Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Juan Cortés
Email	juan.cortes@laas.fr
Institution	LAAS-CNRS
Website	https://www.laas.fr/public/en
Group	Robotics and Interactions

Project

Structural bioinformatics

Project Title:

Studying how protein evolution (re-)shapes local structural preferences

Keywords:

Protein evolution, local structural preferences, structural database, protein flexibility, intrinsically disordered proteins

Summary:

The structural and dynamical properties of a protein are largely determined by its sequence, and strongly influence its function. To maintain protein functions during evolution, these properties must be robust to sequence variations (for example mutations). Nevertheless, structural/dynamical changes may be beneficial, for instance if they improve the way the protein functions or generate functional innovations. Understanding this trade-off between structural stability and malleability in evolution is of great interest for fundamental biology and for protein design. In recent years, we have contributed to better characterize the sequence-structure/dynamics relationship, particularly in the context of highly-flexible protein regions. We have constructed an extensive database of small fragments involving three consecutive residues (called tripeptides) extracted from coil regions in experimentally-determined protein structures. We have shown that this database is useful to accurately sample the conformational variability of protein loops [1] and intrinsically disordered proteins (IDPs) [2]. We have also developed an approach to characterize the structural preferences of each tripeptide sequence, and we have defined metrics to quantify the structural differences between different sequences. The goal of this project is to investigate how mutations occurring during evolution affect protein local structural propensities. To do so, the candidate will exploit our structural database and use the developed approaches and metrics. Starting from a set of currently observed proteins, s/he will infer the evolutionary history relating them and will quantify the correlation between changes in sequence and changes in local structures. We will consider several protein families. The analysis will be particularly focused on proteins in which (local or global) flexibility plays essential roles, such as antibodies, enzymes and IDPs.

References:

[1] Barozet, A., Molloy, K., Vaisset, M., Simeon, T., Cortés, J. (2020) A reinforcement learning approach to enhance protein loop sampling. Bioinformatics, 36(4):1099–1106 [2] Estana, A., Sibille, N., Delaforge, E., Vaisset, M., Cortés, J., Bernado, P. (2019) Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. Structure, 27(2):381-391.E2

Expected skills::

The candidate should have: background in bioinformatics and structural biology, good programming skills, familiarity with Linux, and knowledge about databases. Teamwork skills are also essential for the achievement of the project.

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

The project will be co-supervised by Elodie Laine (Sorbonne Université, Paris)