DEPARTMENT CHEMIE

LEHRSTUHL FÜR BIOCHEMIE

Development of an *in vivo* selection system for the identification of antibacterial cyclic peptides

Sabrina Theresa Harteis

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. Lukas Hintermann
Prüfer der Dissertation:	1.	TUM Junior Fellow Dr. Sabine Schneider
	2.	Univ.-Prof. Dr. A. Stephan Sieber

Die Dissertation wurde am 11.09.2015 an der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 15.10.2015 angenommen.

FÜR MEINE FAMILIE

Acknowledgements

This thesis was realized at the Chair of Biochemistry at the Technische Universität München from november 2011 until september 2015.

I would like to thank Dr. Sabine Schneider for providing me the opportunity to carry out my PhD in her group, her interest in my work and her steady support in this challenging project.

Special Thanks go to

…Prof. Dr. Stephan A. Sieber and his group members for the productive cooperations. Many thanks go to Max Koch for the pleasant teamwork and all the discussions. Furthermore, I would like to thank 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 the unlimited permission to use all of his facilities.

…my colleagues for the good working atmosphere and especially Dr. Veronika Flügel, Christine Oellig, Marion Kirchner and Dr. Andrea Kunfermann for their encouraging words and their advice at all the time.

…the technicians Richard Feicht, Christine Schwarz, Ursula Seidel and the secretary Ute Kashoa and especially Astrid König for the friendly help, support and daily work.

…my students Florian, Anja, Lily, Susi, Patrizia and Matthias.

…Dr. Annika Frank, Dr. Andrea Kunfermann and Dr. Ernie Star for proofreading.

I also want to thank Dr. David Schlereth, Sabine Rittinger, Markus Meier, Sebastian Foraita, and Alexander Henze for our weekly Friday lunch and for sharing good and bad moments.

A special thank goes to my parents, brother and grandparents for all the support and encouragement not only during my PhD but also in all life situations. Last but not least, my major gratuity goes to my husband Uli for the encouraging words and love he is giving me every day.

Zusammenfassung

Riboschalter sind strukturierte RNA Elemente im 5' untranslatierten Bereich der mRNA, die die Expression essentieller Gene regulieren. Da sie hauptsächlich in Bakterien vorkommen, sind sie ein interessantes Angriffsziel für die Entwicklung neuer Medikamente. Mögliche Inhibitoren für Riboschalter sind zyklische Peptide, die sich besonders durch ihre hohe biochemische Stabilität auszeichnen und die aufgrund ihrer starren Konformation gute Bindeaffinitäten aufweisen. Darüber hinaus sind bereits viele bioaktive zyklische Peptide bekannt, die antibakterielle, immunosuppressive oder anti-tumor aktive Eigenschaften besitzen. In den letzten Jahren wurden Methoden entwickelt, um zyklische Peptide genetisch zu kodieren und zu exprimieren.

Ziel dieser Arbeit war die Etablierung eines Selektionssystems, das die Identifizierung von zyklischen Peptiden erlaubt, die selektiv an den Coenzym B_{12} Riboschalter aus *Salmonella typhimurium* binden. Dazu wurden verschiedene Konstrukte kloniert, bei denen der Riboschalter die Expression von Reportergenen kontrolliert. Die Qualität der Regulation wurde mit Hilfe des natürlichen Liganden Coenzym B_{12} geprüft. Es zeigte sich eine erfolgreiche Regulierung der Reportergene *lacZ* und *sacB* mit dem Riboschalter, wodurch ein dreistufiges *in vivo* Selektionsverfahren mit Hilfe einer zyklischen Peptidbibliothek etabliert werden konnte. Darüber hinaus wurde die Effektivität, mit der zyklische Peptide *in vivo* generiert werden, untersucht. Die DnaE Split-Inteine aus *Synechocystis sp. Strain PCC6803 und Nostoc punctiforme* wurden für die *in vivo* Zyklisierung der Peptide herangezogen. Mittels Zellextraktionen und synthetisierten Referenzpeptiden konnte somit erfolgreich die Menge von vier *in vivo* generierten zyklischen Peptiden massenspetrometrisch quantifiziert werden. Die Ergebnisse dieser Arbeit dienen als Grundlage weiterer Forschung zur Untersuchung des Coenzym B_{12} Riboschalters aus *Salmonella typhimurium* und der Identifizierung von biologisch aktiven zyklischen Peptiden *in vivo*.

Abstract

Riboswitches are structured RNA elements in the 5‘untranslated region of mRNA that regulate expression of essential genes. As they predominantly occur in bacteria, riboswitches constitute an interesting target for drug development. A possible compound class to target riboswitches are cyclic peptides. They possess high stability and a conformational rigidity which allows a high binding affinity towards target molecules. There are also many bioactive cyclic peptides with antibacterial, immunosuppressive or anti-tumor activity. In the last years methods have been developed to genetically encode and express cyclic peptides.

The aim of this work was to establish an *in vivo* selection system that allows the identification of cyclic peptides, which selectively bind coenzyme B_{12} riboswitch from *Salmonella typhimurium.* For this purpose, several constructs were cloned in which the expression of reporter genes was controlled by the riboswitch. The quality of regulation was examined by addition of the natural ligand coenzyme B_{12}. A successful regulation of the reporter genes *lacZ* and *sacB* was observed and a three-step *in vivo* screening system using a cyclic peptide library was developed. Furthermore, to optimize the cyclic peptide library, sequence dependency of cyclic peptide generation was investigated. For this purpose, *in vivo* cyclization efficacy of four different cyclic peptide sequences was investigated by using DnaE split inteins from *Synechocystis sp. Strain PCC6803* and *Nostoc punctiforme* for cyclization. Cell extraction and chemically synthesized reference peptides led to the successful quantification of *in vivo* generated cyclic peptides by mass spectrometry.

Results of this work serve as a basis for further research on the coenzyme B_{12} riboswitch from *Salmonella typhimurium* and the identification of biologically active cyclic peptides *in vivo*.

TABLE OF CONTENTS

1. INTRODUCTION

Scientists have always been interested in the question of how genetic information is transmitted. In the 1940s the majority of scientists thought that proteins were the hereditary material. Although Oswald Avery indicated in 1944 that DNA might carry genetic information, his conclusions were widely doubted. [1, 2] It took a further 16 years until Francis Crick published the "central dogma of molecular biology". [3] In this he postulated the information flow going from DNA to proteins within biological systems. Cricks law, as simple and basic as it seems in our days, was a mile stone in molecular biology. Since then, a surprisingly high amount of gene control mechanisms have been identified illustrating the real complexity of gene expression. Regulation for instance, is present at all stages of the gene expression pathway including chromatin modifications, transcription (initiation and termination), posttranscriptional modifications (capping, polyadenylation, splicing), RNA transport to the cytoplasm, stability of mRNA in the cytoplasm and translation. Especially structured RNA elements were found to be involved in many of these regulatory processes and are elaborated on more closely in the following sections.

1.1 RIBOSWITCHES

Riboswitches (RSs) are a widely spread form of gene regulation in bacteria but they have also been identified in archaea and eukaryotes. [4, 5] They are metabolite-sensing RNA elements, predominantly embedded in the 5' untranslated region (UTR) of mRNA. With only few exceptions RSs act in *cis*. [6, 7] They are able to bind small molecules whereupon a conformational change in RS secondary structure affects downstream gene expression. Hence, every RS can be divided into two domains: the ligand binding part or aptamer domain and the downstream expression platform that has an impact on gene expression due to a conformational change (**Figure 1**).

RSs were found to regulate enzymes involved in the biosynthetic pathway of their corresponding RS-binding metabolites and thereby play a role as feedback regulators. A cobalamin RS, for instance, is mainly responsible for the regulation of coenzyme B_{12} (adenosyl cobalamin, AdoCbl) transport in cells. Above a certain level of cobalamin abundancy, AdoCbl binds to the RS and switches gene expression of an AdoCbl transporter to "off". [8, 9]

To date several classes of metabolites have been identified to bind RSs. They range from small ions such as Mg^{2+} to large cofactors like AdoCbl (**Table 1**). The number of metabolites found to trigger RSs is expected to grow in the future.

Table 1 *RS binding metabolites*

Amino acids	**Anions**	**Cofactors and derivates**	**Purines and derivates**	**Metals**
Gln [10]	F^- [11]	AdoCbl [8]	adenine [12]	Mg^{2+} [13, 14]
Gly [15]		FMN [16]	ATP [17]	
Lys [18]		Moco/Tuco [19]	c-di-GMP [20]	
		SAH [21]	dG [22]	
		SAM [23]	guanine [24]	
		TPP [25]	$preQ_1$ [26, 27]	

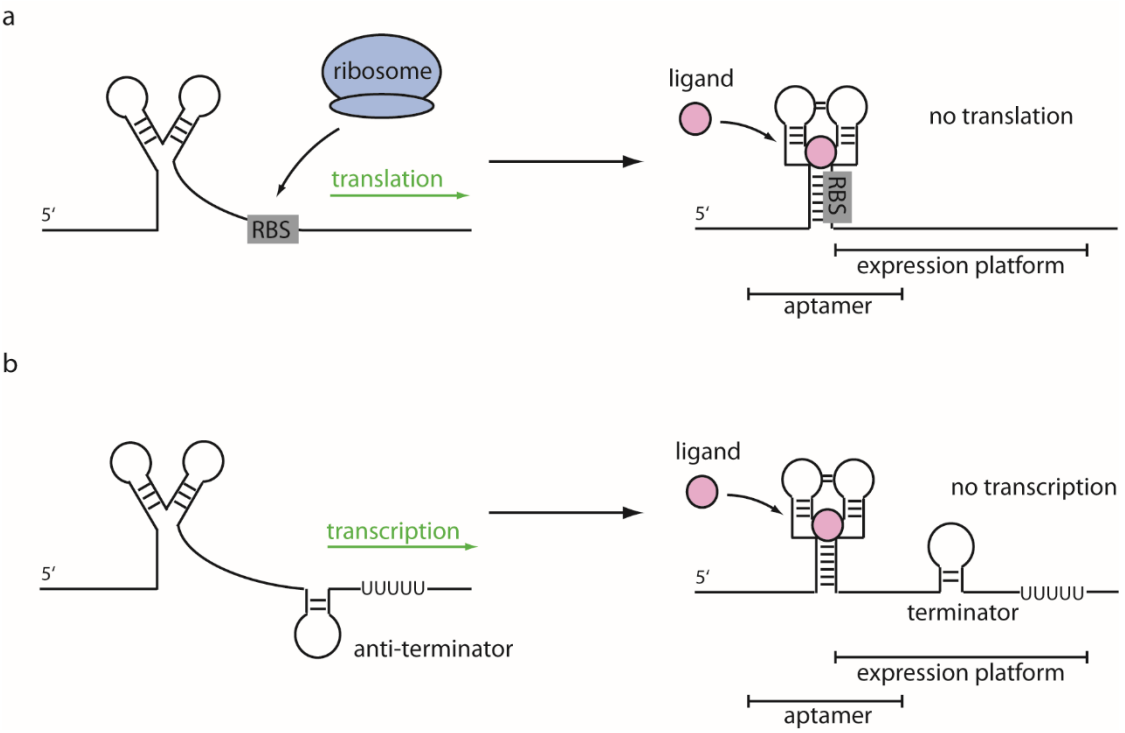
1.1.1 Regulation Mechanisms

RSs are in charge of a large repertoire of regulation mechanisms such as control of translation, transcription, mRNA degradation and splicing. Many RSs regulate gene expression on a translational level by blocking/liberating the ribosome binding site (RBS). **Figure 1a** illustrates this kind of control mechanism such as is found in the previously mentioned AdoCbl RSs. In the ligand-free form the RBS is accessible to the ribosome. With the increase of the ligand concentration above a certain threshold, ligand binding stabilizes a conformation of the RS in which the RBS is sequestered and gene expression is switched off. [28]

A similar and abundant mechanism is found at the regulation on a transcriptional level by forming an anti-terminator or terminator loop. One example is the SAM (*S*-adenosyl methionine) RS which is involved in gene regulation of methionine and cysteine metabolism. If a cognate metabolite binds, a downstream hairpin is formed which serves as a Rho-independent terminator loop. In the absence of the metabolite, folding of a stabilized aptamer-binding domain is disrupted and, as a result of alternative base pairings, an anti-terminator loop forms (**Figure 1b**). [23, 29, 30]

Recently, a Rho-dependent transcription termination by RSs was also found in the Mg^{2+}-sensing mgtA RS from *Salmonella enterica serovar Typhimurium* and the flavin mononucleotide-sensing ribB RS from *Escherichia coli*. [31] It is likely that this newly found mechanism of regulation is widely spread as several RSs lack terminator and RBS-sequestering sequences. [5]

Figure 1 *Schematic representation of two RS control mechanisms. Aptamer and expression platforms are highlighted.* (**a**) *Translation inhibition prompted by ligand binding by sequestration of RBS.* (**b**) *Transcription is terminated after ligand induced formation of a terminator loop.*



A rather exclusive regulation mechanism was found for the *glmS* riboswitch-ribozyme. It is embedded in the 5' UTR of glucosamine-6-phosphate synthase and is the first example of a metabolite responsive ribozyme. The g*lmS*-riboswitch-ribozyme is acting as a negative feedback regulation system, in which glucosamine-6-phosphate (GlcN6P) serves as a cofactor for RNA self-cleavage through internal transesterification. As a result, mRNA is degraded by RNase J1 and the synthesis of GlcN6P is stopped. [32-34] Another mechanism in which RSs induce mRNA degradation was found for a lysine-binding RS regulating *E. coli lysC*. Hence, ligand binding induces a conformational change which not only inhibits translation but also gives access to RNase C cleaving sites. [35]

Eukaryotic RSs were found to be located between introns of pre-mRNA and regulate gene expression by alternative splicing in plants. A thiamine pyrophosphate (TPP) RS for instance controls the accessibility of a 5' splicing site to determine whether an intron in thiamine C synthase (THIC) pre-mRNA is removed or not. Intron removal due to TPP-bound RS decreases mRNA stability and thereby lowers TPP synthesis. [36, 37]
These examples highlight the high diversity in regulation of gene expression by RSs and open a broad field of research possibilities.

1.1.2 Classification and Structure of Riboswitches

The increasing number of identified RSs makes a classification system necessary. For one, RSs can be classified according to their metabolite (**Table 1**). It was found that aptamer domains are evolutionarily more conserved and, within a RS group, several structural motives like the B12 element (B12 RS) or S-box (SAM RS) reoccur. [38] These elements were successfully identified in comparative genome wide studies aiming to find potential regulatory RNA elements. [4, 39-41]

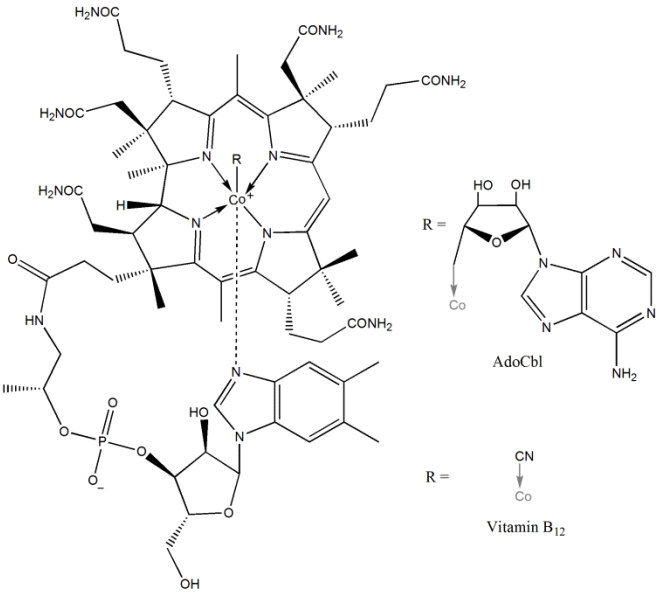
On the other hand, only little structural similarity of RSs can be found. Beside adenine, guanine and deoxyguanosine (dG), the structural diversity of RSs is large and their architecture is not related to the respective ligand structure. Different classes of SAM binding RSs for example adopt dissimilar architectures despite they

are all recognizing the same ligand. [42-44] Therefore, RSs are divided according to their structure into two large groups known as junctional and pseudoknotted RSs. Pseudoknotted RSs are built up by two helices connected by short loops forming a knot-shaped conformation. Base pairings occur between an outside region and a stem loop structure. The metabolite binding region is usually located at the junction of a pseudoknot or in a groove of a helix with nucleotides of the loop trapping it. [42, 45] Junctional RSs display a junction in a central position including the metabolite-sensing domain, which is, for example, found in purine RSs. Alternatively, the junction is located in the periphery and folds backward and stabilizes the RS regulatory domain as observed in a Mg^{2+} RS. Metabolite binding occurs in the junctional and/or close to the regulatory region. [14, 46]

1.1.3 Cobalamin Riboswitch

Cobalamin RSs are some of the most abundant RSs in biology and responsible for the regulation of genes involved in vitamin B_{12} biosynthesis and transport in bacteria. [8, 38, 47-49] These RSs are large in size as they contain aptamer domains of approximately 200 nucleotides, with AdoCbl (**Figure 2**) as their natural ligand. [28] Derived from Vitamin B_{12}, AdoCbl serves as a reversible generator of free radicals through hemolysis in several reactions. These radicals can remove non-acid hydrogen atoms of substrates to facilitate the break of carbon-oxygen, carbon-nitrogen and carbon-carbon bonds. For this purpose, AdoCbl participates, for example, in the glutamate degradation pathway. [50-52] An important component of Vitamin B_{12} transport in bacteria is the cobalamin outer membrane transporter BtuB whose gene expression is controlled by a RS. The BtuB RS was shown to recognize AdoCbl with a high selectivity compared to other cobalamin derivates. [8] Studies on the *btuB* leader region revealed that the initial part of the *btuB* gene itself, namely approximately 100 bases downstream of the start codon, are necessary for efficient regulation of transcription and translation. [53, 54]

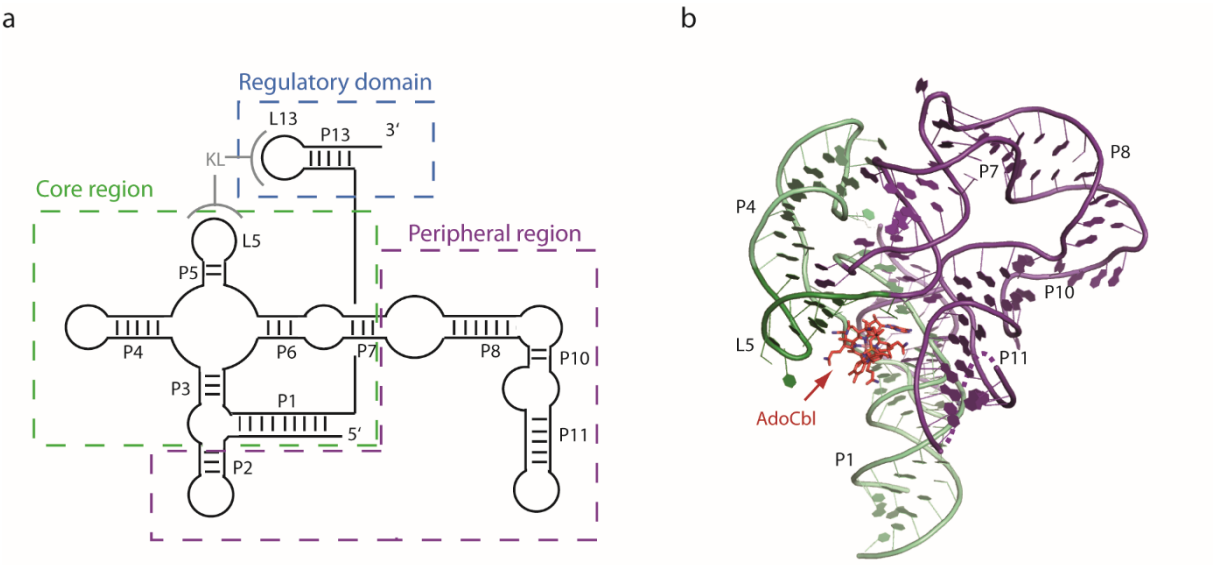
Figure 2 *Chemical composition of AdoCbl. A central hexa-coordinated cobalt atom forms a complex with four nitrogens of a corrin ring. The remaining two sites are coordinated through a dimethylbenzimidazole nitrogen atom and a 5'-deoxyadenosyl group. The latter coordination site is also the center of reactivity. In vitamin B_{12}, this site is substituted by a cyano group.*



The main element of the secondary structure of cobalamin RSs is a central four way junction, which forms the core receptor domain (paired regions P3 - P6) and is responsible for metabolite binding, while the kissing loop interaction between loop 5 (L5) of the core region and L13 regulates gene expression. RSs controlling translation typically contain the RBS in L13 (**Figure 3**). The peripheral domain which is responsible for the recognition of different cobalamin derivates is not conserved. Within the same class, P6 extensions are consistent. Cobalamin RSs are further classified according to the occurrence of peripheral extensions between P1 and P3. Ligand recognition is achieved through hydrogen bonds and van der Waals interactions between the corrin ring and pocket forming nucleotides. To date crystal

structures of only a few cobalamin RSs have been solved: *thermoanaerobacter tengcongensis* AdoCbl RS, *env8* variant of a Pacific Ocean metagenome RS and AdoCbl RS from *Symbiobacterium thermophilum,* all bound to their corresponding ligands. [55, 56] In all of these structures, the cobalamin is sandwiched between the minor grooves of P3 - P6, P5 and P11 (**Figure 3b**). [55, 56]

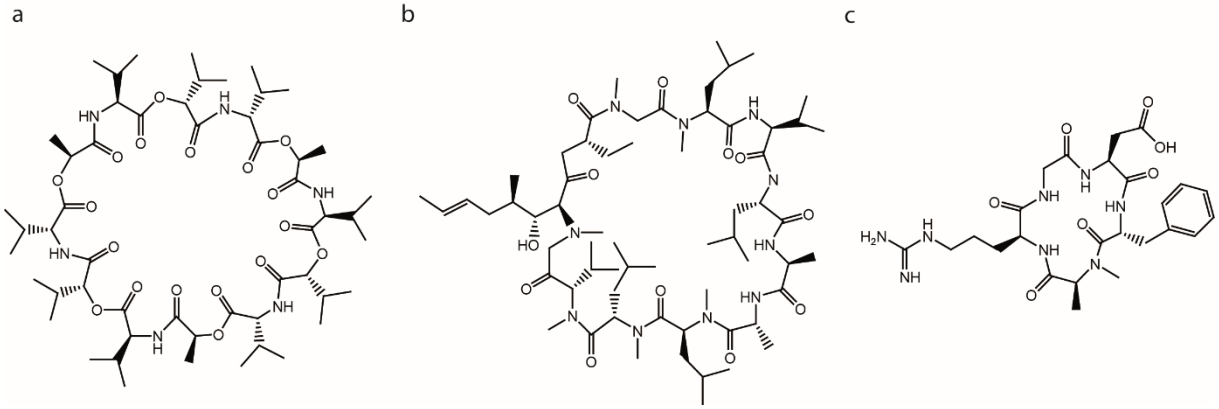
Figure 3 *Structure of cobalamin RSs* (**a**) *Secondary structure of an AdoCbl RS. The conserved core receptor domain (green) is responsible for metabolite binding. The peripheral region (purple) is highly dependent on the RS class and can be absent or truncated. Kissing loops (KLs) are found between L5 and L13 of the regulatory domain (blue).* (**b**) *Ribbon presentation of an overall crystal structure of the S. Thermophilum AdoCbl RS (PDB: 4GXY). The core region is highlighted in green and peripheral extensions in purple. AdoCbl (red) binds the RS in a pocket between P11, P5 and P3 – P6.*



1.2 Cyclic Peptides

For a long time peptides have been regarded as less promising compounds for drug design as they provide little biochemical stability and present difficulties in analogous synthesis compared to other small molecules. However, in recent years research in this area resumed and peptides have become more and more popular. Cyclic peptides (CPs) can be especially potent bioactive compounds and are increasingly viewed as ideal scaffolds. Advantages of using CPs include their lower toxicity, particularly when compared to small synthetic molecules. Since they are composed of amino acids, CP degradation products can be easily metabolized after acting on target molecules, without causing harmful side effects. [57] Furthermore, cyclization overcomes certain disadvantages that are associated with peptide drugs. Lacking free C- and N-termini, biochemical stability is significantly improved as CPs are more resistant to proteolytic degradation. In addition, their conformational rigidity provides CPs with better biological activity because the conformational entropy loss is reduced upon binding. [58, 59] Further stability is observed with respect to high temperatures or extreme pHs, which can be an advantage for application. [60] Some CPs have also been reported to cross the cell membrane. CPs offer a large spectrum of biological activity and have been identified as antibiotics [61], immunosuppressive [62] and anti-tumor agents [63-67] (**Figure 4**). As a result of these features, the interest in drugs based on CPs is increasing and more CPs are expected to emerge as therapeutic agent in the futures. CP libraries will be an especially useful tool in identifying biologically active members of this group.

Figure 4 *Examples of biologically active CPs.* (**a**) *Antibiotic Valinomycin,* (**b**) *immunosuppresive Cyclosporine A and* (**c**) *anti-tumor agent Cilengitide.*



There are several approaches to generate CP libraries including biological, semi-synthetic and synthetic methods. Biological generation of CPs is, for example, achieved by utilizing split inteins that produce CPs during a splicing reaction. Semisynthetic methods combine ribosomal synthesis and an associated genetic tag with the incorporation of unnatural amino acids. A prominent example of a semi-synthetic library is the mRNA display. [68, 69] Completely synthetic CP libraries certainly offer a higher sequence diversity but synthesis is more challenging, time-consuming and the hit identification is more difficult compared to, for example, biological libraries. [70, 71]

1.2.1 Peptide Synthesis

In the 20th century Emil Fischer was the first scientist to synthesize the first dipeptide called glycylglycine. [72] It took a further 50 years until Vigneaud developed a method to produce a polypeptide. His synthesis in homogenous solution was based on the principle that the carboxylic function and the amino group were reversibly blocked. Formation of the peptide bond was achieved after activating the free carboxy group of the N-terminal amino acid. A problematic aspect of this approach was that synthesis was time consuming. All intermediates had to be isolated and purified before the next step could be performed. Furthermore larger peptides tended to be rigid and insoluble and complicated the synthesis. [73]

Bruce Merrifield revolutionized peptide synthesis by performing the reactions under heterogeneous conditions, in which the peptide was C-terminally linked to a polymeric solid resin. He was rewarded the Nobel Prize in 1984 for inventing solid phase peptide synthesis (SPPS). [74] The starting point for SPPS is a solid support which is coupled by a short linker to the C-terminus of an amino acid. All reactive side chain moieties and the N^{α}-amino groups of all amino acids used for SPPS are protected. The first step in SPPS is deprotection of the N^{α}-amino group of the resin bound amino acid. Next, the carboxy group of the amino acid that is going to be coupled is activated with a special auxiliary. Final coupling is performed by formation of a peptide bond between the deprotected N^{α}-amino group and the activated carboxy group. The coupling steps - N^{α}-deprotection, activation and coupling - are repeated until the desired length of peptide is synthesized. Finally, the peptide is removed from the resin and the side chains are deprotected. Most importantly, all reaction steps can be carried out in one reaction vessel and byproducts, as well as a surplus of reactants, can be removed by washing steps.

Merrifield established SPPS by using Boc (*tert*-butyloxycarbonyl) as a N^{α}-amino group and benzyl or related protecting groups for the side chains of trifunctional amino acids. This strategy was exclusively used during the first years and the Boc/benzyl protecting strategy turned out to be useful for difficult synthesis. However, deprotection of the side chains requires HF as a strong acid. In addition to its high toxicity and harsh conditions, the use of HF also has the disadvantage that benzyl removal always leads to Boc deprotection. The introduction of the Fmoc (9-fluorenylmethyloxycarbonyl) strategy made peptide synthesis more convenient and protection group cleavage could now be performed under milder conditions. Furthermore, the Fmoc strategy is orthogonal and allows higher flexibility for complex synthesis strategies. [75]

The buildup of Merrifield's synthesis allowed automation of the process and facilitated peptide synthesis significantly. [76] Microwave-assisted SPPS has been found to further enhance reaction rates and allows the synthesis of rigid and challenging peptide sequences. [77] Today, SPPS is one of the major techniques for therapeutic peptide production and offers, besides the synthesis of natural peptides, the possibility also to incorporate unnatural amino acids or D-amino acids. As such it

is an essential tool and irreplaceable technique in modern laboratories working on peptides.

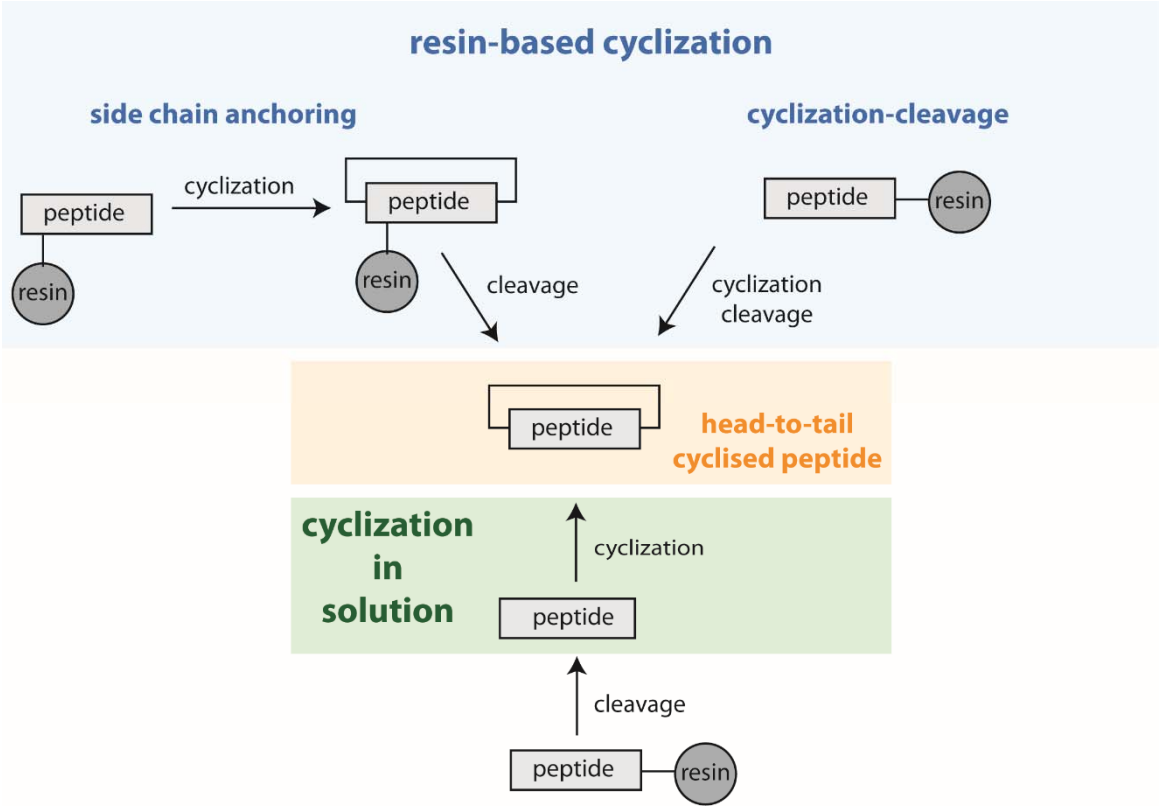
The chemical synthesis of CPs gives access to a high diversity of products. There are three different means of chemical cyclization, namely head-to-side chain cyclization, side chain-to-side chain cyclization and head-to-tail cyclization. The head-to-tail cyclization has especially been investigated intensively. An overview of head-to-tail cyclization methods is shown in **Figure 5**.

Head-to-tail cyclizations can be achieved by the classical reaction of a linear peptide under highly diluted conditions in solution. Here, the intramolecular cyclization is induced by activating the carboxy group of the C-terminus. [78]

Head-to-tail cyclizations can also be performed using resin-based methods. One approach for that is the cyclization of a peptide which is anchored *via* a side chain functional group. [79, 80] The C-terminus of the peptide is orthogonally protected as an ester. After initial deprotection of the N-terminus, the peptide is assembled to the desired length using Fmoc or Boc strategies. In the next step, N- and C-termini are deprotected and activation of the C-terminus induces intramolecular head-to-tail cyclization. Finally, the CP is deprotected and cleaved from the resin. [81]

An alternative way of performing a head-to-tail cyclization on solid support is the cyclization-cleavage approach in which the C-terminus is attached to an electrophilic linker that is labile to free amines. Many strategies for the cyclization-cleavage approach require resins like the Kaiser oxime resin and therefore use Boc chemistry. [82]

Figure 5 *Head-to-tail cyclization methods for peptides.*

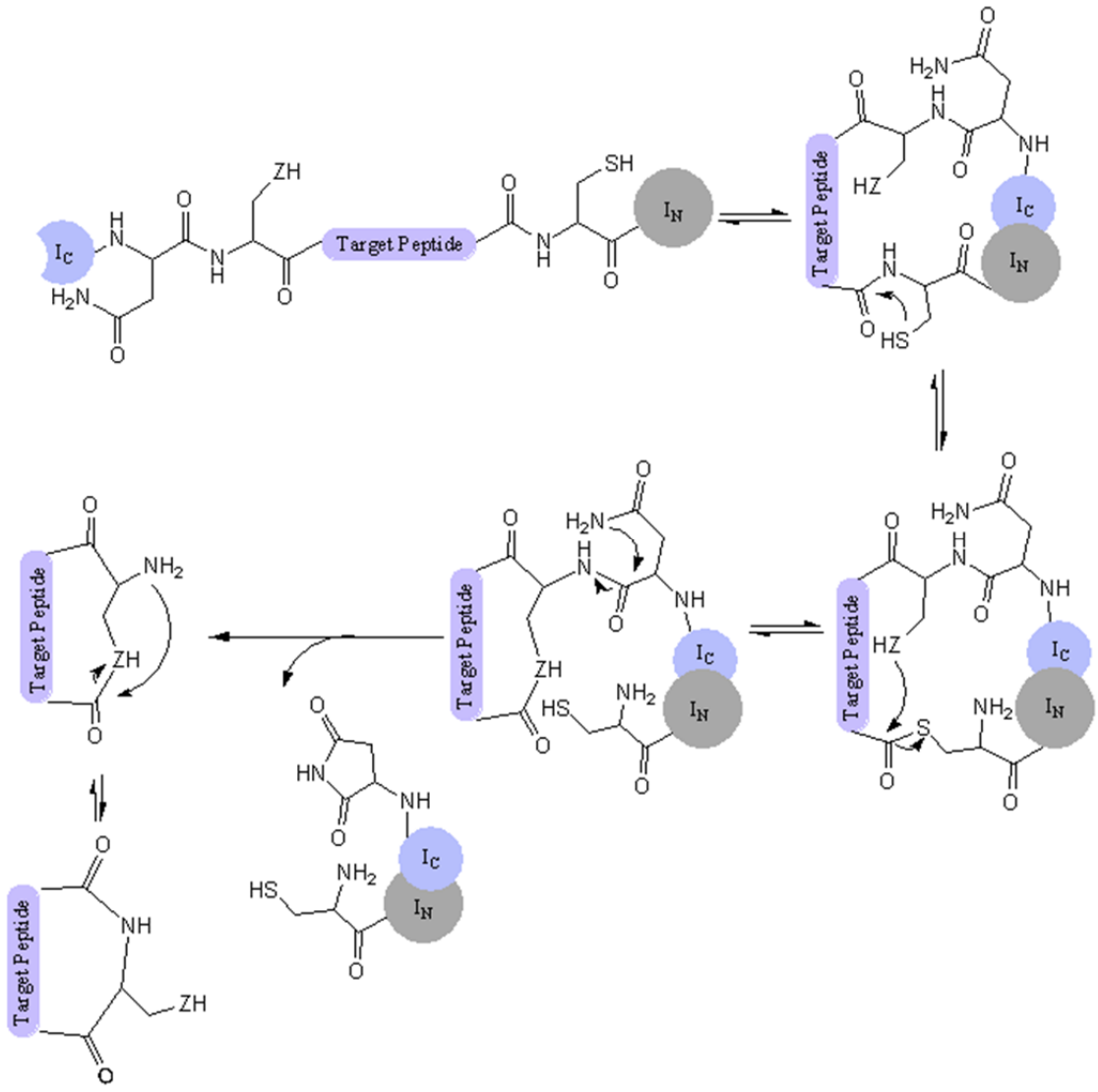


1.2.2 SICLOPPS

CPs can also be produced biologically by a mechanism called split intein mediated circular ligation of peptides and proteins (SICLOPPS). [83] SICLOPPS utilizes split inteins, which are arranged in the following order: I_C - target peptide - I_N. The splicing mechanism produces a head-to-tail cyclization of the target sequence *in vivo* (**Figure 6**). [84-87] In general any target sequence and length can be inserted. The only requirement for SICLOPPS is a nucleophilic side chain (Ser or Cys) in the first position of the target sequence. With SICLOPPS, CP libraries can also be designed by using degenerate oligo nucleotides and polymerase chain reaction (PCR). The oligo nucleotides thereby bear variable segments in the form of NNS, where N can be any DNA base and S represents a C or G. Therefore UAA and UGA stop codons are eliminated from the library and the NNS sequence covers all 20 amino acids. By randomizing amino acids in positions 2 – 5 a hexameric cyclic peptide library of 34 million members can be created on DNA level. Conveniently, this lies within the number of transformants that can be readily achieved in *E. coli*. Another advantage

is that SICLOPPS can be coupled to *in vivo* selection systems in which bioactive peptides can be easily identified by sequencing the plasmids. [84, 88] Therefore SICLOPPS is an elegant method to create and screen CP libraries *in vivo*. SICLOPPS-generated peptide libraries are becoming more and more popular and have so far revealed several active compounds including a number of protease inhibitors [89-91], methyltransferase- [92] and transcription factor inhibitors. [93]

Figure 6 *Mechanism of SICLOPPS. Rearrangement of the intein domains triggers head-to-tail cyclization of the target sequence. First, association of the split inteins results in formation of the active protein. An N-S/N-O acyl shift in the next step leads to a thioester/ester intermediate. Transesterification of the nucleophilic side chain (Z) forms a lariat intermediate. Cyclization of the Asn side chain liberates the cyclic product as a lactone. Finally a Z-to-N acyl shift generates the lactam product in vivo.*



1.3 PROTEIN SPLICING

The SICLOPPS mechanism relies on protein splicing, a post-translational process involved in the regulation of gene expression. It was reported for the first time in 1990 when Stevens and coworkers identified a 69 kDa catalytic subunit of the vacuolar protein-translocating adenosine triphosphatase (H(+)-ATPase) and a 50 kDa protein having their origin in a single translation product. [94] Since that time, protein splicing has been found to occur in all three kingdoms of life. [95]

Such protein splicing reactions are carried out by inteins, whose name is a combination of the words *int*ernal and prot*ein*. These proteins bear the ability to excise themselves from a protein precursor by linkage of the two flanking polypeptide sequences, called exteins by a peptide bond (**Figure 7a**).

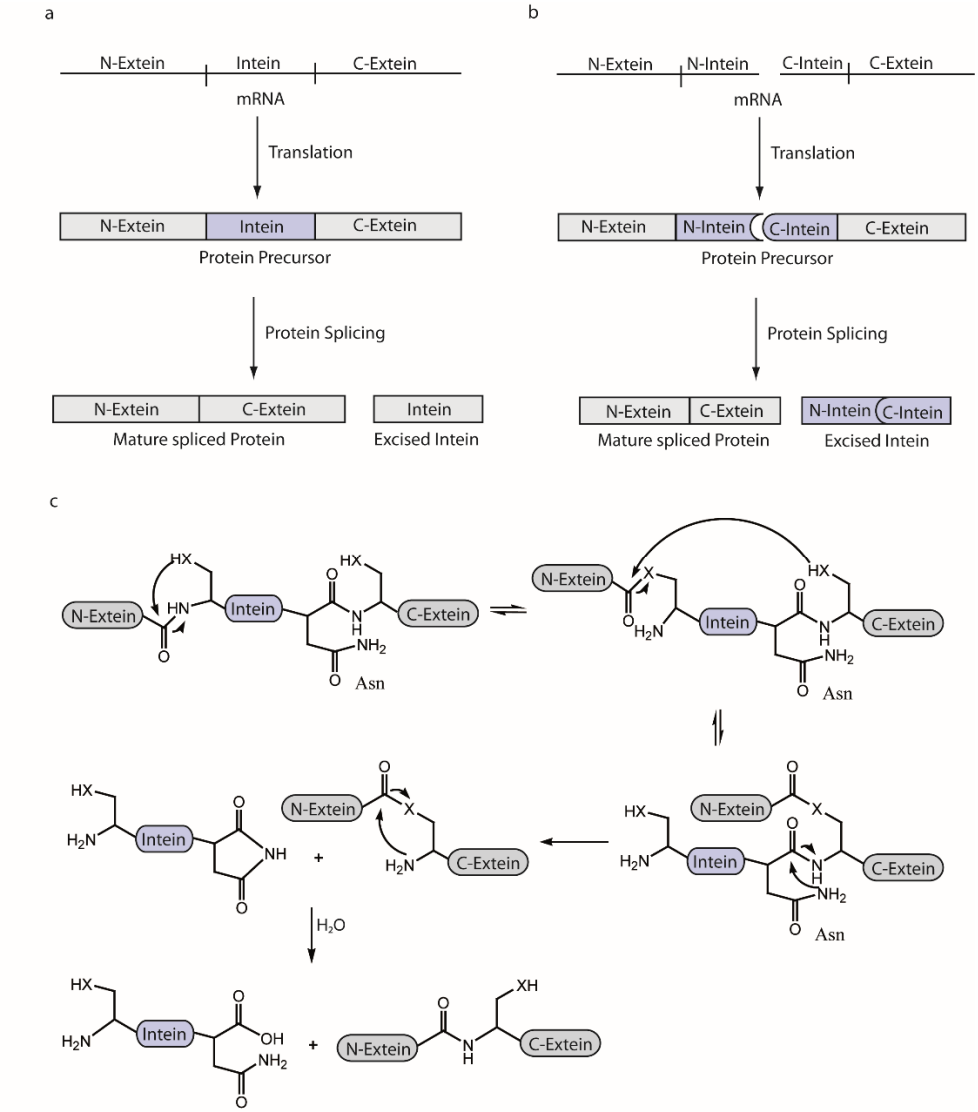
The splicing reaction mechanism can be divided into four steps. First a Cys or Ser undergoes an N-X acyl shift (X = S or O respectively) and converts the peptide bond at the N-terminal splice junction to a thioester/ester. Next, a branched intermediate is formed by the attack of a sulfhydryl/hydroxyl group of the first residue in the C-extein. At the C terminus of the intein a conserved Asn is cyclized and the intein is released. Finally the succinimide is hydrolyzed and a peptide bond is formed by an S-N/O-N acyl shift between the exteins (**Figure 7c**).

Since discovery of protein splicing, several research groups took advantage of its mechanism and inteins quickly became a useful tool in protein chemistry. [96] Artificial splitting of intact inteins soon opened the field of trans-splicing reactions. Therefore it was, for example, possible to perform segmental isotopic labeling studies with NMR spectroscopy. [97]

Proof of natural occurring split inteins was provided by Liu and coworkers when examining the catalytic subunit of DNA polymerase III (DnaE) from the cyanobacterium *Synechocystis sp. Strain PCC6803* (*Ssp*). They discovered that the 123 amino acid long N-terminal domain (I_N) and the C-terminal sequence (I_C) consisting of 36 amino acids are transcribed on two intermitted genes on opposite DNA strands. The splicing mechanism itself is analogous to the mechanism

described above, after I_N and I_C have rearranged to form the active protein (**Figure 7b**). [98]

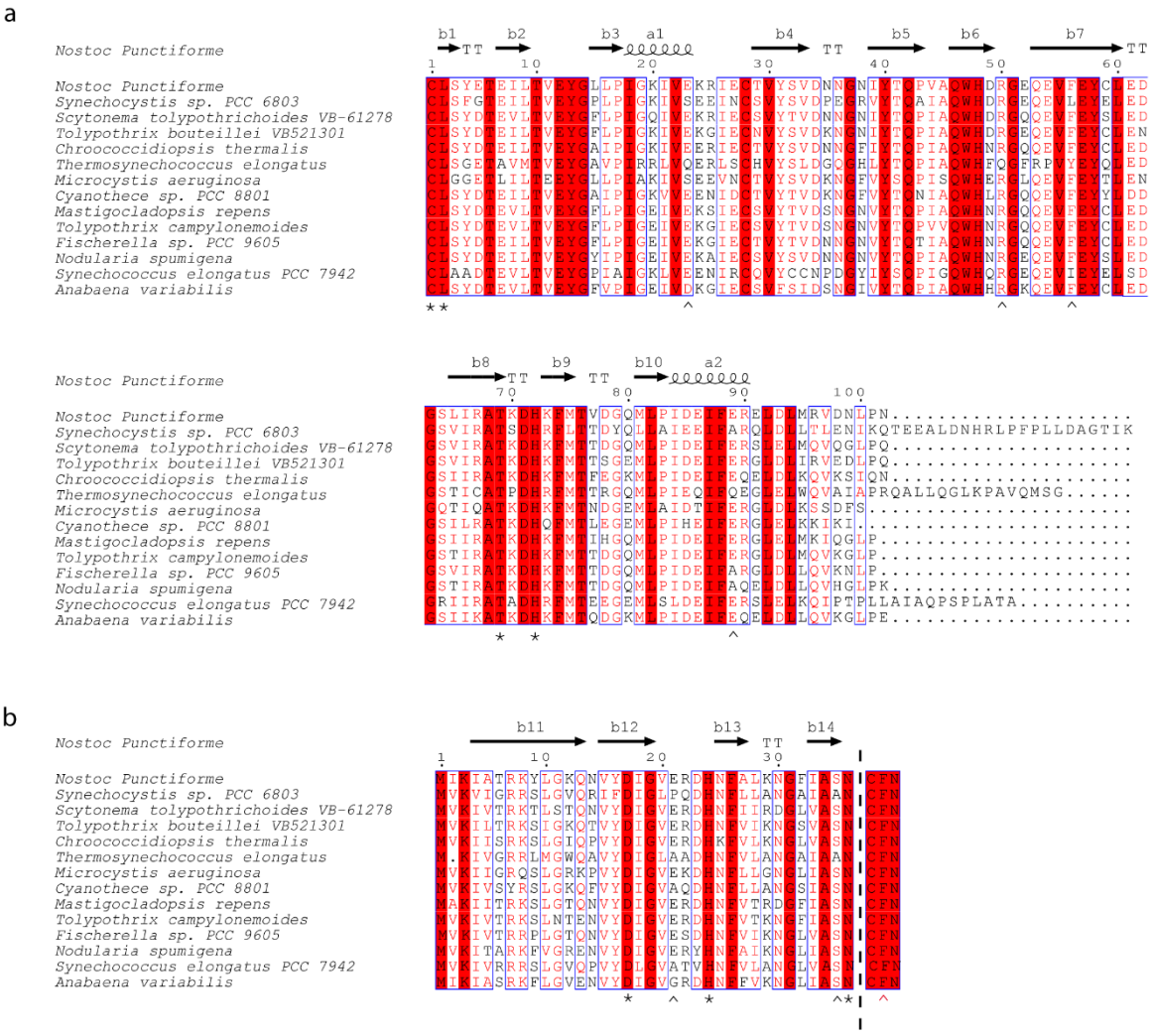
Figure 7 *Protein splicing* (**a**) *Cis-splicing. Translation of mRNA leads to a protein precursor consisting of an N-terminal extein, an intein and a C-terminal extein. Due to protein splicing the intein is excised and the mature protein is formed.* (**b**) *Scheme of trans-splicing. The protein is encoded on two separate genes and intein fragments associate post-translationally, forming an active intein which results in a splicing reaction.* (**c**) *Four step protein splicing pathway. First, an N-S/N-O acyl shift forms a thioester/ester. Next, the thioester/ester bond is cleaved and the branched intermediate is cyclized in step 3. Finally, the aminosuccimide is hydrolyzed and an amide bond is formed between the exteins. X stands for O or S (Ser or Cys, respectively).*



Following the first discovery of a natural occurring split intein, other split inteins have been identified and investigated. For instance the split intein DnaE from the cyanobacterium *Nostoc punctiforme* (*Npu*) turned out to be highly efficient in *trans*-splicing. It was reported that *Npu* possesses a 33 to 170-fold higher trans-splicing rate compared to *Ssp*. Regarding splicing product yields, *Npu* inteins show highest reaction rates at 37°C, whereas *Ssp* DnaE exhibits reduced yields and more side products at temperatures above 30°C. [99]

The DnaE inteins from *Npu* and *Ssp* share a sequence identity of 67% for the I_N and 53% for the I_C, respectively. [100] A C-terminal extension of *Ssp* I_N is missing in *Npu* I_N but it can be removed without loss of activity. [101] Alignments revealed several conserved domains among split inteins. It is noteworthy that the key catalytic residues ^{N}Cys1, ^{N}Leu2, ^{N}Thr69, ^{N}His72, ^{C}Asp17, ^{C}His24 and ^{C}Asn36 are conserved across the family (**Figure 8**). [101-103] ^{C}Asp17 for example facilitates the N-S acyl shift and transesterification. [104-106] It forms hydrogen bonds with ^{N}Cys1 or the C-extein Cys+1 in several steps. [107]

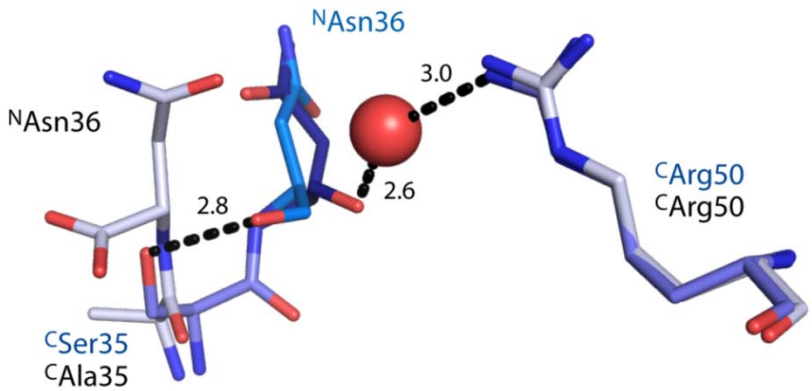
Figure 8 *Sequence alignment of 14 split inteins of DNA Polymerase Subunit III alpha generated with ESPript using the secondary structure of Npu DnaE (PDB: 2KEQ). [108] Catalytic amino acids (*) and non-catalytic residues important for splicing efficacy (^) are highlighted.* (**a**) *Sequence alignment of* I_N. (**b**) *Sequence alignment of* I_C *including the three amino acids (CFN) of the C-terminal extein sequences. The amino acid in extein +2 position (^) is important for splicing efficacy.*



Non-catalytic amino acids are only moderately conserved and therefore seem to modulate splicing efficacy. Interestingly, some of these non-catalytic residues that promote splicing efficacy lie outside of the active side and seem to have stabilizing effects when present in the structure. ^{N}Glu23 for example has no obvious structural role but may be crucial in folding stability. [109] ^{N}Glu89 has been shown to be part of an ion cluster and is therefore important for stabilizing the split intein complex. [110] An aromatic residue at position 56 of I_N, for instance, may stabilize the TXXH

motif (positions 69-72 I_N) by facilitating packing interactions. [109] Structural analysis of the DnaE split intein from *Npu* revealed that ^{C}Asn36 may exhibit two different side chain orientations. In one conformation ^{C}Asn36 forms a hydrogen bond over an ordered water molecule with ^{N}Arg50 and in the other conformation it forms a hydrogen bond with the hydroxyl group of ^{C}Ser35 (**Figure 9**). These two amino acids seem to facilitate cyclization of ^{C}Asn36 in *Npu* DnaE. [103] In the less active intein *Ssp* DnaE, ^{C}Ser35 is substituted with an Ala and ^{N}Arg73 instead might instead promote Asn cyclization. [101] These different stabilization mechanisms during splicing may be an explanation why some inteins such as *Npu* DnaE appear superior over others like *Ssp* DnaE.

Figure 9 *Superimposition of Npu DnaE (blue, PDB 4LX3) and excised Ssp DnaE (white, PDB 1ZD7). The two orientations of Npu DnaE ^{N}Asn36 are coloured in dark blue and cyan.*



Furthermore, the splicing efficacy is influenced by the amino acids close to the splicing region, which is of particular interest for the application of inteins. Amino acids close to the splicing region, such as extein amino acids, were found to have a significant influence on the splicing reaction. A Phe, located at the +2 C-extein position (**Figure 8b**), highly influences the splicing activity of DnaE intein. In many cases a substitution of this Phe abolished the *trans*-splicing reaction completely. [100] It was shown that different inteins possess divergent tolerance upon permutation of the +2 C-extein amino acid. The DnaE split intein from *Npu* was

found to be slightly more tolerant towards substitution of this critical positon compared to *Ssp* DnaE. Mutational analyses of the +2 C-extein position in various inteins to Glu, Gly and Arg by Muir and coworkers suggest negatively charged amino acids to be more easily accommodated in this position, as Glu gave the best splicing yields. Furthermore, bulky amino acids were indicated to be necessary for branched intermediate resolution and efficient splicing, as C**G**N mutants were found to splice more slowly. [109]

The advantage of the split inteins is that reactions can be carried out under native conditions and the problem of refolding is bypassed, which persists when artificial trans-splicing inteins are used. In the last decade split intein based methods like SICLOPPS evolved and have given rise to new opportunities for drug discovery.

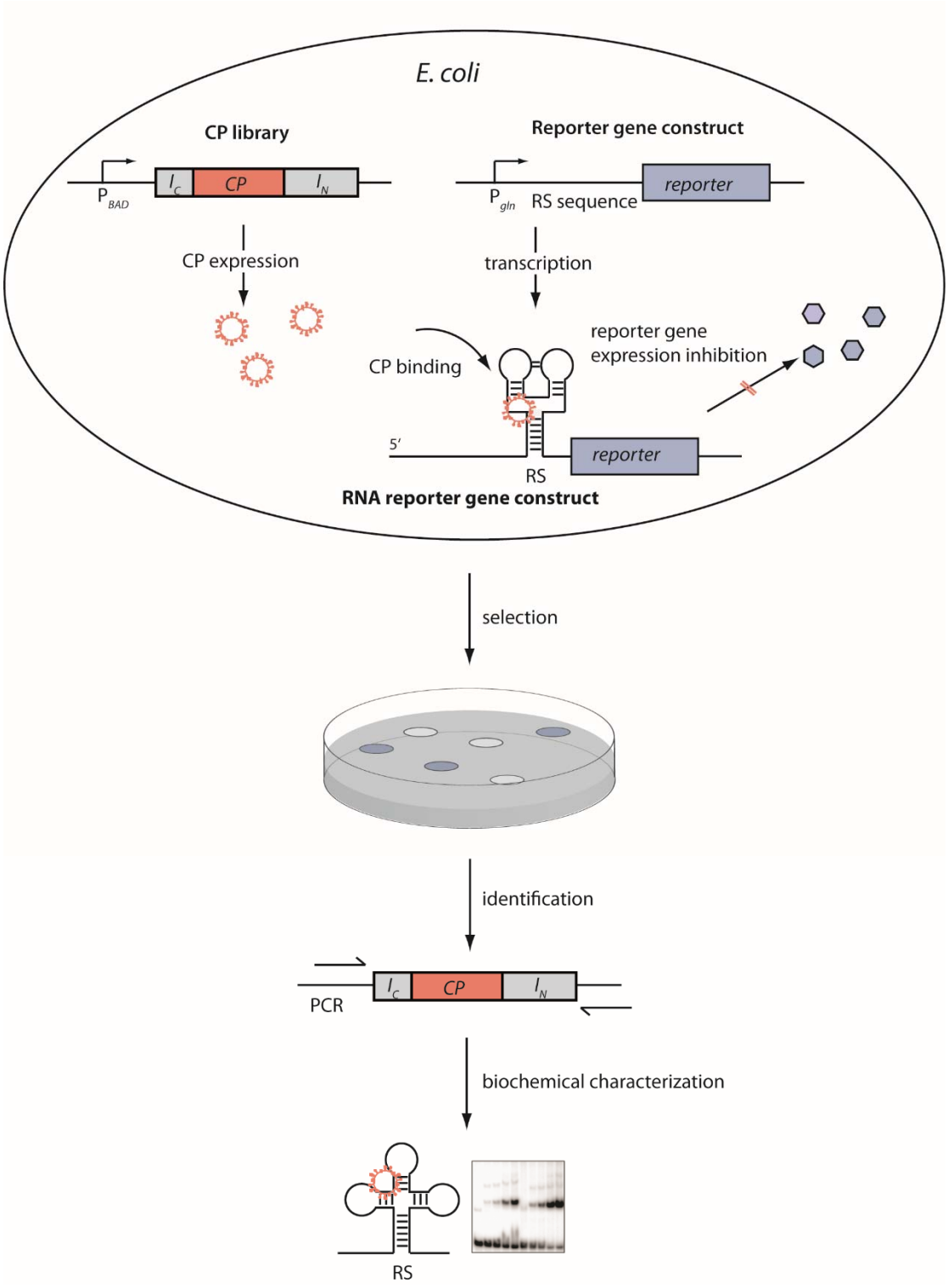
2. OBJECTIVES

Although RNA consists of only four building blocks, it is able to form complex three-dimensional structures. Since its discovery, RNA was thought to act merely as an intermediate, infrastructural component and messenger between genes and proteins. However, in the last decade RNAs have proven to be tremendously versatile molecules playing a pivotal role in many cellular key processes. RNAs fulfill their gene-regulating function by folding into three-dimensional structures that can be targeted by ligands. The aim of this thesis was to develop an *in vivo* selection system to identify bioactive small molecules from genetically encoded and expressed CP libraries. For this purpose, a RS from *Salmonella typhimurium* (*ST*) was used to control the expression of reporter genes, allowing a direct readout of RS-function. The screening assay developed in this project can also be used as model system for other RSs of virulent bacteria. Moreover, the *in vivo* CP library generation was investigated by direct quantification of CPs from *E. coli* cell extracts using liquid chromatography tandem mass spectrometry (LC-ESI-MS/MS).

The project outline is illustrated in **Figure 10**. Achievements of this thesis were

- Development of a reporter system which is suitable for a screening process
- Confirmation of CP generation *in vivo*
- Chemical synthesis of reference CPs
- Quantification of CPs from cell extracts
- Evaluation of systems generating CPs *in vivo*

Figure 10 *Project outline. The in vivo screening system reporter genes are controlled by the target RS. Bioactive CPs from a CP library, which alter reporter gene expression can be identified by sequencing. The binding properties can then be further characterized.*



3. RESULTS AND DISCUSSION

3.1 CYCLIC PEPTIDE LIBRARY

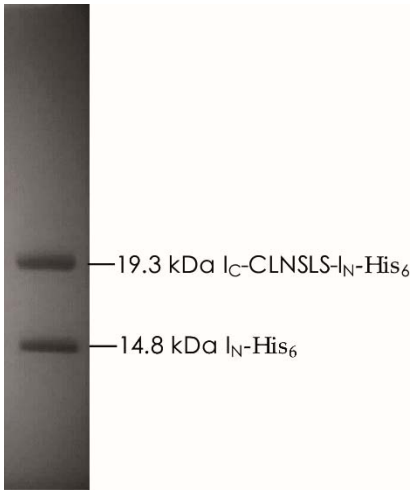
Two hexameric CP libraries (Cys + 5 randomized amino acids), genetically encoded by *pASK Iba5-lib* and *pBAD-lib* were kindly provided by Dr. Sabine Schneider. The *pASK Iba5-lib* carries an ampicillin resistance gene and an anhydrotetracycline promotor. The *pBAD-lib* mediates kanamycin resistance and is inducible by arabinose. Both libraries are based on SICLOPPS, utilizing *Ssp* DnaE split inteins to generate hexameric CPs. A His_6-tag at the I_N domain allows purification by nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography of the split intein constructs. Successful CP synthesis of *pASK Iba5-lib* and a reasonable size of both libraries (10^4 – 10^5 members) was proved previously by Dr. Sabine Schneider.

One expression plasmid of the *pBAD-lib* library was randomly chosen for further analysis of intein expression and CP generation. The CP sequence of this plasmid was CLNSLS with a molecular mass of 617.28 $\frac{g}{mol}$. In this work, the corresponding plasmid will be referred to as *pSSP CLNSLS*.

In order to confirm intein expression, *E. coli* JM109 harboring the *pSSP CLNSLS* plasmid was cultivated at 37°C until the OD_{600} reached 0.6-0.8. Gene expression was induced with 0.02% (w/v) L-arabinose and harvested after 18 hours of incubation at 37°C in order to keep CP production analogous to the preferred screening conditions. The cell lysate from a 0.2 g cell pellet was purified using Ni-NTA affinity spin columns. A 15% SDS gel (**Figure 11**) revealed bands of the spliced I_N-His_6 at 14.8 kDa and the unspliced intein I_C-CLNSLS-I_N-His_6 (19.3 kDa). I_C was not detected due to its small size of 3.9 kDa and the amount of sample loaded onto the gel. The identity of both visible bands was verified by tryptic digest and peptide-mass fingerprinting (Chair of Biotechnology, Prof. Buchner, TU München, data not shown). The presence of spliced I_N indicated CP formation *in vivo*. It was, however, possible that the splicing reaction was not yet complete and had stopped at the branched intermediate. In this case the spliced I_N would have been detected, whereas the

target peptide was still attached to I_C. To address this question, CP formation was verified by mass spectrometry.

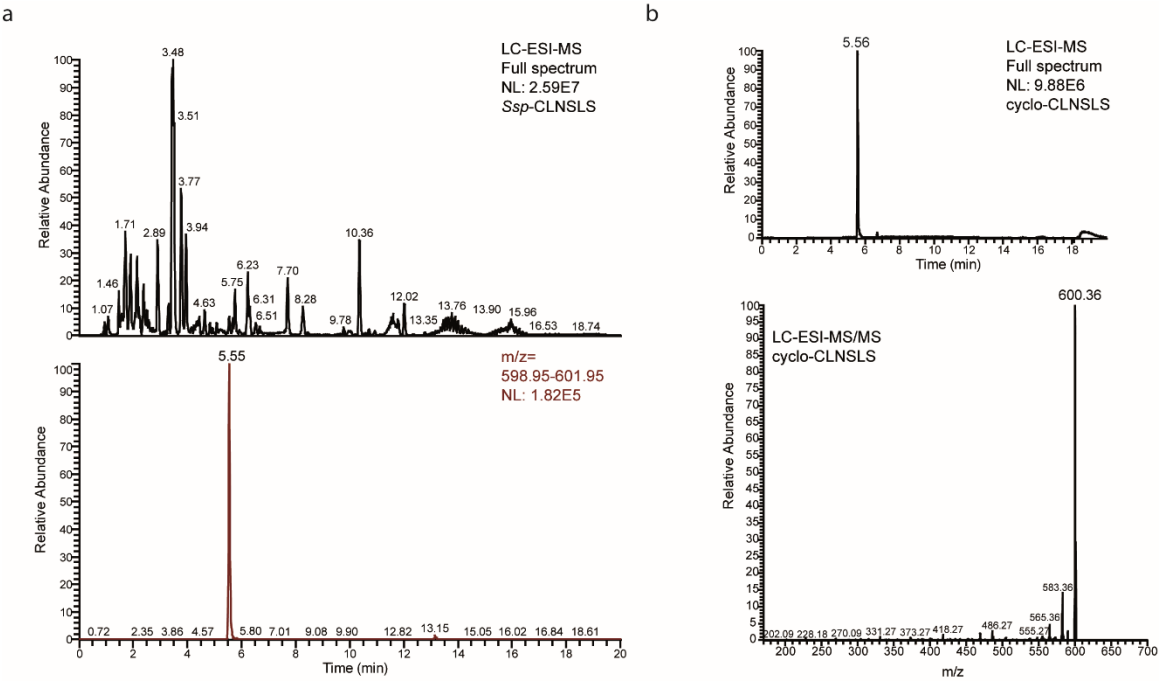
Figure 11 *15% SDS PAGE of the E. coli JM109 pSSP-CLNSLS Ni-NTA affinity chromatography. The elution fraction contained the spliced I_N-His_6 (14.8 kDa) and the unspliced intein I_C-CLNSLS-I_N-His_6 (19.3 kDa).*



For this purpose, 15 g of *E. coli* JM109 *pSSP CLNSLS* cell pellets were disrupted by sonication and extracted, followed by mass analysis. In order to optimize CP extraction, different solvents such as ethyl acetate, *n*-butanol, 100% methanol and 70% methanol were tested in addition to a variety of extraction buffers. The best results were obtained using 1 mM tris(2-carboxyethyl)phosphine (TCEP) in the extraction buffer and *n*-butanol as extraction solvent. After extraction, the organic fraction was removed *in vacuo* and residuals were dissolved in 1.5 mL 70% methanol. Due to the resolution limit of mass spectrometry, solvent volumes had to be minimized to a total amount of 60 µL prior to measurements.

MS analysis of complex cell extracts results in multiple peaks, making identification of the CP solely by mass unfeasible. Therefore, chemically synthesized cyclo-CLNSLS was used as a reference peptide and was measured by LC-MS/MS. Selected reaction monitoring (SRM) was subsequently performed with the cell extracts to identify a cyclo-CLNSLS daughter ion peak ($600.36\,\frac{g}{mol}$). Evaluation of the MS spectra confirmed the presence of the CP in the extract based on its retention time, molecule fragmentation and mass compared to the reference peptide (**Figure 12**).

Figure 12 *Confirmation of cyclo-CLNSLS synthesis in vivo* (**a**) *LC-ESI-MS analysis of the n-butanol cell extract of an E. coli JM109 pSSP CLNSLS cell pellet. A mass filter for the cyclo-CLNSLS daughter ion (600.36* $\frac{g}{mol}$*) was applied and results in a single peak at a retention time of 5.55 ppm.* (**b**) *LC-ESI-MS/MS spectrum of a chemically synthesized cyclo-CLNSLS reference.*



3.2 DEVELOPMENT OF THE *IN VIVO* SELECTION SYSTEM

3.2.1 REPORTER GENE SYSTEM

The reporter gene constructs were cloned according to standard molecular biological techniques described in section **5.2.1**. All reporter constructs are based on pACYCDuet-1-P_{GlnS}-*pylS* as described by Dr. Veronika Flügel. [111] The constructs consisted of a constitutive glutamyl-tRNA synthetase (glnS) promoter, ensuring permanent transcription, and a *ST* AdoCbl RS which controls the expression of reporter genes. It has been shown previously that AdoCbl RSs require the leader region of their naturally regulated gene, *btuB*, to work properly. [9, 53, 54, 112] Therefore, the RS sequence and the first 210 bp of the *btuB* gene were cloned upstream of the reporter genes. Three different reporter genes were chosen for investigations:

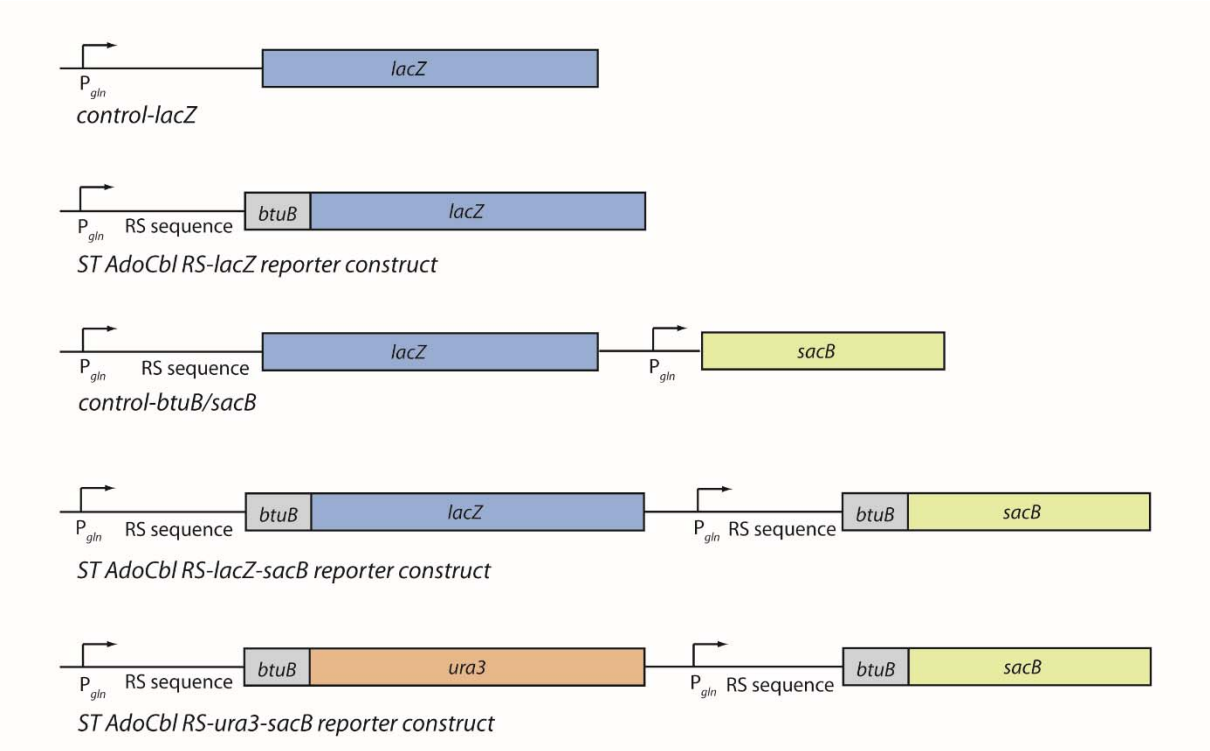
- *lacZ* served for qualitative and quantitative readout of the reporter gene expression. It was derived from *E. coli* K12 genomic DNA and Met3 was used as an alternative start codon in order to avoid mutations in *lacZ* (**Appendix**).
- *sacB* is a reverse selection marker. It was required for the selection as bioactive CPs were expected to constitute a minority of the CP library. In the provided *sacB*, derived from *Bacillus subtilis* genomic DNA, an internal *Hin*dIII restriction site had already been eliminated to facilitate cloning (**Appendix**).
- *ura3* is also a reverse selection marker and was derived from *Saccharomyces cerevisiae* (**Appendix**).

It was shown that expression control is improved when each reporter gene is under the influence of a single glnS promotor. [111] For that reason, only split reporter gene constructs were cloned. An illustration of all reporter gene constructs is shown in **Figure 13**. The functionality of *lacZ* regulation was examined using the *ST AdoCbl RS-btuB-lacZ reporter construct*. *Control-lacZ* lacked the RS sequence and therefore served as a positive control for *lacZ* regulation. The plasmid was kindly provided by Dr. Veronika Flügel. A negative control for *sacB* regulation was the *control-btuB/sacB*, which was kindly provided by Dr. Sabine Schneider. Furthermore, *control-btuB/sacB* lacked the 210 bp *btuB* leader region in front of the *lacZ,* controlling the RS. *SacB* regulation was examined using the *ST AdoCbl RS-lacZ-sacB*

reporter construct and *ura3* regulation was investigated using the *ST AdoCbl RS-ura3-sacB reporter construct.*

In order to proof that the *ST* AdoCbl RS regulates expression of the downstream reporter genes, the extent of reporter gene expression was determined in the presence and absence of the RS's natural ligand AdoCbl.

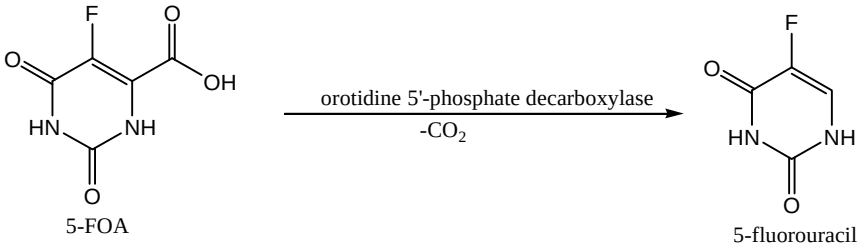
Figure 13 *Illustration of reporter gene constructs*



Regulation of *ura3*

The *ura3* gene encodes for orotidine 5'-phosphate decarboxylase and is involved in the synthesis of pyrimidine ribonucleotides. The addition of 5-fluoroorotic acid (5-FOA) leads to the toxic compound 5-fluorouracil (**Figure 14**) and allows negative selection.

Figure 14 *Reaction of orotidine 5'-phosphate decarboxylase with 5-FOA*



PyrF is the *E. coli* homologue of *ura3* and deletion of *pyrF* can also be complemented by expression of the *ura3* gene of *S. cerevisiae.* [113, 114] In order to investigate the RS mediated regulation of *ura3*, *pyrF* deficient *E.coli* US0ΔpyrF cells were used.

RS control of the *ura3* gene was verified by qualitative determination of orotidine 5'-phosphate decarboxylase expression. For this purpose, 5-FOA-dependent cell growth assays were performed as described in section **5.2.2**. *E. coli* US0ΔpyrF cells were grown in NM minimal medium in the absence and presence of 5-FOA and RS ligand. Without the addition of a RS ligand, cell growth is expected to be inhibited because of the formation of toxic 5-fluorouracil. Upon addition of AdoCbl, cell growth should be restored as expression of orotidine 5'-phosphate decarboxylase is inhibited.
5-FOA-dependent cell growth assays were performed using the *ST AdoCbl RS-ura3-sacB reporter construct.*

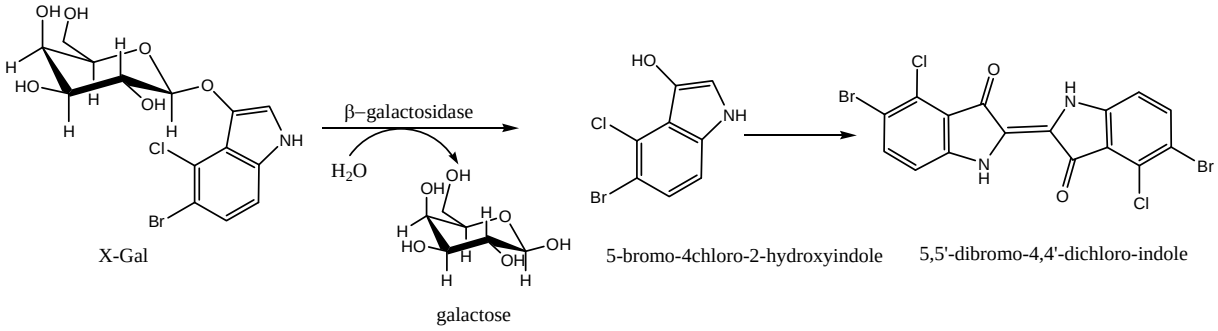
Despite variations in the 5-FOA (0 - 20 mM) and AdoCbl (5 - 10 µM) concentrations, no sufficient gene regulation was observed. At 5-FOA concentrations above 5 mM, the reporter construct did not show ligand-induced recovery of cell growth and at lower concentrations cell growth was not inhibited to a satisfactorily degree (data not shown). Hence further investigations concentrated on the reporter genes *sacB* and *lacZ*.

Regulation of *lacZ*

LacZ is a widely used reporter gene and encodes for β-galactosidase, an enzyme that hydrolyses β-galactosides into monosaccharides. *LacZ* allows qualitative as well as quantitative readout of gene expression. β-Galactosidase cleaves 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) into galactose and 5-bromo-4-chloro-3-

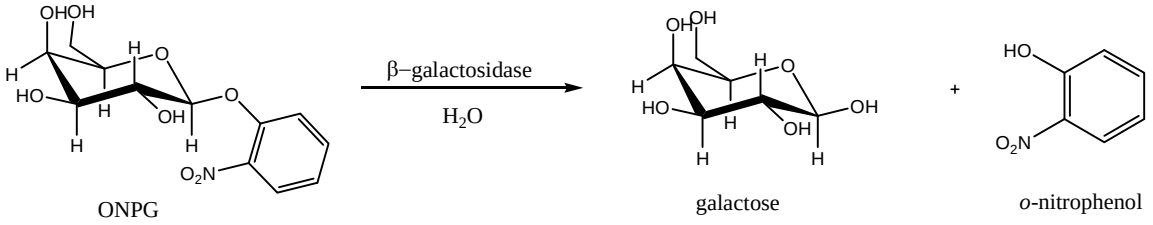
hydroxyindole. Dimerization of the latter leads to 5,5'-dibromo-4,4'-dichloro-indigo, a blue dye (**Figure 15**). Therefore, the addition of 0.04 $\frac{mg}{mL}$ X-Gal to agar plates allows so-called blue-white screenings and a qualitative evaluation of β-galactosidase expression.

Figure 15 *Reaction of β-galactosidase with X-Gal*



Quantitative gene expression of *lacZ* can be determined using *o*-nitrophenyl-β-galactoside (ONPG), a colorless substrate. The reaction of β-galactosidase with ONPG yields galactose and yellow *o*-nitrophenol, which absorbs at 420 nm (**Figure 16**).

Figure 16 *Reaction of β-galactosidase with ONPG*

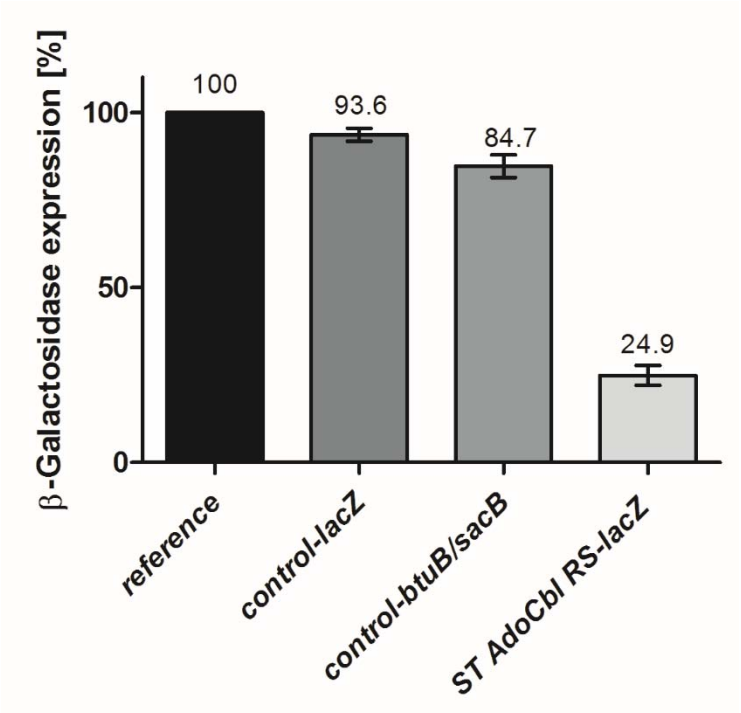


RS control of the *lacZ* gene was verified by quantifying the level of β-galactosidase expression. Has the RS switched the gene expression to "on", β-galactosidase is expressed and converts ONPG to *o*-nitrophenol resulting in a higher absorption at 420 nm. Binding of AdoCbl to the RS inhibits gene expression and hence the absorption level at 420 nm is reduced.

For evaluation of the assay, the β-galactosidase expression in absence of AdoCbl (*reference*) was normalized to 100%. The expression is reduced accordingly when

AdoCbl is present. The β-galactosidase assay confirmed RS controlled gene expression of the *ST AdoCbl RS-lacZ reporter construct* (**Figure 17**). A reduction in relative β-galactosidase expression to 24.9% was accomplished. A positive control, which was lacking the RS sequence (*control-lacZ*), exhibited a reduction of enzyme expression to 93.6%, proving it to be independent of AdoCbl addition. The *control-btuB/sacB* construct lacked the *btuB* leader region and showed a relative β-galactosidase activity of 84.7%. Therefore, *control-btuB/sacB* demonstrated that parts of *btuB* are necessary for efficient downstream regulation of gene expression.

Figure 17 *β-Galactosidase assay of reporter constructs in the presence and absence of AdoCbl. Expression of β-galactosidase in the absence of AdoCbl was normalized to 100% (reference). As a positive control, a lacZ expression construct was used which lacked the RS sequence (control-lacZ). The ST AdoCbl RS-lacZ reporter construct showed a significant reduction of β-galactosidase expression. Only minimal regulation of lacZ through addition of AdoCbl was observed when the btuB leader region was absent (control-btuB/sacB).*



Regulation of *sacB*

As *lacZ* turned out to be a highly suitable reporter gene, a reporter construct was cloned which contained a RS controlled *lacZ* in the first and a RS controlled *sacB* in the second multiple cloning site. *SacB* encodes for levansucrase, an enzyme that catalyzes polymerization of saccharides. In a transglycosylation reaction fructose, originated from sucrose, is condensed to form levan-polysaccharides according to the following equation:

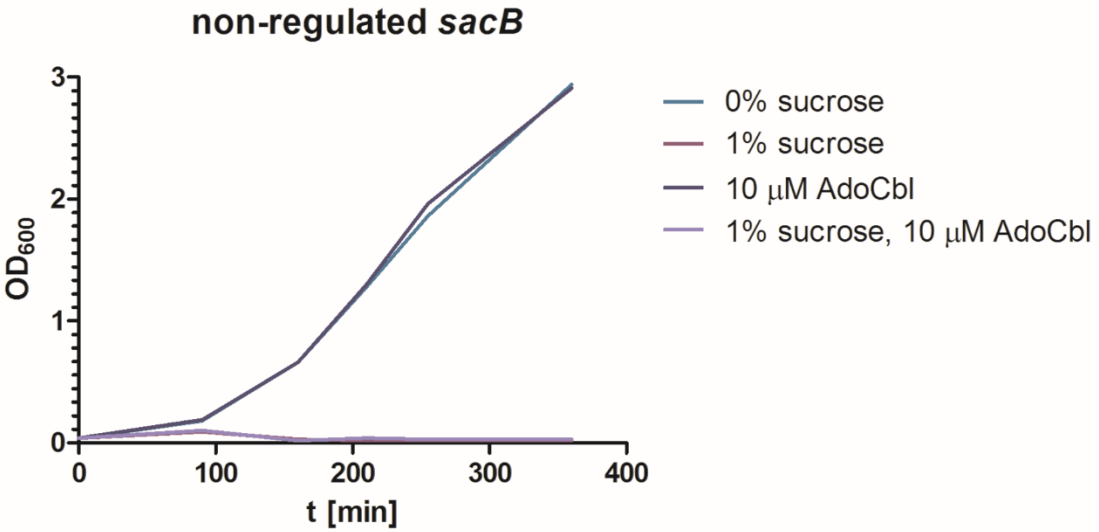
$$sucrose + acceptor \longrightarrow glucose + fructosyl - acceptor$$

Such fructose polymers are lethal for *E. coli* and hence *sacB* can be used as a negative selection marker. RS mediated control of *sacB* was monitored by the cell growth advantage in sucrose containing LB medium supplemented with AdoCbl. *SacB* expression in the presence of sucrose should cause cell death and RS-mediated repression of *sacB* expression should restore viability.

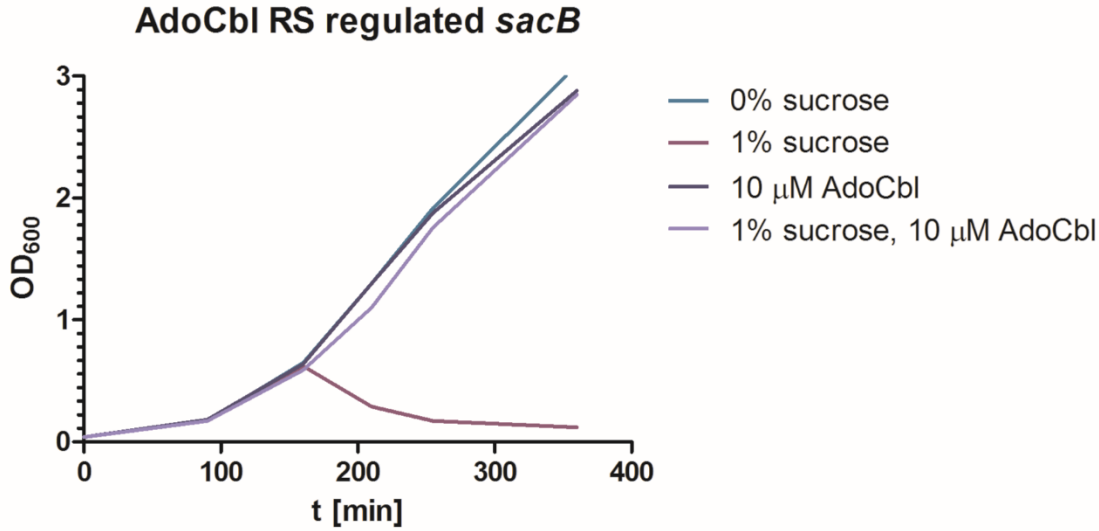
Only freshly transformed *E. coli* JM109 cells bearing the reporter gene constructs were used for overnight cultures. Concentrations of 0.25 - 2% sucrose and 2 – 20 µM AdoCbl were investigated and media containing 1% sucrose and 10 µM AdoCbl gave the best results (**Figure 18**). Cells were not able to grow in the absence of AdoCbl when the medium was supplemented with sucrose. *E. coli ST AdoCbl RS-lacZ-sacB* cells showed nearly full recovery upon addition of 10 µM AdoCbl. A construct in which *sacB* was not RS-controlled (*control-btuB/sacB)* showed no AdoCbl-induced recovery of cell growth.

Figure 18 *Sucrose cell growth assay of reporter constructs in the presence and absence of 10 μM AdoCbl and 1% sucrose.* (**a**) *Cells containing the reference construct control-btuB/sacB without RS showed no viability in the presence of 1% sucrose.* (**b**) *Cells bearing the ST AdoCbl RS-lacZ-sacB reporter construct showed significant growth advantage in media containing 1% sucrose when it was supplemented with 10 μM AdoCbl.*

a



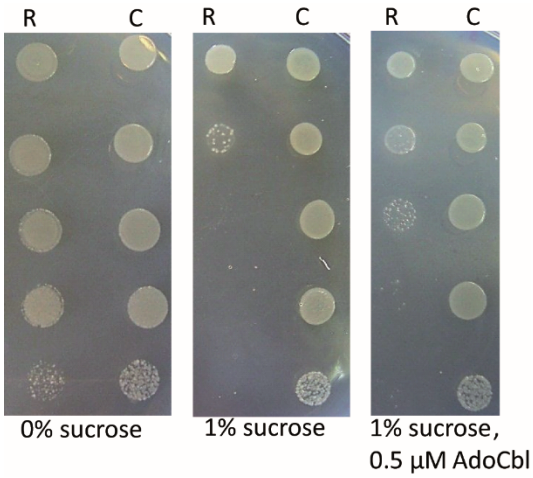
b



Inhibition of *sacB* expression by the ST AdoCbl RS was also shown by serial dilution tests on solid media (**Figure 19**). For agar plates, the addition of 0.5 μM AdoCbl was sufficient for RS regulation. Control cultures without the RS (*ST AdoCbl RS-lacZ)*

remained unaffected upon addition of 1% sucrose and 0.5 µM AdoCbl. In contrast, cell cultures with the *ST AdoCbl RS-lacZ-sacB reporter construct* showed sensitivity in the presence of sucrose, which was decreased when AdoCbl was added.

Figure 19 *Serial dilution tests of cell cultures containing the reporter gene constructs ST AdoCbl RS-lacZ (control, C) and the ST AdoCbl RS-lacZ-sacB reporter construct (R).*



Use of the *ST AdoCbl RS-lacZ-sacB reporter construct* proved RS mediated regulation of the reporter genes *lacZ* and *sacB.* This vector was therefore used for *in vivo* high-throughput screenings of SICLOPPS libraries. Successful RS regulation of a bioactive compound should result in a growth advantage in the presence of 1% sucrose, white colonies in the presence of 0.04 $\frac{mg}{mL}$ X-Gal and colorless samples in a β-galactosidase assay.

For this purpose, electrocompetent *E. coli* cells containing the *ST AdoCbl RS-lacZ-sacB reporter construct* were prepared. *E. coli* US0ΔpyrF and *E. coli* JM109 were chosen due to their ability to grow in minimal media. The transformation efficacy was determined to $2.1 \cdot 10^7$ $\frac{cfu}{\mu g}$ for *E. coli* US0ΔpyrF and $2.6 \cdot 10^8$ $\frac{cfu}{\mu g}$ for *E. coli* JM109.

3.2.2 High-throughput Screening of SICLOPPS Libraries

Two genetically encoded hexameric CP libraries (*pASK Iba5-lib* and *pBAD-lib*) were subjected to the high-throughput screenings as hexameric CPs from SICLOPPS had already proven their ability to act as bioactive compounds. [87] Every screening process consisted of three steps. In the first round, colonies were selected on agar plates with respect to their white color and growth advantage in the presence of 1% sucrose. In the second round of selection, serial dilutions of potential candidates were performed on selective media supplemented with and without 1% sucrose and either arabinose or anhydrotetracycline. This step was included in order to eliminate false positive colonies that may have emerged, for example, due to mutations in *sacB*. Agar plates containing exclusively arabinose or anhydrotetracycline were used to confirm that CPs target the reporter gene construct and no other *E. coli* targets that might be involved in viability. Colonies which passed the first two rounds of selection were further investigated in a third round by a β-galactosidase assay and a sucrose growth assay in solution in order to quantify inhibition by the bioactive CP.

Anhydrotetracycline inducible library

E. coli US0ΔpyrF cells bearing the *ST AdoCbl RS-lacZ-sacB reporter construct* were transformed with the *pASK Iba5-lib* and incubated in 37°C warm SOC medium. Cells were then washed with NM minimal medium and plated on NM minimal medium agar plates supplemented with 1% sucrose, 200 µg/L anhydrotetraycline and 0.04 mg/mL X-Gal. About 130 white colonies were observed after two days of incubation at 37°C. This high number of potential hits was not surprising, since similar observations have been reported previously for other selection systems. [87] In the second round of selection, serial dilutions of all candidates were performed on NM minimal medium supplemented with and without 1% sucrose and 200 µg/L anhydrotetraycline. It became apparent that the addition of 200 µg/L anhydrotetracycline, which was necessary for induction of the library, decreased the cell growth of *E. coli*. Nevertheless, about 13 promising colonies were further investigated in the third selection round by reporter gene assays in solution.

However, all colonies turned out to be false positives due to mutations in the *sacB* gene. Subsequent screenings gave similar results.

The fact that the addition of anhydrotetracycline was influencing cell growth complicated the identification of potential hits. Furthermore, *E. coli* US0ΔpyrF cells turned out to be comparatively sensitive and overnight cultures showed impaired growth. Consequently, all further screenings were performed with the arabinose inducible *pBAD-lib* in *E. coli* JM109. The minimal medium was also changed to M9 minimal medium as it consisted of fewer separate components.

Arabinose inducible library

The arabinose inducible library was investigated using *E. coli* JM109 cells containing the *ST AdoCbl RS-lacZ-sacB reporter construct*. Screenings were performed analogous to previous descriptions. The amount of potential positive colonies increased to over 200 per transformation after the first round of selection. The second round of selection revealed that the addition of arabinose did not have an effect on cell growth.

Unfortunately, all 16 potential hits that reached the third selection round showed a growth advantage resulting from mutations in *sacB*. These mutations may have developed as a bacterial defense mechanism which evolved as a consequence of the long-term selection pressure.
In order to optimize the conditions, further screens were performed in freshly transformed cells pre-incubated with arabinose-containing SOC medium. Beyond that, the arabinose concentration was increased, ranging from 0.0002 to 0.2% (w/v). Despite these optimizations, no bioactive CP was identified. One reason might be that there was no bioactive CP in the library which targeted the *ST* AdoCbl RS. Another explanation could be the fact that levansucrase can also accept arabinose as a fructosyl acceptor. [115] The 4-hydroxyl group of arabinose can be linked to the C2-hydroxyl group of the fructofuranosyl moiety in the oligosaccharide production reaction. Thus, arabinose may be unavailable for the induction of the peptide library. Usage of an arabinose inducible library in combination with *sacB* as a reporter gene may therefore require a much higher concentration of arabinose. An alternative

option to optimize screening conditions could be the replacement of the library's promoter in order to avoid an arabinose inducible system in combination with the reporter gene *sacB.*

In summary, *sacB* turned out to be a suitable reporter gene for the screening process and successfully reduced the number of colonies in the first round of selection. β-Galactosidase assays were advantageous for quantifying the inhibition of gene expression and false positive candidates were identified in the third round of selection. The use of *lacZ* in the initial selection rounds has to be evaluated more critically. Many colonies turned out to be light blue in the first round of selection and a visible distinction between white and blue colonies was challenging. The addition of another negative selection marker to the reporter system may therefore be beneficial in order to identify bioactive compounds.

Another factor that influences the selection in high-throughput screenings is the copy number of plasmids. It is a well-established fact that the plasmid copy number in *E. coli* varies. Depending on the origin of replication, it is possible to distinguish between low copy (1-12 plasmids per cell), medium copy (15-20 plasmids per cell) and high copy plasmids (20-700 plasmids per cell). [116] Factors affecting the copy number include for instance cell growth. [117] Especially for the development of a reporter gene assay, a constant number of the reporter gene is important. Otherwise, bioactive compounds may be overlooked if the number of reporter genes is higher than the number of bioactive compounds. Genome integration of the reporter construct would ensure stable expression and guarantee the excess of CPs compared to the reporter genes. Therefore, the *ST AdoCbl RS-lacZ-sacB reporter construct* was integrated into the *E. coli* genome.

Integration of a reporter construct in the genome of *E. coli*

The *AdoCbl RS-lacZ-sacB reporter construct* was, with its 6146 bp, too large in size for common integration methods such as recombineering. For this reason, integration of the reporter construct into the genome of *E. coli* JM109 was performed according to a method developed by Kuhlman and Cox. [118]. This method uses λ-phage proteins for recombineering and restriction endonucleases

which introduce double strand breaks to promote the integration of the desired fragment. This way, large fragments up to a size of 7,000 bp can be inserted into any desired location.

In the first step of this method, a helper plasmid pTKRED is used for the transformation of competent *E. coli* cells. It harbors I-SceI endonuclease, an arabinose inducible endonuclease from yeast which is not naturally occurring in *E. coli*. In addition pTKRED contains the IPTG inducible λ-Red genes (Redγ (*gam*), Redα (*exo*), Redβ (*bet*)) for homologous recombination as well as a temperature-sensitive ORI, which facilitates plasmid removal after successful integration. pTKRED is utilized to integrate a tetracycline resistance gene (*tetA*) flanked by I-SceI recognition sites and 25 bp landing pad regions into the chromosome by homologous recombination.

After successful integration of *tetA* as a "place holder", larger sequences may be inserted. For this purpose, a donor plasmid is transformed. It carries the insertion fragment which is also flanked by I-SceI recognition sites as well as the 25 bp landing pad regions. Excision of I-SceI triggers the integration of the desired fragment between the landing pad regions. Finally, pTKRED is removed from the cells by increasing the temperature to 42°C and thereby preventing the plasmid's amplification.

The plasmids pTKRED and pTKIP were kindly provided by Prof. Thomas E. Kuhlman from the University of Illinois, US. The DNA template of LP-*tetA* was purchased from GeneArt AG using the same sequence as described in literature instead of utilizing pTKS/CS. [118] The donor plasmid was generated by cloning the whole *ST AdoCbl RS-lacZ-sacB reporter construct* between the landing pads of pTKIP.

The reporter construct was inserted between the *nth-tppB* gene. Strand breaks in this region of the *E. coli* chromosome are lethal. Therefore, only successful integrants with an intact chromosome were able to survive. Successful integration into the *E. coli* genome was confirmed by PCR and sequencing.

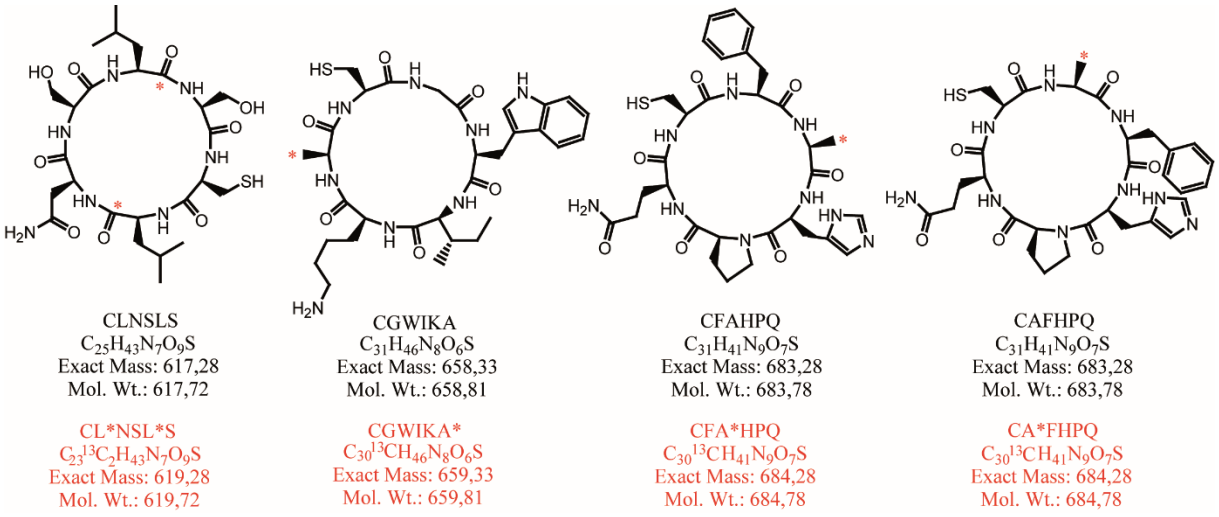
Sucrose cell growth assays in solution and on agar plates as well as β-galactosidase assays were performed with the integrated reporter gene construct. Reporter gene assays revealed that gene expression was close to the detection limit. It remains unclear why insufficient reporter gene expression was observed. The *glnS* promoter may have been too weak for expression and its substitution may be required. An

alternative approach would be the integration of the reporter construct into another location of the genome.

3.3 Impact of Split Inteins and Cyclic Peptide Sequences on SICLOPPS

It has been shown that split inteins from different organisms possess diverse splicing rates, yields and temperature optima. The efficiency and velocity of the splicing reaction strongly depends on the target sequence of the extein (see section **1.3**). Especially for CP library formation in combination with an *in vivo* selection system, it is necessary to have a reliable CP generation system. Otherwise potentially active CPs may be missed due to low intracellular levels. In order to optimize CP generation, the impact of the target sequence and the efficiency of the split inteins *Npu* and *Ssp* DnaE were compared. Therefore the CPs CLNSLS, CGWIKA, CFAHPQ and CAFHPQ were generated by the two inteins *in vivo* and quantified directly from *E. coli* cell extracts using LC-ESI-MS/MS. The complex composition of cell extracts (see **Figure 12 a**), made it necessary to synthesize reference molecules. Furthermore, quantification by mass spectrometry required an internal standard and the generation of a calibration curve. ^{13}C-labeled and unlabeled analogues of cyclo-CLNSLS, cyclo-CGWIKA, cyclo-CFAHPQ and cyclo-CAFHPQ were thus chemically synthesized. (**Figure 20**) The major advantage of using isotopes as internal standards was that they exhibit the same chemical properties as their unlabeled partners but can be differentiated by mass analysis.

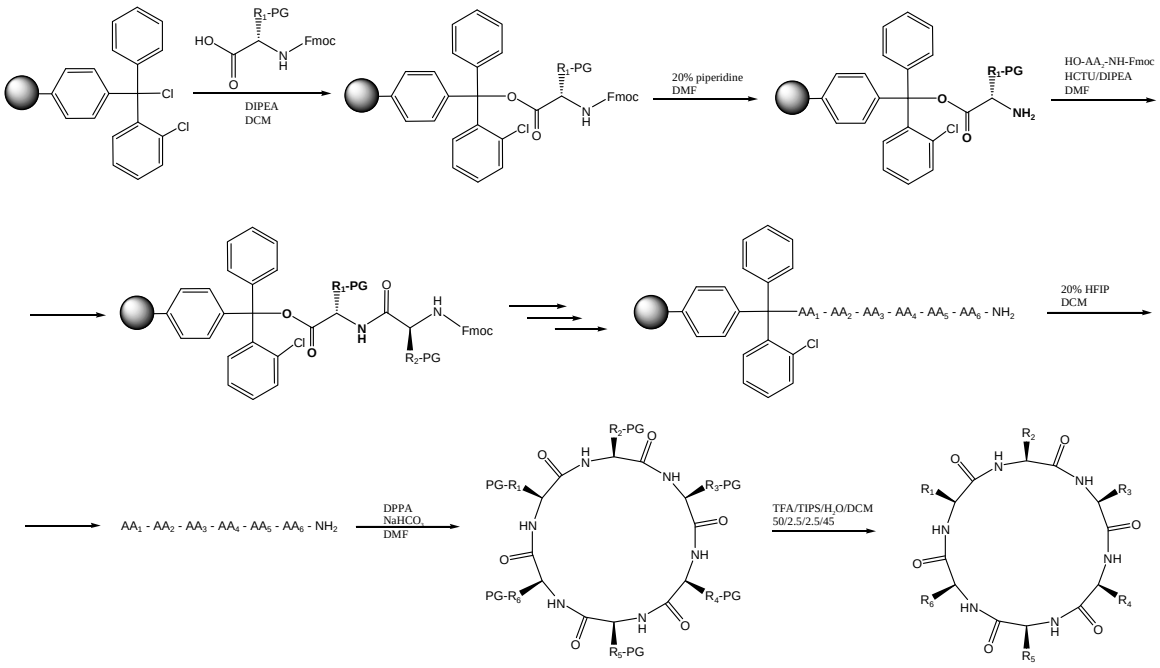
Figure 20 *In vivo generated and chemically synthesized CPs. Positions of the ^{13}C isotopes in the labeled peptides are highlighted with a red asterisk.*



3.3.1 Chemical Synthesis of Cyclic Peptides

Fmoc based solid phase peptide synthesis was used to generate cyclo-CLNSLS, -CGWIKA, -CFAHPQ and -CAFHPQ as well as their ^{13}C labeled isotopes cyclo-CL*NSL*S, -CGWIKA*, -CFA*HPQ and -CA*FHPQ. An overview of the chemical synthesis of CPs is presented in **Figure 21**. As a solid support, 2-chlorotrityl chloride (Trt-Cl) resin was used. This highly acid labile resin offers several advantages. Due to its steric bulk it prevents, for example, formation of diketopiperazine if Pro was added as a C-terminal residue. Amino acids can also be attached without racemization. In addition, and most importantly, it can be cleaved off under mild conditions, retaining the peptide's side chain protecting groups. This property was necessary for a subsequent head-to-tail cyclization of the linear peptide in solution. In the end, trifluoric acid (TFA) removes the CP's side chain protection groups.

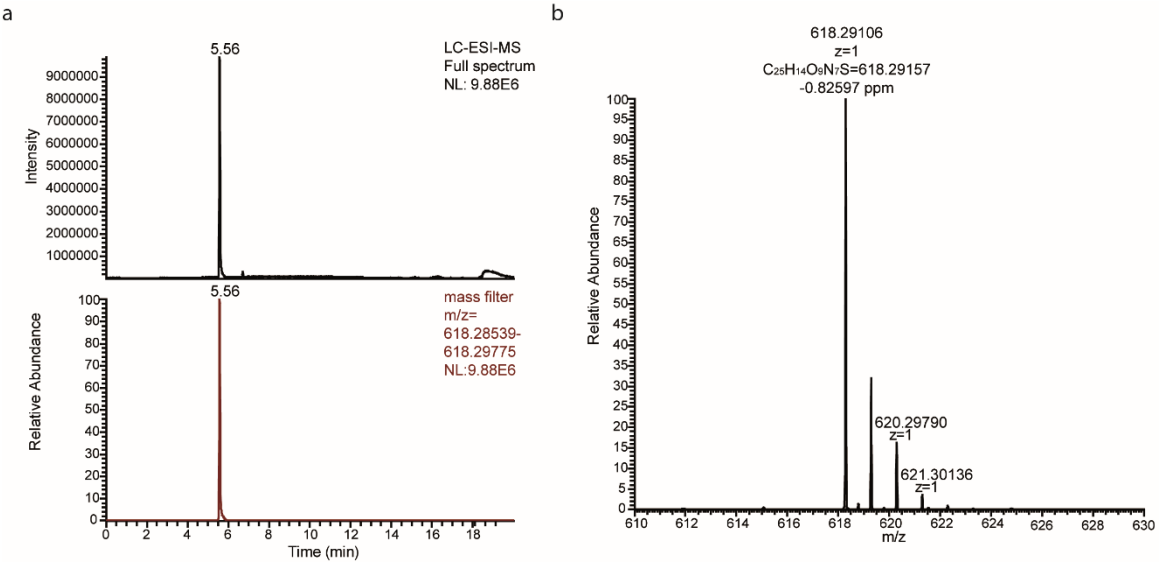
Figure 21 *Overview of the synthesis of CPs*



The crude CP products were purified by reversed phase high-performance liquid chromatography (HPLC) using a C18 column (Xbridge BEH, 10 mm x 250 mm, Waters, Eschborn, GER). Acetonitrile (ACN)/H_2O gradients were adjusted for each peptide, ensuring high purity. Due to dimerization of CPs, 1 mM TCEP was added to the solved crude products prior to purification. Subsequent high resolution mass

analysis further confirmed successful synthesis as judged from the molecular weight and a high purity of each peptide (**Figure 22** and **Appendix**).

Figure 22 *LC-ESI-MS of cyclo-CLNSLS.* (**a**) *A single peak at a retention time of 5.56 min was observed in the full MS spectrum and shows the high purity of the CP. Application of a mass filter corresponding to the molecular weight of cyclo-CLNSLS results in the same peak.* (**b**) *A deviation of -0.82597 ppm between observed and theoretical molecular weight of cyclo-CLNSLS confirmed successful chemical synthesis.*



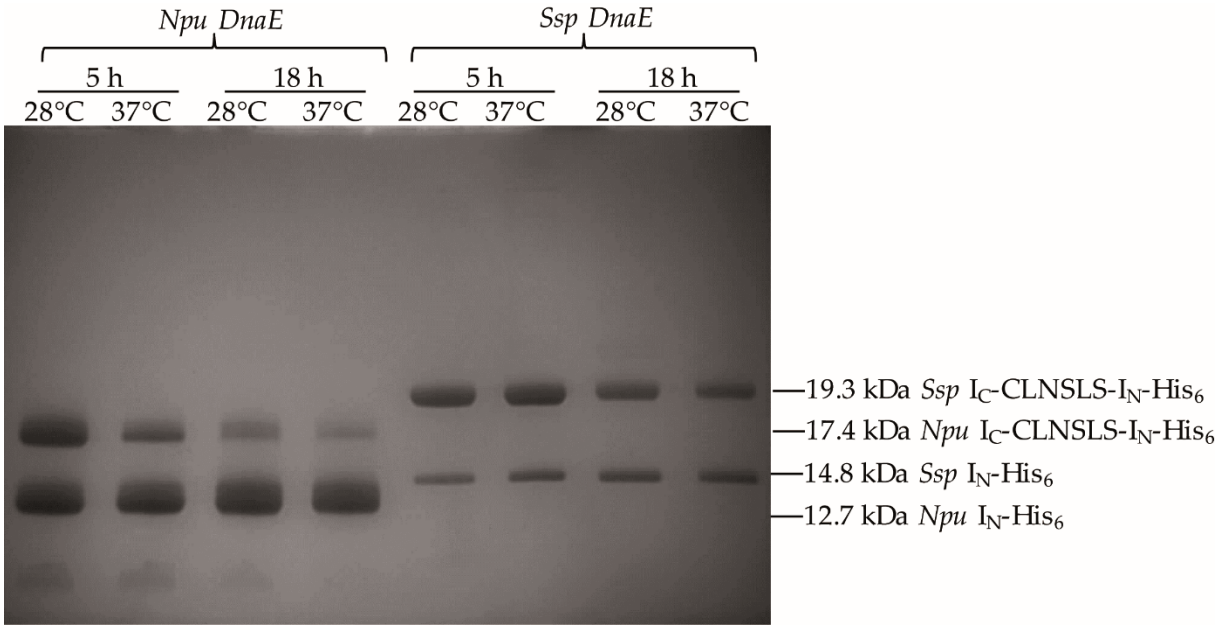
3.3.2 Quantification of in vivo Peptide Synthesis

Ssp and *Npu* DnaE split inteins were compared regarding the impact of the target sequence and the splicing efficiency in SICLOPPS. The *Npu* DnaE sequence was synthesized by GeneArt AG. The plasmids which were used for the investigations were built up analogous to the CP library plasmids of *pBAD-lib* with either *Npu* or *Ssp* DnaE inteins flanking the target peptide sequence. Introduction of the target sequences (CLNSLS, CGWIKA, CFAHPQ and CAFHPQ) between the split inteins was achieved by utilizing SLIM-PCR and standard cloning methods. The construct coding for *Npu*-CGWIKA was cloned in this thesis, all other constructs were kindly provided by Dr. Sabine Schneider. As a result, four pairs of plasmids were used for the cyclization of the target peptide sequence *in vivo* through *Npu* DnaE or *Ssp* DnaE.

Initially, temperature dependence of protein expression was investigated using *E. coli* JM109 *pSSP CLNSLS* and *Npu* I_C-CLNSLS-I_N-His_6 (*pNPU CLNSLS*). For this purpose, expression was carried out under defined growth conditions in M9 minimal medium. After induction with arabinose, the temperature was either kept constantly at 37°C or decreased to 28°C. Samples were taken 5 h and 18 h after induction and the intein expression was detected by Ni-NTA affinity chromatography (**Figure 23**). As expected, the amount of unspliced intein decreased in all samples with longer incubation periods. Both inteins showed best results at 37°C and an incubation time of 18 hours. The good splicing ratio of unspliced:spliced intein of *Npu* DnaE at 37°C is consistent with literature. [99] It has further been reported that *Ssp* DnaE splicing yields are reduced and side product formation is increased at temperatures above 30°C. [99] Here, in this qualitative analysis, no change in *Ssp* intein expression was observed when the temperature was varied.

All further experiments were carried out under defined growth conditions at 37°C and with an incubation period of 18 h before cells were harvested. These conditions also mirror the screening conditions of CP libraries. (see **3.2.2**)

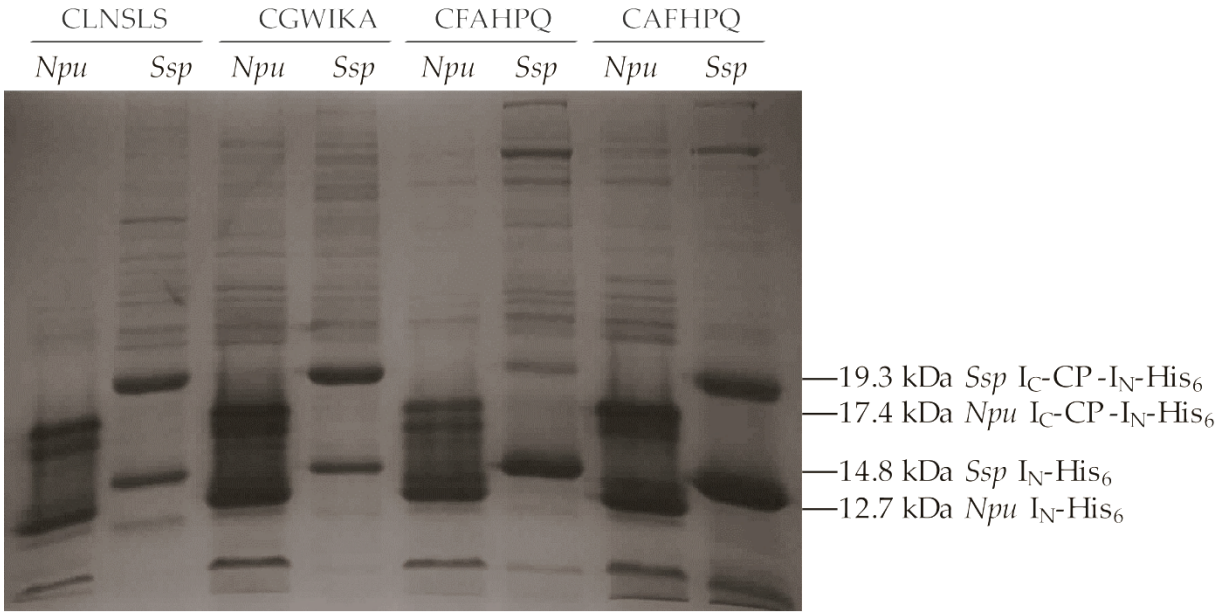
Figure 23 *SDS-Page analysis of temperature dependent intein expression and splicing of Npu-CLNSLS and Ssp-CLNSLS. The split inteins were purified from 0.2 g cell pellets using Ni-NTA affinity chromatography and the elution fractions were visualized on a 15% SDS Gel.*



The intein expression of all eight CP constructs was carried out under conditions previously described in section **3.1**. All inteins were purified *via* Ni-NTA affinity chromatography and analyzed by SDS-PAGE (**Figure 24**). Elution fractions revealed bands that corresponded to spliced I_N-His_6 (14.8 kDa *Ssp*-inteins, 12.7 kDa *Npu*-inteins) and the unspliced product I_C-target peptide-I_N-His_6 (19.3 kDa *Ssp*-inteins, 17.4 kDa *Npu*-inteins). Successful intein expression for all eight constructs was verified by tryptic digest and peptide-mass fingerprinting (Chair of Biotechnology, Prof. Buchner, TU München, data not shown).

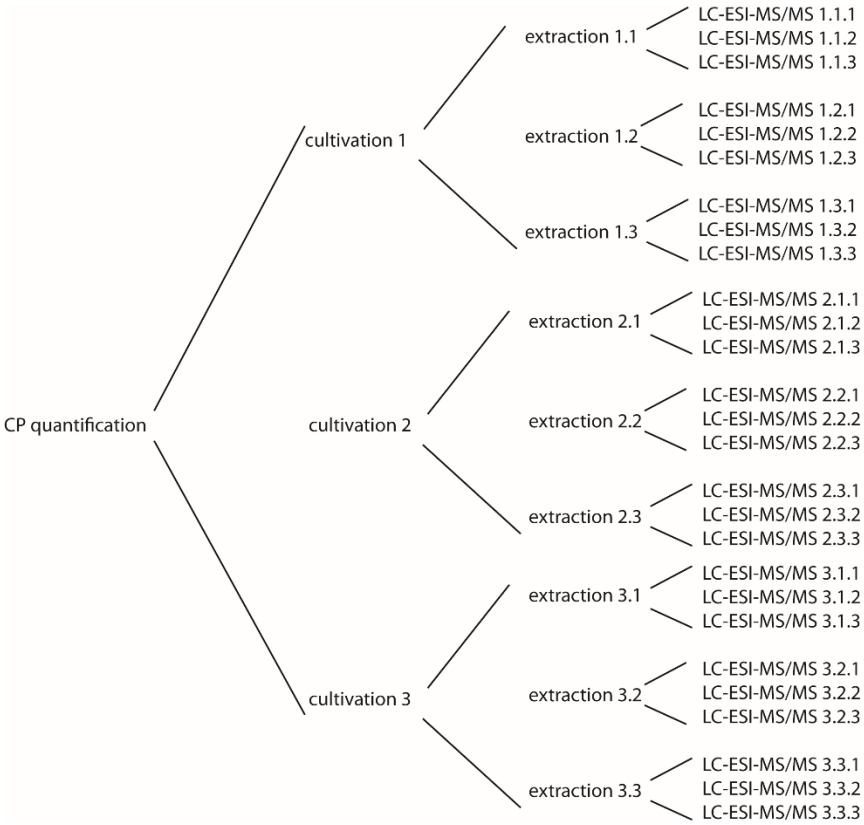
As shown in **Figure 24**, all inteins were expressed in similar amounts. For *Ssp* DnaE, the highest protein expression level was observed with the target sequences CFAHPQ and CAFHPQ. *Ssp*-CFAHPQ exhibited the best spliced:unspliced intein ratio of *Ssp* DnaE inteins. The amounts of unspliced products of *Npu*-inteins were comparable. Differences in protein yields among split inteins may be connected to the CP sequence. Another factor that influences protein yields is the divergent growth of *E.coli* cells which results in variations in protein expression levels.

Figure 24 *SDS-PAGE analysis of the protein expression of investigated inteins. Split inteins from 0.2 g E. coli JM109 pSSP-target peptide or 0.2 g E. coli JM109 pNPU-target peptide cell pellets were purified via Ni-NTA affinity chromatography. Elution fractions contain the spliced I_N-His_6 (14.8 kDa Ssp-inteins, 12.7 kDa Npu-inteins) and the unspliced intein I_C-target peptide-I_N-His_6 (19.3 kDa Ssp-inteins, 17.4 kDa Npu-inteins).*



For a more accurate analysis of the correlation between CP sequence and yields, CPs generated from *Npu* and *Ssp* were quantified from cell extracts using LC-ESI-MS/MS. LC-ESI-MS/MS analysis and quantification of CPs was performed at the Chair of Organic Chemistry II (Prof. Sieber, TUM) in cooperation with Max Koch. After expression under defined growth conditions, extraction of CPs was performed with *n*-butanol. In order to minimize bias resulting from CP degradation during sample preparation, the cell lysates were already supplemented with ^{13}C-labeled internal standards. Any bias due to differences in cell growth was overcome by measuring the extracts from three individual cultures. Furthermore, three independent extractions were prepared from each cultivation and every extraction was measured three times by LC-ESI-MS/MS generating a total of 27 independent values for each CP (**Figure 25**).

Figure 25 *Scheme of CP quantification*



The amount of cell pellet and injection volume was dependent on CP concentration and cell extract composition, and was individually adjusted for each peptide (**Table 15** and **Table 24**).

In order to develop a method for CP quantification, LC-ESI-MS/MS analysis was performed with all synthesized CPs. Daughter ion peaks of reference peptides were utilized for absolute quantification (see sections **3.1**, **5.2.3** and **Appendix**). In the *Ssp*-CAFHPQ cell extracts, another compound eluted at a similar retention time as the internal standard and thus initially interfered with quantification. The method was therefore adjusted by a longer gradient in order to achieve sufficient peak separation. All other CPs could be quantified by using a shorter, standard gradient.

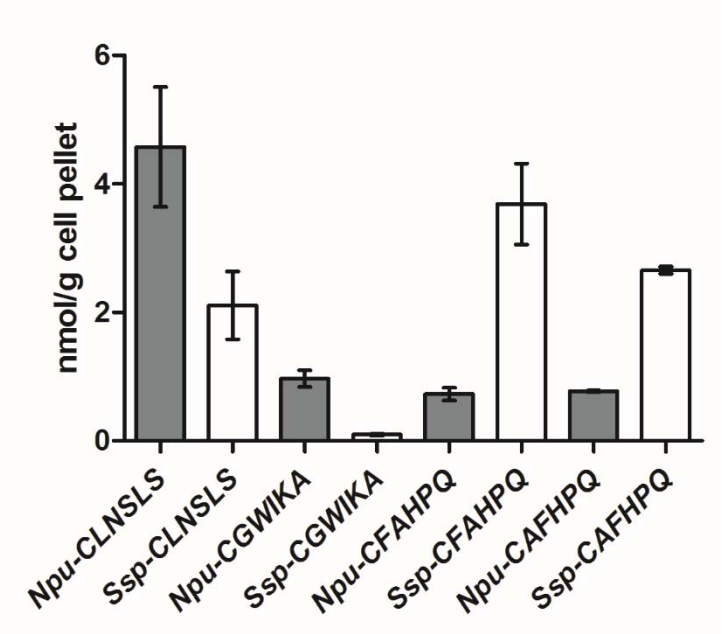
The absolute concentrations of CPs in cell extracts were calculated from obtained ratios (CP:internal standard) and linear equations of the calibration curves (**Appendix**). The quantified amounts of each CP from independent *E. coli* cultures

are listed in **Table 2**. As the culture batch only had a moderate influence on CP yields, all extraction values were plotted in **Figure 26**.

Table 2 *Quantified CP yields from cell extracts by LC-ESI-MS/MS. Every CP was cultivated three times. Triplicates of peptide extractions were also measured three times by LC-ESI-MS/MS.*

	µg/g cell pellet	**nmol/g cell pellet**
***Npu*-CLNSLS**	2.48±0.57	4.02±0.93
	3.54±0.67	5.74±1.08
	2.45±0.23	3.97±0.38
***Ssp*-CLNSLS**	0.95±0.02	1.53±0.20
	1.38±0.61	2.22±0.99
	1.58±0.19	2.56±0.31
***Npu*-CGWIKA**	0.56±0.13	0.86±0.20
	0.74±0.08	1.12±0.12
	0.61±0.04	0.93±0.06
***Ssp*-CGWIKA**	0.07±0.01	0.10±0.02
	0.05±0.01	0.08±0.01
	0.07±0.01	0.11±0.02
***Npu*-CFAHPQ**	0.41±0.03	0.60±0.05
	0.51±0.01	0.75±0.02
	0.57±0.01	0.83±0.01
***Ssp*-CFAHPQ**	1.21±0.04	1.77±0.03
	2.31±0.14	3.38±0.10
	2.1±50.18	3.14±0.13
***Npu*-CAFHPQ**	0.49±0.05	0.71±0.02
	0.53±0.06	0.77±0.02
	0.57±0.13	0.83±0.05
***Ssp*-CAFHPQ**	1.75±0.35	2.56±0.36
	1.74±0.28	2.54±0.29
	1.96±0.44	2.86±0.45

Figure 26 *Calculated amounts of CP synthesis in vivo of cyclo-CLNSLS, cyclo-CGWIKA, cyclo-CFAHPQ and cyclo CAFHPQ by utilization of Ssp- and Npu DnaE inteins.*



The CP yields varied between 0.08±0.01 nmol per gram cell pellet (*Ssp*-CGWIKA) and 5.74±1.08 nmol per gram cell pellet (*Npu*-CLNSLS). For CLNSLS and CGWIKA the splicing of *Npu* inteins resulted in a 10 fold higher yield of CGWIKA and a 2 fold higher yield of CLNSLS. In contrast, *Ssp*-CFAHPQ and *Ssp*-CAFHPQ yields were more than 3 fold higher than those of *Npu*-CPs. It was expected that *Npu*-CFAHPQ exhibits the best yields among all inteins quantified as it combines the potent *Npu* intein with the favored peptide sequence. [99, 100] However, *Npu*-CFAHPQ exhibited the second lowest CP yield with an average of 0.72 nmol per gram cell pellet. The quantification therefore revealed that *Npu* inteins do not necessarily exhibit superior splicing yields when SICLOPPS is used to cyclize these hexameric peptides.

Previous investigations on split inteins showed that higher splicing yields were observed when *Npu* DnaE was utilized instead of *Ssp* DnaE. Furthermore, *Npu* DnaE evolved to be more tolerant towards the extein sequence. These results were obtained by splicing linear fragments or by cyclizing larger peptide sequences. [99, 100, 119, 120]

Thus previous findings cannot be compared with the results that were obtained in this approach. It has to be considered that I_C and I_N are rearranged in SICLOPPS in order to create cyclic extein products. Furthermore, a six amino acid short extein sequence has to be bent strongly for a head-to-tail cyclization. Beyond that, a hexameric peptide is limited in alternative conformations due to its rigidity. Therefore the whole CP's sequence gains a strong influence on cyclization. One of the extein's amino acids could, for example, interrupt the cyclization process due to its side chain. A kink in the extein sequence which is produced by Pro could also decrease CP yields. This might explain the low splicing yields for *Npu*-CAFHPQ and *Npu*-CFAHPQ, in both of which Pro is located close to the splicing junction.

All key catalytic residues within *Npu* and *Ssp* DnaE are conserved and both inteins only differ in their non-catalytic amino acids. Nevertheless, these amino acids contribute to slight differences in the protein's three dimensional structure. *Ssp* DnaE may be able to cyclize a CP sequence for which cyclization is disfavored by *Npu* DnaE and *vice versa*. For this reason it is not possible to determine either *Npu* or *Ssp* DnaE to be more valuable for the generation of a hexameric CP library. It would be of interest to see how these differences in the three dimensional structure affect the cyclization of small CPs. Crystal structures of degenerated DnaE inteins bearing a hexameric CP sequence could shed a light on the question which intein is preferable for small CP libraries and which extein sequences are more favored by which split intein.

4. CONCLUSIONS AND OUTLOOK

In this thesis an *in vivo* selection system was established which allows the identification of *ST* AdoCbl RS targeting compounds. The selection system was subsequently used to screen two hexameric CP libraries. Moreover, *in vivo* CP library generation was investigated by direct quantification of CPs from *E. coli* cell extracts using LC-ESI-MS/MS. Results of this work serve as a basis for further research on the AdoCbl RS and the identification of biologically active CPs *in vivo*.

In the established selection system, the *ST* AdoCbl RS successfully regulates the reporter genes *lacZ* and *sacB.* Further investigations on the RS sequence proved that 0-210 bp of the *btuB* coding region are needed to ensure RS function. In addition, a high-throughput screening was performed with the *ST AdoCbl RS-lacZ-sacB reporter construct* and two hexameric CP libraries. *SacB* was shown to be a suitable reporter gene, minimizing the amount of potential positive hits by negative selection, and *lacZ* was successfully used to quantify inhibition of gene expression in β-galactosidase assays. However, no bioactive CP has been identified during the screening process. A single negative selection marker in the reporter system resulted in a large number of potential positive candidates after the first selection round. All potential positive colonies turned out to be the consequence of mutations in the *sacB* gene. It was furthermore problematic that anhydrotetracycline had an adverse effect on the growth of *E. coli* cells and that arabinose may have been metabolized by levansucrase.

In order to overcome the problem that variations of the plasmid copy number might disfavor the identification of a bioactive compound, the reporter construct was integrated into the genome of *E. coli* JM109 cells. After successful integration, reporter gene expression was reduced to the detection limit. Nevertheless, further studies on genome integration should be pursued in order to generate a more stable platform for the screening process.

Further improvements on the reporter system should involve the introduction of additional reporter genes that influence the growth of *E. coli*. As the number of

negative selection markers is limited, the reporter system should be changed in order to allow the use of a positive selection marker. The RS could, for example, control a gene that encodes a repressor of the promoter upstream of the reporter genes. This method would enlarge the choice of reporter genes to antibiotic resistances or genes encoding fluorescing proteins. Indirect inhibition of genes like *gfp* would allow high-throughput screening methods on a plate reader. This way, a faster and more reliable detection of bioactive compounds would be achieved in addition.

Another approach to facilitate CP screening procedures was the optimization of the CP library. For this purpose, the CP formation of the split inteins *Npu* and *Ssp* DnaE was compared. Both inteins were shown to generate hexameric CPs *in vivo.* Temperature dependent intein expression revealed that the splicing mechanism occurred faster for *Npu* DnaE. The application of *Npu* DnaE may be beneficial when CP libraries are used at lower temperatures. Furthermore the influence of the CP's target sequence on *Ssp*- and *Npu*-SICLOPPS was investigated by quantification of four CPs from cell extracts. Thus, the peptides cyclo-CLNSLS, cyclo-CGWIKA, cyclo-CFAHPQ and cyclo-CAFHPQ and their ^{13}C labeled analogues were successfully synthesized on solid phase. The synthesized peptides served as reference molecules and internal standards for LC-ESI-MS/MS analysis. Absolute quantification revealed CP yields between 0.1 and 5.7 nmol per gram of cell pellet. Splicing yields were shown to be strongly dependent on the target sequence and differences in the tolerance for this sequence were identified for both inteins. These findings do not show one of the two investigated split inteins to be more favorable when SICLOPPS based hexameric CP libraries are used. In order to elucidate the molecular basis for the observed sequence dependency, structures of the *Ssp*- and *Npu* DnaE pre-splicing complexes bearing hexameric peptide sequences should be determined.

5. MATERIALS AND METHODS

5.1 MATERIALS

5.1.1 INSTRUMENTS

Balances

Precision Balance BP3100P, Sartorius (Göttingen, GER)

Analytical Balance TE124S, Sartorius (Göttingen, GER)

Centrifuges

sigma® 1-14K rotor 12084, SIGMA Laborzentrifugen (Osterode am Harz, GER)

sigma® 6-16K rotor 12500, SIGMA Laborzentrifugen (Osterode am Harz, GER)

sigma® 8K rotor 11805, SIGMA Laborzentrifugen (Osterode am Harz, GER)

sigma® 3-30K rotor 12150, SIGMA Laborzentrifugen (Osterode am Harz, GER)

Speedvac, centrifuge RVC 2-25 CO plus, Christ (Osterode am Harz, GER)

Electroporation

2 mm electroporation cuvette, PEQLAB (Erlangen, GER)

1 mm electroporation cuvette, PEQLAB (Erlangen, GER)

MicroPulser™, BioRad (München, GER)

Electrophoresis

Agarose gels, PEQLAB (Erlangen, GER)

peqPOWER 300 V, PEQLAB (Erlangen, GER)

Gel documentation system G:BOX, Syngene (Cambridge, US)

SDS page SE260 Mighty Small™ II Deluxe Mini, Hoefer™ (Holliston, US)

Electrophoresis Power Supply EPS 600, Pharmacia Biotech (Uppsala, SE)

Filter Devices

Sterile Syringe Filter 0.2 µm cellulose Acetate Membrane, VWR (Darmstadt, GER)

Centrifugal Filter Modified Nylon 0.45 µm, 500 µL, VWR (Darmstadt, GER)

HPLC and Mass Spectrometry

Binary pump system 1525, Waters (Eschborn, GER)

Column selector, Waters (Eschborn, GER)

Flex Inject Loop, Waters (Eschborn, GER)

Fraction collector II, Waters (Eschborn, GER)

HPLC system Ultimate 3000, Dionex (Idsetin, GER)

LCQ Fleet MS, Thermo Fisher (Schwerte, GER)

Xbridge BEH C18, 10 mm x 250 mm, Waters (Eschborn, GER)

LTQ FT-ICR, Thermo Finnigan (Schwerte, GER)

Dionex Ultimate 3000 separation module, Thermo Scientific (Sunnyvale, US)

XBridge C18 column, 3.5 µm, 4.6 x 100 mm, Waters (Eschborn, GER)

Peptide Synthesis

MR Hei-Standard Magentic stirrer, Heidolph (Schwabach, GER)

Alpha 2-4 LD plus lyophilisator, Christ (Osterode, GER)

Peptide synthesizer PS3, Protein Technologies (Tucson, US)

Rotavapor R215, Büchi (Essen, GER)

Additional instruments and materials

Thermomixer comfort, Eppendorf (Hamburg, GER)

Thermal Cycler T100™, BioRad (Hercules, US)

Vortex Genie II, Scientific Industries (New York, US)

pH 720 pH-Meter WTW, InoLab (Weilheim, GER)

Incubator, Binder (Tuttlingen, GER)

Benchtop UV Transilluminator, UVP (Upland, US)

Water purification device, TKA (Niederelbert, GER)

Branson Digital Sonifier® 250 G, Heinemann (Schwabisch Gmund, GER)

Varioklav 2 Classic 500+200 EP-Z, HP Medizintechnik (Oberschleißheim, GER)

Shaking Incubator Multitron Standard, INFORS (Bottmingen, CH)

5.1.2 Isolation and Preparation Systems

Table 3 *Isolation and preparation systems*

System	Application	Manufacturer
peqGOLD Cycle-Pure Kit	DNA purification	PEQLAB (Erlangen, GER)
peqGOLD Gel-Extraction Kit	DNA extraction from agarose gels	PEQLAB (Erlangen, GER)
peqGOLD Plasmid Miniprep Kit II	Plasmid DNA isolation from *E. coli*	PEQLAB (Erlangen, GER)
Ni-NTA Spin Columns	Purification of His_6-tagged proteins	Qiagen (Hilden, GER)

5.1.3 Chemicals

Unless stated otherwise, all chemicals were obtained from the following companies:

AppliChem (Darmstadt, GER), Fluka (Neu-Ulm, GER), Merck (Darmstadt, GER), Serva (Heidelberg, GER), Sigma-Aldrich (Steinheim, GER), Roth (Karlsruhe, GER), and VWR (Darmstadt, GER).

5.1.4 MEDIA

All media used for cultivation of bacteria are listed in **Table 4**. Specifications refer to 1 L cultures.

Table 4 *Media*

Medium	Composition
LB medium	10 g pancreatic peptone, 5 g yeast extract, 5 g NaCl in dH_2O
TB medium	12 g pancreatic peptone, 24 g yeast extract in 900 mL dH_2O, 2.31 g KH_2PO_4, 12.54 g K_2HPO_4 in 100 mL dH_2O
SOC medium	20 g pancreatic peptone, 5 g yeast extract, 584.4 mg NaCl, 186.4 mg KCl, 952.1 mg $MgSO_4$, 360.3 mg glucose in dH_2O
Glycerol SOC medium	20 g pancreatic peptone, 5 g yeast extract, 584.4 mg NaCl, 186.4 mg KCl, 952.1 mg $MgSO_4$, 20 mM glycerol in dH_2O
SOB medium	20 g pancreatic peptone, 5 g yeast extract, 10 mM $MgSO_4$, 10 mM NaCl, 10 mM $MgCl_2$, 2 mM KCl in dH_2O
M9 minimal medium	1x M9 salts, 0.1% (v/v) glycerol, 1 mM thiamine, 2 mM $MgSO_4$, 100 µM $CaCl_2$ 10x M9 salts: 477 mM Na_2HPO_4, 220 mM KH_2PO_4, 86 mM NaCl, 187 mM NH_4Cl
MOPS EZ rich defined medium	1x MOPS mixture, 1x Supplement EZ, 1x ACGU, 1.32 mM K_2HPO_4, 0.2% glucose, purchased from Teknova, (Hollister, US)
NM minimal medium	1x M9 salts, 0.4% (w/v) glycerol, 0.1 mM thiamine, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 10 µM $ZnSO_4$, 200 µM Uracil, 1 mg/L Histidine , 1x amino acid mix Amino acid mix (in 100 mL dH_2O each), consisting of Solution I (200x): 0.99 g Phe, 1.1 g Lys, 2.5 g Arg Solution II (200x): 0.2 g Gly, 0.7 g Val, 0.84 g Ala, 0.41 g Trp Solution III (200x): 0.71 g Thr, 8.4 g Ser, 4.6 g Pro, 0.96 g Asn Solution IV (200x): 1.04 g Asp, 14.6g Gln, 3% (v/v) HCl Solution V (200x): 18.7 g K-Glu, 0.36 g Tyr, 0.4 g NaOH Solution VI (200x): 0.79 g Ile, 0.77 g Leu 1.72g Adenine in 500 mL dH_2O

For preparation of agar plates, $14\,\frac{g}{L}$ agar was added before autoclaving. Stock solutions of antibiotics were sterile filtered and added as appropriate (**Table 5**).

Table 5 *Antibiotics*

Antibiotic	Final concentration in media
Ampicillin	$100\,\frac{mg}{L}$
Chloramphenicol	$34\,\frac{mg}{L}$
Kanamycin	$50\,\frac{mg}{L}$
Tetracyclin	$10\,\frac{mg}{L}$
Spectinomycin	$100\,\frac{mg}{L}$

5.1.5 ENZYMES

Enzymatic reactions were performed according to the manufacturers' recommendation. A list of enzymes used in this thesis can be found below.

Table 6 *Enzymes*

Enzyme	Source
Antarctic Phosphatase	New England Biolabs (Frankfurt, GER)
DNA Polymerase I Large (Klenow) Fragment	New England Biolabs (Frankfurt, GER)
GoTaq DNA Polymerase	Promega (Madison, US)
Pfu Turbo Polymerase	Agilent Technologies (Santa Clara, US)
Phusion Polymerase	Finnzymes (Vantaa, FIN)
Restriction enzymes	New England Biolabs (Frankfurt, GER)
T4 DNA Ligase	Life Technologies (Darmstadt, GER)
Taq DNA Polymerase	New England Biolabs (Frankfurt, GER)

5.1.6 E. COLI STRAINS

Table 7 *E. coli strains*

Strain	Genotype	Source/Reference
***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 ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10)*Tet^S *endA1 thi-1 Δ(hsdS-mcrB)5*	New England Biolabs (Frankfurt, GER)
***E. coli* XL1 blue**	endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 $proAB^+$ $lacI^q$ Δ(lacZ)M15] hsdR17(r_K^- m_K^+)	Agilent Technologies (Santa Clara, US)
***E.coli* US0ΔpyrF**	*F- ΔhisB463 Δ(gpt-proAB-arg-lac)XIII zaj::Tn10 ΔpyrF*	[121]

5.1.7 DNA AND PROTEIN STANDARDS

The standard used for agarose gels used in this thesis was the peqGOLD DNA Leiter Mix (PEQLAB Biotechnologie, Erlangen, GER). For SDS gels the Roti®-Mark Standard (Roth, Karlsruhe, GER) was used.

5.2 METHODS

5.2.1 MOLECULAR BIOLOGY METHODS

PCR

The PCR was used to amplify DNA. A typical PCR reaction mixture consisted of:

200 $\frac{ng}{\mu L}$ plasmid DNA or 0.5 – 1 $\frac{ng}{\mu L}$ genomic DNA

1 U polymerase

1 µM forward primer

1 µM reverse primer

10 µM dNTP's

in 1x polymerase buffer

For quantitative PCR, Phusion Polymerase was used, ensuring high fidelity and fast synthesis. All adjustments were performed as specified by the manufacturer. Melting temperatures (T_m) of the used primers were calculated by OligoCalc. [122] A typical PCR temperature scheme can be found below (**Table 8**).

Table 8 *Temperature scheme PCR*

PCR Step	Temperature	Time	
Initial denaturation	98°C	30 s	
Denaturation	98°C	15 s	35 cycles
Annealing	typically 5°C below T_m	30 s	
Elongation	72°C	variable	
Final elongation	72°C	5 min	
Cooling	4°C	∞	

All PCR products were analyzed with by agarose gel electrophoresis and purified using a PeqGOLD Cycle Pure Kit according to the instruction manual. Purified PCR products were stored at -20°C before further use.

Single colony PCR

One colony was used as a DNA template for a single colony PCR and the initial denaturation temperature was increased to 8 min. Furthermore, GoTaq DNA Polymerase was used as specified by the manufacturer.

SLIM PCR

Site-directed, Ligase-Independent Mutagenesis (SLIM) [123] can be used to insert, delete or substitute bases on a plasmid. Mutagenesis was performed in two reactions. One reaction contained a short forward primer (F_S) and a tailed reverse primer (R_T) with the desired mutation. The second reaction mix contained one tailed forward primer (F_T) with the desired mutation and a short reverse primer (R_S). Each reaction mix consisted of:

5 ng template DNA
0.5 U Phusion Polymerase
200 µM dNTP's
2.5 mM $MgSO_4$
100 mM betaine
10 pM forward primer
10 pM reverse primer
In 1x polymerase buffer

A temperature scheme of SLIM PCR can be found in **Table 9**.

Table 9 *Temperature scheme SLIM PCR*

PCR Step	Temperature	Time	
Initial denaturation	98°C	30 s	
Denaturation	98°C	15 s	25 cycles
Annealing	typically 5°C below T_m	20 s	
Elongation	72°C	variable	
Final elongation	72°C	10 min	
Cooling	4°C	∞	

The plasmid template was digested by adding 5 µL D-buffer containing 10 U *Dpn*I and incubated for 60 min at 37°C. Subsequently, the two reaction products were mixed, H-buffer was added (final concentration: 1x) and hybridization was performed under the conditions described in **Table 10** and **Table 11**. The product was transformed in electrocompetent *E. coli* cells.

Table 10 *Temperature scheme of hybridization for SLIM PCR*

PCR Step	Temperature	Time	
Initial denaturation	99°C	3 min	
Denaturation	65°C	5 min	3 cycles
Annealing	30°C	40 min	
Cooling	4°C	∞	

Table 11 *Buffers for SLIM PCR*

Name	Composition
D-buffer	20 mM Tris HCl pH = 8.0 20 mM $MgCl_2$ 5 mM DTT
H-buffer (5x)	125 mM Tris HCl pH = 9.0 100 mM EDTA pH = 8.0 750 mM NaCl

Isolation of plasmid DNA from *E. coli*

Isolation of plasmid DNA from *E. coli* cells was performed using the peqGOLD Plasmid Miniprep Kit II according to the manufacturer's instructions. Purified plasmids were stored at -20°C.

Restriction digest

All restriction digests were carried out according to specifications of the manufacturer. The restriction enzymes were used at a ratio of 2.5 U per µg of DNA. Unspecific restriction was minimized by keeping the amount of glycerol below 5%. A

preparative digest was carried out in a total volume of 50 µL and was incubated for 3 h at 37°C. Analytical digests were performed in a total volume of 10 µL and incubation for 1 h at 37°C.

Plasmid DNA dephosphorylation

After preparative digests of plasmids, Antarctic Phosphatase ($0.1\,\frac{U}{\mu L}$) and Antarctic Phosphatase buffer (final concentration: 1x) were added directly to the digest reaction mixture and the mixture was incubated for 30 min at 37°C. The inactivation of the phosphatase was carried out for 20 min at 65°C. Fragments were purified by agarose gel electrophoresis.

Formation of blunt ends

Blunt ends were created by the removal of 3' overhangs and fill-in of 5'overhangs after a restriction digest. Therefore 0.2 U per µg of DNA Polymerase I (Klenow) Fragment and 33 µM dNTP's were added directly to the restriction digest mixture and the mixture was incubated for 15 min at 25°C. Inactivation was achieved by addition EDTA to a final concentration of 10 µM. Fragments were purified by agarose gel electrophoresis.

Ligation

For the ligation of sticky ends, plasmid DNA and insert fragments were mixed in a molar ratio of 1:3. The ligation of blunt ends was performed in a molar ratio of 1:1. The concentration of DNA was determined by agarose gel electrophoresis. Ligation reactions were performed according to the manufacturer's instruction. A reaction mixture contained:

100 ng linearized vector
3x molar surplus of insert
10 U T4 DNA Ligase
1x T4 DNA Ligase reaction buffer

The reaction mixture was incubated overnight at 4°C and the ligase was heat inactivated for 20 min at 65°C before further use.

Sequence analysis

Analysis of DNA sequences was performed by GATC Biotech AG (Köln, GER).

Preparation of electrocompetent *E. coli* cells

500 mL of LB medium were inoculated with 5 mL of a fresh overnight culture and incubated at 37°C and 130 rpm. At an OD_{600} of 0.5-0.7 cells were transferred into a sterile ice cold beaker and cooled on ice for 20 min. Cells were pelleted by centrifugation for 20 min and 5,000 rpm at 4°C. After resuspending the cell pellet in 500 mL ice cold 10% (v/v) glycerol, the suspension was pelleted again by centrifugation for 20 min and 5,000 rpm at 4°C. This step was repeated twice, whereby cells were resuspended in 20 mL 10% (v/v) glycerol in the last step. Finally, cells were resuspended in 2 mL ice cold 10% (v/v) glycerol. 50-100 µL aliquots were flash frozen in liquid nitrogen and stored at -80°C.

The transformation efficacy was determined using a test plasmid according to the following equation:

$$Transformation\ efficacy\ [\frac{cfu}{\mu g}] = \frac{amount\ of\ colonies}{\mu g\ DNA\ \cdot dilution factor}$$

Transformation of *E. coli* by electroporation

1 µL of plasmid DNA (1-3 $\frac{ng}{\mu L}$) was mixed on ice with 50 µL competent *E. coli* cells and transferred into a 2 mm electroporation cuvette. For transformation of DNA from ligation reactions, 2 µL of the reaction mixture was used instead. After applying a pulse of 2,500 V cells were immediately resuspended in 600 µL SOC medium and incubated for 1 h at 37°C while shaking at 600 rpm. Subsequently, cells were plated on selective LB agar and incubated at 37°C overnight.

Agarose gel electrophoresis

Agarose gel electrophoresis was used for the analytical and preparative separation of DNA fragments. The concentration of gels varied between 0.5 and 1.5% (w/v) agarose in TAE Buffer, depending on the fragment size of the DNA to be separated. Each sample was mixed with DNA loading dye (final concentration: 1x) and loaded onto a gel along with 3 µL of a DNA standard (**Table 12**). Electrophoresis was carried out in 1x TAE buffer for 45 min at 200 V. Gels were stained for 20 min in a solution of 1 $\frac{mg}{L}$ ethidium bromide and the DNA was visualized under UV light (365 nm). When performing a preparative agarose gel, bands of interest were cut out of the gel and purified using a peqGOLD Gel Extraction Kit according to the manufacturer's instructions. Purified DNA was stored at -20°C before further use.

Table 12 *Buffers for agarose gel electrophoresis*

Name	Composition
TAE buffer	0.4 M Tris HCl pH = 8.2 100 mM EDTA
DNA loading dye (6x)	10 mM Tris HCl pH = 8.2 1 mM EDTA 50% (w/v) glycerol 0.25% (w/v) xylene cyanol 0.25% (w/v) bromphenol blue

Cloning of the *ST AdoCbl RS* reporter constructs

The *Control-btuB/sacB* plasmid and the *ST* AdoCbl RS DNA including the btuB leader sequence were kindly provided by Dr. Sabine Schneider. Dr. Veronika Flügel provided the *ura3* DNA and the *control-lacZ* plasmid. The reporter constructs were cloned according to the following strategies:

ST AdoCbl RS-lacZ

The *ST* AdoCbl RS DNA was amplified with the primer pair oST_VitB12RSfor/rev and was subsequently cloned downstream of the *glnS* promotor into the first multiple cloning site (MCS) of the *control-lacZ* plasmid using the restriction sites *Hin*dIII and *Bam*HI. Positive clones were verified by single colony PCR and DNA sequencing.

ST AdoCbl RS-lacZ-sacB

The *ST* AdoCbl RS DNA was amplified with the primer pair oST_VitB12RSfor/rev and cloned into the first MCS of c*ontrol-btuB/sacB* using the restriction sites *Hin*dIII and *Bam*HI. In the second MCS, the *ST* AdoCbl RS DNA was inserted by blunt end ligation *via* the restriction site *Sma*I. Positive clones were verified *via* single colony PCR and DNA sequencing.

ST AdoCbl RS-ura3-sacB

The construct was cloned by a substitution of *lacZ* with *ura3* in the first MCS of *ST AdoCbl RS-lacZ-sacB.* Multiple restriction sites of *Hin*dIII in the plasmid required a three fragment ligation. A *Hin*dIII/*Not*I-restricted *ura3* DNA sequence was thus ligated with a *Hin*dIII/*Nco*I-restricted *ST AdoCbl RS-lacZ-sacB* and a *Nco*I/*Not*I-restricted *ST AdoCbl RS-lacZ-sacB* fragment. Positive clones were verified by single colony PCR and DNA sequencing.

Genome integration of large DNA fragments

Integration of DNA fragments into the *nth-tppB* locus of the *E. coli* JM109 genome was performed according to the procedures published by Kuhlman and Cox. [118]. The plasmids pTKRED and pTKIP were kindly provided by Prof. Thomas E. Kuhlman from the University of Illinois, US. The DNA template of LP-tetA was purchased from GeneArt AG (Thermo Fisher Scientific Inc., Regensburg, GER) using the same sequence as described in literature instead of utilizing pTKS/CS.

In order to generate of the donor plasmid, pTKIP was digested with *Xba*I and *Xho*I restriction endonucleases. The whole *ST AdoCbl RS-lacZ-sacB reporter construct* was

excised with *Eco*NI and *Pac*I. Blunt ends on both digestion products were generated in the next step and the pTKIP backbone was dephosphorylated. Ligation of both fragments led to the pTKIP-reporter plasmid which was analyzed by single colony PCR and sequencing.

LP-*tetA* was amplified in a PCR using the conditions stated below.

0.1 $\frac{ng}{\mu L}$ template DNA

0.02 $\frac{U}{\mu L}$ polymerase

1 µM forward primer

1 µM reverse primer

200 µM dNTP's

10% DMSO

in 1x polymerase buffer

Table 13 *PCR temperature scheme for LP-tetA*

PCR Step	Temperature	Time	
Initial denaturation	98°C	30 s	
Denaturation	98°C	15 s	25 cycles
Annealing	60°C	30 s	
Elongation	72°C	30 s/kbp	
Final elongation	72°C	5 min	
Cooling	4°C	∞	

Electrocompetent *E. coli* JM109 cells carrying the pTKRED plasmid were prepared and grown in SOB medium supplemented with 0.5% (w/v) glucose, 2 mM IPTG and 100 $\frac{\mu g}{mL}$ spectinomycin. pTKRED electrocompetent cells were afterwards transformed by approximately 1 µg LP-*tetA* in a 1 mm electroporation cuvette. Cells were resuspended in 2 ml SOC medium supplemented with 2 mM IPTG. Recovery was performed at room temperature overnight. Cells were subsequently incubated overnight at 30°C on LB agar supplemented with 10 $\frac{\mu g}{mL}$ tetracycline, 100 $\frac{\mu g}{mL}$ spectinomycin and 0.5% (w/v) glucose. Successful integration of LP-*tetA* was confirmed by single colony PCR and DNA sequencing.

Electrocompetent *E.coli* JM109 LP-tetA cells were prepared and transformed with the pTKIP-reporter plasmid. Cells were plated on LB agar supplemented with 10 $\frac{\mu g}{mL}$ tetracycline, 100 $\frac{\mu g}{mL}$ ampicillin, 100 $\frac{\mu g}{mL}$ spectinomycin and 0.5 % (w/v) glucose and grown overnight at 30°C.

Genome integration was performed by inoculating a single colony in 5 ml of MOPS EZ rich defined medium supplemented with 0.5% (w/v) glycerol, 2 mM IPTG, and 0.2% (w/v) L-arabinose. After incubation for 1 h at 37°C, the temperature was reduced to 30°C and cells were shaken for 30 min before 100 $\frac{\mu g}{mL}$ spectinomycin was added. The cells were grown overnight at 30°C and diluted 100 fold before being plated on LB agar plates supplemented with 100 $\frac{\mu g}{mL}$ spectinomycin and 0.004 % (w/v) X-gal. Incubation at 30°C led to blue colonies, which were picked and plated on LB agar plates that contained 10 $\frac{\mu g}{mL}$ tetracycline or 100 $\frac{\mu g}{mL}$ ampicillin, in order 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.

Cloning of the *Npu*-CGWIKA construct

The *Npu*-CLNSLS plasmid served as template for the SLIM PCR that was used to generate the *Npu*-CGWIKA construct. The short primers were oSLIM-Npu-Int_Rs and oSLIM-Npu-Int_Fs and the tailed primers were oFt_Npu_CGWIKA and oRt_Npu_CGWIKA. Positive clones were verified by single colony PCR and DNA sequencing.

5.2.2 Biochemical Methods

Recombinant gene expression in *E. coli*

Medium was inoculated with a fresh overnight culture (dilution 1:100) and incubated at 37°C while shaking at 130 rpm. At an OD_{600} of 0.6-0.8, recombinant gene expression was induced with 0.02% (w/v) L-(+) arabinose and cultures were shaken for a further 18 h at 37°C. Cells were centrifuged for 40 min at 6,000 rpm and stored at -20°C.

Protein purification

A 0.2 g cell pellet was resuspended in 1 mL lysis buffer (**Table 14**) and sonicated (10%, 1 s/1 s on/off, 2x 30 s) on ice. The cell debris was removed by centrifugation (15 min, max. rpm, 4°C) and His_6-tag containing proteins were purified at 4°C using Ni-NTA spin columns (Quiagen, Hilden, GER) according to the manufacturers' protocol:

Equilibration:	600 µL dH_2O, 2 min, 2,900 rpm
	600 µL lysis buffer, 2 min, 2,900 rpm
Loading:	600 µL lysate, 5 min, 1,600 rpm, repeat step
Washing:	600 µL wash buffer, 2 min, 2,900 rpm, repeat step
Elution:	300 µL elution buffer, 2 min, 2,900 rpm, repeat step
Regeneration:	600 µL dH_2O, 2 min, 2,900 rpm, 600 µL 2% (w/v) SDS, 2 min, 2,900 rpm, store in 20% (v/v) ethanol at 4°C.

Table 14 *Buffer composition for protein purification*

Name	Composition
Lysis buffer	20 mM Tris HCl pH = 8.0 500 mM NaCl 2 mM β-mercaptoethanol 5 mM imidazol
Wash buffer	20 mM Tris HCl pH = 8.0 500 mM NaCl 2 mM β-mercaptoethanol 15 mM imidazol
Elution buffer	20 mM Tris HCl pH = 8.0 500 mM NaCl 2 mM β-mercaptoethanol 500 mM imidazol

n-Butanol cell extraction

The cell pellets (**Table 15**) were resuspended in extraction buffer (2 mL per gram of cell pellet, **Table 16**) and lysed by sonication at an amplitude of 10% (1 s/1 s on/off, 2x 45 s, 4°C). After removal of cell debris *via* centrifugation for 20 min at 21,000 rpm and 4°C, 7.5 µL ^{13}C-labeled internal standard ($0.2\,\frac{mg}{mL}$) was added directly to the lysate. Extraction was performed by mixing the same volume of *n*-butanol with the lysate and centrifugation of the samples (20 min at 21,000 rpm, 4°C). The organic phase was carefully removed and dried *in vacuo*. The cell extract was solved in 2x 750 µL 70% (v/v) methanol and reduced to a total volume of 60 µL. Samples were filtered (0.4 µm) prior to mass analysis.

Table 15 *Amount of cell pellets used for quantification*

	***Ssp* intein cell pellet [g]**	***Npu* intein cell pellet [g]**
Cyclo-CLNSLS	5	3
Cyclo-CGWIKA	10	6
Cyclo-CFAHPQ	1.5	3
Cyclo-CAFHPQ	7	2.5

Table 16 *Composition of extraction buffer*

Name	**Composition**
Extraction buffer	20 mM Tris HCl pH = 7.8 500 mM NaCl 1 mM TCEP

SDS-PAGE

SDS polyacrylamide gel electrophoresis (SDS-PAGE), used to analyze proteins according to their approximate size, was carried out per *Lämmli et al.* [124] Protein samples were mixed with 2x SDS loading buffer (**Table 17**), incubated for 5 min at 95°C and loaded onto a 15% SDS gel (**Table 18**) along with a molecular weight

marker. Electrophoresis was performed for 40 min at 30 mA in an electrophoresis buffer (**Table 17**). Protein bands were analyzed after gels were stained in coomassie solution for 45 min followed by destaining in 10% (v/v) acetic acid for at least 1 h (**Table 17**). Bands of interest were subsequently analyzed by in-gel digestion and mass spectrometry by the Chair of Biotechnology (Prof. Buchner, TU München, GER).

Table 17 *Buffers for SDS-PAGE*

Name	Composition
Electrophoresis buffer	25 mM Tris HCl pH = 8.2 – 8.3 192 mM glycine 0.1% (w/v) SDS
Separating gel buffer	1.5 M Tris HCl pH = 8.8 0.4% (w/v) SDS
Stacking gel buffer	1.5 M Tris HCl pH = 6.8 0.4% (w/v) SDS
SDS loading buffer (2x)	60 mM Tris HCl pH = 6.8 30% (v/v) glycerol 10% (w/v) sucrose 5% (w/v) SDS 0.02% (w/v) bromphenol blue 3% (v/v) β-mercaptoethanol
Acrylamide solution	39% (w/v) acrylamide 1.2% (w/v) bisacrylamide
Coomassie solution	0.05% (w/v) coomassie brilliant blue R250 25% (v/v) isopropanol 10% (v/v) acetic acid

Table 18 *Composition of a 15% SDS Gel*

Component	Separating gel (15%)	Stacking gel (4%)
40% (w/v) acrylamide 29:1	4.5 mL	625 µL
Separating gel buffer	3.0 mL	-
Stacking gel buffer	-	1.25 mL
ddH_2O	4.4 mL	3.1 mL
10% (v/v) APS	120 µL	50 µL
1% (v/v) TEMED	12 µL	5 µL

β-Galactosidase assay

β-Galactosidase assays performed in this work were carried out according to a modified protocol by Zhang and Bremer. [125] Selective medium was inoculated with a fresh overnight culture (dilution 1:100) and incubated at 37°C while shaking at 130 rpm. At an OD_{600} of 0.3-0.7, 50 µL cell cultures were mixed with 50 µL permeabilization solution (**Table 19**). Samples were incubated at 30°C for 30 min and mixed with 600 µL substrate solution, warmed to 30°C, which initiated the time recoding for the experiment. After a sufficiently yellow color had developed, 700 µL of stop solution was added and the time was noted. All samples were spun down for 10 min at 13,000 rpm and absorption was measured at 420 nm. Each sample was measured in triplicates and non-inoculated medium was used as a reference. Miller Units were calculated using the following equation:

$$Miller\ Units = \frac{A_{420}}{Volume\ [mL] \cdot reactiontime\ [min] \cdot A_{600}} \cdot 1000$$

Table 19 *Solutions for β-galactosidase assay*

Name	Composition
Permeabilization solution	100 mM Na_2HPO_4 20 mM KCl 2 mM $MgSO_4$ 0.8 $\frac{mg}{mL}$ CTAB 0.4 $\frac{mg}{mL}$ sodium deoxycholate 5.4 $\frac{\mu L}{mL}$ β-mercaptoethanol
Substrate solution	60 mM Na_2HPO_4 40 mM NaH_2PO_4 1 $\frac{mg}{mL}$ ONPG 2.7 $\frac{\mu L}{mL}$ β-mercaptoethanol
Stop solution	1 M Na_2CO_3

Sucrose cell growth assay

A sucrose cell growth assay was performed to monitor the levansucrase expression of *E. coli* cells containing the *sacB* gene. For this purpose, a fresh overnight culture of *E. coli* cells was diluted 1:100 in 7 mL selective M9 minimal medium with or without addition of 1% (w/v) sucrose. Cells were incubated at 37°C while shaking at 130 rpm and growth was followed by measuring the OD_{600} regularly.

5-FOA cell growth assay

A 5-FOA cell growth assay was performed to monitor orotidine 5'-phosphate decarboxylase expression of *pyrF* deficient *E. coli* cells containing the *ura3* gene. For this purpose, a fresh overnight culture of *E. coli* cells was diluted 1:100 in 7 mL selective NM minimal medium with addition of 0 – 20 mM 5-FOA. Cells were incubated at 37°C while shaking at 130 rpm and growth was followed by measuring OD_{600} regularly.

Serial dilutions

Fresh overnight cultures were diluted to an OD_{600} of 0.7. 5 µL samples were mixed with 45 µL of M9 minimal medium. This step was repeated until a dilution of $1{:}10^7$ was reached. 5 µL of the $1{:}10^3$, $1{:}10^4$, $1{:}10^5$, $1{:}10^6$ and $1{:}10^7$ dilutions were transferred onto M9 minimal medium agar plates supplemented with relevant additives.

SICLOPPS library screening process

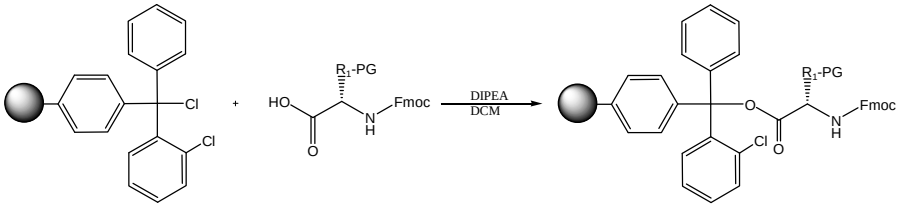
1 µL of the plasmid mixture was transformed into 50 µL of electrocompetent *E. coli* JM109 (*E.coli* US0ΔpyrF) cells containing the *ST AdoCbl RS-lacZ-sacB reporter construct*. Immediately after the electro pulse, cells were allowed to recover in pre-warmed glycerol-SOC medium for 1 h at 37°C. Subsequently, 0.02% (w/v) arabinose ($200\,\frac{\mu g}{mL}$ anhydrotetracycline) was added and cells were shaken for 2 h at 37°C before they were pelleted (5,000 g, 5min) and resuspended in M9 minimal medium (NM medium). Afterwards, the cells were plated onto four selective minimal medium agar plates (diameter 13.5 cm) containing 1% sucrose and $0.04\,\frac{mg}{mL}$ X-Gal. After incubation for 2 days at 37°C, white colonies were picked and used to inoculate 1 mL of LB medium which was incubated overnight at 37°C while shaking. Dilution series of $1{:}10^3$, $1{:}10^4$, $1{:}10^5$ and $1{:}10^6$ were prepared on selective minimal medium plates with and without 0.02% (w/v) arabinose ($200\,\frac{\mu g}{mL}$ anhydrotetracycline). Colonies that showed a growth advantage, when 0.02% (w/v) arabinose ($200\,\frac{\mu g}{mL}$ anhydrotetracycline) was added were further investigated by the sucrose cell growth assay and β-galactosidase assay.

5.2.3 CHEMICAL METHODS

Chemical synthesis of CPs

CP synthesis was carried out utilizing the following strategy:

Coupling



A syringe equipped with a PE Filter was weighed and filled with Trt-Cl resin (loading 1.4 $\frac{\text{mmol}}{g}$, 200 mg, 0.28 mmol, 1 eq). The first amino acid (0.34 mmol, 1.2 eq relative to resin) was solved in 1.6 mL DCM_{dry} and mixed with 231 µL DIPEA (1.36 mmol, 4 eq relative to AA). This solution was sucked directly into the syringe. Remaining residues were taken up with 0.4 mL DCM_{dry} and also sucked into the syringe. The reaction mixture was shaken for 2 h at RT. Subsequently MeOH (40 µL, 1 mmol, 0.2 $\frac{\text{mL}}{g}$ resin) and DIPEA (34 µL, 0.2 mmol, 0.2 eq relative to MeOH) were pre-mixed and sucked into the syringe directly to cap unreacted resin. After shaking at RT for 20 min, the solution was drained 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 finally dried *in vacuo*. The loading yield was calculated according to the following equation:

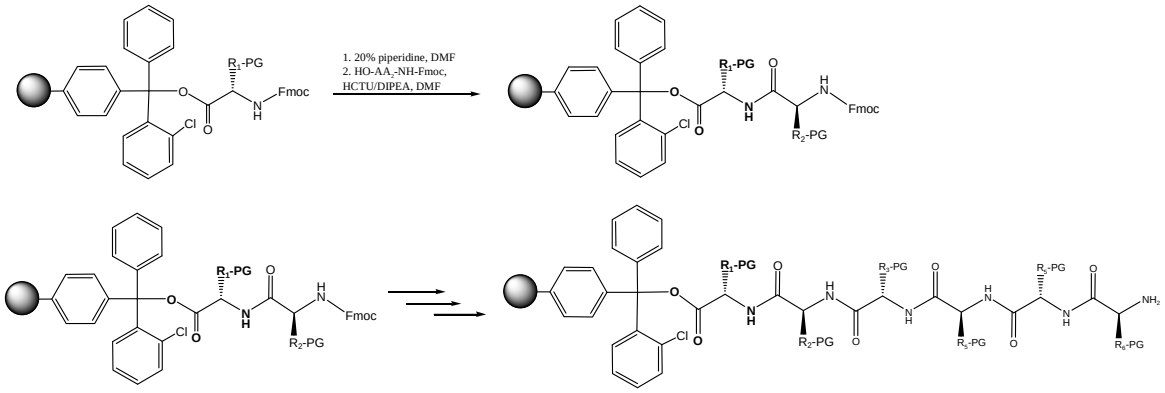
$$c\left[\frac{mmol}{g}\right] = \frac{m(loaded) - m(unloaded)}{[MW(AA) - 36.461\,\frac{g}{mol}] \cdot m(loaded)}$$

m(loaded) = loaded resin [g]

m(unloaded) = unloaded resin [g]

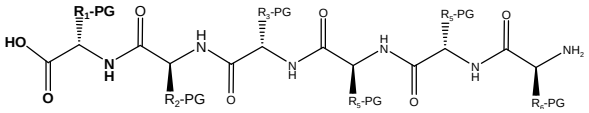
MW(AA) = molecular weight of the immobilized AA [$\frac{g}{mol}$ *]*

Solid phase peptide synthesis



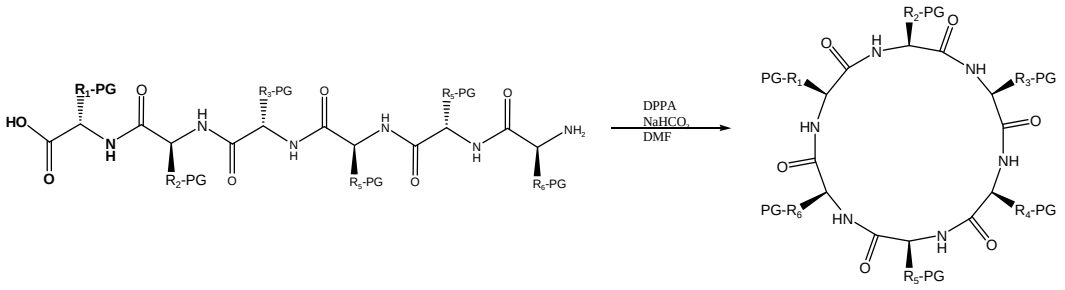
The amino acid coupling steps were carried out on a PS3 peptide synthesizer. Synthesis was performed on a 0.26 mmolar scale. In a first step the resin was swollen in DMF for one hour and deprotected with 20% (v/v) piperidine in DMF for 10 min. The coupling was performed with a fresh solution of Fmoc-protected amino acid (2 eq relative to resin) and HCTU (2 eq relative to resin) in 20% (v/v) N-methyl morpholine for 20 min. After the resin was washed with DMF, the procedure was repeated until all amino acids were coupled. As a final step the N-terminal Fmoc protection group was removed.

Resin cleavage



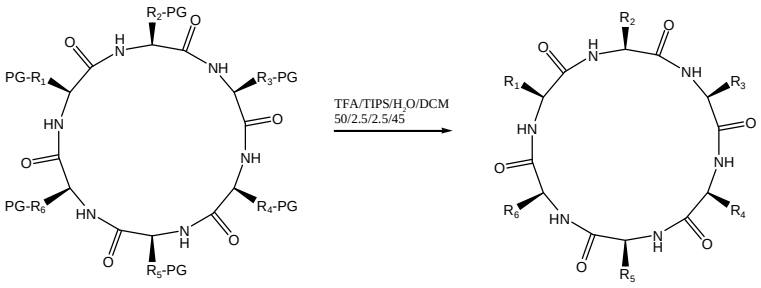
The resin was filled in a syringe equipped with a PE filter and treated with 3 mL 20% (v/v) HFIP in DCM for 15 min while gently shaking at RT. The procedure was repeated three times to cleave the peptide from the resin. Solvent was removed *in vacuo* and residues were co-evaporated twice with toluene.

Backbone cyclization



Cyclization was performed in a round-bottomed flask by adding 50 mL DMF, DPPA (3 eq) and $NaHCO_3$ (5 eq) to the peptide. The reaction was stirred overnight at RT and the solvent was removed *in vacuo*. The CP was extracted in EtOAc/H_2O and the organic layer was dried over Na_2SO_4. The solvent was evaporated *in vacuo*.

Side chain deprotection



A mixture of TFA/TIPS/H_2O/DCM (50:2.5:2.5:45, 5 mL total) was added for final deprotection and stirred for 1 h 40 min at RT. The product was precipitated in ice-cold diethyl ether (20 mL per mL deprotection solution) and pelleted *via* centrifugation for 2 min and 4,000 rpm. The peptide was washed with ice-cold diethyl ether and dried *in vacuo*. The yields can be found in **Table 20**.

Table 20 *Yields of CP synthesis*

		CLNSLS (CL*NSLS)	CGWIKA (CGWIKA*)	CFAHPQ (CFA*HPQ)	CAFHPQ (CA*FHPQ)
Loading		0.24 mmol [86%]	0.28 mmol [100%]	0.24 mmol [86%]	0.25 mmol [89%]
Resin cleavage	MW [$\frac{g}{mol}$]	1232.57 (1234.57)	1119.35 (1120.35)	1428.72 (1429.72)	1428.72 (1429.72)
	Yield	298 mg 0.24 mmol [100%]	319 mg 0.28 mmol [100%]	169 mg 0.12 mmol [49%]	183 mg 0.13 mmol [52%]
Cyclization	MW [$\frac{g}{mol}$]	1214.56 (1216.56)	1101.35 (1102.35)	1409.61 (1410.61)	1409.61 (1410.61)
	Yield	370 mg 0.3 mmol [126%]	366 mg 0.3 mmol [107%]	232 mg 0.16 mmol [133%]	230.2 mg 0.16 mmol [116%]
Deprotection	MW [$\frac{g}{mol}$]	617.58 (619.58)	658.33 (659.33)	683.28 (684.28)	683.28 (684.28)
	Yield (RP)	140 mg 0.21 mmol [70%]	153 mg 0.23 mmol [77%]	79 mg 0.12 mmol [72%]	90 mg 0.13 mmol [82%]

HPLC

Crude products were disssolved in 40-70% (v/v) ACN/H_2O supplemented with 1 mM TCEP and separated by preparative HPLC chromatography using a Xbridge BEH C18 column (10 mm x 250 mm, Waters, Eschborn, GER). The mobile phase for elution consisted of a gradient mixture of 0.1% (v/v) TFA in water (Buffer A, HPLC grade) and 0.1% (v/v) TFA in ACN (Buffer B, HPLC grade). Detailed purification settings can be found in **Table 21**. Fractions were analyzed by ESI-MS with positive

ion detection. Relevant fractions were pooled and the solvent was removed by lyophilization overnight.

Table 21 *HPLC purification of CPs*

Peptide	Flow [ml/min]	Time [min]	Gradient Buffer A [%]	Buffer B [%]	Elution
CLNSLS	4	0	15	85	14-17 min
		45	40	60	23-24%
		55	95	5	
		60	95	5	
CGWIKA	4	0	15	85	28 – 30 min
		50	40	60	29 - 30%
		60	95	5	
		65	95	5	
CFAHPQ	4	0	15	85	18 min
		45	95	5	47%
		50	95	5	
CAFHPQ	4	0	15	85	13-15 min
		45	40	60	22-23%
		55	95	5	
		60	95	5	

Mass spectrometry

Analysis of CP synthesis

Confirmation of successful synthesis was performed by reversed phase HPLC-ESI-MS analysis on a LCQ Fleet (Thermo Fisher) equipped with a Dionex Ultimate 3000 separation module. 5 µL of sample were injected and the analysis was performed either by loop measurements (flow = 0.35 $\frac{\mathrm{mL}}{min}$) or by a gradient of 5% - 95% buffer B in 9 min (Waters XBridge C18 column, 3.5 µm, 4.6 x 30 mm, flow = 1.1 $\frac{\mathrm{mL}}{min}$). The

mobile phase consisted of 0.1% (v/v) formic acid in H_2O (buffer A, HPLC-MS grade) and 0.1% (v/v) formic acid in ACN (buffer B, HPLC-MS grade). Ions were scanned in positive polarity mode.

LC-ESI-MS/MS analysis and quantification of CPs

Reversed-phase HPLC-ESI-MS/MS analysis was performed on a Thermo Finnigan LTQ FT-ICR equipped with a Dionex Ultimate 3000 separation module eluting on a Waters XBridge C18 column (3.5 µm, 4.6 x 100 mm, flow = 1.1 $\frac{\mathrm{mL}}{min}$). The column temperature was maintained at 30 °C. The mobile phase consisted of 0.1% (v/v) formic acid in H_2O (buffer A, HPLC-MS grade) and 0.1% (v/v) formic acid in ACN:water 90:10 (buffer B, HPLC-MS grade). The gradients are listed in **Table 22**. For the quantification of *Ssp*-CAFHPQ a long gradient was used, whereas in all other cases a short, standard gradient was applied.

Table 22 *RP-HPLC gradient settings for quantification*

Gradient short time [min]	**Gradient long time [min]**	**Buffer B [%]**
0 - 15	0 – 32	2 - 98
15 - 17	32 - 34	98
17 – 17.5	34 – 34.5	98 - 2
17.5 - 20	34.5 – 36.5	2

One tenth of the chromatographic eluent was introduced into the ion source. Ions were scanned in positive polarity mode. Analysis and absolute quantification was achieved using selected reaction monitoring (SRM) with the corresponding mass filter listed in **Table 23**. Tandem MS fragmentations were performed *via* collision-induced dissociation (CID) with normalized collision energy of 35 V. ESI source parameters were: spray voltage, 3.5 kV; capillary temperature, 60 °C; capillary voltage, 48 V; and tube lens, 60 V.

All data acquisition and processing were controlled by the LTQ software (Xcalibur 2.1). Mass calibration curves were obtained at different concentration ratios listed in **Table 24**. For absolute quantification, ^{13}C labeled internal standards were added to samples ($25\ \frac{ng}{\mu L}$) and areas of CPs and their ^{13}C labeled standards were determined. Absolute concentrations were calculated from obtained ratios and linear equations of calibration curves. Injection volumes of the calibration curves and the extracts are listed in **Table 24**.

Table 23 *Settings for the quantification of cyclo-CLNSLS, -CGWIKA, -CFAHPQ and -CAFHPQ (mass tolerance: 500 ppm)*

	Mass filter $[\frac{g}{mol}]$	**Corresponding mass** $[\frac{g}{mol}]$
CLNSLS	598.95 – 601.95	600.36
CL*NSL*S	600.95 – 603.95	602.36
CGWIKA	639.85 – 642.85	641.40
CGWIKA*	640.87 – 643.87	642.40
CFAHPQ	664.72 – 667.72	666.30
CFA*HPQ	665.73 – 668.73	667.30
CAFHPQ	664.72 – 667.72	666.30
CA*FHPQ	665.73 – 668.73	667.30

Table 24 *Injection volumes of calibration points and extracts*

	Calibration points	**Extracts**
CLNSLS	250 µM (0.8 µL), 125 µM (1.6 µL), 62.5 µM (3.2 µL), 31.3 µM (6.4 µL), 15.6 µM (10 µL), 7.81 µM (10 µL)	10 µL
CGWIKA	125 µM (1.25 µL), 62.5 µM (2.5 µL), 31.3 µM (5 µL), 15.6 µM (8 µL), 7.81 µM (10 µL), 3.91 µM (10 µL), 1.95 µM (10 µL), 0.98 µM (10 µL)	1.5 µL
CFAHPQ	125 µM (1.25 µL), 62.5 µM (2.5 µL), 31.3 µM (5 µL), 15.6 µM (8 µL), 7.81 µM (10 µL), 3.91 µM (10 µL), 1.95 µM (10 µL), 0.98 µM (10 µL)	10 µL (*Npu*-CFAHPQ) 3 µL (*Ssp*-CFAHPQ)
***Npu*-CAFHPQ**	250 µM (0.1 µL), 125 µM (0.5 µL), 62.5 µM (2.5 µL), 31.3 µM (5 µL), 15.6 µM (8 µL), 7.81 µM (10 µL)	10 µL
***Ssp*-CAFHPQ**	500 µM (1:1 dilution, 1 µL), 375 µM (1:1 dilution 0.25 µL), 250 µM (0.25 µL), 187.5 µM (0.5 µL), 125 µM (0.5 µL), 62.5 µM (1 µL), 31.3 µM (1 µL),	5 µL

6. ABBREVIATIONS

%	percent
Å	Ångstrom
A	absorbance
ACN	acetonitrile
AdoCbl	adenosyl cobalamin
APS	ammonium persulfate
AM	arithmetic mean
Boc	*tert*-butyloxycarbonyl
bp	base pair
BSA	bovine serum albumin
C	C-intein
°C	degree celsius
cfu	colony forming units
CP	cyclic peptide
CTAB	cetyltrimethylammonium bromide
c-di-GMP	cyclic-di-guanosine monophosphate
Da	Dalton
DCM	dichloromethane
dG	deoxyguanosine
DIPEA	*N*,*N*-diisopropylethylamine
DMF	dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DnaE	α-subunit of DNA polymerase III
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
HCTU	2-(6-chlor-1*H*-benzotriazol-1-yl)-*N*,*N*,*N'*,*N'*-tetramethyluronium hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HFIP	1,1,1,3,3,3-hexafluoro-2-propanol
His_6	hexahistidin
HPLC	high-performance liquid chromatography
I_C	C-terminal intein domain
I_N	N-terminal intein domain
k	kilo
l	liter
LB	lysogeny broth
LC	liquid chromatography
µ	micro
M	molar
m	milli
mA	milliampere
MCS	multiple cloning site
MeOH	methanol
min	minute
m/z	mass-to-charge ratio
Moco	molybdenum cofactor
mRNA	messenger RNA
MS	mass spectrometry
MU	Miller unit
MW	molecular weight
^{N}	N-intein
nm	nanometer

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
pNPU CLNSLS	plasmid containing *Npu* I_C-CLNSLS-I_N-His_6
pSSP CLNSLS	plasmid containing *Ssp* I_C-CLNSLS-I_N-His_6
$PreQ_1$	pre-queuosine-1
RBS	ribosome binding site
RNA	ribonucleic acid
RP	raw product
rpm	revolutions per minute
RS	riboswitch
RT	room temperature
RTHS	reverse two-hybrid system
6	standard deviation
SAH	*S*-adenosylhomocysteine
SAM	*S*-adenosylmethionine
SDS	sodium dodecyl sulfate
SRM	selected reaction monitoring
s	second
SICLOPPS	split-intein circular ligation of peptides and proteins
SLIM	site-directed ligase-independent mutagenesis
SOC	super optimal broth with catabolite repression
Ssp	*Synechocystis* sp. strain PCC6803
ST	*Salmonella typhimurium*
TAE	Tris-acetate-EDTA
TCEP	tris(2-carboxyethyl)phosphine
tetA	tetracycline resistance gene
THIC	thiamine C synthase
TPP	thiamine pyrophosphate

Trt-Cl	2-chlorotrityl chloride
TFA	trifluoroacetic acid
TIPS	triisopropylsilane
T_m	melting temperature
Tris	tris(hydroxymethyl)aminomethane
Tuco	tungsten cofactor
U	unit
UTR	untranslated region
V	volt
v/v	volume per volume
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside

7. REFERENCES

1. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii.* The Journal of experimental medicine, 1944. **79**(2): p. 137-58.

2. Rich, A. and C.F. Stevens, *Obituary: Francis Crick (1916-2004).* Nature, 2004. **430**(7002): p. 845-7.

3. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.

4. Sudarsan, N., J.E. Barrick, and R.R. Breaker, *Metabolite-binding RNA domains are present in the genes of eukaryotes.* RNA, 2003. **9**(6): p. 644-7.

5. Serganov, A. and E. Nudler, *A decade of riboswitches.* Cell, 2013. **152**(1-2): p. 17-24.

6. Andre, G., et al., *S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of Clostridium acetobutylicum.* Nucleic acids research, 2008. **36**(18): p. 5955-69.

7. Loh, E., et al., *A trans-acting riboswitch controls expression of the virulence regulator PrfA in Listeria monocytogenes.* Cell, 2009. **139**(4): p. 770-9.

8. Nahvi, A., et al., *Genetic control by a metabolite binding mRNA.* Chemistry & biology, 2002. **9**(9): p. 1043.

9. Nou, X. and R.J. Kadner, *Coupled changes in translation and transcription during cobalamin-dependent regulation of btuB expression in Escherichia coli.* Journal of bacteriology, 1998. **180**(24): p. 6719-28.

10. Ames, T.D. and R.R. Breaker, *Bacterial aptamers that selectively bind glutamine.* RNA biology, 2011. **8**(1): p. 82-9.

11. Baker, J.L., et al., *Widespread genetic switches and toxicity resistance proteins for fluoride.* Science, 2012. **335**(6065): p. 233-5.

12. Mandal, M. and R.R. Breaker, *Adenine riboswitches and gene activation by disruption of a transcription terminator.* Nature structural & molecular biology, 2004. **11**(1): p. 29-35.

13. Cromie, M.J., et al., *An RNA sensor for intracellular Mg(2+).* Cell, 2006. **125**(1): p. 71-84.

14. Dann, C.E., 3rd, et al., *Structure and mechanism of a metal-sensing regulatory RNA.* Cell, 2007. **130**(5): p. 878-92.

15. Mandal, M., et al., *A glycine-dependent riboswitch that uses cooperative binding to control gene expression.* Science, 2004. **306**(5694): p. 275-9.

16. Winkler, W.C., S. Cohen-Chalamish, and R.R. Breaker, *An mRNA structure that controls gene expression by binding FMN.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(25): p. 15908-13.

17. Watson, P.Y. and M.J. Fedor, *The ydaO motif is an ATP-sensing riboswitch in Bacillus subtilis.* Nature chemical biology, 2012. **8**(12): p. 963-5.

18. Sudarsan, N., et al., *An mRNA structure in bacteria that controls gene expression by binding lysine.* Genes & development, 2003. **17**(21): p. 2688-97.

19. Regulski, E.E., et al., *A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism.* Molecular microbiology, 2008. **68**(4): p. 918-32.

20. Nelson, J.W., et al., *Riboswitches in eubacteria sense the second messenger c-di-AMP.* Nature chemical biology, 2013. **9**(12): p. 834-9.

21. Wang, J.X., et al., *Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling.* Molecular cell, 2008. **29**(6): p. 691-702.

22. Kim, J.N., A. Roth, and R.R. Breaker, *Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(41): p. 16092-7.

23. Winkler, W.C., et al., *An mRNA structure that controls gene expression by binding S-adenosylmethionine.* Nature structural biology, 2003. **10**(9): p. 701-7.

24. Mandal, M., et al., *Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria.* Cell, 2003. **113**(5): p. 577-86.

25. Winkler, W., A. Nahvi, and R.R. Breaker, *Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.* Nature, 2002. **419**(6910): p. 952-6.

26. Meyer, M.M., et al., *Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria.* RNA, 2008. **14**(4): p. 685-95.

27. Roth, A., et al., *A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain.* Nature structural & molecular biology, 2007. **14**(4): p. 308-17.

28. Nahvi, A., J.E. Barrick, and R.R. Breaker, *Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes.* Nucleic acids research, 2004. **32**(1): p. 143-50.

29. McDaniel, B.A., et al., *Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(6): p. 3083-8.

30. Whitford, P.C., et al., *Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function.* Biophysical journal, 2009. **96**(2): p. L7-9.

31. Hollands, K., et al., *Riboswitch control of Rho-dependent transcription termination.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(14): p. 5376-81.

32. Collins, J.A., et al., *Mechanism of mRNA destabilization by the glmS ribozyme.* Genes & development, 2007. **21**(24): p. 3356-68.

33. Klein, D.J. and A.R. Ferre-D'Amare, *Structural basis of glmS ribozyme activation by glucosamine-6-phosphate.* Science, 2006. **313**(5794): p. 1752-6.

34. Barrick, J.E., et al., *New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(17): p. 6421-6.

35. Caron, M.P., et al., *Dual-acting riboswitch control of translation initiation and mRNA decay.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(50): p. E3444-53.

36. Wachter, A., et al., *Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs.* The Plant cell, 2007. **19**(11): p. 3437-50.

37. Bocobza, S., et al., *Riboswitch-dependent gene regulation and its evolution in the plant kingdom.* Genes & development, 2007. **21**(22): p. 2874-9.

38. Vitreschak, A.G., et al., *Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element.* RNA, 2003. **9**(9): p. 1084-97.

39. Miranda-Rios, J., M. Navarro, and M. Soberon, *A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(17): p. 9736-41.

40. Gelfand, M.S., et al., *A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes.* Trends in genetics : TIG, 1999. **15**(11): p. 439-42.

41. Grundy, F.J. and T.M. Henkin, *The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria.* Molecular microbiology, 1998. **30**(4): p. 737-49.

42. Gilbert, S.D., et al., *Structure of the SAM-II riboswitch bound to S-adenosylmethionine.* Nature structural & molecular biology, 2008. **15**(2): p. 177-82.

43. Lu, C., et al., *Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism.* Nature structural & molecular biology, 2008. **15**(10): p. 1076-83.

44. Montange, R.K. and R.T. Batey, *Structure of the S-adenosylmethionine riboswitch regulatory mRNA element.* Nature, 2006. **441**(7097): p. 1172-5.

45. Klein, D.J., T.E. Edwards, and A.R. Ferre-D'Amare, *Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase.* Nature structural & molecular biology, 2009. **16**(3): p. 343-4.

46. Batey, R.T., S.D. Gilbert, and R.K. Montange, *Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine.* Nature, 2004. **432**(7015): p. 411-5.

47. Barrick, J.E. and R.R. Breaker, *The distributions, mechanisms, and structures of metabolite-binding riboswitches.* Genome biology, 2007. **8**(11): p. R239.

48. Gardner, P.P., et al., *Rfam: updates to the RNA families database.* Nucleic acids research, 2009. **37**(Database issue): p. D136-40.

49. Rodionov, D.A., et al., *Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes.* The Journal of biological chemistry, 2003. **278**(42): p. 41148-59.

50. Marsh, E.N., *Coenzyme B12 (cobalamin)-dependent enzymes.* Essays in biochemistry, 1999. **34**: p. 139-54.

51. Marsh, E.N., *Review Article Coenzyme-B(12)-Dependent Glutamate Mutase.* Bioorganic chemistry, 2000. **28**(3): p. 176-189.

52. Gruber, K. and C. Kratky, *Coenzyme B(12) dependent glutamate mutase.* Current opinion in chemical biology, 2002. **6**(5): p. 598-603.

53. Ravnum, S. and D.I. Andersson, *Vitamin B12 repression of the btuB gene in Salmonella typhimurium is mediated via a translational control which requires leader and coding sequences.* Molecular microbiology, 1997. **23**(1): p. 35-42.

54. Franklund, C.V. and R.J. Kadner, *Multiple transcribed elements control expression of the Escherichia coli btuB gene.* Journal of bacteriology, 1997. **179**(12): p. 4039-42.

55. Johnson, J.E., Jr., et al., *B12 cofactors directly stabilize an mRNA regulatory switch.* Nature, 2012. **492**(7427): p. 133-7.

56. Peselis, A. and A. Serganov, *Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch.* Nature structural & molecular biology, 2012. **19**(11): p. 1182-4.

57. Loffet, A., *Peptides as drugs: is there a market?* Journal of peptide science : an official publication of the European Peptide Society, 2002. **8**(1): p. 1-7.

58. Horton, D.A., G.T. Bourne, and M.L. Smythe, *Exploring privileged structures: the combinatorial synthesis of cyclic peptides.* Journal of computer-aided molecular design, 2002. **16**(5-6): p. 415-30.

59. Edman, P., *Chemistry of amino acids and peptides.* Annual review of biochemistry, 1959. **28**: p. 69-96.

60. Craik, D.J., *Chemistry. Seamless proteins tie up their loose ends.* Science, 2006. **311**(5767): p. 1563-4.

61. Strieker, M. and M.A. Marahiel, *The structural diversity of acidic lipopeptide antibiotics.* Chembiochem : a European journal of chemical biology, 2009. **10**(4): p. 607-16.

62. Laupacis, A., et al., *Cyclosporin A: a powerful immunosuppressant.* Canadian Medical Association journal, 1982. **126**(9): p. 1041-6.

63. Humphries, M.J., K. Olden, and K.M. Yamada, *A synthetic peptide from fibronectin inhibits experimental metastasis of murine melanoma cells.* Science, 1986. **233**(4762): p. 467-70.

64. Noiri, E., et al., *Cyclic RGD peptides ameliorate ischemic acute renal failure in rats.* Kidney international, 1994. **46**(4): p. 1050-8.

65. Gurrath, M., et al., *Conformation/activity studies of rationally designed potent anti-adhesive RGD peptides.* European journal of biochemistry / FEBS, 1992. **210**(3): p. 911-21.

66. Gehlsen, K.R., et al., *Inhibition of in vitro tumor cell invasion by Arg-Gly-Asp-containing synthetic peptides.* The Journal of cell biology, 1988. **106**(3): p. 925-30.

67. Mas-Moruno, C., F. Rechenmacher, and H. Kessler, *Cilengitide: the first anti-angiogenic small molecule drug candidate design, synthesis and clinical evaluation.* Anti-cancer agents in medicinal chemistry, 2010. **10**(10): p. 753-68.

68. Ma, Z. and M.C. Hartman, *In vitro selection of unnatural cyclic peptide libraries via mRNA display.* Methods in molecular biology, 2012. **805**: p. 367-90.

69. Menegatti, S., et al., *mRNA display selection and solid-phase synthesis of Fc-binding cyclic peptide affinity ligands.* Biotechnology and bioengineering, 2013. **110**(3): p. 857-70.

70. Lian, W., et al., *Screening bicyclic peptide libraries for protein-protein interaction inhibitors: discovery of a tumor necrosis factor-alpha antagonist.* Journal of the American Chemical Society, 2013. **135**(32): p. 11990-5.

71. Liu, T., et al., *Synthesis and screening of a cyclic peptide library: discovery of small-molecule ligands against human prolactin receptor.* Bioorganic & medicinal chemistry, 2009. **17**(3): p. 1026-33.

72. Fischer, E., Ber. Dtsch. Chem. Ges., 1901. **34**: p. 433-454.

73. Zompra, A.A., et al., *Manufacturing peptides as active pharmaceutical ingredients.* Future medicinal chemistry, 2009. **1**(2): p. 361-77.

74. Merrifield, R.B., *Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide.* J. Am. Chem. Soc., 1963. **85**(14): p. 2149–2154.

75. Carpino, L.A. and G.Y. Han, *9-Fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group.* J. Am. Chem. Soc., 1970. **92**(19): p. 5748–5749.

76. Merrifield, R.B., *Automated synthesis of peptides.* Science, 1965. **150**(3693): p. 178-85.

77. Pedersen, S.L., et al., *Microwave heating in solid-phase peptide synthesis.* Chemical Society reviews, 2012. **41**(5): p. 1826-44.

78. Zimmer, S.H., E.; Jung, G.; Kessler H., *"Head-to-Tail" Cyclization of Hexapeptides Using Different Coupling Reagents.* Liebigs Ann. Chem, 1993: p. 497-501.

79. Alsina, J.C., C.; Ortiz, M.; Rabanal, F.; Giralt, E.; Albericio F., *Active carbonate resins for solid-phase synthesis through the anchoring of a hydroxyl function. Synthesis of cyclic and alcohol peptides.* Tetrahedron Lett., 1997. **38**(5): p. 883-886.

80. Alsina, J.R., F.; Giralt, E.; Albericio, F., *Solid-phase synthesis of "head-to-tail" cyclic peptides via lysine side-chain anchoring.* Tetrahedron Lett. , 1994. **35**(51): p. 9633-9636.

81. Valero, M.-V.G., E.; Andreu, D., *Solid Phase-Mediated Cyclization of Head-to-Tail Peptides: Problems Associated with Side Chain Anchoring* Tetrahedron Lett., 1996. **37**(24): p. 4229-4232.

82. Mihara, H.Y., S.; Nildome, T.; Aoyagi, H.; Kumagai, H., *Efficient Preparation of Cyclic Peptide Mixtures by Solid Phase Synthesis and Cyclization Cleavage with Oxime Resin.* Tetrahedron Lett., 1995. **36**: p. 4837-4840.

83. Scott, C.P., et al., *Production of cyclic peptides and proteins in vivo.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(24): p. 13638-43.

84. Tavassoli, A. and S.J. Benkovic, *Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in E. coli.* Nature protocols, 2007. **2**(5): p. 1126-33.

85. Miranda, E., et al., *A cyclic peptide inhibitor of HIF-1 heterodimerization that inhibits hypoxia signaling in cancer cells.* Journal of the American Chemical Society, 2013. **135**(28): p. 10418-25.

86. Tavassoli, A., et al., *Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction.* ACS chemical biology, 2008. **3**(12): p. 757-64.

87. Tavassoli, A. and S.J. Benkovic, *Genetically selected cyclic-peptide inhibitors of AICAR transformylase homodimerization.* Angewandte Chemie, 2005. **44**(18): p. 2760-3.

88. Scott, C.P., et al., *Structural requirements for the biosynthesis of backbone cyclic peptide libraries.* Chemistry & biology, 2001. **8**(8): p. 801-15.

89. Cheng, L., et al., *Discovery of antibacterial cyclic peptides that inhibit the ClpXP protease.* Protein science : a publication of the Protein Society, 2007. **16**(8): p. 1535-42.

90. Verhoeven, K.D., O.C. Altstadt, and S.N. Savinov, *Intracellular detection and evolution of site-specific proteases using a genetic selection system.* Applied biochemistry and biotechnology, 2012. **166**(5): p. 1340-54.

91. Young, T.S., et al., *Evolution of cyclic peptide protease inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(27): p. 11052-6.

92. Naumann, T.A., A. Tavassoli, and S.J. Benkovic, *Genetic selection of cyclic peptide Dam methyltransferase inhibitors.* Chembiochem : a European journal of chemical biology, 2008. **9**(2): p. 194-7.

93. Birts, C.N., et al., *A cyclic peptide inhibitor of C-terminal binding protein dimerization links metabolism with mitotic fidelity in breast cancer cells.* Chem. Sci., 2013. **4**(8): p. 3046-3057.

94. Kane, P.M., et al., *Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase.* Science, 1990. **250**(4981): p. 651-7.

95. Perler, F.B., *InBase: the Intein Database.* Nucleic acids research, 2002. **30**(1): p. 383-4.

96. Muir, T.W., D. Sondhi, and P.A. Cole, *Expressed protein ligation: a general method for protein engineering.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(12): p. 6705-10.

97. Otomo, T., et al., *Improved segmental isotope labeling of proteins and application to a larger protein.* Journal of biomolecular NMR, 1999. **14**(2): p. 105-14.

98. Wu, H., Z. Hu, and X.Q. Liu, *Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(16): p. 9226-31.

99. Zettler, J., V. Schutz, and H.D. Mootz, *The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction.* FEBS letters, 2009. **583**(5): p. 909-14.

100. Iwai, H., et al., *Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme.* FEBS letters, 2006. **580**(7): p. 1853-8.

101. Sun, P., et al., *Crystal structures of an intein from the split dnaE gene of Synechocystis sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing.* Journal of molecular biology, 2005. **353**(5): p. 1093-105.

102. Xu, M.Q. and F.B. Perler, *The mechanism of protein splicing and its modulation by mutation.* EMBO J, 1996. **15**(19): p. 5146-53.

103. Wu, Q., et al., *Conserved residues that modulate protein trans-splicing of Npu DnaE split intein.* Biochem J, 2014. **461**(2): p. 247-55.

104. Singh, P., P. Tripathi, and K. Muniyappa, *Mutational analysis of active-site residues in the Mycobacterium leprae RecA intein, a LAGLIDADG homing endonuclease: Asp(122) and Asp(193) are crucial to the double-stranded DNA cleavage activity whereas Asp(218) is not.* Protein Sci, 2010. **19**(1): p. 111-23.

105. Ding, Y., et al., *Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing.* J Biol Chem, 2003. **278**(40): p. 39133-42.

106. Amitai, G., B. Dassa, and S. Pietrokovski, *Protein splicing of inteins with atypical glutamine and aspartate C-terminal residues.* J Biol Chem, 2004. **279**(5): p. 3121-31.

107. Du, Z., et al., *pK(a) coupling at the intein active site: implications for the coordination mechanism of protein splicing with a conserved aspartate.* J Am Chem Soc, 2011. **133**(26): p. 10275-82.

108. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic acids research, 2014. **42**(Web Server issue): p. W320-4.

109. Shah, N.H., et al., *Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering.* J Am Chem Soc, 2012. **134**(28): p. 11338-41.

110. Shah, N.H., M. Vila-Perello, and T.W. Muir, *Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein.* Angew Chem Int Ed Engl, 2011. **50**(29): p. 6511-5.

111. Fluegel, V., *Expressed Cyclopeptide Libraries as Selective Inhibitors for RNA-Protein Interactions*, in *PhD thesis*, T.U. München, Editor. 2014.

112. Lundrigan, M.D., W. Koster, and R.J. Kadner, *Transcribed sequences of the Escherichia coli btuB gene control its expression and regulation by vitamin B12.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(4): p. 1479-83.

113. Bach, M.L., F. Lacroute, and D. Botstein, *Evidence for transcriptional regulation of orotidine-5'-phosphate decarboxylase in yeast by hybridization of mRNA to the yeast structural gene cloned in Escherichia coli.* Proc Natl Acad Sci U S A, 1979. **76**(1): p. 386-90.

114. Broschard, T.H., A. Lebrun-Garcia, and R.P. Fuchs, *Mutagenic specificity of the food mutagen 2-amino-3-methylimidazo[4,5-f]quinoline in Escherichia coli using the yeast URA3 gene as a target.* Carcinogenesis, 1998. **19**(2): p. 305-10.

115. Tanaka, T., et al., *Structures of heterooligosaccharides synthesized by levansucrase.* J Biochem, 1981. **90**(2): p. 521-6.

116. Mayer, M.P., *A new set of useful cloning and expression vectors derived from pBlueScript.* Gene, 1995. **163**(1): p. 41-6.

117. Mason, C.A. and J.E. Bailey, *Effects of plasmid presence on growth and enzyme activity of Escherichia coli DH5α.* Appl. Microbiol. Biotechnol., 1989. **32**(1): p. 54-60.

118. Kuhlman, T.E. and E.C. Cox, *Site-specific chromosomal integration of large synthetic constructs.* Nucleic Acids Res, 2010. **38**(6): p. e92.

119. Jagadish, K., et al., *Recombinant Expression and Phenotypic Screening of a Bioactive Cyclotide Against alpha-Synuclein-Induced Cytotoxicity in Baker's Yeast.* Angew Chem Int Ed Engl, 2015. **54**(29): p. 8390-4.

120. Naumann, T.A., S.N. Savinov, and S.J. Benkovic, *Engineering an affinity tag for genetically encoded cyclic peptides.* Biotechnol Bioeng, 2005. **92**(7): p. 820-30.

121. Meng, X., et al., *Counter-selectable marker for bacterial-based interaction trap systems.* Biotechniques, 2006. **40**(2): p. 179-84.

122. Kibbe, W.A., *OligoCalc: an online oligonucleotide properties calculator.* Nucleic acids research, 2007. **35**(Web Server issue): p. W43-6.

123. Chiu, J., et al., *Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb.* Journal of microbiological methods, 2008. **73**(2): p. 195-8.

124. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.

125. Zhang, X. and H. Bremer, *Control of the Escherichia coli rrnB P1 promoter strength by ppGpp.* The Journal of biological chemistry, 1995. **270**(19): p. 11181-9.

8. APPENDIX

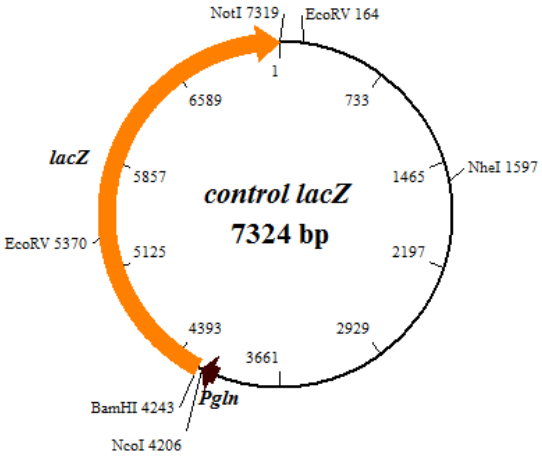
8.1 RESULTS OF REPORTER GENE ASSAYS

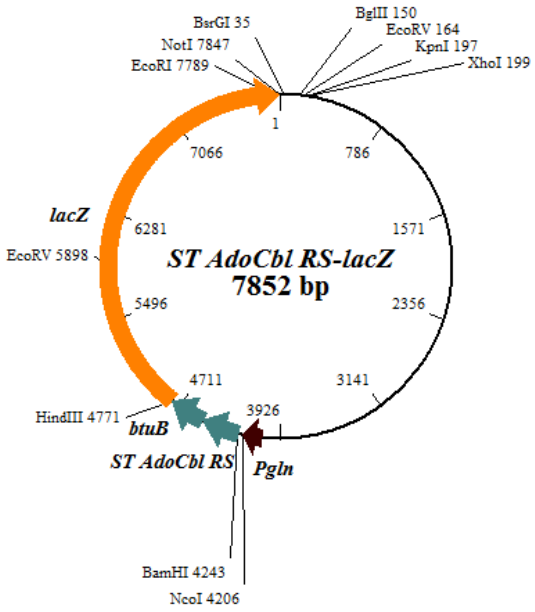
Table 25 *Results of β-galactosidase assay*

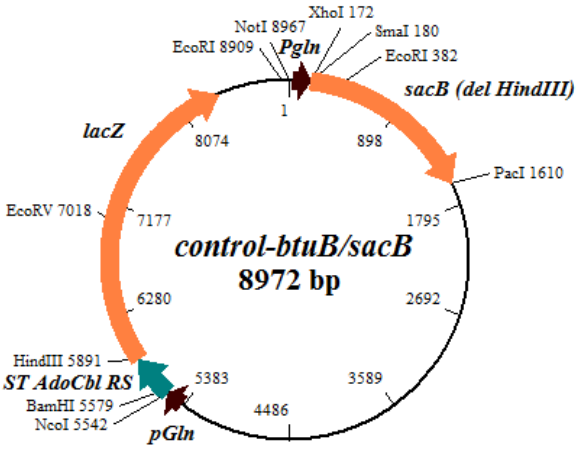
	control-lacZ		***ST AdoCbl RS-btuB-lacZ***		***ST AdoCbl RS-lacZ***	
Medium	LB	LB + 10 µM AdoCbl	LB	LB + 10 µM AdoCbl	LB	LB + 10 µM AdoCbl
OD_{600}	0.8	1.0	0.8	0.8	0.9	0.8
A_{420}	0.5 0.5 0.5	0.6 0.6 0.6	0.2 0.3 0.2	0.1 0.1 0.1	0.5 0.5 0.5	0.4 0.4 0.4
t [min]	63	63	63	63	107	107
MU	204.8 211.0 192.8	188.9 196.8 184.2	98.6 111.5 98.2	21.2 31.0 24.6	105.1 105.9 105.9	96.1 87.5 84.7
MU [%]	100.0	93.1 97.0 90.8	100.0	20.6 30.2 23.9	100.0	90.9 82.9 80.2
6	9.3	6.4	7.6	5.0	0.5	5.9
6 [%]	4.6	3.3	7.4	19.5	0.5	6.6
AM MU	202.9	190.0	102.8	34.3	105.6	89.4
AM MU [%]	100.0	93.6	100.0	24.9	100.0	84.7

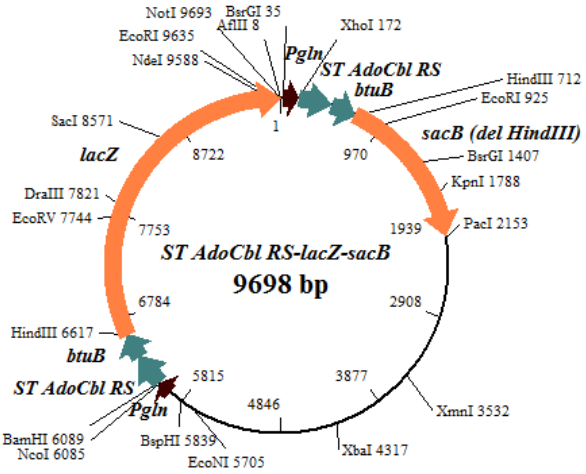
8.2 PLASMIDS

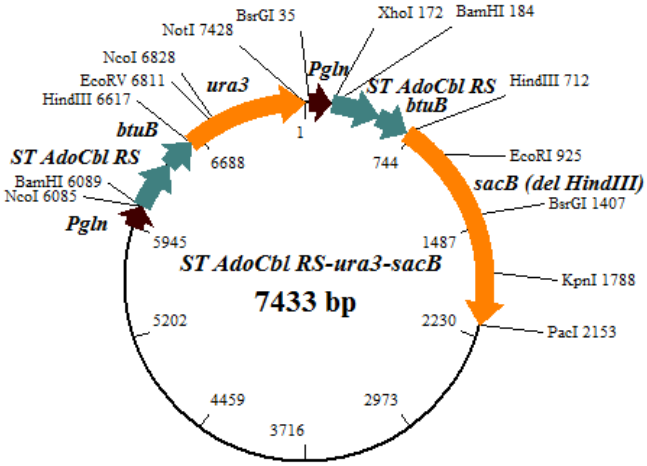
Figure 27 *Plasmids used in this thesis. pNPU CLNSLS and pSSP CLNSLS are representatives for CGWIKA, CFAHPQ and CAFHPQ plasmids. Plasmid maps of* pACYCDuet-1-P_{GlnS}-*pylS*, *pTKRED and pTKIP have been published previously.* [111, 118]

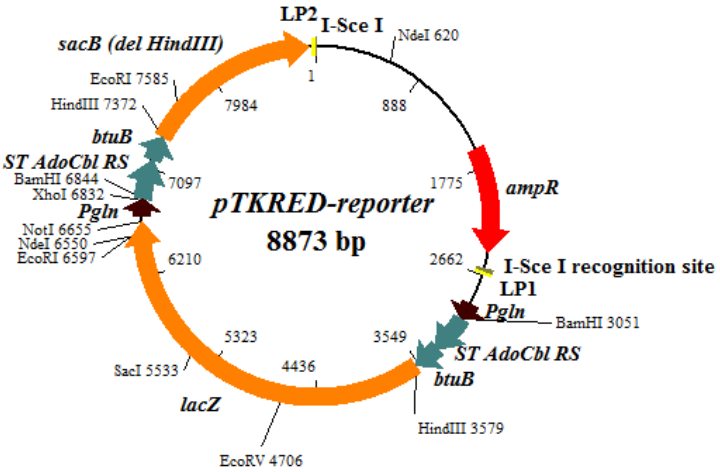


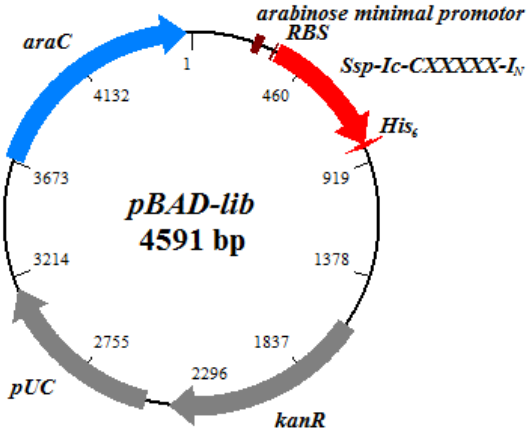


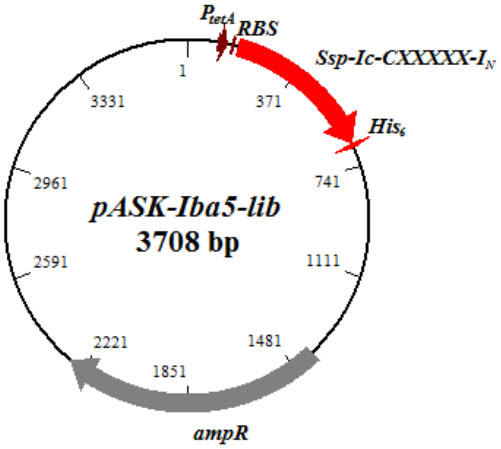

XhoI 172
NotI 8967
SmaI 180
EcoRI 8909
Pgln
EcoRI 382
sacB (del HindIII)
1
lacZ
8074
898
PacI 1610
EcoRV 7018
7177
1795
control-btuB/sacB
8972 bp
6280
2692
HindIII 5891
ST AdoCbl RS
5383
3589
BamHI 5579
4486
NcoI 5542
pGln

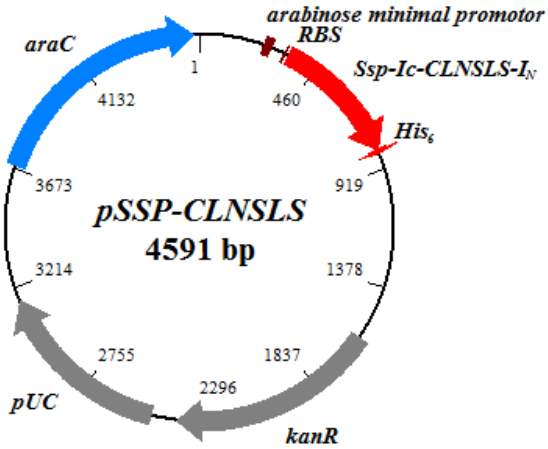

NotI 9693
BsrGI 35
AflII 8
XhoI 172
EcoRI 9635
Pgln
NdeI 9588
ST AdoCbl RS
btuB
HindIII 712
EcoRI 925
SacI 8571
1
sacB (del HindIII)
lacZ
8722
970
BsrGI 1407
KpnI 1788
DraIII 7821
EcoRV 7744
7753
1939
PacI 2153
ST AdoCbl RS-lacZ-sacB
9698 bp
6784
2908
HindIII 6617
btuB
5815
3877
ST AdoCbl RS
4846
XmnI 3532
Pgln
BamHI 6089
BspHI 5839
NcoI 6085
EcoNI 5705
XbaI 4317

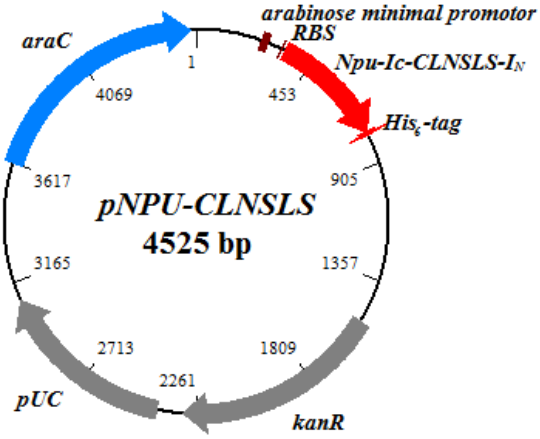

BsrGI 35
XhoI 172
BamHI 184
NotI 7428
NcoI 6828
Pgln
EcoRV 6811
ura3
ST AdoCbl RS
HindIII 712
HindIII 6617
btuB
1
btuB
744
EcoRI 925
ST AdoCbl RS
6688
sacB (del HindIII)
BamHI 6089
BsrGI 1407
NcoI 6085
Pgln
5945
1487
ST AdoCbl RS-ura3-sacB
KpnI 1788
7433 bp
5202
2230
PacI 2153
4459
2973
3716


LP2
I-Sce I
NdeI 620
1
888
sacB (del HindIII)
EcoRI 7585
7984
HindIII 7372
btuB
ST AdoCbl RS
BamHI 6844
XhoI 6832
7097
1775
ampR
Pgln
NotI 6655
NdeI 6550
EcoRI 6597
pTKRED-reporter
8873 bp
6210
2662
I-Sce I recognition site
LP1
Pgln
BamHI 3051
5323
3549
ST AdoCbl RS
SacI 5533
4436
btuB
lacZ
HindIII 3579
EcoRV 4706


arabinose minimal promotor
RBS
araC
1
Ssp-Ic-CXXXXX-I_N
4132
460
His_6
3673
919
pBAD-lib
4591 bp
3214
1378
2755
1837
2296
pUC
kanR


P_tetA
RBS
1
Ssp-Ic-CXXXXX-I_N
3331
371
His_6
2961
741
pASK-Iba5-lib
3708 bp
2591
1111
2221
1481
1851
ampR


arabinose minimal promotor
RBS
araC
Ssp-Ic-CLNSLS-I$_N$
His$_6$
1
4132
460
3673
919
pSSP-CLNSLS
4591 bp
3214
1378
2755
1837
2296
pUC
kanR


arabinose minimal promotor
RBS
araC
Npu-Ic-CLNSLS-I$_N$
His$_6$-tag
1
4069
453
3617
905
pNPU-CLNSLS
4525 bp
3165
1357
2713
1809
2261
pUC
kanR

8.3 PRIMER

Table 26 *List of primers used in this thesis. Melting temperatures were calculated by OligoCalc. [122]*

Name	Sequence (5'->3')	T_m [°C]
Reporter constructs		
oLacZ2_for	GGATTAGCGAAAGCTTATGATTACGGATTCACTGGCCG	67
oLacZ2_rev	GCTACGGTCTGCGGCCGCTTATTTTTGACACCAGACCAACTGG	72
oST_VitB12RSfor	GGCGGATCCTGATTTCTGATTTGTGATGCGCGG	67
oST_VitB12RSrev	GCCGGCAAGCTTGACGCCAGGCAAACGGCG	71
oSacB_for	GCGAGCGGCCGCATGAACATCAAAAAGTTTGC	66
oSacB_rev	GCTACGGTCTCTTAAGTTATTTGTTAACTGTTAATTG	60
oUra3_for	AAGCTTATGTCGAAAGCTACATATAAGGAACG	59
oUra3_rev	GGGGCGGCCGCTTAGTTTTGCTGGCCGCATCTTC	72
Genome Integration		
oPADlanding_nth-tpp_for	CTGCTTTCCGCTCAGGCGACCGATGTCAGTGTTAATAAGGCGACGGCGAATACGGCCCCAAGGTCCAAACGGTGA	79
oPADlanding_nth-tpp_rev	CGGAAAATGTGCGTGTCGACAGCAATAGTCGGCCAGCCGAATGCAGTGTTTTGGCTTCAGGGATGAGGCGCCATC	79
oNth/tppBver_for	ACCACCGAGCTTAATTTCAGTTCGC	58
oNth/tppBver_rev	CCTGTTCGACGTTTTTCCCCGGCGC	64

Table 26 Continuation

Name	Sequence (5'->3')	T_m [°C]
SICLOPPS		
oSLIM-Npu-Int_Fs	TGTCTAAGCTATGAAACGG	47
oSLIM-Npu-Int_Rs	ATTCGAAGCTATGAAGCC	46
oFt_Npu_CGWIKA	TGCGGCTGGATTAAAGCGTGTCTAAGCTATGAAACGG	67
oRt_Npu_CGWIKA	CGCTTTAATCCAGCCGCAATTCGAAGCTATGAAGCC	67

8.4 DNA SEQUENCES

Figure 28 *BtuB UTR including AdoCbl RS. The BtuB sequence included in reportergene constructs (210 bp) is highlighted in grey.*

TGATTTCTGATTTGTGATGCGCGGTCTGGACAGGGTGTTGTGACAGACGGTAGCATCCGTG
GGCCGGTCCTGTGAGTTAATAGGGAATCCAGTGAAAATCTGGAGCTGACGCGCAGCGGTAA
GGAAAGGTGAGATGAGAGCGTAAGCAGACACTGCCTCCGGCGGGAAGTCATCATTTCTGCT
ATCCAGCCAACGGATAACCCTCCAAGCCCGAAGACCTGCCGGCTAACGTCGCATCTGGTTT
TTCATCATCGCGTACTATCGATGAAGCCTGCGGCATCCTTCTTATATTGTGGATGCTTTAC
AATGATTAAAAAAGCTACGCTGCTGACGGCGTTCTCCGTCACGGCCTTTTCCGCTTGGGCG
CAGGACACTAGCCCGGATACCCTGGTTGTCACCGCCAACCGTTTTCAGCAGCCGCGCAGCG
CGGTTCTGGCGCCCGTTACCATCGTGACGCGTCAGGATATTGAACGCTGGCAATCGACCTC
CGTAAATGATGTTCTGCGCCGTTTGCCTGGCGTC

Figure 29 *Bacillus subtilis sacB sequence. (Gene ID 936413) The A->T substitution to delete the internal HindIII restriction site (underlined) is highlighted in red.*

ATGAACATCAAAAAGTTTGCAAAACAAGCAACAGTATTAACCTTTACTACCGCACTGCTG
GCAGGAGGCGCAACTCAAGCGTTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACA
TACGGCATTTCCCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAAAAT
GAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAAATATCTCTTCTGCAAAA
GGCCTGGACGTTTGGGACAGCTGGCCATTACAAAACGCTGACGGCACTGTCGCAAACTAT
CACGGCTACCACATCGTCTTTGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCG
ATTTACATGTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCTGGAAAAACGCTGGC
CGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTATCCTAAAAGACCAAACA
CAAGAATGGTCAGGTTCAGCCACATTTACATCTGACGGAAAAATCCGTTTATTCTACACT
GATTTCTCCGGTAAACATTACGGCAAACAAACACTGACAACTGCACAAGTTAACGTATCA
GCATCAGACAGCTCTTTGAACATCAACGGTGTAGAGGATTATAAATCAATCTTTGACGGT
GACGGAAAAACGTATCAAAATGTACAGCAGTTCATCGATGAAGGCAACTACAGCTCAGGC
GACAACCATACGCTGAGAGATCCTCACTACGTAGAAGATAAAGGCCACAAATACTTAGTA
TTTGAAGCAAACACTGGAACTGAAGATGGCTACCAAGGCGAAGAATCTTTATTTAACAAA
GCATACTATGGCAAAAGCACATCATTCTTCCGTCAAGAAAGTCAAAAACTTCTGCAAAGC
GATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCGGTATGATTGAGCTAAACGAT
GATTACACACTGAAAAAAGTGATGAAACCGCTGATTGCATCTAACACAGTAACAGATGAA
ATTGAACGCGCGAACGTCTTTAAAATGAACGGCAAATGGTACCTGTTCACTGACTCCCGC
GGATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCTTGGTTATGTT
TCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAAAACTGGCCTTGTGTTAAAAATG
GATCTTGATCCTAACGATGTAACCTTTACTTACTCACACTTCGCTGTACCTCAAGCGAAA
GGAAACAATGTCGTGATTACAAGCTATATGACAAACAGAGGATTCTACGCAGACAAACAA
TCAACGTTTGCGCCTAGCTTCCTGCTGAACATCAAAGGCAAGAAAACATCTGTTGTCAAA
GACAGCATCCTTGAACAAGGACAATTAACAGTTAACAAATAA

Figure 30 *The E. coli K12 lacZ sequence (Gene ID 945006). The sequence utilized for reporter genes is highlighted in black.*

ATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT
GGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGC
GAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGC
TTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCT
GAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATC
TACACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCG
ACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACG
CGAATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTC
GGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCC
GGAGAAAACCGCCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGAT
CAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACT
ACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTA
CTGGAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCT
TTATGGCAGGGTGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATC
GATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAA
CTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGAC
GGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAAT
GGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCAT
CATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATG
AAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTAC
ACGCTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGC
ATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGAGCGAACGC
GTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGG
AATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTCGAT
CCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATT
TGCCCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCC
ATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCC
CACGCGATGGGTAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTAT
CCCCGTTTACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAGTCGCTGATTAAATATGAT
GAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGC
CAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAA
GCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACC
AGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGAT
GGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTG
ATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGC
GTAGTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGCAG
TGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCAT
CTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAAC
CGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAACAACTGCTGACGCCG
CTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACC
CGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAA
GCAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCT
CACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGAT
GGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCG
GCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGA
TTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGAT
CTGCCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGC
GGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATC
AGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAA
GAAGGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGG

AGCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTC
TGGTGTCAAAAATAA

Figure 31 *Saccharomyces cerevisiae ura3 sequence (Gene ID S000000747)*

ATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAG
CTATTTAATATCATGCACGAAAAGCAAACAAACTTGTGTGCTTCATTGGATGTTCGTACC
ACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCCAAAATTTGTTTACTAAAAACA
CATGTGGATATCTTGACTGATTTTTCCATGGAGGGCACAGTTAAGCCGCTAAAGGCATTA
TCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATACA
GTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAAT
GCACACGGTGTGGTGGGCCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCGGAAGAAGTA
ACAAAGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATGCAAGGGCTCCCTAGCT
ACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGATTTTGTTATC
GGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTTACGATTGGTTGATTATG
ACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCATTGGGTCAACAGTATAGAACCGTG
GATGATGTGGTCTCTACAGGATCTGACATTATTATTGTTGGAAGAGGACTATTTGCAAAG
GGAAGGGATGCTAAGGTAGAGGGTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTG
AGAAGATGCGGCCAGCAAAACTAA

Figure 32 *LP-tetA fragment synthesized by GeneArt AG (Thermo Fisher Scientific Inc., Regensburg, GER). The landing pads are highlighted in green, the tetA resistance in black and the I-SceI restriction sites in blue.*

GAATTCTACGGCCCCAAGGTCCAAACGGTGATAGGGATAACAGGGTAATATTTACGTTGA
CACCACCTTTCGCGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAAT
ATGAATAGTTCGACAAAGATCGCATTGGTAATTACGTTACTCGATGCCATGGGGATTGGC
CTTATCATGCCAGTCTTGCCAACGTTATTACGTGAATTTATTGCTTCGGAAGATATCGCT
AACCACTTTGGCGTATTGCTTGCACTTTATGCGTTAATGCAGGTTATCTTTGCTCCTTGG
CTTGGAAAAATGTCTGACCGATTTGGTCGGCGCCCAGTGCTGTTGTTGTCATTAATAGGC
GCATCGCTGGATTACTTATTGCTGGCTTTTTCAAGTGCGCTTTGGATGCTGTATTTAGGC
CGTTTGCTTTCAGGGATCACAGGAGCTACTGGGGCTGTCGCGGCATCGGTCATTGCCGAT
ACCACCTCAGCTTCTCAACGCGTGAAGTGGTTCGGTTGGTTAGGGGCAAGTTTTGGGCTT
GGTTTAATAGCGGGGCCTATTATTGGTGGTTTTGCAGGAGAGATTTCACCGCATAGTCCC
TTTTTTATCGCTGCGTTGCTAAATATTGTCACTTTCCTTGTGGTTATGTTTTGGTTCCGT
GAAACCAAAAATACACGTGATAATACAGATACCGAAGTAGGGGTTGAGACGCAATCAAAT
TCGGTGTACATCACTTTATTTAAAACGATGCCCATTTTGTTGATTATTTATTTTTCAGCG
CAATTGATAGGCCAAATTCCCGCAACGGTGTGGGTGCTATTTACCGAAAATCGTTTTGGA
TGGAATAGCATGATGGTTGGCTTTTCATTAGCGGGTCTTGGTCTTTTACACTCAGTATTC
CAAGCCTTTGTGGCAGGAAGAATAGCCACTAAATGGGGCGAAAAAACGGCAGTACTGCTC
GAATTTATTGCAGATAGTAGTGCATTTGCCTTTTTAGCGTTTATATCTGAAGGTTGGTTA
GATTTCCCTGTTTTAATTTTATTGGCTGGTGGTGGGATCGCTTTACCTGCATTACAGGGA
GTGATGTCTATCCAAACAAAGAGTCATGAGCAAGGTGCTTTACAGGGATTATTGGTGAGC
CTTACCAATGCAACCGGTGTTATTGGCCCATTACTGTTTACTGTTATTTATAATCATTCA
CTACCAATTTGGGATGGCTGGATTTGGATTATTGGTTTAGCGTTTTACTGTATTATTATC
CTGCTATCAATGACCTTCATGTTGACCCCTCAAGCTCAGGGGAGTAAACAGGAGACAAGT
GCTTAGTAGGGATAACAGGGTAATGATGGCGCCTCATCCCTGAAGCCAAAAGCTT

Figure 33 *Exemplary Ssp DnaE sequence utilized for cyclization. The CP sequence CLNSLS is highlighted in green, the His_6 sequence in blue.*

ATGGTTAAAGTTATCGGTCGTCGTTCCCTCGGAGTGCAAAGAATATTTGATATTGGTCTT
CCCCAAGACCATAATTTTCTGCTAGCCAATGGGGCGATCGCCCACAATTGTCTGAACAGC
CTGAGCTGCCTCAGTTTTGGCACCGAAATTTTAACCGTTGAGTACGGCCCATTGCCCATT
GGCAAAATTGTGAGTGAAGAAATTAATTGTTCTGTGTACAGTGTTGATCCAGAAGGGAGA
GTTTACACCCAGGCGATCGCCCAATGGCATGACCGGGGAGAGCAGGAAGTATTGGAATAT
GAATTGGAAGATGGTTCAGTAATCCGAGCTACCTCTGACCACCGCTTTTTAACCACCGAT
TATCAACTGTTGGCGATCGAAGAAATTTTTGCTAGGCAACTGGACTTGTTGACTTTAGAA
AATATTAAGCAAACTGAAGAAGCTCTTGACAACCATCGTCTTCCCTTTCCATTACTTGAC
GCTGGGACTATTAAACACCACCACCACCACCACTAA

Figure 34 *Exemplary Npu DnaE sequence synthesized by GeneArt AG (Thermo Fisher Scientific Inc., Regensburg, GER). The CP sequence CLNSLS is highlighted in green, the His_6 sequence in blue.*

ATGGTCAAAATAGCCACACGTAAATATTTAGGCAAACAAAATGTCTATGACATTGGAGTT
GAGCGCGACCATAATTTTGCACTCAAAAATGGCTTCATAGCTTCGAATTGTCTGAACAGC
CTGAGCTGTCTAAGCTATGAAACGGAAATATTGACAGTAGAATATGGATTATTACCGATT
GGTAAAATTGTAGAAAAGCGCATCGAATGTACTGTTTATAGCGTTGATAATAATGGAAAT
ATTTATACACAACCTGTAGCACAATGGCACGATCGCGGAGAACAAGAGGTGTTTGAGTAT
TGTTTGGAAGATGGTTCATTGATTCGGGCAACAAAAGACCATAAGTTTATGACTGTTGAT
GGTCAAATGTTGCCAATCGATGAAATATTTGAACGTGAATTGGATTTGATGCGGGTTGAT
AATTTGCCGAATCACCACCACCACCACCACTAA

8.5 Mass Spectra Peptide Synthesis

Figure 35 *Mass spectra* of *cyclo-CL*NSL*S.* (**a**) *LC-ESI-MS full spectrum and mass filtered spectrum corresponding to the molecular weight of the CP.* (**b**) *Deviation between measured and theoretical molecular weight of cyclo-CL*NSL*S. (**c**) LC-ESI-MS/MS of cyclo-CL*NSL*S.*

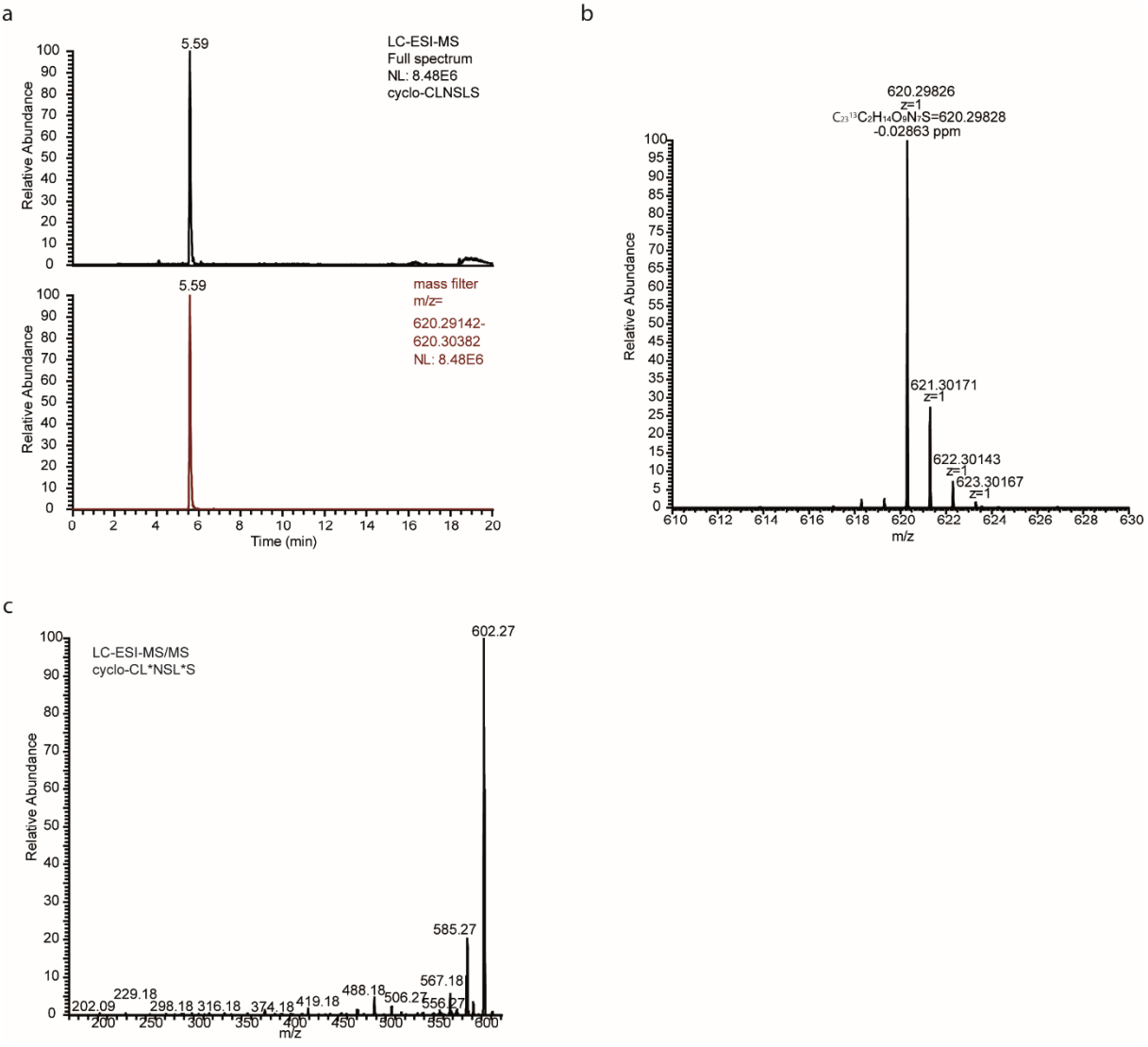


Figure 36 *Mass spectra* of *cyclo-CGWIKA and cyclo-CGWIKA*.* (**a**) *LC-ESI-MS full spectrum and mass filtered spectrum corresponding to the molecular weight of the CP (cyclo-CGWIKA left, cyclo-CGWIKA* right).* (**b**) *Deviation between measured and theoretical molecular weight of cyclo-CGWIKA (left) and cyclo-CGWIKA* (right).* (**c**) *LC-ESI-MS/MS of cyclo-CGWIKA (left) and cyclo-CGWIKA* (right).*

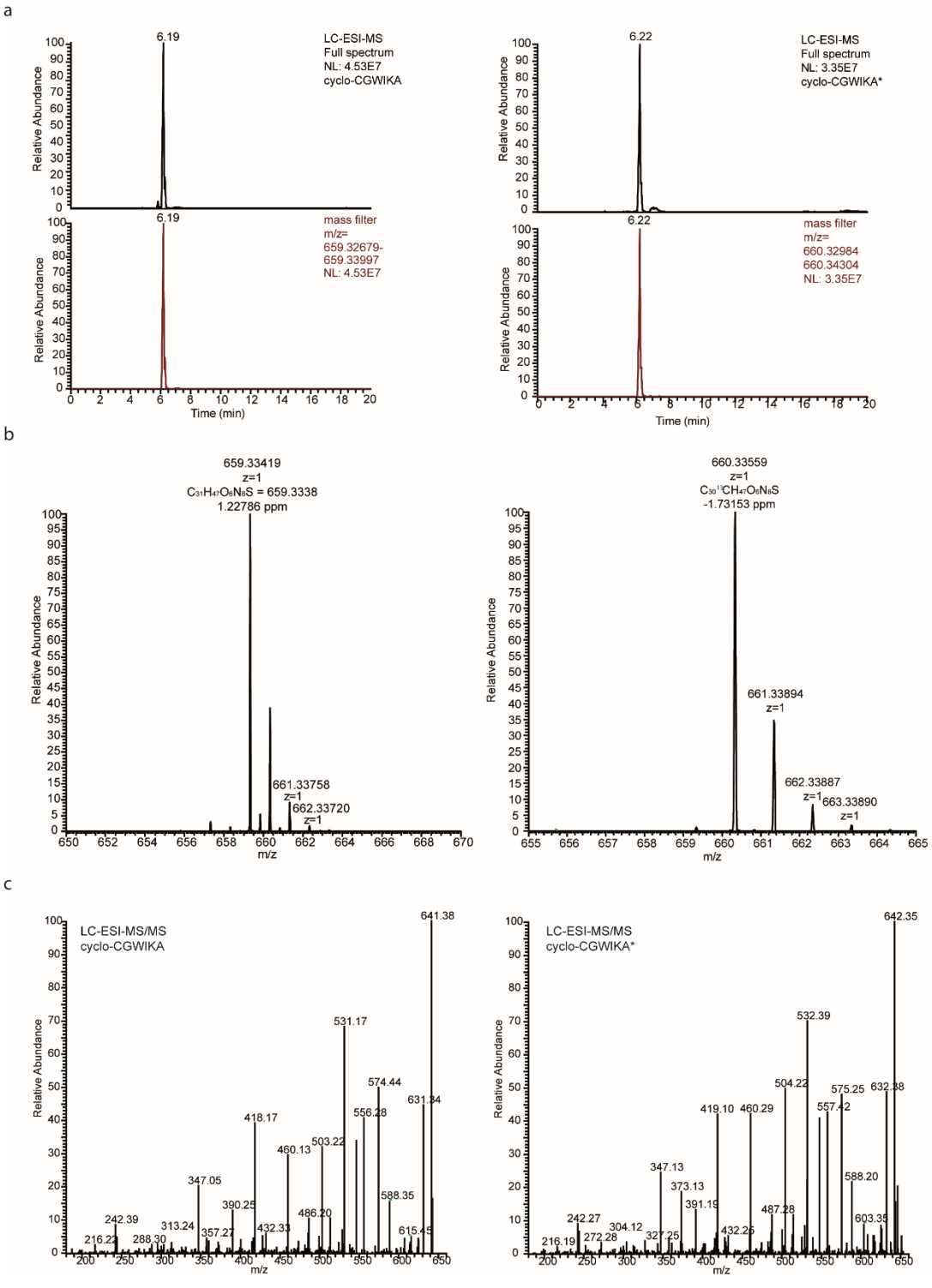


Figure 37 *Mass spectra* of *cyclo-CFAHPQ and cyclo-CFA*HPQ.* (**a**) *LC-ESI-MS full spectrum and mass filtered spectrum corresponding to the molecular weight of the CP (cyclo-CFAHPQ left, cyclo-CFA*HPQ right).* (**b**) *Deviation between measured and theoretical molecular weight of cyclo-CFAHPQ (left) and cyclo-CFA*HPQ (right).* (**c**) *LC-ESI-MS/MS of cyclo-CFAHPQ (left) and cyclo-CFA*HPQ (right).*

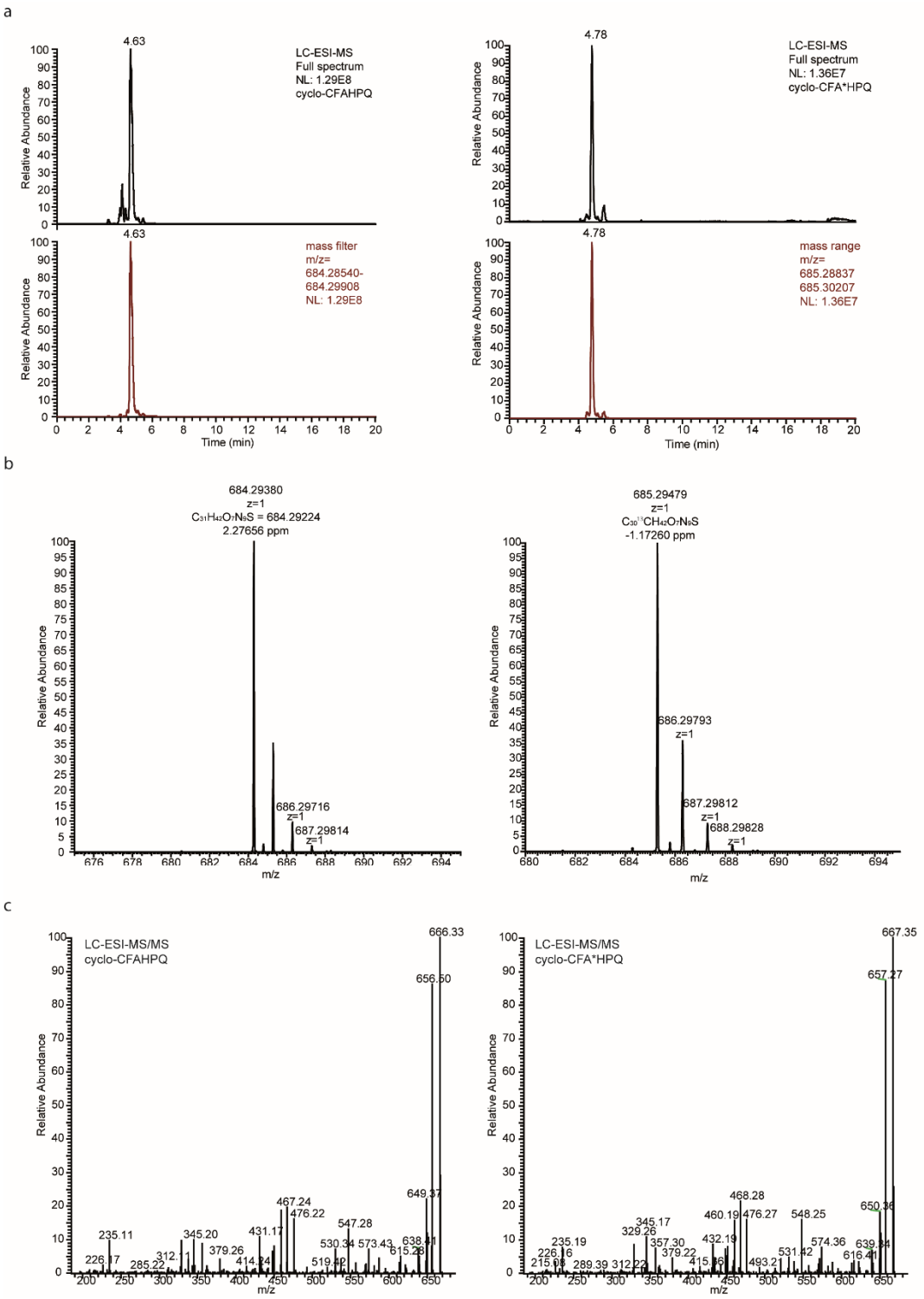
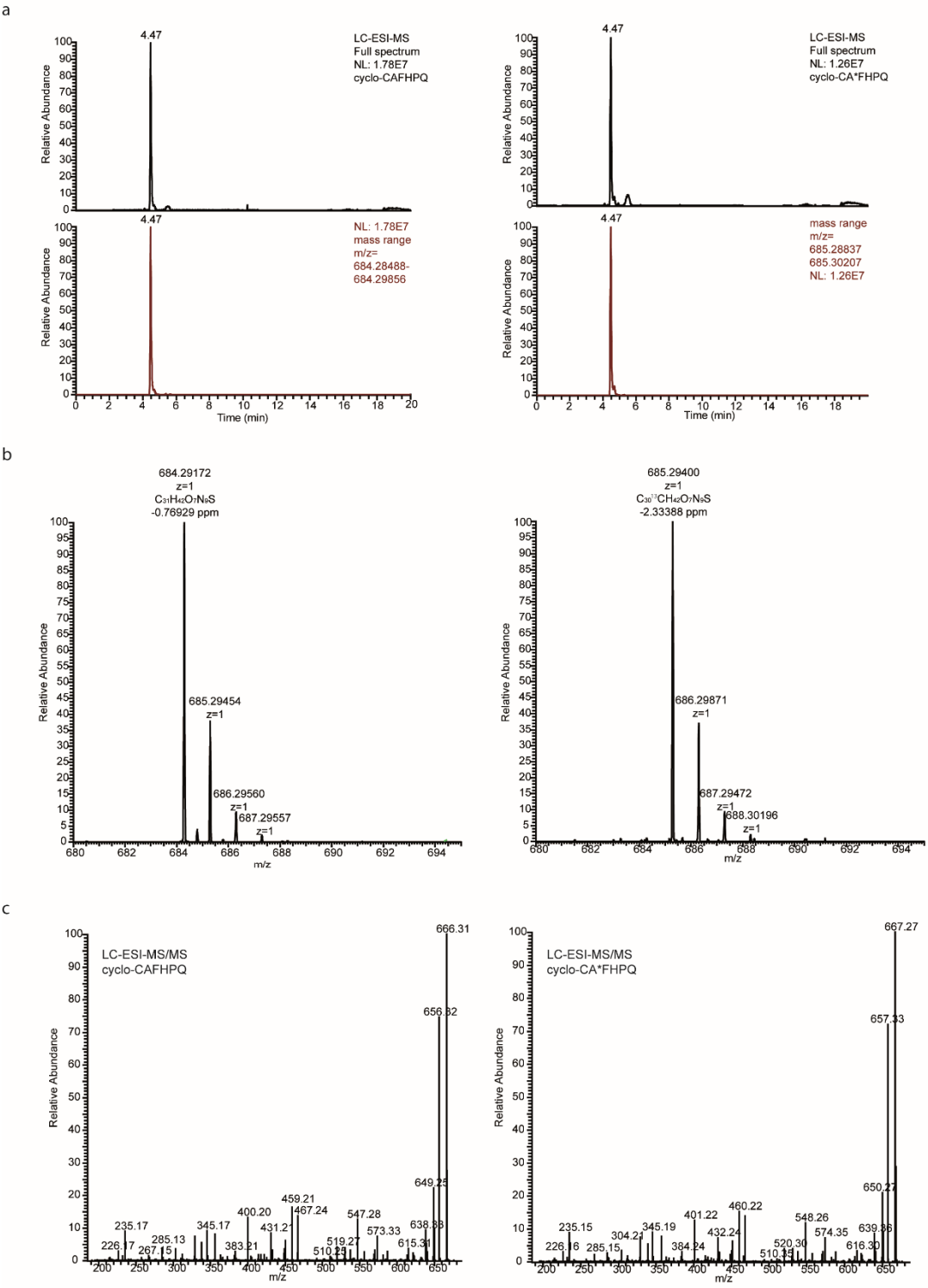
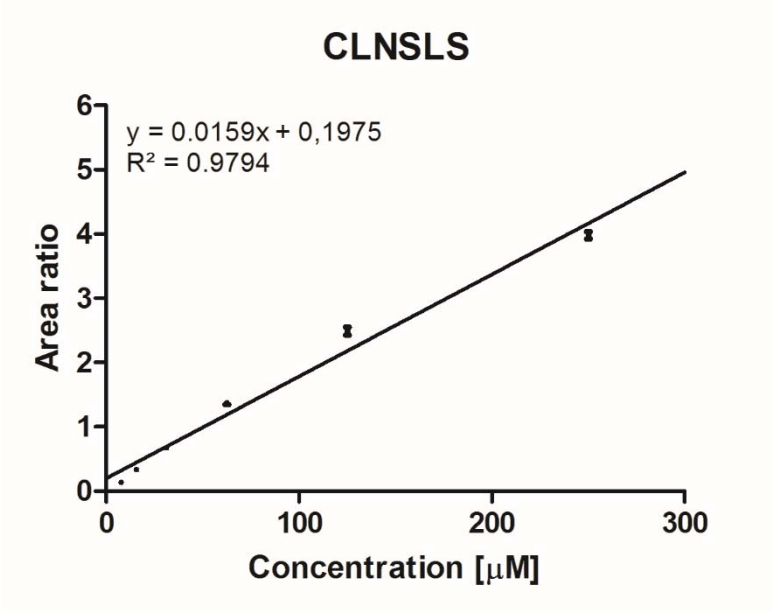


Figure 38 *Mass spectra* of *cyclo-CAFHPQ and cyclo-CA*FHPQ.* (**a**) *LC-ESI-MS full spectrum and mass filtered spectrum corresponding to the molecular weight of the CP (cyclo-CAFHPQ left, cyclo-CA*FHPQ right).* (**b**) *Deviation between measured and theoretical molecular weight of cyclo-CAFHPQ (left) and cyclo-CA*FHPQ (right).* (**c**) *LC-ESI-MS/MS of cyclo-CAFHPQ (left) and cyclo-CA*FHPQ (right).*



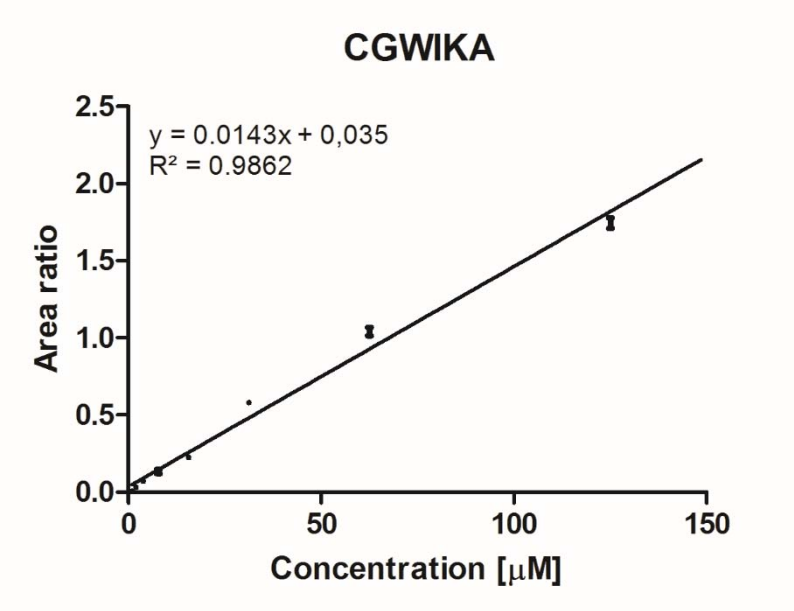
8.6 Data of Cyclic Peptide Quantification

Figure 39 *Calibration curve and data of CLNSLS. All samples were supplemented with 25* $\frac{ng}{\mu L}$ *cyclo CL*NSL*S as internal standard.*



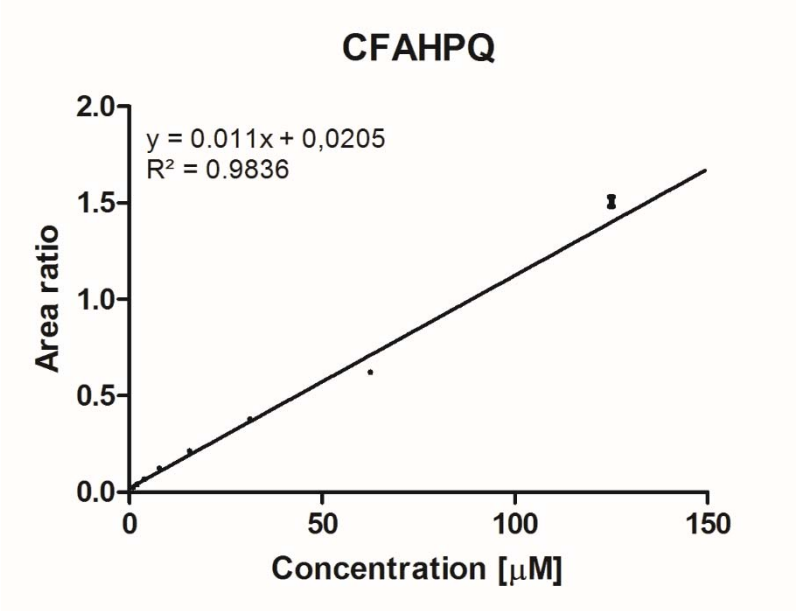
Concentration [µM]	Area ratio
7.81	0.137, 0.133, 0.137
15.6	0.331, 0.342, 0.34
31.3	0.653, 0.665, 0.717
62.5	1.349, 1.431, 1.351
125	2.364, 2.535, 2.566
250	4.062, 3.869, 4.006

Figure 40 *Calibration curve and data of CGWIKA. All samples were supplemented with* 25 $\frac{ng}{\mu L}$ *cyclo CGWIKA* as internal standard.*



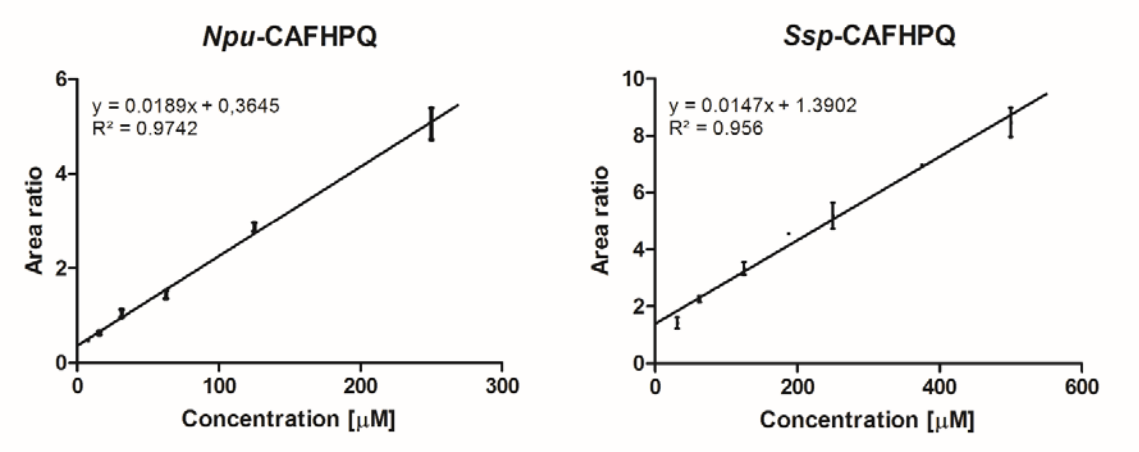
Concentration [µM]	Area ratio
0.98	0.003, 0.003, 0.011
1.95	0.026, 0.035, 0.035
3.91	0.069, 0.072, 0.073
7.81	0.125, 0.117, 0.16
15.6	0.228, 0.225, 0.227
31.3	0.599, 0.569, 0.571
62.5	1.083, 0.99, 1.05
125	1.812, 1.712, 1.705

Figure 41 *Calibration curve and data of CFAHPQ. All samples were supplemented with 25* $\frac{ng}{\mu L}$ *cyclo CFAHPQ* as internal standard.*



Concentration [µM]	Area ratio
0.98	0.02, 0.023, 0.023, 0.028, 0.028, 0.029
1.95	0.043, 0.048, 0.045, 0.034, 0.038, 0.036
3.91	0.067, 0.07, 0.071, 0.057, 0.067, 0.069
7,81	0.118, 0.124, 0.107, 0.139, 0.138, 0.123
15.6	0.222, 0.234, 0.229, 0.206, 0.186, 0.213
31.3	0.363, 0.385, 0.418, 0.376, 0.372, 0.366
62.5	0.631, 0.634, 0.636, 0.626, 0.603, 0.609
125	1.482, 1.531

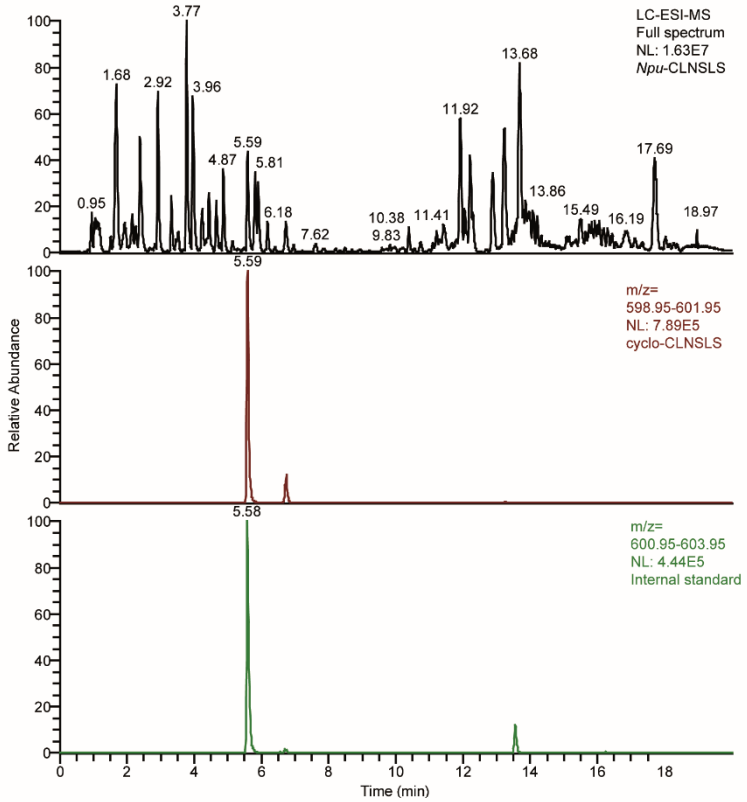
Figure 42 *Calibration curve and data of CAFHPQ for Npu-CAFHPQ (left) and Ssp-CAFHPQ (right). All samples were supplemented with* 25 $\frac{ng}{\mu L}$ *cyclo CA*FHPQ as internal standard.*



	Area ratio	
Concentration [µM]	***Npu*-CAFHPQ**	***Ssp*-CAFHPQ**
7.81	0.470, 0.475, 0.493	-
15.6	0.722, 0.745, 0.696, 0.542, 0.546, 0.565	-
31.3	0.853, 0.878, 0.867, 1.295, 1.194, 1.196	1.813, 1.349, 1.124
62.5	1.427, 1.666, 1.746, 1.209, 1.284, 1.332	2.326, 2.393, 2.066
125	2.701, 2.91, 3.016	2.889, 3.613, 3.484
187	-	4.560, 4.624, 4.554
250	4.415, 5.508, 5.247	5.282, 4.354, 5.939
375	-	7.005, 6.919
500	-	8.758, 7.462, 9.184

Figure 43 *LC-ESI-MS spectra of cell extracts utilized for quantification of* (**a**) *Npu-CLNSLS and* (**b**) *Ssp-CLNSLS.*

a



b

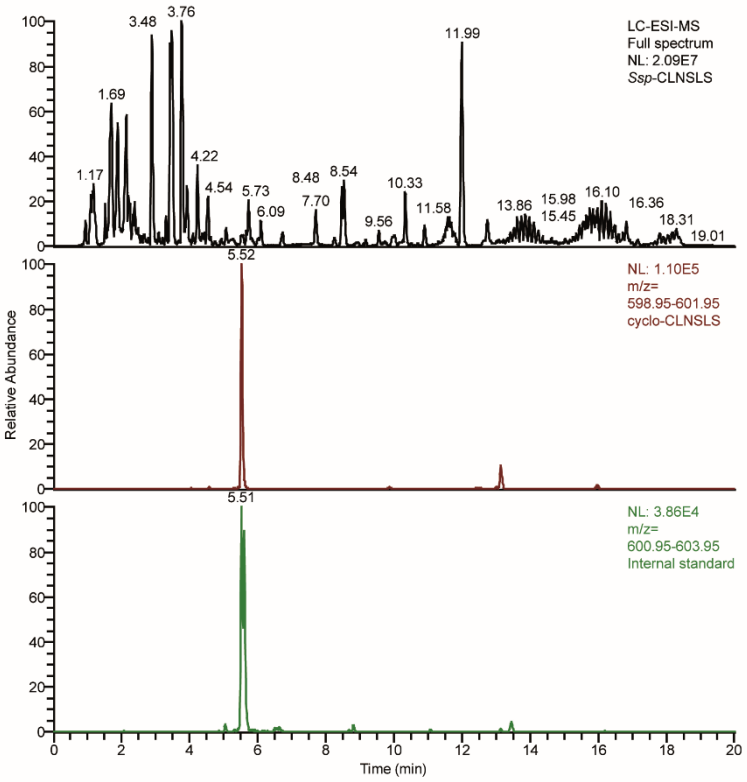


Figure 44 *LC-ESI-MS spectra of cell extracts utilized for quantification of* (**a**) *Npu-CGWIKA and* (**b**) *Ssp-CGWIKA.*

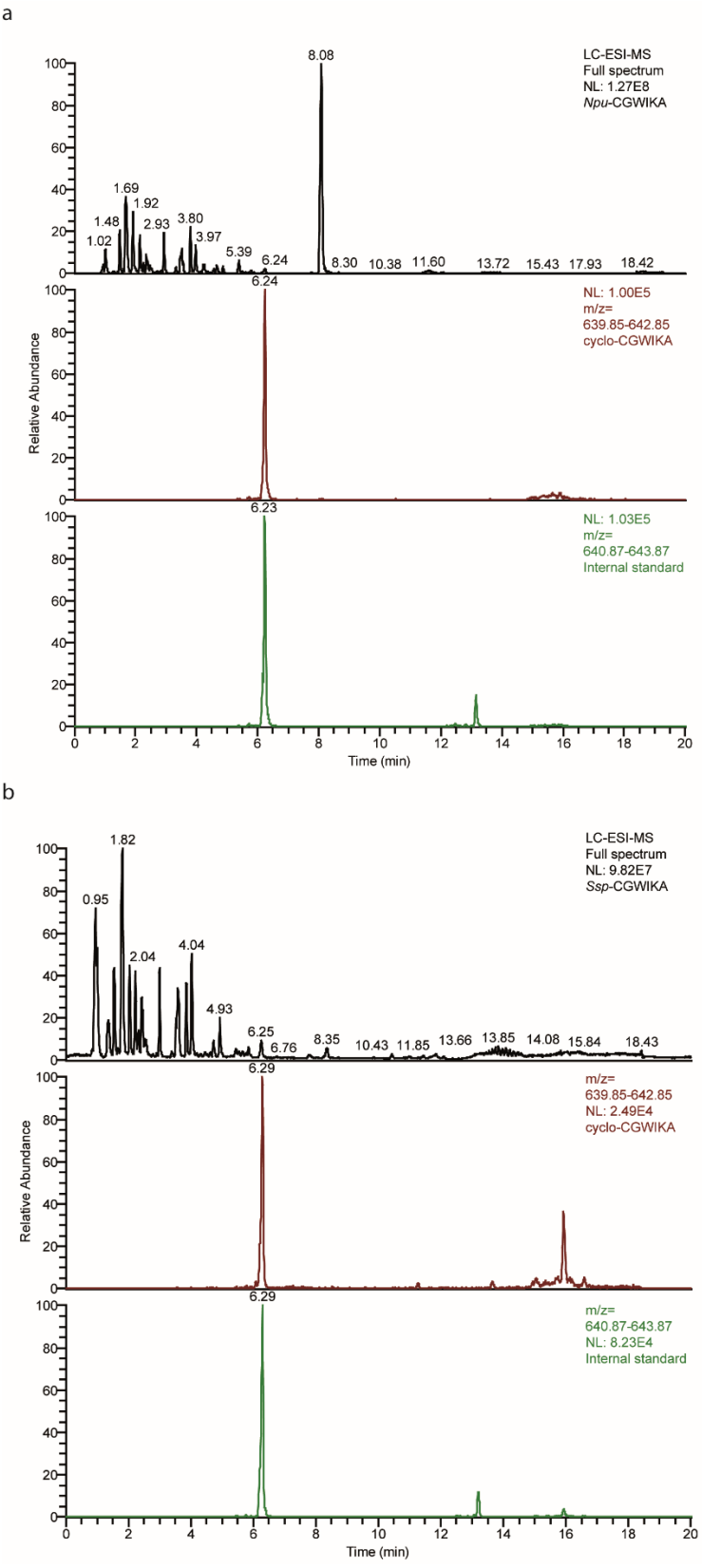


Figure 45 *LC-ESI-MS spectra of cell extracts utilized for quantification of* (**a**) *Npu-CFAHPQ and* (**b**) *Ssp-CFAHPQ.*

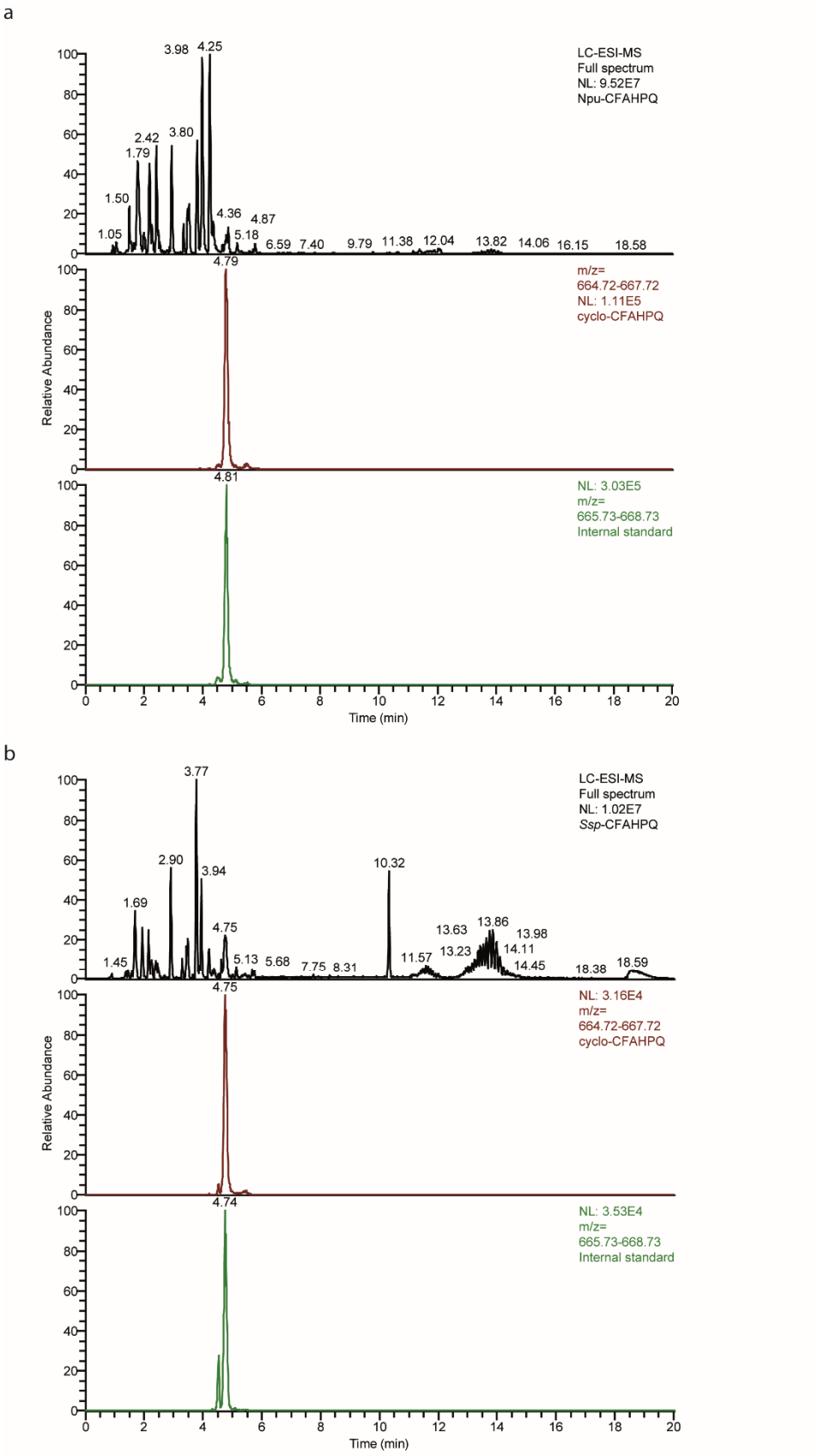
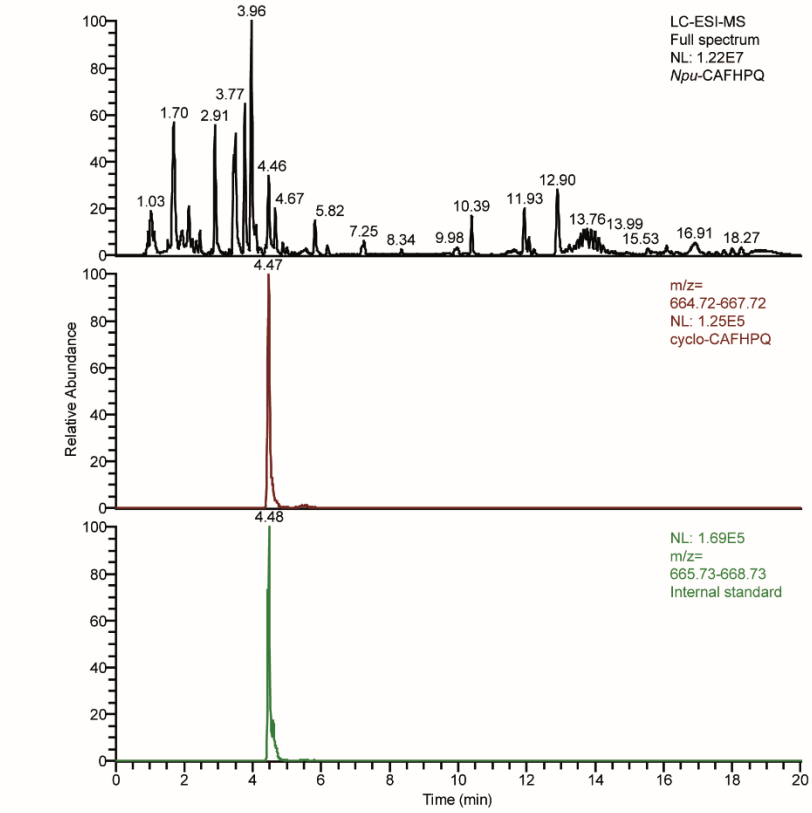


Figure 46 *LC-ESI-MS spectra of cell extracts utilized for quantification of* (**a**) *Npu-CAFHPQ and* (**b**) *Ssp-CAFHPQ.*

a



b

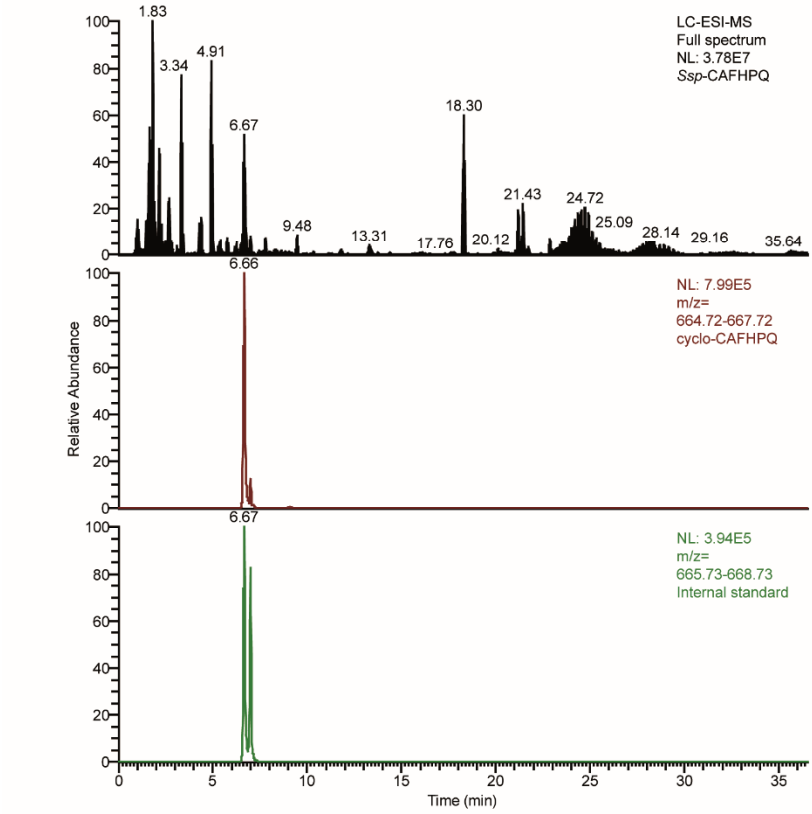


Table 27 *Data of quantification of Npu-CLNSLS, Ssp-CLNSLS, Npu-CGWIKA, Ssp-CGWIKA, Npu-CFAHPQ, Ssp-CFAHPQ, Npu-CAFHPQ and Ssp-CAFHPQ. The data corresponds to concentrations in the samples before processing.*

	Calculated amount [µM]		
	cultivation 1	***cultivation 2***	***cultivation 3***
***Npu*-CLNSLS**	238.54	258.00	202.37
	234.05	258.24	212.81
	214.98	258.30	196.37
	199.18	300.45	187.59
	187.92	336.51	212.85
	191.19	289.52	198.57
	189.27	311.39	188.04
	175.34	277.95	197.59
	177.09	291.65	190.49
***Ssp*-CLNSLS**	133.72	168.48	194.09
	139.56	169.07	215.39
	135.09	175.54	205.45
	134.35	159.63	209.25
	127.07	169.00	209.05
	133.57	162.95	213.02
	124.00	151.96	208.48
	113.49	259.28	231.06
	119.58	256.60	236.12
***Npu*-CGWIKA**	90.66	111.05	84.11
	92.30	115.95	76.83
	97.48	119.28	79.96
	91.50	104.89	90.62
	88.26	116.65	77.15
	94.34	116.69	72.75
	91.77	108.01	91.01
	96.80	101.21	99.62
	92.94	111.57	98.42
***Ssp*-CGWIKA**	16.50	12.41	16.31
	18.84	11.60	19.18
	17.71	12.28	18.65
	19.89	14.60	19.97
	15.72	12.88	17.53
	15.85	11.82	18.07
	16.16	12.36	16.68
	16.10	14.73	19.80
	16.87	12.06	18.64

Table 27 Continuation

	Calculated amount [µM]		
	cultivation 1	*cultivation 2*	*cultivation 3*
***Npu*-CFAHPQ**	28.13	37.65	41.75
	31.82	37.89	41.97
	29.79	37.46	42.08
	30.26	38.16	41.23
	28.69	37.08	42.01
	28.79	37.14	41.77
	30.73	37.18	41.13
	30.10	37.68	41.62
	31.02	36.97	41.85
***Ssp*-CFAHPQ**	44.25	83.86	82.60
	44.37	85.52	80.57
	44.10	89.76	78.05
	42.62	86.63	77.98
	44.36	81.09	83.18
	44.11	84.27	78.44
	44.16	81.81	78.64
	44.71	83.51	74.18
	44.92	84.29	73.48
***Npu*-CAFHPQ**	32.13	31.64	34.89
	27.95	29.22	33.82
	27.47	34.13	36.11
	32.14	34.69	39.10
	29.49	32.26	25.45
	28.59	30.11	37.32
	29.87	32.81	34.53
	29.43	33.84	33.86
	29.30	31.80	36.34
***Ssp*-CAFHPQ**	260.66	282.48	301.72
	306.89	322.49	319.18
	322.00	289.80	353.59
	317.77	319.39	402.84
	282.79	332.27	297.08
	264.14	259.13	305.08
	334.34	291.64	328.46
	266.28	297.38	381.73
	328.90	272.90	317.58

9. List of Publications

Making the bend: DNA tertiary structure and protein DNA interactions
S. Harteis, S. Schneider, *International Journal of Molecular Sciences* **2014**, 15(7):12335-63.

Structural, biochemical and computational studies reveal the mechanism of selective aldehyde dehydrogenase 1A1 inhibition by cytotoxic duocarmycin analogs
M. F. Koch[+], S. Harteis[+], I. D. Blank, G. Pestel, L. F. Tietze, C. Ochsenfeld, S. Schneider*, S. A. Sieber*, *Angewandte Chemie*, **2015**

10. DECLARATION

I, Sabrina Theresa Harteis, 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 will be published in scientific journals.

Garching, September 10, 2015