nature reviews drug discovery

https://doi.org/10.1038/s41573-023-00832-0

Perspective

Check for updates

Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR

Alexander Tropsha [1] ✉, Olexandr Isayev [2], Alexandre Varnek [3], Gisbert Schneider [4] & Artem Cherkasov [5,6] ✉

Abstract

Quantitative structure–activity relationship (QSAR) modelling, an approach that was introduced 60 years ago, is widely used in computer-aided drug design. In recent years, progress in artificial intelligence techniques, such as deep learning, the rapid growth of databases of molecules for virtual screening and dramatic improvements in computational power have supported the emergence of a new field of QSAR applications that we term 'deep QSAR'. Marking a decade from the pioneering applications of deep QSAR to tasks involved in small-molecule drug discovery, we herein describe key advances in the field, including deep generative and reinforcement learning approaches in molecular design, deep learning models for synthetic planning and the application of deep QSAR models in structure-based virtual screening. We also reflect on the emergence of quantum computing, which promises to further accelerate deep QSAR applications and the need for open-source and democratized resources to support computer-aided drug design.

Sections

[1]University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2]Carnegie Mellon University, Pittsburgh, PA, USA. [3]University of Strasbourg, Strasbourg, France. [4]ETH, Zurich, Switzerland. [5]University of British Columbia, Vancouver, BC, Canada. [6]Photonic Inc., Coquitlam, BC, Canada. ✉e-mail: alex_tropsha@unc.edu; artc@mail.ubc.ca

Introduction

The field of quantitative structure–activity relationship (QSAR) modelling — in which a quantitative description of a chemical structure is correlated with its biological activity or other types of chemical properties (Box 1) — can be traced back to a seminal publication by Hansch et al. in 1962 (ref. 1). Since then, the field has progressed with the substantial expansion of biological and chemical data and concurrent use of increasingly more complex machine learning algorithms for model development[2]. Additionally, QSAR modelling concepts have proliferated across multiple areas of data-rich research, including drug design, health care, materials science and education[3].

As the size and complexity of data sets in all areas of research have grown, deep learning has emerged as a type of machine learning that can recognize complex patterns in big data and rationalize them to produce accurate predictions. The origin of modern deep learning dates back to the mid-1960s, when the foundational book by Ivakhnenko and Lapa[4] was published. However, applications of deep learning in areas such as image recognition and natural language processing have progressed especially rapidly in the past decade, in parallel with large increases in computational power, methodological advances and the expansion of data. The initial application of deep learning in QSAR modelling was in the Merck Molecular Activity Challenge hosted by Kaggle in 2012. Entrants were asked to build QSAR models using numerical descriptors generated from the chemical structures of compounds in a training set with known biological activities, and models were assessed by their abilities to predict the biological activities of compounds outside the training set. This has been followed by multiple contributions in a burgeoning area of research we now call deep QSAR, with more than 200 publications since the first paper describing the use of deep learning in QSAR was published in 2015 (ref. 5).

In this Perspective, marking both the sixtieth anniversary of the overall QSAR field and the tenth anniversary of the introduction of deep learning approaches in QSAR modelling, we summarize ongoing developments that have catalysed and enabled the emergence of deep QSAR modelling and the application of these models in virtual screening of greatly expanded chemical space. These developments include advances in artificial intelligence, especially deep learning[6–8], the rapid growth of molecular databases (Box 2) and a new generation of molecular representations based on embeddings of both conventional linear notations, such as SMILES, and chemical graphs (Box 3).

After briefly overviewing the fundamentals of deep QSAR modelling, we focus on the latest advances that are particularly relevant to the discovery of small-molecule drug candidates. We discuss deep QSAR in generative molecular design and highlight the potential for integrating QSAR models used in generative design with deep learning models for synthesis planning and experimental chemistry into closed-circle, fully automated drug discovery platforms. We then describe approaches for virtual screening using deep QSAR models, which could help harness the recent growth in the size of explicitly enumerated chemical libraries available for virtual screening (Box 2). We predominantly highlight unique developments and applications in QSAR modelling that have been enabled or catalysed by the emergence and application of deep learning methods. Traditionally important areas of drug discovery and development, such as physical property prediction and modelling of absorption, distribution, metabolism, excretion and toxicity characteristics, can also benefit from deep QSAR modelling but are not reviewed here, as such a discussion would expand but not enrich the intended scope of the article. Similarly, dramatic improvements in hardware, such as graphics processing units (GPUs), are also supporting advances in deep QSAR but are not covered here as these issues were reviewed recently elsewhere[9]. Finally, we highlight emerging trends in the field, including the need and potential for open-source and democratization initiatives in computer-aided drug design (CADD) and the potential for quantum computing to dramatically accelerate the processing of ultra-large data sets for challenging tasks such as the use of machine learning-accelerated quantum mechanical calculations to improve ligand scoring accuracy in molecular docking.

Fundamentals of deep QSAR modelling

Conventional cheminformatics tasks such as QSAR modelling (Box 1) or chemical similarity searching rely on molecular descriptors designed to numerically characterize molecular structures at different levels of structure representation, from 1D to 3D, or even 4D (Fig. 1). However, the adaptation of deep learning to chemical data sets requires novel types of molecular representation, where descriptor engineering (which involves generation and selection of the most informative numerical molecular descriptors) is replaced by molecular embeddings (where molecules are represented by vectors in artificially created high-dimensional spaces that are employed in learning tasks using neural network architectures) (Fig. 1). Notably, common machine learning approaches use conventional chemical descriptors computed from chemical structures based on defined formulas (Box 1) prior to model development. Conversely, deep learning models can employ molecular embeddings created from standard chemical input data, such as molecular SMILES or chemical graphs, that can be modified as part of the learning process to achieve the most accurate prediction of the property of interest[10] (Fig. 1). Thus, these approaches generally learn feature vectors corresponding to a molecule or an atom through the training of a deep learning model on a specific task.

Thus, unlike traditional QSAR modelling (Box 1), chemical structure embedding (which could be viewed as a process analogous to chemical descriptor calculation in traditional cheminformatics) and learning using this representation are inseparable components of the model optimization process. Ultimately, practitioners have to decide which model architecture is best suited to the task at hand. Models can be designed heuristically based on human experience or semi-automatically with an evolutionary algorithm[11], neural architecture search or meta-learning[12]. A recent study suggested[13] pre-training a deep learning model with one million unlabelled molecules from ChEMBL to learn the representations, followed by model fine-tuning for various downstream quantitative structure–property relationship or QSAR tasks.

An important advantage of deep QSAR methods over traditional QSAR methods is that they may more effectively address multi-objective optimization tasks by using knowledge transfer, where, in practical terms, concurrent use of different data available for different tasks helps improve prediction accuracy for each task[14]. However, it has been shown that model improvement over respective single-task QSAR models is not always assured: better or worse performance can be obtained depending on the level of correlation between activities against individual targets[14]. Recent studies have suggested additional methodological advances to improve the accuracy of deep QSAR application to multi-objective learning[15].

Traditional areas of concern for QSAR modelling, such as data curation, model applicability domains and independent model validation, remain key areas for consideration when building deep QSAR models. Methods for chemical and biological data curation that combine automatic and manual efforts have been extensively described[16–18],

Box 1

An introduction to QSAR modelling in drug discovery

QSAR modelling was originally introduced as a computational tool to relate molecular properties (such as dipole moment or hydrophobicity) to bioactivities measured for small congeneric series of compounds. It has since evolved in terms of the complexity of both molecular characterization and statistical or machine learning approaches used to build models as well as in the size and chemical diversity of the data sets used for model development (reviewed in ref. 25).

QSAR modelling has progressed to become one of the major tools in computer-assisted drug discovery, and key principles and best practices of QSAR modelling have also found multiple applications in other areas of chemistry, materials science and beyond (see ref. 3 for a recent review). In addition to the prediction of bioactivity of compounds, QSAR methods have been employed widely in the prediction of multiple physicochemical properties, such as aqueous solubility, melting point and pK_a, that are routinely considered by medicinal chemists. QSAR approaches have also been employed in the calculation of biological properties of compounds other than target bioactivity such as various absorption, distribution, metabolism, excretion and toxicity end points of critical importance in drug development.

As we emphasize in this Perspective, the accumulation of big data on chemical bioactivity for large numbers of molecular targets has created a need to accelerate different types of chemical computing used in both ligand-based and protein structure-based drug discovery. These challenges have led to the proliferation of QSAR approaches (recently enhanced by deep learning methods) to areas that historically have not employed statistical modelling methods such as quantum chemistry and molecular docking.

Modern QSAR approaches can be generally described as an application of statistical and machine learning techniques to finding empirical relationships of the form $A_i = \hat{k}(D_1, D_2, \ldots, D_n)$, where A_i are biological activities (or other properties of interest) of molecules, $D_1, D_2, \ldots, D_n$ are calculated (or, sometimes, experimentally measured) structural properties (known as molecular descriptors) of compounds, and $\hat{k}$ is some empirically established mathematical transformation that should be applied to the descriptors to calculate the property values for all molecules for which the relationship holds.

Depending on how molecular structure is characterized, QSAR models have been classified as 1D, 2D, 3D or even higher levels (Fig. 1). Examples of 1D descriptors include molecular weight, counts of atom types, counts of hydrogen bond donors or acceptors, number of rings, or number of specific functional groups. Descriptors in the 2D category include multiple molecular indices calculated from molecular graphs. 3D descriptors are calculated from the knowledge of molecular geometry (for instance, polar surface area or 3D atom pairs) and 4D descriptors are calculated for multiple molecular conformations arising from conformational search or molecular dynamics simulations.

With the development of multiple types of molecular descriptors, increasingly more complex machine learning and descriptor sampling approaches (*k*-nearest neighbour, support vector machines, random forest, artificial neural networks and others) have been introduced with prominent use in QSAR modelling. As the complexity of both data and methods has increased, QSAR modelling has gradually transitioned from a simple statistical tool related to concepts in physical organic chemistry to an application of multivariate statistical modelling of data in chemical and materials sciences. Thus, it is not surprising that deep learning methods are increasingly being applied in QSAR modelling.

As we discuss methods and applications of deep QSAR, it is important to recognize that the best practices of model development and validation established in the field and summarized in several key publications[2,18,20] remain valid. These key principles include the importance of chemical and biological data curation, which can be accomplished using respective workflows[16,17], as well as rigorous model validation protocols[20].

but the large size of the data sets to which deep QSAR methods are or could be applied means there is now a need for methods that can effectively address data curation at scale as manual curation is not feasible. Reflecting on this need, automated workflows are being developed to process both training and external data sets, such as those available in KNIME, to ensure the trustworthiness of models[19].

Similarly, deep QSAR models require rigorous external validation. This critical aspect of QSAR model development has been extensively discussed in the literature, and rigorous model validation workflows have been developed[3,20]. These workflows can be employed when building deep QSAR models for big data sets, but their execution requires significant computational resources. Furthermore, with the substantial expansion of external virtual screening sets used to identify novel active compounds, assessing the applicability domain of deep QSAR models (that is, whether external data are within or out of the distribution of the training data set) also becomes more challenging.

Recent studies have expanded the original concept of modelability[21], which was introduced to explain the low accuracy of QSAR models for data sets with a large fraction of "activity cliffs"[22] (pairs of compounds with the highest mutual similarity but different activity classes). This concept was recently expanded by introducing the roughness of molecular property landscapes[23], and the argument was made that estimating the roughness could improve the extrapolative accuracy of QSAR models. Recent studies have also emphasized the importance of addressing prediction confidence, in addition to traditional objectives of predicting the activity class or value for external compounds. For instance, Bosc et al.[24] advocated for the use of conformal prediction methods to provide information on the certainty of predictions. As both chemical bioactivity data sets and virtual screening libraries continue to grow, rigorous assessment of chemical distribution and prediction certainty should be a requisite part of any deep QSAR modelling study.

Deep QSAR and generative modelling

Generative molecular design

One traditional use of QSAR models is the virtual screening of chemical databases to identify molecules of interest that can be purchased and

Box 2

Expansion of searchable chemical space

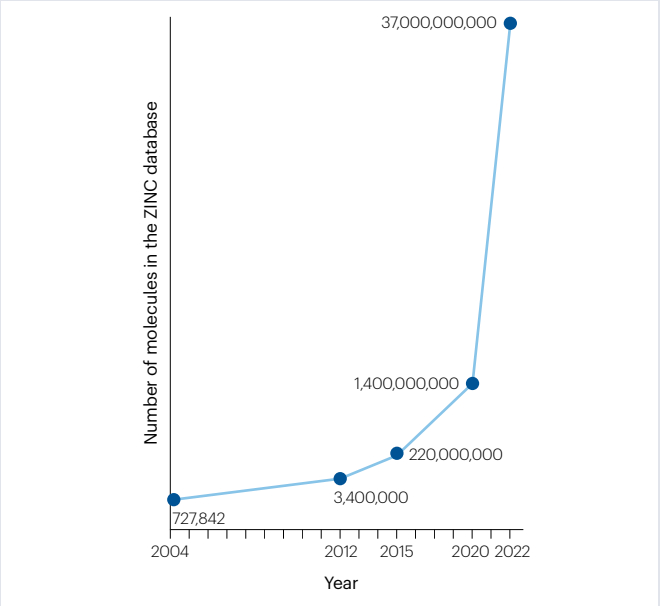
Advances in click chemistry, robotics, synthesis automation and computational planning have led to exponential growth in the sizes of chemical databases in recent years.

One of the most popular databases for sourcing molecules for virtual screening is ZINC. In 2006, this database contained less than a million molecules, but the most recent ZINC22 version has grown to over 37 billion unique chemical entries — a >50,000-fold increase, with the most dramatic growth in the last 2 years (see figure)[152].

Commercial libraries such as WuXi AppTec or CHEMriya by OTAVA have also grown rapidly. The commonly used Enamine REAL database currently offers up to 5.5 billion unique compounds for order, while the Enamine REAL Space on-demand library can be expanded into ~38 billion entities (when isomers are considered). Similarly, the SAVI database encompasses 1.75 billion chemicals that have been generated by applying 53 chemical transformations to 150,000 Enamine building blocks[64]. There is also a growing number of knowledge-based collections, including BioSolveIT KnowledgeSpace, which describes the chemical space of hundreds of billions of structures with a plausible synthetic feasibility.

Recently, a comprehensive collection of 4.2 billion molecules has been aggregated from 23 different sources in an effort to find direct antivirals against SARS-CoV-2 (ref. 153). Many (if not all) pharmaceutical companies have created their own proprietary ultra-large chemical databases, for example, PLC by Eli Lilly, PGVL by Pfizer and XXL collections by GlaxoSmithKline[154].

Further expansion of accessible chemical space could be already projected into trillions of entities. Ruddigkeit et al. generated 166.4 billion virtual molecules containing up to 17 atoms restricted to basic chemical elements (C, N, O, S and halogens) that could be made by following simple synthetic rules[155]. Other virtual databases such as eXplore from eMolecules and MolDB from DeepCure already include trillions of hypothetical entries, with the latter collection already aiming at expanding to 10^{18} synthesizable compounds constructed by generative models.



tested (reviewed in ref. 25). However, progress in medicinal chemistry requires the discovery of new chemical entities. Generating new molecules from scratch (de novo) involves construction, scoring and optimization[26] of molecular structures and advances in these tasks have recently been achieved using deep QSAR coupled with generative chemistry[27].

Methods employed for de novo molecular design include rule-based and rule-free approaches[28,29], both of which have been shown to identify new bioactive compounds that are also synthetically accessible (see ref. 30 for a review). Rule-based methods use sets of molecular building blocks and chemical transformations such as virtual reaction schemes for structure generation[31]. In contrast, rule-free 'generative' (or 'constructive') deep learning methods[32] sample new molecules from a learned statistical distribution of the training data ('latent space'), without explicitly representing their molecular structure in chemical terms, and this molecular design process is difficult (if not impossible) to describe in a way that can be easily interpreted.

Many generative drug design approaches have been built on deep neural networks (Fig. 2a–d). The most prominent methods are chemical language models[33] that employ textual representation of molecules by SMILES strings to learn the intrinsic grammar of the strings and generate new strings corresponding to novel realistic molecules. The majority of the chemical language models reported in the literature have employed recurrent neural networks with long short-term memory[34], variational autoencoders[35] and generative adversarial networks[36], and graph neural networks[37] have also been used to learn and generate molecular graphs. Several other deep learning architectures have been proposed for this purpose, including hybrid approaches combining rule-based and rule-free networks[38]. A recent review provides a good summary of neural network architectures used in generative chemical language models[39].

All of the methods considered above typically sample new molecules from a latent representation of molecular structure learned by the neural network during model training; that is, they act as statistical structure generators. At some point during or after the molecule construction process, the proposed designs are evaluated and prioritized according to the desired function; that is, their biological activity and/or other properties. Notably, it has also recently been demonstrated that the inverse strategy — generating molecules with desired properties that are decoded from a particular area of the latent descriptor space — can be successfully pursued using a type of deep learning method called a conditional recurrent neural network[40].

Virtual assessment of the target property of the generated molecules is the most critical and error-prone part of the design process. Scoring of new molecules can be performed in several ways (Fig. 2e,f).

These include using an external QSAR model or enriching the training data with reference molecules that have the desired activity[41], iteratively adapting the parameters of the generative model to preferentially construct molecules with the desired properties (for example, by one-shot or few-shot learning[42], transfer learning[43], or reinforcement learning[27]), or directly by using the probabilities learned by the generative model as the evaluation criterion[44]. Depending on the nature of the machine learning algorithm employed for adaptive model optimization, this criterion is sometimes referred to as 'reward' or 'fitness'.

Generative drug design with deep QSAR models

Chemical language models combined with external scoring seem to currently dominate the field of generative chemistry, possibly owing to the availability of the respective software tools[45–47]. Thus, deep QSAR models that are built to assess molecular bioactivity have recently been combined with chemical language models, either as separate external tools to rank the generated molecules by their activity or as model-intrinsic scoring functions to guide chemical structure generation towards molecules with the desired properties[39,41,47]. Studies that also include experimental validation of compounds proposed by generative molecular design methods are still scarce but are beginning to emerge[48]. For instance, compound **1** (Fig. 2f) is a new receptor-related orphan receptor gamma (RORγ) inverse agonist (IC_{50} 370 nM), which emerged as a top-scoring design based on a sampling probability estimate[44]. This beam-search approach eliminates the strict need for an explicit QSAR model by relying on the learned statistical distribution of the training data, which included known RORγ modulators. Additional examples of experimentally bioactive compounds designed de novo using generative chemical language approaches have been discussed in a recent review[39]; these include a dual modulator of the retinoid X and peroxisome proliferator-activated receptors and an inhibitor of Moloney murine leukaemia virus kinase 1 and cyclin-dependent kinase 4.

As discussed above, similar to conventional QSAR models, deep QSAR models suffer from a decrease in performance and accuracy when applied to poorly curated data, when they lack appropriate validation or when applied to out-of-domain data. To increase confidence in bioactivity predictions, model ensembles can be used, combining the predictions with a majority voting approach[49]. In a recent study[50], an ensemble of 100 deep QSAR models was trained with the ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately)[51] algorithm. Each model predicted phosphoinositide 3-kinase-γ (PI3Kγ) inhibition with a slightly different performance. Testing of a subset of the computer-generated compounds showed that the difference in votes (0 to 100) for each de novo-generated molecule was approximately, though not perfectly, reflected in their dissociation constants. The top-scoring compound, compound **2** (Fig. 2f), potently inhibited the intended target PI3Kγ (K_i 63 nM).

Computational models can suggest potential new drug candidates to medicinal chemists within seconds to hours; therefore, synthesizing chemicals for subsequent experimental evaluation remains a time-limiting step at this point in the process. The rapid development of robotic platforms for drug and materials design has stimulated the creation of efficient cheminformatics tools for planning and guiding organic synthesis. These tools aim to assess the synthetic accessibility of a compound and suggest feasible synthetic routes between available starting materials and the target molecule, and are discussed in the following section.

Deep learning in synthesis planning

Two strategies are conventionally used to determine a series of reaction steps that lead to a given compound from available starting materials: forward synthesis (starting from a collection of building blocks), and backward synthesis or retrosynthesis (starting from the target molecule and searching for putative precursors and the respective reactions[52]). Generally, a synthetic route contains several one-step reactions for which major products, yield and, ideally, the reaction rate should be assessed. For a given elementary step, reaction conditions (solvent, catalyst, temperature and so on) leading to a reasonable yield should be suggested.

Computer-assisted synthesis planning has long been an important area of research, starting from seminal contributions by Corey[53]. Machine learning methods have been progressively brought into this field, as extensively discussed in a recent review[54]. As in many other disciplines, the growth of information on chemical reactions and the development of large reaction databases stimulated the emergent use of

Box 3

Molecular cartography

The rapid growth of modern chemical databases requires special tools to analyse and visualize this information. In this context, molecular cartography or chemography can be particularly useful. Chemography employs multidimensional scaling methods to project compounds in chemical libraries represented by multiple molecular descriptors (chemical space) onto a 2D map[156]. Chemical space visualization has been an important component of cheminformatics research for many years[157]; however, deep learning approaches brought innovation and acceleration into this area of cheminformatics to enable the visualization of modern ultra-large chemical libraries.

One such recent method — generative topographic mapping (GTM) — allows both compound positions on a map and the data probability distribution function to be obtained, which, in turn, makes it possible to describe (ultra)large compound collections. GTM can be used efficiently for various cheminformatics tasks[158–160], including chemical data visualization and analysis, prediction of properties or biological activities, comparison of large chemical libraries, drug repurposing, and virtual screening. Related options are implemented in the polyfunctional ChemSpace Atlas tool[161], which includes >40,000 hierarchically related maps accommodating several billion compounds.

In combination with SMILES-based or graph-based autoencoders, chemography can be used for automatized generation of chemical structures with desired biological activities, which is an emerging area of cheminformatics. It has been demonstrated that GTM constructed on the autoencoder latent variables provides direct and intuitive access to the autoencoder chemical space. Sampling of this space can be 'driven' by the map towards the highly relevant zones of a drug discovery project[162]. This approach can also be extended to chemical reactions where, for instance, a cartography-enhanced artificial intelligence tool enabled the discovery of several new types of Suzuki coupling reaction[163].

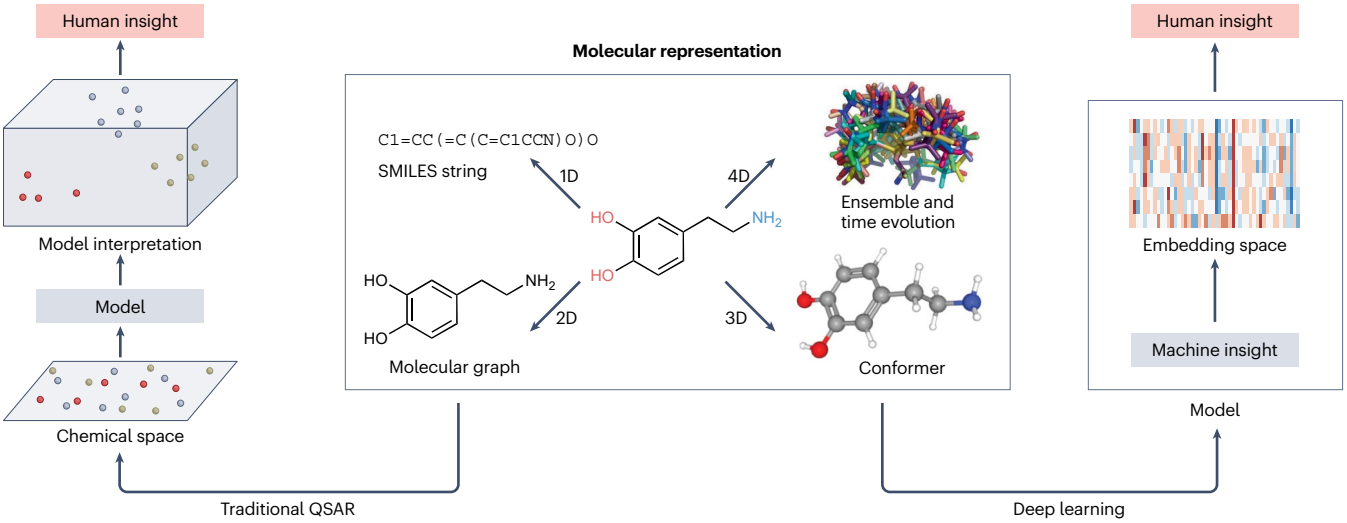


Fig. 1 | Contrasting traditional and deep QSAR models. Both approaches employ similar input data representing molecules in 1D to 4D. However, traditional QSAR methods (shown on the left) require calculation of explicit numerical descriptors from the respective molecular representations that are used for various machine learning tasks (that is, descriptor calculations and machine learning are separated), whereas deep QSAR approaches (shown on the right) learn molecular representations as part of model optimization conducted in the latent chemical space.

deep learning methods in computer-assisted chemical synthesis planning. In their seminal studies, Segler et al.[55,56] reported deep neural network models that assess the probabilities of different transformations, which, in combination with symbolic artificial intelligence methods and the Monte Carlo Tree Search, led to the selection of the most feasible reaction pathways. Two alternative methods — template-based methods and template-free methods[54] — were applied both to retrosynthesis and to outcome prediction.

In template-based methods for retrosynthesis, the templates (transformation rules) can either be suggested by expert chemists or extracted automatically from reaction databases. Retrosynthetic transformation rules provided by expert chemists were used in LHASA, the first expert system for organic synthesis planning, which was developed around 50 years ago by Corey and Wipke[52]. This concept has now been achieved in several retrosynthetic tools, including the popular CHEMATICA (also known as SYNTHIA) programme[57], which integrates more than 50,000 mechanism-based transformation rules deduced manually from the experience of organic chemists. The main trend is currently the automatic extraction of transformation rules (patterns) from chemical reaction databases. Such an approach was initially used by Segler et al.[55], and it was also implemented in the popular AiZynthFinder tool[58].

Alternatively, in template-free methods for retrosynthetic route planning, relationships between structures of reactants and products of chemical reactions are deduced directly. Thus, Jin et al.[59] used a graph-convolutional neural network with a global attention mechanism to predict pairs of atoms belonging to the reaction centre. Another template-free method, sequence-to-sequence learning, is based on natural language processing and is widely used in text translation[60]. In the case of chemical reactions, SMILES strings of the reactants and the products are analogous to the initial text and its translation, respectively. This methodology has been used for the prediction of retrosynthetic reaction routes[61] and of reaction products[62].

In forward synthesis modelling, new molecules are grown in a stepwise manner, and all possible solutions are enumerated at each step, followed by the submission of top-scoring intermediates into a subsequent growing step. The quality of designed products is assessed by the similarity to a target molecule. This way, the forward synthesis approach can be utilized as a de novo design tool coupled with synthesis prediction. One of the most productive forward synthesis tools was developed in the Schneider group and implemented in the DOGS programme[31]. Forward synthesis is also utilized to explore chemical reaction networks originating from given building blocks[63] and to generate virtual compound libraries[64,65].

A synthetic accessibility score (or the opposite, a synthetic complexity score) represents another important scoring metric to aid the prioritization of virtual compounds. It can be used as a quantitative filter for the screening of virtual libraries or de novo design. Among different matrices of this kind[66,67], the most popular is SAscore[68], which considers both a fragment score based on fragment occurrence in PubChem compounds, and a complexity penalty calculated as a function of the number of rings, chiral centres, macrocyclic fragments and the total number of atoms. Recently, Coley et al.[69] proposed SCScore, a synthetic complexity score that always rates products higher than reactants.

In addition to assessing synthetic accessibility, it is important to assess reaction kinetic and thermodynamic characteristics as part of forward synthesis planning. Such predictive models can be built using condensed graphs of reaction (CGR)[70], which combine the information about all reactants and products. Fragment descriptors generated for CGR are combined with solvent and temperature descriptors and then used in the modelling of reaction rate constants for bimolecular nucleophilic substitution[67,71], bimolecular elimination[72] and different types of cycloaddition[73] as well as for equilibrium constants of tautomerization reactions[74]. Recently, some studies on reaction yield prediction were also published, but the results appeared somewhat controversial as the model quality is generally good for the sets of data collected using high-throughput techniques[75,76] but quite poor for diverse reaction data gathered from the literature sources[77,78].

Modelling of reaction conditions has also been attempted in recent studies. CGR-derived fragment descriptors were used by Marcou et al.[79] to obtain classification models able to predict optimal types of solvents and catalysts for the Michael reaction. In another study, Gao et al.[80] described neural network-based models that predict the catalyst, solvent, reagent and temperature for a particular reaction, and reported a close match (69.6%) to experimental conditions within the top-10 predictions. The likelihood ranking approach developed by Afonina et al.[81] outputs several alternative reaction conditions ranked with the help of a neural network-based QSAR model. Lin et al.[82] followed the heuristic principle: similar reactions proceed under similar conditions, allowing the problem of reaction condition assessment to be reduced to a simple similarity search in reaction databases with recorded conditions, which is especially effective with the aforementioned CGR technology[70].

Potential for automated compound design

Developments with machine learning models for molecule construction and synthesis planning have opened an intriguing opportunity for fully automated molecular design[7,83] in which fully robotic platforms could make decisions concerning both the structure of the molecule to be synthesized and the related synthesis plan without human intervention[84].

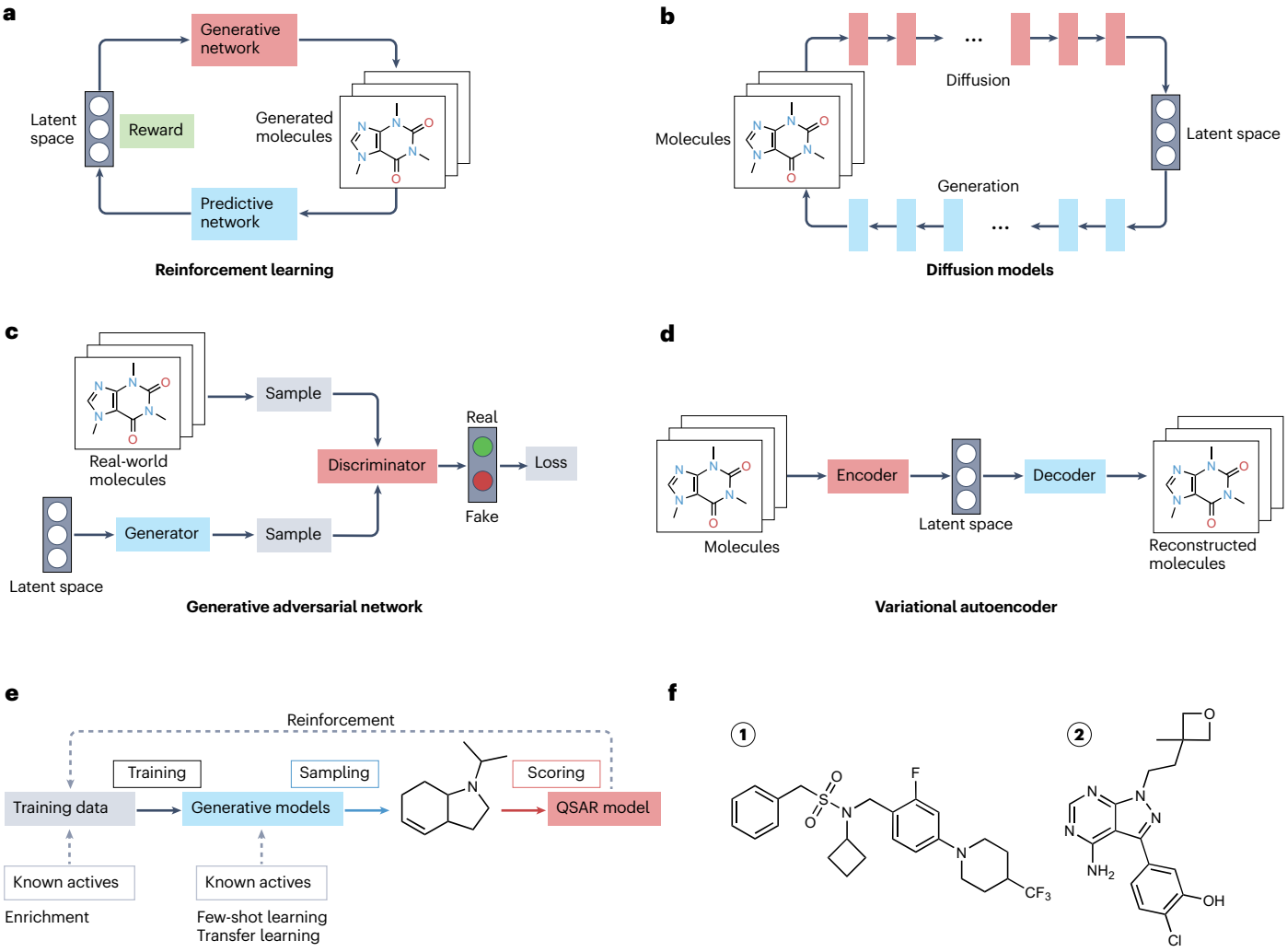


Fig. 2 | Generative molecular design. Commonly used architectures of generative models for chemical design using reinforcement learning (part **a**), diffusion models (part **b**), generative adversarial networks (part **c**) and a variational autoencoder (part **d**). These architectures differ in the specific protocols used to train models using molecular embeddings in the latent space (represented as grey rectangles with white circles) and generate novel compounds. See ref. 32 for a detailed discussion of these architectures. A generative model (part **e**, blue box) is trained with sets of molecules with desired properties, approximating the statistical distribution of the training data (grey box). New molecules are generated by sampling from the learned distribution. The biological activity (and/or other properties) can be predicted with a QSAR model (external scoring, red box) or by directly using the model-intrinsic sampling probability of a new molecule as a quality criterion (internal scoring). In situations with limited training data, few-shot learning or transfer learning can be used to bias a generic generative model towards the molecules of interest. Model reinforcement establishes a feedback loop from the QSAR model to the learning instance, aiming to bias the generative model towards constructing top-scoring new molecules. Examples of molecules designed by generative models with deep QSAR (part **f**). Compound **1** acts as an inverse agonist of receptor-related orphan receptor-γ (RORγ) and compound **2** inhibits phosphoinositide 3-kinase-γ (PI3Kγ).

Importantly, such approaches could be naturally integrated with deep QSAR models, directing the synthesis of compounds with desired biological and physicochemical properties. The unique aspect of model integration within the design, make, test, analyse cycle[83] is the need to process data generated by automated chemistry on the fly and translate model predictions into new synthetic instructions. The obvious necessity to optimize multiple objectives of automated design and synthesis of compounds with the desired properties, purity, yield, and bioactivities along with the large volumes of data generated by such systems suggest the growing use of deep QSAR models that particularly excel for large data sets and multi-task optimization. Emerging prototypes of such fully automatized systems have been reported[75,85,86] (summarized in a recent review[84]). We expect that rapid improvements in the methodology and software for artificial intelligence-driven chemical synthesis[87] will open a new era of artificial intelligence-driven synthetic chemistry assisted by deep QSAR.

Deep QSAR in structure-based screening

Virtual screening of large molecular libraries (often in excess of 1,000,000 compounds) is commonly used to identify potential ligands when suitable structural information on the protein targets is available. Typically, these computational approaches involve two steps: docking the molecules into the binding site to achieve a realistic 'pose' and scoring these poses to rank virtual screening hits and to support decision-making about which compounds to test in experiments. Thus, the common objectives of molecular docking approaches include the prediction of both the pose of the ligands and their binding scores, which are expected to correlate with experimentally measured binding affinity.

In traditional docking approaches, docking scores can only be calculated following the prediction of the ligand poses; therefore, docking and scoring of ultra-large chemical libraries comprising billions of molecules remain very challenging even with current computational power. The challenge of rapidly calculating docking scores for ultra-large compound libraries has been addressed recently with the emergence of an approach termed 'deep docking'[88]. As discussed in the next section, deep docking emerged at the interface between traditional approaches to score calculation and deep QSAR modelling.

Progressive docking and early adaptations of active learning

In 2006, the use of QSAR modelling and active learning to predict docking scores from computationally inexpensive chemical descriptors for small-molecule compounds in a virtual screening library was pioneered in an approach known as progressive docking[89]. This approach utilized scores generated for a fraction of small molecules selected from a virtual screening library by docking these compounds against a particular protein target of interest. Docking scores obtained for this small set were used as a target property (in place of the traditional use of bioactivity measurements) to build a QSAR model that was then used to estimate docking outcomes for as-yet-unprocessed ligands and to iteratively remove entries predicted to have unfavourable scores from the docking database to save computational cost. Progressive docking was tested on ~90,000 molecules screened by the docking programme Glide against a number of protein structures available at the time, including the sex hormone-binding globulin, carbonic anhydrase, corticosteroid-binding globulin, HIV reverse transcriptase and, notably, SARS-1 Main protease (M^{pro}). For each of these targets, protein-independent 'inductive' ligand descriptors[90] were used to approximate continuous Glide SP values for as-yet-undocked compounds, while using only 10% or 20% of the database for training. The application of the resulting linear QSAR models approximating Glide SP scores enabled up to 2.6-fold acceleration of the virtual screen while maintaining an up to 99% hit recovery rate[89].

Published a few years later, the NNscore method introduced non-linear QSAR modelling of docking scores[91]. The authors constructed a shallow artificial neural network consisting of 194 input nodes corresponding to the molecular descriptors of ligands (such as pairwise atom binding, various energy terms and the number of rotatable bonds) connected to a five-unit hidden layer, followed by a layer with 'good or poor binder' classification node. In a later study, the possibility of accurate emulation of docking scores with a conformational predictor was proposed by Svensson et al.[92]. Another methodology was employed in the iterative approach implemented for Apache Spark[93], in which the authors proposed a strategy very similar to progressive docking, wherein they iteratively docked sets of ligands from a large virtual screening library to form progressively changing training sets for model building and used the models to predict docking scores for the remaining ligands from the original library to progressively exclude low-scoring molecules and prioritize the high-scoring ones.

Accelerated docking score prediction with deep docking

While the early studies described above demonstrated an intriguing opportunity for the prediction of computationally intensive docking scores by computing them using simplistic ligand descriptors, the reported twofold to fourfold acceleration in library screening is still insufficient to deal with molecular libraries of a billion compounds or more. Only with the emergence of deep learning methods have such approaches become feasible. A deep docking approach making use of deep QSAR modelling for accelerated screening of ultra-large chemical libraries was developed in early 2020 (ref. 88) (Fig. 3).

The combination of active deep learning with simple, protein-independent 2D fingerprint chemical descriptors makes deep docking particularly suited for virtual screening of emerging giga-sized chemical libraries using standard computational resources. For example, deep docking enabled the evaluation of 1.4 billion compounds from the ZINC15 database against SARS-CoV-2 M^{pro} using only 60 CPU cores employed for Glide docking, and 4 GPU cores for deep neural network training[88]. The top 1,000 hits from the ZINC15 database (corresponding to 585 distinct chemical scaffolds) were disclosed, many of which have been subsequently independently confirmed as active compounds[94]. Importantly, the rapidity of the process allowed initial results to be published on the day the COVID-19 pandemic was declared by the World Health Organization[88].

This original publication[88] was followed by a series of similar deep docking campaigns with ultra-large chemical libraries that employed different docking software and various linear, non-linear, and machine learning approaches to build QSAR-like models to predict docking scores but generally replicated the original deep docking workflow. These recent studies are summarized in Table 1. The most recent examples of experimentally confirmed hits emerging from deep docking approaches include micromolar inhibitors of SARS-CoV-2 M^{pro} (ref. 95) and papain-like protease[96] as candidate antiviral agents and Lin28 inhibitors[97] with potential application as anticancer agents.

Large-scale and consensus deep docking

The emergence of active deep learning approaches, such as deep docking, has significantly reduced the computational cost of structure-based virtual screening campaigns, making it possible to routinely evaluate libraries composed of >1 billion compounds,

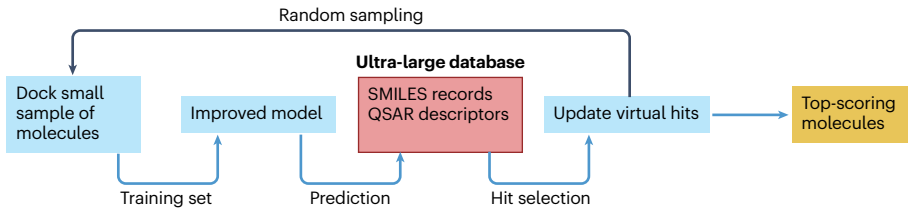


Fig. 3 | Workflow for deep docking. Deep docking uses a QSAR model trained on docking scores of a portion of an ultra-large docking database of molecules to iteratively predict the docking outcomes for the remaining entries. The workflow consists of the following major steps. The first step is model training, in which a small fraction of a database is randomly sampled and the sampled molecules are docked into the target of interest using any conventional protocol. A deep neural network model is then trained, correlating Morgan fingerprint descriptors of compounds in the sample set to their corresponding docking scores translated into a binary form based on a threshold (such as the top 1%). The next step is prediction; after the model is trained, the remaining entries in the ultra-large database get scored by the QSAR model and all 'active' predictions are then considered as emulated virtual hits to be added to the training set for the next iteration. Subsequently, in a database reduction step, all 'inactive' QSAR model predictions are removed from the docking base and the workflow is repeated with the 'active or inactive' threshold becoming more stringent after each iteration. These iterations are repeated until convergence; that is, when the number of retried virtual hits does not change significantly. The emulated virtual hits, scoring favourably within all iterations, are actually docked into the target site.

such as ZINC or Enamine REAL, in their entirety, using standard computational setups that are affordable to many CADD practitioners. For instance, in a recent proof-of-principle study, the authors were able to screen the entire ZINC15 database (1.4 billion molecules) against 12 drug targets, effectively reducing the amount of computation by 100-fold or more while recovering high-scoring molecules confirmed by conventional docking[98].

An intriguing emerging opportunity is the use of multiple docking programmes to build deep docking models, which enables CADD scientists to rely on and, hence, incorporate various consensus protocols into hit selection[99]. The use of consensus docking has a long-standing history[100,101] and is an integral part of best practices in CADD[102]. Furthermore, the application of rigorous consensus docking protocols to ultra-large chemical libraries could enable the development of fully automated hit-calling approaches that would not require any expert involvement. For example, a high-throughput virtual screening pipeline was described recently that utilized deep docking with the Glide, ICM, FRED, GPU-AutoDock and QuickVina2 programmes, which were all used to screen 40 billion molecules (combining ZINC15 and Enamine REAL Space databases) against SARS-CoV-2 M^{pro} (ref. 95). The consecutive deep docking runs with the five programmes took approximately 90 days of computing on 250 GPUs and 640 CPU cores and reduced the initial 40-billion-molecule library by 1,500-fold. Importantly, the use of 5 docking programmes, combined with ligand structural clustering and pharmacophore filtering, allowed the authors to develop and evaluate 28 individual consensus strategies, 26 of which were fully automated and 2 of which involved final examination of generated hit lists by a CADD expert. When up to 100 compounds were ordered and evaluated for each of the 26 automated hit-selection protocols, some reached hit rates of 13%, a respectable performance. Although the two strategies involving expert examination clearly outperformed the rest by achieving hit rates of 16% and 23%, these results suggest that deep docking strategies can already enable fully automated virtual screening of ultra-large chemical libraries.

While deep docking offers an exciting opportunity to accelerate structure-based drug discovery approaches and the use of consensus deep docking improves the results, there are several pitfalls associated with this approach that we hope will be addressed in the near future. Most importantly, deep docking is designed to accelerate the process of finding compounds with high docking scores, but it fully relies on the current scoring functions used in the docking software, and the resulting hit selection is predicated on the choice of software and respective scoring function. Thus, deep docking shares all the limitations and pitfalls concerning the accuracy of the current docking methods. As these methods improve in their accuracy, so will deep docking. Additional improvements may be associated with a more intelligent choice of the initial library to generate docking scores (for instance, by using structure-based pharmacophores to select compounds for initial docking calculations) as well as with the progress in deep learning methods.

Polypharmacology modelling with deep learning

We have already discussed above the use of deep learning in multi-objective optimization tasks using ligand bioactivity data[14,15]. This active area of research has recently been expanded with methods for multi-objective modelling of docking scores in the context of polypharmacology. For instance, Liu et al.[103] proposed joint modelling of docking data of new and known targets using multi-task learning and demonstrated that such an approach outperforms both single-task learning and active learning.

There are emerging examples of the use of deep learning to predict protein–ligand interactions[104,105], providing opportunities for large-scale target profiling for chemicals of interest. Kinome-wide multi-task deep neural network models were developed by Li et al.[106] using ~140,000 data points obtained by probing a panel of 391 kinases with small molecules. The resulting model allowed the authors to map a comprehensive kinome interaction network encompassing a variety of on-target and off-target effects. This tool was further developed into an online platform, KinomeX[107], that can predict kinome-wide polypharmacology of compounds based solely on their 2D structures enumerated through ECFP4 fingerprints.

While experimental data are limited by nature, the use of emulated docking scores could, in principle, be constrained only by the availability of computational resources. Moreover, the generation of docking scores for training deep learning models would not require any prior knowledge of a previously unstudied 'cold' target. For instance, the authors of a recent study[108] implemented a reinforcement learning approach that uses docking scores for compounds created by a generative model for a 'cold' target, while starting the training process with known ligands of homologous proteins. The authors did not augment

Table 1 | Virtual screening approaches using machine learning to predict docking scores

Method	Emulated docking score	Ligand descriptors	QSAR function	Ref.
Deep docking	Glide SP Quick Vina2 FRED GPU-AutoDock ICM	Morgan fingerprints	Deep neural network	95
Pyzer-Knapp approach	AutoDock-Vina	Extended-connectivity fingerprints	Bayesian optimization	138
Jastrzębski et al. approach	Glide XP SMINA	Contact fingerprints	Deep neural network	139
MolPal	AutoDock-Vina	Morgan fingerprints	Neural network; random forest; message passing neural network	140
Martin approach	DOCK	Morgan fingerprints	Linear regression	141
Lean docking	GOLD AutoDock-Vina FRED Glide SP MOE	Unfolded counted atom pairs fingerprints	Regressor model	142
HASTEN	Glide SP FRED	Morgan fingerprints	Message passing neural network	143
MEMES	AutoDock	Extended-connectivity fingerprints; Mol2Vec descriptors; CDDDs	Convolutional neural network; recurrent neural network	144
Yang et al. approach	Glide SP DOCK 3.7	Morgan fingerprints; molecular graphs	Graph-convolutional neural network; random forest	145
V-DOCK	AutoDock-Vina	2048 RDKit fingerprints combined with 166 bits; MACSS fingerprints	PyTorch deep learning library	146
Bucinsky et al. approach	AutoDock	SOAP molecular descriptors; SchNet 128 bits vectors	Keras neural network; deep tensor neural network; gradient-boosted trees	147
NeuralDock	MedusaDock	36 bits atom type vectors with 7 channels for ligands; 10×10×10, 2-angstrom resolution images with 8 channels for protein pockets	TensorFlow neural network	148
MILCDOCK	LeDock PLANT Vina AutoDock 4 rDock	Pose-based RMSD values; metadata from docking programmes	Gradient-boosted trees; random forest; naive Bayes neural network	149
DOCKSTRING	AutoDock-Vina	Various fingerprints	Regressions; gradient-boosted trees; Gaussian processes; graph neural network	150

CDDD, continuous and data-driven descriptors; QSAR, quantitative structure–activity relationship; RMSD, root-mean-square deviation; SOAP, smooth overlap of atomic positions.

their training set with highly scored compounds but the transfer learning approach enabled them to find potential binders for JAK2 kinase when only ligands for the closely related kinases JAK1, JAK3 and TYK2 were used to initiate the generative model.

The use of deep learning methods not only allows larger parts of a chemical space to be processed but also enables large-scale docking campaigns on a much broader range of targets. This could lead to a notable impact on both structure-based and structure-free models in predictive polypharmacology.

Outlook

The combination of the emergence of big data, increases in computational power and advances in deep learning methods is driving the progress in CADD described above. Here, we highlight some of the trends we expect to continue or to emerge in the coming years.

Expansion of accessible chemical space

The remarkable growth of 'make-on-demand' chemical libraries (Box 2) has brought unprecedented opportunities for virtual screening but also poses tremendous challenges. Virtual screening of such ultra-large collections by conventional molecular docking remains practically unfeasible due to the associated costs of both software licenses and access to high-performance computing. There are only a handful of packages that can screen libraries of around a billion compounds, and they often rely on their code scalability across supercomputing clusters to perform 'brute force' virtual screening. For example, the GigaDocking method implemented within the Orion package from OpenEye was used to dock the full Enamine REAL database (1.4 billion molecules) to two targets (purine nucleoside phosphorylase and heat shock protein 90) in less than a day. To do so, the OpenEye team used up to 45,000 Amazon Web Services CPUs to complete the tasks, corresponding

to a staggering ~50 years of continuous computing[109]. Similarly, the AutoDock programme was modified for more effective parallelization with GPUs, and the resulting GPU-AutoDock method was used on the 27,000 GPUs of the Summit supercomputer to process the Enamine REAL library against SARS-CoV-2 M^{pro} in 1 day[110]. In another large-scale study, Gorgulla et al.[111] used the VirtualFlow platform with AutoDock to screen the Enamine REAL library against KEAP1, an E3 ubiquitin ligase substrate adaptor, in 4 weeks using 8,000 CPUs. Venkatraman et al.[112] implemented an end-to-end drug discovery approach called DrugSniffer, which required 40,000 computational hours for the screening of 3.7 billion molecules against 22 target pockets in SARS-CoV-2 viral proteins.

These studies represent rare examples of ultra-large-scale distributed virtual screening campaigns, which are not generally affordable for the global research community. While it has been demonstrated that the use of larger chemical libraries improves the effectiveness of drug discovery[113], most reported virtual screening studies rely on a few million molecules or less. The dramatic expansion of accessible chemical space requires the development of novel approaches to virtual screening. Such modern CADD technologies, on the one hand, should be able to rank billions of potential ligands against any target of interest within a reasonable amount of time, and on the other, should enable automated hit-selection strategies that require minimal human intervention, especially as automated chemistry labs driven by artificial intelligence algorithms are being contemplated and established[114]. Various deep learning-based methodologies discussed above illustrate the practicality of hybrid approaches that combine the predictive power of conventional docking with the utility of ligand-based QSAR modelling. In our opinion, these advances provide opportunities for global democratization of drug discovery owing to their wider accessibility and affordability. In particular, deep learning-enabled methods of docking and predictive polypharmacology could provide affordable means for screening billions of potential drug candidates. Furthermore, open-source-enabled and GUI-enabled implementation of such methods could make virtual screening of ultra-large chemical libraries more accessible.

Deep learning accelerates quantum mechanics calculations

The high computational demands of quantum mechanics calculations have traditionally limited their applicability in QSAR modelling and CADD in general. Consequently, the development of fast, accurate and universal approximations to quantum mechanics has long been a focus in computational chemistry that has recently been enriched by the use of deep learning approaches.

One type of deep learning-based model that has been developed with such a focus is atomistic neural network potentials (NNPs)[115,116]. NNPs can predict energies and other quantum mechanics properties of molecules, generalizing to the same level of accuracy as density-functional theory on a large set of organic molecules while being six orders of magnitude faster[115,116]. Data acquisition for NNPs is based on the concept of active learning, which uses the disagreement between an ensemble of machine learning potentials to infer the reliability of the ensemble prediction. Active learning allows automatic sampling of regions of chemical space where the machine learning potential fails to accurately predict the potential energy, thereby reducing the size of the data set required for training by up to 90% compared to naive random sampling techniques[117].

Notably, the training process for NNPs is analogous to those used to train deep QSAR models, except that the target property, such as energy, is computed with full quantum mechanics methods rather than measured as in a bioactivity prediction task. While NNPs are fast and accurate, the majority do not aim to become universal in their description of chemical interactions, which limits their use in CADD to specific molecular systems. However, this limitation has been addressed with the breakthrough development of the first universal atomistic NNP for organic molecules known as ANI-1 (ref. 118).

The results of ANI-1 and its successors[117,119] have been shown to approach 'chemical accuracy' (errors of ~1 kcal/mol) relative to reference quantum mechanics data for multiple applications. Even the early version of the ANI-1 potential was found to be more accurate than semi-empirical and tight-binding quantum mechanics methods while being much faster. The model correctly predicted the stability of ring-containing structures and captured the large conformational changes. Additionally, the potential accurately predicted shape and smoothness of the potential energy surface (PES), which is traditionally defined by bond stretching, angle bending and torsional rotations. ANI models have been actively used as a proxy for quantum mechanics calculations to parametrize custom ligands in general force fields such as the general Amber force field[120]. It has been shown that PESs described by the ANI models are practically indistinguishable from PESs obtained with quantum mechanics calculations[117,119].

Deep learning improves accuracy of ligand binding affinity and property calculations

The ANI molecular mechanics scheme has been expanded to predict protein–ligand binding free energies[121], which represent another target property for deep QSAR modelling. Binding free energies of ligands can be calculated with molecular dynamics simulations of the protein–ligand complex using molecular mechanics potential energy of the complex and molecular mechanics and machine learning potential energies in a vacuum for the ligand (Fig. 4). The approach has been tested with predictions of the binding affinities of kinase inhibitors using ANI-2x potential and AMBER14SB/TIP3P force fields. Studies with inhibitors for tyrosine-protein kinase 2 (TYK2) from the Schrödinger JACS benchmark set[122] showed that the machine learning–molecular mechanics approach substantially reduced the absolute binding free energy errors obtained with molecular mechanics calculations[121]. While the molecular mechanics approach led to free energies with a root-mean-square error of 0.97 kcal/mol, correcting the molecular mechanics free energy with the machine learning–molecular mechanics approach improved results to a root-mean-square error of 0.47 kcal/mol. Notably, the molecular mechanics to machine learning–molecular mechanics corrections were all positive (up to a magnitude of 4 kcal/mol), whereas aliphatic groups with high conformational degrees of freedom tend to have larger corrections.

The recently developed AIMNet model has a revised architecture that builds on the success of ANI potentials[123]. The overarching principle was inspired by atoms in molecules (AIM) theory[124], which states that an electron density distribution function can be used to partition a molecule into interacting atoms. In the AIMNet model, atoms are characterized by learnable atomic feature vectors to approximate complex interatomic interactions instead of electron density. Knowledge inside the AIM layer could be exploited to learn new atomic properties without effectively retraining the model. For example, following training of the AIMNet model to predict the energy, partial atomic charges and atomic volumes, the Gibbs free energy of solvation could be predicted with an accuracy of 1.8 kcal/mol based on the AIM layer vector only, using just 6% of the training data[123]. Benchmarks for macrocycle conformer

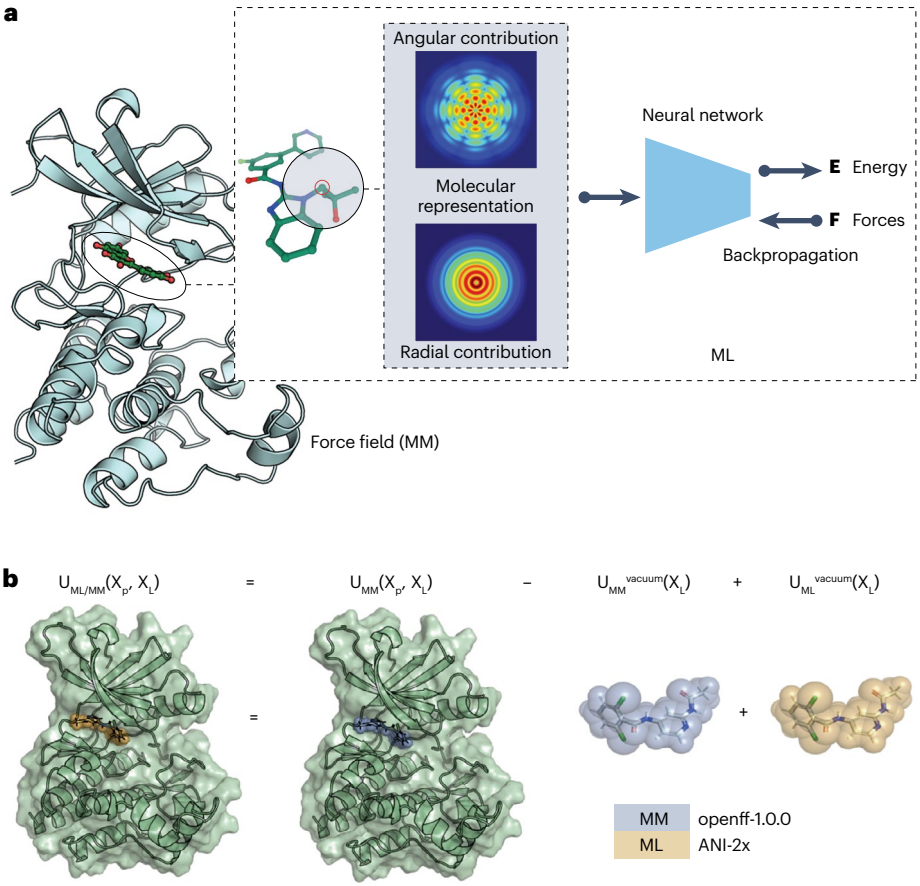


Fig. 4 | Molecular simulations enhanced by deep learning potentials in the calculation of ligand binding affinity. Machine learning (ML) potentials developed with deep learning approaches (panel **a**) improve the accuracy of calculating protein–ligand interaction energies by using a hybrid ML–molecular mechanics (MM) scheme (panel **b**). Calculation of the hybrid, higher accuracy ML–MM energy $U_{ML/MM}(X_P,X_L)$ for the protein–ligand complex is achieved by subtracting the MM energy of the ligand in a vacuum and adding the more accurate ML energy of the ligand in a vacuum. The net effect of this scheme is that the less accurate MM potential energy of the ligand is replaced by its higher accuracy ML potential. Here, the MM model uses the OpenFF small-molecule force field[151], while the ML model uses ANI-2x[119]. Details of the algorithm for the development of the ANI-type ML potentials are described elsewhere[121].

generation and ranking indicated that the model is on par with popular density-functional theory functionals.

Intriguingly, learned molecular representations enable accurate prediction of molecular properties of critical importance for drug discovery. A recently developed machine learning model for protein pK_a values achieved unprecedented accuracy of 0.5 log units for all amino acid types[125]. This model substantially surpassed the accuracy of the popular PROPKA method[126] and demonstrated the capabilities of NNPs to provide pK_a parameters that can be used as chemical descriptors in QSAR models or in mechanistic analysis of protein–ligand interactions[125].

In summary, deep learning methods employed to develop NNPs have not only enabled the acceleration of quantum mechanics calculations but also afforded much higher accuracy in calculating ligand binding affinity and molecular properties of high importance for CADD. For instance, as discussed above, deep docking methods critically depend on the quality of scoring functions used in the docking software. Fast and accurate NNPs, especially those enabled by the recent AIMNet method, promise to improve the accuracy of scoring functions and develop novel deep docking approaches as well as molecular simulations methods with higher efficiency and accuracy.

Potential impact of quantum computing

Challenges in working with both ultra-large databases and with quantum mechanics calculations[127] could also be addressed by a revolutionary advance in high-performance computing: quantum computing[128,129]. While conventional computers operate on binary encoded data represented by discrete states 0 or 1 (bits), quantum computers process information based on the laws of quantum mechanics, where quantum bits (qubits) can achieve states of 0 and 1 simultaneously. Consequently, qubits can exhibit distinct features, including superposition and entanglement, that allow quantum computers to manipulate vast amounts of information with very few operations, which enables unprecedented acceleration of certain computational tasks[128].

It is expected that quantum computers will be able to drastically outperform conventional processing units in several areas traditionally interrelated with CADD, including quantum chemistry computations and machine learning[127,130]. For the former group of approaches, quantum computers should be able to find solutions to the Schrödinger equation for large molecular systems, thereby bringing revolutionary changes into the development of novel synthetic routes, allowing the modelling of metabolic transformations and modes of action of covalent drugs, the study of transition states and coordinates of enzymatic reactions, the development of quantum chemical descriptors to empower QSAR modelling and the computation of accurate thermodynamics of drug–target interactions[131], and supporting other innovations in pharmaceutical discovery[128]. For machine learning methods, application of quantum computing could play a disruptive role as it can vastly outperform conventional supervised and unsupervised methods in both computational efficiency and accuracy[130,132]. For instance, quantum computing techniques implemented on noisy intermediate and scale quantum machines already include quantum

autoencoders, support vector machines, and methods of k-means clustering and principal component analysis[130]. Quantum methods implemented on fault-tolerant quantum computing devices include restricted Boltzmann machines, Bayesian inference, least-squares regression and support vector machines[130]. There are a growing number of studies where quantum computing has already been evaluated for conventional machine learning or deep learning applications in QSAR and CADD, including target discovery, protein folding, target site characterization[133], generative molecular modelling[134], docking and force field refinement[127], lead optimization[135], toxicity risk assessment[136], and molecular matching and similarity searching through the chemical space[132].

The emergence of hybrid CADD architectures integrating big data modelling algorithms into specialized hardware or combining classical and task-specific hardware, such as noisy intermediate and scale quantum computers or GPU platforms, are growing trends in drug discovery. We should expect that components of CADD pipelines will be served in the near future by different hybrid models fine-tuned for particular cheminformatics applications, empowering CADD by enabling dramatically faster computing using much bigger data sets.

Impact on early stage drug discovery

Although it is too early for deep learning methods, and deep QSAR in particular, to have enabled the development of approved drugs[137], there is growing evidence that these methods have accelerated the preclinical research stages for small-molecule drug candidates. After 2020, when Exscientia announced that its first drug candidate designed by artificial intelligence entered a phase I clinical trial, several companies followed with similar announcements. Notably, Exscientia reported that it took 12 months only to complete the exploratory research phase prior to the trial. Similarly, Insilico Medicine reported that it took them 30 months to develop a novel, artificial intelligence-designed, phase I anti-fibrotic clinical candidate, starting with the discovery of a novel target. Both companies employ types of deep learning approaches that we have discussed in this Perspective in the context of deep QSAR as part of their computational platforms. These recent successes suggest that the field of deep QSAR is beginning to reach its 'plateau of productivity' in the Gartner hype cycle. Continued development and use of deep QSAR methods should increasingly enable accelerated discovery of small-molecule drug candidates, which could be especially important in the face of novel and often unpredictable threats posed by emerging infectious diseases such as COVID-19.

Published online: 08 December 2023

References

1. Hansch, C., Maloney, P., Fujita, T. & Muir, R. Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. *Nature* **194**, 178–180 (1962).
2. Cherkasov, A. et al. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **57**, 4977–5010 (2014).
3. Muratov, E. N. et al. QSAR without borders. *Chem. Soc. Rev.* **49**, 3525–3564 (2020).
4. Ivakhnenko, A. G. & Lapa, V. G. *Cybernetics and Forecasting Techniques* (American Elsevier Co, 1967).
5. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. *J. Chem. Inf. Model.* **55**, 263–274 (2015).
6. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. & Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **23**, 1241–1250 (2018).
7. Yang, X., Wang, Y., Byrne, R., Schneider, G. & Yang, S. Concepts of artificial intelligence for computer-assisted drug discovery. *Chem. Rev.* **119**, 10520–10594 (2019).
8. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2**, 573–584 (2020).
9. Pandey, M. et al. The transformational role of GPU computing and deep learning in drug discovery. *Nat. Mach. Intell.* **4**, 211–221 (2022).
10. Bengio, Y., Courville, A. & Vincent, P. Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–1828 (2012).
11. Real, E., Aggarwal, A., Huang, Y. & Le, Q. V. Regularized evolution for image classifier architecture search. Preprint at:*arXiv* https://doi.org/10.48550/arXiv.1802.01548 (2018).
12. Elsken, T., Metzen, J. H. & Hutter, F. Neural architecture search: a survey.*J. Mach. Learn. Res.* **20**, 1–21 (2019).
13. Li, X. & Fourches, D. Inductive transfer learning for molecular activity prediction: next-gen QSAR models with MolPMoFiT. *J. Cheminform.* **12**, 27 (2020).
14. Xu, Y., Ma, J., Liaw, A., Sheridan, R. P. & Svetnik, V. Demystifying multitask deep neural networks for quantitative structure–activity relationships. *J. Chem. Inf. Model.* **57**, 2490–2504 (2017).
15. Moon, C. & Kim, D. Prediction of drug-target interactions through multi-task learning. *Sci. Rep.* **12**, 18323 (2022).
16. Fourches, D., Muratov, E. & Tropsha, A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.* **50**, 1189–1204 (2010).
17. Fourches, D. et al. Trust, but verify II: a practical guide to chemogenomics data curation. *J. Chem. Inf. Model.* **56**, 1243–1252 (2016).
18. Fourches, D., Muratov, E. & Tropsha, A. Curation of chemogenomics data. *Nat. Chem. Biol.* **11**, 535 (2015).
19. Alves, V. M. et al. Curated data in — trustworthy in silico models out: the impact of data quality on the reliability of artificial intelligence models as alternatives to animal testing. *Altern. Lab. Anim.* **49**, 73–82 (2021).
20. Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **29**, 476–488 (2010).
21. Golbraikh, A., Muratov, E., Fourches, D. & Tropsha, A. Data set modelability by QSAR. *J. Chem. Inf. Model.* **54**, 1–4 (2014).
22. Maggiora, G. M. On outliers and activity cliffs — why QSAR often disappoints. *J. Chem. Inf. Model.* **46**, 1535 (2006).
23. Aldeghi, M. et al. Roughness of molecular property landscapes and its impact on modellability. *J. Chem. Inf. Model.* **62**, 4660–4671 (2022).
24. Bosc, N. et al. Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery. *J. Cheminform.* **11**, 4 (2019).
25. Varnek, A. & Tropsha, A. *Chemoinformatics Approaches to Virtual Screening*. https://doi.org/10.1039/9781847558879 (Royal Society of Chemistry, 2008).
26. Schneider, G. & Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **4**, 649–663 (2005).
27. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).
28. Schneider, P. et al. Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug Discov.* **19**, 353–364 (2019).
29. Schneider, G. Mind and machine in drug design. *Nat. Mach. Intell.* **1**, 128–130 (2019).
30. Schneider, G. & Clark, D. E. Automated de novo drug design: are we nearly there yet? *Angew. Chem. Int. Ed. Engl.* **58**, 10792–10803 (2019).
31. Hartenfeller, M. et al. DOGS: reaction-driven de novo design of bioactive compounds. *PLoS Comput. Biol.* **8**, e1002380 (2012).
32. Tong, X. et al. Generative models for de novo drug design. *J. Med. Chem.* **64**, 14011–14027 (2021).
33. Moret, M. et al. Leveraging molecular structure and bioactivity with chemical language models for de novo drug design. *Nat. Commun.* **14**, 114 (2023).
34. Segler, M. H. S., Kogej, T., Tyrchan, C. & Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **4**, 120–131 (2018).
35. Blaschke, T., Olivecrona, M., Engkvist, O., Bajorath, J. & Chen, H. Application of generative autoencoder in de novo molecular design. *Mol. Inform.* **37**, 1700123 (2018).
36. Putin, E. et al. Reinforced adversarial neural computer for de novo molecular design. *J. Chem. Inf. Model.* **58**, 1194–1204 (2018).
37. Atz, K., Grisoni, F. & Schneider, G. Geometric deep learning on molecular representations. *Nat. Mach. Intell.* **3**, 1023–1032 (2021).
38. Button, A., Merk, D., Hiss, J. A. & Schneider, G. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nat. Mach. Intell.* **1**, 307–315 (2019).
39. Grisoni, F. Chemical language models for de novo drug design: challenges and opportunities. *Curr. Opin. Struct. Biol.* **79**, 102527 (2023).
40. Kotsias, P. C. et al. Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks. *Nat. Mach. Intell.* **2**, 254–265 (2020).
41. Korshunova, M. et al. Generative and reinforcement learning approaches for the automated de novo design of bioactive compounds. *Commun. Chem.* **5**, 129 (2022).
42. Baskin, I. I. Is one-shot learning a viable option in drug discovery? *Expert Opin. Drug Discov.* **14**, 601–603 (2019).
43. Simões, R. S., Maltarollo, V. G., Oliveira, P. R. & Honorio, K. M. Transfer and multi-task learning in QSAR modeling: advances and challenges. *Front. Pharmacol.* **9**, 74 (2018).
44. Moret, M., Helmstädter, M., Grisoni, F., Schneider, G. & Merk, D. Beam search for automated design and scoring of novel ROR ligands with machine intelligence. *Angew. Chem. Int. Ed. Engl.* **60**, 19477–19482 (2021).
45. Moret, M., Friedrich, L., Grisoni, F., Merk, D. & Schneider, G. Generative molecular design in low data regimes. *Nat. Mach. Intell.* **2**, 171–180 (2020).
46. Blaschke, T. et al. REINVENT 2.0: an AI tool for de novo drug design. *J. Chem. Inf. Model.* **60**, 5918–5922 (2020).

47. Grisoni, F. & Schneider, G. De novo molecular design with chemical language models. *Methods Mol. Biol.* **2390**, 207–232 (2022).
48. Chen, H. Can generative-model-based drug design become a new normal in drug discovery? *J. Med. Chem.* **65**, 100–102 (2022).
49. Lam, L. & Suen, C. Y. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Trans. Syst. Man Cybern. Part. A Syst. Hum.* **27**, 553–568 (1997).
50. Nippa, D. F. et al. Enabling late-stage drug diversification by high-throughput experimentation with geometric deep learning. Preprint at: *ChemRxiv* https://doi.org/10.26434/CHEMRXIV-2022-GKXM6 (2022).
51. Clark, K., Luong, M.-T., Le, Q. V. & Manning, C. D. ELECTRA: pre-training text encoders as discriminators rather than generators. Preprint at:*arXiv* https://doi.org/10.48550/arxiv.2003.10555 (2020).
52. Corey, E. J. & Wipke, W. T. Computer-assisted design of complex organic syntheses. *Science* **166**, 178–192 (1969).
53. Corey, E. J. General methods for the construction of complex molecules. *Pure Appl. Chem.* **14**, 19–38 (1967).
54. Coley, C. W., Green, W. H. & Jensen, K. F. Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).
55. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
56. Segler, M. H. S. & Waller, M. P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry* **23**, 5966–5971 (2017).
57. Szymkuć, S. et al. Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem. Int. Ed. Engl.* **55**, 5904–5937 (2016).
58. Genheden, S. et al. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. *J. Cheminform.* **12**, 70 (2020).
59. Jin, W., Coley, C. W., Barzilay, R. & Jaakkola, T. In: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* 2608–2617 (Neural Information Processing Systems Foundation, 2017).
60. Sutskever, I., Vinyals, O. & Le, Q. V. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems* 2, 3104–3112 (Neural Information Processing Systems Foundation, 2014).
61. Liu, B. et al. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Cent. Sci.* **3**, 1103–1113 (2017).
62. Schwaller, P., Gaudin, T., Lányi, D., Bekas, C. & Laino, T. "Found in Translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chem. Sci.* **9**, 6091–6098 (2018).
63. Wołos, A. et al. Computer-designed repurposing of chemical wastes into drugs. *Nature* **604**, 668–676 (2022).
64. Patel, H. et al. SAVI, in silico generation of billions of easily synthesizable compounds through expert-system type rules. *Sci. Data* **7**, 384 (2020).
65. Zabolotna, Y. et al. SynthI: a new open-source tool for synthon-based library design. *J. Chem. Inf. Model.* **62**, 2151–2163 (2022).
66. Bonnet, P. Is chemical synthetic accessibility computationally predictable for drug and lead-like molecules? A comparative assessment between medicinal and computational chemists. *Eur. J. Med. Chem.* **54**, 679–689 (2012).
67. Boda, K., Seidel, T. & Gasteiger, J. Structure and reaction based evaluation of synthetic accessibility. *J. Comput. Aided Mol. Des.* **21**, 311–325 (2007).
68. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **1**, 8 (2009).
69. Coley, C. W., Rogers, L., Green, W. H. & Jensen, K. F. SCScore: synthetic complexity learned from a reaction corpus. *J. Chem. Inf. Model.* **58**, 252–261 (2018).
70. Hoonakker, F., Lachiche, N., Varnek, A. & Wagner, A. A representation to apply usual data mining techniques to chemical reactions — illustration on the rate constant of S(N)2 reactions in water. *Int. J. Artif. Intell. Tools* **20**, 253–270 (2010).
71. Gimadiev, T. et al. Bimolecular nucleophilic substitution reactions: predictive models for rate constants and molecular reaction pairs analysis. *Mol. Inform.* **38**, 1800104 (2019).
72. Baskin, I. I., Madzhidov, T. I., Antipin, I. S. & Varnek, A. A. Artificial intelligence in synthetic chemistry: achievements and prospects. *Russ. Chem. Rev.* **86**, 1127–1156 (2017).
73. Glavatskikh, M. et al. predictive models for kinetic parameters of cycloaddition reactions. *Mol. Inform.* **38**, 1800077 (2019).
74. Gimadiev, T. R. et al. Assessment of tautomer distribution using the condensed reaction graph approach. *J. Comput. Aided Mol. Des.* **32**, 401–414 (2018).
75. Granda, J. M., Donina, L., Dragone, V., Long, D. L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).
76. Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* **360**, 186–190 (2018).
77. Skoraczyński, G. et al. Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient? *Sci. Rep.* **7**, 3582 (2017).
78. Probst, D., Schwaller, P. & Reymond, J.-L. Reaction classification and yield prediction using the differential reaction fingerprint DRFP. *Digit. Discov.* **1**, 91–97 (2022).
79. Marcou, G. et al. Expert system for predicting reaction conditions: the Michael reaction case. *J. Chem. Inf. Model.* **55**, 239–250 (2015).
80. Gao, H. et al. Using machine learning to predict suitable conditions for organic reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018).
81. Afonina, V. A. et al. Prediction of optimal conditions of hydrogenation reaction using the likelihood ranking approach. *Int. J. Mol. Sci.* **23**, 248 (2021).
82. Lin, A. I. et al. Automatized assessment of protective group reactivity: a step toward big reaction data analysis. *J. Chem. Inf. Model.* **56**, 2140–2148 (2016).
83. Schneider, G. Automating drug discovery. *Nat. Rev. Drug Discov.* **17**, 97–113 (2018).
84. Abolhasani, M. & Kumacheva, E. The rise of self-driving labs in chemical and materials sciences. *Nat. Synth.* **2**, 483–492 (2023).
85. Reutlinger, M., Rodrigues, T., Schneider, P. & Schneider, G. Combining on-chip synthesis of a focused combinatorial library with computational target prediction reveals imidazopyridine GPCR ligands. *Angew. Chem. Int. Ed. Engl.* **53**, 582–585 (2014).
86. Burger, B. et al. A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
87. Genheden, S., Norrby, P. O. & Engkvist, O. AiZynthTrain: robust, reproducible, and extensible pipelines for training synthesis prediction models. *J. Chem. Inf. Model.* **63**, 1841–1846 (2023).
88. Ton, A.-T., Gentile, F., Hsing, M., Ban, F. & Cherkasov, A. Rapid identification of potential inhibitors of SARS- CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol. Inform.* **39**, e2000028 (2020).
89. Cherkasov, A., Ban, F., Li, Y., Fallahi, M. & Hammond, G. L. Progressive docking: a hybrid QSAR/docking approach for accelerating in silico high throughput screening. *J. Med. Chem.* **49**, 7466–7478 (2006).
90. Hilpert, K., Fjell, C. D. & Cherkasov, A. Peptide-based drug design. *Methods Mol. Biol.* **494**, 127–159 (2008).
91. Durrant, J. D. & McCammon, J. A. NNScore 2.0: a neural-network receptor-ligand scoring function. *J. Chem. Inf. Model.* **51**, 2897–2903 (2011).
92. Svensson, F., Norinder, U. & Bender, A. Improving screening efficiency through iterative screening using docking and conformal prediction. *J. Chem. Inf. Model.* **57**, 439–444 (2017).
93. Ahmed, L. et al. Efficient iterative virtual screening with Apache Spark and conformal prediction. *J. Cheminform.* **10**, 8 (2018).
94. Rossetti, G. G. et al. Non-covalent SARS-CoV-2 Mpro inhibitors developed from in silico screen hits. *Sci. Rep.* **12**, 2505 (2022).
95. Gentile, F. et al. Automated discovery of noncovalent inhibitors of SARS-CoV-2 main protease by consensus deep docking of 40 billion small molecules. *Chem. Sci.* **12**, 15960–15974 (2021).
96. Garland, O. et al. Large-scale virtual screening for the discovery of SARS-CoV-2 papain-like protease (PLpro) non-covalent inhibitors. *J. Chem. Inf. Model.* **63**, 2158–2169 (2023).
97. Radaeva, M. et al. Discovery of novel Lin28 Inhibitors to suppress cancer cell stemness. *Cancers* **14**, 5687 (2022).
98. Gentile, F. et al. Deep docking: a deep learning platform for augmentation of structure based drug discovery. *ACS Cent. Sci.* **6**, 939–949 (2020).
99. Gorgulla, C. et al. VirtualFlow Ants — ultra-large virtual screenings with artificial intelligence driven docking algorithm based on ant colony optimization. *Int. J. Mol. Sci.* **22**, 5807 (2021).
100. Charifson, P. S., Corkery, J. J., Murcko, M. A. & Walters, W. P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **42**, 5100–5109 (1999).
101. Palacio-Rodríguez, K., Lans, I., Cavasotto, C. N. & Cossio, P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Sci. Rep.* **9**, 5142 (2019).
102. Ban, F. et al. Best practices of computer-aided drug discovery: lessons learned from the development of a preclinical candidate for prostate cancer with a new mechanism of action. *J. Chem. Inf. Model.* **57**, 1018–1028 (2017).
103. Liu, Z. et al. In: *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 381–385 https://doi.org/10.1109/BIBM52615.2021.9669513 (2021).
104. McNutt, A. T. & Koes, D. R. Improving ΔΔG predictions with a multitask convolutional siamese network. *J. Chem. Inf. Model.* **62**, 1819–1829 (2022).
105. Wang, J. & Dokholyan, N. V. Yuel: improving the generalizability of structure-free compound-protein interaction prediction. *J. Chem. Inf. Model.* **62**, 463–471 (2022).
106. Li, X. et al. Deep learning enhancing kinome-wide polypharmacology profiling: model construction and experiment validation. *J. Med. Chem.* **63**, 8723–8737 (2020).
107. Li, Z. et al. KinomeX: a web application for predicting kinome-wide polypharmacology effect of small molecules. *Bioinformatics* **35**, 5354–5356 (2019).
108. Krishnan, S. R., Bung, N., Bulusu, G. & Roy, A. Accelerating de novo drug design against novel proteins using deep learning. *J. Chem. Inf. Model.* **61**, 621–630 (2021).
109. Gentile, F. et al. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat. Protoc.* **17**, 672–697 (2022).
110. LeGrand, S. et al. In: BCB '20: *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* https://doi.org/10.1145/3388440.3412472 (Association for Computing Machinery, Inc., 2020).
111. Gorgulla, C. et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* **580**, 663–668 (2020).
112. Venkatraman, V. et al. Drugsniffer: an open source workflow for virtually screening billions of molecules for binding affinity to protein targets. *Front. Pharmacol.* **13**, 1389 (2022).
113. Lyu, J. et al. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
114. Häse, F., Roch, L. M. & Aspuru-Guzik, A. Next-generation experimentation with self-driving laboratories. *Trends Chem.* **1**, 282–291 (2019).
115. Zubatiuk, T. & Isayev, O. Development of multimodal machine learning potentials: toward a physics-aware artificial intelligence. *Acc. Chem. Res.* **54**, 1575–1585 (2021).
116. Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).

117. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: sampling chemical space with active learning. *J. Chem. Phys.* **148**, 241733 (2018).
118. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
119. Devereux, C. et al. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* **16**, 4192–4202 (2020).
120. Galvelis, R., Doerr, S., Damas, J. M., Harvey, M. J. & De Fabritiis, G. A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning. *J. Chem. Inf. Model.* **59**, 3485–3493 (2019).
121. Rufa, D. A. et al. Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning / molecular mechanics potentials. Preprint at: *bioRxiv* https://doi.org/10.1101/2020.07.29.227959 (2020).
122. Wang, L. et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 (2015).
123. Zubatyuk, R., Smith, J. S., Leszczynski, J. & Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Sci. Adv.* **5**, eaav6490 (2019).
124. Matta, C. F. & Boyd, R. J. An introduction to the quantum theory of atoms in molecules. *The Quantum Theory of Atoms in Molecules* https://doi.org/10.1002/9783527610709.ch1 (2007).
125. Gokcan, H. & Isayev, O. Prediction of protein pKa with representation learning. *Chem. Sci.* **13**, 2462–2474 (2022).
126. Bas, D. C., Rogers, D. M. & Jensen, J. H. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins* **73**, 765–783 (2008).
127. Lam, Y. H. et al. Applications. *Org. Process. Res. Dev.* **24**, 1496–1507 (2020).
128. Hassanzadeh, P. Towards the quantum of quantum chemistry in pharmaceutical process development: current state and opportunities-enabled technologies for development of drugs or delivery systems. *J. Control. Rel.* **324**, 260–279 (2020).
129. Li, Q. et al. The role of UNC5C in Alzheimer's disease. *Ann. Transl. Med.* **6**, 178 (2018).
130. Cao, Y., Romero, J. & Aspuru-Guzik, A. Potential of quantum computing for drug discovery. *IBM J. Res. Dev.* **62**, 10.1147/JRD.2018.2888987 (2018).
131. Kirsopp, J. J. M. et al. Quantum computational quantification of protein-ligand interactions. *Int. J. Quantum Chem.* **122**, e26975 (2022).
132. Outeiral, C. et al. The prospects of quantum computing in computational molecular biology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11**, e1481 (2021).
133. Li, J. et al. Drug discovery approaches using quantum machine learning. Preprint at: *arXiv* https://doi.org/10.48550/arxiv.2104.00746 (2021).
134. Romero, J., Olson, J. P. & Aspuru-Guzik, A. Quantum autoencoders for efficient compression of quantum data. *Quantum Sci. Technol.* **2**, 045001 (2017).
135. Cavasotto, C. N. Binding free energy calculation using quantum mechanics aimed for drug lead optimization. *Methods Mol. Biol.* **2114**, 257–268 (2020).
136. Heinen, S. et al. Predicting toxicity by quantum machine learning. *J. Phys. Commun.* **4**, 125012 (2020).
137. Jayatunga, M. K. P., Xie, W., Ruder, L., Schulze, U. & Meier, C. AI in small-molecule drug discovery: a coming wave? *Nat. Rev. Drug Discov.* **21**, 175–176 (2022).
138. Pyzer-Knapp, E. O. Using Bayesian optimization to accelerate virtual screening for the discovery of therapeutics appropriate for repurposing for COVID-19. Preprint at: *arXiv* https://doi.org/10.48550/arxiv.2005.07121 (2020).
139. Jastrzębski, S. et al. Emulating docking results using a deep neural network: a new perspective for virtual screening. *J. Chem. Inf. Model.* **60**, 4246–4262 (2020).
140. Graff, D. E., Shakhnovich, E. I. & Coley, C. W. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci.* **12**, 7866–7881 (2020).
141. Martin, L. J. State of the art iterative docking with logistic regression and Morgan fingerprints. *ChemRxiv* https://doi.org/10.26434/chemrxiv.14348117.v1 (2021).
142. Berenger, F., Kumar, A., Zhang, K. Y. J. & Yamanishi, Y. Lean-docking: exploiting ligands' predicted docking scores to accelerate molecular docking. *J. Chem. Inf. Model.* **61**, 2341–2352 (2021).
143. Kalliokoski, T. Machine learning boosted docking (HASTEN): an open-source tool to accelerate structure-based virtual screening campaigns. *Mol. Inform.* **40**, 2100089 (2021).
144. Mehta, S. et al. MEMES: machine learning framework for enhanced molecular screening. *Chem. Sci.* **12**, 11710–11721 (2021).
145. Yang, Y. et al. Efficient exploration of chemical space with docking and deep learning. *J. Chem. Theory Comput.* **17**, 7106–7119 (2021).
146. Choi, J. & Lee, J. V-Dock: fast generation of novel drug-like molecules using machine-learning-based docking score and molecular optimization. *Int. J. Mol. Sci.* **22**, 11635 (2021).
147. Bucinsky, L. et al. Machine learning prediction of 3CLpro SARS-CoV-2 docking scores. *Comput. Biol. Chem.* **98**, 107656 (2022).
148. Sha, C. M., Wang, J. & Dokholyan, N. V. NeuralDock: rapid and conformation-agnostic docking of small molecules. *Front. Mol. Biosci.* **9**, 244 (2022).
149. Morris, C. J., Stern, J. A., Stark, B., Christopherson, M. & Della Corte, D. MILCDock: machine learning enhanced consensus docking for virtual screening in drug discovery. *J. Chem. Inf. Model.* **62**, 5342–5350 (2022).
150. García-Ortegón, M. et al. DOCKSTRING: easy molecular docking yields better benchmarks for ligand design. *J. Chem. Inf. Model.* **62**, 3486–3502 (2022).
151. Qiu, Y. et al. Development and benchmarking of open force field v1.0.0 — the parsley small-molecule force field. *J. Chem. Theory Comput.* **17**, 6262–6280 (2021).
152. Tingle, B. I. et al. ZINC-22 — a free multi-billion-scale database of tangible compounds for ligand discovery. *J. Chem. Inf. Model.* **63**, 1166–1176 (2023).
153. Babuji, Y. Targeting SARS-CoV-2 with AI- and HPC-enabled lead generation: a first data release. Preprint at: *arXiv* https://doi.org/10.48550/arXiv.2006.02431 (2020).
154. Warr, W. A., Nicklaus, M. C., Nicolaou, C. A. & Rarey, M. Exploration of ultralarge compound collections for drug discovery. *J. Chem. Inf. Model.* **62**, 2021–2034 (2022).
155. Ruddigkeit, L., van Deursen, R., Blum, L. C. & Reymond, J. L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17.*J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
156. Oprea, T. I. & Gottfries, J. Chemography: the art of navigating in chemical space. *J. Comb. Chem.* **3**, 157–166 (2001).
157. Medina-Franco, J., Martinez-Mayorga, K., Giulianotti, M., Houghten, R. & Pinilla, C. Visualization of the chemical space in drug discovery. *Curr. Comput. Aided Drug Des.* **4**, 322–333 (2008).
158. Kireeva, N. et al. Generative topographic mapping (GTM): universal tool for data visualization, structure-activity modeling and dataset comparison. *Mol. Inform.* **31**, 301–312 (2012).
159. Zabolotna, Y. et al. Chemography: searching for hidden treasures. *J. Chem. Inf. Model.* **61**, 179–188 (2021).
160. Casciuc, I. et al. Virtual screening with generative topographic maps: how many maps are required? *J. Chem. Inf. Model.* **59**, 564–572 (2019).
161. Zabolotna, Y. et al. ChemSpace Atlas: multiscale chemography of ultralarge libraries for drug discovery. *J. Chem. Inf. Model.* **62**, 4537–4548 (2022).
162. Sattarov, B. et al. De novo molecular design by combining deep autoencoder recurrent neural networks with generative topographic mapping. *J. Chem. Inf. Model.* **59**, 1182–1196 (2019).
163. Bort, W. et al. Discovery of novel chemical reactions by deep generative recurrent neural network. *Sci. Rep.* **11**, 3178 (2021).

Acknowledgements

The authors acknowledge support of their studies by the National Institutes of Health (grant R01GM140154) for A.T. and National Science Foundation (grant CHE-2154447) for O.I.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Drug Discovery* thanks Esben Jannik Bjerrum, Eric Martin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Related links

BioSolveIT KnowledgeSpace: https://www.biosolveit.de/infiniSee/#knowledgespace
CAS. AI drug discovery: assessing the first AI-designed drug candidates to go into human clinical trials: https://www.cas.org/resources/cas-insights/drug-discovery/ai-designed-drug-candidates
CHEMRriya: https://chemriya.com/
Enamine REAL Database: https://enamine.net/compound-collections/real-compounds/real-database
Enamine REAL Space on-demand library: https://enamine.net/compound-collections/real-compounds/real-space-navigator
eXplore from eMolecules: https://www.biosolveit.de/2022/09/12/introducing-explore-trillion-sized-chemical-space-by-emolecules/
Gartner hype cycle: https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2022-gartner-hype-cycle
Insilico Medicine. From start to phase 1 in 30 months: https://insilico.com/phase1
Kaggle. Merck Molecular Activity Challenge: https://www.kaggle.com/c/MerckActivity
KNIME: https://sites.google.com/site/dtclabdc/
MolDB from Deepcure: https://deepcure.ai/technology/
WuXi AppTec: https://www.wuxiapptec.com/