Protein structure and design: from sequence to function

by the group MASIM

15h-15h15 Overview of the structural Bioinformatics discipline in France. G. André-Leroux *(INRA MaIAGE, Jouy-en-Josas France).*

15h15-15h50 S. Barbe *(Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INRA, INSA, CNRS, Toulouse France).* AI-based Computational Protein Design Methods & Tools to Accelerate the Delivering of Tailor-Made Proteins.
Structure-based Computational Protein Design (CPD) plays a critical role in advancing the field of protein engineering for a wide range of applications, from medicine, biotechnology, and synthetic biology to nanotechnologies. By combining physico-chemical models governing relations between protein amino-acid composition and their 3D structure with optimization algorithms, CPD seeks to identify mutant sequences that fold into a given 3D structure and ultimately possess the targeted biochemical function. Starting from a huge search space, the protein sequence-conformation space, this in silico pre-screening aims to considerably narrow down the number of mutants tested at experimental level while substantially increasing the chances of reaching the desired protein. Herein, we present our methodological advances in the CPD field that enabled overcoming technological bottlenecks and hence propose innovative CPD methods to explore large sequence-conformation spaces while providing more accuracy and robustness than classical approaches. Our CPD methods speed-up search across vast sequence-conformation spaces by several orders of magnitude, find the minimum energy protein design and generate exhaustive lists of near-optimal sequences, defining small mutant libraries. These new methods, in rupture with traditional approaches are based on efficient algorithms issued from recent research in artificial intelligence. In order to make them accessible to a wide range of user profiles in protein engineering, we developed a web-based easy-to-use platform, CUSTOZYME, for the automated design of mutant proteins.

15h50-16h30 B. Bardiaux. *(Structural Unit, Institut Pasteur and CNRS, Paris, France).* Applications of evolutionary contact predictions to integrative structural biology.
Structural prediction methods provide a relatively effective alternative to experimental approaches to provide a first insight into the native fold of a protein. Since the advent of high throughput sequencing technologies, the gap between the number of structures and protein sequences available in databases has steadily increased. This strong growth of genomic information helped bring to light prediction tools using co-evolutionary data. Conservation of a specific function implies strong restraints on interacting residues involved in the folding. Once detected, these interactions can help to model the conformation of a protein. However, some important aspects need to be improved in the 3D modelling process, including the detection of false positive among the predicted contacts and their translation into geometrical restraints. Using our experience in the similarly ill-posed problem of structure determination by Nuclear Magnetic Resonance, we have developed an iterative approach to efficiently predict protein structures de novo, using evolutionary information. Additionally, we will show how this new kind of spatial information can be integrated with other experimental data such as cryo-electron microscopy to build and validate 3D atomic models of macromolecular complexes.

16h30-17h00 Pause café

17h00-17h30 B. Offmann *(Université de Nantes, Unité Fonctionnalité et Ingénierie des Protéines (UFIP) UMR CNRS 6286, Nantes, France).* Structural alphabet structural and intensive functional annotation of proteins by 3D -FORSA.
In the field of protein sciences, computational methods towards fold recognition consists in assessing how well a target amino acid sequence can fit a given protein fold template. FORSA is a such a method that is based on the use of a structural alphabet to represent a fold. A structural alphabet is a limited set of structural prototypes that are recurrently observed locally in protein structures. One such library termed as Protein Blocks (PBs) is composed of a set of 16 structural prototypes each 5-residue long lettered from a to p. As such, any protein fold can be represented in the form a 1D sequence of letters where each letter represents a structural prototype. FORSA proceeds by testing the compatibility of any amino acid sequence with any existing fold represented in the form of a PB sequence. It is based on well-established local or global sequence alignment techniques (Smith-Watermann and Needleman & Wunch) but adapted to the problematic of fold recognition. Because it uses the powerful dynamic programming algorithm, FORSA is as fast as any sequence alignment tool. The basis behind the method and the scoring function of FORSA will be presented as well as example applications in the field of structural and functional annotation of proteins on a genome wide basis, part of which is conducted in collaboration with G. André-Leroux (MIAGE, Jouy-en-Josas).

17h30-18h J. Cortes *(Laboratoire d'Analyse et d'Architecture des Systèmes - CNRS, Toulouse, France).* The importance of local sequence-dependent structural information for modeling intrinsically disordered proteins.

Up to now, protein structure prediction and protein design problems have been mostly formulated assuming that proteins fold into a well-defined three-dimensional form. Nevertheless, there is an increasing corpus of work showing the importance of proteins that do not follow this pattern. They are the so-called Intrinsically Disordered Proteins (IDPs). IDPs are fully functional despite their lack of a permanent secondary or tertiary structure, and they exploit their plasticity to perform highly specialized tasks that are complementary to those of their globular counterparts. Most IDPs are not pure random coils. Very often, IDPs contain short evolutionary-conserved partially-structured fragments that are responsible for partner recognition, and thus, for their function. Malfunction of disordered proteins due to mutations or the dysregulation of homeostatic or post-translational processes can induce severe diseases, such as cancer or neurodegeneration. The structural properties of IDPs are key to decipher the bases of these functional or pathological processes. Nevertheless, structural investigations of IDPs are extremely challenging, and require a tight coupling of experimental and computational methods. In this talk, I will present recent work on IDP modeling in collaboration with P. Bernadó and N. Sibille (CBS, Montpellier). Our approach exploits information extracted from coil regions in high-resolution experimentally-determined protein structures, and shows the importance of local sequence-dependent structural propensities.