Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2018-08-01

Exploration of Antimicrobial Activity in Natural Peptides and High-Throughput Discovery of Synthetic Peptides

Emma Kay Dallon
Brigham Young University

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Life Sciences Commons

BYU ScholarsArchive Citation

Dallon, Emma Kay, "Exploration of Antimicrobial Activity in Natural Peptides and High-Throughput Discovery of Synthetic Peptides" (2018). *All Theses and Dissertations*. 7468.
https://scholarsarchive.byu.edu/etd/7468

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Exploration of Antimicrobial Activity in Natural Peptides and High-Throughput Discovery of Synthetic Peptides

Emma Kay Dallon

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

Master of Science

Joel S. Griffitts, Chair
William Russell McCleary
Brent Nielsen

Department of Microbiology and Molecular Biology

Brigham Young University

Copyright © 2018 Emma Kay Dallon

All Rights Reserved

ABSTRACT

Exploration of Antimicrobial Activity in Natural Peptides and High-Throughput Discovery of Synthetic Peptides

Emma Kay Dallon
Department of Microbiology and Molecular Biology, BYU
Master of Science

Despite many medical advances, antibiotic resistant bacteria increasingly plague the modern world, necessitating discovery of new antibiotics. One area of nature that can provide inspiration for antibiotics is antimicrobial peptides. Many of these peptides exist in nature, with some classes that have not been studied or characterized well. One such class is the defensin-like peptides generated by the plant *Medicago truncatula* as part of their symbiotic relationship with *Sinorhizobium meliloti*. Nodule-specific Cysteine Rich (NCR) peptides are defined by the presence of multiple cysteines, and regulate the growth of *S. meliloti* within plant cells. While some of these NCR peptides have been shown to have antimicrobial properties, hundreds of peptides remain uncharacterized. We have developed an assay for further characterization of these peptides in *E. coli*. Of the seven peptides that have been tested using this assay, three have exhibited antimicrobial properties when expressed in *E. coli*. Additionally, we have developed a system for discovering novel antimicrobial peptides. This platform, called PepSeq, uses the expression of random peptides in *E. coli* combined with deep sequencing to detect antimicrobial activity. This technology is capable of screening through millions of peptide molecules simultaneously. Using this platform, we have discovered and confirmed six novel antimicrobial peptides, with hundreds of additional predicted antimicrobial peptides. The PepSeq platform can be adapted in a number of ways to generate unique peptide scaffolds to discover more potent antimicrobial peptides.

Keywords: antimicrobial peptide, NCR peptide

ACKNOWLEDGMENTS

I would like to thank the members of my committee, Dr. Joel Griffitts, Dr. Bill McCleary, and Dr. Brent Nielsen for their guidance and direction. I would like to particularly thank Dr. Griffitts for all that I learned while working in his lab. I would also like to thank the other members of the lab for their advice and friendship.

TABLE OF CONTENTS

TITLE PAGE i

ABSTRACT ii

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS iv

LIST OF FIGURES x

LIST OF TABLES xii

CHAPTER 1: Antimicrobial peptides designed by nature and scientists 1

1.1 An introduction to antimicrobial peptides 1

1.2 RiPPs: enzymatically modified peptides resulting in complex structures 3

1.3 Cysteine-rich peptides: secondary structure using disulfide bonds 5

1.3.1 Defensins and conotoxins 5

1.3.2 Peptides involved in symbiosis 7

1.4 Strategies for creating synthetic peptides 10

1.4.1 Selective mutation 10

1.5 Computational methods for AMP discovery 11

1.5.1 AMP libraries for high-throughput screening 13

1.6 Summary 15

1.6.1 Promoter optimization 16

1.6.2 Antimicrobial activity of NCR peptides in *E. coli* 16

1.6.3 Massively parallel peptide activity screen (PepSeq) 17

CHAPTER 2: Promoter optimization in *E. coli* 18

2.1 Background 18

2.2 Methods 21

2.2.1 Cloning native P_{ara} and P_{rha} promoters 21

2.2.2 Cloning P_{tet} and P_{lacT5} promoters 22

2.2.3 Cloning P_{rha} with extended ribosome binding site, P_{ara} from pGLO, and P_{lacT5} from pSX2 22

2.2.4 Fluorescence assays 22

2.2.5. Growth curves 22

2.2.6 Miller assays 23

2.3 Results 23

2.3.1 Comparison of P_{rha}, P_{ara}, P_{lacT5} and P_{tet} 23

2.3.2 Comparison of P_{rha}, P_{ara}, and P_{lacT5} in MG1655 25

2.3.3 Carbon sources in minimal media 25

2.4 Discussion 27

CHAPTER 3: Antimicrobial activity of nodule-specific cysteine rich peptides in *E. coli* 29

3.1 Background 29

3.2 Methods 33

3.2.1 Molecular Cloning Protocols 33

3.2.2 Analysis of antimicrobial activity 34
3.2.3 Osmotic shock and PhoA assays 34
3.3 Results 34
3.3.1 Verification of YiaD membrane tethering 34
3.3.2 Antimicrobial activity of NCR peptides 36
3.4 Discussion 38
CHAPTER 4: Massively parallel peptide activity screening 39
4.1 Background 39
4.2 Methods 41
4.2.1 Peptide library cloning 41
4.2.2 Library selection 43
4.2.3 Preparing samples for Illumina sequencing 43
4.2.4 Read counting, quality control, and motif finding 43
4.2.5 Individual toxic peptide cloning 43
4.2.6 Toxic peptide confirmation 44
4.3 Results 44
4.3.1 Library and experimental design 44
4.3.2 PepSeq data quality control 47
4.3.3 Results 47
4.3.4 Peptide confirmation 51

4.3.5 Motif results ... 51

4.3.6 Motif confirmation ... 54

4.4 Discussion ... 61

CHAPTER 5: Future applications and adaptations for PepSeq ... 62

5.1 PepSeq Adaptions ... 62

5.2 Antibiotic adjuvants ... 62

5.3 Cyclic peptide scaffolds ... 63

5.3.1 MP1 and MccJ25 as peptide scaffolds ... 63

5.3.2 Unnatural amino acids to cyclize peptides ... 65

5.3.3 Disulfide bonds and peptide export ... 67

5.4 Conclusion ... 67

BIBLIOGRAPHY ... 69

APPENDIX I ... 79

Promoter primer, insert, and plasmid sequences ... 79

P_{ara} from MG1655 ... 79

P_{ara} from pGLO ... 79

P_{rha} from MG1655 ... 79

P_{rha} with XRBS addition ... 80

P_{tet} ... 81

P_{lacT5} (*lacI*WT) ... 81

P_{lacT5} (*lacI*q) 81
Parent vector sequences 82
NCR peptide primer, insert, and plasmid sequences 85
Insert primers 85
Parent vectors 87
Illumina library primer sequences 90
Peptide library primers 90
Illumina primers 90
PepSeq confirmation primer and plasmid sequences 91
Insert primers 91
Parent vector 92
APPENDIX II 94
Miller Assays 94
phoA Assays 94
Molecular cloning 95
Preparing vector 95
Preparing insert 95
Digestion with restriction enzymes 96
Ligating vector and insert together 97
Transforming plasmid into bacterial cells 97

Library cloning and PepSeq 98
Griffitts Lab Homemade Midiprep 98
Library insert synthesis 99
Vector and insert digests 99
Ligation 100
Transformation 100
Library selection 101
PCR-based preparation of Illumina libraries 101
LB and minimal media (M9M) 102
APPENDIX III 104
Top toxic peptides 104
Motif results 115
3mer, 0 wildcards 115
4mer, 1 wildcard 116
5mer, 1 wildcard 117
5mer, 2 wildcards 118
6mer, 2 wildcards 119

LIST OF FIGURES

Fig. 1. Antibiotic resistance timeline. 2

Fig. 2. Non-ribosomally and ribosomally-synthesized peptide structures. 4

Fig. 3. Ribosomally-synthesized, post-translationally modified peptides. 4

Fig. 4. Three-dimensional structures of cysteine-rich peptides 6

Fig. 5. Symbiosis of *M. truncatula* and *S. meliloti*. 9

Fig. 6. Strategies for high-throughput antimicrobial activity assays. 14

Fig. 7. Inducible promoters. 19

Fig. 8. *lacI* promoter layout. 24

Fig. 9. Miller assays in DH5α. 24

Fig. 10. XRBS data summary. 26

Fig. 11. Glycerol versus glucose as the carbon source in M9M. 26

Fig. 12. Vector maps. 28

Fig. 13. Nodulation in *M. truncatula*. 30

Fig. 14. NCR peptide antimicrobial activity assay. 32

Fig. 15. PhoA activity assay. 35

Fig. 16. NCR peptide antimicrobial activity assay. 37

Fig. 17. PepSeq workflow. 40

Fig. 18. PepSeq methods. 42

Fig. 19. Peptide expression system and library design. 45

Fig. 20. Library growth and peptide expression in M9M. 45

Fig. 21. PepSeq dataset quality control. 46

Fig. 22. PepSeq data summary. 50

Fig. 23. Peptide validation growth curves and fluorescence over time. 53

Fig. 24. Motif finder workflow. 55

Fig. 25. YII motif growth and fluorescence curves. 58

Fig. 26. YFXYF motif summary.. 59

Fig. 27. YFXYF growth and fluorescence curves. 60

Fig. 28. Proposed peptide structural scaffolds. 64

Fig. 29. Cysteine converted to thiazole.. 66

Fig. 30. Cysteine-reactive chemistry forms a cyclopeptide. 66

LIST OF TABLES

Table 1. Antimicrobial peptides.. 2

Table 2. NCR peptide sequences and functions... 9

Table 3. NCR mature peptide sequences 30

Table 4. NCR peptides tested.. 35

Table 5. Peptides from PepSeq dataset 48

Table 6. Toxic peptides identified by PepSeq 49

Table 7. Top six toxic peptides 52

Table 8. Five additional toxic peptides 52

Table 9. Top five motifs of various kmer lengths.. 56

Table 10. YII-containing motifs 57

Table 11. YII-containing peptides for confirmation 58

Table 12. Potential degenerate codons... 64

CHAPTER 1: Antimicrobial peptides designed by nature and scientists

1.1 An introduction to antimicrobial peptides

Despite many medical advances, antibiotic resistant bacteria present a growing threat to human health. Antibiotics cure bacterial infections and have saved hundreds of millions of lives over the last century; however, the relentless emergence of bacterial resistance to antibiotics threatens their effectiveness, in many cases rendering them useless [1-5]. The rate of antibiotic discovery has declined dramatically since the "Golden Era" of antibiotic discovery during the 1940s-1960s (Fig. 1) [6]. Since that time, resistance has steadily progressed, bringing on the current "Era of Resistance" that features "superbugs" that cannot be treated with conventional antibiotics [7-10].

Many of our current antibiotics are derived from natural products that are produced by fungi or bacteria using specialized pathways and enzymes [11]. While there are arguments that we have reached the limits of antibiotic discovery based on natural products of easily cultivated bacteria [12, 13], there is still much we can learn from nature's defense systems, particularly antimicrobial peptides (AMPs).

Peptides are polymers of amino acids and are ubiquitous throughout nature. These small molecules are crucial in many processes for all organisms, from bacteria to humans. Peptides are categorized based on their origin and properties. In the context of antibiotic discovery, AMPs are of particular interest. A well-characterized class of AMPs is cationic peptides, which are typically thought to target bacterial membranes [14-20], though other intracellular targets are possible [21]. AMPs also elicit a variety of stress responses in cells, such as cell envelope stress response [22]. As part of this response, genes involved in cell wall synthesis, genes controlled by certain extracytoplasmic function sigma factors, and genes involved in certain two-component

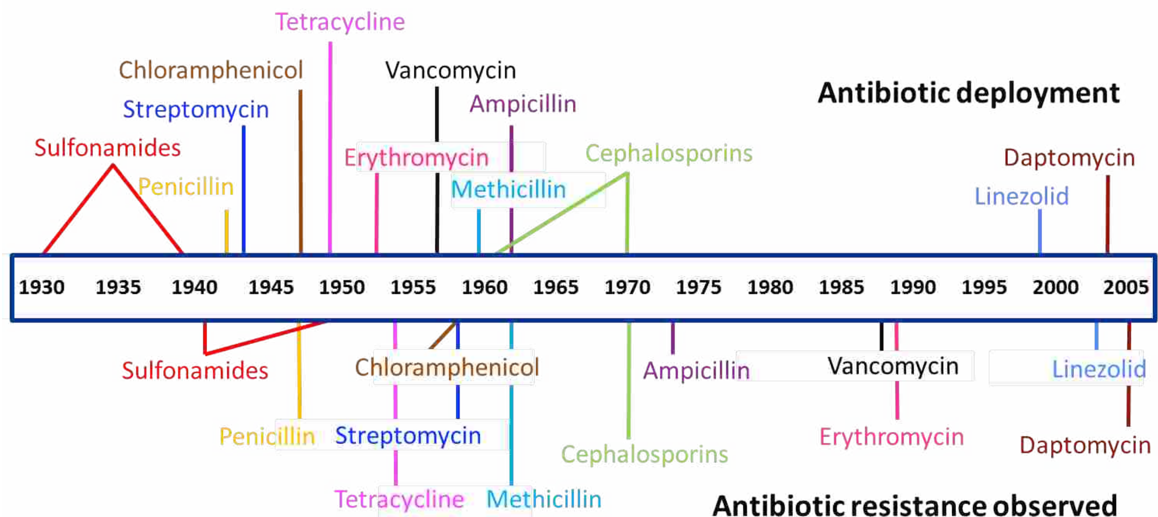


Fig. 1. Antibiotic resistance timeline. Timeline depicting when the antibiotic was first used (top) and when resistance was first observed (bottom). Adapted from: Clatworthy, Anne E., Emily Pierson, and Deborah T. Hung. "Targeting Virulence: A New Paradigm for Antimicrobial Therapy." Nature Chemical Biology 3.9 (2007): 541-48.

Table 1. Antimicrobial peptides

AMP	Origin	Subcategory	Function
vancomycin	non-ribosomal	glycopeptide	inhibits cell wall synthesis
colistin	non-ribosomal	polymyxin	disrupts cell membrane
micrococcin	ribosomal	thiopeptide	inhibits protein synthesis
thiostrepton	ribosomal	thiopeptide	inhibits protein synthesis
microcin J25	ribosomal	lasso	inhibits RNA polymerase
capistruin	ribosomal	lasso	inhibits RNA polymerase
human beta-defensin 3	ribosomal	cysteine-rich	inhibits cell wall synthesis
Ib-AMP4	ribosomal	cysteine-rich	disrupts cell membrane

signaling pathways are upregulated [23, 24]. The results of this upregulation include a variety of changes such as thickening the cell wall or altering net surface charge [25]. AMPs can either be ribosomally-synthesized, by translation of an mRNA template, or produced by non-ribosomal pathways (Table 1). Non-ribosomal peptides, generated by non-ribosomal peptide synthetases, account for multiple currently used antibiotics, and are primarily produced by bacteria and fungi (Fig. 2A) [26]. Ribosomally-synthesized AMPs, however, are expressed by a wide range of organisms from bacteria to plants to humans. While structures and mechanisms of antimicrobial activity may vary, the intent is always the same; to protect the host and its resources from threats. In bacteria, AMPs are exported to prevent other bacteria from using the surrounding resources [27]. In plants and humans, AMPs serve to protect the organism from potential pathogens [27]. While all are produced ribosomally, there are still a variety of intricate and elegant structures that can be achieved (Fig. 2B). Some of these structures require additional post-translational modifications by a variety of enzymes, while others rely on disulfide bonds between cysteine residues to achieve their structure.

1.2 RiPPs: enzymatically modified peptides resulting in complex structures

Ribosomally-synthesized Postranslationally-modified Peptides (RiPPs) are made by the ribosome, but as the name indicates, are heavily modified to generate un-peptide-like structures [28, 29]. RiPPs all undergo a similar maturation process. They begin as a linear precursor peptide, consisting of core and leader portions, then undergo a series of enzymatic reactions that result in a mature peptide containing a variety of unusual structures, such as heterocycles and macrocycles (Fig. 3A) [28]. To better understand this process, we will examine the structures of micrococcin P1 (MP1) and microcin J25 (MccJ25).

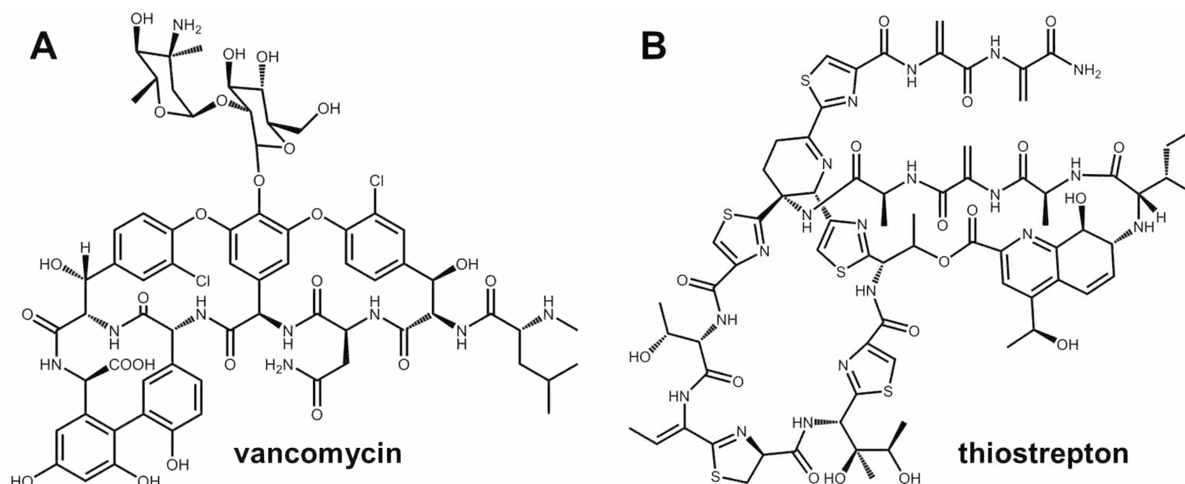


Fig. 2. Non-ribosomally and ribosomally-synthesized peptide structures. A. Structure of non-ribosomal peptide antibiotic vancomycin. B. Structure of ribosomally-synthesized antimicrobial peptide thiostrepton.

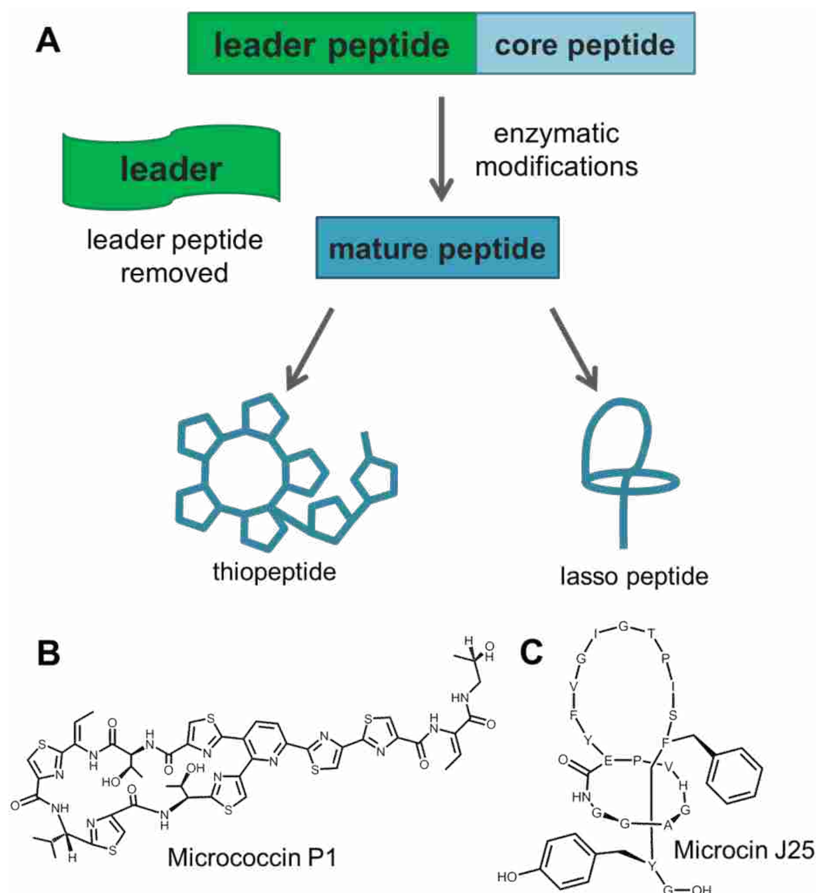


Fig. 3. Ribosomally-synthesized, post-translationally modified peptides. A. The general RiPP pathway begins with a long linear peptide and, through a series of enzymatic modifications, is converted to a mature peptide that has been separated from the leader peptide. B. Micrococcin P1 chemical structure. C. Microcin J25 chemical structure. Only relevant amino acid side chains are shown.

MP1 is a plasmid-encoded compound made by several bacteria, including *Macrococcus caseolyticus* and *Bacillus cereus* [30]. MP1 can be further categorized by its structure as a thiopeptide (containing thiazoles and a macrocycle). In order to go from a linear peptide to a highly modified macrocycle (Fig. 3B), MP1 requires a fairly simple pathway containing a minimum of seven enzymes to make its various chemical modifications [30]. MP1 possesses antimicrobial activity, making it an AMP of particular interest due to the relatively small number of enzymes required to form its elegant and complex chemical structure.

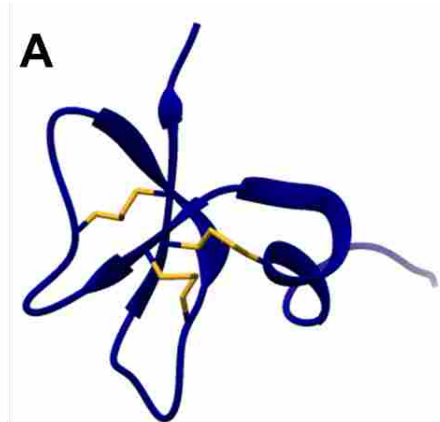
MccJ25 is a member of a specialized class of RiPPs called lasso peptides (Fig. 3C). Lasso peptides have the unique ability to loop their end through a ring, forming a knot structure. It is stabilized by large amino acid side chains on either side of the loop, as shown in Fig. 3C. Like MP1, MccJ25 is plasmid-encoded and requires a small collection of enzymes to achieve its final structure. Unlike MP1 however, most individual amino acids within the structure of MccJ25 are unmodified [31, 32]. Thus its toxicity comes primarily from its looped structure, indicating how crucial secondary structure can be in determining antimicrobial activity.

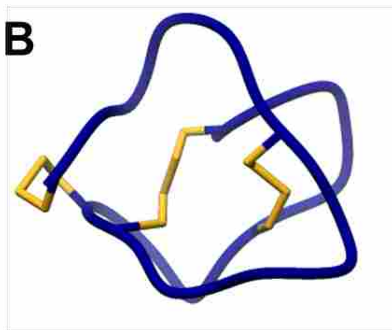
1.3 Cysteine-rich peptides: secondary structure using disulfide bonds

While RiPPs rely on enzymes to achieve higher-order structures, there are simpler ways for peptides to fold. One of the most basic ways is accomplished via disulfide bonds. These bonds depend on carefully positioned cysteines to fold correctly, and notably, they can only be maintained in oxidizing environments. Such cysteine-rich peptides can be found in several organisms, including humans.

1.3.1 Defensins and conotoxins

In humans, there are two main classes of AMPs, one of which is defensins. Defensins are defined by the presence of multiple cysteine residues throughout the peptide (Fig. 4A) [33-35].





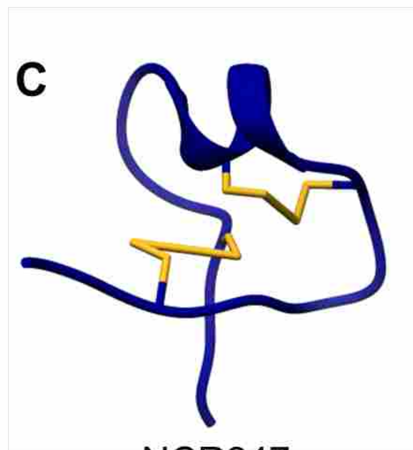


Fig. 4. Three-dimensional structures of cysteine-rich peptides A. Human beta-defensin 3. B. ω-conotoxin C. I-TASSER predicted structure of NCR247. Disulfide bonds are shown in gold.

Defensins are produced by human cells as part of the innate immune system. They primarily interact with bacterial membranes, permeabilizing, destabilizing, and killing the cells [33]. There is also some evidence that once the defensins break open the membrane and enter the cell, they are able to bind to molecules inside to further attack the bacteria [35]. This ability to achieve such a variety of functions is due to their chemical nature and structure, and their structure is reliant upon cysteine-mediated disulfide bonds.

Another cysteine-rich peptide is found in the venom of marine cone snails. These peptides are called conotoxins (Fig. 4B), and they allow the snail to paralyze its prey by attacking the nervous system [36]. These peptides are defined by the presence of several cysteines, similar to the defensins described above. The structural conformation caused by disulfide bonds is deceptively complex, as the structure required for neurotoxicity is very specific and is aided by additional posttranslational modifications to favor the active structure [37]. While in nature conotoxins are not antimicrobial, simple modifications made in the lab lead to antimicrobial properties. For example, macrocyclization of an ω-conotoxin (MVIIA) has produced a potent AMP [38]. This is yet another indication of the importance of specific secondary structures for antimicrobial activity.

1.3.2 Peptides involved in symbiosis

In addition to humans and snails, defensins and other cysteine-rich peptides have also been characterized in plants [39-43]. One class, known as Ib-AMPs, are only 20 residues long with four cysteines present in the peptides [40]. Despite their small size, they have demonstrated antifungal and antimicrobial activity. In addition to defensin-like activities, AMPs in plants can have a variety of functions, as illustrated by a class of specialized peptides involved in symbiosis. This category of defensin-like peptides is found in the leguminous plant *Medicago*

truncatula, and they are called Nodule-specific Cysteine Rich (NCR) peptides. These peptides have structural similarities to defensins (Fig. 4C), while having the unique function of orchestrating the relationship between the host plant and its symbiont bacteria, *Sinorhizobium meliloti* [44, 45].

In this symbiosis, *M. truncatula* forms structures called nodules on its roots (Fig. 5A). *S. meliloti* cells populate the nodules, entering the plant cells to reside, and in return they fix atmospheric nitrogen for the plant, facilitating growth (Fig. 5B). As part of this process, the plant expresses NCR peptides [46]. NCR peptides are transported through a specialized host secretory system to the bacterial cells, and act to convert the bacteria into terminally differentiated, nitrogen-fixing organelles (Fig. 5C) that continue to grow but no longer divide. When the plant dies, these large bacteria, called bacteroids, die with it.

In response to the plant's mobilization of NCR peptides, some *S. meliloti* strains express a peptidase called HrrP. This protein degrades NCR peptides, preventing or diminishing levels of terminal differentiation [47]. The constant back-and-forth conversation between plant and bacteria has led to hundreds of NCR peptides being encoded in the plant's genome, with over 700 NCR peptides annotated (Table 2) [48]. Many of these NCR peptides have been studied in isolation and found to have a variety of functions, including antimicrobial activity [45, 48, 49].

Like defensins and conotoxins, NCR peptides rely on disulfide bonds between cysteines for their structure and subsequent activity [50]. The secondary structure achieved by these relatively simple bonds leads to peptide molecules that more closely mimic the complex structure of antibiotics. This suggests that disulfide bonds can be exploited for the study and production of synthetic peptide antibiotics.

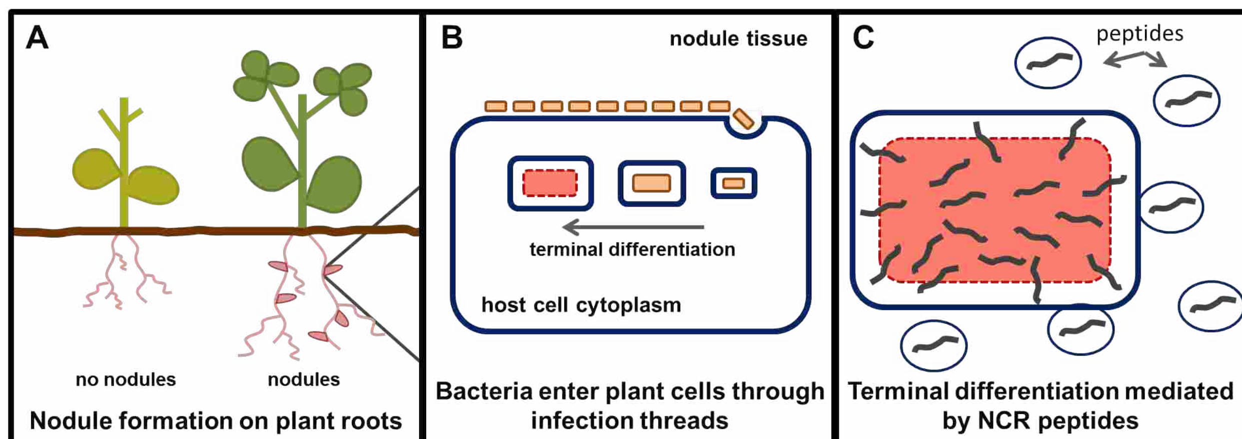


Fig. 5. Symbiosis of *M. truncatula* and *S. meliloti*. A. Nodules form on plant roots. Plants unable to nodulate exhibit inhibited growth. B. Bacterial cells enter the host cytoplasm and begin terminal differentiation. C. The host inoculates bacterial cells with NCR peptides, resulting in bacteroids that are unable to divide.

Table 2. NCR peptide sequences and functions

NCR peptide	Amino Acid Sequence	Function
NCR247	RNGCIVDPRCPYQQCRRPLYCRRR	Antimicrobial activity, blocks cell division, inhibits translation
NCR169	EDIGHIKYCGIVDDCYKSKKPLFKIWKCVENVCVLWYK	Regulation of differentiation, bacteroid maintenance, regulation of nitrogen fixation
NCR211	DRECDTDTECQKKFPGVNAHHLWCDNGNCVSYPK	Bacteroid maintenance, regulation of nitrogen fixation
NCR335	RLNTTFRPLNFKMLRFWGQNRNIMKHRGQKVHFSLILSDCKTNK DCPKLRRANVRCRKSYCVPI	Antimicrobial activity, destabilizes membrane
NFS1	ASPFYCDEDDYFCFGLCLPPMIDHCTLGGQCICITISTEVES	Antimicrobial activity
NFS2	NEECTCAADCYKRYPRWSLLPTYCIEGSCYSDFLNSGKKYLSP	Destabilizes membrane, blocks cell division
NCR035	NSSFLGTFISSCKRDKDCPKLYGANFRCRKGTCVPPI	Blocks cell division, destabilizes membrane

Cysteines are underlined for emphasis.

1.4 Strategies for creating synthetic peptides

While a multitude of AMPs is evident in nature, and many have yet to be characterized, researchers have begun to design synthetic peptides similar to the peptides that have been studied. If we can find the patterns that lead to toxicity, it stands to reason that we can then apply those patterns to design specific novel AMPs. Many techniques have been used to design synthetic peptides, from selective mutation to codon randomization to in silico optimization of natural peptides. Each of these methods relies on a natural compound as its scaffold.

1.4.1 Selective mutation

Among the first ways of designing synthetic AMPs was to take natural AMPs and create a series of variants on that peptide. This technique has been used for a variety of peptides and has been successful in producing increasingly toxic AMPs [51-53], some of which have been shown to give plants resistance to pathogens [54, 55]. For example, making substitutions to a variant of bovine peptide bactenecin led to the discovery of Bac8c, a potent antimicrobial [56]. The result of these substitutions took the peptide from RLARIVVIRVAR to RIWVIWRR [52], a small peptide that is active against a broad spectrum of pathogens. While this method has proven successful, as demonstrated by this example, it requires natural peptides acting as starting points.The easiest method of rational peptide design is to take a group of peptides, find similarities, and then incorporate those patterns into a new peptide. This method has seen great success by manually deriving a pattern from as few as 80 peptides [57]. By calculating the frequency of each residue at each position in a set of AMPs, a consensus sequence can be determined. Comparing peptides from different sources leads to different sequences, such as the consensus GILSKLGKALKKAAKHAAKA from comparing peptides from insects versus GRFRRLGRKFKKLFKKYGP from comparing mammalian peptides [57]. The resulting

peptides, however, focused specifically on hydrophobic regions, a common property of many canonical AMPs. This indicates that the method does not provide a path for designing peptides with unique targets and toxicity mechanisms. This illustrates the difficulty of designing a peptide, and the inability of researchers to explore the infinite options for AMPs.

1.5 Computational methods for AMP discovery

In the past, synthetic peptide design was driven primarily by the researcher and their ability to observe specific patterns in a group of AMPs. This leads to a lack of novel targets and mechanisms in regard to AMPs, as the entirety of peptide possibilities is not explored using the previously mentioned methods. With the advent of high-throughput technologies and machine learning methods, however, it has become possible to take advantage of computational methods to expand on known biological principles and investigate novel AMPs.

Modern methods have begun to take advantage of the power of computation to discover patterns. One example of a novel peptide design relies on statistical analysis of peptides in a database (APD2) [58]. Using peptide analysis tools, the most common amino acids in a group of peptides contained in the database can be identified and used to generate new peptides. 831 AMPs were compared to determine that among the most frequently occurring residues were glycine (Gly), leucine (Leu), and lysine (Lys). Three peptides were then designed based on a 12-amino acid template peptide (KR-12). The first replaced all non Leu/Lys residues with Leu or Lys. The second was a combination of motifs generated by tools associated with the database using the most common amino acids, resulting in a peptide consisting of Leu, Lys, and Gly. The third peptide was designed based on the overall percentages of each amino acid and common

motifs, leading to a more diverse sequence. One of the three designed peptides was more toxic than the original AMP template [58].

Another way in which computational algorithms can expedite the discovery of novel AMPs is through genome mining. In this method, previously unannotated genes can be identified as potential AMPs and subjected to further testing. One sophisticated genome mining tool was recently used to identify over 1,000 lasso peptide gene clusters, including several novel lasso peptides [59]. Further analysis of the data revealed valuable patterns in regard to precursor peptides, conserved motifs, evolutionary lineage, and predicted protein interactions. This demonstrates the ability of computer programs to contribute to the study and discovery of AMPs.

There are also ways that algorithms can contribute to AMP rational design. Computationally-driven codon randomization of thiopeptides has led to the generation of peptides with more potent antimicrobial activity [60]. In addition to yielding a more toxic peptide, this method also led to insights concerning what structural components are necessary for activity. Unfortunately, it also required biological testing of all the possible combinations following randomization. The analysis of these libraries can be time- and labor-intensive. By using increasingly elegant algorithms, however, it is possible to eliminate some of the bench work that has previously been required.

Work by the Franco lab has recently yielded a method of optimizing AMPs in silico, eliminating much of the costly wet lab work [61]. Using a natural AMP from guava, Pg-AMP1, as a template, the program essentially seeks to mimic evolution and natural selection, assigning a fitness score to each peptide derivative (termed guavanins). The peptides with the top fitness scores were synthesized and tested for activity. This resulted in a novel potent AMP, with a sequence significantly different from the original peptide and with reduced bench work. The

limitation of this method, however, lies in its dependence on a template AMP from which to derive more potent variations.

In addition to classical computer algorithms, machine learning algorithms have also been applied to the search for novel AMPs. These methods typically work by providing computers with a set of previously characterized AMPs as "training", then the computers use information gleaned to predict and design novel AMPs in addition to providing insight about conserved motifs, patterns, and mechanisms in the process [62]. One example of machine learning focuses specifically on membrane active peptides [63]. Like the previous example of guavanins, this program explores the sequence space in ways that can be difficult to perform in the lab. Unlike the previous example, however, this method is not dependent on a single template peptide. It relies on large sets of characterized AMPs to do its work. This indicates a need for additional methods to explore the vast number of possibilities to identify candidates that could then be used by programs such as this one in order to generate more potent peptide antibiotics as efficiently as possible. The design and screening of large libraries of AMPs is critical in this process.

1.5.1 AMP libraries for high-throughput screening

In 1995, researchers were contemplating the design of libraries of AMPs. In addition to suggesting conotoxins as therapeutics, it was suggested how researchers might be able to design combinatorial libraries using the lessons learned from how cone snails develop diverse conotoxins with specific targets [64]. Of particular interest in developing libraries are cyclic molecules due to their stability, rigidity, and structural similarities to known antibiotics. Macrocylic peptides have inspired antibiotics currently in clinical trials and are a useful scaffold in the further exploration of peptide antibiotics [65].

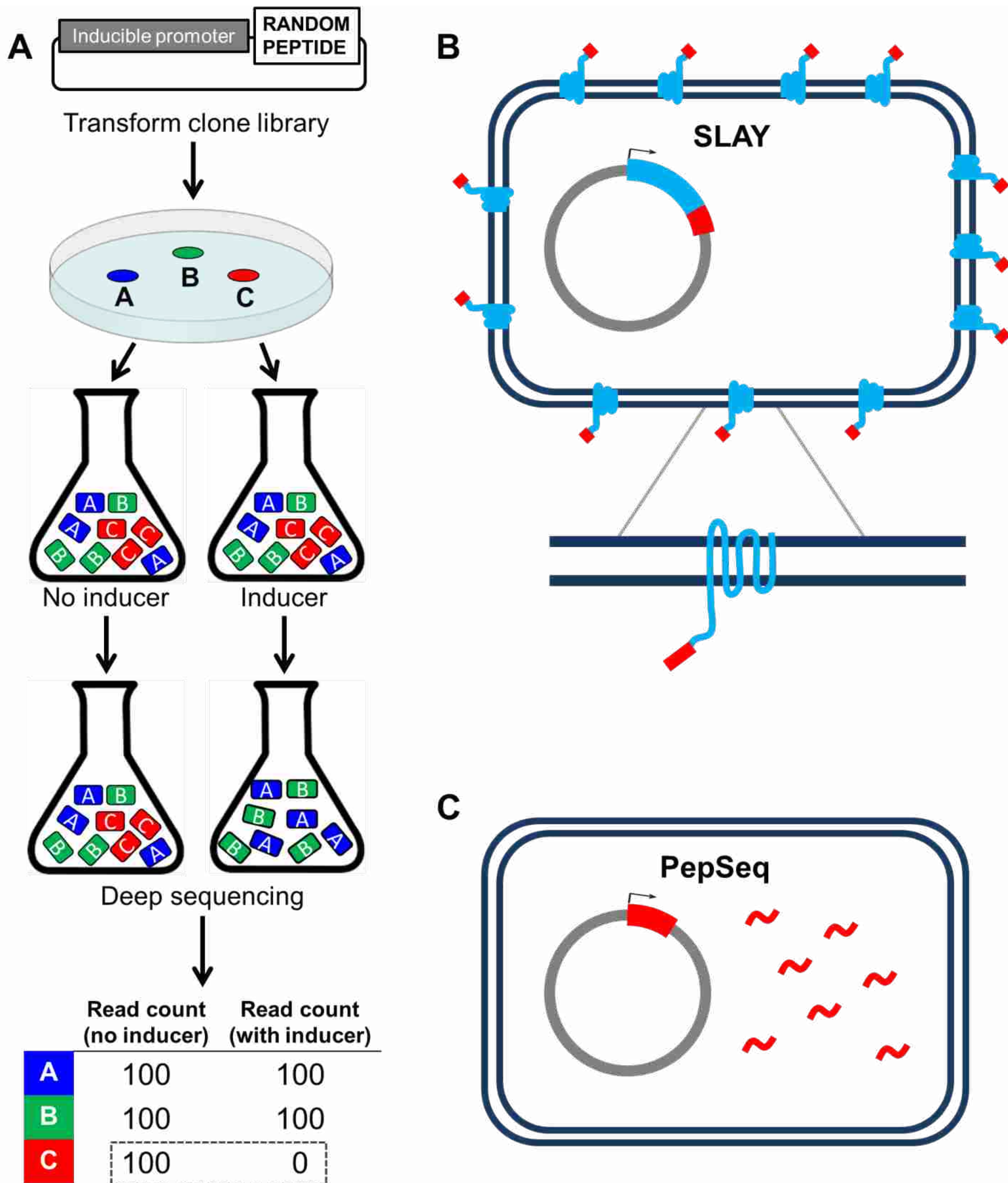


Fig. 6. Strategies for high-throughput antimicrobial activity assays. A. Workflow for SLAY and PepSeq, beginning with plasmid library construction, then growth induced and uninduced. These steps are followed by Illumina sequencing and counting reads to detect antimicrobial peptides. B. For SLAY, peptides are tethered to the outer membrane, allowing additional detection of ability enter the cell. C. For PepSeq, peptides are expressed in the cytoplasm, allowing detection of intracellular targets.

While our ability to generate large libraries of diverse structures has increased, the issue of screening these libraries to find bioactive compounds remains rate-limiting. Recent progress has been made in regard to screening these large libraries. A technique called Surface Localized Antimicrobial Display (SLAY) has been developed, which simultaneously screens a random peptide library for toxicity and cell permeability [66]. This is accomplished by tethering a random peptide to the outer membrane of a cell, just long enough that it can re-enter the two membranes but not so long that it can permeate surrounding cells (Fig. 6A, B). This leashed peptide is placed under the control of an inducible promoter on a plasmid. Libraries are then grown under induced and uninduced conditions, and depletion of certain peptides is detected by deep sequencing of the cultures. Depleted peptides were further tested to determine the extent of their antimicrobial activity. This resulted in the discovery of thousands of AMPs. The tether is only long enough to allow the peptide to barely reenter both membranes, making it likely that the peptides discovered act primarily on the membrane. This thesis presents a complementary version of this high-throughput technique, where peptides remain in the cytoplasm, thus allowing the discovery of peptides with unique intracellular targets.

1.6 Summary

This thesis presents two related projects. One explores the antimicrobial activity of NCR peptides in *E. coli*. The other is the discovery of novel AMPs via a high-throughput platform. Each aspect is highly dependent on plasmid systems of exquisitely inducible promoters driving expression of peptides of interest.

1.6.1 Promoter optimization

To determine the best promoter for each project, a variety of inducible promoters was tested for expression levels when induced and uninduced. The ribosome binding site (RBS) of the promoters was also examined as part of the optimization. Both *lacZ* and green fluorescent protein (msfGFP) were used as reporter genes, and testing was performed in two strain backgrounds grown in rich and minimal media. While labor-intensive, these experiments resulted in valuable data that informed the critical decision of which promoter and strain to use for final screening.

1.6.2 Antimicrobial activity of NCR peptides in *E. coli*

As previously discussed, hundreds of NCR peptides have been annotated, and several have been more thoroughly characterized. While there are several approaches to begin to comprehensively understand NCR peptide properties, one aspect that is easily examined is antimicrobial activity. This thesis discusses the development of an assay for testing antimicrobial activity of NCR peptides in *E. coli*, and potential adjustments for use of the assay in *S. meliloti*.

The assay is a plasmid-based system in which an NCR peptide is placed under the control of an inducible promoter. Cells are then induced to test for a self-killing phenotype, indicating antimicrobial activity of the particular NCR peptide. The system can be further adapted to export the NCR peptide to the periplasm or tether the peptide to the inner membrane. This allows additional insight into the importance of disulfide bonds in antimicrobial activity, and where the peptide targets are located within the cell. While still under development, it is hoped that this assay can be used to test many more NCR peptides for antimicrobial activity to further understand the role of NCR peptides in symbiosis, and their potential to be used in other applications.

1.6.3 Massively parallel peptide activity screen (PepSeq)

This thesis additionally presents the development and optimization of a high-throughput platform for discovering novel AMPs. This platform, known as PepSeq, takes advantage of high-throughput sequencing to determine potential antimicrobial activity of random peptides (Fig. 6A, C). In this system, a plasmid encoding a random segment of DNA under the control of an inducible promoter is transformed into *E. coli* leading to a plate of hundreds of thousands of colonies. Each of these colonies represents a unique peptide-encoding region. These colonies are pooled together to form a library. This library is then grown under induced and uninduced conditions. After a period of growth, each culture is sequenced using Illumina technology.

Following sequencing, we use a computational tool to count the number of sequence reads for each peptide-encoding sequence. In the uninduced population, where no peptides were being produced, we anticipate that the entire spectrum of peptide-encoding sequences will be present. When peptide production is induced, those that encode toxic peptides will drop out of the population, leading to a reduced number of sequences in the induced population. In this way, we anticipate that we can identify hundreds of potentially toxic peptides in a single experiment. These peptides can then be further studied, and the platform is highly adaptable to various peptide scaffolds.

Similar platforms have been used to generate peptides that can infiltrate cells and act as antibiotics, as previously mentioned [66]. The platform described in this thesis provides a cytoplasmic version, and additionally provides and describes standardized Python programs, developed by other graduate students and collaborators from the Computer Science department at Brigham Young University, for easy and rapid analysis of the data obtained from the assay.

CHAPTER 2: Promoter optimization in *E. coli*

2.1 Background

To observe self-killing phenotypes of AMP-producing strains, it is imperative to achieve the lowest rate of basal expression possible. This type of controlled expression can be achieved using a number of inducible promoters [67-72]. A set of inducible promoters was tested to determine the rates of expression with and without inducer, or tightness of the promoter. Each inducible promoter has a slightly different structure, shown in Fig. 7.

The primary components of a bacterial promoter are the -35 and -10 elements where RNA polymerase binds, and binding sites for regulatory proteins, or operators. For each inducible promoter, there is an associated regulatory gene. These genes encode regulatory proteins and are under the control of constitutive promoters that allow constant expression at low levels. In some cases, an additional regulatory protein called Catabolite Activator Protein (CAP) is involved in regulation of the inducible promoter. This molecule binds cyclic-AMP (cAMP), which is made in high concentrations when glucose is absent. Once CAP binds cAMP, it binds the DNA and activates transcription of the desired gene. For some inducible promoters, the CAP/cAMP complex is crucial for high expression of the desired gene.

The lactose promoter (P_{lac}; Fig. 7A) has three operator sites, which are bound by LacI in the absence of lactose, preventing transcription of the desired gene. When lactose is present in the cell, a portion of the molecule, called allolactose, binds LacI. This changes its shape and releases LacI from its binding site, allowing transcription to occur. The CAP/cAMP complex must also bind to initiate transcription from the inducible promoter. Similarly, the arabinose promoter (P_{ara}; Fig. 7B) has three operator sites, bound by a regulatory protein called AraC. In the absence of arabinose, AraC binds at O_2 and I_1, making the CAP binding site inaccessible.

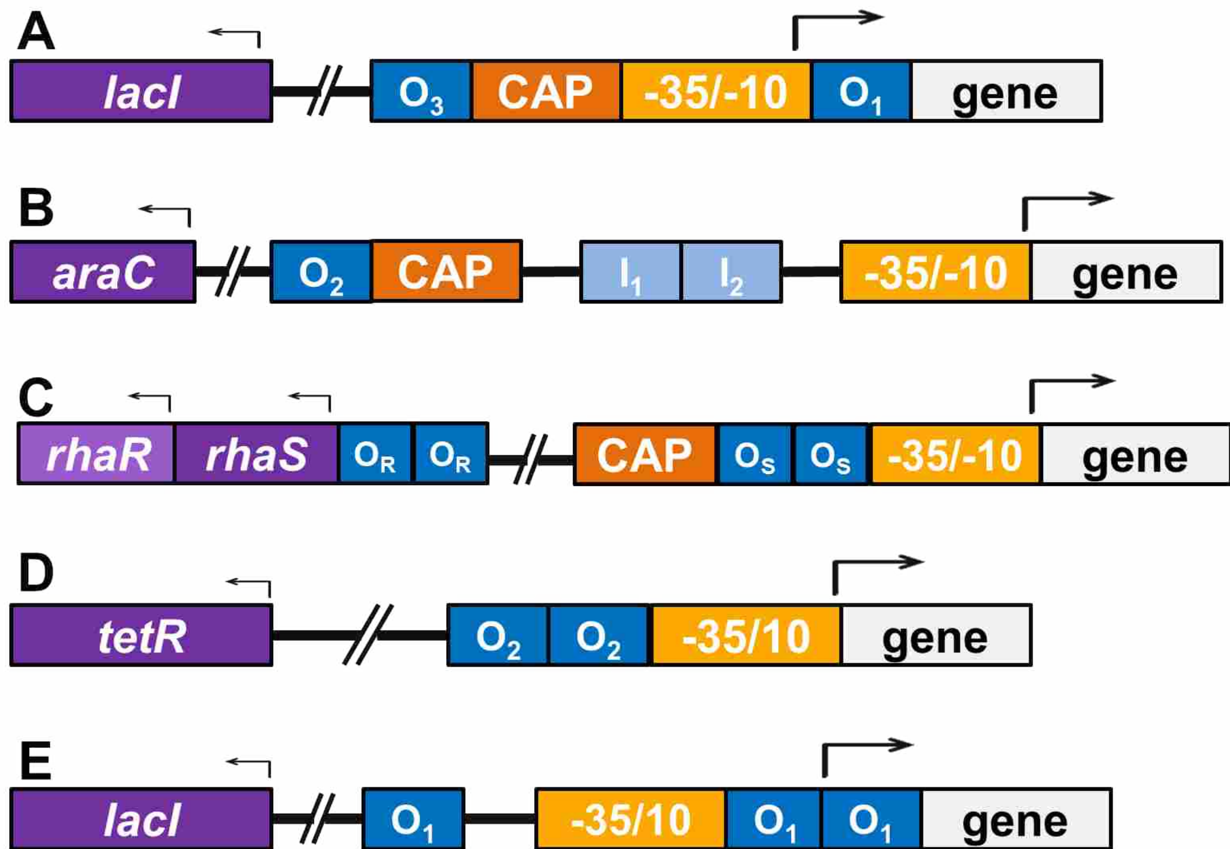


Fig. 7. Inducible promoters. Regulatory protein-encoding genes are shown in purple, with their operating sites in blue. The constitutive promoter driving expression of regulatory genes is indicated by a left-facing arrow. The CAP/cAMP binding site (where applicable) is shown in dark orange, while the RNA polymerase recognition sequence (-35/-10) is shown in light orange. The transcription start site for the gene of interest is indicated by a right-facing arrow. A. P_{lac}. B. P_{ara}. C. P_{rha}. D. P_{tet}. E. P_{lacT5}.

When arabinose is bound to AraC, it binds instead at I_1 and I_2, leaving the promoter accessible to CAP and allowing expression to occur. The rhamnose promoter (P_{rha}; Fig. 7C) has two regulatory proteins, RhaR and RhaS. RhaR binds at the O_R sites to activate transcription of *rhaS*. The product (RhaS) then binds at O_s to induce expression of the genes downstream of the promoter. Like P_{ara} and P_{lac}, P_{rha} is additionally regulated by the CAP/cAMP complex.

While the promoters described above are derived from *E. coli,* some promoters are highly engineered, such as the tetracycline promoter (P_{tet}; Fig. 7D), or P_{lacT5} (Fig. 7E). In the case of $P_{tet,}$ TetR binds the operator sites in the absence of anhydrotetracycline (ATc) to repress expression. When ATc binds, TetR is released and expression can occur. The hybrid promoter P_{lacT5} combines parts of two different promoters, and consists of the promoter from the T5 phage, with operator sites from P_{lac}, making the phage-derived promoter inducible with isopropyl β-D-1-thiogalactopyranoside (IPTG). The differences in regulation mechanisms of each promoter lead to varying expression levels and levels of tightness in each inducible promoter.

To test the expression of different promoters in induced and uninduced conditions, researchers often use reporter genes. A variety of these genes can be used, and each has a corresponding assay. For our plasmid-based assays, one of two reporter genes was added to a multi-copy plasmid containing both an inducible promoter and the appropriate regulatory gene with its associated constitutive promoter. The two reporter genes used were *lacZ* and monomeric superfolder green fluorescent protein (msfGFP). *lacZ* encodes an enzyme typically involved in lactose metabolism in *E. coli*, called β-galactosidase. This enzyme cleaves lactose, breaking it down into smaller parts. To detect the amount of expression of *lacZ,* a Miller assay is used. This assay takes advantage of the ability of β-galactosidase to cleave chemicals similar to lactose, such as *ortho*-Nitrophenyl-β-galactoside (ONPG). When ONPG is cleaved, it results in *o-*

nitrophenol, which has a distinct yellow color. Miller assays measure how much of the yellow product is produced, leading to an indirect measurement of the amount of enzyme. While this assay is very sensitive, one limitation of using this reporter gene is that, in addition to the plasmid-encoded copies, it is also present on the genome of wild-type *E. coli*, and will thus be expressed regardless of promoter activity on the plasmids. It can only be used for strains that do not have *lacZ*, hence the need for the second reporter gene msfGFP. The assay for msfGFP is more straightforward, because the presence of msfGFP can be easily detected by measuring fluorescence in arbitrary fluorescence units (AFUs), and msfGFP is not naturally found in *E. coli*. For either reporter gene, however, the amount of activity when uninduced allows a measure of basal level expression, while the amount of activity when induced allows a measure of expression levels. Characterization of promoters allows researchers to make informed decisions concerning which promoters to use for various projects.

2.2 Methods

Appendix I contains relevant primer and plasmid sequences. Appendix II contains full protocols for cloning, in addition to buffer and media recipes.

2.2.1 Cloning native P_{ara} and P_{rha} promoters

Promoters were amplified from wild-type *E. coli* genomic DNA, using primers 1634 and 1635 (for P_{ara}) or 1636 and 1637 (for P_{rha}). Inserts were ligated into pJG744 and pJG747, which contains *lacZ* as a reporter gene, following digestion with XhoI and KpnI. All plasmids were transformed into DH5α. All clones were sequence-verified using Sanger sequencing.

2.2.2 Cloning P_{tet} and P_{lacT5} promoters

P_{tet} was amplified from pJG717 using primers oDB014 and oDB015. P_{lacT5} was amplified from pJG729 using primers 1642 and 1643. *lacI* was amplified from pJG729 (pSX2 from Scarab Genomics) using primers 1638 and 1639 (for original *lacI*) or 1638 and 1640 (for *lacIq*). Inserts were ligated into pJG744 and pJG747 following digestion with XhoI and KpnI, and additionally BamHI for P_{lacT5} and *lacI*. All plasmids were transformed into DH5α. All clones were sequence verified using Sanger sequencing.

2.2.3 Cloning P_{rha} with extended ribosome binding site, P_{ara} from pGLO, and P_{lacT5} from pSX2

P_{rha} was amplified from pED006 with primers oED096 and oED097 to add an extended ribosome binding site (XRBS). P_{ara} was amplified from pGLO (lab strain C398) with primers oED098 and oED099. P_{lacT5} was amplified from pSX2 (lab strain pJG719), from Scarab Genomics. Each insert was ligated into pED006, which contains msfGFP as a reporter gene, following digestion with XhoI and NdeI. Plasmids were isolated using a Zymo Research Miniprep Plasmid Classic kit and transformed into MG1655 (D050).

2.2.4 Fluorescence assays

Cells were grown with and without inducer (0.3% rhamnose (Rha), 0.3% arabinose (Ara), or 0.1mM IPTG). Cultures were then diluted to a normalized optical density (measured at 600 nm, OD_{600}) based on the lowest measured value. AFUs were determined using a BioTek plate reader. Final AFU values were normalized to the media blank and control cultures.

2.2.5. Growth curves

Cells were grown in 4 ml-cultures (unless otherwise noted), and the OD_{600} was measured at specific time points. In many cases, a 1:10 dilution of culture was used to determine OD_{600}.

2.2.6 Miller assays

Samples were grown in LB with and without inducer (0.3% Ara, 0.3% Rha, 0.3mM IPTG, 0.4 μM ATc) for approximately 4 hours prior to performing Miller assays. The OD_{600} of each culture was measured. For each assay, β-Gal master mix (see Appendix II), chloroform, and diluted culture were combined. Samples were incubated at 32°C until bright yellow color was observed in 1 or more samples. 1M Na_2CO_3 (see Stop Buffer, Appendix II) was added to end the reaction. Samples were briefly centrifuged prior to measuring OD_{420}. Miller units were calculated using the following equation: $(1000*OD_{420})/(OD_{600}*t*v)$ where t is time in minutes and v is volume in ml.

2.3 Results

2.3.1 Comparison of P_{rha}, P_{ara}, P_{lacT5} and P_{tet}

Initial tests began with two native *E. coli* inducible promoters, P_{ara}, and P_{rha}, and two versions of P_{lacT5}. The P_{lacT5} promoter is a combination of the operators from P_{lac} and a T5 phage promoter. By adding in *lacI* and its operators, the T5 promoter becomes IPTG-inducible. One version included *lacI* under the control of its native constitutive promoter, and the other used a stronger constitutive promoter known as *lacIq* (Fig. 8). Previous studies have shown that the higher level of LacI expression that results from this change decreases leakiness of the inducible promoter [72], a compelling reason for its inclusion in the overall comparison. For initial tests, the promoters were tested in DH5α, a strain where *lacZ* has been removed. Miller assays were performed to determine the levels of expression of both induced

Fig. 8. *lacI* promoter layout. -35 and -10 sites, and start codon are in uppercase. The *lacI*q mutation is shown in bold.

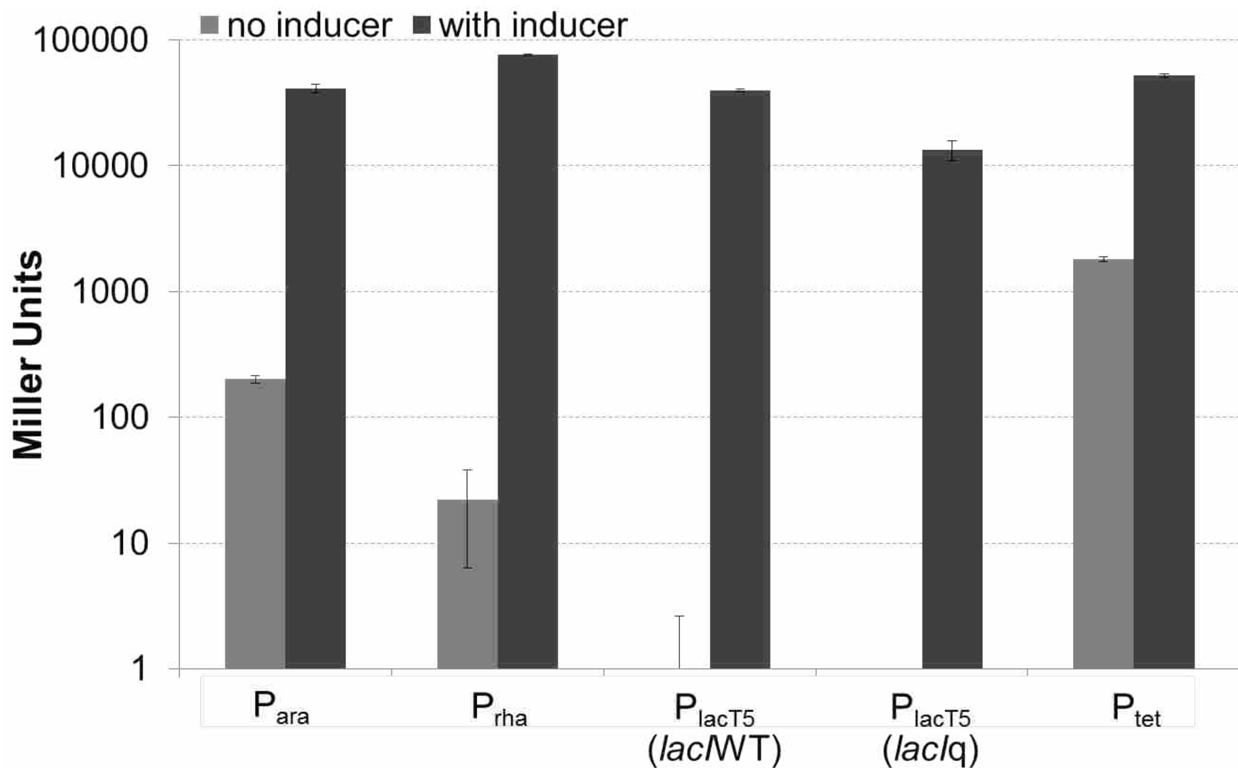


Fig. 9. Miller assays in DH5α. Five promoters were tested using *lacZ* as a reporter gene. Strains were grown in LB with and without inducer prior to performing Miller assays. Cultures were grown in triplicate. Error bars show standard error.

and uninduced cultures. The results of the Miller assays are shown in Fig. 9. From this first round of promoter testing, we found P_{rha} to have the highest levels of expression when induced and minimal levels of expression when uninduced, in *E.coli*.

2.3.2 Comparison of P_{rha}, P_{ara}, and P_{lacT5} in MG1655

Following an initial round of tests, we moved P_{rha} into wild-type *E. coli* (MG1655), where higher levels of expression when no inducer was added were observed. Thus, we tested additional promoters to determine tightness in MG1655. In addition to P_{rha}, P_{ara} from pGLO and P_{lacT5} from pSX2 were used. We adjusted P_{rha} to include the extended RBS region (XRBS, Fig. 10A) that was common to both of the commercial expression vectors. Strains were initially grown in LB to test expression (Fig. 10B). Expression levels were similar when grown in minimal media (M9M, Fig. 10C). While no significant growth defects were noted in LB, when strains expressing msfGFP were grown in M9M, adding rhamnose to the strain containing P_{rha} caused significant loss of growth compared to MG1655 (Fig. 10D).

2.3.3 Carbon sources in minimal media

It is important to note that in the previous tests, it was critical to use glycerol as the carbon source in minimal media. As mentioned previously, many of the promoters tested rely on the CAP/cAMP complex for expression. This means that if glucose were used as the carbon source, the promoters would be expressed at lower levels due to absence of an activator. P_{lacT5} does not rely on CAP/cAMP for activation. Following tests with the XRBS, where P_{lacT5} showed the highest levels of expression when induced and the tightest levels of expression when uninduced, this inducible promoter (expressing msfGFP, in MG1655) was grown in M9M with either glucose or glycerol. Strains grown in glucose showed both improved growth and

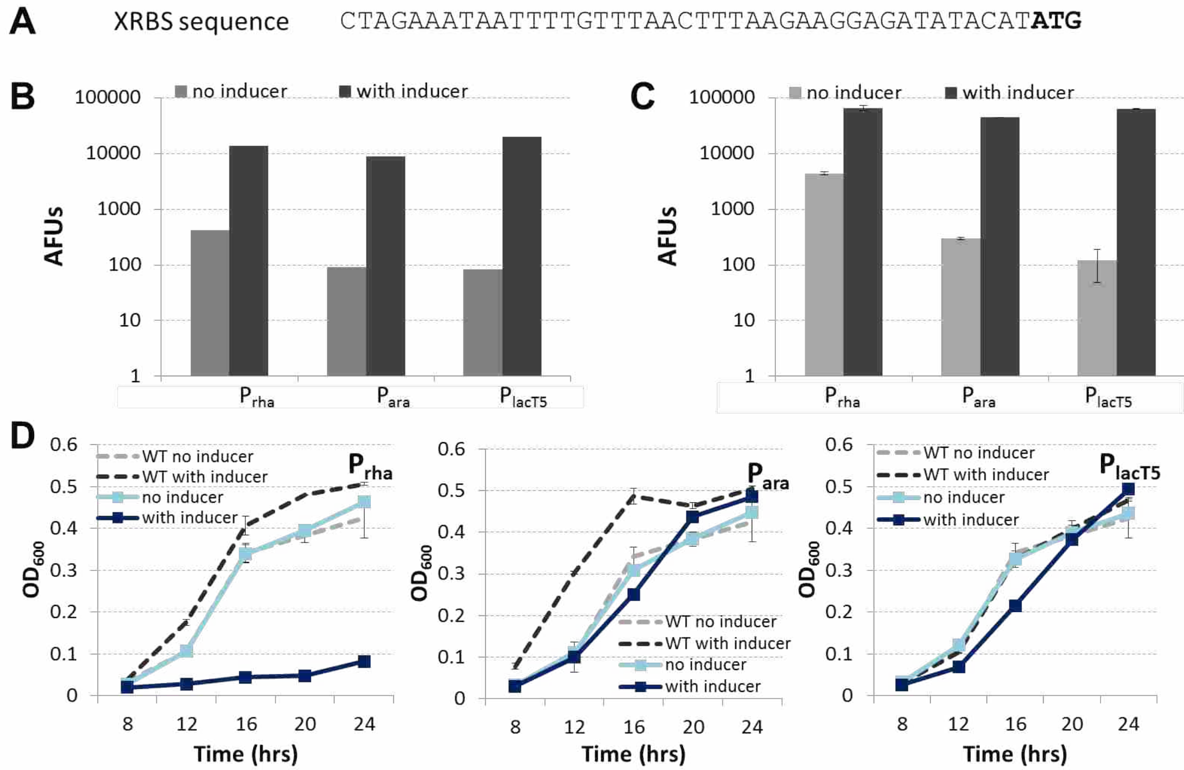


Fig. 10. XRBS data summary. A. XRBS sequence. Start codon is shown in bold. B. Promoter expression comparison, cultures grown in LB. C. Promoter expression comparison, cultures grown in M9M in duplicate. D. Growth curves in M9M, cultures grown in duplicate. Wild-type controls are shown as dashed lines. In each case, error bars show standard error.

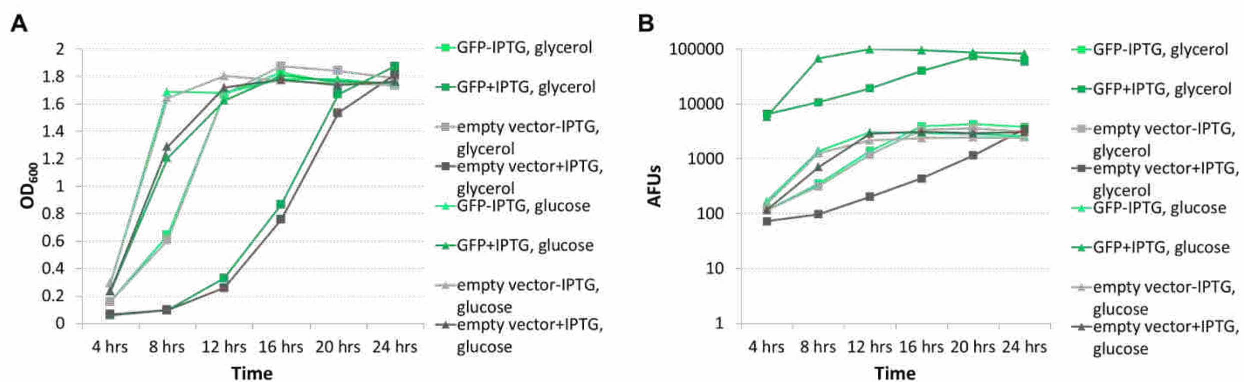


Fig. 11. Glycerol versus glucose as the carbon source in M9M. A. Growth curve of empty vector and vector with msfGFP in M9M+glucose and M9M+glycerol. B. Fluorescence of cultures used for growth curves. In each graph, gray lines correspond to the empty vector and green lines to vector with msfGFP. Square markers indicate glycerol as the carbon source, and triangles indicate glucose as the carbon source.

expression (as measured by fluorescence) over time as compared to those grown in glycerol (Fig. 11).

2.4 Discussion

In this chapter, we've explored a number of promoters that might allow maintenance of strains expressing toxic peptides. To observe a self-killing phenotype, it is crucial to prevent peptide expression during cloning and maintenance of peptide-encoding bacterial strains. Even basal levels of expression of a highly toxic peptide could inhibit cell growth prior to completion of cloning, making it impossible to generate the necessary strains. To ensure controlled expression, an inducible promoter is used. Because of the importance of this aspect, we compared several promoters to determine expression levels, using different reporter genes, to determine the promoter that performed best in all conditions.

Two promoters stood out in terms of tightness and expression: P_{rha} and P_{lacT5} (Fig. 12). P_{rha} and P_{lacT5} both performed very well in most experiments. P_{rha} had leakier expression in wild-type *E. coli* as opposed to DH5α, and often leads to a growth deficit when rhamnose is added in minimal media. P_{lacT5}, while maintaining tight and very high expression in all conditions tested, has its own weaknesses. This promoter relies on *lacI* as its repressor. When a plasmid containing *lacI* is transformed into a cell, there is an interim period when the genes on the plasmid are not yet being expressed. In DH5α, *lacI* has been removed from the genome, leaving a period of time immediately after transformation when repressor is not present in the cell and the gene controlled by the inducible promoter can be expressed freely. This phenomenon is known as the zygotic effect. When expressing highly toxic peptides, this can lead to the growth of few or no colonies following transformation. This problem is solved by transforming instead into MG1655 (wild-

type *E. coli*), which has a genomic copy of *lacI*, and thus immediately represses expression from the plasmid.

For the PepSeq project, we wanted very high and very tight expression, in MG1655 as the background strain. P_{lacT5} is the ideal promoter in that strain, with the condition that all experiments would be conducted solely in MG1655, with no DH5α intermediates. For the NCR peptide project, P_{rha} was chosen as the promoter.

While there are inducible promoters that stand out in terms of high expression and tightness, none behaves perfectly in every condition. Growth media, the peptide being expressed, and strain background for a particular project are all important aspects to be considered in determining which inducible promoter to use.

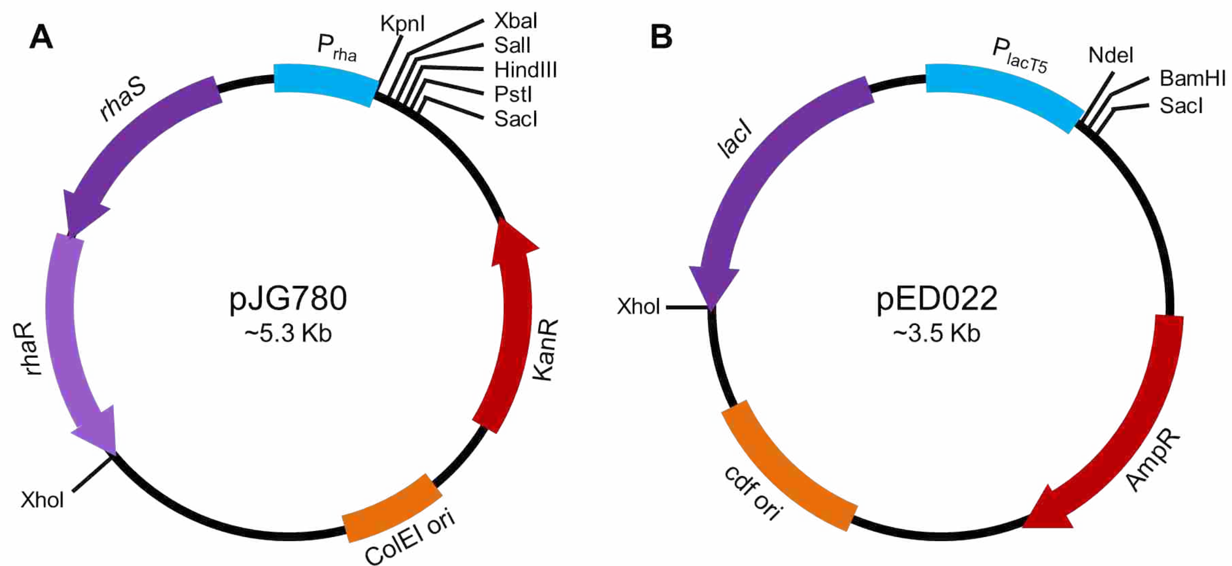


Fig. 12. Vector maps. A. P_{rha} vector, used for expressing NCR peptides. B. P_{lacT5} vector, used for expressing random peptides for PepSeq.

CHAPTER 3: Antimicrobial activity of nodule-specific cysteine rich peptides in *E. coli*

3.1 Background

The symbiotic relationship between the plant *Medicago truncatula* and *Sinorhizobium meliloti* bacteria is one of great interest in agriculture. Atmospheric nitrogen cannot be used by plants until after nitrogen fixation. Typically agricultural plants acquire nitrogen from synthetic fertilizer. In the case of *M. truncatula*, fertilizer is not necessary due to nitrogen fixation by *S. meliloti*. *S. meliloti* enters the plant cells and fixes nitrogen for the plant. The symbiosis is also of great interest due to the complex dialogue between plant and bacteria via a variety of chemical compounds, including peptides. This communication directs the bacteria into nodule tissue, and induces differentiation of the bacterial cells into nitrogen-fixing entities.

The first conversation between plant and bacteria occurs in the soil. The plant releases flavonoids into the soil, where rhizobia can recognize them and return their own signal in the form of Nod factors [73]. Nod factors initiate the nodulation process (Fig. 13A, B). The root nodules are intended to house the bacteria, and provide resources as the bacteria provide fixed nitrogen to the plant. When the nodule is fully formed, bacteria enter through what is known as an infection thread (Fig. 13B) [74]. Once inside the plant cell, another dialogue occurs when the plants send Nodule-specific Cysteine Rich (NCR) peptides to the resident *S. meliloti* (Fig. 13C). These peptides act on and in the bacteria in a variety of ways that, in combination, result in the formation of terminally differentiated bacteroids [45]. The bacteroids continue to grow, but are unable to divide, meaning they are essentially nitrogen-fixing organelles within plant cells. Some strains of *S. meliloti* can mitigate the effects of NCR peptides and prevent their enslavement by producing a peptide degrading enzyme such as HrrP [47].

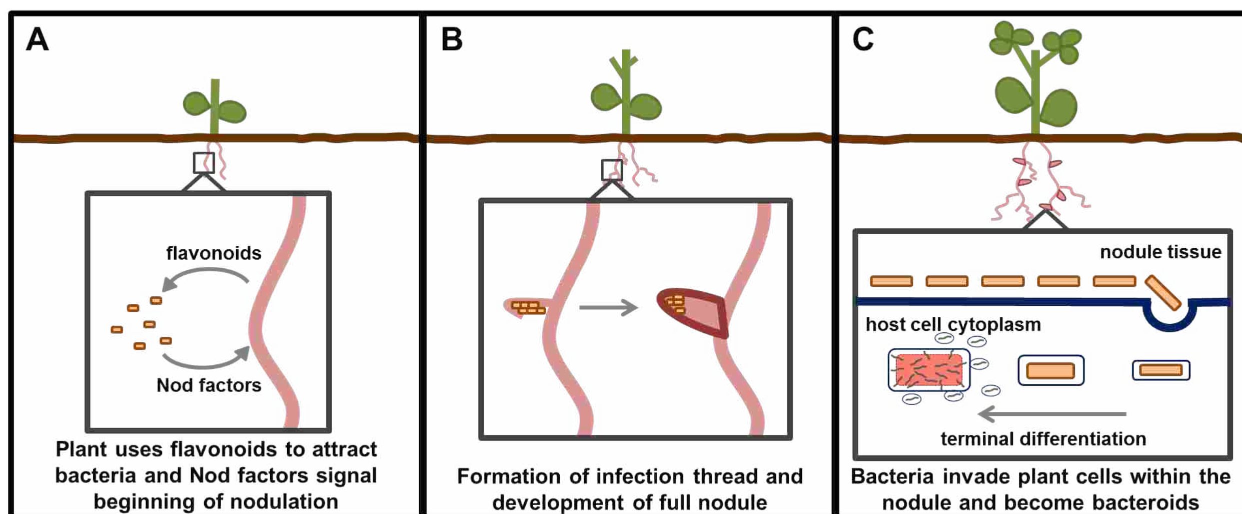


Fig. 13. Nodulation in *M. truncatula*. A. Plant releases flavonoids to attract bacteria. Bacteria signal back using Nod factors, which begin the nodulation process. B. Root hairs curl around bacteria to begin forming an infection thread. C. Inside of a fully developed nodule, bacterial cells invade plant cells. Once in the host cell cytoplasm, bacteria are inundated with NCR peptides and undergo terminal differentiation.

Table 3. NCR mature peptide sequences

NCR peptide	Mature Peptide Sequence	Number of cysteines
NCR055	IRIHCKDDFDCIENRLQVGCRLQREKPRCVNLVCRCLRR	6
NCR111	SDIPNNSNRNSPKEDVFCNSNDDCPTILYYVSKCVYNFCEYW	4
NCR145	EADTSCHSFDDCPWVAHHYRECIEGLCAYRILY	4
NCR224	KDLPFNICEKDEDCLEFCAHDKVAKCMLNICFCF	6
NCR235	DTDPFAFCIKDSNCGQDLCTSPNEVPECRLLKCQCIKS	6
NCR317	LPISCKDHFECRRKINILRCIYRQEKPMCINSICTCVKLL	6
NCR391	NCTFIGFQDNPCKTDKDCRKVRGVNLRCRNGHCVMILQ	5
NCR466	HSFLPCVTKDDCAYDECISPRKPTCYLETCHCL	6
NCR514	SPVLCQRNYECYEQICLPPKKHWCNILELVRINGFYLGLCACI	6
NCR519	YVPCITVADCPPNTWFKIYRCEKGICRYHKLWIV	4
NCR524	YVIIIFLSLFLVATNIEGKLFYRFQISFFYFGHNILLYFCYLILFSF FITTAFFRCKNDFDCVHKRCRGPMRAKCISKAICKCRLAFTLK	7

Over 700 NCR peptides are produced in nodules. These few were selected to demonstrate a variety of sequence length, amino acid composition, and number of cysteine residues. Cysteines are underlined for emphasis.

As is seen in Table 3, NCR peptides are diverse in sequence, related only by the presence of multiple cysteines. Over 700 have been annotated, and yet relatively few have been well characterized. Of those that have been characterized, some have been found to exhibit antimicrobial activity [49, 75-77]. In an effort to better classify NCR peptides, we have developed a rapid and cost-effective assay for antimicrobial activity in *E. coli*. While it has thus far only been used for a handful of NCR peptides, it has the potential to be expanded to other peptides, allowing a better understanding of this unique symbiosis.

The assay depends on a plasmid with an exquisitely inducible promoter. The NCR peptide to be tested is placed under the control of the promoter, and transformed into *E. coli.* To test for antimicrobial effects, the NCR peptide-producing strain is grown both induced and uninduced. If the NCR peptide is antimicrobial, the induced culture will have inhibited growth compared to the uninduced culture. For this assay, we use the rhamnose promoter (P_{rha}), as shown in Fig. 14A.

Adding additional complexity is the consistent presence of cysteines in NCR peptides. Due to the observation that these residues typically occur either 4 or 6 times in each peptide (Table 3), we hypothesized that disulfide bonding would play an important role in peptide function. Thus the assay uses three versions of plasmid scaffolds to address this. One expresses the NCR peptide in the cytoplasm (Fig. 14B), where disulfide bonds cannot occur in *E. coli*. The second expresses the peptide in the periplasm (Fig. 14C) using the signal peptide from *E. coli* protein DsbA. The third scaffold expresses the peptide tethered to the inner membrane (Fig. 14D) using the signal peptide from YiaD, which has three transmembrane regions. This leads to a tethered peptide facing the periplasm. These two platforms allow disulfide bonds to occur, with the tethered peptide additionally concentrated at the membrane. As the membrane is a likely

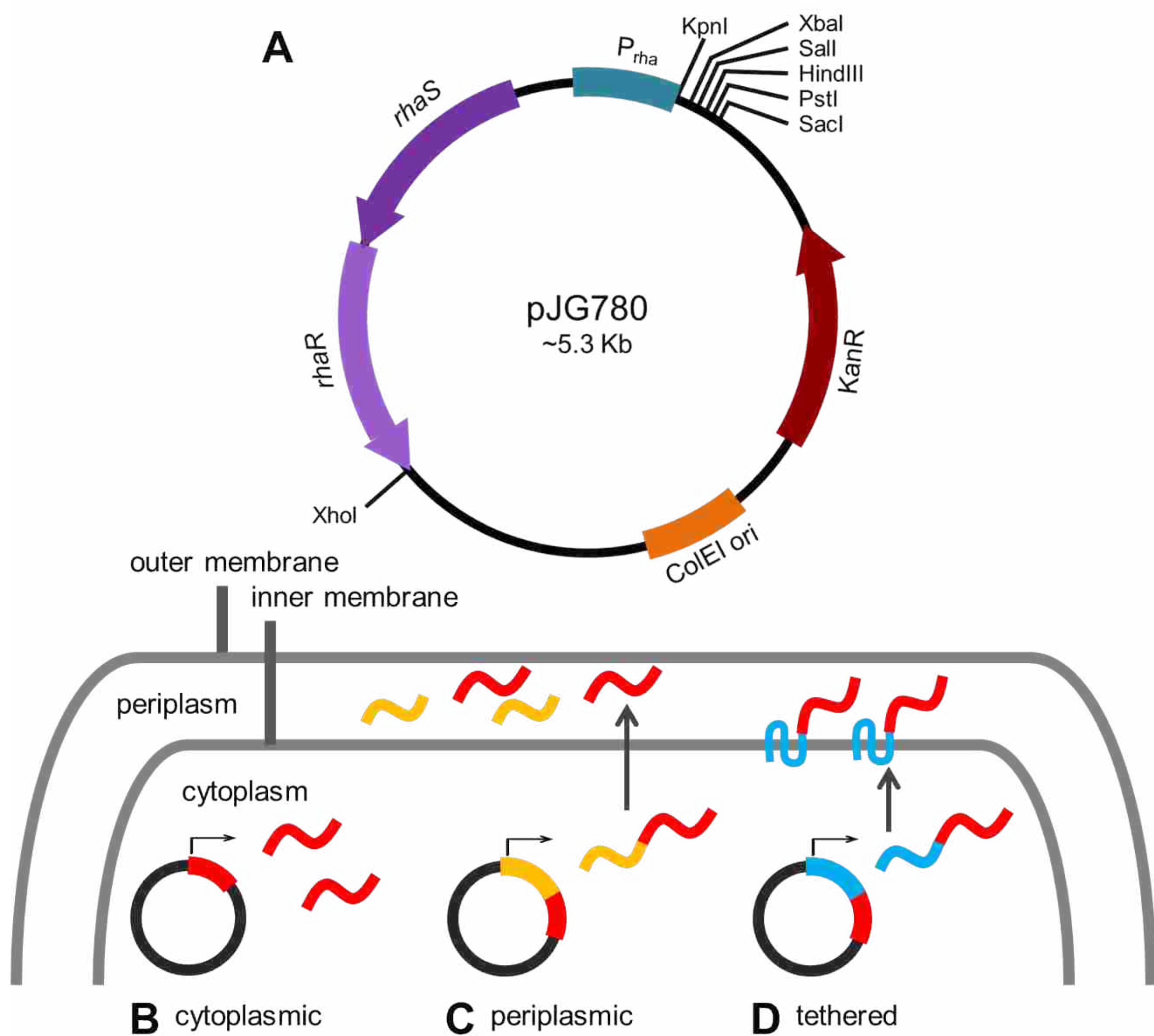


Fig. 14. NCR peptide antimicrobial activity assay. A. Parent vector for NCR assays. B. Depiction of NCR peptides (red) expressed in the cytoplasm. C. Depiction of NCR peptides (red) expressed in the periplasm. Peptides are appended to the signal peptide from DsbA (orange), which signals delivery to the periplasm. D. Depiction of NCR peptides tethered to the inner membrane. Peptides are appended to the signal peptide from YiaD (blue), which has three transmembrane regions. The peptide is tethered to the inner membrane facing into the periplasm.

target of antimicrobial activity, it is expected that this construct will increase toxicity of an NCR peptide.

Seven NCR peptides (Table 4) were tested in each construct in this initial proof of concept experiment. Each of the seven peptides has been previously characterized. NCR247 is known to have antimicrobial properties, impacted in part by its disulfide bonds [50], and it inhibits cell division and protein synthesis [78]. NCR169 and NCR211 are known to play crucial roles in symbiosis [76, 79]. NFS1 and NFS2 each have two versions of the gene that result in either successful symbiosis (Fix+) or defective symbiosis (Fix-) [77, 80]. These two peptides act in a strain-specific manner, and do not result in the same phenotypes when different combinations of plant and bacterial strains are used. Due to their known roles in symbiosis, these seven peptides were selected for testing in this assay.

3.2 Methods

3.2.1 Molecular Cloning Protocols

Genes corresponding to DsbA and YiaD signal peptides were amplified from MG1655 genomic DNA using primers oED020 and oED021 or oED092 and oED093, respectively. The wild-type *phoA* gene was amplified from MG1655 genomic DNA using oED089 and oED090, with an additional insert (sans signal peptide) made using oED091 and oED090. NCR peptide inserts were made by designing overlapping primers, codon-optimized for *E. coli*, annealed and extended using a PCR machine to generate a blunt-end double-stranded DNA insert (see full protocol in Appendix II). Inserts were ligated into pJG780, pJG837, and pJG987 following digestion with KpnI and SacI (for pJG780) or PstI and SacI (for pJG837 and pJG987). Ligation

mixtures were transformed into DH5α and cloned plasmids were sequence-verified using Sanger sequencing. See Appendix I for primer sequences

3.2.2 Analysis of antimicrobial activity

Strains were grown in 4 ml-cultures of LB with kanamycin (15μg/ml), without inducer or with 0.3% rhamnose (Rha). Growth was measured by OD_{600} at 6, 12, and 24 hours following induction.

3.2.3 Osmotic shock and PhoA assays

To prepare samples for PhoA assays, pJG966 and pJG996 50 ml-cultures of LB with 0.3% rhamnose were grown for 6 hours at 37°C, with shaking. Grown cultures were pelleted, supernatant removed and pellets frozen at -80°C for at least 1 hour. Pellets were thawed on ice, and resuspended in periplasmic isolation buffer (see Appendix II). The mixture was then vortexed in 10 second-increments, 6 times each. Four volumes of chilled ddH_2O were added to induce osmotic shock and release the periplasmic contents. Following 10 minute incubation on ice, cell debris was pelleted. The pellet (resuspended in periplasmic isolation buffer) and supernatant were diluted 100-fold prior to PhoA detection assays. For the PhoA assay, diluted sample was combined with 1 M Tris pH 8.0, 0.1% SDS, chloroform, and 5mg/ml p-Nitrophenyl Phosphate (PNPP) for a final concentration of 0.45 μg/ml. The sample mixture was incubated at 37°C for ~30 minutes, until it appeared yellow. The reaction was stopped with 1 M KH_2PO_4. Samples were briefly centrifuged prior to OD_{420} measurements.

3.3 Results

3.3.1 Verification of YiaD membrane tethering

While the signal peptide from DsbA has previously been used to export proteins to the periplasm [81], we wanted to ensure that the tethering region from YiaD would localize peptides

Table 4. NCR peptides tested

NCR peptide	Mature Peptide Sequence	Number of cysteines
NCR247	RNGCIVDPRCPYQQCRRPLYCRRR	4
NCR169	EDIGHIKYCGIVDDCYKSKKPLFKIWKCVENVCVLWYK	4
NCR211	DRECDTDTECQKKFPGVNAHHLWCDNGNCVSYPK	4
NFS1Fix-	ASPFYCVDDDYFCFGLCLPPMIDHCTLRGQCICITISTEVES	6
NFS1Fix+	ASPFYCDEDDYFCFGLCLPPMIDHCTLGGQCICITISTEVES	6
NFS2Fix-	KEECTYAADCYKRYPRWSLLPNYCIEGSCYSDFLNSGKKYLSP	4
NFS2Fix+	NEECTCAADCYKRYPRWSLLPTYCIEGSCYSDFLNSGKKYLSP	5

Differences between Fix- and Fix+ versions of NFS1 and NFS2 are underlined.

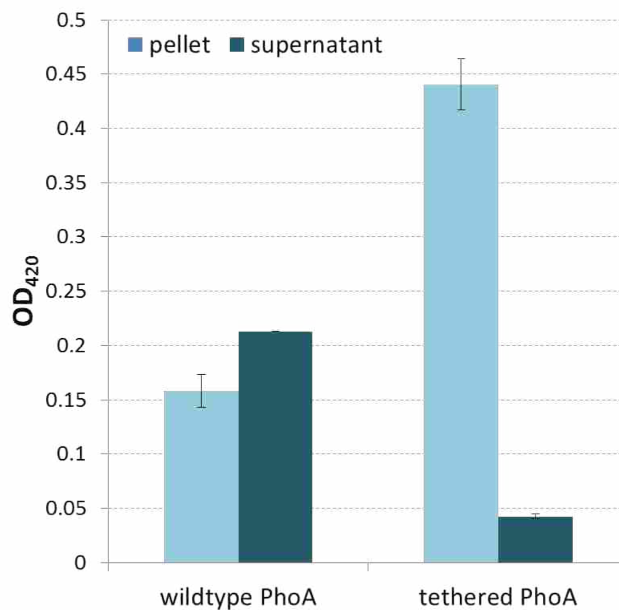


Fig. 15. PhoA activity assay. PhoA activity was detected by measuring OD_{420} of samples containing PNPP substrate. Free periplasmic activity occurs primarily in the supernatant, while tethered or cytoplasmic activity occurs in the pellet. Error bars show standard error.

to the periplasm yet remain tethered to the inner membrane. This was done by fusing the YiaD tether to *phoA*, a reporter gene previously used to determine protein topology [82]. The YiaD fusion protein was compared to the wild-type PhoA which, while located in the periplasm, would be released with the free periplasmic contents leaving tethered proteins with pelleted cells. Each strain was grown in duplicate, and following separation of periplasmic proteins from the rest of the cell by vortexing and osmotic shock, both pellet and supernatant were tested for PhoA activity. The supernatant is predicted to contain free periplasmic proteins and the pellet is predicted to contain membrane-bound proteins and cytoplasmic proteins. Thus, for wild-type PhoA, high activity is expected in the supernatant and low or no activity is expected in the pellet. If the YiaD tether worked as expected, higher activity is predicted in the pellet with low or no activity in the supernatant. While periplasmic protein extraction was not complete for wild-type PhoA, when the YiaD tether was added, there was a great increase in activity in the pellet as compared to the supernatant (Fig. 15).

3.3.2 Antimicrobial activity of NCR peptides

Using the three previously described expression platforms, seven NCR peptides were tested for antimicrobial activity (Table 4). Each was expressed in the cytoplasm, the periplasm, and tethered to the inner membrane. To determine NCR peptide toxicity, cells expressing NCR peptides were grown in LB for 12 hours prior to measuring OD_{600}. When peptides were expressed in the cytoplasm, no significant difference was detected in cell density (Fig. 16A). When peptides were expressed in the periplasm, NCR169, NFS1Fix-, and NFS1Fix+ showed decreased growth when peptide expression was induced (Fig. 16B). Interestingly, when peptides were tethered to the inner membrane, all showed decreased growth (Fig. 16C). It is important to note that toxicity was also observed at 6 hours, although control cultures were not saturated at

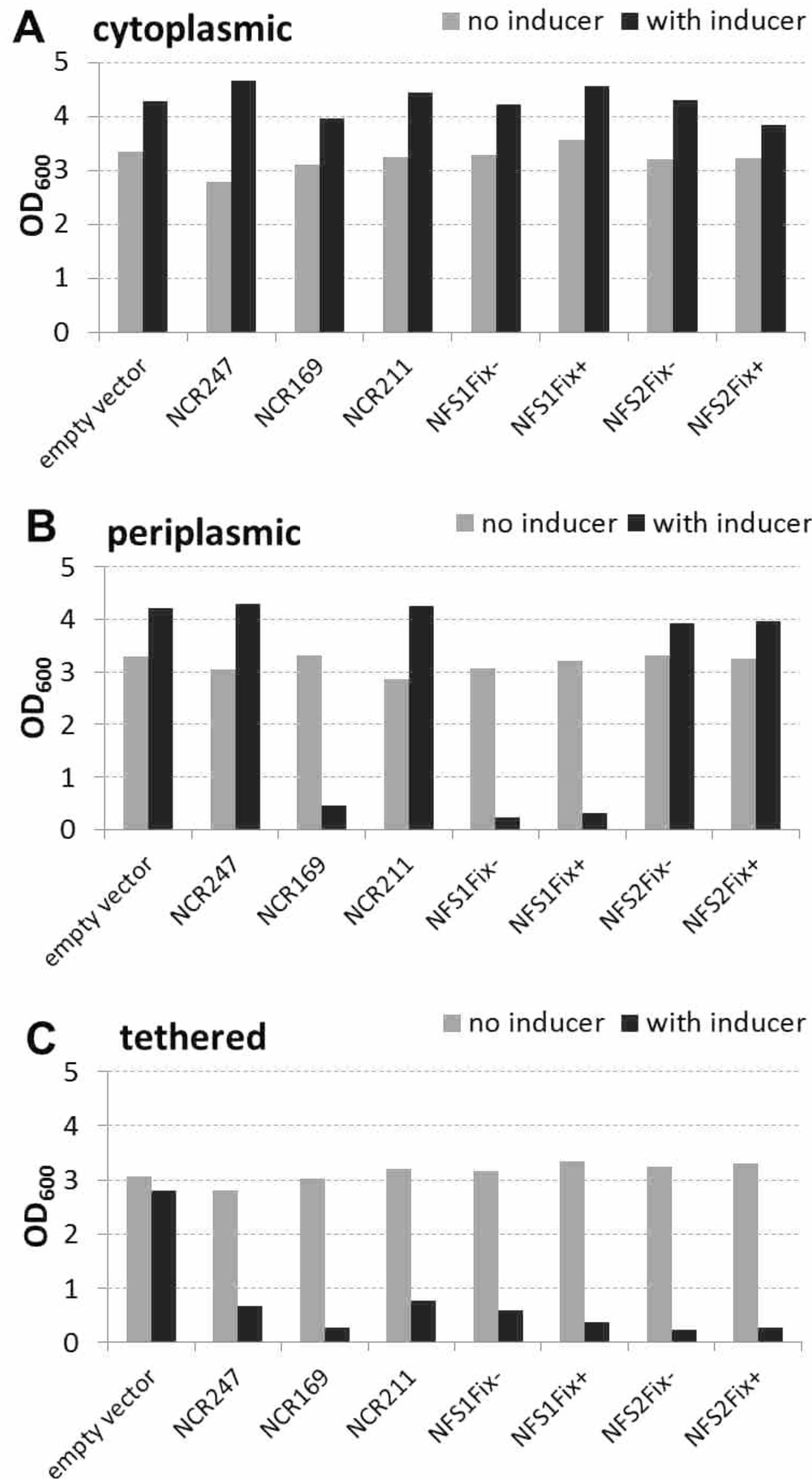


Fig. 16. NCR peptide antimicrobial activity assay. A. OD_{600} with and without inducer when NCR peptides are expressed in the cytoplasm. B. OD_{600} with and without inducer when NCR peptides are expressed in the periplasm. C. OD_{600} with and without inducer when NCR peptides are tethered to the inner membrane.

this time. Additionally, after 24 hours, many induced cultures appeared to overcome toxicity. It is not clear whether this is due to mutations allowing resistance to toxicity or due to depletion of the inducer over time.

3.4 Discussion

While no antimicrobial activity of NCR peptides, even those known to be toxic when applied exogenously, was observed when peptides were expressed in the cytoplasm, activity was observed in both instances when peptides were localized to the periplasm. One possible explanation is the ability of peptides to form disulfide bonds in the periplasmic environment and not in the cytoplasm. It was also interesting to note that all tethered peptides showed some level of antimicrobial activity. The increased toxicity associated with membrane tethering suggests that these peptides are membrane-active, and localizing them to the membrane environment enhances their cytotoxicity. While only a handful of peptides were tested in this proof of concept experiment, there are hundreds more NCR peptides that can be tested for toxicity. Additionally, NCR peptides could also be tested for antimicrobial activity in their native host, *S. meliloti*. While much remains to be done to characterize NCR peptides, the development of this assay will allow a large fraction of NCR peptides to be assayed for toxicity using the periplasmic-localized platforms. This will help us to determine how much stress is imposed on developing bacteroids in the symbiotic relationship between *M. truncatula* and *S. meliloti*.

CHAPTER 4: Massively parallel peptide activity screening

4.1 Background

In the past, the drug discovery industry has relied primarily on compounds derived from nature. While many of these routes have been thoroughly investigated, there are still many ways that nature can be manipulated to produce antimicrobial compounds. A biologically convenient compound is peptides, due to enormous variety of potential chemical properties and functions. Because peptides are genetically encoded, straightforward DNA-based methods can be employed to encode their production in living cells. There are several known antimicrobial peptides (AMPs) throughout nature, including the NCR peptides already discussed. Previous research has sought to optimize antimicrobial activity in specific peptides [57, 58, 60, 61], but have generally been limited by the number of peptides that can be chemically synthesized for analysis.

To expand on the set of characterized peptides, increasing possible structures and targets of antimicrobial activity, a platform called Surface Localized Antimicrobial Display (SLAY) has been developed that allows for simultaneous screening of hundreds of thousands of novel peptides for antimicrobial activity and cell permeability [66]. In this method, random peptides are tethered to the outside of cells, allowing peptides to penetrate the cell membrane. If the peptide is toxic, those cells drop out of the population. This self-killing phenotype can be detected by high-throughput sequencing. Unfortunately, this technique is restricted by the length of the peptide tether, which limits the potential cellular targets of the peptides.

The platform described here provides a cytoplasmic version of high-throughput AMP discovery. This platform, known as PepSeq takes advantage of high-throughput sequencing to determine potential antimicrobial activity of random peptides. In this system, a plasmid encoding

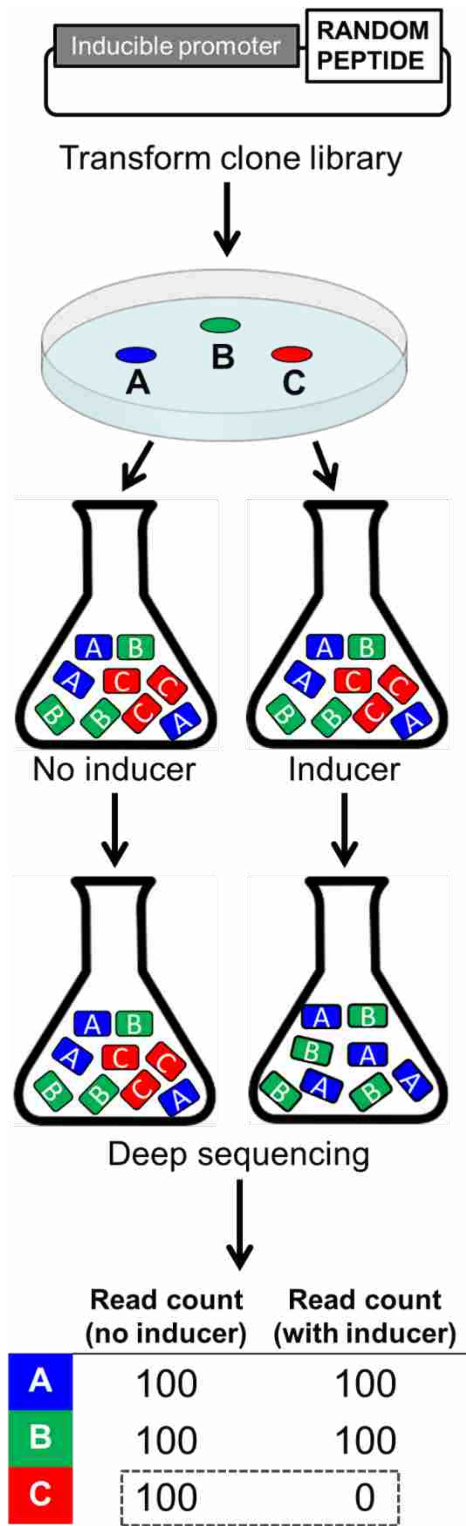


Fig. 17. PepSeq workflow. A bacterial population is grown in uninduced and induced conditions. Cells that fail to grow when induced, as determined by DNA sequencing, are classified as hits for antimicrobial activity.

a random segment of DNA under the control of an inducible promoter is transformed into *E. coli,* leading to hundreds of thousands of colonies. Each of these colonies represents a unique peptide-encoding region. These colonies are pooled to form a library. This library is then grown in parallel under two conditions, induced and uninduced. After a period of growth, each culture is sequenced using Illumina technology (Fig. 17).

Following sequencing, we use a computational tool to count the number of sequence reads for each peptide-encoding sequence. In the uninduced population, where no peptides are being produced, we anticipate that the entire spectrum of peptide-encoding sequences will be present. When peptide production is induced, those that encoded toxic peptides will drop out of the population, leading to reduced sequence counts in the induced population (Fig. 17). In this way, we anticipate that we can identify hundreds of potentially toxic peptides in a single experiment. These peptides can then be further studied, and the platform is highly adaptable to various peptide scaffolds.

4.2 Methods

4.2.1 Peptide library cloning

The random peptide library was generated by ligating inserts from degenerate primers oED110 and oED111 (Fig. 18A) into pED021 using BamHI and SacI. Clones were transformed into electrocompetent MG1655 and plated to obtain libraries of ~280,000 colonies. Colonies were scraped into LB + 15% glycerol and shaken for 30 minutes at 30°C to ensure homogeneity in frozen aliquots. Library aliquots were stored at -80°C.

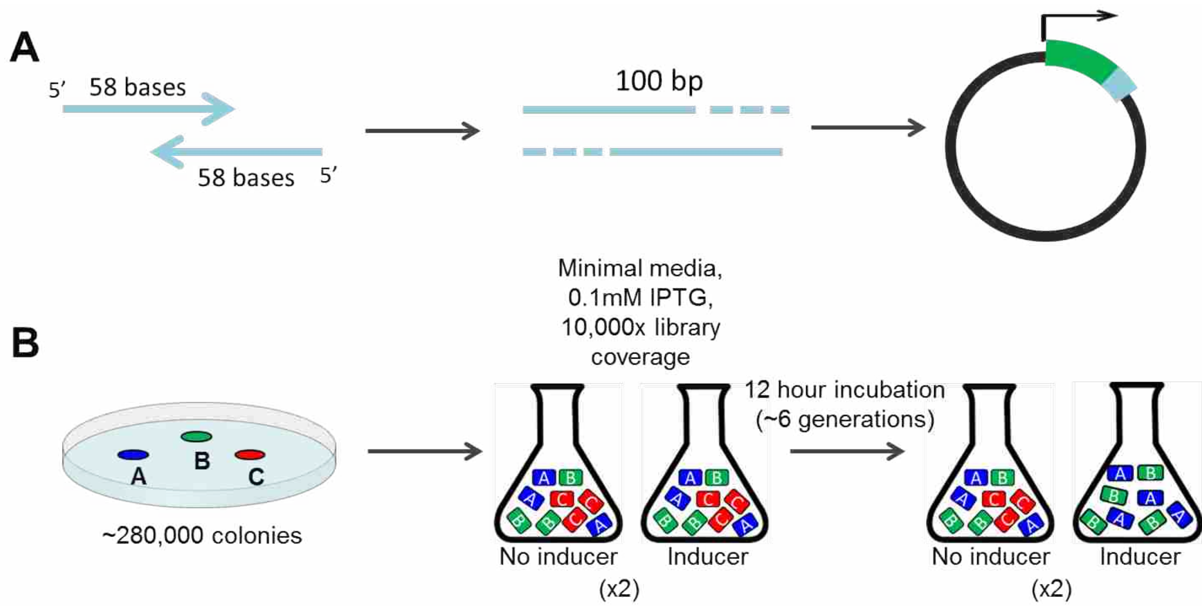


Fig. 18. PepSeq methods. A. Library construction. B. Library selection

4.2.2 Library selection

Libaries were grown in 40 ml of minimal medium (M9M, see Appendix II) with ampicillin (50 µg/ml) and 1:80 dilution of LB total (including the inoculum volume), induced (0.1 mM IPTG) and uninduced for 12 hours, in duplicate, prior to preparing for Illumina sequencing (Fig. 18B). Dilutions of the library and dilutions of the 12 hour-cultures were plated to determine colony forming units (CFUs) and estimate number of generations.

4.2.3 Preparing samples for Illumina sequencing

Following selection, samples were centrifuged and peptide-encoding plasmids were isolated using a Zymo Research Miniprep Plasmid Classic kit. Illumina adapters and barcodes were added by PCR using primers oED112-118. See Appendix II for primer sequences. PCR samples were run on a low-melt gel and subsequently extracted using a QIAquick Gel Extraction Kit from Qiagen.

4.2.4 Read counting, quality control, and motif finding

Read counts were obtained using a Python script written by an undergraduate student (Michael Frampton) in the bioinformatics department. The resulting peptides were trimmed according to total number of counts in the reference cultures, where those with a reference sum of less than 50 counts were removed. The trimmed dataset was used in the motif finder tool written by Alex Erikson.

4.2.5 Individual toxic peptide cloning

Inserts were made using the appropriate overlapping primers (see Appendix II), similar to degenerate primers oED110 and oED111, which were annealed and extended using a PCR machine to generate a blunt-end double-stranded DNA insert. The resulting insert was ligated

into pED021 following digestion with BamHI and SacI. All resulting plasmids were transformed directly into *E. coli* strain MG1655 (lab strain D050).

4.2.6 Toxic peptide confirmation

Toxic peptides were confirmed by growing strains in 4 ml of M9M with ampicillin (50 μg/ml) and 1:80 dilution of LB (including volume of inoculum), induced (0.1 mM IPTG) and uninduced, for 24 hours. OD_{600} and fluorescence measurements of a 1:10 dilution were taken every 2 hours until 12 hours, and a final measurement was taken at 24 hours. Fluorescence was determined using a BioTek plate reader. Final arbitrary fluorescence units (AFUs) were normalized to the media blank. A small amount of culture was additionally removed for serial dilutions and plating to determine CFUs at 6, 12, and 24 hours.

4.3 Results

4.3.1 Library and experimental design

For this version of PepSeq, the degenerate codon 'NDT' was used. As seen in Fig. 19C, this codon allows a variety of chemical properties in amino acids while avoiding stop codons. Additionally, the 8-residue random peptide was appended to the end of msfGFP to provide both stability and an ability to quantify peptide expression when induced (Fig. 19A, B**)**. Library complexity was estimated to be ~280,000 unique peptides, with a library titer of 5.3×10^{10} CFU/ml. To effectively identify toxic peptides, the number of generations the library underwent was crucial. Based on previous attempts, too many generations lead to many peptides with a toxicity score of 0, making it difficult to distinguish levels of toxicity. Too few generations may not allow enough time for toxic peptide-producing cells to disappear from the population, again making it difficult to distinguish toxicity. Ultimately, the library was grown with 10,000x

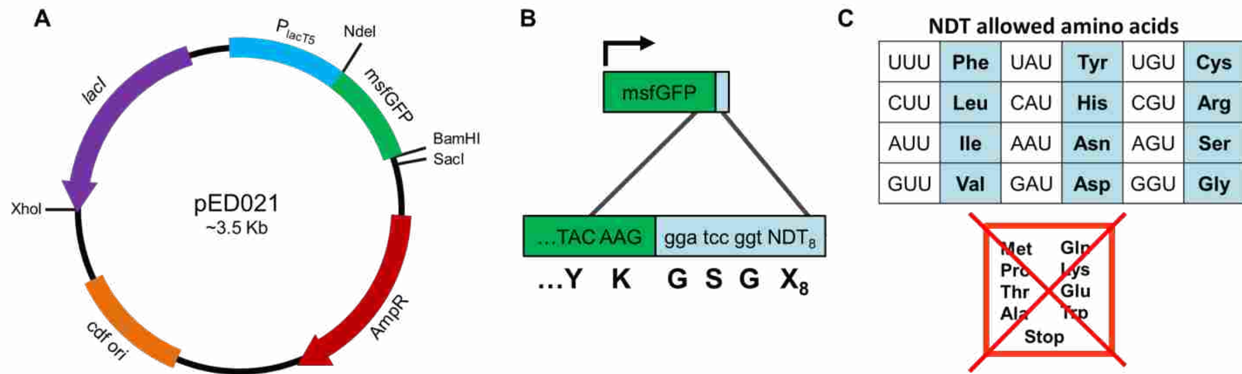


Fig. 19. Peptide expression system and library design. A. Library vector with IPTG-inducible P_{lacT5} and msfGFP carrier. B. Random peptide (blue) appended to msfGFP using degenerate codon NDT. C. Amino acids allowed by NDT (blue) with corresponding codons. Prohibited amino acids are indicated with a red box.

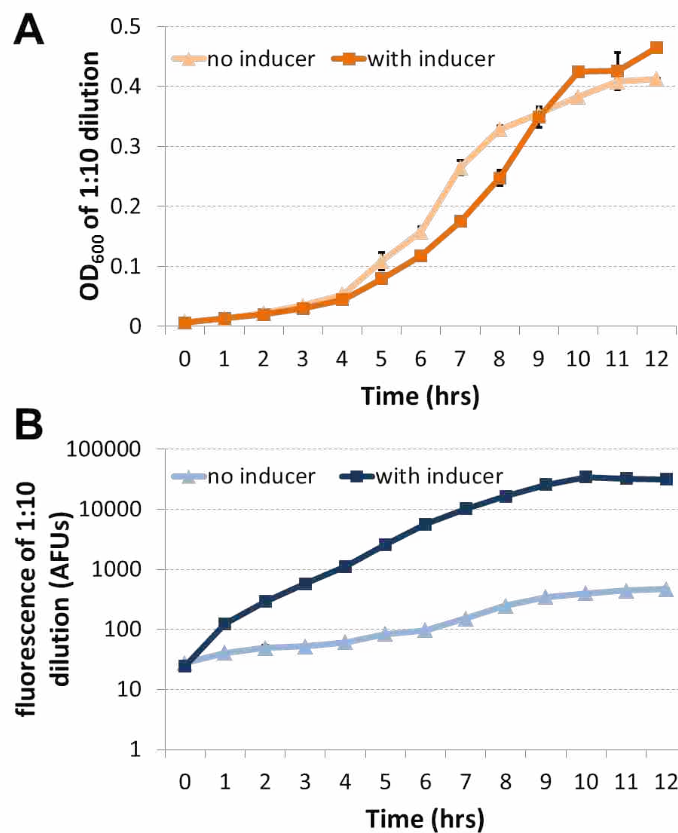


Fig. 20. Library growth and peptide expression in M9M. The peptide library was grown in 40 ml of M9M, induced and uninduced, for 12 hours to determine number of generations and verify peptide expression by measuring fluorescence. Each condition was grown in triplicate. Error bars show standard error. A. The OD_{600} of a 1:10 dilution was taken every hour up to twelve hours. B. Fluorescence of a 1:10 dilution of the same cultures was taken every hour up to twelve hours.

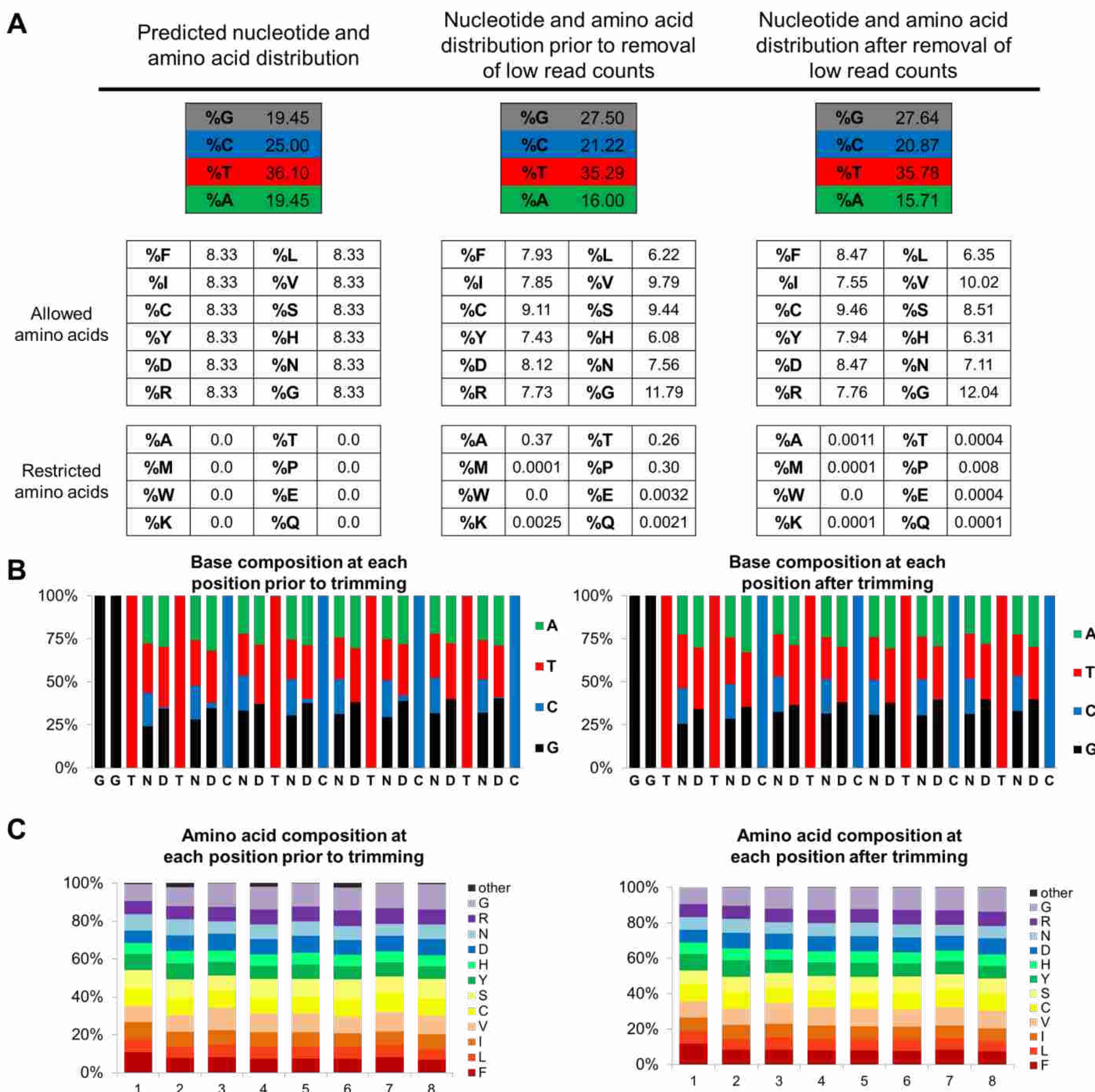


Fig. 21. PepSeq dataset quality control. A. Overall base and amino acid composition. B. Base composition at each position is shown before and after removing low read counts from the dataset. C. Amino acid composition at each position is shown before and after removing low read counts from the dataset.

coverage, in duplicate, induced and uninduced, in M9M-glucose for 12 hours, undergoing an estimated 5-6 generations. No growth difference was detected when inducer was added (Fig. 20).

4.3.2 PepSeq data quality control

Following read counting, 288,870 unique peptides were identified. To ensure the quality of the toxic hits, an additional 111,371 peptides were discarded due to low read counts in reference cultures. This left 177,139 peptides in the final dataset where, on average, each peptide has 100 counts in each reference culture. Overall base and amino acid composition were analyzed before and after data trimming (Fig. 21A). Additionally, base composition at each of the 24 positions was analyzed before and after removing peptides with low read counts (Fig. 21B). Amino acid composition at each of 8 positions was also determined before and after data trimming (Fig. 21C). While the data prior to trimming was close to what was anticipated, occurrences of unexpected nucleotides and thus amino acids decreased after peptides with low read counts were removed.

4.3.3 Results

After low quality peptides were removed from the dataset, 177,138 peptides remained. For each peptide, a hydrophobicity score is calculated, where a higher number indicates a higher number of hydrophobic residues. The charge of each peptide was also determined. To analyze toxicity, the total number of counts in the induced culture is divided by the total number of counts in the reference (uninduced) culture. This generates a toxicity score where a low score indicates high toxicity, and a high score indicates low toxicity. A subset of the data is shown in Table 5, to demonstrate the range of toxicity scores. Peptides with toxicity scores of less than 0.1 are shown in Table 6, with peptides having toxicity scores of 0.2 or less in contained in Appendix III.

Table 5. Peptides from PepSeq dataset

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATAACTTTATCGGTTTCTTTCGC	YNFIGFFR	5	4	80	77	1	1	0.01
AGTCACTTTTTCTATGGCGTTTTC	SHFFYGVF	7	2	47	43	1	1	0.02
CGTTTCCTTTACTTTGGCCATCTC	RFLYFGHL	6	6	49	38	1	2	0.03
GGTTTCTATTACGTTTGCATTATC	GFYYVCII	11	0	50	57	4	1	0.05
TATTACCGTATCATTTGCGTTTTC	YYRIICVF	9	4	40	22	1	2	0.05
TGTCTCTATATCGTTCGCTTTCTC	CLYIVRFL	10	4	64	69	1	6	0.05
CTTTACATTATCCGTCTCATTAGC	LYIIRLIS	8	4	59	55	7	11	0.16
TTTAGCTGTTACGTTGTCCGTTTC	FSCYVVRF	7	4	97	86	23	20	0.23
TTTCGCTTTTACTATTTCCGTTGC	FRFYYFRC	5	8	27	24	5	7	0.24
TATTACTTTGTCCTTATCTTTGAC	YYFVLIFD	10	-4	77	73	24	13	0.25
TTTTACCGTCACTATTTCGTTTTC	FYRHYFVF	7	6	118	113	54	44	0.42
CATAACATTTTCTTTTACTATGTC	HNIFFYYV	7	2	58	48	25	20	0.42
TATGGCGTTTTCTTTGGCCGTGGC	YGVFFGRG	5	4	136	115	75	52	0.51
GTTAACTATCGCCATAGCAATCTC	VNYRHSNL	-3	6	79	72	57	56	0.75
GATAACTGTATCCGTGGCTTTGGC	DNCIRGFG	-1	0	76	57	56	57	0.85
TGTCGCGTTGGCTTTAACTGTGTC	CRVGFNCV	4	4	109	110	113	105	1.00
ATTGTCGTTGTCAATATCAGTAAC	IVVVNISN	5	0	156	149	161	160	1.05
GATTGCCTTGACCATTACGTTAGC	DCLDHYVS	0	-6	76	73	230	208	2.94
GGTCACGGTCACTGTTTCTATTGC	GHGHCFYC	3	4	150	111	436	454	3.41
CGTGGCAGTTTCGTTAGCAGTGGC	RGSFVSSG	-1	4	23	27	83	102	3.70

20 peptides from throughout the dataset are shown here, with their accompanying hydrophobicity score (HPHOB), charge, counts in each culture (REF1, REF2, IND1, IND2, where REF indicates uninduced cultures and IND refers to induced cultures), and toxicity score. Toxicity score is calculated as the sum of IND1 and IND2 divided by the sum of REF1 and REF2. While a range is shown here, the majority of peptides have a toxicity score between 0.75 and 1.25.

Table 6. Toxic peptides identified by PepSeq

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATAACTTTATCGGTTTCTTTCGC	YNFIGFFR	5	4	80	77	1	1	0.01
AGTCACTTTTTCTATGGCGTTTTC	SHFFYGVF	7	2	47	43	1	1	0.02
CGTTTCCTTTACTTTGGCCATCTC	RFLYFGHL	6	6	49	38	1	2	0.03
GGTTTCTATTACGTTTGCATTATC	GFYYVCII	11	0	50	57	4	1	0.05
TATTACCGTATCATTTGCGTTTTC	YYRIICVF	9	4	40	22	1	2	0.05
TGTCTCTATATCGTTCGCTTTCTC	CLYIVRFL	10	4	64	69	1	6	0.05
CATTACTATATCATTTACTTTATC	HYYIIYFI	10	2	88	77	5	4	0.05
CGTTACTATTACATTATCTTTGTC	RYYYIIFV	9	4	97	81	1	9	0.06
TTTTGCCTTTTCCGTGGCTTTGGC	FCLFRGFG	7	4	123	125	7	7	0.06
TATAACATTTTCTATATCCTTTAC	YNIFYILY	9	0	53	48	2	4	0.06
TTTCGCATTTTCGTTTTCCTTATC	FRIFVFLI	12	4	27	23	2	1	0.06
TTTTTCCGTATCGTTCTCGTTGTC	FFRIVLVV	12	4	42	24	3	1	0.06
TATCACCTTGTCTATCGCATTTAC	YHLVYRIY	6	6	40	42	2	3	0.06
CATCGCGTTGTCTATGTCTTTATC	HRVVYVFI	8	6	37	44	3	2	0.06
CGTCACTTTTTCTATGGCGTTTTC	RHFFYGVF	6	6	533	479	25	39	0.06
CGTAACTTTTTCTATGGCGTTTTC	RNFFYGVF	5	4	31	30	4	0	0.07
CGTTTCCGTATTCTTATCCTTCTC	RFRILILL	8	8	24	34	2	2	0.07
GATTTCGATGGCGTTGGCATTAGC	DFDGVGIS	1	-8	165	153	13	9	0.07
TGTTTCTATGTCATTCACATTATC	CFYVIHII	11	2	37	33	4	1	0.07
TTTCGCATTTTCTATCGCATTTTC	FRIFYRIF	7	8	80	88	5	7	0.07
TATTACTATCTCGGTTTCAATTTC	YYYLGFNF	7	0	45	36	4	2	0.07
CGTTACTATATCATTTGCCTTGTC	RYYIICLV	9	4	41	26	3	2	0.07
TTTCGCTGTATCATTGTCCGTCTC	FRCIIVRL	7	8	79	80	8	4	0.08
TATCGCCTTATCATTTTCTTTTAC	YRLIIFFY	10	4	43	46	5	2	0.08
GATTACGTTATCATTGTCTATCGC	DYVIIVYR	6	0	159	157	7	18	0.08
GTTCGCTTTTACGTTTACCTTATC	VRFYVYLI	10	4	46	29	3	3	0.08
CATCGCTTTTTCTATGTCCTTTTC	HRFFYVLF	8	6	81	68	6	6	0.08
TATTACTATGTCTTTTACCTTTGC	YYYVFYLC	11	0	26	35	3	2	0.08
TGTTACCGTCTCTATATCATTTTC	CYRLYIIF	9	4	230	221	24	13	0.08
CTTCGCGTTTACCTTGTCTTTGTC	LRVYLVFV	11	4	55	30	3	4	0.08
AATTACCTTTTCTATTGCTTTCTC	NYLFYCFL	9	0	110	96	9	8	0.08
ATTTACTTTATCTATCTCCGTCGC	IYFIYLRR	6	8	37	35	3	3	0.08
GATTACTATATCGTTTTCCTTGTC	DYYIVFLV	10	-4	49	46	5	3	0.08
TATTACCGTATCTATTGCTATTTC	YYRIYCYF	7	4	36	35	2	4	0.08
CGTTTCATTATCATTCGCCTTCAC	RFIIIRLH	5	10	31	28	5	0	0.08
CGTATCTATTACATTATCATTTGC	RIYYIIIC	9	4	36	45	2	5	0.09
TATAGCTGTTACCTTATCCTTTTC	YSCYLILF	10	0	40	49	5	3	0.09
ATTCGCCATGTCTATATCATTATC	IRHVYIII	8	6	48	52	4	5	0.09
ATTGTCATTATCCGTTGCGTTCGC	IVIIRCVR	7	8	55	33	5	3	0.09
AGTCTCTATATCTTTTTCATTTGC	SLYIFFIC	11	0	34	31	2	4	0.09
CGTCGCATTATCATTGTCTTTCTC	RRIIIVFL	8	8	64	76	9	4	0.09
ATTTACGTTTTCAGTTTCGGTGTC	IYVFSFGV	10	0	85	76	8	7	0.09
TATCGCGTTATCGTTTTCTTTTAC	YRVIVFFY	10	4	23	30	1	4	0.09
TATCGCCTTTTCCTTGTCTATTTC	YRLFLVYF	10	4	225	229	23	20	0.09

All peptides with a toxicity score less than 0.1 are depicted here.

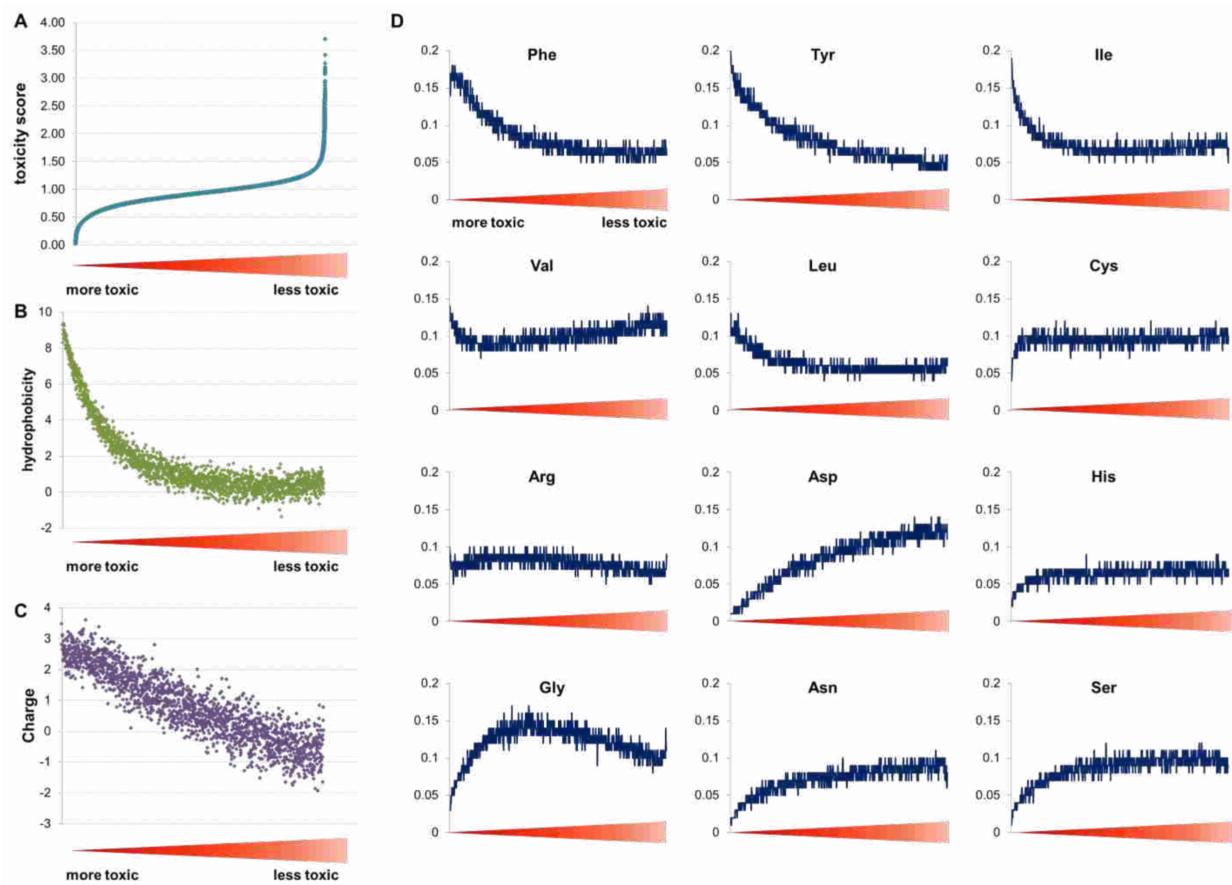


Fig. 22. PepSeq data summary. A. Peptides were ranked by toxicity (low score to high score) prior to graphing. B. Sliding window average of hydrophobicity of peptides ranked by toxicity. Window size=100, step size=100. C. Sliding window average of charge of peptides ranked by toxicity. Window size=100, step size=100. D. Sliding window average of amino acid composition of peptides ranked by toxicity. Window size=200, step size=200. Darker red indicates higher toxicity, while increasing width indicates higher toxicity score.

Several analyses were performed to determine general patterns of toxicity (Fig. 22). The average hydrophobicity and charge of toxic peptides is high, as seen in Fig. 22B and C. Additionally, the amino acid composition of the peptides was analyzed (Fig. 22D). Phenyalanine, tyrosine, and isoleucine appear to be associated with toxicity, while glycine and aspartate appear to be inversely correlated with toxicity. The inverse correlation of aspartate with toxicity is of particular note, as it leads to an increased overall charge in toxic peptides (Fig. 22B).

4.3.4 Peptide confirmation

To confirm whether our predicted antimicrobial peptides were toxic, several peptides were individually cloned for confirmation. The first set of peptides tested individually for toxicity was the top six ranked toxic peptides from the dataset (Table 7). While the majority of toxic peptides in our dataset fall into the category of hydrophobic and cationic, there are also some peptides with more unusual properties. Four of these unusual peptides were selected for additional testing, along with one canonical peptide (Table 8).Both culture density and fluorescence were measured to determine toxicity (Fig. 23). A nontoxic peptide from the dataset showed no decrease in growth or expression compared to vector controls (Fig. 23A). Of the top six peptides, two showed decreased growth when induced (Fig. 23B). Four of the other peptides selected for confirmation showed decreased growth when induced, including the hydrophobic cationic peptide (Fig. 23C). Of particular note are the peptides YNFIGFFR and YDDFSFYH, due to the lack of CFUs when plated at 12 and 24 hours.

4.3.5 Motif results

In addition to the read counting Python tool, another computational tool was developed by a fellow master's student, Alex Erikson. This tool uses the data generated from PepSeq to find motifs associated with toxicity (Fig. 24). It allows searches for various motif lengths, in addition

Table 7. Top six toxic peptides

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATAACTTTATCGGTTTCTTTCGC	YNFIGFFR	5	4	80	77	1	1	0.01
AGTCACTTTTTCTATGGCGTTTTC	SHFFYGVF	7	2	47	43	1	1	0.02
CGTTTCCTTTACTTTGGCCATCTC	RFLYFGHL	6	6	49	38	1	2	0.03
GGTTTCTATTACGTTTGCATTATC	GFYYVCII	11	0	50	57	4	1	0.05
TATTACCGTATCATTTGCGTTTTC	YYRIICVF	9	4	40	22	1	2	0.05
TGTCTCTATATCGTTCGCTTTCTC	CLYIVRFL	10	4	64	69	1	6	0.05

Table 8. Five additional toxic peptides

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TTTCGCATTTTCTATCGCATTTTC	FRIFYRIF	7	8	80	88	5	7	0.07
GATTTCGATGGCGTTGGCATTAGC	DFDGVGIS	1	-8	165	153	13	9	0.07
AGTTACGATCTCTGTGGCAATGAC	SYDLCGND	-3	-8	141	113	12	15	0.11
TATGACGATTTCAGTTTCTATCAC	YDDFSFYH	0	-6	61	59	5	9	0.12
GGTATCGGTTGCGATGGCGATGAC	GIGCDGDD	-3	-12	473	371	60	68	0.15

FRIFYRIF is considered a canonical AMP, while the others exhibit unusual hydrophobicity and charge.

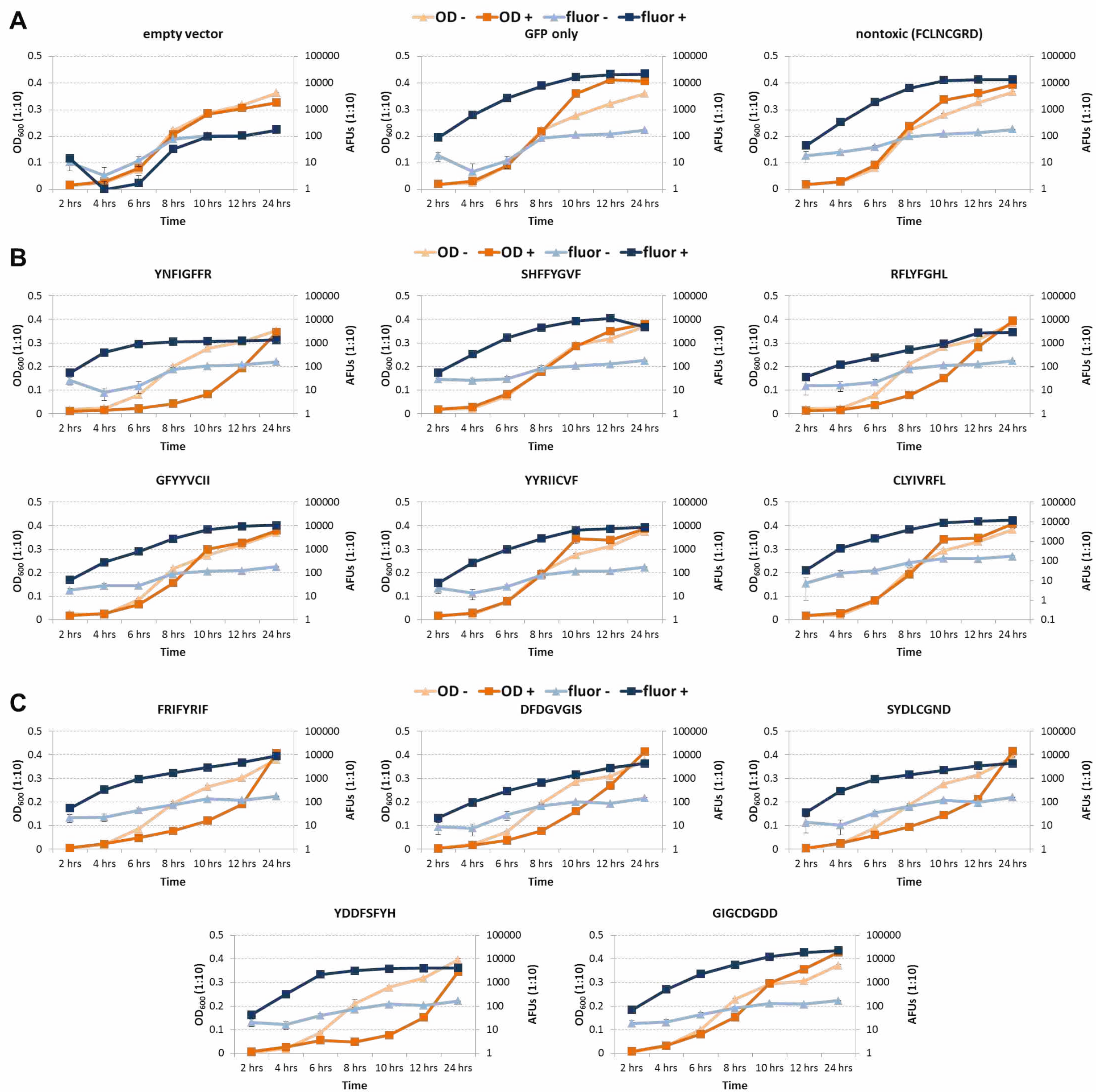


Fig. 23. Peptide validation growth curves and fluorescence over time. A. Empty vector, msfGFP with no peptide, and nontoxic peptide controls. B. Top six toxic peptides. C. One canonical peptide and four unusual peptides. OD_{600} measurements are shown on the primary axis, in orange. Fluorescence measurements are shown on the secondary axis, in blue. Triangle markers indicate uninduced conditions, while square markers indicate induced conditions. All cultures were grown in triplicate. All measurements were taken using a 1:10 dilution. Error bars show standard error.

to allowing for “wildcard” amino acids. This means that there can be a degree of variability within motifs, allowing greater flexibility in finding motifs associated with toxicity. For example, if the wildcard is denoted as X and the motif finder tool is set to search for 5-amino-acid motifs (5mers), it can return a motif such as “YFXYF,” where the middle amino acid is variable among the toxic peptides. Prior to running the motif finder tool, the user defines a set of toxic peptides and a set of nontoxic peptides for comparison. These two sets do not necessarily encompass the entire dataset. It returns each motif, with the number of times it appears in the toxic set, the number of times it appears in the nontoxic set, and various statistics. The tool was run with a variety of parameters to obtain motifs of varying lengths with varying numbers of wildcards. The top five motifs for each motif length are shown in Table 9. A comprehensive list of motifs can be found in Appendix III.

4.3.6 Motif confirmation

Two motifs of interest were selected for further testing and confirmation, in the same manner as the individual toxic peptides. The first motif was YII, chosen due to its prevalence within other motifs in all kmer lengths (Table 10). Three peptides containing this motif were selected for individual testing (Table 11). The growth curves can be seen in Fig. 25. While all three showed decreased growth when induced compared to controls, HYYIIYFI showed the largest difference (Fig. 25).

The other motif that was selected was YFXYF, due to its presence in the top five of all 5mer and 6mer runs (Table 9, Fig. 26A). A consensus sequence was found (Fig. 26B, C), and of all the peptides in the dataset, only one peptide contained the consensus motif (YFCYFG). This

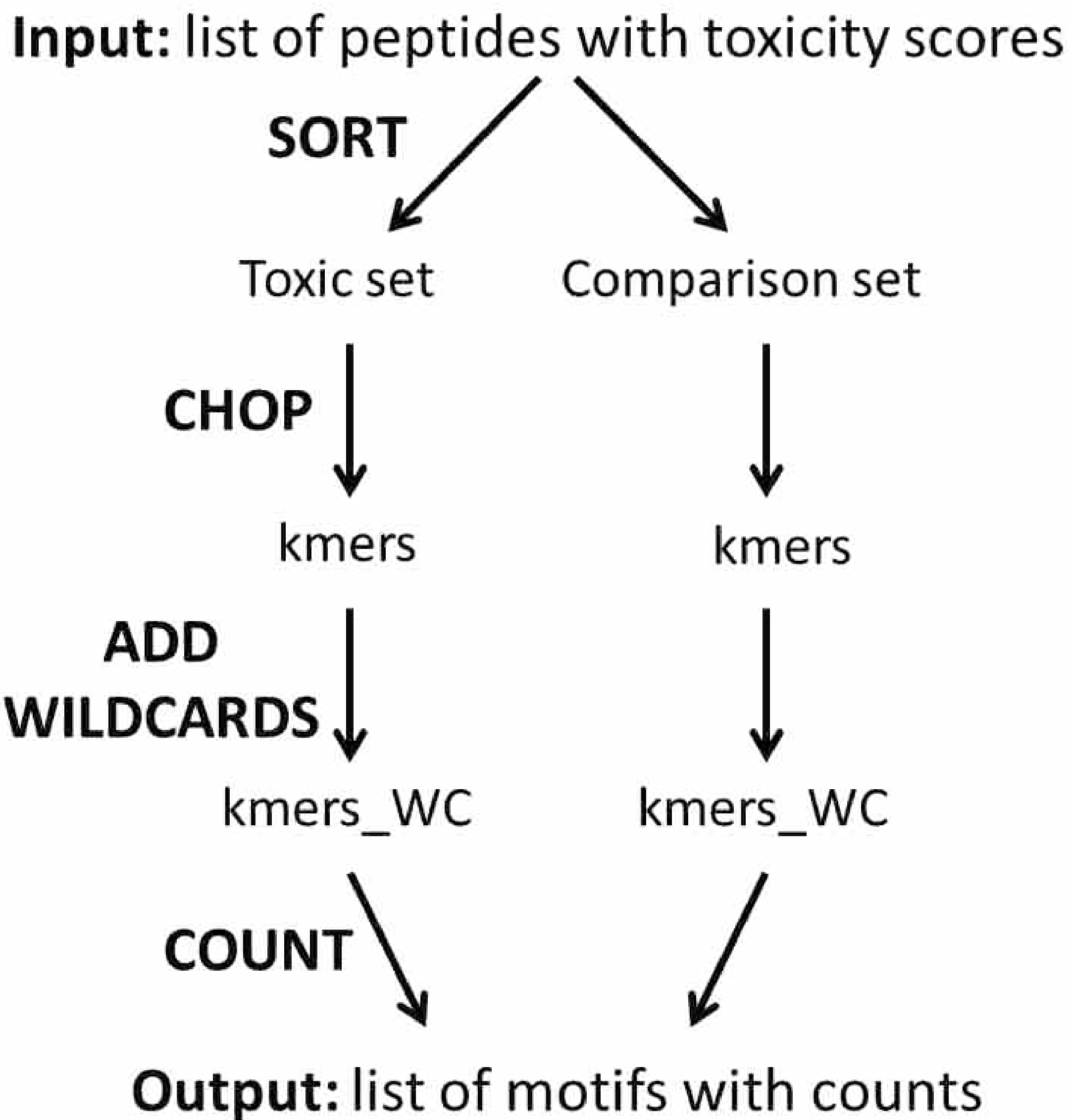


Fig. 24. Motif finder workflow. See text for details.

Table 9. Top five motifs of various kmer lengths

	motif	number of occurrences in toxic group (tox)	number of occurrences in comparison group (comp)	tox/comp	p-value	total run results	
3mer, 0 wild	FYL	201	50	4.02	9.79E-20		
	YYF	251	68	3.70	1.08E-15	**total peptides in toxic set**	12280
	FYF	249	68	3.66	3.09E-19	**total peptides in comparison set**	77316
	YYL	169	48	3.52	3.16E-17	**unique kmers identified**	1782
	YLY	156	47	3.32	5.94E-12		
4mer, 1 wild	YICY	36	0	INF	6.16E-08		
	YCVI	39	1	39	2.49E-05	**total peptides in toxic set**	12280
	YFCY	36	1	36	8.60E-09	**total peptides in comparison set**	77316
	VCYI	31	1	31	2.51E-06	**unique kmers identified**	27876
	YRFF	31	1	31	6.15E-07		
5mer, 1 wild	YFXYF	38	0	INF	1.61E-08		
	YFCYX	31	0	INF	3.50E-08	**total peptides in toxic set**	12280
	YICYX	30	0	INF	6.36E-07	**total peptides in comparison set**	77316
	FYIIX	35	1	35	1.86E-09	**unique kmers identified**	280698
	YIIXY	33	1	33	1.68E-07		
5mer, 2 wild	YFXYF	38	0	INF	1.61E-08		
	YFCYX	31	0	INF	3.50E-08	**total peptides in toxic set**	12280
	YICYX	30	0	INF	6.36E-07	**total peptides in comparison set**	77316
	FYIIX	35	1	35	1.86E-09	**unique kmers identified**	298321
	YIIXY	33	1	33	1.68E-07		
6mer, 2 wild	YFXYFX	36	0	INF	2.01E-08		
	XYFXYF	30	0	INF	5.08E-07	**total peptides in toxic set**	12280
	FYIIXX	29	1	29	2.17E-08	**total peptides in comparison set**	77316
	YYIIXX	30	2	15	1.96E-11	**unique kmers identified**	1484299
	FYXIFX	28	2	14	4.72E-10		

Note that 6mer with one wildcard was also run, but while many unique kmers were identified, none occurred with a high enough frequency to be considered significant.

Table 10. YII-containing motifs

	motif	number of occurrences in toxic group (tox)	number of occurrences in comparison group (comp)	tox/comp	p-value
3mer, 0 wild	YII	249	113	2.20	1.18E-21
4mer, 1 wild	FYII	38	2	19	8.40E-09
	VYII	35	2	17.5	5.68E-08
	YIII	32	3	10.67	6.72E-08
	YYII	32	3	10.67	6.15E-12
	YIIF	43	5	8.6	8.39E-07
	YIIY	30	4	7.5	2.12E-05
	YIIV	31	6	5.17	8.10E-06
	YIIX	236	86	2.74	2.82E-23
	XYII	204	87	2.34	1.39E-19
5mer, 1 wild	FYIIX	35	1	35	1.86E-09
	YIIXY	33	1	33	1.68E-07
	VYIIX	33	1	33	1.03E-07
	XVYII	31	1	31	1.94E-07
	YYIIX	30	2	15	1.96E-11
	XYIII	30	2	15	4.24E-08
	XYIIF	35	3	11.67	2.30E-06
	YIIFX	35	4	8.75	2.33E-05
	YIIXV	36	5	7.2	1.50E-05
5mer, 2 wild	FYIIX	35	1	35	1.86E-09
	YIIXY	33	1	33	1.68E-07
	VYIIX	33	1	33	1.03E-07
	XVYII	31	1	31	1.94E-07
	YYIIX	30	2	15	1.96E-11
	XYIII	30	2	15	4.24E-08
	XYIIF	35	3	11.67	2.30E-06
	YIIFX	35	4	8.75	2.33E-05
	XYIIX	191	60	3.18	1.10E-21
	YIIXX	196	70	2.8	1.25E-19
6mer, 2 wild	FYIIXX	29	1	29	2.17E-08
	YYIIXX	30	2	15	1.96E-11
	YIIXVX	29	4	7.25	2.87E-05

Table 11. YII-containing peptides for confirmation

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
CATTACTATATCATTTACTTTATC	HY**YII**YFI	10	2	88	77	5	4	0.05
CGTTACTATTACATTATCTTTGTC	RYY**YII**FV	9	4	97	81	1	9	0.06
TGTTACCGTCTCTATATCATTTTC	CYRL**YII**F	9	4	230	221	24	13	0.08

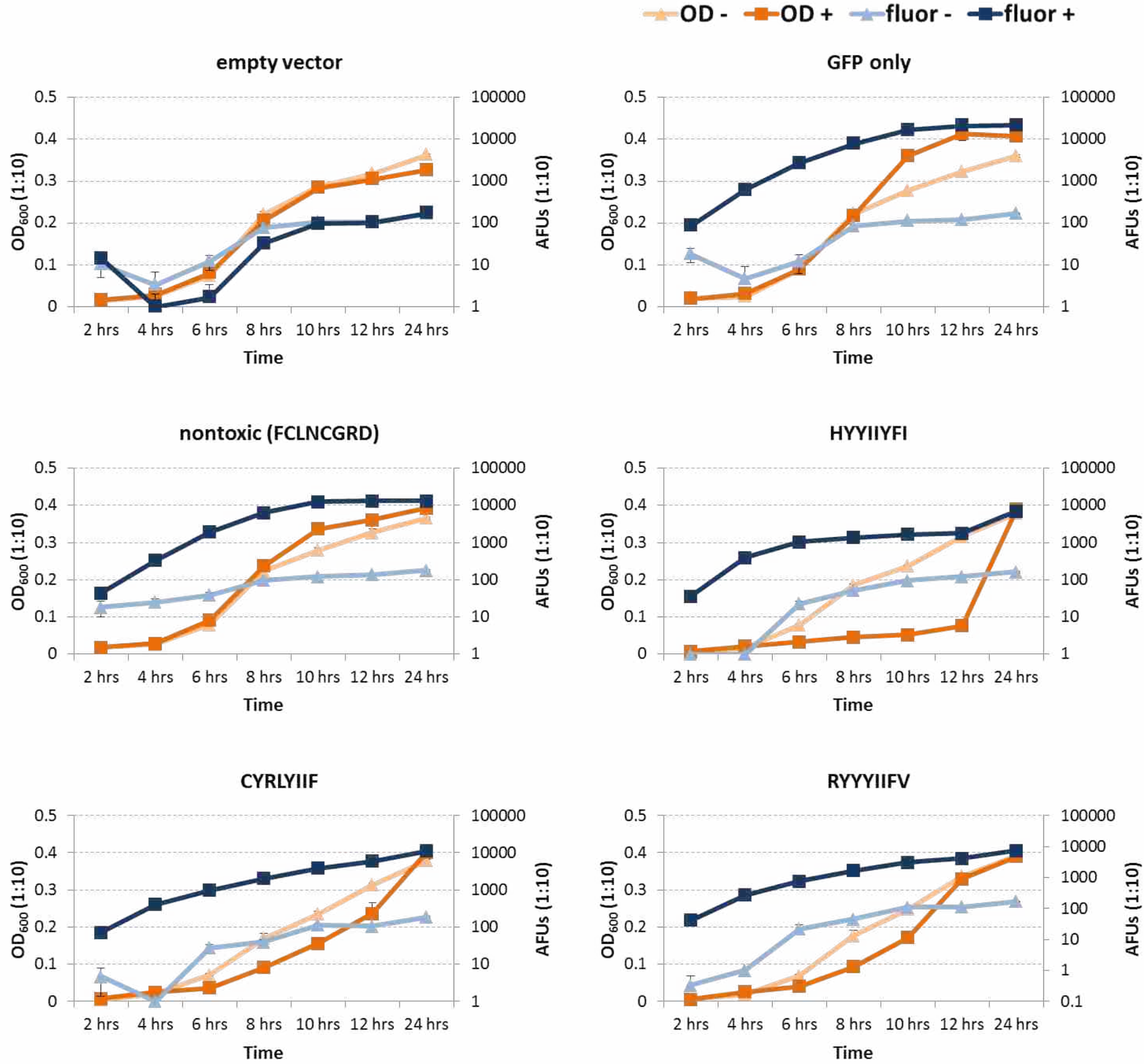
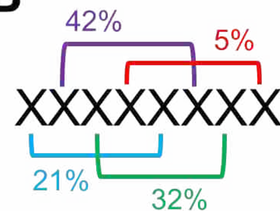


Fig. 25. YII motif growth and fluorescence curves. OD_{600} measurements are shown on the primary axis in orange. Fluorescence measurements are shown on the secondary axis in blue. Triangle markers indicate uninduced conditions. Square markers indicate induced conditions. All cultures were grown in triplicate. All measurements were taken using a 1:10 dilution. Error bars show standard error.

A

	motif	number of occurrences in toxic group (tox)	number of occurrences in comparison group (comp)	tox/comp	p-value
5mer, 1 wild	YFXYF	38	0	INF	1.61E-08
5mer, 2 wild	YFXYF	38	0	INF	1.61E-08
6mer, 2 wild	YFXYFX	36	0	INF	2.01E-08
	XYFXYF	30	0	INF	5.08E-07

B



C **YFXYFX**

%F	%L	%I	%V	%C	%S	%Y	%H	%D	%N	%R	%G
6	14	11	11	8	8	8	6	3	0	6	19

%F	%L	%I	%V	%C	%S	%Y	%H	%D	%N	%R	%G
13	3	4	2	24	5	8	5	5	8	5	8

Fig. 26. YFXYF motif summary. A. YFXYF in top motifs for 5mers and 6mers. B. The percentage of all YFXYF motifs that occur in each position within the peptide. C. The frequency of each amino acid in the wildcard positions.

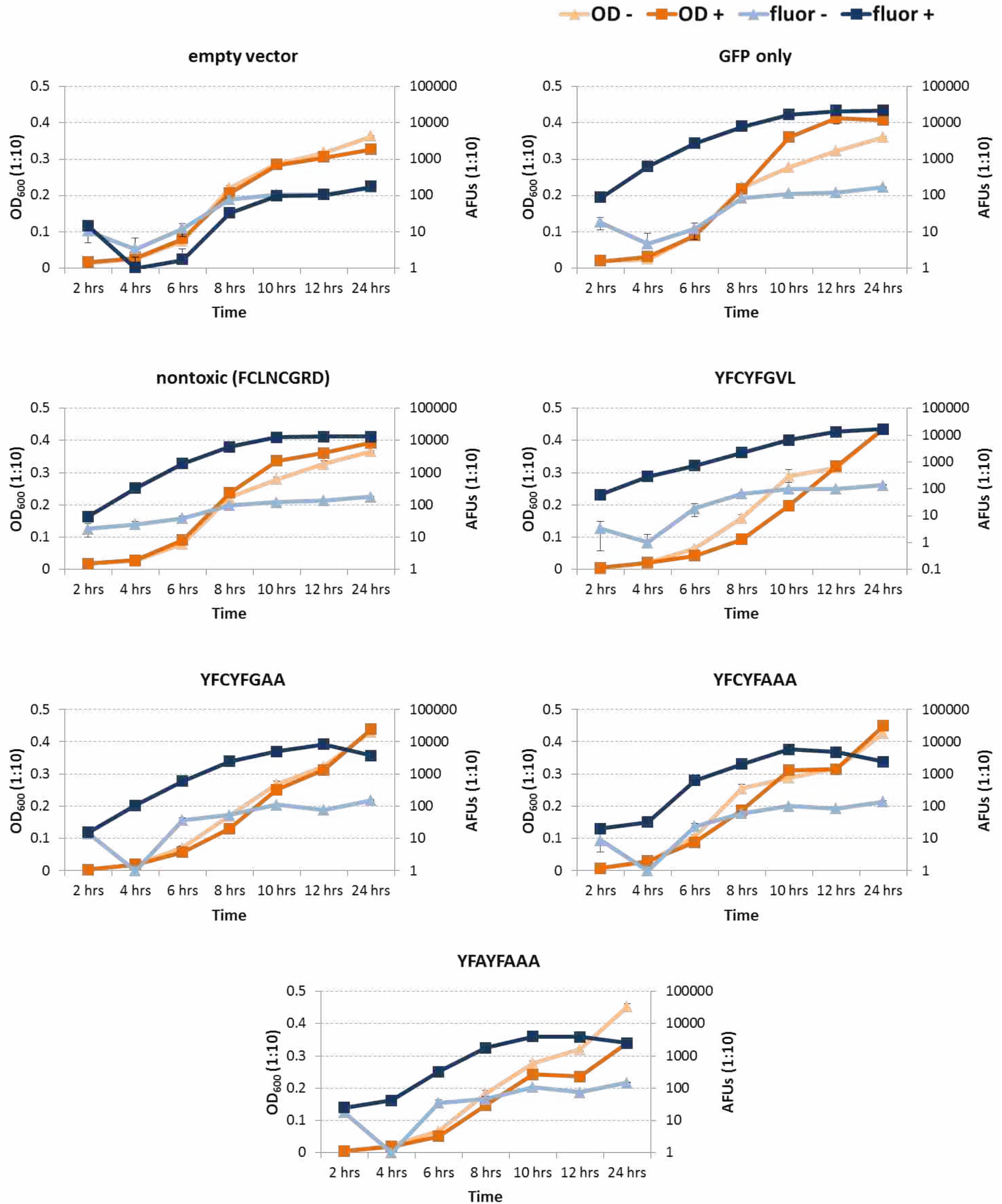


Fig. 27. YFXYF growth and fluorescence curves. OD_{600} measurements are shown on the primary axis, in orange. Fluorescence measurements are shown on the secondary axis, in blue. Triangle markers indicate uninduced conditions, while square markers indicate induced conditions. All cultures were grown in triplicate. All measurements were taken using a 1:10 dilution. Error bars show standard error.

peptide was tested for toxicity, along with a set of variations containing alanines in key positions. This motif did not retest as well as 'YII' (Fig. 27).

4.4 Discussion

The basic platform of PepSeq, while only tested on a linear peptide library, can be easily adapted for a variety of peptide structures. In addition to the obvious changes of peptide length, carrier protein, and number of amino acids in the random portion, there are many other possibilities. More intricate peptide structures can be made with relative ease. For example, as discussed in the previous chapter, a periplasmic-localization signal could be added to allow potential disulfide bonds to form. Additionally, the platform can be adapted to clinically-relevant bacteria, or combined with other factors to search for peptides that increase susceptibility to known antibiotics, increase toxicity of known toxic compounds, or peptides that rescue bacteria from toxic compounds.

CHAPTER 5: Future applications and adaptations for PepSeq

5.1 PepSeq Adaptions

The basic platform of PepSeq, presented in the previous chapter, is highly adaptable. There are two primary types of adaptations for PepSeq. The first is changes in growth conditions, such as type of media and strain background. The second type of adaptation is to change the peptide structure, including length and shape. This chapter will focus on a discussion of a few of these possibilities in more detail.

5.2 Antibiotic adjuvants

The dataset presented here was grown in wild type *E. coli*, in minimal medium. While the initial proof of concept worked well, the platform could be adapted for additional use in clinically relevant pathogens, such as *Klebsiella*. It could also be used for exploring increased susceptibility in antibiotic resistant strains. It has been previously suggested that AMPs could be used in combination with known antibiotics to increase efficacy, particularly against antibiotic resistant bacteria [83]. Using PepSeq, peptides could be expressed in a resistant strain and grown in the appropriate antibiotic. The members that drop out of the population would indicate either peptide toxicity or the peptide could be increasing the susceptibility of the resistant strain to the antibiotic. The difference between the two would become clear in individual confirmation tests, by simply comparing growth with and without antibiotic. Additionally, the platform could be used to find peptides that increase efficacy of existing drugs and antimicrobial compounds outside of antibiotic resistance, in a way similar to the resistance tests just described. By changing growth conditions and the strain background, we can supplement preexisting treatments, in addition to discovering novel AMPs.

Another alternative for adapting the platform is to change the peptide structure. The most straightforward method is to change the peptide length or degenerate codon. By changing the codon, more amino acids would be allowed, expanding the possibilities (Table 12). Changing the peptide length would also increase the number of possibilities in both peptide sequence and potential structures. A longer peptide has a higher probability of folding into a bioactive conformation. While linear peptides may occasionally take on their active shape, there are also techniques that could be used to generate more intricate and stable peptide structures by using multi-plasmid systems, discussed in more detail below.

5.3 Cyclic peptide scaffolds

As indicated in the first chapter of this thesis, there are many platforms that are adaptable for high-throughput screens, such as PepSeq. By incorporating extra features, such as enzymatic changes, tethers, secretion signals, and more, we can expand on the simplistic linear structure focused on up to this point. By producing more complex structures, the peptides will be stabilized, less prone to degradation, and may increase toxic effects and the variety of intracellular targets. Some possible structures include lasso peptides, cyclic peptides, and peptides stabilized by heterocycles (Fig. 28).

5.3.1 MP1 and MccJ25 as peptide scaffolds

As previously discussed, the compound micrococcin P1 (MP1) requires relatively few enzymes to generate its cyclic structure. In the case of MP1, there is a set of three enzymes that work together to convert cysteines to heterocyclized thiazoles [30] (Fig. 29). These enzymes are likely to recognize cysteines in any position [28]. By adding an additional plasmid encoding

Table 12. Potential degenerate codons

	Phe	Ser	Tyr	Cys	Leu	Pro	His
NDT	Arg	Trp	Gln	Ile	Thr	Asn	Lys
	Met	Val	Ala	Asp	Glu	Gly	stop
	Phe	Ser	Tyr	Cys	Leu	Pro	His
YNY	Arg	Trp	Gln	Ile	Thr	Asn	Lys
	Met	Val	Ala	Asp	Glu	Gly	stop
	Phe	Ser	Tyr	Cys	Leu	Pro	His
NNK	Arg	Trp	Gln	Ile	Thr	Asn	Lys
	Met	Val	Ala	Asp	Glu	Gly	Stop

Allowed amino acids for each codon are highlighted in green.

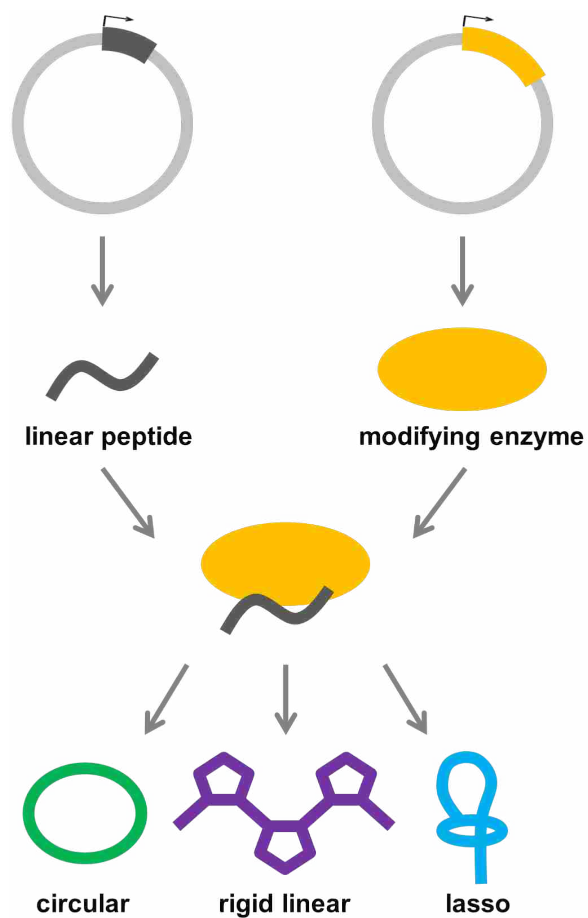


Fig. 28. Proposed peptide structural scaffolds.

these enzymes to the PepSeq system, cysteines in a random peptide can be converted to heterocycles, making a rigid conformation of a linear peptide.Another peptide structure that could be achieved is a lasso peptide. The enzymes involved in generating the lasso of microcin J25 (MccJ25) have been shown to tolerate changes in the amino acid sequence in the loop region [32, 84]. By adjusting the peptide library to include the recognition sequence for these enzymes, a lasso peptide library could be generated and used for PepSeq.

5.3.2 Unnatural amino acids to cyclize peptides

Another possibility for cyclized peptides is to use unnatural amino acids. The incorporation of these chemicals into peptides can generate cyclopeptides [85-87], using amber suppression technology. This technique uses a $tRNA^{Tyr}$ and its corresponding aminoacyl tRNA synthetase (aaRS) from the organism *Methanococcus jannaschii* [88]. The tRNA is modified to correspond to the amber stop codon (TAG). This means that an amino acid can now be incorporated at this codon. The aaRS has also been modified by the Fasan lab to charge the amber-corresponding tRNA with an unnatural amino acid, instead of the typical tyrosine. Altogether, this technology allows us to use cells to create peptides with an unnatural amino acid in specific places within the sequence. One particular system takes advantage of cysteine-reactive chemistry to form a thioether [88]. This innovative system is capable of charging and incorporating two unnatural amino acids, O-(2-bromoethyl)-tyrosine (O2beY) and O-propargyl-tyrosine (OpgY; Fig. 30A). O2beY is a cysteine-reactive amino acid, meaning that in a spontaneous chemical reaction, its side chain and a nearby cysteine side chain form a covalent bond, specifically a thioether, generating a cyclopeptide (Fig. 30B). For this reaction to occur, O2beY and the cysteine must be within 2-10

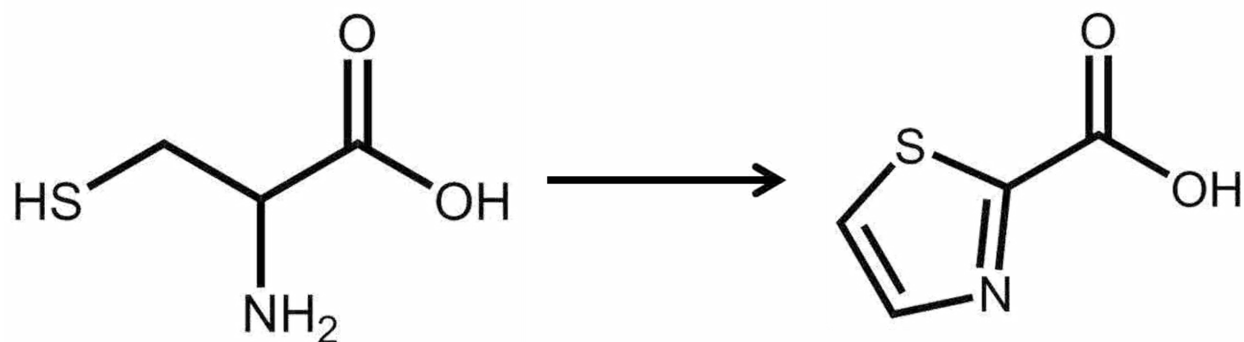


Fig. 29. Cysteine converted to thiazole.

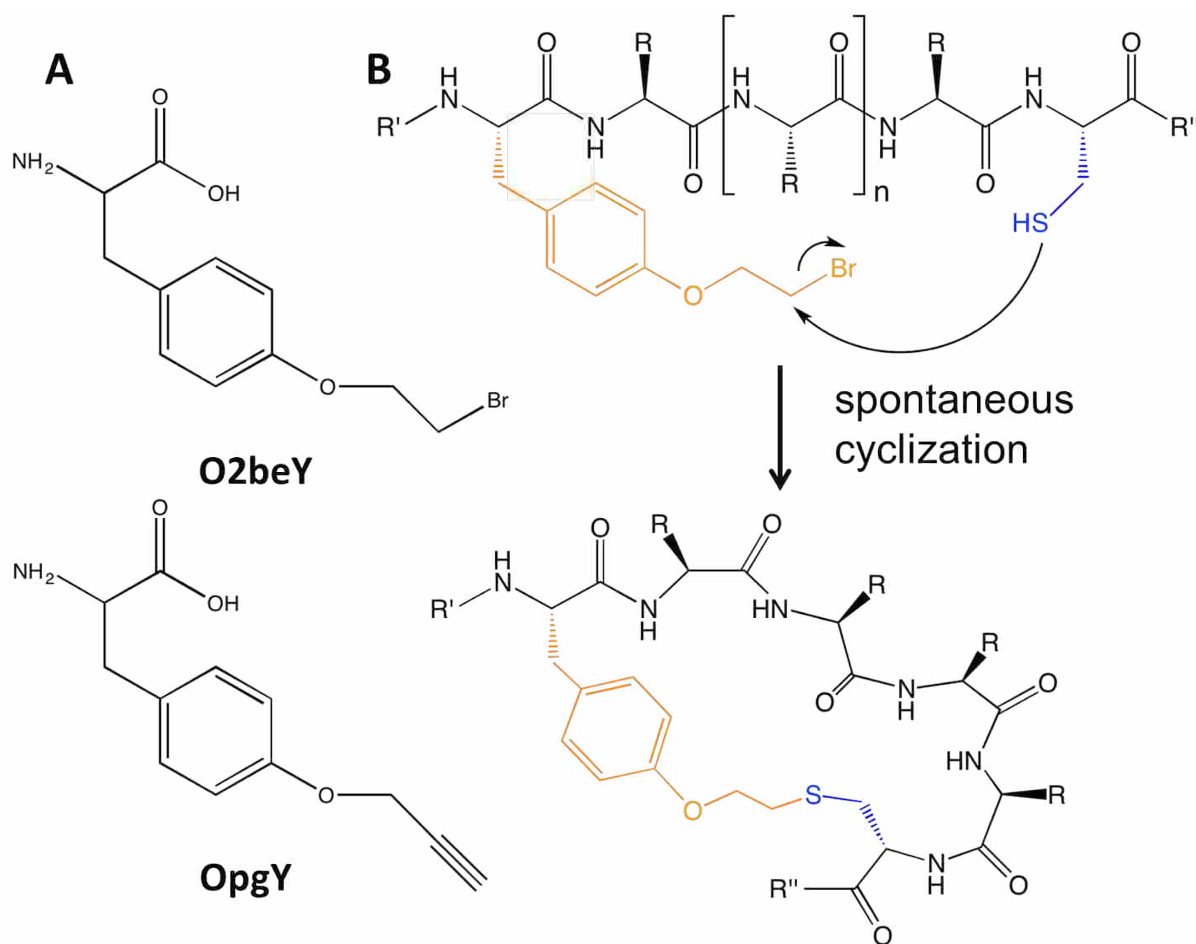


Fig. 30. Cysteine-reactive chemistry forms a cyclopeptide. A. Chemical structures of the unnatural amino acids O2beY and OpgY. B. A spontaneous chemical reaction between O2beY and Cys forms a cyclopeptide.

amino acids of each other [88]. OpgY can also be incorporated using this same system. While it is similar to O2beY in structure, it does not react with cysteine and leaves the peptide in the original linear structure, allowing comparison of toxicity of linear and cyclic peptides.

This system, used in combination with PepSeq, could be used to generate and screen a cyclopeptide library, using a two-plasmid system, to place O2beY near a cysteine. The remaining amino acids in the structure can then be randomized to generate a random library. OpgY can additionally be used to compare the effect of cyclization on toxicity. While this proposed system, along with the lasso peptide and thiazole-containing peptide systems, requires additional enzymes to achieve the final secondary structure, there are also possibilities for secondary structures driven by disulfide bonds.

5.3.3 Disulfide bonds and peptide export

While not possible in the cytoplasm of *E. coli*, disulfide bonds can occur in the periplasm when peptides are exported using the DsbA signal peptide as seen in Chapter 3. This same principle could be applied to PepSeq, with cysteines in specific or random positions. By replacing the msfGFP carrier protein used previously with either the DsbA or YiaD signal peptide, the random peptide could be expressed in the periplasm, allowing disulfide bonds to form. Comparisons between DsbA and YiaD signal peptide libraries could also be used to indicate potential peptide targets, as those active at the membrane would likely increase in toxicity when tethered, while those that act on intracellular targets would be likely to decrease in toxicity when tethered due to inability to reach their target.

5.4 Conclusion

While toxic peptides have been discovered in the linear version of PepSeq, adding in additional secondary structure may increase our ability to discover highly toxic AMPs. Signal

peptides localizing random library peptides to the periplasm would allow higher order structures through disulfide bonds between cysteines. The addition of modifying enzymes on separate plasmids would enable the creation of lasso, cyclic, and rigid peptides. These same peptide structures could also be applied using SLAY [66], allowing a more direct route to potential antibiotic peptides.

BIBLIOGRAPHY

1. Brown, E.D. and G.D. Wright, *Antibacterial drug discovery in the resistance era.* Nature, 2016. 529(7586): p. 336-343.
2. Laxminarayan, R., et al., *Antibiotic resistance—the need for global solutions.* The Lancet infectious diseases, 2013. 13(12): p. 1057-1098.
3. Nathan, C. and O. Cars, *Antibiotic Resistance — Problems, Progress, and Prospects.* New England Journal of Medicine, 2014. 371(19): p. 1761-1763.
4. Piddock, L.J.V., *The crisis of no new antibiotics-what is the way forward?* The Lancet Infectious Diseases, 2011. 12(3): p. 249-253.
5. Spellberg, B., J.G. Bartlett, and D.N. Gilbert, *The Future of Antibiotics and Resistance.* New England Journal of Medicine, 2013. 368(4): p. 299-302.
6. Clatworthy, A.E., E. Pierson, and D.T. Hung, *Targeting virulence: a new paradigm for antimicrobial therapy.* Nature Chemical Biology, 2007. 3: p. 541.
7. Aiken, A.M., et al., *Antibiotic resistance needs global solutions.* The Lancet Infectious Diseases, 2014. 14(7): p. 550-551.
8. Bush, K., et al., *Tackling antibiotic resistance.* Nat Rev Micro, 2011. 9(12): p. 894-896.
9. Nathan, C. and F.M. Goldberg, *The profit problem in antibiotic R&D.* Nat Rev Drug Discov, 2005. 4(11): p. 887-891.
10. Ventola, C.L., *The antibiotic resistance crisis: part 1: causes and threats.* P T, 2015. 40(4): p. 277-83.
11. Clardy, J., M. Fischbach, and C. Currie, *The natural history of antibiotics.* Current biology : CB, 2009. 19(11): p. R437-R441.

12. Lewis, K., *Platforms for antibiotic discovery.* Nature Reviews Drug Discovery, 2013. 12: p. 371.

13. Fenical, W. and P.R. Jensen, *Developing a new resource for drug discovery: marine actinomycete bacteria.* Nature Chemical Biology, 2006. 2: p. 666.

14. Hancock, R.E.W. and G. Diamond, *The role of cationic antimicrobial peptides in innate host defences.* Trends in Microbiology, 2000. 8(9): p. 402-410.

15. Peschel, A. and H.-G. Sahl, *The co-evolution of host cationic antimicrobial peptides and microbial resistance.* Nature Reviews Microbiology, 2006. 4: p. 529.

16. Wu, M., et al., *Mechanism of Interaction of Different Classes of Cationic Antimicrobial Peptides with Planar Bilayers and with the Cytoplasmic Membrane of Escherichia coli.* Biochemistry, 1999. 38(22): p. 7235-7242.

17. Hancock, R.E.W. and A. Patrzykat, *Clinical Development of Cationic Antimicrobial Peptides: From Natural to Novel Antibiotics.* Current Drug Targets - Infectious Disorders, 2002. 2(1): p. 79-83.

18. Bechinger, B. and K. Lohner, *Detergent-like actions of linear amphipathic cationic antimicrobial peptides.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2006. 1758(9): p. 1529-1539.

19. Zhang, L., A. Rozek, and R.E.W. Hancock, *Interaction of Cationic Antimicrobial Peptides with Model Membranes.* Journal of Biological Chemistry, 2001. 276(38): p. 35714-35722.

20. Hancock, R.E.W., *Cationic peptides: effectors in innate immunity and novel antimicrobials.* The Lancet Infectious Diseases, 2001. 1(3): p. 156-164.

21. Hale, J.D.F. and R.E.W. Hancock, *Alternative mechanisms of action of cationic antimicrobial peptides on bacteria.* Expert Review of Anti-infective Therapy, 2007. 5(6): p. 951-959.

22. Omardien, S., et al., *Synthetic antimicrobial peptides delocalize membrane bound proteins thereby inducing a cell envelope stress response.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2018.

23. Pietiäinen, M., et al., *Cationic antimicrobial peptides elicit a complex stress response in Bacillus subtilis that involves ECF-type sigma factors and two-component signal transduction systems.* Microbiology, 2005. 151(5): p. 1577-1592.

24. Poole, K., *Bacterial stress responses as determinants of antimicrobial resistance.* Journal of Antimicrobial Chemotherapy, 2012. 67(9): p. 2069-2089.

25. Omardien, S., S. Brul, and S.A.J. Zaat, *Antimicrobial Activity of Cationic Antimicrobial Peptides against Gram-Positives: Current Progress Made in Understanding the Mode of Action and the Response of Bacteria.* Frontiers in Cell and Developmental Biology, 2016. 4(111).

26. Hancock, R.E.W., *Peptide antibiotics.* The Lancet, 1997. 349(9049): p. 418-422.

27. Maróti, G., et al., *Natural roles of antimicrobial peptides in microbes, plants and animals.* Research in Microbiology, 2011. 162(4): p. 363-374.

28. Bennallack, P.R. and J.S. Griffitts, *Elucidating and engineering thiopeptide biosynthesis.* World Journal of Microbiology and Biotechnology, 2017. 33(6): p. 119.

29. Ortega, Manuel A. and Wilfred A. van der Donk, *New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products.* Cell Chemical Biology. 23(1): p. 31-44.

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

31. Kok-Phen, Y., et al., *Dissecting the Maturation Steps of the Lasso Peptide Microcin J25 in vitro.* ChemBioChem, 2012. 13(7): p. 1046-1052.

32. Pan, S.J. and A.J. Link, *Sequence Diversity in the Lasso Peptide Framework: Discovery of Functional Microcin J25 Variants with Multiple Amino Acid Substitutions.* Journal of the American Chemical Society, 2011. 133(13): p. 5016-5023.

33. De Smet, K. and R. Contreras, *Human Antimicrobial Peptides: Defensins, Cathelicidins and Histatins.* Biotechnology Letters, 2005. 27(18): p. 1337-1347.

34. Hazlett, L. and M. Wu, *Defensins in innate immunity.* Cell and Tissue Research, 2011. 343(1): p. 175-188.

35. Ganz, T., *Defensins: antimicrobial peptides of innate immunity.* Nature Reviews Immunology, 2003. 3: p. 710.

36. Olivera, B.M., et al., *Calcium Channel Diversity and Neurotransmitter Release: The ω-Conotoxins and ω-Agatoxins.* Annual Review of Biochemistry, 1994. 63(1): p. 823-867.

37. Bulaj, G., et al., *Efficient oxidative folding of conotoxins and the radiation of venomous cone snails.* Proceedings of the National Academy of Sciences, 2003. 100(suppl 2): p. 14562.

38. Hemu, X. and J.P. Tam, *Macrocyclic Antimicrobial Peptides Engineered from ω-Conotoxin.* Current Pharmaceutical Design, 2017. 23(14): p. 2131-2138.

39. Terras, F.R., et al., *Small cysteine-rich antifungal proteins from radish: their role in host defense.* The Plant Cell, 1995. 7(5): p. 573.

40. Tailor, R.H., et al., *A Novel Family of Small Cysteine-rich Antimicrobial Peptides from Seed of Impatiens balsamina Is Derived from a Single Precursor Protein.* Journal of Biological Chemistry, 1997. 272(39): p. 24480-24487.

41. Martijn, R., et al., *A small, cysteine-rich protein secreted by Fusarium oxysporum during colonization of xylem vessels is required for I-3-mediated resistance in tomato.* Molecular Microbiology, 2004. 53(5): p. 1373-1383.

42. Gao, A.-G., et al., *Fungal pathogen protection in potato by expression of a plant defensin peptide.* Nature Biotechnology, 2000. 18: p. 1307.

43. Thomma, B.P., B.P. Cammue, and K. Thevissen, *Plant defensins.* Planta, 2002. 216(2): p. 193-202.

44. Wang, D., et al., *A Nodule-Specific Protein Secretory Pathway Required for Nitrogen-Fixing Symbiosis.* Science, 2010. 327(5969): p. 1126.

45. Van de Velde, W., et al., *Plant Peptides Govern Terminal Differentiation of Bacteria in Symbiosis.* Science, 2010. 327(5969): p. 1122.

46. Durgo, H., et al., *Identification of nodule-specific cysteine-rich plant peptides in endosymbiotic bacteria.* PROTEOMICS, 2015. 15(13): p. 2291-2295.

47. Price, P.A., et al., *Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility.* Proceedings of the National Academy of Sciences of the United States of America, 2015. 112(49): p. 15244-15249.

48. Maróti, G., J.A. Downie, and É. Kondorosi, *Plant cysteine-rich peptides that inhibit pathogen growth and control rhizobial differentiation in legume nodules.* Current Opinion in Plant Biology, 2015. 26: p. 57-63.

49. Tiricz, H., et al., *Antimicrobial Nodule-Specific Cysteine-Rich Peptides Induce Membrane Depolarization-Associated Changes in the Transcriptome of Sinorhizobium meliloti.* Applied and Environmental Microbiology, 2013. 79(21): p. 6737-6746.

50. Shabab, M., et al., *Disulfide cross-linking influences symbiotic activities of nodule peptide NCR247.* Proc Natl Acad Sci U S A, 2016. 113(36): p. 10157-62.

51. Powell, W.A., C.M. Catranis, and C.A. Maynard, *Synthetic antimicrobial peptide design.* Mol Plant Microbe Interact, 1995. 8(5): p. 792-4.

52. Spindler, E.C., et al., *Deciphering the Mode of Action of the Synthetic Antimicrobial Peptide Bac8c.* Antimicrobial Agents and Chemotherapy, 2011. 55(4): p. 1706-1716.

53. Gun Lee, D., K.-S. Hahm, and S. Yub Shin, *Structure and fungicidal activity of a synthetic antimicrobial peptide, P18, and its truncated peptides*. Vol. 26. 2004. 337-41.

54. Mentag, R., et al., *Bacterial disease resistance of transgenic hybrid poplar expressing the synthetic antimicrobial peptide D4E1.* Tree Physiology, 2003. 23(6): p. 405-411.

55. Cary, J.W., et al., *Transgenic expression of a gene encoding a synthetic antimicrobial peptide results in inhibition of fungal growth in vitro and in planta.* Plant Science, 2000. 154(2): p. 171-181.

56. Hilpert, K., et al., *High-throughput generation of small antibacterial peptides with improved activity.* Nature Biotechnology, 2005. 23: p. 1008.

57. Alessandro, T., T. Chiara, and R. Domenico, *Design of Synthetic Antimicrobial Peptides Based on Sequence Analogy and Amphipathicity.* European Journal of Biochemistry, 1997. 250(2): p. 549-558.

58. Wang, G., X. Li, and Z. Wang, *APD2: the updated antimicrobial peptide database and its application in peptide design.* Nucleic Acids Research, 2009. 37(suppl_1): p. D933-D937.

59. Tietz, J.I., et al., *A new genome-mining tool redefines the lasso peptide biosynthetic landscape.* Nature chemical biology, 2017. 13(5): p. 470-478.

60. Young, T.S., P.C. Dorrestein, and C.T. Walsh, *Codon randomization for rapid exploration of chemical space in thiopeptide antibiotic variants.* Chemistry & biology, 2012. 19(12): p. 1600-1610.

61. Porto, W.F., et al., *In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design.* Nature Communications, 2018. 9(1): p. 1490.

62. Lee, E.Y., G.C.L. Wong, and A.L. Ferguson, *Machine learning-enabled discovery and design of membrane-active peptides.* Bioorganic & Medicinal Chemistry, 2018. 26(10): p. 2708-2718.

63. Lee, E.Y., et al., *Mapping membrane activity in undiscovered peptide sequence space using machine learning.* Proceedings of the National Academy of Sciences, 2016. 113(48): p. 13588.

64. Olivera, B.M., et al., *Combinatorial peptide libraries in drug design: lessons from venomous cone snails.* Trends in Biotechnology, 1995. 13(10): p. 422-426.

65. Luther, A., C. Bisang, and D. Obrecht, *Advances in macrocyclic peptide-based antibiotics.* Bioorganic & Medicinal Chemistry, 2018. 26(10): p. 2850-2858.

66. Tucker, A.T., et al., *Discovery of Next-Generation Antimicrobials through Bacterial Self-Screening of Surface-Displayed Peptide Libraries.* Cell, 2018. 172(3): p. 618-628.e13.

67. Rodríguez-García, A., et al., *Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria Streptomyces.* Nucleic Acids Research, 2005. 33(9): p. e87-e87.

68. Qin, J.Y., et al., *Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter.* PLOS ONE, 2010. 5(5): p. e10611.

69. Taylor, W.E., et al., *Transcription from a heat-inducible promoter causes heat shock regulation of the sigma subunit of E. coli RNA polymerase.* Cell, 1984. 38(2): p. 371-381.

70. Newman, J.R. and C. Fuqua, *Broad-host-range expression vectors that carry the l-arabinose-inducible Escherichia coli araBAD promoter and the araC regulator.* Gene, 1999. 227(2): p. 197-203.

71. Cardona, S.T. and M.A. Valvano, *An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in Burkholderia cenocepacia.* Plasmid, 2005. 54(3): p. 219-228.

72. Chen, C.Y. and S.C. Winans, *Controlled expression of the transcriptional activator gene virG in Agrobacterium tumefaciens by using the Escherichia coli lac promoter.* Journal of Bacteriology, 1991. 173(3): p. 1139-1144.

73. Morieri, G., et al., *Host-specific Nod-factors associated with Medicago truncatula nodule infection differentially induce calcium influx and calcium spiking in root hairs.* The New Phytologist, 2013. 200(3): p. 656-662.

74. Gage, D.J., *Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes.* Microbiology and Molecular Biology Reviews, 2004. 68(2): p. 280-300.

75. Haag, A.F., et al., *Protection of Sinorhizobium against Host Cysteine-Rich Antimicrobial Peptides Is Critical for Symbiosis.* PLOS Biology, 2011. 9(10): p. e1001169.

76. Kim, M., et al., *An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing symbiosis.* Proceedings of the National Academy of Sciences, 2015. 112(49): p. 15238.

77. Wang, Q., et al., *Host-secreted antimicrobial peptide enforces symbiotic selectivity in Medicago truncatula.* Proceedings of the National Academy of Sciences, 2017. 114(26): p. 6854.

78. Farkas, A., et al., *Medicago truncatula symbiotic peptide NCR247 contributes to bacteroid differentiation through multiple mechanisms.* Proceedings of the National Academy of Sciences, 2014. 111(14): p. 5183-5188.

79. Horváth, B., et al., *Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the Medicago truncatula dnf7 mutant.* Proceedings of the National Academy of Sciences, 2015. 112(49): p. 15232-15237.

80. Yang, S., et al., *Microsymbiont discrimination mediated by a host-secreted peptide in Medicago truncatula.* Proceedings of the National Academy of Sciences, 2017. 114(26): p. 6848.

81. Schierle, C.F., et al., *The DsbA Signal Sequence Directs Efficient, Cotranslational Export of Passenger Proteins to the Escherichia coli Periplasm via the Signal Recognition Particle Pathway.* Journal of Bacteriology, 2003. 185(19): p. 5706-5713.

82. San Millan, J.L., et al., *Use of phoA fusions to study the topology of the Escherichia coli inner membrane protein leader peptidase.* Journal of Bacteriology, 1989. 171(10): p. 5536-5541.

83. Mishra, B., et al., *Host defense antimicrobial peptides as antibiotics: design and application strategies.* Current Opinion in Chemical Biology, 2017. 38: p. 87-96.

84. Pan, S.J., et al., *Computational design of the lasso peptide antibiotic microcin J25.* Protein Engineering, Design and Selection, 2011. 24(3): p. 275-282.

85. Bionda, N. and R. Fasan, *Ribosomal Synthesis of Natural-Product-Like Bicyclic Peptides in Escherichia coli.* ChemBioChem, 2015. 16(14): p. 2011-2016.

86. Frost, J.R., et al., *Ribosomal Synthesis of Macrocyclic Peptides in Vitro and in Vivo Mediated by Genetically Encoded Amino-Thiol Unnatural Amino Acids.* ACS chemical biology, 2015. 10(8): p. 1805-1816.

87. Frost, J.R., et al., *Side-chain-to-tail cyclization of ribosomally derived peptides promoted by aryl and alkyl amino-functionalized unnatural amino acids.* Organic & Biomolecular Chemistry, 2016. 14(24): p. 5803-5812.

88. Bionda, N., A.L. Cryan, and R. Fasan, *Bioinspired Strategy for the Ribosomal Synthesis of Thioether-Bridged Macrocyclic Peptides in Bacteria.* Acs Chemical Biology, 2014. 9(9): p. 2008-2013.

APPENDIX I

Promoter primer, insert, and plasmid sequences

P_{ara} from MG1655

Primers:
1634 CGCctcgagGCTACTCCGTCAAGCCGTC
1635 CGCggtaccAAAAACGGGtATGGAGAAACAGTAG
Template: B079 (MG1655)
Insert:

```
ctcgagGCTACTCCGTCAAGCCGTCAATTGTCTGATTCGTTACCAAttaTGACAACTTGACGGCTACATCATTCACTTTTTCTTCA
CAACCGGCACGAAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACTCGCGAGAAATAGAGTTGATCGTCAAAACCAACATT
GCGACCGACGGTGGCGATAGGCATCCGGGTAGTGCTCAAAAGCAGCTTCGCCTGACTAATGCGTTGGTCCTCGCGCCAGCTTAAGA
CGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAA
TTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGC
TTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGA
TTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAAACCCGTATTGGCAAATATTGACGGCCAGTTA
AGCCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATG
AATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAGACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAA
TGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCGTTAAA
CCCGCCACCAGATGGGCGTTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCcatACTTTTCATACTCCCACCATT
CAGAGAAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCCAACCGGTAA
CCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGA
AAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACG
CTTTTTATCGCAACTCTCTACTGTTTCTCCATaCCCGTTTTTggtacc
```

P_{ara} from pGLO

Primers:
oED098 cgcctcgagGCAAACCCTATGCTACTCCG
oED099 CTCCAGTGAAAAGTTCTTCTCC
Template: C398 (pGLO)
Insert:

```
ctcgagGCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTCTGATTCGTTACCAAttaTGACAACTTGACGGCTACATCATTCA
CTTTTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCA
AAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCG
CCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGG
CGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGAC
TCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCC
GGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTG
ACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGA
CCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCC
CTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAA
TCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCcatACTTTTCATA
CTCCCGCCATTCAGAGAAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAAC
CAAACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATA
ATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGAT
CCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAGAAATAATTTTGTTTAACTTTAAGA
AGGAGATATAcatatg
```

P_{rha} from MG1655

Primers:
1636 CGCctcgagCATCGTCGGCATCGGCATG

1637 CGCggtaccGCTGAATTTCATtACGACCAGTC
Template B079 (MG1655)
Insert:

```
ctcgagCATCGTCGGCATCGGCATGGCGAAttaATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGTTTCCC
GGGTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAATATCACTGATTAACAGGCGGCTATGCTGGAGA
AGATATTGCGCATGACACACTCTGACCTGTCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAAC
GCGCTCACTGCACGATGCCTCATCACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTAT
CCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAATCACTGGTGTAACGATGGCGATTCAGCAACATCACCAACTGCCCGAACAGC
AACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCATGCTCAAGCTGACCGATAACCTGCCGCGCCTGCGCCATCCC
CATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAAATCCCGGAATCGCCCCCTGCCAGTCAAGATTCAGCTTCAGAC
GCTCCGGGCAATAAATAATATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCAGCATGAATGTAAAAGAGA
TCGCCACGGGTAATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAAAATCATG
TGTATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGCAAAAAAATCATCTTTGAGAA
GTTTTAACTGATGCGCcacCGTGGCTACCTCGGCCAGAGAACGAAGTTGATTATTCGCAATATGGCGTACAAATACGTTGAGAAGA
TTCGCGttaTTGCAGAAAGCCATCCCGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGT
GGTTACTGTCGCTGAATCCACAGCGATAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGT
CGCAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCGTTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTG
ATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAGCCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGTTTT
CCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTGCATAAACAAGATCTCGCGACTGGCGGTCGAGGGTAAATCA
TTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGCGCCAGTGAGACGGATA
CTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCACATTGGTCA
GACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTGCCACCGGTGATGGTATAGGGCTGC
CCATTAAACACATGAATACCCGTGCCATGTTCGACAATCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGG
GAGCCGGGGTTCTATCGCCACGGACGCGTTACCAGACGGAAAAAAATCCACACTATGTAATACGGTcatACTGGCCTCCTGATGTC
GTCAACACGGCGAAATAGTAATCACGAGGTCAGGTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCA
AGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCACATCACCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTC
CCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCGTaATGAAATTCAG
Cggtacc
```

P_{rha} with XRBS addition

Primers:
oED096 gagcgtagcgaccgagtg
oED097 cgcgcgCATATGtatatctccttcttaaagttaaacaaaattatttctagccGCTGAATTTCATtACGACCAGTC
Template: pJG780
Insert (XRBS):

```
ctcgagCATCGTCGGCATCGGCATGGCGAAttaATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGTTTCCC
GGGTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAATATCACTGATTAACAGGCGGCTATGCTGGAGA
AGATATTGCGCATGACACACTCTGACCTGTCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAAC
GCGCTCACTGCACGATGCCTCATCACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTAT
CCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAATCACTGGTGTAACGATGGCGATTCAGCAACATCACCAACTGCCCGAACAGC
AACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCATGCTCAAGCTGACCGATAACCTGCCGCGCCTGCGCCATCCC
CATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAAATCCCGGAATCGCCCCCTGCCAGTCAAGATTCAGCTTCAGAC
GCTCCGGGCAATAAATAATATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCAGCATGAATGTAAAAGAGA
TCGCCACGGGTAATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAAAATCATG
TGTATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGCAAAAAAATCATCTTTGAGAA
GTTTTAACTGATGCGCcacCGTGGCTACCTCGGCCAGAGAACGAAGTTGATTATTCGCAATATGGCGTACAAATACGTTGAGAAGA
TTCGCGttaTTGCAGAAAGCCATCCCGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGT
GGTTACTGTCGCTGAATCCACAGCGATAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGT
CGCAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCGTTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTG
ATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAGCCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGTTTT
CCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTGCATAAACAAGATCTCGCGACTGGCGGTCGAGGGTAAATCA
TTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGCGCCAGTGAGACGGATA
CTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCACATTGGTCA
GACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTGCCACCGGTGATGGTATAGGGCTGC
CCATTAAACACATGAATACCCGTGCCATGTTCGACAATCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGG
GAGCCGGGGTTCTATCGCCACGGACGCGTTACCAGACGGAAAAAAATCCACACTATGTAATACGGTcatACTGGCCTCCTGATGTC
GTCAACACGGCGAAATAGTAATCACGAGGTCAGGTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCA
AGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCACATCACCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTC
```

CCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTCGTaATGAAATTCAG
CggctagaaataattttgtttaactttaagaaggagatataCATatg

P_{tet}

Primers:
oDB014 CGCCTCGAGACGTCTTAAGACCCACTTTCAC
oED015 CGCGGTACCTTTTCTCTATCACTGATAGGGA
Template (pJG717)
Insert:
CTCGAgacgtcTTAagacccactttcacatttaagttgtttttctaatccgcatatgatcaattcaaggccgaataagaaggctgg
ctctgcaccttggtgatcaaataattcgatagcttgtcgtaataatggcggcatactatcagtagtaggtgtttccctttcttctt
tagcgacttgatgctcttgatcttccaatacgcaacctaaagtaaaatgccccacagcgctgagtgcatataatgcattctctagt
gaaaaaccttgttggcataaaaaggctaattgattttcgagagtttcatactgtttttctgtaggccgtgtacctaaatgtacttt
tgctccatcgcgatgacttagtaaagcacatctaaaacttttagcgttattacgtaaaaaatcttgccagctttccccttctaaag
ggcaaaagtgagtatggtgcctatctaacatctcaatggctaaggcgtcgagcaaagcccgcttattttttacatgccaatacaat
gtaggctgctctacacctagcttctgggcgagtttacgggttgttaaaccttcgattccgacctcattaagcagctctaatgcgct
gttaatcactttacttttatctaatctagaCATcattaattcctaatttttgTTGACActctatcgttgatagagtTATTTTacca
ctccctatcagtgatagagaaaagGTACC

P_{lacT5} (*lacI*WT)

Primers:
1642 cgcGGATCCcgacgctctcccttatgcg
1643 cgcGGTaccgttgtggAATTGTGAGCG
1638 cgcCTCGAGccggtgcctaatgagtgagc
1639 cgcGgatcccggacaccatcgaatgGCGCAAaacctttc
Template: pJG729
Primers:
oED100 cgcCTCGAGgatcccggtgcctaatgag
oED101 gaaTTGtggATGactCCAgc
Template: pJG719
Insert (lacWT):
CTCGAGccggtgcctaatgagtgagctaacttacattaattgcgttgcgcTCActgcccgctttccagtcgggaaacctgtcgtgc
cagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgaga
cgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaa
tcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaac
gcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccct
cattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcga
gtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacc
caatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacat
caagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagccca
ctgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggc
acccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatca
gcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgc
gttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataa
cgttactggtttCACattcaccaccctgaattgactctcttccgggcgctATCATGccataccgcgaaaggttTTGCGCcattcga
tggtgtccgggaTCC

P_{lacT5} (*lacI*q)

Primers:
1642 cgcGGATCCcgacgctctcccttatgcg

1643 cgcGGTaccgttgtggAATTGTGAGCG
1638 cgcCTCGAGccggtgcctaatgagtgagc
1640 cgcGgatcccggacaccatcgaatgGtGCAAaacctttc
Template: pJG729
Insert (lacIq):

```
CTCGAGccggtgcctaatgagtgagctaacttacattaattgcgttgcgcTCActgcccgctttccagtcgggaaacctgtcgtgc
cagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgaga
cgggcaacagctgattgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaa
tcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaac
gcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccct
cattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgattgcga
gtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacc
caatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacat
caagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagccca
ctgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggc
acccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatca
gcaacgactgtttgcccgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgc
gttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataa
cgttactggtttCACattcaccaccctgaattgactctcttccgggcgctATCATGccataccgcgaaaggtttTTGCACcattcga
tggtgtccgggaTCC
```

Parent vector sequences

pJG744

```
GAATTCccaCTCGAGccaGGTACCaccTCTAGAcccGTCGACaccAAGCTTccaCTGCAGccaGAGCTCggctgctaacaaagcccgaaaggaagct
gagttggctgctgccaccgCTGCTGGTTCGCTCATAAGTAAaaaacggcacctggtgccgtttttttgtctgaaacaagctgagcaataactagcat
aaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcggcgca
ttaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcg
ccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttga
ttagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttc
caaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattt
aacaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattt
ttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggatta
tcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcga
ttccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccgg
tgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgtta
ttcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactg
ccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatc
atcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacg
ctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagccc
atttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacaccccttgtatt
actgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaa
aggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagag
ctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaaga
actctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaag
acgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatac
ctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacga
gggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcagggg
gcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccct
gattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcgg
aagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcaatggtgcactctcagtacaatctgctctgatgccgcatag
ttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtct
gctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctg
cggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtct
ggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtCACtgatgcctccgtgtaagggggatttctgttcatgggggtaatgat
accgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatgg
atgcggcgggaccagagaaaaatcaCTCAGGGTCAATGCCAGCGcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcga
tgcagatccg
```

pJG747

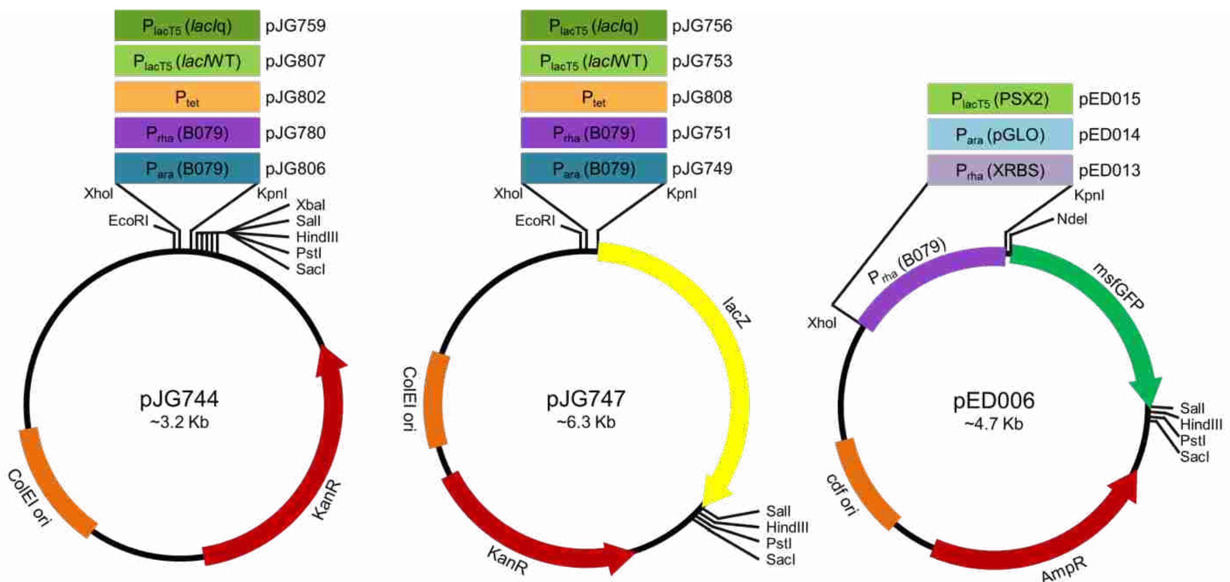
```
GAATTCccaCTCGAGccaGGTaccACAGGAAACAGCTatgACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCT
GGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC
GCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGT
```

CGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACG
GAGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGCGTTAACTCGG
CGTTTCATCTGTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGA
AAACCGCCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCG
TTGCTGCATAAACCGACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTACTGGAGGCTGAAGTTCAGA
TGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAAT
TATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGT
GCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGC
TGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCATCATCCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGA
TATCCTGCTGATGAAGCAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACGGCCTG
TATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGAGCGAACGCG
TAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTA
TCGCTGGATCAAATCTGTCGATCCTTCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCG
CGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCG
AATACGCCCACGCGATGGGTAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGGGCGGCTTCGTCTGGGA
CTGGGTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTTC
TGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCCGGGC
AAACCATCGAAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGG
TGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTA
CGCGTAGTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAGTGTGACGCTCC
CCGCCGCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGG
CTTTCTTTCACAGATGTGGATTGGCGATAAAAAACAACTGCTGACGCCGCTGCGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTA
AGTGAAGCGACCCGCATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCACGGCAG
ATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGATGG
TAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCA
GAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACA
TGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTT
CAACATCAGCCGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATATCGACGGTTTC
CATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTCTGGTGTC
AAAAAtaaTAATAACCGGGCAGGCCATGTCTGCCgtcGACaccAAGCTTccaCTGCAGccaGAGCTCggctgctaacaaagcccgaaaggaagctga
gttggctgctgccaccgCTGCTGGTTCGCTCATAAGTAAaaaacggcacctggtgccgtttttttgtctgaaacaagctgagcaataactagcataa
ccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcggcgcatt
aagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgcc
acgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgatt
agggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttcca
aactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaa
caaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttt
ctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatc
aataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgatt
ccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtg
agaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttatt
cattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgcc
agcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcat
caggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgct
acctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccat
ttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacaccccttgtattac
tgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaag
gatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagct
accaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaac
tctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagac
gatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacct
acagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagg
gagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggc
ggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctga
ttctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaa
gagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcaatggtgcactctcagtacaatctgctctgatgccgcatagtt
aagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgc
tcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcg
gtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctgg
cttctgataaagcgggccatgttaagggcgtttttttcctgtttggtCACtgatgcctccgtgtaagggggatttctgttcatgggggtaatgatac
cgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggat
gcggcgggaccagagaaaaatcaCTCAGGGTCAATGCCAGCGcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatg
cagatccg

pED006

CTCgagCATCGTCGGCATCGGCATGGCGAAttaATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGTTTCCCGGGTAAACACC
ACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAATATCACTGATTAACAGGCGGCTATGCTGGAGAAGATATTGCGCATGACACACTC
TGACCTGTCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAACGCGCTCACTGCACGATGCCTCATCACAAAATTT
ATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTATCCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAATCACTGGTG
TAACGATGGCGATTCAGCAACATCACCAACTGCCCGAACAGCAACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCATGCTCAAGCT

GACCGATAACCTGCCGCGCCTGCGCCATCCCCATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAAATCCCGGAATCGCCCCCTGCCA
GTCAAGATTCAGCTTCAGACGCTCCGGGCAATAAATAATATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCAGCATGAATG
TAAAAGAGATCGCCACGGGTAATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAAAATCATGTG
TATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGCAAAAAAATCATCTTTGAGAAGTTTTAACTGATG
CGCcacCGTGGCTACCTCGGCCAGAGAACGAAGTTGATTATTCGCAATATGGCGTACAAATACGTTGAGAAGATTCGCGttaTTGCAGAAAGCCATC
CCGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGTGGTTACTGTCGCTGAATCCACAGCGATAGGCGATG
TCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGTCGCAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCGTTTGCTGCT
TAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTGATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAGCCAGGCCAG
AAGCAAGTTGAGACGTGATGCGCTGTTTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTGCATAAACAAGATCTCGCGACTG
GCGGTCGAGGGTAAATCATTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGCGCCAGTGAG
ACGGATACTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCACATTGGTCAGACA
CAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTGCCACCGGTGATGGTATAGGGCTGCCCATTAAACACATGA
ATACCCGTGCCATGTTCGACAATCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCGCCACGGACG
CGTTACCAGACGGAAAAAAATCCACACTATGTAATACGGTcatACTGGCCTCCTGATGTCGTCAACACGGCGAAATAGTAATCACGAGGTCAGGTTC
TTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCAAGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCACATCACCAC
AATTCAGCAAATTGTGAACATCATCACGTTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCT
TTTTAGACTGGTCGTaATGAAATTCAGCggtaccGGAGGtcacatatgAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCAC
CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGAC
TTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCG
AGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAG
CCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCCACAACGTCGAGGACGGCAGCGTGCAGCTCGCC
GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCA
ACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGtaaGTCGACaccAAGCT
TccaCTGCAGccaGAGCTCggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccc
tgataaatgcttcaataatattgaaaaaggaagagtATGagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctg
tttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaa
gatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccggg
caagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaa
gagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttt
gcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgta
gcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttg
caggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcact
ggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggt
gcctcactgattaagcattggTAActgtcagaccaagtttactcGTAAAAACCCGCTTCGGCGGGTTTTTTTATGctagggcggttcagtagaaaag
atcaaaggatcttcttgagatcctttttttctgcgcgtaatctttttgccctgtaaacgaaaaaaccacctggggaggtggtttgatcgaaggttaag
tcagttggggaactgcttaaccgtggtaactggctttcgcagagcacagcaaccaaatctgtccttccagtgtagccggactttggcgcacacttca
agagcaaccgcgtgtttagctaaacaaatcctctgcgaactcccagttaccaatggctgctgccagtggcgttttaccgtgcttttccgggttggac
tcaagtgaacagttaccggataaggcgcagcagtcgggctgaacggggagttcttgcttacagcccagcttggagcgaacgacctacaccgagccga
gataccagtgtgtgagctatgagaaagcgccacacttcccgtaagggagaaaggcggaacaggtatccggtaaacggcagggtcggaacaggagagc
gcaagagggagcgacccgccggaaacggtggggatctttaagtcctgtcgggtttcgcccgtactgtcagattcatggttgagcctcacggctccca
cagatgcaccggaaaagcgtctgtttatgtgaactctggcaggagggcggagcctatggaaaaacgccaccggcgcggccctgctgttttgcctcac
atgttagtcccctgcttatccacggaatctgtgggtaactttgtatgtgtccgcagcgcccgccgcagtctcacgcccggagcgtagcgaccgagtg
aa



NCR peptide primer, insert, and plasmid sequences

Insert primers

NCR247
Peptide sequence:
RNGCIVDPRCPYQQCRRPLYCRRR
Primers:
oED0013 `cgcGGTACCGGAGGtcacatATGcgtaatggttgtatcgtcgatccacgttgtccataccaac`
oED0024 `cgcGAGCTCTTAacgacggcggcaatataacggacggcgacactgttggtatggacaacgtgga`
oED0025 `cgcCTGCAGcgcgtaatggttgtatcgtcgatccacgttgtccataccaacagtg`

NCR169
Peptide sequence:
EDIGHIKYCGIVDDCYKSKKPLFKIWKCVENVCVLWYK
Primers:
oED0026
`cgcGGTACCGGAGGtcacatATGgaagatatcggccacattaaatactgcggtatcgtggatgactgttacaaatccaagaaacc`
oED0027 `cgcGAGCTCTTAtttataccacagtacacagacgttttcaacgcacttccagatcttgaacagcggtttcttggatt`
`tgtaac`
oED0028 `cgcCTGCAGcggaagatatcggccacattaaatactgcggtatcgtggatgactgttacaaatccaagaaaccg`

NCR211
Peptide sequence:
DRECDTDTECQKKFPGVNAHHLWCDNGNCVSYPK
Primers:
oED0033 `cgcGGTACCGGAGGtcacatATGgatcgtgagtgcgacacggatactgaatgtcagaaaaagtttccaggtgtcaa`
`tgcgc`
oED0034 `cgcGAGCTCTTActtcggatagctcacgcagttgccattgtcacaccacaggtgatgcgcattgacacctggaaac`
oED0035 `cgcCTGCAGcggatcgtgagtgcgacacggatactgaatgtcagaaaaagtttccaggtgtcaatgcgc`

NFS1Fix-
Peptide sequence:
ASPFYCVDDDYFCFGLCLPPMIDHCTLRGQCICITISTEVES
Primers:
oED0036 `cgcGGTACCGGAGGtcacatATGgcgagcccgttttactgtgttgatgatgactatttctgctttggcctgtgttt`
`accacctatgattg`
oED0037 `cgcGAGCTCTTAcgattcaacctcagtgctaatcgtaatacagatgcactgaccgcgcaaggtgcagtgatcaatc`
`ataggtggtaaacac`
oED0038
`cgcCTGCAGcggcgagcccgttttactgtgttgatgatgactatttctgctttggcctgtgtttaccacctatgattg`

NFS1Fix+
Peptide sequence:
ASPFYCDEDDYFCFGLCLPPMIDHCTLGGQCICITISTEVES
Primers:
oED0039 `cgcGGTACCGGAGGtcacatATGgcgagtccgttttactgtgacgaagatgactatttctgctttggcctgtgt`
`ttaccacctatgattg`

oED0040 cgcGAGCTCTTAcgattcaacctcagtgctaatcgtaatacagatgcactgaccgcccaaggtgcagtgatcaat
cataggtggtaaacac

oED0041 cgcCTGCAGcggcgagtccgttttactgtgacgaagatgactatttctgctttggcctgtgtttaccacctatgat
tg

NFS2Fix-

Peptide sequence:

KEECTYAADCYKRYPRWSLLPNYCIEGSCYSDFLNSGKKYLSP

Primers:

oED0042 cgcGGTACCGGAGGtcacatATGaaagaagagtgcacgtatgcggccgattgttacaagcgctatccgcgttgg
tctctgttacctaactattgc

oED0043 cgcGAGCTCTTAtgggctcagtttcttgccagaatttaagaagtcactgtaacacgaaccttcaatgcaatagt
taggtaacagag

oED0044 cgcCTGCAGcgaaagaagagtgcacgtatgcggccgattgttacaagcgctatccgcgttggtctctgttacct
aactattgc

NFS2Fix+

Peptide sequence:

NEECTCAADCYKRYPRWSLLPTYCIEGSCYSDFLNSGKKYLSP

Primers:

oED0045 cgcGGTACCGGAGGtcacatATGaacgaagagtgcacgtgtgcggccgattgttacaagcgctatccgcgttggt
ctctgttacctacctattgc

oED0046 cgcGAGCTCTTAtgggctcagatatttcttgccagaatttaagaagtcactgtaacacgaaccttcaatgcaata
ggtaggtaacagag

oED0047 cgcCTGCAGcgaacgaagagtgcacgtgtgcggccgattgttacaagcgctatccgcgttggtctctgttaccta
cctattgc

phoA

Primers:

oED089 cgcGGTACCTTAAGGATTAACAAtgAAACAAAGCACTATTGCACTG

oED090 cgcGAGCTCttaTTTCAGCCCCAGAGCGG

oED091 cgcCTGCAGcgCCTGTTCTGGAAAACCGGG

Template: B079 (MG1655)

Insert (full length, pJG966):

```
GGTACCTTAAGGATTAACAAtgAAACAAAGCACTATTGCACTGGCACTCTTACCGTTACTGTTTACCCCTGTGACAAAAGCCCGGA
CACCAGAAATGCCTGTTCTGGAAAACCGGGCTGCTCAGGGCGATATTACTGCACCCGGCGGTGCTCGCCGTTTAACGGGTGATCAG
ACTGCCGCTCTGCGTGATTCTCTTAGCGATAAACCTGCAAAAAATATTATTTTGCTGATTGGCGATGGGATGGGGGACTCGGAAAT
TACTGCCGCACGTAATTATGCCGAAGGTGCGGGCGGCTTTTTTAAAGGTATAGATGCCTTACCGCTTACCGGGCAATACACTCACT
ATGCGCTGAATAAAAAAACCGGCAAACCGGACTACGTCACCGACTCGGCTGCATCAGCAACCGCCTGGTCAACCGGTGTCAAAACC
TATAACGGCGCGCTGGGCGTCGATATTCACGAAAAAGATCACCCAACGATTCTGGAAATGGCAAAAGCCGCAGGTCTGGCGACCGG
TAACGTTTCTACCGCAGAGTTGCAGGATGCCACGCCCGCTGCGCTGGTGGCACATGTGACCTCGCGCAAATGCTACGGTCCGAGCG
CGACCAGTGAAAAATGTCCGGGTAACGCTCTGGAAAAAGGCGGAAAAGGATCGATTACCGAACAGCTGCTTAACGCTCGTGCCGAC
GTTACGCTTGGCGGCGGCGCAAAAACCTTTGCTGAAACGGCAACCGCTGGTGAATGGCAGGGAAAAACGCTGCGTGAACAGGCACA
GGCGCGTGGTTATCAGTTGGTGAGCGATGCTGCCTCACTGAATTCGGTGACGGAAGCGAATCAGCAAAAACCCCTGCTTGGCCTGT
TTGCTGACGGCAATATGCCAGTGCGCTGGCTAGGACCGAAAGCAACGTACCATGGCAATATCGATAAGCCCGCAGTCACCTGTACG
CCAAATCCGCAACGTAATGACAGTGTACCAACCCTGGCGCAGATGACCGACAAAGCCATTGAATTGTTGAGTAAAAATGAGAAAGG
CTTTTTCCTGCAAGTTGAAGGTGCGTCAATCGATAAACAGGATCATGCTGCGAATCCTTGTGGGCAAATTGGCGAGACGGTCGATC
TCGATGAAGCCGTACAACGGGCGCTGGAATTCGCTAAAAAGGAGGGTAACACGCTGGTCATAGTCACCGCTGATCACGCCCACGCC
AGCCAGATTGTTGCGCCGGATACCAAAGCTCCGGGCCTCACCCAGGCGCTAAATACCAAAGATGGCGCAGTGATGGTGATGAGTTA
CGGGAACTCCGAAGAGGATTCACAAGAACATACCGGCAGTCAGTTGCGTATTGCGGCGTATGGCCCGCATGCCGCCAATGTTGTTG
GACTGACCGACCAGACCGATCTCTTCTACACCATGAAAGCCGCTCTGGGGCTGAAAtaaGAGCTC
```

Insert (partial, pJG996):

```
CTGCAgcGCCTGTTCTGGAAAACCGGGCTGCTCAGGGCGATATTACTGCACCCGGCGGTGCTCGCCGTTTAACGGGTGATCAGACT
GCCGCTCTGCGTGATTCTCTTAGCGATAAACCTGCAAAAAATATTATTTTGCTGATTGGCGATGGGATGGGGGACTCGGAAATTAC
TGCCGCACGTAATTATGCCGAAGGTGCGGGCGGCTTTTTTAAAGGTATAGATGCCTTACCGCTTACCGGGCAATACACTCACTATG
CGCTGAATAAAAAAACCGGCAAACCGGACTACGTCACCGACTCGGCTGCATCAGCAACCGCCTGGTCAACCGGTGTCAAAACCTAT
AACGGCGCGCTGGGCGTCGATATTCACGAAAAAGATCACCCAACGATTCTGGAAATGGCAAAAGCCGCAGGTCTGGCGACCGGTAA
CGTTTCTACCGCAGAGTTGCAGGATGCCACGCCCGCTGCGCTGGTGGCACATGTGACCTCGCGCAAATGCTACGGTCCGAGCGCGA
CCAGTGAAAAATGTCCGGGTAACGCTCTGGAAAAAGGCGGAAAAGGATCGATTACCGAACAGCTGCTTAACGCTCGTGCCGACGTT
ACGCTTGGCGGCGGCGCAAAAACCTTTGCTGAAACGGCAACCGCTGGTGAATGGCAGGGAAAAACGCTGCGTGAACAGGCACAGGC
GCGTGGTTATCAGTTGGTGAGCGATGCTGCCTCACTGAATTCGGTGACGGAAGCGAATCAGCAAAAACCCCTGCTTGGCCTGTTTG
CTGACGGCAATATGCCAGTGCGCTGGCTAGGACCGAAAGCAACGTACCATGGCAATATCGATAAGCCCGCAGTCACCTGTACGCCA
AATCCGCAACGTAATGACAGTGTACCAACCCTGGCGCAGATGACCGACAAAGCCATTGAATTGTTGAGTAAAAATGAGAAAGGCTT
TTTCCTGCAAGTTGAAGGTGCGTCAATCGATAAACAGGATCATGCTGCGAATCCTTGTGGGCAAATTGGCGAGACGGTCGATCTCG
ATGAAGCCGTACAACGGGCGCTGGAATTCGCTAAAAAGGAGGGTAACACGCTGGTCATAGTCACCGCTGATCACGCCCACGCCAGC
CAGATTGTTGCGCCGGATACCAAAGCTCCGGGCCTCACCCAGGCGCTAAATACCAAAGATGGCGCAGTGATGGTGATGAGTTACGG
GAACTCCGAAGAGGATTCACAAGAACATACCGGCAGTCAGTTGCGTATTGCGGCGTATGGCCCGCATGCCGCCAATGTTGTTGGAC
TGACCGACCAGAGAGCTC
```

Sequencing primers:
oED022 GCCCATTTTCCTGTCAGTAACG
oED023 CTTATGAGCGAACCAGCAGcg

DsbA
Primers:
oED020 cgcGGTACCGGAGGtcacatatgAAAAAGATTTGGCTGGCGCTGGCTGGTT
oED021 GCGctgcagATGCGCTAAACGCTAAAACTAAACCAGCCAGCGCCAGCCAAA
Template: B079 (MG1655)

YiaD
Primers:
oED092 cgcGGTACCttaaggattaacaatgAAGAAACGTG
oED093 cgcGAGCTCttacgctgcagactgattgttattgttattcggTTTGTCGCGCAGCTTCGCTT
Template: B079 (MG1655)

Parent vectors

pJG780

```
GAATTCccaCTCgagCATCGTCGGCATCGGCATGGCGAAttaATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGTTTCCCGG
GTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAATATCACTGATTAACAGGCGGCTATGCTGGAGAAGATATTGCGCAT
GACACACTCTGACCTGTCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAACGCGCTCACTGCACGATGCCTCATC
ACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTATCCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAA
TCACTGGTGTAACGATGGCGATTCAGCAACATCACCAACTGCCCGAACAGCAACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCAT
GCTCAAGCTGACCGATAACCTGCCGCGCCTGCGCCATCCCCATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAAATCCCGGAATCGC
CCCCTGCCAGTCAAGATTCAGCTTCAGACGCTCCGGGCAATAAATAATATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCA
GCATGAATGTAAAAGAGATCGCCACGGGTAATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAA
AATCATGTGTATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGCAAAAAAATCATCTTTGAGAAGTTT
TAACTGATGCGCcacCGTGGCTACCTCGGCCAGAGAACGAAGTTGATTATTCGCAATATGGCGTACAAATACGTTGAGAAGATTCGCGttaTTGCAG
AAAGCCATCCCGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGTGGTTACTGTCGCTGAATCCACAGCGA
TAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGTCGCAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCG
TTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTGATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAG
CCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGTTTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTGCATAAACAAGATC
TCGCGACTGGCGGTCGAGGGTAAATCATTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGC
GCCAGTGAGACGGATACTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCACATT
GGTCAGACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTGCCACCGGTGATGGTATAGGGCTGCCCATTA
AACACATGAATACCCGTGCCATGTTCGACAATCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCG
CCACGGACGCGTTACCAGACGGAAAAAAATCCACACTATGTAATACGGTcatACTGGCCTCCTGATGTCGTCAACACGGCGAAATAGTAATCACGAG
GTCAGGTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCAAGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCA
CATCACCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAAT
TCAGGCGCTTTTTAGACTGGTCGTaATGAAATTCAGCggtACCaccTCTAGAcccGTCGACaccAAGCTTccaCTGCAGccaGAGCTCggctgctaa
caaagcccgaaaggaagctgagttggctgctgccaccgCTGCTGGTTCGCTCATAAGTAAaaaacggcacctggtgccgtttttttgtctgaaaca
```

gctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatggga
cgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgct
ttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacc
tcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctt
taatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattgg
ttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcgg
aacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaa
tttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaaga
tcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccat
gagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcact
cgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgc
aaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcag
tggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatc
tgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgc
ccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggc
tcataacaccccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcag
accccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggt
ttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtag
ttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtc
ttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgac
ctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtc
ggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttt
tgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgtt
ctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgag
tcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcaatggtgcactctcagtacaat
ctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacg
cgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcacc
gaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttc
tccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtCACtgatgcctccgtgtaagggggatttc
tgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgaggg
taaacaactggcggtatggatgcggcgggaccagagaaaaatcaCTCAGGGTCAATGCCAGCGcttcgttaatacagatgtaggtgttccacagggt
agccagcagcatcctgcgatgcagatccg

pED024

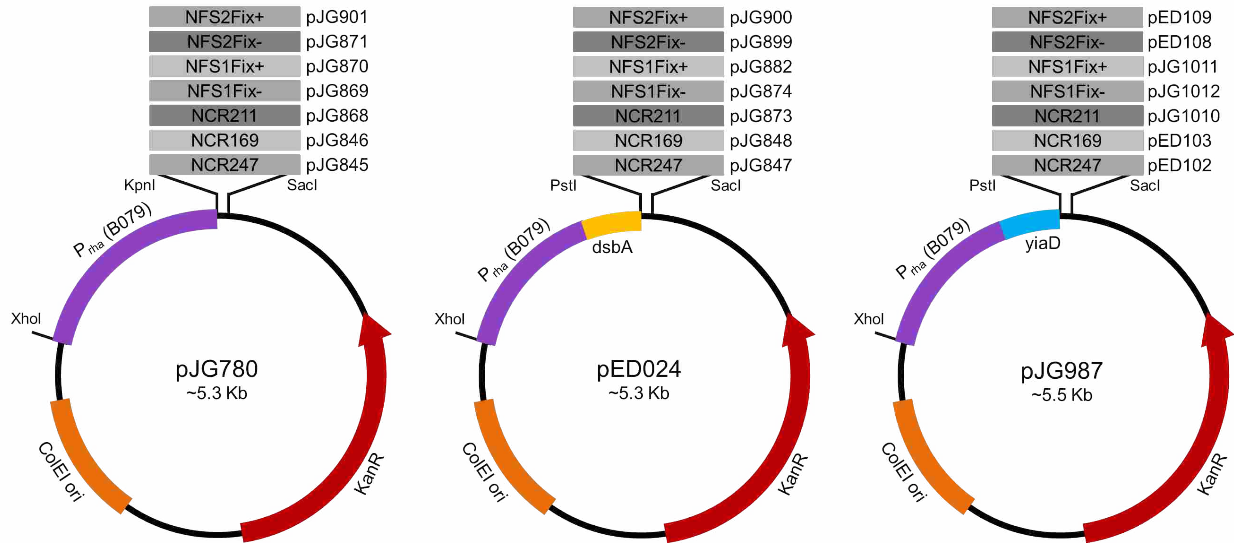
GAATTCccaCTCgagCATCGTCGGCATCGGCATGGCGAAttaATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGTTTCCCGG
GTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAATATCACTGATTAACAGGCGGCTATGCTGGAGAAGATATTGCGCAT
GACACACTCTGACCTGTCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAACGCGCTCACTGCACGATGCCTCATC
ACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTATCCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAA
TCACTGGTGTAACGATGGCGATTCAGCAACATCACCAACTGCCCGAACAGCAACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCAT
GCTCAAGCTGACCGATAACCTGCCGCGCCTGCGCCATCCCCATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAAATCCCGGAATCGC
CCCCTGCCAGTCAAGATTCAGCTTCAGACGCTCCGGGCAATAAATAATATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCA
GCATGAATGTAAAAGAGATCGCCACGGGTAATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAA
AATCATGTGTATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGCAAAAAAATCATCTTTGAGAAGTTT
TAACTGATGCGCcacCGTGGCTACCTCGGCCAGAGAACGAAGTTGATTATTCGCAATATGGCGTACAAATACGTTGAGAAGATTCGCGttaTTGCAG
AAAGCCATCCCGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGTGGTTACTGTCGCTGAATCCACAGCGA
TAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGTCGCAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCG
TTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTGATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAG
CCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGTTTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTGCATAAACAAGATC
TCGCGACTGGCGGTCGAGGGTAAATCATTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGC
GCCAGTGAGACGGATACTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCACATT
GGTCAGACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTGCCACCGGTGATGGTATAGGGCTGCCCATTA
AACACATGAATACCCGTGCCATGTTCGACAATCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCG
CCACGGACGCGTTACCAGACGGAAAAAAATCCACACTATGTAATACGGTcatACTGGCCTCCTGATGTCGTCAACACGGCGAAATAGTAATCACGAG
GTCAGGTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCAAGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCA
CATCACCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAAT
TCAGGCGCTTTTTAGACTGGTCGTaATGAAATTCAGCggtACCGGAGGtcacatatgAAAAAGATTTGGCTGGCGCTGGCTGGTTTAGTTTTAGCGT
TTAGCGCATctgcagcgtaaGAGCTCggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgCTGCTGGTTCGCTCATAAGTAAaa
aacggcacctggtgccgtttttttgtctgaaacaagctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttttgct
gaaaggaggaactatatccggattggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgcta
cacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggct
ccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacg
gtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttg
atttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgcttac
aatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaattaattct
tagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggag
aaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctc
gtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttca
acaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgc
tgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttc
taatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcata

aattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggct
tcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttggaatttaatcg
cggcctagagcaagacgtttcccgttgaatatggctcataacaccccttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaat
cccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgc
ttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgc
agataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgtt
accagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggg
ggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaaggga
gaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgt
cgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacgg
ttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgata
ccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcgg
tatttcacaccgcaatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcat
ggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccggg
agctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgt
ctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctg
tttggtCACtgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgat
gatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcaCTCAGGGTCAATGCCAGC
Gcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccg

pJG987

GAATTCccaCTCgagCATCGTCGGCATCGGCATGGCGAAttaATCTTTCTGCGAATTGAGATGACGCCACTGGCTGGGCGTCATCCCGGTTTCCCGG
GTAAACACCACCGAAAAATAGTTACTATCTTCAAAGCCACATTCGGTCGAAATATCACTGATTAACAGGCGGCTATGCTGGAGAAGATATTGCGCAT
GACACACTCTGACCTGTCGCAGATATTGATTGATGGTCATTCCAGTCTGCTGGCGAAATTGCTGACGCAAAACGCGCTCACTGCACGATGCCTCATC
ACAAAATTTATCCAGCGCAAAGGGACTTTTCAGGCTAGCCGCCAGCCGGGTAATCAGCTTATCCAGCAACGTTTCGCTGGATGTTGGCGGCAACGAA
TCACTGGTGTAACGATGGCGATTCAGCAACATCACCAACTGCCCGAACAGCAACTCAGCCATTTCGTTAGCAAACGGCACATGCTGACTACTTTCAT
GCTCAAGCTGACCGATAACCTGCCGCGCCTGCGCCATCCCCATGCTACCTAAGCGCCAGTGTGGTTGCCCTGCGCTGGCGTTAAATCCCGGAATCGC
CCCCTGCCAGTCAAGATTCAGCTTCAGACGCTCCGGGCAATAAATAATATTCTGCAAAACCAGATCGTTAACGGAAGCGTAGGAGTGTTTATCGTCA
GCATGAATGTAAAAGAGATCGCCACGGGTAATGCGATAAGGGCGATCGTTGAGTACATGCAGGCCATTACCGCGCCAGACAATCACCAGCTCACAAA
AATCATGTGTATGTTCAGCAAAGACATCTTGCGGATAACGGTCAGCCACAGCGACTGCCTGCTGGTCGCTGGCAAAAAAATCATCTTTGAGAAGTTT
TAACTGATGCGCcacCGTGGCTACCTCGGCCAGAGAACGAAGTTGATTATTCGCAATATGGCGTACAAATACGTTGAGAAGATTCGCGttaTTGCAG
AAAGCCATCCCGTCCCTGGCGAATATCACGCGGTGACCAGTTAAACTCTCGGCGAAAAAGCGTCGAAAAGTGGTTACTGTCGCTGAATCCACAGCGA
TAGGCGATGTCAGTAACGCTGGCCTCGCTGTGGCGTAGCAGATGTCGGGCTTTCATCAGTCGCAGGCGGTTCAGGTATCGCTGAGGCGTCAGTCCCG
TTTGCTGCTTAAGCTGCCGATGTAGCGTACGCAGTGAAAGAGAAAATTGATCCGCCACGGCATCCCAATTCACCTCATCGGCAAAATGGTCCTCCAG
CCAGGCCAGAAGCAAGTTGAGACGTGATGCGCTGTTTTCCAGGTTCTCCTGCAAACTGCTTTTACGCAGCAAGAGCAGTAATTGCATAAACAAGATC
TCGCGACTGGCGGTCGAGGGTAAATCATTTTCCCCTTCCTGCTGTTCCATCTGTGCAACCAGCTGTCGCACCTGCTGCAATACGCTGTGGTTAACGC
GCCAGTGAGACGGATACTGCCCATCCAGCTCTTGTGGCAGCAACTGATTCAGCCCGGCGAGAAACTGAAATCGATCCGGCGAGCGATACAGCACATT
GGTCAGACACAGATTATCGGTATGTTCATACAGATGCCGATCATGATCGCGTACGAAACAGACCGTGCCACCGGTGATGGTATAGGGCTGCCCATTA
AACACATGAATACCCGTGCCATGTTCGACAATCACAATTTCATGAAAATCATGATGATGTTCAGGAAAATCCGCCTGCGGGAGCCGGGGTTCTATCG
CCACGGACGCGTTACCAGACGGAAAAAAATCCACACTATGTAATACGGTcatACTGGCCTCCTGATGTCGTCAACACGGCGAAATAGTAATCACGAG
GTCAGGTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAACGTCGATTTTTCAAGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCA
CATCACCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTCCCTGGTTGCCAATGGCCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAAT
TCAGGCGCTTTTTAGACTGGTCGTaATGAAATTCAGCggtACCttaaggattaacaatgAAGAAACGTGTTTATCTTATTGCCGCCGTAGTGAGTGG
CGCTCTGGCGGTATCTGGCTGCACAACTAACCCTTACACCGGCGAACGCGAAGCAGGTAAATCTGCTATCGGCGCAGGTCTGGGCTCTCTCGTGGGC
GCGGGTATTGGTGCGCTCTCTTCTTCGAAGAAAGATCGCGGTAAAGGCGCGCTGATTGGCGCAGCAGCAGGCGCAGCTCTGGGCGGCGGCGTTGGTT
ATTACATGGATGTGCAGGAAGCGAAGCTGCGCGACAAAccgaataacaataacaatcagtctgcagcgtaaGAGCTCggctgctaacaaagcccgaa
aggaagctgagttggctgctgccaccgCTGCTGGTTCGCTCATAAGTAAaaaacggcacctggtgccgtttttttgtctgaaacaagctgagcaata
actagcataaccccttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgccctgta
gcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttc
ctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaa
aaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggac
tcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatga
gctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctattt
gtttatttttctaaatacattcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatat
caggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcg
gtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgact
gaatccggtgagaatggcaaaagtttatgcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaacca
aaccgttattcattcgtgattgcgcctgagcgagacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcag
gaacactgccagcgcatcaacaatattttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaac
catgcatcatcaggagtacggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcat
tggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatc
gcgagcccatttatacccatataaatcagcatccatgttggaatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacaccc
cttgtattactgtttatgtaagcagacagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaa
aagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccgg
atcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccacca
cttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttg
gactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaac
tgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggaga
gcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcg
tcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgt

```
tatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcga
ggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcaatggtgcactctcagtacaatctgctctgatg
ccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgg
gcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcga
ggcagctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgt
taatgtctggcttctgataaagcgggccatgttaagggcggtttttcctgtttggtCACtgatgcctccgtgtaagggggatttctgttcatgggg
gtaatgataccgatgaaacgagagaggatgctcacgatacgggttactgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactgg
cggtatggatgcggcgggaccagagaaaaatcaCTCAGGGTCAATGCCAGCGcttcgttaatacagatgtaggtgttccacagggtagccagcagca
tcctgcgatgcagatccg
```



Illumina library primer sequences

Peptide library primers

oED110 `cgcGGATCCggtNDTndcNDTndcNDTndcNDTndcTGACGACGACGTCGActgagag`
oED111 `cgcGAGCTCaagttgtaagccttaactcaggatcgctagcttctctcagTCGACGTCG`
*Note: while we typically order primers from Invitrogen, these were ordered from IDT due to bias found in the Invitrogen versions of the primers.

Full length insert:

```
cgcGGATCCggtNDTndcNDTndcNDTndcNDTndcTGACGACGACGTCGActgagagaagctagcgatcctgagttaaggcttac
aacttGAGCTCgcg
```

Illumina primers

Reverse primers, forward read:

oED112
`AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNgTCGACGTCGTCGTCAg`

oED113
`AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNgTCGACGTCGTCGTCAg`

oED114
`AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNgTCGACGTCGTCGTCAg`

P5flowcellanchor Illumina Read 1 primer vector seq

*Note: all three forward read primers were combined for the PCR to increase diversity for better cluster separation during Illumina.

Forward primers, index read:
oED115
`CAAGCAGAAGACGGCATACGAGATcgtgatGTGACTGGAGTTCAGACGTGTGCTCTTCCGATcGCCACAACGTCGAGGAC`
oED116
`CAAGCAGAAGACGGCATACGAGATacatcgGTGACTGGAGTTCAGACGTGTGCTCTTCCGATcGCCACAACGTCGAGGAC`
oED117
`CAAGCAGAAGACGGCATACGAGATgcctaaGTGACTGGAGTTCAGACGTGTGCTCTTCCGATcGCCACAACGTCGAGGAC`
oED118
`CAAGCAGAAGACGGCATACGAGATattggcGTGACTGGAGTTCAGACGTGTGCTCTTCCGATcGCCACAACGTCGAGGAC`
*Note: these are the reverse complement of the sequence shown below.

```
GTCCTCGACGTTGTGGC|GATCGGAAGAGCACACGTCTGAACTCCAGTCACnnnnnnATCTCGTATGCCGTCTTCTGCTTG
vector seq         Index primer binding site         INDEX      P7flowcellanchor
```

Final PCR product:
```
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTnnnnnnGTCGACGTCGTCGTCA
ghnAHNghnAHNghnAHNghnAHNaccGGAtccCTTGTACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGA
ACTCCAGCAGGACCATGTGATCGCGCTTCTCGTTGGGGTCTTTGCTCAGCTTGGACTGGGTGCTCAGGTAGTGGTTGTCG
GGCAGCAGCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCCTCGACGTT
GTGGCgATCGGAAGAGCACACGTCTGAACTCCAGTCACnnnnnnATCTCGTATGCCGTCTTCTGCTTG
```

Final Illumina read:
`NNNNNNgTCGACGTCGTCGTCAghnAHNghnAHNghnAHNghnAHNacc`

PepSeq confirmation primer and plasmid sequences

Insert primers

oED123 `cgcGAGCTCaagttgtaagccttaactcaggatcgctagcttctctcagTCGACGTCG`
Reverse primer for all inserts
oED129 `cgcGGATCCggtcattactatatcatttactttatcTGACGACGACGTCGActgagag`
Peptide: HYYIIYFI
oED130 `cgcGGATCCggtgatttcgatggcgttggcattagcTGACGACGACGTCGActgagag`
Peptide: DFDGVGIS
oED131 `cgcGGATCCggttttcgcattttctatcgcattttcTGACGACGACGTCGActgagag`
Peptide: FRIFYRIF
oED132 `cgcGGATCCggtagttacgatctctgtggcaatgacTGACGACGACGTCGActgagag`
Peptide: SYDLCGND
oED133 `cgcGGATCCggtttttgccttaactgtggccgtgacTGACGACGACGTCGActgagag`
Peptide: FCLNCGRD
oED134 `cgcGGATCCggtaatcgccttgacggtagcgttgtcTGACGACGACGTCGActgagag`
Peptide: NRLDGSVV
oED135 `cgcGGATCCggttataactttatcggtttctttcgcTGACGACGACGTCGActgagag`
Peptide: YNFIGFFR
oED136 `cgcGGATCCggtagtcactttttctatggcgttttcTGACGACGACGTCGActgagag`
Peptide: SHFFYGVF

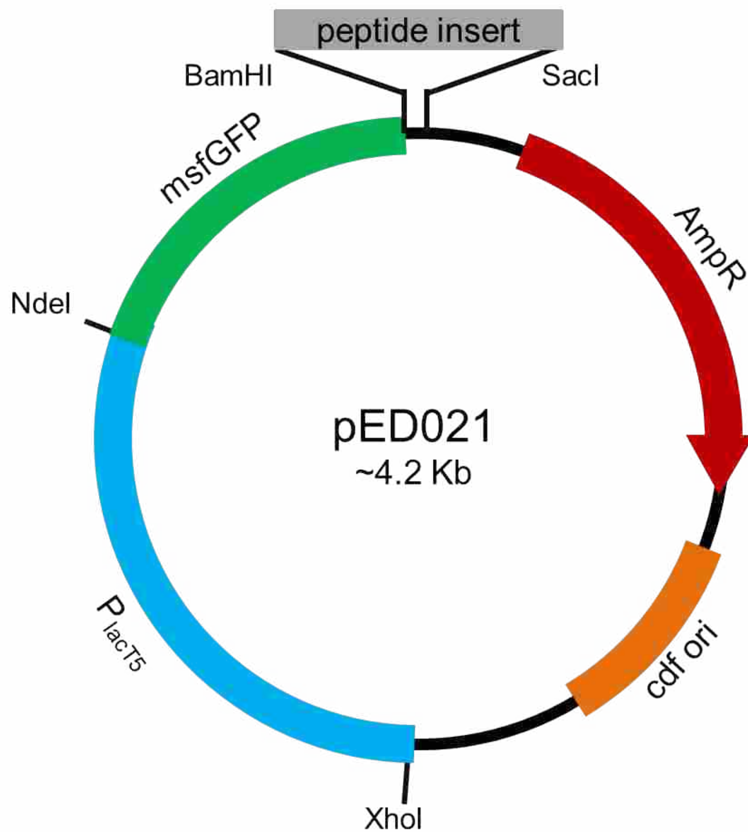
oED137 cgcGGATCCggtcgtttcctttactttggccatctcTGACGACGACGTCGActgagag
Peptide: RFLYFGHL
oED138 cgcGGATCCggtggtttctattacgtttgcattatcTGACGACGACGTCGActgagag
Peptide: GFYYVCII
oED139 cgcGGATCCggttattaccgtatcatttgcgttttcTGACGACGACGTCGActgagag
Peptide: YYRIICVF
oED140 cgcGGATCCggttgtctctatatcgttcgctttctcTGACGACGACGTCGActgagag
Peptide: CLYIVRFL
oED141 cgcGGATCCggttatgacgatttcagtttctatcacTGACGACGACGTCGActgagag
Peptide: YDDFSFYH
oED142 cgcGGATCCggtggtatcggttgcgatggcgatgacTGACGACGACGTCGActgagag
Peptide: GIGCDGDD
oED143 cgcGGATCCggttatttctgttactttggcgttctcTGACGACGACGTCGActgagag
Peptide: YFCYFGVL
oED144 cgcGGATCCggttatttctgttactttggcgcggcaTGACGACGACGTCGActgagag
Peptide: YFCYFGAA
oED145 cgcGGATCCggttatttctgttactttgccgcggcaTGACGACGACGTCGActgagag
Peptide: YFCYFAAA
oED146 cgcGGATCCggttatttcgcgtactttgccgcggcaTGACGACGACGTCGActgagag
Peptide: YFAYFAAA
oED147 cgcGGATCCggtcgttactattacattatctttgtcTGACGACGACGTCGActgagag
Peptide: RYYYIIFV
oED148 cgcGGATCCggttgttaccgtctctatatcattttcTGACGACGACGTCGActgagag
Peptide: CYRLYIIF

Parent vector

pED021

```
CTCGAGgatcgatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgcTCActgcccgctttccagtcgggaaacctgtcgtgccag
ctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctga
ttgcccttcaccgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcg
ggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatgtccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcc
cagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccag
tcgccttcccgttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggc
ccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatggg
tgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatc
agcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcaccca
gttgatcggcgcgagatttaatcgccgcgacaatttgcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcc
cgccagttgttgtgccacgcggttgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctgg
ttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttCACattcaccaccctgaattgactct
cttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattagga
agcagcccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcc
tgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaacc
gcacctgtggcgccggtgagatctcgatcccgcgaaatAATTGTGAGCGGATAACAATTacgagcttcatgcacagtgaaatcatgaaaaatttatT
TGCTTtgtgagcggataacaatTATAATatgtggAATTGTGAGCGCTCACAATTccacaacggtTCTAGAaataattttgtttaactttaagaagga
gatataCATatgAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGC
GCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC
CACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG
GAGCGCACCATCTCCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGG
GCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAA
CGGCATCAAGGCGAACTTCAAGATCCGCCACAACGTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGC
CCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCG
TGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGggatccTAAgagCTCggaaatgtgcgcggaacccctatttgtttatttttcta
aatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtATGagtattcaacatttccgt
gtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtg
cacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagt
```

```
tctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactca
ccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttc
tgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatga
agccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcc
cggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctg
gagccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggc
aactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggTAActgtcagaccaagtttactcGTAAAAACCCGCT
TCGGCGGGTTTTTTTATGctagggcggttcagtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatcttttgccctgtaaac
gaaaaaaccacctggggaggtggtttgatcgaaggttaagtcagttggggaactgcttaaccgtggtaactggctttcgcagagcacagcaaccaaa
tctgtccttccagtgtagccggactttggcgcacacttcaagagcaaccgcgtgtttagctaaacaaatcctctgcgaactcccagttaccaatggc
tgctgccagtggcgttttaccgtgcttttccgggttggactcaagtgaacagttaccggataaggcgcagcagtcgggctgaacggggagttcttgc
ttacagcccagcttggagcgaacgacctacaccgagccgagataccagtgtgtgagctatgagaaagcgccacacttcccgtaagggagaaaggcgg
aacaggtatccggtaaacggcagggtcggaacaggagagcgcaagagggagcgacccgccggaaacggtggggatctttaagtcctgtcgggtttcg
cccgtactgtcagattcatggttgagcctcacggctcccacagatgcaccggaaaagcgtctgtttatgtgaactctggcaggagggcggagcctat
ggaaaaacgccaccggcgcggccctgctgttttgcctcacatgttagtcccctgcttatccacggaatctgtgggtaactttgtatgtgtccgcagc
gcccgccgcagtctcacgcccggagcgtagcgaccgagtgaa
```



Peptide	Plasmid
FCLNGRD	pED078
NRLDGSVV	pED087
YNFIGFFR	pED088
SHFFYGVF	pED089
RFLYFGHL	pED096
GFYYVCII	pED097
YYRIICVF	pED098
CLYIVRFL	pED101
DFDGVGIS	pED080
FRIFYRIF	pED081
SYDLCGND	pED086
GIGCDGDD	pED095
YDDFSFYH	pED099
HYYIIYFI	pED085
CYRLYIIF	pED090
RYYYIIFV	pED100
YFCYFGVL	pED091
YFCYFGAA	pED092
YFCYFAAA	pED093
YFAYFAAA	pED094

APPENDIX II

Miller Assays

β-Gal Master Mix (per reaction)
600 µl Basal Buffer
0.5 mg ONPG
Vortex until ONPG is dissolved.
.5 µl 10% SDS
3 µl BME (β-Mercaptoethanol)
Typically this recipe will be multiplied by 10 or more to obtain a master mix.

Basal Buffer Recipe
500 m H_2O
4.3 g Na_2HPO_4
2.4 g NaH_2PO_4
0.75 g KCl
pH to 7.0 with about 1.5 ml 2N KOH
Add 1 ml of 1M $MgSO_4_7H_2O$
Add 1 ml of $CHCl_3$ to sterilize

Stop Buffer
1M Na_2CO_3
This is 53g/500 ml
Add 1 ml $CHCl_3$ to sterilize

phoA Assays

Add 2 ml of overnight culture to 50 ml of LB+ inducer to induce phoA expression. Incubate on shaker for 6 hours at 37°C. Take OD_{600} readings. Remove cultures and divide into 10 ml aliquots in 50 ml falcon tubes. Centrifuge at 3260 RPM at 4°C for 15 minutes. Dump and tap supernatant. To ensure a dry tube, keep all tubes upside down for 2-5 minutes, occasionally tapping additional moisture out of the tubes. Freeze at -80°C for at least 1 hour.
When prepared to do a periplasmic protein release procedure and subsequent phoA assay (approximately 3 hours total) make Periplasmic Isolation Buffer (PIB) and 0.1% SDS Solution. Scale PIB recipe to accommodate 1.2 ml per pellet. Do not make PNPP solution yet.
After making solutions, remove pellets from -80°C and keep on ice. Resuspend each pellet in 200 µl PIB. Transfer suspension to pre-chilled microcentrifuge tube on ice. Vortex each pellet for 10 seconds on high, returning each pellet to ice after vortexing. Repeat 6 times.
After vortexing add 800 µl pre-chilled ddH_2O to each pellet suspension for osmotic shock. Invert several times to mix. Incubate on ice for 10 minutes.
Centrifuge in cold room for 3 minutes. Carefully pipette supernatant into labeled tubes. *The supernatant contains released periplasmic proteins. Centrifuge pellet again for 1 minute in cold room. Remove and discard excess supernatant. Resuspend pellet in 1 ml PIB. *The resuspended pellet contains periplasmic membrane-bound proteins and proteins that were not released by vortex/osmotic shock.

Dilute each pellet and supernatant sample into ddH_2O to 1:100. Use the sample dilutions for the phoA assay. Now is prepare PNPP solution so it will be fresh (don't leave it sitting for more than an hour before use). Prepare for 75 µl PNPP solution per phoA assay tube.
Prepare microcentrifuge tubes for phoA assay.
Per reaction:
700 µl 1 M Tris pH 8.0
20 µl 0.1% SDS
40 µl chloroform
10 µl of diluted sample
Remember to prepare a blank that has all parts of the recipe except for a sample dilution.
Begin reaction by adding 75 µl PNPP solution to each tube and inverting several times to mix.
Mark starting time and incubate at 37°C for ~30 minutes, until samples turn yellow.
Stop reaction by adding 315 µl stop buffer (1 M KH_2PO_4). Centrifuge at full speed for 1 minute.
Take OD_{420} readings.

PNPP Solution (1 ml)
5 mg PNPP (p-Nitrophenyl Phosphate)
1 ml 1 M Tris pH 8

Periplasmic Isolation Buffer (10 ml)
3.5 ml 50% sucrose
1 ml 1 M Tris pH 8
40 µl 0.25 M EDTA
5.5 ml ddH_2O

0.1% SDS (1 ml)
10 µl 10% SDS
990 µl ddH_2O

Molecular cloning

Preparing vector

Vector is prepared using Zymo Research Miniprep Plasmid Classic kit
Run miniprep on analytical gel to visualize plasmid band.

Preparing insert

PCR amplification of insert from template (as done for promoter cloning)
per rxn (40 µl)
24 µl ddH_2O
8 µl Q5 buffer
1.5 µl 10 mM dNTPs
0.5 µl Q5 polymerase
3 µl 1:10 dilution of 100 µM forward primer
3 µl 1:10 dilution of 100 µM reverse primer
1 µl template

Program
96° 1 min
96° 15 s |
63° 15 s | x30
72° 20 s-2 min (depending on length of insert to be made) |
72° 1 min
4° ∞
Cleanup using Zymo column with 200 µl endowash and 400 µl plasmidwash. Run on gel.

PCR to extend 2 annealed oligos (as done for NCR peptides and PepSeq peptides)
per rxn (40 µl)
25 µl ddH_2O
8 µl Q5 buffer
3 µl dNTPs
1.5 µl Q5 polymerase
1.5 µl 100 µM forward primer
1.5 µl 100 µM reverse primer

96° 1:00
50° 20:00
55° 5:00
60° 5:00
65° 5:00
70° 5:00
4° ∞
Cleanup on Zymo column. Run on gel.

Annealing 2 oligos (often for multiple cloning sites)
per pair of oligos (30 µl rxn)
22 µl ddH_2O
3 µl 10x ligation buffer
2.5 µl 100 µM forward primer
2.5 µl 100 µM reverse primer
Put tube on a 95°C heat block, then cool slowly (take heat block out and set on bench for ~1 hr). When cool (<40°C), dilute into 1x ligation buffer. Use this for the insert in a typical ligation rxn (3 µl). No cleanup necessary.

Digestion with restriction enzymes

per rxn (30 µl)
10 µl ddH_2O
3 µl CutSmart buffer
15 µl DNA (from miniprep or PCR)
1 µl enzyme 1
1 µl enzyme 2

Incubate at 37°C for at least 1 hour. Add 1 µl CIP to vector for last hour if desired. Run on gel.

Purify on a 1% low melt gel. Cut out slices with razor. This slice can then be melted at 65°C and used in a ligation.

Ligating vector and insert together

per rxn (15 µl)
6.5 µl ddH_2O
1.5 µl 10x ligation buffer
1 µl ligase
3 µl vector
3 µl insert
Let sit at room temperature for at least 45 min. Heat inactivate for 5 min at 65°C.

Transforming plasmid into bacterial cells

Heat shock:
Mix 50 µl of thawed (on ice) chemically competent ($CaCl_2$) cells with 5 µl ligation mixture. Flick and incubate on ice 5-15 min. Heat shock at 42°C 35-45 s. Place immediately back on ice for 2-3 minutes. Add 500 µl LB and recover on shaker 30-90 minutes. Plate on appropriate antibiotic.
*note: Amp^R plasmids need only 30 minutes, while Kan^R plasmids need the full 90 minutes.

Chemically competent cells:
Grow cells in LB to OD of .4-.6. Put flask(s) on ice water 15 min. Combine into 1 flask. Pour 50 ml into 2 Falcon tubes (100 ml total). Pellet full speed (4500 rpm) at 4°C for 10 minutes. Dump and tap. Place back on ice.
Resuspend each pellet in 12 ml cold MgCa solution. Incubate on ice 20-30 minutes. Pellet as above. Dump and tap. Place back on ice.
Resuspend each pellet in 1 ml 100 mM $CaCl_2$. Combine in 1 tube. Incubate on ice 20-30 minutes. Add 0.4 ml 80% glycerol (or 0.2 ml per original Falcon tube). Mix well.
Freeze in 0.2 ml aliquots. Can flash freeze with liquid nitrogen if desired.
1 50 ml Falcon tube yields 12-15 tubes of competent cells.

Mg^{2+}/Ca^{2+}
3.25 g $MgCl_2 \cdot 6H_2O$
0.6 g $CaCl_2 \cdot 2H_2O$
200 ml dH_2O
Autoclave

100 mM $CaCl_2$
2.95 g $CaCl_2 \cdot 2H_2O$
200 ml dH_2O
Autoclave

Electroporation:
Cleanup ligation on Zymo column to remove additional salt. Add 4 µl cleaned ligation to 50 µl competent cells. Move 51 µl to cold cuvette, shock using micropulser. IMMEDIATELY add 1 ml LB, put on shaker for 60-90 minutes. Plate on appropriate antibiotic.

Electrocompetent cells:
Pre-chill sorvall and SLA-1500 rotor (in RIC; leave note telling tech that you need the machine on and cold). Add 1 ml of freshly saturated D050 in LB to 250 ml ECM3.
Shake at 37°C to OD of ~0.3-0.4 (~2h for D050). Place the 2 flasks in wet ice slurry and make sure cells stay ice cold for the rest of the protocol
For each 250 ml culture:
Pour cells into autoclaved 250 ml centrifuge bottle. Centrifuge at 8k for 7 minutes. Dump and tap thoroughly.
Resuspend pellet in 20 ml ice cold 5% glycerol (5%gly). Combine the pellets, then add 20 ml more of 5%gly (for 60 ml total). Centrifuge at 8k for 7 minutes. Dump and tap thoroughly.
Resuspend in 20 ml of 5%gly, then add 40 ml more (for 60 ml total). Centrifuge at 8k for 7 minutes. Dump and tap thoroughly. Resuspend in 0.4 ml of 5%gly. Use within 15 minutes.
Electroporate as 50 µl sample in 0.1cm cuvette.

0.5 L ECM3 medium
7 g tryptone
3 g yeast
1 g NaCl
Autoclave as 2x250 ml (in baffled flasks).
After autoclaving, add 2.7 ml of 1 M (18%) glucose (final 0.2%) to each flask.

5% glycerol (532 ml)
500 ml dH_2O
32 ml 80% glycerol
75 µl 2N NaOH
Autoclave. Cool to ice cold before using

Library cloning and PepSeq
Griffitts Lab Homemade Midiprep

Pellet 40 ml of overnight (1 ml overnight culture added to 40 ml LB+Amp, grown ~12 hours) culture at 8,000 rpm in SS-34 rotor (round-bottom tube) for 15 minutes. (2 per vector for 4x 40 ml cultures). Freeze at -20°C for at least 1 hour.
Suspend each pellet in 3 ml of $T_{50}E_{10}$. Add 3.5 ml of lysis buffer allow to lyse for 2 minutes. Add 4 ml of 3M KOAc pH 6. Add 18 µl of 20 mg/ml RNase. Invert to mix. Incubate in 30°C water bath for 15 minutes. Pellet the cells for 12 minutes at 13,000 rpm. Move ~9 ml of supernatant to new round bottom tube. Combine 2 tubes of each vector, leaving 2 total tubes.
Add 1 volume of 100% isopropanol to each vector. Place the tubes on ice for 15 minutes. Pellet the cells at 13,000 rpm for 5 minutes. Dump and tap, then centrifuge again at 13,000 rpm for 1 minute. Pipet off as much of the isopropanol as possible. Place the tubes upside down and let them air dry for 20 minutes.
Dissolve pellet in 200 µl of warm $T_3E_{.3}$. Resuspending should take about 3-5 minutes.

Add 5 µl of RNase, place in 30°C water bath for 15 minutes. Add 1 ml of endowash (5x the volume of TE that was added previously) to the tube. Run endowash down the side of the tube several times. Aliquot 600 µl of the solution into two separate Zymo columns.
Spin for 1 minute at max speed. Dump and tap. Add 400 µl endowash. Spin for 1 minute at max speed. Dump and tap. Add 800 µl of plasmid wash, spin for 1 minute at max speed. Dump and tap. Elute in 65 µl of warm $T_3E_{.3.}$

Lysis buffer
1 ml 10% SDS
1 ml 2N NaOH
8 ml ddH_2O
3M KOAc
45 ml H2O
22.1 g KOAc
pH adjust to 6.0 with acetic acid
Adjust volume to 75 ml

Library insert synthesis

per 40 µl rxn
24 µl H_2O
8 µl Q5 buffer
2 µl 10 mM dNTP mix
1 µl Q5 enzyme
2.5 µl 100 µM forward primer (oED0110)
2.5 µl 100 µM reverse primer (oED0111)

96° 1:00
50° 20:00
55° 5:00
60° 5:00
65° 5:00
70° 5:00
4° ∞

Zymo cleanup (200 µl endowash, 400 µl plasmid wash). Elute in 50 µl $T_3E_{0.3.}$

Vector and insert digests

Vector
per 80 µl reaction
28 µl H_2O
8 µl Cut Smart buffer
40 µl DNA
2 µl BamHI-HF
2 µl SacI-HF

37°C for 12 hours. Add 1.5 µl CIP for last hour. LMP gel cleanup. (check 2 µl on analytical gel, load rest with 15 µl loading dye on low melt gel for ~40 minutes). Cut out gel slices, heat at 65°C to melt. Add 500 µl warm endowash (45°C water bath). Run through Zymo column. Add 400 µl plasmidwash. Elute in 60 µl warm $T_3E_{0.3}$. This was done to get rid of vector only background. Insert was only cleaned with Zymo column.

Insert
per 80 µl reaction
48 µl H_2O
8 µl Cut Smart buffer
20 µl DNA
2 µl BamHI-HF
2 µl SacI-HF
Incubate at 37°C for 12 hrs. Zymo cleanup, elute in 80 µl $T_3E_{0.3}$.

Based on small scale ligations (10 µl), the vector only had approximately 1 colony per 100 colonies of vector plus insert. And the optimal ratio of vector to insert is 1:1 (no dilution of insert).

Ligation

per 80 µl reaction
8 µl H_2O
8 µl T4 ligase buffer
30 µl vector digest
30 µl insert digest
4 µl T4 ligase

PCR program
1. 16°C, 2h |
2. 20°C, 2h | x2
3. 24°C, 2h |
4. 65°C, 5 min
5. 4°C, ∞

Zymo cleanup (400 µl endowash, 800 µl plasmidwash). Elute in 80 µl $T_3E_{0.3}$.

Transformation

To 50 µl of D050 electrocompetent cells add 4 µl cleaned-up ligation, move 52 µl to cold cuvette, shock on EC1 setting, immediately recover in 1 ml of LB, shake in disposable tube for 60 minutes, plate. 10 transformations were done, combining all 10 ml into 1 50 ml Falcon tube. An additional 10 ml of LB was added to the tube for a final volume of 20 ml. The Falcon tubes were shaken at 37°C for 1 hour. 3.75 ml 80% glycerol was added (for final concentration of 15% glycerol) to the tube. 2 0.2 ml aliquots were removed for dilution plating. The tube was then frozen at -80°C.

The msfGFP vector (pED21) dilution plating resulted in 140 colonies in 20 µl of a 1:10 dilution, for a total of 1,400,000 colonies (or 70,000 CFUs/ml).
So for 200 colonies, plate 3 µl (in 50 µl LB). For 15,000 colonies, 0.214 ml were plated on each of 20 large plate LB+Amp plates.

After 24 hours, each of 20 plates was scraped with 6 ml of LB+15% glycerol (~300,000 colonies in the library). The library initially went into a single 125 ml flask, and the flask was shaken for 30 minutes at 30°C to disrupt chunks. Library was frozen down in 500 µl aliquots.

Library selection

Library complexity is 280,000. Library frozen titer is 5.3×10^{10} CFUs/ml. For 1000x coverage, need 5 µl of library (plus 1:80 of LB). 5 µl in 40 ml M9M results in ~9 generations. For 6 generations, increase the inoculum to 50 µl, which is 10,000x coverage, plus 450 µl LB. After 12 hours of growth (37°C, 225 rpm), cultures underwent approximately 5-6 generations (+IPTG (.1mM) are on the lower end).

40 ml M9M (x8 total):
To 40 ml ddH_2O (40 ml H_2O autoclaved in 125 ml flasks) add:
2 ml 20x M9 salts
400 µl 50% glucose
80 µl 1M MgSO4
80 µl M9 Trace
80 µl Amp
20 µl IPTG (half of flasks)

Grow up for ~12 hours. Pellet 3 ml. Freeze at -80°C for at least an hour. Miniprep according to Zymo kit instructions. Elute in 60 µl $T_3E_{0.3}$.

PCR-based preparation of Illumina libraries

For technical reasons, the Illumina reads read the (-) strand.
Each amplification product will be 388 bp (+/- due to stagger)

PCR:
per 50 µl reaction
28.6 µl H_2O
10 µl 5X Q5 buf
1.5 µl 10 mM dNTP
0.7 µl Q5 pol
4 µl 10 µM forward primer (equal volumes of oED112, oED113, and oED114)
4 µl 10 µM reverse primer (oED115, oED116, oED117, or oED118)
1.2 µl DNA (undiluted)

Program:
96° 1 min

96° 20 s |
60° 20 s | x18
72° 20 s |
72° 2 min
4° ∞

To 50 µl reaction, add 9 µl of 8x dye. Check 3 µl on a gel, then do a gel extraction.

Qiagen gel extraction:
Run out 46 µl of sample on 1% LMP agarose gel
Slice out the 390 bp products (band between primers and hybrids)
Melt at 65°C for 5-10 min
Allow sample to return to nearly room temp
Add 500 µl of QG and mix well
Incubate at 37°C, 20-30 minutes with occasional vortexing
Add 150 µl of isopropanol, mix well, and pulse in microfuge
Apply sample to spin column, spin through, 1 minute
Wash with 0.4 ml QG
Wash with 0.7 ml PE
Spin again 1 min
Elute in 52 µl of 3 mM Tris 8.0 (no EDTA)
Run 3 µl on a gel to check with 3 µl of PCR marker

Deliver sample to Ed, with instructions to quantify by qPCR and then load onto 50-bp SE rapid run flow cell (2 lanes) with 5-8% PhiX.
Resulting Illumina reads were processed using Illumina_mapper_jan_5.py

LB and minimal media (M9M)

20X M9M:
1M K_2HPO_4
1M MOPS
0.4M NH_4Cl
pH adjust to 7.8 with NaOH pellets

for 40 ml 20X M9M:
~27 ml H_2O
6.97 g K_2HPO_4
8.37 g MOPS
0.86 g NH_4Cl
pH adjust to 7.8 with ~9 pellets of NaOH
Autoclave

To make 1X M9M:
20X dilution of M9M stock
100X dilution of 50% glycerol (or 50% glucose)

500X dilution of 1M $MgSO_4$
500X dilution of M9 Trace

M9 Trace:
50 ml water
230 mg EDTA
100 mg $FeCl_3$
50 mg $CaCl_2$
50 mg $ZnSO_4$
50 mg $MnSO_4$
5 mg H_3BO_3
5 mg $CuSO_4$
5 mg Na_2MoO_4
5 mg $CoCl_2$
Filter sterilize

For 40 ml M9M:
To 40 ml ddH_2O (40 ml H_2O autoclaved in 125 ml flasks) add:
2 ml 20x M9 salts
400 µl 50% glucose
80 µl 1M MgSO4
80 µl M9 Trace
80 µl Amp (or other antibiotic)

LB
1 L H_2O
10 g Bacto tryptone
5 g Bacto yeast extract
5 g NaCl
1 ml 2N NaOH (for the pH of water in our building)
For plates, 3 g agar per 250 ml bottle
Autoclave

APPENDIX III

Top toxic peptides

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATAACTTTATCGGTTTCTTTCGC	YNFIGFFR	5	4	80	77	1	1	0.01
AGTCACTTTTTCTATGGCGTTTTC	SHFFYGVF	7	2	47	43	1	1	0.02
CGTTTCCTTTACTTTGGCCATCTC	RFLYFGHL	6	6	49	38	1	2	0.03
GGTTTCTATTACGTTTGCATTATC	GFYYVCII	11	0	50	57	4	1	0.05
TATTACCGTATCATTTGCGTTTTC	YYRIICVF	9	4	40	22	1	2	0.05
TGTCTCTATATCGTTCGCTTTCTC	CLYIVRFL	10	4	64	69	1	6	0.05
CATTACTATATCATTTACTTTATC	HYYIIYFI	10	2	88	77	5	4	0.05
CGTTACTATTACATTATCTTTGTC	RYYYIIFV	9	4	97	81	1	9	0.06
TTTTGCCTTTTCCGTGGCTTTGGC	FCLFRGFG	7	4	123	125	7	7	0.06
TATAACATTTTCTATATCCTTTAC	YNIFYILY	9	0	53	48	2	4	0.06
TTTCGCATTTTCGTTTTCCTTATC	FRIFVFLI	12	4	27	23	2	1	0.06
TTTTTCCGTATCGTTCTCGTTGTC	FFRIVLVV	12	4	42	24	3	1	0.06
TATCACCTTGTCTATCGCATTTAC	YHLVYRIY	6	6	40	42	2	3	0.06
CATCGCGTTGTCTATGTCTTTATC	HRVVYVFI	8	6	37	44	3	2	0.06
CGTCACTTTTTCTATGGCGTTTTC	RHFFYGVF	6	6	533	479	25	39	0.06
CGTAACTTTTTCTATGGCGTTTTC	RNFFYGVF	5	4	31	30	4	0	0.07
CGTTTCCGTATTCTTATCCTTCTC	RFRILILL	8	8	24	34	2	2	0.07
GATTTCGATGGCGTTGGCATTAGC	DFDGVGIS	1	-8	165	153	13	9	0.07
TGTTTCTATGTCATTCACATTATC	CFYVIHII	11	2	37	33	4	1	0.07
TTTCGCATTTTCTATCGCATTTTC	FRIFYRIF	7	8	80	88	5	7	0.07
TATTACTATCTCGGTTTCAATTTC	YYYLGFNF	7	0	45	36	4	2	0.07
CGTTACTATATCATTTGCCTTGTC	RYYIICLV	9	4	41	26	3	2	0.07
TTTCGCTGTATCATTGTCCGTCTC	FRCIIVRL	7	8	79	80	8	4	0.08
TATCGCCTTATCATTTTCTTTTAC	YRLIIFFY	10	4	43	46	5	2	0.08
GATTACGTTATCATTGTCTATCGC	DYVIIVYR	6	0	159	157	7	18	0.08
GTTCGCTTTTACGTTTACCTTATC	VRFYVYLI	10	4	46	29	3	3	0.08
CATCGCTTTTTCTATGTCCTTTTC	HRFFYVLF	8	6	81	68	6	6	0.08
TATTACTATGTCTTTTACCTTTGC	YYYVFYLC	11	0	26	35	3	2	0.08
TGTTACCGTCTCTATATCATTTTC	CYRLYIIF	9	4	230	221	24	13	0.08
CTTCGCGTTTACCTTGTCTTTGTC	LRVYLVFV	11	4	55	30	3	4	0.08
AATTACCTTTTCTATTGCTTTCTC	NYLFYCFL	9	0	110	96	9	8	0.08
ATTTACTTTATCTATCTCCGTCGC	IYFIYLRR	6	8	37	35	3	3	0.08
GATTACTATATCGTTTTCCTTGTC	DYYIVFLV	10	-4	49	46	5	3	0.08
TATTACCGTATCTATTGCTATTTC	YYRIYCYF	7	4	36	35	2	4	0.08
CGTTTCATTATCATTCGCCTTCAC	RFIIIRLH	5	10	31	28	5	0	0.08
CGTATCTATTACATTATCATTTGC	RIYYIIIC	9	4	36	45	2	5	0.09
TATAGCTGTTACCTTATCCTTTTC	YSCYLILF	10	0	40	49	5	3	0.09
ATTCGCCATGTCTATATCATTATC	IRHVYIII	8	6	48	52	4	5	0.09
ATTGTCATTATCCGTTGCGTTCGC	IVIIRCVR	7	8	55	33	5	3	0.09
AGTCTCTATATCTTTTTCATTTGC	SLYIFFIC	11	0	34	31	2	4	0.09
CGTCGCATTATCATTGTCTTTCTC	RRIIIVFL	8	8	64	76	9	4	0.09
ATTTACGTTTTCAGTTTCGGTGTC	IYVFSFGV	10	0	85	76	8	7	0.09
TATCGCGTTATCGTTTTCTTTTAC	YRVIVFFY	10	4	23	30	1	4	0.09
TATCGCCTTTTCCTTGTCTATTTC	YRLFLVYF	10	4	225	229	23	20	0.09
CGTAGCGTTTTCTTTGTCTATATC	RSVFFVYI	8	4	23	40	4	2	0.10
TTTTTCCATAGCGTTTTCTATTGC	FFHSVFYC	8	2	97	101	8	11	0.10
TATCGCGTTTACGTTTACTTTTGC	YRVYVYFC	8	4	47	56	7	3	0.10
TTTCGCATTTTCTTTTACATTGTC	FRIFFYIV	11	4	61	52	3	8	0.10

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATCGCGTTATCTATTTCTGTCGC	YRVIYFCR	5	8	74	80	6	9	0.10
AGTTTCCGTTTCTATGTCATTTAC	SFRFYVIY	7	4	84	89	9	8	0.10
TATATCTATTACCTTATCATTGGC	YIYYLIIG	11	0	35	36	4	3	0.10
CGTATCTATTACGTTTACTTTGTC	RIYYVYFV	9	4	98	64	9	7	0.10
CTTCTCTATTGCTTTATCTATCGC	LLYCFIYR	9	4	31	29	2	4	0.10
GGTTACCGTATCATTTTCATTTAC	GYRIIFIY	8	4	85	72	7	9	0.10
GATTACGTTATCCTTGTCTTTATC	DYVILVFI	11	-4	109	97	9	12	0.10
TATCGCGGTTTCTTTAGCTTTTTC	YRGFFSFF	6	4	158	124	14	15	0.10
TATTTCATTAGCTATTACCTTCGC	YFISYYLR	6	4	55	61	6	6	0.10
TATCGCTGTATCTTTATCTATGTC	YRCIFIYV	9	4	89	84	9	9	0.10
AGTTTCTGTCTCTATGTCTTTATC	SFCLYVFI	11	0	85	77	9	8	0.10
AATTTCTATATCGTTTACTATTTC	NFYIVYYF	9	0	45	50	5	5	0.11
TATCGCATTTTCTATTGCTTTCTC	YRIFYCFL	9	4	157	147	23	9	0.11
TTTCGCATTTACTTTCTCCTTTGC	FRIYFLLC	10	4	61	52	3	9	0.11
AGTTACGATCTCTGTGGCAATGAC	SYDLCGND	-3	-8	141	113	12	15	0.11
GGTCGCTTTTTCTATCTCATTGTC	GRFFYLIV	9	4	42	42	3	6	0.11
GTTTTCTATATCATTCGCTTTTAC	VFYIIRFY	10	4	41	42	6	3	0.11
TTTCGCTATTACATTATCTTTCGC	FRYYIIFR	6	8	101	83	10	10	0.11
CGTCGCTGTATCTTTCTCATTATC	RRCIFLII	7	8	119	118	14	12	0.11
TATAACGTTATCTATATCGTTATC	YNVIYIVI	10	0	50	32	5	4	0.11
TGTCGCATTTACTTTTTCATTCGC	CRIYFFIR	6	8	142	140	20	11	0.11
TATCACTTTATCAGTTTCTGTTGC	YHFISFCC	7	2	180	165	17	21	0.11
ATTTACCGTTTCTATATCGTTATC	IYRFYIVI	10	4	85	96	8	12	0.11
GTTCGCTATTACATTTTCGTTGTC	VRYYIFVV	10	4	87	75	7	11	0.11
TTTTGCCGTATCTATTACATTTTC	FCRIYYIF	9	4	28	34	3	4	0.11
CGTGTCATTTACATTAACTTTATC	RVIYINFI	7	4	63	61	7	7	0.11
AGTTACCGTATCCTTCGCCTTATC	SYRILRLI	4	8	57	58	8	5	0.11
TGTTACATTATCTTTTACGTTATC	CYIIFYVI	13	0	22	31	4	2	0.11
CATTACCGTATCTTTGTCTATATC	HYRIFVYI	7	6	51	37	5	5	0.11
CTTATCCGTTTCTATTACTTTATC	LIRFYYFI	10	4	84	82	8	11	0.11
GATCTCGTTTACCTTTACTTTTTC	DLVYLYFF	10	-4	192	199	22	23	0.12
CGTTTCGTTATCTATCTCATTTGC	RFVIYLIC	10	4	102	89	10	12	0.12
TTTCGCTTTTTCTTTTACGTTCTC	FRFFFYVL	11	4	28	24	1	5	0.12
CGTTACTTTGTCATTCACTATTTC	RYFVIHYF	7	6	69	43	7	6	0.12
GGTCTCCGTATCATTCTCTTTTGC	GLRIILFC	9	4	93	105	12	11	0.12
TATGACGATTTCAGTTTCTATCAC	YDDFSFYH	0	-6	61	59	5	9	0.12
TATTTCCTTATCATTCGCCGTCAC	YFLIIRRH	4	10	57	54	5	8	0.12
ATTTTCCGTATCATTTACCTTGTC	IFRIIYLV	11	4	58	53	6	7	0.12
GTTCTCATTTACAATTTCCGTGGC	VLIYNFRG	5	4	59	52	7	6	0.12
TTTCGCATTTACTTTTTCGTTATC	FRIYFFVI	11	4	189	160	19	22	0.12
TTTTTCCGTCTCTATTTCTTTGGC	FFRLYFFG	9	4	112	108	8	18	0.12
TGTTACGGTATCTATATCTTTCTC	CYGIYIFL	11	0	41	35	5	4	0.12
GGTCGCGTTTTCGTTGTCGTTTTC	GRVFVVVF	10	4	101	93	14	9	0.12
ATTTACCTTATCCGTATCGTTTGC	IYLIRIVC	10	4	147	139	15	19	0.12
TATTACATTTTCTGTATCTATCGC	YYIFCIYR	8	4	91	84	11	10	0.12
TATCACTTTATCCATTACGGTTAC	YHFIHYGY	5	4	95	88	8	14	0.12
TATATCGTTATCCTTGTCCGTCGC	YIVILVRR	7	8	141	113	13	18	0.12
CGTTTCATTGTCCTTCTCTATGTC	RFIVLLYV	11	4	171	197	15	30	0.12
TATCGCGTTTTCATTTTCGTTAGC	YRVFIFVS	8	4	91	88	10	12	0.12
AGTTACCGTGTCATTATCTTTTAC	SYRVIIFY	7	4	102	85	11	12	0.12
TTTTACCGTGTCCTTATCGTTCTC	FYRVLIVL	11	4	112	107	17	10	0.12
TATCGCGTTATCTGTATCTTTATC	YRVICIFI	10	4	44	45	5	6	0.12
CGTATCGTTTTCGTTTACTATGTC	RIVFVYYV	10	4	66	55	9	6	0.12
AGTATCGTTTACATTATCTATGGC	SIVYIIYG	9	0	165	157	21	19	0.12

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
GTTTACTATATCCGTCTCTTTAAC	VYYIRLFN	6	4	39	49	8	3	0.13
TATGTCGTTATCCTTCTCCGTCGC	YVVILLRR	7	8	175	161	20	22	0.13
GTTCGCTATTACTTTCTCTTTGTC	VRYYFLFV	10	4	70	73	9	9	0.13
GATTACTGTATCCTTATCTATCTC	DYCILIYL	9	-4	57	61	8	7	0.13
TATGGCTTTGTCATTGTCTATTTC	YGFVIVYF	12	0	88	77	10	11	0.13
CGTTACATTATCCTTCTCCGTGGC	RYIILLRG	5	8	54	40	6	6	0.13
ATTATCATTGTCCGTCGCTTTAAC	IIIVRRFN	4	8	128	99	14	15	0.13
CGTGTCGTTTACGTTTACATTATC	RVVYVYII	10	4	65	60	9	7	0.13
AATTTCTATATCTATTTCGTTATC	NFYIYFVI	10	0	53	56	7	7	0.13
TATTACGTTCTCTTTTACATTATC	YYVLFYII	13	0	54	47	5	8	0.13
AATTACGTTGTCGTTTACTTTCGC	NYVVVYFR	6	4	29	33	6	2	0.13
TGTGTCTATCGCATTATCTTTCTC	CVYRIIFL	10	4	67	57	8	8	0.13
CGTATCTATGTCATTTACGGTTTC	RIYVIYGF	8	4	120	120	11	20	0.13
TATAGCTATGTCTTTTTCTATCTC	YSYVFFYL	10	0	57	59	10	5	0.13
CTTTACCTTTGCTTTTTCCGTCGC	LYLCFFRR	6	8	146	116	18	16	0.13
ATTCGCATTTTCTATGTCCTTGTC	IRIFYVLV	11	4	147	122	16	19	0.13
TTTCGCTGTCTCATTTTCTTTGTC	FRCLIFFV	11	4	99	85	13	11	0.13
GTTCGCTATTACGTTATCATTCGC	VRYYVIIR	6	8	152	169	20	22	0.13
TATTACATTTACTTTATCCGTGTC	YYIYFIRV	9	4	98	70	8	14	0.13
TATCGCCGTTACTTTATCATTTTC	YRRYFIIF	6	8	101	67	14	8	0.13
AGTTACGTTCTCTGTATCTTTATC	SYVLCIFI	11	0	73	72	8	11	0.13
CATCGCGTTTACTTTCTCTTTGTC	HRVYFLFV	8	6	188	163	22	24	0.13
CTTTACATTTACCATTTCATTCGC	LYIYHFIR	7	6	58	41	8	5	0.13
TGTTACTATCTCTATATCTTTGTC	CYYLYIFV	12	0	50	64	8	7	0.13
CATCGCGTTTTCATTGTCGTTTTC	HRVFIVVF	9	6	53	38	2	10	0.13
CATTGCTATTACTTTGTCCTTATC	HCYYFVLI	10	2	51	39	6	6	0.13
TATGGCTTTGACGTTTACTATTTC	YGFDVYYF	7	-4	193	174	20	29	0.13
TTTTACATTTACTATTACCTTCTC	FYIYYYLL	12	0	70	71	8	11	0.13
GGTCACTTTTTCGTTATCCTTTAC	GHFFVILY	10	2	67	51	8	8	0.14
CGTTACTTTGTCATTCTCTTTTGC	RYFVILFC	10	4	71	76	10	10	0.14
GATTTCGTTTGCATTATCTTTCTC	DFVCIIFL	11	-4	119	130	14	20	0.14
TTTATCCGTTACATTTACTTTGTC	FIRYIYFV	10	4	68	56	11	6	0.14
GTTTACTATTGCCTTTTCGTTTTC	VYYCLFVF	13	0	135	113	17	17	0.14
TGTTACCGTTGCTTTATCATTCTC	CYRCFIIL	9	4	25	26	3	4	0.14
CGTAGCTATATCGTTTTCCTTGTC	RSYIVFLV	8	4	101	81	12	13	0.14
GTTTACCTTCTCTATTTCATTTGC	VYLLYFIC	13	0	51	36	3	9	0.14
CGTTACGTTTACATTTGCTTTATC	RYVYICFI	9	4	43	58	8	6	0.14
TATCACTATATCATTCTCTTTTAC	YHYIILFY	10	2	40	46	5	7	0.14
CGTTACTTTCTCATTTACATTATC	RYFLIYII	10	4	91	88	10	15	0.14
TTTTACATTCTCTATATCGTTATC	FYILYIVI	14	0	43	50	6	7	0.14
TATTACATTTGCATTATCTATATC	YYICIIYI	12	0	144	142	19	21	0.14
TATCACATTCTCTATCTCTTTTGC	YHILYLFC	10	2	228	208	37	24	0.14
AGTATCCGTTTCATTCTCTTTATC	SIRFILFI	9	4	20	30	3	4	0.14
CGTCTCTATATCTTTTTCATTTGC	RLYIFFIC	10	4	403	360	48	59	0.14
AGTTACGTTATCTTTCTCGTTCAC	SYVIFLVH	9	2	90	95	12	14	0.14
TATATCGTTCTCGTTCGCTTTTAC	YIVLVRFY	10	4	34	37	4	6	0.14
TTTCGCATTTACTGTTACGTTATC	FRIYCYVI	9	4	78	64	9	11	0.14
GTTTACCATGTCTTTCGCTGTATC	VYHVFRCI	7	6	42	36	4	7	0.14
TATTTCTATGTCATTCGCTGTTAC	YFYVIRCY	8	4	50	42	7	6	0.14
CGTATCTTTGTCCTTGTCTATTGC	RIFVLVYC	10	4	57	42	7	7	0.14
TATCGCATTTTCCTTGTCTGTGGC	YRIFLVCG	8	4	219	195	30	29	0.14
TTTTACCTTTTCATTTGCGGTGGC	FYLFICGG	10	0	37	33	3	7	0.14
CATTTCTATGTCATTTACTATCTC	HFYVIYYL	10	2	43	34	3	8	0.14
TATAGCATTTTCTATCTCATTGTC	YSIFYLIV	11	0	53	52	7	8	0.14

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
ATTCGCTATATCGTTGTCGTTAGC	IRYIVVVS	8	4	107	110	15	16	0.14
TATGTCGTTATCCATTTCCTTGAC	YVVIHFLD	8	-2	121	103	14	18	0.14
CGTTACTATGTCATTCACCTTCTC	RYYVIHLL	7	6	144	121	19	19	0.14
GGTTTCCTTTACCTTCTCTTTATC	GFLYLLFI	13	0	137	121	17	20	0.14
CGTATCGTTTACTTTTGCCGTCTC	RIVYFCRL	6	8	110	127	14	20	0.14
TATCGCCTTTGCTTTCTCCTTATC	YRLCFLLI	10	4	235	218	37	28	0.14
CTTTACCTTCGCTTTCTCTTTCTC	LYLRFLFL	11	4	57	75	10	9	0.14
CGTTACTGTCTCCTTGTCATTCGC	RYCLLVIR	6	8	63	62	10	8	0.14
CTTTTCTTTAACTATAACCTTTTC	LFFNYNLF	7	0	42	41	5	7	0.14
AATTACATTTTCTATTACTTTGGC	NYIFYYFG	7	0	40	36	5	6	0.14
TTTCTCGTTTACATTATCTATGGC	FLVYIIYG	12	0	41	35	6	5	0.14
ATTTACTATTTCCGTGTCATTCTC	IYYFRVIL	10	4	128	100	20	13	0.14
GGTTTCTTTGTCATTTACATTGTC	GFFVIYIV	13	0	127	121	15	21	0.15
GGTTACTTTATCTGTATCGTTTAC	GYFICIVY	11	0	160	163	19	28	0.15
CTTTGCTATCGCTATTTCTATGGC	LCYRYFYG	6	4	115	98	15	16	0.15
ATTTACTATGTCGTTTACTTTCGC	IYYVVYFR	9	4	88	70	13	10	0.15
ATTTACATTTTCCGTTACTTTCAC	IYIFRYFH	7	6	55	48	10	5	0.15
ATTCGCTTTTTCTATATCGTTGTC	IRFFYIVV	11	4	61	69	6	13	0.15
CGTTTCGTTTTCTATTACTTTGTC	RFVFYYFV	10	4	57	59	3	14	0.15
TATCGCATTGTCATTTGCTATCTC	YRIVICYL	9	4	122	89	12	19	0.15
CTTCGCATTCTCGTTTGCATTATC	LRILVCII	11	4	112	119	18	16	0.15
TATTACATTCTCCATGTCATTCTC	YYILHVIL	11	2	112	85	16	13	0.15
CATCGCTGTTACGTTCTCATTATC	HRCYVLII	7	6	91	72	9	15	0.15
TATTTCGTTGTCTATATCTTTGTC	YFVVYIFV	14	0	82	74	13	10	0.15
GGTTACTATGTCCTTTACGTTTTC	GYYVLYVF	11	0	25	36	5	4	0.15
TATATCGTTTACATTCTCAGTCGC	YIVYILSR	7	4	35	32	1	9	0.15
ATTTACCGTTACCTTGGCATTTTC	IYRYLGIF	8	4	39	28	6	4	0.15
CGTATCTTTATCTATATCTGTGGC	RIFIYICG	8	4	175	173	29	23	0.15
GGTCTCATTCACGTTTTCTTTGTC	GLIHVFFV	11	2	101	93	8	21	0.15
CGTAACTATATCATTTTCATTCTC	RNYIIFIL	7	4	253	188	31	35	0.15
CTTATCGTTTACCTTATCGGTGTC	LIVYLIGV	13	0	105	102	16	15	0.15
TGTATCCGTATCATTCTCCTTAGC	CIRIILLS	8	4	59	61	11	7	0.15
CGTTTCCGTCTCATTTACCTTTTC	RFRLIYLF	7	8	163	143	24	22	0.15
CGTAGCATTTACATTTACTATTTC	RSIYIYYF	6	4	50	63	11	6	0.15
TATCGCATTATCTTTTTCTTTTTC	YRIIFFFF	11	4	54	59	6	11	0.15
TATCTCGTTGTCGTTTTCCTTAGC	YLVVVFLS	12	0	71	42	4	13	0.15
TTTGTCCGTCTCTTTTTCATTTTC	FVRLFFIF	12	4	148	124	23	18	0.15
TTTAACATTTACCTTCTCCTTATC	FNIYLLLI	11	0	225	246	31	40	0.15
TTTTACATTATCTATTACTATTTC	FYIIYYYF	12	0	146	152	18	27	0.15
GGTTACTTTTTCCTTATCTTTGTC	GYFFLIFV	13	0	81	71	8	15	0.15
GGTTACGTTCACTTTGTCATTATC	GYVHFVII	10	2	140	144	16	27	0.15
TATTACCTTTGCATTCGCCTTCGC	YYLCIRLR	5	8	42	24	2	8	0.15
GGTATCGGTTGCGATGGCGATGAC	GIGCDGDD	-3	-12	473	371	60	68	0.15
TTTGTCGTTTACGTTATCTATCGC	FVVYVIYR	10	4	90	88	13	14	0.15
CGTCTCATTTACCTTCGCCTTTAC	RLIYLRLY	6	8	115	89	17	14	0.15
GTTTGCGTTTACGTTCTCTATGTC	VCVYVLYV	13	0	94	90	14	14	0.15
CTTTACATTATCCGTTTCCTTCTC	LYIIRFLL	11	4	110	74	10	18	0.15
GTTAGCGTTTACTATCTCTATGTC	VSVYYLYV	10	0	125	105	16	19	0.15
GGTTTCCGTATCTATATCGTTTGC	GFRIYIVC	8	4	87	77	13	12	0.15
GGTGACATTATCTTTGTCATTGTC	GDIIFVIV	10	-4	41	57	7	8	0.15
TGTATCTTTAGCATTTACTATTAC	CIFSIYYY	9	0	111	98	17	15	0.15
CATTTCGTTATCGTTTACCTTGTC	HFVIVYLV	12	2	226	198	28	37	0.15
GGTAACGTTTACGTTATCCTTTTC	GNVYVILF	9	0	119	96	19	14	0.15
GATTTCGTTGTCTATATCATTGTC	DFVVYIIV	11	-4	234	248	38	36	0.15

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
GATAGCTATATCATTATCTATTTC	DSYIIIYF	7	-4	151	129	22	21	0.15
CTTTGCCTTATCCGTTTCCTTCAC	LCLIRFLH	8	6	49	29	5	7	0.15
TATGTCGTTATCGTTGTCCGTGGC	YVVIVVRG	9	4	75	68	16	6	0.15
TATTACATTCTCCTTGTCCGTGGC	YYILLVRG	8	4	114	81	10	20	0.15
GATTTCTATTACATTATCTTTTAC	DFYYIIFY	9	-4	52	58	9	8	0.15
GGTCTCTATATCCTTTTCGTTCGC	GLYILFVR	9	4	52	45	10	5	0.15
GGTTTCCGTATCGTTTTCCTTTAC	GFRIVFLY	9	4	156	122	27	16	0.15
ATTTACTATATCGTTTACTGTATC	IYYIVYCI	12	0	61	68	9	11	0.16
GTTCACATTCTCTATTTCAATTAC	VHILYFNY	7	2	27	31	6	3	0.16
TATTACTTTGTCATTATCCGTGGC	YYFVIIRG	8	4	29	29	5	4	0.16
GGTCGCTTTTACCTTTGCTTTATC	GRFYLCFI	8	4	34	24	6	3	0.16
AATATCATTATCTATTGCTTTATC	NIIIYCFI	10	0	58	58	11	7	0.16
AGTTTCCTTTACGTTATCCTTTAC	SFLYVILY	11	0	95	79	17	10	0.16
TATGTCTGTTTCCGTCTCTATGGC	YVCFRLYG	7	4	98	76	12	15	0.16
TATTACGTTATCGTTTACCGTGGC	YYVIVYRG	7	4	161	148	19	29	0.16
TGTCGCCTTTGCCATATCTATCGC	CRLCHIYR	2	10	560	475	73	88	0.16
ATTGTCAGTTACATTATCATTCGC	IVSYIIIR	8	4	57	65	7	12	0.16
TATCGCATTGTCTTTTTCTGTCGC	YRIVFFCR	6	8	132	112	21	17	0.16
CGTGTCTATATCTTTTACCTTGAC	RVYIFYLD	6	0	57	39	7	8	0.16
CGTCACTTTATCTTTGTCATTGGC	RHFIFVIG	7	6	108	109	24	10	0.16
TATTACGTTCTCTATTACCTTTGC	YYVLYYLC	11	0	67	67	9	12	0.16
ATTCGCTATTTCTATTACTATGGC	IRYFYYYG	6	4	136	132	21	21	0.16
TTTTACGGTGTCTTTCGCTTTATC	FYGVFRFI	9	4	69	52	11	8	0.16
TTTCTCCGTTACGTTCTCTTTTTC	FLRYVLFF	11	4	58	50	9	8	0.16
CGTATCGTTCGCGTTTTCATTGTC	RIVRVFIV	8	8	57	57	6	12	0.16
CTTTACATTATCCGTCTCATTAGC	LYIIRLIS	8	4	59	55	7	11	0.16
TTTCGCATTTACGTTATCTATTTC	FRIYVIYF	10	4	166	144	27	22	0.16
TTTGTCCGTGTCTATGTCATTCTC	FVRVYVIL	11	4	288	249	49	36	0.16
TATAACATTTACGTTGTCGTTATC	YNIYVVVI	10	0	35	28	5	5	0.16
CTTTACCATCTCATTCGCGGTCAC	LYHLIRGH	3	8	151	132	25	20	0.16
TATTACATTTGCTATGTCTTTTTC	YYICYVFF	12	0	49	39	5	9	0.16
TATTACATTGTCTATGTCGTTCGC	YYIVYVVR	9	4	74	83	13	12	0.16
CGTCGCTTTATCTTTGTCATTATC	RRFIFVII	8	8	60	53	9	9	0.16
TATAACTTTTTCTGTTACATTCGC	YNFFCYIR	5	4	38	31	6	5	0.16
CTTTACATTGTCTATTTCATTGTC	LYIVYFIV	14	0	69	69	12	10	0.16
TATTACTGTCACTATTTCTATGTC	YYCHYFYV	8	2	64	55	9	10	0.16
TATAACTATTACCTTGTCATTCGC	YNYYLVIR	5	4	20	30	5	3	0.16
TATTACTTTTACATTTACTTTATC	YYFYIYFI	12	0	57	43	12	4	0.16
TATGTCGATCTCTGTGACGGTGAC	YVDLCDGD	0	-12	217	214	33	36	0.16
TTTTGCTATCGCATTTACCTTCGC	FCYRIYLR	5	8	120	117	15	23	0.16
CTTCGCATTCGCCTTATCGTTATC	LRIRLIVI	8	8	53	53	12	5	0.16
GTTTACATTGTCCGTTTCCATGGC	VYIVRFHG	6	6	57	49	6	11	0.16
TTTAGCATTTACTTTCTCCGTATC	FSIYFLRI	8	4	80	82	14	12	0.16
TTTAACATTCTCTTTTTCATTGTC	FNILFFIV	12	0	91	77	16	11	0.16
AATCGCATTTACATTCTCTATCTC	NRIYILYL	6	4	104	120	23	13	0.16
TATTACATTCTCCGTTTCATTTTC	YYILRFIF	10	4	42	45	4	10	0.16
GGTCGCCGTATCATTATCCTTATC	GRRIIILI	6	8	43	44	8	6	0.16
TATTACGTTATCTATCTCGGTGGC	YYVIYLGG	9	0	93	81	11	17	0.16
TATTACCTTAGCTTTTTCTTTGTC	YYLSFFFV	11	0	105	106	17	17	0.16
TTTTACATTCGCTATTTCCGTCTC	FYIRYFRL	6	8	113	116	18	19	0.16
TGTTTCCGTATCATTTTCGTTTAC	CFRIIFVY	10	4	169	159	26	27	0.16
GTTCGCCTTATCTTTGTCTATTGC	VRLIFVYC	10	4	82	66	13	11	0.16
TTTCGCTTTTTCGTTATCTTTCTC	FRFFVIFL	12	4	57	60	12	7	0.16
CGTATCTATATCCTTTACGTTGGC	RIYILYVG	8	4	69	60	9	12	0.16

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATCACCTTATCCTTGTCCGTTTC	YHLILVRF	8	6	84	69	12	13	0.16
CGTATCTTTCTCTTTTTCTTTTGC	RIFLFFFC	11	4	132	131	23	20	0.16
TATTACCTTATCATTGTCCGTTAC	YYLIIVRY	9	4	456	411	59	83	0.16
CGTGTCATTTTCATTCGCTTTAGC	RVIFIRFS	5	8	89	88	18	11	0.16
TATAGCGTTCGCATTATCATTATC	YSVRIIII	8	4	35	26	6	4	0.16
TATTGCGATATCTGTTACTTTTTC	YCDICYFF	8	-4	98	91	10	21	0.16
CGTGTCGTTCTCCTTTACCTTATC	RVVLLYLI	11	4	118	144	19	24	0.16
TTTCTCTATTGCTATATCTTTGTC	FLYCYIFV	13	0	69	65	10	12	0.16
ATTCGCTTTATCTATGTCATTTTC	IRFIYVIF	11	4	34	39	5	7	0.16
AATTACTTTATCTATCTCGTTTAC	NYFIYLVY	9	0	39	34	2	10	0.16
TATATCCATTACTATCTCTTTGTC	YIHYYLFV	10	2	80	72	11	14	0.16
AGTATCTGTTACGTTATCTTTATC	SICYVIFI	11	0	47	44	8	7	0.16
CATTACTTTGTCGTTGTCGTTTTC	HYFVVVVF	12	2	60	31	9	6	0.16
TATATCGGTGTCCTTATCTTTGTC	YIGVLIFV	13	0	102	92	15	17	0.16
CGTTACATTGTCGTTATCTATGTC	RYIVVIYV	10	4	109	97	17	17	0.17
TATGTCTATTACGTTTGCCTTCTC	YVYYVCLL	12	0	48	61	11	7	0.17
TTTTGCGTTATCGTTCGCGGTCGC	FCVIVRGR	5	8	58	51	7	11	0.17
CTTGTCCGTATCTATTTCTGTCGC	LVRIYFCR	6	8	60	67	11	10	0.17
AATCGCTGTTACATTATCATTGTC	NRCYIIIV	6	4	125	135	17	26	0.17
ATTGTCTATTTCGTTCGCTTTGTC	IVYFVRFV	11	4	97	96	21	11	0.17
GGTTTCCGTGTCTATTTCATTTAC	GFRVYFIY	8	4	101	104	17	17	0.17
AATTACGTTTTCCTTCTCCTTTAC	NYVFLLLY	10	0	33	33	5	6	0.17
TATTACGTTTACTTTATCTATAGC	YYVYFIYS	9	0	65	61	11	10	0.17
CGTATCCGTTGCATTATCTTTATC	RIRCIIFI	7	8	100	92	17	15	0.17
CGTCGCATTCTCATTTACATTCTC	RRILIYIL	7	8	138	168	27	24	0.17
TATCACCTTTTCCTTTTCTGTGTC	YHLFLFCV	11	2	90	83	13	16	0.17
AATCTCTTTCTCATTTACATTTTC	NLFLIYIF	11	0	80	87	11	17	0.17
CGTATCTATATCGTTTGCTATCTC	RIYIVCYL	9	4	99	68	12	16	0.17
GATGTCGTTTTCTATATCGTTATC	DVVFYIVI	11	-4	66	59	14	7	0.17
AATTACCGTTGCTATATCATTTTC	NYRCYIIF	5	4	58	61	6	14	0.17
TATTACATTGTCATTATCTATGTC	YYIVIIYV	13	0	56	51	9	9	0.17
CATATCATTCTCTATCTCCTTGAC	HIILYLLD	8	-2	47	30	7	6	0.17
TGTAGCTTTCGCATTGTCTTTCTC	CSFRIVFL	8	4	47	30	5	8	0.17
AGTTTCGTTGTCGTTGTCTATCGC	SFVVVVYR	8	4	71	77	17	8	0.17
TGTCGCCTTATCTATCTCGTTTAC	CRLIYLVY	9	4	103	116	17	20	0.17
GATTTCTGTATCGTTATCTATTAC	DFCIVIYY	9	-4	193	168	39	22	0.17
GGTTACGTTGTCTTTTTCGTTCGC	GYVVFFVR	9	4	79	63	16	8	0.17
TATTACATTATCGTTTGCTTTTAC	YYIIVCFY	12	0	67	69	14	9	0.17
ATTAACTATGTCATTTTCCTTTAC	INYVIFLY	10	0	31	34	6	5	0.17
TTTGACTATTACATTCTCTATCTC	FDYYILYL	9	-4	36	29	6	5	0.17
AATTACGTTTACGTTGTCGTTCTC	NYVYVVVL	10	0	29	30	1	9	0.17
CTTATCATTCTCCGTGTCGTTCAC	LIILRVVH	9	6	64	48	10	9	0.17
GTTCGCCTTTTCTGTTACGTTCTC	VRLFCYVL	10	4	32	21	4	5	0.17
TGTAACTATTACATTATCTTTGTC	CNYYIIFV	9	0	114	98	18	18	0.17
CGTGTCGTTGTCATTTTCCGTAGC	RVVVIFRS	5	8	138	109	27	15	0.17
ATTCGCTTTATCGTTCTCCGTATC	IRFIVLRI	8	8	47	47	6	10	0.17
ATTCGCTTTATCGTTCTCTATCTC	IRFIVLYL	11	4	188	192	30	35	0.17
TTTATCTATTACTATCTCGTTCTC	FIYYYLVL	13	0	102	85	10	22	0.17
TATATCGGTTACTATCGCCTTTTC	YIGYYRLF	7	4	230	208	37	38	0.17
CGTGTCCTTTTCATTCTCTATAGC	RVLFILYS	8	4	93	88	15	16	0.17
CGTATCATTATCGTTTTCAATCGC	RIIIVFNR	4	8	45	60	7	11	0.17
CGTATCGTTCTCTTTTACTTTGGC	RIVLFYFG	9	4	52	53	10	8	0.17
TTTCACATTCTCCTTCTCCTTTAC	FHILLLLY	12	2	113	126	20	21	0.17
TATTACCATTTCTTTCTCTTTGTC	YYHFFLFV	11	2	86	83	13	16	0.17

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TTTTGCCGTATCATTCTCTTTTAC	FCRIILFY	10	4	46	53	13	4	0.17
CGTGGCTATTTCCTTATCATTGTC	RGYFLIIV	9	4	47	52	8	9	0.17
GTTCGCGTTGTCTTTGTCATTTAC	VRVVFVIY	11	4	50	49	5	12	0.17
GTTATCTTTATCCGTCTCCATCAC	VIFIRLHH	6	8	238	191	36	38	0.17
GATTTCATTCTCATTTTCTTTTTC	DFILIFFF	12	-4	65	74	15	9	0.17
CTTTACCATCTCTATATCGTTTAC	LYHLYIVY	10	2	67	72	10	14	0.17
CGTTTCTATTACCTTTACGTTTTC	RFYYLYVF	9	4	51	59	11	8	0.17
AATTTCCGTCTCTTTTTCGTTATC	NFRLFFVI	8	4	122	98	18	20	0.17
TATGTCCTTTACTGTATCATTCAC	YVLYCIIH	10	2	24	28	3	6	0.17
TATCGCGTTTTCATTGTCATTGTC	YRVFIVIV	11	4	127	127	13	31	0.17
TATTACTTTCGCGTTATCTATATC	YYFRVIYI	9	4	158	165	29	27	0.17
CGTCGCTTTTACATTCTCCATTTC	RRFYILHF	4	10	198	159	28	34	0.17
TATTACTTTTACATTTACCTTTTC	YYFYIYLF	12	0	106	84	12	21	0.17
GTTTACGGTATCTATATCTTTTAC	VYGIYIFY	11	0	587	575	89	113	0.17
TTTGGCTTTCTCTGTTACATTCAC	FGFLCYIH	9	2	32	37	6	6	0.17
CGTATCGTTATCGTTCGCATTGGC	RIVIVRIG	6	8	90	71	15	13	0.17
AATGGCCGTATCATTATCTTTAGC	NGRIIIFS	3	4	55	31	7	8	0.17
TATTACATTCTCTGTATCCTTCTC	YYILCILL	13	0	92	80	17	13	0.17
TTTAACCGTTACATTATCATTTTC	FNRYIIIF	7	4	59	67	12	10	0.17
TTTAACCTTCGCATTATCGTTCTC	FNLRIIVL	8	4	329	278	54	52	0.17
TATGTCATTTACGTTTGCCATCGC	YVIYVCHR	6	6	93	73	15	14	0.17
TATCGCTTTTACTGTATCGTTCGC	YRFYCIVR	5	8	47	33	4	10	0.18
ATTTACCGTCTCTATGTCTATTAC	IYRLYVYY	8	4	57	40	10	7	0.18
TATCGCTATATCATTTTCCGTCTC	YRYIIFRL	6	8	28	29	5	5	0.18
ATTGTCTTTCGCTATATCTATCTC	IVFRYIYL	10	4	40	34	8	5	0.18
TGTTTCCGTCTCTATTTCTTTGGC	CFRLYFFG	8	4	121	95	11	27	0.18
CATCGCTGTTACATTGTCTTTCTC	HRCYIVFL	7	6	68	57	13	9	0.18
TATTTCCGTCTCATTATCGTTAAC	YFRLIIVN	7	4	26	25	5	4	0.18
TGTCGCTTTTACATTTACTTTGTC	CRFYIYFV	9	4	37	31	8	4	0.18
TGTTACGTTTACCTTTTCCTTGTC	CYVYLFLV	13	0	102	102	21	15	0.18
GGTATCATTATCCTTAACGGTGAC	GIIILNGD	4	-4	123	109	19	22	0.18
GTTCGCCTTCTCGTTGTCATTGTC	VRLLVVIV	12	4	99	82	14	18	0.18
TTTTACTGTCTCGTTTGCTATCTC	FYCLVCYL	12	0	215	192	31	41	0.18
GTTATCTATTACATTTGCATTGGC	VIYYICIG	11	0	64	66	13	10	0.18
GTTCACTATATCATTTGCCTTCGC	VHYIICLR	7	6	53	60	12	8	0.18
TATTTCGTTATCGTTAACTATCGC	YFVIVNYR	6	4	50	46	10	7	0.18
TATAGCGTTTACCTTGTCTTTGTC	YSVYLVFV	11	0	101	91	12	22	0.18
TATTACATTCTCCTTCGCCGTTGC	YYILLRRC	5	8	71	70	11	14	0.18
TATCTCTATTTCCTTTACGGTGGC	YLYFLYGG	9	0	120	128	19	25	0.18
TATCGCGGTTTCATTATCTATATC	YRGFIIYI	8	4	86	83	13	17	0.18
TATTACGTTATCTGTTTCATTATC	YYVICFII	13	0	84	68	17	10	0.18
CGTTACTTTTACGTTGTCATTTGC	RYFYVVIC	9	4	84	68	9	18	0.18
TGTCGCTTTGTCATTATCCGTTAC	CRFVIIRY	6	8	99	81	15	17	0.18
CTTCGCGTTTTCTATATCTATCAC	LRVFYIYH	7	6	223	204	46	30	0.18
GTTTTCTATATCTATTTCGTTGGC	VFYIYFVG	12	0	71	58	10	13	0.18
CGTCTCCTTTTCGTTTACTTTATC	RLLFVYFI	11	4	226	211	40	38	0.18
GGTTACGTTTTCTATTGCTATCAC	GYVFYCYH	7	2	115	120	18	24	0.18
TATCGCTTTTTCCGTTTCTTTTAC	YRFFRFFY	6	8	194	175	31	35	0.18
TTTTTCTATCACGTTGTCATTGTC	FFYHVVIV	12	2	59	47	10	9	0.18
TATCGCTATGTCCTTCTCTATATC	YRYVLLYI	9	4	55	62	11	10	0.18
CATCGCGTTATCGTTTTCTATGTC	HRVIVFYV	8	6	122	123	20	24	0.18
TGTCTCTATCGCATTTTCCTTGTC	CLYRIFLV	10	4	85	82	14	16	0.18
GGTATCCTTGTCCTTGTCTATGTC	GILVLVYV	13	0	347	304	52	65	0.18
TTTATCGGTTACTATCTCTATGGC	FIGYYLYG	9	0	51	38	12	4	0.18

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
CGTTACATTGTCTTTGTCTTTCTC	RYIVFVFL	11	4	96	93	10	24	0.18
TATGGCTATATCCTTCTCATTTTC	YGYILLIF	12	0	78	83	16	13	0.18
AGTTTCTATATCTATGTCATTTTC	SFYIYVIF	11	0	106	99	23	14	0.18
CTTATCTATGTCATTTTCCATCGC	LIYVIFHR	8	6	35	37	5	8	0.18
CGTTGCCTTGTCATTTACATTCGC	RCLVIYIR	6	8	82	62	13	13	0.18
TTTTACCGTATCTATTACCTTGGC	FYRIYYLG	7	4	72	44	10	11	0.18
CGTCTCTTTATCATTTTCTTTGTC	RLFIIFFV	12	4	62	65	15	8	0.18
AATATCATTCTCATTTACCGTATC	NIILIYRI	7	4	30	36	4	8	0.18
CATAACTATAGCTTTGACTTTGTC	HNYSFDFV	1	-2	37	29	9	3	0.18
AATTACATTTACTATCTCTATGGC	NYIYYLYG	6	0	42	24	7	5	0.18
AATATCTTTCTCTTTCGCCTTTTC	NIFLFRLF	8	4	62	59	9	13	0.18
CGTATCTATTACTGTCTCTTTTTC	RIYYCLFF	9	4	94	93	23	11	0.18
TATCTCCTTATCTATTACCGTGGC	YLLIYYRG	7	4	144	147	30	23	0.18
TATTACGTTCTCTATGTCTGTCTC	YYVLYVCL	12	0	135	123	24	23	0.18
GGTCTCTATGTCTTTATCCTTCGC	GLYVFILR	9	4	157	139	22	32	0.18
TTTCACGTTTACTATATCCTTTAC	FHVYYILY	10	2	62	36	13	5	0.18
CTTATCATTTACTTTATCGATTGC	LIIYFIDC	10	-4	188	160	31	33	0.18
TATCACGTTTACATTATCGTTTTC	YHVYIIVF	11	2	95	68	20	10	0.18
GTTTGCTATATCCTTCTCATTAGC	VCYILLIS	11	0	39	37	9	5	0.18
AATCGCCTTATCATTTACGTTTGC	NRLIIYVC	6	4	116	101	23	17	0.18
TATTACTTTATCCATGTCTTTATC	YYFIHVFI	11	2	78	63	10	16	0.18
TATTACTTTATCTATATCGTTTTC	YYFIYIVF	13	0	52	51	7	12	0.18
AATTTCTATATCATTTACTGTTGC	NFYIIYCC	8	0	35	30	8	4	0.18
TATATCATTGGCATTATCCGTCTC	YIIGIIRL	9	4	184	184	37	31	0.18
CTTCGCATTTTCATTGTCTTTATC	LRIFIVFI	12	4	49	70	12	10	0.18
ATTTACTGTATCGTTGTCTTTCAC	IYCIVVFH	11	2	54	65	11	11	0.18
TTTCTCCGTCTCATTATCCTTTAC	FLRLIILY	11	4	154	127	23	29	0.19
TGTCGCGGTTTCTATTTCTTTCGC	CRGFYFFR	4	8	25	29	6	4	0.19
TATCTCGGTTTCTATTACAGTCTC	YLGFYYSL	8	0	47	34	8	7	0.19
CTTATCATTATCCGTAGCTTTTTC	LIIIRSFF	9	4	82	80	17	13	0.19
TATCGCATTGTCCATATCTTTTAC	YRIVHIFY	7	6	68	56	10	13	0.19
GGTTTCGTTGTCGTTATCTTTGTC	GFVVVIFV	14	0	175	175	42	23	0.19
GTTTACCTTTTCTATTTCTATCGC	VYLFYFYR	9	4	77	79	13	16	0.19
GTTTACTATATCTATCACCGTTTC	VYYIYHRF	6	6	64	65	11	13	0.19
TATTTCCTTATCTATTGCCGTTGC	YFLIYCRC	8	4	64	54	7	15	0.19
CTTTACCTTTGCTATTACTTTATC	LYLCYYFI	12	0	77	73	11	17	0.19
TGTATCTATTTCGTTTACTGTCGC	CIYFVYCR	8	4	124	117	20	25	0.19
CGTTTCTTTGTCCTTGTCTATCTC	RFFVLVYL	11	4	130	127	27	21	0.19
TATGACATTATCATTCTCCTTTTC	YDIIILLF	11	-4	63	28	8	9	0.19
TATCGCCGTTACATTGTCATTGTC	YRRYIVIV	6	8	103	95	16	21	0.19
CGTTGCGTTCTCGTTGTCTTTATC	RCVLVVFI	11	4	118	96	26	14	0.19
AGTTTCAATCTCTGTTTCGATTAC	SFNLCFDY	3	-4	30	34	5	7	0.19
ATTCGCTTTATCCTTGTCTATCGC	IRFILVYR	7	8	33	31	7	5	0.19
CGTCACATTTTCTATATCATTATC	RHIFYIII	8	6	44	20	10	2	0.19
TATATCCATTACTTTTTCCTTATC	YIHYFFLI	11	2	54	58	11	10	0.19
TTTCACGTTGTCCTTTACCTTGTC	FHVVLYLV	12	2	138	123	20	29	0.19
GGTATCTTTTACATTATCGTTTAC	GIFYIIVY	12	0	99	114	12	28	0.19
TATCGCATTTTCTGTCGCCTTCAC	YRIFCRLH	3	10	82	67	16	12	0.19
TTTTGCGTTTACCTTATCCTTCTC	FCVYLILL	14	0	157	141	32	24	0.19
AGTTACTGTAGCCTTCGCTTTTTC	SYCSLRFF	4	4	64	53	11	11	0.19
TGTCGCTTTTTCATTTTCTATTTC	CRFFIFYF	10	4	51	34	11	5	0.19
CTTCGCTATTTCATTTGCATTCTC	LRYFICIL	10	4	55	30	5	11	0.19
TTTAGCTATTTCGTTGTCTTTGTC	FSYFVVFV	12	0	143	133	22	30	0.19
CGTTTCAGTTTCTTTATCGTTATC	RFSFFIVI	9	4	145	115	26	23	0.19

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATTACGTTCTCTGTTTCTATTAC	YYVLCFYY	11	0	106	85	15	21	0.19
GGTTGCTATTTCATTTTCCTTGGC	GCYFIFLG	10	0	30	23	6	4	0.19
TTTTTCTATCACATTTACCTTGTC	FFYHIYLV	11	2	91	68	11	19	0.19
TGTTACCATGTCATTTACTTTATC	CYHVIYFI	10	2	97	83	19	15	0.19
ATTATCGTTATCCGTTGCTTTGGC	IIVIRCFG	9	4	174	138	33	26	0.19
TATCTCTATCACCTTTACATTATC	YLYHLYII	10	2	56	55	8	13	0.19
TTTCGCTGTATCTTTGTCCGTTAC	FRCIFVRY	6	8	75	73	18	10	0.19
GGTTGCTGTGTCGTTGTCTTTATC	GCCVVVFI	12	0	126	133	25	24	0.19
TATCGCTATCACTTTCGCCTTTAC	YRYHFRLY	2	10	112	115	21	22	0.19
TATTGCCTTATCTATATCATTATC	YCLIYIII	13	0	168	154	34	27	0.19
CTTTTCTATCGCATTTTCATTTTC	LFYRIFIF	11	4	96	94	17	19	0.19
ATTGTCATTATCCGTTGCTGTCTC	IVIIRCCL	10	4	34	24	6	5	0.19
GATGTCATTCTCTTTTACTTTTAC	DVILFYFY	10	-4	61	55	8	14	0.19
AGTGGCATTCTCTTTATCCTTTGC	SGILFILC	10	0	135	149	26	28	0.19
AGTTTCATTGTCATTCTCTATTGC	SFIVILYC	11	0	189	163	29	38	0.19
TTTCGCCTTATCATTGTCTATGGC	FRLIIVYG	9	4	44	40	4	12	0.19
CATTGCTATTTCTATCTCATTTTC	HCYFYLIF	10	2	57	48	11	9	0.19
ATTATCTTTCGCTTTCGCCATTTC	IIFRFRHF	5	10	86	61	12	16	0.19
AATGTCGTTCTCTATATCCTTCGC	NVVLYILR	7	4	87	65	19	10	0.19
CGTATCATTGTCTTTGTCTATTAC	RIIVFVYY	10	4	63	68	11	14	0.19
TATAGCGTTCTCGTTCTCCGTTTC	YSVLVLRF	8	4	47	42	8	9	0.19
CATGGCCGTTTCTTTTTCCTTTAC	HGRFFFLY	6	6	48	41	8	9	0.19
TATATCTATCTCTATGTCCTTCTC	YIYLYVLL	13	0	119	85	20	19	0.19
TATTTCTGTCTCTATGTCATTATC	YFCLYVII	13	0	180	160	27	38	0.19
TATGTCCTTCGCTGTATCATTCTC	YVLRCIIL	10	4	69	93	17	14	0.19
AATTGCCTTTACGTTATCATTTTC	NCLYVIIF	10	0	110	109	14	28	0.19
ATTCGCATTTTCTTTCGCGGTCTC	IRIFFRGL	6	8	63	62	16	8	0.19
GGTGGCATTTACGTTGTCTTTTTC	GGIYVVFF	11	0	68	67	16	10	0.19
AATATCTATTACTGTCGCATTTTC	NIYYCRIF	5	4	58	51	14	7	0.19
GTTCGCTTTTTCTATATCGTTGGC	VRFFYIVG	9	4	61	48	6	15	0.19
GTTGTCTATTACATTCGCTATTGC	VVYYIRYC	8	4	94	98	19	18	0.19
GTTTACTTTGTCATTTTCTATGTC	VYFVIFYV	14	0	113	84	20	18	0.19
GTTATCCGTGTCATTATCTTTGTC	VIRVIIFV	12	4	29	28	8	3	0.19
TATTACGTTTTCTATGTCTATTTC	YYVFYVYF	12	0	88	83	12	21	0.19
CGTATCTTTGTCCTTCTCCGTAGC	RIFVLLRS	5	8	47	41	9	8	0.19
GGTTACATTTTCGTTTACCTTCGC	GYIFVYLR	8	4	89	87	20	14	0.19
TTTGTCATTTACCGTATCAATTTC	FVIYRINF	7	4	94	82	23	11	0.19
TTTGTCGTTCACATTTTCCTTGTC	FVVHIFLV	13	2	124	114	23	23	0.19
GATTACTATTTCGTTTACATTTTC	DYYFVYIF	9	-4	70	80	12	17	0.19
AATTACCTTTTCATTCTCATTTAC	NYLFILIY	10	0	45	48	8	10	0.19
TATTACATTATCCGTCTCTGTTAC	YYIIRLCY	8	4	189	152	28	38	0.19
CGTCGCATTGTCATTTTCTGTGTC	RRIVIFCV	7	8	137	116	22	27	0.19
GGTTGCCTTCACATTATCTTTTTC	GCLHIIFF	10	2	154	135	32	24	0.19
TGTAGCTTTGTCCTTATCATTTAC	CSFVLIIY	11	0	52	46	7	12	0.19
ATTCGCCTTTTCTTTCTCCTTAAC	IRLFFLLN	8	4	201	160	36	34	0.19
CGTGTCTGTAACTGTATCTTTCTC	RVCNCIFL	6	4	36	31	6	7	0.19
TATTGCTTTCTCATTTACGGTTAC	YCFLIYGY	10	0	34	38	5	9	0.19
GGTATCGGTATCTATTGCATTTAC	GIGIYCIY	9	0	45	63	5	16	0.19
TTTCTCTGTGTCATTCGCTGTCGC	FLCVIRCR	6	8	60	48	13	8	0.19
GATCGCGTTTACATTATCTTTCTC	DRVYIIFL	7	0	186	174	37	33	0.19
TATCGCTTTGTCATTCTCGTTATC	YRFVILVI	11	4	91	94	16	20	0.19
CGTCTCTTTATCGTTCGCTATTTC	RLFIVRYF	7	8	35	42	10	5	0.19
GTTGGCATTCTCTTTATCCGTCTC	VGILFIRL	10	4	45	32	7	8	0.19
TTTTACAGTGTCATTATCATTATC	FYSVIIII	12	0	57	61	9	14	0.19

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
ATTTACTATCTCTATCGCTGTGTC	IYYLYRCV	8	4	58	60	14	9	0.19
TATAGCGTTCTCTTTATCGTTTTC	YSVLFIVF	12	0	86	78	18	14	0.20
TATCGCTTTTTCTTTATCGGTGGC	YRFFFIGG	7	4	71	57	15	10	0.20
AGTTTCATTATCATTCGCTATGTC	SFIIIRYV	8	4	75	58	10	16	0.20
TATTACATTATCGTTTTCCATTAC	YYIIVFHY	10	2	93	86	17	18	0.20
TATGTCTTTTACTGTCTCTGTCGC	YVFYCLCR	8	4	111	88	17	22	0.20
TATTACAGTATCCTTCTCATTCGC	YYSILLIR	7	4	48	54	8	12	0.20
TTTTACCTTCTCCTTTACTATCGC	FYLLLYYR	9	4	22	39	3	9	0.20
ATTTGCCTTATCCGTGTCTTTTAC	ICLIRVFY	10	4	587	617	114	123	0.20
TATATCATTCTCTATTTCTGTCGC	YIILYFCR	9	4	35	31	7	6	0.20
TTTTTCCGTGTCCTTATCCTTATC	FFRVLILI	12	4	73	74	13	16	0.20
ATTATCTATGTCTTTATCCGTCAC	IIYVFIRH	8	6	74	73	11	18	0.20
CTTCGCCTTTACCGTATCTTTATC	LRLYRIFI	7	8	122	111	19	27	0.20
TATTGCATTCACGTTTACCTTAAC	YCIHVYLN	6	2	38	43	4	12	0.20
CGTATCTTTTACCTTGTCATTATC	RIFYLVII	11	4	42	44	6	11	0.20
GTTTACATTCTCTTTGTCGGTCTC	VYILFVGL	13	0	98	84	17	19	0.20
ATTTACTATTGCTATCGCCGTCTC	IYYCYRRL	4	8	53	48	11	9	0.20
TATTACTATATCCGTAGCTATTTC	YYYIRSYF	5	4	56	45	7	13	0.20
CATGTCCATATCATTTACATTCTC	HVHIIYIL	9	4	61	50	8	14	0.20
AATTTCTATTTCATTTACTTTATC	NFYFIYFI	10	0	55	66	8	16	0.20
CTTAGCATTTACTATATCCGTTGC	LSIYYIRC	6	4	113	129	18	30	0.20
GGTATCTATATCATTCGCGTTCTC	GIYIIRVL	9	4	65	81	11	18	0.20
CATTACGTTATCGTTTTCTATTAC	HYVIVFYY	10	2	88	68	12	19	0.20
TTTTTCCGTTACTTTCTCCATGGC	FFRYFLHG	6	6	89	72	18	14	0.20
TTTTACAGTGTCTATGTCTTTATC	FYSVYVFI	11	0	81	85	12	21	0.20
TATTTCTTTGTCTATCGCCGTTAC	YFFVYRRY	5	8	306	286	46	72	0.20
AATTACGTTTTCTATTTCGTTCTC	NYVFYFVL	10	0	27	33	7	5	0.20
TTTCACTATTTCAGTCTCTATGGC	FHYFSLYG	6	2	36	34	7	7	0.20
TATTGCATTTACCTTGTCCTTAGC	YCIYLVLS	10	0	53	32	8	9	0.20
GTTAGCCTTCTCATTTTCTTTATC	VSLLIFFI	13	0	63	47	11	11	0.20
TATTACGTTCTCTGTTGCCTTCAC	YYVLCCLH	9	2	59	56	7	16	0.20
CGTTTCCATCTCTTTGTCTTTCTC	RFHLFVFL	9	6	81	54	12	15	0.20
AGTTACGTTCTCATTTGCCTTATC	SYVLICLI	11	0	65	80	17	12	0.20
TTTGACTATTACATTTTCCTTATC	FDYYIFLI	10	-4	82	73	13	18	0.20
CTTCGCGTTCTCCTTTTCCGTTAC	LRVLLFRY	7	8	157	137	24	35	0.20
TATTACATTATCTTTTTCAATAGC	YYIIFFNS	7	0	126	123	23	27	0.20
CTTCTCGTTCTCTATCACTATATC	LLVLYHYI	11	2	115	109	27	18	0.20
TATTACTATATCTGTGTCTTTTAC	YYYICVFY	11	0	107	77	22	15	0.20
GGTCGCCTTATCGTTGTCTTTCGC	GRLIVVFR	6	8	86	68	14	17	0.20
TATATCGGTTTCCGTTTCTATGGC	YIGFRFYG	6	4	68	76	14	15	0.20
TTTAGCTATCTCTTTCTCTATTTC	FSYLFLYF	11	0	83	61	16	13	0.20
AGTTACATTCTCTATGTCTATATC	SYILYVYI	10	0	65	74	10	18	0.20
TGTTACTATCTCTTTTGCTTTGTC	CYYLFCFV	12	0	207	200	37	45	0.20
TATGACATTTTCGTTTTCCTTGTC	YDIFVFLV	11	-4	64	65	11	15	0.20
TGTTACTTTGTCTTTTTCCGTCGC	CYFVFFRR	6	8	66	63	12	14	0.20
ATTTACGTTTACTATCTCGGTTAC	IYVYYLGY	10	0	70	54	11	14	0.20
CATATCGTTCTCTTTTTCTATGGC	HIVLFFYG	10	2	57	62	10	14	0.20
GTTTACTATCTCTTTGTCTGTAAC	VYYLFVCN	9	0	59	55	11	12	0.20
GTTATCTGTCTCCTTCTCCGTCGC	VICLLLRR	7	8	117	96	20	23	0.20
ATTGTCATTTACTTTGTCGGTCGC	IVIYFVGR	9	4	93	105	15	25	0.20
TATTACATTCGCATTTACATTATC	YYIRIYII	9	4	54	30	10	7	0.20
CGTTACGTTGTCATTCTCGTTCAC	RYVVILVH	8	6	310	258	57	58	0.20
TATGACATTATCATTTTCTTTGTC	YDIIIFFV	11	-4	34	45	7	9	0.20
AGTCTCCGTTTCGTTTTCTTTTTC	SLRFVFFF	9	4	39	40	7	9	0.20

DNASEQ	PEPSEQ	HPHOB	CHARGE	REF1	REF2	IND1	IND2	toxicity
TATTACATTTACCATCTCATTGTC	YYIYHLIV	10	2	133	104	24	24	0.20
TATTTCTTTGTCCTTTTCCTTTGC	YFFVLFLC	14	0	34	35	7	7	0.20
GTTATCGTTCTCTGTCGCTATATC	VIVLCRYI	10	4	63	70	14	13	0.20
CATTACTATTACATTTGCGTTTTC	HYYYICVF	9	2	90	82	16	19	0.20
TGTTACCTTATCTGTATCTATCGC	CYLICIYR	8	4	91	81	16	19	0.20
TATTACATTCTCCTTATCAGTTTC	YYILLISF	11	0	52	61	8	15	0.20
CATTACATTATCTATTACTGTATC	HYIIYYCI	9	2	115	106	29	16	0.20
TATCGCTATTACTATGGCCTTTAC	YRYYYGLY	5	4	52	56	12	10	0.20
CTTTACGTTTTCTATATCTATCGC	LYVFYIYR	9	4	145	120	24	30	0.20
TATATCCGTTGCTATTTCTATATC	YIRCYFYI	8	4	84	73	20	12	0.20
AATATCGTTTTCATTTGCTATATC	NIVFICYI	10	0	54	49	5	16	0.20
GTTAACTATCTCATTGTCTTTATC	VNYLIVFI	11	0	99	97	18	22	0.20
CGTCTCATTTACCTTTTCTGTCAC	RLIYLFCH	7	6	91	51	15	14	0.20
ATTTGCTATATCCATCGCCGTCGC	ICYIHRRR	-1	14	45	48	11	8	0.20
ATTGTCATTCTCTATCGCCGTCTC	IVILYRRL	7	8	102	69	21	14	0.20
TATATCGTTATCCGTTTCTATCGC	YIVIRFYR	6	8	251	218	47	49	0.20
ATTTTCTATCTCGTTATCTTTGTC	IFYLVIFV	15	0	82	79	18	15	0.20
TTTTTCTATCGCGTTTACCTTTTC	FFYRVYLF	10	4	85	76	21	12	0.20

Motif results

3mer, 0 wildcards

kmer	tox#	comp#	tox/comp	tox-comp	tox+comp	charge	hphobic	KD_HPHOBIC	ROSE_HPHOBIC	z-value	p-value
FYL	201	50	4.02	151	251	0	3	5.30	2.49	9.02	0.00
YYF	251	68	3.69	183	319	0	3	0.20	2.40	7.93	0.00
FYF	249	68	3.66	181	317	0	3	4.30	2.52	8.89	0.00
YYL	169	48	3.52	121	217	0	3	1.20	2.37	8.36	0.00
YLY	156	47	3.32	109	203	0	3	1.20	2.37	6.78	0.00
FFY	238	72	3.31	166	310	0	3	4.30	2.52	8.75	0.00
LYF	165	50	3.30	115	215	0	3	5.30	2.49	8.23	0.00
FYY	206	65	3.17	141	271	0	3	0.20	2.40	7.34	0.00
IYY	208	67	3.10	141	275	0	3	1.90	2.40	8.43	0.00
YYI	218	72	3.03	146	290	0	3	1.90	2.40	8.41	0.00
YFF	208	70	2.97	138	278	0	3	4.30	2.52	7.88	0.00
YLF	198	67	2.96	131	265	0	3	5.30	2.49	8.50	0.00
YYY	174	59	2.95	115	233	0	3	-3.90	2.28	7.47	0.00
YFY	194	68	2.85	126	262	0	3	0.20	2.40	8.64	0.00
YFL	189	72	2.63	117	261	0	3	5.30	2.49	8.69	0.00
LFF	170	65	2.62	105	235	0	3	9.40	2.61	5.74	0.00
LYL	118	46	2.57	72	164	0	3	6.30	2.46	7.25	0.00
FLY	177	69	2.57	108	246	0	3	5.30	2.49	6.88	0.00
YFI	244	97	2.52	147	341	0	3	6.00	2.52	9.80	0.00
YLL	104	42	2.48	62	146	0	3	6.30	2.46	6.29	0.00
IYF	216	88	2.45	128	304	0	3	6.00	2.52	10.01	0.00
FFL	182	76	2.39	106	258	0	3	9.40	2.61	5.89	0.00
CYF	189	79	2.39	110	268	0	2	4.00	2.55	6.44	0.00
YIL	169	71	2.38	98	240	0	3	7.00	2.49	5.09	0.00
FIY	216	91	2.37	125	307	0	3	6.00	2.52	6.35	0.00
IYL	163	70	2.33	93	233	0	3	7.00	2.49	7.00	0.00
FYI	222	96	2.31	126	318	0	3	6.00	2.52	8.46	0.00
YIY	188	85	2.21	103	273	0	3	1.90	2.40	7.12	0.00
YII	249	113	2.20	136	362	0	3	7.70	2.52	9.49	0.00
FRL	141	64	2.20	77	205	1	2	2.10	2.37	5.16	0.00
CYL	174	79	2.20	95	253	0	2	5.00	2.52	5.99	0.00
LFY	138	63	2.19	75	201	0	3	5.30	2.49	6.42	0.00
CYI	218	100	2.18	118	318	0	2	5.70	2.55	7.87	0.00
FLF	181	84	2.15	97	265	0	3	9.40	2.61	7.25	0.00
LYY	135	63	2.14	72	198	0	3	1.20	2.37	6.32	0.00
RFF	169	79	2.14	90	248	1	2	1.10	2.40	6.52	0.00
FFF	205	96	2.14	109	301	0	3	8.40	2.64	4.50	0.00
RLF	131	62	2.11	69	193	1	2	2.10	2.37	5.87	0.00
YYC	189	92	2.05	97	281	0	2	-0.10	2.43	5.89	0.00
LYI	178	87	2.05	91	265	0	3	7.00	2.49	6.47	0.00
FCY	198	97	2.04	101	295	0	2	4.00	2.55	6.55	0.00
IFY	195	96	2.03	99	291	0	3	6.00	2.52	7.19	0.00
IIY	198	98	2.02	100	296	0	3	7.70	2.52	7.46	0.00
YIF	199	99	2.01	100	298	0	3	6.00	2.52	6.13	0.00
LFI	180	90	2.00	90	270	0	3	11.10	2.61	5.75	0.00
YLI	150	76	1.97	74	226	0	3	7.00	2.49	5.38	0.00
IYI	186	95	1.96	91	281	0	3	7.70	2.52	6.68	0.00
ILY	143	75	1.91	68	218	0	3	7.00	2.49	4.85	0.00
LFL	117	62	1.89	55	179	0	3	10.40	2.58	5.09	0.00
YRL	111	59	1.88	52	170	1	2	-2.00	2.25	4.39	0.00

4mer, 1 wildcard

kmer	tox#	comp#	tox/comp	tox-comp	tox+comp	charge	hphobic	KD_HPHOBIC	ROSE_HPHOBIC	z-value	p-value
YICY	36	0	INF	36	36	0	3	4.40	3.31	5.29	0.00
YCVI	39	1	39.00	38	40	0	3	9.90	3.41	4.06	0.00
YFCY	36	1	36.00	35	37	0	3	2.70	3.31	5.64	0.00
VCYI	31	1	31.00	30	32	0	3	9.90	3.41	4.56	0.00
YRFF	31	1	31.00	30	32	1	3	-0.20	3.16	4.85	0.00
FCRL	31	1	31.00	30	32	1	2	4.60	3.28	2.73	0.00
FYYI	29	1	29.00	28	30	0	4	4.70	3.28	4.49	0.00
FYII	38	2	19.00	36	40	0	4	10.50	3.40	5.64	0.00
VYII	35	2	17.50	33	37	0	4	11.90	3.38	5.30	0.00
ICFI	35	2	17.50	33	37	0	3	14.30	3.55	5.24	0.00
YCIV	32	2	16.00	30	34	0	3	9.90	3.41	2.86	0.00
FCYI	29	2	14.50	27	31	0	3	8.50	3.43	4.32	0.00
YFIY	33	3	11.00	30	36	0	4	4.70	3.28	3.80	0.00
YIII	32	3	10.67	29	35	0	4	12.20	3.40	5.27	0.00
FFYF	32	3	10.67	29	35	0	4	7.10	3.40	4.20	0.00
FCYF	32	3	10.67	29	35	0	3	6.80	3.43	3.25	0.00
FCII	32	3	10.67	29	35	0	3	14.30	3.55	4.47	0.00
YYII	32	3	10.67	29	35	0	4	6.40	3.28	6.78	0.00
FYIY	31	3	10.33	28	34	0	4	4.70	3.28	4.65	0.00
YYCY	31	3	10.33	28	34	0	3	-1.40	3.19	4.11	0.00
VYFI	30	3	10.00	27	33	0	4	10.20	3.38	4.69	0.00
FIIY	30	3	10.00	27	33	0	4	10.50	3.40	4.48	0.00
VFII	30	3	10.00	27	33	0	4	16.00	3.50	4.35	0.00
YVIF	29	3	9.67	26	32	0	4	10.20	3.38	4.93	0.00
YYIY	29	3	9.67	26	32	0	4	0.60	3.16	2.78	0.00
FYFG	29	3	9.67	26	32	0	3	3.90	3.24	4.51	0.00
YVYY	29	3	9.67	26	32	0	4	0.30	3.14	2.96	0.00
IVYI	27	3	9.00	24	30	0	4	11.90	3.38	3.13	0.00
FCFI	27	3	9.00	24	30	0	3	12.60	3.55	3.75	0.00
FVIY	27	3	9.00	24	30	0	4	10.20	3.38	3.69	0.00
YYFI	27	3	9.00	24	30	0	4	4.70	3.28	3.02	0.00
YIIF	43	5	8.60	38	48	0	4	10.50	3.40	4.79	0.00
YCXL	187	23	8.13	164	210	0	2	5.00	2.52	8.82	0.00
IYYF	32	4	8.00	28	36	0	4	4.70	3.28	4.23	0.00
YIIY	30	4	7.50	26	34	0	4	6.40	3.28	4.09	0.00
YIFY	30	4	7.50	26	34	0	4	4.70	3.28	3.14	0.00
FYFI	29	4	7.25	25	33	0	4	8.80	3.40	3.92	0.00
YFIV	36	5	7.20	31	41	0	4	10.20	3.38	5.21	0.00
YXLF	199	28	7.11	171	227	0	3	5.30	2.49	9.77	0.00
YVFY	34	5	6.80	29	39	0	4	4.40	3.26	5.96	0.00
YIVY	27	4	6.75	23	31	0	4	6.10	3.26	4.68	0.00
FYYF	27	4	6.75	23	31	0	4	3.00	3.28	1.37	0.09
LFII	27	4	6.75	23	31	0	4	15.60	3.49	4.39	0.00
FRIF	26	4	6.50	22	30	1	3	5.60	3.28	2.59	0.00
FYVV	38	6	6.33	32	44	0	4	9.90	3.36	5.21	0.00
YCXI	231	37	6.24	194	268	0	2	5.70	2.55	7.26	0.00
FFYG	30	5	6.00	25	35	0	3	3.90	3.24	3.37	0.00
IIFY	30	5	6.00	25	35	0	4	10.50	3.40	4.46	0.00
YXLY	163	28	5.82	135	191	0	3	1.20	2.37	7.96	0.00
FXYF	226	39	5.79	187	265	0	3	4.30	2.52	8.45	0.00

5mer, 1 wildcard

kmer	tox#	comp#	tox/comp	tox-comp	tox+comp	charge	hphobic	KD_HPHOBIC	ROSE_HPHOBIC	z-value	p-value
YFXYF	38	0	INF	38	38	0	4	3.00	3.28	5.53	0.00
YFCYX	31	0	INF	31	31	0	3	2.70	3.31	5.39	0.00
YICYX	30	0	INF	30	30	0	3	4.40	3.31	4.84	0.00
FYIIX	35	1	35.00	34	36	0	4	10.50	3.40	5.90	0.00
YIIXY	33	1	33.00	32	34	0	4	6.40	3.28	5.10	0.00
VYIIX	33	1	33.00	32	34	0	4	11.90	3.38	5.19	0.00
YCVIX	32	1	32.00	31	33	0	3	9.90	3.41	3.56	0.00
XYCVI	32	1	32.00	31	33	0	3	9.90	3.41	3.47	0.00
XVYII	31	1	31.00	30	32	0	4	11.90	3.38	5.08	0.00
YCIVX	30	1	30.00	29	31	0	3	9.90	3.41	2.69	0.00
XYFCY	30	1	30.00	29	31	0	3	2.70	3.31	5.47	0.00
FYXIF	34	2	17.00	32	36	0	4	8.80	3.40	7.12	0.00
YYIIX	30	2	15.00	28	32	0	4	6.40	3.28	6.61	0.00
XYIII	30	2	15.00	28	32	0	4	12.20	3.40	5.36	0.00
VFXIY	29	2	14.50	27	31	0	4	10.20	3.38	4.06	0.00
YFCXF	29	2	14.50	27	31	0	3	6.80	3.43	4.89	0.00
VYFIX	28	2	14.00	26	30	0	4	10.20	3.38	4.25	0.00
YXLFV	28	2	14.00	26	30	0	4	9.50	3.35	3.42	0.00
ICFIX	28	2	14.00	26	30	0	3	14.30	3.55	4.50	0.00
VFIIX	28	2	14.00	26	30	0	4	16.00	3.50	4.38	0.00
XYIIF	35	3	11.67	32	38	0	4	10.50	3.40	4.58	0.00
FYVXF	30	3	10.00	27	33	0	4	8.50	3.38	6.70	0.00
XYIVF	29	3	9.67	26	32	0	4	10.20	3.38	4.65	0.00
XYFIY	29	3	9.67	26	32	0	4	4.70	3.28	3.52	0.00
FCYFX	29	3	9.67	26	32	0	3	6.80	3.43	3.20	0.00
XIIFY	28	3	9.33	25	31	0	4	10.50	3.40	4.37	0.00
IYFVX	28	3	9.33	25	31	0	4	10.20	3.38	3.86	0.00
FFXRY	28	3	9.33	25	31	1	3	-0.20	3.16	3.92	0.00
XIYYF	27	3	9.00	24	30	0	4	4.70	3.28	3.69	0.00
YXIYF	27	3	9.00	24	30	0	4	4.70	3.28	6.15	0.00
FXIIR	27	3	9.00	24	30	1	3	7.30	3.28	3.27	0.00
YYIYX	27	3	9.00	24	30	0	4	0.60	3.16	2.71	0.00
YIIFX	35	4	8.75	31	39	0	4	10.50	3.40	4.07	0.00
FYVVX	34	4	8.50	30	38	0	4	9.90	3.36	4.94	0.00
YIXFI	32	4	8.00	28	36	0	4	10.50	3.40	4.33	0.00
YXIFY	32	4	8.00	28	36	0	4	4.70	3.28	5.20	0.00
YVFYX	30	4	7.50	26	34	0	4	4.40	3.26	5.59	0.00
YXVYV	29	4	7.25	25	33	0	4	5.80	3.24	3.92	0.00
YXFFF	29	4	7.25	25	33	0	4	7.10	3.40	4.38	0.00
FRXFF	29	4	7.25	25	33	1	3	3.90	3.28	4.36	0.00
YIIXV	36	5	7.20	31	41	0	4	11.90	3.38	4.17	0.00
XIYIV	28	4	7.00	24	32	0	4	11.90	3.38	5.78	0.00
FYCXI	27	4	6.75	23	31	0	3	8.50	3.43	1.37	0.08
YFIVX	27	4	6.75	23	31	0	4	10.20	3.38	4.60	0.00
XILFI	26	4	6.50	22	30	0	4	15.60	3.49	4.87	0.00
XYVFY	32	5	6.40	27	37	0	4	4.40	3.26	5.52	0.00
VYXIY	29	5	5.80	24	34	0	4	6.10	3.26	3.02	0.00
YXVFY	29	5	5.80	24	34	0	4	4.40	3.26	3.08	0.00
FYVFX	28	5	5.60	23	33	0	4	8.50	3.38	4.80	0.00
FYXIY	28	5	5.60	23	33	0	4	4.70	3.28	4.64	0.00

5mer, 2 wildcards

kmer	tox#	comp#	tox/comp	tox-comp	tox+comp	charge	hphobic	KD_HPHOBIC	ROSE_HPHOBIC	z-value	p-value
YFXYF	38	0	INF	38	38	0	4	3.00	3.28	5.53	0.00
YFCYX	31	0	INF	31	31	0	3	2.70	3.31	5.39	0.00
YICYX	30	0	INF	30	30	0	3	4.40	3.31	4.84	0.00
FYIIX	35	1	35.00	34	36	0	4	10.50	3.40	5.90	0.00
YIIXY	33	1	33.00	32	34	0	4	6.40	3.28	5.10	0.00
VYIIX	33	1	33.00	32	34	0	4	11.90	3.38	5.19	0.00
YCVIX	32	1	32.00	31	33	0	3	9.90	3.41	3.56	0.00
XYCVI	32	1	32.00	31	33	0	3	9.90	3.41	3.47	0.00
XVYII	31	1	31.00	30	32	0	4	11.90	3.38	5.08	0.00
YCIVX	30	1	30.00	29	31	0	3	9.90	3.41	2.69	0.00
XYFCY	30	1	30.00	29	31	0	3	2.70	3.31	5.47	0.00
FYXIF	34	2	17.00	32	36	0	4	8.80	3.40	7.12	0.00
YYIIX	30	2	15.00	28	32	0	4	6.40	3.28	6.61	0.00
XYIII	30	2	15.00	28	32	0	4	12.20	3.40	5.36	0.00
VFXIY	29	2	14.50	27	31	0	4	10.20	3.38	4.06	0.00
YFCXF	29	2	14.50	27	31	0	3	6.80	3.43	4.89	0.00
VYFIX	28	2	14.00	26	30	0	4	10.20	3.38	4.25	0.00
YXLFV	28	2	14.00	26	30	0	4	9.50	3.35	3.42	0.00
ICFIX	28	2	14.00	26	30	0	3	14.30	3.55	4.50	0.00
VFIIX	28	2	14.00	26	30	0	4	16.00	3.50	4.38	0.00
XYIIF	35	3	11.67	32	38	0	4	10.50	3.40	4.58	0.00
FYVXF	30	3	10.00	27	33	0	4	8.50	3.38	6.70	0.00
XYIVF	29	3	9.67	26	32	0	4	10.20	3.38	4.65	0.00
XYFIY	29	3	9.67	26	32	0	4	4.70	3.28	3.52	0.00
FCYFX	29	3	9.67	26	32	0	3	6.80	3.43	3.20	0.00
YXLFX	172	18	9.56	154	190	0	3	5.30	2.49	9.18	0.00
FFXRY	28	3	9.33	25	31	1	3	-0.20	3.16	3.92	0.00
XIIFY	28	3	9.33	25	31	0	4	10.50	3.40	4.37	0.00
IYFVX	28	3	9.33	25	31	0	4	10.20	3.38	3.86	0.00
XIYYF	27	3	9.00	24	30	0	4	4.70	3.28	3.69	0.00
YXIYF	27	3	9.00	24	30	0	4	4.70	3.28	6.15	0.00
FXIIR	27	3	9.00	24	30	1	3	7.30	3.28	3.27	0.00
YYIYX	27	3	9.00	24	30	0	4	0.60	3.16	2.71	0.00
YIIFX	35	4	8.75	31	39	0	4	10.50	3.40	4.07	0.00
XXFYL	131	15	8.73	116	146	0	3	5.30	2.49	7.55	0.00
XYXLF	163	19	8.58	144	182	0	3	5.30	2.49	8.95	0.00
FYVVX	34	4	8.50	30	38	0	4	9.90	3.36	4.94	0.00
YIXFI	32	4	8.00	28	36	0	4	10.50	3.40	4.33	0.00
YXIFY	32	4	8.00	28	36	0	4	4.70	3.28	5.20	0.00
XYCXL	149	19	7.84	130	168	0	2	5.00	2.52	7.66	0.00
YVFYX	30	4	7.50	26	34	0	4	4.40	3.26	5.59	0.00
YXCYX	179	24	7.46	155	203	0	2	-0.10	2.43	10.07	0.00
YXVYV	29	4	7.25	25	33	0	4	5.80	3.24	3.92	0.00
YXFFF	29	4	7.25	25	33	0	4	7.10	3.40	4.38	0.00
FRXFF	29	4	7.25	25	33	1	3	3.90	3.28	4.36	0.00
YIIXV	36	5	7.20	31	41	0	4	11.90	3.38	4.17	0.00
XIYIV	28	4	7.00	24	32	0	4	11.90	3.38	5.78	0.00
XYCXI	189	27	7.00	162	216	0	2	5.70	2.55	6.00	0.00
FYCXI	27	4	6.75	23	31	0	3	8.50	3.43	1.37	0.08
YFIVX	27	4	6.75	23	31	0	4	10.20	3.38	4.60	0.00

6mer, 2 wildcards

kmer	tox#	comp#	tox/comp	tox-comp	tox+comp	charge	hphobic	KD_HPHOBIC	ROSE_HPHOBIC	z-value	p-value
YFXYFX	36	0	INF	36	36	0	4	3.00	3.28	5.49	0.00
XYFXYF	30	0	INF	30	30	0	4	3.00	3.28	4.89	0.00
FYIIXX	29	1	29.00	28	30	0	4	10.50	3.40	5.48	0.00
YYIIXX	30	2	15.00	28	32	0	4	6.40	3.28	6.61	0.00
FYXIFX	28	2	14.00	26	30	0	4	8.80	3.40	6.12	0.00
FYVVXX	29	3	9.67	26	32	0	4	9.90	3.36	4.71	0.00
YFXVXF	29	3	9.67	26	32	0	4	8.50	3.38	5.09	0.00
YXIVXF	30	4	7.50	26	34	0	4	10.20	3.38	3.11	0.00
YIIXVX	29	4	7.25	25	33	0	4	11.90	3.38	4.02	0.00
XYVFYX	28	4	7.00	24	32	0	4	4.40	3.26	5.13	0.00
YVFXFX	27	4	6.75	23	31	0	4	8.50	3.38	5.37	0.00
FXVYIX	27	4	6.75	23	31	0	4	10.20	3.38	3.64	0.00
FXVYXF	30	5	6.00	25	35	0	4	8.50	3.38	4.10	0.00
YVIVXX	27	5	5.40	22	32	0	4	11.60	3.36	3.40	0.00
FYXIYX	25	5	5.00	20	30	0	4	4.70	3.28	4.10	0.00
XYXIIF	27	6	4.50	21	33	0	4	10.50	3.40	5.41	0.00
YXVIVX	26	6	4.33	20	32	0	4	11.60	3.36	4.30	0.00
XYVIVX	25	6	4.17	19	31	0	4	11.60	3.36	3.74	0.00
YFXXIV	25	6	4.17	19	31	0	4	10.20	3.38	3.06	0.00
XYVFXF	25	6	4.17	19	31	0	4	8.50	3.38	3.84	0.00
YXIVVX	32	8	4.00	24	40	0	4	11.60	3.36	2.59	0.00
YXIIFX	26	7	3.71	19	33	0	4	10.50	3.40	4.93	0.00
YVXYVX	25	7	3.57	18	32	0	4	5.80	3.24	3.18	0.00
XVIVVX	24	7	3.43	17	31	0	4	17.10	3.46	2.79	0.00
FYVIXX	24	7	3.43	17	31	0	4	10.20	3.38	2.61	0.00
XXIVVF	23	7	3.29	16	30	0	4	15.70	3.48	2.28	0.01
FXVYVX	23	8	2.88	15	31	0	4	9.90	3.36	1.94	0.03
XVXYIV	23	8	2.88	15	31	0	4	11.60	3.36	2.21	0.01
XYVIXF	22	8	2.75	14	30	0	4	10.20	3.38	3.21	0.00
YFVVXX	26	10	2.60	16	36	0	4	9.90	3.36	2.14	0.02
XIVYXI	22	9	2.44	13	31	0	4	11.90	3.38	2.69	0.00
YVIXVX	24	10	2.40	14	34	0	4	11.60	3.36	2.55	0.01
FIVYXX	26	11	2.36	15	37	0	4	10.20	3.38	3.36	0.00
XFIVXF	24	11	2.18	13	35	0	4	14.30	3.50	2.60	0.00
FVVYXX	23	11	2.09	12	34	0	4	9.90	3.36	2.96	0.00
YIVVXX	23	11	2.09	12	34	0	4	11.60	3.36	2.50	0.01
XFVVVX	25	12	2.08	13	37	0	4	15.40	3.46	1.55	0.06
XFVIVX	22	11	2.00	11	33	0	4	15.70	3.48	2.58	0.00
FIIXVX	20	10	2.00	10	30	0	4	16.00	3.50	3.48	0.00
XXVIVV	21	11	1.91	10	32	0	4	17.10	3.46	2.51	0.01
YVXXFV	21	11	1.91	10	32	0	4	9.90	3.36	3.27	0.00
FXIVXV	22	12	1.83	10	34	0	4	15.70	3.48	2.07	0.02
YXIICX	19	11	1.73	8	30	0	3	10.20	3.43	2.15	0.02
YXVXIV	20	12	1.67	8	32	0	4	11.60	3.36	2.70	0.00
YVXXVF	20	12	1.67	8	32	0	4	9.90	3.36	1.77	0.04
XIVVXI	23	14	1.64	9	37	0	4	17.40	3.48	1.89	0.03
FVIVXX	22	14	1.57	8	36	0	4	15.70	3.48	2.28	0.01
XVVFVX	18	12	1.50	6	30	0	4	15.40	3.46	1.39	0.08
FXVXFI	19	13	1.46	6	32	0	4	14.30	3.50	2.16	0.02
YXIXVC	20	14	1.43	6	34	0	3	9.90	3.41	1.24	0.11