Rethinking drug design in the artificial intelligence era

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] & Gisbert Schneider[1,*]

[1] ETH Zurich, RETHINK, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland
[2] Relay Therapeutics, 399 Binney St, Cambridge, MA 02139, USA
[3] Sanofi-Aventis Deutschland GmbH, Integrated Drug Discovery, Industriepark Höchst, D-65926 Frankfurt am Main, Germany
[4] Institute for Philosophy, University of Bremen, PF 33 04 40, 28334 Bremen, Germany
[5] University of California, Berkeley, Electrical Engineering and Computer Science, 387 Soda Hall, MC 1776, Berkeley, CA 94720, 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, United Kingdom
[9] Novartis Institutes for BioMedical Research, 5300 Chiron Way, Emeryville, CA 94608, USA
[10] Novartis Institutes for BioMedical Research, 181 Massachusetts Ave, Cambridge MA 02139, USA
[11] ATOM Consortium, 499 Illinois St., San Francisco, CA 94158, USA
[12] Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Str. 65, 88397 Biberach an der Riss, Germany
[13] Healthy Life Consultants, 201 E. 28th St. #4L, New York, NY 10016, USA
[14] Bloomberg Intelligence, Bloomberg LP, 3 Queen Victoria, London EC4A 4TQ, UK
[15] ETH Zurich, Department of Architecture, Stefano-Francini-Platz 5, 8093 Zurich, Switzerland
[16] Genentech, 1 DNA Way, South San Francisco, CA 94080, USA
[17] The University of Tokyo, Department of Chemical System Engineering, Hongo 7−3−1, Bunkyo-ku, Tokyo, 113-8656, Japan
[18] Nara Institute of Science and Technology, Data Science Center, 8916-5 Takayama-cho, Ikoma, Nara 630-0192, Japan
[19] Arup Global Foresight, 560 Mission Street, Suite 700, San Francisco, CA 94105, USA

Artificial intelligence (AI) tools are increasingly being applied in drug discovery. Whilst some protagonists point to vast opportunities potentially offered by such tools, others remain skeptical, waiting for a clear impact to be shown in drug discovery projects. The truth is probably somewhere between these extremes, but 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' for small-molecule drug discovery with AI and 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[3,4,5,6]. However, 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 fueled 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[11,12,13]. Selection of the most appropriate experimentally validated HTS hits for follow-up is critical for 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 physicochemical characteristics that could be important in drug pharmacokinetics and safety. Consequently, medicinal chemists typically face challenging multi-objective optimization 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 prioritize alternatives. This represents at least a partial transfer of decision power to a 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 (Box 1). 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 prohibitive-costly full-deck HTS.

To be successful in the long run, drug design with AI has to provide solutions to several questions, which can be encompassed in five "grand challenges" (Figure 1):

1. Generating and obtaining appropriate datasets,

2. Generating new hypotheses,

3. Optimizing in a multi-objective fashion,

4. Reducing cycle times, and

5. Changing the research culture and creating appropriate mindset.

With the aim to shed light on these challenges and approaches to address them, a select group of international experts from diverse scientific backgrounds and institutions convened in San Francisco in December 2018 for a workshop organized by ETH Zurich, to rethink drug design with AI. This Perspective summarizes these discussions and presents the main conclusions drawn.

Challenge 1: Generating and 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]. In this section, we will detail some of the factors that should be considered when evaluating data to be used in generating predictive models for drug design. We will highlight applications of AI in generating higher quality or more relevant datasets to support predictive modeling, and we will suggest ways in which the scientific community can collaborate to increase understanding of what are good data attributes to support AI-augmented drug design.

One of the most important factors when evaluating data for predictive modeling is whether those data were collected with the ultimate endpoint 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 is an important problem, 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, there is currently a limited amount of *in vivo* toxicology data available. As a result, 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 that are 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 are constructed based on the results of previously designed high-throughput experiments. This approach may offer advantages owing to the larger data set 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 a vastly lower sensitivity than measurements of focused measurement of pre-defined off-target area[25,F26]. 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 data-set relevance, mentioned above, it is important for us 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 ignored[30]. Although automated annotations exist to a certain extent (*e.g.*, machines generating and analyzing data typically add meta-data such as time, date etc.), 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. AI techniques based on probabilistic latent variable models are already helping to extract meaning from multi-origin data sets, 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 U.S Food and Drug Administration (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, just as it has been in statistics. In order to accurately leverage our data, we must have a clear understanding of both the accuracy and the precision of our measurements (Box 2)[37,38].

Another source of difficulty and uncertainty when using experimental data to build predictive models is the accidental misreporting of data. This can take the form of simple typos in reported values, gene identifiers, units, or a number of other parameters which 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 may be another area where AI is 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 the same ideas to identify data which may have been incorrectly reported. Of course, not all outliers are mistakes; they may instead highlight an alternate mechanism of action which 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 can 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]; 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 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 if it is possible and desirable to find lower-dimensional representations with which to do the associations, rather than to 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 toward automating this process[41].

The domains with the most success to date with AI have been in imaging and natural language processing (NLP)[42]. These data differ substantially from those typically found in drug discovery: Whether "a user followed a hyperlink" or "there is a stop sign in an image" tend to be more easily labeled accurately. 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[43,44,45]. The given drug discovery project provides context for the data. It 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 training data, no doubt the reason for their being the first big successes of deep learning. Labeled data sets of this size simply do not exist in drug discovery owing to the more difficult process of obtaining the data. The amount of data required to generate a truly predictive model, is still an open question. Many of the current, 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 labeled data to capture the diversity of the input and obtain a model that generalizes well[46,47]. The number and nature of molecules required to build a predictive model for drug design is still undetermined[48,49].

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 K_D or a cellular EC_{50} value[50]. 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[51,52].

Over the last two decades, we have seen the appearance of a number of public databases containing millions of biological assay results, such as ChEMBL[53] and PubChem[54], which can provide input for machine learning models predicting a variety of biological activities or physical properties for drug-like molecules[55]. While 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 that data closely, and 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 (*e.g.*, phenyl to pyridyl) and the differences in activity (*e.g.*, three-fold reduction in hERG activity)[57]. Another approach would be the EU Innovative Medicines Initiative (IMI) for privacy-preserving federated machine learning[58], or the ATOM Consortium[59], a group of industrial and academic scientists working to streamline the drug discovery process, to provide a means of testing while maintaining confidentiality.

Challenge 2: Generating new hypotheses

Only a tiny fraction of the druglike 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 screening deck (10^6–10^7 compounds)[60,61], even with the advent of DNA-encoded libraries which can typically test 10^7–10^{10} compounds[62,63], 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[64]. 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[65]. Still, this number is dwarfed by the estimated cardinality of druglike chemical space in the order of 10^{18}–10^{200} molecules[66,67], depending on the constraints set[68].

The size of druglike 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[69], which designates a special series of aromatic ring substituents intended to systematically explore design opportunities by optimizing substitution patterns, and 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[70,71].

Chemical design can be considered pattern matching[72,73,74], and indeed, computer-based *de novo* design methods have been explored as idea generators to support drug design since the 1990s[75,76]. Today, however, generative AI offers a fresh approach to *de novo* drug design by providing a statistical framework for decision making[77]. In contrast to earlier molecular design engines that employed a set of explicit chemical transformations (*e.g.*, virtual reaction schemes based on reaction SMILES) and assembly rules (*e.g.*, 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 3).

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[78,79]. The main advantages of these models over previous *de novo* methods are (i) speed of execution (NCEs can be generated on-the-fly to allow for interactive modeling), (ii) rapid re-training or fine-tuning on the project at hand, (iii) scalability by providing access to a virtually infinite chemical space without the need for explicit compound library enumeration, (iv) software availability, and (v) 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[80,81]. One of the base limitations of 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 needs to provide answers flexibly, as drug discovery knowledge develops.

So-called "mechanistic" models, on the other hand, are able to address both challenges by capturing behaviors at different levels of abstraction (genetic, molecular, cellular, etc.) and providing explanations for how these behaviors evolve and interact[82,83,84]. 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.

Challenge 3: 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 multi-objective optimization (MOO; also referred to variously as multi-parameter optimization (MPO), multi-objective programming, vector optimization, multicriteria optimization, multi-attribute optimization or Pareto optimization)[85,86,87]. 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 artificial intelligence[88,89,90,91].

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 which balance the desired properties (Box 4). 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; then those solutions are handed over to human experts to sift through with deep, implicit knowledge and intuition – at least for the time being.

The main challenges and opportunities in MOO for drug design at present include:

(i)   leveraging the power of modern day predictive modelling approaches, including neural networks, for desired properties and uncertainties over them;

(ii)   correspondingly, to robustly deal with the emerging pitfalls associated with predictive modelling methods like deep learning[92];

(iii)   leveraging improvements in solving MOO and related problems;

(iv)   operating in the face of limited data with data-hungry models; and

(v)   adopting these methods by their intended consumers, including medicinal chemists and drug designers, which may include modifying MOO approaches.

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 of interest, such as biological activity in the form of pK_D or $\log(EC_{50})$ values. These models are what will help tell us, in the context of MOO, if 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")[93,94]. Similar to the *de novo* design task discussed above (Challenge 2), certain generative AI models seem to be well-suited to properly address this problem[95]. 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 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 decades-old benchmarks on tasks in vision and audio. Despite their popularity though, a substantial improvement by virtue of their use has not reliably been seen to date in other fields. The reasons for this are not yet known, but one can speculate that they may include some of the following: In vision and audio, one can readily acquire massive amounts of labelled data that are vital for success with current deep neural networks. In general, the fields of biology and chemistry are not 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 fewer data, sometimes referred to as "few shot" learning[96,97,98]. Another potential reason deep neural networks have seen major advances only in the areas of audio and vision is that their development in recent years was tailored to these fields. In vision, the predictive features are typically pixels of an image, and the labels may be classes of objects, such as a stop sign – very specific structure is present, such as piece-wise smoothness of the pixels, invariance of class label to the position of a cat in an image, and so forth. Many advances in so-called general machine learning are leveraging such vision-task-specific structure but are then immediately applied in other domains often without much thought as to the appropriateness (*e.g.*, convolutional filters). Deducing analogous structure in chemistry and biology or ways to encode it are in their infancy, and they are fundamentally more challenging endeavors 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 scaleable and well-suited to the domain (*e.g.*, generate only synthetically accessible molecules)[99,100,101].

Several troubling phenomena have emerged from state-of-the-art deep neural network models in fields such as image processing. For example, 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[102]. While drug design is unlikely to face deliberate adversaries, the existence of these adversarial examples betrays the fact that deep neural network models are full of pathologies that we may be blind to and can only mitigate once we become aware of them. A second troubling phenomenon of deep neural networks is that they are notoriously over-confident in themselves[103]. That is, when a neural network says it thinks an image is 95% likely to be a cat, in reality, we know that it is typically a much 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[104,105]. 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 from "comfort" zones, all the while being completely 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[106]. We emphasize that we are not suggesting that one should not use 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 upon which to conduct an 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[107]. These methods have largely been replaced by more principled methods in which the diversification operations have been replaced by sampling from statistical models within a more coherent statistical framework[108]. These latter approaches, typically in the class of Estimation of Distribution Algorithms (EDA)[109] such as Covariance Matrix Adaptation Evolutionary Strategies (CMA-ES)[110] now overlap with and are synergistic with machine learning methodology such as Information Geometric Optimization and the Expectation-Maximization algorithm (*e.g.*, references 111, 112). Moreover, these methods in turn have connections to machine learning for robotics, namely, reinforcement learning[113,114]. 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. The generative model takes the place of the mutations and cross-over 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 to, through a particular statistical formalism, tune the parameters such that only desired molecules can be sampled from it. So 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[115]. 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 Natural Language Processing (NLP), it has been found that converting sentences, inherently composed of discrete symbols, into real-valued vectors, can provide benefit in downstream tasks[116]. Similar arguments and efforts have been made in molecular design already in the 1990s[7,117,118,119] and have recently been rediscovered in the context of deep learning[120,121]. Again, this is an area which can benefit from advances in related application domains, but it is likely to benefit by adaptation rather than direct application, such as new ways to represent molecular structures and properties for better efficacy in the domain of drug discovery. In starting this section, we described MOO as being used with a "human-in-the-loop" to help decide among different optimal solutions. An area where AI and machine learning could assist would be to enable better encoding of the decisions being made by the human, so that they can feed into the automated part of the system.

Challenge 4: 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[122,123,124]. While these methods have their merits (but also their controversy[125]), 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-analyze (DMTA) cycle[126]. 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 provide greater understanding. This knowledge is then analysed and translated into the development of the design hypothesis for the next cycle (Figure 1).

A number of groups have reported methods to improve the effectiveness of the DMTA cycle (Box 5); for example, with greater use of predicted data[127], improved data analysis tools[128] and increased effectiveness in compound synthesis[129,130], thereby shortening the timelines for the delivery of the critical data to address a hypothesis[131,132,133,134]. Certain aspects of AI potentially offer an alternative to HTS. Instead of compiling and relying on a large screening compound library, small numbers (<1000) of *de novo* generated compounds are 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 four, or even eight weeks to complete. As a result, the required number of cycles to deliver a clinical candidate will take a significant period of time. Whilst *Design* and *Analysis* can be fast, and *Testing* (in particular *in vitro* data generation including potency, selectivity and ADMET 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. If this phase can be shortened, it can significantly reduce the iteration time of a DMTA cycle. Laboratory automation has and will continue to play a decisive role in this context[135,136,137,138]. Fast compound synthesis using robust reactions in batch or flow with automated analytics and purification will help speed-up the *Make* phase. The choice of reactions for automation should primarily focus on the ones frequently employed by medicinal chemists[139,140], so the chemist will be free to perform the more challenging synthetic steps and conceive new chemical reactions[141,142].

Due to the variety of design hypotheses and the different times necessary to synthesize molecules and profile them in various critical 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 multi-parameter optimization[143]. 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 which need to be captured and analyzed in molecular design make it difficult for medicinal chemists and scientists to consistently incorporate a full understanding of thousands of data points and trends and to constantly be on top of and spot all learning 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 like "rules of thumb", efficiency metrics[144], model systems like log$P$ or log$D$, or matched molecular pairs[145]. 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 and decision making and thus, bring about more effective and productive discovery teams[146].

AI provides a range of opportunities to increase the effectiveness of the DMTA cycle including integrating and analyzing 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[147]. 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' to shorten the *Make* phase[148,149,150,151].

Research scientists would greatly benefit from receiving pre-digested and tailored pieces of information or recommendations in the right moment, format and context[152]. Such hints (where to look for what) would reduce the need to plough through raw data. Scientists could instead focus on the critical 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 conveniently when necessary.

In a recent example of the use of computational deep learning to virtually generate of set of molecules and prioritise, based on a range of criteria, a small set of compounds for rapid synthesis and testing was described for the kinase discoidin domain receptor 1 (DDR1)[153]. A comprehensive machine learning model was built using a broad range of data sets, and subsequently trained on certain data including the known DDR1 inhibitor and common kinase inhibitor data sets. The design stage using reinforcement learning was then 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 IC$_{50}$ values <20 nM in an enzymatic kinase assay of DDR1 inhibition. One aspect of this study not highlighted in the paper is the fact that the most prominent molecule (Compound 1) only differs from known DDR1 inhibitors, including the marketed drug Ponatinib (IC$_{50}$ = 9 nM),[154] 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, which will still be longer and the larger costs in drug discovery will come during the most likely required optimization steps. Hence, whilst one of the rare prospective examples, it still has a way to go to achieve and show significant cycle time and cost reductions in the generation of a novel clinical 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. A demonstration of the applicability of AI for drug design from scratch in low-data situations is still to be shown. In this context, the established concept of transfer learning[155] renders "few shot" methods for generative molecular design[156], with first pioneering examples demonstrating its practical applicability[78,79,157]. Notwithstanding, assessing the actual impact on hit and lead generation requires further validation of transfer learning methodology in different low-data situations and projects. An additional challenge will come with biological targets with less associated knowledge and it will be exciting to see this area develop further and be applied in these less, or non-, trodden areas.

In the future, with more accurate predictive models across multiple parameters, the whole DMTA cycle could become virtual, with more intermittent synthesis of molecules to ensure progress remains on track. With more integrated analyses, hypothesis generation becomes faster, the proposed molecules better address the multi-objective challenges of drug design and they could even come with proposed synthesis routes. In such a setting, the AI components would support chemists and project team members to help them be more effective. Ultimately, this could help significantly decrease the time of DMTA cycles and reduce the hit discovery to clinical candidate delivery timeline. In addition, if the AI component is useful for some projects, fewer resources would be needed for them compared with today, and the additional resources could be focused on projects which are less amenable to AI methods.

Challenge 5: Research culture and the appropriate mind-set

Leaving the technical questions around AI aside, possibly the greatest challenge for pharmaceutical discovery with machine intelligence lies in the appropriate mindset and drug discovery "culture". The success of AI in drug design will be decided by the ethos, attitude, and willingness of all stakeholders to apply these computational models and utilize their results[158].

By "culture" we refer to a community of subjects who share certain practices – usually speaking a common language (including formalisms and algorithms) or using a common method or exemplar (*i.e.*, a typical instance, illustration or exemplification). Accordingly, when claiming that a "change of culture" is asked for, what we mean is the demand for a change in the language, methods and exemplars. This understanding of culture is very broad, and it is important in the present context since the relevant "cultures" to be addressed in the context of drug design range from scientists to business people. These cultures might be subdivided further into "subcultures" of, for instance, single lab groups using a very specific and unique method. The challenge, therefore, is to facilitate mutual appreciation and discourse. This process is to be achieved by first recognizing and

acknowledging the different experiences people have and then adapting and developing common terminologies and exemplars. To achieve this vision and see its impact, we need a comprehensive and precise language to create a new narrative that focuses on a clear role for each of the stakeholders in the drug design process and how they interact.

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 own doings to colleagues also from other fields of research and to a wider audience[159,160]. Next, to develop such a common narrative, it is first important to recognize the general drivers that will induce the application of AI in drug design and, then, to describe them on an abstract level before again turning to the individual cultures and subcultures. The three interrelated drivers in this sense are personalization, automation, and materialization. Any change of culture of AI in drug design will be based on a change in at least one of these drivers, and will likely lead to a dynamic reshaping of the whole interaction between the three components. 21st century healthcare is challenged to deliver personalized medicine that is bespoke for every patient or patient cohort, driven by an understanding of genetics and drug response in humans. The envisioned feedback cycle between patients and cure should be rapid, and the cure itself should be readily attainable. Such challenges can be met by a high degree of *Automation* of virtual and physical processes – safely enabled by trustworthy AI and carefully curated by responsible humans. Data needs to be available and secure, algorithms robust, and platform pipelines modular, so that the resulting complexity can be handled efficiently and knowledge and technologies can be shared across individual applications and stakeholders. Personalization and automation in turn have to be complemented by a respective strategy of *Materialization* where drugs must be synthesized in appropriate quantities as needed, and delivered on-time and in the right location in order to achieve highly targeted and effective treatments. Due to the anticipated synergies between personalization, automation and materialization, novel technological developments in any area can trigger surprising developments in the other areas with potentially game-changing impacts on 21st century healthcare.

A key opportunity to leverage AI is to assist the drug discovery scientists, medicinal chemists and design teams in their work: to make their processes more productive while concomitantly increasing the quality of the data and the acceptance of AI. The paradigm to approach this goal is to identify areas where AI could support and augment chemists and drug designers, not replace them. For example, one of the limits 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 the electronic laboratory notebooks could leverage AI to facilitate the process of capturing, annotating, and curating the data, the chemists would be able to focus more time on the innovation and the human insight necessary to make effective drugs. In turn, there would be better training sets to improve the output of the AI components.

Furthermore, if the AI tools could be leveraged to sift through past years of data from other medicinal chemistry programs and connect it to the current program, additional directions to go in drug design space might be highlighted for further analysis by project teams[161]. To make such loops useful, however, drug discovery scientists must accept the value of the AI output and leverage it along with their own experience. At the same time, AI systems will have to be able to interact and cooperate with human experts 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 behavior 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[162,163]. This view would also positively contribute to the current debate on the patentability of AI-generated drugs[164,165].

Finally, turning to the broader patient population, there are three approaches that might leverage its positive engagement and confidence in AI systems: an egoistic, an altruistic, and a playful one. The egoistic approach starts from individual concerns about one's health and from other individual interests and curiosities. The hope for bespoke medicines plays an important role here as well as the idea to learn more about oneself (*e.g.*, gene sequencing as a means to know more about one's ancestry). The altruistic approach appeals to people's solidarity with their fellow human beings and woos people with the idea of becoming part of the scientific endeavor. "Data sharing" and "donating data" are the keywords in this context[166,167]. Finally, there is a playful approach, which starts with the use of smartphones and social media rather than being about healthcare *per se*. The examination of oneself is initiated, as it were, by playing around with some apps, by starting to monitor one's heartbeat or by sharing one's jogging trail[168,169]. Of course, in practice people will usually be guided by a mixture of these three approaches.

Pharma has started to cautiously adopt instances of AI-related technology, specifically various machine learning methods, through partnerships and collaboration, though there is an understandable reluctance to place all bets on AI-based drug discovery. A curious but cautious approach is advisable, given the regulated nature of drug development and the shift from traditional science to science that includes and values the contribution of AI. Hence, a long-term vision of AI in drug design is needed to leverage a proactive, constructive and collaborative machine intelligence that could help reduce barriers among the multiple research cultures and subcultures in the ecosystem of creating new medicines to improve healthcare outcomes.

Acknowledgements

Author Contributions

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

Conflict of Interest Statement

The authors declare the following 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.

Figure 1
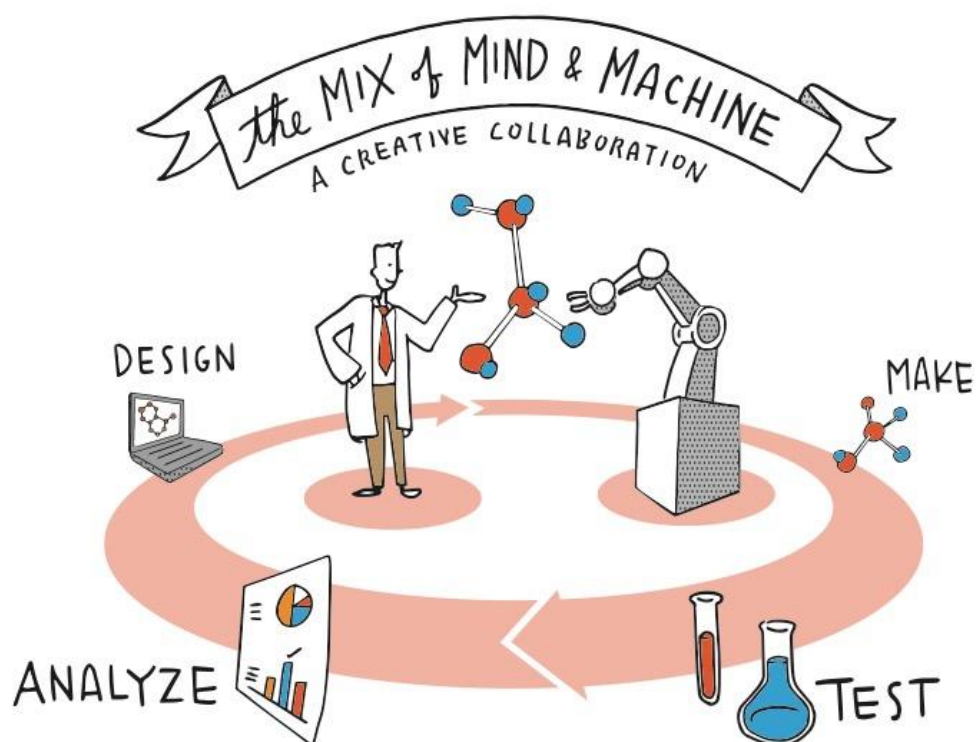
Figure 1. Artificial intelligence and laboratory automation augment human decision making, chemical synthesis and biological testing in the design-make-test-analyze cycle. It is anticipated that this collaborative intelligence emerging from the combination of "mind and machine" will enable better decision making in drug discovery. Credit: Jack Burgess.

Box 1. Glossary of terms.

Adaptive algorithm. An adaptive algorithm implements a problem solving heuristic that changes its behavior 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)[170].

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" (i.e., 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 (e.g., 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 analyze large amounts of natural language data, *e.g.* scientific literature. Deep statistical machine learning models achieve state-of-the-art results in many natural language tasks, for example in language modeling and parsing. NLP can also be used for chemical language analysis and de novo design.

Neural network. A particular type of function approximator wherein functions that predict discrete classes (classifiers) or real-values (regression models) do so by composing a series of (typically non-linear) functions, each one converting the previous layer's outputs to a new "space". These models have been around for decades but came to prominence in the 1990s when the combination of access to large data sets, 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; *i.e.* speaking a common language (including formalisms and algorithms) or sharing typical instances, illustrations or exemplifications (including molecular structures).

Some of the definitions are adapted from ref. 171.

Box 2. 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[172] describe a method which 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 (DLS)[173]. 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 1000 cycles of the simulation above and assume a two-fold experimental error (*i.e.*, 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 five-fold, the maximum possible value of r decreases to 0.96. If the experimental error is 10-fold, 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.

Box 3. Chemical hypothesis generation with constructive machine learning.

Constructive machine learning aims to create examples from its learned domain which 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 molecules. In contrast to rule-based AI models for *de novo* drug design, these generative methods do not learn explicit chemical transformations (*e.g.*, forward- and/or retrosynthetic reaction schemes)[174] 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[148,149,150,175].

Examples of models used to generate *de novo* structures include deep (a) "backtransformation" (*i.e.*, inversion) of supervised feedforward networks[176], (b) supervised and semi-supervised variational autoencoders, (c) recurrent neural networks[177,178], and (d) generative adversarial networks (GAN). Models based on reinforcement methods often play a large role in computational design algorithms[179].
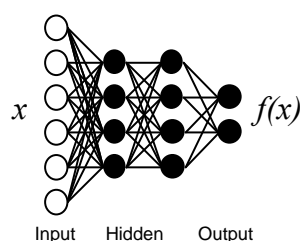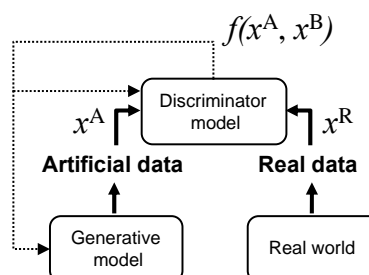

a)   Feedforward Net
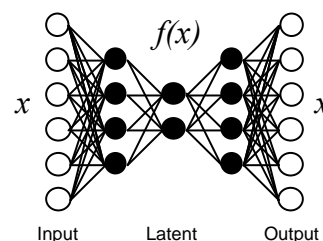
b)   Variational autoencoder
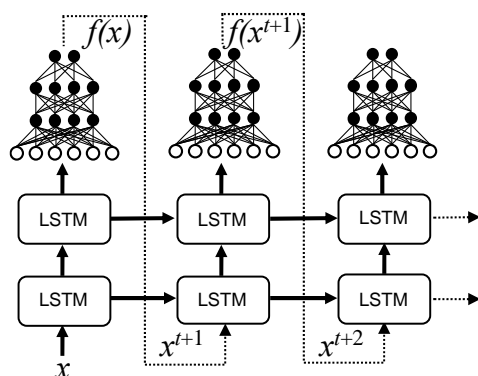
c) Recurrent LSTM network

d) Generative Adversarial Network

All of these neural network models transform a numerical input *x* to a functional f(*x*) as 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. The class of *feedforward nets* (a) constitutes universal function approximators, where each layer of compute steps (circles) performs feature extraction from the training data. A deep network is designated by several such feature extraction layers. Depending of the type of layers, whether the data have corresponding lables, and whether there are cycles of information allowed in the network architecture, one obtains various network types, *e.g. autoencoders* (b). *Recurrent networks* (*e.g.*, long short-term memory (LSTM) models, (c)) perform pattern recognition by sequence

analysis. Here, the input (*e.g.*, 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 GAN (d), one network generates candidate molecules and the other evaluates them, *e.g.*, by comparing the virtual chemical structures to real examples for reinforcement learning.
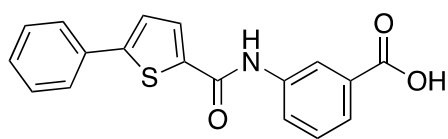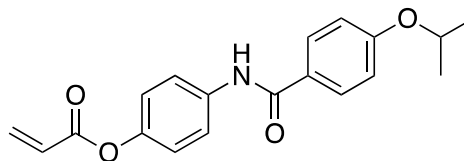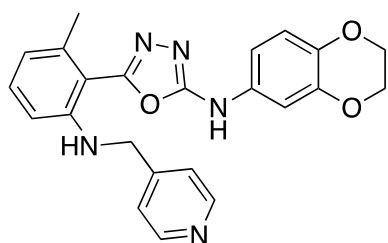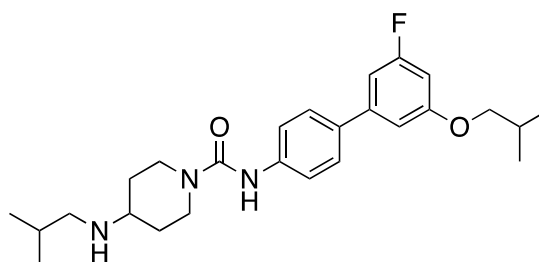
First successful syntheses of *de novo* generated compounds (**1**–**4**) corroborate the practical applicability of generative molecular design to drug discovery. Pioneering prospective designs, RXRγ agonists **1** (EC$_{50}$ = 0.06±0.02 µM) and **2** (EC$_{50}$ = 19.1±0.1 µM), were generated with a deep LSTM network[157]. This model was trained with SMILES strings of bioactive compounds from ChEMBL[28] and fine-tuned on nuclear hormone receptor targets using transfer learning. VEGFR-2 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[151]. Partial 5-HT$_{2B}$ antagonist **4** resulted from *de novo* structure generation with a neural network for virtual forward synthesis[154].

Box 4. Machine learning models applied to multi-objective lead optimization challenges in drug discovery projects.

Multi-objective optimization (MOO) of a chemical series aims to identify compounds meeting simultaneously all the desirable characteristics of a potential pre-clinical drug candidate. Ideally, this 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 ADME/Tox properties that are generally indicative of drug-like properties include the following: Aqueous solubility, metabolic stability, cellular permeability, action by cellular transporters (*e.g.*, P-gp), drug-drug interactions, blood-brain barrier penetration, cardiotoxicity (*e.g.*, 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 (undisclosed target)[180]. With a clear definition of metrics for 11 objectives (phenotypic activity, selectivity against 5-HT$_{2A}$, 5-HT$_{2B}$, α1, and D1 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; further, 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 data set and used to score new virtual structures generated by a proprietary SMILES-based molecular generator built on deep learning LSTM generative models. The molecular generator used QSAR scoring on generated structures to converge iteratively to structures maximizing QSAR scores for all 11 objectives. Finally, 150 virtual structures meeting all objectives were proposed, all of those having rare or absent substructures within the initial data set. 20 compounds were selected for synthesis of which 11 were successfully synthesized. Upon assay, three compounds fulfilled all 11 objectified criteria. An example of a compound (**1**), that satisfied nine of the 11 objectives, was part of the QSAR training set that led to the design, synthesis and assay of compound (**2**), that satisfied all 11 objectives. The [1,2,3]triazolo[1,5-a]pyridine moiety appeared only six times in the initial 880 structure data set.
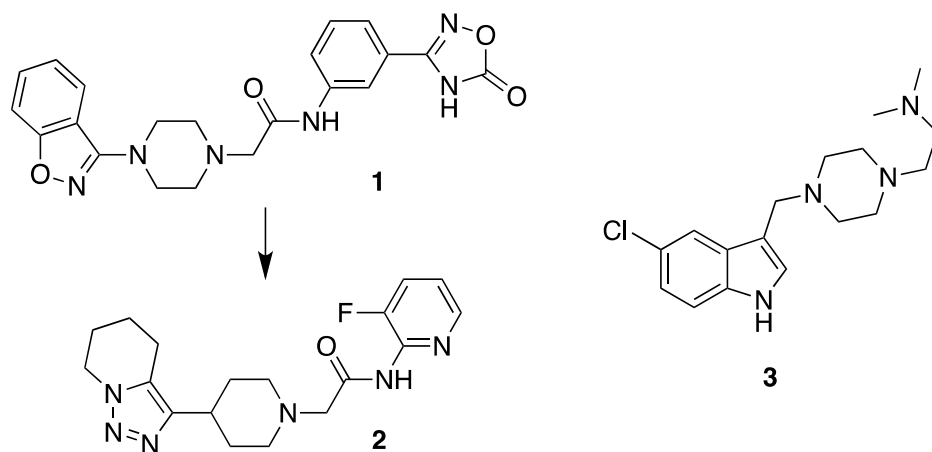
In an earlier study[181], researchers from ETH Zurich and Novartis used an adaptive molecular construction method (MAntA, Molecular Ant Algorithm[182]) to automatically generate small, synthetically accessible *de novo* designs that are selective antagonists of 5HT$_{2B}$ receptor without binding to other 5HT 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 (5HT$_{2B}$ K_i = 251±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 MOO show that there are several ways in which a machine intelligence can support the discovery of novel lead structures with multiple desired properties.

Box 5. Reducing cycle times by automation.

The potential value of automated design-make-test-analyze (DMTA) cycles in drug discovery is substantial[126,183]. Benefits of automation include[131]:

(i)   Diminished measurement errors and reduced material consumption;

(ii)   Fast feedback loops for AI-based hypothesis generation and optimization;

(iii)   Opportunities to expand the medicinal chemistry synthetic toolbox (*e.g.*, microfuidic-assisted synthesis reactions, reactions under extreme conditions)[134];

(iv)   Rigorous compound prioritization by applying sophisticated cell–based assays (*e.g.*, cells/organs-on-chips)[184,185,186], in an effort to more effectively recapitulate disease biology and thereby improve the likelihood of identifying compounds that show efficacy in humans;

(v)   Molecular optimization towards multiple relevant biochemical and biological endpoints without personal bias.

Implementations of integrated DMTA platforms have already proven their efficiency and applicability to drug design and optimization[187]. 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[188]. 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[189].

Despite the advances of machine learning, chemical reactor and assay technology, as of today, 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[190]. In progressing from the known inhibitor **1** to designed compound **2** in automated DMTA cycles, the IC_{50} value against hepsin was improved from ~1 μM to 33 nM, and the selectivity over urokinase-type plasminogen activator (uPA) 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.
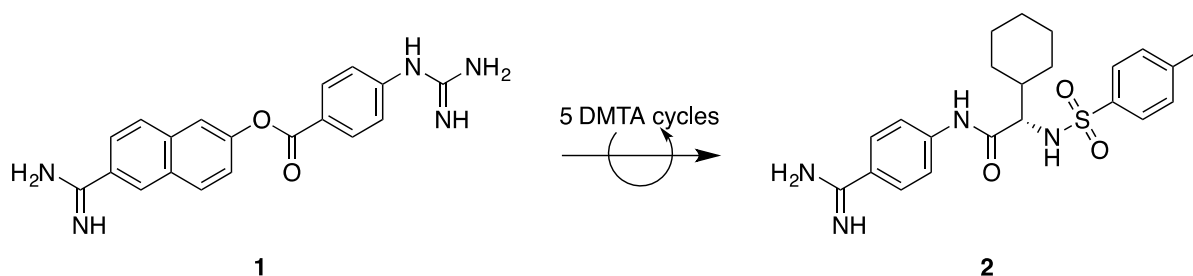
Nanoscale synthesis without the necessity for time-consuming purification in combination with rapid affinity ranking might further reduce DMTA cycle times. In a pioneering study, the coupling of high-throughput nanomole-scale synthesis with a label-free affinity-selection mass spectrometry bioassay facilitated simultaneous optimization of reaction conditions and building blocks as well as protein affinity[191].

References

[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 Disc.* **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**, in press (2019). https://doi.org/10.1021/acs.chemrev.8b00728

[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.*, doi: 10.1002/biot.201800613 (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. *J. Mol. Sci.* **15**, 21136-21154 (2014).

[20] Huang, R. *et al.* Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nature Comm.* **7**, 10425 (2016).

[21] Kirchmair, J. *et al.* Predicting drug metabolism: experiment and/or computation? *Nature Reviews Drug Discovery* **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, London (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 (FDA), *Guidance for Industry: Computerized Systems Used in Clinical Investigations*, U.S. Department of Health and Human Services (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 IC50 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, Philadelphia, PA (2012).

[41] Casale, F. P., Dalca, A. V., Saglietti, L., Listgarten, L. & Fusi, M. Gaussian process prior variational autoencoders. *Neural Information Processing Systems* **32** (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] Gao, Z., Wang, L., Zhou, L. & Zhang, J. HEp-2 cell image classification with deep convolutional neural networks. *IEEE J. Biomed. Health Inform*. **21**, 416–428 (2017).

[47] 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**, doi:10.1109/TIP.2019.2908802 (2019).

[48] Reker, D. & Brown, J. B. Selection of informative examples in chemogenomic datasets. *Methods Mol. Biol*. **1825**, 369–410 (2018).

[49] 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).

[50] 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).

[51] 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).

[52] Schneider, G. & Schneider, P. Macromolecular target prediction by self-organizing feature maps. *Expert Opin. Drug Discov*. **12**, 271–277 (2017).

[53] Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Res*. **45**, D945–D954 (2017).

[54] Kim, S. *et al*. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res*. **47**, D1102–D1109 (2019).

[55] https://www.click2drug.org/

[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] https://www.medchemica.com/the-salt-knowledge-share-consortium/

[58] *The Innovative Medicines Initiative* (IMI) of the European Union and the European Federation for Pharmaceutical Industries and Associations (EFPIA). https://www.imi.europa.eu/

[59] The ATOM Consortium, 499 Illinois Street, San Francisco, CA 94158, USA. www.atomscience.org

[60] Mayr, L. M. & Bojanic, D.Novel trends in high-throughput screening. *Curr. Opin. Pharmacol*. **9**, 580–588 (2009).

[61] Gong, Z. Compound libraries: recent advances and their applications in drug discovery. *Curr. Drug Discov. Technol*. **14**, 216–228 (2017).

[62] Franzini, R. M. & Randolph, C. Chemical space of DNA-encoded libraries. *J. Med. Chem*. **59**, 6629–6644 (2016).

[63] Favalli, N., Bassi, G., Scheuermann, J. & Neri, D. DNA-encoded chemical libraries - achievements and remaining challenges. *FEBS Lett*. **592**, 2168–2180 (2018).

[64] Lucas, X., Grüning, B. A., Bleher, S. & Günther, S. The purchasable chemical space: a detailed picture. *J. Chem. Inf. Model*. **55**, 915–924 (2015).

[65] Lyu, J. *et al*. Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).

[66] Reymond, J. L. *et al*. Chemical space as a source for new drugs. *Med. Chem. Commun*. **1**, 30–38 (2010).

[67] 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).

[68] 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).

[69] 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).

[70] 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).

[71] Atyanarayanajois, S. D. & Hill, R. A. Medicinal chemistry for 2020. *Future Med. Chem.* **3**, 1765–1786 (2011).

[72] 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).

[73] 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).

[74] 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).

[75] Schneider, G. & Fechner, U.Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **4**, 649–663 (2005).

[76] Schneider, G. & Clark, D. E. Automated de novo drug design – "are we nearly there yet?" *Angew. Chem. Int. Ed.* **58**, doi: 10.1002/anie.201814681 (2019).

[77] Schneider, G. Generative models for artificially-intelligent molecular design. *Mol. Inf.* **37**, 1–2 (2018).

[78] Merk, D., Friedrich, L., Grisoni, F. & Schneider, G. De novo design of bioactive small molecules by artificial intelligence. *Mol. Inf.* **37**, 1700153 (2018).

[79] Merk, D., Grisoni, F., Friedrich, L. & Schneider, G. (2018) Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators. *Communications Chemistry* **1**, 68 (2018).

[80] Yang, Y., Adelstein, S. J. & Kassis, A. I. Target discovery from data mining approaches. *Drug Discov. Today* **17**, S16–S23 (2012).

[81] Frigault, M. M. & Barrett, J. C. Is target validation all we need? *Curr. Opin. Pharmacol.* **17**, 81–86 (2014).

[82] Fisher J. & Henzinger T. A. Executable cell biology. *Nat. Biotechnol.* **25**, 1239–1249 (2007).

[83] Moignard, V. *et al*. Decoding the transcriptional program for blood development from whole tissue single cell gene expression measurements. *Nat. Biotechnol.* **33**, 269–276 (2015).

[84] Silverbush, D. *et al*. Cell-specific computational modeling of the PIM pathway in acute myeloid leukemia. *Cancer Res.* **77**, 827–838 (2017).

[85] Miettinen K. *Nonlinear Multiobjective Optimization*. Springer, New York (1999).

[86] Lambrinidis, G. & Tsantili-Kakoulidou, A. Challenges with multi-objective QSAR in drug discovery. *Expert Opin. Drug Discov.* **13**, 851–859 (2018).

[87] Nicolaou, C. A. & Brown, N. Multi-objective optimization methods in drug design. *Drug Discov. Today Technol.* **10**, e427–e435 (2013).

[88] Nicolotti, O. *et al*. Strategies of multi-objective optimization in drug discovery and development. *Expert Opin. Drug Discov.* **6**, 871–884 (2011).

89 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).

90 Kutchukian, P. S. & Shakhnovich, E. I. De novo design: balancing novelty and confined chemical space. *Expert Opin. Drug Discov.* **5**, 789–812 (2010).

91 Grisoni, F., Merk, D., Friedrich L. & Schneider, G. Design of natural-product-inspired multi-target ligands by machine learning. *ChemMedChem* **14**, 1129–1134 (2019).

92 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

93 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).

94 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).

95 Blaschke, T., Olivecrona, M., Engkvist, O., Bajorath, J. & Chen, H. Application of generative autoencoder in de novo molecular design. *Mol. Inf.* **37**, 1700123 (2018).

96 Snell, J., Swersky, K. & Zemel, R. Prototypical networks for few-shot learning. *Neural Information Processing Systems* **31** (2017).

97 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).

98 Baskin, I. I. Is one-shot learning a viable option in drug discovery? *Expert Opin. Drug Discov.* **18**, 1–3 (2019).

99 Duvenaud, D. K. *et al.* Convolutional networks on graphs for learning molecular fingerprints. *Advances in Neural Information Processing Systems* **28**, 2224–2232 (2015).

100 Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. *Proceedings of the 34th International Conference on Machine Learning* **70**, 1263–1272 (2017).

101 Jin, W., Barzilay, R. & Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. *Proceedings of Machine Learning Research* **80**, 2323–2332 (2018).

102 Yuan, X., He, P., Zhu, Q. & Li, X. Adversarial examples: attacks and defenses for deep learning. IEEE Trans. *Neural Netw. Learn. Syst.* doi: 10.1109/TNNLS.2018.2886017 (2019).

103 Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. On calibration of modern neural networks. *Proceedings of Machine Learning Research* **70**, 1321–1330 (2017).

104 Kuleshov, V., Fenner, N. & Ermon, S. Accurate uncertainties for deep learning using calibrated regression. *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80 (2018).

105 Lakshminarayanan, B., Pritzel, A. & Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *31st Conference on Neural Information Processing Systems* (NIPS 2017), Long Beach, CA, USA, NIPS 3205 (2017).

106 Brookes, D. H., Parl, H. & Listgarten, J. Conditioning by adaptive sampling for robust design. *Proceedings 36th Intl. Conf. Mach. Learning Res.* **97**, 773–782 (2019).

107 Gillet, V. J. Designing combinatorial libraries optimized on multiple objectives. *Methods Mol. Biol.* **275**, 335–354 (2004).

108 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).

109 Kramer, O. *Self-Adaptive Heuristic for Evolutionary Computation.* Springer-Verlag, Berlin, Heidelberg (2008).

110 Hansen, N. The CMA evolution strategy: a tutorial. arXiv:1604.00772 (2006).

111 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).

112 Brookes, D. H., Busia, A., Fannkjiang, C., Murphy K. & Listgarten, J. A view of estimation of distribution algorithms through the lens of expectation-maximization. *arXiv*, arXiv:1905.10474 (2019).

113 Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction.* The MIT Press, Cambridge, MA (1998).

114 Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015).

115 Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. *Proceedings of the 2nd International Conference on Learning Representations* (ICLR), Banff, Canada (2014).

116 Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems* (NIPS) **26** (2013).

117 Schneider, G. & Wrede, P. Artificial neural networks for computer-based molecular design. *Prog. Biophys. Mol. Biol.* 70, 175–222 (1998).

118 Schneider G. Neural networks are useful tools for drug design. *Neural Netw.* 13,15–16 (2000).

119 Zupan, J. & Gasteiger, J. *Neural Networks for Chemists.* VCH, Weinheim (1993).

120 Gómez-Bombarelli, R. *et al.* Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).

121 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).

122 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).

123 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).

124 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).

125 Kenny, P. W., Leitão., A. & Montanari, C. A. Ligand efficiency metrics considered harmful. *J. Comput. Aided Mol Des.* **28**, 699–710 (2014).

126 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).

127 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).

128 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).

129 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).

130 Blakemore, D. C. *et al.* Organic synthesis provides opportunities to transform drug discovery. *Nat. Chem.* **10**, 383–394 (2018).

131 Schneider, G. Automating drug discovery. *Nat. Rev. Drug Discov.* **17**, 97–113 (2018).

132 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).

133 Reker, D., Schneider, P., Schneider, G. & Brown, J. B. Active learning for computational chemogenomics. *Future Med. Chem.* **9**, 381–402 (2017).

134 Steiner, S. *et al.* Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, 122–123 (2019).

135 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).

136 Selekman, J. A. *et al.* High-throughput automation in chemical process development. *Annu. Rev. Chem. Biomol. Eng.* **8**, 525–547 (2017).

137 King, R. D. *et al.* Make way for robot scientists. *Science* **325**, 945 (2009).

138 Dimitrov, T., Kreisbeck, C., Becker, J. S., Aspuru-Guzik, A. & Saikin, S. K. Autonomous molecular design: Then and now. *ACS Appl. Mater. Interfaces* **11** (2019) doi: 10.1021/acsami.9b01226.

139 Jordan, A. M. & Roughley, S. D. Drug discovery chemistry: a primer for the non-specialist. *Drug Discov. Today* **14**, 731–744 (2009).

140 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).

141 Boström, J. & Brown, D. G. Stuck in a rut with old chemistry. *Drug Discov. Today* **21**, 701–703 (2016).

142 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).

143 Segall, M. D. Multi-parameter optimization: identifying high quality compounds with a balance of properties. *Curr. Pharm. Des.* **18**, 1292–1310 (2012).

144 Scott, J. S. & Waring, M. J. Practical application of ligand efficiency metrics in lead optimization. *Bioorg. Med. Chem.* **26**, 3006–3015 (2018).

145 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).

146 King, R. D. *et al.* The automation of science. *Science* **324**, 85–89 (2009).

147 Hessler, G. & Baringhaus, K. H. Artificial intelligence in drug design. *Molecules* **23**, 2520 (2018).

148 Szymkuć, S. *et al.* Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem. Int. Ed.* **55**, 5904-5937 (2016).

149 Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).

150 Coley, C. W., Green, W. H. & Jensen, K. F. Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).

151 Bédard, A. C. *et al.* Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018).

152 Rohall, S. L., Pancost-Heidebrecht, M., Shirley, B., Bacon, D. & Tarselli, M. A. Recommendations for chemists: a case study. *RecSys '18*, Vancouver, Britsh Columbia, Canada (2018).

153 Zhavoronkov, A., *et al.* Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040 (2019).

[154] 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).

[155] Thrun, S. & Pratt, L. (eds). *Learning to Learn*, Springer, New York, USA (2012).

[156] Gupta, A. *et al.* Generative recurrent networks for de novo drug design. *Mol. Inf.* **37**, 1700111 (2018).

[157] Bruns, D., Merk, D., Kumar K. S., Baumgartner, M. & Schneider, G. Synthetic activators of cell migration designed by constructive machine learning. *ChemistryOpen* **8**, accepted (2019).

[158] Fleming, N. How artificial intelligence is changing drug discovery. *Nature* **557**, S55–S57 (2018).

[159] Sieroka, N., Otto, V. I., & Folkers, G. Critical thinking in education and research – Why and how? *Angew. Chem. Int. Ed.* **57**, 16574–16575 (2018).

[160] Kut, E., Sieroka, N., Folkers, G., & Otto, V.I. A new course fosters critical thinking on pharmaceutical sciences at ETH Zurich. *ChemMedChem News* (2018). DOI: https://onlinelibrary.wiley.com/page/journal/18607187/homepage/archive

[161] Azzaoui K, *et al.* Scientific competency questions as the basis for semantically enriched open pharmacological space development. *Drug Discov. Today* **18**, 843–852 (2013).

[162] Topol, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **25**, 44–56 (2019).

[163] Goldberg, K. Robots and the return to collaborative intelligence. *Nat. Mach. Intell.* **1**, 2–4 (2019).

[164] Heuer, L. AI could threaten pharmaceutical patents. *Nature* **558**, 519 (2018).

[165] King, R. D. & Courtney, P. Dilemma over AI and drug patenting already under debate. *Nature* **2018** 560, 307 (2018).

[166] Whiddett, R., Hunter, I., Engelbrecht, J. & Handy, J. Patients' attitudes towards sharing their health information. *Int. J. Med. Inform.* **75**, 530–541 (2006).

[167] Mursaleen, L. R., Stamford, J. A., Jones, D. A., Windle, R. & Isaacs. T. Attitudes towards data collection, ownership and sharing among patients with Parkinson's disease. *J. Parkinsons Dis.* **7**, 523–531 (2017).

[168] Lymberis A. Smart personalised health systems: from gadgets to better health. *Stud. Health Technol. Inform.* **117**, 1–5 (2005).

[169] Boulos, M. N. & Yang, S. P. Exergames for health and fitness: the roles of GPS and geosocial apps. *Int. J. Health Geogr.* **12**, 18 (2013).

[170] Dick, S. Artificial intelligence. *Harvard Data Science Review* **1**, DOI: https://doi.org/10.1162/99608f92.92fe150c (2019).

[171] Olley, D. (ed) *Artificial Intelligence: How Knowledge is Created, Transferred, and Used.* Amsterdam, Elsevier (2019). https://www.elsevier.com/research-intelligence/resource-library/ai-report?utm_source=AI-EC

[172] Brown, S. P., Muchmore, S. W & Hajduk, P. J. Healthy skepticism: Assessing realistic model performance. *Drug Discov. Today* **14**, 420–427 (2009).

[173] 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).

[174] Hartenfeller, M. *et al.* DOGS: Reaction-driven de novo design of bioactive compounds. *PLoS Comput. Biol.* **8**, e1002380 (2012).

175 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).

176 Button, A., Merk, D., Hiss, J. A. & Schneider, G. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nature Mach. Intell.* **1**, 307–315 (2019).

177 Yuan, W. *et. al.* Chemical space mimicry for drug discovery. *J. Chem. Inf. Model.* **57**, 875–882 (2017).

178 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).

179 Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).

180 Perron, Q. *et al.* Deep learning for ligand-based de novo design in lead optimization: A real life case study. Presented at XXV *EFMC International Symposium on Medicinal Chemistry*, Ljubljana, Slovenia (2018).

181 Rodriguez, T. *et al.* Multidimensional de novo design reveals 5-HT$_{2B}$ receptor-selective ligands. *Angew. Chem. Int. Ed.* **54**, 1551–1555 (2015).

182 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).

183 Parry, D. M. Closing the loop: developing an integrated design, make, and test platform for discovery. *ACS Med. Chem. Lett.* **10**, 848–856 (2019).

184 Esch, E. W., Bahinski, A. & Huh, D. Organs–on–chips at the frontiers of drug discovery. *Nat. Rev. Drug Discov.* **14**, 248–260 (2015).

185 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).

186 Jones, L. H. & Bunnage, M. E. Applications of chemogenomic library screening in drug discovery. *Nat. Rev. Drug Discov.* **16**, 285–296 (2017).

187 Trobe, M. & Burke, M. D. The molecular industrial revolution: automated synthesis of small molecules. *Angew. Chem. Int. Ed.* **57**, 4192–4214 (2018).

188 Baranczak, A. *et al.* Integrated platform for expedited synthesis–purification–testing of small molecule libraries. *ACS Med. Chem. Lett.* **8**, 461–465 (2017).

189 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).

190 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).

191 Gesmundo, N. J. *et al.* Nanoscale synthesis and affinity ranking. *Nature* **557**, 228–232 (2018).