Proposed Final Work for the Master in Biophysics UAM 2023-2024
Tutor: Dr. Ugo Bastolla, Centro de Biologia Molecular Severo Ochoa (CSIC-UAM)
Improving phylogenetic inference of protein evolution through a structurally constrained model of protein evolution based on torsional normal modes

Phylogenetic inference is essential for reconstructing the history of life and discovering its mechanisms. Nevertheless, most mathematical models of protein evolution that perform this inference do not consider the protein structure, and they are based on the unrealistic assumption that different protein sites evolve independently of each other, despite it is known that their interactions are crucial for maintaining the stability of the native state and its functional dynamics.

Some years ago, we introduced a structure-aware model of protein evolution that computes site-specific substitution processes [1] based on the **stability-constrained model of protein evolution (SCPE)** [2], which imposes global constraint of the conservation of the folding stability of the native state. This model maintains the computational simplicity that different sites evolve independently of each other, and we used it for the computational reconstruction of ancestral sequences [3]. However, this model is still too tolerant to mutations and it overestimates the sequence entropy and substitution rates of protein families [4], probably because it does not consider the structural changes that happen in protein evolution [5].
Some years ago, Julian Echave proposed how to predict such structural changes through the linear response of the elastic network model [6] that approximates the native state of the protein. Based on his proposal, we developed a model that adopts the elastic network model in torsion angle space (TNM) developed in our group [7,8], in order to predict how a given amino acid mutation perturbs the protein structure. We used this model for predicting the fitness consequences of mutations that perturb the native structure, obtaining the **Structure and stability constrained protein evolution (SSCPE) model** [9]. The SSCPE improves our previous model based only on folding stability, it is less tolerant to mutations, and it reproduces well the observed dependence of site-specific substitution rates on the structural properties of the site in the native state.

Despite their success, it is not clear that structure-based models are able to infer correct phylogenetic trees. The main problem is that the maximum likelihood method for tree inference is biased, if the a priori distribution of trees is not uniform. The first proposed work will be dedicated to studying the biases of phylogenetic trees inferred through maximum likelihood, using simulated trees and simulated multiple alignments, and investigating the possibility to infer less biased trees through the Neighbor Joining method based on the SSCPE model.
A second proposed work, more oriented towards bioinformatics, will consists in adopting the SSCPE model for obtaining site-specific protein alignments and test whether they improve the ability to model the protein structure through homology.
A third possible work will consist in developing and testing an Ornstein-Uhlenbeck model of protein structure evolution based on the linear response of the TNM model that we adopted for the SSCPE model.

The work will be carried out in the Center of Molecular Biology “Severo Ochoa” in the UAM campus but, due to the current circumstances, it will take place mostly remotely.

References
1) Arenas M, Sánchez-Cobos A, Bastolla U (2015) Maximum-Likelihood Phylogenetic Inference

with Selection on Protein Folding Stability. Mol Biol Evol. 32:2195-207.
2) Bastolla U, Dehouck Y, Echave J. (2017) What evolution tells us about protein physics, and protein physics tells us about evolution. Curr Opin Struct Biol. 42:59-66. Review.
3) Arenas M, Weber CC, Liberles DA, Bastolla U (2017) ProtASR: An Evolutionary Framework for Ancestral Protein Reconstruction with Selection on Folding Stability. Syst Biol. 66:1054-1064.
4) Jimenez MJ, Arenas M, Bastolla U (2017) Substitution rates predicted by stability-constrained models of protein evolution are not consistent with empirical data. Mol Biol Evol. 2017
5) Pascual-García A, Arenas M, Bastolla U (2019) The Molecular Clock in the Evolution of Protein Structures. Syst Biol. 2019
6) Echave J, Fernández FM (2010) A perturbative view of protein structural variation. Proteins. 78:173-80.
7) Mendez R, Bastolla U (2010) Torsional network model: normal modes in torsion angle space better correlate with conformation changes in proteins. Phys Rev Lett. 104:228103.
8) Alfayate A, Caceres CR, Gomes Dos Santos H, Bastolla U (2019) Predicted dynamical couplings of protein residues characterize catalysis, transport and allostery. Bioinformatics.
doi: 10.1093/bioinformatics/btz301.
9) Lorca I, Arenas M, Bastolla U (2023) Structure- and stability- constrained protein evolution model (SSCPE) improves phylogenetic inference through biophysical properties of the protein native state. Systematic biology, under review.