3BIO-BIOINFO

Genomic and Structural Bioinformatics Unit

BioModeling, BioInformatics & BioProcesses Department

Pr. M. Rooman et D. Gilis

Subject list

1. Rational modification of the thermal resistance of proteins
2. Prediction of protein solubility changes upon mutation
3. Bioinformatic analysis of punctual variations in the human genome and of their relationship with disease
4. Rational design of vaccines: prediction of B-cell epitopes
5. Modeling of the structure of a membrane receptor and of its interactions with a peptide
6. Study of the dynamics of house dust mite allergens from family 5
7. Origin of the genetic code and prebiotic peptides
8. Dynamical modeling of gene expression in yeast and bacteria subject to perturbations
9. Design of synthetic gene circuits and noise control
10. Development of protein-protein docking algorithms
11. Development of an algorithm for side-chain positioning within proteins

---

1. Rational modification of the thermal resistance of proteins [Yves Dehouck, Marianne Rooman]

Proteins carry out a large variety of biological functions in living organisms. Some proteins, for example, act as particularly specific and efficient catalysts. The possibility to exploit the functional properties of proteins in industrial applications (food industry, chemical processes, pharmaceutical developments, etc.) is extremely interesting; however, one of its major limitations is that proteins generally lose their stability and activity under non-physiological conditions. In particular, most proteins are able to carry out their activity only within a very narrow range of temperature. The ability to design modified proteins that remain structured and active at higher (or lower) temperatures is therefore an important objective of both fundamental and industrial research.

The project consists of developing a computational approach to predict the changes in thermal stability of proteins upon mutations, on the basis of statistical energy functions that have been previously created in our group. This software will be applied to rationally modify proteins of industrial interest in view of improving their tolerance of temperature variations and consequently increasing their efficiency within a given application. This theme is part of a project of creation of a spin-off company originating from our research group.

2. Prediction of protein solubility changes upon mutation [Dimitri Gilis, Yves Dehouck, Marianne Rooman]

The properties of proteins are widely exploited in the field of biotechnology. However, the use and the purification of proteins with a low solubility remain difficult, and this low solubility can affect their activity. Mutating the protein sequence can improve the protein solubility. We propose to develop a method aiming at predicting the protein solubility changes upon single-site mutations. This prediction method will rely on approaches that are similar to those used to develop our programs that predict thermodynamic (see http://babylone.ulb.ac.be/popmusic) and thermal stability changes upon mutation. These different programs will be grouped in a suite that aims at helping the design of proteins with modified physical-chemical properties. This subject is also related to the creation of a spin-off company from our research unit.

3. Bioinformatic analysis of punctual variations in the human genome and of their relationship with disease [Yves Dehouck, Marianne Rooman]

During the last few years, high-throughput sequencing technologies have become much less expensive, and have produced massive amounts of data. The analysis and the exploitation of these data is one of the major current challenges in genomics and bioinformatics. In particular, millions of Single Nucleotide Polymorphisms (SNP) have been identified by collaborative projects such as the HapMap Consortium of the Human Variation Project. Each SNP may affect the expression or function of one or several genes, and have in consequence an impact on the functioning of the living organism. Non-synonymous SNPs that occur within coding regions of the DNA are of particular interest since they lead to differences in the sequences of the expressed proteins (SAAP: Single Amino Acid Polymorphism). The annotation of these variations and of their impact on human health is an important step towards a better understanding of the genetic background of hereditary diseases, and towards the development of personalized therapeutic approaches for complex diseases such as cancer or diabetes.

The project consists of developing an automatic approach for the analysis of SAAP data in view of predicting their possible relationship with disease. In contrast with most existing methods that have been developed for this purpose, the considered approach will focus more specifically on the exploitation of structural information about the variant proteins. More specifically, structural models will have to be created for each target protein. These models will be subjected to various structural analysis tools that were previously developed in our group or in others, in view of predicting the effects of SAAPs and their relationship with the development of diseases.

4. Rational design of vaccines: prediction of B-cell epitopes [Marianne Rooman]

The *in silico* prediction of B-cell epitopes through the use of bioinformatics tools is an extremely promising way forward in the rational design of vaccines, as it drastically limits and guides experimental tests - among others on animals - and leads to significant gains in time and cost. The development of a reliable prediction tool of B-cell epitopes is an objective of prime importance.

To develop software for predicting B-cell epitopes, we will use the sequences and structures of experimentally characterized antibody-antigen complexes, which are available in specialized databases. The sequence and structure characteristics of the antibody-antigen interfaces will be identified, as well as their energy characteristics. These data will be rationalized, combined, and will form the basis of the prediction method.

5. Modeling of the structure of a membrane receptor and of its interactions with a peptide [Dimitri Gilis, Yves Dehouck et Marianne Rooman]

VIP is a neuropeptide that is synthetized and secreted by the central nervous system. VIP can present anti-inflammatory properties through its interaction with the VPAC1 membrane receptor. It can also interact with the VPAC2 and PACAP membrane receptors. This project aims at modeling the structure of membrane receptors (VPAC1, VPAC2, PACAP) and to study *in silico* the VIP-receptor interaction by using docking methods. This master thesis is in the general framework of a collaboration with IRIBHM (Prof. I. Langer).

6. Study of the dynamics of house dust mite allergens from family 5 [D. Gilis]

House dust mites allergy represents an important public health problem. House dust mite allergens are grouped in several families. Allergens from family 5 are proteins whose biological function is still unknown, as well as how they provoke allergy. They show a three-helix bundle, and some of them are monomers, whereas others are dimers. In the dimers, one of the helices is kinked, and a hydrophobic cavity that could accommodate a ligand is observed. The relationships between these structural features and the allergenic properties of these allergens are also unknown.

We propose to study by molecular dynamics methods the influence of mutations and of solvent conditions on the dynamics of these allergens.

7. Origin of the genetic code and prebiotic peptides [Marianne Rooman]

Although some stages in the origin of life have been determined with some confidence (prebiotic synthesis of amino acids, formation of peptides and other complex molecules under prebiotic conditions, formation of cell membrane ancestors, etc.), there are still many unsolved issues, such as the origin of the genetic code and the form of the first functional peptides. Regarding the genetic code, it was suggested that specific RNA sequences interact with individual amino acids, and that these sequences contain codons and/or anticodons of the amino acids to which they bind. With regards to the prebiotic peptides, recent studies suggest that small peptides of 5 to 8 residues, built from the simplest amino acids (Gly, Ala, Asp and Val), could constitute the missing link between the individual amino acids synthesized in prebiotic conditions on the primitive earth and the current proteins that are encoded in the genomes.

The project will consist of studying the affinity of individual amino acids for specific RNA sequences, via molecular dynamics simulations, in view of determining what interactions are favored. This could give hints for the understanding of the origin of the genetic code. Moreover, the structure and stability of peptides proposed as prebiotics peptide candidates will also be studied by molecular dynamics simulations. The experimental analysis of their structural and functional features can also be performed, by infrared spectroscopy in collaboration with Prof. E. Goormaghtigh (SFMB), and by SDS-PAGE techniques and mass spectrometry in collaboration with Prof. I. Langer (IRIBHM).

8. Dynamical modeling of gene expression in yeast and bacteria subject to perturbations [Marianne Rooman and Jaroslav Albert]

The mechanisms of gene expression regulation form complex networks that involve a large number of proteins and depend on the cell cycle, stage of development of the organism, tissue, external perturbations, and whether the cells are healthy or not. A growing body of experimental data on transcribed genes in different systems and conditions, obtained among others by DNA microarray techniques, are available in public databases. We will focus on time series, describing yeast cells or bacteria subject to external perturbations such as

radiation, temperature changes, or diauxie. The time evolution of interactions between genes and their products (RNA, proteins) in these systems will be modeled by systems of differential equations, whose parameters will be identified on the basis of available experimental data. The different models will be analyzed on the basis of their stability, robustness, and their evolution towards a fixed point.

9. Design of synthetic gene circuits and noise control [Jaroslav Albert, Marianne Rooman]

Synthetic biology is a relatively new field whose focus is on improving already existing and engineering new gene circuits. In order to implement a new gene circuit one must understand its behavior in terms of the system parameters. This requires that we understand not only the system's averaged deterministic behavior, but also how random variations, e.i. internal and external noise, affect its performance. Since the computational time for increasingly complex gene circuits grows exponentially, it is necessary to find simplified methods of stochastic simulation. One such method involves a separation of a large circuit into smaller motifs which can be understood in terms of a relatively few system parameters. General rules about the system's behavior can then be inferred. Using a combination of deterministic models (e.g. differential equations) and stochastic algorithms, we will be analyzing small circuits with positive autoregulation. Such circuits are of interest due to their ability to increase the randomness of their own expression.

10. Development of protein-protein docking algorithms [Dimitri Gilis, Yves Dehouck, Marianne Rooman]

A large number of proteins are biologically active only when they are associated in multimers or when they are in complex with a ligand. Protein-protein and protein-ligand associations are also related to the regulation of biological processes, to signal transduction or to the catalysis of biochemical reactions. Understanding and predicting how proteins interact is thus very important.

The project aims at developing a protein-protein or a protein-ligand docking algorithm that predicts the relative orientation of two proteins or a protein and a ligand. Several sampling algorithms will be considered, with several scoring functions to compute the interactions. Molecular flexibility will also be modelled. The best algorithm will be implemented and tested. This master thesis is in the framework of a collaboration with the Informatics Department (Pr. J. Cardinal).

11. Development of an algorithm for side-chain positioning within proteins [Dimitri Gilis, Yves Dehouck, Marianne Rooman]

Most methods that have been developed to predict protein structures from their sequences or to dock proteins with their ligands rely heavily on a simplified description of protein structure and frequently neglect the degrees of freedom associated with the conformations of the side-chains of the residues. Such simplifications are necessary to perform efficient searches in the conformational space of proteins within reasonable computation times. Still, a correct positioning of side-chains is necessary in later stages in order to refine the structural models and analyze functional mechanisms. Side-chain positioning is a complex combinatorial problem, which must take into account the intrinsic preferences of each side-chain, the possible interactions with neighboring residues, the optimal packing of the protein chain, and the flexibility of the protein backbone.

The project will be achieved by combining optimization algorithms (linear programming) and different types of energy functions. The relative weights of these energy functions will be determined through parametric identification procedures. This project is in collaboration with the Department of Informatics (Pr. J. Cardinal).

Coordinates
Pr. Marianne Rooman : mrooman@ulb.ac.be
Pr. Dimitri Gilis: dgilis@ulb.ac.be
Dr. Yves Dehouck: ydehouck@ulb.ac.be
Dr. Jaroslav Albert : jalbert@ulb.ac.be