A Structural Alphabet to study Protein Dynamics

Pandini A[1], Fornili A[2], Kleinjung J[1]

Motivation

Protein structures have a hierarchical and partially redundant nature characterized by repeated local conformations. This evidence has inspired the development of several computational methods for Structural Bioinformatics. In particular, structure prediction and protein design have been successfully approached by strategies based on “fragment-assembly”, using culled collections of local structures as building blocks. These collections can be large target-oriented libraries or small sets of carefully selected representatives. The latter are usually called Structural Alphabets. They include less than 30 fragments and form a bridge between the string-oriented methods of sequence analysis and the coordinate-oriented methods of protein structure analysis. [1] The potential of Structural Alphabets has been exploited in structure mining, prediction and design, but not yet for the analysis of molecular motions and conformational transitions. [1] Indeed, previously proposed alphabets were derived with the aim of being accurate in approximating known structures but overlooking stability in the description of their dynamics. Conversely, the ability to correctly capture protein dynamics is now of great interest. The current understanding of protein folding suggests that the native state is better described by an ensemble of alternative substructures in thermodynamic equilibrium, instead of a single well-defined structure. In this framework functional activation is driven by a population-shift among the accessible substructures. [2] In view of this, we developed a new Structural Alphabet specifically suitable for application in studies of protein dynamics. [3]

Methods

A Structural Alphabet has been derived by clustering all four-residue fragments from a high-resolution subset of the Protein Data Bank and extracting the high-density states as representative conformational states. Each fragment was described by three internal angles calculated on the C-alpha representation of the protein. All dataset fragments were mapped as points in a three-dimensional space of these internal angles and clustered with the density-based method OPTICS (Ordering Points To Investigate the Clustering [4]). Cluster representatives were extracted in the order of decreasing density, which is equivalent with decreasing importance, and the unique data ordering corresponds to a minimum distance path, providing a gradual interconversion among the selected states. The Structural

[1] Division of Mathematical Biology, MRC National Institute for Medical Research, London [2] Randall Division of Cell and Molecular Biophysics, King’s College London, London

Alphabet was assessed in terms of its accuracy in reconstructing protein structures. Its suitability in describing protein dynamics was evaluated by encoding a test set of conformation ensembles generated by tCONCOORD [5] and by measuring the correlation between local flexibility (as atomic root mean square fluctuations) and encoding variability (as Shannon entropy). A performance comparison was made with other Structural Alphabets of four-residue fragments.

Results

The derived Structural Alphabet includes 25 states, equivalent to conformational attractors centred on the highly populated regions of the conformational space. The assessment on the test set showed that not only proteins can be reconstructed within the experimental uncertainty, with an average RMSD of 0.70 ± 0.11 Angstrom, but also that the average local flexibility of an ensemble of structures is correctly captured by the Structural Alphabet encoding (Spearman correlation coefficient 0.89 – 0.92). Thus, the Structural Alphabet has the novelty of combining high accuracy in reconstructing protein structures and high stability in encoding ensembles of protein structures from molecular simulations. The selected structural letters correctly describe local conformational states and their transitions. Recently, we started developing a model of allosteric modulation based on the analysis of correlated fragment motions. Preliminary tests were run on a selected number of cases from a benchmark of allosteric proteins. [6] The results suggest the possibility to describe signalling in proteins through the analysis of local deformations

Availability

http://mathbio.nimr.mrc.ac.uk/wiki/Software

Contact e-mail

alessandro.pandini@gmail.com

Supplementary information

References

[1] Offmann B., Tyagi M., de Brevern A.G. “Local structures”, Current Bioinformatics. (2007) 2:165-202. [2] Smock R.G., Gierasch L.M., “Sending Signals Dynamically.”, Science (2009) 324:198-203. [3] Pandini, A., Fornili, A., Kleinjung, J., “Structural alphabets derived from attractors in conformational space.”, BMC Bioinformatics (2010) 11:97. [4] Ankerst M., Breunig M.M., Kriegel H.P., Sander J. “OPTICS: Ordering Points To Identify the Clustering Structure”, In Proceedings ACM SIGMOD, June 1-3, 1999, Philadelphia, PA, USA, ACM Press 1999:49–60. [5] Seeliger D., Haas J., de Groot B.L., “Geometry-based sampling of conformational transitions in proteins.” Structure (2007) 15:1482–1492. [6] Daily M.D., Gray J.J., “Local motions in a benchmark of allosteric proteins.”, Proteins: Structure, Function, and Bioinformatics (2007) 67:385–399.