Novel antimicrobial peptide discovery using machine learning and biophysical selection of minimal bacteriocin domains

Francisco R. Fields[1,2,3], Stefan D. Freed[1,2,3], Katelyn E. Carothers[1,2], Md Nafiz Hamid[4,5], Daniel E. Hammers[1,2], Jessica N. Ross[1,2], Veronica R. Kalwajtys[1], Alejandro J. Gonzalez[1], Andrew D. Hildreth[1], Iddo Friedberg[4,5], Shaun W. Lee*[,1,2,3]

[1]University of Notre Dame. Department of Biological Sciences, Notre Dame, IN 46556

[2]Eck Institute for Global Health, Notre Dame, IN 46556

[3]Chemistry-Biochemistry-Biology Interface Program, Notre Dame, IN 46556

[4]Iowa State University. Department of Veterinary Microbiology and Preventive Medicine, Ames, IA 50011

[5]Bioinformatics and Computational Biology Program, Ames, IA 50011

*Corresponding author E-mail: Shaun.W.Lee.310@nd.edu

First author E-mail: ffields@nd.edu

Abstract

Bacteriocins are ribosomally produced antimicrobial peptides that represent an untapped source of promising antibiotic alternatives. However, inherent challenges in isolation and identification of natural bacteriocins in substantial yield have limited their potential use as viable antimicrobial compounds. In this study, we have developed an overall pipeline for bacteriocin-derived compound design and testing that combines sequence-free prediction of bacteriocins using a machine-learning algorithm and a simple biophysical trait filter to generate minimal 20 amino acid peptide candidates that can be readily synthesized and evaluated for activity. We generated 28,895 total 20-mer peptides and scored them for charge, α-helicity, and hydrophobic moment, allowing us to identify putative peptide sequences with the highest potential for interaction and activity against bacterial membranes. Of those, we selected sixteen sequences for synthesis and further study, and evaluated their antimicrobial, cytotoxicity, and hemolytic activities. We show that bacteriocin-based peptides with the overall highest scores for our biophysical parameters exhibited significant antimicrobial activity against *E. coli* and *P. aeruginosa*. Our combined method incorporates machine learning and biophysical-based minimal region determination, to create an original approach to rapidly discover novel bacteriocin candidates amenable to rapid synthesis and evaluation for therapeutic use.

Introduction

Many bacteria have become resistant to conventional antibiotics, necessitating the discovery of novel antimicrobial compounds[1]. However, pharmaceutical antibiotic development has declined chiefly due to brief usability window of existing antibiotic scaffolds [2]. To combat the lack of novel antimicrobial discovery, many bioinformatic approaches have been developed to mine the genomes of bacteria for natural products[3]. One promising class of natural products are bacteriocins, the ribosomally produced antimicrobial peptides of bacteria[4,5]. These chemically and functionally diverse peptides are divided into two main classes. The class I bacteriocins include extensive post-translational modifications in their final form; for example, nisin is a commonly studied bacteriocin whose features include post-translational modifications such as lanthionine and methyllanthionine[6]. Enterocin AS-48, another class I bacteriocin, undergoes head-to-tail circularization[7]. The class II bacteriocins primarily consist of peptides that do not undergo post-translational modification[4]. Bacteriocins are often located in genetic clusters containing the structural gene encoding the precursor peptide, as well as the context genes necessary for bacteriocin maturation, export, and immunity. The bacteriocin structural gene is often expressed as a prepropeptide, consisting of the unmodified bacteriocin functional domain and an N-terminal leader sequence. Upon installation of the post-translational modifications and cleavage of the leader peptide, the mature peptide is exported by an ABC-cassette type transporter[4,5,8].

Genome mining approaches have taken advantage of the bacteriocin operon-like structure to identify novel bacteriocin candidates through two approaches: identification of the bacteriocin precursor gene or identification of bacteriocin context genes[3,8–10]. Online genome mining tools, such as BAGEL, and bacteriocin databases, such as BACTIBASE, allow

investigators to identify and classify putative bacteriocins based on their homology to other known bacteriocin genes[9,10]. Similar tools, such as anti-SMASH, have been expanded to not only identify putative bacteriocins, but also secondary metabolites and other genetically identifiable antibiotics[3,8,11,12].

Large sequence heterogeneity and a small number of experimentally determined bacteriocins, as well as the small size of most structural genes (30-150aa) have presented challenges in identifying novel bacteriocins using BLAST and other sequence similarity approaches [8]. To overcome these problems, some bacteriocin prediction software identify novel bacteriocins by searching for conserved context genes of the bacteriocin operon[8,13]. The bacteriocin operon and gene block associator (BOA) identifies context genes through homology-based genome searches[8]. BOA has identified 95% of BAGEL annotated bacteriocins in addition to 1,033 putative bacteriocins not identified by BAGEL. ClusterFinder, another context gene based approach, has been used to mine the genomes of human commensal organisms. This approach led to the identification of the novel thiopeptide bacteriocin lactocillin[13]. Another tool, MetaRiPPquest, connects genomic bacteriocin predictions to tandem mass spectrometry data. Peptidogenomic approaches attempt to bridge the gap between computational and *in vitro* identification[14]. While context-based approaches seem to circumvent the need for sequence similarity, novel methods that move away from homology-based mining tools are still needed. Recently, *k*-mer based machine learning approaches have been used successfully to classify protein sequences without the need for homology[15,16].

Regardless of the genome mining approach, *in vitro* verification of the antimicrobial activity of computationally identified putative bacteriocins also remains a major challenge due to several factors. First, bacteriocins have diverse mechanisms of action with most having

specific targets within or on host cells. This is especially true for the class I bacteriocins; for example, microcin B17 (MccB17) inhibits the activity of DNA gyrase while nisin inhibits peptidoglycan synthesis by binding to lipid II[17–19]. Even class II bacteriocins can have extremely specific targets; for example, lactococcin A targets the mannose phosphotransferase system to induce pore formation[5,20]. Secondly, bacteriocins may exhibit a very narrow spectrum of activity and be highly specific against a competitor strain. Many bacteriocins produced by lactic acid bacteria will only kill other closely related species such as *Lactobacillus*, *Enterococcus*, and *Listeria*[5]. Finally, some bacteriocins may not have bacterial targets. Streptolysin S (SLS) is structurally similar to MccB17 as both are thiazole-oxazole modified microcins; however, SLS is a virulence factor that promotes invasion upon Group A *Streptococcus* infection[21,22]. Other bacteriocin-like peptides may act as signaling molecules, as nisin and subtilosin have both been implicated as autoregulators acting as autocrine signaling peptides at distinct concentrations [11,19]. However, some bacteriocins target the bacterial membrane in a non-specific fashion through electrostatic and hydrophobic interactions, including enterocin AS-48 and sakacin[5,7,23].

Recently, we have shown that membrane targeting bacteriocins can serve as templates for the efficient design of synthetic antimicrobial peptides[24]. Using the AS-48 homologue, safencin AS-48, we created a synthetic peptide corresponding to the membrane interacting region of enterocin AS-48[7,25]. This region is a cationic, hydrophobic, alpha-helical peptide that abstracts the full-length 70 amino acid bacteriocin to a 25 amino acid peptide. Interestingly, these biophysical qualities are very similar to synthetic antimicrobial peptides derived from eukaryotic sources, whose activity relies on their overall positive charge, conformation, and amphipathicity[26–29]. To determine if these biophysical guidelines of antimicrobial peptide design could select for regions of putative membrane-targeting bacteriocins, we wrote a script to scan

for 20 amino acid stretches at a time along the length of a putative bacteriocin and score each 20-mer for charge, alpha-helical propensity, and hydrophobic moment. Upon chemical synthesis and antimicrobial testing of a set of these 20-mers, we observed that peptide candidates with the highest scores in all three categories exhibited significant antimicrobial activity. This approach represents a method by which membrane-targeting regions of putative bacteriocins can be rapidly selected, synthesized, and verified *in vitro*. We propose that peptides discovered through this process could then serve as scaffolds for subsequent optimization and eventual therapeutic development.

Materials and Methods

Initial selection of candidate bacteriocins

We selected an initial set of putative novel bacteriocins using a word embedding algorithm, Word2vec, as described previously[16]. Briefly, we created a vocabulary of all possible 8,000 amino-acid trimers. Each trimer is then represented as a vector, which captures the probabilities of that trimer being in the neighborhood of other trimers, also known as the skip-gram model[15]. Each protein sequence was then represented as the sum of vectors representing the trimers comprising the protein. We then trained several supervised learning models with a positive set of 346 known bacteriocins and a negative set of the same size. The best performing method, and support vector machine (SVM) was then used to discover the set of 676 putative bacteriocins used in this study. In essence, the machine learning algorithm employed thus generates a list of new bacteriocin-like sequences that preserve key evolved features of natural bacteriocins products. We eliminated from the list all known bacteriocins which were discovered using BLAST against GenBank with an e-value of 10^{-3} or less, and which were

125 annotated as bacteriocins. The result was a set of 676 putative bacteriocins, not obviously
126 homologous, by sequence similarity, to existing bacteriocins.

127 ***Biophysical selection of 20-mer peptides***

128 Using a sliding window, we generated 28,895 20-mers from the 676 predicted peptides,
129 and calculated the following biophysical parameters for each 20-mer candidate: (1)Charge, (2)
130 Helicity, and (3) Hydrophobic moment (Figure 1)[30–32]. Net charge was calculated as a sum of
131 the charge for each amino acid at pH 7. Helicity was calculated as a sum of the Chou-Fasman
132 probabilities of each amino acid. Finally, the hydrophobic moment was calculated using the
133 hydrophobicity values for each residue assuming that the 20-mer peptides would adopt an alpha
134 helical structure. For each of the biophysical parameters, the 20-mers were ranked as high,
135 middle, or low based on the range of scores within that parameter (Figure 2A).

136 ***Peptide Selection and Synthesis***

137 To evaluate our biophysical parameter scores, we selected a total of sixteen 20-mer
138 peptides for synthesis and further experimentation. Peptides were synthesized by Genscript
139 (Piscataway, NJ), to >95% purity and verified by HPLC and mass spectrometry. All peptides
140 were dissolved in DMSO for subsequent experimentation (Thermo Fischer).

141 ***Bacteria and Growth Conditions***

142 *E. coli* BL-21 (Thermo Fischer) and *P. aeruginosa* PAO1 (gift from J. Shrout at the
143 University of Notre Dame) were grown in LB broth Miller (EMD chemicals, Gibbstown NJ).
144 *Staphylococcus aureus* USA300 was grown in Todd Hewitt broth (Neogen Corporation,
145 Lansing, MI). All cultures were grown at 37 °C.

146 ***Antimicrobial Activity Assays***

147 Minimal inhibitory concentrations (MICs) of the 20-mer peptides were determined via
148 microtiter dilution assay[33]. Briefly, dilute bacterial cultures were added to a series of serial two-
149 fold dilutions of peptide in Mueller-Hinton broth (Thermo Fischer). The lowest concentration at
150 which no bacterial growth was observed after overnight incubation at 37 °C was defined as the
151 MIC. If an MIC could be determined, cultures from the MIC experiment were plated and
152 incubated overnight at 37 °C. The concentration at which no colonies were visible after
153 overnight incubation was defined as the minimal bactericidal concentration (MBC).

154 ***Antibiofilm Formation Assays***

155 Antibiofilm activity of the peptides were assessed using USA 300 and PAO1. For
156 USA300 biofilms, overnight cultures grown in TSB (Sigma-Aldrich) were diluted 1:100 in TSB
157 .1% glucose 1% NaCl with or without peptide[34]. For PAO1 biofilms, overnight cultures grown
158 in LB were diluted 1:100 in M63 1mM MgSO4 and .4% arginine with or without peptide[35].
159 Samples were incubated for 24 hours in a microplate. Planktonic cells were removed from the
160 wells and the biofilms were washed three times with ddH_2O. Biofilms were then stained with
161 .1% crystal violet, washed three times with ddH_2O, and resuspended in 30% acetic acid[36]. These
162 were then quantified by OD 550 reading on an Synergy Microplate Reader (Biotek).

163 ***Peptide Cytotoxicity Assays***

164 Eukaryotic cytotoxicity was determined by ethidium homodimer and hemolysis assays.
165 Ethidium homodimer assays were carried out with HaCaT cells in 24 well culture dishes grown
166 to 90% confluency. Medium was aspirated, and cells were washed with PBS (Thermo Fischer).
167 Peptide in fresh DMEM (Dibco) was added to the cells at the desired concentration. Cells were
168 incubated with peptide for 16 hours. Medium was aspirated, and cells were washed with PBS.
169 Cells were incubated in 4 μM ethidium homodimer (Molecular Probes) in PBS for 30 minutes.

Fluorescence was determined by 528 excitation and 617 nm emission and a cutoff value of 590 nm. Saponin (.1%) was then added to each well and incubated for 20 minutes. The fluorescence was read again. Percent membrane permeabilization was determined by dividing the initial fluorescence by the second fluorescence reading. For hemolysis assays, 100 μL of sheep red blood cells (RBCs) were washed 3 times in cold PBS. Washed cells were resuspended in 25 ml of PBS. Triton, PBS, or peptide in 10% DMSO/PBS were added to 180 μL of resuspended RBCs and incubated at 37°C for 1 hour. Samples were read at 450 nm. Data was expressed as percent hemolysis by relativizing to the Triton and PBS controls.

Results

Design and biophysical selection of 20-mer minimal bacteriocins

From the initial set of 676 putative novel bacteriocins using the word embedding algorithm, Word2vec, 28,895 total 20-mer bacteriocin peptide candidates were generated (Figure 1). Each peptide was then assigned a low, middle, or high ranking for each of the biophysical parameters based on the range of scores within that parameter (Figure 2A). For example, a peptide with a net charge of 5, a helical score of 17, and a μH of 900 would rank middle for charge, low for helicity, and high for μH (Figure 2A).

80% of the 20-mers received a low ranking for charge (a net positive charge between +1 and +3) while only 1% ranked high (Figure 2B). For the hydrophobic moment values, a majority of the peptides also ranked low (any hydrophobic moment value below 333) with only 5% receiving a high score (Figure 2B). However, for the helicity score, a majority of the peptides, 65%, fell into the middle range of scores between 19 and 22 with only 2% scoring high for helicity (Figure 2B). It is important to note that the hydrophobic moment and helicity scores

may not truly represent these parameters for the peptides as the propensity to form a beta sheet was not taken into consideration when calculating these values.

Peptide Selection for Chemical Synthesis

Many cationic antimicrobial peptides will adopt an amphipathic alpha helical conformation. Therefore, we reasoned that of the peptides generated by our script those ranking high in all three biophysical categories would yield the most antimicrobial activity. Of the sixteen peptides selected for synthesis, peptides 1 and 2 ranked low for all three biophysical parameters while peptides 3 and 4 ranked high for the three parameters (Table 1). The remaining 12 peptides were randomly selected from all 20-mers ranking middle in at least one category and high for the remaining parameters (Table 1).

PEP-FOLD prediction of secondary structure

To determine if our biophysical selection criteria were able to accurately predict an amphipathic alpha helical structure of the peptides selected for synthesis, we modeled their secondary structure using the PEP-FOLD online tool. For peptides 1 and 2, which received low scores for helicity and hydrophobic moment, their structures are predicted to exist as a majority random coil (Figure 3A). In contrast, peptides 3 and 4, having high scores for helicity and hydrophobic moment, are predicted to exist as fully extended alpha helices with clear clustering of the polar and charged amino acids to one side of the helix and the hydrophobic residues on the other, indicative of a strong hydrophobic moment (Figure 3B). Peptides 5 through 10 have a high helicity score; however, the modeling predictions expect unstructured regions owing to helix-breakering residues glycine and proline that occur within their sequences (Figure 3 and Table 1). All of these peptides also received middle scores for their hydrophobic moment which is visible as hydrophobic residues within the polar face of the helix, such as peptide 6, and

charged amino acids within the hydrophobic face, such as peptide 9. Interestingly, peptide 11 is predicted to exist as a beta sheet (Figure 3E). The biophysical calculator only takes into account the Chou-Fasman residue helical propensity score and does not calculate the individual likelihood of forming a beta sheet; therefore, peptides with a higher sheet propensity were not excluded from the list of peptides for synthesis. Finally, the rest of the peptides are predicted to adopt various helical structures with differing amphipathic characteristics (Figure 3).

Antimicrobial Properties of Synthetic 20-mers

The peptides were assessed for their minimal inhibitory concentration (MIC) and minimal bactericidal concentration (MBC) on *Escherichia coli, Staphylococcus aureus,* and *Pseudomonas aeruginosa* (Table 2). As expected, peptides 1 and 2, which scored low in all three biophysical parameters, did not have activity against any of the organisms tested. Peptides 3 and 4, which scored high in all three biophysical parameters, exhibited antimicrobial activity against both *E. coli* and *P. aeruginosa* (Table 2). Peptides 5, 6, and 7 scored high in charge and helicity and middle in hydrophobic moment (Table 1). Interestingly, these peptides showed a range of antimicrobial activities (Table 2). Peptide 6 was more efficient at inhibiting the growth of *P. aeruginosa* (MIC = 32 μM) than *E. coli* (MIC =128 μM). Peptides 5 and 7 were much less active than peptide 6 despite having similar values for their biophysical scores (Tables 1 and 2). Peptides 8, 9, and 10 scored high for helicity with middle scores for charge and hydrophobic moment. These peptides did not have any antimicrobial activity against the organisms tested. This overall trend continued for the rest of the peptides tested. Indeed, peptides scoring high in any one of the biophysical parameters with only middle scores for the others (peptides 8-16) did not have any antimicrobial activity. We did not test any of the peptide candidates at

237 concentrations above 128μM, so biological activities at higher concentrations cannot be ruled
238 out.

239 ***Inhibition of Biofilm Formation by the Synthetic 20-mers***

240 Despite not having a true MIC, we observed that peptides 11 and 16 were able to
241 significantly reduce the overnight growth of *S. aureus* cultures (Supplementary Figure 1A-B).
242 To investigate if these peptides were exerting antibiofilm effects, we employed the biofilm
243 formation assay. Upon incubation with peptide 11 for 24 hours in biofilm inducing media, we
244 observed a significant decrease in USA 300 biofilm formation down to a concentration of 4 μM
245 (Supplementary Figure 1C). This trend was also observed for peptide 16; however, this only
246 inhibited biofilm formation down to 16 μM (Supplementary Figure 1D) Finally, to determine if
247 these peptides could inhibit the biofilms of other bacteria we used *P. aeruginosa*. Peptides 11
248 and 16 exhibited no bacteriostatic effects on PAO1 (Supplementary Figure 2A-B). However,
249 these peptides exerted mild antibiofilm formation activity down to 16 μM (Supplementary
250 Figure 2C-D). In addition to identifying peptides with potent antimicrobial activity, we have
251 also identified peptides with antibiofilm activity.

252 ***Peptide mammalian cell cytotoxicity***

253 To determine if our biophysical parameters were able to select for peptides with affinity
254 for bacterial membranes instead of mammalian membranes, we assessed their ability to
255 compromise the membranes of erythrocytes and keratinocytes. Fourteen of the peptides
256 exhibited no hemolytic activity even at high concentrations (Supplementary Table 1). However,
257 peptides 2 and 10 exhibited increased levels of hemolysis at only the highest concentrations
258 (128 μM). Cytotoxicity to keratinocytes was interrogated using the ethidium homodimer assay.
259 We observed that all of the peptides were unable to cause cell death when incubated with

HaCaT cells for 16 hours at the highest concentrations. Together, these data indicate that these peptides generally do not target mammalian membranes.

Discussion

Bacteriocins are a barely-tapped source of highly diverse antimicrobials. However, verifying the antimicrobial activity of putative bacteriocins can be difficult due to the potentially narrow activity spectra and highly diverse mechanisms[4,37]. Additionally, traditional methods of natural bacteriocin isolation as well as heterologous expression strategies are complicated by purification limitations and low yield[38–40]. Here we describe a complete strategy by which *de novo* mining of bacteriocins can be parsed using a biophysical algorithm to identify minimally active bacteriocin peptide candidates. Biophysical selection was done by focusing on three parameters that have been implicated in the activity of membrane active antimicrobial peptides: helicity, charge, and amphipathicity[30–32,41–43]. Our strategy for employing predictive algorithms with biophysical selection and minimal domain candidate design allows for the development of completely novel, highly active, synthetic bacteriocins that have wide applicability as antimicrobial compounds. Previous studies have shown that synthetic peptide variants of full length bacteriocins can be used to approximate their antimicrobial function. For example, linear variants of enterocin AS-48, a circular bacteriocin consisting of five alpha helices, have been shown to retain some of the antimicrobial activity of the parent[44,45]. The antimicrobial action was shown to be dependent upon the cationic and hydrophobic residues present within helices four and five that are designated as the membrane-interacting region[25]. Recently, we published a strategy whereby the membrane-interacting region in an AS-48 homologue was used as a template to create a series of small, optimized antimicrobial peptides[24]. This establishes a precedent by which synthetic peptides can be used to approximate the activity of the full length

283 bacteriocin. We have built upon these previous studies by utilizing the biophysical parameters
284 of synthetic antimicrobial peptide design to select for membrane interacting regions of putative
285 bacteriocins[25]. We observed that peptides with the highest scores for the biophysical parameters
286 of charge, helicity, and hydrophobic moment were the most active against the bacteria tested
287 (Table 2). Interestingly, the only two peptides to meet these criteria were from the same putative
288 bacteriocin. It is therefore highly likely that this putative bacteriocin works in a membrane
289 active manner [29,42,43]. The interpretation of these data becomes confounded for the peptides
290 whose biophysical parameters begin to receive middle scores. For example, peptide 6, with a
291 middle score for hydrophobic moment, is a more effective antimicrobial against *P. aeruginosa*,
292 MIC = 32, than *E. coli*, MIC = 128. This observation is in contrast to the activities of the high
293 scoring peptides, 3 and 4, whose antimicrobial activities were higher against *E. coli*. Therefore,
294 it may be possible to tune antimicrobial *specificity* by modifying the biophysical scores[46,47].
295 While most research has focused on modification of these parameters and their effects on
296 eukaryotic cytotoxicity and overall antimicrobial activity few have examined how these
297 parameters tune the specificity of these compounds to specific bacteria[48,49].
298 There are some drawbacks to this approach. While it seems that our approach has selected for
299 antimicrobial regions of putative bacteriocins, it is also possible that using a minimal synthetic
300 peptide strategy has decoupled the function of the synthetic bacteriocin from the function of the
301 full sequence. Enterocin AS-48 undergoes dimer formation and then subsequent tertiary
302 structural changes before inserting itself into the membrane of target bacteria[50]. However,
303 synthetic AS-48 peptides lose this ability to dimerize and work in a mechanism more akin to
304 carpet or pore models of synthetic antimicrobial peptide activity[25]. Therefore, some of the
305 antimicrobial function and specificity inherent in bacteriocins will be lost by utilizing synthetic

306 minimal versions. Finally, our approach cannot verify the activity of bacteriocins which do not
307 target the bacterial membrane or whose biophysical characteristics change upon post-
308 translational modification[4,5,11].
309 Despite these drawbacks, the techniques described herein have potential for linking *de novo*
310 computational bacteriocin discovery with immediate therapeutic development. With the
311 increasing amount of computational work being done to predict novel antimicrobial compounds
312 there is a mounting need to verify their antimicrobial activity *in vitro*[8–10]. Our method validates
313 the use of machine learning algorithms to further mine genomic information for potential
314 bacteriocins candidates that can be refined using biophysical scripting parameters and size
315 optimization for rapid synthesis and testing. The lack of mammalian cell cytotoxicity in our
316 synthesized peptide set indicates that selecting minimal bacteriocin candidates based on the
317 specific set of biophysical parameters that we have established will select for candidates that
318 specifically target bacterial membranes, a highly valuable outcome from our studies (Table 2
319 and Supplementary Table 1). Many current synthetic antimicrobial peptides used to treat human
320 disease have been built around an existing scaffold from eukaryotes[51]. Omiganan, derived from
321 magainin of the African three-toed frog, is currently being developed as a topical antimicrobial
322 for the treatment of diabetic foot ulcers[52]. In contrast, relatively few bacteriocins have been
323 developed for the treatment of disease[51,53,54]. Our strategy to combine machine learning
324 algorithms for *de novo* bacteriocin discovery along with biophysical refinement and minimal
325 design represent a particularly robust workflow for the development of new antibiotic
326 compounds. These synthetic bacteriocin scaffolds could be further refined via iterative testing
327 and data collection for efficacy and selectivity.
328

329 **Acknowledgements**

330 IF was funded, in part, by NSF awards ABI-1551363 and ABI-1458359. SL was funded by NIH

331 Innovator Award 1DP2 OD008468-01 and an Eck Institute for Global Health Pilot Project

332 Program. FF is supported by an NSF-GRFP fellowship and a National GEM Consortium

333 Fellowship.

Figure and Table Legends

Figure 1: Overall strategy for selection of bacteriocins for synthesis. Machine learning set of 676 putative bacteriocins was used to generate overlapping 20-mer peptide candidates. 28,895 20-mers were scored and ranked for charge, helicity, and hydrophobic moment. A representative sample of 16 peptides were selected for synthesis and in vitro characterization in this study.

Figure 2: Scoring breakdown of biophysical computational parameters of the candidate peptides. **A**. Peptides were divided into high (grey), middle (orange), and low (blue) groups based on their charge, helicity, and hydrophobic moment scores. **B**. Most of the peptides scored low to middle with only a small percentage scoring high for each of the biophysical parameters.

Figure 3: PEP-FOLD models of the peptides selected for synthesis **A.** peptides 1 and 2, **B.** 3 and 4, **C.** 5,6, and 7, **D.** 8, 9, and 10, **E.** 11, 12, and 13, and **F.** 14, 15, and 16. Basic, acidic, and hydrophobic residues are in blue, red, and orange respectively.

Table 1: 20-mers selected for synthesis and their corresponding biophysical scores.

Table 2: MICs and MBCs of the synthetic 20-mer bacteriocins against *S. aureus*, *E. coli*, and *P. aeruginosa.*

References

1. CDC. *Antibiotic resistance threats in the United States, 2013*. (2013). doi:CS239559-B

2. Martens, E. & Demain, A. L. The antibiotic resistance crisis, with a focus on the United States. *J. Antibiot. (Tokyo).* **70,** 520–526 (2017).

3. Fields, F. R., Lee, S. W. & McConnell, M. J. Using bacterial genomes and essential genes for the development of new antibiotics. *Biochem. Pharmacol.* **134,** 74–86 (2017).

4. Alvarez-Sieiro, P., Montalbán-López, M., Mu, D. & Kuipers, O. P. Bacteriocins of lactic acid bacteria: extending the family. *Appl. Microbiol. Biotechnol.* **100,** 2939–2951 (2016).

5. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins - a viable alternative to antibiotics? *Nat. Rev. Microbiol.* **11,** 95–105 (2013).

6. Field, D., Cotter, P. D., Ross, R. P. & Hill, C. Bioengineering of the model lantibiotic nisin. *Bioengineered* **5979,** 37–41 (2015).

7. Sánchez-Hidalgo, M. *et al.* AS-48 bacteriocin: Close to perfection. *Cellular and Molecular Life Sciences* **68,** 2845–2857 (2011).

8. Morton, J. T., Freed, S. D., Lee, S. W. & Friedberg, I. A large scale prediction of bacteriocin gene blocks suggests a wide functional spectrum for bacteriocins. *BMC Bioinformatics* **16,** 1–9 (2015).

9. Van Heel, A. J., De Jong, A., Montalbá N-Ló Pez, M., Kok, J. & Kuipers, O. P. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* **41,** W448–W453 (2013).

10. Hammami, R., Zouhir, A., Hamida, J. Ben & Fliss, I. BACTIBASE: a new web-accessible database for bacteriocin characterization. *BMC Microbiol.* **7,** 1–6 (2007).

11. Flaherty, R. A., Freed, S. D. & Lee, S. W. The Wide World of Ribosomally Encoded

373 Bacterial Peptides. *PLoS Pathog.* **10,** (2014).

374 12. Blin, K. *et al.* antiSMASH 2.0—a versatile platform for genome mining of secondary
375 metabolite producers. *Nucleic Acids Res.* **41,** W204–W212 (2013).

376 13. Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human
377 microbiome reveals a common family of antibiotics. *Cell* **158,** 1402–1414 (2014).

378 14. Mohimani, H. *et al.* MetaRiPPquest : A Peptidogenomics Approach for the Discovery of
379 Ribosomally Synthesized and Post- translationally Modified Peptides. *BioRxiv* (2017).
380 doi:10.1101/227504

381 15. Mikolov, T., Corrado, G., Chen, K. & Dean, J. Efficient Estimation of Word
382 Representations in Vector Space. *Arxiv* (2013). doi:10.1162/153244303322533223

383 16. Hamid, M.-N. & Friedberg, I. Bacteriocin prediction using Word Embedding with Deep
384 Recurrent Neural Networks. *BioRxiv* (2018). doi:10.1101/255505

385 17. Vizan, J. L., Hernandez-Chico, C., Del Castillo, I. & Moreno1, F. The peptide antibiotic
386 microcin Bi 7 induces double-strand cleavage of DNA mediated by E.coli DNA gyrase. *EMBO*
387 *J.* **10,** 467–476 (1991).

388 18. Field, D., Cotter, P. D., Hill, C. & Ross, R. P. Bioengineering lantibiotics for therapeutic
389 success. *Frontiers in Microbiology* **6,** 1–6 (2015).

390 19. Kleerebezem, M. Quorum sensing control of lantibiotic production; nisin and subtilin
391 autoregulate their own biosynthesis. *Peptides* **25,** 1405–1414 (2004).

392 20. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide
393 natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.*
394 **30,** 108–160 (2013).

395 21. Lee, S. W. *et al.* Discovery of a widely distributed toxin biosynthetic gene cluster. *PNAS*

105, 5879–5884 (2008).

22. Molloy, E. M., Cotter, P. D., Hill, C., Mitchell, D. A. & Ross, R. P. Streptolysin S-like virulence factors: The continuing sagA. *Nature Reviews Microbiology* **9,** 670–681 (2011).

23. Grande Burgos, M., Pulido, R., del Carmen López Aguayo, M., Gálvez, A. & Lucas, R. The Cyclic Antibacterial Peptide Enterocin AS-48: Isolation, Mode of Action, and Possible Food Applications. *Int. J. Mol. Sci.* **15,** 22706–22727 (2014).

24. Fields, F. R. *et al.* Rational design of syn-safencin, a novel linear antimicrobial peptide derived from the circular bacteriocin safencin AS-48. *J. Antibiot. (Tokyo).* (2018). doi:10.1038/s41429-018-0032-4

25. Angeles Jiménez, M., Barrachi-Saccilotto, A. C., Valdivia, E., Maqueda, M. & Rico, M. Design, NMR characterization and activity of a 21-residue peptide fragment of bacteriocin AS-48 containing its putative membrane interacting region. *J. Pept. Sci.* **11,** 29–36 (2005).

26. Thaker, H. D., Cankaya, A., Scott, R. W. & Tew, G. N. Role of amphiphilicity in the design of synthetic mimics of antimicrobial peptides with gram-negative activity. *ACS Med. Chem. Lett.* **4,** 481–485 (2013).

27. Jiang, Z. *et al.* Effects of net charge and the number of positively charged residues on the biological activity of amphipathic a-helical cationic antimicrobial peptides. *Biopolym. - Pept. Sci. Sect.* **90,** 369–383 (2008).

28. Ong, Z. Y., Wiradharma, N. & Yang, Y. Y. Strategies employed in the design and optimization of synthetic antimicrobial peptide amphiphiles with enhanced therapeutic potentials. *Advanced Drug Delivery Reviews* **78,** 28–45 (2014).

29. Fjell, C. D., Hiss, J. A., Hancock, R. E. W. & Schneider, G. Designing antimicrobial peptides: form follows function. *Nat. Rev. Drug Discov.* **11,** 37–51 (2012).

30. Monera, O. D., Sereda, T. J., Zhou, N. E., Kay, C. M. & Hodges, R. S. Relationship of sidechain hydrophobicity and α- helical propensity on the stability of the single- stranded amphipathic α- helix. *J. Pept. Sci.* **1,** 319–329 (1995).

31. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci.* **81,** 140–144 (1984).

32. Chou, P. Y. & Fasman, G. D. Prediction of Protein Conformation. *Biochemistry* **13,** 222–245 (1974).

33. Wiegand, I., Hilpert, K. & Hancock, R. E. W. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3,** 163–75 (2008).

34. Cue, D. *et al.* SaeRS-dependent inhibition of biofilm formation in Staphylococcus aureus Newman. *PLoS One* **10,** (2015).

35. Beaudoin, T., Zhang, L., Hinz, A. J., Parr, C. J. & Mah, T. F. The biofilm-specific antibiotic resistance gene ndvB is important for expression of ethanol oxidation genes in Pseudomonas aeruginosa biofilms. *J. Bacteriol.* **194,** 3128–3136 (2012).

36. O'Toole, G. A. Microtiter Dish Biofilm Formation Assay. *JoVE* **47,** (2011).

37. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins - a viable alternative to antibiotics? *Nat. Rev. Microbiol.* **11,** 95–105 (2013).

38. Chen, H. *et al.* Cloning, expression, and identification of a novel class IIa bacteriocin in the Escherichia coli cell-free protein expression system. *Biotechnol. Lett.* **34,** 359–364 (2012).

39. Pal, A., Ramana, K. V. & Bawa, A. S. Simplification and optimization of deMan Rogosa Sharpe (MRS) medium for enhanced production of bacteriocin by Weissella paramesenteroides DFR-8. *J. Food Sci. Technol.* **47,** 258–265 (2010).

442 40. Borrero, J. *et al.* Cloning, production, and functional expression of the bacteriocin
443 enterocin A, produced by Enterococcus faecium T136, by the yeasts Pichia pastoris,
444 Kluyveromyces lactis, Hansenula polymorpha, and Arxula adeninivorans. *Appl. Environ.*
445 *Microbiol.* **78,** 5956–5961 (2012).
446 41. Uggerhøj, L. E. *et al.* Rational Design of Alpha-Helical Antimicrobial Peptides: Do's and
447 Don'ts. *ChemBioChem* **16,** 242–253 (2015).
448 42. Yeaman, M. R. Mechanisms of Antimicrobial Peptide Action and Resistance.
449 *Pharmacol. Rev.* **55,** 27–55 (2003).
450 43. Lee, T., Hall, K. N. & Aguilar, M. Antimicrobial Peptide Structure and Mechanism of
451 Action : A Focus on the Role of Membrane Structure. *Curr. Top. Med. Chem.* **16,** 25–39 (2016).
452 44. Montalbán-López, M. *et al.* Characterization of linear forms of the circular enterocin AS-
453 48 obtained by limited proteolysis. *FEBS Lett.* **582,** 3237–3242 (2008).
454 45. Montalbán-López, M., Martínez-Bueno, M., Valdivia, E. & Maqueda, M. Expression of
455 linear permutated variants from circular enterocin AS-48. *Biochimie* **93,** 549–555 (2011).
456 46. Bagheri, M., Keller, S. & Dathe, M. Interaction of W-substituted analogs of cyclo-
457 RRRWFW with bacterial lipopolysaccharides: The role of the aromatic cluster in antimicrobial
458 activity. *Antimicrob. Agents Chemother.* **55,** 788–797 (2011).
459 47. Datta, A. *et al.* Role of aromatic amino acids in lipopolysaccharide and membrane
460 interactions of antimicrobial peptides for use in plant disease control. *J. Biol. Chem.* **291,**
461 13301–13317 (2016).
462 48. Maher, S. & Mcclean, S. N. Investigation of the cytotoxicity of eukaryotic and
463 prokaryotic antimicrobial peptides in intestinal epithelial cells in vitro. *Biochem. Pharmacol.* **71,**
464 1289–1298 (2006).

49. Fimland, G., Eijsink, V. G. H. & Nissen-Meyer, J. Mutational analysis of the role of tryptophan residues in an antimicrobial peptide. *Biochemistry* **41,** 9508–9515 (2002).

50. Cebrian, R. *et al.* The bacteriocin AS-48 requires dimer dissociation followed by hydrophobic interactions with the membrane for antibacterial activity. *J. Struct. Biol.* **190,** 162–172 (2015).

51. Ashby, M., Petkova, A. & Hilpert, K. Cationic antimicrobial peptides as potential new therapeutic agents in neonates and children: a review. *Curr. Opin. Infect. Dis.* **27,** 258–67 (2014).

52. Rubinchik, E., Dugourd, D., Algara, T., Pasetka, C. & Friedland, H. D. Antimicrobial and antifungal activities of a novel cationic antimicrobial peptide, omiganan, in experimental skin colonisation models. *Int. J. Antimicrob. Agents* **34,** 457–461 (2009).

53. Lamarche, M. J. *et al.* Discovery of LFF571: An investigational agent for Clostridium difficile infection. *Journal of Medicinal Chemistry* **55,** 2376–2387 (2012).

54. Burgos, M. J. G., Aguayo, M. C. L., Pulido, R. P., Gálvez, A. & López, R. L. Inactivation of Staphylococcus aureus in Oat and Soya Drinks by Enterocin AS-48 in Combination with Other Antimicrobials. *J. Food Sci.* **80,** 2030–2034 (2015).

Fig 1

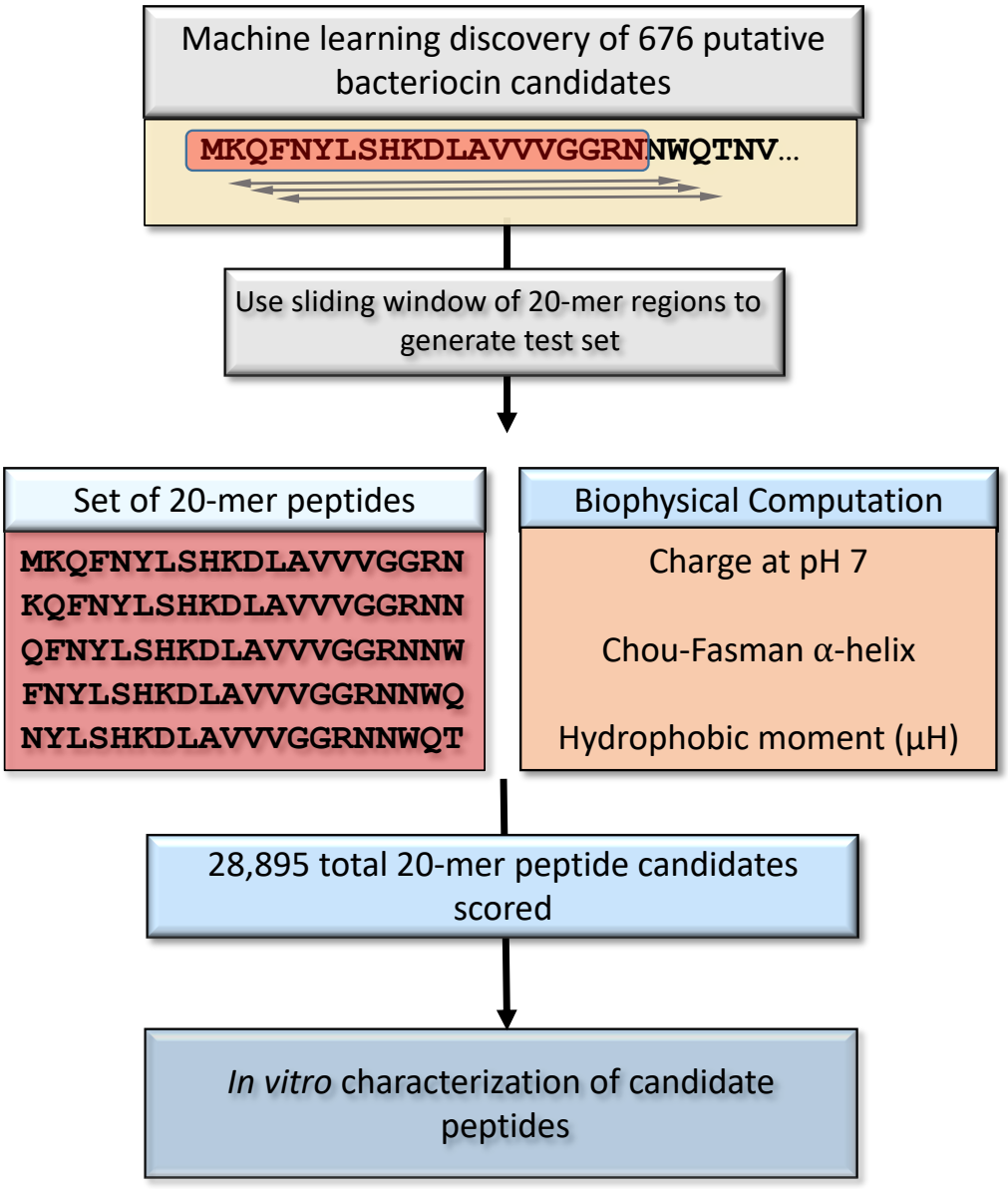


Figure 1: Overall strategy for selection of bacteriocins for synthesis. Machine learning set of 676 putative bacteriocins was used to generate overlapping 20-mer peptide candidates. 28,895 20-mers were scored for charge, helicity, and hydrophobic moment. A representative sample of 16 peptides were selected for synthesis and *in vitro* characterization in this study.

482

Fig 2

A.

Biophysical scoring of peptide candidates

	Charge	α - helix	μH
LOW	1-3	16-19	1-333
MIDDLE	4-6	19.01-22	333.01-667
HIGH	7-10	22.01-25	667.01-1000

B.

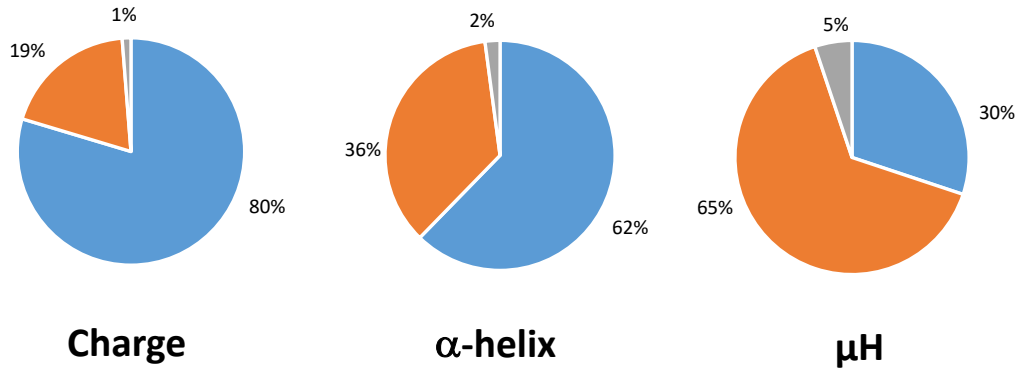


Figure 2: Scoring breakdown of biophysical computational parameters of the candidate peptides. **A.** Peptides were divided into high (grey), middle (orange), and low (blue) groups based on their charge, helicity, and hydrophobic moment scores. **B.** Most of the peptides scored low to middle with only a small percentage scoring high for each of the biophysical parameters.

Fig 3

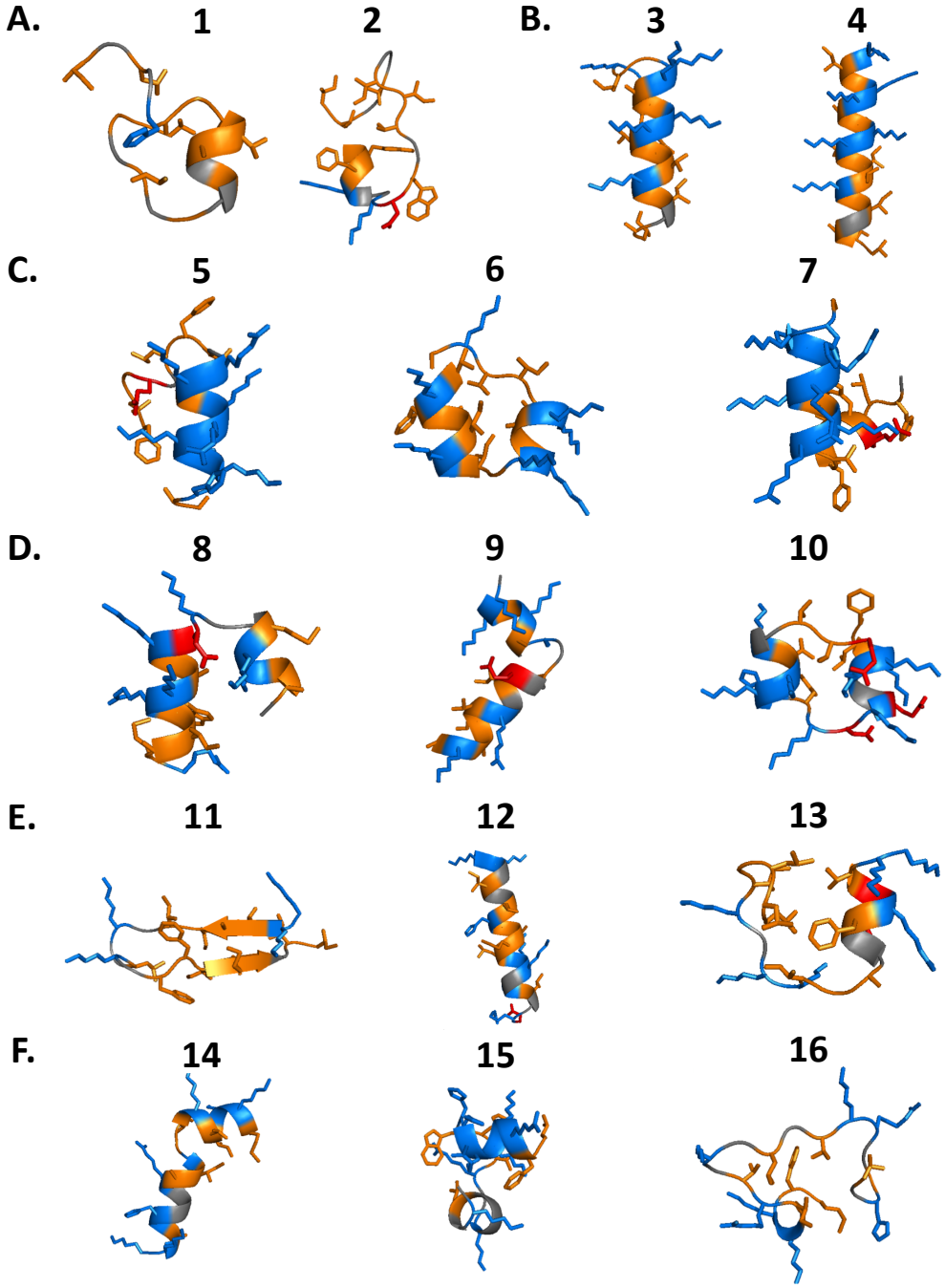


Figure 3: PEP-FOLD models of the peptides selected for synthesis **A.** peptides 1 and 2, **B.** 3 and 4, **C.** 5,6, and 7, **D.** 8, 9, and 10, **E.** 11, 12, and 13, and **F.** 14, 15, and 16. Basic, acidic, and hydrophobic residues are in blue, red, and orange respectively.

484

Table 1

Bacteriocin Accession #	Peptide sequence	Peptide #	Charge	Helicity	Hydrophobic Moment
REF_PRJNA222774:T285_02970	VGGTICGPACAVAGAHYLPI	1	1	18.93	322.53
YP_003061758.1	GWFKTYKDWPVIISNVPGVI	2	1	18.13	222.87
YP_004033496.1	IKKIGKKAAKKVIVKAIQAI	3	7	22.27	835.73
YP_004033496.1	KKIGKKAAKKVIVKAIQAIV	4	7	22.41	758.92
YP_008203939.1	FLGEWLFVTRKAKRKHKHAA	5	7	22.7	599.8
YP_004033496.1	VKFAIKKIGKKAAKKVIVKA	6	8	22.43	503.81
YP_003175051.1	SFLGEWLFVTRKAKRKHKHA	7	7	22.11	566.58
YP_001578036.1	QAAKKMSNKEKAAKRWIAMR	8	6	22.94	611.8
YP_005852536.1	AKAIVRTIEQAPKTAKAKKS	9	5	22.25	467.52
YP_005854059.1	EVVPKVKKHFKDKENKKKLF	10	6	22.16	466.17
YP_007937902.1	KKIAIITKNLKNFFIMLNII	11	4	20.86	912.76
YP_007988223.1	KKSINSLHLAMLKKQRLNRD	12	6	21.3	673.77
YP_003602376.1	LLLGIKYKKIGANNFREVKK	13	5	20.53	680.63
YP_004034256.1	FMKKIVKIGKIARYSHKAKK	14	9	20.81	567.158
YP_004840669.1	LFRKKRHHWFVRTKNSLSKN	15	9	20.19	357.31
YP_004032450.1	VNHGYRKVTLTKTVRIKKFM	16	7	19.51	471.63

Table 1: 20-mers selected for synthesis and their corresponding biophysical scores.

485

Table 2

MICs Table

Peptide #	*Pseudomonas aeruginosa*		*Escherichia coli*		*Staphylococcus aureus*	
	MIC (μM)	MBC (μM)	MIC (μM)	MBC (μM)	MIC (μM)	MBC (μM)
1	> 128	> 128	> 128	> 128	> 128	> 128
2	> 128	> 128	> 128	> 128	> 128	> 128
3	64	> 128	32	64	> 128	> 128
4	16	32	2	4	> 128	> 128
5	> 128	> 128	128	> 128	> 128	> 128
6	32	128	128	> 128	> 128	> 128
7	128	> 128	128	> 128	> 128	> 128
8	> 128	> 128	> 128	> 128	> 128	> 128
9	> 128	> 128	> 128	> 128	> 128	> 128
10	> 128	> 128	> 128	> 128	> 128	> 128
11	> 128	> 128	> 128	> 128	> 128	> 128
12	> 128	> 128	> 128	> 128	> 128	> 128
13	> 128	> 128	> 128	> 128	> 128	> 128
14	> 128	> 128	128	> 128	> 128	> 128
15	> 128	> 128	> 128	> 128	> 128	> 128
16	> 128	> 128	> 128	> 128	> 128	> 128

Table 2: MICs and MBCs of the synthetic 20-mer bacteriocins against *S. aureus*, *E. coli*, and *P. aeruginosa.*

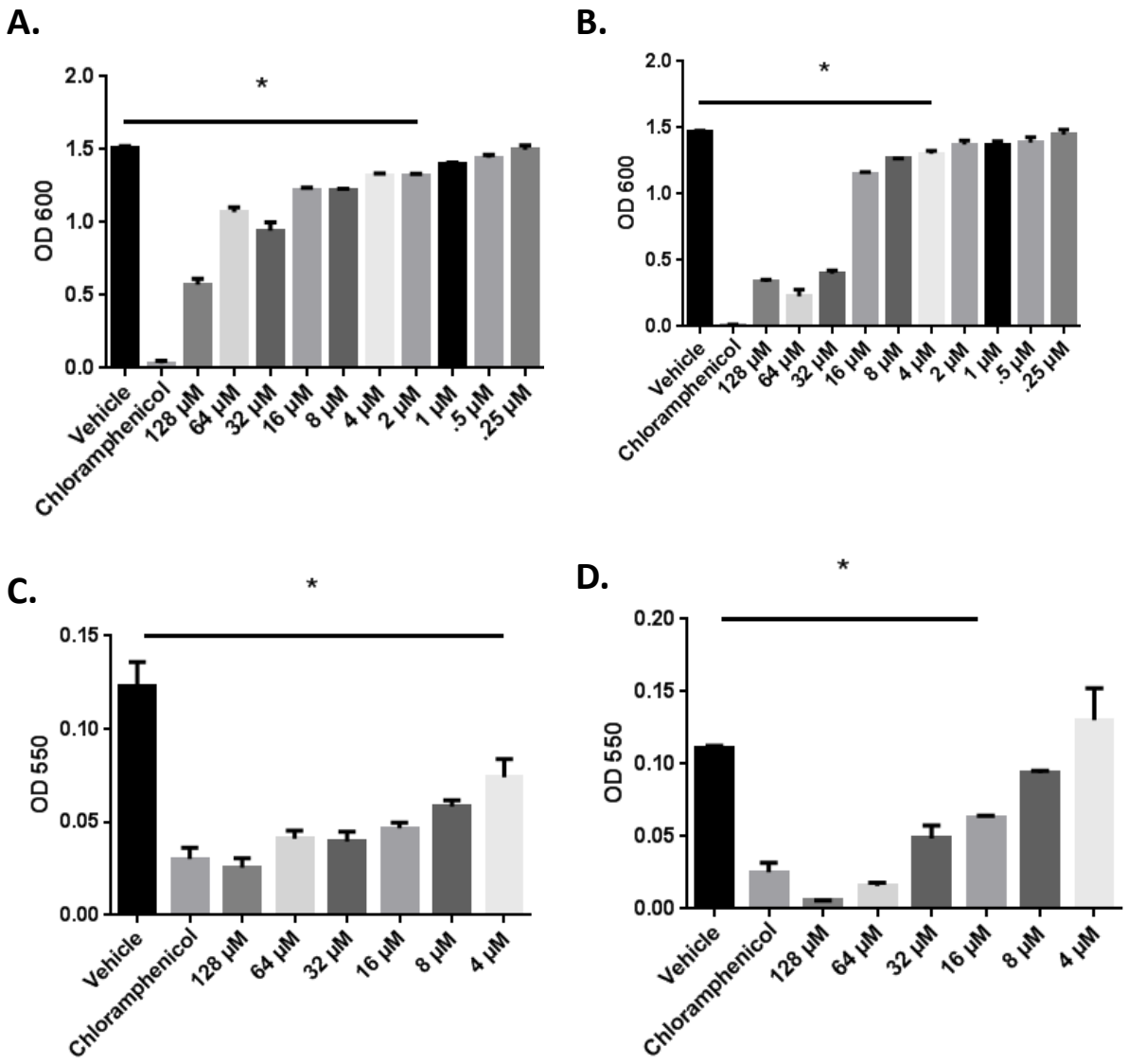
Supplemental Material

Supplemental Figure 1: Antibiofilm activities of peptides 11 and 16 on *S. aureus*. A. Peptide 11 exhibits a bacteriostatic effect and B. peptide 16 exhibits a bacteriostatic effect C. Peptide 11 inhibits biofilm formation at all concentrations tested. D. Peptide 16 inhibits biofilm formation to 16 μM. Data is representative of 3 biological replicates. P-values were determined via one-way ANOVA. A * indicates a significant difference determined via Tukey HSD compared to the vehicle control.

Supplemental Figure 2: Antibiofilm activities of peptides 11 and 16 on *P. aeruginosa*. A. Peptide 11 and B. peptide 16 exhibit no bacteriostatic activity. C. Peptide 11 exhibits mild antibiofilm activities D. Peptide 16 exhibits mild antibiofilm activity. Data is representative of 3 biological replicates. A * represents a p-value < .05 as determined via one-way ANOVA (A,B). A * represents a significant difference as determined via Tukey HSD compared to the vehicle control (C,D).

Supplemental Table 1: Cytotoxicity of 20-mer bacteriocins at 128 μM. Y indicates an increase in hemolysis or cytotoxicity at 128 μM. N indicates no increase in hemolysis or cytotoxicity at 128 μM.

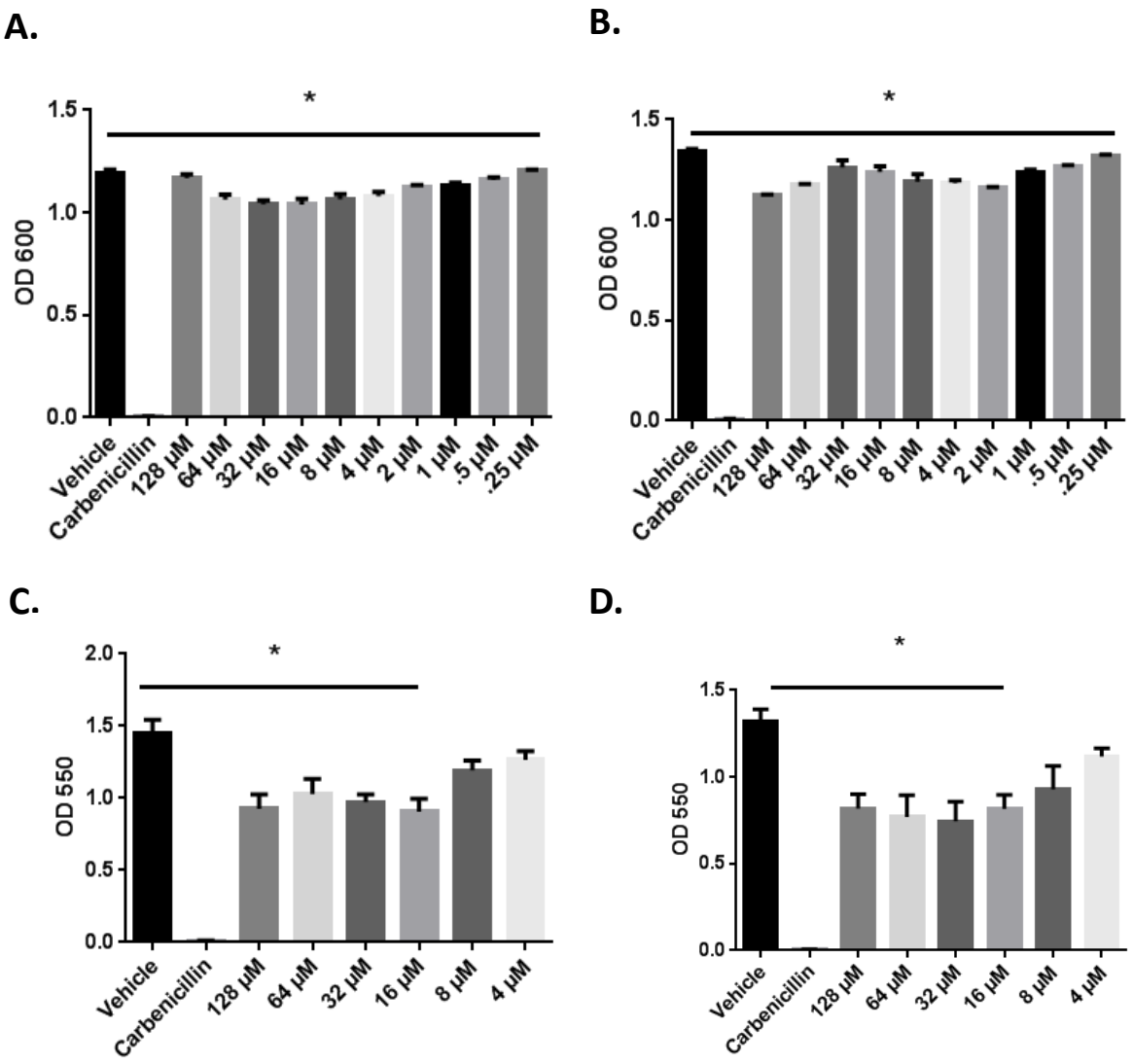
Supp Fig 1



Supplemental Figure 1: Antibiofilm activities of peptides 11 and 16 on *S. aureus*. **A.** Peptide 11 exhibits a bacteriostatic effect and **B.** peptide 16 exhibits a bacteriostatic effect **C.** Peptide 11 inhibits biofilm formation at all concentrations tested. **D.** Peptide 16 inhibits biofilm formation to 16 µM. Data is representative of 3 biological replicates. P-values were determined via one-way ANOVA. A * indicates a significant difference determined via Tukey HSD compared to the vehicle control.

504

Supp Fig 2



Supplemental Figure 2: Antibiofilm activities of peptides 11 and 16 on *P. aeruginosa*. **A.** Peptide 11 and **B.** peptide 16 exhibit no bacteriostatic activity. **C.** Peptide 11 exhibits mild antibiofilm activities **D.** Peptide 16 exhibits mild antibiofilm activity. Data is representative of 3 biological replicates. A * represents a p-value < .05 as determined via one-way ANOVA **(A,B)**. A * represents a significant difference as determined via Tukey HSD compared to the vehicle control **(C,D)**.

505

Supp Table 1

Peptide #	Charge	Helicity	Hydrophobic Moment	Cytotoxic?	Hemolytic?
1	1	18.93	322.53	N	N
2	1	18.13	222.87	N	Y
3	7	22.27	835.73	N	N
4	7	22.41	758.92	N	N
5	7	22.7	599.8	N	N
6	8	22.43	503.81	N	N
7	7	22.11	566.58	N	N
8	6	22.94	611.8	N	N
9	5	22.25	467.52	N	N
10	6	22.16	466.17	N	Y
11	4	20.86	912.76	N	N
12	6	21.3	673.77	N	N
13	5	20.53	680.63	N	N
14	9	20.81	567.158	N	N
15	9	20.19	357.31	N	N
16	7	19.51	471.63	N	N

Supplemental Table 1: Cytotoxicity of 20-mer bacteriocins at 128 µM. Y indicates an increase in hemolysis or cytotoxicity at 128 µM. N indicates no increase in hemolysis or cytotoxicity at 128 µM.

506