Protein Loop Modeling Using Fragment Assembly

Dong-Seon LEE and Chaok SEOK*

Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-747

Julian LEE†

Department of Bioinformatics and Life Science, Soongsil University, Seoul 156-743

We model loop regions of proteins based on a fragment assembly method. The fragments that comprise candidates of the local structure of a protein loop are collected from a structure database, and all loop conformations possible from a smooth assembly of the fragments are generated. For each of the fragment-assembled conformations, a Monte Carlo simulation in the conformational subspace that satisfies the loop closure constraint is performed to minimize the root-mean-square deviation of the backbone dihedral angles from the fragment angles. The side-chains are then built using a rotamer library, and the backbone and the side-chain conformations are optimized locally with the AMBER 96 force field without solvation terms to remove steric clashes. The resulting conformations are then ranked using the DFIRE potential. A test prediction for eight protein loops with sizes ranging from 8 to 12 residues is presented to show the feasibility of our method. Tests with further optimization using Monte Carlo with minimization show that extensive conformational optimization leading to deviation from the original fragment-assembled structures tend to deteriorate the prediction accuracy, suggesting that the utilization of fragment information is superior to purely energy-based methods.

I. INTRODUCTION

Prediction of the native structure of a protein is one of the utmost goals in theoretical biophysics because information on the native structure of a protein is crucial in understanding and regulating its biological function [1]. When homologous sequences with known structures are available, comparative modeling can be used to predict the native structure for the query sequence by using the structures of homologous sequences in the structure database as templates [2,3]. However, there are gap regions in the query sequence that cannot be aligned with the sequences of the template structures. These are regions of insertions of amino acids that occurred during the evolution process. They usually form no definite secondary structures, and are called loops. Protein loops are often involved in functionally important regions, such as molecular recognition sites, and contribute to functional specificity of homologous proteins.

A protein loop has to be modeled such that it is geometrically consistent with the rest of the protein struc-

ture. Therefore, there are constraints that must be satisfied by the dihedral angles of the loop: the atoms at the two ends of the loop must be connected with the atoms in the two stem regions with adequate bond lengths and bond angles [4]. Given an accurate free energy function, the problem is reduced to finding the set of dihedral angles that minimizes the free energy of the whole protein structure and that satisfies the constraint of loop closure simultaneously.

In a fragment assembly method, fragments of local structures are collected from a structural database, and the full structure is obtained by assembling these fragments. This method can alleviate the burden of a heavy conformational search, which becomes evident in an exhaustive conformational search of loops longer than 8 residues [5–7].

Fragment assembly methods have been used for modeling the global structures of proteins with no apparent sequence similarities to those in the structure database [3,8–15], as well as structurally variable regions when homologous templates exist in the database [16]. However, when conformations of a loop are generated from fragment assembly, they violate the loop closure constraint in general because it is very unlikely that the fragment structures drawn from the database fit into the frame of

---

*E-mail: chaok@snu.ac.kr; Fax: +82-2-871-8119;
†E-mail: jul@ssu.ac.kr; Fax: +82-2-824-4383

the protein of interest exactly. In Ref. 16, conformations were allowed to violate the loop closure constraint initially, and a gap penalty for chain discontinuity in the energy function was used to close the loops.

In contrast, in this work, we impose the loop closure constraint at the initial stage of geometric construction by finding a set of dihedral angles satisfying the constraint that is as close to the fragment-assembled conformation as possible. Our method is distinct from the fragment assembly method of Ref. 16 in that the loop closure problem is dealt with in a purely geometric way without relying on minimization of an energy function. This purely geometric method is more efficient because it does not require expensive energy minimization and is simpler because relative strengths of different energy terms need not be considered.

Conformations generated from fragment assembly and loop closure are optimized with the AMBER 96 potential [17] without solvation energy to relieve steric clashes and are scored with the knowledge-based DFIRE potential [18] as an effective free energy function. The DFIRE potential has been shown to be as successful in scoring loop decoy conformations as force fields such as AMBER or OPLS with a generalized Born solvation free energy [19]. DFIRE has the advantage of being more efficient in terms of computation time. We perform test predictions on eight protein loops with sizes 8 to 12 residues. We find that if too much optimization using the DFIRE energy function is performed and thus the conformations deviate far away from the original fragment-assembled conformation, the result deteriorates, suggesting that the utilization of fragment information is superior to purely energy-based methods.

II. METHODS

The protein loop modeling method presented here takes the steps of fragment assembly to generate the candidate loop structures, adjustment of the structures to satisfy the loop closure constraint, and optimization and scoring of the loop conformations. The flowchart of these steps is shown in Figure 1, and each step is discussed below in detail.

1. Fragment Assembly

We consider windows of size 15, and for each window we collect a set of kNN fragments from a structure database, corresponding to the most probable conformations of the sequence lying inside the window. We used $kNN = 25$ in this work. The structural database of non-redundant proteins is constructed by clustering an ASTRAL SCOP (version 1.63) set [20] so that no two proteins in the database have more than 25 % sequence identity with each other [21,22]. The fragment selection
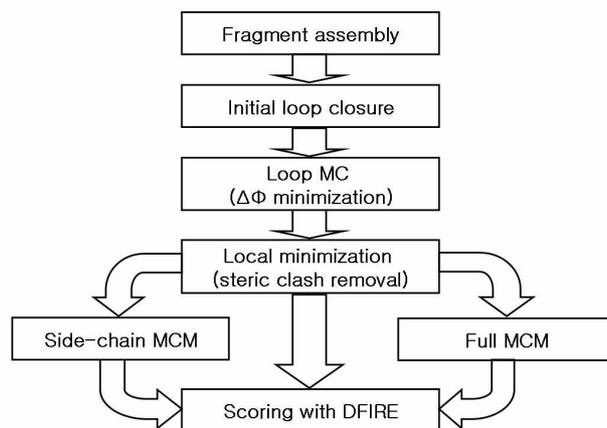

Fig. 1. Flowchart of the protein loop modeling method by fragment assembly.

is based on similarity of sequence features. Instead of comparing raw sequences directly, PSI-BLAST [23] profiles that contain evolutionary information are generated. For a given segment of the query sequence, 25 fragments with similar sequence profiles are selected from the structural database by using the k-nearest neighbor method [21,22,24,25]. The details can be found in Ref. 14.

We assemble fragments to construct loop conformations. The fragments are joined only when they overlap and share at least one residue with similar values of backbone dihedral angles. Two sets of dihedral angles (ϕ_1, ψ_1) and (ϕ_2, ψ_2) in each of the two fragments are considered to be similar to each other if

$$|\phi_1 - \phi_2| + |\psi_1 - \psi_2| \leq 5°. \qquad (1)$$

If we find such a residue pair in the two fragments, then the second fragment is joined smoothly to the first one starting from this residue [14]. Since the joining occurs in the middle of the fragments, only a part of the 15-residue-long fragment is used for the fragment assembly. The constraint of smooth-joining keeps conformational space within a manageable size. We generate all possible conformations by this method, whose number is denoted by N_{conf}, as displayed in Table 1 for the test set of eight protein loops.

2. Generation of Closed-loop Structures

Conformations for a protein loop generated from fragment assembly violate the loop closure constraint in general, i.e., the loop structures are not connected with the stem regions of the protein with the correct geometry. Therefore, the backbone dihedral angles of the loop must be modified so that the loop structures correctly fit into the protein. This is performed by randomly selecting three residues and rotating the six backbone dihedral angles (three ϕ and three ψ angles) of the residues by

the solution values of the analytic loop closure equation [4]. This step is the 'Initial loop closure' part in Figure 1. We then perform a Monte Carlo search in the conformational subspace that consists of closed loops to find a set of dihedral angles as close as possible to the original angles in the fragments, which is referred to as 'Loop MC' (Figure 1).

To elaborate on the Monte Carlo method, a backbone dihedral angle, called a driver angle, is rotated randomly within 10 degrees; then, six other dihedral angles of randomly chosen three residues are adjusted to compensate for the loop opening introduced by the driver angle rotation. The new conformation is accepted or rejected according to the Metropolis criterion, where the objective function is defined as the root-mean-square deviation of the backbone dihedral angles from the fragment angles as follows:

$$\Delta\phi = \sum_{i=1}^{N} \sqrt{\frac{(\phi_i - \phi_{0i})^2 + (\psi_i - \psi_{0i})^2}{2N}}, \qquad (2)$$

where N is the sequence length of the loop, and ϕ_{0i} and ψ_{0i} are the dihedral angles of the ith residue in the original fragment-assembled conformation. The temperature parameter, $k_B T$, is kept at a constant value of 0.5° so that local barriers can be overcome when finding a low-lying local minimum of the objective function. For each conformation, 2000 Monte Carlo steps are taken.

Since the resulting conformations depend on the initial choice of the three residues that are used to solve the loop closure equation, we perform 20 independent Monte Carlo simulations with different selections of the three residues for the initial loop closure. For each of the fragment-assembled loop conformation, the set of backbone dihedral angles that result in the minimum $\Delta\phi$ is selected to generate the initial backbone structure for further optimization and scoring.

Since the fragments are collected from proteins whose sequences are different from that of the query, only backbone dihedral angles are collected from the fragments. Side-chain dihedral angles are selected from the backbone-dependent rotamer library [26] with probabilities proportional to the rotamer probabilities.

3. Optimization and Scoring of the Loop Structures

The initial loop structures generated as explained above are optimized in three different ways and are scored with the knowledge-based potential, DFIRE [18], which is derived from the distribution of inter-atomic distances found in the structural database. DFIRE was shown to produce as a good prediction in scoring loop conformations as the AMBER or OPLS force field with a generalized Born solvation free energy [19].

The initial loop structures have steric clashes with the rest of the protein structure because the only structural constraint imposed on them is the loop closure condition. Therefore, local energy minimizations are carried out to relieve the clashes before scoring. This is the step 'Local minimization' in Figure 1. The DFIRE potential is a discrete function of atomic distances, so the AMBER 96 force field [17] without any solvation terms was employed for local optimization. The DFIRE potential was then used to score the resulting loop conformations.

We also tried two different Monte Carlo with Minimization (MCM) methods to further optimize the loop structures, in addition to the local minimization ('Side-chain MCM' and 'Full MCM' in Figure 1). If $N_{conf} > 10$, we selected 10 best-scoring locally minimized conformations for the MCM optimization. The difference between the two methods is whether only the side-chain conformations are perturbed (side-chain MCM) or both the side-chain and the backbone conformations are perturbed (full MCM) in the Monte Carlo move step. In the side-chain MCM, the side-chains are perturbed by drawing the conformations from the backbone-dependent rotamer library [26] with probabilities proportional to the rotamer probabilities. In the full MCM, the side-chains are perturbed in the same way with a probability of 20 % at each MCM step, and the backbone is perturbed at each MCM step by rotating at most two driver angles randomly within 30 degrees and maintaining the closed state of the loop by compensating rotations of the six dihedral angles of three randomly chosen residues. Both the backbone and the side-chain degrees of freedom are changed in the subsequent minimization with the AMBER potential without solvation terms. The Metropolis criterion was then applied with the DFIRE potential as the objective function. The temperature was kept at a constant value, $k_B T = 5$ kcal/mol, and 200 MCM steps were tried both for side-chain MCM and full MCM. The conformation with the minimum DFIRE potential was taken as the final model conformation.

III. RESULTS

Loop modeling tests were performed for the eight protein loops listed in Table 1, with sizes ranging from 8 to 12 residues. The loop structures were reconstructed after deleting the loop regions from the crystal structures of the whole proteins. The loop conformations generated by fragment assembly and loop closure are compared with the native crystal loop conformations in terms of RMSD (Root-Mean-Square Deviation) in the Cartesian coordinates of the backbone heavy atoms. We expect that the results will not change qualitatively if a different definition of RMSD, such as RMSD in C_α coordinates only is used because RMSDs in different sets of atoms are roughly correlated with each other [27]. The loop conformations are not superimposed for optimal rotation and translation but are compared in the fixed frame of the protein structure instead.

Table 1. The RMSD values of the fragment-assembled loop conformations before and after loop closure.

protein	length	residues	N_{conf}	initial			loop MC		
				R_{ave} [a]	R_{min} [b]	$\Delta\phi$ [c]	R_{ave} [a]	R_{min} [b]	$\Delta\phi$ [c]
135l	8	84 − 91	88	8.5	2.7	34.3	2.5	1.0	17.8
154l	12	153 − 164	28	10.6	1.0	32.2	7.2	0.5	16.6
1cyo	12	12 − 23	9	4.4	1.2	23.2	3.1	0.9	12.9
1eco	12	35 − 46	157	9.6	2.4	28.0	3.8	1.3	15.7
1rcf	12	88 − 99	13	10.8	2.8	25.4	3.9	1.2	14.0
2pia	11	74 − 84	13	7.7	0.7	20.7	2.8	0.3	12.0
5pti	10	23 − 32	20	3.5	0.6	17.2	2.3	0.3	10.6
6taa	11	150 − 160	55	13.6	1.3	32.5	5.8	0.7	17.0

[a] Average RMSD (in Å) over N_{conf} conformations.
[b] Minimum RMSD (in Å) among N_{conf} conformations.
[c] The average deviation (in degrees) over N_{conf} conformations.

1. Generation of Loop Structures from Fragment Assembly and Loop Closure

The average and minimum RMSD values of the fragment-assembled loop conformations before and after loop closure are summarized in Table 1. The number of fragment-assembled loop conformations, N_{conf}, for each protein loop varies widely, from 9 to 157. The number depends on the loop sequence rather than on the loop length for this set of loops. The RMSD value averaged over N_{conf} conformations for each protein loop ranges from 3.5 to 13.6 Å for the 8-loop test set. Note that the fragment-assembled loops do not form a closed loop geometry when attached to the original protein. Therefore, the loop RMSD is calculated after constructing the geometry by using the fragment angles starting from the N-terminal end of the loop and leaving the C-terminal end of the loop open. The minimum RMSD among the N_{conf} conformations ranges from 0.6 to 2.8 Å, indicating that native-like loop conformations are generated by fragment assembly. Therefore, it would be worthwhile to try to keep the dihedral angles as close to the fragment angles as possible during loop closure. The loop closure equation [4] is used to close the loop conformations initially and to minimize the deviation from the fragment angles. The initial closure can result in an arbitrary change of the six dihedral angles because only the loop closure constraint is satisfied. Then, the deviation from the fragment angles, $\Delta\phi$ in Eq. (2), is minimized with a Monte Carlo procedure as described in Methods. The average deviation $\Delta\phi$ of the dihedral angles is between 17 and 34 degrees when the fragment-assembled structures are initially closed, but the deviation decreases by about half on average (11 − 18 degrees) after minimizing $\Delta\phi$. The loop conformations also become closer to native, as can be seen from the decreased RMSD values in Table 1: the average RMSD over the generated conformations is between 2.3 and 7.2 Å, and the minimum RMSD is between 0.3 and 1.3 Å.

2. Optimization and Scoring of the Loop Structures

The results of optimization and scoring of the loop conformations generated by fragment assembly and loop closure are summarized in Table 2. The prediction accuracy is reasonably good when the DFIRE potential is used for scoring after local minimization or side-chain MCM: the RMSD values of the best-scoring conformations range from 1.6 to 3.2 Å for local optimization and from 1.6 to 2.8 Å for side-chain MCM, except for the case of 154l. The side-chain MCM produces slightly better predictions than local minimization alone. More optimization results in a better DFIRE score: the DFIRE score improves as the optimization method goes from local minimization to side-chain MCM to full MCM. However, the RMSD values of the best-scoring conformation are larger for those generated by the full MCM. Note that the DFIRE score is relative to that of the native loop conformation, and a positive (negative) score means a worse (better) score than the native loop. It can be seen that the full MCM with both side-chain and loop perturbation is very effective in finding conformations with better DFIRE scores, and sometimes better than the native. However, this test illustrates a limitation of the DFIRE score. Extensive searches exploring the energy landscape of DFIRE result in worse predictions, and this means that the global shape of the DFIRE potential does not accurately represent the nature's free energy surface.

As representative examples, modeled structures for the loops of 135l and 2pia are illustrated in Figure 2 and Figure 3, respectively. The structures of the 8-residue loop of 135l in Figure 2 predicted with the three optimization methods are similar to each other and deviate from the native structure to similar degrees (1.5 − 2.0 Å). However, in the case of the 11-residue loop of 2pia shown in Figure 3, the loop structure obtained from the full MCM optimization deviates much more from the native structure (5.6 Å) than the other two structures obtained

Table 2. The results of the loop reconstruction test for eight protein loops. The RMSD values of the conformation with the best score is displayed for each protein for three different optimization methods: local minimization without additional MCM (local Min), MCM with side-chain perturbation (SC MCM), and full MCM.

protein	length	residues	local Min		SC MCM		full MCM	
			RMSD [a]	score [b]	RMSD [a]	score [b]	RMSD [a]	score [b]
135l	8	84 − 91	2.0	49	1.8	36	1.5	4
154l	12	153 − 164	4.5	281	4.5	161	4.4	51
1cyo	12	12 − 23	2.8	413	2.1	244	3.3	185
1eco	12	35 − 46	2.2	188	2.8	26	3.7	−146
1rcf	12	88 − 99	3.2	264	2.7	164	3.3	−1
2pia	11	74 − 84	1.8	299	2.0	9	5.6	−91
5pti	10	23 − 32	1.6	212	1.6	116	2.2	98
6taa	11	150 − 160	2.9	130	2.7	31	2.3	−7

[a]RMSD (in Å) of the best scoring conformation.
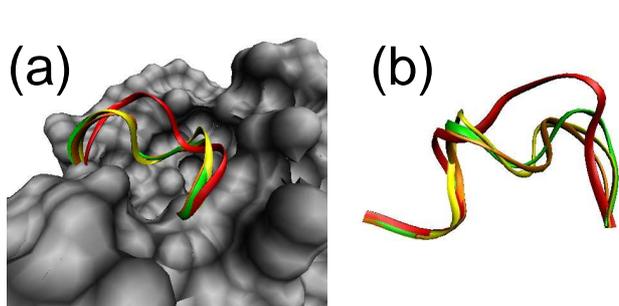[b]DFIRE score of the best scoring conformation in kcal/mol.

Fig. 2. The structures of the 8-residue loop of 135l modeled using the three different optimization methods compared with the native structure. The native loop is colored in red, and those obtained from local minimization, side-chain MCM, and loop MCM in yellow, orange, and green, respectively. The loop structures are drawn with and without the rest of the protein in (a) and (b), respectively.
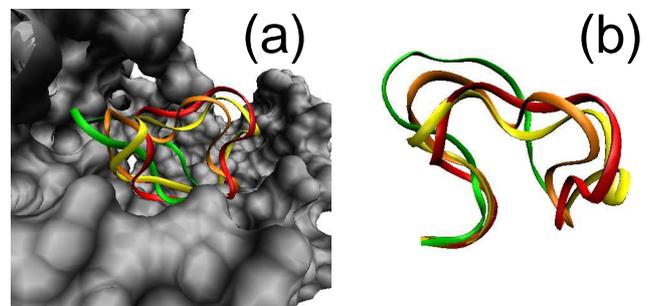

Fig. 3. The structures of the 11-residue loop of 2pia modeled using the three different optimization methods compared with the native structure. The native loop is colored in red, and those obtained from local minimization, side-chain MCM, and loop MCM in yellow, orange, and green, respectively. The loop structures are drawn with and without the rest of the protein in (a) and (b), respectively.

from local minimization and side-chain MCM (5.7 and 4.4 Å from the full MCM structure) which are 1.8 and 2.0 Å away from native. It is notable that the structure from the full MCM is more compactly packed onto the rest of the protein than the native structure, as can be seen from Figure 3 (a). This result may be a consequence of the fact that the DFIRE potential does not have an accurate balance of the packing interaction and hydration, considering the fact that the DFIRE score depends on the separation of atoms only and not on the environment of the atoms, for example, whether the atoms are on the surface of the protein structure or in the core. In reality, atoms on the surface interact with solvent molecules in native condition, but such effects are averaged out, together with the core interactions, in the current parametrization of the DFIRE potential.

In conclusion, a moderate optimization of the MCM with side-chain perturbation that does not much perturb the backbone conformation generated from fragment assembly gave the best result in this study.

IV. DISCUSSIONS

Fragment assembly methods have been widely used for predicting protein structures when templates of homologous proteins do not exist. They reduce the conformational space of the local backbone structures to those found in the database of similar sequence features. Since the conformational space of loop structures is smaller than that of the whole protein, we employed an exhaustive enumeration method [14] instead of a stochastic sampling method [16] for conformational search. The fragment assembly method used in this study was very effective in finding native-like loop conformations for loops of 8 to 12 residues for the eight different loops in the test set.

The analytic loop closure method was applied to close the loop conformations generated from fragment assembly with minimum perturbation to the fragment angles. The method is based purely on geometry and does not rely on minimization of a penalty function for unclosed loops and, thus, does not depend on arbitrary penalty function parameters. The average root-mean-square deviation of the dihedral angles from those of the fragment angles could be reduced almost to half the initial deviation.

The knowledge-based DFIRE potential was able to select the loop conformations within 3 Å in seven out of the eight loops in the test set when an optimization was performed only on side-chains. However, an extensive search for better-scoring loop conformations optimizing both side-chains and backbones results in worse predictions as the loop conformations deviated away from the fragment structures, demonstrating a limitation of the DFIRE score in describing the free energy of protein conformations.

The current method of combined use of fragment assembly and analytic loop closure is expected to be feasible in predicting conformations of longer protein loops, too. However, an improvement in the scoring function is necessary for more accurate prediction. Better accounts of solvation effects or conformational entropy [28] could be explored. A more extensive conformational search, use of a larger number of fragments, and refinement of the resulting conformations are also possible given the efficient implementation of the current conformational search algorithm.

ACKNOWLEDGMENTS

REFERENCES

[1] X. Shi, H. Xie, S. Zhang and B. Hao, J. Korean Phys. Soc. **50**, 118 (2007); H. Y. Kim, H. Y. Kang, J. W. Ryu, C. N. Yoon and S. K. Han, J. Korean Phys. Soc. **50**, 290 (2007); P. Holme, J. Korean Phys. Soc. **50**, 300 (2007); J. W. Ryu, H. Y. Kim, T. H. Kang, J. S. Yoo and J. S. Chung, J. Korean Phys. Soc. **51**, 1805 (2007).

[2] J. Moult, K. Fidelis, B. Rost, T. Hubbard and A. Tramontano, Proteins **61 (S7)**, 3 (2005).

[3] D. Baker and A. Sali, Science **294**, 93 (2001).

[4] E. A. Coutsias, C. Seok, M. P. Jacobson and K. Dill, J. Comput. Chem. **25**, 510 (2004); E. A. Coutsias, C. Seok, M. J. Wester and K. A. Dill, Int. J. Quantum Chem. **106**, 176 (2006).

[5] R. E. Bruccoleri and M. Karplus, Biopolymers **26**, 137 (1987).

[6] P. I. W. de Bakker, M. A. DePristo, D. F. Burke and T. L. Blundell, Proteins **51**, 21 (2002); M. A. DePristo, P. I. W. de Bakker, S. C. Lovell and T. L. Blundell, Proteins **51**, 41 (2002).

[7] M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw and R. A. Friesner, Proteins **55**, 351 (2004).

[8] A. M. Lest, L. Lo Conte and T. J. P. Hubbard, Proteins **45 (S5)**, 98 (2001); P. Aloy, A. Stark, C. Hadley and R. B. Russell, Proteins **53 (S6)**, 436 (2003); J. J. Vincent, C. H. Tai, B. K. Sathyanarayana and B. Lee, Proteins **61 (S7)**, 67 (2005).

[9] D. T. Jones, Proteins **45 (S5)**, 127 (2001); D. T. Jones, K. Bryson, A. Coleman, L. J. McGuffin, M. I. Sadowski, J. S. Sodhi and J. J. Ward, Proteins **61 (S7)**, 143 (2005).

[10] K. T. Simons, C. Kooperberg, E. Huang and D. Baker, J. Mol. Biol. **268**, 209 (1997); C. Rohl, C. Strauss, K. Misura and D. Baker, Methods Enzymol. **383**, 66 (2004).

[11] G. Chikenji, Y. Fujitsuka and S. Takada, J. Chem. Phys. **119**, 6895 (2003); Y. Fujitsuka, G. Chikenji and S. Takada, Proteins **62**, 381 (2006); G. Chikenji, Y. Fujitsuka and S. Takada, J. Proc. Acad. Natl. Sci. U.S.A. **103**, 3141 (2006).

[12] J. Lee, S.-Y. Kim, K. Joo, I. Kim and J. Lee, Proteins **56**, 704 (2004); J. Lee, S.-Y. Kim and J. Lee, Biophys. Chem. **115**, 209 (2005); *ibid*, J. Korean Phys. Soc. **46**, 707 (2005).

[13] S.-Y. Kim, W. Lee and J. Lee, J. Chem. Phys. **125**, 194908 (2006).

[14] T.-K. Kim and J. Lee, J. Korean Phys. Soc. **52**, 137 (2008).

[15] K.-H. Cho, T.-K. Kim and J. Lee, J. Korean Phys. Soc. **52**, 143 (2008).

[16] C. A. Rohl, C. E. M. Strauss, D. Chivian and D. Baker, Proteins **55**, 656 (2004).

[17] 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, J. Am. Chem. Soc. **117**, 5179 (1995).

[18] H. Zhou and Y. Zhou, Protein Sci. **11**, 2714 (2002).

[19] C. Zhang, S. Liu and Y. Zhou, Protein Sci. **13**, 391 (2004).

[20] S. E. Brenner, P. Koehl and M. Levitt, Nuc. Acids Res. **28**, 254 (2000).

[21] J. Sim, S.-Y. Kim and J. Lee, Bioinformatics **21**, 2844 (2005).

[22] S.-Y. Kim, J. Sim and J. Lee, Lecture Notes in Bioinformatics **4115**, 562 (2006).

[23] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, Nuc. Acids Res. **25**, 3389 (1997).

[24] K. Joo, J. Lee, S.-Y. Kim, I. Kim, S. J. Lee and J. Lee, J. Korean Phys. Soc. **44**, 599 (2004).

[25] K. Joo, I. Kim, J. Lee, S.-Y. Kim, S. J. Lee and J. Lee, J. Korean Phys. Soc. **45**, 1441 (2004).

[26] R. L. Dunbrack and M. Karplus, J. Mol. Biol. **230**, 543 (1993).

[27] A. Fiser, R. K. Do and A. Sali, Protein Sci. **9**, 1753 (2000).

[28] Z. Xiang, C. S. Soto and B. Honig, Proc. Natl. Acad. Sci. U.S.A. **99**, 7432 (2002); J. Lee and C. Seok, Proteins **70**, 1074 (2008).