Geometric Machine Learning for Enhanced Modeling of Elongation and Theoretical Inquiries

Grégoire Sergeant-Perthuis*
Elodie Laine †

The ribosome is responsible for protein synthesis (translation). Bacterial ribosomes are among the biomolecules targeted for antibiotic design. The ribosome is composed of ribosomal RNA (rRNA) and ribosomal proteins (r-proteins), and its movement along the mRNA, known as elongation, enables the sequential attachment of amino acids to the protein undergoing biosynthesis. The current view is that most of the ribosome's functions are accomplished by rRNA, the role of r-proteins being essentially thought of in terms of stabilization and assembly [1–3]. However, this view has been challenged by recent work [4–6], which highlights the role of r-proteins in regulating the movements during elongation. Focusing on r-proteins rather than rRNA, and vice versa, changes the qualitative description of the ribosome's movements. A quantitative understanding of the conformational changes of the ribosome during elongation and of what drives these movements would help settle the debate, depending on the level of granularity one wants in the description of the movements. Molecular dynamics simulations theoretically offer a fine-grained characterization of these movements but are limited in practice in their ability to sample large conformational changes of the ribosome due to its size (number of atoms), structural complexity, and the wide range of time scales of the dynamics that are relevant to translation [7, 8]. Furthermore, as exemplified by the debate on the role of rRNA and r-proteins, an a priori focus on one or the other should be accounted for in simulations to allow biasing the sampling of certain movements rather than others and exacerbate differences. In this project, we propose to speed up molecular dynamics simulations of the ribosome during translation by incorporating a wide range of prior information that we want to account for.

A growing amount of work in machine learning and deep learning focuses on exploiting geometric and topological structures of signals and models to improve processing and to further the theoretical understanding of such processing (see geometric deep learning [9, 10], sheaves in data science [11–19],

*CQSB, Sorbonne Université gregoireserper@gmail.com
†CQSB, Sorbonne Université elodie.laine@sorbonne-universite.fr

and categorical approaches in machine learning and probability, specifically for theoretical inquiries [18, 20–27]). In particular, generalizations of graph neural networks (GNNs), prominent in learning on biomolecules such as proteins, to heterogeneous structures have gained recent focus [11–14] and have shown promising results for better handling of heterogeneous signals over graphs. In [17, 19], the supervisors of this proposal extended the framework of graphical models and Markov Random Fields (MRFs), which are central for inference of energies in interacting systems, to allow the fusion of heterogeneous descriptions of signals that incorporate dynamic and geometric priors and proposed a message-passing algorithm for inference in such models. Both approaches, the generalization of GNNs and of MRFs, rely on introducing geometric-combinatorial objects—specifically, sheaves over partially ordered sets (see for sheaf theory [28], cellular sheaves [11], and sheaves in optimization [17, 19])—to account for a larger range of geometric and topological prior knowledge on signals. Topological prior is understood in a broad sense, where topology can arise from different perspectives or 'views' on the molecule and the way these views interact. An example of this is when these perspectives arise from coarse-grained models of specific regions of the molecule that interact through overlapping sub-parts. Rather than introducing a coarse-graining of the entire biomolecule, one allows for more flexibility in the descriptions of these regions and their junctions with other regions by considering all these descriptions collectively to reconstruct plausible conformations of the whole biomolecule.

Leveraging the combined expertise of the supervisors in the molecular dynamics and geometric machine learning, we aim, first, to exploit recent methodological advances in geometric machine learning and deep learning to accelerate and enhance these simulations, in particular for elongation. This will be achieved by interpolating high-resolution snapshots of the ribosome during key transitions [29] while incorporating diverse geometric and topological priors. Prior knowledge of small molecules has proven effective in accelerating molecular dynamics [30]. The objective of this PhD will be to benchmark architectures inspired by geometric machine learning and deep learning for accelerating molecular dynamics, with a focus on elongation. The second aim of the project is to provide results in quantifying the impact of the topology of these sheaves on the quality of inference and learning, and towards a classification of different critical points of the associated losses in order to obtain better interpretability of inferred or learned dynamics (see [31]).

The PhD student will be co-supervised by Grégoire Sergeant-Perthuis and Élodie Laine, the research will take place at CQSB, Sorbonne Université. The PhD student will also benefit from discussions with Youri Timsit (MIO) and Daniel Bennequin (Université Paris-Cité).

References

[1] H. F. Noller, L. Lancaster, S. Mohan, and J. Zhou, "Ribosome structural dynamics in translocation: yet another functional role for ribosomal rna," *Quarterly Reviews of Biophysics*, vol. 50, p. e12, 2017.

[2] T. A. Steitz, "A structural understanding of the dynamic ribosome machine," *Nature reviews Molecular cell biology*, vol. 9, no. 3, pp. 242–253, 2008.

[3] D. N. Wilson and K. H. Nierhaus, "Ribosomal proteins in the spotlight," *Critical reviews in biochemistry and molecular biology*, vol. 40, no. 5, pp. 243–267, 2005.

[4] Y. Timsit and S.-P. Grégoire, "Towards the idea of molecular brains," *International journal of molecular sciences*, vol. 22, no. 21, p. 11868, 2021.

[5] Y. Timsit, G. Sergeant-Perthuis, and D. Bennequin, "Evolution of ribosomal protein network architectures," *Scientific reports*, vol. 11, no. 1, p. 625, 2021.

[6] Y. Timsit, G. Sergeant-Perthuis, and D. Bennequin, "The role of ribosomal protein networks in ribosome dynamics," *Nucleic Acids Research*, vol. 53, no. 1, p. gkae1308, 2025.

[7] L. V. Bock, M. H. Kolář, and H. Grubmüller, "Molecular simulations of the ribosome and associated translation factors," *Current opinion in structural biology*, vol. 49, pp. 27–35, 2018.

[8] C. Gupta, D. Sarkar, D. P. Tieleman, and A. Singharoy, "The ugly, bad, and good stories of large-scale biomolecular simulations," *Current Opinion in Structural Biology*, vol. 73, p. 102338, 2022.

[9] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric deep learning: Grids, groups, graphs, geodesics, and gauges," 2021.

[10] G. Sergeant-Perthuis, J. Maier, J. Bruna, and E. Oyallon, "On non-linear operators for geometric deep learning," in *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds.), vol. 35, pp. 10984–10995, Curran Associates, Inc., 2022.

[11] J. Curry, *Sheaves, cosheaves and applications*. PhD thesis, The University of Pennsylvania, 2013. arXiv:1303.3255.

[12] J. Hansen and R. Ghrist, "Distributed optimization with sheaf homological constraints," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 565–571, 2019.

[13] J. Hansen and R. Ghrist, "Toward a spectral theory of cellular sheaves," *Journal of Applied and Computational Topology*, vol. 3, pp. 315 – 358, 2018.

[14] C. Bodnar, F. D. Giovanni, B. P. Chamberlain, P. Lio, and M. M. Bronstein, "Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs," in *Advances in Neural Information Processing Systems* (A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, eds.), 2022.

[15] O. Peltre, "Message passing algorithms and homology," 2020. Ph.D. thesis, Link.

[16] O. Peltre, "A homological approach to belief propagation and bethe approximations," in *Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds.), (Cham), pp. 218–227, Springer International Publishing, 2019.

[17] G. Sergeant-Perthuis, "Regionalized optimization," 2022. `https://arxiv.org/abs/2201.11876`.

[18] G. Sergeant-Perthuis, "Compositional statistical mechanics, entropy and variational free energy (Extended abstract)," in *7th International Conference on Applied Category Theory (ACT 7)*, (Oxford, UK, France), David Jaz Myers and Michael Johnson, June 2024.

[19] G. Sergeant-Perthuis and N. Ruet, "Inference on diagrams in the category of Markov kernels (Extended abstract)," in *7th International Conference on Applied Category Theory (ACT 7)*, (Oxford (UK), United Kingdom), David Jaz Myers and Michael Johnso, June 2024.

[20] T. Fritz, "A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics," *Advances in Mathematics*, vol. 370, p. 107239, 2020.

[21] T. Fritz and P. Perrone, "A probability monad as the colimit of spaces of finite samples," *Theory and Applications of Categories*, 2019.

[22] P. Baudot and D. Bennequin, "The homological nature of entropy," *Entropy*, vol. 17, no. 5, pp. 3253–3318, 2015.

[23] Y. Jia, G. Peng, Z. Yang, and T. Chen, "Category-theoretical and topos-theoretical frameworks in machine learning: A survey," 2024.

[24] J. P. Vigneaux, "Information structures and their cohomology," *Theory and Applications of Categories*, vol. 35, no. 38, pp. 1476–1529, 2020.

[25] M. Marcolli, "Gamma spaces and information," *Journal of Geometry and Physics*, 2019.

[26] Y. Manin and M. Marcolli, "Homotopy Theoretic and Categorical Models of Neural Information Networks," *Compositionality*, vol. Volume 6 (2024), Sept. 2024.

[27] G. Sergeant-Perthuis, "Compositional statistical mechanics, entropy and variational inference," in *Twelfth Symposium on Compositional Structures (SYCO 12)*, 2024.

[28] J. Gallier and J. Quaintance, *Homology, Cohomology, and Sheaf Cohomology for Algebraic Topology, Algebraic Geometry, and Differential Geometry*. World Scientific, 2022.

[29] J. P. Rickgauer, H. Choi, A. S. Moore, W. Denk, and J. Lippincott-Schwartz, "Structural dynamics of human ribosomes in situ reconstructed by exhaustive high-resolution template matching," *Molecular Cell*, vol. 84, no. 24, pp. 4912–4928.e7, 2024.

[30] Z. Smith and P. Tiwary, "Making high-dimensional molecular distribution functions tractable through belief propagation on factor graphs," *The Journal of Physical Chemistry B*, vol. 125, no. 40, pp. 11150–11158, 2021.

[31] G. Sergeant-Perthuis, "Functoriality of inference on diagrams in the category of Markov kernels.," in *Non Commutative Geometry & Higher Structures*, (Thessaloniki, Greece), P. Batakidis and F. Bonechi and A. S. Cattaneo and N. Ciccoli and F. D' Andrea and D. Fiorenza and M. Jotz and F. Petalidou and M. Schiavina and P. Xu, Sept. 2024. `https://hal.sorbonne-universite.fr/hal-04706267v2/file/hal-presentation.pdf`.