Proposed TFM (Trabajo Final de Master) in Biophysics, year 2021-2022

Protein structure evolution and protein structure aware substitution models for phylogenetic inference.

Tutor: Ugo Bastolla ubastolla@cbm.csic.es

Proteins evolve under the constraint of functional conservation throughout most of their natural history, except in key moments of functional innovation. Some years ago, we proposed a measure of protein structure divergence that strongly correlates with sequence divergence, and showed that structure divergence is slower than sequence divergence, which suggests that protein structure is more constrained by negative natural selection than other sequence-dependent properties like protein folding stability. More recently, we verified that protein structure evolution follows an approximate molecular clock during function conservation, so that it can be used to infer phylogenetic trees, but undergo evolutionary accelerations that violate the molecular clock when the function changes, which also implies stronger positive selection acting on protein structures than on sequences (Pascual-García, Arenas, Bastolla. The Molecular Clock in the Evolution of Protein Structures. Syst Biol. 2019 68:987-1002).

The proposed work can follow one of two orthogonal lines:

(1) Test of a novel method for inferring phylogenetic trees of protein superfamilies (proteins with common origin and similar structure but different function) through our structure divergence measure and the Neighbor joining method (Saitou and Nei, 1985) for reconstructing phylogenetic trees.

(2) Test of a novel amino acid substitution process with selection on structure conservation that can be applied to phylogenetic inference through the maximum likelihood method. We recently developed an amino acid substitution process with selection on protein folding stability and approximations that allow its use for phylogenetic inference. However, this model computes protein stability under the simplifying assumption that protein structure does not change upon mutation, and therefore it cannot constraint structure conservation, and it tolerates too many mutations. We later developed a method for predicting the structural effect of a mutation based on the linearly forced elastic network model proposed by Julian Echave, adopting our own elastic network model in torsion angle coordinates (torsional network model) and a new mapping of the structural perturbations induced by amino acid mutations. The fitness function predicted from this model on proteins with known structure will be implemented by the tutor into the stability constrained amino acid substitution process, and the predictions of the resulting substitution model will be tested by the student against observations derived from multiple sequence alignments that represent the natural evolutionary process of a protein family.

The study will use the computational infrastructure of the Centro de Biologia Molecular Severo Ochoa (CBMSO) in the UAM campus but, due to the Covid-19 pandemics, part of the work may take place remotely.