Published in final edited form as:

Curr Opin Struct Biol. 2010 June ; 20(3): 333–334. doi:10.1016/j.sbi.2010.04.004.

Sequences and Topology: Evolution of proteins and evolution of computational approaches

Sarah A Teichmann and

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0EH, UK

Nick Grishin

Howard Hughes Medical Institute, Department of Biochemistry, University of Texas, Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9050, USA

Sarah A Teichmann: sat@mrc-lmb.cam.ac.uk; Nick Grishin: grishin@chop.swmed.edu

There are eight exciting reviews in this issue of Sequences & Topology, spanning from computational approaches for protein annotation in the era of personal genomes, to new analysis of protein conformational changes in the era of structural genomics. A common theme is evolution: both evolution of proteins and evolution of the field in the wake of a deluge of sequence, structure and other types of data.

The impact of next generation sequencing technology means that there are now many variations on a particular protein sequence. This poses new challenges for the structural and functional annotation of proteins. Blicher and colleagues present a review of the ways that sequence variation can be taken into account and variants can be compared in terms of their structural and functional features. In particular, they present their integrated tool called EPipe for display of differential features across sequence variants.

Jordan, Ramensky and Sunyaev review various methods, from phylogenetic to structural, to predict effects of non-synonymous alleles on the functions of human proteins, and applications of such analyses to genetics, from population to medical. Development of these approaches is particularly timely today, as tomorrow we will be flooded with many thousands human genomes, and many of the current approaches might not be up to the task of handling such massive amounts of data.

A different approach of tapping into available genetic diversity for functional studies is to compare proteins between species and to use phylogenetic tools for functional annotation. Lichtarge and Wilkins discuss the latest developments of the Evolutionary Tracing (ET) method to predict locations of functional sites utilizing non-random sequence variation revealing itself as a functional diversification between subfamilies. The authors show both the wide applicability of the method to large-scale analysis and illustrate some case studies, for instance functional networks of residues within G-protein coupled receptors.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

To further understand the interplay between structure, function and evolution, interesting approaches reviewed by Harms and Thornton have recently emerged. “Resurrection” of prehistoric proteins through ancestral sequence reconstruction and experimental studies of proteins encoded by such sequences bridges biochemistry with evolutionary biology and offers unprecedented opportunities. For instance, probing mutations in the ancestral sequence background will explore possible trajectories not taken by evolution and might help answering questions about the balance between chance and function.

One of the characteristics of proteins that the above techniques can inform on is conformational change. Perica & Chothia use the many three-dimensional crystal complexes of ubiquitin with ubiquitin-binding proteins to analyse in detail the nature of conformational changes in ubiquitin when it interacts with other proteins. They discuss the importance of conformational selection in protein recognition in this system. It emerges that there are three binding residues in ubiquitin that are invariant in their positions, while loop regions take on many different conformations in different contexts of interactors.

The protein ubiquitin is also used as a test case by Friedland & Kortemme in investigating the relationship between divergence in sequence and divergence in protein conformation. The context of this finding is their computational sequence design method that takes into account sampling of conformational space when generating diverse sequences. They discuss the wide range of applications that have been demonstrated using sequence design, and emphasize the impact of new sequencing and gene synthesis technologies on this field.

The mode and tempo of protein sequence divergence is one of the oldest themes in computational biology, and we end with two reviews that look at different aspects of sequence divergence.

Wilke and Drummond discuss various factors affecting evolution of coding regions. The primary interest of their review is not the direct functional constraints, but biophysical factors, such as protein synthesis, folding and interactions between proteins. These effects could be dominating protein evolution. Does evolutionary rate depend on a protein structure or the level of expression? Why is evolutionary rate of a protein not strongly affected by the number of its interacting partners? Do silent (synonymous) mutations influence the fitness of a protein? These tantalising questions are at the centre of this review.

Protein evolution in both an abstract and very specific sense is reviewed by Bornberg-Bauer and colleagues. They describe recent findings on the emergence of new genes from non-coding DNA in organisms, and the possibility that many of these may initially be intrinsically unstructured proteins. These may then evolve into globular proteins eventually. How do such transitions take place? Theoretical models as well as artificial evolution experiments suggest evolutionary intermediates are often multi-functional and –structural. These flexible intermediates provide a route for transitions from one sequence and phenotype to another.

Molecular evolution generates sequence diversity that reflects protein structures and functions. Analysis of intra- and inter-specific sequence variation is increasingly productive in terms of functional annotations and understanding mechanisms of protein functions and

diseases, interactions between proteins and, finally, mechanisms of evolution itself. The evolution of computational approaches themselves is shaped by the newly accumulating vast amounts of data, ensuring that researchers are offered the best predictive tools available to aid in experimental exploration of proteins, organisms and populations.

Biographies

Sarah works on the evolution of protein–protein and protein–nucleic acid interactions using protein structure and interaction network data. Over the past few years, a particular focus has been on the prediction, evolution and dynamics of transcription factors and transcriptional regulatory networks, and a further focus is the classification and analysis of structures of protein complexes.

Nick is using theoretical methods to study protein molecules. The major efforts are devoted to understanding of evolutionary relationships between protein domains and development of computational methods to assist in expert-driven comparative analysis of protein sequences, structures and functions.