Coevolutionary modeling of protein sequences: Predicting structure, function, and mutational landscapes

MARTIN WEIGT, Universite Pierre et Marie Curie, Paris

Over the last years, biological research has been revolutionized by experimental high-throughput techniques, in particular by next-generation sequencing technology. Unprecedented amounts of data are accumulating, and there is a growing request for computational methods unveiling the information hidden in raw data, thereby increasing our understanding of complex biological systems. Statistical-physics models based on the maximum-entropy principle have, in the last few years, played an important role in this context. To give a specific example, proteins and many non-coding RNA show a remarkable degree of structural and functional conservation in the course of evolution, despite a large variability in amino acid sequences. We have developed a statistical-mechanics inspired inference approach - called Direct-Coupling Analysis - to link this sequence variability (easy to observe in sequence alignments, which are available in public sequence databases) to bio-molecular structure and function. In my presentation I will show, how this methodology can be used (i) to infer contacts between residues and thus to guide tertiary and quaternary protein structure prediction and RNA structure prediction, (ii) to discriminate interacting from non-interacting protein families, and thus to infer conserved protein-protein interaction networks, and (iii) to reconstruct mutational landscapes and thus to predict the phenotypic effect of mutations. References [1] M. Figliuzzi, H. Jacquier, A. Schug, O. Tenaillon and M. Weigt "Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1", Mol. Biol. Evol. (2015), doi: 10.1093/molbev/msv211 [2] E. De Leonardis, B. Lutz, S. Ratz, S. Cocco, R. Monasson, A. Schug, M. Weigt "Direct-Coupling Analysis of nucleotide coevolution facilitates RNA secondary and tertiary structure prediction", Nucleic Acids Research (2015), doi: 10.1093/nar/gkv932 [3] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. Marks, C. Sander, R. Zecchina, J.N. Onuchic, T. Hwa, M. Weigt, "Direct-coupling analysis of residue co-evolution captures native contacts across many protein families", Proc. Natl. Acad. Sci. 108, E1293-E1301 (2011).