

Description and comparison of protein quaternary structures

Majid Akbary-Gasany and Jean-Christophe Nebel*

Faculty of Computing, Information Systems & Mathematics, Kingston University,
Kingston-Upon-Thames, KT1 2EE, UK

j.nebel@kingston.ac.uk

Abstract

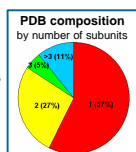
Oligomeric proteins are proteins which are composed of two or more polypeptide chains, or subunits. The arrangement of subunits is known as quaternary structure and varies in term of composition and topography. These proteins are essential to key biological processes such as proteolytic cleavage, protein folding and translation. Moreover, it is suggested they provide evolutionary benefit. Therefore, their systematic study could contribute to protein annotation, a better understanding of some biological processes and the evaluation of organism complexity. In order to facilitate comparisons of protein quaternary structures, we developed a tool which automatically analyses the 3D structure of oligomeric proteins to extract key information such as their composition, topography and topology.

Rational

▪ 43% of proteins in the PDB** [1] are composed of several subunits

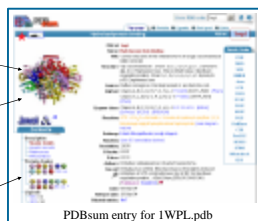
▪ Many biological functions rely on proteins in an oligomeric state

▪ Analysis of quaternary structures requires a new representation including subunit positions, interactions and homogeneity



Interactions between subunits are unclear
No information regarding possible symmetries

Homogeneity status of the protein is not specified



Cyclic symmetry detection

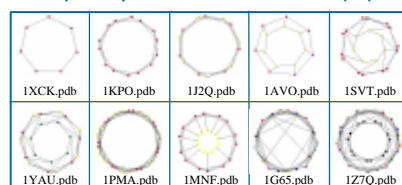
Method

▪ According to eigenvalues of subunit positions, a reference plane is defined

▪ Based on the reference plane, a cyclic string matching algorithm [3] is applied on 3D positions and type of subunits

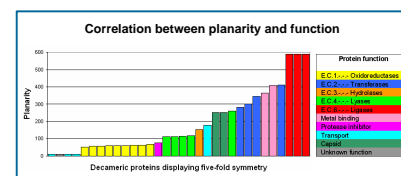
▪ Rotational symmetries based on the 3D geometry and/or type of subunits are detected

Examples of proteins with 7-fold rotation (C7)



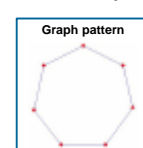
Towards function prediction

▪ Protein function may be correlated to symmetry, planarity and homogeneity



▪ Graph patterns may be associated to specific functions

▪ 7-node cyclic graph with beta-sheet edges



34 hits in PDB:

- 24 chaperone proteins (C7)
- 9 SM-like proteins (C7)
- 1 DNA replication protein (C1)

Quaternary structure representation

Data extraction from PDB file

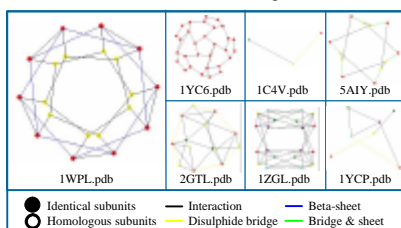
▪ Each subunit belonging to the biologically functional molecule is represented by the 3D position of the centre of mass of the atoms of the associated polypeptide chain

▪ Homogeneity of sequences of the subunits is evaluated by pairwise alignment [2] - chains are then clustered according to their similarity status (identical, homologous or unrelated)

▪ Atomic interactions between subunits are detected including disulphide bridges (SS) and inter chain beta-sheets (ICBS)

2D cartoon generation

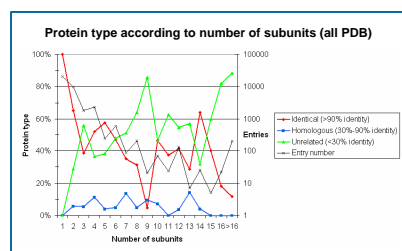
- Subunit positions are projected in the most suitable 2D plane according to eigenvalues of their positions
- Subunits are connected according to interactions



It allows visual analysis of protein topology

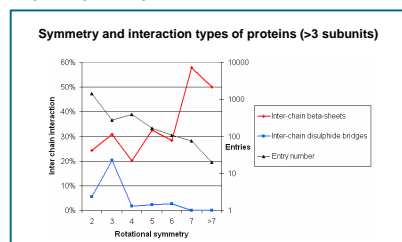
Results

Oligomer homogeneity



Oligomers are either composed of identical (>90% identity) or unrelated subunits (<30% identity)

Cyclic symmetry & interactions between subunits



Majority of C7 proteins contain beta-sheets (ICBS)

Future work

Oligomer comparison

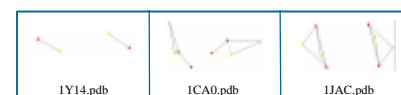
▪ Distance metric for comparison between graphs of oligomers

Pattern/motif recognition

▪ Detection of common subgraphs

Detection of multiple biological molecules

▪ Detection of disconnected graphs



References

- [1] Berman, H.M. et al. (2000) The Protein Data Bank, Nucleic Acids Research, 28, 235-242
- [2] Needleman, S.B. and Wunsch, C.S. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins, J. Molecular Biol., 48:443-453
- [3] Liadós, J., Bunke, H. and Martí, E. (1997) Finding rotational symmetries by cyclic string matching, Pattern Recognition Letters, 18(14):1435-1442

** All data based on the PDB release of 15th August 2006