TECHNISCHE UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR CHEMIE
LEHRSTUHL FÜR BIOCHEMIE

Expressed Cyclopeptide Libraries as Selective Inhibitors for RNA-Protein Interactions

Structural Basis for the Site-Specific Incorporation of Non-Natural Amino Acids into Peptides and Proteins

Veronika M. Flügel

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. L. Hintermann
Prüfer der Dissertation:	1. TUM Junior Fellow Dr. S. Schneider
	2. Univ.-Prof. Dr. A. Itzen

Die Dissertation wurde am 28.08.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 15.10.2014 angenommen.

„Krise kann ein produktiver Zustand sein. Man muss ihr nur den Beigeschmack der Katastrophe nehmen.“

Max Frisch

Acknowledgements

I would like to thank…

… Dr. Sabine Schneider for giving me the opportunity to carry out my PhD in her group and for supporting me during this challenging project. I am very thankful that she taught me everything what is necessary in order to survive in a biochemistry lab.

… the *Fonds der Chemischen Industrie* for the 2-year financial support during my PhD.

… Dr. Milan Vrabel for the synthesis of the non-natural amino acids.

… Dr. Florian Rechenmacher from the group of Prof. Kessler for showing me tricks in peptide synthesis.

… Katja Bäuml for introducing me into the handling of the mass spectrometer.

… Prof. Michael Groll for integrating me into his group and for giving me unlimited permission to use all of his facilities.

… Astrid König and Ute Kashoa for the great organization in and around the labs, for sharing their food with me and for always having a piece of advice.

… my lab colleagues Sabrina Harteis, Marion Kirchner and Christine Oellig for being great co-workers and for creating such a very friendly working atmosphere in our little group.

… all group members for their daily support and especially Nadine Gillmaier, Stephanie Grubmüller, Sabrina Harteis, Andrea Kunfermann, Anja List and Ingrid Span for the great time inside and outside the lab.

… Nerea Gallastegui for our legendary lunches and for being such an amazing friend!

… Grażyna Ewert, Janusz Ewert, Karolina Ewert and the rest of the Ewert Clan for integrating me into the family and for taking care of me during my countless visits.

… most of all: my family and Michael Ewert:

I am deeply thankful for what you all have done for me in the past and especially within these last four years. At any time I felt unconditionally supported, sheltered and loved. No matter if via telephone or life in person, you encouraged me whenever I needed it the most and I am grateful that you have always been honest with me.

Michi, you are the person who witnesses my emotions in the most unfiltered way, and I especially thank you for being strong for me when I can't and for putting a smile on my face every day. *Po prostu kocham cię!*

Contents

I. Summary

Expressed Cyclopeptide Libraries as Selective Inhibitors for RNA-Protein Interactions

Since the 1940s, antibiotics have been effective drugs to treat bacterial infections, but within the last decades, a vast amount of antibiotic-resistant bacteria strains have developed. Therefore, it is crucial to find new drug targets in order to develop innovative antibacterial drugs with novel modes of action. Riboswitches (RSs), which are structured RNA elements in the 5'UTR of mRNAs, regulate the expression of essential genes in bacterial metabolism and are therefore highly potential drug targets. In this thesis, studies on RSs from *Staphylococcus aureus* and *Escherichia coli* provide a basis for the study of a novel antibiotic mode of action that specifically targets RNA and not proteins. Hereby, cyclic peptides (CPs) represent a substrate class, which could be evolved to serve as new generation antibiotics. They are able to bind to structured RNA elements and due to their extremely stable cyclic structure they offer a great variety of biochemical and biophysical benefits. In this project, a genetically encoded and expressed CP library is constructed in *E. coli*.

The aim of this thesis was to develop an *in vivo* CP screening assay that allows for the identification of CPs that specifically bind to a RS of interest and thus inhibit expression of the downstream genes. In order to investigate the interplay between RSs and CPs, different constructs have been cloned, in which RSs should control the expression of two reporter genes, *lacZ* and *sacB*. The RS activities and qualities of the different constructs were examined using the RSs' respective natural ligands. Hereby, one of the constructs carrying the coenzyme B_{12}-binding *E. coli btuB* AdoCbl RS showed a convincing

regulation of gene expression. In the course of this work, a two-step *in vivo* CP screening assay, which is based on the abovementioned *E. coli btuB* AdoCbl RS, was successfully established. In spite of the fact, that during the timeframe of this thesis an active CP has not yet been identified, this system still presents a powerful tool for the straightforward selection of single active members from CP libraries of several million members.

Structural Basis for the Side-Specific Incorporation of Non-Natural Amino Acids into Proteins and Peptides

Posttranslational modifications (PTMs) on histones epigenetically regulate normal and disease related processes. Hereby, lysine is the main target for modifications like acetylation or methylation. However, it has been shown recently, that lysine PTMs are more versatile and include as well malonylation, succinylation, butyrylation, propionylation or crotonylation, which differ only marginally in their chemical structures but result in diverse, yet significant alteration of gene expression patterns. However, the functional effects of these PTMs are not yet understood. Therefore, it is important to generate proteins, which contain these modified lysine derivatives site-specifically. In the last decade, methods for the translational incorporation of non-natural amino acids using orthogonal aminoacyl-tRNA synthetase (aaRS):tRNAaaCUA pairs were developed. A number of studies show that aaRS like the pyrrolysyl-tRNA synthetase (PylRS) from *Methanosarcina mazei* or *Methanosarcina barkeri* can be genetically evolved to utilize non-natural amino acids and thus help to expand the genetic code. Nevertheless, the wild type PylRS (WT-PylRS) from *M. mazei* readily accepts a variety of lysine derivatives as substrates without the need for additional mutations.

The aim of this project was to elucidate the X-ray crystal structures of the *M. mazei* WT-PylRS in complex with the four non-natural amino acids ε–*N*-propargyloxy-carbonyl- (Kalk), ε-*N*-butyryl- (Kbu), ε-*N*-crotonyl- (Kcr) and ε-*N*-propionyl-lysine (Kpr). The crystallographic data provided insights into the plasticity of the PylRS active site and allowed the deduction of certain key features that are essential for the non-natural amino acids in order to be recognized as a substrate. Results presented in this part of the thesis have been published earlier (Flügel *et al.*, 2014).

Zusammenfassung

Exprimierte Zyklopeptid-Bibliotheken als selektive Inhibitoren für RNA-Protein Interaktionen

Seit den 40er Jahren werden Antibiotika erfolgreich genutzt, um bakterielle Infektionen zu behandeln. Allerdings hat sich in den letzten Jahrzehnten eine Vielzahl von antibiotikaresistenten Bakterienstämmen entwickelt. Daher ist es essentiell, neue Wirkstofftargets zu finden, um innovative Antibiotika mit neuartigen Wirkmechanismen entwickeln zu können. Riboswitche (RSs) sind strukturierte RNA Elemente in der 5'UTR von mRNAs und regulieren die Expression essentieller Gene im bakteriellen Metabolismus. Daher stellen sie Wirkstofftargets mit großem Potential für die Arzneimittelforschung dar. Forschungsarbeiten an den RSs aus *Staphylococcus aureus* und *Escherichia coli*, die in dieser Arbeit durchgeführt wurden, bieten eine wissenschaftliche Grundlage für die Untersuchung eines neuen antibiotischen Wirkmechanismus, der spezifisch RNA involviert und nicht Proteine. Hierbei stellen zyklischen Peptide (CPs) eine Substanzklasse dar, die zu neuartigen Antibiotika weiterentwickelt werden könnten. Sie sind in der Lage an strukturierte RNA zu binden und aufgrund ihrer äußerst stabilen zyklischen Struktur bieten sie eine Vielzahl von biochemischen und biophysikalischen Vorteilen. In diesem Projekt wurde eine Bibliothek von genetisch codierten und exprimierten zyklischen Peptiden in *E. coli* generiert.

Ziel der Arbeit war die Entwicklung eines *in vivo* CP Screeningverfahrens, das die Identifikation von CPs erlaubt, die spezifisch an einen RS von Interesse binden und dadurch die Genexpression nachgeschalteter Gene inhibieren. Um das Zusammenspiel von RSs und CPs zu untersuchen, wurden verschiedene Konstrukte kloniert, in denen die

RSs die Expression von zwei Reportergenen, *lacZ* und *sacB*, kontrollieren sollen. Die Aktivität der RSs und die Qualität der verschiedenen Konstrukte wurden mit Hilfe der natürlichen Liganden der RSs untersucht. Eines der Konstrukte, welches den Coenzym B_{12}-bindenden *E. coli btuB* AdoCbl RS trägt, zeigte hierbei eine überzeugende Regulation der Genexpression. Im Laufe dieser Arbeit wurde ein zweistufiges *in vivo* CP Screeningverfahren etabliert, welches auf dem obenerwähnten *E. coli btuB* AdoCbl RS basiert. Trotz der Tatsache, dass im Rahmen dieser Arbeit noch kein aktives CP identifiziert werden konnte, stellt dieses System dennoch eine leistungsfähige Methode dar, um einzelne aktive CPs aus CP-Bibliotheken mit mehreren Millionen Mitglieder zu selektieren.

Strukturelle Grundlagen für den ortspezifischen Einbau unnatürlicher Aminosäuren in Proteine und Peptide

Posttranslationale Modifikationen (PTMs) an Histonen regulieren normale und krankheitsbezogene Prozesse auf epigenetischer Ebene. Hierbei sind Lysine die Haupttargets für Modifikationen wie Acetylierungen oder Methylierungen. Unlängst wurde jedoch gezeigt, dass Lysin-PTMs vielfältiger sind als bisher angenommen und auch Malonylierungen, Succinylierungen, Butyrylierungen, Propionylierungen oder Crotonylierungen beinhalten. Diese unterscheiden sich zwar nur geringfügig in ihrer chemischen Struktur voneinander, rufen aber unterschiedliche, jedoch erhebliche Veränderungen von Genexpressionsmustern hervor. Allerdings sind die funktionellen Auswirkungen dieser PTMs noch nicht vollständig aufgeklärt. Daher ist es wichtig Proteine zu generieren, die diese modifizierten Lysinderivate ortsspezifisch enthalten. In den letzten zehn Jahren wurden Methoden für den translationellen Einbau unnatürlicher Aminosäuren entwickelt, die das orthogonale aminoacyl-tRNA Synthetase (aaRS):tRNAaaCUA Paar verwenden. Eine Vielzahl an Studien zeigt, dass aaRS wie die Pyrrolysyl-tRNA Synthetase (PylRS) aus *Methanosarcina mazei* oder *Methanosarcina barkeri* gentechnisch weiterentwickelt werden können, sodass sie unnatürliche Aminosäuren verwenden und somit helfen den genetischen Code zu erweitern. Jedoch akzeptiert auch die Wildtyp PylRS (WT-PylRS) aus *M. mazei* eine Vielzahl an

Lysinderivaten als Substrate, sodass eine gerichtete Evolution des Enzyms nicht notwendig ist.

Ziel dieses Projektes war die Aufklärung der Röntgenkristallstrukturen der *M. mazei* WT-PylRS im Komplex mit den vier unnatürlichen Aminosäuren ε–*N*-Propargyloxy-carbonyl- (Kalk), ε-*N*-Butyryl- (Kbu), ε-*N*-Crotonyl- (Kcr) und ε-*N*-Propionyl-Lysin (Kpr). Die kristallographischen Daten bieten Einsicht in die Plastizität des aktiven Zentrums der PylRS und erlaubten die Ableitung bestimmter Schlüsselmerkmale, die für unnatürliche Aminosäuren essentiell sind, um als Substrate erkannt zu werden. Ergebnisse, die in diesem Teil der Arbeit präsentiert werden, sind bereits veröffentlicht (Flügel *et al.*, 2014).

II. Expressed Cyclopeptide Libraries as Selective Inhibitors for RNA-Protein Interactions

II.1. Introduction

Within the past decades, antibiotic-resistant bacteria have become a serious worldwide problem. Ever since the mid-1940s, when antibiotics were first used to fight bacterial infections, microorganisms have started to develop a noticeable amount of mechanisms to protect themselves against this kind of medication. Currently, the most harmful antibiotic-resistant bacteria are methicillin-resistant *S. aureus* (MRSA), vancomycin-resistant enterococci (VRE) and Enterobacteriaceae with extended-spectrum β-lactamases (Dennesen *et al.*, 1998). *E. coli*, a member of the Enterobacteriaceae family, can cause severe infections like gastroenteritis, diarrhea, urinary tract infection, peritonitis, meningitis, and septicemia (von Baum & Marre, 2005; Sodha *et al.*, 2011; Tadesse *et al.*, 2012). One important reason for the increasing development of resistance mechanisms in bacteria is that antibiotics in clinical use have a limited variety regarding their chemical structure and address only a small range of cellular mechanisms (Blount & Breaker, 2006). Therefore, the identification of new drug targets is crucial for the development of novel antibiotics with innovative chemical structures and modes of action.

One of such novel approaches is to specifically target the regulation of genes that are vital or impart virulence to pathogenic bacteria. Therefore, structured RNA elements, like riboswitches, that are able to regulate gene expression are considered to be potent cellular targets in the struggle against multi-drug resistance (Blount & Breaker, 2006). Hereby, it is crucial to identify small molecules that are able to bind to RSs and thus inhibit the expression of essential genes in the bacterial metabolism.

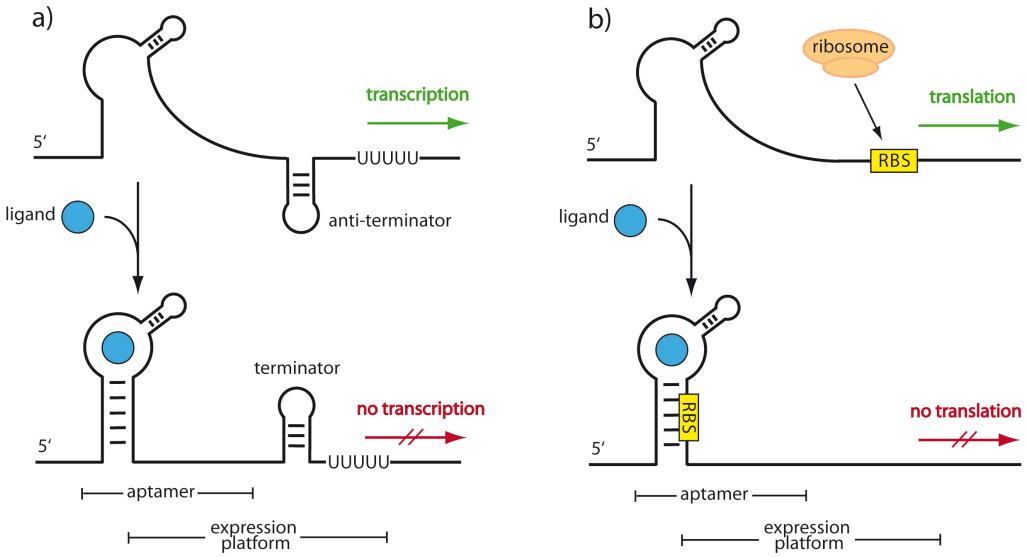
II.1.1. Riboswitches

In the early 1960s, it was a common opinion that biopolymer functions were clearly separated. Proteins, on the one hand, were considered the only molecules capable of catalyzing chemical reactions. Even control of gene expression was assumed to be implemented by proteins. DNA and RNA, on the other hand, were thought to merely store and transform genetic information (Crick, 1970; Mandal & Breaker, 2004b). However, the discovery of ribozymes has proven that RNA can actually be a catalytically active molecule, due to the formation of complex three-dimensional structures (Kruger *et al.*, 1982; Guerrier-Takada *et al.*, 1983). Since then, RNA has turned out to be tremendously versatile and to control a myriad of cellular key processes, such as metabolite sensing, regulation of gene expression, development and differentiation (Mattick, 2009).

Riboswitches (RSs), for instance, are structured and conserved *cis*-regulatory RNA elements in the 5' UTR of mRNAs. They are able to bind small molecules and metabolites and fold into complex three-dimensional structures in order to regulate gene expression. RSs consist of two functional domains. The *aptamer domain* is the higher conserved domain and selectively binds effector molecules by forming a three-dimensional binding site. Upon ligand binding, the downstream *expression platform* undergoes conformational changes and thus enhances or inhibits gene expression (Mandal & Breaker, 2004b; Garst *et al.*, 2011; Breaker, 2012). Hereby, inhibition of gene expression is of particular interest, regarding an application as antibacterial drug target. The two most common mechanisms to inhibit gene expression involve either the formation of a rho-independent terminator stem, which results in inhibition of

transcription or the withdrawal of the ribosome binding site (RBS) from RNA polymerase access, resulting in inhibition of translation (see **Figure 1**) (Kim & Breaker, 2008). The three-dimensional structures of these highly complex RNA molecules are stabilized by tertiary motives such as A-minor motifs, base-triples, kink-turns, loop-loop interactions, pseudoknots and ribose zippers (see **II.1.1.1** and **II.1.1.2)** (Appasamy *et al.*, 2013).

Figure 1. Mechanisms of transcription and translation inhibition (modified after Kim & Breaker, 2008). a) Inhibition of transcription. In absence of a RS ligand, the expression platform adopts an anti-terminator stem structure, which is followed by a poly-uracil RNA sequence. Upon ligand binding, a terminator stem is formed, which causes RNA polymerase to temporarily stall. The weak interactions between the RNA poly-uracil and DNA poly-adenine chain further destabilize the DNA-RNA duplex, which subsequently unwinds, dissociates from RNA polymerase (Wilson & von Hippel, 1995) and causes transcription to stop. b) Inhibition of translation. In absence of a RS ligand, the RBS is accessible for the ribosome to bind and initiate translation. Upon ligand binding, the expression platform structurally rearranges in a way that makes the RBS inaccessible, which prevents translation.



So far, over a dozen classes of RSs have been identified in various bacteria (Montange & Batey, 2008). In human pathogens, for instance, thiamine pyrophosphate (TPP), flavin mononucleotide (FMN), adenosylcobalamin (AdoCbl), purine (adenine or guanine), S-adenosylmethionine (SAM), lysine, glucosamine-6-phosphate (GlcN6P), glycine or pre-queuosine 1 ($preQ_1$) binding RS were found (see **Table 1**) (Blount & Breaker, 2006).

Table 1. Selection of human pathogenic bacteria (modified after Blount & Breaker, 2006). Black ticks indicate presence of a RS. Red ticks indicate that at least one of the RS-regulated genes is crucial for viability or pathogenicity.

Riboswitch target metabolite	TPP	FMN	AdoCbl	Purine	SAM1	SAM2	SAM3	Lysine	GlcN6P	Glycine	$PreQ_1$
Acinetobacter baumannii	✓	✓	✓	-	-	-	-	-	-	✓	-
Bacillus anthracis	✓	✓	✓	✓	✓	-	-	✓	✓	✓	✓
Brucella melitensis	✓	✓	✓	-	-	✓	-	-	-	✓	-
Enterococcus faecalis	✓	-	✓	✓	-	-	✓	✓	✓	-	✓
Escherichia coli	✓	✓	✓	-	-	-	-	✓	-	-	-
Francisella tularensis	✓	✓	-	-	-	✓	-	-	-	-	✓
Haemophilus influenza	✓	✓	-	-	-	-	-	✓	-	✓	✓
Helicobacter pylori	✓	-	-	-	-	-	-	-	-	-	-
Listeria monocytogenes	✓	✓	✓	✓	✓	-	-	✓	✓	✓	✓
Mycobacterium tuberculosis	✓	-	✓	-	-	-	-	-	-	✓	-
Pseudomonas aeruginosa	✓	✓	✓	-	-	-	-	-	-	-	-
Salmonella enterica	✓	✓	✓	-	-	-	-	-	-	-	-
Staphylococcus aureus	✓	✓	-	✓	✓	-	-	✓	✓	✓	✓
Streptococcus pneumoniae	✓	✓	-	✓	-	-	✓	-	-	✓	-
Vibrio cholerae	✓	✓	✓	-	-	-	-	✓	-	✓	-
Yersinia pestis	✓	✓	✓	-	-	-	-	-	-	-	-

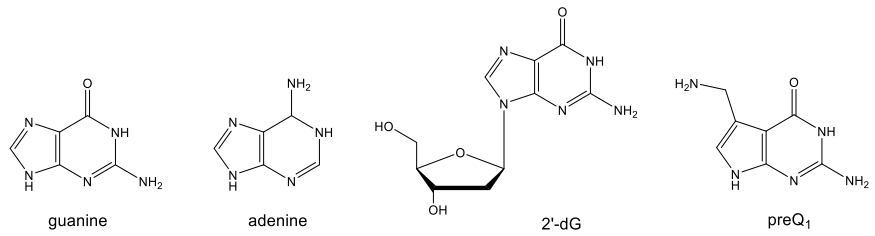
Some antimicrobial compounds with previously unknown mode of action recently have been shown to target some of these RSs (Deigan & Ferre-D'Amare, 2011). The thiamine analog pyrithiamine, for example, has an antimicrobial effect on bacteria and fungi, which is based on pyrithiamine binding to a TPP RS and subsequent repression of thiamine biosynthesis and transport (Sudarsan *et al.*, 2005). Similarly, roseoflavin, an analog of riboflavin and FMN, inhibits growth of Gram-positive bacteria by binding to an FMN RS and causing a reduced downstream gene expression (Serganov *et al.*, 2009). Even though, most of these RSs have been found in prokaryotic species, the TPP RS has also been identified in plants and fungi (Bocobza & Aharoni, 2008) and it is expected, that more eukaryotic RSs will be identified in the near future (Sudarsan *et al.*, 2003).

These examples show the significance of RSs as targets in antimicrobial drug development and the arising need for innovative molecules that are able to specifically bind to structured RNA elements.

II.1.1.1. Purine Riboswitches

Purine RSs are currently the most extensively studied RSs (Stoddard *et al.*, 2008). They can be subcategorized into RSs that sense guanine, adenine, 2'-deoxyguanosine (2'-dG) or the queuosine precursor pre-queuosine 1 ($preQ_1$) (Kim & Breaker, 2008) (see **Figure 2**).

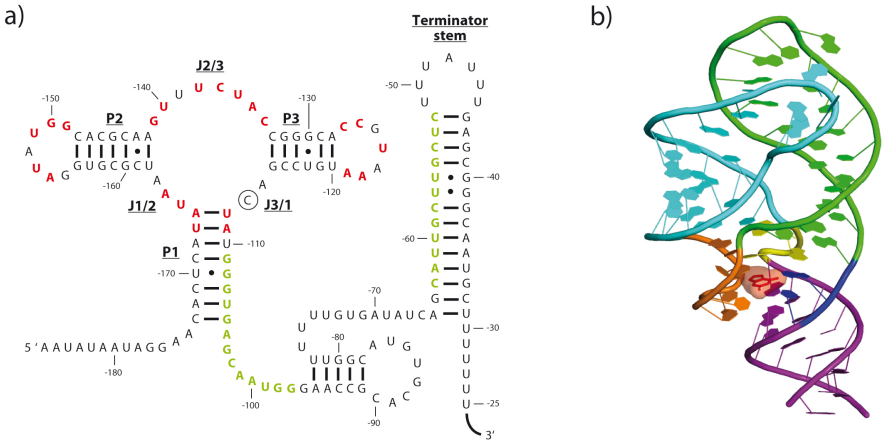
Figure 2. Purine RS ligands.



The $preQ_1$ RS aptamer is the smallest known so far and can be formed of less than 35 nucleotides (Roth *et al.*, 2007). In guanine- and adenine-sensing RSs, the aptamer consists of a three-way junction which is formed by the pairing nucleotide stems P1, P2 and P3 with P2 and P3 usually encompassing 7 bp (Kim & Breaker, 2008) (see **Figure 3**). Ligand recognition is promoted by the highly conserved nucleotides within the hairpin loops L2 and L3 and the junctions J1/2, J2/3 and J3/1. The aptamer nucleotides form a cavity in which the ligand is almost completely enveloped (Kim *et al.*, 2009). Both, guanine and adenine RSs, are very similar and show several interactions between the RS and the ligand. But only one nucleotide is crucial for the differentiation between guanine and adenine sensing. Within the J3/1 junction there is a critical cysteine or uracil residue (C-113 in the *B. subtilis xpt* guanine RS and U-113 in the *B. subtilis add* adenine RS), which forms a Watson-Crick base pairing interaction with the ligand and thus strictly discriminates between the binding of guanine or adenine (Serganov *et al.*, 2004). In the past years, this structural knowledge has allowed the specific construction of non-natural

compounds that selectively bind to guanine (Gilbert *et al.*, 2006; Mulhbacher *et al.*, 2010) and adenine (Mandal & Breaker, 2004a; Gilbert *et al.*, 2009) RSs.

Figure 3. Purine RS secondary and crystal structure (modified after Serganov *et al.*, 2004). a) Secondary structure of the *B. subtilis xpt* guanine RS. Highly conserved nucleotides are shown in red. Nucleotides that are involved in the formation of an anti-terminator stem in the absence of a RS ligand are highlighted in green. The circled cytidine is the critical nucleotide for ligand recognition and is exchanged for a uracil in adenine-sensing RSs (C-113 → U-113; numbering relative to *xpt* translation start, which is denoted as +1). b) Crystal structure of the *B. subtilis xpt* guanine aptamer domain in complex with guanine (red) (PDB code 1Y27). The different pairing regions and associated loops and junctions are colored according to the following scheme: P1 purple, J1/2 yellow, P2 cyan, J2/3 orange, P3 green, J3/1 blue.

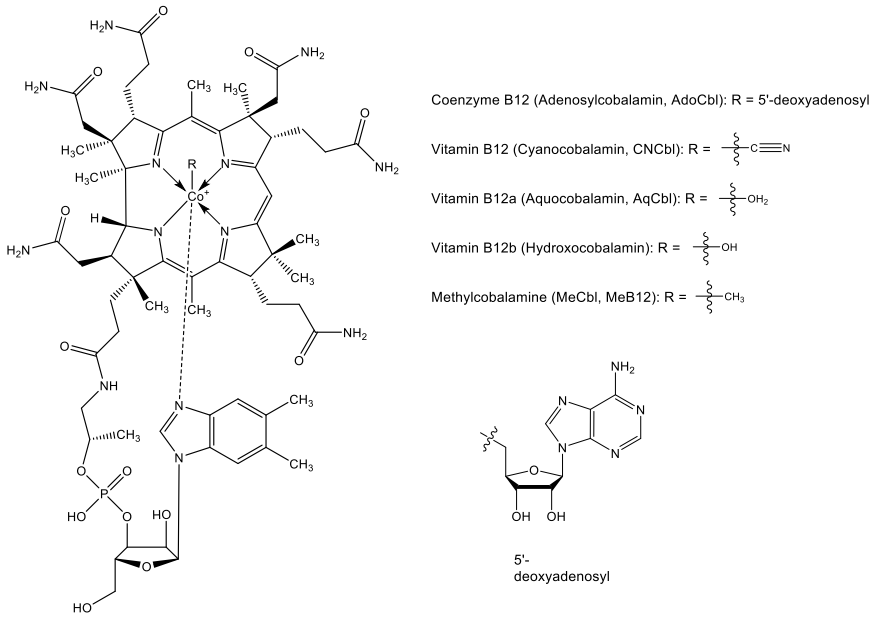


II.1.1.2. Cobalamin Riboswitches

Cobalamin (Cbl) RS are among the largest RSs known and possess aptamer domains that consist of more than 200 nucleotides. They are widely distributed among various prokaryotes (Nahvi *et al.*, 2004; Blount & Breaker, 2006) (see **Table 1**) which can be explained by the ancient origin of Cbl-derived coenzymes and their critical role in prokaryotic metabolism (Lawrence & Roth, 1996; Roth *et al.*, 1996). Already in 1991, Kadner and coworkers described, that a nucleotide sequence within the leader region of the *E. coli btuB* gene is responsible for the regulation of gene expression on a

transcriptional and translational level (Lundrigan *et al.*, 1991). In 2002, this nucleotide sequence has been referred to as cobalamin (Cbl) RS for the first time. It has been shown, that the *E. coli* Cbl RS upstream of the *btuB* gene, which encodes the Cbl outer membrane transporter, recognizes adenosylcobalamin (AdoCbl) with great selectivity against other Cbl derivatives like cyanocobalamin (CNCbl) or methylcobalamin (MeCbl) (Nahvi *et al.*, 2002) (see **Figure 4**).

Figure 4. Cobalamin derivatives.

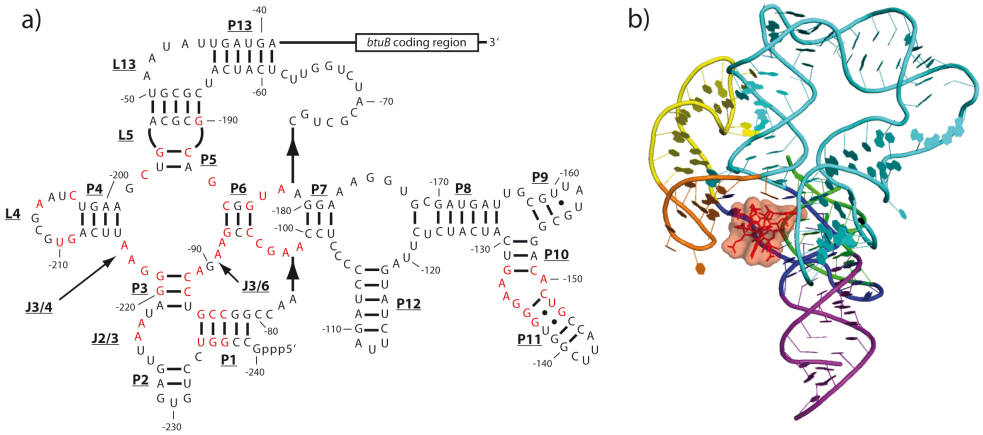


Phylogenetic analysis and secondary structure prediction allowed the proposition of secondary structure models, but three-dimensional structures of cobalamin riboswitches bound to their effector molecules have not been accessible until recently. In 2012, crystal structures of Cbl RS from *Symbiobacterium thermophilum* (Peselis & Serganov, 2012), *Thermoanaerobacter tengcongensis* and the *env8* variant of a Pacific Ocean metagenome (Johnson *et al.*, 2012) in complex with their respective ligands, revealed structural and mechanistic details of Cbl RS-mediated regulation of gene expression.

In general, Cbl RS structures are comprised of a core domain, which contains a conserved four-way junction (P3-P6) and a peripheral domain, which contains a conserved stem

(P10-P11) and a non-conserved accessory hairpin (P13, L13) (see **Figure 5**). As expected, the binding pocket for the respective Cbl molecule is constructed by the most conserved sections of the aptamers (P3-P6, P11). Ligand recognition is achieved primarily through van der Waals interactions and only a few hydrogen bonds between the pocket-forming nucleotides and the moieties protruding from the corrin ring. Especially, gene regulation by AdoCbl-binding RSs depends on the formation of a kissing-loop (KL) interaction between L5 of the core domain and L13 of the accessory hairpin which typically carries the RBS in Cbl RSs that regulate gene expression on a translational level. It has been shown, that formation of the ligand-bound core domain orients the P5-L5 hairpin in a way that makes L5 accessible for tertiary pairing with L13. Mutational studies on L13 of the *E. coli btuB* AdoCbl RS that allow perfect Watson-Crick base pairing with L5 showed an AdoCbl-independent KL formation, corresponding to a constitutively repressed gene expression (Johnson *et al.*, 2012).

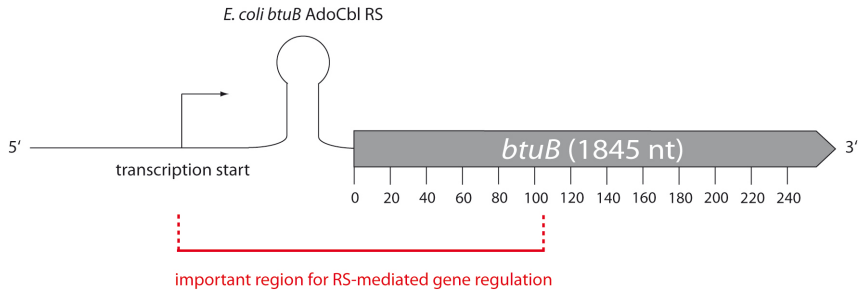
Figure 5. Cbl RS secondary and crystal structure. a) Secondary structure of the *E. coli btuB* AdoCbl RS (Nahvi *et al.*, 2004). Highly conserved nucleotides are shown in red. The nucleotide numbering is relative to the *btuB* translation start, which is denoted as +1. b) Crystal structure of the *S. thermophilum* AdoCbl RS in complex with AdoCbl (red) (PDB code: 4GXY) (Peselis & Serganov, 2012). The different pairing regions and associated loops and junctions are colored according to the following scheme: P1 purple, P2 green (background), P3 blue, P4 yellow, P5 orange, P6-P11 cyan.



It has been reported that the initial part of the *E. coli btuB* gene up to around nucleotide +110 (numbering relative to start of *btuB* translation, which is denoted as +1) is necessary

for efficient regulation of transcription (Lundrigan *et al.*, 1991; Franklund & Kadner, 1997; Ravnum & Andersson, 1997; Nou & Kadner, 1998) (see **Figure 6**). In spite of this fact, there are no crystal structures available that contain RS sequences including this translated regulatory region (TRR).

Figure 6. Important regions for RS-mediated regulation of gene expression. Besides the RS itself, the *btuB* TRR up to around nucleotide +110 is necessary for efficient gene regulation.

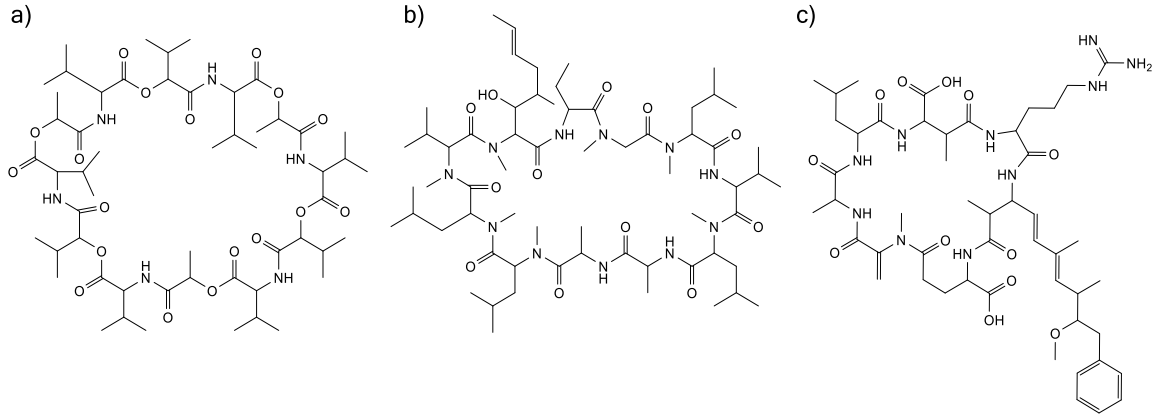


II.1.2. Cyclic Peptides

Cyclic peptides (CPs) are highly potent bioactive compounds which are of great interest for modern drug development. Due to their cyclic structure, they offer a wide range of bio- and physicochemical benefits. First of all, CPs have no free C- and N-termini. This confers increased structural rigidity and makes especially head-to-tail cyclized peptides remarkably stable against proteolytic degradation, compared to their linear or disulfide-cyclized counterparts (Satoh *et al.*, 1996). Moreover, reduced conformational flexibility can increase binding affinity of CPs towards their target molecules because of lowered conformational entropy loss upon binding (Satoh *et al.*, 1996; Kinsella *et al.*, 2002; Craik, 2006; Oyston *et al.*, 2009; Joo, 2012). Furthermore, cyclized peptide structures resist high temperatures and conditions of extreme pH (Craik, 2006), which is not only an advantage for chemical synthesis but also for application. These multifaceted molecules exhibit a wide range of biological activities and can act as antibiotics (Strieker & Marahiel, 2009), toxins (van Apeldoorn *et al.*, 2007) or immunosuppressives (Laupacis *et al.*, 1982) (see

Figure 7). Even lead compounds against HIV (Hayouka *et al.*, 2010) and hepatitis C (Manna *et al.*, 2013) have already been identified.

Figure 7. Selection of naturally occurring CPs. a) valinomycin (antibiotic), b) cyclosporine A (immunosuppressive). c) microcystin-LR (toxin).



In spite of the large spectrum of biological activities, CP drugs are still relatively unexplored. This may be due to the fact that generation and screening of CP libraries have not been trivial. Until recently, CP synthesis was mainly conducted via conventional combinatorial chemistry. Admittedly, this technique allows generation of libraries with great chemical diversity (Scott *et al.*, 2001). But small library sizes of only ~10^2 members (Kohli *et al.*, 2002; Qin *et al.*, 2004; Liu *et al.*, 2009) and the fact, that combinatorial synthesis requires laborious post-synthesis identification of active compounds, makes this procedure inadequate for high-throughput screening in modern drug discovery.

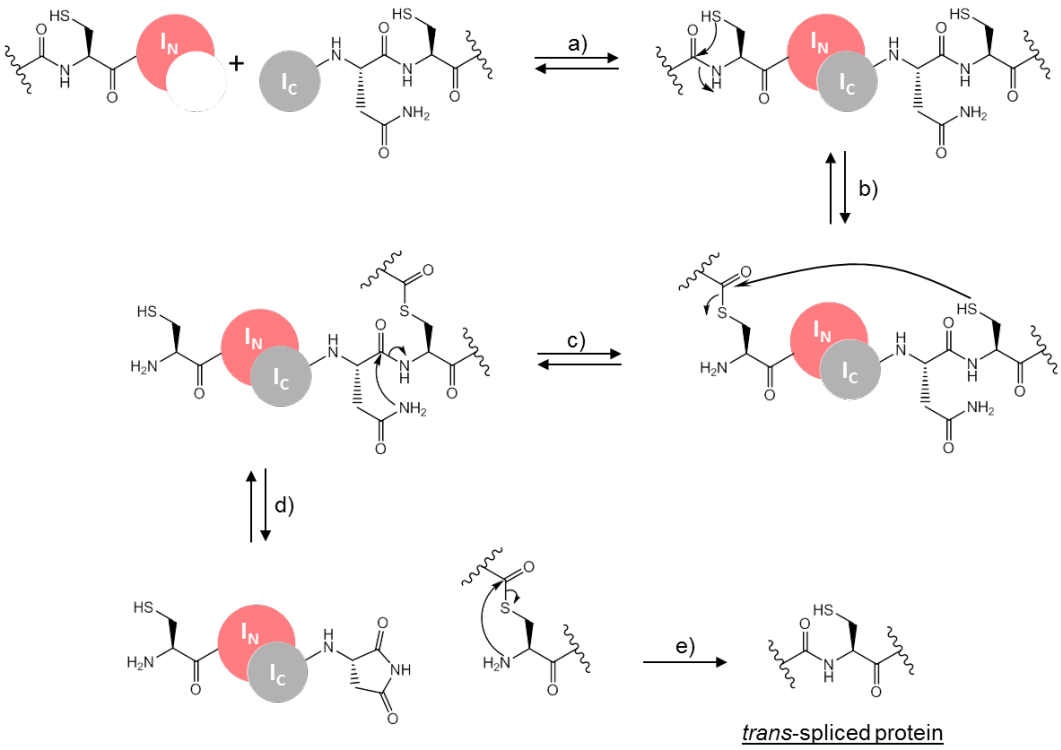
Recent methods, however, use a more elegant approach to create and screen CP libraries. They benefit from the possibility to genetically encode CP libraries in bacteria (Scott *et al.*, 1999), yeast (Kritzer *et al.*, 2009) or human cells (Kinsella *et al.*, 2002), which allows the generation of considerably larger libraries (see **II.1.3**). Moreover, biological synthesis can be readily combined with *in vivo* screening assays and therefore enable a straightforward identification of active members towards a chosen target (Tavassoli & Benkovic, 2007).

Bio-stable CP libraries that are expressed *in vivo* can be constructed via split-intein circular ligation of peptides and proteins (SICLOPPS) (Scott *et al.*, 1999). This method has been engineered from a naturally occurring split-intein mediated *trans*-splicing mechanism for the head-to-tail cyclization of peptides and proteins.

II.1.3. Split-Intein Circular Ligation of Peptides and Proteins (SICLOPPS)

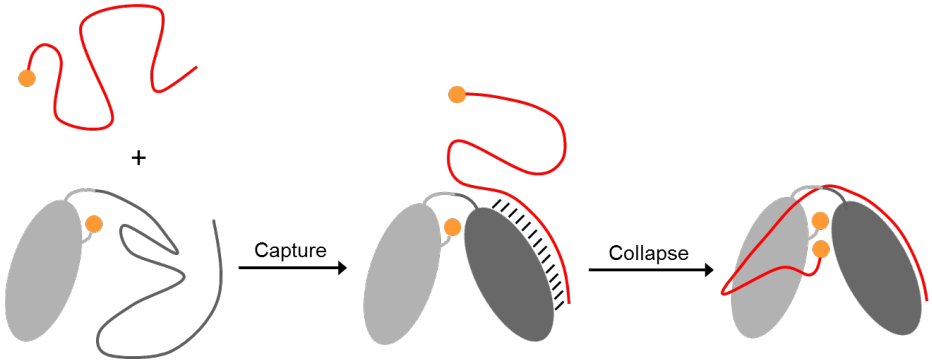
Inteins (*int*ernal prot*eins*) are protein-intervening sequences that catalyze their own removal from a precursor fusion protein while ligating the adjacent polypeptide sequences (exteins) through a native peptide bond (Saleh & Perler, 2006). There are two kinds of inteins: the commonly occurring contiguous inteins and the rarer split-inteins (Shah *et al.*, 2013). The first split-intein was discovered in 1998 by Liu and coworkers (Wu *et al.*, 1998). They found that the N- and C-terminal domains of the catalytic α subunit of DNA polymerase III (DnaE) from the cyanobacterium *Synechocystis* sp. strain PCC6803 (*Ssp*) are located on two intermitted genes on opposite DNA strands. The N-terminal and C-terminal domain include a 123 (Int_N/I_N) and 36 (Int_C/I_C) amino acid intein sequence, respectively (Wu *et al.*, 1998). The functional DnaE subunit is excised from this precursor protein via a *trans*-splicing mechanism (see **Figure 8**) (Shah & Muir, 2011) in which two separately transcribed and translated fragments (I_N and I_C) spontaneously associate to form an active protein ligase.

Figure 8. *Trans*-splicing mechanism (modified after Shah & Muir, 2011). a) Formation of an active intein. b) An N-to-S acyl shift at the extein-I_N junction results in the formation of a thioester intermediate. c) The thioester undergoes a transesterification reaction with a serine side chain at the I_C-extein junction. d) Succinimide formation within the Asn side chain promotes the release of the *trans*-spliced protein as a thiolactone. e) The final protein in its lactam form is generated via a spontaneous S-to-N acyl shift.



While the chemical mechanism of *trans*-splicing reactions is well understood, it has remained unclear, how the spatially divided N- and C-terminal inteins form an active protein ligase. Only recently, the mechanism of split-intein assembly was elucidated on the basis of biophysical studies on the *Nostoc punctiforme* (*Npu*) split-intein. It is postulated that the intein fragments associate via a "Capture and Collapse" mechanism (see **Figure 9**) (Shah *et al.*, 2013) in which the N-intein fragment has been shown to possess a unique two-part structure with one well-folded region (I_{N1}) tethered to an intrinsically disordered region (I_{N2}) of equal length. The extended and highly anionic I_{N2} captures the completely disordered and highly cationic I_C via electrostatic interactions, which leads to the formation of a structured and native-like intermediate. The flexible regions of I_C further collapse and intercalate between the two I_N regions. Hydrophobic interactions between I_C and the well-ordered I_{N1} and electrostatic interactions between I_C and I_{N2} stabilize this intertwined final structure. The entanglement of I_C and I_N does not only prevent the intein fragments from rapidly dissociating, but also brings the extein fragments in close proximity, which allows the *trans*-splicing reaction to take place.

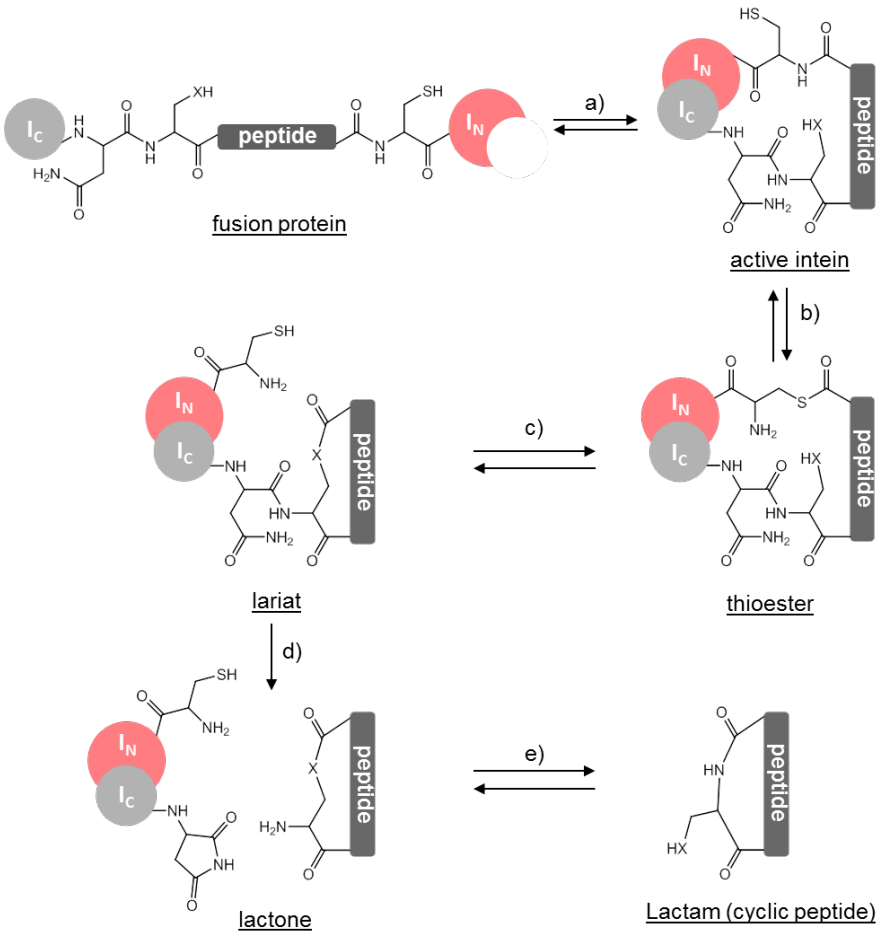
Figure 9. "Capture and collapse" mechanism of split-intein assembly (modified after Shah et al., 2013). The disordered I_C (red) is captured via electrostatic interactions by the extended I_{N2} region (dark grey). This capturing results in the formation of a structured and native-like intermediate. The flexible regions of I_C collapse and intercalate between I_{N1} (light grey) and I_{N2} resulting in an intertwined complex, which prevents rapid dissociation.



Split-intein mediated splicing reactions can also be used to generate backbone cyclized peptides. In 1999, Benkovic and coworkers showed that an inverted order of intein fragments around a polypeptide of interest, (I_C-target peptide-I_N) results in the formation of an active *cis*-intein that splices the precursor protein into a head-to-tail cyclized peptide product (Scott *et al.*, 1999) (see **Figure 10**).

In order to create CP libraries, the target peptide needs to be encoded by degenerate oligonucleotides and incorporated between the sequences for I_C and I_N. In general, any target sequence can be introduced into the SICLOPPS vector. The only prerequisite to ensure peptide cyclization is a serine or cysteine residue in the position of the first amino acid within the target peptide. For instance, a hexameric CP library with five variable amino acids can theoretically result in a library size of 34 million members at DNA level. Conveniently, this number lies in within the transformation efficiency of *E. coli* (Tavassoli & Benkovic, 2007). Screening of SICLOPPS-generated peptide libraries have already resulted in the discovery of protease inhibitors (Young *et al.*, 2011), methyltransferase inhibitors (Naumann *et al.*, 2008) and molecules that reduce the cellular pathology of Parkinson's disease (Kritzer *et al.*, 2009) (Shah & Muir, 2011).

Figure 10. SICLOPPS mechanism (modified after: Scott *et al.*, 1999; Tavassoli & Benkovic, 2007). a) Formation of an active intein. Folding of the expressed fusion protein brings the N- and C-terminal intein fragments in close proximity, allowing an active protein ligase to form. b) An N-to-S acyl shift at the junction between I_N and target peptide (TP) results in the formation of a thioester intermediate. c) The thioester undergoes a transesterification reaction with the nucleophilic side chain residue of the first TP amino acid (X = S (Cys) or O (Ser)) resulting in a lariat intermediate. d) A cyclization reaction within the Asn side chain promotes the release of the cyclic peptide as a lactone. e) The final cyclic peptide product in its thermodynamically favored lactam form is generated via a spontaneous X-to-N acyl shift.

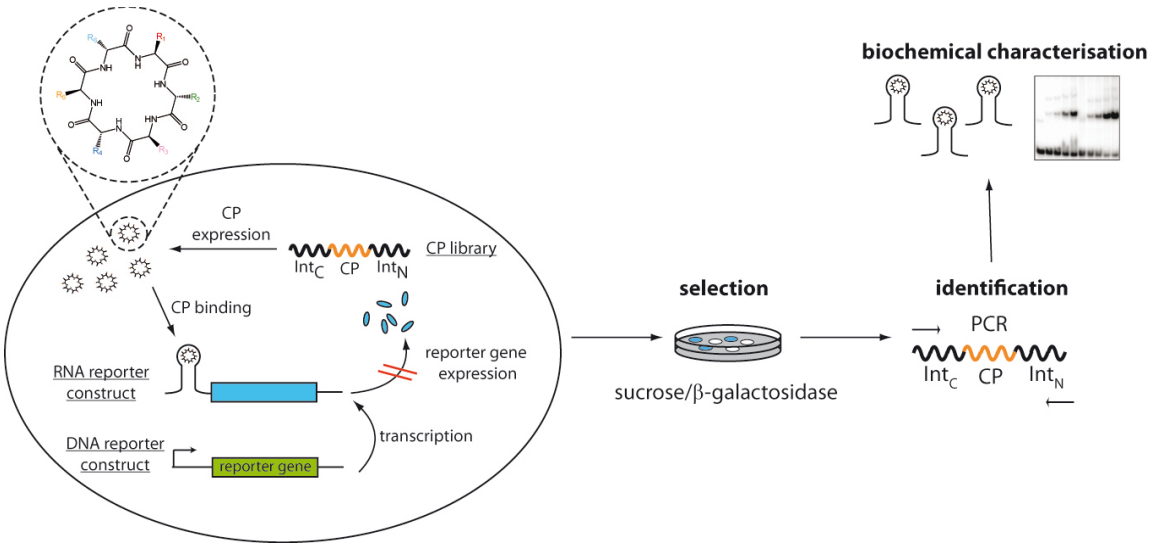


II.1.4. Objectives

Due to the rising number of antibiotic resistant bacteria, there is a high demand for innovative antibiotics with novel modes of action. The aim of this project was to provide a basis for the discovery of antibiotics with a novel mode of action that specifically target RNA and not proteins. In order to achieve this, an *in vivo* screening assay in *E. coli* was developed, which allows for the identification of CPs from an expressed SICLOPPS library that specifically bind to RSs of pathogenic bacteria. The CP-RS interactions cause an inhibition downstream gene expression. Considered the fact, that bacterial RSs often regulate genes that are vital or confer virulence to the bacterium, this inhibition of gene expression is the basis for the potential antibiotic effect. The efficiency of RS-mediated regulation of gene expression was monitored with the help of reporter genes that are under RS control. By the use of this *in vivo* screening system it should be possible to identify single active members from a library of 34 million CPs at DNA level. The project outline is depicted in **Figure 11**. Defined subgoals of the thesis were:

1.) Cloning of a hexameric expressed CP library in *E. coli*
2.) Cloning of a reporter construct carrying the reporter genes
3.) Development of an *in vivo* screening assay
4.) Chemical synthesis of selected CPs

Figure 11. Project outline.



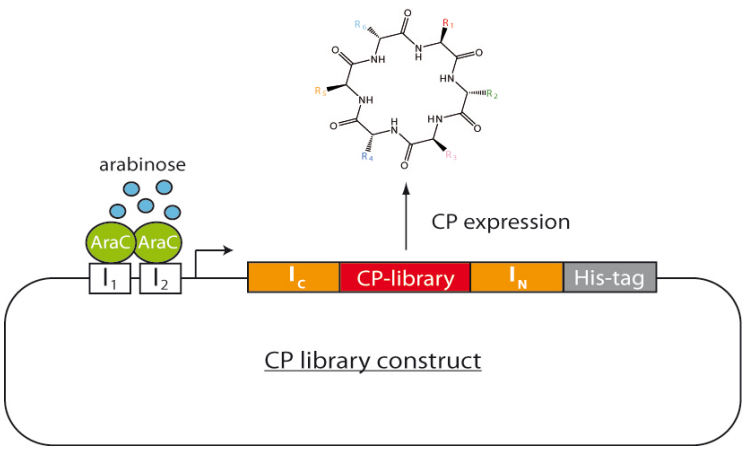
II.2. Results and Discussion

II.2.1. Expressed CP Library

II.2.1.1. Generation of the CP Library Construct

The expressed CP library was generated as explained in detail by (Tavassoli & Benkovic, 2007) and as indicated in section **IV.2.1.13**. The CP library construct consists of an arabinose-inducible araBAD promoter (P_{BAD}), followed by in frame sequences of the *Ssp* DnaE split inteins I_C (102 nt, 737˙312-737˙211, Gene ID: 953863) and I_N (369 nt, 3˙560˙359-3˙560˙727, Gene ID: 951978), which are intermitted by the CP library sequence. A hexahistidin-tag (His_6-tag) is cloned in frame with the I_N sequence in order to allow purification of the construct via Ni-NTA chromatography (see **Figure 12**).

Figure 12. CP library construct. The *Ssp* DnaE I_C, the CP library and the *Ssp* DnaE I_N are cloned in frame downstream of the araBAD promoter. Addition of arabinose induces the expression of the CP library construct. A His_6-tag at the 3' region of I_N allows for a purification via Ni-NTA chromatography.

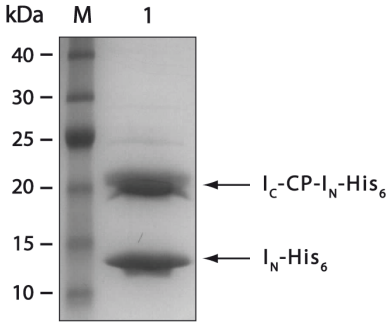


The expressed CP library consists of 6 amino acids with the first amino acid being cysteine, followed by 5 randomized amino acids. A hexameric CP library with five variable amino acids was chosen because the theoretical number of 34 million library members at DNA level lies within the transformation efficiency of *E. coli* (Tavassoli & Benkovic, 2007).

II.2.1.2. Verification of CP expression in *E. coli*

The expression of the CP library construct and formation of CPs in *E. coli* was verified by purification of the His_6-tagged intein. Therefore, cells carrying the CP library construct were cultivated and the expression was induced with 0.2 % (w/v) L-arabinose as described in **IV.2.2.1**. After overnight expression, cells were harvested, and lysed. The intein was purified via Ni-NTA Spin Columns (Qiagen, Hilden, DE) and the elution fractions were analyzed by SDS-PAGE (see **Figure 13**). The results show two protein bands with ~19 and ~14 kDa which match the sizes of the 19 kDa (exact size depending on CP sequence) unspliced CP library construct (I_C-CP-I_N-His_6) and the 14.8 kDa fully spliced product (I_N-His_6).

Figure 13. SDS-PAGE analysis of CP library construct expression in *E. coli*. In case of an incomplete CP formation, the whole CP library construct (I_C-CP-I_N-His_6, ~19 kDa) is visible on the SDS-PAGE gel. In case of a complete cyclization and release of a CP, the His_6-tagged I_N part of the intein is visible on the gel (I_N-His_6, 14.8 kDa). Bands for both constructs are shown in lane 1.



A direct verification of CP formation was possible via mass spectrum (MS) analysis. This method, however, required the expression of only one distinct CP of known molecular mass. Therefore, *E. coli* cells were transformed with the vector carrying the CP library construct as described in **IV.2.1.8**. Single colonies were picked, cultivated and their plasmids were isolated. The DNA and thus amino acid sequence of the CP was determined by sequencing of the CP library reporter construct. Exemplary amino acid sequences of expressed CPs were: CLNSLS, CVFTAD, CVGLLG, CASFDR, CCNFRN, CVGTVA and CPNCLQ.

The CP with the sequence CLNSLS (cyclo-CLNSLS, exact mass: 617.28 Da) was used as a representative peptide for further analysis. Cells carrying the CP library construct of cyclo-CLNSLS were cultivated and induced with 0.2 % (w/v) L-arabinose as described in **IV.2.2.1**. After overnight expression, cells were harvested, and lysed. The crude lysate was subject to an extraction with *n*-BuOH as described in **IV.2.2.2**. Detection of cyclo-CLNSLS via LC-ESI-MS analysis was achieved by comparing the distinct retention time of a chemically synthesized version of cyclo-CLNSLS to the retention time of the expressed cyclo-CLNSLS in the *n*-BuOH extract. The LC-ESI-MS analysis showed both target molecular ion peaks (*m/z* = 618.29, $M + H^+$, see **Figure 14**) at identical retention times (see **Figure 15**), which confirmed the formation of cyclo-CLNSLS *in vivo*.

Figure 14. LC-ESI-MS spectrum of the synthesized cyclo-CLNSLS recorded on a Thermo Scientific LTQ Orbitrap XL.

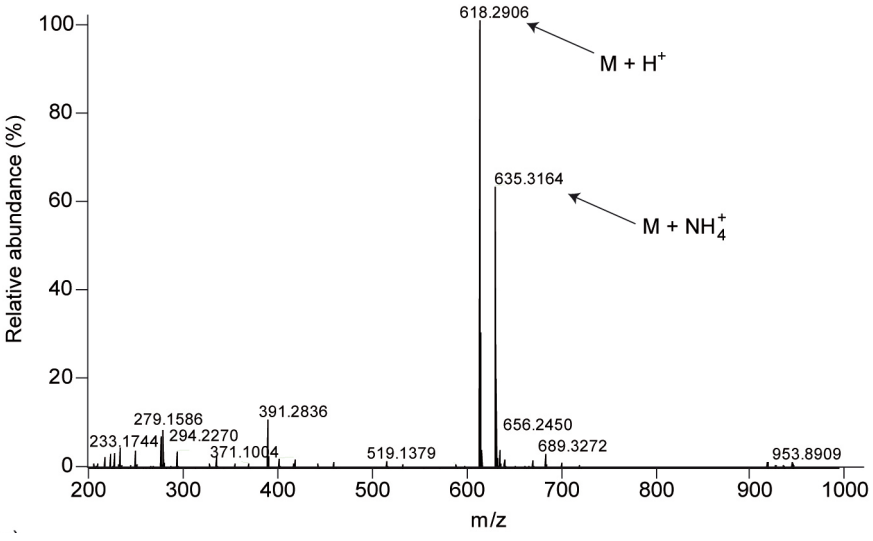
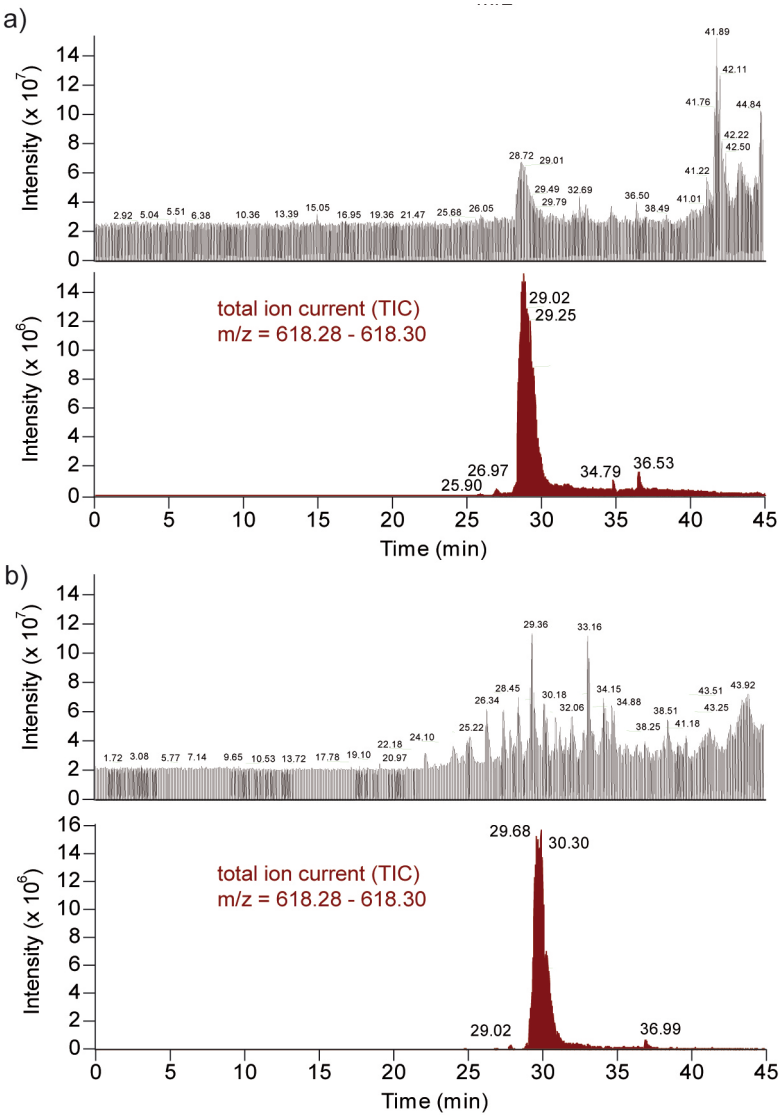


Figure 15. Comparison of cyclo-CLNSLS retention times on a Thermo Scientific LTQ Orbitrap XL. a) Synthesized cyclo-CLNSLS, b) *n*-BuOH extract, containing expressed cyclo-CLNSLS.



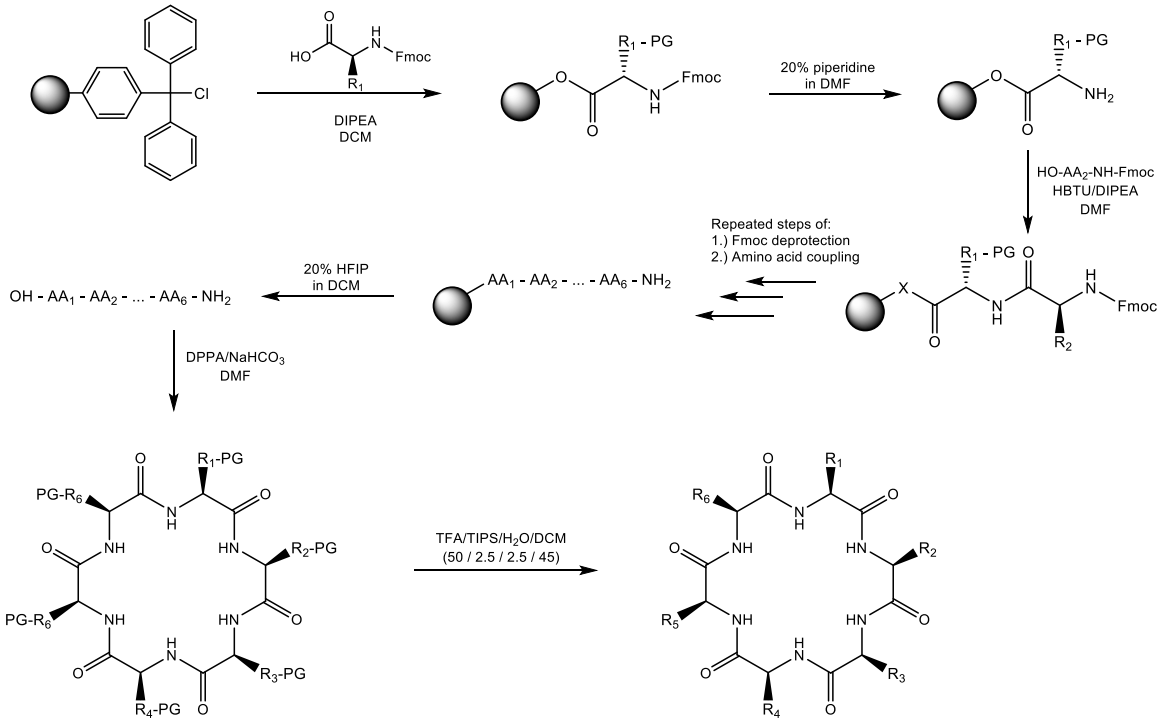
II.2.2. Chemical Synthesis of cyclo-CLNSLS

Since the *n*-BuOH cell extract of cyclo-CLNSLS-expressing cells contained a large number of low molecular weight compounds, there was the need for a synthesized reference CP with the same amino acid sequence that allows for the determination of the

CP's distinct retention time and thus facilitates MS analysis of the expressed cyclo-CLNSLS.

The method of choice for the chemical synthesis of peptides is solid-phase peptide synthesis (SPPS). In the special case of CP synthesis, a linear peptide precursor is first synthesized on a solid support, then cleaved and subsequently cyclized in solution. In SPPS, Boc (*tert*-butyloxycarbonyl) and Fmoc (9-fluorenylmethyloxycarbonyl) SPPS strategies (referring to the protection of the Nα-amino group) are reliable standard procedures. Even though both have their advantages and disadvantages, Fmoc SPPS was chosen for the synthesis of cyclo-CLNSLS, since the Nα-amino deprotection and resin cleavage take place under very mild reaction conditions, unlike in Boc SPPS, which requires rather harsh treatment with hydrogen fluoride. TCP (trityl chloride polystyrene) resin was chosen as a solid support because it is compatible with the Fmoc strategy, suitable to immobilize carboxylic acids and enables cleavage of completely protected peptides under very mild acidic conditions using dilute acetic acid, HFIP (1,1,1,3,3,3-hexafluoro-2-propanol) or TFA (trifluoroacetic acid). The first amino acid, Fmoc-L-Ser(OtBu)-OH, was coupled to the TCP resin in the presence of DIPEA (*N*,*N*-diisopropylethylamine) in DCM (dichloromethane). Unreacted sites on the resin were capped by the addition of MeOH. Repeated steps of Fmoc deprotection (20 % (v/v) piperidine in DMF (dimethylformamide)) and amino acid coupling (Fmoc-L-Leu, Fmoc-L-Ser(OtBu), Fmoc-L-Asn(Trt), Fmoc-L-Leu, Fmoc-L-Cys(Trt)) were performed in DMF on a PS3 peptide synthesizer (Protein Technologies, Tucson, US) using HBTU (*O*-(1*H*-benzotriazol-1-yl)-*N*,*N*,*N'*,*N'*-tetramethyluronium hexafluorophosphate) as an activator and DIPEA as a base. Cleavage of the linear hexapeptide from the resin was performed under mild conditions using 20 % (v/v) HFIP in DCM. The fully side-chain protected linear hexapeptide was cyclized in DMF at a concentration of 3.4×10^{-3} M using DPPA (diphenylphosphoryl azide) and $NaHCO_3$. The final side-chain deprotection was achieved using TFA/TIPS (triisopropylsilane)/H_2O/DCM in volume ratios of 50 : 2.5 : 2.5 : 45. During this process highly reactive cationic species are generated from the protecting groups, which can, unless trapped, react with nucleophilic moieties of the peptide and result in unwanted side products. Therefore TIPS is added as a scavenger to the side-chain deprotection mixture in order to trap these ions and prevent side reactions.

Figure 16. Fmoc SPPS of cyclo-CLNSLS.



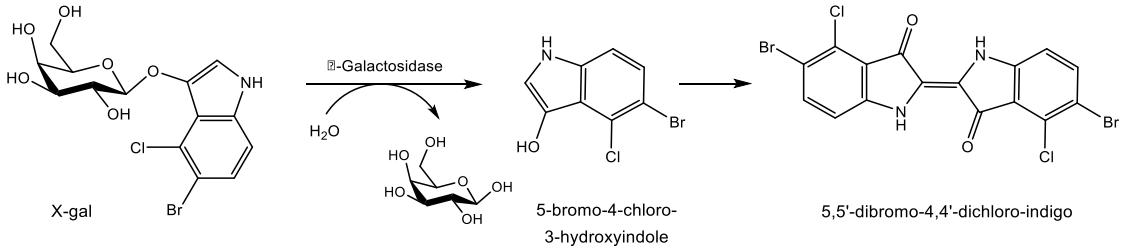
II.2.3. Reporter Constructs

II.2.3.1. Principle

The reporter constructs consist of the constitutive *glnS* (glutamyl-tRNA synthetase) promoter, the *S. aureus xpt* guanine RS or the *E. coli btuB* AdoCbl RS and the reporter genes *lacZ* and *sacB*. Hereby, *lacZ* encodes the enzyme β-galactosidase, which catalyzes the hydrolysis of β-galactosides, e.g. lactose, into monosaccharides. Artificial substrates can be cleaved into dyes, which allow the determination of β-galactosidase expression levels. The colorless substrate X-gal (5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside) is an analog of lactose and is cleaved into galactose and 5-bromo-4-chloro-3-

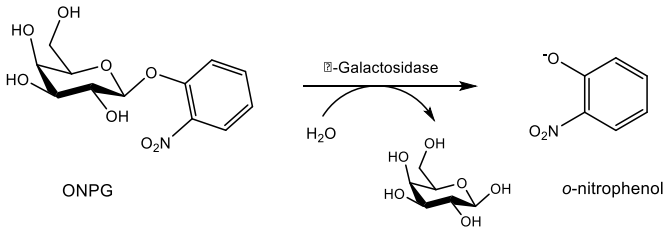
hydroxyindole. The latter spontaneously dimerizes and is oxidized into 5,5'-dibromo-4,4'-dichloro-indigo, an intensely blue product which is insoluble in water (Juers *et al.*, 2012). Therefore, X-gal is used in blue-white screenings on agar plates, allowing the qualitative detection of expressed β-galactosidase.

Figure 17. Reaction of β-galactosidase and X-gal.

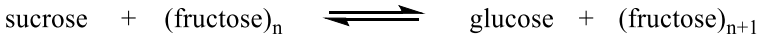


Another artificial substrate, the colorless ONPG (*ortho*-nitrophenyl-β-galactoside), is cleaved into galactose and *ortho*-nitrophenol, which is a yellow dye. The absorption of *o*-nitrophenol can be measured at a wavelength of 420 nm, which allows the quantitative determination of β-galactosidase expression levels.

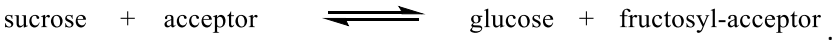
Figure 18. Reaction of β-galactosidase and ONPG.



The *B. subtilis sacB* gene encodes the enzyme levansucrase, which belongs to the family of glycosyltransferases. It catalyzes a transglycosylation reaction, in which the fructose moieties originating from the disaccharide sucrose are condensed to form levan polysaccharides (Dedonder, 1966). The latter are fructose polymers and toxic to *E. coli* cells. The transglycosylation reaction can be described by the equation



or by the general equation



When cells are grown without the addition of a RS ligand, the reporter genes *lacZ* and *sacB* are constitutively expressed and thus colonies on X-gal agar plates should be blue, β-galactosidase assay should show development of yellow *o*-nitrophenol and cells should not be able to grow on sucrose containing medium. On the other hand, when grown upon addition of a RS ligand, expression of the reporter genes should be inhibited. Thus, colonies on X-gal agar plates should be white, β-galactosidase assay should not show development of yellow *o*-nitrophenol and cells should be able to grow on sucrose containing medium.

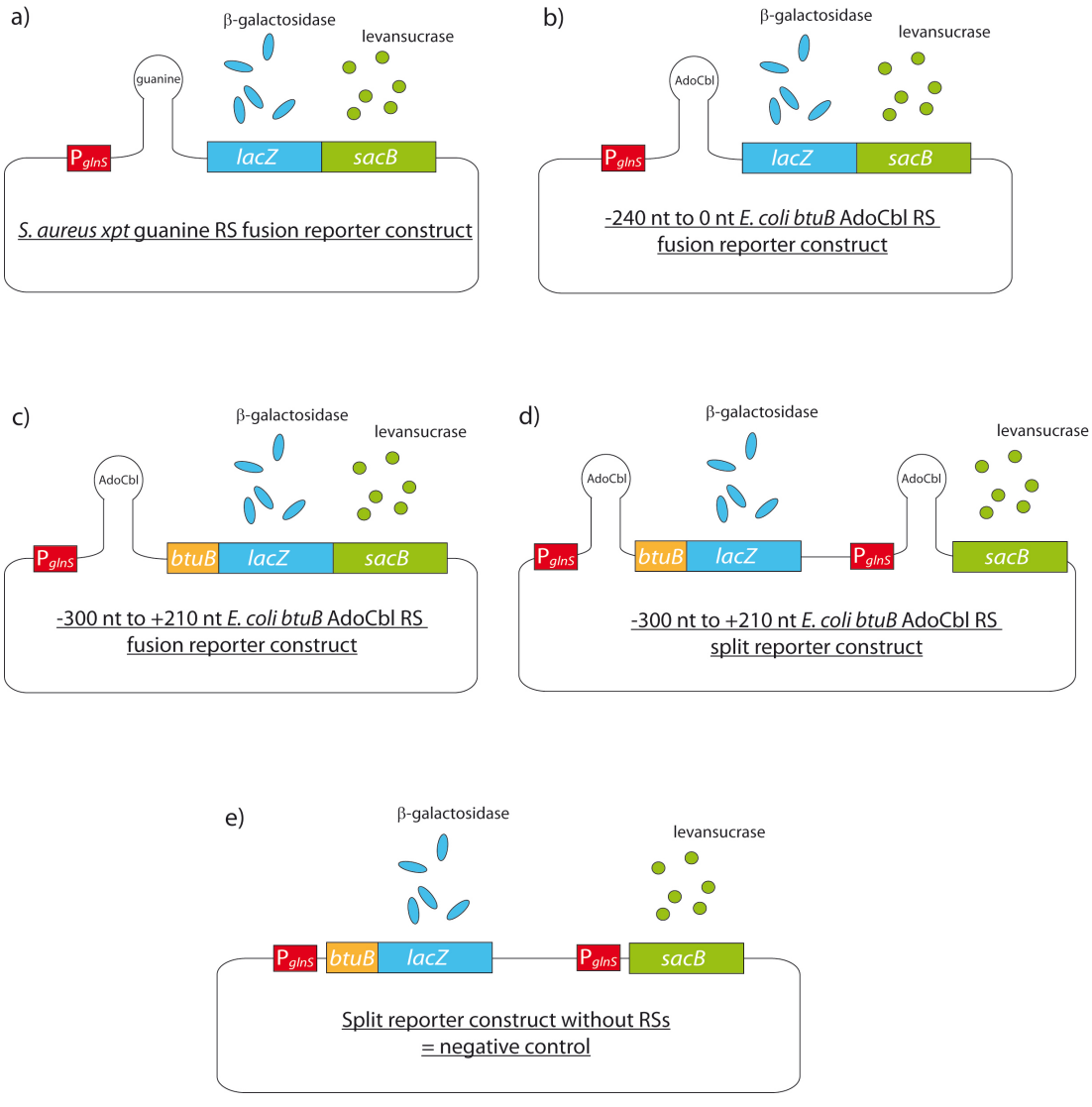
LacZ and *sacB* were chosen as reporter genes because they are well studied and widely used reporter genes in molecular biology.

II.2.3.2. Generation of Reporter Constructs

Two different compositions of reporter constructs were examined. In one approach, a *lacZ*/*sacB* fusion gene is under control of a single constitutive *glnS* promoter and RS, whereas in the second approach, each reporter gene is under control of its own *glnS* promoter and RS (see **Figure 19**). The *S. aureus xpt* guanine RS was tested with the downstream *lacZ*/*sacB* fusion gene, whereas the *E. coli btuB* AdoCbl RS was tested with both the fusion and the split reporter genes. The 3075 nt *lacZ* gene, which was PCR amplified from *E. coli* K12 genomic DNA (Gene ID: 945006), has an alternative Met start codon only six nucleotides after the original start codon. As the *lacZ* gene showed repeated mutations within this beginning region, the alternative Met was used as a start codon in all reporter constructs. The 1422 nt *sacB* gene, which was PCR amplified from *B. subtilis* genomic DNA (Gene ID: 936413), has an internal HindIII restriction site. In order to facilitate cloning, this restriction site was eliminated by site-directed mutagenesis

(see **IV.2.1.10**). The cloning strategy of all constructs is described in sections **IV.2.1.14**, **IV.2.1.15** and **IV.2.1.16**.

Figure 19. Reporter constructs. a) *S. aureus xpt* guanine RS fusion reporter construct, b) -240 nt to 0 nt *E. coli btuB* AdoCbl RS fusion reporter construct, c) -300 nt to +210 nt *E. coli btuB* AdoCbl RS fusion reporter construct, d) -300 nt to +210 nt *E. coli btuB* AdoCbl RS split reporter construct. e) The split reporter construct without RSs serves as a negative control in the reporter gene assays.



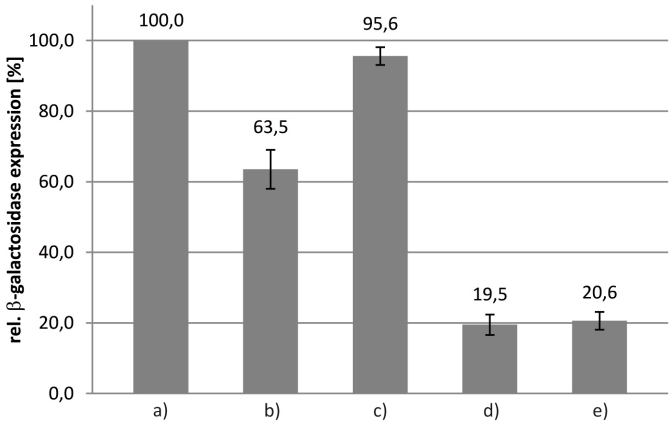
II.2.3.3. RS-mediated Regulation of Reporter Genes

In order to prove that the chosen RSs regulate expression of the downstream reporter genes, the strength of reporter gene expression was determined in presence and absence of the RSs' natural ligands guanine (250 μg/ml) and AdoCbl (10 μM).

II.2.3.3.1. Regulation of lacZ

RS-control of the *lacZ* gene was verified by spectroscopically quantifying the level of β-galacosidase expression. Therefore, β-galactosidase assays were performed as described in **IV.2.3.1**. The calculated values for relative β-galactosidase expression given in percent were obtained by normalization against the Miller Units from assay conditions without RS ligands. This was done on the assumption that cell growth in absence of a RS ligand results in constitutive, maximum β-galactosidase expression (100 %), whereas the presence of a RS ligand correspondingly reduces β-galactosidase expression (see **Figure 20**). Results of the assay show that the *S. aureus xpt* guanine RS controls gene expression on a moderate level, as it reduces β-galactosidase expression only to 63.5 %. A comparison of the -240 nt to 0 nt and the -300 nt to +210 nt *E. coli btuB* AdoCbl RS clearly showed that parts of the *btuB* gene are necessary for efficient downstream regulation of gene expression, as reported previously (Lundrigan *et al.*, 1991; Franklund & Kadner, 1997; Ravnum & Andersson, 1997; Nou & Kadner, 1998). The -240 nt to 0 nt *E. coli btuB* AdoCbl RS still showed 95.6 % β-galactosidase expression, whereas the -300 nt to +210 nt *E. coli btuB* AdoCbl RS controls gene expression in both the fusion and the split reporter construct on a high level, reducing β-galactosidase expression to about 20 %. Since nucleotides +0 to +210 of the *btuB* gene are clearly necessary for the RS function, only the -300 nt to +210 nt *E. coli btuB* AdoCbl RS fusion and split reporter constructs were used for further experiments.

Figure 20. Relative β-galactosidase expression in the presence or absence of a RS ligand. a) In absence of a RS ligand, the β-galactosidase expression is at maximum and was normalized to 100 %. b) Experiments with the *S. aureus xpt* guanine RS fusion reporter construct showed a reduction of β-galactosidase expression to 63.5 % in presence of 250 µg/ml guanine in the culture medium. c) In presence of 10 µM AdoCbl, the -240 nt to 0 nt *E. coli btuB* RS fusion reporter construct still has a β-galactosidase expression of 95.6 %. d) In contrast, the -300 nt to +210 nt *E. coli btuB* AdoCbl RS fusion construct had a reduced β-galactosidase expression of 19.5 % which is consistent with the value of e) 20.6 %, which was obtained for the -300 nt to +210 nt *E. coli btuB* RS split reporter construct.



II.2.3.3.2. Regulation of sacB

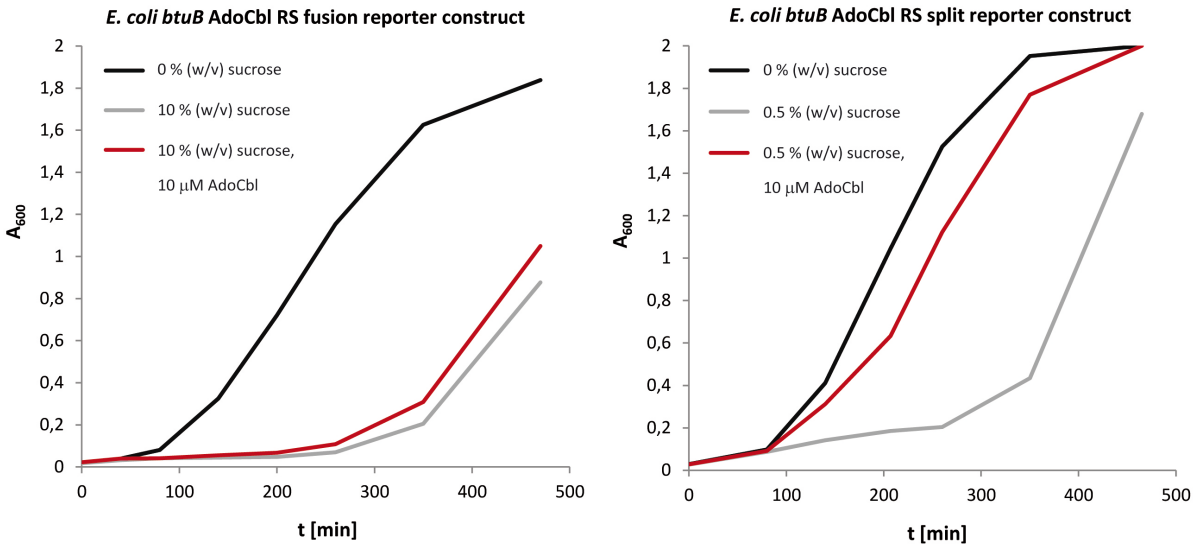
RS-control of the *sacB* gene was verified by qualitative determination of levansucrase expression. Therefore, sucrose-dependent cell growth assays were performed as described in **IV.2.3.2**. *E. coli* JM109 cells were grown in LB medium in absence and presence of sucrose and RS ligand. Without the addition of a RS ligand, cell growth is expected to be inhibited, because of the formation of toxic fructose polymers. Upon addition of a RS ligand, cell growth should be restored, because expression of levansucrase is inhibited.

First of all, it has to be mentioned that the poor solubility of guanine made it impossible to determine the level of RS-controlled levansucrase expression in the *S. aureus xpt* guanine RS fusion reporter construct with the sucrose cell growth assay. The guanine stock solution was prepared in DMSO (2.5 mg/ml) and had to be diluted in a ratio of 1:10 with culture medium in order to yield the required concentration of 250 µg/ml in the

assay. This, however, adds 10 % (v/v) DMSO to the medium. A control without sucrose and guanine, but 10 % (v/v) DMSO in the culture medium showed, that cell growth is extremely slowed down (data not shown) and therefore a determination of levansucrase expression was not possible under the stated experimental conditions. Taking this into consideration, it seems possible that the low levels of gene regulation in the β-galactosidase assay are as well owing to the high DMSO concentrations, as cultures had to be cultivated over 24 hours before β-galactosidase expression could be determined.

In contrast to the β-galactosidase assays, the sucrose cell growth assay with the -300 nt to +210 nt *E. coli btuB* AdoCbl RS showed differences between the fusion and the split reporter constructs. Whereas the fusion reporter construct required addition of 10 % (w/v) sucrose to the medium to cause sufficient inhibition of cell growth, the split reporter construct showed sucrose sensitivity already at a mass concentration of 0.5 %. Probably the expression of levansucrase or proper protein folding in the fusion reporter construct is lower than in the split reporter construct and therefore requires a higher concentration of sucrose in the medium. As a result of the high sucrose concentrations, the fusion reporter construct did not show ligand-induced recovery of cell growth. Moreover, it could be observed, that in sucrose-containing probes an exponential cell-growth sets in after 5-6 h. This is probably due to spontaneous mutations in the *sacB* gene that are developed as a defense mechanism by bacteria that are exposed to long-time selection pressure. In contrast to the plasmid with the fusion reporter genes, the -300 nt to +210 nt *E. coli btuB* AdoCbl RS split reporter construct shows almost full recovery upon addition of 10 μM AdoCbl to the medium (see **Figure 21**). The fact, that cell growth is not restored to 100 % matches the results from the β-galactosidase assay, which showed that genes downstream of the ligand-bound AdoCbl RS are still expressed to about 20 %.

Figure 21. Sucrose cell growth assay with the *E. coli btuB* AdoCbl RS fusion and split reporter construct. In assays with the fusion reporter construct, cell growth is inhibited upon addition of 10 % (w/v) sucrose to the culture medium and couldn't be restored in medium containing 10 μM AdoCbl. Assays with the split reporter construct showed inhibition of cell growth already at a concentration of 0.5 % (w/v) sucrose and almost full recovery upon addition of 10 μM AdoCbl.



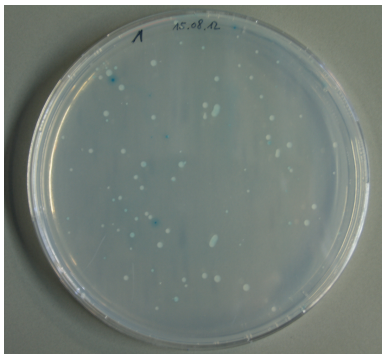
II.2.4. *In vivo* Screening Assay

II.2.4.1. First Round of CP screening

The best RS-mediated regulation of reporter gene expression was achieved with the -300 nt to +210 nt *E. coli btuB* AdoCbl RS split reporter construct, which is from now further referred to as AdoCbl RS construct. This vector was used for the *in vivo* screening of CPs. For this purpose, electrocompetent *E. coli* JM109 cells carrying the abovementioned plasmid were transformed with the CP library and the cells were plated on selective CP screening medium (see **Figure 22**). Cells that express a CP binding to the AdoCbl RS should form white colonies and have a growth advantage on sucrose containing medium because of the repressive effect on reporter gene expression. Besides white colonies, also blue colonies were able to grow on the selection medium, which

indicated the presence of false positive clones. This, however, is not uncommon at this stage of selection. Only white colonies were subjected to the next selection round.

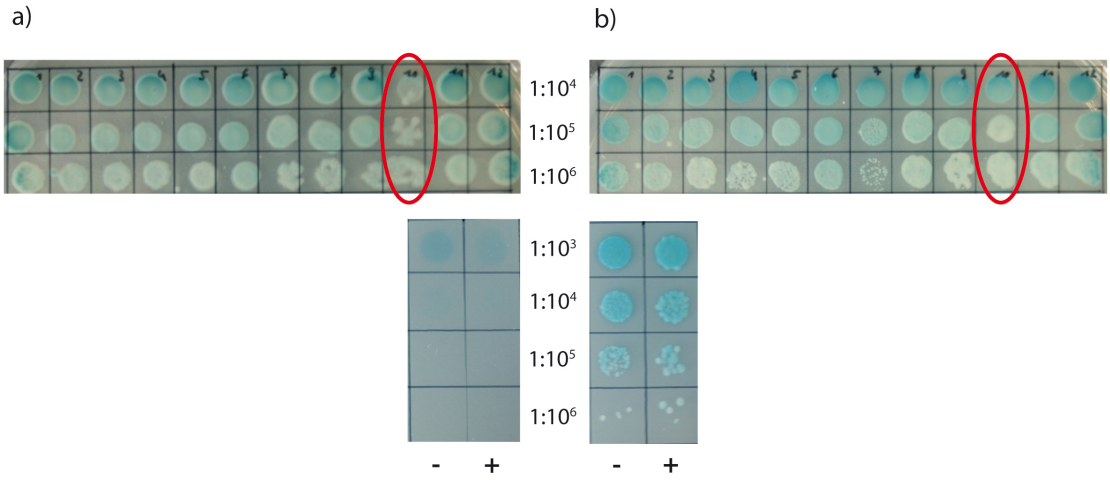
Figure 22. Colonies on selective CP screening medium after the first round of CP screening.



II.2.4.2. Second Round of CP screening

In the next round of CP screening, it was necessary to confirm, that white colonies from the first selection round where true positive hits. Therefore, white colonies from the first round of CP screening where picked and cultivated. Drops of serial dilutions ($1:10^4$-$1:10^6$) were spotted on selective CP screening medium with and without L-arabinose, which induces CP expression. CPs that bind to the AdoCbl RS and inhibit gene expression of the reporter genes *lacZ* and *sacB* should have a growth advantage on CP screening medium that contains L-arabinose. From about 500 tested colonies, there was one hit (colony 10) that showed increased growth upon CP expression (see **Figure 23**). To confirm this result, precultures of colony 10 were prepared with and without the addition of L-arabinose and a new set of serial dilutions ($1:10^3$-$1:10^6$) was spotted on selective CP screening medium with and without L-arabinose. The result from the initial screening could be reproduced, whereas there was no additional growth advantage, whether or not the precultures already contained L-arabinose.

Figure 23. Serial dilutions of positive colonies from the first round of CP screening. a) CP screening medium without L-arabinose and b) supplemented with L-arabinose. Colony number 10 showed growth advantage on medium supplemented with L-arabinose. Presence of L-arabinose in the preculture (+) did not give an additional growth advantage compared to precultures without L-arabinose (-).



In order to determine the CP sequence from colony 10, plasmid DNA was isolated and sequenced to be 5'-TGC ATC CAC GGG TAG AGC-3'. Hereby, the nucleotide triplet TAG stands for the amber stop codon. However, the *E. coli* JM109 strain carries an amber suppressor tRNA (*supE44* mutation, see **IV.1.6**) and therefore translates TAG into glutamine (Singaravelan *et al.*, 2010). Thus the CP, which is expressed in colony 10 is cyclo-CIHGQS.

II.2.4.3. Reporter gene assays with cyclo-CIHGQS

In order to confirm the previous positive results with cyclo-CIHGQS, β-galactosidase, as well as sucrose cell growth assays were performed with varying L-arabinose concentrations and with the following negative controls:

1) D-glucose instead of L-arabinose: Glucose is known to act as a repressor for the araBAD promoter (Guzman *et al.*, 1995). This control should illustrate the dependence of the assay results on the level of CP expression.

2) Reporter construct without RS (see **Figure 19)**

3) No CP construct

4) No CP construct + reporter construct without RS

The β-galactosidase expression could be reduced to 10.1 % upon addition of 0.2 % (w/v) L-arabinose. Surprisingly, in sucrose cell growth assays there was no measurable recovery of cell growth in presence of 0.5 % (w/v) sucrose and 0.2 % (w/v) L-arabinose. This could be explained by the fact, that levansucrase also accepts L-arabinose as a fructosyl acceptor, linking the C4-hydroxyl group of L-arabinose to the C2-hydroxyl group of the fructofuranosyl moiety of the cleaved sucrose molecule and resulting in the production of oligosaccharides (Tanaka *et al.*, 1981). This leads to the assumption that the expressed levansucrase completely consumes L-arabinose in the presence of sucrose and that an excess of L-arabinose has to be added to ensure sufficient concentrations for full induction of CP expression. An increased L-arabinose concentration of 5 % (w/v) led to full recovery of cell growth (see **Figure 25**). In retrospective on step one and two of the initial CP screening assay, an increase of the arabinose concentration in the CP screening medium would probably improve the selection process and help to reduce the growth of false positive colonies. Interestingly, controls that contained 0.2 % (w/v) and 5 % (w/v) L-arabinose without sucrose, reached their plateau phase at a lower OD_{600} than the control without any supplements. This is shown exemplarily in the first diagram of **Figure 25**.

Unfortunately, results of β-galactosidase and sucrose cell-growth assays with negative controls 1-4 were in the same range as the results of cyclo-CIHGQS and 5 % L-arabinose and are thus false positive results. They show that the reduction of reporter gene expression is not dependent on the presence of RSs or expressed CPs. These false positive results prompt speculations, whether the monosaccharides L-arabinose and D-glucose generally influence cell growth and gene expression levels in a negative way. Of course,

it is known that the presence of glucose in the cultivation medium lowers intracellular cAMP levels, which prevents formation of a cAMP-CAP complex, its binding to the DNA, recruitment of RNA polymerase and thus transcription initiation (Zheng *et al.*, 2004). But the *glnS* promoter region does not carry a CAP-binding site and therefore reporter gene transcription should not be affected by glucose or cAMP levels. In the end, it remains elusive why the addition of high concentrations of arabinose and glucose result in a reduction of reporter gene expression. One possible way to circumvent the monosaccharide-based difficulties would be to exchange the araBAD promoter for a promoter, which is not inducible by a monosaccharide.

Figure 24. β-Galactosidase assays of the AdoCbl RS construct upon induction/repression of cyclo-CIHGQS expression. a) 0 % (w/v) arabinose/glucose. b) 0.2 % (w/v) arabinose. c) 0.2 % (w/v) glucose. d) 0.2 % (w/v) arabinose, no CP construct present. e) 0.2 % (w/v) arabinose, no RS in reporter construct, no CP construct present.

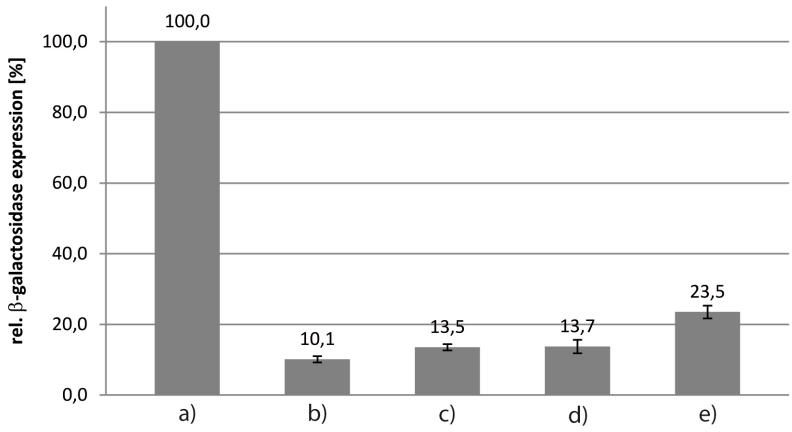
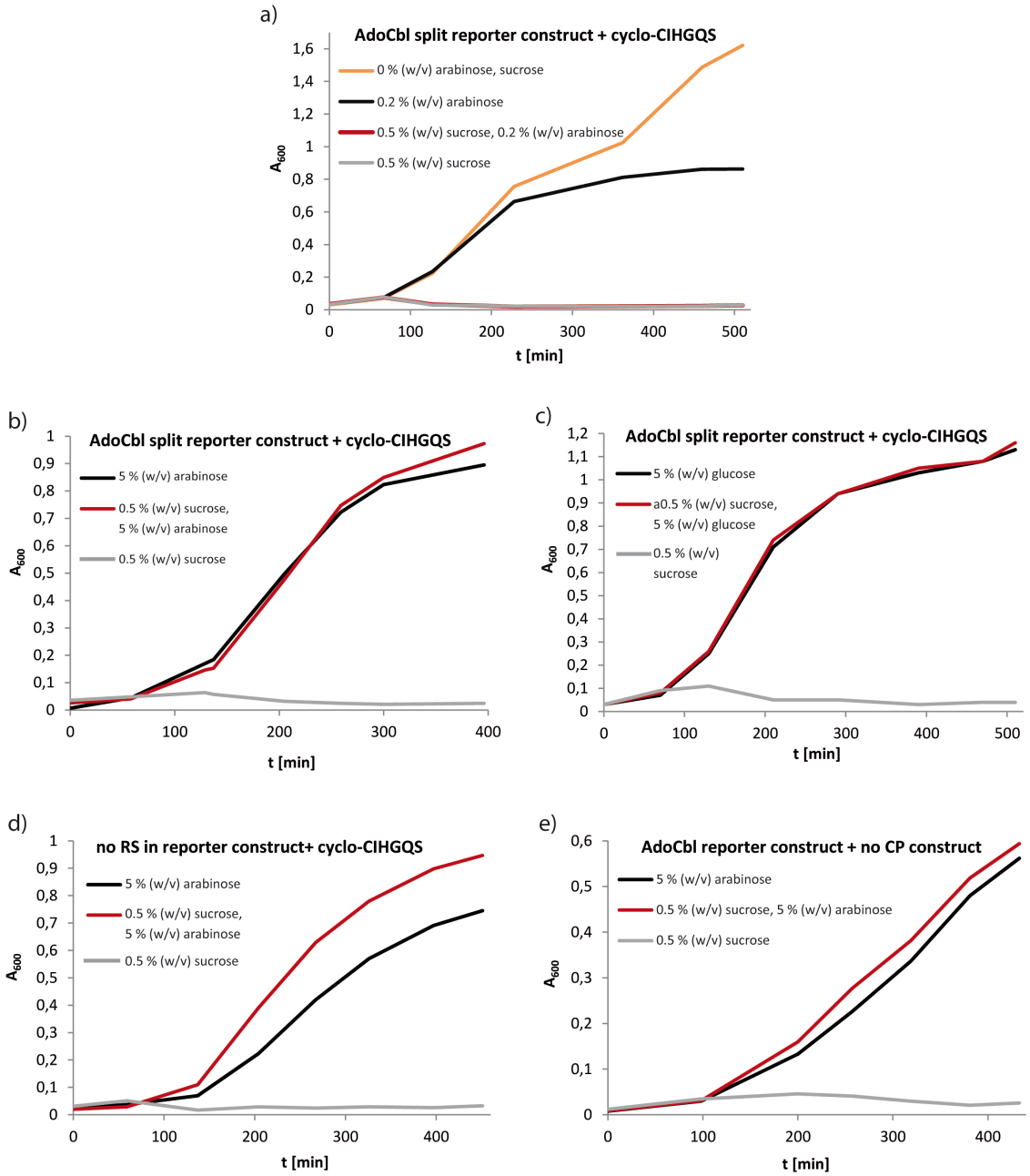


Figure 25. Sucrose cell growth assays of the AdoCbl RS construct upon induction/repression of cyclo-CIHGQS expression. a) 0.2 % (w/v) arabinose. b) 5 % (w/v) arabinose. c) 5 % (w/v) glucose. d) 5 % (w/v) arabinose, no RS in reporter construct. e) 5 % (w/v) arabinose, no CP construct present.



II.2.5. Integration of a Reporter Construct into the Genome of *E. coli*

Plasmids are tremendously versatile tools in molecular biology. Usually their application is straight forward, but sometimes working with plasmids can have disadvantages. Plasmids exist in several copies within the cell, depending on the origin of replication. The copy number determines whether the plasmid is a low-copy (~1-12), medium-copy (~15-20) or a high-copy plasmid (~20-700). But despite this categorization, the plasmid copy number within the cell can vary depending on the growth state of the cell (Lin-Chao & Bremer, 1986). Even under stable growth conditions, there are cell-to-cell copy number fluctuations (Paulsson & Ehrenberg, 2001).

When screening CP libraries for members that specifically bind to a RS of interest, these instabilities in copy numbers can lead to variations in reporter gene expression and thus give inconsistent results during the selection process. In the *in vivo* screening assay, the plasmid copy numbers have been chosen, so that the reporter construct is a low copy plasmid and the CP library construct is a high copy plasmid, which produces an excess of CPs compared to the number of RSs. But regarding the above-mentioned fluctuations of plasmid copy numbers, it seemed reasonable to integrate the AdoCbl RS construct into the genome of *E. coli*, while the CP library construct remains on the plasmid. This ensures a stable expression of reporter genes from a single genetic copy within the cell and an excess of CP molecules.

Because of the great length of the reporter construct (6125 nt), conventional chromosomal integration methods like simple homologous recombination could not be applied. Instead a system described by (Kuhlman & Cox, 2010) was used to achieve the integration of this large DNA fragment (see **IV.2.1.12**).

This system is based on the assumption, that recombination of large DNA fragments is more likely to occur when strand breaks have been introduced into the genomic DNA of *E. coli*. The strand breaks are introduced via the I-SceI restriction endonuclease, which is not naturally occurring in *E. coli* and needs to be co-expressed in the respective bacteria strain. The small tetracycline resistance gene *tetA,* which is flanked by two I-SceI restriction sites and two landing pads (LPs), is first integrated via homologous

recombination into the location of interest, and serves as a selectable placeholder for larger DNA fragments. One striking advantage of this mechanism is, that once the *tetA* gene is inserted, it can be conveniently replaced by any other sequence with up to 7 000 bp. This is possible by introducing strand breaks at the I-SceI restriction sites via restriction digestion and subsequent homologous recombination, which occurs at the previously introduced LPs and the LPs of the fragment to be inserted. There are certain locations in the genome of *E. coli* where those strand breaks are lethal, for example between the genes *atpI-gidB*, *yieN-trkB*, *ygcE-ygcF*, *ybbd-ylbG*, *essQ-cspB* or *nth-tppB*. When the final fragment is inserted into one of these locations, successful integrants can be identified without the need of antibiotic selection, as only cells with an intact chromosome can survive. The AdoCbl RS construct was integrated into the *nth-tppB* location of the *E. coli* JM109 genome. In order to achieve this, two additional plasmids express the genes that are necessary to help integrate the DNA fragment. The plasmids are equipped with the following features:

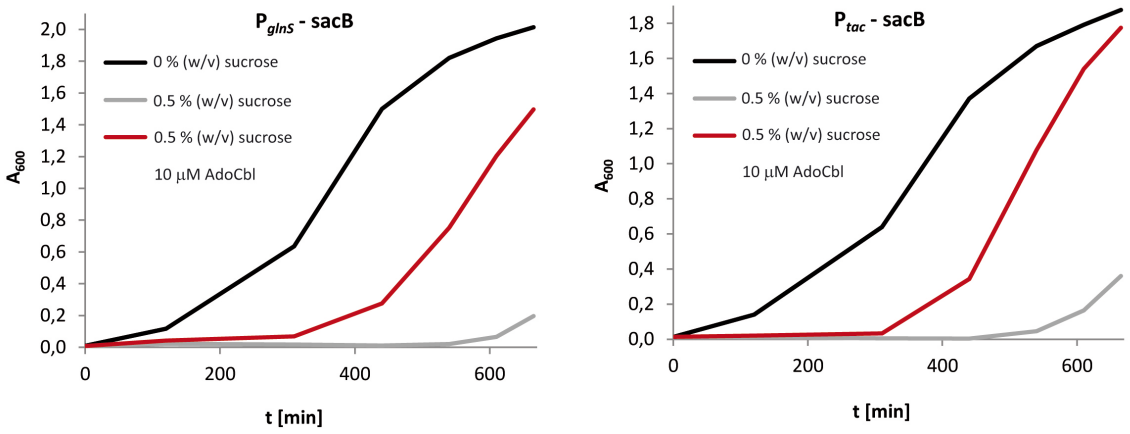
1) The pTKRED (helper plasmid) carries:
    - λ-Red genes (Redγ (*gam*), Redα (*exo*), Redβ (*bet*)), which are mediating homologous recombination (IPTG-inducible),
    - an I-SceI restriction endonuclease, which is a yeast mitochondrial homing endonuclease and not naturally occurring in *E. coli* (arabinose-inducible),
    - a spectinomycin resistance and
    - a temperature-sensitive origin of replication, which facilitates plasmid curing after successful DNA insertion into the genome.

2) The pTKIP (donor plasmid) carries:
    - an ampicillin resistance,
    - the DNA fragment to be inserted, framed by
    - two landing pads and
    - two I-SceI restriction sites.

After successful insertion of the entire AdoCbl RS construct into the genome of *E. coli* JM109, the functionality of the reporter genes was tested with β-galactosidase and sucrose cell growth assays. Unfortunately, the results of these experiments showed that

the reporter gene expressions were reduced almost to the detection limit (data no shown). Of course, the reporter genes now exist as one single copy per cell and therefore, the concentrations of β-galactosidase and levansucrase within the cell are lower. But *E. coli* strains like BL21 also have an integrated *lacZ* gene on the genome and β-galactosidase can be detected easily. As Kuhlman and Cox have used the strong synthetic and IPTG-inducible promoter P_{LlacO1} (Kuhlman & Cox, 2010) upstream of the genome-inserted genes, it seemed most likely that the *glnS* promoter upstream of each reporter gene is not strong enough to ensure sufficient expression levels. In order to prove this theory, the *glnS* promoter upstream of the *sacB* gene on the donor plasmid pTKIP was exchanged for the *tac* promoter via two successive rounds of site-directed ligase-independent mutagenesis (SLIM) as described in **IV.2.1.11**. Levansucrase expression levels were compared by performing sucrose-dependent cell growth assays with the P_{glnS}- and P_{tac}-containing donor plasmids (see **Figure 26**). Since there was no difference regarding levansucrase expression of the two plasmids, a genome integration of the P_{tac} construct was not taken into consideration.

Figure 26. Comparison of the promoters P_{glnS} and P_{tac} in a sucrose cell growth assay. On donor plasmid pTKIP-reporter, the promoter P_{glnS} (left) upstream of *sacB* has been replaced by the stronger synthetic promoter P_{tac} (right). Sucrose cell growth assays with both constructs show similar results and the *tac* promoter does not significantly improve gene expression.



II.2.6. Conclusion and Outlook

Initial tests with the *S. aureus xpt* guanine RS and the *E. coli btuB* AdoCbl RS together with their natural ligands gave the best RS-dependent regulation of gene expression when both reporter genes, *lacZ* and *sacB,* are under control of their own promoter and RS (split reporter construct). The *S. aureus xpt* guanine RS did not show sufficient gene regulation abilities in the β-galactosidase assay and failed to give any result in the sucrose cell growth assay. It is likely that this is owing to the poor solubility of the RS's natural ligand guanine. Furthermore, the guanine RS from the gram-positive bacterium *S. aureus* has, so far, only been studied in other gram-positive model organisms, like *B. subtilis*. It is known, that rho-independent transcription termination, as occurring in the *S. aureus xpt* guanine RS, is generally less common in gram-negative than in gram-positive bacteria. Comparing rho-independent terminator hairpins of *E. coli* and *B. subtilis* have revealed that there are significant differences regarding the typical lengths of the terminator stems as well as the typical sequences found in the terminator stems and loops (de Hoon *et al.*, 2005). Thus, it can be speculated, whether the lacking functionality of the *S. aureus xpt* guanine RS may be due to a missing recognition of the terminator hairpin structure by the gram-negative host organism *E. coli*.

Comparison of the -240 nt to 0 nt and the -300 nt to +210 nt *E. coli btuB* AdoCbl RS revealed that indeed parts of the *btuB* gene are necessary to efficiently control gene expression of the downstream genes. Upon binding of the natural ligand AdoCbl, the split reporter construct showed that the expression of both reporter genes, *lacZ* and *sacB,* could be reduced to about 20 %. Therefore the -300 nt to +210 nt *E. coli btuB* AdoCbl RS split reporter construct was used for the *in vivo* CP screening assay. However, no bioactive CP compound could be identified. False positive results were obtained from the addition of the monosaccharide arabinose, which is inducing the CP library expression. Furthermore, it was shown that the effect of reduced reporter gene expression did not depend on the presence of a RS or the CP library. In the future, a different promoter should replace the araBAD promoter of the CP library construct. Preferably, the new promoter does not have a monosaccharide as an inducing compound. Furthermore, attempts to integrate the *E. coli btuB* AdoCbl RS split reporter construct into the genome of *E. coli* JM109 were successful but reporter gene expression in the integrated state was reduced to the

detection limit. Consequently, the strong synthetic *tac* promoter replaced the *glnS* promoter upstream of *sacB*. Preliminary experiments that compared both promoters on the donor plasmid pTKIP-reporter could not provide an improvement of gene expression levels. Nevertheless, genome integration seems to be a more stable platform for screening experiments and should therefore be further pursued in this project. It could also be necessary to reconsider the structure of the reporter construct. In a different approach, for example, the RS could control a gene that encodes a repressor of the promoter upstream of the reporter genes. Furthermore, the reporter gene construct should be extended to include an antibiotic resistance gene. This reverse RS control mechanism would allow the forward selection with antibiotic resistance genes.

III. Structural Basis for the Site-Specific Incorporation of Non-Natural Amino Acids into Peptides and Proteins

III.1. Introduction

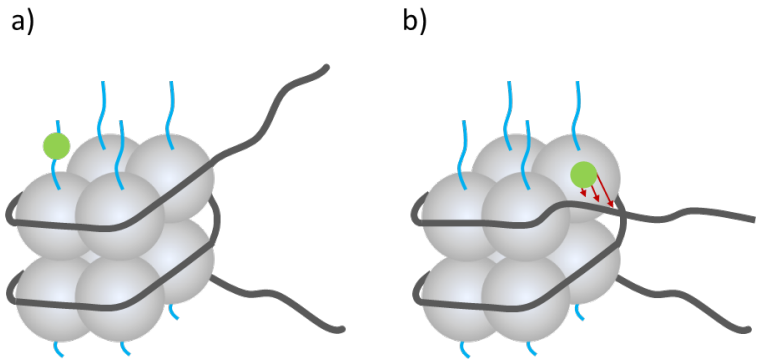
III.1.1. Posttranslational Modifications (PTMs) in Histones

Chromatin is a complex of DNA and protein and is used by eukaryotes to pack their large genomes into the nucleus while still enabling accessibility to it. Hereby, nucleosomes are the basic repeating unit of chromatin. They are comprised of a barrel shaped nucleoprotein, consisting of each two copies of the four core histones H2A, H2B, H3 and H4, and of about 147 base pairs of DNA that are wound around the histone octamer (Tropberger & Schneider, 2013). Chromatin structure and thus gene expression is epigenetically regulated by posttranslational modifications (PTMs) at the N-terminal tails protruding from the histone octamers or at the core regions of histones (see **Figure 27**).

Histone PTMs include among others acetylation and methylation of lysine residues and have led to the so called 'histone code' hypothesis. These modifications change the net charge of the residue from positive to negative or neutral as well as alter the hydrophobicity and flexibility of the modified protein. The interaction properties, landscape and partners of the protein are therefore altered, ultimately impacting on chromatin structure and gene expression (Tropberger & Schneider, 2013). As various

human diseases like cancer (Li, 2013), disorders of the central nervous system (Pena-Altamira *et al.*, 2013), and autoimmune diseases (Quintero-Ronderos & Montoya-Ortiz, 2012) are associated with misregulation of histone PTMs, histone acetyltransferases (HATs) and histone deacetylases (HDACs), which are enzymes responsible for lysine modifications, are regarded as important drug targets (Cole, 2008).

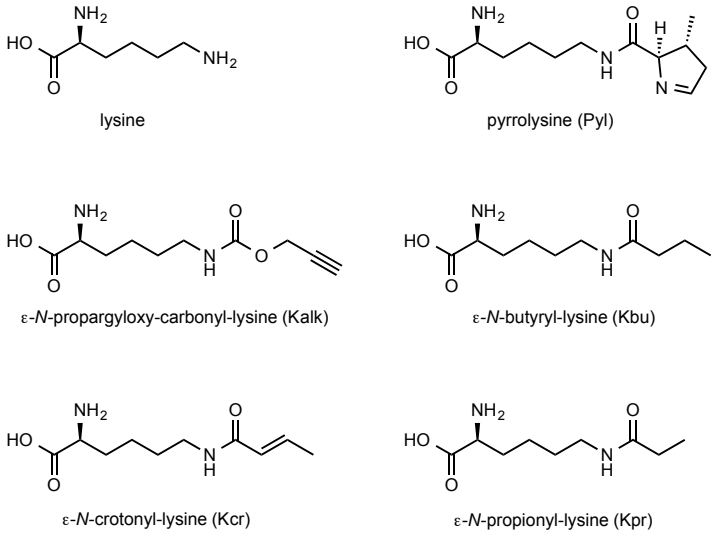
Figure 27. Modification regions on histones (modified after: Tropberger & Schneider, 2013). DNA (dark grey) is wrapped around the barrel-shaped histone octamer (grey). a) PTMs (green spheres) at the histone tails are readily accessible by effector molecules. b) PTMs can also be found in the core domains of histones, which directly influence histone-DNA interactions (red arrows).



Recently it became clear that lysine residues in histones can not only be acetylated or methylated, but also malonylated, succinylated, butyrylated, propionylated and crotonylated (Chen *et al.*, 2007; Garrity *et al.*, 2007; Tan *et al.*, 2011; Zhang *et al.*, 2011; Xie *et al.*, 2012; Arnaudo & Garcia, 2013). Despite the fact that some lysine modifications like butyrylation, crotonylation and propionylation are very similar in their chemical structure (see **Figure 28**), their effects on gene expression differ depending on the context. For instance, a direct link between activation of gene expression through crotonylation of Lys on histone H3 has been demonstrated (Tan *et al.*, 2011; Montellier *et al.*, 2012). In order to be able to understand the epigenetic regulation of cellular key processes, it is essential to elucidate the distinct temporal and spacial patterns of PTMs. Thus, lysine modifying enzymes, interaction partners and their regulatory proteins need to

be identified. It is therefore crucial to generate the target protein containing site-specific modifications.

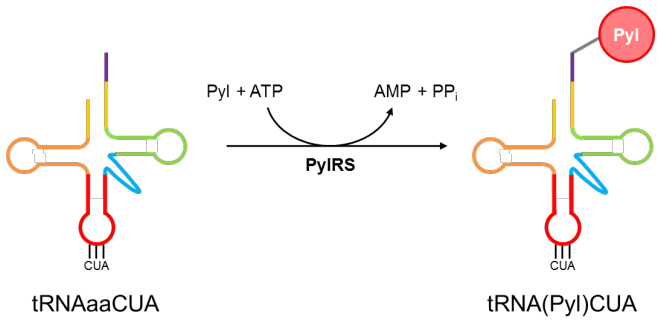
Figure 28. Lysine, pyrrolysine and analogs that are recognized by WT-PylRS.



III.1.2. Pyrrolysyl-tRNA synthetase (PylRS) – Expanding the Genetic Code

Pyrrolysine (Pyl) is a lysine derivative and known as the 22nd natural amino acid. It was discovered during structural elucidations of the methylamine methyltransferases MtmB, MtbB, and MttB of *Methanosarcina* species (Hao *et al.*, 2002; Srinivasan *et al.*, 2002; Soares *et al.*, 2005). The genes coding for these methyltransferases contain a single in-frame UAG amber codon, which is a stop codon in most organisms and usually terminates protein synthesis. In *Methanosarcina* species, however, UAG is coding for the unusual amino acid Pyl. In order to incorporate Pyl into the emerging protein chain, there is the need for a special tRNA, which carries the appropriate CUA anticodon (tRNAaaCUA), and for an aminoacyl-tRNA synthetase (aaRS), in this case pyrrolysyl-tRNA synthetase (PylRS), which loads the tRNAaaCUA with Pyl (see **Figure 29**).

Figure 29. Role of PylRS and tRNAPylCUA in the decoding of the UAG amber stop codon. Incorporation of Pyl into a protein requires a special tRNA with CUA as anticodon and PylRS, which charges the tRNAaaCUA with Pyl under the consumption of ATP.



These aaRS:tRNAaaCUA pairs are orthogonal to the genetic code and have been used to encode amino acids beyond the canonical 20 amino acids. Those pairs include not only the PylRS from *Methanosarcina mazei* and *Methanosarcina barkeri*, which incorporates Pyl, but also for example the tyrosyl-tRNA synthetase (TyrRS) from *Methanococcus jannaschii*, which incorporates tyrosine into a peptide of interest.

As the 20 proteinogenic amino acids offer only a limited variety of functional groups, researchers have been interested in expanding the genetic code particularly with regard to site-specific functionalization of peptides and proteins. Various studies on genetic code expansion have shown that PylRS can be evolved to accept a variety of non-natural, functionalized or chemically modifyable amino acids. By introducing an amber stop codon into a gene of interest and providing the respective non-natural amino acid in the cultivation medium, the host organism is able to use the orthogonal PylRS:tRNAaaCUA pair to site-specifically insert the respective non-natural amino acid (Wang & Rizzo, 2001; Santoro *et al.*, 2002; Zhang *et al.*, 2005; Chen *et al.*, 2009; Wang *et al.*, 2011; Kaya *et al.*, 2012; Kim *et al.*, 2012; Perona & Hadd, 2012; Wang *et al.*, 2012). Structures of PylRS in complex with its natural substrate (Kavran *et al.*, 2007; Yanagisawa *et al.*, 2008a; Yanagisawa *et al.*, 2013) as well as evolved mutants containing non-natural amino acids are already available (Yanagisawa *et al.*, 2008b; Takimoto *et al.*, 2011; Schneider *et al.*, 2013). The PylRS triple-mutant (L274A C313A Y349F) from *M. barkeri* allows for the introduction of ε-*N*-crotonyl-lysine (Kcr) into histones (Kim *et al.*, 2012). However, it

was discovered that also the wild type PylRS (WT-PylRS) from *M. mazei* recognizes Kcr, ε-*N*-butyryl-(Kbu), ε-*N*-propionyl-(Kpr) (Gattner *et al.*, 2013) and ε–*N*-propargyloxy-carbonyl-lysine (Kalk) (see **Figure 28**) as substrates and can therefore be used to integrate these non-natural amino acids into histones. The latter of the mentioned amino acids allows further chemical functionalization after heterologous expression using Cu(I)-catalyzed azide-alkyne cycloaddition reaction (Rostovtsev *et al.*, 2002; Tornoe *et al.*, 2002; Kaya *et al.*, 2009; Nguyen *et al.*, 2009; Wan *et al.*, 2010; Milles *et al.*, 2012). The extended substrate tolerance of the *M. mazei* WT-PylRS can be most likely attributed to its high sequence identity (71 %; sequence similarity = 85 %) to the *M. barkeri* triple-mutant PylRS. Previous studies on the WT-PylRS show that evolution of PylRS is not strictly necessary in order to introduce non-natural amino acids into a protein of interest.

III.1.3. Objectives

Lysine is the main target for histone PTMs like acetylation or methylation. But lysine PTMs emerged to be more versatile and include as well malonylation, succinylation, butyrylation, propionylation or crotonylation, which result in significant alteration of gene expression patterns. In order to be able to understand the epigenetic regulation of cellular key processes, it is necessary to identify lysine modifying enzymes, interaction partners and their regulatory proteins. Therefore it is crucial to generate proteins that carry these modifications site-specifically.

The site-specific translational incorporation of non-natural amino acids can be achieved by using the orthogonal PylRS:tRNAaaCUA pair. Structures of evolved mutants of PylRS in complex with non-natural amino acids are already available, but so far it remained unknown what key features a substrate should possess in order to be tolerated by WT-PylRS without the need for directed evolution.

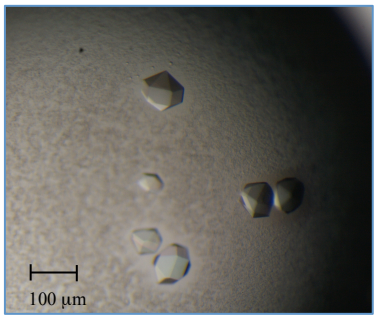
Therefore, the aim of this project was to elucidate the X-ray crystal structure of the *M. mazei* WT-PylRS in complex with the non-natural amino acids Kalk, Kbu, Kcr and Kpr. The crystallographic data should provide insight into the structure-function relationship and enable to deduce the above-mentioned key features for potential PylRS substrates and to understand why directed evolution of the protein is not always necessary for substrate tolerance.

III.2. Results and Discussion

III.2.1. Crystallization and structure determination of WT-PylRS in complex with Kalk, Kbu, Kcr and Kpr

PylRS was co-crystallized with the non-natural amino acids Kalk, Kbu, Kcr and Kpr in the presence of ATP. However, addition of the respective amounts of amino acid and ATP to the concentrated protein led to precipitation of PylRS. Therefore, the previous separately diluted compounds were mixed and subsequently concentrated to a protein concentration of 10 mg/ml (see **IV.2.4.1**). The published crystallization conditions for PylRS in complex with the ATP analog AMP-PNP (adenylyl imidodiphosphate) (100 mM Tris (pH 7.0-8.0) and 8-14% (w/v) PEG2000 monomethyl ether) (Kavran *et al.*, 2007) were not applicable on the co-crystallization of PylRS with Kalk, Kbu, Kcr and Kpr. Therefore, high-throughput sitting-drop vapor diffusion screenings were performed in 96-well Intelli-Plates using buffer kits from Qiagen (Hilden, DE). The initial hit condition, in which all of the four protein complexes crystalized, was found in the PEGs Suite buffer kit and contained 200 mM lithium acetate and 20 % (w/v) PEG3350 with a 1:2 mixture of protein and well solution. In order to optimize crystal quality, manual fine screens around the hit condition were performed in 24-well plates using the hanging-drop vapor diffusion method. Hexagonal shaped crystals (see **Figure 30**) were obtained in 100 mM lithium acetate and 10-14 % (w/v) PEG3350. After cryoprotection and freezing in liquid nitrogen, datasets of the crystals were collected using synchrotron X-ray irradiation. The diffraction data were processed to 2.1-2.3 Å resolution (see **Table 22**) and the structures were solved by molecular replacement using the coordinates of the previously reported WT-PylRS structure (PDB code: 4BW9).

Figure 30. Hexagonal shaped WT-PylRS crystals.



The electron densities clearly show that all four adenylated amino acids occupy the same space in the active site of the aaRS domain (see **Figure 31**). While the adenosine and phosphate moieties superimpose almost perfectly, the complex structures of Kalk, Kcr and Kpr show that the amino group can take up alternative conformations by a 180° rotation around the C-Cα bond. This was previously observed as well in the structure of the engineered enzyme (PylRS Y306A Y384F) that binds ε–*N*-*o*-azidobenzyloxycarbonyl lysine (PDB code 2ZIO) (Yanagisawa *et al.*, 2008b). In the Pyl-bound PylRS complex, the α–amino group of Pyl interacts with the hydroxyl group of Tyr 384, which is located in the β7–β8 hairpin (Kavran *et al.*, 2007). However, the Pyl analogs used in this study lack the ability to form a comparable H-bond and therefore this loop is flexible and not defined in the electron density. This flexibility is independent from the presence of bound substrate and it was postulated that this hairpin protects the unstable pyrrolysyladenylate intermediate (Yanagisawa *et al.*, 2008a; Yanagisawa *et al.*, 2013). In addition, mutation of Tyr 384 to Phe showed that this H-bond interaction between the loop and the substrate is not essential for enzyme function, and its absence may even enhance utilization of non-natural substrates (Yanagisawa *et al.*, 2008b; Kaya *et al.*, 2012).

In close proximity of the α-phosphate some positive difference density is visible, which could correspond to a partly occupied pyrophosphate, which is a remnant from the adenylation reaction. Only in the Kpr complex structure, the observed density allowed modeling of the pyrophosphate at 60 % occupancy. Compared to the pyrroline ring in Pyl the alkyne, butyryl, crotonyl or propionyl functional groups in Kalk, Kbu Kcr and Kpr are smaller. Thus more space and a higher degree of flexibility is possible, which explains

that not all atoms of the methylene spacer carrying the functional group are fully defined in the electron density (**Figure 32**).

Figure 31. Cartoon representation of the overall structure of the catalytic domain of PylRS. a) Type II tRNA synthetase folding topology of the tRNA synthetase domain from WT-PylRS, shown as cartoon representation and overlaid with its semi-transparent surface. The adenylated Kbu is highlighted as green stick model in the active site. b)-f) Zoom in the active site: PylRS in complex with Kalk (orange), Kbu (green), Kcr (blue) and Kpr (red), drawn as stick models. (r.m.s.d ~0.35 Å). All four non-natural amino acids bind in the same position.

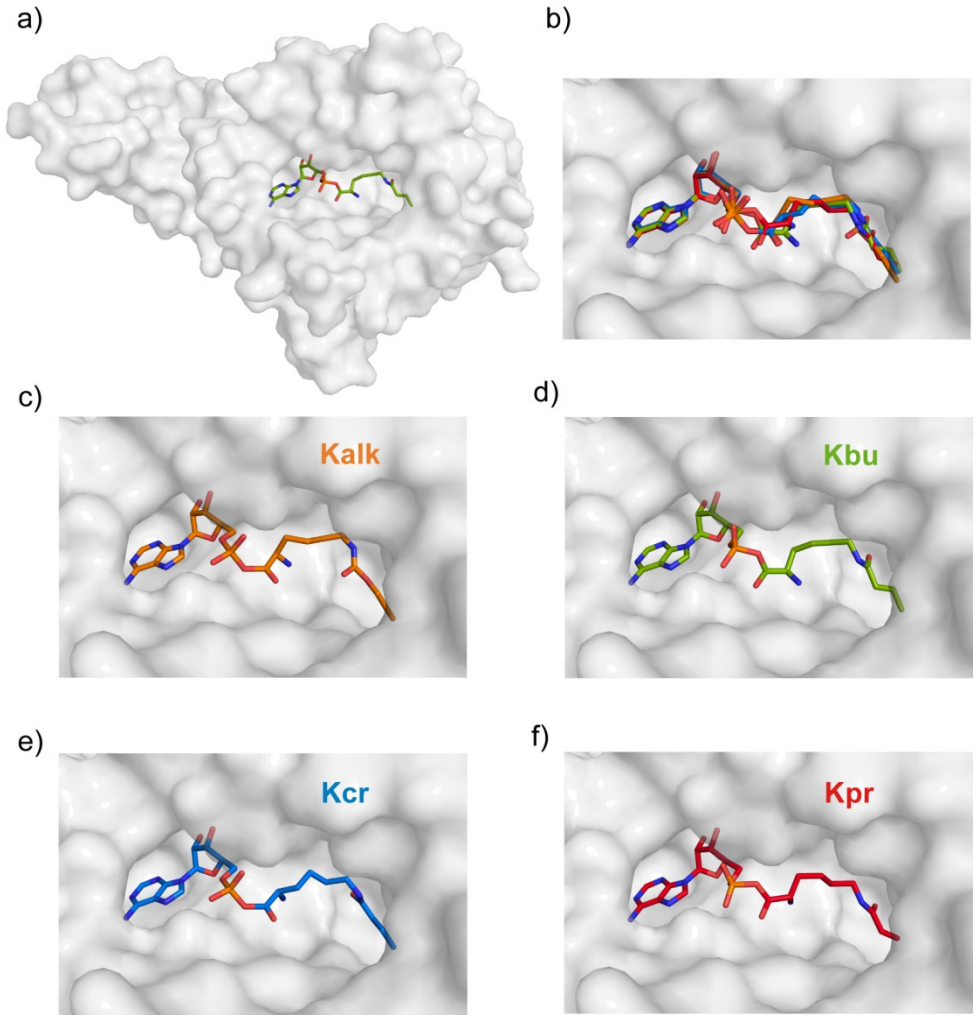
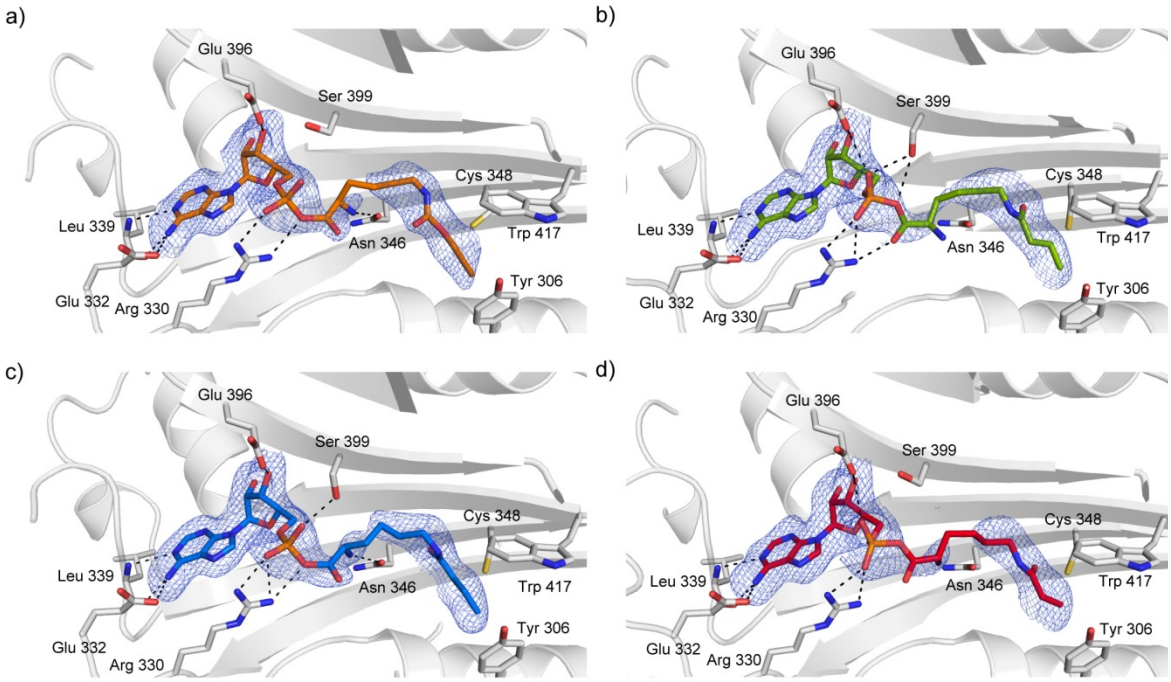


Figure 32. Simulated annealing-omit Fo-DFc electron density contoured at 2 σ of PylRS in complex with a) Kalk (orange), b) Kbu (green), c) Kcr (blue) and d) Kpr (red). The protein is shown as cartoon, overlaid with its semi-transparent surface representation. Amino acids providing key interactions are drawn as sticks, hydrogen-bonds as dashed lines.

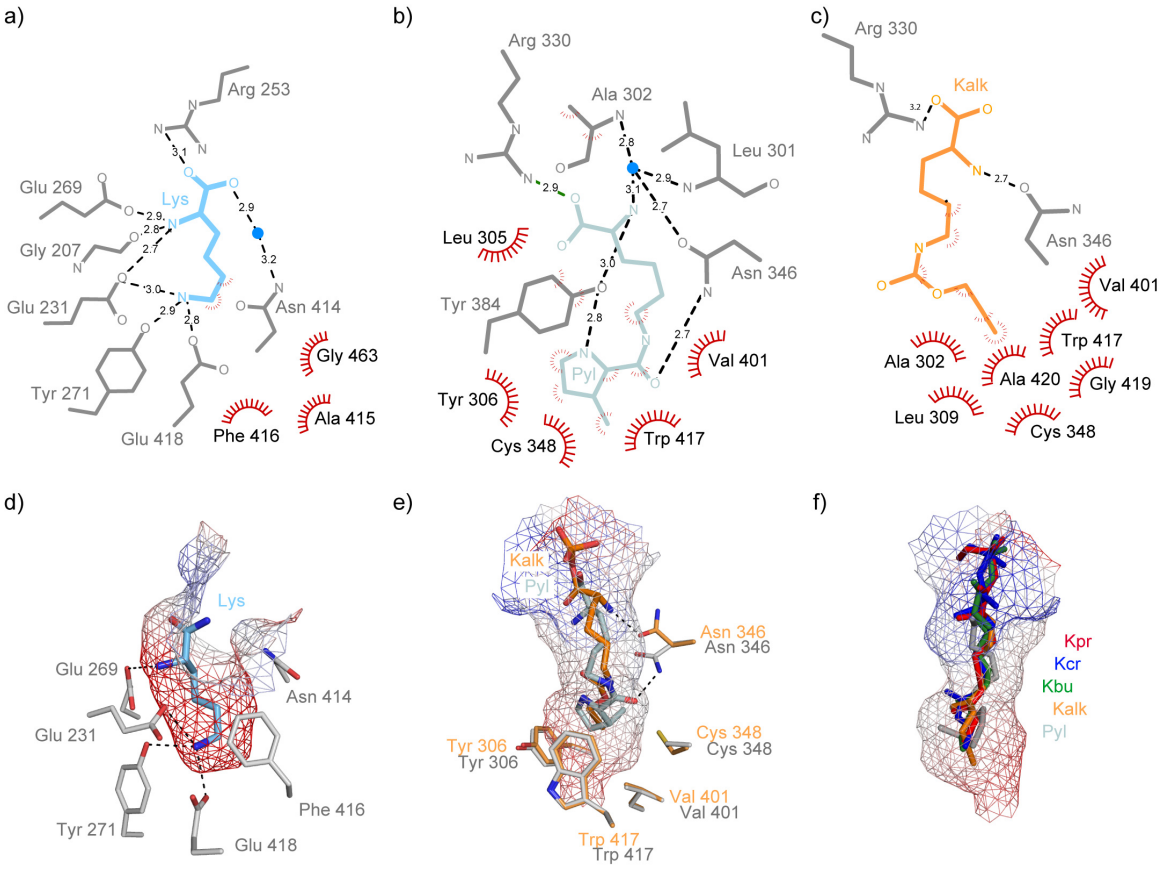


III.2.2. Key Features for Substrate Recognition by WT-PylRS

While PylRS seems to accept a variety of substrates, the lysyl-tRNA synthetase (LysRS) is very specific for the amino acid lysine. A comparison of both active sites elucidates the differences in substrate acceptance and helps to deduce key features for potential PylRS ligands. Both PylRS and LysRS belong to the type II family of aaRS and thus share the same overall folding topology. Structures of the LysRS from *Bacillus stearothermophilus* (PDB code 3A74), *Bulkholderia thailandensis* (PDB code 4EX5 (Baugh *et al.*, 2013)) and *Escherichia coli* (PDB code 1E22 (Desogus *et al.*, 2000)) are available. The three LysRS share about 53 % sequence identity, the catalytic domains superimpose with an r.m.s.d. of 1.2 Å and the residues that are lining the active site are highly conserved. In comparison, the sequence identities to the catalytic domain of PylRS from *M. mazei* are 17 %, 19 %

and 20 %, respectively, and superimpose with an r.m.s.d. of about 2.1 Å. The ATP binding sites in both PylRS and LysRS are comparable. The LysRS employs an elaborate interaction network with Lys and forms a tight binding pocket (**Figure 33a and 33d**).

Figure 33. Comparison of the binding pockets of PylRS and LysRS. Two-dimensional plot (Wallace *et al.*, 1995) of residues interacting with Lys (a), Pyl (b) and Kalk (c). Van-der-Waals contacts are shown as red half-spheres. Surface representation of the binding pockets of LysRS (d) and PylRS (e) with Lys, Pyl and Kalk, respectively. f) Superposition of Pyl (grey), Kalk (orange), Kbu (green), Kcr (blue) and Kpr (red). The surface of the binding pocket is shown as mesh, with the charge distribution indicated by coloring (red = negative, blue = positive). The surrounding residues are drawn as sticks. (PDB codes PylRS: 2Q7H, LysRS: 3A74).



In contrast, PylRS provides only some key interactions with Pyl in a more spacious active site that is lined with hydrophobic residues (**Figure 33b and e**). The alkyne group of Kalk, the least flexible of the used derivatives, protrudes slightly deeper into the pocket than Pyl but lies in the same plane as the pyrroline ring. Thus favorable π–π interactions

between the amino acid side chain with Tyr 306 as well Trp 417 are possible (**Figure 33e**). PylRS can therefore accommodate a wider range of substrates, as long as they possess certain key prerequisites: If present, the carbamate, carbonyl or amide moiety, can either interact with Asn 346 or Cys 348, even if this interaction is not permanent, as can be seen in a Pyl (Yanagisawa *et al.*, 2006; Kavran *et al.*, 2007; Yanagisawa *et al.*, 2013) and a norbornene-containing Pyl analog (Knorb) (Schneider *et al.*, 2013). For Pyl, an additional hydrogen bond accepting imine nitrogen in the pyrroline ring further increases the activation efficiency (Li *et al.*, 2009). Our data show that analogs containing only the amide motif, can efficiently bind to the active site of the PylRS. Thus Kalk, Kbu, Kcr, and Kpr are recognized as substrates and fit well in the pocket (**Figure 33f**). The preferred substrate head group is hydrophobic and up to four atom bonds in length can be accepted in the binding pocket. Additionally, branched functional groups like a tert-butyloxycarbonyl group (Yanagisawa *et al.*, 2008b) or small ring systems, such as cyclopentane can be accommodated in the spacious pocket.

However, reengineering of the enzyme active site would be required to accommodate non-natural amino acids lacking these key features. The hot spots which were targeted so far by directed evolution include amino acids whose side chains are either lining the binding pocket (Tyr 306, Leu 309, Asn 346, Cys 348, Trp 417) or resides in the active site closing loop (Ile 417, Phe 384) (Corresponding residues in *M. bakeri*: Leu 274, Asn 311, Cys 313 and Tyr 349) (Yanagisawa *et al.*, 2008b; Kaya *et al.*, 2009; Takimoto *et al.*, 2011; Kim *et al.*, 2012; Wang *et al.*, 2012). For instance in order to efficiently use larger and more bulky Pyl analogues like norbornene or ε-*N*-benzyloxycarbonyl-L-lysine, more room is needed. This can be made available by replacing Tyr 306 at the bottom of the binding pocket by a Gly or Ala (Yanagisawa *et al.*, 2008b; Schneider *et al.*, 2013). Structural data on PylRS mutants show that even replacing this large amino acid appears not to perturb the overall structural architecture.

III.2.3. Conclusion and Outlook

The X-ray crystallographic data from the *M. mazei* WT-PylRS in complex with the adenylated non-natural amino acids Kalk, Kbu, Kcr and Kpr revealed that all four amino acids are enzyme substrates and occupy the same space in the active site. As there is no H-bond formation possible between the Pyl analogs and Tyr 384 of the β7-β8 hairpin, this loop is flexible, which may even enhance substrate tolerance. Furthermore, a high degree of flexibility is observed for the branchless head groups of the amino acids. In addition to this, key features for the non-natural PylRS substrates could be deduced. The preferred head group is hydrophobic and up to 4 atom bonds in length. Branched and unbranched as well as smaller ring systems are tolerated in the binding pocket. Furthermore, the head group is connected to the methylene spacer of the lysine side chain via a carbamate or amide group, which enables stabilizing interactions with Asn 346 and Cys 348. Overall, it could be shown that directed evolution of the enzyme is not necessary for substrate tolerance.

Knowing about these key features, more PylRS substrates could be designed in the future in order to expand the genetic code. Hereby, non-natural amino acids, which offer the possibility to be further functionalized, are of particular interest. Chemical methods like click chemistry (Kaya *et al.*, 2012; Jagadish *et al.*, 2013; Schneider *et al.*, 2013) or Staudinger ligation (Yanagisawa *et al.*, 2008b) have proven to be powerful tools for the straightforward modification of peptides and proteins that have been generated via the utilization of orthogonal aaRS:tRNAaaCUA pairs. Taking this into account, the PylRS:tRNAaaCUA pair also seems a suitable tool for the site-directed functionalization of expressed CPs. In a previous project, fluorescent labeling of the norbornene-containing Pyl analogue Knorb was possible by applying very mild, copper-free click chemistry with tetrazole-, tetrazine- or hydrazonyl chloride-based fluorophores (Kaya *et al.*, 2012; Schneider *et al.*, 2013). Applying similar techniques to CPs that contain the clickable non-natural amino acids Kalk or Knorb could generate CPs that possess a higher binding affinity towards a RS of choice or other cellular targets.

IV. Materials and Methods

IV.1. Materials

IV.1.1. Chemicals

All chemicals were purchased from the following companies:

AppliChem (Darmstadt, DE), Fluka (Neu-Ulm, DE), Merck (Darmstadt, DE), Sigma-Aldrich (Steinheim, DE), Protein Technologies, Inc. (Tucson, US), Roth (Karlsruhe, DE) and VWR (Darmstadt, DE).

The non-natural amino acids Kalk, Kbu, Kcr and Kpr were synthesized as described earlier (Gattner *et al.*, 2013) by Dr. Milan Vrabel in the group of Prof. Carell (LMU Munich).

IV.1.2. Enzymes

All enzymes were purchased from New England Biolabs (Ipswich, US) unless stated otherwise. They were used as specified by the manufacturer. Variations from the instructions are indicated in the following chapters.

IV.1.3. Isolation and Preparation Systems

Table 2. Isolation and purification systems.

System	Application	Manufacturer
peqGOLD Plasmid Miniprep Kit II	Plasmid DNA isolation from *E. coli*	PEGLAB (Erlangen, DE)
peqGOLD Cycle-Pure Kit	DNA purification	PEGLAB (Erlangen, DE)
peqGOLD Gel Extraction Kit	DNA extraction from agarose gels	PEGLAB (Erlangen, DE)
Ni-NTA Spin Columns	Purification of His_6-tagged proteins	Qiagen (Hilden, DE)
Strep•Tactin® SpinPrep™ Kit	Purification of Strep-tagged CPs	Novagen®/Merck (Darmstadt, DE)

IV.1.4. Antibiotics

Table 3. Mass concentrations of antibiotic solutions.

	Ampicillin	Chloramphemicol	Kanamycine	Spectinomycine	Tetracycline
β (final)	100 mg/l	34 mg/l	50 mg/l	100 mg/l	20 mg/l

IV.1.5. Media

For the preparation of agar plates, 1.8 % (w/v) agar was added to the respective medium.

Table 4. Rich media for bacterial growth.

LB	SOC	Glycerol-SOC	SOB
10 g/l peptone	20 g/l peptone	20 g/l peptone	20 g/l peptone
5 g/l yeast extract	5 g/l yeast extract	5 g/l yeast extract	5 g/l yeast extract
5 g/l NaCl	10 mM $MgSO_4$	10 mM $MgSO_4$	10 mM $MgSO_4$
	10 mM NaCl	10 mM NaCl	10 mM NaCl
	10 mM $MgCl_2$	10 mM $MgCl_2$	10 mM $MgCl_2$
	2.5 mM KCl	2.5 mM KCl	2.5 mM KCl
	20 mM glucose	20 mM glycerol	

Table 5. Minimal media for bacterial growth.

M9-Minimal-Medium	10 x M9 Salts	CP Screening Medium
1 x M9 salts	477 mM Na_2HPO_4	1 x M9 salts
0.4 % (w/v) glycerol	220 mM KH_2PO_4	0.4 % (w/v) glycerol
10 g/l casamino acids	86 mM NaCl	10 g/l casamino acids
1 mM thiamine	187 mM NH_4Cl	1 mM thiamine
2 mM $MgSO_4$		2 mM $MgSO_4$
100 µM $CaCl_2$		100 µM $CaCl_2$
		0.2 % (w/v) L-arabinose
		0.004 % (w/v) X-gal
		1 % (w/v) sucrose

IV.1.6. Strains

Table 6. *E. coli* strains used in this thesis.

Organism	Strain	Genotype	Source
E. coli	JM109	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F´ *tra*D36, *pro*AB, *laq*I^qZΔM15]	Promega (Madison, US)
E. coli	NEB Turbo	*F' proA*$^+$*B*$^+$ *lacI*^{q} Δ*lacZM15 / fhuA2* Δ*(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)*Tet^S *endA1 thi-1* Δ*(hsdS-mcrB)5*	New England Biolabs (Ipswich, US)
E. coli	XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr)]*	Stratagene/Agilent Technologies (Santa Clara, US)
E. coli	Rosetta DE3	*F- ompT hsdS*$_B$*(*r_B^- m_B^-*) gal dcm (DE3) pRARE (*Cam^R*)*	Novagen®/Merck (Darmstadt, DE)

IV.2. Methods

IV.2.1. Cloning Methods

IV.2.1.1. PCR

For preparative DNA amplification, PCRs were performed with Phusion® High-Fidelity DNA Polymerase as specified by the manufacturer. The PCRs were optimized for annealing temperature and DMSO concentration. Typical component concentrations and thermocycling conditions are shown below:

Table 7. Component concentration and thermocycling conditions of a typical PCR with Phusion® High-Fidelity DNA Polymerase.

Component	Final concentration		Thermocycling conditions
Template DNA	0.01-0.3 ng/µl		98 °C, 30 sec
Forward primer	1 µM		98 °C, 15 sec
Reverse primer	1 µM	25 cycles	50-68 °C, 30 sec
dNTPs	200 µM		72 °C, 30 s per kb
Phusion HF or GC Buffer	1 x		72 °C, 5 min
DMSO	0-10 %		4 °C, ∞
Phusion DNA Polymerase	0.02 U/µl		

IV.2.1.2. Single Colony PCR

For analytical DNA amplification, single colony PCRs were performed with either *Taq* DNA Polymerase with ThermoPol Buffer (New England Biolabs (Ipswich, US)) or GoTaq® DNA Polymerase (Promega (Madison, US)) as specified by the manufacturer. A PCR master mix was split into 10 µl aliquots. Individual colonies were picked and resuspended in the reaction mixture. Typical component concentrations and thermocycling conditions are shown below:

Table 8. Component concentration and thermocycling conditions of a typical PCR with *Taq* DNA Polymerase with ThermoPol Buffer.

Component	Final concentration		Thermocycling conditions
Forward primer	1 µM		98 °C, 8 min
Reverse primer	1 µM		95 °C, 30 sec
dNTPs	200 µM	35 cycles	50-68 °C, 30 sec
ThermoPol Buffer	1 x		68 °C, 1 min per kb
Taq DNA Polymerase	0.025 U/µl		68 °C, 7 min
			4 °C, ∞

Table 9. Component concentration and thermocycling conditions of a typical PCR with GoTaq® DNA Polymerase.

Component	Final concentration		Thermocycling conditions
Forward primer	1 µM		98 °C, 5 min
Reverse primer	1 µM		95 °C, 30 sec
dNTPs	200 µM	35 cycles	50-68 °C, 30 sec
Green GoTaq Reaction Buffer	1 x		72 °C, 1 min per kb
GoTaq DNA Polymerase	0.05 U/µl		72 °C, 5 min
			4 °C, ∞

IV.2.1.3. Restriction Digestion

45-150 ng/µl DNA was digested with 0.6-1.0 U/µl of the respective restriction endonuclease in a recommended 1 x reaction buffer and 1 x BSA, if required. Preparative restriction digestions were incubated at an enzyme specific temperature for 3 h, analytical restriction digestions for 1 h.

IV.2.1.4. Formation of Blunt Ends

Fill-in of 5' overhangs and removal of 3' overhangs from digested DNA fragments was performed by adding 0.2 U/µl DNA Polymerase I, Large (Klenow) Fragment and 33 µM dNTPs directly to the restriction digestion mixture. After incubation at 25 °C for 15 min, the enzyme was inactivated by addition of 10 mM EDTA. Subsequently the DNA fragment was purified by agarose gel electrophoresis (see **IV.2.1.9**).

IV.2.1.5. Plasmid DNA Dephosphorylation

Previous to ligation, plasmid DNA was dephosphorylated by adding 0.1 U/µl Antarctic Phosphatase and Antarctic Phosphatase Reaction Buffer (final concentration: 1 x) directly to the restriction digest mixture. After incubation at 37 °C for 30 min, the enzyme was inactivated by incubation at 65 °C for 5 min. Subsequently the DNA fragment was purified by agarose gel electrophoresis (see **IV.2.1.9**).

IV.2.1.6. Ligation

Plasmid and insert DNA fragments were previously purified via agarose gel electrophoresis (see **IV.2.1.9**). They were ligated overnight at 4 °C in a molar ratio of 1:3 (plasmid:insert) with 20 U/µl T4 DNA Ligase in 1 x T4 DNA Ligase Reaction Buffer.

IV.2.1.7. Preparation of Electrocompetent *E. coli* Cells

500 ml SOB medium was inoculated with 5 ml of a fresh overnight culture and incubated at 37 °C and 130 rpm shaking. At an OD_{600} of 0.5-0.7, cells were transferred to a sterile ice cold centrifuge beaker and cooled on ice for 20-30 min. Cells were spun down at 4 °C and 5 000 rpm for 20 min. The supernatant was decanted and the cell pellet resuspended in 500 ml ice cold 10 % (v/v) glycerol. This centrifugation and washing step was repeated with 500 ml and then with 20 ml. After final centrifugation, the cell pellet was resuspended in 2-3 ml ice cold 10 % (v/v) glycerol. 50-100 µl aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

IV.2.1.8. Transformation of Electrocompetent *E. coli* Cells

For transformation of *E. coli* with plasmid DNA, 1 µl plasmid with a concentration of 1-3 ng/µl was mixed with 40 µl electrocompetent cells which had been previously thawed on ice. For transformation of *E. coli* with DNA from ligation reactions, 1-2 µl of the ligation mixture was used instead. The mixture of cells and DNA was transferred into a precooled 2 mm electroporation cuvette (PEQLAB (Erlangen, DE)) and electroporated in a Bio-Rad (Hercules, US) MicroPulser™ at a voltage of 2.5 kV. Cells were immediately resuspended in 1 ml SOC medium and allowed to recover at 37 °C for 1 h. 30 µl (plasmid DNA) or the entire (DNA from ligation reactions) SOC medium was plated on selective LB agar medium and incubated at 37 °C overnight.

IV.2.1.9. Agarose Gel Electrophoresis

Depending on DNA fragment size, agarose gels with 0.6-2 % (w/v) agarose in 1 x TAE buffer were used to preparatively or analytically separate DNA fragments. Samples were mixed with DNA loading dye (final concentration: 1 x) and loaded on the gel, which was subsequently run for 45 min at 175 V in a PEQLAB PerfectBlue™ Mini M electrophoresis chamber.

Table 10. Buffers and solutions used for agarose gel electophoresis.

TAE Buffer	DNA Loading Dye (6x)
0.4 M Tris-HCl, pH 8.2	10 mM Tris-HCl, pH 8.2
100 mM EDTA	1 mM EDTA
	50 % (w/v) glycerol
	0.25 % (w/v) xylene cyanol
	0.25 % (w/v) bromphenol blue

IV.2.1.10. Site-Directed Mutagenesis

Site-directed point mutations were performed in 50 µl scale PCRs with varying amounts of template DNA (10-50 ng) while keeping the amount of primer constant at 125 ng. The conversion of nanograms of primer into picomoles was done using the following equation:

$$pmol\ of\ primer = \frac{ng\ of\ primer \times 1000}{330 \times number\ of\ bases}$$

Table 11. Component concentration and thermocycling conditions of a typical PCR with *PfuTurbo* DNA Polymerase (Stratagene/Agilent Technologies (Santa Clara, US)).

Component	Final concentration		Thermocycling conditions
Template DNA	0.2-1 ng/µl		95 °C, 30 sec
Forward primer	2.5 ng/µl	12 cycles	95 °C, 30 sec
Reverse primer	2.5 ng/µl		55 °C, 1 min
dNTPs	200 µM		68 °C, 1 min per kb
Cloned *Pfu* DNA Polymerase Reaction Buffer	1 x		4 °C, ∞
PfuTurbo DNA Polymerase	0.05 U/µl		

In a total volume of 50 µl

After completion of the PCR, the parental plasmid was digested by adding 0.2 U/µl DpnI to the reaction mixture and incubating at 37 °C for 1 h. Subsequently, 40 µl of XL1-Blue electrocompetent *E. coli* cells were transformed with 1 µl of reaction mixture as described in **IV.2.1.8**. The transformed cells were plated on selective LB agar medium and incubated at 37 °C overnight. Precultures were prepared from single colonies. Their DNA was isolated and sequenced in order to confirm a successful insertion of the point mutation.

IV.2.1.11. Site-Directed Ligase-Independent Mutagenesis (SLIM)

SLIM was performed as described by (Chiu *et al.*, 2008). Two separate SLIM-PCRs were set up, one with the primer pair F_T+R_S and one with F_S+R_T.

Table 12. Component concentration and thermocycling conditions of a typical SLIM-PCR with Phusion® High-Fidelity DNA Polymerase.

Component	Final concentration		Thermocycling conditions
Template DNA	0.2 ng/µl		98 °C, 30 sec
Forward primer (F_T/F_S)	0.4 µM		98 °C, 15 sec
Reverse primer (R_S/R_T)	0.4 µM	25 cycles	55 °C, 30 sec
dNTPs	200 µM		72 °C, 30 s per kb
Phusion HF Buffer	1 x		72 °C, 10 min
$MgSO_4$	2.5 mM		4 °C, ∞
Betaine	100 mM		
Phusion DNA Polymerase	0.02 U/µl		

In a total volume of 25 µl

After completion of the PCRs, the template DNAs were digested for one hour at 37 °C in 1 x D-Buffer containing 10 U DpnI. Subsequently, the reaction products were mixed and

set up for hybridization. Furthermore, two negative controls were set up, containing only one of the digestion products.

Table 13. Pipetting scheme for SLIM hybridization.

SLIM Hybridization		Control 1	Control 2	Thermocycling conditions	
Digestion product 1	10 µl	10 µl	-		99 °C, 3 min
Digestion product 2	10 µl	-	10 µl	25 cycles	65 °C, 5 min
5 x H-Buffer	10 µl	10 µl	10 µl		30 °C, 40 min
ddH_2O	20 µl	30 µl	30 µl		4 °C, ∞
	50 µl	50 µl	50 µl		

Finally, 40 µl of electrocompetent JM109 *E. coli* cells were transformed with either 1 µl of the SLIM-hybridization product, control 1 or control 2 as described in **IV.2.1.8**. The transformed cells were plated on selective LB agar medium and incubated at 37 °C overnight. Precultures were prepared from single colonies. Their DNA was isolated and sequenced in order to confirm a successful insertion of the point mutation.

Table 14. Composition of D- and H-Buffer used for SLIM.

D-Buffer	H-Buffer
20 mM Tris pH 8.0	125 mM Tris pH 9.0
20 mM $MgCl_2$	750 mM NaCl
5 mM DTT	100 mM EDTA pH 8.0

IV.2.1.12. Genome Integration of Large DNA Fragments

Integration of the *E. coli btuB* AdoCbl RS split reporter construct into the genome of *E. coli* JM109 was performed as described by (Kuhlman & Cox, 2010). The plasmids pTKRED and pTKIP were kindly provided by Prof. Thomas E. Kuhlman from the

Princeton University (US) (now University of Illinois, US). Instead of using the pTKS/CS plasmid, the tetracycline resistance (LP-*tetA*) was amplified from a template DNA with identical sequence, which was purchased from GeneArt AG (Thermo Fisher Scientific Inc., Regensburg, DE) (see **Figure 34**).

Generation of the donor plasmid pTKIP-reporter:

The entire *E. coli btuB* AdoCbl RS split reporter construct was excised from its original plasmid via digestion with EcoNI and PacI. pTKIP was digested with XbaI and XhoI. Blunt ends on both digestion products were generated as described in **IV.2.1.4**. The pTKIP backbone was dephosphorylated (see **IV.2.1.5**) and ligated with the reporter construct as described in **IV.2.1.6**.

Integration of LP-*tetA* into the genome of *E. coli* JM109:

Electrocompetent *E. coli* JM109 cells carrying the pTKRED plasmid were generated as described in **IV.2.1.7**. They were grown in SOB medium supplemented with 0.5 % (w/v) glucose, 100 µg/ml spectinomycin and 2mM IPTG. LP-*tetA* was amplified in a PCR under the following conditions:

Table 15. Component concentration and thermocycling conditions for the amplification of LP-*tetA*.

Component	Final concentration		Thermocycling conditions
Template DNA	0.1 ng/µl		98 °C, 30 sec
Forward primer	1 µM		98 °C, 15 sec
Reverse primer	1 µM	25 cycles	60 °C, 30 sec
dNTPs	200 µM		72 °C, 30 s per kb
Phusion HF Buffer	1 x		72 °C, 5 min
DMSO	10 %		4 °C, ∞
Phusion DNA Polymerase	0.02 U/µl		

pTKRED electrocompetent cells were transformed with about 1 μg LP-tet in a 1 mm electroporation cuvette (PEQLAB (Erlangen, DE)) as described in **IV.2.1.8**. Cells were resuspended in 2 ml SOC medium supplemented with 2 mM IPTG. After recovery at room temperature overnight, cells were plated on LB agar plates supplemented with 10 μg/ml tetracycline, 100 μg/ml spectinomycin and 0.5 % (w/v) glucose and grown overnight at 30 °C. Positive LP-*tetA* integrants that were verified via single colony PCR and DNA sequencing were made electrocompetent as described in **IV.2.1.7**.

Integration of the *E. coli btuB* AdoCbl RS reporter construct into the genome of *E. coli* JM109:

LP-*tetA* JM109 electrocompetent cells were transformed with pTKIP-reporter as described in **IV.2.1.8**. They were plated on LB agar plates supplemented with 10 μg/ml tetracycline, 100 μg/ml spectinomycin, 100 μg/ml ampicillin and 0.5 % (w/v) glucose and grown overnight at 30 °C. 5 ml of MOPS EZ rich defined medium (purchased from Teknova (Hollister, US)) supplemented with 0.5 % (w/v) glycerol, 2 mM IPTG, and 0.2 % (w/v) L-arabinose were inoculated with a single colony. After growing at 37 °C for 1 h, the temperature was reduced to 30 °C. After 30 min at 30 °C, 100 μg/ml spectinomycin was added and the cells were grown overnight at 30 °C. The next day, cultures were diluted $1:10^5$ and plated on LB agar plates supplemented with 100 μg/ml spectinomycin and 0.004 % (w/v) X-gal and grown at 30 °C. Potential integrants (blue colonies) were picked and plated on LB agar plated that contained 10 μg/ml tetracycline or 100 μg/ml ampicillin, respectively, to confirm the loss of the landing pad and the donor plasmid. Positive reporter construct integrants were verified via single colony PCR and DNA sequencing.

IV.2.1.13. Cloning of the CP Library Construct

The expressed CP library was generated as explained in detail by (Tavassoli & Benkovic, 2007). The I_C region of the *Ssp* DnaE split intein (108 nt, Gene ID: 951978) was

amplified with from *Synechocystis* sp PCC6803 genomic DNA with primers **oIntC-EcoRI_for** and **oIntC-SalI_rev** and inserted into the plasmid pBAD202/d/LacZ downstream of the araBAD promoter region between the NcoI/XhoI restriction sites. As XhoI cuts pBAD202/d/LacZ at the 3' end of *lacZ*, this procedure allowed the simultaneous excision of the *lacZ* gene from the vector, which is important regarding further applications of *lacZ* as a reporter gene. The I_N region (369 nt, Gene ID: 953863) was amplified similarly with primers **oIntN_for** and **oIntN-Hisrev**, introducing the His_6-tag via the reverse primer. The I_N-His_6 fragment was inserted into the plasmid pSGAT2 via HindIII/XhoI restriction sites. The CP library was generated by performing a PCR on I_N-His_6 with the library primer **oC+5_lib** and **oIntN-Hisrev.** As mismatches in the random amino acid sequence of the CP are likely to occur, an additional PCR with the 'zipper' primer **oZipper** and **oIntN-Hisrev** was performed on the lib-I_N-His_6 PCR fragment. Cloning of the CP library construct was completed by in frame insertion of the final lib-I_N-His_6 fragment downstream of I_C into the plasmid pBAD202/d/LacZ.

IV.2.1.14. Cloning of the *S. aureus xpt* guanine RS Fusion Reporter Construct

The 333 nt *S. aureus xpt* guanine RS (-333 nt to 0 nt region; relative to *xpt* translation start) was PCR amplified from *S. aureus* Mu50 genomic DNA with primers **oPurineRS_for** and **oPurineRS_rev**. The *lacZ* gene (Gene ID: 945006) was PCR amplified from *E. coli* K12 genomic DNA (Gene ID: 945006) with primers **oLacZ5_for** and **oLacZ2_rev.** The *sacB* gene was amplified from *B. subtilis* genomic DNA (Gene ID: 936413) with primers **oSacB_for** and **oSacB_rev**. The fragments were sequentially cloned into the plasmid pACYCDuet-1-P_{glnS}-*pylS* downstream of the *glnS* promoter between the restriction sites of BamHI/AflII.

IV.2.1.15. Cloning of the *E. coli btuB* AdoCbl RS Fusion Reporter Constructs

In order to prove the above statement (see **II.1.1.2**) that parts of the *btuB* gene are necessary to efficiently regulate gene expression, the *E. coli btuB* AdoCbl RS was PCR amplified in two different lengths. One version of the RS comprised the region from -240 nt to 0 nt (numbering relative to start of *btuB* translation, which is denoted +1) and another version included the region from -300 nt to +210 nt (210 nt of the *btuB* gene). In both constructs, the genes for *lacZ* and *sacB* were amplified as denoted in **IV.2.1.14**. The *E. coli btuB* AdoCbl RSs were PCR amplified from *E. coli* K12 genomic DNA with the primer pair **oB12RS_for**/ **oB12RS_rev** (-240 nt to 0 nt) and the primer pair **oB12RS2_for/oB12RS2_rev** (-300 nt to +210 nt). The fragments were sequentially cloned into the plasmid pACYCDuet-1-P_{glnS}-*pylS* downstream of the *glnS* promoter between the restriction sites of BamHI/AflII.

IV.2.1.16. Cloning of the *E. coli btuB* AdoCbl RS Split Reporter Construct

The *E. coli btuB* AdoCbl split reporter construct was only generated in one version with the -300 nt to +210 nt RS. The first RS and the *lacZ* gene were amplified and inserted into the first multiple cloning site of the pACYCDuet-1-P_{glnS}-*pylS* vector as described for the corresponding fusion reporter construct (see **IV.2.1.15**). A second *glnS* promoter was amplified from the same plasmid with primers **oGlnBsrI_for/oGlnXhoI_rev**. A second *E. coli btuB* AdoCbl RS was amplified from *E. coli* K12 genomic DNA with primers **oB12RS2_for/oB12RS3_rev**. *SacB* was amplified from *B. subtilis* genomic DNA (Gene ID: 936413) with primers **SacB_XhoI-SmaI_for/SacBPacI_rev**. P_{glnS}, RS and *sacB* were cloned into the second multiple cloning site of the pACYCDuet-1-P_{glnS}-*pylS* vector via restriction sites BsrGI/PacI.

IV.2.2. Protein Expression and Analysis

IV.2.2.1. Expression and Analysis of the CP Library Construct

E. coli JM109 cells were transformed with the CP library construct carrying the CLNSLS reference CP as described in section **IV.2.1.8**. 5 ml selective LB medium were inoculated with a single colony from the transformation and incubated overnight at 37 °C and 130 rpm shaking. This preculture was diluted in a ratio of 1:50 with selective LB medium and incubated at 37 °C and 130 rpm shaking. 50 ml cultures were prepared for SDS-PAGE analysis of I_N, and 1 l cultures for mass spectrum analysis of the reference CP CLNSLS. At an OD of 0.7-0.9, cells were induced with 0.2 % (w/v) L-arabinose and further incubated overnight at 37 °C and 130 rpm shaking. Cells were spun down at 20 °C and 5 000 x g for 15 min. The supernatant was decanted and the cell pellets were stored at -20 °C until further use. *E. coli* cell pellets were resuspended in a minimal amount of lysis buffer and lysed by sonication with 10 % amplitude for an overall duration of 1 min (alternating 1 sec pulse, 1 sec pause). Cells debris was pelleted by centrifugation at 12 000 x g for 15 min and the crude cell lysate was purified with Ni-NTA Spin Columns (Qiagen, Hilden, DE) as specified by the manufacturer. The elution fractions were analyzed via SDS-PAGE (see **IV.2.2.3**).

Table 16. Composition of Lysis, Wash and Elution Buffer used for Ni-NTA purification of His_6-tagged proteins.

Lysis Buffer	Wash Buffer	Elution Buffer
20 mM Tris-HCl, pH 8.0	20 mM Tris-HCl, pH 8.0	20 mM Tris-HCl, pH 8.0
500 mM NaCl	500 mM NaCl	500 mM NaCl
10 mM imidazole	15 mM Imidazole	500 mM Imidazole
2 mM β-mercaptoethanol	2 mM β-mercaptoethanol	2 mM β-mercaptoethanol

IV.2.2.2. *n*-BuOH Extraction of Expressed CPs

E. coli cells were lysed as described in 2.3.1. The crude cell lysate was extracted with an equal amount of *n*-BuOH upon vigorous mixing. The organic and the aqueous layer were separated by centrifugation at 21 000 rpm at 4 °C. The organic phase was carefully removed and dried in vacuo. The cell extract was dissolved in a mixture of MeOH/H_2O and analyzed via HPLC-MS.

IV.2.2.3. SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Discontinous SDS-PAGE was performed in a Hoefer™ SE260 Mighty Small™ II Deluxe Mini Vertical Electrophoresis Unit according to Laemmli (Laemmli, 1970). A current of 35-40 mA per gel was applied for 60 min or until the frontline has reached the bottom of the gel. Protein bands were stained for 30 min in preheated Coomassie solution and destained in 10 % acetic acid.

Table 17. Pipetting scheme for two SDS-gels.

Component	Separating Gel (15 %)	Stacking Gel (4 %)
ddH_2O	4.4 ml	3.1 ml
Separating Gel Buffer	3.0 ml	-
Stacking Gel Buffer	-	1.25 ml
40 % (w/v) Acrylamide	4.5 ml	0.625 ml
10 % (v/v) APS	120 µl	50 µl
1 % (v/v) TEMED	12 µl	5 µl

Table 18. Buffers used in SDS-PAGE.

Separating Gel Buffer	Stacking Gel Buffer	Running Buffer
1.5 M Tris-HCl, pH 8.8	0.5 M Tris-HCl, pH 6.8	25 mM Tris-HCl, pH 8.2-8.3
0.4 % (w/v) SDS	0.4 % (w/v) SDS	192 mM glycine 0.1 % (w/v) SDS

Table 19. Solutions used in SDS-PAGE.

40 % (w/v) Acrylamide	Coomassie Solution	SDS Loading Dye
39 % (w/v) acrylamide	0.05 % (w/v) Coomassie brilliant blue R250	60 mM Tris-HCl, pH 6.8
1.2 % (w/v) bisacrylamide	25 % (v/v) isopropanol	30 % (v/v) glycerol
	10 % (v/v) acetic acid	10 % (w/v) sucrose
		5 % (w/v) SDS
		0.02 % (w/v) bromphenol blue
		3 % (v/v) β-mercaptoethanol

IV.2.2.4. Expression and Analysis of WT-PylRS

The plasmid pET28a that carries the catalytic domain (residues 185-454) of the PylRS from *M. mazei* Go1 (Gene ID: 1479787) was a kind gift of the group of Prof. Carell from the LMU Munich. *E. coli* Rosetta DE3 cells were transformed with the plasmid as described in **IV.2.1.8**. 5 ml selective LB medium were inoculated with a single colony from the transformation and incubated overnight at 37 °C and 130 rpm shaking. This preculture was diluted in a ratio of 1:50 with selective LB medium and incubated at 37 °C and 130 rpm shaking. At an OD of 0.7-0.9, cells were induced with 1mM IPTG and further incubated for four hours at 30 °C and 130 rpm shaking. Cells were harvested by centrifugation at 20 °C and 5 000 x g for 15 min, the supernatant was decanted and the cell pellets were stored at -20 °C until further use. The protein was purified as described previously (Kavran *et al.*, 2007; Schneider *et al.*, 2013).

IV.2.3. Reporter Gene Assays

IV.2.3.1. β-Galactosidase Assay

β-Galactosidase expression was assayed as described elsewhere (Zhang & Bremer, 1995) with the following modifications. The assay was performed with cultures that had an OD_{600} value above 0.3. Each sample was measured as triplicate. Depending on expression strength, 80-95 µl of permeabilization solution were mixed with 5-20 µl (total volume: 100 µl) of *E. coli* cells and incubated at 20-30 min at 30 °C. Addition of 600 µl preheated (30 °C) substrate solution started the time measurement. After sufficient color had developed (A_{420} between 0.2-1.0), 700 µl of stop solution were added, mixed well and the stop time was noted. Samples were cleared by centrifugation for 5 min at full speed. The supernatant was carefully transferred into disposable cuvettes and A_{420} was recorded. Miller Units were calculated according to the following equation:

$$Miller\ Units = 1000 \times \frac{\mathrm{A420}}{A600 \times sample\ volume\ (ml) \times reaction\ time\ (min)}$$

Table 20. Solutions used in β-galactosidase assay.

Permeabilization Solution	Substrate Solution	Stop Solution
100 mM Na_2HPO_4	60 mM Na_2HPO_4	1 M Na_2CO_3
20 mM KCl	40 mM NaH_2PO_4	
2 mM $MgSO_4$	1 mg/ml ONPG	
0.8 mg/ml CTAB	2.7 µl/ml β-mercaptoethanol	
0.4 mg/ml sodium deoxycholate		
5.4 µl/ml β-mercaptoethanol		

IV.2.3.2. Sucrose Cell Growth Assay

Levansucrase expression was analyzed by determining the ability of *E. coli* cells to grow on sucrose containing medium. For this purpose, 5 ml selective LB medium were inoculated with a single colony of *E. coli* cells carrying the *sacB* gene and incubated overnight at 37 °C and 130 rpm shaking. This preculture was diluted in a ratio of 1:100 with selective LB medium with or without 0.5 % (w/v) sucrose and incubated at 37 °C and 130 rpm shaking. In regular time intervals, medium was removed from the culture and the OD_{600} was determined. The values for the OD_{600} were plotted against the time of culture growth.

IV.2.3.3. *In vivo* CP Screening Assay

50 µl of electrocompetent *E. coli* JM109 cells carrying the *E. coli btuB* AdoCbl *lacZ*/*sacB* split reporter construct were transformed with 1 µl of the CP library construct as described in 2.1.9. After electroporation, the cells were immediately resuspended in 5 ml Glycerol-SOC medium and allowed to recover at 37 °C for 1 h upon agitation. After recovery, 0.2 % (w/v) L-arabinose and appropriate antibiotics were added to the medium and the cells were further incubated overnight at 37 °C upon agitation. The next day, cells were pelleted by centrifugation at 5 000 x g for 5 min and resuspended in 0.9 % (w/v) saline. The entire amount of cells was distributed on 4 petri dishes (diameter 13.5 cm) with selective CP screening medium and incubated for 48 h at 37 °C. White colonies where picked, inoculated in 1 ml LB medium and cultivated overnight at 37 °C and 130 rpm shaking. Cells were diluted in a ratio of $1:10^3$, $1:10^4$, $1:10^5$ and $1:10^6$ with 0.9 % saline. 5 µl drops from each dilution were pipetted on selective CP screening medium with and without 0.2 % L-arabinose. Colonies that showed a growth advantage on plates supplemented with 0.2 % L-arabinose were subject to further testing with β-galactosidase assay and sucrose cell growth assay in the absence and presence of 0.2 % L-arabinose.

IV.2.4. Protein Cristallography

IV.2.4.1. Co-Cristallization of PylRS in Complex with Kalk, Kbu, Kcr and Kpr

In order to co-crystallize PylRS with the non-natural amino acids Kalk, Kbu, Kcr and Kpr, the protein was diluted to 1 mg/ml and incubated for 2 h with 1 mM of the respective amino acid and 1 mM of ATP in protein storage buffer (10 mM HEPES pH 7.4, 300 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT). The protein was concentrated to 10 mg/ml prior to crystallisation using Vivaspin® 6 Centrifugal Concentrators (Sartorius, Göttingen, DE) with 10 kDa cutoff. Initial crystallization screens were performed with crystallization buffer kits from Qiagen (Hilden, DE) in 96-well Intelli-Plates (Art Robbins Instruments (Sunnyvale, US)) using the sitting-drop vapor diffusion method. The crystallization experiments were set up with a Phoenix Liquid Handling System (Art Robbins Instruments (Sunnyvale, US)) with ratios of 1:1, 2:1 and 1:2 of protein and well solution. The hit conditions were further optimized in SuperClear Pregreased 24-well plates (Crystalgen (New York, US)) using the hanging-drop vapor diffusion method. Crystals appeared overnight at 20 °C in drops consisting of a 1:2 mixture of protein and well solution (100 mM lithium acetate, 10-14 % (w/v) PEG3350). All crystals were cryoprotected with well solution supplemented with 30 % (w/v) ethylene glycol before flash-freezing and storing in liquid nitrogen.

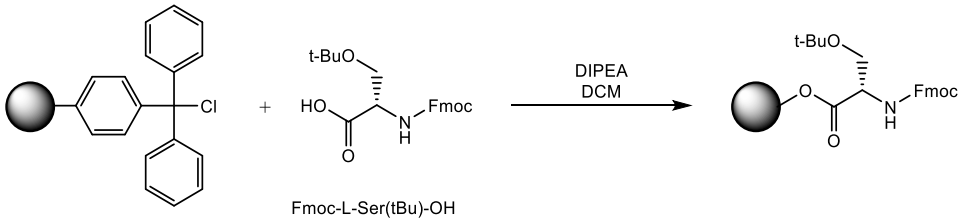
IV.2.4.2. Data Collection and Structure Determination

Diffraction data were collected at the synchrotron beam lines PXI and PXIII (Swiss Light Source, Villigen, Switzerland). The data were processed with XDS (Kabsch, 2010) to 2.05 Å (Kalk), 2.28 Å (Kbu), 2.15 Å (Kcr) and 2.2 Å (Kpr) spacing, respectively, ensuring consistent indexing and choosing the same set of free reflections. The Crystals belong to the same space group as reported previously for the WT-PylRS (Kavran *et al.*, 2007) ($P6_4$, unit cell dimensions: a = b = 105 Å, c = 71 Å). The structure was solved by molecular replacement using the PylRS coordinates (PDB code: 4BW9) in PHASER

(McCoy *et al.*, 2007; Winn *et al.*, 2011). For the Kalk, Kcr and Kpr complexes the coordinates of the protein atoms of the PylRS-adenylated Kbu complex were used in rigid body refinement in REFMAC (Murshudov *et al.*, 2011). In order to reduce model bias, all non-protein atoms as well as the loop region around Phe 384 were removed from the model prior to molecular replacement/rigid body refinement, and the temperature factors were reset, followed by simulated annealing in PHENIX (Adams *et al.*, 2010). Clear peaks for AMP-Kalk, AMP-Kbu, AMP-Kcr and AMP-Kpr were visible in the simulated-annealing omit Fo-DFc map. Rounds of model building and refinement were carried out in COOT (Emsley *et al.*, 2010) and REFMAC. The refinement parameter file for the adenylated amino acids was generated with prodrg2 (Schuttelkopf & van Aalten, 2004), as implemented in COOT, and the TLSMD server (Painter & Merritt, 2006) was used to determine TLS groups for TLS refinement (Winn *et al.*, 2003). Diffraction data and refinement statistics are summarized in **Table 22**. Structural superpositions were done with SSM (Krissinel & Henrick, 2004). Two-dimensional interaction plots were carried out with LIGPLOT (Wallace *et al.*, 1995). All structural figures were prepared with PyMol (Delano Scientific, San Carlos, CA). Atomic coordinates were submitted to the Protein Data Bank (http://www.ebi.ac.uk/pdbe/) with the PDB codes: 4CH6 (PylRS-Kalk), 4CH3 (PylRS-Kbu), 4CH4 (PylRS-Kcr) and 4CH5 (PylRS-Kpr).

IV.2.5. Chemical Synthesis of Cyclic Peptides

IV.2.5.1. Resin Loading



Fmoc-L-Ser(tBu)-OH

TCP resin (loading 1.4 mmol/g, 200 mg, 0.28 mmol, 1 eq) was filled into a suitable syringe (5 ml for 200 mg resin) equipped with a PE-filter. Fmoc-L-Ser(tBu)-OH (130 mg,

0.34 mmol, 1.2 eq) was dissolved in 2 ml DCM_{dry} and mixed with DIPEA (134 µl, 176 mg, 1.36 mmol, 4 eq rel. to AS). This mixture was sucked directly into the syringe and shaken for 2 h.

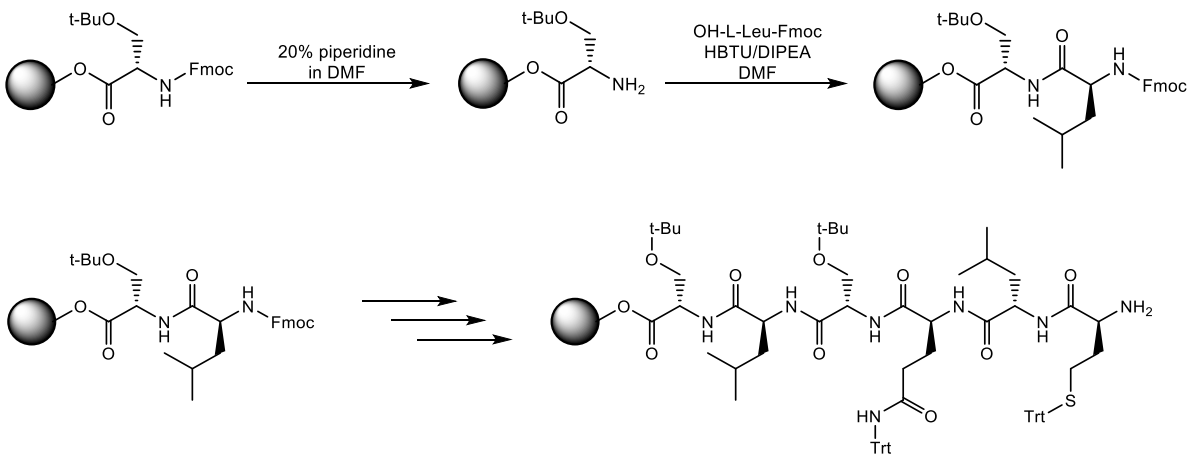
For capping unreacted resin, MeOH (0.2 ml per g resin, 40 µl, 32 mg, 1 mmol) was mixed with DIPEA (34 µl, 26 mg, 0.20 mmol, 0.2 eq rel. to MeOH), sucked into the syringe with the reaction solution and shaken for 20 min. The solution was drained from the syringe and the resin was washed three times with DCM, three times with DMF, two times with DMF/MeOH (1:1), three times with MeOH and dried under vacuum.

The real loading was calculated according to the following equation and used as loading for coupling further amino acids:

$$loading\ \left[\frac{mmol}{g}\right] = 1000\times\frac{m(loaded) - m(unloaded)}{(MW - 36.461\frac{g}{mol})\times m(loaded)}$$

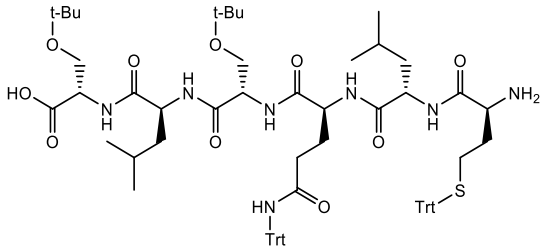
m(loaded) = mass of loaded resin [g]
m(unloaded) = mass of unloaded resin [g]
MW = molecular weight of immobilized amino acid [g/mol]

IV.2.5.2. Solid-Phase Synthesis of Linear Peptides



Linear peptides were synthesized on a PS3 peptide synthesizer (Protein Technologies, Tucson, US) in 0.2 mmol scale with double excess of Fmoc-amino acids (0.4 mmol) relative to TCP resin. Fmoc deprotection was performed using 20 % (v/v) piperidine in DMF and coupling was performed using amino acid/HBTU/DIPEA in molar ratios of 1 : 1 : 2 in DMF. After synthesis, the resin-bound linear peptide was dried under vacuum.

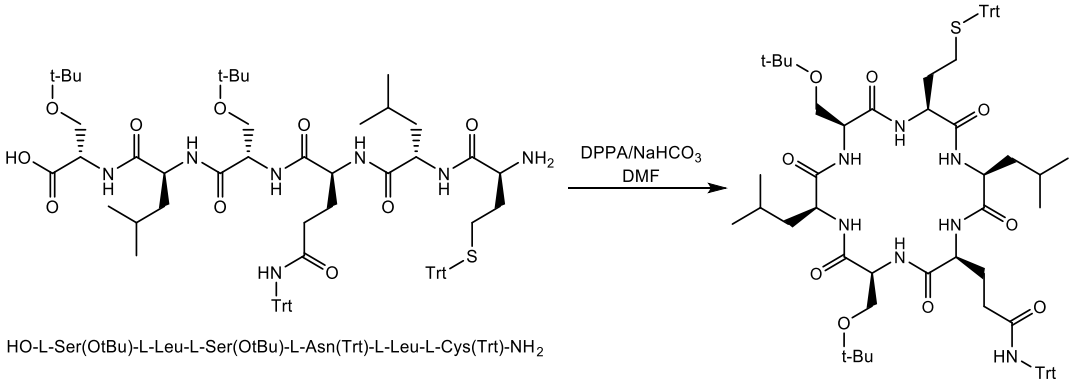
IV.2.5.3. Cleavage



HO-L-Ser(OtBu)-L-Leu-L-Ser(OtBu)-L-Asn(Trt)-L-Leu-L-Cys(Trt)-NH_2

The resin was transferred from the peptide synthesizer reaction vessel into a suitable syringe (5 ml for 200 mg resin) equipped with a PE-filter. HFIP (20 % in DCM, 3 ml) was sucked directly into the syringe and the reaction mixture was shaken for 15 min. The solvent was drained and the cleavage procedure was repeated three times more. Finally the resin was washed three times with cleavage solution. The solvent was evaporated under reduced pressure and then two times co-evaporated with toluene, yielding the linear peptide with the sequence H_2N-CLNSLS-OH (217 mg, 0.17 mmol, 85 %).

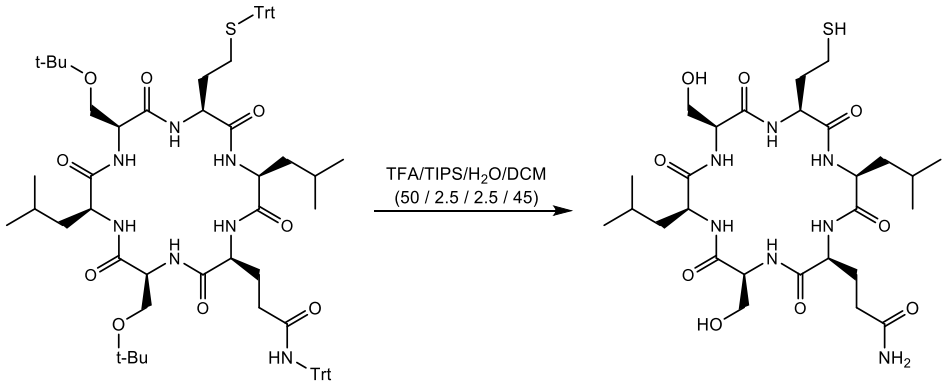
IV.2.5.4. Backbone Cyclization



HO-L-Ser(OtBu)-L-Leu-L-Ser(OtBu)-L-Asn(Trt)-L-Leu-L-Cys(Trt)-NH_2

50 ml DMF, DPPA (110 µl, 140 mg, 0.51 mmol, 3 eq) and $NaHCO_3$ (71.4 mg, 0.85 mmol, 5 eq) were added to the linear peptide (217 mg, 0.17 mmol, 1 eq) and stirred at room temperature overnight. The solvent was evaporated under reduced pressure followed by an extraction in EtOAc/H_2O. The organic layer was dried on Na_2SO_4 and the solvent evaporated *in vacuo*. The final purification was achieved by means of HPLC (RP18, gradient 50-70 % acetonitrile/0.1 % TFA) (Zimmer *et al.*, 1993).

IV.2.5.5. Side-Chain Deprotection



5 ml of a mixture of TFA/TIPS/H_2O/DCM (50/2.5/2.5/45) was added to the side-chain protected cyclic peptide and stirred at room temperature for 2 h. The side-chain deprotected cyclic peptide was precipitated by addition of diethyl ether (10 ml per 1 ml deprotection mixture). The precipitate was spun down in a centrifuge, washed twice with diethyl ether and dried under vacuum.

V. Abbreviations

%	percent
Å	angstrom
A	absorbance
aaRS	aminoacyl-tRNA synthetase
AdoCbl	adenosylcobalamin
AMP	adenosine monophosphate
APS	ammonium persulfate
b	base
Boc	*tert*-butyloxycarbonyl
bp	base pair
BSA	bovine serum albumin
B. subtilis	*Bacillus subtilis*
°C	degree Celsius
cAMP	3'-5'-cyclic adenosine monophosphate
CAP	catabolite activator protein
Cbl	cobalamin
CNCbl	cyanocobalamin
CP	cyclic peptide
CTAB	cetyltrimethylammonium bromide
Da	Dalton
DCM	dichloromethane
DIPEA	*N*,*N*-diisopropylethylamine
DMF	dimethylformamide

DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DPPA	diphenylphosphoryl azide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
E. coli	*Escherichia Coli*
eq	equivalent
EtOAc	ethyl acetate
ESI	electrospray ionization
FMN	flavin mononucleotide
Fmoc	9-fluorenylmethyloxycarbonyl
g	gram
GlcN6P	glucosamine-6- phosphate
h	hour
HAT	histone acetyltransferases
HBTU	*O*-(1*H*-benzotriazol-1-yl)-*N*,*N*,*N'*,*N'*-tetramethyluronium Hexafluorophosphate
HDAC	histone deacylases
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFIP	1,1,1,3,3,3-hexafluoro-2-propanol
His_6	hexahistidin
HIV	human immunodeficiency virus
HPLC	high-performance liquid chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
IntC/I_C	C-terminal split-intein domain
IntN/I_N	N-terminal split-intein domain
k	kilo
Kalk	ε–*N*-propargyloxy-carbonyl-lysine
Kbu	ε–*N*-butyryl-lysine
Kcr	ε–*N*-crotonyl-lysine
Kpr	ε–*N*-propionyl-lysine

l	liter
LB	lysogeny broth
LC	liquid chromatography
LP	landing pad
LysRS	lysyl-tRNA synthetase
µ	micro
M	molar
m	milli
mA	milliampere
MeCbl	methylcobalamin
MeOH	methanol
min	minute
M. mazei	*Methanosarcina mazei*
m/z	mass-to-charge ratio
mRNA	messenger RNA
MRSA	methicillin-resistant *S. aureus*
MS	mass spectrometry
MtbB	dimethylamine methyltransferase
MtmB	monomethylamine methyltransferase
MttB	trimethylamine methyltransferase
MW	molecular weight
n	nano
Ni-NTA	nickel-nitrilotriacetic acid
Npu	*Nostoc punctiforme*
OD_{600}	optical density at 600 nm wavelength
ONPG	*ortho*-nitrophenyl-β-galactoside
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PE	polyethylene
$preQ_1$	pre-queuosine 1
PTM	posttranslational modification
Pyl	pyrrolysine

PylRS	pyrrolysyl-tRNA synthetase
RBS	ribosome binding site
RNA	ribonucleic acid
rpm	revolutions per minute
RS	riboswitch
SAM	S-adenosylmethionine
S. aureus	*Staphylococcus aureus*
SDS	sodium dodecyl sulfate
sec	second
SICLOPPS	split-intein circular ligation of peptides and proteins
SLIM	site-directed ligase-independent mutagenesis
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
Ssp	*Synechocystis* sp. strain PCC6803
TAE	Tris-acetate-EDTA
TCP	trityl chloride polystyrene
Tet	tetracycline
TFA	trifluoroacetic acid
TIPS	triisopropylsilane
Tm	melting temperature
TPP	thiamine pyrophosphate
tRNAaaCUA	aminoacyl tRNA carrying a CUA anticodon
Tris	tris(hydroxymethyl)aminomethane
TyrRS	tyrosyl-tRNA synthetase
U	unit
UTR	untranslated region
V	volt
VRE	vancomycin-resistant enterococci
v/v	volume per volume
WT	wild type
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside

VI. References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. & Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, **66**, 213-221.

Appasamy, S.D., Ramlan, E.I. & Firdaus-Raih, M. (2013) Comparative sequence and structure analysis reveals the conservation and diversity of nucleotide positions and their associated tertiary interactions in the riboswitches. *PLoS One*, **8**, e73984.

Arnaudo, A.M. & Garcia, B.A. (2013) Proteomic characterization of novel histone post-translational modifications. *Epigenetics Chromatin*, **6**, 24.

Baugh, L., Gallagher, L.A., Patrapuvich, R., Clifton, M.C., Gardberg, A.S., Edwards, T.E., Armour, B., Begley, D.W., Dieterich, S.H., Dranow, D.M., Abendroth, J., Fairman, J.W., Fox, D., 3rd, Staker, B.L., Phan, I., Gillespie, A., Choi, R., Nakazawa-Hewitt, S., Nguyen, M.T., Napuli, A., Barrett, L., Buchko, G.W., Stacy, R., Myler, P.J., Stewart, L.J., Manoil, C. & Van Voorhis, W.C. (2013) Combining functional and structural genomics to sample the essential Burkholderia structome. *PLoS One*, **8**, e53851.

Blount, K.F. & Breaker, R.R. (2006) Riboswitches as antibacterial drug targets. *Nat Biotechnol*, **24**, 1558-1564.

Bocobza, S.E. & Aharoni, A. (2008) Switching the light on plant riboswitches. *Trends Plant Sci*, **13**, 526-533.

Breaker, R.R. (2012) Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol*, **4**.

Chen, P.R., Groff, D., Guo, J., Ou, W., Cellitti, S., Geierstanger, B.H. & Schultz, P.G. (2009) A facile system for encoding unnatural amino acids in mammalian cells. *Angew Chem Int Ed Engl*, **48**, 4052-4055.

Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S.C., Falck, J.R., Peng, J., Gu, W. & Zhao, Y. (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol Cell Proteomics*, **6**, 812-819.

Chiu, J., Tillett, D., Dawes, I.W. & March, P.E. (2008) Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *J Microbiol Methods*, **73**, 195-198.

Cole, P.A. (2008) Chemical probes for histone-modifying enzymes. *Nat Chem Biol*, **4**, 590-597.

Craik, D.J. (2006) Chemistry. Seamless proteins tie up their loose ends. *Science*, **311**, 1563-1564.

Crick, F. (1970) Central dogma of molecular biology. *Nature*, **227**, 561-563.

de Hoon, M.J., Makita, Y., Nakai, K. & Miyano, S. (2005) Prediction of transcriptional terminators in Bacillus subtilis and related species. *PLoS Comput Biol*, **1**, e25.

Dedonder, R. (1966) [86] Levansucrase from Bacillus subtilis. In Elizabeth F. Neufeld, V.G. (ed) *Methods in enzymology*. Academic Press, pp. 500-505.

Deigan, K.E. & Ferre-D'Amare, A.R. (2011) Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs. *Acc Chem Res*, **44**, 1329-1338.

Dennesen, P.J., Bonten, M.J. & Weinstein, R.A. (1998) Multiresistant bacteria as a hospital epidemic problem. *Ann Med*, **30**, 176-185.

Desogus, G., Todone, F., Brick, P. & Onesti, S. (2000) Active site of lysyl-tRNA synthetase: structural studies of the adenylation reaction. *Biochemistry*, **39**, 8418-8425.

Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.

Flügel, V., Vrabel, M. & Schneider, S. (2014) Structural basis for the site-specific incorporation of lysine derivatives into proteins. *PLoS One*, **9**, e96198.

Franklund, C.V. & Kadner, R.J. (1997) Multiple transcribed elements control expression of the Escherichia coli btuB gene. *J Bacteriol*, **179**, 4039-4042.

Garrity, J., Gardner, J.G., Hawse, W., Wolberger, C. & Escalante-Semerena, J.C. (2007) N-lysine propionylation controls the activity of propionyl-CoA synthetase. *J Biol Chem*, **282**, 30239-30245.

Garst, A.D., Edwards, A.L. & Batey, R.T. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, **3**.

Gattner, M.J., Vrabel, M. & Carell, T. (2013) Synthesis of epsilon-N-propionyl-, epsilon-N-butyryl-, and epsilon-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem Commun (Camb)*, **49**, 379-381.

Gilbert, S.D., Mediatore, S.J. & Batey, R.T. (2006) Modified pyrimidines specifically bind the purine riboswitch. *J Am Chem Soc*, **128**, 14214-14215.

Gilbert, S.D., Reyes, F.E., Edwards, A.L. & Batey, R.T. (2009) Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure*, **17**, 857-868.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849-857.

Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol*, **177**, 4121-4130.

Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A. & Chan, M.K. (2002) A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science*, **296**, 1462-1466.

Hayouka, Z., Hurevich, M., Levin, A., Benyamini, H., Iosub, A., Maes, M., Shalev, D.E., Loyter, A., Gilon, C. & Friedler, A. (2010) Cyclic peptide inhibitors of HIV-1 integrase derived from the LEDGF/p75 protein. *Bioorg Med Chem*, **18**, 8388-8395.

Jagadish, K., Borra, R., Lacey, V., Majumder, S., Shekhtman, A., Wang, L. & Camarero, J.A. (2013) Expression of fluorescent cyclotides using protein trans-splicing for easy monitoring of cyclotide-protein interactions. *Angew Chem Int Ed Engl*, **52**, 3126-3131.

Johnson, J.E., Jr., Reyes, F.E., Polaski, J.T. & Batey, R.T. (2012) B12 cofactors directly stabilize an mRNA regulatory switch. *Nature*, **492**, 133-137.

Joo, S.H. (2012) Cyclic Peptides as Therapeutic Agents and Biochemical Tools. *Biomol Ther (Seoul)*, **20**, 19-26.

Juers, D.H., Matthews, B.W. & Huber, R.E. (2012) LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. *Protein Sci*, **21**, 1792-1807.

Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr*, **66**, 133-144.

Kavran, J.M., Gundllapalli, S., O'Donoghue, P., Englert, M., Soll, D. & Steitz, T.A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci U S A*, **104**, 11268-11273.

Kaya, E., Gutsmiedl, K., Vrabel, M., Muller, M., Thumbs, P. & Carell, T. (2009) Synthesis of threefold glycosylated proteins using click chemistry and genetically encoded unnatural amino acids. *Chembiochem*, **10**, 2858-2861.

Kaya, E., Vrabel, M., Deiml, C., Prill, S., Fluxa, V.S. & Carell, T. (2012) A genetically encoded norbornene amino acid for the mild and selective modification of proteins in a copper-free click reaction. *Angew Chem Int Ed Engl*, **51**, 4466-4469.

Kim, C.H., Kang, M., Kim, H.J., Chatterjee, A. & Schultz, P.G. (2012) Site-specific incorporation of epsilon-N-crotonyllysine into histones. *Angew Chem Int Ed Engl*, **51**, 7246-7249.

Kim, J.N., Blount, K.F., Puskarz, I., Lim, J., Link, K.H. & Breaker, R.R. (2009) Design and antimicrobial action of purine analogues that bind Guanine riboswitches. *ACS Chem Biol*, **4**, 915-927.

Kim, J.N. & Breaker, R.R. (2008) Purine sensing by riboswitches. *Biol Cell*, **100**, 1-11.

Kinsella, T.M., Ohashi, C.T., Harder, A.G., Yam, G.C., Li, W., Peelle, B., Pali, E.S., Bennett, M.K., Molineaux, S.M., Anderson, D.A., Masuda, E.S. & Payan, D.G. (2002) Retrovirally delivered random cyclic Peptide libraries yield inhibitors of interleukin-4 signaling in human B cells. *J Biol Chem*, **277**, 37512-37518.

Kohli, R.M., Walsh, C.T. & Burkart, M.D. (2002) Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature*, **418**, 658-661.

Krissinel, E. & Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr*, **60**, 2256-2268.

Kritzer, J.A., Hamamichi, S., McCaffery, J.M., Santagata, S., Naumann, T.A., Caldwell, K.A., Caldwell, G.A. & Lindquist, S. (2009) Rapid selection of cyclic peptides that reduce alpha-synuclein toxicity in yeast and animal models. *Nat Chem Biol*, **5**, 655-663.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. & Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.

Kuhlman, T.E. & Cox, E.C. (2010) Site-specific chromosomal integration of large synthetic constructs. *Nucleic Acids Res*, **38**, e92.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Laupacis, A., Keown, P.A., Ulan, R.A., McKenzie, N. & Stiller, C.R. (1982) Cyclosporin A: a powerful immunosuppressant. *Can Med Assoc J*, **126**, 1041-1046.

Lawrence, J.G. & Roth, J.R. (1996) Evolution of coenzyme B12 synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. *Genetics*, **142**, 11-24.

Li, G.M. (2013) Decoding the histone code: Role of H3K36me3 in mismatch repair and implications for cancer susceptibility and therapy. *Cancer Res*, **73**, 6379-6383.

Li, W.T., Mahapatra, A., Longstaff, D.G., Bechtel, J., Zhao, G., Kang, P.T., Chan, M.K. & Krzycki, J.A. (2009) Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *J Mol Biol*, **385**, 1156-1164.

Lin-Chao, S. & Bremer, H. (1986) Effect of the bacterial growth rate on replication control of plasmid pBR322 in Escherichia coli. *Mol Gen Genet*, **203**, 143-149.

Liu, T., Joo, S.H., Voorhees, J.L., Brooks, C.L. & Pei, D. (2009) Synthesis and screening of a cyclic peptide library: discovery of small-molecule ligands against human prolactin receptor. *Bioorg Med Chem*, **17**, 1026-1033.

Lundrigan, M.D., Koster, W. & Kadner, R.J. (1991) Transcribed sequences of the Escherichia coli btuB gene control its expression and regulation by vitamin B12. *Proc Natl Acad Sci U S A*, **88**, 1479-1483.

Mandal, M. & Breaker, R.R. (2004a) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol*, **11**, 29-35.

Mandal, M. & Breaker, R.R. (2004b) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*, **5**, 451-463.

Manna, A.K., Kumar, A., Ray, U., Das, S., Basu, G. & Roy, S. (2013) A cyclic peptide mimic of an RNA recognition motif of human La protein is a potent inhibitor of hepatitis C virus. *Antiviral Res*, **97**, 223-226.

Mattick, J.S. (2009) Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Ann N Y Acad Sci*, **1178**, 29-46.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. & Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.

Milles, S., Tyagi, S., Banterle, N., Koehler, C., VanDelinder, V., Plass, T., Neal, A.P. & Lemke, E.A. (2012) Click strategies for single-molecule protein fluorescence. *J Am Chem Soc*, **134**, 5187-5195.

Montange, R.K. & Batey, R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys*, **37**, 117-133.

Montellier, E., Rousseaux, S., Zhao, Y. & Khochbin, S. (2012) Histone crotonylation specifically marks the haploid male germ cell gene expression program: post-meiotic male-specific gene expression. *Bioessays*, **34**, 187-193.

Mulhbacher, J., Brouillette, E., Allard, M., Fortier, L.C., Malouin, F. & Lafontaine, D.A. (2010) Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog*, **6**, e1000865.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. & Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr*, **67**, 355-367.

Nahvi, A., Barrick, J.E. & Breaker, R.R. (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res*, **32**, 143-150.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L. & Breaker, R.R. (2002) Genetic control by a metabolite binding mRNA. *Chem Biol*, **9**, 1043.

Naumann, T.A., Tavassoli, A. & Benkovic, S.J. (2008) Genetic selection of cyclic peptide Dam methyltransferase inhibitors. *Chembiochem*, **9**, 194-197.

Nguyen, D.P., Lusic, H., Neumann, H., Kapadnis, P.B., Deiters, A. & Chin, J.W. (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J Am Chem Soc*, **131**, 8720-8721.

Nou, X. & Kadner, R.J. (1998) Coupled changes in translation and transcription during cobalamin-dependent regulation of btuB expression in Escherichia coli. *J Bacteriol*, **180**, 6719-6728.

Oyston, P.C., Fox, M.A., Richards, S.J. & Clark, G.C. (2009) Novel peptide therapeutics for treatment of infections. *J Med Microbiol*, **58**, 977-987.

Painter, J. & Merritt, E.A. (2006) TLSMD web server for the generation of multi-group TLS models. *J Appl Crystallogr*, **39**, 109-111.

Paulsson, J. & Ehrenberg, M. (2001) Noise in a minimal regulatory network: plasmid copy number control. *Q Rev Biophys*, **34**, 1-59.

Pena-Altamira, L.E., Polazzi, E. & Monti, B. (2013) Histone post-translational modifications in Huntington's and Parkinson's diseases. *Curr Pharm Des*, **19**, 5085-5092.

Perona, J.J. & Hadd, A. (2012) Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry*, **51**, 8705-8729.

Peselis, A. & Serganov, A. (2012) Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nat Struct Mol Biol*, **19**, 1182-1184.

Qin, C., Bu, X., Zhong, X., Ng, N.L. & Guo, Z. (2004) Optimization of antibacterial cyclic decapeptides. *J Comb Chem*, **6**, 398-406.

Quintero-Ronderos, P. & Montoya-Ortiz, G. (2012) Epigenetics and autoimmune diseases. *Autoimmune Dis*, **2012**, 593720.

Ravnum, S. & Andersson, D.I. (1997) Vitamin B12 repression of the btuB gene in Salmonella typhimurium is mediated via a translational control which requires leader and coding sequences. *Mol Microbiol*, **23**, 35-42.

Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. (2002) A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl*, **41**, 2596-2599.

Roth, A., Winkler, W.C., Regulski, E.E., Lee, B.W., Lim, J., Jona, I., Barrick, J.E., Ritwik, A., Kim, J.N., Welz, R., Iwata-Reuyl, D. & Breaker, R.R. (2007) A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat Struct Mol Biol*, **14**, 308-317.

Roth, J.R., Lawrence, J.G. & Bobik, T.A. (1996) Cobalamin (coenzyme B12): synthesis and biological significance. *Annu Rev Microbiol*, **50**, 137-181.

Saleh, L. & Perler, F.B. (2006) Protein splicing in cis and in trans. *Chem Rec*, **6**, 183-193.

Santoro, S.W., Wang, L., Herberich, B., King, D.S. & Schultz, P.G. (2002) An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat Biotechnol*, **20**, 1044-1048.

Satoh, T., Li, S., Friedman, T.M., Wiaderkiewicz, R., Korngold, R. & Huang, Z. (1996) Synthetic peptides derived from the fourth domain of CD4 antagonize off function and inhibit T cell activation. *Biochem Biophys Res Commun*, **224**, 438-443.

Schneider, S., Gattner, M.J., Vrabel, M., Flügel, V., Lopez-Carrillo, V., Prill, S. & Carell, T. (2013) Structural insights into incorporation of norbornene amino acids for click modification of proteins. *Chembiochem*, **14**, 2114-2118.

Schuttelkopf, A.W. & van Aalten, D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*, **60**, 1355-1363.

Scott, C.P., Abel-Santos, E., Jones, A.D. & Benkovic, S.J. (2001) Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem Biol*, **8**, 801-815.

Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C. & Benkovic, S.J. (1999) Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A*, **96**, 13638-13643.

Serganov, A., Huang, L. & Patel, D.J. (2009) Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature*, **458**, 233-237.

Serganov, A., Yuan, Y.R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R. & Patel, D.J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol*, **11**, 1729-1741.

Shah, N.H., Eryilmaz, E., Cowburn, D. & Muir, T.W. (2013) Naturally split inteins assemble through a "capture and collapse" mechanism. *J Am Chem Soc*, **135**, 18673-18681.

Shah, N.H. & Muir, T.W. (2011) Split Inteins: Nature's Protein Ligases. *Isr J Chem*, **51**, 854-861.

Singaravelan, B., Roshini, B.R. & Munavar, M.H. (2010) Evidence that the supE44 mutation of Escherichia coli is an amber suppressor allele of glnX and that it also suppresses ochre and opal nonsense mutations. *J Bacteriol*, **192**, 6039-6044.

Soares, J.A., Zhang, L., Pitsch, R.L., Kleinholz, N.M., Jones, R.B., Wolff, J.J., Amster, J., Green-Church, K.B. & Krzycki, J.A. (2005) The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases. *J Biol Chem*, **280**, 36962-36969.

Sodha, S.V., Lynch, M., Wannemuehler, K., Leeper, M., Malavet, M., Schaffzin, J., Chen, T., Langer, A., Glenshaw, M., Hoefer, D., Dumas, N., Lind, L., Iwamoto, M., Ayers, T., Nguyen, T., Biggerstaff, M., Olson, C., Sheth, A. & Braden, C.

(2011) Multistate outbreak of Escherichia coli O157:H7 infections associated with a national fast-food chain, 2006: a study incorporating epidemiological and food source traceback results. *Epidemiol Infect*, **139**, 309-316.

Srinivasan, G., James, C.M. & Krzycki, J.A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science*, **296**, 1459-1462.

Stoddard, C.D., Gilbert, S.D. & Batey, R.T. (2008) Ligand-dependent folding of the three-way junction in the purine riboswitch. *Rna*, **14**, 675-684.

Strieker, M. & Marahiel, M.A. (2009) The structural diversity of acidic lipopeptide antibiotics. *Chembiochem*, **10**, 607-616.

Sudarsan, N., Barrick, J.E. & Breaker, R.R. (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *Rna*, **9**, 644-647.

Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G.M. & Breaker, R.R. (2005) Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem Biol*, **12**, 1325-1335.

Tadesse, D.A., Zhao, S., Tong, E., Ayers, S., Singh, A., Bartholomew, M.J. & McDermott, P.F. (2012) Antimicrobial drug resistance in Escherichia coli from humans and food animals, United States, 1950-2002. *Emerg Infect Dis*, **18**, 741-749.

Takimoto, J.K., Dellas, N., Noel, J.P. & Wang, L. (2011) Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids. *ACS Chem Biol*, **6**, 733-743.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B. & Zhao, Y. (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, **146**, 1016-1028.

Tanaka, T., Yamamoto, S., Oi, S. & Yamamoto, T. (1981) Structures of heterooligosaccharides synthesized by levansucrase. *J Biochem*, **90**, 521-526.

Tavassoli, A. & Benkovic, S.J. (2007) Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli. *Nat Protoc*, **2**, 1126-1133.

Tornoe, C.W., Christensen, C. & Meldal, M. (2002) Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem*, **67**, 3057-3064.

Tropberger, P. & Schneider, R. (2013) Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nat Struct Mol Biol*, **20**, 657-661.

van Apeldoorn, M.E., van Egmond, H.P., Speijers, G.J. & Bakker, G.J. (2007) Toxins of cyanobacteria. *Mol Nutr Food Res*, **51**, 7-60.

von Baum, H. & Marre, R. (2005) Antimicrobial resistance of Escherichia coli and therapeutic implications. *Int J Med Microbiol*, **295**, 503-511.

Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*, **8**, 127-134.

Wan, W., Huang, Y., Wang, Z., Russell, W.K., Pai, P.J., Russell, D.H. & Liu, W.R. (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli. *Angew Chem Int Ed Engl*, **49**, 3211-3214.

Wang, Y.S., Fang, X., Wallace, A.L., Wu, B. & Liu, W.R. (2012) A rationally designed pyrrolysyl-tRNA synthetase mutant with a broad substrate spectrum. *J Am Chem Soc*, **134**, 2950-2953.

Wang, Y.S., Russell, W.K., Wang, Z., Wan, W., Dodd, L.E., Pai, P.J., Russell, D.H. & Liu, W.R. (2011) The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives. *Mol Biosyst*, **7**, 714-717.

Wang, Z. & Rizzo, C.J. (2001) Synthesis of the C8-deoxyguanosine adduct of the food mutagen IQ. *Org Lett*, **3**, 565-568.

Wilson, K.S. & von Hippel, P.H. (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci U S A*, **92**, 8793-8797.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., McNicholas, S.J., Murshudov,

G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A. & Wilson, K.S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, **67**, 235-242.

Winn, M.D., Murshudov, G.N. & Papiz, M.Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol*, **374**, 300-321.

Wu, H., Hu, Z. & Liu, X.Q. (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc Natl Acad Sci U S A*, **95**, 9226-9231.

Xie, Z., Dai, J., Dai, L., Tan, M., Cheng, Z., Wu, Y., Boeke, J.D. & Zhao, Y. (2012) Lysine succinylation and lysine malonylation in histones. *Mol Cell Proteomics*, **11**, 100-107.

Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. & Yokoyama, S. (2008a) Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J Mol Biol*, **378**, 634-652.

Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K. & Yokoyama, S. (2008b) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol*, **15**, 1187-1197.

Yanagisawa, T., Ishii, R., Fukunaga, R., Nureki, O. & Yokoyama, S. (2006) Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon Methanosarcina mazei. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, **62**, 1031-1033.

Yanagisawa, T., Sumida, T., Ishii, R. & Yokoyama, S. (2013) A novel crystal form of pyrrolysyl-tRNA synthetase reveals the pre- and post-aminoacyl-tRNA synthesis conformational states of the adenylate and aminoacyl moieties and an asparagine residue in the catalytic site. *Acta Crystallogr D Biol Crystallogr*, **69**, 5-15.

Young, T.S., Young, D.D., Ahmad, I., Louis, J.M., Benkovic, S.J. & Schultz, P.G. (2011) Evolution of cyclic peptide protease inhibitors. *Proc Natl Acad Sci U S A*, **108**, 11052-11056.

Zhang, X. & Bremer, H. (1995) Control of the Escherichia coli rrnB P1 promoter strength by ppGpp. *J Biol Chem*, **270**, 11181-11189.

Zhang, Y., Wang, L., Schultz, P.G. & Wilson, I.A. (2005) Crystal structures of apo wild-type M. jannaschii tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein Sci*, **14**, 1340-1349.

Zhang, Z., Tan, M., Xie, Z., Dai, L., Chen, Y. & Zhao, Y. (2011) Identification of lysine succinylation as a new post-translational modification. *Nat Chem Biol*, **7**, 58-63.

Zheng, D., Constantinidou, C., Hobman, J.L. & Minchin, S.D. (2004) Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Res*, **32**, 5874-5893.

Zimmer, S., Hoffmann, E., Jung, G. & Kessler, H. (1993) "Head-to-Tail" Cyclization of Hexapeptides Using Different Coupling Reagents. *Liebigs Ann. Chem.*, **1993**, 497–501.

VII. Appendix

VII.1. List of Primers

Table 21. List of primers that have been used during this thesis.

Name of the primer	Nucleotide sequence (5' → 3')	Tm [°C]
CP library construct		
oIntC-EcoRI_for	GTCTGAATTCATGGTTAAAGTTATCGGTCGTCGTTCCC	66
oIntC-SalI_rev	GGATTAGCGAGTCGACATTGTGCGCAATCGCCCCATTG	70
oIntN_for	GCTACGGTCTAAGCTTGCCTCAGTTTTGGC	64
oIntN-Hisrev	TAGCGACTCGAGTTAGTGGTGGTGGTGGTGGTGTTTAATAGTCCCAGC GTC	73
oC+5_lib	GGAATTCGCCAATGGGGCGATCGCCCACAATTGCNNSNNSNNSNNSN NSTGCCTCAGTTTTGGC	75
oZipper	GGAATTCGCCAATGGGGCGATCGCC	64
oIntC-NcoI_for	GTCTCCATGGTTAAAGTTATCGGTCGTCGTTC	63
Reporter constructs		
oPurineRS_for	GCGAGGATCCGTTAAATAATTTACATAAAC	56
oPurineRS_rev	CTACGGTCTAAGCTTTCTTAAAACCTCCTCAG	62
oB12RS_for	GGATTAGCGAGGATCCGCCGGTCCTGTGAGTTAATAGG	70
oB12RS_rev	GCTACGGTCTAAGCTTTGTAAAGCATCCACAATAGAAGAAGG	66
oB12RS2_for	GGATTAGCGAGGATCCGAATTCCTATTTGTGAGCTACGTC	68
oB12RS2_rev	GCGGAAGACGGCGCAGC	57

oB12RS3_rev	GCCCGGAAGACGGCGCAGC	62
oLacZ5_for	GGATTAGCGAAAGCTTATGATTACGGATTCACTGGCCG	67
oLacZ2_rev	GCTACGGTCTGCGGCCGCTTATTTTTGACACCAGACCAACTGG	72
oSacB_for	GCGAGCGGCCGCATGAACATCAAAAAGTTTGC	66
oSacB_rev	GCTACGGTCTCTTAAGTTATTTGTTAACTGTTAATTG	60
oSacB-A13357T_for	CAATCAACGTTTGCGCCTAGCTTCCTGCTGAACAT	66
oSacB-A13357T_rev	ATGTTCAGCAGGAAGCTAGGCGCAAACGTTGATTG	66
oSacB_XhoI-SmaI_for	GATTAGCGACTCGAGCCCGGGATGAACATCAAAAAGTTTGC	69
oSacBPacI_rev	GCTACGGTCTTTAATTAATTATTTGTTAACTGTTAATTG	58
oGlnBsrI_for	GATTAGCGATGTACACGGCCGCATAATCTCATCAATCATCCCCATAATCCTTG	71
oGlnXhoI_rev	CTACGGTCTCTCGAGGGCTTCCTCAAAGCGTAAACAAC	69
Genome Integration		
oPADlanding_nth-tpp_for	CTGCTTTCCGCTCAGGCGACCGATGTCAGTGTTAATAAGGCGACGGCGAATACGGCCCCAAGGTCCAAACGGTGA	79
oPADlanding_nth-tpp_rev	CGGAAAATGTGCGTGTCGACAGCAATAGTCGGCCAGCCGAATGCAGTGTTTTGGCTTCAGGGATGAGGCGCCATC	79
oNth/tppBver_for	ACCACCGAGCTTAATTTCAGTTCGC	58
oNth/tppBver_rev	CCTGTTCGACGTTTTTCCCCGGCGC	64
P_{glnS} – P_{tac} Exchange		
oP*tac*_SLIM_FT	TTGACAATTAATCATCGGCTCGTATAATGAAGCCCTCGAGCCCGATC	70
oP*tac*_SLIM_RT	CATTATACGAGCCGATGATTAATTGTCAACGATTACTTTCTGTTCGACTTAAGG	68
oP*tac*_SLIM_FS	AAGCCCTCGAGCCCGATC	55
oP*tac*_SLIM_RS	CGATTACTTTCTGTTCGACTTAAGG	54
oP*tac*-lacO_FS	GAATTCCAATAATTTTGTTTAACTTTAAG	50
oP*tac*-lacO_FT	TGTGGAATTGTGAGCGGATAACAATTGAATTCCAATAATTTTGTTTAACTTTAAG	65
oP*tac*-lacO_RS	CATTATACGAGCCGATGATTAATTG	53
oP*tac*-lacO_RT	AATTGTTATCCGCTCACAATTCCACACATTATACGAGCCGATGATTAATTG	67

Tm is the basic melting temperature of a primer and is calculated for primer < 14 nt:

$$Tm = 4\ °C \times (number\ of\ G\ +\ C\ in\ the\ primer) + 2\ °C \times (number\ of\ A\ +\ T\ in\ the\ primer)$$

For primer > 14 nt, the following equation is used, in which N is the length of the primer:

$$Tm = 64.9\ °C + 41\ °C \times \frac{(number\ of\ G\ +\ C\ in\ the\ primer - 16.4)}{N}$$

VII.2. DNA Sequences

Figure 34. DNA sequence of the LP-*tetA* fragment synthesized by GeneArt AG (Thermo Fisher Scientific Inc., Regensburg, DE). The ORF of the tetracycline resistance is highlighted in black, landing pads in red and I-SceI restriction sites in green.

```
1    GAATTCTACGGCCCCAAGGTCCAAACGGTGATAGGGATAACAGGGTAATATTTACGTTGA   60
61   CACCACCTTTCGCGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAAT  120
121  ATGAATAGTTCGACAAAGATCGCATTGGTAATTACGTTACTCGATGCCATGGGGATTGGC  180
181  CTTATCATGCCAGTCTTGCCAACGTTATTACGTGAATTTATTGCTTCGGAAGATATCGCT  240
241  AACCACTTTGGCGTATTGCTTGCACTTTATGCGTTAATGCAGGTTATCTTTGCTCCTTGG  300
301  CTTGGAAAAATGTCTGACCGATTTGGTCGGCGCCCAGTGCTGTTGTTGTCATTAATAGGC  360
361  GCATCGCTGGATTACTTATTGCTGGCTTTTTCAAGTGCGCTTTGGATGCTGTATTTAGGC  420
421  CGTTTGCTTTCAGGGATCACAGGAGCTACTGGGGCTGTCGCGGCATCGGTCATTGCCGAT  480
481  ACCACCTCAGCTTCTCAACGCGTGAAGTGGTTCGGTTGGTTAGGGGCAAGTTTTGGGCTT  540
541  GGTTTAATAGCGGGGCCTATTATTGGTGGTTTTGCAGGAGAGATTTCACCGCATAGTCCC  600
601  TTTTTTATCGCTGCGTTGCTAAATATTGTCACTTTCCTTGTGGTTATGTTTTGGTTCCGT  660
661  GAAACCAAAAATACACGTGATAATACAGATACCGAAGTAGGGGTTGAGACGCAATCAAAT  720
721  TCGGTGTACATCACTTTATTTAAAACGATGCCCATTTTGTTGATTATTTATTTTCAGCG  780
781  CAATTGATAGGCCAAATTCCCGCAACGGTGTGGGTGCTATTTACCGAAAATCGTTTTGGA  840
841  TGGAATAGCATGATGGTTGGCTTTTCATTAGCGGGTCTTGGTCTTTTACACTCAGTATTC  900
901  CAAGCCTTTGTGGCAGGAAGAATAGCCACTAAATGGGGCGAAAAAACGGCAGTACTGCTC  960
961  GAATTTATTGCAGATAGTAGTGCATTTGCCTTTTTAGCGTTTATATCTGAAGGTTGGTTA 1020
1021 GATTTCCCTGTTTTAATTTTATTGGCTGGTGGTGGGATCGCTTTACCTGCATTACAGGGA 1080
1081 GTGATGTCTATCCAAACAAAGAGTCATGAGCAAGGTGCTTTACAGGGATTATTGGTGAGC 1140
1141 CTTACCAATGCAACCGGTGTTATTGGCCCATTACTGTTTACTGTTATTTATAATCATTCA 1200
1201 CTACCAATTTGGGATGGCTGGATTTGGATTATTGGTTTAGCGTTTTACTGTATTATTATC 1260
1261 CTGCTATCAATGACCTTCATGTTGACCCCTCAAGCTCAGGGGAGTAAACAGGAGACAAGT 1320
1321 GCTTAGTAGGGATAACAGGGTAATGATGGCGCCTCATCCCTGAAGCCAAAAGCTT       1375
```

VII.3. Plasmid Maps

Figure 35. Plasmid map of pBAD202/d/LacZ

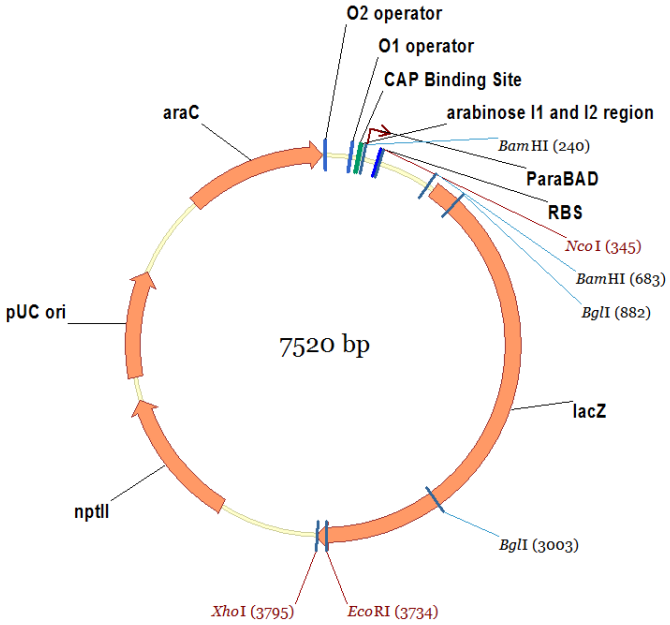


Figure 36. Plasmid map of the CP library construct carrying cyclo-CIHGQS.

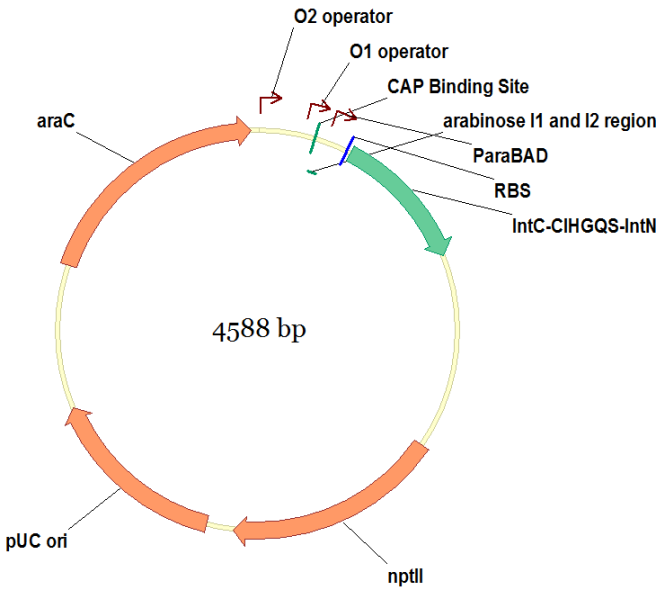


Figure 37. Plasmid map of pACYCDuet-1.

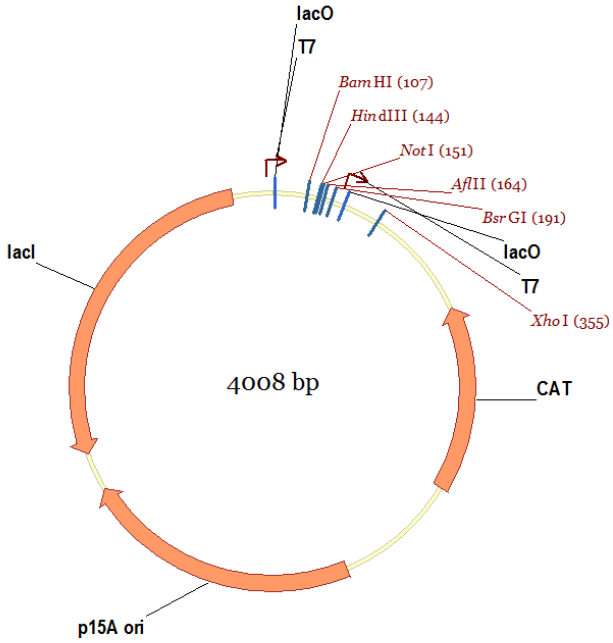


Figure 38. Plasmid map of pACYCDuet-1-P$_{glnS}$-*pylS*.

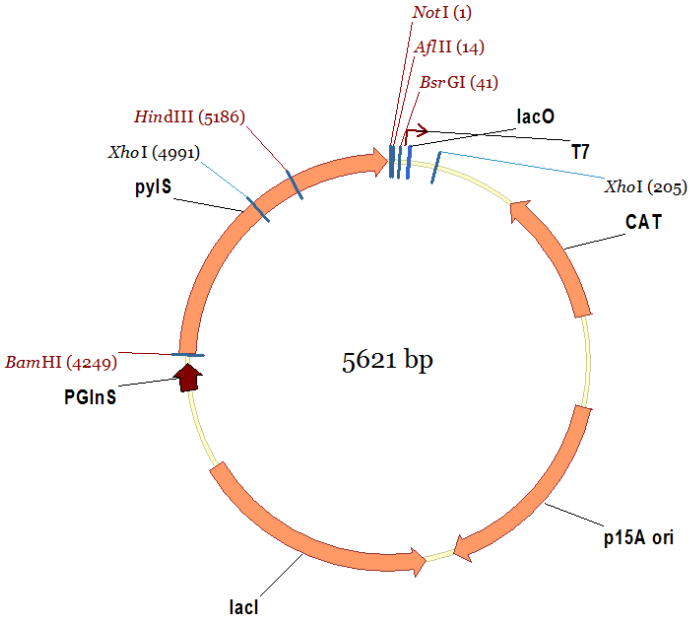


Figure 39. Plasmid map of the *S. aureus xpt* guanine RS fusion reporter construct.

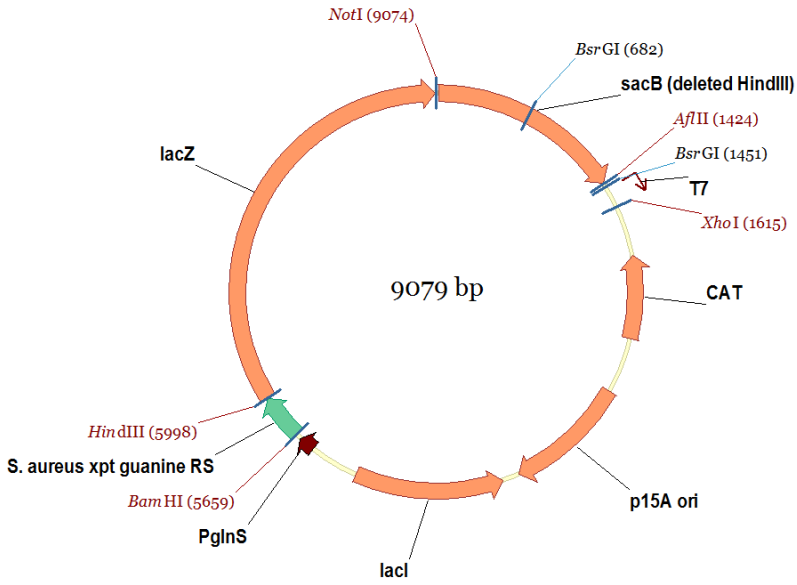


Figure 40. Plasmid map of the -240 nt to 0 nt *E. coli btuB* AdoCbl RS fusion reporter construct.

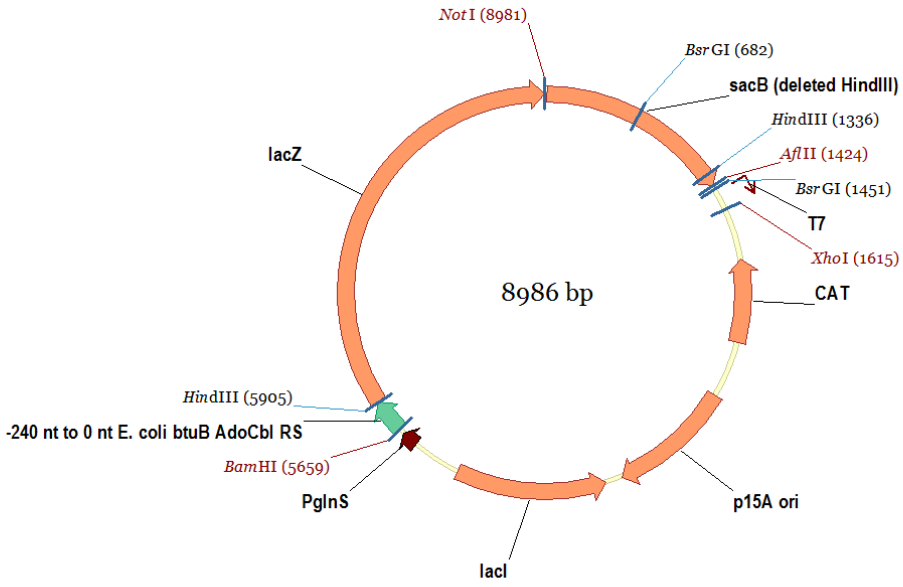


Figure 41. Plasmid map of the -300 nt to +210 nt *E. coli btuB* AdoCbl RS fusion reporter construct.

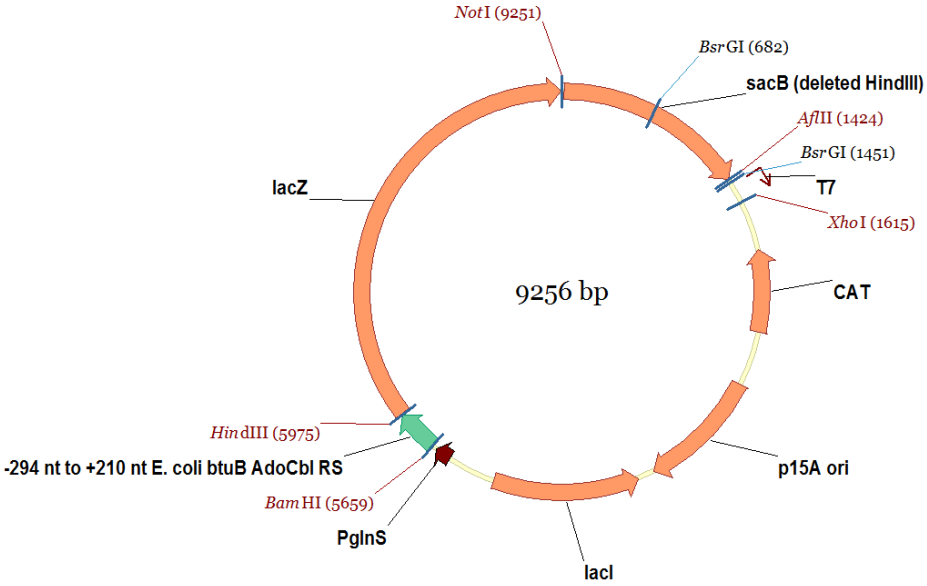


Figure 42. Plasmid map of the -300 nt to +210 nt *E. coli btuB* AdoCbl RS split reporter construct.

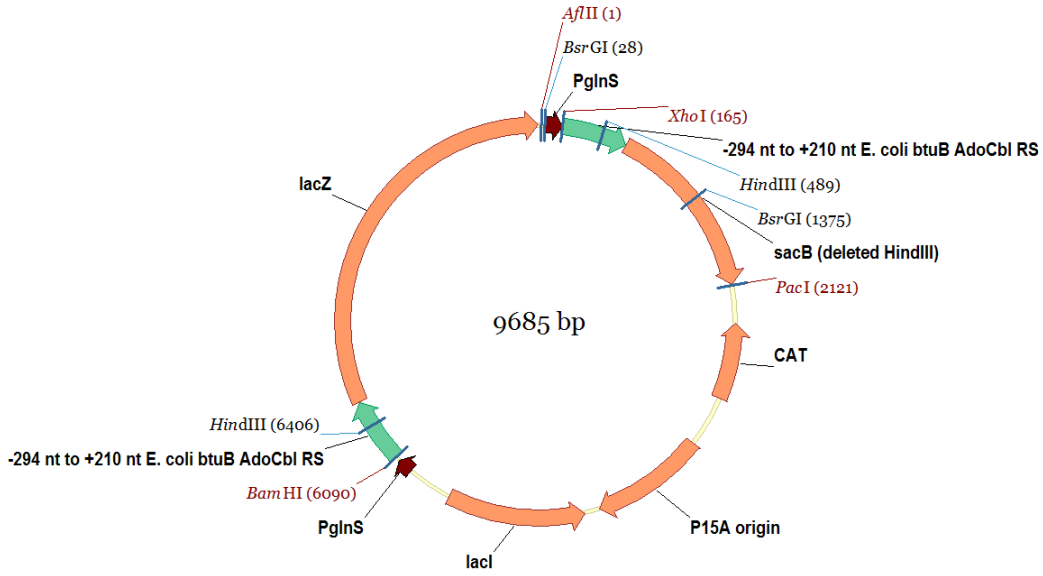


Figure 43. Plasmid map of the donor plasmid pTKIP-reporter (P_{glnS} promoter)

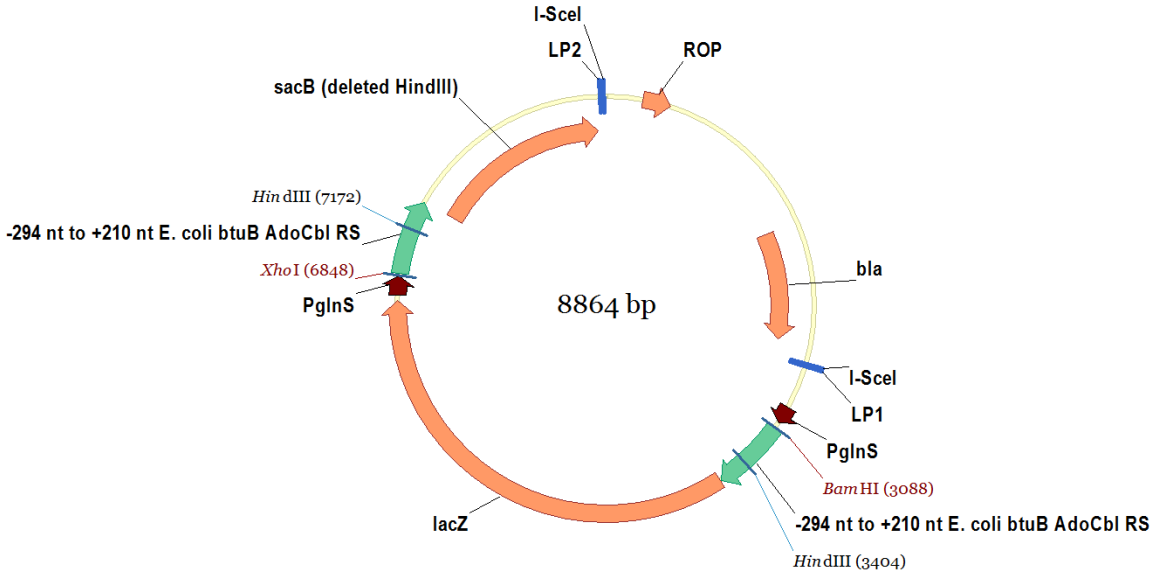
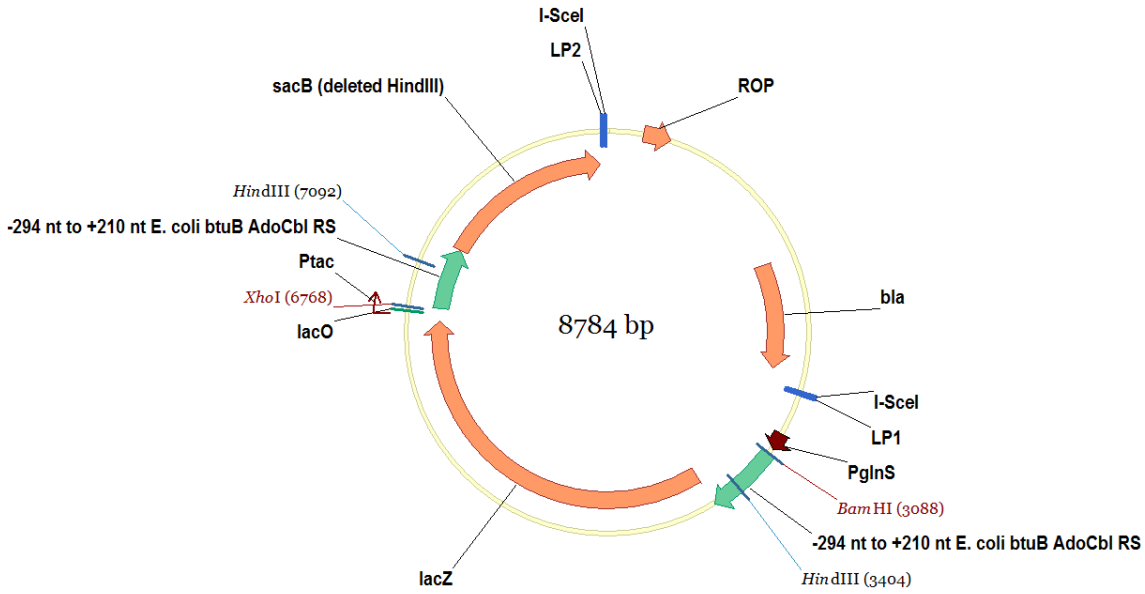


Figure 44. Plasmid map of the donor plasmid pTKIP-reporter (P_{tac} promoter).



Plasmid maps of pTKRED and pTKIP have been published elsewhere (Kuhlman & Cox, 2010).

VII.4. Crystallographic Data

Table 22. Data collection and refinement statistics.

	Kalk (PDB code 4CH6)	Kbu (PDB code 4CH3)	Kcr (PDB code 4CH4)	Kpr (PDB code 4CH5)
Wavelength (Å)	1.0	1.0	1.0	1.0
Resolution range (Å)	38.29-2.05 (2.12-2.05)	38.44-2.28 (2.36-2.28)	42.42-2.16 (2.24-2.16)	42.3-2.2 (2.28-2.20)
Space group	$P6_4$	$P6_4$	$P6_4$	$P6_4$
Unit cell	104.9 104.9 71.0	105.3 105.3 71.4	105.0 105.0 71.8	105.4 105.4 70.8
Total reflections	271274 (18606)	123272 (12794)	247942 (23845)	228986 (21051)
Unique reflections	28018 (2789)	20634 (2045)	24308 (2439)	22651 (2185)
Multiplicity	9.7 (6.7)	6.0 (6.3)	10.2 (9.8)	10.1 (9.6)
Completeness (%)	99.98 (99.82)	99.94 (100.00)	100.00 (100.00)	99.68 (96.90)
Mean I/sigma(I)	45.70 (3.29)	24.69 (3.86)	27.19 (4.53)	23.80 (2.96)
Wilson B-factor	43.9	44.3	36.4	43.0
R-merge	0.0272 (0.593)	0.0440 (0.450)	0.0611 (0.534)	0.0622 (0.699)
R-meas	0.0288	0.0483	0.0643	0.0656
CC1/2	1 (0.877)	1 (0.889)	0.999 (0.923)	0.999 (0.837)
CC*	1 (0.967)	1 (0.97)	1 (0.98)	1 (0.955)
R-work	0.174 (0.271)	0.176 (0.231)	0.166 (0.330)	0.174 (0.225)
R-free	0.186 (0.296)	0.197 (0.245)	0.209 (0.349)	0.190 (0.257)
Number of atoms	2296	2252	2310	2276
macromolecules	2158	2151	2120	2157
ligands	21	21	21	30
water	117	80	169	89
Protein residues	264	268	259	264
RMS(bonds)	0.011	0.010	0.010	0.010
RMS(angles)	1.51	1.38	1.40	1.44
Ramachandran favored (%)	97	96	97	98
Ramachandran outliers (%)	0	0	0	0.38
Clashscore	0.23	1.17	0.71	1.62
Average B-factor	53.1	49.6	40.1	47.8
macromolecules	52.9	49.4	39.3	47.6
ligands	58.5	60.3	50.2	58.2
solvent	56.0	50.1	49.6	50.1

Statistics for the highest-resolution shell are shown in parentheses.

VIII. Declaration

I, Veronika Martina Flügel, hereby declare that I independently prepared the present thesis, using only the references and resources stated. This work has not been submitted to any examination board yet. Parts of this work have been or will be published in scientific journals.

Garching, August 2014