Normal Mode Diffusion: Towards Dynamics-Informed Protein Design

Anonymous Authors[1]

Abstract

We introduce a new method that uses normal mode analysis (NMA) to condition diffusion models for protein design to create proteins with specific dynamical properties – that is, their lowest non-trivial normal mode moves a selected set of residues in a targeted way. We demonstrate that our approach is feasible by conditioning an equivariant graph-diffusion model for protein backbone generation to design molecules with a pre-defined lowest normal mode. Our work represents a first step towards incorporating dynamical behaviour in protein design, and may open the door to designing more flexible and functional proteins in the future.

1. Introduction

Generative artificial intelligence (AI) has demonstrated significant advances in molecular design. In protein design, denoising diffusion models allow the controllable generation of a range of proteins (Watson et al., 2022; Ingraham et al., 2022). Current models allow researchers to condition protein designs on desirable characteristics such as shape, scaffolding functional motifs, and other sequence and structural properties.

However, despite the extensive range of properties that can be conditioned on, no methods to condition generative models on protein dynamical properties have been introduced. Here, we present a novel method to condition diffusion models on protein dynamics data derived from normal mode analysis (NMA) (Bahar et al., 2010). Our approach expands the potential for generating proteins with specific dynamical properties, crucial for functional applications.

NMA offers unique advantages for generative AI-driven protein design. First, it presents a fast route to a rough hypothesis of *dynamics* for any protein purely based on structure, without the need for expensive molecular dynamics information (Bahar et al., 2010). Second, despite the fact that NMA is less sophisticated than molecular dynamics simulations, the lowest (5-15) non-trivial normal modes have been shown to capture a substantial portion of functional motions in proteins (Bahar et al., 2010). Prominent examples include dihydrofolate reductase (DHFR) (Bahar et al., 1997), lysozyme (Gibrat & Gō, 1990), and adenylate kinase (AdK) (Tama & Sanejouand, 2001), among others.

Our results show how protein backbone generation can be conditioned on such that their lowest normal mode moves a select set of residues move in a targeted way. This provides a route to designing proteins with tailored dynamical properties which may open new avenues for protein engineering and drug discovery. Our contributions are:

1. To our knowledge, this is the first proposal to introduce explicit conditioning on dynamical properties into diffusion models for protein design.
2. We evaluate our method in a proof-of-concept setting on its ability to generate protein structures that are (1) realistic, (2) novel and (3) follow the specified dynamics condition. Our experiments demonstrate that conditioning on NMA-dynamics is viable and can be incorporated in current diffusion models.

2. Background and related work

Diffusion models Denoising diffusion probabilistic models (DDPM) (Sohl-Dickstein et al., 2015; Ho et al., 2020) have been applied in a variety of tasks within and outside biology, such as image synthesis (Dhariwal & Nichol, 2021; Kong et al., 2021), drug design (Schneuing et al., 2023) and protein design (Watson et al., 2022; Ingraham et al., 2022). Given a data sample x_0, DDPMs learn are trained to remove random noise that is added to the sample. By doing so, they learn to reconstruct data samples from noise. We refer the reader to Ho et al. (2020) for a good introduction.

Normal mode analysis (NMA) NMA is a technique to investigate the vibrational dynamics of a system of molecules (Bahar et al., 2010). It is based on a harmonic approximation of the potential energy around a given protein structure.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by *the 2023 ICML Workshop on Computational Biology*. Do not distribute.

From this approximation, the vibrational frequencies and corresponding directions of oscillation (normal modes) arise as eigenvectors and eigenvalues of the potential energy Hessian matrix. Compared to molecular dynamics simulations, NMA is computationally efficient and can often distill complex dynamics into a few dominant modes (more in App. B).

3. Methods

Setup We consider the simplified setting of monomeric protein chains which we represent with only the most relevant features as being an *ordered* point cloud $x_0 \in \mathbb{R}^{n\times 3}$ constituted only of the N-to-C ordered list of alpha-carbon coordinates of its residues, without any side-chain information. As training and validation data we extract 10'000 high-resolution (< 3 Å) samples from CATHv4.3 (Orengo et al., 1997) of lenghts between 25-100 amino acids (Details in App. C).

Diffusion loop, denoiser and training We follow Hoogeboom et al. (2022) and use the Markovian DDPM fomulation of Ho et al. (2020) for our forward process. We use the noise schedule from Hoogeboom et al. (2022) with 500 steps and subtract the center of mass from the noise for an equivariant diffusion process (Hoogeboom et al., 2022).
As denoising model $\epsilon_\theta(x_t, t)$, we use the Geometric Vector Perceptron (GVP-GNN) (Jing et al., 2021), where the normalised time-step t is added as node-feature, and the initial layer norms in the embeddings are dropped. We perform message passing on a fully connected C_α graph with positional encoding to capture the chain structure. The denoiser is trained using using the denoising objective from Ho et al. (2020) for 500 epochs with learning rate 1e-4.

Unconditional sampling We follow (Hoogeboom et al., 2022), and start by sampling $x_T \sim \mathcal{N}(0, 1)$, subtracting the center of mass, and applying the denoising steps using

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\sqrt{1-\alpha_t}}{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t)\right) + \eta(1-\alpha_t)z, \tag{1}$$

with $z \sim \mathcal{N}(0, 1)$ with subtracted center of mass. Motivated by Song et al. (2020), we introduce an empirical noise scale η, with $\eta = 1$ being the DDPM formulation in Ho et al. (2020). We find that the quality of the unconditioned samples is generally improved for small η. We therefore use a deterministic reverse process ($\eta = 0$) during all experiments below.

NMA-conditioned sampling We consider the following situation: Given a set $\mathcal{C}$ of residues of interest – for example a functional motif – we would like to specify their dominant relative motion.
To specify the relative motion, we define a *target matrix* $y \in \mathbb{R}^{|\mathcal{C}|\times 3}$ of normal mode component vectors $y_i \in \mathbb{R}^3$ for each residue i in $\mathcal{C}$. The target matrix may be manually specified, or extracted from the normal modes of a functional motif of interest in a target protein. In this work, we wanted to evaluate against diverse motions and therefore choose samples according to a *strain-energy* calculation: We randomly sample a target protein in the holdout data, perform NMA and identify the most flexible part of the protein by calculating the strain energy of each node (Hinsen & Kneller, 1999). We then choose the lowest normal mode component of the 10 consecutive nodes for which the summed energy is largest as our target matrix y.

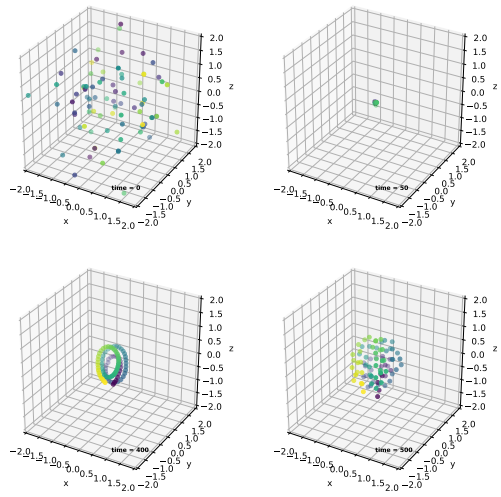

Figure 1: Snapshots of the reverse process for unconditional backbone generation. Point-cloud samples from a Gaussian prior with center-of-mass zero are progressively denoised. We observe that the reverse process firstly aligns the backbone residues in a chain, before expanding the structure to biological sizes. Structural details such as alpha helices or beta strands are formed towards the end of the process.

Conditioning with gradient descent. We take inspiration from classifier-based guidance (Dhariwal & Nichol, 2021), where the score of the data is conditioned on the target y by Bayes rule:

$$\nabla_{x_t} \ln p(x_t|y) = \nabla_{x_t} \ln p(y|x_t) + \nabla_{x_t} \ln p(x_t) \tag{2}$$

The key idea is that instead of training a model to predict $p(y|x_t)$, we exploit that coarse-grained NMA is robust to slight variations in structure (Bahar et al., 2010) and can be computed differentiably. Therefore, once a rough structural hypothesis is formed we can use gradient descent on an NMA based loss to induce a probability flow towards samples that satisfy the condition y:

$$\Delta x = -\gamma(x_t, t)\nabla_{x_t} l(y, v(x_t)) \tag{3}$$

Here $v(x_t)$ is the lowest non-trivial normal mode of x_t, $\gamma(x_t, t)$ is a guidance scale and $l(y, v(x_t))$ is the loss. To ensure meaningful NMA results, we require a somewhat protein-like structure x_t. We therefore start conditioning only from $t < t_{\text{start}}$ onward. As seen in Figure 1, a meaningful chain structure emerges in the middle of our reverse process and the physical distances between residues are only recovered towards the last quarter of the diffusion process. Based on these observations, we found a $t_{\text{start}} = T/5$ to work well for our model.

Upon sampling from the reverse process and starting from $t < t_{\text{start}}$, we condition in the following way. We (1) inflate the current structure to physical size such that mean chain distance is 3.8 Å, (2) extract the lowest non-trivial mode of the structure using Hinsen force-field (Hinsen & Kneller, 1999) parametrisation with a 16Å cut-off, (3) subset the *current mode* to the mode components for conditioned residues $v(x_t) \in \mathbb{R}^{|\mathcal{C}|\times 3}$.
The structure x_t is then optimised towards the target y by computing the loss (Eq. 4) and performing gradient descent. In practice we found it beneficial to split each conditioning time step into $r = 5$ smaller steps, interleaving a partial denoising (with rescaled amplitudes) with a gradient descent step with recalculated loss. To balance the magnitudes of the denoising and the conditioning update, we set $\gamma(x_t, t)$ to the maximal magnitude of the denoising update.

The conditioning loss (Equation 4) is chosen as a simple combination of amplitude and angle terms between all pairwise residues.

$$l_{\text{NMA}}(y, v(x_t)) = l_{\text{angle}}(y, v(x_t)) + 2l_{\text{ampl}}(y, v(x_t)) \quad (4)$$

$$l_{\text{angle}} = \sum_{i,j\in\mathcal{C}} |\cos(y_i, y_j) - \cos(v(x_t)_i, v(x_t)_j)| \quad (5)$$

$$l_{\text{ampl}} = \sum_{i\in\mathcal{C}} \left| \frac{||y_i||}{||y||} - \frac{||v(x_t)_i||}{||v(x_t)||} \right| \quad (6)$$

This choice extracts invariants from the target matrix y which are independent of the reference frame. Further, it ensures rotational equivariance of the gradient with respect to rotations of x_t. The amplitude terms are normalised such that only their relative sizes matter, consistent with the fact that amplitude information from NMA can only make relative statements about the participation of a given residue in a mode (Bahar et al., 2010). For the combined loss, the l_{ampl} is scaled by 2, such that its contribution is similar in magnitute to l_{angle}. This simple loss does not consider higher order correlations (e.g. among the motion of triplets of residues), but could readily be extended to do so.

4. Results

Our goal is to evaluate the efficacy of NMA Diffusion in generating realistic backbones with specific dynamic properties. This prompts us to address three central questions: (1) Do the generated samples represent realistic proteins? (2) If they represent realistic proteins, do they display the desired normal mode dynamics? (3) Are the samples novel compared to the training set?
To assess these questions, we use 300 hold-out target structures of various lengths from our CATH dataset and extract an NMA target condition as explained above. We then generate 3 unconditioned and 4 NMA-conditioned samples for each target condition. We then filter the samples for realistic backbones, requiring that the mean C_α distance along the backbone is within 0.05 Å of the average C_α distance of 3.8 Å(Voet & Voet, 2010). This reduces the number of samples by approximately 25% and 50% for unconditional and conditional samples respectively. Of these, we select the sample with the lowest NMA loss (Eq. 4), leaving with approximately 300 samples per sampling procedure. Two samples are shown in Figure 2 (more in App. F).

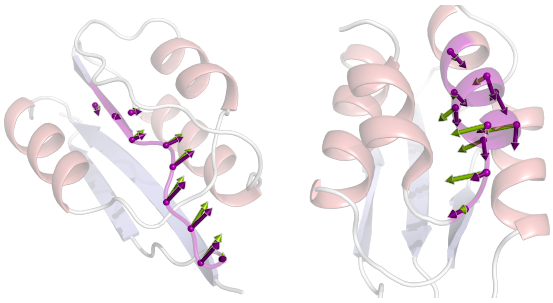

Figure 2: **Left:** conditioned sample with NMA-loss $l_{\text{NMA}} = 0.11$. **Right:** not conditioned sample with $l_{\text{NMA}} = 0.51$. Purple arrows represent the target, green displacements in the novel protein.

Filtered samples represent realistic proteins For the filtered samples, we compute the C_α backbone distances (Fig. 3). The unconditioned samples follow the C_α distance distrubtion of the training set well, indicating that the denoiser can generate realistic backbones. The NMA-conditioned samples show a similar distribution, but have heavy tails extending to occasionally unrealistic distances. When filtering samples with unrealistic distances, we lose another 3x of samples compared to the unconditioned case. We attribute the distortion of backbone distances to the NMA-loss, which does not actively promote realistic backbone distances. Currently, the backbone distances are only corrected by the denoiser, which for our simple denoiser appears to be insufficient. Since we did not spend much time in tuning and training our denoiser for this work, we believe that a more sophisticated denoiser (Ingraham et al., 2022) or an extra term in the conditioning to encourage maintaining realistic backbone distances can fix the drop in sample quality. Overall, we conclude that the NMA-conditioned

sampling can generate realistic backbones, but at the cost of requiring more samples to find a realistic backbone.

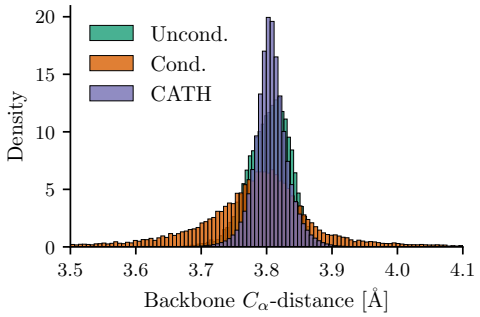

Figure 3: C_α backbone distances of the generated samples compared to the training set.

We further compute secondary structure features from the C_α positions with the P-SEA algorithm (Labesse et al., 1997) and compare it to the distribution of true proteins in our dataset (Fig. 4). The unconditioned samples have a similar secondary structure distribution as the training set, with a slight overrepresentation of alpha helices. This is in line with previous work (Watson et al., 2022) Interestingly, the NMA-conditioned sampling remedies this bias and reproduces the secondary structure distribution of the training distribution better. This shows that the NMA conditioning does destroy the secondary structure and indeed may allow correcting for the bias of unconditioned sampling. When looking at generated samples (App. F) we find that the generated samples still often show slightly unrealistic packing, but that there are also samples which are plausible combination of structural motifs (e.g. Fig. 2).

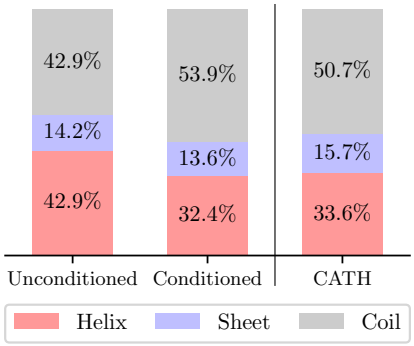

Figure 4: Secondary structure distribution of the generated samples compared to the training set.

Conditioned samples exhibit the targeted normal mode To examine whether conditioning the selected residues to have the targeted lowest normal mode component, we compute the NMA loss (Eq. 4) for the realistic, generated samples and compare it to the unconditioned samples (Fig. 5).

We see that the NMA-conditioned samples are significantly enriched towards low loss values compared to the unconditional samples. In the examples in Fig. 2 and App. F we also see that lowest normal mode of the generated samples (green) is in good agreement with the target (purple). This indicates that NMA-conditioning is effective in generating samples with the targeted lowest normal mode component.

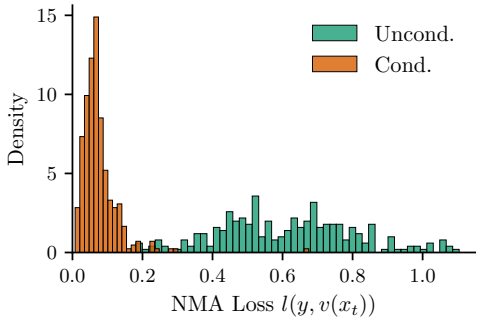

Figure 5: Histograms of the NMA-loss (Eq. 4) for conditioned and not conditioned samples.

Samples are novel. Finally, to ensure that the model did not overfit we compute the TM-score (Zhang & Skolnick, 2005) between the generated samples and the best matching target fragment of the same size in the training set (App. E). Both, conditioned and unconditioned samples, have TM-scores of about 0.3-0.4, indicating that they are novel compared to the training set and do not simply memorise the training data. Importantly, NMA-conditioning does not have a negative effect on the novelty of the samples.

5. Conclusion and further work

We introduced a novel framework to condition diffusion models for protein design on protein dynamics information with normal mode analysis. Our analysis demonstrates that with our *NMA-Diffusion* sampling procedure, it is possible to condition a generative model of protein structure such that its lowest normal mode moves a selected set of residues in a targeted way, while still generating realistic and novel proteins. While the current study represents a proof-of-concept with a simple denoiser, relatively short proteins and omitting side-chains, future work is underway to extend this appraoch to full protein design (Watson et al., 2022; Ingraham et al., 2022) and to investigate whether the so-conditioned samples behave as expected in molecular dynamics simulations. To the best of our knowledge, this is the first time a diffusion model for protein structure has been conditioned on protein dynamics information. We believe this approach holds potential for protein design, where it is desirable to design proteins that are stable and have a desired conformational flexibility.

References

Adamovic, I., Mijailovich, S. M., and Karplus, M. The elastic properties of the structurally characterized myosin ii s2 subdomain: A molecular dynamics and normal mode analysis. *Biophysical Journal*, 94(10):3779–3789, 2008. ISSN 0006-3495. doi: https://doi.org/10.1529/biophysj.107.122028. URL https://www.sciencedirect.com/science/article/pii/S0006349508703813.

Bahar, I., Atilgan, A. R., and Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173–181, 1997.

Bahar, I., Lezon, T. R., Bakan, A., and Shrivastava, I. H. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chemical Reviews*, 110(3):1463–1497, 2010.

Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis. *ArXiv*, abs/2105.05233, 2021.

Gibrat, J.-F. and Gō, N. Normal mode analysis of human lysozyme: study of the relative motion of the two domains and characterization of the harmonic motion. *Proteins: Structure, Function, and Bioinformatics*, 8(3):258–279, 1990.

Hinsen, K. and Kneller, G. R. A simplified force field for describing vibrational protein dynamics over the whole frequency range. *The Journal of Chemical Physics*, 111(24): 10766–10769, 12 1999. ISSN 0021-9606. doi: 10.1063/1.480441. URL https://doi.org/10.1063/1.480441.

Ho, J., Jain, A., and Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv*, 2020.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3D. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S. (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8867–8887. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/hoogeboom22a.html.

Ingraham, J., Baranov, M., Costello, Z., Frappier, V., Ismail, A., Tie, S., Wang, W., Xue, V., Obermeyer, F., Beam, A., and Grigoryan, G. Illuminating protein space with a programmable generative model. *bioRxiv*, 2022.

Jing, B., Eismann, S., Soni, P. N., and Dror, R. O. Equivariant graph neural networks for 3d macromolecular structure, 2021.

Kong, Z., Ping, W., Huang, J., Zhao, K., and Catanzaro, B. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=a-xFK8Ymz5J.

Labesse, G., Colloc'h, N., Pothier, J., and Mornon, J.-P. P-sea: a new efficient assignment of secondary structure from cα trace of proteins. *Bioinformatics*, 13(3):291–295, 1997.

Levitt, M., Sander, C., and Stern, P. S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *Journal of Molecular Biology*, 181(3):423–447, 1985. ISSN 0022-2836. doi: https://doi.org/10.1016/0022-2836(85)90230-X. URL https://www.sciencedirect.com/science/article/pii/002228368590230X.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. Cath–a hierarchic classification of protein domain structures. *Structure*, 5(8):1093–1109, 1997.

Schneuing, A., Du, Y., Harris, C., Jamasb, A. R., Igashov, I., weitao Du, Blundell, T. L., Lio, P., Gomes, C. P., Welling, M., Bronstein, M. M., and Correia, B. Structure-based drug design with equivariant diffusion models, 2023. URL https://openreview.net/forum?id=uKmuzIuVl8z.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In Bach, F. and Blei, D. (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. *arXiv:2010.02502*, October 2020. URL https://arxiv.org/abs/2010.02502.

Tama, F. and Sanejouand, Y. Conformational change of proteins arising from normal mode calculations. *Protein engineering*, 14 1:1–6, 2001.

Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters*, 77 9:1905–1908, 1996.

Voet, D. and Voet, J. G. *Biochemistry*. John Wiley & Sons, 2010.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock,

S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., Lauko, A., Bortoli, V. D., Mathieu, E., Barzilay, R., Jaakkola, T. S., DiMaio, F., Baek, M., and Baker, D. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *biorXiv*, 2022.

Zhang, Y. and Skolnick, J. Tm-align: a protein structure alignment algorithm based on the tm-score. *Nucleic acids research*, 33(7):2302–2309, 2005.

A. Why Normal Mode Analysis (NMA)?

NMA offers unique advantages for generative AI-driven protein design. First, it presents a fast route to a rough hypothesis of *dynamics* for any protein purely based on structure, without the need for expensive molecular dynamics information (Bahar et al., 2010). Second, despite the fact that NMA is less sophisticated than molecular dynamics simulations, the lowest (5-15) non-trivial normal modes have been shown to capture a substantial portion of functional motions in proteins (Bahar et al., 2010). Prominent examples include dihydrofolate reductase (DHFR) (Bahar et al., 1997), lysozyme (Gibrat & Gō, 1990), and adenylate kinase (AdK) (Tama & Sanejouand, 2001), among others.

B. What are NMAs assumptions, limitations and merits?

The simplicity of NMA comes from its strong assumptions (Bahar et al., 2010). It assumes that the given protein structure represents a potential energy minimum, is in thermal equilibrium, and ignores solvent effects. Despite its simplicity, NMA has been a key tool in protein dynamics research since the 1970s (Bahar et al., 2010). Its successful application spans a variety of proteins, providing key insights into their functional mechanisms. For instance, it elucidated the hinge-bending motion in trypsin (Levitt et al., 1985), and the large-scale conformational changes in myosin (Adamovic et al., 2008). Moreover, NMA has been instrumental in decoding the opening and closing mechanism of AdK (Tama & Sanejouand, 2001), and the functional dynamics of DHFR (Bahar et al., 1997). Empirical studies such as these consistently found that the lowest few (typically 5-15) normal modes often capture the predominant part of the functional motions in proteins, beyond the theoretical confines of the harmonic approximation. This makes NMA a practical approach to distill complex protein dynamics into a manageable set of normal modes.

There is an array of NMA models, from coarse-grained, where residues or domains are nodes (Tirion, 1996), to highly detailed models, with atoms as nodes and structures refined in molecular force fields (Bahar et al., 2010). Despite the diverse methods, a common finding is that the specific choice of the NMA model does not radically influence the general trend about protein dynamics (Bahar et al., 2010). This suggests that the essential dynamics of proteins can be captured with reasonable accuracy regardless of the model choice such that our method can be adapted in problems requiring complex force-fields.

We refer the reader to (Bahar et al., 2010) for a comprehensive review of NMA and its applications.

C. Training data

Since this work is meant of a proof of concept, we restricted ourselves to short protein sequences to work with limited computational resources. To obtain an interesting sample of short protein snippets, we filtered CATHv4.3 domains (Orengo et al., 1997) for structures with high resolution (< 3Å), between 20-100 amino acids long. To remove redundancy, we clustered the sequences at 95% sequence similarity. The resulting dataset contains 9'220 protein structures. For these, we set aside 300 for extracting normal mode condition targets that could not have been observed in the training data and use the remaining 8'920 for training.

D. Pre-filtering of samples

We then generate 3 unconditioned and 4 NMA-conditioned samples for each target condition. We then filter the samples for realistic backbones, requiring that the mean C_α distance along the backbone is within 0.05 Å of the average C_α distance of 3.8 Å(Voet & Voet, 2010). This reduces the number of samples by approximately 25% and 50% for unconditional and conditional samples respectively. Of these, we select the sample with the lowest NMA loss (Eq. 4), leaving with approximately 300 samples per sampling procedure.

E. Novelty of samples

To ensure that the model did not overfit we compute the TM-score (Zhang & Skolnick, 2005) between the generated samples and the best matching target fragment of the same size in the training set (Fig. 6, App.). Both, conditioned and unconditioned samples, have TM-scores of about 0.3-0.4, indicating that they are novel compared to the training set and do not simply memorise the training data. Importantly, NMA-conditioning does not have a negative effect on the novelty of the samples.

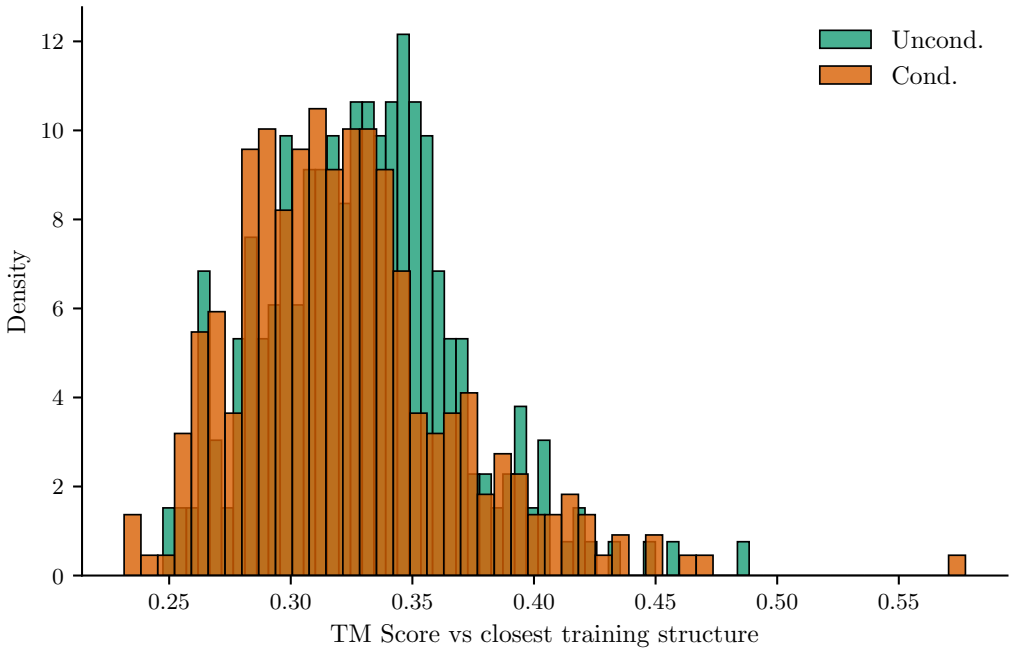

Figure 6: Histogram of the TM-score between the generated samples and the best matching target in the training set. With TM scores around 0.3-0.4, unconditioned and conditioned samples both generate novel samples and NMA-conditioning does not have a negative effect on novelty.

F. Extra samples

Below we provide a few more samples from the unconditional and conditional sampling process, to give a better idea of the diversity and quality of the samples.

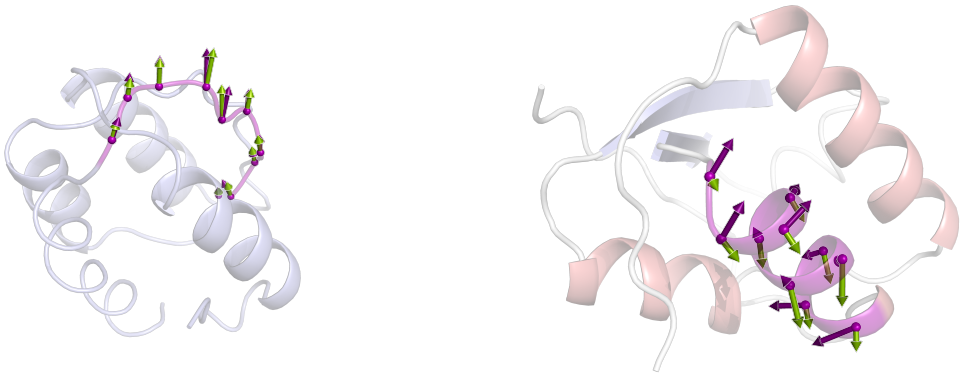

Figure 7: **Left:** conditioned sample with L_1 loss 0.05. **Right:** not conditioned sample L_1 loss 0.89.

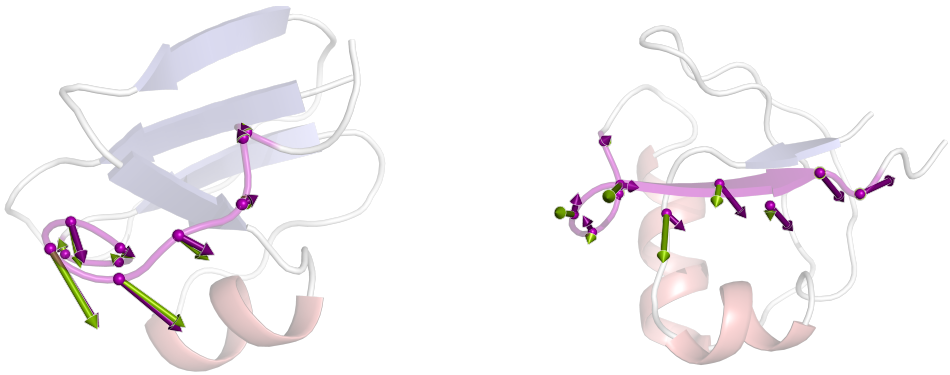

Figure 8: **Left:** conditioned sample with L_1 loss 0.04. **Right:** not conditioned sample L_1 loss 1.02

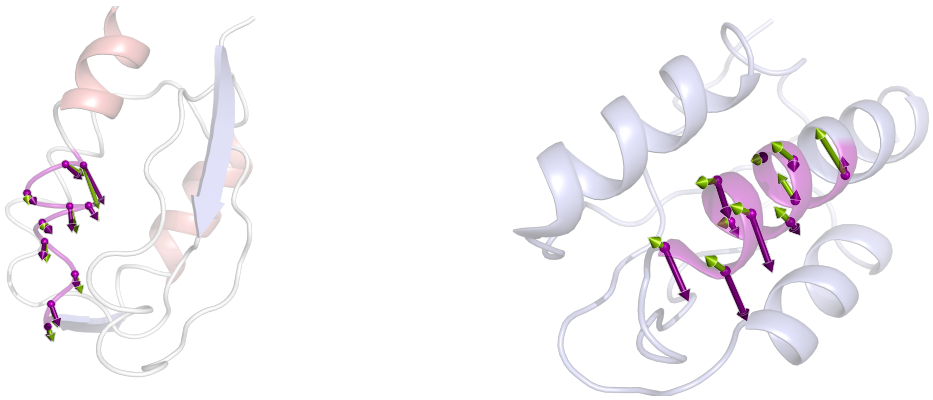

Figure 9: **Left:** conditioned sample, loss 010. **Right:** not conditioned sample, loss 0.82