Poster C-10

Evolutionary History of Multi-Domain Proteins

Authors:
Diana Ekman (*Stockholm Bioinformatics Center, Stockholm University*)
Åsa Björklund (*Stockholm Bioinformatics Center, Stockholm University*)
Arne Elofsson (*Stockholm Bioinformatics Center, Stockholm University*)

Short Abstract: A majority of the eukaryotic proteins are multidomain proteins. Here, we present a study of the evolution of new domain architectures through fusion and internal duplication. We found that architectures are created mainly through insertions/deletions of domains at the terminals. Expansion of domain repeats is also an important contribution.

Long Abstract:

The eukaryotic proteomes consist mainly of multi-domain proteins that have been created through fusion, deletion and internal repetition of genes, or parts of genes. The fraction of multi-domain proteins is similar in eukaryotes of different complexity, from yeast to human, and significantly higher than in prokaryotes. The expansion of protein domain repeats is believed to have been important particularly for the evolution of higher eukaryotes since large portions of their proteomes consist of repeating protein domains. Here we investigate the creation of new domain architectures through these different evolutionary events.

First, a method for identifying the domain architecture from which each protein originates is required. For that we used the domain distance measure, which is defined as the number of domains that differ between two domain architectures. The domain distance relates to the number of domain insertions, deletions and repetitions that are required to transform one architecture to another. Domains that are not detected by the assignment procedure are problematic in studies of domain architectures, since non-existing domain architectures may be found and other existing architectures missed. To reduce this effect we included domains predicted with low confidence if the domain was part of a domain repeat or if it resulted in a domain architecture that was previously identified in another protein. Regions with no detected domains were treated as domains with no homologs, i.e. orphan domains.

The evolutionary events that distinguish a domain architecture from its most similar architectures were counted. We found that both indels and internal repetitions are frequent, whereas exchanges of domains are rare. Further, indels occur mainly at the N- and C-terminals, while repeated domains may also be inserted between domains. Also in many other respects, domain repeats behave differently from indels, for example, while indels mainly involve a single domain, repeat expansion frequently involve multiple domains. We found, using internal sequence similarity, that some domain families show distinct duplication patterns, e.g. nebulin domains have mainly been expanded seven domains at a time, while duplications of many other domain families involve varying numbers of domains. However, these duplication patterns show no dependence on the size of the domains. Finally, the rapid expansion of domain repeats may to some extent be explained by exon shuffling.

The evolution of architectures was studied in a number of eukaryotes, including yeast and a plant. Obviously, new domain architectures have evolved in every major evolutionary stage. Considerable expansions of domain architectures have occurred, for example, at the emergence of multicellularity and vertebrates.