In silico comparative function prediction of enzymes, applied to fatty acid metabolism in microalgae

The study of biology associates pattern recognition of diversity with modeling of functional and evolutionary processes. Pleiade addresses the double challenge of measuring dissimilarity between biological objects quickly and precisely, and exploring the relations between diversity in traits and diversity in function at multiple scales. We develop algorithms, models, and software frameworks for applications in ecology, evolution, and biotechnology.

In silico function prediction of enzymes is a key challenge in biotechnology. As the essential link from genome to metabolism, it makes it possible to explain, select, and modify the behavior of existing organisms used as cell factories. In synthetic biology, it can be used inversely, to design new metabolic pathways to introduce into a synthetic organism.

The Inria project teams Pleiade and BioCore, in collaboration with the team of F. Correlou of the CNRS LBM, are developing in silico methods for linking sequence and functional diversity in fatty-acid metabolism of microalgae. The potential of microalgae to produce specific, and even designed, polyunsaturated fatty acids, is the focus of considerable interest in biotechnology for food additives, biofuels, pharmaceutics, and industrial chemistry.

Successful development of these methods will have a direct impact on industrial applications of existing microalgae, as well as laying the groundwork for the design, through synthetic biology, of cell factories that produce specific molecules.

Within the framework of a public partnership with academic partners:

- Inria Project Lab « In silico Algae »
- The objective is to develop predictive methods linking sequence and functional diversity of desaturases, using large-scale comparative analysis of genomic, biochemical, structural, and experimental data.
- Is regular travel foreseen for this post? Quarterly meetings of the IPL. Pleiade and LBM are in Bordeaux; BioCore is near Nice.

**Mission confiée**

Desaturases are key enzymes in fatty acid metabolism that are remarkable for their diversity and their specificity to different substrates. Desaturases are responsible for the biochemical transformation that determines how and where a fatty acid will be (poly)unsaturated. Fatty acids are bonded to head-groups and desaturases can specifically recognize both fatty acids and head groups. Significantly, some parts of the enzyme that recognize the substrate specifically are distinct from the active site that performs the desaturation.

The primary mission is to design and implement in silico methods for large-scale comparative analysis of a large set of candidate desaturase genes from all branches of life, with the goal of defining methods for predicting function, specificity (fatty acid chain), and regiospecificity (head group). The foundation of this mission will be a large-scale, systematic, multi-criteria comparison of candidate proteins, integrating sequences, annotations, structures and experimental results when available. Using classification and pattern recognition methods developed in Inria, predictive methods will be defined. These predictions will be evaluated by biological experts in the CNRS and submitted to experimental validation.

We will then exploit these results to define compact descriptions of functional diversity, either qualitatively or quantitatively. Modeling this diversity will make it possible to establish links between observed sequence diversity and functional diversity. If successful, we can then consider the choice of desaturase as an optimization problem. The feasibility of using these classifications to reduce the search space of in silico protein structure prediction will be assessed.
Principales activités

Main activities:

- Choose, in collaboration with biologists, a panel of in silico methods for comparing candidate gene sequences: sequence motifs and transmembrane regions, physico-chemical properties, predictions of secondary structure
- Systematically apply these comparisons to produce high-dimension distance matrices; analyze these matrices using both unsupervised classification and geometric methods. Pattern recognition and cluster extraction will be applied. Comparison to genes with known specificity, as well as consensus clustering, will be used to refine classification methods
- Submit predictions of selected candidate genes to experimental validation, and potentially to in silico protein structure prediction
- Develop qualitative or quantitative models of desaturase functional diversity, with the goal of linking it to mathematical descriptions of gene sequence diversity

Additional activities:

- Publish results, computer models, and program codes
- Present results to academic partners

Compétences

Technical skills and level required: PhD-level experience in large-scale comparative analyses or machine learning. Ideally bioinformatic analyses, especially comparative genomics or grammatical methods, or less specifically knowledge engineering. Familiarity with mathematical modeling OR structure prediction would be a plus.

Languages: English or French

Relational skills: Must be ready to integrate a pluridisciplinary project with biologists, computer scientists, and mathematicians. The post-doc will work closely with biologists.

Other skills appreciated: Programming (ideally Python, Scala, or C++) and familiarity with computing clusters.

Avantages sociaux

- Subsidised catering service
- Partially-reimbursed public transport

Rémunération

2653€ / month (before taxes)