Drug Discovery Today • Volume xxx, Number xx • xxxx 2021 REVIEWS

De novo molecular design and generative models

Joshua Meyers *, Benedek Fabian, Nathan Brown

BenevolentAI, 4–8 Maple Street, London W1T 5HD, UK

Molecular design strategies are integral to therapeutic progress in drug discovery. Computational approaches for *de novo* molecular design have been developed over the past three decades and, recently, thanks in part to advances in machine learning (ML) and artificial intelligence (AI), the drug discovery field has gained practical experience. Here, we review these learnings and present *de novo* approaches according to the coarseness of their molecular representation: that is, whether molecular design is modeled on an atom-based, fragment-based, or reaction-based paradigm. Furthermore, we emphasize the value of strong benchmarks, describe the main challenges to using these methods in practice, and provide a viewpoint on further opportunities for exploration and challenges to be tackled in the upcoming years.

Keywords: De novo design; Generative models; Generative chemistry; Molecular representation; Artificial intelligence; Molecular design; Automated design; Fragment-based; Atom-based; Reaction-based

Introduction

De novo molecular design is the process of automatically proposing novel chemical structures that optimally satisfy a desired molecular profile. Often in drug discovery, the objective molecular profile is to provoke a desirable biological response, while maintaining acceptable pharmacokinetic properties. Recently, *de novo* design has also been referred to as generative chemistry, arising from the increased popularity of generative models in AI.

Traditionally, virtual screening (VS) is undertaken to identify molecules likely to exhibit desirable experimental outcomes. A key difference, compared with *de novo* design, is the source of the molecules considered: where structures are known *a priori* in VS, in *de novo* design we seek to generate the structures to be evaluated.

Chemical space, the expanse spanning all possible molecules, is vast. Although enumerated VS libraries are becoming enormous by drug discovery standards, with many containing upward of a billion molecules, these libraries correspond to a miniscule proportion of chemical space. Additionally, when considering compound libraries of such size, evaluation methods might necessarily sacrifice predictive validity.[1] By generating compounds in a directed manner using *de novo* design, computational practitioners hope to traverse chemical space more effectively, reaching optimal chemical solutions while considering fewer molecules than allowed by brute-force screening of large chemical libraries (Fig. 1). Furthermore, there are likely many acceptable regions of chemical space for a given objective and, to this end, molecular design methods are tasked with balancing the exploration of global solutions with the exploitation of local minima.

De novo design has a rich history in chemoinformatics and has received recent attention as ML methodologies continue to open new possibilities for navigating and sampling large search spaces.[2,3] In this review, we consider de novo design methodologies from the perspective of the coarseness of molecular representation; specifically, we distinguish atom-based, fragment-based, and reaction-based approaches for generating novel structures. We begin by reviewing methods to measure the comparative

* *Corresponding author.* Meyers, J. (joshua.meyers@benevolent.ai)

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

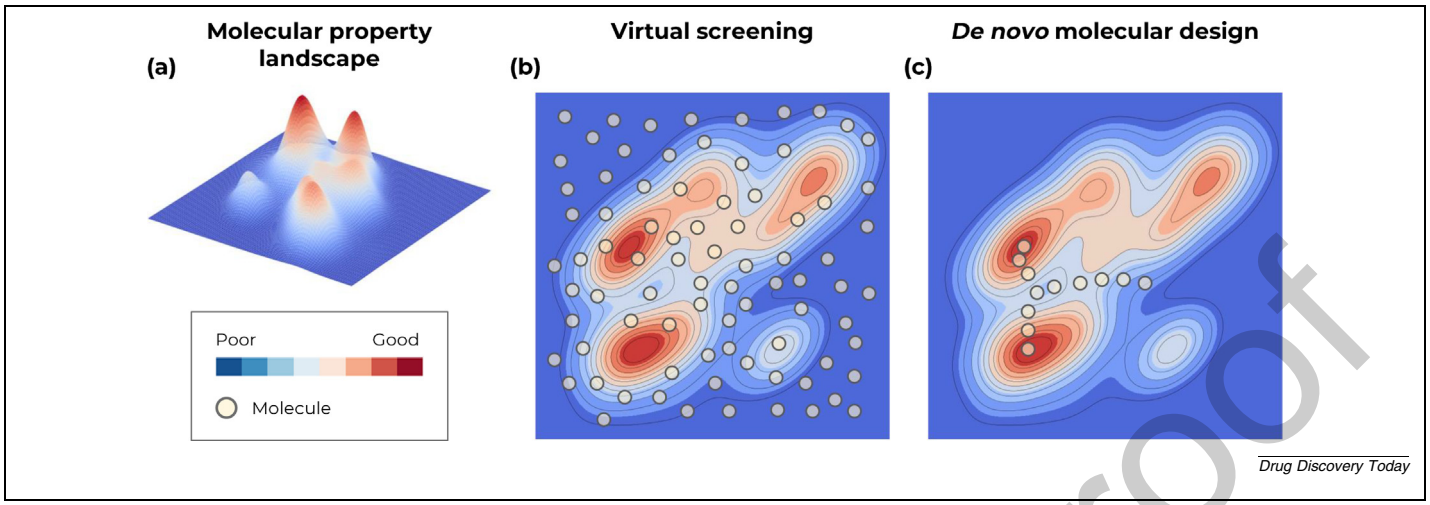

FIGURE 1

Title. A schematic of chemical space, with colors signifying the optimality of molecules residing in that locality given an arbitrary objective molecular property profile **(a)**. The conceptual difference is illustrated between **(b)** virtual screening of large pre-existing chemical libraries and **(c)** traversal of chemical space by an effective *de novo* molecular design program. *De novo* design results in fewer molecules being considered to reach optimal molecular structures through a more effective traversal of chemical space.

aptitude of *de novo* design approaches before discussing both established methods and new frontiers in generative chemistry. Finally, we appraise the successes of the technology and highlight potential hurdles yet to be leapt on the path toward *de novo* molecular design realizing its full potential in drug discovery.

Molecular design

Assessing methods for de novo molecular design

To measure the progress of automated methods for chemical structure generation consistently, the establishment of standard benchmark suites is vital.

De novo design methods are often evaluated by their performance on standalone toy tasks, such as maximizing the quantitative estimate of drug-likeness (QED)[4] or a penalized form of the calculated octanol–water partition coefficient (ClogP).[5] Although such objectives are trivial to calculate and showcase the ability of optimizers to generate molecules toward an objective, they fail to capture the complexities of real-world drug discovery. By contrast, another approach to evaluate *de novo* design methodologies is to demonstrate their use experimentally. For example, Firth *et al.*[6] synthesized and tested novel inhibitors of cyclin-dependent kinase 2 (CDK2). Although compelling, this represents an anecdotal evaluation dependent on numerous factors unrelated to the *de novo* design algorithm.

The publication and availability of suites of standard benchmark tasks has standardized evaluation of *de novo* design methods from an *in silico* perspective. The Molecular Sets (MOSES) benchmark includes a set of distribution learning tasks along with measures of molecule validity and uniqueness.[7] The aim of distribution learning tasks is to measure the structural diversity and relevance of proposed compounds by comparing the generated chemical space to known chemical structures; MOSES also considers scaffold and fragment diversity. The GuacaMol benchmark suite includes, in addition to distribution benchmarks, a more applied set of goal-directed tasks, which imitate discrete uses of *de novo* design tools.[8]

Although benchmark suites for *de novo* design are invaluable, many methods are still evaluated on bespoke standalone tasks or even subsets of available benchmark suites, making direct comparison between methods challenging. Although clearly the usage of benchmarks should be encouraged, we recognize the need to improve benchmarks to include tasks that better capture the requirements of intended use cases. We return to this point later.

Molecular representation

Computational methods for evaluating chemical structures must rely upon a suitable molecular representation, that is, the form in which a molecular structure is seen by a subsequent algorithm. Molecular representation is a broad topic[9]; for example, methods can encode the presence or absence of functional groups, express a molecule as its topological graph, or include 3D information describing bond angles.

Among *de novo* design methods, common molecular representations are text based, such as the simplified molecular input line entry system (SMILES), and graph based where the molecular generator might operate explicitly on the molecular topology. Text-based methods benefit from the huge volume of active research in natural language processing (NLP), whereas graph-based approaches embody a more natural representation of molecular structure. Other influences on choice of representation include whether the molecular representation is discrete (e.g., bitvector), continuous (e.g., vector of floating points), and invertible.[10] Recent reviews of *de novo* design methods[9–11] have focused on molecular representation through the lens of generative model architecture, whereas we focus here on the granularity of molecular representation (Fig. 2) because this translates directly to practical aspects of molecular design.

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

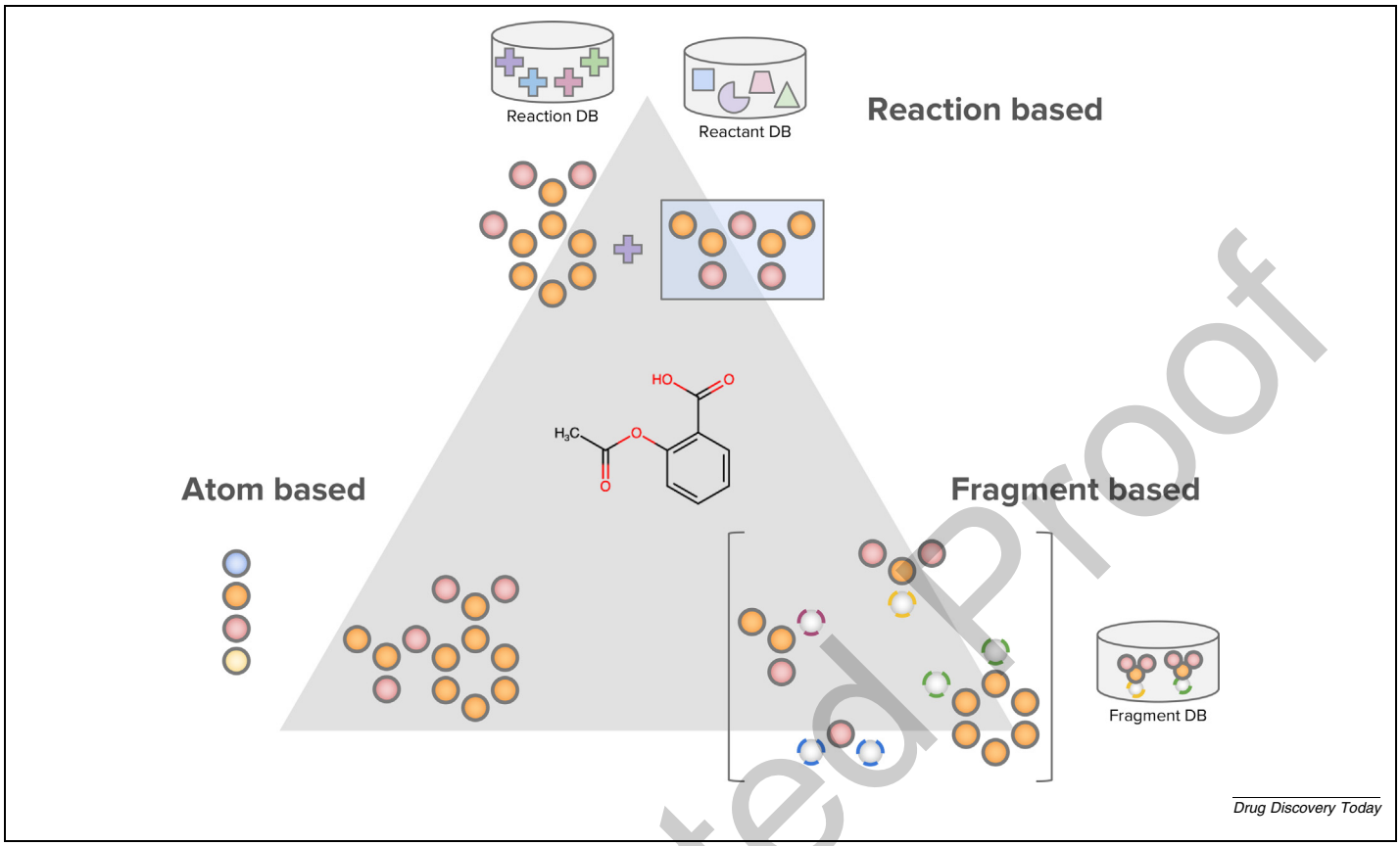

FIGURE 2

Illustration of the continuum between atom-based, fragment-based, and reaction-based molecular representation paradigms, shown for aspirin. The atom-based approach is supported by a vocabulary containing a small number of atoms and bonds. The reaction-based approach is supported by dual sets of reactants and reaction rules. Finally, the fragment-based approach is supported by a fragmentation scheme, and a set of interchangeable fragments; gray atoms denote attachment points annotated with a disconnection type (color).

SMILES has become commonplace as a molecular representation for generative models; however, a shortcoming is that each SMILES string is not a unique description of the molecular graph. A SMILES is constructed by following a linear walk through the molecular structure; thus, different start positions and paths through the molecule produce different SMILES.[12] Canonical SMILES represents a standardized traversal of the molecular graph[13]; however, generative models trained on canonical SMILES can capture nuisance aspects of the SMILES grammar, rather than the underlying molecular structure. Studies have shown a benefit when including non-canonical SMILES during training of generative models.[14–16] Furthermore, adaptations to SMILES have been described that are more suited for use with ML, including DeepSMILES[17] and SELFIES.[18]

A chemical structure can be represented at the atomic level, by encoding each atom and bond in a molecule, or at a coarser representation where functional groups, or substructures and their connectivities remain fixed; for example, a benzene group with a 1,3 substitution pattern can be considered a single group. A further extension is to encode reactions, whereby a target molecule is considered a product of reactants and reaction conditions. In practice, atom-based, fragment-based, and reaction-based approaches have distinct strengths and weaknesses, and many methods blur the boundaries between these classifications.

Gradient-free molecular optimization

Given a molecular representation, optimization algorithms guide generation toward optimal molecules according to a calculable objective function. Metaheuristic ('gradient-free') methods for *de novo* design use population-based stochastic optimization procedures to navigate chemical space, such as evolutionary algorithms[19] or swarm intelligence.[20] Briefly, we highlight exemplar works from the recent literature with respect to the granularity of their chosen molecular representation.

Atom based

An example of an atom-based approach for *de novo* design is the graph-based genetic algorithm (GB-GA),[21] which uses reaction SMARTS to perform mutation and crossover on a pool of candidates while an elitist natural selection procedure ensures that the most optimal molecules are maintained in the population; and ChemGE,[22] which uses grammatical evolution to optimize a population of SMILES following a context-free grammar. GB-GA was included in the GuacaMol benchmark and achieved state-of-the-art performance, although the benchmark authors commented on the 'quality' of the compounds by measuring the number of reactive and unstable groups among proposed molecules. Winter *et al.*[23] described Molecular Swarm Optimization (MSO), a method that uses particle swarm optimization to

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

identify desirable regions in a continuous embedding space before decoding discrete molecular structures. Although the representation used by MSO is learnt, the optimization procedure is a gradient-free approach and achieves state-of-the-art performance on the GuacaMol goal-directed benchmarks. A key issue for population-based approaches is maintaining structural diversity in the population. MolFinder[24] uses a minimum topological distance between molecules in the pool to ensure this, whereas graph-based elite patch illumination (GB-EPI) extends GB-GA to maintain a population of feature-based niches.[25]

Fragment based

Fragment-based approaches constrain the generation of new compounds to comprise known relevant substructures, such as those in the medicinal chemistry literature. Fragmentation schemes deconstruct molecules using either simple rules (e.g., all acyclic single bonds) or retrosynthetically inspired disconnections.[26] A library of fragments each containing one or more atoms can then be used to construct novel molecules. MOARF[6] is a fragment-based *de novo* design method that makes use of a set of retrosynthetic disconnection rules (SynDiR) and an evolutionary algorithm. More recently developed, the CReM framework[27] uses chemically reasonable mutations using a fragmentation scheme adapted from matched molecular pairs, showing comparable performance to MSO on the GuacaMol goal-directed benchmark tasks.

Reaction based

Arguably the most practical strategy for *de novo* design is to perform forward reactions *in silico*. In 2003, Vinkers *et al*. described SYNOPSIS,[28] an approach to apply virtual reactions iteratively to maximize a desired fitness function, demonstrated by synthesizing and testing compounds designed to inhibit HIV reverse transcriptase (HIV-RT). More recently developed, AutoGrow4[29] makes use of a genetic algorithm and a library of reactions derived from robust organic reactions[30] to mutate molecules in a population. A shortcoming of reaction templates is that they match reaction handles irrespective of other reactive groups in a molecule, which can affect the reaction in reality. Ghiandoni *et al*.[31] recently reported a reaction class recommender taht allows filtering of reactions from undesired classes.

Gradient-based molecular optimization

Although population-based metaheuristic approaches for *de novo* design have been shown to be robust at finding optimization minima, deep learning approaches for molecular design have seen widespread adoption over the past 3 years. Gradient-based ML approaches are often pretrained on large corpuses of existing molecular structures and learn to navigate arbitrary property surfaces toward optimal solutions.[2,3]

Several deep learning architectures have been proposed for learning to generate molecule structures, including variational autoencoders (VAEs), generative adversarial networks (GANs), and recurrent neural networks (RNNs). Once trained, generative models allow the user to sample molecules from the learnt chemical space and, when coupled with an optimization process, such as Bayesian optimization (BO) or reinforcement learning (RL), can efficiently identify desirable molecular profiles. We refer the reader to recent reviews[9–11] for more details on the ML processes involved. Here, as for metaheuristic approaches, we highlight exemplar works.

Atom based

Many atom-based generative models make use of SMILES as a molecular representation. Given that SMILES is a text-based representation, generative chemistry methods are able to take advantage of deep learning architectures suited to sequences, such as RNNs. By pretraining on a large corpus of molecular structures, generative models can learn *a priori*, which encapsulates the grammar and syntax of valid SMILES. Whereas early works used transfer learning to bias generation toward the chemical space of interest,[32] it is now commonplace to couple the generation task to an RL algorithm, which learns to navigate the search space to obtain higher rewards (more optimal molecules).[33]

In addition to SMILES-based generative models, there has been much interest in models that directly consider the topology of the molecular graph, where atoms and bonds are regarded as nodes and edges, respectively. By operating on a more natural representation of molecular structure, graph-based models seek to sidestep artificial aspects of the SMILES syntax. GraphVAE[34] and MolGAN[35] are generative graph-based approaches that learn to generate the entire graph adjacency matrix at once. Methods have also been described that learn to generate molecules in a stepwise fashion by modifying the molecular graph iteratively.[36,37] More recently, RL methods have shown promising results in the graph setting.[38,39]

Fragment based

Whereas pretrained atom-based generative models maintain a strong prior toward substructures present in their training data, they are nonetheless able to individually modify each atom in a molecule. Although this flexibility encourages learnt models to be maximally expressive, resulting in wider coverage of chemical space, fragment-based approaches constrain the search space using a coarser molecular representation. An early example of a coarse molecular representation, initially developed to represent patent Markush structures, is the reduced graph.[40] Over 20 years later, researchers at GlaxoSmithKline reported the use of a sequence-to-sequence (Seq2Seq) model that learns to translate between reduced graphs and SMILES.[41] Jin *et al*.[42] described JT-VAE, a two-step generation process in which a junction tree is first constructed to represent the composition of molecular substructures in a molecule (much like a reduced graph) and, subsequently, a graph message passing network is used to decode the final molecular structure. DeepFMPO[43] showed improved performance by constraining optimization according to the similarity of fragments considered.

Reaction based

Several works have reported reaction-based generative models for *de novo* molecular design. DINGOS[44] builds on the US Patent and Trademark Office (USPTO) reaction data set (D.M. Lowe, PhD thesis, University of Cambridge, 2012) to produce new chemical entities that are structurally similar to a molecular template using a hybrid of ML and rule-based methodologies. Molecule Chef[45]

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

uses a VAE to embed reactant structures and optimizes the molecular properties of the resulting product by biasing reactant selection (for a single-step reaction). ChemBO[46] represents an algorithmically simpler approach, whereby a random selection of reactants and conditions is used to generate candidate structures stochastically, which are then subjected to property evaluation. This workflow produces molecules through multistep chemical synthesis, although the selection of reactants is not biased toward the optimization objective.

Recent studies have reported the use of RL to navigate the vast space of possible reactions by modeling forward synthesis as a Markov decision process (MDP). REACTOR[47] uses a set of two-reactant reaction templates (encoded as reaction SMARTS) and selects missing reactants according to which ones will maximally improve the reward of the next state. When more than one reaction product is generated, the product associated with the largest reward is chosen. Similarly, Policy gradient for forward synthesis (PGFS)[48] combines biased reactant selection with multistep reactions and the directed acyclic graphs (DAGs) of (molecular) graphs (DoGs) approach iteratively generates a DAG for the forward synthesis route.[49] All these approaches have shown impressive performance optimizing toward desired molecular properties.

One of the breakthrough successes in ML for drug discovery was the development of synthesis-planning protocols, which learn to predict synthetic routes from large reaction databases.[50] A current limitation of reaction-based generative models is their dependence on hand-crafted reaction templates. We expect future research to combine the use of learned reaction schemas with universal optimizers over chemical space.

3D *de novo* molecular design

Consideration of the 3D environment of a molecule can be beneficial, and can be easily achieved by including 3D scorers in the optimization objective, such as shape similarity,[6] molecular docking,[29,51] quantum mechanics calculations,[52] or free energy perturbation (FEP) calculations.[53] However, these approaches are predicated on conformational assumptions of both the molecule and the protein (where relevant) and are frequently coupled with inaccurate representations of physical forces; therefore, the inclusion of 3D scorers is not always a superior approach. In addition to conceptual difficulties, there can be practical challenges, such as the computational and licensing costs associated with scoring thousands or even millions of molecules.

3D molecular structures can also be produced directly by a generative model through conditioned generation. Direct 3D generation has potential advantages for use cases such as optimizing protein–ligand binding (thus bypassing the docking search algorithm), or prediction of crystal packing. Indeed, early examples of *de novo* design constructed molecules in 3D (e.g., SPROUT[54]) use an iterative tree search of molecular fragments to build molecules within a protein-binding site. Modern approaches include DeLinker,[55] a method for 3D scaffold hopping and fragment growing; and LigDream,[56] which makes use of a conditional VAE and an image-captioning module to generate SMILES. Simm *et al.*[57] describe a generative framework based on RL that uses a rotationally invariant internal coordinate system, capable of generating molecules in 3D (see Table 1).

Successes and challenges

Synthesis and testing of de novo molecules

Ultimately, the impact of *de novo* design methods is seen through their use in drug discovery programs. A previous review has collected experimental validations of *de novo* design[58]; here, we discuss selected recent examples with a focus on generative models.

Possibly the most widely reported case of automated molecular design for drug discovery was published by Zhavoronkov *et al.*[59] wherein the authors use a GAN-based generative approach, GENTRL, to select 40 compounds for synthesis and testing against the discoidin domain receptor 1 (DDR1) kinase. The synthesized compounds were followed up with pharmacokinetic studies in mice, resulting in the identification of a lead compound with a favorable property profile; the authors acknowledged the potential for further optimization before progressing to candidate selection.

Assmann *et al.*[60] described the practical challenges of deploying *de novo* design to aid the discovery of novel inhibitors of CDK9. A refined VS strategy is described where molecules proposed by a molecular generator are used as seeds in a similarity search of the Enamine REAL library. Of 69 tested compounds,

TABLE 1

Summary of exemplar methods for *de novo* molecular design broken down by the coarseness of molecular representation.

	Atom based	Fragment based	Reaction based
Gradient free	GB-GA[21] ChemGE[22] MSO[23]	MOARF[6] CReM[27] CoG[19]	SYNOPSIS[28] AutoGrow4[29]
Gradient based	Segler *et al.*[32] REINVENT[33] GraphVAE[34] MolGAN[35] CG-VAE[36] Li *et al.*[37] MolDQN[38] GraphINVENT[39] Optimol[51]	Pogány *et al.*[41] JT-VAE[42] DeepFMPO[43]	DINGOS[44] Molecule Chef[45] ChemBO[46] REACTOR[47] PGFS[48] DoGs[49]

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

seven showed activity against CDK9. Another practical demonstration for deploying generative methods to identify optimal solutions to multiparameter objectives was recently reported by Perron *et al.*, using an RNN-based generative model.[61]

Li *et al.*[62] investigated the ability of RNN-based *de novo* design methods to generate novel molecule inhibitors in well-studied areas of chemical space. The authors described efforts to find novel inhibitors of the well-studied proto-oncogene serine/threonine-protein kinase 1 (PIM1) and CDK4 kinases. After testing four compounds, they reported one potent PIM1 inhibitor and two lead compounds that inhibit CDK4.

Grisoni *et al.*[63] recently demonstrated the combination of one-step reaction-based generative design combined with automated on-chip synthesis for identifying agonists of the liver X receptor (LXR). In total, 25 compounds were successfully synthesized and subsequent *in vitro* activity screening and follow-up studies revealed 12 of these to be potent hits with up to 60-fold LXR activation.

Generative models have also been used to propose novel small molecules for synthesis and testing as potential treatments for Coronavirus 2019 (COVID-19) infection.[64]

Practical artificial intelligence

Generative methods should be flexible in their usage such that they can complement routine design strategies in medicinal chemistry. The coarseness of molecular representation is closely linked with the actionability of proposed molecules. Although atom-based methods are expressive, and can feasibly explore the largest number of molecules, their synthesis can be impractical. The fragment-based paradigm restricts the expressivity of molecule generation, but has advantages with respect to the average actionability of the proposed molecules. Reaction-based methods are practical and actionable, but the search space over both reactants and reactions is vast and current methods do not offer an indication of whether synthesis is likely to be successful.

A common medicinal chemistry design strategy is to explore the impact of modifications constrained to a single region of a molecular series or surrounding a fixed scaffold. Such targeted modifications allow practitioners to build an understanding of relevant structure–property relationships. For *de novo* workflows to produce a set of molecules guaranteed to contain a given substructure, either the molecular generator must be aware of the molecular graph or a post-processing filter can be applied to remove molecules without the desired motif. The former strategy is preferable because it is more sample efficient and guarantees output molecules, given that it is possible that all final molecules might be filtered. For molecular generators relying on the SMILES (text-based) representation, fixed scaffold *de novo* design poses a challenge because a generator must learn which non-contiguous syntactic characters correspond to fixed substructure atoms. Arús-Pous *et al.*[65] overcome this representational shortcoming by training using fragmented molecules and allowing the generator to elaborate a molecule starting from attachment points on a molecular scaffold. Graph representations can achieve fixed scaffold *de novo* design more naturally.[66,67]

Published *de novo* design methods have displayed an inconsistent consideration of chirality. Whereas 3D methods explicitly output a chiral molecule, many approaches fail to consider chirality at all, and the user is forced to post rationalize the relevant enantiomer for synthesis. A simple solution is to enumerate and score all possible stereoisomers of a proposed molecule.[42] We expect that future benchmark developments should highlight this shortcoming among proposed methods.

Finally, the synthesizability of molecules proposed by *de novo* design methods has received much attention.[68] Although we have discussed the merits of atom-based, fragment-based, and reaction-based molecular representations in this regard, it remains a challenge to balance the expressivity, ease of optimization, and actionability of generated molecules. Beyond the synthetic accessibility of proposed chemistry, approaches that allow the user to specify a set of available building blocks offer an additional practical advantage. We expect future reaction-based workflows to prioritize high-yielding, reliable reactions at a lower cost and with fewer individual reaction steps.[69] In a similar practical vein, Vaucher *et al.*[70] recently described a method for the translation of unstructured reaction procedures written in text to an actionable sequence of synthetic steps; such an action sequence will be crucial for efforts toward automated synthesis.

Challenges designing an objective function

An outstanding challenge for *de novo* design is for desired property profiles to reflect the needs of medicinal chemistry more accurately. Although it is useful to demonstrate that methods can optimize molecules toward calculated molecular property profiles, similarity measures or quantitative structure–activity relationship (QSAR) models, drug discovery is multifaceted and current *de novo* design efforts are limited by a narrow view of the overall process.

Although there is an ongoing need to improve predictive models of complex biological responses, multi-objective optimization (MOO) aims to coalesce signals from several weak scorers using data fusion concepts, such as Pareto optimality or standardized *z*-scores.[71] The design of effective MOO profiles is nontrivial and often makes use of normalization functions and scaling protocols when combining multiple objectives.[72] It is usually necessary to experiment with several iterations between scoring function refinement and molecular generation. Gruenif.ai[23] demonstrates a human-in-the-loop workflow where the user can provide feedback interactively while molecules are generated. In the authors' opinion, an increased focus on such tools for guiding the development of effective scoring functions for molecular design would be beneficial.[23,73,74]

MOO guides molecular generators in their exploration of chemical space and often involves the use of one or more QSAR models. These models predict molecular properties of prospective compounds given a set of training examples. A recent study[75] exploring the failure modes of molecular generators showed that, although it is natural to rely on predictive models, caution should be exercised because molecular generators can exploit features unique to the model arising from model-specific and data-specific biases, resulting in the generation of molecules that are numerically superior but not practically useful. This undesired exploitation behavior is not specific to QSAR models and can also be observed for 3D models, such as those

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

discussed earlier. It is also important to consider the confidence of predictions made by QSAR models, because predictive accuracy is highly dependent on the domain of applicability model of a model.[76]

Improving benchmarks

Improved benchmarks of *de novo* design methods will encourage the development of more useful methodologies. Although *de novo* design has learnt a lot from computer vision with respect to the importance of standard benchmarks, there is now a need to branch out and create benchmarks specific to drug discovery.

Current benchmark metrics that measure the diversity of generated molecules can be misleading. In a study by Renz *et al.*[75] the authors experimented with a dummy generative model, AddCarbon, which randomly adds carbon to molecules in the training set. By design, the model only produces novel, valid molecules and, thus, performs extremely well on the GuacaMol distribution benchmarks. Given that this model is obviously useless in practice, this experiment demonstrates the ease with which current benchmarks can be fooled and shows the need to focus attention toward improving measures of generative performance. In particular, the systematic measurement of the novelty of chemical structures is challenging because, in reality, novelty implies a nonobvious inventive step, as judged by one skilled in the art of medicinal chemistry. Bush *et al.*[77] measured the overlap between chemists' ideas and molecules suggested by different design methods. Although not practical for measuring the aptitude of new molecular design tools, this study gives an indication of the true novelty of molecules suggested by generative approaches.

Whereas the GuacaMol goal-directed benchmark suite represents toy tasks for a molecular generator to achieve, many of which are trivially solved as remarked by Zhang *et al.*[78], we believe that more challenging tasks could be added to measure the usefulness of *de novo* methods in practical drug discovery programmes. For example, tasks relating to average synthesizability, susceptibility to local minima traps, sample efficiency,[79] generation within the constraints of a predefined building block library, or fixed scaffold molecular generation would provide direction for the development of future methodologies. Extensions could also be made for more specialized settings, such as 3D methods,[80] or those conditioned on biological context.

Despite improvements in standard benchmarks, there is clearly still a need to find a middle ground between the toy tasks *in silico* and true *in vitro* validation. We hope that the field will revisit the evaluation of *de novo* methods by experimental demonstration.[58]

Opportunities for de novo method development

There remain ample opportunities for algorithmic improvements to *de novo* design workflows, particularly toward making *de novo* methods more practically useful in drug discovery. In particular, reaction-based *de novo* workflows[49] and adoption of more natural graph-based representations[39] are promising directions for future research. We also believe fragment-based approaches offer an appealing compromise between molecular expressivity and practicality.

We also would like to remind the field that, despite the great attention and progress brought by gradient-based approaches for *de novo* design, gradient-free approaches are effective optimizers of chemical structure showing robustness, a reduced computational burden, and state-of-the-art benchmark performance; in addition, significant gains might be possible using methods that include the strengths of both families of techniques.[81]

In addition to continued advancements for existing families of generative models, we are excited about the potential for other statistically robust methods to efficiently generate molecules, particularly those that focus on a graph representation, such as flow-based autoregressive models, which map simple distributions to chemical space via a series of invertible transformations.[82,83]

Another interesting direction is the development of specialized generative models that incorporate additional context during generation, such as the 3D environment,[57] 3D protein binding sites,[84,85] or gene expression signatures.[86,87]

Finally, we comment that many generative models are still black-box technologies and, therefore, diagnostic studies that aim to better understand these approaches will help to make practical recommendations for their use in the field. For example, Grebner *et al.*[72] studied the impact of the choice of training set on generated molecules and explored different molecular scoring protocols. These studies enabled the authors to make suggestions for how to support programs practically at both lead generation and lead optimization.

Concluding remarks

The role of ML and AI in drug discovery is growing and there is an appetite for *de novo* molecular design methods because of their ability to navigate extremely large chemical spaces more effectively than either VS or a human expert. Despite early concerns regarding the use of automated methods for molecular design, often relating to the instability, reactivity, actionability, and synthetic feasibility of the molecules suggested,[68] we now have a variety of tools at our disposal that are proficient generators of sensible molecule structures.

Now the challenge for the field is to evaluate whether our generators and optimization objectives are useful for the tasks at hand. To this end, we reinforce the categorization between atom-based generators, the maximally expressive paradigm that could encourage us to make molecules that have not been previously described; fragment-based methods, being practical and constrained to a predefined set of building blocks; and reaction-based *de novo* design tools, which have a native grammar but a more challenging optimization problem.

De novo molecular design and generative chemistry models remain a controversial topic in the field,[88–91] but we believe there is potential to learn from collective experience and add these methods to the medicinal chemistry toolbox.

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

References

1 J.W. Scannell, J. Bosley, When quality beats quantity: decision theory, drug discovery, and the reproducibility crisis, PLoS ONE 11 (2016) e0147215.
2 T.J. Sejnowski, The unreasonable effectiveness of deep learning in artificial intelligence, Proc Natl Acad Sci U S A 117 (2020) 30033–30038.
3 Emmert-Streib F, Yang Z, Feng H, Tripathi S, Dehmer M. An introductory review of deep learning for prediction models with big data. Front Artif Intell Published online February 28, 2020. https://doi.org/10.3389/frai.2020.00004.
4 G.R. Bickerton, G.V. Paolini, J. Besnard, S. Muresan, A.L. Hopkins, Quantifying the chemical beauty of drugs, Nat Chem 4 (2012) 90–98.
5 Kusner MJ, Paige B, Hernández-Lobato JM. Grammar variational autoencoder. arXiv [statML]. Published online March 6, 2017. http://arxiv.org/abs/1703.01925.
6 N.C. Firth, B. Atrash, N. Brown, J. Blagg, MOARF, an integrated workflow for multiobjective optimization: implementation, synthesis, and biological evaluation, J Chem Inf Model 55 (2015) 1169–1180.
7 Polykovskiy D, Zhebrak A, Sanchez-Lengeling B, Golovanov S, Tatanov O, Belyaev S, et al. Molecular Sets (MOSES): a benchmarking platform for molecular generation models. arXiv [csLG]. Published online November 29, 2018. http://arxiv.org/abs/1811.12823.
8 N. Brown, M. Fiscato, M.H.S. Segler, A.C. Vaucher, GuacaMol: Benchmarking models for de novo molecular design, J Chem Inf Model 59 (2019) 1096–1108.
9 B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering, Science 361 (2018) 360–365.
10 D.C. Elton, Z. Boukouvalas, M.D. Fuge, P.W. Chung, Deep learning for molecular design—a review of the state of the art, Mol Syst Des Eng 4 (2019) 828–849.
11 Y. Xu, K. Lin, S. Wang, L. Wang, C. Cai, C. Song, et al., Deep learning for molecular generation, Future Med Chem 11 (2019) 567–597.
12 Fabian B, Edlich T, Gaspar H, Segler M, Meyers J, Fiscato M, et al. Molecular representation learning with language models and domain-relevant auxiliary tasks. arXiv [csLG]. Published online November 26, 2020. http://arxiv.org/abs/2011.13230.
13 N. Schneider, R.A. Sayle, G.A. Landrum, Get your atoms in order–an open-source implementation of a novel and robust molecular canonicalization algorithm, J Chem Inf Model 55 (2015) 2111–2120.
14 E.J. Bjerrum, B. Sattarov, Improving chemical autoencoder latent space and molecular de novo generation diversity with heteroencoders, Biomolecules 8 (2018) 131.
15 O. Prykhodko, S.V. Johansson, P.-C. Kotsias, J. Arús-Pous, E.J. Bjerrum, O. Engkvist, et al., A de novo molecular generation method using latent vector based generative adversarial network, J Cheminform 11 (2019) 74.
16 J. Arús-Pous, S. Johansson, O. Prykhodko, E.J. Bjerrum, C. Tyrchan, J.-L. Reymond, et al., Randomized SMILES strings improve the quality of molecular generative models, J Cheminform 11 (2019) 71.
17 O'Boyle N, Dalke A. DeepSMILES: an adaptation of SMILES for use in machine-learning of chemical structures. ChemRxiv, Published online September 18, 2018. https://doi.org/10.26434/chemrxiv.7097960.v1.
18 M. Krenn, F. Häse, A. Nigam, P. Friederich, SELFIES: a robust representation of semantically constrained graphs with an example application in chemistry, Mach Learn Sci Technol 1 (2020) 045024.
19 N. Brown, B. McKay, F. Gilardoni, J. Gasteiger, A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules, J Chem Inf Comput Sci 44 (2004) 1079–1087.
20 M. Reutlinger, T. Rodrigues, P. Schneider, G. Schneider, Multi-objective molecular de novo design by adaptive fragment prioritization, Angew Chem Int Ed Engl 53 (2014) 4244–4248.
21 J.H. Jensen, A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space, Chem Sci 10 (2019) 3567–3572.
22 N. Yoshikawa, K. Terayama, M. Sumita, T. Homma, K. Oono, K. Tsuda, Population-based de novo molecule generation, using grammatical evolution, Chem Lett 47 (2018) 1431–1434.
23 R. Winter, F. Montanari, A. Steffen, H. Briem, F. Noé, D.-A. Clevert, Efficient multi-objective molecular optimization in a continuous latent space, Chem Sci 10 (2019) 8016–8024.
24 Y. Kwon, J. Lee, MolFinder: an evolutionary algorithm for the global optimization of molecular properties and the extensive exploration of chemical space using SMILES, J Cheminform 13 (2021) 24.
25 J. Verhellen, J. Van den Abeele, Illuminating elite patches of chemical space, Chem Sci 11 (2020) 11485–11491.
26 X.Q. Lewell, D.B. Judd, S.P. Watson, M.M. Hann, RECAP–retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry, J Chem Inf Comput Sci 38 (1998) 511–522.
27 P. Polishchuk, CReM: chemically reasonable mutations framework for structure generation, J Cheminform 12 (2020) 28.
28 H.M. Vinkers, M.R. de Jonge, F.F.D. Daeyaert, J. Heeres, L.M.H. Koymans, J.H. van Lenthe, et al., SYNOPSIS: SYNthesize and OPtimize system in silico, J Med Chem 46 (2003) 2765–2773.
29 J.O. Spiegel, J.D. Durrant, AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization, J Cheminform 12 (2020) 25.
30 M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.-H. Altmann, G. Schneider, et al., A collection of robust organic synthesis reactions for in silico molecule design, J Chem Inf Model 51 (2011) 3093–3098.
31 G.M. Ghiandoni, M.J. Bodkin, B. Chen, D. Hristozov, J.E.A. Wallace, J. Webster, et al., Enhancing reaction-based de novo design using a multi-label reaction class recommender, J Comput Aided Mol Des 34 (2020) 783–803.
32 M.H.S. Segler, T. Kogej, C. Tyrchan, M.P. Waller, Generating focused molecule libraries for drug discovery with recurrent neural networks, ACS Cent Sci 4 (2018) 120–131.
33 Blaschke T, Arús-Pous J, Chen H, Margreitter C, Tyrchan C, Engkvist O, u REINVENT 2.0 – an AI Tool for De Novo Drug Design. ChemRxiv. Published online April 3, 2020. https://doi.org/10.26434/chemrxiv.12058026.v1.
34 Simonovsky M, Komodakis N. GraphVAE: towards generation of small graphs using variational autoencoders. In: Artificial Neural Networks and Machine Learning – ICANN 2018. London: Springer International Publishing; 2018: 412–422.
35 De Cao N, Kipf T. MolGAN: An implicit generative model for small molecular graphs. arXiv [statML]. Published online May 30, 2018. http://arxiv.org/abs/1805.11973.
36 Q. Liu, M. Allamanis, M. Brockschmidt, A. Gaunt, Constrained graph variational autoencoders for molecule design, Adv Neural Inf Process Syst 31 (2018) 7795–7804.
37 Li Y, Vinyals O, Dyer C, Pascanu R, Battaglia P. Learning deep generative models of graphs. arXiv [csLG]. Published online March 8, 2018. http://arxiv.org/abs/1803.03324.
38 Z. Zhou, S. Kearnes, L. Li, R.N. Zare, P. Riley, Optimization of molecules via deep reinforcement learning, Sci Rep 9 (2019) 10752.
39 R. Mercado, T. Rastemo, E. Lindelöf, G. Klambauer, O. Engkvist, H. Chen, et al., Graph networks for molecular design, Mach Learn Sci Technol 2 (2020) 025023.
40 M.F. Lynch, J.D. Holliday, The Sheffield generic structures project: a retrospective review, J Chem Inf Comput Sci 36 (1996) 930–936.
41 P. Pogány, N. Arad, S. Genway, S.D. Pickett, De novo molecule design by translating from reduced graphs to SMILES, J Chem Inf Model 59 (2019) 1136–1146.
42 Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. arXiv [csLG]. Published online February 12, 2018. http://arxiv.org/abs/1802.04364.
43 N. Ståhl, G. Falkman, A. Karlsson, G. Mathiason, J. Boström, Deep reinforcement learning for multiparameter optimization in de novo drug design, J Chem Inf Model 59 (2019) 3166–3176.
44 A. Button, D. Merk, J.A. Hiss, G. Schneider, Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis, Nat Mach Intell 1 (2019) 307–315.
45 J. Bradshaw, B. Paige, M.J. Kusner, M. Segler, J.M. Hernández-Lobato, A model to search for synthesizable molecules, Adv Neural Inf Process Syst 32 (2019) 7937–7949.
46 K. Korovina, S. Xu, K. Kandasamy, W. Neiswanger, B. Poczos, J. Schneider, E. Xing, ChemBO: Bayesian optimization of small organic molecules with synthesizable recommendations, Proc Machine Learn Res 108 (2020) 3393–3403.
47 Horwood J, Noutahi E. Molecular design in synthetically accessible chemical space via deep reinforcement learning. arXiv [physics.chem–ph]. Published online April 29, 2020. http://arxiv.org/abs/2004.14308.
48 Gottipati SK, Sattarov B, Niu S, Pathak Y, Wei H, Liu S, et al. Learning to navigate the synthetically accessible chemical space using reinforcement learning. arXiv [csLG]. Published online April 26, 2020. http://arxiv.org/abs/2004.12485.
49 Bradshaw J, Paige B, Kusner MJ, Segler MHS, Hernández-Lobato JM. Barking up the right tree: an approach to search over molecule synthesis DAGs. arXiv [csLG]. Published online December 21, 2020. http://arxiv.org/abs/2012.11522.
50 M.H.S. Segler, M. Preuss, M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI, Nature 555 (2018) 604–610.

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019

51 J. Boitreaud, V. Mallet, C. Oliver, J. Waldispühl, OptiMol: Optimization of binding affinities in chemical space for drug discovery, J Chem Inf Model 60 (2020) 5658–5666.
52 O.A. von Lilienfeld, O.A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, Exploring chemical compound space with quantum-based machine learning, Nat Rev Chem 4 (2020) 347–358.
53 P. Ghanakota, P.H. Bos, K.D. Konze, J. Staker, G. Marques, K. Marshall, et al., Combining cloud-based free-energy calculations, synthetically aware enumerations, and goal-directed generative machine learning for rapid large-scale chemical exploration and optimization, J Chem Inf Model 60 (2020) 4311–4325.
54 V.J. Gillet, W. Newell, P. Mata, G. Myatt, S. Sike, Z. Zsoldos, et al., SPROUT: recent developments in the de novo design of molecules, J Chem Inf Comput Sci 34 (1994) 207–217.
55 F. Imrie, A.R. Bradley, M. van der Schaar, C.M. Deane, Deep generative models for 3D linker design, J Chem Inf Model 60 (2020) 1983–1995.
56 M. Skalic, J. Jiménez, D. Sabbadin, G. De Fabritiis, Shape-based generative modeling for de novo drug design, J Chem Inf Model 59 (2019) 1205–1214.
57 Simm GNC, Pinsler R, Hernández-Lobato JM. Reinforcement learning for molecular design guided by quantum mechanics. arXiv [statML]. Published online February 18, 2020. http://arxiv.org/abs/2002.07717.
58 G. Schneider, D.E. Clark, Automated de novo drug design: are we nearly there yet?, Angew Chem Int Ed Engl 58 (2019) 10792–10803
59 A. Zhavoronkov, Y.A. Ivanenkov, A. Aliper, M.S. Veselov, V.A. Aladinskiy, A.V. Aladinskaya, et al., Deep learning enables rapid identification of potent DDR1 kinase inhibitors, Nat Biotechnol 37 (2019) 1038–1040.
60 Assmann M, Bal M, Craig M, D'Oyley J, Phillips L, Triendl H, et al. A novel machine learning approach uncovers new and distinctive inhibitors for cyclin-dependent kinase 9. bioRxiv. Published online March 19, 2020. http://dx.doi/org/10.1101/2020.03.18.996538.
61 Perron Q, Mirguet O, Tajmouati H, Skiredj A, Rojas A, Gohier A, et al. Deep generative models for ligand-based de Novo design applied to multi–parametric optimization. ChemRxiv. Published online January 25, 2021. http://dx.doi/org/10.26434/chemrxiv.13622417.v2.
62 X. Li, Y. Xu, H. Yao, K. Lin, Chemical space exploration based on recurrent neural networks: applications in discovering kinase inhibitors, J Cheminform 12 (2020) 42.
63 Grisoni F, Huisman B, Button A, Moret M, Atz K, Merk D, et al. Combining generative artificial intelligence and on-chip synthesis for de novo drug design. ChemRxiv. Published online December 30, 2020. http://dx.doi/org/10.26434/chemrxiv.13498587.v1.
64 A.C. Kaushik, U. Raj, AI-driven drug discovery: a boon against COVID-19?, AI Open 1 (2020) 1–4
65 J. Arús-Pous, A. Patronov, E.J. Bjerrum, C. Tyrchan, J.-L. Reymond, H. Chen, et al., SMILES-based deep generative scaffold decorator for de-novo drug design, J Cheminform 12 (2020) 38.
66 W. Jin, D.R. Barzilay, T. Jaakkola, Multi-objective molecule generation using interpretable substructures, Proc Mach Learn Res 119 (2020) 4849–4859.
67 J. Lim, S.-Y. Hwang, S. Moon, S. Kim, W.Y. Kim, Scaffold-based molecular design with a graph generative model, Chem Sci 11 (2020) 1153–1164.
68 W. Gao, C.W. Coley, The synthesizability of molecules proposed by generative models, J Chem Inf Model 60 (2020) 5714–5723.
69 G.M. Makara, L. Kovács, I. Szabó, G. Pőcze, Derivatization design of synthetically accessible space for optimization: in silico synthesis vs deep generative design, ACS Med Chem Lett 12 (2021) 185–194.
70 A.C. Vaucher, F. Zipoli, J. Geluykens, V.H. Nair, P. Schwaller, T. Laino, Automated extraction of chemical synthesis actions from experimental procedures, Nat Commun 11 (2020) 3601.
71 C.A. Nicolaou, N. Brown, Multi-objective optimization methods in drug design, Drug Discov Today Technol 10 (2013) e427–e435.
72 C. Grebner, H. Matter, A.T. Plowright, G. Hessler, Automated de novo design in medicinal chemistry: which types of chemistry does a generative neural network learn?, J Med Chem 63 (2020) 8809–8823
73 M.I. Ecemis, J. Wikel, C. Bingham, E. Bonabeau, A drug candidate design environment using evolutionary computation, IEEE Trans Evol Comput 12 (2008) 591–603.
74 E.-W. Lameijer, J.N. Kok, T. Bäck, A.P. Ijzerman, The molecule evoluator. an interactive evolutionary algorithm for the design of drug-like molecules, J Chem Inf Model 46 (2006) 545–552.
75 P. Renz, D. Van Rompaey, J.K. Wegner, S. Hochreiter, G. Klambauer, On failure modes in molecule generation and optimization, Drug Discov Today Technol 32–33 (2019) 55–63.
76 N. Brown, B. McKay, J. Gasteiger, A novel workflow for the inverse QSPR problem using multiobjective optimization, J Comput Aided Mol Des 20 (2006) 333–341.
77 J.T. Bush, P. Pogany, S.D. Pickett, M. Barker, A. Baxter, S. Campos, et al., A Turing test for molecular generators, J Med Chem 63 (2020) 11964–11971.
78 Zhang J, Mercado R, Engkvist O, Chen H. Comparative study of deep generative models on chemical space coverage. ChemRxiv. Published online November 16, 2020. https://doi.org/10.26434/chemrxiv.13234289.v3.
79 C.W. Coley, N.S. Eyke, K.F. Jensen, Autonomous discovery in the chemical sciences part II: outlook, Angew Chem Int Ed Engl 59 (2020) 23414–23436.
80 Cieplinski T, Danel T, Podlewska S, Jastrzebski S. We should at least be able to design molecules that dock well. arXiv [q-bioBM]. Published online June 20, 2020. http://arxiv.org/abs/2006.16955.
81 Nigam A, Friederich P, Krenn M, Aspuru-Guzik A. Augmenting genetic algorithms with deep neural networks for exploring the chemical space. arXiv [csNE]. Published online September 25, 2019. http://arxiv.org/abs/1909.11655.
82 Shi C, Xu M, Zhu Z, Zhang W, Zhang M, Tang J. GraphAF: a flow-based autoregressive model for molecular graph generation. arXiv [csLG]. Published online January 26, 2020. http://arxiv.org/abs/2001.09382.
83 Madhawa K, Ishiguro K, Nakago K, Abe M. GraphNVP: an invertible flow model for generating molecular graphs. arXiv [statML]. Published online May 28, 2019. http://arxiv.org/abs/1905.11600.
84 Xu M, Ran T, Chen H. De Novo molecule design through molecular generative model conditioned by 3D information of protein binding sites. ChemRxiv. Published online December 30, 2020. https://doi.org/10.26434/chemrxiv.13498332.v1.
85 M. Simonovsky, J. Meyers, DeeplyTough: learning structural comparison of protein binding sites, J Chem Inf Model 60 (2020) 2356–2366.
86 O. Méndez-Lucio, B. Baillif, D.-A. Clevert, D. Rouquié, J. Wichard, De novo generation of hit-like molecules from gene expression signatures using artificial intelligence, Nat Commun 11 (2020) 10.
87 J. Born, M. Manica, A. Oskooei, J. Cadow, G. Markert, Martínez M. Rodríguez, PaccMannRL: de novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning, iScience 24 (2021) 102269.
88 A. Zhavoronkov, A. Aspuru-Guzik, Reply to "Assessing the impact of generative AI on medicinal chemistry", Nat Biotechnol 38 (2020). 146–146.
89 Bajorath J. Faculty Opinions Recommendation of "Assessing the impact of generative AI on medicinal chemistry". Faculty Opinions. Published online February 11, 2020; https://doi.org/10.3410/f.737286584.793570630.
90 W.P. Walters, W. Patrick Walters, M. Murcko, Assessing the impact of generative AI on medicinal chemistry, Nat Biotechnol 38 (2020) 143–145.
91 H. Chen, O. Engkvist, Has drug design augmented by artificial intelligence become a reality?, Trends Pharmacol Sci 40 (2019) 806–809

Please cite this article in press as: J. Meyers et al., Drug Discovery Today (2021), https://doi.org/10.1016/j.drudis.2021.05.019