Protein structure prediction based on fragment assembly and parameter optimization

Julian Lee[a], Seung-Yeon Kim[b], Jooyoung Lee[b],*

[a]Department of Bioinformatics and Life Science, Computer Aided Molecular Design Research Center,
Bioinformatics and Molecular Design Technology Innovation Center, Soongsil University, Seoul 156-743, South Korea
[b]School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, South Korea

Abstract

We propose a novel method for ab-initio prediction of protein tertiary structures based on the fragment assembly and global optimization. Fifteen residue long fragment libraries are constructed using the secondary structure prediction method PREDICT, and fragments in these libraries are assembled to generate full-length chains of a query protein. Tertiary structures of 50 to 100 conformations are obtained by minimizing an energy function for proteins, using the conformational space annealing method that enables one to sample diverse low-lying local minima of the energy. Then in order to enhance the performance of the prediction method, we optimize the linear parameters of the energy function, so that the native-like conformations become energetically more favorable than the non-native ones for proteins with known structures. We test the feasibility of the parameter optimization procedure by applying it to the training set consisting of three proteins: the 10–55 residue fragment of staphylococcal protein A (PDB ID 1bdd), a designed protein betanova, and 1fsd.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Protein folding; Tertiary structure prediction; Ab initio prediction; Fragment assembly; Global optimization; Parameter optimization

1. Introduction

The prediction of the unique tertiary (three-dimensional) structure of a protein from its amino-acid sequence alone is one of the most important and challenging problems in biophysical chemistry today. The information on the tertiary structure of a protein is quite crucial in understanding the function and biological role of the protein. Popular approaches to this problem include comparative modeling [1–5] and fold recognition [6–9], which are classified as knowledge-based methods [5,10,11]. These methods use statistical information on sequences and their three-dimensional structures, in structural databases such as Protein Data Bank (PDB), to predict the unknown structure of a protein. Obviously, these methods can be used only when the amino-acid sequence of a target protein with unknown structure is related to those of one or more proteins with known structures.

On the other-hand, when homologous or weakly homologous sequences with known structures are not available, we turn to new fold (or ab initio) methods [5,13–26], which include energy-based methods. Energy-based methods are based on the thermodynamic hypothesis [27] that the native structure of a protein corresponds to the global minimum of its free energy for its physiological environment. However, although much progress has been made in energy-based methods [14–17], successful prediction of protein structure solely from the potential energy function still remains as a challenging problem. For this reason, most of recent new fold prediction methods use information on known structures to some degree. One of the popular trends among such methods is to determine the tertiary structure of a target protein by assembling fragments generated from the protein data bank (PDB). The effect of the short-range interactions is incorporated by using fragments from the PDB, and only long-range interaction terms are included in the energy

function, which are minimized in order to find conformations with optimal tertiary packing [18–21].

As in pure energy-based methods, there are two crucial elements for successful application of the fragment-based method for structure prediction, which are an accurate energy (or score) function and a powerful optimization algorithm for finding low-energy conformations. In a previous paperprofesy, we have introduced a fragment-based protein structure prediction method PROFESY (PROFile Enumerating SYstem), which utilizes the fragment library obtained from the secondary structure prediction method PREDICT (PRofile Enumeration DICTionary) [28,29]. In contrast to earlier methods where only simple sampling algorithms such as the simulated annealing method are used for generating low-energy conformations, PROFESY applies a powerful global optimization algorithm, conformational space annealing (CSA) [30–34], sykim, jul1 for sampling low-energy conformations. However, the energy function used was rather primitive partly due to the fact that the solvent effect was not properly incorporated. Moreover, various parameters in energy terms were set by crude guesses. Although some promising results were obtained from the benchmark tests of PROFESY, which is believed to be mainly due to the high efficiency of the sampling method, it is necessary to construct a reasonably accurate energy function for successful application of PROFESY to the protein structure prediction.

In this work, we address this problem by incorporating the solvation effect indirectly, by building C^{β} atoms and introducing a pairwise interaction term between them, whose strength depends on the types of the amino-acids in contact. Also, various parameters of the energy function are optimized in a systematic manner.

In fact, an iterative procedure which systematically refines the parameters of a given potential energy function was recently proposed [35] in the context of the energy-based method of protein structure prediction, and it was successfully applied to the parameter optimization of a coarse-grained potential energy function [35–39]. The method exploits the high efficiency of the CSA method in finding distinct low energy conformations. For a given set of proteins, whose low-lying local minimum-energy conformations for a given energy function is found by the CSA method, values of the parameters are modified so that native-like conformations of these proteins would have lower energies than those of non-native ones. Since the CSA method is also used for sampling the low energy conformations in PROFESY, it is straightforward to apply the above-mentioned parameter optimization procedure to PROFESY. In this work, in order to test the feasibility of such an idea, we optimize the parameters of the energy terms of PROFESY for the training set consisting of three proteins, the 10–55 residue fragment of staphylococcal protein A (PDB ID 1bdd), a designed protein betanova, and 1fsd.

2. Fragment assembly

We first briefly describe the way one generates conformations using fragment assembly PROFESY. The fragment libraries used in PROFESY are constructed using the recently proposed secondary structure prediction method PREDICT [28,29]. For each residue of a query sequence, a window of size fifteen is considered, where the center of the window is located on the residue under consideration. The fragment library of this residue is the collection of twenty backbone structures of the corresponding twenty nearest patterns in the pattern database of PREDICT. After constructing fragment libraries for all residues of a query sequence, full-length chain conformations can be constructed by assembling fragments in these libraries.

3. Energy function

The energy function is given by

$$U = \sum_{i,j} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \right) - w_h N_{\text{hb}} + \sum_{i,j} w_{a(i),a(j)} E_{a(i),a(j)},$$

(1)

where the summation index i, j are residue indices, $a(i)$ indicates the amino-acid type of the i-th residue, and r_{ij} is the $C_{\beta}-C_{\beta}$ distance between i-th and j-th residues. The first term is the Lennard–Jones 6–12 Van der Waals energy introduced to avoid steric clashes. In order to incorporate possible quantum effects, we use separate values of the Van der Waals interaction strength between the neighboring residues, which we call A' and B'. In the second term, N_{hb} is the number of hydrogen bonds between residues, which are at least five residues apart in sequence. This term is introduced to facilitate the β-strand pairing between extended fragments. A hydrogen bond is assumed to exist when an amide hydrogen atom and a carboxyl oxygen atom are placed within 2.24 from each other. The last term is a Miyazawa–Jernigan type contact term [40,41] between C_{β} atoms, introduced to incorporate the solvation effect in an indirect manner. Its functional form is given by

$$E_{a(i)a(j)} = f\left(\frac{r_{ij} - r_{a(i)a(j)}^{(c)}}{\Delta_{a(i)a(j)}} \right)$$

(2)

where

$$f(x) = \begin{cases} 1 & (x \leq -1) \\ \frac{1}{2} - \frac{15}{16}x + \frac{5}{8}x^3 - \frac{3}{16}x^5 & (-1 < x < 1) \\ 0 & (x \geq 1) \end{cases}$$

(3)

is a smoothed step function defining the contact. We used this function instead of the sharp step function, so that we can take the derivative with respect to the parameters, which is required for optimizing parameters. The numerical

coefficients in the polynomial are determined so that the function $f(x)$ and its derivative become continuous up to second order.

The first and second term in the energy function (1) are the same as those in Ref. [26], whereas the last term is the one we introduce in this work in order to account for the effect of solvation properly. We optimize the parameters A, B, A', B', w_h, w_{ab}, $r_{ab}^{(c)}$, and Δ_{ab}, total number of them being $5+210\times3=635$.

4. Local minimization

In order to apply the CSA method to the fragment assembly of protein tertiary structure prediction method such as PROFESY, one must define the concept of local minimization. In PROFESY, a conformation generated from a fragment assembly is locally minimized with respect to the energy by randomly selecting a residue and attempting to replace a part of the fifteen residue long fragment of the chain by another one in the corresponding library. If the new fragment can be inserted smoothly into the existing chain and if the new conformation is lower in energy than the existing one, the former replaces the latter. This process is continued either for 10 N_{seq} times, where N_{seq} is the length of the protein, or until the update attempts fail for N_{seq} consecutive times, whichever is encountered first, which completes the local minimization.

4.1. Conformational sampling

The CSA [30–32] is a powerful global optimization algorithm that has played the integral role in the recent success of the energy-based method for protein structure prediction [14–17]. A population of local minimum-energy conformations is maintained in the CSA method, which is called the bank. The diversity of the bank is directly controlled in CSA by introducing a distance measure $D(A,B)$ between two conformations A and B, and comparing it to a cutoff value, D_{cut}. As the algorithm proceeds, D_{cut} is gradually reduced, playing the role of temperature in simulated annealing. Hence the name "conformational space annealing". The annealing of D_{cut} amounts to shifting the emphasis from diversity of sampling at the early stage of the algorithm, to obtaining low-energy conformations at later stages, enabling efficient sampling of low-lying local minimum-energy conformations.

In order to check the performance of a potential energy function for a given set of parameters, one has to sample native-like and non-native conformations for each protein in the training set. To obtain native-like conformations, we have performed local energy minimization of the experimental native structure, and we have collected all the conformations obtained by successful fragment replacements during this local minimization. It should be noted that most of these native-like conformations are not local minimum-energy conformation. The conformations obtained from CSA search and local minimization of the native conformation are added to the structural database of local minimum-energy conformations for each protein.

5. Parameter refinement using linear programming

The procedure of the parameter optimization, which is described in this section, is almost the same as described in Ref. [39]. First, the changes of energy gaps are estimated by the linear approximation of the potential energy in terms of parameters. Since a potential can be considered to describe the nature correctly if the native-like structure has lower energy than non-native ones, the parameters are optimized to minimize the energy gap E_{gap},

$$E_{\text{gap}} = E^{\text{N}} - E^{\text{NN}} \tag{4}$$

for each protein in the training set, where E^{N} and E^{NN} are the energy of the native conformation and the lowest energy of the non-native conformations, respectively. We add to the energy a term proportional to the RMSD values of the conformations:

$$E = U + w_{\text{RMSD}} \text{ RMSD}. \tag{5}$$

The additional term is introduced in order to make the conformations with large RMSD to have high energies compared to ones with small RMSD values after the parameter optimization [39]. However, in contrast to Ref. [39] where the numerical value of w_{RMSD} was fixed to an arbitrary value of 0.3, in this work we determine its value at the initial stage of every iteration of the parameter optimization, which depends on the energy scale of the conformations:

$$w_{\text{RMSD}} = 0.5 \text{ RMSD}_0/|E_{\text{gap}}|, \tag{6}$$

where RMSD_0 is the RMSD value of the conformation with the lowest energy.

The parameter optimization is carried out by minimizing the energy gap E_{gap} of each protein in turn, while imposing the constraints that the energy gaps of the other proteins do not increase. Changing the parameters by small amounts, the energy with the new parameters can be estimated by the linear approximation:

$$E(\mathbf{x_{min}};p^{\text{new}}) \approx E(\mathbf{x_{min}};p^{\text{old}}) + \sum_i (p_i^{\text{new}} - p_i^{\text{old}}) \frac{\partial E(\mathbf{x_{min}};p^{\text{old}})}{\partial p_i}$$

where the p_i^{old} and p_i^{new} terms represent the parameters before and after the modification, respectively. The parameter dependence of the position of the local minimum can be neglected in the linear approximation, since the derivative in the conformational space vanishes at a local minimum [35]. The additional term w_{RMSD} RMSD of Eq. (5) vanishes in these expressions due to the same reason. The magnitude of the parameter change
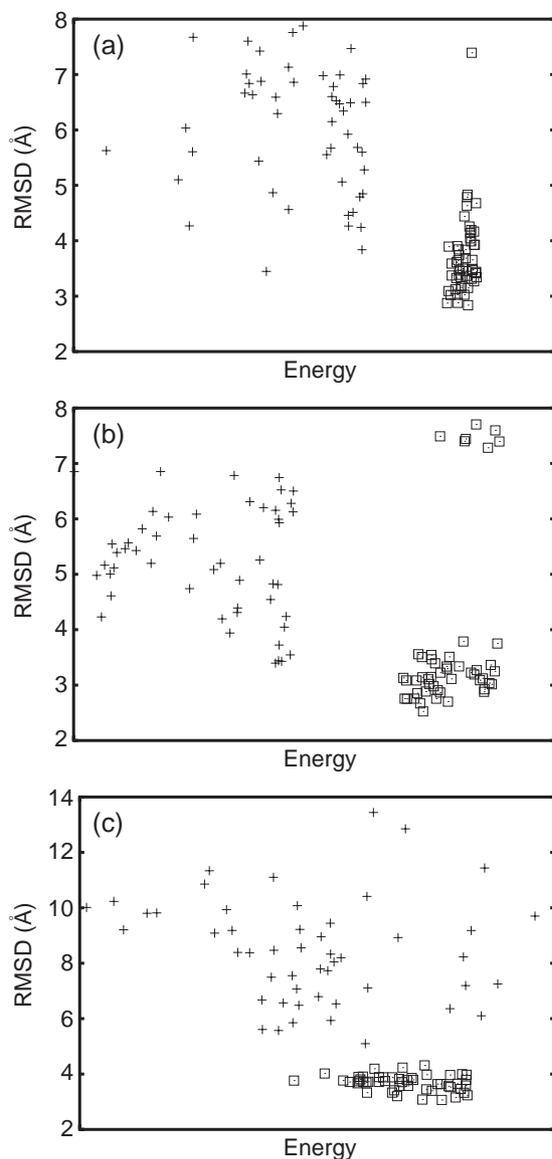
Fig. 1. Plots of the energy function and C^α RMSD (from the native structure) of three proteins obtained from CSA search using the initial and refined parameters. The crosses and squares denote the results obtained using the parameters before the optimization and after the 27-th iteration, respectively. The numerical values of energies are not displayed since they have no physical meaning, due to the fact that the overall scale of the energy is arbitrary in our procedure. The results are shown for (a) betanova, (b) 1fsd, and (c) 1bdd.

$\delta p_j \equiv p_j^{\mathrm{new}} - p_j^{\mathrm{old}}$ is bounded by a certain fraction ϵ of p_j^{old}. We use $\epsilon = 0.001$ for linear parameters and $\epsilon = 0.00001$ for nonlinear parameters in this study. The resulting optimization problem is a Linear Programming [39], which is solved by the primal-dual method with supernodal Cholesky factorization [42]. We select each protein in the training set in turn, and repeat this procedure (300 times in this work) of minimizing ΔE_{gap}.

It should be noted that we do not put any constraints on the overall scale of the parameters. Since the energies are proportional to the overall parameter scale, and since this

scale changes freely during the optimization process, the overall energy scale is not determined in our protocol. Therefore the numerical value of energy has no physical meaning in our work.

6. Iterative refinement of parameters

Since the change of energy gaps after the parameter change was estimated using the linear approximations, we now have to evaluate the true energy gaps using the newly obtained parameter set. Therefore, we reminimize the conformations in the structural database with the new parameter set. We also perform the CSA search with the new parameters [38,39]. The low-lying local energy minima found in the new conformational searches are added into the
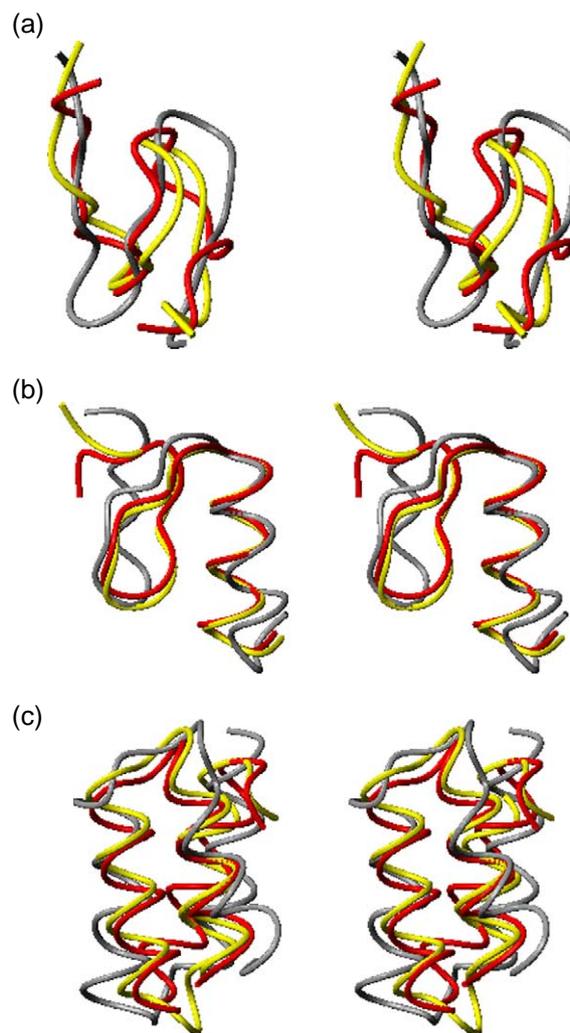

Fig. 2. The C^α trace of GMEC found with the optimized parameters is shown together with the native structure for each of the three proteins in the training set, (a) betanova, (b) 1fsd, and (c) 1bdd. The native structure is shown in grey, the GMEC in red, and the conformation with the smallest RMSD value in yellow. The figures are prepared with the program MOLMOL [43]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structural database of local energy minima. The conformations in the database are used to obtain the energy gaps, which are used for the new round of parameter refinement. As the procedure of [CSA→parameter refinement→energy reminimization] is repeated, the number of conformations in the structural database increases [38,39]. This iterative procedure is continued until sufficiently good native-like conformations are found from the CSA search.

7. Results

We have applied our protocol to a training set consisting of three proteins. They are the designed protein betanova, 1fsd, and the 10–55 fragment of the B-domain of staphylococcal protein A (1bdd), which are 20, 28, and 46 residues long, respectively. The betanova is a β protein, 1fsd is an á/β protein, and 1bdd is an á protein, which represent structural classes of small proteins. The initial parameter set is the one used in CASP5 [26] except those for the contact terms. The initial parameters for the contact terms are those obtained by Miyazawa and Jernig [41].

Fifty conformations were sampled in each CSA search, and the global minimum-energy conformations (GMECs) found with the initial parameters have RMSD values of 5.6, 6.9, and 10.0 Å, respectively, and the smallest values of RMSD found from the CSA search are 3.4, 3.4 and 5.1 Å, respectively. After the 27-th iteration of the parameter refinement, the conformations with smaller values of RMSD are found from the global CSA search. The GMECs have RMSD values of 2.9, 3.1, and 3.8 Å and the smallest values of RMSD found are 2.8, 2.5, and 3.1 Å, respectively. The results of the global search with the initial and optimized parameter set for the three proteins are plotted using different symbols in terms of energy and RMSD in Fig. 1. The C^α traces of the GMECs of the three proteins found using the parameters obtained after the 27-th iteration of optimization are shown in Fig. 2 along with the native conformations.

8. Discussion

In this work, we have improved PROFESY, a novel method for the prediction of protein tertiary structure based on the fragment assembly, by introducing solventation energy terms and systematically optimizing the energy parameters. The parameter optimization was performed by applying the general protocol for the force field parameter optimization and landscape design which have been used previously only in the context of the pure energy-based method. Using this procedure, we optimized the 635 parameters so that they correctly describe the energetics of three selected proteins simultaneously. This optimized parameter set yielded GMECs with RMSD values of 1.9, 3.1, and 3.8 Å for betanova, 1fsd, and 1bdd, respectively.

Further tests of parameter optimization on training sets containing a larger number of proteins, as well as jackknife tests on proteins not included in the training sets, are necessary for successful application of our method to the structure prediction of unknown proteins. These are currently under investigation.

References

[1] A. Fiser, R.K.G. Do, A. Sali, Modeling of loops in protein structures, Protein Sci. 9 (2000) 1753–1773.

[2] A. Kolinski, M.R. Betancourt, D. Kihara, P. Rotkiewicz, J. Skolnick, Generalized Comparative Modeling (GENECOMP): a combination of sequence comparison, threading, lattice and off-lattice modeling for protein structure prediction and refinement, Proteins 44 (2001) 133–149.

[3] P.A. Bates, L.A. Kelley, R.M. MacCallum, M.J.E. Sternberg, Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM, Proteins (Suppl. 5) (2001) 39–46.

[4] C. Venclovas, Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment, Proteins (Suppl. 5) (2001) 47–54.

[5] D. Baker, A. Sali, Protein structure prediction and structural genomics, Science 294 (2001) 93–96.

[6] K.K. Koretke, R.B. Russell, A.N. Lupas, Fold recognition from sequence comparisons, Proteins (Suppl. 5) (2001) 68–75.

[7] A.G. Murzin, A. Bateman, CASP2 knowledge-based approach to distant homology recognition and fold prediction in CASP4, Proteins (Suppl. 5) (2001) 76–85.

[8] K. Karplus, R. Karchin, C. Barrett, S. Tu, M. Cline, M. Diekhans, L. Grate, J. Casper, R. Hughey, What is the value added by human intervention in protein structure prediction? Proteins (Suppl. 5) (2001) 86–91.

[9] M.G. Williams, H. Shirai, J. Shi, H.G. Nagendra, J. Mueller, K. Mizuguchi, R.N. Miguel, S.C. Lovell, C.A. Innis, C.M. Deane, L. Chen, N. Campillo, D.F. Burke, T.L. Blundell, P.I.W. de Bakker, Sequence–structure homology recognition by iterative alignment refinement and comparative modeling, Proteins (Suppl. 5) (2001) 92–97.

[10] J. Moult, T. Hubbard, K. Fidelis, J.T. Pedersen, Critical assessment of methods of protein structure prediction (CASP): round III, Proteins (Suppl. 3) (1999) 2–6.

[11] J. Moult, K. Fidelis, A. Zemla, T. Hubbard, Critical assessment of methods of protein structure prediction (CASP): round IV, Proteins (Suppl. 5) (2001) 2–7.

[12] A.M. Lesk, L.L. Conte, T.J.P. Hubbard, Assessment of novel fold targets in CASP4: predictions of three-dimensional structures, secondary structures, and interresidue contacts, Proteins (Suppl. 5) (2001) 98–118.

[13] J. Lee, A. Liwo, H.A. Scheraga, Energy-based de novo protein folding by conformational space annealing and an off-lattice united-residue force field: application to the 10–55 fragment of staphylococcal protein A and to apo calbindin D9K, Proc. Natl. Acad. Sci. 96 (1999) 2025–2030.

[14] J. Lee, A. Liwo, D.R. Ripoll, J. Pillardy, H.A. Scheraga, Calculation of protein conformation by global optimization of a potential energy function, Proteins (Suppl. 3) (1999) 204–208.

[15] J. Lee, A. Liwo, D.R. Ripoll, J. Pillardy, J.A. Saunders, K.D. Gibson, H.A. Scheraga, Hierarchical energy-based approach to protein-structure prediction: blind-test evaluation with CASP3 targets, Int. J. Quant. Chem. 77 (2000) 90–117.

[16] A. Liwo, J. Lee, D.R. Ripoll, J. Pillardy, H.A. Scheraga, Protein structure prediction by global optimization of a potential energy function, Proc. Natl. Acad. Sci. 96 (1999) 5482–5485.

[17] K.T. Simons, C. Kooperberg, E. Huang, D. Baker, Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, J. Mol. Biol. 268 (1997) 209–225.

[18] K.T. Simons, C. Strauss, D. Baker, Prospects for ab initio protein structural genomics, J. Mol. Biol. 306 (2001) 1191–1199.

[19] R. Bonneau, J. Tsai, I. Ruczinski, D. Chivian, C. Rohl, C.E.M. Strauss, D. Baker, Rosetta in CASP4: progress in ab initio protein structure prediction, Proteins (Suppl. 5) (2001) 119–126.

[20] D.T. Jones, Predicting novel protein folds by using FRAGFOLD, Proteins (Suppl. 5) (2001) 127–132.

[21] D.M. Standley, V.A. Eyrich, Y. An, D.L. Pincus, J.R. Gunn, R.A. Friesner, Protein structure prediction using a combination of sequence-based alignment, constrained energy minimization, and structural alignment, Proteins (Suppl. 5) (2001) 133–139.

[22] D. Xu, O.H. Crawford, P.F. LoCascio, Y. Xu, Application of PROSPECT in CASP4: characterizing protein structures with new folds, Proteins (Suppl. 5) (2001) 140–148.

[23] J. Skolnick, A. Kolinski, D. Kihara, M. Betancourt, P. Rotkiewicz, M. Boniecki, Ab initio protein structure prediction via a combination of threading, lattice folding, clustering, and structure refinement, Proteins (Suppl. 5) (2001) 149–156.

[24] D. Kihara, H. Lu, A. Kolinski, J. Skolnick, Touchstone: an ab initio protein structure prediction method that uses threading-based tertiary restraints, Proc. Natl. Acad. Sci. 98 (2001) 10125–10130.

[25] J. Lee, S.-Y. Kim, K. Joo, I. Kim, J. Lee, Prediction of protein tertiary structure using PROFESY, a novel method based on fragment assembly and conformational space annealing, Proteins: Struct., Funct., Bioinform. 56 (2004) 704–714.

[26] C.B. Anfinsen, Studies on the principles that govern the folding of protein chains, Science 181 (1973) 223–230.

[27] K. Joo, J. Lee, S.-Y. Kim, I. Kim, S.J. Lee, J. Lee, Profile-based nearest neighbor method for pattern recognition, J. Korean Phys. Soc. 44 (2004) 599–604.

[28] K. Joo, I. Kim, S.-Y. Kim, J. Lee, J. Lee, S.J. Lee, Prediction of the secondary structures of proteins using PREDICT, a nearest neighbor method on pattern space, J. Korean Phys. Soc. 45 (2004) 1441.

[29] J. Lee, H.A. Scheraga, S. Rackovsky, New optimization method for conformational energy calculations on polypeptides: conformational space annealing, J. Comput. Chem. 18 (1997) 1222–1232.

[30] J. Lee, H.A. Scheraga, S. Rackovsky, Conformational analysis of the 20-residue membrane-bound portion of melittin by conformational space annealing, Biopolymers 46 (1998) 103–115.

[31] J. Lee, H.A. Scheraga, Conformational space annealing by parallel computations: extensive conformational search of met-enkephalin and the 20-residue membrane-bound portion of melittin, Int. J. Quant. Chem. 75 (1999) 255–265.

[32] S.-Y. Kim, S.J. Lee, J. Lee, Conformational space annealing and an off-lattice frustrated model protein, J. Chem. Phys. 119 (2003) 10274–10279.

[33] J. Lee, I.H. Lee, J. Lee, Unbiased global optimization of Lennard–Jones clusters for $N \leq 201$ using conformational space annealing method, Phys. Rev. Lett. 91 (2003) 0802011–0802014.

[34] J. Lee, D.R. Ripoll, C. Czaplewski, J. Pillardy, W.J. Wedemeyer, H.A. Scheraga, Optimization of parameters in macromolecular potential energy functions by conformational space annealing, J. Phys. Chem., B 105 (2001) 7291–7298.

[35] J. Pillardy, C. Czaplewski, A. Liwo, W.J. Wedemeyer, J. Lee, D.R. Ripoll, P. Arlukowicz, S. Oldziej, Y.A. Arnautova, H.A. Scheraga, Development of physics-based energy functions that predict medium-resolution structures for proteins of the α, beta, and α/β structural classes, J. Phys. Chem., B 105 (2001) 7299–7311.

[36] A. Liwo, P. Arlukowicz, C. Czaplewski, S. Oldziej, J. Pillardy, H.A. Scheraga, A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape: application to the UNRES force field, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 1937–1942.

[37] J. Lee, K. Park, J. Lee, Full optimization of linear parameters of a united residue protein potential, J. Phys. Chem., B 106 (2002) 11647–11657.

[38] J. Lee, S.-Y. Kim, J. Lee, Design of a protein potential energy landscape by parameter optimization, J. Phys. Chem., B 108 (2004) 4525–4534.

[39] S. Miyazawa, R.L. Jernigan, Estimation of effective inter-residue contact energies from protein crystal structures: quasi-chemical approximation, Macromolecules 18 (1985) 534–552.

[40] S. Miyazawa, R.L. Jernigan, Residue–residue potentials with a favorable contact pair term and an unfavorable high packing density term for simulation and threading, J. Mol. Biol. 256 (1996) 623–644.

[41] C.A. Mészáros, Fast cholesky factorization for interior point methods of linear programming, Comput. Math. Appl. 31 (1996) 49–51.

[42] R. Koradi, M. Billeter, K. Wuthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph. 14 (1996) 51–55.