Towards a genome-scale coevolutionary landscape of the bacteria

Résumé du projet de recherche (Langue 1)

In the course of evolution, genes and genomes change continuously. However, changes cannot be random, a gene has to code for a functional protein, a genome for a viable organism. Thanks to next-generation sequencing, more than 35 000 genomes are sequenced, offering ample data to assess genetic and genomic variability. Novel computational approaches are required to extract information contained in raw data, and to understand the evolutionary and biological constraints shaping their variability. The project shall capture this variability for the bacteria, starting with Escherichia coli as a reference (but using thousands of other species in the variability analysis). It is based on the Direct Coupling Analysis (DCA), a computational approach developed by the first proponent’s team. The Calsimlab has recently financed a joint postdoc of the proposing teams, who has developed a DCA-based approach to infer mutational landscapes integrating protein sequence and structural data. The aim of the current project is to exploit these developments and carry them from few example proteins to a large-scale view in the bacteria. Due to the inherent computational complexity of genome-wide studies, we propose a three-step approach: 1. A genome-wide application of DCA to all protein domain families present in E. coli, and an assessment of the effects on protein function of thousands of mutations discovered in experimental evolution (unpublished data provided by co-supervisor O. Tenaillon). 2. A genome-wide analysis of inter-domain coevolution in all multi-domain proteins of E. coli, and the assembly of full-length protein structures for proteins where only separate domains have been structurally resolved. 3. A genome-wide analysis of inter-protein coevolution between E. coli proteins, using genomic co-localization and phylogenetic profiling to select interesting and potentially interacting protein pairs.

Résumé du projet de recherche (Langue 2)

The project relies on a tight collaboration between computational research and biological research, represented by the two proponents, and partially uses algorithmic approaches inspired by one proponent’s original statistical-physics background. Results of computational studies shall be tested in wet-lab experiments. This places the project firmly at the interface between computer science, physics and biology.