19. Is there a relationship between protein connectivity and evolutionary rate?

Ramazan Saeed

ramazan.saeed@wolfson.oxford.ac.uk
University of Oxford

In an attempt to investigate the contentious relationship between the evolutionary rate of a protein and its connectivity we performed a large scale study on a number of datasets, looking at expression levels, error rates and set overlaps. Our results show that the relationship is not as straightforward as some claim.

New computational and experimental techniques have begun to produce enormous datasets representing the interactions between proteins. This data allows us to explore how a proteins functional environment (number of interacting partners) is controlled by or effects a proteins other properties.

It has been suggested many a time that proteins which participate in a large number of interactions, evolve at a slower rate (Dickerson, 1971; Ingram, 1961; Wilson et al., 1977; Brookfield, 2000). Fraser et al (2002) demonstrated the negative correlation, which this theory would suggest between protein-protein interactions and evolutionary rate.

However this relationship has since proved to be controversial. The correlation was challenged by a number of different studies, on the basis that the datasets being used were biased (Bloom and Adami, 2003) and the observed association was in fact due to a stronger relationship between evolutionary distance and protein abundance (Jordan et al 2003).

The primary crux of contention in all these investigations could be attributed to the fact that contradicting studies ran their analysis on different interaction datasets often using overlapping data obtained from several experimental studies and some computational methods. Here we scrutinise the relationship, taking data from a variety of sources and measuring any overlap and and error rates for each set.

Datasets we used included DIP, BIND, IntAct, MIPS, GRID and MINT. By finding Best Reciprocal Hit orthologs in the Mus Musculus genome, we modelled the number of amino acid substitutions per site in a protein and recorded the weight and significance of a correlation using Spearman's rank correlation coefficients.

We calculated the overlap between each dataset and examined the expression levels of each interacting protein so as to elucidate any relationship between the abundance of a protein and its connectivity. In order to assess the error rate in each dataset we used the Expression Profile Reliability index (Deane et al 2003), a localisation similarity index and comparison to a good reference set.

We found that statistically significant yet weak correlations (Spearmans rank ~ -0.2) between evolutionary rate and the number of interactions were observed in a few datasets and no correlation was found in others. Error rate and overlap analysis showed that where the correlation was not observed it was probably because of a high number of false positives or true negatives in the datasets. The apparent weakness of the correlation can be attributed to low quality, low coverage interaction data.

A stronger relationship between evolutionary rate and abundance was also observed. Highly abundant proteins tend evolve at a slower rate. Proteins that are abundant would be expected to participate in many interactions, whereas scarce proteins would have fewer. We also demonstrated such a relationship. A fourth factor that constrains the number of interactions is protein age. The younger a protein, the fewer interactions it has, whereas older proteins participate in more interactions.