

## Research Article

# Statistical Analysis of Terminal Extensions of Protein $\beta$ -Strand Pairs

Ning Zhang,<sup>1</sup> Shan Gao,<sup>2</sup> Lei Zhang,<sup>3</sup> Jishou Ruan,<sup>2</sup> and Tao Zhang<sup>3</sup>

<sup>1</sup> Department of Biomedical Engineering, Tianjin University, Tianjin Key Lab of BME Measurement, Tianjin 300072, China

<sup>2</sup> College of Mathematical Sciences and LPKM, Nankai University, Tianjin 300071, China

<sup>3</sup> College of Life Sciences, Nankai University, Tianjin 300071, China

Correspondence should be addressed to Jishou Ruan; [jsruan@nankai.edu.cn](mailto:jsruan@nankai.edu.cn) and Tao Zhang; [zhangtao@nankai.edu.cn](mailto:zhangtao@nankai.edu.cn)

Received 15 July 2012; Revised 30 December 2012; Accepted 30 December 2012

Academic Editor: Bhaskar Dasgupta

Copyright © 2013 Ning Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The long-range interactions, required to the accurate predictions of tertiary structures of  $\beta$ -sheet-containing proteins, are still difficult to simulate. To remedy this problem and to facilitate  $\beta$ -sheet structure predictions, many efforts have been made by computational methods. However, known efforts on  $\beta$ -sheets mainly focus on interresidue contacts or amino acid partners. In this study, to go one step further, we studied  $\beta$ -sheets on the strand level, in which a statistical analysis was made on the terminal extensions of paired  $\beta$ -strands. In most cases, the two paired  $\beta$ -strands have different lengths, and terminal extensions exist. The terminal extensions are the extended part of the paired strands besides the common paired part. However, we found that the best pairing required a terminal alignment, and  $\beta$ -strands tend to pair to make bigger common parts. As a result, 96.97% of  $\beta$ -strand pairs have a ratio of 25% of the paired common part to the whole length. Also 94.26% and 95.98% of  $\beta$ -strand pairs have a ratio of 40% of the paired common part to the length of the two  $\beta$ -strands, respectively. Interstrand register predictions by searching interacting  $\beta$ -strands from several alternative offsets should comply with this rule to reduce the computational searching space to improve the performances of algorithms.

## 1. Introduction

The issue of protein structure prediction is still extremely challenging in bioinformatics [1, 2]. Usually, structural information for protein sequences with no detectable homology to a protein of known structure could be obtained by predicting the arrangement of their secondary structural elements [3]. As we know, the two predominant protein secondary structures are  $\alpha$ -helices and  $\beta$ -sheets. However, a combination of the early suitable  $\alpha$ -helical model systems and sustained researches have resulted in a detailed understanding of  $\alpha$ -helix, while comparatively little is known about  $\beta$ -sheet [4]. Tertiary structures of  $\beta$ -sheet-containing proteins are especially difficult to simulate [3, 5]. Unlike  $\alpha$ -helices,  $\beta$ -sheets are more complex resulting from a combination of two or more disjoint peptide segments, called  $\beta$ -strands. Therefore, the  $\beta$ -sheet topology is very useful for elucidating protein folding

pathways [6, 7] for predicting tertiary structures [3, 8–11], and even for designing new proteins [12–14].

As fundamental components,  $\beta$ -sheets are plentifully contained in protein domains. In a  $\beta$ -sheet, multiple  $\beta$ -strands held together linked by hydrogen bonds and can be classified into parallel and antiparallel direction styles. Adjacent  $\beta$ -strands bring distant residues on sequences into close special contact with one another and constitute a specific mode of amino acid pairing [1, 15–17], interactions (like DNA base pairing). There is a growing recognition of the importance of the strand-to-strand interactions among  $\beta$ -sheets [18]. Several studies, including statistical studies examining frequencies of nearest-neighbor amino acids in  $\beta$ -sheets, found a significantly different preference for certain interstrand amino acid pairs at nonhydrogen-bonded and hydrogen-bonded sites [1, 17, 19, 20], Dou et al. [21] created a comprehensive database of interchain  $\beta$ -sheet (ICBS)

interactions. We also developed the SheetsPair database [22] to compile both the interchain and the intrachain amino acid pairs.

Generally speaking, previous work on  $\beta$ -sheets mainly focused on the interresidue contacts or amino acid partners [23–28]. Prediction of inter-residue contacts in  $\beta$ -sheets is interesting, while the prediction by ab initio structure is also useful to understand protein folding [29, 30]. Our previous studies showed that the interstrand amino acid pairs played a significant role to determine the parallel or antiparallel orientation of  $\beta$ -strands [15], and the statistical results could possibly be used to predict the  $\beta$ -strand orientation [16]. Cheng and Baldi [11] introduced BETAPRO method to predict and assemble  $\beta$ -strands into a  $\beta$ -sheet, in which a single misprediction of one amino acid pairing from the first stage could be amplified by the next stages and results in serious wrong set of partner assignments between  $\beta$ -strands. However, those studies can be viewed as initial steps of  $\beta$ -sheet studies relative to predict strand level pairing [25]. In this paper, to go one step further, we investigate the  $\beta$ -strand pairing on the strand level for exploring the rules of how  $\beta$ -strands form a  $\beta$ -sheet.

Many results have shown the importance of statistical analysis in protein structure studies [15, 16]. In particular, statistical information could provide a starting point for de novo computational design methods that are now becoming successful for short, single-chain proteins [14], as well as methods of protein structure predictions and understanding of protein folding mechanisms [31, 32]. Fooks et al. [1] also indicated that such statistical analysis results would be useful for protein structure prediction. Therefore, we advocate using the tools of statistics and informatics to study  $\beta$ -sheet and generate new rules for algorithm development. In this study, we focused on the terminal extensions of paired  $\beta$ -strands.

## 2. Results

**2.1. Dataset.** All protein structure data used in this study were taken from a PISCES [33, 34] dataset generated on May 16, 2009. In the dataset, the percentage identity cutoff is 25%, the resolution cutoff is 2.0 angstroms, and the  $R$ -factor cutoff is 0.25. Secondary structures were assigned from the experimentally determined tertiary structures by using the DSSP program. Besides proteins containing disordered regions [35–37], all data were further preprocessed according to the following criteria: (i) no  $\beta$ -sheet-containing protein chains were removed; (ii) protein chains with nonstandard three-letter residue names (such as DPN, EFC, ABA, C5C, PLP, etc.) were removed, since these indicate that the protein chains have covalently bounded ligands or modified residues; (iii) protein chains with uncertain structures or incorrect data were removed. Since  $\beta$ -bulges tend to be isolated and rare [11], we did not consider  $\beta$ -bulges in this study either, as several previous studies did [1, 3]. Finally, 2,315 protein chains were extracted, containing 19,214  $\beta$ -strand pairs. Note that in the special case of  $\beta$ -bulges, no amino acid pair is assigned.

**2.2. The  $\beta$ -Sheet Structure.** The  $\beta$ -sheets, where two or more  $\beta$ -strands are arranged in a specific conformation, are illustrated in Figure 1(a), by a protein example (PDB code 1HZT). Adjacent strands, or the so-called strand pairs, can either run in the same (parallel) or in the opposite (antiparallel) direction styles. In protein 1HZT, there are 3  $\beta$ -sheets called A, B, and C, formed by 10 different  $\beta$ -strands numbered from 1 to 10, making 7 different  $\beta$ -strand pairs, respectively. The 10  $\beta$ -strands can be named by the  $\beta$ -sheet each belongs to and the index numbers in the order of partnership. For example, the 3  $\beta$ -strands forming  $\beta$ -sheet A can be called “A1,” “A2,” and “A3,” while other 4  $\beta$ -strands forming  $\beta$ -sheet B can be called “B1,” “B2,” “B3,” and “B4,” respectively. “A1-A2,” “A2-A3,” “B1-B2,” “B2-B3,” and “B3-B4” are all  $\beta$ -strand pairs. Sequences of the 10  $\beta$ -strands with their initial and ending residue numbers are also given in Figure 1(b).

**2.3. Different Lengths of Paired  $\beta$ -Strands.** For a  $\beta$ -strand pair, the terminal of one  $\beta$ -strand does not always align with the terminal of the other (Figure 2), making “terminal extensions” besides the common paired parts. Note that only amino acids in the common part construct amino acid pairs.

Why “terminal extensions” exist widely in  $\beta$ -strand pairs? We firstly investigated the lengths of two paired  $\beta$ -strands and then calculated the percent of each case whether the “terminal extensions” exist or not. Results are shown in Table 1.

As shown in Table 1, the two paired  $\beta$ -strands having the same length only account for 29.53% of all samples. In other 70.47% percent of samples, lengths of the two paired  $\beta$ -strands are different.

**2.4. Statistical Results of Variables.** We define the following variables.

- (1) Let  $SL_1$  and  $SL_2$  represent the lengths of two paired  $\beta$ -strands, respectively. Length of the  $\beta$ -strand with smaller strand number (strand numbers can be obtained from PDB database) is defined as  $SL_1$ , while length of the other  $\beta$ -strand is defined as  $SL_2$ .
- (2) Let  $PL$  stand for the length of the common part, which is often smaller than  $SL_1$  and  $SL_2$ .
- (3) Terminal extensions can be found in either of the two  $\beta$ -strands. We define the lengths of the two terminal extensions  $Et_1$  and  $Et_2$ , respectively. Length of the terminal extension of the  $\beta$ -strand with length  $SL_1$  is defined as  $Et_1$  while the other as  $Et_2$ .
- (4) Let  $EL$  represent the whole length;  $EL = PL + Et_1 + Et_2$ .

Then, the pairing ratio  $R$  could be calculated by

$$R = \frac{PL}{EL} \times 100\% = \frac{PL}{PL + Et_1 + Et_2} \times 100\%. \quad (1)$$

The ratio of the common paired part to the length of each  $\beta$ -strand ( $i = 1, 2$ ) could be calculated by

$$Rt_i = \frac{PL}{SL_i} \times 100\%, \quad i = 1, 2. \quad (2)$$

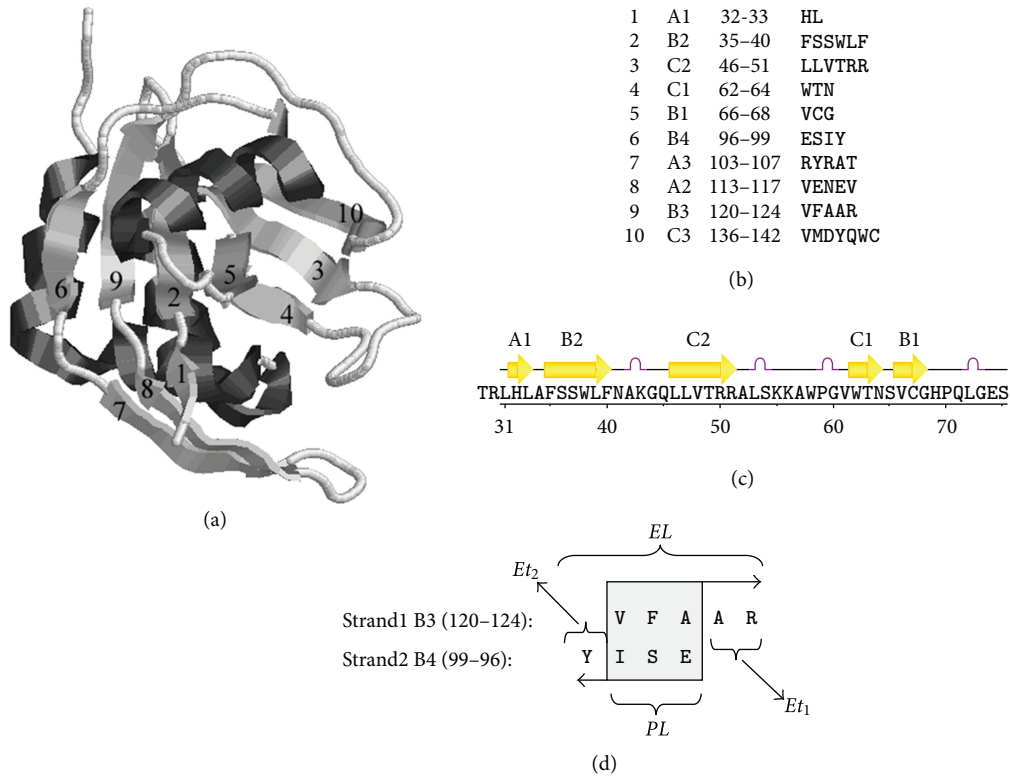


FIGURE 1: An illustrated example of  $\beta$ -strand pairing in a  $\beta$ -sheet (PDB code: 1HZT). (a) The sketch of the tertiary structure of the protein produced by using RASMOL. Protein 1HZT is an  $\alpha/\beta$  protein with 10  $\beta$ -strands numbered from 1 to 10, forming seven different strand pairs. (b) The sequences of the 10  $\beta$ -strands with their initial and ending residue numbers. (c) The 10  $\beta$ -strands in the linear primary sequence. (d) An example of a  $\beta$ -strand partnership graph. The pairing is between strands “B3” and “B4,” with the light gray box representing the common pairing part.

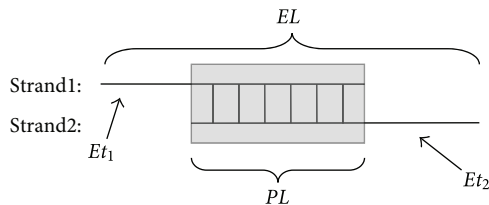


FIGURE 2: A schematic diagram of terminal extensions of  $\beta$ -strand pairs. The two blank lines represent the two  $\beta$ -strands, respectively. The light gray box represents the common pairing part of the two  $\beta$ -strands with amino acid pairing.

A small percent of  $\beta$ -strand pairs have no “terminal extensions,” the  $R$ ,  $Rt_1$ , and  $Rt_2$  values for which will be 100%.

We calculated  $PL$ ,  $Et_1$ ,  $Et_2$ ,  $EL$  for all  $\beta$ -strand pairs in the present dataset. Table 2 gives the range of these variables as well as the averages and standard deviations.

We also calculated  $R$ ,  $Rt_1$ , and  $Rt_2$  for all  $\beta$ -strand pairs in the present dataset. The distribution of these variables is shown in Figure 3.

### 3. Discussion

3.1. *Strands Tend to Align Their Terminals.* For the 70.47% of samples with different strand lengths, although they have

different lengths, the differences are not big for most of them. Only a small percent of samples (below 2.09%) have the difference above 5. In these cases, it is obvious that they cannot align the terminals (with both  $Et_1 = 0$  and  $Et_2 = 0$ ). They have two ways to choose from: either align to only one terminal making another “terminal extension”, or align to none of the two terminals making both “terminal extensions.” However, it can be seen from Table 1 that most  $\beta$ -strands tend to be in the former case. For example, in case of the length difference 1, the former case accounts for 85.18% while the latter only 14.82%. It is consistent with the case of same-length strand pairs, in which  $\beta$ -strands tend to align their terminals with each other. Interestingly, it is suggested that  $\beta$ -strands tend to align their terminals. In different-length strand pairs, they still retain one terminal alignment, although they can not align both ends.

3.2. *Small “Terminal Extensions”.* From Table 2, it can be seen that lengths of  $\beta$ -strands are not very long, ranging from 1 to 25 with an average length about 4-5 amino acids. The averages and the standard deviations are similar between lengths of the two paired  $\beta$ -strands ( $SL_1$  and  $SL_2$ ).

The length of the common part  $PL$  has a range similar to that of lengths of  $\beta$ -strands. This indicates that although “terminal extensions” exist, common pairing parts occupy most of  $\beta$ -strands, while “terminal extensions” occupy least.

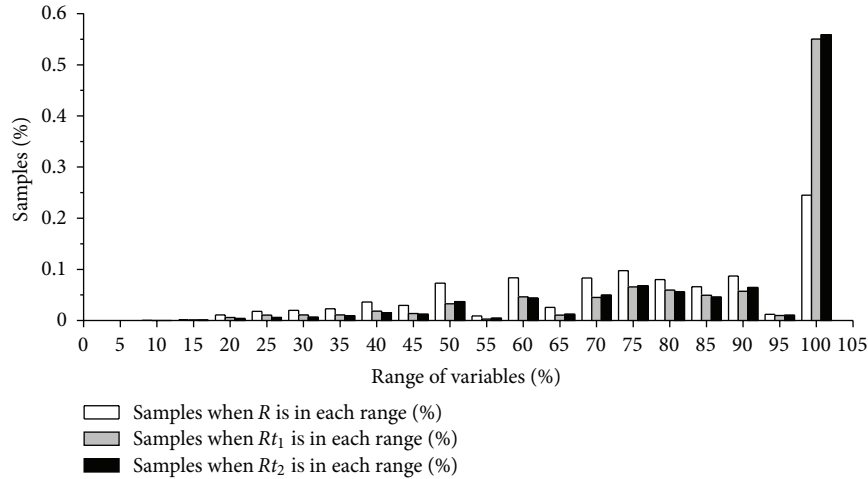


FIGURE 3: Distribution of  $R$ ,  $Rt_1$ , and  $Rt_2$  variables in the current dataset.

TABLE 1: Statistical results of lengths of two paired  $\beta$ -strands and percent of samples in each case whether the two “terminal extensions” exist or not.

Abs ( $SL_1 - SL_2$ )*	Number of pairs	Percent	Percent of $Et_1 = 0$ and $Et_2 = 0$	Percent of $Et_1 = 0$ and $Et_2 > 0$	Percent of $Et_1 > 0$ and $Et_2 = 0$	Percent of $Et_1 > 0$ and $Et_2 > 0$
0	5673	29.53%	82.95%	0.00%	0.00%	17.05%
1	5633	29.32%	0.00%	40.48%	44.70%	14.82%
2	3170	16.50%	0.00%	30.57%	41.10%	28.33%
3	1798	9.36%	0.00%	28.59%	31.15%	40.27%
4	1016	5.29%	0.00%	26.77%	29.82%	43.41%
5	618	3.22%	0.00%	29.13%	27.51%	43.37%
6	401	2.09%	0.00%	30.42%	25.69%	43.89%
7	323	1.68%	0.00%	25.39%	32.51%	42.11%
8	247	1.29%	0.00%	16.19%	30.36%	53.44%
9	101	0.53%	0.00%	26.73%	24.75%	48.51%
10	69	0.36%	0.00%	20.29%	39.13%	40.58%
>10	165	0.86%	0.00%	15.15%	27.27%	57.58%

\* Absolute value of the difference of  $SL_1 - SL_2$ .

TABLE 2: Statistical results of variables of  $\beta$ -strand pairs in the current dataset.

	Minimum value	Maximum value	Average	Standard deviation
$SL_1$	1	25	4.99	2.82
$SL_2$	1	25	4.90	2.80
$PL$	1	23	4.86	2.26
$Et_1$	0	18	1.15	1.79
$Et_2$	0	22	1.03	1.64
$EL$	2	29	7.03	3.09

The fact that the maximum value of  $EL$  is 29, only a little bigger than that of lengths of  $\beta$ -strands, and the fact that in average both the “terminal extensions” only have about 1 amino acid ( $Et_1 = 1.05$  and  $Et_2 = 1.03$ ) also support this assumption.

Figure 3 gives percent of samples for  $R$ ,  $Rt_1$ , and  $Rt_2$  in each range of their possible values (from 0% to 100%), respectively. It can be seen that the distributions of  $Rt_1$  and  $Rt_2$  are similar. More than half of the  $\beta$ -strand pairs have these two variables above 95% (or in the range (95–100)). Big  $Rt_1$  or  $Rt_2$  means big common part of  $\beta$ -strands, or small “terminal extensions.” Rare  $\beta$ -strand pairs have smaller values of  $R$ ,  $Rt_1$ , and  $Rt_2$ , which indicates that most  $\beta$ -strands do not pair by means of small “common part” or big “terminal extensions.” It could be concluded from the results that  $\beta$ -strands tend to pair with bigger pairing common parts, leaving smaller “terminal extensions.”

3.3. Possible Reasons for  $\beta$ -Strand Extensions. Why “terminal extensions” exist so widely in  $\beta$ -strand pairs? The fact that lengths of two paired  $\beta$ -strands are not the same in most cases as shown in Table 1 may be one of the possible reasons. If paired  $\beta$ -strands have the same lengths, most of them

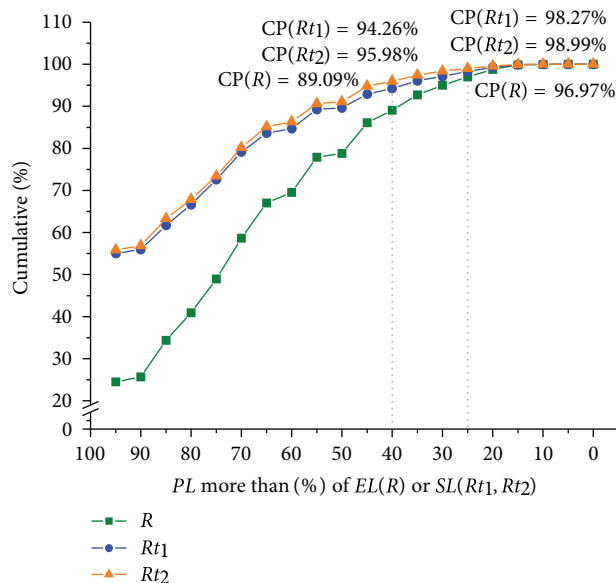


FIGURE 4: Cumulative percentages (CPs) of  $R$ ,  $Rt_1$ , and  $Rt_2$  calculated from the present dataset. The horizontal axis denotes the percentage of common paired region  $PL$  to  $EL$  (for curve  $R$ ) or to  $SL$  (for curves  $Rt_1$  and  $Rt_2$ ). Points on the  $R$  curve denote the cumulative percentages of samples whose  $R = PL/EL$  equals or is bigger than the corresponding abscissa value. Points on the  $Rt_1$  and  $Rt_2$  curves denote the cumulative percentages of samples whose  $Rt_1 = PL/Rt_1$  or  $Rt_2 = PL/Rt_2$  equals or is bigger than the corresponding abscissa value, respectively.

(82.95%) tend to align their terminals with each other, leaving no “terminal extensions.”

A  $\beta$ -strand is led to pair with another by several kinds of potential forces. Steward and Thornton [3] indicated that a single  $\beta$ -strand was still able to recognize a noninteracting  $\beta$ -strand with greater accuracy than that in the case of between two random sequences. The potential forces include hydrogen bonds, van der Waals forces, electrostatic interaction, ionic bonds, hydrophobic effects, and so forth. Parisien and Major [38] revealed that among all the forces, the most important one was the construction of a hydrophobic face. It is conceivable that one residue of a  $\beta$ -strand prefers to pair with the residue of another resulting in a stable state of hydrophobic effects. Optimizing such interactions may result in extensions, which could be the second reason, since more often than not the “terminal alignment” is not the case of optimized pairing style.

A third possible reason could be due to the nucleation events that initiate the  $\beta$ -sheet folding. Amino acids in the central part could pair firstly and then fold to extend to terminals.

Another reason is the roles of the nonpaired terminal amino acids in stabilizing the  $\beta$ -sheet structure. Several other studies have identified their key roles in modulating protein folding rates, stability, and folding mechanism [39–43]. Therefore, the  $\beta$ -strand terminals could also be important factors for a  $\beta$ -sheet formation.

**3.4. Ratio Rule of Pairing Strand Alignment.** To quantify the pairing common part of paired  $\beta$ -strands, we calculated the cumulative percent of variables  $R$ ,  $Rt_1$ , and  $Rt_2$  and depicted them in Figure 4.

From Figure 4, it can be seen that when  $Rt_1 \geq 40\%$  and  $Rt_2 \geq 40\%$ , the cumulative percentages reach 94.26% and 95.98%, respectively, while when  $R \geq 40\%$  only 89.89%. When  $R \geq 25\%$ , the cumulative percentages reach up to 96.97%. Therefore, a rule can be made of the alignment of  $\beta$ -strand pair as follows:

$$R \geq 25\%, \quad Rt_i \geq 40\%, \quad i = 1, 2. \quad (3)$$

Almost all samples (above 94%) obey this rule.

In a  $\beta$ -strand alignment prediction algorithm, all possible pairings should be examined and scored; it is a time-consuming task. Kato et al. [44] stated that prediction of planar  $\beta$ -sheet structures was NP-hard in the present state of our knowledge (<http://en.wikipedia.org/wiki/NP-hard>). However, this previous rule should be used as a constraint of the relative positions in  $\beta$ -strand alignment to reduce the computational searching space, which could be used to develop high-speed  $\beta$ -strand topology prediction algorithms.

## 4. Conclusion

At the most straightforward level, full “identification” of a  $\beta$ -strand pair could consist of (i) finding the interacting partner  $\beta$ -strand(s), (ii) predicting the relative orientation (i.e. parallel or antiparallel), and (iii) shifting the relative positions of the two interacting  $\beta$ -strands [15, 16]. In this study, we focused on the third aspect. The formation of protein structure and protein folding mechanism are very complex, and the mechanisms of  $\beta$ -sheet formation are unclear [45]. However, simple rules could contribute to developing new algorithms in the step of full prediction of  $\beta$ -sheet and understanding of protein folding pathways in ongoing research.

In this study, to go one step further, we studied  $\beta$ -sheets on the strand level instead of amino acid level. Statistical analyses of the terminal extensions of paired  $\beta$ -strands were performed and a simple rule “ $R \geq 25\%$  and  $Rt_i \geq 40\%$ ,  $i = 1, 2$ ” was made. Steward and Thornton [3] developed an information theory approach to predict the relative offset positions by shifting one  $\beta$ -strand up to 10 residues either side of that observed. Such a rule could be used in similar studies. We certainly believe that the conclusions presented in this study could contribute to predict protein structures and to develop  $\beta$ -sheet prediction methods.

## Conflict of Interests

The authors have declared that no conflict of interests exists.

## Acknowledgment

This work was supported by Grants from the National Natural Science Foundation of China (nos. 31171053, 11232005, 81171342, 68075049, and 10671100).

## References

- [1] H. M. Fooks, A. C. R. Martin, D. N. Woolfson, R. B. Sessions, and E. G. Hutchinson, "Amino acid pairing preferences in parallel  $\beta$ -sheets in proteins," *Journal of Molecular Biology*, vol. 356, no. 1, pp. 32–44, 2006.
- [2] M. Dorn and O. N. de Souza, "A3N: an artificial neural network n-gram-based method to approximate 3-D polypeptides structure prediction," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7497–7508, 2010.
- [3] R. E. Steward and J. M. Thornton, "Prediction of strand pairing in antiparallel and parallel  $\beta$ -sheets using information theory," *Proteins*, vol. 48, no. 2, pp. 178–191, 2002.
- [4] M. Jäger, M. Dendle, A. A. Fuller, and J. W. Kelly, "A cross-strand Trp-Trp pair stabilizes the hPin1 WW domain at the expense of function," *Protein Science*, vol. 16, no. 10, pp. 2306–2313, 2007.
- [5] M. Kuhn, J. Meiler, and D. Baker, "Strand-loop-strand motifs: prediction of hairpins and diverging turns in proteins," *Proteins*, vol. 54, no. 2, pp. 282–288, 2004.
- [6] J. S. Merkel and L. Regan, "Modulating protein folding rates in vivo and in vitro by side-chain interactions between the parallel  $\beta$  strands of green fluorescent protein," *Journal of Biological Chemistry*, vol. 275, no. 38, pp. 29200–29206, 2000.
- [7] Y. Mandel-Gutfreund, S. M. Zaremba, and L. M. Gregoret, "Contributions of residue pairing to  $\beta$ -sheet formation: conservation and covariation of amino acid residue pairs on antiparallel  $\beta$ -strands," *Journal of Molecular Biology*, vol. 305, no. 5, pp. 1145–1159, 2001.
- [8] S. M. Zaremba and L. M. Gregoret, "Context-dependence of amino acid residue pairing in antiparallel  $\beta$ -sheets," *Journal of Molecular Biology*, vol. 291, no. 2, pp. 463–479, 1999.
- [9] I. Ruczinski, C. Kooperberg, R. Bonneau, and D. Baker, "Distributions of beta sheets in proteins with application to structure prediction," *Proteins*, vol. 48, no. 1, pp. 85–97, 2002.
- [10] B. Rost, J. Liu, D. Przybylski et al., "Prediction of protein structure through evolution," in *Handbook of Chemoinformatics: From Data to Knowledge*, J. Gasteiger and T. Engel, Eds., pp. 1789–1811, John Wiley & Sons, New York, NY, USA, 2003.
- [11] J. Cheng and P. Baldi, "Three-stage prediction of protein  $\beta$ -sheets by neural networks, alignments and graph algorithms," *Bioinformatics*, vol. 21, supplement 1, pp. i75–i84, 2005.
- [12] C. K. Smith and L. Regan, "Construction and design of betasheets," *Accounts of Chemical Research*, vol. 30, no. 4, pp. 153–161, 1997.
- [13] T. Kortemme, M. Ramirez-Alvarado, and L. Serrano, "Design of a 20-amino acid, three-stranded  $\beta$ -sheet protein," *Science*, vol. 281, no. 5374, pp. 253–256, 1998.
- [14] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker, "Design of a novel globular protein fold with atomic-level accuracy," *Science*, vol. 302, no. 5649, pp. 1364–1368, 2003.
- [15] N. Zhang, J. Ruan, G. Duan, S. Gao, and T. Zhang, "The inter-strand amino acid pairs play a significant role in determining the parallel or antiparallel orientation of  $\beta$ -strands," *Biochemical and Biophysical Research Communications*, vol. 386, no. 3, pp. 537–543, 2009.
- [16] N. Zhang, G. Duan, S. Gao, J. Ruan, and T. Zhang, "Prediction of the parallel/antiparallel orientation of beta-strands using amino acid pairing preferences and support vector machines," *Journal of Theoretical Biology*, vol. 263, no. 3, pp. 360–368, 2010.
- [17] E. G. Hutchinson, R. B. Sessions, J. M. Thornton, and D. N. Woolfson, "Determinants of strand register in antiparallel  $\beta$ -sheets of proteins," *Protein Science*, vol. 7, no. 11, pp. 2287–2300, 1998.
- [18] J. S. Nowick, "Exploring  $\beta$ -sheet structure and interactions with chemical model systems," *Accounts of Chemical Research*, vol. 41, no. 10, pp. 1319–1330, 2008.
- [19] A. G. Cochran, R. T. Tong, M. A. Starovasnik et al., "A minimal peptide scaffold for  $\beta$ -turn display: optimizing a strand position in disulfide-cyclized  $\beta$ -hairpins," *Journal of the American Chemical Society*, vol. 123, no. 4, pp. 625–632, 2001.
- [20] S. J. Russell and A. G. Cochran, "Designing stable  $\beta$ -hairpins: energetic contributions from cross-strand residues," *Journal of the American Chemical Society*, vol. 122, no. 50, pp. 12600–12601, 2000.
- [21] Y. Dou, P. F. Baisnée, G. Pollastri, Y. Pécout, J. Nowick, and P. Baldi, "ICBS: a database of interactions between protein chains mediated by  $\beta$ -sheet formation," *Bioinformatics*, vol. 20, no. 16, pp. 2767–2777, 2004.
- [22] N. Zhang, J. Ruan, J. Wu, and T. Zhang, "Sheetspair: a database of amino acid pairs in protein sheet structures," *Data Science Journal*, vol. 6, no. 15, pp. S589–S595, 2007.
- [23] Q. Zhang, S. Yoon, and W. J. Welsh, "Improved method for predicting  $\beta$ -turn using support vector machine," *Bioinformatics*, vol. 21, no. 10, pp. 2370–2374, 2005.
- [24] J. Cheng and P. Baldi, "Improved residue contact prediction using support vector machines and a large feature set," *BMC Bioinformatics*, vol. 8, article 113, 2007.
- [25] P. Baldi, G. Pollastri, C. A. Andersen, and S. Brunak, "Matching protein beta-sheet partners by feedforward and recurrent neural networks," in *Proceedings of International Conference on Intelligent Systems for Molecular Biology (ISMB '00)*, vol. 8, pp. 25–36, 2000.
- [26] O. Grana, D. Baker, R. M. MacCallum et al., "CASP6 assessment of contact prediction," *Proteins*, vol. 61, no. 7, pp. 214–224, 2005.
- [27] I. Halperin, H. Wolfson, and R. Nussinov, "Correlated mutations: advances and limitations. A study on fusion proteins and on the Cohesin-Dockerin families," *Proteins*, vol. 63, no. 4, pp. 832–845, 2006.
- [28] P. J. Kundrotas and E. G. Alexov, "Predicting residue contacts using pragmatic correlated mutations method: reducing the false positives," *BMC Bioinformatics*, vol. 7, article 503, 2006.
- [29] G. Z. Zhang, D. S. Huang, and Z. H. Quan, "Combining a binary input encoding scheme with RBFNN for globulin protein inter-residue contact map prediction," *Pattern Recognition Letters*, vol. 26, no. 10, pp. 1543–1553, 2005.
- [30] J. Cheng and P. Baldi, "Improved residue contact prediction using support vector machines and a large feature set," *BMC Bioinformatics*, vol. 8, article 113, 2007.
- [31] C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, and D. Baker, "Protein structure prediction using rosetta," *Methods in Enzymology*, vol. 383, pp. 66–93, 2004.
- [32] J. Lee, S. Y. Kim, and J. Lee, "Protein structure prediction based on fragment assembly and parameter optimization," *Biophysical Chemistry*, vol. 115, no. 2-3, pp. 209–214, 2005.
- [33] G. Wang and R. L. Dunbrack, "PISCES: a protein sequence culling server," *Bioinformatics*, vol. 19, no. 12, pp. 1589–1591, 2003.
- [34] G. Wang and R. L. Dunbrack, "PISCES: recent improvements to a PDB sequence culling server," *Nucleic Acids Research*, vol. 33, no. 2, pp. W94–W98, 2005.

- [35] F. Ferron, S. Longhi, B. Canard, and D. Karlin, "A practical overview of protein disorder prediction methods," *Proteins*, vol. 65, no. 1, pp. 1–14, 2006.
- [36] R. Linding, L. J. Jensen, F. Diella, P. Bork, T. J. Gibson, and R. B. Russell, "Protein disorder prediction: implications for structural proteomics," *Structure*, vol. 11, no. 11, pp. 1453–1459, 2003.
- [37] B. Liu, L. Lin, X. Wang, X. Wang, and Y. Shen, "Protein long disordered region prediction based on profile-level disorder propensities and position-specific scoring matrixes," in *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BIBM '09)*, pp. 66–69, November 2009.
- [38] M. Parisien and F. Major, "Ranking the factors that contribute to protein  $\beta$ -sheet folding," *Proteins*, vol. 68, no. 4, pp. 824–829, 2007.
- [39] M. S. Searle and B. Ciani, "Design of  $\beta$ -sheet systems for understanding the thermodynamics and kinetics of protein folding," *Current Opinion in Structural Biology*, vol. 14, no. 4, pp. 458–464, 2004.
- [40] K. S. Rotondi and L. M. Gierasch, "Local sequence information in cellular retinoic acid-binding protein I: specific residue roles in  $\beta$ -turns," *Biopolymers*, vol. 71, no. 6, pp. 638–651, 2003.
- [41] J. Kim, S. R. Brych, J. Lee, T. M. Logan, and M. Blaber, "Identification of a key structural element for protein folding within  $\beta$ -hairpin turns," *Journal of Molecular Biology*, vol. 328, no. 4, pp. 951–961, 2003.
- [42] J. Karanicolas and C. L. Brooks, "The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design?" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 7, pp. 3954–3959, 2003.
- [43] K. S. Rotondi, L. F. Rotondi, and L. M. Gierasch, "Native structural propensity in cellular retinoic acid-binding protein I 64–88: the role of locally encoded structure in the folding of a  $\beta$ -barrel protein," *Biophysical Chemistry*, vol. 100, no. 1-3, pp. 421–436, 2003.
- [44] Y. Kato, T. Akutsu, and H. Seki, "Dynamic programming algorithms and grammatical modeling for protein beta-sheet prediction," *Journal of Computational Biology*, vol. 16, no. 7, pp. 945–957, 2009.
- [45] B. Wathen and Z. Jia, "Protein  $\beta$ -sheet nucleation is driven by local modular formation," *Journal of Biological Chemistry*, vol. 285, no. 24, pp. 18376–18384, 2010.