Predicting Ligand Binding Sites of Uncharacterized Protein

Mizuki Morita
mizuki@bi.a.u-tokyo.ac.jp

Shugo Nakamura
shugo@bi.a.u-tokyo.ac.jp

Kentaro Shimizu
shimizu@bi.a.u-tokyo.ac.jp

Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

Keywords: binding site, active site, protein-ligand interaction, van der Waals potential energy

1 Introduction

We have developed a computational algorithm for the accurate identification of ligand binding sites on three-dimensional protein structures.

A variety of computational methods for identifying and characterizing binding sites have been proposed over the past two decades (see [1-3] for a recent review). Because the ultimate goal of the binding site prediction methods is to find active sites on uncharacterized structures, it is important to test and validate the algorithm on data sets of the unbound (apo) structures. Most of these methods, however, only tested ligand-bound (horo) structures. We developed a method called StickyFingers, which involves a calculation of the van der Waals interaction energy between protein and probes placed around the protein surface and clustering the probes with attractive interaction to find most favorable locus, and performed a realistic test using unbound (apo) structures. Our method is simple, accurate and fast enough to predict protein-ligand binding site of uncharacterized protein structures.

2 Method and Results

2.1 Method

In our method, called StickyFingers, the protein surface is coated with a layer of probes to calculate van der Waals interaction energies with the protein. Probes with favorable interaction energies are retained and clusters of these probes are ranked according to the number of probes in a cluster or their total interaction energies. The largest cluster or the energetically most favorable cluster is then ranked first.

Figure 1: Steps of probe clustering with double thresholds (PDB ID: 2RTA).
(a) seed clusters (gray), (b) accreted clusters around the seeds, (c) ranked clusters (from yellow to blue), (d) the first ranked cluster (yellow) with superimposed ligand (orange)

2.2 Results
The success rate in the first predicted site was 91% (32/35) for the ligand-bound state and 86% (30/35) for the unbound state. The percentages of proteins with at least one success in the first two sites were 100% (35/35) for the ligand-bound state and 94% (33/35) for the unbound state.

A number of papers predicting locations of binding sites have been published, and accuracy evaluation methods differ from paper-to-paper. It is quite difficult to compare ours with all those methods. In this work, we used Laurie and Jackson's dataset and measure[4] to evaluate our approach and compare our results with those of Q-SiteFinder[4] and Pocket-Finder[4]. The results are summarized in Table 1.

The average speed of overall process is 0.5 minutes on a Dual 2.5 GHz Apple PowerPC G5.

Table 1: Comparison of results from StickyFingers with those from Q-SiteFinder and Pocket-Finder.

<table>
<thead>
<tr>
<th></th>
<th>ligand in the first site (Precision* ≥ 0.25)</th>
<th>average Precision*</th>
</tr>
</thead>
<tbody>
<tr>
<td>StickyFingers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ligand-bound</td>
<td>0.914</td>
<td>0.797</td>
</tr>
<tr>
<td>unbound</td>
<td>0.857</td>
<td>0.602</td>
</tr>
<tr>
<td>Q-SiteFinder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ligand-bound</td>
<td>0.743</td>
<td>0.739</td>
</tr>
<tr>
<td>unbound</td>
<td>0.514</td>
<td>0.619</td>
</tr>
<tr>
<td>Pocket-Finder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ligand-bound</td>
<td>0.714</td>
<td>0.375</td>
</tr>
<tr>
<td>unbound</td>
<td>0.514</td>
<td>0.354</td>
</tr>
</tbody>
</table>

*Precision is the proportion of probes in a single cluster that could predict locus of ligand atoms correctly (0 < Precision < 1).

3 Discussion

The success of ligand binding site prediction by our simple method, which uses only van der Waals potential energy and requires no information about ligand or substrate, imply an important property that might help to understand the mechanism of ligand binding. Namely, the coupling of a ligand to a protein is to a large extent stabilized not by electrostatic interactions but by van der Waals attractions. It would be the reason why our algorithm works well.

References


