2019-03-01

Ribosomally Synthesized and Post-Translationally Modified Peptides as Potential Scaffolds for Peptide Engineering

Devan Bursey
Brigham Young University

Ribosomally Synthesized and Post-Translationally Modified Peptides as Potential Scaffolds for Peptide Engineering

Devan Bursey

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Joel S. Griffitts, Chair
William R. McCleary
David L. Erickson

Department of Microbiology and Molecular Biology

Brigham Young University

ABSTRACT

Ribosomally Synthesized and Post-Translationally Modified Peptides as Potential Scaffolds for Peptide Engineering

Devan Bursey
Department of Microbiology and Molecular Biology, BYU
Master of Science

Peptides are small proteins that are crucial in many biological pathways such as antimicrobial defense, hormone signaling, and virulence. They often exhibit tight specificity for their targets and therefore have great therapeutic potential. Many peptide-based therapeutics are currently available, and the demand for this type of drug is expected to continue to increase. In order to satisfy the growing demand for peptide-based therapeutics, new engineering approaches to generate novel peptides should be developed. Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a group of peptides that have the potential to be effective scaffolds for *in vivo* peptide engineering projects. These natural RiPP peptides are enzymatically endowed with post-translational modifications (PTMs) that result in increased stability and greater target specificity. Many RiPPs, such as microcin J25 and micrococcin, can tolerate considerable amino acid sequence randomization while still being capable of receiving unique post-translational modifications. This thesis describes how we successfully engineered *E. coli* to produce the lasso peptide microcin J25 using a two-plasmid inducible expression system. In addition, we characterized the protein-protein interactions between PTM enzymes in the synthesis of micrococcin. The first step in micrococcin synthesis is the alteration of cysteines to thiazoles on the precursor peptide TclE. This step is accomplished by three proteins: TclI, TclJ, and TclN. We found that a 4-membered protein complex is formed consisting of TclI, TclJ, TclN, and TclE. Furthermore, the TclI protein functions as a central adaptor joining two other enzymes in the Tcl pathway with the substrate peptide.

Keywords: peptide engineering, post-translational modifications, thiazole, lasso peptides, thiopeptides

ACKNOWLEDGEMENTS

I would like to acknowledge the support of my advisor Joel Griffitts who has been a great mentor and has helped me see what it requires to become a good scientist. I would also like to thank the other members of my committee William McCleary and David Erickson for their insight, encouragement and expertise with my academic endeavors. In addition, I would like to thank the members of the Griffitts lab as well as my cohort for their friendship and reassurance which has allowed me to enjoy my time spent here.

I would also like to acknowledge the help and support from our collaborators Kathryn Bewley and Sue Miller who have worked at UCSF on the micrococcin pathway and whose work has greatly influenced my science. In addition, I would like to acknowledge Phil Bennallack, whose previous work in the Griffitts lab on micrococcin has laid the groundwork for my research.

Lastly, I would like to thank my family, especially my parents for always encouraging and supporting me. In addition, I would like to thank my friends for their help.

TABLE OF CONTENTS

TITLE PAGE i

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER 1: Introduction 1

1.1 The power of peptides as potential therapeutic molecules 1

Global peptide-based therapeutic market 1

1.2 Engineering peptides with designer modifications 3

D- vs L- amino acid incorporation 4

Stapled peptides 5

Need for innovative peptide library design 5

Synthesizing peptides chemically versus in cells 6

Post-translational modification installation 7

1.3 Natural post-translationally modified peptide pathways as potential library scaffolds 9

Introduction to ribosomally synthesized and post-translationally modified peptides 9

Introduction to lasso peptides 10

Introduction to thiopeptides 11

Microccocin, a model thiopeptide 12

1.4 Heterocycle formation using TOMM biosynthetic enzymes 14

Heterocycle installation 14

Processing requirements for TOMM enzymes 15

1.5 Summary of subsequent chapters 18

CHAPTER 2: Lasso peptides as scaffolds for a high throughput screen of novel antimicrobial peptides 20

2.1 Introduction 20

Lasso peptides as peptide scaffolds 20

Massively parallel screening for bioactivity of randomized RiPPs *in vivo* 21

2.2 Experimental Methods 24

Strains and culture conditions 24

Plasmid construction 25

Fluorescent assays 25

Miller assays 25

Growth curves 26

2. 3 Results 26

Promoter optimization 26

A synthetic inducible MccJ25 expression system 27

Preparing McjA for easy core peptide randomization 27

2.4 Discussion 30

CHAPTER 3: Mapping interactions between micrococcin biosynthetic enzymes 33

3.1 Abstract 33

3.2 Introduction 33

The microccocin biosynthetic pathway 34

3.3 Experimental Procedures 35

Strains and culture conditions 35

Plasmid construction 37

Tcl protein expression and purification 37

Mass spectrometry analysis of TclE processing 38

Modeling of Tcl protein structures 40

3.4 Results 40

Functional Tcl proteins can be expressed in *E. coli* 40

TclI acts as a scaffolding protein in a TclIJNE complex 42

Predicted structural models 44

Further characterization of TclI interaction surfaces with TclJ, TclN, and TclE 46

3.5 Discussion 48

REFERENCES 50

SUPPLEMENTARY INFORMATION 59

CHAPTER 2: Vector sequences 59

Parent Vectors..... 59

Inserts..... 65

CHAPTER 3: Vector sequences 70

Parent vectors 70

Inserts..... 73

L501A2 vector sequence 77

LIST OF TABLES

Table 1. Approved Therapeutic Peptides 2

Supplementary Table 2. Plasmid Table 69

Supplementary Table 3. Strain Table 69

Supplementary Table 4. Plasmid Table 79

Supplementary Table 5. Strain Table 80

LIST OF FIGURES

Fig. 1. Examples of ways to engineer peptides with unnatural amino acids 4

Fig. 2. Some post-translational modifications found in nature 7

Fig. 3. Depiction of three levels of selectivity exhibited by PTM enzymes 8

Fig. 4. General biosynthetic pathway of RiPPs 10

Fig. 5. Lasso peptide structure 11

Fig. 6. Four representative thiopeptide structures 12

Fig. 7. Micrococcin biosynthesis 13

Fig. 8. Step-wise formation of Cys-derived heterocycles. 16

Fig. 9. Comparison of biosynthetic gene clusters for various 17

Fig. 10. Microcin J25 biosynthesis 22

Fig. 11. Single amino acid mutation study of McJA from Pavlova, et al. 22

Fig. 12. Overview of massively parallel activity screening (MPAS) 23

Fig. 13. Promoter optimization for *E. coli* expression 28

Fig. 14. Reconstitution of MccJ25 expression in *E. coli* 29

Fig. 15. Altering the McjA core peptide to accommodate a unique BamHI site 30

Fig. 16. Lasso Peptide Library Design 32

Fig. 17. Micrococcin Synthesis 36

Fig. 18. Functional expression and purification of thiazole installation proteins from *E. coli* 41

Fig. 19. TclI as a central docking protein in the TclEIJN complex 43

Fig. 20. Hypothetical TclIJN complex models 45

Fig. 21. Structural models for TclI, TclJ, and TclN 45

Fig. 22. Domain analysis of TclI binding to TclJ, N, and E 47

CHAPTER 1: Introduction

1.1 The power of peptides as potential therapeutic molecules

Peptides are short chains of amino acids that are ubiquitous in nature and are of great interest in the pharmaceutical industry. Peptides are crucial in a wide array of biological pathways and therefore possess a myriad of functions, including hormone signalling, antimicrobial activity, virulence, and cell-cell communication (2, 3). Since peptides are involved in many natural pathways across various species, they may easily be adapted into therapeutics in a way that closely mimics nature (3, 4). In addition to having vast target potential, peptides make excellent drug candidates because they typically are highly effective and discriminatory at binding to specific proteins. For example, some natural peptides can bind to various G protein-coupled receptors (GPCRs) and the elongation factor EF-Tu (3-5). Thus, peptides are innately primed with low off-target effects in hosts. Peptides are composed of amino acids, which are abundant in all living organisms, allowing them to also have low-immunogenicity to human hosts (6). This optimal pharmacological profile grants them a unique niche in between small molecules and larger biological entities like monoclonal antibodies (2, 3, 6, 7).

Global peptide-based therapeutic market

Peptide based therapeutics are found across various fields such as oncology, metabolic diseases, cardiovascular health, and infectious diseases (6). Currently there are over 60 peptide drugs approved for use in the United States (Table 1) and over 500 more peptide-based drugs currently in clinically trials (4, 8). In 2017, peptide-based therapeutics sales brought in more than $20 billion US dollars alone (3, 8). Some approved peptide-based pharmaceuticals such as calcitonin and ecallantide are described in Table 1. Many of these peptide-based drugs such as liraglutide (trademarked Victoza) are based on naturally occurring peptides (3). Victoza is an analog to a naturally occurring peptide called glucagon-like peptide-1 (GLP-1) (3). In humans

GLP-1 binds to the GLP-1 receptor triggering a biological cascade that enhances the secretion of insulin. Victoza also stimulates increased insulin expression and is available for treatment of treatment type 2 diabetes mellitus (3).

Table 1. Approved Therapeutic Peptides

Peptide	Mode of action	Therapeutic use	Year of Approval
Calcitonin (4, 6)	Calcitonin is a hormone that increases calcium retention	Treatment for bone diseases such as osteoporosis and to reduce high blood pressure	1971
Bivalirudin (6)	It inhibits both circulating and clot-bound thrombin	Thins the blood to prevent blood clots	2000
Daptomycin (6)	Disrupts the bacterial cell membrane	Treatment of gram-positive bacterial infections	2003
Enfuvirtide (6)	HIV fusion inhibitor	Treatment of HIV	2003
Ecallantide (9)	Inhibits kallikrein	Treatment for hereditary angioedema	2009
Victoza (6)	Derivative of human incretin, glucagon-like peptide 1 (GLP) and stimulates insulin secretion	Treatment of type 2 diabetes	2010
Afamelanotide (6)	Synthetic peptide analogue of α-melanocyte stimulating hormone and drives melanogenesis	Treatment to prevent skin damage from sun in people with erythropoietic protoporphyria	2015
Tymlos (8)	Selective activation of parathyroid hormone 1 receptor	Treatment for osteoporosis	2017

Giapreza (8)	Acts on the CNS to increase ADH production	Control of blood pressure in adults in critical condition	2017
Macrilen (8)	Mimics the endogenous ligand for Ghrelin	Diagnosis of adult growth hormone deficiency	2017
Trulance (8)	Mimics uroguanylin and guanylin and increases intestinal fluids	Used to treat irritable bowel syndrome with constipation	2017

1.2 Engineering peptides with designer modifications

Despite the success of peptide therapeutics, there are several limitations of native peptides that have tempered their development into drugs. Peptides have low bioavailability because they are easily degraded by digestive enzymes. Thus, most peptide drugs are administered parenterally, a method of drug delivery less attractive to patients (4, 6). Once inside the body, peptides have a short half-life because they are cleaved by cellular peptidases and are easily excreted through the kidneys (4, 6). Peptides are also often hydrophobic which can make crossing the cell membrane and reaching their molecular targets difficult (2, 6). Therefore, scientists have been developing tactics to overcome these apparent handicaps. Many of these solutions consist of particular chemical features such as the incorporation of unnatural amino acids to create D-peptides and stapled peptides (4, 7, 10).

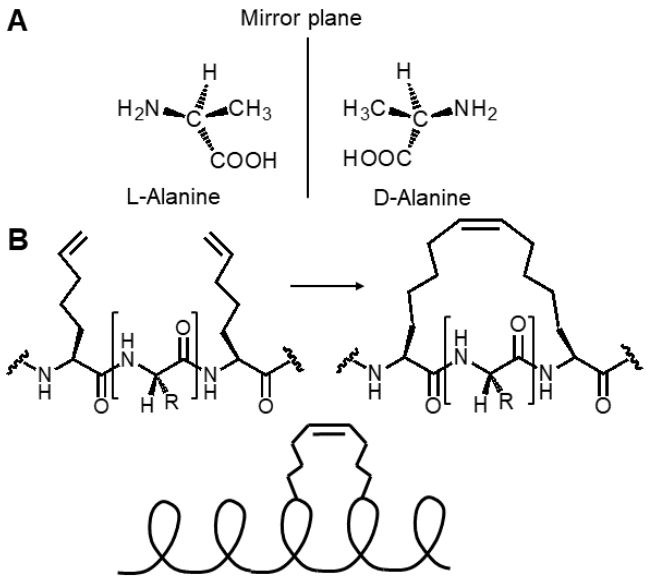


Fig. 1. Examples of ways to engineer peptides with unnatural amino acids. *A*, depiction of L-alanine compared to D-alanine. *B*, simple representation of how a stapled peptide is made. Two unnatural amino acid with a hydrocarbon chain are incorporated into the peptide. Those two hydrocarbons react to form a heterocyle. The brackets on the peptide chain represent that the bracket region can be expanded to show more residues that are inbetween the two unnatural amino acids.

D- vs L- amino acid incorporation

There are two isoforms for each amino acid (excluding glycine) which are designated as either a D- (right) amino acid or L- (left) amino acid (Fig. 1*A*.) In nature, biology uses almost exclusively L-amino acids, with a notable exception of the incorporation of D-amino acids into the peptidoglycan of bacterial cell walls (11, 12). Therefore, naturally occurring peptides are also composed of L-amino acids which makes them intrinsically susceptible to proteolysis by the many proteases present *in vivo* thereby decreasing their biostability (11, 12). For example, P113D is an antimicrobial drug derived from histatin that has D-amino acids incorporated into it. With the incorporation of D-amino acids P113D has increased stability and becomes biologically active in the sputum of patients with cystic fibrosis unlike the natural L-amino acid version of P113.

Stapled peptides

Another solution to rigidify the structure of a peptide is to create a stapled peptide (5) (Fig. 1*B*). Linear peptides possess numerous rotatable bonds that allow them conformational freedom to alternate different structures. This physical fluidity can weaken peptide stability (10, 13, 14). A stapled peptide is locked into a fixed conformational structure by hydrocarbon crosslinks. This can be accomplished by incorporating distinct unnatural amino acids into successive turns in the peptide (10, 13, 14). These amino acids have been modified to contain special hydrocarbon side chains which interact with each other to form a ring structure, effectively securing the α-helix structure of the peptide Fig. 1*B* (10, 13, 14). These unnatural amino acids get incorporated into the peptide when the peptide is being chemically synthesized. Then cyclization of the hydrocarbon side chains is initiated by adding various chemical catalysts. One of the first examples of this method was done on an HIV-1 fusion inhibitor T649. When staples were introduced into this peptide it was resistant to chymotrypsin cleavage and demonstrated increased anti-viral activity (14).

Need for innovative peptide library design

Mining peptides from nature is advantageous because they have been evolutionarily selected to bind to their specific target and maintain a certain level of stability; however, there is a limit to the number of naturally occurring peptides (3, 6), and natural peptides are not typically perfectly adapted to therapeutic needs. Since the global market for peptide therapeutics is expected to almost double by 2025 reaching approximately $50 billion US dollars, it is imperative that novel and innovative peptides be discovered and developed (3, 6, 8). One way of discovering novel peptide drugs is by creating large peptide libraries. These novel peptide libraries can be created chemically or in living cells such as bacteria.

Synthesizing peptides chemically versus in cells

The standard method for chemically synthesizing peptides is a process pioneered by Dr. Robert Merrifield known as solid-phase synthesis (2, 6). Solid-phase synthesis is a step-wise process for the C- to N- terminus construction of a polypeptide chain attached to a resin bead (2, 6). One of the advantages of chemically synthesizing peptide libraries is that it is possible to incorporate designer features such as unnatural amino acids which can increase peptide stability (e.g. D-amino acids and stapled peptides). In addition, chemically synthesized peptides can be easily purified since the resin bead acts as a purification tag (2, 6). Some disadvantages to chemically synthesized peptide libraries are low product yield, inability to synthesize large peptides, and peptides are synthesized in individual reactions thus limiting library size. (2, 6, 15, 16).

An alternative method to creating peptides is to genetically engineer bacteria, this can be accomplished by having the desired peptide sequence encoded on a plasmid in the cell and having the ribosome synthesize the product (2). A great advantage to synthesizing peptides in cells is the scalable product yield (6). Another ideal quality is that there is potential to create vast peptide libraries simply by randomizing the amino acid sequence of the peptide encoded on a plasmid. By randomizing five amino acids in a peptide, one can generate approximately 20^5 (3.2 million) novel peptides. Some major disadvantages to generating peptide libraries in living cells is that overexpressing certain proteins can be toxic to cells, or some proteins may not express well (17). One major drawback of synthesizing peptides produced by bacteria compared to chemical synthesis is that historically bacterially produced peptides cannot be made to incorporate special chemical modifications onto their side chains (2, 6). Therefore, developing

methods to make post-translational modifications on peptides is important to enhance the potential of bacteria to generate peptide libraries for therapeutic studies (18, 19).

Post-translational modification installation

There are many types of post-translational modifications (PTMs) that occur in a cell, such as glycosylation, lysine acetylation, phosphorylation, ubiquitinylation, heterocycle formation, dehydration, etc. (20, 21) (Fig. 2). These PTMs get incorporated onto peptides through the activity of various enzymes. All PTM incorporating enzymes exhibit a certain affinity for their substrates (22). However, from an engineering perspective, there is a trade-off between substrate-specificity and substrate-permissiveness (23, 24). There are varying levels to which an enzyme modulates its selectivity. An enzyme might recognize its substrate based on adjacent or nearby sequence (regioselectivity), identity of the modified side chain (chemoselectivity), or through remote sites on the substrate protein (alloselectivity) (Fig. 3). Some enzymes, such as

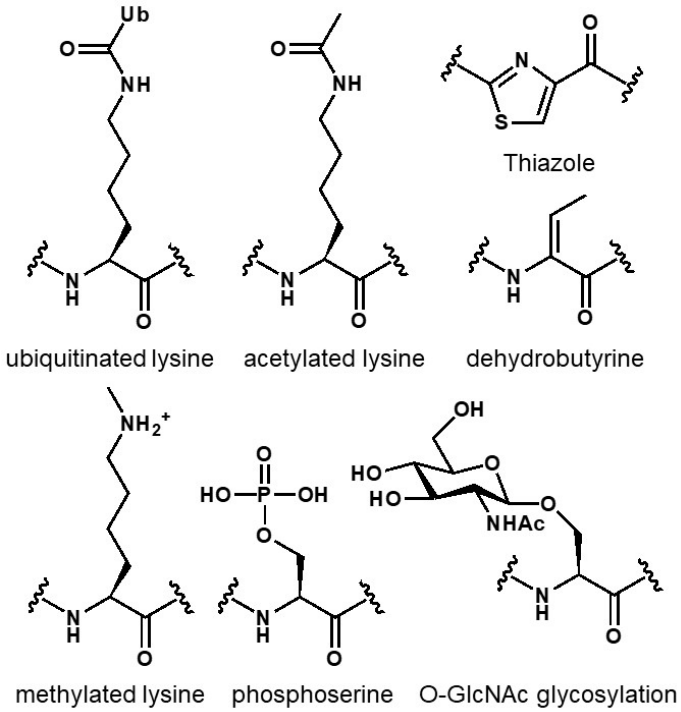


Fig. 2. Some post-translational modifications found in nature. Chemical structures for several different post translational modifications.

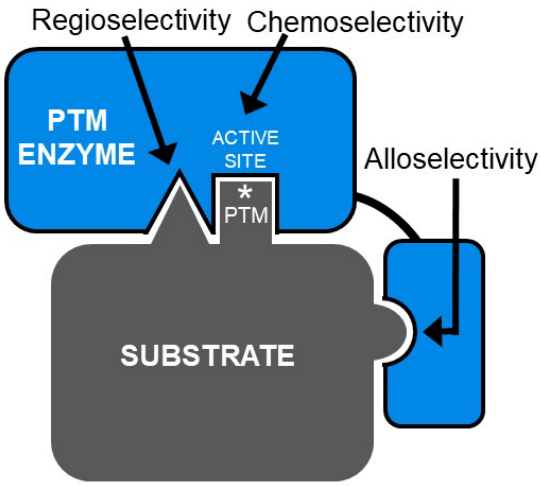


Fig. 3. Depiction of three levels of selectivity exhibited by PTM enzymes. See text for details.

kinases, exhibit remarkable specificity for their substrate based on these three levels. Many kinases require serine/threonine (or tyrosine) at the phosphorylation site (chemoselectivity), preferring side chains in close proximity to residues of phosphorylation (regioselectivity) and often binding with the substrate peptide on far-away regions from the enzyme active site (alloselectivity) (25-27). Other enzymes are far looser in their level of specificity. For example, the cyanobactin subtilisin-like serine protease PatG is an enzyme responsible for the macrocyclization in the cyanobacterial patellamide pathway. It exhibits a fairly relaxed substrate specificity so long as a short C-terminal sequence AYDE (E) is present (28, 29). This allows PatG to act on 29 known natural variant patellamide-precursor peptides (29). Interestingly, PatG can form macrocyles anywhere between 5 to 22 amino acids in size. In addition, a study was carried out using synthetic peptides and PatG, which showed that PatG can act on peptides of varying length and even those that contain unnatural amino acids and D-amino acids (30). Substrate promiscuity does come at a cost, as PatG reactions tend to happen at a very slow rate (30). Most widely tolerant PTM enzymes have a decreased kinetic rate (19). These loose PTM

enzymes with low regioselectivity are of great interest. They can be developed into a powerful tool that would use a constant handle that is sufficient to recruit the PTM enzyme but allows for adjacent peptide sequences to be modified in diverse sequence contexts (18, 19). By pairing a promiscuous PTM enzyme with a variety of random peptides bearing the appropriate PTM enzyme-recognition handle, it would be possible to generate a large post-translationally modified peptide library in bacterial cells (18, 19, 31, 32). There is also the potential that one could pair two promiscuous enzymes from different biosynthetic pathways together to generate even more complex and structurally diverse peptide libraries (18, 19). An ideal platform would use enzymes with minimized regioselectivity (i.e. many sequence contexts are tolerated) but maintain high chemoselectivity and alloselectivity. An enticing natural template for this type of bioengineering project are ribosomally synthesized and post-translationally modified peptide biosynthetic pathways, such as thiopeptides and lasso peptides.

1.3 Natural post-translationally modified peptide pathways as potential library scaffolds

Introduction to ribosomally synthesized and post-translationally modified peptides

Ribsomally synthesized and post-translationally modified peptides (RiPPs) possess many enzymatic chemical alterations which drastically alter their structures and functions. RiPPs are produced by many different organisms and there are currently over 20 different classes of RiPPs such as lanthipepides, cyanobactins, thiopeptides, proteucins, microviridins, bottromycins, lasso peptides and linaridians (33, 34). Several of these families lend themselves to *in vivo* randomization, therefore they have the potential to be used as scaffolds to generate thousands of new potentially biologically active peptides (18, 19, 31, 32). While each RiPP family possesses a unique structural arrangement with a distinctive profile of PTMs, RiPPs generally follow the same biosynthetic logic (Fig. 4) (35). RiPPs are synthesized on the ribosome as linear precursor

peptides which are then subsequently enzymatically modified to their final structure. The precursor peptide is composed of two functional parts: the leader region and the core region. The leader region is usually involved in initial recognition by modifying enzymes while the core region receives the modifications and eventually becomes the active peptide product after leader removal (Fig. 4) (33, 35). Since most PTM enzymes recognize the leader peptide, RiPP core peptides can be mutated and the enzymes can still modify them provided the leader peptides remains intact (18, 19). Lasso peptides and thiopeptides are the two families of RiPPs that will be elaborated on further (31, 32).

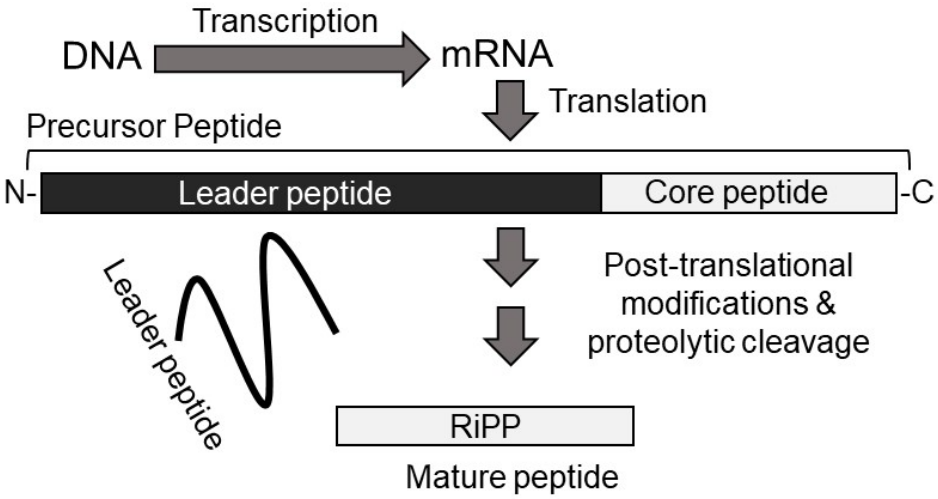


Fig. 4. General biosynthetic pathway of RiPPs. RiPP maturation begins with the production of a bipartite precursor peptide. This peptide consists of an N-terminal leader and a C-terminal core. The core gets extensively modified by PTM enzymes and the leader gets cleaved from the altered core to result in the mature RiPP.

Introduction to lasso peptides

Lasso peptides are a RiPP family that were first discovered in 1991 and are primarily produced by *Proteobacteria* and *Actinobacteria* (36). There are currently 35 known lasso peptides (36). Lasso peptides follow the same general maturation pathway as other RiPPs, beginning as a bipartite precursor peptide that gets transformed into a mature peptide by enzymes (32, 36, 37). These lasso peptides exhibit a wide array of functions, including antimicrobial

activity, and are characterized by their unique, knot-like shape (Fig. 5) (32, 36, 37) . The C-terminal tail gets threaded through a 7-9 amino acid N-terminal ring, forming a loop (36, 38-40). The loop topology is usually stabilized by steric interactions and their structure makes them highly resistant to proteases. There is evidence in the literature that some lasso peptides, like microcin J25, can be modified to produce thousands of unique peptides (1, 32, 37).

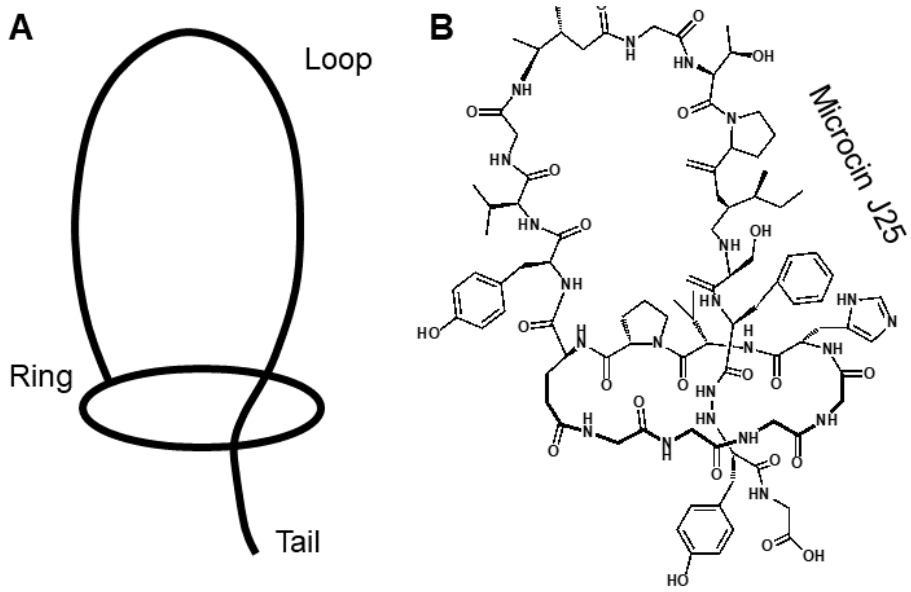


Fig. 5. Lasso peptide structure. A, a simple graphic to show the general structure of lasso peptide. B, the structure of lasso peptide microcinJ25.

Introduction to thiopeptides

Thiopeptides are a large RiPP family that boasts over 100 members (Fig. 6) (41, 42). They make attractive scaffolds for engineering novel peptides due to their wide variety of biological properties (31). Many possess antimicrobial activity, specifically against Gram-positive bacteria usually through inhibition of translation either by interfering with the elongation factor EF-Tu or the 50S ribosomal subunit (41, 43). In addition, they have demonstrated anti-viral, anti-parasitic, immunosuppressive, and anti-proliferative effects (41, 43). The hallmark of a thiopeptide is a six-membered nitrogenous ring that closes a peptide macrocycle. Certain thiopeptides such as thiostrepton A contain two macrocycles. In addition, thiopeptides also

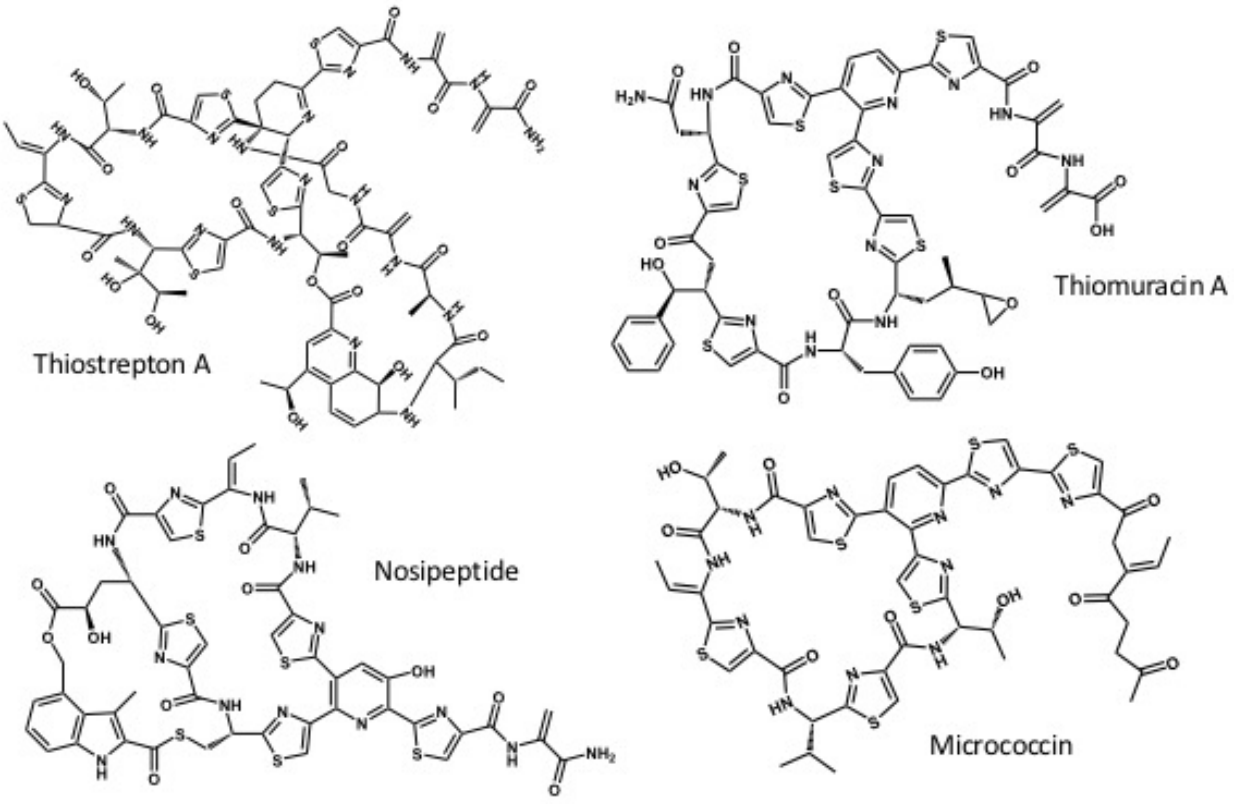


Fig. 6. Four representative thiopeptide structures. Thiostrepton A and nosipeptide are bimacrocyclic thiopeptide structures. Micrococcin and thiomuracin are both monomacrocylic thiopeptide structures.

contain azoles rings (thiazoles and oxazoles), dehydroamino acids, and other pathway-specific alterations (41, 43, 44). Thiopeptides begin as linear precursor peptides 20-110 amino acids in length which are then enzymatically endowed with chemical alterations. Canonically, the C-terminal core peptide is used as the substrate for all the subsequent PTMs. The N-terminal leader region is usually involved in initial recognition by modifying enzymes, but eventually it is cleaved and only the core remains in the final peptide architecture.

Microccocin, a model thiopeptide

We have been studying the archetypal thiopeptide micrococcin that is produced by gram positive bacteria species like *Macrococcus caseolyticus* and *Bacillus cereus* (43) (Fig. 6.). In recent years, a plasmid-based gene cluster for the production of micrococcin from *Macrococcus*

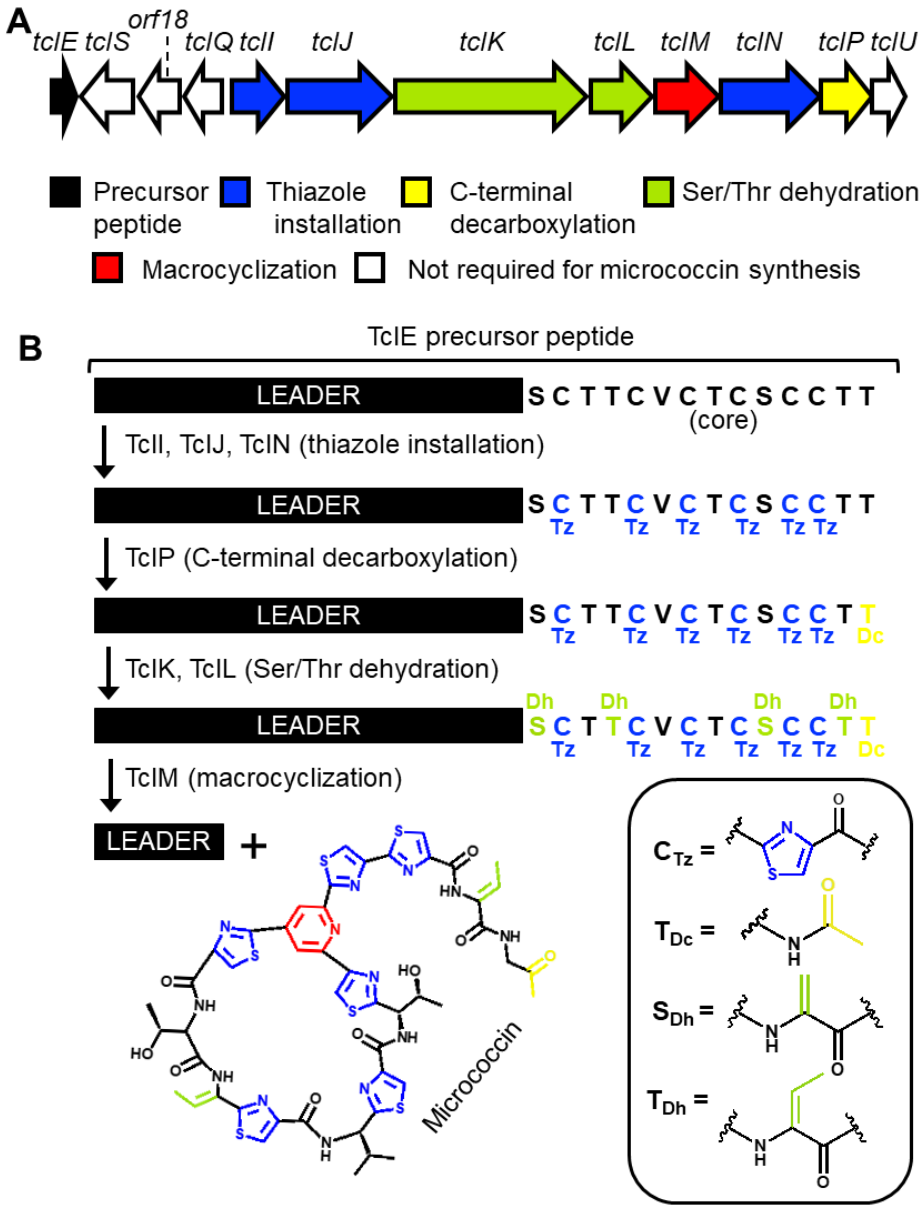


Fig. 7. Micrococcin biosynthesis. *A*, Gene cluster for micrococcin synthesis. *B*, biosynthetic pathway of micrococcin synthesis.

caseolyticus has been described (42, 45). The gene cluster for complete micrococcin synthesis in the *M. caseolyticus* system has been identified, and consists of 12 protein-coding genes (8 of which are required for micrococcin biosynthesis), making it a relatively simple model thiopeptide biosynthetic pathway (Fig. 7*A*) (42). TclE is the precursor peptide which acts as the substrate that gets manipulated by the PTM enzymes to eventually become the mature

thiopeptide. It follows typical RiPP precursor peptide composition possessing a 35 amino acid N-terminal leader and a 14 amino acid C-terminal core (Fig. 7*B*). The maturation of micrococcin begins with the installation of six thiazoles from the six cysteines present in the core region. This step is carried out by three proteins: TclI, TclJ, and TclN. Thiazole installation is followed by a C-terminal decarboxylation step, carried out by TclP (46). Next, serine/threonine dehydration is brought about by TclK and TclL on several serine and threonine residues within the core region. Micrococcin maturation is culminated by a macrocyclization event initiated by TclM forming a pyridine ring out of the two dehydrated serine side chains (Fig. 7). Of greatest interest to us for the future development of creating *in vivo* systems capable of installing PTMs on peptide libraries are those proteins involved heterocycle formation: TclJ, TclN, TclI, and TclE (42, 43, 45, 46).

1.4 Heterocycle formation using TOMM biosynthetic enzymes

Heterocycle installation

The formation of heterocycles such as thiazoles (derived from cysteines) or oxazoles (derived from serine/threonine) in thiopeptides is a two-step process. This process begins with a cyclodehydration event to form an azoline, followed by a dehydrogenation step to create the more stable aromatic azole (Fig. 8) (47, 48). Heterocycle formation is not unique to thiopeptides and is a pattern that appears in many other RiPP families such as cyanobactins and lanthipeptides (49, 50). The process involves three different proteins: a heterocyclase, a precursor peptide recognition protein and a dehydrogenase (Fig. 8). The heterocyclase, dehydrogenase, and precursor peptide recognition protein are classified into a group called thiazole/oxazole-modified microcins (TOMM) biosynthetic enzymes (47). Typically, both the heterocyclase and precursor peptide recognizing protein are involved in azoline formation. The heterocyclase is

primarily responsible for catalysis of cyclodehydration. The heterocyclase consists of an ATP-dependent YcaO domain with a proline rich C-terminus (PXPXP) which bestows upon it azoline-incorporating capability (51-53). The protein involved in precursor peptide recognition has one domain that usually contains some sort of N-terminal winged-helix-turn-helix structure which is defined as the RiPP precursor peptide recognition element (RRE). This RRE allows the protein to bind to the leader peptide (35, 54). This binding presumably helps to accurately guide the core of the precursor peptide to the active site of the heterocyclase. In approximately half of the known TOMM biosynthetic gene clusters the heterocyclase and precursor peptide recognition protein are fused together (e.g. LynD of the aestuaramide pathway and TruD of the trunkamide pathway) showcasing the importance of the involvement of both proteins in cyclodehydration (30, 47, 52). The dehydrogenases are FMN-dependent and oxidize azolines to azoles (Fig. 8). There is great sequence diversity amongst dehydrogenase proteins but they all belong to the nitroreductase superfamily and appear to have a conserved Lys-Tyr motif near the FMN binding site (55, 56). Dehydrogenases can function as a single protein or fused to another protein (e.g. PatG involved in patellamide biosynthesis) (57, 58). In pathways where the terminal heterocycle form is a thiazoline or oxazoline then a dehydrogenase may be absent (47, 48).

Processing requirements for TOMM enzymes

The enzymatic requirements for substrate processing, including complex formation, direction of processing, chemo-, regio-, and allo-specificity vary amongst TOMM gene clusters. Some other studied TOMM enzymatic pathways include microcin B17 (59-61), steptolysin S (62, 63), hakacin (54, 64), patellamide A (57, 65, 66), and thiomuracin A (67, 68) (Fig. 9). The order in which post-translational modifications are installed along the core varies from system to system. For example, processing of the core can happen from the N- to C- terminus like in the

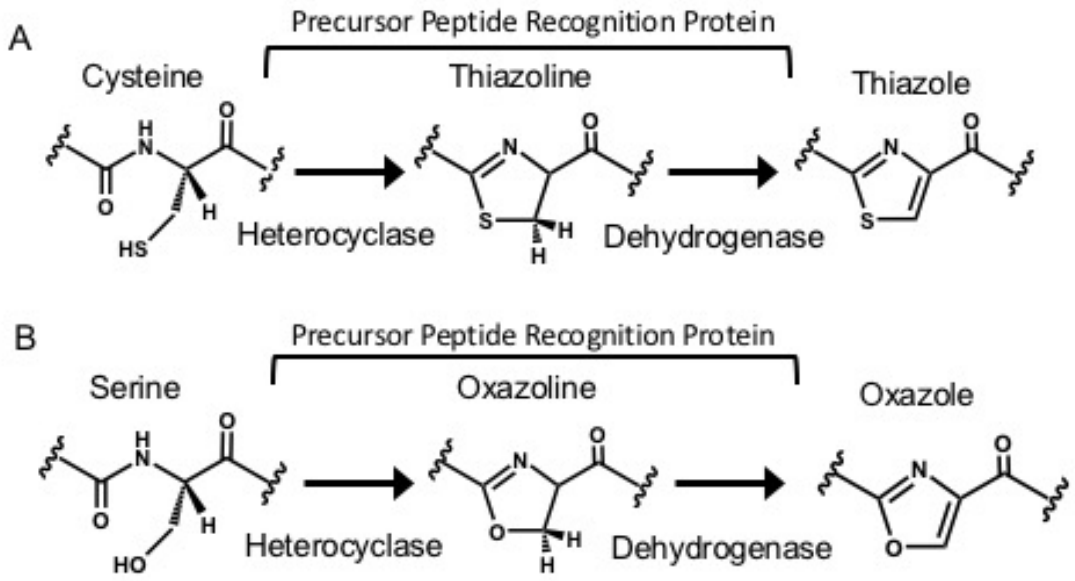


Fig. 8. Step-wise formation of Cys-derived heterocycles**.** *A*, diagram of two step conversion of cysteine to thiazole. First an ATP-dependent heterocyclase converts cysteine to thiazoline, then an FMN-dependent hydrogenase oxidizes the thiazoline into thiazole. *B*, diagram of conversion of two step conversion of serine to oxazole. First an ATP-dependent heterocyclase converts cysteine to thiazoline, then an FMN-dependent hydrogenase oxidizes the thiazoline into thiazole.

microcin B17 pathway (59, 64). However, in other systems, like BalA1, C- to N- terminal processing can be observed(64) and in some systems, like thiamuracin, heterocycles can be installed in a non-linear fashion (68). The level of promiscuity of the processing enzymes also ranges greatly amongst specific pathways (30, 55, 56, 64). RiPP PTM enzymes from cyanobacteria are known to have less stringent regio-, chemo-, and allo- specificity due to the fact that several hypervariable precursor peptide cores are often present in the gene cluster and are each post-translationally modified (19, 29, 58, 66, 69). For example, the cyclodehydratase PatD, involved in patallemide synthesis, has been shown to modify cysteines to thiazolines in many core sequence contexts and lengths just as long as the C-terminal region of the leader peptide is present (70). However, this promiscuity comes at a cost as PatD is slow in converting thiazolines onto cysteines, especially on non-native cores. Conversely, the cyclodehydratase and dehydrogenase in microcin B17 synthesis have rather tight specificity. When mutations are made

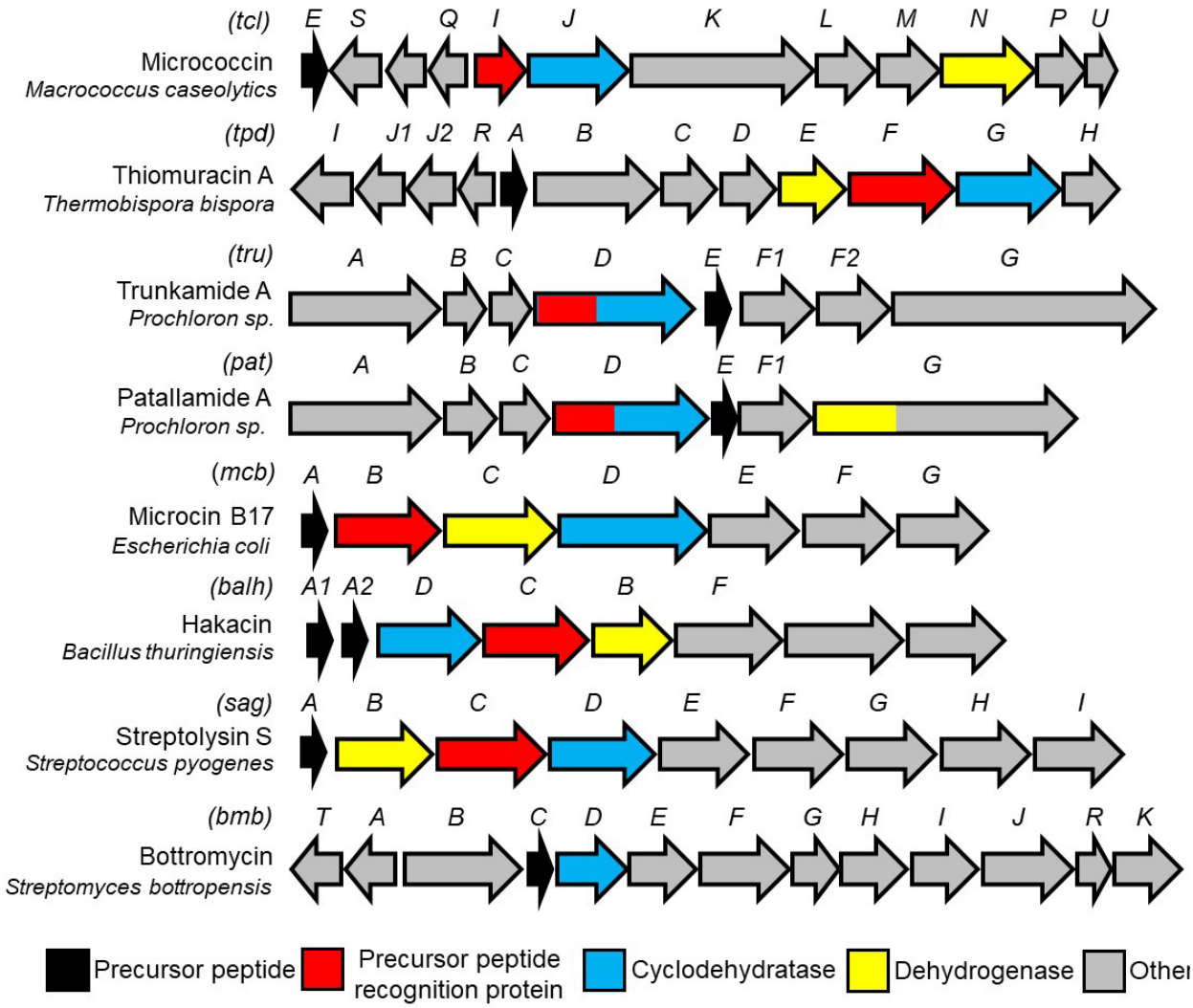


Fig. 9. Comparison of biosynthetic gene clusters for various. Precursor peptide is represented in black. Precursor peptide recognition peptide is depicted in red. Cyclodehydratase genes are shown in blue, dehydrogenase genes are shown in yellow, and all other genes not necessary for heterocycle formation are shown in gray.

to heterocycle flanking residues in the microcin B17 precursor peptide, then azole formation is thwarted (60). The hakacin biosynthetic gene cluster encodes two precursor peptides BalhA1 and BalhA2, similar to cyanobacteria systems (64). This suggests that the heterocycle installing enzymes would be required to be more permissive. In previous studies it was shown that the cyclodehydratase in the hakacin pathway is very efficient at modifying a scrambled BalHA1 core and the microcin B17 core, but that the residue C40 in the leader region of the hakacin precursor peptide is necessary (64).

There are thus no definitive rules for enzyme assembly and processing mechanics for TOMM proteins, and so these systems must be studied independently. As a consequence, many pathway-specific details remain elusive. In some cases, like Microcin B17, the cyclodehydratase, precursor recognition protein and dehydrogenase have been found to form a complex (32, 60). In TOMM systems there is the potential to mix and match different cyclodehyratases, dehydrogenases, and precursor peptide recognition proteins in order to maximize the number of novel peptides that can be generated. Therefore, evaluating how these enzymes work together (whether in a complex or alone) is important. When diverse noncognate dehydrogenases replaced the cognate dehydrogenase in Microcin B17 synthesis, they were not capable of oxidizing any heterocycles (56). This is perhaps a result of the inability of these new dehydrogenases to form a functional complex with the cyclodehydratase and precursor recognition protein. However, in other pathways such as the BalhA1, BalhD (dehydrogenase) was able to act in isolation to install azoles regardless of the presence of BalhC (precursor peptide recognition protein) or BalhB (cyclodehydratases). In addition, non-native dehydrogenases were able to be exchanged for the native dehydrogenase in the hakacin pathway with thiazoles being formed (56). Many details of how the heterocycle-forming enzymes work together in the micrococcin are unknown. This reinforces the remarkably variable nature of these systems and how each system's properties must be characterized on an individual basis.

1.5 Summary of subsequent chapters

This thesis evaluates the use of RiPP biosynthetic systems and their use in engineering novel peptides. It focuses on two different peptide systems: the lasso peptide microcin J25 and the thiopeptide micrococcin. In chapter two I explore the use of microcin J25 as a peptide scaffold for the discovery of novel antimicrobial peptides through an *in vivo* high-throughput

screen. Then, in the last chapter I explore the system specific protein-protein interactions between micrococcin heterocycle installing proteins TclI, TclJ, and TclN with the hope that in future the crucial components of heterocycle formation can be used to generate novel peptides.

CHAPTER 2: Lasso peptides as scaffolds for a high throughput screen of novel antimicrobial peptides

2.1 Introduction

Lasso peptides as peptide scaffolds

Lasso peptides are a family of RiPPs that hold great potential as scaffolds for novel peptide generation. One of the most well-studied yet biosynthetically simple lasso peptides is microcin J25 (MccJ25). MccJ25 possesses antimicrobial activity and is plasmid-encoded in certain *Escherichia coli* strains. Its biosynthetic gene cluster consists of four genes: *mcjA, mcjB, mcjC* and *mcjD.* The gene *mcjA* encodes the precursor peptide, while *mcjB/C* encode the two proteins that comprise the enzymatic knot-tying complex responsible for modifying the precursor peptide (1, 32). McjB acts like an ATP-dependent cysteine protease, cleaving the leader peptide from the core and McjC is involved in the formation of the ring (37). The MccJ25 precursor peptide (McjA) is 58 amino acids long, with a 37- amino acid N-terminal leader peptide and a 21-amino acid C-terminal core peptide (37). A linkage between Glu8 and Gly1 forms the 8-amino acid ring of the lasso peptide while the bulky side chains of Phe19 and Tyr20 provide the steric hindrance necessary to keep the loop in place (37, 71, 72) (Fig. 10).

Previous research has shown that the loop region of MccJ25 can be varied considerably and still be recognized by the knot-tying machinery. This possibility for variability makes lasso peptides ideal for bioengineering. For example, one study created all possible single amino acid substitutions in all positions of the 21-amino acid core peptide excluding Glu8 (one of the amino acids involved directly in ring formation) (Fig. 11) (1). They found that most of the positions of the core peptide, except Gly1, Gly2, Glu8, Phe19, and Tyr20, were able to tolerate various changes to the amino acid sequence while still producing a stable, mature, and exported lasso peptide (73). Several of these positions in the core peptide were able to tolerate over fifteen

different amino acid substitutions. This illustrates that there is a lot of flexibility allowed in the amino acid sequence of the loop region. Lastly, another study investigated the effect that various 3-amino acid substitutions in the loop region of the core peptide had on the function of MccJ25. They found that over half of the mutants retained their function (74). This flexibility was further confirmed when three amino acids in the loop region of MccJ25 were swapped out for an integrin binding motif Arg-Gly-Asp (RGD) in order to generate an MccJ25-RGD mutant that was capable of inhibiting integrin, a potential target for cancer therapy (38, 39). The resulting peptide was successfully converted into an integrin antagonist with nanomolar effectiveness. This demonstrates that lasso peptides can be bioengineered into compounds with pharmacological properties. It has also been shown that you can shorten the loop region and extend the C-terminal tail below the ring without altering the lasso structure (71). All of this past research highlights the vast potential MccJ25 has to be modified and its ability to be used to create novel lasso peptides.

Massively parallel screening for bioactivity of randomized RiPPs *in vivo*

MccJ25 seems to possess the flexibility to be modified to produce thousands of unique lasso structures. In addition, like many RiPPs it possesses antimicrobial activity. Due to the increase of antibiotic resistant bacteria, development of novel antimicrobial peptides is ideal. However, one problem when generating thousands of novel peptides is that only a few would be expected to exhibit antimicrobial activity. Therefore, it is necessary to be able to screen through vast numbers of peptides for those that appear to be toxic. We have developed a method to screen through hundreds of thousands of peptides using massively parallel peptide activity screening (MPAS) (Fig. 12). The basis of this platform is to generate a library of hundreds of thousands of variant peptides in a plasmid-based inducible expression system. Random-insert

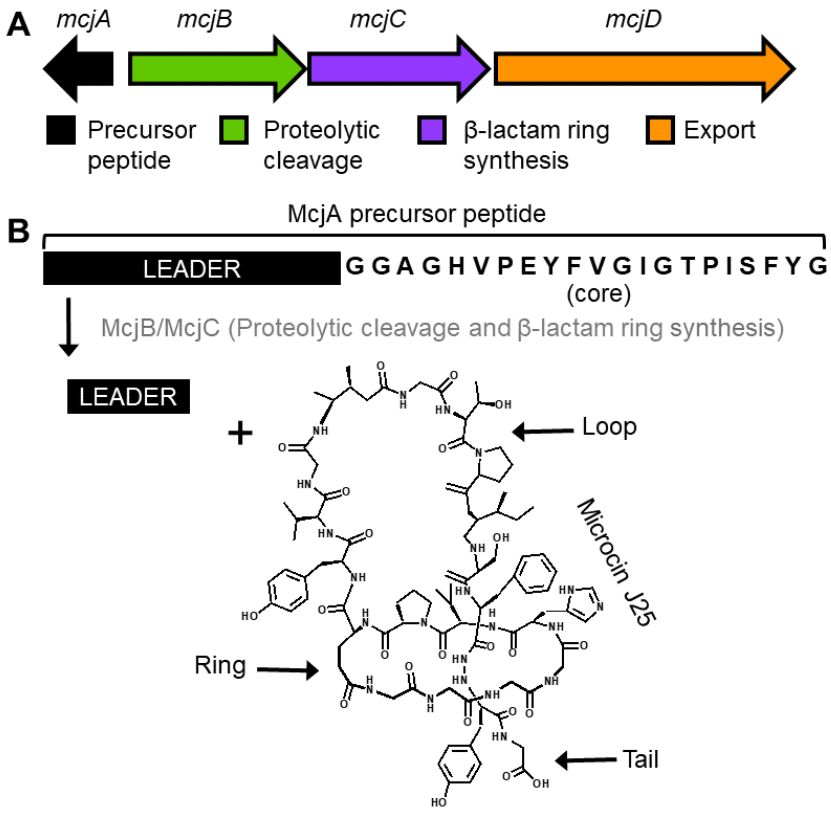


Fig. 10. Microcin J25 biosynthesis. *A*, Microcin J25 biosynthetic gene cluster. *B*, Simple diagram depicting the microcin J25 maturation pathway.

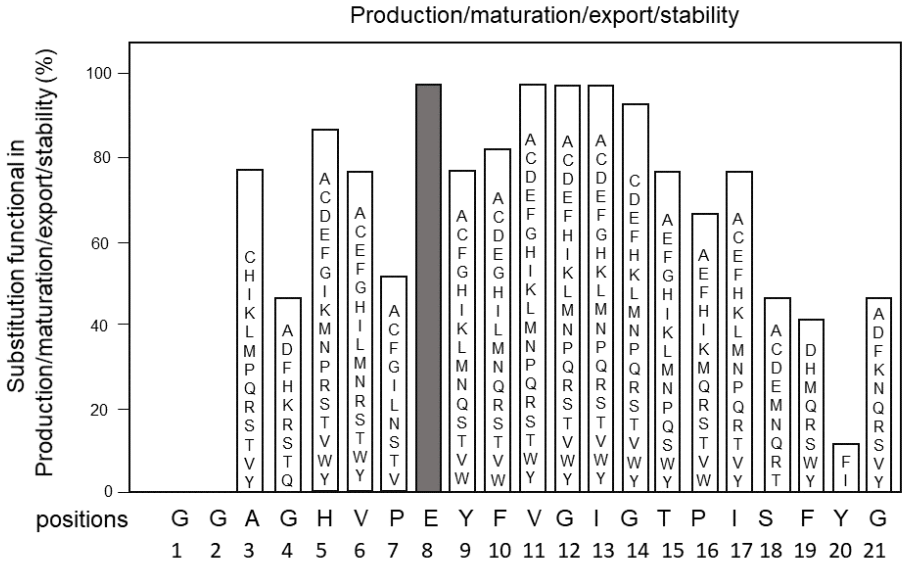
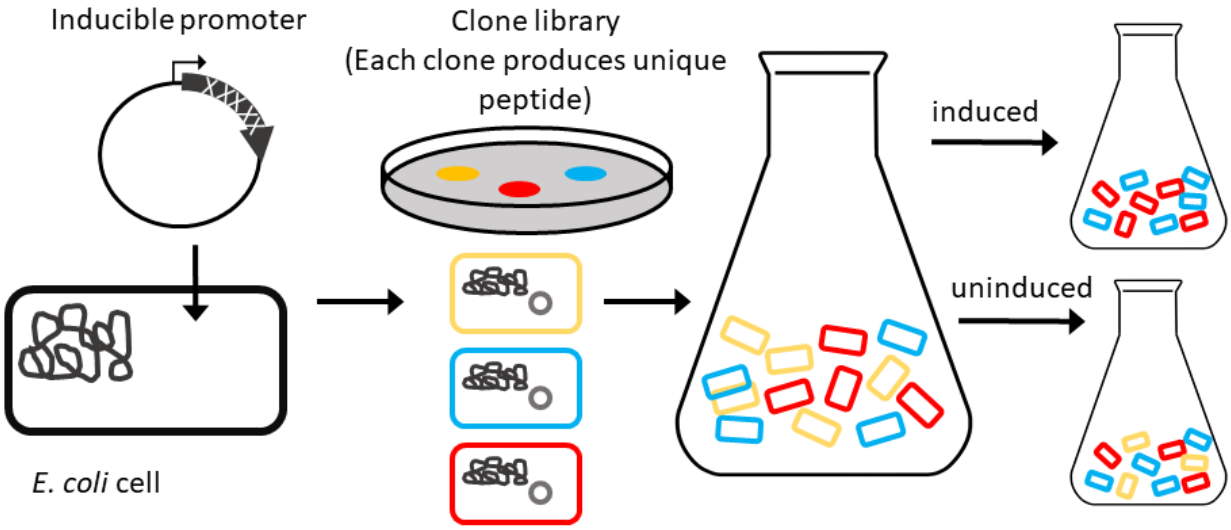


Fig. 11. Single amino acid mutation study of McJA from Pavlova, et al**.** Researchers in Pavlova's group performed 380 single-amino acid substitutions. Amino acid noted in the bars are tolerated mutations for production, maturation and export of MccJ25. The native sequence of McjA core is shown below the chart. (1)

plasmids are transformed into *E. coli* cells to create a library, where each clone is effectively 'assigned' to produce only one peptide variant. As a result, the *E. coli* library contains thousands of different peptide variants. Once the library is created, it is mixed and grown under conditions with or without inducer, turning on or off the expression of the potential antimicrobial peptides. Next, Illumina sequencing is performed to obtain DNA sequence data for surviving peptide-encoding plasmids. When a cell is assigned a toxic peptide, it will be depleted in the population, eventually causing the DNA sequence for that peptide to be lost from the culture. By comparing the sequences of the induced culture to the un-induced culture, one can identify many potentially toxic peptides, or "hits", by looking for peptide-encoding sequences that are depleted in the induced culture compared to the un-induced. In our lab, we have carried out proof-of-principle work using this MPAS platform on a linear peptide library and obtained hundreds of hits.



PEPSEQ	REF1	REF2	TEST1	TEST2
GLSCYHDVV	126	138	110	93
GLFINNCGI	126	134	119	106
GVIIVYGIR	126	131	0	2
GVDYYGSCH	126	130	162	125
GVVHRCDFR	126	97	390	295
...	...		...	

Fig. 12. Overview of massively parallel activity screening (MPAS). See text for more details.

In order to perform the future screen to generate non-linear peptides based on RiPP scaffolds (such as a lasso peptide scaffold), an inducible plasmid system must first be developed which allows expression of variant peptides. It is important that the inducible system is extremely tightly controlled so that there is next to no expression of the peptides when uninduced. This is desirable because the peptides that are of the greatest interest are those that will kill the cell. Therefore, if the system is too leaky then we may get antimicrobial peptide activity in un-induced conditions particularly if a peptide is especially toxic, making it difficult for us to identify those highly sought-after toxic peptides. We discuss below a system in which microcin J25 can be used as a scaffold to generate thousands of novel lasso peptides that can then be screened for antimicrobial activity by MPAS methods. We evaluate the strength and tightness of six different *E. coli*-derived promoters as well as develop a two-plasmid system for lasso peptide expression using molecular cloning techniques.

2.2 Experimental Methods

Strains and culture conditions

The bacterial strains and plasmids used in this chapter can be found in Tables 2 and 3. Plasmids were constructed and maintained in *Escherichia coli* strain DH5α. For growth curves, strains were either based in DH5α or MDS™42 Meta ΔrecA *E. coli* (Scarab genomics). All bacterial cultures were grown in Luria broth (LB: per liter, 10 g Bacto tryptone, 5 g Bacto yeast extract, 5 g NaCl, 1 ml 2N NaOH) or minimal defined media (MDM: 0.06 M NaCl, 0.01 M KH_2PO_4, 0.01 M NH_4Cl, 0.5% glycerol, 0.002 M $MgSO_4$, 0.02 mL 10mM $CaCl_2$, 0.2 ml Trace minerals (200 ml H_2O, 400 mg EDTA, 300 mg $FeSO_4$, 50 mg $MnSO_4$, 50 mg $ZnSO_4$, 25 mg H_3BO_3, 25 mg $CoCl_2$, 5 mg Na_2MoO_4, 5 mg $CuSO_4$). Antibiotics used were kanamycin (30 mg/ml) and chloramphenicol (30 mg/ml). Cultures were induced for protein expression using 0.3

mM isopropyl β-D-1-thiogalactopyranoside (IPTG), 0.3% L-rhamnose, 0.3% L-arabinose, or 0.4 µM anhydrotetracycline (ATc).

Plasmid construction

The *mcj* genes were synthesized with codon optimization for expression in *E. coli*. Plasmid construction was carried out using standard molecular cloning procedures. Annotated vector and insert sequence files are given in supporting information file. Complete vector sequences are given in the Supplementary Information. Plasmids and strains employed in this study are given in Supplementary Table 1 and 2.

Fluorescent assays

Cultures of promoter strains were grown overnight and subsequently 20 µl were used to inoculate 4 ml LB cultures. These cultures were grown with and without inducer for 8 hours. Optical density of the cultures was measured at 600 nm (OD_{600}). Cultures were then diluted in LB and normalized to the lowest OD_{600}. Arbitrary fluorescent units (AFU) were determined by reading a white polystyrene 96-well plate on a BioTek plate reader at 509 nm. Final AFU values were normalized to vector-only and blank LB controls.

Miller assays

Strains were grown in LB with and without inducer for approximately 4 hours. Then OD_{600} measurements were recorded for individual cultures. For each assay, 600 µl of β-Gal master mix consisting of 600 µl Basal Buffer (Basal Buffer: 500 ml H_2O, 4.3 g Na_2HPO_4, 2.4 g NaH_2PO_4, 0.75 g KCl, pH to 7.0 with about 1.5 ml 2N KOH, 1 ml 1M $MgSO_4\_7H_2O$, 1 ml of $CHCl_3$), 0.5 mg ortho-Nitrophenyl-β-galactosidase (ONPG), 0.5 µl 10% sodium dodecyl sulfate (SDS), 3 µl β-mercaptoethanol (BME), 50 µl chloroform, and diluted culture were combined and timer was started. Samples were incubated at 32°C until bright yellow color was observed in 1 or

more samples. Stop buffer (1M Na_2CO_3 with 1 ml $CHCl_3$) was added to stop reactions. Samples were then vortexed and centrifuged to pellet chloroform and other cellular debris. OD_{420} measurements were taken and Miller units were calculated using the equation $(1000*OD_{420})/(OD_{600}*t*v)$ where t= assay time in minutes, v=volume of cells added in ml corrected for the dilution factor that was performed.

Growth curves

Strains were grown overnight, then a 20 µl sample was used to inoculate 4 ml cultures of either LB (Fig 14*BC*) or MDM (Fig. 14*B*) with and without their appropriate inducers and with antibiotics. Then OD_{600} measurements were taken at specific time points using a spectrophotometer.

2. 3 Results

Promoter optimization

There are various options in choosing an inducible promoter for vector expression systems. For our MPAS experiments we wanted an inducible promoter with extremely low background activity when uninduced, but that also yields high expression when induced. Therefore, we evaluated the expression profile of several common *E. coli*-derived promoters: P_{araBAD}, P_{rhaBAD}, two variations on P_{lac} (P_{lacT5}(*lacIWT*), P_{lacT5} (*lacIq*)), and P_{tet} (Fig 13). The two different lacT5 promoters vary in the constitutive promoter region of the lac repressor (LacI). P_{lacT5} (lacIq) has an altered -35 region which makes it more canonical which should increase expression of the repressor. We tested each promoter in the context of two different reporters: β-galactosidase (originally encoded by the gene *lacZ* of the *lac operon)* and monomeric superfolder green fluorescent protein (msfGFP). First, in the data from the Miller assay measuring β-galactosidase, we see that P_{rhaBAD} and the two lac promoters appear to give the

lowest amount of background uninduced expression while still leading to a large amount of expression when induced (Fig. 13*F*). Considering the fluorescent data control of expression for P_{rhaBAD} is extremely tight yet very strong (Fig. 13*G*). By fluorescence-based assay, the lacT5 promoters yield very low expression of msfGFP while giving the most background expression when uninduced. Interestingly, the P_{tet} promoter shines as being a very strong promoter with low uninduced expression when fluorescence is measured. P_{araBAD} is arguably the leakiest and least strong promoter of the group.

A synthetic inducible MccJ25 expression system

We wanted to develop a two-plasmid system with distinct promoters in order to selectively turn on expression of either the precursor peptide McjA or the modifying enzymes McjBC or both. From the promoter data, P_{rhaBAD} and P_{tet} promoters were chosen to express McjA and McjBC respectively. While MccJ25 does originate from a single *E. coli* plasmid we wanted to confirm that our two-plasmid based system still yielded functional MccJ25. When McjA and McjBC were co-expressed, cultures did not appear to grow compared to control strains (Fig. 14*A*). This is indicative that microcin J25 is being produced and its antimicrobial activity is killing the cells. More detailed growth curves confirm that when expression of all three Mcj proteins is induced, growth of cultures is inhibited compared to uninduced and control strains (Fig. 14*BC*). This also confirms that our promoters are tight because we are getting normal growth when uninduced.

Preparing McjA for easy core peptide randomization

In future MPAS experiments the core region of McjA will be randomized. In order to make future cloning steps easier we wanted to add a unique BamHI site into the beginning of the core

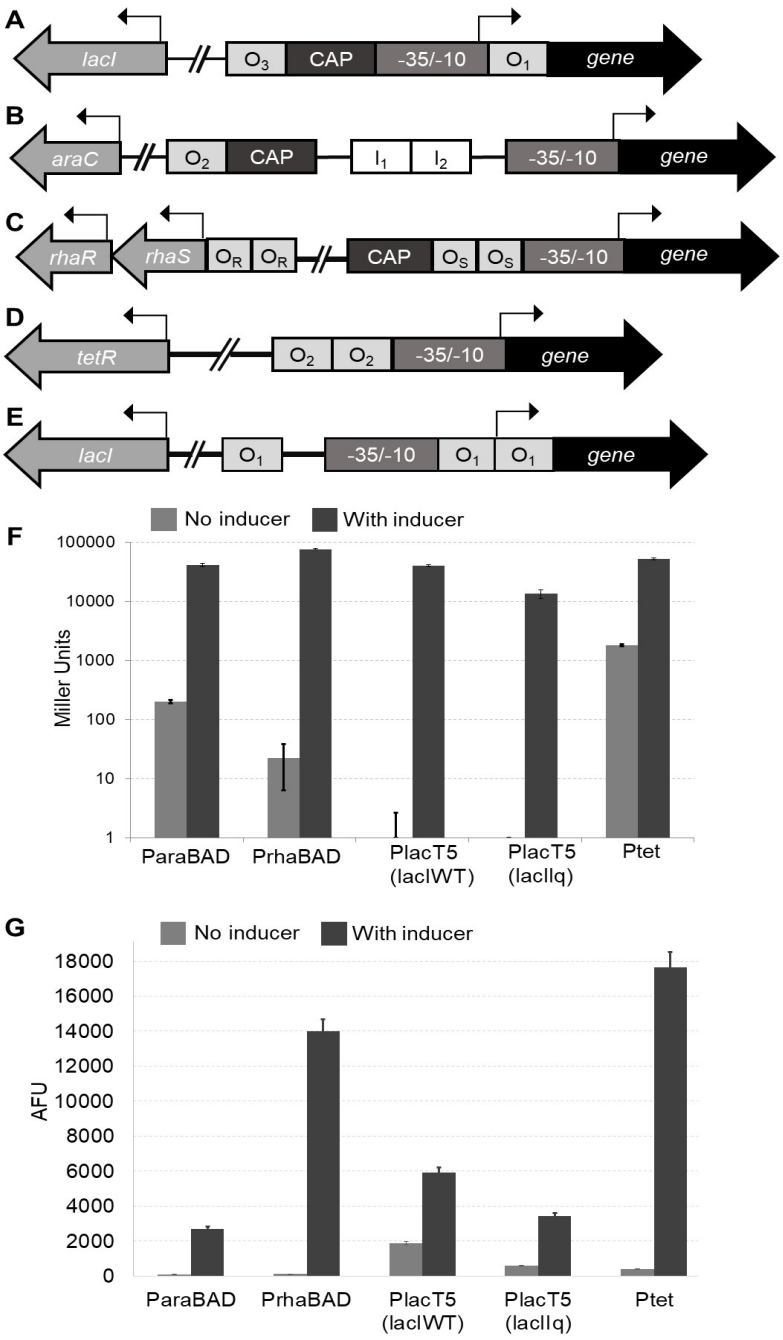


Fig. 13. Promoter optimization for *E. coli* expression**.** Inducible promoters. lacI, araC, rhaR, rhaS, and tetR all encode regulatory proteins. These regulatory proteins bind to operating sites indicated by O_x. Promoters are indicated by right angled arrows above genes. CAP represents the CAP/cAMP binding site. RNA polymerase recognition sequence is represented as the -35/-10 region. A, Lac operon. B, arabinose operon. C, rhamnose operon. D, tet operon E, Lac T5 promoter F, Results from Miller assay on promoter strains driving the production of β-galatosidase. Each condition for individual strains was done in triplicate and in LB. Plasmids were in DH5α. F, Results from Miller assay on promoter strains driving the production of β-galatosidase. Each condition for individual strains was done in triplicate and in LB. Strains were in DH5α. G, Results from fluorescent assay on promoter strains driving the production of msfGFP. Each condition for individual strains was done in triplicate and in LB. Strains were in DH5α.

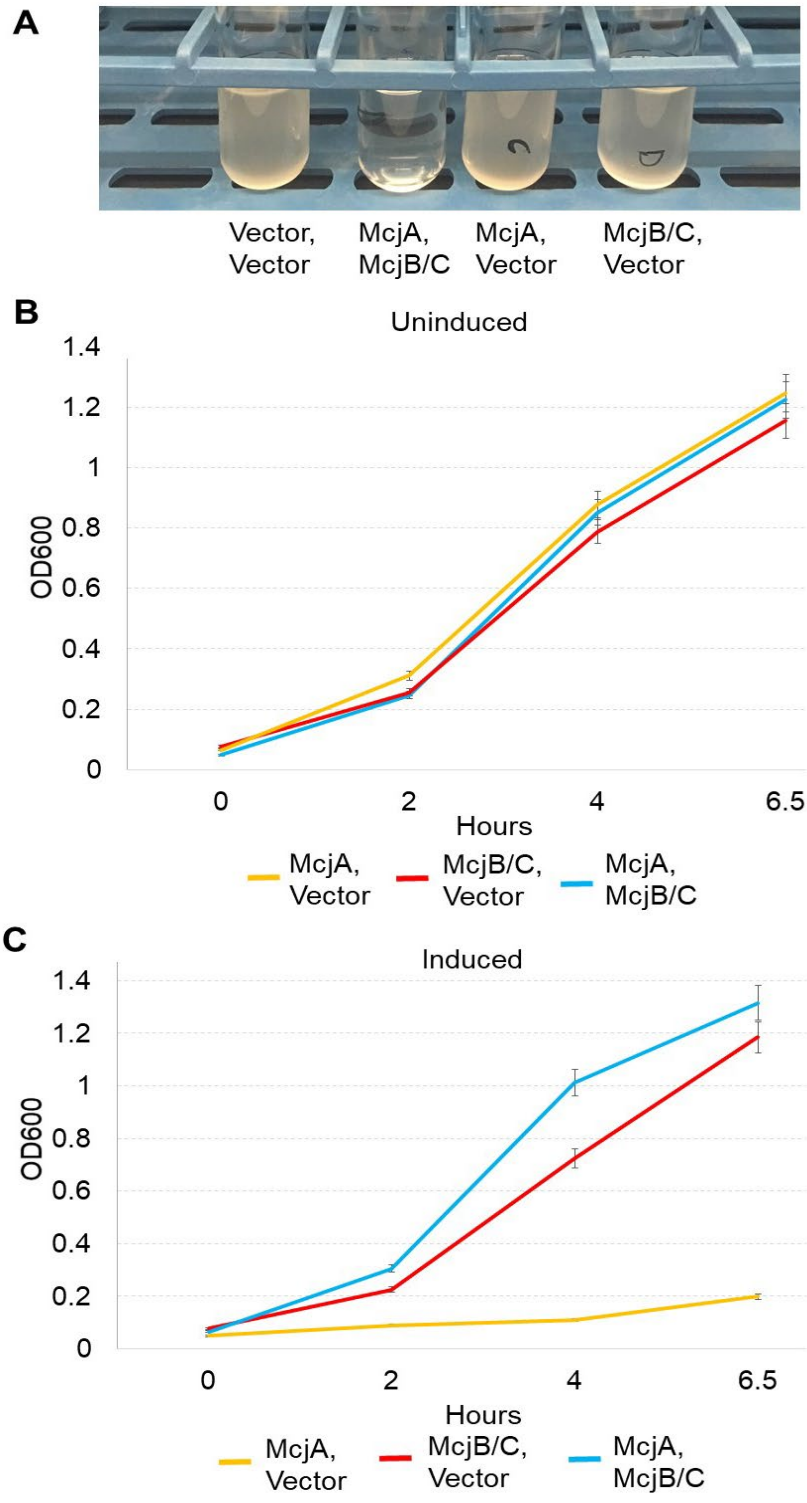


Fig. 14. Reconstitution of MccJ25 expression in *E. coli*. *A*, picture of induced cultures in MDM after 24hrs of growth. When McjABC are co-expressed growth is inhibited. *B*, growth curve of uninduced cultures in LB. *C*, growth curve of induced cultures in LB.

(Fig. 15*A*). This would allow us to easily switch in and out new core sequences. Altering the core to include a BamHI site causes an alanine to serine mutation on the third residue in the core region of McjA. Although we believed that this mutation would be tolerated based on previously described data (Fig. 11) (1), we wanted to confirm this to be true. Growth curves performed with the expressed altered McjA in conjunction with McjBC indicated that active MccJ25 was being created compared to the uninduced and vectors only strain control (Fig. 15*B*).

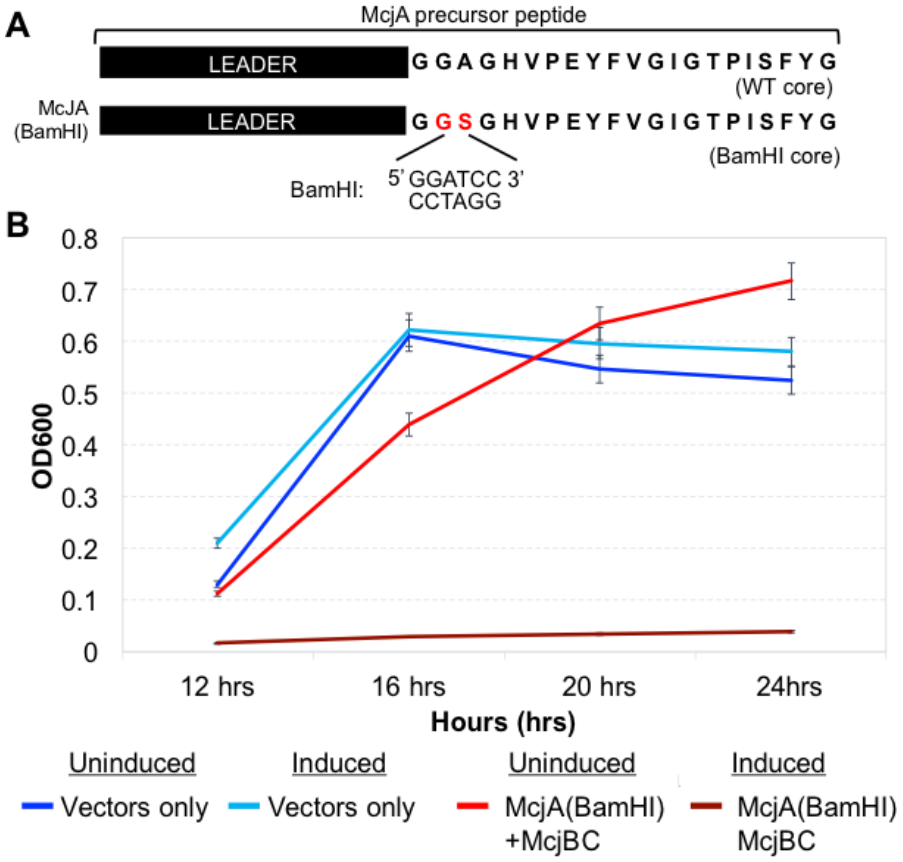


Fig. 15. Altering the McjA core peptide to accommodate a unique BamHI site. *A*, graphical representation of where the BamHI site wlll be inserted into the core and the subsequent mutation it will cause. *B*, Growth curves confirming production of MccJ25 with altered core including BamHI site. Cultures were grown in MDM.

2.4 Discussion

In future, a small lasso peptide library of hundreds of thousands of clones could be generated by randomizing a nine amino acid section of the loop region of MccJ25. The loop

region can be randomized by genetically modifying *mcjA* (Fig. 16*A*). The library insert randomizes the amino acids from tyrosine to isoleucine (Y9 to I17) in the core portion of the peptide. The nucleotide sequence of the random region of the loop is VVTNHTVVTNHTVVT NHTVVTNHTVVT (Fig. 16*B*). This sequence avoids getting peptides that possess long string of hydrophobic amino acids. Long strings of hydrophobic amino acids are naturally more antimicrobial due to their ability to disrupt cell membranes. We want to avoid these peptides because we want to generate peptides with potentially novel and specific mechanisms of action. This sequence also avoids any stop codons and allows for a variety of charged, polar and nonpolar amino acids. The plasmid containing the modified *mcjA* would be transformed into a strain of *E. coli* which has already been transformed with the plasmid containing the modifying enzymes. Clones would then be selected on LB without inducer and combined into a mixed population for subsequent MPAS experiments.

Some of the hits generated from the MPAS experiment indicating that a peptide is toxic would subsequently need to be confirmed as toxic and assessed for lasso-ization by McjB or McjC. This lasso formation could be assessed by either mass spectrometry or potentially simple gel-shift analysis. When gel electrophoresis is performed, there should be a characteristic band shift when the precursor peptide gets cleaved by McjB. The strep-GB1 tag is located on the N-terminus of the precursor peptide, and the cleaved leader peptide is missing the 21 amino acid core peptide (2.1 kDa loss) which should lead to a shift in the gel. This method assumes that if McjB is capable of recognizing and cleaving the precursor peptide then McjC should be able to modify the core to produce the lasso peptide. This system is ready to be used in an MPAS experiment and library insert oligos have been purchased; however, MPAS technology was

being further validated using linear peptides for the second time and I moved onto work that will be presented in Chapter 3.

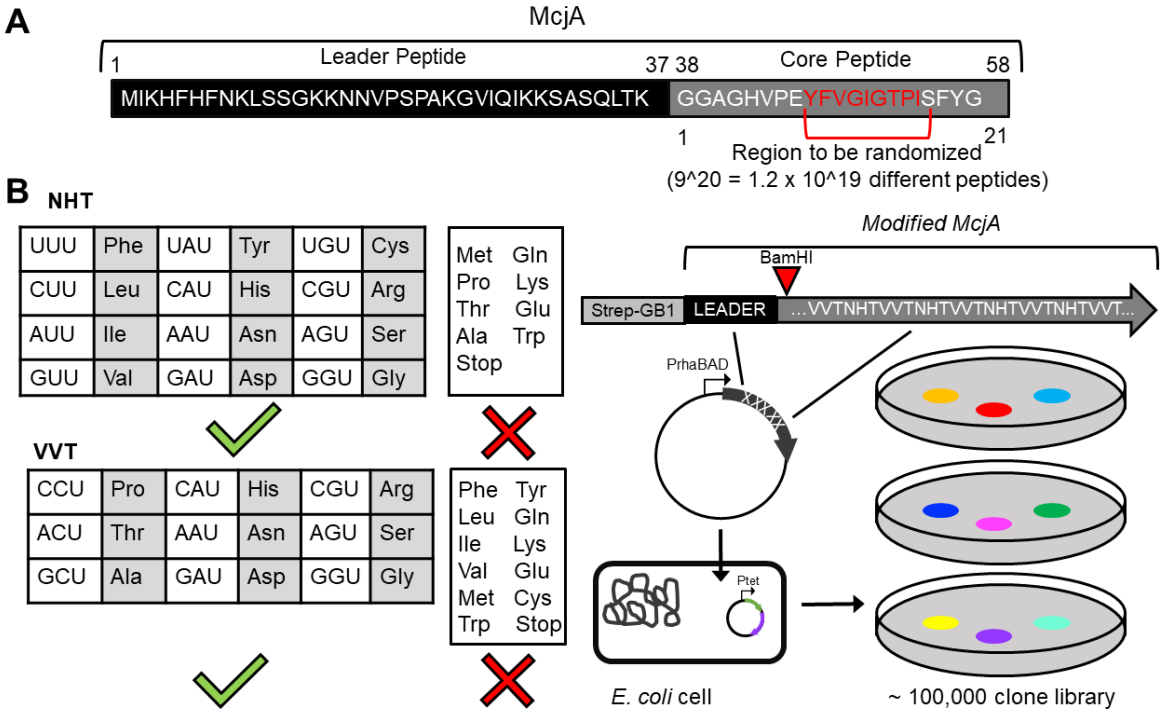


Fig. 16. Lasso Peptide Library Design. A, Detailed description of the exact region of the McjA core that will be randomized. B, Chart illustrates which amino acids are allowed with NHT and VVT. Allowed amino acids are shown in the table with the checkmark below while amino acids that are not allowed are present in the box with the x below. VVTNHT allows for various amino acids, but eliminates the ability to have a long string of hydrophobic amino acids in a row or any stop codons.

CHAPTER 3: Mapping interactions between micrococcin biosynthetic enzymes

Devan M Bursey[1], Kathryn D Bewley[2], Philip R Bennallack[1], Susan S Miller[2], Joel S Griffitts[1]

[1]Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602 and [2]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

3.1 Abstract

Thiopeptides are bacterially produced bioactive compounds derived from ribosomally synthesized peptides that are extensively post-translationally modified. We have recently elucidated the biosynthetic pathway for the antibacterial thiopeptide micrococcin. This pathway is initiated when six Cys residues in the TclE precursor peptide are converted to heterocyclic thiazoles, through the action of the ATP-dependent cyclodehydratase enzyme TclJ (to form thiazolines), followed by the FMN-dependent dehydrogenase TclN (to form thiazoles). A non-enzymatic protein, TclI, is involved in heterocycle formation; however, its exact role has been unclear. Here we employ structural modeling and biochemical co-purification methods to investigate TclI as an adaptor protein that couples the TclJ and TclN enzyme to the TclE substrate peptide. TclI appears to be a two-domain protein, with the N-terminus binding to TclE and the C-terminus binding to TclJ and TclN. TclI binds to each of its three partners through independent interaction surfaces. We also show that an N-terminal portion of TclJ, in isolation, can associate with TclI.

3.2 Introduction

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are microbially produced natural products that are initially synthesized on the ribosome, and subsequently decorated with numerous post-translational modifications (PTMs) to yield bioactive compounds (33, 35). RiPP precursor peptides generally contain an N-terminal leader sequence involved in recruitment of the enzymes that carry out PTM installation, as well as a C-

terminal “core peptide” region that is the site of the modifications. The final step in RiPP biosynthesis often involves removal of the leader peptide from the modified core. Conversion of core peptide Cys, Ser, and Thr residues to azolines, and their subsequent oxidation to azoles are common steps in RiPP biosynthesis, particularly for cyanobactins (e.g. patallamide and trunkamide), linear azol(in)e containing peptides (LAPs) (e.g. microcin B17 and streptolysin), and thiopeptides (e.g. thiomuracin, and micrococcin) (28, 47, 50). Thiopeptides constitute a large RiPP family with over 100 known members possessing diverse biological activities, including antimicrobial functions (41, 43). A typical thiopeptide contains several azol(in)e structures, with Cys-derived thiazoles being most prominent. Thiopeptides also feature at least one macrocycle anchored by a six-membered nitrogenous ring derived from dehydroamino acids such as dehydroalanine (41, 43, 44). Thiopeptides may also contain other diverse modifications, including C-terminal decarboxylation (micrococcin (46)), methylation (nocardithiocin (75)), hydroxylation(thiocillin (76), nosiheptide (77)), and glycosylation (thiazomycin (78), glycothiohexide (53)).

The microccocin biosynthetic pathway

Microccocin was the first characterized thiopeptide produced by *Bacillus* and *Macrococcus sp*. and its biosynthesis involves several PTMs including thiazoles, dehydroamino acids, C-terminal decarboxylation, and formation of a pyridine ring (43, 45). The *M. caseolyticus* micrococcin gene cluster is plasmid-encoded and consists of 12 *tcl* genes, suggesting that it is simpler than the analogous gene cluster in *Bacillus cereus* ATCC (Fig 17*A*) (45). The micrococcin precursor peptide TclE is divided into the 35 amino acid residue leader and the 14 amino acid core. Micrococcin biosynthesis begins with the chemical alteration of all six cysteine residues present in the core of TclE to thiazoles. This conversion of Cys to thiazole licenses the

core for subsequent modifications (Fig 17*B*) (42, 45). Each thiazole conversion is a two-step process that necessitates the activity of three proteins: TclI, TclJ, and TclN (42, 45). First TclJ, a Thiazole/Oxazole-Modified Microcin (TOMM) cyclodehytradase, alters cysteines to thiazolines in an ATP-dependent manner (52, 53). Subsequently, TclN, an FMN-dependent TOMM dehydrogenase, oxidizes the heterocycles to thiazoles (Fig 1*C*). TclI is a required non-enzymatic auxiliary protein in the heterocycle forming process (48, 79).

TOMM biosynthetic genes are prevalent in many phyla of bacteria and as a result are often genetically diverse (48). Therefore, the specific details of each biosynthetic pathway must be elucidated individually. While the general function of these thiazole incorporating enzymes in micrococcin synthesis is known, information on how these proteins in this system interact with one another is still missing. In this paper, we investigate the general protein-protein architecture of enzymes TclI, TclJ, TclN, and TclE in micrococcin biosynthesis, focusing on the predicted role of TclI as a potential docking protein that interacts independently with TclJ, TclN and TclE. We used a plasmid-based expression system in *E. coli* to study the interactions of TclIJNE primarily by His_6-TclI co-purification, SDS-PAGE and predicted structural modeling.

3.3 Experimental Procedures

Strains and culture conditions

The bacterial strains and plasmids used in this study are summarized in Tables S4 and S5 Plasmids were constructed and maintained in *Escherichia coli* strain DH5α. For protein expression, plasmids were transformed into strain BL21, Nico 21 (DE3) or DH5α. All bacterial cultures were grown in Luria broth (LB: per liter, 10 g Bacto tryptone, 5 g Bacto yeast extract, 5 g NaCl, 1 ml 2N NaOH). Antibiotics used were kanamycin 30 mg/ml, ampicillin 100 mg/ml, and

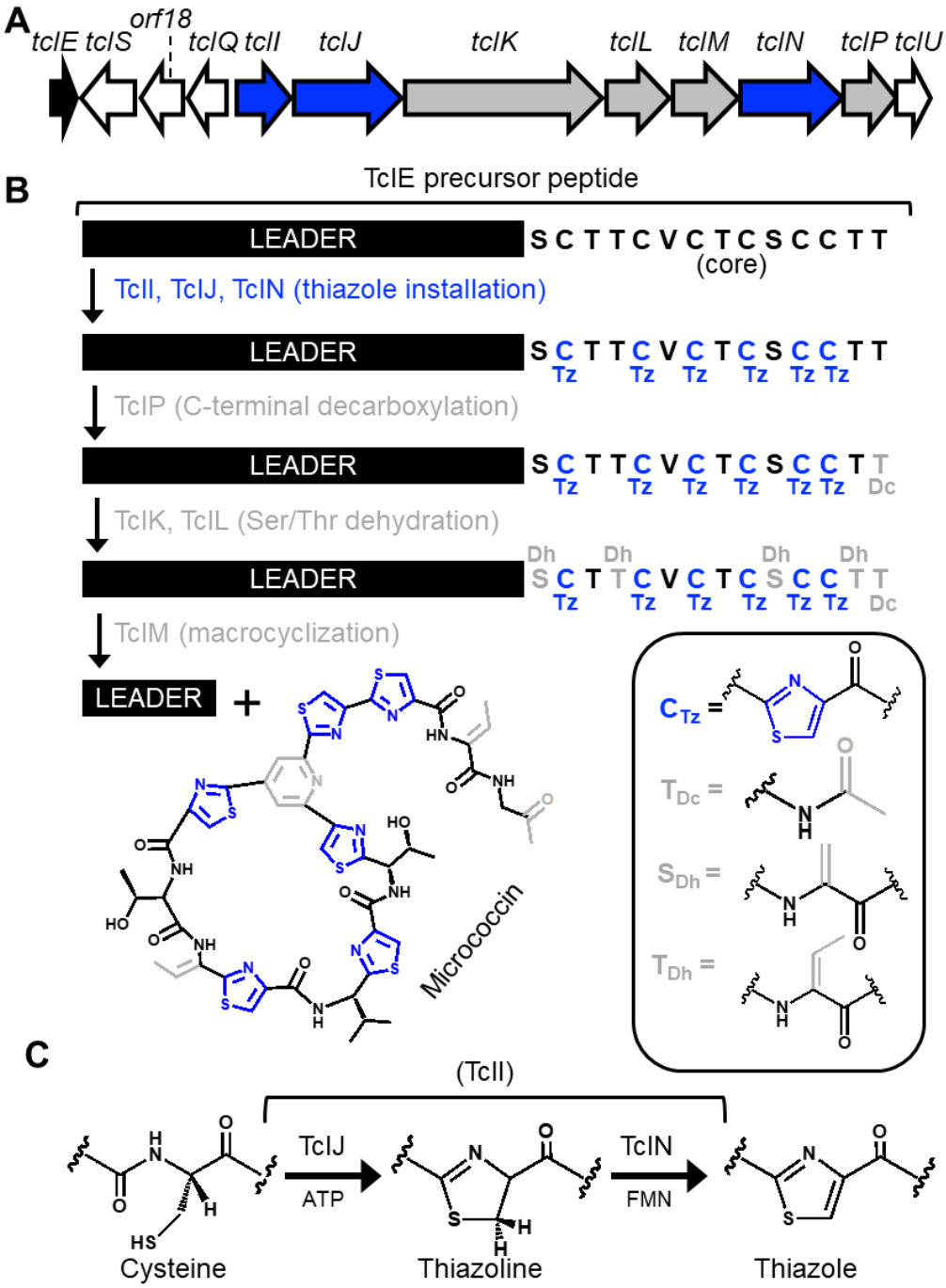


Fig. 17. Micrococcin Synthesis. A, A map of the native *tcl* gene cluster from *M. caseolyticus*. The gene encoding the precursor peptide (TclE) is colored black, and the genes encoding proteins for thiazole installation (TclI, TclJ, TclN) are colored blue. Other essential proteins for complete micrococcin production are shown in gray. Nonessential proteins are shown in white. B, Overview of the micrococcin biosynthetic pathway, with TclE core peptide modifications brought about by the thiazole installation machinery indicated in blue, and all subsequent modifications indicated in gray. Abbreviations: Tz, thiazolyl; Dc, decarboxyl; Dh, dehydro. C, Two-step conversion of TclE Cys residues to thiazole via TclI, TclJ, and TclN.

chloramphenicol 30 mg/ml. Cultures were induced for protein expression using 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Plasmid construction

The *tcl* genes were synthesized with codon optimization for expression in *E. coli*. Plasmid construction was carried out using standard molecular cloning procedures. Annotated vector and insert sequence files are given in supporting information file. Complete vector sequences are given in the Supplementary Information. Plasmids and strains employed in this study are given in Supplementary Table 3 and 4.

Tcl protein expression and purification

To prepare samples for SDS-PAGE analysis, overnight liquid cultures (4 ml) were grown from single colonies in the presence of appropriate antibiotics. 100 ml cultures were inoculated with 2 ml of overnight culture, allowed to grow at 30 °C for 1 h, followed by induction for another 8 h at 30 °C. Cells were collected by centrifuging 50 ml of the culture, and cell pellets were frozen at -80 °C for a minimum of 1 h. Cell pellets were then subjected to protein purification with either Ni-NTA-linked (for the His6 tag) or glutathione-linked (for the GST tag) resin.

For His_6 purification cell pellets were thawed on ice and re-suspended in 1.5 ml of lysis buffer (50 mM HEPES, 300 mM NaCl, 0.2 % Triton X-100, 0.5 mg/ml lysozyme, 40 mM imidazole, 1 mM EDTA). Lysis took place for 1 h at 4 °C. Then cell lysates were sonicated 2 × 20 sec at power 3 using a probe sonicator to ensure complete lysis. Samples were centrifuged at 13,000 rpm for 9 min (4 °C) and approximately 1 ml of supernatant was transferred to a new micro centrifuge tube. Supernatant was incubated end over end with 50 µl of NTA-nickle

agarose beads at 4 °C for 30 min. Unless stated otherwise in figure legends the beads were washed 3 × 1 ml wash buffer (60mM imidazole, 300 mM NaCl, 50 mM HEPES) and eluted in 50 µl of 2x SDS sample loading dye (20% glycerol, 83 mM Tri pH 6.8, 40 mg/ml sodium dodecyl sulfate (SDS), 0.01% bromophenol blue, 0.03 µl/ml 2-mercaptoethanol).

For GST purifications cell pellets were thawed on ice and re-suspended in 1.2 ml lysis buffer (GST buffer (50 mM Tris 8.0,150 mM NaCl), 0.5 mg/ml lysozyme, 2 mM EDTA, 0.2% Triton X-100, 1/100 protease inhibitor (Sigma)). Cells were lysed at room temperature for 15 min then placed on ice where dithiothreitol (DTT) was added to a final concentration of 1.5 mM. Samples were sonicated 2 × 20 sec at power level 3. Cell lysates were centrifuged at 13,000 rpm at 4 °C for 9 min to pellet cell debris. Approximately 1 ml of supernatant was transferred to a new microfuge tube. 50 µl of unwashed glutathione-agarose beads were added and samples were rotated end over end for 45 min at 4 °C. Slurry was pelleted at 13,000 rpm for 10 seconds. Supernatant was removed and beads were washed 3 × 1 ml GST buffer. GST buffer was completely removed and proteins were eluted from resin in 75 µl of 2X SDS loading dye.

Purified samples were heated at 100°C for 2 min then centrifuged. Unless stated otherwise 5 µl of supernatant was loaded onto a 12% resolving Laemmli gel with a 4% stacking gel. Gels were run using 1x Laemmli running buffer and were stained overnight in coomassie blue stain then destained before pictures were taken.

Mass spectrometry analysis of TclE processing

For purification of His tagged enzymes for mass spectrometry analysis 25 ml overnight cultures were grown from single colonies. These overnight cultures were then used to inoculate 1 L growths (30 °C). After 1 h, IPTG was added and the cultures were grown for an additional 6 hours. The cells were harvested by centrifugation and the cell pellets were frozen at –80 ºC

overnight. For co-purification of His6-TclIJ/His6-TclIN/His6-TclIJN the cells were then thawed on ice with the addition of lysis buffer (50 mM HEPES, 150 mM NaCl, pH 7.8). A protease inhibitor tablet (Roche), 0.2% Triton X-100 and 0.5 mg/ml lysozyme were added and the cells were incubated on ice for 1 h. Complete lysis was achieved by sonication for 2 min on ice using a Branson Sonifier 450, followed by centrifugation for 20 min at 32,539 × g. The supernatant was incubated with 1 ml of Talon resin for 30 min at 4 °C. Resin was washed with 3 × 10 ml lysis buffer, followed by elution with lysis buffer plus 75 mM imidazole (4 × 1 ml). The elution fractions containing protein were buffer exchanged back into lysis buffer, concentrated and flash frozen with 10% glycerol and stored at –80 ºC.

For purification of GST-tagged TclE, 30 ml cultures were inoculated with 1 mL overnight culture, grown at 37°C until an OD_{600} = 0.6, then IPTG was added and the cells were grown for an additional 20 hrs at 25°C. The cells were harvested via centrifugation and the cell pellets were frozen at –80 ºC for at least 30 min. The cells were then thawed and resuspended in 1 ml lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 0.5 mg/ml lysozyme, 2 mM EDTA and one Roche protease inhibitor tablet per 10 ml). Complete lysis was achieved after a 15-min incubation at room temperature (RT) in lysis buffer, followed by addition of 1.5 mM dithiothreitol (DTT) and several short sonication pulses with a microtip. The disrupted cells were pelleted at 7,000 × g to remove insoluble material and the supernatant was combined with 30 µl of glutathione-agarose resin (slurry) at 4 ºC for 45 min (rotating). The resin was pelleted and the beads were washed with 50 mM Tris pH 8, 150 mM NaCl three times and the peptide was eluted with 40 µl buffer plus 10 mM reduced glutathione. The eluant was either frozen at –80 ºC for later use, or directly treated with tobacco etch virus (TEV) protease and ZipTipped (using the manufacturer's instructions.)

Activity of Tcl enzymes was tested in vitro. 20 µl reactions containing 20 mM GST-TclE, 5 mM DTT, 2 mM ATP, 20 mM MgCl2, 1 mM enzymes, and 1 mg TEV protease, were allowed to react for 40 min at RT. Reactions were zip-tipped (using the manufacturer's instructions) and analyzed by matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF) abbbreviation MS.

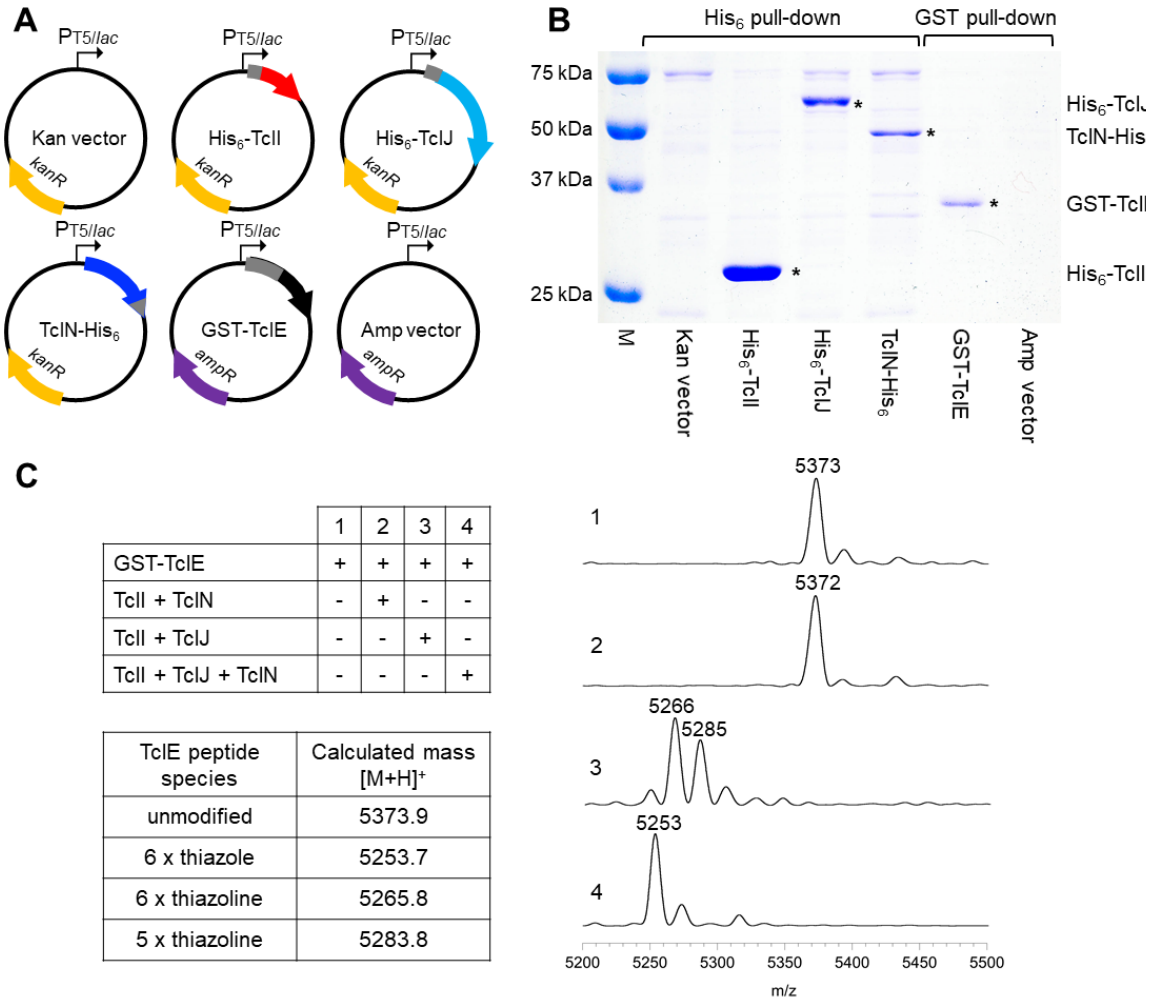
Modeling of Tcl protein structures

Sequences for TclI, TclJ, and TclN were submitted online to the I-TASSER (Iterative Threading ASSEmbly Refinement) server (80) to predict protein structure. The models with the highest C-score were chosen for each Tcl protein. The top-ranked PDB structural analogs for each each Tcl protein (4BS9, 3E07) were used as a comparative reference. The structural models and reference models .pdb files were downloaded and viewed in USF Chimera(81).

3.4 Results

Functional Tcl proteins can be expressed in *E. coli*

While an expression system for *tcl* genes has been described previously in *B. subtilis*, for simple analysis of the microccocin biosynthetic pathway we wanted to develop a system to express functional Tcl proteins in the gram-negative *E. coli.* In this change of systems, we wanted to first confirm that our engineered Tcl proteins were capable of being expressed in *E. coli.* Testing for expression necessitated the ability to be able to purify each protein. Thus, TclE, TclI, TclJ, and TclN were engineered with affinity tags based on previously published results in which the functionality and expression of each protein was tested after the inclusion of an N- or C-terminal affinity tag in *B. subtilus*(46). Bands corresponding to the correct molecular weight of TclE, TclI, TclJ, TclN appear on a gel when they are individually expressed and affinity purified. All of these bands are absent in the lanes of the control strains, which verifies that all Tcl



	1	2	3	4
GST-TclE	+	+	+	+
TclI + TclN	-	+	-	-
TclI + TclJ	-	-	+	-
TclI + TclJ + TclN	-	-	-	+

TclE peptide species	Calculated mass $[M+H]^+$
unmodified	5373.9
6 x thiazole	5253.7
6 x thiazoline	5265.8
5 x thiazoline	5283.8

Fig. 18. Functional expression and purification of thiazole installation proteins from *E. coli*. A, Maps of plasmids used to express affinity-tagged Tcl proteins. Plasmid names are according to the inserted affinity-tagged Tcl protein. Locations of His_6 or GST tags are indicated in gray. Genes encoding Tcl proteins are color coded consistent with other figures in this report. *B*, SDS-PAGE analysis with coomassie blue staining to detect Tcl protein expression and purification. Whether proteins were purified with nickel-NTA (His_6 pull-down) or glutathione-linked (GST pull-down) resins is indicated. M, molecular weight markers; asterisks indicate tagged Tcl proteins, determined by expected molecular weight and comparison to the empty vector lane. *C*, MS analysis (MALDI-TOF) of thiazole installation on TclE using purified components from *E. coli*. Reactions included TEV protease to cleave the leader-core region of TclE from the GST tag prior to mass analysis. The modified N-terminus of TclE (after TEV cleavage) is (GGSEFQT...). See Fig. 5*C* for the complete TclE peptide sequence.

proteins are capable of being expressed under our conditions (Fig. 18*AB*).

Although we were able to detect our expressed proteins, it remained unclear whether or not our engineered proteins were functional. To determine whether or not the Tcl proteins were functional we evaluated the conversion of cysteines to heterocycles in the core of TclE in the presence of Tcl proteins purified from *E. coli*. A major peak corresponding to 6-thiazolines was observed when TclI and TclJ where incubated with TclE. A minor peak representing the near complete 5-thiazolines product was also evident suggesting that in our reaction conditions TclJ was capable of catalyzing complete and near complete conversion to thiazolines in the presence of TclI. TclE in the presence of TclI and TclN appeared the same as the unmodified version. This is because TclN requires pre-existing thiazolines, which are formed by TclJ, as substrate. When TclI, TclN, TclJ, and TclE were incubated together the data showed complete conversion to 6-thiazoles compared to the control (Fig. 18*C*). This demonstrated that all proteins were functional *in vitro*.

TclI acts as a scaffolding protein in a TclIJNE complex

Analysis of the biosynthetic gene cluster for microcin B17 indicated that a trimeric complex formed containing a cyclodehydratase, a dehydrogenase, and a docking protein (47, 60). As a result, it is believed that many TOMM biosynthetic enzymes form a complex, although in many cases co-purification data is lacking as typically each protein is individually expressed then combined together *in vitro* (82). To determine if TclIJNE form a complex, we co-expressed all four proteins, with TclI possessing a His_6 affinity tag (Fig. 19*A*). After purification of TclI, we were able to see three co-purifying bands that appear to be TclJ, TclN and TclE (Fig. 19*B*). When the His_6 tag is removed from TclI and the experiment is repeated, all four bands disappear,

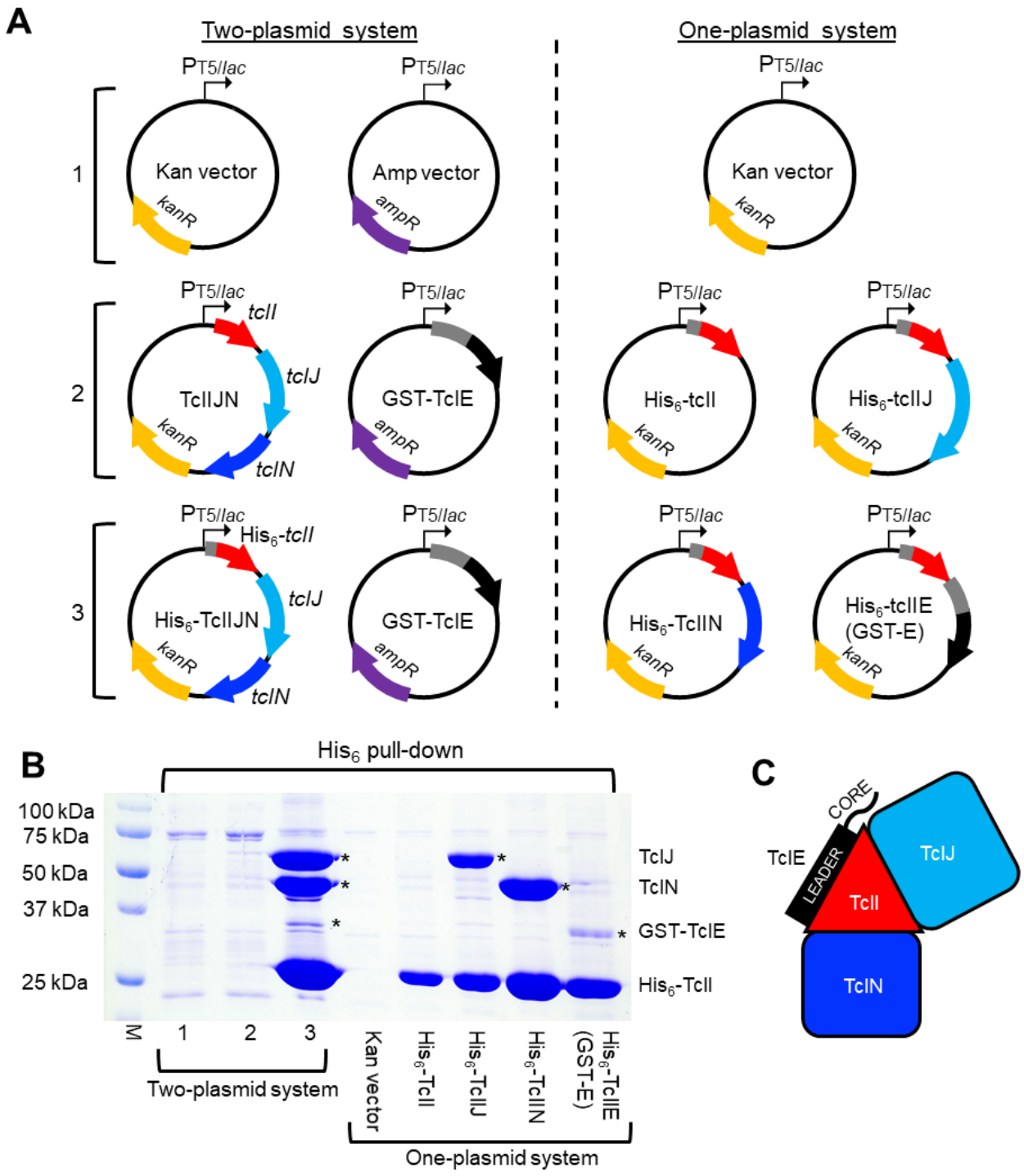


Fig. 19. TclI as a central docking protein in the TclEIJN complex. *A*, Maps of plasmid used in the two-plasmid and one-plasmid analyses indicated in (*B*). These maps are rendered similarly to those shown in Fig. 2*A*. *B*, SDS-PAGE analysis with coomassie blue staining to detect Tcl protein expression and co-purification. Asterisks indicate Tcl proteins that co-purified with His_6-tagged TclI. *C*, a model depicting the connectivity between Tcl proteins during thiazole installation.

indicating that the inducibly expressed Tcl proteins do not bind to nickel-NTA resin beads non-specifically in our system. Our data show that a four membered TclIJNE complex is readily co-purified.

The general architecture of the TclIJNE protein complex was unclear. The Tcl proteins could interact in many different combinations (Fig. 20). TclI exhibits no enzymatic activity and can be used to isolate the TclIJNE complex; therefore, we predict it plays an important role as the scaffolding protein for this complex. We were interested in characterizing the binding partners of TclI. It is possible that TclI could bind directly to one, two or all of the proteins in the complex. We expressed TclI, TclJ, and TclN each individually with His_6-TclI (Fig 3*A*). TclJ, TclN, and TclE each co-purify independently with TclI (Fig. 3*B*), leading us to conclude that TclI has three unique interaction surfaces and plays a key role in bringing together the TclJ and TclN enzymes with their substrate. (Fig. 19*C*).

Predicted structural models

With evidence for a general model in which TclI interacts with TclE, TclJ, and TclN across three separate surfaces we wished to consult structural models of the proteins when making further predictions of the architecture of this complex. Unfortunately, the structures for Tcl proteins are unknown; therefore, we generated predicted structures using an online protein modeler. Interestingly, both the TclI and TclJ models most closely resembled two separate domains of the single polypeptide TruD (PDB:4BS9) (83). TruD is a cyanobactin heterocylase in the trunkamide pathway (Fig. 21*A*). TclI appears most similarly to the N-terminal domain (NTD) of TruD while TclJ parallels the C-terminal domain (CTD) of TruD. In addition, the N-terminal domain of TclI appears to have an N-terminal winged-helix-turn-helix structure characteristic of a RiPP precursor peptide recognition element (RRE) (35, 54). The putative TclN model showed

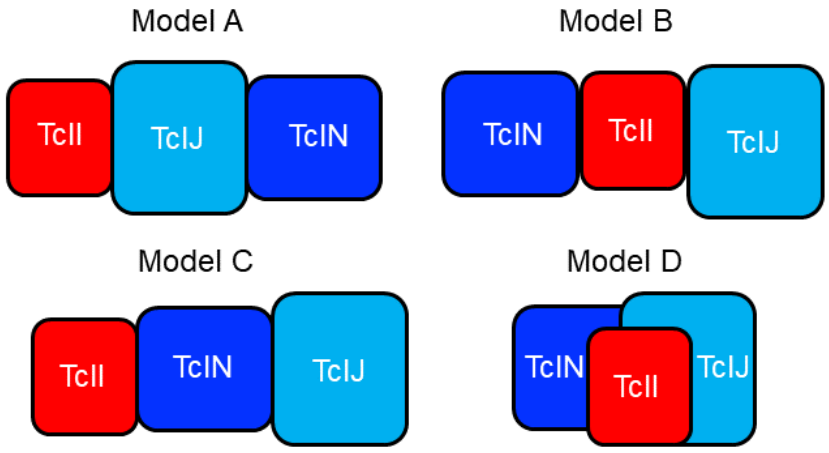


Fig. 20. Hypothetical TclIJN complex models**.** Simple depiction of some of the different combinations TclIJN could interact to form a complex.

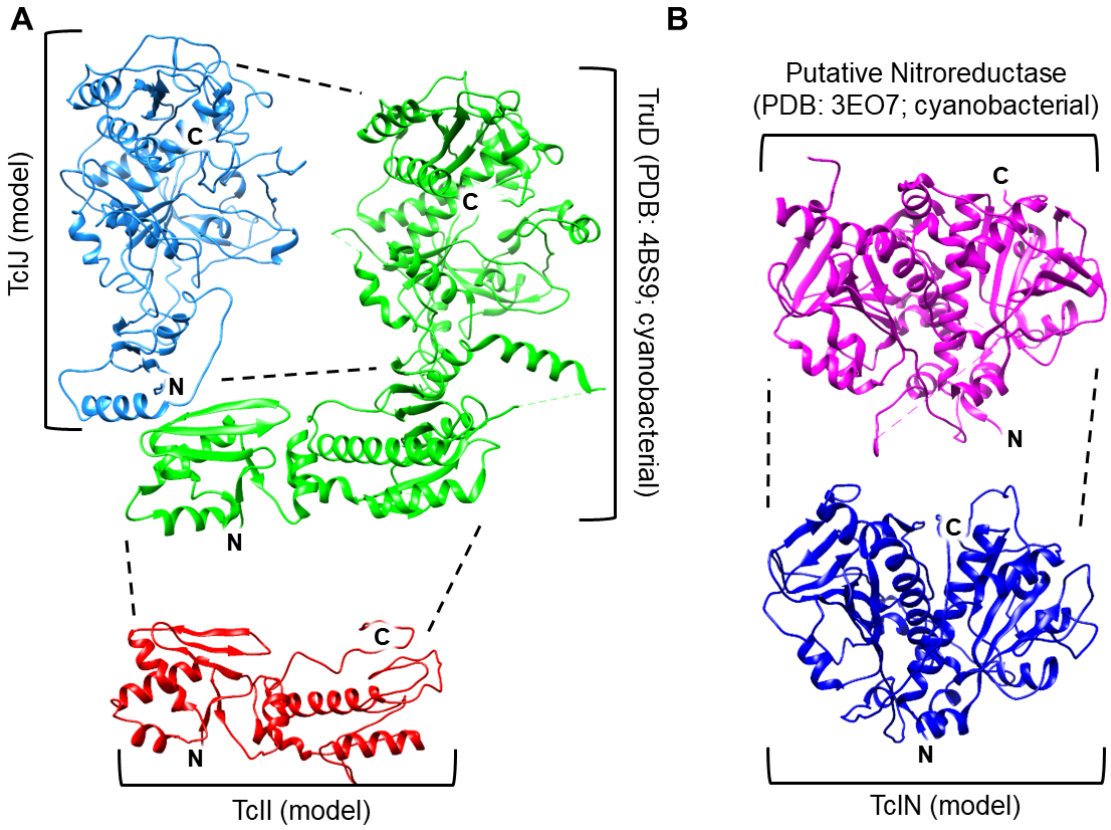


Fig. 21. Structural models for TclI, TclJ, and TclN. *A,* Predicted I-TASSER models for TclI and TclJ compared to TruD a cyclodehydratase fromthe trunkamide A pathway. Dashed lines represent regions of potential homology. *B,* Predicted I-TASSER model for TclN compared to a putative nitroreductase from cyanobacteria. Dashed lines represent regions of potential homology.

the highest similarity with a cyanobactin nitroreductase (3EO7) (Fig. 21*B*), which folds into a single globular domain.

Further characterization of TclI interaction surfaces with TclJ, TclN, and TclE

Based on analysis of the predicted structural models we hypothesized that due to the presence of the RRE in the NTD of TclI it should bind to the leader of TclE. In comparison to TruD, it appears perhaps that the CTD of TclI interacts with TclJ. We postulated that TclN might also interact with the CTD of TclI. Co-purification of TclJ, TclN, TclE and $TclE_{leader}$ with the affinity tagged NTD or CTD of TclI were performed to evaluate these predictions. TclE does co-purify with $TclI_{NTD}$, and $TclI_{NTD}$ appears to be stabilized in the presence of TclE compared to the control (Fig. 22*A*). However, it should be noted that when TclE and $TclI_{NTD}$ are co-expressed, cells grow much slower compared to when TclE is absent, requiring six times the amount of cell culture to reach the same pellet density as all other cultures. Consistent with these observations, TclE does not appear to be pulled down with the $TclI_{CTD}$ confirming that TclE appears to bind specifically to the NTD of TclI. Furthermore, $TclE_{leader}$ is sufficient for binding to $TclI_{NTD}$ (Fig. 22*B*)**.** In addition, our results support a model in which $TclI_{CTD}$ is capable of binding to both TclJ and TclN, since bands representing the correct molecular weight for TclJ and TclN appear when they are co-expressed with the CTD of TclI compared to control strains where TclN and TclJ are absent(Fig. 22*A*). Structural models for both TclJ and TclI closely resemble TruD. TclJ shows similarity to the CTD of TruD while TclI is similar to the NTD of TruD (Fig. 21). Therefore, we hypothesized that the CTD of TclJ is the region of the protein that binds to TclI. To test this, two versions of the $TclJ_{NTD}$ were co-expressed with His_6-TclI. From the data, it appears that the NTD of TclJ is adequate for binding to full length TclI (Fig. 22*C*). Combining this data with our previous data we suggest that the N-terminal domain of TclJ interacts through a surface on the CTD of TclI.

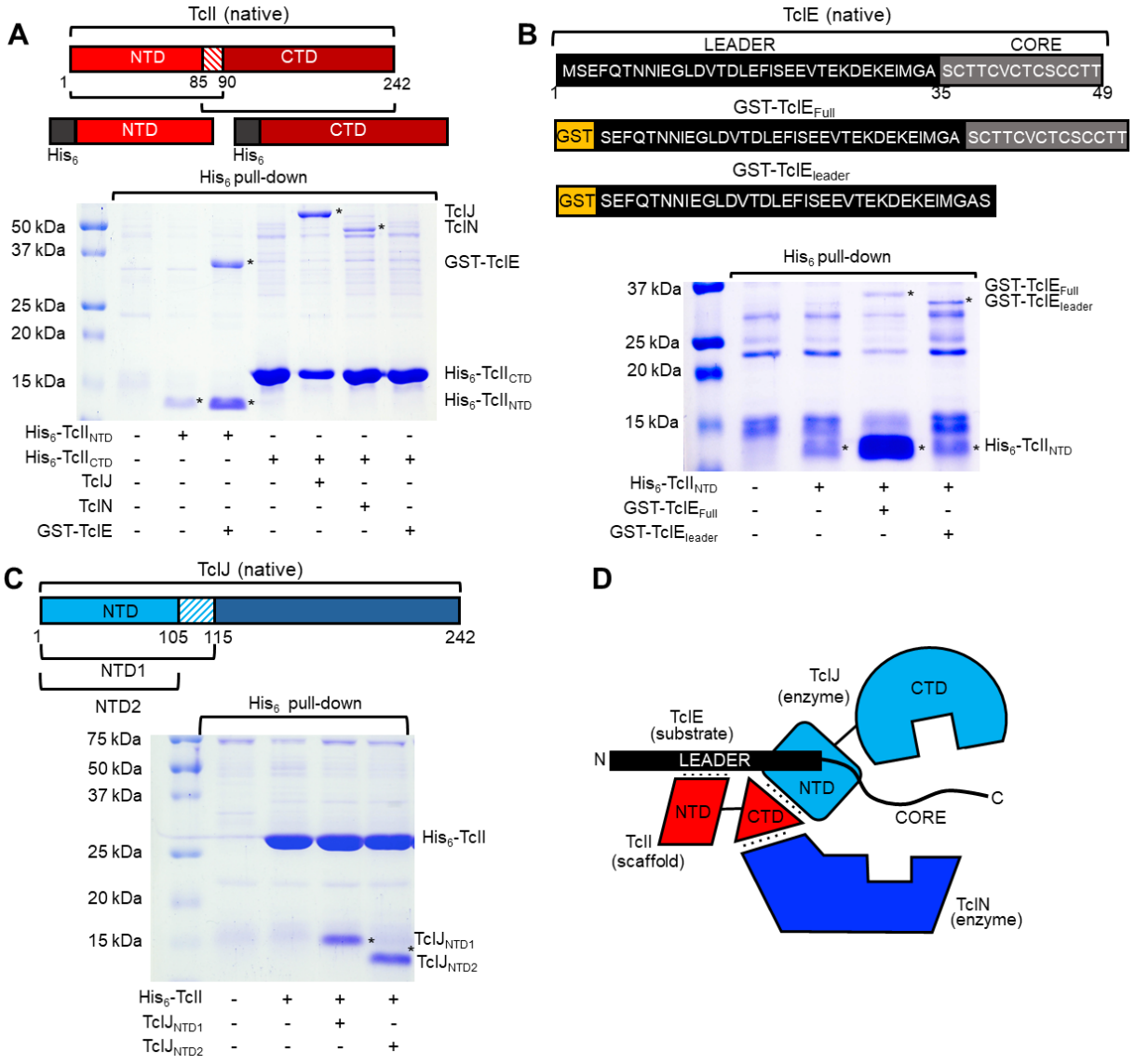


Fig. 22. Domain analysis of TclI binding to TclJ, N, and E. *A*, Simple diagram of His_6-$TclI_{NTD}$ and His_6-$TclI_{CTD}$ compared to native TclI. SDS-PAGE analysis with coomassie blue staining to detect Tcl protein expression and co-purification. Asterisks indicate His_6-$TclI_{NTD}$ and Tcl proteins that co-purified with His_6-tagged $TclI_{NTD}$/$TclI_{CTD}$ *B*, Diagram of the two NTD versions of TclJ used. SDS-PAGE analysis with coomassie blue staining to detect Tcl protein expression and co-purification. Asterisks indicate TclJ proteins that co-purified with His_6-tagged TclI. *C*, Diagram of GST-$TclE_{full}$ and GST-$TclE_{leader}$. SDS-PAGE analysis with coomassie blue staining to detect Tcl protein expression and co-purification. Asterisks indicate His_6-TclINTD and Tcl proteins that co-purified with His_6-tagged $TclI_{NTD}$. 55 mM Imidazole, Ni (NTA) beads washed twice *D*, a more detailed model for how TclI, TclJ, TclN and TclE interact with each other.

3.5 Discussion

We have confirmed that TclIJNE combine to form a four membered complex, and based on the findings in this paper, we have proposed a more detailed model for how these four proteins interact (Fig. 22*D*). In this model TclI, while not possessing any known enzymatic activity, is the central scaffolding protein for this thiazole installing complex. TclI NTD interacts with the leader region of the precursor peptide TclE. The NTD of TclI is where the RiPP precursor peptide recognition element is (RRE) is proposed to be, therefore it is not surprising that TclE interacts with $TclI_{NTD}$. $TclE_{leader}$ is sufficient for binding to the NTD of TclI. This corroborates work that has been done with LynD, NisB, and McbB (35, 54). However, our data also suggests that having the core peptide present stabilized the NTD of TclI. TclI binding to TclE most likely positions TclE in a way that directs the core towards modifying enzymes, TclJ and TclN. In our model, the NTD of TclJ binds to the CTD of TclI. It is interesting that TclJ and TclI both mapped to TruD. This lends support to previous research which has described that in approximately half known TOMM biosynthetic gene clusters the cyclodehydratase is fused with an auxiliary protein (30, 47, 52). Most research on TOMM biosynthetic gene clusters focus on the cyclodehydratases while dehydrogenases have been characterized less (48, 52, 53, 56). With the lack of information on dehydrogenases and the limited structural models we proposed that TclN bound to the CTD of TclI due to the fact that the NTD contains primarily the RRE. In our co-purification experiments we did see that the CTD of TclI was sufficient to bind TclN. Work should continue to be done to determine specific amino acid residues that are required on the interaction surfaces of all proteins in the TclIJNE complex. A greater understanding of TOMM protein complex architecture in micrococcin synthesis as well as other RiPP biosynthetic pathways can be useful for future endeavors to generate novel combinatorial peptides by

potentially allowing researchers to mix and match various cyclodehydratases, dehydrogenases and RRE-containing auxiliary proteins.

REFERENCES

1. Pavlova O, Mukhopadhyay J, Sineva E, Ebright RH, & Severinov K (2008) Systematic structure-activity analysis of microcin J25. *J Biol Chem* 283(37):25589-25595.
2. Furman JL, Chiu M, & Hunter MJ (2015) Early engineering approaches to improve peptide developability and manufacturability. *AAPS J* 17(1):111-120.
3. Fosgerau K & Hoffmann T (2015) Peptide therapeutics: current status and future directions. *Drug Discov Today* 20(1):122-128.
4. Lau JL & Dunn MK (2018) Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorg Med Chem* 26(10):2700-2707.
5. Kaspar AA & Reichert JM (2013) Future directions for peptide therapeutics development. *Drug Discov Today* 18(17-18):807-817.
6. Vlieghe P, Lisowski V, Martinez J, & Khrestchatisky M (2010) Synthetic therapeutic peptides: science and market. *Drug Discov Today* 15(1-2):40-56.
7. Bakail M & Ochsenbein F (2016) Targeting protein–protein interactions, a wide-open field for drug design. *Comptes Rendus Chimie* 19(1-2):19-27.
8. Al Musaimi O, Al Shaer D, de la Torre B, & Albericio F (2018) 2017 FDA Peptide Harvest. *Pharmaceuticals* 11(2):42.
9. Farkas H & Varga L (2011) Ecallantide is a novel treatment for attacks of hereditary angioedema due to C1 inhibitor deficiency. *Clin Cosmet Investig Dermatol* 4:61-68.
10. Verdine GL & Hilinski GJ (2012) Stapled peptides for intracellular drug targets. *Methods Enzymol* 503:3-33.
11. Feng Z & Xu B (2016) Inspiration from the mirror: D-amino acid containing peptides in biomedical approaches. *Biomol Concepts* 7(3):179-187.

12. Garton M, *et al.* (2018) Method to generate highly stable D-amino acid analogs of bioactive helical peptides using a mirror image of the entire PDB. *Proc Natl Acad Sci U S A* 115(7):1505-1510.

13. Tan YS, Lane DP, & Verma CS (2016) Stapled peptide design: principles and roles of computation. *Drug Discov Today* 21(10):1642-1653.

14. Bird GH, *et al.* (2010) Hydrocarbon double-stapling remedies the proteolytic instability of a lengthy peptide therapeutic. *Proc Natl Acad Sci U S A* 107(32):14093-14098.

15. Eichler J, *et al.* (1995) Peptide, peptidomimetic, and organic synthetic combinatorial libraries. *Med Res Rev* 15(6):481-496.

16. Houghten RA (1993) Peptide libraries: criteria and trends. *Trends Genet* 9(7):235-239.

17. Rosano GL & Ceccarelli EA (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* 5:172.

18. Sardar D, Lin Z, & Schmidt Eric W (2015) Modularity of RiPP Enzymes Enables Designed Synthesis of Decorated Peptides. *Chemistry & Biology* 22(7):907-916.

19. Sardar D & Schmidt EW (2016) Combinatorial biosynthesis of RiPPs: docking with marine life. *Curr Opin Chem Biol* 31:15-21.

20. Duan G & Walther D (2015) The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol* 11(2):e1004049.

21. Prabakaran S, Lippens G, Steen H, & Gunawardena J (2012) Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 4(6):565-583.

22. Ivry SL, *et al.* (2018) Global substrate specificity profiling of post-translational modifying enzymes. *Protein Sci* 27(3):584-594.

23. Kontaxi C, Piccardo P, & Gill AC (2017) Lysine-Directed Post-translational Modifications of Tau Protein in Alzheimer's Disease and Related Tauopathies. *Front Mol Biosci* 4:56.

24. Dai C & Gu W (2010) p53 post-translational modification: deregulated in tumorigenesis. *Trends in Molecular Medicine* 16(11):528-536.

25. ElizabethJ. Goldsmith RA, Xiaoshan Min, Tianjun Zhou, John M. Humphreys (2007) Substrate and docking interactions in Serine/Threonine Protein Kinases. *Chemical Reviews* 107:5065-5081.

26. Endicott JA, Noble ME, & Johnson LN (2012) The structural basis for control of eukaryotic protein kinases. *Annu Rev Biochem* 81:587-613.

27. Garai A, *et al.* (2012) Specificity of linear motifs that bind to a common mitogen-activated protein kinase docking groove. *Sci Signal* 5(245):ra74.

28. John A. McIntosh CRR, Vinayak Agarwal, Satish K. Nair, Grzegorz W. Bulaj, Eric W. Schmidt (2010) Circular logic: nonribosomal peptide-like macrocyclization with a ribosomal peptide catalyst. *Journal of American Chemical Society* 132(44):15499-15501.

29. Donia MS, Ravel J, & Schmidt EW (2008) A global assembly line for cyanobactins. *Nat Chem Biol* 4(6):341-343.

30. McIntosh JA, Donia MS, & Schmidt EW (2010) Insights into heterocyclization from two highly similar enzymes. *J Am Chem Soc* 132(12):4089-4091.

31. Li C, Zhang F, & Kelly WL (2011) Heterologous production of thiostrepton A and biosynthetic engineering of thiostrepton analogs. *Mol Biosyst* 7(1):82-90.

32. Pan SJ & Link AJ (2011) Sequence diversity in the lasso peptide framework: discovery of functional microcin J25 variants with multiple amino acid substitutions. *J Am Chem Soc* 133(13):5016-5023.

33. Arnison PG*, et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(1):108-160.

34. Zhang Y, Chen M, Bruner SD, & Ding Y (2018) Heterologous Production of Microbial Ribosomally Synthesized and Post-translationally Modified Peptides. *Front Microbiol* 9:1801.

35. Ortega MA & van der Donk WA (2016) New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem Biol* 23(1):31-44.

36. Zhao N, Pan YX, Cheng Z, & Liu HG (2016) Lasso peptide, a highly stable structure and designable multifunctional backbone. *Amino Acids* 48(6):1347-1356.

37. Yan KP*, et al.* (2012) Dissecting the Maturation Steps of the Lasso Peptide Microcin J25 in vitro. *Chembiochem* 13(7):1046-1052.

38. Hegemann JD*, et al.* (2014) Rational Improvement of the Affinity and Selectivity of Integrin Binding of Grafted Lasso Peptides. *Journal of Medicinal Chemistry* 57(13):5829-5834.

39. Knappe TA*, et al.* (2011) Introducing Lasso Peptides as Molecular Scaffolds for Drug Design: Engineering of an Integrin Antagonist. *Angewandte Chemie-International Edition* 50(37):8714-8717.

40. Oman TJ & van der Donk WA (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nat Chem Biol* 6(1):9-18.

41. Just-Baringo X, Albericio F, & Alvarez M (2014) Thiopeptide antibiotics: retrospective and recent advances. *Mar Drugs* 12(1):317-351.

42. Bennallack PR*, et al.* (2016) Reconstitution and Minimization of a Micrococcin Biosynthetic Pathway in Bacillus subtilis. *Journal of Bacteriology* 198(18):2431-2438.

43. Bennallack PR & Griffitts JS (2017) Elucidating and engineering thiopeptide biosynthesis. *World J Microbiol Biotechnol* 33(6):119.

44. Zheng Q, Fang H, & Liu W (2017) Post-translational modifications involved in the biosynthesis of thiopeptide antibiotics. *Org Biomol Chem* 15(16):3376-3390.

45. Philip R. Bennallack SRB, Michael J. Heder, Richard A. Robison, Joel S. Griffitts (2014) Characterization of a Novel Plasmid-Borne Thiopeptide Gene Cluster in *Staphylococcus epidermidis* Strain 115. *Journal of Bacteriology*:4344-4350.

46. Bewley KD*, et al.* (2016) Capture of micrococcin biosynthetic intermediates reveals C-terminal processing as an obligatory step for in vivo maturation. *Proc Natl Acad Sci U S A* 113(44):12450-12455.

47. Melby JO, Nard NJ, & Mitchell DA (2011) Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Curr Opin Chem Biol* 15(3):369-378.

48. Cox CL, Doroghazi JR, & Mitchell DA (2015) The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. *BMC Genomics* 16:778.

49. Ruffner DE, Schmidt EW, & Heemstra JR (2014) Assessing the Combinatorial Potential of the RiPP Cyanobactin tru Pathway. *ACS Synthetic Biology* 4(4):482-492.

50. Truman AW (2016) Cyclisation mechanisms in the biosynthesis of ribosomally synthesised and post-translationally modified peptides. *Beilstein J Org Chem* 12:1250-1268.

51. Dunbar KL*, et al.* (2014) Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. *Nat Chem Biol* 10(10):823-829.

52. Dunbar KL, Melby JO, & Mitchell DA (2012) YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. *Nat Chem Biol* 8(6):569-575.

53. Burkhart BJ, Schwalen CJ, Mann G, Naismith JH, & Mitchell DA (2017) YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem Rev* 117(8):5389-5456.

54. Burkhart BJ, Hudson GA, Dunbar KL, & Mitchell DA (2015) A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nature Chemical Biology* 11(8):564-570.

55. Deane CD*, et al.* (2016) In Vitro Biosynthesis and Substrate Tolerance of the Plantazolicin Family of Natural Products. *ACS Chem Biol* 11(8):2232-2243.

56. Melby JO, Li X, & Mitchell DA (2014) Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry* 53(2):413-422.

57. Czekster CM, Ge Y, & Naismith JH (2016) Mechanisms of cyanobactin biosynthesis. *Curr Opin Chem Biol* 35:80-88.

58. McIntosh JA & Schmidt EW (2010) Marine molecular machines: heterocyclization in cyanobactin biosynthesis. *Chembiochem* 11(10):1413-1421.

59. Neil L. Kelleher CLH, and Christopher T. Walsh (1999) Posttranslational Heterocyclization of Cysteine and Serine Residues in Antibiotic Microcin B17: Distributivity and Directionality. *Biochemistry*:15623-15630.

60. Ranabir Sinha Roy PJB, and Christopher T. Walsh (1998) Mutational Analysis of Posttranslational Heterocycle Biosynthesis in the Gyrase inhibitor Microcin B17: Distance Dependence from propetide and tolerance for substitution in a GSCG cyclizable sequence. *Biochemistry* 37(12):4125-4136.

61. Shkundina I, Serebryakova M, & Severinov K (2014) The C-terminal part of microcin B is crucial for DNA gyrase inhibition and antibiotic uptake by sensitive cells. *J Bacteriol* 196(9):1759-1767.

62. Molloy EM*, et al.* (2015) Identification of the minimal cytolytic unit for streptolysin S and an expansion of the toxin family. *BMC Microbiol* 15:141.

63. Molloy EM, Cotter PD, Hill C, Mitchell DA, & Ross RP (2011) Streptolysin S-like virulence factors: the continuing sagA. *Nat Rev Microbiol* 9(9):670-681.

64. Melby JO, Dunbar KL, Trinh NQ, & Mitchell DA (2012) Selectivity, directionality, and promiscuity in peptide processing from a Bacillus sp. Al Hakam cyclodehydratase. *J Am Chem Soc* 134(11):5309-5316.

65. Eric W. Schmidt JTN, David A. Rasko, Sebastian Sudek, Jonathan A. Eisen, Margo G. Haygood, and Jacques Ravel (2005) Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissooclinum patella*. *PNAS* 102(20):7315-7320.

66. Koehnke J*, et al.* (2014) The structural biology of patellamide biosynthesis. *Curr Opin Struct Biol* 29:112-121.

67. Hudson GA, Zhang Z, Tietz JI, Mitchell DA, & van der Donk WA (2015) In Vitro Biosynthesis of the Core Scaffold of the Thiopeptide Thiomuracin. *J Am Chem Soc* 137(51):16012-16015.

68. Zhang Z*, et al.* (2016) Biosynthetic Timing and Substrate Specificity for the Thiopeptide Thiomuracin. *J Am Chem Soc* 138(48):15511-15514.

69. Huo L, Rachid S, Stadler M, Wenzel SC, & Muller R (2012) Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem Biol* 19(10):1278-1287.

70. Goto Y, Ito Y, Kato Y, Tsunoda S, & Suga H (2014) One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. *Chem Biol* 21(6):766-774.

71. Ducasse R*, et al.* (2012) Sequence Determinants Governing the Topology and Biological Activity of a Lasso Peptide, Microcin J25. *Chembiochem* 13(3):371-380.

72. Assrir N*, et al.* (2016) Initial Molecular Recognition Steps of McjA Precursor during Microcin J25 Lasso Peptide Maturation. *Chembiochem* 17(19):1851-1858.

73. Pavlova O, Mukhopadhyay J, Sineva E, Ebright RH, & Severinov K (2008) Systematic structure-activity analysis of microcin J25. *Journal of Biological Chemistry* 283(37):25589-25595.

74. Pan SJ & Link AJ (2011) Sequence Diversity in the Lasso Peptide Framework: Discovery of Functional Microcin J25 Variants with Multiple Amino Acid Substitutions. *Journal of the American Chemical Society* 133(13):5016-5023.

75. Mukai A*, et al.* (2009) Nocardithiocin, a novel thiopeptide antibiotic, produced by pathogenic Nocardia pseudobrasiliensis IFM 0757. *J Antibiot (Tokyo)* 62(11):613-619.

76. Bowers AA, Acker MG, Koglin A, & Walsh CT (2010) Manipulation of thiocillin variants by prepeptide gene replacement: structure, conformation, and activity of heterocycle substitution mutants. *J Am Chem Soc* 132(21):7519-7527.

77. Weiying Liu YX, Min Ma, Shuzhen Wang, Nan Liu and Yijun Chen (2013) Multiple Oxidative Routes towards the Maturation of Nosiheptide. *Chem Bio Chem*.

78. Elshahawi SI, Shaaban KA, Kharel MK, & Thorson JS (2015) A comprehensive review of glycosylated bacterial natural products. *Chem Soc Rev* 44(21):7591-7697.

79. Dunbar KL, Tietz JI, Cox CL, Burkhart BJ, & Mitchell DA (2015) Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J Am Chem Soc* 137(24):7672-7677.

80. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.

81. Pettersen EF*, et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605-1612.

82. Shaun W. Lee DAM, Andrew L. Markley, Mary E. Hensler, David Gonzalez, Aaron Wohlrab, Pieter C. Dorrestein, Victor Nizet, Jack E. Dixon (2008) Discovery of a widely distributed toxin biosynthetic gene cluster. *PNAS*.

83. Koehnke J*, et al.* (2013) The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction. *Angew Chem Int Ed Engl* 52(52):13991-13996.

SUPPLEMENTARY INFORMATION

CHAPTER 2: Vector sequences

Parent Vectors

Plasmid features:

Multiple Cloning Site

ColE1 ori

LacZ ORF

KanR ORF

RhaR ORF

RhaS ORF

RhaBAD promoter

P15A ori

TetR ORG

Tet promoter

CmR ORF

pJG744

GAATTCccaCTCGAGccaGGTACCaccTCTAGAcccGTCGACaccAAGCTTccaCTGCAGccaGAGCTCg
gctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgCTGCtggttcgctcataagtaaaa
aacggcacctggtgccgtttttttgtctgaaacaagctgagcaataactagcataacccctttggggcctc
taaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccct
gtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccct
agcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctcta
aatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagg
gtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgtt
ctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgattta
taagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaatt
ttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgt
ttatttttctaaatacattcaaatatgtatccgctcatgaattaattcTTAGAAAAACTCATCGAGCATC
AAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATG
AAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCG
TCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGA
GTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGC
CATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAG
ACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACT
GCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGG

GGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCAT
AAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGT
TTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACAT
TATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGA
CGTTTCCCGTTGAATATGGCTCATaacacccttgtattactgtttatgtaagcagacagttttattgtt
catgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaagga
tcttcTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGG
TGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT
ACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACA
TACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGG
ACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAG
CTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCC
GAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC
CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTT
GTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAaacgccagcaacgcggccttttacggttcctggcc
ttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccg
cctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagc
ggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcaatggtgcactc
tcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtc
atggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccg
cttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacg
cgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatcc
gcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaaggg
cggtttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgatacc
gatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgt
gagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatgccagcgct
tcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccg

pJG747

GAATTCccaCTCGAGccaGGTACCacaggaaacagctATGACCATGATTACGGATTCACTGGCCGTCGTT
TTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCG
CCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGA
ATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAG
GCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGA
CCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGGTTGTTACTCGCTCACATT
TAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTAACTCGGCGTTT
CATCTGTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGA
GCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGCAGTTATCT
GGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTACA
CAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTACTGGAGGCTGAAG
TTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGCAGGT
CGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACA
CTACGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTG
AACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGAT
TGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCATCAT
CCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATGAAGCAGAACAACT
TTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGGCCT

GTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGATCCG
CGCTGGCTACCGGCGATGAGCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGA
TCATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATC
TGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGC
CCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGC
TTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTCTTGG
CGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGACTGG
GTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCG
ATACGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCT
GACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGC
GAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGG
CAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCA
GCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCGCATGGTCAGAA
GCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGT
CCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCA
ATTTAACCGCCAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAACAACTGCTGACGCCGCTG
CGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGCATTGACCCTA
ACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGC
AGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTT
ATCAGCCGGAAAACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGA
GCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTG
GCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTG
CCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCGAAT
TGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGCCGCTACAGTCAACAGCAACT
GATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTCCAT
ATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCT
ACCATTACCAGTTGGTCTGGTGTCAAAAATAAtaataaccgggcaggccatgtctgccgtcgacaccaag
cttccactgcagccagagctcggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgct
gctggttcgctcataagtaaaaaacggcacctggtgccgtttttttgtctgaaacaagctgagcaataac
tagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccgg
attggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtga
ccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgc
cggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctc
gaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcc
ctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctat
ctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatt
taacaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaa
tgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattc
TTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTT
TGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGT
ATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTT
ATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTC
CAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCA
TTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGA
ATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAAT
ACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAAT
GCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATT
GGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATT

GTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAAT
TTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATaacacccttgtattactgtttat
gtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcaga
ccccgtagaaaagatcaaaggatcttcTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACA
AAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAA
CTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAA
GAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGAT
AAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGG
GGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCT
ATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACA
GGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACC
TCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAaacgccagcaacgc
ggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgat
tctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgca
gcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtat
ttcacaccgcaatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactcc
gctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgg
gcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggt
tttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattc
acagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctg
ataaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttc
tgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaaca
tgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatc
actcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctg
cgatgcagatccg

pJG780

gaattcccactcgagcatcgtcggcatcggcatggcgaaTTAATCTTTCTGCGAATTGAGATGACGCCAC
TGGCTGGGCGTCATCCCGGTTTCCCGGGTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATT
CGGTCGAAATATCACTGATTAACAGGCGGCTATGCTGGAGAAGATATTGCGCATGACACACTCTGACCTG
TCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAACGCGCTCACTGCAC
GATGCCTCATCACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCT
TATCCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAATCACTGGTGTAACGATGGCGATTCAGCAACAT
CACCAACTGCCCGAACAGCAACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCATGCTCA
AGCTGACCGATAACCTGCCGCGCCTGCGCCATCCCCATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGC
TGGCGTTAAATCCCGGAATCGCCCCCTGCCAGTCAAGATTCAGCTTCAGACGCTCCGGGCAATAAATAAT
ATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCAGCATGAATGTAAAAGAGATCG
CCACGGGTAATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCT
CACAAAAATCATGTGTATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTG
GTCGCTGGCAAAAAAATCATCTTTGAGAAGTTTTAACTGATGCGCCACcgtggctacctcggccagagaa
cgaagttgattattcgcaatatggcgtacaaatacgttgagaagattcgcgTTATTGCAGAAAGCCATCC
CGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGTGGTTACTGT
CGCTGAATCCACAGCGATAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTT
CATCAGTCGCAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCGTTTGCTGCTTAAGCTGCCGATGTAGC
GTACGCAGTGAAAGAGAAAATTGATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCA
GCCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGTTTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAG
CAAGAGCAGTAATTGCATAAACAAGATCTCGCGACTGGCGGTCGAGGGTAAATCATTTTCCCCTTCCTGC

TGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGCGCCAGTGAGACGGAT
ACTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCG
ATACAGCACATTGGTCAGACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAA
CAGACCGTGCCACCGGTGATGGTATAGGGCTGCCCATTAAACACATGAATACCCGTGCCATGTTCGACAA
TCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCGCCAC
GGACGCGTTACCAGACGGAAAAAAATCCACACTATGTAATACGGTCATactggcctcctgatgtcgtcaa
cacggcgaaatagtaatcacgaggtcaggttcttaccttaaattttcgacggaaaaccacgtaaaaaacg
tcgatttttcaagatacagcgtgaattttcaggaaatgcggtgagcatcacatCACCACAATTCAGCAAA
TTGTGAACATCATCACGTTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGT
CGCGAATTCAGGCGCTTTTTAGACTGGTCGTAatgaaattcagcGGTACCaccGTCGACaccAAGCTTcc
aCTGCAGccaGAGCTCggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgctgg
ttcgctcataagtaaaaaacggcacctggtgccgtttttttgtctgaaacaagctgagcaataactagca
taacccctttggggcctctaaacgggtcttgaggggtttttgctgaaaggaggaactatatccggattgg
cgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgct
acacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggct
ttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccc
caaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttg
acgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgg
tctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaaca
aaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgc
gcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcTTAGA
AAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAA
AAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGG
TCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAA
GTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGAC
TTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGT
GATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCA
ACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTG
GAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTG
ATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAA
CGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGC
ACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAAT
CGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATaacacccttgtattactgtttatgtaag
cagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccg
tagaaaagatcaaaggatcttcTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGC
TTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACT
CTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC
GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGT
TCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAG
AAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGA
GCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGA
CTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAaacgccagcaacgcggcct
ttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgt
ggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgag
tcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcac
accgcaatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctat
cgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttg

tctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttca
ccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacaga
tgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaa
gcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaaggggatttctgttc
atgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgccc
ggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactca
gggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatg
cagatccg

pJG795

gaattcccactcgagacgtcTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCA
ATTCAAGGCCGAATAAGAAGGCTGGCTCTGCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAA
TGGCGGCATACTATCAGTAGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAAT
ACGCAACCTAAAGTAAAATGCCCCACAGCGCTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTT
GGCATAAAAAGGCTAATTGATTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTAC
TTTTGCTCCATCGCGATGACTTAGTAAAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGC
CAGCTTTCCCCTTCTAAAGGGCAAAAGTGAGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGA
GCAAAGCCCGCTTATTTTTTACATGCCAATACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTT
ACGGGTTGTTAAACCTTCGATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACTTTACTTTTA
TCTAATCTAGACATcattaattcctaatttttGTTGACACTCTATCGTTGATAGAGTTATTTTACCACTC
CCTATCAGTGATAGAGAAaaGGTACCgcgTCTAGAcccGTCGACgccAAGCTTccgCTGCAGcccGAGCT
Cgcttcctggtgtccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcg
gcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttgagttatcg
agattttcaggagctaaggaagctaaaATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCC
AATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCA
GCTGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCAC
ATTCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATAT
GGGATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGA
ATACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTG
GCCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCA
GTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATAC
GCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTC
GGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAttttttaaggca
gttattggtgcccttaaacgcctggtgtaaaaacccgcttcggcgggtttttttatgctgattaagcatt
ggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaag
gatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactga
gcgtcagacccccttaataagatgatcttcTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAA
ACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAA
CTGGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAG
ACTAACTCCTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGAC
TCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCT
TGGAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAA
TGACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTA
TCTTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGG
CGGAGCCTATGGAAAaacggctttgccgcggccctctcacttccctgttaagtatcttcctggcatcttc
caggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcgagtca
gtgagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttttctcc

tgccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagc
gctgaggtcagcttgcccacgtagacc

pJG745

GAATTCccaCTCGAGccaGGTACCaccTCTAGAcccGTCGACgccAAGCTTccgCTGCAGcccGAGCTCg
cttcctggtgtccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgttgatcggc
acgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtatttttgagttatcgag
attttcaggagctaaggaagctaaaATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAA
TGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGC
TGGATATTACGGCCTTTTTAAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACAT
TCTTGCCCGCCTGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGG
GATAGTGTTCACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAAT
ACCACGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGC
CTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCACCAGT
TTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAAATATTATACGC
AAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTCTGTGATGGCTTCCATGTCGG
CAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAAttttttaaggcagt
tattggtgcccttaaacgcctggtgtaaaaacccgcttcggcgggtttttttatgctgattaagcattgg
taactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaagga
tctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagc
gtcagacccctaataagatgatcttcTTGAGATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAAC
GAAAAAACCGCCTTGCAGGGCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACT
GGCTTGGAGGAGCGCAGTCACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGAC
TAACTCCTCTAAATCAATTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTC
AAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTG
GAGCGAACTGCCTACCCGGAACTGAGTGTCAGGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATG
ACACCGGTAAACCGAAAGGCAGGAACAGGAGAGCGCACGAGGGAGCCGCCAGGGGGAAACGCCTGGTATC
TTTATAGTCCTGTCGGGTTTCGCCACCACTGATTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCG
GAGCCTATGGAAAaacggctttgccgcggccctctcacttccctgttaagtatcttcctggcatcttcca
ggaaatctccgccccgttcgtaagccatttccgctcgccgcagtcgaacgaccgagcgtagcgagtcagt
gagcgaggaagcggaatatatcctgtatcacatattctgctgacgcaccggtgcagccttttttctcctg
ccacatgaagcacttcactgacaccctcatcagtgccaacatagtaagccagtatacactccgctagcgc
tgaggtcagcttgcccacgtagacc

Inserts

PlacT5 (LacIq)(XhoI/SalI)

CTCGAGccggtgcctaatgagtgagctaacttacattaattgcgttgcgcTCActgcccgctttccagtc
gggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattggg
cgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctga
gagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacg
gcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcag
cccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtggga
acgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgtt
ccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagac
agaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagt
cgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacg

ccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcag
cccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctacc
atcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcg
cgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccac
gcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtgg
ctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacg
ttactggtttCACattcaccaccctgaattgactctcttccgggcgctATCATGccataccgcgaaaggt
tTTGCACcattcgatggtgtccgggaTCC

PlacT5 (LacIWT) (XhoI/SalI)

CTCGAGccggtgcctaatgagtgagctaacttacattaattgcgttgcgcTCActgcccgctttccagtc
gggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattggg
cgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctga
gagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacg
gcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcag
cccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtggga
acgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgtt
ccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagac
agaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagt
cgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacg
ccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcag
cccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctacc
atcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcg
cgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccac
gcggttgggaatgtaattcagctccgccatcgccgcttccacttttcccgcgttttcgcagaaacgtgg
ctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacg
ttactggtttCACattcaccaccctgaattgactctcttccgggcgctATCATGccataccgcgaaaggt
tTTGCGCcattcgatggtgtccgggaTCC

Ptet (XhoI/SalI)

ctcgagACGTCTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTCAAGGC
CGAATAAGAAGGCTGGCTCTGCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAATGGCGGCAT
ACTATCAGTAGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCT
AAAGTAAAATGCCCCACAGCGCTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCATAAAA
AGGCTAATTGATTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGCTCC
ATCGCGATGACTTAGTAAAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCC
CCTTCTAAAGGGCAAAAGTGAGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCC
GCTTATTTTTTACATGCCAATACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGGGTTGT
TAAACCTTCGATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACTTTACTTTTATCTAATCTA
GACATCATTAATTCCTAATTTTTGTTGACACTCTATCGTTGATAGAGTTATTTTACCACTCCCTATCAGT
GATAGAGAAAAggtacc

PrhaBAD (XhoI/SalI)

ctcgagCATCGTCGGCATCGGCATGGCGAAttaATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGC
GTCATCCCGGTTTCCCGGGTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAA
TATCACTGATTAACAGGCGGCTATGCTGGAGAAGATATTGCGCATGACACACTCTGACCTGTCGCAGATA
TTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAACGCGCTCACTGCACGATGCCTCA

TCACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTATCCAGCA
ACGTTTCGCTGGATGTTGGCGGCAACGAATCACTGGTGTAACGATGGCGATTCAGCAACATCACCAACTG
CCCGAACAGCAACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCATGCTCAAGCTGACCG
ATAACCTGCCGCGCCTGCGCCATCCCCATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAA
ATCCCGGAATCGCCCCCTGCCAGTCAAGATTCAGCTTCAGACGCTCCGGGCAATAAATAATATTCTGCAA
AACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCAGCATGAATGTAAAAGAGATCGCCACGGGTA
ATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAAAAT
CATGTGTATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGC
AAAAAAATCATCTTTGAGAAGTTTTAACTGATGCGCcacCGTGGCTACCTCGGCCAGAGAACGAAGTTGA
TTATTCGCAATATGGCGTACAAATACGTTGAGAAGATTCGCGttaTTGCAGAAAGCCATCCCGTCCCTGG
CGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGTGGTTACTGTCGCTGAATC
CACAGCGATAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGTCG
CAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCGTTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGT
GAAAGAGAAAATTGATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAGCCAGGCCA
GAAGCAAGTTGAGACGTGATGCGCTGTTTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAG
TAATTGCATAAACAAGATCTCGCGACTGGCGGTCGAGGGTAAATCATTTTCCCCTTCCTGCTGTTCCATC
TGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGCGCCAGTGAGACGGATACTGCCCAT
CCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCAC
ATTGGTCAGACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTG
CCACCGGTGATGGTATAGGGCTGCCCATTAAACACATGAATACCCGTGCCATGTTCGACAATCACAATTT
CATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCGCCACGGACGCGTT
ACCAGACGGAAAAAAATCCACACTATGTAATACGGTcatACTGGCCTCCTGATGTCGTCAACACGGCGAA
ATAGTAATCACGAGGTCAGGTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTT
CAAGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCACATCACCACAATTCAGCAAATTGTGAACA
TCATCACGTTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTC
AGGCGCTTTTTAGACTGGTCGTaATGAAATTCAGCggtacc

ParaBAD (XhoI/KpnI)

ctcgagGCTACTCCGTCAAGCCGTCAATTGTCTGATTCGTTACCAAttaTGACAACTTGACGGCTACATC
ATTCACTTTTTCTTCACAACCGGCACGAAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACTCGC
GAGAAATAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTAGTGCTCA
AAAGCAGCTTCGCCTGACTAATGCGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCG
GAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTG
TCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACT
CGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCG
CCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCC
GGGCGAAAGAAACCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGGAC
GAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGG
CGGGAACAGCAAAATATCACCCGGTCGGCAGACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCG
CGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGT
TGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCGTTAAACGAGTATCCCGGCAGCAGGGGATCATT
TTGCGCTTCAGCcatACTTTTCATACTCCCACCATTCAGAGAAGAAACCAATTGTCCATATTGCATCAGA
CATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCCAACCGGTAACCCCGCTTATTAAAAGCA
TTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAA
AGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCG
GATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATaCCCGTTTTTggtacc

msfGFP (KpnI/SalI)

ggtaccGGAGGTCACATATGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCT
GGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAG
CTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGA
CCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT
GCCCGAAGGCTACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCA
ACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAA
GAACGGCATCAAGGCGAACTTCAAGATCCGCCACAACGTCGAGGACGGCAGCGTGCAGCTCGCCGACCAC
TACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
CCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG
GATCACTCTCGGCATGGACGAGCTGTACAAGTAAgtcgac

McjB & McjC (KpnI/SalI)

ggtaccTACAGGAGGGAGTGTGCAAAATGATCCGTTACTGCTTAACCAGTTATAGAGAGGATCTTGTTAT
CCTGGATATAATTAATGATAGTTTCAGCATAGTGCCTGACGCAGGTAGCTTGCTAAAAGAAAGAGATAAA
TTGCTTAAAGAATTCCCACAACTATCTTACTTTTTTGACAGTGAATATCATATTGGAAGTGTTTCTCGTA
ATAGTGACACTTCTTTTCTTGAAGAACGCTGGTTTCTACCAGAACCTGACAAAACATTATATAAGTGTTC
TCTATTTAAACGATTTATATTATTACTCAAAGTCTTTTACTATAGCTGGAATATTGAAAAAAAAGGGATG
GCATGGATTTTCATAAGTAATAAAAAAGAGAATAGGCTATACTCCTTGAATGAAGAGCATCTTATCCGGA
AAGAAATTAGTAATCTTTCCATTATCTTTCATCTTAATATTTTTAAATCTGACTGTCTTACCTATTCATA
CGCACTAAAAAGAATTCTTAATTCCAGAAATATTGATGCTCATCTTGTTATTGGTGTAAGGACACAACCT
TTTTATAGCCACTCTTGGGTGGAGGTTGGGGGACAAGTTATCAATGATGCTCCCAATATGCGGGATAAAT
TATCTGTTATTGCAGAGATATAGTTATGGAAATATTTAATGTCAAGTTAAATGATACTTCAATTAGAATT
ATTTTCTGTAAAACGCTTTCTGCCTTCCGGACAGAAAATACCATCGTTATGCTCAAAGGAAAAGCAGTTT
CAAATGGCAAACCTGTATCCACAGAGGAGATTGCCAGAGTAGTGGAAGAAAAAGGTGTTTCAGAAGTAAT
AGAAAATTTAGATGGTGTTTTCTGTATCCTAATTTATCATTTTAATGATCTCCTTATAGGGAAAAGCATT
CAATCAGGCCCCGCTCTATTTTATTGTAAAAAGAATATGGATATTTTTGTTTCGGATAAAATTTCTGATA
TCAAATTTTTGAATCCAGATATGACATTCAGTCTAAATATAACAATGGCAGAACATTATCTGTCAGGAAA
TCGAATAGCAACCCAGGAATCACTAATCACTGGCATTTACAAAGTAAATAATGGTGAGTTTATAAAATTT
AATAATCAGTTGAAACCTGTGCTACTTCGTGATGAGTTTAGTATTACCAAAAAGAACAATTCAACTATCG
ACAGTATCATTGATAATATTGAGATGATGCGGGATAATAGAAAAATAGCCCTATTATTTTCCGGAGGATT
GGATTCTGCATTAATTTTTCACACACTTAAAGAATCAGGTAACAAATTCTGCGCTTATCATTTTTTTTCT
GATGAATCTGATGACAGTGAAAAGTATTTTGCTAAGGAATACTGTTCAAAATATGGAGTTGATTTTATAT
CTGTTAATAAAAACATCAACTTTAATGAAAAACTTTATTTCAATTTAAATCCTAATAGTCCGGACGAAAT
CCCTTTGATATTTGAACAGACAGATGAAGAAGGTGAAGGTCAGCCCCCCATAGACGATGATTTATTATAT
CTATGTGGTCACGGTGGAGATCATATTTTCGGACAAAATCCTTCAGAACTTTTTGGCATTGATGCATATC
GAAGTCATGGCTTGATGTTTATGCATAAAAAAATAGTAGAATTTTCCAATCTCAAGGGAAAGAGATATAA
AGATATCATATTTTCAAATATTTCCGCATTCATTAATACATCCAACGGATGTTCTCCAGCAAAGCAAGAG
CACGTATCAGATATGAAACTTGCCTCTGCTCAGTTTTTTGCAACTGATTATACAGGAAAAATTAATAAAC
TAACTCCATTCCTGCATAAAAATATTATCCAGCATTATGCTGGCTTACCAGTTTTTAGTCTATTTAACCA
GCACTTTGATCGTTATCCCGTTCGTTATGAAGCGTTTCAACGATTTGGTTCAGATATTTTCTGGAAAAAA
ACCAAACGGTCATCTTCACAGCTAATATTCAGAATTCTATCCGGTAAAAAGGATGAACTAGTGAATACAA
TAAAACAGTCAGGATTAATTGAAATATTAGGCATTAACCATATTGAATTGGAAAGCATTTTGTATGAAAA
TACGACTACACGTCTGACAATGGAACTACCATATATACTTAACTTATACCGTCTGGCAAAATTCATTCAA
CTTCAATCCATTGATTATAAAGGTTAAgtcgac

McjA (KpnI/SalI)

GGTACCaaataaggaggctaaaaATGATTAAGCATTTTCATTTTAATAAACTGTCTTCTGGTAAAAAAAA
TAATGTTCCATCTCCTGCAAAGGGGGTTATACAAATAAAAAAATCAGCATCGCAACTCACAAAAGGTGGT
GCAGGACATGTGCCTGAGTATTTTGTGGGGATTGGTACACCTATATCTTTCTATGGCTGAgtcgac

McjA with BamHI modification (KpnI/SalI)

GGTACCaaataaggaggctaaaaATGATTAAGCATTTTCATTTTAATAAACTGTCTTCTGGTAAAAAAAA
TAATGTTCCATCTCCTGCAAAGGGGGTTATACAAATAAAAAAATCAGCATCGCAACTCACAAAAGGTGGA
TCCGGACATGTGCCTGAGTATTTTGTGGGGATTGGTACACCTATATCTTTCTATGGCTGAgtcgac

Supplementary Table 2. Plasmid Table

Plasmid Name	Parent Plasmid	Insert
pJG756	pJG747	PlacT5(lacIq)
pJG753	pJG747	PlacT5(LacIWT)
pJG808	pJG747	Ptet
pJG751	pJG747	PrhaBAD
pJG749	pJG747	ParaBAD
pJG804	pJG744	Ptet, msfGFP
pJG783	pJG744	PrhaBAD, msfGFP
pJG805	pJG744	PlacT5(lacIQ), msfGFP
pJG809	pJG744	ParaBAD, msfGFP
pJG810	pJG744	PlacT5(lacIWT), msfGFP
pJG798	pJG780	McjA
pJG803	pJG780	McjA with BamHI
pJG797	pJG795	McjB & McjC

Supplementary Table 3. Strain Table

Strain name	Background	Plasmid (s)
C893	DH5α	pJG797, pJG798
C894	DH5α	pJG795, pJG798
C895	DH5α	pJG797, pJG745
C896	MDS™42 Meta ΔrecA	pJG797, pJG798
C897	MDS™42 Meta ΔrecA	pJG795, pJG780
C901	MDS™42 Meta ΔrecA	pJG797, pJG803
C902	MDS™42 Meta ΔrecA	pJG795, pJG780

CHAPTER 3: Vector sequences

Parent vectors

Plasmid Features :

LacI ORF

Lac/T5 promoter

Multiple cloning site

T7 terminator

AmpR ORF

KanR ORF

CmR ORF

pCDF ori

ColE1 ori

pJG743

tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgacc
gctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccg
gctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcga
ccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccct
ttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatct
cggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattta
acaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatg
tgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcTT
AGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTG
AAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTAT
CGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTAT
CAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCA
GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATT
CGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAAT
GCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATAC
CTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGC
TTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGG
CAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT
CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTT
AATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATaacacccttgtattactgtttatgt
aagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagacc
ccgtagaaaagatcaaaggatcttcTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA
AAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT

GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGA
ACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAA
GTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG
GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTAT
GAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGG
AGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTC
TGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAaacgccagcaacgcgg
cctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattc
tgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagc
gagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtattt
cacaccgcaatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgc
tatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggc
ttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttt
tcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcac
agatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgat
aaagcgggccatgttaagggcggttttttcctgtttggtcactgatgcctccgtgtaagggggatttctg
ttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatg
cccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcac
tcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcg
atgcagatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaac
cgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcg
tatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagc
acgatcatgcgcacccgtggggccgccatgccggcgataatggcctgcttctcgccgaaacgtttggtgg
cgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaagcgacaggccgatcat
cgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagt
tgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctga
ctgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacattaat
tgcgttgcgcTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA
CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAAC
AGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCA
GGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCC
CACTACCGAGATGTCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATC
TGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAAC
CGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATG
CCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGA
CCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGG
GTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATC
CTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCT
TTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAG
ATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAA
CGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCC
ACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGA
CACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACattcaccaccctgaattgactctcttc
cgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcc
cttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaagg
aatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccga
aacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgcc
agcaaccgcacctgtggcgccggtgagatctcgatcccgcgaaataaTTGTGAGCGGATAACAAttacga

gcttcatgcacagtgaaaTCATGAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAtgtggAA
TTGTGAGCGCTCACAATTccacaacggttctaGGTACCaccTCTAGAcccGTCGACaccAAGCTTccaCT
GCAGccaGAGCTCggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctgctggttc
gctcataagtaaaaaacggcacctggtgccgttttttgtctgaaacaagctgagcaataaCTAGCATAA
CCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGctgaaaggaggaactatatccggat

pED022

ctcgaggatcgatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgcTCACTGCCCGCT
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGC
GTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCT
GGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGT
GGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCA
ACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCG
CAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCC
TTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGC
GCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCA
CGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAG
AAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTA
ATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTC
GTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTG
CGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGT
TGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAG
AAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATC
GTATAACGTTACTGGTTTCACattcaccaccctgaattgactctcttccgggcgctatcatgccataccg
cgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagg
aagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatgg
cgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaa
gtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccg
gtgagatctcgatcccgcgaaatAATTGTGAGCGGATAACAATTacgagcttcatgcacagtgaaaTCAT
GAAAAATTTATTTGCTTTGTGAGCGGATAACAATTATAATAtgtggAATTGTGAGCGCTCACAATTccac
aacggtTCTAGAaataattttgtttaactttaagaaggagatataCATATGGGATCCTGAGAGCTCggaa
atgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaata
accctgataaatgcttcaataatattgaaaaaggaagagtaTGAGTATTCAACATTTCCGTGTCGCCCTT
ATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATG
CTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAG
TTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCC
CGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACT
CACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCAT
GAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTG
CACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACG
ACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACT
TACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGC
TCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCA
TTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAAC
TATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAActgtcagac
caagtttactcgtaaaaacccgcttcggcgggtttttttatgctagggcggttcagtagaaaaGATCAAA
GGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTTTTGCCCTGTAAACGAAAAAACCACCTGGGGAG

GTGGTTTGATCGAAGGTTAAGTCAGTTGGGGAACTGCTTAACCGTGGTAACTGGCTTTCGCAGAGCACAG
CAACCAAATCTGTCCTTCCAGTGTAGCCGGACTTTGGCGCACACTTCAAGAGCAACCGCGTGTTTAGCTA
AACAAATCCTCTGCGAACTCCCAGTTACCAATGGCTGCTGCCAGTGGCGTTTTACCGTGCTTTTCCGGGT
TGGACTCAAGTGAACAGTTACCGGATAAGGCGCAGCAGTCGGGCTGAACGGGGAGTTCTTGCTTACAGCC
CAGCTTGGAGCGAACGACCTACACCGAGCCGAGATACCAGTGTGTGAGCTATGAGAAAGCGCCACACTTC
CCGTAAGGGAGAAAGGCGGAACAGGTATCCGGTAAACGGCAGGGTCGGAACAGGAGAGCGCAAGAGGGAG
CGACCCGCCGGAAACGGTGGGGATCTTTAAGTCCTGTCGGGTTTCGCCCGTACTGTCAGATTCATGGTTG
AGCCTCACGGCTCCCACAGATGCACCGGAAAAGCGTCTGTTTATGTGAACTCTGGCAGGAGGGCGGAGCC
TATGGAAAAACGCCACCGGCGCGGCCCTGCTGTTTTGCCTCACATGTTAGTCCCCTGCTTATCCACGGAA
TCTGTGGGTAACTTTGTATGTGTCCGCAGCGCccgccgcagtctcacgcccggagcgtagcgaccgagtg
aa

Inserts

TclI (KpnI/XbaI) insert

ggtaccGGAGGTCACATATGTACCTGCTGTTCAAGAAAGACACCTTTTATATCAAGACCCACAACGAGGG
TATTCTGTTCAAAAACAACTTTACCAACCTGGAAGTGAAGACCCATAAAAGCTACTATGTTTTCGAGAAC
CTGATCGAATACCTGAACGGCAGCTATACCGAGAACCAGATCATCGAAAACATCAAGAACAAGAAAGTGG
CGCTGTTCTGCAAAAACATCATTAACGTTCTGAAGGAGAAAAACTTTATCTACGCGAGCGAAAACCGTCT
GGATAACCTGAGCGAGCTGGAAATCAAGATTCTGTATCTGAACAGCAAGAACATCCAACTGAGCAAAGAC
GCGTTCCTGAACAACAACAAGATCGATATTAAAACCTACAACCTGATGGAGAACGAGTTCGCGAACACCA
TGCTGAAGGAGTACTATTTTAGCCCGAGCGAAACCGCGAACATCAAGATTAGCCTGGGTAAATTTGAGAG
CGACATCGGCTACTATATCATTCCGGATGGTGAATACCTGATTGTGGGCAAAAGCAAGAAACAGCACACC
CACACCGTTAGCCTGAGCACCTATGAGATCCCGCTGCACGCGTGGCTGATCTGCCTGAACATTCTGCTGA
GCGAACTGTTCCTGTACTGCACCAGCATTTATAGCGAGGACAAGTTCGATGAAAACTACTACAAGTTCAA
CGTGCAAACCTTTGAGGGTGACTTCTTTAAGAAAGAAAACAAATAAtctaga

TclJ (XbaI/SalI) insert

tctagaGGAGGTCACATATGGAAATCTTCGAAAGCCCGGAGTTTAACATCATTCGTTACCTGAGCAACAG
CTATATTTTCAAGAGCACCAGCAAAGAACCGGACTTTCTGAACAGCATCAAAGAGCGTATTCTGCTGAAT
ACCAACGATATCATTTACAAGAAAAACATCCCGGAACGTAACATTAGCTTCCTGAGCCGTATCGAGAACA
GCATTGGTAACTTTATCTGCAACCAGATTAGCGAATACGACTACATCATCTACATCTACAACCACAAGAC
CAACAAGATCAAGAAAGTGAACTACTTCATCAGCCCGCACTACTATAAGAAACTGCCGAACGACACCACC
AACTGCTTCAACATCATTATCAACGAGCTGAAGAACAGCAAAAGCCACAACTTTCGTAGCCGTAACATTG
ATAGCATCTACAACCAGCTGGACAAATATTTCCTGGATAAGAACCTGGGTATCTGCAACATGCTGCTGGA
CAACTACGATGGCCCGTTTCCGATTAGCGTTGCGATGCTGCCGCTGGACAACGGTAAAGAGGAACCGGGT
GTGGGCCGTACCCGTAAGATCCAACGTAGCCGTGCGGTTGCGCTGCTGGAAGCGTACGAGCGTTATAGCG
GTCTGGAACCGCGTGGCAAGAAAACCATCAACCACACCGAGGATCTGAAGAGCAAGAAAGTGAACATGAA
CAGCCTGATTCTGCACAACAACCCGTTCCTGATCAGCCACGGCATTAAAAACAGCAACTTTACCTACAAC
GACCTGCTGGATAACGAAATCAGCTGGGTTAAGTGCCTGAACCTGAACACCTTCCAGACCCTGCTGATTC
CGGAGCAATATGCGTACTATGGTATTAACATCAGCAACCACAAGGAAAAAGCGATTAACATCGCGTACGA
GATCAGCAACGGTTGCAGCGTGGGCAACAACTATCTGGAAGTGGTTTACTATGGCCTGATGGAAGTTATC
GAGCGTGACAGCTTCCTGTGCAGCTGGTACTTTAACACCCCGAAGGATAAAATTAGCCTGAAAAACGCGA
GCACCAGCATCAAGAACCTGATTCAGCAATTCACCAGCTACTATAACGACTATAAACTGGAACTGTTCTA
CCTGTATAACGAGTTTAACATCCCGGTGGTTCTGGCGACCGTGACCCTGAAGGAAAGCAGCACCAAGAAA
ATGAACTTTATGTGCGCGGCGGCGGCGGACATTAACATCGAAGATGCGATCGAGAAAAGCATTCACGAGA
TCGGTGGCATTCTGTTCGGTCTGAACAAGAAATTTATCGATCGTTACCACGAGCTGGAAGCGATTCGTAA

AAACAACCTGGACGTGAAGACCATGGAAGATCACACCCTGGTTTATGGCCTGCCGGAGCACCGTACCTAC
ATCCAGCAAAAGATGAACTACGAAAACATCTACGACTATGATAAGGAGCTGACCCCGAAAATCTTCTACA
AAGAGGTGCAGAAGCTGATCAAGAAAATTAGCACCACCAAGGACATTCTGCTGGTTGATCAAACCCCGCT
GATCAGCAAGAAAATTGACCTGAAAGTGGGTAAAATTATCGTTCCGGGCCTGCTGCCGATGACCTTTGGC
AAATACAACATCCGTGTTAGCGAAAACCGTTACCACGAGCTGTGCCACTTCTATAGCAAGGACCTGATTA
TCGATCTGAACCCGCACCCGTTTCCGTAAgtcgac

TclN (SalI/SacI) insert

gtcgacGGAGGTCACATATGGATATCAGCAAATTCCTGTACAACCTGCACTATAACCCGGGCGAGGTGGT
TAGCGCGAGCTACACCATCGAGGACACCATTCAGCGTAACAGCGAAGGTTTTTACAAGGGCTATGGTATC
GATTTCCTGAAACTGCAACAAAAGAGCCCGATCGTGAAAGTTATTCTGAAGAGCTACGGCGACATCTTCT
TTAACCGTGTGGAAAACAAGAAAAAGCCGCTGATCTTCTGCCGTAAAATGACCCCGAGCGGTGGCGGTCT
GTATCCGATCAACATTTTTATCTGCACCAACTTCAAGAACCGTATCGCGCTGTTCCAATTTGATTTCAAA
CGTAACCTGCTGATCTTCATCAAGTACATCAACATCGAGATCAACAACGAATGCACCAAACTGTACCTGG
TTCCGTGCTATACCCGTAACTACTTCAAGTACAAGGAGTTCAGCTATCGTCTGTGCCCGCTGGACACCGG
CTACCTGATCAGCACCCTGCTGTATAACTTTAGCGTGGAGAACATCACCTTCAAGCTGAGCATTAAACTG
AACAAGAACAGCGACATCACCGATGTTCTGAACGAAATCGGTTGCGAGGAAATTCCGTACAGCATCATTG
AGCTGAACGAAAGCCTGAACCTGGACAACCTGAGCCTGGAGCACTACGATACCGAAAGCTATTTCTTTAA
CCCGAACAAAGTGCGTAACCTGCTGGAGATCGACACCCTGATTCACCAGGAATACCACAAGGATATCAAC
ATCAACTTCAATAACGAGAACAAGCTGTTCGAGAAGTTCGAAATCCAAAAACGTATTAGCCCGGGTGGTG
AGTTCATCCAGAACAGCAAAGTGGAGCAAGAAAGCATTAACAAGTTTATCAGCCTGATTATGCAGTACAA
AAACAAGAGCAACTTCCTGAGCGAATATATCCTGCTGAACCTGATTGACGTTCAAAACAAGCGTATCATT
AACCTGAGCGCGAGCGAGTTTCTGAGCTACAAAAACAACGTGAGCATCGAATTTATTGATAAGCAGCTGA
CCCGTCGTAACTTCAACCTGCTGGCGGTTCCGTACATCCTGTATGTGGGCGTTAACGAGGAAAAAATTAA
GGAGAACTACAGCAACAACTACTTCAAAATCAGCCGTATCATCGCGGGCTTCTGGAGCGGTGTGGTTAGC
ATTCTGAGCGCGCAGTGCGGTCTGAGCACCCACCCGATGATGAGCTACAACGCGCGTGAACTGGAGGAAT
ACATCTTCAAGAACCGTTATAGCATTCTGAACCAAATCGTTATTGGCGGTAACATCACCACCAACCGTAT
GGACTCCATGCTGATTCGTAGCGATCTGTAAAAGCTGCCCgagctc

His_6-TclI (KpnI/XbaI) insert

ggtaccGGAGGTCACATATGCATCATCACCATCACCATGGCGGAAGCTACCTGCTGTTCAAGAAAGACAC
CTTTTATATCAAGACCCACAACGAGGGTATTCTGTTCAAAAACAACTTTACCAACCTGGAAGTGAAGACC
CATAAAAGCTACTATGTTTTCGAGAACCTGATCGAATACCTGAACGGCAGCTATACCGAGAACCAGATCA
TCGAAAACATCAAGAACAAGAAAGTGGCGCTGTTCTGCAAAAACATCATTAACGTTCTGAAGGAGAAAAA
CTTTATCTACGCGAGCGAAAACCGTCTGGATAACCTGAGCGAGCTGGAAATCAAGATTCTGTATCTGAAC
AGCAAGAACATCCAACTGAGCAAAGACGCGTTCCTGAACAACAACAAGATCGATATTAAAACCTACAACC
TGATGGAGAACGAGTTCGCGAACACCATGCTGAAGGAGTACTATTTTAGCCCGAGCGAAACCGCGAACAT
CAAGATTAGCCTGGGTAAATTTGAGAGCGACATCGGCTACTATATCATTCCGGATGGTGAATACCTGATT
GTGGGCAAAAGCAAGAAACAGCACACCCACACCGTTAGCCTGAGCACCTATGAGATCCCGCTGCACGCGT
GGCTGATCTGCCTGAACATTCTGCTGAGCGAACTGTTCCTGTACTGCACCAGCATTTATAGCGAGGACAA
GTTCGATGAAAACTACTACAAGTTCAACGTGCAAACCTTTGAGGGTGACTTCTTTAAGAAAGAAAACAAA
TAAtctaga

His_6-TclJ (XbaI/SalI) insert

tctagaGGAGGTCACATATGCATCATCACCATCACCATGGCGGAAGCGAAATCTTCGAAAGCCCGGAGTT
TAACATCATTCGTTACCTGAGCAACAGCTATATTTTCAAGAGCACCAGCAAAGAACCGGACTTTCTGAAC
AGCATCAAAGAGCGTATTCTGCTGAATACCAACGATATCATTTACAAGAAAAACATCCCGGAACGTAACA

TTAGCTTCCTGAGCCGTATCGAGAACAGCATTGGTAACTTTATCTGCAACCAGATTAGCGAATACGACTA
CATCATCTACATCTACAACCACAAGACCAACAAGATCAAGAAAGTGAACTACTTCATCAGCCCGCACTAC
TATAAGAAACTGCCGAACGACACCACCAACTGCTTCAACATCATTATCAACGAGCTGAAGAACAGCAAAA
GCCACAACTTTCGTAGCCGTAACATTGATAGCATCTACAACCAGCTGGACAAATATTTCCTGGATAAGAA
CCTGGGTATCTGCAACATGCTGCTGGACAACTACGATGGCCCGTTTCCGATTAGCGTTGCGATGCTGCCG
CTGGACAACGGTAAAGAGGAACCGGGTGTGGGCCGTACCCGTAAGATCCAACGTAGCCGTGCGGTTGCGC
TGCTGGAAGCGTACGAGCGTTATAGCGGTCTGGAACCGCGTGGCAAGAAAACCATCAACCACACCGAGGA
TCTGAAGAGCAAGAAAGTGAACATGAACAGCCTGATTCTGCACAACAACCCGTTCCTGATCAGCCACGGC
ATTAAAAACAGCAACTTTACCTACAACGACCTGCTGGATAACGAAATCAGCTGGGTTAAGTGCCTGAACC
TGAACACCTTCCAGACCCTGCTGATTCCGGAGCAATATGCGTACTATGGTATTAACATCAGCAACCACAA
GGAAAAAGCGATTAACATCGCGTACGAGATCAGCAACGGTTGCAGCGTGGGCAACAACTATCTGGAAGTG
GTTTACTATGGCCTGATGGAAGTTATCGAGCGTGACAGCTTCCTGTGCAGCTGGTACTTTAACACCCCGA
AGGATAAAATTAGCCTGAAAAACGCGAGCACCAGCATCAAGAACCTGATTCAGCAATTCACCAGCTACTA
TAACGACTATAAACTGGAACTGTTCTACCTGTATAACGAGTTTAACATCCCGGTGGTTCTGGCGACCGTG
ACCCTGAAGGAAAGCAGCACCAAGAAAATGAACTTTATGTGCGCGGCGGCGGCGGACATTAACATCGAAG
ATGCGATCGAGAAAAGCATTCACGAGATCGGTGGCATTCTGTTCGGTCTGAACAAGAAATTTATCGATCG
TTACCACGAGCTGGAAGCGATTCGTAAAAACAACCTGGACGTGAAGACCATGGAAGATCACACCCTGGTT
TATGGCCTGCCGGAGCACCGTACCTACATCCAGCAAAAGATGAACTACGAAAACATCTACGACTATGATA
AGGAGCTGACCCCGAAAATCTTCTACAAAGAGGTGCAGAAGCTGATCAAGAAAATTAGCACCACCAAGGA
CATTCTGCTGGTTGATCAAACCCCGCTGATCAGCAAGAAAATTGACCTGAAAGTGGGTAAAATTATCGTT
CCGGGCCTGCTGCCGATGACCTTTGGCAAATACAACATCCGTGTTAGCGAAAACCGTTACCACGAGCTGT
GCCACTTCTATAGCAAGGACCTGATTATCGATCTGAACCCGCACCCGTTTCCGTAAgtcgac

GST-TclE (XbaI/SacI) insert

tctagaAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCCCCTATACTAGGTTATTGGAAAA
TTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTA
TGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCT
TATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACA
ACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAG
ATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTA
CCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCC
ATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTT
CCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGC
AAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAAGGAG
GAGAAAACCTGTATTTTCAAGGCGGATCCGAGTTCCAGACCAACAACATCGAAGGTCTGGACGTGACCGA
TCTGGAGTTTATTAGCGAGGAAGTTACCGAAAAGGACGAGAAAGAAATCATGGGTGCTTCTTGCACCACC
TGCGTGTGCACCTGTAGCTGCTGCACCACCTAACTGCAGCCAgagctc

TclN-His$_6$ (SalI/SacI) insert

gtcgacGGAGGTCACATATGGATATCAGCAAATTCCTGTACAACCTGCACTATAACCCGGGCGAGGTGGT
TAGCGCGAGCTACACCATCGAGGACACCATTCAGCGTAACAGCGAAGGTTTTTACAAGGGCTATGGTATC
GATTTCCTGAAACTGCAACAAAAGAGCCCGATCGTGAAAGTTATTCTGAAGAGCTACGGCGACATCTTCT
TTAACCGTGTGGAAAACAAGAAAAAGCCGCTGATCTTCTGCCGTAAAATGACCCCGAGCGGTGGCGGTCT
GTATCCGATCAACATTTTTATCTGCACCAACTTCAAGAACCGTATCGCGCTGTTCCAATTTGATTTCAAA
CGTAACCTGCTGATCTTCATCAAGTACATCAACATCGAGATCAACAACGAATGCACCAAACTGTACCTGG
TTCCGTGCTATACCCGTAACTACTTCAAGTACAAGGAGTTCAGCTATCGTCTGTGCCCGCTGGACACCGG
CTACCTGATCAGCACCCTGCTGTATAACTTTAGCGTGGAGAACATCACCTTCAAGCTGAGCATTAAACTG

AACAAGAACAGCGACATCACCGATGTTCTGAACGAAATCGGTTGCGAGGAAATTCCGTACAGCATCATTG
AGCTGAACGAAAGCCTGAACCTGGACAACCTGAGCCTGGAGCACTACGATACCGAAAGCTATTTCTTTAA
CCCGAACAAAGTGCGTAACCTGCTGGAGATCGACACCCTGATTCACCAGGAATACCACAAGGATATCAAC
ATCAACTTCAATAACGAGAACAAGCTGTTCGAGAAGTTCGAAATCCAAAAACGTATTAGCCCGGGTGGTG
AGTTCATCCAGAACAGCAAAGTGGAGCAAGAAAGCATTAACAAGTTTATCAGCCTGATTATGCAGTACAA
AAACAAGAGCAACTTCCTGAGCGAATATATCCTGCTGAACCTGATTGACGTTCAAAACAAGCGTATCATT
AACCTGAGCGCGAGCGAGTTTCTGAGCTACAAAAACAACGTGAGCATCGAATTTATTGATAAGCAGCTGA
CCCGTCGTAACTTCAACCTGCTGGCGGTTCCGTACATCCTGTATGTGGGCGTTAACGAGGAAAAAATTAA
GGAGAACTACAGCAACAACTACTTCAAAATCAGCCGTATCATCGCGGGCTTCTGGAGCGGTGTGGTTAGC
ATTCTGAGCGCGCAGTGCGGTCTGAGCACCCACCCGATGATGAGCTACAACGCGCGTGAACTGGAGGAAT
ACATCTTCAAGAACCGTTATAGCATTCTGAACCAAATCGTTATTGGCGGTAACATCACCACCAACCGTAT
GGACTCCATGCTGATTCGTAGCGATCTGAGCGGCGGACATCATCACCATCACCATTAACTGCAGCCCgag
ctc

His_6TclI_{NTD} (KpnI/XbaI) insert

ggtaccGGAGGTCACATATGCATCATCACCATCACCATGGCGGAAGCTACCTGCTGTTCAAGAAAGACAC
CTTTTATATCAAGACCCACAACGAGGGTATTCTGTTCAAAAACAACTTTACCAACCTGGAAGTGAAGACC
CATAAAAGCTACTATGTTTTCGAGAACCTGATCGAATACCTGAACGGCAGCTATACCGAGAACCAGATCA
TCGAAAACATCAAGAACAAGAAAGTGGCGCTGTTCTGCAAAAACATCATTAACGTTCTGAAGGAGAAAAA
CTTTATCTACGCGAGCGAAAACCGTCTGGATAACTAAtctaga

His_6TclI_{CTD} (KpnI/XbaI) insert

ggtaccGGAGGTCACATATGCATCATCACCATCACCATGGCGGAAGCGAAAACCGTCTGGATAACCTGAG
CGAGCTGGAAATCAAGATTCTGTATCTGAACAGCAAGAACATCCAACTGAGCAAAGACGCGTTCCTGAAC
AACAACAAGATCGATATTAAAACCTACAACCTGATGGAGAACGAGTTCGCGAACACCATGCTGAAGGAGT
ACTATTTTAGCCCGAGCGAAACCGCGAACATCAAGATTAGCCTGGGTAAATTTGAGAGCGACATCGGCTA
CTATATCATTCCGGATGGTGAATACCTGATTGTGGGCAAAAGCAAGAAACAGCACACCCACACCGTTAGC
CTGAGCACCTATGAGATCCCGCTGCACGCGTGGCTGATCTGCCTGAACATTCTGCTGAGCGAACTGTTCC
TGTACTGCACCAGCATTTATAGCGAGGACAAGTTCGATGAAAACTACTACAAGTTCAACGTGCAAACCTT
TGAGGGTGACTTCTTTAAGAAAGAAAACAAATAAtctaga

$TclJ_{NTD1}$ (XbaI/SacI) insert

tctagaGGAGGTCACATATGGAAATCTTCGAAAGCCCGGAGTTTAACATCATTCGTTACCTGAGCAACAG
CTATATTTTCAAGAGCACCAGCAAAGAACCGGACTTTCTGAACAGCATCAAAGAGCGTATTCTGCTGAAT
ACCAACGATATCATTTACAAGAAAAACATCCCGGAACGTAACATTAGCTTCCTGAGCCGTATCGAGAACA
GCATTGGTAACTTTATCTGCAACCAGATTAGCGAATACGACTACATCATCTACATCTACAACCACAAGAC
CAACAAGATCAAGAAAGTGAACTACTTCATCAGCCCGCACTACTATAAGAAACTGCCGAACGACACCACC
AACTGCTTCAACTAAgagctc

$TclJ_{NTD2}$ (XbaI/SacI) insert

tctagaGGAGGTCACATATGGAAATCTTCGAAAGCCCGGAGTTTAACATCATTCGTTACCTGAGCAACAG
CTATATTTTCAAGAGCACCAGCAAAGAACCGGACTTTCTGAACAGCATCAAAGAGCGTATTCTGCTGAAT
ACCAACGATATCATTTACAAGAAAAACATCCCGGAACGTAACATTAGCTTCCTGAGCCGTATCGAGAACA
GCATTGGTAACTTTATCTGCAACCAGATTAGCGAATACGACTACATCATCTACATCTACAACCACAAGAC
CAACAAGATCAAGAAAGTGAACTACTTCATCAGCCCGCACTACTATAAGAAATAAgagctc

GST-TclE$_{leader}$ (XbaI/SacI) insert

tctagaAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGTCCCCTATACTAGGTTATTGGAAAA
TTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCTTGAAGAAAAATATGAAGAGCATTTGTA
TGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAATTGGGTTTGGAGTTTCCCAATCTTCCT
TATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCATCATACGTTATATAGCTGACAAGCACA
ACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAG
ATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCTTAGCAAGCTA
CCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCC
ATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATACATGGACCCAATGTGCCTGGATGCGTT
CCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGC
AAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAAGGAG
GAGAAAACCTGTATTTTCAAGGCGGATCCGAGTTCCAGACCAACAACATCGAAGGTCTGGACGTGACCGA
TCTGGAGTTTATTAGCGAGGAAGTTACCGAAAAGGACGAGAAAGAAATCATGGGTGCTTCTTgagctc

L501A2 vector sequence

Features:

LacI ORF

TAC promoter

Multiple cloning site

AmpR ORF

ColE1 ori

GST-TclE

gtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgta
tccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtATGAGTATTCAA
CATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGC
TGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAG
CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTA
TGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGA
ATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATG
CAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAG
GAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACT
ATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT
GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGC
GTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACAC
GACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAG
CATTGGTAActgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaattta
aaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttcca
ctgagcgtcagaccccgtagaaaagatcaaaggatcttcTTGAGATCCTTTTTTTCTGCGCGTAATCTGC
TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTT
TTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCT

GCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGT
CGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCT
ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGC
AGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCG
GGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAaa
cgccagcaacgcggccttttacggttcctggcctttgctggcctttgctcacatgttctttcctgcg
ttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaa
cgaccgagcgcagcgagtcagtgagcgaggaagcggaagtctagtccgacaccatcgaatggtgcaaaac
ctttcgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatGTGAAACCAGTAACG
TTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCC
ACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACCGCGT
GGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCG
CCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGG
TAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCT
GATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCGGCG
TTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGAC
TGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGT
CTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAA
CGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTC
CCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCT
GCGCGTTGGTGCGGATATCTCAGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCG
TTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTC
AGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCC
CAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGA
CTGGAAAGCGGGCAGTGAgcgcaacgcaattaatgtaagttagctcactcattaggcacaattctcatgt
ttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcggaagctgtggt
atggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttctggataatgt
tttttgcgccgacatcataacggttctggcaaatattctgaaatgagctgTTGACAATTAATCATCGGCT
CGTATAATGtgtggaattgtgagcggataacaatttcacacaggaaactctagAAAGGAGGTGGAGGTCA
TGTCCCCTATACTAGGTTATTGGAAAATTAAGGGCCTTGTGCAACCCACTCGACTTCTTTTGGAATATCT
TGAAGAAAAATATGAAGAGCATTTGTATGAGCGCGATGAAGGTGATAAATGGCGAAACAAAAAGTTTGAA
TTGGGTTTGGAGTTTCCCAATCTTCCTTATTATATTGATGGTGATGTTAAATTAACACAGTCTATGGCCA
TCATACGTTATATAGCTGACAAGCACAACATGTTGGGTGGTTGTCCAAAAGAGCGTGCAGAGATTTCAAT
GCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACT
CTCAAAGTTGATTTTCTTAGCAAGCTACCTGAAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAA
CATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTTATA
CATGGACCCAATGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCA
CAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTG
GTGGTGGCGACCATCCTCCAAAAGGAGGAGAAAACCTGTATTTTCAAGGCGGATCAGAATTCCAAACAAA
CAATATCGAAGGTTTAGATGTCACTGATTTAGAATTTATCAGTGAAGAAGTTACTGAAAAAGACGAGAAA
GAAATCATGGGTGCTTCTTGTACTACATGTGTTTGTACATGCAGTTGTTGTACAACTTAAggatccggct
gacgcgtacaggaaacacagaaaaaagcccgcacctgacagtgcgggctttttttttcgaccaaaggtaa
cgaggtaacaaccatgcgagtgttgaagtggcagaactgctgaacgcaggtctgggcggttctgataacg
agtaatcgttaatccgcaaataacgtaaaaacccgcttcggcgggtttttttatgggggggagtttaggga
aagagccagatccctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcg
ccgctacagggcgcgtcag

Supplementary Table 4. Plasmid Table

Plasmid name	Parent Plasmid	Name in figure	Insert
pJG1013	pJG743	His_6-TclI	His_6-TclI
pJG1066	pJG743	His_6-TclJ	His_6-TclJ
pJG1068	pJG743	TclN-His_6	TclN-His_6
pJG1070	pJG743	TclIJN	TclI, TclJ, TclN
pJG958	pJG743	His_6-TclIJN	His_6-TclI, TclJ, TclN
pJG964	pJG743	His_6-tclIJ	His_6-TclI, TclJ
pJG965	pJG743	His_6-TclIN	His_6-TclI, TclN
pJG968	pJG743	His_6-tclIE(GST-E)	His_6-TclI, GST-TclE
pJG1047	pJG743	N/A	His_6-$TclI_{NTD}$
pJG1051	pJG743	N/A	His_6-$TclI_{NTD}$, TclJ
pJG1053	pJG743	N/A	His_6-$TclI_{NTD}$, TclN
pJG1052	pJG743	N/A	His_6-$TclI_{NTD}$, GST-TclE
pJG1038	pJG743	N/A	His_6-$TclI_{CTD}$,
pJG1040	pJG743	N/A	His_6-$TclI_{CTD}$, TclJ
pJG1042	pJG743	N/A	His_6-$TclI_{CTD}$, TclN
pJG1041	pJG743	N/A	His_6-$TclI_{CTD}$, GST-TclE
pJG1062	pJG743	N/A	His_6-TclI, $TclJ_{NTD1}$
pJG1061	pJG743	N/A	His_6-TclI, $TclJ_{NTD2}$
pJG1065	pJG743	N/A	His_6-TclI, GST-$TclE_{leader}$
pJG985	pED022	GST-TclE	GST-TclE
L501A2	N/A	N/A	GST-TclE

Supplementary Table 5. Strain Table

Strain	Background	Plasmid(s)
D104	BL21	pJG743
D158	BL21	pJG1012
D148	BL21	pJG1066
D149	BL21	pJG1068
D143	BL21	pJG985
D142	BL21	pED022
D091	BL21	pED022, pJG743
D147	BL21	pJG985, pJG1070
D094	BL21	pJG985. pJG958
D101	BL21	pJG964
D102	BL21	pJG965
D103	BL21	pJG968
D136	BL21	pJG1047
D132	BL21	pJG1051
D133	BL21	pJG1052
D134	BL21	pJG1053
D135	BL21	pJG1039
D116	BL21	pJG1040
D118	BL21	pJG1041
D117	BL21	pJG1042
D074	Nico 21	pJG958
D075	Nico 21	pJG964
D076	Nico 21	pJG965
D154	BL21	pJG1061
D155	BL21	pJG1062
D153	BL21	pJG1063
PB524	DH5α	L501A2