F^{3}LOW: FRAME-TO-FRAME COARSE-GRAINED MOLECULAR DYNAMICS WITH SE(3) GUIDED FLOW MATCHING

Shaoning Li[1,2]*, **Yusong Wang**[3,4]*, **Mingyu Li**[5]*, **Jian Zhang**[5], **Bin Shao**[3],
Nanning Zheng[4], **Jian Tang**[1,6,7]†
[1] Mila - Québec AI Institute [2] Université de Montréal [3] Microsoft Research AI4Science
[4] National Key Laboratory of Human-Machine Hybrid Augmented Intelligence,
National Engineering Research Center for Visual Information and Applications,
and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
[5] Medicinal Chemistry and Bioinformatics Center, Shanghai Jiao Tong University School of Medicine
[6] HEC Montréal [7] Canadian Institute for Advanced Research (CIFAR)

ABSTRACT

Molecular dynamics (MD) is a crucial technique for simulating biological systems, enabling the exploration of their dynamic nature and fostering an understanding of their functions and properties. To address exploration inefficiency, emerging enhanced sampling approaches like coarse-graining (CG) and generative models have been employed. In this work, we propose a Frame-to-Frame generative model with guided Flow-matching (**F^{3}low**) for enhanced sampling, which (a) extends the domain of CG modeling to the SE(3) Riemannian manifold; (b) retreating CGMD simulations as autoregressively sampling guided by the former frame via flow-matching models; (c) targets the protein backbone, offering improved insights into secondary structure formation and intricate folding pathways. Compared to previous methods, F^{3}low allows for broader exploration of conformational space. The ability to rapidly generate diverse conformations via force-free generative paradigm on $\mathrm{SE}(3)$ paves the way toward efficient enhanced sampling methods.

1 INTRODUCTION

Molecular dynamics (MD) is a computational technique that simulates the motion and behavior of biological macromolecules systems without the need for wet lab experiments. By employing physics-based force fields and algorithms, MD provides valuable insights into the dynamic aspects of protein structures, such as protein folding event (Lindorff-Larsen et al. (2011), Fig. 1). This approach is instrumental in elucidating the functions and interactions of proteins at atomic level.

However, MD simulations faces challenges in efficiently exploring the conformation space due to the difficulty of crossing the high energy-barriers, which requires milliseconds of simulation steps and renders infeasible. To address this issue, *enhanced sampling* methods are developed to encourage the feasible exploration of conformation space via MD simulations. One of the enhanced sampling methods is *coarse-grained* molecular dynamics (CGMD) (Kmiecik et al. (2016)). It involves creating a coarse-grained representation by slicing out atoms from each residue up to its C_α, termed as *beads* shown in Fig. 1(a). In this process, atomistic forces generated by potential functions or machine learning force fields are applied to beads on the CG space to perform simulations. This facilitates the smoothing of energy landscapes, preventing the occurrence of traps in local minima. Another emerging method utilizes *probabilistic generative models* to generate conformations targeting the Boltzmann distribution, without the dependency of empirical forces and motion integration (Fig. 1(b)). Klein et al. (2023) employs normalizing flow to map structures of small peptides from Gaussian noises for the next step, highlighting transferability between states with relatively larger timesteps.

[*] Equal contribution.
[†] Corresponding author.

In this paper, we leverage the advantages of both CGMD and generative models, presenting F^{3}low, for enhanced sampling. Firstly, we extend the common modeling resolution of CGMD from C_α level to *backbone* level ($C - C_\alpha - N - O$, Fig. 1(b)). In essence, the space of conformation poses is expanded from $\mathbb{R}^{3N}$ (or $\mathrm{E}(3)^N$) to $\mathrm{SE}(3)^N$, with N denoting the number of residues. In such way, it preserves the structural geometry, enabling the direct observation of the secondary structure and folding pathway without the need for any reconstruction. Secondly, we leverage the guided flow on the SE(3) manifold (Lipman et al. (2022); Chen & Lipman (2023); Zheng et al. (2023)) to perform the simulation in a backbone-resolution. Concretely, as illustrated in Fig. 1(b), F^{3}low samples the next frame $\mathcal{C}_{s+1}$ from prior distribution (e.g., zero center of mass Gaussian) while being guided by the preceding frame $\mathcal{C}_s$ at step s. We term this sampling based on conditional probability as *frame-to-frame* simulation. This process can be repeated for S steps to conduct complete simulations. Compared with Klein et al. (2023), F^{3}low is not confined to small peptides (under 4AA) but can handle larger fast folding proteins (Lindorff-Larsen et al. (2011)) spanning from Chignolin (10AA) to Homeodomain (54AA) benefiting from the advantage of coarse-graining.

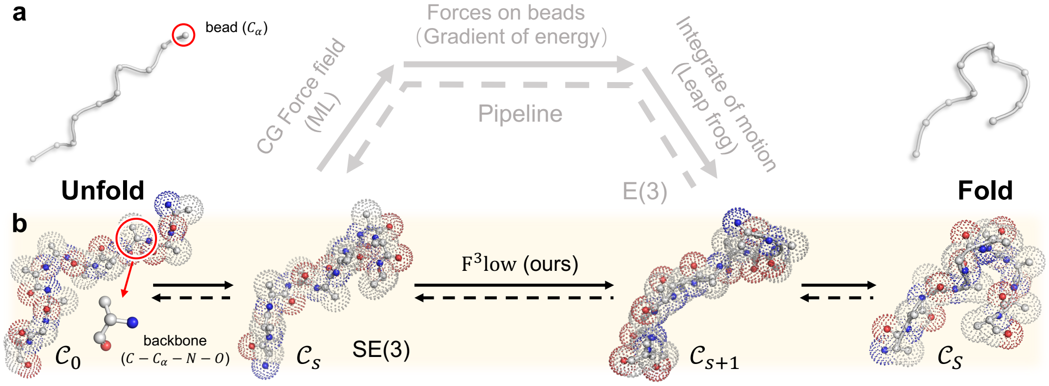

Figure 1: Overview of the enhanced sampling methods, traditional CGMD pipeline (panel **a**) and generative models (e.g., F^{3}low, panel **b**). Traditional CGMD relies on empirical forces applied to each CG bead to calculate the next frame. In contrast, generative models operate by directly sampling the next frame from a prior distribution, bypassing the need for explicit force calculations.

2 Methods

Given a conformation frame $\mathcal{C}_s$ at step s, we leverage a probabilistic generative model P_θ to sample the next frame $\mathcal{C}_{s+1}$ guided by $\mathcal{C}_s$ from a prior source distribution z:

$$\mathcal{C}_{s+1} \sim P_\theta(\mathcal{C}_{s+1}; z, \mathcal{C}_s). \tag{1}$$

This iterative process, starting from an initial frame $\mathcal{C}_0$ and repeated S times, results in the generation of a complete trajectory $\Gamma = [\mathcal{C}_0, \ldots, \mathcal{C}_S]$. We refer to this procedure based on conditional probability as *frame-to-frame* simulation illustrated in Fig. 1(b). After the simulations, we can conduct various trajectory analysis of protein dynamics, including comparing free energy surfaces and calculating transition probabilities between meta-stable states.

2.1 SE(3) guided flow matching

In this context, we provide a brief introduction to the general concept of Riemannian flow matching, followed by its extension to the SE(3) space. Then we introduce the basic concept of guided flow.

From Riemannian to SE(3) flow matching. We consider complete connected, smooth *Riemannian manifolds* $\mathcal{M}$ endowed with *metric* g. At each point $x \in \mathcal{M}$ can attach to a *Tangent space* $\mathcal{T}_x$ and the metric g is defined as inner product over $\mathcal{T}_x\mathcal{M}$. The *continuous normalizing flow* (CNF) $\psi_t : \mathcal{M} \to \mathcal{M}$ is defined as the solution of the following ordinary differential equation (ODE) along with a time-dependent vector field $u_t \in \mathcal{T}_x\mathcal{M}$: $\frac{d}{dt}\psi_t(x) = u_t(\psi_t(x))$, with initial conditions $\psi_0(x) = x$. Density distribution on Riemannian manifolds is continuous non-negative functions

$p : \mathcal{M} \rightarrow \mathbb{R}_+$ the integrate to $\int_{\mathcal{M}} p(x)dx = 1$. Given two distributions p_0 and p_1 on $\mathcal{M}$, we can define a *probability path* $p_t : t \in [0, 1]$ as their interpolation. In other word, p_t is a sequence of probability distributions generated by vector field u_t through pushing forward p_0 to p_1. To learn a CNF, *flow-matching* (FM) is proposed (Albergo & Vanden-Eijnden (2022); Lipman et al. (2022)) as a simulation-free method by regressing vector field u_t with a parametric one v_θ. Since computing u_t is intractable and cannot be directly targeted, we always adopt *conditional flow matching* (CFM), attaching with a *conditional probability path* $p_t(x|x_1)$ satisfying $p_0(x|x_1) = p_0(x)$ and $p_1(x|x_1) \approx \delta_{x_1}(x)$. $\delta_{x_1}(x)$ is the delta probability concentrating all its mass at x_1. Subsequently, the unconditional p_t can be recovered by calculating marginal distribution $p_t = \int_{\mathcal{M}} p_t(x|x_1)p_1(x_1)dx_1$. The objective of training CFM is to minimize v^t_θ with conditional vector field $u_t(x|x_1)$. When plugging the conditional flow $x_t = \psi_t(x_0|x_1)$ with initial condition $\psi_0(x_0|x_1) = x_0$ into the CFM loss, we can reparameterize $u_t(x|x_1)$ with $\frac{d}{dt}\psi_t(x)$, and rewrite the final CFM loss as:

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,p_1(x_1),p_0(x_0)} \left\| v^t_\theta(x_t) - \dot{x_t} \right\|^2_g, \tag{2}$$

where $t \sim \mathcal{U}(0, 1)$, $\dot{x_t} = \frac{d}{dt}x_t = u_t(x|x_1)$ and $\|\cdot\|^2_g$ is the norm induced by Riemannian metric g.

The basic structure of protein backbone is characterized by a rigid transformation $T = (r, x) \in \mathrm{SE}(3)$ (Jumper et al. (2021)), where this transformation can be decomposed into a rotation $r \in \mathrm{SO}(3)$ and translation vector $x \in \mathbb{R}^3$. Following Yim et al. (2023b;a), the metric g on $\mathrm{SO}(3)$ is $\langle a, a' \rangle_{\mathrm{SO}(3)} = \mathrm{Tr}(aa'^\top)/2$ where a is the vector in tangent space of $\mathrm{SO}(3)$, referring to $\mathfrak{so}(3)$, and $\mathrm{Tr}(\cdot)$ denotes the trace of the matrix. The metric g on $\mathbb{R}^3$ is $\langle y, y' \rangle_{\mathbb{R}^3} = \sum_{i=1}^3 y_i y'_i$ where $y \in \mathbb{R}^3$. Combining these two metrics, we derive the metric on $\mathrm{SE}(3)$ as $\langle (a, y), (a', y') \rangle_{\mathrm{SE}(3)} = \langle a, a' \rangle_{\mathrm{SO}(3)} + \langle y, y' \rangle_{\mathbb{R}^3}$. It allows us construct the conditional flow ψ_t connecting source distribution p_0 and target distribution p_1. Given $T_0(r_0, x_0) \sim p_0$ and $T_1(r_1, x_1) \sim p_1$, $\psi_t(T_0|T_1)$ is denoted as the geodesic interpolant path represented on $\mathrm{SO}(3)$ and $\mathbb{R}^3$ respectively:

$$\begin{aligned} r_t &= \exp_{r_0}(t \log_{r_0}(r_1)) \\ x_t &= (1 - t)x_0 + t x_1, \end{aligned} \tag{3}$$

where $\log$ is logarithm map converting elements in $\mathrm{SO}(3)$ to $\mathfrak{so}(3)$, and $\exp$ is exponential map for inverting the logarithm map. Referring to Yim et al. (2023a), the objective can be written as directly predicting the target:

$$\mathcal{L}_{\mathrm{CFM-SE(3)}}(\theta) = \mathbb{E}_{t,p_0(T_0),p_1(T_1)} \frac{1}{(1-t)^2} \left[\|\hat{x}_1 - x_1\|^2_{\mathbb{R}^3} + \left\|\log_{r_t}(\hat{r}_1) - \log_{r_t}(r_1)\right\|^2_{\mathrm{SO}(3)} \right]. \tag{4}$$

where $\hat{r}_1$ and $\hat{x}_1$ are the prediction of $v^t_\theta(T_t)$.

Guided flow. Guided flow is first introduced by Zheng et al. (2023), which is an adaptation of classifier-free guidance Ho & Salimans (2022) to FM models for conditional generation. Given the condition y, we can extend the unconditional p_t to $p_t(x|y) = \int_{\mathcal{M}} p_t(x|x_1)p_1(x_1|y)dx_1$ and corresponding vector field $u_t(x|y) = \int_{\mathcal{M}} u_t(x|x_1)\frac{p_t(x|x_1)p_1(x_1|y)}{p_t(x)}dx_1$. Following Zheng et al. (2023), the guided vector field can be derived from the combination of $p_1(x)^{1-\omega}p_1(x|y)^\omega$:

$$\tilde{u}_t(x|y) = (1 - \omega)u_t(x) + \omega u_t(x|y). \tag{5}$$

where $\omega \in \mathbb{R}$ is a manually selected weight. Zheng et al. (2023) also showcases that $u_t(x|y)$ is related to score function $\nabla \log p_t(x|y)$ by $u_t(x|y) = a_t x + b_t \nabla \log p_t(x|y)$, with a_t and b_t as schedulers [1].

2.2 F^{3}LOW FOR COARSE-GRAINED MOLECULAR DYNAMICS

In this part, we introduce our methods, F^{3}low, which provides a new perspective on coarse-grained molecular dynamics by iteratively generating conformation frames in a trajectory using a guided flow-matching model. We outline the basic concept, elaborate on integrating conformation guidance into the flow-matching models based on initial guesses, and provide an overview of the training and sampling procedures.

[1] Zheng et al. (2023) only showcases the guided flow on $\mathbb{R}^3$. We will add the proof of its generalization on $\mathrm{SO}(3)$ in the future work.

Basic concept. As defined in Eq. 1, we leverage a trained guided flow-matching model to sample the next frame. To be consistent with the aforementioned definitions, we denote $p_0 := z$, $p_1 := \mathcal{C}_{s+1}$ and $y := \mathcal{C}_s$. The conformation $\mathcal{C}$ is modeled with coarse-graining at the protein backbone level, consisting of 4 heavy atoms $C - C_\alpha - N - O$. Let N denote the number of residue within $\mathcal{C}$, the space of conformation poses is $\mathrm{SE}(3)^N$, where each residue is mapped by a rigid transformation T_i with $i \in [N]$. And hence $\mathcal{C} = [T_0, \ldots, T_N] \in \mathrm{SE}(3)^N$. To preserve the $\mathrm{SE}(3)$ symmetries, all the translations are carried out within the zero center of mass (CoM) subspace. This ensures that the prior source distribution z^x and the intermediate x_t remain $\mathrm{SE}(3)$-invariant, and the vector field u_t maintains $\mathrm{SE}(3)$-equivariance. Here we denote the prior distribution in Eq. 1 as $z = [z^r, z^x]$ where z^x represents the Gaussian distribution on $\mathbb{R}^3$ subtracted the CoM. Following Yim et al. (2023a); Watson et al. (2023), z^r serves as the isotropic Gaussian distribution $\mathcal{IG}_{\mathrm{SO}(3)}$.

It is noteworthy that we *cannot* directly define a flow from $\mathcal{C}_s$ to $\mathcal{C}_{s+1}$ since they belong to the same trajectory and, therefore, are under the same Boltzmann distribution, rather than two separate distributions. Instead, we formulate the problem as a guided flow, where we sample $\mathcal{C}_{s+1}$ guided by $\mathcal{C}_s$ from a normal prior distribution. This aligns with the definition of flow in this context.

Integration of conformation guidance. As demonstrated in Eq. 3, r_t and x_t represent the linear interpolation of z and $\mathcal{C}_{s+1}$. Then these values are utilized as the input for v_θ^t, which predicts the target $\hat{r}_1$ and $\hat{x}_1$. In the context of classifier-free guidance, incorporating the condition y (the former conformation $\mathcal{C}_s$) into the FM model poses challenges in the case of protein structures. Unlike image or text labels, representing 3D protein structures is non-trivial. Various attempts have been proposed, including methods such as template embedding (Jumper et al. (2021)) or treating the structures as *initial guesses* before recycling (Bennett et al. (2023)). Motivated by Bennett et al. (2023), we also consider $\mathcal{C}_s$ as a form of initial guess, and integrate it into z as a weighted average in space:

$$\begin{aligned} \tilde{r}_0 &= \exp_{r_0}(\gamma \log_{r_0}(z^r)) \\ \tilde{x}_0 &= (1-\gamma) z^x + \gamma x_0, \end{aligned} \tag{6}$$

where γ is a hyperparameter termed as initial guess weight. During training, it is necessary to compute the optimal transport (OT) between x_1 and z^x. And hence the derivation of $\tilde{x}_0$ can be re-written:

$$\tilde{x}_0 = (1-\gamma)\,\mathrm{OT}(z^x, x_1) + \gamma x_0. \tag{7}$$

Then we subtract the CoM of $\tilde{x}_0$ to ensure equivariance. Such linear interpolation helps better preserve the structure information while keeping the model architecture unchanged. Detailed training and sampling procedures are provided in Appendix Algorithm 1 and Algorithm 2.

3 Experiment

3.1 Free Energy Surface

To validate the capabilities of the F^{3}low model for conformational space exploration, we conducted training and sampling procedures (see details on Appendix D) on three representative fast-folding proteins: Chignolin, a β-hairpin structure; Trpcage, dominated by its dominant α-helix; and Homeodomain, a complex assembly of three helices. Further comparative analysis with the coarse-grained machine learning force field (CG-MLFF) simulations (Majewski et al. (2023)), which focus solely on the C_α atom, has demonstrated the superior performance of our approach in accurately depicting secondary structures by considering all backbone atoms.

Table 1: Minimum RMSD value with respect to the crystal structure for all simulations.

		Reference	CG-MLFF	CG-F^{3}low
Chignolin	C_α	0.15	0.17	0.14
	Backbone	0.27	0.89	0.36
Trpcage	C_α	0.45	1.03	0.56
	Backbone	0.58	2.41	0.62
Homeodomain	C_α	0.56	2.17	0.51
	Backbone	0.54	2.11	0.53

For the purpose of dimensionality reduction, we applied time-lagged independent component analysis (Pérez-Hernández et al. (2013), TICA) to both the CG-MLFF and CG-F^{3}low simulation trajectories. This analysis projected the data onto the leading three or four components, utilizing covariance matrices derived from the reference MD simulations. Additionally, we constructed Markov state models (Husic & Pande (2018), MSMs) and applied a reweighting technique to the free energy surfaces, leveraging the stationary distribution of the MSMs. Comprehensive details regarding the TICA and MSM parameters for each protein are provided in the Appendix B.

As illustrated in Fig. 2, our evaluation of two relatively simple proteins, Chignolin and Trpcage, demonstrates that both models achieve commendable coverage of the free energy surface. Notably, F^{3}low presented a more precise energy distribution in comparison to the reference data. In the case of the larger protein, Homeodomain, we observed that the CG-MLFF simulations rapidly converged to the native structure and did not sufficiently sample the unstructured conformations. In contrast, F^{3}low effectively explored these less-defined regions, achieving this with a reduced number of starting points (8 as opposed to 32). A broader exploration of the conformational space allows us to go even further in our in-depth analysis of Homeodomain, see Appendix C for more details.

Furthermore, we have superimposed the crystal structure and the AlphaFold-predicted (Jumper et al. (2021)) structure onto the free energy surface plots. The AlphaFold predictions align remarkably well with the actual crystal structure. Nevertheless, it represents a singular structure. The generative models, on the other hand, are capable of exploring a vastly expanded conformational space, thereby offering more profound insights into the dynamic behavior of proteins.

We also visualize an individual simulation trajectory for each protein in Fig. 3. The frames from diluted trajectories imply that our approach is capable of exploring a broader space of potential energy surfaces within a single trajectory, as opposed to solely depending on trajectory ensemble.

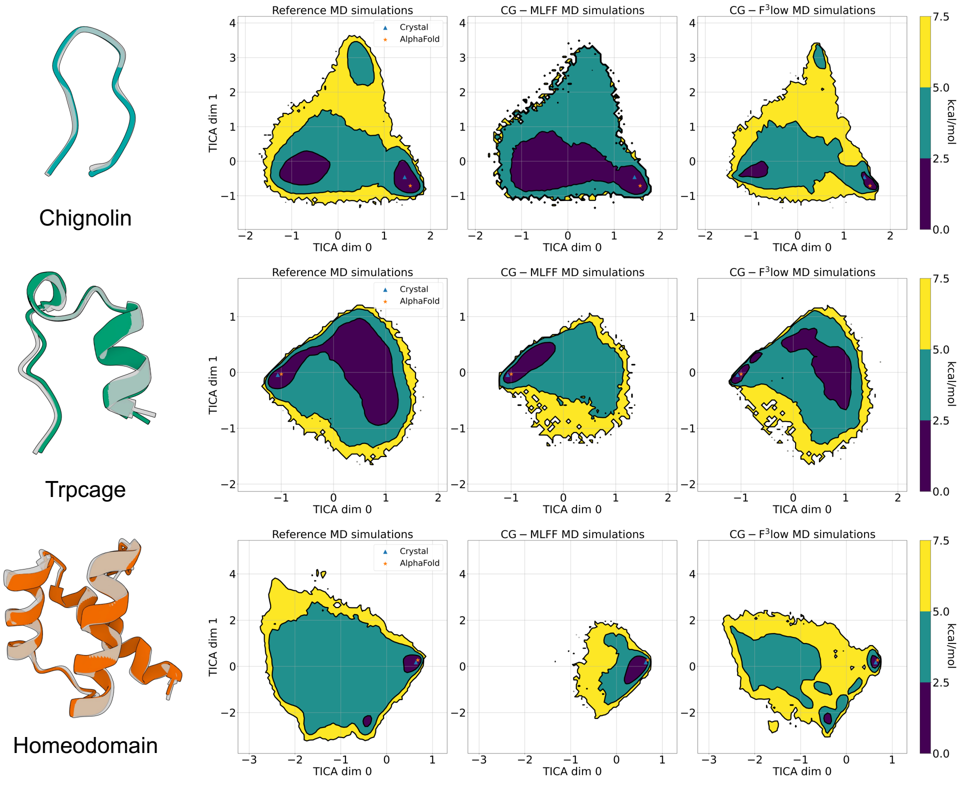

Figure 2: Comparison of free energy surface across reference, CG-MLFF and CG-F^{3}low simulations. The crystal structure (gray) and the lowest RMSD structure in CG-F^{3}low simuluation (colored) are presented on the left, respectively. The corresponding Min. RMSD can be found in Table 1.

3.2 Comparison of simulated and crystal structures.

All simulations successfully identified the location of the global minimum on the free energy surface across all proteins. Subsequently, we conducted a comparative analysis of the minimum rootmean-square deviation (RMSD) values relative to the crystal structures, as detailed in Table 1. Our CG-F^{3}low simulations consistently sampled structures with lower RMSD values than CG-MLFF simulations and showed comparable RMSD to the reference simulations. The efficacy of our algorithm is particularly pronounced when comparing in terms of the backbone. In contrast, CG-MLFF simulations require additional steps to reconstruct the backbone, highlighting the inherent advantages of our approach.

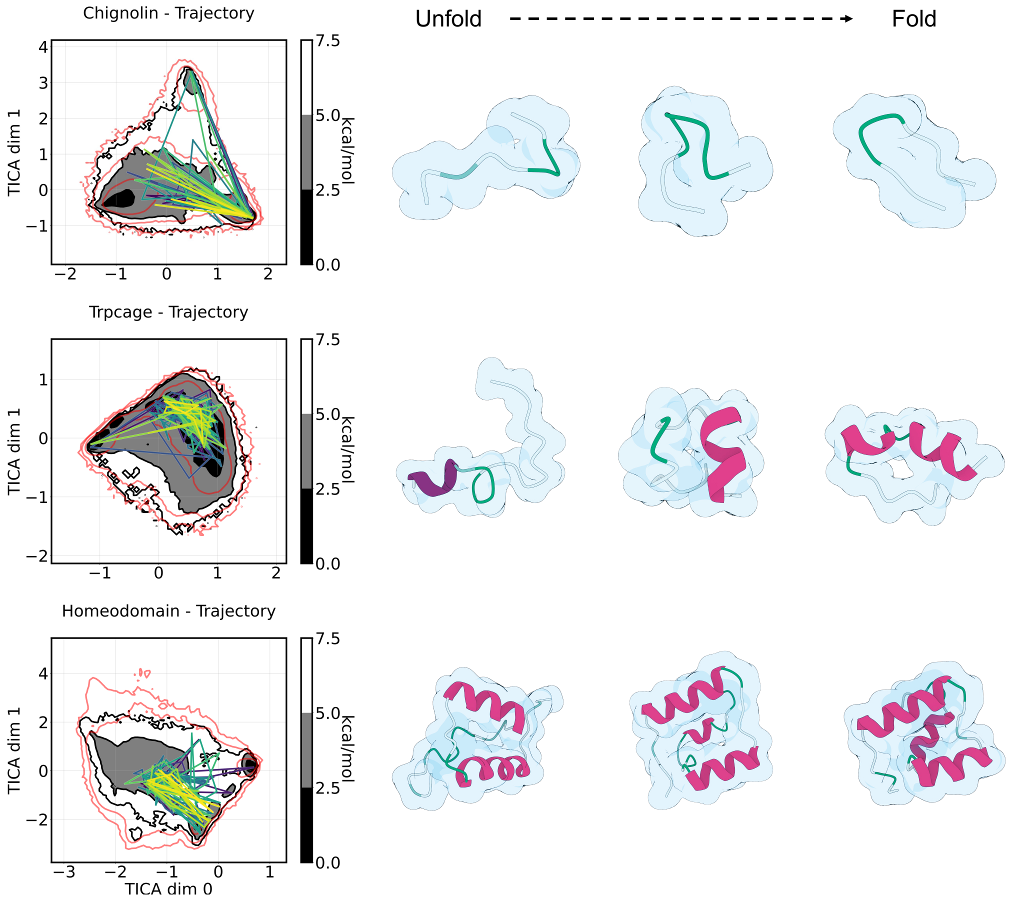

Figure 3: Individual trajectory visualization for Chignolin, Trpcage and Homeodomain. In each visual representation, the simulation trajectory traverses the free energy landscape, depicted in a gradient from purple to yellow.

4 Conclusion

In this work, we present a novel enhanced sampling method named F^{3}low, which is a frame-to-frame generative model with guided flow matching on $\mathrm{SE}(3)$. F^{3}low extends the modeling domain to $\mathrm{SE}(3)$ with a backbone level-resolution, providing improved insights into secondary structure information and intricate folding pathways. By analyzing the simulation trajectories of 3 representative proteins and comparing with traditional MLFF, we illustrate the F^{3}low's capacity for broadly exploring conformational spaces within a generative paradigm. In our future work, we intend to incorporate the physical analysis of the simulation trajectories, and conduct the remaining fast folding proteins. Additionally, the flow process on side-chain torsion angles would be included to perform all-atom protein molecule dynamics.

ACKNOWLEDGEMENT

We thank the reviewers of GEM workshop for their valuable comments. We thank Adrià Pérez, the author of Majewski et al. (2023) for providing the detailed settings of free energy surface. We thank Le Zhuo for valuable discussions of flow matching methods.

This paper is supported by National Key Research and Development Program of China (2023YFF1205103), National Natural Science Foundation of China (81925034) and National Natural Science Foundation of China (62088102). This paper is also supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant, the Canada CIFAR AI Chair Program, collaboration grants between Microsoft Research and Mila, Samsung Electronics Co., Ltd., Amazon Faculty Research Award, Tencent AI Lab Rhino-Bird Gift Fund, a NRC Collaborative R&D Project (AI4D-CORE-06) as well as the IVADO Fundamental Research Project grant PRF-2019-3583139727.

REFERENCES

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

Nathaniel R Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, et al. Improving de novo protein binder design with deep learning. *Nature Communications*, 14(1):2625, 2023.

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

Mari L DeMarco, Darwin OV Alonso, and Valerie Daggett. Diffusing and colliding: the atomic level folding/unfolding pathway of a small helical protein. *Journal of molecular biology*, 341(4): 1109–1124, 2004.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Brooke E Husic and Vijay S Pande. Markov state models: From an art to a science. *Journal of the American Chemical Society*, 140(7):2386–2396, 2018.

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.

Leon Klein, Andrew YK Foong, Tor Erlend Fjelde, Bruno Mlodozeniec, Marc Brockschmidt, Sebastian Nowozin, Frank Noé, and Ryota Tomioka. Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics. *arXiv preprint arXiv:2302.01170*, 2023.

Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chemical reviews*, 116(14):7898–7936, 2016.

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

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

Maciej Majewski, Adrià Pérez, Philipp Thölke, Stefan Doerr, Nicholas E Charron, Toni Giorgino, Brooke E Husic, Cecilia Clementi, Frank Noé, and Gianni De Fabritiis. Machine learning coarse-grained potentials of protein thermodynamics. *Nature Communications*, 14(1):5739, 2023.

Ugo Mayor, Christopher M Johnson, Valerie Daggett, and Alan R Fersht. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proceedings of the National Academy of Sciences*, 97(25):13518–13522, 2000.

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.

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.

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

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

Qinqing Zheng, Matt Le, Neta Shaul, Yaron Lipman, Aditya Grover, and Ricky TQ Chen. Guided flows for generative modeling and decision making. *arXiv preprint arXiv:2311.13443*, 2023.

A Training and Sampling

We describe the precise training and simulation process in Algorithm 1 and 2 over distribution in $\mathrm{SE}(3)^N$. Different from Yim et al. (2023b), *frame* in this paper is the snapshot within a trajectory, rather than the spatial structure of the protein backbone. In the training process line 7, when combining the initial guess, an alignment for x_0 to x_1 is performed before adding it to the transported z^x. This step is implemented to maintain training stability. In the simulation process, the notation $\mathrm{ODEStep}$ (line 10) refers to the Euler method.

Algorithm 1: F^{3}low training on $\mathrm{SE}(3)$

```
Input  : Trajectory dataset $\mathcal{C}_\Gamma$, flow network $v_\theta$, initial guess weight $\gamma$, probability of
         unconditional training $p_{\mathrm{uncond}}$
Output : Trained F$^3$low $v_\theta$
1  while Training do
2      $t \sim \mathcal{U}(0,1)$;
3      $\mathcal{C}_s, \mathcal{C}_{s+1} \sim \mathcal{C}_\Gamma$;
4      $z^r, z^x \sim \mathcal{IG}_{\mathrm{SO(3)}},\ \mathcal{N}(0,1)$;
5      $z^x \leftarrow z^x - \mathrm{CoM}(z^x)$;
6      if Uniform$(0, 1.0) > p_{\mathrm{uncond}}$ then
7          $\tilde{r}_0, \tilde{x}_0 \leftarrow \exp_{r_0}(\gamma \log_{r_0}(z^r)),\ (1-\gamma)\,\mathrm{OT}(z^x, x_1) + \gamma x_0$;
8          $\tilde{x}_0 \leftarrow \tilde{x}_0 - \mathrm{CoM}(\tilde{x}_0)$;
9      else
10         $\tilde{x}_0 \leftarrow \mathrm{OT}(z^x, x_1)$;
11     $r_t, x_t \leftarrow \exp_{r_0}(t \log_{r_0}(r_1)),\ (1-t)\tilde{x}_0 + t x_1$;
12     $\hat{r}_1, \hat{x}_1 \leftarrow v_\theta(r_t, x_t | r_0, x_0)$;
13     $\mathcal{L}_{\mathrm{CFM-SE(3)}} \leftarrow \frac{1}{(1-t)^2}\left[\left\|\hat{x}_1 - x_1\right\|^2_{\mathbb{R}^3} + \left\|\log_{r_t}(\hat{r}_1) - \log_{r_t}(r_1)\right\|^2_{\mathrm{SO(3)}}\right]$;
14     $\theta \leftarrow \mathrm{Update}(\theta, \nabla_\theta \mathcal{L}_{\mathrm{CFM-SE(3)}})$;
15 return $v_\theta$
```

Algorithm 2: Coarse-grained frame-to-frame simulation on $\mathrm{SE}(3)$ via F^{3}low

```
Input  : Trained F$^3$low $v_\theta$, initial start frame $\mathcal{C}_0$, guidance parameter $\omega$, initial guess weight $\gamma$,
         number of ODE steps $n_{\mathrm{ode}}$, number of simulation steps $S$
Output : Trajectory $\mathcal{C}_\Gamma^S$
1  $\mathcal{C}_{\mathrm{cur}} \leftarrow \mathcal{C}_0$ ;
2  $\mathcal{C}_\Gamma^S \leftarrow List[\mathcal{C}_0]$;
3  $h \leftarrow \frac{1}{n_{\mathrm{ode}}}$;
4  for $s$ in $[1, \ldots, S]$ do
5      $r_0, x_0 \leftarrow \mathcal{C}_{\mathrm{cur}}$;
6      $\tilde{x}_0 \leftarrow (1-\gamma) z^x + \gamma x_0$;
7      $\tilde{r}_0 \leftarrow \exp_{r_0}(\gamma \log_{r_0}(z^r))$;
8      for $t = 0, h, \ldots, 1-h$ do
9          $\tilde{v}_\theta(r_t, x_t) \leftarrow (1-\omega) v_\theta(r_t, x_t) + \omega v_\theta(r_t, x_t | \tilde{r}_0, \tilde{x}_0)$;
10         $x_{t+h} \leftarrow$ ODEStep $(\tilde{v}_\theta(r_t, x_t), x_t)$;
11     $\mathcal{C}_s \leftarrow r_1, x_1$;
12     Append$(\mathcal{C}_\Gamma^S, \mathcal{C}_s)$;
13 return $\mathcal{C}_\Gamma^S$
```

B Markov state model estimation

Markov state models (MSMs) were constructed for the reference MD simulations, as well as for the CG-MLFF and CG-F^{3}low simulations. For the reference simulations, in line with the methodologies outlined in previous research Majewski et al. (2023), simulation data were transformed into pairwise C_α distances (excluding the terminal five residues of Trpcage). TICA was then employed to reduce the dimensionality of the featurized data, capturing the first 4, 3, and 3 principal components for Chignolin, Trpcage, and Homeodomain, respectively. The resulting components were clustered using the K-means algorithm, and this discretized data was utilized for MSM estimation. The CG-MLFF and CG-F^{3}low simulations underwent an identical process. However, for these simulations, covariance matrices derived from the reference MD simulations were used during the TICA projection to ensure a consistent basis for comparison of the first 3 or 4 components. This step was crucial for assessing how accurately the coarse-grained simulations reproduce the free energy surfaces of proteins. To enhance the interpretability of the MSMs, the PCCA algorithm was applied to consolidate the microstates into macrostates. The comparative free energy surfaces were generated by partitioning the first two TICA components into an 80×80 grid and adjusting them through reweighting using the MSM stationary distribution. The detailed parameters are listed in the table below.

Table 2: Parameters of the TICA and MSM analysis for reference, CG-MLFF and CG-F^{3}low simulations.

		TICA lag time (steps)	# of TICA projection components	# of K-means clusters	MSM lag time (ns)	# of MSM macrostates
Reference	Chignolin	20	4	1200	20	3
	Trpcage	100	3	500	10	2
	Homeodomain	100	3	600	10	3
CG-MLFF	Chignolin	/	3	200	0.001	3
	Trpcage	/	3	200	0.01	2
	Homeodomain	/	3	200	0.01	3
CG-F^{3}low	Chignolin	/	4	1200	20	3
	Trpcage	/	3	500	10	2
	Homeodomain	/	3	600	10	3

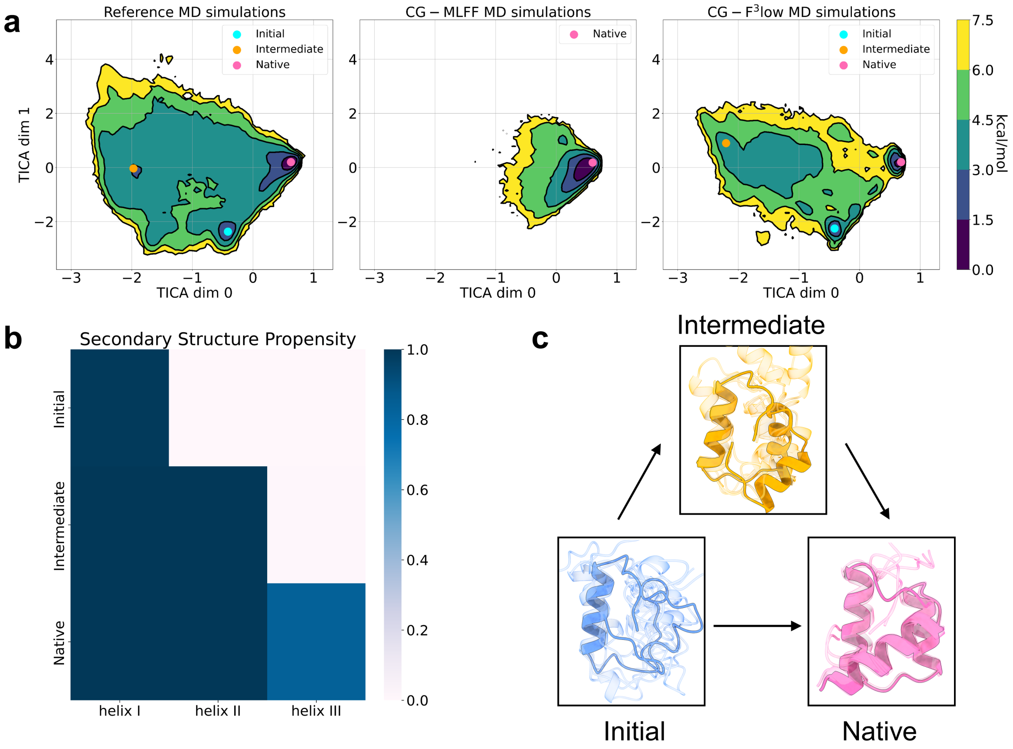

Figure 4: Simulation analysis of Homeodomain. **a.** The free energy surface of Homeodomain mapped onto the first two principal components derived from tICA for the reference all-atom MD simulations (left), the CG-MLFF simulations (center) and the CG-F^{3}low simulations (right). Distinct macrostates on the landscape are highlighted by circles in different colors: cyan for the initial state, orange for the intermediate state, and pink for the native state. **b.** The propensity of the three secondary structural elements of the Homeodomain across the macrostates, was measured by the percentage of conformational ensembles with an RMSD threshold of 2Å for each helix. **c.** The representative conformations from the macrostates identified in the CG-F^{3}low simulations correspond to the free energy minima indicated by the same color coding. Transparent structures reveal additional diverse conformations from the same state. Arrows represent the main pathways transitioning from the initial unstructured state to the native folded state.

C Structural analysis of Homeodoamin simulations

To demonstrate the capabilities of the F^{3}low in exploring conformational space, we took the Homeodomain as a case study. The conformational landscape explored by the CG-F^{3}low exhibited a similar feature as the reference all-atom MD simulations, successfully identifying three metastable states. On the contrary, the CG-MLFF was confined to a single metastable state (Figure 4(a)). Furthermore, the Homeodomain (54 AA in total) is structured into a three-helix bundle, specifically helix I (8-20 AA), helix II (26-36 AA), and helix III (40-53 AA). Notably, helices I and III together form a Helix-Turn-Helix (HTH) motif, a signature motif of DNA-binding proteins and known for its fast-folding properties. Experiment results showed that the HTH motif is more prone to instability than helix I, and the unfolding pathway preferentially disrupts the HTH motif before helix I (Mayor et al. (2000)). Our simulations corroborate these observations, further offering insights into its folding pathway (Figure 4(b) and (c)). From the representative conformational ensembles, it is inferred that helix I appears in all states, even in the unstructured initial microstate. Subsequently, it could transition to an intermediate state, wherein the secondary structures of helix I and helix II are nearly formed, and the tertiary interactions are partially established. Such findings support the hypothesis that the Homeodomain employs a diffusion-collision mechanism for folding. In the mechanism, the secondary structural elements first form independently and later assemble to the tertiary structure, aligning with prior studies of Homeodomain (Mayor et al. (2000); DeMarco et al. (2004)).

D Experiment details

D.1 Trajectory Dataset

We adopt the large-scale all-atom molecular dynamics dataset in Majewski et al. (2023), which contains twelve fast-folding proteins studied by Lindorff-Larsen et al. (2011). These proteins exhibit diverse secondary structural elements, including α-helices and β-strands, with lengths spanning from 10 to 80 amino acids. The evaluation of the proposed methods is conducted on three specific protein trajectories selected from the dataset, which are Chignolin (10AA), Trpcage (20AA), and Homeodomain (54AA). Other settings are shown in Table 3.

Table 3: All-atom MD simulation trajectory dataset from Majewski et al. (2023).

Protein	Sequence length (#AA)	Aggregated time (μs)	Min. C_α RMSD (Å)
Chignolin	10	186	0.15
Trpcage	20	195	0.45
Homeodomain	54	198	0.56

D.2 Training details

We employ the FramePred model architecture from Yim et al. (2023b), which stacks multiple Invariant Point Attention layers introduced by AlphaFold2 Jumper et al. (2021), which encodes the structural features and decodes the rigid postures in an equivariant way. Additionally, we augment the model with sequence embedding as single features for input. All other model hyperparameters follow Yim et al. (2023a). For guided flow, we set initial guess weight γ to 0.5, guidance weight ω to 1 and unconditional probability p_{uncond} to 0, in order to conduct a fully guided frame-to-frame simulation. We would explore the effect of γ and p_{uncond} in the future. The training configuration includes a maximum of 100 training epoch, a learning rate of 0.0001. For Chignolin and Trpcage, a batch size of 512 is employed, while for Homeodomain, a batch size of 128 is utilized. All training experiments are performed individually on a single NVIDIA GeForce RTX 4090 GPU.

D.3 Simulation details

After training, we executed 8 parallel coarse-grained simulations (sampling procedures) with 150,000 samples per protein. These simulations were initiated from conformational states that were randomly selected from the reference free energy surface within the test set. To remove the effect of the starting conformations, we removed the first 10% of sample points per trajectory when plotting the free energy surfaces. All simulation experiments are performed individually on a single NVIDIA GeForce RTX 4090 GPU. The simulation time per step is reported in Table 4.

	Chignolin	Trpcage	Homeodomain
Simulation time per step (s)	0.32 ± 0.01	0.34 ± 0.01	0.38 ± 0.01

Table 4: Simulation time per step for each protein.