[Invited Speech/ CBRC2012:The Impact of Next Generation Sequencers]

On the evolution of orphan proteins as domains

Arne Elofsson
Science for Life Laboratory, Stockholm University, Sweden

During gene evolution various insertion and deletion (indel) events occur. The fate of any given indel is subject to constrains at the DNA, mRNA and protein levels. For instance, previous studies show that such events are more likely to take place within region tandem nucleotide repeats. Clearly, the fate of the indel depends upon its location in the protein sequence as insertions in secondary structure elements are more likely to be deleterious than indels in loop regions.

Intrinsically disordered regions are regions that do not assume known secondary structures. These malleable regions are important in signalling and protein-protein interactions and, per definition, are not present inside protein domains. We have found that indels located in the disordered regions are more prevalent than expected. There is a striking relationship between variation in length and variation in the number of intrinsically disordered residues. Indeed, up to seven out of ten inserted/deleted residues are disordered. This relationship holds true for several different eukaryotic sets of homologous protein pairs, regardless of whether the proteins are paralogs or orthologs.
Further, indel events seem to be under a lower selective pressure within disordered regions. The disordered indels are also on avarage slightly longer than those in ordered regions.

The number of indels per point mutation is at least twice as high in proteins with more than 20% disorder than in fully ordered proteins. Disordered regions are known to mediate binding between proteins and they are also often situated in loops. The contribution of binding properties and the lack of effect on the protein domain modules may explain why indels in these regions are more prevalent than expected.