SP5: Improving Protein Fold Recognition by Using Torsion Angle Profiles and Profile-Based Gap Penalty Model

Author
Zhang, Wei, Liu, Song, Zhou, Yaoqi

Published
2008

Journal Title
PLoS One

DOI
https://doi.org/10.1371/journal.pone.0002325

Copyright Statement
Copyright 2013 Zhang et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License CCAL. (http://www.plos.org/journals/license.html)

Downloaded from
http://hdl.handle.net/10072/57587
SP5: Improving Protein Fold Recognition by Using Torsion Angle Profiles and Profile-Based Gap Penalty Model

Wei Zhang1,2, Song Liu3,4,5, Yaoqi Zhou1

1 Indiana University School of Informatics and Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, United States of America, 2 Institute of Applied Physics and Computational Mathematics, Beijing, People’s Republic of China, 3 Department of Biostatistics, Center of Excellence in Bioinformatics & Life Sciences, University at Buffalo, State University of New York, Buffalo, New York, United States of America, 4 Department of Biostatistics, Roswell Park Cancer Institute, Buffalo, New York, United States of America, 5 Howard Hughes Medical Institute Center for Single Molecule Biophysics, Department of Physiology & Biophysics, University at Buffalo, State University of New York, Buffalo, New York, United States of America

Abstract

How to recognize the structural fold of a protein is one of the challenges in protein structure prediction. We have developed a series of single (non-consensus) methods (SPARKS, SP2, SP3, SP4) that are based on weighted matching of two to four sequence and structure-based profiles. There is a robust improvement of the accuracy and sensitivity of fold recognition as the number of matching profiles increases. Here, we introduce a new profile-profile comparison term based on real-value dihedral torsion angles. Together with updated real-value solvent accessibility profile and a new variable gap-penalty model based on fractional power of insertion/deletion profiles, the new method (SP5) leads to a robust improvement over previous SP method. There is a 2% absolute increase (5% relative improvement) in alignment accuracy over SP3 based on two independent benchmarks. Moreover, SP5 makes 7% absolute increase (22% relative improvement) in success rate of recognizing correct structural folds, and 32% relative improvement in model accuracy of models within the same fold in Lindahl benchmark. In addition, modeling accuracy of top-1 ranked models is improved by 12% over SP4 for the difficult targets in CASP 7 test set. These results highlight the importance of harnessing predicted structural properties in challenging remote-homolog recognition. The SP5 server is available at http://sparks.informatics.iupui.edu.

Introduction

Fold recognition refers to recognizing the structural fold of a protein, given its sequence information. Fold recognition is one of the key bottlenecks for protein structure predictions as the protein data bank now appears to contain the complete (or near complete) set for all possible structural folds of proteins, at least for small domain proteins [1,2].

Recently completed assessment of automated servers for protein structure prediction (CASP 7) [3] reveals the power of post-treatment of models predicted by individual fold recognition methods through consensus predictions (For example, ROBETTA [4], Pmodeller [5], Fams-ace [6] and or constrained template-fragment recombination and refinement (For example, Chunk-TASSER [7], I-TASSER [8]). The prediction quality of these methods, however, relies heavily on the accuracy of initial models generated by individual fold recognition methods in the first step. Another observation is that the accuracy of top single servers can rival with most consensus methods. Thus, the accuracy and sensitivity of fold recognition as the number of matching profiles increases [10,11], and predicted versus actual solvent accessible surface area [12]. There is a robust improvement of the accuracy and sensitivity of fold recognition as the number of matching profiles increases [10,11], and [12]. SPARKS, SP5, and SP were ranked among the top performers for automatic servers in recent CASP 6 [13,14] and 7 [12,3]. This exemplifies the importance and effectiveness of multiple-dimensional use of the structural information of templates in developing fold-recognition techniques.

In this paper, we introduce the fifth “dimension” for fold recognition by incorporating predicted backbone torsion angles (SP5). The backbone torsion angles (ϕ and ψ) are two rotation angles about the Cα—N bond (ϕ) and the Cα—C bond (ψ). Because the polypeptide backbone of a protein is a linked sequence of rigid planar peptide groups, these two angles essentially determine the backbone conformation of proteins. While a three-state classification of secondary structures is a coarse-grained one-dimensional representation of local backbone conformation, backbone torsion angles encode the backbone tertiary structure, at least in principle.
Traditionally, dihedral torsion angles are predicted as a few discrete states based on local (fragment) structural patterns using either machine-learning techniques or classification schemes [15–22]. However, there were only a few limited applications of predicted angle states to fold recognition [10] and sequence alignment [23]. The former uses torsion-angle states as a replacement of simple three-state secondary structures to build an iterated alignment hidden Markov model [10]. The latter [23] predicts angle states by hidden Markov model and employs the predicted angles to build structural context-based substitution matrices. Here, we propose to match predicted and actual torsion angles as a new profile term in a multi-dimensional profile-profile alignment. This represents a novel use of predicted torsion angles as a complementary to rather than a replacement of secondary structures for fold recognition. The angle profile used in this work is built on a recent advancement in real-value prediction of torsion angles [24]. By taking advantage of angle periodicity and using integrated neural networks, we have obtained ten-fold cross-validated maximum absolute errors of 38° for ψ and 25° for φ [24]. This accuracy of real-value prediction was found comparable to or more accurate than those based on multi-state classification of the φ–ψ map.

In SP5, the effect of solvation was taken into consideration by matching the predicted and actual solvent accessibility (SA). The SA profiles are based on two states (exposed and buried) classified according to an arbitrary threshold of 25%. The two-state classification increases the accuracy of prediction by reducing number of states in SA. This is at the cost of losing the detailed fluctuation pattern of SA along the sequence. We recently have developed method (called Real-SPINE) for real value SA prediction, which yields a 10-fold cross-validated Pearson’s correlation coefficient (PCC) of 0.74 between predicted and actual solvent SA [25]. We thus have updated the original two-state SA profile with the new real-value one in developing SP5 scoring function.

In addition to the torsion angle and real-value SA term, we will introduce a new variable gap-penalty model to replace the original constant gap-penalty model. The new model is based on insertion and deletion probability profiles generated from PSIBLAST. Several studies [26–28] have indicated the usefulness of these context-dependent profiles for improving alignment accuracy. Here, we propose an implementation by using insertion and deletion probability profiles to a fractional power.

The above-proposed algorithm leads to the new method called SP5. SP5 is tested in two alignment benchmarks and two structure-deletion probability profiles to a fractional power. The ultimate purpose of improving alignment is to make more accurate fold recognition and structure prediction. Lindahl Benchmark is a large data set of 976 proteins, with 555, 434, and 321 pairs of proteins in the same family, superfamly, and fold, respectively [44]. However, DSSP [30] failed to produce results for 9 proteins. Thus, the actually used proteins in this study are 967 and the number of proteins in family, superfamly and fold is 550, 430, and 317, respectively. Here, the fold recognition sensitivity of each method is tested by aligning each protein with the rest 966 proteins, and checking whether or not the method can recognize the member of same family, superfamly or fold as the first rank or within the top 5 ranks. Thus, the benchmark tests both modeling accuracy and ranking methods of proposed methods.

Table 2 shows the fraction of correctly recognized match of proteins in the same family, superfamly, fold as first rank or within top 5 rank of the templates given by various SP methods and several other methods. Although many published methods have been applied to this benchmark [45–47,10], we only list most recent ones [10,11,48,12]. This is because of the time dependent nature of sequence database for sequence profiles. For facilitating

<table>
<thead>
<tr>
<th>Table 1. The alignment accuracies for Prosup and SALIGN benchmark.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SP5</strong></td>
</tr>
<tr>
<td>Prosup</td>
</tr>
<tr>
<td>SALIGN</td>
</tr>
</tbody>
</table>

*One-to-one match given by the method and Prosup.
*One-to-one match given by the method and TMalign.
*Mean value and the standard error (estimated by bootstrap simulation on 10,000 re-sampling of the data set).

doi:10.1371/journal.pone.0002325.0001
the comparison within SP methods, we used original sequence profiles from Ref. [11].

Table 2 indicates that the improvement over SP3 and SP4 in success rate of fold recognition by SP5 exists in all three levels (family, superfamly, and fold). The largest improvement over SP4 is observed in fold level (7% absolute increase in Top 1 and 5% absolute increase for the best in Top 5; 22% relative increase in Top 1, 9.5% in top 5). This is somewhat expected because the method was trained for remote homolog recognition (structurally similar protein with less than 30% sequence identity, PREFAB benchmark). Again the relative improvement of SP5 over SP3 and SP4 is significantly larger than the standard errors estimated from bootstrap simulations. We further removed 43 proteins that have >30% sequence identity with the training sequences in the PREFAB benchmark. Their effect on the final result is negligible. For comparison, we also include the results of PSIBLAST [9], SPARKS® [10], HHsearch / HHpred [27] and FOLDpro [48]. The performance of SPAKRS and Foldpro was from Ref. [10] and Ref. [48], respectively. We further performed PSIBLAST and HHpred locally with their default parameters. Among all methods listed [9–12,27,48], SP5 method has the highest success rate on the fold level (both first and top 5 ranks) and the superfamly for the first rank.

Above success rates of matching sequences within the same SCOP classification are based on somewhat subjective SCOP definition of family, superfamly and fold [49]. A more direct measurement of accuracy is to calculate the accuracy of the first-ranked model built from the fold-recognition alignment. The model is first built by transferring the Cβ coordinates of the template structures to the aligned residues in the query sequence. The constructed model is then assessed by using the MaxSub score of the models built by SP3, SP4 and SP5 methods averaged over the number of proteins. Again SP5 improves over SP4 and SP3 in all levels. The relative improvement of SP5 over SP4 in MaxSub score is 1.4%, 3.1% and 32.2% in family, superfamly and fold levels, respectively.

### CASP7 test set

We use CASP 7 targets [50] as an additional test set for SP5 method. The test set consists of 95 targets and was released between May and July of 2006. The 95 targets were officially classified into 109 template-based-modeling (TBM) domains and 19 free-modeling (FM) domains, based on whether or not the structurally similar template (deposited in PDB) had been identified and used in prediction.

We test SP3, SP4 and SP5 methods on the CASP7 test set. The template library for SP methods was built in the same way. This was done by using the 40% representative domains of SCOP 1.61. The entire chains of multiple-domain proteins are also contained in the library. The library was then updated with new proteins released after SCOP 1.61 if they have less than 40% sequence identity with the sequences already in the library. To make a strict test, we only include template proteins released before May of 2006 for this test, and we also excluded the templates with sequence identity >20% to the query. The performance of SP5 over SP4 in MaxSub score is 1.4%, 3.1% and 32.2% in family, superfamly and fold levels, respectively.

### Table 3. The model quality of top-1 ranked models in Lindahl benchmark per protein.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Total</th>
<th>Family</th>
<th>Superfamily</th>
<th>Fold</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP3</td>
<td>0.358 (±0.03%)</td>
<td>0.529 (±0.05%)</td>
<td>0.232 (±0.05%)</td>
<td>0.107 (±0.05%)</td>
</tr>
<tr>
<td>SP4</td>
<td>0.361 (±0.03%)</td>
<td>0.532 (±0.05%)</td>
<td>0.251 (±0.05%)</td>
<td>0.116 (±0.05%)</td>
</tr>
<tr>
<td>SP5</td>
<td>0.374 (±0.03%)</td>
<td>0.538 (±0.05%)</td>
<td>0.257 (±0.05%)</td>
<td>0.153 (±0.06%)</td>
</tr>
</tbody>
</table>

*All 976 proteins.
*Family only.
*Superfamily only.
*Fold only.
*The mean MaxSub score and the standard error (estimated by bootstrap simulation on 10,000 re-sampling of the data set) for the first-ranked models. 

Table 2. The success rate for recognizing proteins within the same family, superfamily, or fold in the Lindahl benchmark.
Table 4. The model quality of top-1 ranked models for CASP7 test set.

<table>
<thead>
<tr>
<th></th>
<th>Fulla</th>
<th>ALLe</th>
<th>TBMf</th>
<th>FMg</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP3</td>
<td>0.364 (±0.20%)</td>
<td>0.375 (±0.17%)</td>
<td>0.408 (±0.19%)</td>
<td>0.152 (±0.37%)</td>
</tr>
<tr>
<td>SP4</td>
<td>0.373 (±0.20%)</td>
<td>0.387 (±0.17%)</td>
<td>0.420 (±0.19%)</td>
<td>0.153 (±0.32%)</td>
</tr>
<tr>
<td>SP5</td>
<td>0.383 (±0.21%)</td>
<td>0.397 (±0.17%)</td>
<td>0.431 (±0.18%)</td>
<td>0.171 (±0.38%)</td>
</tr>
</tbody>
</table>

a95 full chain targets.
All 124 domains (There are 4 targets belonging to both TBM and FM categories).
109 Template-based Modeling domains.
19 Free Modeling domains.
*The mean Maxsub score and the standard error (estimated by bootstrap simulation on 10,000 re-sampling of the data set) for top 1 model.

doi:10.1371/journal.pone.0002325.t004

different SP method is evaluated by the Maxsub score of the first ranked Cα model, which is transferred from the alignment.

Table 4 compares the model quality predicted by SP3, SP4, and SP5. Overall, there is a consistent 3% (5% to 6%) improvement from SP3 to SP5 (SP5) for the CASP 7 targets regardless the evaluation based on domains or full chains. For the 109 TBM domains, SP5 is 3% (6%) better than SP3 (SP4). For the most difficult free-modeling targets, there is a 12% improvement from SP3 to SP5. This pattern of improvement is consistent with that from Lindahl benchmark. That is, the most significant improvement from SP3 to SP5 is on the most challenging targets.

Discussion

This paper reports several significant changes over previously developed SP method: the torsion-angle term for profile-profile matching, real-value-based SA profile, and variable gap-penalty model based on fractional-powered insertion/deletion profiles. We showed that by integrating these new features with existing sequence-derived profile, secondary structure profile, residue depth-dependent structure-based profile, the new method SP5 makes a robust improvement over previously developed SP serial methods. Comparing with SP3 and SP4, there is a 2-6% absolute improvement in one-to-one match of alignment accuracy depending on benchmarks. Application of SP5 to the large Lindahl benchmark reveals 1%, 2% and 7% improvements over SP3 in success rates in recognizing proteins within the same family, superfamily and fold, respectively. The improvement in recognition leads to 1%, 3% and 32% improvement in modeling accuracy based on the top-1 ranked, family, superfamily and fold-level models, respectively. Additional test on CASP 7 targets yields 3-6% improvement in 109 template-based modeling targets and 12% improvement in 19 free-modeling targets. Thus, SP5 marks a significant improvement over SP3 and SP4 in fold-recognition, as designed.

This paper represents a full exploitation of predicted torsion-angles for fold recognition. Previous similar studies [18,23] are limited to view discrete torsion-angle states as an expansion of secondary structures. This paper, however, treats predicted angles as complementary information to predicted three-state secondary structures. The two quantities are complementary because three-state secondary structures represent a coarse-grained description of local structures while torsion angles contain detailed local and nonlocal structural information if they are predicted accurately. Indeed, our limited initial test indicates that removing secondary structures from SP3 will reduce its alignment accuracy. Obviously, the success of SP5 is made possible because of reasonably accurate real-value prediction of torsion angles [24].

Recent progress in sequence alignment and structure prediction has suggested the importance of variable gap penalties in protein sequence alignment [51]. Different form of context (either structure or sequence context or both)-dependent gap-penalty model has been proposed [52,53]. Employing fractional-powered gap insertion/deletion profiles is another new feature introduced in SP5. While these insertion/deletion profiles were used, previously [26–28], our trial-and-error analysis indicates that the fractional-powered gap insertion/deletion profiles with a power of 0.1 seem to be more suitable for improving alignment accuracy. However, more systematic comparative studies are needed to check if any other functional forms are more appropriate.

To analyze the usefulness of the new gap model, we made a version of SP5 with the previously used gap model and found that new gap model leads to a small but positive increase in alignment accuracy (0.5% in PREFAB, 1.5% in ProSup and 0.1% in SALIGN). Thus, the main contribution for improved ability in fold recognition by SP5 is due to introduction of torsion angles.

SP3 and SP4 were among the top performers in automatic servers in CASP 6 and 7 [13,12]. It is noted that in CASP7, SP3 scored higher than SP4 according to GDT-HA, TMscore, and ALIO for all targets. A close examination [12] indicates that SP4 is slightly more accurate than SP5 in hard targets (FM category), but slightly worse than SP3 in other targets (TBM category). This is perhaps because all parameters were optimized for fold recognition targets. On the other hand, SP5 performs consistently better than SP3 at both FM and TBM categories if the cumulative Z-score is used [12]. The development of SP5 continues our emphasis on searching a more sensitive method for fold recognition. Significant improvement of SP5 over SP4 and SP3 indicates that SP5 is among the most accurate automatic servers for fold recognition.

In the SP serial methods, the alignment generated for fold recognition is used directly in modeling. It is quite possible that a separate alignment method optimized for modeling may further improve the accuracy of predicted model. This will be a subject of future studies.

Methods

Alignment Score

The alignment score of SP4 for aligning query position i with the template position j is

\[ S(i,j) = - (1 - w_{\text{struc}}) F_{\text{seq}}(i) \odot M_{\text{template}}(j) \]

\[ - w_{\text{struc}} F_{\text{seq}}(j) \odot M_{\text{query}}(i) \]

\[ - w_{\text{sym}} \delta_{ij} - w_{\text{align}} (1 - 2|sa(i) - sa(j)|) \]

\[- w_{c}(1 - d/90) + s_{\text{shift}} \]

with four weight parameters \(w_{\text{struc}}, w_{\text{sym}}, w_{\text{align}}, w_{c}\) and a constant shift \(s_{\text{shift}}\). This score represents weighted matching of five profiles that are described in detail below.

The first term in Eq. (1) is the profile-profile comparison between the sequence profile from the query sequence and that from the template sequence. \(F_{\text{seq}}(i)\) is the sequence-derived frequency profile of the query sequence. \(M_{\text{template}}(j)\) and \(M_{\text{query}}(i)\) are the sequence-derived log odd profile of the template sequence and that of query sequence, respectively. These sequence profiles are constructed by three iterations of PSI-BLAST [9] searching \((E\ value\ cutoff\ 0.001)\) against non-redundant (NR) sequence database, which was filtered to remove low-complexity regions, transmembrane regions, and coiled-coil segments [29].
The second term in Eq. (1) compares the sequence profile from the query sequence and that derived from the template sequence (sequence profiles that would “fit” to the structure). \( F_{\text{template}}(j) \) is a depth-dependent sequence profile generated from the sequences of those structural fragments that are similar to 9-residue segment structures of the template [11].

The second term in Eq. (1) measures the difference between the predicted secondary structure of the query sequence and the actual secondary structure of the template. \( \delta_{si,j} \) is a simple function of the secondary structure element \( si \) of the query at sequence position \( i \) and \( sj \) of the template at sequence position \( j \). If \( si = sj \), then \( \delta_{si,j} = 1 \); otherwise \( \delta_{si,j} = -1 \). We use a three-state definition of secondary structures: \( H \) for helix, \( E \) for strand, and \( C \) for coil. The secondary structure for query sequences is predicted by SPINE [31]. The first three terms constitute the method SP3 [11] except that SPINE was employed to predict the secondary structure of the query sequence. DSSP [30] is used for analyzing template structures because SPINE was trained based on the DSSP definition of secondary structures.

The fourth term in Eq. (1) is the matching score between the predicted solvent accessibility of the query sequence and solvent accessibility of the template structure. \( s\text{a}(i) \) and \( s\text{a}(j) \) are the predicted residue solvent accessibility of query sequence and that of the template structure, respectively. The residue solvent accessibilities of query sequence are predicted by Real-SPINE [25] while residue solvent accessible areas of template structures are calculated from DSSP [30] and normalized by unfolded solvent accessible surface areas [32]. The first four terms constitute the method SP5 [33] except that in SP5, PSIPRED [29] rather than SPINE [31] was employed to predict the secondary structure of the query sequence, and the real values of solvent accessibility from Real-SPINE [25] rather than two-state classifications by SABLE [34] are used to predict the residue solvent accessibility of the query sequence.

The fifth term in Eq. (1) is a new addition in SP5. It characterizes the difference between predicted angles (\( \psi(j) \) and \( \phi(j) \)) of the query sequence and actual angles (\( \psi(j) \) and \( \phi(j) \)) of the template structure with

\[
\Delta = \sqrt{\frac{1}{2} \left[ (\psi(i) - \psi(j))^2 + (\phi(i) - \phi(j))^2 \right]}
\]

Real values of angles for the query sequence are from Real-SPINE 2.0 [24] while these angles are calculated by DSSP [30] for the template structure. Real-SPINE 2.0 is a method for real-value prediction of torsion angles by using back-propagation neural networks trained with a sliding 21-residue window of sequence profiles, representative amino acid properties, and predicted secondary structures. The ten-fold-cross-validated mean absolute errors are 38° for \( \psi \) and 25° for \( \phi \), respectively.

Profile-based Gap Model

SP3 [11] and SP5 [12] employ a simple secondary-structure dependent gap penalty. No gaps are allowed if \( si = sj = \alpha \) (helix) or \( si = sj = \beta \) sheet). The gap opening (\( \alpha \text{gap} \)) and gap extension (\( \alpha \text{ext} \)) penalties are applied to other regions. In this paper, we construct a profile-based gap model from the multiple sequence alignment made by PSIBLAST [9]. The multiple sequence alignment allows us to calculate the probability of deletion at sequence position \( i, P_{\text{del}}(i) \), and the probability of insertion at sequence position \( i, P_{\text{ins}}(i) \). \( P_{\text{del}}(i) = n_{\text{del}}^i / N \) and \( P_{\text{ins}}(i) = n_{\text{ins}}^i / N \) where \( n_{\text{del}}^i \) and \( n_{\text{ins}}^i \) are number of deletions in sequence position \( i \), and total number of aligned sequences, respectively.

Thus, we have four profiles: two for query sequences and two for template sequences \( P_{\text{del(query)}}(i), P_{\text{ins(query)}}(i), P_{\text{del(template)}}(i), \) and \( P_{\text{ins(template)}}(i) \).

The gap penalty is calculated as follows. We still use \( \omega_{\text{gap}} \) as the gap opening penalty. The extension gap penalty is modified by

\[
w_1 = 1 - \frac{\left( \left( P_\text{del(query)}(i) \right) + \left( P_\text{ins(template)}(i) \right) \right)^2}{2} \quad \text{for residue } i \text{ in the query sequence that is aligned with a gap after residue } j \text{ in template.}
\]

Similarly, the extension gap penalty is modified by

\[
w_1 = 1 - \frac{\left( \left( P_\text{ins(query)}(i) \right) + \left( P_\text{del(template)}(i) \right) \right)^2}{2} \quad \text{or residue } j \text{ in template that is aligned with a gap after residue } i \text{ in query.}
\]

Here, \( w_1 \) is a to-be-optimized weight factor. Usually, \( \ln(P_{\text{del/insert(query/template)}}(i)) \) is an energetic term. Here, we use \( \ln(P_{\text{del/insert(query/template)}}(i)) \) rather than \( \ln(P_{\text{del/insert(query/template)}}(i)) \) to avoid singularity at \( P_{\text{del/insert(query/template)}}(i) = 0 \).

\[
w_1 = 0.1 \text{ by trials and errors.}
\]

Dynamic Programming and Template Ranking

Similar to SP3 and SP5, we used the Smith-Waterman local alignment algorithm [35] to optimize the score that matches the query profiles with template profiles based on Eq. (1) with the revised gaping method described above. Note that the optimization of alignment is to minimize the total alignment score due to the negative signs in Eq. (1).

The templates are ranked based on the difference score between the raw alignment score and the reverse alignment raw score in which the alignment is made with the reversed query sequence [36]. The results of fold-recognition alignment are used to build Cα models based on native template structure. This is done by directly transferring the Cα coordinates of the template structures to the aligned residues in the query sequence. If there is no structural similarity between first two models (defined as zero MaxSub score [37]), templates will be re-ranked by the greater one of two Z-scores, which are calculated based on the raw alignment score normalized by the full alignment length and the non-end-gap alignment length, respectively. Here, the Z-score for a template \( i \) is given by

\[
Z(i) = \frac{S(i) - S_{\text{ave}}}{S_{\text{sd}}},
\]

where \( ave \) and \( sd \) denotes the average and standard deviation of normalized score for all the templates. This ranking mechanism was based on an empirical observation. We found that ranking based on the difference score between the raw alignment score and the reverse alignment raw score works well only if there is some structural similarity between the top two ranked models (i.e. a significant structural cluster detected). Otherwise, ranking based on Z-scores works better [11].

Author Contributions

Conceived and designed the experiments: YZ. Performed the experiments: WZ SL. Analyzed the data: WZ SL. Wrote the paper: YZ WZ SL.

References
