# FUSING NEURAL AND PHYSICAL: AUGMENT PROTEIN CONFORMATION SAMPLING WITH TRACTABLE SIMULATIONS

**Jiarui Lu[1,2], Zuobai Zhang[1,2], Bozitao Zhong[1,2], Chence Shi[1,2], Jian Tang[1,3,4]**
[1]Mila - Québec AI Institute, [2]Université de Montréal
[3]HEC Montréal, [4]CIFAR AI Chair
{jiarui.lu, zuobai.zhang, bozitao.zhong, chence.shi}@mila.quebec,
jian.tang@hec.ca

## ABSTRACT

The protein dynamics are common and important for their biological functions and properties, the study of which usually involves time-consuming molecular dynamics (MD) simulations *in silico*. Recently, generative models has been leveraged as a surrogate sampler to obtain conformation ensembles with orders of magnitude faster and without requiring any simulation data (a "zero-shot" inference). However, being agnostic of the underlying energy landscape, the accuracy of such generative model may still be limited. In this work, we explore the few-shot setting of such pre-trained generative sampler which incorporates MD simulations in a tractable manner. Specifically, given a target protein of interest, we first acquire some seeding conformations from the pre-trained sampler followed by a number of physical simulations in parallel starting from these seeding samples. Then we fine-tuned the generative model using the simulation trajectories above to become a target-specific sampler. Experimental results demonstrated the superior performance of such few-shot conformation sampler at a tractable computational cost.

## 1 INTRODUCTION

Elucidating protein dynamics stands as a significant yet challenging problem during the study of protein functionality and regulation. Such macromolecule can go through transitions between multiple conformational states for different length or time scales. For example, the SARS-Cov-2 spike protein was found to have transitions between its open and closed states (Gur et al., 2020) for its function. Experimental instruments including crystallographic B-factors and NMR spectroscopy can be leveraged to probe such dynamics but only available to a limited spatial and temporal scale.

Traditionally, computational simulation methods especially molecular dynamics (MD) are used to simulate the dynamic behavior of these biological molecules. MD operates by simply evolving the Newtonian equation on the whole system of particles where the accelerations, or the gradients of velocity, are determined by a pre-specified force field (energy, or unnormalized density). To study the protein dynamics, the time scale of MD simulations can reach micro- to milli-seconds in order to completely capture the behaviors such as the transition from unfolding to folding (Lindorff-Larsen et al., 2011). However, simulating protein systems in such time scales is very computationally intensive, which can take several hundreds of GPU days , depending on the size of the system.

With the emergence of accurate structure prediction models (Jumper et al., 2021; Baek et al., 2021; Lin et al., 2023), deep generative models are developed to serve as efficient surrogate for MD simulations, such as Str2Str (Lu et al., 2024). Str2Str learns to explore the conformation space by training on protein structures from the Protein Data Bank (PDB) in an amortized way, and is ready to perform zero-shot conformation sampling for any unseen test proteins. The ability of direct sampling makes such sampler become orders of magnitude more efficient than traditional MD simulations. The inference of such diffusion sampler is based on successful transfer learning of the naturally occurred protein geometries. However, training solely on PDB data can make it fail to take care of

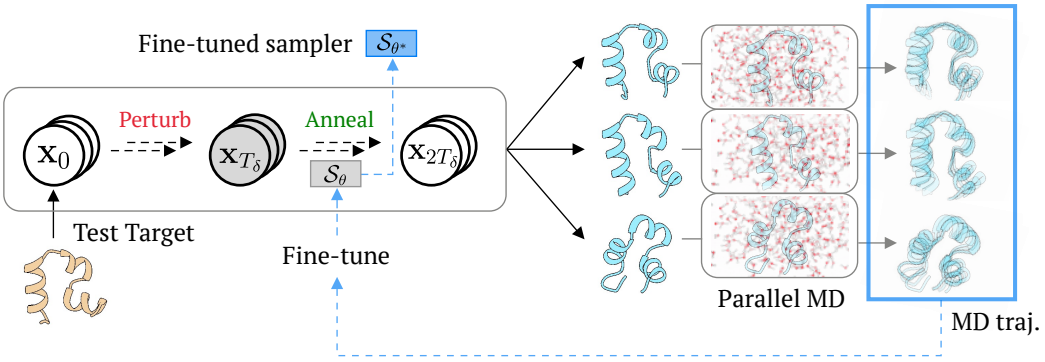



Figure 1: Illustrative diagram of fine-tuning the pre-trained diffusion sampler Str2Str (Lu et al., 2024). Firstly, initial conformation samples of the target protein are generated from a pre-trained network $\mathcal{S}_\theta$ parameterized by $\theta$, followed by parallel MD simulations respectively for each sample. The production trajectories are leveraged to make a target-specific sampler $\mathcal{S}_{\theta^*}$ via fine-tuning.

the underlying energy landscape, which is, on the contrary, the fundamental consideration in MD simulation to generate Boltzmann-distributed ensembles.

In this work, we aim to improve the diffusion sampler with tractable MD simulations via fine-tuning, forming a "few-shot"[1] setting of Str2Str (Lu et al., 2024). To be specific, given a target protein during inference, zero-shot samples are firstly generated from the pre-trained conformation sampler, followed by a number of "roll-out", i.e. short MD simulations respectively initialized by each sample, which is highly parallelizable and thus efficient. Then the production trajectories are leveraged to fine-tune the sampler *ad hoc* for the target protein and reinforce the sampling accuracy. Experimental results demonstrate that such process significantly improve the quality of conformation sampling, at the cost of tractable computations compared to conventional simulations from scratch.

## 2 Related Work

**Protein conformation sampling** Prior to this work, several methods have been proposed to sample protein conformations. Depending on the requirement of simulation data for training, these methods can be divided into two categories: **(1) Require simulation data.** Boltzmann generator (Noé et al., 2019) uses normalizing flow to approach the underlying Boltzmann distribution by learning from simulation data, which is one of the pioneer works among this research direction. Arts et al. (2023) instead adopts the denoising diffusion model to learn such distribution over the coarse-grained protein conformations. idpGAN (Janson et al., 2023) involves a conditional generative adversarial network trained on intrinsic disordered proteins. DiG (Zheng et al., 2023) trains a conditional diffusion model with both PDB and in-house simulation data and can sample conformations conditioned on input sequence. **(2) Require no simulation data.** AlphaFold2-RAVE (Vani et al., 2023) injects stochasticity to the structure output by making modifications to the input channel of AlphaFold2 (Jumper et al., 2021). EigenFold (Jing et al., 2023) revisits the structure prediction (folding) problem in a generative view and is able to sample from the distribution over conformations given the input sequence. Str2Str (Lu et al., 2024) is an equivariant diffusion sampler trained solely on PDB data and formulates the conformation sampling in a structure-to-structure manner.

***De novo* design of protein structure** Despite different purpose, protein structure design methods can also be relevant because of similar techniques used for modeling. Anand & Achim (2022) and Shi et al. (2022) leverage the backbone architecture of AlphaFold2 but feed forward from the input conditions such as contact map. Afterwards, generative models are applied to learn from PDB and *de novo* design novel backbone structures: ProtDiff (Trippe et al., 2022), FoldingDiff (Wu et al., 2022), RFDiffusion (Watson et al., 2023), Chroma (Ingraham et al., 2023), FrameDiff (Yim et al., 2023) and FoldFlow (Bose et al., 2023) to name a few. These sequence-agnostic generative models can

[1]Here we define "few-shot" in the sense that a limited quota of physical simulations for the test protein can be acquired, in contrast to the zero-shot setting where no simulation sample is present.

design non-native backbone structures, based on which de novo sequences can further be designed by an inverse folding model, such as ProteinMPNN (Dauparas et al., 2022).

# 3 METHODOLOGY

## 3.1 PRELIMINARIES

**Protein structure.** Protein structure consists of a sequence of amino acids (or residues) which respectively defines a set of atoms existing in three-dimension (3D) space. Formally, a protein with $L$ residues, while the residue $i$ ($1 \leq i \leq L$) is composed of $n_i$ atoms, can be represented by the Euclidean coordinates of atoms $\mathbf{x} \in \mathbb{R}^{N\times 3}$, where $N = \sum_{i=1}^{L} n_i$ is the totally number of atoms.

**Score-based generative models.** Score-based generative models (SGMs) (Song et al., 2020) aim to model the data distribution via a diffusion process defined by the Itô stochastic differential equation (SDE):

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}, \tag{1}$$

where continuous time $t \in [0, T]$, $\mathbf{f}(\mathbf{x}, t) \in \mathbb{R}^n$ and $g(t) \in \mathbb{R}$. The $\mathbf{w} \equiv \mathbf{w}(t) \in \mathbb{R}^n$ represents the standard Wiener process. Then, the corresponding backward SDE can be expressed as follow (Anderson, 1982):

$$\mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}, \tag{2}$$

where $\mathrm{d}t$ is negative timestep and $\bar{\mathbf{w}}$ is the standard Wiener process with time $t$ flowing from $T \to 0$.

**Zero-shot conformation sampler.** Str2Str (Lu et al., 2024) leveraged SGMs to sample protein backbone conformation via a perturbation followed by an annealing process. Str2Str was trained solely with the protein structures in PDB and requires no simulation data during both training and inference. Str2Str demonstrated the effectiveness of SGMs for conformation sampling in a transfer learning manner, which shows promise when long MD simulation data is not readily available.

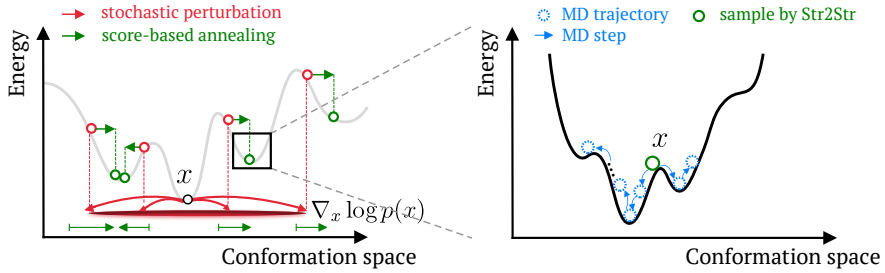
**Molecular dynamics simulation.** Molecular dynamics (MD) simulations amount to evolve the whole particle system through time $T^{\text{sim}} > 0$ by the physical dynamics $d\mathbf{x} = \mathbf{v}(\mathbf{x}, \tau)d\tau$, where $\mathbf{v} \in \mathbb{R}^{N\times 3}$ is the velocity field. With some force field defining $U(\mathbf{x})$ as potential energy, the velocity can be updated through some dynamics such as Newtonian equation[2]: $d\mathbf{v}(\mathbf{x}, \tau) = \bar{\mathbf{m}}\nabla_{\mathbf{x}} U(\mathbf{x})\ d\tau$, where $\bar{\mathbf{m}}$ is the inverse mass of atoms in the system. At each time step, the position and velocity can be updated by an integrator such as Verlet (Verlet, 1967), yielding a simulation trajectory.

## 3.2 NEIGHBORHOOD EXPLOITATION WITH SHORT MD

**The dilemma in direct sampling and simulation.** One disadvantage of Str2Str is that it has no prior knowledge about what do good conformations of the test protein look like during inference. On the other hand, MD simulations simply query a universal oracle (force field) at each time step, being generalizable and training-free, but have a hard time sufficiently exploring the conformation space within tractable computation for simulation. Previous methods, such as Arts et al. (2023), proposed to train using the simulation trajectory of a target protein and sample like a generative model. Although achieving good performance, such method however requires expensive simulation data beforehand and cannot quickly generalize to other proteins beyond the training target, which in practice may have restricted potential utility. Here arises an interesting question: *how to organically combine the advantages of both generative models and physical simulations?*

**Exploration-exploitation balance.** To answer this question, we propose to sample protein ensembles via a two-stage sampling, or exploration-exploitation of the conformation space. Given a target protein of interest, we firstly leverage the Str2Str to sample a set of plausible conformations $X \equiv \{\mathbf{x}^{(i)}\}_{i=1}^{m}$, which can be seemingly good but not energy favorable. On top of that, a short MD simulation can be run for each sample $\mathbf{x}^{(i)}$ to produce (locally) equilibrated conformations. The resulting ensemble $X'$ will be the collection of conformations from each individual trajectory.

[2]Sometimes referred to as Hamiltonian dynamics. Besides, stochastic dynamics can also be applied such as Langevin dynamics to update the velocity states.




(a) Str2Str sampling (w/o force field) (b) Local simulation (w/ force field)

Figure 2: Illustration of the conformation sampling scenario of Str2Str. In (a), sampling is performed in two steps to obtain independent samples of the target protein, where the energy landscape (or information of the force field) is unknown and colored in **gray**. In (b), a hypothetical zoom-in neighborhood of a sample is shown. Due to the complex conformation landscape, the samples directly generated by Str2Str are probably not potential energy-optimal. Short MD simulation can be run to obtain locally equilibrated samples which can be used to fine-tune the pre-trained sampler.

The short MD simulations in the second stage exploits the initial swarm explored by the pre-trained diffusion sampler. The sampling with neighborhood exploitation (**Str2Str-NE**) enjoys (1) better exploration than conventional serial MD simulations under limited computation quota, as well as (2) better exploitation than diffusion sampler alone due to the awareness of force field.

### 3.3 Fine-tuning of diffusion sampler

The two-stage sampling proposed above can itself serve as a good sampler. Alternatively, given the refined conformations $X'$ as dataset, one can incorporate the supervision from the force field of MD simulation into the diffusion sampler Str2Str via fine-tuning (**Str2Str-FT**). In contrast to Arts et al. (2023) which involves expensive long MD trajectory for training, the acquisition of the production trajectories is very efficient: (1) The sampling (exploration) cost of diffusion sampler is negligible to MD simulations; (2) The computation of the local exploitation from short simulations can be readily paralleled according to the number of prior samples and enjoy perfect scalability.

$$\mathcal{L}_{\text{FT}} = \mathbb{E}_{t\in[0,T]}\, \mathbb{E}_{\mathbf{x}_0\sim p(\mathbf{x})} \left\{ \omega(t,\mathbf{x}_0)\, \mathbb{E}_{\mathbf{x}_t\sim p(\mathbf{x}_t|\mathbf{x}_0)} \left[ \|\mathbf{s}_\theta(\mathbf{x}_t,t) - \nabla_{\mathbf{x}_t} \log p_{t|0}(\mathbf{x}_t|\mathbf{x}_0)\|^2 \right] \right\}, \tag{3}$$

where $\omega(t,\mathbf{x}_0) > 0$ is a positive loss reweighting function depending both on time $t$ and sample $\mathbf{x}_0$, and $\mathbf{x}_t \sim p_{t|0}(\mathbf{x}_t|\mathbf{x}_0)$ is defined by the corresponding perturbation kernel. Among them, $\omega(t,\mathbf{x}_0)$ is a natural generalization of the importance sampling (IS) scheme for time in the training of diffusion model (Song et al., 2020). The pseudo-code of the whole process is described in Algorithm 1.

**Algorithm 1** Fine-tuning of a pre-trained conformation sampler.

**Require:** Target protein X, pre-trained diffusion sampler $\mathcal{S}_\theta$, MD simulator $\mathcal{M}$, simulation temperature $\beta$, simulation time $T^{\text{sim}}$.
$\{\mathbf{x}^{(i)}\}_{i=1}^m \sim \mathcal{S}_\theta(X)$ // Generate $m$ initial samples
$D_X = \varnothing$ // Initialize fine-tuning dataset
**for** $i = 1$ to $m$ **do**
    $\{\hat{\mathbf{x}}_j^{(i)}\}_{j=1}^{n_i} \sim \mathcal{M}(\mathbf{x}^{(i)}; \beta, T^{\text{sim}})$ // Run short MD simulation starting from $x_i$
    $D_X \leftarrow D_X \cup \{\hat{\mathbf{x}}_j^{(i)}\}_{j=1}^{n_i}$ // Update dataset with the trajectory of size $n_i$
Update $\theta \to \theta^*$ based on $D_X$ by training on Eq. 3.
**return** $s_{\theta^*}$

## 4 Experiments

**Setup.** To validate the effectiveness of the proposed fine-tuning pipeline, we evaluate the sampling performance on the fast-folding protein set (Lindorff-Larsen et al., 2011). For baselines, we compare

the fine-tuned Str2Str (namely Str2Str-FT) against: MSA subsampling (Del Alamo et al., 2022), EigenFold (Jing et al., 2023), idpGAN (Janson et al., 2023), and Str2Str (pre-trained only) (Lu et al., 2024). The size of sampled ensembles is aligned to be 1,000 to make fair comparison. We follow the evaluation metrics used in Lu et al. (2024) and list them as below: (1) Validity: *Val-Clash*, *Val-Bond* evaluates whether the conformations obey basic geometric constraints w.r.t. steric clash and bonding association. These metrics are the higher the better (↑). (2) Distance: *JS-PwD*, *JS-TIC*, *JS-Rg* assess the distance between the model distribution and reference distribution, which includes the pairwise distance (PwD), time-lagged independent components (TIC) (Naritomi & Fuchigami, 2011; Pérez-Hernández et al., 2013) and radius of gyration (Rg), which are the lower the better (↓).

**Benchmark results.** In this study, we limit the few-shot budget of MD production simulations to be *100ns* in total, which is computationally tractable and lasts only several GPU hours. Under this budget, we firstly sample for each target 100 conformations from a pre-trained Str2Str followed by 100 short simulations per 1 ns long. As shown in Table 1, both the sampler with neighborhood exploitation (Str2Str-NE) and the fine-tuned version (Str2Str-FT) achieve the state-of-the-art performance across all three distance metrics while keeping good validity.

Table 1: Benchmark results of conformation sampling methods on fast folding proteins (Lindorff-Larsen et al., 2011). Metrics are averaged across all fast-folding targets for each method. The results of Str2Str is reproduced by running the open-source code, while the results of other baselines are taken from Lu et al. (2024). The best result for each section is **bolded**.

| Methods | Val-Clash(↑) | Val-Bond(↑) | JS-PwD(↓) | JS-TIC(↓) | JS-Rg (↓) |
|---|---|---|---|---|---|
| MSA subsampling | **0.999** | 0.997 | 0.634 | 0.624 | 0.656 |
| EigenFold | 0.812 | 0.874 | 0.530 | 0.497 | 0.666 |
| idpGAN | 0.960 | 0.032 | 0.480 | 0.517 | 0.661 |
| Str2Str | 0.977 | 0.982 | 0.348 | 0.400 | 0.365 |
| Str2Str-NE (ours) | 0.990 | **1.000** | **0.294** | 0.369 | **0.331** |
| Str2Str-FT (ours) | 0.966 | 0.948 | 0.303 | **0.366** | 0.350 |

**Ablation study.** To better demonstrate the effectiveness of the proposed Str2Str-FT, we also consider ablation experiments. The base model is the (1) Str2Str-FT sampler which is fine-tuned by samples from Str2Str-NE. We (2) replace the fine-tune data with conventional MD simulation of 100ns; and (3) (4) ablate the pre-training from both settings. The results are shown in Table 2.

Table 2: Ablation results on different strategies of training. "NE" indicates the training set is from the samples of Str2Str-NE introduced above while "MD" from the conventional MD simulations. "FT" means model parameters are initialized from the PDB pre-trained checkpoint (i.e. Str2Str).

| Settings | PDB | MD | NE | Val-Clash(↑) | Val-Bond(↑) | JS-PwD(↓) | JS-TIC(↓) | JS-Rg (↓) |
|---|---|---|---|---|---|---|---|---|
| FT on NE data | ✓ | ✗ | ✓ | 0.966 | 0.948 | **0.303** | **0.366** | **0.350** |
| FT on MD data | ✓ | ✓ | ✗ | **0.999** | **0.961** | 0.538 | 0.570 | 0.544 |
| NE data | ✗ | ✗ | ✓ | 0.227 | 0.562 | 0.398 | 0.406 | 0.473 |
| MD data | ✗ | ✓ | ✗ | 0.835 | 0.632 | 0.549 | 0.534 | 0.603 |

## 5 Conclusion

In this work, we explore the few-shot settings of protein conformation sampling by combining score-based generative models and physical MD simulations. Generative models explore the conformation space efficiently while the physical force field refine the unreasonable samples, achieving good exploration-exploitation balance. The fine-tuned diffusion sampler has shown significant improvement on conformation sampling, which exemplified the domain adaption of Str2Str with efficient short simulations. Future works can focus on the study of more effective fine-tuning strategies to enable sampling approximately from Boltzmann distribution.

### ACKNOWLEDGMENTS

This project is supported by Twitter, Intel, the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant, the Canada CIFAR AI Chair Program, Samsung Electronics Co., Ltd., Amazon Faculty Research Award, Tencent AI Lab Rhino-Bird Gift Fund, an NRC Collaborative R&D Project (AI4D-CORE-06) as well as the IVADO Fundamental Research Project grant PRF-2019-3583139727.

## REFERENCES

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

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Marloes Arts, Victor Garcia Satorras, Chin-Wei Huang, Daniel Zuegner, Marco Federici, Cecilia Clementi, Frank Noé, Robert Pinsler, and Rianne van den Berg. Two for one: Diffusion models and force fields for coarse-grained molecular dynamics. *arXiv preprint arXiv:2302.00600*, 2023.

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.

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.

Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

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.

Peter Eastman, Raimondas Galvelis, Raúl P Peláez, Charlles RA Abreu, Stephen E Farr, Emilio Gallicchio, Anton Gorenko, Michael M Henry, Frank Hu, Jing Huang, et al. Openmm 8: Molecular dynamics simulation with machine learning potentials. *The Journal of Physical Chemistry B*, 2023.

Mert Gur, Elhan Taka, Sema Zeynep Yilmaz, Ceren Kilinc, Umut Aktas, and Mert Golcuk. Conformational transition of sars-cov-2 spike glycoprotein between its closed and open states. *The Journal of chemical physics*, 153(7), 2020.

Ulrich HE Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281(1-3):140–150, 1997.

Moritz Hoffmann, Martin Scherer, Tim Hempel, Andreas Mardt, Brian de Silva, Brooke E Husic, Stefan Klus, Hao Wu, Nathan Kutz, Steven L Brunton, et al. Deeptime: a python library for machine learning dynamical models from time series data. *Machine Learning: Science and Technology*, 3(1):015009, 2021.

John B Ingraham, Max Baranov, Zak Costello, Karl W Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana M Lord, Christopher Ng-Thow-Hing, Erik R Van Vlack, et al. Illuminating protein space with a programmable generative model. *Nature*, pp. 1–9, 2023.

Giacomo Janson, Gilberto Valdes-Garcia, Lim Heo, and Michael Feig. Direct generation of protein conformational ensembles via machine learning. *Nature Communications*, 14(1):774, 2023.

Bowen Jing, Ezra Erives, Peter Pao-Huang, Gabriele Corso, Bonnie Berger, and Tommi Jaakkola. Eigenfold: Generative protein structure prediction with diffusion models. *arXiv preprint arXiv:2304.02198*, 2023.

William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

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.

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

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.

Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical programming*, 45(1-3):503–528, 1989.

Jiarui Lu, Bozitao Zhong, Zuobai Zhang, and Jian Tang. Str2str: A score-based framework for zero-shot protein conformation sampling. In *The Twelfth International Conference on Learning Representations*, 2024.

James A Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E Hauser, and Carlos Simmerling. ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, 11(8):3696–3713, 2015.

Yusuke Naritomi and Sotaro Fuchigami. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *The Journal of chemical physics*, 134(6), 2011.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

Guillermo Pérez-Hernández, Fabian Paul, Toni Giorgino, Gianni De Fabritiis, and Frank Noé. Identification of slow molecular order parameters for markov model construction. *The Journal of chemical physics*, 139(1), 2013.

Chence Shi, Chuanrui Wang, Jiarui Lu, Bozitao Zhong, and Jian Tang. Protein sequence and structure co-design with equivariant translation. *arXiv preprint arXiv:2210.08761*, 2022.

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.

Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

Robert H Swendsen and Jian-Sheng Wang. Replica monte carlo simulation of spin-glasses. *Physical review letters*, 57(21):2607, 1986.

Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

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

Bodhi P Vani, Akashnathan Aranganathan, Dedi Wang, and Pratyush Tiwary. Alphafold2-rave: From sequence to boltzmann ranking. *Journal of Chemical Theory and Computation*, 2023.

Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Kevin E Wu, Kevin K Yang, Rianne van den Berg, James Y Zou, Alex X Lu, and Ava P Amini. Protein structure generation via folding diffusion. *arXiv preprint arXiv:2209.15611*, 2022.

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.

Zhijun Zhang, Xinzijian Liu, Kangyu Yan, Mark E Tuckerman, and Jian Liu. Unified efficient thermostat scheme for the canonical ensemble with holonomic or isokinetic constraints via molecular dynamics. *The Journal of Physical Chemistry A*, 123(28):6056–6079, 2019.

Shuxin Zheng, Jiyan He, Chang Liu, Yu Shi, Ziheng Lu, Weitao Feng, Fusong Ju, Jiaxi Wang, Jianwei Zhu, Yaosen Min, et al. Towards predicting equilibrium distributions for molecular systems with deep learning. *arXiv preprint arXiv:2306.05445*, 2023.

# A EXPERIMENTAL DETAILS

## A.1 SIMULATION PROTOCOL

The short molecular dynamics (MD) simulations were respectively initiated from each sample from the diffusion sampler. The sampled protein models are firstly pre-processed by the PDBFixer of OpenMM (Eastman et al., 2023) before simulation. Hydrogen atoms are added under neutral environment pH=7.0. All simulations were run on the NVIDIA Tesla V100-SXM2-32GB GPU using the Amber ff14SB force field (Maier et al., 2015) and the TIP3P water model (Jorgensen et al., 1983) compatible with the corresponding Amber 14 force field. $Na^+$ and $Cl^-$ ions were added to achieve the electrical neutrality of the solvent. The particle mesh Ewald (PME) method (Darden et al., 1993) was used to evaluate the long-range electrostatic interaction. Each system was locally energy minimized using the L-BFGS algorithm (Liu & Nocedal, 1989) followed by equilibration in the NVT (canonical) ensemble for 1 ns. Langevin Middle Integrator (Zhang et al., 2019) with a friction coefficient of 1 $ps^{-1}$ was adopted as the thermostat (same below). Simulation temperatures for each benchmark target follows the original setting in Lindorff-Larsen et al. (2011), specifically, unit in Kelvin (K): Chignolin: 340K, Trp-cage: 290K, BBA: 325K, Villin: 360K, WW domain: 360K, NTL9: 355K, BBL: 298K, Protein B: 340K, Homeodomain: 360K, Protein G: 350K, $\alpha$3D: 370K and lambda-repressor: 350K. Then the system was equilibrated in the NPT (isothermal-isobaric) ensemble for 1 ns while coupled with the Monte Carlo barostat at constant 1 bar pressure. All production simulations were performed using OpenMM v8.0.0 (Eastman et al., 2023) in the NPT ensemble integrated with time step of 2.5 fs and trajectory frames were saved per 10 ps of simulation.

## A.2 FINE-TUNING DATASET

To acquire the fine-tuning dataset (ES) for each target, we firstly generate the seeding structure using ESMFold (Lin et al., 2023) with default configuration and ran Str2Str as described in Lu et al. (2024) to generate 100 conformations from the seeding structure. Then MD simulations was performed in parallel initialized respectively from these conformations according to the protocol above in Appendix A.1. We ignored those simulation threads failed due to possible bad initial geometries. The resulting conformations as well as their associated potential energies were saved to construct the fine-tuning dataset. We randomly split the dataset into training and validation set with ratio $0.95 : 0.05$ while the latter is used for early stopping of training.

## A.3 SETUP OF TRAINING AND INFERENCE

To optimize the network parameters of the pre-trained Str2Str, the Adam optimizer (Kingma & Ba, 2014) was used with $\mathrm{lr} = 10^{-4}$ and $\beta_1 = 0.9, \beta_2 = 0.999$. As scheduler, the learning rate was reduced in the factor of 0.1 when the loss has stopped decreasing for 10 epochs. The training process was set to be at least 100 epochs and mediated the early stopping strategy by monitoring the validation loss with 10 epochs patience. The fine-tuning time for different targets was on-average $\sim$ 2 GPU hours on the NVIDIA Tesla V100-SXM2-32GB GPU. For simplicity, we use uniform weighting in this study for the training objective, i.e. $\omega(x, t) \equiv 1$. Other model configurations and evaluation pipelines were kept the same as in Lu et al. (2024). After fine-tuning, we sampled the target protein from the prior distribution $p_T(\mathbf{x})$ with discretization of 1,000 timesteps by reversing the diffusion process via Langevin dynamics. In practice, a minimum time $t_\epsilon = 0.01$ instead of zero is used for numerical stability. The SDE mode of sampling was used for both Str2Str and Str2Str-FT.

## A.4 FAST-FOLDING TARGETS

The fast-folding targets with the reference MD trajectories are listed as follows (Lindorff-Larsen et al., 2011): Chignolin, Trp-cage, BBA, Villin, WW domain, NTL9, BBL, Protein B, Homeodomain, Protein G, $\alpha$3D and Lambda-repressor. The reference MD trajectories for evaluation were obtained by sending requests to the authors of Lindorff-Larsen et al. (2011).

## B EVALUATION METRICS

In this section, we elaborate the definition of the evaluation metrics. For validity, the Val-Clash is defined by the ratio of samples in the predicted ensemble of size 1,000 which do not contain any clash. Clash is detected by examining whether there exists pairwise distance between C$\alpha$ atoms less than a distance threshold of $\delta_c = 3.0$Å; the Val-Bond is defined similarly as the ratio of samples having no (pseudo) bond dissociation. The dissociation is counted if there is any distance between adjacent C$\alpha$ atoms that exceeds certain threshold $\delta_b$, where $\delta_b$ is defined by the maximum value of distances between adjacent C$\alpha$ in the reference MD trajectory of the target system. For distribution divergence metrics, we calculate three features including pairwise distance (PwD), time-lagged independent component (TIC) coordinates, and radius of gyration (Rg) for each conformation in the predict ensemble. To transform the continuous feature values into distribution, histograms are built with $N_{\text{bin}} = 50$ bins to represent the categorized distribution over which the JS divergence is calculated. For each channel of histogram, a pseudo-count value of $\epsilon = 10^{-6}$ was added to offset zero frequencies and slightly smooth the distribution. The pairwise distance is computed by enumerating all pairs of C$\alpha$ atoms with an offset three; the TICA dimension reduction is performed using the Deeptime library (Hoffmann et al., 2021) and the first two slowest dimensions are selected for evaluation; and the radius of gyration is defined by the root mean square distance of each C$\alpha$ atoms relative to the center of mass of the protein. All of three distributions are compared respective to the reference long MD simulation to calculate the Jensen-Shannon (JS) divergence metrics.

## C EXTENDED EXPERIMENTAL RESULTS

**Compare neighborhood exploitation with MD simulation** We demonstrate the effectiveness of the neighborhood exploitation (NE) over conventional MD simulation in Table 3. The exploration over conformation space of MD simulation (under limited budget) can be strongly enhanced by firstly sampling from a pre-trained diffusion sampler. The simulation protocol of both Str2Str-NE and MD are kept the same as described in A.1.

Table 3: Comparison between samples obtained from neighborhood exploitation (parallel short simulations) and computation-equivalent conventional MD simulation.

| Setting | Val-Clash(↑) | Val-Bond(↑) | JS-PwD(↓) | JS-TIC(↓) | JS-Rg (↓) |
|---|---|---|---|---|---|
| MD (100 ns) | **1.000** | **1.000** | 0.518 | 0.547 | 0.523 |
| Str2Str-NE | 0.990 | **1.000** | **0.294** | **0.369** | **0.331** |

**Runtime Analysis** To illustrate the efficiency of proposed few-shot sampling over conventional MD simulation, we here profile the *wall clock* running time of each type of "sampler" for each benchmark target, shown in Table 3. Note that the runtime can vary from different proteins due to the different number of atoms. The profiling results are all reported as the wall clock hours based on 4× NVIDIA Tesla V100-SXM2-32GB GPUs.

## D LIMITATIONS AND DISCUSSION

Similar to Str2Str (Lu et al., 2024), the proposed methods cannot guarantee the sampled ensemble are Boltzmann distributed (Noé et al., 2019) due to the nature of local simulation. It is not designed for replacing canonical MD simulations for the study of protein dynamics. Incorporating short MD simulations in the Str2Str sampler takes the advantages from both physical and neural sides: the neural sampler (eg., diffusion) is good at proposing data-like conformation hypothesis very efficiently in the vast conformation space while the physical simulation can equilibrate the samples onto some local minima. Both neighborhood exploitation (NE) and fine-tuning (FT) provide a simple yet effective way to refine the resulting conformations sampled from a neural sampler.

From another perspective, neighborhood exploitation coincides with the purpose of enhanced sampling methods. Enhanced sampling methods aim to overcome the energy barrier and accelerate the exploration of long MD simulations. For example, the umbrella sampling (Torrie & Valleau, 1977),

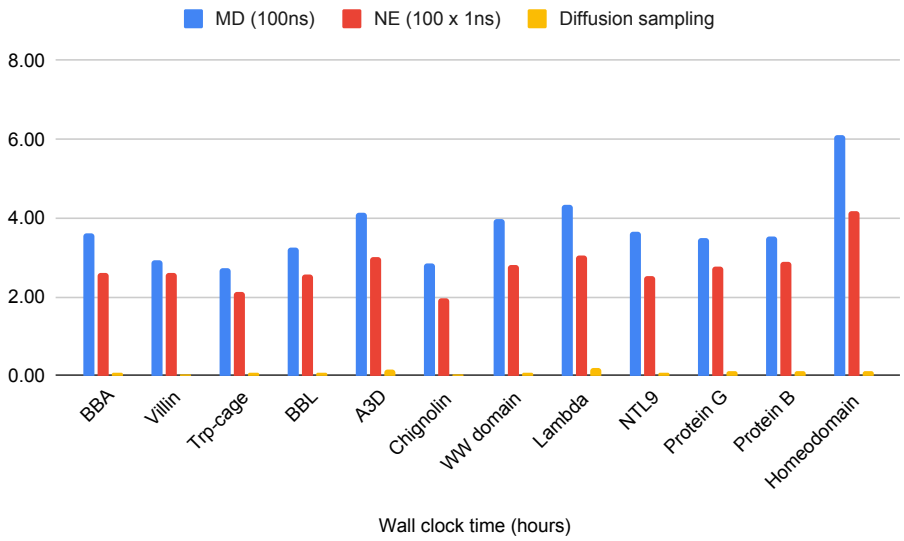



Figure 3: The runtime profile of different samplers across each fast folding target used in this study.

and parallel tempering (synonymous replica exchange molecular dynamics, REMD) (Hansmann, 1997; Sugita & Okamoto, 1999; Swendsen & Wang, 1986). The neural sampler such as Str2Str can provide diverse conformation seeds to augment the exploration of MD simulation similar to Vani et al. (2023). We find it a promising research direction to develop the neural enhanced sampling methods in the future.