Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2024-2025

Personal Information

Supervisor	R. Gonzalo Parra
Email	gonzalo.parra@bsc.es
Institution	Barcelona Supercomputing Center
Website	https://www.bsc.es/parra-rodrigo-gonzalo
Group	Computational Biology Group

Project

Structural bioinformatics

Project Title:

Evolution of Energetic Constraints from Ancestral Proteins to Extant Families

Keywords:

protein function, protein evolution, networks, statistics, biophysics

Summary:

Natural proteins fold by minimizing their internal energetic conflicts satisfying the “minimum frustration principle” 1. However, we can find that around 10-15% of the interactions between residues in the native state of proteins are in energetic conflict with their local environment 2,3. These conflictive interactions, a.k.a frustrated interactions, maintained over evolutionary timescales, are not artifacts but a functional, adaptive response. These energetic conflicts allow proteins to interconvert between the different conformers that constitute their native state and hence perform their function. Local frustration has been linked to many different functional aspects of proteins such as protein-protein interactions 2,3, allosterism 4, order-disorder transitions 5 or catalytic activity 6, among others. A method to localize and quantify local frustrated interactions has been publicly available as a web-server for a few years 7 and has recently been implemented as an R package 8. In parallel, we have recently developed FrustraEvo 9, a tool that measures the degree of conservation of local frustration within proteins that belong to the same protein family. The basic idea is that if a given energetic feature, either minimized or in conflict, is conserved across all family members it may hint to an evolutionary conserved functional requirement of the family. Additionally, differences between evolutionary related subfamilies may refer to specific adaptations that have occurred within each lineage since they started to diverge from their common ancestor. In our recent publication 9 we have applied FrustraEvo to a few examples that were important to show the conceptual value of the designed strategy. To better understand the emergence of the energetic patterns and their functional consequences we propose to analyze how these signatures emerge over evolutionary timescales. We will focus on specific protein families such as the beta lactamases because of their importance in antibiotic resistance and hemoglobins as models for protein-protein interactions. For these, we will reconstruct ancestral protein sequences at different timepoints in the past, using state-of-the-art methodologies, and compare them to the extant protein families instances. We will obtain a time dependent understanding of how energetic constraints arose over time fulfilling different stability and functional requirements. This project will allow the student to learn how to use different state-of-the-art Artificial Intelligence (AI) and other types of computational techniques such as protein structure prediction with AlphaFold2, local energetic frustration calculations, homologs searching, multiple sequence alignments, ancestral sequence reconstructions, Large Language Models to generate low dimensional embeddings among others. The candidate will be working in collaboration with PhD students and postdocs that are contributing different types of complementary expertise to the overall project. Dr. R. Gonzalo Parra is an expert in protein biophysics (structures, folding, function) and computational tools development. A summary of his research can be found at: https://scholar.google.com/citations?user=TRGamd0AAAAJ&hl=en. In addition, both Dr. Parra and Prof. Valencia are highly connected to the leadership of different international organizations in the field of bioinformatics and therefore have collaborations to ensure the success of the project. The candidate will work in a highly sophisticated HPC environment, will have access to systems and computational infrastructures.

References:

1. Bryngelson, J. D. & Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. Proc. Natl. Acad. Sci. U. S. A. 84, 7524–7528 (1987). 2. Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. Localizing frustration in native proteins and

protein assemblies. Proc. Natl. Acad. Sci. U. S. A. 104, 19819–19824 (2007). 3. Parra, R. G., Espada, R., Verstraete, N. & Ferreiro, D. U. Structural and Energetic Characterization of the Ankyrin Repeat Protein Family. PLoS Comput. Biol. 11, e1004659 (2015). 4. Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. On the role of frustration in the energy landscapes of allosteric proteins. Proc. Natl. Acad. Sci. U. S. A. 108, 3499–3503 (2011). 5. Freiberger, M. I., Wolynes, P. G., Ferreiro, D. U. & Fuxreiter, M. Frustration in Fuzzy Protein Complexes Leads to Interaction Versatility. J. Phys. Chem. B 125, 2513–2520 (2021). 6. Freiberger, M. I., Guzovsky, A. B., Wolynes, P. G., Parra, R. G. & Ferreiro, D. U. Local frustration around enzyme active sites. Proc. Natl. Acad. Sci. U. S. A. 116, 4037–4043 (2019). 7. Parra, R. G. et al. Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. Nucleic Acids Res. 44, W356–60 (2016). 8. Rausch, A. O. et al. FrustratometeR: an R-package to compute local frustration in protein structures, point mutants and MD simulations. Bioinformatics (2021) doi:10.1093/bioinformatics/btab176. 9. Freiberger, M. I. et al. Local energetic frustration conservation in protein families and superfamilies. Nat. Commun. 14, 8379 (2023).

Expected skills:

Skills required: - Background in molecular biology and proteins - Basic computational skills (R, python and bash programming) - Basic statistical knowledge - Ability to access and evaluate scientific literature - Fluency in spoken and written English. - Good communication and presentation skills.

Possibility of funding:

Yes

Possible continuity with PhD:

To be discussed