Refinement on O Atom Positions for Protein Backbone Prediction

HSIAO-YEN CHANG
National Sun Yat-sen University
Dept. of Computer Sci. and Eng.
Kaohsiung, TAIWAN
changhy@par.cse.nsysu.edu.tw

CHANG-BIAU YANG*
National Sun Yat-sen University
Dept. of Computer Sci. and Eng.
Kaohsiung, TAIWAN
cbyang@cse.nsysu.edu.tw

HSING-YEN ANN
National Sun Yat-sen University
Dept. of Computer Sci. and Eng.
Kaohsiung, TAIWAN
annhy@par.cse.nsysu.edu.tw

Abstract: For given the 3D coordinates of C_α in a protein, the all-atom protein backbone reconstruction problem (PBRP) is to rebuild the 3D coordinates of all major atoms (N, C and O) on the backbone. The previous works show that the prediction accuracy of the 3D positions of the O atoms is not so good, compared with the other two atoms N and C. Thus, we aim to refine the positions of the O atoms after the initial prediction of N, C and O atoms on the backbone has been done by the previous works. Based on the AMBER force field, we modify the energy function to a simplified one with the statistical data on the bond lengths and bond angles of amino acids. We find that the position of each O atom can be calculated independently. Thus, we propose a two-phase refinement method (TPRM) to efficiently tune the position of each O atom based on the modified energy function. The experimental results show that our reconstruction results are more accurate than others. Moreover, the solutions obtained by our method are more stable than the others. Besides, our method also runs much faster than the famous prediction tool, SABBAC.

Key–Words: Bioinformatics, protein structure, backbone prediction, RMSD

1 Introduction

A protein is a linear chain of amino acids. Usually, we say that there are 20 kinds of standard amino acids. Sometimes the nonstandard amino acids may be involved as the 21st kind of amino acid. To reveal its corresponding function, the linear chain has to be folded correctly. Therefore, one of the most important things for biologists is to analyze the 3D conformations of proteins. X-ray crystallographic and *nuclear magnetic resonance* (NMR)[17] are two major ways to determine the 3D structure of a given protein. However, they are slow and costly. Therefore, the protein structure prediction becomes one of important topics in the field of structure biology.

Homology modeling [1, 8, 9, 15, 20] and *ab initio* [8, 10–12, 19] are two major strategies for protein structure prediction. Homology modeling is also called knowledge-based modeling which utilizes known structures to produce a template set and determines the appropriate templates according to the similarity of conformations or sequences. On the other hand, ab initio strategies do not require propaedeutic of protein structures, but they depend on a principle of molecular mechanics, thermodynamic and the experiential potential functions. *Force fields* are sets of parameters and equations, which are derived from both experimental work and high-level quantum mechanical calculations, that can be used in molecular mechanics simulations. Some methods or software packages of force fields are often used, such as AMBER [5], CHARMM [4, 13] and OPEP [18].

Protein Data Bank (PDB) [3] is a famous database which collects lots of solved 3D protein structures. However, some of the proteins in PDB lack for the important information, including the coordinates of the major atoms N, C and O on the backbone and the atoms on the side chains. That is, they have only α-carbon coordinates. Many researches focus on how to reconstruct (predict) other missed atoms' coordinates based on the given atom positions, e.g. α-carbon coordinates. In this paper, we shall only study the prediction of all-atom coordinates on the protein backbone.

Given the 3D coordinates of the α-carbons in protein and its corresponding amino acid sequence, the all-atom *protein backbone reconstruction problem* (PBRP) is to rebuild the 3D coordinates of all major atoms (N, C, and O) on the backbone. In general, there are two major strategies for solving PBRP, one employs the similarity of small fragments, and the other is to minimize the molecular energy. Some approaches combine the above two kinds of methods [2, 7]. There are also many methods proposed to solve the PBRP, such as MaxSprout [8], Adcock's method [1], SABBAC [14] and Wang's method [20].

SABBAC [14] is a famous online service for all-

*Corresponding author

atom PBRP. First, they used the encoding of the protein tracing in *hidden Markov model* (HMM) derived structure alphabet to generate the collection of fragments. There is a collection of candidate fragments at each position on the backbone. Totally, only 155 fragments are used to describe the 27 letters of the structure alphabet. Then, they applied a greedy algorithm to search for the optimal combination. To drive the search, they used the score function inspired from OPEP force field [18]. Finally, SABBAC gave, on average, better results than MaxSprout [8] and Adcock's method [1]. For the execution time, the typical calculation of the small protein is in the order of few seconds. For one protein of larger size, the calculation time may increase up to several minutes or tens of minutes, depending on the server load.

Another method to solve all-atom PBRP was proposed by Wang *et al.* [20], which is based on the homology modeling. First, they extract all consecutive four-residue fragments from the structures of all proteins in PDB, which are based on the six inner distances and the residue group classification, to create their fragment library. Because every fragment is identified by its second, third and fourth residues, the fragments are classified into $20^3 = 8,000$ residue groups. In each residue group, the fragments with similar structures are clustered together. One representative fragment is used to represent one cluster, and then these representative fragments form the fragment library. Afterward, they seek out possible candidates from the fragment library to reconstruct the backbone of the target protein. The most noticeable contribution of Wang's method is that it requires much less execution time than the previous works, such as SABBAC. Moreover, its reconstruction accuracy is also comparable to these previous works.

In this paper, we employ Wang's method [20] as an initial prediction method. Based on the AMBER force field [5, 6], we modify the energy function to a simplified one with the statistical data on the bond lengths and bond angles of the 21 distinct amino acids (including the nonstandard one). Then, we propose a *two-phase refinement method* (TPRM) to tune the position of each O atom independently that optimizes the modified energy function. The experimental results show that the prediction accuracy of our method is better than the previous works. Besides, like Wang's method [20], our method runs much faster than the famous prediction tool, SABBAC.

The rest of this paper is organized as follows. In Section 2, we present our method for refining the position of each O atom on the protein backbone. Next, Section 3 shows the experimental results of our method, which is more accurate than those obtained by the previous works. Finally, the conclusion of this paper is given in Section 4.

Table 1: The RMSD of the N, C and O atoms of Wang's prediction result [20].

Set of proteins	RMSD (Å)		
	N atoms	C atoms	O atoms
Set 1	0.26	0.25	0.71
Set 2	0.26	0.27	0.77

2 Refinement of O Atoms' Positions

Wang's method [20] can be used for predicting the 3D coordinates of all atoms (N, C and O) on the backbone by given a sequence of amino acids and their α-carbon coordinates. However, Wang's method does not consider the condition of potential energy. Thus, we aim to improve the reconstruction accuracy based on Wang's method by adding the measurement of potential energy.

Wang *et al.* [20] performed experiments on two test sets of proteins. The first test set consists of 32 proteins, referring to the experimental results proposed by Maupetit *et al.* [14]. The second test set consists of 66 proteins, a subset extracted from CASP7 [16] targets. We compute the RMSD of the N, C and O atoms, respectively, between the real protein backbone structures and the predicted ones of these two test sets, as shown in Table 1. We find that the RMSD of the O atoms is large, very different from those of the N and C atoms. Intuitively, we may gain some improvement if we can refine the positions of the O atoms on the backbone.

Figure 1 shows the flowchart of our refinement method. The input data of our method is the coarse grained protein chain, an amino acid sequence and its α-carbon 3D coordinates. First, we acquire the initial backbone structure by Wang's method [20]. Then, we refine the O atom positions on the backbone. Finally, the output of our method is the 3D coordinates of all atoms on the backbone.

Now, we will explain why the positions of all O atoms on the backbone can be calculated independently. Based on the AMBER force field [5], the potential energy function is defined as follows:

$$E_{total} = E_{bonds} + E_{angles} + E_{torsion} + E_{van} + E_{elec},$$

where, E_{bonds}, E_{angles}, $E_{torsion}$, E_{van} and E_{elec} denote the energy items of bond stretching, angle bending, torsion angle changes, van der Waals force and electrostatic force, respectively. If we ignore the weak interactions, van der Waals force and electrostatic

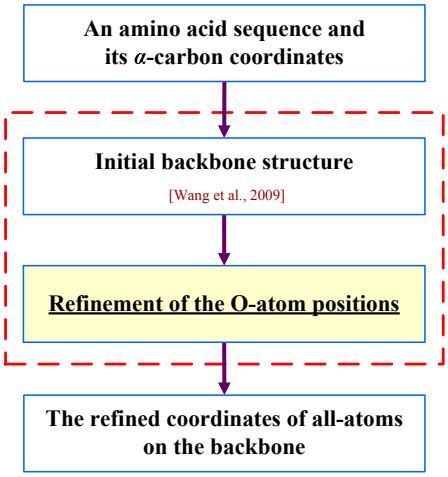

Figure 1: The flowchart of our refinement method.

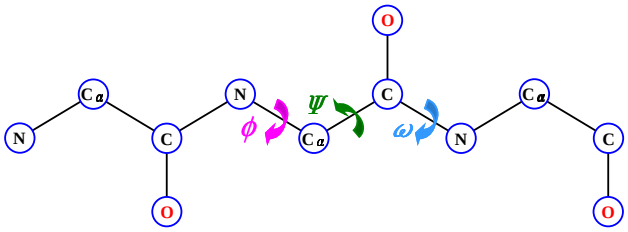

Figure 2: The illustration of three backbone dihedral angles.

force, the energy items remain bond stretching, angle bending and torsion angle changes. For torsion angle changes, there are three dihedral angles on the protein backbone, ϕ, ψ and ω, as shown in Figure 2. The backbone dihedral angles ϕ, ψ, and ω involve atoms C-N-C_α-C, N-C_α-C-N, and C_α-C-N-C_α respectively. One can see that these angles do not depend on the positions of O atoms on the backbone. Therefore, the only energy items that have to be considered are bond stretching and angle bending. Accordingly, we define the fitness function to involve only the bond stretching (C-O) and the angle bending (O-C-C_α and O-C-N). Consequently, we can calculate the positions of all O atoms on the backbone independently when coordinates of N, C and C_α atoms are given.

Based on the independency of the O atoms, we propose a *two-phase refinement method* (TPRM) to refine all O atoms on the backbone. Our idea is illustrated as Figure 3. We first define *coarse moving scope*, denoted as ρ, to be the boundary of the cube centering at the initial O position. The *coarse resolution* within the coarse moving scope, denoted as δ, is the number of grid points on each side of the cube, where each grid point represents one candidate position of the O atom. Accordingly, there are δ^3 pos-

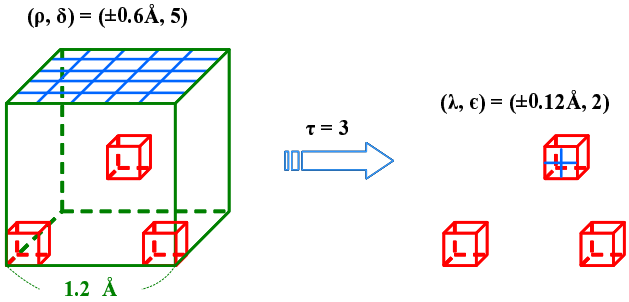

Figure 3: The illustration of our two-phase refinement method.

sible candidates. In the first phase of TPRM, within the coarse moving scope, we try all δ^3 possible candidates exhaustively. Then, we select τ candidate positions with smaller, i.e. better, energy, each as the center of one fine cube bounded by the *fine moving scope*, denoted as λ, to be further investigated. Let the *fine resolution* of the fine cube be denoted by ϵ. In the second phase, in each of the fine cube, we also try all ϵ^3 possible positions exhaustively to seek out the structure with minimum energy. In summary, for each O atom, the amount of modified energy computations (fitness function) is $\delta^3 + \tau\epsilon^3$.

In the reconstruction of all O atoms, the goodness of each candidate position is measured by the fitness function. The solution with a lower score is the better one. As mentioned above, our fitness function only considers the bonded potential energy which involves the O atom on the backbone. Thus, we define the fitness function as follows:

$$E_O = K_b(b - \overline{b})^2 + K_{\theta_1}(\theta_1 - \overline{\theta_1})^2 + K_{\theta_2}(\theta_2 - \overline{\theta_2})^2,$$

where K_b, K_{θ_1} and K_{θ_2} are the force constants for the bond stretching (C-O) and the angle bending (O-C-C_α and O-C-N), respectively; b, θ_1 and θ_2 are the input bond length (C-O) and bond angles (O-C-C_α and O-C-N), respectively; $\overline{b}$, $\overline{\theta_1}$ and $\overline{\theta_2}$ are the standard (average) values of bond length (C-O) and bond angles (O-C-C_α and O-C-N) extracted from PDB, respectively. After calculation, the force constants and standard values of the fitness function are obtained.

3 Experimental Results

The experiments are performed on a PC with AMD Athlon^{TM} 64 X2 dual core processor 3800+ 2.01 GHz and 1.46 GB RAM. The operating system is Microsoft Windows XP Professional Version 2002 Service Pack 2.

Our method has been tested on 32 proteins, referring to the experimental results proposed by Maupetit

et al. [14]. Those are also referred to the experimental results of Wang's method [20]. The refined results are compared with the real crystallographic structures obtained from PDB, and then the RMSD values are calculated. If the RMSD value is smaller, the predicted protein backbone structure is considered as more similar to the real one.

We compare the refined results of our method with some previous works, including MaxSprout [8], Adcock's method [1], SABBAC [14] and Wang's method [20]. For SABBAC, it is an online service devoted to protein backbone reconstruction with α-carbon. In 2008, only a new implementation, SABBAC v1.2, is presented on their online service to rebuild the protein backbone structure. And their first version, SABBAC v1 [14], is no longer available. The results of SABBAC v1.2 are better and faster than SABBAC v1, but these results are not yet published officially in any journal or conference. Here, we compare the results of our method with both SABBAC v1 and SABBAC v1.2.

Table 2 shows the experimental results of our method and the previous works. The solutions of SABBAC v1 are referred to the paper [14], and the solutions of SABBAC v1.2 are generated by the online service. In Table 2, "♭ ", "♮ ", "♯ ", "† " and "§ " denote the previous methods MaxSprout, Adcock's method, SABBAC v1, SABBAC v1.2 and Wang's method, respectively. The RMSD value of the test protein of our method is marked with the corresponding sign, if it is better than the corresponding previous one. Table 2 shows that the prediction results of our method are more accurate and more stable than all previous results.

We summarize our results according to Tables 2. First, because our fitness function is suitable for measuring biological energy, one can see that our method greatly improves Wang's result. Second, we want to address that the accuracy of our method would depend on the initial structure, such as Wang's result. Thus, some of our solutions are worse than SABBAC, especially when the reconstructed structure of Wang's method is less accurate than that of SABBAC by more than 0.5 $\mathring{A}$ RMSD. As one can see, the overall accuracy of our results is superior to SABBAC.

Tables 3 shows the execution time of SABBAC v1.2, Wang's method and our method. According to the publishers of SABBAC v1.2, their machine is simply a dual core processor PC, similar to ours. Therefore, the required time of our prediction and SABBAC would reflect the efficiency of algorithms, rather than the performance of machines. As one can see that the execution time required for our method is much less than that for SABBAC v1.2. Note that, our execution includes the execution time of Wang's method. As

Table 2: Comparison of our method with the previous works. "μ" and "σ" denote the average and the standard deviation of the RMSD values of all test proteins. Let *L*, *MS*, *A*, *W* and *Our* denote the length, MaxSprout, Adcock's method, Wang's method and our method, respectively.

Protein		Main chain RMSD (Å)					
		Prior works					
				SABBAC			
PDB ID	*L*	*MS*$^{\flat}$	*A*$^{\natural}$	v1$^{\sharp}$	v1.2†	*W*§	*Our*
111M	154	0.42	0.31	0.29	0.21	0.26	0.25$^{\flat\natural\sharp\S}$
1CTF	68	0.73	0.41	0.43	0.33	0.42	0.38$^{\flat\natural\sharp\S}$
1IGD	61	0.44	0.34	0.36	0.37	0.36	0.35$^{\flat\sharp\dagger\S}$
1OMD	107	0.41	0.39	0.35	0.40	0.39	0.34$^{\flat\natural\sharp\dagger\S}$
1SEMA	58	0.34	0.5	0.48	0.48	0.45	0.40$^{\sharp\dagger\S}$
1TIMA	247	0.60	0.56	0.59	0.58	0.54	0.52$^{\flat\natural\dagger\S}$
1UBQ	76	0.38	0.37	0.35	0.34	0.37	0.36$^{\flat\S}$
2CTS	437	0.45	0.37	0.40	0.38	0.34	0.31$^{\flat\natural\dagger\S}$
2LYM	129	0.44	0.32	0.38	0.38	0.29	0.28$^{\flat\natural\dagger\S}$
2MHR	118	0.54	0.33	0.5	0.44	0.39	0.35$^{\flat\sharp\S}$
2PCY	99	0.54	0.48	0.42	0.44	0.33	0.30$^{\flat\natural\dagger\S}$
2OZ9	104	0.42	0.24	0.3	0.29	0.22	0.19$^{\flat\natural\dagger\S}$
4PTI	58	0.44	0.51	0.53	0.37	0.42	0.42$^{\flat\natural}$
5CPA	307	-	0.48	0.41	0.41	0.34	0.32$^{\sharp\dagger\S}$
5NLL	138	0.46	0.42	0.37	0.38	0.39	0.34$^{\flat\natural\dagger\S}$
PDB newcomers subset							
1PXZA	346	0.54	-	0.55	0.48	0.53	0.50$^{\flat\sharp\S}$
1RKIA	101	0.44	-	0.58	0.56	0.50	0.46$^{\sharp\dagger\S}$
1S7LA	177	0.36	-	0.29	0.34	0.38	0.33$^{\flat\dagger\S}$
1T70A	255	0.50	-	0.42	0.44	0.48	0.44$^{\flat\S}$
1TXOA	235	0.38	-	0.41	0.39	0.44	0.40$^{\sharp\S}$
1V0ED	666	0.45	-	0.48	0.51	0.40	0.36$^{\flat\sharp\dagger\S}$
1V7BA	175	0.41	-	0.3	0.25	0.37	0.34$^{\flat\S}$
1VB5B	275	0.42	-	0.34	0.30	0.41	0.38$^{\flat\S}$
1VKCA	149	0.33	-	0.28	0.34	0.37	0.34§
1VR4A	103	0.59	-	0.47	0.46	0.47	0.45$^{\flat\sharp\dagger\S}$
1VR9A	121	0.45	-	0.42	0.45	0.49	0.45§
1WMHA	83	0.28	-	0.27	0.27	0.38	0.36§
1WMIA	88	0.42	-	0.41	0.42	0.50	0.46§
1WPBG	168	0.35	-	0.37	0.34	0.43	0.39§
1X6JA	88	0.36	-	0.43	0.40	0.49	0.46§
1XB9A	108	0.51	-	0.46	0.48	0.53	0.50$^{\flat\S}$
1XE0B	107	0.62	-	0.61	0.58	0.55	0.54$^{\flat\sharp\dagger\S}$
μ		0.45	0.4	0.41	0.40	0.41	0.38
σ		0.09	0.09	0.09	0.09	0.08	0.08

Table 3: The execution time (seconds) of SABBAC v1.2, Wang's method and our method.

PDB ID	SABBAC v1.2	Wang's method	Our method
111M	51	4	10
1CTF	39	2	5
1IGD	38	1	3
1OMD	38	3	7
1SEMA	38	1	3
1TIMA	131	7	18
1UBQ	38	2	5
2CTS	488	13	32
2LYM	50	3	8
2MHR	39	3	8
2PCY	38	3	7
2OZ9	38	3	7
4PTI	37	1	3
5CPA	257	9	22
5NLL	50	4	10
1PXZA	374	10	26
1RKIA	39	2	6
1S7LA	76	4	11
1T70A	180	8	19
1TXOA	110	8	18
1V0ED	2327	19	49
1V7BA	63	5	13
1VB5B	192	8	20
1VKCA	50	4	10
1VR4A	40	3	7
1VR9A	50	3	8
1WMHA	38	2	5
1WMIA	38	2	5
1WPBG	83	4	11
1X6JA	37	2	5
1XB9A	53	3	7
1XE0B	54	3	7
Total	5174	149	375

shown in Tables 3, the additional time spent by our method after Wang's method is not too much.

4 Conclusion

In this paper, we propose a two-phase refinement method to establish the 3D coordinates of all O atoms on the protein backbone. First, we try all possible grid positions in a coarse cube exhaustively. Then, we select some candidate positions with smaller energy, each as the center of one fine cube to be further considered. In each of the fine cubes, we also try to test all possible grid positions exhaustively to seek out the structure with minimum energy.

Furthermore, we modify the energy function to a simplified one with the statistical data on the bond lengths and bond angles of the 21 distinct amino acids (including the nonstandard one), to improve the reconstruction accuracy. The experimental results show that the reconstruction accuracy of our method is better than most of the previous works, such as MaxSprout and Adcock's method. In addition, we not only improve Wang's results but also outperform SABBAC in most of the proteins. The solutions of our method gain the same standard deviations of Wang's method and are more stable than the remaining previous works. Besides, our method runs much faster than SABBAC v1.2.

On the other hand, our method cannot deal with all kinds of protein structures, such as nonstandard residues. The future works include the modification of the force constant of the energy function, and the improvement of the prediction accuracy. Besides, the refinement of all N and C atoms on protein backbone might be achieved with other evolutionary algorithms based on the AMBER force field.

References:

[1] S. A. Adcock, "Peptide backbone reconstruction using dead-end elimination and a knowledge-based forcefield," *Journal of Computational Chemistry*, Vol. 25, pp. 16–27, 2004.

[2] J. Arunachalam, V. Kanagasabai, and N. Gautham, "Protein structure prediction using mutually orthogonal latin squares and a genetic algorithm," *Biochemical and Biophysical Research Communications*, Vol. 342, pp. 424–433, 2006.

[3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissing, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Research*, Vol. 28, pp. 235–242, 2000.

[4] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," *Journal of Computational Chemistry*, Vol. 4, pp. 187–217, 1983.

[5] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules," *Journal of American Chemical Society*, Vol. 117, pp. 5179–5197, 1995.

[6] Y. Duan, C. WU, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based

on condensed-phase quantum mechanical calculations," *Journal of Computational Chemistry*, Vol. 24, No. 16, pp. 85–97, 2003.

[7] M. Feig, P. Rotkiewicz, A. Kolinski, J. Skolnick, and C. L. B. III, "Accurate reconstruction of all-atom protein representations from side-chain-based low-resolution models," *Proteins: Structure, Function, and Genetics*, Vol. 41, pp. 86–97, 2000.

[8] L. Holm and C. Sander, "Database algorithm for generating protein backbone and side-chain coordinates from a c alpha trace application to model building and detection of coordinate errors," *Journal of Molecular Biology*, Vol. 21, No. 1, pp. 183–194, 1991.

[9] J. L. Hsin, C. B. Yang, K. S. Huang, and C. N. Yang, "An ant colony optimization approach for the protein side chain packing problem," *Proceedings of the 6th WSEAS International Conference on Microelectronics, Nanoelectronics, Optoelectronics*, Istanbul, Turkey, pp. 44–49, 2007.

[10] R. Kazmierkiewicz, A. Liwo, and H. A. Scheraga, "Energy-based reconstruction of a protein backbone from its α-carbon trace by a Monte-Carlo method," *Journal of Computational Chemistry*, Vol. 23, pp. 715–723, 2002.

[11] C. Lee and S. Subbiah, "Prediction of protein side-chain conformation by packing optimization," *Journal of Molecular Biology*, Vol. 217, No. 2, pp. 373–388, 1991.

[12] S. Liang and N. V. Grishin, "Side-chain modeling with an optimized scoring function," *Protein Science*, Vol. 11, No. 2, pp. 322–331, 2002.

[13] A. D. MacKerell, B. Brooks, C. L. B. III, L. Nilsson, B.Roux, Y. Won, and M. Karplus, "Charmm: The energy function and its parameterization with an overview of the program," *The Encyclopedia of Computational Chemistry*, Vol. 1, pp. 271–277, 1998.

[14] J. Maupetit, R. Gautier, and P. Tuffery, "SABBAC: online structural alphabet-based protein backbone reconstruction from alpha-carbon trace," *Nucleic Acids Research*, Vol. 34, pp. W147–W151, 2006.

[15] M. Milik, A. Kolinski, and J. Skolnick, "Algorithm for rapid reconstruction of protein backbone from alpha carbon coordinates," *Journal of Computational Chemistry*, Vol. 18, pp. 80–85, 1997.

[16] M. Milostan, A. Kryshtafovych, K. Fidelis, P. Daniluk, V. Nakonechnyy, and Z. Dmytriv, "Critical Assessment of Techniques for Protein Structure Prediction." http://predictioncenter.org/, Genome Center, University of California, Davis.

[17] I. Ruczinski, C. Kooperberg, R. Bonneau, and D. Baker, "Distributions of beta sheets in proteins with application to structure prediction," *Proteins: Structure, Function, and Genetics*, Vol. 48, No. 1, pp. 85–97, 2002.

[18] S. Santini, G. Wei, N. Mousseau, and P. Derreumaux, "Exploring the folding pathways of proteins through energy landscape sampling: Application to alzheimer's β-amyloid peptide," *Internet Electronic Journal of Molecular Design*, Vol. 2, No. 9, pp. 564–577, 2003.

[19] A. A. Tantar, N. Melab, and E. G. Talbi, "A comparative study of parallel metaheuristics for protein structure prediction on the computational grid," *International Parallel and Distributed Processing Symposium (IPDPS 2007)*, Long Beach, California, USA, pp. 1–10, 2007.

[20] J. H. Wang, C. B. Yang, and C. T. Tseng, "Reconstruction of protein backbone with the α-carbon coordinates," *Accepted by Journal of Information Science and Engineering*, 2009.