CASP2 Molecular Docking Predictions
With the LIGIN Software

Vladimir Sobolev, 1* Theodore M. Moallem, 1 Rebecca C. Wade, 2 Gert Vriend, 2 and Marvin Edelman 1
1Department of Plant Genetics, Weizmann Institute of Science, Rehovot, Israel
2EMBL, Heidelberg, Germany

ABSTRACT
Seven docking predictions were made with the LIGIN program. In six cases the location of the binding pocket was identified correctly by systematically docking everywhere within the protein structure. In two cases the ligand was docked to within 1.8 Å RMSD of the experimentally determined structure. LIGIN has not been optimized to deal with highly flexible ligands that dock at the surface of proteins. Consequently, in three cases the exposed part of the ligand was docked poorly, although the buried parts were docked well, and made similar atomic contacts with the protein as in the experimentally determined structure. Proteins, Suppl. 1:210–214, 1997. © 1998 Wiley-Liss, Inc.

Key words: molecular recognition; ligand-receptor contacts; complementarity function; ligand flexibility; drug design

INTRODUCTION
Most methods for predicting the structure of ligand-receptor complexes calculate either interaction energy or shape complementarity to estimate ligand fitness (see 1–4 and refs. 1–23 in 5). In our approach, surface complementarity between a ligand and receptor is the guiding principle for ligand docking. 5 Surface complementarity incorporates information about the shape and chemical nature of the atoms of the interacting molecules. 6 The advantages of using surface complementarity are particularly apparent when ligands are docked into spacious receptor pockets (Fig. 1). In such cases, our method performs well because the definition of contact surface 6 allows loose contacts (up to a solvent-separated distance) to be considered and it optimizes favorable contacts, both loose and tight. Our approach provides lists of residues in contact with the ligand and major contacts (including putative hydrogen bonds) between receptor and ligand. These lists permit an analysis of all contributions to the complementarity function and assist in the design of improved ligands.

In the examples suggested by CASP2, we tested our approach for predicting binding pocket location, ligand orientation and the major interactions stabilizing the ligand-receptor complex.

METHOD
For our predictions, we used the LIGIN software for docking of molecules into protein binding sites. The original version of this program has been described. 5 The additional modifications employed for the present predictions include a subroutine to treat ligand flexibility and to identify de novo the location of potential binding sites on a macromolecular receptor. The software calculates a normalized complementarity function (NC), which is given by

\[ \text{NC} = \frac{(S_1 - S_i - E)}{S_{ac}} \]

where \( S_i \) and \( S_1 \) are sums of legitimate and illegitimate contact surfaces between ligand atoms and atoms in the receptor binding site, \( E \) is a repulsion term, and \( S_{ac} \) is the solvent-accessible surface of the uncomplexed ligand. The repulsion term prevents strong interatomic overlaps during optimization of the ligand position. Contacts up to solvent-separated distances (~6 Å) contribute to the NC. Thus, the method is suited for docking of ligands both in loose and tight binding pockets, and is reasonably robust for small, induced conformational changes in the receptor upon ligand binding.

We have divided the atom types into 8 classes: hydrophilic, N and O that can donate and accept hydrogen bonds; acceptor, N or O that can only accept a hydrogen bond; donor, N that can only donate a hydrogen bond; hydrophobic, Cl, Br, I, and all C atoms that are not in aromatic rings and do not have a covalent bond to hydrophilic, donor or acceptor atoms; aromatic, C atoms in aromatic rings; neutral, C atoms that have a covalent bond to at least one hydrophilic atom or two or more acceptor or donor ones; neutral-donor, C atoms that have a covalent bond with only one donor atom; neutral-acceptor, C atoms that have a covalent bond with only one acceptor atom. Legitimacy depends on the complementarity of the contacting atoms (see Table I

© 1998 WILEY-LISS, INC.
We use the complementarity function as an efficient measure to predict ligand position. At this stage, however, it clearly lacks the detail of an energy function and is not very sensitive to the ligand coordinates.

The entire protein was searched for binding sites by dividing it into nonoverlapping cubes with sides of 10–15 Å. Then, a number of random ligand positions and orientations were generated within each cube so that the number of starting points corresponded to a density of 4 per Å³. No constraints were placed on the position of the ligands during optimization. Thus, the center of geometry of the docked ligand sometimes fell outside of the cube searched. Complete searching of a cube took 40–200 minutes on a DEC (TURBO-LASER) SERVER 8200 5/300.

To treat ligand flexibility, we considered all rotamers for small ligands. For large ligands, we could only choose a few rotamers to represent the most important degrees of freedom and the range of conformational variability. Searching was performed using the 6 df (rotational and translational) available to the ligand, extended by the number of ligand single bonds (i.e., the complementarity function depended both on ligand position and conformation). During optimization of the ligand position, the program allowed 10–20° rotations around freely rotatable bonds. Such adjustments were introduced without penalty. To avoid being stuck in a wrong local minimum, several rotamer combinations were tried for every ligand. CPU time limitations precluded a rigorous search of all potential rotamer combinations.

### RESULTS AND DISCUSSION

#### Docking of Methyl α-D-Arabinofuranoside to Concanavalin A (T0013)

Methyl α-D-Arabinofuranoside (Fig. 2) is small and rather symmetric. The small size meant an increased probability of finding several cavities of proper dimension in addition to the correct one. However, the experimenters provided the approximate location of the binding pocket, thus resolving this potential difficulty.

Due to the chemical symmetry, multiple orientations of the ligand within the binding site can be expected during docking. For docking we used the ring conformation given in and considered two different rotamers for the single bonds C9—C10 and C6—O1 (i.e., four conformers in total). Within the binding pocket, there were many ligand orientations with practically equal normalized complementarity values and we submitted the top three (NC = 0.84, 0.83, 0.82). The second one is closest to that determined experimentally (RMSD = 1.4 Å). Residues in contact with the ligand and putative hydrogen bonds are listed in Table I for the predicted and experimental structures.

#### Docking of Pentamidine to Trypsin (T0033)

In the case of the trypsin–pentamidine complex, the ligand (Fig. 3) has a large number of rotamers as there are eight rotatable bonds. The number of rotamers is too large to test each of them by screening the whole protein. We restricted ourselves by examining only the extended structure of the ligand as well as structures differing from it by rotations around the single bonds C11—C12 and C13—C14. The comparison of contacts in the experimental and
best-predicted structure (Table II) shows a high degree of correspondence. The list of residues forming the binding site (i.e., those showing a contact surface area in the column labeled “Surf”) are similar. The solvent-accessible surface of the ligand in the crystal structure was, however, relatively large (about 150 Å², while in the uncomplexed ligand it was 620 Å²) and we failed to correctly predict the position of the solvent-exposed part of the ligand. As a result, we obtained an overall RMSD of 7.2 Å. However, the RMSD for the buried part of the ligand in the best-docked example is about 3.0 Å. The fitness of the buried part is reflected in the correspondence of the stabilizing contacts (HB, A–P and h–h columns) for the experimental and predicted structures. Protein residues shown in bold in Table II are in contact with the buried part of the ligand. Rigid-body docking, using the correct ligand conformation, gives the correct ligand position with NC = 0.69 versus 0.62 found in the flexible docking.

Docking of Amiloride to Trypsin (T0034)

The ligand of the trypsin–amiloride complex is shown in Figure 4. Rotation around the single bond C₃—C₁₀ was allowed during the docking procedure. The best predicted structure (RMSD = 1.8 Å) has a similar, but not identical list of major contacts and hydrogen bonds as the experimental structure (for details, see www page http://sgedg.weizmann.ac.il/casp2/t0034.html). Differences mainly involve the O₁₁ atom. The explanation for this is that we considered only the cis orientation of the O₁₁—C₁₀—N₁₂—H fragment. Docking of the trans orientation would have given a structure closer to the experimental one (RMSD = 0.3 Å, although with NC = 0.76 versus 0.78 for the cis conformation).

Docking of SBB Inhibitor to Pancreatic Elastase (T0036)

Within the context of CASP2, information was provided that SBB inhibitor is covalently bound to the protein. We therefore allowed strong overlap of the ligand with any single protein residue. Using this procedure, both the binding pocket and the residue of strong overlap (Ser 195) were correctly determined. However, our prediction of ligand orientation was incorrect (RMSD = 9.2 Å). Perhaps intramolecular forces at the ligand–protein linkage (which are not considered by LIGIN) play an essential role in ligand orientation.

Docking of Aica-riboside Phosphate to Human Fructose-1,6-bisphosphatase (T0039)

The aica-riboside phosphate ligand is schematically presented in Figure 5. Both rings were considered as rigid. Twelve rotamers were taken into account (3, 2, and 2 for the C₇—C₁₅, C₉—N₂ and C₁₀—C₁₁ bonds, respectively). LIGIN predicted a position for the ligand which is very far from the experimental binding pocket (RMSD = 35 Å). Subsequently, docking of the correct ligand conformation into the known binding pocket gave NC = 0.64 versus 0.70 for the predicted structure. In trying to understand the failure of the program in this case, we noticed that there are two large, illegitimate (hydrophilic–hydrophobic) contacts: N₁₄ with the Cβ atom of Leu30 (3.6 Å apart) and O₁₇ with the Cβ atom of Glu20 (2.8 Å apart). Even reassignment of atom classes for the C atom of the Glu side chain did not lead to prediction of the correct binding pocket. Our method is currently not able to predict locations of binding sites involving unfavorable contacts such as these.

Docking of INH and INI to Trypsin (T0040 and T0041)

In the trypsin–INH and trypsin–INI complexes, the ligands (Figs. 6 and 7) have eight single bonds each. We considered all rotamers of bonds C₂—C₄, N₁—C₂ and N₁₉—S₂₈ (INH inhibitor) or N₂₁—S₃₃ (INI inhibitor). We obtained only 8 rotamers for each ligand by discarding structures having internal bumping. All these structures were used in the searching procedure. For each ligand, part of the molecule in the experimental structure is surface-located: solvent-accessible surfaces in the crystal structures are 200 Å² for INH and 290 Å² for INI, while for the uncomplexed ligands they are 670 Å² and 740 Å², respectively. Analysis of the contact lists (for details, see www pages http://sgedg.weizmann.
ac.il/casp2/t0040.html and http://sgedg.weizmann.ac.il/casp2/t0041.html) revealed that in the protein embedded parts of the ligands the predicted structures occupy the same positions as in the experimental ones, but the surface located parts take different orientations. The resultant overall RMSD for our predictions were 8.0 Å and 7.5 Å, respectively, while for the buried parts, they were a respectable 1.4 Å and 2.1 Å. Docking of the correct ligand conformations gave NC50.63 for the INH and 0.53 for INI, versus NC50.63 and 0.58, respectively, for the predicted structures. Thus, even if we had considered all possible conformations for the ligands we would have had difficulties in predicting surface located parts, where the protein residues form neither pocket nor niche.

### CONCLUSIONS

We used the LIGIN software for docking the CASP2 ligand-receptor pairs. In cases where the experiment did not provide an approximate location for the binding pocket, we made no assumption about the pockets location and screened the entire protein molecule to determine its position. In this, our strategy differed fundamentally from that of other groups who preassigned the locations of the binding site. We submitted seven predictions. In six of seven cases our program correctly found the binding pocket location; in one case it failed (subsequent to CASP2, we have added an automatic cavity determination module to the WHAT IF package for this purpose). In principle, any software that can

### TABLE II. Comparison of Contacts in the Experimental and Predicted Structure of the Trypsin-Pentamidine Complex

<table>
<thead>
<tr>
<th>Protein residue</th>
<th>Experimental structure</th>
<th>Predicted structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Surf</td>
<td>HB</td>
</tr>
<tr>
<td>Ala56</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>His57</td>
<td>27</td>
<td>+</td>
</tr>
<tr>
<td>Tyr59</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Tyr94</td>
<td>-</td>
<td>26</td>
</tr>
<tr>
<td>Ser96</td>
<td>3</td>
<td>28</td>
</tr>
<tr>
<td>Asn97</td>
<td>24</td>
<td>33</td>
</tr>
<tr>
<td>Thr98</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td>Leu99</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>Gln175</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>Asp189</td>
<td>35</td>
<td>11</td>
</tr>
<tr>
<td>Ser190</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>Cys191</td>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>Gln192</td>
<td>14</td>
<td>+</td>
</tr>
<tr>
<td>Ser195</td>
<td>28</td>
<td>+</td>
</tr>
<tr>
<td>Val213</td>
<td>13</td>
<td>+</td>
</tr>
<tr>
<td>Ser214</td>
<td>19</td>
<td>+</td>
</tr>
<tr>
<td>Trp215</td>
<td>101</td>
<td>+</td>
</tr>
<tr>
<td>Gly216</td>
<td>44</td>
<td>+</td>
</tr>
<tr>
<td>Gly219</td>
<td>18</td>
<td>N24</td>
</tr>
<tr>
<td>Cys220</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Gly226</td>
<td>30</td>
<td>11</td>
</tr>
<tr>
<td>Tyr228</td>
<td>7</td>
<td>N25</td>
</tr>
</tbody>
</table>

*In this table, Surf = contact surface area (Å²) between atoms of the ligand and the residue; HB = ligand atoms hydrogen bonded to the corresponding residue; A–P = aromatic-polar contacts; h–h = hydrophobic-hydrophobic contacts. Protein residues in bold are in contact with the buried part of the ligand in the experimental structure.

---

![Fig. 4](T0034 ligand: amiloride.)

![Fig. 5](T0035 ligand: aica-riboside phosphate.)

---

ac.il/casp2/t0040.html and http://sgedg.weizmann.ac.il/casp2/t0041.html) revealed that in the protein embedded parts of the ligands the predicted structures occupy the same positions as in the experimental ones, but the surface located parts take different orientations. The resultant overall RMSD for our predictions were 8.0 Å and 7.5 Å, respectively, while for the buried parts, they were a respectable 1.4 Å and 2.1 Å. Docking of the correct ligand conformations gave NC50.63 for the INH and 0.53 for INI, versus NC50.63 and 0.58, respectively, for the predicted structures. Thus, even if we had considered all possible conformations for the ligands we would have had difficulties in predicting surface located parts, where the protein residues form neither pocket nor niche.

### CONCLUSIONS

We used the LIGIN software for docking the CASP2 ligand-receptor pairs. In cases where the experiment did not provide an approximate location for the binding pocket, we made no assumption about the pockets location and screened the entire protein molecule to determine its position. In this, our strategy differed fundamentally from that of other groups who preassigned the locations of the binding site. We submitted seven predictions. In six of seven cases our program correctly found the binding pocket location; in one case it failed (subsequent to CASP2, we have added an automatic cavity determination module to the WHAT IF package for this purpose). In principle, any software that can
identify protein binding pockets (such as g) can be used in conjunction with LIGIN.

For the two cases with buried ligands, orientation was also correctly predicted. We were unsuccessful in predicting the orientation of the covalently bound ligand (one case), maybe due to omitting intramolecular forces at the ligand–protein linkage from consideration. For the three cases where part of the ligand was located on the surface of the protein and part was buried, our program had difficulty in predicting the orientations of the more flexible surface-exposed sections but gave correct predictions for the buried parts. Treatment of flexibility in the LIGIN software has yet to be optimized.

Our approach allows detailed analysis of protein–ligand contacts. In five of our six correctly found pockets, the lists of residues forming the binding pockets and those for stabilizing contacts, were very similar for the experimental and predicted structures. Indeed, our main goal in the CASP2 exercise was to test our approach in predicting the major favorable interactions that stabilize ligand–receptor complexes. Such predictions are important in protein engineering and drug design, and should form part of the criteria used to determine the efficacy of a docking procedure.

Details on our CASP2 predictions can be found in the www page http://sgedg.weizmann.ac.il/casp2/

ACKNOWLEDGMENTS
V.S. and M.E. acknowledge the support of the Forschheimer and Wilstatter Centers at the Weizmann Institute of Science. G.V. acknowledges support from the BMBF (German Science Ministry) for the RELIWE project.

REFERENCES