

## Master 2 internship proposal

### Distance metrics for the analysis of 3D shape variability in 2D cryo electron microscopy images

**Supervisor:** Slavica Jonic, Research Director CNRS, IMPMC – UMR 7590 CNRS, Sorbonne Université, MNHN, IRD, 4 Place Jussieu, 75005 Paris ([slavica.jonic@upmc.fr](mailto:slavica.jonic@upmc.fr), <http://www.impmc.upmc.fr/~jonic>)

**Co-supervisor :** Jean Feydy, Research Fellow at INRIA, HeKA team - INRIA Paris, INSERM, Université Paris-Cité, Paris-Santé Campus, 2 - 10 Rue d'Oradour-sur-Glane, 75015 Paris ([jean.feydy@inria.fr](mailto:jean.feydy@inria.fr), <http://jeanfeudy.com>)

This Master 2 internship targets the **reconstruction of the manifold of 3D conformations of a biomolecule from a large collection of 2D images** acquired using cryo electron microscopy (cryo-EM). This project will leverage geometric machine learning tools and can lead to a PhD thesis.

The 3D structures of proteins and protein complexes are directly linked to their biological functions. In order to accomplish various tasks, proteins undergo changes of their 3D structures, which are referred to as conformational changes. **There is significant interest in determining the full conformational landscapes (manifolds) of protein complexes, which should help us to understand their mechanisms of action in health and disease.** But unfortunately, fast, robust and user-friendly structural biology methods to extract this information are still lacking.

Cryo-EM is currently the mainstream data collection technique for structural biology. As illustrated in the picture below, it enables the observation of protein structures present in a purified sample at an atomic scale: in as little time as 1-2 days, a cryo-EM device can acquire millions of noisy 2D images at a resolution of approximately  $1 \cdot 10^{-10}$  m, each corresponding to an independent projection of a protein at a random and unknown orientation. **Understanding the intrinsic, low-dimensional structure of such a dataset (e.g.,  $N = 1 \cdot 10^6$  noisy images with a size of  $200 \times 200$  pixels) is key to reconstruct the average 3D conformation of a biomolecule as well as its main modes of variations.**

This internship aims at developing the foundational building blocks of a geometric method to tackle this problem. As a first step, the intern will study several **distances between 3D shapes** that can be used with methods such as UMAP to reconstruct a continuous conformational manifold from a set of discrete poses, and benchmark them on both synthetic and experimental 3D data. We will notably study kernel metrics, the Wasserstein distance that is induced by optimal transport and some shape metrics from the computer graphics and medical imaging literatures. Then, the intern will **interface these 3D metrics with structure estimation methods for 2D cryo-EM images:** the aim will be to increase the reliability of these (state-of-the-art) methods on challenging biomolecules that exhibit a wide, continuous range of conformations. Finally, time permitting, the intern may start thinking about an **end-to-end model** to learn a conformational manifold directly from a set of cryo-EM images.

**Duration:** Up to 6 months.

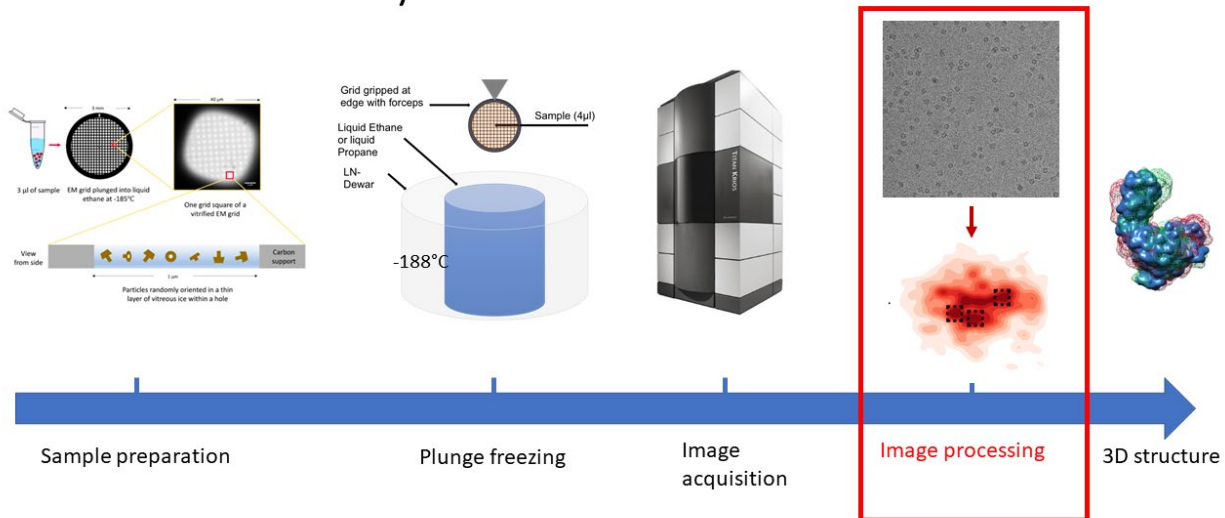
**Gratification:** ≈600€/month (standard for internships in French public research labs)

**Location:** The internship student will be located at the IMPMC, Tour 23, Campus Jussieu (Sorbonne Université, 4 Place Jussieu, 75005 Paris).

**PhD thesis after the internship:** This work could be continued in the framework of a PhD thesis.

**Contact for applications and required documents:** The applications should be sent to [slavica.jonic@upmc.fr](mailto:slavica.jonic@upmc.fr) and [jean.feydy@inria.fr](mailto:jean.feydy@inria.fr). They should include a CV, a motivation letter, and grade transcripts.

## Cryo-EM workflow



### References in relation with the project:

1. Sverrisson, F., **Feydy, J.**, Correia, B.E. and Bronstein, M.M. (2021). Fast end-to-end learning on protein surfaces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15272-15281), CVPR 2021. DOI: 10.1109/CVPR46437.2021.01502 [Link](#)
2. **Feydy, J.**, Glaunès, A., Charlier, B. and Bronstein, M., (2020). Fast geometric learning with symbolic matrices. Advances in Neural Information Processing Systems, 33, pp.14448-14462. NeurIPS 2020. [Link](#)
3. Shen, Z., **Feydy, J.**, Liu, P., Curiale, A.H., San Jose Estepar, R., San Jose Estepar, R. and Niethammer, M. (2021). Accurate Point Cloud Registration with Robust Optimal Transport. Advances in Neural Information Processing Systems, 34, pp.5373-5389. NeurIPS 2021. [Link](#)
4. **Feydy J.** Geometric data analysis, beyond convolutions. PhD thesis, 2020. Université Paris-Saclay. [Link](#)
5. Hamitouche, I., and **Jonic, S.** (2022). DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. Front Mol Biosci 9, 965645. DOI: 10.3389/fmolb.2022.965645. [Link](#)
6. Harastani, M., Eltsov, M., Leforestier, A., and **Jonic, S.** (2022). TomoFlow: Analysis of Continuous Conformational Variability of Macromolecules in Cryogenic Subtomograms based on 3D Dense Optical Flow. J Mol Biol 434, 167381. DOI: 10.1016/j.jmb.2021.167381. [Link](#)
7. **Jonic, S.** (2017). Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. Curr Opin Struct Biol 43, 114-121. DOI: 10.1016/j.sbi.2016.12.011. [Link](#)
8. Harastani, M., Vuillemot, R., Hamitouche, I., Moghadam, N.B., and **Jonic, S.** (2022). ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. J Struct Biol 214, 107906. DOI: 10.1016/j.jsb.2022.107906. [Link](#)