Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Toward computer-made artificial antibiotics

Marcelo Der Torossian Torres[1,2] and Cesar de la Fuente-Nunez[1,2]

Merging concepts from synthetic biology and computational biology may yield antibiotics that are less likely to elicit resistance than existing drugs and that yet can fight drug-resistant infections. Indeed, computer-guided strategies coupled with massively parallel high-throughput experimental methods represent a new paradigm for antibiotic discovery. Infections caused by multidrug-resistant microorganisms are increasingly deadly. In the current post-antibiotic era, many of these infections cannot be treated with our existing antimicrobial arsenal. Furthermore, we may have already exhausted the category of large molecules produced in nature having antimicrobial activity: the antibiotic scaffolds we have discovered so far may represent the majority of those that exist. The rise in drug-resistant bacteria and lack of new antibiotic classes clearly call for out-of-the-box strategies. Recent advances in computational synthetic biology have enabled the development of antimicrobials. New molecular descriptors and genetic and pattern recognition algorithms are powerful tools that bring us a step closer to developing efficient antibiotics. We review several computational tools for drug design and a number of recently generated antibiotic candidates, with an emphasis on peptide-based molecules. Design strategies can generate a diversity of synthetic antimicrobial peptides, which may help to mitigate the spread of resistance and combat multidrug-resistant microorganisms.

Addresses
[1] Machine Biology Group, Departments of Psychiatry and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, United States of America
[2] Department of Bioengineering, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

Corresponding author: de la Fuente-Nunez, Cesar (cfuente@pennmedicine.upenn.edu)

Current Opinion in Microbiology 2019, **51**:30–38

This review comes from a themed issue on **Antimicrobials**

Edited by **Mattew I Hutchings, Andrew W Truman and Barrie Wilkinson**

For a complete overview see the Issue and the Editorial

Available online 11th May 2019

https://doi.org/10.1016/j.mib.2019.03.004

Introduction

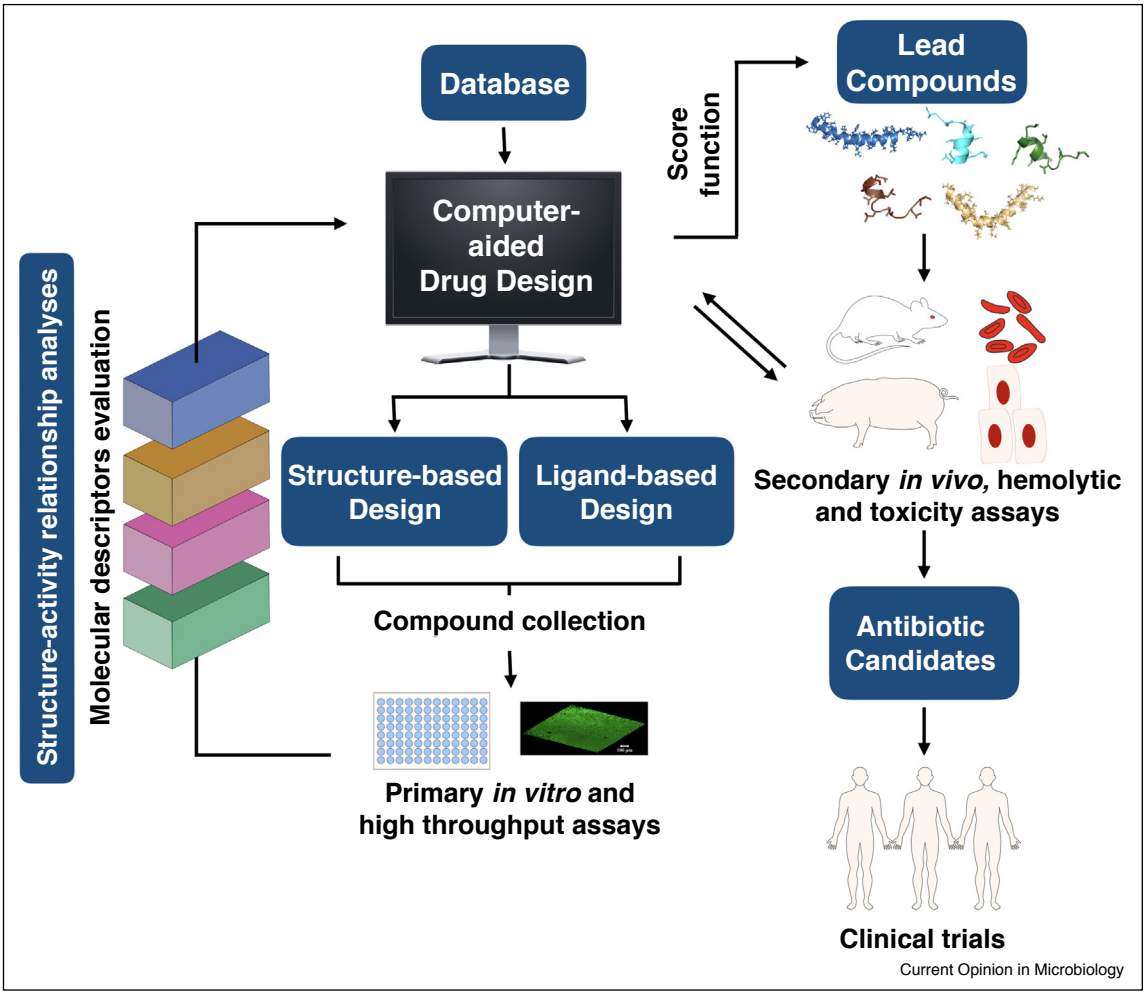
Computer-guided drug design has been proposed as a promising approach for identifying new drugs [1]. The emergence of improved combinatorial synthesis and high-throughput screening methods has enabled systematic modifications of template molecules for drug discovery, requiring minimal compound design or prior knowledge. However, such approaches are not always accurate and, so far, have had low yields [2]. Recent developments have improved both the synthesis of combinatorial libraries of compounds and the screening of such molecules for biological function so that now these can be done in a time-effective and cost-effective manner. Such technological advances currently enable production of large datasets that can be analyzed by computer-based methodologies [3]. The lack of truly new antibiotic classes calls for inventive strategies to combat multidrug-resistant organisms [4], which have been predicted to kill 10 million people annually by 2050 [5]. The urgent need for more effective antibiotics [6] has already prompted the development of biological activity descriptors and methodologies such as support vector machines (SVM) and neural networks for their application [7] (Figure 1). Such approaches may enable the creation of sturdier antibiotics that do not elicit resistance on the part of the targeted bacteria.

Computer-guided antibiotic design

Novel compounds can be made from templates, by slightly modifying the functional groups of existing molecules or by attaching fragments to produce new active chemotypes [8]. The computationally guided design of molecules usually requires datasets with considerable information as input for filtering the most relevant pharmacological properties, that is, those that define specific functions and the potency of these compounds [3]. In fact, the determination of which properties should be optimized is the key step in designing novel compounds. Parameters that guide the optimization of lead molecules, whether to increase their selectivity and affinity or tune other biological functions, include pharmacokinetic properties such as absorption, distribution, metabolism, excretion, and the potential for toxicity [9].

Depending on the presence or absence of reliable structural data, computational-based drug design can be classified into two general categories: ligand-based and structure-based designs. The ligand-based

Figure 1



Computer-aided drug design. Large databases serve as input for filtering the most relevant pharmacological properties that define the specific functions and potency of these compounds. Independently of the approach, the data generated are used to plan a set of novel compounds that are subsequently tested, and the results obtained are used to analyze new properties. The steps are repeated until a reliable score function leads to optimized antibiotic candidates, which are then evaluated according to their *in vivo* activity and toxicity. The lead compounds generated by this process are promising candidates for clinical trials.

approach is generally preferred when little structural information, or none, is available. This method consists of indirectly designing molecules by having prior information that informs what might bind to the target of interest. Preferably, molecular design is followed by quantitative structure–activity relationship studies, correlating the most important biological descriptors to some extent of biological data. The structure-based approach, contrastingly, relies on the knowledge of the structure, generally, to calculate interaction energies [1]. Small organic molecules, peptides, proteins, and other classes of compounds may be optimized by both ligand and structure-based methods [10–12]. Computational-made design techniques have focused on peptides and proteins, and these molecules still likely stand as the primary class of biomacromolecules to be developed and explored [13], mostly because they are versatile and have multifunctional properties [14]. However, other notable antimicrobial agents have been recently reported, such as the 32 new 3,9-disubstituted eudistomin U derivatives reported by Dai *et al.* [15], which were designed and synthesized based on computer-aided drug discovery. Sixteen of the 3,9-disubstituted eudistomin U derivatives were shown to exhibit potent antibacterial activity. The most active compound, 6p, displayed higher activity (1.56 μmol L^{-1}) than the commercial drugs fosfomycin, ciprofloxacin, and propineb. The antibacterial mechanism of the 3,9-disubstituted eudistomin U derivatives involves a bactericidal effect caused by simultaneously damaging the bacterial cell membrane and disrupting the function of DNA gyrase.

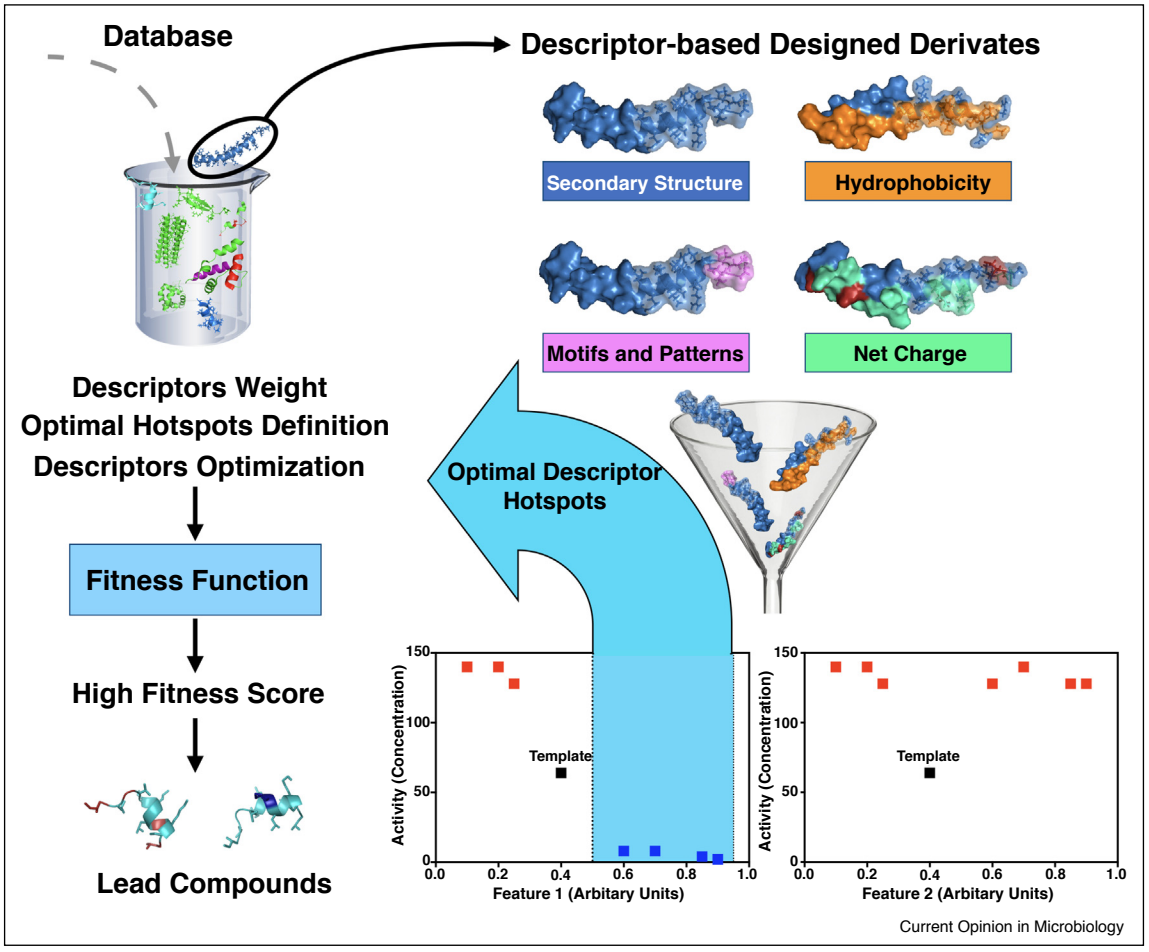
Computationally guided exploration of peptide antibiotics

Peptides are natural or synthetic moleules that present a wide range of biological functions and so are not easily described as a single category. Currently several interesting approaches are being developed with the overarching goal of elucidating the modes of action of biologically active peptides [11]. Peptides have been reported as promising antibiotic candidates because of their antimicrobial [6,16], immunomodulatory [17], and synergistic activities with other classes of antibiotics [18].

Several techniques are being used to describe the features of peptides using computational algorithms that balance efficiency and information content [19]. The optimal descriptor set depends on the biological function predicted, as well as on the ligand-based technique used [8], and, therefore, many different algorithms for deriving chemical information have been developed and used [10,19–21] (Figure 2). Molecular descriptors can be structural as well as physicochemical and have multiple levels of increasing complexity. Information described includes molecular weight, geometry, volume, surface areas, aromatic ring content, rotatable bonds, interatomic distances, bond distances, atom types, planar and nonplanar systems, molecular walk counts, electronegativities, polarizabilities, symmetry, atom distribution, topological charge indices, functional group composition, aromaticity indices, and solvation properties, among other characteristics [22]. These descriptors are generated through knowledge-based, graph-theoretical methods, molecular mechanical, or quantum-mechanical tools [23,24]. The descriptors are classified according to scalar physicochemical properties such as molecular weight and bi-dimensional and three-dimensional molecular constitution-derived descriptors.

The various levels of complexity, however, overlap with the more complex descriptors, often incorporating information from simpler ones. For example, many

Figure 2



Importance of physicochemical descriptors for antibiotic development. Physicochemical features of each element of a representative dataset are extracted, and their influence on biological activity is reported. Once optimal hotspots are defined, fitness function prediction can be made more accurate, and lead high-ranked compounds can be generated that are more selective and have other enhanced biological activities.

bi-dimensional and three-dimensional descriptors use physicochemical properties to weight their functions and to describe the overall distribution of these properties. Lee *et al.* [25] reported the importance of having different types of descriptors in a machine learning-enabled discovery and design model that generates antimicrobial peptides (AMPs) that disrupt bacterial membranes. According to the authors, net positive charge, length, composition, and solvent access to residues are the most important descriptors to favor peptide–membrane interactions and promote membrane destabilization. Similar structural and physicochemical descriptors have also been shown to be important for biological function in other structure-function design studies in the literature [16,26••,27••].

Several computational methods are being used to discover, design, and optimize AMPs (Figure 3). Among these, comparative modeling, predictive models, and hybrid models currently stand out because of their accuracy and usefulness.

Comparative modeling

Encrypted templates might be identified from complex natural structures, such as proteins and enzymes. Frequently, proteolytic processes of large precursors are potential sources of encrypted bioactive molecules; however, most of the cleaved parts do not necessarily present biological activity. Pane *et al.* [27••], for instance, described a computational-experimental framework for the discovery of novel cryptic AMPs after validating the antimicrobial scoring functions of the activation peptide of pepsin A, the main human stomach protease, and its N-terminal and C-terminal fragments as AMPs. The three peptides from the pepsinogen A3 isoform were prepared in a recombinant form using a fusion carrier specifically developed to express potentially toxic peptides in *Escherichia coli*. Recombinant pepsinogen A3-derived peptides proved to be wide-spectrum antimicrobial agents with MIC values ranging from 1.56 to 50 $\mu mol\,L^{-1}$, comparable to those of the whole activation peptide (1.56–12.5 $\mu mol\,L^{-1}$), against the same organisms. Moreover, the activation peptide was bactericidal at pH 3.5 for relevant foodborne pathogens, suggesting that this new class of previously unexplored AMPs may contribute to microbial surveillance within the human stomach. The cryptic peptides showed no toxicity toward human cells but exhibited anti-infective activity *in vivo*, reducing by up to four orders of magnitude the bacterial load of *Pseudomonas aeruginosa* PAO1 in a mouse skin infection model.

Predictive models

Several approaches to predicting the biological functions of proteins and peptides have been reported over the last few decades [12]. Numerous studies regarding protein/peptide folding, for example, have contributed substantially to unraveling the role of biophysical, biochemical, and physicochemical properties in structure–activity related processes. Understanding these features should enable the design of customized proteins and peptides and lead to greater knowledge of how proteins exert their biological functions. The *de novo* protein design technique detailed by Huang *et al.* [28] epitomizes how the exploration of the full sequence space, guided by appropriate physicochemical principles and descriptors, can contribute to the understanding of protein folding. The authors report the most relevant methods for designing proteins: structure prediction, fixed-backbone, and *de novo*, highlighting *de novo* protein design. In structure prediction, the sequence is fixed and the backbone structure is unknown; in fixed backbone protein design, the sequence is unknown but the structure is fixed; and in *de novo* protein design, neither is known. The *de novo* design approach might be an alternative for the design of new functional protein and peptides without involving modification and optimization of natural templates.

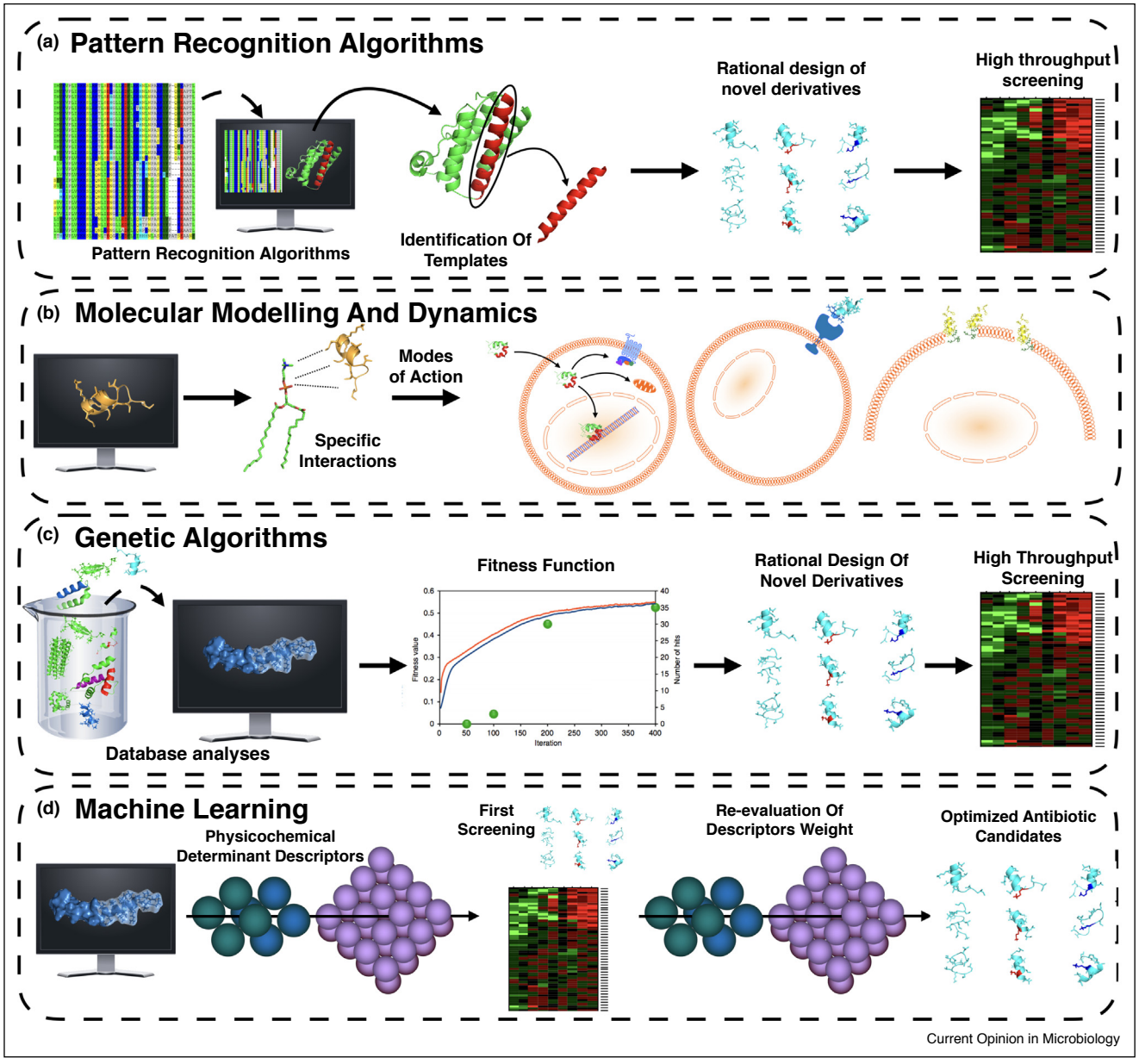
Recent computer science advances such as genetic algorithms, machine learning, and deep learning have enabled other useful tools for the rational prediction of antimicrobial proteins and peptides. Here, we briefly review some of the practical applications of these techniques.

Predictive models: genetic algorithms

Another approach for designing AMPs is by using genetic algorithms that harness the basic principles of Darwinian natural selection to 'evolve' molecules toward a desired biological function [29]. This approach may be used to classify virtually any new AMP sequences through fitness functions based on activity descriptors and information collected from AMP databases. Despite the redundancy of sequences generated by genetic algorithms, this technique is capable of identifying novel artificially generated AMPs with distinct composition and function. Porto *et al.* [26••] described the use of plant-derived templates for the computer-aided design of the guava glycine-rich peptide, Pg-AMP1, which was used as a template to generate synthetic guavanin peptides by means of a genetic algorithm. The specific modifications determined by the algorithm, including a descriptive fitness function, as well as the interruption of the algorithm before reaching a plateau (i.e. an optimal solution), generated truly innovative peptides whose design was driven by the exploration of another parcel of combinatorial sequence space. The most promising analog designed, guavanin 2, was bactericidal at low concentrations, causing the disruption of *P. aeruginosa* bacterial membranes by triggering their hyperpolarization.

Genetic algorithms based on simple fitness functions may also be coupled to other design models such as the one described by Pane *et al.* [30], with which the authors predicted cationic AMPs based on linear correlations

Figure 3



Computational design principles for antibiotic peptides. Peptides with antibiotic properties have been recently explored through multiple computational techniques. Combinations of these methodologies might also be applied. The **(a)** pattern recognition algorithms are important tools used to discover bioactive templates that are encrypted into natural biomolecules by comparing them to known bioactive molecules. This technique enables the template-based rational design of novel compounds that is usually conducted by **(b)** molecular modelling and molecular dynamics, which are very effective ways of analyzing structure–activity relationship. **(c)** Genetic algorithms are also used for generating antibiotic candidates from databases through physicochemical descriptors that are ranked in a fitness function. This function classifies the novel molecules generated and points to lead compounds for high-throughput screening. **(d)** Machine learning is one of the most recent techniques being applied to computational drug design. Machine learning uses statistical techniques to give computer systems the ability to learn and progressively improve their performance as far as generating bioactive compounds from physicochemical and biological activity data.

among net charge, hydrophobicity, and length. Output datasets generated in such studies may then be used as an initial population that can be evolved and biologically optimized via genetic algorithms. The authors also attribute the relative contributions of each physicochemical parameter to the antimicrobial potency in a species-specific manner. The most important findings from this work are that some bacterial strains are sensitive to highly charged peptides, whereas others are particularly susceptible to more hydrophobic peptides. Similarly, molecular structure

and, specifically for peptides and proteins, conformer disposition might be assessed through algorithms. Supady *et al.* [31] reported the identification of low-energy conformers by employing a genetic algorithm that searched segments of the conformation space for molecules with lower energy profiles. The authors aimed to predict all conformers within an energy window above the global minimum instead of finding the global minimum energy. Well-established first principles methods of molecular structure for drug design were used to evaluate a dataset extracted from a database consisting of amino acid dipeptide conformers, and the performance of a systematic search was compared with that of a random conformer generator. The algorithm was capable of accurately reproducing the reference data and the low-energy conformational space coverage of the genetic algorithm exceeded two competing methods selected by the authors significantly at similar effort.

Predictive models: machine learning

Effective approaches are needed for interpreting and integrating data increasingly available in massive data collections generated by high-throughput screens [32]. Machine learning is currently one of the most active areas of research in computer science, with broad applications in the fields of biological engineering and synthetic biology [33]. Yoshida *et al.* [19] presented a proof-of-concept methodology for efficiently optimizing the efficacy of AMPs. In particular, the authors used a closed-loop approach that combines a genetic algorithm, machine learning, and *in vitro* evaluation to improve the antimicrobial activity of peptides against *E. coli*. The authors identified 44 lead peptides from a small cationic natural template. The hits were up to 160-fold more active than the wild-type molecule after three cycles of predictions. The authors were also able to convert unstructured peptides into well-defined helical molecules, which were more active than the wild-type. The results obtained by the authors exemplify how machine learning can provide tools and accelerate the discovery of AMPs with promising antimicrobial activities, allowing the exploration of structural patterns and indicating how particular descriptors influence biological activity.

Machine learning-based design models might also be aimed at creating AMPs with certain mechanisms of action, as shown by Lee *et al.* [34•], who developed a support vector machine (SVM)-based classifier to investigate the activity of α-helical AMPs with activity on bacterial membranes. The model related functional commonality to sequence homologies. The authors used a SVM to design new AMPs; the machine took into consideration α-helicity as determined by X-ray scattering and *in vitro* antimicrobial activity. Lee *et al.* also observed that their model could generate negative Gaussian membrane curvature as an indirect measure of antimicrobial activity, which may be a very useful tool as it would provide a topological basis for the membrane activity common to many AMPs.

Predictive models: deep learning

Hierarchically embedded neural networks, also known as deep learning [35], might be the most promising of currently available artificial intelligence and computer-guided design techniques. Whereas physicochemical or structural patterns used in genetic algorithms and machine learning are aimed at influencing antimicrobial activity only indirectly, the design of AMPs generated by deep learning usually result in direct prediction of AMP activity, a distinct advantage. Yet, only a few studies have been done so far to advance the use of deep learning in this context. In these studies, the layers of the neural network are grouped in a hierarchy generating related concepts or decision trees, so that the output data for one of the layers lead to a set of more deeply related layers that measure the direct influence, especially from sequence data, on antimicrobial activity or structure, as well as on physicochemical features.

One disadvantage of this method is that reliable biological data in large amounts serve as inputs for the self-learning process, which is an issue for AMPs because we currently lack standard procedures for testing peptides and because purified peptides needed to avoid prediction errors are expensive to synthesize. Veltri *et al.* [36•] reported a simple deep neural network with just a few layers that provided very accurate AMP recognition. The authors proposed a model for identifying position variant patterns along the peptide sequence and obtained results with higher precision than those achieved by other well-known models currently used in bioinformatics, such as modlAMP [24], AntiBP2 [37], CAMP [38], and iAMPpred [39].

Muller *et al.* [40••] presented an alternative approach for computationally-guided peptide design, which consisted of generating long short-term memory recurrent neural networks. Using this model, the authors were able to propose combinatorial *de novo* peptide design. The model acted by capturing patterns present in α-helical AMP sequences and generating new peptides from the learned context. The authors reported that 82% of the AMPs that were predicted to be active were, in fact, active, compared to 65% of randomly sampled sequences with the same amino acid distribution as the training set. These strategies offer alternatives to peptide and protein design and eliminate the need for exhaustive high-throughput screens for the generation of bioactive sequence libraries.

Hybrid models

Hybridization of the models described above has also generated new models, which tend to exhibit improved accuracy as they capture the advantages of each constituent model. The hybrid model proposed by Schneider *et al.*

[41] combines network and deep learning models. The architecture allows the use of diverse and multi-dimensional descriptors in a self-organizing map that converts the descriptors to two-dimensional images for further processing; the two-dimensional images are used as an input layer for feed-forward neural networks. This model has greater classification accuracy and predictive robustness compared to the feed-forward network classifiers that lack the layer of self-organizing maps.

Hybrid models have been also developed for protein design and prediction. Yan *et al.* [42] described a model that aims to appropriately incorporate biological information into *ab initio* docking. The authors have developed a hybrid docking protocol using template-based and template-free approaches. Briefly, the authors modeled the structures of individual components based on the template complex or by regular homology modeling, when a template was not available. The structure was then predicted by traditional protein–protein docking. By using this methodology, the authors predicted correct models for 16 CASP-CAPRI targets. They also accurately predicted protein–peptide binding for six out of eight targets.

Taking a different route, Gangopadhyay and Datta [43] coupled structure-based and ligand-based approaches, identifying ligandable sites and using the identified inhibitors as reference ligands for the design of potential inhibitors of active cholera toxin. The authors identified potential ligandable sites of the active cholera toxin by employing an energy-based approach to identify ligand binding sites; and then by comparing their findings with results of computational solvent mapping, the authors identified two potential ligandable sites that might be potential targets against cholera. Similar approaches could be used for structure-based drug design. The authors also reported the interactions of indole-based alkaloids and phosphates with residues of the ligandable region of the toxin.

Future perspectives

The extensive variety of computational tools used in drug discovery suggests that there is not one fundamentally superior design method. The performance of methods varies greatly depending on the desired target, available data, and available resources. Improvement in scoring functions is a key factor for the large-scale application of computationally aided drug design techniques, and it will involve the use (and development) of accurate descriptors and precise free energy measurements. Most likely, the combination of the structure-based and ligand-based computational approaches will provide the accuracy necessary for the successful design of novel antibiotics.

Additionally, with the development of elaborate structure and physicochemical descriptors, computational design methods are becoming more and more complex — and concomitantly less accessible. There is an urgent need to couple these methods with new advances in molecular dynamics and molecular modeling. Melo *et al.* [44], for example, created an integrated, comprehensive, customizable, and easy-to-use suite, which makes it possible to yield high-quality analyses of target proteins or peptides that generate precise output data. Combining quantum mechanics and molecular mechanics, Melo *et al.* merged the two widely used molecular dynamics and visualization softwares (NAMD and VMD) with the quantum chemistry packages ORCA and MOPAC. The authors demonstrated that interface, setup, execution, visualization, and analysis can be straightforward procedures for all levels of expertise and can be applied to generate high-quality analyses of target proteins and peptides.

Initiatives to find or build molecules that have precision and selectivity against microorganisms, and even specificity against particular strains, are the next step for predictive models; these models will play a key role in describing complex correlations between descriptors and biological activities. Vishnepolsky *et al.* [45•] recently described a predictive model of small linear AMPs that present antimicrobial activity against particular Gram-negative strains. The study showed that it is possible to accurately distinguish peptides, especially a subset composed of 18–27 residues, with specific activity against *E. coli* ATCC 25922 and *P. aeruginosa* ATCC 27853 by means of a semi-supervised machine-learning approach coupled to a density-based clustering algorithm. Veltri *et al.* [46] also reported a method that constructs and selects complex descriptors of AMPs based on sequence patterns. This method also provides a transparent summary of antibacterial activity at the sequence level by recognizing the antimicrobial activity of a peptide and predicting its target selectivity based on models of activity against Gram-positive bacteria, Gram-negative bacteria, or both.

Conclusions

There is currently an urgent need to design truly novel antibiotics while at the same time increasing the accuracy and decreasing the time and cost related to the discovery and development process. Among the available methods for the precise design of antibiotics, computational and synthetic biology tools have the most promise. The value of these tools is even greater when they are coupled to practical methodologies such as combinatorial synthesis and high-throughput screening.

Peptide design offers various opportunities not only in terms of biological applications but also for understanding how to represent the physiochemical principles underlying stability, folding mechanisms, and interactions with other molecules. Currently, the available scoring functions are biased toward hydrophobic and electrostatic

interactions to the neglect of hydrogen bonding interactions, leading to toxic, aggregation-prone sequences, with non-selective activity. Nevertheless, computational-aided drug design methods such as the ones outlined above, when guided by accurate genetic algorithms, show promising results, as they have been able to successfully generate AMPs with novel sequences, folds, topologies, and functions. Progress has accelerated at a rapid pace in the wake of new computational technologies, more affordable computational resources, and breakthroughs in the development of more accurate scoring functions and optimization algorithms. We may be heading toward a new era where we will be able to build computationally designed peptide antibiotics with selective, specific, and augmented activities for the treatment of infectious disease.

Conflict of interest statement

Nothing declared.

Acknowledgements

The figures shown here were prepared using the Motifolio drawing toolkit and PyMOL (v1.8.2.3). This work was supported by the Ramon Areces Foundation (to CFN).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Lionta E, Spyrou G, Vassilatis D, Cournia Z: **Structure-based virtual screening for drug discovery: principles, applications and recent advances**. *Curr Top Med Chem* 2014, **14**:1923-1938.
2. Macalino SJY, Gosu V, Hong S, Choi S: **Role of computer-aided drug design in modern drug discovery**. *Arch Pharm Res* 2015, **38**:1686-1701.
3. Wang C, Xu P, Zhang L, Huang J, Zhu K, Luo C: **Current strategies and applications for precision drug design**. *Front Pharmacol* 2018, **9**:787.
4. Review on Antimicrobial Resistance, Review on Antimicrobial Resistance (London), Grande-Bretagne: *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations*. Review on Antimicrobial Resistance; 2014 https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf.
5. CDC: **Antibiotic resistance threats in the United States, 2013**. *Current* 2013. doi:CS239559-B.
6. de la Fuente-Nunez C, Torres MD, Mojica FJ, Lu TK: **Next-generation precision antimicrobials: towards personalized treatment of infectious diseases**. *Curr Opin Microbiol* 2017, **37**:95-102.
7. Grisoni F, Consonni V, Todeschini R: In *Impact of Molecular Descriptors on Computational Models BT - Computational Chemogenomics*. Edited by Brown JB. New York: Springer; 2018: 171-209.
8. Beltran JA, Aguilera-Mendoza L, Brizuela CA: **Optimal selection of molecular descriptors for antimicrobial peptides classification: an evolutionary feature weighting approach**. *BMC Genomics* 2018, **19**:672.
9. Kore PP, Mutha MM, Antre RV, Oswal RJ, Kshirsagar SS: **Computer-aided drug design: an innovative tool for modeling**. *Open J Med Chem* 2012, **02**:139-148.
10. Porto WF, Pires AS, Franco OL: **Computational tools for exploring sequence databases as a resource for antimicrobial peptides**. *Biotechnol Adv* 2017, **35**:337-349.
11. Liu S, Fan L, Sun J, Lao X, Zheng H: **Computational resources and tools for antimicrobial peptides**. *J Pept Sci* 2016, **23**:4-12.
12. Porto WF, Pires ÁS, Franco OL: **Antimicrobial activity predictors benchmarking analysis using shuffled and designed synthetic peptides**. *J Theor Biol* 2017, **426**:96-103.
13. Setiawan D, Brender J, Zhang Y: **Recent advances in automated protein design and its future challenges**. *Expert Opin Drug Discov* 2018, **13**:587-604.
14. Lee EY, Lee MW, Fulan BM, Ferguson AL, Wong GCL: **What can machine learning do for antimicrobial peptides, and what can antimicrobial peptides do for machine learning?** *Interface Focus* 2017, **7**.
15. Dai J, Dan W, Li N, Wang J: **Computer-aided drug discovery: novel 3,9-disubstituted eudistomin U derivatives as potent antibacterial agents**. *Eur J Med Chem* 2018, **157**:333-338.
16. Torres MDT, Pedron CN, Higashikuni Y, Kramer RM, Cardoso MH, Oshiro KGN, Franco OL, Silva PI Junior, Silva FD, Oliveira VX Junior *et al.*: **Structure-function-guided exploration of the antimicrobial peptide polybia-CP identifies activity determinants and generates synthetic therapeutic candidates**. *Commun Biol* 2018, **1**:221.
17. Hancock REW, Haney EF, Gill EE: **The immunology of host defence peptides: beyond antimicrobial activity**. *Nat Rev Immunol* 2016, **16**:321.
18. Lázár V, Martins A, Spohn R, Daruka L, Grézal G, Fekete G, Számel M, Jangir PK, Kintses B, Csörgo B *et al.*: **Antibiotic-resistant bacteria show widespread collateral sensitivity to antimicrobial peptides**. *Nat Microbiol* 2018, **3**:718-731.
19. Yoshida M, Hinkley T, Tsuda S, Abul-Haija YM, McBurney RT, Kulikov V, Mathieson JS, Galiñanes Reyes S, Castro MD, Cronin L: **Using evolutionary algorithms and machine learning to explore sequence space for the discovery of antimicrobial peptides**. *Chem* 2018, **4**:533-543.
20. Bhadra P, Yan J, Li J, Fong S, Siu SWI: **AmPEP: sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest**. *Sci Rep* 2018, **8**:1697.
21. Porto WF, Fensterseifer ICM, Ribeiro SM, Franco OL: **Joker: an algorithm to insert patterns into sequences for designing antimicrobial peptides**. *Biochim Biophys Acta - Gen Subj* 2018, **1862**:2043-2052.
22. Mauri A, Consonni V, Todeschini R: In *Molecular Descriptors BT - Handbook of Computational Chemistry*. Edited by Leszczynski J, Kaczmarek-Kedziera A, Puzyn TG, Papadopoulos M, Reis HK, Shukla M. Springer International Publishing; 2017:2065-2093.
23. Chen Z, Zhao P, Li F, Leier A, Marquez-Lago TT, Wang Y, Webb GI, Smith AI, Daly RJ, Chou K-C *et al.*: **iFeature: a python package and web server for features extraction and selection from protein and peptide sequences**. *Bioinformatics* 2018, **34**:2499-2502.
24. Müller AT, Gabernet G, Hiss JA, Schneider G: **modlAMP: python for antimicrobial peptides**. *Bioinformatics* 2017, **33**:2753-2755.
25. Lee EY, Wong GCL, Ferguson AL: **Machine learning-enabled discovery and design of membrane-active peptides**. *Bioorg Med Chem* 2018, **26**:2708-2718.
26. •• Porto WF, Irazazabal L, Alves ESF, Ribeiro SM, Matos CO, Pires ÁS, Fensterseifer ICM, Miranda VJ, Haney EF, Humblot V *et al.*: **In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design**. *Nat Commun* 2018, **9**:1490.

The authors reported the design of active peptides generated by an optimized genetic algorithm that explored new parcels of peptide combinatorial space.

27. •• Pane K, Cafaro V, Avitabile A, Torres MDT, Vollaro A, De Gregorio E, Catania MR, Di Maro A, Bosso A, Gallo G *et al.*: **Identification of novel cryptic multifunctional antimicrobial**

peptides from the human stomach enabled by a computational–experimental platform. *ACS Synth Biol* 2018, **7**:2105-2115.
The authors reported a computaional-experimental framework for the discovery of novel cryptic antimicrobial peptides.

28. Huang P-S, Boyken SE, Baker D: **The coming of age of de novo protein design**. *Nature* 2016, **537**:320-327.

29. De Magalhães CS, Almeida DM, Barbosa HJC, Dardenne LE: **A dynamic niching genetic algorithm strategy for docking highly flexible ligands**. *Inf Sci (Ny)* 2014, **289**:206-224.

30. Pane K, Durante L, Crescenzi O, Cafaro V, Pizzo E, Varcamonti M, Zanfardino A, Izzo V, Di Donato A, Notomista E: **Antimicrobial potency of cationic antimicrobial peptides can be predicted from their amino acid composition: application to the detection of "cryptic" antimicrobial peptides**. *J Theor Biol* 2017, **419**:254-265.

31. Supady A, Blum V, Baldauf C: **First-principles molecular structure search with a genetic algorithm**. *J Chem Info Model* 2015, **55**:2338-2348.

32. Lavecchia A: **Machine-learning approaches in drug discovery: methods and applications**. *Drug Discov Today* 2015, **20**:318-331.

33. Lima AN, Philot EA, Trossini GHG, Scott LPB, Maltarollo VG, Honorio KM: **Use of machine learning approaches for novel drug discovery**. *Expert Opin Drug Discov* 2016, **11**:225-239.

34. • Lee EY, Fulan BM, Wong GCL, Ferguson AL: **Mapping membrane activity in undiscovered peptide sequence space using machine learning**. *Proc Natl Acad Sci U S A* 2016, **113**:13588-13593.
The authors reported the generation of active peptides using machine learning tools to target bacterial membranes.

35. LeCun Y, Bengio Y, Hinton G: **Deep learning**. *Nature* 2015, **521**:436.

36. • Veltri D, Kamath U, Shehu A: **Deep learning improves antimicrobial peptide recognition**. *Bioinformatics* 2018, **34**:2740-2747.
The authors reported the use of deep learning to optimize recognition of antimicrobial peptides.

37. Lata S, Mishra NK, Raghava GPS: **AntiBP2: improved version of antibacterial peptide prediction**. *BMC Bioinform* 2010, **11**:S19.

38. Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S: **CAMP: a useful resource for research on antimicrobial peptides**. *Nucleic Acids Res* 2010, **38**:D774-D780.

39. Meher PK, Sahu TK, Saini V, Rao AR: **Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC**. *Sci Rep* 2017, **7**:42362.

40. •• Müller AT, Hiss JA, Schneider G: **Recurrent neural network model for constructive peptide design**. *J Chem Inf Model* 2018, **58**:472-479.
The authors reported computationally guided peptide design by neural networks using structural patterns for optimized design of active peptides.

41. Schneider P, Müller AT, Gabernet G, Button AL, Posselt G, Wessler S, Hiss JA, Schneider G: **Hybrid network model for "Deep Learning" of chemical data: application to antimicrobial peptides**. *Mol Informatics* 2016, **36**:1600011.

42. Yan Y, Wen Z, Wang X, Huang S-Y: **Addressing recent docking challenges: a hybrid strategy to integrate template-based and free protein-protein docking**. *Proteins Struct Funct Bioinform* 2016, **85**:497-512.

43. Gangopadhyay A, Datta A: **Identification of inhibitors against the potential ligandable sites in the active cholera toxin**. *Comput Biol Chem* 2015, **55**:37-48.

44. Melo MCR, Bernardi RC, Rudack T, Scheurer M, Riplinger C, Phillips JC, Maia JDC, Rocha GB, Ribeiro JV, Stone JE *et al.*: **NAMD goes quantum: an integrative suite for hybrid simulations**. *Nat Methods* 2018, **15**:351.

45. • Vishnepolsky B, Gabrielian A, Rosenthal A, Hurt DE, Tartakovsky M, Managadze G, Grigolava M, Makhatadze GI, Pirtskhalava M: **Predictive model of linear antimicrobial peptides active against gram-negative bacteria**. *J Chem Inf Model* 2018, **58**:1141-1151.
Reported a predictive model for the generation of peptides with selective activity against Gram-negative bacteria.

46. Veltri D, Kamath U, Shehu A: **Improving recognition of antimicrobial peptides and target selectivity through machine learning and genetic programming**. *IEEE/ACM Trans Comput Biol Bioinform* 2017, **14**:300-313.