
Conditioned Protein Structure Prediction

Tengyu Xie*
Westlake University
xietengyu@westlake.edu.cn

Zilin Song*
Westlake University
songzilin@westlake.edu.cn

Jing Huang[†]
Westlake University
huangjing@westlake.edu.cn

Abstract

Deep learning based protein structure prediction has facilitated major breakthroughs in biological sciences. However, current methods struggle with alternative conformation prediction and offer limited integration of expert knowledge on protein dynamics. We introduce AFEXplorer, a generic approach that tailors AlphaFold predictions to user-defined constraints in coarse coordinate spaces by optimizing embedding features. Its effectiveness in generating functional protein conformations in accordance with predefined conditions were demonstrated through comprehensive examples. AFEXplorer serves as a versatile platform for conditioned protein structure prediction, bridging the gap between automated models and domain-specific insights.

1 Introduction

Contemporary structural biology examines not only the architecture of biomacromolecules but also discerns their functional implications, which are of great scientific and therapeutic importance. In recent years, the field has witnessed transformative advancements in computational structural biology empowered by deep learning (DL) approaches such as AlphaFold2 (AF) and comparable methods [1, 2, 3, 4, 5, 6, 7]. These tools have leveraged protein structures determined by laborious experiments such as X-ray crystallography, NMR spectroscopy, or cryogenic electron microscopy, facilitating the widely accepted interim breakthroughs in the highly accurate prediction of *static* protein structures [8]. However, physiologically relevant protein motions, especially in functional response to external perturbations or interactions, are highly dynamic.

Despite successful attempts have been made to capture the ensemble and dynamic nature of protein conformations, effective applications of AF on modern protein science research have yet to be fully realized due to the lack of interoperability between these data-driven models and human intelligence [9]. Current DL models could not generate protein structures with specific local or global structural features, as the domain knowledge from human experts on specific conformational descriptors that distinguish different protein functional states is not embedded [10]. For example, AF lacks the capability to specifically generate the conformational structure of a membrane transporter in its outward-facing state. Such discrepancy impedes the robust and productive applications of AF and related methods in crucial research practices such as the developments of novel therapeutic agents or biosynthetic pathways.

Here, we present AFEXplorer (AFEX), a versatile approach that offers highly accessible user-conditioned protein structure prediction with AF. In brief, AFEX serves as a surrogate protein structure

*Equal contribution

[†]Correspondence: huangjing@westlake.edu.cn

synthesizer that learns the optimal feature representation for AF to generate protein conformations subjected to spatial constraints in a coarse coordinate space.

2 Related Works

2.1 Physics-based modeling

The conformational change of proteins associated with the transition between functional states is often a rare event in protein dynamics. Theoretical understanding of rare event sampling predominantly comes from physics-based modeling. Among various rare event enhanced sampling methods, Targeted Molecular Dynamics (TMD) [11] is especially useful for directing the time evolution of protein conformations towards specific configurations. Basically, TMD introduces to the N -body subsystem an extra penalty potential, U_{bias} , on the instantaneous coordinates $\mathbf{x}(t)$:

$$U_{\text{bias}}(\mathbf{x}(t), \hat{\mathbf{x}}, t) = \frac{k}{N} \|\mathbf{x}(t) - \mathbf{R}\hat{\mathbf{x}} - \mathbf{T}\|_2^2 + U_s(\mathbf{x}(t), \hat{\mathbf{x}}, t) \quad (1)$$

, where k is the force constant, $\hat{\mathbf{x}}$ represents the target subsystem configuration, U_s is a time-scheduled smoothing potential, and the alignment operators $(\mathbf{R}, \mathbf{T}) = \text{argmin}_{(\mathbf{R}, \mathbf{T})} \|\mathbf{x}(t) - \mathbf{R}\hat{\mathbf{x}} - \mathbf{T}\|_2$. As long as the Boltzmann statistics apply, time evolution of the system would approach the targeted configuration under this bias potential. A closely related method for studying protein unfolding dynamics is Steered Molecular Dynamics (SMD), which exerts a pulling force of fictitious origin on the terminal region to distort the folded structure [12]. In this approach, equilibrium thermodynamic observables can be retrieved via the Jarzynski equation in the case of irreversible pulling [13].

2.2 Generating alternative structures with AF

Interest in exploring the conformational ensemble learned by AF has catalyzed an evolving area of research [14, 15]. Indeed, AF or equivalent models can be viewed as a black-box function that maps the multiple sequence alignment (MSA) featurization of amino acid sequences to the corresponding protein spatial coordinates. One could thus characterize multiple protein conformations by feeding AF with randomly perturbed or specifically curated feature representations. For example, Heo and Feig proposed to substitute the default template database with a curated set of GPCR homologues in the active state, aiming to mitigate AF’s bias toward predicting inactive GPCR conformations [16]. Similarly, a strategy that combines MSA subsampling [17] and template manipulation has been exploited to model the functional conformations of kinases and GPCRs [18].

Different ways to subsample the raw MSA inputs have been proposed [17, 19, 20]. Del Alamo et al. adopted stochastic sampling to reduce the depths of MSAs, aiming to drive AF to produce alternative structures with some likelihood of differing from the default AF predictions [17]. Wayment-Steele et al. developed the AF-Cluster method, which clusters MSAs to maximize the coevolutionary decoupling between distinct folds and use these fold-aware MSA clusters to generate alternative structures via AF [19]. A recent benchmark on 16 membrane proteins indicates limited success of these two methods in alternative conformation predictions [21]. Zhang et al. developed an MSA generative model that, by feeding random noises, can generate various MSA sets with the potential to drive AF to produce alternative conformations [22]. Moreover, the AFsample approach proposed by Wallner uniquely sets the dropout layers in AF-Multimer to training mode during inference; models stochasticity induced by random dropouts in the internal AF latent representation naturally lead to diverse protein complex conformation predictions [23]. Recently, Bryant and Noe introduced AFPprofile, which uses AF-Multimer confidence scores as the objective to learn an offset bias that denoises the input MSA for improving AF-Multimer predictions [24].

3 Method

We redefine the conditioned structure prediction as an optimization problem, where the objective is to identify the optimal MSA feature that forces AF to generate the structures that align with our *a priori* knowledge on specific states (Figure 1).

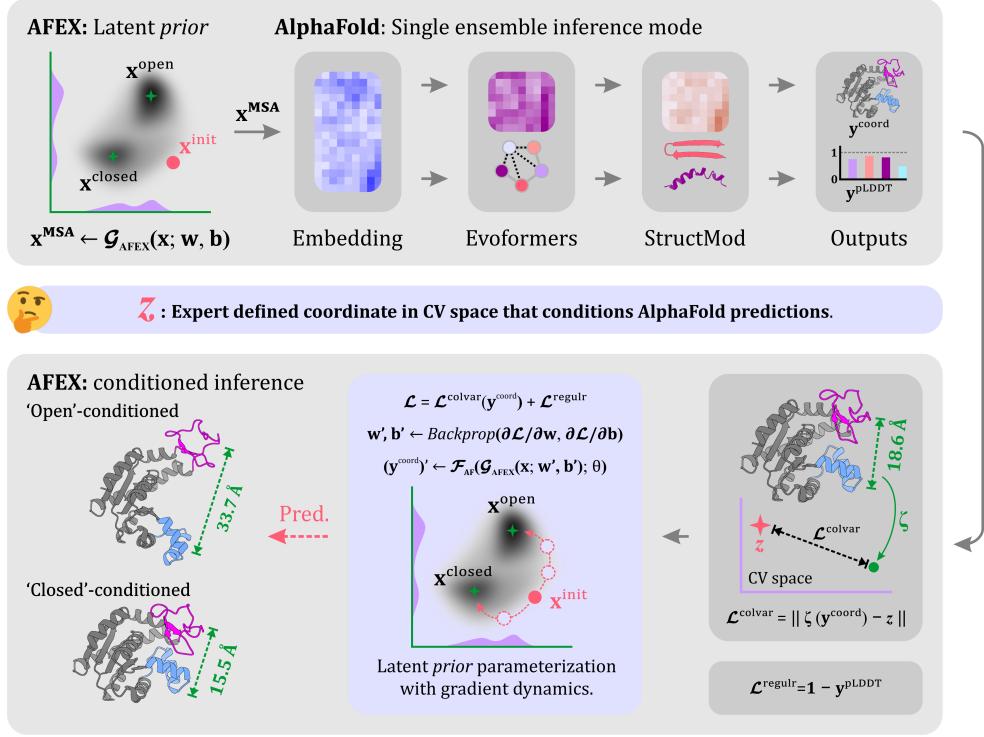


Figure 1: The AFEX protocol is designed to identify MSA features tailored to specific conformational states. It leverages domain knowledge, encoded as a loss function, to optimize features that guide AF in generating the targeted state. The AdK protein serves as an example to demonstrate AFEX’s capability in generating both its open and closed states.

3.1 Algorithm

We begin with the AF inference function, \mathcal{F}_{AF} , whose input could be generally categorized into features of the predicting target (x^{targ}), the MSA profiles (x^{MSA}), and the template profiles (x^{temp}):

$$\mathcal{F}_{\text{AF}} : (x^{\text{targ}}, x^{\text{MSA}}, x^{\text{temp}}) \mapsto (y^{\text{coord}}, y^{\text{PLDDT}}) \quad (2)$$

, where y^{coord} and y^{PLDDT} denotes the predicted Cartesian coordinates and the per-residue confidence of the prediction. Let us introduce the *a prior* knowledge from human experts that a reduced representation $\xi : y^{\text{coord}} \mapsto z^{\text{colvar}}$ could *reasonably* capture the functional evolution of the target protein conformations and we are interested in a specific state, y_z^{coord} , characterized by the Dirac’s δ function $y_z^{\text{coord}} = y^{\text{coord}} \delta(\xi(y^{\text{coord}}) - z)$. To find y_z^{coord} through \mathcal{F}_{AF} , we look for x_z^{MSA} as we keep the target features x^{targ} fixed and drop the template inputs to \mathcal{F}_{AF} :

$$x_z^{\text{MSA}} = \underset{x^{\text{MSA}}}{\operatorname{argmin}} (\|\xi(y^{\text{coord}}) - z\| + \mathcal{L}^{\text{regular}}) \quad (3)$$

, where $\|\cdot\|$ denotes a generic distance measure that defines the collective variable (CV) loss $\mathcal{L}^{\text{colvar}}$, and $\mathcal{L}^{\text{regular}}$ represents the regularization loss. We find that x_z^{MSA} can be parametrized via a generator, $\mathcal{G}_{\text{AFEX}}$:

$$\mathcal{G}_{\text{AFEX}} : x \mapsto x_z^{\text{MSA}} \quad (4)$$

, where x denotes *some* latent prior. This concludes the AFEX framework with Eqs. (2), (3), and (4). While extensions for learning optimal feature representations other than MSAs (e.g., templates) are feasible, we note that many previous methods for diversifying AF predictions based on feature perturbations can be viewed as special cases of AFEX with specific implementations of $\mathcal{G}_{\text{AFEX}}$.

Although more complex differentiable architectures can be parametrized as the latent feature generator, in the present work, we simply implement $\mathcal{G}_{\text{AFEX}}$ as linear activation applied selectively to the `cluster_profiles` of the MSA features:

$$\mathcal{G}_{\text{AFEX}}(\mathbf{x}^{\text{MSA}}; \mathbf{w}, \mathbf{b}) = \mathbf{w}\mathbf{x}^{\text{MSA}} + \mathbf{b} \quad (5)$$

Based on such a linear mapping, the inference graph

$$\mathcal{F}_{\text{AF}} : (\mathcal{G}_{\text{AFEX}}(\mathbf{x}^{\text{MSA}}; \mathbf{w}, \mathbf{b})) \mapsto (\mathbf{y}^{\text{coord}}, \mathbf{y}^{\text{pLDDT}}) \quad (6)$$

naturally casts to the optimization problem:

$$\mathbf{w}, \mathbf{b} = \underset{\mathbf{w}, \mathbf{b}}{\operatorname{argmin}} (\|\xi(\mathbf{y}^{\text{coord}}) - z\| + \mathcal{L}^{\text{regulr}}) \quad (7)$$

with the regularization loss defined using pLDDT: $\mathcal{L}^{\text{regulr}} = 1 - \mathcal{F}_{\text{AF}}(\mathcal{G}_{\text{AFEX}}(\mathbf{x}^{\text{MSA}}; \mathbf{w}, \mathbf{b}))^{\text{pLDDT}}$.

3.2 Implementation details

The implementation of AFEX requires no modifications to the AF source codes (as per Eq. (6)). The AF runtime is kept in inference mode, with a single MSA feature ensemble. The CV losses used for each tested system are detailed in their respective Results sections. We parametrized each linear AFEX learner with the Adam optimizer [25]. The implementation leverages the Optax library within the JAX ecosystem [26]. AFEX is publicly available at: <https://github.com/JingHuangLab/AFEXplorer>.

4 Results

4.1 AFEX generates AdK open and closed states

Adenylate kinase (AdK) is a classic system for studying large-scale conformational changes [27]. We define the CV as the distance (D) between the $\text{C}\alpha$ atoms of A37 in the AMP lid and R124 in the ATP lid (*E. coli* AdK sequence). The corresponding loss function is defined using the sigmoid function $\sigma(x) = 1/(1 + \exp(-x))$ as $\mathcal{L}_{\text{open}}^{\text{colvar}} = -\log[\sigma(D - z)]$ and $\mathcal{L}_{\text{closed}}^{\text{colvar}} = -\log[\sigma(z - D)]$. z is set to 25 Å and $\mathcal{L}^{\text{regulr}}$ is added to $\mathcal{L}_{\text{open}}^{\text{colvar}}$ or $\mathcal{L}_{\text{closed}}^{\text{colvar}}$ to formulate the final loss function of AFEX. A 500-step optimization of \mathbf{w} and \mathbf{b} drives the AF prediction towards the open conformation ($D = 33.7$ Å) and the closed conformation ($D = 15.5$ Å), compared to the initial D value of 18.6 Å (Figure 1). The overlap with experimental AdK structures 4AKE ($D = 33.6$ Å) and 1AKE ($D = 16.1$ Å) confirms that highly accurate structure predictions for these two functional states were obtained. The $\text{C}\alpha$ RMSDs are 1.60 Å for the open and 1.83 Å for the closed conformation, respectively.

4.2 AFEX generates DFG-in/out conformations for all human kinases

Kinases are important drug targets due to their central role in regulating cellular signaling pathways. They are highly dynamic, capable of adopting various conformations such as the DFG-in and DFG-out states [28]. The DFG-out conformation is important because it creates a binding pocket that facilitates the design of type-II inhibitors, with imatinib (Gleevec) serving as a prominent example. Capturing the DFG-out conformation for a kinase remains a challenge even with the advent of AF, as AF exhibits a strong bias towards the DFG-in state [29, 30]. Modi and Dunbrack identified two CVs, D_1 and D_2 , for distinguishing DFG-in and DFG-out states. D_1 is the distance between the $\text{C}\zeta$ atom of the Phe residue in the DFG motif and the $\text{C}\alpha$ atom of the fourth residue following the conserved Glu in the C-helix, while D_2 is the distance between the Phe $\text{C}\zeta$ atom and the $\text{C}\alpha$ atom of the conserved Lys in the $\beta 3$ strand [28]. DFG-in is indicated by $D_1 \leq 11$ Å and $D_2 \geq 11$ Å, while $D_1 > 11$ Å and $D_2 \leq 14$ correspond to DFG-out.

Accordingly, we perform AFEX calculations using the CV loss defined as $\mathcal{L}_{\text{DFG-in}}^{\text{colvar}} = -\log[\sigma(z_1 - D_1)] - \log[\sigma(D_2 - z_2)]$ for DFG-in and $\mathcal{L}_{\text{DFG-out}}^{\text{colvar}} = -\log[\sigma(D_1 - z_3)] - \log[\sigma(z_4 - D_2)]$ for DFG-out, with $z_1 = z_2 = z_3 = 11$ Å and $z_4 = 14$ Å. Combined with $\mathcal{L}^{\text{regulr}}$, we generated specific DFG-in and DFG-out conformations for 453 human kinases. As shown in Fig. 2, 99.3% of DFG-in and 99.1% of DFG-out conditioned predictions were successful. In contrast, AF generates

predominantly DFG-in states with very narrow distributions (Fig. 2d). For a specific kinase, the two conformations were generally similar, differing mainly in the local structural arrangement we conditioned on. Associated variations in adjacent structural elements, such as the activation loop, were also observed (Fig. 2e). While AFEX expands the accessible conformational space of human kinases, the distributions of D_1 and D_2 differ from those in the PDB as plotted in Fig. 2c using 7,978 experimental kinase structures. Further refinement may be achievable through more sophisticated definition of CV loss.

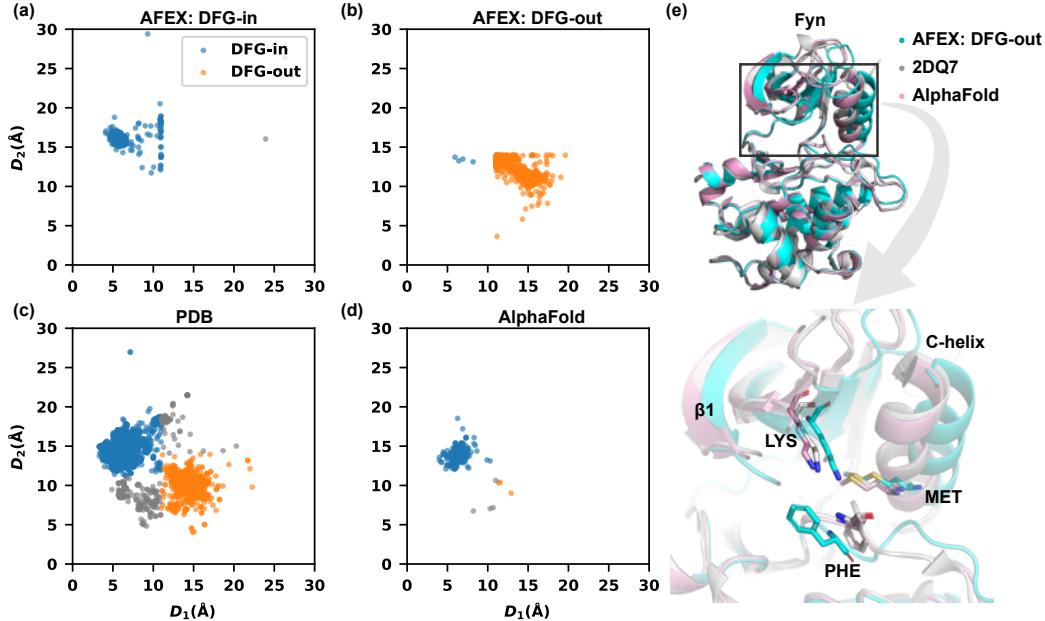


Figure 2: AFEX successfully generated alternative states (DFG-in and DFG-out) for human kinases. The Fyn kinase is shown as an example.

4.3 AFEX generates inward-facing (IF) and outward-facing (OF) states of transporters

Obtaining alternative conformations for membrane transporters is notoriously difficult for both experimental and computational techniques. This is easily achievable with AFEX, as demonstrated using four transporters from the IOMemP dataset [21]. Similarly, we define the CV loss as $\mathcal{L}_{\text{IF}}^{\text{colvar}} = -\log[\sigma(z - D)]$ and $\mathcal{L}_{\text{OF}}^{\text{colvar}} = -\log[\sigma(D - z)]$, in which D is the $C\alpha$ distance between a pair of extracellular gating residues and $z = 10$ Å. With 500-step AFEX optimizations, conformational changes were induced towards either the IF or the OF states for all four proteins (Fig. 3). Take mdfA as an example: its initial structure with AF (grey) is in the IF state, and thus shows minimal change under AFEX conditioned on IF. However, when conditioned on OF, AFEX guides the protein into a OF conformational state. For PF0807, the initial AF structure corresponds to the OF state, and its IF state was successfully deduced by AFEX. Such a straightforward way to generate alternative states for membrane proteins will facilitate mechanistic studies of their functions.

5 Conclusion

In this study, we introduce AFEX as a versatile method that optimizes MSA embedding features under predefined conditions, guiding AF to explore conformational spaces for structures that meet these conditions. Using kinases and transporters as examples, we demonstrated that AFEX provides a controllable way for both the exploration and exploitation of conformational diversity. The condition, defined using CV loss, incorporates domain-specific insights. While simple, nearly binary variables were employed in this study, AFEX can condition on any local or global structure features that can be described by atomic coordinates. Adding scoring functions to the CV loss could further extend AFEX into tools for fully flexible protein-protein, peptide-protein, or ligand-protein docking.

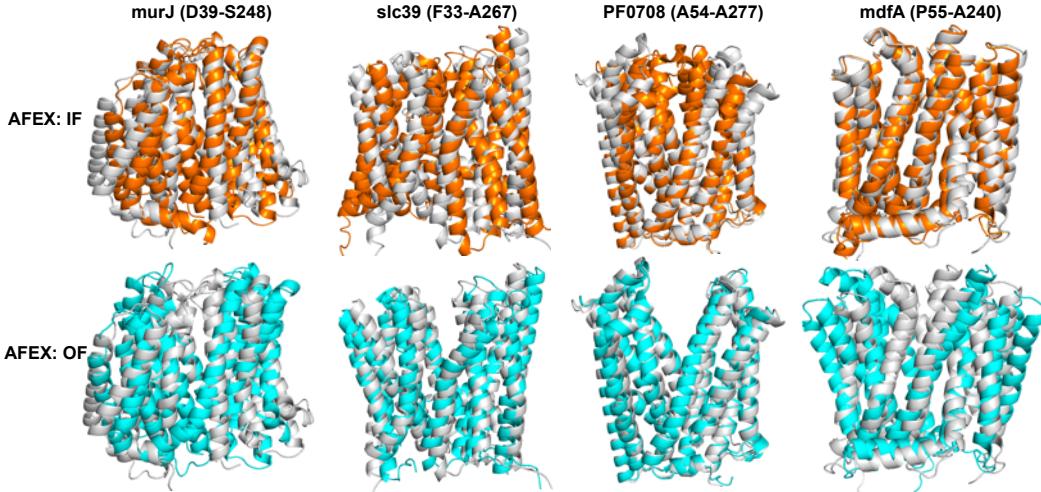


Figure 3: IF (orange) and OF (cyan) states generated by AFEX optimization, each aligned with the initial structure predicted by AF (grey). Protein names and selected gating residues are labeled.

Acknowledgments and Disclosure of Funding

The work is supported by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (2023C03109) and the National Natural Science Foundation of China (32171247, 21803057). We thank the Westlake University Supercomputer Center for computational resources and related assistance.

References

- [1] Andrew W. Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander W. R. Nelson, Alex Bridgland, Hugo Penedones, Stig Petersen, Karen Simonyan, Steve Crossan, Pushmeet Kohli, David T. Jones, David Silver, Koray Kavukcuoglu, and Demis Hassabis. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020.
- [2] 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.
- [3] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021.
- [4] Gustaf Ahndritz, Nazim Bouatta, Sachin Kadyan, Qinghui Xia, William Gerecke, Timothy J O’Donnell, Daniel Berenberg, Ian Fisk, Niccolò Zanichelli, Bo Zhang, et al. OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. *bioRxiv*, 2022. DOI: 10.1101/2022.11.20.517210.
- [5] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: Making protein folding accessible to all. *Nature methods*, 19(6):679–682, 2022.
- [6] Ruidong Wu, Fan Ding, Rui Wang, Rui Shen, Xiwen Zhang, Shitong Luo, Chenpeng Su, Zuofan Wu, Qi Xie, Bonnie Berger, et al. High-resolution de novo structure prediction from primary sequence. *bioRxiv*, 2022. DOI: 10.1101/2022.07.21.500999.
- [7] 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.

- [8] Thomas J Lane. Protein structure prediction has reached the single-structure frontier. *Nature Methods*, 20(2):170–173, 2023.
- [9] Alexander B Tong, Jason D Burch, Daniel McKay, Carlos Bustamante, Michael A Crackower, and Hao Wu. Could AlphaFold revolutionize chemical therapeutics? *Nature Structural & Molecular Biology*, 28(10):771–772, 2021.
- [10] Carmen Al-Masri, Francesco Trozzi, Marcel Patek, Anna Cichońska, Balaguru Ravikumar, and Rayees Rahman. Investigating the conformational landscape of AlphaFold2-predicted protein kinase structures. *bioRxiv*, 2022. DOI: 10.1101/2022.12.02.518928.
- [11] J Schlitter, M Engels, P Krüger, E Jacoby, and A Wollmer. Targeted molecular dynamics simulation of conformational change-application to the T ↔ R transition in insulin. *Molecular Simulation*, 10(2-6):291–308, 1993.
- [12] Barry Isralewitz, Mu Gao, and Klaus Schulten. Steered molecular dynamics and mechanical functions of proteins. *Current opinion in structural biology*, 11(2):224–230, 2001.
- [13] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.
- [14] D Sala, F Engelberger, HS Mchaourab, and J Meiler. Modeling conformational states of proteins with AlphaFold. *Current Opinion in Structural Biology*, 81:102645, 2023.
- [15] Yingnan Hou, Tengyu Xie, Liuqing He, Liang Tao, and Jing Huang. Topological links in predicted protein complex structures reveal limitations of AlphaFold. *Commun. Biol.*, 6:1098, 2023.
- [16] Lim Heo and Michael Feig. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins: Structure, Function, and Bioinformatics*, 90(11):1873–1885, 2022.
- [17] 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.
- [18] Davide Sala, Peter W Hildebrand, and Jens Meiler. Biasing AlphaFold2 to predict GPCRs and kinases with user-defined functional or structural properties. *Frontiers in Molecular Biosciences*, 10:1121962, 2023.
- [19] Hannah K. Wayment-Steele, Adedolapo Ojoawo, Renee Otten, Julia M. Apitz, Warintra Pit-sawong, Marc Hömberger, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Prediction of multiple conformational states by combining sequence clustering with AlphaFold2. *Nature*, 2023.
- [20] Richard A. Stein and Hassane S. Mchaourab. SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with alphafold2. *PLOS Computational Biology*, 18(8):1–16, 08 2022.
- [21] Tengyu Xie and Jing Huang. Can protein structure prediction methods capture alternative conformations of membrane proteins? *bioRxiv*, 2023. DOI: 10.1101/2023.08.04.552045.
- [22] Jun Zhang, Sirui Liu, Mengyun Chen, Haotian Chu, Min Wang, Zidong Wang, Jialiang Yu, Ningxi Ni, Fan Yu, Diqing Chen, Yi Isaac Yang, Boxin Xue, Lijiang Yang, Yuan Liu, and Yi Qin Gao. Few-shot learning of accurate folding landscape for protein structure prediction. *arXiv preprint arXiv:2208.09652*, 2022.
- [23] Björn Wallner. AFsample: Improving multimer prediction with AlphaFold using aggressive sampling. *Bioinformatics*, 39:btad573, 2023.
- [24] Patrick Bryant and Frank Noe. Improved protein complex prediction with AlphaFold-multimer by denoising the MSA profile. *bioRxiv*, 2023. DOI: 10.1101/2023.07.04.547638.
- [25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

- [26] Igor Babuschkin, Kate Baumli, Alison Bell, Surya Bhupatiraju, Jake Bruce, Peter Buchlovsky, David Budden, Trevor Cai, Aidan Clark, Ivo Danihelka, Antoine Dedieu, Claudio Fantacci, Jonathan Godwin, Chris Jones, Ross Hemsley, Tom Hennigan, Matteo Hessel, Shaobo Hou, Steven Kapiturowski, Thomas Keck, Iurii Kemaev, Michael King, Markus Kunesch, Lena Martens, Hamza Merzic, Vladimir Mikulik, Tamara Norman, George Papamakarios, John Quan, Roman Ring, Francisco Ruiz, Alvaro Sanchez, Laurent Sartran, Rosalia Schneider, Eren Sezener, Stephen Spencer, Srivatsan Srinivasan, Miloš Stanojević, Wojciech Stokowiec, Luyu Wang, Guangyao Zhou, and Fabio Viola. The DeepMind JAX Ecosystem. <https://github.com/google-deepmind/jax>, 2020.
- [27] Katherine A. Henzler-Wildman, Vu Thai, Ming Lei, Maria Ott, Magnus Wolf-Watz, Tim Fenn, Ed Pozharski, Mark A. Wilson, Gregory A. Petsko, Martin Karplus, Christian G. Hübner, and Dorothee Kern. Intrinsic motions along an enzymatic reaction trajectory. *Nature*, 450(7171):838–844, 2007.
- [28] Vivek Modi and Roland L. Dunbrack. Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences*, 116(14):6818–6827, 2019.
- [29] Bodhi P. Vani, Akashnathan Aranganathan, and Pratyush Tiwary. Exploring kinase DFG loop conformational stability with AlphaFold2-RAVE. *arXiv preprint arXiv:2309.03649*, 2023.
- [30] Noah B. Herrington, David Stein, Yan Chak Li, Gaurav Pandey, and Avner Schlessinger. Exploring the druggable conformational space of protein kinases using AI-generated structures. *bioRxiv*, 2023. DOI: 10.1101/2023.08.31.555779.