Rethinking drug design in the artificial intelligence era

Petra Schneider, W. Patrick Walters, Alleyn T. Plowright, Norman Sieroka, Jennifer Listgarten, Robert A. Goodnow Jr., Jasmin Fisher, Johanna M. Jansen, José S. Duca, Thomas S. Rush, Matthias Zentgraf, John Edward Hill, Elizabeth Krutoholow, Matthias Kohler, Jeff Blaney, Kimito Funatsu, Chris Luebkemann and Gisbert Schneider

Abstract | Artificial intelligence (AI) tools are increasingly being applied in drug discovery. While some protagonists point to vast opportunities potentially offered by such tools, others remain sceptical, waiting for a clear impact to be shown in drug discovery projects. The reality is probably somewhere in-between these extremes, yet it is clear that AI is providing new challenges not only for the scientists involved but also for the biopharma industry and its established processes for discovering and developing new medicines. This article presents the views of a diverse group of international experts on the 'grand challenges' in small-molecule drug discovery with AI and the approaches to address them.

Despite advances in the understanding of disease biology and impressive leaps in technology, bringing new drugs to market remains a time-consuming and expensive process, largely owing to the substantial costs associated with the high proportion of failures in clinical trials[1,2]. Consequently, there is a need for fresh thinking, new and revised conceptions of the drug discovery process, and innovative approaches to deliver medicines for more patients at a lower cost-to-market. In this context, computer-assisted small-molecule drug design has long been considered a potential opportunity[3–6]. The field is now in the midst of a surge of interest, catalysed by advances in data processing power and the development of new artificial intelligence (AI) tools[7,8]. The key question is whether such approaches can help us design better small-molecule drug candidates faster[9,10].

For the past two decades, small-molecule drug discovery has been fuelled by high-throughput screening (HTS), with estimated hit rates of 0–0.01%, depending on several aspects, including the definition of a 'hit', the nature of the biological target, the assay type and readout, and the quality and diversity of the screening compound pool[11–13]. Selection of the most appropriate experimentally validated HTS hits for follow-up is critical to the success of a drug discovery project[14]. Many parameters need to be considered in hit selection and subsequent optimization, including potency and selectivity at the desired pharmacological targets and potential off-targets as well as the physicochemical characteristics that could be important in drug pharmacokinetics and safety. Consequently, medicinal chemists typically face challenging multi-objective optimization (MOO) problems, with far more potential choices than are possible to explore systematically as well as increasingly large and complex datasets to analyse when making their choices.

Part of the appeal of applying AI in drug design thus lies in the potential to develop data-driven, implicit model-building processes to navigate vast datasets arising from HTS and to prioritize alternatives. This represents at least a partial transfer of decision power to machine intelligence and could be viewed as synergistic with human intelligence, that is, a domain-specific implicit AI that would augment the capabilities of medicinal chemists in drug design and selection. More ambitiously, the ultimate challenge for drug design with AI is to autonomously generate new chemical entities (NCEs) with the desired properties from scratch (de novo), without the need for the often prohibitively costly full-deck HTS.

To be successful in the long run, drug design with AI (FIG. 1) has to provide solutions to several questions, which can be encompassed in five 'grand challenges': obtaining appropriate datasets, generating new hypotheses, optimizing in a multi-objective manner, reducing cycle times, and changing the research culture and creating an appropriate mindset. This Perspective is based on discussions around each of these five grand challenges among a group of diverse international experts at a meeting in 2018 on rethinking drug design with AI and presents the main conclusions drawn as well as a discussion on advances since then.

Obtaining appropriate datasets

Appropriate input data are crucial for building useful predictive models for decision-making and generation of NCEs[15,16]. Without an appropriate dataset and an understanding of the scope and limitations of those data, even a seemingly sophisticated model will not be able to produce useful results[17,18].

One of the most important factors when evaluating data for predictive modelling is whether those data were collected with the ultimate end point in mind and, if not, what might go wrong. For instance, many groups have built models to predict whether molecules will be toxic[19]. This problem is important, as a reliable toxicology model may be able to reduce the time and cost of drug discovery as well as the need for animal testing[20,21]. However, limited in vivo toxicology data are available, and therefore many toxicology models are built based on surrogate in vitro outcomes[22] and, in most cases, the relationship between these in vitro outcomes and the ultimate in vivo toxicology response has not been clearly established[23].

This issue is not specific to toxicology models. Animal models typically used in drug discovery may have a limited relationship with the outcome that will ultimately be seen in patients[24]. The relevance of data can also be a factor with in vitro experiments. The protein constructs used in biochemical assays may poorly reflect the native protein found in cells. In many cases, predictive models

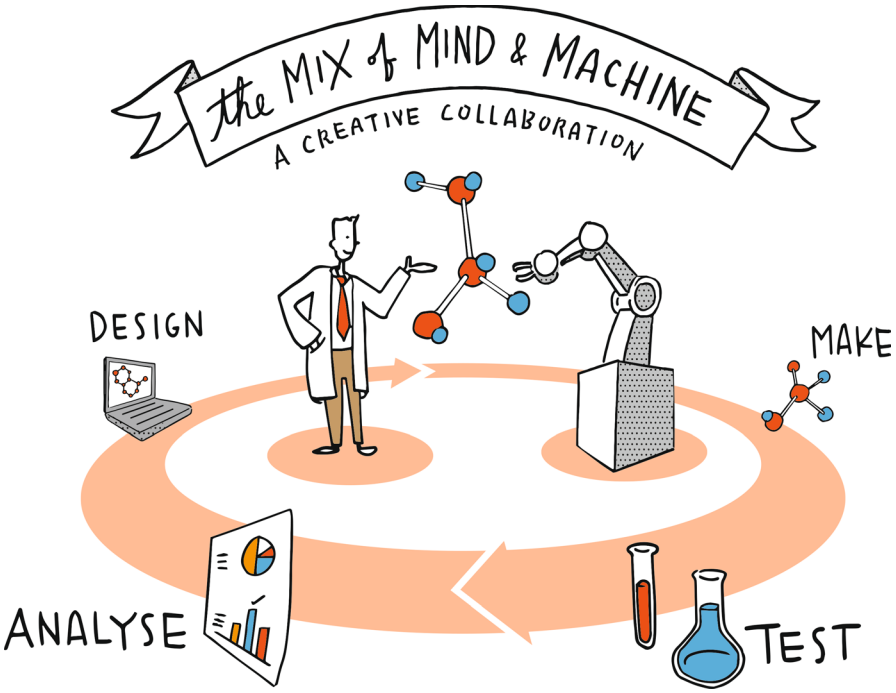


Fig. 1 | **Integrating mind and machine in drug discovery.** Artificial intelligence and laboratory automation could augment human decision-making, chemical synthesis and biological testing in design–make–test–analyse cycles involved in drug discovery. It is anticipated that this collaborative intelligence emerging from the combination of 'mind and machine' will enable better decision-making. Definitions of selected terms related to artificial intelligence are provided in the glossary. Some of the definitions are adapted from REF.[161]. Image courtesy of Jack Burgess, Jack Burgess Studios.

are constructed based on the results of previously designed high-throughput experiments. This approach may offer advantages owing to the larger dataset size — an important factor in modelling of complex phenomena. On the other hand, relying on existing large-scale data can be problematic because of compromises in design decisions made in the development of high-throughput assays. Adjustments made to increase assay throughput may diminish the relationship between the high-throughput assay and its more accurate, lower-throughput counterpart. For example, genome-wide off-target screening has much lower sensitivity than analysis of a pre-defined off-target area[25,26]. In order for data to be most useful, the context of the experiment used to capture the data and its relevance to the final outcome must be clearly understood[27].

Different levels of uncertainty exist between and within datasets. AI can be applied to address some of these uncertainties and generate higher-quality datasets. Related to the concept of dataset relevance, mentioned above, it is important to understand the provenance of the data, so that, if questions arise, the appropriate meta-data can be found[28,29]. This requires appropriate annotation, possibly by humans, which is a tedious process and, as such, is frequently not done or is ignored[30]. Although automated annotations exist to a certain extent (for example, machines generating and analysing data typically add metadata such as time and date), AI may be able to ease this burden by inferring context, providing a starting point for a human annotator and auto-detecting likely erroneous annotations from inconsistencies.

Another challenge to annotation is the rapidly changing and inconsistently described landscape of biology, that is, the lack of a coherent ontology[31,32]. AI techniques for language translation may be able to provide a mapping between terms in a rapidly evolving nomenclature, and AI techniques based on probabilistic latent variable models are already helping to extract meaning from multi-origin datasets in a clinical setting without shared institutional ontologies[33]. Additionally, scientific data stewardship and management should aim for Findable, Accessible, Interoperable, Reusable (FAIR) data[34]. Data should also follow the simple ALCOA (Attributable, Legible, Contemporaneous, Original and Accurate) principles defined by US FDA guidance[35,36]. Such guidelines should be updated as appropriate.

A critical factor in building predictive models is an understanding of the technical error and biological variability associated with the underlying data. In order to accurately leverage our data, we must have a clear understanding of both the accuracy and the precision of our measurements[37,38] (BOX 1).

Another source of difficulty and uncertainty when using experimental data to build predictive models is the accidental misreporting of data. Misreporting can take the form of simple typos in reported values, gene identifiers, units or other parameters that are reported in the scientific literature and stored in databases. Even one or two misreported data points have the potential to skew the results of a predictive model. Data curation and the identification of potential mistakes in data reporting is another area where AI might be relevant to drug discovery. AI techniques used for fraud detection are capable of spotting patterns that lie outside the commonly observed norm. It may be possible to apply some of these same ideas to identify data that may have been incorrectly reported. Of course, not all outliers are mistakes; they may instead highlight an alternative mechanism of action that could provide new insights. The way in which we search for outliers or potential errors may depend on the scale on which a model is being built[39].

Drug discovery is inherently an optimization problem. In order to generate a drug, a team must identify a compound which, among a plethora of criteria, is active against a biological target of interest, has an appropriate pharmacokinetic profile and does not produce adverse outcomes when dosed in vivo. As a result, drug discovery datasets often contain data for dozens of assays. However, this compound/assay matrix typically has missing values owing to time and money constraints. Because only compounds that perform well in higher-throughput in vitro or cellular assays are tested in more expensive in vivo experiments, data are also not missing completely at random[40] and, consequently, special care must be taken with such data. Even when datasets are complete, they are often imbalanced with either large numbers of inactive compounds and small numbers of active compounds, or vice versa. This imbalance can be particularly acute when taking data from the scientific literature, where the reporting of negative results is rare.

Increasingly, data modalities are generating richer representations — from dose–response curves to MRI images over

time — that are not immediately amenable to the current methods of analysis originally developed for simpler data. For example, most genetic studies to date focus on finding associations to a scalar (single-number) trait. With high-throughput imaging, one must consider whether it is possible or desirable to find lower-dimensional representations with which to do the associations, rather than with millions of pixels simultaneously, which is likely to yield under-powered tests. In brain imaging scans, for example, manual extraction of features, such as brain volume, is the norm. However, AI methods are working towards automating this process[41].

The domains with the most success to date with AI are imaging and natural language processing (NLP)[42]. These data differ substantially from those typically found in drug discovery, since the accurate labelling of whether 'a user followed a hyperlink' or 'there is a stop sign in an image' tends to be easier. The questions of whether a compound is 'active against a target' or 'toxic' are much more complex and labelled with much greater difficulty and nuances[43–45]. The given drug discovery project provides context for the data and enables the project members to draw conclusions from data analysis. However, if such data are pooled across multiple discovery projects or laboratories, the relevant context is often lost. Moreover, the domains of imaging processing and NLP have access to millions of data points for training, which is the reason why the first big successes of deep learning[46] have been in these areas. Indeed, many of the current and successful applications of AI, such as image classification, have required large training datasets, often in combination with data augmentation and model pre-training with weakly labelled data, to capture the diversity of the input and obtain a model that generalizes well[47,48]. Labelled datasets of this size do not exist in drug discovery owing to the more difficult process of obtaining the data, and the number and nature of molecules required to build a predictive model for drug design is still undetermined[49,50].

Another challenge related to the availability of data for model building is the fact that, in many cases, experiments do not generate data that can be easily translated to a single number like a biochemical dissociation constant (K_d) or a cellular effective concentration for half-maximum response (EC_{50}) value[51]. AI is now being used to develop representations of such experiments in ways that enable categorization of data that can ultimately be used to build predictive models. For instance, artificial neural networks are being employed to classify complex cellular phenotypes and build predictive models that are used for drug repurposing[52,53].

Over the past two decades, we have seen the appearance of a number of public databases containing millions of biological assay results, such as ChEMBL[54] and PubChem[55], which can provide input for machine learning models predicting a variety of biological activities or physical properties for drug-like molecules (see Click2Drug for a list). Although these databases are useful, the data contained in them only represent a small fraction of what has been measured because many of the larger datasets are proprietary to pharmaceutical companies or publishers and are not publicly and freely available. Most companies view their data as a competitive advantage and guard it closely; additionally, data sharing can be a complex undertaking. However, there are approaches, typically based on cryptography, that may enable the community to share specific parameters without disclosing proprietary information[56]. One way would be to collect metadata on many predictive models, ensuring that it does not de-identify inappropriate information. Such information might include the number of molecules used to build the model, some objective measure of molecular diversity, data range, data balance and so forth. An example is the SALT Knowledge Share Consortium, which examined the impact of particular chemical changes to molecules on a variety of in vitro outcomes. Rather than sharing the chemical structures of proprietary molecules, the companies shared the chemical differences between pairs of molecules (for example, phenyl to pyridyl) and the differences in activity (for example, threefold reduction in hERG activity). Other examples include the EU Innovative Medicines Initiative's projects for privacy-preserving federated machine learning and the ATOM Consortium's projects to provide a means for compound testing while maintaining confidentiality.

Generating new hypotheses

Only a tiny fraction of the drug-like chemical universe has been sampled in our search for new therapeutic agents, despite the advances in HTS technologies and the inventiveness of medicinal and synthetic chemists. A typical HTS deck (10^6–10^7 compounds)[57,58], even with the

Box 1 | The impact of experimental error on predictive model performance

In drug discovery, machine learning models are used to predict the physical properties, or biological activity, of molecules proposed for synthesis. These models are often used to rank order and ultimately prioritize ideas. The model generation process typically begins by calculating a set of characteristics, known as descriptors, for a set of molecules and developing a classification or regression model that relates the descriptors to some experimental observable such as a physical property or biological activity. For instance, a drug discovery team might use the aqueous solubility of a set of molecules to train a predictive model that is used to evaluate a set of ideas for new molecules and select the most promising candidates for improving solubility.

One critical factor, which is often overlooked in this process, is the impact of experimental error on the performance of the model. In a regression model, which predicts the value of the experimental observable, model performance is often assessed using a correlation coefficient such as Pearson's ρ, Spearman's ρ or Kendall's τ. These values, which range between –1 and 1, provide a quantitative measure of the relationship between the predicted and experimental values. Typically, models with correlation coefficients <0.5 are considered to have limited predictive value.

In a 2009 paper, Brown, Muchmore and Hajduk[162] described a method that uses simulation to estimate the maximum correlation that can be achieved with a regression model. This error is estimated by adding normally distributed variance to each experimental datapoint and calculating the correlation between the experimental data and the experimental data with the added variance. This simulation is then repeated several thousand times to obtain an estimate of the maximum observable correlation. As an example, we can consider the DLS100 dataset, a set of 100 aqueous solubility values determined using dynamic light scattering[163]. This data spans a range from 1.6 nM to 1.7 M, with a mean of 0.9 mM and standard deviation of 1.7 log units. If we carry out 1,000 cycles of the simulation above and assume a twofold experimental error (that is, the estimate of the range of values within which the true value of the quantity is likely to lie), the maximum achievable value for the Pearson correlation coefficient r is 0.98. If the error increases to fivefold, the maximum possible value of r decreases to 0.96. If the experimental error is tenfold, the best correlation that can be achieved decreases to $r = 0.86$.

One can see that experimental error can have a significant impact on correlation, even in the case of 'perfect' models such as those illustrated above. In an actual case, experimental error can make a difference between a useful model and one with limited practical value. Inflations of correlation are often due to overfitting of the datasets. These factors show that a proper treatment of experimental error and selection of datasets is critical when building a machine learning model.

advent of DNA-encoded libraries that can typically test 10^7–10^{10} compounds[59,60], is still sampling a small percentage of total chemical space. In 2015, the purchasable chemical space contained approximately 125 million compounds, and is still growing[61]. Automated 'ultra-large' chemical library docking has recently been performed for 170 million virtual compounds assembled from 130 chemical reactions — the largest in silico ligand–receptor docking run to date[62]. Still, this number is dwarfed by the estimated cardinality of drug-like chemical space in the order of 10^{18}–10^{200} molecules[63,64], depending on the constraints set[65].

The size of drug-like chemical space thus renders exhaustive enumeration impossible, and so drug design essentially boils down to the central question 'what to make next'. Medicinal chemists draw inspiration from their experience, from synthetic guidelines such as the non-mathematical Topliss Batchwise Scheme[66], which designates a special series of aromatic ring substituents intended to systematically explore design opportunities by optimizing substitution patterns, as well as from human creativity and loosely defined 'chemical intuition'. Given the complexity of human disease and the challenges that medicinal chemists and drug designers need to overcome, a more thorough approach for hypothesis generation in drug design may be beneficial[67,68].

Chemical design can be considered pattern matching[69–71] and, indeed, computer-based de novo design methods have been explored as idea generators to support drug design since the 1990s[72,73]. Today, however, generative AI offers a fresh approach to de novo drug design by providing a statistical framework for decision-making[74]. In contrast to earlier molecular design engines that employed a set of explicit chemical transformations (for example, virtual reaction schemes based on reaction SMILES) and assembly rules (for example, fragment growing and linking), these generative models represent chemical knowledge implicitly in terms of the statistical probabilities of data distributions. In other words, the language used by these two different concepts is no longer textbook chemistry (as we know it) but a new language learned from the training data (BOX 2).

This type of approach deserves further discussion as it directly relates to the interpretability problem of AI systems in chemistry. Nevertheless, pioneering prospective applications have shown that generative de novo design produces synthetically accessible molecules with desired properties and activities[75,76]. The main advantages of these models over previous de novo methods are: the speed of execution (NCEs can be generated on-the-fly to allow for interactive modelling); rapid re-training or fine-tuning on the project at hand; scalability by providing access to a virtually infinite chemical space without the need for explicit compound library enumeration; software availability; and synthetic accessibility of the designs, which has troubled many of the earlier de novo approaches.

Drug design will be confronted with increasingly more complex data and target hypotheses[77,78]. A key limitation in the drug discovery process is the lack of fundamental knowledge about human biology. While this Perspective focuses on drug design, the former point implies the need for adaptability during the optimization and design processes as the biological assays often rapidly change with the evolution of knowledge through the lifetime of the system being studied. AI therefore needs to provide answers flexibly, as drug discovery knowledge develops.

On the other hand, 'mechanistic' models are able to address both challenges by capturing behaviours at different levels of abstraction (for example, genetic, molecular and cellular) and providing explanations for how these behaviours evolve and interact[79–81]. In silico mechanistic models are complementary to AI-based approaches as they can add the interpretation (mechanistic explanation) for the associations found by machine learning models. Hence, with such models to provide new hypotheses and machine learning models to provide further data to test these hypotheses and improve the models, a virtual cycle is formed that creates a complete learning system.

Multi-objective optimization

NCE discovery requires the balancing of several criteria during the design process, including target potency, selectivity, clearance and permeability. However, optimizing for one of these properties may be to the detriment of others. Such a problem of potentially conflicting goals can be cast in the computational framework of MOO (also referred to variously as multi-parameter optimization, multi-objective programming, vector optimization, multicriteria optimization, multi-attribute optimization or Pareto optimization)[82–84]. The field of drug discovery and development is already well on its way to leveraging these computational techniques, which currently overlap with machine learning as a sub-branch of AI[85–88].

In the in silico MOO setting, one requires access to a set of computational predictive models for each desired property, and can then apply one of many existing MOO algorithms to attempt to solve the underlying optimization problem, that is, to find a molecule or set of molecules that balance the desired properties (BOX 3). As these properties are frequently in conflict, the goal is to generate a set of possible solution leads, each of which makes a trade-off in a different way, but where no member of the solution set could be improved in one property without giving something away in a different property. In this latter sense, each solution is therefore optimal. The set of solutions can be thought to trace out a frontier of optimality, where moving along the frontier yields a set of optimal solutions, each with its own way of trading off the properties.

Inherent to the goal of tracing out such a frontier is the fact that we are performing an optimization in the face of missing information. In particular, if we knew precisely how we were willing to trade-off various drug design criteria, which sometimes is the case at the start of a project, one could instead use more conventional computational optimization approaches to find a molecule that optimized the precisely known trade-off function. However, in drug discovery, as in many domains, the development process is iterative rather than analytical, with a substantial 'human-in-the-loop' component that is unlikely to disappear in the near future. This human component, for example, a medicinal chemist, imparts expert knowledge and decision-making that is not yet amenable to being encoded by statistical and machine learning models, either because of a paucity of relevant data or because the problem is just inherently difficult. Thus, the goal of MOO is to generate a set of different but practically optimal solutions to a particular molecular design challenge; subsequently, those solutions are handed over to human experts to sift through with deep, implicit knowledge and intuition — at least for the time being.

The goal of predictive modelling in the context of MOO for drug design is to reduce and even replace a laboratory measurement for a property of interest (or proxy to it), which may be categorial (classification) or continuous (regression), with a predictive model. For example, quantitative structure–activity relationship (QSAR) modelling seeks to model the mapping from predictive

features, such as physicochemical properties or other molecular representations, to a response variable such as biological activity in the form of pK_d or $\log(EC_{50})$ values. In the context of MOO, these models will help inform whether a particular molecule is on the optimality frontier or not.

The challenge of MOO is to take these models and, in a sense, back-engineer them to find the predictive features that correspond to optimal response variable settings ('inverse QSAR')[89,90]. Similar to the de novo design task discussed above in the generating new hypotheses section, certain generative AI models seem to be well-suited to properly address this problem[91]. As in standard machine learning, the basis for building such predictive models is to acquire a set of known feature–response pairs ('labelled data'), and then to 'train' a posited class of machine learning

Box 2 | Chemical hypothesis generation with constructive machine learning

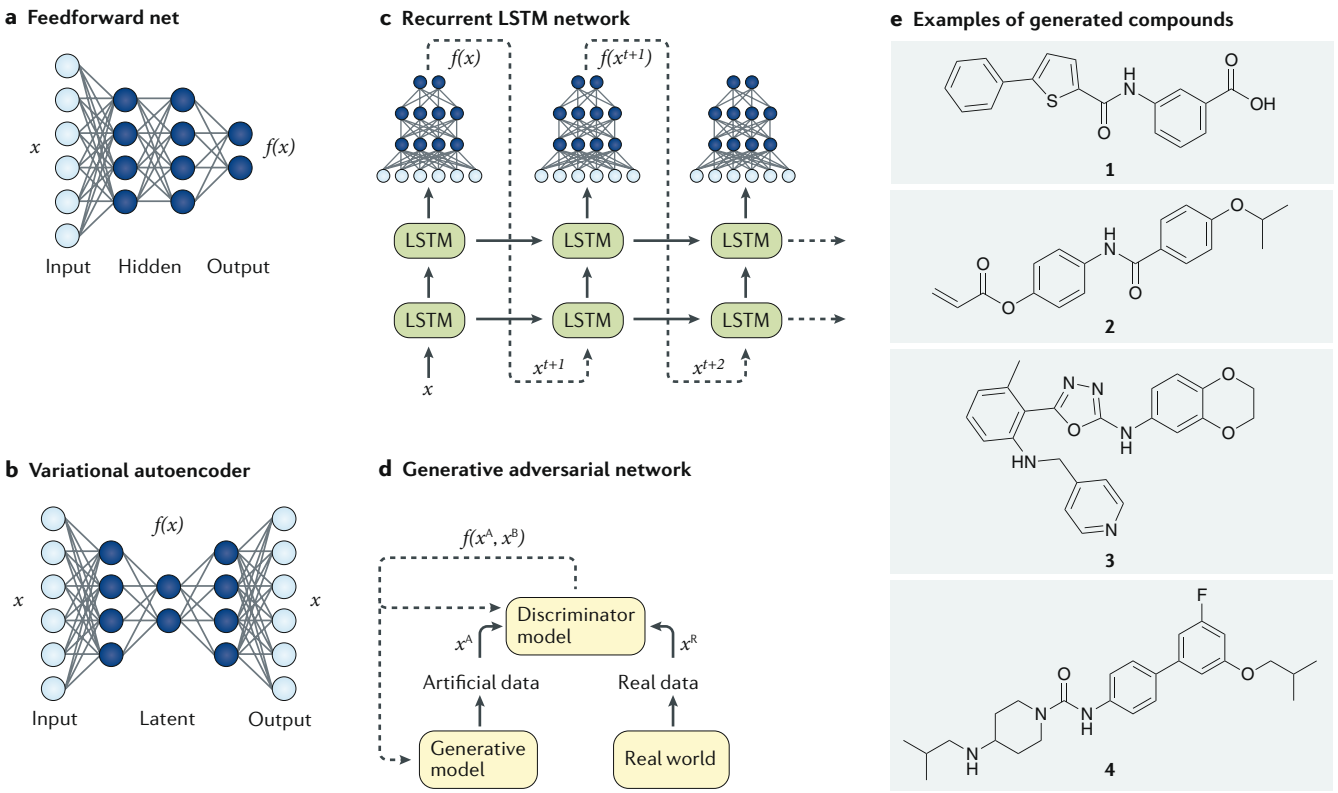
Constructive machine learning aims to create examples from its learned domain that are likely to exhibit desired properties. Generative deep neural networks belong to this class of algorithms and can be an important component in 'designing' new molecules from scratch or optimizing existing molecules. In contrast to rule-based artificial intelligence (AI) models for de novo drug design, these generative methods do not learn explicit chemical transformations (for example, forward and/or retrosynthetic reaction schemes)[164] to construct new molecules but instead model the distribution of chemical features of the training data and thereby implicitly capture important aspects of chemical 'synthesizability'. AI systems for explicit retrosynthetic analysis or synthesis planning can be used in concert with de novo design models[144–146,165].

Examples of models used to generate de novo structures are shown in the figure. These include deep 'back-transformation' (that is, inversion) of supervised feedforward networks[166] (part **a**), supervised and semi-supervised variational autoencoders (part **b**), recurrent neural networks[167,168] (part **c**) and generative adversarial networks (part **d**).

Models based on reinforcement methods often play a large role in computational design algorithms[169]. All of these neural network models transform a numerical input x to a function $f(x)$ as the output. Generative models can be used for back-transformation of $f(x)$ to x. When x is a suitable molecular representation, and the approach allows for some degree of ambiguity, then these systems generate new molecular structures.

Feedforward nets (part **a**) are universal function approximators, in which each layer of computational steps (circles) performs feature extraction from the training data. A deep network has several such data processing layers. Depending on the type of layers, whether the data have corresponding labels, and whether there are cycles of information allowed in the network architecture, one obtains various network types such as autoencoders (part **b**). Recurrent networks (for example, long short-term memory (LSTM) models; part **c**) perform pattern recognition by sequence analysis. Here, the input (for example, a SMILES string representing the atom connectivity of a molecular graph) is analysed in a step-wise token-based fashion, thereby enabling the detection of complex patterns in strings of text or other sequential data representations. In a generative adversarial network (part **d**), one network generates candidate molecules and the other evaluates them, for example, by comparing the virtual chemical structures with real examples for reinforcement learning.

The first successful syntheses of de novo-generated compounds (**1–4**; shown in part **e**) corroborate the practical applicability of generative molecular design to drug discovery. Pioneering prospective designs for the RXRγ agonists **1** (effective concentration for half-maximum response (EC_{50}) = 0.06 ± 0.02 μM) and **2** (EC_{50} = 19.1 ± 0.1 μM) were generated with a deep LSTM network[75]. This model was pre-trained with SMILES strings of bioactive compounds from ChEMBL[28] and fine-tuned on nuclear hormone receptor targets using transfer learning. The VEGFR2 kinase inhibitor **3** was constructed with a related SMILES-based approach, in which the bioactivity of the computer-generated molecules was estimated using ligand–receptor docking[161]. The partial 5-HT_{2B} antagonist **4** resulted from de novo structure generation with a neural network for virtual forward synthesis[167].



models, such as a particular neural network architecture, to fit the observed data while ignoring the 'noise' in the system, so that only relevant information, which could be applied to new, unseen data, is stored. The ability to also model uncertainty in the predictions will be of critical importance.

Neural networks, a popular class of predictive models at present, have far surpassed decade-old benchmarks on tasks in vision and audio. However, despite their popularity, their use thus far has not reliably resulted in a substantial improvement in other fields. The reasons for this are not yet known, though there are several possibilities. First, in vision and audio, one can readily acquire massive amounts of labelled data that are vital for success with current deep neural networks, whereas the fields of biology and chemistry are not typically yet sufficiently data-rich to make use of these neural networks, at least not above and beyond many other classes of models. However, the field of machine learning is actively pursuing how to do better with less data, sometimes referred to as 'few shot' learning[92–94]. Another potential reason is that the development of deep neural networks in recent years has been tailored to characteristics of data in the fields of audio and vision such as the highly specific structure of images based on pixels. Many advances in so-called general machine learning, such as convolutional filters, are leveraging such a vision task-specific structure but are then immediately applied in other domains, often without much thought as to the appropriateness. Deducing analogous structure in chemistry and biology or ways to encode it are in their infancy, and they are fundamentally more challenging endeavours than the analysis of vision and audio data. Both supervised and unsupervised graph-based neural networks are emerging as plausible approaches to tackling chemistry, although much work remains, including how to make these computationally scalable and well-suited to the domain (for example, by generating only synthetically accessible molecules)[95–97].

Several troubling phenomena have emerged from state-of-the-art deep neural network models in fields such as image processing. First, one can 'fool' a neural network to classify something that looks entirely like a cat, as, say, a dog or anything one wishes, by making tiny, imperceptible changes to the input image — these are known as adversarial examples[98]. While drug design is unlikely to face deliberate adversaries, the existence of these adversarial examples highlights the fact that deep neural network models have pitfalls that we may be blind to and can only mitigate once we become aware of them. Second, deep neural networks are notoriously over-confident in themselves[99]; that is, when a neural network says it thinks an image is 95% likely to be a cat, in reality, it is typically a much lower probability such as 20%. Especially in conjunction with drug design and MOO, it is important to understand and address these issues of uncertainty calibration[100,101]. Finally, MOO fundamentally relies on these models being reasonably accurate. However, it is thought by many that predictive models in general, and possibly more so deep neural networks, are extremely poor at extrapolation. During the course of MOO, these models will tend to veer away

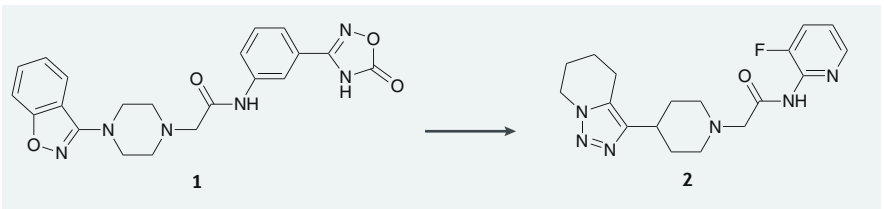
Box 3 | **Multi-objective lead optimization using machine learning**

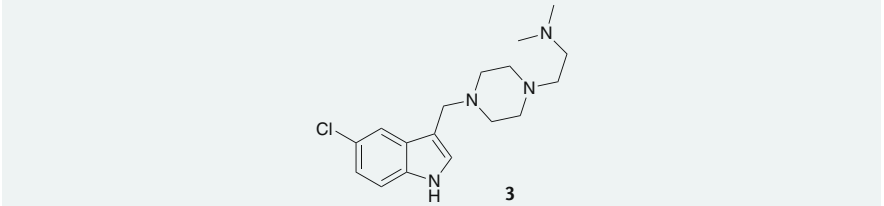
Multi-objective optimization of a chemical series aims to identify compounds that simultaneously meet all the desirable characteristics of a potential preclinical drug candidate. Ideally, this artificial intelligence (AI)-driven or AI-guided process takes the form of simultaneous optimization of a molecule's affinity for the desired biological target or targets, of selectivity against anti-targets and of various drug-like property measures. Examples of absorption, distribution, metabolism, excretion and toxicology (ADMET) properties that are generally indicative of drug-like properties include the following: aqueous solubility, metabolic stability, cellular permeability, action by cellular transporters (such as P-glycoprotein), drug–drug interactions, blood–brain barrier penetration and cardiotoxicity (such as hERG channel binding).

Recently, scientists at IKTOS and Servier described the use of deep learning for ligand-based optimization of compounds in a late-stage lead optimization context for an undisclosed target[170]. With a clear definition of metrics for 11 objectives (phenotypic activity, selectivity against 5-HT_{2A}, 5-HT_{2B}, α_1 and D_1 receptors, Na_V 1.2 and hERG ion channels, liver microsomal stability in rat and human, and Caco-2 Fabs and Efflux), no molecule within an initial dataset of 880 molecules satisfied all 11 objectives; furthermore, only 48 of the molecules had completed data with respect to all these objectives. Quantitative structure–activity relationship (QSAR) models were developed from the project dataset and used to score new virtual structures generated by a proprietary SMILES-based molecular generator built on deep learning long short-term memory (LSTM) generative models. The molecular generator used QSAR scoring on the generated structures to converge iteratively to structures that maximized QSAR scores for all 11 objectives. Finally, 150 virtual structures meeting all objectives were proposed, all of which had rare or absent substructures within the initial dataset. Of these, 20 compounds were selected for synthesis, and 11 were successfully synthesized. Upon assay, 3 compounds fulfilled all 11 objectified criteria. Compound **1** in the figure, which satisfied 9 of the 11 objectives, was part of the QSAR training set that led to the design, synthesis and assay of compound **2**, which satisfied all 11 objectives. The [1,2,3]triazolo[1,5-a]pyridine moiety appeared only six times in the initial dataset of 880 structures.

In an earlier study[171], researchers from ETH Zurich and Novartis used an adaptive algorithm for molecular construction (Molecular Ant Algorithm, known as MAntA)[172] to automatically generate small, synthetically accessible de novo designs that are selective antagonists of the 5-HT_{2B} receptor without binding to other 5-HT receptor subtypes or to a panel of 18 other protein targets, including the hERG potassium channel. QSAR predictions were based on multidimensional Gaussian regression models. The software suggested compound **3**, which exhibited the desired selectivity (5-HT_{2B} $K_i = 251 \pm 2$ nM), and was synthesized in flow from commercially available starting materials in one reductive amination step, in agreement with the synthetic route suggested by the MAntA molecular design method.

These selected examples of AI-guided multi-objective optimization show that there are several ways that machine intelligence can support the discovery of novel lead structures with multiple desired properties.





from 'comfort' zones, all the while being unaware of it and thereby possibly leading the design cycle dramatically astray. One can mitigate such problems by implicitly encoding prior information about what 'reasonable' molecules/proteins may look like and incorporating this into the MOO process[102]. We emphasize that we are not suggesting that one should avoid using deep neural networks: rather, that one must remain vigilant about their possible pitfalls and differences in their application between fields.

Assuming that one has access to reasonable predictive models to conduct a MOO, the question remains of how to attempt to solve the MOO problem. Historically, algorithms for MOO were dominated by 'genetic algorithms', which use an analogy to mutation and crossover diversification operations as well as a concept of fitness, to perform optimization[103]. These methods have been largely substituted by those in which the diversification operations have been replaced by sampling from statistical models within a more coherent statistical framework[104]. These latter approaches, typically in the class of estimation of distribution algorithms (EDA)[105], such as Covariance Matrix Adaptation Evolutionary Strategies[106], now overlap and are synergistic with machine learning methodology such as Information Geometric Optimization[107] and the Expectation-Maximization algorithm[108]. Moreover, these methods in turn have connections to machine learning for robotics, namely reinforcement learning[109,110]. Thus, cross-pollination between these areas of optimization and machine learning may lead to more rapid progress.

One fundamental ingredient to modern day MOO algorithms, such as EDA, is a generative model that takes the place of mutations and crossover operations in genetic algorithms. Perhaps the simplest possible generative model one can think of (for continuous data) would be a normal distribution with a mean and a variance parameter. As one changes these parameters, the samples from the normal correspondingly change in nature. Effectively, the way EDAs work is to have a sufficiently 'rich' generative model (that is, one that can generate a broad set of objects in the design class such as molecules) and then, using a particular statistical formalism, to tune the parameters such that only desired molecules can be sampled from it. Therefore, the ingredients of MOO are not only the predictive models used, and the MOO algorithm, but also the class of generative models. Modern day machine learning has seen a convergence of two technical fields — that of graphical models and neural networks — each with pros and cons, to achieve particularly rich classes of probabilistic generative models such as the variational autoencoders[111]. Drug design-based MOO stands to gain from making use of such advances and those in the future.

A related note is that of how to represent molecules and proteins in a way that is most amenable to leveraging the full power of current machine learning tasks such as predictive modelling and generative models. The problem of representation is in turn connected to the generative models just mentioned — these also generally provide representations of the inputs as a useful side effect or even *desideratum*. In the field of NLP, converting sentences, inherently composed of discrete symbols, into real-value vectors has been shown to provide benefit in downstream tasks[112]. Similar arguments and efforts were made in molecular design in the 1990s[7,113–115] and have recently been rediscovered in the context of deep learning[116,117]. Again, this area is one that could benefit from advances in related application domains but it is likely to benefit by adaptation rather than direct application. In starting this section, we described MOO as being used with a 'human-in-the-loop' to help decide between different optimal solutions; an area where AI and machine learning could be valuable is in enabling better encoding of the decisions being made by the human, such that these can be fed into the automated part of the system.

Reducing cycle times

Timelines and investments required for lead identification and optimization of NCEs are substantial and the risk of failure at all stages of the drug discovery process is high. To address this, the pharmaceutical industry has continuously invested in its compound profiling capabilities. The resulting growth in the number of generated data points is desirable but creates a number of challenges. Increasingly, the information processing capacity of the human brain becomes a limiting factor. In their efforts to keep up with the size, complexity and dimensionality of drug discovery projects (and to translate their findings into designs for new compounds), scientists often resort to simple heuristics and efficiency metrics[118–120]. While these methods have their merits (but also their controversy)[121], they have not led to significant reductions in the number of learning cycles or overall timelines needed to generate NCEs.

The central process in drug discovery to improve the profiles of lead molecules towards the required profile for a candidate drug is known as the design–make–test–analyse (DMTA) cycle[122]. This classical hypothesis-based method first uses available data to develop hypotheses and design molecules (or select existing molecules from libraries). The designed compounds are subsequently synthesized (or taken from libraries) and tested in the appropriate assays to investigate whether the design hypothesis was correct and improves understanding. This knowledge is then analysed and translated into the development of the design hypothesis for the next cycle (FIG. 1).

A number of groups have reported methods to improve the effectiveness of the DMTA cycle (BOX 4), for example, with the greater use of predicted data[123], improved data analysis tools[124] and increased effectiveness in compound synthesis[125,126], thereby shortening the timelines for the delivery of the critical data to address a hypothesis[127–130]. Certain aspects of AI potentially offer an alternative to HTS. Instead of compiling and relying on a large screening compound library, small numbers (<1,000) of de novo-generated compounds can be synthesized in each iteration of the DMTA cycle and only in the amount needed for testing, or sourced by in-house compound repositories or on-demand synthesis facilities that have cherry-picking capability, until a desired assay readout is obtained. However, this 'active learning' approach to hit and lead identification, despite its conceptual appeal, has its own issues, for example, the type of chemistry is limited to reactions that are amenable to automated microfluidics-assisted synthesis and analytics.

Even with these improvements, the cycle time of a DMTA iteration is still slow and can often take more than 4–8 weeks to complete. As a result, the required number of cycles to deliver a clinical candidate will require substantial time. Whilst the 'design' and 'analysis' phases can be fast and the 'test' phase (in particular in vitro data generation, including potency, selectivity and ADMET (absorption, distribution, metabolism, excretion and toxicology) profiling) can be optimized and streamlined to be fast and predictable, the 'make' phase is often slow, taking weeks to complete the synthesis of novel and complex molecules. Therefore, shortening this phase could substantially reduce the iteration time of a DMTA cycle. Laboratory automation, such as fast

Box 4 | **Reducing cycle times by automation**

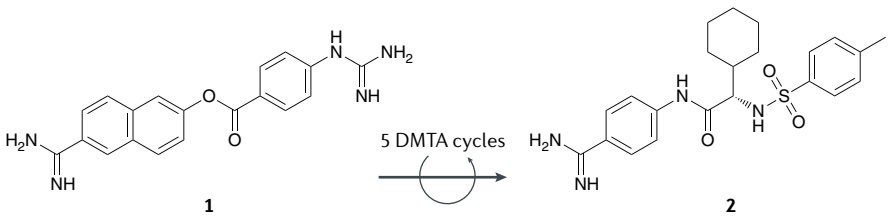
The potential value of automated design–make–test–analyse (DMTA) cycles in drug discovery is substantial[122,173]. Benefits of automation include[131]:

- Diminished measurement errors and reduced material consumption
- Fast feedback loops for artificial intelligence (AI)-based hypothesis generation and optimization
- Opportunities to expand the medicinal chemistry synthetic toolbox (for example, microfluidic-assisted synthesis reactions, reactions under extreme conditions)[134]
- Rigorous compound prioritization by applying sophisticated cell-based assays (for example, cells/organs-on-chips)[174–176] in an effort to more effectively recapitulate disease biology and thereby improve the likelihood of identifying compounds that show efficacy in humans
- Molecular optimization towards multiple relevant biochemical and biological endpoints without personal bias

Implementations of integrated DMTA platforms have already proved their efficiency and applicability to drug design and optimization[177]. For example, researchers at AbbVie have developed an integrated robotic platform for the automated parallel synthesis of small focused compound libraries, built mainly from commercially available components[178]. Their system is able to perform liquid handling and evaporation for inline analytics, purification and activity testing. Short turnaround times were reported, which allows the project teams involved to obtain results from hypothesis testing within a day or two[179].

Despite the advances of machine learning, chemical reactor and assay technology, so far, there are few published examples of fully integrated, automated, AI-driven discovery platforms. One such example is the automated closed-loop multiparameter design of selective hepsin inhibitors[180]. In progressing from the known hepsin inhibitor compound (**1** in the figure) to the designed compound (**2**) in automated DMTA cycles, the half-maximal inhibitory concentration (IC_{50}) value against hepsin was improved from ~1 μM to 33 nM, and the selectivity over urokinase-type plasminogen activator was increased from 30-fold to 100-fold. In each cycle, which took approximately 1.5 hours, a machine-learning model for activity prediction guided the selection of molecular building blocks for synthesis.

Nanoscale synthesis without the need for time-consuming purification in combination with rapid affinity ranking might further reduce DMTA cycle times. In a pioneering study, the coupling of high-throughput nanomole-scale synthesis with a label-free affinity-selection mass spectrometry bioassay facilitated simultaneous optimization of reaction conditions and building blocks as well as protein affinity[181].



1 **2**

compound synthesis using robust reactions in batch or flow with automated analytics and purification, has and will continue to play a decisive role in this context[131–134]. The choice of reactions for automation should primarily focus on the ones frequently employed by medicinal chemists[135,136], so the chemist is free to perform the more challenging synthetic steps and conceive new chemical reactions[137,138].

Owing to the variety of design hypotheses and the different times necessary to synthesize molecules and profile them in various assays, multiple design cycles are often run in parallel. Design cycles are often left incomplete as the next round of chemical synthesis is initiated before all of the data have arrived to allow effective answering of design hypotheses and true MOO[139]. Furthermore, all learnings from the previous cycle(s) cannot be factored into the design process. Often, the DMTA cycle becomes a 'workflow' more than an information generation cycle. The ever-increasing amounts of data that need to be captured and analysed in molecular design make it difficult for medicinal chemists and scientists to consistently incorporate a full understanding of thousands of data points and trends as well as to spot all lessons that the data can provide. In their efforts to keep up with the size, complexity and dimensionality of modern drug discovery project datasets (and to translate their findings into new compounds), scientists often have to resort to simple heuristics such as rules of thumb, efficiency metrics[140], model systems like log*P* or log*D*, or matched molecular pairs[141]. Searching for patterns, trends and insights (and translating them into new hypotheses and designs) is intellectually challenging and poses the risks of becoming overwhelming and missing important conclusions. AI has the potential to support scientists and enable better use of data in greater volumes for decision-making and, thus, lead to more productive discovery teams[142].

AI provides a range of opportunities to increase the effectiveness of the DMTA cycle, including integrating and analysing all available experimental and predicted data to support the chemists and design teams with insights and data analysis as well as the de novo design of molecules[143]. By providing improved synthesis routes and optimized reaction conditions, AI models could enable the chemist to follow the most effective route and ideally be 'right first time', ultimately shortening the 'make' phase[144–147].

Researchers would greatly benefit from receiving pre-digested and tailored pieces of information or recommendations in the right moment, format and context[148]. This would reduce their need to plough through raw data so they could instead focus on the assessment of the presented information using their chemical intuition and extensive background knowledge. To be successful, this scenario will require sufficient information on the data provenance, standardized data ontologies and the possibility to drill-down to the original raw data when necessary.

In a recent example focused on inhibitors of the kinase discoidin domain receptor 1 (DDR1)[149], deep learning was used to virtually generate a set of molecules and prioritize a small set of compounds for rapid synthesis and testing. A comprehensive machine learning model was built using a broad range of datasets and subsequently trained on data including the known DDR1 inhibitor and common kinase inhibitor datasets. Using reinforcement learning, the design stage was applied to provide 30,000 virtual molecules, which were filtered by applying various commonly employed criteria and chemical diversity sorting. Further prioritization using a range of conventional computational methods, including pharmacophore models, as well as selection by the team, eventually led to six compounds being synthesized and experimentally validated. Two of the compounds displayed half-maximal inhibitory concentration (IC_{50}) values of <20 nM in a DDR1 inhibition assay. One aspect of this study, not highlighted in the paper, is the fact that the most prominent molecule only differs from known DDR1 inhibitors, including the marketed drug

ponatinib (IC_{50} = 9 nM)[150], by a simple isosteric replacement (isoxazole for amide carbonyl). Furthermore, the reported timelines of 46 days to complete this DMTA cycle are not those required to deliver a drug candidate; the optimization steps that would typically be required would take long and incur larger costs. Hence, whilst this study is one of the rare examples in which machine learning has been applied prospectively in medicinal chemistry, it is still some way from demonstrating cycle time and cost reductions in the generation of a novel drug candidate.

The selection of a well-known biological target with an associated high degree of background knowledge, including a range of active chemical scaffolds and X-ray structural information, is clearly an advantage and a valid starting point to showcase deep learning technology. The applicability of AI for drug design from scratch in low-data situations is yet to be shown. In this context, the established concept of transfer learning[151] can provide 'few shot' methods for generative molecular design[152], and pioneering examples have demonstrated its practical applicability[78,79,153]. However, assessing the impact on hit and lead generation requires further validation of transfer learning methodology in different low-data situations and projects. An additional challenge will arise when working with biological targets with less associated knowledge.

In the near future, with more accurate predictive models across multiple parameters, the whole DMTA cycle could become virtual, with more intermittent synthesis of molecules to ensure that progress remains on track. With more integrated analyses, hypothesis generation would become faster and the proposed molecules would better address MOO challenges; molecules could even come with proposed synthesis routes. Ultimately, this could help decrease the time required for DMTA cycles and the clinical candidate delivery timeline. Moreover, if the AI component is useful for some projects, this would allow more resources to be focused on projects less amenable to AI methods.

Research culture and mindset

Aside from technical questions, possibly the greatest challenge to the success of applications of AI in drug discovery and development overall lies in nurturing the appropriate mindset and 'culture' of stakeholders, such that they are willing to apply these computational models and use their results. In the context of drug design, stakeholders include researchers from multiple disciplines as well as businesspeople, and it will be critical to facilitate mutual appreciation and discourse. This could be achieved by first recognizing and acknowledging the different experiences of the various stakeholders, and then adapting and developing common terminologies and exemplars to establish a clear role for each (and their interactions) in an AI-assisted drug design process. One important way to foster such a development already at the university level is to educate and guide students in critical thinking — in stepping back and becoming self-reflective and aware of other mindsets, and in becoming 'response-able' in the sense of being able to explain one's actions to colleagues (including those from other fields of research) and to a wider audience[154,155].

A key opportunity to encourage the uptake of AI approaches is to identify areas in which AI can augment and support (rather than replace) chemists and drug designers to make their processes more productive, while concomitantly increasing the quality of the data and the acceptance of AI. For example, one limit of AI is having well-curated data to build appropriate training sets, but the process to annotate and curate the data is one that many chemists find onerous. If electronic laboratory notebooks could leverage AI to facilitate the process of capturing, annotating and curating the data, chemists would be able to focus more time on the innovation and the human insight necessary to develop effective drugs. In turn, there would be better training sets to improve the output of the AI components. Furthermore, if AI tools could be leveraged to sift through past years of data from other medicinal chemistry programmes and connect the data to the current programme, additional directions in drug design space might be highlighted for further analysis by project teams[156]. However, to make such loops useful, drug discovery scientists must accept the value of the AI output and leverage it along with their own experience.

AI systems will also have to be able to interact and cooperate with human experts

Glossary

Adaptive algorithm
An adaptive algorithm implements a problem-solving heuristic that changes its behaviour at the time it is run, based on information available and a reward mechanism.

Artificial intelligence
(AI). The various definitions and interpretations of this term agree on three essential capabilities of an AI (most often referring to a computer or machine): (i) problem solving, (ii) learning from experience (memory and adaptation) and (iii) coping with new challenges (generalization).

Deep learning
A set of machine learning techniques that utilize multi-layer neural networks to derive relationships from data, specifically the use of neural networks (see below) with many layers. Neural networks with many layers are called 'deep neural networks', which corresponds to having many layers of function compositions. Typically, the deeper the layer, the more abstract the semantics of its 'feature space' (that is, the implicit representation created by the neural network at that layer).

Hypothesis
A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation, without any assumption of its truth. In the context of drug design, a molecular structure can serve as a hypothesis.

Machine learning
The science (and art) of programming computers so that they can learn from data; also a branch of artificial intelligence focused on one of several tasks, typically all function approximators. The most common task is the construction and training of classifier models, followed by regression models — both forms of 'supervised learning', wherein pairs of 'inputs' and 'labels' are used to train the model to then make label predictions for cases where only the inputs are observed. Also common in machine learning is 'unsupervised learning', wherein only 'inputs' are used (for example, a list of molecules numerically encoded such as by way of SMILES strings) and general properties of these are learned by the model, which can then tell you how likely a new input is to have belonged to this set of objects, or can be used to generate 'new' such objects. More nuanced mixing and matching of tasks is also possible, yielding 'semi-supervised learning'.

Natural language processing
(NLP). NLP is concerned with the interactions between computers and human (natural) languages, in particular how to process and analyse large amounts of natural language data, for example, scientific literature. Deep statistical machine learning models achieve state-of-the-art results in many natural language tasks, for example, in language modelling and parsing. NLP can also be used for chemical language analysis and de novo design.

Neural networks
A particular type of function approximators wherein functions that predict discrete classes (classifiers) or real-values (regression models) do so by composing a series of (typically nonlinear) functions, each one converting the previous layer's outputs into a new 'space'. These models have been around for decades but came to prominence in the 1990s when the combination of access to large datasets, along with the ability to train 'deep' models (see Deep learning) and more powerful computers, enabled them to break benchmarks in computational audio and vision tasks.

Research culture
A community sharing certain practices or using a common method or exemplar, that is, speaking a common language (including formalisms and algorithms) or sharing typical instances, illustrations or exemplifications (including molecular structures).

to conduct complex and only partially defined tasks. It will be essential to design such systems in a way that allows them to observe and learn from human behaviour in feedback cycles that are deemed beneficial for both sides. In addition, the uptake of AI-based systems will benefit if a comprehensible rationale is delivered alongside the recommendation or prediction itself. Considering the AI a collaborative partner rather than a competitor might be advisable[157,158]. This view would also positively contribute to the current debate on the patentability of AI-generated drugs[159,160].

Outlook

Pharmaceutical companies have started to adopt instances of AI-related technology, specifically various machine learning methods, through partnerships and collaboration, but there is an understandable reluctance to place all bets on AI-based drug discovery. A curious but cautious approach is advisable, given the highly complex and regulated nature of drug development. A long-term vision is needed when developing AI applications in drug design that could increase efficiency in the various processes involved and reduce barriers between the multiple research cultures in the ecosystem to create new medicines.

*Petra Schneider[1], W. Patrick Walters[2], Alleyn T. Plowright[3], Norman Sieroka[4], Jennifer Listgarten[5], Robert A. Goodnow Jr.[6], Jasmin Fisher[7,8], Johanna M. Jansen[9], José S. Duca[10], Thomas S. Rush[11], Matthias Zentgraf[12], John Edward Hill[13], Elizabeth Krutoholow[14], Matthias Kohler[15], Jeff Blaney[16], Kimito Funatsu[17,18], Chris Luebkemann[1,19] and Gisbert Schneider[1]**

[1]*ETH Zurich, RETHINK, Department of Chemistry and Applied Biosciences, Zurich, Switzerland.*

[2]*Relay Therapeutics, Cambridge, MA, USA.*

[3]*Sanofi-Aventis Deutschland GmbH, Integrated Drug Discovery, Frankfurt am Main, Germany.*

[4]*Institute for Philosophy, University of Bremen, Bremen, Germany.*

[5]*University of California, Berkeley, Electrical Engineering and Computer Science, Berkeley, CA, USA.*

[6]*Pharmaron USA LLC, Boston, MA, USA.*

[7]*Department of Biochemistry, University of Cambridge, Cambridge, UK.*

[8]*UCL Cancer Institute, University College London, London, UK.*

[9]*Novartis Institutes for BioMedical Research, Emeryville, CA, USA.*

[10]*Novartis Institutes for BioMedical Research, Cambridge, MA, USA.*

[11]*ATOM Consortium, San Francisco, CA, USA.*

[12]*Boehringer Ingelheim Pharma GmbH & Co. KG, Medicinal Chemistry, Biberach an der Riss, Germany.*

[13]*Healthy Life Consultants, New York, NY, USA.*

[14]*Bloomberg Intelligence, Bloomberg LP, London, UK.*

[15]*ETH Zurich, Department of Architecture, Zurich, Switzerland.*

[16]*Genentech, 1 DNA Way, South San Francisco, CA, USA.*

[17]*The University of Tokyo, Department of Chemical System Engineering, Tokyo, Japan.*

[18]*Nara Institute of Science and Technology, Data Science Center, Nara, Japan.*

[19]*Arup Global Foresight, San Francisco, CA, USA.*

**e-mail: gisbert@ethz.ch*

https://doi.org/10.1038/s41573-019-0050-3

Published online 4 December 2019

1. Smietana, K., Siatkowski, M. & Møller, M. Trends in clinical success rates. *Nat. Rev. Drug Discov.* **15**, 379–380 (2016).
2. Mullard, A. 2018 FDA drug approvals. *Nat. Rev. Drug Discov.* **18**, 85–89 (2019).
3. Hopfinger, A. J. Computer-assisted drug design. *J. Med. Chem.* **28**, 1133–1139 (1985).
4. Martin, Y. C. Computer-assisted rational drug design. *Methods Enzymol.* **203**, 587–613 (1991).
5. Yu, W. & MacKerell, A. D. Jr. Computer-aided drug design methods. *Methods Mol. Biol.* **1520**, 85–106 (2017).
6. Baig, M. H. et al. Computer aided drug design: success and limitations. *Curr. Pharm. Des.* **22**, 572–581 (2016).
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. Ching, T. et al. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **15**, 20170387 (2018).
9. Mignani, S., Huber, S., Tomás, H., Rodrigues, J. & Majoral, J. P. Why and how have drug discovery strategies in pharma changed? What are the new mindsets? *Drug Discov. Today* **21**, 239–249 (2016).
10. Jordan, A. M. Artificial intelligence in drug design – the storm before the calm? *ACS Med. Chem. Lett.* **9**, 1150–1152 (2018).
11. Bender, A. et al. Which aspects of HTS are empirically correlated with downstream success? *Curr. Opin. Drug Discov. Devel.* **11**, 327–337 (2008).
12. Gilad, Y., Nadassy, K. & Senderowitz, H. A reliable computational workflow for the selection of optimal screening libraries. *J. Cheminform.* **7**, 61 (2015).
13. Bajorath, J. Extending accessible chemical space for the identification of novel leads. *Expert Opin. Drug Discov.* **11**, 825–829 (2016).
14. Holenz, J. & Stoy, P. Advances in lead generation. *Bioorg. Med. Chem. Lett.* **29**, 517–524 (2019).
15. Oliveira, A. L. Biotechnology, big data and artificial intelligence. *Biotechnol. J.* **14**, e1800613 (2019).
16. Brown, N. et al. Big data in drug discovery. *Prog. Med. Chem.* **57**, 277–356 (2018).
17. Esaki, T. et al. Data curation can improve the prediction accuracy of metabolic intrinsic clearance. *Mol. Inf.* **38**, 1800086 (2019).
18. Fourches, D., Muratov, E. & Tropsha, A. Trust, but verify II: a practical guide to chemogenomics data curation. *J. Chem. Inf. Model.* **56**, 1243–1252 (2016).
19. Cases, M. et al. The eTOX data-sharing project to advance in silico drug induced toxicity prediction. *Int. J. Mol. Sci.* **15**, 21136–21154 (2014).
20. Huang, R. et al. Modelling the Tox21 10K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nat. Commun.* **7**, 10425 (2016).
21. Kirchmair, J. et al. Predicting drug metabolism: experiment and/or computation? *Nat. Rev. Drug Discov.* **14**, 387–404 (2015).
22. Knudsen, T. B. et al. FutureTox II: in vitro data and in silico models for predictive toxicology. *Toxicol. Sci.* **143**, 256–267 (2015).
23. Wetmore, B. A. Quantitative in vitro-to-in vivo extrapolation in a high-throughput environment. *Toxicology* **332**, 94–101 (2015).
24. Gorelick, F. S. & Lerch, M. M. Do animal models of acute pancreatitis reproduce human disease?. *Cell Mol. Gastroenterol. Hepatol.* **4**, 251–262 (2017).
25. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148 (2016).
26. Rao, M. et al. Novel computational approach to predict off-target interactions for small molecules. *Front. Big Data* **2**, 25 (2019).
27. Bittker, J. A. & Ross, N. T. (Eds) *High Throughput Screening Methods: Evolution and Refinement* (Royal Society of Chemistry, 2017).
28. Papadatos, G., Gaulton, A., Hersey, A. & Overington, J. P. Activity, assay and target data curation and quality in the ChEMBL database. *J. Comp. Aided Mol. Des.* **29**, 885–896 (2015).
29. Tang, J. et al. Drug target commons: a community effort to build a consensus knowledge base for drug-target interactions. *Cell Chem. Biol.* **25**, 224–229 (2018).
30. Mazzolari, A. et al. Prediction of UGT-mediated metabolism using the manually curated MetaQSAR database. *ACS Med. Chem. Lett.* **10**, 633–638 (2019).
31. Santos, R. et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
32. Lin, Y. et al. Drug target ontology to classify and integrate drug discovery data. *J. Biomed. Sem.* **8**, 50 (2017).
33. Halpern, Y., Choi, Y., Horng, S. & Sontag, D. Using anchors to estimate clinical state without labeled data. *AMIA Annu. Symp. Proc.* **2014**, 606–615 (2014).
34. Wilkinson, M. D. et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
35. Food and Drug Administration. Computerized systems used in clinical investigations (FDA, 2007).
36. Rattan, A. K. Data integrity: history, issues, and remediation of issues. *PDA J. Pharm. Sci. Technol.* **72**, 105–116 (2018).
37. Shockley, K. R. Quantitative high-throughput screening data analysis: challenges and recent advances. *Drug Discov. Today* **20**, 296–300 (2015).
38. Mpindi, J. P. et al. Impact of normalization methods on high-throughput screening data with high hit rates and drug testing with dose–response data. *Bioinformatics* **31**, 3815–3821 (2015).
39. Kalliokoski, T., Kramer, C., Vulpetti, A. & Gedeck, P. Comparability of mixed IC_{50} data – a statistical analysis. *PLOS ONE* **8**, e61007 (2013).
40. Polit, D. F. & Beck, C. T. *Nursing Research: Generating and Assessing Evidence for Nursing Practice* (Wolters Kluwer, 2012).
41. Casale, F. P., Dalca, A. V., Saglietti, L., Listgarten, L. & Fusi, M. Gaussian process prior variational autoencoders. in *Adv. Neural Inf. Process Syst.* (NIPS, 2018).
42. Goldberg, Y. A primer on neural network models for natural language processing. *J. Artif. Intell. Res.* **57**, 345–420 (2016).
43. Peck, M., Moffat, A., Latham, B. & Badrick, T. Review of diagnostic error in anatomical pathology and the role and value of second opinions in error prevention. *J. Clin. Pathol.* **71**, 995–1000 (2018).
44. Lindsey, R. et al. Deep neural network improves fracture detection by clinicians. *Proc. Natl Acad. Sci. USA* **115**, 11591–11596 (2018).
45. Miller, D. D. & Brown, E. W. Artificial intelligence in medical practice: the question to the answer? *Am. J. Med.* **131**, 129–133 (2018).
46. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
47. Gao, Z., Wang, L., Zhou, L. & Zhang, J. HEp-2 cell image classification with deep convolutional neural networks. *IEEE J. Biomed. Health Inf.* **21**, 416–428 (2017).
48. Liu, D., Cheng, B., Wang, Z., Zhang, H. & Huang, T. S. Enhance visual recognition under adverse conditions via deep networks. *IEEE Trans. Image Process.* **28**, 4401–4412 (2019).
49. Reker, D. & Brown, J. B. Selection of informative examples in chemogenomic datasets. *Methods Mol. Biol.* **1825**, 369–410 (2018).
50. Korotcov, A., Tkachenko, V., Russo, D. P. & Ekins, S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. *Mol. Pharm.* **14**, 4462–4475 (2017).
51. Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and challenges in phenotypic drug discovery: an industry perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
52. Aliper, A. et al. Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.* **13**, 2524–2530 (2016).
53. Schneider, G. & Schneider, P. Macromolecular target prediction by self-organizing feature maps. *Expert Opin. Drug Discov.* **12**, 271–277 (2017).

54. Gaulton, A. et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **45**, D945–D954 (2017).
55. Kim, S. et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **47**, D1102–D1109 (2019).
56. Jagadeesh, K. A., Wu, D. J., Birgmeier, J. A., Boneh, D. & Bejerano, G. Deriving genomic diagnoses without revealing patient genomes. *Science* **357**, 692–695 (2017).
57. Mayr, L. M. & Bojanic, D. Novel trends in high-throughput screening. *Curr. Opin. Pharmacol.* **9**, 580–588 (2009).
58. Gong, Z. et al. Compound libraries: recent advances and their applications in drug discovery. *Curr. Drug Discov. Technol.* **14**, 216–228 (2017).
59. Franzini, R. M. & Randolph, C. Chemical space of DNA-encoded libraries. *J. Med. Chem.* **59**, 6629–6644 (2016).
60. Favalli, N., Bassi, G., Scheuermann, J. & Neri, D. DNA-encoded chemical libraries - achievements and remaining challenges. *FEBS Lett.* **592**, 2168–2180 (2018).
61. Lucas, X., Grüning, B. A., Bleher, S. & Günther, S. The purchasable chemical space: a detailed picture. *J. Chem. Inf. Model.* **55**, 915–924 (2015).
62. Lyu, J. et al. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
63. Reymond, J. L. et al. Chemical space as a source for new drugs. *Med. Chem. Commun.* **1**, 30–38 (2010).
64. Drew, K. L., Baiman, H., Khwaounjoo, P., Yu, B. & Reynisson, J. Size estimation of chemical space: how big is it? *J. Pharm. Pharmacol.* **64**, 490–495 (2012).
65. Follmann, M. et al. An approach towards enhancement of a screening library: the Next Generation Library Initiative (NGLI) at Bayer – against all odds? *Drug Discov. Today* **24**, 668–672 (2019).
66. Richter, L. Topliss batchwise schemes reviewed in the era of open data reveal significant differences between enzymes and membrane receptors. *J. Chem. Inf. Model.* **57**, 2575–2583 (2017).
67. Paul, S. M. et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).
68. Satyanarayanajois, S. D. & Hill, R. A. Medicinal chemistry for 2020. *Future Med. Chem.* **3**, 1765–1786 (2011).
69. Lusher, S. J., McGuire, R., van Schaik, R. C., Nicholson, C. D. & de Vlieg, J. Data-driven medicinal chemistry in the era of big data. *Drug Discov. Today* **19**, 859–868 (2014).
70. Shatsky, M., Shulman-Peleg, A., Nussinov, R. & Wolfson, H. J. The multiple common point set problem and its application to molecule binding pattern detection. *J. Comput. Biol.* **13**, 407–428 (2006).
71. Wolber, G., Seidel, T., Bendix, F. & Langer, T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today* **13**, 23–29 (2008).
72. Schneider, G. & Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **4**, 649–663 (2005).
73. Schneider, G. & Clark, D. E. Automated de novo drug design – "are we nearly there yet?". *Angew. Chem. Int. Ed.* **58**, 10792–10803 (2019).
74. Schneider, G. Generative models for artificially-intelligent molecular design. *Mol. Inf.* **37**, 1880131 (2018).
75. Merk, D., Friedrich, L., Grisoni, F. & Schneider, G. De novo design of bioactive small molecules by artificial intelligence. *Mol. Inf.* **37**, 1700153 (2018).
76. Merk, D., Grisoni, F., Friedrich, L. & Schneider, G. Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators. *Commun. Chem.* **1**, 68 (2018).
77. Yang, Y., Adelstein, S. J. & Kassis, A. I. Target discovery from data mining approaches. *Drug Discov. Today* **14**, 147–154 (2009).
78. Frigault, M. M. & Barrett, J. C. Is target validation all we need? *Curr. Opin. Pharmacol.* **17**, 81–86 (2014).
79. Fisher, J. & Henzinger, T. A. Executable cell biology. *Nat. Biotechnol.* **25**, 1239–1249 (2007).
80. Moignard, V. et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33**, 269–276 (2015).
81. Silverbush, D. et al. Cell-specific computational modeling of the PIM pathway in acute myeloid leukemia. *Cancer Res.* **77**, 827–838 (2017).
82. Miettinen K. *Nonlinear Multiobjective Optimization* (Springer, 1999).
83. Lambrinidis, G. & Tsantili-Kakoulidou, A. Challenges with multi-objective QSAR in drug discovery. *Expert Opin. Drug Discov.* **13**, 851–859 (2018).
84. Nicolaou, C. A. & Brown, N. Multi-objective optimization methods in drug design. *Drug Discov. Today Technol.* **10**, e427–e435 (2013).
85. Nicolotti, O. et al. Strategies of multi-objective optimization in drug discovery and development. *Expert Opin. Drug Discov.* **6**, 871–884 (2011).
86. Ekins, S., Honeycutt, J. D. & Metz, J. T. Evolving molecules using multi-objective optimization: applying to ADME/Tox. *Drug Discov. Today* **15**, 451–460 (2010).
87. Kutchukian, P. S. & Shakhnovich, E. I. De novo design: balancing novelty and confined chemical space. *Expert Opin. Drug Discov.* **5**, 789–812 (2010).
88. Grisoni, F., Merk, D., Friedrich, L. & Schneider, G. Design of natural-product-inspired multi-target ligands by machine learning. *ChemMedChem* **14**, 1129–1134 (2019).
89. Wong, W. W. & Burkowski, F. J. A constructive approach for discovering new drug leads: using a kernel methodology for the inverse-QSAR problem. *J. Cheminform.* **1**, 4 (2009).
90. Miyao, T., Kaneko, H. & Funatsu, K. Inverse QSPR/QSAR analysis for chemical structure generation (from y to x). *J. Chem. Inf. Model.* **56**, 286–299 (2016).
91. Blaschke, T., Olivecrona, M., Engkvist, O., Bajorath, J. & Chen, H. Application of generative autoencoder in de novo molecular design. *Mol. Inf.* **37**, 1700123 (2018).
92. Snell, J., Swersky, K. & Zemel, R. Prototypical networks for few-shot learning. *Neural Inf. Process. Syst.* **31** (2017).
93. Altae-Tran, H., Ramsundar, B., Pappu, A. S. & Pande, V. Low data drug discovery with one-shot learning. *ACS Cent. Sci.* **3**, 283–293 (2017).
94. Baskin, I. I. Is one-shot learning a viable option in drug discovery? *Expert Opin. Drug Discov.* **14**, 601–603 (2019).
95. Duvenaud, D. K. et al. Convolutional networks on graphs for learning molecular fingerprints. *Proc. Adv. Neural Inf. Process. Syst.* **28**, 2215–2223 (2015).
96. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. *Proc. Mach. Learn. Res.* **70**, 1263–1272 (2017).
97. Jin, W., Barzilay, R. & Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. *Proc. Mach. Learn. Res.* **80**, 2323–2332 (2018).
98. Yuan, X., He, P., Zhu, Q. & Li, X. Adversarial examples: attacks and defenses for deep learning. *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 2805–2824 (2019).
99. Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. On calibration of modern neural networks. *Proc. Mach. Learn. Res.* **70**, 1321–1330 (2017).
100. Kuleshov, V., Fenner, N. & Ermon, S. Accurate uncertainties for deep learning using calibrated regression. *Proc. Mach. Learn. Res.* **80**, 2796–2804 (2018).
101. Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv. Neural Inf. Process. Syst.* **30**, 6402–6413 (2017).
102. Brookes, D. H., Park, H. & Listgarten, J. Conditioning by adaptive sampling for robust design. *Proc. Int. Conf. Mach. Learn.* **97**, 773–782 (2019).
103. Gillet, V. J. Designing combinatorial libraries optimized on multiple objectives. *Methods Mol. Biol.* **275**, 335–354 (2004).
104. Shim, V. A., Tan, K. C., Chia, J. Y. & Al Mamun, A. Multi-objective optimization with estimation of distribution algorithm in a noisy environment. *Evol. Comput.* **21**, 149–177 (2013).
105. Kramer, O. *Self-Adaptive Heuristic for Evolutionary Computation* (Springer-Verlag, 2008).
106. Hansen, N. The CMA evolution strategy: a tutorial. Preprint at *arXiv* https://arxiv.org/abs/1604.00772 (2016).
107. Ollivier, Y., Arnold, L., Auger, A. & Hansen, N. Information-geometric optimization algorithms: a unifying picture via invariance principles. *J. Mach. Lern. Res.* **18**, 1–65 (2017).
108. Brookes, D. H., Busia, A., Fannjiang, C., Murphy K. & Listgarten, J. A view of estimation of distribution algorithms through the lens of expectation-maximization. Preprint at *arXiv* https://arxiv.org/abs/1905.10474 (2019).
109. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
110. Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015).
111. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. Preprint at *arXiv* https://arxiv.org/abs/1312.6114 (2014).
112. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. Distributed representations of words and phrases and their compositionality. Preprint at *arXiv* https://arxiv.org/abs/1310.4546 (2013).
113. Schneider, G. & Wrede, P. Artificial neural networks for computer-based molecular design. *Prog. Biophys. Mol. Biol.* **70**, 175–222 (1998).
114. Schneider, G. Neural networks are useful tools for drug design. *Neural Netw.* **13**, 15–16 (2000).
115. Zupan, J. & Gasteiger, J. *Neural Networks for Chemists* (VCH, 1993).
116. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
117. 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).
118. Hopkins, A. L., Keserü, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nat. Rev. Drug Discov.* **13**, 105–121 (2014).
119. Cavalluzzi, M. M., Mangiatordi, G. F., Nicolotti, O. & Lentini, G. Ligand efficiency metrics in drug discovery: the pros and cons from a practical perspective. *Expert Opin. Drug Discov.* **12**, 1087–1104 (2017).
120. Meanwell, N. A. Improving drug design: an update on recent applications of efficiency metrics, strategies for replacing problematic elements, and compounds in nontraditional drug space. *Chem. Res. Toxicol.* **29**, 564–616 (2016).
121. Kenny, P. W., Leitão., A. & Montanari, C. A. Ligand efficiency metrics considered harmful. *J. Comput. Aided Mol. Des.* **28**, 699–710 (2014).
122. Plowright, A. T. et al. Hypothesis driven drug design: improving quality and effectiveness of the design-make-test-analyse cycle. *Drug Discov. Today* **17**, 56–62 (2012).
123. Cumming, J. G., Davis, A. M., Muresan, S., Haeberlein, M. & Chen, H. Chemical predictive modelling to improve compound quality. *Nat. Rev. Drug Discov.* **12**, 948–962 (2013).
124. Harrison, S. et al. Extending 'predict first' to the design-make-test cycle in small-molecule drug discovery. *Future Med. Chem.* **9**, 533–536 (2017).
125. Andersson, S. et al. Making medicinal chemistry more effective – application of Lean Sigma to improve processes, speed and quality. *Drug Discov. Today* **14**, 598–604 (2009).
126. Blakemore, D. C. et al. Organic synthesis provides opportunities to transform drug discovery. *Nat. Chem.* **10**, 383–394 (2018).
127. Schneider, G. Automating drug discovery. *Nat. Rev. Drug Discov.* **17**, 97–113 (2018).
128. Reker, D., Schneider, P. & Schneider, G. Multi-objective active machine learning rapidly improves structure-activity models and reveals new protein-protein interaction inhibitors. *Chem. Sci.* **7**, 3919–3927 (2016).
129. Reker, D., Schneider, P., Schneider, G. & Brown, J. B. Active learning for computational chemogenomics. *Future Med. Chem.* **9**, 381–402 (2017).
130. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
131. Nettekoven, M. & Thomas, A. W. Accelerating drug discovery by integrative implementation of laboratory automation in the work flow. *Curr. Med. Chem.* **9**, 2179–2190 (2002).
132. Selekman, J. A. et al. High-throughput automation in chemical process development. *Annu. Rev. Chem. Biomol. Eng.* **8**, 525–547 (2017).
133. King, R. D. et al. Make way for robot scientists. *Science* **325**, 945 (2009).
134. Dimitrov, T., Kreisbeck, C., Becker, J. S., Aspuru-Guzik, A. & Saikin, S. K. Autonomous molecular design: then and now. *ACS Appl. Mater. Interfaces* **11**, 28 (2019).
135. Jordan, A. M. & Roughley, S. D. Drug discovery chemistry: a primer for the non-specialist. *Drug Discov. Today* **14**, 731–744 (2009).
136. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54**, 3451–3479 (2011).
137. Boström, J. & Brown, D. G. Stuck in a rut with old chemistry. *Drug Discov. Today* **21**, 701–703 (2016).
138. Boström, J., Brown, D. G., Young, R. J. & Keserü, G. M. Expanding the medicinal chemistry synthetic toolbox. *Nat. Rev. Drug Discov.* **17**, 709–727 (2018).
139. Segall, M. D. Multi-parameter optimization: identifying high quality compounds with a balance of properties. *Curr. Pharm. Des.* **18**, 1292–1310 (2012).

140. Scott, J. S. & Waring, M. J. Practical application of ligand efficiency metrics in lead optimization. *Bioorg. Med. Chem.* **26**, 3006–3015 (2018).
141. Griffen, E., Leach, A. G., Robb, G. R. & Warner, D. J. Matched molecular pairs as a medicinal chemistry tool. *J. Med. Chem.* **54**, 7739–7750 (2011).
142. King, R. D. et al. The automation of science. *Science* **324**, 85–89 (2009).
143. Hessler, G. & Baringhaus, K. H. Artificial intelligence in drug design. *Molecules* **23**, 2520 (2018).
144. Szymkuć, S. et al. Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).
145. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
146. Coley, C. W., Green, W. H. & Jensen, K. F. Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).
147. Bédard, A. C. et al. Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018).
148. Rohall, S. L., Pancost-Heidebrecht, M., Shirley, B., Bacon, D. & Tarselli, M. A. Recommendations for chemists: a case study. in *Proc. 12th ACM Conf. Recom. Syst*. 347–351 (ACM, 2018).
149. Zhavoronkov, A. et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040 (2019).
150. Canning, P. et al. Structural mechanisms determining inhibition of the collagen receptor DDR1 by selective and multi-targeted type II kinase inhibitors. *J. Mol. Biol.* **426**, 2457–2470 (2014).
151. Thrun, S. & Pratt, L. (eds). *Learning to Learn* (Springer, 2012).
152. Gupta, A. et al. Generative recurrent networks for de novo drug design. *Mol. Inf.* **37**, 1700111 (2018).
153. Bruns, D., Merk, D., Kumar, K. S., Baumgartner, M. & Schneider, G. Synthetic activators of cell migration designed by constructive machine learning. *ChemistryOpen* **8**, 1303–1308 (2019).
154. Sieroka, N., Otto, V. I. & Folkers, G. Critical thinking in education and research – why and how? *Angew. Chem. Int. Ed.* **57**, 16574–16575 (2018).
155. Kut, E., Sieroka, N., Folkers, G., & Otto, V. I. A new course fosters critical thinking on pharmaceutical sciences at ETH Zurich. *ChemMedChem News* https://onlinelibrary.wiley.com/page/journal/18607187/homepage/archive (2018).
156. Azzaoui, K. et al. Scientific competency questions as the basis for semantically enriched open pharmacological space development. *Drug Discov. Today* **18**, 843–852 (2013).
157. Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).
158. Goldberg, K. Robots and the return to collaborative intelligence. *Nat. Mach. Intell.* **1**, 2–4 (2019).
159. Heuer, L. AI could threaten pharmaceutical patents. *Nature* **558**, 519 (2018).
160. King, R. D. & Courtney, P. Dilemma over AI and drug patenting already under debate. *Nature* **560**, 307 (2018).
161. Olley, D. (ed.) Artificial intelligence: how knowledge is created, transferred, and used (Elsevier, 2019).
162. Brown, S. P., Muchmore, S. W. & Hajduk, P. J. Healthy skepticism: assessing realistic model performance. *Drug Discov. Today* **14**, 420–427 (2009).
163. McDonagh, J. L., Nath, N., De Ferrari, L., van Mourik, T. & Mitchell, J. B. O. Uniting cheminformatics and chemical theory to predict the intrinsic aqueous solubility of crystalline druglike molecules. *J. Chem. Inf. Model.* **54**, 844–856 (2014).
164. Hartenfeller, M. et al. DOGS: Reaction-driven de novo design of bioactive compounds. *PLOS Comput. Biol.* **8**, e1002380 (2012).
165. 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).
166. 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).
167. Yuan, W. et al. Chemical space mimicry for drug discovery. *J. Chem. Inf. Model.* **57**, 875–882 (2017).
168. 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).
169. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).
170. Perron, Q. et al. Deep learning for ligand-based de novo design in lead optimization: a real life case study. Presented at the XXV EFMC International Symposium on Medicinal Chemistry (2018).
171. Rodriguez, T. et al. Multidimensional de novo design reveals 5-HT_{2B} receptor-selective ligands. *Angew. Chem. Int. Ed.* **54**, 1551–1555 (2015).
172. Reutlinger, M., Rodrigues, T., Schneider, P. & Schneider, G. Multi-objective molecular de novo design by adaptive fragment prioritization. *Angew. Chem. Int. Ed.* **53**, 4244–4248 (2014).
173. Parry, D. M. Closing the loop: developing an integrated design, make, and test platform for discovery. *ACS Med. Chem. Lett.* **10**, 848–856 (2019).
174. Esch, E. W., Bahinski, A. & Huh, D. Organs–on–chips at the frontiers of drug discovery. *Nat. Rev. Drug Discov.* **14**, 248–260 (2015).
175. Eglen, R. M. & Randle, D. H. Drug discovery goes three-dimensional: goodbye to flat high-throughput screening? *Assay. Drug Dev. Technol.* **13**, 262–265 (2015).
176. Jones, L. H. & Bunnage, M. E. Applications of chemogenomic library screening in drug discovery. *Nat. Rev. Drug Discov.* **16**, 285–296 (2017).
177. Trobe, M. & Burke, M. D. The molecular industrial revolution: automated synthesis of small molecules. *Angew. Chem. Int. Ed.* **57**, 4192–4214 (2018).
178. Baranczak, A. et al. Integrated platform for expedited synthesis–purification–testing of small molecule libraries. *ACS Med. Chem. Lett.* **8**, 461–465 (2017).
179. Vasudevan, A., Bogdan, A. R., Koolman, H. F., Wang, Y. & Djuric, S. W. Enabling chemistry technologies and parallel synthesis–accelerators of drug discovery programmes. *Prog. Med. Chem.* **56**, 1–35 (2017).
180. Pant, S. M. et al. Design, synthesis, and testing of potent, selective hepsin inhibitors via application of an automated closed-loop opptimization platform. *J. Med. Chem.* **61**, 4335–4347 (2018).
181. Gesmundo, N. J. et al. Nanoscale synthesis and affinity ranking. *Nature* **557**, 228–232 (2018).

Acknowledgements
This article is based on a meeting involving a group of international experts from diverse scientific backgrounds and institutions convened in San Francisco in December 2018 for a workshop organized by the RETHINK think-and-do tank of ETH Zurich to rethink drug design with artificial intelligence. Figure 1 was created and contributed by Jack Burgess, who also acted as a visual scribe during the workshop. Jürg Brunnschweiler and the ETH Global team are thanked for excellent organizational support. This research was financially supported by the RETHINK initiative of ETH Zurich.

Author contributions
All authors contributed equally to the content and approved the final version of the manuscript.

Competing interests
G.S. and P.S. declare a potential financial conflict of interest in their role as life science industry consultants and cofounders of inSili.com GmbH, Zurich. The remaining authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

RELATED LINKS
Click2Drug: https://www.click2drug.org/
The ATOM Consortium: https://www.atomscience.org
The Innovative Medicines Initiative: https://www.imi.europa.eu/
The SALT Knowledge Share Consortium: https://www.medchemica.com/the-salt-knowledge-share-consortium/