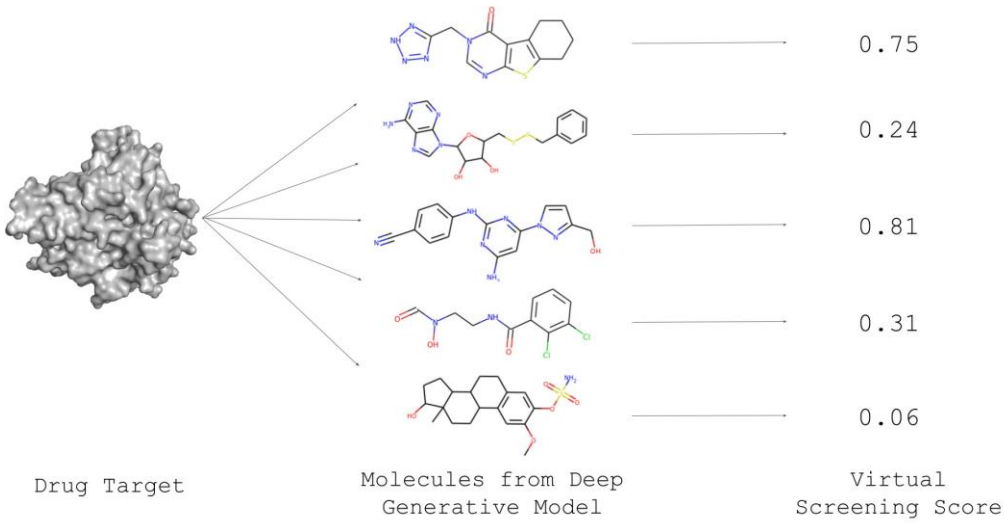
Graphical Abstract

AI in 3D Compound Design

Thomas E. Hadfield, Charlotte M. Deane



Highlights

AI in 3D Compound Design

Thomas E. Hadfield, Charlotte M. Deane

- Deep generative models are increasingly being proposed as a method for compound design.
- Models which incorporate 3D target-specific information are better able to design molecules which are complementary to the binding site.
- Molecules generated *in-silico* can rapidly be assessed by structure-based virtual screening models.
- To allow integration into drug discovery campaigns, generative and virtual screening models need to be interpretable.

AI in 3D Compound Design

Thomas E Hadfield[†], Charlotte M. Deane[†*]

[†]*Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, UK*

*deane@stats.ox.ac.uk

Abstract

The success of Artificial Intelligence (AI) across a wide range of domains has fuelled significant interest in its application to designing novel compounds and screening compounds against a specific target. However, many existing AI methods either do not account for the 3D structure of the target at all or struggle to capture meaningful spatial information from the target. In this Opinion, we highlight a range of recent structure-aware approaches which utilise deep learning for compound design and virtual screening. We discuss how such methods can be better integrated into existing drug discovery pipelines by facilitating the design of compounds which conform to a specified design hypothesis and by uncovering key protein-ligand interactions which can be used to aid molecule design.

Keywords: Drug Discovery, 3D Compound Design, Deep Generative Models, Structure-Based Virtual Screening

Introduction

A recent study estimated that the median cost to develop a drug was $985 million and that, on average, it took 8.3 years to conduct the research and development needed before clinical trials [1]. The development of techniques which can deliver medicines to patients more quickly and at a lower cost is therefore of the utmost importance. Enabled by recent advances in computer hardware, machine learning (ML) algorithms have been successfully employed in a broad range of fields, including natural language processing, computer vision and protein structure prediction [2, 3, 4]. ML algorithms, in particular those based on deep neural networks, have a number of attractive properties: They are capable of identifying complex relationships between features with minimal feature preprocessing, they scale well to large datasets and can perform at a level which has been shown in some cases to be on par with or better than

those of human experts [5]. There is significant interest in the application of such deep learning algorithms to the drug discovery pipeline. One area where deep learning is just starting to show potential is 3D compound design.

Deep Generative Models for Compound Design

In drug discovery campaigns, molecule design relies heavily on the expertise of medicinal chemists, who leverage their understanding of chemical space and the target's structure to design ligands which form promising interactions. However, even with access to existing computational techniques, such as docking algorithms and molecular dynamics simulations, this process is costly and time consuming. Deep generative models are increasingly being proposed as an alternative method for molecular design. These methods aim to be able to rapidly generate sets of high-affinity binders; a schematic of a deep generative model for drug discovery is shown in Figure 1.

Early ML models for molecule generation represented molecules as SMILES strings [6], a series of ASCII characters which completely describe 2D chemical structure. These methods were frequently able to generate SMILES strings corresponding to valid molecules [7, 8, 9], but in early examples the lack of grammar constraints imposed on the model meant that a high proportion of strings produced by the model were chemically invalid [7]. More recent models have proposed using graphs to represent molecules (e.g. [10, 11, 12]), either constructing molecules 'atom-by-atom' or from a vocabulary of molecular building blocks; Jin et al. [10] demonstrated that their graph-based model generated molecules with superior properties to SMILES-based approaches [7, 8, 9].

An attractive property of deep generative models is that they allow the optimisation of molecules against a specific property, either by applying a search algorithm such as Bayesian Optimisation (e.g. [7, 10, 11]) or by training the model using reinforcement learning (e.g. [13, 14, 15]). Olivecrona et al. [13] and Zhavoronkov et al. [14] attempted to generate molecules which would be active against a specific protein target, but their methods relied heavily on the existence of molecules known to be active against the protein, limiting their applicability when designing

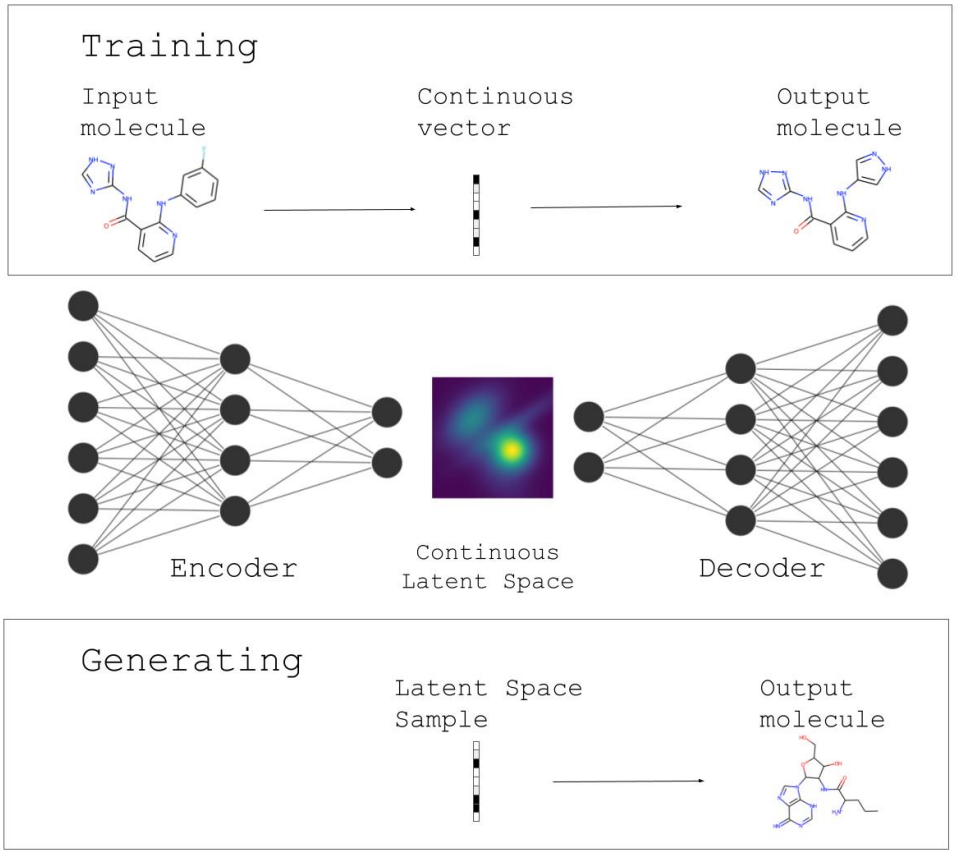


Figure 1: Schematic of a deep generative model for drug discovery. During training a molecule is encoded into a low-dimensional latent space and a decoder attempts to recreate the original molecule from the latent space representation. After the model is trained one can use the decoder to transform random samples in the latent space into novel molecules, and leverage search algorithms to identify latent space regions which correspond to molecules with desirable properties.

molecules for a novel target and potentially inducing the risk of designing molecules which are strikingly similar to existing actives used to train the model [16].

Instead of using known-actives as a basis for designing molecules against a particular target, a small number of methods have explicitly included protein specific 3D information into their generative processes, making them applicable to a far wider range of targets than models which require a library of known-actives. The first structure-aware generative model, LiGANN [17], generated a set of ligand shapes complementary to the binding pocket, using a shape captioning network to generate ligands with pharmacophores capable of binding to the protein. Compared to a random sample obtained from ZINC15 [18], LiGANN generated ligands with superior docking scores on a test set comprising 31 targets.

Masuda et al [19] also incorporated 3D information by encoding atomic density grids into separate latent representations for ligand and

protein and training a model to generate 3D ligands conditional on the protein structure.
Unlike other models, which typically generate molecules as a 2D graph or as a SMILES string, their model directly generated 3D atomic densities and translated them to 3D structures. The authors showed that conditioning upon the protein structure when generating molecules generally improved the 3D properties of the resulting molecules compared to those made by a similar model which did not account for protein structure.

In contrast to the approaches described above, which explicitly provided a representation of protein structure as an input to the model, Jeon and Kim [20] utilised reinforcement learning in the generation of molecules by incorporating a docking score into their reward function to incentivise the generation of molecules which bound with high affinity to the protein. Similarly, Bai et al. [21] proposed a genetic algorithm for generating molecules with a fitness score [22] based on Autodock Vina [23], which predicts the binding affinity between a protein and ligand. Whilst Jeon and Kim [20] demonstrated that their algorithm was able to make modifications to a molecule which improved its docking score, it is well established that docking scores do not correlate perfectly with binding affinities [24] potentially hindering the ability of the model to learn which protein-ligand interactions are most appropriate.

The approaches described above [17, 19, 20] are analogous to high-throughput screening, since any two molecules generated by the model can be completely dissimilar. A different strategy is to explicitly incorporate/build on known binders in the context of the protein. DeLinker, proposed by Imrie et al. [26], is a generative model for fragment linking which takes two fragments and 3D information describing the relative positions of the fragments as input and returns a linked molecule. DeLinker exhibited superior performance compared to a database-search baseline on recovering the ground-truth linker and was able to generate a greater proportion of molecules which were highly 3D similar to the ground-truth. In a follow-up study, Imrie et al [25] demonstrated that the imposition of user-specified 3D pharmacophoric constraints (Figure 2) yielded substantial improvements over the original DeLinker model and allowed a greater degree of control over the pharmacophoric profile of the generated molecules, allowing easier integration into real-world drug discovery campaigns in which a medicinal chemist typically formulates a design hypothesis and seeks to enumerate molecules which conform to that hypothesis.

The rapid proliferation of deep generative models has meant that it is difficult for potential users to assess which method, if any, they should use for a specific task. This difficulty is exacerbated by the fact that different

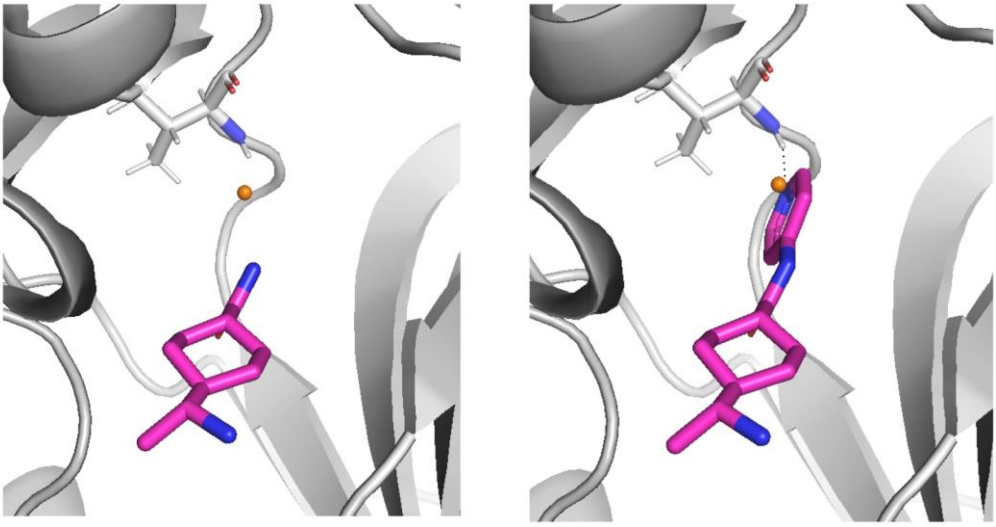


Figure 2: Example of a 3D Pharmacophoric Constraint employed by Imrie et al [25] on a Scaffold Elaboration task. Left: The orange ball represents the Hydrogen Bond Acceptor constraint provided to the model. Right: An elaborated molecule where a Hydrogen Bond Acceptor is placed close to the constraint, allowing it to make a Hydrogen Bond with the protein.

models are usually trained and tested using different sets of molecules and are evaluated using different metrics. There is therefore a clear need for a set of benchmarks which allow practitioners to easily assess which generative model is most suitable for their design task. MOSES [27] and GuacaMol [28] are recently published platforms which facilitate the comparison of different generative models across a range of distribution-learning and goal-directed benchmarks. However, existing goal-directed benchmarks focus on the generation of molecules which are highly similar to existing ligands rather than molecules which are likely to bind to a specific protein: In the next section we discuss several recently published virtual screening models which could potentially be incorporated into generative model assessment platforms.

Structure-Based Virtual Screening

The question of whether a specific molecule will bind to a protein of interest is a central question of drug discovery. When considering large sets of molecules, as is often the case when using generative models, it is important

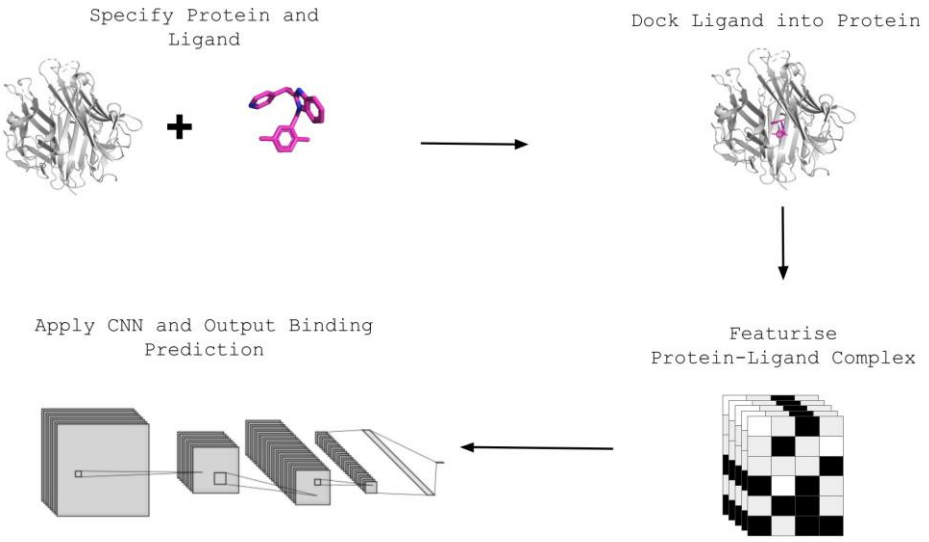


Figure 3: Schematic of protein-ligand scoring using Convolutional Neural Networks. A ligand is docked into the protein of interest, featurised and passed to a CNN. The CNN captures important spatial information and predicts whether the ligand is likely to bind to the protein in the pose provided to the model.

to be able to cheaply and accurately assess which molecules will bind to the protein in order to avoid synthesising large numbers of inactive molecules. Molecular docking tools (e.g. [29, 30, 31, 23]) utilise empirical or knowledge-based scoring functions to predict the correct binding mode of a ligand and estimate the binding affinity of the protein-ligand complex. Docking protocols are widely used in virtual screening campaigns to distinguish binders from non-binders and rapidly prioritise a small number of promising molecules from a much larger library. There has been significant interest in using machine learning methods for virtual screening; their ability to learn complex relationships from data without explicit parametric assumptions makes them an attractive alternative to classical scoring functions. Early applications of ML algorithms (e.g. [32, 33]) took a protein-ligand interaction fingerprint as input and classified the ligand as a binder or non-binder; whilst these models were often reported as exhibiting superior predictive accuracy to classical scoring functions, subsequent studies raised concerns surrounding their lack of sensitivity to changes in pose and their ability to generalise to new targets [34, 35].

Inspired by the ability of convolutional neural networks (CNNs) to capture important spatial information on image recognition tasks, several authors sought to predict protein-ligand binding by using a docked protein ligand complex as the input to a CNN. Compared to existing fingerprint based virtual screening models, it was hypothesized that CNN-based models, illustrated in Figure 3, would be better able to predict

binding affinity in a similar way to how an expert might, by identifying ligand pharmacophores which have the potential to form interactions with protein residues. Pereira et al [36] computed a set of features (distances, partial atomic charges etc.) for each atom and employed a CNN to capture the spatial information in these features and classify molecules as actives or decoys. Three recent methods have utilised voxelised representations of the protein and ligand, with each voxel containing information regarding the presence or absence of different atom types [37, 38, 39]. Two of these methods, [37, 39], demonstrated superior predictive performance on a virtual screening task compared to Autodock Vina [23] and existing machine learning scoring functions when the test set was derived from the DUD-E set [40]. However, the models performed comparably with existing scoring functions when tested on the MUV [41] dataset, suggesting that, like other methods, these CNN models struggled to generalise to novel targets. A follow-up study [42] illustrated that the strong performance exhibited by models trained and tested on the DUD-E dataset may well be driven by hidden biases in DUD-E which had allowed the model to detect systematic differences between the actives and decoys, rather than any ability of such models to learn protein-ligand interactions from the data. In fact, it was shown that if the protein structure was removed from the input to these models only minimal degradation in performance was observed [42, 43]. To address this issue, Scantlebury et al. [43] proposed a dataset augmentation technique where the set of decoys used to train the model was augmented by active ligands which had been assigned random conformations and randomly rotated and translated. Under this formulation, to discriminate between the true actives and the perturbed actives labelled as decoys, the model would be forced to consider local protein structure. The performance of the model proposed by Scantlebury et al. was degraded substantially by the omission of the protein structure, illustrating that this model was more dependent on protein structure for making predictions.

Whilst protein structure is playing an increasingly central role in deep learning-based approaches to both compound design and protein-ligand scoring, in general the models are provided with a single, static structure,

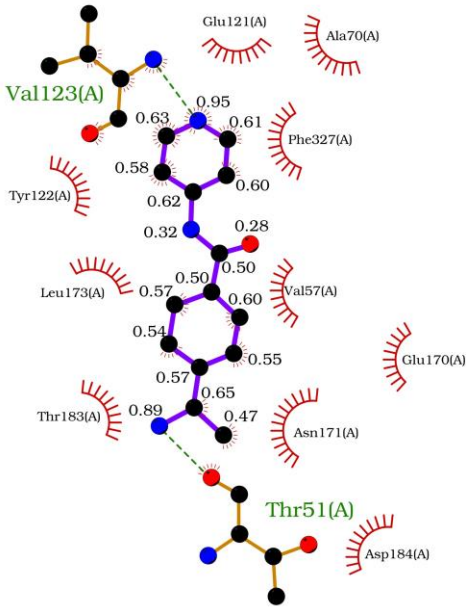


Figure 4: Example visualisation of attribution in protein-ligand scoring for cAMP-dependent Protein Kinase A in complex with ligand Y27 (PDB ID: 1Q8T), generated using LIGPLOT [45]. Atoms with scores close to 1 are considered to make an important contribution to the ligand binding to the protein; conversely, atoms with a score close to 0 are considered to reduce the probability of binding.

restricting the ability of models to account for binding site flexibility. There has been considerable interest in the application of machine learning algorithms to conducting rapid molecular simulations in order to account for protein flexibility (see [44] for a recent review). The incorporation of such simulations into compound design and protein-ligand scoring pipelines will be an important avenue of future research.

A key step in building greater trust in structure-based virtual screening methods is the development of methods which allow users to understand which features of a ligand meant that the model predicted it would bind (Figure 4 illustrates an example protein-ligand complex which has been assigned such attributions). Hochuli et al [46] proposed several methods for visualising the influence specific atoms had on a CNN model's classification, for example by predicting the binding probability using the whole protein-ligand complex and using the whole complex with one atom removed; the difference in the two scores is then considered to be the 'contribution' of that atom to the model's binding decision. Brown et al [47] proposed a method which takes a congeneric ligand series as input and predicts which substructures improve or degrade binding affinity relative to other ligands in the series, allowing

easier iterative refinement. Additional methods for attribution have been proposed [48, 49], although they do not consider the problem of protein-ligand binding and instead restrict their focus to the identification of pre-defined substructures within a molecule. Sanchez-Lengeling et al. [50] recently proposed a set of metrics for the evaluation of attribution methods, with the aim of facilitating the development of more interpretable deep-learning models. The development of such methods is an important step in the further integration of protein-ligand scoring models into drug discovery campaigns, as existing models currently only predict whether a specific ligand will bind to a protein but give no insights as to which motifs play a key role in allowing binding to occur; accurate methods for attribution would more easily allow practitioners to retain important motifs and discard those which do not help the ligand to bind, and could be used to inform the kinds of molecules made by generative models.

Conclusion

Although the long-term hope of those developing generative models is to fully automate the drug discovery process by inputting a protein structure and receiving a highly potent ligand a short time later, it is clear that for the foreseeable future human experts will remain an indispensable part of molecule design. Generative models therefore need to be easy to access and use and able to tailor the molecules they generate to conform to one or more specified design hypotheses. Similarly, as well as being able to distinguish binders from non-binders, virtual screening models need to highlight important interactions which can allow chemists and generative models to develop more promising molecules. A key step in this direction is the creation of large, unbiased virtual screening datasets, to allow models to learn important biophysical interactions, rather than overfitting to ligand-specific biases.

Acknowledgements

Funding: This work was supported by the Engineering and Physical Sciences Research Council, LifeArc, F. Hoffman-La Roche AG, and UCB Pharma (grant number EP/L016044/1).

References

[1] O. J. Wouters, M. McKee, J. Luyten, Estimated research and development investment needed to bring a new medicine to market, 2009-2018, Jama 323 (9) (2020) 844–853.

[2] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, arXiv preprint arXiv:2005.14165 (2020).

[3] K. Simonyan, A. Zisserman, Very deep convolutional networks for largescale image recognition, arXiv preprint arXiv:1409.1556 (2014).

[4] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Z´ıdek, A. W. Nelson, A. Bridgland, et al., Improved protein structure prediction using potentials from deep learning, Nature 577 (7792) (2020) 706–710.

[5] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al., Mastering the game of Go without human knowledge, Nature 550 (7676) (2017) 354–359.

[6] D. Weininger, Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules, Journal of Chemical Information and Computer Sciences 28 (1) (1988) 31–36.

[7] R. Gomez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. HernandezLobato, B. Sanchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules, ACS Central Science 4 (2) (2018) 268–276.

[8] M. J. Kusner, B. Paige, J. M. Hernandez-Lobato, Grammar variational autoencoder, in: International Conference on Machine Learning, PMLR, 2017, pp. 1945–1954.

[9] H. Dai, Y. Tian, B. Dai, S. Skiena, L. Song, Syntax-directed variational autoencoder for structured data, arXiv preprint arXiv:1802.08786 (2018).

[10] W. Jin, R. Barzilay, T. Jaakkola, Junction tree variational autoencoder for molecular graph generation, in: International Conference on Machine Learning, PMLR, 2018, pp. 2323–2332.

[11] Q. Liu, M. Allamanis, M. Brockschmidt, A. L. Gaunt, Constrained graph variational autoencoders for molecule design, arXiv preprint arXiv:1805.09076 (2018).

[12] Y. Li, J. Hu, Y. Wang, J. Zhou, L. Zhang, Z. Liu, Deepscaffold: A comprehensive tool for scaffold-based de novo drug discovery using deep learning, Journal of Chemical Information and Modeling 60 (1) (2019) 77–91.

[13] M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, Molecular de-novo design through deep reinforcement learning, Journal of Cheminformatics 9 (1) (2017) 1–14.

[14] A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev, et al., Deep learning enables rapid identification of potent DDR1 kinase inhibitors, Nature Biotechnology 37 (9) (2019) 1038–1040.

[15] Z. Zhou, S. Kearnes, L. Li, R. N. Zare, P. Riley, Optimization of molecules via deep reinforcement learning, Scientific Reports 9 (1) (2019) 1–10.

[16] W. P. Walters, M. Murcko, Assessing the impact of generative AI on medicinal chemistry, Nature Biotechnology 38 (2) (2020) 143–145.

[17] M. Skalic, D. Sabbadin, B. Sattarov, S. Sciabola, G. De Fabritiis, From target to drug: Generative modeling for the multimodal structure-based ligand design, Molecular Pharmaceutics 16 (10) (2019) 4282–4291.

The first generative model to incorporate 3D structural information extracted directly from the protein.

[18] T. Sterling, J. J. Irwin, ZINC 15–ligand discovery for everyone, Journal of Chemical Information and Modeling 55 (11) (2015) 2324–2337.

[19] T. Masuda, M. Ragoza, D. R. Koes, Generating 3d molecular structures conditional on a receptor binding site with deep generative models, arXiv preprint arXiv:2010.14442 (2020).

[20] W. Jeon, D. Kim, Autonomous molecule generation using reinforcement learning and docking to develop potential novel inhibitors, Scientific Reports 10 (1) (2020) 1–11.

[21] Q. Bai, S. Tan, T. Xu, H. Liu, J. Huang, X. Yao, Molaical: a soft tool for 3d drug design of protein targets by artificial intelligence and classical algorithm, Briefings in bioinformatics 22 (3) (2021)

[22] R. Quiroga, M. A. Villarreal, Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening, PloS one 11 (5) (2016).

[23] O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, Journal of Computational Chemistry 31 (2) (2010) 455–461.

[24] T. Pantsar, A. Poso, Binding affinity via docking: fact and fiction, Molecules 23 (8) (2018) 1899.

[25] F. Imrie, T. E. Hadfield, A. R. Bradley, C. M. Deane, Deep Generative Design with 3D Pharmacophoric Constraints, Chemical Science, 12 (2021) 14577-14589.

Demonstrated how the specification of pharmacophoric constraints allowed greater control over the molecules generated.

[26] F. Imrie, A. R. Bradley, M. van der Schaar, C. M. Deane, Deep generative models for 3d linker design, Journal of Chemical, Information and Modeling 60 (4) (2020) 1983–1995.

[27] D. Polykovskiy, A. Zhebrak, B. Sanchez-Lengeling, S. Golovanov, O. Tatanov, S. Belyaev, R. Kurbanov, A. Artamonov, V. Aladinskiy, M. Veselov, et al., Molecular sets (MOSES): a benchmarking platform for molecular generation models, Frontiers in Pharmacology 11 (2020).

[28] N. Brown, M. Fiscato, M. H. Segler, A. C. Vaucher, GuacaMol: benchmarking models for de novo molecular design, Journal of Chemical Information and Modeling 59 (3) (2019) 1096–1108.

Benchmarking platform which allows comparison between different generative models.

[29] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, Improved protein–ligand docking using GOLD, Proteins: Structure, Function, and Bioinformatics 52 (4) (2003) 609–623.

[30] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, Journal of Medicinal Chemistry 47 (7) (2004) 1739–1749.

[31] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, Journal of Computational Chemistry 30 (16) (2009) 2785–2791.

[32] P. J. Ballester, J. B. Mitchell, A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking, Bioinformatics 26 (9) (2010) 1169–1175.

[33] J. D. Durrant, J. A. McCammon, NNScore 2.0: a neural-network receptor–ligand scoring function, Journal of Chemical Information and Modeling 51 (11) (2011) 2897–2903.

[34] J. Gabel, J. Desaphy, D. Rognan, Beware of Machine Learning-Based Scoring Functions On the Danger of Developing Black Boxes, Journal of Chemical Information and Modeling 54 (10) (2014) 2807–2815.

[35] C. Kramer, P. Gedeck, Leave-cluster-out cross-validation is appropriate for scoring functions derived from diverse protein data sets, Journal of Chemical Information and Modeling 50 (11) (2010) 1961–1969.

[36] J. C. Pereira, E. R. Caffarena, C. N. Dos Santos, Boosting docking-based virtual screening with deep learning, Journal of Chemical Information and Modeling 56 (12) (2016) 2495–2506.

[37] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, D. R. Koes, Protein–ligand scoring with convolutional neural networks, Journal of Chemical Information and Modeling 57 (4) (2017) 942–957.

[38] J. Jiménez, M. Skalic, G. Martinez-Rosell, G. De Fabritiis, K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks, Journal of Chemical Information and Modeling 58 (2) (2018) 287–296.

[39] F. Imrie, A. R. Bradley, M. van der Schaar, C. M. Deane, Protein family specific models using deep neural networks and transfer learning improve virtual screening and highlight the need for more data, Journal of Chemical Information and Modeling 58 (11) (2018) 2319–2330.

[40] M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking, Journal of Medicinal Chemistry 55 (14) (2012) 6582– 6594.

[41] S. G. Rohrer, K. Baumann, Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data, Journal of Chemical Information and Modeling 49 (2) (2009) 169–184.

[42] L. Chen, A. Cruz, S. Ramsey, C. J. Dickson, J. S. Duca, V. Hornak, D. R. Koes, T. Kurtzman, Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening, PloS One 14 (8) (2019) e0220113.

Demonstrated that several structure-based virtual screening models did not perform markedly worse when structural information was removed, suggesting they were learning to classify ligands based on biases in the training datasets rather than by learning important protein-ligand interactions.

[43] J. Scantlebury, N. Brown, F. Von Delft, C. M. Deane, Data Set Augmentation Allows Deep Learning-Based Virtual Screening to Better Generalize to Unseen Target Classes and Highlight Important Binding Interactions, Journal of Chemical Information and Modeling 60 (8) (2020) 3722–3730.

[44] F. Noé, A. Tkatchenko, K.-R. Muller, C. Clementi, Machine learning for molecular simulation, Annual Review of Physical Chemistry 71 (2020) 361–390.

[45] A. C. Wallace, R. A. Laskowski, J. M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions, Protein Engineering, Design and Selection 8 (2) (1995) 127–134.

[46] J. Hochuli, A. Helbling, T. Skaist, M. Ragoza, D. R. Koes, Visualizing convolutional neural network protein-ligand scoring, Journal of Molecular Graphics and Modelling 84 (2018) 96–108.

[47] B. P. Brown, J. Mendenhall, A. R. Geanes, J. Meiler, General Purpose Structure-Based drug discovery neural network score functions with human-interpretable pharmacophore maps, Journal of Chemical Information and Modeling 61 (2) (2021) 603–620.

[48] K. McCloskey, A. Taly, F. Monti, M. P. Brenner, L. J. Colwell, Using attribution to decode binding mechanism in neural network models for chemistry, Proceedings of the National Academy of Sciences 116 (24) (2019) 11624–11629.

Graph-based model which allowed the specification of how important each atom was in making a prediction.

[49] V. Sundar, L. Colwell, Attribution Methods Reveal Flaws in Fingerprint-Based Virtual Screening, arXiv preprint arXiv:2007.01436 (2020).

B. Sanchez-Lengeling, J. Wei, B. Lee, E. Reif, P. Wang, W. Qian, K. McCloskey, L. Colwell, A. Wiltschko, Evaluating Attribution for Graph Neural Networks, in: H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, Vol. 33, Curran Associates, Inc., 2020, pp. 5898–5910