Bioorganic & Medicinal Chemistry Letters 28 (2018) 2807–2815

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Digest

Recent applications of machine learning in medicinal chemistry

Jane Panteleev[a,*,1], Hua Gao[a,*,1], Lei Jia[b,*,1]

[a] Amgen Discovery Research, 360 Binney St., Cambridge, MA 02141, USA
[b] Amgen Discovery Research, One Amgen Center Dr., Thousand Oaks, CA 91320, USA

A B S T R A C T

In recent decades, artificial intelligence and machine learning have played a significant role in increasing the efficiency of processes across a wide spectrum of industries. When it comes to the pharmaceutical and biotechnology sectors, numerous tools enabled by advancement of computer science have been developed and are now routinely utilized. However, there are many aspects of the drug discovery process, which can further benefit from refinement of computational methods and tools, as well as improvement of accessibility of these new technologies. In this review, examples of recent developments in machine learning application are described, which have the potential to impact different parts of the drug discovery and development flow scheme. Notably, new deep learning-based approaches across compound design and synthesis, prediction of binding, activity and ADMET properties, as well as applications of genetic algorithms are highlighted.

General concepts in the field of artificial intelligence (AI) increasingly feature in discussions of predictive modeling and optimization of the medicinal chemistry processes in drug discovery. One of the goals of AI is the establishment of machine learning (ML) platforms that allow progressive improvement of model performance. For instance, a recent class of ML algorithms based on complex architectures of neural networks, deep learning (DL), is prominent in the current literature as a possible tool for improving existing models or developing novel computational platforms (Fig. 1). This perspective aims to give an overview of examples showcasing how new ML methods, including DL, are applied in different areas of the drug discovery process.

Recent advances in ML have led to significant expansion and influence of technologies utilizing AI on our everyday lives. Through countless examples of applications empowered by rapid growth of data (big data), ranging from Apple's SIRI and Tesla's self-piloting Model S vehicle to the Google Photo auto classification feature and advanced risk assessment by the credit card issuers, AI has led to significant changes across a broad spectrum of industries. Although application of many of these methods to research and development in pharmaceutical and biotechnology industries is still in the early phase, the potential of AI to revolutionize the drug discovery process has already been demonstrated by the uptake and prevalent use of absorption, distribution, metabolism, excretion, and toxicity (ADMET) predictive tools, virtual screening, and quantitative structure activity relationship (QSAR) modeling.[1] Development and implementation of AI methods could considerably benefit more aspects of early drug discovery. This could include more accurately identifying hits from high throughput screening (HTS) and hit expansion, effectively designing novel drug like compounds and synthetic routes toward them, improving efficiency of lead optimization through design of more accurate ADMET and QSAR models, and lowering the overall cost and time of the drug discovery process (Fig. 2).

Development and application of *in silico* QSAR models to predict drug activity, has become increasingly utilized in drug development and discovery over the past few decades.[2] Simple machine learning algorithms such as multiple linear regression (MLR) and partial least squares (PLS) have been used for model development for relatively small data sets. With the application of HTS assays in drug discovery, large amounts of data on activity of diverse chemical matter have been generated,[3] and more sophisticated machine learning algorithms[i] such as random forests (RF),[4] support vector machines (SVM),[5] artificial neural network (ANN),[6] and Cubist[7] have been developed and used successfully to develop various types of *in silico* ADMET and QSAR predictive models (Fig. 3).

The current era of AI is led by advancement of deep learning (DL) technology and the methods around its implementation, which have demonstrated advantages in many areas including drug discovery.[8–10] DL expands on the application of artificial neural networks in ML,

* Corresponding authors.
E-mail address: ljia@amgen.com (L. Jia).
[1] All three authors contributed equally to this article.
[i] While detailed description of existing ML algorithms and statistical learning models falls outside of the scope of this review, numerous publications offer a good introduction to statistical learning principles. For an example see: James G., Witten D., Hastie T., Tibshirani R. An Introduction to Statistical Learning. New York, NY: Springer; 2013.

https://doi.org/10.1016/j.bmcl.2018.06.046
Received 10 April 2018; Received in revised form 24 June 2018; Accepted 26 June 2018
Available online 28 June 2018
0960-894X/

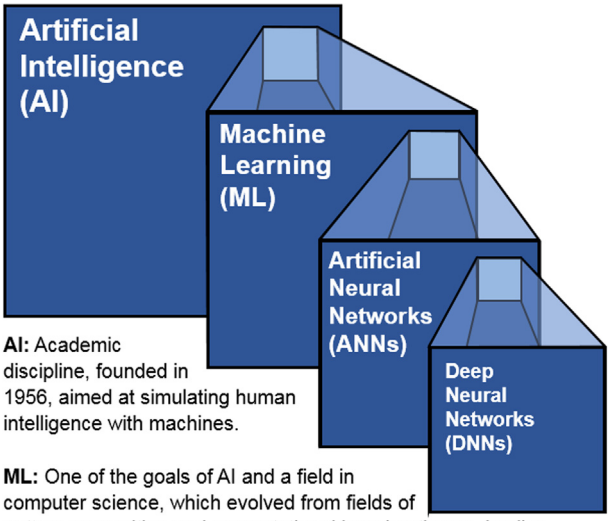

Fig. 1. Key definitions and relationships between artificial intelligence and deep learning.

where more complex network architectures are used (Fig. 4). Typically, multiple fully connected or convolutional hidden layers are utilized and interact in different ways to define platforms such as deep neural networks (DNNs), recurrent neural networks (RNNs), and convolutional deep neural networks (CNNs). DNNs are artificial neural networks with multiple hidden layers, typically with data flow in the forward direction, with capability to model complex non-linear relationships. RNNs can use internal memory to process arbitrary sequences of inputs, with data flow in any direction. RNNs are frequently used for text and speech recognition. The key feature of CNNs compared to DNNs is that CNNs can learn to recognize patterns across space and are effectively used in image and pattern recognition.

The use of DL methods has some advantages over traditional ML methods, namely in the capability of handling larger datasets, being able to utilize larger numbers of descriptors (or features), and obviating the need for costly feature engineering (Fig. 3). A predictive modeling and analytics contest (Kaggle competition) hosted by Merck, Inc. in 2012 demonstrated the high performance of deep neural nets, with methods utilizing DNNs winning the contest by predicting binding and ADMET properties of 15 data sets of small molecules.[11] DL-based methods have since been actively developed in the context of drug discovery,[10] especially when building more sophisticated and presumably more robust ADMET models.[12–18] In recent years, several start-up companies, including Accutar Biotechnology, Atomwise, Berg, Exscientia, Insilico, Insili.com, and Numerate, have been founded based on DL or non-DL based AI technology platforms, which are being actively explored in collaboration with pharmaceutical and biotechnology companies.[19]

Although a complete review of machine learning, and its applications in medicinal chemistry would require a more extensive series of publications, herein we focus on selected topics in early drug discovery which have been influenced by the most recent advancements in ML, including examples in structure generation and *de novo* design, synthetic assessment, ligand binding and activity prediction, ADMET predictions, and application of genetic algorithms in medicinal chemistry.

Recent applications of ML in structure generation

Although numerous computational approaches exist to generate and enrich virtual libraries for screening or hit expansion purposes, application of AI methods for the generation of novel chemical structures during lead optimization has been fairly confined to fragment growing or scaffold hopping strategies. Searching of extensive libraries of existing or virtual compounds by using pharmacophore models or virtual docking have been central to Ligand-Based Drug Design (LBDD) and

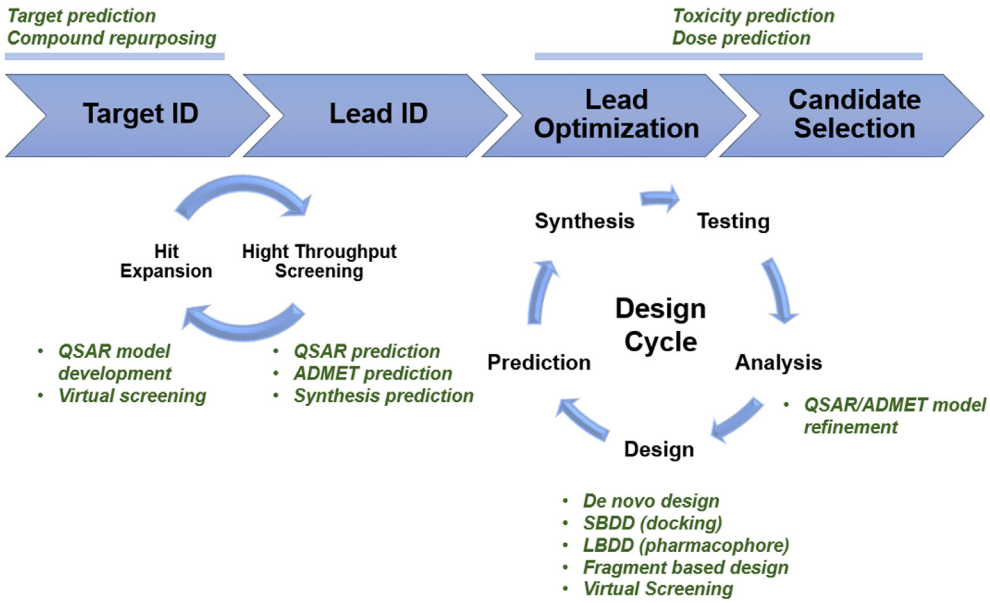

Fig. 2. Application of artificial intelligence and machine learning tools in the early stages of the drug discovery process. SBDD: structure-based drug design, LBDD: ligand-based drug design.

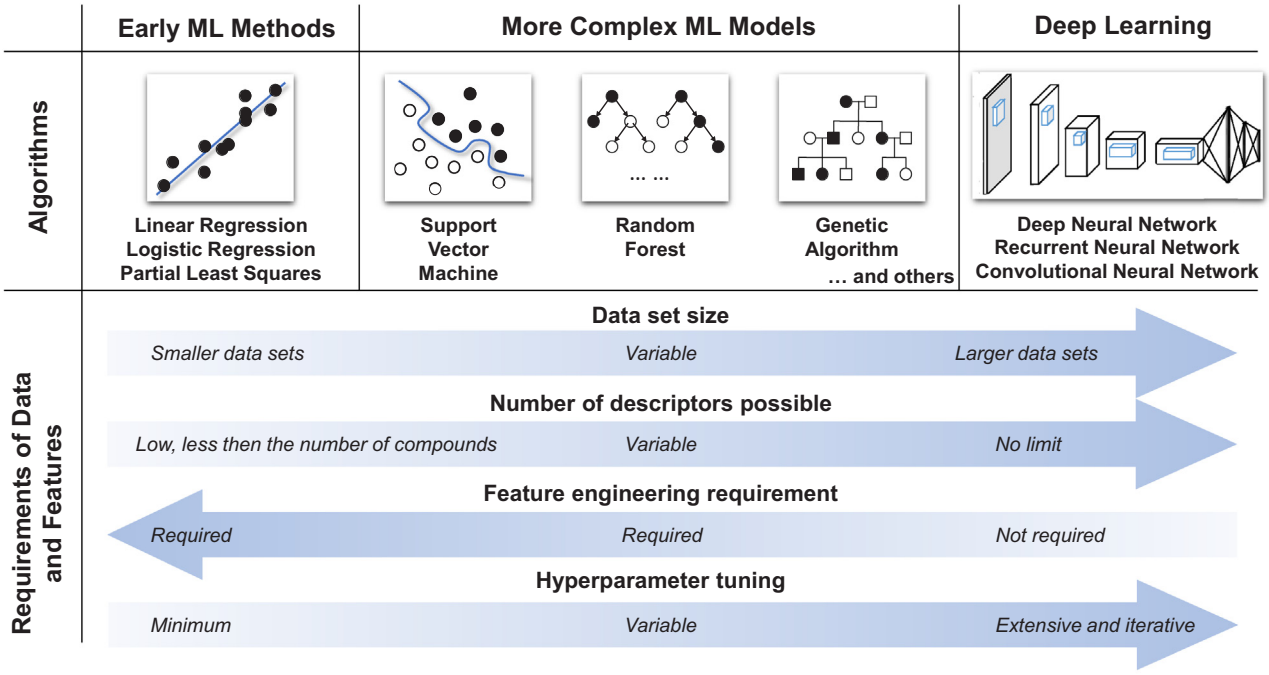

Fig. 3. Evolution of machine learning methods employed in predictive model development and associated data requirements[ii].

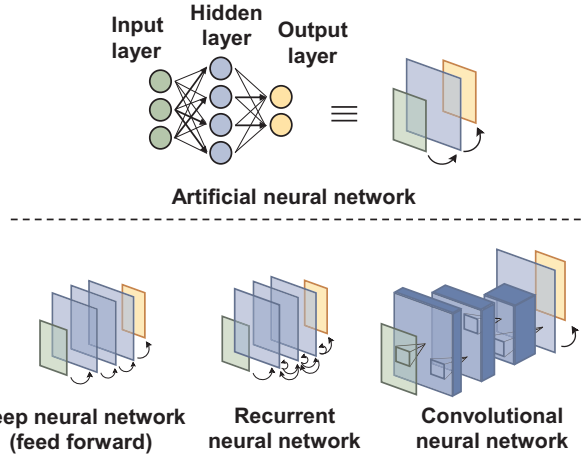

Fig. 4. Representation of selected neural network architectures. Different architectures can be used individually, or in combination[ii].

Structure-Based Drug Design (SBDD), respectively, with numerous reviews summarizing these well-established techniques and programs that apply them.[20,21] Most of the standard approaches to structure generation utilize manually defined reaction- or medicinal chemistry-based transformation templates on collections of monomers or a starting fragment hit (e.g. DrugGuru[22]). Although these methods can be quite powerful, as exemplified by impressively large virtual compound collections generated using enumerative approaches,[23–26] they can be somewhat rigid, requiring adherence to the initial rule set. In the recent literature, algorithms based on neural network frameworks and deep learning, aiming to generate new chemical structures without predefined reaction or transformation templates have been described. At this time, real world applications of these methods are limited, with the examples below lacking validation through synthesis and biochemical testing of the generated structures. However, the differentiation from more classic methods and the potential to offer an alternative method of structure generation warrants some discussion of these approaches.

In numerous recent reports of applications of DL to *de novo* design the text-based SMILES representation was shown to be effective in application of neural network-based learning frameworks (Fig. 5). Segler and Waller recently disclosed a model for design of molecules using recurrent neural networks (RNNs).[27] The featured RNN is designed to utilize computational functions analogous to those used in language-processing tasks, which are used to learn the SMILES string syntax and to assign probability to characters when building a new SMILES string, similar to language models which can assign probabilities to potential words in sentence prediction tasks. This specific model was trained on a set of 1.4 million molecules from the ChEMBL database in order to learn the grammar of SMILES, and the method was then used to generate novel structures. Impressively the SMILES output validity of 97.7% was obtained following parsing with a cheminformatics toolkit (chemistry development kit (CDK)).[28] The generated compounds covered a similar chemical property space to the training dataset. To assess whether the new molecules were valid from a perspective of hit finding, the compound set was subjected to an AstraZeneca filtering scheme intended for high-throughput screening collection enrichment, with 75% of the generated molecules passing this scheme. The model could then be used to generate structures similar to known biologically active compounds through a ML method called transfer learning, wherein the model is retrained or 'fine-tuned' on a smaller set of SMILES with known desired properties. Following fine-tuning the algorithm was able to generate a molecule set with > 50% of the compounds having predicted activity against 5-HT_{2a} receptor. In a similar fashion, the model was used to generate compounds with predicted activity against *Plasmodium falciparum* and *Staphylococcus aureus*. Schneider and coworkers described a similar application of RNNs for structure generation and subsequent use of transfer learning for fine-tuning.[29] In a proof of concept study, the authors fine-tuned their model on a set of molecules which were known inhibitors of peroxisome proliferator-activated receptor (PPAR) or trypsin, and then generated molecules with high similarity to the active

[ii] Graphics for algorithm and network architectures are intended for illustration purposes only. Technical details are not fully represented in these depictions.

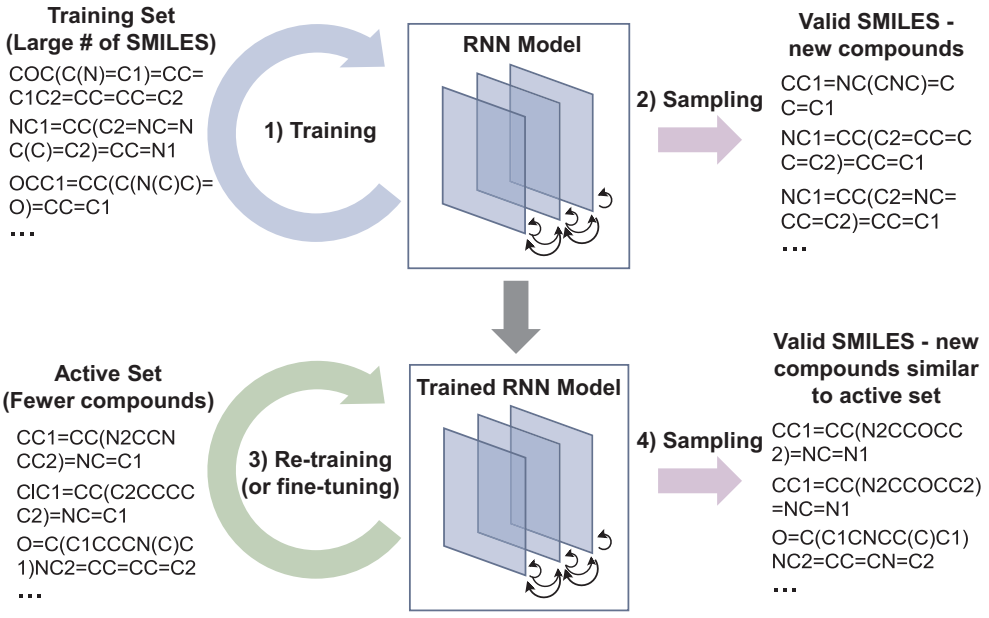

Fig. 5. General strategy for generation of new SMILES using RNN models.

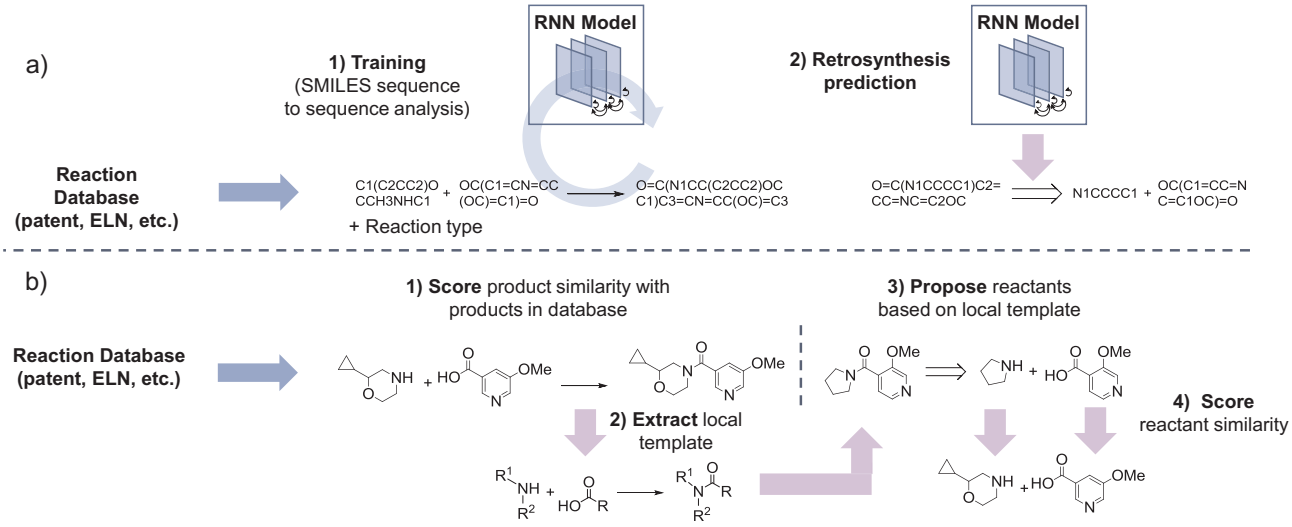

Fig. 6. Simplified schematic representation of retrosynthesis predictive algorithms not utilizing a template database. a) RNN-based sequence-to-sequence model applied to learn SMILES to SMILES substrate to product relationship; b) Structure similarity-based approach comparing product and substrate similarity with examples found in a database.

compounds.[30] This work also described a fragment growing approach with a user-defined starting fragment SMILES.

In another application of RNNs to a structure generation task, Olivercrona *et al.* described a similar framework using an alternative machine learning strategy for fine-tuning the model, called reinforcement learning; a method inspired by behaviorist psychology.[31,32] This model was trained to propose structures similar to the drug Celecoxib, and generate structures with predicted activity against dopamine receptor type 2 (DRD2).

In further applications of deep learning principles, examples featuring more complex architectures have been recently described. In these examples, neural network modules are trained on an input of SMILES to extract the most important features (encoder module), which can then be decoded back into SMILES (decoder module). These extracted features are encoded into a vector format which can be manipulated as a continuous representation. By further designing models which can differentiate vectors with high and low probability of having desired properties, generation of vectors, and sorting based on predicted property can allow new structure generation after decoding. This architecture was initially featured in a report from Aspuru-Guzik *et al.*,[33] and then adapted by Blaschke, Chen *et al.*[34] to generate novel compounds with structural similarity to a given structure, or predicted activity against dopamine receptor type 2 (DRD2) akin to the earlier work by the group.[31]

The recent advances in generative models based on neural networks have the potential to be complementary to the more established ligand- and structure-based computational design methods. While lacking constraints from template- or rule-based approaches, structures generated through these methods could have some liabilities, including limited structural diversity or poor synthetic accessibility, which would require further analysis of diversity or some forms of scoring and filtering based on synthetic feasibility and other drug-like properties. It is conceivable that the effectiveness of models used to generate structures with desired biological activity could be highly subject to the integrity and size of the dataset of active compounds, which is a common limitation of real-world early phase programs.

Recent advances in synthetic assessment of compounds

The necessity for evaluation of synthetic accessibility during *in silico* drug design stems from two common design problems. Novel chemical entities designed through non-reaction based means, for instance reverse QSAR methods[35] or structure generative methods (*vide supra*), can often be inaccessible from a synthetic standpoint. Alternatively, *de novo* design of new chemical entities is often accomplished through enumeration of virtual compounds by using curated reaction templates. Models capable of retrosynthesis or forward reaction prediction without the need for manually curated templates have been examined recently, and could further advance this field.

Pande and coworkers described an application of a model based on recurrent neural networks (RNNs) for retrosynthesis prediction (Fig. 6a).[36] Building on similar work reported by Nam and Kim who developed a forward predicting tool,[37] the model uses a training reaction set derived from a patent database utilizing reaction SMILES and associated reaction types. By using an encoder and a decoder architecture for feature extraction, the model learns to predict which types of reactants form which products through comparison of their extracted descriptors. The model performance was evaluated by comparison to a rule-based algorithm adopted from work reported by Green and Jensen.[38] The platform was found to perform significantly better on certain reaction types (such as heterocycle formations, protections, and deprotections) while the rule-based model was superior in other reaction classes (including heteroatom alkylations, arylations, and acylations). Some advantages over template-based approaches include incorporation of both chemical rule learning and candidate ranking into a single procedure, scalability, and capture of overall molecular structure, versus only atoms specific to the reaction sites. Waller *et al.* reported a NN-based model which can be used for both retrosynthesis and forward reaction predictions.[39] The model was trained on millions of known reactions and reaction rules, applying the extended connectivity fingerprint 4 (ECFP4) vector representation for the substrates and products. Initially 103 manually entered rules are applied in the exercise, where a neural network was used to learn which rules apply to the input molecule. Subsequently a probability of any given rule being valid for a disconnection was assigned. For retrosynthesis, accuracy of up to 98% was achieved within top 3 rule selections, or 78% for the highest ranked rule. The effort was then repeated using an algorithmically extracted set of 8720 rules, with the accuracy remaining high in the top 10 selections (95%), or the highest rated rule (64%). The main limitations of this system stem from the rule definitions, which can either be too general or too specific, as well as the loss of stereochemical information. As with most rule-based models, prediction of unusual reactions or disconnections was challenging. In a more recent report, the Waller group described another data-driven approach utilizing automatically extracted rule sets.[40] Their method featured a search algorithm often utilized in gaming (Monte Carlo tree search) in combination with three neural networks, which were trained on the Reaxis database of chemical reactions. In a double-blind evaluation of proposed and literature synthetic routes, participant organic chemists preferentially selected the machine-generated routes in most cases.

In a different approach, Green, Jensen, and coworkers report a method for proposing retrosynthesis based on similarity to know reactions (Fig. 6b).[41] A database of 40,000 reactions drawn from patent literature was used for reference. Initially a molecular similarity score was calculated between the desired product and the products in the reaction database, measuring the extent to which overlapping substructure elements exist between the reference and the template molecule. Similarity scores were calculated using Morgan circular fingerprints and Tanimoto similarity metrics. Secondly, transformations were extracted from the reference procedures containing the most similar products. By only including the immediate reaction atoms involved, possible reactant structures were generated using these templates. Thirdly, the template reactants were subjected to similarity scoring against the proposed reactants. The overall transformations were then scored as a product of substrate and product similarity scores and ranked. The method compared favorably to alternative retrosynthetic predictive algorithms. Because the approach is inherently data-driven, a possible limitation of the method is the disfavoring of the more unusual or "creative" disconnections in favor of operating within the scope of known chemistry.

More recent algorithms applying template-based approaches in forward reaction prediction have leveraged ML and NN-based models to train and extract template information from large databases of reaction records, such as patent reaction databases or electronic notebooks. In the work described by Green, Jensen, and coworkers a two-step process was developed, including enumeration of possible reactions from a given substrate using reaction templates and subsequent reaction scoring.[38] The reaction templates were extracted from a set of atom-mapped reaction SMILES (> 1 million) derived from published US patents between 1976 and 2013. Upon testing, the correct reaction product was predicted within the top 10 ranked reactions 91% of the time, and 72% of the time as the top ranked reaction. While template-free approaches in forward reaction prediction are rare, a few methods based on deep learning have been recently reported.[42,43]

Unlike predictive models for ligand binding and ADMET, retrosynthesis and reaction prediction models have not been widely adopted in medicinal chemistry. Over the last decade more sophisticated manually-encoded template-based applications are aiming to change this fact, with newer programs, such as Millipore Sigma's Chematica, showcasing impressive planning capabilities.[44,45] It is interesting to observe if the recent applications of NNs and DL will further improve the applicability and increase the uptake of synthesis planning software on a larger scale.

Recent advances of QSAR modeling for ligand binding activity driven by deep learning

The application of neural networks and DL to ligand binding prediction has become a rapidly expanding field in biological and pharmaceutical chemistry. The potential to improve the accuracy of QSAR models is particularly noteworthy.

Pande's group at Stanford University in collaboration with Pfizer has published a report which describes the use of different machine learning methods to predict the inhibitory activity against beta-Secretase 1 (BACE-1).[46] A relatively small training set was used containing only 205 compounds as well as a prospective test set containing 69 compounds to build the machine learning models. Among those, DNN-based models slightly outperformed other models when using a set of descriptors generated by a Schrödinger application, Canvas. The DNN models yielded ~70% accuracy for classification and a root mean squared error of about one log unit for regression when predicting a relatively large external validation set containing 1273 compounds. A crystal structure of hBACE-1/ligand complex (PDB ID: 1QP8) was used for guiding the QSAR model development, by using the bound ligand conformations as templates. This work demonstrates that DNNs can be effective in the absence of a very large training data set. In this case, the crystal structures of the receptor/ligand complex play important role in guiding the ligand binding pose to make the predictive modeling effective without a large amount of data points.

In a more recent publication from the Koes group, convolutional neural networks (CNN) were applied to score protein ligand binding.[47] CNNs are a type of neural networks used in deep learning methods, which show significant improvement in image recognition. A key feature of CNNs is the emphasis on the context information in the data set. For instance, in the image recognition case, the local pixel groups are detected for granular image features, and different CNN layers can be used to train models based on different levels of granularity. In this work, the authors aimed to detect the context of the molecular interactions by using a CNN-based model, with features of protein-ligand

interactions, which correlate with binding, being automatically captured by the scoring function. The authors demonstrated that this scoring function outperformed the more traditional scoring tools, such as AutoDock Vina, when ranking conformations for pose prediction and virtual screening. The authors were also able to extend CNNs from 2D image recognition, to capturing three-dimensional aspects of molecular structures. With independent test sets, the CNN-based method furnished improved classification versus Vina in a pose prediction experiment (0.792 AUC[iii] vs 0.682), and in virtual screening using ChEMBL data (0.779 AUC vs 0.665). However, when using a more limited dataset (maximum unbiased validation (MUV) dataset) none of the methods furnished highly predictive results. This work shows that the CNN method, when used with proper property descriptors, can lead to improvements over the traditional non-neural network methods. Although it remains to be determined whether CNN-based methods have reached a plateau in terms of the prediction accuracy.

Liu, Vedadi, Liu, Jin, and coworkers reported the use of a DL-assisted docking application to successfully design two micromolar inhibitors of Nicotinamide N-Methyltransferase (NNMT).[48] This work demonstrated an application of the more modern AI technology to a real-life drug discovery program. A co-crystal structure with one ligand showed that the binding pose was in general agreement with the docking outcome. A potential caveat of this work is a lack of direct comparison to traditional docking applications, making it challenging to assess whether this AI-based docking algorithm is superior.

In silico in vitro ADMET modeling for optimizing molecular properties

A plethora of *in silico* models have been developed using different machine learning algorithms.[7,49–52] *In silico* ADMET models have been widely used in HTS hits triage, prioritization of singleton and library design ideas, and lead optimization. State-of-the-art tools developed for ADMET modeling have been extensively reviewed.[53] There are several important factors which need to be considered when building *in silico* ADMET models. One of the first considerations is the understanding of the ADMET property to be analyzed and how the research team intends to use this property to make design decisions. Distinguishing between categorical use, such as permeable or non-permeable, or interpretation as a continuous spectrum, such as fraction unbound for calculating *in vivo* clearance or dose, is valuable for model construction. Next, the variability of the experimental data should be examined. Since *in silico* modeling is intended to simulate an experimental assay, the models are only as good as the quality of the data on which they are based. Following this, the machine learning method(s) to be used to analyze the structure-property/activity relationship should be examined in the context of the structural diversity, SAR linearity, and size of the data set to be analyzed. For small data sets, especially for a congeneric series of compounds, simple multi-linear regression analysis or partial least squares can be sufficient. For large and structurally diverse sets of data with non-linear SAR relationships, more sophisticated methods such RF, ANN, and Cubist can be more practical. As was previous published, Cubist is a good method for analyzing large datasets, especially those with highly non-linear SAR.[7] The next aspect to be considered is the available molecular descriptor set, as accuracy, interpretability, reproducibility, and speed need to be evaluated. In our experience, conformation-dependent three-dimensional molecular descriptors should be avoided when feasible. *In silico* models built based on interpretable molecular descriptors can be more useful for end users for understanding the underlying molecular properties when designing new molecules. Finally, the application domain or prediction confidence needs to be examined if the model is meant to be applied for prospective property predictions. In practice, fingerprint-based molecular similarity can serve as a reasonable and rapid surrogate metric of prediction confidence. Some recent examples of systematic utilization of *in silico* tools include the work of Doherty and coworkers, who developed a platform for combinatory library design by incorporating *in silico* ADME screening and modular library synthesis technologies.[54] Recently Gurjar *et al.* designed imidazole analogues as potential cholinesterase inhibitors as neuroprotective agents for Alzheimer's disease by incorporating calculated ADME properties including permeability and blood-brain penetration.[55]

Recently, deep learning has had some success in improving *in vitro* ADME prediction accuracy compared to traditional machine learning methods, such as random forest and Cubist.[8,11,13,56] The multiple task deep neural network[57] and graph convolutional[58,59] methods, which rely on molecular feature extraction directly from graph depictions, play important roles in the accuracy boost. In our own work, we revealed that in multiple cases, *in silico* ADME models developed using graph convolutional network were significantly better than models developed using Cubist as baseline method.[60] Performance of models developed by different algorithms is highly dependent on size of data sets, type of endpoints, type of models, and type of molecular descriptors used. When comparing different statistical algorithms, these factors should be considered. In the work of Ramsundar *et al.*,[61] multitask deep networks were surprisingly robust and offered strong improvement over random forests.

From *in vitro* ADME to *in vivo* PK modeling

In recent years, significant efforts have been devoted to modeling *in vivo* PK parameters, such as clearance, volume of distribution,[62] bioavailability, and biliary excretion.[63] There are two basic approaches to modeling *in vivo* PK properties: (1) constructing models from physicochemical properties of molecules; (2) constructing models by incorporating relevant experimental or calculated *in vitro* endpoints as well as related molecular properties. *In silico* models developed using the latter approach are typically more physiologically or pharmacologically relevant, are often considered to be mechanism-based, and are useful for identifying underlying ADME properties needing to be optimized. These types of *in silico* models are simpler to understand and interpret, and when possible, are preferable.

In an example of an *in vivo* model based on physicochemical properties, Gao and coworkers published a model for rat biliary excretion developed directly from several molecular properties.[63] In their modeling, Moriguchi descriptors,[64] Abraham solvation parameters,[65] clogD at pH 7.4, and descriptors representing the summation of Van der Waals surface area (VSA) with partial atomic charges in a range of −0.05 to 0.00 and a range of −0.20 to −0.15, respectively[66] were used. The derived model with an r^2 of 0.85 and a q^2 of 0.84, was successfully used to design cholecystokinin A receptor agonists.

In a recent example of a model based on *in vitro* as well as physicochemical properties, rat *in vivo* clearance was modeled for a set of inhibitors of the voltage-gated sodium channel, Nav1.7.[67] In this report, rat total clearance was calculated from a set of relevant *in vitro* endpoints either experimentally or obtained from *in silico* predictions. Experiments have indicated that a large component of the total clearance of this series of compounds was transporter mediated, especially through BCRP (Breast Cancer Resistance Protein) and MRP2 (Multidrug Resistance-Associated Protein 2). The *in vivo* clearance of the compounds in rat was also affected by passive permeability and plasma protein binding. Based on *in vitro* experimental data, predictive *in silico* models were built for cell permeability, BCRP and MRP2 transporter efflux, and rat plasma protein binding. Rat *in vivo* clearance was estimated by mathematically combining these predicted *in vitro* ADME endpoints. Utilizing this model, it was clear that compounds with higher efflux ratio, and higher free fraction would have higher total *in*

[iii] AUC (area under the curve) is a performance metric based on receiver operating characteristic curves (ROCCs). AUC = 1 represents a perfect classifier and AUC = 0.5 represents a random classifier.

vivo clearance. By measuring *in vitro* ADME endpoints and building robust *in silico* models, experimental or predicted *in vitro* ADME values could be used to estimate *in vivo* clearance. Since *in vivo* clearance can be an expensive and time-consuming endpoint to obtain experimentally, using models such as this can not only help in reducing animal usage and cost, but can also facilitate quicker design decisions, potentially decreasing design cycle times. The model was successfully used to prioritize synthetic targets and select compounds for *in vitro* and *in vivo* testing.

Another important *in vivo* PK parameter is volume of distribution. Lombardo and coworkers have reported on a number of *in silico* models of volume of distribution over the years. Recently, the group analyzed a set of 1096 compounds and, by carefully selecting a set of molecular descriptors and statistical methods, they were able to develop an *in silico* model utilizing a random forest-based algorithm. The results were comparable with those of *in vivo* scaling approaches, and the group has recommended random forest model as the optimal method for volume of distribution prediction.[68]

Oral bioavailability is defined as the fraction of a drug that reaches the systemic circulation. It can be one of the most important properties determining whether a compound can be progressed. Since many factors including gut metabolism, first pass metabolism, permeability, solubility, and formulation contribute to oral bioavailability, it can be a very difficult parameter to analyze and develop robust *in silico* predictive models for. Although different methods have been developed to date,[69–71] significant efforts are still needed to improve the quality and robustness of such *in silico* predictions, especially with different deep learning algorithms.

Applying *in silico* ADMET models in multi-parameter optimization of molecules

Over the course of the drug discovery process many factors, such as lack of efficacy, poor pharmacokinetic profile, and various toxicities may contribute to the attrition of molecules. In practice, many molecular attributes need to be optimized to improve the likelihood of success of a molecule progressing through the pipeline.[72] The drug discovery process, especially during the lead optimization (LO) stage, can be looked at as a multi-parameter optimization process,[73] where predicted human dose can be used as a key metric capturing many PKPD parameters important for compound progression.[74] Integration of properties such as potency, clearance, and brain exposure for CNS targets into human dose prediction as multi-parameter optimization (MPO) scoring function with rationally weighed metrics can make compound optimization and prioritization more objective and transparent. Furthermore, using human dose projection to triage compounds for *in vivo* efficacy studies and other resource demanding *in vivo* experiments can improve discovery efficiency by reducing research costs and shortening design cycle time. The dose projection method has been successfully used in developing opioid receptor antagonist as anti-obesity molecules.[2,75] In this example, the authors built several *in silico* ADME models including PgP efflux, brain impairment, *in vitro* liver microsomal stability, and liver microsomal binding. Then by combining experimental potency (or estimated potency for virtual compounds) and quantitative pharmacology, human dose was calculated and dose prediction and associated *in silico* models were fully integrated into the compound design and screening paradigm. As a result of the extensive use of these *in silico* predictions in combination with *in vitro* assay data, fewer compounds were submitted for *in vivo* efficacy and exposure studies. It was estimated that *in vivo* studies were decreased by 75%.

Recent advances of genetic algorithm application in medicinal chemistry

A genetic algorithm (GA) is a simple but elegant search technique which finds approximate solutions to optimization problems based on the principles of Darwinian evolution. It is a machine learning method which has been applied as an effective optimization method in cheminformatics since 1990s.[76] In nature, most organisms evolve by means of the two primary processes of natural selection and breeding. The selection process determines the elite population which survives to breeding. The breeding or reproduction process conducts crossover and random mutations of the gene to generate offspring. Both processes alternate iteratively to produce many generations.[77] GAs were developed by simulating this process of natural evolution. They can be applied in models where the best combinations of multiple parameters or features need to be determined in searching tasks, for instance when selecting the most relevant pharmacophore features in virtual screening or selecting the best set of variables for QSAR modeling. Recent advances of GA features coupling GAs with other machine learning applications (e.g. Principal Component Analysis (PCA); Support Vector Machine (SVM); Multiple Linear Regression (MLR) etc.) to take advantage of this optimization power.[78]

In recent work, Visco's group reported using a PCA-GA-SVM algorithm to identify novel factor XIIa inhibitors through virtual screening. A search against the PubChem 72 million compound library resulted 123 hits. Following experimental validation of 14 compounds among the 123 hits, 6 were found to be active.[79] Ahmadi and coworkers used GA-MLR to predict butyrylcholinesterase inhibitors, and they demonstrated that GA-MLR outperformed stepwise MLR with cross validation.[80] Cao *et al.* applied nondominated sorting genetic algorithm-II (NSGA-II) to select 30 molecular descriptors for Caco-2 cell permeability prediction. When coupled with a boosting method, the model reached 0.81 R^2 with the test set which was randomly selected (20% of the data set).[81] Moreover, Nekoei et al. compared GA-MLR and GA-SVM directly in a VEGFR-2 inhibitor QSAR study, finding that, GA-MLR had slightly better predictive power compared to GA-SVM, which was slightly more overfitted.[82]

Caflisch *et al.* applied GA for automatic fragment-based design of molecules within a protein binding site of known structure.[83] Their GA named GANDI (Genetic Algorithm-based de Novo Design of Inhibitors) was used to design 1809 CDK2 inhibitors based on a library of 14,000 fragments. Some of the designed compounds shared the substructure of known kinase inhibitors. Roth, Hopkins and coworkers developed an adaptive optimization algorithm, which is very similar to GA and successfully applied it to ligand design.[84] Overall accuracy of 75% was achieved over 800 newly designed GPCR ligands. Such an achievement can significantly shorten the drug discovery process turn-around time and save associated costs. Exscientia Ltd was founded based on several inventions including this one.[85]

Coupling GAs with deep neural networks[86,87] has been reported in other research fields including robotic system development,[88] video editing,[89] high-speed rail system development,[90] and medical science,[91,92] just to name a few. However, such applications are rarely seen in medicinal chemistry to date. Lusci, Baldi and coworkers reported using GA to select descriptor sets for solubility prediction with DNN.[18] We believe that further applications of GAs in combination with state-of-the-art machine learning algorithms could be a promising direction for advancing the AI performance in drug discovery.

Conclusions

Recent advancements in AI, including development of more sophisticated machine learning algorithms, have made a significant impact on the drug discovery process, including medicinal chemistry. The major advantage of applying various AI methods is their potential to reduce cost, cycle time, and labor demands during the early stages of drug discovery, by leveraging *in silico* methods for compound design, synthetic route assessment, and modeling ADMET and binding properties. The improvements in prediction accuracy of *in vivo* properties and toxicological events could lead to a lower attrition rate of drug molecules progressing through the pipeline. This review highlights

some recent advances utilizing newer machine learning methods in design of tools capable of structure generation and synthetic assessment, as well as ADMET and ligand binding models. While some of these tools and models are more prevalently utilized in the drug discovery process than others, further improvements in either accuracy or accessibility stand to further impact the efficiency of the drug discovery process.

Acknowledgements

The authors would like to acknowledge Angel Guzman-Perez, Margaret Chu-Moyer, David St. Jean, and Daniel Horne for their contribution to the preparation of this manuscript.

References

1. Leelananda SP, Lindert S. Computational methods in drug discovery. *Beilstein J Org Chem.* 2016;12:2694–2718. http://dx.doi.org/10.3762/bjoc.12.267.
2. Inge R, Leif J. Machine Learning: Advances in Research and Applications. https://www.novapublishers.com/catalog/product_info.php?products_id=63154. Accessed 16.01.18.
3. Cheng A, Diller DJ, Dixon SL, Egan WJ, Lauri G, Merz KM. Computation of the physio-chemical properties and data mining of large molecular collections. *J Comput Chem.* 2002;23:172–183. http://dx.doi.org/10.1002/jcc.1164.
4. Breiman L. Random forests. *Mach Learn.* 2001;45:5–32.
5. Cortes C, Vapnik V. Support-Vector Networks. *Mach Learn.* 1995;20:273–297. http://dx.doi.org/10.1023/A:1022627411411.
6. Hagan MT, Demuth HB, Beale MH. Neural Network Design. Distributed by Campus Pub. Service. University of Colorado Bookstore.
7. Gao H, Yao L, Mathieu HW, et al. In Silico Modeling of Nonspecific Binding to Human Liver Microsomes. *Drug Metab Dispos.* 2008;36:2130–2135. http://dx.doi.org/10.1124/dmd.107.020131.
8. Gawehn E, Hiss JA, Schneider G. Deep Learning in Drug Discovery. *Mol Inform.* 2016;35:3–14. http://dx.doi.org/10.1002/minf.201501008.
9. Ekins S. The next era: deep learning in pharmaceutical research. *Pharm Res.* 2016;33:2594–2603. http://dx.doi.org/10.1007/s11095-016-2029-7.
10. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. *Drug Discov Today.* 2018. http://dx.doi.org/10.1016/j.drudis.2018.01.039.
11. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep neural nets as a method for quantitative structure-activity relationships. *J Chem Inf Model.* 2015;55:263–274. http://dx.doi.org/10.1021/ci500747n.
12. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of Deep Learning in Biomedicine. *Mol Pharm.* 2016;13:1445–1454. http://dx.doi.org/10.1021/acs.molpharmaceut.5b00982.
13. Kearnes S, Goldman B, Pande V. Modeling Industrial ADMET Data with Multitask Networks. *arXiv*; June 2016. 1606.08793. https://arxiv.org/pdf/1606.08793.pdf.
14. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discovery Today.* 2017 https://ac.els-cdn.com/S1359644616304366/1-s2.0-S1359644616304366-main.pdf?_tid=a8a9afba-9fd8-11e7-b71e-00000aacb362&acdnat=1506114035_9278bac113b8363a2ec62d4b934fa012. Published 2017. Accessed September 22, 2017.
15. Goh GB, Hodas NO, Vishnu A. Deep learning for computational chemistry. *J Comput Chem.* 2017;38:1291–1307. http://dx.doi.org/10.1002/jcc.24764.
16. Aliper A, Plis S, Artemov A, Ulloa A, Mamoshina P, Zhavoronkov A. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol Pharm.* 2016;13:2524–2530. http://dx.doi.org/10.1021/acs.molpharmaceut.6b00248.
17. Ma J, Sheridan RP, Liaw A, Dahl GE, Svetnik V. Deep neural nets as a method for quantitative structure-activity relationships. *J Chem Inf Model.* 2015;55:263–274. http://dx.doi.org/10.1021/ci500747n.
18. Lusci A, Pollastri G, Baldi P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *J Chem Inf Model.* 2013;53:1563–1575. http://dx.doi.org/10.1021/ci400187y.
19. Smalley E. AI-powered drug discovery captures pharma interest. *Nat Biotechnol.* 2017;35:604–605.
20. Kutchukian PS, Shakhnovich EI. De novo design: balancing novelty and confined chemical space. *Expert Opin Drug Discovery.* 2010;5:789–812. http://dx.doi.org/10.1517/17460441.2010.497534.
21. Ramírez D. Computational methods applied to rational drug design. *Open Med Chem J.* 2016;10:7–20. http://dx.doi.org/10.2174/1874104501610010007.
22. Stewart KD, Shiroda M, James CA. Drug guru: a computer software program for drug design using medicinal chemistry rules. *Bioorganic Med Chem.* 2006;14:7011–7022. http://dx.doi.org/10.1016/j.bmc.2006.06.024.
23. Nicolaou CA, Watson IA, Hu H, Wang J. The proximal lilly collection: mapping, exploring and exploiting feasible chemical space. *J Chem Inf Model.* 2016;56:1253–1266. http://dx.doi.org/10.1021/acs.jcim.6b00173.
24. Hu Q, Peng Z, Sutton SC, et al. Pfizer global virtual library (PGVL): a chemistry design tool powered by experimentally validated parallel synthesis information. *ACS Comb Sci.* 2012;14:579–589. http://dx.doi.org/10.1021/co300096q.
25. Peng Z. Very large virtual compound spaces: construction, storage and utility in drug discovery. *Drug Discov Today Technol.* 2013;10:e387–e394. http://dx.doi.org/10.1016/j.ddtec.2013.01.004.
26. REAL Database – 16 Million Tangible Screening Compounds – Pharmaceutical Business Review. http://www.pharmaceutical-business-review.com/suppliers/enamine/products/real-database-16-million-tangible-screening-compounds. Accessed 5.03.18.
27. Segler MHS, Kogej T, Tyrchan C, Waller MP. Generating focussed molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci.* 2018;4:120–131. http://dx.doi.org/10.1021/acscentsci.7b00512.
28. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E. The chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics. *J Chem Inf Comput Sci.* 2003;43:493–500. http://dx.doi.org/10.1021/ci025584y.
29. Gupta A, Müller AT, Huisman BJH, Fuchs JA, Schneider P, Schneider G. Generative recurrent networks for de novo drug design. *Mol Inform.* 2017. ;36http://dx.doi.org/10.1002/minf.201700111 ASAP.
30. Merk D, Friedrich L, Grisoni F, Schneider G. De novo design of bioactive small molecules by artificial intelligence. *Mol Inform.* 2018. ;37http://dx.doi.org/10.1002/minf.201700153 ASAP.
31. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. *J Cheminform.* 2017;9:1–14. http://dx.doi.org/10.1186/s13321-017-0235-x.
32. Van Otterlo M, Wiering M. Reinforcement learning and markov decision processes. *Reinf Learn.* 2012:3–42. http://dx.doi.org/10.1007/978-3-642-27645-3_1.
33. Duvenaud D, Maclaurin D, Aguilera-Iparraguirre J, et al. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv*; 2018. 1509.09292v2. https://arxiv.org/pdf/1509.09292.pdf.
34. Blaschke T, Olivecrona M, Engkvist O, Bajorath J, Chen H. Application of generative autoencoder in de novo molecular design; 2017. 1700123. doi:10.1002/minf.201700123.
35. Miyao T, Kaneko H, Funatsu K. Inverse QSPR/QSAR analysis for chemical structure generation (from y to x). *J Chem Inf Model.* 2016;56:286–299. http://dx.doi.org/10.1021/acs.jcim.5b00628.
36. Liu B, Ramsundar B, Kawthekar P, et al. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Cent Sci.* 2017;3:1103–1113. http://dx.doi.org/10.1021/acscentsci.7b00303.
37. Nam J, Kim J. Linking the Neural Machine Translation and the Prediction of Organic Chemistry Reactions. arXiv; 2016. 1–19.
38. Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF. Prediction of organic reaction outcomes using machine learning. *ACS Cent Sci.* 2017. http://dx.doi.org/10.1021/acscentsci.7b00064.
39. Segler MHS, Waller MP. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chem A Eur J.* 2017;23:5966–5971. http://dx.doi.org/10.1002/chem.201605499.
40. Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature.* 2018;555:604–610. http://dx.doi.org/10.1038/nature25978.
41. Coley CW, Rogers L, Green WH, Jensen KF. Computer-assisted retrosynthesis based on molecular similarity. *ACS Cent Sci.* 2017;3:1237–1245. http://dx.doi.org/10.1021/acscentsci.7b00355.
42. Jin W, Coley CW, Barzilay R, Jaakkola T. Predicting Organic Reaction Outcomes with Weisfeiler-Lehman Network. *arXiv*; September 2017. 1709.04555. http://arxiv.org/abs/1709.04555.
43. Kayala MA, Azencott C-A, Chen JH, Baldi P. Learning to predict chemical reactions. *J Chem Inf Model.* 2011;51:2209–2222. http://dx.doi.org/10.1021/ci200207y.
44. Szymkuć S, Gajewska EP, Klucznik T, et al. Computer-assisted synthetic planning: the end of the beginning. *Angew Chem Int Ed.* 2016;55:5904–5937. http://dx.doi.org/10.1002/anie.201506101.
45. Klucznik T, Mikulak-Klucznik B, McCormack MP, et al. Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. *Chem.* 2018:1–11. http://dx.doi.org/10.1016/j.chempr.2018.02.002.
46. Subramanian G, Ramsundar B, Pande V, Denny RA. Computational modeling of $β$-secretase 1 (BACE-1) inhibitors using ligand based approaches. *J Chem Inf Model.* 2016;56:1936–1949. http://dx.doi.org/10.1021/acs.jcim.6b00290.
47. Ragoza M, Hochuli J, Idrobo E, Sunseri J, Koes DR. Protein-ligand scoring with convolutional neural networks. *J Chem Inf Model.* 2017;57:942–957. http://dx.doi.org/10.1021/acs.jcim.6b00740.
48. Babault N, Allali-Hassani A, Li F, et al. Discovery of bisubstrate inhibitors of nicotinamide N-methyltransferase (NNMT). *J Med Chem.* 2018;61:1541–1551. http://dx.doi.org/10.1021/acs.jmedchem.7b01422.
49. Gao H, Shanmugasundaram V, Lee P. Estimation of aqueous solubility of organic compounds with QSPR approach. *Pharm Res.* 2002;19(4):497–503 http://www.ncbi.nlm.nih.gov/pubmed/12033386. Accessed 16.01.18.
50. Gao H, Steyn SJ, Chang G, Lin J. Assessment of *in silico* models for fraction of unbound drug in human liver microsomes. *Expert Opin Drug Metab Toxicol.* 2010;6:533–542. http://dx.doi.org/10.1517/17425251003671022.
51. Lee PH, Cucurull-Sanchez L, Lu J, Du YJ. Development of in silico models for human liver microsomal stability. *J Comput Aided Mol Des.* 2007;21:665–673. http://dx.doi.org/10.1007/s10822-007-9124-0.
52. Stoner C, Troutman M, Gao H, et al. Moving in silico screening into practice: a minimalist approach to guide permeability screening!!. *Lett Drug Des Discovery.* 2006;3:575–581. http://dx.doi.org/10.2174/157018006778194736.
53. Alqahtani S. *In silico* ADME-Tox modeling: progress and prospects. *Expert Opin Drug Metab Toxicol.* 2017;13:1147–1158. http://dx.doi.org/10.1080/17425255.2017.1389897.
54. Bryan MC, Hein CD, Gao H, et al. Disubstituted 1-aryl-4-aminopiperidine library

synthesis using computational drug design and high-throughput batch and flow technologies. *ACS Comb Sci.* 2013;15:503–511. http://dx.doi.org/10.1021/co400078r.
55. Gurjar AS, Darekar MN, Yeong KY, Ooi L. In silico studies, synthesis and pharmacological evaluation to explore multi-targeted approach for imidazole analogues as potential cholinesterase inhibitors with neuroprotective role for Alzheimer's disease. *Bioorg Med Chem.* 2018. http://dx.doi.org/10.1016/j.bmc.2018.01.029 In Press.
56. Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery datasets. *Mol Pharm.* 2017. http://dx.doi.org/10.1021/acs.molpharmaceut.7b00578 acs.molpharmaceut.7b00578.
57. Xu Y, Ma J, Liaw A, Sheridan RP, Svetnik V. Demystifying multitask deep neural networks for quantitative structure-activity relationships. *J Chem Inf Model.* 2017;57:2490–2504. http://dx.doi.org/10.1021/acs.jcim.7b00087.
58. Kearnes S, McCloskey K, Berndl M, Pande V, Riley P. Molecular graph convolutions: moving beyond fingerprints. *J Comput Aided Mol Des.* 2016. http://dx.doi.org/10.1007/s10822-016-9938-8.
59. Ramsundar B, Liu B, Wu Z, et al. Is multitask deep learning practical for pharma? *J Chem Inf Model.* 2017. http://dx.doi.org/10.1021/acs.jcim.7b00146 acs.jcim.7b00146.
60. Liu K, Sun X, Jia L, et al. Chemi-net: a graph convolutional network for accurate drug property prediction. March 2018. http://arxiv.org/abs/1803.06236. Accessed 26.03.18.
61. Ramsundar B, Liu B, Wu Z, et al. Is multitask deep learning practical for pharma? *J Chem Inf Model.* 2017;57:2068–2076. http://dx.doi.org/10.1021/acs.jcim.7b00146.
62. Lombardo F, Obach RS, Shalaeva MY, Gao F. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. *J Med Chem.* 2002;45:2867–2876. http://dx.doi.org/10.1021/jm0200409.
63. Chen Y, Cameron K, Guzman-Perez A, Perry D, Li D, Gao H. Structure-pharmacokinetic relationship of in vivo rat biliary excretion. *Biopharm Drug Dispos.* 2010;31:82–90. http://dx.doi.org/10.1002/bdd.692.
64. Moriguchi I, Hirono S, Liu Q, Nakagome I. MAatsushita Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem Pharm Bull (Tokyo).* 1992;40:127–130. http://dx.doi.org/10.1248/cpb.40.127.
65. Platts JA, Butina D, Abraham MH, Hersey A. Estimation of molecular linear free energy relation descriptors using a group contribution approach. *J Chem Inf Comput Sci.* 1999;39:835–845. http://dx.doi.org/10.1021/ci980339t.
66. Labute P. A widely applicable set of descriptors. *J Mol Graph Model.* 2000;18:464–477. http://dx.doi.org/10.1016/S1093-3263(00)00068-1.
67. Weiss MM, Dineen TA, Marx IE, et al. Sulfonamides as selective Na $_{V}$ 1.7 inhibitors: optimizing potency and pharmacokinetics while mitigating metabolic liabilities. *J Med Chem.* 2017;60:5969–5989. http://dx.doi.org/10.1021/acs.jmedchem.6b01851.
68. Lombardo F, Jing Y. In silico prediction of volume of distribution in humans. extensive data set and the exploration of linear and nonlinear methods coupled with molecular interaction fields descriptors. *J Chem Inf Model.* 2016;56:2042–2052. http://dx.doi.org/10.1021/acs.jcim.6b00044.
69. Ahmed SSSJ, Ramakrishnan V. Systems Biological Approach of Molecular Descriptors Connectivity: Optimal Descriptors for Oral Bioavailability Prediction. Khan AU, ed. *PLoS One.* 2012;7:e40654. http://dx.doi.org/10.1371/journal.pone.0040654.
70. Tian S, Li Y, Wang J, Zhang J, Hou T. ADME, evaluation in drug discovery. 9. prediction of oral bioavailability in humans based on molecular properties and structural fingerprints. *Mol Pharm.* 2011;8:841–851. http://dx.doi.org/10.1021/mp100444g.
71. Olivares-Morales A, Hatley OJD, Turner D, Galetin A, Aarons L, Rostami-Hodjegan A. The use of ROC analysis for the qualitative prediction of human oral bioavailability from animal data. *Pharm Res.* 2014;31:720–730. http://dx.doi.org/10.1007/s11095-013-1193-2.
72. Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol.* 2004;1:337–341. http://dx.doi.org/10.1016/j.ddtec.2004.11.007.
73. Glen RC, Galloway WRJD, Spring DR, Liwiki G. Multiple-parameter optimization in drug discovery: example of the 5-HT1B GPCR. *Mol Inform.* 2016;35:599–605. http://dx.doi.org/10.1002/minf.201600056.
74. Page KM. Validation of early human dose prediction: a key metric for compound progression in drug discovery. *Mol Pharm.* 2016;13:609–620. http://dx.doi.org/10.1021/acs.molpharmaceut.5b00840.
75. Gao H. Multi-parameter optimization and in silico modeling in lead optimization. 18th European Symposium on quantitative structure-activity relationships. In: European Symposium on Quantitative Structure-Activity Relationships; 2010.
76. Kubiny H. Variable selection in QSAR studies. i. an evolutionary algorithm. *QSAR Comb Sci.* 1994;13:285–294. http://dx.doi.org/10.1002/qsar.19940130306.
77. Jia L, Zou J, So S-S, Sun H. Automated pharmacophore query optimization with genetic algorithmsa case study using the MC4R system. *J Chem Inf Model.* 2007;47:1545–1552. http://dx.doi.org/10.1021/ci700089w.
78. Martinez-Lopez Y, Caballero Y, Barigye SJ, et al. State of the art review and report of new tool for drug discovery. *Curr Top Med Chem.* 2017;17:2957–2976. http://dx.doi.org/10.2174/1568026617666170821123856.
79. Chen JJF, Visco DP. Identifying novel factor XIIa inhibitors with PCA-GA-SVM developed vHTS models. *Eur J Med Chem.* 2017;140:31–41. http://dx.doi.org/10.1016/j.ejmech.2017.08.056.
80. Ahmadi S, Ganji S. Genetic algorithm and self-organizing maps for qspr study of some N-aryl derivatives as butyrylcholinesterase inhibitors. *Curr Drug Discov Technol.* 2016;13:232–253. http://dx.doi.org/10.2174/1570163813666160725114241.
81. Wang N-N, Dong J, Deng Y-H, et al. ADME properties evaluation in drug discovery: prediction of caco-2 cell permeability using a combination of NSGA-II and boosting. *J Chem Inf Model.* 2015;56:763–773. http://dx.doi.org/10.1021/acs.jcim.5b00642.
82. Nekoei M, Mohammadhosseini M, Pourbasheer E. QSAR study of VEGFR-2 inhibitors by using genetic algorithm-multiple linear regressions (GA-MLR) and genetic algorithm-support vector machine (GA-SVM): a comparative approach. *Med Chem Res.* 2015;24:3037–3046. http://dx.doi.org/10.1007/s00044-015-1354-4.
83. Dey F, Caflisch A. Fragment-based de Novo Ligand design by multiobjective evolutionary optimization. *J Chem Inf Model.* 2008;48:679–690. http://dx.doi.org/10.1021/ci700424b.
84. Besnard J, Ruda GF, Setola V, et al. Automated design of ligands to polypharmacological profiles. *Nature.* 2012;492:215–220. http://dx.doi.org/10.1038/nature11691.
85. Besnard J. Design of molecules. 2011;(Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.). 138pp.
86. Sremath Tirumala S. Implementation of Evolutionary Algorithms for Deep Architectures. Vol 1315.; 2014.
87. Young SR, Rose DC, Karnowski TP, Lim S-H, Patton RM. Optimizing deep learning hyper-parameters through an evolutionary algorithm. In: Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments – MLHPC '15. ACM Press: New York, USA; 2015:1–5. doi:10.1145/2834892.2834896.
88. Hossain D, Capi G. Genetic Algorithm Based Deep Learning Parameters Tuning for Robot Object Recognition and Grasping. In: *International Scholarly and Scientific Research & Innovation*; 2017:629–633. doi:10.1999/1307-6892/10006712.
89. Fujino S, Hatanaka T, Mori N, Matsumoto. *The Evolutionary Deep Learning based on Deep Convolutional Neural Network for the Anime Storyboard Recognition.* Cham: Springer; 2018:278–285 10.1007/978-3-319-62410-5_34.
90. Kaicheng Li, Qingpeng Gan, Lei Yuan, Qiang Fu. Optimized generation of test sequences for high-speed train using deep learning and genetic algorithm. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE; 2016:784–789. doi:10.1109/ITSC.2016.7795644.
91. Jiang S, Chin KS, Wang L, Qu G, Tsui KL. Modified genetic algorithm-based feature selection combined with pre-trained deep neural network for demand forecasting in outpatient department. *Expert Syst Appl.* 2017;82:216–230. http://dx.doi.org/10.1016/j.eswa.2017.04.017.
92. Sharma A, Rani R. An optimized framework for cancer classification using deep learning and genetic algorithm. *J Med Imaging Heal Inf.* 2017;7:1851–1856. http://dx.doi.org/10.1166/jmihi.2017.2266.