Protein surface descriptions and function

James Bradford, Nicola D. Gold, Steven J. Pickering and David R. Westhead

Macromolecular docking
We are currently developing a novel machine vision-based algorithm to help solve the "docking problem", which can be defined as follows: given the coordinates of two proteins, predict whether they interact and the structure of their complex. Our method uses graph theory to identify regions of shape complementarity between two protein surfaces, analogous to finding areas of similarity in the protein surface comparison project (see below). Recently, we have been looking at ways of using biological information to facilitate our algorithm and one strategy has been to predict the binding sites on the two proteins. Most binding site prediction methods assume that the largest cluster of highly conserved residues on the surface of the protein corresponds to the interface. With this in mind, we have carried out a thorough and systematic sequence-structure study of the pattern of conservation on the surface of seven protease-inhibitor docking test cases and related the results to binding site prediction. The proteases all displayed a significantly higher concentration of conserved vertices at the interface compared to the exposed surface in contrast to the inhibitors where the difference between interface and exposed surface was much less distinct and varied from case to case. Indeed, some inhibitors had a greater proportion of unconserved vertices at their interface than the exposed surface. We conclude that binding site prediction methods relying on a highly conserved interface would succeed when applied to our protease test cases but complications could arise with the inhibitors.

Protein surface comparisons
Locating a protein sequence or structure common to two or more proteins often implies that these proteins possess similar functionality. With this in mind, protein function is usually predicted through protein sequence or backbone structure comparison. However, protein functions such as catalysis or molecular recognition occur predominantly on or near the protein surface. Our current work therefore concentrates on protein surface matching, a novel technique that has the power to reveal further functional relationships between proteins, not necessarily apparent from comparisons of the protein fold, by using protein surface comparisons to identify similarly shaped regions of surface.

We have identified residue coordinates important in protein function and stored this information in a database SITESDB containing approximately 50,000 functional sites (ligand binding and/or active sites). This information has also been used to create a database of protein functional site surfaces. We have also developed three methods to compare a query functional site in a pair-wise fashion to the database of known templates all based on techniques in graph theory. In addition, to reduce program execution time the methods have been implemented using MPI (Message Passing Interface) for use on parallel machines such as a Beowulf cluster.

In each of the methods the first step is to find identifying features to be used by the graph comparison algorithm.

Method 1: Each functional site residue is represented with a co-ordinate at the residue's centre of mass.
Method 2: Following the generation of a Connolly protein surface for a functional site each functional site residue occurring on the protein surface is represented with a co-ordinate on the protein surface.
Method 3: Following the generation of a Connolly protein surface for a functional site a co-ordinate is used to represent certain shape properties (concave, convex, radius of curvature) or charge or hydrophobicity.

The co-ordinates of these features are then represented by vertices in mathematical graphs. Common vertices between two graphs are found using distance criteria and clique detection algorithms. The more vertices the two graphs have in common, the more similar the functional sites.

We believe that residues that lie on the protein surface at a functional site are more useful in assigning biochemical function than the use of the whole sequence. Using these residues allows for the detection of functional sites from evolutionarily related proteins where the sequence remains conserved at the functional site (method 2). For detecting more distantly related, or unrelated proteins, with similar function, we deduce the overall shape of the functional site from the orientation and position of various concave and convex regions within the site (method 3).

Early results indicate that we can successfully match related proteins within the same SCOP family as the query (Fig. 1). There are some unrelated matches that will be investigated further. We are also able to predict the function of sites that have not yet been classified in SCOP.

Fig. 1: Selected surface features of three eukaryotic serine proteases are shown in blue.

Trypsin (1btx)  Elastase (1bma)  Chymotrypsin (2gct)

**Publications**

**Funding**
We thank the BBSRC for funding this work.