Poster I-5

Protein structure topology comparison based on contact maps

Authors:
Raquel C. Melo (*UFMG*)
Ivan Mazoni (*Embrapa*)
Goran Neshich (*Embrapa*)
Marcelo M. Santoro (*UFMG*)
Wagner Meira Jr (*UFMG*)

Short Abstract: Topologs (Topology Homologous) is a tool which is part of the Star Sting Suite (http://sms.cbi.cnptia.embrapa.br/SMS). It is a structural classification of proteins totally based on the pattern of intrachain interactions. Its precision is higher than 85% for all tested fold types and has been precomputed for all ASTRAL 40 chains.

Long Abstract:

Many protein structures have been determined and reveal that protein molecules can adopt the same fold despite having very dissimilar sequences. The PDB has about 80,000 protein chains, although, the majority of this data is quite redundant in sequence and structure. Chains with more than 30% of sequence in common, generally, adopt the same fold. Proteins fall into families of related folds. What can we learn about protein folding and evolution observing how proteins cluster into families? To address these questions, a suitable measure of the structural similarity between proteins is needed.

In this work, we use a metric proposed in [1] to classify the structures. As we showed, contact maps of electrostatic and hydrophobic interactions and hydrogen bonds are the reliable representation of protein structure, are highly conserved within a fold family and can be used to compare structures with high confidence. We present a tool called Topologs (Topology Homologous) which is part of the Star Sting Suite [2], available at the http://sms.cbi.cnptia.embrapa.br/SMS. To the best of our knowledge, there is no any other available package that uses contacts to classify protein structures.

In this work we used a list of proteins in ASTRAL (40%, version 1.69) and for any particular protein form this ensemble, we calculate and present to a user a list of the 100 most similar chains based on contact maps. The tests showed a precision in fold classification higher than 85%. In Topologs users can submit a set of up to 100 PDB chain identifiers to be analyzed against a specific PDB chain or even against a local file containing a modeled structure in PDB format. This tool can be used to predict family and function of recently discovered proteins, to compare a set of mutants with the wild protein, to study a pattern of contacts which describes a fold family, to enhance the phylogenetic analysis based on the contacts, among other applications.

[1] F.A. Fernandes Jr and C.E.R. Lopes and R.C. Melo and R.L. Carceroni and M.M. Santoro and C.H. Silveira and W. Meira Jr (2004). An Image-Matching Approach to Protein Similarity

Analysis. XVII Brazilian Symposium on Computer Graphics and Image Processing, 17-24.

[2] G. Neshich and I. Mazoni and S.R.M. Oliveira and M.E.B. Yamagushi and P.R. Kuser-Falcão and L.C. Borro and D.U. Morita and K.R.R. Souza and G.V. Almeida and D.N. Rodrigues and J.G. Jardine and R.C. Togawa and A.L. Mancini and R.H. Higa and S.A.B. Cruz and F.D. Vieira and E.H. dos Santos and R.C. Melo and M.M. Santoro (2006). Star STING Server: A multiplatform environment for protein structure analysis. Submited.