ResiCon: a method for the identification of dynamic domains, hinges and interfacial regions in proteins.

Maciej Dziubinski, Paweł Daniluk* and Bogdan Lesyng

Department of Biophysics and CoE BioExploratorium, Faculty of Physics, University of Warsaw, Żwirki i Wgury 93, 02-089 Warsaw, Poland, and Bioinformatics Laboratory, Mossakowski Medical Research Centre, Polish Academy of Sciences, Pawliczkiego 5, 02-106 Warsaw, Poland

pawel@bioexploratorium.pl

Introduction

With current simulation methods and computational power it is possible to acquire Molecular Dynamics (MD) trajectories in microsecond time-scales in which major conformational changes can be observed. Because of the large amount of data produced in such simulations, it is necessary to use fast, automatic procedures to track and describe structural motions. We propose a new approach called ResiCon which performs a data-mining analysis of MD trajectories, and finds dynamic domains (quasi-rigid parts). ResiCon uses a time-generalized concept of a contact between amino-acid residues to construct a virtual scaffold characterizing the rigidity of a protein (see [1] for the description of local structural properties, including contacts). Stiffness of edges reflects variability of corresponding contacts observed during simulation. Next, ResiCon performs a spectral clustering, treating resulting clusters as putative quasi-rigid regions. We propose employing ResiCon to study proteins with flexible structural domains. ResiCon is available as a web server at http://dworkowa.imdik.pan.pl/EP/ResiCon

Contact between residues

We say that two residues are in contact, if for at least one configuration in the trajectory the following condition is satisfied:

\[ (d_{il} \leq 6.5\text{Å}) \lor (d_{il} > 8\text{Å} \land d_{il} - d_{ij} \geq 0.75\text{Å}) \]

Geometrical variability of a pair of residues \( i \) and \( j \) is quantified using the following metric:

\[ G(i,j) := \frac{\text{max}}{\text{pairs of states} (\cdot)} \text{RMSD}(E'_{ij}, E_{ij}) \]

where \( E' \) denotes the configuration of a contact between residues \( i \) and \( j \) in the \( r \)th trajectory frame. Small values of geometrical variability correspond to strong contacts.

The algorithm

As INPUT ResiCon requires a sample of representative structures of a protein in the PDB format. For each pair of residues in contact the geometrical variability is calculated. From this, values of parameters \( \alpha \) and \( \beta \) are estimated, and the similarity matrix is constructed. ResiCon carries out a spectral clusterings into 2, 3, ... 10 clusters. For each such partitioning an indicator \( \gamma_{\text{opt}} \) is calculated, which is used to choose the optimal number of clusters, \( k \) (see [2]). OUTPUT of ResiCon contains information on partitioning of the studied protein molecule into quasi-rigid structural parts, which correspond to clusters. Result of a clustering procedure is typically shown as a colorful stripe. This, along with a coloring of the protein structure, provides a clear visualization of the dynamic domains. In the case of the HIV-1 protease, flaps and two larger subdomains emerge when a partitioning into \( k = 4 \) clusters is carried out.

Sequential discontinuity

According to ResiCon, ASN-83 (magenta) belongs to a different cluster (cyan), than its sequential neighbours (yellow). This can be explained by the fact that ASN-83 plugs a cavity in the cyan domain, and it remains there throughout the trajectory.

HIV-1 clustered into 6

The optimal number of clusters (according to ResiCon) is 4. It should be noted, however, that the indicator suggests another interesting partitioning for \( k = 6 \) (right). It distinguishes structures resembling “arms” connecting flaps with core subdomains.

Note that there is a sequential discontinuity in the cyan/yellow regions. Singular residues were assigned to different parts than their sequential neighbours.

References


Acknowledgements

These studies were supported by the Biocentrum- Ochota project (POIG.02.03.00-00-003/09), the research grant (DEC-2011/30/D/NZ2/02004) of the National Science Centre, and the BST 163500/BF (Z. b. 14).