Q-SiteFinder: An online tool for ligand-protein binding site prediction

Alasdair T. R. Laurie* and Richard M. Jackson

University of Leeds

1 Introduction

Identifying the location of ligand binding sites on a protein is of fundamental importance for a range of applications including molecular docking, *de novo* drug design and structural identification and comparison of functional sites. Q-SiteFinder is a new energy-based method for predicting protein-ligand binding sites (Laurie and Jackson, 2005). The method analyses interaction energies of a methyl probe with a protein using software developed by Jackson (2002). Probes with favourable interaction energies are retained and clusters of these probes are ranked according to their total interaction energies. The energetically most favourable cluster is then ranked first. When tested on 134 ligand-bound protein structures, there was at least one successful prediction in the top three predicted sites in 90% of cases. The method was also found to be successful on unbound protein structures.

Q-SiteFinder uses several separate procedures to perform ligand binding site prediction described in detail previously (Laurie and Jackson, 2005). Briefly, these processes include the addition of hydrogen atoms to the protein as described by Jackson *et al.* (1998). It then calculates the non-bonded interaction energy of a methyl probe (-CH_3) with the protein at each position on a 3D grid of resolution 0.9Å, using the GRID force field parameters as previously described (Jackson, 2002). The probes with the most favourable binding energy are retained based on an interaction energy threshold of -1.4 kcal/mol. Individual probe coordinates are then clustered according to their spatial proximity, and the total interaction energies of probes within each cluster are calculated. The probe clusters are ranked according to their total interaction energies, with the most favourable being identified as the first predicted binding site. The speed of the overall process is dependent on protein size, but it is usually 10-15 seconds on the current server (1.8GHz CPU).

A variety of useful information is calculated by Q-SiteFinder and is shown in figure 1. Some of the information is site-dependent, and changes dynamically according to the selected site(s). The predicted sites can be represented by the methyl probe clusters or by the atoms that lie within 5Å these clusters. These atoms are listed, and are individually selectable for display using CHIME. The predicted ligand-binding sites are colour coded for ease of identification.

Minimum and maximum coordinates are calculated when predicted sites are selected. They describe the minimum binding box around the selected probe clusters. This binding box is extended by 5Å at each of the box faces. A ligand binding site can be described by several adjacent predicted sites, so selection of the appropriate sites will give binding box coordinates that can usefully be applied to docking and *de novo* design studies.

The precision value is a measure of how well a predicted site maps onto ligand coordinates. The term "precision" used here defines the percentage of probe sites in a single cluster that are within 1.6Å of a ligand atom. We used a precision threshold of 25% to define success when we tested Q-SiteFinder on 134 proteins. The average precision of first predicted sites was 68% excluding total failures (i.e. a precision of 0%). We suggest that such predicted sites are precise enough to enable accurate identification of the atoms that form the ligand binding site while

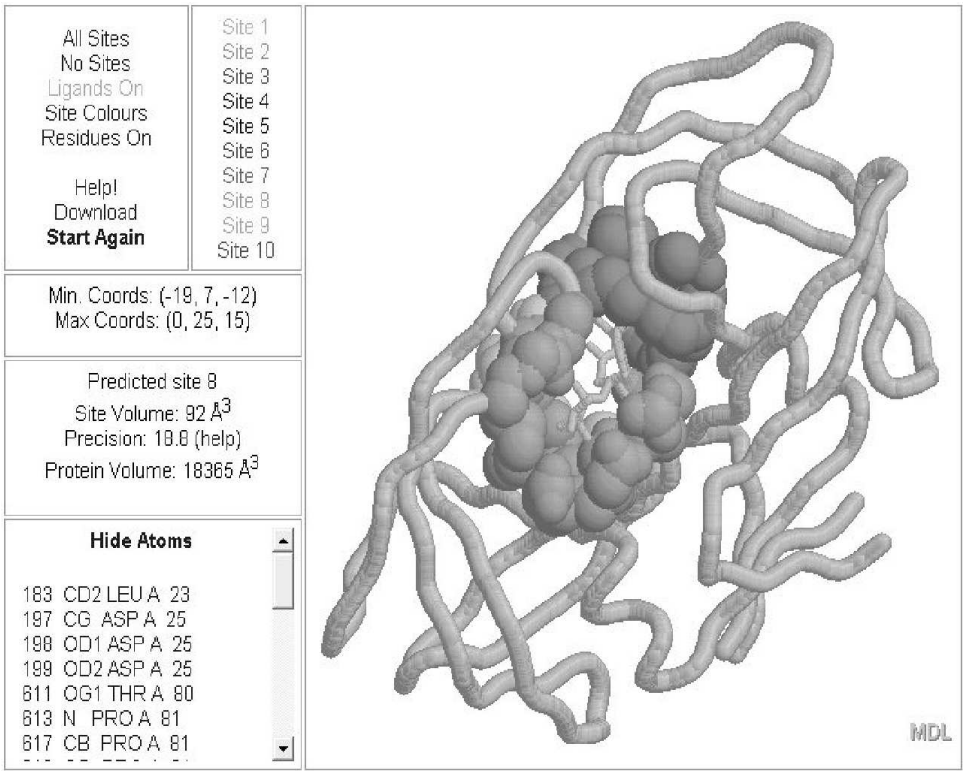

Figure 1: The CHIME-based output from Q-SiteFinder.

restricting the number of atoms that are false-positively identified as being part of a ligand binding site.

We have presented an online tool, Q-SiteFinder, for ligand binding site prediction that is based on determining energetically favourable binding sites on the surface of a protein. The tool provides a convenient first step in the drug design process and for structural identification and comparison of functional sites.

Availability: `http://www.bioinformatics.leeds.ac.uk/qsitefinder`

Contact: jackson@bmb.leeds.ac.uk

References

Jackson, R. M. (2002). Q-fit: a probabilistic method for docking molecular fragments by sampling low energy conformational space. *J. Comput. Aided Mol. Des.*, **16**, 43-57.

Jackson, R. M., Gabb, H. A. and Sternberg, M. J. (1998). Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.*, **276**, 265-285.

Laurie, A. T. R. and Jackson, R. M. (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, in press.