Conformational diversity: Relationship with protein evolution and designability

Diego Zea¹, Cristina Marino Buslje², Gustavo Parisi³


The study of evolutionary rates is a central issue to understand the mechanisms underlying protein molecular evolution. Several factors have been associated to the modulation of the evolutionary rate such as: genomic location, functional importance of the protein, expression level, structural constraints, protein stability and developmental time(Pal, Papp, and Lercher 2006). However, it was recently established that the gene expression level is the property showing one of the strongest and consistent correlation between genomic data and evolutionary rate (Pearson correlation r of 0.53 and p-value less than 1e-9)(Drummond et al. 2005). Previous estimations have established that protein structural constraints could explain as much as 10% of the evolutionary rate in proteins(Bloom et al. 2006) but recent findings indicate that structure-functional features and translation rates could have comparable contributions to explain evolutionary rates(Wolf, Wolf, and Koonin 2008).

In this work we study how the presence of conformational diversity in proteins could influence the rate of evolution. As we mentioned before, structural constraints could have an important role to define the rate of a given protein. However, most of these studies have been done describing the native state of a protein with a single structure. It is well established that native state of proteins are better represent by an ensemble of different conformers in dynamical equilibrium(Tsai, Ma, and Nussinov 1999). The conformational ensemble is a key concept to explain essential properties of proteins like function(del Sol et al. 2009; Hilser 2010), enzyme and antibody promiscuity(James, Roversi, and Tawfik 2003; Khersonsky, Roodveldt, and Tawfik 2006), enzyme catalytic power (Boehr et al. 2006), signal transduction(Smock and Gierasch 2009), protein-protein recognition(Yogurtcu et al. 2008) and the origin of new functions(Tokuriki and Tawfik 2009).

To study this relationship we used the PCDB database (Protein Conformational Data Base)(Juritz, Fernandez-Alberti, and Parisi 2010). This database contains almost 8000 proteins with different degrees of structural diversity measured as the maximum RMSD (RMSDmax) found between the different conformers for each protein. We have used a subset of the total number of proteins deposited in PCDB. This set consists in those proteins showing conformers with and without binding ligands. Each of these proteins was linked to OMA database(Altenhoff et al.) (The Orthologs Matrix Project) of orthologs to estimate the evolutionary rate. Each set of orthologs was further associated with its corresponding coding DNA and the evolutionary rates were estimated using PAML 4 (Phylogenetic Analysis by Maximum Likelihood) (Yang 1997; Yang 2007) using different codon models. For this purpose we used a tree constructed with Protpars from Phylib package(Felsenstein 1989) for each set of orthologs.

The final set contains 58 proteins corresponding to 67 domains in PCDB with an average RMSD of 1.23Å. The maximum RMSD in the set is 5.07Å. Using this benchmark we have determined that the evolutionary rate positive correlates with the degree of conformational diversity measured by the RMSDmax between conformers (Spearman correlation with a rho of 0.37 and a p-value of 1e-4).

It is interesting to note that a positive correlation has been found between evolutionary rates and protein designability(Bloom et al. 2005). The designability is the number of
sequences adopting a given fold in the global minimum energy (England and Shakhnovich 2003). It was found that high designable folds have a high average of inter residues contacts (England, Shakhnovich, and Shakhnovich 2003; Zeldovich, Chen, and Shakhnovich 2007). Our results support the idea that proteins with larger native conformational space could have a higher average of inter residues contacts giving rise to an increased evolutionary rate. In this sense we have found a positive correlation between the rate of protein evolution (measured by dN/dS) and the number of contacts per residue considering the extension of conformational diversity (rho of 0.44 and a p-value of 5.4e-6).

Our results indicate that conformational diversity relates with protein designability and that both properties have an important role modulating protein evolutionary rates. We think that our findings could have important implications in the understanding of protein evolution process.

References


