Post-doc position available at the Université Pierre et Marie Curie, Paris

Coevolution analysis of viral sequences: from HBV to a general framework

Context & Biological Question. A particular focus has been drawn in recent years to coevolving residues, within a protein and among proteins. Coevolving residues in a protein structure, possibly a complex, correspond to groups of residues whose mutations have arisen simultaneously during the evolution of different species, and this is due to several possible reasons involving the three-dimensional shape of the protein: functional interactions, conformational changes and folding. Several studies addressed the problem of extracting signals of coevolution between residues. All these methods provide sets of coevolved residues that are usually close in the three-dimensional structure, form connected networks covering roughly a third of the entire structure, and have been demonstrated for a few protein complexes to play a crucial role in allosteric mechanisms, to maintain short paths in network communication and to mediate signaling. These methods are applicable to protein families displaying a large number of evolutionarily related sequences and sufficient divergence, these characteristics constituting the bottleneck of today coevolution analysis methods. In this project we shall go beyond this bottleneck with a further development of BIS^2, a fast algorithm for the coevolution analysis of relatively small sets of sequences (where "small" means < 50 sequences) displaying high similarity [1,2,4]. BIS^2 can be applied to many proteins, characteristic of vertebrate or viral species, where other coevolution methods cannot be applied because of the reduced number of sequences (either coming from species or from populations) and their conservation.

In this project, we aim at decrypting mechanistic features of the hepatitis B virus (HBV) fusion process and at proposing an experimentally supported model for HBV membrane fusion mechanism. This will be done through a multidisciplinary approach, combining bioinformatics analysis of molecular coevolution and experimental assays. Due to the study of sequences coming from HB viral populations (identified in geographical regions, or even in single patients),

Previous studies. The BIS method was recently used to address two questions on hepatitis C virus (HCV). The first concerns the reconstruction of the HCV protein-protein interaction network [2] and the second the understanding of the HCV fusion mechanism [3]. These two questions could not have been addressed by other current state-of-the-art coevolution methods, due to their requirement for large numbers of sequences with high divergence. This pioneering work constitutes the first coevolution analysis of an entire viral genome and opens the way to coevolution-based reconstruction of protein-protein interaction networks in viruses.

Directions developed by the post-doc. Within the project, the specific objectives of the postdoc are: 1. the development of a coevolution analysis tool for HBV viral populations; 2. the computational identification of the residues being involved in membrane fusion in HBV.

1. The development of a coevolution analysis tool for HBV viral populations.

We propose to extend the BIS^2 method in several directions taking into consideration the non standard genomes of viral species and viral populations. Among the several aspects that we will explicitly handle, we highlight a systematic inclusion of independent studies of several genotypes in the predictions of intra and inter-protein signals, a nucleotide analysis of the viral sequences leading to a distinction of protein functions and protein interactions, viral gene overlapping and genome recombination. We will develop an improved method, specifically conceived for viral sequences, that leads to the prediction of coevolving residues with high statistical significance. Suitable statistical scores (p-value), taking into consideration the various constraints that we shall include in our study, should be defined.

2. The computational identification of the residues being involved in HBV membrane fusion. We shall extend the use of our coevolution analysis tool (see Objective 1), to predict intra- and inter-protein interactions underlying the HBV entry process, including the binding and membrane fusion mechanisms. We shall tackle the problem of the interactions among surface proteins, even for proteins where no structure is available. We shall reconstruct coevolution signals on independent analysis of genotype sequence families based on which we shall identify pairs of physically proximal coevolved clusters of residues and use them to possibly support hypotheses of large

collective movements (driving conformational changes). Depending on the evolution of the project, we might be able to extend the analysis to different actors in the fusion process.

The project demands the development of efficient algorithms.

References

1 Dib, L., Carbone, A. Protein fragments: Functional and structural roles of their coevolution networks. *PLoS ONE* 7(11), e48124 (2012).

2 Champeimont, E., Laine, E., Hu, S.-W., Penin, F. & Carbone, A. Coevolution analysis of Hepatitis C virus genome to identify the structural and functional dependency network of viral proteins. *Scientific Reports* 6:26401 (2016).

3 Douam, F. *et al.* A protein coevolution method designed for conserved sequences uncovers critical features of the original HCV fusion mechanism and provides molecular basis for the design of effective antiviral strategies. *In revision* (2016).

4 Baussand, J. & Carbone, A. A combinatorial approach to detect co-evolved amino-acid networks in protein families with variable divergence. *PLoS Computational Biology* 5, e1000488, (2009).

Environment:

The postdoc will benefit from the interaction with the experimental virologists at the Ecole Normale Supérieure de Lyon (Francois-Loic Cosset Lab – Centre International de Recherche en Infectiologie ; http://ciri.inserm.fr/le-ciri/presentation/#) to include in the formal model all aspects of the complexity of HBV genome that will be necessary to obtain accurate predictions. A fundamental aspect of the project is the back-and-forth between the experimentalists and the computer scientists/mathematicians involved; it will be fundamental to design the experimental strategies based on an appropriate understanding of genomic coevolution signals and to conceive an appropriate formal tool for the coevolution analysis of viral genomes. There are two distinguished outputs for the project: the identification of the mechanistic features of the HBV fusion process and the novel tool for coevolution analysis of viral genomes that will be made available to the community for exploitation with other viral genomes. The results obtained by the postdoc will be made available to the community. They will be of interest for the entire virology community because we do expect them to be applicable to other viruses.

The research team (Analytical Genomics team) is part of the Laboratory of Computational and Quantitative Biology (LCQB) UMR7238 CNRS-UPMC, headed by A. Carbone. The LCQB has a very dynamic environment where 8 teams work, in different area, at the frontier of computational biology. Information on the lab and the teams can be found at http://www.lcqb.upmc.fr. For the team, see http://www.lcqb.upmc.fr/AnalGenom/ and http://www.lcqb.upmc.fr/AnalGenom/projects.html

Period: The postdoc will last 2 years. It is available from March 1st, 2017. It is funded by ANRS (Agence Nationale pour la Recherche contre le SIDA et les Hépatites Virales).

Postdoc profile. We look for a postdoc with a very successful PhD experience and an excellent publication record in either

1. sequence analysis and evolution.

or

2. mathematics/computer science with knowledge in graph theory and combinatorics, pattern matching, and in protein structures.

Contacts:

A.Carbone – alessandra.carbone@lip6.fr

Laboratory of Computational and Quantitative Biology (LCQB) UMR7238 CNRS-UPMC
Université Pierre et Marie Curie, Paris.