Single-Body Residue-Level Knowledge-Based Energy Score Combined With Sequence-Profile and Secondary Structure Information for Fold Recognition

Hongyi Zhou and Yaoqi Zhou

Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology & Biophysics, State University of New York at Buffalo, New York

ABSTRACT

An elaborate knowledge-based energy function is designed for fold recognition. It is a residue-level single-body potential so that highly efficient dynamic programming method can be used for alignment optimization. It contains a backbone torsion term, a buried surface term, and a contact-energy term. The energy score combined with sequence profile and secondary structure information leads to an algorithm called SPARKS (Sequence, secondary structure Profiles and Residue-level Knowledge-based energy Score) for fold recognition. Compared with the popular PSI-BLAST, SPARKS is 21% more accurate in sequence-sequence alignment in ProSup benchmark and 10%, 25%, and 20% more sensitive in detecting the family, superfamily, fold similarities in the Lindahl benchmark, respectively. Moreover, it is one of the best methods for sensitivity (the number of correctly recognized proteins), alignment accuracy (based on the MaxSub score), and specificity (the average number of correctly recognized proteins whose scores are higher than the first false positives) in LiveBench 7 among more than twenty servers of non-consensus methods. The simple algorithm used in SPARKS has the potential for further improvement. This highly efficient method can be used for fold recognition on genomic scales. A web server is established for academic users on http://theory.med.buffalo.edu. Proteins 2004;55:1005–1013.

© 2004 Wiley-Liss, Inc.

Key words: knowledge-based potential; structure selections; fold recognition; threading; protein structure prediction

INTRODUCTION

Proteins adopt a limited number of unique structural folds. As a result, unrelated (or remotely related) sequences may encode similar structure. The aim of fold recognition is to recognize the structure similarities of proteins. Although significant advances have been made in the last 20 years, fold recognition continues to be one of the bottlenecks for accurate structure prediction in the CASP (Critical Assessment of Techniques for Protein Structure Prediction) experiments.1

There are two main approaches for fold recognition. One approach is sequence-structure threading2 in which the compatibility of a sequence with each known structure is assessed by a score function (or structural profile).3–8 (For recent reviews, see References 9–13.) The other approach involves direct sequence-sequence comparison on the pairwise9–14 or multiple sequence15–18 levels. (For recent reviews, see References 12, 26–28.) Because the two approaches use different input information, a combined approach19–33 is superior than either approach alone. However, this combined approach often increases detection sensitivity but does not improve significantly the accuracy of alignment. Panchenko et al.34 showed that improving both sensitivity and accuracy of alignment requires an optimal combination of structural and sequence profile terms.

In this paper, we develop a single-body knowledge-based energy score for fold recognition. A single-body energy score (sometimes called profile-energy score35–37) allows the use of efficient dynamic programming for optimal alignment. This is in contrast to the pairwise interaction potential which would require extensive computational time38–40 or a frozen environment approximation41 for optimal alignment. Previous single-body scores are limited to one or more of the properties such as hydrophobicity, solvent exposure, secondary structure, and contact energies. For example, Shan et al.36 (COBLATH) and Kelley et al.37 (3D-PSSM) used a score function that only takes the secondary structure and solvent exposure into account whereas Panchenko et al.38 employed a contact potential. (For a review, see David et al.39) In this paper, we establish a more elaborate knowledge-based score that contains a torsion-angle term for backbone interaction and a combined buried surface and contact-energy term for residue-residue and residue-solvent interactions. We found that a torsion-angle term is important for correctly identifying
the native structure from decoys and this term together with a buried surface and contact energy terms provides one of the best residue-level potentials for structure selections from decoys.

The proposed knowledge-based, structure-derived score function is combined with the sequence profiles generated from PSI-BLAST\textsuperscript{21} and the secondary structure information predicted from PSIPRED\textsuperscript{44} for fold recognition. The highly efficient method called SPARKS (Sequence, secondary-structure Profiles And Residue-level Knowledge-based Score) is found to improve over most existing, nonconsensus, fold-recognition methods in sensitivity, alignment accuracy, and specificity based on several well-established benchmarks.

**METHODS**

**The Knowledge-Based Score**

The knowledge-based, structure-derived score function for residue type I at the position i is a sum of three terms (a backbone torsion term $u_{\text{torsion}}(I, \phi_i, \psi_i)$, a buried surface term $u_{\text{surface}}(I, dASA_i)$, and a contact energy term $u_{\text{contact}}(I, N_{i\text{contact}})$):

$$u_{i\text{structure}}(I) = u_{\text{torsion}}(I, \phi_i, \psi_i) + u_{\text{surface}}(I, dASA_i) + u_{\text{contact}}(I, N_{i\text{contact}}), \quad (1)$$

where $u_{\text{surface}}$ and $u_{\text{contact}}$ are weight factors for buried surface and contact energy terms relative to the torsion term, respectively, $dASA_i$ and $N_{i\text{contact}}$ are the buried accessible surface area (ASA) and the number of contacts for the $C_{\alpha}$ atom of the residue type I at position i, respectively, and $\phi_i$ [$= \phi(C_{i-1}, C_{i}, C_{i+1})$] and $\psi_i$ [$= \psi(C_{i}, C_{i-1}, C_{i+1})$] are the backbone torsion angles at each side of a residue at the sequence position i. [For the first and last residues, $u_{\text{torsion}}(I, \phi_i, \psi_i)$ is set to zero because torsion angles are not defined.]

The torsion potential $u_{\text{torsion}}(I, \phi_i, \psi_i)$ is given by

$$u_{\text{torsion}}(I, \phi_i, \psi_i) = -RT \ln \frac{N_{\text{obs}}(I, \phi_i, \psi_i)}{\sum_{\phi, \psi} N_{\text{obs}}(I, \phi, \psi)/N_{\text{bin}}} \quad (2)$$

where $N_{\text{obs}}(I, \phi, \psi)$ is the number of observed occurrence of the residue type I at torsion angles of $\phi$ and $\psi$. $N_{\text{obs}}$ is the number of contacts of residue type I as cited from the 1011-protein database\textsuperscript{45}, and $N_{\text{bin}}$ is the number of bins divided, in this case, into 36 bins. That is, $N_{\text{bin}} = 36$. The potential parameters contained in a table of $20 \times 36 \times 36$ are available on http://theory.med.buffalo.edu. The statistical distribution is obtained by using a structural database of 1011 non-homologous soluble proteins.\textsuperscript{45}

The buried accessible surface term of residue type I is given by

$$u_{\text{surface}}(I, dASA_i) = -B_i^i \times dASA_i, \quad (3)$$

where $B_i^i$ is the burial of residue type I and $dASA_i$ is the buried accessible surface area at the position i. This term is introduced because it was shown that the average contribution of a residue to the stability of a protein is linearly correlated with its buried accessible surface area with a slope called buriability.\textsuperscript{46} The value of buriability was derived from an all-atom distance-dependent potential of mean force using a distance-scaled infinite-ideal gas reference state (DFIRE)\textsuperscript{47} and the structural database of 1011 proteins.\textsuperscript{45} The value of buriability for each amino acid,\textsuperscript{46} listed in Table I, can quantitatively take into account the fact that hydrophobic residues are more likely to be buried into the core of proteins.\textsuperscript{46}

To increase the computational efficiency, we take advantage of the fact that the buried ASA ($dASA_i$) normalized by the average buried ASA ($\bar{dASA}$) correlates strongly with the number of contacts ($N_{i\text{contact}}$) normalized by the average number of contacts ($\bar{N}_{i\text{contact}}$). As shown in Figure 1, the correlation is not perfect but significant enough for the purpose of this paper. That is, Eq. (3) can be rewritten as

$$u_{\text{surface}}(I, dASA_i) = u_{\text{surface}}(I, N_{i\text{contact}}) = -B_i^i \times \frac{dASA_i}{N_{i\text{contact}}} \times \frac{N_{i\text{contact}}}{N_{i\text{contact}}}, \quad (4)$$

where $dASA_i$ and $N_{i\text{contact}}$ are average buried ASA obtained from Zhou and Zhou\textsuperscript{48} and average C-atom contact number of a residue type calculated from the 1011-protein database,\textsuperscript{45} respectively. The values of $dASA_i$ and $N_{i\text{contact}}$ are also listed in Table I.

The contact energy term is given by

$$u_{\text{contact}}(I, N_{i\text{contact}}) = -RT \ln \frac{N_{\text{obs}}(I, k)}{\sum_k N_{\text{obs}}(I, k)/N_{\text{obs}}}, \quad (5)$$

where $N_{\text{obs}}(I, k)$ is the number of observed contacts of residue I with other residues at $k$th bin and $N_{\text{bin}}\times$ the number of contact bins, is set to 25. In other words, the number of contacts allowed in the program is 25. In the rare occasions of more than 25 contacts, the statistics is included in the bin for 25 contacts. Here, we set $T = 300K$ ($RT = 0.6$ kcal/mol). When $N_{\text{obs}}(I, k) = 0$, the knowledge-
FOLD RECOGNITION BY SPARKS

1007

1

based score $u^\text{structure}(I)$ is set to a constant of 10 kcal/mol. A
contact is defined by the $C_a-C_a$ distance within 8.5 Å. The
20 × 25 matrix for $u^\text{contact}(I, N_{\text{contact}})$ is available in the
SPARKS web server.

It should be noted that buried surface and contact-energy terms are two different ways of calculating the potential of mean force for residue-residue and residue-solvent interactions (See, e.g. Melo et al.49). One is expressed in terms of buried surface area and is proportional to the number of contacts whereas the other is proportional to the logarithm of the number of contacts. Because it is not clear which expression is a more accurate representation for the potential of mean force, we include both terms by introducing adjustable weight factors [Eq. (1)]. These weight factors are to be determined by optimizing the performance for selecting structures from decoys (see below). The total energy score of a protein is given by

$$u^\text{structure} = \sum_{i=2}^{N_c-1} u^\text{structure}(I),$$

where $N_c$ is the total number of residues, $I$ is the residue type of residue at position $i$. Note that the first and last residues are excluded from the calculations. Using the decoy sets of 4state_reduced,50 fisa,51 fisa casp3,51 lmds, and lattice ssfit,52 we found that optimal values of $u^\text{contact}$ and $u^\text{surface}$ are 1.0 and 2.0, respectively. These parameters lead to an energy score that correctly identified 22 native proteins from a 32 protein decoy sets. This success rate was similar or better than some distance-dependent

all-atom pair potentials57 (RAPDF53 and atomic KBP54), which are more accurate than the corresponding residue-level potentials. Furthermore, the best distance-dependent potential (TE-13) among several residue-level potentials compared by Tobi and Elber55 identified 14 native structures out of 25 decoy sets. The success rate of the new score function [Eq. (6)] is 17 out of the same 25 proteins. Thus, the new single-body, distance-independent, potential is one of the best residue-level potentials for structure discrimination.

Sequence Profiles

The log-likelihood (position specific score matrix PSSM) profile $P^b_j(t)$ of a sequence $b$ of residue type $t$ at position $j$ was generated by PSI-BLAST21 after three iterations and an E value cutoff of 0.001. This is the same setting used by PSIPRED for secondary structure prediction.44 The sequence database was obtained from NCBI at ftp://

A frequency profile $F^\text{query}(i)$ of the query sequence at position $i$ was also constructed from the output of the multiple sequence alignment of PSI-BLAST for those sequences that have less than 98% identity and less than 0.001 E-value. The profile was normalized to one at each sequence position. The mutation score is expressed by profiles as follows:

$$u^\text{mutation}(i, j) = -\sum_{t=1}^{20} F^\text{query}(t) \cdot P^\text{template}(t),$$

where $F^\text{query}$ ($P^\text{template}$) is the frequency (log-likelihood) profile of query sequence (template sequence), and the summation is over the 20 residue types. This mutation score was also used by others for fold recognition (e.g., Fischer56).

Secondary Structures

The log-likelihood profile of query sequence, $P^\text{query}(i)$, was used for the secondary structure prediction by PSIPRED. The score from the predicted secondary structure information is given by

$$u^\text{2ndary}(i, j) = -\delta_{\text{query}(j), \text{template}(j)},$$

where $\text{query}(i)$ and $\text{template}(j)$ are the predicted secondary structure of the query sequence at position $i$ and the actual secondary structure of the template at position $j$, respectively and $\delta_{\text{query}(j), \text{template}(j)}$ is 1 if there is a match between predicted and actual secondary structures and -1, if otherwise. Similar score function for secondary structure was used by Fischer and Eisenberg.31 The secondary structures of templates were obtained by H-bonds (DSSP-like) criteria.57 Three states (helix, strand, coil) were used for secondary structure. A built-in simplified PSIPRED algorithm of secondary structure prediction is used to speed up computation. The original PSIPRED used two rounds of neural networks and provided a confidence level (C. L.) for the prediction at each position. Our implementation employed only one round of neural network without evaluation of C. L. This is because we found that a more
accurate prediction of the secondary structure by the second neural network did not further improve the accuracy of fold recognition.

The Fold Recognition Algorithm and Assessment

The sequence profile and secondary structure information is combined with the knowledge-based score for fold recognition. The total score is given by

$$u_{\text{total}} = u_{\text{rotation}}(i, j) + w_{\text{2ndary}} u_{i, j}^{\text{2ndary}} + u_{\text{structure}}(t_i),$$

(9)

where $u_{\text{structure}}(t_i)$ is the knowledge-based score of the template at position $j$ with the residue type $t_i$ of the query position $i$, and $w_{\text{2ndary}}$ and $w_{\text{structure}}$ are the weight factors for secondary structure and structure scores, respectively. As with Fischer and Eisenberg, we let $w_{\text{2ndary}}$ equal to the gap extension penalty $w_1$ to reduce the number of adjustable parameters.

A profile-profile alignment method using a global-local dynamic programming algorithm was employed to find the minimum of the total score that aligns the query sequence with a template in the template library. That is, the raw score of an alignment is $u_{\text{template}}$. The $Z$ score for a given template is defined as $(u_{\text{total}} - u_{\text{ave}}) / u_{\text{sd}}$, where $u_{\text{ave}}$ and $u_{\text{sd}}$ are the average and the standard deviation of the raw scores for all templates, respectively. The initial ranking is given by the raw score normalized by the alignment length ($S_{\text{template}}$) in order to remove sequence-length dependence. To determine the specificity of the fold recognition method, the significance score of a template is calculated by the multiplication of its normalized raw score ($S_{\text{template}}$) and the $Z$ score. A bonus score of $-1$ is also added to the significance score if the model has a structural similarity with one or more of the next nine models based on initial ranking with normalized raw scores. (A structure is similar if MaxSub score is greater than 0.01). This simple method for calculating significance of an alignment is based on the rational that a good model should not only have a very low normalized score ($S_{\text{template}}$) but also have a score that is lower relative to the scores based on other templates ($Z$ score). A similar method is used in Ref. 36.

All weight factors and gap parameters were obtained by optimizing the performance of the method in Fischer’s dataset (see Results). $w_0 = 4.5$, $w_1 = 0.5$, $w_{\text{2ndary}} = w_1$, and $w_{\text{structure}} = 0.175$ whereas $u_{\text{contact}} = 1.0$ and $u_{\text{surface}} = 2.0$ by optimizing native-structure selections from decoys (see above).

The accuracy of alignment was assessed by MaxSub score, which is a measure of similarity between 0.0 (no similarity) and 1.0 (perfect similarity) between the predicted (model) structure and the native structure. The value is calculated by searching the largest subset of well-superimposed residues ($\leq 3.5$ Å). The model structure was built by using the first-ranking template with MODeller 6 v1.66. We use MaxSub score because it is the official evaluation method used in the CAFASP/Critical assessment of fully automated protein structure predic-

RESULTS

Training Set: Fischer’s Dataset for Sensitivity and Alignment Accuracy

The initial gap penalty ($w_0$) and the extension gap penalty ($w_1$) of the SPARKS method were first optimized by using Fischer’s dataset without the knowledge-based, structure-derived score. The dataset contains 68 probe sequences and 301 library structures. A match occurs when the expected match was ranked as the number one or only below its superfamily members based on SCOP 1.61 classification (i.e., no incorrect folds have a better score than the expected match). The highest success rate is 56/68 when $w_0 = 4.5$ and $w_1 = 0.5$. The weight factor for the secondary structure $w_{\text{2ndary}}$ is set to be same as $w_1$ (see Methods).

The weight factor for the structure-derived knowledge-based score $w_{\text{structure}}$ was then obtained by including the single-body knowledge-based score and further optimizing the performance for Fischer’s dataset. We found that the number of correct matches is not very sensitive to the value of $w_{\text{structure}}$. Thus, we optimized $w_{\text{structure}}$ for alignment accuracy (MaxSub score) instead. The total MaxSub score for Fischer’s dataset was maximized to 26.12 at $w_{\text{structure}} = 0.175$. This score increases from 24.68 given by SPARKS without the structure-derived score. The number of correctly recognized proteins for SPARKS, however, remains 56. This success rate (56/68) is the same as that of COBLATH (a combined approach of sequence profile with a score matrix for secondary structure and solvent exposure) but is lower than that of a computationally more intensive, hierarchical threading method called PROSPECTOR (58-61/68), which involves pair interaction potentials and separate profiles for closely-related and distant-related sequences.

Test Set 1: ProSup Benchmark for Sequence-Accuracy

ProSup benchmark was prepared by Sippl’s group to test the alignment accuracy of fold recognition methods. The set consists of 127 pairs of proteins with correct alignments obtained by structural alignment program ProSup. The accuracy of an alignment was obtained by calculating the percent of matches between the correct alignment and the alignment made by a fold recognition method. Table II compares the performance of several methods. It should be emphasized, however, that the performance of our method is not directly comparable with the published results for some of the previous methods because available sequence and structure databases at the time is different. Nevertheless, the comparison can be served as an approximate indicator for the accuracy of SPARKS. For the direct pairwise sequence alignment, the optimized substitution matrix STROMA performs the best. It is remarkable that an optimized substitution matrix can perform even slightly better than multiple-sequence alignment method of PSI-BLAST.
method given by Sippl et al. makes a significant improvement over sequence-based methods. However, the sequence profile and secondary structure information in SPARKS alone is already more accurate than the threading method (51.4% versus 48.0%). Incorporation of the single-body knowledge-based score increases the accuracy by another 5.3% to 57.2%. The latter is more than 21% better than the accuracy provided by PSI-BLAST. If the criterion for the accuracy is relaxed to allow a shift of residues from the “correct” alignment, the overall accuracy of alignment is 78.5% for SPARKS (69.2% for the SPARKS without the knowledge-based score). Note that the improvement upon the use of the structure-derived score is 8.9% for a relaxed criterion for the accuracy. Clearly, introducing the knowledge-based score brings in the significant improvement in alignment accuracy.

Test Set 2: Lindahl Benchmark for Fold-Recognition Sensitivity

The Lindahl set was designed to assess the fold recognition sensitivity. It has 976 proteins. Each protein is aligned with the rest 975 proteins. There are 555, 434, and 321 pairs of proteins in the same family, superfamily, and fold, respectively. The fold-recognition method is tested by checking whether or not the method can recognize the member of same family, superfamily, or fold as the first rank or within the top five ranks. The results of SPARKS are compared with several well-established methods in Table III. We stress that the comparison only serves an approximate guide because the sequence and structure database available for previous methods are smaller than the one used in SPARKS. This sensitivity test, however, may not reflect the true alignment accuracy because the result is based on some-what subjective SCOP classification. To address this question more quantitatively, we calculated the MaxSub score between the model built from first-ranking template and the known native structure. It is found that the total number of recognized proteins (MaxSub > 0.01) increases from 616 to 630 (a 2.0% increment) whereas the total MaxSub score increases from 322.72 to 334.47 (a 3.6% increment). Thus, the knowledge-based score makes a small but significant improvement in both sensitivity and accuracy of fold recognition for the Lindahl benchmark set.

Test Set 3: LiveBench 7

Sixty proteins of LiveBench targets released between February 1, 2003 and April 10, 2003 are used to further test the performance of SPARKS (http://bioinfo.pl/LiveBench/). This partial LiveBench set is what was available prior to the preparation and final submission of this paper. The LiveBench targets were new entries in the Protein Data Bank which do not have sequence similarity to previously released proteins and thus provide an unbiased set for testing the performance of fold recognition methods. Currently, more than 30 servers are tested on LiveBench automatically.

The template library for SPARKS was built by using the 40% representative domains of SCOP 1.61. The entire
chains of multiple-domain proteins are contained in the library. The library was then updated with new proteins released before February 3, 2003 if they have less than 40% sequence identity with the sequences already in the library. (This was done by protein sequence culling server PISCES\(^\text{55}\)).

The results of SPARKS were compared with the then top three single-method performers of more than 20 servers of non-consensus methods in Table IV. The number of recognized proteins (sensitivity) increases from between 39 and 41 for the top three performers (a sequence-based method: FFAS03\(^\text{23}\) and two combined methods: 3D-PSSM\(^\text{34}\) and PROSPECT II\(^\text{38}\)) to 46 for SPARKS. SPARKS also provides the most accurate model structures based on the total MaxSub score and has the highest specificity. The total MaxSub score increases from between 13.36 and 13.98 for the top three performers to 14.40 for SPARKS. SPARKS also reduces the standard deviation of the total MaxSub score to 0.60, which is comparable to the other top three performers in sensitivity and overall alignment accuracy but a slightly lower value for specificity. The use of the structure-derived score leads to a marked improvement in all three categories (15%, 5%, and 21% in sensitivity, model accuracy, and specificity, respectively). It should be noted that the ranking score based on \(S^{\text{template}} Z_{\text{score}}\) (see Methods) is perhaps the simplest score function for determining specificity among the top performers. For example, FFAS03 uses a sophisticated transformation of \(Z_{\text{score}}\). The 3D-PSSM calculates a theoretical error rate per query (ET).

The above results are limited to 60 targets. The performance of SPARKS for the entire 115 targets on LiveBench 7\(^\text{49}\) (http://bioinf.pl/LiveBench/) is available at the resubmission. Table V compares the results of SPARKS with those of the top performers of single servers. All results were directly taken from LiveBench server. SPARKS provides one of the most accurate model structures based on the total MaxSub score and has one of the highest specificities. The total MaxSub score is 23.85 behind SHGU (25.56), SHUM (25.27), FFAS03 (24.78), and RAPTOR (24.31). The specificity of SPARKS (55.8) is also ranked #5 behind FFAS03 (63.3), FUG2 (57.9), SHGU (57.4) and INBGU (56.3). The number of recognized proteins (sensitivity) is 65. This is rank #7 behind SHUM (71), FFAS03 (70), RAPTOR (69), SHGU (69), 3DPS (67) and PRO2 (66). (The results for the last two methods are not shown.) Among the top performers on the LiveBench server, INBGU, SHGU and SHUM are single servers with built-in multiple methods. INBGU\(^{56}\) is a combination of five methods which exploit sequence and structure information in different ways and produces one consensus prediction of the five.

\(3\text{D-SHOTGUN consensus method}\)\(^{69}\).

**Table IV. Performance for the 60 LiveBench 7 Targets\(^\text{1}\)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Overall accuracy</th>
<th>Specificity</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D-PSSM</td>
<td>39</td>
<td>13.36</td>
<td>19</td>
<td>31</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>36</td>
<td>30.0</td>
</tr>
<tr>
<td>PROSPECT II</td>
<td>41</td>
<td>13.52</td>
<td>33</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>35</td>
<td>34.8</td>
</tr>
<tr>
<td>FFAS03</td>
<td>41</td>
<td>13.98</td>
<td>32</td>
<td>34</td>
<td>34</td>
<td>36</td>
<td>38</td>
<td>38</td>
<td>35.6</td>
</tr>
<tr>
<td>SPARKS</td>
<td>46(40)</td>
<td>14.40(13.68)</td>
<td>31(26)</td>
<td>40(30)</td>
<td>40(34)</td>
<td>41(34)</td>
<td>41(35)</td>
<td>38.4(31.8)</td>
<td>38.4</td>
</tr>
<tr>
<td>3DS3(^{2})</td>
<td>49</td>
<td>17.19</td>
<td>39</td>
<td>42</td>
<td>43</td>
<td>43</td>
<td>46</td>
<td>47</td>
<td>43.4</td>
</tr>
</tbody>
</table>

\(^1\)List of 60 proteins: 1gvnb, 1ghea, 1na5a, 1nxa, 1nb9a, 1n91a, 1l9ja, 1mg5a, 1ms3a, 1mg5a, 1mx0a, 1nx3a, 1rv9a, 1n59a, 1n7aa, 1n3ka, 1nkva, 1kyta, 1g5sa, 1h97a, 1nswa, 1lyga, 1j70a, 1ap6sa, 1nkd, 1nek5a, 1l85a, 1ls5a, 1n81a, 1n67a, 1n60a, 1omia, 1lyga, 1nxxa, 1nxua, 1l1sa, 1hhkea, 1nlyla, 1n57a, 1ngna, 1m5qa, 1kkga, 1gqqa, 1o7de, 1b9ka, 1nqj9a, 1nafa, 1n2ma, 1gwka, 1o22a, 1o29a, 1o1z1a, 1o1xa, 1nt2b, 1tznz, 1lznz

\(^2\)The specificity is defined as the number of recognized proteins with scores better than 1–5 false positives.

\(^3\)Results of other methods are taken from the top three performers of non-consensus methods in LiveBench 7 (not including the single server with built-in multiple methods). Their MaxSub scores are calculated locally based on the predicted models.

\(^4\)Sensitivity is the number of targets whose first-ranking models with a MaxSub score of greater than 0.01.

\(^5\)Total MaxSub score.

\(^6\)The numbers in parenthesis are the results of SPARKS without the structure-derived score.

\(^7\)List of 60 proteins: 1gvnb, 1ghea, 1na5a, 1nxa, 1nb9a, 1n91a, 1l9ja, 1mg5a, 1ms3a, 1mg5a, 1mx0a, 1nx3a, 1rv9a, 1n59a, 1n7aa, 1n3ka, 1nkva, 1kyta, 1g5sa, 1h97a, 1nswa, 1lyga, 1j70a, 1ap6sa, 1nkd, 1nek5a, 1l85a, 1ls5a, 1n81a, 1n67a, 1n60a, 1omia, 1lyga, 1nxxa, 1nxua, 1l1sa, 1hhkea, 1nlyla, 1n57a, 1ngna, 1m5qa, 1kkga, 1gqqa, 1o7de, 1b9ka, 1nqj9a, 1nafa, 1n2ma, 1gwka, 1o22a, 1o29a, 1o1z1a, 1o1xa, 1nt2b, 1tznz, 1lznz

**Table V. Performance for the 115 LiveBench 7 Targets\(^\text{1}\)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity</th>
<th>Total MaxSub score</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>INBGU(^{56})</td>
<td>65</td>
<td>22.34</td>
<td>56.3</td>
</tr>
<tr>
<td>SHGU</td>
<td>69</td>
<td>25.56</td>
<td>57.4</td>
</tr>
<tr>
<td>SHUM</td>
<td>71</td>
<td>25.27</td>
<td>55.1</td>
</tr>
<tr>
<td>FFAS03</td>
<td>70</td>
<td>24.78</td>
<td>63.3</td>
</tr>
<tr>
<td>FUG2(^{24})</td>
<td>63</td>
<td>21.90</td>
<td>57.9</td>
</tr>
<tr>
<td>SPARKS</td>
<td>65</td>
<td>23.85</td>
<td>55.8</td>
</tr>
<tr>
<td>3DS3(^{2})</td>
<td>79</td>
<td>29.96</td>
<td>68.3</td>
</tr>
</tbody>
</table>

\(^1\)Taken from LiveBench server. Listed are individual servers having either better MaxSub score or specificity or both than SPARKS. Meta server 3DS5 is included for comparison.

\(^2\)Sensitivity is the number of targets whose first-ranking models with a MaxSub score of greater than 0.01.

\(^3\)The specificity is define as the average number of recognized proteins that have scores better than 1–10 false positives.

\(^4\)INBGU is a combination of five methods which exploit sequence and structure information in different ways and produces one consensus prediction of the five.

\(^5\)ShotGun-INBGU uses the ShotGun method to create alternative consensus models for the INBGU components. SHUM is an experimental version.

\(^6\)3D-SHOTGUN meta server.\(^{69}\)
for the INBGU components. SHUM is an experimental version. FFAS is a sequence-based method. RAPTOR, FUG2, 3DPS, and PRO2, similar to SPARKS, use a combined sequence and structural information.

It should be noted that all single methods are not as accurate as some consensus methods. The results of the best performer 3D-SHOTGUN\textsuperscript{74} were also shown in Tables III and IV for comparison.

A careful review of our results found that MaxSub scores for some models are reported as zero on the LiveBench server but are nonzero when calculated locally. (We are able to reproduce all nonzero values of MaxSub scores.) This happens not only to SPARKS but also to FFAS03 or other methods. For example, the MaxSub scores for the target 1kkga are zero for SPARKS and FFAS03 on the server but are 0.177 (template 1fxxa) for SPARKS and 0.195 (template 1jmta) for FFAS03, respectively, when calculated locally. We found that the difference is mostly caused by the way to handle missing structural regions in the native structures and nonstandard residues. Because all query sequences are treated as continuous, the corresponding pdb files are changed accordingly. Table VI compares the results of sensitivity, accuracy, and specificity that are reevaluated for FFAS03, RAPTOR, SHGU, and SPARKS. (We only picked a few from Table IV that are mostly likely to perform better than SPARKS.) For FFAS03, RAPTOR and SHGU, we downloaded all the models predicted by them from the LiveBench server. As the table shows, the results for all four methods improve over the published values on the server. SPARKS now has a higher sensitivity (82) and specificity (71.6) than FFAS03 (80 and 66.7, respectively) but have a lower accuracy (25.76) than FFAS03 (26.22) in term of total MaxSub scores. The performance of SPARKS are better than that of RAPTOR in all categories. Compared to SHGU, SPARKS continues to have the highest specificity but has lower sensitivity and accuracy. Thus, the reevaluation of the results significantly improves the rank of SPARKS relative to others.

**DISCUSSION**

In this paper, we have designed a single-body knowledge-based score for fold recognition. The score function contains a backbone torsion term, a buried surface term, and a contact-energy term. Each term is a statistical potential derived from 1011 non-homologous structure database of proteins.\textsuperscript{45} The first and last two terms are designed to take the backbone and side-chain interactions into account, respectively. The relative weights of the three terms were optimized by using well-established 32 decoy sets (Methods). The structure-derived score function is then combined with the sequence profile from PSI-BLAST and the secondary structure information from PSIPRED to obtain a new method for fold recognition (SPARKS). SPARKS is shown to improve sensitivity, alignment accuracy, and specificity over many well-established non-consensus methods. The improvement over other methods is significant in all testing benchmarks.

The success of SPARKS illustrated the power of combining the sequence-based and threading-based approaches in fold recognition.\textsuperscript{29–39} Much of the success is owing to the fact that the sequence profile and the secondary structure information alone yielded a method that has comparable accuracy with many existing well-developed algorithms. Another possible reason is that the SPARKS method, for the first time, incorporates sophisticated backbone and side interactions into the score function. We found that the lack of torsion angle term would significantly reduce the ability of the residue-based potential to discriminate against decoys. Moreover, there is a buried-surface term which quantitatively takes into account the fact that hydrophobic residues are more likely to be buried.\textsuperscript{46}

To further analyze the relative contributions of different terms in the structural score, we test our potential in all other six combinations (torsion only, ASA only, contact only, ASA + contact, torsion + contact, ASA + torsion) without adjusting the weight factors. The results on Prosop benchmark are shown in Table VII. The contact term does not seem to help for alignment accuracy. There is no change in alignment accuracy with the addition of contact term (51.4% versus 51.5%). However, when it was combined with ASA term, the two-term structural score (ASA + contact) becomes the second best (56.8%) in alignment accuracy behind SPARKS (57.2%). The ASA term appears to be the most important. It has the highest accuracy in single-term structure-score function and its combination with either contact or torsion terms yields the best and second best alignment accuracy in the two-term structure-score function. SPARKS (all three terms) provides a small but nontrivial improvement. If the criterion for the accuracy is relaxed to allow a shift of ≲4 residues from the “correct” alignment, the overall picture remains the same, i.e., the ASA term is the most important in the structural score. SPARKS is perhaps one of the simplest methods among the top performers. INBGU, SHGU, and SHUM are single servers with built-in multiple methods. FFAS03 uses a sophisticated transformation of Z\textsubscript{score} to rank models. RAPTOR uses pair (two-body) interactions used in PROSPECT II and thus is computationally more expensive. Thus SPARKS has the potential for further improvement in accuracy of folding recognition.

<table>
<thead>
<tr>
<th>Method</th>
<th>Sensitivity\textsuperscript{a}</th>
<th>Total MaxSub score</th>
<th>Specificity\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAPTOR\textsuperscript{68}</td>
<td>74</td>
<td>24.84</td>
<td>54.8</td>
</tr>
<tr>
<td>FFAS03\textsuperscript{23}</td>
<td>80</td>
<td>26.22</td>
<td>66.7</td>
</tr>
<tr>
<td>SHGU\textsuperscript{c}</td>
<td>88</td>
<td>28.15</td>
<td>67.3</td>
</tr>
<tr>
<td>SPARKS</td>
<td>82(80)\textsuperscript{d}</td>
<td>25.76(24.98)</td>
<td>71.6(59.1)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Sensitivity is the number of targets whose first-ranking models with a MaxSub score of greater than 0.01.

\textsuperscript{b}The specificity is defined as the average number of recognized proteins that have scores better than 1–10 false positives.

\textsuperscript{c}ShotGun-INBGU uses the ShotGun method to create alternative consensus models for the INBGU components.

\textsuperscript{d}The numbers in parenthesis are the results of SPARKS without the structure-derived score.
The use of a single-body potential in SPARKS leads to an efficient algorithm for fold recognition on a genomic scale. We find that the completion of fold recognition of M. genitalium genome (480 genes or proteins, ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/) takes only about a week on a PC with an Athlon 1.4GHz processor without serious effort on code optimization. Thus, an analysis of human proteins can be completed within a few weeks using an inexpensive, small cluster of PCs. Thus, SPARKS can serve as a powerful new tool in structural genomics. Fold recognition is increasingly important as more and more structures with new folds become available. For example, we can resolve 71% of proteins in M. genitalium genome based on a cutoff of -2.9 for specificity ranking score that allows zero false positive in LiveBench 7 (Table VI).

ACKNOWLEDGMENT

We thank Professor D. Fischer for providing us his benchmark and Drs. Y. Xu and D. Xu for their help for installing and using their PROSPECT II program.

REFERENCES