Artificial intelligence in drug discovery: recent advances and future perspectives

Review Article

Author(s):
Jiménez-Luna, José; Grisoni, Francesca (iD); Weskamp, Nils; Schneider, Gisbert (iD)

Artificial intelligence in drug discovery: recent advances and future perspectives

José Jiménez-Luna, Francesca Grisoni, Nils Weskamp & Gisbert Schneider

Taylor & Francis
Taylor & Francis Group

REVIEW

🔓 OPEN ACCESS   Check for updates

Artificial intelligence in drug discovery: recent advances and future perspectives

José Jiménez-Luna , Francesca Grisoni , Nils Weskamp[b] and Gisbert Schneider

aDepartment of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland; bBoehringer Ingelheim Pharma GmbH & Co. KG, Biberach an Der Riss, Germany

ABSTRACT

Introduction: Artificial intelligence (AI) has inspired computer-aided drug discovery. The widespread adoption of machine learning, in particular deep learning, in multiple scientific disciplines, and the advances in computing hardware and software, among other factors, continue to fuel this development. Much of the initial skepticism regarding applications of AI in pharmaceutical discovery has started to vanish, consequently benefitting medicinal chemistry.

Areas covered: The current status of AI in chemoinformatics is reviewed. The topics discussed herein include quantitative structure-activity/property relationship and structure-based modeling, de novo molecular design, and chemical synthesis prediction. Advantages and limitations of current deep learning applications are highlighted, together with a perspective on next-generation AI for drug discovery.

Expert opinion: Deep learning-based approaches have only begun to address some fundamental problems in drug discovery. Certain methodological advances, such as message-passing models, spatial-symmetry-preserving networks, hybrid de novo design, and other innovative machine learning paradigms, will likely become commonplace and help address some of the most challenging questions. Open data sharing and model development will play a central role in the advancement of drug discovery with AI.

1. Introduction

Machine learning algorithms have been widely applied for computer-assisted drug discovery [1–3]. Deep learning approaches, that is, artificial neural networks with several hidden processing layers [4,5], have recently gathered renewed attention owing to their ability to perform automatic feature extractions from the input data, and their potential to capture nonlinear input–output relationships. These properties of deep learning techniques complement traditional machine learning approaches that rely on human-crafted molecular descriptors [6,7]. Drug discovery has experienced a relatively late resurgence in the interest for deep learning[8], which already led to an unprecedented explosion of novel modeling approaches and applications [9–12]. Many areas of the chemical sciences have already benefited from the ever-advancing developments in deep learning [13–15]. This opinion article delineates, through examples, some of the aspects that have allowed deep learning methodologies to flourish, and in some cases, outperform the existing approaches in chemoinformatics. Specifically, ligand-based quantitative structure-activity/property relationship (QSAR/QSPR), as well as structure-based modeling, de novo molecular design, and synthesis prediction are addressed (Figure 1). We also highlight the limitations of contemporary artificial intelligence (AI) in each of the considered topics and predict how it could shape the future landscape of computer-assisted drug discovery.

2. State of the art applications of artificial intelligence

2.1. QSAR/QSPR and structure-based modeling with artificial intelligence

QSAR/QSPR modeling has come a long way since its inception more than 50 years ago [16]. The impact of these computational models on drug discovery is undeniable, evidenced by the successful prediction of biological activity and pharmacokinetic parameters, viz. absorption, distribution, metabolism, excretion, and toxicity (ADMET) [17–21]. For ligand-based QSAR/QSPR modeling, the structural features of molecules (e. g. as pharmacophore distribution, physicochemical properties, and functional groups) are commonly converted into machine-readable numbers using the so-called molecular descriptors [7]. The spectrum of hand-crafted molecular descriptors is wide[7], aiming to capture a variety of aspects of the underlying chemical structure. In general, QSAR/QSPR approaches have transitioned from the use of simpler models, such as linear regression and k-nearest neighbors, toward more universally applicable machine learning techniques, such as support vector machines (SVM) and gradient boosting methods (GBM) [15], aiming to address more complex and potentially nonlinear relationships between the chemical

Article highlights

Ligand-based drug discovery built on novel deep learning techniques, such as message-passing neural networks, could facilitate the discovery of new chemical entities.

Deep learning techniques embedding three-dimensional symmetries as well as chemical information into their architecture bear promise for structure-based and conformation-aware modelling.

The combination of rule-based and rule-free approaches will further the capability of AI to deliver synthesizable bioactive molecules and explore new regions of chemical space.

Explainable AI, multitask, and meta-learning will pave the way for a new generation of predictive models with increased interpretability and robust performance in low-data regimes.

Natural language processing models will become commonplace solutions in both retrosynthesis and forward synthesis prediction. Additional effort will be drawn towards related problems such as the prediction of reaction conditions.

This box summarizes key points contained in the article.

structure and its physicochemical/biological properties, often at the expenses of intepretability [22].

Deep learning is not a new technique [23]. Artificial neural networks in chemoinformatics had their first heyday in the 1990s when many of the current concepts were pioneered [24–26], including deep and adaptive network architectures, self-organizing maps, recurrent systems for sequence and time-series analysis, and autoencoders. However, deep networks had their final breakthrough arguably after their success in the Merck Molecular Activity Challenge in 2012 [27]. While there is some controversy as to whether the latter type of models are superior performance-wise to other approaches (e. g. gradient boosting machines [28]) when using the same set of descriptors [29], deep learning methods offer several advantages. Arguably, the most important one is that deep networks can perform automatic feature extraction during the training procedure. Graph neural networks [12] (also referred to as message-passing approaches) and recurrent neural networks [30] in particular, are able to generate internal context-specific representations of molecular structures. In the specific case of graph neural networks, this is achieved by learning latent atom and bond representations during the training process. Therefore, deep learning approaches are promising for modeling tasks for which classical descriptors had not been initially engineered. Examples include the modeling of peptides [31], macrocycles, and proteolysis-targeting chimeras (PROTACs) [32]. Another potential advantage of deep architectures is their applicability to multitask learning [33–35], which aims to find a common internal representation that is useful for a set of related endpoints (not to be confused with multi-output learning which does not explicitly exploit related information between the tasks to be learned). As drug discovery is a multiparameter optimization challenge [36], multitask learning might make more efficient use of correlated data in common scenarios where the entirety of a molecular library is not exhaustively tested on all endpoints of interest, and without

the need for prior imputation. The idea of multi-output QSAR modeling, aiming to relate a set of predefined chemical descriptors to observable endpoints, had been explored before the rise of popularity of deep learning approaches [37–42]. Despite the promise of multitask learning, to date, only modest performance improvements over single-task models have been reported [43–46].

A well-known drawback of deep learning is its poor performance in medium-to-low data scenarios [47]. Some chemogenomic-based approaches might provide further insight in these scenarios by exploiting additional genomic or biological interactome data sources [48]. In addition, recent advances in 'few-shot' learning [49] (i.e. a set of approaches that can use prior knowledge to obtain better generalization when data is scarce) and meta-learning [50] (i.e. a family of methods that aims to develop a set of learnable parameters that can quickly adapt to new, unseen tasks) hold promise in mitigating some of these issues. Along those lines, purely data-driven approaches for molecular property predictions are, in contrast to techniques that are (fully or partially) physics-based, fundamentally limited in their ability to extrapolate and make reliable predictions for unseen compound classes. Physics-inspired machine learning approaches and additional active learning strategies (i.e. approaches where the model has a role in requesting specific training data for improved generalization) provide additional tools to overcome these limitations [51,52]. The success of these strategies, will furthermore critically depend on how well their specific implementations cope with data sparsity, given that suitable sources that would allow for efficient data imputation are often scarce [53].

Deep-learning models have also been widely criticized for their notorious debugging difficulty and 'black-box' character [54]. In contrast, the manual development of domain-specific features [55,56] (i.e. descriptors specifically engineered with a specific task in mind) still holds the potential to integrate background knowledge in a more human-intelligible way. Explainable AI techniques could offer partial solutions to these problems by providing comprehensible interpretations of the decision-making process undertaken by deep learning approaches [57]. Continued development of feature attribution techniques [58] (i. e. approaches that aim to highlight the overall importance of an input) instance-based explanations (e.g. counterfactuals, model-generated examples that are conditioned on user-defined queries) [54], and attention-based networks [59] will help narrow the gap between deep learning and drug-discovery specialists. Hence, close collaboration between these fields is imperative.

Another commonly-claimed disadvantage of deep-learning approaches is their high computational cost. Without specialized hardware such as consumer-grade graphical processing or tensor-processing units, deep learning typically entails longer training and evaluation times than many other machine-learning approaches. While the previous statement holds true under most scenarios, deep-learning models can learn in an online setting by naturally taking advantage of its most popular training strategy, i.e. stochastic gradient-descent optimization [60]. This has the advantage of scaling linearly with respect to the size of the training dataset, and thus it does not require the latter to be entirely loaded into the system's memory. We argue that the capability to train deep learning models stochastically
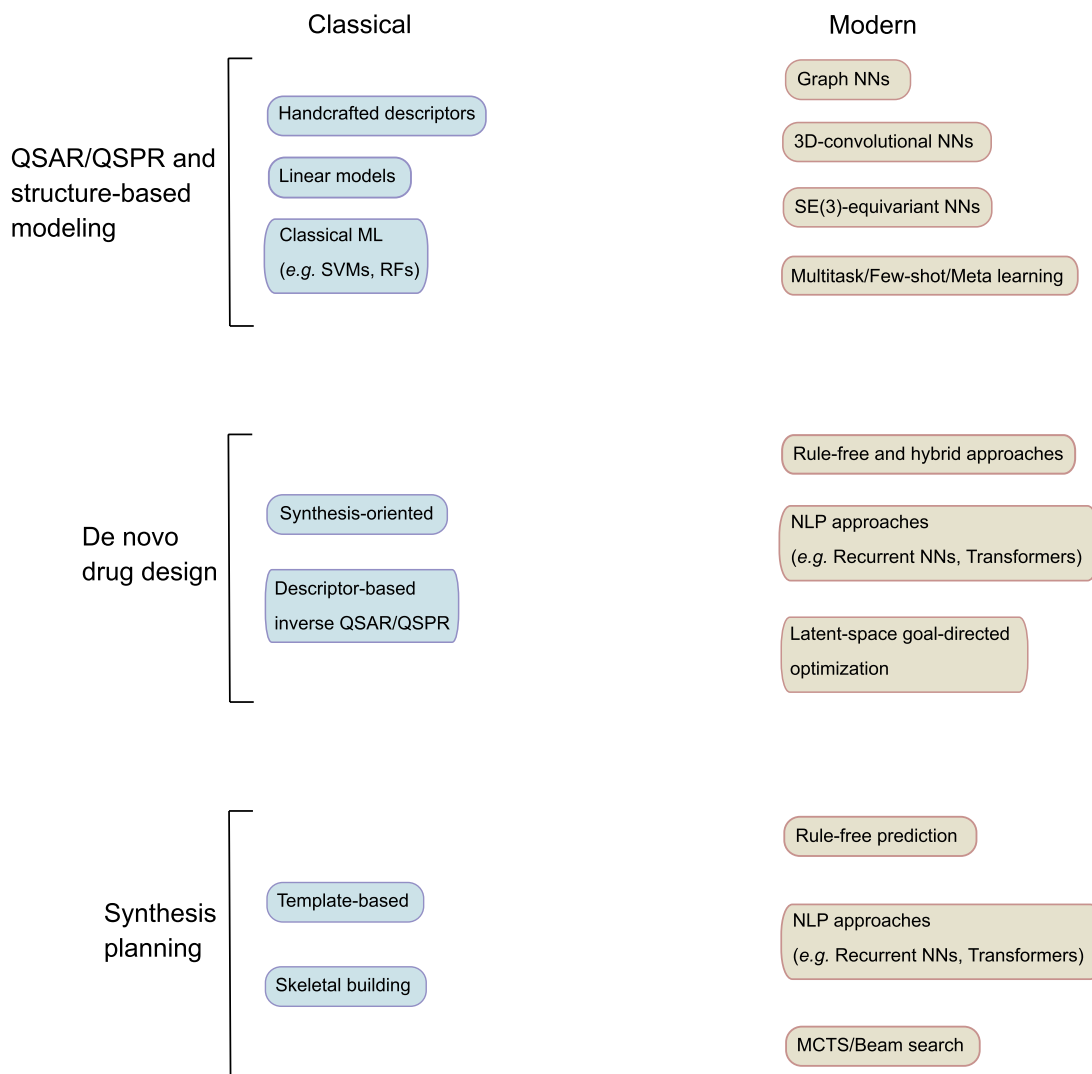
Figure 1. Schematic diagram of the transition between classical and modern methodologies for some relevant problems in drug discovery, such as QSAR/QSPR modeling, de novo drug design, and synthesis planning. Abbreviations: ML, machine learning; SVM, support vector machine; RF, random forest; QSAR/QSPR, quantitative structure-activity/property relationship; NN, neural network; SE(3), special Euclidean group in three-dimensions; NLP, natural language processing; MCTS, Monte Carlo tree search.

on sequential, random, batches of data can make them more suitable than other alternatives in big data scenarios [61]. A related issue is that predictive deep learning tends to require significantly more human expertise in many practical scenarios compared to other, more thoroughly tested approaches. For instance, while one can train a well-performing random forest model with a relatively small effort for hyperparameter tuning, our understanding of contemporary deep learning approaches is not yet at the level of reliable defaults, although recent theory suggests that this may change in the near future [62].

Furthermore, neural networks might provide the right answers for misleading reasons (*i.e.* the so-called Clever Hans effect [63]) and have a tendency to produce overly confident predictions, even when these are evidently wrong [64]. This is further exacerbated in the context of property prediction in drug discovery, as experiments under similar conditions can provide significantly different measurements. This drawback might be alleviated in the next few years with the wider adoption of uncertainty estimation techniques, either with deep learning approaches that have uncertainty directly embedded into their design, such as Bayesian neural networks [65], or *post hoc* techniques such as ensemble learning [66].

Remarkable progress has also been made in the structure-based prediction of protein-ligand activities which, unlike classical QSAR, requires a co-crystal or a docked pose to generalize over different targets. Many classical approaches modeled an explicit, predefined

mathematical relationship of the protein-ligand complex via partial least squares or multiple linear regression, in order to accurately consider the contribution of individual descriptors (*e.g.* physico-chemical properties) for a target property [67–71]. Approaches making use of more advanced as well as more flexible non-linear models, such as random forests or support vector machines, appeared in the 2000s [72], and became popular in the early 2010s, coupled with a wide range of descriptors such as protein-ligand atom pair counts [73], property-encoded shape distributions [74], or basic atomic interactions [75]. Similar to its purely ligand-based counterpart, this particular subfield has recently witnessed the advent of deep learning and used it to its advantage. Early approaches were inspired by the advances of computer vision and image recognition which were mostly driven by convolutional neural networks [76], and ultimately adapted for bioactivity prediction [77–80]. Others studies used graph-based approaches in conjunction with distance and angle-based featurizations toward the same goal [10]. Some of these were reported to provide incremental performance improvements over previous approaches in structure-based virtual screening and lead optimization competitions [81,82], although it is argued whether some well-known benchmarks tend to favor ML-based scoring functions over classical ones [83]. In keeping with an increased interest in interpretable AI, recent attempts toward explaining structure-based convolutional neural network models have shown that these are able to highlight relevant protein-ligand interactions in comprehensible terms, such as hydrogen bridging and π-π stacking [79].

Approaches based on three-dimensional convolutional neural networks, however, possessed certain theoretical limitations, namely the lack of rotational invariance with respect to the input, a desirable property when modeling atomistic systems. How to overcome this issue has recently become a very active area of research, with newly developed neural network architectures such as Euclidean Neural Networks [84–86] and SchNet [87], featuring equivariance with respect to the special Euclidean group in three-dimensions (SE(3)) (*i.e.* rotations and translations) directly embedded into their design. These architectures have already been applied to several molecular tasks, such as the prediction of electronic properties of molecules [88]. Research in this direction is expected to intensify in the near future, opening up fresh modeling opportunities.

Given the growth of deep learning applications in drug discovery and the fact that these methods benefit from large training sets, diligent data curation and proper benchmarking of newly developed models is mandatory. The availability and size of chemical compound libraries has improved over the past few years, with databases such as ZINC [89] and ChEMBL [90] representing a commonly-used starting point for ligand-based projects. A similar trend was observed for structure-based modeling, for which databases such as PDBbind [91] and BindingDB [92] provide highly detailed structural information on protein-ligand complexes, as well as their associated bioactivity data. Recent progress in the field of protein structure prediction and determination warrants optimism that structural information for many more drug targets will become available in the future [93,94].

Much has already been invested in open and standardized assessments of machine learning methodologies in the context of cheminformatics. For example, the MoleculeNet benchmarking suite [9] aims to facilitate model testing by providing a timely evaluation of many popular deep learning architectures for drug-related property predictions in well-curated datasets from areas such as biophysics, physical chemistry, and physiology. Nonetheless, while we have argued that the amount of public data is increasing at a fast pace, most of the structural activity/property relationship data are still generated by commercial research organizations, publishers, and pharmaceutical companies [95–97], which often consider the generated data as a differentiating asset to be kept confidential. Recent developments have shown that molecular structures can often be partially recovered from molecular descriptors, which may further complicate data sharing even at the latent feature level [98]. Attempts to overcome such limitations, *for example*, by developing federated and intellectual property (IP)-preserving learning techniques, are underway [99].

With regards to model evaluation, it is now known that testing performance on sets drawn from a database in a pseudo-random fashion can produce overly confident results. Alternatives such as scaffold-based [100] or time-based splits [101], which try to approximate the course of a lead optimization project, are potentially more informative. However, one should keep in mind that there is no such thing as a 'one-split-fits-all' strategy, as each evaluation portrays model performance in a different applicability domain. Furthermore, even though prospective applications should be considered the gold standard for model benchmarking, we note that these are not necessarily straightforward and are not devoid of biases either [102]. However, while the lack of benchmarking consensus is not ideal, it has not prevented machine-learning scoring functions [103] to be surprisingly predictive in some virtual screening campaigns [104]. Additionally, many efforts have been dedicated to the use of proper performance metrics for classification and regression models, and the limitations thereof [105–108].

2.2. De novo drug design with artificial intelligence

De novo design, the generation of novel molecular entities with desired pharmacological properties from scratch [109], can be considered as one of the most challenging computer-assisted tasks in drug discovery, due to the cardinality of the chemical space of drug-like molecules (estimated to range in the order of 10^{60}–10^{100}) [110,111]. De novo molecule generation faces the problem of combinatorial explosion due to the number of different atomic types and molecular topologies one could investigate [112]. Depending on the information used to guide the de novo design, the respective approaches can either be ligand-based, structure-based, or a mixture of both.

Ligand-based methodologies may be divided into two major categories: (i) rule-based approaches, which use a set of construction rules for molecule assembly from a set of 'building blocks' (*i.e.* reagents or molecular fragments), and (ii) rule-free approaches, which do not employ explicit construction rules. One of the ancestors of contemporary rule-based de novo design is the Topliss scheme [113], for the

stepwise generation of analogs of an active lead compound to maximize potency [113]. Contemporary approaches are based on applying a given set of molecular transformations for optimization, such as matched molecular pairs [114,115], or rules-of-thumb for functional group and molecular framework modification [116]. Synthesis-oriented approaches explicitly include synthesis rules for building block assembly and ligand generation. These approaches are useful, for instance, to design synthetically accessible libraries [117,118], such as BI CLAIM [119] and CHIPMUNK [120]. Since the late 1990s, hybrid approaches, such as TOPAS [121], DOGS [122], and DINGOS [123], have been developed to steer the generation of novel compounds by simultaneously maximizing both their similarity to known bioactive ligands and the chemical synthesizability of the designs.

'Rule-free' approaches aim to directly generate molecules with desired properties without the need for molecular construction rules. Contemporary approaches are often based on generative deep learning models [124], which sample new molecules from a learned latent molecular representation. Although these approaches have gained popularity in the last few years, the idea of sampling from a numerical representation of molecules for de novo design dates back to the 'inverse QSAR' problem formulated in the pioneering work of Skvortsova and Zefirov in the early 1990s [125–127]. Inverse QSAR leverages an existing QSAR model to identify the descriptor values corresponding to a desired property, and uses this information for molecule generation [128–131]. The latter approaches pose several challenges, such as the existence of multiple solutions for any given property and the issue of reverse-decoding molecular descriptors into valid structures. Generative deep learning overcomes some of these issues by modeling the underlying distribution of a given set of molecules, and then generating novel compounds by sampling from the learned distribution [132]. The most commonly used generative models are those borrowed from the field of natural language processing, coupled with Simplified Molecular Input Line Entry Systems (SMILES) [133]. These models are trained to learn the SMILES 'syntax' (*i.e.* how to generate a chemically valid string) on chosen 'semantics' (*i.e.* bioactivity or other desired molecular properties). They are mostly based on recurrent neural networks [134,135], coupled with transfer [134,135] or reinforcement learning [136–138]. Other popular deep-learning-based generative learning models, such as variational autoencoders [139], or generative adversarial networks [140,141] have also been commonly reported, as well as others based on graph convolutions [142,143]. Recently, instances of conditional generative approaches have been suggested, which leverage additional information guiding the design, such as three-dimensional shape [144], drug-likeness [142], synthesizability [142,145], molecular descriptors values [146], and gene expression signatures [147]. A major upcoming challenge in this context will be the definition of balanced objective functions that enable complex and constrained multi-parameter optimizations, similarly to those used in Pareto [148] or in desirability-based approaches [149–152], that are typically required in drug discovery.

Fueled in part by the rapid development of novel generative neural network approaches, the number of ligand-based design methods has skyrocketed. A recent review [153] reported more than 40 new models that were developed in the last couple of years. This explosion of potential drug-design tools has motivated researchers to evaluate and benchmark generative approaches in a fair and standardized manner. Recent efforts include the MOSES [154] and GuacaMol platforms [155], which implement several popular neural generative models, as well as more classical models (*e.g.* genetic algorithms [156]), and provide several metrics for their comparison. While de novo design tools are in general more difficult to evaluate retrospectively than predictive methods, some of the commonly used metrics are: (i) validity of the generated molecular representations and novelty of the corresponding molecules, (ii) similarity to known compounds in terms of chemical and biological properties [157], and (iii) scaffold and fragment diversity.

Rule-based and rule-free approaches have different advantages. Rule-based methods, by relying on preexisting knowledge, such as building blocks and reaction rules, can generate molecules that are often readily synthesizable and possess the desired physicochemical properties. However, the chemical diversity of the designs is influenced by the hard-coded rules and the chosen building block libraries. Rule-free approaches learn directly from the data without the need of hard-coded design/similarity rules, thus theoretically allowing a broader exploration of the chemical space. As a downside, this freedom of exploration risks the generation of compounds that are more difficult, if not impossible, to synthesize. Mixed approaches combining rule-free and rule-based methods might represent a promising middle ground for the design of novel bioactive and synthesizable molecular entities. Recently, a mixed strategy showed promise in designing bioactive molecules in a rule-free manner, while at the same time retaining synthesizability within a microfluidics system, thanks to a set of predefined virtual reactions [145].

To date, most of the deep-learning-based de novo design studies have focused on ligand-based approaches. Structure-based generative design constitutes a promising complementary research direction for targeting orphan receptors and hitherto unexplored macromolecules [158]. These approaches, which typically leverage information about the ligand-binding site (*e.g.* by fragment linking or growing[109]), to the best of our knowledge, have not been yet permeated extensively by deep learning. However, initial developments for ligand design have emerged by taking into account the shape and properties of the binding pocket [159–161].

2.3. Automated synthesis planning with artificial intelligence

The majority of all known organic compounds can be synthesized with a limited number of robust reactions [162]. However, reliable and fully automated synthesis planning in chemistry is a challenge that is yet to be met [163]. Part of the reason is owing to the extensive chemistry expertise that is required for efficient forward and retrosynthetic planning [164]. Synthesis planning with AI has a rich history, dating back to the 1970s in the field of computer-aided retrosynthetic prediction [165]. Increased computational capabilities, the advent of big data, and the development of novel algorithms for deep learning and optimization, have resulted in a resurgence of AI for synthetic organic chemistry. In retrosynthesis,

where the main goal is to recursively design efficient synthetic routes for a molecule of interest, rule-based methods [122,123,166] are indisputably valuable. These aim to suggest retro-synthetic pathways via reaction mechanism encoding and skeletal building. One of their main limitations is their dependency on explicit chemical transformations/reactions. These usually entail manual construction and curation. The field has recently drawn inspiration from natural language processing methods, such as sequence-to-sequence models [167] and transformer models [168]. This line of research is motivated by the observation that the rank distribution of fragments in organic molecules is similar to that of words in the English language [169]. Rule-free approaches typically consider products in a text-based representation (e.g. SMILES), and process them via an encoder-decoder architecture, which is subsequently used to predict the corresponding synthetic precursors at a one-step reaction distance [170]. Improvements over this architecture feature the use of tiered neural networks [171], whose goal is to partition the retrosynthesis prediction problem into reaction type classification and reaction rule selection steps. This separation, inspired by a previously reported molecular similarity method [172], was shown to achieve performance gains over previous baselines.

While most of the above-described methods focus on the linear one-step retrosynthesis problem, a more realistic scenario faces a rapidly exploding combinatorial problem. Inspired by progress in reinforcement learning, one of the most important breakthroughs in the last few years came from the widespread usage of sophisticated search methods, such as Monte Carlo Tree Search [173], to efficiently navigate through chemical reaction spaces [174]. A recent study [175] has attempted to elucidate both reactants and reagents via the use of transformer models, employing one-step precursor predictions in combination with the construction of hyper-graphs (i.e. directed acyclic graphs where edges can link multiple nodes simultaneously), which represent synthetic pathways. In order to find a reasonable synthetic pathway, these hyper-graphs are explored with beam search, aided by a Bayesian-like probability scheme that is biased toward the suggestion of chemically simpler precursors [172].

Forward synthesis planning distinguishes itself from retrosynthesis. While the latter might be solvable by leveraging existing reaction databases, forward synthesis would require information from reactions that yield no product whatsoever. The current chemical reaction databases are heavily skewed toward productive reaction data [176]. There is a critical demand for additional data, such as experimental conditions (e.g. solvent and temperature) or side-product information. With the aim of addressing some of these limitations, several steps have been taken to expand known reaction databases with negative reaction outcomes [177], and thereby, create new customized data compilations for automated synthesis planning [178].

Some of the earlier approaches ranked candidate products using hard-coded reaction templates derived from data [177,179]. Proof-of-concept machine learning ranked reaction templates, when the details of reactants and reagents were given [180]. Newer approaches aim to directly rank products by viewing the chemical reaction prediction problem as a graph transformation task [181,182]. Driven by advances in quantum mechanics, another set of approaches opted for using first-principle calculations (e.g. density functional theory) to evaluate the energy barriers of a particular reaction. However, this approach is computationally prohibitive for medium-to-large systems. The accurate prediction of energies and forces [183] via quantum-mechanical machine learning might help bridge this gap in the near future. With regard to template-free forward synthesis prediction, natural language processing approaches based on transformer [184] or recurrent neural network architectures [185] are also becoming popular. These have reported a top-1 reactant accuracy above 90%. Other recent alternative deep learning approaches [186,187] have opted to encode reaction prediction as an electron rearrangement exercise alongside the usage of message-passing neural networks. The latter approach, however, requires filtering reactions where electron flow is not directly identifiable, which excludes many relevant organic ones.

3. Expert opinion

Evidence suggests that AI applications are starting to become ubiquitous in drug discovery and design. These techniques are slowly living up to some of the community's expectations, with remarkable advances in QSAR modeling, de novo molecular design, and synthesis planning, among others. However, whether these techniques will actually prove useful by aiding researchers to design and synthesize 'better drug candidates faster' still remains to be demonstrated [188,189].

In the context of ligand-based property prediction, methods relying on more 'raw' chemical representations (e.g. graph neural networks and SMILES-based recurrent neural networks) can be anticipated to perform at least on-par with standard descriptor-based models. Moreover, deep learning approaches are easily adaptable to a wider class of chemical entities and modeling tasks, and allow for a more efficient use of data, for example, via multitask and online learning. In contrast, conformation-aware deep learning, especially considering methodologies that embed three-dimensional symmetries into their design, are still in their infancy. Nonetheless, rapid progress can be expected toward their application in drug discovery and related areas, such as quantum mechanics and material science, particularly as a proxy for first-principle calculations, which are computationally more demanding.

In de novo drug design, we have been slowly witnessing an augmentation of rule-based approaches along with rule-free approaches in the past few years. While the latter hold promise in exploring unseen regions of the chemical space, they also come with limitations, such as limited synthesizability. Mixing rule-free and rule-based methods (i.e. 'hybrid' methods) might provide a pragmatic solution. Particular attention will be drawn toward generative approaches that can exploit additional sources of information, such as some of the pioneering works including gene expression [120], conformational space [123], and ligand binding site information [131,132].

For automated synthesis planning and reaction prediction, advanced natural language processing will continue to inspire and drive innovation. Much needed attention will be drawn to commonly

less explored topics, such as yield estimation, side-product formation, and the prediction of suitable reaction conditions. Advances in robotics and reinforcement learning will lay the groundwork for fully automated synthesis in the next few years.

The newfound interest in explainable AI [57], with methodologies such as feature attribution [190], instance-based molecular counterfactual explanation [191], and uncertainty estimation [64], will increase the acceptance of AI-supported drug discovery. The development and validation of these techniques will require further interdisciplinary research. Special consideration will also be given to approaches that can exploit information in low-data regimes, such as transfer learning [192], as well as multi-task and meta-learning [193]. The barriers against learning and prospectively applying deep learning approaches have been greatly lowered for interested practitioners in the last few years. The current trend suggests that these methods will be increasingly accessible in the foreseeable future, with the continued development of general high-level research and deployment software packages [194,195], as well as comprehensible documentation.

Funding

Declaration of interest

N Weskamp is an employee of Boehringer Ingelheim. G Schneider is a cofounder of inSili.com LLC, Zurich and a consultant to the pharmaceutical industry. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

ORCID

José Jiménez-Luna 🔵 http://orcid.org/0000-0002-5335-7834
Francesca Grisoni 🔵 http://orcid.org/0000-0001-8552-6615
Gisbert Schneider 🔵 http://orcid.org/0000-0001-6706-1084

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat. Rev. Drug Discov. 2019;18(6):463–477.
2. Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. Drug Discov Today. 2015;20(3):318–331.
3. Lo Y-C, Rensi SE, Torng W, et al. Machine learning in chemoinformatics and drug discovery. Drug Discov Today. 2018;23(8):1538–1546.
4. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117.
5. Yang X, Wang Y, Byrne R, et al. Concepts of artificial intelligence for computer-assisted drug discovery. Chem. Rev. 2019;119(18):10520–10594.
6. Xue L, Bajorath J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. Comb. Chem. High Throughput Screen. 2000;3(5):363–372.
7. Todeschini R, Consonni V. Molecular descriptors for chemoinformatics: volume i: alphabetical listing/volume ii: appendices, references. Weinheim, Germany: John Wiley & Sons; 2009.
8. Schneider G. Mind and machine in drug design. Nat. Mach. Intell. 2019;1(3):128–130.
9. Wu Z, Ramsundar B, Feinberg E, et al. MoleculeNet: a benchmark for molecular machine learning. Chem. Sci. 2018;9(2):513–530.
10. Feinberg EN, Sur D, Wu Z, et al. PotentialNet for molecular property prediction. ACS Cent Sci. 2018;4(11):1520–1530.
11. Kearnes S, McCloskey K, Berndl M, et al. Molecular graph convolutions: moving beyond fingerprints. J. Comput. Aided Mol. Des. 2016;30(8):595–608.
12. Gilmer J, Schoenholz SS, Riley PF et al. Neural message passing for quantum chemistry. arXiv [cs.LG] (2017).
13. Gawehn E, Hiss JA, Schneider G. Deep learning in drug discovery. Mol Inform. 2016;35(1):3–14.
14. Zhang L, Tan J, Han D, et al. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today. 2017;22(11):1680–1685.
15. Chen H, Engkvist O, Wang Y, et al., The rise of deep learning in drug discovery. Drug Discov Today. 23(6): 1241–1250. 2018..
   (••) An excellent introduction to the basics of deep learning in the context of drug discovery pipelines.
16. Hansch C, Maloney PP, Fujita T, et al. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature. 1962;194(4824):178–180.
17. Göller A, Kuhnke L, Montanari F, et al. Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. Drug Discov Today. 2020;25(9):1702–1709.
18. Winiwarter S, Ahlberg E, Watson E, et al. In silico ADME in drug design – enhancing the impact. ADMET DMPK. 2018;6(1):15–33.
19. Beresford AP, Segall M, Tarbit MH. In silico prediction of ADME properties: are we making progress? Curr. Opin. Drug Discov. Devel. 2004;7(1):36–42.
20. Norinder U, Bergström CAS. Prediction of ADMET properties. ChemMedChem. 2006;1(9):920–937.
21. Beck B, Geppert T. Industrial applications of in silico ADMET. J. Mol. Model. 2014;20(7). DOI:10.1007/s00894-014-2322-5
22. Fujita T, Winkler DA. Understanding the roles of the 'two QSARs'. J. Chem. Inf. Model. 2016;56(2):269–274.
   (•) This paper provides a good starting point for understanding the relationship between explainable and performant QSAR/QSPR models
23. Rumelhart DE, McClelland JL. Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations. Cambridge, MA: MIT Press; 1986.
24. Zupan J, Gasteiger J. Neural networks for chemists: an introduction. New York, NY: John Wiley & Sons, Inc.; 1993.
25. Devillers J. Neural networks in QSAR and drug design. Lyon, France: Academic Press; 1996.
26. Schneider G. Adaptive systems in drug design. Boca Raton, FL: CRC Press; 2002.
27. Unterthiner T, Mayr A, Klambauer G et al. Deep learning as an opportunity in virtual screening. Proceedings of the deep learning workshop at NIPS, Montreal, Canada. 27, 1–9 (2014).
28. Sheridan RP, Wang WM, Liaw A, et al. Extreme gradient boosting as a method for quantitative structure–activity relationships. J. Chem. Inf. Model. 2016;56(12):2353–2360.
29. Winkler DA, Le TC. Performance of deep and shallow neural networks, the universal approximation theorem, activity cliffs, and QSAR. Mol. Inform. 2017;36(1–2):1600118.

30. Chakravarti SK, Alla SRM. Descriptor-free QSAR modeling using deep learning with long short-term memory neural networks. Front. Artif. Intell. Appl 2019;2:17.

31. Henninot A, Collins JC, Nuss JM. The current state of peptide drug discovery: back to the future? J. Med. Chem. 2018;61(4):1382–1414.

32. Zou Y, Ma D, The WY. PROTAC technology in drug development. Cell Biochem. Funct. 2019;37(1):21–30.

33. Ramsundar B, Kearnes S, Riley P et al. Massively multitask networks for drug discovery. arXiv [stat.ML] (2015).

34. Simões RS, Maltarollo VG, Oliveira PR, et al. Transfer and multi-task learning in QSAR modeling: advances and challenges. Front Pharmacol. 2018;9(74). DOI:10.3389/fphar.2018.00074

35. Lee K, Kim D. In-Silico molecular binding prediction for human drug targets using deep neural multi-task learning. Genes (Basel). 2019;10(906):906.

36. Sosnin S, Karlov D, Tetko IV, et al. Comparative study of multitask toxicity modeling on a broad chemical space. J. Chem. Inf. Model. 2019;59(3):1062–1072.

37. Vilar S, Santana L, Uriarte E. Probabilistic neural network model for the in silico evaluation of anti-HIV activity and mechanism of action. J. Med. Chem. 2006;49(3):1118–1124.

38. Prado-Prado FJ, García-Mera X, González-Díaz H. Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. Bioorg. Med. Chem. 2010;18 (6):2225–2231.

39. Speck-Planche A, Kleandrova VV, Luan F, et al. Rational drug design for anti-cancer chemotherapy: multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. Bioorg. Med. Chem. 2012;20(15):4848–4855.

40. Speck-Planche A, Kleandrova VV, Cordeiro MNDS. 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(10):2727–2732.

41. Speck-Planche A, Cordeiro MNDS. Multitasking models for quantitative structure–biological effect relationships: current status and future perspectives to speed up drug discovery. Expert Opin. Drug Discov. 2015;10(3):245–256.

42. Ambure P, Halder AK, González Díaz H, et al. QSAR-Co: an open source software for developing robust multitasking or multitarget classification-based QSAR models. J. Chem. Inf. Model. 2019;59 (6):2538–2544.

43. Montanari F, Kuhnke L, Ter Laak A, et al. Modeling physico-chemical ADMET endpoints with multitask graph convolutional networks. Molecules. 2019;25(1):44.

44. Wenzel J, Matter H, Schmidt F. Predictive multitask deep neural network models for ADME-Tox properties: learning from large data sets. J. Chem. Inf. Model. 2019;59(3):1253–1268.

45. Rodríguez-Pérez R, Bajorath J. Prediction of compound profiling matrices, part II: relative performance of multitask deep learning and random forest classification on the basis of varying amounts of training data. ACS Omega. 2018;3(9):12033–12040.

46. Merget B, Turk S, Eid S, et al. Profiling prediction of kinase inhibitors: toward the virtual assay. J. Med. Chem. 2017;60 (1):474–485.

47. Lenselink EB, Ten Dijke N, Bongers B, et al. Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. J Cheminform. 2017;9(45). DOI:10.1186/s13321-017-0232-0.

48. Chu Y, Kaushik AM, Wang X, et al. DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. Brief. Bioinform. 2021;22(1):451–462.

49. Altae-Tran H, Ramsundar B, Pappu AS, et al. Low data drug discovery with one-shot learning. ACS Cent Sci. 2017;3(4):283–293.

50. Finn C, Abbeel P, Levine S Model-agnostic meta-learning for fast adaptation of deep networks. arXiv [cs.LG] (2017).

51. Reker D, Schneider G. Active-learning strategies in computer-assisted drug discovery. Drug Discov Today. 2015;20 (4):458–465.

52. Reker D, Schneider P, Schneider G, et al. Active learning for computational chemogenomics. Future Med. Chem. 2017;9(4):381–402.

53. Irwin BWJ, Levell JR, Whitehead TM, et al. Practical applications of deep learning to impute heterogeneous drug discovery data. J. Chem. Inf. Model. 2020;60(6):2848–2857.

54. Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv. JL & Tech 2017;31:841–887.

55. Mauri A. alvaDesc: a tool to calculate and analyze molecular descriptors and fingerprints. In: Roy K, editor. Ecotoxicological QSARs. New York, NY: Springer US; 2020. p. 801–820.

56. Moriwaki H, Tian Y-S, Kawashita N, et al. A molecular descriptor calculator. J Cheminform. 2018;10(4). DOI:10.1186/s13321-018-0258-y

57. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat. Mach. Intell. 2020;2 (10):573–584.

58. Sundararajan M, Taly A, Yan Q Axiomatic attribution for deep networks. arXiv [cs.LG] (2017).

59. Preuer K, Klambauer G, Rippmann F, et al. Interpretable deep learning in drug discovery. In: Samek W, Montavon G, Vedaldi A, et al., editors. Explainable AI: interpreting, explaining and visualizing deep learning. Cham, Switzerland: Springer International Publishing; 2019. p. 331–345.

60. Zhang Y, Lee AA. Bayesian semi-supervised learning for uncertainty-calibrated prediction of molecular properties and active learning. Chem. Sci. 2019;10(35):8154–8163.

61. Gawehn E, Hiss JA, Brown JB, et al. Advancing drug discovery via GPU-based deep learning. Expert Opin. Drug Discov. 2018;13 (7):579–582.

62. Belkin M, Hsu D, Ma S, et al. Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proc Natl Acad Sci USA. 2019;116(32):15849–15854.

63. Lapuschkin S, Wäldchen S, Binder A, et al. Unmasking Clever Hans predictors and assessing what machines really learn. Nat. Commun. 2019;10(1):1096.

64. Nguyen A, Yosinski J, Clune J Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. in Proceedings of the IEEE conference on computer vision and pattern recognition,Boston, MA. 427–436 (2015).

65. Graves A. Practical variational inference for neural networks. In: Shawe-Taylor J, Zemel RS, Bartlett PL, et al., editors. Adv. Neural Inf. Process. Syst. 24. Granada, Spain: Curran Associates, Inc.; 2011. p. 2348–2356.

66. Lakshminarayanan B, Pritzel A, Blundell C, et al. Simple and scalable predictive uncertainty estimation using deep ensembles. In: Guyon I, editors. Adv. Neural inf. process. syst. 30. Long Beach, CA: Curran Associates, Inc.; 2017. p. 6402–6413.

67. Cao Y, Li L. Improved protein–ligand binding affinity prediction by using a curvature-dependent surface-area model. Bioinformatics. 2014;30(12):1674–1680.

68. Wang R, Lai L, Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J. Comput. Aided Mol. Des. 2002;16(1):11–26.

69. Verdonk ML, Cole JC, Hartshorn MJ, et al. Improved protein-ligand docking using GOLD. Proteins. 2003;52(4):609–623.

70. Böhm HJ. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. J. Comput. Aided Mol. Des. 1998;12(4):309–323.

71. Wang R, Liu L, Lai L, et al. SCORE: a new empirical method for estimating the binding affinity of a protein-ligand complex. Mol. Mod. Annu. 1998;4(12):379–394.

72. Ain QU, Aleksandrova A, Roessler FD, et al. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2015;5(6):405–424.

73. Ballester PJ, Mitchell JBO. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics. 2010;26(9):1169–1175.

74. Sato T, Honma T, Yokoyama S. Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening. J. Chem. Inf. Model. 2010;50(1):170–185.

75. Zilian D, Sotriffer CA. SFCscoreRF: a random forest-based scoring function for improved affinity prediction of protein–ligand complexes. J. Chem. Inf. Model. 2013;53(8):1923–1933.

76. Pereira JC, Caffarena ER. dos Santos, C. N. boosting docking-based virtual screening with deep learning. J. Chem. Inf. Model. 2016;56 (12):2495–2506.

77. Ragoza M, Hochuli J, Idrobo E, et al. Protein–ligand scoring with convolutional neural networks. J. Chem. Inf. Model. 2017;57 (4):942–957.

78. Jiménez J, Škalič M, Martínez-Rosell G, et al. K DEEP: protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. J. Chem. Inf. Model. 2018;58(2):287–296.

79. Hochuli J, Helbling A, Skaist T, et al. Visualizing convolutional neural network protein-ligand scoring. J Mol Graph Model. 2018;84:96–108.

80. Skalic M, Martínez-Rosell G, Jiménez J, et al. PlayMolecule BindScope: large scale CNN-based virtual screening on the web. Bioinformatics. 2019;35:1237–1238.

81. Sunseri J, King JE, Francoeur PG, et al. Convolutional neural network scoring and minimization in the D3R 2017 community challenge. J. Comput. Aided Mol. Des. 2019;33(1):19–34.

82. Li H, Sze K, Lu G, et al. Machine-learning scoring functions for structure-based virtual screening. Wiley Interdiscip Rev Comput Mol Sci. 2021;11(1). DOI:10.1002/wcms.1478

83. Sieg J, Flachsenberg F, Rarey M. In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening. J. Chem. Inf. Model. 2019;59(3):947–961.

84. Thomas N, Smidt T, Kearnes S et al. Tensor field networks: rotation-and translation-equivariant neural networks for 3d point clouds. arXiv preprint arXiv (2018).
   (••) This study showcases the development of one of the first SE(3)-equivariant neural networks for atomistic systems.

85. Cohen TS, Geiger M, Köhler J, et al. Spherical CNNs. arXiv preprint arXiv (2018).

86. Anderson B, Hy TS, Kondor R. Cormorant: covariant molecular neural networks. Adv. Neural Inf. Process. Syst. 2019;32.

87. Schütt KT, Sauceda HE, Kindermans PJ, et al. SchNet–A deep learning architecture for molecules and materials. J. Chem. Phys. 2018;148(241722):241722.

88. Qiao Z, Welborn M, Anandkumar A, et al. OrbNet: deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. J. Chem. Phys. 2020;153(124111):124111.

89. Irwin JJ, Shoichet BK. ZINC - A free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005;45:177–182.

90. Gaulton A, Hersey A, Nowotka M, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945–D954.

91. Liu Z, Li Y, Han L, et al. PDB-wide collection of binding data: current status of the PDBbind database. Bioinformatics. 2015;31(3):405–412.

92. Liu T, Lin Y, Wen X, et al. BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. Nucleic Acids Res. 2006;35(Database):D198–D201.

93. Senior AW, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. Nature. 2020;577 (7792):706–710.

94. Si D, Moritz SA, Pfab J, et al. Deep learning to predict protein backbone structure from high-resolution cryo-EM density maps. Sci Rep. 2020;10(4282). DOI:10.1038/s41598-020-60598-y.

95. Wassermann AM, Lounkine E, Hoepfner D, et al. Dark chemical matter as a promising starting point for drug lead discovery. Nat Chem Biol. 2015;11:958–966.

96. Engels MFM, Gibbs AC, Jaeger EP, et al. A cluster-based strategy for assessing the overlap between large chemical libraries and its application to a recent acquisition. J. Chem. Inf. Model. 2006;46(6):2651–2660.

97. Kogej T, Blomberg N, Greasly PJ, et al. Big pharma screening collections: more of the same or unique libraries? The AstraZeneca-Bayer Pharma AG case. Drug Discov Today. 2013;18 (19–20):1014–1024.

98. Le T, Winter R, Noé F, et al. Neuraldecipher – reverse-engineering extended-connectivity fingerprints (ECFPs) to their molecular structures. Chem. Sci. 2020;11(38):10378–10389.

99. The HC. MELLODDY project. Fierce Biotech. 2019;1.

100. Kramer C, Gedeck P. Leave-cluster-out cross-validation is appropriate for scoring functions derived from diverse protein data sets. J. Chem. Inf. Model. 2010;50(11):1961–1969.

101. Sheridan RP. Time-split cross-validation as a method for estimating the goodness of prospective prediction. J. Chem. Inf. Model. 2013;53(4):783–790.

102. Ortwine DF, Aliagas I. Physicochemical and DMPK in silico models: facilitating their use by medicinal chemists. Mol. Pharm. 2013;10 (4):1153–1161.

103. Ballester PJ. Selecting machine-learning scoring functions for structure-based virtual screening. Drug Discov Today Technol. 2019;32-33:81–87.

104. Durrant JD, Carlson KE, Martin TE, et al. Neural-network scoring functions identify structurally novel estrogen-receptor ligands. J. Chem. Inf. Model. 2015;55(9):1953–1961.

105. Alexander DLJ, Tropsha A, Winkler DA. Beware of R2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. J. Chem. Inf. Model. 2015;55(7):1316–1322.

106. Todeschini R, Ballabio D, Grisoni F. Beware of unreliable Q2! A comparative study of regression metrics for predictivity assessment of QSAR models. J. Chem. Inf. Model. 2016;56(10):1905–1913.

107. Brown JB. Classifiers and their metrics quantified. Mol. Inform. 2018;37(1–2):1700127.

108. Ballabio D, Grisoni F, Todeschini R. Multivariate comparison of classification performance measures. Chemometrics Intellig. Lab. Syst. 2018;174:33–44.

109. Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. Nat. Rev. Drug Disc. 2005;4(8):649–663.

110. Dobson CM. Chemical space and biology. Nature. 2004;432 (7019):824–828.

111. Lipinski C, Hopkins A. Navigating chemical space for biology and medicine. Nature. 2004;432(7019):855–861.

112. Ruddigkeit L, Van Deursen R, Blum LC, et al. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. J. Chem. Inf. Model. 2012;52(11):2864–2875.

113. Topliss JG. Utilization of operational schemes for analog synthesis in drug design. J. Med. Chem. 1972;15(10):1006–1011.

114. O'Boyle NM, Boström J, Sayle RA, et al. Using matched molecular series as a predictive tool to optimize biological activity. J. Med. Chem. 2014;57(6):2704–2713.

115. Griffen E, Leach AG, Robb GR, et al. Matched molecular pairs as a medicinal chemistry tool. J. Med. Chem. 2011;54(22):7739–7750.

116. Stewart KD, Shiroda M, James CA. Drug Guru: a computer software program for drug design using medicinal chemistry rules. Bioorg. Med. Chem. 2006;14(20):7011–7022.

117. Patel H, Bodkin MJ, Chen B, et al. Knowledge-based approach to De Novo design using reaction vectors. J. Chem. Inf. Model. 2009;49 (5):1163–1184.

118. Hoffmann T, Gastreich M. The next level in chemical space navigation: going far beyond enumerable compound libraries. Drug Discov Today. 2019;24(5):1148–1156.

119. Lessel U, Wellenzohn B, Lilienthal M, et al. Searching fragment spaces with feature trees. J. Chem. Inf. Model. 2009;49(2):270–279.

120. Humbeck L, Weigang S, Schäfer T, et al. CHI PMUNK: a virtual synthesizable small-molecule library for medicinal chemistry, exploitable for protein-protein interaction modulators. ChemMedChem. 2018;13(6):532–539.

121. Schneider G, Lee ML, Stahl M, et al. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. J. Comput. Aided Mol. Des. 2000;14(5):487–494.

122. Hartenfeller M, Zettl H, Walter M, et al. DOGS: reaction-driven de novo design of bioactive compounds. PLoS Comput. Biol. 2012;8(2): e1002380.

123. Button A, Merk D, Hiss JA, et al. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. Nat. Mach. Intell. 2019;1(7):307–315.

124. Salakhutdinov R. Learning deep generative models. Annu. Rev. Stat. Appl. 2015;2(1):361–385.

125. Gordeeva EV, Molchanova MS, Zefirov NS. General methodology and computer program for the exhaustive restoring of chemical structures by molecular connectivity indexes. Solution of the inverse problem in QSAR/QSPR. Tetrahedron Comput. Methodol. 1990;3(6):389–415.

126. Skvortsova MI, Stankevich IV, Zefirov NS. Generation of molecular structures of polycondensed benzenoid hydrocarbons using the randic index. J.Struct. Chem. 1992;33(3):416–422.

127. Skvortsova MI, Baskin II, Slovokhotova OL, et al. Inverse problem in QSAR/QSPR studies for the case of topological indexes characterizing molecular shape (Kier indices). J Chem Inf Model. 1993;33:630–634.

128. Miyao T, Kaneko H, Funatsu K. Ring-system-based exhaustive structure generation for inverse-QSPR/QSAR. Mol. Inform. 2014;33 (11–12):764–778.

129. Wong WW, Burkowski FJ. A constructive approach for discovering new drug leads: using a kernel methodology for the inverse-QSAR problem. J Cheminform. 2009;1(4). DOI:10.1186/1758-2946-1-4

130. Cho SJ, Zheng W, Tropsha A. Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches. J. Chem. Inf. Comput. Sci. 1998;38(2):259–268.

131. Miyao T, Kaneko H, Inverse FK. Inverse QSPR/QSAR analysis for chemical structure generation (from y to x). J. Chem. Inf. Model. 2016;56(2):286–299.

132. Vanhaelen Q, Lin Y-C, Zhavoronkov A. The advent of generative chemistry. ACS Med. Chem. Lett. 2020;11(8):1496–1505.

133. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Model. 1988;28(1):31–36.

134. Segler MHS, Kogej T, Tyrchan C, et al. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018;4(1):120–131.

135. Merk D, Friedrich L, Grisoni F, et al. De Novo design of bioactive small molecules by artificial intelligence. Mol. Inform. 2018;37 (1–2):1700153.

136. Olivecrona M, Blaschke T, Engkvist O, et al. Molecular de-novo design through deep reinforcement learning. J Cheminform. 2017;9(48). DOI:10.1186/s13321-017-0235-x

137. Blaschke T, Engkvist O, Bajorath J, et al. Memory-assisted reinforcement learning for diverse molecular de novo design. J Cheminform. 2020;12(68). DOI:10.1186/s13321-020-00473-0

138. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci. Adv. 2018;4:eaap 7885.

139. Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268–276.

140. Barigye SJ, García de la Vega, JM & Perez-Castillo, Y Generative Adversarial Networks (GANs) based synthetic sampling for predictive modeling. Mol. Inform. 2020;39(10):e2000086.

141. Maziarka Ł, Pocha A, Kaczmarczyk J, et al. Mol-CycleGAN: a generative model for molecular optimization. J. Cheminform. 2018;12:2.

142. Li Y, Zhang L, Liu Z. Multi-objective de novo drug design with conditional graph generative model. J Cheminform. 2018;10(33). 10.1186/s13321-018-0287-6

143. Khemchandani Y, O'Hagan S, Samanta S, et al. DeepGraphMolGen, a multi-objective, computational strategy for generating molecules with desirable properties: a graph convolution and reinforcement learning approach. J. Cheminform. 2020;12(1):1–17.

144. Skalic M, Jiménez J, Sabbadin D, et al. Shape-based generative modeling for de novo drug design. J. Chem. Inf. Model. 2019;59 (3):1205–1214.

145. Grisoni F, Huisman B, Button A, et al. Combining generative artificial intelligence and on-chip synthesis for de novo drug design. ChemRxiv. 2020. 10.26434/chemrxiv.13498587.v1.

146. Kotsias P-C, Arús-Pous J, Chen H, et al. Direct steering of de novo molecular generation using descriptor-conditional recurrent neural networks (cRNNs). Nat. Mach. Intell. 2020;2(5):254–265.

147. Méndez-Lucio O, Baillif B, Clevert D-A, et al., De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. Nat. Commun. 11(1): 10. 2020..
    (•) **This study showcases how to leverage additional biological information in the training of generative neural network models.**

148. Nicolaou CA, Brown N. Multi-objective optimization methods in drug design. Drug Discov. Today Technol. 2013;10(3):e427–35.

149. Cruz-Monteagudo M, Borges F, Cordeiro MNDS. Desirability-based multiobjective optimization for global QSAR studies: application to the design of novel NSAIDs with improved analgesic, antiinflammatory, and ulcerogenic profiles. J. Comput. Chem. 2008;29(14):2445–2459.

150. Cruz-Monteagudo M, Borges F, Cordeiro MNDS. Jointly handling potency and toxicity of antimicrobial peptidomimetics by simple rules from desirability theory and chemoinformatics. J. Chem. Inf. Model. 2011;51(12):3060–3077.

151. Sánchez-Rodríguez A, Pérez-Castillo Y, Schürer SC, et al. From flamingo dance to (desirable) drug discovery: a nature-inspired approach. Drug Discov Today. 2017;22(10):1489–1502.

152. Perez-Castillo Y, Sánchez-Rodríguez A, Tejera E, et al. A desirability-based multi objective approach for the virtual screening discovery of broad-spectrum anti-gastric cancer agents. PLoS One. 2018;13(2):e0192176.

153. Elton DC, Boukouvalas Z, Fuge MD, et al. Deep learning for molecular design—A review of the state of the art. Mol. Syst. Des. Eng. 2019;4(4):828–849.

154. Polykovskiy D, Zhebrak A, Sanchez-Lengeling B et al. Molecular Sets (MOSES): a benchmarking platform for molecular generation models. arXiv [cs.LG] (2018).

155. Brown N, Fiscato M, Segler MHS, et al. GuacaMol: benchmarking models for de novo molecular design. J. Chem. Inf. Model. 2019;59 (3):1096–1108.
    (•) **This paper details the development of a benchmarking platform for de novo design artificial intelligence methods**

156. Jensen JH. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. Chem. Sci. 2019;10(12):3567–3572.

157. Preuer K, Renz P, Unterthiner T, et al. Fréchet ChemNet distance: a metric for generative models for molecules in drug discovery. J. Chem. Inf. Model. 2018;58(9):1736–1741.

158. Spiegel JO, Durrant JD. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. J. Cheminf. 2020;12(1):25.

159. Skalic M, Sabbadin D, Sattarov B, et al. From target to drug: generative modeling for the multimodal structure-based ligand design. Mol. Pharm. 2019;16(10):4282–4291.

160. Xu M, Ran T, Chen H. De novo molecule design through molecular generative model conditioned by 3D information of protein binding sites. ChemRxiv. 2020. 10.26434/chemrxiv.13498332.v1

161. Jeon W, Kim D. Autonomous molecule generation using reinforcement learning and docking to develop potential novel inhibitors. Sci. Rep. 2020;10(1):22104.

162. Christ CD, Zentgraf M, Kriegl JM. Mining electronic laboratory notebooks: analysis, retrosynthesis, and reaction based enumeration. J. Chem. Inf. Model. 2012;52(7):1745–1756.

163. De Almeida AF, Moreira R, Rodrigues T. Synthetic organic chemistry driven by artificial intelligence. Nat. Rev. Chem. 2019;3 (10):589–604.
    (•) **A comprehensive review on the challenges and opportunities of machine learning approaches in chemical synthesis**

164. Ley SV. The engineering of chemical synthesis: humans and machines working in harmony. Angew. Chem. Int. Ed Engl. 2018;57(19):5182–5183.

165. Wipke WT, Ouchi GI, Krishnan S. Simulation and evaluation of chemical synthesis—SECS: an application of artificial intelligence techniques. Artif. Intell. 1978;11(1–2):173–193.

166. Friedrich L, Rodrigues T, Neuhaus CS, et al. From complex natural products to simple synthetic mimetics by computational de novo design. Angew. Chem. Int. Ed Engl. 2016;55(23):6789–6792.

167. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Ghahramani Z, Welling M, Cortes C, et al., editors. Adv. neural inf. proc. sys. Vol. 27. Montreal, Canada: Curran Associates, Inc.; 2014. p. 3104–3112.

168. Devlin J, Chang M-W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv [cs. CL] (2018).

169. Cadeddu A, Wylie EK, Jurczak J, et al. Organic chemistry as a language and the implications of chemical linguistics for structural and retrosynthetic analyses. Angew. Chem. Int. Ed Engl. 2014;53(31):8108–8112.

170. Liu B, Ramsundar B, Kawthekar P, et al. Retrosynthetic reaction prediction using neural sequence-to-sequence models. ACS Cent Sci. 2017;3(10):1103–1113.

171. Baylon JL, Cilfone NA, Gulcher JR, et al. Enhancing retrosynthetic reaction prediction with deep learning using multiscale reaction classification. J. Chem. Inf. Model. 2019;59(2):673–688.

172. Coley CW, Rogers L, Green WH, et al. Computer-assisted retrosynthesis based on molecular similarity. ACS Cent Sci. 2017;3(12):1237–1245.

173. Gelly S, Silver D. Monte-Carlo tree search and rapid action value estimation in computer Go. Artif. Intell. 2011;175(11):1856–1875.

174. Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018;555 (7698):604–610.

175. Schwaller P, Petraglia R, Zullo V, et al. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. Chem. Sci. 2020;11(12):3316–3325.

176. Raccuglia P, Elbert KC, Adler PDF, et al. Machine-learning-assisted materials discovery using failed experiments. Nature. 2016;533 (7601):73–76.

177. Coley CW, Barzilay R, Jaakkola TS, et al. Prediction of organic reaction outcomes using machine learning. ACS Cent Sci. 2017;3(5):434–443.

178. Coley CW The open reaction database. [cited 2020 Dec 15]. Available from: https://docs.open-reaction-database.org/.

179. Satoh H, Funatsu K. SOPHIA, a knowledge base-guided reaction prediction system - Utilization of a knowledge base derived from a reaction database. J. Chem. Inf. Comput. Sci. 1995;35(1):34–44.

180. Wei JN, Duvenaud D, Aspuru-Guzik A. Neural networks for the prediction of organic chemistry reactions. ACS Cent Sci. 2016;2 (10):725–732.

181. Jin W, Coley C, Barzilay R, et al. Predicting organic reaction outcomes with Weisfeiler-Lehman network. In: Guyon I, editors Adv. neural inf. proc. sys. Vol. 30, Long Beach, CA: Curran Associates, Inc.; 2017. p. 2607–2616.

182. Coley CW, Jin W, Rogers L, et al. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem. Sci. 2019;10(2):370–377.

183. Smith JS, Isayev O, Roitberg AE. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chem. Sci. 2017;8(4):3192–3203.

184. Schwaller P, Laino T, Gaudin T, et al. Molecular Transformer: a model for uncertainty-calibrated chemical reaction prediction. ACS Cent Sci. 2019;5(9):1572–1583.

185. Schwaller P, Gaudin T, Lányi D, et al., 'Found in Translation': predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. Chem. Sci. 9(28): 6091–6098. 2018..
(•) One of the pioneering studies that treated retrosynthesis as a language translation problem.

186. Bradshaw J, Kusner MJ, Paige B et al. J. M. A generative model for electron paths. arXiv [physics.chem-ph] (2018).

187. Do K, Tran T, Venkatesh S Graph transformation policy network for chemical reaction prediction. in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 750–760, Anchorage, AK. (Association for Computing Machinery, 2019).

188. Schneider P, Walters P, Plowright AT, et al. Rethinking drug design in the artificial intelligence era. Nat. Rev. Drug Discov. 2020;19 (5):353–364.

189. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. Drug Discov Today. 2020;26(2):511–524.

190. Riniker S, Landrum GA. Similarity maps - A visualization strategy for molecular fingerprints and machine-learning methods. J. Cheminform. 2013;5(1):43.

191. Numeroso D, Bacciu D Explaining deep graph networks with molecular counterfactuals. arXiv [q-bio.QM] (2020).

192. Raina R, Battle A, Lee H et al. Self-taught learning: transfer learning from unlabeled data. In Proceedings of the 24th International Conference on Machine Learning 759–766, Corvalis, OR (ACM, 2007).

193. Nguyen CQ, Kreatsoulas C, Branson KM Meta-learning GNN initializations for low-resource molecular property prediction. arXiv [cs.LG] (2020).

194. Bradbury J, Frostig R, Hawkins P et al. JAX: composable transformations of Python+NumPy programs. [cited 2020 Dec 15]. Available from: http://github.com/google/jax (2018).

195. Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 2019;32.