On the conformational diversity of proteins and its relationship with biological properties

Alexander Monzón, Ezequiel Juritz and Gustavo Parisi

Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bs. As. Argentina

Protein native state is better represented by an ensemble of conformers in equilibrium describing the conformational diversity or dynamism of a protein [1]. Protein conformational diversity is a key feature to understand essential properties of proteins like function [2-4], enzyme and antibody promiscuity [5, 6], enzyme catalytic power [7], signal transduction [8], protein-protein recognition [9] and the origin of new functions [10]. Protein conformers describing the native state of a protein exist in a dynamic equilibrium that change in response to the presence of ligands such as substrate or allosteric modulators shifting the relative conformational population [11, 12]. Despite that, the characterization of the ensemble of conformers in equilibrium, involving the study of the structural and thermodynamic features of each individual conformer, represents a major challenge to overcome. Crystallographic structures of the same protein obtained in different instances can be considered as representatives conformers of protein native state. This view is supported by the correlation found between the observed structural diversity determined by solution experiments such as NMR measurements and those coming from crystallographic structures of proteins obtained in different conditions [13-19]. Furthermore, significant correlations were found when computational methods, such as molecular dynamics, were used to simulate protein dynamism and then compared with NMR structures [20, 21].

We have previously developed a database of protein conformational diversity (PCDB; http://www.pcdb.unq.edu.ar[22]) based on CATH database structural hierarchy[23]. PCDB consist of redundant collections of domain structures crystallized in different instances. However, some protein biological properties are related not with domains, but with the whole protein. In order to study how biological properties of proteins are associated with the extension of their conformational diversity, in this work we present the generation of a protein conformational database based of entire proteins as derived from PDB[24]. For this purpose we recruited the redundant collection of crystallized structures from 3,600 monomeric proteins, accounting a total of 12,221 structures, representing different conformers for each corresponding protein (an average of 3.4 different conformers per protein). An all vs. all structural alignment between the corresponding conformers of each protein was done for each protein using different programs (TMscore[25], Profit[http://www.bioinf.org.uk/software/profit/], Mammoth[26]). Using these alignments we obtained that the average RMSD between conformers of a same protein is 0.7Å and a maximum of 38 Å. This average increases notably when considering only those pairs crystallized under different conditions, as ligands (1.9Å), mutations (1.0Å), pH (3.8Å), and temperature (4.1Å), among other. By cross linking our proteins with several servers and databases we recruited a broad spectrum of biological and physical-chemical information (as taxonomy, GO terms, presence of ligands, presence of mutations, etc.). Our dataset include 866 taxonomic units, where 23% of the proteins correspond to humans and 7% to E. coli. Using GO terms, different interesting correlations were found between the extension of conformational diversity and protein function, location and process. One interesting result is that the conformational diversity of thermophilic organisms, even when crystallized under different conditions, shows an average maximum RMSD of 0.8Å which is significantly minor compared to the rest of our data set. Also, significant differences at the extension of conformational diversity were observed for proteins with same GO terms at the process level or even at the same molecular function for well populated proteins. In some of these examples taxonomic specific differences were found.

We think that this extension of PCDB, through incorporation of whole proteins instead of protein domains, could help us in the understanding of how conformational diversity is related with function, structural folds and even with sequence divergence during evolution.
References