Interdisciplinary Research (IDR) Origination Awards - Cover Page

Project Title: Development of a high-throughput antimicrobial peptide discovery toolkit

Principal Investigator(s):

Name (PI listed first)	*Department*	*College*
Joel Griffitts (PI)	Microbiology and Molecular Biology	Life Sciences
Perry Ridge	Biology	Life Sciences
JC Price	Chemistry and Biochemistry	Computational, Mathematical, and Physical Sciences

Track: 2

Abstract:
Antimicrobial resistance is a growing global health problem as bacterial pathogens evolve beyond the reach of existing antibiotics. In the search for new antimicrobial agents, the natural reserve of compounds is largely exhausted, and synthetic chemical libraries have yielded only rare successes. We propose an innovative strategy for discovering antibacterial peptides by engineering *Escherichia coli* cells to both produce and test candidate molecules. Our plan departs from most other research on antimicrobial peptides in that we will identify novel antimicrobial peptides that behave like small-molecule drugs, by focusing on those that are small (8-10 amino acids in length) and by coaxing cells to produce them as circular (cyclic) molecules with better chemical stability and bioactivity than their linear counterparts. Because the peptides we screen will be DNA-encoded and synthesized by *E. coli* cells, we can evaluate millions of distinct molecules for low cost, using next-generation sequencing to readily detect when bacteria inhibit their own growth by self-intoxication. This two-year project will establish a mature DNA-based screening system, opening avenues for expansion and more focused study to be funded externally. The work will be enabled by crucial interdisciplinary expertise in molecular genetics (Griffitts), biochemistry (Price), and computational science (Ridge), providing students with excellent opportunities for discovery at the leading edge of antimicrobial science.

Summary of Plans for External Funding
Primary potential funding source: National Institutes of Health - National Institute of Allergy and Infectious Diseases (NIAID). This institute has been a major source of funding for work on antimicrobial agents. The National Science Foundation may be compatible with a study that focuses on mechanisms of antimicrobial action on a laboratory strain of *E. coli*. Submission of grant applications is proposed for Summer of 2027 and Spring of 2028.

Project Narrative

Introduction to the antimicrobial resistance conundrum: Their genes vs. our wit

Disease-causing bacteria constitute a grand challenge to human health, and our pipeline of antimicrobial drugs is not keeping pace with their rapid evolutionary innovations (1). The classical approach to discovering antibiotics is to adopt molecules from nature: Penicillin, streptomycin, erythromycin, tetracycline, and many others, are derived from such efforts, but the well of naturally occurring antibiotics has nearly run dry, and contemporary discovery pipelines rely largely on synthetic chemical libraries for identifying lead compounds. In a rare example of the success of this approach, the antibiotic Zosurabalpin (ZAB; Fig. 1), currently in clinical trials, was initially discovered from a synthetic library of 45,000 cyclic peptides. ZAB inhibits growth of the feared hospital pathogen *Acinetobacter baumannii* via a novel mechanism of action (2,3). This promising development, however, is only a drop in a large bucket of our antibiotic discovery deficit.

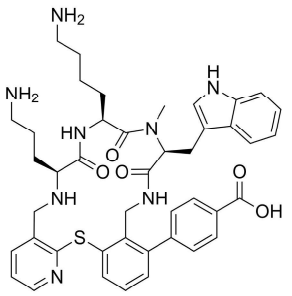

Fig 1. ZAB, a cyclic peptide antibiotic in the clinical pipeline.

Future antimicrobial therapies will emerge from an enormous mobilization of numerous and innovative research teams from industry and academia. **We as a team of interdisciplinary scientists at BYU are dedicated to alleviating the worldwide problem of antimicrobial resistance by giving our undergraduate and graduate students meaningful roles in the discovery of novel antimicrobial compounds.** We are going to do this in an unusual way: Rather than testing synthetic libraries of chemicals on target bacteria, **we will engineer bacteria to both synthesize and test the compounds.** Our strategy, which requires expertise in genetics (Dr. Griffitts), biochemistry (Dr. Price), and computation (Dr. Ridge), will allow for many millions of distinct molecules to be tested. For reasons outlined below, **we will specifically pursue antibacterial compounds in the peptide class of molecules.**

The appealing attributes of DNA-encoded peptide libraries

Peptides rule the biological world. They are polymers possessing a regular “backbone” pattern of atoms: N-C_1-C_2-N-C_1-C_2…, with each N-C_1-C_2 unit representing an amino acid. Linked to each C_1 atom is a variable side chain. In nature, there are 20 different side chains, and these define the 20 naturally occurring amino acids, which may be polymerized in any order. Consequently, consider the astronomical number of possible peptide molecules: For a rather short peptide of length10 amino acids (aa), there are 20^{10} (~10 trillion) possible sequences. Of these, it is likely that at least tens of billions would have antimicrobial activity, and that most of these have never existed in nature. Peptides in this size range thus constitute an exciting realm of chemical possibilities for antimicrobial discovery.

In the biological world, peptides are the products of genes (DNA). The peptide-synthesizing machinery of living cells reads DNA sequence in a linear, directional fashion, translating each DNA triplet (“codon”) into a single amino acid unit along the peptide polymer. For example, the DNA sequence ATG-GGT-CAT-CCT-TCA-CGT, which is 18 base pairs (bp) in length, encodes the 6-aa peptide sequence M-G-H-P-S-R. Short lengths of DNA (10-100 bp) are very cheap to synthesize, and when introduced into living bacterial cells, they live on indefinitely with the cells,

poised to produce whichever peptides they encode. DNA synthesis companies can easily randomize the DNA letters, and so for around $20, a researcher can obtain a mixed population of trillions of different DNA sequences. If cells are programmed with synthetic DNA encoding a wide variety of random peptides, those that are antimicrobial will reveal themselves by virtue of inhibiting the cells assigned to produce them. We will refer to this strategy as **"screening by self-inhibition" or SSI**: We give each cell one random gene to translate into peptides, and if the cell dies, we know that gene encodes a promising antimicrobial lead (see Fig. 2).

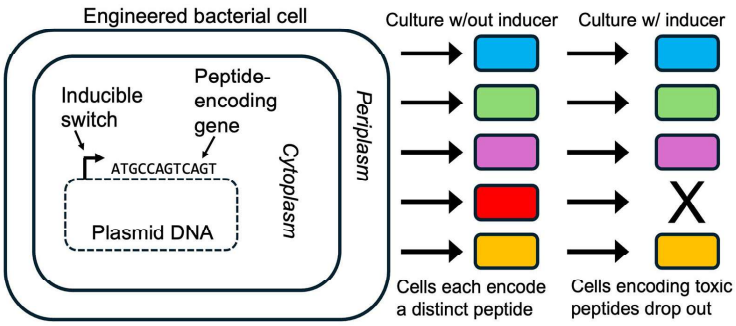

Fig 2. Illustration of the SSI approach. See text (left) for details.

SSI is off to a good start, with much unrealized potential

We are not the first to propose peptide-SSI using genetically tractable bacteria, but only a few papers have been published on the topic over the past 8 years (4,5,6). The most cited papers come from the lab of Bryan Davies at UT-Austin. Their 2018 paper in *Cell* reported a screen in which they could quantify the toxicity of 800,000 distinct peptides, identifying hundreds with promising antimicrobial activity (4). Their system was used to study peptides that are 20+ aa in length, but it has not been used to explore smaller ZAB-sized molecules (see Fig. 1).

Prior to 2018, the Griffitts and Ridge labs collaborated on a proof-of concept SSI screen using *E. coli*. Random 8-aa linear peptides were encoded on plasmid DNA molecules, and plasmids were introduced into *E. coli*, forming a population representing 200,000 different peptides (one peptide type for a given cell). We designed the system such that cells express their assigned peptide only when we add an "inducer" chemical to the growth medium to switch on expression. One copy of the population could thus be grown *without inducer* for assessment of all potential peptides in the library, and another copy could be grown *with inducer*, thus causing any cells encoding toxic peptides to die off and exit the population (see Fig. 2). Comparative deep DNA sequencing of both populations **allowed us to discern hundreds of peptides that appear to cause bacterial death or cessation of growth.** This prototype experiment provided justification for continuing the project, but lack of funding and competing research priorities prevented it.

We describe here a plan to develop an updated SSI strategy for identifying antibacterial peptides from massive DNA-encoded libraries. We will briefly describe our innovative DNA-based peptide expression strategy, as well as unique attributes of the interdisciplinary team that will make the work possible. **The proposed 2-year project will support development and initial deployment of the discovery pipeline (a short-term goal), with future external funding enabling more scaled-up screening and comprehensive testing of lead molecules (long-term goals).**

Aim 1: Building DNA tools for expressing linear and cyclic peptides (Griffitts and Price)

To make large *E. coli*-based peptide expression libraries, a robust inducible genetic platform is being built and tested. DNA libraries are best maintained in *E. coli* as circular molecules called plasmids. Students in the Griffitts lab are creating plasmids that facilitate peptide expression that

only occurs when the benign chemical, salicylate (Sal), is present in the growth medium. This feature is indicated as the "inducible switch" in Fig. 2. The *E. coli* cell has two major compartments: the cytoplasm and the periplasm (see Fig. 2). We want to be able to send peptides to either compartment, as each contains bacterial machinery that may be targeted by inhibitory peptides. While the cytoplasm is the default localization of expressed peptides, inclusion in the plasmid of a periplasmic localization sequence (PLS) will send peptides to the periplasm. **We will design and rigorously test plasmids for peptide delivery to either the cytoplasm or the periplasm.**

Biological machinery produces peptides in their linear configuration by default. While we know from experience that linear peptides may display good antimicrobial activity, **we desire to make peptides in cyclic (circular) form to increase rigidity and stability against cellular peptide-degrading enzymes.** Cyclic peptides will likely exhibit more drug-like properties in terms of cell permeability and half-life in the presence of degrading enzymes (7,8). We envision two ways to approach the difficult challenge of *in vivo* peptide cyclization, and these are summarized in Fig. 3. First, the *E. coli* periplasm naturally contains machinery that will cyclize peptides if they contain two cysteine amino acids (via disulfide bond formation); this is a useful but less stable side-chain mediated cyclization strategy. Second, there is a system drawn from cyanobacteria that enables very stable backbone-level cyclization, called split-intein cyclization. For this to work, a peptide chain needs to be flanked by two cyclization modules (each ~80 aa in length) that catalyze cyclization of the embedded peptide within the cytoplasm (9,10). While the genetic constructions for peptide cyclization are easily accomplished in the Griffitts lab, essential biochemical analyses to confirm appearance of cyclic peptides will require expertise from the Price lab.

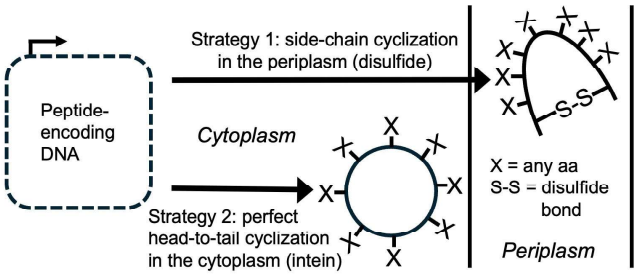

Fig 3. Diagram of our *in vivo* peptide cyclization strategies.

Aim 2: SSI peptide discovery trials with linear and cyclic peptides (Griffitts and Ridge)

Aim 1 above consisted of constructing and testing at least four peptide expression plasmids: cytoplasmic/linear, periplasmic/linear, cytoplasmic/cyclic, and periplasmic/cyclic. This second aim focuses on building actual molecular libraries with these plasmids and using these libraries to make initial antimicrobial peptide discoveries. It is within this aim that the computational expertise of our third faculty participant, Dr. Ridge, will be critical.

We will build our first SSI libraries with our simpler linear-peptide platforms. As an example, we will outline how library construction using our periplasmic expression platform would look: To encode random 8-aa periplasmic peptides, random DNA sequences 24-bp in length will be placed to the right of the Sal-inducible promoter and the periplasmic localization sequence in a population of billions of plasmid molecules, such that each molecule encodes a distinct peptide (see Fig. 4).

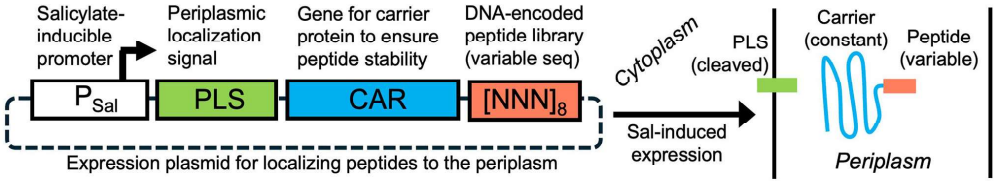

Fig 4. Example of how a plasmid would be engineered to send peptides to the periplasm.

These plasmid molecules will then be introduced into *E. coli* cells by electroporation, with each transformed cell receiving one plasmid molecule. Cells can then be selectively grown as a large population representing around 1 million distinct peptides. Because the Sal-inducible promoter is not activated at this point, every cell is viable; but after addition of Sal, when every cell actively produces its assigned peptide in the periplasm, the small fraction of cells producing antimicrobial peptides will drop out of the population. At this point, the peptide-encoding DNA from remaining viable cells will be extracted and sequenced. This enables us to infer all peptide sequences from the library that *are not toxic* when expressed. By comparing these with DNA from before induction, we can infer the peptide sequences that *are toxic*. A simple set of mock data are given in Fig. 5 to illustrate this analysis. This induced/uninduced comparison involving millions of DNA sequences requires sophisticated computation, critically supported by our collaboration with the Ridge group.

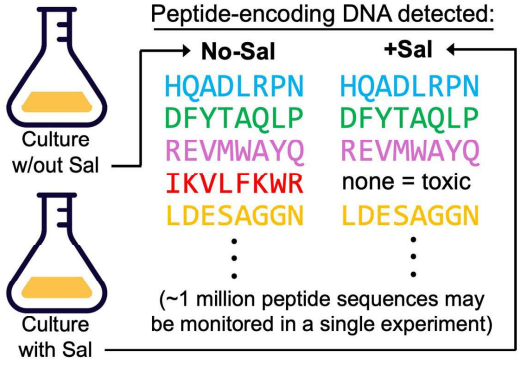

Fig 5. Sequence-based analysis to detect antimicrobial peptides

As these initial SSI studies using linearly configured peptides near completion, we will be prepared to carry out a similar process using plasmids arranged to express random 8-aa peptides as cyclic molecules. With prototype screens performed on just a few million peptide sequences, we are sure to identify hundreds of strong antimicrobial peptide candidates. **In fact, in recent preliminary work where students hand-screened bacterial colonies representing only a few hundred different linear peptide sequences, several robust hits emerged.**

Aim 3: Testing effects of antimicrobial peptide candidates when exogenously applied

Many peptides that inhibit bacterial growth when synthesized within the cells (identified in Aim 2) will likely not prove effective when synthesized chemically and then applied to cells from the outside. As a final aim of this 2-year project, **we will test a subset of the antimicrobial peptide sequences inferred from our SSI studies for their efficacy as synthetic molecules applied exogenously.** Synthesis of short linear peptides is rather routine and inexpensive, while cyclic peptide synthesis is more complicated. Our plan will be to have approximately 20 linear peptides (inexpensive), 20 C-C cyclized peptides (inexpensive), and 5 head-to-tail cyclized peptides (expensive, but most promising) synthesized. These 45 synthetic peptides will be tested on *E. coli* cells in small-scale culture to determine how frequently cell permeability presents a significant barrier to efficacy. Based on a recent publication in which a low-throughput SSI approach was used to identify lead peptides, only three candidates were chemically synthesized and one of them exhibited strong antimicrobial activity when applied exogenously (6). Thus, we are confident that from our ~45 candidates, several will exhibit exogenous activity.

Summary of team member contributions and timeline

Aims 1 and 2 above will overlap and occupy most of the 2-year project (see timeline, below), with Aim 3 occupying the final few months. **The mass spectrometry-based analytical expertise in the Price lab** will be vital for determining whether the DNA-encoded expression platforms built as part of Aim 1 in fact send peptides to the intended cellular locations (cytoplasm vs. periplasm)

in the right configuration (linear vs. cyclic). This same expertise will be important for characterizing peptide candidates emerging in Aim 2 and for confirming synthetic peptide structures in Aim 3. **The development of bioinformatic pipelines for big-data analysis (Aim 2) will be the stewarded by the Ridge lab.** Computational tool development will begin in mid-2026, and refinements will continue throughout the two-year project. **The Griffitts lab (lead) will be responsible for design and production of the cell-based peptide expression platforms.** At least six Griffitts lab members will be responsible for DNA engineering (Aims 1 and 2), cell culture and DNA sequencing (Aim 2), and exogenous testing of peptide candidates (Aim 3). **The three principal investigators possess technical expertise that is diverse and exceptionally complementary for accomplishing this interdisciplinary project. Dr. Griffitts will be responsible for scheduling monthly team meetings to facilitate communication and strategic planning.**

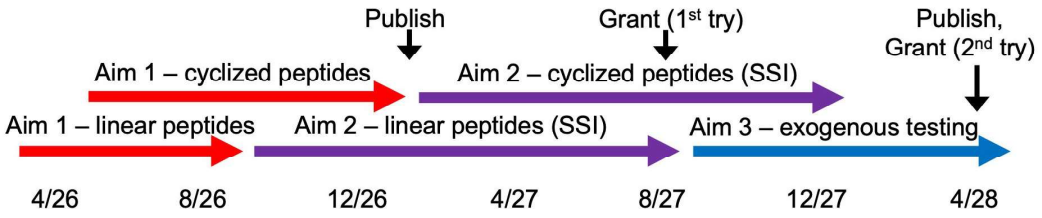

Fig 6. Project timeline and milestones

Measures of success and longer-term objectives

Tracking of project success will be straightforward (see Fig. 6 above): First, we expect publication reporting the ensemble of peptide expression plasmids near the end of the first year, which will clearly indicate completion of Aim 1. A second publication at the end of Year 2 will report a large compendium of linear and cyclic antimicrobial peptides identified using our high-throughput approach. A crucial indicator of project success will be submission of a proposal for external funding toward the start of the second year, and (as a second attempt) at the conclusion of the second year. Another indicator of project success will be measured by our fidelity to a monthly team meeting schedule, which will allow continual communication and strategic planning. A final indicator of project success will be the number of students consequentially involved in the scientific work. The number of student project participants, and the number of students with authorship on peer-reviewed publications, will be important signifiers of the difference made for BYU students.

We anticipate this two-year "germination" period will be just the beginning of a 7-10 year externally funded antimicrobial discovery project that oversees very large-scale screening of cyclic peptides—hundreds of millions of molecular species—generating sufficiently large datasets that can leverage powerful AI-based algorithms to predict antimicrobial peptide sequences and structures not found in the limited peptide sequence space accessible to experimental testing. **We believe the project will deliver lead molecules that pave the way for entirely new antibiotic classes targeting diverse bacterial processes, significantly impacting human health and well-being in the long term.**