Chem

CellPress

Article Type

Exploring Sequence Space for Antimicrobial Peptides using Evolutionary Algorithms and Machine Learning

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] Leroy Cronin[1, 2, *]

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

[2]Lead Contact

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

SUMMARY

We present a proof of concept methodology to efficiently optimize a chemical trait using an artificial evolutionary workflow. We demonstrate this by optimizing the efficacy of antimicrobial peptides (AMPs). In particular, a closed-loop approach that combines genetic algorithm, machine learning, and *in vitro* evaluation was employed to improve the antimicrobial activity of peptides against Escherichia coli. Starting with a 13-mer natural AMP, 44 highly potent peptides were identified 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 α-helical form. This strategy not only establishes the potential of evolution of molecules *in vitro* using an algorithmic genetic system, but also accelerates the discovery of antimicrobial peptides and other functional molecules within a relatively small number of experiments, and to explore broad sequence and structural space.

Keywords artificial intelligence, genetic algorithm, machine learning, *in vitro* bacterial assay, antibiotics, antimicrobial peptides, peptide conformation, bacterial multi-drug resistance

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 gene expressions 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 that they have evolved to cope with bacterial invasion and infection. AMPs have been investigated as new antibiotic agents,[5] and more than 2600 peptides have been isolated that display antimicrobial activity from a broad range of organisms including bacteria and mammalian cells.[6] While the main mode of action causing bacterial death is due to the 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 due to the massive size of possible sequence space[9]. Conventionally *in vitro* evolutionary methods, such as mRNA display and liposome display[10,11], are employed 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], 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, analogues 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 AMPs design since 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 improving knowledge-based AMP designs, machine learning-based approaches, such as artificial neural networks model combined with chemoinformatic methods, were employed to capture more complex motifs of antimicrobial activities.[7,17,18] Indeed, such approaches have proved important applied to drug design.[19] While 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 an experimentally characterized feedback have allowed optimization of molecules with desired biological or chemical activities from small starting libraries.[21,22] While this method has been applied to optimize relatively short peptides,[23,24] the optimization of longer peptides 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 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 object (Scheme 1) including functional biomolecules. The method stats with a (population of) target object(s), which are physically synthesized. They are then ranked according to evaluation results by a assay method, and the top ones are used to produce a new population for the next generation.

To demonstrate its capability, we apply this method to discover effective AMPs starting from a natural AMP. Although we used a similar methodology to evolve physico-chemical properties of oil-in-water droplets successfully,[25] application to AMP evolution is not straightforward as it could be hampered by other factors, such as high costs for peptide synthesis. To circumvent this issue, we here combine 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 coupling evolutionary algorithms and machine learning for discovery of AMPs as follows: Firstly, the interactive process of *in silico* prediction by a machine learning model and experimental assay was employed to screen better AMP candidates. This is initiated by the selection of a target sequence, generation of a population, followed by physical synthesis of each of the members of the population for testing. In contrast, the virtual screening of potential candidates was performed using existing knowledge or database (e.g. structure-activity relationship) for AMPs during evolutionary optimization in the previously proposed algorithms[18]. Due to this all-*in-silico* optimization process, such algorithms are inevitably database-dependent. On the other 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 online *in vitro* assay through the iterative process. For this reason, our algorithm can be a general methodology applicable to any functional polymers, not just for AMPs, that can perform optimization by bootstrapping the assay to synthesis process. Furthermore, as a control experiment, we compare the evolutionary algorithm-based optimizations guided by predictions from 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 accelerates the evolutionary 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 number of chemical synthesis required, which is generally the most expensive part of the optimization process.

As a proof-of-concept study, we started from a natural AMP and show that the algorithm can successfully identify peptide sequences with improved antimicrobial activity against *Escherichia coli* (*E. coli*). The optimization of AMPs proceeds as follows (Scheme 2): In the pre-optimization process (left), two peptide libraries, Generation A (G-A) and B (G-B) were identified *in silico* based on a natural AMP, which we refer to as wild-type (WT) sequence hereafter. The peptides were synthesized by automated solid-phase method and evaluated *in vitro*. The results were analyzed to determine approximate fitness values (i.e. expected improvements in antimicrobial activity) of individual amino acid substitutions using generalized linear model. In the optimization process (right), a fitness matrix was prepared using the calculated fitness values and used 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 goes as follows: At each round, a new library of peptides was generated based on predictions from the generalized linear model and evaluated experimentally. The antimicrobial activity of peptides improved rapidly using this method, within a relatively small number of experimental evaluations. 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 optimization process. In the pre-optimization process (Scheme 2, left column), a library of peptide sequences, called Generation A (G-A), was created using *position specific interactive basic local alignment search tool* (PSI-BLAST), which iteratively searches sequences in the protein databases similar to a sequence of interest (see Supplemental Experimental Procedures) .[27] We used PSI-BLAST rather than BLAST because the former may find distantly related, functionally similar sequences. We selected 93 sequences through 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 as the WT sequence were selected. All amino acid substitutions in G-A were ranked by the frequency and Generation B (G-B) was generated by applying the most frequent 93 single amino acid substitutions to the WT sequence. Generation A and B were then synthesized and their antimicrobial activities were evaluated *in vitro* by measuring the half maximal inhibitory concentration (IC_{50}) against the *Escherichia coli* (MG1655 strain). The *in vitro* evaluation data was used 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 expected decrease in IC_{50} value and is presented as a value in a fitness matrix.

In the optimization process (Scheme 2, right column), a library of new 90 peptide sequences was generated *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 synthesized (Figure S7-S8) and evaluated experimentally (refer to the Supplemental Experimental Procedures for details). Then, the fitness matrix was updated by performing regression on all the *in vitro* data. The optimization process was repeated until the IC_{50} improvement saturated, for three generations. In order to assess the effect of the fitness matrix, a control optimization experiment was performed for three generations by introducing random amino acid substitutions instead of predicting advantageous substitutions using the matrix. The IC_{50} values of peptides in the optimization experiment are shown as box plots in Figure 1A along with that of the WT peptide (81.0 μM, dashed line). The average changes in IC_{50} values relative to the WT peptide are shown as bar plots in Figure 1B, and the WT and the most potent sequences in individual generations are in Table 1. In Generation A and

B, most peptides were less potent than the WT peptide, demonstrating 9.3 and 6.7-fold higher average IC_{50} values 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, while most of the peptides in G1 showed weaker antimicrobial activities than the WT peptide. In the second generation (G2), more peptides demonstrated improved antimicrobial activities, and the average IC_{50} value decreased to 71.5 μM. The lowest IC50 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 activities. The average IC_{50} value was significantly improved (12.4 μM), achieving 6.5-fold decrease relative to the WT peptide. The most potent peptide was found in G2 with 162 times lower IC_{50} value (0.50 μM) than the WT peptide. This peptide is referred to as the 'best peptide' hereafter. Other peptides with strong antimicrobial activities were also identified (Table S1); new 44 peptides (29.1%) out of 141 tested peptides 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). At each generation, the fitness matrix was re-calculated to predict which amino acid substitution (at which locus) would be likely to improve the antimicrobial activity. When predicting G3 substitutions using the data from G-A, G-B, and G1-2, the maximum of the fitness matrix values sharply dropped compared to the previous generation. The average value also decreased. As the fitness matrix values can be interpreted as "expectation" for each amino acid substitution to improve the activity, this data suggests that the model predicted lesser 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 are likely to improve as AMP, or less likely to decrease the efficacy. Owing to the implementation of 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, new potentially good substitutions were introduced into the peptide to generate a set of peptides for the next generation. As a result, the efficacy data for G3 peptides showed smaller standard deviation compared to the previous one. To further investigate the convergence of the optimization process, we performed additional experiment. First, 20 amino acid substitutions that are likely to increase the antimicrobial activity was identified by screening the IC_{50} data for all the sequences synthesized in the previous generations (see Table S3). A new set of 39 different sequences were generated by introducing the selected substitutions randomly to the WT sequence. We refer to this peptide set as "strong substitutions". We found that the IC_{50} values of these peptides varied from 1.73 to 1155.3 μM with the average IC_{50} value 40.8 μM (2-fold increase compared to the WT peptide). As the new peptide set did not show any improvements compared to the best peptide in G2, this would support that the optimization converged within the three rounds of optimization, together with 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 tested peptides showed IC_{50} values lower than 4.1 μM, up to *c.a.* 60 times 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 times higher than the WT peptide. However, IC_{50} of the most potent peptides in the control generations were comparable (~1.3 μM, Table 1) with the ones in the 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 to identify a peptide with high antimicrobial activity. However, due to the nonlinear effect of amino acid substitution, the entire population in the control generation showed a diverse antimicrobial activity levels. This also illustrates that the fitness matrix indeed played an important role in the search for multiple potent AMPs, not just a best one.

The physicochemical properties of all the peptides evaluated were calculated to investigate the mechanism behind the improvement of 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 (Figure S1-3), the results indicated that finding explicit correlations between them are not trivial task, especially because of many outliers in the plots (Figure S3). Among others, net charge and hydrophobic moment, the physicochemical properties known to be important for antimicrobial activities, showed relatively good correlation with IC_{50} values. Average changes in net charge and hydrophobic moment compared to 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, respectively,[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 takes a helical structure. 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 in hydrophobic moment (0.74) compared to the WT sequence (Table 1 and Figure S3B). To confirm this, *de novo* structural reconstructions were performed 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 (Figure 3A, Figure S4). This secondary structure changes were confirmed using circular dichroism (CD) spectroscopy. While the CD spectrum of the WT sequence and other sequences in G1 and G2 indicated a random coil structure, that of the best peptide showed a typical profile of an α-helical structure (Figure 3B, Figure 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 a 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 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, opposite of the hydrophobic sector, while 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 location 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 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, the low hemolytic activity was confirmed experimentally using the best peptide and red blood cells, which showed only ~1% lysis even at higher concentration than the IC_{50} (Figure S9) and ToxinPred also predicted that all the peptides in Table 1 were non-toxic. We speculate this improved peptide activity might be related to a fact that this optimization process showed similar evolutionary steps 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 obtained by serially culturing the WT strain (the same strain used for *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, a new approach to the design of AMPs was developed by combining an evolutionary algorithm, machine learning-based prediction, and *in vitro* bacterial assays. This method rapidly improved antimicrobial activity, achieving 162 times increase compared to the original peptide within three generations. The best peptide identified here has one of the lowest IC_{50} 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 a 20-fold decrease in IC_{50} values compared to the WT peptide. This discovery of multiple potent peptides is remarkable considering the number of peptides tested through the pre-optimization and the optimization processes (291 out of 8×10^{16} possible peptides). As the other factors such as low cytotoxicity are also crucial as 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 owing to its 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, it also should be noted that the optimization method can adapt other measures, such as IC_{50} of multi-drug pathogenic bacteria, hemolytic potency and cytotoxicity, to calculate fitness matrix. In addition, it is also possible to employ multiple measures, not just a single measure, by performing multivariate regression with 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 owing to the iterative nature of the process, coupling *in silico* algorithm and experimental validation. This makes a stark contrast with preceding all *in silico* algorithms where a database for optimized materials (e.g. AMPs) 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 using 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 purposes, because they have only moderate direct antimicrobial activity which work 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 explore broader sequence space of peptides or other polymers.[39] For example, incorporation of new components such as non-canonical amino acids (ncAA) would improve the chemical diversity of AMPs.[40,41] Discovery and optimization of AMPs containing ncAA with conventional *in silico* methods are currently challenging as they rely on existing database for physicochemical properties or efficacy. However, our method proposed here does not require any prior information to perform evolutionary optimization, and hence may open up a path to incorporate 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 for AMPs and other properties from entirely random sequences.

EXPERIMENTAL PROCDURES

Since the details of *in silico* optimization algorithm and *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 and Supplemental Data Items consisting of 11 Figures and 3 Tables.

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.G.M., V.K. T.H. and S.T. programmed the *in silico* algorithms. M.Y., S.T. and S.G.R. performed *in vitro* bacterial and hemolysis assay. M.Y. and Y.M.A. performed CD spectroscopy assay. R.G.M., V.K., Y.M.A. D.C. and J.S.M. synthesized peptides and analyzed in HPLC and MS. M.Y., S.T., S.G.R. analyzed the collected experimental data. L.C., S.T., M.Y., and Y.M.A. co-wrote the manuscript with inputs from all co-authors. All authors co-wrote the supplementary information.

DECLARATION OF INTERESTS

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

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the EPSRC (Grant Nos EP/H024107/1, EP/I033459/1, EP/J00135X/1, EP/J015156/1, EP/K021966/1, EP/K023004/1, EP/K038885/1, EP/L015668/1, EP/L023652/1), BBSRC (Grant No. BB/M011267/ 1), EC (projects 610730 EVOPROG, 611640 EVOBLISS), ERC (project 670467 SMART-POM), University of Glasgow for LKAS fellowship, and Honjo International Scholarship Foundation. The authors thank Prof. Sharon Kelly for using the CD instrument.

REFERENCES AND NOTES

1. Wilson, R.C., and Doudna, J.A. (2013). Molecular Mechanisms of RNA Interference. Annu. Rev. Biophys. *42*, 217–39.

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*, 845–8.

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., and Gallo, R.L. (2016). Antimicrobial Peptides. Curr. Biol. *26*, R14–9.

5. Brogden, KA. (2005) Antimicrobial Peptides: Pore Formers or Metabolic Inhibitors in Bacteria? Nat. Rev. Microbiol. *3*, 238–50.

6. Chaudhary, K., Kumar, R., Singh, S., Tuknait, A., Gautam, A., Mathur, D., et al. (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., et al. (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*, 146–50.

9. Currin, A., Swainston, N., Day, P.J., Kell, D.B. (2015). Synthetic Biology for the Directed Evolution of Protein Biocatalysts: Navigating Sequence Space Intelligently. Chem. Soc. Rev. *44*, 1172–239.

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. U. S. A. *98*, 3750–5.

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. *110*, 16796–801.

12. Bornscheuer, U.T., and Pohl, M. (2001). Improved Biocatalysts by Directed Evolution and Rational Protein Design. Curr. Opin. Chem. Biol. *5*, 137–43.

13. Blondelle, S.E., and Houghten, R.A. (1992). Design of Model Amphipathic Peptides Having Potent Antimicrobial Activities. Biochemistry. *31*, 12688−94. doi:10.1021/bi00165a020.

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*, 358–64.

15. Tossi, A., Tarantino, C., and Romeo, D. (1997). Design of Synthetic Antimicrobial Peptides Based on Sequence Analogy and Amphipathicity. Eur. J. Biochem. *250*, 549–58.

16. Loose, C., Jensen, K., Rigoutsos, I., and Stephanopoulos, G. (2006) A Linguistic Model for the Rational Design of Antimicrobial Peptides. Nature. *443*, 867–9.

17. Fjell, C.D., Jenssen, H., Hilpert, K., Cheung, W.A., Pant, N., Hancock, R.E.W., et al. (2009). Identification of Novel Antibacterial Peptides by Chemoinformatics and Machine Learning. J. Med. Chem. *52*, 2006−15.

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*, 131−41.

20. Small, B.G., McColl, B.W., Allmendinger, R., Pahle, J., López-Castejón, G., Rothwell, N.J., et al. (2011). Efficient Discovery of Anti-Inflammatory Small-Molecule Combinations using Evolutionary Computing. Nat. Chem. Biol. *7*, 902–8.

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., 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*, 3128−32.

23. Singh, J., Ator, M.A., Jaeger, E.P., Allen, M.P., Whipple, D.A., Soloweij, J.E., et al. (1996). Application of Genetic Algorithms to Combinatorial Synthesis: A Computational Approach to Lead Identification and Lead Optimization. J. Am. Chem. Soc. *118*, 1669−76.

24. Yokobayashi, Y., Ikebukuro, K., McNiven, and S., Karbe. (1996). Directed Evolution of Trypsin Inhibiting Peptides using a Genetic Algorithm. J. Chem. Soc. Perkin. Trans. *20*, 2435−7.

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., et al. (2010). Five Novel Antimicrobial Peptides from Skin Secretions of the Frog, Amolops Loloensis. Comp. Biochem. Physiol. Part B Biochem. Mol. Biol. *155*, 72–6.

27. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. Nucleic Acids Res. *25*, 3389–402.

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, 475–85.

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*, 371–4.

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*, W449–54.

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*, 1008–12.

32. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G.P.S., et al. (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. Chemie. Int. Ed. *50*, 10686–9.

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., et al. (2016). DBAASP v.2: An Enhanced Database of Structure and Antimicrobial/Cytotoxic Activity of Natural and Synthetic Peptides. Nucleic Acids Res. *44*, D1104–12.

36. Ganz, T. (2003). Defensins: Antimicrobial Peptides of Innate Immunity. Nat. Rev. Immunol. *3*, 710–20.

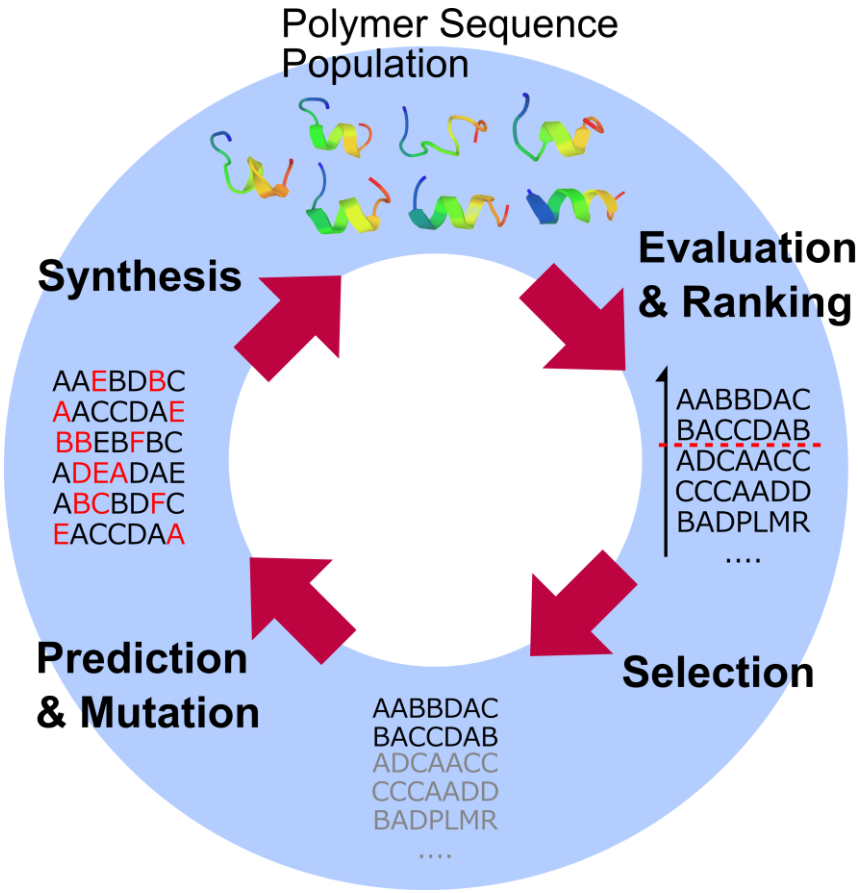
37. Hancock, R.E.W., and Sahl, H.G. (2006). Antimicrobial and Host-Defense Peptides as New Anti-Infective Therapeutic Strategies. Nat. Biotechnol. *24*, 1551–7.

38. Mansour, S.C., Pena. O.M., Hancock, R.E.W. (2014). Host Defense Peptides: Front-Line Immunomodulators. Trends Immunol. *35*, 443–50.

39. Frederix, P.W.J.M., Scott, G.G., Abul-Haija, Y.M., Kalafatovic, D., Pappas, C.G., Javid, N., et al. (2014). Exploring the Sequence Space for (Tri-)Peptide Self-Assembly to Design and Discover New Hydrogels. Nat. Chem. *7*, 30–7.

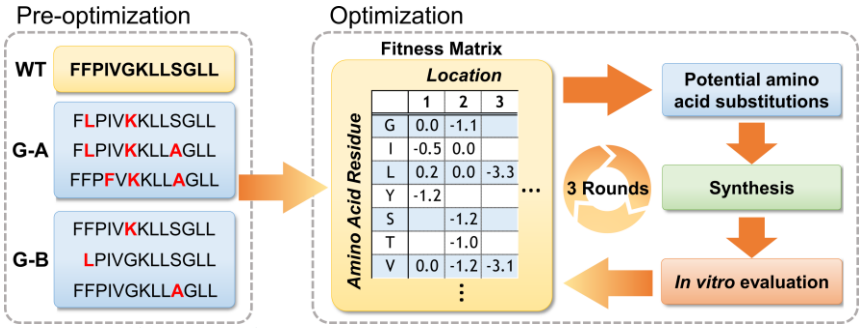
40. Molhoek, E.M., Van Dijk, A., Veldhuizen, E.J.A., Haagsman, H.P., and Bikker, F.J. (2011). Improved Proteolytic Stability of Chicken Cathelicidin-2 Derived Peptides by D-Amino Acid Substitutions and Cyclization. Peptides. *32*, 875–80.

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.



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 are used as the starting 'Polymer Sequence Population' and this forms the sequences experimentally synthesized in a peptide synthesis robot. The polymer sequence activities recorded then undergo analysis against a user desired property (*e.g.* IC_{50}) in the 'Evaluation' step. The sequences are ranked in terms of desired property automatically, and the least good rejected in the 'Ranking' step allowing a new population to be 'Selected'. Meanwhile the accepted formulations are used as a basis to create a new 'Sequence Populations' after random 'Mutation' and 'Crossover'.



Scheme 2. Workflow to search for potent antimicrobial peptides.

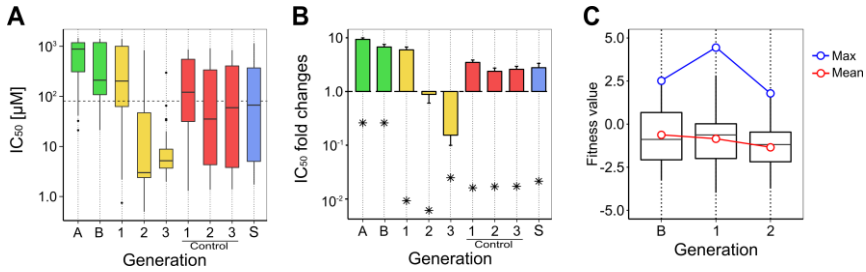


Figure 1. Optimization of antimicrobial peptides.

(**A**) IC_{50} values of all the peptides in each generation are shown in a box plot. Each solid line in the box plot represents the median of IC_{50} values. Generation A and B, generation one to three, strong substitutions, and control generation one to three are shown in green, yellow, blue and red, respectively. The IC_{50} value of the WT peptide is shown in a dashed line. (**B**) Fold changes of IC_{50} values compared to the WT sequence. Bar plots indicate average changes of IC_{50} values with standard errors. The range of activity of the most potent peptides in each generation is shown as asterisk. (**C**) A box plot 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.

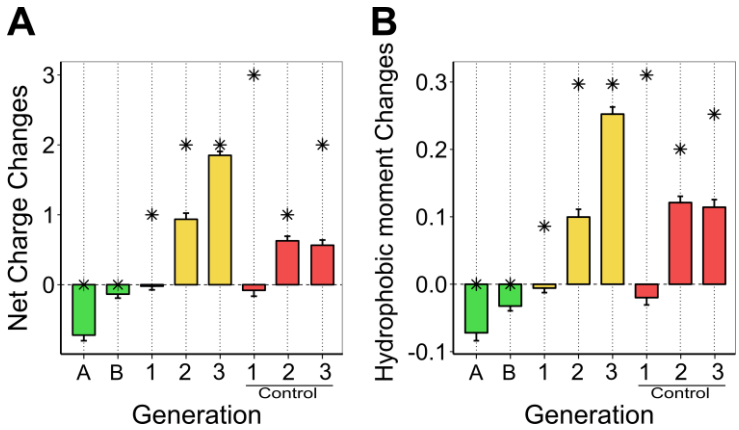


Figure 2. Changes of physicochemical properties of peptides.

The properties of peptides were subtracted by those of the WT peptide. Average changes of (**A**) net charge and (**B**) hydrophobic moment in individual generations are shown in bar plots with standard errors. Changes of the most potent peptides in individual generations are shown as asterisks. Generation A and B, 1 to 3, and control generation 1 to 3 are shown in green, yellow and red, respectively.

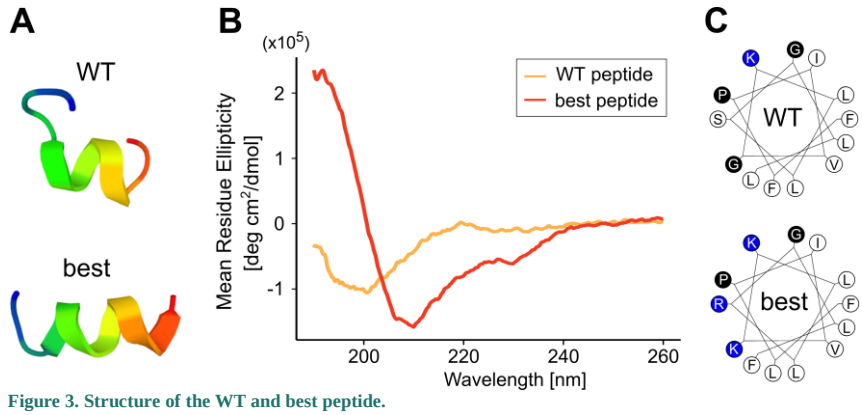


Figure 3. Structure of the WT and best peptide.

(**A**) *De novo* structural reconstruction using PEP-FOLD3. (**B**) Structural characterization using circular-dichroism spectroscopy. The spectrum 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. The one letter codes for amino acids were used.

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	FFPIVGKLLSGL**F**	21.1	0.98	0.45	0.54
B	FFPIVGKLLSGL**F**	21.1	0.98	0.45	0.54
1	FFPIV**K**KLLSGL**F**	0.75	1.98	0.53	0.58
2	F**L**PIV**K**KLL**R**GL**F**	0.50	2.98	0.74	0.55
3	**VL**PIV**K**KLL**K**GL**F**	2.01	2.98	0.64	0.50
1C	F**L**PIV**K**KLL**RK**L**F**	1.30	3.98	0.76	0.52
2C	FFPI**F**GKLL**R**GL**F**	1.37	1.98	0.65	0.52
3C	FFPIVGKLLR**K**L**F**	1.39	2.98	0.70	0.57

The amino acid residues substituted from the WT sequence are underlined and shown in bold. Net charge and hydrophobic moment were calculated using Peptides package on R. Hemolytic potency was predicted using HemoPI[29] and lower values indicate that peptides are less hemolytic.