

# Learning Protein Motions

S. Grudinin<sup>1</sup>, R. Vuillemot<sup>1</sup>, K. Olechnovič<sup>1</sup>, V. Lombard<sup>2</sup>, E. Laine<sup>2,3</sup>

<sup>1</sup>Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France, <sup>2</sup>Department of Computational, Quantitative, and Synthetic Biology (CQSB), UMR 7238, IBPS, Sorbonne Université, CNRS Paris, 75005, France, <sup>3</sup>Institut universitaire de France (IUF)

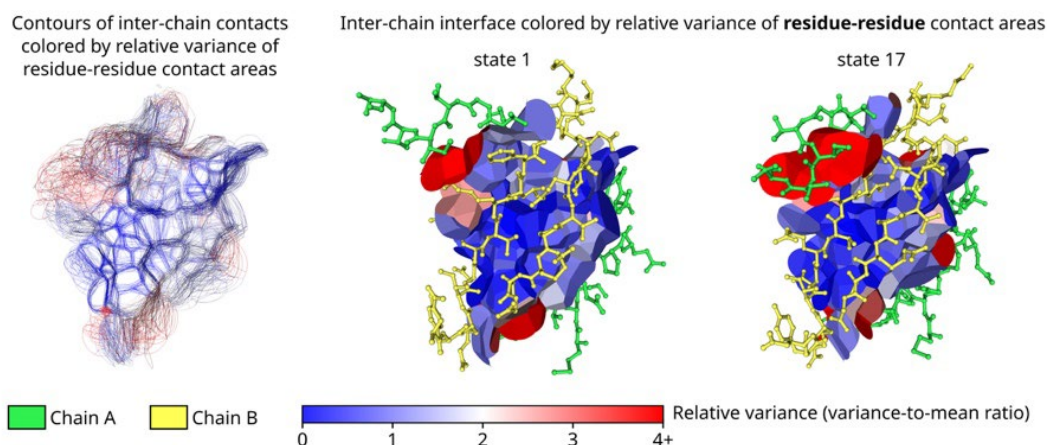
*Sergei.Grudinin@univ-grenoble-alpes.fr*

Proteins move and deform to ensure their biological functions. Despite significant progress in protein structure prediction, approximating conformational ensembles at physiological conditions remains a fundamental open problem. I will present novel perspectives on this problem and our contributions.

One can directly target continuous compact representations of protein motions inferred from sparse experimental observations. To do so, we collected protein single-chain statistics from the PDB for a systematic and comprehensive description of protein families conformational variability [1]. Beyond descriptive analysis, we assessed classical dimensionality reduction techniques for sampling unseen states on a representative benchmark. Then, we developed a task-specific loss function enforcing data symmetries, including scaling and permutation operations. Our method PETIMOT (Protein sEquence and sTructure-based Inference of MOTions) leverages transfer learning from pre-trained protein language models through an SE(3)-equivariant graph neural network [2]. When trained and evaluated on the Protein Data Bank, PETIMOT shows superior performance in capturing protein dynamics, particularly large/slow conformational changes, compared to state-of-the-art flow-matching approaches and traditional physics-based models. We have also explored whether continuous compact representations of protein motions could be predicted directly from sequences, without exploiting 3D structures. Our sequence-only method, SeaMoon, leverages protein Language Model (pLM) embeddings as input to a lightweight convolutional neural network [3]. When assessed against ~ 1 000 collections of experimental conformations exhibiting diverse motions, SeaMoon predicts at least one ground-truth motion with reasonable accuracy for 40% of the test proteins.

I will also present our recent work on reconstructing protein motions from Atomic Force Microscopy (AFM) images. Indeed, AFM offers a unique opportunity to study the conformational dynamics of proteins in near-physiological conditions at the single-molecule level. However, interpreting the two-dimensional molecular surfaces of multiple molecules measured in AFM experiments as 3D conformational dynamics poses a significant challenge. To address this, we developed AFMfit, a flexible fitting procedure that deforms an input atomic model to match multiple AFM observations [4]. The procedure results in a set of fitted atomic models that form a conformational ensemble describing the AFM experiment. Our method uses a new fast fitting algorithm based on the nonlinear Normal Mode Analysis (NMA) method NOLB to associate each molecule with its conformational state. AFMfit processes conformations of hundreds of AFM images of a single molecule in a few minutes on a single workstation, enabling analysis of larger datasets, including high-speed (HS)-AFM. I will demonstrate the applicability of our method to the membrane-embedded transient receptor potential channel TRPV3 and bound and unbound forms of alpha-actinin [5].

Finally, one more way to computationally explore protein dynamics is to analyze 3D Voronoi tessellations of the snapshots from conformational ensembles. Figure 1 shows Voronota-LT derived tessellations applied to an NMR protein ensemble [6].



**Figure 1.** An NMR ensemble of 20 conformational states taken from the Protein Data Bank entry 1CIR (chymotrypsin inhibitor 2). We computed the contact areas and, for every unique contact, we calculated the mean and the variance values of the corresponding observed area distribution. Inter-chain contact surfaces are shown colored by the corresponding variance-to-mean ratios (relative variances) that indicate how much every contact area is dispersed relative to its mean. 3D visualizations were generated using Voronota-GL.

[1] Lombard V., S. Grudin, and E. Laine (2024). *Sci. Data* **11**:752. doi: 10.1038/s41597-024-03524-5.

[2] Lombard V., S. Grudin, and E. Laine (2025). In *Learning Meaningful Representations of Life workshop at the thirteenth International Conference on Learning Representations (ICLR), Singapore, SG*.

[3] Lombard V., D. Timsit, S. Grudin, and E. Laine (2025). *Structure, in Press*. doi: 10.1101/2024.09.23.614585.

[4] Vuillemot, R., Pellequer, J. L., & Grudin, S. (2025). *Communications Biology*, in Press.

[5] Ngo, K. X., Sumikama, T., Vuillemot, R., Nguyen, H. G., Uyeda, T. Q., Le, N. T. P., & Grudin, S. (2025). *bioRxiv*, 2025-04.

[6] Olechnovič, K., & Grudin, S. (2025). *Journal of Computational Chemistry*, in Press.