Bending and Binding: Predicting Protein Flexibility upon Ligand Interaction using Diffusion Models

Xuejin Zhang*
VantAI
xuejin@vant.ai

Tomas Geffner*
VantAI & UMass Amherst
tomasgeffner@gmail.com

Matt McPartlon
VantAI
matt@vant.ai

Mehmet Akdel
VantAI
mehmet@vant.ai

Dylan Abramson
VantAI
dylan@vant.ai

Graham Holt
VantAI
graham@vant.ai

Alexander Goncearenco
VantAI
alex@vant.ai

Luca Naef
VantAI
luca@vant.ai

Michael M. Bronstein
VantAI & Oxford University
michael.bronstein@cs.ox.ac.uk

Abstract

Predicting protein conformational changes driven by binding of small molecular ligands is imperative to accelerate drug discovery for protein targets with no established binders. This work presents a novel method to capture such conformational changes: given a protein *apo* conformation (unbound state), we propose an equivariant conditional diffusion model to predict its *holo* conformations (bound state with external small molecular ligands). We design a novel variant of the EGNN architecture for the score network (score-informed EGNN), which is able to exploit conditioning information in the form of the reference (*apo*) structure to guide the diffusion's sampling process. Learning from experimentally determined *apo/holo* conformations, we observe that our model can generate conformations close to *holo* conditioned only on *apo* state.

1 Introduction

Protein conformational changes upon ligand binding is a common phenomenon in drug discovery and design [1–3]. Such conformational changes are frequently thought to stem from "conformational selection," in which the protein conformation distribution is biased by the presence of a ligand towards *existing* conformations that are compatible with ligand binding [3, 4]. If this hypothesis is true, we might expect to be able to predict ligand-induced conformational changes without incorporating information about the ligand identity or structure.

To test this hypothesis, we design a method to predict conformational changes conditioned on protein structure and sequence, excluding the ligand. Since our focus is on capturing protein conformational diversity, we build on diffusion models [5–7], a powerful method for approximating distributions (diffusion models are introduced briefly in Appendix A). Specifically, we propose ApolloDiff (Apo-to-holo diffusion, Fig. 1), an equivariant conditional diffusion model [8, 9] to sample a protein's *holo* conformations (i.e., bound) given its *apo* structure (i.e., unbound) and sequence as conditioning variables. The score network used by ApolloDiff has two main components: an Evoformer block [10] that produces a protein representation, followed by a *score-informed* EGNN, and a novel variant of the EGNN architecture [11] that leverages the reference *apo* structure provided to produce the score required by the diffusion to generate samples.

Generative AI and Biology Workshop (NeurIPS 2023).

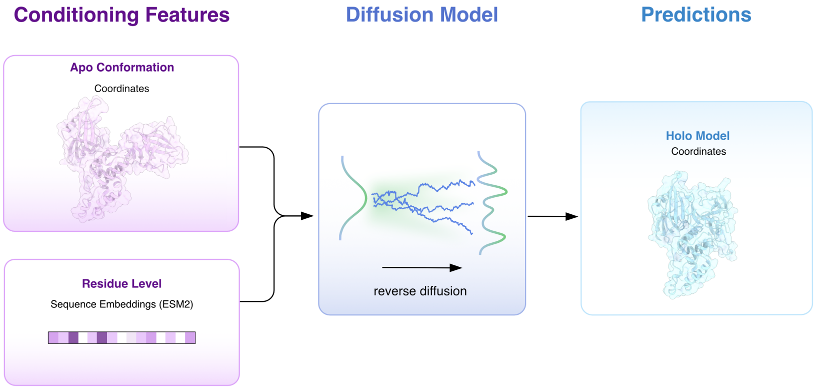

Figure 1: APOLLODIFF overview. Given an *apo* structure, APOLLODIFF uses a diffusion model to produce an ensemble of predicted conformations approximating the protein structure in the *holo* state.

Our methodology presents a significant advantage over ligand-based strategies, as it enables the enumeration of alternative backbone conformations prior to the knowledge of the ligand, facilitating more efficient virtual screenings. This computational efficiency derives from the capacity to substitute the redundant process of enumerating protein backbone conformation changes for each distinct ligand during the docking phase with a single, a-priori enumeration step. Furthermore, we find that our method outperforms the previous state-of-the-art, suggesting that it represents a promising alternative to model ligand-induced conformational changes, obviating the need for direct incorporation of ligand information.

2 Apo-Conditioned Diffusion for Holo Protein Structures

As mentioned above, our aim is to predict *holo* conformations given an *apo* structure. We frame this as approximating the distribution $p(\mathbf{x}_{\text{holo}} \,|\mathbf{x}_{\text{apo}}, s)$, where $\mathbf{x}_{\text{holo}}$ and $\mathbf{x}_{\text{apo}}$ represent a protein's aligned *holo* and *apo* structures, and s the protein sequence. We represent protein structures as 3D point clouds $\mathbf{x}_{\text{holo}}, \mathbf{x}_{\text{apo}} \in \mathbb{R}^{n_{\text{res}} \times 3}$, where each residue is identified with the C_α coordinates. Since the *apo* and *holo* coordinates are aligned (and centered), the target distribution is equivariant under rigid body transformations, i.e. $p(\mathbf{x}_{\text{holo}} \,|\, \mathbf{x}_{\text{apo}}, s) = p(\mathbf{R}\mathbf{x}_{\text{holo}} + \mathbf{t} \,|\, \mathbf{R}\mathbf{x}_{\text{apo}} + \mathbf{t}, s)$ for any $(\mathbf{t}, \mathbf{R})$.

We use a rotation-equivariant conditional diffusion model to approximate this distribution. We handle translations following Hoogeboom et al. [8], Yim et al. [12], centering all structures and defining the diffusion in the zero center-of-mass linear subspace. As noted by Igashov et al. [9], two conditions are sufficient to guarantee that the marginal distribution defined by the diffusion is rotation-equivariant: an equivariant reference distribution, and an equivariant score network. We use the variance-preserving SDE formulation for diffusion models (presented in Appendix A), which yields a standard Gaussian reference, satisfying the first condition. Additionally, we use a rotation-equivariant score network with two main components (Fig. 4):

Evoformer block [10]. This block takes as input the sequence ESM [13] embedding $\mathbf{r}$, an $n_{\text{res}} \times c_s$ matrix, and the pair features $\mathbf{p}$, an $n_{\text{res}} \times n_{\text{res}} \times c_p$ tensor, where $\mathbf{p}_{ij}$ is obtained combining the sequence separation between residues i, j and their pairwise distance in the *apo* structure (both features are binned, producing one-hot vectors which are concatenated). Each layer updates $(\mathbf{r}, \mathbf{p}) \leftarrow \text{Evoformer}(\mathbf{r}, \mathbf{p})$, allowing the exchange of information between $\mathbf{r}$ and $\mathbf{p}$.[1]

SI-EGNN. A novel variant of the EGNN [11] we call *score-informed equivariant GNN* (SI-EGNN), which produces the predicted score leveraging the *apo* structure provided as input. Each node i in our GNN represents a residue in the protein, and consists of a tuple $(\mathbf{x}^t_{\text{holo}}(i), \mathbf{r}^t_i)$, where $\mathbf{x}^t_{\text{holo}}$ denotes the diffused *holo* coordinates at diffusion time t, and $\mathbf{r}^t_i$ denotes $\mathbf{r}_i$ concatenated with the sinusoidal encoding of t. The edges between nodes are assigned the features in $\mathbf{p}$, with $\mathbf{p}_{ij}$ corresponding to the

[1] The evoformer layers do not include triangle attention (for efficiency) nor column attention (no MSA).

edge connecting nodes i and j. Each layer $(\mathbf{x}^t_{\text{holo}}, \mathbf{r}^t) \leftarrow \text{SI-EGNN}(\mathbf{x}^t_{\text{holo}}, \mathbf{x}_{\text{apo}}, \mathbf{r}^t, \mathbf{p}, t)$ is given by

$$m_{ij} \leftarrow \phi_p\left(\mathbf{r}^t_i, \mathbf{r}^t_j, \|\mathbf{x}^t_{\text{holo}}(i) - \mathbf{x}^t_{\text{holo}}(j)\|^2, \mathbf{p}_{ij}\right), \qquad m_i \leftarrow \frac{\sum_{j \neq i} m_{ij}}{n_{\text{res}} - 1}, \qquad \mathbf{r}^t_i \leftarrow \phi_r\left(\mathbf{r}^t_i, m_i\right),$$
$$\underline{\mathbf{s}_{\text{apo}} \leftarrow \nabla \log p_t(\mathbf{x}^t_{\text{holo}} \,|\, \mathbf{x}_{\text{apo}}),} \quad \mathbf{x}^t_{\text{holo}}(i) \leftarrow \mathbf{x}^t_{\text{holo}}(i) + \sum_{j \neq i}\left(\mathbf{x}^t_{\text{holo}}(i) - \mathbf{x}^t_{\text{holo}}(j)\right)\phi_x(m_{ij}) \underline{+\mathbf{s}_{\text{apo}}\,\phi_s\left(\mathbf{r}^t_i\right)}, \tag{1}$$

where ϕ_* are MLPs, $p_t(\mathbf{x} \,|\, \mathbf{x}_{\text{apo}})$ is the distribution obtained by applying the diffusion forward kernel (Appendix A) to the *apo* conformation, and m_{ij} represents the "message" sent from node j to node i. The predicted score is then computed as the difference between the initial and the updated *holo* coordinates (i.e. difference between input and output of the SI-EGNN).

It can be observed that the SI-EGNN follows the EGNN architecture with some additional terms, underlined in Eq. (1). Empirically, we observe this variant of the EGNN, applicable due the available reference *apo* structure, yields improved training convergence and final results. An intuitive explanation for this stems from the fact that, by fixing $\phi_p, \phi_r, \phi_x = 0$ and $\phi_s = 1/L$ (with L the number of SI-EGNN layers), the predicted score results in the score required to generate the reference *apo* structure (with these choices, the reverse diffusion approximately produces the reference *apo* structure without any training). If the *holo* and *apo* structures display structural similarities, as they typically do [14], this provides useful guidance to the network. Critically, the SI-EGNN maintains the rotation equivariance property from the original EGNN, as $\nabla \log p_t(\mathbf{x}|\mathbf{x}_{\text{apo}})$ is equivariant to joint rotations $(\mathbf{x}, \mathbf{x}_{\text{apo}}) \mapsto (\mathbf{R}\mathbf{x}, \mathbf{R}\mathbf{x}_{\text{apo}})$ for the Gaussian forward kernel from Appendix A.

Acknowledgments and Disclosure of Funding

Do **not** include this section in the anonymized submission, only in the final paper. You can use the ack environment provided in the style file to automatically hide this section in the anonymized submission.

3 Results

We evaluate ApolloDiff on a subset of the D3PM dataset [14] using C_α RMSD as our evaluation metric (see Appendix C for data processing details). We compare against the "Aligned Diffusion Schrödinger Bridge" approach (SBalign) [15], described in Appendix B. Briefly, this method trains a diffusion to transport proteins from their *apo* to their *holo* conformations (i.e. the diffusion starts from the *apo* conformation, not from random noise). To ensure a fair comparison, we re-train SBalign on our variant of the D3PM dataset.

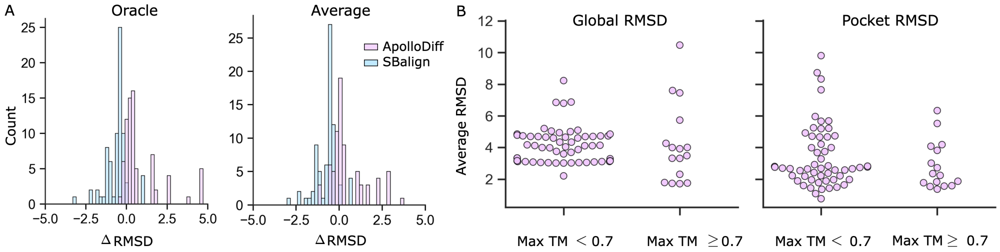

Figure 2: (A) RMSD improvement per UniProtID accession ($\Delta\text{RMSD} > 0$ indicates the model produces structures closer to the *holo* than the reference *apo*). (B) Global and Pocket RMSD between generated conformations by our model and *holo* structures (one dot per *holo* structure in the test set).

ApolloDiff vs SBalign We begin by studying whether the methods are able to produce structures closer to the *holo* conformation than the respective *apo* used as reference. We evaluate this measuring $\Delta\text{RMSD} = \text{RMSD}_{\text{apo}-\text{holo}} - \text{RMSD}_{\text{gen}-\text{holo}}$, which compares the RMSD between the models' predictions and *holo* conformations ($\text{RMSD}_{\text{gen}-\text{holo}}$) against the RMSD between *apo* and *holo* conformations ($\text{RMSD}_{\text{apo}-\text{holo}}$). We note that positive values for ΔRMSD indicate good performance, while negative values are indicative of poor performance. We report two variants of this metric for each UniProt accession (using 5 generated samples for each model): *oracle* ΔRMSD (Fig. 2A, left), computed using the best generated sample (minimum RMSD w.r.t. *holo*), and *average* ΔRMSD (Fig. 2A, right), computed using the average RMSD across all generated samples. Results in Fig. 2 shows that our model is able to produce conformations closer to the *holo* than the reference *apo* (i.e. $\Delta\text{RMSD} > 0$) for most cases, outperforming SBALIGN.

	APOLLODIFF		APOLLODIFF (*apo* masked)		SBALIGN	
ΔRMSD	Average ↑	Oracle ↑	Average ↑	Oracle ↑	Average ↑	Oracle ↑
≥0.1Å	31	55	30	54	4	7
≥0.5Å	21	23	21	24	4	4
≥1Å	16	19	16	19	0	4
≥2Å	10	10	10	10	0	0

Table 1: $\Delta RMSD$ for each of the 74 holo structures in the test set. Average and Oracle are computed as explained in the main text for Fig. 2A. All 74 structures have $\Delta\text{RMSD}_{\text{apo}-\text{holo}}$ within [3Å, 10Å]. APOLLODIFF (*apo* masked) stands for our method when the *apo* structure has 10 consecutive residues artificially masked, included to evaluate the APOLLODIFF's ability to impute unresolved residues.

We further study APOLLODIFF and SBALIGN by reporting the number of cases for which each one is able to generate predictions with $\Delta\text{RMSD} > \delta$, for $\delta \in$\{0.1Å, 0.5Å, 1Å, 2Å\}. Results are shown in Table 1. Using the oracle metric, we observe that SBALIGN generates conformations that are better than the reference *apo* for 7 cases out of the 74 in the test set, with this number dropping to 4 for the average metric. On the other hand, APOLLODIFF generates improved conformations for 55 cases out of the 74 in the test set (oracle metric, 31 for average), outperforming SBALIGN. We show structures predicted by APOLLODIFF for four *holo* structures from the test set in Fig. 3.

APOLLODIFF and missing residues To evaluate APOLLODIFF's ability to handle missing residues in the reference *apo* structure, we use it with artificially masked *apo* conformations (masking 10 consecutive residues with location chosen at random). We show results in Table 1, where it can be observed that APOLLODIFF's performance in this scenario is on par with the performance obtained without masked residues.

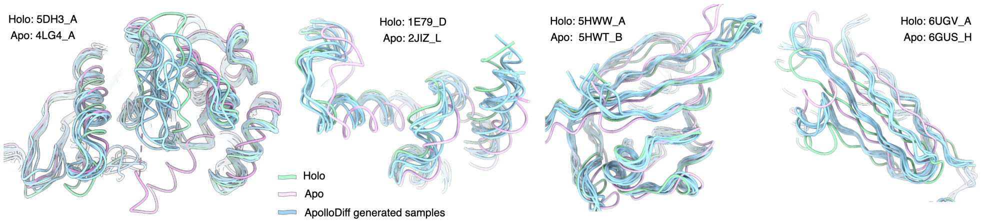

Figure 3: Structures generated by APOLLODIFF without masks for four proteins in the test set. *Apo* structures and all generated samples are aligned with the *holo* structure under consideration.

Global vs Local (Pocket) Conformational Changes To further evaluate our model, we split the test set into two categories, measuring the TMscore[2] for each *holo* structure in the test against all structures in the training set. The structures for which the maximum observed TMscore is less than 0.7 were labeled as *Difficult*, while the remaining structures (maximum TM score greater than 0.7) were labeled as *Easy*. Figure 2 shows average RMSD values between APOLLODIFF predictions and *holo* structures in the test set. The left plot shows the global RMSD (computed using all C_α atoms), and the right plot shows pocket RMSD (computed using C_α atoms within 8Å from the ligand, provided in D3PM database). We observe that in several cases APOLLODIFF generated backbones display lower pocket RMSD than global RMSD.

[2] Briefly, the TMscore [16] ranges between 0 and 1, and measures the similarity between two proteins, taking into account the sequences and structure. Scores close to 1 indicate high similarity.

4 Conclusions

We introduce a novel generative method, APOLLODIFF, to predict protein conformations associated with the binding of small molecules, i.e. transitions from *apo* to *holo* state. While docking of small molecules and virtual screening has a significantly higher success rate if *holo* conformations are used, the associated *holo* structures are rarely experimentally solved. Our method provides a new framework for modelling apo-to-holo conformational changes through a diffusion process with explicit task-optimized priors in the network architecture, such as an implicit bias to fall back on the conditioning *apo* structure. We show despite lack of knowledge of the ligand, APOLLODIFF generates conformations that are closer to *holo* than *apo* in 74% of the cases, compared to 9% in the current state of the art method. Our method thus has the potential to improve the performance of downstream screening campaigns that may rely on accurate modeling of holo structures. While this presents a significant improvement over state of the art, there is further room for improvement, for instance, by augmenting training data with simulated conformations (e.g. obtained via MD simulations), adding additional relevant conditions such as pocket position, and incorporating auxiliary losses penalizing unrealistic conformations in our training pipeline (e.g. steric clashes).

References

[1] Steven Hayward Herman JC Berendsen. Collective protein dynamics in relation to function. *Current Opinion in Structural Biology*, 10:165–169, 2000.

[2] Ferran Feixas, Steffen Lindert, William Sinko, and J. Andrew McCammon. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophysical Chemistry*, 186: 31–45, 2014. ISSN 0301-4622. doi: https://doi.org/10.1016/j.bpc.2013.10.007. URL `https://www.sciencedirect.com/science/article/pii/S0301462213001919`. Special issue : conformational selection.

[3] Enrico Di Cera. Mechanisms of ligand binding. *Biophysics Reviews*, 1(1):011303, 11 2020. ISSN 2688-4089. doi: 10.1063/5.0020997. URL `https://doi.org/10.1063/5.0020997`.

[4] Kei ichi Okazaki and Shoji Takada. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proceedings of the National Academy of Sciences*, 105(32):11182–11187, 2008. doi: 10.1073/pnas.0802524105. URL `https://www.pnas.org/doi/abs/10.1073/pnas.0802524105`.

[5] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015.

[6] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[7] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[8] Emiel Hoogeboom, Vıctor 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.

[9] 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.

[10] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021.

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

[12] 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*, 2023.

[13] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[14] Zhang X. Xu Z. Chen Z. Yang Y. Cai T. Zhu W Peng, C. D3pm: a comprehensive database for protein motions ranging from residue to domain. *BMC bioinformatics doi:10.1186/s12859-022-04595-0*, 23(1):70, 2022.

[15] Vignesh Ram Somnath, Matteo Pariset, Ya-Ping Hsieh, Maria Rodriguez Martinez, Andreas Krause, and Charlotte Bunne. Aligned diffusion schrodinger bridges. *arXiv preprint arXiv:2302.11419*, 2023.

[16] Yang Zhang and Jeffrey Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57(4):702–710, 2004.

[17] 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.

[18] Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

[19] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

[20] 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.

[21] Yeqing Lin and Mohammed AlQuraishi. Generating novel, designable, and diverse protein structures by equivariantly diffusing oriented residue clouds. 2023.

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

[23] Matt McPartlon, Ben Lai, and Jinbo Xu. A deep se (3)-equivariant model for learning inverse protein folding. *bioRxiv*, pages 2022–04, 2022.

[24] Namrata Anand, Raphael Eguchi, Irimpan I Mathews, Carla P Perez, Alexander Derry, Russ B Altman, and Po-Ssu Huang. Protein sequence design with a learned potential. *Nature communications*, 13(1):746, 2022.

[25] Alexey Strokach, David Becerra, Carles Corbi-Verge, Albert Perez-Riba, and Philip M Kim. Fast and flexible protein design using deep graph neural networks. *Cell systems*, 11(4):402–411, 2020.

[26] Chloe Hsu, Robert Verkuil, Jason Liu, Zeming Lin, Brian Hie, Tom Sercu, Adam Lerer, and Alexander Rives. Learning inverse folding from millions of predicted structures. In *International Conference on Machine Learning*, pages 8946–8970. PMLR, 2022.

[27] Jianyi Yang, Ivan Anishchenko, Hahnbeom Park, Zhenling Peng, Sergey Ovchinnikov, and David Baker. Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117(3):1496–1503, 2020.

[28] Diego Del Alamo, Davide Sala, Hassane S Mchaourab, and Jens Meiler. Sampling alternative conformational states of transporters and receptors with alphafold2. *Elife*, 11:e75751, 2022.

[29] Patrick Bryant. Structure prediction of alternative protein conformations. *bioRxiv*, pages 2023–09, 2023.

[30] Christian Léonard. A survey of the schr\" odinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.

[31] Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021.

[32] Wei Lu, Ji-Xian Zhang, Weifeng Huang, Ziqiao Zhang, Xiangyu Jia, Zhenyu Wang, Leilei Shi, Chengtao Li, Peter Wolynes, and Shuangjia Zheng. Dynamicbind: Predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model. 2023.

[33] Greg Landrum et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. *Greg Landrum*, 8:31, 2013.

[34] Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

[35] Jinze Zhang, Hao Li, Xuejun Zhao, Qilong Wu, and Sheng-You Huang. Holo protein conformation generation from apo structures by ligand binding site refinement. *Journal of Chemical Information and Modeling*, 62(22):5806–5820, 2022.

[36] "https://www.uniprot.org/".

[37] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Sal Candido, et al. Language models of protein sequences at the scale of evolution enable accurate structure prediction. *bioRxiv*, 2022.

[38] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

A Diffusion Models

Diffusion models are generative modelling techniques trained to produce samples from a target distribution. They work by learning to reverse a diffusion process that gradually converts clean samples to noise. In this work, we use the variance-preserving formulation from Song et al. [17], which defines this forward process as

$$\mathrm{d}x_t = -\frac{1}{2}\beta(t)x_t\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w_t, \quad \text{where} \quad t \in [0,1] \quad \text{and} \quad x_0 \sim p_{\text{data}}. \tag{2}$$

While many choices for $\beta(t)$ are possible, a widely used one [17] is given by $\beta(t) = t(\beta_{\max} - \beta_{\min}) + \beta_{\min}$. For a sufficiently large $\beta_{\max}$, it can be shown that samples x_1 obtained by simulating the forward process from Eq. (2) approximately satisfy $x_1 \sim \mathcal{N}(0, I)$. Therefore, new samples from the data distribution can be generated by simulating the time-reversed process, given by

$$\mathrm{d}x_t = -\frac{\beta(t)}{2}\big(x_t + 2\nabla \log p_t(x_t)\big)\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}\bar{w}_t, \quad x_1 \sim \mathcal{N}(0, I), \tag{3}$$

where $\nabla \log p_t(x_t)$ represents the *scores* of the marginal density of the forward process at time t. Since these densities and their scores are often unavailable, a *score network* $s_\theta(x_t, t)$ is trained to approximate them minimizing the denoising score matching objective [18, 19]

$$\mathcal{L}(\theta) = \mathbb{E}_{t, p_{\text{data}}(x_0), p_t(x_t \,|\, x_0)} \left[\lambda(t) \left\| s_\theta(x_t, t) - \nabla \log p_t(x_t \,|\, x_0) \right\|^2\right]. \tag{4}$$

Here $p_t(x_t \,|\, x_0)$ represents the transition kernel from the forward SDE, given by $p_t(x_t \,|\, x_0) = \mathcal{N}(x_t \,|\, x_0 e^{-\tilde{\beta}(t)/2}, I - Ie^{-\tilde{\beta}(t)})$, where $\tilde{\beta}(t) = t^2(\beta_{\max} - \beta_{\min})/2 + t\beta_{\min}$. After training, samples are produced by simulating the reverse SDE from Eq. (3) with the score network in place of the true scores.

Diffusion models can be naturally extended to handle conditional distributions. In this case, the dataset consists on pairs (x^i, c^i), the score network takes the conditioning variable c as additional input, $s_\theta(x_t, t, c)$, and the reverse diffusion produces samples from $p(x \,|\, c)$.

B Related Work

Diffusion models are being increasingly used for molecule synthesis [20] and (conditional) protein generation tasks. For instance, many approaches for backbone generation use diffusion models, effectively treating the problem as approximating a distribution over backbones [12, 21, 22]. These approaches often do not take a sequence as input (but a sequence can be produced using inverse folding methods [23–26]).

Diffusion models have also been used for protein folding. Most machine learning methods for this task aim to predict single structures [10, 27], not capturing proteins inherent flexibility. While multiple approaches addressed this by introducing stochastic subsampling (e.g. dropout) and reducing MSA depth [28, 29] withing AlphaFold2 [10], another line of work involves using SE(3) equivariant diffusion models. Given a sequence, the model is trained to approximate the conditional distribution over protein structures compatible with the sequence.

Recently, techniques used for diffusion models (e.g. score matching) have been used to solve the Schrödinger bridge (SB) problem [30, 31]. Briefly, the SB problem involves finding a bridge between two distributions under some prior condition. This is related to the traditional formulation behind diffusion models, which often bridge a tractable Gaussian and the data distribution. The appeal behind SB involves its capacity to bridge two distributions without requiring a tractable reference. However, score matching based approaches for this task often build on Iterative Proportional Fitting (an iterative algorithm to solve the Schrödinger bridge problem), which may impact their efficiency.

Somnath et al. [15] proposed SBALIGN, an alternative approach to solve the SB problem under the assumption that joint samples from the two distributions of interest are available. Their method bypasses the iterative proportional fitting algorithm, and can be trained as efficiently as traditional diffusion models. One of the applications they explore involves predicting *apo*-*holo* conformational changes. They attempt to predict *holo* conformations initializing the diffusion from the reference *apo* structure. The method's main benefit is that, if the *apo* and *holo* conformations are close to each other, the reverse diffusion can be simulated accurately with a relatively small number of steps. However, it relies on the availability of a complete reference *apo* structure (this is in contrast to our model, that can impute unresolved residues).

Concurrently with our work, Lu et al. [32] proposed a flexible molecular docking method based on equivariant diffusion models. Their approach follows a SB-like formulation, producing protein+ligand complexes starting from AlphaFold2 [10] conformations and a specific ligand specification (with an initial conformation obtained with RDkit [33]). Their score network builds on the e3nn architecture [34] to predict the necessary transformations (translations and rotations) of the input protein residues and ligand. This method, however, is focused on a different set of problems, as it cannot handle missing residues in the reference structure, and requires a fully-specified ligand to perform docking (while our approach predicts *holo* conformations without ligand information).

Finally, Zhang et al. [35] designed an improved ensemble docking protocol to generate holo-like conformations of protein from *apo* structures via metadynamics simulations. Unlike our approach, this protocol aims to predict the physical movements of protein over time.

C Dataset and data processing

We use the subset of D3PM [14] used by Somnath et al. [15], which groups proteins by UniProt accession [36] and keeps proteins for which there exists *apo-holo* conformations that differ by more than 3Å. However, we chose a different processing pipeline. While Somnath et al. [15] chose a random set of apo-holo pairs for train, validation and test, we argue that at inference time, frequently, apo-holo pairs are unknown for a given protein. Indeed, a suitable method should be able to predict not only proteins where no corresponding apo-holo pairs were available at training time, but ideally also generalize towards structures with significant structural differences. We find that for 93% of the test data points in Somnath et al. [15], there exists an apo-holo pair in train or validation with an identical UniProt accession, inadequately capturing these scenarios. We thus create our data set splits from two fractions to capture these two scenarios, and provide a breakdown of performance for both fractions in Fig. 2. It is of note that this is a significantly more challenging evaluation scenario than proposed in Somnath et al. [15], but we believe captures real-life usage more accurately.

The D3PM dataset contains 2152 unique PDB chains of which 783 are apo and 1369 are holo, stemming from 589 unique proteins. To have a significant fraction of the test data with sufficient structural and sequence difference from train and validation set (required to assess the second scenario), we calculate pairwise sequence similarity and structural alignment via TMscore. The sequence similarity is defined as $\mathrm{SeqSim}(s_1, s_2) = \frac{1}{L}\sum_i \mathbb{1}\{s_1[i] = s_2[i]\}$, where $\mathbb{1}\{\cdot\}$ stands for the indicator function (it is one if the condition inside the brackets is true, zero otherwise). We cluster by structural and sequence similarity as well as UniProt accession which yields clusters of proteins where the sequence and structure difference is larger than >0.7. We select 20 of these clusters at random as validation and test sets, respectively. To assess the first scenario, i.e. inference on similar but not identical proteins to what is observed during training, we also add apo-holo pairs from an additional 10 UniProt accessions to the validation and test data sets, respectively. In total, this results in 74 unique PDB chains from 32 UniProt accessions in the test set. The statistics of this data set are shown in Table 2.

	# PDB chains	# Apo PDB chains	# Holo PDB chains	# Unique Proteins
Train	1924	690	1234	518
Val	112	51	61	39
Test	116	42	74	32

Table 2: Overview of training, validation, and test sets.

D Model Details

The ESM embeddings, of size 1280, were obtained with the ESMFold [37] esm2_t33_650M_UR50D, and then projected down to a dimension of 256 using a learned linear layer, resulting in $\mathbf{r}$. The pair representation $\mathbf{p}$ was initialized using relative sequence separation between residues (as a one

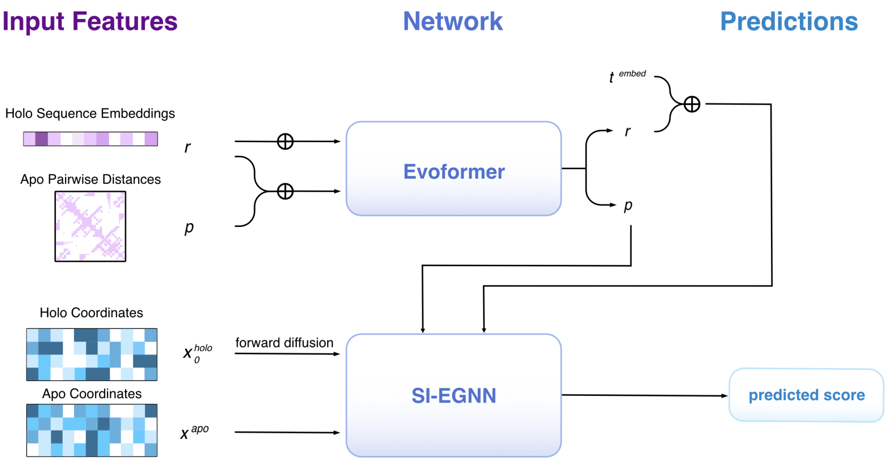

Figure 4: Overview of the score network used by APOLLODIFF.

hot vector of length 65, with sequence separation values ± 32), and the pairwise distances between residues in the reference *apo* structure (binned into bins of width 2Å between 2-16Å, and an additional bin for distances $>$16Å). These two one-hot vectors are concatenated and fed through a learned linear layer, which produces the pair representation of length 128. We use a sinusoidal encoding for the diffusion time t of size 48.

E Training and sampling

During APOLLODIFF training, we randomly mask residues in the reference *apo* structure, to train the model's capabilities to impute missing residues, a common problem in PDB structures. We do this through two random masking mechanisms: (i) randomly masking each residue with a probability p, with $p \sim U(0, L/10)$, and randomly masking a continuous sub-sequence of residues of length $L_{\text{mask}} \sim U(0, 15)$. We first train our model on crops of 100 residues, and finetune on 500 residues. Training took a total of two days on a A100 (80 GB) GPU. We train using Adam [38] with a learning rate of 10^{-4}. For all models we generate five samples per *apo-holo* pair provided.