Postdoctoral position at the interface of machine learning, computational biology and statistical physics, in Paris-Saclay

An 18 months postdoc position is available from March 2025, to work with Jorge FERNANDEZ-DE-COSSIO-DIAZ (CEA, IPhT) and Martin WEIGT (Sorbonne University). The project aims at using methods based in statistical physics and machine learning / artificial intelligence to unveil signatures of protein functional specificity encoded in amino acid sequences. While many standard modeling approaches of protein sequences are based on multiple-sequence alignments (MSA), functional specificity is frequently encoded in hardly-alignable regions of variable length, and therefore discarded in standard sequence models. The goal of our project is to develop principled approaches able to handle variable sequence length and amino-acid insertions, and to relate them to protein function. We will, in particular, explore hierarchical models for sequence specific insertions, and protein language models.

We are looking for candidates with a strong quantitative background in machine learning, statistical physics or bioinformatics. We expect a strong motivation for interdisciplinary work; prior experience in biology is not required but considered a plus.

If you are interested, please contact jorge.fdcd@ipht.fr and martin.weigt@sorbonne-universite.fr by email.

An 18 months postdoc position is available from March 2025, to work with Jorge FERNANDEZ-DE-COSSIO-DIAZ and Martin WEIGT. Homology-aware models of protein sequences are often trained on multiple-sequence alignments (MSA), grouping together evolutionarily related proteins that perform closely related functions. Members of a family can exhibit varying degrees of specificity, e.g., favoring particular molecular interactions over other similar ones. Specificity-determining amino-acids often sit in loops, flexible structural elements of variable length, treated as insertions in MSAs and often discarded by downstream models. The goal of this project is to develop new generative modelling approaches that can handle insertions in a more principled manner. The candidate will have the opportunity to work on topics related to machine learning, bioinformatics, and statistical physics. We will explore state-of-the art transformer-based models.