Structure-based Drug Design with Equivariant Diffusion Models

Arne Schneuing[1*], **Yuanqi Du**[2*], **Charles Harris**[3], **Arian Jamasb**[3], **Ilia Igashov**[1], **Weitao Du**[4], **Tom Blundell**[3], **Pietro Liò**[3], **Carla Gomes**[2], **Max Welling**[5], **Michael Bronstein**[6] **& Bruno Correia**[1]
[1]École Polytechnique Fédérale de Lausanne, [2]Cornell University, [3]University of Cambridge, [4]USTC, [5]Microsoft Research AI4Science, [6]University of Oxford

Abstract

Structure-based drug design (SBDD) aims to design small-molecule ligands that bind with high affinity and specificity to pre-determined protein targets. Traditional SBDD pipelines start with large-scale docking of compound libraries from public databases, thus limiting the exploration of chemical space to existent previously studied regions. Recent machine learning methods approached this problem using an atom-by-atom generation approach, which is computationally expensive. In this paper, we formulate SBDD as a 3D-conditional generation problem and present DiffSBDD, an E(3)-equivariant 3D-conditional diffusion model that generates novel ligands conditioned on protein pockets. Furthermore, we curate a new dataset of experimentally determined binding complex data from Binding MOAD to provide a realistic binding scenario that complements the synthetic CrossDocked dataset. Comprehensive in silico experiments demonstrate the efficiency of DiffSBDD in generating novel and diverse drug-like ligands that engage protein pockets with high binding energies as predicted by in silico docking.

1 Introduction

Structure-based drug design (SBDD), which aims to design small-molecule ligands that bind to target protein pockets, is a vital problem in drug discovery. Traditionally, SBDD is handled either by high-throughput experimental or virtual screening [1, 2] of large chemical databases. Not only is this expensive and time consuming but it also limits the exploration of chemical space to the historical knowledge of previously studied molecules, with a further emphasis usually placed on commercial availability [3]. Moreover, the optimization of initial lead molecules is often a biased process, with heavy reliance on human intuition [4]. Recent advances in geometric deep learning, especially in modeling geometric structures of biomolecules [5, 6], provide a promising direction for structure-based drug design [7]. Even though utilizing deep learning as surrogate docking models has achieved remarkable progress [8, 9], deep learning-based design of ligands that bind to target proteins is still an open problem. Early attempts [10–12] either rely on nontrivial post-processing steps or formulate this as an atom-by-atom generation problem which makes a sequence-conditioning assumption over the generation process of molecules and model inference inefficient.

In this work, we develop an equivariant diffusion model for structure-based drug design (DiffSBDD) which, to the best of our knowledge, is the first of its kind. Specifically, we formulate SBDD as a 3D-conditioned generation problem where we aim to generate diverse ligands with high binding affinity for specific protein targets. We propose an $E(3)$-equivariant 3D-conditional diffusion model that

*Equal contribution. Correspondence to arne.schneuing@epfl.ch, yd392@cornell.edu, bruno.correia@epfl.ch

Machine Learning for Structural Biology Workshop, NeurIPS 2022.

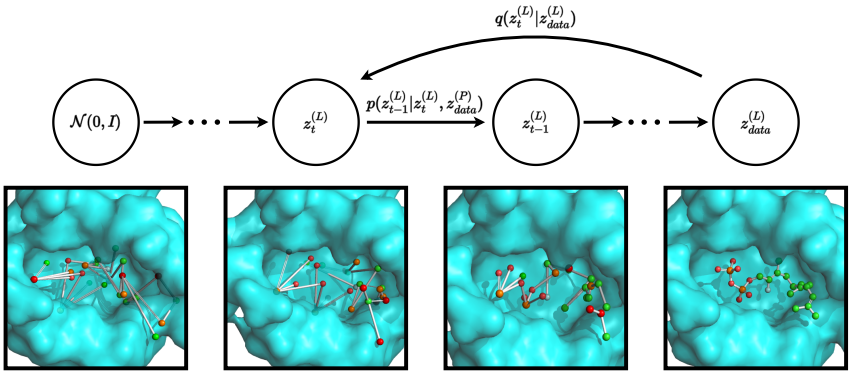


Figure 1: DiffSBDD in the protein-conditioned scenario. The conditional model learns to denoise molecules $\boldsymbol{z}^{(L)}$ in the fixed context of protein pockets $\boldsymbol{z}^{(P)}_{\text{data}}$.

respects translation, rotation, reflection, and permutation equivariance. We introduce two strategies, *protein-conditioned generation* and *ligand-inpainting generation* producing new ligands conditioned on protein pockets. Specifically, protein-conditioned generation considers the protein as a fixed context, while ligand-inpainting models the joint distribution of the protein-ligand complex and new ligands are inpainted during inference time. We also demonstrate that our model can be used for out-of-the-box for molecular optimization. We further curate an experimentally determined binding dataset derived from Binding MOAD [13], which supplements the commonly used synthetic CrossDocked [14] dataset to validate our model performance under realistic binding scenarios. The experimental results demonstrate that DiffSBDD is capable of generating novel, diverse and drug-like ligands with predicted high binding affinities to given protein pockets.

2 Equivariant Diffusion Models for SBDD

We utilize an equivariant Denoising Diffusion Probabilistic Model (DDPM) to generate molecules and binding conformations jointly with respect to a specific protein target. We represent protein and ligand point clouds as fully-connected graphs that are further processed by EGNNs [15]. We consider two distinct approaches to 3D pocket conditioning: (1) a conditional DDPM that receives a fixed pocket representation as context in each denoising step, and (2) a model that approximates the joint distribution of ligand-pocket pairs combined with inpainting at inference time.

2.1 Pocket-conditioned small molecule generation

In the conditional molecule generation setup, we provide a fixed three-dimensional context in each step of the denoising process. To this end, we supplement the ligand node point cloud $\boldsymbol{z}_t^{(L)}$, denoted by superscript L, with protein pocket nodes $\boldsymbol{z}^{(P)}_{\text{data}}$, denoted by superscript P, that remain unchanged throughout the reverse diffusion process (Figure 1). All nodes $\boldsymbol{z} = [\boldsymbol{x}, \boldsymbol{h}]$ comprise coordinates $\boldsymbol{x} \in \mathbb{R}^3$ and categorical features $\boldsymbol{h} \in \mathbb{R}^d$. We embed atom types and residue types in a joint node embedding space by separate learnable MLPs to perform denoising steps with a single EGNN [15, 16]. The EGNN's message-passing scheme for node i at layer l is slightly modified so that the coordinate update step is not applied to pocket nodes:

$$\boldsymbol{x}_i^{l+1} = \boldsymbol{x}_i^l + \begin{cases} \sum_{j \neq i} \frac{\boldsymbol{x}_i^l - \boldsymbol{x}_j^l}{||\boldsymbol{x}_i^l - \boldsymbol{x}_j^l|| + 1} \phi_x(\boldsymbol{h}_i^l, \boldsymbol{h}_j^l, ||\boldsymbol{x}_i^l - \boldsymbol{x}_j^l||^2), & \text{if } i \text{ belongs to ligand} \\ \mathbf{0}, & \text{if } i \text{ belongs to pocket} \end{cases} \tag{1}$$

In this way, the three-dimensional protein context remains fixed throughout the EGNN layers.

Equivariance In the probabilistic setting with 3D-conditioning, we would like to ensure $E(3)$-equivariance in the following sense[2]: Evaluating the likelihood of a molecule $\boldsymbol{x}^{(L)} \in \mathbb{R}^{3 \times N_L}$ given the three-dimensional representation of a protein pocket $\boldsymbol{x}^{(P)} \in \mathbb{R}^{3 \times N_P}$ should not depend on global

[2]Here we ignore node type features, which transform invariantly, for simpler notation.

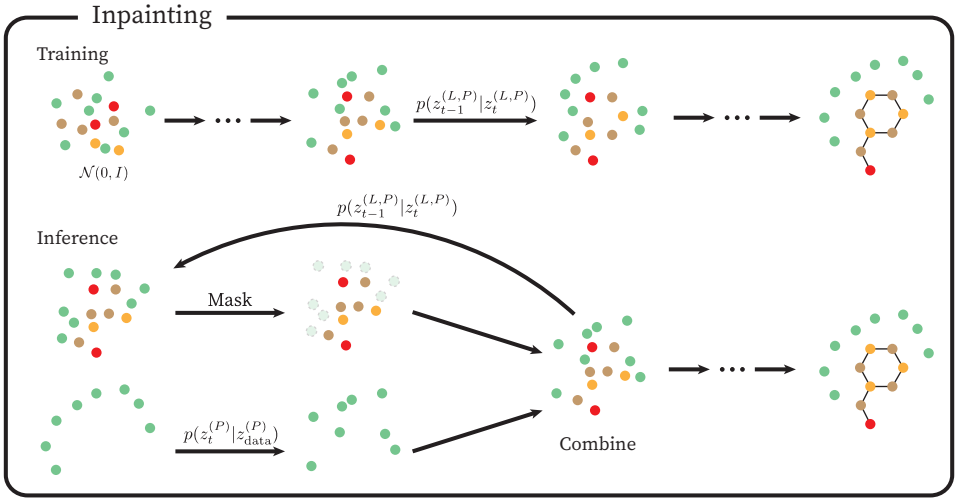


Figure 2: Schematic of the inpainting approach. The model first learns to approximate the joint distribution of ligand and pocket nodes $\boldsymbol{z}_{\text{data}}^{(L,P)}$. For sampling, context is provided by combining the latent representation of the ligand with a forward diffused representation of the pocket in each denoising step.

$E(3)$-transformations of the system, i.e. $p(\boldsymbol{R}\boldsymbol{x}^{(L)} + \boldsymbol{t}|\boldsymbol{R}\boldsymbol{x}^{(P)} + \boldsymbol{t}) = p(\boldsymbol{x}^{(L)}|\boldsymbol{x}^{(P)})$ for orthogonal $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ with $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}$ and $\boldsymbol{t} \in \mathbb{R}^3$ added column-wise. At the same time, it should be possible to generate samples $\boldsymbol{x}^{(L)} \sim p(\boldsymbol{x}^{(L)}|\boldsymbol{x}^{(P)})$ from this conditional probability distribution so that equivalently transformed ligands $\boldsymbol{R}\boldsymbol{x}^{(L)} + \boldsymbol{t}$ are sampled with the same probability if the input pocket is rotated and translated and we sample from $p(\boldsymbol{R}\boldsymbol{x}^{(L)} + \boldsymbol{t}|\boldsymbol{R}\boldsymbol{x}^{(P)} + \boldsymbol{t})$.

Equivariance to the orthogonal group $O(3)$ (comprising rotations and reflections) is achieved because we model both prior and transition probabilities with isotropic Gaussians where the mean vector transforms equivariantly w.r.t. rotations of the context (see Hoogeboom et al. [16] and Appendix C). Ensuring translation equivariance, however, is not as easy because the transition probabilities $p(\boldsymbol{z_{t-1}}|\boldsymbol{z}_t)$ are not inherently translation-equivariant. In order to circumvent this issue, we follow previous works [17, 18, 16] by limiting the whole sampling process to a linear subspace where the center of mass (CoM) of the system is zero. In practice, this is achieved by subtracting the center of mass of the system before performing likelihood computations or denoising steps.

2.2 Joint distribution with inpainting

As an extension to the conditional approach described above, we also present a ligand-*inpainting* approach. Originally introduced as a technique for completing masked parts of images [19, 20], inpainting has been adopted in other domains, including biomolecular structures [21]. Here, we extend this idea to three-dimensional point cloud data.

We first train an unconditional DDPM to approximate the joint distribution of ligand and pocket nodes $p(\boldsymbol{z}_{\text{data}}^{(L)}, \boldsymbol{z}_{\text{data}}^{(P)})$[3]. This allows us to sample new pairs without additional context. To condition on a target protein pocket, we then need to inject context into the sampling process by modifying the probabilistic transition steps. The combined latent representation $\boldsymbol{z}_{t-1}^{(L,P)}$ of protein pocket and ligand at diffusion step $t-1$ is assembled from a forward noised version of the pocket that is combined with ligand nodes predicted by the DDPM based on the previous latent representation at step t

$$\boldsymbol{z}_{t-1,\text{known}}^{(P)} \sim p(\boldsymbol{z}_{t-1}^{(P)}|\boldsymbol{z}_{\text{data}}^{(P)}) \tag{2}$$

$$\boldsymbol{z}_{t-1,\text{unknown}}^{(L,P)} \sim p_\theta(\boldsymbol{z}_{t-1}^{(L,P)}|\boldsymbol{z}_t^{(L,P)}) \tag{3}$$

$$\boldsymbol{z}_{t-1}^{(L,P)} = \left[\boldsymbol{z}_{t-1,\text{unknown}}^{(L)}, \boldsymbol{z}_{t-1,\text{known}}^{(P)}\right]. \tag{4}$$

[3]We use notations $\boldsymbol{z}^{(L,P)}$ and $[\boldsymbol{z}^{(L)}, \boldsymbol{z}^{(P)}]$ interchangeably to describe the combined system of ligand and pocket nodes.

Table 1: Evaluation of generated molecules for targets from the CrossDocked and Binding MOAD test sets. * denotes that we re-evaluate the generated ligands provided by the authors. The inference times are taken from their papers.

	Vina Score (kcal/mol, ↓)	QED (↑)	SA (↑)	Lipinski (↑)	Diversity (↑)	Time (s, ↓)
CrossDocked test set	-6.871 ± 2.32	0.476 ± 0.20	0.728 ± 0.14	4.340 ± 1.14	—	—
3D-SBDD (AR) [22]*	-5.888 ± 1.91	0.502 ± 0.17	0.675 ± 0.14	4.787 ± 0.51	0.742 ± 0.09	19659 ± 14704
Pocket2Mol [12]*	-7.058 ± 2.80	0.572 ± 0.16	0.752 ± 0.12	4.936 ± 0.27	0.735 ± 0.15	2504 ± 2207
DiffSBDD-cond (C_α)	-5.540 ± 1.57	0.460 ± 0.14	0.357 ± 0.09	4.821 ± 0.45	0.815 ± 0.06	324 ± 189
DiffSBDD-inpaint (C_α)	-5.735 ± 1.80	0.427 ± 0.15	0.343 ± 0.09	4.789 ± 0.49	0.807 ± 0.07	329 ± 177
DiffSBDD-cond	-6.584 ± 2.06	0.495 ± 0.15	0.336 ± 0.09	4.795 ± 0.49	0.730 ± 0.11	1634 ± 769
Binding MOAD test set	-8.103 ± 2.26	0.602 ± 0.15	0.336 ± 0.08	4.838 ± 0.37	—	—
DiffSBDD-cond (C_α)	-6.220 ± 1.83	0.516 ± 0.16	0.325 ± 0.09	4.855 ± 0.40	0.719 ± 0.07	414 ± 151
DiffSBDD-inpaint (C_α)	-5.981 ± 5.38	0.486 ± 0.17	0.324 ± 0.09	4.697 ± 0.63	0.716 ± 0.08	417 ± 151

In this manner, we traverse the Markov chain in reverse order from $t = T$ to $t = 0$, replacing the predicted pocket nodes with their forward noised counterparts in each step (Figure 2). Equation (3) conditions the generative process on the given protein pocket. Thanks to the noise schedule, which decreases the variance of the noising process to almost zero at $t = 0$, the final sample is guaranteed to contain an unperturbed representation of the protein pocket.

Since the model is trained to approximate the unconditional joint distribution of ligand-pocket pairs, the training procedure is identical to the unconditional molecule generation procedure developed by Hoogeboom et al. [16] aside from the fully-connected neural networks that embed protein and ligand node features in a common space as described in Section 2.1. The conditioning on known protein pockets is entirely delegated to the sampling algorithm, which means this approach is not limited to ligand-inpainting but, in principle, allows us to mask and replace arbitrary parts of the ligand-pocket system without retraining.

Equivariance Similar desiderata as in the conditional case apply to the joint probability model, where we desire $E(3)$-invariance that can be obtained from invariant priors via equivariant flows [17]. The main complications compared to the previous approach are the missing reference frame and impossibility of defining a valid translation-*invariant* prior noise distribution $p(\boldsymbol{z}_T)$ as such a distribution cannot integrate to one. Consequently, it is necessary to restrict the probabilistic model to a CoM-free subspace as described in previous works [17, 18, 16]. While the reverse diffusion process is defined for a CoM-free system, substituting the predicted pocket node coordinates with a new diffused version of the known pocket as described in Equations (2) - (4) can lead to non-zero CoM. To prevent this, we translate the known pocket representation so that its center of mass coincides with the predicted representation: $\tilde{\boldsymbol{x}}^{(P)}_{t-1,\text{known}} = \boldsymbol{x}^{(P)}_{t-1,\text{unknown}} - \boldsymbol{x}^{(P)}_{t-1,\text{known}}$ before creating the new combined representation $\boldsymbol{z}^{(L,P)}_{t-1} = [\boldsymbol{z}^{(L)}_{t-1,\text{unknown}}, \tilde{\boldsymbol{z}}^{(P)}_{t-1,\text{known}}]$ with $\tilde{\boldsymbol{z}}^{(P)}_{t-1,\text{known}} = [\tilde{\boldsymbol{x}}^{(P)}_{t-1,\text{known}}, \boldsymbol{h}^{(P)}_{t-1,\text{known}}]$.

3 Experiments

3.1 Datasets

CrossDocked We use the CrossDocked dataset [14] and follow the same filtering and splitting strategies as in previous work [22, 12]. This results in 100,000 high-quality protein-ligand pairs for the training set and 100 proteins for the test set. The split is done by 30% sequence identity using MMseqs2 [23].

Binding MOAD We also evaluate our method on experimentally determined protein-ligand complexes found in Binding MOAD [13] which are filtered and split based on the proteins' enzyme commission number as described in Appendix B. This results in 40,354 protein-ligand pairs for training and 130 pairs for testing.

3.2 Evaluation

For every experiment, we evaluated all combinations of all-atom and C_α level graphs with conditional and inpainting-based approaches respectively (with the exception of the all-atom inpainting approach due to computational limitations). Full details of model architecture and hyperparameters are given

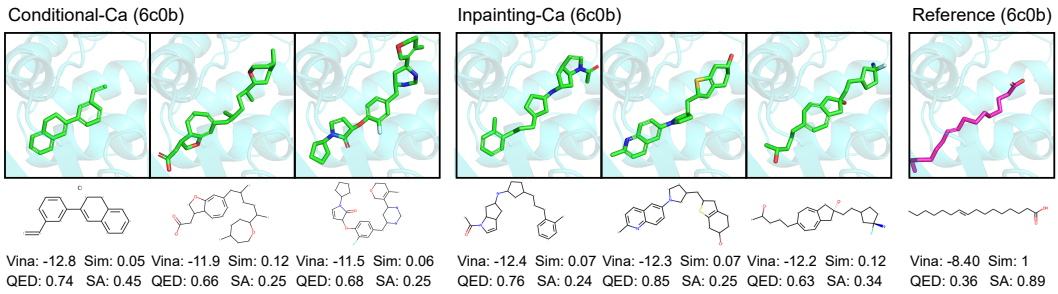


Figure 3: DiffSBDD models trained on Binding MOAD evaluated against a human receptor protein (PDB: 6c0b). Conditional and inpainting approaches are compared (C_α for both) and the three highest affinity molecules from each model are presented. Further details of the molecules shown here are explained in Appendix D.1.

in Appendix A. We sampled 100 valid molecules[4] for each target pocket with ground truth ligand sizes and remove all atoms that are not bonded to the largest connected fragment. Evaluation metrics are described in Appendix A.

Baselines We compare with two recent deep learning methods for structure-based drug design. *3D-SBDD* [22] and Pocket2Mol [12] are autoregressive schemes relying on graph representations of the protein pocket and previously placed atoms to predict probabilities based on which new atoms are added. *3D-SBDD* uses heuristics to infer bonds from generated atomic point clouds while *Pocket2Mol* directly predicts them during the sequential generation process.

3.3 Results

Overall, the experimental results in Table 1 suggest that DiffSBDD can generate diverse small-molecule compounds with predicted high binding affinity, matching state-of-the-art performance. We do not see significant differences between the conditional model and the inpainting approach. The diversity score is arguably the most interesting, as this suggests our model is able to sample greater amounts of chemical space when compared to previous methods, while maintaining high binding performance, one of the most important requirements in early-stage, structure-based lead discovery. Specifically, DiffSBDD aims to generate ligands that bind to protein pockets and learn the probability density of ligands interacting with protein pockets. While it does not optimize for other molecular properties, such as QED and Lipinski, it generates molecules similar to the test set distributions. Only SA scores are significantly lower on average. While it is unclear why the model fails to approximate the distribution of synthetic accessibility scores successfully, simple techniques can be used for downstream optimization of this property once promising candidates are found (Section D.4). Generally, presenting the full atomic context to the model constrains the space of outputs considerably, leading to higher Vina scores but lower diversity compared to the C_α-only models. The all-atom model consistently beats C_α-based models on a per target basis (Appendix Figure 11).

Generated molecules for a representative target from the Binding MOAD dataset are shown in Figure 3. The target (PDB: 6c0b) is a human receptor which is involved in microbial infection [24] and possibly tumor suppression [25]. The reference molecule, a long fatty acid (see Figure 3) that aids receptor binding [24], has too high a number of rotatable bonds and low a number of hydrogen bond donors/acceptors to be considered a suitable drug (QED of 0.36). Our model however, generates drug-like (QED between 0.63-0.85) and suitably sized molecules by adding aromatic rings connected by a small number of rotatable bonds, which allows the molecules to adopt a complementary binding geometry and is entropically favourable (by reducing the degrees of freedom), a classic technique in medicinal chemistry [26]. More examples and additional analyses can be found in Appendix D.

[4]Due to occasional processing issues the actual number of available molecules is slightly lower on average (see Appendix D.1).

4 Conclusion

In this work, we propose DiffSBDD, an $E(3)$-equivariant 3D-conditional diffusion model for structure-based drug design. We demonstrate the effectiveness and efficiency of DiffSBDD in generating novel and diverse ligands with predicted high-affinity for given protein pockets on both a synthetic benchmark and a new dataset of experimentally determined protein-ligand complexes. We demonstrate that an inpainting-based approach can achieve competitive results to direct conditioning on a wide range of molecular metrics. Extending this more versatile strategy to an all atom pocket representation therefore holds promise to solve a variety of other structure-based drug design tasks, such as lead optimization or linker design, without retraining.

References

[1] Paul D Lyne. Structure-based virtual screening: an overview. *Drug discovery today*, 7(20): 1047–1055, 2002.

[2] Brian K Shoichet. Virtual screening of chemical libraries. *Nature*, 432(7019):862–865, 2004.

[3] John J Irwin and Brian K Shoichet. Zinc- a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling*, 45(1):177–182, 2005.

[4] Leonardo G Ferreira, Ricardo N Dos Santos, Glaucius Oliva, and Adriano D Andricopulo. Molecular docking and structure-based drug design strategies. *Molecules*, 20(7):13384–13421, 2015.

[5] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

[6] Kenneth Atz, Francesca Grisoni, and Gisbert Schneider. Geometric deep learning on molecular representations. *Nature Machine Intelligence*, 3(12):1023–1032, 2021.

[7] Thomas Gaudelet, Ben Day, Arian R Jamasb, Jyothish Soman, Cristian Regep, Gertrude Liu, Jeremy B R Hayter, Richard Vickers, Charles Roberts, Jian Tang, David Roblin, Tom L Blundell, Michael M Bronstein, and Jake P Taylor-King. Utilizing graph machine learning within drug discovery and development. *Briefings in Bioinformatics*, 22(6), May 2021. doi: 10.1093/bib/bbab159. URL `https://doi.org/10.1093/bib/bbab159`.

[8] Wei Lu, Qifeng Wu, Jixian Zhang, Jiahua Rao, Chengtao Li, and Shuangjia Zheng. Tankbind: Trigonometry-aware neural networks for drug-protein binding structure prediction. *bioRxiv*, 2022.

[9] Hannes Stärk, Octavian Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola. Equibind: Geometric deep learning for drug binding structure prediction. In *International Conference on Machine Learning*, pages 20503–20521. PMLR, 2022.

[10] Yibo Li, Jianfeng Pei, and Luhua Lai. Structure-based de novo drug design using 3d deep generative models. *Chemical science*, 12(41):13664–13675, 2021.

[11] Matthew Ragoza, Tomohide Masuda, and David Ryan Koes. Generating 3d molecules conditional on receptor binding sites with deep generative models. *Chemical science*, 13(9): 2701–2713, 2022.

[12] Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. *arXiv preprint arXiv:2205.07249*, 2022.

[13] Liegi Hu, Mark L Benson, Richard D Smith, Michael G Lerner, and Heather A Carlson. Binding moad (mother of all databases). *Proteins: Structure, Function, and Bioinformatics*, 60 (3):333–340, 2005.

[14] Paul G Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B Iovanisci, Ian Snyder, and David R Koes. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of Chemical Information and Modeling*, 60(9):4200–4215, 2020.

[15] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pages 9323–9332. PMLR, 2021.

[16] Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International Conference on Machine Learning*, pages 8867–8887. PMLR, 2022.

[17] Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. In *International conference on machine learning*, pages 5361–5370. PMLR, 2020.

[18] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

[19] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020.

[20] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022.

[21] Jue Wang, Sidney Lisanza, David Juergens, Doug Tischer, Joseph L Watson, Karla M Castro, Robert Ragotte, Amijai Saragovi, Lukas F Milles, Minkyung Baek, et al. Scaffolding protein functional sites using deep learning. *Science*, 377(6604):387–394, 2022.

[22] Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. A 3d generative model for structure-based drug design. *Advances in Neural Information Processing Systems*, 34:6229–6239, 2021.

[23] Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, October 2017. doi: 10.1038/nbt.3988. URL `https://doi.org/10.1038/nbt.3988`.

[24] Peng Chen, Liang Tao, Tianyu Wang, Jie Zhang, Aina He, Kwok-ho Lam, Zheng Liu, Xi He, Kay Perry, Min Dong, et al. Structural basis for recognition of frizzled proteins by clostridium difficile toxin b. *Science*, 360(6389):664–669, 2018.

[25] Lin-Can Ding, Xiao-Yu Huang, Fei-Fei Zheng, Jian Xie, Lin She, Yan Feng, Bo-Hua Su, Da-Li Zheng, and You-Guang Lu. Fzd2 inhibits the cell growth and migration of salivary adenoid cystic carcinomas. *Oncology Reports*, 35(2):1006–1012, 2016.

[26] Timothy J Ritchie and Simon JF Macdonald. The impact of aromatic ring count on compound developability–are too many aromatic rings a liability in drug design? *Drug discovery today*, 14(21-22):1011–1020, 2009.

[27] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

[28] Christopher A Lipinski, Franco Lombardo, Beryl W Dominy, and Paul J Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 64:4–17, 2012.

[29] Amr Alhossary, Stephanus Daniel Handoko, Yuguang Mu, and Chee-Keong Kwoh. Fast, accurate, and reliable molecular docking with quickvina 2. *Bioinformatics*, 31(13):2214–2216, 2015.

[30] Greg Landrum et al. Rdkit: Open-source cheminformatics software. 2016.

[31] Noel M O'Boyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3 (1):1–14, 2011.

[32] Scott A Wildman and Gordon M Crippen. Prediction of physicochemical parameters by atomic contributions. *Journal of chemical information and computer sciences*, 39(5):868–873, 1999.

[33] Shitong Luo, Yufeng Su, Xingang Peng, Sheng Wang, Jian Peng, and Jianzhu Ma. Antigen-specific antibody design and optimization with diffusion-based generative models. *bioRxiv*, 2022.

[34] Matthias Barone, Matthias Müller, Slim Chiha, Jiang Ren, Dominik Albat, Arne Soicke, Stephan Dohmen, Marco Klein, Judith Bruns, Maarten van Dinther, et al. Designed nanomolar small-molecule inhibitors of ena/vasp evh1 interaction impair invasion and extravasation of breast cancer cells. *Proceedings of the National Academy of Sciences*, 117(47):29684–29690, 2020.

[35] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *Advances in neural information processing systems*, 34:21696–21707, 2021.

[36] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*, 2021.

[37] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2837–2845, 2021.

[38] Yuanqi Du, Tianfan Fu, Jimeng Sun, and Shengchao Liu. Molgensurvey: A systematic survey in machine learning models for molecule design. *arXiv preprint arXiv:2203.14500*, 2022.

[39] Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *arXiv preprint arXiv:2206.01729*, 2022.

[40] Namrata Anand and Tudor Achim. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. *arXiv preprint arXiv:2205.15019*, 2022.

[41] Brian L Trippe, Jason Yim, Doug Tischer, Tamara Broderick, David Baker, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

[42] Amy C Anderson. The process of structure-based drug design. *Chemistry & biology*, 10(9): 787–797, 2003.

[43] Lawrence A Kelley, Stefans Mezulis, Christopher M Yates, Mark N Wass, and Michael JE Sternberg. The phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10(6):845–858, 2015.

[44] Subha Kalyaanamoorthy and Yi-Ping Phoebe Chen. Structure-based drug design to augment hit discovery. *Drug discovery today*, 16(17-18):831–839, 2011.

[45] Pavol Drotár, Arian Rokkum Jamasb, Ben Day, Cătălina Cangea, and Pietro Liò. Structure-aware generation of drug-like molecules. *arXiv preprint arXiv:2111.04107*, 2021.

[46] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. *Advances in neural information processing systems*, 28, 2015.

[47] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

[48] Kostiantyn Lapchevskyi, Benjamin Miller, Mario Geiger, and Tess Smidt. Euclidean neural networks (e3nn) v1. 0. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2020.

[49] Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Se (3) equivariant graph neural networks with complete local frames. In *International Conference on Machine Learning*, pages 5583–5608. PMLR, 2022.

[50] Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.

[51] Johannes Klicpera, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

[52] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E Smidt, and Boris Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1): 1–11, 2022.

Appendix for "Structure-based Drug Design with Equivariant Diffusion Models"

A Implementation Details

Molecule size As part of a sample's overall likelihood, we compute the empirical joint distribution of ligand and pocket nodes $p(N_L, N_P)$ observed in the training set and smooth it with a Gaussian filter ($\sigma = 1$). In the conditional generation scenario, we derive the distribution $p(N_L|N_P)$ and use it for likelihood computations.

For sampling, we can either fix molecule sizes manually or sample the number of ligand nodes from the same distribution given the number of nodes in the target pocket:

$$N_L \sim p(N_L|N_P). \tag{5}$$

Preprocessing All molecules are expressed as graphs. For the C_α only model the node features for the protein are set as the one hot encoding of the amino acid type. The full atom model uses the same one hot encoding of atom types for ligand and protein nodes. We refrain from adding a categorical feature for distinguishing between protein and ligand atoms in this case and continue using two separate MLPs for embedding the node features instead.

Noise schedule We use the pre-defined polynomial noise schedule introduced in [16]:

$$\tilde{\alpha}_t = 1 - \left(\frac{t}{T}\right)^2, \quad t = 0, ..., T \tag{6}$$

Following [27, 16], values of $\tilde{\alpha}^2_{t|s} = \left(\frac{\tilde{\alpha}_t}{\tilde{\alpha}_s}\right)^2$ are clipped between 0.001 and 1 for numerical stability near $t = T$, and $\tilde{\alpha}_t$ is recomputed as

$$\tilde{\alpha}_t = \prod_{\tau=0}^{t} \tilde{\alpha}_{\tau|\tau-1}. \tag{7}$$

A tiny offset $\epsilon = 10^{-5}$ is used to avoid numerical problems at $t = 0$ defining the final noise schedule:

$$\alpha_t^2 = (1 - 2\epsilon) \cdot \tilde{\alpha}_t^2 + \epsilon. \tag{8}$$

Feature scaling We scale the node type features $\boldsymbol{h}$ by a factor of 0.25 relative to the coordinates $\boldsymbol{x}$ which was empirically found to improve model perfomance in previous work [16].

Hyperparameters Hyperparameters for all presented models are summarized in Table 2. Training takes about $11\,\mathrm{h}$ (BindingMOAD) and $24\,\mathrm{h}$ (CrossDocked) per 100 epochs on a single NVIDIA V100 GPU in the C_α scenario and $96\,\mathrm{h}$ (CrossDocked) per 100 epochs on a single NVIDIA A100 GPU with all atom pocket representation.

Evaluation Metrics We employ widely-used metrics to assess the quality of our generated molecules [12, 10]: (1) **Vina Score** is a physics-based estimation of binding affinity between small molecules and their target pocket; (2) **QED** is a simple quantitative estimation of drug-likeness combining several desirable molecular properties; (3) **SA** (synthetic accessibility) is a measure estimating the difficulty of synthesis; (4) **Lipinski** measures how many rules in the Lipinski rule of five [28], which is a loose rule of thumb to assess the drug-likeness of molecules, are satisfied; (5) **Diversity** is computed as the average pairwise dissimilarity (1 - *Tanimoto similarity*) between all generated molecules for each pocket; (6) **Inference Time** is the average time to sample 100 molecules for one pocket across all targets. All docking scores and chemical properties are calculated with QuickVina2 [29] and RDKit [30].

Postprocessing For postprocessing of generated molecules, we use a similar procedure as in [22]. Given a list of atom types and coordinates, bonds are first added using OpenBabel [31]. We then use RDKit to sanitise molecules, filter for the largest molecular fragment and finally remove steric clashes with 200 steps of force-field relaxation.

Table 2: DiffSBDD hyperparameters.

	CrossDocked			Binding MOAD	
	Cond	Cond (C_α)	Inpaint (C_α)	Cond (C_α)	Inpaint (C_α)
No. layers	6	6	6	6	6
Joint embedding dim.	32	32	32	32	32
Hidden dim.	256	256	256	256	256
Learning rate	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
Weight decay	10^{-12}	10^{-12}	10^{-12}	10^{-12}	10^{-12}
Diffusion steps	1000	1000	1000	1000	1000
Edges	< 7 Å	fully connected	fully connected	fully connected	fully connected
Epochs	1000	1000	1000	800	800

B Binding MOAD Dataset

We curate a dataset of experimentally determined complexed protein-ligand structures from Binding MOAD [13]. We keep pockets with valid[5] and moderately 'drug-like' ligands with QED score > 0.3. We further discard small molecules that contain atom types $\notin \{C, N, O, S, B, Br, Cl, P, I, F\}$ as well as binding pockets with non-standard amino acids. We define binding pockets as the set of residues that have any atom within 8 Å of any ligand atom. Ligand redundancy is reduced by randomly sampling at most 50 molecules with the same chemical component identifier (3-letter-code). After removing corrupted entries that could not be processed, $40\,354$ training pairs and 130 testing pairs remain. A validation set of size 246 is used to monitor estimated log-likelihoods during training. The split is made to ensure different sets do not contain proteins from the same Enzyme Commission Number (EC Number) main class.

C Proofs

In the following proofs we do not consider categorical node features $\boldsymbol{h}$ as only the positions $\boldsymbol{x}$ are subject to equivariance constraints. Furthermore, we do not distinguish between the zeroth latent representation $\boldsymbol{x}_0$ and data domain representations $\boldsymbol{x}_{\text{data}}$ for ease of notation, and simply drop the subscripts.

C.1 $O(3)$-equivariance of the prior probability

The isotropic Gaussian prior $p(\boldsymbol{x}_T^{(L)}|\boldsymbol{x}^{(P)}) = \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{x}^{(\boldsymbol{P})}), \sigma^2\boldsymbol{I})$ is equivariant to rotations and reflections represented by an orthogonal matrix $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ as long as $\boldsymbol{\mu}(\boldsymbol{R}\boldsymbol{x}^{(P)}) = \boldsymbol{R}\boldsymbol{\mu}(\boldsymbol{x}^{(P)})$ because:

$$
\begin{aligned}
p(\boldsymbol{R}\boldsymbol{x}_T^{(L)}|\boldsymbol{R}\boldsymbol{x}^{(P)}) &= \frac{1}{\sqrt{(2\pi)^{N_L}\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}||\boldsymbol{R}\boldsymbol{x}_T^{(L)} - \boldsymbol{\mu}(\boldsymbol{R}\boldsymbol{x}^{(P)})||^2\Big) \\
&= \frac{1}{\sqrt{(2\pi)^{N_L}\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}||\boldsymbol{R}\boldsymbol{x}_T^{(L)} - \boldsymbol{R}\boldsymbol{\mu}(\boldsymbol{x}^{(P)})||^2\Big) \\
&= \frac{1}{\sqrt{(2\pi)^{N_L}\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}||\boldsymbol{R}\big(\boldsymbol{x}_T^{(L)} - \boldsymbol{\mu}(\boldsymbol{x}^{(P)})\big)||^2\Big) \\
&= \frac{1}{\sqrt{(2\pi)^{N_L}\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}||\boldsymbol{x}_T^{(L)} - \boldsymbol{\mu}(\boldsymbol{x}^{(P)})||^2\Big) \\
&= p(\boldsymbol{x}_T^{(L)}|\boldsymbol{x}^{(P)}).
\end{aligned}
$$

Here we used $||\boldsymbol{R}\boldsymbol{x}||_2 = ||\boldsymbol{x}||_2$ for orthogonal $\boldsymbol{R}$.

[5] as defined in `http://www.bindingmoad.org/`

C.2 $O(3)$-equivariance of the transition probabilities

The denoising transition probabilities from time step t to $s < t$ are defined as isotropic normal distributions:

$$p_\theta(\boldsymbol{x}_{t-1}^{(L)}|\boldsymbol{x}_t^{(L)}, \hat{\boldsymbol{x}}^{(L)}, \boldsymbol{x}^{(P)}) = \mathcal{N}(\boldsymbol{x}_{t-1}^{(L)}|\boldsymbol{\mu}_{t\to s}(\boldsymbol{x}_t^{(L)}, \hat{\boldsymbol{x}}^{(L)}, \boldsymbol{x}^{(P)}), \sigma_{t\to s}^2 \boldsymbol{I}). \tag{9}$$

Therefore, $p_\theta(\boldsymbol{x}_{t-1}^{(L)}|\boldsymbol{x}_t^{(L)}, \hat{\boldsymbol{x}}^{(L)}, \boldsymbol{x}^{(P)})$ is $O(3)$-equivariant by a similar argument to Section C.1 if $\boldsymbol{\mu}_{t\to s}$ is computed equivariantly from the three-dimensional context.

Recalling the definition of $\boldsymbol{\mu}_{t\to s} = \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2}\boldsymbol{x}_t^{(L)} + \frac{\alpha_s \sigma_{t|s}^2}{\sigma_t^2}\hat{\boldsymbol{x}}^{(L)}$, we can prove its equivariance as follows:

$$\begin{aligned}
\boldsymbol{\mu}_{t\to s}(\boldsymbol{R}\boldsymbol{x}_t^{(L)}, \boldsymbol{R}\boldsymbol{x}^{(P)}) &= \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2}\boldsymbol{R}\boldsymbol{x}_t^{(L)} + \frac{\alpha_s\sigma_{t|s}^2}{\sigma_t^2}\hat{\boldsymbol{x}}^{(L)}(\boldsymbol{R}\boldsymbol{x}_t^{(L)}, \boldsymbol{R}\boldsymbol{x}^{(P)}) \\
&= \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2}\boldsymbol{R}\boldsymbol{x}_t^{(L)} + \frac{\alpha_s\sigma_{t|s}^2}{\sigma_t^2}\boldsymbol{R}\hat{\boldsymbol{x}}^{(L)}(\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)}) \quad \text{(equivariance of } \hat{\boldsymbol{x}}^{(L)}) \\
&= \boldsymbol{R}\Big(\frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2}\boldsymbol{x}_t^{(L)} + \frac{\alpha_s\sigma_{t|s}^2}{\sigma_t^2}\hat{\boldsymbol{x}}^{(L)}(\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)})\Big) \\
&= \boldsymbol{R}\boldsymbol{\mu}_{t\to s}(\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)}),
\end{aligned}$$

where $\hat{\boldsymbol{x}}^{(L)}$ defined as $\hat{\boldsymbol{x}}^{(L)} = \frac{1}{\alpha_t}\boldsymbol{x}_t^{(L)} - \frac{\sigma_t}{\alpha_t}\hat{\boldsymbol{\epsilon}}$ is equivariant because:

$$\begin{aligned}
\hat{\boldsymbol{x}}^{(L)}(\boldsymbol{R}\boldsymbol{x}_t^{(L)}, \boldsymbol{R}\boldsymbol{x}^{(P)}) &= \frac{1}{\alpha_t}\boldsymbol{R}\boldsymbol{x}_t^{(L)} - \frac{\sigma_t}{\alpha_t}\hat{\boldsymbol{\epsilon}}(\boldsymbol{R}\boldsymbol{x}_t^{(L)}, \boldsymbol{R}\boldsymbol{x}^{(P)}, t) \\
&= \frac{1}{\alpha_t}\boldsymbol{R}\boldsymbol{x}_t^{(L)} - \frac{\sigma_t}{\alpha_t}\boldsymbol{R}\hat{\boldsymbol{\epsilon}}(\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)}, t) \quad (\hat{\boldsymbol{\epsilon}} \text{ predicted by equivariant neural network}) \\
&= \boldsymbol{R}\Big(\frac{1}{\alpha_t}\boldsymbol{x}_t^{(L)} - \frac{\sigma_t}{\alpha_t}\hat{\boldsymbol{\epsilon}}(\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)}, t)\Big) \\
&= \boldsymbol{R}\hat{\boldsymbol{x}}^{(L)}(\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)}).
\end{aligned}$$

C.3 $O(3)$-equivariance of the learned likelihood

Let $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$ be an orthogonal matrix representing an element g from the general orthogonal group $O(3)$. We obtain the marginal probability density of the Markovian denoising process as follows

$$\begin{aligned}
p_\theta(\boldsymbol{x}_0^{(L)}|\boldsymbol{x}^{(P)}) &= \int p(\boldsymbol{x}_T^{(L)}|\boldsymbol{x}^{(P)}) p_\theta(\boldsymbol{x}_{0:T-1}^{(L)}|\boldsymbol{x}_T^{(L)}, \boldsymbol{x}^{(P)}) \mathrm{d}\boldsymbol{x}_{1:T} \\
&= \int p(\boldsymbol{x}_T^{(L)}|\boldsymbol{x}^{(P)}) \prod_{t=1}^{T} p_\theta(\boldsymbol{x}_{t-1}^{(L)}|\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)}) \mathrm{d}\boldsymbol{x}_{1:T}
\end{aligned}$$

and the sample's likelihood is $O(3)$-equivariant:

$$\begin{aligned}
p_\theta(\boldsymbol{R}\boldsymbol{x}_0^{(L)}|\boldsymbol{R}\boldsymbol{x}^{(P)}) &= \int p(\boldsymbol{R}\boldsymbol{x}_T^{(L)}|\boldsymbol{R}\boldsymbol{x}^{(P)}) \prod_{t=1}^{T} p_\theta(\boldsymbol{R}\boldsymbol{x}_{t-1}^{(L)}|\boldsymbol{R}\boldsymbol{x}_t^{(L)}, \boldsymbol{R}\boldsymbol{x}^{(P)}) \mathrm{d}\boldsymbol{x}_{1:T} \\
&= \int p(\boldsymbol{x}_T^{(L)}|\boldsymbol{x}^{(P)}) \prod_{t=1}^{T} p_\theta(\boldsymbol{R}\boldsymbol{x}_{t-1}^{(L)}|\boldsymbol{R}\boldsymbol{x}_t^{(L)}, \boldsymbol{R}\boldsymbol{x}^{(P)}) \mathrm{d}\boldsymbol{x}_{1:T} \quad \text{(equivariant prior)} \\
&= \int p(\boldsymbol{x}_T^{(L)}|\boldsymbol{x}^{(P)}) \prod_{t=1}^{T} p_\theta(\boldsymbol{x}_{t-1}^{(L)}|\boldsymbol{x}_t^{(L)}, \boldsymbol{x}^{(P)}) \mathrm{d}\boldsymbol{x}_{1:T} \quad \text{(equivariant transition probabilities)} \\
&= p_\theta(\boldsymbol{x}_0^{(L)}|\boldsymbol{x}^{(P)})
\end{aligned}$$

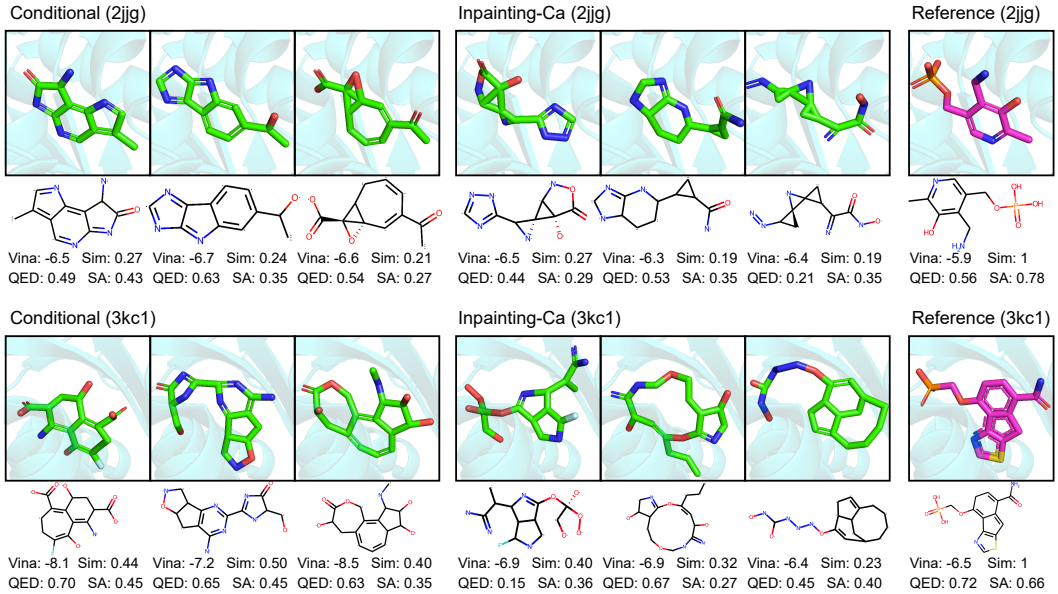


Figure 4: DiffSBDD models trained on CrossDocked and evaluated against a aminotransferase (top, PDB: 2jjg) and hydrolase (bottom, PDB: 3kc1). Conditional and inpainting approaches are compared (using all-atom and C_α level protein presentations respectively) and three high affinity molecules from each model are presented. 'Sim' is the Tanimoto similarity between the generated and reference ligand.

D Extended results

D.1 Additional Experimental Details

The numbers of available molecules differ slightly between different methods due to computational issues or missing molecules in the available baseline sets. More precisely, on average 93.5, 92.8, and 98.3 molecules have been evaluated per pocket for DiffSBDD-cond, DiffSBDD-inpaint (C_α), and DiffSBDD-cond (C_α), respectively. For Pocket2Mol, 98.4 molecules are available per pocket. The set of 3D-SBDD molecules does not contain generated ligands for two test pockets. For the remaining 98 pockets, 89.9 molecules are available on average.

All Figures show molecules generated where the starting number of nodes equals the number of nodes in the reference ligands, with the exception of Figure 3, which employs the sampling strategy outlined in Appendix A.

D.2 Additional Molecular Metrics

In addition to the molecular properties discussed in Section 3 we assess the models' ability to produce novel and valid molecules using four simple metrics: validity, connectivity, uniqueness, and novelty. **Validity** measures the proportion of generated molecules that pass basic tests by RDKit–mostly ensuring correct valencies. **Connectivity** is the proportion of valid molecules that do not contain any disconnected fragments. We convert every valid and connected molecule from a graph into a canonical SMILES string representation, count the number unique occurrences in the set of generated molecules and compare those to the training set SMILES to compute **uniqueness** and **novelty** respectively.

Table 3 shows that only a small fraction of all generated molecules is invalid and must be discarded for downstream processing. The DiffSBDD models trained on CrossDocked with C_α pocket representation generate fragmented molecules about 50% of the time. Since we can simply select and process the largest fragments in these cases, low connectivity does not necessarily affect the efficiency of the generative process. Moreover, all models produce diverse sets of molecules unseen in the training set.

Table 3: Basic molecular metrics for generated small molecules given a C_α and full atom representation of the protein pocket.

Model	Validity	Connectivity	Uniqueness	Novelty
CrossDocked Training data	100%	100%	–	–
DiffSBDD-cond (C_α)	97.75%	48.02%	96.95%	100%
DiffSBDD-inpaint (C_α)	91.62%	51.38%	98.64%	100%
DiffSBDD-cond	93.23%	83.46%	97.46%	100%
Binding MOAD Training data	96.38%	100%	–	–
DiffSBDD-cond (C_α)	94.02%	66.46%	99.55%	99.81%
DiffSBDD-inpaint (C_α)	94.98%	70.21%	99.75%	99.80%

D.3 Octanol-water partition coefficient

The octanol-water partition coefficient ($\log P$) is a measure of lipophilicity and is commonly reported for potential drug candidates [32]. We summarize this property for our generated molecules in Table 4.

Table 4: LogP values of generated molecules.

	CrossDocket	Binding MOAD
Test set	0.894 ± 2.73	0.456 ± 1.15
3D-SBDD (AR) [22]	0.273 ± 2.01	—
Pocket2Mol [12]	1.720 ± 1.97	—
DiffSBDD-cond (C_α)	-0.184 ± 1.01	0.090 ± 1.02
DiffSBDD-inpaint (C_α)	-0.519 ± 1.09	-0.366 ± 1.04
DiffSBDD-cond	-0.328 ± 1.18	—

D.4 Optimization

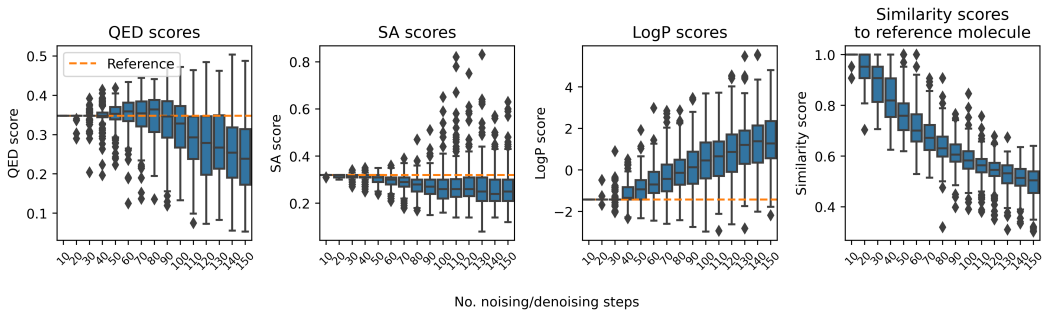


Figure 5: Effect of number of noising/denoising steps on molecule properties.

We use our model to optimize exciting candidate molecules, a common task in drug discovery called lead optimization. This is when we take a compound found to have high binding affinity and optimize it for better 'drug-like' properties. We first noise the atom features and coordinates for t steps (where t is small) using the forward diffusion process. From this partially noised sample, we can then denoise the appropriate number of steps with the reverse process until $t = 0$. This allows us to sample new candidates of various properties whilst staying in the same region of active chemical space, assuming t is small (Figure 5). This approach is inspired by [33] but note this does not allow for direct optimization of specific properties, rather directed exploration around the local chemical space according to what was learnt from the training distribution. We demonstrate the effect the number of noising/denoising steps (t) has on various molecular properties in Figure 5. We test all values of t at

intervals of 10 steps and 200 molecules are sampled at every timestep. Note this does not allow for explicit optimization of any particular property unless combined with the evolutionary algorithm.

We extend this idea by combining the partial noising/denoising procedure with a simple evolutionary algorithm that optimizes for specific molecular properties. During the evolutionary algorithm, at the end of every generation the top 10 docking molecules are used to seed the next population. Every seed molecule is elaborated into 20 new candidates with a randomly chosen t between 10 and 150. To make the first population, we start with the single reference molecule and sample 200 new molecules with t chosen as above. We find that our model performs well at this task out-of-the-box without any additional fine-tuning. As a showcase, we optimize a molecule in the test set targeting PDB:5ndu, a cancer therapeutic [34], which has low SA and QED scores, 0.31 and 0.35 respectively, but high binding affinity. Over a number of rounds of optimization, we can observe significant increases in QED (from 0.35 to mean of 0.43) whilst still maintaining high similarity to the original molecule (Figure 6a). We can also rescue the low synthetic accessibility score of the seed molecule by producing a battery of highly accessible molecules when selecting for SA. Finally, we observe that we can perform significant optimization of binding affinity after only a few rounds of optimization. Figure 6b shows 3 representative molecules with substantially optimized scores (QED, SA or Vina) whilst maintaining comparable binding affinity and globally similar structures.

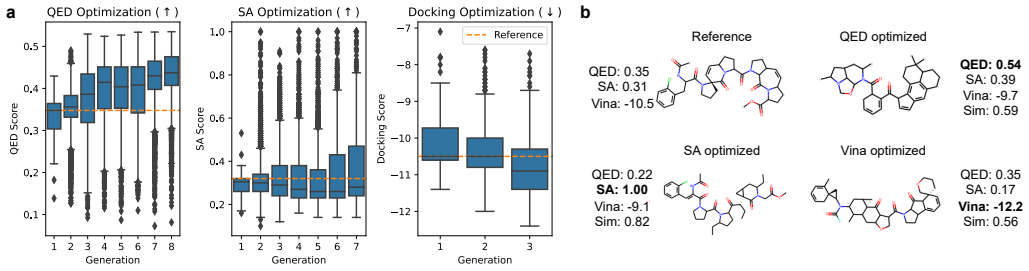


Figure 6: (a) Optimizing for various properties. (b) Examples of optimized molecules.

D.5 Agreement of generated and docked conformations

All docking scores reported in Table 1 are within one standard deviation of each other, which poses challenges for the discrimination of the best models. To verify successful pocket-conditioning, we therefore discuss an alternative way of using QuickVina for assessing the quality of the conditional generation procedure besides its *in silico* docking score. We compare the generated raw conformations (before force-field relaxation) to the best scoring QuickVina docking pose and plot the distribution of resulting RMSD values in Figures 7 and 8. As a baseline, the procedure is repeated for RDKit conformers of the same molecules with identical center of mass. For a large percentage of molecules generated by the all-atom CrossDocked model, QuickVina agrees with the predicted bound conformations, leaving them almost unchanged (RMSD below 2 Å). This demonstrates successful conditioning on the given protein pockets and showcases the success of our method to model protein-drug interactions at the atomic level.

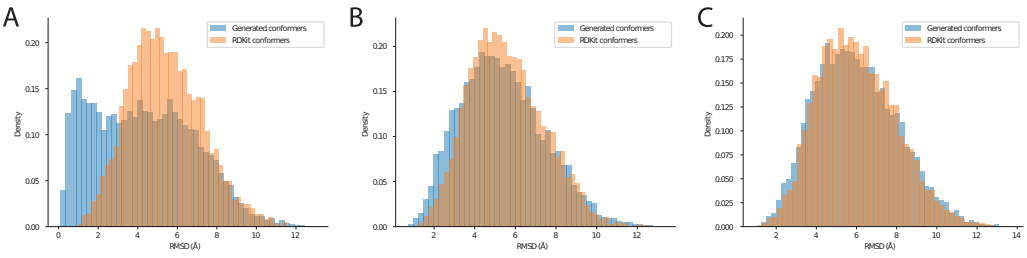


Figure 7: RMSD between original and docked conformations for CrossDocked dataset. (A) DiffSBDD-cond, sample size 8804. (B) DiffSBDD-cond (C_α), sample size 9611. (C) DiffSBDD-inpaint (C_α), sample size 8641.

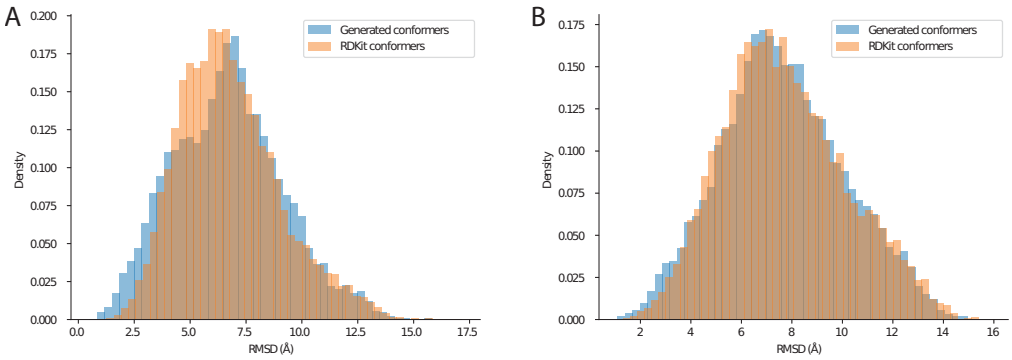


Figure 8: RMSD between original and docked conformations for BindingMOAD dataset. (A) DiffSBDD-cond (C_α), sample size 12 315. (B) DiffSBDD-inpaint (C_α), sample size 12 550.

For the C_α-only models results are less convincing. They produce poses that only slightly improve upon conformers lacking pocket-context. Likely, this is caused by atomic clashes with the proteins' side chains that QuickVina needs to resolve. Notably, however, there is a clear enrichment of molecules with less than 3 Å RMSD for both conditional models (Binding MOAD and CrossDocked) showing the advantage over unconditional conformer generation.

D.6 Random generated molecules

Randomly selected molecules generated with our method and 3 baseline methods (LiGAN, SBDD-3D and Pocket2Mol) when trained with CrossDocked are presented in Figure 9. Randomly selected molecules generated wby our method when trained with Binding MOAD are show in Figure 10.

D.7 Distribution of docking scores by target

We present extensive evaluation of the docking scores for our generated molecules in Figure 11. We evaluate all models trained with a given dataset first against all targets (Figure 11A+C) and 10 randomly chosen targets (Figure 11B+D). We note that the all-atom model trained using CrossDocked data outperforms all other methods. Unsurprisingly, model performance is highly target dependent, likely varying with properties like pocket geometry, size, charge and hydrophbicity, which would affect the propensity of generating high affinity molecules.

E Related Work

Diffusion Models for Molecules Inspired by non-equilibrium thermodynamics, diffusion models have been proposed to learn data distributions by modeling a denoising (reverse diffusion) process and have achieved remarkable success in a variety of tasks such as image, audio synthesis and point cloud generation [35–37]. Recently, efforts have been made to utilize diffusion models for molecule design [38]. Specifically, Hoogeboom et al. [16] propose a diffusion model with an equivariant network that operates both on continuous atomic coordinates and categorical atom types to generate new molecules in 3D space. Torsional Diffusion [39] focuses on a conditional setting where molecular conformations (atomic coordinates) are generated from molecular graphs (atom types and bonds). Similarly, 3D diffusion models have been applied to generative design of larger biomolecular structures, such as antibodies [33] and other proteins [40, 41].

Structure-based Drug Design Structure-based Drug Design (SBDD) [4, 42] relies on the knowledge of the 3D structure of the biological target obtained either through experimental methods or high-confidence predictions using homology modelling [43]. Candidate molecules are then designed to bind with high affinity and specificity to the target using interactive software [44] and often human-based intuition [4]. Recent advances in deep generative models have brought a new wave of research that model the conditional distribution of ligands given biological targets and thus enable *de novo* structure-based drug design. Most of recent work consider this task as a sequential generation

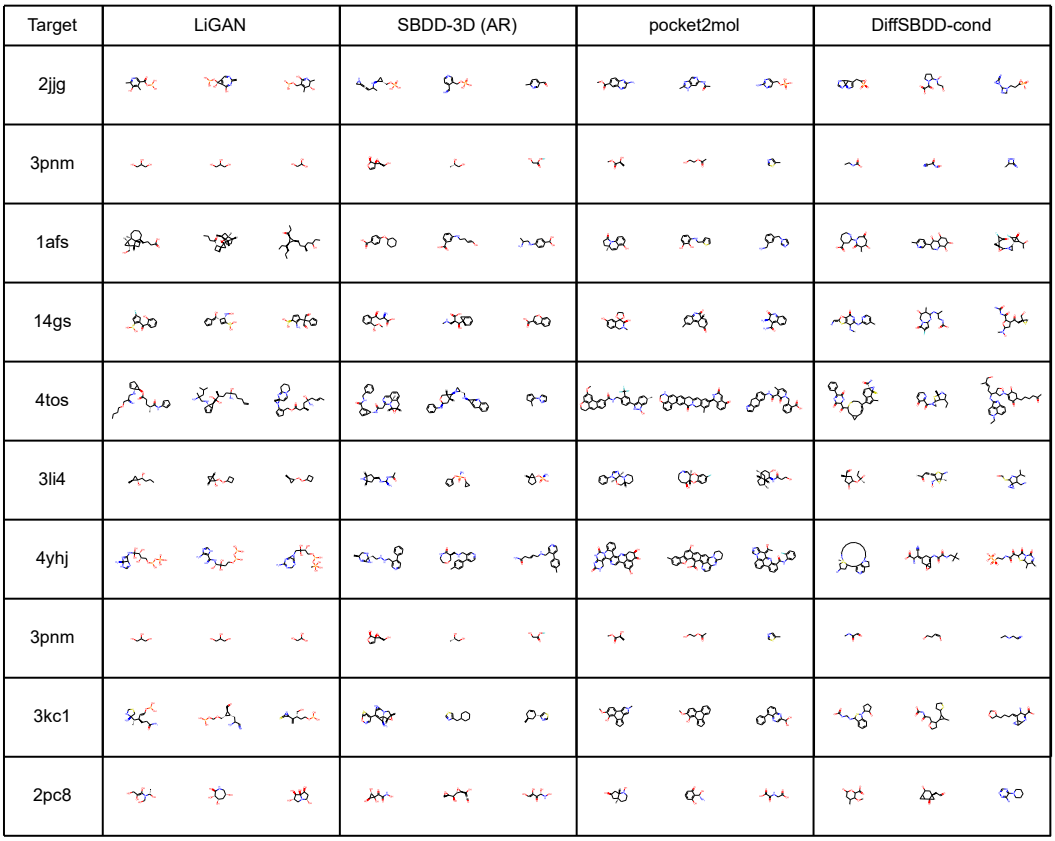


Figure 9: Generated molecules for 10 randomly chosen targets in the CrossDocked test set. For each target, 3 randomly selected generated molecules from 4 models are shown.

problem and design a variety of generative methods including auto-regressive models, reinforcement learning, etc., to generate ligands inside protein pockets atom by atom [45, 22, 10, 12].

Geometric Deep Learning for Drug Discovery Geometric deep learning refers to incorporating geometric priors in neural architecture design that respects symmetry and invariance, thus reduces sample complexity and eliminates the need for data augmentation [5]. It has been prevailing in a variety of drug discovery tasks from virtual screening to de novo drug design as symmetry widely exists in the representation of drugs. One line of work introduces graph and geometry priors and designs message passing neural networks and equivariant neural networks that are permutation- and translation-, rotation-, reflection-equivariant, respectively [46, 47, 15, 48, 49], and has been widely used in representing biomolecules from small molecules to proteins [6] and solving downstream tasks such as molecular property prediction [50, 51], binding pose prediction [9], molecular dynamics [52], etc. Another line of work focuses on generative design of new molecules [38]. Specifically, they formulate molecule design as a graph or geometry generation problem and there are two strategies: one-shot generation that generates graphs (atom and bond features) in one step and sequential generation that generates them in a sequence of steps.

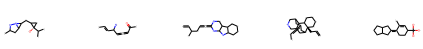
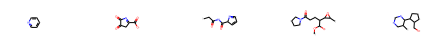
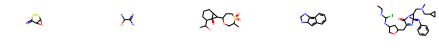
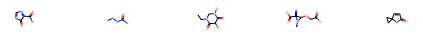
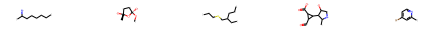
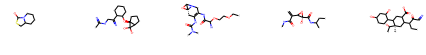
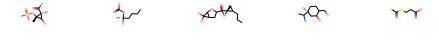
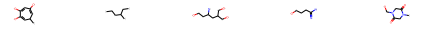
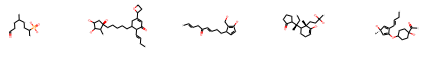
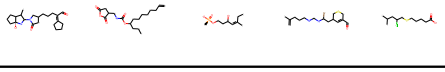
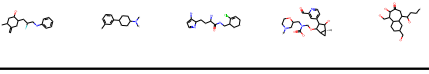
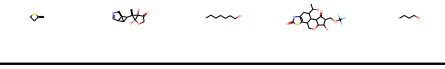
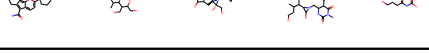
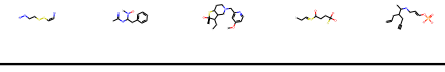
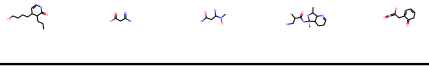
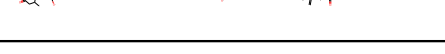
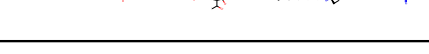
Target	DiffSBDD-cond	DiffSBDD-inpaint
2fky		
3zjx		
3gt9		
5ndu		
2vl8		
1j78		
3eks		
5zzb		
1fd7		
2a5x		

Figure 10: Generated molecules for 10 randomly chosen targets in the Binding MOAD test set. For each target-model pair, 5 randomly selected generated molecules are shown. C_α level proteins were used for both models.

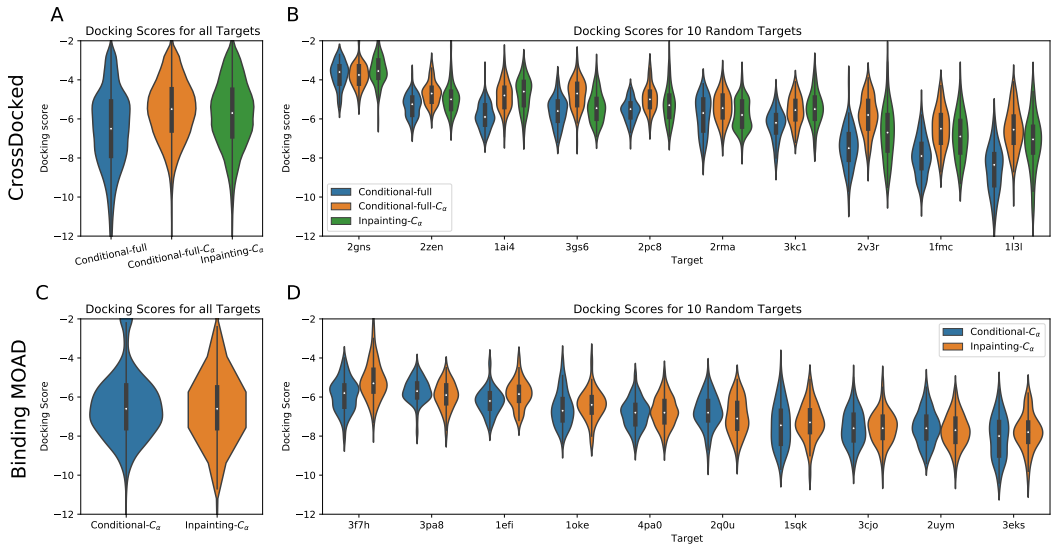


Figure 11: Docking scores of generated molecules for various methods trained on the CrossDocked (A-B) and Binding MOAD (C-D) datasets. (A) Violin plot of docking scores for all 3 methods trained using CrossDocked. (B) Same as before but for 10 randomly chosen targets sorted by mean score. (C) Violin plot of docking scores for all 2 methods trained using Binding MOAD. (D) Same as before but for 10 randomly chosen targets sorted by mean score.