

Meeting report

Evolving proteins at Darwin's bicentenary

John W Pinney and Michael PH Stumpf

Address: Centre for Bioinformatics, Division of Molecular Biosciences, Imperial College London, Wolfson Building, London SW7 2AZ, UK.

Correspondence: Michael PH Stumpf. Email: m.stumpf@imperial.ac.uk

Published: 17 April 2009

Genome Biology 2009, **10**:307 (doi:10.1186/gb-2009-10-4-307)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2009/10/4/307>

© 2009 BioMed Central Ltd

A report of the Biochemical Society/Wellcome Trust meeting 'Protein Evolution - Sequences, Structures and Systems', Hinxton, UK, 26-27 January 2009.

The effects of natural selection are ultimately mediated through protein function. The traditional view that selection on proteins is primarily due to the effects of mutations on protein structure has, however, in recent years been replaced by a much richer picture. This modern perspective was in evidence at a recent meeting on protein evolution in Hinxton, UK. Here we report some of the highlights.

Unsurprisingly, Charles Darwin featured at lot at the meeting. Evolutionary arguments are all-pervasive in the biomedical and life sciences and this is particularly true for the analysis of proteins and their role in cell and molecular biology. From initial investigations of individual proteins in the 1940s and 1950s, which were motivated by even earlier work on blood groups, we can now routinely collect information from a large number of sequenced genomes to help us understand the evolution of proteins in terms of their sequences, structures and functions, and their roles as parts of biological systems.

Comparative evolution

The primacy of comparative, and thus evolutionary, arguments in the analysis of proteins and their structure was emphasized by Tom Blundell (University of Cambridge, UK), who reviewed almost 40 years of structural bioinformatics. He noted that in the early studies of insulin structure, the common ancestry of all life on Earth meant that lessons learned in the context of one species were transferable to other species. This in turn meant that sequence data could

be linked to structure more directly through comparative arguments than would have been possible using biophysical or biochemical arguments. Despite vast increases in computational power and experimental resolution, this continues to be the case to the present day.

The explosion in available whole-genome data has provided us with a much richer understanding of genomic aspects of protein evolution. This was highlighted by Chris Ponting (University of Oxford, UK), who contrasted the distributions of proteins and protein family members in the human and mouse genomes. Such a comparison reveals high levels of sequence duplication - probably in line with what might be expected, given recent findings of copy-number variation - and suggests a scenario where ancient single-copy genes are only rarely gained or lost. Members of larger gene families, however, have experienced much more frequent gene duplication and loss; this may reflect the role of such gene families in adaptive evolution, as seen in the rapid evolution of the androgen-binding proteins in mouse.

The theme of adaptation was elaborated on by Bengt Mannervik (Uppsala University, Sweden), who focused on the evolution of enzymes, a class of proteins with perhaps uniquely well-characterized functionality. Here, he argued, the relative trade-off between substrate specificity and enzymatic activity has given rise to a quasi-species-like evolutionary scenario: abundant protein polymorphisms underlie a complex population of functional enzymatic variants. Such diversity in the metabolic functions available within the population may presumably help to buffer changes in the environment encountered during evolution.

Araxi Urrutia (University of Bath, UK) addressed predominantly the link between gene and protein expression and evolutionary conservation and adaptation. As she pointed out, there is clear emerging evidence that highly

expressed genes in humans share certain characteristics such as short intron lengths and higher codon-usage bias and favor less metabolically expensive amino acids. This affects the rate at which protein-coding genes evolve in a manner independent of protein structure. Moreover, this level of selection also appears to depend on the genomic context, as patterns of expression of neighboring genes are statistically correlated.

Insights from structure

Also fundamental to protein activity is post-translational modification, notably phosphorylation. This is a field of enormous biomedical importance, as kinase and phosphatase activities crucially regulate signaling and metabolic processes. The structural work of Louise Johnson (University of Oxford, UK) and colleagues bridges 'classical' structural biology and systems biology, and she discussed the structural factors underlying the regulation of kinases and phosphorylation. These comprehensive analyses are now also beginning to reveal how biochemical compounds can affect kinase regulation in a manner that may become clinically exploitable.

Keeping to the structural theme, Christine Orengo (University College London, UK) discussed the phenomenal insights that have been gained recently into the evolution of protein domain superfamilies and the ensuing effects that this can have on protein structure, active sites, and ultimately, function. For example, the analysis clearly reveals common structural cores that are shared across the members of the same superfamily but may be modified in individual members. Orengo documented how such differences in the HUP superdomain family lead to differences in the participation of paralogs in protein complexes and biological processes following duplication.

Alex Bateman (Wellcome Trust Sanger Institute, UK) further elaborated on the evolution of families of protein domains. Such a domain-centric point of view adds a valuable and useful perspective. Yet even at the level of shuffling these protein building blocks, the picture becomes more detailed as the available evolutionary resolution increases: for example, the frequency of changes in domain architecture is seen to approximately double following a gene duplication event as compared with a speciation event.

Protein evolution *in vitro* and *in vivo*

Using extensive and genome-wide data from yeast and humans, Laurence Hurst (University of Bath, UK) demonstrated the substantial role of non-structural selection pressures, such as those imposed by transcription and translation, on the evolutionary dynamics of proteins.

Taking these into account results in a much richer picture of protein evolution, with the contribution of splicing-related constraints being particularly pronounced in mammals. Surprisingly, perhaps, these constraints show the same relative importance for protein evolution as aspects of gene expression do, as discussed by Urrutia. This is in stark contrast to the traditional amino-acid-centered view of protein evolution.

Using analogies with mountaineering, Dan Tawfik (Weizmann Institute, Rehovot, Israel) covered the exciting opportunities afforded by experimental studies of protein evolution. Evolution has sometimes been viewed previously as an observational and mathematical discipline rather than one characterized by experimental work. Tawfik showed how it is possible to explore evolutionary trajectories through the space of possible protein folds or functions in far more detail than had previously been thought possible. One of the exciting possibilities emerging from this work is that we will be able to study the interplay between neutral evolution and the various factors influencing selection. There is already good direct experimental detail from these laboratory studies that demonstrate the link between the rate of protein evolution and 'functional promiscuity' and conformational variability.

One of us (MPHS) described the phage-shock stress response in *Escherichia coli* as an example in which the loss and gain of proteins across bacterial species can only be understood in the context of mechanistic models of the system itself. Loss of individual genes can compromise the functionality of the stress response, which can only be tolerated under certain ecological conditions. As a result, it appears that either the complete set of proteins contributing to the stress response is maintained in bacterial genomes, or all are lost together. This all-or-nothing scenario is probably inextricably linked to the ecological niches inhabited by the bacteria.

David Robertson (University of Manchester, UK) discussed how patterns of gene duplication and diversification have shaped the global structure of protein-protein interaction networks, as well as many of their detailed features. In contrast to previous work, this detailed analysis of the protein-interaction network in *Saccharomyces cerevisiae* clearly shows that the coevolution of interacting proteins cannot simply be explained by observed protein-protein interactions. What emerges from this and related studies is that many of the high-level models of network evolution proposed only a few years ago are too simplistic for dealing with such highly contingent and complex processes. Robertson concluded with a discussion of the evolutionary history of human disease genes, which also highlights the importance of historical levels of gene duplication, and reinforces the need for nuanced assessment of the different factors affecting protein evolution.

Discussing the physical interactions of kinases, Mike Tyers (University of Edinburgh, UK) described an exciting new experimental mapping study of physical protein-protein interactions of kinases. The experimental determination of these, frequently weak, protein interactions poses many challenges, requiring considerable reworking of existing platforms for proteomics, but the information produced is expected to be of great value to systems biologists. Preliminary results already suggest that the wealth of material expected from this survey will aid our understanding of the molecular mechanisms involved in these processes.

Two hundred years after the birth of Charles Darwin, we understand a great deal about the processes of evolution and how they have shaped the diversity of life on Earth. The application of the simple idea of "descent with modification" to proteins, their structures, expression patterns, interactions and ultimately their emergent functions continues to produce fundamental insights into how biological systems evolve. But the picture emerging from this unprecedented access to molecular data at all levels of cellular organization is much more nuanced than we would have thought possible only a few years ago.