



Published in final edited form as:

Chem Biol Drug Des. 2015 January ; 85(1): 14–21. doi:10.1111/cbdd.12423.

Machine-Learning Techniques Applied to Antibacterial Drug Discovery

Jacob D. Durrant^{1,§} and Rommie E. Amaro¹

¹Department of Chemistry & Biochemistry and the National Biomedical Computation Resource, University of California, San Diego, La Jolla, CA, 92093

Abstract

The emergence of drug-resistant bacteria threatens to catapult humanity back to the pre-antibiotic era. Even now, multi-drug-resistant bacterial infections annually result in millions of hospital days, billions in healthcare costs, and, most importantly, tens of thousands of lives lost. As many pharmaceutical companies have abandoned antibiotic development in search of more lucrative therapeutics, academic researchers are uniquely positioned to fill the resulting vacuum.

Traditional high-throughput screens and lead-optimization efforts are expensive and labor intensive. Computer-aided drug discovery techniques, which are cheaper and faster, can accelerate the identification of novel antibiotics in an academic setting, leading to improved hit rates and faster transitions to pre-clinical and clinical testing. The current review describes two machine-learning techniques, neural networks and decision trees, that have been used to identify experimentally validated antibiotics. We conclude by describing the future directions of this exciting field.

Introduction

Addressing the threat of drug-resistant bacteria is one of modern medicine's greatest challenges. The excitement surrounding Alexander Fleming's discovery of penicillin in 1928, which has rightfully been described as a "turning point in history" (1), was quickly followed by the disheartening realization that bacteria can mount a counterassault. Penicillinase, a β -lactamase capable of degrading penicillin, was identified even before penicillin had been applied clinically (2). Following widespread use in hospitals, sulfonamide-resistant *S. pyogenes* and penicillin-resistant *S. aureus* emerged in the 1930s (3) and 1940s (4), respectively. Many other bacterial strains have subsequently developed resistance, including some that are impervious to multiple antibiotics (1, 5).

In retrospect, this development is hardly surprising. Humans use hundreds of thousands of tons of antibiotics per year (6) for medical, veterinary, and agricultural purposes (5), thereby applying tremendous anthropogenic evolutionary pressure that favors resistance. Many resistance-conferring bacterial proteins existed even before the medical use of antibiotics

[§]Corresponding author: JDD: jdurrant@ucsd.edu. REA: ramaro@ucsd.edu.

Conflict of Interest

The authors have no potential conflicts of interest.

(7), and novel mutations in modern times have produced additional resistance genes. To complicate matters further, gene exchange, often plasmid mediated (8), is a “universal property of bacteria” (1) that does not respect even taxonomic and ecological boundaries (5), allowing resistance to spread quickly. As a single example of this phenomenon, consider the fact that 40–60% of nosocomial *S. aureus* in the U.S. and U.K. is now methicillin-resistant (MRSA), and many strains are multi-drug resistant (MDR) (5).

The economic and social burdens associated with treating resistant bacterial infections are substantial. Each year in Europe and the United States alone, these contagions result in ~11 million additional hospital days and over \$20 billion in additional health care costs (9, 10). Europe reports ~400,000 annual MDR infections that result in 25,000 deaths (9). While the development of novel therapeutics might initially appear to be profitable given the magnitude of the threat, in fact pharmaceutical companies have shied away from antibiotic development in recent years. New antibiotics are typically only used after more traditional medicines have failed. Rather than developing “drugs of last resort” with short-term utility, industry has shifted its focus to more lucrative long-term treatments to manage chronic conditions (10, 11).

A Unique Opportunity for Academia and Computer-Aided Drug Design

Given industry’s reluctance to develop novel antibiotics, academia is uniquely positioned to play a leading role in the earliest stages of lead identification and optimization (1). In response to this and other opportunities, academic drug-discovery centers have already been established at universities in Belgium, Sweden, the United Kingdom, and the United States (12). Success in these new settings depends on adapting industry approaches to the constraints of university research. For example, in industry high-throughput screens (HTS) are used to identify pharmacologically active lead antibiotic compounds by testing hundreds of thousands of compounds in highly automated assays (13, 14). Unfortunately, although robotics and miniaturization have led to increased efficiency, traditional HTS is beyond the reach of most academic researchers due to its high costs and labor requirements.

To make high-throughput testing more tractable, many have sought to complement large-scale experimental testing with software that predicts molecular recognition (i.e., ligand binding). Computer-aided drug design (CADD) techniques, though still in their infancy, have already contributed to the discovery and development of a number of drugs, including captopril, dorzolamide, boceprevir, aliskiren, nelfinavir, saquinavir, zanamivir, oseltamivir, and raltegravir, among others (15). By applying predictive CADD techniques to entire compound databases, computational biologists can recommend sets of compounds that are typically far more likely to bind to a given antibiotic drug target than compounds selected at random, thus requiring fewer subsequent *in vitro* and *in vivo* experiments. Further compound optimization can then be performed to ensure that any identified hits are capable of traversing the bacterial cell wall to reach those targets, if necessary.

Both ligand- and receptor-based methods for predicting molecular recognition have been developed. Ligand-based methods (e.g., Quantitative Structure-Activity Relationships, or QSAR) seek to identify molecules that are similar to known binders by algorithmically

mapping small-molecule descriptors to biological activity, independent of the receptor (16). QSAR can be as simple as determining whether or not a candidate ligand is structurally similar to known binders (e.g., through substructure or Tanimoto similarity searching). In more complex implementations, mathematical/statistical analyses of the molecular properties of known binders (e.g., LogP, molecular weight, polarizability, etc.) are used to build predictive models that can then be applied to new compounds, including models that simultaneously predict ADMET/Tox properties (17–20) and activity against multiple bacterial strains (21, 22). Some QSAR approaches even align 3D models of known small-molecule binders in order to find consistent patterns in the locations of key interacting groups (e.g., hydrogen-bond donors or acceptors, aromatic moieties, etc.). Novel compounds are then evaluated based on how well their 3D chemical configurations match these patterns, or pharmacophores (23, 24).

Receptor-based techniques, which require crystallographic, NMR, or homology models of the drug target (receptor), similarly seek to formulaically predict binding affinity. Rather than considering molecular descriptors, receptor-based approaches evaluate ligand-receptor interactions using “scoring functions.” If required, “docking programs” can be used to first predict the binding pose of a potential ligand within the drug-target binding pocket.

Notably, some receptor-centric scoring functions are generalized, meaning they can be applied to a novel drug target without any prior knowledge of specific ligands, though positive controls are certainly useful for pre-experimental validation. Furthermore, generalized docking/scoring protocols are arguably more likely to identify ligands that are structurally distinct from known binders, as they are not trained on a limited set of experimentally validated, target-specific modulators.

Machine Learning Applied to Computer-Aided Drug Design

Both ligand- and receptor-based techniques rely on intuitive, mathematical, or statistical “mappings” between 1) structural, molecular, or pharmacophoric data that can be evaluated *in silico*, and 2) experimentally verified enzymatic or biological activity. Traditional CADD techniques generally attempt to reduce binding to a single formulaic or statistical form, neglecting the subtle and perhaps nonlinear synergistic interplay between the many molecular factors that govern binding. In contrast, newer nonparametric machine-learning techniques learn directly from crystallographic and assay data without requiring explicit, programmatic instruction. They find patterns in observations of nature herself without the constraints of formulas or even human theories. By learning directly from the natural world, machine-learning techniques can often achieve accuracies not possible with more conventional approaches.

The remainder of this review describes specific antibiotic drug-discovery projects that have used advanced nonparametric machine-learning techniques to map chemical/binding properties (i.e., “descriptors”) to metrics of biological activity. We will focus on recent studies that have used neural networks or decision trees to identify potential antibiotics that were subsequently experimentally validated.

Neural Networks Applied to Antibiotic Drug Discovery: Ligand-Based Approaches

Artificial neural networks (ANNs) attempt to mimic the cellular architecture of the brain. Neurons and synapses are represented by virtual “neurodes” and “connections,” respectively. Network behavior is governed by both neurode organization and connection strength. An initial training phase involves systematically modifying the strengths of the connections in order to optimize the network’s ability to accurately predict experimentally measured activities when given corresponding vectors of molecular, structural, or pharmacophoric descriptors. The trained networks can subsequently be used to predict the activity of other potential ligands not included in the training set (25–27).

In recent years, ANNs have been used in conjunction with ligand-based QSAR to identify novel antibiotics. A 2004 study by Murcia-Soler et al. (28) considered a diverse set of 217 antibiotics spanning multiple classes and targets, as well as 216 decoys. This library of small molecules was divided into a training set (70%) and a testing set (30%). An ANN that mapped 62 structure-based molecular descriptors to the biological activity of the training-set compounds had an accuracy of 91.4% on the testing set, validating the network’s utility. The same descriptors were then generated for all the compounds of the Available Chemicals Directory and similarly fed into the trained ANN. Ten of the top predicted antibiotics were experimentally tested against both gram positive and gram negative bacteria (*E. faecalis*, *S. aureus*, *E. coli*, and *P. aeruginosa*). Of these, four had low micromolar potency against one or more strains, including one that was more effective against *S. aureus* and *E. faecalis* than the two known inhibitors included as positive controls (cephalosporin C and nalidixic acid).

In 2012, Sabet et al. (29) studied thirty-one 3-hydroxypyridine-4-one and 3-hydroxypyran-4-one antibiotics, known chelators that act by depriving bacteria of iron cations. An ANN was trained to use bond-based physiochemical (MOLMAP) molecular descriptors to classify these 31 compounds by their experimentally determined MIC values. In conjunction with other techniques, this trained network was then used to prioritize 302 novel 3-hydroxypyridine-4-one compounds. The ANN successfully predicted the activity of 84% of the 19 compounds ultimately tested; ten of these were able to kill *S. aureus*.

ANNs have also been used to identify novel bacteriocidal cationic peptides. In 2009, two similar papers were published describing the application of neural networks to peptide design (30, 31). Forty-four “atomic-resolution” descriptors were generated for ~1,400 peptides, both active and inactive against *P. aeruginosa*. This data set was used to train a predictive ANN, which then prioritized 100,000 novel “virtual peptides.” Rather than selecting only the best of these 100,000 peptides, the researchers synthesized and tested compounds with a wide range of predicted potencies, allowing them to confirm that the QSAR scores and experimentally measured IC₅₀ values were highly correlated. Furthermore, a number of the best predicted peptides were in fact effective against *P. aeruginosa* in the low micromolar range, more potent, even, than the best peptide from the training set. Some peptides were effective against several multi-drug resistant strains, and two showed *in vivo* efficacy in a mouse model of invasive Staphylococcal infection.

Neural Networks Applied to Antibiotic Drug Discovery: Receptor-Based Approaches

Aside from using neural networks in ligand-based QSAR, ANNs have also been used to create generalized receptor-based scoring functions. The distinction between ligand- and receptor-based techniques is important. Unlike ligand-based approaches, which consider small-molecule physical, chemical, and structural properties, receptor-based scoring functions consider ligand-receptor interactions. As they are interaction centric rather than ligand centric, in theory these scoring functions 1) can be applied to a novel drug target without the need for prior validated small-molecule binders, and 2) are more likely to identify inhibitors with novel scaffolds.

In 2010 and 2011, Durrant et al. published two papers describing a novel class of scoring functions based on ANNs: NNScore 1.0 (26) and NNScore 2.0 (27) (hereafter called NN1 and NN2, respectively). Briefly, a database containing thousands of models of diverse ligand-receptor complexes with associated experimentally determined binding affinities was derived from the MOAD (32) and PDBbind-CN databases (33, 34). Descriptors of each of these complexes were calculated by considering the structural features of the models themselves. In NN1, these descriptors included tallies of 1) ligand atoms, 2) juxtaposed ligand/receptor atoms, 3) and pairwise electrostatic energies, categorized by AutoDock atom type. Additionally, the number of ligand rotatable bonds was also considered. In NN2, the terms of the AutoDock Vina scoring function (35), as well as the more complex ligand-receptor interactions identified using the BINANA algorithm (36), were additionally considered. Neural networks were trained so that they could accurately predict experimentally measured binding affinities from a vector containing the appropriate descriptors. Though only recently developed, these scoring functions have already been used in a number of studies, in both industry and academia (personal communications, as well as refs. (37–42), for example).

In 2013, Lindert et al. used NNScore to identify inhibitors of farnesyl diphosphate synthase (FPPS), a critical enzyme in the mevalonate isoprenoid biosynthetic pathway (43). FPPS inhibitors have potential application as antiparasitic, antitumor, and antibiotic therapeutics. In bacteria, FPPS is essential for cell-wall biosynthesis. Lindert et al. docked a small-molecule library of 1,008 compounds into an allosteric FPPS pocket using AutoDock Vina (35) and Schrödinger's Glide. When the Vina-docked poses were rescored with NN1, the rankings of the known inhibitors improved substantially. By experimentally testing compounds that ranked well in both the NN1 and Glide screens, a single low micromolar hit was ultimately identified. Chemically similar compounds were then subjected to the same computational protocol in a secondary screen that identified ten additional low micromolar inhibitors, including one with an IC_{50} value of 1.8 μ M. Although Lindert et al.'s screen targeted human FPPS, subsequent studies demonstrated that the lead compound was effective against both *E. coli* and *S. aureus*, likely due to dual inhibition of bacterial FPPS and undecaprenyl pyrophosphate synthase, a second enzyme in the same pathway (44, 45).

In a separate 2013 study, Durrant et al. demonstrated that NNScore is well suited to several other bacterial drug targets (25). To compare NNScore performance to that of AutoDock

Vina and Schrödinger's Glide, Durrant et al. turned to the Directory of Useful Decoys (46), a benchmark set of forty diverse receptors with associated validated ligands. These known ligands were docked into their appropriate targets with Vina and Glide and optionally rescored with NN1 and NN2. A set of 1,560 structurally diverse presumed non-binders (the NCI Diversity Set III) was also included in the screen, and screen performance was evaluated by comparing the rankings of the known inhibitors to those of the decoys. In this benchmark study, NNScore performed particularly well against three antibiotic targets: *M. tuberculosis* enoyl-acyl-carrier-protein reductase (ENR), *L. casei* dihydrofolate reductase (DHFR), and *E. coli* AmpC β -lactamase.

M. tuberculosis ENR is the drug target of isoniazid, a first-line medication in the treatment of tuberculosis (47). Durrant et al. found that rescoring Vina-docked poses with NN2 was particularly effective at identifying true ENR inhibitors, as judged by a ROC-curve metric of early performance. In brief, given that only the top-ranking compounds found in a virtual screen are typically submitted for experimental validation, the goal of any screen should be to minimize the number of recommended compounds that are not ligands (the false positive rate, FPR), while maximizing the number of recommended compounds that are ligands (the true positive rate, TPR). To judge virtual-screen performance, Durrant et al. found the number of top-ranking compounds from the Vina-NN2 screen that would have to be recommended in order to achieve a FPR of no more than 5%. Recommending this same number of top compounds would have yielded a TPR of 47%. In contrast, this same TPR would have been only 35%, 21%, and 26% had Vina, Glide HTVS, or a three-tiered Glide HTVS/SP/XP protocol been used.

An NNScore-based protocol for identifying *L. casei* DHFR inhibitors was also particularly predictive. DHFR is the drug target of trimethoprim and its derivatives. Durrant et al. found that when the poses generated by the Glide HTVS/SP/XP docking/scoring protocol were rescored with NN1 and NN2, the TPR was 88% when the FPR was fixed at 5%. In contrast, the TPR was 8% and 0% when Vina and Glide HTVS were used for docking/scoring, respectively. (For technical reasons, it was not possible to obtain the TPR at 5% FPR for the HTVS/SP/XP protocol in this case.)

This same protocol (HTVS/SP/XP/NNScore) was also effective against *E. coli* AmpC β -lactamase, a bacterial enzyme targeted by drugs like clavulanic acid, sulbactam, tazobactam, and avibactam in order to make otherwise resistant bacteria susceptible to β -lactam antibiotics. The TPR of the HTVS/SP/XP/NNScore protocol was 29% when the FPR was fixed at 5%. The same TPR was 24%, 24%, and 2% when HTVS/SP/XP without NNScore, Glide HTVS, and Vina were used, respectively.

It is important to note that the screens performed by Durrant et al. were retrospective, not prospective. Nevertheless, they suggest that these neural-network scoring functions have great potential against antibacterial and other targets, so much so that the authors are currently using NNScore to pursue novel DHFR-inhibiting antibiotics.

Decision Trees Applied to Antibiotic Drug Discovery: Ligand-Based Approaches

A decision tree is a hierarchical representation of rules that can map a vector of descriptors to a given classification (48). Modern techniques allow decision trees to be automatically generated from large training sets without requiring direct human input. Suppose, for example, that a training set contains both known ligands and decoys, and that a number of descriptors (e.g., lipophilicity, molecular weight, polarizability, etc.) have been calculated for each compound. During the training phase, a decision tree is constructed by first considering many different descriptor cutoffs in order to determine the single best rule for separating the compounds according to their activity.

For the purpose of illustration, suppose this single best rule is to partition the compounds into two groups, with molecular weights greater than 400 daltons (likely ligands) and less than 400 daltons (likely decoys). It is not that this rule is perfect; many true ligands may have molecular weights less than 400 daltons. It is only that this rule happens to be better than any other at separating ligands from decoys. The same procedure is then applied separately to the two partitions, generating an additional layer of “rules” for further dividing the data. Applying the same procedure recursively again and again eventually yields a hierarchical set of rules capable of accurately classifying compounds as binders or nonbinders. These same sets of rules can then be applied to novel compounds not included in the training set.

A random forest is comprised of an ensemble of decision trees (49). A given training set is randomly partitioned many different ways. Additionally, the descriptor space can also be randomly partitioned. Diverse compound and descriptor subsets are then used to train many different individual decision trees. A given compound is classified by considering the output of these many trees (e.g., by calculating the average output, the mode, etc., of the whole forest), rather than considering the output of a single tree. These ensemble-based techniques often lead to improved accuracy.

Like neural networks, decision trees have been used extensively in ligand-based QSAR methods. In 2013, Lira et al. used a decision tree to identify novel antimicrobial peptides (50). Sixty known peptides, each with 53 calculated physicochemical molecular descriptors, were used for training. The decision tree was ultimately able to classify the antibacterial activity of each peptide as either none, low, medium, and high. This same tree was then used to prioritize five peptides similar to colossomin, a known antibiotic. Two peptides, colossomin C and colossomin D, were ultimately tested against both Gram-positive and Gram-negative bacteria. The peptides were particularly effective against *S. aureus*.

In a second study from 2007, Debeljak et al. sought to identify novel antibiotic coumarin derivatives (51). A random forest was generated using 64 quantum-mechanical and chemical descriptors calculated for 33 known coumarin-based antibiotics. Using a consensus score that combined the predictions of the forest and two other methods, Debeljak et al. ultimately identified two promising 4-morpholino coumarin derivatives that were effective against *S.*

aureus and *C. albicans*. Improvements in antifungal efficacy over the original compounds were particularly notable.

Decision Trees Applied to Antibiotic Drug Discovery: Receptor-Based Approaches

Receptor-based generalized rescoring functions have also benefited from the random-forest approach. In 2010, Ballester et al. published RF-Score, a random-forest scoring function analogous in many ways to NNScore (52). RF-Score was trained on over 1,000 ligand-receptor complexes with known binding affinities, taken from the PDBbind-CN database (33, 34). Like NNScore, RF-Score binding-event descriptors were determined based on the distances between juxtaposed receptor-ligand atom pairs. Ballester et al. found that the predictive accuracy of RF-Score was better than that of 16 well-established classical scoring functions that do not rely on machine learning.

In a separate 2012 study, Ballester et al. used RF-Score to identify novel inhibitors of *M. tuberculosis* and *S. coelicolor* type II dehydroquinase, a member of the shikimate biosynthetic pathway required for bacterial synthesis of aromatic amino acids, folic acid, ubiquinone, etc. (53). They initially identified ~4,000 compounds with shapes similar to those of known ligands from among the nine million compounds in the ZINC repository (54). These compounds were then 1) docked into several type II dehydroquinase structures using the GOLD docking program (55), 2) rescored using a consensus score comprised of three scoring functions (ChemScore, GoldScore, and ASP), and 3) rescored with RF-Score. The top predicted compounds were then tested experimentally to confirm enzymatic inhibition, ultimately identifying 100 new inhibitors with 50 new active molecular scaffolds. Importantly, RF-Score had a better hit rate than the consensus score, and RF-Score-identified compounds had better median potency.

Future Directions

Traditional drug-discovery paradigms have failed to keep up with the growing need for novel antibiotics. Many pharmaceutical companies have abandoned antibiotic research entirely in search of more lucrative markets. Even those companies that have sought novel therapeutics have faced great challenges. For example, between 1995 and 2001 GlaxoSmithKline performed a total of 70 high-throughput screens in search of novel antibiotics with Gram-positive or broad-spectrum activity, resulting in only five leads (a success rate four to five times lower than that of other targets at the time) (56). CADD may be able to address some of these challenges. Far more compounds can be tested *in silico* than *in vitro*, and at a much reduced cost. By prioritizing the compounds of a library using computational techniques, fewer compounds need be subsequently tested experimentally.

That having been said, CADD faces challenges of its own. Ligand- and receptor-based methods must find a balance between speed and accuracy (57). On one hand, rigorous physics-based computational techniques for binding-energy prediction, such as thermodynamic integration, single-step perturbation, and free energy perturbation (58), are too time and calculation intensive for use in large-scale virtual screens. In contrast, faster

techniques like QSAR and computer docking sometimes lack the accuracy required to truly enrich a set of compounds for candidate ligands. We have personally seen computational techniques yield fewer true ligands than would have been obtained by picking library compounds at random.

Machine-learning methods have the potential to further improve the accuracy of high-throughput ligand- and receptor-based screening without sacrificing speed. They permit more nuanced binding estimates by freeing affinity prediction from predetermined formulaic or statistical forms. Rather, these techniques find patterns in observations of nature herself, independent of formulas or human theories.

While we expect that machine-learning QSAR will continue to be widely applied to antibiotic research, more recent artificial-intelligence receptor-based methods, which are applicable to a wide range of targets, seem particularly promising. In principle, these general scoring functions don't require known ligands or complex receptor-specific training protocols, thus facilitating widespread adoption. In practice, most of the generalized receptor-based machine-learning scoring functions described in the literature have not been publically released, and so are mere proofs of concept. To our knowledge, only NNScore 1.0 (26), NNScore 2.0 (27), and RF-Score (52, 59) are publically available. Nevertheless, though still in its infancy, the application of machine learning to antibiotic research specifically, and drug design generally, clearly has a bright future.

Acknowledgments

This work was funded through a NIH Director's New Innovator Award (DP2-OD007237) and an NSF XSEDE Supercomputer resources grant (RAC CHE060073N) to R.E.A. Support from the National Biomedical Computation Resource (NBCR, P41 GM103426) is also gratefully acknowledged.

References

1. Davies J, Davies D. Origins and Evolution of Antibiotic Resistance. *Microbiol Mol Biol R*. 2010; 74:417.
2. Abraham EP, Chain E. An enzyme from bacteria able to destroy penicillin. *Nature*. 1940; 146:837.
3. Levy SB. Microbial Resistance to Antibiotics - an Evolving and Persistent Problem. *Lancet*. 1982; 2:83-8. [PubMed: 6123819]
4. Barber M, Rozwadowskadowzenko M. Infection by Penicillin-Resistant Staphylococci. *Lancet*. 1948; 255:641-4. [PubMed: 18890505]
5. Levy SB, Marshall B. Antibacterial resistance worldwide: causes, challenges and responses. *NatMed*. 2004; 10:S122-S9.
6. Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Microbiol*. 2010; 8:260-71. [PubMed: 20208551]
7. Datta N, Hughes VM. Plasmids of the Same Inc Groups in Enterobacteria before and after the Medical Use of Antibiotics. *Nature*. 1983; 306:616-7. [PubMed: 6316165]
8. Norman A, Hansen LH, Sorensen SJ. Conjugative plasmids: vessels of the communal gene pool. *Philos T R Soc B*. 2009; 364:2275-89.
9. Group EEJW. ECDC/EMA Joint Technical Report: The bacterial challenge: time to react. Stockholm: European Centre for Disease Prevention and Control; 2009. ECDC/EMA Joint Technical Report: The bacterial challenge: time to react.
10. Bush K, Courvalin P, Dantas G, Davies J, Eisenstein B, Huovinen P, et al. Tackling antibiotic resistance. *Nat Rev Microbiol*. 2011; 9:894-6. [PubMed: 22048738]

11. Projan SJ. Why is big Pharma getting out of antibacterial drug discovery? *Curr Opin Microbiol.* 2003; 6:427–30. [PubMed: 14572532]
12. Wyatt PG. The emerging academic drug-discovery sector. *Future Med Chem.* 2009; 1:1013–7. [PubMed: 21425989]
13. Mishra KP, Ganju L, Sairam M, Banerjee PK, Sawhney RC. A review of high throughput technology for the screening of natural products. *Biomed Pharmacother.* 2008; 62:94–8. [PubMed: 17692498]
14. Pereira DA, Williams JA. Origin and evolution of high throughput screening. *Br J Pharmacol.* 2007; 152:53–61. [PubMed: 17603542]
15. Talele TT, Khedkar SA, Rigby AC. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Curr Top Med Chem.* 2010; 10:127–41. [PubMed: 19929824]
16. Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol Inform.* 2010; 29:476–88.
17. Speck-Planche A, Kleandrova VV, Cordeiro MN. New insights toward the discovery of antibacterial agents: multi-tasking QSBER model for the simultaneous prediction of anti-tuberculosis activity and toxicological profiles of drugs. *Eur J Pharm Sci.* 2013; 48:812–8. [PubMed: 23376211]
18. Speck-Planche A, Cordeiro MN. Simultaneous virtual prediction of anti-*Escherichia coli* activities and ADMET profiles: A chemoinformatic complementary approach for high-throughput screening. *ACS combinatorial science.* 2014; 16:78–84. [PubMed: 24383958]
19. Speck-Planche A, Cordeiro MN. Simultaneous modeling of antimycobacterial activities and ADMET profiles: a chemoinformatic approach to medicinal chemistry. *Curr Top Med Chem.* 2013; 13:1656–65. [PubMed: 23889052]
20. Speck-Planche A, Kleandrova VV, Cordeiro MN. Chemoinformatics for rational discovery of safe antibacterial drugs: simultaneous predictions of biological activity against streptococci and toxicological profiles in laboratory animals. *Bioorg Med Chem.* 2013; 21:2727–32. [PubMed: 23582445]
21. Prado-Prado FJ, Gonzalez-Diaz H, Santana L, Uriarte E. Unified QSAR approach to antimicrobials. Part 2: predicting activity against more than 90 different species in order to halt antibacterial resistance. *Bioorg Med Chem.* 2007; 15:897–902. [PubMed: 17084086]
22. Prado-Prado FJ, Uriarte E, Borges F, Gonzalez-Diaz H. Multi-target spectral moments for QSAR and Complex Networks study of antibacterial drugs. *Eur J Med Chem.* 2009; 44:4516–21. [PubMed: 19631422]
23. Tosco P, Balle T. A 3D-QSAR-Driven Approach to Binding Mode and Affinity Prediction. *J Chem Inf Model.* 2012; 52:302–7. [PubMed: 22087561]
24. Clark RD, Norinder U. Two personal perspectives on a key issue in contemporary 3D QSAR. *Wires Comput Mol Sci.* 2012; 2:108–13.
25. Durrant JD, Friedman AJ, Rogers KE, McCammon JA. Comparing neural-network scoring functions and the state of the art: applications to common library screening. *J Chem Inf Model.* 2013; 53:1726–35. [PubMed: 23734946]
26. Durrant JD, McCammon JA. NNScore: A Neural-Network-Based Scoring Function for the Characterization of Protein-Ligand Complexes. *J Chem Inf Model.* 2010; 50:1865–71. [PubMed: 20845954]
27. Durrant JD, McCammon JA. NNScore 2.0: A Neural-Network Receptor-Ligand Scoring Function. *J Chem Inf Model.* 2011; 51:2897–903. [PubMed: 22017367]
28. Murcia-Soler M, Perez-Gimenez F, Garcia-March FJ, Salabert-Salvador T, Diaz-Villanueva W, Castro-Bleda MJ, et al. Artificial neural networks and linear discriminant analysis: A valuable combination in the selection of new antibacterial compounds. *J Chem Inf Comp Sci.* 2004; 44:1031–41.
29. Sabet R, Fassihi A, Hemmateenejad B, Saghaei L, Miri R, Gholami M. Computer-aided design of novel antibacterial 3-hydroxypyridine-4-ones: application of QSAR methods based on the MOLMAP approach. *J Comput-Aided Mol Des.* 2012; 26:349–61. [PubMed: 22456859]

30. Fjell CD, Jenssen H, Hilpert K, Cheung WA, Pante N, Hancock REW, et al. Identification of Novel Antibacterial Peptides by Chemoinformatics and Machine Learning. *J Med Chem.* 2009; 52:2006–15. [PubMed: 19296598]
31. Cherkasov A, Hilpert K, Jenssen H, Fjell CD, Waldbrook M, Mullaly SC, et al. Use of Artificial Intelligence in the Design of Small Peptide Antibiotics Effective against a Broad Spectrum of Highly Antibiotic-Resistant Superbugs. *ACS Chemical Biology.* 2009; 4:65–74. [PubMed: 19055425]
32. Hu L, Benson ML, Smith RD, Lerner MG, Carlson HA. Binding MOAD (Mother Of All Databases). *Proteins.* 2005; 60:333–40. [PubMed: 15971202]
33. Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J Med Chem.* 2004; 47:2977–80. [PubMed: 15163179]
34. Wang R, Fang X, Lu Y, Yang CY, Wang S. The PDBbind database: methodologies and updates. *J Med Chem.* 2005; 48:4111–9. [PubMed: 15943484]
35. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2009; 31:455–61. [PubMed: 19499576]
36. Durrant JD, McCammon JA. BINANA: A novel algorithm for ligand-binding characterization. *J Mol Graphics Modell.* 2011; 29:888–93.
37. Wang YT. Insights from modelling the 3D structure of the 2013 H7N9 influenza A virus neuraminidase and its binding interactions with drugs. *MedChemComm.* 2013; 4:1370–5.
38. Lu SJ, Chong FC. Combining Molecular Docking and Molecular Dynamics to Predict the Binding Modes of Flavonoid Derivatives with the Neuraminidase of the 2009 H1N1 Influenza A Virus. *International Journal of Molecular Sciences.* 2012; 13:4496–507. [PubMed: 22605992]
39. Waingeh VF, Groves AT, Eberle JA. Binding of Quinoline-Based Inhibitors to Plasmodium falciparum Lactate Dehydrogenase: A Molecular Docking Study. *Open Journal of Biophysics.* 2013; 3:285–90.
40. Gajo, GC. Seleção virtual de substâncias inibidoras da corismato mutase de *Meloidogyne incognita* e estudo fitoquímico e avaliação da toxicidade de *Croton floribundus* SPRENG para *Atta sexdens*. Lavras, Minas Gerais, Brazil: Universidade Federal de Lavras; 2013. Seleção virtual de substâncias inibidoras da corismato mutase de *Meloidogyne incognita* e estudo fitoquímico e avaliação da toxicidade de *Croton floribundus* SPRENG para *Atta sexdens*.
41. Nunes AS, Campos VP, Mascarello A, Stumpf TR, Chiaradia-Delatorre LD, Machado ART, et al. Activity of chalcones derived from 2,4,5-trimethoxybenzaldehyde against *Meloidogyne exigua* and in silico interaction of one chalcone with a putative caffeic acid 3-O-methyltransferase from *Meloidogyne incognita*. *Experimental Parasitology.* 2013; 135:661–8. [PubMed: 24140596]
42. Shi J, Chen J, Serradji N, Xu XM, Zhou H, Ma YX, et al. PMS1077 Sensitizes TNF-alpha Induced Apoptosis in Human Prostate Cancer Cells by Blocking NF-kappa B Signaling Pathway. *PLoS One.* 2013; 8
43. Lindert S, Zhu W, Liu YL, Pang R, Oldfield E, McCammon JA. Farnesyl Diphosphate Synthase Inhibitors from In Silico Screening. *Chem Biol Drug Des.* 2013; 81:742–8. [PubMed: 23421555]
44. Zhu W, Zhang Y, Sinko W, Hensler ME, Olson J, Molohon KJ, et al. Antibacterial drug leads targeting isoprenoid biosynthesis. *Proc Natl Acad Sci U S A.* 2013; 110:123–8. [PubMed: 23248302]
45. Liu YL, Lindert S, Zhu W, Wang K, McCammon JA, Oldfield E. Taxodione and arenarone inhibit farnesyl diphosphate synthase by binding to the isopentenyl diphosphate site. *Proc Natl Acad Sci U S A.* 2014; 111:E2530–9. [PubMed: 24927548]
46. Huang N, Shoichet BK, Irwin JJ. Benchmarking sets for molecular docking. *J Med Chem.* 2006; 49:6789–801. [PubMed: 17154509]
47. Rieder HL. Fourth-generation fluoroquinolones in tuberculosis. *Lancet.* 2009; 373:1148–9. [PubMed: 19345815]
48. Rokach, L. *Data Mining with Decision Trees: Theory and Applications.* Singapore: World Scientific; 2008.

49. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comp Sci.* 2003; 43:1947–58.
50. Lira F, Perez PS, Baranauskas JA, Nozawa SR. Prediction of Antimicrobial Activity of Synthetic Peptides by a Decision Tree Model. *Appl Environ Microb.* 2013; 79:3156–9.
51. Debeljak Z, Skrbo A, Jasprica I, Mornar A, Plecko V, Banjanac M, et al. QSAR study of antimicrobial activity of some 3-nitrocoumarins and related compounds. *J Chem Inf Model.* 2007; 47:918–26. [PubMed: 17489552]
52. Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. *Bioinformatics.* 2010; 26:1169–75. [PubMed: 20236947]
53. Ballester PJ, Mangold M, Howard NI, Robinson RLM, Abell C, Blumberger J, et al. Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification. *J R Soc Interface.* 2012; 9:3196–207. [PubMed: 22933186]
54. Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model.* 2005; 45:177–82. [PubMed: 15667143]
55. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. *Proteins-Structure Function and Genetics.* 2003; 52:609–23.
56. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature Reviews Drug Discovery.* 2007; 6:29–40.
57. Huang SY, Grinter SZ, Zou XQ. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys Chem Chem Phys.* 2010; 12:12899–908. [PubMed: 20730182]
58. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol.* 2011; 9:71. [PubMed: 22035460]
59. Li HJ, Leung KS, Ballester PJ, Wong MH. istar: A Web Platform for Large-Scale Protein-Ligand Docking. *Plos One.* 2014; 9