A new procedure to detect similarities among distant homologous proteins based on the comparison of domain flexibilities

Pandini A (1), Mauri G (2), Bonati L (1)

(1) Department of Environmental Sciences, University of Milano-Bicocca, Milano
(2) Department of Informatics, Systems and Communication,
University of Milano-Bicocca, Milano

Motivation

Due to the improvement of molecular simulation codes and the availability of faster computational resources, a large amount of data on protein dynamics has been generated in the latest years. It is of great interest to investigate the informativeness of the data on dynamics and to effectively assess their usefulness in approaching and solving some classical problems of protein-protein comparison. When tertiary structures are available, the employment of structural alignment, supported by accurate statistical estimates, allows to detect similarities and derive accurate alignments. However the accuracy of the most sensitive methods remains comparable to reliable sequence-based methods, with a similar tendency in reporting false-positives in the case of difficult structural comparisons. This evidence suggests the need to employ some kind of additional information to improve protein comparison. Our proposal is that protein flexibilities derived from molecular dynamics simulations could be promising candidates in supporting detection of similarities among distant homologous proteins. This could be achieved through annotation of the dynamical properties of proteins with known structures and development of a fast and reliable procedure for large-scale comparisons. The proposed procedure could enhance homology search and alignment, as well as improve function detection and annotation.

Methods

The use of CONCOORD as a fast conformational sampling method is proposed and its reliability is assessed by comparing the results with those from Molecular Dynamics simulations. Essential Dynamics analysis is employed to extract a meaningful subspace of informative motions from the ensembles of structures generated by CONCOORD and the Root Mean Square Fluctuation (RMSF) of alpha carbons in the essential subspace is employed as a measure of the local flexibility of protein domains. On the basis of this representation, a synthetic index of similarity between domain flexibilities is proposed and the informativeness of this index is verified.

Results

To assess the reliability of the procedure, the dynamics of a collection of protein domains from a ASTRAL/SCOP40 fold is analyzed and the possibility to identify relationships at the family level on the basis of the dynamical features is discussed. The overall picture obtained is in good agreement with the SCOP classification, and suggests the presence of a distinguishable familiar trend in the flexibility profiles. Moreover the results obtained also support the complementarity of the dynamical and the structural information. The results of this first test support the hypothesis that the additional level of information provided by flexibility annotation, inaccessible by simple structural comparison, can be employed to detect functional similarities otherwise unrecoverable.

Contact email: alessandro.pandini@unimib.it