
Learning residue level protein dynamics with multiscale Gaussians

Mihir Bafna¹ Bowen Jing¹ Bonnie Berger^{1,2}

¹ CSAIL, Massachusetts Institute of Technology

² Dept. of Mathematics, Massachusetts Institute of Technology
{mihirb14, bjing, bab}@mit.edu

Abstract

Understanding the structure dynamics of proteins is essential for elucidating biological function. We present DYNAPROT, a lightweight, SE(3)-invariant framework that predicts rich descriptors of protein dynamics directly from static structures. By casting the problem through the lens of multivariate Gaussians, DYNAPROT estimates dynamics at two complementary scales: (1) per-residue marginal anisotropy as 3×3 covariance matrices capturing local flexibility, and (2) joint scalar covariances encoding pairwise dynamic coupling across residues. DYNAPROT achieves high accuracy in predicting residue flexibility, marginal anisotropy, and, remarkably, enables reasonable reconstruction of the full covariance matrix for fast ensemble generation using orders of magnitude fewer parameters than prior methods.

1 Introduction

Proteins rarely exist in static conformations. Due to interactions with ligands, other biomolecules, and external factors such as temperature and pH, protein structures continuously fluctuate. Many enzymes rely on loop motions or domain rearrangements to form catalytically active sites (Zinovjev et al., 2024), allostery often involves shifting backbone or side chain conformations propagating signals over long distances (Yu and Koshland Jr, 2001), and membrane proteins such as GPCRs switch between inactive and active conformational states (Zhang et al., 2024). Capturing this behavior computationally has long been the domain of molecular dynamics (MD); however, the enormous computational cost of MD limits its usability, especially for proteome-wide applications. Recent work has explored the use of deep learning to approximate and accelerate this process (Jing et al., 2025), but all of these approaches necessitate large scale PDB pretraining and focus on generating *ensembles* of structures, requiring multiple stochastic forward passes. On the other hand, compact and interpretable representations of dynamics, akin to Normal Mode Analysis (Cui and Bahar, 2005; Skjaerven et al., 2009) often suffice for many practical settings. This motivates the development of models that can *explicitly* predict such dynamics descriptors in a supervised manner.

We introduce DYNAPROT, a lightweight, interpretable, and expressive framework for predicting protein dynamics through the lens of Gaussian distributions over structure. Specifically, DYNAPROT predicts: (1) per-residue marginal anisotropy as 3×3 covariance matrices capturing local dynamics, and (2) joint scalar $N \times N$ covariances encoding pairwise dynamic coupling across residues. Remarkably, while DYNAPROT was not explicitly trained to directly model the full $3N \times 3N$ joint distribution, we find that its marginal and pairwise outputs can be composed into a reasonable approximation, enabling extremely fast ensemble generation in \mathbb{R}^{3N} . DYNAPROT is trained on only $\sim 1,000$ MD-derived proteins and improves upon Normal Mode Analysis (NMA) in both predictive accuracy and efficiency, while remaining dramatically smaller and faster than existing ensemble generation approaches. To our knowledge, DYNAPROT is the first model to *explicitly* learn both marginal and pairwise Gaussian representations of protein dynamics, and the first to predict the full $3N \times 3N$ covariance structure—akin to NMA—in a data-driven, learnable fashion.

2 Method

We propose a perspective for modeling protein dynamics through distributions over atomic coordinates, relying on tractable approximations such as Gaussians. Formally, we model a protein structure with N residues as a random variable $\mathbf{X} \in \mathbb{R}^{3N}$, where each residue contributes the three-dimensional Cartesian coordinates of its C_α atom. The distribution over conformational states is then represented as a multivariate normal distribution: $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\text{joint}})$, $\boldsymbol{\mu} \in \mathbb{R}^{3N}$, $\boldsymbol{\Sigma}_{\text{joint}} \in \mathbb{R}^{3N \times 3N}$. Here, $\boldsymbol{\mu}$ corresponds to the average (or equilibrium) structure—typically the minimum energy conformation—and $\boldsymbol{\Sigma}_{\text{joint}}$ captures the full covariance across all C_α positions, encoding both local fluctuations and long-range correlated motions. The Gaussian formulation provides a principled way to decompose protein dynamics across different levels of granularity (Figure 1 and Table 3), depending on the modeling objective. DYNAPROT focuses on levels 2 and 3 of this hierarchy—explicitly predicting both 3×3 marginal Gaussians per-residue and a $N \times N$ matrix of residue-residue couplings. From the 3×3 marginals, we can easily derive RMSF (level 1, A.1.1). Interestingly, utilizing both the marginals and the pairwise coupling, we can retrieve a reasonable approximation of the full joint $3N \times 3N$ (level 4). This design strikes a balance between local interpretability and global coordination, while avoiding the intractability of directly learning the full joint covariance. See A.1.1 and Table 3 for more.

DYNAPROT (Figure 2) consists of two models, each taking as input a protein structure but designed to *explicitly* capture different granularities of protein dynamics: (i) marginal Gaussian blobs per residue, and (ii) pairwise covariance across residues. DYNAPROT is given the input structure as a set of local C_α residue frames. Additionally, an initial embedding layer is included to encode the amino acid sequence $\mathbf{s} \in \mathbb{R}^{N \times D}$. Both models share a common architectural backbone composed of eight Invariant Point Attention (IPA) blocks (Jumper et al., 2021).

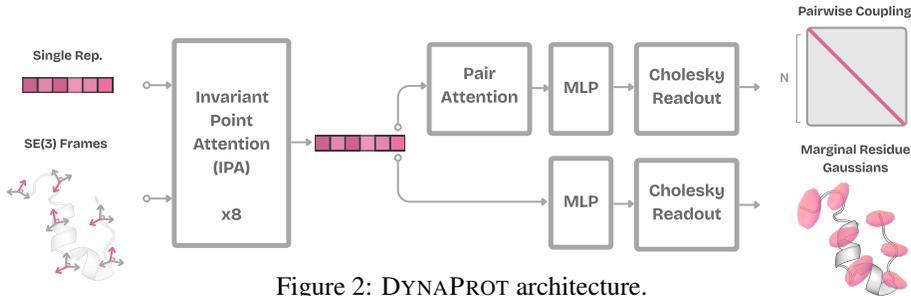


Figure 2: DYNAPROT architecture.

Learning marginal Gaussians After the input sequence representations and residue frames are processed through the IPA backbone, a simple MLP readout is used for marginal prediction of $\boldsymbol{\Sigma}_{\text{marginal}}^{(i)} \in \mathbb{R}^{3 \times 3}$, modeling the local position (xyz) covariance of residue i . These outputs are trained to match empirical marginal distributions derived from the MD data. Note that the mean of each Gaussian is not learned. Instead, we take the input structure’s C_α coordinate $\mathbf{t}_i \in \mathbb{R}^3$ as the fixed mean $\boldsymbol{\mu}_i$ of the distribution: $\boldsymbol{\mu}_i := \mathbf{t}_i$, $\mathbf{X}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{\text{marginal}}^{(i)})$.

Because covariance matrices are required to be symmetric and positive definite (SPD), predicting all 9 elements of a 3×3 matrix would be overparameterized. We instead enforce SPD constraints directly by parameterizing the covariance via its Cholesky factor. Specifically, the model predicts the entries $\{a_j\}_{j=1}^6$ of a lower triangular matrix $\mathbf{L}_i \in \mathbb{R}^{3 \times 3}$, enforces positivity along the diagonal with the Softplus activation function (Glorot et al., 2011), and recovers the covariance:

$$\boldsymbol{\Sigma}_{\text{marginal}}^{(i)} = \mathbf{L}_i \mathbf{L}_i^\top, \quad \text{where } \mathbf{L}_i = \begin{bmatrix} \text{softplus}(a_1) & 0 & 0 \\ a_2 & \text{softplus}(a_3) & 0 \\ a_4 & a_5 & \text{softplus}(a_6) \end{bmatrix} \quad (1)$$

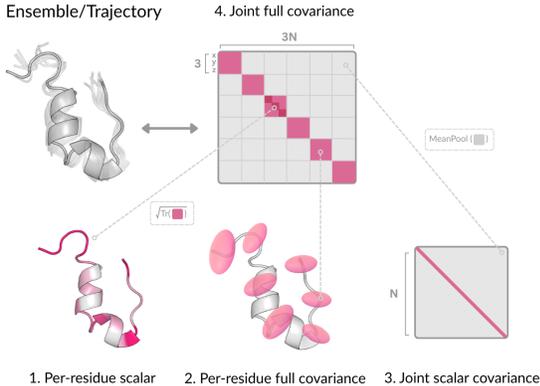


Figure 1: Gaussian view of protein dynamics

This factorization ensures that the predicted covariance matrix is SPD by construction. Since SPD matrices lie on a Riemannian manifold with non-Euclidean geometry, using loss functions that respect this structure is critical for meaningful comparison. Standard Euclidean distances (e.g., MSE or Frobenius norm) ignore the curvature of this space and can lead to unstable or distorted gradients (see ablations in Appendix A.4.2). We instead employ the log-Euclidean (or log-Frobenius) distance (Vemulapalli and Jacobs, 2015; Huang et al., 2015) that reflects the intrinsic geometry of the SPD manifold: $\mathcal{L}_{\text{LogFrob}} = \|\log(\Sigma_{\text{pred}}) - \log(\Sigma_{\text{true}})\|_F^2$ where $\log(\Sigma) = Q \log(\Lambda) Q^T$.

Learning pairwise couplings Using the output representations h from the IPA backbone, the pairwise dynamics module produces a scalar-valued $N \times N$ covariance matrix C , where each entry C_{ij} captures the dynamical coupling between residue pairs. These scalar couplings trained to reproduce MD-derived pairwise fluctuations. To predict the global pairwise covariance structure, we construct pairwise features by concatenating their residue-level embeddings and pass these features through a stack of AlphaFold-style pairwise attention blocks based on the Evoformer architecture (Jumper et al., 2021), which include triangle updates and residue-wise message passing, followed by a scalar readout head. Following the same procedure as the previous section, we enforce SPD constraints on this covariance matrix by populating the lower-triangle entries of $L \in \mathbb{R}^{N \times N}$ with the values of z_{ij} and applying the Softplus activation when $i = j$. Finally, the pairwise covariance matrix is reconstructed via Cholesky composition $C = LL^T$ and the same loss is used for optimization.

Joint reconstruction heuristic. Given a predicted scalar coupling $C \in \mathbb{R}^{N \times N}$ and a set of marginal covariances $\{\Sigma_{\text{marginal}}^{(i)} \in \mathbb{R}^{3 \times 3}\}_{i=1}^N$, we propose a heuristic to reconstruct an approximate full joint covariance matrix $\Sigma_{\text{joint}} \in \mathbb{R}^{3N \times 3N}$. Each marginal covariance $\Sigma_{\text{marginal}}^{(i)}$ is SPD by construction, and thus admits a Cholesky factorization $\Sigma_{\text{marginal}}^{(i)} = L_i L_i^T$, where $L_i \in \mathbb{R}^{3 \times 3}$. We then define a block-diagonal matrix $L_{\text{marginal}} \in \mathbb{R}^{3N \times 3N}$ as $L_{\text{marginal}} = \bigoplus_{i=1}^N L_i$. By construction, L_{marginal} is lower triangular with positive diagonal entries, since each L_i satisfies these properties. Drawing from the univariate identity $\text{Cov}(i, j) = \text{Corr}(i, j) \cdot \sigma_i \sigma_j$, we define the multivariate cross-covariance block between residues i and j as $\Sigma_{\text{joint}}^{(i, j)} = L_i \tilde{C}_{ij} L_j^T$. Here, the Cholesky factor L_i serves as a matrix square root of the covariance $\Sigma_{\text{marginal}}^{(i)}$, analogous to standard deviation in the univariate case. And, \tilde{C} is a correlation matrix found by standardizing C . Using the Kronecker product, we can denote this heuristic cleanly as $\Sigma_{\text{joint}} = L_{\text{marginal}} (\tilde{C} \otimes I_3) L_{\text{marginal}}^T$. This can be shown to be SPD and thus a valid covariance matrix (Appendix A.1.2). We find this approximation reconstructs the joint covariance to a reasonable degree and reduces ensemble sampling to sampling from a Gaussian, which is extremely fast. See Appendix A.1.3 for more.

3 Experiments

We train DYNAPROT on the ATLAS dataset (Vander Meersche et al., 2024) in the same manner as (Jing et al., 2024). Refer to Appendix A.3 for more information regarding training and baselines.

Residue-level flexibility Since DYNAPROT-M is trained to predict marginal Gaussians per residue, it inherently captures residue-level flexibility, as RMSF is defined as the square root of the trace of each marginal covariance (see Section A.1.1). To evaluate DYNAPROT-M’s ability to recover this, we compare against what is, to our knowledge, the only method that *explicitly* predicts residue flexibility from structure: FLEXPERT-3D. For fair comparison, we train and evaluate DYNAPROT-M under the same ATLAS train/val/test split defined in Kouba et al. (2024). Despite solving the more challenging task of predicting marginal anisotropy rather than scalar fluctuations alone, DYNAPROT-M achieves a substantially higher Pearson correlation with MD-derived RMSF (median $r = 0.865$, 75th percentile $r = 0.930$) than FLEXPERT-3D (Table 1), while using three orders of magnitude fewer parameters (955K vs. 1.2B) and without NMA as input. This allows DYNAPROT-M to generalize better while being more parameter efficient. See Appendix A.5 for DYNAPROT-M additional RMSF plots.

Table 1: RMSF Pearson correlation (r) against ATLAS MD-derived RMSF (FlexPert test split). Median and 75th percentile reported.

Method	RMSF r (\uparrow)	# Params
DYNAPROT-M	0.865 / 0.930	955 K
FLEXPERT-3D	0.830 / 0.899	1.2 B
NMA (ANM)	0.697 / 0.784	–

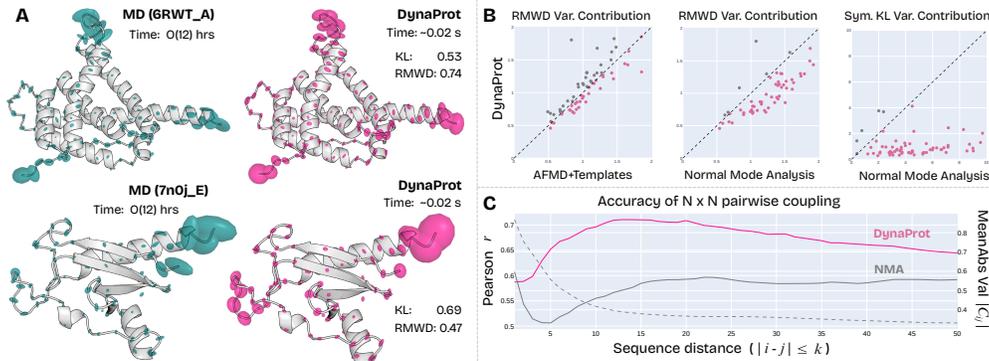


Figure 3: **DYNAPROT marginal Gaussian and residue coupling analysis.** **A.** Renderings of predicted marginal Gaussians compared to ATLAS MD constructed Gaussians (mean symmetric KL divergence and RMWD are reported). **B.** Joint distribution (within 75th percentile) of DYNAPROT performance vs. (AFMD+T, NMA). **C.** Band-wise Pearson correlation between predicted and ground-truth residue–residue coupling matrices as a function of sequence distance.

Residue-level Gaussians To assess the faithfulness of DYNAPROT-M’s predicted marginals, we compare against both physics-based (NMA) and learned ensemble methods (AFMD+T). To quantify the accuracy, we compute the variance contribution of the *symmetric KL divergence* (see Appendix A.2) and the *root mean 2-Wasserstein distance* (RMWD) as described in Jing et al. (2024), compared to the ground truth marginal Gaussians computed from the ATLAS test set (AFMD split).

Despite being orders of magnitude faster and smaller (955k vs. 95M parameters), DYNAPROT-M achieves competitive accuracy. DYNAPROT-M attains a median RMWD of 1.18 and symmetric KL divergence of 0.91, both substantially better than NMA (1.45 and 4.56, respectively), and comparable to AFMD+TEMPLATES’s 1.10 and 0.60.

Table 2: Comparison of methods on anisotropic blob prediction (ATLAS test split). Runtime for a length 271 protein (71ao_A). 25th %ile / Median reported (\downarrow is better).

Method	RMWD Var	Sym. KL Var	# Params	Time
DYNAPROT-M	0.84 / 1.18	0.53 / 0.91	955 K	~0.02 s
AFMD+T	0.87 / 1.10	0.37 / 0.60	95 M	~7000 s
NMA (ANM)	1.14 / 1.45	3.03 / 4.56	–	~5.37 s

Predicting scalar coupling To evaluate DYNAPROT-J in modeling residue–residue couplings, we compare its predicted $N \times N$ scalar covariance matrices against those derived from NMA. We obtain pairwise correlation matrices by normalizing the entries to be unit diagonal and constrained to the range $[-1, 1]$. We observe that in the ground truth $N \times N$ correlations, magnitude of entries diminish rapidly with distance from main diagonal, indicating weak long-range coupling. This is indicated by the dotted gray line in Figure 3C, which shows the mean absolute value of the entries from the principal diagonal up to the k th diagonal band. To focus on meaningful and prominent interactions, we define a *diagonal band* of width $k = 50$ residues ($|i - j| \leq 50$), which captures local and medium-range interactions. This essentially measures residue-residue coupling as a function of *sequence distance* (how distal are i and j along the backbone). We compute the Pearson correlation between predicted and ground-truth residue–residue correlation matrices for the entries along each diagonal band k , by iteratively extracting the upper-triangular entries satisfying $|i - j| \leq k$ for $k = 1$ to 50. This is repeated for each of the 82 test set proteins and the median Pearson correlation is reported (each point in Figure 3C). This band-wise analysis enables us to compare the accuracy of coupling signals at increasing residue distances, and we find that DYNAPROT-J (peak correlation of $r = 0.71$) strongly outperforms NMA (peak correlation of $r = 0.59$) particularly at short to mid-range coupling distances, where the coupling is the strongest.

Ensemble generation As described in Section 2, given the output 3×3 marginal covariances and $N \times N$ residue coupling from DYNAPROT-M and DYNAPROT-J respectively, we reconstruct a full joint covariance using the heuristic defined in Eq. 3. This direct access to the joint distribution enables extremely fast sampling of diverse structures. For evaluation, we sample 250 structures with DYNAPROT, AFMD+TEMPLATES, and NMA to form ensembles for each of the 82 test set proteins in the AFMD split and follow evaluation metrics from Jing et al. (2024). Some examples of where DYNAPROT outperforms AFMD+TEMPLATES on ensemble flexibility correlation are

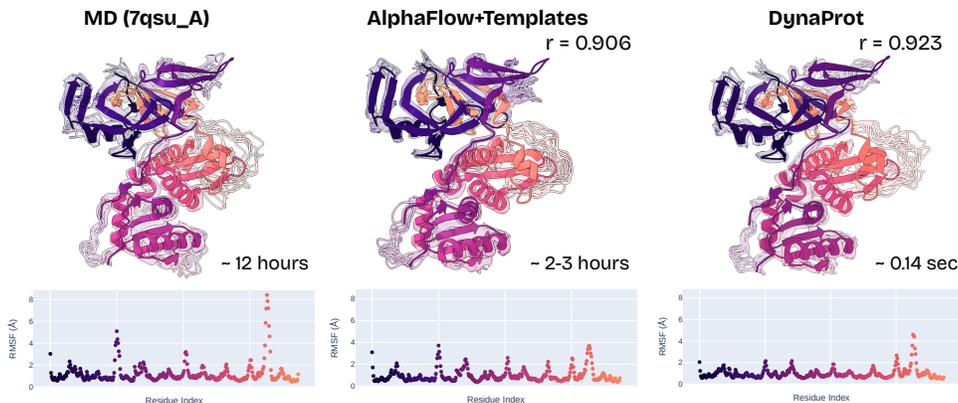


Figure 4: Comparison of DYNAPROT generated ensemble vs. AFMD+T to ATLAS MD simulation (PDB 7qsu_A) overlaid on reference. RMSF Pearson correlation r and sample time reported.

visualized in Figure 4. Moreover, DYNAPROT consistently outperforms NMA across nearly all evaluations—except for transient contact prediction—particularly excelling in measures of local flexibility and pairwise distance preservation (Table 5). Notably, DYNAPROT requires only 2.86 million parameters (vs. 95 million for AFMD+TEMPLATES) and samples ensembles over $70,000\times$ faster on average (~ 0.14 s vs. $\sim 10,000$ s), all while being trained only to predict marginal and scalar covariances. Finally, we assess DYNAPROT’s generalization to longer timescale dynamics, by comparing its zero-shot ensemble of BPTI to the 1ms trajectory from Shaw et al. (2010). Even with these larger conformational changes, DYNAPROT performs reasonably well. It achieves RMSF correlation of 0.88 (c.f. 86 on ATLAS), anisotropy with RMWD of 0.52 \AA (c.f. 1.18 \AA on ATLAS), and strong recovery of transient contacts ($J = 0.54$, c.f. 0.29 on ATLAS). See appendix A.4.1.

Zero-shot cryptic pocket discovery Beyond accuracy, DYNAPROT-M’s marginals can also provide functional insight. Many proteins are considered to be undruggable as their *apo* form may not display a clear binding pocket. However, the pocket may only become apparent after the drug is bound (*holo* form)—a so called “cryptic pocket.” Identification of such pockets is important in drug discovery (Mou et al., 2025). As a case study, we sought to investigate DYNAPROT’s ability in cryptic pocket identification for the enzyme adenylosuccinate synthetase, as it is known to exhibit a cryptic pocket (Meller et al.). We applied DYNAPROT-M to zero shot predict the marginal Gaussians on the *apo* form. The residues with the largest variance are exactly those that encompass the binding pocket. Further, the directionality is consistent with a potential pocket opening motion (when compared with the *holo* form 1CIB). These early results suggest the potential of DYNAPROT’s utility in cryptic pocket discovery.

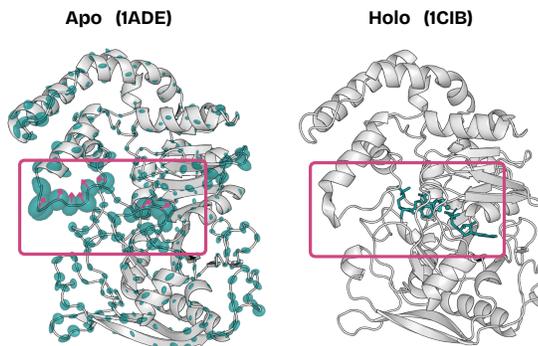


Figure 5: DYNAPROT-M predicted residue Gaussians (ellipsoids) overlaid the *apo* form.

4 Conclusion

We introduce DYNAPROT, a lightweight and data-driven alternative akin to Normal Mode Analysis (NMA), but designed to directly predict structured dynamics descriptors in the form of per-residue and pairwise Gaussian representations. This formulation enables extreme parameter efficiency while outperforming traditional baselines on key metrics, including flexibility estimation, marginal anisotropy, and residue–residue coupling. Remarkably, DYNAPROT’s outputs also support ultra-fast ensemble sampling with reasonable structural fidelity—offering a compelling alternative to conventional ensemble generation methods. Our approach highlights the promise of explicitly learning structured representations of dynamics.

Acknowledgements

We would like to thank Alexander Amini, Jeffrey Li, Yuseong (Nick) Oh, Ryan Mei, Manoj Niverthi, Renzo Soatto, and Kabir Doshi for their useful ideas and discussions.

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 1R35GM141861-01.

References

- Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019.
- Qiang Cui and Ivet Bahar. *Normal mode analysis: theory and applications to biological and chemical systems*. CRC press, 2005.
- Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.
- Chao Hou and Yufeng Shen. Seqdance: A protein language model for representing protein dynamic properties. *bioRxiv*, 2024.
- Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xianqiu Li, and Xilin Chen. Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *International conference on machine learning*, pages 720–729. PMLR, 2015.
- Harold Jeffreys. *The theory of probability*. OuP Oxford, 1998.
- Bowen Jing, Bonnie Berger, and Tommi Jaakkola. AlphaFold meets flow matching for generating protein ensembles. *arXiv preprint arXiv:2402.04845*, 2024.
- Bowen Jing, Bonnie Berger, and Tommi Jaakkola. Ai-based methods for simulating, sampling, and predicting protein ensembles. *arXiv preprint arXiv:2509.17224*, 2025.
- 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.
- Petr Kouba, Joan Planas-Iglesias, Jiri Damborsky, Jiri Sedlar, Stanislav Mazurenko, and Josef Sivic. Learning to engineer protein flexibility. *arXiv preprint arXiv:2412.18275*, 2024.
- Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.
- Valentin Lombard, Dan Timsit, Sergei Grudin, and Elodie Laine. Seamoons: Prediction of molecular motions based on language models. *bioRxiv*, pages 2024–09, 2024.
- Jiarui Lu, Xiaoyin Chen, Stephen Zhewen Lu, Chence Shi, Hongyu Guo, Yoshua Bengio, and Jian Tang. Structure language models for protein conformation generation. *arXiv preprint arXiv:2410.18403*, 2024.
- Artur Meller, Michael Ward, Jonathan Borowsky, Meghana Kshirsagar, Jeffrey M. Lotthammer, Felipe Oviedo, Juan Lavista Ferres, and Gregory R. Bowman. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. 14(1):1177. ISSN 2041-1723. doi: 10.1038/s41467-023-36699-3.
- Minyue Mou, Weicheng Yang, Guangyi Huang, Xiaoyan Yang, Xiao Zhang, Wasala Mudiyansele, Wishwajith Wickramabahu Kandegama, Charles R Ashby Jr, Gefei Hao, and Yangyang Gao. The discovery of cryptic pockets increases the druggability of “undruggable” proteins. *Medicinal Research Reviews*, 2025.

- David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.
- Lars Skjaerven, Siv M Hollup, and Nathalie Reuter. Normal mode analysis for proteins. *Journal of Molecular Structure: THEOCHEM*, 898(1-3):42–48, 2009.
- Yann Vander Meersche, Gabriel Cretin, Aria Gheeraert, Jean-Christophe Gelly, and Tatiana Galochkina. ATLAS: protein flexibility description from atomistic molecular dynamics simulations. *Nucleic Acids Research*, 52(D1):D384–D392, 2024.
- Raviteja Vemulapalli and David W Jacobs. Riemannian metric learning for symmetric positive definite matrices. *arXiv preprint arXiv:1501.02393*, 2015.
- Yan Wang, Lihao Wang, Yuning Shen, Yiqun Wang, Huizhuo Yuan, Yue Wu, and Quanquan Gu. Protein conformation generation via force-guided se (3) diffusion models. *arXiv preprint arXiv:2403.14088*, 2024.
- Edward W Yu and Daniel E Koshland Jr. Propagating conformational changes over long (and short) distances in proteins. *Proceedings of the National Academy of Sciences*, 98(17):9517–9520, 2001.
- Mingyang Zhang, Ting Chen, Xun Lu, Xiaobing Lan, Ziqiang Chen, and Shaoyong Lu. G protein-coupled receptors (gpcrs): advances in structures, mechanisms and drug discovery. *Signal Transduction and Targeted Therapy*, 9(1):88, 2024.
- She Zhang, James M Krieger, Yan Zhang, Cihan Kaya, Burak Kaynak, Karolina Mikulska-Ruminska, Pemra Doruker, Hongchun Li, and Ivet Bahar. Prody 2.0: increased scale and scope after 10 years of protein dynamics modelling with python. *Bioinformatics*, 37(20):3657–3659, 2021.
- Kirill Zinovjev, Paul Guénon, Carlos A Ramos-Guzmán, J Javier Ruiz-Pernía, Damien Laage, and Iñaki Tuñón. Activation and friction in enzymatic loop opening and closing dynamics. *Nature Communications*, 15(1):2490, 2024.

A Appendix

A.1 Method details

A.1.1 Gaussian view of MD

At the local level, the marginal distribution (Figure 1.2) for a single residue i is obtained by integrating out all other residue coordinates: $p(\mathbf{x}_i) = \int p(\mathbf{x}_1, \dots, \mathbf{x}_n) d\mathbf{x}_{-i}$, where $d\mathbf{x}_{-i} := \prod_{j \neq i} d\mathbf{x}_j$. This results in a 3D Gaussian distribution over the C_α coordinates of residue i :

$$\mathbf{X}_i \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{\text{marginal}}^{(i)}), \quad \boldsymbol{\mu}_i \in \mathbb{R}^3, \quad \boldsymbol{\Sigma}_{\text{marginal}}^{(i)} \in \mathbb{R}^{3 \times 3} \quad (2)$$

where $\boldsymbol{\Sigma}_{\text{marginal}}^{(i)}$ is the 3×3 diagonal block of $\boldsymbol{\Sigma}_{\text{joint}}$. These marginals can be interpreted as *Gaussian blobs* encoding anisotropic local fluctuations—i.e., spatial variance of where each residue may reside.

Notably, this formulation allows for simple derivation of scalar flexibility metrics such as the *root-mean-square fluctuation (RMSF)* as $\text{RMSF}_i = \sqrt{\text{Tr}(\boldsymbol{\Sigma}_{\text{marginal}}^{(i)})}$. RMSF (Figure 1.1) represents a simple notion of dynamics: a single scalar per residue quantifying positional fluctuation. However, it discards directional and covariance structure captured by the full marginal.

To capture dynamics beyond residue-local fluctuations, we also consider a covariance matrix $\mathbf{C} \in \mathbb{R}^{N \times N}$ of scalar pairwise coupling (Figure 1.3). Each entry C_{ij} summarizes the dynamical coupling between residues i and j , typically computed as a scalar projection of the corresponding 3×3 block in the full joint covariance: $\boldsymbol{\Sigma}_{\text{joint}}[3i : 3i + 3, 3j : 3j + 3]$. We choose MeanPooling as the scalar projection to compute each C_{ij} . This compact representation enables efficient modeling of residue-residue coupling.

Table 3: Taxonomy of protein dynamics representations under a Gaussian view.

Level	Description	Notation	Space	Captures
1	Per-residue scalar (i.e. RMSF)	$\sqrt{\text{Tr}(\boldsymbol{\Sigma}_{\text{marginal}}^{(i)})}$	\mathbb{R}^N	Magnitude of fluctuation per residue
2	Per-residue full (Gaussian blob)	$\boldsymbol{\Sigma}_{\text{marginal}}^{(i)}$	$\mathbb{R}^{N \times 3 \times 3}$	Anisotropic local covariance per residue
3	Joint scalar (pairwise coupling)	C_{ij}	$\mathbb{R}^{N \times N}$	Scalar covariance across all residues
4	Joint full covariance	$\boldsymbol{\Sigma}_{\text{joint}}$	$\mathbb{R}^{3N \times 3N}$	Full spatial covariance across all residues

A.1.2 SPD closure of joint reconstruction heuristic

Given the marginal Gaussians and the scalar coupling, we want to define a heuristic that reconstructs the full joint covariance matrix (to be used for ensemble generation).

Proposition A.1 (SPD Closure). *Given marginal covariances $\{\boldsymbol{\Sigma}_{\text{marginal}}^{(i)} \in \mathbb{R}^{3 \times 3}\}_{i=1}^N$ and correlation matrix $\tilde{\mathbf{C}} \in \mathbb{R}^{N \times N}$ to be symmetric and positive definite, then the reconstructed joint covariance $\boldsymbol{\Sigma}_{\text{joint}} = \mathbf{L}_{\text{marginal}}(\tilde{\mathbf{C}} \otimes \mathbf{I}_3)\mathbf{L}_{\text{marginal}}^\top$ is also symmetric and positive definite.*

Proof. Let $\mathbf{L}_{\text{marginal}} \in \mathbb{R}^{3N \times 3N}$ be the block-diagonal matrix defined as

$$\mathbf{L}_{\text{marginal}} = \bigoplus_{i=1}^N \mathbf{L}_i,$$

where each $\mathbf{L}_i \in \mathbb{R}^{3 \times 3}$ is the Cholesky factor (or any valid matrix square root) of the positive definite matrix $\boldsymbol{\Sigma}_{\text{marginal}}^{(i)}$. Since each $\boldsymbol{\Sigma}_{\text{marginal}}^{(i)} \succ 0$, it follows that each \mathbf{L}_i is full rank, lower triangular, and has positive diagonal entries. Consequently, $\mathbf{L}_{\text{marginal}}$ is full rank and lower triangular with positive diagonal blocks.

Now consider the matrix $\tilde{C} \otimes I_3 \in \mathbb{R}^{3N \times 3N}$. Since $\tilde{C} \succ 0$ and $I_3 \succ 0$, the Kronecker product $\tilde{C} \otimes I_3 \succ 0$ as well (Kronecker product of two SPD matrices is also SPD). Finally, the product

$$\Sigma_{\text{joint}} = \mathbf{L}_{\text{marginal}} \left(\tilde{C} \otimes I_3 \right) \mathbf{L}_{\text{marginal}}^\top$$

is a congruence transformation of the SPD matrix $\tilde{C} \otimes I_3$ by the full-rank matrix $\mathbf{L}_{\text{marginal}}$. Since congruence preserves positive definiteness, we conclude:

$$\Sigma_{\text{joint}} \succ 0$$

Moreover, Σ_{joint} is symmetric because it is of the form $\mathbf{A}\mathbf{B}\mathbf{A}^\top$. □

A.1.3 Ensemble sampling.

Drawing from the univariate identity $\text{Cov}(i, j) = \text{Corr}(i, j) \cdot \sigma_i \sigma_j$, we define the multivariate cross-covariance block between residues i and j as $\Sigma_{\text{joint}}^{(i, j)} = \mathbf{L}_i \tilde{C}_{ij} \mathbf{L}_j^\top$. Here, the Cholesky factor \mathbf{L}_i serves as a matrix square root of the covariance $\Sigma_{\text{marginal}}^{(i)}$, analogous to standard deviation in the univariate case. And, \tilde{C} is a correlation matrix found by standardizing C . Using the Kronecker product, we can denote this heuristic cleanly as follows,

$$\Sigma_{\text{joint}} = \mathbf{L}_{\text{marginal}} \left(\tilde{C} \otimes I_3 \right) \mathbf{L}_{\text{marginal}}^\top \quad (3)$$

Given the reconstructed joint covariance Σ_{joint} and our assumption that the mean $\boldsymbol{\mu}$ corresponds to the coordinates of the input structure (e.g., the PDB), we have now retrieved our Gaussian distribution over conformations $\mathcal{N}(\boldsymbol{\mu}, \Sigma_{\text{joint}})$. To sample from this distribution, we apply a multivariate generalization of the reparameterization trick used in univariate Gaussian sampling.

Property A.1 (Multivariate Gaussian Sampling). *Given $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where $\Sigma \in \mathbb{R}^{d \times d}$ is SPD and $\Sigma = \mathbf{L}\mathbf{L}^\top$ is its Cholesky decomposition. Then,*

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{L}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d) \quad \Rightarrow \quad \mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$

Proof. Since Gaussian distributions are fully characterized by their first two cumulants (mean and covariance), it suffices to show that the transformed variable has the desired mean and covariance.

Mean of \mathbf{x} :

$$\mathbb{E}[\mathbf{x}] = \mathbb{E}[\boldsymbol{\mu} + \mathbf{L}\boldsymbol{\epsilon}] = \boldsymbol{\mu} + \mathbf{L} \cdot \mathbb{E}[\boldsymbol{\epsilon}] = \boldsymbol{\mu}$$

Covariance of \mathbf{x} :

$$\begin{aligned} \text{Cov}[\mathbf{x}] &= \mathbb{E} [(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^\top] \\ &= \mathbb{E} [(\mathbf{L}\boldsymbol{\epsilon})(\mathbf{L}\boldsymbol{\epsilon})^\top] \\ &= \mathbb{E} [\mathbf{L}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top \mathbf{L}^\top] \\ &= \mathbf{L} \cdot \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top] \cdot \mathbf{L}^\top \\ &= \mathbf{L} \cdot \mathbf{I}_d \cdot \mathbf{L}^\top \\ &= \mathbf{L}\mathbf{L}^\top = \Sigma \end{aligned}$$

□

Note that this sampling relies directly on the Cholesky factor (similar to a matrix square root), mirroring the scalar case. Utilizing DYNAPROT predictions and this heuristic, ensemble sampling becomes extremely fast with minimal computational overhead.

A.2 Evaluation Metrics

RMWD Variance Contribution. To evaluate the efficacy of the marginal Gaussian predictions, we adopt distributional similarity metrics used in [Jing et al. \(2024\)](#). The first of these is the *root mean 2-Wasserstein distance* (RMWD), specifically its variance contribution term. The 2-Wasserstein distance between two multivariate Gaussians has a closed-form expression as follows.

Let $\mathcal{N}_0 = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\mathcal{N}_1 = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ be two d -dimensional Gaussian distributions. The squared 2-Wasserstein distance between them is given by:

$$\mathcal{W}_2^2(\mathcal{N}_0, \mathcal{N}_1) = \|\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1\|_2^2 + \text{Tr} \left(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1 - 2 \left(\boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\Sigma}_0 \boldsymbol{\Sigma}_1^{1/2} \right)^{1/2} \right) \quad (4)$$

This expression consists of two additive components: a mean contribution and a covariance (variance) contribution. This metric is also referred to as the Bures–Wasserstein distance ([Bhatia et al., 2019](#)). Since our method predicts only the covariances, we isolate and evaluate only the second term. We define the RMWD variance contribution across N residues as follows:

$$\text{RMWD}_{\text{var}}(\mathcal{N}_0, \mathcal{N}_1) = \sqrt{\frac{1}{N} \sum_{i=1}^N \text{Tr} \left(\boldsymbol{\Sigma}_{0,i} + \boldsymbol{\Sigma}_{1,i} - 2 \left(\boldsymbol{\Sigma}_{1,i}^{1/2} \boldsymbol{\Sigma}_{0,i} \boldsymbol{\Sigma}_{1,i}^{1/2} \right)^{1/2} \right)} \quad (5)$$

Symmetric KL Divergence Variance Contribution. Alongside the Wasserstein-based metric, we also evaluate the discrepancy between predicted and ground-truth marginal distributions using the *symmetric Kullback–Leibler (KL) divergence*, defined as the mean of the two directed KL divergences mentioned in ([Kullback and Leibler, 1951](#); [Jeffreys, 1998](#)):

$$\text{KL}_{\text{sym}}(\mathcal{N}_0 \parallel \mathcal{N}_1) = \frac{1}{2} [\text{KL}(\mathcal{N}_0 \parallel \mathcal{N}_1) + \text{KL}(\mathcal{N}_1 \parallel \mathcal{N}_0)]$$

For two d -dimensional Gaussian distributions $\mathcal{N}_0 = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $\mathcal{N}_1 = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, the KL divergence from \mathcal{N}_0 to \mathcal{N}_1 is given by:

$$\text{KL}(\mathcal{N}_0 \parallel \mathcal{N}_1) = \frac{1}{2} \left[\text{Tr}(\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_0) + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - d + \log \frac{\det \boldsymbol{\Sigma}_1}{\det \boldsymbol{\Sigma}_0} \right] \quad (6)$$

This expression consists of both a *mean contribution*—the Mahalanobis term—and a *variance contribution*, comprising the trace and log-determinant terms. Since our method predicts only covariances (and uses the input structure coordinates as means), we isolate the variance terms by omitting $(\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0)$ the mean term:

$$\text{KL}_{\text{var}}(\mathcal{N}_0 \parallel \mathcal{N}_1) = \frac{1}{2} \left(\text{Tr}(\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_0) - d + \log \frac{\det \boldsymbol{\Sigma}_1}{\det \boldsymbol{\Sigma}_0} \right) \quad (7)$$

To symmetrize the variance contribution of the divergence, we define the symmetric variance KL as:

$$\text{KL}_{\text{symvar}}(\mathcal{N}_0, \mathcal{N}_1) = \frac{1}{2} (\text{KL}_{\text{var}}(\mathcal{N}_0 \parallel \mathcal{N}_1) + \text{KL}_{\text{var}}(\mathcal{N}_1 \parallel \mathcal{N}_0)) \quad (8)$$

$$= \frac{1}{4} (\text{Tr}(\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_0) + \text{Tr}(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1) - 2d) \quad (9)$$

A.3 Experimental Setup

Preprocessing. We construct ground-truth dynamics labels from the ATLAS molecular dynamics dataset, which comprises 1,390 proteins selected based on structural diversity using the ECOD domain classification (Vander Meersche et al., 2024). Following ALPHAFLOW (Jing et al., 2024) for preprocessing consistency, we concatenate each of the three replicate simulations of 100 ns per protein and extract the C_α coordinates. From each ensemble, we compute the empirical full joint covariance matrix over time and extract the relevant dynamics labels (3×3 marginals per residue and $N \times N$ residue coupling). We evaluate under two train/val/test split regimes. The primary matches ALPHAFLOW’s (1265/39/82), while comparisons to FLEXP3D use DYNAPROT trained on their topology-based split (1112/139/139). For naming, we refer to DYNAPROT-M for the model trained for marginals, DYNAPROT-J for the coupling predictions, and DYNAPROT for both.

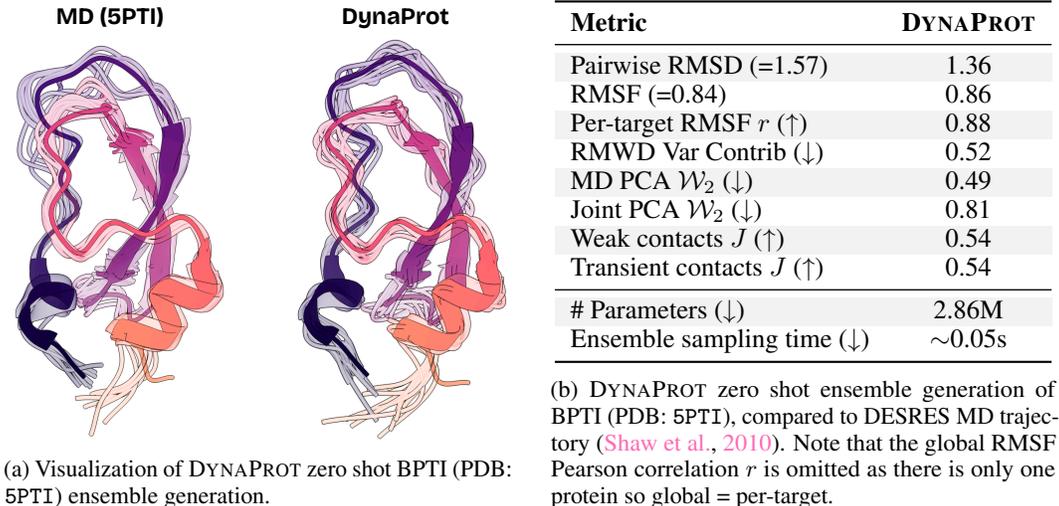
Baselines. For a faithful comparison, we mainly choose baseline methods that take a protein structure as input and predict dynamics descriptors either *implicitly* (AFMD+TEMPLATES) or *explicitly* (FLEXP3D, NMA). For NMA, we utilize the ProDy package (Zhang et al., 2021), specifically the Anisotropic Network Model instantiation. For a broader set of baselines, we also compare against some sequence based methods in Appendix A.4.4.

A.4 Additional Experiments

A.4.1 DYNAPROT zero-shot ensemble generation of BPTI

We note that DYNAPROT was trained on the ATLAS MD dataset comprising 100 ns per replicate trajectories. In contrast, D.E. Shaw Research performed simulations of BPTI (PDB: 5PTI) at millisecond-scale revealing structurally distinct conformational states (Shaw et al., 2010). Thus, in an effort to understand DYNAPROT’s ability to generalize to long-timescale dynamics, we applied it to BPTI and compared to the DESRES trajectory.

Listed in Table 6b, we compute the ensemble evaluation metrics from Jing et al. (2024) and observe that DYNAPROT performs remarkably well: e.g., RMSF correlation of 0.88 (c.f. 86 on ATLAS), local anisotropy with RMWD of 0.52 Å (c.f. 1.18 Å on ATLAS), and strong recovery of transient contacts (Jaccard similarity 0.54, c.f. 0.29 on ATLAS). These metrics emphasize that DYNAPROT is able to model larger conformational changes at high fidelity.



A.4.2 DYNAPROT ablations

To test both the importance of DYNAPROT’s Riemannian aware loss (log Frobenius norm) and the SE(3) invariance from the IPA layers, we have performed the following ablations listed in Table 4. Unsurprisingly, replacing the log Frobenius norm objective with standard Mean Squared Error loss significantly degrades performance as the optimization is over the space of positive definite covariance matrices, which lies on a well-studied Riemannian manifold. Replacing the IPA blocks with standard MLPs also degrades performance, suggesting that SE(3) invariance is crucial in this low-data, low-parameter regime.

Table 4: DYNAPROT-M ablations of the log Frobenius loss and SE(3) invariance.

Metric	DynaProt	No LogFrob Loss	No SE(3) Invariance
RMWD Var (\downarrow)	1.18	2.70	1.92
Sym KL Var (\downarrow)	0.91	9.26	4.46
RMSF r (\uparrow)	0.87	0.38	0.48

A.4.3 Structure input baselines

Table 5 summarizes the ensemble evaluation results across AFMD+TEMPLATES, DYNAPROT, and NMA. DYNAPROT achieves performance comparable to AFMD+TEMPLATES on key flexibility metrics such as pairwise RMSD and per-target RMSF correlation, while lagging slightly behind on distributional similarity and observable recovery. Moreover, DYNAPROT consistently outperforms NMA across nearly all evaluations, particularly excelling in measures of local flexibility and pairwise distance preservation.

Table 5: Comparison of C_α ensemble evaluation metrics on ATLAS MD Dataset between AFMD+Templates, DYNAPROT, and NMA. Underlined is second best.

Metric	AFMD+T	DYNAPROT	NMA
Pairwise RMSD (=2.89)	2.18	<u>2.17</u>	0.91
RMSF (=1.48)	1.17	<u>1.10</u>	0.52
Global RMSF r (\uparrow)	0.91	<u>0.71</u>	0.54
Per-target RMSF r (\uparrow)	0.92	<u>0.86</u>	0.76
MD PCA \mathcal{W}_2 (\downarrow)	1.25	<u>1.74</u>	1.86
Joint PCA \mathcal{W}_2 (\downarrow)	1.58	<u>2.39</u>	2.45
Weak contacts J (\uparrow)	0.62	<u>0.51</u>	0.43
Transient contacts J (\uparrow)	0.47	0.29	<u>0.33</u>
# Parameters (\downarrow)	95 M	2.86 M	-
Ensemble sampling time (\downarrow)	$\sim 10,000$ s	~ 0.14 s	~ 5.69 s

A.4.4 Sequence input baselines

There are many methods worth noting that aim to predict ensembles or dynamics descriptors from sequence itself: standard ALPHAFLOW (AFMD), MSA-SUBSAMPLING, FLEXPART-SEQ, ESMDIFF (Lu et al., 2024), CONFDIFF (Wang et al., 2024), SEQDANCE (Hou and Shen, 2024), and SEAMOON (Lombard et al., 2024). Though DYNAPROT’s true comparison is NMA as it is a data driven and learnable alternative, we still compare against some of these sequence based methods in ensemble generation. DYNAPROT outperforms these methods on local RMSF correlation and marginal anisotropy prediction and is comparable with other distributional metrics. Moreover, the efficiency advantage is clear with DYNAPROT’s sub-second sample time.

Table 6: Comparison of DYNAPROT generated with ensemble generation methods that take in sequence as input. ESMDiff, ESM3 entries reported from Lu et al. (2024).

Metric	DynaProt	ConfDiff OF-r3-MD	AlphaFlow -MD	BioEmu	ESM3 (ID)	ESMDiff (ID)
Pairwise RMSD (=2.89)	2.17	3.43	2.89	3.57	-	-
RMSF (=1.48)	1.10	2.21	1.68	2.47	-	-
Global RMSF r (\uparrow)	0.71	0.67	0.60	0.63	0.19	0.49
Per-target RMSF r (\uparrow)	0.86	0.85	0.85	0.77	0.67	0.68
RMWD Var Contrib (\downarrow)	1.18	1.40	1.30	2.04	4.35	3.37
MD PCA \mathcal{W}_2 (\downarrow)	1.74	1.44	1.52	2.05	2.06	2.29
Joint PCA \mathcal{W}_2 (\downarrow)	2.39	2.25	2.25	4.22	5.97	6.32
Weak contacts J (\uparrow)	0.51	0.59	0.62	0.33	0.45	0.52
Transient contacts J (\uparrow)	0.29	0.36	0.41	0.19	0.26	0.26
# Parameters (\downarrow)	2.86M	12.64M	95M	31M	1.4B	1.4B
Sampling time (\downarrow)	~ 0.14s	~ 570 s	$\sim 10,000$ s	~ 240 s	~ 70 s	~ 70 s

A.5 Supplementary Figures

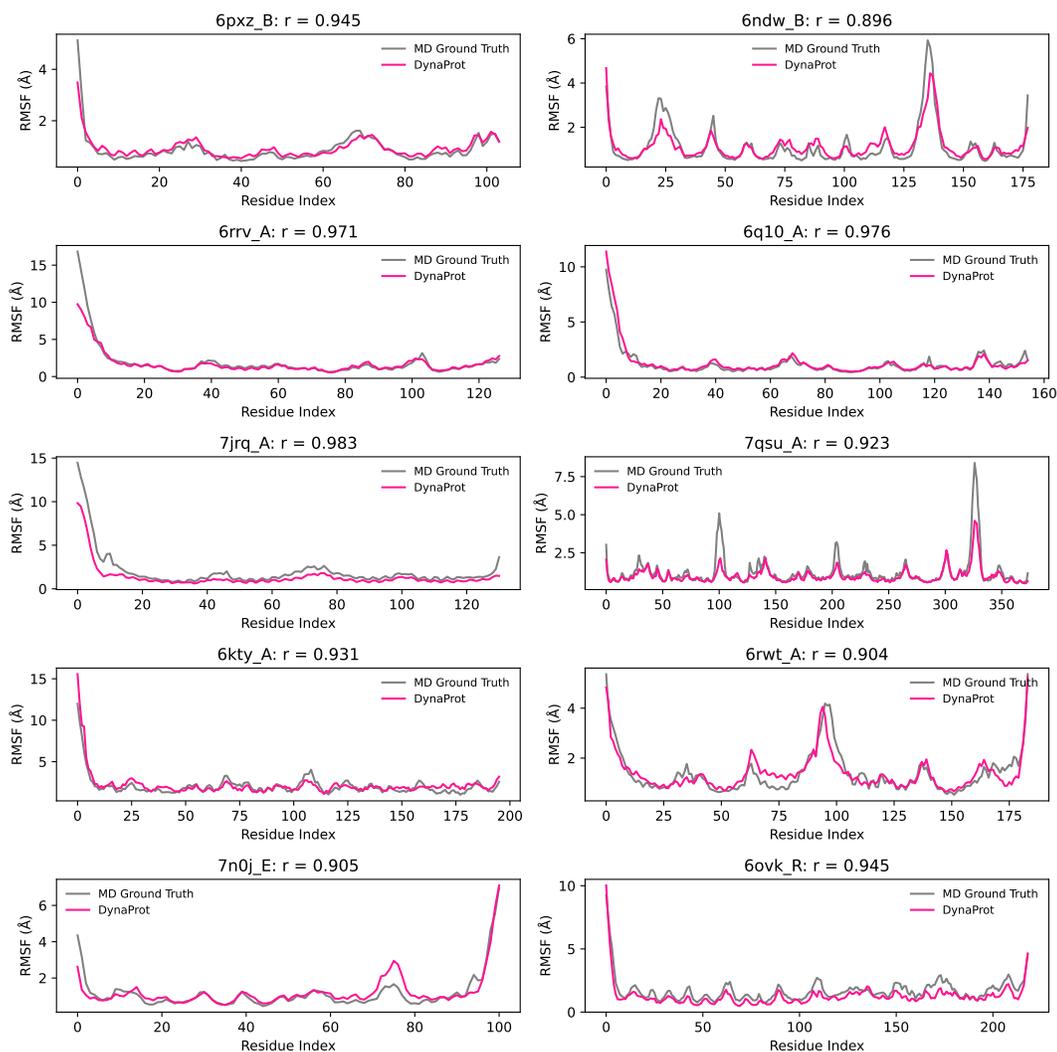


Figure 7: **DYNAPROT-M predicted RMSF correlations.** Visualized test set examples of predicted RMSF per residue (derived from the predicted marginal Gaussians) compared to ground truth RMSF derived from MD trajectories. Pearson correlation coefficient (r) between predicted and ground truth RMSF is reported.