JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Review

Advances and Challenges in De Novo Drug Design Using Three-Dimensional Deep Generative Models

Weixin Xie, Fanhao Wang, Yibo Li, Luhua Lai, and Jianfeng Pei*

Cite This: J. Chem. Inf. Model. 2022, 62, 2269−2279 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: A persistent goal for *de novo* drug design is to generate novel chemical compounds with desirable properties in a labor-, time-, and cost-efficient manner. Deep generative models provide alternative routes to this goal. Numerous model architectures and optimization strategies have been explored in recent years, most of which have been developed to generate two-dimensional molecular structures. Some generative models aiming at three-dimensional (3D) molecule generation have also been proposed, gaining attention for their unique advantages and potential to directly design drug-like molecules in a target-conditioning manner. This review highlights current developments in 3D molecular generative models combined with deep learning and discusses future directions for *de novo* drug design.

KEYWORDS: *de novo drug design, deep learning, generative model, three-dimentional generation, structure-based generation, structure-based drug design*

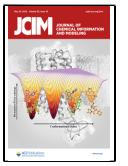
INTRODUCTION

The goal of computational *de novo* drug design is the rational discovery of novel and potent drug compounds at a reduced experimental cost in the validation stage. Traditional structure-based *de novo* design methods, one of the key players in this field, have been developed over decades and their effectiveness has been demonstrated many times.[1] Recently, generative models combined with deep learning techniques have gained increasing attention and injected new vitality into drug discovery research. Through efficient modeling, deep generative models accelerate the hit discovery phase by sampling from and exploring the chemical space of drug-like compounds.

Chemical structures are commonly described as graphs in which the atoms and bonds are represented as nodes and edges, respectively. Simplified molecular input line entry specification (SMILES), an equivalent string representation first introduced by Weininger,[2] can be derived from molecular graphs. Inspired by the generative performance of recurrent neural networks (RNNs) in natural language processing, generative models based on a RNN trained with SMILES were first developed to sample novel entities from chemical space,[3−5] followed by graph-based generative models trained on molecular graphs.[6,7] SMILES- and graph-based generative models are commonly referred to as two-dimensional (2D) models, as they produce only topological (or 2D) structures of molecules without three-dimensional (3D) coordinates. Since the initial success of 2D generative models, numerous generative architectures have been explored and combined with various optimization strategies to achieve targeted compound design.[8]

Many attempts have been made to bias 2D generative models toward the design of molecules that are likely to interact with target proteins using either ligand- or receptor-based approaches. However, due to the nature of 2D molecular representations, encoding atomic interactions between a ligand and the protein binding site during the generation process can be challenging. Usually, researchers resort to taking advantages of known actives for the protein target. For example, many studies have employed transfer learning by fine-tuning the generative model with known actives to force the model to focus on related regions in chemical space.[5,9,10] Techniques such as beam search sampling have been used to enhance the prioritization and selection of promising candidates in this process.[11] Some approaches use known actives for the target receptor to initiate searches in the latent chemical space.[12,13] Others build quantitative structure−activity relationship models from these known actives and use them as a reward function in reinforcement learning to optimize the generative

Received: January 19, 2022
Published: May 11, 2022



https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

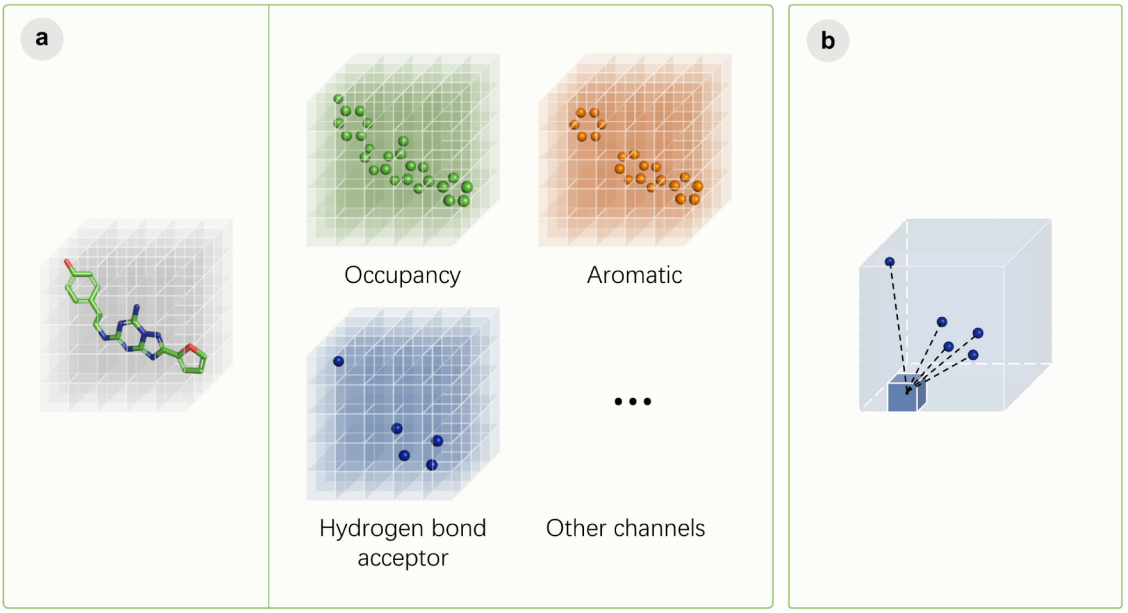


Figure 1. Cubic grid-based molecular representation. The antagonist ZM241385 of the human adenosine A2A receptor, taken from the PDB entry 3EML, is used for illustration purposes. (a) Atoms with the same pharmacophoric properties are assigned to a separate channel and are enclosed by a cubic box of a standard size. The pharmacophore here is identified by RDKit. (b) The value of each voxel is determined by all the atoms included in the channel and is most affected by the nearest atom. Here, the hydrogen-bond-acceptor channel is used as an example.

models.[14−16] These methods are known as ligand-based approaches, as they rely on the availability of compounds with validated bioactivity. By contrast, some studies have alleviated the necessity of obtaining known actives beforehand by developing receptor-based approaches. These methods condition molecular generative models on protein-related information, such as amino acid sequences,[17] the Coulomb matrix of protein pockets,[18] and pharmacophoric constraints.[19,20] Although some of these approaches are conditioned on 3D structures of the receptors, they still generate molecules in SMILES or graph forms. These advances have been comprehensively reviewed before.[21,22]

To obtain a 3D conformer of a generated molecule, it is effective to combine 2D generative models with conformation generation modules to embed it into 3D space or predict its binding pose in the protein pocket.[12,23] It seems there is no need to develop a 3D generative model. However, a major distinction between 3D generative and conformation generation models is that 3D generative models produce topological and conformational structures simultaneously, whereas a conformation generation model generates a conformer based on a known molecular graph. Although 3D generative models are harder to train, they possess some unique advantages, which are elaborated in the next paragraph. The potential for deep learning applications in 3D generative modeling is also of interest. Thus, this review focuses on 3D generative models that directly produce 3D molecular structures.

Representing a molecule using its 3D conformer is a natural approach that has been widely adopted by computational structure-based drug design (SBDD) programs, such as LigBuilder,[24−26] before the recent rise of deep learning methods. Representing molecules in 3D space and developing a 3D generative model have multiple advantages. First, 3D generative models are better suited for receptor (structure)-based drug design. Unlike a 2D molecular graph, which is a simplified representation to describe molecular structures, molecules with 3D conformation allow consideration of intra- and intermolecular interactions during the generation process. More importantly, all local interaction constraints from the protein pocket can be incorporated into the iterative generation of a molecule, which mimics the processes applied by human experts during the structure-based optimization process. Second, the generalizability of 3D generative models is not limited by the sparsity or absence of known ligands that bind to the target protein because design knowledge obtained from other proteins can be transferred to new targets. This meets the requirement of *de novo* drug design, and the novelty of generated molecules may be less restricted than those based on known actives. Last, 3D molecular generative models may achieve a higher level of automation with less human interference. Currently, to evaluate the binding probability, molecules produced by most 2D generative models are first embedded in a 3D space and docked into the binding sites of the protein. Then, the binding modes and scores are analyzed to decide whether the important atomic interaction is recovered. However, choosing suitable embedding algorithms and docking programs can be tricky, especially when deep learning-based methods come into play.[27,28] The ideal approach is a 3D generative model that is able to combine these processes end-to-end to produce reasonable binding poses.

The present paper provides a comprehensive review of recently reported 3D molecular generative models. We first introduce current techniques for 3D molecular structure generation and categorize them into three types, depending on the featurization methods. Then, we provide an overview of the performances of these models when generating in free space or conditioned on the binding sites of the protein targets. Finally, we discuss the unsolved issues and challenges in this field and how they may be addressed in the future.

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

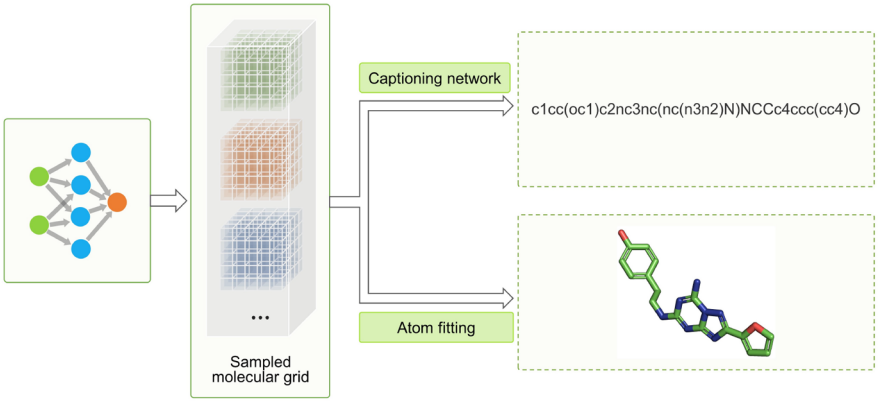


Figure 2. Generation using cubic grid-based molecular representation. To sample a 3D molecule, a generative model first produces the corresponding cubic grids of all channels. Then, these cubic grids are used to infer the exact molecular structures by other postprocessing modules, such as a captioning network or an atom-fitting algorithm.

FEATURIZATION METHODS OF 3D DEEP GENERATIVE MODELS

We categorize the featurization of 3D molecular structures into three types: cubic grid-based, Euclidean distance matrix (EDM)-based, and Cartesian coordinate-based. Each type of featurization requires distinct generative architectures and optimization strategies.

Cubic Grid-Based Molecular Featurization. Representing molecular structures as cubic grids has been successfully exploited in related areas, such as the design of protein−ligand scoring functions[29] and the detection of protein druggable binding sites.[30] The popularity of this method arises from its compatibility with convolutional neural networks, which enables automatic feature extraction and is supported by major advances in computer vision. The ability of small molecules and macro-molecules, such as proteins, to pass through a similar featurization process is also desirable.

The cubic grid representation usually contains several channels, which are analogous to the RGB channels of raster images. These channels are concatenated into a four-dimensional tensor as the direct input for a neural network. Each channel spans a separate cubic grid and focuses on one of the atomic properties, including occupancy (whether it is a nonhydrogen/heavy atom or not) and pharmacophoric properties, such as hydrophobicity, aromaticity, hydrogen-bond acceptor or donor, ionizability, and metallicity. Each atom is assigned to a channel if it has the given property, according to its type and a predefined rule (Figure 1a). Therefore, only atoms with the desired property are considered in each channel, while all heavy atoms are included in the occupancy channel.

Next, the molecule is voxelized into discretized cubic grids centered around it. The value at each grid point, or voxel, in each channel is determined by an atomic smearing function, which quantitatively defines the influence each atom has on its surrounding voxels. Generally, an atomic smearing function depends on atomic radius and assigns a larger influence value to voxels closer to any atom, exemplified by previously explored pair correlation functions[30] and Gaussian-like densities.[31,32] The resulting value at each voxel is the summation or maximum of all influences exerted upon it by every atom in the channel (Figure 1b).

The voxelized molecular structure is in a natural input form for convolutional blocks. Previous work has explored several generative architectures with convolutional layers, including convolutional neural networks (CNNs),[33] autoencoders (AEs),[31] variational autoencoders (VAEs),[31,32,34] and generative adversarial networks (GANs).[35] Generators take voxelized protein binding sites or latent variables encoded from ligand grids as input and sample new binders in a 3D grid form (Figure 2).

An additional optimization step is needed because the transformation between the original molecular structure and the cubic grid representation is not invertible. Generally, when a molecule is converted to its cubic grid form, information such as atom types and the exact positions of atoms and bonds are lost, making it impossible to directly reconstruct the original structure. Therefore, when a molecule in cubic grid form is sampled by a generator, determining the corresponding structure is a nontrivial task. Some studies have attempted to train an RNN to parse the sampled cubic grids into SMILES.[34,35] Others have tried to fit specific atoms into the generated grid using additional optimization algorithms,[31,32] but only obtained atom types and atomic positions, whereas bond information required inference using other programs, such as OpenBabel[36] (Figure 2).

EDM-Based Molecular Featurization. The second type of representation describes a molecular structure by its atomic type vector/matrix and its EDM. A bond matrix accounting for the connectivity of atoms and bond orders is also sometimes included. To circumvent the permutation problem, EDM-based approaches are constrained to fixed chemical compositions or use a specific ordering, such as the InChI-based canonical labeling.

A key requirement of EDM-based featurization is ensuring that the EDM sampled by the generator is in a reasonable form, which requires the application of additional constraints during the training process. The Gram matrix of the EDM is forced to be positive semidefinite, which requires the eigenvalues of the Gram matrix to be non-negative, and the matrix must have a rank of 3 because the atomic coordinates exist in a maximally three-dimensional embedding. However, applying these constraints is nontrivial in practice because the EDM directly sampled by the model at the beginning of training may be unfavorable for backpropagation through eigenvalue decomposition, making model updates problematic. To address these issues, Hoffmann et al. have proposed a novel reparameterization scheme using a non-negative function to

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

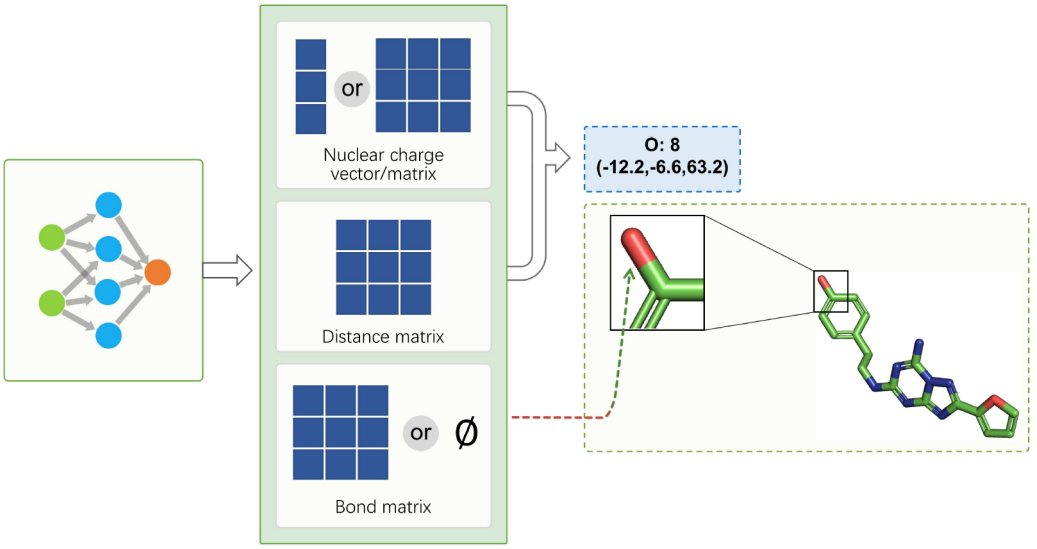


Figure 3. Generation using EDM-based molecular representation. Multiple matrices encoding molecular structures are first sampled by a generative model. Then, the atomic types and coordinates are constructed from the nuclear charge matrix and distance matrix, respectively. The connectivity of the atoms is completed by the bond matrix or other toolkits, such as OpenBabel.

transform the outputs of the model into a positive semidefinite Gram matrix and then a valid EDM.[37]

To construct molecular structures from multiple sampled matrices, the EDM must be converted into the three-dimensional coordinates of the atoms. A multidimensional scaling algorithm is usually adopted for this purpose. Software such as OpenBabel[36] is then used to infer connectivity if the EDM-based representations do not contain the bond information and add hydrogen atoms if the molecule has undergone hydrogen removal (Figure 3).

Cartesian Coordinate-Based Molecular Featurization. The third type of method directly samples molecular structures embedded in 3D space, which is very different from the two featurization methods mentioned above. Cartesian coordinate-based representation contains the element type and Cartesian coordinates for each atom in the compound. Generators construct molecular structures in an autoregressive manner, by adding atoms to the partial structures one-by-one.[38−42] Some work also attempt to sample the entire structure in a single step.[43]

Building such a model is not straightforward. When the molecular structure is rotated by a certain degree, the generator should place the next atom at a correspondingly rotated position, such that generation process is covariant to spatial transformation. Therefore, special attention should be paid to two aspects of the generative model when using Cartesian coordinate-based representation.

The first is how to obtain a rotation-, translation-, and permutation-invariant or covariant state embedding that encodes an abstract understanding of the current state. Invariant state embedding is necessary to determine actions invariant to transformation, such as the selection of the next atom type, in addition to covariant state embedding to determine where to place the next atom. A handful of works have based their generative models on SchNet,[44,45] a deep learning architecture consisting of continuous-filter convolutional layers and originally designed for predicting molecular properties with leading accuracy. The SchNet module produces an invariant embedding for each atom that captures information regarding distances from neighboring atoms.[38−40,42] Recently, Li et al.[46] devised a novel state embedding layer by extending a massage passing neural network with an internal coordinate system to obtain an invariant embedding. Others used different covariant embedding modules, such as the E(n) transformation equivariant graph neural network variants[43] or models based on spherical harmonics.[41]

The second aspect to be aware of is the policy for constructing molecular structures. Assembling a molecular structure involves multiple events, such as focusing on a placed atom and choosing the next atom type, and each event requires a concrete policy. The positional policy, which determines where to place the next atom based on the current situation, requires special attention. An invariant state embedding requires an invariant positional policy, and a covariant state embedding requires a covariant one. Information regarding changes in the orientation of a molecule cannot propagate through an invariant state embedding module; therefore, a new atom is unlikely to be placed accordingly. The major difference between approaches also lies in the positional policy. When a state embedding is obtained with SchNet,[38,39,42] an invariant positional policy introduced by Gebauer et al.[38] can be applied, which predicts the pairwise distances between the next atom and all preceding atoms and then constructs a distribution of candidate locations on a small grid centered around the focal atom prior to the placement of the next atom. In addition to this approach, Simm et al.[40] have modeled the position of an atom using internal coordinates that are invariant to rotation and translation based on a combination of bond length, bond angle, and torsion angle. Equipped with a novel embedding module, Li et al.[46] assembled molecular structures in 3D space equivariantly by predicting the internal coordinates of each atom. For covariant state embedding, Simm et al.[41] have used the local coordinates (d, θ, ϕ) to locate an atom, where d is the distance between the focal atom and the next atom and (θ, ϕ) is the orientation on a unit sphere around the focal atom. Others researchers have forced this requirement for covariance

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

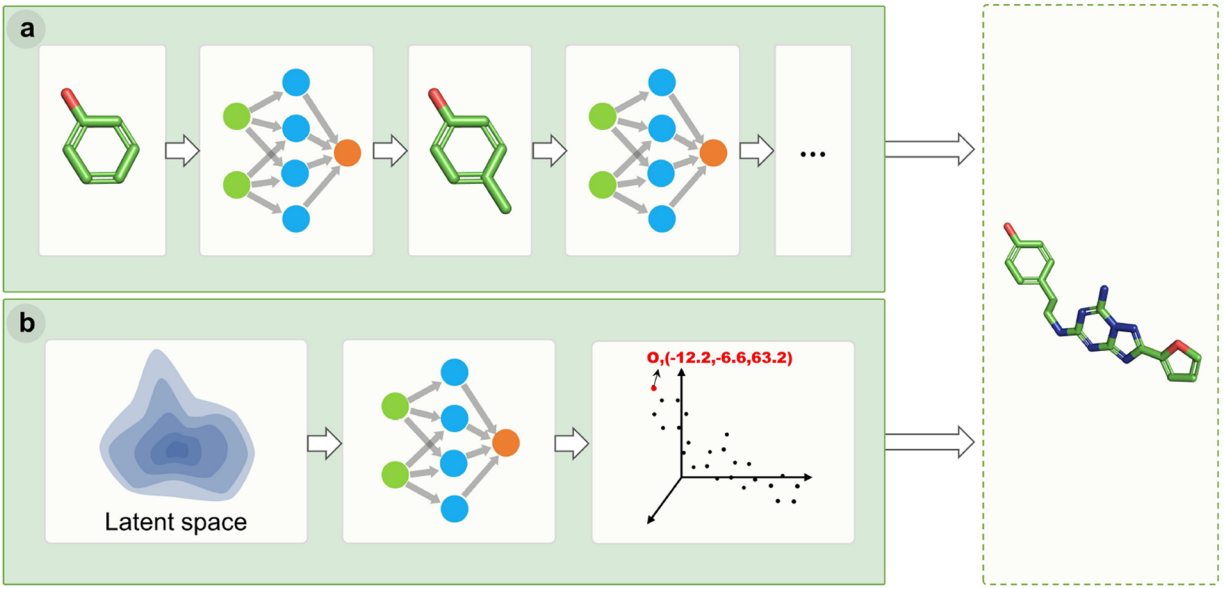


Figure 4. Generation using Cartesian coordinate-based molecular representation. A molecule could either be (a) assembled sequentially by placing atoms in an autoregressive way or (b) decoded from a latent space in one shot. The connectivity between atoms is sometimes absent and requires further inference.

through an equivariant function parametrized by neural networks.[43]

Instead of obtaining the entire molecular structure at the end of the generation process, sometimes bonds must be inferred from the atom types and the 3D coordinates outputted by the generator (Figure 4).

3D DEEP GENERATIVE MODELS AND THEIR APPLICATIONS

This section provides an overview of current 3D molecular generative models, including their model architectures, training data, and model performance. We divided these methods into two types depending on whether they are more suitable for ligand-based or protein structure-based design. Moreover, because no specialized benchmark exists for 3D generative models, training data were collected from various sources, and 3D conformers were generated by different methods with varying levels of precision. Some data were provided by data set maintenance teams, such as the QM9 data set, whereas others were generated by researchers with chemoinformatic toolkits, like RDKit[47] or OpenBabel.[36] The data sets used and the detailed results of each work reviewed below can be found in Table S1.

For Ligand-Based Design. *Cubic Grid or EDM-Based. Case 1.* Skalic et al.[34] have proposed a pharmacophore-based generative model utilizing 3D molecular shapes. Molecules were voxelized into a discretized cubic grid with five channels: hydrophobicity, aromaticity, hydrogen-bond donors, hydrogen-bond acceptors, and heavy atoms (occupancy). Corresponding compound pharmacophores were similarly featurized using only three property channels: hydrogen-bond donors, hydrogen-bond acceptors, and aromatic ring centers. Their model consisted of a shape VAE using 3D convolutional layers that autoencodes the molecules conditioned on the pharmacophore features and a shape captioning network that parsed the molecular shape sampled by the shape VAE into a SMILES string. The model was trained on drug-like compounds from the ZINC15 database, whose 3D conformers were generated via RDKit.

The authors reported a reconstruction rate of 1.74%, with an overall validity of 99.5% when seed molecules from the test set were used to generate latent vectors. They also benchmarked their unconditional version (without the pharmacophore location) of the model on the MOSES[48] platform and achieved performance comparable to other published methods. When tasked with identifying molecules analogous to known binders (used as seed structures) of three pharmaceutical targets, adenosine A2A receptor, thrombin, and stem cell growth factor receptor (KIT), their model tended to propose larger structural modifications relative to the seed structures than Chemical VAE,[49] a SMILES-based method. This difference may be due to the similarity between the cubic grid representation of 3D molecules and the pharmacophore model, and generated molecules decoded from that representation may introduce more structural variations. However, the molecules generated by their model still retained drug-like properties and had higher shape similarity to known binders than random samples from ZINC15. The binding affinities or probabilities of the generated molecules were evaluated by two virtual screening tools and were consistently higher than those of the corresponding decoys from the DUD-E database.[50] Therefore, their model was believed to have good potential for lead discovery. However, none of the generated compounds have been validated experimentally.

Case 2. Ragoza et al.[31] have transformed molecules from the MolPort database into 3D grids prior to training a VAE with convolutional blocks, followed by an adversarial discriminative network. This method was evaluated using a test set collected from the PubChem database. To address the issue of structural inference, they devised an optimization algorithm that fits a set of atoms into the sampled 3D grid with beam search and gradient descent. Their approach was validated by reconstructing molecular structures from the real 3D grids of the test molecules, achieving an overall validity of 99% and an average root-mean-square deviation (RMSD) of 0.011 Å for those that

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

successfully recovered the exact atom types. When the 3D grids of the test molecules from AE posterior and VAE posterior were used, the validity dropped to 92% and 91%, with average Tanimoto similarities between the recovered molecules and the inputted ones of 0.43 and 0.24 and mean quantitative estimate of drug-likeness (QED) values of 0.49 and 0.43, respectively. When sampling molecules from the VAE prior distribution, they reported validity of 91% and significantly smaller sizes for the sampled molecules, with a mean QED of 0.38. Furthermore, shape continuity was found when exploring and interpolating in the latent space.

Case 3. Hoffmann et al.[37] have introduced EDMnet, which utilizes the second type of representation of 3D molecules. The authors trained EDMnet as Wasserstein GANs with a gradient penalty in a multitask fashion on 6095 isomers with the chemical formula $C_7O_2H_{10}$ collected from the QM9 data set to generate molecules in the form of an atom type vector and an EDM. Their model captured atomic pairwise distance distributions, except for the O−O pair. After using OpenBabel to add bonds and removing the invalid structures, the authors reported a validity of 7.5% and a uniqueness rate for cannonical SMILES of 1−2%, which may imply mode collapse. The authors demonstrated the ability of the model to sample diverse conformers for new isomers (338 out of nearly 4000 valid samples). They also found the total energies of the relaxed structures were close to those of the molecules from the QM9 data set. However, the stability of the new conformers before relaxation was uncertain.

Case 4. Nesterov et al., the authors of 3DMolNet,[51] have represented molecules with hydrogens removed as a concatenation of nuclear charge matrix, EDM, and bond order matrix. The 3DMolNet adopted the architecture of VAE and was trained on molecules from the QM9 data set. The model achieved a high reconstruction accuracy on the test set, with median RMSDs of 0.048 and 0.16 Å for heavy and full atoms, respectively. A smooth transition was observed when interpolating in the latent space, especially between structurally similar molecules consisting of different chemical compositions. The intermediate structures were also found to be close to their relaxed ones. Moreover, over 20,000 structures sampled from the latent space were found to be novel isomers not contained in the QM9 data set, showing the ability of 3DMolNet to discover novel molecules.

Cartesian Coordinate-Based. Case 1. Gebauer et al.[38] have proposed an adapted SchNet generative architecture combined with a rotation and translation invariant position sampling policy to construct molecular structures through sequential generation. The model was trained on the equilibrium geometries of 6095 constitutional isomers of $C_7O_2H_{10}$ from the QM9 data set. They reported 4392 unique molecules out of 10000 generated structures, and hundreds of molecules resembled unseen samples in the test set. Furthermore, 20 generated structures with low predicted potential energy were close to the equilibrium configurations, with median RMSDs of 0.2 and 0.36 Å for heavy atoms and all atoms, respectively, comparing before and after relaxation.

Case 2. In another work, Gebauer et al.[39] have used a similar generative pipeline, G-SchNet.[38] The major difference between their two studies was the introduction of two auxiliary tokens, the origin token and the focus token, which label the positions of the center of mass of the generated molecule and the current focal atom, respectively. The training data included 50,000 molecules randomly selected from the QM9 data set. After the generation of molecules was complete, they used OpenBabel[36] to add bonds for the chemical valency check.

Among the 20,000 molecules generated, 77% of them were valid, 87% were novel, and 92% were unique. The radial distributions for pairs of atoms and the angular distributions for bonded triplets were properly reproduced, and the averaged counts of atoms, bonds, and rings per molecule were also close to those of the training data. They generated 3D structures aligned well with their relaxed states, obtained by energy minimization, with a median RMSD of 0.25 Å. The authors successfully biased the generated distribution of highest occupied molecule orbital−lowest occupied molecule orbital gaps toward smaller value regions through transfer learning using the generative model.

Joshi et al.[42] have proposed a generative model largely based on G-SchNet. The major difference from the previous work was the use of a given scaffold to initiate the generation process. Their model was applied to lead optimization tasks for two enzymes of the SARS-CoV-2, the main protease (Mpro) and endoribonuclease (NSP15). Starting from the scaffolds of common electrophilic warheads or bioactive compounds, their model generated molecules with good binding affinity, supported by docking simulations. However, the representative generated structures showed high similarity to a known inhibitor of the NSP15. The bioactivity of the generated molecules against the SARS-CoV-2 enzymes also requires further experimental confirmation.

Case 3. Simm et al.[40] have considered the molecular generation task as a sequential decision-making problem, in which atoms are drawn from a bag of given atoms and placed one by one onto the 3D canvas. They employed SchNet to produce an invariant state embedding of the current configuration and used internal coordinates to specify the relative atomic positions. The sampling policy was iteratively updated through a proximal policy optimization reinforcement learning algorithm. The reward that the agent received was formulated as the negative difference in energy before and after the new atom was placed on the canvas, which encouraged the design of stable molecules.

When given a single atoms bag or multiple atoms bags that required the agent to construct several molecules simultaneously, their model was able to construct diverse 3D structures with a validity of 60%−90% and find nearly stable configurations at the end of the exploration phase. Generally, the agent performed worse when more heavy atoms or element types were included in the bag of given atoms. The model was also evaluated on a solvation task, in which water molecules were generated in the vicinity of a solute. The agent managed to arrange water molecules to allow them to form hydrogen bonds with the solute. Moreover, the agent could generalize to unseen bags of atoms of the same size or a larger size by applying transfer learning.

Case 4. Similar to their previous reinforcement learning pipeline generating molecules on a 3D canvas, Simm et al.[41] have presented a rotation-covariant generative model that specifies the bond length and orientation relative to the focal atom when placing a new atom, which addressed the ambiguity of using internal coordinates. They employed CORMORANT, a neural network architecture that predicts the properties of a chemical system and works in the representation of the SO(3) group, to obtain both invariant and covariant state embeddings for the structures being generated. To alleviate the constraint

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

of a fixed chemical formula, they allowed the model to assemble molecular structures from a stochastic bag of atoms.

The model showed superior stability for the final molecular structures and the training process compared with their previous internal coordinates-based model, especially when building highly symmetric molecules, such as SOF_4. They also reported that the generated molecules had better average validity, diversity, and stability than the baseline over different generation tasks. However, this model yielded large RMSDs in the task using a larger bag containing more heavy atoms and a more complex constitution.

Case 5. Satorras et al.[43] have proposed a probabilistic model obeying Euclidean symmetry based on continuous-time normalizing flows, which was applied to 3D molecular generation by sampling the atom types, charges and Cartesian coordinates in parallel. Specifically, they utilized E(n) equivariant graph neural networks, a neural architecture that is E(n) equivariant and enables invertible transformation, to model the dynamics function and enforce additional translation invariance. In the context of molecular generation, the charge and atom type were converted to the continuous form, and the size of the molecule was determined prior to the sampling of structures.

The model was trained on molecules from the QM9 data set. Compared with the nonequivariant variants, this model generalized well to unseen samples in the test set with a much lower negative log likelihood. However, the validity or "mol stability" of the structures generated was only 4.2%.

For Protein Structure-Based Design. *Case 1.* Skalic et al.[33] have introduced LigVoxel, a 3D convolutional neural network, which takes as inputs 3D grid representations of protein pockets to predict binding ligand grids. Their model could further depend on the amounts of the different pharmacophoric features required. The training set was 8808 protein−ligand complex structures curated from the 2013 release of the sc-PDB database.[52] One of the shortcomings of their approach is the failure to infer the molecular structures from the predicted grids, which limited their evaluation procedures to the grid-level.

When evaluated on the complexes in the test set, their model could predict ligand grids in the protein binding sites with better volumetric overlap with known binders than those from random baseline and a molecular interaction fields method. Their model also demonstrated virtual screening capability to recover close-to-crystal poses of the real binders using the predicted grids via their screening pipeline. Although the generated grids showed some similarity to those of known binders, the performance in designing novel ligands was not specified; therefore, the effectiveness of this model in *de novo* design scenarios remains unclear.

Case 2. Partially to extend their previous approach,[33] Skalic et al.[35] further proposed LiGANN, another structure-based ligand design pipeline. It contained a BicycleGAN with 3D convolutional blocks for generating proper ligand grids given the pocket grids and a shape captioning network, similar to that in their generative model in free space,[34] to parse the 3D ligand grids into SMILES strings. The model was trained on top-ranking redocked poses of 11,256 binding compounds from the DUD-E database. The binders were transformed into 3D grids with five channels, including hydrophobicity, aromaticity, hydrogen-bond donors, hydrogen-bond acceptors, and heavy atoms, whereas the pockets were featurized similarly with two additional channels: positive and negative Gasteiger partial charge. The independent test set consisted of three pharmaceutically interested targets: the delta opioid receptor (7TM) and two serine/threonine-protein kinases Chk1 and TNNI3K.

Compared with Ligvoxel,[33] their model generated more diverse and distinguishable shapes from the target pockets. For the three test proteins, they reported a success rate of 86.5%−93.8% when decoding SMILES strings from the ligand shapes. The distributions of the different property counts of the compounds inferred from their shapes were in good agreement with those of the recorded binders. To evaluate the binding probability of the generated molecules, the authors used both structure-based and ligand-based virtual screening methods. For the three test proteins, the generated molecules were docked into the protein pocket with smina.[53] The docking performance of the generated molecules was superior to that of decoy compounds randomly selected from the ZINC15 database. For another 31 targets from the DUD-E database, the authors used a LightGBM gradient boosting decision tree model for ligand-based virtual screening. Significant enrichment in the generated molecules relative to the decoys was found in the top predicted pK_d decile, with an enrichment factor between 1.5 and 1.9. Some examples of generated ligands for the adenosine receptor A2 were shown to have a similar scaffold as known binders. However, none of the binding candidates were experimentally validated.

Case 3. Ragoza et al.[32] have proposed a conditional VAE model combined with 3D convolutional layers to generate ligands in protein binding sites. Similar to the authors' previous approach,[31] molecules were voxelized into 3D grids with six property channels. The latent codes encoded from the protein−ligand complexes, together with the conditional labels encoded from the receptors alone, were provided to the decoder to generate 3D ligand grids. The ligand grids were then converted into molecular structures through atom-fitting and bond inference algorithms. The whole model was trained on the CrossDocked2020 data set and evaluated on 10 complex structures reserved for testing.

Two sampling modes from the latent space, posterior sampling and prior sampling, were tested. These modes differed according to whether the latent code was obtained from a real protein−ligand complex or a standard normal distribution. For posterior sampling, a validity of 98.5%, a novelty of 100%, and a uniqueness of 77% were obtained for the generated molecules. Posterior molecules were also quite dissimilar from the reference molecules (from the encoded complexes), with an average Tanimoto fingerprint similarity of 0.33. For prior sampling, the authors reported much higher uniqueness and dissimilarity (in both molecular fingerprint and 3D shape) for the generated samples relative to the reference molecules. A detailed comparison is provided in Table S1. The molecular weight distributions for the two sampling modes had a large overlap, with medians <300, but these were not directly compared with the distribution of the reference molecules. Concerning the quality of the 3D conformation, 81% and 91% of the conformational changes measured by RMSD were <2 Å during the universal force field (UFF) minimization. The authors also reported the distributions of some common bond lengths, bond angles and torsion angles. However, all of these geometric properties were measured after the generated molecules completed UFF minimization, and the performance before minimization is unknown. To assess the relative stability of the generated molecules in the receptor binding site, the

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

molecules were docked into the pockets using Vina, and the binding affinity was additionally predicted by an ensemble of CNN scoring functions. For molecules from posterior and prior sampling, 30.8% and 17.3% had lower minimized Vina energy, and 15.4% and 15.9% had greater predicted affinity than the reference molecules, respectively. Unfortunately, detailed performance across the 10 tested targets was not reported.

Interestingly, the authors tested the generation performance during conditioning on the mutated Shikimate kinase from *Mycobacterium tuberculosis*. They created 19 single-residue mutants, including known interacting and noninteracting residues, and four multiresidue mutants. Compared with the reference binder, the generated molecules showed reasonable changes in response to the mutated pocket environment. However, the authors also observed some discrepancies in the structural similarity and changes in binding potential with respect to the known binders. In addition, the representative generated molecules showed in this work were quite small, likely due to the CrossDocked2020 data set they used.

Case 4. Our group (Li et al.)[46] has introduced a novel model architecture, L-Net, to perform both unconditional and structure-based molecular design. Molecules with conformation information were processed by a state encoder consisting of mainly graph convolutional layers organized into a U-Net structure. The continuous representation outputted by the state encoder was used by a subsequent policy network to perform several types of edits on partially completed molecular structures. Because of the introduction of an internal coordinate system, L-Net can generate 3D molecules regardless of positional or orientational variation.

After being trained on drug-like compounds from the ChEMBL data set, L-Net achieved a validity of 94.3%, whereas this value for a test set from on the QM9 data set was 96%. The uniqueness of generated molecules approached 100%. Other molecular property distributions such as QED, molecular weight, and 3D shape also agreed with those of the molecules in the test set. After optimization with the MMFF94s force field, we reported an average RMSD value of 0.613 Å, which shows the high quality of the generated conformers. We also found close matches between the bond lengths, bond angles, and torsion angles of various substructures in generated molecules and those in the test set. The imperfections in the ring structures generated were also discussed.

To design molecules inside target binding sites using L-Net, we developed DeepLigBuilder, which combined our model with a Monte Carlo tree search (MCTS) with the docking score from smina as the reward. DeepLigBuilder was evaluated for the task of designing potential inhibitors for SARS-CoV-2 MPro. Starting from a seed fragment of a validated potent nonpeptide inhibitor that covalently binds SARS-CoV-2 MPro, DeepLigBuilder found novel, drug-like, and synthetically accessible molecules with improved docking scores compared with the original inhibitor. We also found that three key pharmacophoric features were properly captured, indicating a similar binding mode with the seed inhibitor.

We have also demonstrated the ability of DeepLigBuilder to design noncovalent inhibitors targeting SARS-CoV-2 MPro. A small fragment containing three heavy atoms in a newly reported noncovalent inhibitor was used as a starting point. DeepLigBuilder generated compounds with high affinity scores (i.e., smina score lower than -9 kcal mol^{-1}) with a success rate of 78.1%. The average QED of these generated molecules improved from 0.36 for the original inhibitor to 0.53, whereas the synthetic accessibility score of 2.7 was maintained. Similarly, three important pharmacophoric features were inspected and found to be covered well by the generated molecules. Bemis−Murcko scaffolds analysis showed that DeepLigBuilder could generate molecules with the privileged structural patterns found in the original inhibitor.

Challenges and Prospects. *Structural Inference.* Three types of featurization and representation of 3D molecules have been established. Room remains for the improvement and development of new representations. In the work mentioned above, the generative model learns to assemble a 3D molecule either atom-by-atom or in one shot. When the connectivity of atoms is absent from the learning process, the direct output of the model is a set of discrete atoms, and structural inference is needed before further evaluation, which is the current state of most 3D generative models. However, reliable structural inference is still an open question and most studies have resorted to suboptimal solutions with obvious limitations. For example, a captioning network is useful for parsing the sampled ligand grids into SMILES strings, but aligning each substructure in the decoded molecule with the original grid is challenging.[34,35] Other optimization algorithms can be used to fit a set of atoms into the ligand grids.[31] Toolkits, such as OpenBabel,[36] or rule-based methods using empirical bond lengths[43] have also been widely used in EDM- and Cartesian coordinate-based models for bonds completion. However, these approaches do not guarantee a 100% success rate, even when reconstructing real molecules from collections of discrete atoms or grid points.

Several methods may exist to address this urgent issue. First, new evaluation methods can be established to verify the chemical validity of generated molecules without relying on knowledge of bond information, which is a prerequisite for toolkits like RDKit. Second, new invertible representations of 3D molecules can be developed. For example, bond information can be explicitly encoded with the adjacency matrix, as Nesterov et al. did in 3DMolNet,[51] or an additional "bond channel" can be used for 3D grid representation. Third, directly handling molecules with 3D coordinates is another promising method for 3D molecular generation, exemplified by L-Net, which combines a graph neural network with a local coordinate system.[46]

Application to Protein Structure-Based Drug Design. Unlike SMILES-based or graph-based representations of a compound, the 3D structure is a more realistic physical model from which properties, such as intrinsic energy, can be derived. The use of a 3D molecule representation also allows consideration of the intermolecule interactions and binding affinity of the compound in complex with biological targets, which is the basis of SBDD. Therefore, when a generative model utilizing 3D molecular representation is used to perform SBDD, it benefits from not only the knowledge provided by known actives, which most SMILES-based or graph-based generative models rely on, but also the general binding rules of protein−ligand complexes.

The most important goal for 3D molecular generative models is the design of ligands that bind to specific targets, which is the reason for using a more challenging method instead of combining a SMILES- or graph-based model with conformation generation modules. However, most of the 3D molecular generative models reviewed above design com-

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

pounds in free space; only a few have tried to generate compounds in the target binding sites. Nonetheless, lessons can be learned from these early stage explorations which can be summarized into two plausible approaches to do SBDD with a 3D molecular generative model. The first approach uses supervised learning and builds a model conditioned on protein pockets. The works by Skalic et al.[33,35] and Ragoza et al.[32] fall into this category. However, their models require further structural inference, and the novel compounds generated require experimental validation.

The other approach to do SBDD with a 3D molecular generative model is to use reinforcement learning (RL), which does not require the preparation of a complex structure data set. The reward in RL training can be chosen from well-developed scoring functions to bias structure generation toward the desired targets, such as the estimated affinity score from smina or descriptor-based scoring functions boosted with deep learning algorithms. We have discussed some early efforts in this direction. Simm et al.[40,41] have trained a 3D generative model from scratch, receiving rewards from the stability of the molecular system evaluated by the semiempirical parametrized method 6, which could be replaced by a scoring function to provide a target-related reward for conditional generation. Our previous work[46] showed the effectiveness of biasing an unconditional 3D generative model toward specific targets through fine-tuning with MCTS.

Evaluation and Benchmarks. There is no unified and comprehensive benchmark for 3D molecule generative models. Researchers have evaluated the performances of their models and the quality of the generated molecules using many metrics, largely depending on the molecular representation. For example, LigVoxel[33] generated ligands in grid form, which limited its evaluation to the grid-level. The training data were curated from various sources (see Table S1 for data set summary); 3D conformers with the precision level of quantum chemistry were used in some work, whereas conformers generated by heuristic methods were used in others. Different receptors were chosen as targets when generating potential ligands in protein-binding sites. All of these discrepancies create difficulty in directly comparing any two models.

The evaluation of 3D generative models could follow the benchmarks for SMILES or graph-based generative models, such as MOSES[48] and GaucaMol,[54] for better comparison of 2D and 3D models. However, emphasis should be placed on the 3D aspects.

On the one hand, quality assessments of generated 3D conformers are important and may include comparisons of the distributions of bond lengths, bond angles, and torsion angles with those of the molecules from the test set. Many studies report the median RMSD of conformation changes after the generated molecules are optimized by molecular force fields, such as MMFF94.[55] However, reporting energy changes would represent a more rigorous assessment, as RMSD may underestimate, for example, an unnatural bond length, which should be penalized for high energy. This approach is also adopted by most advanced studies of molecular conformation generation to directly assess the quality of the generated conformations.[56]

On the other hand, a 3D generative model that produces diverse conformers is more desirable and more generalizable, which is often omitted by most 3D generative models. Currently, these models are trained to produce stable conformations of molecules, which allows them to learn the rules of stereochemistry, but they may perform poorly when conformational transition is needed, as is observed when molecules are sampled from a trained latent space intended to encode both topological and conformational information.[34,57]

The generation of diverse conformers by 3D generative models, especially those conditioned on receptor structures, is necessary to allow these models to be generalized to different target receptors. It is because ligands typically adopt different conformations to bind different receptors, which is also essential for multitarget ligand design.[26] Moreover, ligands tend to reorganize their conformations upon binding, deviating from global or local minimum conformations.[58] This behavior also requires that 3D generative models have sufficient capacity to produce appropriate conformers.

Finally, a complete evaluation procedure should include the validation of generated binders by wet-lab experiments. None of the molecules obtained from current 3D generative models have been tested in experiments, partly because the generative models tend to produce molecules with low synthetic accessibility, which likely results from the training data they used that is less druglike and hard to synthesize. This has been shown as one of the key factors to influence the synthesizability of generated molecules.[59]

DATA AND SOFTWARE AVAILABILITY

We have listed the code accessibility of each method reviewed in Table S2 and have provided links, but we make no guarantees to their functionalities. Where accessible links are absent, we encourage readers to contact the original authors for more information about the software and how to access it.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00042.

A summary of 3D molecular generative models is shown in Table S1. Code accessibility of each method reviewed is in Table S2 (PDF)

AUTHOR INFORMATION

Corresponding Author

Jianfeng Pei − *Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China;* orcid.org/0000-0002-8482-1185; Email: jfpei@pku.edu.cn

Authors

Weixin Xie − *Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China*

Fanhao Wang − *Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China*

Yibo Li − *Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China*

Luhua Lai − *Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China; Peking-Tsinghua Center for Life Science at BNLMS, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China;* orcid.org/0000-0002-8343-7587

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.2c00042

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (grants 22033001 and 21673010) and the National Science and Technology Major Project "Key New Drug Creation and Manufacturing Program", China (grant 2018ZX09711002).

■ REFERENCES

(1) Schneider, G.; Clark, D. E. Automated De Novo Drug Design: Are We Nearly There Yet? *Angew. Chem., Int. Ed.* **2019**, *58*, 10792−10803.
(2) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comp. Sci.* **1988**, *28*, 31−36.
(3) Bjerrum, E. J.; Threlfall, R. Molecular Generation with Recurrent Neural Networks (RNNs). *arXiv (Machine Learning)*, May 17, 2017, arXiv:1705.04612, ver. 2. https://arxiv.org/abs/1705.04612 (accessed 2022-05-11).
(4) Gupta, A.; Müller, A. T.; Huisman, B. J. H.; Fuchs, J. A.; Schneider, P.; Schneider, G. Generative Recurrent Networks for De Novo Drug Design. *Mol. Inform.* **2018**, *37*, 1700111.
(5) Segler, M. H.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Sci.* **2018**, *4*, 120−131.
(6) Li, Y.; Vinyals, O.; Dyer, C.; Pascanu, R.; Battaglia, P. Learning Deep Generative Models of Graphs. *arXiv (Machine Learning)*, March 8, 2018, 1803.03324, ver. 1. https://arxiv.org/abs/1803.03324 (accessed 2022-05-11).
(7) Simonovsky, M.; Komodakis, N. GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders. *arXiv (Machine Learning)*, February 8, 2018, 1802.03480, ver. 1. https://arxiv.org/abs/1802.03480 (accessed 2022-05-11).
(8) Xu, Y.; Lin, K.; Wang, S.; Wang, L.; Cai, C.; Song, C.; Lai, L.; Pei, J. Deep learning for molecular generation. *Future Med. Chem.* **2019**, *11*, 567−597.
(9) Tan, X.; et al. Automated design and optimization of multitarget schizophrenia drug candidates by deep learning. *Eur. J. Med. Chem.* **2020**, *204*, 112572.
(10) Amabilino, S.; Pogány, P.; Pickett, S. D.; Green, D. V. S. Guidelines for Recurrent Neural Network Transfer Learning-Based Molecular Generation of Focused Libraries. *J. Chem. Inf. Model.* **2020**, *60*, 5699−5713.
(11) Moret, M.; Helmstädter, M.; Grisoni, F.; Schneider, G.; Merk, D. Beam Search for Automated Design and Scoring of Novel ROR Ligands with Machine Intelligence**. *Angew. Chem., Int. Ed.* **2021**, *60*, 19477−19482.
(12) Grow, C.; Gao, K.; Nguyen, D. D.; Wei, G.-W. Generative network complex (GNC) for drug discovery. *Commun. Inf. Syst.* **2019**, *19*, 241.
(13) Sattarov, B.; Baskin, I. I.; Horvath, D.; Marcou, G.; Bjerrum, E. J.; Varnek, A. De Novo Molecular Design by Combining Deep Autoencoder Recurrent Neural Networks with Generative Topographic Mapping. *J. Chem. Inf. Model.* **2019**, *59*, 1182−1196.
(14) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminformatics* **2017**, *9*, 48.
(15) Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: An AI Tool for de Novo Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 5918−5922.
(16) Zhavoronkov, A.; Ivanenkov, Y. A.; Aliper, A.; Veselov, M. S.; Aladinskiy, V. A.; Aladinskaya, A. V.; Terentiev, V. A.; Polykovskiy, D. A.; Kuznetsov, M. D.; Asadulaev, A.; et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. biotechnol.* **2019**, *37*, 1038−1040.
(17) Grechishnikova, D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Sci. Rep.* **2021**, *11*, 321.
(18) Xu, M.; Ran, T.; Chen, H. De novo molecule design through the molecular generative model conditioned by 3D information of protein binding sites. *J. Chem. Inf. Model.* **2021**, *61*, 3240−3254.
(19) Imrie, F.; Bradley, A. R.; van der Schaar, M.; Deane, C. M. Deep generative models for 3D linker design. *J. Chem. Inf. Model.* **2020**, *60*, 1983−1995.
(20) Imrie, F.; Hadfield, T. E.; Bradley, A. R.; Deane, C. M. Deep generative design with 3D pharmacophoric constraints. *Chem. Sci.* **2021**, *12*, 14577−14589.
(21) Sousa, T.; Correia, J.; Pereira, V.; Rocha, M. Generative deep learning for targeted compound design. *J. Chem. Inf. Model.* **2021**, *61*, 5343−5361.
(22) Merz, K. M.; De Fabritiis, G.; Wei, G.-W. Generative Models for Molecular Design. *J. Chem. Inf. Model.* **2020**, *60* (12), 5635−5636.
(23) Nguyen, D. D.; Gao, K.; Wang, M.; Wei, G.-W. MathDL: mathematical deep learning for D3R Grand Challenge 4. *J. Comput. Aid. Mol. Des.* **2020**, *34*, 131−147.
(24) Wang, R.; Gao, Y.; Lai, L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. *Molecular modeling annual* **2000**, *6*, 498−516.
(25) Yuan, Y.; Pei, J.; Lai, L. LigBuilder 2: a practical de novo drug design approach. *J. Chem. Inf. Model.* **2011**, *51*, 1083−1091.
(26) Yuan, Y.; Pei, J.; Lai, L. Ligbuilder v3: a multi-target de novo drug design approach. *Front. Chem.* **2020**, *8*, 142.
(27) Gentile, F.; Agrawal, V.; Hsing, M.; Ton, A.-T.; Ban, F.; Norinder, U.; Gleave, M. E.; Cherkasov, A. Deep docking: a deep learning platform for augmentation of structure based drug discovery. *ACS Central Sci.* **2020**, *6*, 939−949.
(28) Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; Tang, J. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. Proceedings from the *International Conference on Learning Representations*, virtual, April 25, 2022; ICLR: La Jolla,CA, 2022.
(29) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein−ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* **2017**, *57*, 942−957.
(30) Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A. S.; De Fabritiis, G. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics* **2017**, *33*, 3036−3042.
(31) Ragoza, M.; Masuda, T.; Koes, D. R. Learning a Continuous Representation of 3D Molecular Structures with Deep Generative Models. *arXiv (Quantitative Methods)*, November 15, 2020, 2010.08687, ver. 3. https://arxiv.org/abs/2010.08687 (accessed 2022-05-11).
(32) Ragoza, M. T.; Masuda, T.; Koes, D. R. Generating 3D Molecules Conditional on Receptor Binding Sites with Deep Generative Models. *Chem. Sci.* **2022**, *13*, 2701.
(33) Skalic, M.; Varela-Rial, A.; Jiménez, J.; Martínez-Rosell, G.; De Fabritiis, G. LigVoxel: inpainting binding pockets using 3D-convolutional neural networks. *Bioinformatics* **2019**, *35*, 243−250.
(34) Skalic, M.; Jiménez, J.; Sabbadin, D.; De Fabritiis, G. Shape-based generative modeling for de novo drug design. *J. Chem. Inf. Model.* **2019**, *59*, 1205−1214.
(35) Skalic, M.; Sabbadin, D.; Sattarov, B.; Sciabola, S.; De Fabritiis, G. From target to drug: generative modeling for the multimodal structure-based ligand design. *Mol. Pharmaceut.* **2019**, *16*, 4282−4291.
(36) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminformatics* **2011**, *3*, 33.
(37) Hoffmann, M.; Noé, F. Generating valid Euclidean distance matrices. *arXiv (Machine Learning)*, November 15, 2019, 1910.03131, ver. 2. https://arxiv.org/abs/1910.03131 (accessed 2022-05-11).

(38) Gebauer, N. W.; Gastegger, M.; Schütt, K. T. Generating equilibrium molecules with deep neural networks. *arXiv (Machine Learning)*, 1810.11347, October 26, 2018, ver. 1. https://arxiv.org/abs/1810.11347 (accessed 2022-05-11).
(39) Gebauer, N. W.; Gastegger, M.; Schütt, K. T. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. *arXiv (Machine Learning)*, January 9, 2020, 1906.00957, ver. 3. https://arxiv.org/abs/1906.00957 (accessed 2022-05-11).
(40) Simm, G.; Pinsler, R.; Hernández-Lobato, J. M. Reinforcement learning for molecular design guided by quantum mechanics. Proceedings from the *37th International Conference on Machine Learning*, online, July 13−18, 2020; Microtome Publishing: Brookline, MA, 2020; pp 8959−8969.
(41) Simm, G. N.; Pinsler, R.; Csányi, G.; Hernández-Lobato, J. M. Symmetry-aware actor-critic for 3d molecular design. *arXiv (Machine Learning)*, November 25, 2020, 2011.12747, ver. 1. https://arxiv.org/abs/2011.12747 (accessed 2022-05-11).
(42) Joshi, R. P.; Gebauer, N. W.; Bontha, M.; Khazaieli, M.; James, R. M.; Brown, J. B.; Kumar, N. 3D-Scaffold: A Deep Learning Framework to Generate 3D Coordinates of Drug-like Molecules with Desired Scaffolds. *J. Phys. Chem. B* **2021**, *125*, 12166−12176.
(43) Garcia Satorras, V.; Hoogeboom, E.; Fuchs, F.; Posner, I.; Welling, M. E(n) Equivariant Normalizing Flows. *arXiv (Machine Learning)*, January 14, 2022, 2105.09016v4, ver. 4. https://arxiv.org/abs/2105.09016 (accessed 2022-05-11).
(44) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *arXiv (Machine Learning)*, December 19, 2017, 1706.08566m ver. 5. https://arxiv.org/abs/1706.08566 (accessed 2022-05-11).
(45) Schutt, K.; Kessel, P.; Gastegger, M.; Nicoli, K.; Tkatchenko, A.; Muller, K.-R. SchNetPack: A deep learning toolbox for atomistic systems. *J. Chem. Theory Comput.* **2019**, *15*, 448−455.
(46) Li, Y.; Pei, J.; Lai, L. Structure-based de novo drug design using 3D deep generative models. *Chem. Sci.* **2021**, *12*, 13664−13675.
(47) *RDKit: Open-source cheminformatics.* https://www.rdkit.org (accessed 2022-05-11).
(48) Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M.; et al. Molecular sets (MOSES): a benchmarking platform for molecular generation models. *Front. Pharmacol.* **2020**, *11*, 1931.
(49) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Sci.* **2018**, *4*, 268−276.
(50) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.
(51) Nesterov, V.; Wieser, M.; Roth, V. 3DMolNet: a generative network for molecular structures. *arXiv (Biomolecules)*, October 8, 2020, 2010.06477, ver. 1. https://arxiv.org/abs/2010.06477 (accessed 2022-05-11).
(52) Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E. sc-PDB: a 3D-database of ligandable binding sites—10 years on. *Nucleic Acids Res.* **2015**, *43*, D399−D404.
(53) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.* **2013**, *53*, 1893−1904.
(54) Brown, N.; Fiscato, M.; Segler, M. H.; Vaucher, A. C. GuacaMol: benchmarking models for de novo molecular design. *J. Chem. Inf. Model.* **2019**, *59*, 1096−1108.
(55) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(56) Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; Tang, J. GeoDiff: a Geometric Diffusion Model for Molecular Conformation Generation. *arXiv (Machine Learning)*, March 6, 2022, 2203.02923, ver. 1. https://arxiv.org/abs/2203.02923 (accessed 2022-05-11).
(57) Masuda, T.; Ragoza, M.; Koes, D. R. Generating 3D molecular structures conditional on a receptor binding site with deep generative models. *arXiv (Chemical Physics)*, November 23, 2020, 2010.14442, ver. 3. https://arxiv.org/abs/2010.14442 (accessed 2022-05-11).
(58) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.
(59) Gao, W.; Coley, C. W. The synthesizability of molecules proposed by generative models. *J. Chem. Inf. Model.* **2020**, *60*, 5714−5723.

https://doi.org/10.1021/acs.jcim.2c00042
J. Chem. Inf. Model. 2022, 62, 2269−2279