Inference in Genetic Data
Raganathan Rama

Summary
Proteins can fold spontaneously into well-defined three-dimensional structures and can carry out complex biochemical reactions such as molecular recognition, catalysis, and allosteric communication. The precision required for these properties is somehow achieved while also preserving evolvability — the capacity for adaptive variation in response to ever-changing selection pressures. How are proteins built in Nature to support all of these properties? This is a question where physics has made headway. I will review several approach, in particular the statistical coupling analysis, or SCA, for deducing the pattern of constraints on amino acid residues in proteins through statistical analysis of the evolutionary divergence of a protein family. This approach reveals a novel decomposition of proteins into sparse groups of co-evolving amino acids that we term “protein sectors”. The sectors comprise physically connected networks in the tertiary structure and can be modular — with different sectors in a single protein delivering different functional properties. Experiments in several protein systems demonstrate the importance of the sectors and importantly, the SCA information was shown to be necessary and sufficient to design functional artificial proteins without the use of any direct structural or chemical information. These results suggest that sectors are the conserved units of folding, function, and adaptability in natural proteins. Two key problems will further be discussed: (1) understanding the physical mechanisms underlying sectors, and (2) defining how the dynamics of the evolutionary process controls the emergence of this structural architecture in proteins.

Note
Host: M. Wyart (EPFL SB IPHYS PCSL)
Swiss Universities

Learning Prerequisites
Recommended courses
stat mech 1 & 2 Biophysics

Expected student activities
The student must be able to understand physics-based inference methods for genetic sequence data.