PROPOSITION DE SUJET DE STAGE DE M2 ET/OU DE THESE

Statistical Protein Evolution

Proteins recapitulate at a molecular scale much of our difficulty to develop a theory of living matter: we know precisely what constitutes a protein (a sequence of amino acids) but we do not generally understand how the many functions that proteins perform (specific binding, catalysis, information transmission…) emerge from the interactions between their constituents.

Nature built proteins through the dynamical process of evolution by natural selection and our approach is to study the relation between protein sequence and function via this process. Our goal is thus to understand **how evolution encodes function into protein sequences**.

To achieve this goal, we combine three complementary approaches: statistical inference from natural and experimental sequence data, high-throughput experiments of protein evolution and theoretical models based on statistical physics.

Our experimental workflow consists in constructing libraries of millions of mutants of an enzyme, which we encapsulate one by one in mono-disperse droplets. The proteins in each droplet are expressed, assayed for enzymatic function using fluorescent assays and sorted into bins that correspond to each level of enzymatic activity. High-throughput sequencing of genes that encode the proteins in each bin yields quantitative information on the relation between sequence (genotype) and function (phenotype).

We are looking for a candidate with a strong background in either physics, mathematics or computer science and a strong interest for biological problems. Prior experiences with molecular biology and micro-fluidics are not required but the candidate should be ready to learn these techniques. Funding is available for a M2 internship and a PhD. The **M2 internship** will focus on learning the experimental workflow. The **PhD work** will combine experiments and data analysis and/or theoretical modeling.

The project will take place in an interdisciplinary team of physicists and biologists, theoreticians and experimentalists, located at Collège de France in Paris.

Keywords: quantitative biology; statistical physics; machine learning; protein evolution; microfluidics

References:

- A. Fallah-Araghi, J.-C. Baret, M. Ryckelynck, A.D. Griffiths (2012). *A completely in vitro ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution.* Lab Chip 12, 882.
- O. Rivoire, K. Reynolds, R. Ranganathan (2016). *Evolution-based functional decomposition of proteins*. PLoS Comput Biol, 12 : e1004817.
- S. Boyer, D. Biswas, A. K. Soshee, N. Scaramozzino, C. Nizak, O. Rivoire (2016). *Hierarchy and extremes in selections from pools of randomized proteins.* PNAS 113, 3482.

Contacts:

Olivier Rivoire olivier.rivoire@college-de-france.fr
Clément Nizak clement.nizak@college-de-france.fr
Statistical Biology Team
Center for Interdisciplinary Research in Biology (CIRB)
Collège de France, 11 place Marcelin Berthelot, 75005 Paris
Webpage: http://statbio.net