Cell-free protein synthesis of secondary metabolite ribosomal peptides

Samuel Merryn Jones

School of Biosciences

University of Kent

Supervisor: Dr Simon Moore

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification at the University of Kent, or any other University or Institution of learning.

Samuel Merryn Jones

17/12/20

Abstract

Recently, there has been an increased emergence of antimicrobial resistant organisms and a decrease in antibiotic discovery, together producing a public health threat. To overcome this problem, novel antimicrobial scaffolds require development or existing structures need diversifying.

Natural products are a group of diverse chemicals many of which display antimicrobial activity. The biosynthetic gene clusters encoding these products are often termed 'cryptic' due to not being expressed under normal laboratory conditions. Thus, making isolation of novel products difficult.

Cell-free protein synthesis has recently been explored for its potential application in facilitating heterologous expression of natural product clusters as well as peptide diversification via incorporation of non-canonical amino acids. As it's non-living, directly manipulatable nature lends itself to expression and alteration of toxic products.

In this thesis we studied the potential of *E. coli* cell-free protein synthesis for expression of 3 putative lasso peptides, a class of natural product characterised by their lasso structure, isolated from *Streptomyces* genomes. The natural product clusters were amplified from their respective genomes using an optimized PCR protocol for the amplification of high GC sequences and cloned into plasmids for expression in cell-free systems. However, further experiments are required to allow for their expression.

Moreover, steps were taken towards producing a *E. coli* cell-free system for incorporation of tryptophan analogues into peptides with a direct read out of successful incorporation. A deCFP marker was produced to allow for identification of ncAA incorporation, however supporting mutations within its sequence are required to increase quantum yields to suitable levels.

Furthermore, amino acid depletion steps are required during cell-free lysate production to allow for efficient incorporation of ncAAs.

Finally, bioinformatic investigation of RiPP biosynthetic gene cluster intergenic regions was carried out to probe for the presence of intrinsic terminators involved in their regulation. This identified secondary structures present within the intergenic regions of all RiPP clusters tested, par one. Furthermore, MSA of intergenic regions of clusters homologous to the 3 clusters of interest suggested that these secondary structures may be conserved across lasso peptides.

Acknowledgements

I would like to take this opportunity to firstly thank Dr Simon Moore for his guidance, patience and support through my time in his lab. I feel that this research experience has allowed me to develop greatly as a scientist. Furthermore, I would like to also thank all the members of the lab (Agata, Sarah and Seth) for your support throughout the year and for making the experience so enjoyable.

Table of contents

Abbreviations

aaRS	aminoacyl-tRNA synthetases
AMR	antimicrobial resistance
BGCs	biosynthetic gene clusters
cAA	canonical amino acid
CFPS	Cell-free protein synthesis
CFP	cyan fluorescent protein
eCFP	enhanced CFP
gDNA	genomic DNA
GFP	Green fluorescent protein
GSM	global suppression method
His_6	hexahistidine
MIBiG	minimum information about a biosynthetic gene cluster
MSA	multiple sequence alignments
ncAA	non-canonical amino acid
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthases
OTS	orthogonal translation system
OD	optical density
PTMs	post translational modification enzymes
RiPPs	Ribosomally synthesised and post-translationally modified peptides
rSAP	Shrimp alkaline phosphatase
(v/v)	Volume/Volume
(w/v)	Weight/Volume

1 Introduction

1.1 Secondary Metabolites

Natural products can be simply defined as small molecules that are produced by biological sources [1]. In addition, secondary metabolites are a subset of natural products which are a group of structurally and chemically diverse molecules that are not essential for organism growth [2]. Instead, they allow organisms to adapt to their environment by facilitating signalling, defence, and competition for resources[2]. They are widely produced in nature by animals, plants and microorganisms and have been utilised by humans due to their applications as anti-tumour drugs, pigments, pesticides, and antibiotics, allowing for development in health, agriculture, and economy [2].

1.2 Antibiotics and Resistance

The earliest records of natural products are from Mesopotamia in 2600 B.C. which included oils produced by *Commiphora* species (myrrh) and *Cupressus sempervirens* (Cypress) which are still in used today as treatments for inflammation, colds, and coughs[3]. However, before the 20th century infectious diseases were responsible for higher morbidity and mortality around the world compared to the present day, exemplified by the Bubonic plague epidemic accounting for death of approximately 1/3 of Europe's population from 1347 and 1350.

The discovery of penicillin (a β-lactam antibiotic from the fungus Penicillium rubens [4]) by Sir Alexander Fleming in 1928 [5] marked a turning point in modern medicine, by opening the door to antibiotic exploration, ultimately saving millions of lives worldwide.

From 1940 – 1970 a whole plethora of natural antimicrobials and chemically synthesised derivatives were produced such as actinomycin, from *Streptomyces spp* [6]., neomycin from *Streptomyces fradiae* [7], fumigacin from *Aspergillus fumigatus* [8] and streptomycin [9],

producing a defence against a variety of previously highly dangerous pathogens. This development was highly influenced by the work of Selman Waksman who developed the first systematic research platform for the discovery of antimicrobial activity of soil bacteria especially *Streptomyces* strains. This platform gave rise to all the examples of antibiotics given above and encouraged the pharmaceutical industries to pursue antimicrobial research, together leading to the development of a variety of antibiotic scaffolds during this period [10].

However, even during the 'golden age' of antimicrobial discovery clinical isolates showing resistance to antibiotics were seen. The emergence of antibiotic resistance is [4] famously exemplified by Methicillin-resistant *Staphylococcus aureus* (MRSA), which becomes resistant to methicillin upon clinical usage and was estimated by the centres for disease control and prevention (CDC) to kill 10,600 people in the US in 2017 [11]. In addition, following the end of the 'golden age', there was a 38-year gap between 1962 and 2000 [12], where no new antibiotic scaffolds, apart from carbapenems in 1985, were approved for clinical use [13]. Thus, exacerbating the growing problem of AMR [14]. Now, due to resistance and reduced production antibiotics, antimicrobial resistance (AMR) has become a global health concern with an estimated 33,000 deaths across Europe accounted for by AMR in 2015 [15].

This increase in AMR combined with the deceleration in antibiotic discovery emphasises the need for the development of new discovery platforms to enable the exploration of untapped chemical spaces to produce novel scaffolds and more chemically diversified derivatives.

1.3 Traditional approaches to antibiotic discovery

1.3.1 Waksman platform

In 1937 Waksman observed actinomyces soil bacteria inhibiting the growth of other bacteria, due to secretion of antimicrobial compounds as a product of the competition based selective

pressure [16]. This lead to the development of a method that utilises co-culturing a test strain (believed to have antimicrobial activity) and an indicator strain [10], inducing the expression of antimicrobial molecules by the test strain, thus allowing for the discovery of novel scaffolds. However, this era of discovery stopped when the platform hit a rediscovery problem, where there were no new novel antibiotics isolated due to the methods requirement for culturable and rapidly growing strains which produce large amounts of antibiotic [17], limiting the scope of exploration.

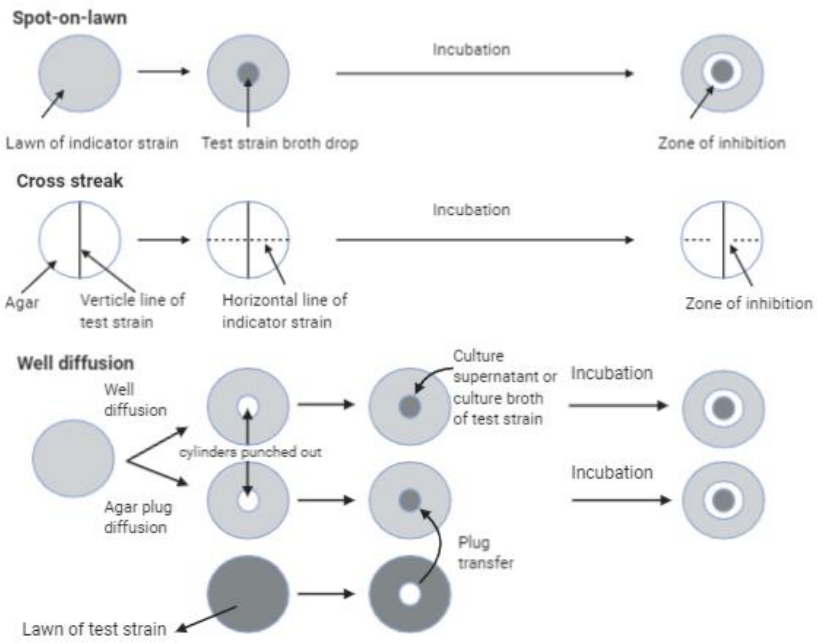


Figure 1 – Solid culture approaches to the Waksman platform, including spot-on-lawn, cross streak and well diffusion. Adapted from [10].

1.3.2 Semi-synthesis

A limitation of the Waksman platform was the pharmaceutical ineffectiveness and safety problems of some of the compounds produced. This was improved upon by semi-synthesis techniques carried out to chemically modify scaffolds to increase stable and remove

undesirable side effects [18]. This was first exemplified by the production of dihydrostreptomycin b via the catalytic hydrogenation of streptomycin in 1946 [19], which resulted in higher stability with similar activity [18]. Furthermore, this technique allowed for the development of a whole class of antibiotics, called Beta lactams. This class contains drugs such as ampicillin, carbenicillin and methicillin and in total is responsible for 60% of antibiotics used by humans [18].

1.3.3 Synthetic chemistry

The first example of an antibiotic produced by total synthesis was in 1949 when chloramphenicol was synthesised – a natural product isolated from *Streptomyces venezuelae*. This was further derivatised by replacing the nitro group with methanesulfonyl to produce thiamphenicol in 1952. Thiamphenicol has increased potency and a reduced toxicity, thus improving the clinical effectiveness of the drug [18]. The importance of this method is further supported by the production of fully synthetic beta-lactams as it allowed more intricate antibiotics to be synthesised leading to the development of a variety of subclasses such as synthetic monobactams and carbapenems. Synthetic carbapenems, first exemplified by the amine to N-formimidoyl transformation of thienamycin from *Streptomyces cattleya* to produce the more stable compound imipenem in 1979 [20], are currently our defence against multidrug resistant infections due to their widespread activity among B-lactams [18].

Together the methods described above have founded a new era of medicinal chemistry, discovering a range of clinically used drug classes such as aminoglycosides, macrolides, tetracyclines with reduced side effects and increased potency [18].

1.4 Genomic approaches to antibiotic discovery

There has been a slow in novel antimicrobial discovery by traditional approaches due to the rediscovery of known scaffolds, thus requiring progression into an age of genome mining to allow for discovery of novel scaffolds.

This progression was facilitated by the development of genome sequencing technologies and the discovery that the biosynthesis of these antimicrobial natural products being encoded for by biosynthetic gene clusters (BGCs). These clusters are physical groupings of genes required for the biosynthesis of the natural product, such as biosynthetic enzymes, post translational modification enzymes (PTMs), regulatory genes [21] and resistance genes. For example, before its complete genome had been sequenced and analysed in 2002, six BGCs for the production of distinct natural products in the organism *Streptomyces coelicolor* A3 (2) had been identified by traditional genetic approaches. However, after genomic analysis an additional 16 BGCs encoding specialised natural products were identified [22]. Thus, showing a deeper unexplored biosynthetic capability of microorganisms for the production of natural products [21].

Furthermore, the high level of sequence conservation of natural product biosynthetic machinery such as non-ribosomal peptide synthases (NRPS) or polyketides being polyketide synthases, allow for the identification of natural product BGCs within genomes [23]. Therefore, bioinformatic technologies such as antiSMASH and ClustScan have been developed to identify BGCs potentially encoding antimicrobial molecules, based on sequence similarity to known natural product enzymes. For example, referring back to the 16 additional clusters identified post genomic analysis, further bioinformatic analysis predicted several clusters may encode novel structures [22]. These clusters can then be induced to allow for the expression of the products of interest.

The importance of this chemical space is further exemplified by the genomic data-based prediction of 33,351 putative BGCs (false positive rate of 5%) in 1154 prokaryotic genomes [24]. This presents an opportunity to find a variety of untapped novel natural products, off which many may be antibiotic or anti-tumour encoded by gene clusters that are not expressed under normal laboratory conditions. This technique thus overcomes some of the limitations of the Waksman approach by being able to identify BGC's that potentially encode bioactive natural products, without the requirement of culturing.

Thus far, genome mining has procured a variety of novel antibiotics such as Brevicidine a non-ribosomal peptide (NRP) and laterocidine a cationic peptide which both exhibit antimicrobial activity against Gram-negative bacteria, including colistin resistant *E. coli* (mcr-1) via disrupting the cell membrane [25].

The features of biosynthetic gene clusters, the natural products they encode, regulatory elements and the application of their 'awakening' is described later.

1.5 Awakening cryptic BGCs

One limitation of traditional approaches is that natural product biosynthesis is tightly regulated meaning their expression is only induced under specific conditions, which are often unknown. This highlights why many natural products have been identified via genomics but were unexpressed via high-throughput antimicrobial screening projects. This is owed to the complex regulation pathways controlling gene expression in response to signal transduction pathways activated by environmental queues. Thus, enabling microbes to react to a changing environment, increasing survivability.

This has led to the phrases ‘cryptic’ or ‘silent’ being used to describe these types of clusters. To overcome this problem a variety of techniques have been developed and can be split into 2 broad categories: pleiotropic and pathway specific techniques, however there is some overlap.

Examples of pleiotropic techniques include: Ribosome/RNA polymerase engineering, growth condition manipulation/chemical elicitors, co-culturing, manipulation of global regulators and epigenetic perturbation, and SARP overexpression (semi-pleiotropic).Whereas pathways specific techniques contain: heterologous expression, local regulator manipulation, refactoring, reporter guided mutant selection, transcription factor decoys, and HiTES. Not all these techniques will be discussed here, but interested readers are directed to these papers to learn more [26][27][28][22][29].

1.5.1 Pleiotropic techniques

These techniques aim to induce the expression of a variety of BGCs within the natural host organism to give rise to multiple natural products. This therefore involves the use of manipulating environments to contain chemical elicitors of BGC gene induction or manipulating global gene expression to allow for the expression of multiple clusters, rather than specific clusters of interest.

Growth condition variation:

This involves changing conditions such as pH and temperature or the addition of competing species or chemical elicitors to induce expression[22]. For example, before this approach was utilised *Aspergillus ochraceus* DSM 7428 only had one known product, whereas by altering the growth conditions i.e. shaking flasks, static penicillin containing liquid cultures and different fermenters lead to the production of 15 metabolites based on polyketide synthases

[30]. Furthermore, in 2006 four novel prenylated quinoline alkaloids from *Aspergillus nidulans* were identified by screening 40 different culture conditions (including 6 different medias, submerged and stationary with varying cultivating periods) [31].

Chromosome remodelling:

This technique is based on the concept that when eukaryotic chromatin is in heterochromatin form it is transcriptionally silent whereas when it is in euchromatin form it has the potential to be transcriptionally active. Furthermore, eukaryotic natural product BGC are usually located towards the end of chromosomes where this regulation is common. Thus, DNMT and HDAC (enzymes that act to pack chromatin) inhibitors can elicit fungal natural product induction[14]. For example, a HDAC inhibitor (suberoylanilide hydroxamic acid) was used to induce the expression of multiple cryptic perylenequinones from *Cladosporium cladosporioides*, some of which were novel analogs of cladochrome[32]. Interestingly work has been carried out which shows that the use of this technique has been successful in *Streptomyces*, due to the presence of bacterial versions of HDAC proteins, although the exact mechanism of how this works has not been fully characterised. For example, expression of 5 previously cryptic gene clusters *S. coelicolor* were induced via sodium butyrate (class I and II HDAC inhibitor)[33].

Global regulator manipulation:

This approach utilises the deletion of repressors and the overexpression of activators of global gene expression which may be associated with secondary metabolism. This can also be utilised in a pathway specific approach if the regulators of specific clusters are identified by genomics. For example, a novel polythioamide called closthioamide was identified via the addition of aqueous soil extract to a *C. cellulolyticum* culture. However, due to the variable nature of soil

cultures some closthioamide congeners were not able to be isolated. In addition, the loci of the closthioamide locus was unknow ruling out the possibility of a specific approach. To overcome this, they utilised the overexpression of a global regulator, nusG, which resulted in the production of the original closthioamide and seven novel analogs[34].

An example of manipulating global regulators for the induction of target clusters is exemplified by a study where two genes encoding transcriptional regulators of the LuxR family, astG1 and vemR, were constitutively expressed in *Streptomyces sp.XZQH13* and *Streptomyces venezuelae*, isolating two known ansatrienins and a new biaryl polyketide venemycin[35].

1.5.2 Specific techniques

This involves removing specific regulation of individual BGCs, by genetic manipulation within the natural producer or in heterologous hosts.

Pathway specific regulator manipulation:

Natural product expression is often partly controlled by pathway specific regulators that tend to be found within the BGC. Therefore, removal of negative regulators and upregulation of positive regulators can be utilised to induce expression[22]. This is exemplified by a study where the genome sequence of *A. nidulans* was analysed and found a BGC encoding a hybrid PKS-NRPS assembly line which remained cryptic under a range of laboratory conditions. Analysis also identified an adjacent hypothesised transcriptional activator of the target cluster, which when integrated into the *A. nidulans* genome under the control of an inducible promoter lead to the production of novel metabolites, aspyridones A and B, which demonstrate moderate cytotoxicity[36].

Heterologous expression and refactoring:

Heterologous expression involves the induction of a specific biosynthetic gene cluster in a non-natural host, which can be engineered to enhance the approach. This is often combined with the refactoring of the pathway, i.e. the replacement of natural promoter with a constitutive or inducible promoter [22]. For example, analysis of *S. leeuwenhoekii* C34^{T} revealed the presence of 3 BGCs potentially capable of producing lasso peptides. From this one of the clusters lasso peptides, LP2, was undetectable in culture supernatants of the native host, however, when cloned and expressed under the control of a constitutive ermE* promoter in *S. coelicolor* it was isolated. This novel lasso peptide was subsequently named Leepeptin[37].

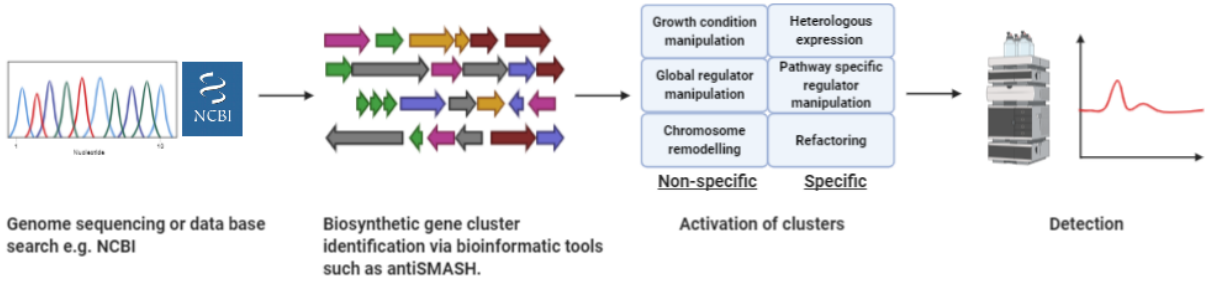


Figure 2 – A generalised workflow of genome mining for natural products, including BGC identification, methods of induction and detection via High Performance Liquid Chromatography (HPLC)/ Mass spectrometry.

1.6 Natural product biosynthesis

Natural products can be classified in terms of their biosynthesis into polyketides, ribosomally synthesised and post-translationally modified peptides (RiPPs) non-ribosomal peptides (NRPs), terpenoids, saccharides and hybrids [38].

There are 2 main classes of natural products that produce peptide secondary metabolites the ribosomally synthesised and post-translationally modified peptides (RiPPs) and non-ribosomal peptides (NRPs) [39]. They are distinguished by the mechanism of synthesis of the peptide with RiPPs being synthesised by the ribosome [40] and NRPs by non-ribosomal peptide synthases [41].

1.6.1 Ribosomally synthesised and post-translationally modified peptides

Introduction:

They are a group of structurally diverse natural products which have been identified in all 3 domains of life. This diversity is impart owed to the extensive post-translational modifications imparted on the peptide, furthermore, they make the peptides typically less conformationally flexible than natural ribosomal peptides, which increases their metabolic and chemical stability and allows for better target recognition [42].

Biosynthesis:

RiPP BGCs contain the genes encoding the precursor peptide, post-translational enzymes and sometimes export/resistance proteins. Their biosynthesis nearly always begins with a ribosomal peptide of ~20-110 residues being synthesised [42], which usually contains a N-terminal leader region and a C-terminal core region [43]. The leader region is important for recognition by post-translational modification enzymes and export [42]. The core region is subjected to post-translational modifications, maturing the peptide [43] into its active form. A signal sequence at the N-terminal domain is often present in eukaryotic RiPPs and directs the peptide to cellular compartments where the PTMs will take place. Furthermore, some peptides

have an additional C-terminal recognition sequence that is important in excision and cyclisation. The mature peptides are then proteolytically cleaved from the non-core regions [42].

Classification:

The different RiPPs can be classified based on their structural features and biosynthetic machineries into a variety of subgroups such as lanthipeptides, thiopeptides, microcins and lasso peptides [40]. The RiPP peptides investigated in this thesis are lasso peptides, so the subsequent section will focus on their structural features and biosynthesis.

1.6.2 Lasso peptides

Introduction:

Lasso peptides are 16-21 residues long [42] and are characterised by their 'lasso' structure formed by the threading of the C-terminal tail through a N-terminal right-handed macrolactam [40]. The macrolactam is achieved by the condensation of the carboxylate group of an aspartate/ glutamate at residue 8 or 9 with the N-terminal amine [42]. The presence of bulky side chain residues in the C-terminal tail then locks it within the macrocycle [40].

Classification:

They are classified by the presence and number of disulfide bonds in their structures. Class I lasso peptides contain 2 disulfide bonds, 1 is produced by the N-terminal cystine residue and the other connects the ring to the tail. Class II contain no disulfide bonds, instead the structure is stabilised by steric interactions. They are also the most common class. Furthermore, class 3 and 4 both have only 1 disulfide bond each – class 3 have theirs connecting the ring to the tail and in class 4 [44] the bond is just in the tail [45].

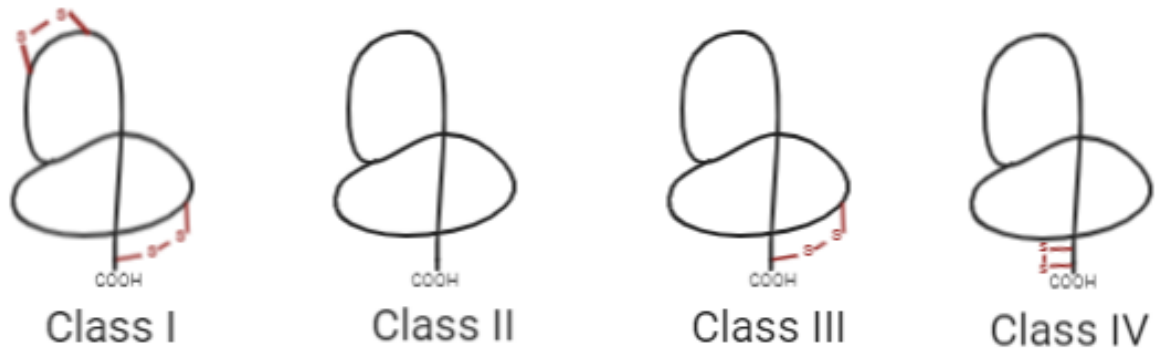


Figure 3 – Classes of Lasso peptides.

Activity:

They tend to be receptor agonists or enzyme inhibitors, providing some with antibacterial properties against Gram-negative or Gram-positive bacteria. For example, microcin J25 [46][47] and Capistruin [48][49] are RNA polymerase inhibitors and are antibacterial. Furthermore, siamycin I (MS-271) is a myosin light chain kinase inhibitor [50] and an anti-HIV agent [51][52]. RP-71955 [53] and siamycin II [51] are also both anti-HIV agents.

1.6.3 Microcin J25:

Biosynthesis:

Microcin J25 is the most characterised lasso peptide and belongs to class II [42]. Its biosynthetic gene cluster is composed of 4 genes, McjABCD. McjA encodes a 58 aa precursor peptide, in which the N-terminal 37 aa makes up the leader peptide [54]. Furthermore, McjB encodes an ATP-dependent cysteine protease which is suggested to complex with McjCs lactam synthase, which together covert the precursor peptide into the active microcin J25 [54][55]. MjcD encodes an ABC transporter for export of the active product [54][56]. The active final product is produced firstly by the action of McjB which cleaves the core peptide

from the leader peptide, exposing the N-terminal amine of Gly 1 of the core peptide [40]. The ATP hydrolysis mechanism of McjB is believed to be coupled with a pre-folding step required before lactam formation, rather than the proteolysis. Then, McjC activates the carboxylate side chain of Glu 8 via ATP-dependent adenylation [55], allowing for the nucleophilic attack of the N-terminal amine, forming the macrolactam [40]. The threaded 13 aa C-terminal tail is irreversibly held in place by the bulky side chains of the aromatic Phe19 and Try20 residues [57].

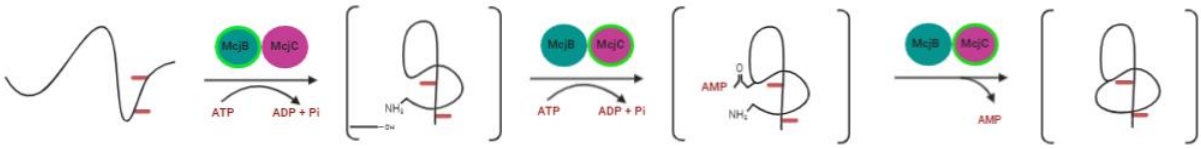


Figure 4 – Microcin J25 Biosynthesis. McjA (precursor peptide) seen in left panel, with Phe19 and Trp20 side chains represented by red lines at the C-terminal. The McjB region of the McjB and C complex, active region represented by green outline, catalyses the cleavage of the leader peptide from the precursor exposing the N- terminal amine of Gly1 and folding into required structure for next steps. Subsequent, activation of the carboxylate side chain of Glu8 by ATP-depended adenylation allows the nucleophilic attack of the N-terminal amine for macrolactam formation.

Adapted from Yan KP, Li Y, Zirah S, et al (2012) [55].

1. 7 Cell-Free Protein Synthesis

1.7.1 Introduction:

Cell-free protein synthesis (CFPS) was first utilised by Nirenberg and Matthaei around 60 years ago to decipher the genetic code and discover the connection between mRNA and protein

production [58]. It is a platform that uses crude cell extracts from the organism of choice that have undergone lysis, washing and other preparation steps to remove genomic DNA and cell debris. However, they retain the main components required for transcription and translation, such as ribosomes and aminoacyl-tRNA synthetases. Furthermore, when the crude extracts are supplemented with the necessary substrates, e.g. amino acids, an energy source, salts and DNA, synthesis of the desired proteins can occur [59].

1.7.2 Advantages:

The main advantages of this platform are facilitated by absence of a functional genome and cellular membrane reducing the constraints usually inflicted on live cells. This means that the reaction environment is directly manipulatable and problems of toxicity can be avoided. Together, this allows for higher production titers of some proteins, the synthesis of large proteins, membrane proteins, virus-like particles and difficult to synthesise proteins [60].

Furthermore, the systems can be arranged in different reactions set ups such as continuous flow, continuous exchange, and batch to allow for optimal conditions for the production of specific proteins. In addition, the scalability of the platforms allows for high-throughput screening and large scale biomanufacturing [60]. It has also been adopted as a tool for synthetic biology techniques such as metabolic engineering [60], natural product synthesis [61] and genetic code expansion [60].

1.7.3 Limitations:

It has been observed that cell-free systems total protein concentration is around 20 times lower than within cells and the rate of protein synthesis is usually slower in CFPS than in vivo. This is thought to be in part caused by the reduced translation rate due to unavailability of tRNA

and elongation factor Tu and the difference between transcription and translation rates, which are synchronous in vivo. Furthermore, generally the artificiality of the protein synthesis environment can make the transferability and predictability of CFPS challenging [62].

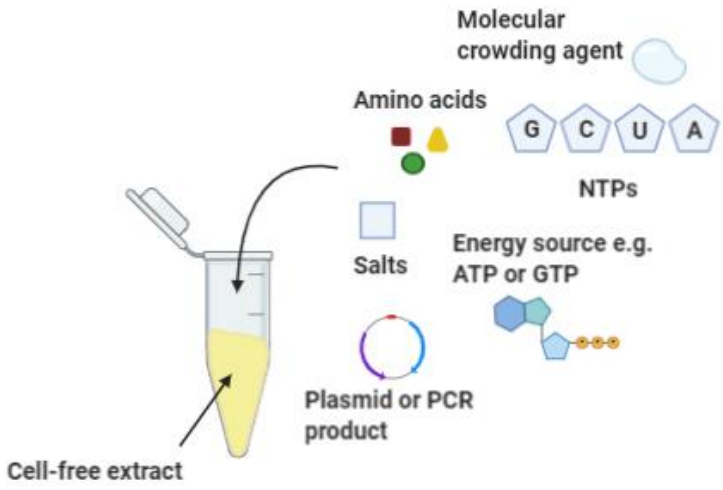


Figure 5 – Cell-free protein synthesis reaction set up.

1.7.4 Applications:

Genetic part and circuit prototyping:

One challenge for researchers is understanding the complexity and interplay of genetic elements. CFPS has multiple advantages for allowing for the testing of genetic element dynamics over in vivo approaches. Firstly, a tight control over the concentration of DNA and polymerase in the reactions. The rapid and quantitative report systems available and the larger parameter space that can be assessed via high throughput methods [60].

For example, CFPS has allowed for the prototyping of regulatory elements such as allosteric transcription factors, riboswitches and light-sensing promoters. The main method of this approach is to design libraries of single variants of a genetic element controlling a measurable

reporter protein and then using liquid handling robots, screen the function of the different elements [63].

High-throughput protein synthesis:

The platform has also been adopted to allow for the high-throughput synthesis of proteins due to the cost-effectiveness of the reactions, potential for automation and ability to use PCR products to avoid cloning [64]. For example, in a study by Goshima et al., 2008 12,996 clones (97.2% of the 13,364 tested) of human proteins were expressed and proteins synthesised via a wheat germ cell-free system. Furthermore, 57 out of the 365 clones that did not produce any protein were cloned into another plasmid with a fluorescent tag, pEW-5SG3Ven. This time all of these produced fluorescence and i.e. protein of intertest. This was tested further by 29 out of the 314 clones that expressed but didn't produce protein were put into the same plasmid above, and the same result was obtained, they all produced protein. In addition, 58 out of the 75 (77%) tested phosphatases were biologically active and several disulphide bond containing cytokines produced in a non-reducing wheat germ cell-free system were seen in an active form [65].

Natural product biosynthesis:

As previously mentioned, heterologous expression on natural products has been very useful for the discovery of novel molecules. However, problems have been encountered when trying to obtain high yields due to the toxicity, unavailability of precursors and metabolic burden it places on whole cells. Using CFPS has been explored to overcome this problem [66]. Furthermore, it has additional advantages over cell-based methods, such as being able to directly control the levels of DNA, precursors and co-factors in the system and the ability to run design-build- test cycles without engineering the host [67].

This cell free approach can be subdivided into 2 categories, purified enzyme biosynthesis and CFPS [67].

Purified enzyme approach:

This method involves expressing and purifying the enzymes involved in a biosynthetic pathway, potentially by using a heterologous host, then adding the enzymes, precursors, cofactors and supplements required to a 'one-pot reaction' to enable the synthesis of the desired product. This was first tested by Professor Ian Scott, who synthesised a late stage vitamin B_{12} precursor, hydrogenobyrinic acid, at around 20% yield via a 12 enzyme one-pot reaction [68]. However, the full synthesis of vitamin B_{12}, a 30-enzyme reaction, was not completed, it did represent a proof-of-concept for one pot enzymatic biosynthesis [69].

CFPS approach:

This approach uses cell-free extracts supplemented with the DNA of the desired biosynthetic gene cluster and the necessary precursors and cofactors to allow for the biosynthesis of the desired product. For example, *E. coli* based cell-free extracts have been utilised to synthesise the nonribosomal peptide molecule D-Phe-L-Prodiketopiperazine (DKP), which is a natural shunt product in the biosynthetic pathway for Gramicidin S via the cell free expression and synthesis of the two required nonribosomal synthases GrsA and GrsB1 in a one-pot reaction. This yielded approximately 12 mg/L, which is higher than previously recorded cell-based production. Furthermore, this experiment showed the ability of the platform to produce active enzymes of >120kDa [70]. Furthermore, CFPS has been utilised for the synthesis of the lanthibiotic nisin [61]and four lasso peptides (known burhizin, fusilassin, capistruin and cellulassin) [71].

Non-canonical amino acid incorporation:

The ability to incorporate non-canonical amino acids (ncAA) into proteins of interest has been a growing field over the last 60 years. This is due to the inability of canonical proteins to meet the demands of scientific inquiry due to their limited diversity and functionality. Incorporation of non-canonical amino acids into proteins allows for the diversification of their biological and physiochemical properties, allowing for potentially novel functionality. This technique has been explored for use in a variety of applications such as mimicking PTMs, studying molecular interactions, labelling, unnatural enzymes and biomaterials. Now, more attention is moving to the use of cell-free systems to overcome challenges associated with in vivo expression. For example, in vivo techniques are often compromised by toxicity, getting ncAAs across membranes and low incorporation efficiencies [72].

There are 2 main methods that have been used for ncAA incorporation, the global suppression method (GSM) and the orthogonal translation system (OTS) [72].

The GSM in CF works by preparing crude cell extracts from auxotrophic strains depleted of the natural amino acids, e.g. by size-exclusion chromatography or growth in minimal media. When the system is supplied with the ncAAs of interest, as long as they are similar enough to the canonical amino acid (cAA) that its recognised by the native machinery, it will be incorporated globally (residue specific) into proteins [72]. For example, Singh-Blom et al. 2014 incorporated 5- and 6- fluorotryptophan into streptavidin via a cell-free global suppression method with 100% efficiency using E. *coli* BL21 (DE3) ΔtrpC where the canonical amino acids were depleted via growth in the minimal media, Medium G-PG, for 50 minutes prior to extract production [73].

The OTS utilises exogenous orthogonal aminoacyl-tRNA synthetases (aaRS) and tRNAs pairs and can be subdivided into 4 methods, stop codon suppression, the sense codon reassignment,

frame-shift suppression, and unnatural base pair. The stop codon suppression method works by choosing one of the 3 non-sense stop codons, UAG, UAA and UGA, to encode the ncAA. However, there is significant competition with the tRNA and release factor 1 (RF-1) [72], so deleting RF-1 results in higher yields [74][75]. For example, cell-free extracts produced from genomically recoded *E. coli* lacking RF-1 allowed for 2.5 X increased expression of sfGFP containing p-propargyloxy-L-phenylalanine (pPaF) or p-acetyl-L-phenylalanine. In addition, the TAG stop codons of 13 essential genes were mutated to TAA stop codons, to reduce inhibition to strain growth [76]. The sense codon reassignment method works by assigning the sense codons to ncAAs, however, there is competition between the native tRNA and artificial tRNAs. Frame-shift suppression works by creating enlarged codons with 4 or 5 nucleotides and the unnatural base pair system works by creating new non-natural codons that are assigned to ncAAs. The advantage of these methods over GSM is that they can incorporate ncAAs into specific positions rather than just at the residue specific level [72].

1.8 Summary

Over recent times there has been an increase in the incidence of antimicrobial resistance and a slow in antibiotic production, producing a public health care crisis. The development of genome sequencing technologies has opened a new avenue for identifying natural products via genome mining – however, large amounts of natural product encoding BGCs remain ‘cryptic’ due to complex regulatory systems. To overcome this problem and allow for expression and isolation of products, a variety of techniques have been developed such as growth condition manipulation, regulator manipulation and heterologous expression. Cell-free protein synthesis has recently been explored as a platform for such heterologous expression and has advantages over traditional cellular techniques such as toxicity tolerance and easy manipulation.

In addition to discovering novel natural products another potential route of combating the reduced rate of drug discovery is the diversification of current scaffolds. One potential method of this is via the incorporation of halogenated non-canonical amino acids. CFPS provides an advantageous platform for such studies and has been utalised to allow for analogue incorporation.

1.9 Aims

1: Utilisation of *E. coli* cell-free protein synthesis for lasso peptide expression.

2: Production of an *E. coli* cell-free platform for efficient incorporation of tryptophan analogues into peptides of interest with a direct read out of successful incorporation.

3: Bioinformatic assessment of gene regulation within the intergenic regions of RiPP BGCs

2 Materials and Methods

2.1 Bacterial strains and plasmids

2.1.1 Table 1 - Bacterial strains

Strain	Description	Source
E. coli		
DH10B	N/A	N/A
Rosetta	N/A	N/A
JM109	N/A	N/A
BL21 Gold (DE3) or BL21 Gold (DE3) PLysS	N/A	N/A
RF4 (aspC tyrB)	*BL21(DE3), aspC tyrB genes deleted*	Robert Gennis & Toshio Iwasaki (Addgene plasmid # 62072)[77]
Streptomyces		
Streptomyces venezuelae	ATCC 10712	Dr Simon Moore
Streptomyces rimosus	DSM40260	Dr Simon Moore

2.1.2 Table 2 - Plasmids

Plasmid	Description	Reference
Pr-deGFP-Mgapt	pBEST vector, OR2-OR1-Pr promoter, UTR1-deGFP-MGapt-T500, AmpR	Pr-deGFP-MGapt was a gift from Richard Murray (Addgene plasmid # 67734 ; http://n2t.net/addgene:67734 ; RRID:Addgene_67734) [78]

Pr-deCFP-Mgapt	pBEST vector, OR2-OR1-Pr promoter, UTR1-deCFP-MGapt-T500, AmpR	This work
T7 sfGFP	T7His promoter, PET RBS, sfGFP, Bba_B0015 terminator, AmpR and pUC19 origin. Streptomyces codon optimised.	Dr Simon Moore
T7 eGFP	T7His promoter, PET RBS, eGFP, Bba_B0015 terminator, AmpR and pUC19 origin. Streptomyces codon optimised.	Dr Simon Moore
pTU1 SP44 mVenus	sp44 promoter, pETRBS, mVenus, rrnB T1 terminator, T7te terminator, AmpR	Dr Simon Moore
pTU1-A-T7RBS_Albusnodin_H	pTU1-A vector backbone (AmpR) with the Albusnodin homolog precursor peptide gene under the control of a T7 promoter and RBS	This work
pTU1-A-T7RBS_Lagmysin_H	pTU1-A vector backbone (AmpR) with the Lagmysin homolog precursor peptide gene under the control of a T7 promoter and RBS	This work

pTU1-A-T7RBS_Moomysin_H	pTU1-A vector backbone (AmpR) with the Moomysin homolog precursor peptide gene under the control of a T7 promoter and RBS	This work
pTU1-A-T7RBS_Albusnodin_H_His6	pTU1-A vector backbone (AmpR) with the Albusnodin homolog precursor peptide_His6 gene under the control of a T7 promoter and RBS	This work
pTU1-A-T7RBS_Lagmysin_H_His6	pTU1-A vector backbone (AmpR) with the Lagmysin homolog precursor peptide_His6 gene under the control of a T7 promoter and RBS	This work
pTU1-A-T7RBS_Moomysin_H_His6	pTU1-A vector backbone (AmpR) with the Moomysin homolog precursor peptide_His6 gene under the control of a T7 promoter and RBS	This work

pSF1C-A-SP44a_pETRBS_Albusnodin_H_PTMs	pSF1C-A vector backbone (ApmR) with the Albusnodin homolog PTM genes under the control of a sp44a promoter and pETRBS	This work
pSF1C-A-SP44a_pETRBS_Lagmysin_H_PTMS	pSF1C-A vector backbone (ApmR) with the Lagmysin homolog PTM genes under the control of a sp44a promoter and pETRBS	This work
pSF1C-A-SP44a_pETRBS_Moomysin_H_PTMs	pSF1C-A vector backbone (ApmR) with the Moomysin homolog PTM genes under the control of a sp44a promoter and pETRBS	This work
T7- BovT150M	BovT and M under control of a T7 promoter and pETRBS with an AmpR cassette.	Dr Risat Haque, University of Warwick
T7-BovA	BovA under control of a T7 promoter and pETRBS with an AmpR cassette.	Dr Risat Haque, University of Warwick
pKD13	KanR gene flanked by FRT regions. R6K γ ori, λ tl3 terminator, AmpR	Datsenko, KA, BL Wanner 2000. [79]

pKD46	Lambda Red recombinase expression plasmid; AraBAD promoter, alpha, beta and gama genes, λ tl3 terminator, pSC101 ori, AmpR, AraC	Datsenko, KA, BL Wanner 2000. [79]

2.2 Bacterial growth conditions

2.2.1 Media

Media used within this study included **Lysogeny broth (LB)** (10 g / L Tryptone, 5 g / L Yeast extract and 5 g / L NaCl made up with miliQ water and autoclaved), **LB24** (10 g / L Tryptone, 24 g / L Yeast extract and 10 g / L NaCl made up with miliQ water and autoclaved), **Terrific broth (TB)** (12 g/ L Tryptone, 24 g / L Yeast extract, 0.4 % V/V Glycerol, made up to 900 mL in miliQ water, autoclaved and topped up to 1 L with filter sterilised 0.72 M K_2HPO_4 and 0.17 M KH_2PO_4. **Cell-free autoinduction media** (3.2 g /L NaCL, 12.8 g /L Tryptone, 3.2 mg / L Yeast extract, 8.96 g /L Potassium Phosphate dibasic (K_2HPO_4), 3.84 g / L Potassium phosphate monobasic (KH_2PO_4) up to 960 mL miliQ water and pH adjusted to 7.2 with 5 M KOH and autoclaved [80]. **M9 media** (200 mL M9 salts (90.36 g / L Na_2HPO_4-$7H_20$, 30 g / L KH_2PO_4,5 g /L NaCl, 10 g NH_4Cl in miliQ water and autoclaved), filter sterilised 2 mM $MgS0_4$, 0.4 % glycerol and 0.1 mM $CaCl_2$ topped up with miliQ water.

2.2.2 Overnight culture (O/N)

O/N cultures containing 1 X appropriate antibiotic were produced via inoculation of 5 mL media with 1 colony of bacteria and incubated overnight at 37oC, 180 rpm for 16-24 hours.

2.2.3 Antibiotics

Antibiotic	Working concentration (µg/mL)
Ampicillin	100
Carbenicillin	100
Kanamycin	25
Apramycin	20
Chloramphenicol	35

Table 3 - antibiotics

2.3 General molecular biology techniques

2.3.1 Table 4 - Primers

Primer name	**Sequence 5' – 3'**
Cloning	
Albusnodin_RiPP_ NdeI_F	CACCATATGACCGATCTGCCGCGTACG
Albusnodin_RiPP_ BamHI_R	CACGGATCCTCAGCAGTTGTAGGCCCGCCGCTTGTCC
Albusnodin_RiPP_ His6_BamHI_R	CACGGATCCTCAGTGGTGGTGGTGGTGGTGGCAGTTGTAG GCCCGCCGCTTGTCC
Albusnodin_PTM_ NdeI_F	CACCATATGCCCATCGGAGGATTCAGC

Albusnodin_PTM_R	CACTTAATTAACTACTTCTCAGGGAACAGCGCCTTCACC
Lag_RiPP_NdeI_F	CACCATATGATGCGCTCAGCGCTCTTACG
Lag_RiPP_BamHI_R	CACGGATCCTCAGAGGAGCTGCTTGCCGCCGAG
Lag_RiPP_His6_BamHI_R	CACGGATCCTCAGTGGTGGTGGTGGTGGTGGAGGAGCTGCTTGCCGCCGAG
Lag_PTM_NdeI_F	CACCATATGGAAGAGTCCCTGTCCTTGTCG
Lag_PTM_PacI_R	CACTTAATTAATCACTCCACGGTGAAGAAGCGCTG
MM_NdeI_F	CACCATATGCGGGAAGCGCCGGAAGC
MM_BamHI_R	CACGGATCCTCACCAGAAGGTGCTGTGCTGG
MM_BamHI_R_6His	CACGGATCCTCAGTGGTGGTGGTGGTGGTGCCAGAAGGTGCTGTGCTGG
MM_NdeI_PTM_F	CACCATATGGATTTCGTCATCTTCCCCGACC
MM_PacI_PTM_R	CACTTAATTAATCAGCGCGGGACGCGCAGCAGC
Cloning sequencing	
VF2	TGCCACCTGACGTCTAAGAA
VR	ATTACCGCCTTTGAGTGAGC
A_PTMs_sequencing_1	TGGGCATGGTCCACGTCC
A_PTMs_sequencing_2	TGGCGACAGCGTCCTGTTC

A_PTMs_sequencing_3	TGGCGTCCCTGGATCTCATC
A_PTMs_sequencing_4	TGCCTGGATCACCACTCACG
S_PTMs_sequencing_1	ATCTGCGGCGGGTGTTC
S_PTMs_sequencing_2	TCTGGATACGCCCAGCATGC
S_PTMs_sequencing_3	TGATCGTCGAGGCCATGCG
S_PTMs_sequencing_4	TCATCGCACCGTAGCCACC
M_PTMs_sequencing_1	ACGATCCGGCAGGTCCAC
M_PTMs_sequencing_2	GGAACACCTGTCGCTGTCC
M_PTMs_sequencing_3	TCCTGGACGACACCGTCATC
Lambda red mutagenesis	
TrpL-FRT_F	CGAACTAGTTAACTAGTACGCAAGTTCACGTAAAAAGGGTATCGACAATGTGTAGGCTGGAGCTGCTTCG
TrpC-FRT_R	GTTAAGTAATGTTGTCATTGTTCCTTTCCTTAATATGCGCGCAGCGTCTGATTCCGGGGATCCGTCGACC

TnaA_F	AATATTCACAGGGATCACTGTAATTAAAATAAATGAAGGATTATGTAATGTGTAGGCTGGAGCTGCTTCG
TnaA_R	ACATCCTTATAGCCACTCTGTAGTATTAATTAAACTTCTTTCAGTTTTGCATTCCGGGGATCCGTCGACC
Colony PCR verification	
FRT_F	AGATGACAGGAGATCCTGC
FRT_R	CGTGACCCATGGCGATGC
Trp_gDNA_F	CGCACTCCCGTTCTGGATAATG
Trp_gDNA_R	GCCAAACTCACCAAAATAGG
deGFP to deCFP golden gate mutagenesis	
deGFP-deCFP_F	AAAGGTCTCACTGGGGCGTGCAGTGCTTC
deGFP-deCFP_R	AAAGGTCTCACCAGGTCAGGGTGGTCAC

2.3.2 PCR

Polymerase chain reactions (PCR) to amplify DNA fragments were produced using the reaction set up shown in component and thermocycler tables shown below.

Component	**Final Concentration**
Reaction Buffer (5X)	1X
Forward Primer	0.5 µM
Reverse Primer	0.5 µM
High GC enhancer (5X)	1X
dNTPs	200 µM
Plasmid/ DNA	Variable
Sterile ddH_2O	Up to either 25 or 50 µL

2.3.3 Colony PCR

Colony PCR was carried out following the general PCR protocol, however instead of plasmid/DNA 1 µL of cell mix was used (1 colony aspirated in 50 µL sterile ddH_2O).

2.3.4 PCR Thermocycling protocol

Stage	Temperature (°C)	Time	No. Cycles
1	98	30 sec	1
2	98	10 sec	30
	57 - 72	20 sec	
	72	30 sec per kB	
3	72	30 sec per kB + 1 min	1
4	8	Infinite	1

2.3.5 Colony PCR thermocycler settings:

Stage	Temperature (°C)	Time	No. Cycles
1	98	2 minutes	1
2	98	10 sec	30
	57 - 72	20 sec	
	72	30 sec per kB	
3	72	30 sec per kB + 1 min	1
4	8	Infinite	1

2.3.6 DNA agarose gels

For the analytical separation of DNA fragments, 1% (w/v) agarose gels were used, containing 1X TAE buffer, 1/25,000 SYBR safe and are run at 100 V for 33 minutes. 7.5 µL ladder is

loaded to the first available lane and samples (mixed with 5X or 6X loading dye) are added to adjacent wells.

2.3.7 PCR clean up

5X volume of buffer PB was added to PCR products, transferred to QIAquick Gel Extraction spin columns, centrifuged at 13,000 rpm for 1 minute and supernatant discarded. 500 µL QG buffer added, centrifuged and supernatant discarded, as above. 750 µL PE buffer added, centrifuged for 10 seconds and supernatant discarded. Column centrifuged for 3 minutes and subsequently transferred to a sterile 1.5 mL Eppendorf. 30 µL sterile ddH_2O added to column, left for 1-5min and centrifuged for 1 minute.

2.3.8 Gel extraction

Regions of agarose gel containing PCR fragment of interest were excised, placed in sterile Eppendorf's, and weighed. Volume of Buffer QC buffer equal to weight of excised gel pieces was added and incubated at 42°C for 5-10 minutes. If DNA fragment <200bp or >6000bp, 1 volume of isopropanol was added. Sample transferred to column, spun at 13,000 rpm for 1 minute and supernatant discarded. 500 µL buffer QC added, spun and supernatant discarded. 750 µL PE buffer added, spun for 10 seconds and supernatant discarded. Column spun for a further 3 minutes, transferred to an Eppendorf, 30 µL sterile ddH_2O, incubated for 1-5 minutes and spun for 1 minute.

2.3.9 Miniprep DNA purification

O/N cultures centrifuged at 4000 rpm, 4°C for 10 minutes, and supernatant discarded. Pellet resuspended in 250 µL Buffer P1 and transferred to a 1.5 mL Eppendorf. 250 µL Buffer P2

added, gently mixed via inversion, and incubated at RT for 2 minutes. 250 µL of Buffer N3 then added, gently mixed via inversion, and centrifuged at 13,000 rpm, RT for 10 minutes. Supernatant transferred to QIAprep 2.0 Spin Column, centrifuged at 13,000 rpm, RT for 1 minute and flow-through discarded. 500 µL PB buffer then added to column, centrifuged at 13,000 rpm, RT for 10-30 seconds and flow-through discarded. 750 µL PE buffer added to column, centrifuged at 13,000 rpm, RT for 10-30 seconds and flow-through discarded. Column centrifuged at 13,000 rpm, RT for 3 min and transferred to a sterile 1.5 mL Eppendorf. 50 µL sterile ddH_2O added to column, left for 1 min at RT then centrifuged at 13,000 rpm, RT for 1 minute.

2.3.10 Midiprep DNA purification:

O/N cultures centrifuged at 4000 rpm, 4°C for 15 minutes, and supernatant discarded. Pellet resuspended in 4 mL Buffer P1, 4 mL Buffer P2 added, mixed by inverting and incubated at room temperature for 3 minutes. QIAfilter Cartridge placed in a suitable tube leaving space for buffer BB addition. 4 mL S3 added to lysate and mixed by inverting 4-6 times. Lysate transferred to QIAfilter Cartridge and incubated at room temperature for 10 minutes. Lysate filtered into falcon, 2 mL Buffer BB added and mixed by inverting 4-6 times. Lysate transferred to QIAGEN Plasmid Plus spin column and vacuum applied. Once lysate passed through column, 0.7 mL Buffer ETR added, and vacuum applied. 0.7 mL Buffer PE added, vacuum applied, and column centrifuged at 10,000 g for 1 minute. Column transferred to sterile 1.5 mL Eppendorf, 200 µL sterile ddH_2O added to column, left to stand for 1 minute and centrifuged for 1 minute.

2.3.11 Standard restriction enzyme digest

A standard RE digest contained 30% plasmid/DNA, 1X reaction buffer, 0.5 units of each enzyme and the required volume of ddH_2O to fill reaction volume to 100%. Reaction then incubated for 1 hour at 37°C unless enzymes require different reaction conditions.

2.3.12 Golden gate ligation

Golden gate ligation reactions for golden gate mutagenesis of Pr-deGFP-Mgapt to Pr-deCFP-Mgapt were set up following the component mixture and thermocycler settings shown in tables below. Resulting plasmids were then digested with 0.5 µL DpnI and incubated at 37°C for 75-90 minutes.

Golden Gate ligation reaction mixture:

Component	Final Concentration
10X T4 DNA Ligase buffer	1X
10X BSA	1X
BsaI (10,000 units/mL)	0.67 units
T4 DNA ligase (1-3 units)	0.067 - 0.2 units
PCR product	208 ng
Sterile ddH_2O	Make up to 15 mL

Golden Gate ligation reaction thermocycler protocol:

Stage	Temperature (°C)	Time (minutes)	No. cycles
1	37	5	25
	16	10	
2	50	5	1
3	80	5	1
4	8	Infinite	1

2.3.13 Heat shock transformation

Competent cells were defrosted on ice, then 20 µL cells were mixed with variable amounts of transformation and incubated on ice for 20 minutes. Mixture subsequently incubated at 42°C for 40 seconds, then incubated on ice for 2 minutes, mixed with 100 µL LB or and recovered at 37°C for 1 hr. 100 µL of mixture was then spread on an agar plate, with relevant antibiotic (if applicable), and incubated at 37°C overnight.

2.3.14 Electrocompetent cell production

Exact protocol used for transformations presented in figure 6 was unrecorded, however was either:

A culture of BL21 (DE3) gold, ampicillin was grown at 28°C, shaking (unknown rpm) overnight and subcultured 1:100 into LB or M9, ampicillin and 20 mM L-arabinose (total volume 50 mL) in the morning. Culture then incubated at 28°C, shaking (unknown rpm) until an OD 0.4-0.6 was reached, transferred to chilled 50 mL falcon, and spun at 3000 g, 4°C for 10 minutes. Supernatant was then discarded and pellet resuspended in 5 mL cold 10% (v/v) glycerol. Centrifugation and resuspension steps were repeated a total of three times before being resuspended in a final volume of 200 µL cold 10% (v/v) glycerol. Final solution was then aliquoted into sterile Eppendorf's and stored at -80°C.

Or

LB overnight culture of BL21 Gold (DE3) pKD46, ampicillin was grown at 28°C, shaking (unknown rpm). Overnight culture was then diluted 1:100 into fresh LB or M9, ampicillin and 20 mM L-arabinose in a baffled flask and incubated at 28°C, shaking (unknown rpm) until OD 0.4 – 5 reached. No more than 40 mL culture transferred to a 50 mL falcon and chilled on ice for 30 minutes. Culture centrifuged at 4000 rpm for 15 min at 4°C, supernatant discarded and

resuspended in equivalent volume of pre-chilled sterile ddH_2O. Sample then centrifuged at 4000 rpm for 15 min at 4°C, supernatant discarded and re-suspended in just enough volume (e.g. 2 mL) of pre-chilled 20% glycerol. If multiple falcon tubes were used, samples were pooled together. Previous centrifugation followed, supernatant discarded and pellet resuspended in around 3 mL pre-chilled 20% glycerol. Cells were then transferred to pre-chilled Eppendorf's and stored at -80°C.

2.3.15 Electro transformation

50 µL electrocompetent BL21 (DE3) Gold were transferred to a pre chilled 1 mm gap electroporation cuvette, on ice. Variable amounts of lambda red PCR fragment added to cuvette and gently pipette mixed. Cuvette then pulsed at 1.8 kV, 200 Ω resistance, and 25 µF capacitance for 5 ms (not 100% certain on resistance and capacitance settings). 950 µL RT outgrowth medium then immediately added to cuvette, sample transferred to an appropriate sterile tube and incubated at 37°C, shaking for 60 minutes. 100 µL cells spread on pre-warmed (in 37°C incubator) kanamycin plates and incubated overnight at 37°C.

2.4 Cell-free

2.4.1 Cell free extract production

Buffers and solutions:

3 main solutions were required for cell-free extract production: S30A, S30B and a sugar solution. 1.33 L of **S30A** was produced containing 14 mM Mg-glutamate, 60 mM K-glutamate, 50 mM Tris, dissolved in ddH_2O water and pH adjusted to 7.7 with acetic acid. 2 L **S30A** (14 mM Mg-glutamate, 60 mM K-glutamate, in ddH_2O and pH adjusted to 8.2 with 2 M Tris) and

a 32 mL **sugar solution** (15% V/V glycerol, 100 g / D-lactose, 12.5 g / D-Glucose in ddH_2O and filter sterilised).

Cell harvest:

400 mL autoinduction media + 13 sugar solution was inoculated with BL21 Gold (DE3) and incubated overnight at 30°C, 180 rpm. In the morning, the culture was transferred evenly into 2 x 1 L centrifuge bottles and centrifuged at 3010 g for 19:30 minutes. 2.66 mL 1 M DTT was then added to the S30A buffer and kept on ice. Supernatant discarded and pellets resuspended in 200 mL S30A. Samples were then centrifuged for 19:30 min at 3010 g, supernatant discarded and pellets resuspended in 200 mL S30A. Samples further centrifuged for 19:30 min at 3010 g, supernatant removed and pellets resuspended in 40 mL S30A. Samples then transferred to pre-chilled 50 mL falcon tubes and centrifuged at 4000 rpm for 10 minutes. Supernatant removed and pellets centrifuged at 4000 rpm for 2 minutes. The weight of the pellets then recorded and pellets resuspended in volume of S30A equal to 1.1 mL S30A per gram of pellet. Samples centrifuged for 10 seconds at 200 rpm, frozen with dry ice and stored at -80°C.

Sonication:

Pellets thawed on ice and sonicated with 154 joules for around 30 seconds per wet cell mass (until samples went dark brown with reduced viscosity, around 6:30 – 7 min for 10 mL sample). Samples transferred to 2 mL Eppendorf's and centrifuged for 10 minutes at 12,000 g, 4°C. Supernatants then transferred to clean 2 mL Eppendorf's and incubated at 37°C for 80 minutes. The samples were then centrifuged for 10 minutes, 12,000 g, 4°C. Sample supernatants of same extract then pooled into falcon tubes. Dialysis tubing pre-equilibrated in 1 L S30B + 1 mM DTT for 2 minutes. Half of each pooled sample was then stored on ice whilst other was

transferred to dialysis tubing in separate 1 L S30B + 1 mM DTT solutions and incubated at 4°C, stirring for 1 hr and 10 minutes. Samples from dialysis tubing then transferred to clean 2 mL Eppendorf's and both dialysed and un-dialysed samples centrifuged at 12,000 g for 10 minutes, 4°C. Respective samples pooled together into clean falcons, mixed by inversion and centrifuged for 20 seconds, 1000 g, 4°C. Samples then aliquoted into sterile Eppendorf's on dry ice and stored at -80°C.

2.4.2 Cell free reactions

Standard single reaction setup:

Cell-free reaction mixtures (33.3% E. coli extract, 6.23 mM Mg-Glutamate, 175 mM K-Glutamate, 1.35 mM NTPs, 0.59 M HEPES pH 8, 1.44% ddH_2O, 2.36 mg/mL tRNA, 3.07 mM CoA, 3.95 mM NAD, 0.8 mM folinic acid, 11.46 mM spermidine, 0.37 M 3-PGA, 1.34 mM amino acids (or 0.96 mM every amino acid apart from leucine which was 0.8mM) and 3.04% PEG-8000 to a total volume of 24.75 µL) were produced and supplemented with the desired concentration of plasmid and topped up to 33 µL with sterile ddH_2O.

19 amino acid mix

35 mL ddH_2O was incubated at 42°C, stirring. 5.2 mM L-leucine and 6.24 mM of each other natural amino acid (apart from tryptophan) were then added one at a time. 840 µL 4 M KOH was added and solution adjusted with ddH_2O to a final volume of 48.04 mL.

2.5 Sequencing

2.5.1 Genewiz:

Samples prepared for sequencing by genewiz, by mixing 5 µL 22.9 ng/ µL pPr-deCFP-Mgapt with 5 µL 5 µM VF2 or VR.

2.6 Bioinformatics

2.6.1 Secondary structure prediction:

Secondary structure prediction was carried out using the MaxExpect web server produced by the Mathews group at University of Rochester Medical Center[81] [82].

2.6.2 DNA to mRNA conversion:

To covert DNA sequences to mRNA the online Transcription and Translation Tool was used [83].

2.6.3 Identification of protein homologs:

NCBI BlastP was used for the identification of protein homologs [84].

2.6.4 Multiple sequence alignments (MSAs):

MSAs of mRNA intergenic regions were caried out using EMBL-EBI Clustal Omega tool [85] and visualised in Jalview [86] and pPr-deCFP-Mgapt sequencing MSA used multalign [87].

2.6.5 Genome sequences:

Genome sequences were provided by NCBI [88].

3 Identification, cloning and cell-free expression of lasso peptide BGCs

3.1 Introduction

Recently there has been an increased demand for an expanded repertoire of antimicrobials due to an increased occurrence of antimicrobial resistance (AMR). One way to identify new antimicrobials is via the bioinformatic analysis of novel BGCs within genomes of sequenced organisms (commonly known as genome mining). These clusters can then be activated by techniques such as heterologous expression to allow expression of the encoded natural product. This method is hit and miss in its ability to acquire antimicrobials, however, additional biological activities such as anti-HIV can be exploited [51].

In this study, using the bioinformatic program antiSMASH, two *Streptomyces* genomes (*Streptomyces venezuelae* and *Streptomyces rimosus*) were searched for the presence of natural product BGCs, of which three lasso peptide clusters were chosen for investigation. The chosen clusters, albusnodin, lagmysin and moomysin homologs, were subsequently copied via PCR, cloned into plasmids, and expressed in cell-free lysates using bovicin as a positive control. Lysates were then tested for antimicrobial activity.

Cell-free protein synthesis was the heterologous platform of choice due to having a tolerance to toxic products [89] that may potentially allow for higher yields of active natural products to be produced compared to live cell-based platforms. Furthermore, it has been previously demonstrated that CFPS can express the lanthipeptide (RiPP) Nisin [61] suggesting lasso peptide production via this method could also be achieved. In addition, more recently Yuanyuan Si, et al. 2020 produced four known lasso peptides burhizin, fusilassin, capistruin and cellulassin via a cell-free platform [71] further supporting the feasibility of the method.

3.2 Results

3.2.1 Identification of lasso peptides:

antiSMASH analysis of *Streptomyces venezuelae* ATCC 10712 (NZ_CP029197.1) and *Streptomyces rimosus* ATCC 10970 (NZ_CP023688.1) genomes revealed a variety of different natural product BGCs within both genomes, such as terpene, non-ribosomal peptide synthetase, type 1 polyketide synthase, lanthipeptides and lasso peptides clusters. From these, three lasso peptide clusters were chosen for further investigation – albusnodin, lagmysin and moomysin cluster homologs. The percentage of genes within each cluster that have a significant BLAST hit to their closest known cluster (similarity) varied between homologs, with values of 100, 80 and 50% respectively.

Organism	**Cluster name**	**Most similar cluster**	**Encoding organism of most similar cluster**	**Similarity (%)**	**Peptide Percentage Identity**
Streptomyces venezuelae	albusnodin Homolog	Albusnodin	(*Streptomyces albus*)	100	55.56
Streptomyces rimosus	lagmysin homolog	Lagmysin	(*Streptomyces sp. NRRL S-118*)	80	29.11
	moomysin homolog	Moomysin	(Streptomyces cattleya NRRL 8057)	50	27.66

Table 5 – Lasso peptide clusters selected for investigation. Table showing the organisms containing the clusters of interest, the cluster names, their most similar clusters, cluster similarity and peptide percentage identity.

Furthermore, multiple sequence alignments (MSA) were performed for peptides of interest and closest known cluster peptides. From this, the percentage identities for the peptide homologs

compared to original clusters, within the MSA frame, were calculated using the equation: (number of identical bases / length of longest sequence) x 100. The albusnodin homolog precursor peptide had highest homology to their closest related cluster peptide with a percentage identity of 55.56**.** Whereas, both lagmysin homolog and moomysin homolog peptides vary more from their most similar cluster peptides, with percentage identity values of 29.11 and 27.66, respectively.

3.2.2 PCR of selected lasso peptide clusters:

Once the clusters had been identified by antiSMASH, primers were designed to amplify each clusters peptide and PTMs separately to allow for cloning into individual plasmids. Furthermore, additional primers were designed to incorporate an N-terminal hexahistidine (His_6) tag to the cluster peptides, allowing for purification by immobilized metal ion affinity chromatography.

The chosen clusters were amplified from the genomic DNA (gDNA) of the producing strain using the primers listed in Table 4. All the cluster peptides were obtained using the general PCR protocol and thermocycler settings, with an annealing temperature of 72°C and an elongation time of 30 seconds and the albusnodin homolog PTMs with a 70 °C annealing temperature and 2-minute elongation time. However, lagmysin and moomysin homolog PTMs were obtained using touchdown protocols and altered master mix compositions. A touchdown protocol with an annealing temperature of 58°C and time of 10 sec, decreasing by 0.5°C each cycle (30 x cycles), an elongation time of 1 min 30 sec and a master mix containing 1 M betaine and 3% (v/v) DMSO was used to achieve lagmysin homolog PTMs. To achieve moomysin homolog PTMs the annealing temperature was set at 68°C, decreasing by 0.5°C per cycle (25

x cycles), for 10 seconds, elongation time of 1 min 30 sec and a master mix containing 1 M betaine, and no DMSO.

To visualise the PCR products, DNA gel electrophoresis was carried out following the standard protocol using 25 µL PCR fragment mixed with 5 µL 6X loading dye, with all 30 µL being loaded. Figure 6 shows bands within the 75-400 and 3000-4000bp region of the gel, suggesting correct amplification of BGCs precursor peptide and Post-translational modification (PTM) enzyme sequences.

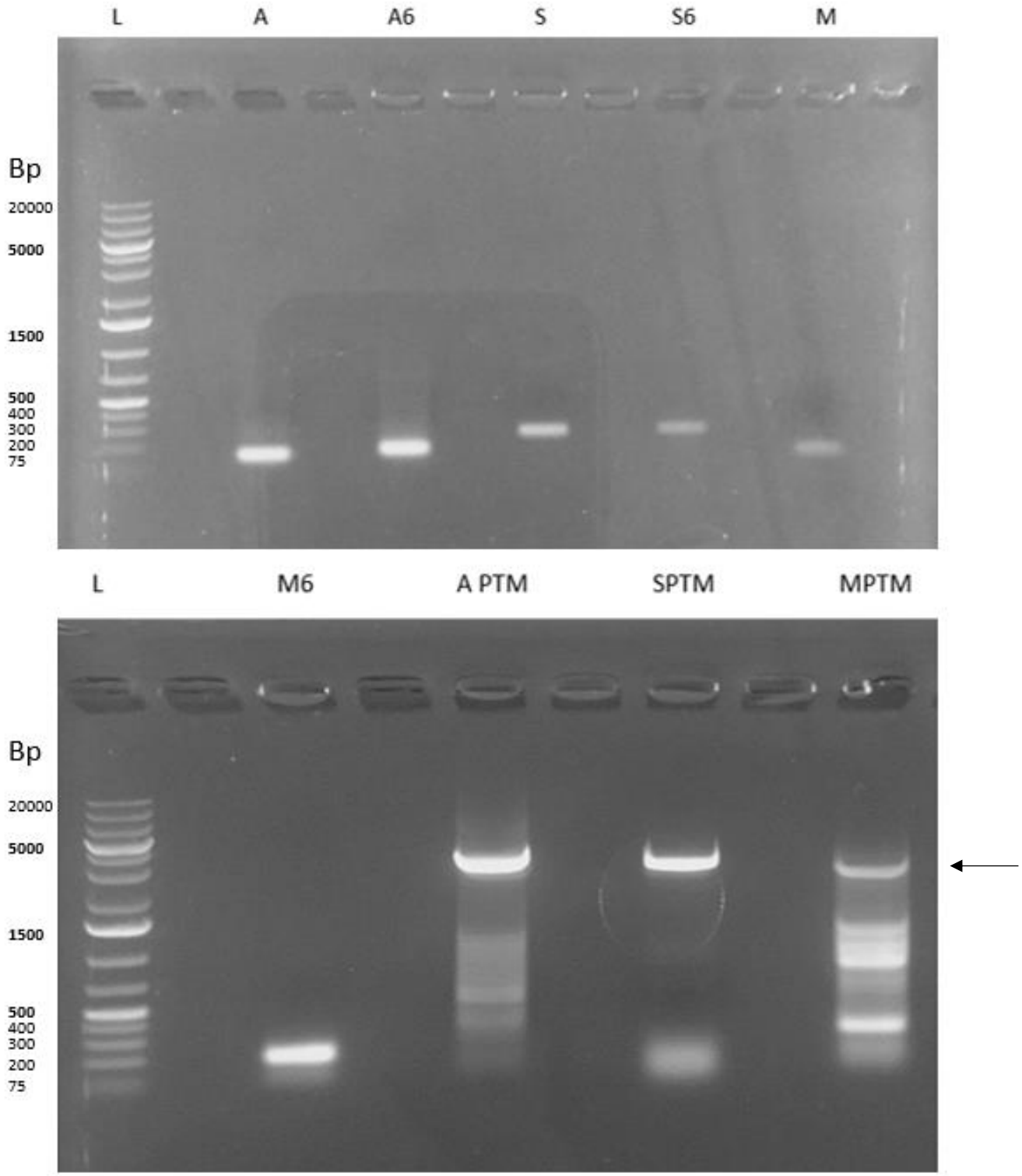


Figure 6 – PCR fragments of cluster precursor peptides, N-terminal His$_6$ peptides and PTMs visualised by DNA gel electrophoresis. PCR fragments produced via above method were analysed by DNA gel electrophoresis. A = albusnodin homolog, L= lagmysin homolog, M= moomysin homolog and 6 = N-terminal his$_6$. Arrow indicates PTM fragments of interest that were later extracted. Bands expected within 147 – 276 for precursor peptides and 2565 – 2952bp for PTM enzyme fragments. GeneRuler 1kbp plus DNA ladder used.

7.2.3 Cloning of fragments into destination vectors:

After the peptides, his_6 peptides and their respective PTMs had been amplified by PCR the resulting fragments were isolated by gel extraction and inserted into their destination plasmids, pTU1-A-T7RBS_OxyJ_His6_B0015 and pSF1C-A-SP44a_pETRBS_AcfP-His6_B0015, using restriction enzyme cloning.

500 ng pTU1-A-T7RBS_OxyJ_His6_B0015 and pSF1C-A-SP44a_pETRBS_AcfP-His6_B0015 were digested overnight at 37°C with NdeI and BamHI and NdeI and PacI with the addition of 0.5 units Shrimp alkaline phosphatase (rSAP). The complete gel extraction elution (around 28 µL) of each peptide and PTM fragment were then mixed with 3.5 µL reaction buffer and digest overnight at 37°C with 1.25 units NdeI and 0.5 units BamHI and 1.25 units NdeI and 0.5 units PacI, respectively. Digested peptide and His_6 peptide fragments were then ligated with pTU1-A-T7RBS_OxyJ_His6_B0015 following the standard protocol, whereas ligations of pSF1C-A-SP44a_pETRBS_AcfP-His6_B0015 and PTM fragments were carried out at room temperature over the weekend using a 1:3.16 ratio of vector to insert. 2 µL of each ligation mixture was then transformed into DH10β, recovered in SOC and incubated at either 37°C overnight on carbenicillin for peptide ligations or 30°C, on apramycin for 2 days for PTM ligations. Overnight cultures of transformants were produced following the standard protocol with transformants of peptide ligations containing carbenicillin and incubated at 30°C, 180 rpm and transformants of PTM ligations containing apramycin and incubated at 30°C or 37°C, 160 or 180 rpm, respectively. Resulting plasmid DNA, table 2, was mini prep purified and digested for 30 minutes following the standard protocol using NdeI and BamHI for peptide containing plasmids and 1 unit XbaI for PTM plasmids and visualised using gel electrophoresis. Plasmids then sent for sequencing using respective primers by Eurofins.

3.2.4 Cell-free expression of RiPP plasmids:

After the PCR fragments of both the peptides and PTMs had been incorporated into their corresponding vectors, they were midiprep purified (using 50 mL LB24 overnights) to produce stock concentrations for dilution to 10nM final concentrations in *E. coli* cell-free protein synthesis reactions.

To assess the ability of the plasmids to express precursor peptides or PTM enzymes cell-free reactions were initiated using the standard protocol using either BL21 (no dialysis) or Rosetta extracts and 10 nM of each plasmid, apart from those indicated via volumes, (44 reactions in total – supplementary table 2). Reactions were produced to express each peptide, His_6 peptide and PTM enzymes separately (9 reactions) and each peptide and his tagged peptide in combination with their respective PTMs (6 reactions), to allow for peptide modification. Furthermore, a positive control cluster (bovicin – with known antimicrobial activity) was also expressed following the previously described format; peptide and PTM separately and then in combination (3 reactions). In addition, 3 positive control reactions were set up: PC 1 (8.25 µL T7 eGFP), PC 2 (8.25 µL T7 sfGFP), PC 3 (10 nM SP44a mVenus) and a NC reaction containing 8.25 µL sterile ddH_2O, to assess the ability of the extracts to promote expression from T7 and sp44 promoters, as positive control plasmids will produce measurable levels of fluorescence. BL21 No dialysis extract containing reactions were incubated at 28°C, 160 rpm in 2 mL Eppendorf's overnight and end point readings of controls were taken after 16hrs of incubation at 28°C, 160 rpm by aliquoting, in triplicate, 10 µL into a 384 well plate and measuring fluorescence in a BMG Omega plate reader using 485-12 excitation and 520 nm emission filters. Whereas Rosetta extract containing reactions were loaded onto a 384 well plate in 10 µL aliquots and incubated at 30°C overnight with intermittent shaking in a BMG Omega plate reader, with fluorescent readings (485-12 excitation, 520nm emission) taken every 10 minutes for 8 hours.

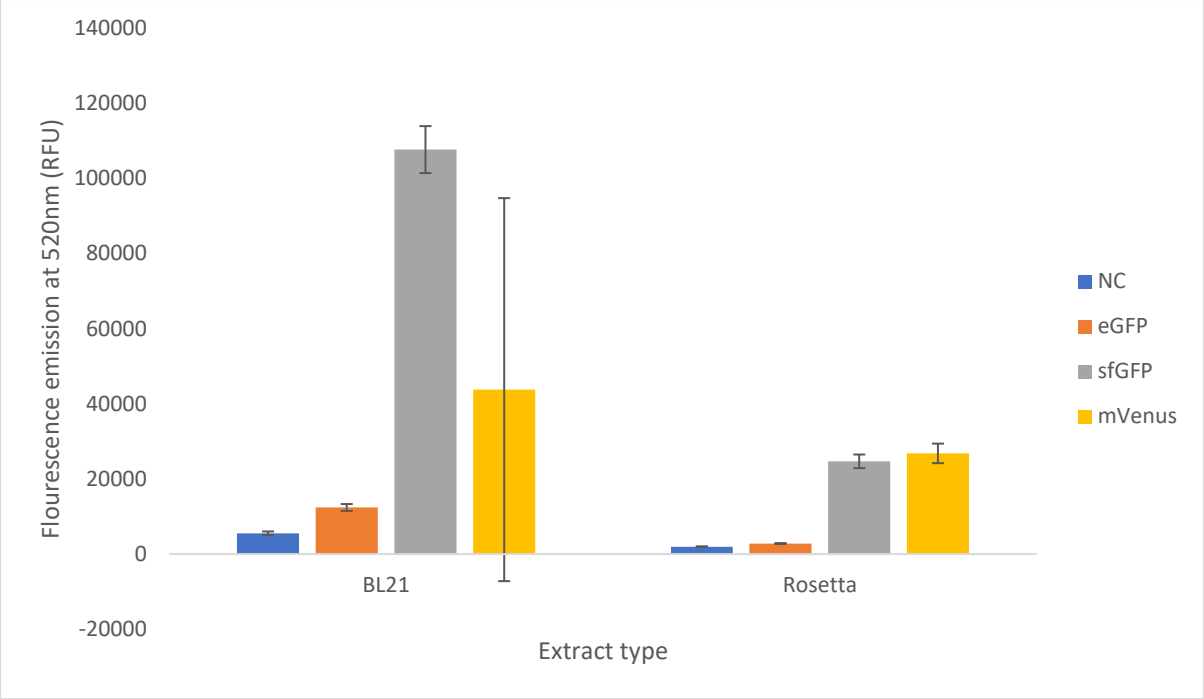


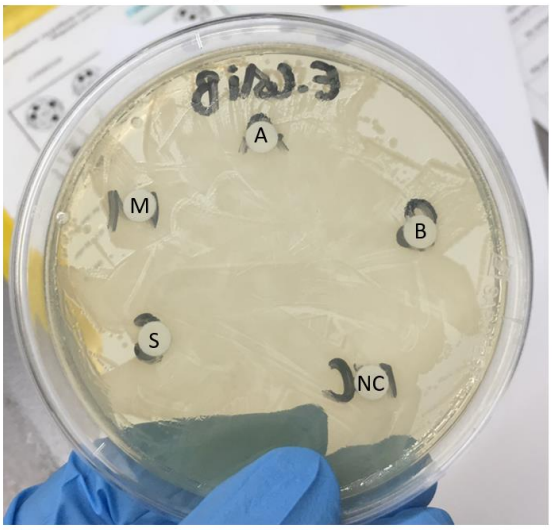
Figure 7 – BL21 no dialysis and Rosetta RiPP Cell-free controls end point fluorescence intensities. Average end point fluorescence intensity values (excitation 485-12 nm, emission 520nm) for the controls of the RiPP cell free assays recorded using a BMG omega plate reader. BL21 values were recorded after 16 hours of incubation at 28°C, 160 rpm and Rosettes taken after 8 hours of incubation in plate reader at 30°C, with intermittent shaking.

Figure 7 showed all BL21 PCs had higher levels of fluorescence than the NC, with T7 sfGFP producing the strongest signal followed by mVenus and eGFP respectively. Furthermore, p-values for 2 tailed, 2 sample T-tests, assuming unequal variance, for RFUs of NC compared to the PCs were calculated, yielding: NC:eGFP = 0.0011, NC:sfGFP = 0.0011 and NC:mVenus = 0.32. Furthermore, all Rosetta extracts PCs reached higher RFUs compared to NC, with mVenus being the highest, followed by sfGFP and eGFP respectively. Furthermore, P-values for 2 tailed, 2 sample T-tests, assuming unequal variance, for RFUs of NC compared to the PCs and sfGFP vs mVenus were calculated, generating: NC:eGFP = 0.0062, NC:sfGFP = 0.0021 and NC:mVenus = 0.0036.

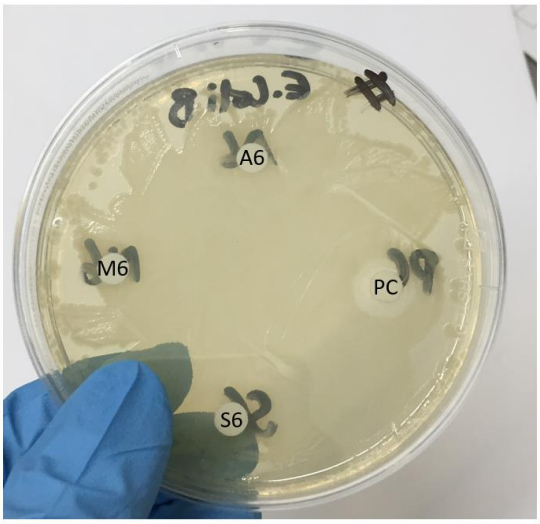
3.2.5 Bioactivity Assays:

To test if any of the modified peptides or his-tagged peptides were antimicrobial, bioassays against *E. coli* MG1655 and *B. subtilis* were carried out. LB agar plates were spread with 50 µL LB suspended *E. coli* MG1655 or *B. subtilis* and 4 or 5 disks were added to each plate. 5 µL aliquots of BL21 or Rosetta CF extracts containing peptide/His_6-tagged peptide and PTM expressed simultaneously were aliquoted on separate disks. Furthermore, for negative controls 5 µL of NC extracts from BL21 and Rosetta RiPP CF experiments were aliquoted on separate disks. Moreover, a PC disk of 1 mg/mL carbenicillin was produced. Plates were then incubated at 37°C overnight.

E. coli MG1655, BL21

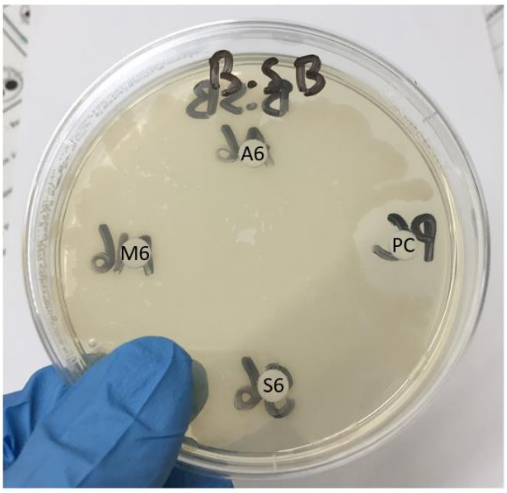


E. coli MG1655, BL21

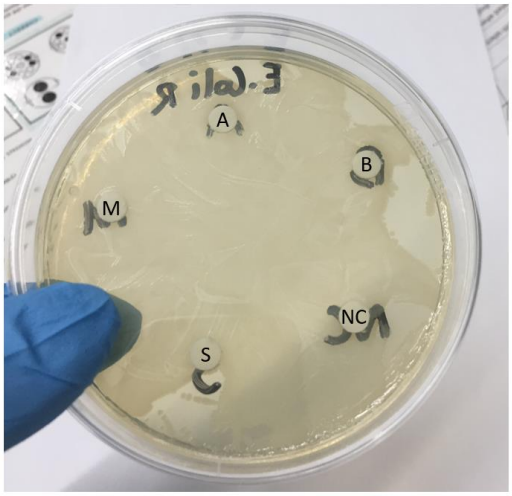


B. subtilis, BL21

B. subtilis, BL21



E. coli MG1655, Rosetta



E. coli MG1655, Rosetta

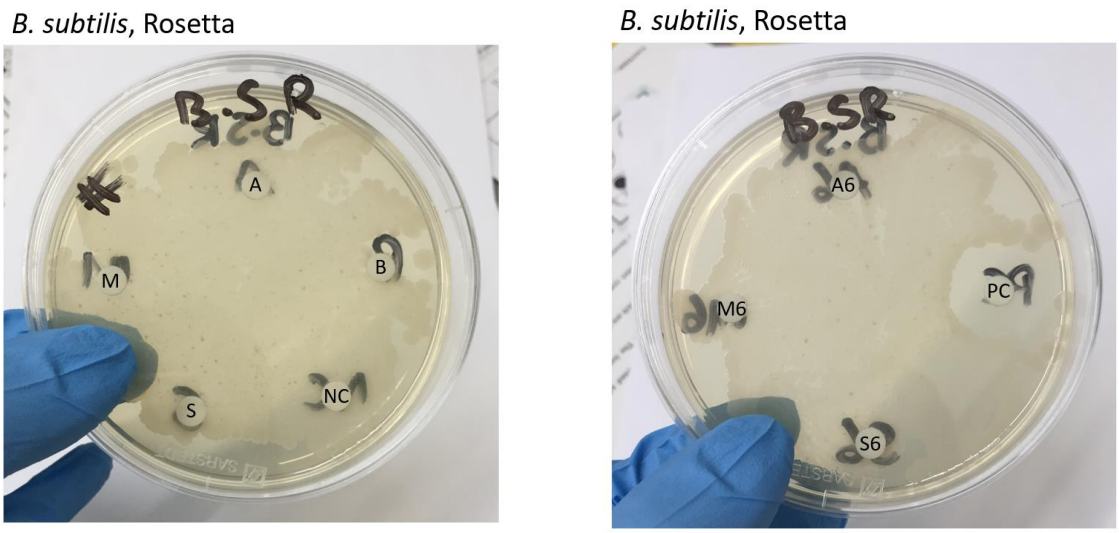


Figure 8 – CF RiPP bioassays. LB agar plates were spread with either 50 mL *E. coli* MG1655 or *B. subtilis*. 5 µL aliquots of either BL21, Rosetta CF RiPP extract containing putative modified peptide/his-tagged peptide, NC, or 1 mg/ml carbenicillin (PC) were added to disks placed on the surface. Plates were then incubated overnight at 37°C. 6 = His-tag

Figure 8 shows only the PC assays produced zones of inhibition.

3.3 Discussion

3.3.1 antiSMASH identification of natural products of interest:

antiSMASH analysis of *Streptomyces venezuelae* ATCC 10712 and *Streptomyces rimosus* ATCC 10970 found a variety of natural product BGCs with varying levels of homology to their respective closest known clusters. However, from the generated outputs only two of the clusters

identified in *Streptomyces rimosus* encoded lassopeptides (lagmysin and moomysin homologs) and one in *Streptomyces venezuelae* (albusnodin homolog), all of which were chosen for further analysis.

From the literature no antimicrobial activity has been shown for both moomysin and lagmysin against an array of bacteria [44], whereas the antimicrobial activity of albusnodin has not been tested. However, the albusnodin, lagmysin and moomysin homolog clusters differentiate in terms of sequence compared to their original clusters, for example the cluster peptides only had 55.56, 29.11, and 27.66 % similarity, respectively, to their original clusters. Therefore, this variance in sequence may allow them to exhibit different antimicrobial activities. Furthermore, due to the antimicrobial activity of the original albusnodin peptide being unassessed, the homolog of interest may present antimicrobial activity. Moreover, even if the peptides do not exhibit antimicrobial activity, they may be involved in other processes such as quorum sensing or have interesting properties such as increased stability which may have future applications or anti-HIV properties. This therefore made them interesting for further analysis. However, due to time constraints we were limited to a single experiment to explore their biological activities. Therefore, we decided the most immediate experiment was to test their antimicrobial activity against strains we had available.

To improve the selection method in future studies deeper searches on a wider array of *streptomyces* species should be carried out to find a larger variety of clusters. This would increase the likelihood of finding clusters with homology to known antimicrobial clusters, which would therefore increase the chance of finding a product with antimicrobial activity.

In addition, further studies could benefit from the use of the program Antibiotic Resistance Target Seeker (ARTS) which uses the presence of resistance genes to identify BGCs potentially encoding antibiotics. This is because bacteria often contain resistance mechanisms against the

antibiotics they produce to protect themselves [90] . This would limit time spent on clusters that are less likely to yield bioactive products and therefore increase the yield of products with medicinal application. Furthermore, future studies could involve studying the motifs in the active lasso peptide sequence that are involved in their antimicrobial activity and then using tools such as NCBI BLAST to search for these motifs as a way of identifying potential antimicrobial peptides. Furthermore, this could be used in combination with antiSMASH and related programs, such as NeuRiPP and BAGEL, as a way of screening the clusters identified for potential antimicrobial peptides.

Together, these additional steps when deciding on which clusters to study could increase the likelihood of finding products with medicinal application. However, I also feel that studying clusters which are not like clusters with known antimicrobial activity is important due to the opportunity of finding more novel products and those with alternative useful properties. For example, lasso peptides often have high resistance to proteolytic degradation due to their knotted structure and sometimes have elevated heat resistance [91]. This shows the potential for finding other natural products with properties with applications within other fields, such as biomanufacturing.

3.3.2 Sequencing results:

The sequencing results from Eurofins showed that the pTU1-A-T7RBS_Albusnodin_H, pTU1-A-T7RBS_Albusnodin_H_His6, pTU1-A-T7RBS_Lagmysin_H, pTU1-A-T7RBS_Lagmysin_H_His6, pTU1-A-T7RBS_Moomysin_H, pTU1-A-T7RBS_Moomysin_H_His6 and pSF1C-A-SP44a_pETRBS_Albusnodin_H_PTMs contained the correct sequences. However, pSF1C-A-SP44a_pETRBS_Lagmysin_H_PTMS sequence showed a one base pair mutation in one of the genes and the

SP44a_pETRBS_Moomysin_H_PTMS sequencing was poor and needs repeating. Furthermore, there was also an ABC transporter ATP binding protein in the lagmysin homolog cluster, which was missed in the cloning process, which may be required for expression in future studies.

3.3.3 Rosetta codon optimisation:

Rosetta based cell-free extracts were used in this study due to containing tRNAs for codons that are more rarely used in *E. coli*, as it was thought it may aid with expressing clusters that originated from *Streptomyces* species. For example, when codon usage was compared between *E. coli* K12 and *Streptomyces albus* (provided by kazusa codon usage database [92]), it was observed that the CCC and AGG rare codon were used more in albus than K-12, suggesting that it may be useful to use Rosetta to allow for higher expression of these codons. Furthermore, to overcome this problem DNA fragments of respective genes with rare codons replaced by more commonly used codons in *Streptomyces* species could have been produced by gene synthesis.

3.3.4 Cell-free fluorescent protein expression from T7 and SP44a promoters:

From the 2 tailed, 2 sample T-tests, assuming unequal variance, comparing RFU values for both BL21 and Rosetta NC vs PC reactions the p values calculated showed that all reactions (apart from BL21 mVenus) RFU values are significantly different to their respective NCs. Thus, demonstrating that both extracts were able to induce expression of the T7 promoters, but only the SP44a promoter in the Rosetta extract. However, the statistically insignificant value observed for BL21 mVenus compared to its respective negative control is suspected to be due to an experimental error/sample drying out reducing the volumes added to the plate for analysis.

Therefore, this reaction should be repeated. Together. this suggests that the peptide/His_6-tagged peptides and PTMs should be expressed in their respective reactions, however BL21 mVenus should be repeated to confirm this.

One limitation to this experiment was the unknown concentrations of T7 sfGFP and SP44a mVenus used in the cell-free reaction controls. Therefore, they do not directly indicate that 10 nM of precursor peptides/PTM enzyme plasmids are sufficient to induce significant levels of protein. Therefore, the experiment should be repeated using 10 nM of control plasmids. Moreover, to further increase reproducibility the experiment should be repeated 2 additional times to obtain a biological triplicate.

3.3.5 BL21 vs Rosetta extract T7 and SP44a expression:

Furthermore, from figure 7 the end point RFU value for the Rosetta extract sfGFP reaction was only 22.9% of value for the BL21 extract sfGFP, suggesting the BL21 extract may allow for higher levels of expression of the peptides/his_6 peptides. However, due to the insignificant result obtained by the BL21 mVenus result, comparison of the extracts ability to induce expression from SP44a promoters cannot be reliably carried out.

Furthermore, the difference in fluorescence could be attributed to the different strength of spectra of sfGFP and mVenus and the wavelength recorded at. Therefore, to more accurately assess the ability of each extract to induce expression from each promoter, the same fluorescence protein should be put under control of each promoter and the same concentration of each plasmid should be used.

3.3.6 Bioassays:

The bioassays presented in figure 8 only showed a zone of inhibition for the 1 mg/mL carbenicillin control. In addition, no zone of inhibition was produced for the bovicin positive control – which has been shown to exhibit antimicrobial activity when produced in a cell-free reaction (personal correspondence with Dr Simon Moore and Dr Emzo De Los Santos, University of Warwick). The exact strains used for the antimicrobial assay of bovicin produced via cell-free is unknown, however, it has been previously shown to inhibit *Bacillus subtilis* ATCC 6537 [93] . Therefore, we would have expected to see antimicrobial activity on the *B. subtilis* plates used here if produced and activated correctly. This therefore suggests that the CF reactions were unable to synthesise required amounts of precursor peptide / PTM enzymes or allow the PTM enzymes to modify and activate the bovicin precursor peptide to allow for it to exhibit antimicrobial activity. However, both positive control plasmids (peptide and PTM) were under the control of T7 promoters which differs from the T7 peptide, SP44a PTM set up used in this study, thus making the control less comparable to test plasmids.

To investigate whether the problem is absent/low levels of protein synthesis tricine and glycine SDS-PAGE should be performed on each sample to see if the peptides and PTMs, respectively, have been synthesised. This could also be further quantified via anti-His western blotting for the precursor peptides or mass spectrometry. If the problem does not appear to be due to protein synthesis levels nuclear magnetic resonance (NMR) and mass spectrometry should be utilised to assess if the peptides have been modified to form their active state.

Furthermore, if the problem appears to be the inability of the PTMs to become active and modify the peptides, it may suggest they require additional co-factors such as Mg^{2+}. For example, the McjC enzyme required to cyclise microcin J25 is ATP/Mg^{2+} dependent [94]. However, information on the exact components of the cell-free reactions used to produce active bovicin is required to properly utilise it as a positive control for this.

If the problem is suggested to be due to too low levels of protein synthesis, it could suggest the *E. coli* extracts are not active enough – which could be overcome by repeating and optimising the extract production protocol. Furthermore, we intended on testing a *streptomyces venezuelae* cell-free system for the production of the natural products as it was thought that the extract being more closely related to the natural producing organism may allow for higher levels of product to be obtained.

Moreover, optimisation of extract production and reaction condition could be carried out to improve yield and activity of products. For example, the type of media used to grow the cells for cell lysate production or the temperature the reactions were carried out at. This is supported by the fact that the natural cluster for Mccj25 (microcin J 25) can only be expressed in *E. coli* grown in minimal media not LB [95] .

Due to time constraints the above experiments could not be carried out to identify the factors limiting natural product expression and how to overcome them. However, given the time, once the cell-free expression had been optimised, if antimicrobial activity of the chosen clusters was still not observed extracts should be tested on a wider array of organisms to test more broadly their antimicrobial activity. Furthermore, they should be tested on HIV as lasso peptides with anti-HIV properties [96] [51] have been demonstrated previously. This would provide a wider picture of the spectrum of activity.

4 Cell-free ncAA incorporation of tryptophan analogues

4.1 Introduction

One potential method of combating slowed antibiotic discovery is ncAA facilitated peptide diversification, as incorporation of ncAAs into peptide natural products could potentially yield bioactive peptides with increased potency or altered functionality.

A cell-free platform for non-canonical tryptophan incorporation has already been established by Singh-Blom et al. 2014 [73], however, here steps are taken towards producing a similar platform with an alternative method of read out of successful non-canonical incorporation. Firstly, lambda red mutagenesis is attempted to produce a tryptophan auxotrophic strain of *E. coli* capable of allowing efficient analogue incorporation. Furthermore, a cyan fluorescent protein (CFP) containing plasmid is produced via golden gate mutagenesis as a potential method of detecting successful analogue incorporation. Finally, the ability of non-tryptophan depleted BL21 (DE3) gold cell-free extract to allow efficient ncAA incorporation is assessed.

E. coli cell-free lysates were chosen for non-canonical amino acid incorporation due to being able to overcome some limitations of live cell-based approaches. For example, higher resilience against toxic analogues such as 5’fluoro tryptophan and absence of cell wall and membrane for easy supplementation of analogues.

4.2 Results:

4.2.1 Lambda red mutagenesis:

We wanted to produce a tryptophan auxotrophic strain of *E. coli* to enable the production of cell-free lysates depleted of canonical tryptophan. This would then remove competition between canonical and non-canonical tryptophan for Tryptophanyl tRNA synthetase binding, thus allowing for increased ncAA incorporation efficiency. To do this, lambda red mutagenesis

deletion of TrpLEDC genes, which are required for tryptophan biosynthesis, was attempted in BL21(DE3) Gold. Furthermore, an additional deletion of tnaA, which encodes a tryptophanase, was required to prevent degradation of tryptophan analogues.

Firstly, 2 sets of primers were designed to produce DNA fragments containing regions homologous to genes to be removed, flippase recognition target (FRT) sites and a kanamycin cassette. The forward primers (F) contained a sequence homologous to the 5' region flanking and the start of the first gene of clusters to be removed e.g. TrpE (H1) followed by a sequence complementary to the 5' FRT site flanking a kanamycin cassette in the plasmid pKD13 (P1). The reverse primers (R) are composed of a sequence complementary to the 3' FRT (P2) site followed by a sequence complementary to the end and 3' flanking region of last gene of cluster to be deleted (H2).

In addition, 100 µL BL21 (DE3) gold were transformed with 182.5 ng pKD46 (lambda red recombinase expression plasmid) following the standard protocol but recovered in 200 µL SOC at 30°C, and 100 µL plated on Ampicillin plates and incubated overnight at 30°C. The cells were then made electrocompetent to allow for transformation with the PCR fragments required for recombination.

PCR of the ΔTrpLEDC fragment was carried out using the TrpL-FRT_F and TrpC-FRT_R primers, table 4, following the standard protocol using 1 µL pKD13 in a total reaction volume of 50 µL and the standard thermocycler protocol with an annealing temperature of 62°C (decreasing each cycle). Furthermore, the TnaA fragment was produced using 1 µL pKD13 in a 50 µL reaction and the general thermocycler protocol, with an annealing temperature of 67°C and elongation time of 1 minute. ΔTrpLEDC Fragments were then purified following the PCR clean up protocol and electro transformed into BL21 (DE3) pKD46 and plated on kanamycin

agar plates. Putative ΔTrpLEDC Colonies subsequently screened via colony PCR for the presence of the FRT-Kan-FRT cassette.

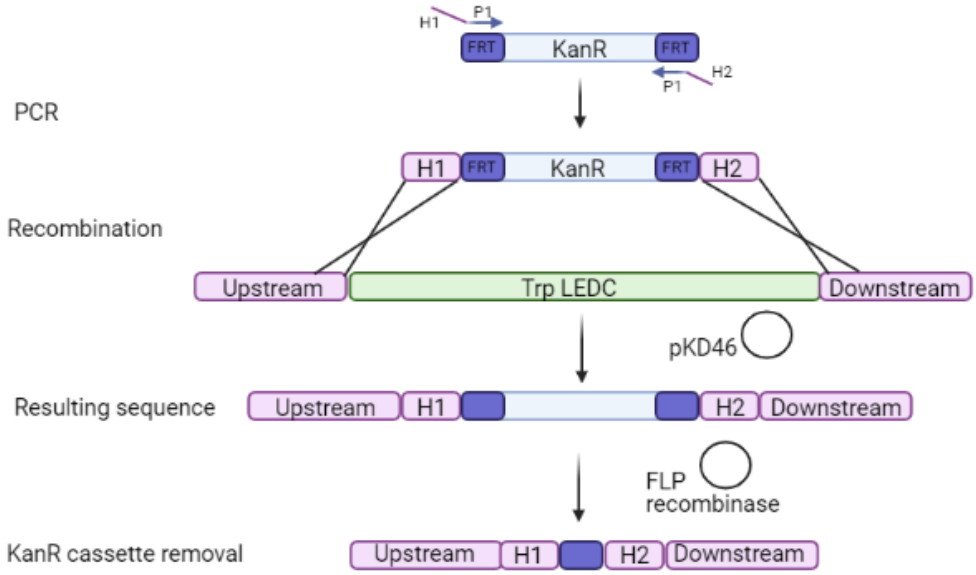
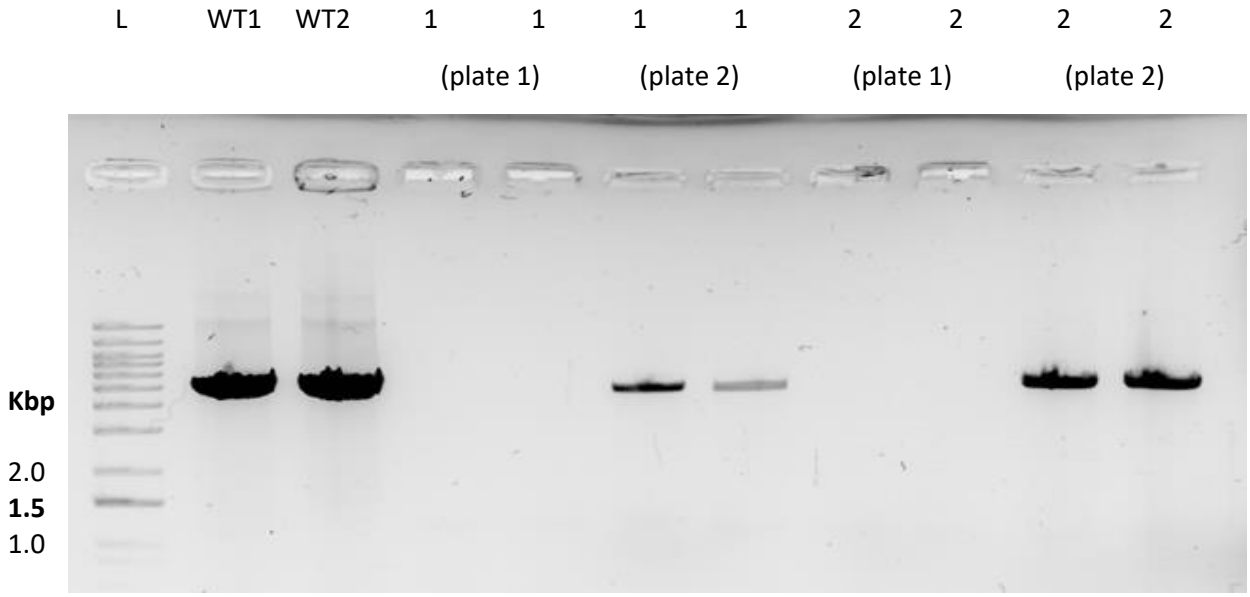


Figure 9 - Schematic representation of the method of lambda red mutagenesis of TrpLEDC attempted in this work.

4.2.2 PCR colony verification of putative BL21 (DE3) Gold ΔTrpLEDC

(a)



(b)

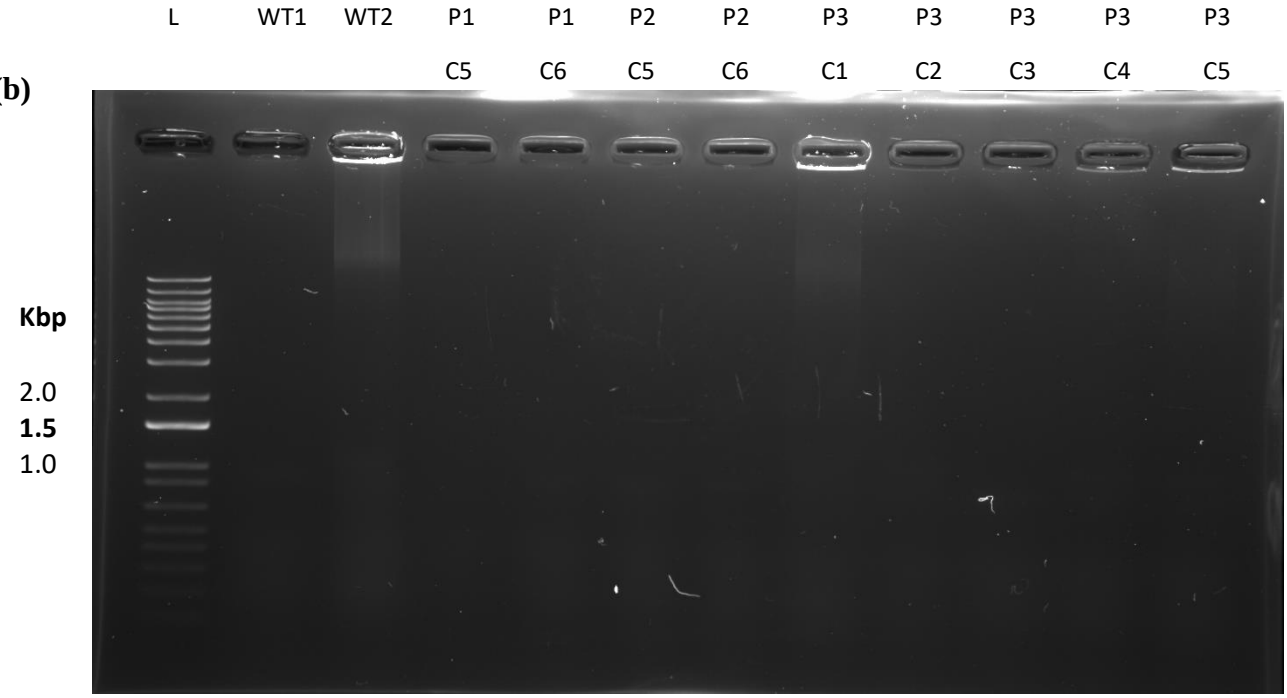


Figure 10 –Putative BL21(DE3) Gold ΔTrpLEDC PCR colony verification polyacrylamide gels. (a) PCR colony verification of colonies from plate 1 and 2 of BL21(DE3) Gold pKD46 transformed with 4033 ng (1) and 2241 ng (2) ΔTrpLEDC PCR fragment, using the Trp_gDNA_F and R primers with the thermocycler protocol being un recorded. **(b)** PCR colony verification of colonies (labelled C) from plate 1, 2 and 3 (transformants incubated at RT for 24 hr before plating) (P1, 2 and 3) of BL21(DE3) Gold transformed with 4033 ng ΔTrpLEDC PCR fragment following the standard thermocycler protocol, with an annealing temperature of 62°C (decreasing each cycle) for 20 seconds, elongation at 72°C for 30 seconds and final elongation at 72°C for 1 minute. Wild type (WT) bands expected at 4.6 Kbp and ΔTrpLEDC at 1.25 Kbp.

Figure 10 (a) shows wild type band patterns from colony PCRs of BL21 (DE3) Gold pKD46 electro transformed with 2241 and 4033 ng ΔTrpLEDC PCR fragments using Trp_gDNA_F and R primers from plates 2 and no bands for plate 1. (b) shows no bands from BL21 (DE3) Gold pKD46 electro transformed with 4033 ng ΔTrpLEDC PCR fragments from plates 1, 2 and 3 (transformants that had been incubated at room temperature for 24 hours before plating) using Trp_gDNA forward and FRT_R primers.

4.2.3 pPr-deCFP-Mgapt biomarker

To enable a direct read out of ncAA incorporation the pPr-deCFP-Mgapt plasmid was produced by golden gate mutagenesis. deCFP differs from deGFP by having a tryptophan at residue 66 (in chromophore) rather than a tyrosine. Therefore, when non-canonical tryptophan has been incorporated, there will be a shift in absorption/emission spectrum.

Firstly, the pPr-deCFP-Mgapt oligonucleotide was produced by PCR following the standard protocol using 1 ng pPr-deGFP-Mgapt and deGFP-deCFP_F and _R primers in a 50 µL reaction and standard thermocycler protocol with and annealing temperature of 68°C (decreasing each cycle) and elongation times of 2 and 3 min in stages 2 and 3, respectively. After verification by gel electrophoresis using the standard protocol with 2 µL PCR product mixed with 2.5X loading dye the PCR product underwent golden gate `ligation and was treated with DpnI to produce pPr-deCFP-Mgapt. JM109 were then transformed with 9.9 ng pPr-deCFP-Mgapt and incubated for 1 or 2 days on an ampicillin plate at 30°C, O/N cultures produced following the standard protocol but using terrific broth and incubated at 30°C, 200 rpm and plasmid miniprepped using double volumes of buffers P1, P2 and N3. Plasmid then digested with EcoRI and XhoI (40% pPr-deCFP-Mgapt, 0.8X reaction buffer, 0.4 units of both

EcoRI and XhoI, 44% sterile ddH_2O and incubated for 60-75 min at 37°C), analysed by gel electrophoresis using 2 µL digested PCR product mixed with 3 µL loading dye, vacuum-dried and resuspended in 32 µL sterile ddH_2O and sequenced by genewiz.

4.2.4 deCFP expression:

Agar plates containing DH10β or JM109 cells expressing pPr-deGFP-Mgapt or pPr-deCFP-Mgapt were imaged on a blue light box, identifying a brighter fluorescence of deCFP compared to deGFP (figure 11).

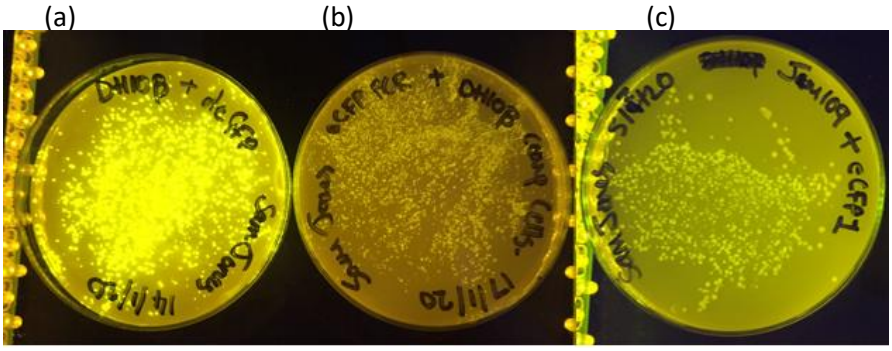


Figure 11 – deGFP vs deCFP fluorescence (a) DH10β pPr-deGFP-Mgapt, (b) DH10β pPr-deCFP-Mgapt and (c) JM109 pPr-deCFP-Mgapt. Dh10β were transformed with 1 µL pPr-deGFP-Mgapt or 2 µL pPr-deCFP-Mgapt respectively for (a) and (b) and JM109 with 9.9 ng pPr-deCFP-Mgapt.

4.2.5 pPr-deCFP-Mgapt sequencing:

The pPr-deCFP-Mgapt purified from JM109 was analysed by sanger sequencing to produce fragment sequences, of both the top and bottom strands, for the region of the plasmid encoding deCFP. These sequences were then aligned with the gene encoding deGFP in pPr-deGFP-Mgapt. This showed guanine bases at positions 599 and 600bp in the coding strand of the

sequenced region of pPr-eCFP-Mgapt, whereas adenine and cytosine at the respective positions in the deGFP gene.

```
            391    400       410       420       430       440       450       460       470       480       490       500       510       520
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
deCFP_VF2;  TTAACTTTAAGAAGGAGATATACCATGGAGCTTTTCACTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACG
 deCFP_VR;  TTAACTTTAAGAAGGAGATATACCATGGAGCTTTTCACTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACG
    deGFP;                          ATGGAGCTTTTCACTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACG
 Consensus  ttaacttttaagaaggagatatacc ATGGAGCTTTTCACTGGCGTTGTTCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACG

            521    530       540       550       560       570       580       590       600       610       620       630       640       650
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
deCFP_VF2;  GCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTGGGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
 deCFP_VR;  GCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTGGGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
    deGFP;  GCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT
 Consensus  GCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTggGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTT

            651    660       670       680       690       700       710       720       730       740       750       760       770       780
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
deCFP_VF2;  CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
 deCFP_VR;  CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
    deGFP;  CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG
 Consensus  CTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG

            781    790       800       810       820       830       840       850       860       870       880       890       900       910
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
deCFP_VF2;  GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACA
 deCFP_VR;  GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACA
    deGFP;  GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACA
 Consensus  GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACA

            911    920       930       940       950       960       970       980       990       1000      1010      1020      1030      1040
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
deCFP_VF2;  ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
 deCFP_VR;  ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
    deGFP;  ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
 Consensus  ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA

            1041   1050      1060      1070      1080      1090      1100      1110      1120      1130      1140      1150      1160      1170
            |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
deCFP_VF2;  GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCTAAACAATAACTGAATAGGGGATCCCGACTGGCGAGAGCCAGGTAACGAATGGATCCCCGAGCTCGAGCAAAGCCCGCCGA
 deCFP_VR;  GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCTAAACAATAACTGAATAGGGGATCCCGACTGGCGAGAGCCAGGTAACGAATGGATCCCCGAGCTCGAGCAAAGCCCGCCGA
    deGFP;  GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCTAA
 Consensus  GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCTAAacaataactgaataggggatcccgactggcgagagccaggtaacgaatggatccccgagctcgagcaaagcccgccga
```

Figure 12 – deCFP sequencing results vs deGFP gene.

Shows the aligned sequencing results from deCFP_VF2 on the top row, then deCFP_VR, the sequence of deGFP gene from pPr-deGFP-Mgapt and the consensus sequence on the bottom. Blue = low consensus, Red = high consensus. TGG codon at 598-600 bp codes for tryptophan producing CFP and TAC at the same positions encodes tyrosine producing GFP.

4.2.6 Characterisation of pPr-deCFP-Mgapt in CF:

To characterise pPr-deCFP-Mgapt, 5, 10 and 20 nM pPr-deCFP-Mgapt and pPr-deGFP-Mgapt were expressed in CF. Reactions were set up following the general cell free protocol using both BL21 no dialysis and Rosetta extracts. 10 µL aliquots were added in triplicate to a 384 well plate and incubated, shaking at 30°C for 4 hours. Fluorescence readings were then taken with 430-20nm excitation and 480-20 emission (optimised for CFP) filters with 1600 gain. 2-tailed T-tests, assuming unequal variance, were performed to compare the GFP and CFP RFU values,

for both Rosetta and BL21, to their respective NCs. All the samples fluorescence intensity values were significantly different to their respective NCs for BL21 (apart from 20 mM CFP) and only the deGFP intensities for Rosetta (supplementary table 1) (p-value of ≤ 0.05 = significant).

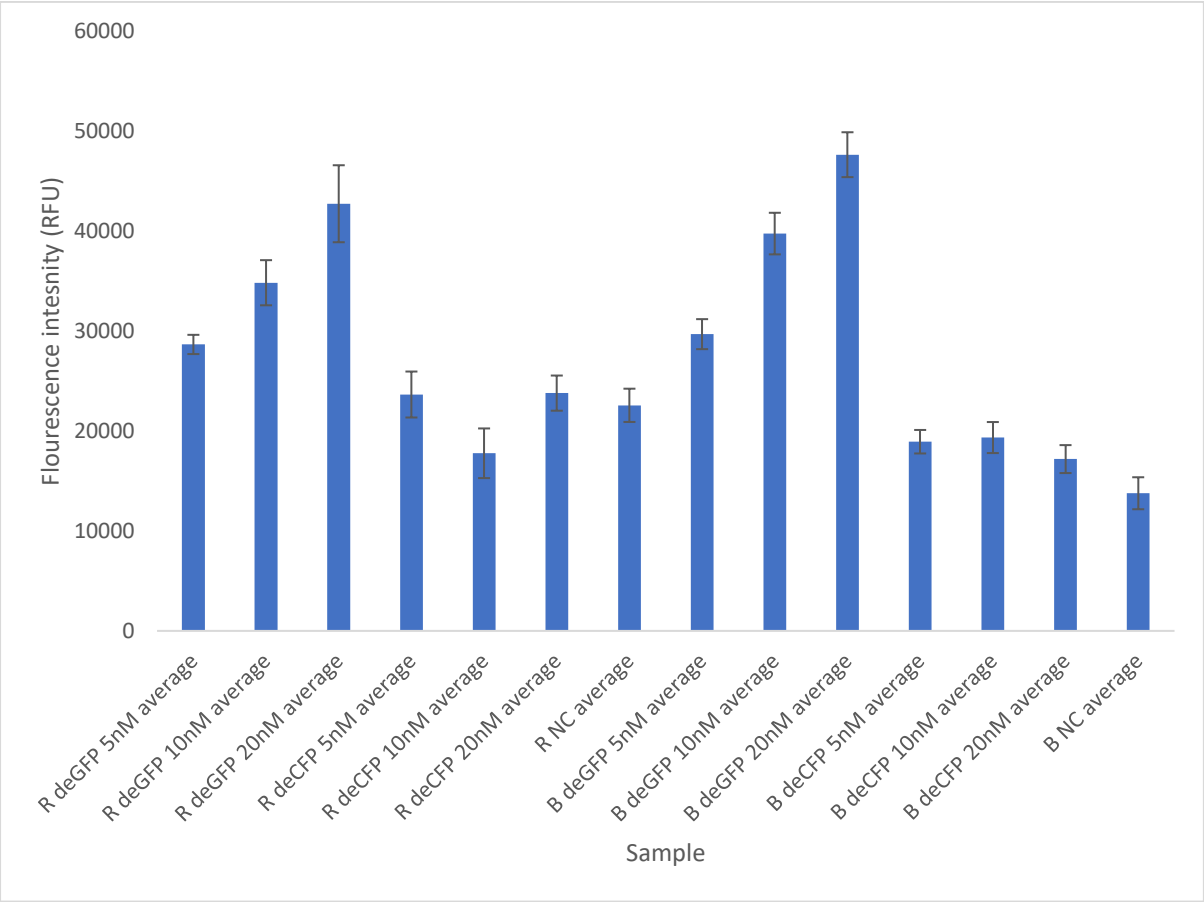


Figure 13 – deGFP vs deCFP fluorescence in Rosetta and BL21 No dialysis extract CF. 5, 10 and 20 nM pPr-deGFP-Mgapt and pPr-deCFP-Mgapt were expressed in Rosetta (R) and BL21 (B) no dialysis containing CF reactions for 4 hours shaking at 30°C. Fluorescence was then measured with 430-20nm excitation and 480-20 emission filters (optimised for CFP) with 1600 gain.

Figure 13 shows that out of the concentrations tested 20 nM deGFP statistically obtained the highest fluorescence intensity, for both Rosetta and BL21. On the other hand, 20 nM deCFP reached the highest fluorescence intensity in Rosetta extract, while 10 nM deCFP reached the highest in BL21. However, there was no statistical difference between the 1st and 2nd highest achieving CFP samples, in both Rosetta and BL21 (p-values of 0.94 and 0.73, respectively).

In addition, for both extracts the GFP values were significantly greater than the CFP values (minus average NC), with R GFP 20 nM reaching an intensity 16.5X greater than R CFP 20 (p value = 0.0057) and B GFP 20 with an intensity 8.1X greater than B CFP 10 (p value 0.00013).

4.2.7 Characterisation of BL21 no dialysis extract free amino acid levels, amino acid mixes and SP44a mVenus:

To allow for incorporation of tryptophan analogues, an alternative cell-free amino acid master mix containing all 19 canonical amino acids, minus tryptophan, was produced following the protocol outlined in the general methods section. This reduces the competition between canonical and non-canonical tryptophan for tRNA usage. In addition, a 6 mM tryptophan solution was produced to allow for supplementation of the 19AA mix to supply extracts with all 20 canonical amino acids. Furthermore, we also wanted to explore the levels of canonical tryptophan present in the BL21 no dialysis extract to assess the requirement for a depleted extract for ncAA incorporation.

Due to the previous experiment showing low level fluorescence intensity produced by the pPr-deCFP-Mgapt plasmid, it was decided that it was unsuitable for accurate and reliable measurement of ncAA incorporation. Therefore, another plasmid, SP44a mVenus, was used in its place. This plasmid was chosen due to mVenus containing a tryptophan residue, however not within its chromophore.

To analyse canonical tryptophan levels within BL21 No dialysis extract and the suitability of the 19AA mix, tryptophan solution and mVenus plasmid for ncAA incorporation and monitoring, CF reactions containing 10 nM sp44 mVenus (contains 1 tryptophan residue) + either; 0AAs, 20AAs, 19AAs and 19AA + tryptophan were produced. These reactions were set up following the standard protocol using 10 nM mVenus and BL21 no dialysis extract, however, the 1.34 mM amino acids (or 0.96 mM every amino acid apart from leucine which was 0.8mM) were replaced with either; 15.9% sterile ddH_2O (0AA), 15.1% 19AA mix and 0.8% sterile ddH_2O, 15.1% 19AA mix + 0.05 mM tryptophan or kept the standard 1.34 mM amino acids (or 0.96 mM every amino acid apart from leucine which was 0.8mM) (20AA). Reactions were then aliquoted in 10 µL triplicates onto a 384 well plate, incubated at 30°C, intermittent shaking for 8 hours in a clariostar plate reader. Fluorescence readings were then taken every 10 minutes using 485-12 excitation and 520 emission filters at 1600 gain.

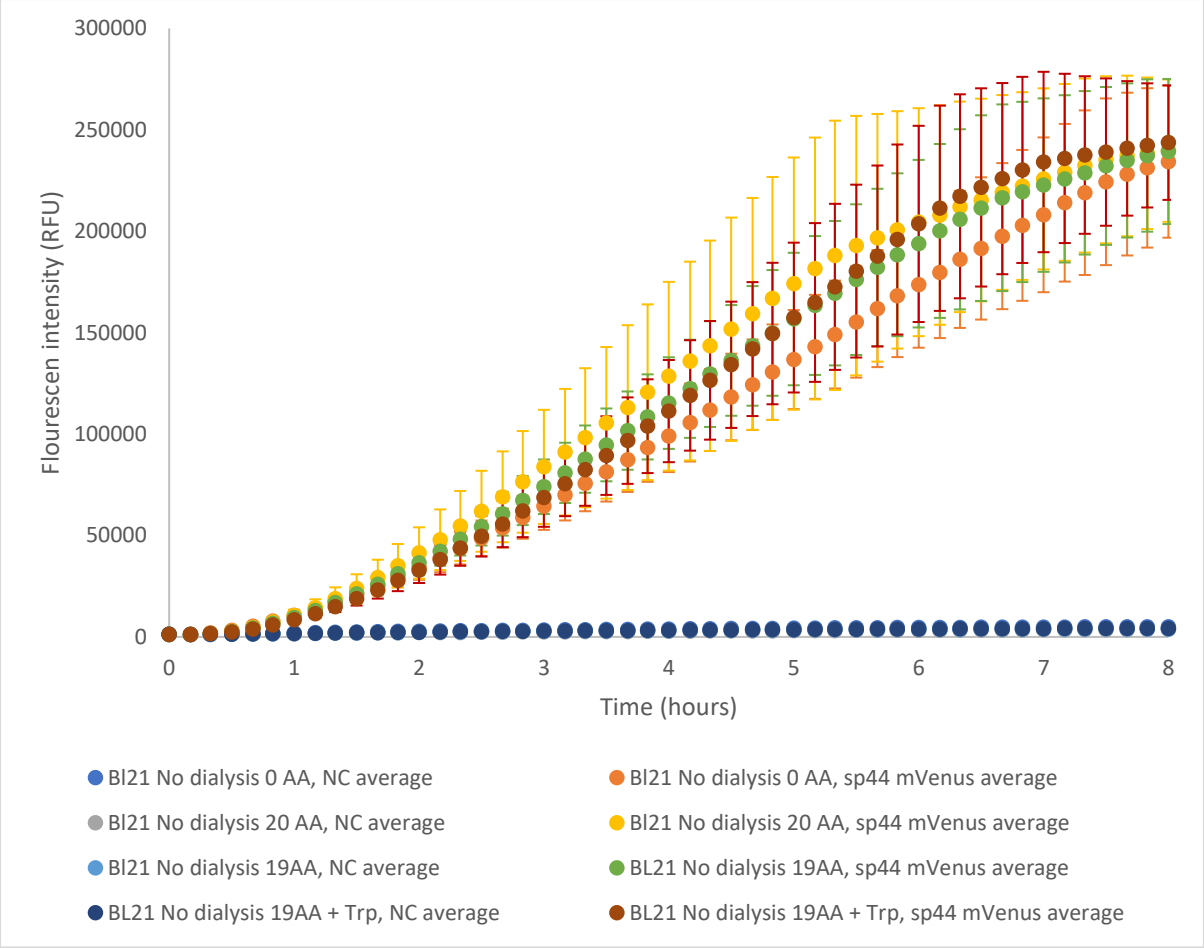


Figure 14 – 10 nM sp44 mVenus expression in BL21 No dialysis extract supplemented with either 0, 20, 19 or 19 + tryptophan amino acid mixes. BL21 No dialysis extract containing cell-free reactions supplemented with either 0, 20, 19 or 19 + tryptophan amino acid mixes and 10 nM sp44a mVenus were incubated at 30°C for 8 hours with fluorescent readings taken every 10 minutes using 485-12 excitation and 520 emission filters, 1600 gain.

Each sample produced statistically significant RFU values compared to their respective NCs with p values of 0.0088, 0.0073, 0.0076 and 0.0046 for 0AAs, 20AAs, 19AAs and 19AA + tryptophan samples, respectively. Furthermore, mVenus containing reactions reached an average fluorescence intensity of 240,000, with no mVenus containing sample reaching a statistically higher intensity than another (p values of: 0:20 AA = 0.86, 0:19AA = 0.88, 0:19

AA + T = 0.75, 20:19 AA = 0.99, 20:19 AA + T = 0.89 and 19:19 + T = 0.88). All p values were calculated using a 2 sample, 2 tailed T-test, assuming unequal variance.

4.3 Discussion:

4.3.1 Lambda red mutagenesis

Figure 10 shows that transforming BL21(DE3) Gold with 2241 and 4033 ng ΔTrpLEDC PCR fragment was unable to produce the desired recombination to remove TrpLEDC. However, the absence of any bands from transformants on plates 1 for both amounts of DNA suggests interference preventing amplification of either the FRT-Kan_FRT insert or the native TrpLEDC genes has occurred. Colonies from 2 plates containing 4033 ng transformants and a plate with transformants plated after incubation at room temperature for 24 hours were tested further using colony PCR with different primers (Trp_gDNA_F and FRT_R). However, the same outcome was seen, no bands. Therefore, a suitable tryptophan auxotrophic strain, RF4 (aspC tyrB), was purchased to be used for future cell free ncAA incorporation experiments.

4.3.2 Shift in Fluorescence emission spectrum:

As seen in figure 11 there is a difference in the emission spectrum of pPr-deGFP-Mgapt compared to pPr-deCFP-Mgapt. This is shown by the more intense emission seen by deGFP containing DH10β compared to the weaker emission by the deCFP containing strains. This indicates that the golden gate mutagenesis reactions worked to enable the mutation of tyrosine 62 (TGG) to tryptophan (TAC) thus altering the conjugated bond system of the chromophore causing a shift in fluorescence emission spectrum. However, in the future it would be better to do all the transformations in parallel using the same concentrations of pPr-deGFP-Mgapt and pPr-deCFP-Mgapt to allow for a more direct comparison in spectrum.

4.3.3 pPr-deCFP-Mgapt sequencing:

It has been previously demonstrated that mutation of tyrosine 66 to tryptophan 66 in GFP produces a shift in spectrum, producing CFP. When this sequence of GFP was aligned with deGFP it was shown that deGFP is truncated by 4 residues at the N-terminal, meaning that residue 66 in GFP is residue 62 in deGFP.

The results obtained from the sequencing provided by GENEWIZ when aligned with the known sequence of deGFP from pPr-deGFP-Mgapt, as represented in figure 12, shows the deCFP gene containing the sequence TGG at positions 598, 599 and 600 respectively, thus encoding a tryptophan at residue 62 of the protein, whereas the deGFP gene contains a tyrosine codon at the respective residue. This therefore further supports that the golden gate mutagenesis reactions allowed produced deCFP from deGFP.

4.3.4 Characterisation of pPr-deCFP-Mgapt via cell-free:

Figure 13 and 2-tailed T-tests, assuming unequal variance, show that both extracts were able to produce a statistically significant amount of deGFP, with both extracts optimal concentration being 20 nM. Although, to test this more accurately, excitation and emission filter should also be set to optimise for GFP (485-12 excitation and 520 emission), and a larger range of concentrations should be tested.

On the other hand, none of the deCFP concentrations were able to produce significant fluorescence intensity values in Rosetta extracts, but could in BL21 (apart from at 20 nM). Furthermore, unlike deGFP, the highest intensity producing deCFP concentration varied between extracts (20 nM deCFP in Rosetta and 10 nM in BL21). Moreover, the highest CFP

values were not statistically higher than the 2nd highest. In addition, in both extracts the highest GFP intensity value was significantly more than the highest CFP.

Together these results suggest that either deGFP is more easily synthesised than deCFP, or the general fluorescence of deCFP is weaker than deGFP. The latter hypothesis is supported by the literature which shows that supporting mutations within the CFP sequence are required to allow for higher levels of fluorescence e.g. enhanced CFP (eCFP). For example, when deCFP was aligned with the eCFP (fluorescence quantum yield = 0.36 [97]) sequence (provided by FPbase [98]), 3 main mutations were observed, N142I, M149T and V159A. Furthermore, Siegal-Gaskins, D et al. 2014 [78], used Pr-deCFP-MGapt as a reporter for cell free protein synthesis quantification, which when aligned with Pr-deGFP-MGapt contained N142I, M149T and V159A mutations. This shows the requirement of these extra mutations to allow for significant fluorescence intensities.

In addition, when the deCFP sequence was compared to mTurquoise2 (sequence from FPbase), the following main mutations were observed T61S, S68A, N142F, H144D, M149T, V159A, S171G, A202K. Therefore, as mTurqoise2 has been observed to obtain a quantum yield of 93% [99] , I expect the T61S, S68A, N142F, H144D, M149T, V159A, S171G and A202K mutations would increase quantum yield of deCFP. However, this may not hold true as the final quaternary structure of the optimised deCFP may be different to optimised eCFP, therefore the mutations detailed above may not have the same positive effects on deCFP as eCFP.

4.3.5 Suitability of pPr-eCFP-Mgapt as a read out for ncAA incorporation:

Due to the low-level signal produced by the deCFP plasmid, it is not suitable to use as a marker to indicate ncAA incorporation, as significant reliable excitation/emission spectrums will be

hard to obtain. For future experiments, supporting mutations should be produced to allow for sufficient levels of fluorescence to be reached.

4.3.6 BL21 Extract characterisation and amino acid mix testing:

Due to the deCFP biomarker not reaching suitable RFU values, mVenus was selected as a substitute, due to only containing 1 tryptophan residue. However, due to the tryptophan residue residing outside of the chromophore it will be unable to provide a direct indication of ncAA by the form of a shift in excitation/emission spectrum but will suggest incorporation if fluorescence observed in an extract not containing natural tryptophan. However, a potential shift in fluorescence may be observed for mVenus if unnatural amino acids are incorporated due to changes in protein structure.

Figure 14 shows that BL21 CF reactions supplemented with either 0AAs, 19AAs, 19 + T or 20AA all produced significant intensity values compared to NCs. Furthermore, none of the reactions produced a significantly higher RFU value than another. This indicates that: 1 – the extract contains enough of all amino acids to synthesis mVenus for 8 hours, including canonical tryptophan, shown by 0AA and 19AA mix reaching same intensity as standard 20AA. 2 – extract cannot be used to assess the ability of the 19AA and tryptophan mixes to allow for protein synthesis.

Due to the BL21 no dialysis extract containing enough of each amino acid to synthesise mVenus for 8 hours, future experiments should attempt the same set up in the Trp auxotrophic strain (RF4), to test for the presence of free canonical amino acids in this strain. If the same outcome is obtained (extract contains free amino acids) amino acid depletion steps will be required. For example, Singh-Blom et al. 2014 achieved efficient ncAA via a cell-free system using an auxotrophic cell extract which underwent depletion for 50 minutes in a rich media

(medium G) without supplementation with tryptophan and dialysis steps latter in extract production [73]. This would then remove free canonical tryptophan and allow for efficient incorporation of tryptophan analogues.

5 Bioinformatic investigation of lasso peptide BGC regulation

5.1 Introduction

Antimicrobial biosynthetic gene clusters are induced in response to environmental queues such as starvation to allow the host to survive and reproduce. Tight global and local genetic regulation ensures expression is only induced when required, allowing energy to otherwise be used for essential processes. To study antibiotic biosynthesis this regulation must be overcome. Therefore, studying the genetic regulation controlling biosynthesis and developing synthetic expression systems to circumvent control is essential.

During the process of identifying and cloning the RiPP biosynthetic gene clusters of interest, non-coding intergenic regions between precursor peptides and post-translational modification enzymes were identified. Their location within the clusters suggests a potential role in regulating the ratio of precursor to enzyme. The literature surrounding this area has identified intergenic regions similarly located in different RiPP biosynthetic gene clusters, some of which containing putative or characterised intrinsic terminators.

5.2 Hypothesis:

The non-coding intergenic regions observed in the RiPP BGCs of interest contain intrinsic transcription terminators, which are conserved in other *Streptomyces* species.

To test this hypothesis, analysis of RiPP BGCs was carried out to investigate the conservation of these intergenic regions within the class of lasso peptides. In addition, mRNA secondary structure prediction was utilised to identify potential regulatory elements and combined with sequence alignments of cluster homologs to assess the conservation of the structure forming sequences.

5.3 Results:

5.3.1 RiPP biosynthetic gene clusters containing intergenic regions:

Firstly, the biosynthetic gene clusters of characterised lasso peptides were identified in the minimum information about a biosynthetic gene cluster (MIBiG) database [100] , cluster sequences taken from NCBI genome sequence of respective host and inserted into benchling [101]. One exception to this was Microcin J25 who's sequence was identified in Solbiati J et al. 1996 [102] and Solbiati J et al. 1999 [54] and subsequently copied from NCBI into Benchling. Another was Microcin C (a microcin not a lasso peptide), which was identified using UniProt [103] , sequence taken from NCBI and inserted into benchling. Genes were then annotated using antiSMASH and NCBI annotations or by most similar proteins via NCBI blast. Chosen clusters and the albusnodin, lagmysin and moomysin homolog clusters of intertest are schematically represented (figure. 14).

Each identified cluster contains these intergenic regions. The length of these regions varies, with the shortest being Astexin-1 (54bp) and the longest being moomysin homolog (343bp). Furthermore, putative stem loop terminators identified in the literature have been displayed [104][105][106].

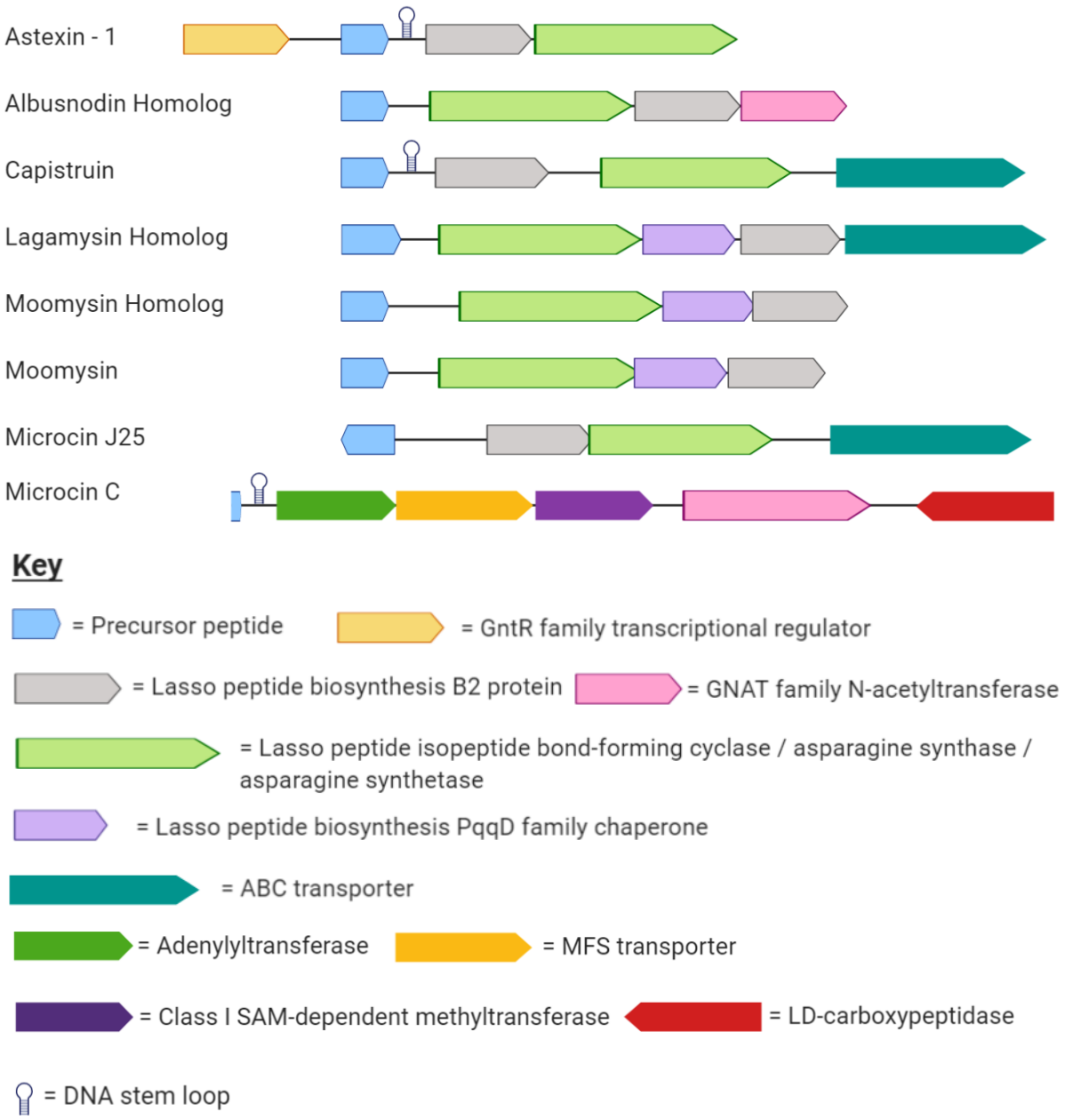


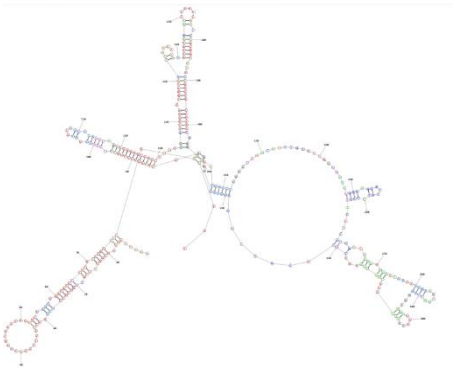
Figure 15 – a schematic representation of the biosynthetic gene clusters of a variety of lasso peptides and microcin C. Genes have been annotated by descriptions provided literature, antiSMASH or closest related protein blast. Furthermore, putative terminators described in the literature have been indicated by the DNA stem loop symbol.

5.3.2 Max Expect Intergenic Regions mRNA secondary structure prediction:

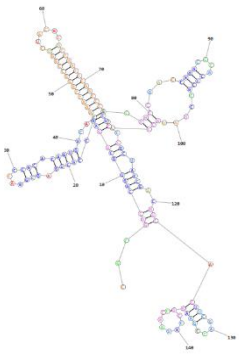
The presence of the non-coding intergenic regions between precursor peptide and PTM enzymes in a variety of different lasso peptide clusters, further supports the hypothesis that these regions may play a role in the regulation of the clusters. Therefore, the secondary structures of the mRNA sequences produced by these regions were predicted using the bioinformatic tool MaxExpect and presented in figure 16, to check for the presence of structures such as stem loops or riboswitches.

As seen in figure 16 there are regions of high probability base pair (>95%) containing stem loop structures present in all outputs, apart from Microcin J25, however the size sequences of loops vary.

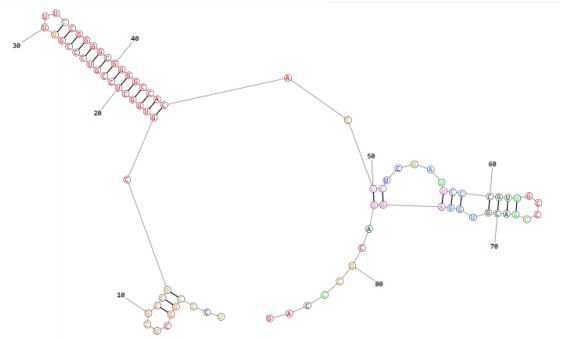
Moomysin Homolog



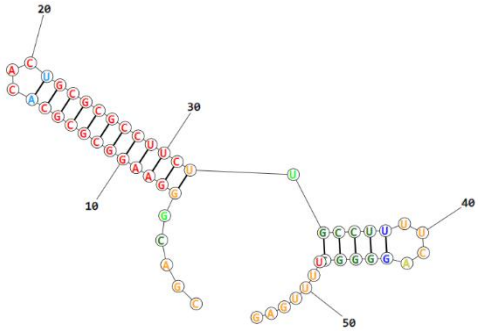
Lagmysin Homolog



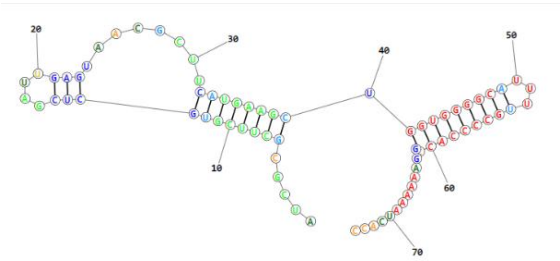
Albusnodin Homolog



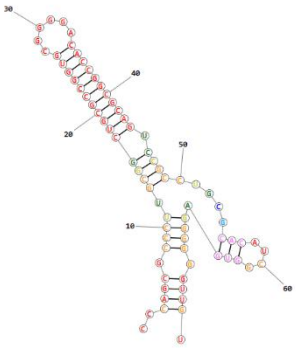
Astexin -1



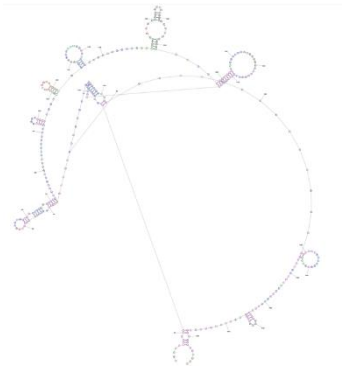
Capistruin



Moomysin



Microcin J25



Microcin C

Figure 16 – MaxExpect predicted mRNA secondary structures of the intergenic regions between the precursor peptide and PTM enzymes for a number of RiPP biosynthetic gene clusters. The colour of the base pairs represents the probability of the pair occurring, with red = >=99%, orange = >=95 - <99%, yellow = >=90 - <95%, dark green = >=80 - <90%, light green = >=70 - <80%, light blue = >=60 - <70%, dark blue = >=50 - <60 and pink = <50%.

5.3.3 Comparison of Capistruin intergenic terminator sequence predicted in the literature and MaxExpect prediction:

To assess the ability of MaxExpect to detect putative stem loop structures, secondary structures predictions were produced for intergenic regions of clusters that contain putative stem loop structures identified in the literature.

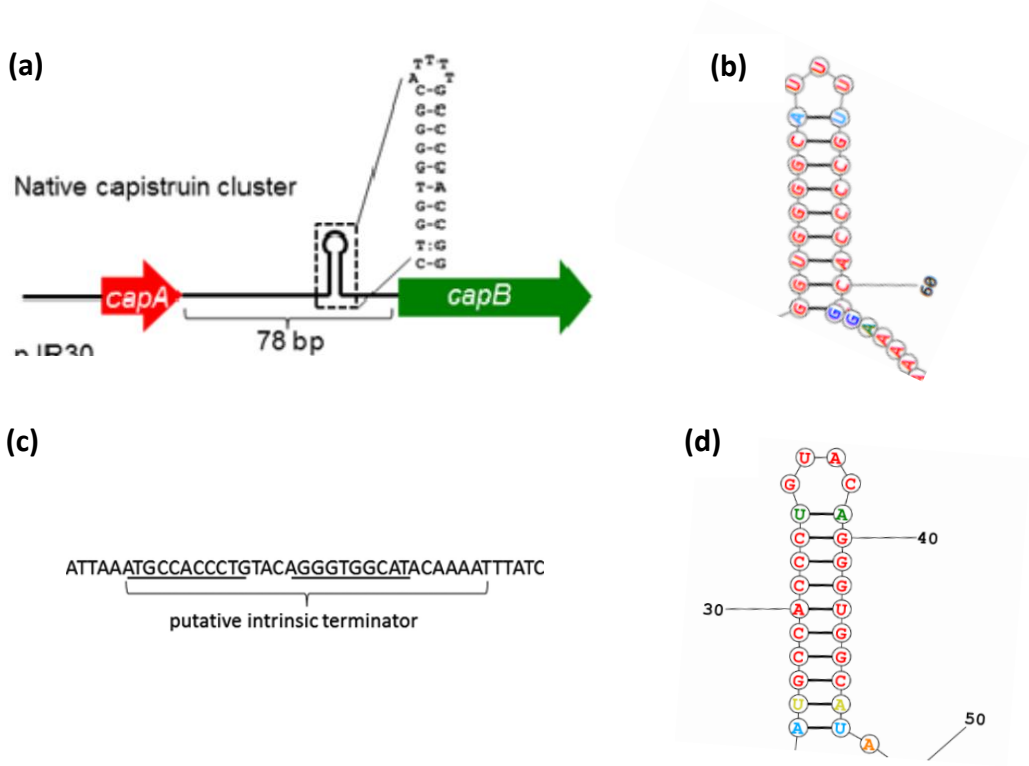


Figure 17 – Comparison of literature and MaxExpect stem loop prediction. (a) The sequence of the stem loop terminators predicted in the intergenic region between capA and capB by Pan S et al 2012 and (b) the high probability stem loop predicted by MaxExpect. (c) the stem loop structure predicted by I. Zukher et al. 2014 and (d) the high probability structure by MaxExpect for the terminator in the intergenic region between mccA and mccB of the microcin C BGC.

The hairpin loop structure predicted by Pan S et al 2012 [105] and the highest probability stem loop structure predicted by MaxExpect for the intergenic region of capA and capB in the capistruin BGC are identical, apart from 2 extra bases at the start (CT) and the end (GG) seen in the former. Furthermore, the same stem loop structure, apart from 1 base in the loop, is predicted by both I. Zukher et al. 2014 [106]and MaxExpect for microcin C.

5.3.4 albusnodin, moomysin and lagmysin precursor peptide homolog homologs:

Due to high probability stem loop structures being present in all the MaxExpect mRNA secondary structure predictions par one, analysis of the homologs of albusnodin, moomysin and lagmysin homolog clusters precursor peptides was undertaken to assess structure conservation across different *Streptomyces* species. Firstly, proteins homologous to the albusnodin, moomysin and lagmysin homolog clusters precursor peptides were searched for using NCBI BlastP and results presented in table 6.

Cluster Peptide	**Organism**	**Max Score**	**Query Cover**	**E value**	**% Identity**	**Accession**	**Full Genome Accession**
Albusnodin Homolog	*Streptomyces venezuelae strain ATCC 10595*	86.7	100%	2.00E-21	97.67%	WP_150505192.1	NZ_CP029195.1
	Streptomyces sp. SID9727	55.5	62%	5.00E-09	88.89%	WP_164399979.1	NZ_JAAGNI010000158.1
	Streptomyces sp. CS081A	55.1	100%	5.00E-09	61.36%	WP_109185061.1	NZ_KZ819170.1
	Streptomyces sp. NRRL F-6491	55.1	100%	6.00E-09	61.36%	WP_107091980.1	NZ_LGEE01000058.1
	Streptomyces sp. NRRL F-6492	55.1	100%	6.00E-09	61.36%	WP_107091980.2	NZ_LGEG01000211.1

	Streptomyces sp. SID9124	54.3	62%	1.00E-08	85.19%	WP_164640004.1	NZ_JAAGNG010000339.1
	Streptomyces sp. SID4913	53.9	86%	2.00E-08	68.42%	WP_018550185.1	NZ_WWJR01000034.1
	Streptomyces sp. CB01249	53.5	62%	3.00E-08	85.19%	WP_107422344.1	NZ_LISW01000003.1
	Streptomyces sp. AC1-42T	53.1	62%	3.00E-08	85.19%	WP_111332389.1	NZ_QKWX01000002.1
	Streptomyces sp. AC1-42W	53.1	62%	3.00E-08	85.19%	WP_111332389.1	NZ_QKWY01000001.1
	Streptomyces sp. CB02460	52.8	88%	5.00E-08	64.29%	WP_107426807.1	NZ_LIVY01000002.1
	Streptomyces atratus strain OK807	52.4	86%	1.00E-07	67.57%	WP_107408175.1	NZ_FPJO01000002.1
Moomysin homolog	*Streptomyces sp. WAC 06783*	95.5	97%	8.00E-25	100%	RSO05921.1	QHJP01000023.1
	Streptomyces albofaciens	86.7	97%	3.00E-21	89.36%	WP_167538269.1	NZ_PDCM01000001.1
	Streptomyces sp. NRRL F-5755	86.7	97%	3.00E-21	89.36%	WP_107088550.1	NZ_LGCW01000271.1
	Streptomyces rimosus subsp. rimosus strain NRRL WC-3904	85.9	97%	6.00E-21	85.11%	WP_107058775.1	NZ_JOCQ01000106.1
	Streptomyces rimosus subsp. paromomycinus strain NBRC 15454	77.4	97%	1.00E-17	76.60%	WP_125056043.1	NZ_BHZD01000001.1
	Streptomyces monomycini NRRL B-24309	69.3	97%	3.00E-14	70.21%	WP_106968949.1	NZ_KL571162.1
Lagmysin Homolog	*Streptomyces alboniger strain ATCC 12461*	75.5	100%	2.00E-15	52.38%	WP_150477720.1	CP023695.1

	Streptomyces sp. TLI_146	72	44%	4.00E-15	97.14%	WP_107503689.1	NZ_PJMX01000001.1
	Streptomyces albofaciens JCM 4342	72	44%	5.00E-15	97.14%	KAA6212439.1	PDCM01000002.1
	Streptomyces chrestomyceticus JCM 4735	71.2	44%	9.00E-15	97.14%	WP_125043397.1	NZ_BHZC01000001.1

Table 6 – NCBI BLAST results of the selected homologs of the precursor peptides from each RiPP homolog cluster being investigated and the NCBI accession number for the full genome sequences of the peptide containing organisms.

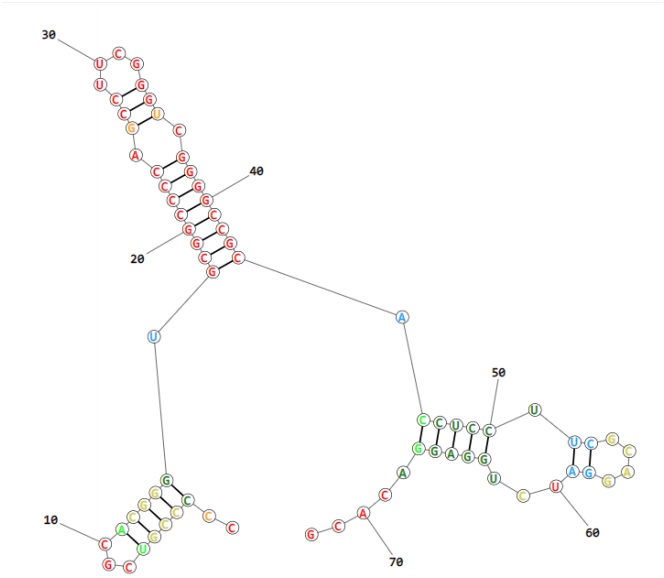
Each precursor peptide from the clusters of interest have homologs in different *Streptomyces* species, however the level of homology (indicated by E-values) varies.

5.3.5 Multiple mRNA sequence alignment of the intergenic regions between precursor peptide and PTM enzymes of the precursor peptide homologs:

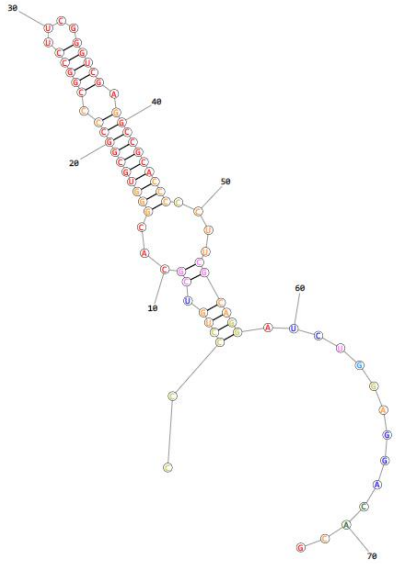
To assess the conservation of the high probability secondary structures within albusnodin, lagmysin and moomysin homolog clusters, MaxExpect secondary structure analysis was conducted on the intergenic regions between precursor peptide and PTMs for all the homologs of the albusnodin, lagmysin and moomysin cluster homologs precursor peptides identified in table 6. Furthermore, multiple sequence alignments of the intergenic regions were carried out and high probability secondary structures were cross referenced with regions of homology within alignments.

Albusnodin homolog precursor peptide homolog clusters:

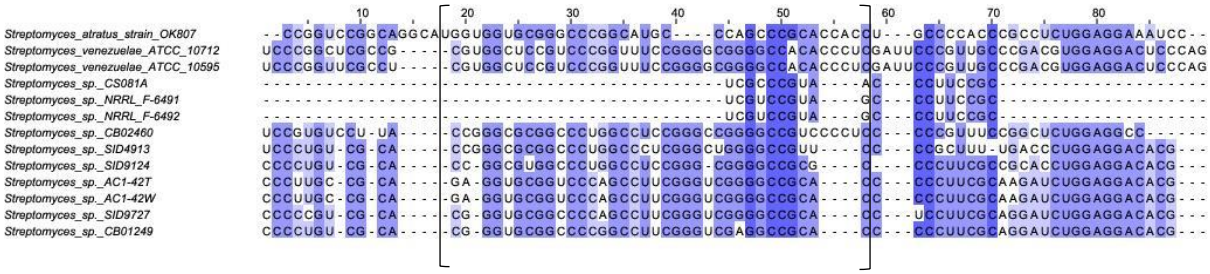
(a)



(b)



(c)



(d)

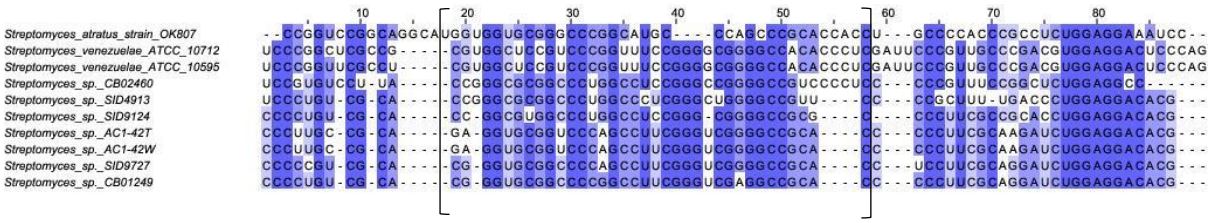
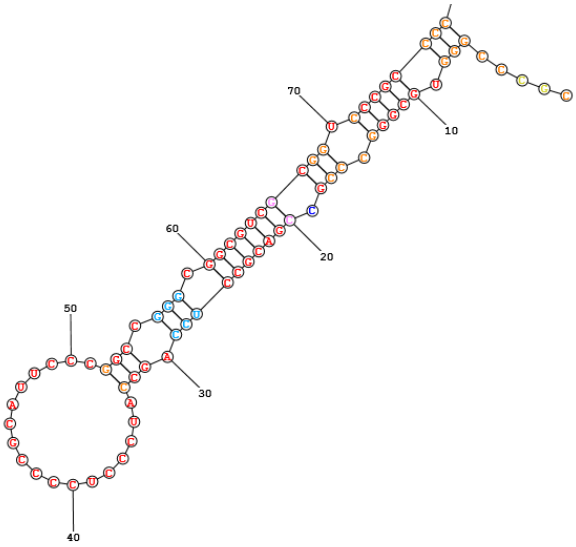


Figure 18 – The MaxExpect secondary structure predictions and multiple sequence alignment of the mRNA produced by the intergenic region between precursor peptide and PTM genes within the albusnodin homolog cluster precursor peptide homolog clusters. (a) mRNA secondary structure prediction for SID9727 and (b) CB01249. (c) Albusnodin homolog precursor peptide homologs sequence alignment and (d) Albusnodin homolog precursor peptide homologs, without the truncated sequences produced by *Streptomyces sp.* CS081A, NRRL F-6491, NRRL F-6492 and NRRL B-24572. Percentage sequence identity represented by the shade of blue, darker blue = higher homology. Region of high probability (>=95%) hair pin loop structures between 18 and 58bp.

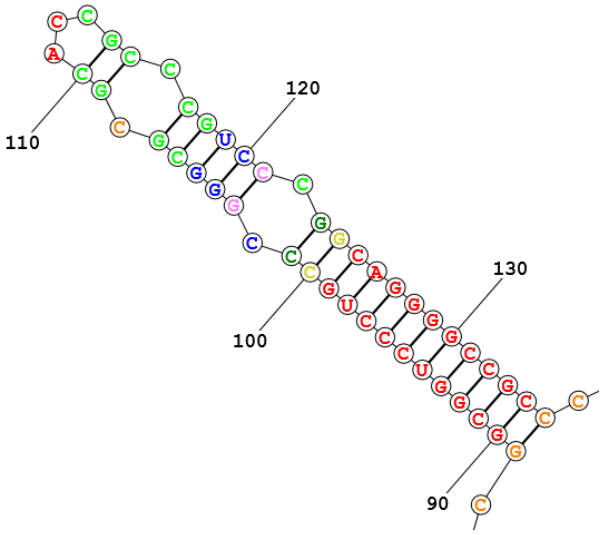
Sections (a) and (b) from figure 18. show the 2 main types of structures seen throughout the predicted structures, apart from the truncated sequences, with ATCC 10712, ATCC 10595, SID9727 and SID9124 falling in type (a) and the remainder in type (b). Furthermore, in each prediction, apart from the truncated sequences and CBO2460, there is a high probability stem loop (>=95%) structure encoded for by conserved regions between 18 and 58 in sections (c) and (d) of figure 18. Moreover, there is a highly conserved region at the 3’ end seen between regions 77-84 of (c) and (d). In addition, the truncated sequences align with region 45-70 in (c), of which some of the residues are highly conserved.

Moomysin homolog precursor peptide homolog clusters:

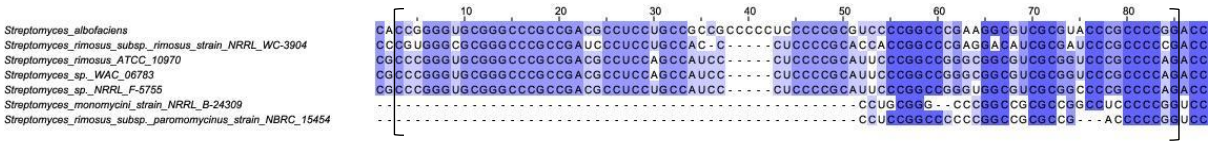
(a)



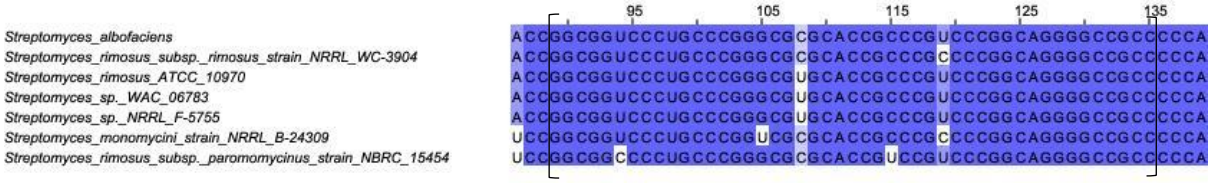
(b)



(c) **Region 1**



(d) **Region 2**



(e)

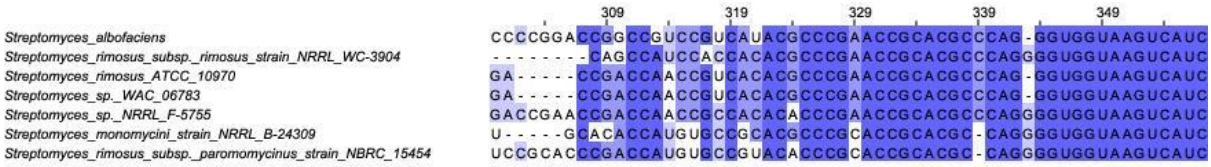


Figure 19 - The 2 main high probability region containing mRNA secondary structures seen throughout the generated predictions for the Moomysin homolog precursor peptide homologs intergenic regions and the multiple mRNA sequence alignments of these regions. (a) high probability stem loop structure from *Streptomyces Rimosus ATCC 10970* and (b) high probability region containing stem loop structure from *Streptomyces Albofaciens*. (c) region containing stem loop structures of type 1 (region 1) and (d) type 2 (region 2). Exact regions indicated by [ ]. (3) MSA 3' end.

In all secondary structure predictions, apart from *Streptomyces monomycini NRRL B-24309* and *Streptomyces rimosus subsp. paromomycinus strain NBRC 15454* there was a high probability stem loop structure from either 3-85 for *Albofaciens and WC-3904* or 6-83 bp in the sequence alignment for *ATCC 10970, WAC 06783 and NRLL-F 5755* as seen in section (a) of figure 19. Furthermore, it was seen that in all predictions there was a shorter, high probability stem loop structure formed by the high homology region 89-135 (region 2) in the alignment.

Moreover, there were a few additional high probability structures observed in *Streptomyces_rimosus_ATCC_10970*, *Streptomyces sp. WAC 06783* and *Streptomyces_sp._NRRL_F-5755* (supplementary materials). However, they were excluded from comparison to MSA due to only being seen in a couple of the sequence, showing they were not conserved across selected peptide homolog clusters.

Lagmysin homolog precursor peptide homolog clusters:

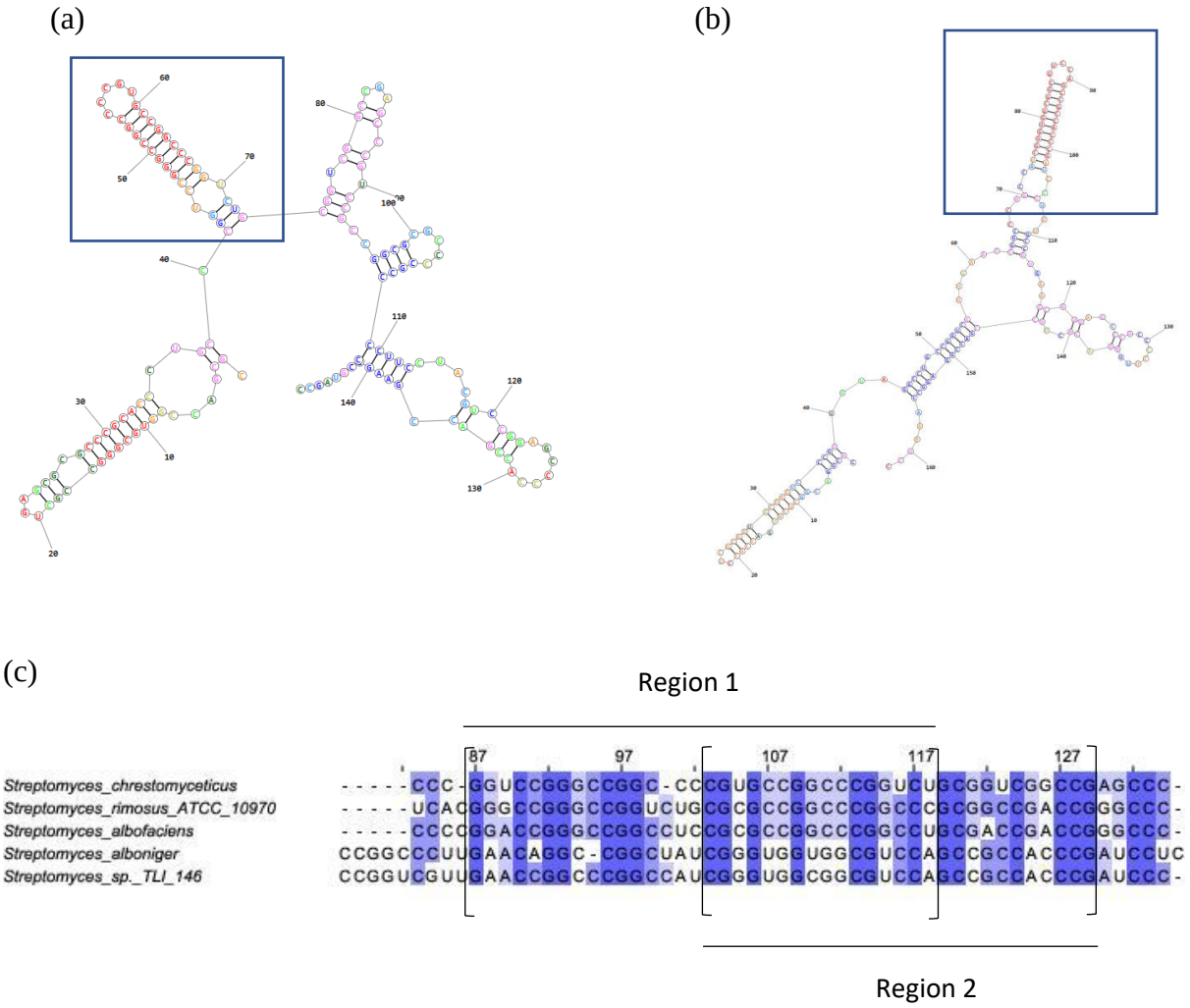


Figure 20 - The MaxExpect secondary structure predictions and multiple sequence alignment of the mRNA produced by the intergenic region between precursor peptide and PTM genes in Lagmysin homolog cluster peptide homolog clusters. (a) Max Expect prediction for Streptomyces Chrestomyceticus (b) Max Expect prediction for Streptomyces sp. TLI_146. (c) alignment from 78 - 135, with main loop containing regions indicated by [ ].

From analysis of the Max Expect predicted mRNA secondary structures for each of the lagmysin homolog cluster peptide homologs (supplementary materials) it was seen that each prediction had a high probability stem loop structure. Furthermore, the location of this stem loop structure was either between 87-118 in the alignment (figure 20) (region 1) for *Streptomyces Rimosus ATCC 10970*, *Chrestomycetius*, *Albofaciens* or 103-129 (region 2) for *TLI146*. In addition, *Alboniger* had stem loop structures in 2 regions, however, one of which in

78-98 region was lower probability than stem loop 2 in region 2. Moreover, *Chrestomyceticus* had a stem loop structure from 2-83 which contains a high probability stem structure.

5.4 Discussion:

5.4.1 Intergenic region conservation across lasso peptides:

Each lasso peptide biosynthetic gene cluster represented in figure 15 have an intergenic region between precursor peptide and PTMs. This suggests they are conserved across the class, supporting the hypothesis they serve a function.

5.4.2 MaxExpect secondary structure prediction of intergenic region mRNA of lasso peptides:

MaxExpect secondary structure prediction was selected over minimum free energy (MFE) prediction due to Zhi John Lu et al. 2009 [82]which demonstrates the higher accuracy of prediction of MaxExpect over traditional MFE.

5.4.3 Comparison of stem loops predicted in the literature and by MaxExpect:

Figure 17 shows that in both the cases of capistruin and microcin c precursor to PTM intergenic regions the stem loop structures predicted both in the literature and by MaxExpect are highly similar only varying slightly in start and end positions and loop size. Furthermore, in the papers where the structures were predicted, removal of the regions lead to differences in gene expression, suggesting they act as regulatory elements. Together, this suggests that the high-probability stem loop structures predicted by MaxExpect in the other clusters seen in figure 16 may also act as intrinsic regulators of gene expression.

5.4.4 Absence of polyU tail after stem loop structures:

It has been observed that none of the high probability stem loop structures are followed by a polyU tail, which is often required for termination. However, E. Laing et al. 2006 [107] analysed the genome of *Streptomyces coelicolor* using the bioinformatic tool GeSTer, which predicted 3,365 terminators of which only 8% contained a polyU tail. This therefore supports that the stem loop structures observed in this study could facilitate transcription termination even though they mostly didn't have polyU tails.

5.4.5 Identification of albusnodin, moomysin and lagmysin cluster homologs precursor peptide homologs:

NCBI BlastP of albusnodin, moomysin and lagmysin precursor peptide identified a variety of homologs from different *streptomyces* species, with varying levels of homology (indicated by E-values) (table 6). For example, the highest and second highest E-value for moomysin and lagmysin homologs were and 3.00E-21, 6.00E-21, and 2.00E-15, 4.00E-15, respectively, whereas albusnodins were 2.00E-21 and 5.00E-09.

However, some errors were made during the selection process. Firstly, some higher homology peptides were not selected due to believing they were from the original producing strain; however, they were from different organisms. Furthermore, some homologs were negated due to already selecting a homolog with the same E-value. However, the homolog peptides may have been in clusters with different intergenic regions which may have contained different secondary structures, which were not analysed. Moreover, there were more homologs with the same power of E-value as *Streptomyces atratus strain OK807* (n^{-7}), therefore all the homologs within this power should have been taken or ignored for consistency.

Furthermore, there were also some limitations in the method used for identification of homologs. One of which is due to the multiple sequence alignments used to produce homology scores by NCBI. For example, when searching for lagmysin homolog homologs a protein (hypothetical protein DMH15_05855, accession RSO47783.1) from streptomyces sp. WAC 06725 was identified as having a query coverage of 100% and a percentage identity of 98.73. However, this peptide is 187 amino acids in length, whereas lagmysin homolog is only 79. Therefore, DMH15_05855 is 108 base pairs longer, thus the percentage similarity of the peptides is only 41.7%. This peptide was negated from further analysis, however other peptides like this may have been missed (like *Streptomyces alboniger strain ATCC 12461* from lagmysin homolog homologs), thus reducing the reliability of the results. Therefore, future experiments could be made more accurate by calculating the percentage identity of each homolog to remove this error and to produce a more reliable value for homology rather than E-values. In addition, another limitation was the inability to identify the BGCs and intergenic regions of homologs due to inaccessibility of genomic data, meaning some homologs were missed. Furthermore, the amount of homologs taken or the range of percentage identities used for selection of homologs for further analysis could have been standardised for all 3 homologs to allow for a more accurate assessment of secondary structure conservation in the latter experiments.

5.4.6 MaxExpect predictions and multiple sequence alignment of peptide homolog clusters intergenic regions:

Conservation of high probability stem loop structures across homologs:

From the MaxExpect predictions of the intergenic mRNA secondary structures of albusnodin, lagmysin and moomysin homolog precursor peptide homolog clusters (supplementary materials), it was observed that there was usually at least one high probability (>=95%) region containing stem loop structure present. Furthermore, these stem loops, when compared to the multiple sequence alignments (figures 18, 19 and 20), often appeared in areas of high sequence conservation, thus suggesting they have been conserved for function e.g. as intrinsic terminators.

Multiple conserved high probability stem loops in homologs:

In addition, for the longer sequences (moomysin and lagmysin homolog homologs) (supplementary materials) it was seen that there were two main high probability structures identified that correlated to different conserved regions of the alignments (figures 19 and 20). In the case of moomysin homolog homologs it was observed that one of the stem loop structures was seen in all predictions apart from Streptomyces monomycini NRRL B-24309 and Streptomyces rimosus subsp. paromomycinus strain NBRC 15454, whereas another shorter high probability stem loop structure was present in all of the predictions. In addition, all loop 2 structures began at base 89 of MSA and ended at 139, whereas loop 1 start and end bases varied by 3 and 2 bases respectively. These observations in combination with the higher sequence conservation in the region of the alignment containing this shorter structure suggests that stem loop 2 is required for function rather than stem loop 1. However, to confirm this,

experiments probing these structures and their ability to regulate transcription are required (these types of experiments will be discussed in detail later).

Moreover, in the case of the lagmysin homolog homologs there were 2 main high probability stem loop structures identified, however unlike in the moomysin homolog homologs they either contained loops in region 1 or 2, apart from in *Streptomyces alboniger strain ATCC 12461* which had one partly in region 1 and one in region 2. Furthermore, both stem loop structures appear to have similar levels of conservation within the alignments, therefore suggesting that both types of stem loop may be involved in transcription regulation within their respective clusters. Again, this requires experimental data for confirmation.

Conserved sequence at 3' end of intergenic regions mRNA:

Furthermore, in each sequence alignment there is high homology at the 3' end, thus suggesting they contain a highly conserved RBS and surrounding region.

Limitations:

One limitation of the method used to assess the conservation of the of secondary structures seen across figures 18, 19 and 20 is the subjective nature of how structures and sequences are classed as 'high probability' and 'high homology'. This stems from variable base pair probabilities seen within the structure prediction, therefore making analysis of what structures are likely to occur difficult and therefore allowing errors and subconscious bias to occur in the results obtained. To improve on this for future work a more quantitative system must be developed. One suggestion for this would be to clone a variety of clusters containing putative stem loop structures and assess the secondary structure formed e.g. by mRNA crystallography. This data could then be cross referenced with bioinformatic secondary structure predictions to

analyse what total percentage probability a structure requires for its occurrence to be classed a 'high', 'medium' or 'low'. Furthermore, the level of sequence homology could be quantified as a percentage of bases that are the same, and then split into categories e.g. 80 – 100% = high homology.

Further experiments:

Experimental data is required to confirm the presence of these structures in vivo and to assess their potential role in regulation. For example, as previously mentioned mRNA crystallography could be used to assess the structures formed by these regions. Furthermore, a simple cell-free experiment using a plasmid containing a synthetic promoter, followed by the intergenic region of interest connected to a reporter gene such as GFP could be used to assess their ability to regulate expression, as the strength of termination will be inversely proportional to the level of GFP fluorescence produced.

Furthermore, after using the data from the experiments described above to address some of the limitations of the current method, the experiment could be repeated on a larger amount of lassopeptide clusters, taking a larger amount of cluster precursor peptide homologs for analysis. This would then build a stronger picture of the conservation of secondary structures within lasso peptide BGCs and their ability repress translation.

5.5.7 Effect of regulators on product yield:

Previous literature has provided conflicting reports on the function of intergenic regions and the structures they may contain. For example, Knappe et al. 2008 [48] obtained a yield of 700 μg/L of capistruin from its native producer *Burkholderia thailandensis* E264 whereas only a yield of 200 μg/L by heterologous expression in *E. coli* under the control of an inducible T7

promoter and T7 terminator. The CapA gene was translated from a vector based artificial RBS where as CapBCD were translated from their native intrinsic RBS. However, Pan et al. 2012 [105] noticed an inverted repeat in the intergenic region sequence between CapA and CapBCD which could potentially function as a regulator. Therefore, they produced a modified method of heterologous expression of capistruin where CapABCD were put under the control of an inducible tetracycline promoter, but the intergenic region between CapA and CapBCD is replaced with an optimised *E. coli* RBS. When expressed in *E. coli* a yield of 1.6 mg/L was obtained, which is >2 times increase compared to native expression and 8-fold higher than the previous heterologous expression method. This therefore suggests that the structure in the intergenic region is regulating the ability of the cluster to produce active capistruin.

Whereas a study by Zukher et al. 2014 [106]identified an intrinsic terminator within the intergenic region between mccA (precursor) and mccB genes of the microcin C biosynthetic gene cluster, which is enhanced by ribosome binding to the mccA open reading frame (orf). However, when the terminator was removed the yield of microcin C, via expression in *E. coli,* decreased 30-fold. This follows an opposite mechanism to what was observed in capistruin biosynthesis. However, microcin C is a microcin rather than a lasso peptide so the difference in effect may be due to differences in the biosynthetic pathways.

Due to the papers reporting terminators influencing gene expression in opposite ways, it suggests that terminators from different RiPP classes may affect gene expression in different ways. Therefore, the effect of the putative terminators shown here, if any, on gene expression may be different from cluster to cluster. To investigate this further, a larger dataset of BGCs with similarly located putative terminators should be generated, cloned and then the effects of each terminator on expression of their cluster should be assessed, e.g. by removal of intergenic regions. This will allow for analysis of the function of these terminators in terms of regulating product yield and any variations from RiPP class to class or from cluster to cluster within the

same class. This information will be very useful for the application of increasing heterologous expression yields of natural products.

Furthermore, the methods used in the different studies mentioned differ slightly, for example, the promoter used or if the intergenic region was deleted or replaced with an RBS. Therefore, when carrying out the further experiment discussed above, it should be done in a more comparable way. For example, by using the same plasmid type, same heterologous host or induced in native host, same constitutive promoters and replacement of the intergenic region being identical in each plasmid. This will allow for a deeper understanding of how terminators in different clusters may affect expression in different ways.

6 Conclusions

6.1 Cell-free protein synthesis of putative lasso peptides:

antiSMASH analysis identified three new lasso peptide clusters within *streptomyces* species, which have varying homologies to their closest known clusters. The natural products were then amplified and cloned into plasmids to allow for heterologous expression. The cell-free platform presented here was unable to synthesise active natural products, indicated by the absence of a zone of inhibition around the bovicin positive control disk. However, due to time constraints the limitations preventing sufficient expression were unable to be explored, however, further experiments have been suggested to identify and overcome these problems. In addition, for future experiments the exact protocol used for expression of active bovicin and the strains tested against is required to allow it to be used as a reliable positive control. If unobtainable, an alternative positive control such as nisin or capistruin (which has been expressed via cell-free) should be utalised.

6.2 Cell-free incorporation of non-canonical amino acids:

To allow for deCFP to be utilised as a reliable marker for non-canonical amino acid incorporation it's quantum yield should be increased. This could be attempted by producing the supporting mutations outlined above which may facilitate higher fluorescent intensities in cell-free. This increased fluorescence yield coupled with a shift in spectrum upon ncAA incorporation should allow for its use as an efficient marker.

In addition, BL21 no dialysis extract is unsuitable to allow for efficient non-canonical incorporation. Therefore, cell-free lysates produced from *E. coli* RF4 should be assessed for the presence of free amino acids and if required be depleted of canonical amino acids following the protocol outlined in Singh-Blom et al. 2014. Together, these extra steps may facilitate the

production of cell-free lysates for ncAA incorporation with a direct read out of successful incorporation.

6.3 Bioinformatic investigation of lasso peptide cluster regulation:

The results showed that the non-coding intergenic regions between precursor peptide and PTMs are seen in a variety of lasso peptide BGCs and microcin c. Furthermore, MaxExpect predicted high probability stem loop structures in each of these BGCs, apart from microcin J25, which are similar to putative terminators seen in the literature. In addition, removal of the intergenic regions containing the putative terminators has been seen to alter gene expression and natural product yield. This suggests that the high probability stem loop structures identified by MaxExpect function as intrinsic terminators in vivo. However, the effect these structures have on gene expression and yield requires experimental examination.

Furthermore, MaxExpect secondary structure prediction and multiple sequence alignment of mRNA of the intergenic regions between precursor peptide and PTMs from clusters containing precursor peptide homologs of the clusters of interest, suggest that these structures are conserved across different *Streptomyces* species. Furthermore, some homolog intergenic regions contained multiple high probability stem loops, however, levels of conservation of these structures indicated some may be more essential than others. However, experimental evidence is required to support this.

Together, the results support the hypothesis that the non-coding intergenic regions observed in the RiPP BGCs of interest contain intrinsic transcription terminators, which are conserved in other *streptomyces* species. However, to increase the accuracy and reliability of the results obtained in future experiments the limitations outlined here must be overcome. Furthermore, this optimisation requires combination with experimental data to further investigate the form

and function of intergenic secondary structures across lasso peptide BGCs. These experiments could then be extended to other RiPP classes due to observing a putative terminator in microcin C which has been demonstrated to effect product yield.

7 References

1. (2007) All natural. Nat. Chem. Biol. 3:351

2. Pham J V., Yilma MA, Feliz A, et al (2019) A review of the microbial production of bioactive natural products and biologics. Front Microbiol 10:1404. https://doi.org/10.3389/fmicb.2019.01404

3. Dias DA, Urban S, Roessner U (2012) A Historical overview of natural products in drug discovery. Metabolites 2:303–336

4. Baker K V., Takano E, Breitling R (2018) The “Three Cs” of Novel Antibiotic Discovery and Production through Synthetic Biology: Biosynthetic Gene Clusters, Heterologous Chassis, and Synthetic Microbial Consortia. Adv. Biosyst. 2:1800064

5. Fleming A (1929) ON THE ANTIBACTERIAL ACTION OF CULTURES OF A PENICILLIUM, WITH SPECIAL REFERENCE TO THEIR USE IN THE ISOLATION OF B. INFLUENZAE. Br J Exp Pathol 10:226–236

6. Waksman SA, Boyd Woodruff H (1940) Bacteriostatic and Bactericidal Substances Produced by a Soil Actinomyces. Soc Exp Biol Med 45:609–614

7. Waksman SASA, Lechevalier2 A, Lechevalier HA (1949) Neomycin, a New Antibiotic Active against Streptomycin-Resistant Bacteria, including Tuberculosis Organisms

8. SELMAN A. WAKSMAN, ELIZABETH S. HORNING, ERNEST L. SPENCER (1942) THE PRODUCTION OF TWO ANTIBACTERIAL SUBSTANCES, FUMIGACIN AND CLAVACIN. Science (80-) 96:202–203

9. Schatz A, Bugle E, Waksman SA (1944) Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria. Proc Soc Exp Biol Med 55:66–69. https://doi.org/10.3181/00379727-55-14461

10. Durand GA, Raoult D, Dubourg G (2019) Antibiotic discovery: history, methods and perspectives. Int. J. Antimicrob. Agents 53:371–382

11. Biggest Threats and Data | Antibiotic/Antimicrobial Resistance | CDC. https://www.cdc.gov/drugresistance/biggest-threats.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fdrugresistance%2Fbiggest_threats.html#mrsa. Accessed 16 Sep 2020

12. Seyedsayamdost MR (2014) High-Throughput platform for the discovery of elicitors of silent bacterial gene clusters. Proc Natl Acad Sci U S A 111:7266–7271. https://doi.org/10.1073/pnas.1400019111

13. Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science (80-.). 325:1089–1093

14. Okada BK, Seyedsayamdost MR (2017) Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. FEMS Microbiol Rev 035:19–33. https://doi.org/10.1093/femsre/fuw035

15. Cassini A, Diaz Högberg L, Plachouras D, et al (2019) Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. www.thelancet.com/infection. https://doi.org/10.1016/S1473-3099(18)30605-4

16. Waksman SA (1937) Associative and antagonistic effects of microorganisms: I. Historical review of antagonistic relationships. Soil Sci 43:51–68. https://doi.org/10.1097/00010694-193701000-00007

17. Terekhov SS, Osterman IA, Smirnov I V. (2018) High-throughput screening of biodiversity for antibiotic discovery. Acta Naturae 10:23–29

18. da Cunha BR, Fonseca LP, Calado CRC (2019) Antibiotic discovery: Where have we come from,

where do we go? Antibiotics 8

19. Robert L. Peck CEHJ. and KF (1946) STREPTOMYCES ANTIBIOTICS. IX. DIHYDROSTREPTOMYCIN. UTC

20. Leanza WJ, Wildonger KJ, Miller TW, Christensen BG (1979) N-Acetimidoyl-and N-Formimidoylthienamycin Derivatives: Antipseudomonal β-Lactam Antibiotics. J. Med. Chem. 22:1435–1436

21. Chen R, Wong H, Burns B (2019) New Approaches to Detect Biosynthetic Gene Clusters in the Environment. Medicines 6:32. https://doi.org/10.3390/medicines6010032

22. Rutledge PJ, Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat. Rev. Microbiol. 13:509–523

23. Ziemert N, Alanjary M, Weber T (2016) The evolution of genome mining in microbes-a review. https://doi.org/10.1039/c6np00025h

24. Cimermancic P, Medema MH, Claesen J, et al (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158:412–421. https://doi.org/10.1016/j.cell.2014.06.034

25. Li YX, Zhong Z, Zhang WP, Qian PY (2018) Discovery of cationic nonribosomal peptides as Gram-negative antibiotics through global genome mining. Nat Commun 9:. https://doi.org/10.1038/s41467-018-05781-6

26. Nguyen CT, Dhakal D, Pham VTT, et al (2020) Recent Advances in Strategies for Activation and Discovery/Characterization of Cryptic Biosynthetic Gene Clusters in Streptomyces. Microorganisms 8:616. https://doi.org/10.3390/microorganisms8040616

27. Ochi K, Hosaka T New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. https://doi.org/10.1007/s00253-012-4551-9

28. Zhu H, Sandiford SK, Van Wezel GP (2014) Triggers and cues that activate antibiotic production by actinomycetes. J. Ind. Microbiol. Biotechnol. 41:371–386

29. Ren H, Wang B, Zhao H (2017) Breaking the silence: new strategies for discovering novel natural products. Curr. Opin. Biotechnol. 48:21–27

30. Bode HB, Bethe B, Höfs R, Zeeck A (2002) Big effects from small changes: Possible ways to explore nature's chemical diversity. ChemBioChem 3:619–627

31. Scherlach K, Hertweck C (2006) Discovery of aspoquinolones A-D, prenylated quinoline-2-one alkaloids from Aspergillus nidulans, motivated by genome mining. Org Biomol Chem 4:3517–3520. https://doi.org/10.1039/b607011f

32. Williams RB, Henrikson JC, Hoover AR, et al (2008) Epigenetic remodeling of the fungal secondary metabolome. Org Biomol Chem 6:1895–1997. https://doi.org/10.1039/b804701d

33. Moore JM, Bradshaw E, Seipke RF, et al (2012) Use and discovery of chemical elicitors that stimulate biosynthetic gene clusters in streptomyces bacteria. In: Methods in Enzymology. Academic Press Inc., pp 367–385

34. Behnken S, Lincke T, Kloss F, et al (2012) Antiterminator-mediated unveiling of cryptic polythioamides in an anaerobic bacterium. Angew Chemie - Int Ed 51:2425–2428. https://doi.org/10.1002/anie.201108214

35. Thanapipatsiri A, Gomez-Escribano JP, Song L, et al (2016) Discovery of Unusual Biaryl Polyketides by Activation of a Silent Streptomyces venezuelae Biosynthetic Gene Cluster. ChemBioChem 17:2189–2198. https://doi.org/10.1002/cbic.201600396

36. Bergmann S, Schümann J, Scherlach K, et al (2007) Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans. Nat Chem Biol 3:213–217. https://doi.org/10.1038/nchembio869

37. Gomez-Escribano JP, Castro JF, Razmilic V, et al (2019) Heterologous expression of a cryptic gene cluster from streptomyces leeuwenhoekii C34T yields a novel Lasso Peptide, Leepeptin. Appl Environ Microbiol 85:. https://doi.org/10.1128/AEM.01752-19

38. Eustáquio AS, Ziemert N (2018) Identification of Natural Product Biosynthetic Gene Clusters from Bacterial Genomic Data

39. Mohimani H, Pevzner PA (2016) Dereplication, sequencing and identification of peptidic natural products: From genome mining to peptidogenomics to spectral networks. Nat. Prod. Rep. 33:73–86

40. 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. 23:31–44

41. Martínez-Núñez MA, López VEL y (2016) Nonribosomal peptides synthetases and their applications in industry. Sustain Chem Process 4:13. https://doi.org/10.1186/s40508-016-0057-6

42. Arnison PG, Bibb MJ, Bierbaum G, et al (2013) Ribosomally synthesized and post-translationally modified peptide natural products: Overview and recommendations for a universal nomenclature. Nat. Prod. Rep. 30:108–160

43. Hudson GA, Mitchell DA (2018) RiPP antibiotics: biosynthesis and engineering potential. Curr. Opin. Microbiol. 45:61–69

44. Tietz JI, Schwalen CJ, Patel PS, et al (2017) A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat Chem Biol 13:470–478. https://doi.org/10.1038/nchembio.2319

45. Martin-Gómez H, Tulla-Puche J (2018) Lasso peptides: chemical approaches and structural elucidation. Org. Biomol. Chem. 16:5065–5080

46. Delgado MA, Rintoul MR, Farías RN, Salomón RA (2001) Escherichia coli RNA polymerase is the target of the cyclopeptide antibiotic microcin J25. J Bacteriol 183:4543–4550. https://doi.org/10.1128/JB.183.15.4543-4550.2001

47. Salom6n RA, Farias RN (1992) Microcin 25, a Novel Antimicrobial Peptide Produced by Escherichia coli

48. Knappe TA, Linne U, verine Zirah S, et al Isolation and Structural Characterization of Capistruin, a Lasso Peptide Predicted from the Genome Sequence of Burkholderia thailandensis E264. https://doi.org/10.1021/ja802966g

49. Kuznedelov K, Semenova E, Knappe TA, et al (2011) The antibacterial threaded-lasso peptide capistruin inhibits bacterial RNA polymerase. J Mol Biol 412:842–848. https://doi.org/10.1016/j.jmb.2011.02.060

50. Yano K, Toki S, Nakanishi S, et al (1996) MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from streptomyces sp. - I. Isolation, structural determination and biological properties of MS-271. Bioorganic Med Chem 4:115–120. https://doi.org/10.1016/0968-0896(95)00175-1

51. Tsunakawa M, Hu SL, Hoshino Y, et al (1995) Siamycins I and II, New Anti-HIV Peptides: I. Fermentation, Isolation, Biological Activity and Initial Characterization. J Antibiot (Tokyo) 48:433–434. https://doi.org/10.7164/antibiotics.48.433

52. Lin PF, Samanta H, Bechtold CM, et al (1996) Characterization of siamycin I, a human immunodeficiency virus fusion inhibitor. Antimicrob Agents Chemother 40:133–138. https://doi.org/10.1128/aac.40.1.133

53. Helynck G, Dubertret C, Mayaux JF, Leboul J (1993) Isolation of RP 71955, a new Anti-Hiv-1 peptide secondary metabolite. J Antibiot (Tokyo) 46:1756–1757. https://doi.org/10.7164/antibiotics.46.1756

54. Solbiati JO, Ciaccio M, Farías RN, et al (1999) Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. J Bacteriol 181:2659–2662. https://doi.org/10.1128/jb.181.8.2659-2662.1999

55. Yan KP, Li Y, Zirah S, et al (2012) Dissecting the Maturation Steps of the Lasso Peptide Microcin J25 in vitro. ChemBioChem 13:1046–1052. https://doi.org/10.1002/cbic.201200016

56. Choudhury HG, Tong Z, Mathavan I, et al (2014) Structure of an antibacterial peptide ATP-binding cassette transporter in a novel outward occluded state. Proc Natl Acad Sci U S A 111:9145–9150. https://doi.org/10.1073/pnas.1320506111

57. Rosengren KJ, Clark RJ, Daly NL, et al (2003) Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. J Am Chem Soc 125:12464–12474. https://doi.org/10.1021/ja0367703

58. Tissikres A, Schlessinger D, Gros ; Matthaei F, et al (1960) THE DEPENDENCE OF CELL-FREE PROTEIN SYNTHESIS IN E. COLI UPON NATURALLY OCCURRING OR SYNTHETIC POLYRIBONUCLEOTIDES

59. Dondapati SK, Stech M, Zemella A, Kubick S (2020) Cell-Free Protein Synthesis: A Promising Option for Future Drug Development. BioDrugs 34:327–348

60. Gregorio NE, Levine MZ, Oza JP (2019) A User’s Guide to Cell-Free Protein Synthesis. Methods Protoc 2:24. https://doi.org/10.3390/mps2010024

61. Liu R, Zhang Y, Zhai G, et al (2020) A Cell-Free Platform Based on Nisin Biosynthesis for Discovering Novel Lanthipeptides and Guiding their Overproduction In Vivo. Adv Sci 7:2001616. https://doi.org/10.1002/advs.202001616

62. Rolf J, Rosenthal K, Lütz S (2019) Application of cell-free protein synthesis for faster biocatalyst development. Catalysts 9

63. Silverman AD, Karim AS, Jewett MC (2020) Cell-free gene expression: an expanded repertoire of applications. Nat. Rev. Genet. 21:151–170

64. Khambhati K, Bhattacharjee G, Gohil N, et al (2019) Exploring the Potential of Cell-Free Protein Synthesis for Extending the Abilities of Biological Systems. Front. Bioeng. Biotechnol. 7:248

65. Goshima N, Kawamura Y, Fukumoto A, et al (2008) Human protein factory for converting the transcriptome into an in vitro-expressed proteome. Nat Methods 5:1011–1017. https://doi.org/10.1038/nmeth.1273

66. Liu WQ, Zhang L, Chen M, Li J (2019) Cell-free protein synthesis: Recent advances in bacterial extract sources and expanded applications. Biochem Eng J 141:182–189. https://doi.org/10.1016/j.bej.2018.10.023

67. Li J, Zhang L, Liu W (2018) Cell-free synthetic biology for in vitro biosynthesis of pharmaceutical natural products. Synth Syst Biotechnol 3:83–89. https://doi.org/10.1016/j.synbio.2018.02.002

68. A. Roessner C, B. Spencer J, J. Stolowich N, et al (1994) Genetically engineered synthesis of precorrin-6x and the complete corrinoid, hydrogenobyrinic acid, an advanced precursor of vitamin B12. Chem Biol 1:119–124. https://doi.org/10.1016/1074-5521(94)90050-7

69. Moore SJ (2019) Enzyme alchemy: cell-free synthetic biochemistry for natural products. Emerg Top Life Sci 3:529–535. https://doi.org/10.1042/etls20190083

70. Goering AW, Li J, McClure RA, et al (2017) In vitro reconstruction of nonribosomal peptide biosynthesis directly from DNA using cell-free protein synthesis. ACS Synth Biol 6:39–44. https://doi.org/10.1021/acssynbio.6b00160

71. Si Y, Kretsch AM, Daigh LM, et al (2020) Cell-Free Biosynthesis to Evaluate Lasso Peptide Formation and Enzyme-Substrate Tolerance. bioRxiv 2020.11.25.399105.

https://doi.org/10.1101/2020.11.25.399105

72. Gao W, Cho E, Liu Y, Lu Y (2019) Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. Front. Pharmacol. 10

73. Singh-Blom A, Hughes RA, Ellington AD (2014) An amino acid depleted cell-free protein synthesis system for the incorporation of non-canonical amino acid analogs into proteins. J Biotechnol 178:12–22. https://doi.org/10.1016/j.jbiotec.2014.02.009

74. Johnson DBF, Xu J, Shen Z, et al (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nat Chem Biol 7:779–786. https://doi.org/10.1038/nchembio.657

75. Lajoie MJ, Rovner AJ, Goodman DB, et al (2013) Genomically recoded organisms expand biological functions. Science (80-) 342:357–360. https://doi.org/10.1126/science.1241459

76. Hong SH, Ntai I, Haimovich AD, et al (2014) Cell-free protein synthesis from a release factor 1 deficient escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. ACS Synth Biol 3:398–409. https://doi.org/10.1021/sb400140t

77. Lin MT, Fukazawa R, Miyajima-Nakano Y, et al (2015) Escherichia coli Auxotroph Host Strains for Amino Acid-Selective Isotope Labeling of Recombinant Proteins. In: Methods in Enzymology. Academic Press Inc., pp 45–66

78. Siegal-Gaskins D, Tuza ZA, Kim J, et al (2014) Gene circuit performance characterization and resource usage in a cell-free “breadboard.” ACS Synth Biol 3:416–425. https://doi.org/10.1021/sb400203p

79. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645. https://doi.org/10.1073/pnas.120163297

80. Levine MZ, So B, Mullin AC, et al (2019) Redesigned upstream processing enables a 24-hour workflow from E. coli cells to cell-free protein synthesis. bioRxiv 729699

81. Welcome to the MaxExpect Web Server. https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/MaxExpect/MaxExpect.html. Accessed 20 Dec 2020

82. Lu ZJ, Gloor JW, Mathews DH (2009) Improved RNA secondary structure prediction by maximizing expected pair accuracy. RNA 15:1805–1813. https://doi.org/10.1261/rna.1643609

83. DNA<->RNA->protein. http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm. Accessed 20 Dec 2020

84. Protein BLAST: search protein databases using a protein query. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. Accessed 20 Dec 2020

85. Sievers F, Wilm A, Dineen D, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

86. Waterhouse AM, Procter JB, Martin DMA, et al (2009) Sequence analysis Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinforma Appl NOTE 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033

87. Corpet F Nucleic Acids Research Multiple sequence alignment with hierarchical clustering

88. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/. Accessed 20 Dec 2020

89. Zhang P, Wang J, Ding X, et al (2019) Exploration of the Tolerance Ability of a Cell-Free

Biosynthesis System to Toxic Substances. Appl Biochem Biotechnol 189:1096–1107. https://doi.org/10.1007/s12010-019-03039-5

90. Mungan MD, Alanjary M, Blin K, et al (2020) ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. Nucleic Acids Res 48:W546–W552. https://doi.org/10.1093/nar/gkaa374

91. Hegemann JD (2020) Factors Governing the Thermal Stability of Lasso Peptides. ChemBioChem 21:7–18. https://doi.org/10.1002/cbic.201900364

92. Codon Usage Database. https://www.kazusa.or.jp/codon/. Accessed 20 Dec 2020

93. Mantovani HC, Hu H, Warobo RW, Russell JB (2002) Bovicin HC5, a bacteriocin from Streptococcus bovis HC5. Microbiology 148:3347–3352

94. Duquesne S, Destoumieux-Garzón D, Zirah S, et al (2007) Two Enzymes Catalyze the Maturation of a Lasso Peptide in Escherichia coli. Chem Biol 14:793–803. https://doi.org/10.1016/j.chembiol.2007.06.004

95. Cheng C, Hua ZC (2020) Lasso Peptides: Heterologous Production and Potential Medical Application. Front. Bioeng. Biotechnol. 8:1131

96. Kaweewan I, Hemmi H, Komaki H, et al (2018) Isolation and structure determination of a new lasso peptide specialicin based on genome mining. Bioorganic Med Chem 26:6050–6055. https://doi.org/10.1016/j.bmc.2018.11.007

97. Kremers GJ, Goedhart J, Van Munster EB, Gadella TWJ (2006) Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET förster radius. Biochemistry 45:6570–6580. https://doi.org/10.1021/bi0516273

98. Lambert TJ (2019) FPbase: a community-editable fluorescent protein database. Nat. Methods 16:277–278

99. Goedhart J, Von Stetten D, Noirclerc-Savoye M, et al (2012) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nat Commun 3:1–9. https://doi.org/10.1038/ncomms1738

100. Kautsar SA, Blin K, Shaw S, et al (2020) MIBiG 2.0: A repository for biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458. https://doi.org/10.1093/nar/gkz882

101. Cloud-Based Informatics Platform for Life Sciences R&D | Benchling. https://www.benchling.com/. Accessed 20 Dec 2020

102. Solbiati JO, Ciaccio M, Fari´as RN, et al (1996) Genetic Analysis of Plasmid Determinants for Microcin J25 Production and Immunity

103. Bateman A (2019) UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res 47:D506–D515. https://doi.org/10.1093/nar/gky1049

104. Maksimov MO, Pelczer I, Link AJ (2012) Precursor-centric genome-mining approach for lasso peptide discovery. Proc Natl Acad Sci U S A 109:15223–15228. https://doi.org/10.1073/pnas.1208978109

105. Pan SJ, Rajniak J, Maksimov MO, Link AJ (2012) The role of a conserved threonine residue in the leader peptide of lasso peptide precursors. Chem Commun 48:1880–1882. https://doi.org/10.1039/c2cc17211a

106. Zukher I, Novikova M, Tikhonov A, et al (2014) Ribosome-controlled transcription termination is essential for the production of antibiotic microcin C. Nucleic Acids Res 42:11891–11902. https://doi.org/10.1093/nar/gku880

107. Laing E, Mersinias V, Smith CP, Hubbard SJ (2006) Analysis of gene expression in operons of Streptomyces coelicolor. Genome Biol 7:. https://doi.org/10.1186/gb-2006-7-6-r46

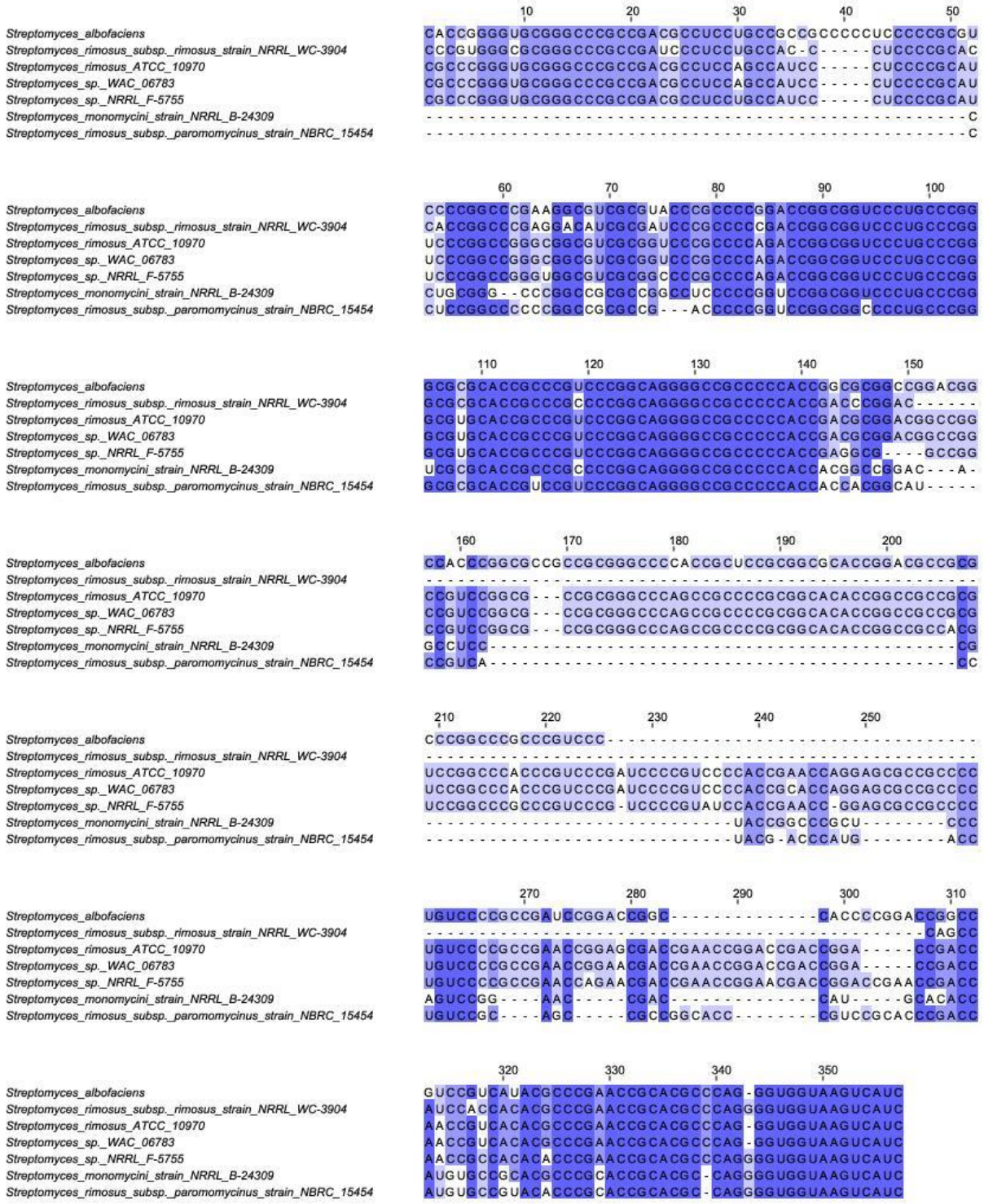
8 Supplementary materials

Extract	Sample	p-value	Significant (p value ≤ 0.05)
Rosetta	NC:GFP 5nM	0.01001903	Y
	NC:GFP 10nM	0.00226264	Y
	NC:GFP 20nM	0.00517587	Y
	NC:CFP 5nM	0.5487012	N
	NC:CFP 10nM	0.05808909	N
	NC:CFP 20nM	0.43199166	N
BL21	NC:GFP 5nM	0.00023649	Y
	NC:GFP 10nM	0.00010551	Y
	NC:GFP 20nM	6.272E-05	Y
	NC:CFP 5nM	0.01317325	Y
	NC:CFP 10nM	0.01236406	Y
	NC:CFP 20nM	0.05047734	N

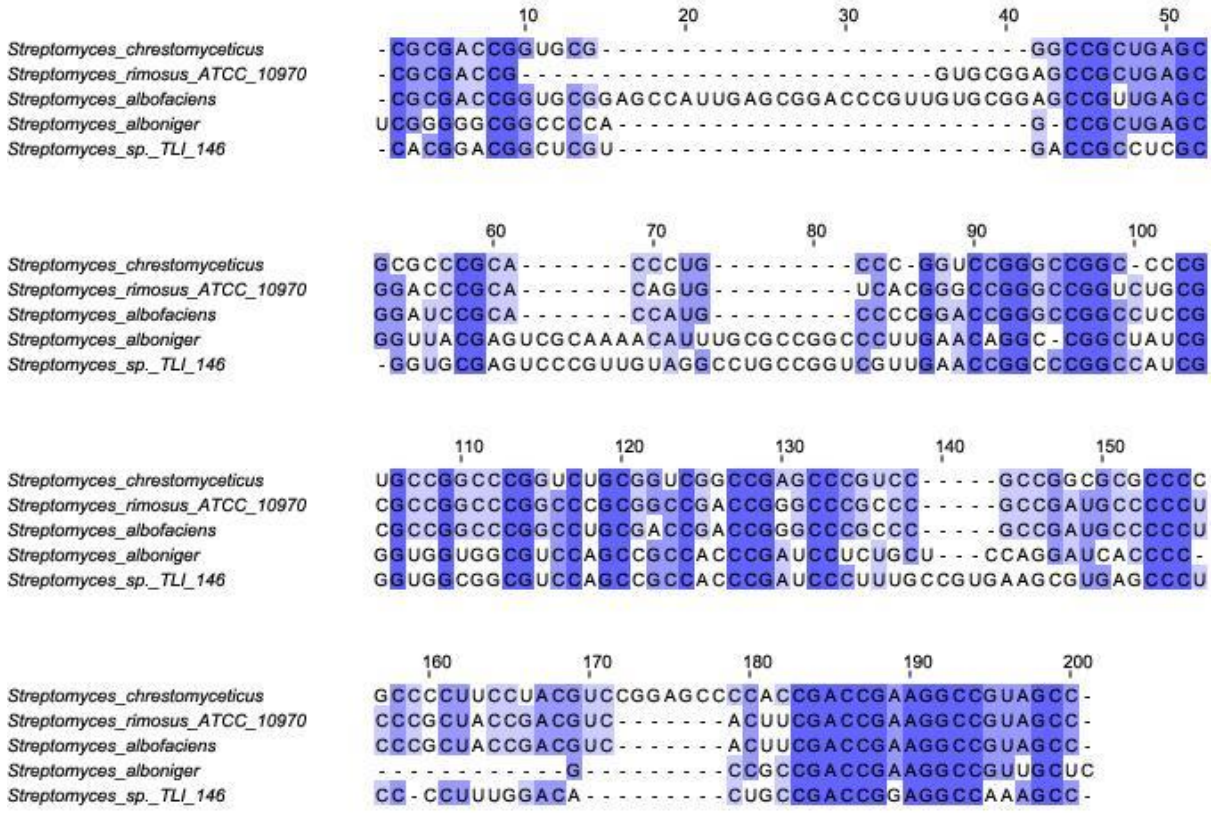
Supplementary table 1 – P values of Rosetta and BL21 cell-free deGFP and deCFP fluorescence intensities vs respective controls. Y = yes, N = no

```
                 1        10        20        30        40        50        60        70        80        90       100       110       120       130
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
My_Pr-deCFP-MGapt; MELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID
       eCFP; MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID
  Consensus .....eELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID

                 131      140       150       160       170       180       190       200       210       220       230       239
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
My_Pr-deCFP-MGapt; FKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI
       eCFP; FKEDGNILGHKLEYNYISHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
  Consensus FKEDGNILGHKLEYNYnSHNVYIwADKQKNGIKaNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI.........

                 1        10        20        30        40        50        60        70        80        90       100       110       120       130
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
   Pr-deCFP-MGapt; MELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG
My_Pr-deCFP-MGapt; MELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG
  Consensus MELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG

                 131      140       150       160       170       180       190       200       210       220  225
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------+----|
   Pr-deCFP-MGapt; NILGHKLEYNYISHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI
My_Pr-deCFP-MGapt; NILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI
  Consensus NILGHKLEYNYnSHNVYIwADKQKNGIKaNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI

                 1        10        20        30        40        50        60        70        80        90       100       110       120       130
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
My_Pr-deCFP-MGapt; MELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTWGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID
 mTurquoise; MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLSWGVQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID
  Consensus .....eELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLsWGVQCFaRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGID

                 131      140       150       160       170       180       190       200       210       220       230       239
                 |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
My_Pr-deCFP-MGapt; FKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI
 mTurquoise; FKEDGNILGHKLEYNYFSDNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
  Consensus FKEDGNILGHKLEYNYnSdNVYIwADKQKNGIKaNFKIRHNIEDGgVQLADHYQQNTPIGDGPVLLPDNHYLSTQSaLSKDPNEKRDHMVLLEFVTAAGI.........
```

Supplementary figure 1 – amino acid sequence alignments of the deCFP gene from the Pr-deCFP-MGapt plasmid produced in this study with eCFP (FPbase), Pr-deCFP-MGapt produced by Siegal-Gaskins, D et al. 2014 and mTurquoise (FPbase).



Supplementary figure 2 - Full length MSA of Moomysin homolog homologs



Supplementary figure 3 - Full length MSA of Lagmysin homolog homologs

<u>Max expect secondary structure predictions of all clusters of interest precursor peptide homolog clusters intergenic regions</u>

The key provided by figure 16 shows the colour code for base pair binding probability.

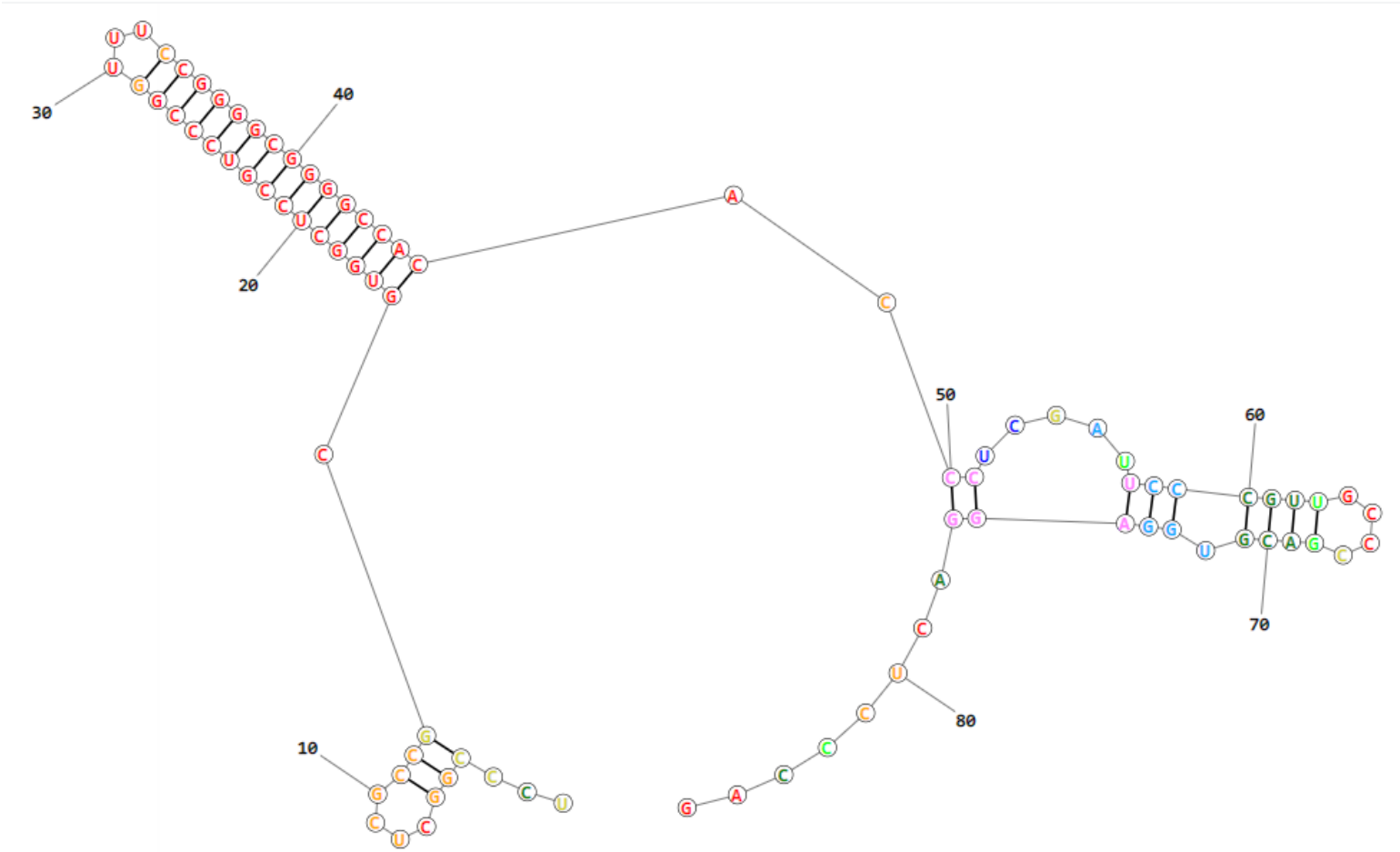
Albusnodin Homolog Cluster Peptide Homologs intergenic regions secondary structures

Intergenic regions RNA sequence provided.

Streptomyces Venezuela ATCC_10712 (Original Albusnodin Homolog):

UCCCGGCUCGCCGCGUGGCUCCGUCCCGGUUUCCGGGGCGGGGCCACACCCUC

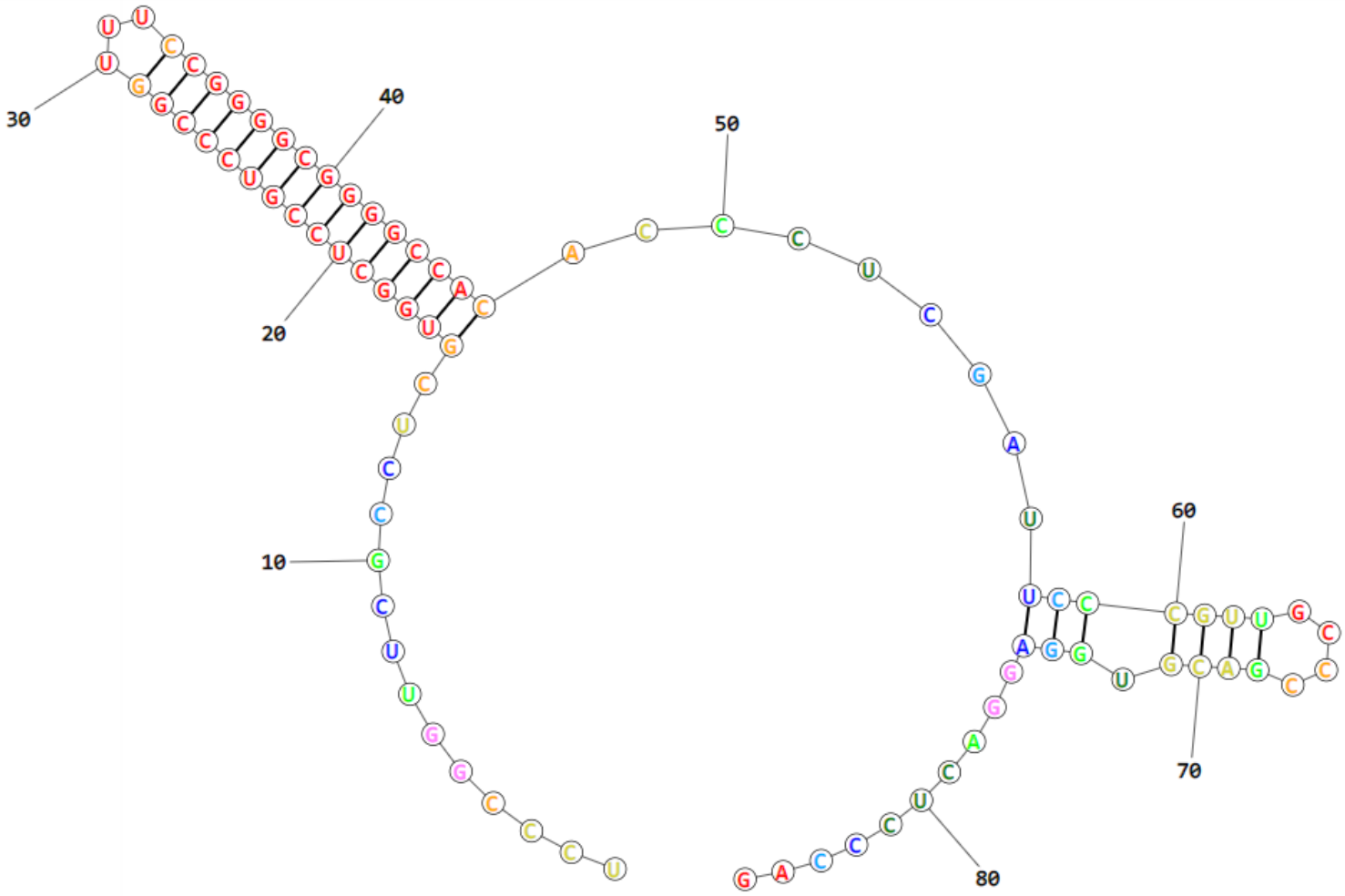
GAUUCCCGUUGCCCGACGUGGAGGACUCCCAG



Streptomyces Venezuela ATCC 10595:

UCCCGGUUCGCCUCGUGGCUCCGUCCCGGUUUCCGGGGCGGGGCCACACCCUC

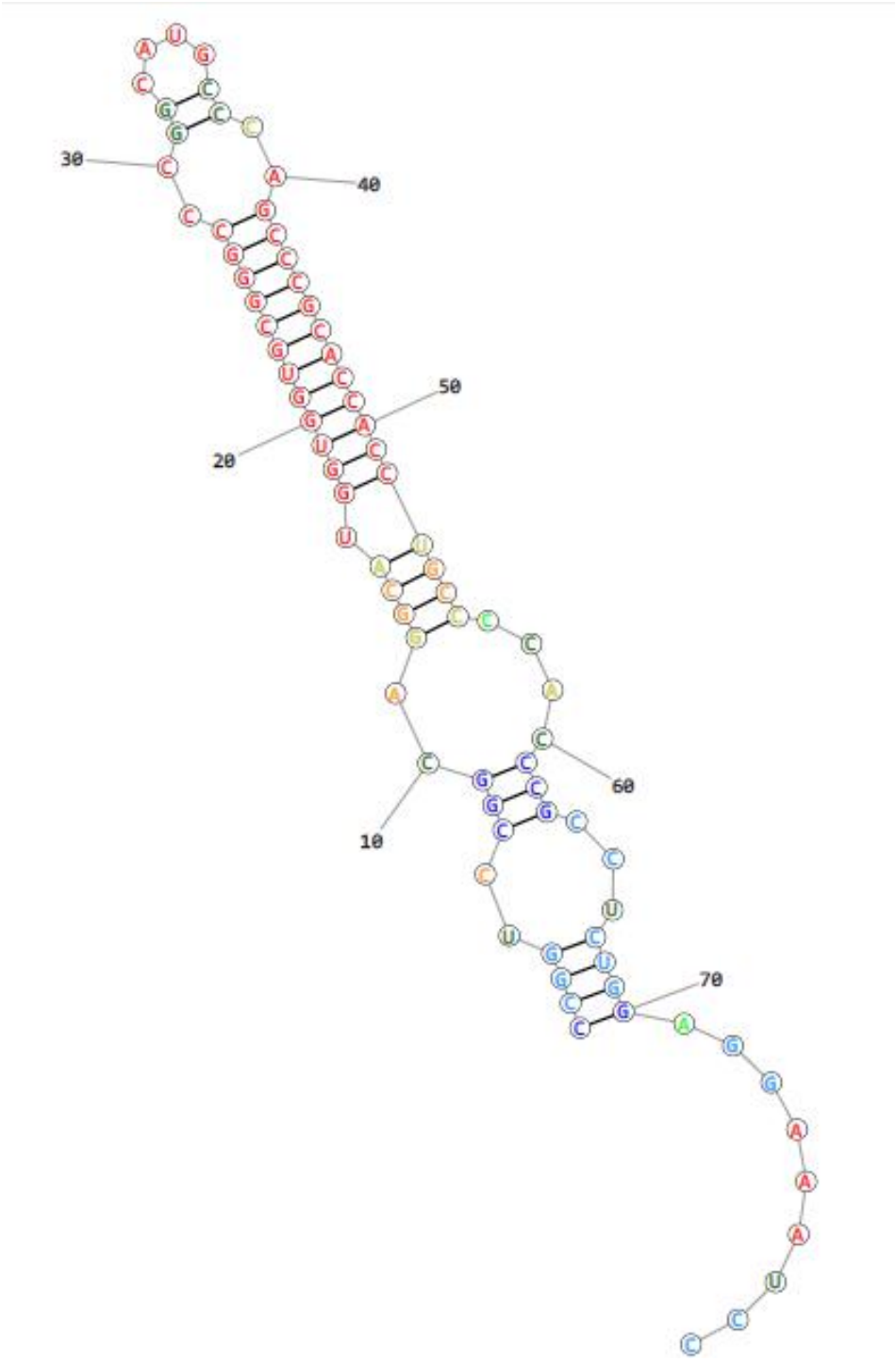
GAUUCCCGUUGCCCGACGUGGAGGACUCCCAG



Streptomyces Atratus OK807:

CCGGUCCGGCAGGCAUGGUGGUGCGGGCCCGGCAUGCCCAGCCCGCACCACCU

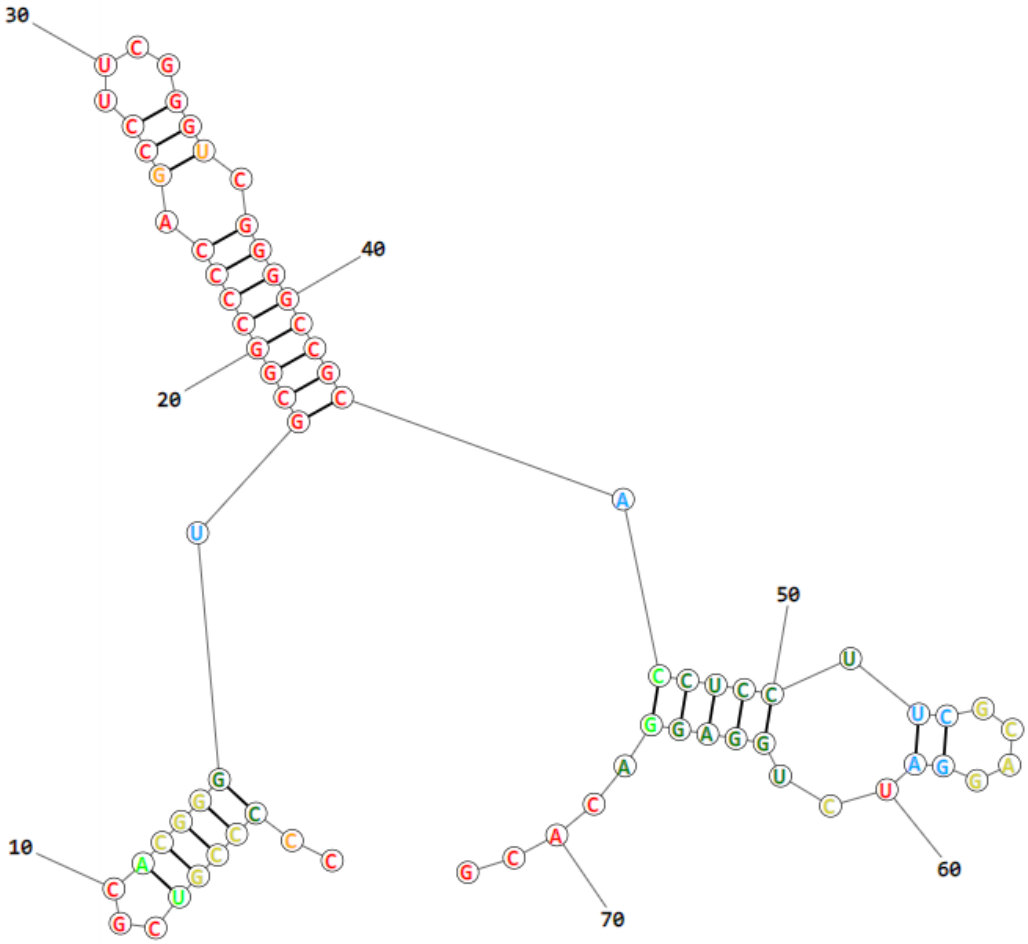
GCCCCACCCGCCUCUGGAGGAAAUCC



SID9727:

CCCCCGUCGCACGGGUGCGGCCCCAGCCUUCGGGUCGGGGCCGCACCUCCUUC

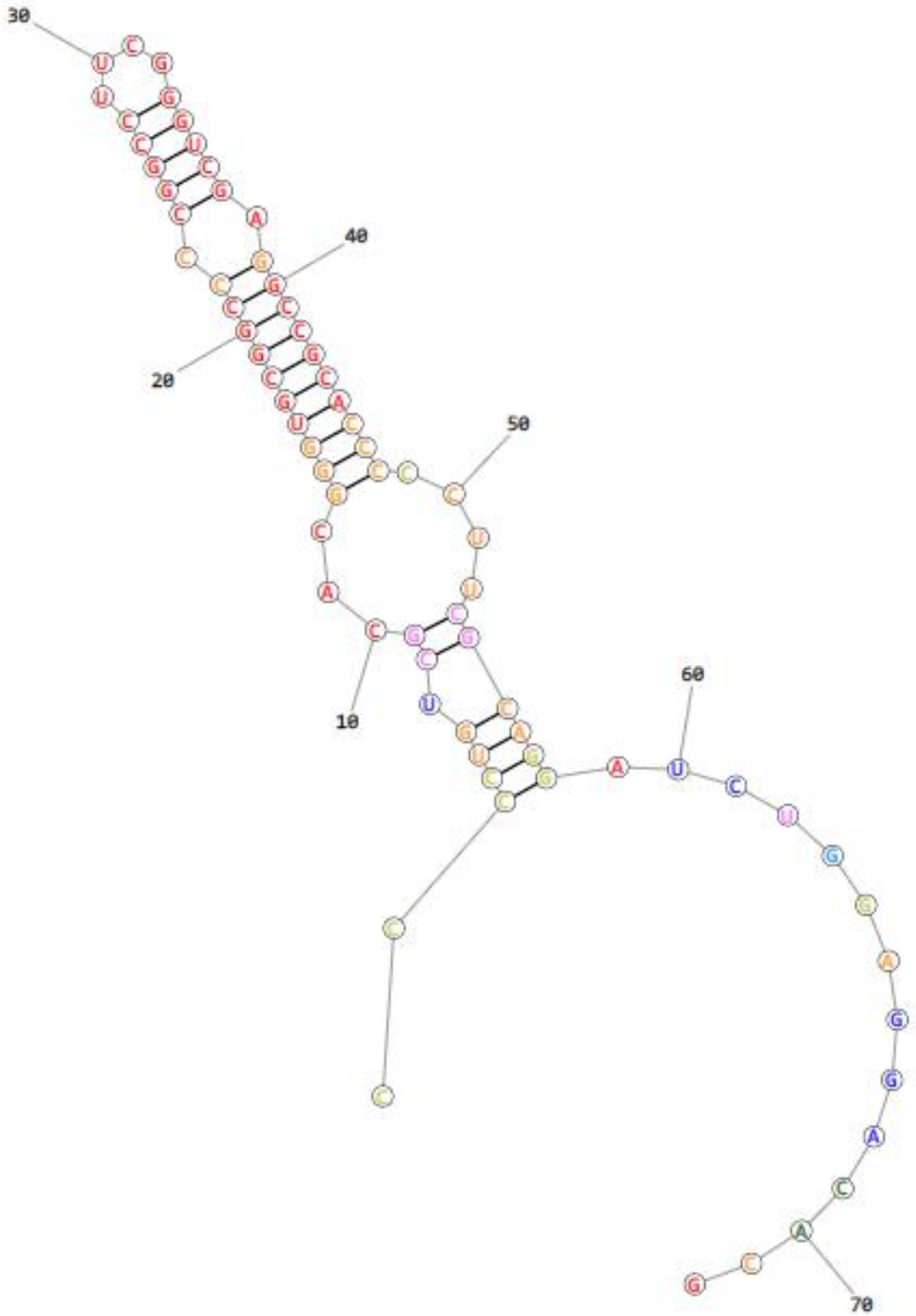
GCAGGAUCUGGAGGACACG



CB01249:

CCCCUGUCGCACGGGUGCGGCCCCGGCCUUCGGGUCGAGGCCGCACCCCCUUC

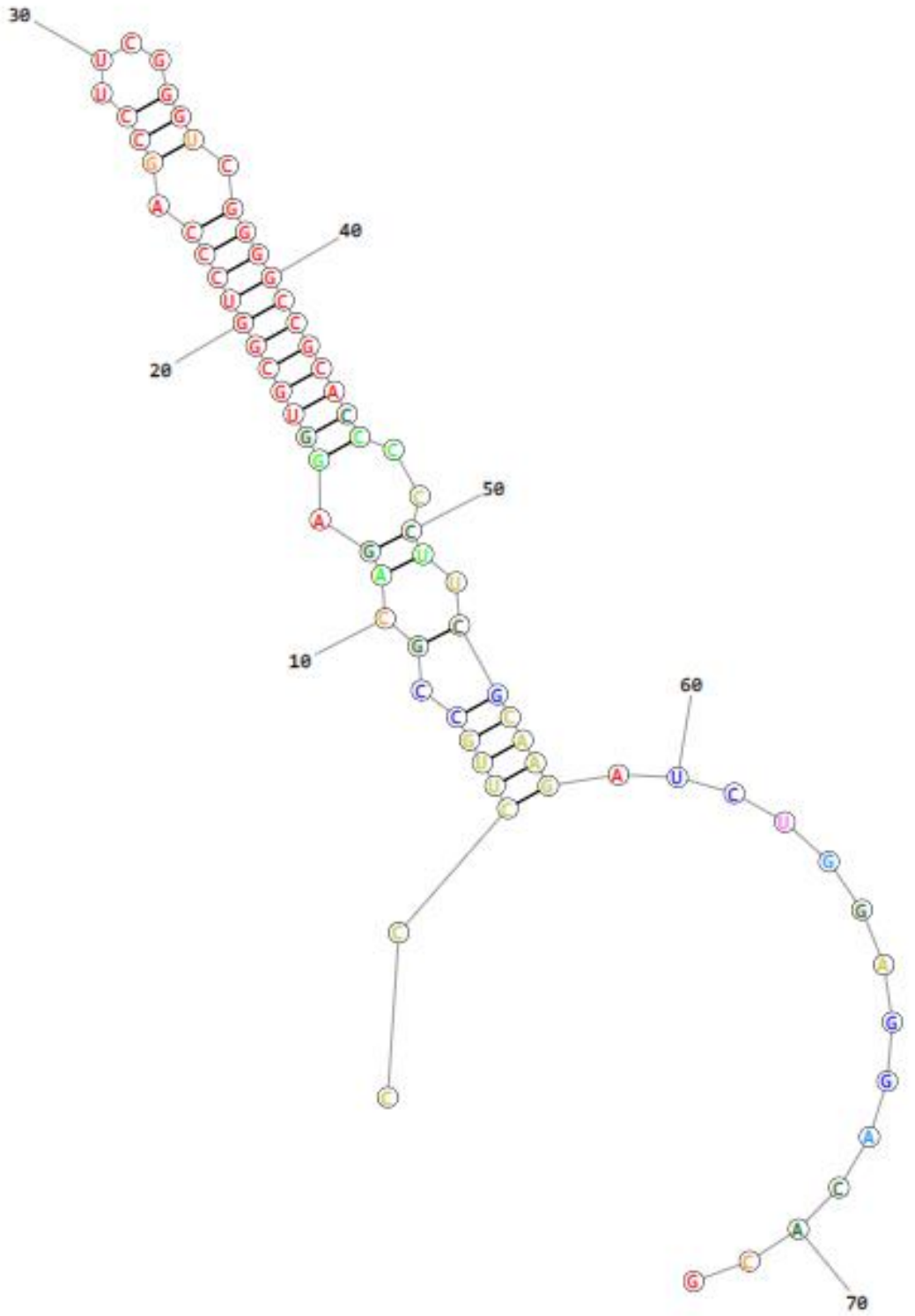
GCAGGAUCUGGAGGACACG



AC1-42T:

CCCUUGCCGCAGAGGUGCGGUCCCAGCCUUCGGGUCGGGGCCGCACCCCCUUC

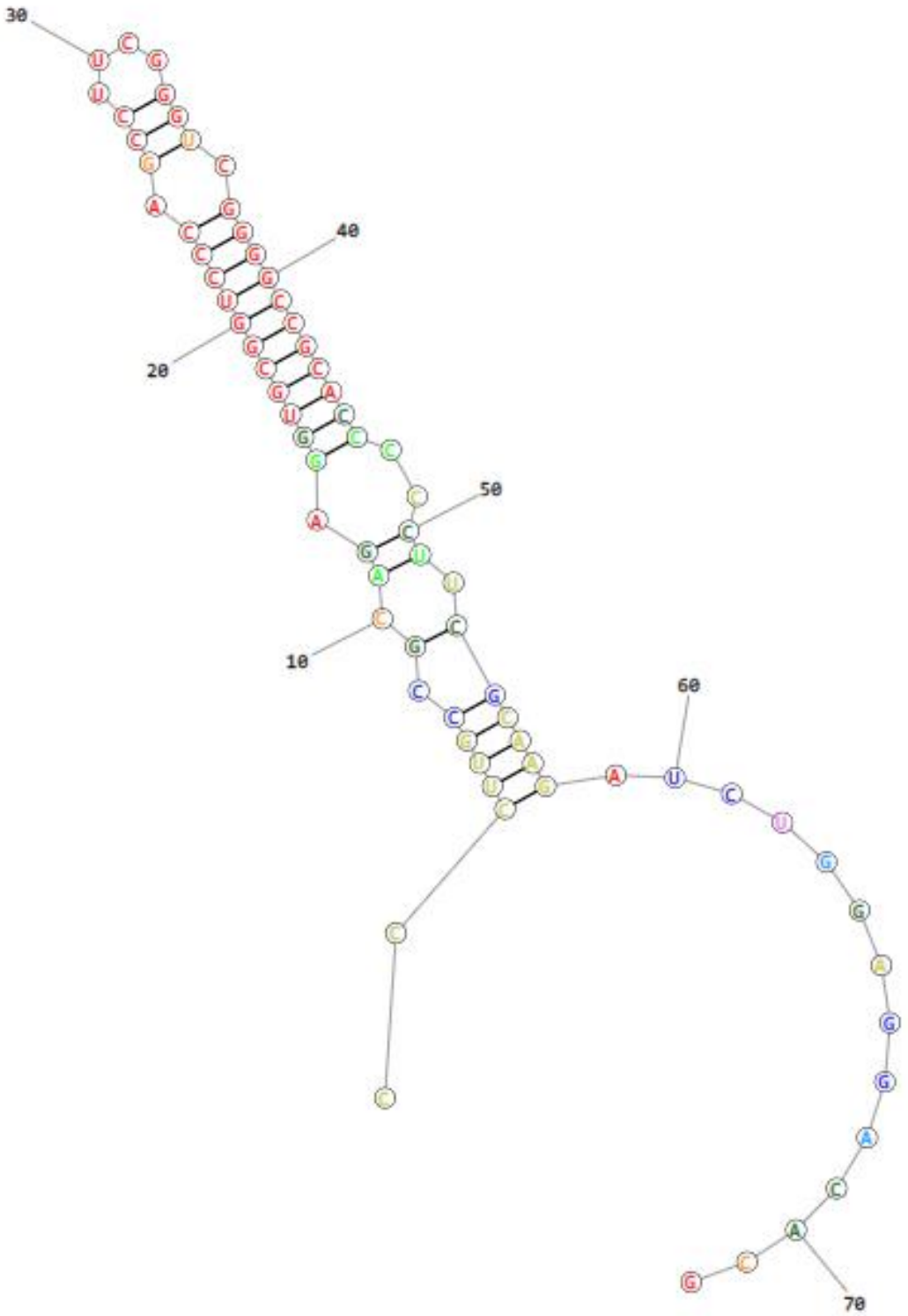
GCAAGAUCUGGAGGACACG



AC1-42W:

CCCUUGCCGCAGAGGUGCGGUCCCAGCCUUCGGGUCGGGGCCGCACCCCCUUC

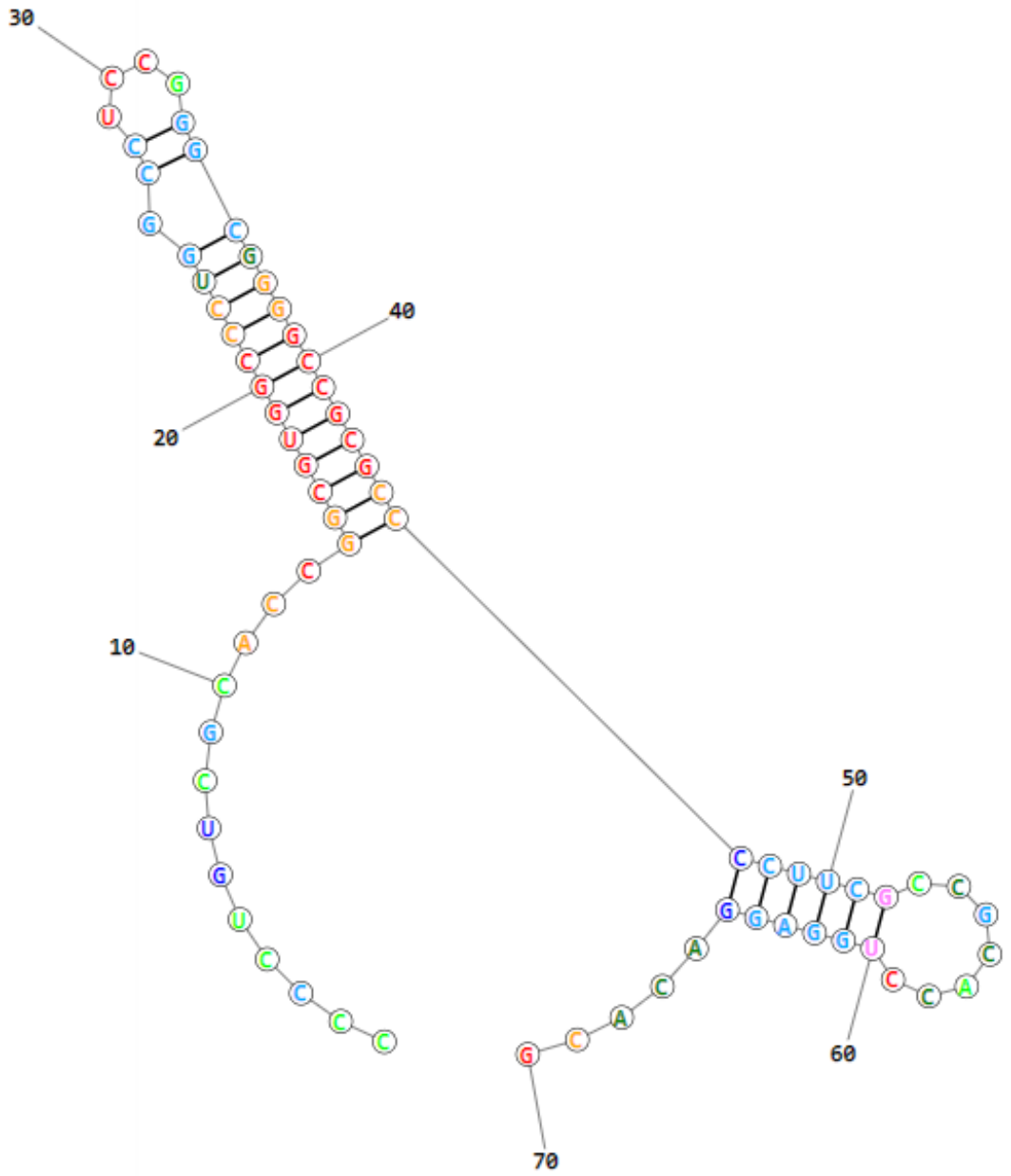
GCAAGAUCUGGAGGACACG



SID9124:

CCCCUGUCGCACCGGCGUGGCCCUGGCCUCCGGGCGGGGCCGCGCCCCUUCGCC

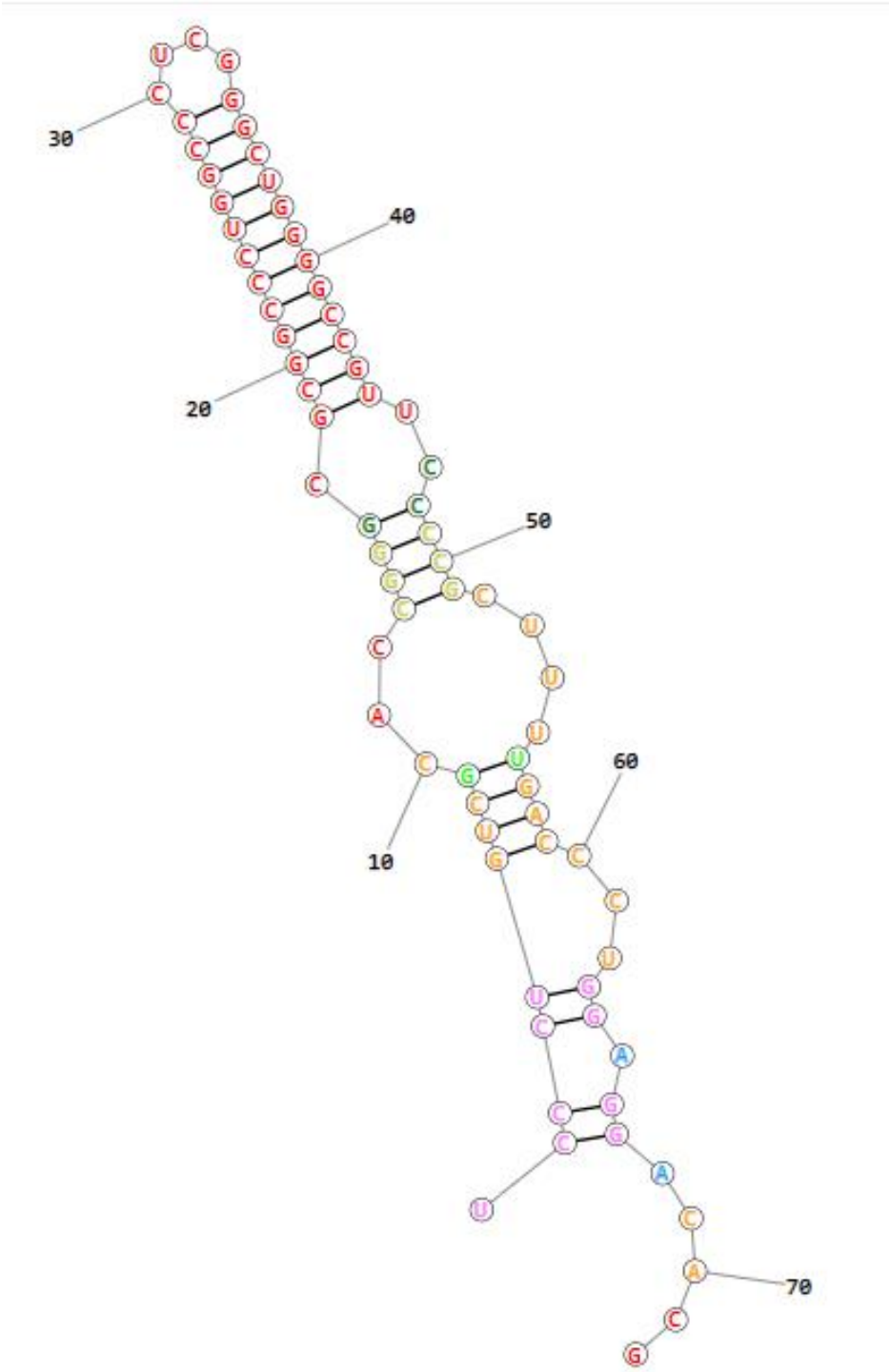
GCACCUGGAGGACACG



SID4913:

UCCCUGUCGCACCGGGCGCGGCCCUGGCCCUCGGGCUGGGGCCGUUCCCCGCU

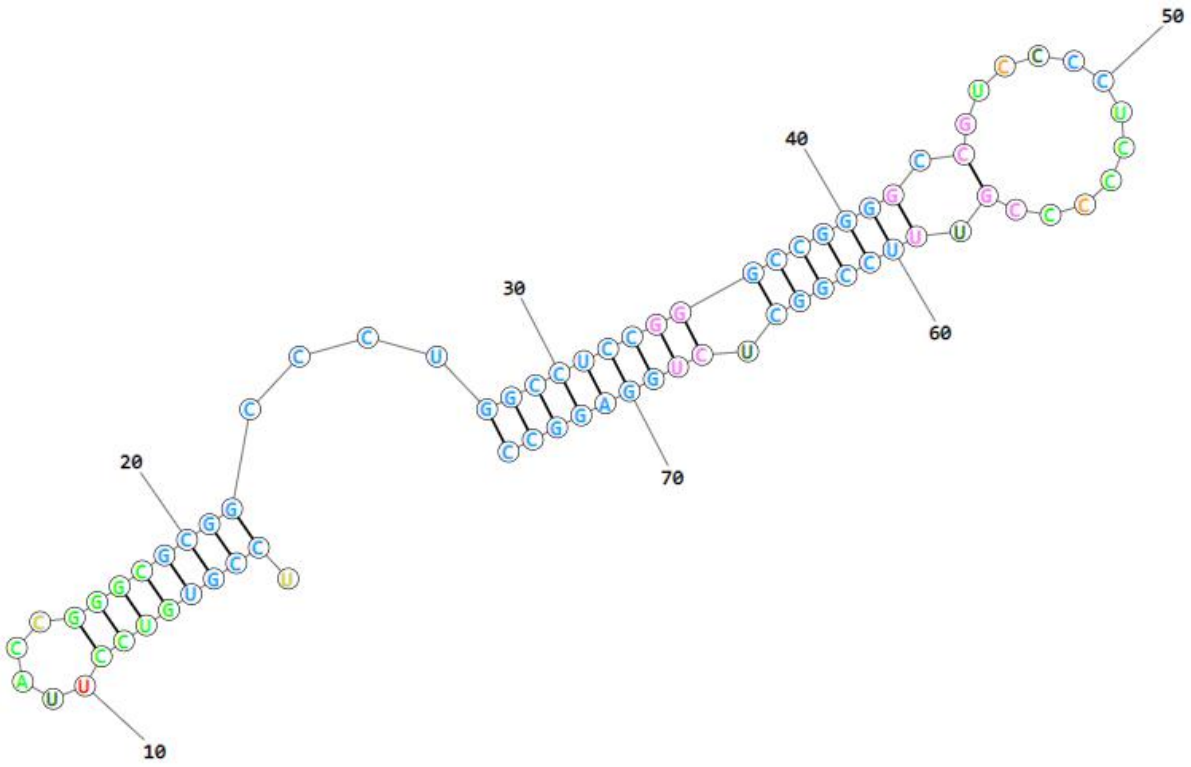
UUUGACCCUGGAGGACACG



CB02460:

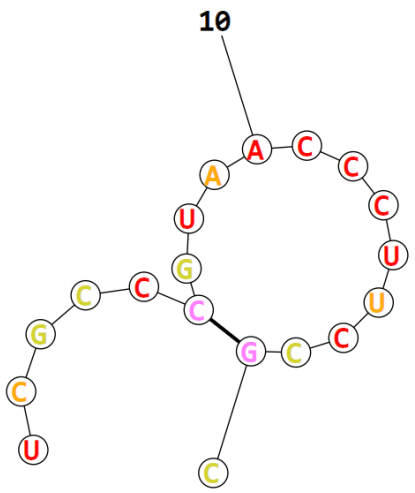
UCCGUGUCCUUACCGGGCGCGGCCCUGGCCUCCGGGCCGGGGCCGUCCCCUCCC

CCGUUUCCGGCUCUGGAGGCC



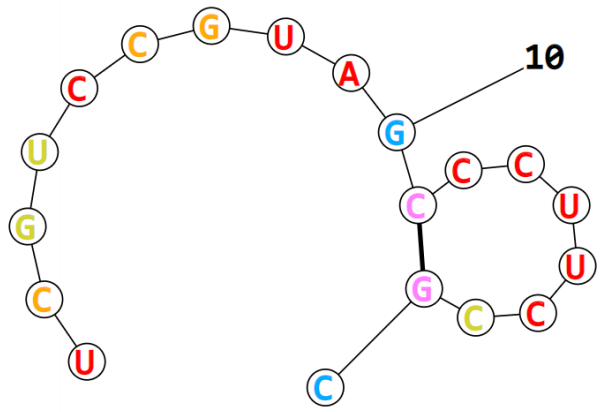
CS081A:

UCGCCCGUAACCCUUCCGC



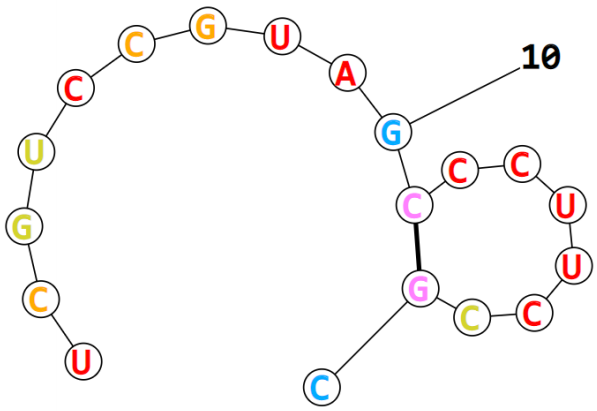
NRRL_F-6491:

UCGUCCGUAGCCCUUCCGC



NRRL_F-6492:

UCGUCCGUAGCCCUUCCGC



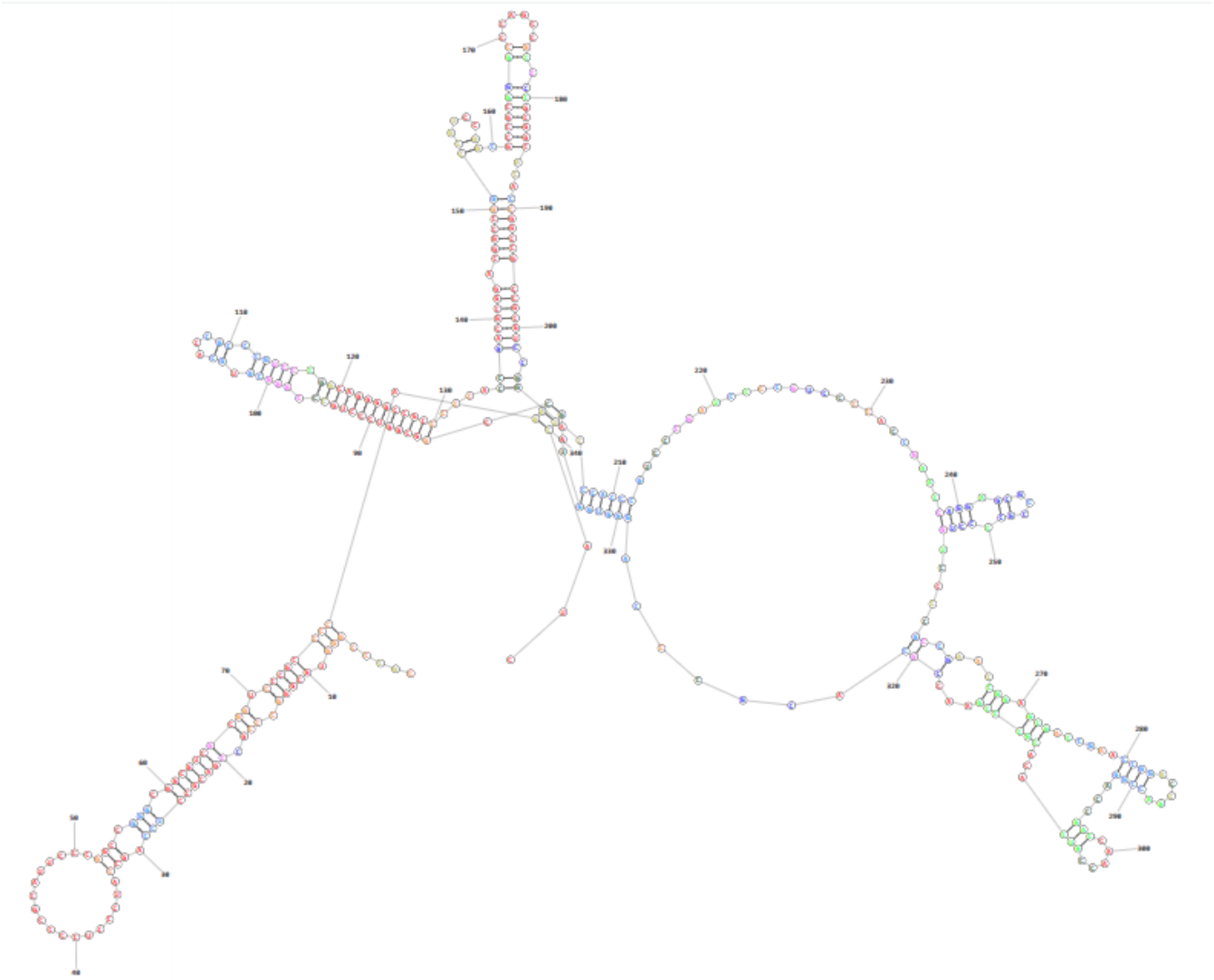
NRRL_B-24572:

CCGCUCGUCC

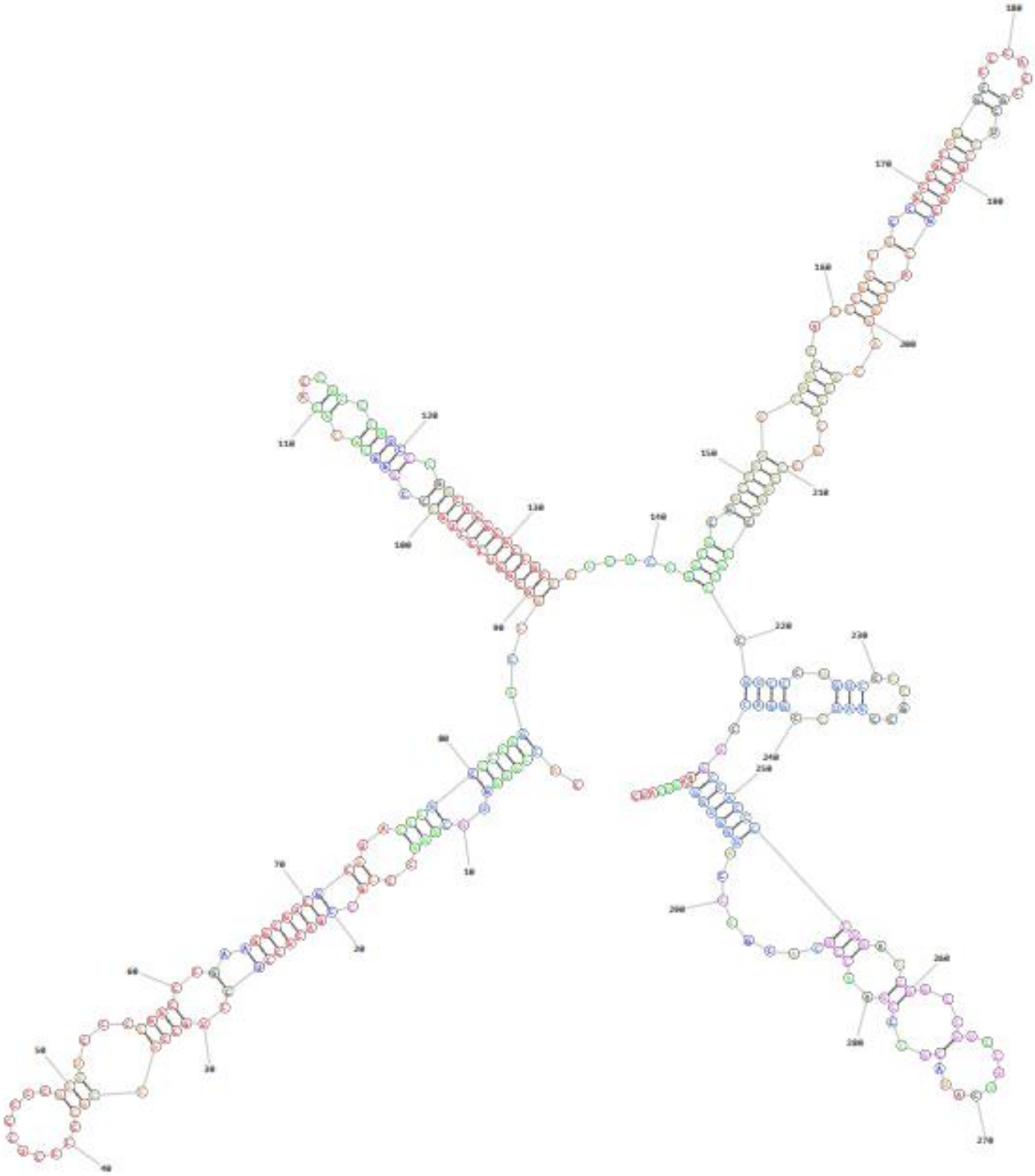
No structure generated.

Moomysin Homolog Cluster Peptide Homologs intergenic regions mRNA secondary structures:

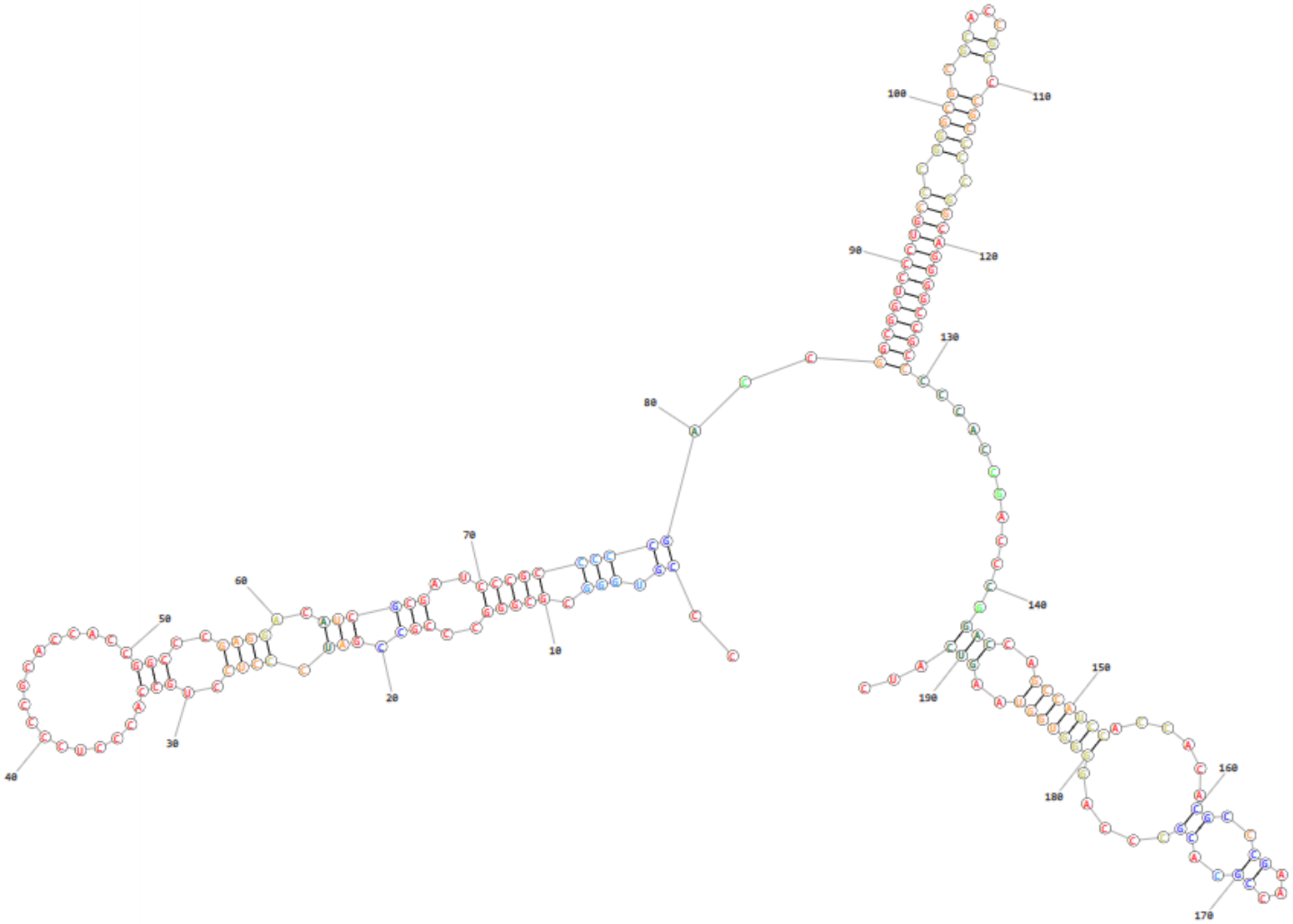
Streptomyces_rimosus_ATCC_10970:



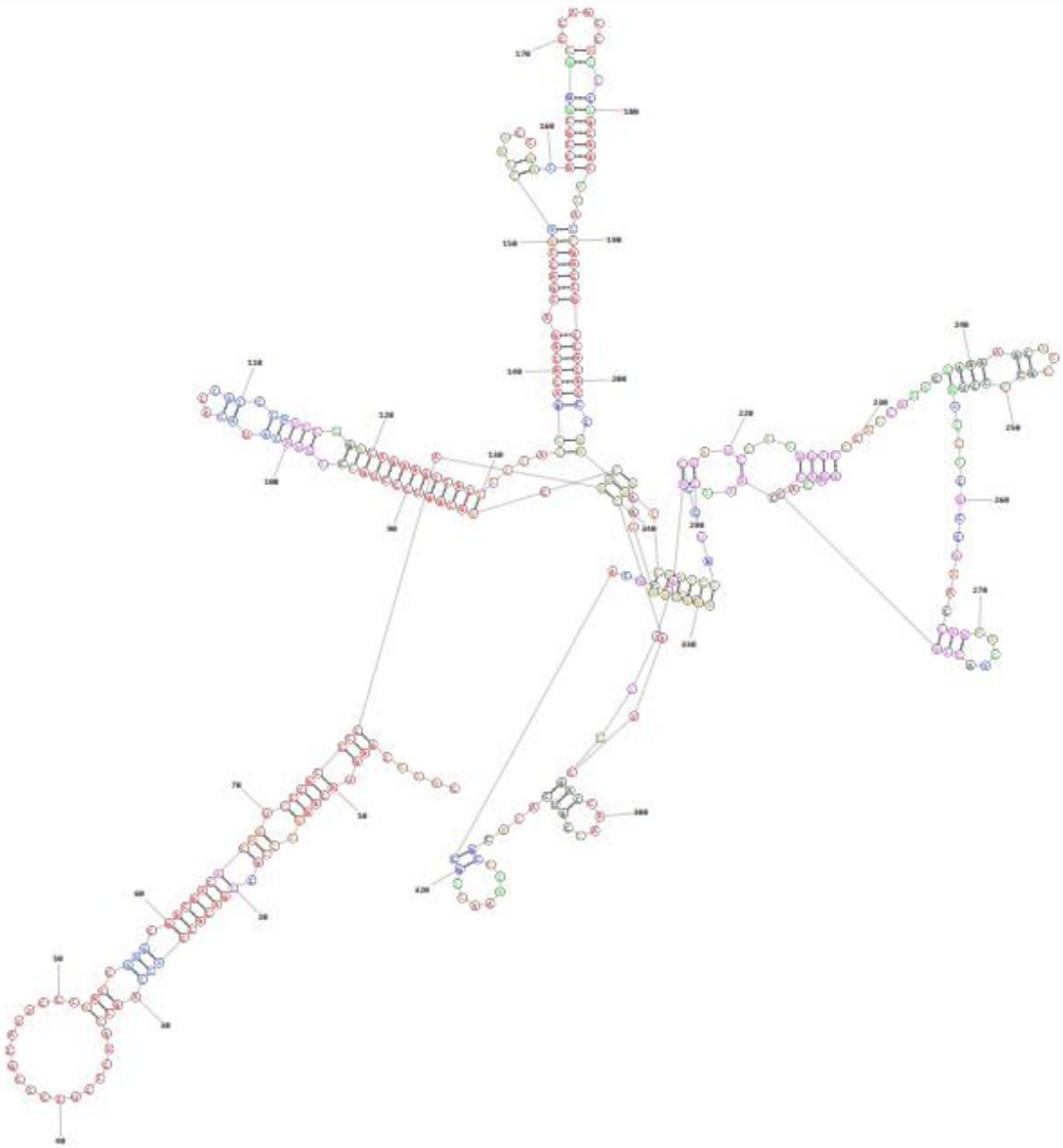
Streptomyces albofaciens:



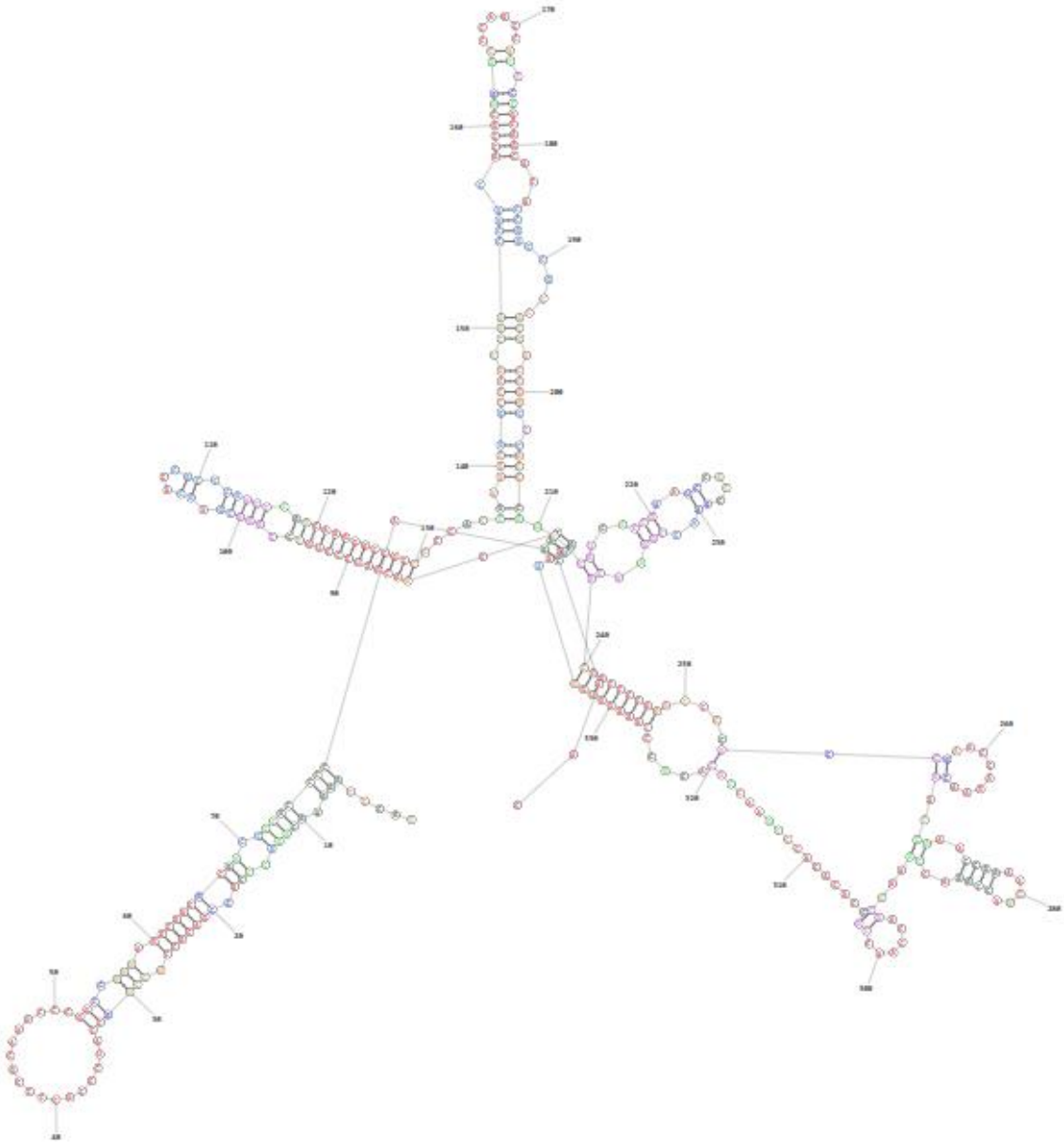
Streptomyces_rimosus_subsp._rimosus_strain_NRRL_WC-3904:



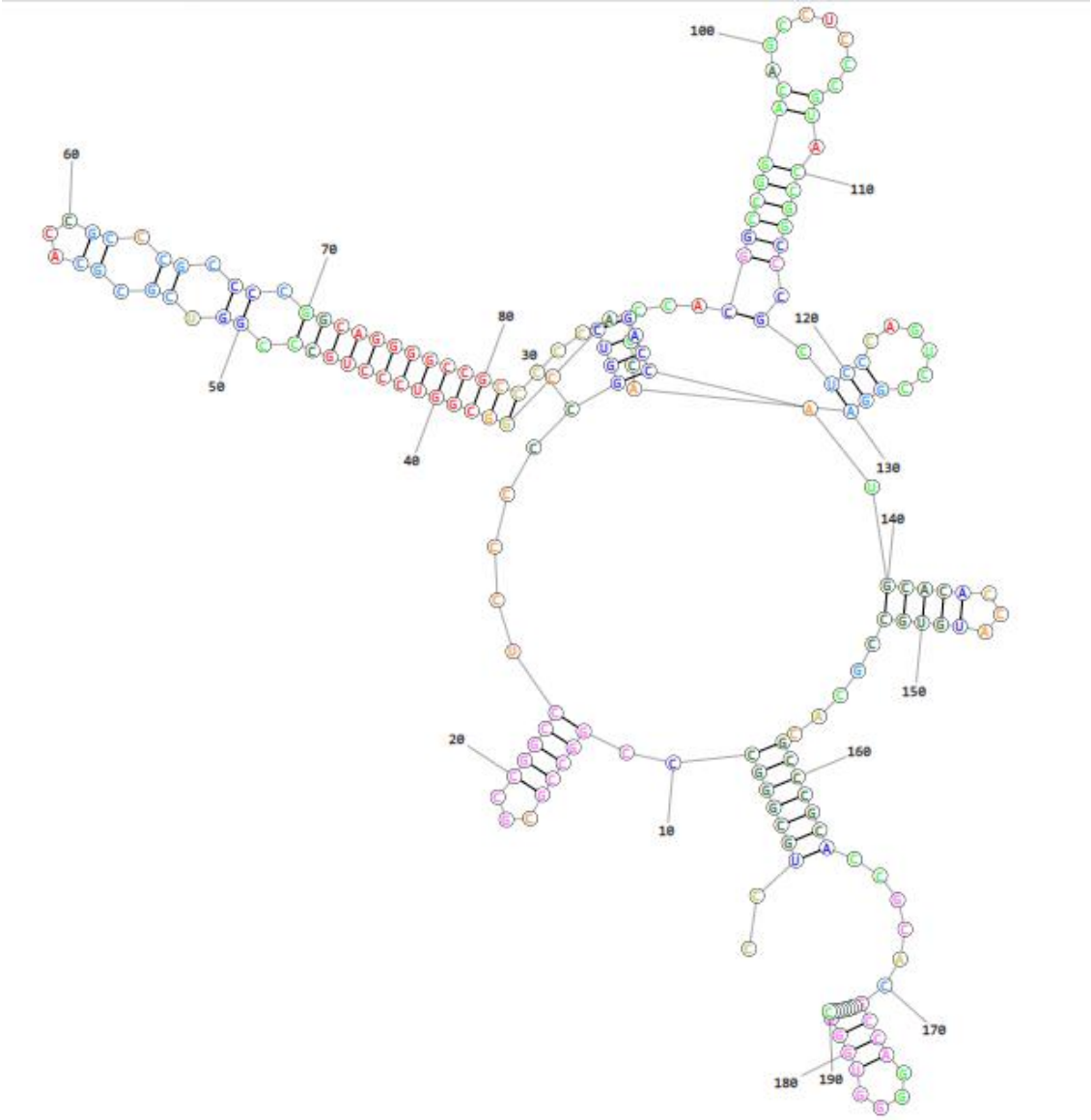
Streptomyces sp. WAC 06783:



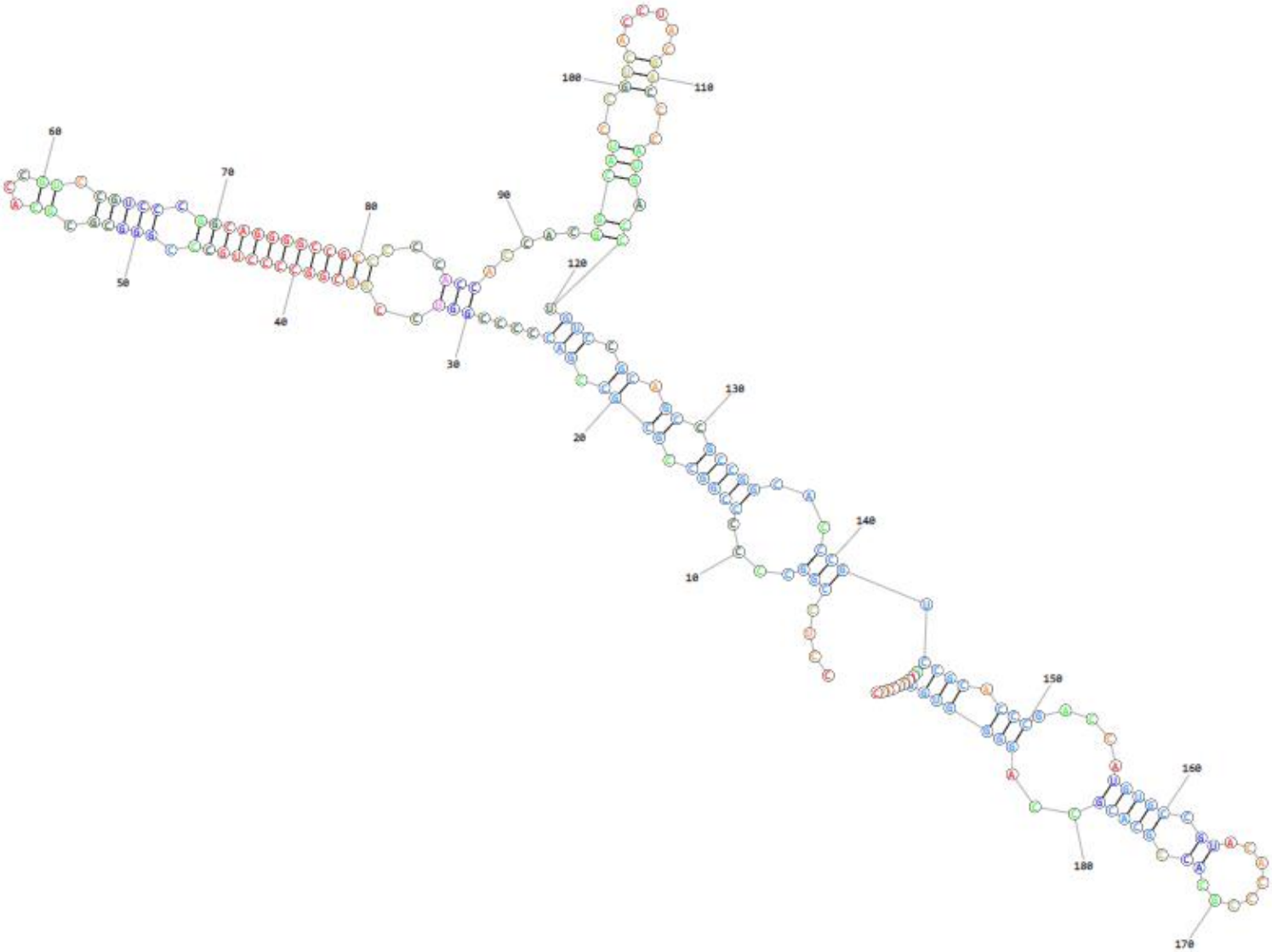
Streptomyces_sp._NRRL_F-5755:



Streptomyces_monomycini_strain_NRRL_B-24309:

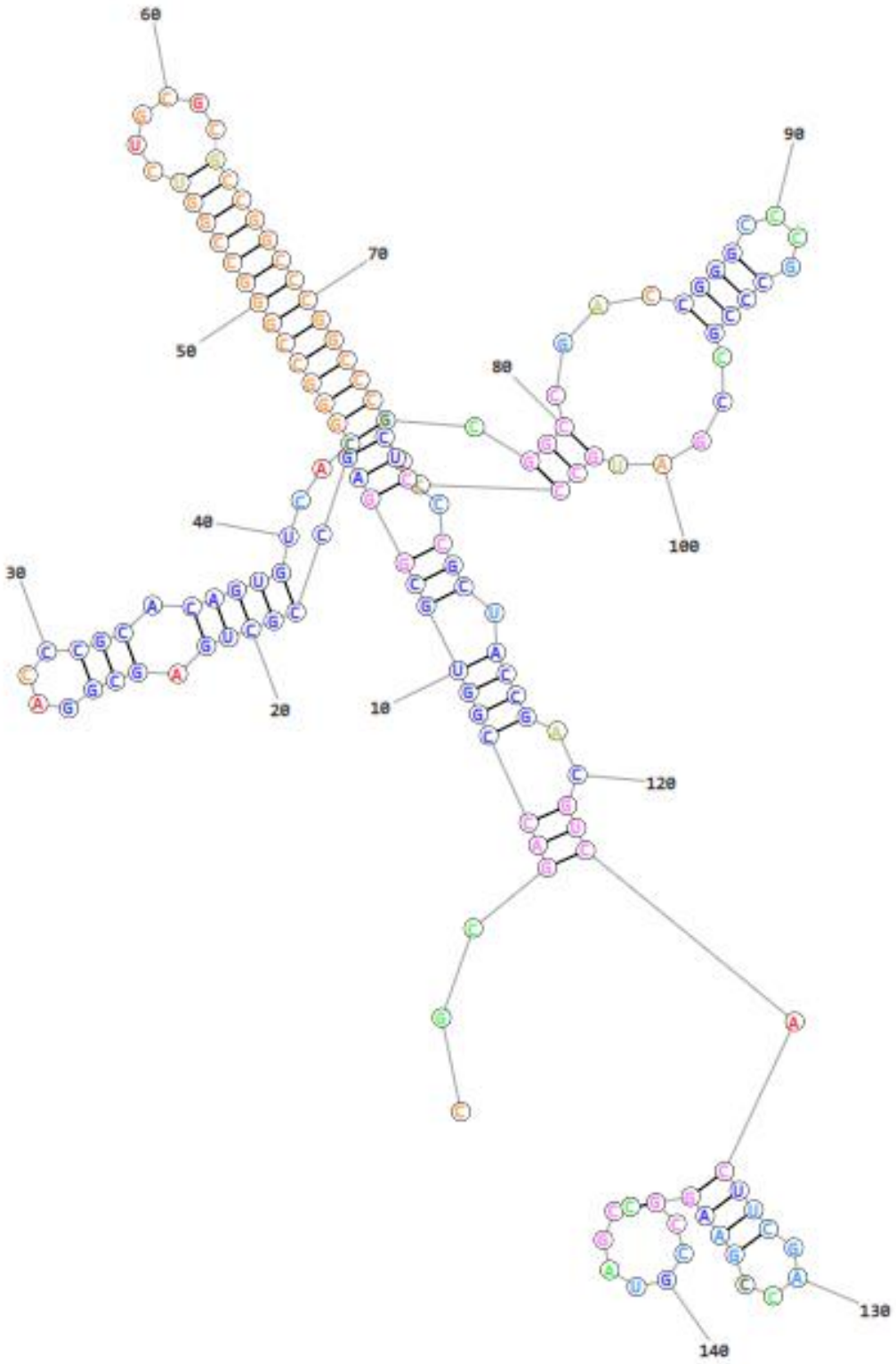


Streptomyces_rimosus_subsp._paromomycinus_strain_NBRC_15454:

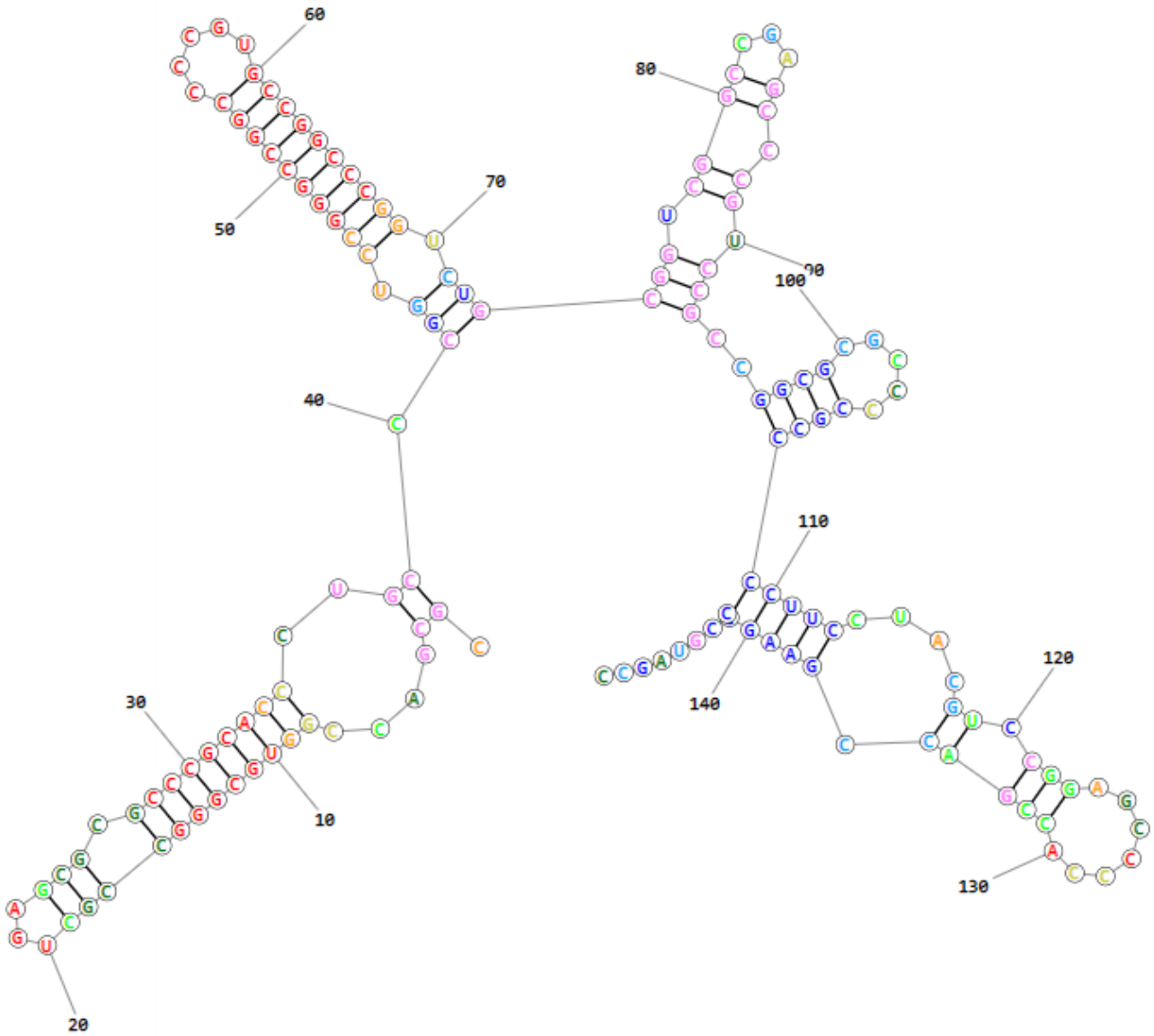


Lagmysin Homolog Cluster Peptide Homologs intergenic regions secondary structures:

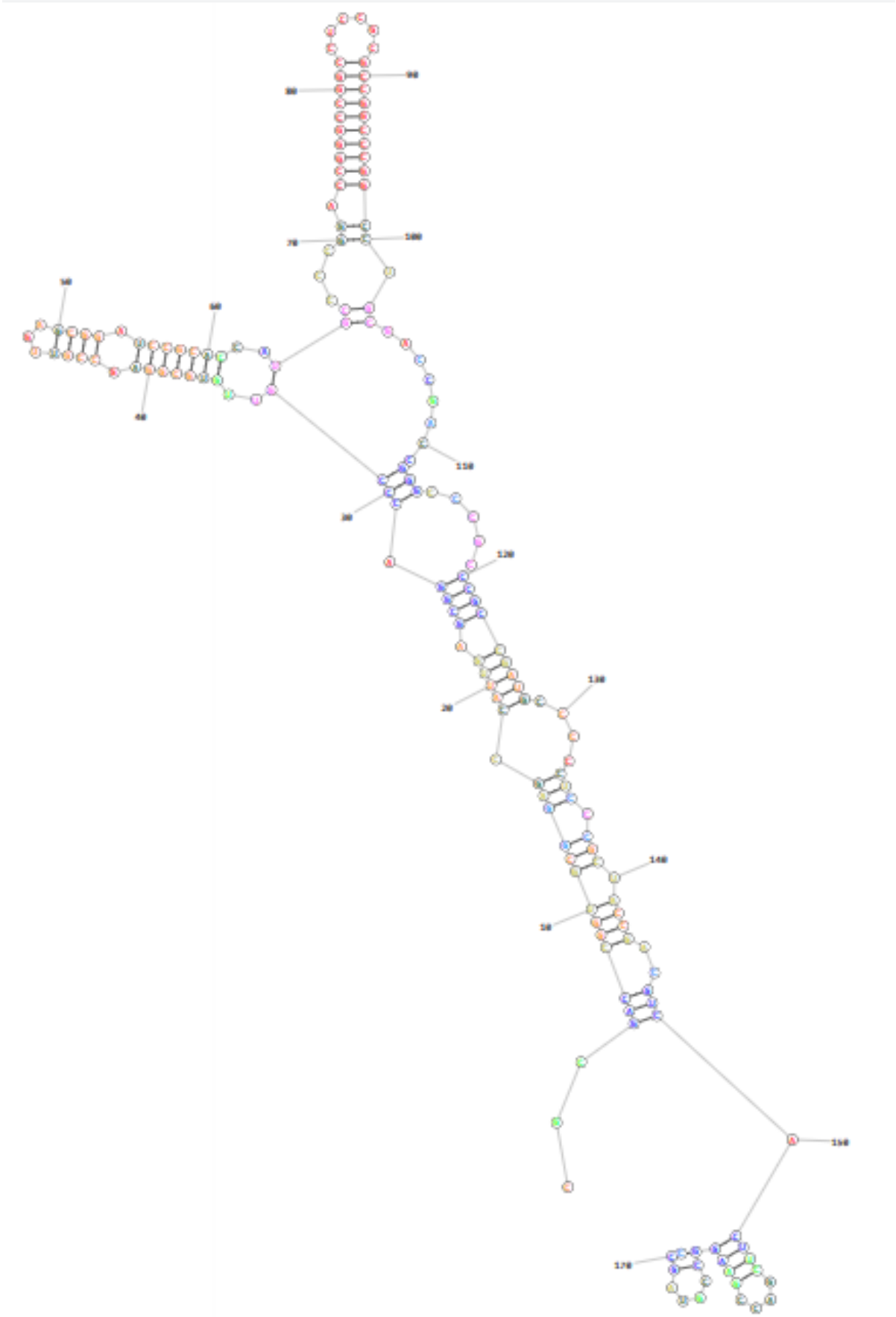
Streptomyces_rimosus_ATCC_10970 (original Lagmysin Homolog):



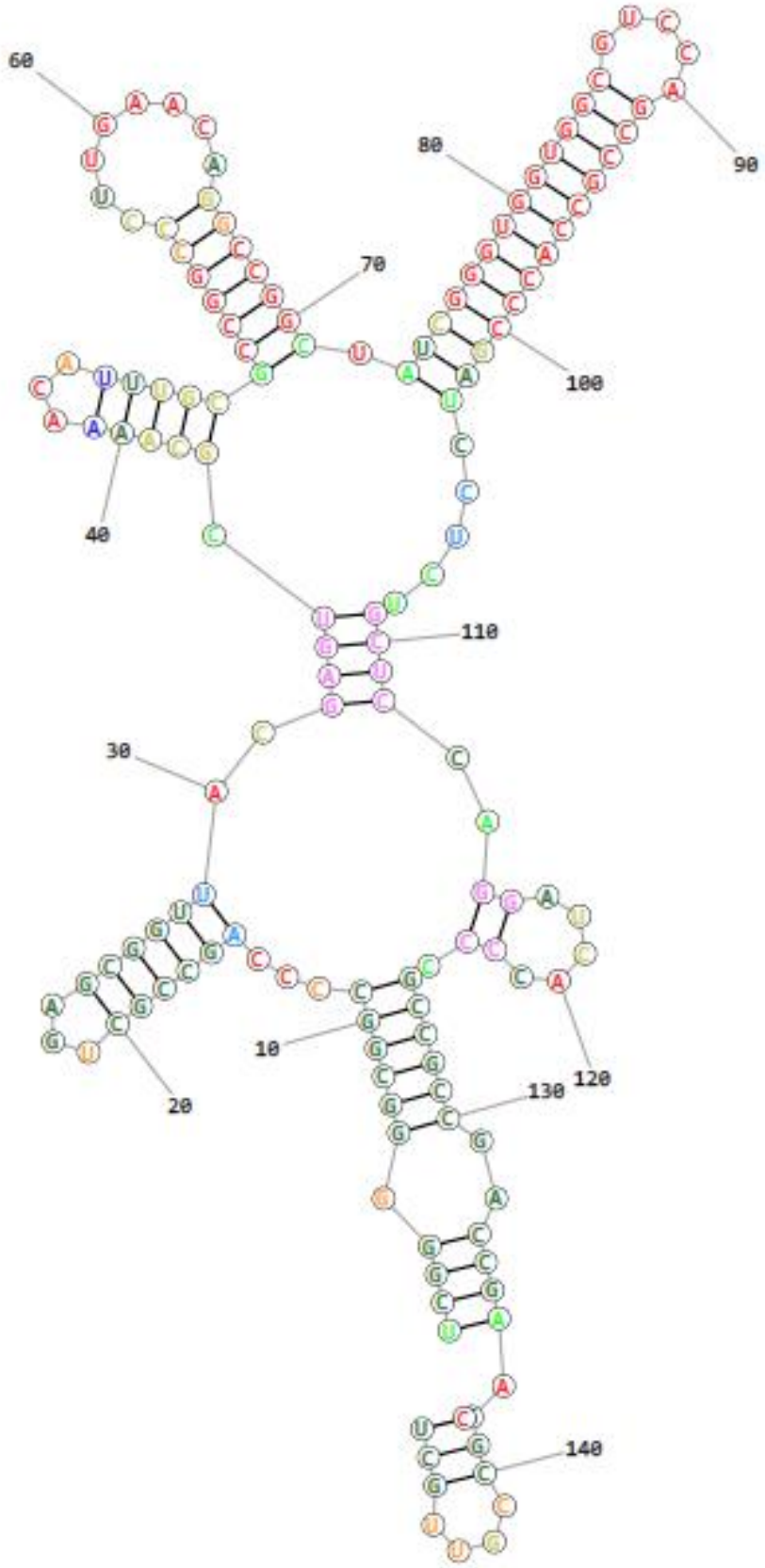
Streptomyces_chrestomyceticus:



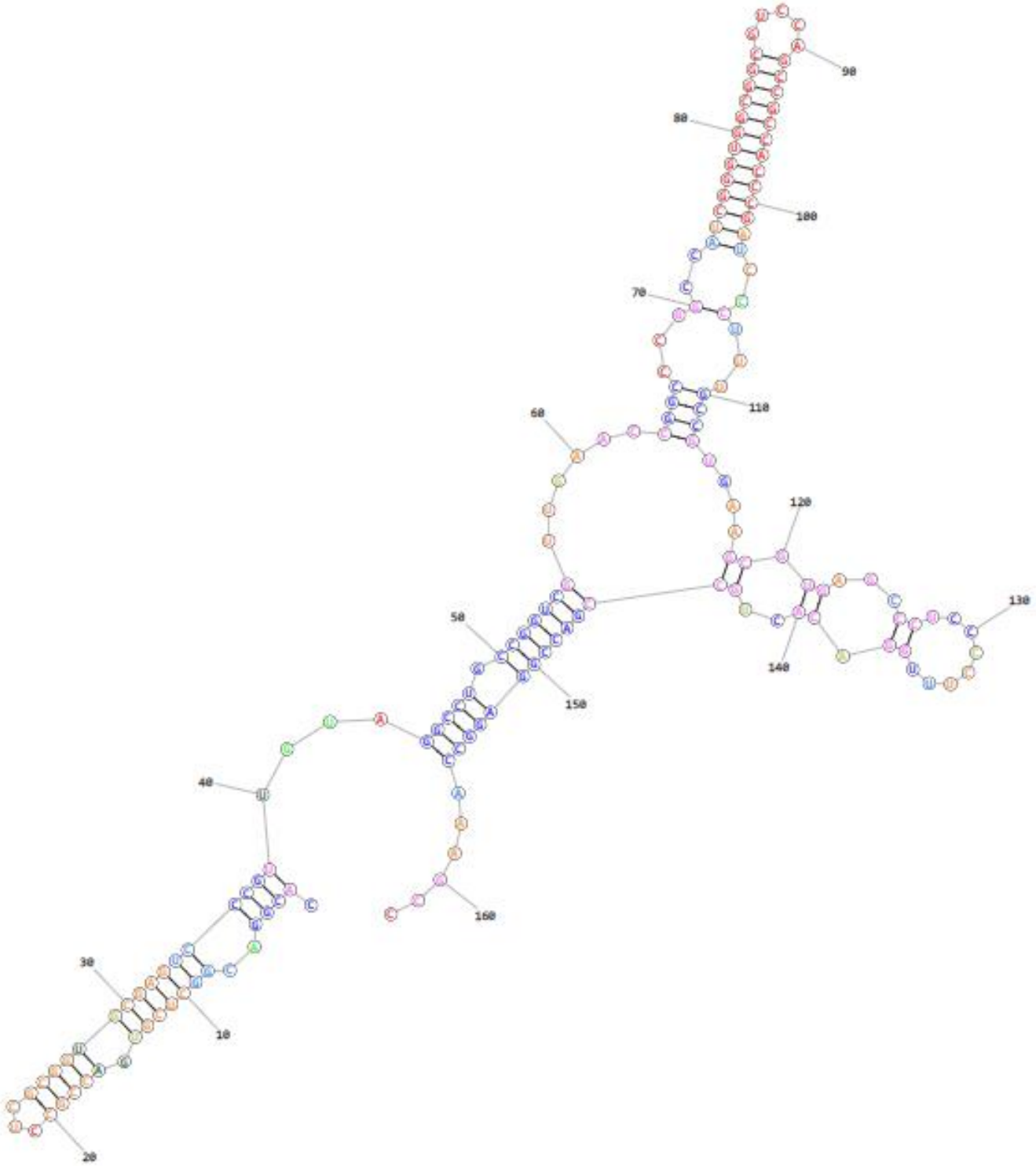
Streptomyces_albofaciens:



Streptomyces_alboniger:



Streptomyces_sp._TLI_146:



Reactions	**Plasmid concentration (nM)**
Individual expression of peptides, his tagged peptides and PTM enzymes separately	
pTU1-A-T7RBS_Albusnodin_H	10
pTU1-A-T7RBS_Lagmysin_H	10
pTU1-A-T7RBS_Moomysin_H	10
pTU1-A-T7RBS_Albusnodin_H_His6	10
pTU1-A-T7RBS_Lagmysin_H_His6	10
pTU1-A-T7RBS_Moomysin_H_His6	10
pSF1C-A-SP44a_pETRBS_Albusnodin_H_PTMs	10
pSF1C-A-SP44a_pETRBS_Lagmysin_H_PTMS	10
pSF1C-A-SP44a_pETRBS_Moomysin_H_PTMs	10
Peptides in combination with PTMs	
pTU1-A-T7RBS_Albusnodin_H + pSF1C-A-SP44a_pETRBS_Albusnodin_H_PTMs	10 of each
pTU1-A-T7RBS_Lagmysin_H + pSF1C-A-SP44a_pETRBS_Lagmysin_H_PTMS	10 of each
pTU1-A-T7RBS_Moomysin_H + pSF1C-A-SP44a_pETRBS_Moomysin_H_PTMs	10 of each
pTU1-A-T7RBS_Albusnodin_H_His6 + pSF1C-A-SP44a_pETRBS_Albusnodin_H_PTMs	10 of each
pTU1-A-T7RBS_Lagmysin_H_His6 + pSF1C-A-SP44a_pETRBS_Lagmysin_H_PTMS	10 of each

pTU1-A-T7RBS_Moomysin_H_His6 + pSF1C-A-SP44a_pETRBS_Moomysin_H_PTMs	10 of each
Positive controls	
T7-BovA	10
T7- BovT150M	10
T7-BovA + T7- BovT150M	10 of each
T7 eGFP	N/A (volume = 8.25ul)
T7 sfGFP	N/A (volume = 8.25ul)
pTU1 sp44 mVenus	10
NC	
sterile ddH20	N/A (volume = 8.25ul)

Supplementary table 2 – RiPP cell-free plasmid combinations and concentrations