An Iterative Knowledge-Based Scoring Function to Predict Protein–Ligand Interactions: II. Validation of the Scoring Function

SHENG-YOU HUANG, XIAOQIN ZOU
Department of Biochemistry, Dalton Cardiovascular Research Center, University of Missouri,
Columbia, Missouri 65211

Received 16 March 2006; Accepted 23 May 2006
DOI 10.1002/jcc.20505
Published online 18 September 2006 in Wiley InterScience (www.interscience.wiley.com).

Abstract: We have developed an iterative knowledge-based scoring function (ITScore) to describe protein–ligand interactions. Here, we assess ITScore through extensive tests on native structure identification, binding affinity prediction, and virtual database screening. Specifically, ITScore was first applied to a test set of 100 protein–ligand complexes constructed by Wang et al. (J Med Chem 2003, 46, 2287), and compared with 14 other scoring functions. The results show that ITScore yielded a high success rate of 82% on identifying native-like binding modes under the criterion of rmsd ≤ 2 Å for each top-ranked ligand conformation. The success rate increased to 98% if the top five conformations were considered for each ligand. In the case of binding affinity prediction, ITScore also obtained a good correlation for this test set (R = 0.65). Next, ITScore was used to predict binding affinities of a second diverse test set of 77 protein–ligand complexes prepared by Muegge and Martin (J Med Chem 1999, 42, 791), and compared with four other widely used knowledge-based scoring functions. ITScore yielded a high correlation of $R^2 = 0.65$ (or $R = 0.81$) in the affinity prediction. Finally, enrichment tests were performed with ITScore against four target proteins using the compound databases constructed by Jacobsson et al. (J Med Chem 2003, 46, 5781). The results were compared with those of eight other scoring functions. ITScore yielded high enrichments in all four database screening tests. ITScore can be easily combined with the existing docking programs for the use of structure-based drug design.

Key words: scoring function; protein–ligand interactions; knowledge-based; molecular docking; virtual screening

Introduction

In the accompanying study, we have described in detail the derivation of our iterative knowledge-based scoring function (ITScore), based on the structures of the protein–ligand complexes in the protein data bank. The affinity prediction test with ITScore on 140 diverse protein–ligand complexes yielded a high correlation of $R = 0.74$. Here, we present further validation of ITScore through extensive evaluations.

There are three commonly used criteria for evaluation of a scoring function. First, can the scoring function reproduce the experimentally determined binding modes? Second, can the scoring function predict the known binding affinities well? Third, can the scoring function identify the known inhibitors from a large database of compounds through virtual screening? It is a great challenge for a scoring function to satisfy all the three criteria for diverse ligand–protein test sets. In this work, we assess ITScore with these three criteria.

Materials and Methods

Calculation of Binding Energy Scores

As elaborated in the accompanying paper, the knowledge-based scoring function, referred to as ITScore, is based on the pair potentials derived using a novel iterative method. The iterative method circumvents the notorious reference state problem by improving the pair potentials, until they correctly discriminate native binding modes from decoy ligand structures. Only intermolecular interactions are considered in ITScore. The score is obtained by summing

Correspondence to: X. Zou; e-mail: zoux@missouri.edu
Contract/grant sponsor: NIH; contract/grant number: DK61529
Contract/grant sponsor: AHA; contract/grant number: 0265293Z

© 2006 Wiley Periodicals, Inc.
Validation of Iterative Knowledge-Based Scoring Function

Identifying Native Binding Modes

The most straightforward method for evaluating a scoring function is to test its ability of reproducing near-native binding modes. As described in Materials and Methods, the test set used for native-like binding mode prediction is a diverse set of 100 ligand–protein complexes constructed by Wang et al.3 This test set was extensively used to evaluate 13 scoring functions, including six empirical, four knowledge-based, and three force field-based scoring functions.4–7,22–33 There is no overlap between this set and our training set.

In the present work, we followed the definition that the binding mode of a complex is successfully identified if the rmsd between the best-scored conformation and the experimentally determined structure is below 2.0 Å.3–5 This is the default criterion unless specified again. Besides running ITScore on this test set, we also tested the force field scoring function provided in the DOCK 4.0 software (referred to as DOCK/FF),21,22 even though the SYBYL version of this force field-based scoring function (referred to as SYBYL/D-Score) had been tested before.3 The reason is because different software versions of a scoring function may yield different results.

Table 1 lists the success rates of ITScore and the force field scoring function of DOCK 4.0. The reported success rates of the aforementioned 13 scoring functions are also listed in the table.3–5 It can be seen that ITScore yields a high success rate (82%) in reproducing native binding modes, which is only slightly lower than that of DrugScoreCD. It is interesting to notice that SYBYL/D-Score, which is drawn from the molecular docking program DOCK by SYBYL, gives the lowest success rate of only 26%, whereas the same force field scoring function in the original DOCK 4.0 version performs much better and gives a success rate of 58%. This phenomenon illustrates the variations in performances with the same scoring function but different software packages. Table 1 also shows that the success rate of ITScore is always high, no matter which rmsd
Table 1. Success Rates of ITScore and 14 Other Scoring Functions on Wang et al.’s Test Set of 100 Complexes with Different rmsd Criteria.a

<table>
<thead>
<tr>
<th>Scoring functionb</th>
<th>Function type</th>
<th>References</th>
<th>( \text{rmsd} \leq 1.0 \text{ Å} )</th>
<th>( \text{rmsd} \leq 1.5 \text{ Å} )</th>
<th>( \text{rmsd} \leq 2.0 \text{ Å} )</th>
<th>( \text{rmsd} \leq 2.5 \text{ Å} )</th>
<th>( \text{rmsd} \leq 3.0 \text{ Å} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DrugScoreCSD</td>
<td>Knowledge-based</td>
<td>5</td>
<td>83</td>
<td>85</td>
<td>87</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ITScore</td>
<td>Iterative score</td>
<td>72</td>
<td>79</td>
<td>82</td>
<td>85</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>Cerius2/PLP</td>
<td>Empirical</td>
<td>25, 26</td>
<td>63</td>
<td>69</td>
<td>76</td>
<td>79</td>
<td>80</td>
</tr>
<tr>
<td>SYBYL/F-Score</td>
<td>Empirical</td>
<td>27</td>
<td>56</td>
<td>66</td>
<td>74</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td>Cerius2/LigScore</td>
<td>Empirical</td>
<td></td>
<td>64</td>
<td>68</td>
<td>74</td>
<td>75</td>
<td>76</td>
</tr>
<tr>
<td>DrugScorePDB</td>
<td>Knowledge-based</td>
<td>30</td>
<td>63</td>
<td>68</td>
<td>72</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>Cerius2/LUDI</td>
<td>Empirical</td>
<td>23, 24</td>
<td>43</td>
<td>55</td>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>X-Score</td>
<td>Empirical</td>
<td>31</td>
<td>37</td>
<td>54</td>
<td>66</td>
<td>72</td>
<td>74</td>
</tr>
<tr>
<td>AutoDock</td>
<td>Force-field-based</td>
<td>6</td>
<td>34</td>
<td>52</td>
<td>62</td>
<td>68</td>
<td>72</td>
</tr>
<tr>
<td>DFIRE</td>
<td>Knowledge-based</td>
<td>4</td>
<td>37</td>
<td>52</td>
<td>58</td>
<td>61</td>
<td>64</td>
</tr>
<tr>
<td>DOCK/FF</td>
<td>Force-field-based</td>
<td>22</td>
<td>37</td>
<td>47</td>
<td>58</td>
<td>66</td>
<td>69</td>
</tr>
<tr>
<td>Cerius2/PMF</td>
<td>Knowledge-based</td>
<td>7</td>
<td>40</td>
<td>46</td>
<td>52</td>
<td>54</td>
<td>57</td>
</tr>
<tr>
<td>SYBYL/G-Score</td>
<td>Force-field-based</td>
<td>29</td>
<td>24</td>
<td>32</td>
<td>42</td>
<td>49</td>
<td>56</td>
</tr>
<tr>
<td>SYBYL/ChemScore</td>
<td>Empirical</td>
<td>28</td>
<td>12</td>
<td>26</td>
<td>35</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>SYBYL/D-Score</td>
<td>Force-field-based</td>
<td>22</td>
<td>8</td>
<td>16</td>
<td>26</td>
<td>30</td>
<td>41</td>
</tr>
</tbody>
</table>

a ITScore stands for our current scoring function. The references for 14 other scoring functions are listed in the reference column, except for Cerius2/LigScore, which is provided in the Cerius2 user manual.33
b The scoring functions are ranked by their success rates at \( \text{rmsd} \leq 2.0 \text{ Å} \). The results of the scoring functions other than ITScore and DOCK/FF were obtained from the corresponding literatures.3–5 There are no available data for DrugScoreCSD at the criteria of \( \text{rmsd} \leq 2.5 \text{ Å} \) and \( \text{rmsd} \leq 3.0 \text{ Å} \).

To further evaluate ITScore, we also examined its performance when considering several top conformations in addition to the best-scored one. This is useful because usually multiple top conformations are saved for each ligand for further evaluation in hierarchical database screenings.34, 35 We defined that a prediction is successful if any of the considered conformations satisfies the specified \( \text{rmsd} \) criterion. Table 2 summarizes the results. The table shows that the success rate greatly improves when the number of considered conformations increases. If using \( \text{rmsd} \leq 2 \text{ Å} \) as the criterion, the success rate increases to 98% when the top five conformations were considered. That is, although a few native structures may not be identified as the best-scored structure, they are still ranked among the top few conformations. Namely, the best scored structure is not far from the native structure in terms of binding scores.3

Table 2. Success rates of ITScore on Wang et al.’s Test Set with Different \( \text{rmsd} \) Criteria When the Top Ranked Conformations were Considered.

<table>
<thead>
<tr>
<th>% of ranked conformations</th>
<th>( \text{rmsd} )</th>
<th>( \text{rmsd} \leq 1.0 \text{ Å} )</th>
<th>( \text{rmsd} \leq 1.5 \text{ Å} )</th>
<th>( \text{rmsd} \leq 2.0 \text{ Å} )</th>
<th>( \text{rmsd} \leq 2.5 \text{ Å} )</th>
<th>( \text{rmsd} \leq 3.0 \text{ Å} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
<td>79</td>
<td>82</td>
<td>85</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>82</td>
<td>86</td>
<td>91</td>
<td>90</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>85</td>
<td>91</td>
<td>94</td>
<td>95</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>92</td>
<td>95</td>
<td>96</td>
<td>98</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>95</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>

There are a total of 101 conformations for each complex (i.e., native structure + 100 putative conformations).

Predicting Binding Affinities

A second important criterion to evaluate scoring functions is to predict real binding affinities of ligand–protein complexes. A good scoring function should obtain satisfactory correlations between calculated energy scores and experimentally measured binding affinities. In the present work, the performance of binding affinity prediction was measured in terms of the widely used Pearson’s correlation coefficient:38

\[
R = \frac{\sum_{k=1}^{N}(x_k - \langle x \rangle)(y_k - \langle y \rangle)}{\sqrt{\sum_{k=1}^{N}(x_k - \langle x \rangle)^2}\left[\sum_{k=1}^{N}(y_k - \langle y \rangle)^2\right]}
\]
with Wang et al.’s test set using ITScore and the force field scoring function of DOCK 4.0. The reported correlations for 13 other scoring functions are also quoted in the table.3–5 All the calculated energy scores were based on the experimentally determined complex structures. It was reported that there exists a moderate correlation between experimental binding affinities and molecular weight ($R \sim 0.56$ for this test set).5–5 This inherent correlation needs to be considered in evaluating the performance of a scoring function.

Table 3 shows that ITScore gives the highest correlation coefficient of 0.65. It is also impressive that besides ITScore, several other scoring functions such as DFIRE, DrugScore, Cerius2/PLP, and SYBYL/G-Score also obtain good correlation coefficients of over 0.56, considering that all these scoring functions were developed based on “structure” information rather than “binding affinity” data. It is also noticed from Tables 1 and 3 that ITScore and several other scoring functions such as X-Score, DrugScore, and Cerius2/PLP are effective on both binding mode prediction and affinity prediction with this test set, which is important in the development of a scoring function.

Muegge and Martin’s Test Set

We also tested ITScore on affinity prediction using a diverse set of 77 ligand–protein complexes prepared by Muegge and Martin.7 The complexes were classified into five subsets. The energy scores were calculated using the crystal structures of the complexes extracted from the protein data bank. The results are shown in Table 4 and Figure 1. The reported correlation results for four other well-known knowledge-based scoring functions (PMP99,7 DrugScorePDB,39 BLEEP,40 and SMoG200141) are also shown in the table and figure.

It can be seen from Table 4 and Figure 1 that ITScore obtains good correlations for different sets (ranging from $R^2 = 0.35$ to $R^2 = 0.87$), suggesting its robustness in affinity prediction. It is also interesting to notice that all the listed knowledge-based scoring functions perform less satisfactory on subset 4, the endothiapepsin complexes. We found that ligands in this subset are relatively large and contain many rotatable bonds, which might account for the less well correlation for subset 4 and need to be considered for future improvement of knowledge-based scoring functions. For the whole

### Table 3. Correlations Between Experimentally Determined Binding Affinities and Calculated Binding Scores from ITScore and 14 Other Scoring Functions on Wang et al.’s Test Set.

<table>
<thead>
<tr>
<th>Scoring function</th>
<th>Function type</th>
<th>Correlation coefficient ($R$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITScore</td>
<td>Iterative</td>
<td>0.65</td>
</tr>
<tr>
<td>X-score</td>
<td>Empirical</td>
<td>0.64</td>
</tr>
<tr>
<td>DFIRE</td>
<td>Knowledge-based</td>
<td>0.63</td>
</tr>
<tr>
<td>DrugScoreCSD</td>
<td>Knowledge-based</td>
<td>0.62</td>
</tr>
<tr>
<td>DrugScorePDB</td>
<td>Knowledge-based</td>
<td>0.60</td>
</tr>
<tr>
<td>Cerius2/PLP</td>
<td>Empirical</td>
<td>0.56</td>
</tr>
<tr>
<td>SYBYL/G-Score</td>
<td>Force-field-based</td>
<td>0.56</td>
</tr>
<tr>
<td>SYBYL/D-Score</td>
<td>Force-field-based</td>
<td>0.48</td>
</tr>
<tr>
<td>SYBYL/ChemScore</td>
<td>Empirical</td>
<td>0.47</td>
</tr>
<tr>
<td>Cerius2/PMF</td>
<td>Knowledge-based</td>
<td>0.40</td>
</tr>
<tr>
<td>DOCK/FF</td>
<td>Force field</td>
<td>0.40</td>
</tr>
<tr>
<td>Cerius2/LUDI</td>
<td>Empirical</td>
<td>0.36</td>
</tr>
<tr>
<td>Cerius2/LigScore</td>
<td>Force-field-based</td>
<td>0.35</td>
</tr>
<tr>
<td>SYBYL/F-Score</td>
<td>Empirical</td>
<td>0.30</td>
</tr>
<tr>
<td>AutoDock</td>
<td>Force-field-based</td>
<td>0.05</td>
</tr>
</tbody>
</table>

*The scoring functions are ranked by Pearson’s correlation coefficients calculated using the experimentally determined structures. The results of the scoring functions other than ITScore and DOCK/FF were obtained from the corresponding literatures.3–5

where $N$ is the number of ligand–protein complexes in the test set, $k$ represents the $k$th complex, $x$ and $y$ stand for the measured binding affinities and calculated energy scores, and $\langle \rangle$ denotes an arithmetic average.

In this section, we first tested the performance of ITScore on Wang et al.’s test set.3 We then gave a detailed comparison between ITScore and several other knowledge-based scoring functions on the test set prepared by Muegge and Martin.7

**Wang et al.’s Test Set**

Table 3 lists the correlation coefficients between the calculated energy scores and the experimental binding affinities we obtained with Wang et al.’s test set using ITScore and the force field scoring function of DOCK 4.0. The reported correlations for 13 other scoring functions are also quoted in the table.3–5 All the calculated energy scores were based on the experimentally determined complex structures. It was reported that there exists a moderate correlation between experimental binding affinities and molecular weight ($R \sim 0.56$ for this test set).5–5 This inherent correlation needs to be considered in evaluating the performance of a scoring function.

Table 3 shows that ITScore gives the highest correlation coefficient of 0.65. It is also impressive that besides ITScore, several other scoring functions such as DFIRE, DrugScore, Cerius2/PLP, and SYBYL/G-Score also obtain good correlation coefficients of over 0.56, considering that all these scoring functions were developed based on “structure” information rather than “binding affinity” data. It is also noticed from Tables 1 and 3 that ITScore and several other scoring functions such as X-Score, DrugScore, and Cerius2/PLP are effective on both binding mode prediction and affinity prediction with this test set, which is important in the development of a scoring function.

**Muegge and Martin’s Test Set**

We also tested ITScore on affinity prediction using a diverse set of 77 ligand–protein complexes prepared by Muegge and Martin.7 The complexes were classified into five subsets. The energy scores were calculated using the crystal structures of the complexes extracted from the protein data bank. The results are shown in Table 4 and Figure 1. The reported correlation results for four other well-known knowledge-based scoring functions (PMP99,7 DrugScorePDB,39 BLEEP,40 and SMoG200141) are also shown in the table and figure.

It can be seen from Table 4 and Figure 1 that ITScore obtains good correlations for different sets (ranging from $R^2 = 0.35$ to $R^2 = 0.87$), suggesting its robustness in affinity prediction. It is also interesting to notice that all the listed knowledge-based scoring functions perform less satisfactory on subset 4, the endothiapepsin complexes. We found that ligands in this subset are relatively large and contain many rotatable bonds, which might account for the less well correlation for subset 4 and need to be considered for future improvement of knowledge-based scoring functions. For the whole

### Table 4. Comparison of ITScore and Four Published Knowledge-Based Scoring Functions in Terms of Binding Affinity Prediction on Muegge and Martin’s Test Set.a

<table>
<thead>
<tr>
<th>No.</th>
<th>Set</th>
<th>No. of complexes</th>
<th>ITScore</th>
<th>PMF99</th>
<th>DrugScorePDB</th>
<th>BLEEP</th>
<th>SMoG2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Serine protease</td>
<td>16</td>
<td>0.87</td>
<td>0.87</td>
<td>0.86</td>
<td>0.79</td>
<td>0.81</td>
</tr>
<tr>
<td>2</td>
<td>Metalloprotease</td>
<td>15</td>
<td>0.70</td>
<td>0.58</td>
<td>0.70</td>
<td>0.59</td>
<td>0.64</td>
</tr>
<tr>
<td>3</td>
<td>l-Arabinose binding prot.</td>
<td>18(9)</td>
<td>0.49</td>
<td>0.48</td>
<td>0.22</td>
<td>0.14</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>Endothiapepsin</td>
<td>11</td>
<td>0.35</td>
<td>0.22</td>
<td>0.30</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>5</td>
<td>Others</td>
<td>17</td>
<td>0.70</td>
<td>0.69</td>
<td>0.43</td>
<td>0.49</td>
<td>0.50</td>
</tr>
<tr>
<td>6</td>
<td>Sets 1–5</td>
<td>77</td>
<td>0.65</td>
<td>0.61</td>
<td>n/a</td>
<td>0.28</td>
<td>0.46</td>
</tr>
</tbody>
</table>

*aThe results for PMF99, DrugScorePDB, BLEEP, and SMoG2001 were extracted from their original papers.7,39–41

bNote that the correlation parameter in this table is the square of correlation coefficient ($R^2$) rather than the correlation coefficient itself ($R$), for consistency with original data.

cThe crystal structures of the nine l-arabinose complexes containing two ligand conformations were treated separately.
test set (last row in Table 4), ITScore yields the highest correlation ($R^2 = 0.65$), closely followed by PMF99 (0.61).

Enrichment Tests for Virtual Database Screening

A third important criterion for evaluating a scoring function is its ability to identify known inhibitors out of a large database of compounds. This discrimination is crucial for the commonly used strategy in virtual screening of chemical databases. A scoring function that yields a good correlation in binding affinity prediction does not necessarily perform well in database ranking, because these are two completely different tasks. The enrichment test is a good criterion to quantify the performance of a scoring function in virtual screening of chemical databases. The enrichment is defined as the accumulated rate of active inhibitors found in a certain top percentage of the ranked database. As described in Materials and Methods, we ran virtual screening against four databases prepared by Jacobsson et al. The four target proteins are ERα, MMP3, fXa, and AChE. The inactive compound database spans the range of physical properties of the active compounds as aforementioned. Both ITScore and the force field scoring function of DOCK 4.0 were used. The ligands were treated as flexible molecules by using OMEGA to generate multiple conformers. The results are plotted in Figure 2.

We also obtained virtual screening data of other seven scoring functions from Dr. Jacobsson. Those scoring functions are ICM Score, ICM-PMF Score, SYBYL/F-Score, SYBYL/D-Score, SYBYL/PMF-Score, SYBYL/G-Score, and SYBYL/ChemScore, respectively. As aforementioned, SYBYL/D-Score is a different version of the DOCK/FF scoring function. SYBYL/PMF-Score is the SYBYL version of the knowledge-based scoring function by Muegge and Martin, and ICM-PMF Score is the MolSofts implementation of a potential of mean force scoring function in ICM, derived using the similar method to that by Muegge and Martin. The enrichment data are also plotted in Figure 2 as an indirect validation of ITScore. It is difficult to assess different scoring functions from Figure 2, because different ligand preparation and scoring/docking combination may affect the performance of the tested scoring functions.

It is usually thought that the top 5–10% of the database is of interest in virtual screening. Thus, a histogram of the enrichment factors at the top 5 and 10% levels of the ranked database for each

![Figure 2](image-url)
the ranked database (Figs. 2 and 3). DOCK/FF also performs very well, and identifies 37% of the active compounds among the top 5% of the screened compounds, respectively (Figs. 2, Tables 1 and 3). Finally, because ITScore and DOCK/FF were used with the same protocol and docking program, comparison of the performances between ITScore and DOCK/FF is direct and shows that ITScore outperforms DOCK/FF in the present study.

**Conclusion**

We have extensively evaluated our scoring function ITScore through three important criteria on diverse test sets. The scoring function was first tested on a set of 100 protein–ligand complexes, with known crystal structures and binding affinities, prepared by Wang et al.3 Compared with 14 other scoring functions, ITScore obtains the second highest success rate (82%) with the criterion of rmsd <2.0 Å in identifying correct binding modes, and yields the highest correlation coefficient in affinity prediction (R = 0.65). With the test set of 77 complexes from Muegge and Martin’s paper,7 the performance of ITScore was compared with that of other four knowledge-based scoring functions (PMF99, DrugScorePDB, BLEEP, and SMoG2001) in terms of binding affinity prediction. Overall, ITScore (R² = 0.65) matches PMF99 (R² = 0.61), and performs better than the other three knowledge-based scoring functions. Enrichment tests of virtual database screening against four target proteins show that ITScore is the only scoring function that gives high enrichments for every target protein, in comparison with the other eight scoring functions, indicating a success of ITScore in identifying known inhibitors out of a database.

ITScore is as fast as empirical functions, yet, unlike empirical functions, it requires no prior knowledge on measured binding affinities to derive the interaction potentials. This feature together with the large, diverse training set of ligand–protein complex structures account for the robustness of ITScore in binding affinity prediction. The new iterative method for derivation of pair potentials circumvents the long-time reference state problem, and not only helps improve affinity prediction but also significantly improves binding mode prediction and virtual screening enrichment. ITScore uses SYBYL’s atom type definitions and is therefore readily applicable to molecular files prepared by SYBYL or converted by programs such as BABEL.19 ITScore can be applied to ligand binding studies with application to structure-based drug design.

Future studies to improve the performance of ITScore are discussed later. First, it has been reported recently that the better
resolved small molecule structures in the Cambridge structural database (CSD) produce pair potentials of superior statistical significance and more detailed shape than do the ligand–protein complex structures in the protein data bank.\textsuperscript{5} We will apply the idea of our iterative method to CSD\textsuperscript{44} to see if we can derive improved pair potentials for ITScore. Second, the current ITScore does not consider the extended extent of atoms and thus may not account for important solvation energies. ITScore may also not be effective to describe some complicated interactions such as hydrogen bonding and hydrophobic effects. All these issues await further studies. Third, it is surprising that the inclusion of metal ions does not improve the performance of ITScore much. The same observation was also reported with other knowledge-based scoring functions.\textsuperscript{4, 45} Therefore, further studies need to be done to well-characterize the metal ions in knowledge-based scoring functions. Finally, it will be interesting to test combinations of ITScore and other scoring functions using the consensus scoring strategy to further improve the scoring efficiency.\textsuperscript{46, 47} It is also necessary to keep in mind that such efficiency tests are not simply a scoring test, but rather a test of the scoring/docking combination.

Acknowledgments

We thank Prof. Irwin D. Kuntz and Dr. Hao-Yang Liu for helpful discussions, and Prof. Kevin Gillis for valuable comments on the manuscript. We also thank Dr. Michael Jacobsson for providing us with their virtual screening data. Support to XZ from OpenEye Scientific Software Inc. (Santa Fe, NM), Tripos, Inc. (St. Louis, MO), and MDL Information Systems, Inc. (San Leandro, CA) is gratefully acknowledged.

References

32. SYBYL, version 6.9; Tripos Inc.; www.tripos.com.