Comparison of proteins flexibilities by self-organizing maps: the test case of SH3 and its mutants

Fraccalvieri D[(1)], Pandini A[(2)], Bonati L[(1)]

[(1)] DISAT, Università degli Studi di Milano-Bicocca, Milano, IT.
[(2)] Division of Mathematical Biology, National Institute for Medical Research, London, UK.

Motivation

Conformational motions in proteins are often essential for biological function. Whatever is the source of flexibility information, experimentally determined structures or computational simulations, one challenging problem is the development of accurate methods for comparing flexibilities of different protein domains. These can be useful in different cases: for mutants of a protein or proteins belonging to the same family, to observe the influence of limited modifications on the biological function; for proteins with different structures, to analyze functional relationships among evolutionary divergent proteins. While in the former case the comparison could be driven by structural superimposition, in the latter, a set of new techniques should be developed to run comparisons independently of the structural alignment, in order to avoid errors from incorrect alignments. To this purpose in this study we present a use of self-organizing maps (SOM) in the analysis of molecular dynamics (MD) data. The method was tested on the Src-Homology region-3 (SH3) domain and a group of its single mutants[1,2].

Methods

For each structure, after a minimization with steepest descent, MD simulations were run using the GROMACS[3] (version 3.3.3) package. All the structures were inserted in a SPC water octahedral box, with periodic boundary conditions. Simulations of 40ns were performed with the Gromos 43a2 force field, in the NPT ensemble, with a time step of 2fs. The extraction of the data on local flexibility was performed after essential dynamics (ED) analysis.[4] This technique, based on principal component analysis (PCA) of conformational ensembles, is aimed to extract informative direction of motion in a multidimensional space. SOM analysis is a powerful method for visualization and treatment of high dimensional data, based on an unsupervised Neural Network approach[5]. A SOM consists of a low dimensional (usually one- or two-dimensional) grid of so-called neurons, and a model vector associated. Each model vector represents a group of similar data vectors. The SOM analysis was performed in five phases:
1) Selection of data representation: each domain conformation in a MD trajectory, that represents a data vector, was synthesized by geometric descriptors.
2) Selection of map structure: a sheetlike (non-periodic) hexagonal grid of neurons was chosen.
3) Initialization of the model vectors.
4) Training of the map: a length of 300 epochs was chosen (i.e. each data vector was read by the map 300 times).
5) Analysis of the trained map, performed by using the k-means clustering algorithm.

Results

The test-case includes the wild-type a-spectrin SH3 (spc-SH3) domain (PDB code: 1SHG) and a group of its single mutants. The SH3 domain is a small protein module (about 60 residues) that mediates protein-protein interactions in a large variety of cellular processes. Several studies demonstrated the importance of conformational dynamics in defining the binding specificity of SH3 domains. The mutants employed in this study are: R21A, R21G, N47G, N47A, A56G, and A56S. These mutations involved functional and structural change.[1,2] Initially, a comparison of flexibilities was driven by standard sequence and structural alignment followed by annotation of the Root Mean Square Fluctuation (RMSF) for each residue in the equivalent positions. Comparison of the RMSF profiles demonstrated that it is possible to directly correlate the dynamical behavior with experimental information on both stability and functionality of these domains. In a second time, a SOM analysis was performed, employing a set of 10000 conformations extracted from each MD simulation. SOMs allow to compare equal dimension vectors and in this study they were used to compare both the conformations of each mutant and the conformations of the whole SH3 test set. The former analysis confirmed that SOMs are able to analyze a large sample of conformations taken from a MD trajectory and to extract clusters representing relevant regions in the conformational space. The latter analysis confirmed that this method allows

to compare data from different simulations and to identify clusters. Each cluster can be explored by one or more mutants during the MD. This allows to easily recover information about similarities in the domain flexibilities. Future directions involve the extension of SOM analysis to comparison of proteins flexibilities without a preliminary structural superimposition. To this end a different approach to represent protein conformations is needed.

References

[1] Casares S. et al., Protein 2007, 67:531-47.
[2] Casares S. et al., FEBS Letter 2003, 539:125-130.
[3] Lindahl E. et al., J. Mol. Mod. 2001, 7:306-317.
[4] Amadei A. et al., Proteins 1993, 17:412-25. 5. Kohonen, Self-organizing Maps, Springer Series in Information Science, Vol. 30, 3rd ed. Springer, Berlin, 2001.

Contact : d.fraccalvieri@campus.unimib.it