synthese auch von sehr langen Transkripten ermöglicht [Uptain *et al.*, 1997]. Weiterhin wird die RNA-Polymerase während der Elongationsphase von zusätzlichen Transkriptionsfaktoren (wie z.B. Nus- und Gre-Faktoren) begleitet, die es der RNA-Polymerase erlauben die Transkription effizient fortzusetzen oder ggf. zu korrigieren, falls diese in der Transkription geblockt wird, nicht-komplementäre Nukleotide verknüpft oder falls die Transkription verlangsamt werden soll, um die DNA zu entspiralisieren oder die Transkriptionstermination einzuleiten [Uptain *et al.*, 1997; Gusarov & Nudler, 2001; Borukhov *et al.*, 2005; Lewis *et al.*, 2008].

2.3 Transkriptionstermination

Die Transkriptionstermination findet wie die Transkriptionsinitiation an bestimmten Stellen im Genom statt. Die Termination in Bakterien wird entweder durch die RNA-Helikase ρ (ρ -abhängige Termination) oder durch intrinsische Terminatoren ohne zusätzliche Faktoren (ρ -unabhängige Termination) eingeleitet [Wilson & von Hippel, 1995]. Das ρ -Protein besteht in seiner aktiven Form aus sechs identischen Untereinheiten [Ciampi, 2006], welches stromaufwärts der transkribierenden RNA-Polymerase an die synthetisierte RNA bindet. Die Bindestelle des ρ -Faktors ist etwa 70 – 80 nt lang, weist einen hohen C- sowie niedrigen G-Gehalt auf und muss frei von Ribosomen sein [Richardson, 2002]. Nach der Bindung des ρ -Faktors an die RNA bewegt sich dieses auf dem synthetisierten RNA-Strang auf die RNA-Polymerase zu [Ciampi, 2006] und bei Erreichung des Transkriptionskomplexes wird die RNA innerhalb des geöffneten Komplexes von der DNA getrennt und entlassen, was vermutlich u.a. über komplexe Strukturausbildungen der RNA um das ρ -Protein erreicht wird [Richardson, 2002; Peters *et al.*, 2011]. Außerdem ist ρ in der Lage, mit den Nus-Elongationsfaktoren zu interagieren, die u.a. die Transkription verlangsamen, um die Termination durch den ρ -Faktor zu erleichtern [Richardson, 2002; Lewis *et al.*, 2008]. Bei der ρ -abhängigen Termination wird die Transkription also durch Verdrängen der RNA-Polymerase durch das ρ -Protein beendet. Auch im Genom von *C. glutamicum* ist das ρ -Protein kodiert (*cg1354/rho*), zu welchem jedoch keine Untersuchungen vorliegen.

Die intrinsische, ρ -unabhängige Termination ist durch bestimmte Sequenzabschnitte auf der DNA gegeben. Diese sind durch einen G+C-reichen Abschnitt gefolgt von einer Poly(T)-Sequenz auf dem Leitstrang der DNA gekennzeichnet [Gusarov & Nudler, 1999; Peters *et al.*, 2011]. Sobald die RNA-Polymerase diesen Bereich transkribiert, bildet das Sequenzmotiv den sogenannten Terminator, eine Haarnadelstruktur mit einem G+C-reichen Stamm und einer Poly(U)-Sequenz, auf der synthetisierten RNA aus [Gusarov & Nudler, 1999; Peters *et al.*, 2011]. Die Ausbildung dieser Haarnadelstruktur hat zur Folge, dass Scherkräfte auf die RNA-Polymerase wirken, die die Dissoziation des

Transkriptionskomplexes und die Lösung des RNA-Stranges mit sich bringt [Carafa *et al.*, 1990; Wilson & von Hippel, 1995; Peters *et al.*, 2011]. Außerdem lässt sich vermuten, dass das RNA-DNA-Hybrid Poly(U)-Poly(A) zusätzlich destabilisierend wirkt und den Zerfall des RNA-Polymerase-Komplexes fördert [Carafa *et al.*, 1990; Peters *et al.*, 2011]. Da die intrinsischen Terminatoren durch ausgeprägte Sequenzabschnitte und Sekundärstrukturen gekennzeichnet sind, lassen sich diese durch bioinformatische Tools (wie z.B. *TransTermHP* [Kingsford *et al.*, 2007]) vorhersagen.

3 Methoden der Transkriptomanalyse

Die Transkriptomanalyse (*Transcriptomics*) ist ein Fachgebiet, welches sich mit der Analyse der Gesamtheit aller Transkripte einer Zelle oder eines Organismus zu einem bestimmten Zeitpunkt oder unter bestimmten Bedingungen befasst. Klassische *Transcriptomics*-Methoden, wie *Northern Blot*, Microarrays und quantitative Reverse Transkriptase-PCR (qRT-PCR), erlauben eine Untersuchung von Expressionsmustern von bekannten Genen oder Genabschnitten und den Vergleich mehrerer Zustände. Die Entwicklung der Transkriptom-Hochdurchsatz-Sequenzierung (RNAseq) hat dieses Fachgebiet revolutioniert und ermöglicht nun eine genomweite Transkriptomanalyse. In den folgenden Abschnitten soll somit die Methode der RNAseq, die anwendbaren Sequenzierungsplattformen und die genomweite Transkriptomanalyse näher beschrieben werden.

3.1 Prinzipien der Transkriptom-Hochdurchsatz-Sequenzierung

Das Ziel des *Transcriptomic*-Fachgebiets ist die Analyse und Katalogisierung aller Transkript-Spezies, aber auch gleichzeitig die Bestimmung des Aufbaus dieser RNAs wie die Transkript-Starts und –Stopps und die Quantifizierung der Transkript-Spezies zu einem bestimmten Zeitpunkt. Verschiedenste Technologien wurden in der Vergangenheit entwickelt, um das Transkriptom zu kartieren und zu quantifizieren, wobei es sich grundsätzlich um hybridisierungsbasierte oder sequenzierungsbasierte Ansätze handelt. Die hybridisierungsbasierten Methoden, wie *Northern Blots* oder Microarrays, die zur Transkriptlevel-Bestimmung genutzt werden, beinhalten in der Regel die Hybridisierung von fluoreszenzmarkierter, zu untersuchender cDNA mit speziell angefertigten DNA-Proben, die das gewünschte Genom bzw. den Genomabschnitt abdecken. Auch die qRT-PCR, eine nicht hybridisierungsbasierte Methode, kann zur relativen Transkriptomquantifizierung verwendet werden. Diese genannten Methoden setzen die Kenntnis der Genomsequenz voraus, besitzen aufgrund von ungewünschten Kreuz-Hybridisierungen zum Teil ein Hintergrundrauschen [Okoniewski & Miller, 2006], sind durch die Sättigung des Signals in der Detektion limitiert und benötigen für die Quantifizierung als auch für den Transkriptlevel-Vergleich komplizierte Normalisierungen. Mit den klassischen sequenzierungsbasierten Methoden, wie *PEX (Primer Extension)* 5'- oder 3'-RACE (*Rapid Amplification of cDNA Ends*), die überwiegend zur Transkriptomkartierung benutzt werden, lassen sich zwar unbekannte cDNA-Sequenzen bestimmen, diese Methoden sind jedoch nicht quantitativ, relativ teuer, liefern nur eine geringe Sequenzausbeute und ermöglichen nur eine bis wenige Kartierungen pro Experiment.

Mit der Entwicklung von Hochdurchsatz-Sequenzierungsmethoden wie der Plattformen von *454 Life Sciences (GS-FLX-Titanium, GS Junior)*, *Illumina, Inc. (GA Iix, MiSeq, HiSeq)*, *Applied Biosystems (SOLiD)* und *Life Technologies (Ion Torrent)*, haben diese Methoden die Anwendung nicht nur zur

Genomsequenzierung, sondern auch Transkriptomsequenzierung gefunden. Insbesondere für Eukaryoten wurde die genomweite Transkriptomsequenzierung im letzten Jahrzehnt extensiv durchgeführt, da dadurch insbesondere die Intron-Exon-Struktur und das alternative Spleißen aufgeklärt werden kann [Cloonan *et al.*, 2008; Lister *et al.*, 2008; Nagalakshmi *et al.*, 2008; Yang *et al.*, 2012]. Für Prokaryoten wurde die Transkriptom-Sequenzierung lange Zeit nicht beachtet und ist erst kürzlich interessant geworden. Möglicherweise stufte man das prokaryotische Transkriptom im Vergleich zu dem eukaryotischen Transkriptom als simpel ein, da prokaryotische Transkripte keine Introns besitzen und kein alternatives Spleißen stattfindet [Sorek & Cossart, 2010]. Ein weiterer Grund mag die anspruchsvolle mRNA-Anreicherung in Prokaryoten sein, da hier die mRNA im Gegensatz zu eukaryotischer mRNA am 3'-Ende nicht polyadenyliert vorliegt und die mRNA-Aufreinigung erschwert [Sorek & Cossart, 2010]. Da sich mehr als 95 % der Gesamt-RNA einer Zelle aus rRNA zusammensetzen kann [Peano *et al.*, 2013], stellt die Abreicherung der rRNA einen essentiellen Schritt bei der Vorbereitung der RNA zur Transkriptomsequenzierung dar. Eine Transkriptomsequenzierung ohne vorherige Abreicherung der rRNA führt zu einem hohen Anteil an redundanten Sequenzen und nur einem sehr geringen Anteil an relevanten mRNA- und sRNA-Sequenzen.

Die Transkriptom-Sequenzierung mittels Hochdurchsatz-Sequenzierungsmethoden (RNAseq) ermöglicht eine umfassende Kartierung sowie Quantifizierung ganzer Transkriptome und Transkriptom-Elemente, besitzt deutliche Vorteile gegenüber klassischen Transkriptom-Methoden und hat somit das Fachgebiet *Transcriptomics* revolutioniert. Mit den erhaltenen Daten lassen sich Transkripte unabhängig ihrer Genomsequenz detektieren und aufgrund der Einzelnukleotid-Auflösung in Bezug auf ihren Anfang und das Ende sehr exakt bestimmen. Das Hintergrundrauschen wird minimiert, da Kreuzhybridisierungen nicht auftreten und die erhaltenen *Reads* eindeutig der Referenzsequenz zugeordnet werden können. Weiterhin besitzt die RNAseq einen nahezu unbeschränkten Detektionsbereich, der von den sequenzierten Fragmenten bestimmt wird. Somit wird durch die Methode RNAseq eine umfassende Analyse aller auftretenden Transkripte zu einem gegebenen Zeitpunkt genomweit und im Hochdurchsatz erlaubt [Wang *et al.*, 2009; Sorek & Cossart, 2010].

Allgemein umfasst die Vorbereitung der RNA zur Sequenzierung die Isolierung der Gesamt-RNA, das Umschreiben der RNA in cDNA und die anschließende Sequenzierung der cDNA-Bibliothek. Einen wichtigen Schritt nach der RNA-Isolierung stellt die Abreicherung der rRNA dar. Wie bereits erwähnt, macht die rRNA- und tRNA-Fraktion die Mehrheit aller Transkripte einer Zelle aus und um eine ausreichende Abdeckung an mRNA-Sequenzen zu erhalten und die redundanten rRNA-Sequenzen zu minimieren, sollte die rRNA vor der weiteren Bearbeitung abgereichert werden. Verschiedenste Methoden zur Abreicherung von rRNA bzw. zur Anreicherung von mRNA wurden für Prokaryoten und sind derzeit kommerziell bei verschiedensten Unternehmen erhältlich. Grundsätzlich lassen sich drei verschiedene Ansätze unterscheiden (Abbildung II-3; nach [Sorek & Cossart, 2010]):

(i) Die Abreicherung der rRNA durch Hybridisierung und darauffolgende selektive Entfernung, bei der die rRNA an spezifische Fragmente hybridisiert wird, die komplementär zu stark konservierten Bereichen der rRNAs sind. Die komplementären Fragmente sind typischerweise über Biotinylierung an magnetische Beads gekoppelt, die nach der Hybridisierung der rRNA entfernt werden. Diese Abreicherungsart wird derzeit u.a. von *Life Technologies (MICROBExpress Kit)*, *Epicentre Illumina (Ribo-Zero- rRNA Removal Kit)* und *Life Technologies (RiboMinus rRNA Depletion Kit)* angeboten.

(ii) Die Abreicherung der rRNA durch Abbau prozessierter Transkripte, wobei die Gesamt-RNA-Lösung mit einer 5'-Phosphat-abhängigen Exonuklease behandelt wird. Dieses Enzym erkennt und verdaut spezifisch prozessierte Transkripte mit 5'-Monophosphatenden, wie die von rRNAs und tRNAs. Primäre, unprozessierte Transkripte mit 5'-Triphosphatenden werden dabei von der Exonuklease nicht angegriffen und verbleiben in der Lösung. Dieses Enzym wird derzeit bei *Epicentre Illumina (Terminator 5'-Phosphat-abhängige Exonuklease)* erhältlich.

(iii) Die selektive Polyadenylierung von mRNAs, bei der die *E. coli*-Poly(A)-Polymerase I verwendet wird, die spezifisch die mRNAs am 3'-Ende polyadenyliert [Wendisch *et al.*, 2001]. Anschließend kann die mRNA über Oligo(dT)-Fragmente aufgereinigt oder über Oligo(dT)-Primer revers transkribiert werden. Bisher wird dieser Ansatz nur in Verbindung mit anderen Abreicherungsverfahren angeboten, z.B. von *Life Technologies (MessageAmp II-Bacteria Kit)* und von *Epicentre Illumina (mRNA-ONLY-Prokaryotic mRNA Isolation Kit)*.

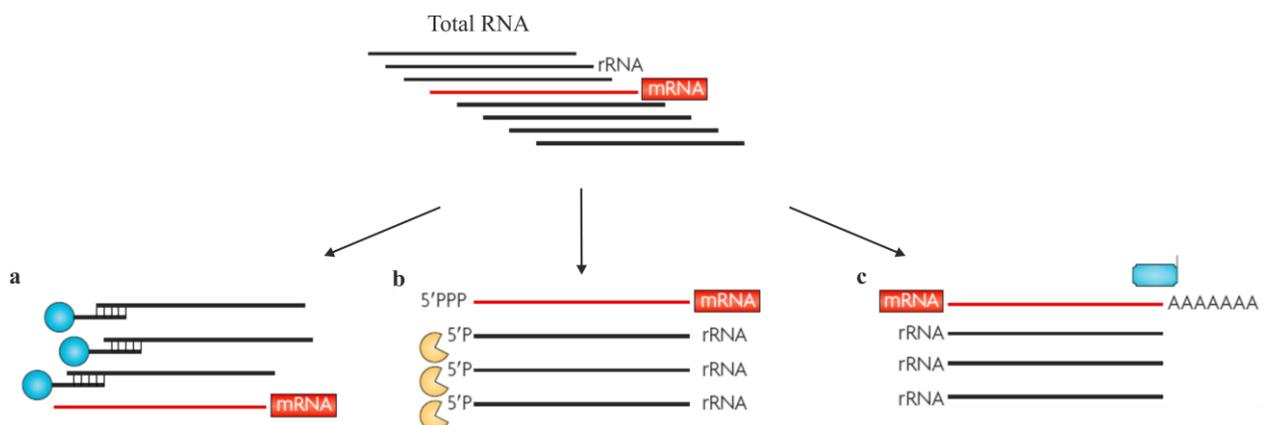


Abbildung II-3: Schematische Darstellung der Möglichkeiten zur rRNA-Abreicherung aus einer Gesamt-RNA-Lösung (modifiziert nach [Sorek & Cossart, 2010]).

a. Abreicherung der rRNA durch Hybridisierung und selektive Entfernung: Bei dieser Methode wird die rRNA an spezifische Fragmente hybridisiert, die komplementär zu stark konservierten Bereichen der rRNAs sind. Die Hybride können dann über magnetische Beads und mit einem Magnet-Separator selektiv entfernt werden.

b. Abreicherung der rRNA durch Abbau prozessierter Transkripte: Bei dieser Methode wird eine 5'-Phosphat-abhängige Exonuklease verwendet, die prozessierte, nicht native 5'-Transkriptenden mit 5'-Monophosphatenden erkennt und verdaut. Native mRNAs mit 5'-Triphosphatenden bleiben dabei intakt.

c. Selektive Polyadenylierung von mRNAs: Bei dieser Methode wird die *E. coli*-Poly(A)-Polymerase I verwendet, die selektiv nur mRNAs in einer Gesamt-RNA-Lösung polyadenyliert. Darauffolgend können die polyadenylierten mRNAs an Oligo(dT)-Fragmente hybridisiert und aufgereinigt oder direkt mit einem Oligo(dT)-Primer revers transkribiert werden.

Ein weiterer wichtiger Punkt bei der Vorbereitung der RNA für die Sequenzierung ist die Erhaltung der Strangspezifität. Die Transkripte einer Zelle können dem Vorwärts- oder dem Rückwärtsstrang der DNA entstammen. Je nach dem gewählten RNAseq-Vorbereitungsprotokoll kann die Orientierung der Transkripte verloren gehen oder beibehalten werden. So kann die Transkriptorientierung nicht nachvollzogen werden, wenn die für die Sequenzierung benötigten Adapter an die umgeschriebene cDNA ligiert werden. Dahingegen kann die Strangspezifität aufrechterhalten werden, indem die Adapter direkt an die RNA ligiert werden [Wang *et al.*, 2009; Sorek & Cossart, 2010]. Protokolle, die die Strangspezifität beibehalten, wurden zunächst für die Sequenzierung von microRNAs in Eukaryoten entwickelt [Cloonan *et al.*, 2008; Lister *et al.*, 2008] und später für prokaryotische Transkripte adaptiert [Sharma *et al.*, 2010; Vockenhuber *et al.*, 2011].

Für bakterielle RNAseq-Studien wurden bisher die Sequenzierungsplattformen von *Illumina*, *454 Life Sciences*, *Applied Biosystems* und *Life Technologies* verwendet [Passalacqua *et al.*, 2009; Perkins *et al.*, 2009; Sharma *et al.*, 2010; Frank *et al.*, 2011; Høvik *et al.*, 2012; Castro *et al.*, 2013]. Jede dieser Sequenzierungsplattformen variiert in ihren technischen Details, der Länge sowie der Anzahl der sequenzierten Fragmente (*Reads*; Tabelle II-3) und eignet sich für das Sequenzieren prokaryotischer Transkriptome. Die langen *Reads* von *454 Life Sciences* sind insbesondere bei einer *de novo*-Transkriptomsequenzierung, bei der keine Referenzsequenz vorliegt, hilfreich, wohingegen die kürzeren *Reads* der anderen Sequenzierungsplattformen eine enorme Sequenziertiefe bei vorliegender Referenz erzeugen können.

Tabelle II-3: Vergleich der Länge und Anzahl der *Reads* zwischen den Sequenzierungsplattformen

	max. Read-Länge	Gesamt-Reads
<i>454 Life Sciences</i>	bis 1.000 bp	> 0,2 Mio. pro Lauf [Sharma <i>et al.</i> , 2010]
<i>Illumina</i>	bis 2 × 300 bp	> 20 Mio. pro Lane [Pfeifer-Sancar <i>et al.</i> , 2013]
<i>Applied Biosystems</i>	bis 100 bp	> 50 Mio. pro Lauf [Passalacqua <i>et al.</i> , 2009]
<i>Life Technologies</i>	bis 400 bp	> 60 Mio. pro Lauf (Produkthomepage)

Das Ergebnis einer typischen RNA-Sequenzierung wird von mehreren Millionen *Reads* bestimmt, die entweder auf einer Referenzsequenz kartiert oder bei fehlender Referenz zu Transkriptsequenzen angeordnet (assembliert) werden. Diese Fülle an erhaltenen Daten stellt insbesondere für die Bioinformatik eine Herausforderung dar. Zum einen erfordert die Transkriptom- aber auch die Genom-Hochdurchsatz-Sequenzierung effiziente Möglichkeiten der Speicherung, Abrufung und Prozessierung von großen Datenmengen, wobei hierfür gute Lösungen von den Sequenzierungsplattformen geliefert werden. Ein weiterer Punkt ist die möglichst schnelle Kartierung der relativ kurzen und vielen *Reads* auf das Referenzgenom bzw. das Assembling dieser *Reads* zu Transkriptsequenzen. Hierfür liegt der Ansatz in der Komprimierung und der parallelen Bearbeitung der erhaltenen Daten (z.B. *SARUMAN* [Blom *et al.*, 2011], *Bowtie* [Langmead *et al.*, 2009]). Bei der Analyse eukaryotischer Transkriptome müssen die bioinformatischen Algorithmen zudem mit Exon-Intron- bzw. Intron-Exon-Übergängen umgehen und auch alternatives Spleißen erkennen können. Außerdem stellen repetitive Genomregionen ein zusätzliches Problem dar, da dadurch ein *Read* an mehreren Stellen im Genom kartiert werden kann. Hierbei gilt es die *Reads* möglichst gleichmäßig auf die repetitiven Region zu verteilen oder die Zuweisung unter Berücksichtigung der Verteilung der benachbarten, nicht-repetitiven *Reads* [Hashimoto *et al.*, 2009].

Dank der voranschreitenden Entwicklung ist es heute möglich die Fülle an erhaltenen Sequenzierdaten zu handhaben und für die Transkriptomanalyse einzusetzen. Die Möglichkeiten und der informationsreiche Gewinn der Transkriptomanalyse mithilfe von RNAseq-Daten soll im nächsten Kapitel näher beschrieben werden.

3.2 Genomweite Transkriptomanalyse mittels RNAseq

Die genomweite Transkriptomsequenzierung und die anschließende Analyse der erhaltenen Daten hat in den letzten Jahren Anwendung für verschiedene Prokaryoten und Eukaryoten gefunden [Lister *et al.*, 2008; Nagalakshmi *et al.*, 2008; Perkins *et al.*, 2009; Sharma *et al.*, 2010; Chaudhuri *et al.*, 2011; Høvik *et al.*, 2012; Yang *et al.*, 2012] und offenbarte insbesondere für prokaryotische Transkriptome eine enorme Komplexität [Sorek & Cossart, 2010].

Die meisten prokaryotischen Genome sind über *in silico* Vorhersagen annotiert. Diese Vorhersage-Softwares sind jedoch fehleranfällig in der Bestimmung von Gen-Starts, in der Vorhersage kleiner Proteine limitiert und sind auf die Findung von Kodierregionen spezialisiert, sodass nicht-kodierende RNAs, TSSs und folglich 5'-UTRs unerkant bleiben [McHardy *et al.*, 2004; Overbeek *et al.*, 2007]. Mittels der RNAseq-Analyse lassen sich nun Gen-Strukturen aufdecken, indem Transkriptstarts und -stopps bestimmt und auch kleine neue Transkripte detektiert werden können. Somit bietet die Transkriptom-Sequenzierung und die darauffolgende Analyse der erhaltenen Daten eine Möglichkeit zur Verbesserung bzw. Erweiterung der Genomannotation für Prokaryoten, aber auch für Eukaryoten, insbesondere durch die Erkennung von Intron-Exon-Strukturen und des alternativen Spleißens [Denoeud

et al., 2008; Wang *et al.*, 2009; Frank *et al.*, 2011]. In Prokaryoten können 5'-UTRs länger als hundert Basen lang sein, was einen Hinweis auf regulatorische, transkriptionelle und translationale Funktionen, wie *Riboswitches*, RNA-Thermometer oder Bindestellen für andere regulatorische RNAs, liefert [Güell *et al.*, 2009; Perkins *et al.*, 2009; Sorek & Cossart, 2010]. Solche regulatorischen Regionen besitzen oft ausgeprägte Sekundärstrukturen, durch die die zugehörigen Gene reguliert werden [Barrick *et al.*, 2004; Mandal & Breaker, 2004]. Aufgrund dieser sekundären Strukturen lassen sich regulatorische Elemente zwar rechnerisch vorhersagen (z.B. *Rfam* [Gardner *et al.*, 2011]), eine eindeutige Position und die Validierung der regulatorischen 5'-UTRs ist jedoch durch die Bestimmung von TSSs und die anschließende Funktionsanalyse gegeben. Durch die RNAseq lassen sich genomweit mehrere Hundert bis Tausend TSSs mit nur einer Sequenzierung identifizieren [Sharma *et al.*, 2010; Pfeifer-Sancar *et al.*, 2013] und somit 5'-UTRs kartieren. Die identifizierten TSSs sind außerdem in der Bestimmung von Promotormotiven, wie z.B. von *housekeeping* Sigmafaktoren, hilfreich [McGrath *et al.*, 2007; Sharma *et al.*, 2010; Pfeifer-Sancar *et al.*, 2013], die sich etwa 10 bp und 35 bp stromaufwärts der TSSs befinden [Pribnow, 1975, 1975; Takanami *et al.*, 1976; Seeburg *et al.*, 1977]. Neben der Identifizierung von Promotormotiven, lassen sich innerhalb der gefundenen 5'-UTRs Bindestellen für Ribosomen suchen, die wichtige Informationen über die Translation liefern können. Demgegenüber haben RNAseq-Analysen u.a. für *Helicobacter pylori*, *Salmonella typhi* und *Pseudomonas putida* gezeigt, dass einige Transkripte ohne 5'-UTRs, d.h. als sogenannte *leaderless* Transkripte, transkribiert werden und einen anderen Translationsmechanismus ohne die Erkennung und Bindung einer RBS innehaben [Perkins *et al.*, 2009; Sharma *et al.*, 2010; Frank *et al.*, 2011]. Des Weiteren kann die RNAseq einen Aufschluss über die Anordnung der Gene zu Operons geben. Das klassische Operon besitzt mehrere Gene, die gemeinsam ausgehend von einem Promotor transkribiert werden [Jacob & Monod, 1961a; Shine & Dalgarno, 1975]. Typischerweise besitzen Operons funktionsverwandte Gene, deren Produkte z.B. zu einem Stoffwechselweg zugeordnet werden [Lawrence & Roth, 1996; Osbourn & Field, 2009]. Die Aufklärung der Operon-Strukturen kann sich somit als hilfreich für die Vorhersagen für Genfunktionen erweisen. Kürzlich durchgeführte Transkriptomanalysen konnten außerdem zeigen, dass zahlreiche Operons aufgrund von internen TSSs in kleinere Sub-Operons unterteilt vorliegen können, die in Antwort auf unterschiedliche Bedingungen auftreten [Güell *et al.*, 2009; Sharma *et al.*, 2010]. Somit können die recht komplexen Operon-Strukturen mittels RNAseq-Analysen entschlüsselt werden. Da mittlerweile bekannt ist, dass neben den Eukaryoten auch Prokaryoten zahlreiche kleine, regulatorische, kodierende und nicht-kodierende RNAs (sRNAs) besitzen [Fröhlich & Vogel, 2009], lassen sich diese durch Transkriptomsequenzierung genomweit erfassen und näher charakterisieren [Vockenhuber *et al.*, 2011; Mentz *et al.*, 2013]. Auch das Vorkommen von massenhafter *antisense* Transkription in Prokaryoten ist durch RNAseq-Analysen aufgedeckt worden, die vermutlich eine Regulation auf deren *sense* Transkripte ausüben [Dornenburg *et al.*, 2010; Thomason & Storz, 2010; Lasa *et al.*, 2011].

Es hat sich gezeigt, dass die genomweite RNAseq-Analyse einen wichtigen Beitrag zum Verständnis von RNA-basierten Mechanismen auf Transkriptions- und Translationsebene liefert [Sorek & Cossart, 2010] und die Dynamik von prokaryotischen Transkriptomen aufzeigt. Weiterhin kann diese einen wertvollen Beitrag zur Genom-Annotation beitragen und bereits vorhandene Annotationen verbessern. Nichtsdestotrotz stößt die Transkriptomanalyse mittels RNAseq an ihre Grenzen: Derzeit werden für diese Methode als Ausgangsmaterial mehrere Millionen Zellen benötigt. Dieses stellt insbesondere ein Problem für nicht-kultivierbare Bakterien dar. Weiterhin wird die Interpretation der erhaltenen Daten erschwert, da es unklar bleibt, ob der erhaltene Zustand nur einem Teil oder der gesamten Population entstammt. Somit ist die Bestimmung von Einzelzell-Transkriptomen zukunftsweisend, die dann eine exakte Transkriptomanalyse einzelner und auch nicht-kultivierbarer Zellen ermöglicht. Für Eukaryoten sind bereits erste Einzelzell-Transkriptomanalysen durchgeführt worden [Tang *et al.*, 2009; Deng *et al.*, 2014], für die Anwendung in Prokaryoten bedarf es jedoch noch weiterer Entwicklung.

4 Zielsetzung der Arbeit

Corynebacterium glutamicum ist ein natürlich vorkommendes Bodenbakterium, welches zur industriellen, biotechnologischen Produktion von Aminosäuren genutzt wird. Da dieses Bakterium unter natürlichen Bedingungen nur die von ihm selbst benötigten Mengen an Aminosäuren herstellt, muss für einen ausreichenden Produktionsbedarf der Aminosäure-Metabolismus z.B. durch genetische Modifizierungen verbessert werden. Ferner stellt *C. glutamicum* einen Modellorganismus für nahe verwandte pathogene Bakterien von medizinischem Interesse dar, wie z.B. *Corynebacterium diphtheriae* und *Mycobacterium tuberculosis*. Um die ständige Optimierung der Produktionsstämme und eine umfassende Erforschung dieses Organismus gewährleisten zu können, ist die vollständige Kenntnis der relevanten Stoffwechselwege sowie zugehöriger Regulationsmechanismen von besonderer Wichtigkeit. Bakterielle Regulationsmechanismen können auf der Ebene des Genoms, Metaboloms, Proteoms und auch des Transkriptomts fungieren. Die vollständige Genomsequenz inklusive der Annotation von *C. glutamicum* liegt nun seit einem Jahrzehnt vor, das Metabolom als auch das Proteom sind weitestgehend untersucht. Analysen zur Regulation durch Transkriptionsregulatoren und die zusammenhängende Expression proteinkodierender Gene sowie die Promotor-Charakterisierung sind ebenfalls durchgeführt worden, allerdings fehlt bis heute eine umfassende Untersuchung zur transkriptionellen Organisation von *C. glutamicum*.

Die Kernaufgabe dieser Dissertationsarbeit bestand somit darin, eine möglichst vollständige Transkriptom-Karte für *C. glutamicum* zu erstellen. Infolgedessen war die Etablierung der RNAseq für das Transkriptom *C. glutamicum* das Ziel dieser Arbeit. Für die Etablierung der RNAseq-Methode standen die folgenden Aspekte im Vordergrund:

- 1) Entwicklung zweier RNAseq-Protokolle, um die Startpunkte und Gesamt-Transkripte als wichtige Elemente eines Transkriptomts separat sequenzieren und analysieren zu können.
- 2) Konstruktion von durchschaubarer, schrittweise zugänglicher und möglichst auch für andere Organismen anwendbarer RNAseq-Protokolle.
- 3) Sicherung der Strangspezifität während der Transkriptom-Aufbereitung bis zur Sequenzierung, da beide DNA-Stränge genetische Informationen enthalten.
- 4) Auswahl geeigneter Kultivierungsbedingungen, da die Transkription abhängig von den ausgesetzten Bedingungen abläuft.
- 5) Festlegung einer Auswertungsstrategie für erlangte RNAseq-Daten.
- 6) Anwendung der entwickelten RNAseq-Protokolle auf das Transkriptom von *C. glutamicum* und die darauffolgende Auswertung erhaltener RNAseq-Daten.

III Materialien und Methoden

1 Verwendete Materialien und Geräte

1.1 Bakterienstamm

In dieser Arbeit wurde der Stamm *Corynebacterium glutamicum* ATCC 13032 verwendet, der im Folgenden *C. glutamicum* genannt wird.

1.2 Nährmedien

Folgende Nährmedien wurden für die Kultivierung von *C. glutamicum* eingesetzt:

Definiertes CGXII-Minimalmedium (modifiziert nach [Keilhauer *et al.*, 1993]) oder LB-Komplexmedium (nach [Bertani, 1951]) als Flüssigmedium zur Anzucht im Erlenmeyerkolben; angepasstes CGXII-Minimalmedium (modifiziert nach [Keilhauer *et al.*, 1993]) zur Kultivierung im Fermenter-System; CASO-Komplexmedium zur Anzucht auf Agar-Platten.

CASO-Agar:

CASO-Bouillon	30 g·l ⁻¹
Agar	15 g·l ⁻¹

In deionisiertem H₂O lösen und autoklavieren. Nach dem Autoklavieren Zugabe von:

Nalidixin (in 0,1 N NaOH, steril filtriert)	50 mg·l ⁻¹
---	-----------------------

CGXII-Minimalmedium für Schüttelkolben:

MOPS-Natriumsalz	42 g·l ⁻¹
NH ₄ (SO ₄) ₂	20 g·l ⁻¹
K ₂ HPO ₄	1 g·l ⁻¹
KH ₂ PO ₄	1 g·l ⁻¹
MgSO ₄ ·7H ₂ O	250 mg·l ⁻¹

In 900 ml deionisiertem H₂O lösen, pH 7,4 mit NaOH (10 M) einstellen und autoklavieren.

Nach dem Autoklavieren Zugabe von:

CaCl ₂ ·2H ₂ O (autoklaviert)	10 mg·l ⁻¹
FeSO ₄ ·7H ₂ O (autoklaviert)	10 mg·l ⁻¹
MnSO ₄ ·H ₂ O (autoklaviert)	10 mg·l ⁻¹
ZnSO ₄ ·7H ₂ O (autoklaviert)	1 mg·l ⁻¹
CuSO ₄ (autoklaviert)	200 µg·l ⁻¹
NiCl ₂ ·6H ₂ O (autoklaviert)	20 µg·l ⁻¹
D-Glukose (autoklaviert)	20 g·l ⁻¹
Harnstoff (steril filtriert)	5 g·l ⁻¹
Protocatechuat (steril filtriert)	30 mg·l ⁻¹
Biotin (steril filtriert)	200 µg·l ⁻¹
Thiamin (steril filtriert)	420 µg·l ⁻¹

Mit deionisiertem H₂O auf 1 l auffüllen.

CGXII-Minimalmedium für das Fermenter-System:

NH ₄ (SO ₄) ₂	25 g·l ⁻¹
K ₂ HPO ₄	1 g·l ⁻¹
KH ₂ PO ₄	1 g·l ⁻¹
MgSO ₄ ·7H ₂ O	250 mg·l ⁻¹

In 750 ml deionisiertem H₂O lösen und autoklavieren.

Nach dem Autoklavieren Zugabe von:

CaCl ₂ ·2H ₂ O (autoklaviert)	10 mg·l ⁻¹
FeSO ₄ ·7H ₂ O (autoklaviert)	10 mg·l ⁻¹
MnSO ₄ ·H ₂ O (autoklaviert)	10 mg·l ⁻¹
ZnSO ₄ ·7H ₂ O (autoklaviert)	1 mg·l ⁻¹
CuSO ₄ (autoklaviert)	200 µg·l ⁻¹
NiCl ₂ ·6H ₂ O (autoklaviert)	20 µg·l ⁻¹
D-Glukose (autoklaviert)	40 g·l ⁻¹
Harnstoff (steril filtriert)	5 g·l ⁻¹
Protocatechuat (steril filtriert)	30 mg·l ⁻¹
Biotin (steril filtriert)	2 mg·l ⁻¹
Thiamin (steril filtriert)	420 µg·l ⁻¹

Mit deionisiertem H₂O auf 800 ml auffüllen.

LB-Komplexmedium:

LB Broth Base 20 g·l⁻¹

In deionisiertem H₂O lösen und autoklavieren.

1.3 Oligonukleotide**Tabelle III-1: Für die RNA-Sequenzierung verwendete Oligonukleotide**

Name	Sequenz (5'-3')
RNA-Adapter	CCCUACACGACGCUCUCCGAUCGAG
DNA-loop-Adapter	AGATCGGAAGAGAGACGTGTGCTCTTCCGATCTNNNNNNN
RNAseq_PCR_a	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACA-CGACGCTCTTCCGATCGAG
RNAseq_PCR_b	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAG-TTCAGACGTGTGCTCTTCCGATCT
RNAseq_Seq_a	ACTCTTCCCTACACGACGCTCTTCCGATCGAG
RNAseq_Seq_b	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Alle verwendeten Oligonukleotide wurden von Metabion International AG synthetisiert.

Tabelle III-2: Für den Nachweis von DNA verwendete Oligonukleotide

Name	Sequenz (5'-3')
DNA_180_a	TCGCAGGCAAGACCGTTATC
DNA-180_b	ATCGCCAACCTTCCGGTAG
DNA_570_a	CCACAGCAACTGGGTCTAAC
DNA_570_b	CGACAACCAGCGTTTCAG

Alle verwendeten Oligonukleotide wurden von *Metabion International AG* synthetisiert.

1.4 Labor-Verbrauchsmaterialien, Enzyme und Kits**Tabelle III-3: In dieser Arbeit verwendete Labor-Verbrauchsmaterialien mit Bezugsquelle**

Produkt	Bezugsquelle
Filter-Pipettenspitzen	<i>Biozym</i>
Glaswaren	<i>Schott</i>
Nitrilhandschuhe	<i>VWR</i>
Parafilm	<i>American National Can Company</i>
PCR-Stripes	<i>VWR</i>

Produkt	Bezugsquelle
Petrischalen	<i>Greiner bio-one</i>
Photometer-Küvetten	<i>Brand</i>
Pipettenspitzen	<i>Starlab</i>
Reaktionsgefäße (1,5 ml)	<i>Sarstedt</i>
Reaktionsgefäße (13 ml)	<i>Sarstedt</i>
Reaktionsgefäße (50 ml)	<i>Greiner bio-one</i>
<i>Ribotubes (Lysing Matrix B)</i>	<i>MP Biomedicals</i>
Spritzen	<i>Starlab</i>
Sterilfilter	<i>GE Healthcare</i>

Tabelle III-4: In dieser Arbeit verwendete Kits mit Bezugsquelle

Kit	Bezugsquelle
<i>Agencourt AMPure XP</i>	<i>Beckman Coulter</i>
<i>Agilent High Sensitivity DNA Kit</i>	<i>Agilent Technologies</i>
<i>Agilent RNA 6000 Nano Kit</i>	<i>Agilent Technologies</i>
<i>Agilent RNA 6000 Pico Kit</i>	<i>Agilent Technologies</i>
<i>DNase I Kit</i>	<i>Roche</i>
<i>Ribo-Zero rRNA Removal Kit (Gram-positive Bacteria)</i>	<i>Epicentre</i>
<i>MICROBExpress Bacterial mRNA Enrichment Kit</i>	<i>Life Technologies</i>
<i>RNase-free DNase Set</i>	<i>Qiagen</i>
<i>RNeasy Mini Kit</i>	<i>Qiagen</i>
<i>ThermoScript RT-PCR System</i>	<i>Life Technologies</i>
<i>TruSeq PE Cluster Kit</i>	<i>Illumina</i>
<i>TruSeq SBS Kit</i>	<i>Illumina</i>

Tabelle III-5: In dieser Arbeit verwendete Enzyme und DNA-Längenstandards mit Bezugsquelle

Enzym/DNA-Leiter	Bezugsquelle
100 bp DNA-Längenstandard (P-805)	<i>MBBL</i>
<i>BIOTAQ DNA-Polymerase</i>	<i>Bioline</i>
<i>Phusion High-Fidelity DNA-Polymerase</i>	<i>Thermo Scientific</i>
<i>RNA 5'-Polyphosphatase</i>	<i>Epicentre</i>
<i>RiboLock RNase Inhibitor</i>	<i>Thermo Scientific</i>
<i>T4 Polynukleotid-Kinase</i>	<i>Thermo Scientific</i>

Enzym/DNA-Leiter	Bezugsquelle
<i>T4 RNA Ligase I</i>	<i>New England Biolabs</i>
<i>Terminator 5'-Phosphat-abhängige Exonuklease</i>	<i>Epicentre</i>

1.5 Chemikalien

Alle hier nicht aufgeführten Chemikalien entstammen einem der aufgeführten Kits (Tabelle III-4).

Tabelle III-6: In dieser Arbeit verwendete Chemikalien mit Bezugsquelle

Chemikalie	Bezugsquelle
Agar	<i>Life Technologies</i>
Agarose	<i>Peqlab</i>
Ammoniumsulfat ($\text{NH}_4(\text{SO}_4)_2$)	<i>Roth</i>
Bromphenolblau	<i>Serva</i>
Biotin	<i>Sigma</i>
CASO	<i>Roth</i>
Calciumchlorid ($\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$)	<i>Merck Millipore</i>
Chloroform	<i>Roth</i>
D-Glukose	<i>Merck Millipore</i>
DEPC-behandeltes Wasser	<i>Roth</i>
Diamid	<i>Roth</i>
Dikaliumhydrogenphosphat (K_2HPO_4)	<i>Merck Millipore</i>
dNTPs	<i>Thermo Scientific</i>
EDTA	<i>Amersham</i>
Eisen(II)-Sulfat-Heptahydrat ($\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$)	<i>Merck Millipore</i>
Eisessig	<i>Roth</i>
Ethanol (EtOH)	<i>Roth</i>
Ethidiumbromid	<i>Sigma</i>
Glycerin	<i>Roth</i>
Glykogen, RNA-grade	<i>Thermo Scientific</i>
Harnstoff	<i>Roth</i>
Isopropanol	<i>Roth</i>
Kaliumacetat	<i>Merck Millipore</i>
Kaliumdihydrogenphosphat (KH_2PO_4)	<i>Merck Millipore</i>
Kupfersulfat (CuSO_4)	<i>Merck Millipore</i>

Chemikalie	Bezugsquelle
LB Broth Base	<i>Life Technologies</i>
Magnesiumacetat	<i>Merck Millipore</i>
Magnesiumsulfat-Heptahydrat ($\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$)	<i>Merck Millipore</i>
Mangan(II)-Sulfat-Monohydrat ($\text{MnSO}_4 \cdot \text{H}_2\text{O}$)	<i>Merck Millipore</i>
MOPS-Natriumsalz	<i>Roth</i>
Nalidixinsäure	<i>Roth</i>
Natriumacetat	<i>Merck Millipore</i>
Natriumchlorid	<i>Merck Millipore</i>
Natronlauge	<i>Merck Millipore</i>
Nickel(II)-Chlorid-Hexahydrat ($\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$)	<i>Merck Millipore</i>
Phenol-Chloroform-Isoamylalkohol (PCI)	<i>Roth</i>
PEG 8000	<i>Sigma</i>
Protocatechuat	<i>Sigma</i>
Salzsäure	<i>Merck Millipore</i>
Thiamin	<i>Sigma</i>
TRIS (10 mM)–EDTA (1 mM), pH 8 (RNase-frei)	<i>Fisher Scientific</i>
TRIS-HCl (1 M), pH 8 (RNase-frei)	<i>Fisher Scientific</i>
<i>TRIzol</i> -Reagenz	<i>Life Technologies</i>
Wasserstoffperoxid	<i>Merck Millipore</i>
Xylencyanol FF	<i>Sigma</i>
Zinksulfat-Heptahydrat ($\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$)	<i>Merck Millipore</i>
β -Mercaptoethanol	<i>Merck Millipore</i>

1.6 Geräte

Tabelle III-7: In dieser Arbeit verwendete Geräte und Apparaturen mit Bezugsquelle

Gerät/Apparatur	Bezugsquelle
<i>Agilent 2100 Bioanalyzer</i>	<i>Agilent Technologies</i>
Autoklav	<i>Systec</i>
Brutschränke	<i>Memmert</i>
<i>Cluster Station</i>	<i>Illumina</i>
Eismaschine	<i>Scotsman</i>
Elektrophorese-Kammern	<i>Eigenbau Universitätswerkstatt</i>
Feinwaagen	<i>Sartorius</i>

Gerät/Apparatur	Bezugsquelle
Fermenter-System (<i>BIOSTAT Q</i>)	<i>Sartorius</i>
Gel-Fotodrucker	<i>Mitsubishi</i>
Gefrierschränke (-80 °C)	<i>Sanyo</i>
Geldokumentationsanlage	<i>AlphaInnotech</i>
Homogenisator (<i>Precellys 24</i>)	<i>Peqlab</i>
Kühlschrank-Gefrier-Kombination	<i>Siemens</i>
Kühltischzentrifuge	<i>Eppendorf</i>
Luftschüttler (<i>Innova 4430</i>)	<i>New Brunswick Scientific</i>
Magnet-Separator für Reaktionsgefäße (1,5 ml)	<i>Life Technologies</i>
Magnetrührer mit Heizplatte	<i>IKA</i>
Mikrowelle	<i>Cinex</i>
Photospektrometer (<i>NanoDrop 1000</i>)	<i>Peqlab</i>
PCR-Cycler	<i>Eppendorf</i>
pH-Meter	<i>Schott</i>
Photometer	<i>Eppendorf</i>
Pipetten	<i>Eppendorf</i>
Reinraumwerkbank	<i>Heraeus HeraSafe</i>
Reinstwasseranlage	<i>TKA</i>
<i>Sequenzierer (Genome Analyzer IIx)</i>	<i>Illumina</i>
Spannungsgeber	<i>Bio-Rad</i>
Thermomixer für Reaktionsgefäße (1,5 ml)	<i>Eppendorf</i>
Tischzentrifuge	<i>Eppendorf</i>
Vortex	<i>Scientific Industries</i>
Wasserbäder	<i>GFL</i>

1.7 Software

Tabelle III-8: In dieser Arbeit verwendete Software/Datenbanken mit deren Verwendungszweck und Referenz

Software/Datenbank	Verwendungszweck	Bezugsquelle/Referenz
<i>Adobe Reader 7.1.0</i>	Verfassen dieser Arbeit	<i>Adobe</i>
<i>BLASTP</i>	Proteinsequenz-Vergleich	[Altschul et al., 1990]
<i>calculateExpression Values</i>	Bestimmung von TSSs, neuen Transkripten und Operon-Strukturen	Christian Rückert, CeBiTec, Universität Bielefeld
<i>CloneManager 7.11</i>	Primer-Design	<i>Sci-Ed Software</i>

Software/Datenbank	Verwendungszweck	Bezugsquelle/Referenz
<i>CorelDRAW X5</i>	Erstellung von Grafiken	<i>Corel</i>
<i>CoryneRegNet 6.0</i>	Annotations- und Analyse-Datenbank für <i>C. glutamicum</i>	[Pauling et al., 2012]
<i>eggNOG 3.0</i>	Einteilung der Gen-Funktionen	[Powell et al., 2012]
<i>Excel 2010</i>	Daten-Analyse	<i>Microsoft</i>
<i>GenDB</i>	Gen-Annotation für <i>C. glutamicum</i>	[Meyer et al., 2003]
<i>Improbizer</i>	Suche von Motiven	[Ao et al., 2004]
<i>ORF finder</i>	Suche von Kodierregionen	http://www.gen-info.osaka-u.ac.jp/~uhmin/study/orf/index.cgi
<i>ReadXplorer</i>	Visualisierung der Sequenzier-Daten	Hilker <i>et al.</i> , eingereicht
<i>Rfam 11.0</i>	Vorhersage der regulatorischen Regionen	[Gardner et al., 2011]
<i>RNAfold</i>	Erstellung von RNA-Faltungsstrukturen	[Gruber et al., 2008]
<i>SARUMAN</i>	Kartierung der Sequenzier- <i>Reads</i>	[Blom et al., 2011]
<i>TransTermHP</i>	Vorhersage der ρ -unabhängigen Terminatoren	[Kingsford et al., 2007]
<i>WebLogo 3</i>	Erstellung von Sequenz-Logos	[Crooks et al., 2004]
<i>Word 2010</i>	Verfassen dieser Arbeit	<i>Microsoft</i>

1.8 Puffer und Lösungen

Soweit nicht anders angegeben, werden alle Puffer und Lösungen in deionisiertem Wasser angesetzt.

Elektrophorese

TAE-Puffer (50fach):

TRIS	2	M
Natriumacetat	500	mM
EDTA	50	mM

Mit Eisessig auf pH 7,8 einstellen.

Bromphenolblau-Xylencyanol (BPB-XC)-Ladepuffer:

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

Bromphenolblau 0,1 % (w/v)

Xylencyanol FF 0,1 % (w/v)

Mit HCl titrieren bis sich die Lösung dunkelgrün färbt.

RNA-Sequenzierung

Fragmentierungspuffer:

Kaliumacetat 100 mM

Magnesiumacetat 30 mM

In RNase-freiem TRIS-HCl (200 mM, pH 8) lösen.

Fragmentierungsstopp-Puffer:

TRIS 10 mM

EDTA 1 mM

Als TRIS-EDTA-Lösung (pH 8) von Fisher Scientific bezogen.

Größenausschluss-Puffer (zur Präzipitation von RNA-Fragmenten > 150 nt):

PEG 8000 25,0 % (w/v)

Natriumacetat 2,5 M

In RNase-freiem Wasser lösen.

2 Mikrobiologische Methoden

2.1 Kultivierung von *C. glutamicum*

Die Anzucht von *C. glutamicum* auf Festmedium-Platten erfolgte durch Einzelkolonie-Ausstrich auf CASO-Agar mit 50 mg·l⁻¹ Nalidixin und anschließender Inkubation für 1 - 2 Tage bei 30 °C im Brutschrank.

Die Anzucht als Flüssigkultur wurde durch Inokulation mit einer Einzelkolonie von einer frisch kultivierten Festmedium-Platte in 25 ml entsprechendes Flüssigmedium in 250 ml-Schüttelkolben bei 300 rpm und 30 °C im Luftschüttler durchgeführt.

Für die Kultivierung von *C. glutamicum* im Fermenter-System wurde eine Vor-Kultur im CGXII-Minimalmedium im Schüttelkolben angezogen und der Fermenter mit dieser Vor-Kultur und einer Start-OD_{600 nm} von 0,2 inokuliert. Die Fermentation wurde bei pH 7, 30 °C und 30 % gelöstem Sauerstoff durchgeführt.

2.2 Lagerung von *C. glutamicum*

Bakterienzellen können durch Zugabe von Glycerin für mehrere Jahre bei -80 °C gelagert werden. Dazu wird die gewünschte Kultur von einer frischen Festmedium-Platte mit einer sterilen Glaspipette oder Pipettierspitze abgenommen, in 300 µl LB-Medium resuspendiert und mit 900 µl Glycerin (87 %) gut vermischt.

2.3 Bestimmung des Bakterientiters

Der Bakterientiter wird durch die photometrische Messung bei einer Wellenlänge von 600 nm ermittelt. Eine OD_{600 nm} von 1 entspricht bei *C. glutamicum* einer Anzahl von etwa 10⁹ Zellen pro ml. Aufgrund des begrenzten Messbereichs des Photometers wurde die Bakterienkultur für die Messung in einer geeigneten Konzentration mit Wasser verdünnt. Als Referenz diente eine entsprechende Verdünnung mit dem jeweiligen Nährmedium.

Mit dieser Methode wird der Gesamt-Titer (lebende und abgestorbene Zellen) bestimmt. Für eine Bestimmung des Lebendtiters wird von der Kultur eine Verdünnungsreihe angelegt, ausplattiert und die Kolonien nach der Inkubation ausgezählt.

2.4 Anzucht von *C. glutamicum* für die Transkriptom-Sequenzierung

Um die Gesamt-RNA für die Erstellung der Bibliothek der nativen 5'-Transkriptenden zu erhalten, wurde *C. glutamicum* als Flüssigkultur unter folgenden Bedingungen angezogen:

- CGXII-Minimalmedium, bis zur exponentiellen Wachstumsphase ($OD_{600\text{ nm}} \approx 10$)
- LB-Komplexmedium, bis zur exponentiellen Wachstumsphase ($OD_{600\text{ nm}} \approx 2$)

Bei Erreichen der exponentiellen Wachstumsphase wurden die Zellen für die RNA-Isolierung geerntet.

Zur Erstellung der Vollängen-Bibliothek wurde *C. glutamicum* ebenfalls als Flüssigkultur unter folgenden Bedingungen kultiviert:

- CGXII-Minimalmedium, bis zur exponentiellen Wachstumsphase ($OD_{600\text{ nm}} \approx 10$)
- LB-Komplexmedium, bis zur exponentiellen Wachstumsphase ($OD_{600\text{ nm}} \approx 2$)
- Salzstress: Zugabe von 10 % Natriumchlorid als Endkonzentration in CGXII-Minimalmedium bei $OD_{600\text{ nm}} \approx 10$ für 15 min.
- Hitzeschock: Temperatur-Veränderung von 30 °C auf 50 °C in CGXII-Minimalmedium bei $OD_{600\text{ nm}} \approx 10$ für 15 min.
- Kälteschock: Temperatur-Veränderung von 30 °C auf 4 °C in CGXII-Minimalmedium bei $OD_{600\text{ nm}} \approx 10$ für 15 min.
- Alkohol-Stress: Zugabe von 10 % Ethanol als Endkonzentration in CGXII-Minimalmedium bei $OD_{600\text{ nm}} \approx 10$ für 15 min.
- Oxidativer Stress: Zugabe von 2 mM Diamid als Endkonzentration in CGXII-Minimalmedium bei $OD_{600\text{ nm}} \approx 10$ für 15 min.
- Oxidativer Stress: Zugabe 1 % von Wasserstoffperoxid als Endkonzentration in CGXII-Minimalmedium bei $OD_{600\text{ nm}} \approx 10$ für 15 min.
- Oxidativer Stress: Änderung der Sauerstoffzufuhr von 30 % auf 90 % bei $OD_{600\text{ nm}} \approx 10$ für 2 h.

Bei Erreichen der exponentiellen Wachstumsphase bzw. nach Aussetzung der Stressoren wurden die Zellen für die RNA-Isolierung geerntet.

Für die Zell-Ernte wurden *C. glutamicum*-Zellen unter den genannten Bedingungen kultiviert, 800 µl der entsprechenden Kultur entnommen, durch kurze Zentrifugation (15 sek., 13.000 rcf) pelletiert und die Zellen nach Entfernung des Überstandes in flüssigem Stickstoff in ihrem physiologischen Zustand eingefroren. Die Zellen wurden anschließend bei -80 °C bis zur RNA-Isolierung gelagert.

3 Molekularbiologische Methoden

3.1 Isolierung von Gesamt-DNA aus *C. glutamicum*

Für eine Gesamt-DNA-Isolierung werden die Zellen mechanisch aufgeschlossen, sodass die so gewonnene DNA nicht für alle molekularbiologischen Arbeiten geeignet ist, jedoch gut als Template für PCR verwendet werden kann. Für die Entfernung von Proteinen und anderen Verunreinigungen wird die Lösung mit PCI und Chloroform behandelt und zur Präzipitation der DNA Natriumacetat und Ethanol verwendet (modifiziert nach [Chomczynski & Sacchi, 1987]).

Protokoll:

- ¼ der Zellen von einer frisch kultivierten Festmedium-Platte abnehmen, in 1 ml TRIS-HCl (25 mM, pH 8) aufnehmen, resuspendieren und in mit Silikatkügelchen (0,1 mm) gefüllte Ribotubes überführen
- Zellen im Homogenisator bei einer Geschwindigkeit von 6.500 rpm für 2 × 30 sek. aufschließen (Suspension zwischenzeitig 1 min. auf Eis abkühlen)
- 15 min. bei 4 °C und 14.000 rcf zentrifugieren
- Überstand mit 1 Vol. PCI (pH 7,5) versetzen, gut vermischen und 5 min. bei 13.000 rcf zentrifugieren
- obere, wässrige Phase abnehmen, mit 1 Vol. Chloroform versetzen, gut vermischen und 5 min. bei 13.000 rcf zentrifugieren
- obere, wässrige Phase abnehmen, mit 0,3 Vol. Natriumacetat (3 M, pH 5,2) und 2,7 Vol. Ethanol (99 %) versetzen, gut mischen und für mindestens 20 min. bei -20 °C
- präzipitierte DNA für 5 min. bei 14.000 rcf und 4 °C pelletieren
- Überstand verwerfen, das Pellet 2 × mit 500 µl Ethanol (70 %) waschen, trocknen und in 150 µl Wasser aufnehmen

3.2 Primer-Design

Die Effizienz und die Spezifität der PCR sind von der Qualität der verwendeten Primer abhängig. Um eine ausreichende Spezifität zu gewährleisten, sollte die Länge der Primer mindestens 18 und max. 30 Nukleotide betragen. Die beiden eingesetzten PCR-Primer sollten möglichst die gleiche Länge sowie Schmelztemperatur besitzen und ihr G+C-Gehalt sollte zwischen 50 und 60 % liegen. Um die Ausbildung von Haarnadelstrukturen oder die Dimerisierung der Primer zu vermeiden, ist für das Primer-Design der

Einsatz von speziellen Computerprogrammen zu empfehlen. Die in dieser Arbeit verwendeten Primer wurden mit dem Programm *CloneManager* erstellt.

3.3 Polymerase-Kettenreaktion (PCR)

Die Polymerasen-Kettenreaktion erlaubt es, definierte DNA-Abschnitte *in vitro* zu vervielfältigen und auch nachzuweisen [Sambrook *et al.*, 1989]. Dazu müssen zumindest die flankierenden Sequenzen des gewünschten DNA-Fragmentes bekannt sein, um für die PCR geeignete Primer generieren zu können.

Der Ablauf der PCR besteht aus drei grundlegenden Schritten, die mehrfach hintereinander ablaufen:

- 1) Denaturierung: Erhitzen der Matrizen-DNA (Template) bei Temperaturen über 90 °C, sodass die DNA danach einzelsträngig vorliegt.
- 2) Annealing: Hybridisierung der Primer an flankierende, komplementäre Bereiche der DNA-Einzelstränge über Wasserstoffbrückenbildung. Der Aufbau und die Länge der Primer beeinflussen die Annealing-Temperatur.
- 3) Elongation: Synthese der komplementären DNA-Stränge ausgehend von den 3'-OH-Enden der Primer durch eine thermostabile DNA-Polymerase. Die optimale Temperatur der Elongation hängt von der Reaktionstemperatur des eingesetzten Enzyms ab, liegt jedoch meistens bei 72 °C.

Das wiederholte Durchlaufen dieser Schritte führt zur exponentiellen Anreicherung des gewünschten DNA-Abschnittes.

In dieser Arbeit wurden die *BIOTAQ DNA-Polymerase* sowie die *Phusion High-Fidelity DNA-Polymerase* verwendet. Die *Phusion High-Fidelity DNA-Polymerase* verfügt im Gegensatz zur *BIOTAQ DNA-Polymerase* über eine 3'-5'-Exonuklease-Aktivität (*proof-reading*), die zu einer geringeren Fehlerrate bei der Amplifikation führt. Ein weiterer Unterschied besteht darin, dass die *BIOTAQ DNA-Polymerase* keine glatten Enden (*blunt ends*) erzeugt, sondern unabhängig vom Aufbau der Matrizen-DNA ein Adenosintriphosphat als letztes Nukleotid anhängt, sodass Amplifikate mit 3'-überhängenden Enden entstehen.

Die *BIOTAQ DNA-Polymerase* wurde in dieser Arbeit zur Überprüfung der RNA auf DNA-Verunreinigung eingesetzt und die Amplifikation der zu sequenzierenden cDNA-Bibliotheken mit der *Phusion High-Fidelity DNA-Polymerase* durchgeführt. Die PCR-Ansätze und -Programme sind im Folgenden an entsprechenden Stellen angegeben.

3.4 Agarose-Gelelektrophorese

Die Agarose-Gelelektrophorese ist eine Methode zur Auftrennung von DNA-Molekülen unterschiedlicher Längen [Sambrook *et al.*, 1989]. DNA-Moleküle können durch ihre negative Ladung (aufgrund der Phosphatgruppen) im elektrischen Feld aufgetrennt werden. Die Laufgeschwindigkeit ist abhängig von der Molekülgröße, der DNA-Konformation (linear, zirkulär, überspiralisiert), der angelegten Feldstärke und der Agarosekonzentration. Durch Vergleiche mit einem Längenstandard, bestehend aus DNA-Fragmenten definierter Länge, kann auf die Länge unbekannter DNA-Fragmente rückgeschlossen werden. Vor dem Auftragen der Probe auf das Gel wird diese mit dem Glycerin-haltigen BPB-XC-Ladepuffer versetzt, wodurch die Dichte der Probe erhöht und ein Absinken in die Geltasche erreicht wird. Gleichzeitig dient die Bromphenolblau-Bande als Marker für die zurückgelegte Laufstrecke während der Elektrophorese. Die Auftrennung der DNA-Probe wird bei einer Spannung von 50 - 100 V durchgeführt, bei einer max. Stromstärke von 300 mA. Anschließend wird die DNA im Gel mit dem DNA-interkalierenden Fluoreszenzfarbstoff Ethidiumbromid angefärbt und unter UV-Licht sichtbar gemacht.

Das 1-2 %ige Agarose-Gel wurde mit TAE-Puffer angesetzt, aufgeköcht und in eine vorbereitete Gelkammer mit Kamm gegossen. Nach dem Erkalten wurde das Gel mit TAE-Puffer überschichtet, der Kamm vorsichtig entfernt und die Proben 1:1 mit BPB-XC-Ladepuffer vermischt. Die angefärbten Proben und ein Längenstandard wurden vorsichtig in die Geltaschen gefüllt und bei 50 - 100 V aufgetrennt. Anschließend wurde das Gel etwa 5 min. mit Ethidiumbromid (1 µg/ml) gefärbt, kurz in Wasser entfärbt und die DNA-Fragmente unter UV-Licht kontrolliert.

3.5 RNA-Aufreinigung mit Phenol-Chloroform-Isoamylalkohol (PCI)

Nach jedem enzymatischen Schritt im RNAseq-Protokoll wurde die RNA mittels PCI aufgereinigt. Bei dieser Methode werden RNA und Enzyme aufgrund ihrer unterschiedlichen Löslichkeiten in Wasser und organischer Lösung in der zweiphasigen Emulsion voneinander getrennt. Anschließend wird die RNA mit Natriumacetat-Ethanol präzipitiert und in mehreren Waschschritten gereinigt (modifiziert nach [Chomczynski & Sacchi, 1987]).

Protokoll:

- 1 Vol. PCI (pH 4,8) zur RNA-Enzym-Reaktion zugeben, gut vermischen und 5 min. bei 13.000 rcf zentrifugieren
- obere, wässrige Phase abnehmen, 0,3 Vol. Natriumacetat (3 M, pH 5,2), 2,7 Vol. Ethanol (99 %) und 20 µg Glykogen versetzen, gut mischen und für mindestens 2 h bei -20 °C inkubieren
- präzipitierte RNA für 15 min. bei 13.000 rcf und 4 °C pelletieren

- Überstand verwerfen, das Pellet $2 \times$ mit 500 μ l Ethanol (70 %) waschen, trocknen und im für die darauffolgende Reaktion benötigten Volumen DEPC-Wasser aufnehmen

3.6 Kit-basierte Gesamt-RNA-Isolierung aus *C. glutamicum*

Zur Kit-basierten Gesamt-RNA-Isolierung wurde in dieser Arbeit das *RNeasy Mini Kit* (*Qiagen*) verwendet. Bei dieser Methode (modifiziert nach [Hüser *et al.*, 2003]) wird die RNA nach dem Zellaufschluss an eine Silikat-Säulenmatrix spezifisch gebunden, in mehreren Waschschriften gereinigt und eluiert. Die Zellen werden im Guanidin-Isothiocyanat-haltigen Lyse-Puffer aufgenommen, der für die Inaktivierung von RNasen sorgt. Aufgrund der sehr stabilen Zellhülle von *C. glutamicum* erfolgt der Zellaufschluss mechanisch. Um die RNA für Sequenzierungen einsetzen zu können, bei denen schon kleinste DNA-Mengen stören können, müssen zur Entfernung verbleibender DNA-Reste Behandlungen mit DNase erfolgen.

Protokoll:

- gefrorene Zellpellets in 800 μ l RLT-Puffer (+ 10 μ l β -Mercaptoethanol auf 1 ml RLT-Puffer) resuspendieren und in mit Silikatkügelchen (0,1 mm) gefüllte *Ribotubes* überführen
- Zellen im *Homogenisator* bei einer Geschwindigkeit von 6.500 rpm für 2×30 sek. aufschließen (Suspension zwischenzeitig 1 min. auf Eis abkühlen)
- 15 min. bei 4 °C und 14.000 rcf zentrifugieren, Überstand mit 250 μ l EtOH (99 %) mischen und auf eine *Qiagen*-Säule geben
- 15 sec. bei 10.000 rcf zentrifugieren und Durchfluss verwerfen
- 700 μ l RW1-Puffer auf die Säule geben und 15 sek. bei 10.000 rcf zentrifugieren
- Säule in ein neues 2 ml-Gefäß setzen
- Säule $2 \times$ mit 500 μ l RPE-Puffer waschen, 15 sek. bei 10.000 rcf zentrifugieren und Durchfluss verwerfen
- Säule 2 min. bei 10.000 rcf trocknen und in ein RNase-freies Reaktionsgefäß setzen
- zur Elution der RNA 2×50 μ l RNase freies Wasser auf die Säulen-Membran geben, jeweils 5 min. bei RT inkubieren und 1 min. bei 10.000 rcf zentrifugieren
- zugeben von 10 μ l DNase I (*Roche*) und 12 μ l Puffer (10fach, *Roche*), vorsichtig mischen und 1 h bei 37 °C inkubieren
- Suspension mit 250 μ l Ethanol (99 %) und 300 μ l RLT-Puffer mischen und auf eine *Qiagen*-Säule geben
- 15 sec. bei 10.000 rcf zentrifugieren und Durchfluss verwerfen
- 350 μ l RW1-Puffer auf die Säule geben und 15 sek. bei 10.000 rcf zentrifugieren

- 10 µl DNase I (*Qiagen*) und 70 µl RDD-Puffer vorsichtig mischen, auf die Säulen-Membran geben und 1 h bei 30 °C inkubieren
- 300 µl RW1-Puffer auf die Säule geben und 15 sek. bei 10.000 rcf zentrifugieren
- Säule 2 × mit 500 µl RPE-Puffer waschen, 15 sek. bei 10.000 rcf zentrifugieren und Durchfluss verwerfen
- Säule 2 min. bei 10.000 rcf trocknen und in ein RNase-freies Reaktionsgefäß setzen
- zur Elution der RNA 2 × 50 µl RNase-freies Wasser auf die Säulen-Membran geben, jeweils 5 min. bei RT inkubieren und 1 min. bei 10.000 rcf zentrifugieren

Die isolierte RNA wurde anschließend mittels PCR auf DNA-Verunreinigung kontrolliert. Die Qualität und Quantität der RNA wurde mit dem *NanoDrop*-Spektrophotometer und dem *Agilent Bioanalyzer* (*Agilent RNA 6000 Nano Kit* oder *Agilent RNA 6000 Pico Kit*) bestimmt. Für eine längerfristige Lagerung wurde die RNA bei -80 °C, für eine kurzfristige Lagerung bei -20 °C eingefroren.

Dieses Protokoll wurde für die Herstellung der nativen 5'-Transkriptenden-Bibliothek verwendet.

3.7 Zwei-Phasen-Extraktion von Gesamt-RNA aus *C. glutamicum*

Zur Zwei-Phasen-Extraktion von Gesamt-RNA wurde in dieser Arbeit das *TRIzol*-Reagenz und Chloroform verwendet. Bei dieser Methode werden Nukleinsäuren und andere Zellbestandteile (wie Proteine und Lipide), aufgrund ihrer unterschiedlichen Löslichkeiten in Wasser und organischer Lösung, in der zweiphasigen Emulsion voneinander getrennt. Dazu werden die Zellen in *TRIzol*-Reagenz aufgenommen, welches für die Denaturierung von Proteinen sorgt (u.a. Inaktivierung von RNasen). Daraufhin erfolgt der Zellaufschluss aufgrund der sehr stabilen Zellhülle von *C. glutamicum* mechanisch. Anschließend werden die Nukleinsäuren nochmals mit Chloroform von restlichen Zellbestandteilen separiert, mit Isopropanol präzipitiert und in mehreren Waschschrinen gereinigt (modifiziert nach [Chomczynski & Sacchi, 1987]). Auch bei dieser Methode müssen anschließend verbliebene DNA-Mengen mit DNase-Behandlungen entfernt werden.

Protokoll:

- gefrorene Zellpellets in 1 ml *TRIzol*-Reagenz resuspendieren und in mit Silikatkügelchen (0,1 mm) gefüllte *Ribotubes* überführen
- Zellen im Homogenisator bei einer Geschwindigkeit von 6.500 rpm für 2 × 30 sek. aufschließen (Suspension zwischenzeitig 1 min. auf Eis abkühlen)
- 15 min. bei 4 °C und 14.000 rcf zentrifugieren, Überstand mit 200 µl Chloroform versetzen, gut vortexen und 1 min. bei RT inkubieren

- 10 min. bei 13.000 rcf zentrifugieren
- obere, wässrige Phase abnehmen, mit 1 Vol. Isopropanol versetzen und durch Invertieren mischen
- 10 min. auf Eis stehen lassen
- präzipitierte RNA für 15 min. bei 4 °C und 14.000 rcf pelletieren
- Überstand verwerfen, das Pellet 2 × mit 500 µl Ethanol (70 %) waschen, trocknen und in 125,5 µl DEPC-Wasser aufnehmen
- 5 min. bei 65 °C inkubieren und auf Eis stellen
- 9 µl DNase I (*Roche*), 15 µl Puffer (10fach, *Roche*) und 0,5 µl *RiboLock RNase Inhibitor* vorsichtig mischen, zur RNA-Suspension geben und 1 h bei 37 °C inkubieren
- RNA-Enzym-Reaktion mit PCI aufreinigen
- 10,5 µl DNase I (*Qiagen*), 15 µl RDD-Puffer (10fach, *Qiagen*) und 0,5 µl *RiboLock RNase Inhibitor* vorsichtig mischen, zur RNA-Suspension geben und 1 h bei 30 °C inkubieren
- RNA-Enzym-Reaktion mit PCI aufreinigen

Die isolierte RNA wurde anschließend mittels PCR auf DNA-Verunreinigung kontrolliert. Die Qualität und Quantität der RNA wurde mit dem *NanoDrop*-Spektrophotometer und dem *Agilent Bioanalyzer* (*Agilent RNA 6000 Nano Kit* oder *Agilent RNA 6000 Pico Kit*) bestimmt. Für eine längerfristige Lagerung wurde die RNA bei -80 °C, für eine kurzfristige Lagerung bei -20 °C eingefroren.

Dieses Protokoll wurde für die Herstellung der Vollängentranskript-Bibliothek verwendet.

3.8 Überprüfung der RNA auf DNA-Verunreinigung mittels PCR

Im Anschluss an die RNA-Isolierung wird eine Kontroll-PCR mit der *BIOTAQ DNA-Polymerase* durchgeführt, um die Probe auf DNA-Verunreinigung zu überprüfen. Dazu werden Testprimer eingesetzt, die ein etwa 200 bp und 550 bp langes Amplifikat auf dem *C. glutamicum*-Genom erzeugen. Wird mit den Testprimern mindestens ein Amplifikat in den RNA-Proben erzielt, muss ein erneuter DNase-Verdau und eine weitere Aufreinigung der RNA durchgeführt werden.

25 µl PCR-Ansatz für die Taq-Polymerase:

deionisiertes, autoklaviertes Wasser	17,75	µl
Polymerase-Puffer (<i>NH₄ Reaction Buffer</i> , 10fach)	2,5	µl

MgCl ₂ (50 mM)	0,75	μl
dNTPs (je 2,5 mM)	1,5	μl
Primer-Mix* (je 10 μM)	1,0	μl
Template **	1,0	μl
<i>BIOTAQ DNA-Polymerase</i> (1U/μl)	0,5	μl

* Verwendete Primer: DNA_180_a, DNA_180_b, DNA_570_a, DNA_570_b

** Template: isolierte RNA; Positiv-Kontrolle: genomische DNA (~5 ng/μl); Negativ-Kontrolle: deionisiertes Wasser

PCR-Programm für die *BIOTAQ DNA-Polymerase*:

1.	95 °C	2 min.	Denaturierung
2.	95 °C	30 sek.	Denaturierung
3.	58 °C	30 sek.	Annealing
4.	72 °C	30 sek.	Elongation
5.	Schritte 2. - 4. 35 × wiederholen		
6.	72 °C	2 min.	finale Elongation
7.	8 C∞		Lagerung

Die PCR-Proben wurden anschließend mittels Agarose-Gelelektrophorese analysiert.

3.9 Qualitäts- und Quantitätskontrolle von RNA

Die Quantität und Qualität von RNA kann mit dem *NanoDrop*-Spektrophotometer und dem *Agilent Bioanalyzer* ermittelt werden.

Für die Messung mit dem *NanoDrop*-Spektrophotometer werden 2 μl Probe und Wasser als Referenz eingesetzt. Einen Hinweis auf mögliche Verunreinigungen der RNA-Proben gibt der Quotient der Absorptionen bei 260 und 280 nm (A260/A280) und bei 260 und 230 nm (A260/A230). Ein Wert A260/A280 von ~2,0 und A260/A230 von 1,8 – 2,2 deutet auf eine reine RNA-Probe hin.

Für die Messung mit dem *Agilent Bioanalyzer* (*Agilent RNA 6000 Nano Kit* oder *Agilent RNA 6000 Pico Kit*) wird die RNA-Probe entsprechend verdünnt und 1 μl der Verdünnung eingesetzt. Ein erfolgreicher Lauf zeichnet sich durch die Erkennung aller Marker-Fragmente (25, 200, 500, 1.000, 2.000, 4.000 und 6.000 nt) im Leiter-Elektropherogramm und die Erkennung des internen Markers (25 nt) in jedem Proben-Elektropherogramm aus. Bei der Kontrolle der Gesamt-RNA sollten weiterhin die 16S und 23S rRNA-Peaks (bei der Isolierung mit dem *RNeasy Mini Kit*) und 5S, 16S und 23S rRNA-Peaks (bei der Isolierung mittels *TRIZOL*-Reagenz) deutlich zu erkennen sein. Für jede Probe wird das ribosomale Verhältnis berechnet und als RIN (*RNA Integrity Number*) ausgegeben. Ein Wert von RIN ~10 weist auf

intakte RNA hin. Außerdem werden für jede Probe im Elektropherogramm die Verteilung der Größen und die Konzentration ausgegeben. Durch Einberechnung der Verdünnung der gemessenen Probe auf die ausgegebene Konzentration kann auf die Konzentration der RNA-Probe geschlossen werden.

3.10 Abreicherung von rRNA

Der Hauptteil eines bakteriellen Transkriptoms setzt sich aus stabilen RNA-Spezies (rRNA, tRNA) zusammen und macht $\geq 95\%$ aus. Daher stellt die Abreicherung der rRNA einen essentiellen Vorbereitungsschritt einer effektiven Transkriptom-Sequenzierung dar und wurde nach der Gesamt-RNA-Isolierung für die Herstellung beider Bibliotheken durchgeführt.

Zur Abreicherung von rRNA aus einer Gesamt-RNA-Lösung wurde in dieser Arbeit das *Ribo-Zero rRNA Removal Kit (Gram-positive Bacteria, Epicentre)* und die anschließende Aufreinigung der abgereicherten RNA mittels NaAc-Ethanol-Präzipitation nach Angaben des Herstellers verwendet. Bei dieser Methode wird zunächst die rRNA (23S, 16S und 5S rRNA) spezifisch an biotinylierte Fragmente hybridisiert, die zu stark konservierten Bereichen der rRNAs komplementär sind. Die Hybride (biotinylierte Fragmente:rRNA) werden im nächsten Schritt an magnetische *Beads* gebunden und mit einem Magnet-Separator selektiv entfernt.

Für die Herstellung der Vollängen-Bibliothek wurde die isolierte RNA aus neun Kultivierungsbedingungen, für die nativen 5'-Transkriptenden-Bibliothek aus zwei Kultivierungsbedingungen, äquimolar gemischt und je Bibliothek 10 μg der gemischten Gesamt-RNA zur Abreicherung eingesetzt.

Für Test-Zwecke wurde zur Abreicherung von rRNA aus einer Gesamt-RNA-Lösung aus *C. glutamicum* auch das *MICROBExpress Kit (Life Technologies; entfernt die 16S und 23S rRNA durch Hybridisierung von Fragmenten, die komplementär zu stark konservierten Bereichen der rRNAs sind)* und die *Terminator 5'-Phosphat-abhängigen Exonuklease (Epicentre; baut alle prozessierten, somit 5'-Monophosphat-Transkripte ab)* nach Angaben der Hersteller getestet.

Die Effizienz der Abreicherung wurde mit dem *Agilent Bioanalyzer (Agilent RNA 6000 Pico Kit)* überprüft.

3.11 RNA-Fragmentierung

Da zur Sequenzierung der RNAseq-Bibliotheken die *Illumina*-Plattform sowie eine *Read*-Länge von nur 27 Basen verwendet wurde, jedoch eine hohe Abdeckung der Genomsequenz durch die *Reads* gewährleistet werden sollte, wurde bei der Vorbereitung beider Bibliotheken die abgereicherte RNA

mittels Metallhydrolyse auf eine Größe von 200 – 500 nt fragmentiert. Dabei wird durch Metall-Ionen das Zucker-Phosphat-Rückgrat des RNA-Stranges (durch Angriff der Sauerstoffe sowie Hydroxygruppen) an zufälligen Stellen aufgebrochen [Forconi & Herschlag, 2009]. Im Anschluss an die Fragmentierung wurde die RNA mittels Natriumacetat-Ethanol-Präzipitation aufgereinigt und bei der darauffolgenden Fragmentierung entstandene Fragmente < 150 nt, die bei der weiteren Vorbereitung stören können, durch Größenausschluss-Präzipitation mit PEG-haltiger Lösung entfernt.

Protokoll:

- RNA-Lösung mit 0,25 Vol. Fragmentierungspuffer versetzen und gut mischen
- Inkubation bei 94 °C für 2,5 min.
- nach Hitzeinkubation direkt auf Eis, 1 Vol. Fragmentierungsstopp-Puffer (eiskalt) zugeben, gut mischen und 5 min. auf Eis inkubieren
- RNA-Probe mit 0,3 Vol. Natriumacetat (3 M, pH 5,2), 2,7 Vol. Ethanol (99 %) und 20 µg Glykogen versetzen, gut mischen und für mindestens 2 h bei -20 °C inkubieren
- präzipitierte RNA für 15 min. bei 14.000 rcf und 4 °C pelletieren
- Überstand verwerfen, das Pellet 2 × mit 500 µl Ethanol (70 %) waschen, trocknen und in 40 µl DEPC-Wasser aufnehmen
- RNA-Lösung mit 10 µl Ausschluss-Puffer versetzen, gut mischen und für 30 min. auf Eis inkubieren
- präzipitierte RNA (Fragmente > 150 nt) für 20 min. bei 14.000 rcf und 4 °C pelletieren
- Überstand verwerfen, das Pellet 2 × mit 500 µl Ethanol (70 %) waschen, trocknen und im für die darauffolgende Reaktion benötigten Volumen (+1 µl für die *Agilent Bioanalyzer*-Messung) DEPC-Wasser aufnehmen

Die Qualität der Fragmentierung und die RNA-Konzentration wurden im Anschluss an die Fragmentierung mit dem *Agilent Bioanalyzer (Agilent RNA 6000 Pico Kit)* überprüft. Der Großteil der RNA-Fragmente sollte eine Größe von 200 – 500 nt aufweisen, andernfalls ist eine weitere Fragmentierung vorzunehmen.

3.12 Anreicherung von nativen Transkriptfragmenten

Im Anschluss an die Fragmentierung erfolgte bei der Herstellung der nativen 5'-Transkriptenden-Bibliothek eine Anreicherung von nativen Transkriptfragmenten mit der *Terminator 5'-Phosphat-abhängigen Exonuklease (Epicentre)* nach Angaben des Herstellers. Dieses Enzym erkennt und verdaut

prozessierte, nicht native 5'-Transkriptenden mit 5'-Monophosphatenden. Somit werden in einer RNA-Lösung enthaltene native 5'-Transkriptfragmente angereichert.

Nach der Anreicherungsreaktion von nativen Transkriptfragmenten wurde die RNA-Enzym-Reaktion mittels PCI aufgereinigt.

3.13 Prozessierung von 5'-Tri- und Diphosphatenden zu 5'-Monophosphatenden

Im Anschluss an die Anreicherung von nativen Transkriptfragmenten bei der Herstellung der nativen 5'-Transkriptenden-Bibliothek und im Anschluss an die Fragmentierung bei der Herstellung der Vollängentranskript-Bibliothek werden die 5'-Fragment-Enden repariert, sodass eine 5'-RNA-Adapter-Ligation erfolgen kann. Dazu wird die *RNA 5'-Polyphosphatase (Epicentre)* nach Angaben des Herstellers verwendet. Dieses Enzym erkennt 5'-Tri- und Diphosphatenden und prozessiert diese zu 5'-Monophosphatenden.

Nach der Prozessierung von 5'-Tri- und Diphosphatenden zu 5'-Monophosphatenden wird die RNA-Enzym-Reaktion mit PCI aufgereinigt.

3.14 Phosphorylierung der 5'-Transkriptenden

Im Anschluss an die Prozessierung von 5'-Tri- und Diphosphatenden zu 5'-Monophosphatenden bei der Herstellung der Vollängentranskript-Bibliothek wurden die unphosphorylierten 5'-Transkriptenden phosphoryliert und phosphorylierte 3'-Transkriptenden dephosphoryliert. Aufgrund der RNA-Fragmentierung erhalten einige 5'-Transkriptenden eine Hydroxygruppe und einige 3'-Transkriptenden eine Phosphatgruppe. Für eine erfolgreiche und effektive 5'-RNA-Adapter-Ligation ist es wichtig, dass am 5'-Transkriptende eine Phosphatgruppe vorliegt. Dazu wurde die *T4 Polynukleotid-Kinase (Thermo Scientific)* nach Angaben des Herstellers verwendet. Dieses Enzym phosphoryliert unphosphorylierte 5'-Transkriptenden und dephosphoryliert phosphorylierte 3'-Transkriptenden.

Nach der Reparatur der 5'-Transkriptenden wurde die RNA-Enzym-Reaktion mittels PCI aufgereinigt.

3.15 RNA-Adapter-Ligation

Im Anschluss an die Reparatur der 5'-Transkriptenden bei der Herstellung der nativen 5'-Transkriptenden-Bibliothek und im Anschluss an die Phosphorylierung der 5'-Transkriptenden bei der Herstellung der Vollängentranskript-Bibliothek wurden RNA-Adapter an die 5'-Fragmentenden ligiert. Dazu wurde die *T4 RNA Ligase I (Thermo Scientific)* nach Angaben des Herstellers und 100 µM des RNA-Adapters pro Reaktion (bzw. Bibliothek) verwendet.

Nach der 5'-RNA-Adapter-Ligation wurde das Enzym durch Hitzebehandlung (20 min. bei 65 °C) inaktiviert und die Reaktion direkt für die reverse Transkription verwendet.

3.16 Reverse Transkription

Die Methode der reversen Transkription ermöglicht durch den Einsatz von Reversen-Transkriptasen die Umschreibung von RNA in DNA. Dieses Enzym benötigt einen DNA-Primer als Startpunkt der cDNA-Synthese, der am 3'-Ende der Template-RNA hybridisiert. Somit entsteht ein RNA-DNA-Hybridstrang, wobei der RNA-Strang nach der Synthese mit einer RNase oder der RNase-Aktivität der Reversen-Transkriptase abgebaut wird.

Der Ablauf der reversen Transkription besteht aus zwei Schritten:

- 1) **Annealing:** Die Hybridisierung des DNA-Primers an den RNA-Einzelstrang über Wasserstoffbrückenbildung. Der Aufbau und die Länge des Primers beeinflussen die Annealing-Temperatur.
- 2) **Synthese der cDNA:** Die Synthese des DNA-Stranges ausgehend vom 3'-OH-Ende des DNA-Primers durch eine Reverse-Transkriptase. Die optimale Temperatur der Synthese hängt von der Reaktionstemperatur des eingesetzten Enzyms ab, liegt jedoch meistens bei 50 °C.

Im Anschluss an die 5'-RNA-Adapter-Ligation bei der Herstellung der nativen 5'-Transkriptenden- und Vollängentranskript-Bibliothek wurden die RNA-Fragmente in cDNA umgeschrieben. Dazu wurde das *ThermoScript RT-PCR System (Life Technologies)* nach Angaben des Herstellers verwendet. Als Synthese-Primer wurde pro Reaktion (bzw. Bibliothek) 1 µl vorbereiteter DNA-*loop*-Adapter (100 µM) verwendet. Nach der Zusammensetzung aller Bestandteile für die reverse Transkription wurde die Reaktion zunächst für 30 min. bei 16 °C (Annealing des Synthese-Primers) und dann für 1 h bei 50 °C (Synthese der cDNA) inkubiert. Nach der Hitzeinaktivierung der Reversen-Transkriptase (5 min. bei 85 °C) wurde die Reaktion direkt für die Amplifikation der cDNA-Bibliothek genutzt.

Vorbereitung des DNA-*loop*-Adapters:

Bevor der DNA-*loop*-Adapter für die reverse Transkription verwendet werden kann, muss seine Struktur geöffnet und richtig gefaltet werden. Dazu wurden 20 µl des DNA-*loop*-Adapters (100 µM) für 3 min. bei 98 °C inkubiert und dann schrittweise bei einer Rate von 1 °C/10 sek. auf 25 °C abgekühlt.

3.17 Amplifikation der cDNA-Bibliotheken nach der reversen Transkription

Direkt nach der reversen Transkription wurde bei der Erstellung beider Bibliotheken die synthetisierte cDNA mit der *Phusion High-Fidelity DNA-Polymerase* in 18 Zyklen vervielfältigt.

100 µl PCR-Ansatz (pro Bibliothek) für die Phusion High-Fidelity DNA-Polymerase:

deionisiertes, autoklaviertes Wasser	44,5	µl
Polymerase-Puffer (<i>HF Buffer</i> , 5fach)	20,0	µl
DMSO (100 %)	1,5	µl
dNTPs (je 2,5 mM)	10,0	µl
Primer-Mix* (je 10 µM)	2,0	µl
cDNA	20,0	µl
Phusion High-Fidelity DNA-Polymerase (0,5U/µl)	2,0	µl

* Verwendete Primer: RNAseq_PCR_a, RNAseq_PCR_b

Für eine bessere Wärmeleitung innerhalb des Reaktionsansatzes wurde der Reaktionsansatz von 100 µl auf zwei 50 µl-Reaktionen aufgeteilt.

PCR-Programm für die Phusion High-Fidelity DNA-Polymerase:

1.	98 °C	30 sek.	Denaturierung
2.	98 °C	10 sek.	Denaturierung
3.	60 °C	30 sek.	Annealing
4.	72 °C	15 sek.	Elongation
5.	Schritte 2. - 4. 18 × wiederholen		
6.	72 °C	2 min.	finale Elongation
7.	8 C∞		Lagerung

3.18 Aufreinigung von cDNA-Bibliotheken

Im Anschluss an die PCR wurden die cDNA-Bibliotheken aufgereinigt und Fragmente < 150 bp (z.B. Primer-Dimere) entfernt. Dazu wurde das System *Agencourt AMPure XP* (*Beckman Coulter*) nach Angaben des Herstellers verwendet. Diese Methode beruht auf nicht-spezifischer Bindung von DNA-Fragmenten an sogenannte *SPRI-Beads* (*Solid Phase Reversible Immobilization Beads*) in PEG-NaCl-haltigem Puffer. Die *SPRI-Beads* sind mit Carboxygruppen beschichtet, welche in Anwesenheit von PEG und Salz reversibel die negativ-geladene DNA binden. Durch das Mischungsverhältnis von PCR-Probe zu *SPRI-Beads* lässt sich die Länge der zu entfernenden DNA-Fragmente bestimmen. Zusätzlich enthalten die *SPRI-Beads* einen magnetischen Kern, der es erlaubt, gebundene DNA-Fragmente von den PCR-Reagenzien zu separieren und mit Ethanol zu waschen.

Die Qualität und Konzentration der aufgereinigten cDNA-Bibliotheken wurde im Anschluss der Aufreinigung mit dem *Agilent Bioanalyzer (Agilent High Sensitivity DNA Assay)* überprüft. Der Großteil der cDNA-Bibliothek sollte eine Größe von 300 – 600 bp und möglichst keine Fragmente < 200 bp aufweisen, andernfalls ist eine weitere Aufreinigung der cDNA-Bibliothek vorzunehmen.

3.19 Quantifizierungs- und Qualitätskontrolle von cDNA-Bibliotheken

Die Konzentration und Qualität der cDNA-Bibliotheken wurde mit dem *Agilent Bioanalyzer* ermittelt. Für die Messung mit dem *Agilent Bioanalyzer (Agilent High Sensitivity DNA Assay)* wird die DNA-Probe entsprechend verdünnt und 1 µl der Verdünnung eingesetzt. Ein erfolgreicher Lauf zeichnet sich durch die Erkennung aller Marker-Fragmente (35, 50, 100, 150, 200, 300, 400, 500, 600, 700, 1.000, 2.000, 3.000, 7.000 und 10.380 bp) im Leiter-Elektropherogramm und die Erkennung des internen Markers (35 und 10.380 bp) in jedem Proben-Elektropherogramm aus. Für jede Probe werden im Elektropherogramm die Verteilung der Größen und die Konzentration ausgegeben. Durch Einberechnung der Verdünnung der gemessenen Probe auf die ausgegebene Konzentration kann auf die Konzentration der DNA-Probe geschlossen werden.

3.20 Sequenzierung von cDNA-Bibliotheken

Die Sequenzierung der cDNA-Bibliotheken wurde mithilfe des *TruSeq PE Cluster Kit (Illumina)* und der *Cluster Station (Illumina)* nach Angaben des Herstellers vorbereitet und die Sequenzierung mit dem *TruSeq SBS Kit (Illumina)* und dem *Genome Analyzer Iix (Illumina)* nach Angaben des Herstellers durchgeführt.

Die Bibliothek der nativen 5'-Transkriptenden wurde im *Single-Read*-Modus mit einer *Read*-Länge von 27 Basen und die Bibliothek der Vollängentranskripte im *Paired-End*-Modus mit einer *Read*-Länge von 2 × 27 Basen jeweils auf einer *Lane* einer *Flow Cell* sequenziert.

Die Datenanalyse, das *Base Calling* und das *Read-Trimming* wurden (automatisch) mithilfe der *Illumina* Software durchgeführt. Dabei wurde bei dem *Read-Trimming* jeweils eine Base am 3'-Ende jeden *Reads*, die aufgrund der Sequenzierungstechnologie eine sehr schlechte Qualität aufweist, entfernt. Somit wurden für die Kartierung und die weitere Analyse *Reads* mit einer Länge von 26 Basen verwendet.

Die Sequenzierungsdaten beider Bibliotheken sind über das *European Nucleotide Archive* (<http://www.ebi.ac.uk/ena/>) unter der Projekt-ID PRJEB4788 erhältlich.

4 Bioinformatische Analysen

Die automatisierten bioinformatischen Analysen wurden in dieser Arbeit mit der Unterstützung von Dr. Christian Rückert (CeBiTec, Universität Bielefeld) und der *Bioinformatics Resource Facility* (CeBiTec, Universität Bielefeld) durchgeführt. Von der *Bioinformatics Resource Facility* wurde für die *Read*-Kartierung *SARUMAN* und für die Visualisierung *ReadXplorer* entwickelt und zur Verfügung gestellt. Dr. Christian Rückert entwickelte ein *Perl-Script* (*calculateExpressionValues*) zur automatischen Erkennung von TSSs, neuen Transkripten und Operon-Strukturen, und stellte dies zur Verfügung.

4.1 *Read*-Kartierung und Visualisierung der RNAseq-Daten

Die *Read*-Kartierung wurde in dieser Arbeit mit *SARUMAN* [Blom *et al.*, 2011] durchgeführt. Dabei wurden die getrimmten *Reads* (je 26 Basen) auf das Genom von *C. glutamicum* ATCC 13032 [Kalinowski *et al.*, 2003] kartiert, wobei max. ein Fehler pro *Read* erlaubt war.

Kartierung der *Reads* der Vollängentranskript-Bibliothek:

Die sequenzierten Vorwärts- und Rückwärts-*Reads* (als *Paired-End-Reads*), die mit einer max. Distanz von 1 kb auf dem Genom von *C. glutamicum* kartiert wurden, wurden zu einem *Combined-Paired-Read* (CPR) verknüpft, wobei die Referenz-Sequenz von *C. glutamicum* als *Insert* eingefügt wurde (Abbildung VIII-1 a.). Falls die Vorwärts- und Rückwärts-*Reads* (als *Paired-End-Reads*) mit einer Distanz > 1 kb kartiert wurden, wurden beide *Reads* verworfen und nicht für weitere Analysen genutzt (Abbildung VIII-1 b.). Falls *Paired-End-Reads* nur einen Vorwärts- oder Rückwärts-*Reads* enthielten, wurden diese als *Single-Reads* für weitere Analysen verwendet (Abbildung VIII-1 c.).

Kartierung der *Reads* der nativen 5'-Transkriptenden-Bibliothek:

Nur die sequenzierten Vorwärts-*Reads* (*Single-Reads*) wurden für die Kartierung sowie für weitere Analysen eingesetzt.

Für die Visualisierung der Kartierungsdaten wurde *ReadXplorer* (Hilker *et al.*, *in preparation*) genutzt.

4.2 Detektion von Transkriptionsstarts

Die automatische Identifizierung von TSSs wurde mit den Kartierungsdaten der Bibliothek der nativen 5'-Transkriptenden durchgeführt. Dazu zählt das *Perl-Script* (*calculateExpressionValues*) zunächst für den Vorwärts- sowie den Rückwärtsstrang und jede Position im Genom von *C. glutamicum* die Anzahl der Start-*Mapping*-Ereignisse und addiert zu jeder Position ein Pseudocount von 1 (um später die Division durch 0 zu verhindern). Ein möglicher TSS an einer Position i auf einem Strang wird dann erkannt, wenn für diese Position i die folgenden Kriterien gelten (Abbildung VIII-2):

- $(x_i > T)$ die Anzahl der Start-*Mapping*-Ereignisse (plus Pseudocount von 1) x_i an der Position i übersteigt einen Hintergrund-Schwellenwert T und
- $(x_i/x_{i-1} > R)$ das Verhältnis von der Anzahl der Start-*Mapping*-Ereignisse (plus Pseudocount von 1) x_i an der Position i zu der Anzahl der Start-*Mapping*-Ereignisse x_{i-1} (plus Pseudocount von 1) an der Position $i-1$ übersteigt einen Schwellenwert R

Nach manueller Begutachtung der Kartierungsdaten der Bibliothek der nativen 5'-Transkriptenden im *ReadXplorer* haben sich $T = 14$ und $R = 5$ als sinnvoll erwiesen.

Zusätzlich wurden alle TSSs, die eine Distanz zum Translationsstart (TLS) > 300 nt ergaben, manuell auf falsch-positiv erkannte TSSs überprüft. Ein TSS wird als falsch-positiv bewertet, wenn es an dieser Position keinen klaren Anstieg an Start-*Mappings* gibt und im Bereich vor und hinter dem (falsch-positiven) TSS die Start-*Mapping*-Anzahl ansteigt und wieder abfällt. Dieses gilt insbesondere für TSSs innerhalb von Kodierregionen, was auf sekundäre Faltungsstrukturen hindeutet, die die Terminator 5'-Phosphat-Dependent Exonuklease nicht abbauen kann, und/oder im Bereich von einer relativ hohen *Read*-Abdeckung > 100 , was auf einen niedrigen Hintergrund-Schwellenwert T für stark exprimierte Bereiche hindeutet.

4.3 Suche von Promotor- und Ribosomenbindestellen-Motiven

Für die Suche von Promotor- und Ribosomenbindestellen-Motiven der Kartierungsdaten der Bibliothek der nativen 5'-Transkriptenden wurde das Web-Programm *Improbizer* [Ao *et al.*, 2004] verwendet. Für beide Suchen wurde die Standardeinstellung angewandt.

Für die -10-Region hat *Improbizer* ein erweitertes Motiv mit 12 bzw. 13 Basen (vier nicht konservierte *leading* Basen, ein Kern-Hexamer und zwei bzw. drei *lagging* Basen) identifiziert. Zur Vereinfachung wird in dieser Arbeit nur das Kern-Hexamer als gekürztes Motiv verwendet.

Zusätzlich werden alle identifizierten Motive in ihrer Konservierung mit Groß- und Kleinbuchstaben beschrieben. Für eine Base wird ein Großbuchstabe verwendet, wenn diese an der bestimmten Position

innerhalb des Motivs in $\geq 80\%$ der Fälle vorkommt. Die Base wird als Kleinbuchstabe dargestellt, wenn diese an der bestimmten Position innerhalb des Motivs in $> 40\%$, jedoch $< 80\%$ der Fälle gefunden wurde. Wenn die Base an der bestimmten Position innerhalb des Motivs in nur $\leq 40\%$ der Fälle vorkommt, wird diese mit einem kleinen n abgebildet.

4.4 Detektion neuer Transkripte

Die automatische Identifizierung von neuen Transkripten wurde mit den Kartierungsdaten der Bibliothek der Vollängentranskripte durchgeführt. Dazu wird zunächst für den Vorwärts- sowie den Rückwärtsstrang und jede Position in Genom von *C. glutamicum* die Anzahl aller *Mapping*-Ereignisse gezählt. Ein mögliches neues Transkript wurde durch folgende Kriterien definiert (Abbildung VIII-3):

- ($c_i > T$) für jede Position i auf dem Vorwärts- sowie den Rückwärtsstrang muss die *Read*-Abdeckung c_i an der Position i einen Hintergrund-Schwellenwert T übersteigen (Start-Position des neuen Transkripts) und
- ($c_{i+x} > T$) wenn die *Read*-Abdeckung der nachfolgenden Position $i+x$ den Hintergrund-Schwellenwert T übersteigt, dann wird die Einheit/das Transkript bis zur Position $i+x$ erweitert oder
- ($c_{i+x} < T$) wenn die *Read*-Abdeckung den Hintergrund-Schwellenwert T an der Position $i+x$ nicht mehr übersteigt, dann wird die Position $i+(x-1)$ als Transkriptende gesetzt und
- das neue Transkript überlappt auf seiner Gesamtlänge (von Position i bis zur Position $i+(x-1)$) mit keinem bereits annotierten auf dem selben Strang befindlichen Gen

Nach manueller Begutachtung der Kartierungsdaten der Bibliothek der Vollängentranskripte im *ReadXplorer* hat sich $T = 14$ als sinnvoll erwiesen.

Die Transkriptenden von Transkripten mit einer Länge von > 1 kb wurden manuell überprüft und ggf. nach folgenden Kriterien korrigiert:

- wenn das zugehörige Transkript zu einer Transkriptionseinheit gehört, die weitere TSSs aufweist, dann wird eine Position vor dem nächstliegenden TSS als Stopp definiert (jedoch muss der nächstliegende TSS weiter als 200 bp entfernt liegen bzw. zu einem anderen Transkript gehören) oder
- wenn der Durchschnitt der *Read*-Abdeckung der ersten max. 50 Basen (X) des neuen Transkripts > 500 ist, dann wird der Hintergrund-Schwellenwert $T = \sim X/15$ gesetzt, wenn $X > 1.000$ ist, dann wird $T = \sim X/25$ gesetzt, und sobald $c_{i+x} < T$ gilt, wird $i+(x-1)$ als Transkriptende festgelegt oder

- wenn der Durchschnitt der *Read*-Abdeckung der ersten max. 50 bp (X) des neuen Transkripts < 20 ist, dann wird der Hintergrund-Schwellenwert $T = 3$ gesetzt und sobald $c_{i+x} < T$ gilt, wird $i+(x-1)$ als Transkriptende festgelegt.

Zusätzlich wurden neue Transkripte unter Zuhilfenahme der Kartierungsdaten der Bibliothek der nativen 5'-Transkriptenden manuell definiert. Dazu ergab sich die Start-Position aus den TSSs, die keinem annotierten Gen, jedoch einem neuen Transkript zugeordnet wurden. Die Transkriptenden wurden manuell mithilfe der Kartierungsdaten der Bibliothek der Volllängentranskripte und nach folgenden Kriterien und definiert:

- für $X > 10.000$: $T = \sim X/100$, für $X > 1.000$: $T = \sim X/25$, für $X > 500$: $T = \sim X/15$, für $X > 50$: $T = \sim X/5$, für $X > 15$: $T = \sim X/2$, für $2 < X < 15$: $T = 3$, und sobald $c_{i+x} < T$ gilt, wird $i+(x-1)$ als Transkriptende festgelegt

Weiterhin wurden neue Transkripte manuell auf ihren genomischen Kontext untersucht. Ein Transkript wird als intergenisch definiert, wenn dieses mit keinem anderen Transkript auf dem Vorwärts- oder Rückwärtsstrang überlappt und sich zwischen zwei annotierten Genen, zwischen zwei Transkripten oder zwischen einem annotierten Gen und einem Transkript auf dem selben Strang befindet. Ein Transkript wird als *antisense* definiert, wenn sich dieses mit einem anderen Transkript auf dem Gegen-Strang überlappt. Ein Transkript gilt als intragenisch, wenn sich der zugehörige TSS innerhalb der Kodierregion eines annotierten Gens befindet. Die Transkriptenden von intragenischen Transkripten werden nicht definiert, da diese mit dem Transkript des annotierten Gens überlappen (Abbildung IV-2 a).

4.5 Identifizierung von offenen Leserastern und die Suche nach homologen Proteinen

Die Sequenzen von neuen, intergenischen Transkripten wurden mit dem Web-Programm *ORF finder* (<http://www.gen-info.osaka-u.ac.jp/~uhmin/study/orf/index.cgi>) nach offenen Leserastern durchsucht. Als Startcodons wurden ATG, GTG, TTG und CTG als Stopp-Codons TAA, TGA und TAG erlaubt.

Proteinsequenzen der identifizierten offenen Leseraster wurden zusätzlich in der NCBI-Datenbank *reference protein database (refseq_protein)* unter Zuhilfenahme von *BLASTP* [Altschul *et al.*, 1990] nach potentiellen homologen Proteinen durchsucht.

4.6 Detektion und Definition von Operon-Strukturen

Die automatische Identifizierung von Operon-Strukturen wurde mit den Kartierungsdaten der Bibliothek der Volllängen Transkripte durchgeführt. Hierbei wurde für jedes annotierte Gen getestet, ob dieses mit dem nächstliegenden annotierten Gen auf demselben Strang mit $> T$ *Paired-End-Reads* verbunden

werden konnte. Wenn die beiden Gene verbunden werden konnten, wurden diese zu einem Operon zusammengefasst und anschließend analog ein Zusammenhang zum nächstliegenden Gen überprüft.

Nach manueller Begutachtung der Kartierungsdaten der Bibliothek der Vollängentranskripte im *ReadXplorer* hat sich $T = 14$ als sinnvoll erwiesen.

In der vorliegenden Arbeit wurden die erhaltenen Operon-Strukturen wie folgt definiert:

Primär-Operon: enthält durch mindestens 15 *Paired-End-Reads* konsekutiv verbundene Gene (Abbildung IV-9)

Sub-Operon: wird durch TSSs innerhalb eines Primär-Operons definiert (Abbildung IV-9)

Monocistronisches Transkript: enthält nur ein Gen; kann auch als Sub-Operon innerhalb eines Primär-Operons vorliegen (Abbildung IV-9)

4.7 Bestimmung und Definition von Transkriptenden

Die automatische Bestimmung von Transkriptenden wurde mit den Kartierungsdaten der Bibliothek der nativen 5'-Transkriptenden und unter Zuhilfenahme der vorhergesagten ρ -unabhängigen Terminatoren von *TransTermHP* [Kingsford *et al.*, 2007] durchgeführt. Ein Transkriptende wird durch folgende Kriterien definiert (Abbildung VIII-4):

- $(X > 10)$ der Durchschnitt der *Read*-Abdeckung der 5 Basen vor dem ersten Terminator-Stamm (X) muss > 10 sein und
- $(X/t_i \geq R)$ wenn das Verhältnis vom Durchschnitt der *Read*-Abdeckung der 5 Basen vor dem ersten Terminator-Stamm (X) zu der *Read*-Abdeckung t_i an der Position i innerhalb des vorhergesagten Terminators gleich einem Schwellenwert R ist oder diesen übersteigt, wird die Position i als Transkriptende definiert

Nach manueller Begutachtung der Kartierungsdaten der Bibliothek der Vollängentranskripte im *ReadXplorer* hat sich $R = 5$ als sinnvoll erwiesen.

IV Ergebnisse

1 Die Entwicklung zweier RNAseq-Vorbereitungsprotokolle: Sequenzierung nativer 5'-Transkriptenden und Vollängentranskripte

Die Genomsequenz von *Corynebacterium glutamicum* ist nun seit mehr als zehn Jahren ermittelt [Ikeda & Nakagawa, 2003; Kalinowski *et al.*, 2003], jedoch ist die transkriptionelle Organisation des Genoms seitdem kaum untersucht worden. Um das Transkriptom von *C. glutamicum* umfangreich analysieren zu können, wurden zwei RNAseq-Vorbereitungsprotokolle entwickelt: Ein Protokoll zur Sequenzierung von Vollängentranskripten, bei dem die Gesamtheit aller Transkripte (ausgenommen der rRNA-Transkripte) zu einer cDNA-Bibliothek verarbeitet und anschließend sequenziert wird (Abbildung IV-1 a). Das zweite Protokoll dient der Sequenzierung nativer 5'-Transkriptenden, bei dem nur die zuvor angereicherten primären Transkriptenden verarbeitet und sequenziert werden (Abbildung IV-1 b). Die grundlegenden Schritte der entwickelten Protokolle setzen sich aus der RNA-Isolierung, Abreicherung der rRNA, der Fragmentierung der RNA, dem Abbau nicht-nativer Transkripte für die Sequenzierung nativer 5'-Transkriptenden (für die Sequenzierung der Vollängentranskripte bleibt dieser Schritt aus), der Vorbereitung der 5'-Enden für die Adapter-Ligation, der reversen Transkription der ligierten RNA-Fragmente in cDNA, der Amplifikation der cDNA-Fragmente und der Sequenzierung der cDNA-Bibliothek zusammen.

Diese entwickelten RNAseq-Protokolle ermöglichen es das primäre und Vollängentranskriptom separat zu sequenzieren und zu analysieren, sind durchschaubar, schrittweise zugänglich und eignen sich für die Transkriptom-Sequenzierung vieler Organismen. Weitere, detaillierte Vorteile der entwickelten zwei Protokolle sind im Folgenden dargestellt:

1) Ein essentieller Schritt für eine effektive Transkriptom-Sequenzierung ist die Abreicherung der stabilen RNAs, welche den größten Teil eines Transkriptoms ausmachen ($\geq 95\%$). Für die Abreicherung der rRNA aus *C. glutamicum*-Gesamt-RNA wurden drei kommerziell erhältliche Anwendungen getestet: die Terminator 5'-Phosphat-abhängigen Exonuklease, das *MICROBExpress Bacterial mRNA Enrichment Kit* und das *Ribo-Zero rRNA Removal Kit*. Alle getesteten Anwendungen zeigten eine Reduzierung der RNA-Menge, wobei für *C. glutamicum* und sowohl für die Kit-basierte Gesamt-RNA-Isolierung als auch für die Zwei-Phasen-Extraktion die Anwendung des *Ribo-Zero rRNA Removal Kits* die effizienteste Abreicherung ergeben hat (Tabelle III-1).

2) Durch die Metallhydrolyse werden die Transkripte auf eine Länge von 200 – 500 nt fragmentiert. Dadurch wird sichergestellt, dass möglichst alle RNA-Fragmente revers transkribiert und anschließend sequenziert werden. Bei der Vorbereitung der Bibliothek der nativen 5'-Transkriptenden wird dadurch

zusätzlich erreicht, dass für den darauffolgenden Schritt (Anreicherung nativer 5'-Transkriptenden) Fragmente mit 5'-Triphosphat- oder 5'-Diphosphatenden (Start nativer Transkripte), unphosphorylierte 5'-Enden (Ergebnis der Metallhydrolyse) oder 5'-Monophosphatenden (prozessierte Transkripte) zurückbleiben. Durch die Behandlung der Fragmente mit der Terminator 5'-Phosphat-abhängigen Exonuklease werden dann die 5'-Monophosphat-tragenden Fragmente abgebaut.

3) Die Anwendung der *RNA 5'-Polyphosphatase* ermöglicht eine Umwandlung der 5'-Triphosphat- und 5'-Diphosphatenden zu 5'-Monophosphatenden. Somit werden die Fragmente für eine effiziente Ligation der 5'-RNA-Adapter vorbereitet. Bei der Vorbereitung der Bibliothek der nativen 5'-Transkriptenden werden dadurch nur native 5'-Enden mit Adaptern versehen. Bei der Vorbereitung der Vollängentranskript-Bibliothek erfolgt vor der RNA-Adapter-Ligation eine *T4 Polynukleotid-Kinase*-Behandlung, bei der alle unphosphorylierten 5'-Enden phosphoryliert werden, um an diese RNA-Adaptern ligieren zu können.

4) Bei der cDNA-Synthese wird ein *DNA-loop-Adapter* als Primer verwendet, der via eines 3'-NNNNNNN-Schwanzes bevorzugt an das 3'-Ende der RNA-Fragmente hybridisiert, wobei die Hybridisierung durch die Wasserstoffbrückenbindung im RNA-DNA-Duplex stabilisiert wird. Da bei diesem Schritt die Markierung des 3'-Endes mit Sequenzier-Primer-Sequenzen sowie die cDNA-Synthese in einem Schritt erfolgen, wird bei diesem Ansatz ein vorbereitender Schritt für die reverse Transkription eingespart (z.B. *Poly(A)-Tailing*, Ligation eines 3'-Adapters).

5) Für eine effiziente Vorbereitung der Bibliotheken finden mehrere Größen-Selektionsschritte statt. Durch die Ausschuss-Präzipitation werden die bei der Fragmentierung entstandenen Fragmente < 150 nt, die sich bei der weiteren Vorbereitung als störend erweisen können, entfernt. Durch das *Agencourt AMPure XP*-System werden die fertigen Bibliotheken von Adapter-Dimeren separiert.

6) Die Sequenzierung der cDNA-Bibliotheken wird mittels der *Illumina*-Plattform durchgeführt, was eine hohe Sequenziertiefe liefert und das *Paired-End*-Sequenzieren erlaubt.

Die Bibliothek der Vollängentranskripte deckt mit dem aufgeführten Protokoll Transkripte > 150 nt ab und ermöglicht somit die Charakterisierung der transkriptionellen Organisation, insbesondere die Bestimmung von Operon-Strukturen, neuen, bisher unbekanntem Transkripten sowie von Transkriptenden, für *C. glutamicum*. Bei der Vorbereitung dieser Bibliothek werden primäre Transkriptenden nicht angereichert, sodass sowohl prozessierte als auch native Transkripte sequenziert werden. Weiterhin werden durch die Sequenzierung der Vollängen-Bibliothek gesamte Transkripte (einschließlich dem 3'-Ende) mit Sequenzier-*Reads* abgedeckt. Da es bei diesem Ansatz wichtig war, die Repräsentation von möglichst vielen Transkripten zu erhalten, wurde die RNA aus *C. glutamicum*-Zellen isoliert, die unter 9 verschiedenen Bedingungen kultiviert wurden (im CGXII-Minimalmedium und LB-Komplexmedium ohne Stressoren, im CGXII-Minimalmedium gestresst mit 10 % Ethanol, 10 % NaCl, 90 % gelöstem Sauerstoff, 1 % Wasserstoffperoxid, 2 mM Diamid, Hitzeschock bei 50 °C und

Kälteschock bei 4 °C für je 15 min. vor Zellernte nach Erreichen der exponentiellen Wachstumsphase), um eine umfassende Expression zu gewährleisten.

Die Bibliothek der nativen 5'-Transkriptenden enthält mit dem aufgeführten Protokoll auftretende 5'-Enden der Primärtranskripte und erlaubt die Identifizierung von TSSs und somit die Analyse von Promotoren, RBS und 5'-UTRs in *C. glutamicum*. Die Zielsetzung hierbei war die Sequenzierung der TSSs, deren Promotoren durch den *housekeeping* Sigmafaktor σ^A erkannt werden. Dazu wurde die RNA aus *C. glutamicum*-Zellen isoliert, die im CGXII-Minimalmedium und LB-Komplexmedium kultiviert wurden. Diese Kultivierungsbedingungen sollten die Aktivität des Sigmafaktors σ^A und somit die Transkription der meisten *housekeeping* Gene gewährleisten.

Tabelle IV-1: Vergleich der rRNA-Abreicherung aus *C. glutamicum*-Gesamt-RNA mit drei kommerziell erhältlichen Anwendungen

	TEN ^a	ME ^b	RZ ^c
Kit-isolierte Gesamt-RNA	~95 %	~86 %	~99 %
Zwei-Phasen-extrahierte Gesamt-RNA	~45 %	~70 %	~99 %

Die Prozentangaben beziehen sich auf die abgereicherte RNA-Menge (gemessen mit dem Agilent Bioanalyzer).

- a TEN: Terminator 5'-Phosphat-abhängige Exonuklease
 b ME: MICROBExpress Bacterial mRNA Enrichment Kit
 c RZ: Ribo-Zero rRNA Removal Kit

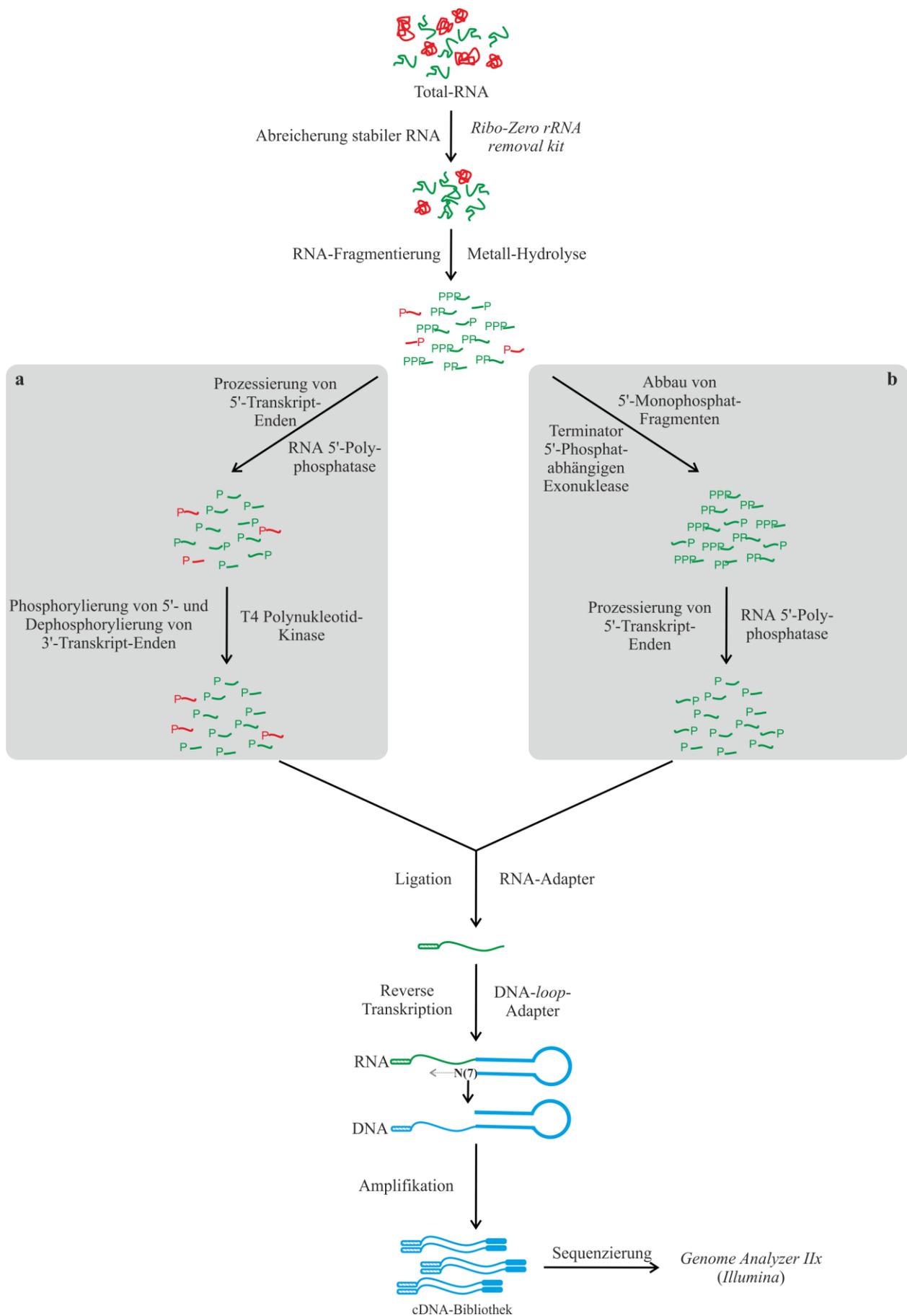


Abbildung IV-1: Experimenteller Vorbereitungsablauf der Vollängentranskript- und der nativen 5'-Transkriptenden-Bibliothek

Beide Protokolle starten mit der Isolierung der Total-RNA. Stabile rRNA wird anschließend mit dem *Ribo-Zero rRNA Removal Kit* abgereichert und die erhaltene RNA mittels Metall-Hydrolyse auf 200 – 500 nt fragmentiert. Bei der Vorbereitung der Vollängentranskript-Bibliothek (a.) werden dann die vorhandenen 5'-Triphosphat- und 5'-Diphosphatenden zu 5'-Monophosphatenden mit der *RNA 5'-Polyphosphatase* prozessiert und darauffolgend mit der *T4 Polynukleotid-Kinase* unphosphorylierte 5'-Enden phosphoryliert und phosphorylierte 3'-Enden dephosphoryliert. Bei der Vorbereitung der nativen 5'-Transkriptenden-Bibliothek (b.) werden nach der Fragmentierung alle nicht nativen Fragmente (mit einem 5'-Monophosphat) mit der *Terminator 5'-Phosphat-abhängigen Exonuklease* abgebaut und anschließend mit der *RNA 5'-Polyphosphatase* die 5'-Triphosphat- und 5'-Diphosphatenden zu 5'-Monophosphatenden prozessiert. Als Nächstes werden bei der Vorbereitung beider Bibliotheken RNA-Adapter an die vorbereiteten 5'-Monophosphatenden ligiert. Daraufhin werden die RNA-Fragmente mittels reverse Transkription in cDNA-Fragmente umgeschrieben, wobei das Markieren des 3'-Endes mit einem DNA-*loop*-Adapter erfolgt, der gleichzeitig als Primer für die Reverse Transkriptase fungiert. Der DNA-*loop*-Adapter enthält am 3'-Ende sieben freie Wobble-Basen, die mit dem 3'-Ende der RNA-Fragmente hybridisieren. Der Rest des DNA-*loop*-Adapters ist zu einer Haarnadelstruktur gefaltet. Nach der reversen Transkription werden die cDNA-Fragmente mittels PCR vervielfältigt und gleichzeitig mit Sequenzier-Linkern versehen. Anschließend werden die Bibliotheken sequenziert. Stabile RNA (rRNA, tRNA) ist in rot, andere RNA-Spezies in grün und DNA in blau dargestellt.

2 Die Generierung der Transkriptom-Daten und die Kartierung der erhaltenen *Reads* auf das Genom von *Corynebacterium glutamicum*

Um das primäre sowie das Gesamt-Transkriptom von *C. glutamicum* zu charakterisieren, wurden die zwei beschriebenen RNAseq-Bibliotheken (Volllängentranskript- und native 5'-Transkriptenden-Bibliothek) generiert und mit dem *Genome Analyzer Iix* (*Illumina*) sequenziert. Insgesamt konnten $2 \times 20,53$ Millionen *Reads* mit der Volllängentranskript-Bibliothek und 20,76 Millionen *Reads* mit der 5'-Transkriptenden-Bibliothek sequenziert sowie 15,96 Millionen *Reads* bei der Volllängentranskript-Bibliothek und 2,65 Millionen *Reads* bei der 5'-Transkriptenden-Bibliothek eindeutig an nur einer Position kartiert werden (Tabelle IV-2). Für weitere Analysen wurden multiple *Reads*, die an mehreren Positionen kartiert werden konnten, ggf. auf einen *Read* reduziert und zufällig im Genom verteilt.

Tabelle IV-2: Zusammenfassung der Sequenzier- und Kartierungsstatistiken für die Bibliotheken der Volllängentranskripte und nativen 5'-Transkriptenden

	Volllängentranskripte ^a	Native 5'-Transkript Enden ^a
Gesamt- <i>Reads</i>	$2 \times 20,53$	20,76
Gesamt-Kartierungen ^b	22,19	5,07
Gesamt-Kartierungen der CPRs ^c	13,91	-
<i>Reads</i> an multiplen Positionen ^b	1,19	0,68
CPRs an multiplen Positionen	0,72	-
<i>Reads</i> an nur einer Position ^b	15,96	2,65
CPRs an nur einer Position	10,20	-

a in Millionen angegeben

b für die Volllängentranskript-Bibliothek: Summe aus CPR, Vorwärts-*Reads*, die ohne Rückwärts-*Reads* und Rückwärts-*Reads*, die ohne Vorwärts-*Reads* auf dem Genom von *C. glutamicum* kartiert werden; für die Bibliothek der nativen 5'-Transkriptenden: Vorwärts-*Reads*, die auf dem Genom von *C. glutamicum* kartiert werden

c CPR: *Combined-Paired-End-Read* (Abbildung VIII-1)

3 Die Auswertung des Datensatzes der nativen 5'-Transkriptenden

Die umfassende Analyse des Datensatzes der nativen 5'-Transkriptenden beinhaltet die Detektion und Klassifizierung von TSSs, eine umfassende Motivsuche von σ^A -abhängigen Promotoren, die Reannotation von proteinkodierenden Genen, die Charakterisierung von 5'-UTRs und die Analyse und Motivsuche von Ribosomenbindestellen.

3.1 Detektion und Klassifizierung von Transkriptionsstarts

Ein wichtiges Element eines Transkriptoms stellt der TSS dar, das Anfangsnukleotid eines Transkripts. Somit bildet die Detektion der TSSs innerhalb des Datensatzes der nativen 5'-Transkriptenden die Basis aller weiteren RNAseq-Analysen. Als Grundlage hierfür diente der Kartierungsdatensatz der Bibliothek der nativen 5'-Transkriptenden. Zunächst erfolgte die Detektion von TSSs systematisch für den Vorwärts- und Rückwärtsstrang unter Einhaltung folgender Kriterien (Abbildung VIII-2):

- ($x_i > 14$) die Anzahl der Start-*Mapping*-Ereignisse (plus Pseudocount von 1) x_i an der Position i übersteigt einen Hintergrund-Schwellenwert T und
- ($x_i/x_{i-1} > 5$) das Verhältnis von der Anzahl der Start-*Mapping*-Ereignisse (plus Pseudocount von 1) x_i an der Position i zu der Anzahl der Start-*Mapping*-Ereignisse x_{i-1} (plus Pseudocount von 1) an der Position $i-1$ übersteigt den Schwellenwert 5
- die Position i wird als möglicher TSS erkannt

Unter Einbehaltung dieser Kriterien für den Datensatz der nativen 5'-Transkriptenden konnten insgesamt 3.163 TSSs detektiert werden. Diese Anzahl beinhaltet auch alternative TSSs, falsch-positive TSSs und TSSs, die zu rRNA- und tRNA-Genen gehören, welche für weitere Analysen entfernt wurden. Ein TSS wird als alternativer TSS definiert, wenn dieser in einem Bereich von einer Base vor oder hinter einem anderen TSS detektiert wird, wobei die Position mit der höheren Anzahl an Start-*Mapping*-Ereignissen als TSS in weitere Analysen eingeht und der andere als alternativer TSS entfernt wird. Bei alternativen TSSs ist anzunehmen, dass diese demselben Promotor entstammen und somit auf einen TSS reduziert werden können. Falsch-positive TSSs sind TSSs, die innerhalb von Kodierregionen detektiert wurden (spricht für sekundäre Faltungsstrukturen, die die *Terminator 5'-Phosphat-Dependent Exonuklease* nicht abbauen kann) und/oder im Bereich von einer relativ hohen *Read*-Abdeckung > 100 (ein Indiz, dass der Hintergrund-Schwellenwert 14 für stark exprimierte Bereiche zu niedrig angesetzt ist).

Die Gesamtanzahl der detektierten TSSs konnte um 81 alternative TSSs, 289 falsch-positive TSSs und 202 TSSs, die zu rRNA- und tRNA-Genen gehören, auf 2.591 verbleibende TSSs reduziert werden (Abbildung IV-2).

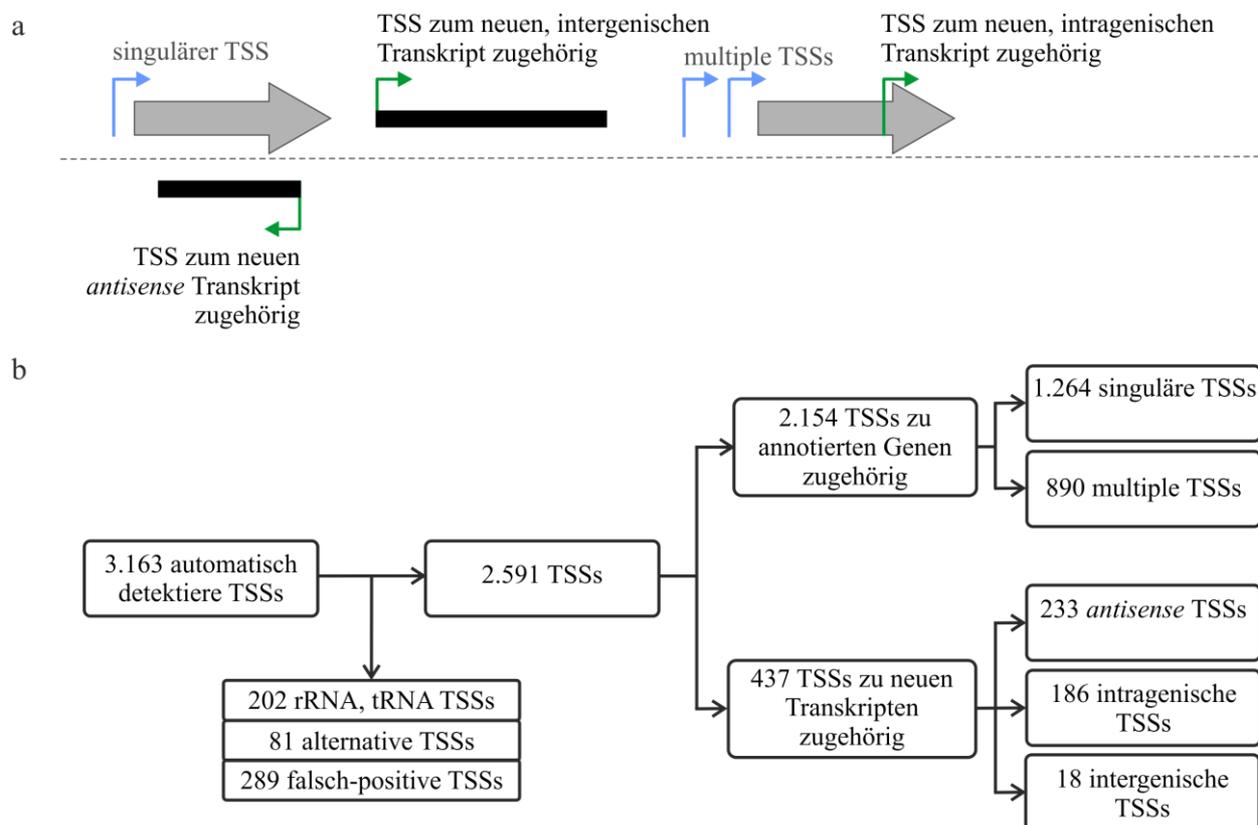


Abbildung IV-2: Klassifizierung und Einteilung der detektierten TSSs

a. Darstellung der TSS-Kategorien im genomischen Kontext. Die erste Stufe der TSS-Klassen wird in zwei Kategorien aufgeteilt: TSSs, die annotierten Genen angehören (blaue Pfeile) und TSSs, die neuen Transkripten angehören (grüne Pfeile). Als weitere Klassifizierungsebene werden TSS, die zu bereits annotierten Genen gehören, in singuläre oder multiple TSS unterteilt. TSS, die zu neuen Transkripten gehören, werden in *antisense*, intragenische oder intergenische TSS gruppiert.

b. Ablauf der Identifizierung, Filterung und Klassifizierung der TSS mittels RNAseq. Nach der automatischen Detektion von TSSs werden TSSs, die zu rRNA und tRNA, zu alternativen und falsch-positiven TSSs gehören, aussortiert. Die verbleibenden TSSs werden dann zu bereits annotierten Genen, welche als singuläre oder multiple TSS klassifiziert werden, oder zu neuen Transkripten, welche in *antisense*, intragenisch oder intergenisch unterteilt werden, zugeordnet.

Diese identifizierten 2.591 TSSs konnten nun in zwei Kategorien und weitere Unterkategorien unterteilt werden:

1) 2.154 TSSs, die zu bereits annotierten Genen gehören. Diese können weiter in 1.264 singuläre (ein einzelner TSS pro Gen) und 890 multiple TSSs (mehr als ein TSS pro Gen), die zu insgesamt 365 Genen gehören, unterteilt werden. Die max. TSS-Anzahl, die pro Gen detektiert wurde, beträgt sechs und wurde u.a. für *cmt1* (*cg0413*; Trehalose-Corynomycolyl-Transferase) detektiert (Tabelle S1).

2) 437 TSSs, die neuen, unbekanntem Transkripten zugeteilt werden. Diese umfassen 233 *antisense* (zugehöriges Transkript überlappt auf dem Gegenstrang mit einem annotierten Gen), 186 intragenische (TSS befindet sich innerhalb eines annotierten Gens auf demselben Strang) und 18 intergenische TSSs (zugehöriges Transkript befindet sich zwischen zwei annotierten Genen auf demselben Strang; Abbildung IV-2).

Von den 2.154 TSSs, die zu annotierten Genen gehören, konnten 2.147 zu proteinkodierenden und 7 zu sRNAs (ein 4.5S RNA-, M1 RNA- und 6C RNA-TSS als auch vier tmRNA-TSSs) zugewiesen werden.

Insgesamt konnten für 1.629 der derzeit 3.043 annotierten Gene in *C. glutamicum* [Meyer *et al.*, 2003] mindestens ein TSS gefunden werden.

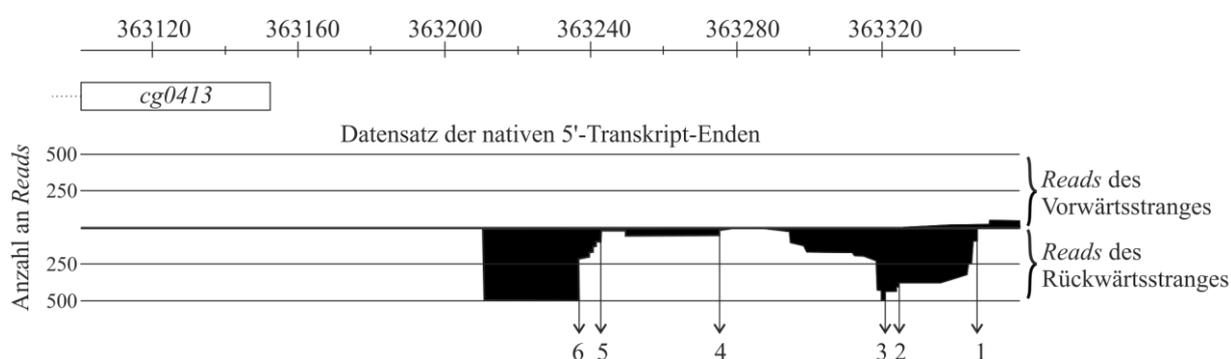


Abbildung IV-3: Veranschaulichung der Detektion von TSS anhand eines Beispiels für das Gen *cmt1/cg0413*

Die schwarz hinterlegten Kurven bzw. Stapel kennzeichnen kumulierte *Reads* der Bibliothek mit nativen 5'-Transkriptenden. Auf der y-Achse ist die *Read*-Abdeckung und auf der x-Achse die Genom-Position aufgeführt. Signifikante *Start-Mapping*-Ereignisse wurden an sechs Positionen bestimmt (1 - 6), was der TSSs für *cmt1 (cg0413)* entspricht.

Zusammenfassend lässt sich sagen, dass mittels des Datensatzes der nativen 5'-Transkriptenden zahlreiche TSSs identifiziert und zu bereits annotierten Genen als auch zu neuen, bisher unbekanntem Transkripten zugeordnet werden konnten. Weiterhin konnte somit gezeigt werden, dass unter den gewählten Bedingungen $\frac{2}{3}$ der Gene in *C. glutamicum* nur einen TSS und die übrigen $\frac{1}{3}$ Gene 2 – 6 TSSs besitzen.

3.2 Reannotation von proteinkodierenden Genen

Die erhaltenen RNAseq-Daten mit den identifizierten TSSs konnten außerdem für die Überprüfung und ggf. die Korrektur der Gen-Annotation, insbesondere von TLSs von proteinkodierenden Genen, angewandt werden. Für diese Überprüfung wurden die Positionen der Gen-Starts in *C. glutamicum* [Kalinowski *et al.*, 2003] mit den Positionen der detektierten TSSs verglichen. Für 2.834 von 3.039 der annotierten, proteinkodierenden Gene lagen die detektierten TSSs stromaufwärts zum annotierten Gen-Start und überlappten nicht mit einem Startcodon.

Sobald jedoch der TSS vor oder hinter dem annotierten Gen-Start lag sowie genau mit der ersten Base eines Startcodons überlappte, wurde überprüft, ob das überlappende Startcodon im Leseraster (*in-frame*) zum zugehörigen annotierten Stoppcodon vorliegt (Abbildung IV-4 a und b). In allen Fällen, bei denen die TSSs mit einem Startcodon überlappen, konnte festgestellt werden, dass es sich bei dem Startcodon um ein ATG oder GTG handelt, wobei das gefundene Startcodon immer im Leseraster zum zugehörigen Stoppcodon vorliegt. Somit wurde die Position des annotierten Gen-Starts mit der Position des dazugehörigen TSS korrigiert, womit es sich bei solchen Transkripten um *leaderless* mRNAs handelt. Insgesamt konnten so 185 Gen-Starts korrigiert und als *leaderless* mRNAs annotiert werden (Tabelle S2). Weiterhin konnten 20 Gen-Starts von mRNAs mit 5'-UTRs korrigiert werden. Bei diesen Genen wurde der TSS innerhalb der Kodierregion identifiziert, welches aber nicht mit einem Startcodon überlappt, das im Leseraster zum zugehörigen Stoppcodon vorliegt (Abbildung IV-4 c). Somit wurde ein neuer TLS stromabwärts von dem detektierten TSS gesucht. Dazu wurden die DNA-Sequenzen (von der TSS-Position bis zum annotierten Stoppcodon) der 20 Gene *in-silico* extrahiert und nach einem neuen Startcodon, welches im Leseraster zum Stoppcodon liegt, gescannt. Für diese Suche wurde die Web-Software *ORF finder* (<http://www.gen-info.osaka-u.ac.jp/~uhmin/study/orf/index.cgi>) mit den Einstellungen ATG, GTG, CTG und TTG als Startcodons und TAA, TGA und TAG als Stoppcodons, verwendet. Für alle 20 Gene konnte ein vom TSS stromabwärts liegender TLS gefunden werden, der zugleich im Leseraster zum Stoppcodon vorliegt (Tabelle S2). Die neu festgelegten TLSs liegen in einem Abstand von 3 – 504 Basen entfernt zum annotierten TLS.

Für alle weiteren Analysen, wie die Charakterisierung der 5'-UTRs, die RBS-Analyse, die Identifizierung von Operon-Strukturen und neuen Transkripten, wurden die hier beschriebenen, reannotierten Gen-Starts berücksichtigt.

Anhand der erhaltenen TSSs konnte die Gen-Annotation für *C. glutamicum* überprüft und korrigiert werden. Mittels des Positionsvergleichs der detektierten TSSs und der annotierten TLSs konnte die TLS-Annotation für 93 % der proteinkodierenden Gene bestätigt und für 7 % korrigiert werden.

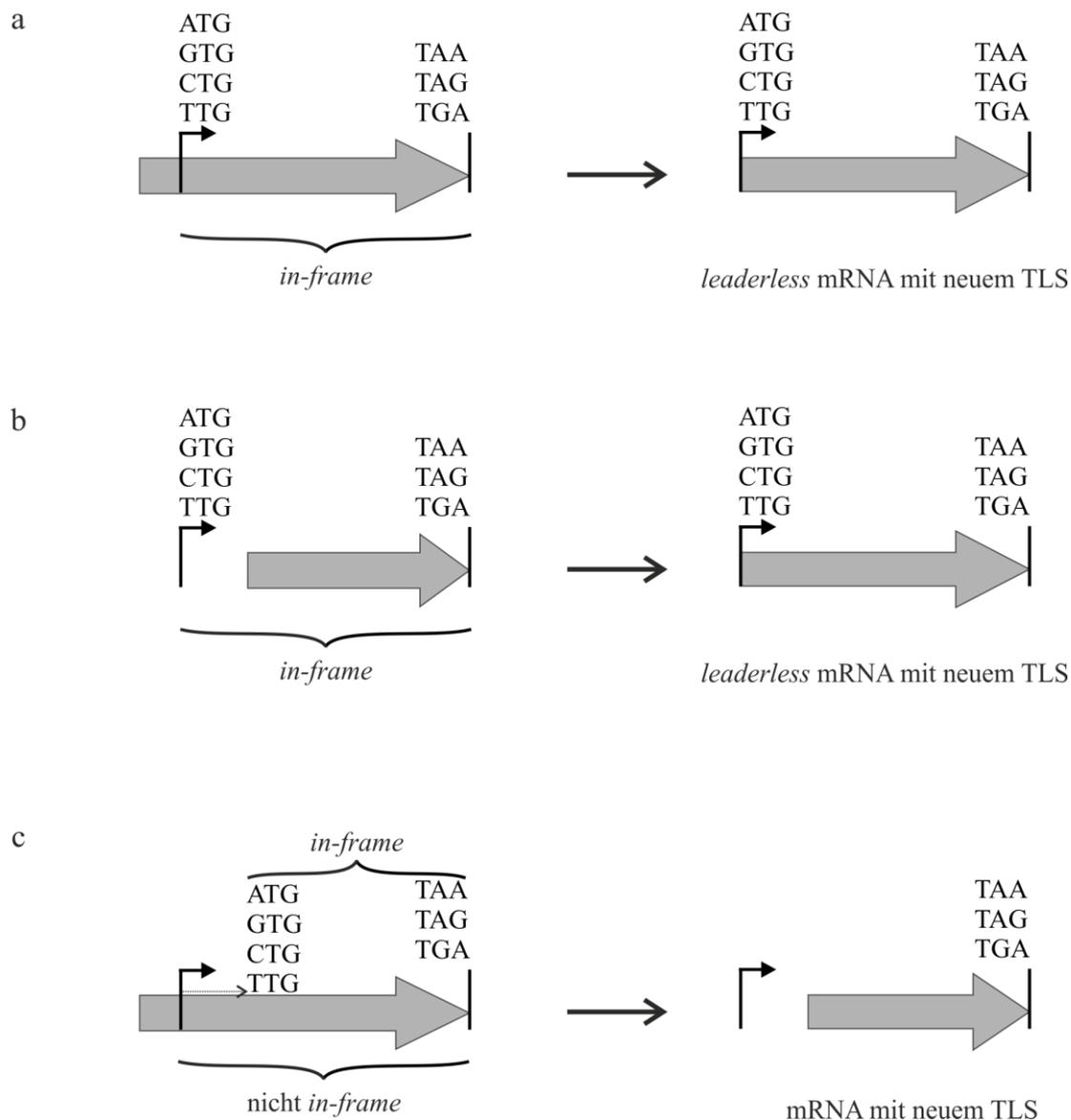


Abbildung IV-4: Schematische Darstellung der Reannotation von Gen-Starts mithilfe der RNAseq-Daten

a. und b. Korrektur von *leaderless* mRNA-Gen-Starts. Der vor oder hinter dem Startcodon detektierte TSS, der zugleich mit der ersten Base eines Startcodons überlappt, welches im Leseraster zum zugehörigen Stoppcodon vorliegt, wird als neuer TLS festgelegt.

c. Gen-Start-Korrektur von mRNAs mit 5'-UTRs. Der innerhalb der Kodierregion detektierte TSS überlappt entweder mit keinem Startcodon oder dieses liegt nicht im Leseraster zum Stoppcodon. Das nächstliegende Startcodon, welches im Leseraster zum Stoppcodon vorliegt, wird als neuer TLS festgelegt.

3.3 Umfassende Motivsuche von σ^A -abhängigen Promotoren

Promotoren sind Sequenzmotive auf der DNA, die das RNA-Polymerase-Holoenzym erkennt und bindet, um die Transkription des stromabwärts liegenden DNA-Abschnitts zu starten [Jacob, 1966; Szybalski *et al.*, 1966]. In Bakterien wurden verschiedene Sigmafaktoren als Komponenten des RNA-Polymerase-

Holoenzym identifiziert, welche unterschiedliche Promotormotive sowohl erkennen als auch binden und somit an der Transkription bestimmter Gen-Sets mitwirken [Ishihama, 1988, 2000; Paget & Helmann, 2003]. Der Sigmafaktor σ^{70} ist in allen bakteriellen Spezies zu finden, ist dort für die Transkription von *housekeeping* Genen verantwortlich [Ishihama, 1988, 2000; Pátek & Nešvera, 2011] und gilt somit als Haupt-Sigmafaktor. Dieser zeigt eine hohe Bindeaffinität zu spezifischen Promotormotiven, der sogenannten -10- und -35-Region [D Pribnow, 1975; Seeburg *et al.*, 1977; Hawley & McClure, 1983; Paget & Helmann, 2003; Pátek & Nešvera, 2011]. In *C. glutamicum* ist σ^A für die Transkription von *housekeeping* Genen verantwortlich und zählt dort als primärer Sigmafaktor, der eine hohe Proteinsequenz-Homologie zu den Haupt-Sigmafaktoren in *C. diphtheriae* (σ^A), *M. tuberculosis* (σ^A) und *E. coli* (σ^{70}) besitzt [Pátek & Nešvera, 2011].

Mithilfe der auf Basis des Datensatzes der nativen 5'-Transkriptenden gefundenen TSSs ist es möglich, nach potentiellen stromaufwärts liegenden Promotormotiven zu suchen. Für die Analyse wurde das Web-Programm *Improbizer* [Ao *et al.*, 2004] verwendet, um 60 Basen stromaufwärts der erhaltenen 2.591 TSSs nach möglichen Promotormotiven zu suchen. Die gewählten Kultivierungsbedingungen von *C. glutamicum* für die Erstellung der Bibliothek der nativen 5'-Transkriptenden sollten die Transkription der meisten *housekeeping* Gene gewährleisten. In 97 % (insgesamt 2.522 von 2.591) der durchsuchten zum TSS stromaufwärts liegenden Regionen konnte das konservierte -10-Motiv „TAnnnT“ mit einer Distanz von 3 – 11 nt zum TSS (Abbildung IV-5; Tabelle S1) gefunden werden. Das gefundene Motiv sowie die Distanz zum TSS sind mit der bereits publizierten Konsensussequenz des Promotors von σ^A in *C. glutamicum* vergleichbar [Pátek *et al.*, 2003; Pátek & Nešvera, 2011].



Abbildung IV-5: Relative Verteilung der Nukleotide innerhalb der identifizierten -10- und -35-Promotorregion in *C. glutamicum* als Sequenzlogo

Die Höhe eines einzelnen Nukleotids repräsentiert das relative Auftreten dieses Nukleotids an der aufgeführten Position. Die Kernregionen der -10- und -35-Motive sind grau hinterlegt. Die Darstellung basiert auf 2.522 -10- und 704 -35-Motiven, die mithilfe von *Improbizer* [Ao *et al.*, 2004] detektiert wurden. Die Visualisierung des Sequenzlogos erfolgte mit *WebLogo* [Crooks *et al.*, 2004].

Als Nächstes wurde in den stromaufwärts liegenden Regionen, in denen ein -10-Motiv gefunden wurde, nach einem potentiellen -35-Motiv gesucht. Als Kriterium für ein mögliches -35-Motiv wurde außerdem die Position des Motivs berücksichtigt, wobei die Distanz zwischen der -10- und -35-Region auf 16 - 19 nt gesetzt wurde [Pátek & Nešvera, 2011]. Gemäß diesen Anforderungen wurde für 704 stromaufwärts liegenden Regionen eine -35-Region mit dem schwach konservierten Motiv „ttgnca“ gefunden (Abbildung IV-5; Tabelle S1).

Die beiden gefundenen Motive für die -10- sowie -35-Region decken sich mit den bereits veröffentlichten Motiven für *C. glutamicum*, die innerhalb von 159 stromaufwärts liegenden Regionen gefunden werden konnten [Pátek & Nešvera, 2011].

Mittels dieser Analyse konnte gezeigt werden, dass mithilfe des erzeugten Datensatzes der nativen 5'-Transkriptenden eine umfassende Suche nach Promotormotiven ermöglicht wird. Außerdem konnte mittels 2.591 TSSs umfassend gezeigt werden, dass das Motiv „TAnnnT“ der -10-Region in *C. glutamicum* stark konserviert, wohingegen das Motiv „ttgnca“ der -35-Region nur sehr schwach ausgeprägt ist.

3.4 Definition und Charakterisierung von 5'-UTRs

Ein weiteres Element, das zur transkriptionellen Organisation beiträgt, ist der Transkript-Abschnitt vom TSS bis zum TLS, die untranslatierte Region am 5'-Ende eines Transkripts. Unter Zuhilfenahme der zu proteinkodierenden Genen gehörenden TSSs und der annotierten bzw. korrigierten Gen-Starts ist es möglich, die Länge der 5'-UTRs zu bestimmen und weiter zu analysieren.

Um die Längen der 5'-UTRs in *C. glutamicum* festzustellen, wurden die Positionen der 2.147 detektierten, zu proteinkodierenden Genen zugehörigen TSSs mit den entsprechenden Gen-Starts verglichen und nach Häufungen in bestimmten 5'-UTR-Längen untersucht (Abbildung IV-6; Tabelle S3).

Diese Analyse zeigte, dass der größte Anteil (~33 %, 707 von 2.147) der gefundenen, zu proteinkodierenden Genen zugehörigen Transkripte in *C. glutamicum* keine 5'-UTR besitzen (eine 5'-UTR-Länge = 0). Diese mRNAs werden in dieser Arbeit als *leaderless* Transkripte definiert. Solche Transkripte sind in allen Domänen zu finden, treten jedoch in Gram-positiven Bakterien häufiger als in Gram-negativen Bakterien auf [Moll *et al.*, 2002]. Besonders für *C. glutamicum* wurde zuvor vermutet, dass *leaderless* Transkripte vermehrt auftreten [Pátek *et al.*, 2003]. *Leaderless* mRNAs werden ohne stromaufwärts liegender Elemente wie RBSs, die die Bindung von Ribosomen an die mRNA und eine effiziente Translation begünstigen [Shine & Dalgarno, 1974; Moll *et al.*, 2002; Brock *et al.*, 2008], translatiert. Dennoch konnte für *E. coli* gezeigt werden, dass das Initiationscodon AUG und die Initiator-tRNA (fMet-tRNA) für eine stabile Bindung des Ribosoms an die mRNA wichtig sind [Moll *et al.*, 2002;

Brock *et al.*, 2008]. Die Betrachtung der Startcodons solcher mRNAs in *C. glutamicum* ergab, dass ~79 % der *leaderless* Transkripte mit einem AUG und ~21 % mit einem GUG beginnen. Der Mechanismus, wie das Startcodon GUG für die Initiation der Proteintranslation von *leaderless* Transkripten in *C. glutamicum* verwendet wird, bleibt unklar und muss weiter untersucht werden.

Eine weitere Akkumulierung von 5'-UTR ist mit Längen zwischen 26 – 40 nt (~13 %, 278 von 2.147) zu finden. Diese kurzen 5'-UTRs bieten Platz für eine RBS und enthalten wahrscheinlich keine weiteren regulatorischen Elemente.

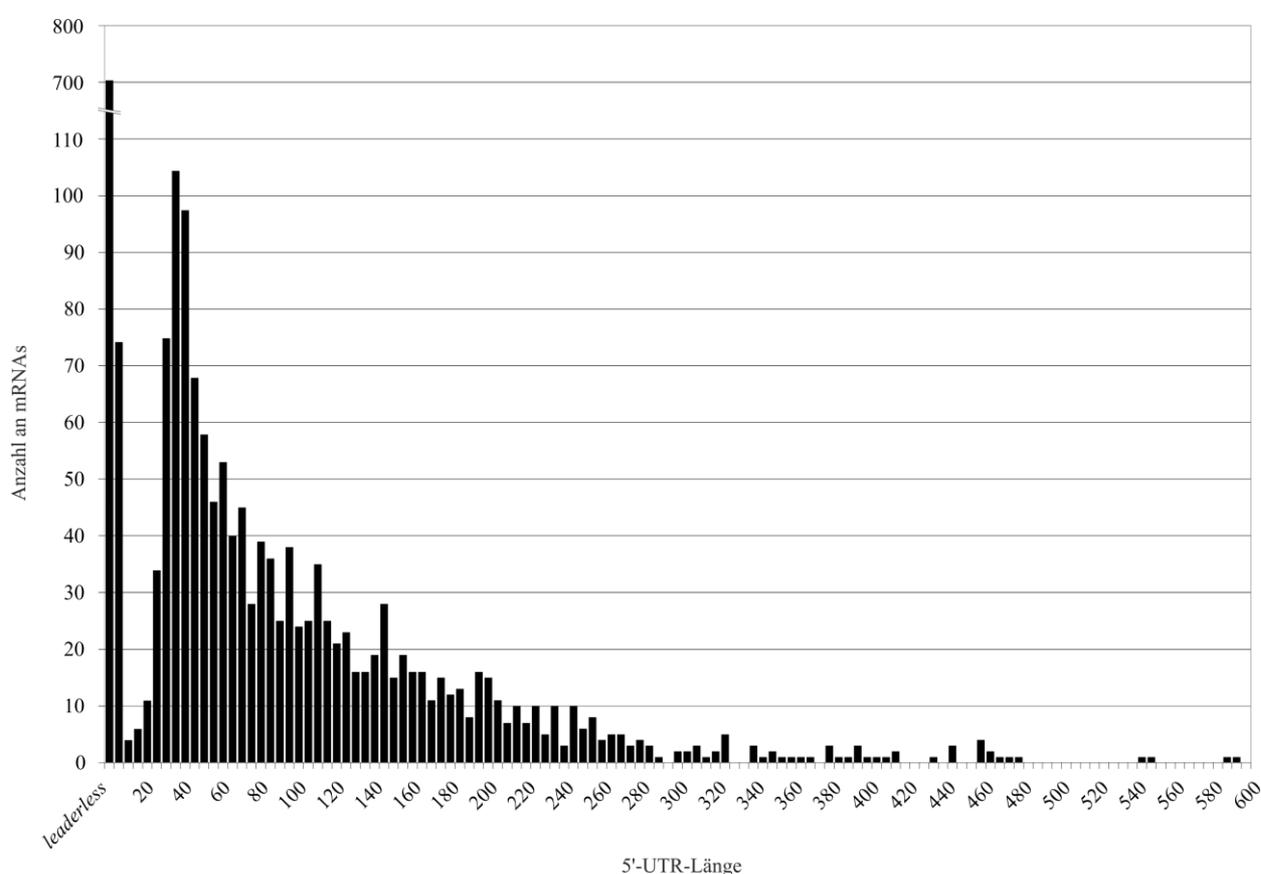


Abbildung IV-6: Diagramm mit der Verteilung der 5'-UTR-Längen von mRNAs in *C. glutamicum*

Die Darstellung basiert auf 2.147 TSSs, die proteinkodierenden Genen zugeordnet sind. Der Balken, der mit „*leaderless*“ markiert ist, repräsentiert eine UTR-Länge von null. Alle weiteren Balken stellen UTR-Längen als Intervall von 5 nt Längen (1 - 5, 6 - 10, 11 - 15, usw.) dar.

Basierend auf einer Unterteilung in Intervalle von 5 nt Länge (1 – 5, 6 – 10, 11 – 15, usw.) ergab die Analyse der Verteilung der 5'-UTR-Längen, dass die Anzahl mit länger werdenden 5'-UTRs im

Allgemeinen exponentiell abnimmt (Abbildung IV-6). Allerdings ergeben die Intervalle mit den 5'-UTR-Längen 1 – 5, 6 – 10, 11 – 15, 16 – 20, 21 – 25 und 26 – 30 nt einen deutlichen „Einbruch“ in der Verteilung, was den relativ geringen Anteil von kurzen 5'-UTRs (1 – 30 nt) in *C. glutamicum* verdeutlicht. Weiterhin stellen die 5'-UTRs mit einer Länge von 1 – 10 nt (~3,7 %, 80 von 2.147) eine interessante Ausnahme dar. Diese 5'-UTRs beinhalten nicht genügend Platz für wichtige Translationselemente, wie eine RBS und einen gewissen Abstand zum Initiationscodon. Infolgedessen bleibt der Initiationsmechanismus für solche mRNAs unklar und es kann vermutet werden, dass diese Transkripte entweder für keine Proteine kodieren, nicht-funktionelle Proteine ergeben oder ähnlich translatiert werden wie *leaderless* Transkripte.

Weiterhin wird durch diese Analyse gezeigt, dass eine relativ hohe Anzahl von 5'-UTRs eine Länge von > 100 nt besitzt (~24,7 %, 531 von 2.147), welche sich auf 384 Gene verteilt. Hierbei kann man vermuten, dass solche langen 5'-UTRs eine Funktion in der translationalen Regulation ihrer mRNAs über sekundäre Strukturformierungen übernehmen. Solche *cis*-regulatorischen 5'-UTRs wurden zuvor für viele Bakterien beschrieben, welche kodierende Sequenzen für *Leader*-Peptide, *Riboswitches*, RNA-Thermometer oder Bindestellen für andere *trans*-regulatorische RNAs beinhalten können [Winkler & Breaker, 2005; Fröhlich & Vogel, 2009; Güell *et al.*, 2011]. Um mögliche *cis*-regulatorische 5'-UTRs für *C. glutamicum* zu finden, wurden die mittels RNAseq detektierten 5'-UTRs mit den *Rfam*-vorhergesagten regulatorischen Regionen für *C. glutamicum* [Gardner *et al.*, 2011] verglichen. Insgesamt sind für *C. glutamicum* 16 regulatorische Regionen von *Rfam* vorhergesagt, wobei es bei 12 Regionen um putative *Riboswitches*, 1 putatives RNA-Thermometer und um 3 Motive mit unbekannter Funktionsweise handelt [Barrick *et al.*, 2004; Nudler & Mironov, 2004; Weinberg *et al.*, 2007, 2010; Giuliadori *et al.*, 2010; Gardner *et al.*, 2011; Watson & Fedor, 2012]. Für 13 der 16 vorhergesagten regulatorischen Regionen konnten innerhalb des Datensatzes für native 5'-Transkriptenden 5'-UTRs identifiziert werden (Tabelle IV-3), was einen Hinweis auf die Existenz dieser regulatorischen Elemente liefert.

Regulatorische Regionen besitzen oft ausgeprägte sekundäre Strukturen, durch die es dem regulatorischen Element ermöglicht wird, die zugehörigen Gene zu regulieren [Barrick *et al.*, 2004; Mandal & Breaker, 2004]. Um einen weiteren Hinweis auf die Existenz der vorhergesagten und mittels RNAseq identifizierten RNA-Elemente zu bekommen, wurden Sekundärstrukturen mittels *RNAfold* [Gruber *et al.*, 2008] berechnet. In allen Fällen können die 5'-UTRs der vorhergesagten regulatorischen Elemente stabile Haarnadelstrukturen falten. Darüber hinaus wird bei den meisten Strukturen (außer bei der 5'-UTR von *cg0215* und *cg2236/thiE*) die RBS zumindest teilweise sequestriert (Abbildung IV-7).

Das mini-*ykkC*-RNA-Motiv vor *cg1231* und zwei *yybP-ykoY-Leader* vor *cg1660* und *cg2785* konnten mit den erhaltenen RNAseq-Daten nicht bestätigt werden, da vor den genannten Genen kein TSS und folglich

keine 5'-UTR identifiziert werden konnte. Es ist jedoch möglich, dass diese Transkripte unter anderen Bedingungen auftreten könnten.

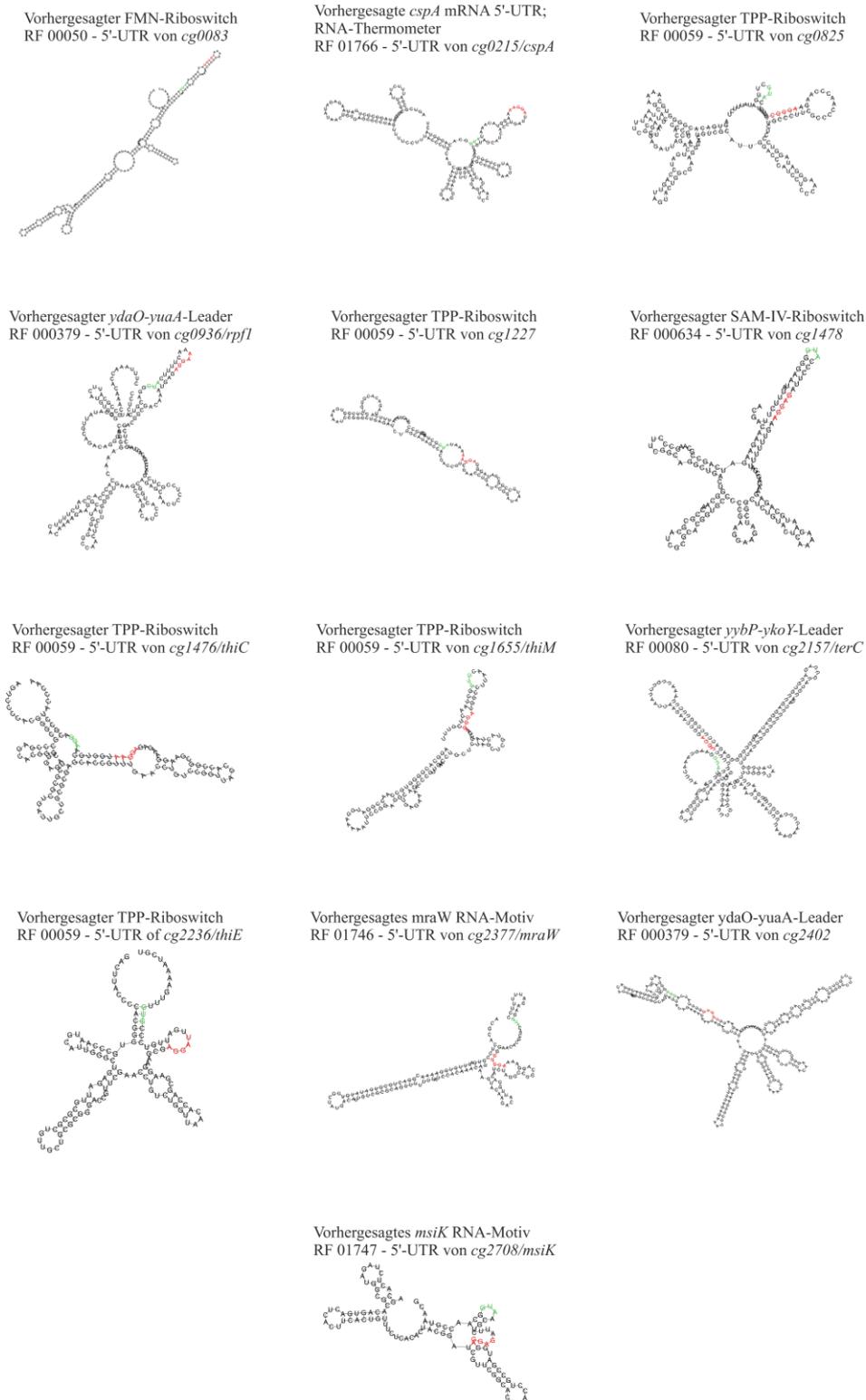


Abbildung IV-7: Sekundäre Strukturen von *Rfam*-vorhergesagten und mit RNAseq betätigten 5'-UTRs in *C. glutamicum*

Die Strukturen wurden mit *RNAfold* [Gruber *et al.*, 2008] generiert. Das Startcodon ist in grün, die mögliche RBS in rot dargestellt.

Tabelle IV-3: Vergleich von *Rfam*-vorhergesagten regulatorischen 5'-UTRs mit den RNAseq-Daten für *C. glutamicum*

Name	Rfam-Vorhersage ^a					Strang	RNAseq-Detektion ^b			
	RNA-Element	ID	Start	Stopp	Bit Score		RNAseq	Start	Stopp	stromabwärts liegendes Gen
FMN-Riboswitch	Riboswitch	RF00050	66.442	66.279	111,60	-	identifiziert	66.438	66.198	<i>cg0083</i>
<i>cspA</i> -mRNA-5'-UTR	Thermometer	RF01766	186.399	186.766	60,19	+	identifiziert	186.328	186.508	<i>cg0215 (cspA)</i>
TPP-Riboswitch	Riboswitch	RF00059	742.654	742.547	63,54	-	identifiziert	742.651	742.490	<i>cg0825</i>
<i>ydaO-yuaA-Leader</i>	Riboswitch	RF00379	870.027	869.859	69,93	-	identifiziert	870.047	869.853	<i>cg0936 (rpf1)</i>
TPP-Riboswitch	Riboswitch	RF00059	1.127.774	1.127.883	51,03	+	identifiziert	1.127.765	1.127.874	<i>cg1227</i>
mini- <i>ykkC</i> -RNA-Motiv	unbekannt	RF01068	1.131.047	1.131.094	33,52	+	nicht identifiziert	-	-	<i>cg1231</i>
TPP-Riboswitch	Riboswitch	RF00059	1.373.213	1.373.103	55,87	-	identifiziert	1.373.210	1.373.105	<i>cg1476 (thiC)</i>
SAM-IV-Riboswitch	Riboswitch	RF00634	1.374.007	1.374.123	70,47	+	identifiziert	1.374.005	1.374.139	<i>cg1478</i>
TPP-Riboswitch	Riboswitch	RF00059	1.544.490	1.544.383	52,11	-	identifiziert	1.544.485	1.544.390	<i>cg1655 (thiM)</i>
<i>yypP-ykoY-Leader</i>	Riboswitch	RF00080	1.550.030	1.550.196	43,71	+	nicht identifiziert	-	-	<i>cg1660</i>
<i>yypP-ykoY-Leader</i>	Riboswitch	RF00080	2.043.157	2.042.989	49,13	-	identifiziert	2.043.151	2.042.955	<i>cg2157 (terC)</i>
TPP-Riboswitch	Riboswitch	RF00059	2.120.271	2.120.383	62,55	+	identifiziert	2.120.271	2.120.384	<i>cg2236 (thiE)</i>
<i>mraW</i> -RNA-Motiv	unbekannt	RF01746	2.267.021	2.266.916	56,64	-	identifiziert	2.266.932	2.266.800	<i>cg2377 (mraW)</i>
<i>ydaO-yuaA-Leader</i>	Riboswitch	RF00379	2.292.467	2.292.279	59,37	-	identifiziert	2.292.509	2.292.267	<i>cg2402</i>
<i>msiK</i> -RNA-Motiv	unbekannt	RF01747	2.582.375	2.582.317	52,13	-	identifiziert	2.582.404	2.582.315	<i>cg2708 (msiK1)</i>
<i>yypP-ykoY-Leader</i>	Riboswitch	RF00080	2.649.004	2.648.890	49,90	-	nicht identifiziert	-	-	<i>cg2785</i>

a Name, ID, Genom-Positionen und *Bit Score* der vorhergesagten regulatorischen 5'-UTRs von *C. glutamicum* sind der *Rfam*-Datenbank [Gardner *et al.*, 2011] entnommen

b Start-Positionen beziehen sich auf die detektierten TSS und die Stopp-Positionen auf die letzte Position des 5'-UTR. Die sekundären Strukturen der identifizierten regulatorischen Regionen sind in Abbildung IV-7 gezeigt.

Weiterhin sollte erwähnt werden, dass mittels RNAseq auch kleine *Leader*-Peptide innerhalb von 5'-UTRs von Genen, die in die Aminosäure-Synthese involviert sind (*cg3359/trpE*, *cg0303/leuA*, *cg1435/ilvB* und *cg1129/aroF*), gefunden wurden [Mentz *et al.*, 2013]. Diese *Leader*-Peptide regulieren die zugehörigen Genen vermutlich mittels Transkription-Attenuation auf Aminosäure-Ebene [Mandal & Breaker, 2004; Nudler & Mironov, 2004] und sind Gegenstand einer anderen Arbeit [Mentz, 2013].

Die Untersuchung und Charakterisierung der 5'-UTRs zeigte, dass über 30 % der gefundenen Transkripte in *C. glutamicum* keine UTR und dass fast 25 % der identifizierten Transkripte eine 5'-UTR-Länge > 100 nt besitzen. Außerdem liefert die überdurchschnittliche Länge der 5'-UTRs einen Hinweis auf regulatorische Elemente innerhalb der 5'-UTRs, wobei 13 von 16 mittels *Rfam* [Gardner *et al.*, 2011] vorhergesagten RNA-Elemente innerhalb der RNAseq-Daten gefunden werden konnten. Die Existenz aller vorhergesagten regulatorischen Regionen wird durch die Annotation der zugehörigen Gene und durch potentielle Faltungsstrukturen unterstützt.

3.5 Analyse und Motivsuche von Ribosomenbindestellen

Ein wichtiges Element, welches zur transkriptionellen Organisation gehört und Auswirkungen auf die Translation hat, ist die Ribosomenbindestelle, die sich stromabwärts vom TSS und in der Regel innerhalb von 20 Basen stromaufwärts des Initiationscodons als Purin-reiche Region vorkommt [Shine & Dalgarno, 1974; Ringquist *et al.*, 1992; McCarthy & Brimacombe, 1994]. Weiterhin ist bekannt, dass das 3'-Ende der 16S rRNA mit der RBS der mRNA interagieren und somit die Translation initiieren [Shine & Dalgarno, 1974; McCarthy & Brimacombe, 1994]. Für *C. glutamicum* fehlen bisher solche Untersuchungen, die sich mithilfe der UTR-Sequenz realisieren lassen.

Diese Untersuchung befasst sich mit RBSs, die innerhalb einer 5'-UTR eines proteinkodierenden Gens lokalisiert sind. Intercistronische RBSs, die zwischen zwei proteinkodierenden Genen auf einem Transkript bzw. Operon liegen, werden hier nicht betrachtet. Da für die Ermittlung dieser RBSs die Kenntnis von Operon-Strukturen notwendig ist, wird die Analyse der intercistronischen RBSs im Kapitel 4.1.1 „Analyse und Motivsuche von internen Ribosomenbindestellen“ behandelt.

Für einen ersten Hinweis auf die Lokalisation der RBS wurde das Verhältnis von Purinen (Guanine und Adenine) zu Pyrimidinen (Thymine und Cytosine) innerhalb von 20 Basen stromaufwärts des Startcodons der identifizierten UTRs von proteinkodierenden Genen berechnet und verglichen (Abbildung IV-8 a). Um eine Verzerrung des Ergebnisses zu vermeiden, wurden für diese Analyse nur singuläre UTRs verwendet. Dazu wurden mehrfach vorkommende UTRs, die einem Gen zugeordnet sind, auf eine UTR reduziert, wobei die längste UTR als singuläre UTR behalten wurde. Mit dem Datensatz der nativen 5'-Transkriptenden konnten für 1.625 proteinkodierende Gene 5'-UTRs bestimmt werden, wobei 922 der proteinkodierenden Gene eine 5'-UTR-Länge ≥ 20 Basen besitzen und für die Bestimmung des Purin-

Pyrimidin-Verhältnisses berücksichtigt werden. Diese Analyse offenbart eine Akkumulation von Purinen (> 55 %) in der Region zwischen der 6. und 18. Position mit einem Maximum von 78 % an der 10. Position der 5'-UTR stromaufwärts des Initiationscodons (Abbildung IV-8 a). Diese Häufung an Purinen liefert ein Indiz für das Vorkommen der RBS in dieser Region für *C. glutamicum*.

Da eine hohe Anzahl an singulären 5'-UTRs mit einer Mindestlänge von 20 Basen verfügbar ist, wurde als Nächstes eine Suche für ein RBS-Motiv durchgeführt. Für die Suche wurde das Web-Programm *Improbizer* [Ao *et al.*, 2004] verwendet, um 20 Basen stromaufwärts des Startcodons nach möglichen RBS-Motiven zu ermitteln. In 92,5 % (insgesamt 853 von 922) der durchsuchten 20 Basen stromaufwärts des Initiationscodons konnte das konservierte RBS-Motiv „AGGag“ gefunden werden (Abbildung IV-8 b; Tabelle S4). Dieses gefundene RBS-Motiv „AGGag“ für *C. glutamicum* stellt ein perfektes, reverses Komplement des 3'-Endes der 16S rRNA „5'-UCACCUCCUUUCU-3'“ dar (unterstrichen).

Als Nächstes wurde die Verteilung der Distanz zwischen der RBS und des Startcodons untersucht. Diese Untersuchung zeigte eine Distanz von 4 – 12 Basen für > 90 % aller gefundenen RBS, wobei ein Abstand von $7,7 \pm 2,7$ Basen als Durchschnittsdistanz berechnet werden konnte (Abbildung IV-8 c; Tabelle S4).

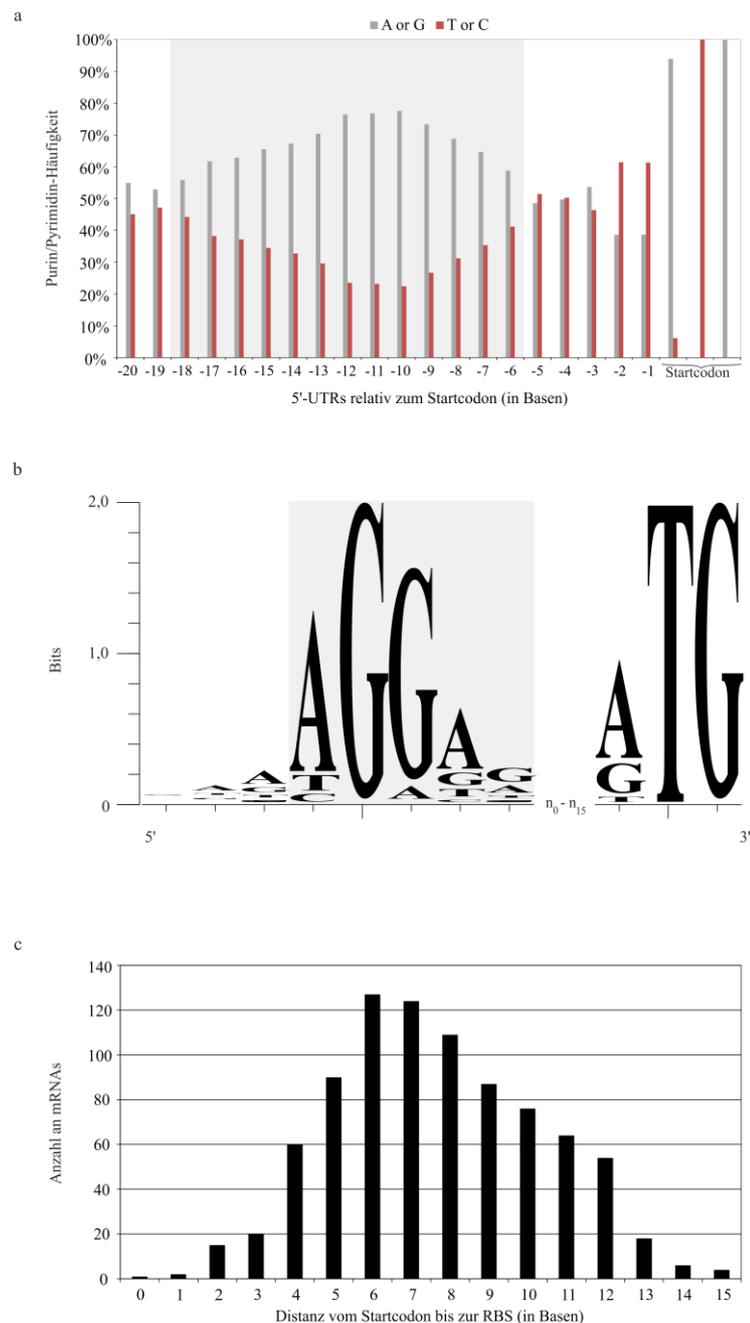


Abbildung IV-8: Analyse der RBS in *C. glutamicum*

a. Diagramm mit den Häufigkeiten von Purinen (A und G; graue Balken) im Vergleich zu den Häufigkeiten von Pyrimidinen (C und T; rote Balken) innerhalb der ersten 20 UTR-Basen (bezogen auf das Startcodon). Der grau hinterlegte Bereich (von -6 bis -18) weist eine Erhöhung in der Purin-Häufigkeit auf (> 55 %). Die Darstellung basiert auf 922 5'-UTRs.

b. Relative Verteilung der Nukleotide innerhalb der identifizierten RBS (grau hinterlegt) mit drei zusätzlichen Nukleotiden am Anfang des Motivs, einer Distanz von 0 – 15 Basen und dem Startcodon. Die Höhe eines einzelnen Nukleotids repräsentiert das relative Auftreten dieses Nukleotids an der aufgeführten Stelle. Die Analyse basiert auf 853 detektierten RBS-Motiven. Das Sequenzlogo wurde mit *WebLogo* [Crooks *et al.*, 2004] erstellt.

c. Diagramm mit der Analyse der Distanz zwischen Startcodon und RBS, die mittels *Improbizer* [Ao *et al.*, 2004] detektiert wurden. Für die Analyse wurden 853 mRNAs verwendet.

4 Die Auswertung des Datensatzes der Vollängentranskripte

Die umfassende Analyse des Datensatzes der Vollängentranskripte beinhaltet die Detektion und Beschreibung von Operon-Strukturen mit der Analyse und Motivsuche von intercistronischen Ribosomenbindestellen, die Detektion neuer Transkripte und die Analyse von Transkriptenden im Vergleich zu vorhergesagten ρ -unabhängigen Terminatoren.

4.1 Detektion und Beschreibung von Operon-Strukturen

Operons bilden eine Funktionseinheit aus Genen, die eng beieinander auf demselben Strang lokalisiert sind, koordiniert von einem Promotor aus transkribiert und ggf. reguliert werden sowie typischerweise für Enzyme eines Biosynthesewegs kodieren [Jacob & Monod, 1961a, 1961b]. Die erhaltenen RNAseq-Daten mit *Paired-End-Reads* der Vollängentranskripte ermöglichen es, u.a. annotierte Gene in Operon-Strukturen anzuordnen. Für die Bestimmung der Operon-Strukturen wurde die aktuell hinterlegte Version des annotierten *C. glutamicum*-Genoms in *GenDB* [Meyer *et al.*, 2003] mit 3.039 proteinkodierenden Genen, 4 nicht-kodierenden RNAs (6C RNA, 4.5S RNA, M1 RNA und tmRNA) und 60 tRNAs verwendet. Dabei wurden Gene zu einem polycistronischen Primär-Operon zusammengefasst, wenn ein annotiertes Gen mit dem nächstliegenden annotierten Gen auf demselben Strang mit mindestens 15 *Paired-End-Reads* verbunden werden konnte. Weiterhin wurden nebeneinanderliegende, überlappende Gene einem Primär-Operon zugeordnet, da solche Gene gekoppelt translatiert [Oppenheim & Yanofsky, 1980; Normark *et al.*, 1983; Kozak, 1999] und folglich gemeinsam transkribiert werden. Auf diese Weise konnten 2.039 Gene zu 648 Primär-Operons zusammengefasst werden. Weitere 935 Gene konnten keinem Primär-Operon zugeordnet werden, folglich treten diese als monocistronische Transkripte auf. Da außerdem mithilfe des Datensatzes der nativen 5'-Transkriptenden TSSs detektiert werden konnten, wurde festgestellt, dass die detektierten polycistronischen Operons interne TSSs enthalten und somit kleinere Untereinheiten der Primär-Operons, sogenannte Sub-Operons, bilden. Somit konnten 1.403 Gene zu alternativen 565 Sub-Operons zugeordnet werden (Abbildung IV-9; Tabelle S5). Weiterhin wurden u.a. die nicht-kodierende 6C RNA und 4.5S RNA als monocistronische Transkripte identifiziert. Die M1 RNA, tmRNA und 8 tRNAs wurden dahingegen zu Primär-Operons zugeordnet, die mit proteinkodierenden Genen co-transkribiert werden (Tabelle S5). Insgesamt konnten somit 2.966 von 3.043 annotierten Genen mithilfe der RNAseq-Daten Operon-Strukturen zugeordnet werden. Die verbleibenden 77 Gene konnten auf Grund des niedrigen Transkript-Levels (Durchschnitt der *Read*-Abdeckung < 1) keiner Operon-Struktur zugewiesen werden.

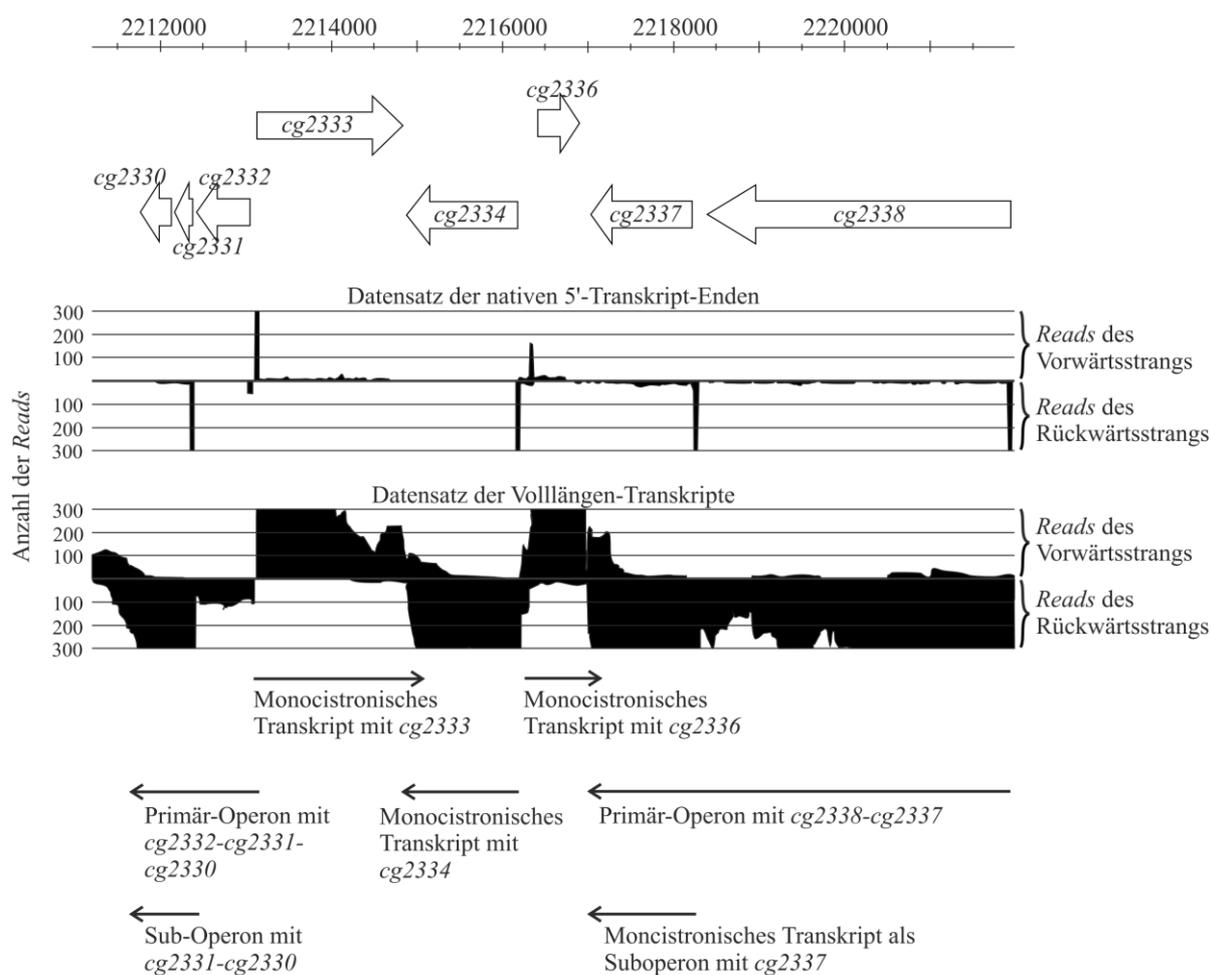


Abbildung IV-9: Identifizierung und Klassifizierung von Operon-Strukturen in *C. glutamicum*

Die schwarz hinterlegten Kurven bzw. Stapel kennzeichnen kumulierte *Reads*, die der Bibliothek mit nativen 5'-Transkriptenden (oberer Bereich) oder der Bibliothek mit Volllängentranskripten (unterer Bereich) entstammen. Auf der y-Achse ist die *Read*-Abdeckung und auf der x-Achse die Genom-Position aufgeführt. Primär-Operons werden durch kombinierte *Reads* der Bibliothek mit Volllängentranskripten, die benachbarte Gene überspannen, detektiert: *cg2332-cg2331-cg2330* und *cg2338-cg2337*; monocistronische Transkripte werden durch *Reads* der Bibliothek mit Volllängentranskripten, die nur ein Gen abdecken, identifiziert: *cg2333*, *cg2334* und *cg2336*; Sub-Operons werden durch TSSs (in der Bibliothek mit nativen 5'-Transkriptenden) innerhalb von Primär-Operons detektiert: *cg2331-cg2330*, *cg2337*.

Allgemein nimmt für Primär-Operons sowie Sub-Operons die Operon-Anzahl mit der steigenden Gen-Anzahl ab (Abbildung IV-10), sodass nur 9 Primär-Operons 10 oder mehr Gene (Tabelle IV-4) und 338 Primär-Operons 2 Gene enthalten. Das längste Primär-Operon besitzt 16 Gene, die für Proteine mit diversen zellulären Funktionen kodieren (Translation, Ribosomen-Struktur und –Biogenese, Transkription, Replikation, Rekombination und Reparatur, Kohlenhydrat-Transport und –Metabolismus sowie Coenzym-Transport und –Metabolismus; nach der *eggNOG*-Klassifizierung [Powell *et al.*, 2012]). Die manuelle Überprüfung zeigte, dass das Operon-Cluster mit 16 Genen Gen-Produkte mit Funktionen

der Riboflavin-Biosynthese (*cg1796*, *cg1797*, *cg1798*, *cg1799*, *cg1800* und *cg1801*), Protein-Modifizierung (*cg1795*, *cg1802*, *cg1803*, *cg1804*, *cg1805* und *cg1807*), die SAM-Synthese (*cg1806*) sowie unbekannte Funktionen (*cg1792*, *cg1793* und *cg1794*) enthält. Weiterhin stellte sich heraus, dass Gene, deren Produkte in die Biosynthese anderer Cofaktoren involviert sind, ebenfalls in langen Operons organisiert sind. So sind Gene der Häm- und Cytochrom C-Biosynthese in einem Operon mit 14 Genen (*cg0510* bis *cg0524*), der Folat-Biosynthese (Nukleotid-Stoffwechsel und -Recycling) in einem Operon mit 13 Genen (*cg2974* bis *cg2987*) und der NAD- und Prolin-Biosynthese in einem Operon mit 12 Genen (*cg2577* bis *cg2589*) organisiert. Außerdem zeigte sich, dass Gene, die für Proteine der Zellwand- und Glykansynthese (*cg0414* bis *cg0424*), der Zellteilung (*cg2363* bis *cg2377*), des Tat-Sekretionssystems und der Proteasom-Funktionen (*cg1683* bis *cg1693*), des Proteinstoffwechsels (*cg3011* bis *cg3020*) und für ribosomale Proteine (*cg0593* bis *cg0604*) kodieren, auch in langen Gen-Clustern organisiert sind (Tabelle IV-4). Es sei jedoch erwähnt, dass sich die in diesen langen Operon-Strukturen organisierte Gene oft nicht nur auf eine zelluläre Funktion beschränken, sondern mehrere zelluläre Funktionen innehaben.

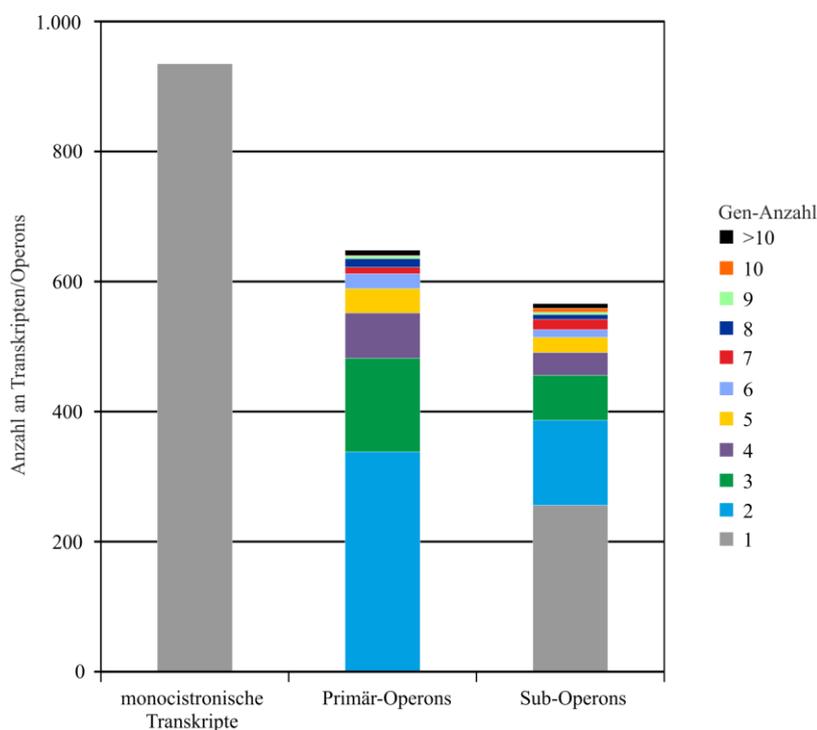


Abbildung IV-10: Analyse der Genanzahl innerhalb monocistronischer Transkripte, Primär-Operons und Sub-Operons in *C. glutamicum*

Die Genanzahl innerhalb der Klassen monocistronischer Transkripte, Primär-Operons und Sub-Operons sind in unterschiedlichen Farben dargestellt (s. Legende).

Tabelle IV-4: Liste der detektierten Operons mit ≥ 10 Genen

Gene	Gen-Anzahl	Strang	TSS ^a	<i>eggNOG</i> -Klassifizierung der Primär-Operon-Gene ^b
<i>cg3011</i> bis <i>cg3020</i>	10	-	detektiert	O, S, X (Proteinstoffwechsel, Membranproteine)
<i>cg0593</i> bis <i>cg0604</i>	11	+	detektiert	J (Ribosomale Proteine)
<i>cg1683</i> bis <i>cg1693</i>	11	-	detektiert	J, K, L, O, U, E, S (Tat-Sekretionssystem, Proteasom-Funktionen)
<i>cg0414</i> bis <i>cg0424</i>	11	+	detektiert	T, M, E, S, X (Zellwand-, Glykansynthese)
<i>cg2577</i> bis <i>cg2589</i>	12	-	detektiert	L, G, E, H, R, S, X (Prolin- und NAD-Biosynthese)
<i>cg2974</i> bis <i>cg2987</i>	13	-	detektiert	J, D, M, O, F, H, S, X (Lysyl-tRNA-Synthase, Nukleotid-Stoffwechsel und -Recycling)
<i>cg0510</i> bis <i>cg0524</i>	14	+	detektiert	O, G, H, P, X (Häm- und Cytochrom C-Biosynthese)
<i>cg2363</i> bis <i>cg2377</i>	15	-	detektiert	D, M, S (Peptidoglykan-Biosynthese, Zellteilung)
<i>cg1792</i> bis <i>cg1807</i>	16	-	detektiert	J, L, K, G, H, R, S (Riboflavin-Biosynthese, Protein-Modifikation, SAM-Synthese)

a Bezieht sich jeweils auf das erste Gen im Operon; Detektiert innerhalb des nativen 5'-Transkriptenden-Datensatzes

b *eggNOG*-Funktionen im Buchstaben-Code [Powell *et al.*, 2012]; Funktionen, die durch manuelle Überprüfung festgestellt wurden, stehen in Klammern

Zusammenfassend lässt sich sagen, dass mittels des Datensatzes der Vollängentranskripte Operon-Strukturen bestimmt werden konnten. Somit konnte gezeigt werden, dass $\sim \frac{1}{3}$ der Gene in *C. glutamicum* als monocistronische Transkripte transkribiert werden und die übrigen $\sim \frac{2}{3}$ der Gene in 648 Primär-Operons organisiert sind. Weiterhin konnten mithilfe der nativen 5'-Transkriptenden zahlreiche TSSs innerhalb von Primär-Operons detektiert werden, die 565 alternative Sub-Operons generieren.

4.1.1 Analyse und Motivsuche von internen Ribosomenbindestellen

Im Kapitel IV3.5 wurde die Analyse und Motivsuche von RBSs, die sich innerhalb einer 5'-UTR eines proteinkodierenden Gens befinden, aufgeführt. Weiterhin sind auch zwischen zwei Genen, die zu einem Operon gehören, sogenannte intercistronische RBSs lokalisiert, die hier näher betrachtet werden sollen.

Dazu wurden 20 Basen stromaufwärts der Gene, die zu einem Primär-Operon, jedoch aber nicht zum ersten Gen im Operon gehören und für die kein TSS gefunden wurde, extrahiert. Außerdem wurden stromaufwärts liegende Sequenzen, die zu überlappenden Genen gehören verworfen, da solche Gene gekoppelt translatiert werden. Bei der Translation solcher Gene wird angenommen, dass das Ribosom an dem überlappenden Stopp-Startcodon hin- und hergleitet, bis die Translation wieder eingeleitet wird [Kozak, 1999]. Somit konnten insgesamt 582 intercistronische Regionen gefunden und für eine Analyse und Motivsuche von internen RBSs verwendet werden.

Hierfür wurde zunächst ebenfalls das Verhältnis von Purinen (Guanine und Adenine) zu Pyrimidinen (Thymine und Cytosine) berechnet und verglichen (Abbildung IV-11 a). Diese Analyse zeigte eine Akkumulation von Purinen (> 55 %) in der Region zwischen der 7. und 18. Position mit einem Maximum von 75 % an der 11. Position stromaufwärts des Initiationscodons (Abbildung IV-11 a). Die darauffolgende Motivsuche mit dem Web-Programm *Improbizer* [Ao *et al.*, 2004] lieferte für 92 % (insgesamt 553 von 582) intercistronischer Sequenzen das Motiv „aAGga“ (Abbildung IV-11 b). Auch dieses Motiv stellt ein perfektes, reverses Komplement des 3'-Endes der 16S rRNA „5'-UCACCUCCUUUCU-3'“ dar (fett), jedoch ist dieses im Vergleich zu dem RBS-Motiv innerhalb der 5'-UTRs um eine Base Richtung 3'-Ende der 16S rRNA verschoben und besitzt eine andere Basen-Konservierung (vgl. „AGGag“; 5'-UCACCUCCUUUCU-3'). Auch die durchschnittliche Distanz der RBS bis zum Startcodon von $8,8 \pm 2,3$ Basen ist um etwa eine Base größer (vgl. $7,7 \pm 2,7$ Basen; Abbildung IV-11 c), was aus dem verschobenen RBS-Motiv resultiert.

Die Analyse und Motivsuche von intercistronischen RBSs in *C. glutamicum* zeigte ein ähnliches, jedoch um eine Base Richtung 3'-Ende der 16S rRNA verschobenes RBS-Motiv „aAGga“, welches einen mittleren Abstand von $8,8 \pm 2,3$ Basen zum Initiatorcodon aufweist. Es lässt sich also mutmaßen, dass die Initiation der Translation intercistronischer Gene ähnlich abläuft wie die der Gene, die sich am Anfang eines Operons befinden.

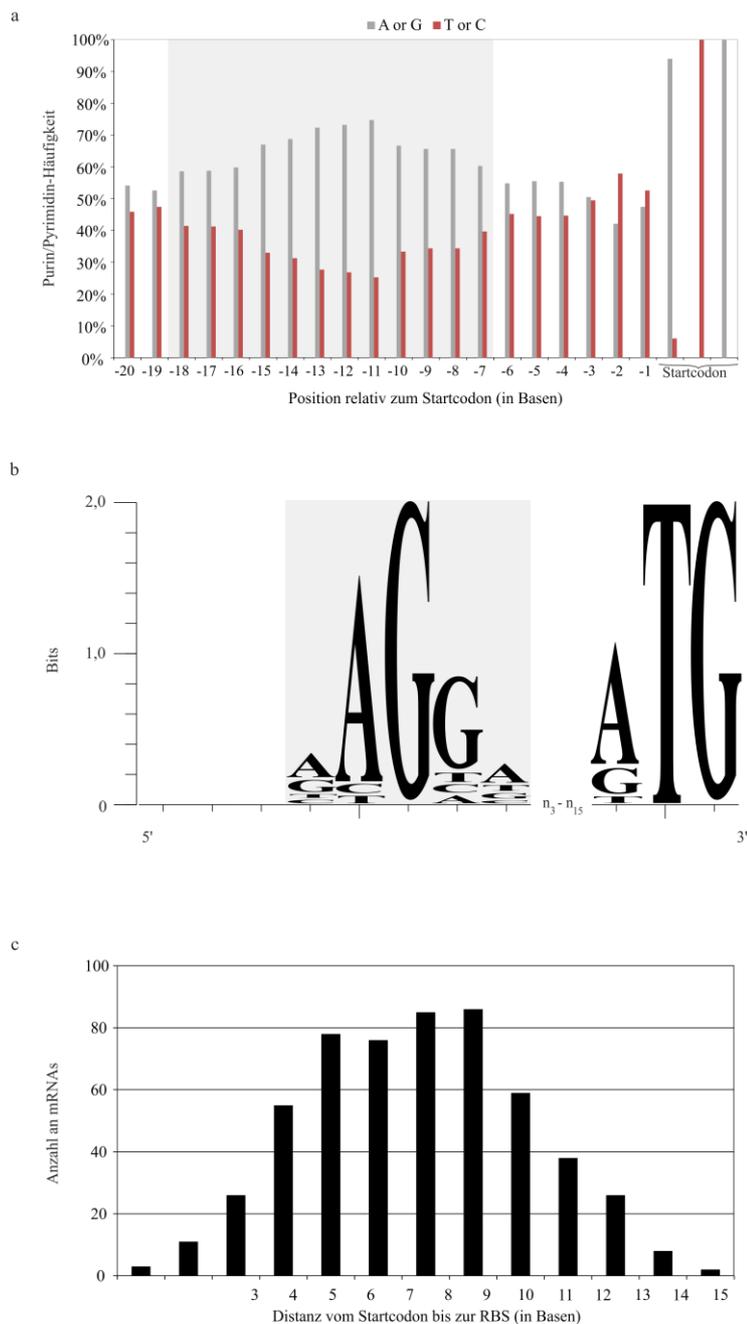


Abbildung IV-11: Analyse der intercistronischen RBS in *C. glutamicum*

a. Diagramm mit den Häufigkeiten von Purinen (A und G; graue Balken) im Vergleich zu den Häufigkeiten von Pyrimidinen (C und T; rote Balken) innerhalb der ersten 20-UTR-Basen (bezogen auf das Startcodon). Der grau hinterlegte Bereich (von -7 bis -18) weist eine Erhöhung in der Purin-Häufigkeit auf (> 55 %). Die Darstellung basiert auf 582 5'-UTRs.

b. Relative Verteilung der Nukleotide innerhalb der identifizierten RBS (grau hinterlegt) mit drei zusätzlichen Nukleotiden am Anfang des Motivs, einer Distanz von 3 – 15 Basen und dem Startcodon. Die Höhe eines einzelnen Nukleotids repräsentiert das relative Auftreten dieses Nukleotids an der gezeigten Stelle. Die Analyse basiert auf 553 detektierten RBS-Motiven. Das Sequenzlogo wurde mit *WebLogo* [Crooks *et al.*, 2004] erstellt.

c. Diagramm mit der Analyse der Distanz zwischen Startcodon und RBS, die mittels *Improbizer* [Ao *et al.*, 2004] detektiert wurden. Für die Analyse wurden 553 mRNAs verwendet.

4.2 Detektion und Klassifizierung neuer Transkripte

Neben der Validierung von bereits bekannten Genen und dessen Einteilung in Operon-Strukturen, wird mit dem erhaltenen Datensatz die Detektion neuer, bisher unbekannter Transkripte für *C. glutamicum* ermöglicht. Für die Beschreibung der Start- und Stopp-Positionen der neuen Transkripte wurden die Daten der Vollängen-Bibliothek verwendet. Zusätzlich wurden Start-Positionen neuer Transkripte mithilfe der identifizierten TSSs, die zu neuen Transkripten zugeordnet werden konnten, definiert, wobei die zugehörigen Stopp-Positionen mit Daten der Vollängen-Bibliothek beschrieben wurden. Durch die Verwendung des aufgeführten Protokolls für die Erstellung der Bibliothek der Vollängentranskripte, enthielt der Datensatz überwiegend Transkripte > 150 nt. Die Identifizierung und Charakterisierung von Transkripten < 150 nt, insbesondere sRNAs, ist in einer anderen Arbeit näher beschrieben [Mentz *et al.*, 2013; Mentz, 2013].

Insgesamt konnten mit dem vorliegenden Datensätzen 916 neue, bisher unbekannte Transkripte für *C. glutamicum* detektiert werden. Weiterhin konnten diese neuen Transkripte in drei Kategorien eingeteilt werden: (i) 30 intergenische Transkripte, die zwischen zwei annotierten Genen lokalisiert sind, (ii) 700 *antisense* Transkripte, die auf dem Gegenstrang mit einem annotierten Gen überlappen und (iii) 186 intragenische Transkripte, die sich innerhalb eines annotierten Gens befinden (Tabellen S6 – S8).

Für 18 der 30 intergenischen Transkripte konnten TSSs innerhalb des Datensatzes für primäre 5'-Transkriptenden identifiziert werden, von denen 17 ein konserviertes σ^A -abhängiges Promotor-Motiv besitzen. Für die restlichen 12 Transkripte konnten unter den gewählten Kultivierungsbedingungen (CGXII-Minimalmedium und LB-Komplexmedium, wobei hauptsächlich σ^A -abhängige Promotoren induziert sind) keine TSSs identifiziert werden. Da die Detektion der neuen Transkripte auf Grundlage der Vollängen-Bibliothek erfolgte, die auf Basis verschiedener Kultivierungsbedingungen - u.a. auch Stress-Bedingungen – erstellt wurde, lässt sich für die restlichen 12 TSSs-freien Transkripte eine σ^A -unabhängige Transkription vermuten. Weiterhin können insbesondere intergenische Transkripte für Proteine kodieren. Um potentielle Kodierregionen innerhalb der intergenischen Transkripte festzustellen, wurden die zugehörigen Sequenzen nach offenen Leserahmen durchsucht. Dazu wurde das Web-Programm *ORF finder* (<http://www.gen-info.osaka-u.ac.jp/~uhmin/study/orf/index.cgi>) mit den Einstellungen ATG, GTG, CTG und TTG als Startcodons und TAA, TGA und TAG als Stoppcodons, verwendet. Für 29 der 30 Transkripte konnte mindestens ein offener Leserahmen gefunden werden, wobei die 17 längsten gefundenen Kodierregionen zumindest ein homologes Protein in einer anderen Spezies besitzen (*BLASTP* [Altschul *et al.*, 1990], *Reference Protein Database (refseq_protein)*). Um eine weitere Evidenz der proteinkodierenden Sequenzen für die intergenischen Transkripte zu erhalten, wurden die stromaufwärts liegenden Regionen nach potentiellen RBS durchsucht. Während ein Transkript als *leaderless* Transkript identifiziert werden konnte, wurde für 13 der restlichen 28 verbleibenden putativ proteinkodierenden Transkripte eine mögliche RBS gefunden (Tabelle S6).

Die 700 detektierten *antisense* Transkripte konnten darüber hinaus ihrem genomischen Kontext zugeordnet werden: 678 Transkripte liegen *antisense* zu einem, zwei oder drei annotierten Genen, 14 Transkripte liegen *antisense* zur 5'- oder 3'-UTR eines annotierten Gens, 3 Transkripte überspannen auf dem Gegenstrang jeweils ein annotiertes Gen und ein neues Transkript und 5 Transkripte sind jeweils *antisense* zu einem neuen Transkript. Weiterhin konnten für 233 der 700 *antisense* Transkripte TSSs innerhalb des Datensatzes für primäre 5'-Transkriptenden identifiziert werden, von denen 230 ein konserviertes σ^A -abhängiges Promotor-Motiv besitzen. Für die restlichen 467 Transkripte konnten unter den gewählten Kultivierungsbedingungen (CGXII-Minimalmedium und LB-Komplexmedium, wobei hauptsächlich σ^A -abhängige Promotoren induziert sind) keine TSSs identifiziert werden. Jedoch lässt sich auch hier eine σ^A -unabhängige Transkription mutmaßen, die wahrscheinlich auf die eingesetzten Stress-Bedingungen zurückgeht, da die Transkripte innerhalb der Bibliothek für Vollängentranskripte, aber nicht innerhalb der Bibliothek für primäre 5'-Transkriptenden, identifiziert werden konnten (Tabelle S7). Für alle 186 intragenischen Transkripte konnten TSSs innerhalb des Datensatzes für primäre 5'-Transkriptenden identifiziert werden, von denen 170 ein konserviertes σ^A -abhängiges Promotor-Motiv tragen (Tabelle S8). Zudem stellen die intragenischen Transkripte eine Ausnahme der neu detektierten Transkripte dar. Diese konnten nicht in ihrer Länge bestimmt oder weiter analysiert werden, da diese mit den Transkripten der annotierten Gene überlappen. Diese liefern lediglich einen Hinweis auf intragenische Transkription und lassen die Komplexität eines Transkriptoms erkennen.

Die Suche und Klassifizierung von neuen, bisher unbekanntem Transkripten in *C. glutamicum* zeigte eine Vielfalt an *antisense* Transkripten (insgesamt 700). Weiterhin konnten mit den RNAseq-Daten 30 intergenische und 186 intragenische Transkripte gefunden und näher charakterisiert werden.

4.3 Detektion von Transkriptenden und Vergleich mit vorhergesagten ρ -unabhängigen Terminatoren

Die ρ -abhängige und -unabhängige Termination sind zwei wesentliche Mechanismen, welche die bakteriellen Transkriptionseinheiten bestimmen [Wilson & von Hippel, 1995]. Die ρ -abhängige Termination benötigt dabei ein Protein, den ρ -Faktor, welches die RNA-Polymerase durch allosterische Effekte verdrängt und somit die Transkription beendet [Ciampi, 2006; Peters *et al.*, 2011]. Diese Art von Termination wird in der Regel als Antwort auf verschiedenste (Stress-) Bedingungen reguliert und stellt somit nicht immer ein Transkriptende dar [Ciampi, 2006]. Da weiterhin die ρ -abhängige Termination für *C. glutamicum* nicht untersucht ist, wird die Detektion der Transkriptenden in *C. glutamicum* nur für ρ -unabhängige bzw. intrinsische Terminatoren durchgeführt. Diese Terminatoren stellen ein Sequenzmotiv aus einer G+C-reichen- und Poly(T)-Sequenz auf dem nicht-transkribierten DNA-Strang dar, welches eine Haarnadelstruktur mit einem G+C-reichen Stamm gefolgt von einer Poly(U)-Sequenz ergibt, sobald dieses Motiv von der RNA-Polymerase in RNA transkribiert wurde [Gusarov & Nudler, 1999; Peters *et*

al., 2011]. Es wird angenommen, dass der RNA-Polymerase-Komplex nach der Transkription und der ausgebildeten Haarnadelstruktur durch Scherkräfte und/oder Konformationsänderungen destabilisiert wird, sodass das RNA-Molekül im Bereich der Poly(U)-Sequenz vom RNA-Polymerase-Komplex gelöst wird [Carafa *et al.*, 1990; Wilson & von Hippel, 1995; Peters *et al.*, 2011].

Da die intrinsischen, ρ -unabhängigen Terminatoren ein bestimmtes Struktur- (Haarnadelstruktur) sowie Sequenzmotiv aufweisen, ist es in einem Genom möglich, speziell hinter annotierten Genen, nach solchen Struktur- und Sequenzmotiven (Poly(A)-Sequenz, darauffolgend eine Haarnadelstruktur mit einem G+C-reichen Stamm gefolgt von einer Poly(T)-Sequenz) zu suchen und Vorhersagen aufzustellen. Für *C. glutamicum* liegt eine solche Vorhersage von *TransTermHP* mit 1.814 Terminatoren jeweils stromabwärts von aktuell annotierten Genen in *GenDB* vor [Meyer *et al.*, 2003; Kingsford *et al.*, 2007].

Zur Vervollständigung der Transkriptomanalyse von *C. glutamicum* wurden nun neben der Detektion von TSSs mittels des Datensatzes der nativen 5'-Transkriptenden sowie von Operon-Strukturen und neuen Transkripten mithilfe des Datensatzes der Vollängentranskripte, auf Basis der Vollängentranskript-Bibliothek die 3'-Transkriptenden untersucht und ferner mit vorhergesagten, ρ -unabhängigen Terminatoren [Kingsford *et al.*, 2007] verglichen. Dafür wurden alle detektierten monocistronischen Transkripte sowie die jeweils am Ende eines Primär-Operon liegenden Gene extrahiert und mit den 1.814 vorhergesagten Terminatoren abgeglichen. Insgesamt konnten somit 1.389 vorhergesagte, intrinsische Terminatoren für die am Ende eines Transkripts liegenden Gene erhalten und für eine Bestimmung der Transkriptenden verwendet werden. Ein Transkriptende wurde als ein deutlicher *Read*-Abfall mit den folgenden Kriterien definiert (Abbildung VIII-4):

- ($X > 10$) der Durchschnitt der *Read*-Abdeckung von den 5 Basen vor dem Terminator (X) muss > 10 sein und
- ($X/t_i \geq 5$) das Verhältnis vom Durchschnitt der *Read*-Abdeckung von den 5 Basen vor Terminator (X) zu der *Read*-Abdeckung t_i an der Position i innerhalb des vorhergesagten Terminators muss ≥ 5 sein

Weiterhin wurden die intrinsischen, vorhergesagten Terminatoren in vier Bereiche eingeteilt, in denen das Transkriptende bestimmt wurde: Stamm 1, Schleife, Stamm 2 und Poly(T)-Sequenz, wobei Stamm 1 und 2 hybridisieren und zur Ausbildung der Terminator-Haarnadelstruktur beitragen können (Abbildung IV-12).

Unter Zuhilfenahme dieser Kriterien konnten für den Datensatz der Vollängentranskripte nur 322 (~23,2 %) Transkriptenden im Bereich der vorhergesagten, ρ -unabhängigen Terminatoren bestimmt werden. Von den übrigen 1.067 Terminator-Bereichen zeigten 291 (~21 %) eine unzureichende *Read*-Abdeckung, um ein Transkriptende bestimmen zu können, und 776 (~55,9 %) wiesen keinen deutlichen *Read*-Abfall auf, sondern vielmehr eine graduelle *Read*-Abnahme. Interessanterweise konnte für die 322 Terminator-Bereiche mit den eindeutigen Transkriptenden festgestellt werden, dass der Hauptteil der Transkripte (232 von 322; 72 %) innerhalb des Stamms 1 der Terminator-Haarnadelstruktur endet.

Folglich endet nur ein geringer Anteil an Transkripten innerhalb der Haarnadelschleife (60 von 322; 18,6 %), des Stamms 2 (24 von 322; 7,5 %) und der Poly(T)-Sequenz (6 von 322; 1,9 %; Abbildung IV-12, Tabelle S9). Somit wird bei den Genen, bei denen das Transkript innerhalb der Terminator-Struktur endet, der Terminator nicht mittranskribiert, sodass die Haarnadelstruktur am Ende der mRNA nicht ausgebildet werden kann.

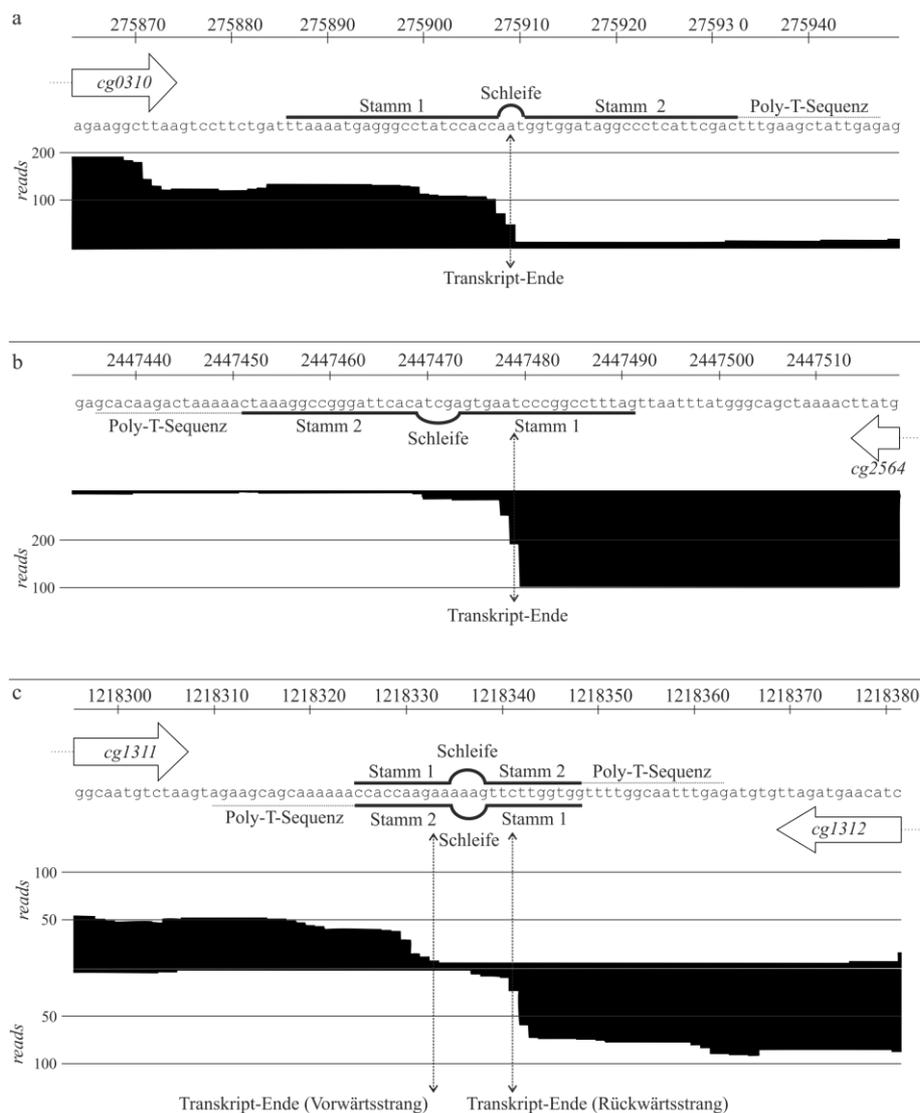


Abbildung IV-12: Beispiele von RNAseq-detektierten Transkriptenden im Vergleich zu ρ -unabhängigen Terminatoren

Die schwarz hinterlegten Kurven bzw. Stapel kennzeichnen kumulierte Reads, die der Bibliothek mit Vollängentranskripten entstammen. Auf der y-Achse ist die Read-Abdeckung und auf der x-Achse die Genom-Position aufgeführt. Proteinkodierende Gene sind durch große Pfeile mit der Gen-ID dargestellt. Die ρ -unabhängigen Terminatoren wurden von *TransTermHP* [Kingsford *et al.*, 2007] vorhergesagt.

a. Unidirektionaler Terminator von *cg0310* mit einem Read-Abfall innerhalb der Terminator-Schleife.

b. Unidirektionaler Terminator von *cg2564* mit einem Read-Abfall innerhalb des ersten Terminator-Stamms.

c. Bidirektionaler Terminator von *cg1311* und *cg1312* mit einem Read-Abfall jeweils innerhalb des ersten Terminator-Stamms.

V Diskussion

1 Die zwei etablierten RNAseq-Protokolle ermöglichen eine genomweite Analyse der transkriptionellen Organisation mit Einzelnukleotid-Auflösung

Die vollständige Genomsequenz von *C. glutamicum* und dessen Annotation ist seit 2003 publiziert [Ikeda & Nakagawa, 2003; Kalinowski *et al.*, 2003]. Die Kenntnis über die transkriptionelle Organisation, wie Promotor-Elemente und nicht-kodierende Transkripte, ist jedoch nur auf einige Beispiele beschränkt [Zemanová *et al.*, 2008; Pátek & Nešvera, 2011]. *C. glutamicum* stellt ein industriell bedeutsames Bakterium dar und ist gleichzeitig ein apathogener Modellorganismus der Unterordnung *Corynebacterineae*, speziell für die Gattungen *Mycobacterium*, *Nocardia* sowie *Corynebacterium* und ihre pathogenen Vertreter sowie andere Aminosäure-Produzenten [Wendisch *et al.*, 2006; Bott, 2007].

Diese Arbeit präsentiert die Anwendung zweier etablierter RNAseq-Protokolle und die entsprechende Analyse der daraus resultierenden Datensätze für *C. glutamicum*. Ferner wird verdeutlicht, dass eine umfangreiche Transkriptomanalyse mit hervorstechender Auflösung mithilfe dieser Protokolle ermöglicht wird.

Der Einsatz von RNAseq zur Transkriptomanalyse und -charakterisierung in Prokaryoten als auch in Eukaryoten, findet insbesondere seit jüngster Zeit Anwendung [Hillier *et al.*, 2009; Oliver *et al.*, 2009; Filichkin *et al.*, 2010; Jacob *et al.*, 2010; Sharma *et al.*, 2010; Frank *et al.*, 2011; Høvik *et al.*, 2012]. Während in den oben genannten Studien zur Erstellung von RNAseq-Bibliotheken für gewöhnlich kommerzielle Kits Verwendung finden, werden in dieser Arbeit zwei optimierte, transparente RNAseq-Vorbereitungsprotokolle vorgestellt, die es erlauben, in jeden Schritt einzugreifen, um diese ggf. anzupassen. Weiterhin lässt sich diese Vorgehensweise einfach für andere Spezies adaptieren, indem ggf. ein passendes RNA-Isolierungs-, rRNA-Abreicherungs- sowie PCR-Protokoll verwendet wird.

Für eine umfassende Transkriptomanalyse stellt diese Arbeit ein Protokoll zur Sequenzierung von Vollängentranskripten, welches die Gesamtheit aller Transkripte > 150 nt liefert, und ein Protokoll zur Sequenzierung nativer 5'-Transkriptenden, wobei spezifisch primäre Transkripte angereichert und deren 5'-Ende sequenziert werden, entwickelt. Die Validierung der entwickelten RNAseq-Protokolle erfolgte durch die Sequenzierung und anschließende Analyse des Transkriptoms von *C. glutamicum*. Für das Vollängentranskript-Protokoll wurde Gesamt-RNA aus *C. glutamicum*-Zellen isoliert, die unter 9 verschiedenen Bedingungen, vor allem Stressbedingungen, kultiviert wurden, um eine umfassende Expression und Repräsentation von möglichst vielen Transkripten zu erhalten. Für das Protokoll der

nativen 5'-Transkriptenden wurde *C. glutamicum* im CGXII-Minimalmedium und LB-Komplexmedium kultiviert, da die Analyse der Promotoren, die durch den *housekeeping* Sigmafaktor σ^A erkannt werden, im Vordergrund stand. Die anschließende Sequenzierung der erstellten Bibliotheken wurde mittels der *Illumina*-Plattform durchgeführt, was eine hohe Sequenziertiefe mit Einzelnukleotid-Auflösung erlaubte. Für die Beschreibung des Transkriptom-Profiles von *C. glutamicum* wurden die erhaltenen Analyse-Ergebnisse für das Primär- und Vollängentranskriptom kombiniert. Auf diese Weise war es nun möglich, zahlreiche 5'- und 3'-Enden von Transkripten zu beschreiben, neue, unbekannte Transkripte, Operon-Strukturen sowie 5'-UTRs zu detektieren, als auch umfassend Promotoren und RBS zu charakterisieren. Ferner konnten Starts von proteinkodierenden Genen korrigiert werden, weshalb sich die Transkriptom-Sequenzierung auch für eine Verbesserung von Genomannotationen eignet.

2 Der Datensatz der nativen 5'-Transkriptenden liefert eine enorme Anzahl an Transkriptionsstarts und ermöglicht die Identifizierung zugehöriger σ^A -abhängiger Promotoren in *C. glutamicum*

Mithilfe des Datensatzes für native 5'-Transkriptenden konnten unter Einhaltung bestimmter Kriterien (Anzahl an *Read*-Starts an einer Position muss über einem Hintergrund-Schwellenwert liegen und Verhältnis von *Read*-Starts an einer Position zu *Read*-Starts an vorheriger Position muss einen Schwellenwert übersteigen) 2.591 TSSs detektiert werden, wobei knapp über 80 % der TSSs annotierten Genen und die restlichen neuen Transkripten zugeordnet wurden [Pfeifer-Sancar *et al.*, 2013]. Weiterhin konnte gezeigt werden, dass 365 Gene mehr als nur einen TSS (2 – 6 TSSs) besitzen, was in multiplen 5'-UTRs mit unterschiedlicher Länge resultiert. Das Auftreten der multiplen TSSs ist eine Folge von mehrfach auftretenden Promotoren vor einem Gen, was allgemein eine flexible Genexpression ermöglicht [Schibler & Sierra, 1987]. Der Gebrauch von mehreren Promotoren kann sich in der umgebungsbedingten Adaption als sehr nützlich erweisen, indem beispielsweise ein Promotor für eine konstitutive Genexpression zuständig ist und die Übrigen eine induzierbare Transkription als Antwort auf bestimmte Stimuli gewährleisten [Ponnambalam *et al.*, 1986; Schibler & Sierra, 1987]. Außerdem können somit die mit unterschiedlicher Länge entstandenen 5'-UTRs eine differente mRNA-Stabilität zur Folge haben, indem eine mRNA eine stabilere Struktur ausbildet und die andere nicht [Belin *et al.*, 1987; Schibler & Sierra, 1987]. Ferner konnte gezeigt werden, dass die Benutzung von multiplen Promotoren eine unterschiedliche Translationseffizienz durch Vorkommen von Sekundärstrukturen, die die RBS sequestrieren oder nicht, zur Folge haben kann [McPheeters *et al.*, 1986; Schibler & Sierra, 1987]. Weiterhin können in den unterschiedlich langen 5'-UTRs andersartige Signalpeptide mitkodiert werden, die über die Bestimmungsorte in der Zelle (z.B. extrazellulär oder intrazellulär) eines Proteins entscheiden [Carlson & Botstein, 1982].

Im Anschluss an die TSSs-Detektion wurde stromaufwärts jedes TSSs *in-silico* nach wiederholenden Motiven gesucht. Bei dieser Suche konnten eine -10-Region mit dem konservierten Motiv „TAnnnT“ und eine -35-Region mit dem schwach-konservierten Motiv „ttgnca“ als σ^A -abhängige Promotormotive für *C. glutamicum* gefunden werden [Pfeifer-Sancar *et al.*, 2013]. Der Sigmafaktor σ^A stellt einen Haupt-Sigmafaktor dar, welcher in allen bekannten Bakterien präsent ist [Ishihama, 1988, 2000; Pátek & Nešvera, 2011]. In *C. glutamicum* besitzt der σ^A -Faktor eine Erkennungssequenz aus zwei 6 Basen langen Regionen, „TTGNCA“ und „TANANT“, die sich um die -35- und -10-Position relativ zum TSSs befinden [Pátek & Nešvera, 2011] und beide mit den gefundenen Motiven übereinstimmen. Diese Analyse zeigt somit einerseits die Validierung des RNAseq-Protokolls für native 5'-Transkriptenden sowie die Möglichkeit zur exakten, gleichzeitigen und ausführlichen Bestimmung von Tausenden TSSs

und Promotoren, die zuvor einzeln über 5'-RACE [Frohman, 1993] oder PEX [Thompson *et al.*, 1979] identifiziert werden mussten.

Bereits zuvor wurde für *C. glutamicum* beobachtet, dass die -35-Region der σ^A -abhängigen Promotoren schlecht konserviert vorliegt [Pátek *et al.*, 2003; Pátek & Nešvera, 2011]. Die schwache Konservierung des -35-Motivs lässt sich möglicherweise durch das Vorkommen eines erweiterten -10-Promotor-Elements erklären, da viele identifizierte -10-Regionen ein erweitertes Element besitzen. Für viele prokaryotische Promotoren konnte gezeigt werden, dass ein erweitertes Promotor-Element („TGn“ direkt vor der -10-Region) ein starkes Erkennungssignal für den Haupt-Sigmafaktor darstellt und eine fehlende -35-Region kompensieren kann [Seeburg *et al.*, 1977; Keilty & Rosenberg, 1987; Wösten, 1998]. Die Funktionsweise des erweiterten Elements in *C. glutamicum* muss allerdings noch näher untersucht werden.

Neben dem primären σ^A -Faktor besitzt *C. glutamicum* einen dem σ^A -ähnlichen, nicht essentiellen σ^B -Faktor, der vermutlich dieselben Promotoren wie σ^A erkennt, und fünf alternative ECF-Sigmafaktoren σ^C , σ^D , σ^E , σ^H und σ^M . Jeder einzelne ECF-Sigmafaktor erkennt dabei spezielle, deutlich von den σ^A/σ^B unterscheidbaren, Promotor-Sequenzen, dessen zugehörige Gene in spezifische Funktionen und Stress-Antworten involviert sind [Pátek & Nešvera, 2011]. Diese Sigmafaktoren sind also an der Transkriptionsaktivierung als Antwort auf Stress-Bedingungen beteiligt. Sieben unterschiedliche Stress-Bedingungen sind zwar in der Bibliothek für Vollängentranskripte abgedeckt, fehlen jedoch in der Bibliothek für native 5'-Transkriptenden. Folglich wurde hier die Suche auf σ^A -abhängige Promotormotive beschränkt. Eine Untersuchung von σ^H -abhängigen Promotoren in *C. glutamicum* mittels des hier etablierten Protokolls für native 5'-Transkriptenden wurde durchgeführt und lieferte 45 Promotormotive vor annotierten Genen [Busche, 2013] und 44 Promotormotive vor kleinen RNA-Genen [Mentz *et al.*, 2013; Mentz, 2013]. Promotoren, die von anderen Sigmafaktoren erkannt werden, müssen in weiteren Studien analysiert werden.

3 Die Transkriptomsequenzierung zeigt eine hohe Abundanz von *leaderless* mRNAs in *C. glutamicum*

Die Untersuchung bezüglich der 5'-UTR-Längen hat für *C. glutamicum* ergeben, dass etwa 33 % aller mRNAs, denen mindestens ein TSS zugeordnet werden konnte, *leaderless* sind und somit keine 5'-UTR besitzen [Pfeifer-Sancar *et al.*, 2013]. Zuvor haben Transkriptionsanalysen ergeben, dass in *H. pylori* ~1,5 %, 26 von 1.907, [Sharma *et al.*, 2010], für *Streptomyces coelicolor* ~28 %, 53 5'-UTRs < 11 nt von 192, [Vockenhuber *et al.*, 2011] und für *Pseudomonas putida* KT2440 ~4,7 %, 8 5'-UTRs < 11 nt von 170, [Frank *et al.*, 2011] *leaderless* mRNAs gefunden wurden. Somit ist der gefundene Anteil an *leaderless* mRNAs in *C. glutamicum* (707 von 2.147) der bisher höchst beschriebene Wert für Bakterien. *Leaderless* Transkripte sind in Eukaryoten, Archaeen und Prokaryoten bekannt [Wu & Janssen, 1996] und dabei als besonders abundant in Gram-positiven Bakterien beschrieben [Moll *et al.*, 2002]. Interessanterweise wurde kürzlich für die meisten Gene in Archaeen gezeigt, dass diese als *leaderless* Transkripte exprimiert werden [Wurtzel *et al.*, 2010; Kixmüller *et al.*, 2011], wobei es auch Archaeen gibt, die hauptsächlich lange 5'-UTRs besitzen [Jäger *et al.*, 2009]. Nichtsdestotrotz ist es anzuzweifeln, dass es Ähnlichkeiten zwischen der Transkriptionsmaschinerie in Archaeen und *C. glutamicum* gibt, da der Transkriptionsvorgang in Archaeen dem der Eukaryoten gleicht [Langer *et al.*, 1995].

Da Bakterien, Archaeen und Eukaryoten sequenzverwandte Translationsfaktoren besitzen und bakterielle *leaderless* mRNAs auf zuverlässige Weise in Archaeen und Eukaryoten translatiert werden können, lässt sich vermuten, dass vorkommende *leaderless* mRNAs alle einen Ursprung haben [Kyrpides & Woese, 1998; Moll *et al.*, 2002]. Die Gemeinsamkeit und Einfachheit der Translationsinitiation von *leaderless* mRNAs in Eukaryoten, Archaeen und Prokaryoten ist ein weiterer Hinweis darauf, dass dieser Translationsmechanismus einen frühen, evolutionären Ursprung hat [Malys & McCarthy, 2011].

Der Translationsmechanismus von *leaderless* mRNAs, die keine RBS tragen, und die Erkennung solcher Transkripte durch Ribosomen waren lange Zeit unbekannt. Für *E. coli* konnte dann allerdings gezeigt werden, dass das AUG-Startcodon einer *leaderless* mRNA und die passende Initiator-tRNA (fMet-tRNA) eine stabile Bindung zwischen der mRNA und dem Ribosom begünstigen kann [Moll *et al.*, 2002; Brock *et al.*, 2008]. Ferner konnte gezeigt werden, dass *leaderless* Transkripte im Gegensatz zu RBS-tragenden mRNAs, die zunächst 30S-Ribosomenuntereinheiten rekrutieren, eine größere Präferenz gegenüber 70S Ribosomen-Monomeren als zu 30S-Untereinheiten aufweisen und somit möglicherweise eine alternative Translationsinitiation für solche mRNAs existiert [Moll *et al.*, 2004]. Diese Arbeit zeigt jedoch auf, dass *leaderless* Transkripte in *C. glutamicum* AUG (~79 %) sowie GUG (~21 %) als Startcodons haben können [Pfeifer-Sancar *et al.*, 2013]. Obwohl es unklar ist, wie GUG als Startcodon bei *leaderless* mRNAs fungieren kann, konnte gezeigt werden, dass das GUG-Startcodon auch in anderen Organismen vorkommt und vermutlich eine geringere Translationseffizienz bietet [Moll *et al.*, 2002; Vockenhuber *et*

al., 2011]. Es sei erwähnt, dass die fMet-tRNA auch für die Translationsinitiation von nicht-*leaderless* mRNAs mit den alternativen Startcodons GUG, UUG und CUG verantwortlich ist (interessanterweise durch Wobble-Basen an der 1. Position) [Arora *et al.*, 2013]. Weshalb diese jedoch die Translation von *leaderless* mRNAs mit den Startcodons AUG und GUG, nicht aber UUG und CUG, initiiert, bleibt unklar [Moll *et al.*, 2002].

Um mögliche Hinweise auf zelluläre Funktionen der Gene von *leaderless* Transkripten zu erhalten, wurden diese in die funktionellen Gen-Kategorien nach *eggNOG* [Powell *et al.*, 2012] eingeteilt und mit der genomweiten Einteilung für die 3.039 annotierten, proteinkodierenden Gene in *C. glutamicum* verglichen. Für die detektierten *leaderless* mRNAs zeigte sich jedoch für keine Gen-Kategorie eine Präferenz, wie bereits für *E. coli* [Cho *et al.*, 2009] und *G. sulfurreducens* [Qiu *et al.*, 2010] beschrieben. Für *S. cerevisiae* wurde allerdings festgestellt, dass viele Gene mit kurzen bzw. ohne 5'-UTRs *housekeeping* Gene sind, die keine Regulation auf posttranskriptioneller Ebene benötigen [David *et al.*, 2006]. Da die *leaderless* mRNAs in *C. glutamicum* sowie in *E. coli* und *G. sulfurreducens* keinen bestimmten funktionellen Gen-Gruppen angehören, sondern über alle Gen-Gruppen vertreten sind, könnte man als evolutionären Ursprung eine Gen-Struktur ohne 5'-UTR vermuten.

4 Die Charakterisierung von 5'-untranslatierten Regionen gibt Hinweise auf neue regulatorische Elemente

Mit der Charakterisierung und Analyse von 5'-UTRs konnten 13 regulatorische 5'-UTRs, die von *Rfam* vorhergesagt wurden [Gardner *et al.*, 2011], verifiziert werden. Diese regulatorischen 5'-UTRs stellen überwiegend *Riboswitches* dar (insgesamt 10 von 13 detektierten regulatorischen 5'-UTRs), die u.a. vor *cg0083* (FMN-sensierender *Riboswitch*), *cg0825*, *cg1227*, *cg1476*, *cg1655*, *cg2236* (TPP-sensierende *Riboswitches*) und *cg1478* (SAM-sensierender *Riboswitch*) gefunden werden konnten [Pfeifer-Sancar *et al.*, 2013]. *Riboswitches* können Konformationsänderungen eingehen und somit die Genexpression direkt auf Transkript-Ebene kontrollieren. Diese Konformationsänderung geht mit der Bindung von kleinen Molekülen einher, die Produkte der zu regulierenden Stoffwechselwege sind und die chemische Eigenschaft zur Interaktion mit Nukleinsäuren besitzen [Mandal & Breaker, 2004; Nudler & Mironov, 2004; Winkler & Breaker, 2005]. *Riboswitches* in der 5'-UTR von Stoffwechsel-Genen sind für Gram-positive und Gram-negative Bakterien beschrieben und regulieren die Genexpression auf Transkriptionsebene (z.B. durch Ausbildung einer Terminator-Struktur, die die vorzeitige Transkriptionstermination auslöst, oder durch Spaltung des Transkripts als Folge eines aktivierenden Ribozyms) als auch auf Translationsebene (z.B. durch Sequestrieren oder Freigeben der RBS) [Winkler *et al.*, 2002; Mandal & Breaker, 2004; Nudler & Mironov, 2004; Vitreschak *et al.*, 2004].

Für *C. glutamicum* ist die Existenz der 10 detektierten *Riboswitches* tatsächlich sehr wahrscheinlich, da viele Gene, die ein *Riboswitch* innerhalb ihrer 5'-UTR tragen, als Enzyme mit metabolischer Funktion, die in Verbindung zur *Riboswitch*-Funktionsweise steht, annotiert sind. Andererseits kann das Vorkommen eines *Riboswitches* einen Hinweis auf die Gen-Funktion geben oder diese bestätigen.

Vor dem Gen *cg0083* konnte mittels der RNAseq-Daten ein FMN-sensierender *Riboswitch* gefunden werden, der vermutlich durch die Bindung von FMN und die darauffolgende Konformationsänderung die Transkription von *cg0083* terminiert oder die Translationsinitiation von *cg0083* blockiert [Winkler *et al.*, 2002]. Das Gen *cg0083* kodiert für einen putativen Mononukleotid-Transporter, wobei bereits gezeigt werden konnte, dass dieser Energie-unabhängig Riboflavin transportieren kann [Vogl *et al.*, 2007]. Riboflavin ist eine Vorstufe von FMN und FAD und kann in *C. glutamicum* neben der Aufnahme durch *Cg0083* durch die Produkte von *cg1797*, *cg1798* und *cg1799* aus Ribulose-5'-Phosphat synthetisiert werden. Für *C. glutamicum* ist zu vermuten, dass bei ausreichender FMN-Konzentration der Import von Riboflavin durch den FMN-sensierenden *Riboswitch* unterbunden wird. Die genaue Regulation der Riboflavin-Biosynthese im Zusammenhang mit der Funktionsweise des FMN-*Riboswitches* muss jedoch weiter untersucht werden.

Vor den Genen *cg1476* (*thiC*), *cg1655* (*thiM*) und *cg2236* (*thiE*) konnte ein TPP-sensierender *Riboswitch* gefunden werden, wobei die zugehörigen Gen-Produkte in die Biosynthese von Thiamin involviert sind. Auch hier lässt sich für *C. glutamicum* vermuten, dass bei ausreichender TPP-Konzentration die Synthese von TPP inaktiviert wird. Auch vor den Genen *cg0825* (kodiert für eine putative kurzkettige Oxidoreduktase) und *cg1227* (kodiert für eine Domäne eines ABC-Transporters) konnte ein TPP-*Riboswitch* gefunden werden. Die Annotation dieser Gene liefert keinen Zusammenhang zu der Thiamin-Biosynthese, das Vorkommen eines TPP-*Riboswitches* vor diesen Genen lässt jedoch eine Thiamin-abhängige Funktion der zugehörigen Produkte vermuten. Auch hier bleibt die genaue Funktionsweise des TPP-*Riboswitches* unklar und muss in weiteren Untersuchungen geklärt werden.

In der 5'-UTR von *cg1478* (kodiert für ein hypothetisches Protein) konnte ein SAM-IV-*Riboswitch* gefunden werden, wobei die Annotation keinen Hinweis auf einen Zusammenhang zum Methionin- oder Cystein-Metabolismus liefert. Allerdings konnte für *cg1478* festgestellt werden, dass dieses zusammen mit vielen Genen der Methionin-Biosynthese bei pH 6 induziert vorliegt [Follmann *et al.*, 2009]. Dieser Anhaltspunkt sowie der SAM-IV-sensierende *Riboswitch* lassen eine Verbindung zum Methionin- oder Cystein-Metabolismus annehmen.

Die 5'-UTR von *cg0936* und *cg2402* enthalten jeweils einen vorhergesagten *ydaO-yuaA-Leader*, der mit der Zellwand-Remodellierung und/oder osmotischem Schock im Zusammenhang steht. Das Produkt von *cg2402* (kodiert für ein sekretiertes Protein der NPL/P60-Familie) trägt die Funktion einer Zellwand-assoziierten Hydrolase inne und das Produkt von *cg0936* (*rpfl*; kodiert für einen putativen *Resuscitation-promoting* Faktor (RPF)) ist bei der Belebung inaktiver Zellen durch die Hydrolyse von Peptidoglykan beteiligt. So sind RPFs in *M. tuberculosis* u.a. für die Intaktheit der äußeren Membran verantwortlich [Wivagg & Hung, 2012]. Der Ligand von dem *ydaO-yuaA-Leader* ist unbekannt und auch die Funktionsweise und Regulation des *ydaO-yuaA-Leaders* muss für *C. glutamicum* durch weitere Analysen geklärt werden.

Ein *yybP-ykoY-Leader* wurde in der 5'-UTR von *cg2157* (*terC*), welches für ein Tellurium-Resistenz-Membran-Protein kodiert, detektiert und wird vor Genen für Tellurium-Resistenz, ATPasen und Membran-Proteinen gefunden [Barrick *et al.*, 2004]. Auch hier ist das sensierte Metabolit sowie die Funktionsweise und Regulation des *yybP-ykoY-Leaders* unklar.

Neben den Metabolit-sensierenden *Riboswitches* konnte in der 5'-UTR von *cg0215* (kodiert für ein Kälteschock-Protein) ein RNA-Thermometer mithilfe der RNAseq-Daten gefunden werden. RNA-Thermometer agieren wie *Riboswitches* über ihre Sekundärstruktur auf Translationsinitiationsebene, jedoch in Antwort auf Temperaturveränderungen [Narberhaus *et al.*, 2006]. In *E. coli* ist beschrieben, dass der *cspA*-Thermosensor die Translation der *cspA*-mRNA (kodiert ebenfalls für ein Kälteschock-Protein) bei niedrigen Temperaturen effizienter moduliert als bei höheren Temperaturen, als Anpassung der zellulären Mechanismen bei einer Kälte-Schock-Antwort [Giuliodori *et al.*, 2010]. Der vorhergesagte

und mittels RNAseq validierte RNA-Thermometer vor *cg0215* sowie das kodierte Kälte-Schock-Protein Cg0215 deuten auf eine ähnliche Funktion dieses RNA-Thermometers hin.

Auch die zwei gefundenen, regulatorischen 5'-UTRs (*mraW*-RNA-Motiv und *msiK*-RNA-Motiv), die weder als *Riboswitch* noch als RNA-Thermometer vorhergesagt sind, liegen vor Genen, deren Annotation einen Hinweis für die Existenz der regulatorischen Elemente liefert. Vor dem Gen *cg2377*, welches für eine SAM-abhängige Methyltransferase, die in die Zellwand Biogenese involviert ist, wurde das *mraW*-RNA-Motiv vorhergesagt. Dieses Motiv reguliert möglicherweise Gene der Peptidoglykan-Synthese, obwohl die Funktionsweise dieses Elements unklar bleibt [Weinberg *et al.*, 2010]. Das *msiK*-RNA-Motiv, welches vor *cg2708* (*msiK1*), das für einen putativen ABC-Zucker-Transporter kodiert, gefunden wurde, ist an der Regulation von Zucker-Transportern beteiligt, wobei auch hier die Funktionsweise unbekannt bleibt [Weinberg *et al.*, 2010].

Neben den 13 vorhergesagten, mittels der RNAseq-Daten validierten regulatorischen Elementen, konnten das mini-*ykkC*-RNA-Motiv vor *cg1231* und zwei *yybP-ykoY-Leader* vor *cg1660* und *cg2785* mit den erhaltenen RNAseq-Daten nicht bestätigt werden, da vor den genannten Genen kein TSS und folglich keine 5'-UTR identifiziert werden konnten. Nichtsdestotrotz lassen die zugehörigen Gen-Annotationen auch hier das Vorkommen dieser regulatorischen Elemente vermuten. Das Gen *cg1231* kodiert für einen putativen Antiporter für $\text{Na}^+/\text{Ca}^{2+}$, wobei das mini-*ykkC*-RNA-Motiv für die Regulation von Efflux-Pumpen und Detoxifikationssystemen verantwortlich ist [Barrick *et al.*, 2004; Weinberg *et al.*, 2007]. Auffällig ist jedoch, dass das vorhergesagte mini-*ykkC*-RNA-Motiv im 3'-Ende eines stromaufwärts liegenden Gens (*cg1230*, konserviertes hypothetisches Protein) vorkommt. Es ist also auch möglich, dass dieses Motiv eine Regulation auf *cg1230* vom zugehörigen 3'-Ende aus vollzieht. Die Gene *cg1660* und *cg2785* kodieren beide für putative Membran-Proteine, die beide jeweils einen *yybP-ykoY-Leader* in der 5'-UTR beinhalten, der für die Regulation von Membran-Proteinen bekannt ist [Barrick *et al.*, 2004].

Da die Annotation der Gene *cg1231* bzw. *cg1230*, *cg1660* und *cg2785* in Bezug auf die vorhergesagten regulatorischen Elemente schlüssig scheint, ist es denkbar, dass unter anderen Bedingungen TSSs detektiert und dadurch das Vorkommen dieser RNA-Elemente bestätigt werden könnten.

Eine weitere Evidenz für Existenz der 13 vorhergesagten, validierten regulatorischen Elemente wird durch die entsprechenden sekundären Strukturen gestützt, da diese die Regulation der zugehörigen Gene ermöglichen [Barrick *et al.*, 2004; Mandal & Breaker, 2004]. Für alle 13 validierten regulatorischen 5'-UTRs konnten stabile Haarnadelstrukturen festgestellt. Außerdem wird bei allen Strukturen, außer bei den 5'-UTRs von *cg0215* (RNA-Thermometer) und *cg2236/thiE* (TPP-*Riboswitch*), die RBS mindestens teilweise RBS sequestriert. Das RNA-Thermometer bildet hierbei eine Ausnahme, da eine Temperaturveränderung eine Konformationsänderung der Struktur und folglich die Sequestrierung der RBS mit sich führen kann. Die Sequestrierung der RBS innerhalb der Sekundärstruktur liefert ferner einen Hinweis auf eine Funktionsweise der regulatorischen Elemente auf Translationsinitiationsebene, anstatt auf Transkriptionsebene. Somit ist für den TPP-*Riboswitch* vor *cg2236/thiE* denkbar (bei dem die

RBS nicht sequestriert vorliegt), dass dieses während der Transkription und in Antwort auf bestimmte Stimuli interne Terminatorstrukturen ausbilden kann und eine vollständige Transkription von *cg2236/thiE* verhindert. Somit würde dieser TPP-*Riboswitch* eine Regulation auf Transkriptionsebene übernehmen. Diese Hypothesen zur Funktionsweise der regulatorischen Elemente muss jedoch durch weitere Untersuchungen belegt werden. Weiterhin bleibt bei diesen regulatorischen Elementen unklar, ob diese bei Interaktion mit spezifischen Liganden oder externen Stimuli die Genexpression aktivieren oder inaktivieren. Das Sequestrieren der RBS innerhalb der sekundären Strukturen kann lediglich einen Anhaltspunkt liefern. Somit kann vermutet werden, dass bei Interaktion mit Liganden oder externen Stimuli die regulatorischen 5'-UTRs, bei denen die RBS sequestriert vorliegt, die Genexpression aktivieren und bei nicht-sequestrierten RBS, die Genexpression inaktivieren.

Neben den vorhergesagten, validierten regulatorischen Elementen konnte eine Vielzahl an weiteren Kandidaten, die möglicherweise *cis*-regulatorische RNA-Elemente darstellen, identifiziert werden. So besitzen viele Gene, deren Produkte in den Aminosäure-Metabolismus involviert sind (z.B. Arginin, Cystein, Methionin, Histidin), 5'-UTRs > 60 nt. Zuvor wurden einige RNA-Elemente beschrieben, die insbesondere die Regulation von Genen des Aminosäure-Metabolismus in Bakterien beeinflussen [Mandal & Breaker, 2004; Nudler & Mironov, 2004; Seliverstov *et al.*, 2005]. Ein derartiges RNA-Element konnte beispielsweise vor *hisD* (*cg2305*) gefunden werden, welches in die Histidin-Biosynthese involviert ist [Kulis-Horn *et al.*, 2014]. Die entsprechende Sekundärstruktur enthält Bindestellen für das Anticodon (CAC für das Anticodon GUG) als auch für ungeladene 3'-Enden der Histidyl-tRNA (UGG für das 3'-Ende CCA) und übt möglicherweise eine Genregulation auf Translationsinitiationsebene aus, indem ungeladene Histidyl-tRNAs an die Sekundärstruktur binden, die RBS und folglich die Translation freigeben [Kulis-Horn *et al.*, 2014].

Solche *cis*-regulatorischen RNA-Elemente erlauben eine Fein Anpassung der Genexpression [Michalodimitrakis & Isalan, 2009], was einen energetischen Vorteil für ein schnell-wachsendes, metabolisch vielfältiges Bakterium wie *C. glutamicum* ausmachen kann.

Eine Alternative zu *Riboswitches* und RNA-Thermometer innerhalb von 5'-UTRs stellen stabile Haarnadelstrukturen dar. Solche Strukturen können insbesondere einen Schutz gegen die Degradation von mRNAs, z.B. durch RNasen, bieten [Emory *et al.*, 1992; Bricker & Belasco, 1999] und eine mögliche Erklärung für die relativ hohe Anzahl an langen 5'-UTRs in *C. glutamicum* sein.

5 Die umfassende Analyse der Transkriptomsequenzierung liefert das Motiv „AGGag“ als Ribosomenbindestelle für *C. glutamicum*

Mithilfe der vorliegenden RNAseq-Daten war es möglich, eine höchst umfassende und präzise Analyse der RBS durchzuführen. Die RBS ist auf der mRNA als Purin-reiche Region stromabwärts des TSS bekannt, die mit dem 3'-Ende der 16S rRNA interagiert, die Bindung des Ribosoms an die mRNA ermöglicht und somit an der Translationsinitiation beteiligt ist [Shine & Dalgarno, 1974; Ringquist *et al.*, 1992; McCarthy & Brimacombe, 1994].

Für die Analyse der RBS in *C. glutamicum* wurden detektierte singuläre 5'-UTRs (mehrfache 5'-UTRs pro Gen wurden auf eine reduziert) ≥ 20 nt von proteinkodierenden Genen verwendet. Der Abgleich von 20 Basen innerhalb der 5'-UTRs stromaufwärts des Startcodons zeigte ein erhöhtes Purin-Vorkommen ($> 55\%$) zwischen der 6. und 18. Base, was auf eine Erkennungsstelle des Ribosoms hinweist [Shine & Dalgarno, 1975]. Die Suche nach einem Motiv innerhalb dieser Purin-reichen Region lieferte für 89 % der durchsuchten 5'-UTR-Abschnitte das konservierte Motiv „AGGag“. Dieses Motiv stellt außerdem ein perfektes, reverses Komplement zum 3'-Ende der 16S rRNA von *C. glutamicum* dar. Diese Entdeckung korreliert auch mit den publizierten Daten für *E. coli* [Shine & Dalgarno, 1974, 1975; Ringquist *et al.*, 1992; McCarthy & Brimacombe, 1994]. Die Translationseffizienz hängt einerseits vom Startcodon und der ausgebildeten Sekundärstrukturen innerhalb der 5'-UTR ab, wird aber auch u.a. durch die Hybridisierung der RBS auf der mRNA mit dem 3'-Ende der 16S rRNA und somit durch die Konservierung der RBS-Sequenz bestimmt [Chapon, 1982; Ozbudak *et al.*, 2002]. Es konnte gezeigt werden, dass die Translationseffizienz durch die Benutzung einer optimalen RBS, die als perfektes, reverses Komplement zum 3'-Ende der 16S rRNA vorliegt, gesteigert werden kann [Chapon, 1982; Ozbudak *et al.*, 2002]. Für *C. glutamicum* lässt sich nun vermuten, dass die ersten drei besonders stark konservierten Basen „AGG“ des gefundenen RBS-Motivs „AGGag“ essentiell für den Translationsinitiationsmechanismus sind und die übrigen Basen die Translationseffizienz modulieren. Durch die Verwendung einer stärker oder schwächer konservierten RBS kann somit die Effizienz der Bindung des Ribosoms an die mRNA verstärkt oder vermindert werden. Weiterhin konnte gezeigt werden, dass die Distanz zwischen der RBS und des Initiationscodons ebenfalls eine starke Auswirkung auf die Translation besitzt, wenn diese < 4 oder > 12 Basen beträgt [Gold, 1988; Ringquist *et al.*, 1992]. Der physische Abstand zwischen dem hybridisierten 3'-Ende der 16S rRNA und dem Initiationscodon ist somit maßgebend, da dadurch die passende Positionierung der mRNA im Ribosom gewährleistet wird.

In *C. glutamicum* hat sich durch die Analyse der RBS gezeigt, dass eine Distanz zwischen der RBS und des Initiationscodons von 4 – 12 Basen bei 90 % der mRNAs, bei denen ein RBS-Motiv detektiert werden konnte, verwendet wird. Der berechnete mittlere Abstand von $7,7 \pm 2,7$ Basen ist mit dem mittleren Abstand bei *E. coli* von $6,9 \pm 2$ Basen vergleichbar [Stormo *et al.*, 1982]. Aufgrund dessen dient

wahrscheinlich eine Distanz zwischen der RBS und des Startcodons von 5 - 10 Basen in *C. glutamicum* als die für die Translation optimale Distanz.

6 Der Datensatz der Vollängentranskripte klärt die Operon-Strukturen für *C. glutamicum* auf

Bei diesem RNAseq-Ansatz konnten etwa 96 % der aktuell annotierten Gene unter mindestens einer der neun verwendeten Wachstumsbedingungen als transkribiert detektiert und zu Operon-Strukturen zugeordnet werden. Klassischerweise bilden Gene, die sich nebeneinander auf dem selben Strang befinden und gemeinsam, koordiniert von einem Promotor aus transkribiert werden, eine genetische Einheit, ein sogenanntes Operon [Jacob & Monod, 1961a, 1961b]. Für *C. glutamicum* lieferte die Analyse der Operon-Strukturen 648 Primär-Operons, die zwischen 2 und 16 Gene und weitere 935 monocistronische Transkripte, die jeweils nur ein Gen beinhalten. Somit zeigte sich, dass etwa $\frac{1}{3}$ der Gene in *C. glutamicum* als monocistronische Transkripte transkribiert werden und die übrigen $\frac{2}{3}$ Gene als Operons organisiert sind. Die relativ hohe Anzahl an monocistronischen Transkripten deutet an, dass es vermutlich vorteilhaft ist, einzelne Gene individuell zu transkribieren und folglich zu regulieren, z.B. durch Transkriptionsregulatoren. Weiterhin wird dadurch eine Feinanpassung in der Transkription eines einzelnen Gens erlaubt. Andererseits wächst mit der Zahl an Monocistrons der regulatorische Ballast, wobei genügend Regulatoren und Energie von der Zelle aufgebracht werden müssen. Man glaubte außerdem lange Zeit, dass bakterielle Gene, die einem Biosynthese-Weg angehören, oft als genetische Einheit organisiert sind und als Folge dasselbe Expressionslevel besitzen [Jacob & Monod, 1961a, 1961b]. Es konnte jedoch gezeigt werden, dass es aufgrund von Abbau der mRNA, internen Terminatoren und regulatorischen Elementen wie sRNAs und *Riboswitches* zur ungleichen Genexpression innerhalb eines Operons kommen kann [Murakawa *et al.*, 1991; Perkins *et al.*, 1996; McDaniel *et al.*, 2003]. Weiterhin konnten durch die Kombination der Daten für native 5'-Transkriptenden und Vollängentranskripte Operon-interne TSSs detektiert werden, die die 648 Primär-Operons in weitere 565 Sub-Operons aufteilen. Es ist naheliegend, dass auch solche internen TSSs für ein ungleiches Transkript-Level innerhalb eines Operons sorgen können. Die Existenz von zusätzlichen Sub-Operons wurde auch zuvor für verschiedene Bakterien beschrieben: so konnten z.B. durch die Kartierung der TSSs in *H. pylori* 337 primäre Operons, 127 Sub-Operons und 66 Monocistrons innerhalb der primären Operons gefunden werden [Sharma *et al.*, 2010], wobei für *Mycoplasma pneumoniae* 341 Operons detektiert werden konnten, die durch die Analyse von 173 unterschiedlichen Kulturbedingungen in weitere 447 alternative Operons unterteilt wurden [Güell *et al.*, 2009]. Die Verwendung von internen TSSs kann auf enorme Weise die Transkriptom-Komplexität und die regulatorische Kapazität erhöhen, ohne dabei das Genom zu vergrößern [Cho *et al.*, 2009; Qiu *et al.*, 2010; Sorek & Cossart, 2010]. Darüber hinaus konnten andere Transkriptomanalysen eine Stimulus-abhängige Modulation der Operon-Strukturen zeigen [Güell *et al.*, 2009]. Somit kann ein Gen, welches in einer Bedingung als Polycistron organisiert ist, unter anderen Bedingungen als Monocistron vorkommen

[Güell *et al.*, 2009; Koide *et al.*, 2009]. Diese Transkriptom-Flexibilität und das Umschalten zwischen Operon-Strukturen zeigt sich auch für *C. glutamicum*.

Weiterhin hat die Operon-Analyse gezeigt, dass Gene, die in die Synthese von Cofaktoren, in den Protein-Metabolismus oder in die Zellwand-Synthese involviert sind, in besonders langen Operons organisiert sind (≥ 10 Gene). Diese Feststellung lässt mutmaßen, dass insbesondere in solchen Fällen eine koordinierte Expression von Vorteil ist. Andererseits sollte erwähnt werden, dass lange Operon-Strukturen auch Gruppen von Genen beinhalten, die sich nicht nur auf eine zelluläre Funktion beschränken, sondern vielfältige Funktionen innehaben. Solche Operons mit vielfältigen Gen-Gruppen sind vermutlich evolutionsbedingt entstanden (z.B. durch horizontalen Gentransfer).

Allgemein lässt sich sagen, dass ein Operon kein, wie lange gedacht, statisches Element ist, sondern vielmehr dynamisch ist. Insbesondere interne TSSs und die Bildung von alternativen Sub-Operons ermöglichen eine auf äußere Einflüsse anpassungsfähige Transkription von benötigten Abschnitten eines Operons.

Die Aufklärung der Operon-Strukturen ermöglichte es auch intercistronische RBS zu analysieren. Intercistronische RBS befinden sich zwischen zwei Genen auf einem Operon und sind nicht Bestandteil von 5'-UTRs. Diese Analyse lieferte ein ähnliches RBS-Motiv, wie das innerhalb der 5'-UTR, welches jedoch um eine Base Richtung 3'-Ende der 16S rRNA verschoben ist („aAGga“ im Vergleich zu „AGGag“). Auch der mittlere Abstand der RBS zum Initiationscodon ist um etwa eine Base größer ($8,8 \pm 2,3$ Basen im Vergleich zu $7,7 \pm 2,7$ Basen). Diese Ergebnisse lassen vermuten, dass die Initiation der Translation intercistronischer Gene und Gene, die sich am Anfang eines Transkripts befinden, auf ähnliche Art und Weise ablaufen muss. Die Detektion eines intercistronischen RBS-Motivs ist ein Hinweis darauf, dass sich die Ribosomen jeweils vor einer Kodierregion (am Anfang und innerhalb einer mRNA) neu assoziieren und die Translation einleiten. Außerdem hängt vermutlich auch hier die unterschiedliche Konservierung der RBS mit der Bindungseffizienz des Ribosoms zusammen, wobei die stark konservierten Basen „AG“ möglicherweise essentiell sind und die anderen Basen das Translationslevel modulieren. Inwieweit diese Hypothese jedoch zutrifft, muss in weiteren Analysen festgestellt werden.

7 Die Detektion neuer, unbekannter Transkripte liefert eine erhebliche Anzahl von *antisense* RNAs

Mithilfe der RNAseq-Daten konnten 916 transkriptionell aktive, bisher unbekannte Regionen für *C. glutamicum* entdeckt werden. Davon sind 30 intergenisch, befinden sich also zwischen zwei annotierten Genen, 186 sind intragenisch und 700 konnten *antisense* zu anderen annotierten Genen oder neuen Transkripten gefunden werden. Intergenische Transkripte stellen höchstwahrscheinlich kodierende Gene dar, da für 29 der 30 intergenischen Transkripte ein potentieller offener Leserahmen und für 20 zumindest ein homologes Protein in einer anderen Spezies gefunden werden konnte. Weiterhin ist es möglich, dass die intergenischen Transkripte nicht-kodierende Gene abbilden und regulatorische Funktionen übernehmen. Kleine, nicht-kodierende RNAs sind in Bakterien weit verbreitet [Storz, 2002; Thomason & Storz, 2010] und sind auch zahlreich in *C. glutamicum* vorhanden [Mentz *et al.*, 2013]. Da für die Sequenzierung von kleinen RNAs eine spezielle Vorbereitung notwendig ist, ist die Sequenzierung und Analyse solcher RNAs der Bestandteil einer anderen Arbeit [Mentz, 2013]. In dieser Arbeit werden solche kleinen RNAs nicht behandelt, da die vorgestellten Protokolle Transkripte vornehmlich > 150 nt abdecken.

Es wurde eine recht hohe Anzahl (insgesamt 186) an intragenisch-lokalisierten Transkripten für *C. glutamicum* durch die Kartierung von TSSs festgestellt. Da solche Transkripte mit bereits annotierten Gen-Transkripten überlappen, ist es nicht möglich, die intragenischen Transkripte in ihrer vollen Länge zu erfassen. Weiterhin bleibt unklar, ob diese TSSs alternative, kürzere Proteine, neue proteinkodierende oder nicht-kodierende Gene ergeben. Solche intragenischen TSSs wurden ebenfalls für *H. pylori* [Sharma *et al.*, 2010] und *E. coli* [Cho *et al.*, 2009] beschrieben. Außerdem wurden solche Gen-internen TSSs in Viren und im Menschen gefunden, die ein weiteres, kürzeres Gen-Produkt ergeben [Löchelt *et al.*, 1993; Schoenfeld *et al.*, 1998; Denoeud *et al.*, 2007]. In Bakterien ist die Aufgabe solcher Gen-interner TSSs unklar, es kann aber vermutet werden, dass diese internen Transkripte die Varietät und den genetischen Informationsgehalt erhöhen, indem diese nur Stimulus-spezifisch vorkommen. Weiterhin ist es denkbar, dass diese Transkripte keine Funktion übernehmen, sondern vielmehr Produkte zufälliger Transkription sind.

Des Weiteren konnten mit den RNAseq-Daten 700 neue *antisense* Transkripte für *C. glutamicum* gefunden werden, wobei für 233 von diesen 700 ein TSS kartiert werden konnte. Da die *antisense* Transkripte mit den Daten der Vollängentranskripte identifiziert wurden (Stress-Bibliothek) und nicht mit den Daten der nativen 5'-Transkriptenden (ungestresste Bibliothek), ist es nicht auszuschließen, dass die übrigen TSS-losen Transkripte durch σ^A -unabhängige Transkription entstanden sind, die nur unter passenden Stress-Bedingungen detektiert werden kann. Diese *antisense* Transkripte sind vermutlich nicht-kodierende RNAs, die regulatorische Funktionen auf der Ebene der Genexpression übernehmen.

Die Transkriptom-Sequenzierung führte zu der Beschreibung von *antisense* Transkripten in verschiedenen Bakterien, Eukaryoten und Archaeen, welche vermutlich die gemeinsame Form der Regulation in allen Domänen darstellt [Sorek & Cossart, 2010; Thomason & Storz, 2010]. *Antisense* RNAs nehmen durch verschiedenste Mechanismen Einfluss auf die Genexpression als Antwort auf unterschiedliche Stimuli. Einerseits können diese in ihrer gesamten oder Teillänge an ein *sense* Transkript hybridisieren, dadurch (i) Strukturveränderungen herbeiführen, welche dann zur Termination oder Anti-Termination der Transkription des Ziel-Transkripts führen (Attenuation), (ii) Endo- oder Exoribonuklease-Schnittstellen in die Ziel-RNA einbringen oder blockieren (RNA-Abbau), oder (iii) einen Effekt auf die Translation des Ziel-Transkripts ausüben, indem die RBS sequestriert oder freigegeben wird (Beeinflussung der Translation) [Thomason & Storz, 2010]. Andererseits können *antisense* Transkripte die *sense* Transkription beeinflussen, ohne dabei direkt mit der Ziel-RNA zu interagieren, indem durch Polymerase-Kollisionen die Produktion der Ziel-RNA unterdrückt wird (transkriptionelle Interferenz) [Thomason & Storz, 2010]. Der Gebrauch von *antisense* RNAs in Bakterien zur Kontrolle der Genexpression erlaubt eine zu den Protein-Regulatoren zusätzliche und strenge Regulation von Ziel-RNAs [Thomason & Storz, 2010]. Dadurch wird außerdem eine Adaption der Genexpression an viele verschiedene Bedingungen ermöglicht, ohne Einfluss auf die Genom-Größe oder die Aufwendung zusätzlicher Energie zur Produktion von Protein-Regulatoren zu nehmen.

Kürzlich wurden für *C. glutamicum* zwei *antisense* RNAs beschrieben: Die *asaC*-RNA (detektiert mithilfe der vorliegenden RNAseq-Daten), die *antisense* zu *argC* liegt und durch transkriptionelle Interferenz die Transkription von *argC* beeinflusst [Haasner, 2013], und die *arnA*-RNA, die *antisense* zur 5'-UTR von *gntR2* (Regulator der GntR-Familie; reguliert den Gluconat-Katabolismus und die Glukose-Aufnahme [Frunzke *et al.*, 2008]) ist und durch σ^H -abhängige Transkription synthetisiert wird [Zemanová *et al.*, 2008]. Diese beschriebenen *antisense* Transkripte zeigen auf, dass diese in *C. glutamicum* eine wichtige Regulation im zentralen Metabolismus übernehmen und verdeutlichen die Komplexität der Genregulation.

Die Funktion und Regulation der hier detektierten *antisense* als auch intergenischen und intragenischen Transkripte bleiben jedoch unbekannt, diese stellen aber eine interessante Forschungsbasis für weitere Studien dar, um dessen biologische Rolle in der Zelle aufzuklären.

8 Die Auswertung des Vollängentranskript-Datensatzes ermöglicht eine genaue Untersuchung der Transkriptenden

Die ρ -abhängige und ρ -unabhängige Termination sind die prinzipiellen Mechanismen, durch die die Transkriptionseinheiten definiert bzw. abgeschlossen werden [Wilson & von Hippel, 1995]. ρ -abhängige Termination benötigt das ρ -Protein, welches für die regulierte und nicht-regulierte Transkriptionstermination zuständig ist [Ciampi, 2006; Peters *et al.*, 2009]. Intrinsische, ρ -unabhängige Terminatoren sind dagegen Sequenzabschnitte auf der DNA, wobei ein G+C-reicher Abschnitt gefolgt von einer Poly(T)-Sequenz wichtige Merkmale darstellen [Gusarov & Nudler, 1999]. Es wird vermutet, dass der RNA-Polymerase-Komplex nach der Transkription der Terminator-Sequenz auf der DNA und der Ausbildung einer Haarnadelstruktur mit einem G+C-reichen Stamm auf der RNA innerhalb der Poly(U)-Sequenz destabilisiert wird, sodass als Folge die Transkription beendet und das transkribierte RNA-Molekül entlassen wird [Carafa *et al.*, 1990; Wilson & von Hippel, 1995].

Die vorliegenden RNAseq-Daten ermöglichen durch die Basen-genaue Auflösung mithilfe der Vollängentranskript-Bibliothek die Bestimmung der 3'-Transkriptenden. Die Analyse der Daten hat im Gegenzug zu der Literatur-Lage ergeben, dass nur eine geringe Anzahl an Transkripten (1,9 %; 6 von 322) innerhalb der Poly(T)- bzw. Poly(U)-Sequenz enden. Die Mehrheit der detektierten 3'-Transkriptenden wurde im Bereich des G+C-reichen Stamm der Haarnadelschleife gefunden, was einen Widerspruch zu der vorliegenden Literatur bildet [Carafa *et al.*, 1990; Wilson & von Hippel, 1995]. Eine mögliche Erklärung für dieses Ergebnis könnten RNA-Prozessierungsstellen darstellen, die innerhalb des Haarnadel-Stamms entstehen und Angriffspunkte für RNasen darstellen. Das Genom von *C. glutamicum* kodiert für verschiedenste Endoribonukleasen: die RNase III (*cg2273*), RNase J (*cg2160*) und RNase E/G (*cg2579*), welche die Phosphodiesterbindung innerhalb des RNA-Moleküls hydrolytisch spalten [Martín *et al.*, 2003; Maeda & Wachi, 2012], wobei RNase III für die Spaltung von doppelsträngiger RNA verantwortlich ist [Kharrat *et al.*, 1995] und als Kandidat für die Spaltung der Haarnadelschleife relevant sein könnte. So konnte für *E. coli* gezeigt werden, dass ein Durchlesen des *int*-Gens (kodiert für ein Protein, das in die DNA-Rekombination involviert ist) eine Formation einer RNase III-Schnittstelle in der Terminator-Struktur nach sich zieht, um die Genexpression dieses Gens zu regulieren [Schmeissner *et al.*, 1984], was die Hypothese einer Ausbildung einer RNase III-Schnittstelle in der Haarnadelschleife des Terminators in *C. glutamicum* stützt. Somit könnten die gefundenen 3'-Transkriptenden innerhalb des Terminator-Stamms durch RNasen prozessierte Transkripte darstellen. Als weitere Erklärung für die detektierten 3'-Transkriptenden innerhalb des Haarnadelschleife-Stamms könnte die Ausbildung von kreuzförmigen Strukturen auf beiden DNA-Strängen innerhalb der Terminator-Struktur darstellen. Solch eine Struktur könnte für die Verdrängung des RNA-Polymerase-Komplexes verantwortlich sein und somit die Transkription terminieren. Obwohl *in vitro* Studien

belegen, dass eher A+T-reiche als G+C-reiche Sequenzen für die Ausbildung solcher kreuzförmiger Strukturen zuständig sind, konnte dieses *in vivo* nicht nachgewiesen werden [Schroth & Ho, 1995].

Weiterhin sei hier erwähnt, dass es nicht ausgeschlossen werden kann, dass die Nutzung des DNA-*loop*-Adapters mit den sieben freien Wobble-Basen am 3'-Ende keine optimale Hybridisierung im Bereich des Terminators bietet und somit keine genaue Auflösung der 3'-Transkriptenden erreicht. Um dies zu überprüfen, ist die Vorbereitung der RNA über andere Protokolle notwendig. Eine geeignete Methode ist für eukaryotische Poly(A)-RNA beschrieben: das RNA-PET (*RNA-paired end tagging*), bei dem die 5'- und 3'-Transkriptenden fusioniert und anschließend sequenziert werden [Ruan & Ruan, 2012].

Als eine wichtige Erkenntnis zählt hier aber, dass für etwa 80 % der vorhergesagten, ρ -unabhängigen Terminatoren in *C. glutamicum*, aufgrund unzureichender *Read*-Abdeckung (291 von 1389) oder aufgrund einer nur graduellen *Read*-Abnahme (776 von 1389), kein 3'-Transkriptende bestimmt werden konnte. Dieses Ergebnis lässt einerseits durch die kontinuierliche Neusynthese der RNA erklären, andererseits liefert es ein starkes Indiz für einen sehr schnellen Abbau der RNA durch 3'-5'-Exoribonukleasen. *C. glutamicum* besitzt einige Exoribonukleasen, jedoch katalysieren die meisten spezifische Reaktionen wie die tRNA-Prozessierung und –Reifung (RNase D und RNase PH; [Ishii *et al.*, 2003; Zuo *et al.*, 2005]) oder den Abbau von sehr kurzen (2 – 5 nt) RNA-Fragmenten (OrnA; [Ghosh & Deutscher, 1999]). Allerdings kodiert das Genom von *C. glutamicum* für eine putative Exoribonuklease (*cg2453*), die in einen schnellen RNA-Abbau involviert sein könnte. Jedoch ist dieses Enzym nicht charakterisiert und um weitere Belege für solch eine Aktivität von Cg2453 zu erhalten, bedarf es weiterer Untersuchungen.

Die hohe Anzahl an graduell auslaufenden 3'-Transkriptenden sowie die Transkriptabbrüche innerhalb des G+C-reichen Stamms des Terminators sind ein Hinweis darauf, dass die Prozesse, die in die ρ -unabhängige Termination in *C. glutamicum* bzw. allgemein in Aktinobakterien involviert sind, noch nicht vollständig aufgeklärt sind.

VI Ausblick

In dieser Arbeit konnten zwei RNAseq-Protokolle für die Sequenzierung nativer 5'-Transkriptenden und Vollängentranskripte entwickelt und erfolgreich auf das Transkriptom von *C. glutamicum* angewandt werden. Die so erhaltenen und analysierten RNAseq-Daten präsentieren ein umfassendes Transkriptom-Profil für diesen industriell relevanten Modellorganismus. Die generierten Transkriptom-Daten sind strangspezifisch, besitzen eine Einzelnukleotid-Auflösung und erlauben einen Einblick in die transkriptionelle Organisation.

Die vorliegenden Daten erlaubten eine nahezu vollständige Detektion der σ^A -abhängigen Promotormotive. Neben dem primären Sigmafaktor σ^A besitzt *C. glutamicum* einen dem σ^A -ähnlichen Sigmafaktor σ^B und fünf alternative ECF-Sigmafaktoren (σ^C , σ^D , σ^E , σ^H und σ^M), die jeweils spezifische Promotor-Sequenzen als Antwort auf Stress-Bedingungen erkennen. Diese Promotormotive sind bisher nur wenig untersucht und stellen eine wichtige Erforschungsgrundlage dar. Eine erste Analyse von σ^H -abhängigen Promotoren in *C. glutamicum* mittels des hier etablierten Protokolls für native 5'-Transkriptenden wurde durchgeführt und lieferte 45 σ^H -Promotormotive [Busche, 2013]. Das entwickelte RNAseq-Protokoll eignet sich somit auch für die Untersuchung weiterer Sigmafaktoren.

Die Untersuchung der 5'-UTR-Längen hat für *C. glutamicum* ergeben, dass etwa 33 % aller mRNAs als *leaderless* mRNAs transkribiert werden. Diese überraschend hohe Anzahl liefert einen Nachweis für die Wichtigkeit der Erforschung der Translationsinitiation solcher Transkripte, da die Translation die direkte Verbindung zum herstellenden Gen-Produkt darstellt. Der Translationsmechanismus von *leaderless* mRNAs in *C. glutamicum* ist bisher unbekannt, die erhaltenen RNAseq-Daten liefern allerdings einen Hinweis darauf, dass das AUG- sowie GUG-Startcodon und die Initiator-tRNA fMet-tRNA daran beteiligt sind. Dieses sollte jedoch durch weitere Analysen bestätigt werden.

Durch die Charakterisierung und Analyse der 5'-UTRs konnten 13 vorhergesagte, regulatorische 5'UTRs (überwiegend *Riboswitches*) verifiziert werden. Diese regulatorischen 5'-UTRs sind in der Lage, die Genexpression durch direkte Sensierung eines Stimulus (z.B. Metabolit, Temperaturveränderung) und ohne Einfluss von Protein-Regulatoren, zu regulieren. Da solche 5'-UTR-Elemente eine schnelle, effiziente und energiesparende Genregulation erlauben, eignet sich deren Einsatz beispielsweise in der Generierung von Produktionsstämmen. Weiterhin wurden zahlreiche 5'-UTRs > 100 nt, auch vor Genen, dessen Produkte in den Aminosäure-Metabolismus involviert sind, gefunden. Diese stellen ebenfalls geeignete Kandidaten für regulatorische 5'-UTRs dar. Es ist also sinnvoll, diese und die vorhergesagten 5'-UTR-Kandidaten und die zugehörigen sensierten Stimuli experimentell zu bestimmen bzw. zu bestätigen sowie deren Regulations- und Wirkmechanismus aufzuklären bzw. ebenfalls zu verifizieren.

Die umfassende Analyse der RBS innerhalb der detektierten 5'-UTRs lieferte das konservierte Motiv „AGGag“, welches ein perfektes, reverses Komplement zum 3'-Ende der 16S rRNA von *C. glutamicum* zeigt. Da die Translationseffizienz von der Hybridisierung der RBS auf der mRNA mit dem 3'-Ende der 16S rRNA abhängt, kann somit durch eine veränderte RBS-Sequenz die Translation beabsichtigt bestimmt werden, was sich z.B. ebenfalls in der Generierung von Produktionsstämmen als sehr hilfreich erweisen kann. In Bezug darauf könnte es von Interesse sein, den Zusammenhang zwischen der RBS-Sequenz und der Translationseffizienz zu untersuchen.

Die Analyse der Operon-Strukturen zeigte für *C. glutamicum* eine Organisation von 648 Primär-Operons, die zwischen 2 und 16 Gene und 935 monocistronische Transkripte, die jeweils nur ein Gen beinhalten. Weiterhin stellte sich heraus, dass ein Operon kein, wie lange gedacht, statisches Element ist, sondern vielmehr dynamisch ist, da zahlreiche alternative Sub-Operons gefunden werden konnten, die sich aus den längeren Primär-Operons ergeben. Solch eine Bildung von alternativen Sub-Operons ermöglicht eine auf äußere Einflüsse anpassungsfähige Transkription von benötigten Abschnitten eines Operons. Somit wird hier nur ein Ausschnitt von möglichen Operon-Strukturen für *C. glutamicum* abgebildet, sodass sich Operon-Strukturen unter verschiedenen Bedingungen unterschiedlich verhalten. Die Bestimmung von Operon-Strukturen unter unterschiedlichen Bedingungen kann einerseits eine weitere Aufklärung der genetischen Information in *C. glutamicum* liefern und sich andererseits für eine gewollte, koordinierte Genexpression als wertvoll erweisen.

Die erlangten RNAseq-Daten lieferten 916 transkriptionell aktive, bisher unbekannte Regionen für *C. glutamicum*, wobei sich 30 zwischen zwei annotierten Genen, 186 intragenisch und 700 *antisense* zu anderen annotierten Genen oder neuen Transkripten befinden. Da insbesondere *antisense* RNAs durch verschiedenste Mechanismen in Antwort auf unterschiedliche Stimuli Einfluss auf die Genexpression nehmen, stellen diese ein interessantes Forschungsfeld dar, wie das Beispiel der *asaC*-RNA zeigt: Diese liegt *antisense* zu *argC* und beeinflusst durch transkriptionelle Interferenz die Transkription von *argC* (involviert in die Argininbiosynthese). Die neuen, transkriptionell aktiven Regionen sind also geeignete Kandidaten zur Aufklärung derer Funktions- und Regulationsweise.

Weiterhin konnte gezeigt werden, dass die Mehrheit der Transkripte in *C. glutamicum* kein klares 3'-Transkriptende, sondern ein graduelles Transkript-Auslaufen, besitzen und nur ein geringer Anteil der 3'-Transkriptenden innerhalb der Poly(T)- bzw. Poly(U)-Sequenz oder innerhalb der Terminator-Struktur enden. Diese hohe Anzahl an 3'-Transkriptenden, die ein solches End-Muster aufweisen, dass die Prozesse, die in die ρ -unabhängige Termination in *C. glutamicum* noch nicht vollständig aufgeklärt sind. Hier sind also weitere Analysen nötig, um die erhaltenen 3'-Transkriptenden zu validieren und die ρ -unabhängige Termination zu untersuchen, da unklar bleibt, welche Rolle die vorhergesagten intrinsischen Terminatoren in *C. glutamicum* spielen.

Somit zeigt sich beispielhaft für *C. glutamicum*, dass die hier entwickelten Protokolle zur Transkriptom-Sequenzierung letztendlich informationsreiche Daten für eine vollständige Genomannotation liefern und

eine interessante Anwendung für weitere Organismen bieten. Die daraus erhaltenen Daten bringen vielfältige, genomweite Informationen, die viele weitere Forschungsmöglichkeiten auf Transkriptions- sowie auf Translationsebene darstellen. Neben der qualitativen Beschreibung von Transkripten und Transkriptelementen eröffnet insbesondere das Protokoll der Vollängentranskripte die Möglichkeit zur quantitativen Messung von bekannten, aber auch unbekanntem Transkripten mit herausragender Einzelnukleotid-Auflösung.

VII Literatur

- ALTSCHUL, S.F., GISH, W., MILLER, W., MYERS, E.W. & LIPMAN, D.J. (1990)** Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**(3): 403–410.
- AO, W., GAUDET, J., KENT, W.J., MUTTUMU, S. & MANGO, S.E. (2004)** Environmentally Induced Foregut Remodeling by PHA-4/FoxA and DAF-12/NHR. *Science* **305**(5691): 1743–1746.
- ARORA, S., BHAMIDIMARRI, S.P., BHATTACHARYYA, M., GOVINDAN, A., WEBER, M.H.W., VISHVESHWARA, S. & VARSHNEY, U. (2013)** Distinctive Contributions of the Ribosomal P-Site Elements m2G966, m5C967 and the C-Terminal Tail of the S9 Protein in the Fidelity of Initiation of Translation in *Escherichia coli*. *Nucleic Acids Research* **41**(9): 4963–4975.
- BABU, M.M. & TEICHMANN, S.A. (2003)** Evolution of Transcription Factors and the Gene Regulatory Network in *Escherichia coli*. *Nucleic Acids Research* **31**(4): 1234–1244.
- BADGER, J.H. & OLSEN, G.J. (1999)** CRITICA: Coding Region Identification Tool Invoking Comparative Analysis. *Molecular Biology and Evolution* **16**(4): 512–524.
- BAIRD, G.J. & FONTAINE, M.C. (2007)** *Corynebacterium pseudotuberculosis* and its Role in Ovine Caseous Lymphadenitis. *Journal of Comparative Pathology* **137**(4): 179–210.
- BALDRIAN, P. & LÓPEZ-MONDÉJAR, R. (2014)** Microbial Genomics, Transcriptomics and Proteomics: New Discoveries in Decomposition Research Using Complementary Methods. *Applied Microbiology and Biotechnology* **98**(4): 1531–1537.
- BANSAL-MUTALIK, R. & NIKAIDO, H. (2011)** Quantitative Lipid Composition of Cell Envelopes of *Corynebacterium glutamicum* Elucidated Through Reverse Micelle Extraction. *Proceedings of the National Academy of Sciences of the United States of America* **108**(37): 15360–15365.
- BARKSDALE, L. (1970)** *Corynebacterium diphtheriae* and Its Relatives¹. *Bacteriological Reviews* **34**(4): 378–422.
- BARRICK, J.E., CORBINO, K.A., WINKLER, W.C., NAHVI, A., MANDAL, M., COLLINS, J., LEE, M., ROTH, A., SUDARSAN, N., JONA, I., WICKISER, J.K. & BREAKER, R.R. (2004)** New RNA Motifs Suggest an Expanded Scope for Riboswitches in Bacterial Genetic Control. *Proceedings of the National Academy of Sciences of the United States of America* **101**(17): 6421–6426.
- BELIN, D., MUDD, E.A., PRENTKI, P., YI-YI, Y. & KRISCH, H.M. (1987)** Sense and Antisense Transcription of Bacteriophage T4 Gene 32 - Processing and Stability of the mRNAs. *Journal of Molecular Biology* **194**(2): 231–243.
- BERGEY, D.H., HARRISON, F.C., BREED, R.S., HAMMER, B.W. & HUNTOON, F.M. (1925)** Bergey's Manual Of Determinative Bacteriology. *The Williams & Wilkins Company*, Baltimore.
- BERTANI, G. (1951)** Studies on Lysogenesis - I. The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *Journal of Bacteriology* **62**(3): 293–300.
- 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**(10): 1351–1358.

- BORUKHOV, S., LEE, J. & LAPTENKO, O. (2005)** Bacterial Transcription Elongation Factors: New Insights Into Molecular Mechanism of Action. *Molecular Microbiology* **55**(5): 1315–1324.
- BOTT, M. (2007)** Offering Surprises: TCA Cycle Regulation in *Corynebacterium glutamicum*. *Trends in Microbiology* **15**(9): 417–425.
- BREED, R.S., BERGEY, D.H., HITCHENS, A.P. & MURRAY, E.G.D. (1948)** Bergey's Manual Of Determinative Bacteriology. *The Williams & Wilkins Company*, Baltimore.
- BRENNAN, N.M., BROWN, R., GOODFELLOW, M., WARD, A.C., BERESFORD, T.P., SIMPSON, P.J., FOX, P.F. & COGAN, T.M. (2001)** *Corynebacterium mooreparkense* sp. nov. and *Corynebacterium casei* sp. nov., Isolated from the Surface of a Smear-Ripened Cheese. *International Journal of Systematic and Evolutionary Microbiology* **51**(Pt 3): 843–852.
- BRICKER, A.L. & BELASCO, J.G. (1999)** Importance of a 5' Stem-Loop for Longevity of *papA* mRNA in *Escherichia coli*. *Journal of Bacteriology* **181**(11): 3587–3590.
- BRINKROLF, K., BRUNE, I. & TAUCH, A. (2007)** The Transcriptional Regulatory Network of the Amino Acid Producer *Corynebacterium glutamicum*. *Journal of Biotechnology* **129**(2): 191–211.
- BROCK, J.E., POURSHAHIAN, S., GILIBERTI, J., LIMBACH, P.A. & JANSSEN, G.R. (2008)** Ribosomes Bind Leaderless mRNA in *Escherichia coli* Through Recognition of Their 5'-Terminal AUG. *RNA* **14**(10): 2159–2169.
- BROWN, K.L. & HUGHES, K.T. (1995)** The Role of Anti-Sigma Factors in Gene Regulation. *Molecular Microbiology* **16**(3): 397–404.
- BROWNING, D.F. & BUSBY, S.J.W. (2004)** The Regulation of Bacterial Transcription Initiation. *Nature Reviews Microbiology* **2**(1): 57–65.
- BURKOVSKI, A. (2008)** *Corynebacteria: Genomics and Molecular Biology*. *Caister Academic Press*, Norfolk.
- BUSBY, S. & EBRIGHT, R.H. (1994)** Promoter Structure, Promoter Recognition, and Transcription Activation in Prokaryotes. *Cell* **79**(5): 743–746.
- BUSCHE, T. (2013)** Analyse von Regulationsnetzwerken der Extracytoplasmic Function (ECF)-Sigmafaktoren in *Corynebacterium glutamicum*. *Universität Bielefeld*.
- CALABRESE, V.G.M., ELSAVAGE, R.E., BISSONNETTE, G.K. & SEXSTONE, A.J. (1993)** Mobility and Enhanced Biodegradation of a Dilute Waste Oil Emulsion During Land Treatment. *Journal of Industrial Microbiology* **12**(1): 13–20.
- CARAFI, Y., D'AUBENTON, BRODY, E. & THERMES, C. (1990)** Prediction of Rho-Independent *Escherichia coli* Transcription Terminators: A Statistical Analysis of their RNA Stem-Loop Structures. *Journal of Molecular Biology* **216**(4): 835–858.
- CARLSON, M. & BOTSTEIN, D. (1982)** Two Differentially Regulated mRNAs with Different 5'-Ends Encode Secreted with Intracellular Forms of Yeast Invertase. *Cell* **28**(1): 145–154.
- CASTRO, T.L.P., SEYFFERT, N., RAMOS, R.T.J., BARBOSA, S., CARVALHO, R.D.O., PINTO, A.C., CARNEIRO, A.R., SILVA, W.M., PACHECO, L.G.C., DOWNSON, C., SCHNEIDER, M.P.C., MIYOSHI, A., AZEVEDO, V. & SILVA, A. (2013)** Ion Torrent-Based Transcriptional Assessment of a *Corynebacterium pseudotuberculosis equi* Strain Reveals Denaturing High-Performance

- Liquid Chromatography a Promising rRNA Depletion Method. *Microbial Biotechnology* **6**(2): 168–177.
- CHAPON, C. (1982) Expression of *maltT*, the Regulator Gene of the Maltose Region in *Escherichia coli*, Is Limited Both at Transcription and Translation. *The EMBO Journal* **1**(3): 369–374.
- CHAUDHURI, R.R., YU, L., KANJI, A., PERKINS, T.T., GARDNER, P.P., CHOUDHARY, J., MASKELL, D.J. & GRANT, A.J. (2011) Quantitative RNA-Seq Analysis of the *Campylobacter jejuni* Transcriptome. *Microbiology* **157**(Pt 10): 2922–2932.
- CHEN, H.-H., LI, W.-J., TANG, S.-K., KROPPESTEDT, R.M., STACKEBRANDT, E., XU, L.-H. & JIANG, C.-L. (2004) *Corynebacterium halotolerans* sp. nov., Isolated from Saline Soil in the West of China. *International Journal of Systematic and Evolutionary Microbiology* **54**(Pt 3): 779–782.
- CHO, B.-K., ZENGLER, K., QIU, Y., PARK, Y.S., KNIGHT, E.M., BARRETT, C.L., GAO, Y. & PALSSON, B.Ø. (2009) The Transcription Unit Architecture of the *Escherichia coli* Genome. *Nature Biotechnology* **27**(11): 1043–1049.
- CHOMCZYNSKI, P. & SACCHI, N. (1987) Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. *Analytical Biochemistry* **162**(1): 156–159.
- CIAMPI, M.S. (2006) Rho-Dependent Terminators and Transcription Termination. *Microbiology* **152**(9): 2515–2528.
- CLOONAN, N., FORREST, A.R.R., KOLLE, G., GARDINER, B.B.A., FAULKNER, G.J., BROWN, M.K., TAYLOR, D.F., STEPTOE, A.L., WANI, S., BETHEL, G., ROBERTSON, A.J., PERKINS, A.C., BRUCE, S.J., LEE, C.C., RANADE, S.S., PECKHAM, H.E., MANNING, J.M., MCKERNAN, K.J. & GRIMMOND, S.M. (2008) Stem Cell Transcriptome Profiling via Massive-Scale mRNA Sequencing. *Nature Methods* **5**(7): 613–619.
- COLLINS, M.D., BURTON, R.A. & JONES, D. (1988) *Corynebacterium amycolatum* sp. nov. A New Mycolic Acid-Less *Corynebacterium* Species from Human Skin. *FEMS Microbiology Letters* **49**(3): 349–352.
- COLLINS, M.D., FALSEN, E., AKERVALL, E., SJÖDEN, B. & ALVAREZ, A. (1998) *Corynebacterium kroppenstedtii* sp. nov., a Novel *Corynebacterium* that Does Not Contain Mycolic Acids. *International Journal of Systematic Bacteriology* **48 Pt 4**: 1449–1454.
- COLLINS, M.D., HOYLES, L., FOSTER, G. & FALSEN, E. (2004) *Corynebacterium caspium* sp. nov., From a Caspian Seal (*Phoca caspica*). *International Journal of Systematic and Evolutionary Microbiology* **54**(Pt 3): 925–928.
- CONSTANTINIDES, A. (1980) Steroid Transformation at High Substrate Concentrations Using Immobilized *Corynebacterium simplex* Cells. *Biotechnology and Bioengineering* **22**(1): 119–136.
- CROOKS, G.E., HON, G., CHANDONIA, J.-M. & BRENNER, S.E. (2004) WebLogo: A Sequence Logo Generator. *Genome Research* **14**(6): 1188–1190.
- DAVID, L., HUBER, W., GRANOVSKAIA, M., TOEDLING, J., PALM, C.J., BOFKIN, L., JONES, T., DAVIS, R.W. & STEINMETZ, L.M. (2006) A High-Resolution Map of Transcription in the Yeast Genome. *Proceedings of the National Academy of Sciences of the United States of America* **103**(14): 5320–5325.
- DECRÉAU, R.A., MARSON, C.M., SMITH, K.E. & BEHAN, J.M. (2003) Production of Malodorous Steroids from Androsta-5,16-Dienes and Androsta-4,16-Dienes by *Corynebacteria* and Other

- Human Axillary Bacteria. *The Journal of Steroid Biochemistry and Molecular Biology* **87**(4–5): 327–336.
- DELCHER, A.L., HARMON, D., KASIF, S., WHITE, O. & SALZBERG, S.L. (1999)** Improved Microbial Gene Identification with GLIMMER. *Nucleic Acids Research* **27**(23): 4636–4641.
- DENG, Q., RAMSKÖLD, D., REINIUS, B. & SANDBERG, R. (2014)** Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. *Science* **343**(6167): 193–196.
- DENOEUDE, F., AURY, J.-M., DA SILVA, C., NOEL, B., ROGIER, O., DELLEDONNE, M., MORGANTE, M., VALLE, G., WINCKER, P., SCARPELLI, C., JAILLON, O. & ARTIGUENAVE, F. (2008)** Annotating Genomes with Massive-Scale RNA Sequencing. *Genome Biology* **9**(12): R175.
- DENOEUDE, F., KAPRANOV, P., UCLA, C., FRANKISH, A., CASTELO, R., DRENKOW, J., LAGARDE, J., ALIOTO, T., MANZANO, C., CHRAST, J., DIKE, S., WYSS, C., HENRICHSEN, C.N., HOLROYD, N., DICKSON, M.C., TAYLOR, R., HANCE, Z., FOISSAC, S., MYERS, R.M., ROGERS, J., HUBBARD, T., HARROW, J., GUIGÓ, R., GINGERAS, T.R., ANTONARAKIS, S.E. & REYMOND, A. (2007)** Prominent Use of Distal 5' Transcription Start Sites and Discovery of a Large Number of Additional Exons in ENCODE Regions. *Genome Research* **17**(6): 746–759.
- DONDRUP, M., ALBAUM, S.P., GRIEBEL, T., HENCKEL, K., JÜNEMANN, S., KAHLKE, T., KLEINDT, C.K., KÜSTER, H., LINKE, B., MERTENS, D., MITTARD-RUNTE, V., NEUWEGER, H., RUNTE, K.J., 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.
- DONDRUP, M., HÜSER, A.T., MERTENS, D. & GOESMANN, A. (2009)** An Evaluation Framework for Statistical Tests on Microarray Data. *Journal of Biotechnology* **140**(1-2): 18–26.
- DORNENBURG, J.E., DEVITA, A.M., PALUMBO, M.J. & WADE, J.T. (2010)** Widespread Antisense Transcription in *Escherichia coli*. *mBio* **1**(1): e00024–10.
- EBRIGHT, R.H. (2000)** RNA Polymerase: Structural Similarities Between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. *Journal of Molecular Biology* **304**(5): 687–698.
- EGGELING, L. & BOTT, M. [eds] (2005)** Handbook of *Corynebacterium glutamicum*. CRC Press, Boca Raton.
- EMORY, S.A., BOUVET, P. & BELASCO, J.G. (1992)** A 5'-Terminal Stem-Loop Structure Can Stabilize mRNA in *Escherichia coli*. *Genes & Development* **6**(1): 135–148.
- EPSHTEIN, V. & NUDLER, E. (2003)** Cooperation Between RNA Polymerase Molecules in Transcription Elongation. *Science* **300**(5620): 801–805.
- ESCHERICH, T. (1894)** Aetiologie und Pathogenese der Epidemischen Diphtherie. Alfred Hölder, K.U.K. Hof- und Universitätsbuchhändler, Wien.
- EUZÉBY, J.P. (1997)** List of Bacterial Names with Standing in Nomenclature: a Folder Available on the Internet. *International Journal of Systematic Bacteriology* **47**(2): 590–592.
- FERNÁNDEZ-GARAYZÁBAL, J.F., VELA, A.I., EGIDO, R., HUTSON, R.A., LANZAROT, M.P., FERNÁNDEZ-GARCÍA, M. & COLLINS, M.D. (2004)** *Corynebacterium ciconiae* sp. nov., Isolated from the Trachea of Black Storks (*Ciconia nigra*). *International Journal of Systematic and Evolutionary Microbiology* **54**(Pt 6): 2191–2195.

- FILICHKIN, S.A., PRIEST, H.D., GIVAN, S.A., SHEN, R., BRYANT, D.W., FOX, S.E., WONG, W.-K. & MOCKLER, T.C. (2010) Genome-Wide Mapping of Alternative Splicing in *Arabidopsis thaliana*. *Genome Research* **20**(1): 45–58.
- FOLLMANN, M., OCHROMBEL, I., KRÄMER, R., TRÖTSCHEL, C., POETSCH, A., RÜCKERT, C., HÜSER, A., PERSICKE, M., SEIFERLING, D., KALINOWSKI, J. & MARIN, K. (2009) Functional Genomics of pH Homeostasis in *Corynebacterium glutamicum* Revealed Novel Links Between pH Response, Oxidative stress, Iron Homeostasis and Methionine Synthesis. *BMC Genomics* **10**: 621.
- FORCONI, M. & HERSCHLAG, D. (2009) Metal Ion-Based RNA Cleavage as a Structural Probe. *Methods in Enzymology* **468**: 91–106.
- FRANK, S., KLOCKGETHER, J., HAGENDORF, P., GEFFERS, R., SCHÖCK, U., POHL, T., DAVENPORT, C.F. & TÜMLER, B. (2011) *Pseudomonas putida* KT2440 Genome Update by cDNA Sequencing and Microarray Transcriptomics. *Environmental Microbiology* **13**(5): 1309–1326.
- FRÖHLICH, K.S. & VOGEL, J. (2009) Activation of Gene Expression by Small RNA. *Current Opinion in Microbiology* **12**(6): 674–682.
- FROHMAN, M.A. (1993) Rapid Amplification of Complementary DNA Ends for Generation of Full-Length Complementary DNAs: Thermal RACE. *Methods in Enzymology* **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. *Molecular Microbiology* **67**(2): 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. *International Journal of Systematic and Evolutionary Microbiology* **52**(4): 1127–1131.
- GARDNER, P.P., DAUB, J., TATE, J., MOORE, B.L., OSUCH, I.H., GRIFFITHS-JONES, S., FINN, R.D., NAWROCKI, E.P., KOLBE, D.L., EDDY, S.R. & BATEMAN, A. (2011) Rfam: Wikipedia, Clans and the ‘Decimal’ Release. *Nucleic Acids Research* **39**(Database Issue): D141–145.
- GHOSH, S. & DEUTSCHER, M.P. (1999) Oligoribonuclease Is an Essential Component of the mRNA Decay Pathway. *Proceedings of the National Academy of Sciences* **96**(8): 4372–4377.
- GIULIODORI, A.M., DI PIETRO, F., MARZI, S., MASQUIDA, B., WAGNER, R., ROMBY, P., GUALERZI, C.O. & PON, C.L. (2010) The *cspA* mRNA is a Thermosensor that Modulates Translation of the Cold-Shock Protein CspA. *Molecular Cell* **37**(1): 21–33.
- GOLD, L. (1988) Posttranscriptional Regulatory Mechanisms in *Escherichia coli*. *Annual Review of Biochemistry* **57**: 199–233.
- GOPINATH, V., MURALI, A., DHAR, K.S. & NAMPOOTHIRI, K.M. (2012) *Corynebacterium glutamicum* as a Potent Biocatalyst for the Bioconversion of Pentose Sugars to Value-Added Products. *Applied Microbiology and Biotechnology* **93**(1): 95–106.
- GRUBER, A.R., LORENZ, R., BERNHART, S.H., NEUBÖCK, R. & HOFACKER, I.L. (2008) The Vienna RNA Websuite. *Nucleic Acids Research* **36**: W70–74.

- GÜELL, M., VAN NOORT, V., YUS, E., CHEN, W.-H., LEIGH-BELL, J., MICHALODIMITRAKIS, K., YAMADA, T., ARUMUGAM, M., DOERKS, T., KÜHNER, S., RODE, M., SUYAMA, M., SCHMIDT, S., GAVIN, A.-C., BORK, P. & SERRANO, L. (2009) Transcriptome Complexity in a Genome-Reduced Bacterium. *Science* **326**(5957): 1268–1271.
- GÜELL, M., YUS, E., LLUCH-SENAR, M. & SERRANO, L. (2011) Bacterial Transcriptomics: What Is Beyond the RNA Horiz-ome? *Nature Reviews Microbiology* **9**(9): 658–669.
- GUSAROV, I. & NUDLER, E. (1999) The Mechanism of Intrinsic Transcription Termination. *Molecular Cell* **3**(4): 495–504.
- GUSAROV, I. & NUDLER, E. (2001) Control of Intrinsic Transcription Termination by N and NusA: The Basic Mechanisms. *Cell* **107**(4): 437–449.
- HAASNER, K. (2013) Untersuchung des Argininmetabolismus und dessen Regulation in *Corynebacterium glutamicum* ATCC 13032. *Universität Bielefeld*.
- HALL, V., COLLINS, M.D., HUTSON, R.A., LAWSON, P.A., FALSEN, E. & DUERDEN, B.I. (2003) *Corynebacterium atypicum* sp. nov., from a Human Clinical Source, Does not Contain Corynomycolic Acids. *International Journal of Systematic and Evolutionary Microbiology* **53**(Pt 4): 1065–1068.
- HASHIMOTO, T., HOON, M.J.L. DE, GRIMMOND, S.M., DAUB, C.O., HAYASHIZAKI, Y. & FAULKNER, G.J. (2009) Probabilistic Resolution of Multi-Mapping Reads in Massively Parallel Sequencing Data Using MuMRescueLite. *Bioinformatics* **25**(19): 2613–2614.
- HAWLEY, D.K. & MCCLURE, W.R. (1983) Compilation and Analysis of *Escherichia coli* Promoter DNA Sequences. *Nucleic Acids Research* **11**(8): 2237–2255.
- HERMANN, T. (2003) Industrial Production of Amino Acids by Coryneform Bacteria. *Journal of Biotechnology* **104**(1-3): 155–172.
- HILLIER, L.W., REINKE, V., GREEN, P., HIRST, M., MARRA, M.A. & WATERSTON, R.H. (2009) Massively Parallel Sequencing of the Polyadenylated Transcriptome of *C. elegans*. *Genome Research* **19**(4): 657–666.
- HIRSCHMAN, J., WONG, P.K., SEI, K., KEENER, J. & KUSTU, S. (1985) Products of Nitrogen Regulatory Genes *ntxA* and *ntxC* of Enteric Bacteria Activate *glNA* Transcription *in Vitro*: Evidence That the *ntxA* Product is a Sigma Factor. *Proceedings of the National Academy of Sciences* **82**(22): 7525–7529.
- HOURLY, W.A., FRISHMAN, D., ECKERSKORN, C., LOTTSPEICH, F. & HARTL, F.U. (1999) Identification of *in Vivo* Substrates of the Chaperonin GroEL. *Nature* **402**(6758): 147–154.
- HØVIK, H., YU, W.-H., OLSEN, I. & CHEN, T. (2012) Comprehensive Transcriptome Analysis of the Periodontopathogenic Bacterium *Porphyromonas gingivalis* W83. *Journal of Bacteriology* **194**(1): 100–114.
- HÜSER, A.T., BECKER, A., BRUNE, I., DONDRUP, M., KALINOWSKI, J., PLASSMEIER, J., PÜHLER, A., WIEGRÄBE, 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. *Journal of Biotechnology* **106**(2-3): 269–286.

- IKEDA, K. (2002) New Seasonings - Translated (and shortened to 75%) by Yoko Ogiwara and Yuzo Ninomiya from Journal of the Chemical Society of Tokyo, No. 30, 820–836 (1909). *Chemical Senses* **27**(9): 847–849.
- IKEDA, M. & NAKAGAWA, S. (2003) The *Corynebacterium glutamicum* Genome: Features and Impacts on Biotechnological Processes. *Applied Microbiology and Biotechnology* **62**(2-3): 99–109.
- IKEDA, M. & TAKENO, S. (2013) Amino Acid Production by *Corynebacterium glutamicum*. In *Corynebacterium glutamicum - Biology and Biotechnology* [eds H. YUKAWA & M. INUI] pp. 107–147. Springer Berlin Heidelberg, Berlin, Heidelberg.
- INUI, M. & YUKAWA, H. [eds] (2013) *Corynebacterium glutamicum - Biology and Biotechnology*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- ISHIHAMA, A. (1980) Subunit of Assembly of *Escherichia coli* RNA Polymerase. *Advances in Biophysics* **14**: 1–35.
- ISHIHAMA, A. (1988) Promoter Selectivity of Prokaryotic RNA Polymerases. *Trends in Genetics* **4**(10): 282–286.
- ISHIHAMA, A. (2000) Functional Modulation of *Escherichia coli* RNA Polymerase. *Annual Review of Microbiology* **54**: 499–518.
- ISHII, R., NUREKI, O. & YOKOYAMA, S. (2003) Crystal Structure of the tRNA Processing Enzyme RNase PH From *Aquifex aeolicus*. *The Journal of Biological Chemistry* **278**(34): 32397–32404.
- JACKMAN, P.J.H., PITCHER, D.G., PELCZYNSKA, S. & BORMAN, P. (1987) Classification of *Corynebacteria* Associated with Endocarditis (Group JK) as *Corynebacterium jeikeium* sp. nov. *Systematic and Applied Microbiology* **9**(1–2): 83–90.
- JACOB, F. (1966) Genetics of the Bacterial Cell. *Science* **152**(3728): 1470–1478.
- JACOB, F. & MONOD, J. (1961a) On the Regulation of Gene Activity. *Cold Spring Harbor Symposia on Quantitative Biology* **26**: 193–211.
- JACOB, F. & MONOD, J. (1961b) Genetic Regulatory Mechanisms in the Synthesis of Proteins. *Journal of Molecular Biology* **3**: 318–356.
- JACOB, N.M., KANTARDJEFF, A., YUSUFI, F.N.K., RETZEL, E.F., MULUKUTLA, B.C., CHUAH, S.H., YAP, M. & HU, W.-S. (2010) Reaching the Depth of the Chinese Hamster Ovary Cell Transcriptome. *Biotechnology and Bioengineering* **105**(5): 1002–1009.
- JÄGER, D., SHARMA, C.M., THOMSEN, J., EHLERS, C., VOGEL, J. & SCHMITZ, R.A. (2009) Deep Sequencing Analysis of the *Methanosarcina mazei* Gö1 Transcriptome in Response to Nitrogen Availability. *Proceedings of the National Academy of Sciences of the United States of America* **106**(51): 21878–21882.
- KALINOWSKI, J., BATHE, B., BARTELS, D., BISCHOFF, N., BOTT, M., BURKOVSKI, A., DUSCH, N., EGGELING, L., EIKMANN, B.J., GAIGALAT, L., GOESMANN, A., HARTMANN, M., HUTHMACHER, K., KRÄMER, R., LINKE, B., MCHARDY, A.C., MEYER, F., MÖCKEL, B., PFEFFERLE, W., PÜHLER, A., REY, D.A., RÜCKERT, C., RUPP, O., SAHM, H., WENDISCH, V.F., WIEGRÄBE, I. & 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. *Journal of Biotechnology* **104**(1-3): 5–25.

- KEILHAUER, C., EGGELING, L. & SAHM, H. (1993) Isoleucine Synthesis in *Corynebacterium glutamicum*: Molecular Analysis of the *ilvB-ilvN-ilvC* Operon. *Journal of Bacteriology* **175**(17): 5595–5603.
- KEILTY, S. & ROSENBERG, M. (1987) Constitutive Function of a Positively Regulated Promoter Reveals New Sequences Essential for Activity. *The Journal of Biological Chemistry* **262**(13).
- KESSLER, N., NEUWEGER, H., BONTE, A., LANGENKÄMPER, G., NIEHAUS, K., NATTKEMPER, T.W. & GOESMANN, A. (2013) MeltDB 2.0-Advances of the Metabolomics Software System. *Bioinformatics* **29**(19): 2452–2459.
- KHARRAT, A., MACIAS, M.J., GIBSON, T.J., NILGES, M. & PASTORE, A. (1995) Structure of the dsRNA Binding Domain of *E. coli* RNase III. *The EMBO Journal* **14**(14): 3572–3584.
- KINGSFORD, C.L., AYANBULE, K. & SALZBERG, S.L. (2007) Rapid, Accurate, Computational Discovery of Rho-Independent Transcription Terminators Illuminates their Relationship to DNA Uptake. *Genome Biology* **8**(2): R22.
- KINOSHITA, S. (2005) A Short History of the Birth of the Amino Acid Industry in Japan. In *Handbook of Corynebacterium glutamicum* [eds L. EGGELING & M. BOTT] pp. 3–5. CRC Press, Boca Raton.
- KINOSHITA, S., NAKAYAMA, K. & AKITA, S. (1958) Taxonomical Study of Glutamic Acid Accumulating Bacteria, *Micrococcus glutamicus* nov. sp. *Bulletin of the Chemical Society of Japan* **22**(3): 176–185.
- KIXMÜLLER, D., STRAHL, H., WENDE, A. & GREIE, J.-C. (2011) Archaeal Transcriptional Regulation of the Prokaryotic KdpFABC Complex Mediating K(+) Uptake in *H. salinarum*. *Extremophiles Life Under Extreme Conditions* **15**(6): 643–652.
- KOIDE, T., REISS, D.J., BARE, J.C., PANG, W.L., FACCIOTTI, M.T., SCHMID, A.K., PAN, M., MARZOLF, B., VAN, P.T., LO, F.-Y., PRATAP, A., DEUTSCH, E.W., PETERSON, A., MARTIN, D. & BALIGA, N.S. (2009) Prevalence of Transcription Promoters within Archaeal Operons and Coding Sequences. *Molecular Systems Biology* **5**: 285.
- KOZAK, M. (1999) Initiation of Translation in Prokaryotes and Eukaryotes. *Gene* **234**(2): 187–208.
- KULIS-HORN, R.K., PERSICKE, M. & KALINOWSKI, J. (2014) Histidine Biosynthesis, its Regulation and Biotechnological Application in *Corynebacterium glutamicum*. *Microbial Biotechnology* **7**(1): 5–25.
- KYRPIDES, N.C. & WOESE, C.R. (1998) Universally Conserved Translation Initiation Factors. *Proceedings of the National Academy of Sciences of the United States of America* **95**(1): 224–228.
- LANGER, D., HAIN, J., THURIAUX, P. & ZILLIG, W. (1995) Transcription in Archaea: Similarity to that in Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **92**(13): 5768–5772.
- LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S.L. (2009) Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biology* **10**(3): R25.
- LARISCH, C., NAKUNST, D., HÜSER, A.T., 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**(1): 4.

- LASA, I., TOLEDO-ARANA, A., DOBIN, A., VILLANUEVA, M., MOZOS, I.R. DE LOS, VERGARA-IRIGARAY, M., SEGURA, V., FAGEGALTIER, D., PENADÉS, J.R., VALLE, J., SOLANO, C. & GINGERAS, T.R. (2011) Genome-Wide Antisense Transcription Drives mRNA Processing in Bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **108**(50): 20172–20177.
- LAWRENCE, J.G. & ROTH, J.R. (1996) Selfish Operons: Horizontal Transfer May Drive the Evolution of Gene Clusters. *Genetics* **143**(4): 1843–1860.
- LECHEVALIER, M.P. & LECHEVALIER, H. (1970) Chemical Composition as a Criterion in the Classification of Aerobic Actinomycetes. *International Journal of Systematic Bacteriology* **20**(4): 435–443.
- LEHMANN, K.B. & NEUMANN, R.O. (1896) Atlas und Grundriss der Bakteriologie und Lehrbuch der speziellen bakteriologischen Diagnostik. *J.F. Lehmann*, München.
- LETEK, M., FIUZA, M., ORDÓÑEZ, E., VILLADANGOS, A.F., RAMOS, A., MATEOS, L.M. & GIL, J.A. (2008) Cell Growth and Cell Division in the Rod-Shaped Actinomycete *Corynebacterium glutamicum*. *Antonie van Leeuwenhoek* **94**(1): 99–109.
- LEWIS, P.J., DOHERTY, G.P. & CLARKE, J. (2008) Transcription Factor Dynamics. *Microbiology* **154**(7): 1837–1844.
- LISTER, R., O'MALLEY, R.C., TONTI-FILIPPINI, J., GREGORY, B.D., BERRY, C.C., MILLAR, A.H. & ECKER, J.R. (2008) Highly Integrated Single-Base Resolution Maps of the Epigenome in *Arabidopsis*. *Cell* **133**(3): 523–536.
- LIU, J., BARRY, C.E., BESRA, G.S. & NIKAIDO, H. (1996) Mycolic Acid Structure Determines the Fluidity of the Mycobacterial Cell Wall. *Journal of Biological Chemistry* **271**(47): 29545–29551.
- LÖCHELT, M., MURANYI, W. & FLÜGEL, R.M. (1993) Human Foamy Virus Genome Possesses an Internal, Bel-1-Dependent and Functional Promoter. *Proceedings of the National Academy of Sciences of the United States of America* **90**(15): 7317–7321.
- MAEDA, H., FUJITA, N. & ISHIHAMA, A. (2000) Competition Among Seven *Escherichia coli* σ Subunits: Relative Binding Affinities to the Core RNA Polymerase. *Nucleic Acids Research* **28**(18): 3497–3503.
- MAEDA, T. & WACHI, M. (2012) 3'-Untranslated Region-Dependent Degradation of the *aceA* mRNA, Encoding the Glyoxylate Cycle Enzyme Isocitrate Lyase, by RNase E/G in *Corynebacterium glutamicum*. *Applied and Environmental Microbiology* **78**(24): 8753–8761.
- MALYS, N. & MCCARTHY, J.E.G. (2011) Translation Initiation: Variations in the Mechanism Can Be Anticipated. *Cellular and Molecular Life Sciences* **68**(6): 991–1003.
- MANDAL, M. & BREAKER, R.R. (2004) Gene Regulation by Riboswitches. *Nature Reviews Molecular Cell Biology* **5**(6): 451–463.
- MARTÍN, J.F., BARREIRO, C., GONZÁLEZ-LAVADO, E. & BARRIUSO, M. (2003) Ribosomal RNA and Ribosomal Proteins in *Corynebacteria*. *Journal of Biotechnology* **104**(1–3): 41–53.
- MATEOS, L.M., ORDÓÑEZ, E., LETEK, M. & GIL, J.A. (2010) *Corynebacterium glutamicum* as a Model Bacterium for the Bioremediation of Arsenic. *International Microbiology* **9**(3): 207–215.

- MATHEW, R. & CHATTERJI, D. (2006) The Evolving Story of the Omega Subunit of Bacterial RNA Polymerase. *Trends in Microbiology* **14**(10): 450–455.
- MCCARTHY, J.E. & BRIMACOMBE, R. (1994) Prokaryotic Translation: The Interactive Pathway Leading to Initiation. *Trends in Genetics* **10**(11): 402–407.
- MCDANIEL, B.A.M., GRUNDY, F.J., ARTSIMOVITCH, I. & HENKIN, T.M. (2003) Transcription Termination Control of the S Box System: Direct Measurement of S-Adenosylmethionine by the Leader RNA. *Proceedings of the National Academy of Sciences of the United States of America* **100**(6): 3083–3088.
- MCGRATH, P.T., LEE, H., ZHANG, L., INIESTA, A.A., HOTTES, A.K., TAN, M.H., HILLSON, N.J., HU, P., SHAPIRO, L. & MCADAMS, H.H. (2007) High-Throughput Identification of Transcription Start Sites, Conserved Promoter Motifs and Predicted Regulons. *Nature Biotechnology* **25**(5): 584–592.
- MCHARDY, A.C., GOESMANN, A., PÜHLER, A. & MEYER, F. (2004) Development of Joint Application Strategies for Two Microbial Gene Finders. *Bioinformatics* **20**(10): 1622–1631.
- MCPHEETERS, D.S., CHRISTENSEN, A., YOUNG, E.T., STORMO, G. & GOLD, L. (1986) Translational Regulation of Expression of the Bacteriophage T4 Lysozyme Gene. *Nucleic Acids Research* **14**(14): 5813–5826.
- MENTZ, A. (2013) Identifizierung und Funktionsanalysen von kleinen RNAs in *Corynebacterium glutamicum*. Dissertation, *Universität Bielefeld*, Bielefeld.
- 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**(1): 714.
- MEYER, F., GOESMANN, A., MCHARDY, A.C., BARTELS, D., BEKEL, T., CLAUSEN, J., KALINOWSKI, J., LINKE, B., RUPP, O., GIEGERICH, R. & PÜHLER, A. (2003) GenDB--An Open Source Genome Annotation System for Prokaryote Genomes. *Nucleic Acids Research* **31**(8): 2187–2195.
- MICHALODIMITRAKIS, K. & ISALAN, M. (2009) Engineering Prokaryotic Gene Circuits. *FEMS Microbiology Reviews* **33**(1): 27–37.
- MIGULA, W. (1900) System der Bakterien. *Gustav Fischer*, Jena.
- MITSUHASHI, S. (2014) Current Topics in the Biotechnological Production of Essential Amino Acids, Functional Amino Acids, and Dipeptides. *Current Opinion in Biotechnology* **26**: 38–44.
- MOLL, I., GRILL, S., GUALERZI, C.O. & BLÄSI, U. (2002) Leaderless mRNAs in Bacteria: Surprises in Ribosomal Recruitment and Translational Control. *Molecular Microbiology* **43**(1): 239–246.
- MOLL, I., HIROKAWA, G., KIEL, M.C., KAJI, A. & BLÄSI, U. (2004) Translation Initiation with 70S Ribosomes: An Alternative Pathway for Leaderless mRNAs. *Nucleic Acids Research* **32**(11): 3354–3363.
- MUKHERJEE, K., NAGAI, H., SHIMAMOTO, N. & CHATTERJI, D. (1999) GroEL is Involved in Activation of *Escherichia coli* RNA Polymerase Devoid of the ω Subunit *in Vivo*. *European Journal of Biochemistry* **266**(1): 228–235.
- MURAKAWA, G.J., KWAN, C., YAMASHITA, J. & NIERLICH, D.P. (1991) Transcription and Decay of the Lac Messenger: Role of an Intergenic Terminator. *Journal of Bacteriology* **173**(1): 28–36.

- NAGALAKSHMI, U., WANG, Z., WAERN, K., SHOU, C., RAHA, D., GERSTEIN, M. & SNYDER, M. (2008) The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science* **320**(5881): 1344–1349.
- NAKAMURA, Y., NISHIO, Y., IKEO, K. & GOJOBORI, T. (2003) The Genome Stability in *Corynebacterium* Species Due to Lack of the Recombinational Repair System. *Gene* **317**(1-2): 149–155.
- NARBERHAUS, F., WALDMINGHAUS, T. & CHOWDHURY, S. (2006) RNA Thermometers. *FEMS Microbiology Reviews* **30**(1): 3–16.
- NORMARK, S., BERGSTROM, S., EDLUND, T., GRUNDSTROM, T., JAURIN, B., LINDBERG, F.P. & OLSSON, O. (1983) Overlapping Genes. *Annual Review of Genetics* **17**(1): 499–525.
- NUDLER, E. & MIRONOV, A.S. (2004) The Riboswitch Control of Bacterial Metabolism. *Trends in Biochemical Sciences* **29**(1): 11–17.
- OKONIEWSKI, M.J. & MILLER, C.J. (2006) Hybridization Interactions Between Probesets in Short Oligo Microarrays Lead to Spurious Correlations. *BMC Bioinformatics* **7**(1): 276.
- OLIVER, H.F., ORSI, R.H., PONNALA, L., KEICH, U., WANG, W., SUN, Q., CARTINHO, S.W., FILIATRAULT, M.J., WIEDMANN, M. & BOOR, K.J. (2009) Deep RNA Sequencing of *L. monocytogenes* Reveals Overlapping and Extensive Stationary Phase and Sigma B-Dependent Transcriptomes, Including Multiple Highly Transcribed Noncoding RNAs. *BMC Genomics* **10**: 641.
- OPPENHEIM, D.S. & YANOFSKY, C. (1980) Translational Coupling During Expression of the Tryptophan Operon of *Escherichia coli*. *Genetics* **95**(4): 785–795.
- OSBOURN, A.E. & FIELD, B. (2009) Operons. *Cellular and Molecular Life sciences* **66**(23): 3755–3775.
- OSHIMA, T., ISHIKAWA, S., KUROKAWA, K., AIBA, H. & OGASAWARA, N. (2006) *Escherichia coli* Histone-Like Protein H-NS Preferentially Binds to Horizontally Acquired DNA in Association with RNA Polymerase. *DNA Research* **13**(4): 141–153.
- OVERBEEK, R., BARTELS, D., VONSTEIN, V. & MEYER, F. (2007) Annotation of Bacterial and Archaeal Genomes: Improving Accuracy and Consistency. *Chemical Reviews* **107**(8): 3431–3447.
- OZBUDAK, E.M., THATTAI, M., KURTSE, I., GROSSMAN, A.D. & VAN OUDENAARDEN, A. (2002) Regulation of Noise in the Expression of a Single Gene. *Nature Genetics* **31**(1): 69–73.
- PAGET, M.S.B. & HELMANN, J.D. (2003) The Sigma70 Family of Sigma Factors. *Genome Biology* **4**(1): 203.
- PARTE, A.C. (2014) LPSN-List of Prokaryotic Names with Standing in Nomenclature. *Nucleic Acids Research* **42**(D1): D613–D616.
- PASCUAL, C., LAWSON, P.A., FARROW, J. A. E., GIMENEZ, M.N. & COLLINS, M.D. (1995) Phylogenetic Analysis of the Genus *Corynebacterium* Based on 16S rRNA Gene Sequences. *International Journal of Systematic Bacteriology* **45**(4): 724–728.
- PASSALACQUA, K.D., VARADARAJAN, A., ONDOV, B.D., OKOU, D.T., ZWICK, M.E. & BERGMAN, N.H. (2009) Structure and Complexity of a Bacterial Transcriptome. *Journal of Bacteriology* **191**(10): 3203–3211.

- PÁTEK, M. & NEŠVERA, J. (2011) Sigma Factors and Promoters in *Corynebacterium glutamicum*. *Journal of Biotechnology* **154**(2-3): 101–113.
- PÁTEK, M., NESVERA, J., GUYONVARCH, A., REYES, O. & LEBLON, G. (2003) Promoters of *Corynebacterium glutamicum*. *Journal of Biotechnology* **104**(1-3): 311–323.
- 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 Research* **40**(Database issue): D610–614.
- PEANO, C., PIETRELLI, A., CONSOLANDI, C., ROSSI, E., PETITI, L., TAGLIABUE, L., DE BELLIS, G. & LANDINI, P. (2013) An efficient rRNA removal Method for RNA Sequencing in GC-rich Bacteria. *Microbial Informatics and Experimentation* **3**: 1.
- PERKINS, J.B., BOWER, S., HOWITT, C.L., YOCUM, R.R. & PERO, J. (1996) Identification and Characterization of Transcripts from the Biotin Biosynthetic Operon of *Bacillus subtilis*. *Journal of Bacteriology* **178**(21): 6361–6365.
- PERKINS, T.T., KINGSLEY, R.A., FOOKES, M.C., GARDNER, P.P., JAMES, K.D., YU, L., ASSEFA, S.A., HE, M., CROUCHER, N.J., PICKARD, D.J., MASKELL, D.J., PARKHILL, J., CHOUDHARY, J., THOMSON, N.R. & DOUGAN, G. (2009) A Strand-Specific RNA-Seq Analysis of the Transcriptome of the Typhoid Bacillus *Salmonella typhi*. *PLOS Genetics* **5**(7): e1000569.
- PETERS, J.M., MOONEY, R.A., KUAN, P.F., ROWLAND, J.L., KELEŞ, S. & LANDICK, R. (2009) Rho Directs Widespread Termination of Intragenic and Stable RNA Transcription. *Proceedings of the National Academy of Sciences of the United States of America* **106**(36): 15406–15411.
- PETERS, J.M., VANGELOFF, A.D. & LANDICK, R. (2011) Bacterial Transcription Terminators: The RNA 3'-End Chronicles. *Journal of Molecular Biology* **412**(5): 793–813.
- PETERSEN, C., MØLLER, L.B. & VALENTIN-HANSEN, P. (2002) The Cryptic Adenine Deaminase Gene of *Escherichia coli*. Silencing by the Nucleoid-Associated DNA-Binding Protein, H-NS, and Activation by Insertion Elements. *Journal of Biological Chemistry* **277**(35): 31373–31380.
- 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.
- PONNAMBALAM, S., WEBSTER, C., BINGHAM, A. & BUSBY, S. (1986) Transcription Initiation at the *Escherichia coli* Galactose Operon Promoters in the Absence of the Normal -35-Region Sequences. *The Journal of Biological Chemistry* **261**(34): 16043–16048.
- POWELL, S., SZKLARCZYK, D., TRACHANA, K., ROTH, A., KUHN, M., MULLER, J., ARNOLD, R., RATTEI, T., LETUNIC, I., DOERKS, T., JENSEN, L.J., VON MERING, C. & BORK, P. (2012) eggNOG v3.0: Orthologous Groups Covering 1133 Organisms at 41 Different Taxonomic Ranges. *Nucleic Acids Research* **40**(Database Issue): D284–289.
- PRIBNOW, D. (1975) Nucleotide Sequence of an RNA Polymerase Binding Site at an Early T7 Promoter. *Proceedings of the National Academy of Sciences of the United States of America* **72**(3): 784–788.
- PRIBNOW, D. (1975) Bacteriophage T7 Early Promoters: Nucleotide Sequences of Two RNA Polymerase Binding Sites. *Journal of Molecular Biology* **99**(3): 419–443.

- PUECH, V., CHAMI, M., LEMASSU, A., LANÉELLE, M.A., SCHIFFLER, B., GOUNON, P., BAYAN, N., BENZ, R. & DAFFÉ, M. (2001) Structure of the Cell Envelope of *Corynebacteria*: Importance of the Non-Covalently Bound Lipids in the Formation of the Cell Wall Permeability Barrier and Fracture Plane. *Microbiology* **147**(Pt 5): 1365–1382.
- QIU, Y., CHO, B.-K., PARK, Y.S., LOVLEY, D., PALSSON, B.Ø. & ZENGLER, K. (2010) Structural and Operational Complexity of the *Geobacter sulfurreducens* Genome. *Genome Research* **20**(9): 1304–1311.
- RAMULU, H.G., ADINDLA, S. & GURUPRASAD, L. (2006) Analysis and Modeling of Mycolyl-Transferases in the CMN Group. *Bioinformatics* **1**(5): 161–169.
- RICHARDSON, J.P. (2002) Rho-Dependent Termination and ATPases in Transcript Termination. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* **1577**(2): 251–260.
- RINGQUIST, S., SHINEDLING, S., BARRICK, D., GREEN, L., BINKLEY, J., STORMO, G.D. & GOLD, L. (1992) Translation Initiation in *Escherichia coli*: Sequences within the Ribosome Binding Site. *Molecular Microbiology* **6**(9): 1219–1229.
- RUAN, X. & RUAN, Y. (2012) Genome Wide Full-Length Transcript Analysis Using 5'- and 3'-Paired-End-Tag Next Generation Sequencing (RNA-PET). In *Transcriptional Regulation* [ed A. VANCURA] pp. 535–562. Springer New York.
- RUIMY, R., RIEGEL, P., BOIRON, P., MONTEIL, H. & CHRISTEN, R. (1995) Phylogeny of the Genus *Corynebacterium* Deduced from Analyses of Small-Subunit Ribosomal DNA Sequences. *International Journal of Systematic Bacteriology* **45**(4): 740–746.
- 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. *Applied and Environmental Microbiology* **65**(5): 1973–1979.
- SAMBROOK, J., FRITSCH, E.F. & MANIATIS, T. (1989) Molecular cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- SCHIBLER, U. & SIERRA, F. (1987) Alternative Promoters in Developmental Gene Expression. *Annual Review of Genetics* **21**: 237–257.
- SCHMEISSNER, U., MCKENNEY, K., ROSENBERG, M. & COURT, D. (1984) Removal of a Terminator Structure by RNA Processing Regulates in Gene Expression. *Journal of Molecular Biology* **176**(1): 39–53.
- SCHNEIDER, J., VORHÖLTER, F.-J., TROST, E., BLOM, J., MUSA, Y.R., NEUWEGER, H., NIEHAUS, K., SCHATSCHEIDER, S., TAUCH, A. & GOESMANN, A. (2010) CARMEN - Comparative Analysis and *In Silico* Reconstruction of Organism-Specific Metabolic Networks. *Genetics and Molecular Research* **9**(3): 1660–1672.
- SCHOENFELD, A., DAVIDOWITZ, E.J. & BURK, R.D. (1998) A Second Major Native von Hippel-Lindau Gene Product, Initiated from an Internal Translation Start Site, Functions as a Tumor Suppressor. *Proceedings of the National Academy of Sciences of the United States of America* **95**(15): 8817–8822.
- 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**(1): 545.

- SCHROTH, G.P. & HO, P.S. (1995) Occurrence of Potential Cruciform and H-DNA Forming Sequences in Genomic DNA. *Nucleic Acids Research* **23**(11): 1977–1983.
- SEEBURG, P.H., NÜSSLEIN, C. & SCHALLER, H. (1977) Interaction of RNA Polymerase with Promoters from Bacteriophage fd. *European Journal of Biochemistry* **74**(1): 107–113.
- SELIVERSTOV, A.V., PUTZER, H., GELFAND, M.S. & LYUBETSKY, V.A. (2005) Comparative Analysis of RNA Regulatory Elements of Amino Acid Metabolism Genes in Actinobacteria. *BMC Microbiology* **5**: 54.
- SHARMA, C.M., HOFFMANN, S., DARFEUILLE, F., REIGNIER, J., FINDEISS, S., SITTKA, A., CHABAS, S., REICHE, K., HACKERMÜLLER, J., REINHARDT, R., STADLER, P.F. & VOGEL, J. (2010) The Primary Transcriptome of the Major Human Pathogen *Helicobacter pylori*. *Nature* **464**(7286): 250–255.
- SHINE, J. & DALGARNO, L. (1974) The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences of the United States of America* **71**(4): 1342–1346.
- SHINE, J. & DALGARNO, L. (1975) Determinant of Cistron Specificity in Bacterial Ribosomes. *Nature* **254**(5495): 34–38.
- SILVERSTONE, A.E., GOMAN, M. & SCAIFE, J.G. (1972) ALT: A New Factor Involved in the Synthesis of RNA by *Escherichia coli*. *Molecular and General Genetics* **118**(3): 223–234.
- SOREK, R. & COSSART, P. (2010) Prokaryotic Transcriptomics: A New View on Regulation, Physiology and Pathogenicity. *Nature Reviews Genetics* **11**(1): 9–16.
- SORIANO, F. & TAUCH, A. (2008) Microbiological and Clinical Features of *Corynebacterium urealyticum*: Urinary Tract Stones and Genomics as the Rosetta Stone. *European Journal of Clinical Microbiology and Infectious Diseases* **14**(7): 632–643.
- STACKEBRANDT, E., RAINEY, F.A. & WARD-RAINEY, N.L. (1997) Proposal for a New Hierarchic Classification System, Actinobacteria classis nov. *International Journal of Systematic Bacteriology* **47**(2): 479–491.
- STACKEBRANDT, E. & WOESE, C.R. (1981) The Evolution Of Prokaryotes. In *Molecular and Cellular Aspects of Microbial Evolution* [eds M.J. CARLILE, J.F. COLLINS & B.E.B. MOSELEY]. Cambridge University Press, Cambridge.
- STORMO, G.D., SCHNEIDER, T.D. & GOLD, L.M. (1982) Characterization of Translational Initiation Sites in *E. coli*. *Nucleic Acids Research* **10**(9): 2971–2996.
- STORZ, G. (2002) An Expanding Universe of Noncoding RNAs. *Science* **296**(5571): 1260–1263.
- SZYBALSKI, W., KUBINSKI, H. & SHELDRIK, P. (1966) Pyrimidine Clusters on the Transcribing Strand of DNA and their Possible Role in the Initiation of RNA Synthesis. *Cold Spring Harbor Symposia on Quantitative Biology* **31**: 123–127.
- TAKANAMI, M., SUGIMOTO, K., SUGISAKI, H. & OKAMOTO, T. (1976) Sequence of Promoter for Coat Protein Gene of Bacteriophage fd. *Nature* **260**(5549): 297–302.
- TANG, F., BARBACIORU, C., WANG, Y., NORDMAN, E., LEE, C., XU, N., WANG, X., BODEAU, J., TUCH, B.B., SIDDIQUI, A., LAO, K. & SURANI, M.A. (2009) mRNA-Seq Whole-Transcriptome Analysis of a Single Cell. *Nature Methods* **6**(5): 377–382.

- TAUCH, A., HOMANN, I., MORMANN, S., RÜBERG, S., BILLAULT, A., BATHE, B., BRAND, S., BROCKMANN-GRETZA, O., RÜCKERT, C., SCHISCHKA, N., WRENGER, C., HOHEISEL, J., MÖCKEL, B., HUTHMACHER, K., PFEFFERLE, W., PÜHLER, A. & KALINOWSKI, J. (2002) Strategy to Sequence the Genome of *Corynebacterium glutamicum* ATCC 13032: Use of a Cosmid and a Bacterial Artificial Chromosome Library. *Journal of Biotechnology* **95**(1): 25–38.
- THANKY, N.R., YOUNG, D.B. & ROBERTSON, B.D. (2007) Unusual Features of the Cell Cycle in Mycobacteria: Polar-Restricted Growth and the Snapping-Model of Cell Division. *Tuberculosis* **87**(3): 231–236.
- THOMASON, M.K. & STORZ, G. (2010) Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annual Review of Genetics* **44**: 167–188.
- THOMPSON, J.A., RADONOVICH, M.F. & SALZMAN, N.P. (1979) Characterization of the 5'-Terminal Structure of Simian Virus 40 Early mRNA's. *Journal of Virology* **31**(2): 437–446.
- UDAKA, S. (2008) The Discovery of *Corynebacterium glutamicum* and Birth of Amino Acid Fermentation Industry in Japan. In *Corynebacteria: Genomics and Molecular Biology* [ed A. BURKOVSKI]. Caister Academic Press, Norfolk.
- UPTAIN, S.M., KANE, C.M. & CHAMBERLIN, M.J. (1997) Basic Mechanisms of Transcript Elongation and Its Regulation. *Annual Review of Biochemistry* **66**(1): 117–172.
- VASSYLYEV, D.G., SEKINE, S., LAPTENKO, O., LEE, J., VASSYLYEVA, M.N., BORUKHOV, S. & YOKOYAMA, S. (2002) Crystal Structure of a Bacterial RNA Polymerase Holoenzyme at 2.6 Å Resolution. *Nature* **417**(6890): 712–719.
- VENTURA, M., CANCHAYA, C., TAUCH, A., CHANDRA, G., FITZGERALD, G.F., CHATER, K.F. & VAN SINDEREN, D. (2007) Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. *Microbiology and Molecular Biology Reviews* **71**(3): 495–548.
- VITRESCHAK, A.G., RODIONOV, D.A., MIRONOV, A.A. & GELFAND, M.S. (2004) Riboswitches: The Oldest Mechanism for the Regulation of Gene Expression? *Trends in Genetics* **20**(1): 44–50.
- VOCKENHUBER, M.-P., SHARMA, C.M., STATT, M.G., SCHMIDT, D., XU, Z., DIETRICH, S., LIESEGANG, H., MATHEWS, D.H. & SUESS, B. (2011) Deep Sequencing-Based Identification of Small Non-Coding RNAs in *Streptomyces coelicolor*. *RNA Biology* **8**(3): 468–477.
- VOGL, C., GRILL, S., SCHILLING, O., STÜLKE, J., MACK, M. & STOLZ, J. (2007) Characterization of Riboflavin (Vitamin B₂) Transport Proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *Journal of Bacteriology* **189**(20): 7367–7375.
- WACHI, M. (2013) Amino Acid Exporters in *Corynebacterium glutamicum*. In *Corynebacterium glutamicum - Biology and Biotechnology* [eds H. YUKAWA & M. INUI] pp. 335–349. Springer Berlin Heidelberg, Berlin, Heidelberg.
- WANG, Z., GERSTEIN, M. & SNYDER, M. (2009) RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nature Reviews Genetics* **10**(1): 57–63.
- WATSON, P.Y. & FEDOR, M.J. (2012) The ydaO Motif is an ATP-Sensing Riboswitch in *Bacillus subtilis*. *Nature Chemical Biology* **8**(12): 963–965.
- WEINBERG, Z., BARRICK, J.E., YAO, Z., ROTH, A., KIM, J.N., GORE, J., WANG, J.X., LEE, E.R., BLOCK, K.F., SUDARSAN, N., NEPH, S., TOMPA, M., RUZZO, W.L. & BREAKER, R.R. (2007)

- Identification of 22 Candidate Structured RNAs in Bacteria Using the CMfinder Comparative Genomics Pipeline. *Nucleic Acids Research* **35**(14): 4809–4819.
- WEINBERG, Z., WANG, J.X., BOGUE, J., YANG, J., CORBINO, K., MOY, R.H. & BREAKER, R.R. (2010)** Comparative Genomics Reveals 104 Candidate Structured RNAs from Bacteria, Archaea, and their Metagenomes. *Genome Biology* **11**(3): R31.
- WENDISCH, V.F., BOTT, M., KALINOWSKI, J., OLDIGES, M. & WIECHERT, W. (2006)** Emerging *Corynebacterium glutamicum* Systems Biology. *Journal of Biotechnology* **124**(1): 74–92.
- WENDISCH, V.F., ZIMMER, D.P., KHODURSKY, A., PETER, B., COZZARELLI, N. & KUSTU, S. (2001)** Isolation of *Escherichia coli* mRNA and Comparison of Expression Using mRNA and Total RNA on DNA Microarrays. *Analytical Biochemistry* **290**(2): 205–213.
- WILSON, K.S. & VON HIPPEL, P.H. (1995)** Transcription Termination at Intrinsic Terminators: The Role of the RNA Hairpin. *Proceedings of the National Academy of Sciences of the United States of America* **92**(19): 8793–8797.
- WINKLER, W., NAHVI, A. & BREAKER, R.R. (2002)** Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression. *Nature* **419**(6910): 952–956.
- WINKLER, W.C. & BREAKER, R.R. (2005)** Regulation of Bacterial Gene Expression by Riboswitches. *Annual Review of Microbiology* **59**: 487–517.
- WINKLER, W.C., COHEN-CHALAMISH, S. & BREAKER, R.R. (2002)** An mRNA Structure that Controls Gene Expression by Binding FMN. *Proceedings of the National Academy of Sciences of the United States of America* **99**(25): 15908–15913.
- WIVAGG, C.N. & HUNG, D.T. (2012)** Resuscitation-Promoting Factors Are Required for Beta-Lactam Tolerance and the Permeability Barrier in *Mycobacterium tuberculosis*. *Antimicrobial Agents and Chemotherapy* **56**(3): 1591–1594.
- WOHLLEBEN, W., MUTH, G. & KALINOWSKI, J. (1993)** Genetic Engineering of Gram Positive Bacteria. In *Biotechnology Set* [eds H.-J. REHM & G. REED] pp. 455–505. VCH, Weinheim.
- WÖSTEN, M.M. (1998)** Eubacterial Sigma-Factors. *FEMS Microbiology Reviews* **22**(3): 127–150.
- WU, C.J. & JANSSEN, G.R. (1996)** Translation of *vph* mRNA in *Streptomyces lividans* and *Escherichia coli* after Removal of the 5'-Untranslated Leader. *Molecular Microbiology* **22**(2): 339–355.
- WURTZEL, O., SAPRA, R., CHEN, F., ZHU, Y., SIMMONS, B.A. & SOREK, R. (2010)** A Single-Base Resolution Map of an Archaeal Transcriptome. *Genome Research* **20**(1): 133–141.
- YANG, D., FU, Y., WU, X., XIE, Y., NIE, H., CHEN, L., NONG, X., GU, X., WANG, S., PENG, X., YAN, N., ZHANG, R., ZHENG, W. & YANG, G. (2012)** Annotation of the Transcriptome from *Taenia pisiformis* and Its Comparative Analysis with Three *Taeniidae* Species. *PLOS One* **7**(4): e32283.
- ZEMANOVÁ, M., KADERÁBKOVÁ, P., PÁTEK, M., KNOPPOVÁ, M., SILAR, R. & NESVERA, J. (2008)** Chromosomally Encoded Small Antisense RNA in *Corynebacterium glutamicum*. *FEMS Microbiology Letters* **279**(2): 195–201.
- ZUO, Y., WANG, Y. & MALHOTRA, A. (2005)** Crystal Structure of *Escherichia coli* RNase D, an Exoribonuclease Involved in Structured RNA Processing. *Structure* **13**(7): 973–984.

VIII Anhang

1 Abkürzungsverzeichnis

A	Adenin
bp	Basenpaar(e)
BPB-XC	Bromphenolblau-Xylencyanol
bzw.	Beziehungsweise
C	Cytosin
<i>C. diphtheriae</i>	<i>Corynebacterium diphtheriae</i>
<i>C. glutamicum</i>	<i>Corynebacterium glutamicum</i>
CASO	Casein-Soja-Pepton
cDNA	<i>Complementary Deoxyribonucleic Acid</i> (komplementäre Desoxyribonukleinsäure)
CPR(s)	<i>Combined-Paired-Read(s)</i>
DEPC	Diethylpyrocarbonat
DMSO	Dimethylsulfoxid
DNA	<i>Deoxyribonucleic Acid</i> (Desoxyribonukleinsäure)
DNase(n)	Desoxyribonuklease(n)
dNTP(s)	Desoxyribonukleosidtriphosphat(e)
<i>E. coli</i>	<i>Escherichia coli</i>
ECF	Extracytoplasmatische Funktion
EDTA	Ethylendiamintetraacetat
EtOH	Ethanol
FAD	Flavin-Adenin-Dinukleotid
G	Guanin
<i>G. sulfurreducens</i>	<i>Geobacter sulfurreducens</i>
ggf.	Gegebenenfalls
ID	Identifikationsnummer
kDa	Kilodalton
LB	<i>Lysogeny Broth</i>
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
max	Maximal
min	Minute(n)
Mio	Million(en)
mm	Millimeter
MOPS	3-Morpholinopropan-sulfonsäure
NAD	Nicotinamidadenindinukleotid
nt	Nukleotid(e)
NTP(s)	Nukleosidtriphosphat(e)
PCI	Phenol-Chloroform-Isoamylalkohol
PCR	Polymerase Chain Reaction (Polymerase-Kettenreaktion)
PEG	Polyethylenglykol
PEX	<i>Primer Extension</i>
RACE	<i>Rapid Amplification of cDNA Ends</i>
RBS(s)	Ribosomenbindestelle(n)
rcf	<i>Relative Centrifugal Force</i>
RIN	<i>RNA Integrity Number</i>
RNA	<i>Ribonucleic Acid</i> (Ribonukleinsäure)
RNase(n)	Ribonuklease(n)
RNAseq	Transkriptom-Hochdurchsatz-Sequenzierung

rRNA(s).....	<i>Ribosomal Ribonucleic Acid(s)</i> (Ribosomale Ribonukleinsäure(n))
s.....	Siehe
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
sek.....	Sekunde(n)
<i>SOLiD</i>	<i>Sequencing by Oligonucleotide Ligation and Detection</i>
sRNA(s)	<i>Small Ribonucleic Acid(s)</i> (kleine Ribonukleinsäure(n))
T.....	Tyrosin
TLS(s)	Translationsstart(s)
tmRNA.....	<i>Transfer-Messenger Ribonucleic Acid</i> (Transfer-Boten-Ribonukleinsäure)
TRIS.....	Tris-aminomethan
tRNA(s).....	<i>Transfer Ribonucleic Acid(s)</i> (Transfer Ribonukleinsäure(n))
TSS(s)	Transkriptionsstart(s)
u.a.....	Unter anderem
UTR(s)	<i>Untranslated Region(s)</i> (Untranslatierte Region)
UV.....	Ultraviolett(strahlung)
vgl.	Vergleiche
Vol.	Volumen
z.B.....	Zum Beispiel

2 Ergänzende Abbildungen

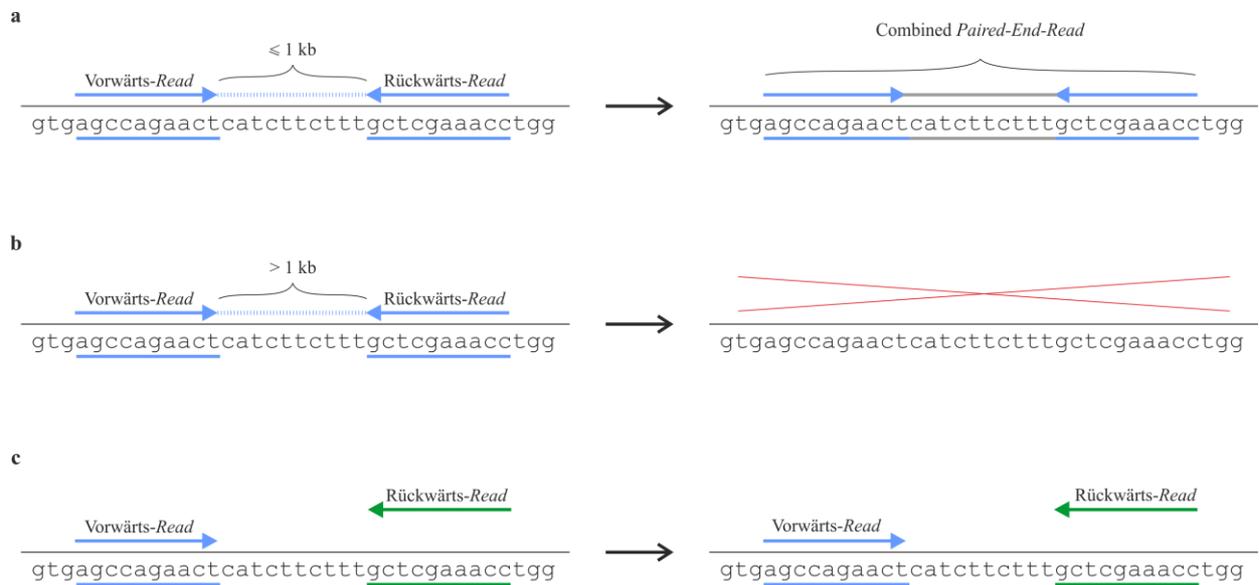


Abbildung VIII-1: Schematische Darstellung der Read-Kartierung der Bibliothek der Vollängentranskripte auf das Genom von *C. glutamicum*

- a.** Vorwärts- und Rückwärts-Reads als *Paired-End-Reads* (blaue Pfeile), die mit einer max. Distanz von 1 kb auf dem Genom von *C. glutamicum* kartiert werden, werden zu einem *Combined Paired-End-Read* (CPR) verknüpft, wobei die Genomsequenz als *Insert* (graue Linie) zwischen den beiden *Paired-End-Reads* eingefügt wird.
- b.** Vorwärts- und Rückwärts-Reads als *Paired-End-Reads* (blaue Pfeile), die mit einer Distanz von > 1 kb auf dem Genom von *C. glutamicum* kartiert werden, werden beide verworfen und nicht für weitere Analysen verwendet.
- c.** Vorwärts- oder Rückwärts-Reads, die ohne einen passenden *Paired-End-Read* auf dem Genom von *C. glutamicum* kartiert werden (blauer und grüner Pfeil), werden als einzelne *Single-Reads* für weitere Analysen verwendet.

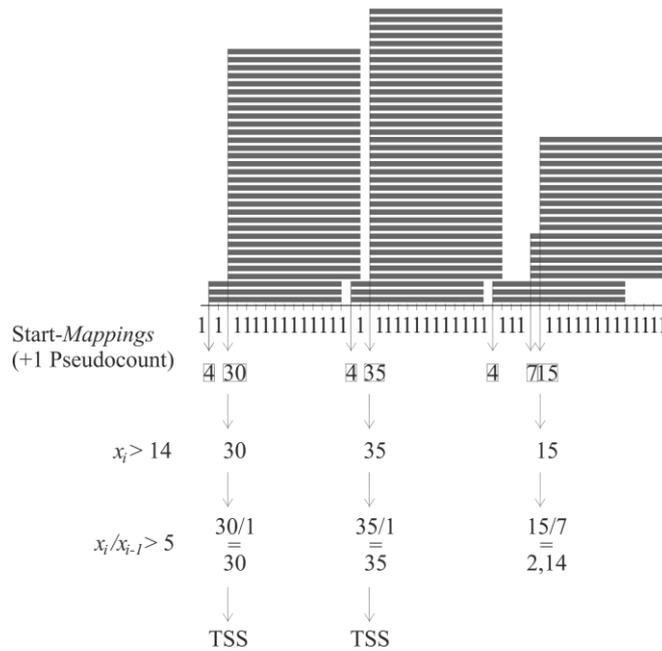


Abbildung VIII-2: Schematische Darstellung der automatischen TSS-Detektion innerhalb der Kartierungsdaten der Bibliothek der nativen 5'-Transkriptenden

Zunächst wird an jeder Position im Genom die Anzahl der Start-Mapping-Ereignisse gezählt und ein Pseudocount von 1 hinzuaddiert. Wenn dann die Anzahl der Start-Mapping-Ereignisse (plus Pseudocount von 1) den Schwellenwert 14 übersteigt sowie das Verhältnis von der Anzahl der Start-Mapping-Ereignisse (plus Pseudocount von 1) an der Position i zu der Anzahl der Start-Mapping-Ereignisse (plus Pseudocount von 1) an der vorherigen Position $i - 1$ den Schwellenwert 5 übersteigt, wird die Position i als TSS erkannt.

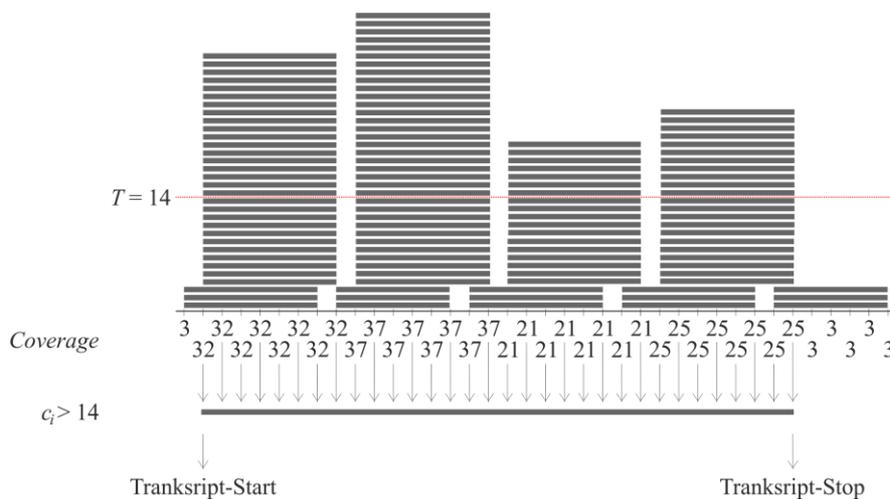


Abbildung VIII-3: Schematische Darstellung der automatischen Detektion neuer Transkripte innerhalb der Kartierungsdaten der Bibliothek der Vollängentranskripte

Zunächst wird die Read-Abdeckung an jeder Position im Genom bestimmt. Wenn die Read-Abdeckung einen Schwellenwert von 14 an der Position i übersteigt, wird diese Position als Transkript-Start gesetzt. Wenn der Schwellenwert von 14 auch an Position $i + x$ überstiegen wird, wird das Transkript um die Position verlängert. Wenn an der Position $i + x$ der Schwellenwert von 14 nicht mehr überstiegen wird, wird die Position $i + (x - 1)$ als Transkriptende gesetzt.

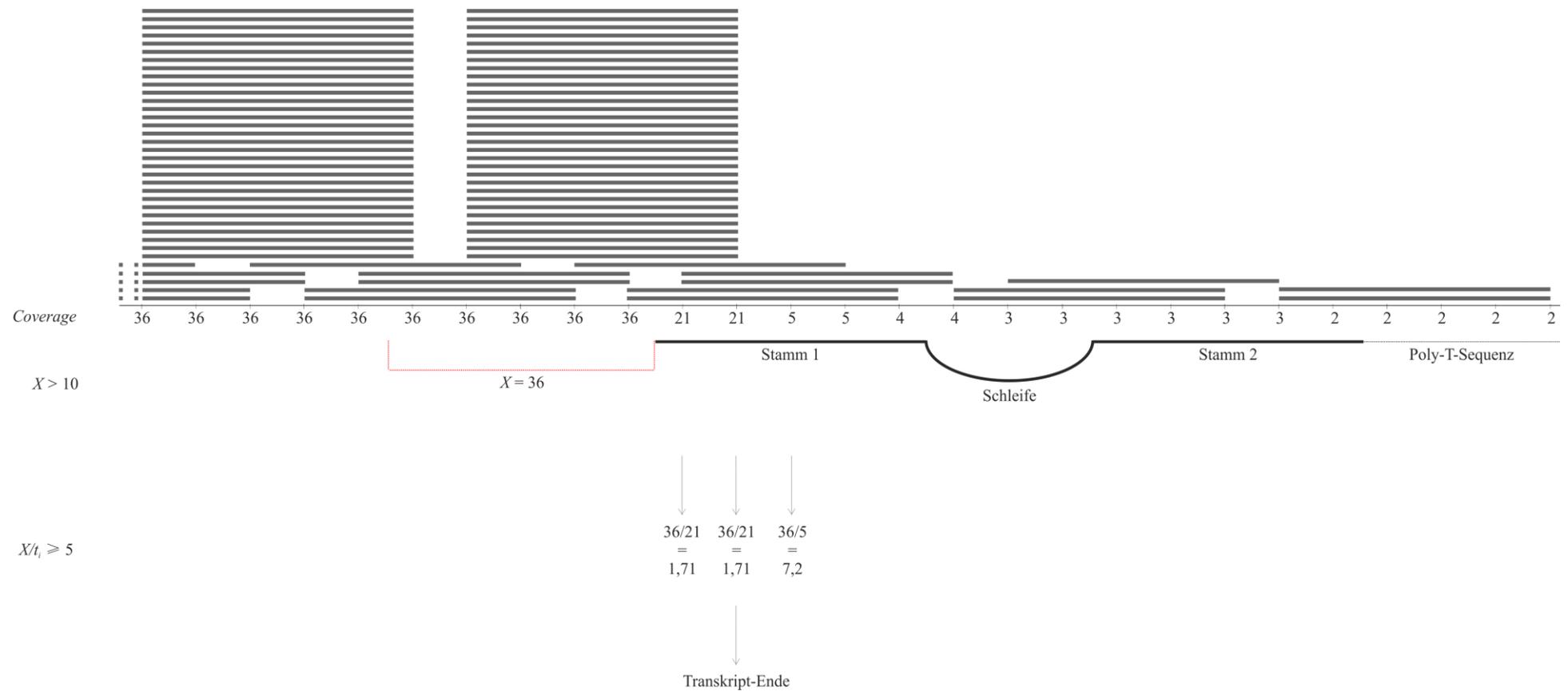


Abbildung VIII-4: Schematische Darstellung der automatischen Detektion von Transkriptenden von annotierten Genen mit vorhergesagten p-unabhängigen Terminatoren innerhalb der Kartierungsdaten der Bibliothek der Vollängentranskripte

Zunächst wird der Durchschnitt der *Read*-Abdeckung (X) der fünf Positionen vor dem ersten, vorhergesagten Terminator-Stamm berechnet. Wenn diese > 10 ist, dann wird versucht ein Transkriptende zu bestimmen. Dazu wird an jeder Position des vorhergesagten Terminators das Verhältnis des Durchschnitts der *Read*-Abdeckung X zu der *Read*-Abdeckung t_i innerhalb des Terminators bestimmt (X/t_i). Sobald $X/t_i \geq 5$ gilt, wird die Position i als Transkriptende gesetzt.

3 Ergänzende Tabellen

Die folgenden Tabellen sind aufgrund ihrer Größe auf der beigefügten Daten-CD als .xlsx-Dokumente hinterlegt. Zum Öffnen dieser Dokumente wird Microsoft Excel, OpenOffice Calc oder eine ähnliche Software benötigt.

- Tabelle S1: Liste mit detektierten TSSs, zugehörigen -10- und -35-Promotormotiven des Sigmafaktors σ^A innerhalb 60 Basen stromaufwärts der TSSs.
- Tabelle S2: Liste mit korrigierten Gen-Starts und zugehörigen Genen.
- Tabelle S3: Liste mit Längen und Sequenzen der detektierten 5'-UTRs proteinkodierender Gene.
- Tabelle S4: Liste mit detektierten Ribosomenbindestellen-Motiven innerhalb 20 Basen der 5'-UTRs stromaufwärts des Initiatorcodons.
- Tabelle S5: Liste mit den identifizierten Operon-Strukturen: Primär-Operons, Sub-Operons und monocistronische Transkripte.
- Tabelle S6: Liste mit den neu gefundenen, bisher unbekanntem intergenischen Transkripten.
- Tabelle S7: Liste mit den detektierten, bisher unbekanntem antisense Transkripten.
- Tabelle S8: Liste mit den identifizierten, bisher unbekanntem intragenischen Transkripten.
- Tabelle S9: Liste mit den detektierten Transkriptenden im Vergleich zu den vorhergesagten, ρ -unabhängigen Terminatoren.

4 Abbildungsverzeichnis

Abbildung II-1:	Morphologie und Zellwandaufbau von <i>C. glutamicum</i>	5
Abbildung II-2:	Schematische Darstellung des Aufbaus der prokaryotischen RNA-Polymerase mit einer σ^{70} -Untereinheit, die Interaktion mit Promotorregionen und der Mechanismus der Transkriptionsinitiation (modifiziert nach [Browning & Busby, 2004])	13
Abbildung II-3:	Schematische Darstellung der Möglichkeiten zur rRNA-Abreicherung aus einer Gesamt-RNA-Lösung (modifiziert nach [Sorek & Cossart, 2010]).	19
Abbildung IV-1:	Experimenteller Vorbereitungsablauf der Vollängentranskript- und der nativen 5'-Transkriptenden-Bibliothek	61
Abbildung IV-2:	Klassifizierung und Einteilung der detektierten TSSs	64
Abbildung IV-3:	Veranschaulichung der Detektion von TSS anhand eines Beispiels für das Gen <i>cmt1/cg0413</i>	65
Abbildung IV-4:	Schematische Darstellung der Reannotation von Gen-Starts mithilfe der RNAseq-Daten	67
Abbildung IV-5:	Relative Verteilung der Nukleotide innerhalb der identifizierten -10- und -35-Promotorregion in <i>C. glutamicum</i> als Sequenzlogo	68
Abbildung IV-6:	Diagramm mit der Verteilung der 5'-UTR-Längen von mRNAs in <i>C. glutamicum</i>	70
Abbildung IV-7:	Sekundäre Strukturen von <i>Rfam</i> -vorhergesagten und mit RNAseq betätigten 5'-UTRs in <i>C. glutamicum</i>	73
Abbildung IV-8:	Analyse der RBS in <i>C. glutamicum</i>	77
Abbildung IV-9:	Identifizierung und Klassifizierung von Operon-Strukturen in <i>C. glutamicum</i>	79
Abbildung IV-10:	Analyse der Genanzahl innerhalb monocistronischer Transkripte, Primär-Operons und Sub-Operons in <i>C. glutamicum</i>	80
Abbildung IV-11:	Analyse der intercistronischen RBS in <i>C. glutamicum</i>	83
Abbildung IV-12:	Beispiele von RNAseq-detektierten Transkriptenden im Vergleich zu ρ -unabhängigen Terminatoren	87
Abbildung VIII-1:	Schematische Darstellung der <i>Read</i> -Kartierung der Bibliothek der Vollängentranskripte auf das Genom von <i>C. glutamicum</i>	131
Abbildung VIII-2:	Schematische Darstellung der automatischen TSS-Detektion innerhalb der Kartierungsdaten der Bibliothek der nativen 5'-Transkriptenden	132
Abbildung VIII-3:	Schematische Darstellung der automatischen Detektion neuer Transkripte innerhalb der Kartierungsdaten der Bibliothek der Vollängentranskripte	132
Abbildung VIII-4:	Schematische Darstellung der automatischen Detektion von Transkriptenden von annotierten Genen mit vorhergesagten ρ -unabhängigen Terminatoren innerhalb der Kartierungsdaten der Bibliothek der Vollängentranskripte	133

5 Tabellenverzeichnis

Tabelle II-1:	Vier mittels RNAseq verifizierten [Mentz <i>et al.</i> , 2013; Mentz, 2013] und <i>Rfam</i> vorhergesagten [Gardner <i>et al.</i> , 2011] sRNAs für <i>C. glutamicum</i>	9
Tabelle II-2:	Untereinheiten der RNA-Polymerase in <i>C. glutamicum</i> und dessen Funktion	12
Tabelle II-3:	Vergleich der Länge und Anzahl der <i>Reads</i> zwischen den Sequenzierungsplattformen	20
Tabelle III-1:	Für die RNA-Sequenzierung verwendete Oligonukleotide	28
Tabelle III-2:	Für den Nachweis von DNA verwendete Oligonukleotide	28
Tabelle III-3:	In dieser Arbeit verwendete Labor-Verbrauchsmaterialien mit Bezugsquelle	28
Tabelle III-4:	In dieser Arbeit verwendete Kits mit Bezugsquelle	29
Tabelle III-5:	In dieser Arbeit verwendete Enzyme und DNA-Längenstandards mit Bezugsquelle..	29
Tabelle III-6:	In dieser Arbeit verwendete Chemikalien mit Bezugsquelle.....	30
Tabelle III-7:	In dieser Arbeit verwendete Geräte und Apparaturen mit Bezugsquelle.....	31
Tabelle III-8:	In dieser Arbeit verwendete Software/Datenbanken mit deren Verwendungszweck und Referenz.....	32
Tabelle IV-1:	Vergleich der rRNA-Abreicherung aus <i>C. glutamicum</i> -Gesamt-RNA mit drei kommerziell erhältlichen Anwendungen.....	59
Tabelle IV-2:	Zusammenfassung der Sequenzier- und Kartierungsstatistiken für die Bibliotheken der Volllängentranskripte und nativen 5'-Transkriptenden	62
Tabelle IV-3:	Vergleich von <i>Rfam</i> -vorhergesagten regulatorischen 5'-UTRs mit den RNAseq-Daten für <i>C. glutamicum</i>	74
Tabelle IV-4:	Liste der detektierten Operons mit ≥ 10 Genen.....	81

Danksagung

An erster Stelle bedanke ich mich bei apl. Prof. Dr. Jörn Kalinowski für die Vergabe dieses interessanten Themas und die ausgezeichnete wissenschaftliche Betreuung. Ich danke Ihm für die schöne Zeit in seiner Arbeitsgruppe „Die Corynes“, die vielen spannenden Diskussion und vielen nützlichen Tipps.

Ein besonderer Dank geht auch Dr. Christian Rückert, der mir immer zur Seite stand, mich bei meiner Arbeit mit Rat und Tat unterstützt und bis zuletzt mit mir konstruktive Diskussion geführt hat.

Außerdem bedanke ich mich bei allen Mitgliedern der „Coryne“-Gruppe für das schöne Arbeitsklima und den abwechslungsreichen Spaß außerhalb der Laborarbeit bei dem einen oder anderen Grillen, den gemeinsamen Partys, dem Weihnachts-Bowling, ... Insbesondere möchte ich mich bei meinen Labor- und Büro-Mitbewohnern Almut Mentz, Kathrin Petri, Johanna Milse, Eugenie Fredrich und Armin Neshat bedanken, die mich durch die Höhen und Tiefen meiner Doktorarbeit begleitet haben und für meine Anliegen stets ein offenes Ohr hatten.

Abschließend möchte ich mich bei meiner Familie und meinen Freunden bedanken, die mich in vielen Entscheidungen inspiriert und maßlose Geduld entgegengebracht haben. Vor allem gilt dieser Dank meinem Mann James, meiner Schwester Christina und meinen Eltern, die während dieser spannenden Zeit an meiner Seite waren und in jeglicher Hinsicht unterstützt haben.

Danke!

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertation selbst angefertigt und nur die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen der Dissertation, die anderen Quellen im Wortlaut oder dem Sinn nach entnommen wurden, sind durch Zitate kenntlich gemacht.

Weiterhin erkläre ich, dass die vorliegende Dissertation weder vollständig noch in Auszügen einer anderen Fakultät zum Erwerb eines akademischen Titels vorgelegt worden ist.

Die Anfertigung der vorliegenden Dissertation mit dem Titel „Entwicklung der Transkriptomsequenzierung und Anwendung zur Analyse des Transkriptoms von *Corynebacterium glutamicum*“ erfolgte unter der Betreuung von apl. Prof. Dr. Jörn Kalinowski in der Arbeitsgruppe "Mikrobielle Genomik und Biotechnologie" und im Rahmen des *CLIB Graduate Cluster* Promotionsstudiengangs „*Industrial Biotechnology*“ an der Fakultät für Biologie der Universität Bielefeld.

Rheda-Wiedenbrück,

Katharina Pfeifer-Sancar