TOWARDS STRUCTURE-BASED DRUG DESIGN WITH PROTEIN FLEXIBILITY

Arne Schneuing[1*], **Ilia Igashov**[1*], **Thomas Castiglione**[1], **Michael Bronstein**[2] **& Bruno Correia**[1]
[1]École Polytechnique Fédérale de Lausanne, [2]University of Oxford
{arne.schneuing,ilia.igashov,bruno.correia}@epfl.ch

ABSTRACT

Incorporating flexibility and dynamics of protein targets is a frontier of computational drug design. A machine learning model that jointly generates ligands and bound conformations of binding pockets holds promise to access a larger chemical space by removing unnecessary structural constraints, and opens the door to design campaigns in which only the unbound structure of the therapeutic target is known. Here we report on progress we made towards this goal and present two models: DrugFlow, a new generative model for structure-based drug design with static protein structures that shows strong performance compared to previous methods, and FlexFlow, an extension of this model that also predicts side chain torsion angles together with preliminary empirical data.

1 INTRODUCTION

Despite a growing importance of biologics, chemical entities are still the predominant class of FDA-approved drugs with a share of 85% (Santos et al., 2017). Furthermore, over 95% of known drugs target human or pathogen proteins (Santos et al., 2017). At the same time cost and duration of the development of new drugs are skyrocketing (Simoens & Huys, 2021), which explains the growing interest in computational design of small molecular compounds that bind specifically to disease-associated proteins. In recent years, the machine learning community has contributed a plethora of generative tools tackling the so-called structure-based drug design problem. Some of these models generate molecules unconditionally without knowledge about the target protein. They either create molecular graphs (Jo et al., 2022; Vignac et al., 2022), 3D atomic point clouds (Gebauer et al., 2019; Garcia Satorras et al., 2021; Hoogeboom et al., 2022) or both Vignac et al. (2023). More closely related to the work presented here, another family of models attempts to generate novel chemical matter conditioned on three-dimensional context, typically a structural model of a target protein. Among these most of the models sample atom positions and types without providing explicit information about covalent bonds (Liu et al., 2022; Ragoza et al., 2022; Schneuing et al., 2022; Igashov et al., 2022; Guan et al., 2023a; Lin et al., 2022; Xu et al., 2023). Others generate the full molecular graph structure and binding pose jointly (Peng et al., 2022; Guan et al., 2023b; Zhang et al., 2023). Notably, most recent structure-based drug design models belong to the family of diffusion probabilistic models (Schneuing et al., 2022; Guan et al., 2023a; Lin et al., 2022; Xu et al., 2023; Guan et al., 2023b; Weiss et al., 2023).

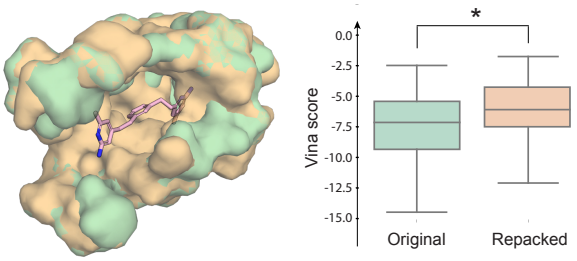

Figure 1: The geometry of unbound ligand binding pockets (orange) differs from their bound state (green). This leads to significantly weaker docking scores when ligands are relaxed in pockets with repacked side chains (Wilcoxon rank-sum test with p-value of 3.5×10^{-4}).

*These authors contributed equally.

Even though these approaches have shown promising performance on computational benchmarks, all of them make the important assumption of static proteins. However, this assumption only holds in few cases, and incorporating the dynamics and flexibility of protein structures is one of the key open challenges for structure-based drug design (Fraser & Murcko, 2024). What is more, drug design methods are typically trained and tested on protein structures in complex with known ligands, thereby making the additional assumption that the bound conformation of the pocket is available (or, to be more precise, one of several possible bound conformations). If medicinal chemists target new sites on proteins without previously known binder even this requirement may be too strict as the drug design campaign might be launched from a structural model that is not amenable to ligand binding without a rearrangement of the pocket residues.

To illustrate this, we relaxed 100 small molecules from the CrossDocked dataset (Francoeur et al., 2020) and computed the Vina docking score (McNutt et al., 2021). Then, we repacked the side chains of all pocket residues in absence of the molecules and repeated relaxation and scoring. The data in Figure 1 shows that even this minor structural change with a fixed protein backbone leads to significantly worse predicted binding affinity of the same ligands.

Based on these findings we set out to develop a structure-based drug design model that permits full side chain flexibility while keeping backbone atoms fixed. Such a model will sample probabilistic ensembles of possible binding modes and enable drug design for targets in unbound conformation as long as they have a relatively stable backbone structure.

To this end we first introduce DrugFlow, a new generative model for structure-based drug design. Our method combines Euclidean flow matching for atom coordinates and a Markov bridge model (Igashov et al., 2023) for atom and bond types. We then extend our model to the flexible side chain scenario and introduce FlexFlow which additionally utilizes Riemannian flow matching to operate on side chain torsion angles. While DrugFlow demonstrates strong empirical performance in the rigid setting, FlexFlow is currently still struggling to create high-quality residue conformations and protein-ligand interactions.

2 Methods

2.1 Generative modeling

In this work, we use flow matching methods for all continuous data types, that is atom coordinates and side chain torsion angles, because they are conceptually simple for Euclidean data and can be easily adapted to manifolds like the hypertorus of the torsion angles. Flow matching has already found some biomolecular applications (Stärk et al., 2023; Yim et al., 2023a; Bose et al., 2023) but has not been well explored as a tool for small molecule design. To generate the discrete molecular graphs, we use Markov bridge models Igashov et al. (2023). More details about these generative modeling frameworks are presented in Appendix A.1.

2.2 Side chain reconstruction

To provide the full-atomic information more explicitly, we convert the side chain dihedral angles back to atom positions in every training and sampling step. This operation is performed efficiently using the Natural Extension Reference Frame (NERF) algorithm (Parsons et al., 2005; Alcaide et al., 2022) in parallel for each residue.

2.3 Computational aspects

Input graph definition While the computational graph of the generated small molecule must necessarily be complete so that bond types can be generated freely, we improve the computational efficiency by removing edges between pocket residues or between residues and ligand atoms based on a predefined cutoff distance (10Å).

Nodes in this graph correspond either to a ligand atom or a pocket residue. Both node types are passed through separate neural encoders before being processed jointly with a GVP-GNN (Jing et al., 2020; 2021). To retain the full atomic information while adopting this coarse-grained representation for the protein pocket (one computational node per residue), we include difference vectors

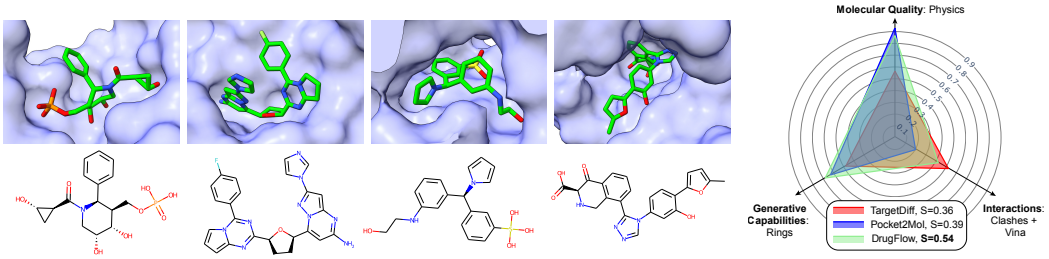

Figure 2: (left) Randomly picked four molecules that pass all the filters from Table 1 and have between 15 and 25 heavy atoms. (right) Overall performance of the methods. While TargetDiff better captures interactions, it remarkably underperforms in molecular quality and generative capabilities. On the other hand, Pocket2Mol demonstrates competitive results in terms of the two latter aspects but underperforms in the quality of interactions. DrugFlow is the most balanced method that successfully optimizes all three directions, as reflected by the area S of its triangle.

to each atom of the residue in addition to the C_α coordinate and amino acid type as node features similar to Zhang et al. (2023).

Featurization We consider the atom types {C, N, O, S, B, Br, Cl, P, I, F, NH, N+, O-} where +/- indicate charges and NH is a nitrogen atom with explicit hydrogen. In all other cases, hydrogens are assumed to be implicit following normal valence assumptions. Furthermore, DrugFlow generates single, double, triple, aromatic, and "None" as bond types. FlexFlow additionally outputs five torsion angles $\{\chi_1, \chi_2, \chi_3, \chi_4, \chi_5\}$ for each residue. Since not all angles are present in every residue we mask predictions where appropriate.

We also use two extra input features. (1) As a rudimentary way to encode information about protein dynamics, we endow the residue nodes with normal vectors from an anisotropic network model using ProDy Zhang et al. (2021). (2) For ligand nodes we include node-level cycle counts up to size 5 following (Vignac et al., 2022; Igashov et al., 2023).

Self-conditioning Self-conditioning (Chen et al., 2022) is a sampling strategy in which the neural network takes its previous prediction as additional input during iterative sampling. Like previous works (Yim et al., 2023b; Stärk et al., 2023) we observe significant performance improvements using this technique.

Size prediction Both DrugFlow and FlexFlow generate molecules of prespecified size. To decide on the number of heavy atoms we generate we consider two strategies. (1) We compute the categorical distribution $p(N|M)$ (histogram) of molecule sizes N given the number of residues M in the target pocket based on the training set, and sample from it. (2) We train a simple classifier on the same training set to predict the molecule size given a coarse-grained C_α representation of the pocket and sample from its output after applying a softmax activation. The DrugFlow results presented in this paper were obtained with the neural size predictor whereas the work-in-progress FlewFlow model uses the histogram approach.

3 Results

Dataset & Baselines We use the CrossDocked dataset (Francoeur et al., 2020) with 100,000 high-quality protein-ligand pairs for training and 100 proteins for testing, following previous works (Luo et al., 2021; Peng et al., 2022). The data split was done by 30% sequence identity using MMseqs2 (Steinegger & Söding, 2017). We compare DrugFlow with an autoregressive method, Pocket2Mol (Peng et al., 2022), and two diffusion-based methods, TargetDiff (Guan et al., 2023a) and DiffSBDD (Schneuing et al., 2022). We generated 50 samples for each target with DrugFlow and selected only molecules that passed validity and connectivity filters. As Pocket2Mol generates connected molecules by design, and DiffSBDD outputs only valid molecules, applying these filters to our samples before evaluation enables a fair comparison. Because DrugFlow produces $n < 50$

Table 1: Evaluation metrics that assess the ability of the models to learn the training distribution, to sample physically and chemically plausible molecules, and to build strong interactions with the target protein. Top-2 results for each metric are highlighted in bold.

	Generative Capabilities			Molecular Quality		Interactions	
Method	FCD ↓	Ring Coverage ↑	Diversity ↑	Physics ↑	MedChem ↑	Clashes ↑	Docking ↑
DiffSBDD	**12.016**	0.522	**0.725**	0.501	0.121	0.699	0.444
TargetDiff	14.371	0.477	0.716	0.575	0.123	**0.883**	**0.582**
Pocket2Mol	13.323	**0.617**	**0.840**	**0.939**	**0.494**	**0.919**	0.225
DrugFlow	**5.383**	**0.681**	0.676	**0.866**	**0.289**	0.867	**0.499**

valid and connected molecules (out of 50 samples for each target), we randomly select the same amount of samples from other methods to ensure equal sample sizes for all targets and baselines.

Metrics An open problem in computational drug discovery is the lack of high-quality metrics that adequately assess the quality of the generated compounds. While several attempts have been recently made to standardize the evaluation procedures (Buttenschoen et al., 2024; Harris et al., 2023), there are still fundamental issues with the robustness of the existing scoring functions and quality of the heuristics commonly used for approximating different molecular properties such as drug-likeness or synthetic availability. Moreover, in our task we aim to optimize several orthogonal properties that can be formulated as follows,

- **Generative capabilities**: how well does the model learn the training distribution?
- **Molecular quality**: how "good" are the generated molecules?
- **Interactions**: how do the molecules interact with the target?

3.1 De novo molecule generation

Generative capabilities To assess how well the model learns the training distribution, we compute the Fréchet ChemNet Distance Preuer et al. (2018) (FCD) and Jensen-Shannon distance between the histograms of ring systems (Walters, 2022; 2021) in the training set and generated samples (ring coverage). The FCD computes the distance between the latent representations of the reference molecules (from the training set) and those of the generated molecules. These latent representations are obtained from the penultimate activation layer of the ChemNet model (Goh et al., 2018), which is trained to predict chemical and biological activity. FCD is given as the Wasserstein-2 (Fréchet) distance between the two distributions. Lower FCD values indicate that the generated molecules are more similar to the reference molecules in the high-dimensional space defined by ChemNet's features, suggesting that the model has effectively learned the distribution of the training data. As provided in Table 1, DrugFlow remarkably outperforms other methods in both metrics.

Molecular quality To estimate the overall quality of the generated molecules, we use different filters that assess the geometric and physical relevance of the molecules as well as their suitability as a drug candidate (MedChem). For geometric and physical relevance, we compute the fraction of the samples that pass all PoseBusters intramolecular filters (Buttenschoen et al., 2024). To assess the chemical relevance of the samples, we compute the fraction of the samples that are better than the reference molecules in terms of drug-likeness (QED), synthetic accessibility (SA), and Lipinski's rule of five, and report it as the MedChem ratio. As shown in Table 1, DrugFlow performs better than most baselines except Pocket2Mol.

Interactions To assess the quality of interactions between the molecules and target pockets, we consider two aspects. First, we report the fractions of the samples that pass PoseBusters intermolecular filters (Buttenschoen et al., 2024) that account for steric clashes and volumetric overlaps between molecule and protein atoms. Next, we compute the fractions of the samples that improve Vina scores compared to the reference molecules. While TargetDiff and Pocket2Mol produce fewer clashes than DrugFlow, after minimization our molecules improve docking scores more often than Pocket2Mol samples (Table 1).

Overall performance Structure-based drug discovery is a multi-objective task where one aims to optimize various chemical and physical properties of a drug candidate as well as its interactions with the target protein. We summarize the performance of DrugFlow, Pocket2Mol, and TargetDiff in these three directions in Figure 2. While both TargetDiff and Pocket2Mol strongly perform in one specific direction, both show a considerable imbalance in the overall performance. On the contrary, DrugFlow achieves competitive performance in each of the three evaluation directions and demonstrates the highest performance balance across all of them. The overall performance can be quantified by the area of the triangle on the diagram, and according to this metric DrugFlow outperforms other methods by a large margin.

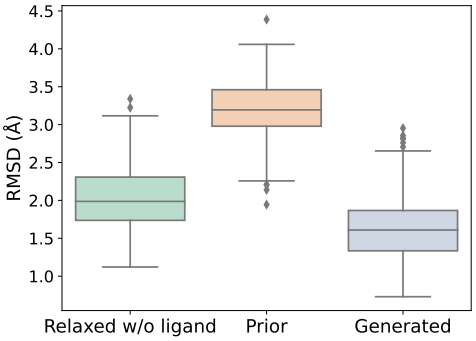

Figure 3: Side chain root-mean-square deviation (RMSD) for relaxed pockets in absence of the ligand (left), random samples from the model's prior (middle), and FlexFlow-generated side chain conformers for a fixed ligand structure (right).

3.2 Flexible design

In this section, we provide preliminary results on sampling side chain torsion angles and molecules for repacked CrossDocked pockets as a proxy for unbound protein structures. We repacked side chains with the Rosetta repack protocol (Conway et al., 2014) and achieved an RMSD of about 2Å for side chain heavy atoms (Figure 3, left). As a first test, we sampled 10 sets of side chain torsion angles per target with FlexFlow while keeping the ligand fixed, i.e. using the ground truth vector field instead of the predicted vector field for all ligand-related variables. Figure 3 shows that the model samples pocket structures close to the original bound conformations. This is expected because fixing the ligand binding pose substantially constrains the space of feasible solutions. For comparison, we also include the distribution of RMSD values that results from simply taking random angles from the prior distribution (Figure 3, middle). Next, we generated 50 samples for repacked pockets with Pocket2Mol, DrugFlow and FlexFlow, and selected only valid and connected molecules. We report Vina docking scores and PoseBusters intermolecular metrics in Table 2. While FlexFlow performs on par with DrugFlow according to most of the molecular and generative metrics, it underperforms its rigid counterpart in terms of docking score and introduces markedly more steric clashes than both static baselines. This result is particularly surprising as it contradicts our assumption that the flexible side chain model will discover better bound configurations that lead to improved binding. For these experiments, we adapted a DrugFlow model without increasing the number of trainable weights. While there may be a fundamental limitation due to our side chain modeling approach, we speculate that improved performance might be achieved by purely scaling up its capacity to match the increased complexity of the problem.

Table 2: Evaluation on repacked targets.

Method	Docking	Clashes
Pocket2Mol	-5.123 ± 1.567	**0.898**
DrugFlow	$\mathbf{-6.204 \pm 2.496}$	0.837
FlexFlow	-6.179 ± 2.820	0.551

4 Conclusion & Outlook

In this work, we introduce DrugFlow, a generative model for structure-based drug design that consistently demonstrates state-of-the-art performance across various orthogonal metrics. We further extend our method to the flexible side chain scenario and introduce FlexFlow. The work on FlexFlow is still in progress, and our goal is to demonstrate the effectiveness of the generative process with the flexible protein side chains over the rigid modeling in cases when only unbound protein structures are available. Another interesting question for further research is the ability of FlexFlow to recover bound side chain conformations and to sample side chains according to the rotamer distributions of natural binders. Encouraging preliminary results are included in Figure 3 and Appendix Figure 8. Finally, we note that drug design with flexible side chains is the first step toward full-fledged in-

duced fit modeling which often involves backbone movements. In future work, we aim to address this modality and extend FlexFlow to operate on the protein backbone atoms as well.

ACKNOWLEDGMENTS

The authors would like to thank Charles Harris, Rebecca Neeser, Yuanqi Du and Evgenia Elizarova for their helpful feedback. A.S. was supported by Microsoft Research AI4Science. I.I. has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Sklodowska-Curie grant agreement No. 945363.

REFERENCES

Michael S Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *arXiv preprint arXiv:2209.15571*, 2022.

Eric Alcaide, Stella Biderman, Amalio Telenti, and M Cyrus Maher. Mp-nerf: A massively parallel method for accelerating protein structure reconstruction from internal coordinates. *Journal of Computational Chemistry*, 43(1):74–78, 2022.

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993, 2021.

Avishek Joey Bose, Tara Akhound-Sadegh, Kilian Fatras, Guillaume Huguet, Jarrid Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael Bronstein, and Alexander Tong. Se (3)-stochastic flow matching for protein backbone generation. *arXiv preprint arXiv:2310.02391*, 2023.

Martin Buttenschoen, Garrett M Morris, and Charlotte M Deane. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 2024.

Ricky TQ Chen and Yaron Lipman. Riemannian flow matching on general geometries. *arXiv preprint arXiv:2302.03660*, 2023.

Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog bits: Generating discrete data using diffusion models with self-conditioning. *arXiv preprint arXiv:2208.04202*, 2022.

Patrick Conway, Michael D Tyka, Frank DiMaio, David E Konerding, and David Baker. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein science*, 23(1): 47–55, 2014.

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.

James S Fraser and Mark A Murcko. Structure is beauty, but not always truth. *Cell*, 187(3):517–520, 2024.

Victor Garcia Satorras, Emiel Hoogeboom, Fabian Fuchs, Ingmar Posner, and Max Welling. E (n) equivariant normalizing flows. *Advances in Neural Information Processing Systems*, 34:4181–4192, 2021.

Niklas Gebauer, Michael Gastegger, and Kristof Schütt. Symmetry-adapted generation of 3d point sets for the targeted discovery of molecules. *Advances in neural information processing systems*, 32, 2019.

Garrett B. Goh, Charles Siegel, Abhinav Vishnu, and Nathan O. Hodas. Using rule-based labels for weak supervised learning: A chemnet for transferable chemical property prediction. *arXiv preprint arXiv:1712.02734*, 2018.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543*, 2023a.

Jiaqi Guan, Xiangxin Zhou, Yuwei Yang, Yu Bao, Jian Peng, Jianzhu Ma, Qiang Liu, Liang Wang, and Quanquan Gu. Decompdiff: Diffusion models with decomposed priors for structure-based drug design. 2023b.

Charles Harris, Kieran Didi, Arian Jamasb, Chaitanya Joshi, Simon Mathis, Pietro Lio, and Tom Blundell. Posecheck: Generative models for 3d structure-based drug design produce unrealistic poses. In *NeurIPS 2023 Workshop on New Frontiers of AI for Drug Discovery and Development*, 2023.

Emiel Hoogeboom, Vıctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, pp. 8867–8887. PMLR, 2022.

Ilia Igashov, Hannes Stärk, Clément Vignac, Victor Garcia Satorras, Pascal Frossard, Max Welling, Michael Bronstein, and Bruno Correia. Equivariant 3d-conditional diffusion models for molecular linker design. *arXiv preprint arXiv:2210.05274*, 2022.

Ilia Igashov, Arne Schneuing, Marwin Segler, Michael Bronstein, and Bruno Correia. Retrobridge: Modeling retrosynthesis with markov bridges. *arXiv preprint arXiv:2308.16212*, 2023.

Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael JL Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons. *arXiv preprint arXiv:2009.01411*, 2020.

Bowen Jing, Stephan Eismann, Pratham N Soni, and Ron O Dror. Equivariant graph neural networks for 3d macromolecular structure. *arXiv preprint arXiv:2106.03843*, 2021.

Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International Conference on Machine Learning*, pp. 10362–10383. PMLR, 2022.

Haitao Lin, Yufei Huang, Meng Liu, Xuanjing Li, Shuiwang Ji, and Stan Z Li. Diffbp: Generative diffusion of 3d molecules for target protein binding. *arXiv preprint arXiv:2211.11214*, 2022.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Meng Liu, Youzhi Luo, Kanji Uchino, Koji Maruhashi, and Shuiwang Ji. Generating 3d molecules for target protein binding. *arXiv preprint arXiv:2204.09410*, 2022.

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.

Andrew T McNutt, Paul Francoeur, Rishal Aggarwal, Tomohide Masuda, Rocco Meli, Matthew Ragoza, Jocelyn Sunseri, and David Ryan Koes. Gnina 1.0: molecular docking with deep learning. *Journal of cheminformatics*, 13(1):1–20, 2021.

Jerod Parsons, J Bradley Holmes, J Maurice Rojas, Jerry Tsai, and Charlie EM Strauss. Practical conversion from torsion space to cartesian space for in silico protein synthesis. *Journal of computational chemistry*, 26(10):1063–1068, 2005.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, pp. 17644–17655. PMLR, 2022.

Kristina Preuer, Philipp Renz, Thomas Unterthiner, Sepp Hochreiter, and Günter Klambauer. Fréchet chemnet distance: A metric for generative models for molecules in drug discovery. *arXiv preprint arXiv:1803.09518*, 2018.

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.

Rita Santos, Oleg Ursu, Anna Gaulton, A Patrícia Bento, Ramesh S Donadi, Cristian G Bologa, Anneli Karlsson, Bissan Al-Lazikani, Anne Hersey, Tudor I Oprea, et al. A comprehensive map of molecular drug targets. *Nature reviews Drug discovery*, 16(1):19–34, 2017.

Arne Schneuing, Yuanqi Du, Charles Harris, Arian Jamasb, Ilia Igashov, Weitao Du, Tom Blundell, Pietro Lió, Carla Gomes, Max Welling, Michael Bronstein, and Bruno Correia. Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695*, 2022.

Steven Simoens and Isabelle Huys. R&d costs of new medicines: A landscape analysis. *Frontiers in Medicine*, 8:760762, 2021.

Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Harmonic self-conditioned flow matching for multi-ligand docking and binding site design. *arXiv preprint arXiv:2310.05764*, 2023.

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.

Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian Fatras, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*, 2023.

Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022.

Clement Vignac, Nagham Osman, Laura Toni, and Pascal Frossard. Midi: Mixed graph and 3d denoising diffusion for molecule generation. *arXiv preprint arXiv:2302.09048*, 2023.

Pat Walters. Useful rdkit utils, 2021. URL https://github.com/PatWalters/useful_rdkit_utils.

Pat Walters. Mining ring systems in molecules for fun and profit, 2022. URL https://practicalcheminformatics.blogspot.com/2022/12/identifying-ring-systems-in-molecules.html.

Tomer Weiss, Eduardo Mayo Yanes, Sabyasachi Chakraborty, Luca Cosmo, Alex M Bronstein, and Renana Gershoni-Poranne. Guided diffusion for inverse molecular design. *Nature Computational Science*, pp. 1–10, 2023.

Kevin Eric Wu, Kevin K Yang, Rianne van den Berg, James Zou, Alex Xijie Lu, and Ava P Amini. Protein structure generation via folding diffusion. 2022.

Minkai Xu, Alexander S Powers, Ron O Dror, Stefano Ermon, and Jure Leskovec. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, pp. 38592–38610. PMLR, 2023.

Jason Yim, Andrew Campbell, Andrew YK Foong, Michael Gastegger, José Jiménez-Luna, Sarah Lewis, Victor Garcia Satorras, Bastiaan S Veeling, Regina Barzilay, Tommi Jaakkola, et al. Fast protein backbone generation with se (3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023a.

Jason Yim, Brian L Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. Se (3) diffusion model with application to protein backbone generation. *arXiv preprint arXiv:2302.02277*, 2023b.

Odin Zhang, Jintu Zhang, Jieyu Jin, Xujun Zhang, RenLing Hu, Chao Shen, Hanqun Cao, Hongyan Du, Yu Kang, Yafeng Deng, et al. Resgen is a pocket-aware 3d molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence*, 5(9):1020–1030, 2023.

She Zhang, James M Krieger, Yan Zhang, Cihan Kaya, Burak Kaynak, Karolina Mikulska-Ruminska, Pemra Doruker, Hongchun Li, and Ivet Bahar. Prody 2.0: increased scale and scope after 10 years of protein dynamics modelling with python. *Bioinformatics*, 37(20):3657–3659, 2021.

A APPENDIX

A.1 GENERATIVE MODEL

Flow matching Flow matching describes a class of deep generative models that approximate a time-dependent vector field $u_t(x)$ which generates a sequence of probability distributions $\{p_t : t \in [0,1]\}$ pushing a prior p_0 towards the data distribution p_1. This learned vector field describes an ordinary differential equation (ODE) that can be integrated to transform a sample from the prior into a sample from the data distribution

$$\frac{d}{dt}\psi_t(x) = u_t(\psi_t(x)). \tag{1}$$

Efficient training of flow matching models is only possible because we do not need to define the true vector field $u_t(x)$ but can instead match the conditional flow $u_t(x|x_1)$ which is much easier to parameterize (Lipman et al., 2022) based on a data point x_1. Thus, the conditional flow matching loss amounts to

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,q(x_1),p_t(x|x_1)}\|v_\theta(t,x) - u_t(x|x_1)\|^2 \tag{2}$$

$$= \mathbb{E}_{t,q(x_1),p(x_0)}\|v_\theta(t,x_t) - \dot{x}_t\|^2. \tag{3}$$

For sampling, we obtain a sample $\psi_0(x) = x_0$ from the prior and simulate the ODE in equation 1 replacing the true vector field with the learned vector field $v_\theta(t, x_t)$.

In this work, we build on a variant called Independent-coupling Conditional Flow Matching (ICFM) (Albergo & Vanden-Eijnden, 2022; Tong et al., 2023) and consider a Gaussian conditional probability path

$$p_t(x|x_1) = \mathcal{N}(x|\mu_t(x_1), \sigma_t(x_1)^2) \tag{4}$$

with generating vector field

$$u_t(x|x_1) = \frac{\sigma_t'(x_1)}{\sigma_t(x_1)}\left(x - \mu_t\left(x_1\right)\right) + \mu_t'(x_1). \tag{5}$$

We use this setup with

$$\mu_t(x_1) = tx_1 + (1-t)x_0, \quad \sigma_t(x_1) = \sigma \tag{6}$$

to model the flow for ligand coordinates. This results in a constant velocity vector field $\dot{x}_t = \frac{x_1 - x_t}{1-t} = x_1 - x_0$ and the flow matching loss:

$$\mathcal{L}_{\text{coord}}(\theta) = \mathbb{E}_{t,q(x_1),p(x_0)}\|v_\theta(t,x_t) - (x_1 - x_0)\|^2 \tag{7}$$

Riemannian conditional flow matching (RCFM) For side chain torsion angles, we need to define a flow on the torus $[-\pi,\pi)^N$. Fortunately, all components of the flow matching framework can be computed in a simulation-free manner on this simple manifold. We use the explicit RCFM loss derived by Chen & Lipman (2023):

$$\mathcal{L}_{\text{RCFM}}(\theta) = \mathbb{E}_{t,q(x_1),p(x_0)}\|v_\theta(t,x_t) + d(x_0,x_1)\frac{\nabla_{x_t} d(x_t,x_1)}{\|\nabla_{x_t} d(x_t,x_1)\|_g^2}\|_g^2 \tag{8}$$

which in our case with premetric $d(x,y) = |w(y-x)|$ amounts to

$$\mathcal{L}_\chi(\theta) = \mathbb{E}_{t,q(x_1),p(x_0)} w(\|v_\theta(t,x_t) - w(x_1 - x_0)\|)^2 \tag{9}$$

where function $w(\alpha) = ((\alpha+\pi) \mod 2\pi) - \pi$ wraps values within the range $[-\pi,\pi)$ as in Wu et al. (2022). The norm $\|\cdot\|_g$ on the tangent space to $\mathcal{M}$ is induced by the standard inner product $\langle u,v\rangle_g = \langle u,v\rangle,\ u,v \in T_x\mathcal{M}$ in this particular case. Here, the learned vector field $v_\theta(t,x_t)$ outputs vectors on the tangent plane.

We obtain x_t as a point on the geodesic connecting x_0 and x_1 in closed-form using the exponential and logarithm maps (Chen & Lipman, 2023)

$$x_t = \exp_{x_0}\left(t\log_{x_0}\left(x_1\right)\right) \tag{10}$$

with $\exp_x(u) = w(x+u)$ and $\log_x(y) = \text{atan2}(\sin(y-x), \cos(y-x))$ [1].

[1] $x, y \in \mathcal{M}$ and $u, v \in T_x\mathcal{M}$.

Markov Bridges The molecular graph consists of discrete entities (node and edge types) and can therefore not be easily modeled in the flow matching framework. While discrete diffusion formulations (Austin et al., 2021; Vignac et al., 2022) can be used in principle, we decided to employ the Markov bridge model (Igashov et al., 2023) instead which is conceptually more similar to the flow matching scheme used for the continuous variables.

The Markov Bridge Model captures the stochastic dependency between two discrete-valued spaces $\mathcal{X}$ and $\mathcal{Y}$. It defines a Markov process between fixed start and end points $z_0 = x$ and $z_1 = y$, respectively, through a sequence of $N + 1$ random variables $(z_{t=i/N})_{i=0}^{N}$ for which

$$p(z_t|z_0, z_{0+\Delta t}, ..., z_{t-\Delta t}, z_1 = y) = p(z_t|z_{t-\Delta t}, z_1 = y) \tag{11}$$

with $\Delta t = 1/N$. Additionally, since the process is pinned at its end point, we have

$$p(z_1 = y|z_{1-\Delta t}, y) = 1. \tag{12}$$

Each transition is given by

$$p(z_{t+\Delta t}|z_t, z_1 = y) = \text{Cat}(\boldsymbol{z}_{t+\Delta t}; \boldsymbol{Q}_t \boldsymbol{z}_t) \tag{13}$$

where $\boldsymbol{z}_t \in \{0, 1\}^K$ is a one-hot representation of the current category and $\boldsymbol{Q}_t$ is a transition matrix parameterised as

$$\boldsymbol{Q}_t \coloneqq \boldsymbol{Q}_t(\boldsymbol{y}) = \beta_t \boldsymbol{I} + (1 - \beta_t)\boldsymbol{y}\boldsymbol{1}_K^T. \tag{14}$$

Any intermediate state of the Markov chain can be probed in closed form:

$$p(z_t|z_0, z_1) = \text{Cat}(\boldsymbol{z}_t; \bar{\boldsymbol{Q}}_{t-\Delta t}\boldsymbol{z}_0) \tag{15}$$

with

$$\bar{\boldsymbol{Q}}_t = \boldsymbol{Q}_t \boldsymbol{Q}_{t-\Delta t} ... \boldsymbol{Q}_0 = \bar{\beta}_t \boldsymbol{I} + (1 - \bar{\beta}_t)\boldsymbol{y}\boldsymbol{1}_K^T. \tag{16}$$

In this work, we choose a linear schedule for $\bar{\beta} = 1 - t$ which implies $\beta_t = \bar{\beta}_t / \bar{\beta}_{t-\Delta t} = (1 - t)/(1 - t + \Delta t)$.

The neural network predicts $\hat{y} = \varphi_\theta(z_t, t)$ so that we can sample from the Markov bridge without knowing the true final state. It is trained by maximizing the following lower bound of the log-likelihood

$$\begin{aligned} \log q_\theta(y|x) &\geq -T \cdot \mathbb{E}_{t, z_t \sim p(z_t|x,y)} D_{\text{KL}}(p(z_{t+1}|z_t, y)||q_\theta(z_{t+1}|z_t)) \\ &=: -\mathcal{L}_{\text{MBM}}. \end{aligned} \tag{17}$$

Our overall loss function is a weighted sum of the previously introduced loss terms:

$$\mathcal{L} = \mathcal{L}_{\text{coord}} + \lambda_\chi \mathcal{L}_\chi + \lambda_a \mathcal{L}_{\text{MBM, atom}} + \lambda_b \mathcal{L}_{\text{MBM, bond}} \tag{18}$$

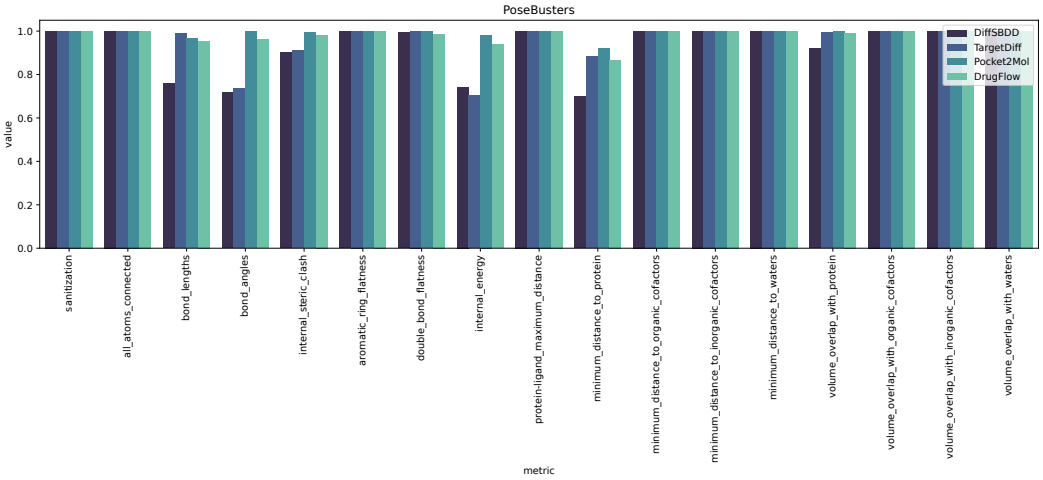

Figure 4: Detailed PoseBusters results on the original CrossDocked dataset.

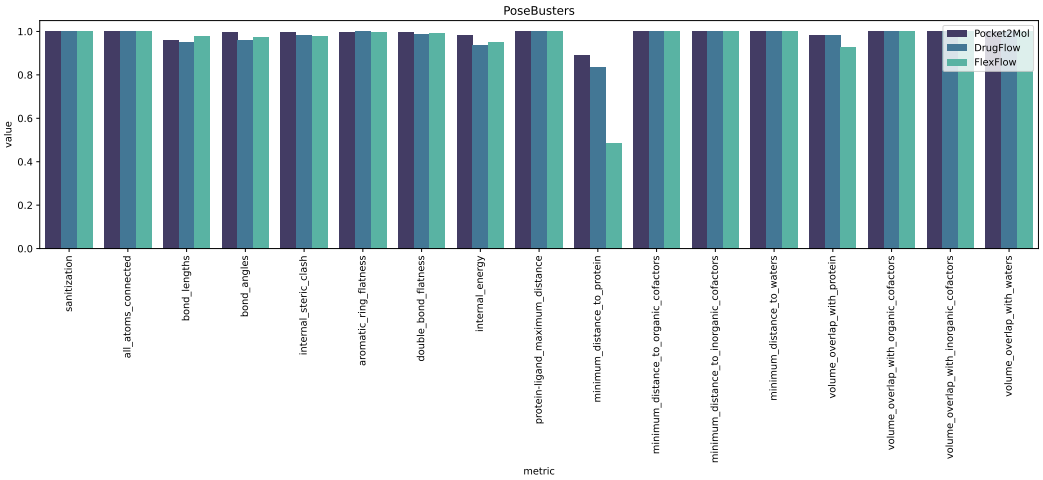

Figure 5: Detailed PoseBusters results on the repacked CrossDocked dataset.

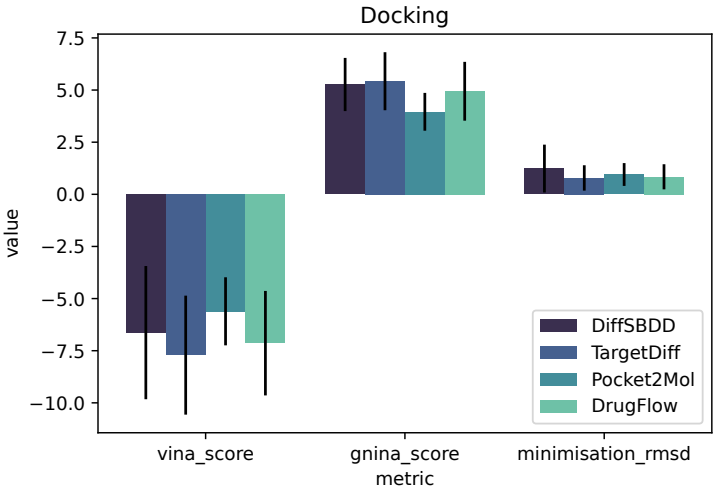

Figure 6: Detailed docking results on the original CrossDocked dataset.

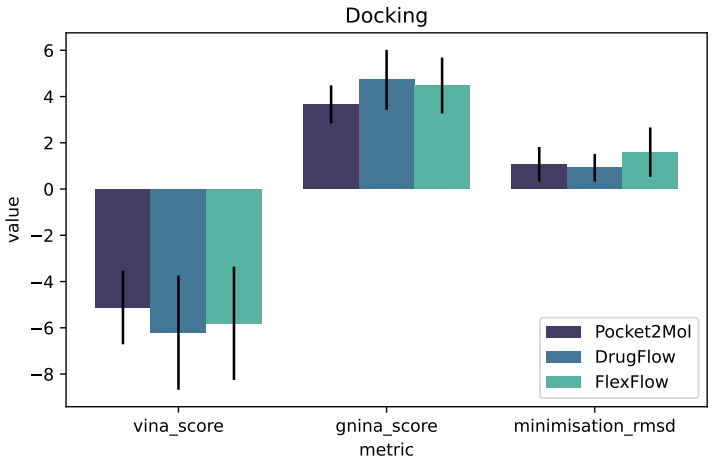

Figure 7: Detailed docking results on the repacked CrossDocked dataset.

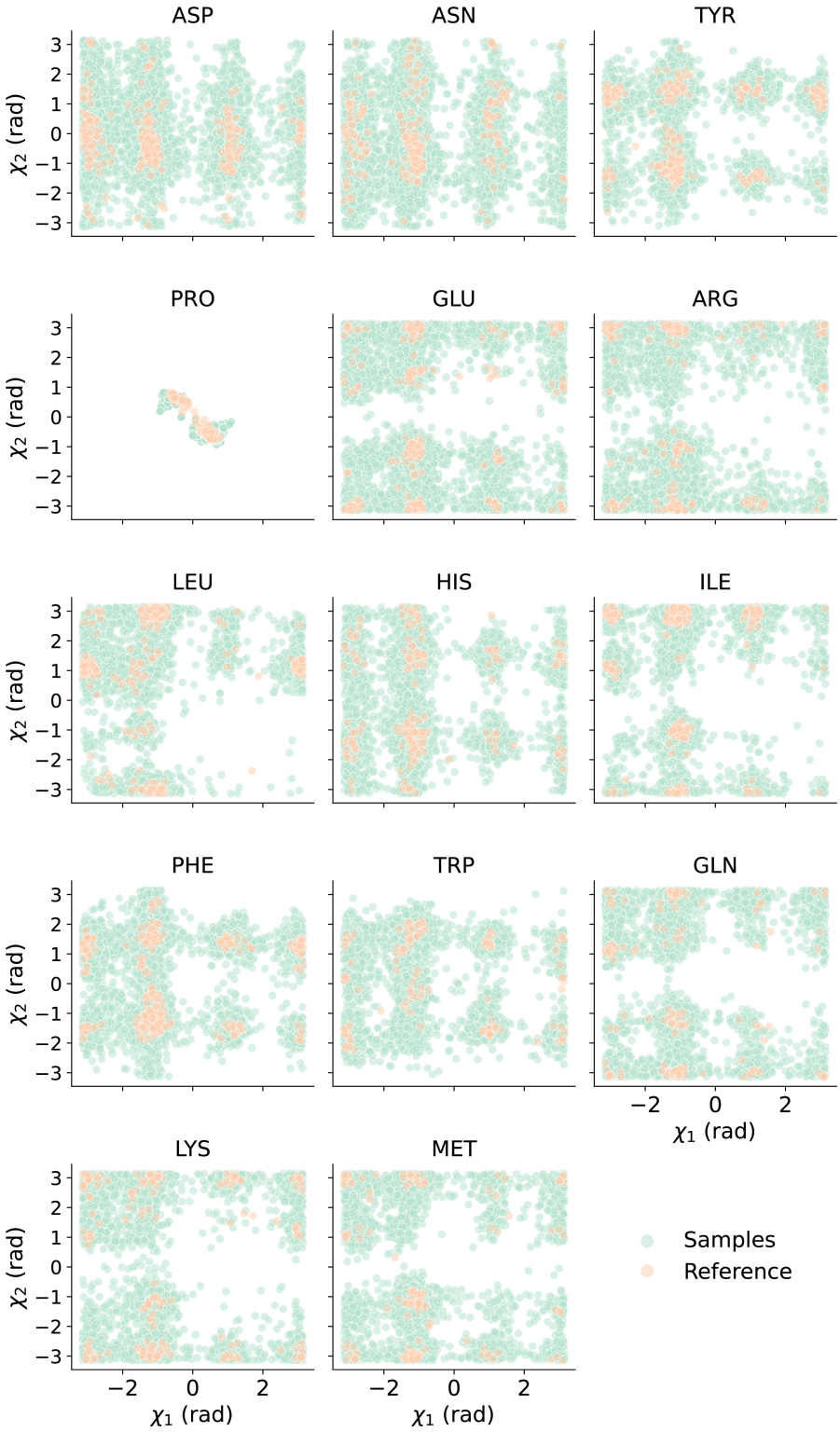

Figure 8: Distributions of χ_1 and χ_2 angles for the 14 amino acids that have at least two side chain torsion angles. We compare FlexFlow samples to the bound pocket conformations from the test set.