



Master project 2024-2025

Personal Information

Supervisor	Ben Lehner
Email	albert.escobedo@crg.eu
Institution	CRG
Website	https://www.crg.eu/ben_lehner
Group	Genetic Systems

Project

Computational systems biology

Project Title:

A complete model of a protein-protein interaction: evolutionary biophysics through deep sequence space

Keywords:

protein-protein interactions, biophysics, thermodynamics, modeling, deep learning

Summary:

The ability of proteins to interact with each other is a major feature defining their function. The interfaces involved in protein-protein interactions (PPIs) are largely fine-tuned to provide the right affinity and specificity for a particular process to take place. Mutations leading to aberrant PPIs are often involved in disease, and the ability to manipulate PPIs by editing protein sequences offers the possibility to engineer biomolecular systems. Thus, modeling PPIs for accurate genetic prediction of the effects of mutations has vast applications in disease variant classification, drug development for precision medicine and bioengineering. Leucine zippers are a class of coiled-coils, one of the most abundant PPI modules in nature, and also one of the simplest. They consist of two amphipathic alpha-helices coiled to one another, each presenting a periodic structure every heptad (every 7 residues), with 2 apolar residues in each helix at the interaction interface core (one often a leucine, thus the name), 2 polar residues at the interaction boundaries and 3 residues at the opposite side of the interaction. Several positions (particularly those at the core and the boundaries) are involved in inter-helix contacts that are key for the interaction, while others are structurally relevant. Using deep mutational scanning (DMS) on several combinatorial variant libraries each sampling a set of key interactions or structural positions, we have explored sequence evolution by measuring the effects that hundreds of thousands of mutations have on the leucine zipper PPI established by the proto-oncogenes FOS and JUN [1]. This allows us to infer the energetic effects associated with each mutation, which is key for accurate modeling and genetic prediction [2,3]. Using the neural network MoCHI [4], the student will aim at integrating all these datasets into one single complete model of the FOS-JUN interaction. She/he will explore how much data is required to build a fully predictive model of mutational effects in a PPI by testing how well smaller models trained on subsets of data predict increasingly distant mutational effects: how well does a model trained on mutational data in one heptad predict mutational effects in another heptad of the same leucine zipper? How well does a model trained on mutational data in one zipper (i.e. FOS-JUN) predict the mutational effects in a different zipper (i.e. MYC-MAX)? The performance of these models trained on large DMS datasets will be compared to zero-shot protein large language models (LLMs), so that the student will evaluate the advantages of fitting thermodynamic models on datasets that capture the underlying biophysics of mutational effects as compared to sequence co-evolution information alone. This project is offered as either fully computational or combining both wet and dry labs depending on the student's interests.

References:

[1] The genetic landscape of a physical interaction. Diss & Lehner, eLife 2018. [2] Mapping the energetic and allosteric landscapes of protein binding domains. Faure, Domingo, Schmiedel, Hidalgo-Carcedo, Diss & Lehner, Nature 2022. [3] The energetic and allosteric landscape for KRAS inhibition. Weng, Faure, Escobedo & Lehner, Nature 2022. [4] MoCHI: neural networks to fit interpretable models and quantify energies, energetic couplings, epistasis and allostery from deep mutational scanning data. Faure & Lehner, bioRxiv 2024.

Expected skills:

python, machine learning, R, statistics

Possibility of funding:

To be discussed

Possible continuity with PhD:

To be discussed

Comments:

The Lehner lab provides a highly collaborative and multidisciplinary atmosphere with opportunities to learn from experts in many different fields. You will take part in the lab routine and benefit from PI and postdoc mentorship along your project. This position will be located at the Lehner lab PRBB site in Barcelona. We work together and have permanent and extensive exchange with the Lehner lab at the Wellcome Sanger Institute at Hinxton, Cambridge, UK. Remuneration and remote work options are available at the CRG and will be explored. This project is offered as either fully computational or combining both wet and dry labs depending on the student's interests.