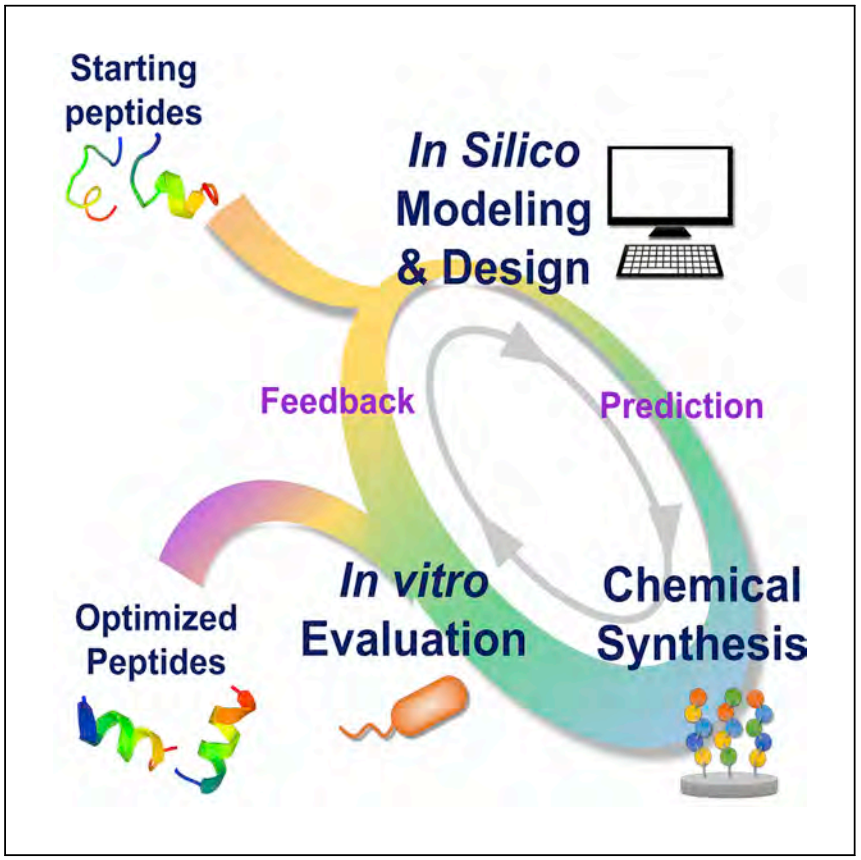
Chem

CellPress

Article

Using Evolutionary Algorithms and Machine Learning to Explore Sequence Space for the Discovery of Antimicrobial Peptides



Here, we use a closed-loop discovery and optimization approach for searching the peptide sequence space. Combining an evolutionary algorithm with machine learning and *in vitro* assay allowed for rapid development of new antimicrobial peptides.

Mari Yoshida, Trevor Hinkley, Soichiro Tsuda, ..., Sabrina Galiñanes Reyes, Maria D. Castro, Leroy Cronin

lee.cronin@glasgow.ac.uk

HIGHLIGHTS

Presents a general closed-loop approach for evolution of functional molecules

Couples machine learning and artificial intelligence with *in vitro* assay

Demonstrates quick identification of a number of potent antimicrobial peptides

Selects for α-helical conformation, a common motif of potent antimicrobial peptides

Yoshida et al., Chem 4, 533–543
March 8, 2018 © 2018 Published by Elsevier Inc.
https://doi.org/10.1016/j.chempr.2018.01.005

Chem

CellPress

Article

Using Evolutionary Algorithms and Machine Learning to Explore Sequence Space for the Discovery of Antimicrobial Peptides

Mari Yoshida,[1] Trevor Hinkley,[1] Soichiro Tsuda,[1] Yousef M. Abul-Haija,[1] Roy T. McBurney,[1] Vladislav Kulikov,[1] Jennifer S. Mathieson,[1] Sabrina Galiñanes Reyes,[1] Maria D. Castro,[1] and Leroy Cronin[1,2,*]

SUMMARY

We present a proof-of-concept methodology for efficiently optimizing a chemical trait by using an artificial evolutionary workflow. We demonstrate this by optimizing the efficacy of antimicrobial peptides (AMPs). In particular, we used a closed-loop approach that combines a genetic algorithm, machine learning, and *in vitro* evaluation to improve the antimicrobial activity of peptides against *Escherichia coli*. Starting with a 13-mer natural AMP, we identified 44 highly potent peptides, achieving up to a ca. 160-fold increase in antimicrobial activity within just three rounds of experiments. During these experiments, the conformation of the peptides selected was changed from a random coil to an α-helical form. This strategy not only establishes the potential of *in vitro* molecule evolution using an algorithmic genetic system but also accelerates the discovery of antimicrobial peptides and other functional molecules within a relatively small number of experiments, allowing the exploration of broad sequence and structural space.

INTRODUCTION

Natural biological polymers, such as peptides and RNAs, play a crucial role to maintain cellular functions. It is widely known that short RNA molecules interfere with gene expression within eukaryotic cells.[1] RNA molecules with enzymatic functions (ribozymes) are involved in various intracellular processes, such as RNA self-splicing.[2] Peptides are known to function as signaling molecules, such as hormones and neurotransmitters.[3] Antimicrobial peptides (AMPs) are yet another example of such natural functional biopolymers.[4] They are essential to the immune systems of all multicellular organisms as they have evolved to cope with bacterial invasion and infection. AMPs have been investigated as new antibiotic agents,[5] and more than 2,600 peptides that display antimicrobial activity from a broad range of organisms, including bacteria and mammalian cells, have been isolated.[6] Although the main mode of action causing bacterial death is disruption of the integrity of the bacterial membrane, AMPs have a wide variety of effective antimicrobial mechanisms, including the inhibition of DNA, RNA, and protein synthesis, to increase their efficacy to combat invading pathogens.[7,8] Methodologies to artificially discover such functional biopolymer sequences have been challenging because of the massive size of possible sequence space.[9] Conventionally, *in vitro* evolutionary methods, such as mRNA display and liposome display,[10,11] are used to tackle this combinatorial problem by generating billions of variants and screening them in a high-throughput fashion. However, there are limitations in these methodologies,[12]

The Bigger Picture

Biological evolution is a powerful way to produce new form and function but requires a fully functional biological organism. Here, we developed a closed-loop artificial evolution system and applied it to the exploration of antimicrobial peptides (AMPs). AMPs are a promising class of antibiotics to combat this issue because of their diverse mechanisms of action. However, discovery of antimicrobial peptides has been difficult because of a massive number of possible peptide sequences. Using this approach, we identified AMPs with improved potency. This method employs a genetic algorithm with peptide sequence as the "gene" and *in vitro* bacterial assay as "fitness." In addition, efficient predictions by machine learning further accelerated the process, showing up to a 160-fold potency increase within only three optimization rounds. This demonstrates a possibility of optimizing peptides without relying on an existing physicochemical database.

such as complex target-specific assay methods that are generally not transferrable to different experiments.

As an alternative method, rational design has long been studied to obtain molecules with a designed property. For example, in the case of AMPs, analogs of known natural AMPs were designed by modifying their physicochemical properties in order to improve antimicrobial activity and decrease toxicity to human cells.[8,13,14] This approach has identified peptides with improved antimicrobial activity. However, it has also illustrated the difficulty of rational AMP design because the physicochemical properties of peptides do not always correlate with antimicrobial activity. Indeed, *de novo* designs have been studied by defining sequence patterns of known AMPs to predict potent peptide sequences, which has revealed that specific motifs in a sequence are crucial for potent antimicrobial activity.[15,16] To overcome the difficulty in improving knowledge-based AMP designs, machine-learning-based approaches, such as artificial neural network models combined with chemoinformatic methods, were used to capture more complex motifs of antimicrobial activities.[7,17,18] Indeed, such approaches have proved important applied to drug design.[19] Although these designs have demonstrated promising predictive power, they require relatively large datasets for accurate predictions as well as careful parameter adjustments depending on the starting peptide library.

In contrast, genetic algorithms[20] with experimentally characterized feedback have allowed optimization of molecules with desired biological or chemical activities from small starting libraries.[21,22] This method has been applied to optimize relatively short peptides,[23,24] but optimizing longer peptides by using genetic algorithms can be challenging. This is because the number of possible amino acid combinations is huge (e.g., $20^{13} \approx 8 \times 10^{16}$ combinations for 13-mer peptides) and a search within the combinatorial space would require a large amount of experimental validation. This would limit optimization of AMPs considering that the synthesis of peptides is still costly. In this study, we propose a general methodology for evolution of physical objects (Scheme 1), including functional biomolecules. The method starts with a (population of) target object(s), which are physically synthesized. They are then ranked according to the results of the assay that is used for evaluating the effectiveness of the AMPs, whereby the best ones are used to produce a new population for the next generation.

To demonstrate its capability, we applied this method to discover effective AMPs by starting from a natural AMP. Although we used a similar methodology to evolve the physico-chemical properties of oil-in-water droplets successfully,[25] the application to AMP evolution is not straightforward because it could be hampered by other factors, such as the high cost of peptide synthesis. To circumvent this issue, here we combined the evolutionary method with machine learning, which provides more efficient predictions when generating the next generation.

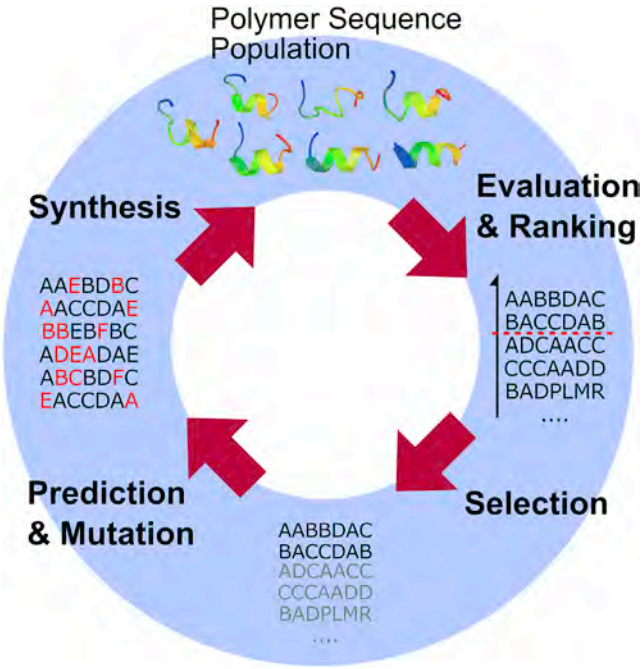
Our method differs from previously proposed *in silico* optimization algorithms that couple evolutionary algorithms and machine learning to discover AMPs as follows. Firstly, we used the interactive process of *in silico* prediction by a machine-learning model and experimental assay to screen better AMP candidates. This was initiated by the selection of a target sequence, the generation of a population, and physical synthesis of each of the members of the population for testing. In contrast, the previously proposed algorithms performed virtual screening of potential candidates by using existing knowledge or an existing database (e.g., structure-activity relationship) of AMPs during evolutionary optimization.[18] Because of this all-*in-silico* optimization process, such algorithms are inevitably database dependent. On the other

[1]WestCHEM, School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

[2]Lead Contact

*Correspondence: lee.cronin@glasgow.ac.uk

https://doi.org/10.1016/j.chempr.2018.01.005

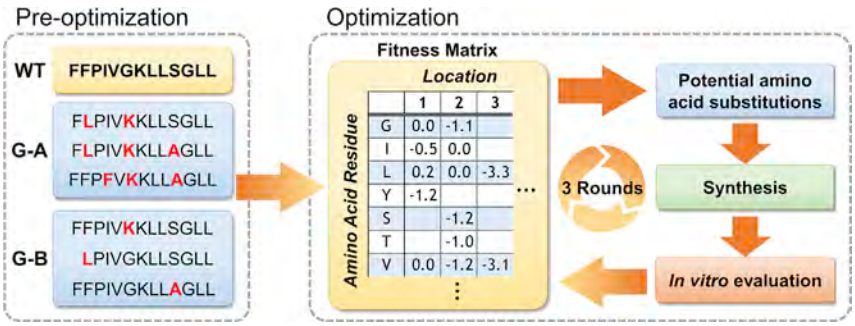


Scheme 1. General Schematic Describing the Evolutionary Process

The circle represents the robotic process with the computational algorithm. In the first step, a random selection of the polymer sequence population is used as the starting polymer sequence population, and this forms the sequences experimentally synthesized in a peptide synthesis robot. The recorded polymer sequence activities then undergo analysis against a user-desired property (e.g., IC_{50}) in the evaluation step. The sequences are ranked in terms of the desired property automatically, and those of poorest quality are rejected in the ranking step, allowing a new population to be selected. Meanwhile, the accepted sequences are used as a basis for creating a new sequence population after random mutation and crossover.

hand, our method is capable of optimizing a desirable feature of target compounds (in this case peptides) even if no preceding information or databases are available. This is because the evolutionary algorithm constructs its own database as it performs an online *in vitro* assay through the iterative process. For this reason, our algorithm can be a general methodology that applies to any functional polymers—not just AMPs—and can perform optimization by bootstrapping the assay to the synthesis process. Furthermore, as a control experiment, we compared the evolutionary algorithm-based optimizations guided by predictions from a machine-learning model and by random mutations. This is because we hypothesized that predictions by machine learning would allow us to navigate such massive peptide sequence space efficiently without synthesizing a large number of peptides and thus accelerate the evolutionarily driven search through such a large space. In theory, this would enable the discovery of potent AMPs in a more efficient manner and thus reduce the amount of chemical synthesis required, which is generally the most expensive and most time-consuming part of the optimization process.

As a proof-of-concept study, we started from a natural AMP to show that the algorithm can successfully identify peptide sequences with improved antimicrobial activity against *Escherichia coli*. The optimization of AMPs proceeded as follows (Scheme 2): In the pre-optimization process (left), two peptide libraries, generations A (G-A) and B (G-B) were identified *in silico* on the basis of a natural AMP, which we refer to as the wild-type (WT) sequence hereafter. The peptides were synthesized by



Scheme 2. Searching for Potent Antimicrobial Peptides through a Workflow that Combines the Genetic Algorithm with Machine Learning

an automated solid-phase method and evaluated *in vitro*. We analyzed the results to determine the approximate fitness values (i.e., expected improvements in antimicrobial activity) of individual amino acid substitutions by using a generalized linear model. In the optimization process (right), we prepared a fitness matrix by using the calculated fitness values and used it to predict potential amino acid substitutions. The *in silico* library was synthesized and evaluated *in vitro*. The fitness matrix was then updated at each round of optimization.

RESULTS

The optimization process went as follows. At each round, a new library of peptides was generated on the basis of predictions from the generalized linear model and evaluated experimentally. The antimicrobial activity of peptides improved rapidly with this method within a relatively small number of experimental cyles. The optimization of antimicrobial peptides was conducted as illustrated in Scheme 1. A 13-mer peptide, Temporin-Ali (FFPIVGKLLSGLL-NH_2) was chosen as a starting WT peptide because of its known moderate antimicrobial activity.[26] The optimization comprises two sub-processes, a pre-optimization and an optimization process. In the pre-optimization process (Scheme 2, left column), a library of peptide sequences, called G-A, was created with the position-specific interactive basic local alignment search tool (PSI-BLAST), which iteratively searches for sequences in the protein databases as it would a sequence of interest (see Supplemental Experimental Procedures).[27] We used PSI-BLAST rather than BLAST because the former can find distantly related, functionally similar sequences. We selected 93 sequences from the search for assay with 96-well plates. The WT sequence was compared with sequences in the PSI-BLAST database, and sequences with high identity scores, such as the WT sequence, were selected. We ranked all amino acid substitutions in G-A by frequency and generated G-B by applying the 93 most frequent single-amino-acid substitutions to the WT sequence. We then synthesized G-A and G-B and evaluated their antimicrobial activities *in vitro* by measuring the half maximal inhibitory concentration (IC_{50}) against *E. coli* (MG1655 strain). We used the *in vitro* evaluation data to train a generalized linear model that performs a regression analysis on amino acid substitutions (see Supplemental Experimental Procedures) and changes in IC_{50}. A coefficient for each substitution corresponds to the expected decrease in IC_{50} value and is presented as a value in a fitness matrix.

In the optimization process (Scheme 2, right column), we generated a library of new 90 peptide sequences *in silico* by substituting amino acid residues of the most potent peptide in the previous generation. Substitutions were introduced with a probability proportional to values in the fitness matrix. Those peptides were then

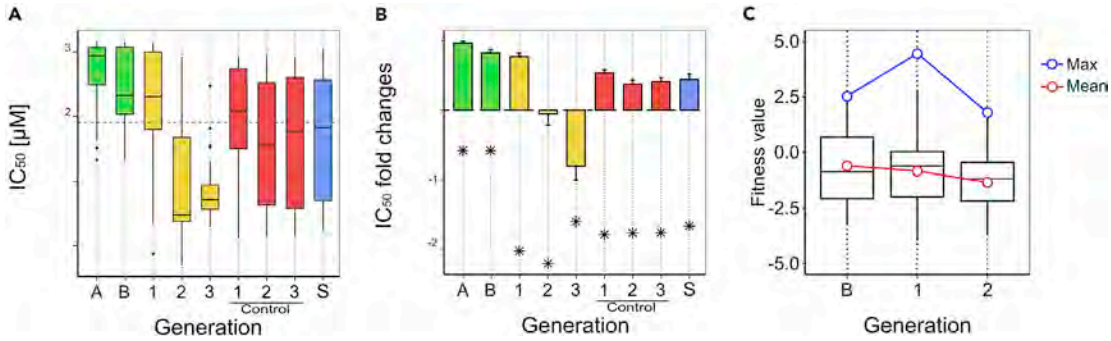


Figure 1. Optimization of Antimicrobial Peptides

(A) IC_{50} values of all the peptides in each generation are shown in a boxplot. Each solid line in the boxplot represents the median IC_{50} value. Boxes show the first and third quartiles. The upper and lower whiskers indicate 50% of the values higher and lower, respectively, than the median. Black dots represent outliers. Generations A and B, generations 1–3, strong substitutions, and control generations 1–3 are shown in green, yellow, blue, and red, respectively. The IC_{50} value of the WT peptide is shown by a dashed line.

(B) Fold changes of IC_{50} values in comparison with the WT sequence. Barplots indicate average changes in IC_{50} values with standard errors. The range of activity of the most potent peptides in each generation is shown by an asterisk.

(C) A boxplot of the fitness matrix values for each generation (G-B, G1, and G2). The maximum and average values for each generation are shown as blue and red circles, respectively.

synthesized (Figures S7 and S8) and evaluated experimentally (refer to the Supplemental Experimental Procedures for details). Then, we updated the fitness matrix by performing regression on all the *in vitro* data. The optimization process was repeated for three generations until the IC_{50} improvement saturated. In order to assess the effect of the fitness matrix, we performed a control optimization experiment for three generations by introducing random amino acid substitutions instead of predicting advantageous substitutions by using the matrix. The IC_{50} values of peptides in the optimization experiment, along with that of the WT peptide (81.0 μM, dashed line), are shown as boxplots in Figure 1A. The average changes in IC_{50} values in relation to the WT values are shown as bar plots in Figure 1B, and the WT and the most potent sequences in individual generations are in Table 1. In G-A and G-B, most peptides were less potent than the WT peptide, whereby they demonstrated 9.3- and 6.7-fold higher average IC_{50} values, respectively, than the IC_{50} of the WT peptide. In the first generation (G1), only a few peptides presented improved antimicrobial activity, of which the lowest IC_{50} value was 0.75 μM; most of the peptides in G1 showed weaker antimicrobial activity than the WT peptide. In the second generation (G2), more peptides demonstrated improved antimicrobial activity, and the average IC_{50} value decreased to 71.5 μM. The lowest IC_{50} value also decreased to 0.5 μM. In the third generation (G3), although there was no improvement in the lowest IC_{50} value, most peptides demonstrated very strong antimicrobial activity. The average IC_{50} value significantly improved (12.4 μM), such that it was 6.5-fold lower than that of the WT peptide. The most potent peptide was found in G2, which had an IC_{50} value (0.5 μM) 162-fold lower than that of the WT peptide. This peptide is referred to as the "best peptide" hereafter. Other peptides with strong antimicrobial activity were also identified (Table S1); 44 new peptides (29.1%) of 141 peptides tested showed IC_{50} values lower than 4.1 μM (i.e., 20-fold decrease in IC_{50}) within three generations.

The fact that there were no improvements in terms of IC_{50} in G3 suggests that the optimization of AMPs was converging within three generations. A possible explanation can be found in the fitness matrix (Figure 1C). After each generation, we re-calculated the fitness matrix to predict which amino acid substitution (at which locus) would be likely to improve the antimicrobial activity. When we predicted

Table 1. Experimental IC_{50} Values, Physicochemical Properties, and Predicted Hemolytic Potency of the WT and Most Potent Peptides in Individual Generations

Generation	Sequence	IC_{50} (μM)	Net Charge	Hydrophobic Moment	Hemolytic Potency
WT	FFPIVGKLLSGLL	81.0	0.98	0.45	0.66
A	FFPIVGKLLSGLF	21.1	0.98	0.45	0.54
B	FFPIVGKLLSGLF	21.1	0.98	0.45	0.54
1	FFPIVKKLLSGLF	0.75	1.98	0.53	0.58
2	FLPIVKKLLRGLF	0.50	2.98	0.74	0.55
3	VLPIVKKLLKGLF	2.01	2.98	0.64	0.50
1C	FLPIVKKLLRKLF	1.30	3.98	0.76	0.52
2C	FFPIFGKLLRGLF	1.37	1.98	0.65	0.52
3C	FFPIVGKLLRKLF	1.39	2.98	0.70	0.57

The amino acid residues substituted from the WT sequence are underlined. Net charges and hydrophobic moments were calculated with the Peptides package in R. Hemolytic potency was predicted with HemoPI,[29] and lower values indicate that peptides are less hemolytic.

G3 substitutions by using the data from G-A, G-B, G1, and G2, the maximum of the fitness matrix values dropped sharply in comparison with the previous generation. The average value also decreased. Because the fitness matrix values can be interpreted as the "expectation" for each amino acid substitution to improve the activity, the data suggest that the model predicted a lower increase in the antimicrobial activity (i.e., converging). The reason for a smaller standard deviation can be simply attributed to the model prediction. As the model collects more experimental data as the optimization proceeds, predictions by the model and the data become more and more accurate. Thus, in the later generations, the model tended to select amino acid substitutions that were likely to improve as AMP or less likely to decrease the efficacy. With the implementation of the genetic algorithm, a peptide with the strongest activity in the previous generation was carried over to the next generation, and the core effective substitutions were maintained. On the other hand, we introduced new potentially good substitutions into the peptide to generate a set of peptides for the next generation. As a result, the efficacy data for the G3 peptides showed a smaller standard deviation than the previous one. To investigate the convergence of the optimization process further, we performed additional experiments. First, we identified 20 amino acid substitutions that were likely to increase the antimicrobial activity by screening the IC_{50} data for all the sequences synthesized in the previous generations (see Table S3). We then generated a new set of 39 different sequences by introducing the selected substitutions randomly into the WT sequence. We refer to this peptide set as "strong substitutions." As a result, we found that the IC_{50} values of these peptides varied from 1.73 to 1,155.3 μM with an average IC_{50} value of 40.8 μM (2-fold increase over that of the WT peptide). Because the new peptide set did not show any improvements over the best peptide in G2, this would support the notion that the optimization converged within the three rounds of optimization, as well as the analysis of the fitness matrix values (Figure 1C).

Peptides from the control experiment demonstrated relatively poor improvements in antimicrobial activity (Figure 1, red). Although 44 peptides (17.2%) within the 256 peptides tested showed IC_{50} values lower than 4.1 μM, up to a ca. 60-fold decrease, the others showed a diverse range of antimicrobial activity. In fact, the average IC_{50} value of the third control generation was still 2.6-fold higher than that of the WT peptide. However, the IC_{50} values of the most potent peptides in the control generations were comparable (~1.3 μM; Table 1) with those in the

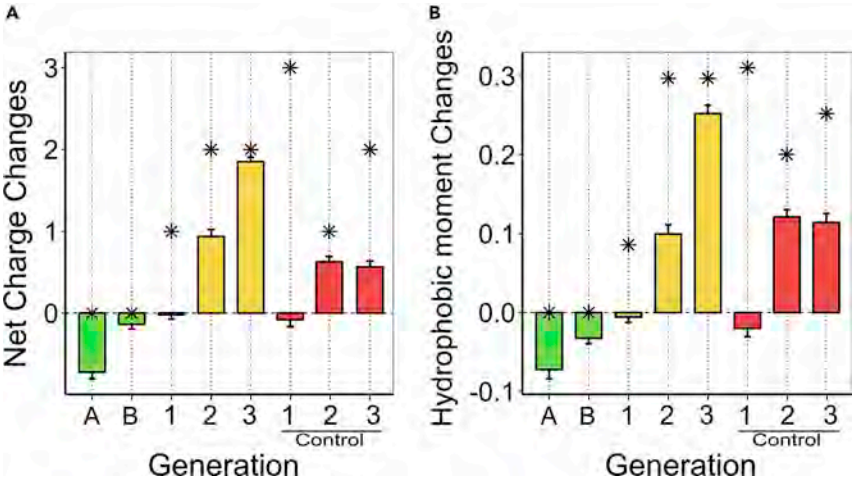


Figure 2. Changes in the Physicochemical Properties of Peptides

Values of the properties of peptides were subtracted from those of the WT peptide. Average changes in (A) net charge and (B) hydrophobic moment in individual generations are shown in barplots with standard errors. Changes in the most potent peptides in individual generations are shown as asterisks. Generations A and B, generations 1–3, and control generation 1–3 are shown in green, yellow, and red, respectively.

non-control generations (0.5–2.0 μM). In addition, the optimized peptide sequences in both cases were relatively similar. This result indicates that the genetic algorithm used to optimize the WT peptide was indeed effective at identifying a peptide with high antimicrobial activity. However, because of the nonlinear effect of amino acid substitution, the entire population in the control generation showed diverse antimicrobial activity levels. This also illustrates that the fitness matrix indeed played an important role in the search for multiple potent AMPs and not just the best one.

We calculated the physicochemical properties of all the peptides evaluated to investigate the mechanism behind the improvement in antimicrobial activities. The following peptide properties were calculated: molecular weight (g/mol), net charge, hydrophobicity, hydrophobic moment, isoelectric point (pI), aliphatic index, instability index, and Boman index. Although moderate correlations between the calculated parameter values and IC_{50} were observed in some of the properties (Figures S1–S3), the results indicated that finding explicit correlations between them is not a trivial task, especially because of the many outliers in the plots (Figure S3). Among others, net charge and hydrophobic moment, the physicochemical properties known to be important for antimicrobial activity, showed relatively good correlation with IC_{50} values. Average changes in net charge and hydrophobic moment in relation to those of the WT peptide are shown in bar plots in Figure 2. The average net charge was increased through the optimization process (Figure 2A). As a result, the peptides in the later generations were positively charged, which indicates better binding to the negatively charged bacterial membrane than the WT peptide.[28] The hydrophobic moment for the peptide rotation angle of 100° also increased through the optimization process (Figure 2B). This suggests that the peptides in the latter generation may have formed amphipathic helical structures with hydrophilic and hydrophobic amino acid residues on each side of the helix,[29] which facilitated the attachment and insertion into bacterial membranes.[5]

The increased hydrophobic moment indicated that the peptides in latter generations improved the spatial amphiphilicity if a peptide took a helical structure.

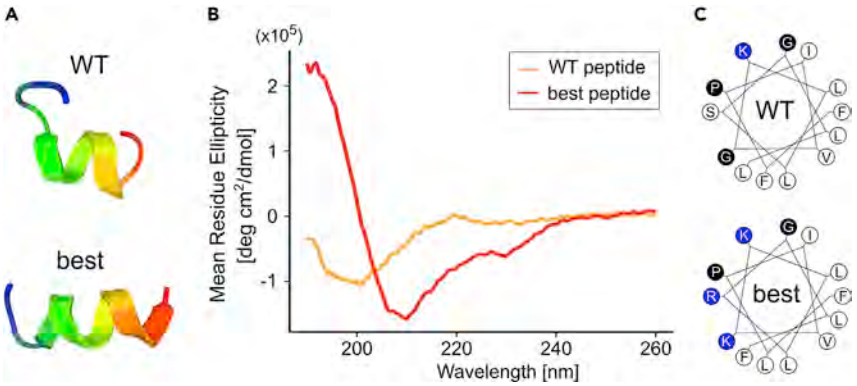


Figure 3. Structure of the WT and Best Peptide

(A) *De novo* structural reconstruction using PEP-FOLD3.

(B) Structural characterization using CD spectroscopy. The spectra of the WT and best peptides are shown in yellow and red, respectively.

(C) Helical wheel projection of peptides. Positively charged residues, hydrophobic residues, and other residues are shown in blue, white, and black, respectively. One-letter codes are used for amino acids.

Amphipathic α-helical peptides are known to be more potent than peptides with less-defined secondary structures.[5] In fact, the best peptide had a higher value for hydrophobic moment (0.74) than the WT sequence (Table 1 and Figure S3B). To confirm this, we performed *de novo* structural reconstructions by using PEP-FOLD3,[30] and they indicated that the best peptide formed a longer helical structure than the WT peptide and other sequences in G1 and G2 (Figures 3A and S4). These changes in secondary structure were confirmed by circular dichroism (CD) spectroscopy. The CD spectrum of the WT sequence and other sequences in G1 and G2 indicated a random coil structure, whereas that of the best peptide showed a profile typical of an α-helical structure (Figures 3B and S4B). These results strongly suggest that the best peptide with a high hydrophobic moment and helical structure allowed better disruption of the bacterial membrane.[28] However, care must be taken when considering these two measures (hydrophobic moment and helicity) in relation to the antimicrobial activity. Although the hydrophobic moment showed moderate correlation with the antimicrobial activity (Figure S3B), it does not mean that peptides with high hydrophobic moment always have helical structures. In fact, some of the peptides with a high hydrophobic moment (Figure S4C) were predicted to have less-defined structures despite the antimicrobial activity (IC_{50} < 2.5 μM).

In terms of the peptide sequence, the helical wheels showed that there were more positively charged amino acid residues located on one side of the helix in the best peptide, the opposite of the hydrophobic sector, whereas the WT sequence had only one positively charged residue (Figure 3C). This change occurred not only for the best peptide but also for other peptides in G2 and G3 (Figure S5). They had multiple positively charged residues at locations six, seven, and ten (Figure S6). The result illustrates that the model-based predictions successfully identified favorable substitutions,[31] which significantly improves the antimicrobial activity.

As potent AMPs, it is important that the peptides selectively attack bacterial cells while keeping host cells intact. To confirm this, we examined potential hemolytic activity and cytotoxicity of the optimized peptides by using *in silico* predictive models, called HemoPI[6] and ToxinPred,[32] respectively (see Table 1). The predicted hemolytic potency was gradually improved over the generations. In fact, we

experimentally confirmed the low hemolytic activity by using the best peptide and red blood cells, which showed only ~1% lysis even at a higher concentration than the IC_{50} (Figure S9), and ToxinPred also predicted that all the peptides in Table 1 were non-toxic. We speculate that this improved peptide activity might be related to the fact that this optimization process showed evolutionary steps similar to the natural evolution of antimicrobial peptides (Figure S10).[33] Furthermore, we tested the efficacy of the best peptide against drug-resistant bacterial strains (Figure S11). To assess the efficacy in a comparable manner, we used drug-resistant *E. coli* MG1655 strains that we obtained by serially culturing the WT strain (the same strain used for the *in vitro* assay above) under single or multiple antibiotic stress.[34] Despite the fact that the bacterial strains are resistant to various types of antibiotics, including a peptide antibiotic polymyxin B or combinations of them, IC_{50} values of the best peptide against the drug-resistant strains were still low (approximately 1.5–2.0 μM). Although the efficacy against clinically isolated drug-resistant bacteria may be different, this result suggests that the optimized peptide is also effective against drug-resistant bacteria that are difficult to treat with existing antibiotics.

DISCUSSION

In summary, we developed a new approach to the design of AMPs by combining an evolutionary algorithm, machine-learning-based prediction, and *in vitro* bacterial assays. This method rapidly improved antimicrobial activity, such that it achieved 162-fold more than the original peptide within three generations. The best peptide identified here has one of the lowest IC_{50} values among the previously known AMPs, despite the short sequence length.[34] In addition to the best peptide, 44 new peptides were found to be highly potent with 20-fold lower IC_{50} values than the WT peptide. This discovery of multiple potent peptides is remarkable considering the number of peptides tested through the pre-optimization and optimization processes (291 of 8×10^{16} possible peptides). Because the other factors such as low cytotoxicity are also crucial for AMPs, identification of multiple potent AMPs by the model prediction would be useful for further screening of antimicrobial agents. In fact, all the peptides in G3 were predicted to be non-toxic partially because of the natural origin of the WT peptides (i.e., Temporin-Ali isolated from a frog).[26] Although we evaluated the synthesized peptides with the antimicrobial activity on *E. coli* in this study, the optimization method can adapt other measures—such as the IC_{50} of multi-drug-resistant bacteria, hemolytic potency, and cytotoxicity—to calculate the fitness matrix. In addition, it is also possible to use multiple measures by performing multivariate regression with the generalized linear model.

One of the advantages of this optimization method is that it does not require any prior knowledge before starting the process. This is because of the iterative nature of the process that couples *in silico* algorithm with experimental validation. This makes a stark contrast with previous all-*in-silico* algorithms, where a database of optimized materials (e.g., AMPs)[35] is crucial. Our approach constructs its own database as the algorithm runs the optimization process. Experimental validation, however, generally comes with a high cost for chemical synthesis. This can be mitigated by machine-learning-based efficient prediction. Any peptide sequences can thus be used as a starting peptide for optimization by this method, although only one natural AMP was tested in this study. Most of the known natural AMPs have not yet been utilized in clinical or industrial processes because they have only moderate direct antimicrobial activity that works efficiently at the site of infection in harmony with other immune systems.[36–38] Our method can potentially identify a series of potent peptides by using those natural AMPs as starting peptides.

Furthermore, we also anticipate that this method would be applicable to exploring the broader sequence space of peptides or other polymers.[39] For example, incorporation of new components such as non-canonical amino acids (ncAAs) would improve the chemical diversity of AMPs.[40,41] Discovery and optimization of AMPs containing ncAAs with conventional *in silico* methods are currently challenging because they rely on existing databases of physicochemical properties or efficacy. However, our method proposed here does not require any prior information to perform evolutionary optimization and hence could open up a path for incorporating a wider range of molecular building blocks to discover novel AMPs as well as other functional polymers. In future work, we will aim to explore sequence space more broadly to see how algorithm-based evolutionary systems with a digital genome can be used to explore AMPs and other properties from entirely random sequences.

EXPERIMENTAL PROCEDURES

Because the details of the *in silico* optimization algorithm and the *in vitro* bacterial assay require a long and detailed technical description, please refer to the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, 11 figures, and 3 tables and can be found with this article online at https://doi.org/10.1016/j.chempr.2018.01.005.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (grants EP/H024107/1, EP/I033459/1, EP/J00135X/1, EP/J015156/1, EP/K021966/1, EP/K023004/1, EP/K038885/1, EP/L015668/1, and EP/L023652/1), BBSRC (grant BB/M011267/1), EC (projects 610730 EVOPROG and 611640 EVOBLISS), the European Research Council (project 670467 SMART-POM), the University of Glasgow for a LKAS fellowship, and the Honjo International Scholarship Foundation. The authors thank Prof. Sharon Kelly for using the CD instrument.

AUTHOR CONTRIBUTIONS

L.C. conceived the original idea, and L.C., M.Y., and T.H. adapted and designed the project. L.C. coordinated the efforts of the research team with help from M.Y., T.H., R.T.M., and V.K. T.H. and S.T. programmed the *in silico* algorithms. M.Y., S.T., and S.G.R. performed the *in vitro* bacterial and hemolysis assays. M.Y. and Y.M.A.-H. performed the CD spectroscopy assays. R.T.M., V.K., Y.M.A.-H., M.D.C., and J.S.M. synthesized peptides and analyzed the HPLC and MS results. M.Y., S.T., and S.G.R. analyzed the collected experimental data. L.C., S.T., M.Y., and Y.M.A.-H. co-wrote the manuscript with input from all co-authors. All authors co-wrote the Supplemental Information.

DECLARATION OF INTERESTS

L.C. is the inventor of a patent, EP2855008 A1, related to this manuscript and is the founding scientific director of CroninGroupPLC.

Received: August 4, 2017
Revised: September 7, 2017
Accepted: January 8, 2018
Published: February 8, 2018

REFERENCES AND NOTES

1. Wilson, R.C., and Doudna, J.A. (2013). Molecular mechanisms of RNA interference. Annu. Rev. Biophys. *42*, 217–239.
2. Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., and Breaker, R.R. (2010). An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science *329*, 845–848.
3. Hook, V., Funkelstein, L., Lu, D., Bark, S., Wegrzyn, J., and Hwang, S.R. (2008). Proteases for processing proneuropeptides into peptide neurotransmitters and hormones. Annu. Rev. Pharmacol. Toxicol. *48*, 393–423.
4. Zhang, L.J., and Gallo, R.L. (2016). Antimicrobial peptides. Curr. Biol. *26*, R14–R19.
5. Brogden, K.A. (2005). Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol. *3*, 238–250.
6. Chaudhary, K., Kumar, R., Singh, S., Tuknait, A., Gautam, A., Mathur, D., Anand, P., Varshney, G.C., and Raghava, G.P. (2016). A web server and mobile app for computing hemolytic potency of peptides. Sci. Rep. *6*, 22843.
7. Cherkasov, A., Hilpert, K., Jenssen, H., Fjell, C.D., Waldbrook, M., Mullaly, S.C., Volkmer, R., and Hancock, R.E. (2009). Use of artificial intelligence in the design of small peptide antibiotics effective against a broad spectrum of highly antibiotic-resistant superbugs. ACS Chem. Biol. *4*, 65–74.
8. Dathe, M., Nikolenko, H., Meyer, J., Beyermann, M., and Bienert, M. (2001). Optimization of the antimicrobial activity of magainin peptides by modification of charge. FEBS Lett. *501*, 146–150.
9. Currin, A., Swainston, N., Day, P.J., and Kell, D.B. (2015). Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. Chem. Soc. Rev. *44*, 1172–1239.
10. Wilson, D.S., Keefe, A.D., and Szostak, J.W. (2001). The use of mRNA display to select high-affinity protein-binding peptides. Proc. Natl. Acad. Sci. USA *98*, 3750–3755.
11. Fujii, S., Matsuura, T., Sunami, T., Kazuta, Y., and Yomo, T. (2013). *In vitro* evolution of α-hemolysin using a liposome display. Proc. Natl. Acad. Sci. USA *110*, 16796–16801.
12. Bornscheuer, U.T., and Pohl, M. (2001). Improved biocatalysts by directed evolution and rational protein design. Curr. Opin. Chem. Biol. *5*, 137–143.
13. Blondelle, S.E., and Houghten, R.A. (1992). Design of model amphipathic peptides having potent antimicrobial activities. Biochemistry *31*, 12688–12694.
14. Zelezetsky, I., Pag, U., Antcheva, N., Sahl, H.G., and Tossi, A. (2005). Identification and optimization of an antimicrobial peptide from the ant venom toxin pilosulin. Arch. Biochem. Biophys. *434*, 358–364.
15. Tossi, A., Tarantino, C., and Romeo, D. (1997). Design of synthetic antimicrobial peptides based on sequence analogy and amphipathicity. Eur. J. Biochem. *250*, 549–558.
16. Loose, C., Jensen, K., Rigoutsos, I., and Stephanopoulos, G. (2006). A linguistic model for the rational design of antimicrobial peptides. Nature *443*, 867–869.
17. Fjell, C.D., Jenssen, H., Hilpert, K., Cheung, W.A., Pant, N., Hancock, R.E., and Cherkasov, A. (2009). Identification of novel antibacterial peptides by chemoinformatics and machine learning. J. Med. Chem. *52*, 2006–2015.
18. Fjell, C.D., Hiss, J.A., Hancock, R.E.W., and Schneider, G. (2012). Designing antimicrobial peptides: form follows function. Nat. Rev. Drug Discov. *11*, 37–51.
19. Wedge, D.C., Rowe, W., Kell, D.B., and Knowles, J. (2009). *In silico* modelling of directed evolution: implications for experimental design and stepwise evolution. J. Theor. Biol. *257*, 131–141.
20. Small, B.G., McColl, B.W., Allmendinger, R., Pahle, J., López-Castejón, G., Rothwell, N.J., Knowles, J., Mendes, P., Brough, D., and Kell, D.B. (2011). Efficient discovery of anti-inflammatory small-molecule combinations using evolutionary computing. Nat. Chem. Biol. *7*, 902–908.
21. Pickett, S.D., Green, D.V.S., Hunt, D.L., Pardoe, D.A., and Hughes, I. (2011). Automated lead optimization of MMP-12 inhibitors sing a genetic algorithm. ACS Med. Chem. Lett. *2*, 28–33.
22. Kreutz, J.E., Shukhaev, A., Du, W., Druskin, S., Daugulis, O., and Ismagilov, R.F. (2010). Evolution of catalysts directed by genetic algorithms in a plug-based microfluidic device tested with oxidation of methane by oxygen. J. Am. Chem. Soc. *132*, 3128–3132.
23. Singh, J., Ator, M.A., Jaeger, E.P., Allen, M.P., Whipple, D.A., Soloweij, J.E., Chowdhary, S., and Treasurywala, A.M. (1996). Application of genetic algorithms to combinatorial synthesis: a computational approach to lead identification and lead optimization. J. Am. Chem. Soc. *118*, 1669–1676.
24. Yokobayashi, Y., Ikebukuro, K., McNiven, S., and Karbe. (1996). Directed evolution of trypsin inhibiting peptides using a genetic algorithm. J. Chem. Soc. Perkin. Trans. *20*, 2435–2437.
25. Gutierrez, J.M.P., Hinkley, T., Taylor, J.W., Yanev, K., and Cronin, L. (2014). Evolution of oil droplets in a chemorobotic platform. Nat. Commun. *5*, 5571.
26. Wang, M., Wang, Y., Wang, A., Song, Y., Ma, D., Yang, H., Ma, Y., and Lai, R. (2010). Five novel antimicrobial peptides from skin secretions of the frog, *Amolops loloensis*. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. *155*, 72–76.
27. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
28. Hall, K., Mozsolits, H., and Aguilar, M. (2003). Surface plasmon resonance analysis of antimicrobial peptide-membrane interactions: affinity & mechanism of action. Lett. Pept. Sci. *10*, 475–485.
29. Eisenberg, D., Weiss, R.M., and Terwilliger, T.C. (1982). The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature *299*, 371–374.
30. Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derreumaux, P., and Tufféry, P. (2016). PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. Nucleic Acids Res. *44*, W449–W454.
31. Hilpert, K., Volkmer-Engert, R., Walter, T., and Hancock, R.E.W. (2005). High-throughput generation of small antibacterial peptides with improved activity. Nat. Biotechnol. *23*, 1008–1012.
32. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., and Raghava, G.P.; Open Source Drug Discovery Consortium (2013). *In silico* approach for predicting toxicity of peptides and proteins. PLoS One *8*, e73957.
33. Torrent, M., Valle, J., Nogués, M.V., Boix, E., and Andreu, D. (2011). The generation of antimicrobial peptide activity: a trade-off between charge and aggregation? Angew. Chem. Int. Ed. *50*, 10686–10689.
34. Yoshida, M., Reyes, S.G., Tsuda, S., Horinouchi, T., Furusawa, C., and Cronin, L. (2017). Time-programmable drug dosing allows the manipulation, suppression and reversal of antibiotic drug resistance *in vitro*. Nat. Commun. *8*, 15589.
35. Pirtskhalava, M., Gabrielian, A., Cruz, P., Griggs, H.L., Squires, R.B., Hurt, D.E., Grigolava, M., Chubinidze, M., Gogoladze, G., Vishnepolsky, B., et al. (2016). DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. Nucleic Acids Res. *44*, D1104–D1112.
36. Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. Nat. Rev. Immunol. *3*, 710–720.
37. Hancock, R.E.W., and Sahl, H.G. (2006). Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. Nat. Biotechnol. *24*, 1551–1557.
38. Mansour, S.C., Pena, O.M., and Hancock, R.E.W. (2014). Host defense peptides: front-line immunomodulators. Trends Immunol. *35*, 443–450.
39. Frederix, P.W., Scott, G.G., Abul-Haija, Y.M., Kalafatovic, D., Pappas, C.G., Javid, N., Hunt, N.T., Ulijn, R.V., and Tuttle, T. (2014). Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. Nat. Chem. *7*, 30–37.
40. Molhoek, E.M., Van Dijk, A., Veldhuizen, E.J.A., Haagsman, H.P., and Bikker, F.J. (2011). Improved proteolytic stability of chicken cathelicidin-2 derived peptides by D-amino acid substitutions and cyclization. Peptides *32*, 875–880.
41. Baumann, T., Nickling, J.H., Bartholomae, M., Buivydas, A., Kuipers, O.P., and Budisa, N. (2017). Prospects of in vivo incorporation of non-canonical amino acids for the chemical diversification of antimicrobial peptides. Front. Microbiol. *8*, 124.