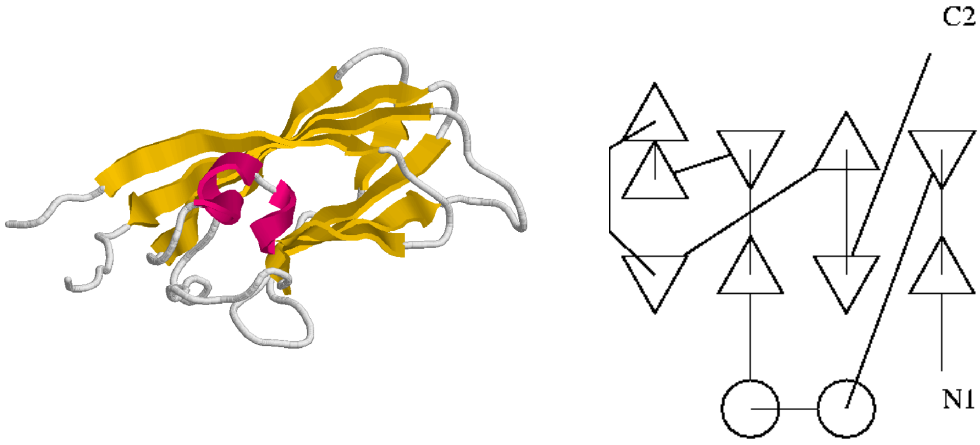
Protein topology and structural alignment

Ioannis Michalopoulos, Amy Williams and David R Westhead

An understanding of the similarities and differences between protein structures is very important for the study of the relationship between sequence, structure and function, and for the analysis of possible evolutionary relationships. This has led to the need for computational methods of structure comparison, and algorithms for searching structural databases. Furthermore, the rapid increase in the size of structural databases means that search and comparison techniques should be both fast and accurate.

Topological abstractions of protein structures consider a sequence of secondary structure elements (SSEs), i.e. helices or strands, together with relationships like spatial adjacency within the fold and approximate orientation, neglecting details like the lengths of SSEs and loops. This simplification enables the implementation of very fast algorithms. TOPS cartoons are pseudo-2D schematic abstractions, where the third dimension is implied, since SSEs are considered to have an approximate direction of "up" or "down" (depending on the way the lines connecting the symbols are drawn). Adjacent strand pairs are connected by H-bonds, being parallel or antiparallel. Chiralities between parallel strands are also implicit. Topological information relating sequential order and relative spatial position is easier to deduce from the TOPS cartoon than the three-dimensional structure.

As part of this project, we have been developing a database of protein structures modelled at the level of topology and enhanced with amino-acid sequence information, ligand binding information and EC number where appropriate. We have developed a rich scheme for describing these structures and implemented this in a MySQL relational database held at Leeds, which we are now in the process of populating. In addition at City University we are developing a suite of search, machine learning and structure comparison algorithms to



Structure of 1stm views using Rasmol and TOPS. 1stm has a β jelly-roll structure. It consists of four Greek key motifs that adopt an eight-stranded beta sandwich structure.

compute over this data.

Protein domain superfolds are unusual in that they support a wide variety of different biological and biochemical functions. This suggests multiple evolutionary origins. To support this, there was no evidence that SCOP superfamilies that share the same fold have a common evolutionary origin. However, PSI-BLAST provided statistically reliable evidence indicating that at least 12 of the 23 SCOP (α/β) TIM barrel superfamilies might share a common origin. Here we extend this work to the β Jelly Roll superfamilies. As well as using PSI-BLAST, a structural approach was used. Hidden Markov Models were built for each superfamily based on a multiple structural alignment. The results did not provide any evidence to suggest that any of the β Jelly Roll superfamilies share a common ancestor.

Another conclusion of this work was that existing structural alignment packages struggled to align diverse β Jelly Roll structures, in certain cases within the same superfamily. The aim therefore was to create a method that will do this. The method uses topological descriptions of proteins. It uses pattern matching software that produces secondary structure equivalencies and uses these to produce the best possible structural alignment. The output is a structure-based sequence alignment and a 3D superposition of the structures. Current work aims to optimise the alignment using a genetic algorithm.

Collaborators

David R Gilbert, City University, London.
Gilleain M Torrance, City University, London.

References

Gilbert D., Westhead D., Viksna J. & Thornton J. (2001) A computer system to perform structure comparison using TOPS representations of protein structure. *Computers & Chemistry* **26**, 23-20.

Funding

We thank BBSRC and EPSRC for funding.