Distribution and extension of protein conformational diversity
Ezequiel Iván Juritz, Nicolás Palopoli, Sebastián Fernández Alberti, Gustavo Parisi
Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes, Buenos Aires, Argentina

Background
Protein conformational diversity is increasingly recognized as an essential field to understand protein function and evolution. Although several methods have been developed to simulate proteins motions, we are now interested in analyzing this information from proteins structure databases.

To explore the distribution and extension of protein motions, here we present a curated database of conformational diversity constituted by 36,581 domain structures as derived from CATH database. To study conformational diversity using crystallographic structures, we search for proteins that has been crystallized at least two times in different conditions. For each protein we calculate different structural similarity scores (RMSD, GDT, MAXSUB, LG) between all the structures found for each particular protein. We registered the maximum dissimilarity value based in the parameters mentioned above, as a measure of conformational diversity for each protein. We then associate this value with diverse information coming from different databases in order to find structural, functional and physico-chemical factors correlated with the measures of the movements.

Results and Conclusions
This analysis allows us identifying those factors that promote the expression of the conformational diversity of a protein and those that, on the contrary, constrain the movements of the polypeptide. Among the factors we evaluated are pH variations, presence and type of mutations, presence and type of ligands, changes in oligomeric state, presence and amount of disulphide bonds, taxonomy and structural classification according to CATH. We then split the database in a set of protein families that do not present any factor to promote the expression of the conformational diversity (same protein crystallized in the same condition different times) and one that contains at least one structure with a factor possibly responsible of the structure diversity (for example at least one structure with a bound ligand and one without or with a different one). Both distribution showed to be different according to Mann-Whitney test, with Z-score = 20 (Figure 1).

Within the group of proteins that express conformational diversity, it is possible to discern those factors with higher incidence on it expression and to infer it possible causes.

Figure 1. Discrepancy in Conformational Diversity between set of proteins crystallized several times under different and set of proteins crystallized several times under similar conditions.

Comparing these distributions we estimated a cutoff of 2.6 RMSD to define conformational diversity. Using this value we found that the 16% of the proteins are above this 2.6 RMSD cutoff and have up to 25.6 RMSD units.

Oligomeric state variations, pH variations, presence and type of mutations, presence and type of ligands, are some of the features studied, in order according to its relevance, that predispose the protein to express its conformational diversity.