Atelier
« Les mathématiques de l'évolution des séquences : modèles biologiques et applications »
23-27 septembre 2013

Workshop
"Mathematics of Sequence Evolution: Biological Models and Applications"
September 23–27, 2013

Modelling structural constraints on protein evolution: current progress and open challenges

Claudia Kleinman*

claudia.kleinman@mcgill.ca

WEB: www.genomequebec.mcgill.ca/compgen/majewskilab/ClaudiaKleinman

---

Protein sequences are the net result of a complex evolutionary process combining mutation, stochastic variation due to genetic drift and natural selection on a number of diverse phenotypic traits. One such trait is the protein three dimensional structure. While its importance in shaping natural sequences is generally accepted, the incorporation of structural constraints into evolutionary models has proven to be a difficult task, due to the theoretical and computational complexities of both the structural modeling and the phylogenetic methods involved. Effective approaches to interdisciplinary protein modeling are only beginning to emerge.

In this talk, we will review the progress in the development of probabilistic models of sequence evolution that explicitly describe structural constraints, in particular those relaxing the assumption of independence between sites required by common phylogenetic methods. Statistical inference with site-interdependency will be discussed, as well as the different approaches proposed to date to describe the protein structure.

Compared to models of sequence evolution that do not attempt to mechanistically describe the underlying biological processes, structurally constrained models hold an explanatory power that purely phenomenological ones cannot provide. A preliminary study case will be presented illustrating this point, where selective constraints imposed to maintain the structure are investigated in relationship with protein expression level, a factor known to affect evolutionary rate. Patterns of model fit and posterior distributions of parameters associated with selection obtained on a set of highly expressed genes in *E. coli* are compared to those obtained on the lowest expressed genes. A higher strength of selection for compatibility with the folded state is found in the most abundant genes. In particular,

*McGill University and Génome Québec Innovation Centre, 740 Dr. Penfield Ave., Montréal, QC H3A 1A4, CANADA.

sequence changes disrupting amino acid propensities for solvent accessibility are significantly penalized in these abundant proteins, suggesting that avoidance of misfolding and aggregation plays a role in the long-standing correlation between evolutionary rate and expression level.

Finally, we will move beyond the description of an evolutionary process acting on a static protein structure to discuss recent literature investigating constraints not exclusively restricted to the stability of isolated molecules, such as flexibility, cellular crowding, avoidance of misfolding and protein-protein interactions.