Using protein conformational diversity increase protein stability estimation
Ezequiel Juritz, Maria Silvina Fornasari and Gustavo Parisi

The understanding of the mutational effects on protein stability has an outstanding importance in structural biology of proteins. Mutations are essential to explain key processes in molecular evolution such as evolvability (Tokuriki and Tawfik 2009a), protein robustness to mutations (Taverna and Goldstein 2002), the nature of neutral mutations (Bloom et al. 2005), evolutionary rates (Zeldovich, Chen, and Shakhnovich 2007) and the origin of new functions (Tokuriki et al. 2008). The close relationship between mutations and the change in protein stability has been used in biotechnology to design in silico enzymes with enhanced solubility and thermal stability (Cabrita et al. 2007) and to the development of different approaches to disease prediction (Bromberg and Rost 2007). Protein stability can be estimate by the differences in the free energies between the folded and unfolded state of a protein ($\Delta G$). The effects of mutations on protein stability can then be measured using the differences between $\Delta G$ for the wild and the mutation form of the protein ($\Delta \Delta G = \Delta G_{\text{mut}} - \Delta G_{\text{wild}}$). These changes in free energy can be estimated using experimental approaches like those using thermal denaturation or the use of denaturant agents along with calorimetric measurements. Also, several computational methods have been developed to estimate stability changes caused by mutations in proteins. Most of them rely in the analysis of the energetic and/or structural perturbation introduced by the mutation in the protein native structures (Lee and Levitt 1991; Lee 1994; Sippl 1995). The correlation between experimentally determined $\Delta \Delta G$ and those predicted using computational tools, is generally a measure of the reliability on the predictive method used. In this sense, different methods with different datasets have reached a rather good correlation in the range of 0.45 to 0.78 (Gillis and Rooman 1996; Hoppe and Schomburg 2005). Besides the accuracy of the cited methods to estimate free energies changes, predictions on the effect of mutations are commonly estimated using a single structure of the corresponding protein. This approach apparently underestimate the well established concept that the native state of a protein is better represented by a ensemble of conformers (Tsai et al. 2001; James and Tawfik 2003; Lange et al. 2008). The conformational ensemble is a key concept to explain essential properties of proteins like function (del Sol et al. 2009; Hilser 2010), 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 2009b). Here we explore how the presence and extension of conformational diversity affects the estimation of the $\Delta \Delta G$ for a given mutation. Considering the structural rearrangements conformers could have, it is evident that the effect of a mutation would be different in each of the conformers explored. For that propose, we used 1993 mutations in 90 proteins showing different degrees of conformational diversity. These proteins were taken from PCDB database (Juritz, Fernandez-Alberti, and Parisi 2010) a redundant collection of protein structures linked with biological information. The $\Delta \Delta G$ for each mutation was estimated using FOLDX (Guerois, Nielsen, and Serrano 2002) and the different structures (or conformers) for each protein. We have found that the $\Delta \Delta G$ estimation for an individual mutation highly depends on the different conformers used in the estimation. In 83% of the studied proteins we found that at least one mutation could have a maximum difference among conformers of at least 1 kcal/mol in its predicted $\Delta \Delta G$ and in 30% of the proteins have mutations that could be considered as stabilizing as well as destabilizing. We have also found that the consideration of conformational diversity notably increases the correlation of estimated $\Delta \Delta G$ with those
obtained experimentally. However, it is interesting to note that mapping those mutations which better correlates with experimental data in 70% of the mutations studied occur in a single conformer. This could indicate that the experimental determination of $\Delta G$ could have a bias towards those conformers predominating in the chemical equilibrium. Our results indicate that the use of conformational diversity is a key issue to understand the effects of mutation in protein function and possibly in the origin of diseases.

References


