Unbound Protein-Protein Docking Using Conformational Space Annealing

Kyoungrim Lee, Kee-Hyoung Joo and Jooyoung Lee
School of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea
Email : jlee@kias.re.kr

ABSTRACT: We have studied unbound docking for 12 protein-protein complexes using conformational space annealing (CSA) combined along with statistical pair potentials. The CSA, a powerful global optimization tool, is used to search the conformational space represented by a translational vector and three Euler amgles between two proteins. The energy function consists of three statistical pair-wise energy terms; one from the distance-scaled finite ideal-gas reference state (DFIRE) approach by Zhou and the other two derived from residue-residue contacts. The residue-residue contact terms describe both attractive and repulsive interactions between two residues in contact. The performance of the CSA docking is compared with that of ZDOCK, a well-established protein-protein docking method. The results show that the application of CSA to the protein-protein docking is quite successful, indicating that the CSA combined with a good scoring function is a promising method for the study of protein-protein interaction.

1 INTRODUCION

Issues in protein-protein docking include the prediction of the 3-dimensional structure of a protein-protein complex from its component proteins as well as the study of its protein-protein interaction. Understanding the protein-protein interaction can provide insights for the mechanism of molecular interactions in biological systems. Computational protein docking procedure typically involves the following two steps. First, decoys are generated using conformational search algorithms. Second, among these decoys near-native complex structures are selected using a score/energy function. More detailed information on the docking methods can be found in recent reviews.[1-7]

One of the difficulties in protein-protein docking is that both receptor and ligand molecules may undergo conformational changes upon protein association. Often, the side-chain conformational changes and/or large backbone movements are observed. To develop an efficient docking method, one should consider a test set of protein complexes containing the unbound structures of both the receptor and the ligand, and the bound complex structure. The docking efficiency can be evaluated by measuring the structural accuracies of the modeled complexes relative to their native complexes. In most docking procedures, a set of protein complex models is generated from the given unbound structures of two component proteins, and near-native complex models are determined (if any) by comparing them with the native complex.

If one is equipped with an accurate energy function which can efficiently discriminate the native association of the component proteins from a variety of non-native associations, then the protein docking can be considered as one of global optimization problems since the success of the docking application depends on the identification of the most stable association of the proteins using a given energy function. For this reason, successful docking procedures involve a rigorous conformational search of a given system considering the relative position and orientation of the component proteins as well as their flexibilities.

Conformational space annealing (CSA),[8,9] one of the most efficient global optimization, has been successfully applied to various problems including ab initio protein structure prediction,[10-12] 3D-structure prediction of multi-chain homo-oligomer proteins[13,14] and small-molecule docking.[15] The CSA combined with all-atom AMBER94 potential[16] was also used to the Critical Assessment of Prediction of Interactions (CAPRI) blind experiment[17] on the comparative evaluation of protein-protein docking for structure prediction.[18] The results[18] show that the CSA method has a potential for the study of protein-protein interaction. The basic idea of CSA is that it enforces a broad conformational sampling in early stages of simulation and gradually directs the search into various narrow regions populated with low-energy conformations. The major advantage of the CSA is that it can find many distinct families of low-energy conformations. This makes it possible to search the whole intermolecular phase space of the protein-protein association for a given energy function. The CSA deals with the population containing diverse solutions by directly controlling the diversity of the population. Consequently, it can generate many distinct low-energy solutions, one of which may correspond to the true solution if the energy function used is reasonably accurate.

How accurately we can describe the interaction between the receptor and ligand depends on the energy function in use. The energy function can be of various forms. It may include geometric and chemical complementarities, as well as electrostatic interaction, hydrogen-binding interaction, and solvation energy terms. All-atom empirical potentials and/or database-derived score functions can be also used. Recently, a residue-specific all-atom, distance-dependent potential of mean-force was developed by by Zhou.[19] They introduced a reference state, namely the distance-scaled, finite ideal-gas reference (DFIRE) state, to construct the potential of mean force from a database of 1011 non-homologous (less than 30% homology) protein structures with resolution less than 2 Å. The energy function is shown to effectively select native structures from decoys[19] and to predict the mutation-induced change in stability[19] and loop conformations.[19,20] The successful applications of DFIRE-based energy function to the single-chain proteins motivated us to consider it for the protein-protein docking problems.

In this work, we combine the CSA with an energy function consisting of the DFIRE-based energy term and two additional residue-based contact energy terms developed in our group to carry out unbound protein-protein docking. The two residue-residue contact energy terms represent both attractive and repulsive interactions. The test set of 12 protein complexes used in this docking study is a subset of 54 unbound protein-protein pairs from the benchmark0.0 investigated by Chen et al.[21] The criteria for selecting this subset of 12 complexes were; (i) complexes should be a protein dimer and (ii) their sequence lengths are less than 300 amino acid residues. The goal for this study is to investigate the applicability of the CSA coupled with the statistical pairwise potentials for the protein-protein docking. The CSA results will be compared with those from ZDOCK,[22] a well-known protein-docking method. The rest of this article contains computational details including algorithms and implementation followed by the results. Analysis and discussions are provided by highlighting key findings and suggestion for further improvement.

2 COMPUTATIONAL DETAILS

2.1 Conformational Space Annealing

We provide only a brief description of essential implementation of the CSA algorithm to the protein-protein docking study. Details of the CSA and its applications are available in the references.[8-15] The general mechanism of CSA algorithm is shown in Figure. 1.

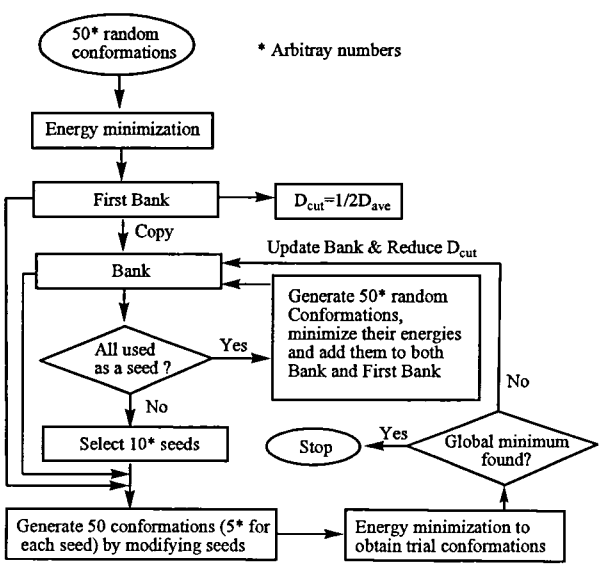

Figure 1. Flow chart of the CSA algorithm

In this work, we have considered only rigid-body variables which describe translational and rotational moves between the receptor and the ligand. In this work, we follow the CSA procedure described in reference 18. Initially, we generate 200 random conformations by assigning random translation vectors (x,y,z) and rotational Euler angles (ϕ, θ, ψ) to the ligand with respect to its receptor which is fixed. Energies of these random conformational complexes are subsequently minimized by quenching the energy through small random moves of the ligand relative to the receptor. The details of energy-quenching are discussed in the section 2.3. Throughout the paper, the term of *minimization* refers to the local energy-minimization by quenching. We call the set of these minimized complexes the *first bank* and make a copy of this first bank to the bank. The conformations in the bank are updated in later stages, while those in the first bank are kept unchanged. The initial value of the structural difference cutoff, D_{cut}, is set as $D_{ave}/2$ where D_{ave} is the average pairwise distance between all existing conformation pairs in the first bank. The conformational variables of protein complexes are the rigid-body variables (3 for the translational vector and the other 3 for Euler angles). To generate new conformations, we first choose 100 seed conformations from the bank and then replace a subset of the variables from a seed conformation with the corresponding one from a conformation in the bank or in the first bank. Their corresponding energies are subsequently minimized, and these minimized conformations become trial conformations.

The distance $D(\alpha,A)$ between a trial conformation α and a conformation A in the bank can be defined by

$$D(\alpha,A) = \left|\Delta T_{\alpha A}(x,y,z)\right| + \omega_{\Theta}\Theta_{\alpha A}(\phi,\theta,\psi) \qquad (1)$$

where $\Delta T_{\alpha A}(x,y,z)$ is the translation vector from A to α, and $\Theta_{\alpha A}(\phi,\theta,\psi)$ is the angle between the two Euler angle (ϕ,θ,ψ) vectors of A and α. The weight factor ω_{Θ} is set so that two terms in the right-hand side of eq. (1) contribute equally on average. Let's suppose that, among all bank conformations, A is the closest to α. If $D(\alpha,A) < D_{cut}$, we consider that α is similar to A. Then, the conformation with a lower energy between α and A is kept in the bank and the other is discarded. On the contrary, if $D(\alpha,A) > D_{cut}$, α is regarded as distinct from A and consequently from all conformations in the bank. In this case, the conformation with the highest energy among all bank conformation plus α is discarded, and the rest are kept in the bank so that the total number of the bank conformations remains unchanged through this procedure. We perform this operation until the bank is updated using all available trial conformations.

The value of D_{cut} is reduced by a fixed ratio after the bank updating takes place. The value of D_{cut} is set to $D_{ave}/2$ and reaches to its final value $D_{ave}/5$ after 200,000 minimization steps. After the value of D_{cut} is reduced, new seeds are selected from the bank conformations that have not yet been used as seeds before, to repeat the aforementioned procedure. The value of D_{cut} is kept constant after it reached the final value. When all conformations in the bank are used as seeds, one round of iteration is completed. We perform additional iterations of search by first erasing the record of bank conformations having been used as seeds, and by starting a new round of iteration.

For a problem with known global minimum, the whole procedure stops when the global minimum is found which is examined immediately after the bank is updated by all trial conformations. Since in this work, we do not know the value of the global minimum energy, we have set an arbitrary stopping criterion. That is, after 3 iterations are completed, we add additional 200 randomly-generated and minimized conformations to the first bank and to the bank,

and repeat the same procedure above for additional 3 iterations to finish the search. Therefore, the conformational search continues until the bank size becomes larger than the preset maximum number (400 in this work). It should be noted that since one iteration is completed only after all bank conformations have been used as seeds, and since we add an additional number of conformations whenever our search reaches a deadlock, there is no loss of generality for using particular values for the number of seeds, the number of bank conformations, and so forth.

2.2 Energy Function

The energy function used in this work consists of three terms; DFIRE-based statistical pair potential developed by Zhou[19] and residue-based attractive and repulsive contact energies that we have developed in our lab. Details of these terms are described below.

2.2.1 DFIRE-based statistical pair potential

The details of DFIRE-based statistical pair potential and its successful applications can be found in the references.[19,20] We briefly discuss about the essential points of the DFIRE energy function and its implementation in this work. The key of the DFIRE potential is to use a physical references state of ideal gases, which appears to make the DFIRE energy function more accurate. The reference state of uniformly distributed points in finite spheres is called distance-scaled finite, ideal gas reference (DFIRE) state and used as the zero-interaction reference state. The residue-specific all-atom, distance-dependent potential of mean force was constructed from the structures of single-chain proteins by utilizing the DFIRE state.
The DFIRE reference state can be derived using simple statistical equations[23] as follows:

$$N_{\exp}(i,j,r) = (r/r_{cut})^{\alpha}(\Delta r/\Delta r_{cut})N_{obs}(i,j,r_{cut}) \qquad (2)$$

where $N_{exp}(i,j,r)$ is the expected number of non-interacting atomic pairs (i,j) in a same distance shell $r - \Delta r/2$ to $r + \Delta r/2$ only by considering a short-range interaction with a cutoff distance of r_{cut} in the reference state; $N_{obs}(i,j,r)$ is the observed number of the same atomic pairs (i,j) with the same distance shell in the database used; and α is a given constant. The estimation of the α values is explained well in the reference 19. Using eq. (2), the DFIRE-based potential is written as:

$$\bar{u}(i,j,r) = \eta RT \ln \frac{N_{obs}(i,j,r)}{\left(\frac{r}{r_{cut}}\right)^{\alpha}\frac{\Delta r}{\Delta r_{cut}}N_{obs}(i,j,r_{cut})} \qquad (3)$$

where η is a constant needed for mutation-induced stability change, R the gas constant, and T is temperature (300K).
The DFIRE-based potential has been successfully applied to selecting native structures from protein decoys,[19] predicting mutation-induced change in stability[19] and loop conformation.[19,20] We have used the DFIRE 1.0 program[19] into CSA to study protein-protein docking.

2.2.2 Residue-based attractive energy

We have developed a set of residue-based contact energy parameters by extracting the distances of all residue pairs from a set containing 6,204 proteins which have been selected from PDB_SELECT90 database.[24,25] The protein set of 6,204 proteins was prepared by removing the proteins with missing residues among 7,217 proteins in the PDB_SELECT90. Contacts are defined between residues i and j where the C_β-C_β distance is within 7.0 Å considering only those residues at least 4 residues apart long the sequence (i.e. $|i - j| \geq 4$). The way we developed the residue-base contact energy function is as follows. First, we count the number of residue pairs, e.g. (A,B), which are in contact, and the number of their component residues of the pair in a protein of the protein set. Second, to obtain the frequency of the (A,B) pair, we divide the number of the (A,B) pair by the value of the number of A times the number of B, and sum the divided value over all chains in the set. The frequency calculation can be defined by eq. (4)

$$P_{(A,B)} = \sum_{i}^{N_T} \frac{N_i^{(A,B)}}{N_i^A \cdot N_i^B} \qquad (4)$$

where $N_i^{(A,B)}$, N_i^A and N_i^B are the numbers of (A,B) pair, residue A and residue B in a protein i, and N_T the total number (=6,204) of proteins in the set. Third, we construct a residue contact matrix by dividing the frequency values of all possible pairs by the smallest frequency value and then by taking the logarithm of the values. The elements of the contact matrix have all non-negative values so that all residue contacts basically produce favorable interactions. For example, the matrix element $M_{(A,B)}$ of the contact pair (A,B) can be derived below in eq. (5).

$$M_{(A,B)} = \ln \frac{P_{(A,B)}}{P_{\min}} \qquad (5)$$

In this protein-protein docking, since the interactions between the receptor and the ligand are only taken into account, A should belong to the receptor and B to the ligand, or vice versa. The attractive contact energy resulting from the residue pairs in contact for protein docking is defined by eq. (6)

$$E_{attractive} = -\frac{1}{2}\sum_{i,j}^{20} M_{(i_R,j_L)} \qquad (6)$$

where (i_R, j_L) indicates the pair of contact residues i_R and j_L which belong to the receptor and the ligand, respectively and a factor of 1/2 is used to avoid double counting of the residue-residue contacts.

2.2.3 Residue-based repulsive energy

The third term of the energy function is residue-based repulsive contact energy. This energy term is used to avoid any possible clashes between two residues in contact. The clashes may take place because the residue-based contact is always favored regardless of the types of the residue pairs in contact. To derive the repulsive energy, we used the same protein set prepared for obtaining the residue-based attractive contact energy in section 2.2.2. We describe the

repulsive interactions between two contact residues using the C_β atoms and the nitrogen (N) and oxygen (O) atoms of the protein backbones. We measured the distances of the C_β-C_β atoms, N-N atoms and O-O atoms and found their probable smallest distances d_{min}'s to be 3.4 (C_β-C_β), 3.6 (N-N), and 2.8 (O-O) Å by excluding the unrealistically small distances when their corresponding residues are in contact. We set the maximum contact distances d_{max}'s of the three atomic pairs to d_{min}'s + 0.5 Å so that they can have an allowable movement of 0.5 Å. The repulsive energy contains three terms since we deal with three different types of atomic pairs. Each term has the same form as defined for C_β-C_β repulsion below in eq. (7)

$$E_{repulsive}^{C_\beta-C_\beta} = \sum_{ij}\left(\frac{\left(d_{\max} - d_{ij}^{C_\beta-C_\beta}\right)}{\left(d_{\max} - d_{\min}\right)}\right)^8 \quad (7)$$

where $d_{ij}^{C_\beta-C_\beta}$ is the distance between the C_β atoms of the contact residues i and j.

2.2.3 Combining all energy terms

The energy function is built by combining all energy terms as below;

$$E = E_{DFIRE} + w_1 E_{attractive} + w_2 E_{repulsive} \quad (8)$$

where the weight factors, w_1 and w_2, are set to 0.2 and 1.0, respectively.

The DFIRE-based potential is derived from all-atom interactions while the last two terms are calculated from residue-based interactions. The interacting proteins usually undergo conformational changes of their side-chains and/or backbones, which can not be described by DFIRE energy function. We have added the residue-based contact energy terms to the DFIRE potential in order to give softness to the rigid unbound protein-protein docking. We hope that the near-native complexes remain in the final bank with help of favored residue-contact interactions, even though their DFIRE-based interaction is unfavorable.

2.3 Unbound Protein-Protein Docking

To evaluate the performance of CSA docking coupled with the statistical pairwise energy function, we prepared the test set containing 12 unbound protein complexes of relatively small size (less than 300 residues) and dimer molecules from the benchmark0.0 developed by Chen et al.[21] The CSA docking was initiated by generating the first bank containing 200 randomly generated and minimized docked complexes. The number of seed conformations was 100. For each seed conformation, 5 perturbed conformations were generated (3 by replacing the translational vector and 2 by replacing the Euler angles). Therefore, a total of 500 perturbed conformations were generated, and they were energy-minimized to obtain trial conformations. Using these trial conformations, the bank was updated. The rest of procedure is as explained above in the section 2.1. The CSA search continued until the bank size became larger than the maximum size of 400 so that we could have 400 decoys for each complex.

The energies of the complex conformers generated during the CSA search were minimized by the quenching method. The energy function used in this work is a discrete statistical pairwise potential which can not be minimized using the energy gradient-based methods. The energies were quenched to be lower through small random moves such as translations (±1Å) and rotations (±3°). Only the moves which were able to lower the energy were accepted and the minimization process was ended when 50 moves were successively rejected. The maximum allowed number of moves is set to 200 for each conformer generated by the CSA operation.

3 RESULTS

The performance of the CSA docking is compared with that of ZDOCK (specifically ZDOCK1.3),[22] one of well-known protein-protein docking tools. The energy function in ZDOCK1.3 has three terms: grid-based shape complementarity, desolvation, and electrostatics. For simplicity, the term ZDOCK is just used to refer to ZDOCK1.3. The evaluation of protein docking performance is typically made based on the root mean squared deviation (RMSD) values of decoy complexes from the native complex structure, and the fractions of the native residue-residue contacts of the decoy structures. The near-native structures are defined as the structures whose RMSD values are less than 10 Å, or whose native residue-residue contact fractions are greater than 25%. Another criterion for determining the near-native structure is the RMSD calculated over the C_α atoms of interface residues, which are residue pairs between receptor and ligand with at least one inter-residue heavy atom pairs less than 10 Å. Decoy with the C_α-RMSD less than 2.5 Å is considered a near-native structure. We have used the C_α-RMSD value for performance evaluation of the CSA docking.

PDB ID	CSA dock (400-decoy set)		ZDOCK1.3 (1000/2000-decoy set)[a]	
	No. of native hits[b]	Highest rank[c]	No. of native hits[b]	Highest rank[c]
1A0O	1	116	4/9	619
1ACB	1	47	154/199	3
1AVZ	-	-	-	-
1BRC	1	71	9/24	52
1BRS	3	2	-/3	-/1019
1CGI	1	27	43/77	3
1CHO	1	43	53/93	22
1MEL	2	58	19/32	9
1PPE	6	1	257 /318	1
1TAB	2	40	-	-
1TGS	2	12	60 /86	5
2PTC	1	248	38 /62	65

Table 1. Performance comparison between the CSA dock and ZDOCK1.3

[a] Only top-1000 decoys and top-2000 decoys (1000decoys/2000 decoys) are taken out of the first 54,000 decoys predicted by ZDOCK1.3 using the energy function consisting of grid-based shape complementarity, desolvation, and electrostatics. [b] No. of native hits are defined as the number of docked complexes

with interface Cα-RMSD less than 2.5Å. The first value is the number from the top-1000 decoy set and the second is the number from the top-2000 decoy set. [c] Highest rank represents the first energetic rank of the near-native structures obtained from the CSA dock and ZDOCK1.3. The numbers are shown first from the top 1000 decoy set and second from the top 2000 decoy set.

Table 1 shows the comparison of docking performance between the ZDOCK and the CSA methods. The CSA docking produced 400 docking decoys in the final bank, which are compared with the top-1000 decoys and the top-2000 decoys out of the first 54,000 decoys predicted by the ZDOCK method. The CSA gave at least one near-native decoy in the final 400-decoy set except one case of 1AVZ whereas the ZDOCK were unable to find near-natives structure from the 1000-decoy set for three cases (1AVZ, 1BRS and 1TAB) and from 2000-decoy set for two cases (1AVZ and 1TAB) even though the sizes of decoy sets of ZDOCK were much larger than that of CSA. For some cases such as 1ACB, 1CGI, 1CHO, 1MEL, 1PPE, 1TGS and 2PTC, however, ZDOCK shows relatively much more hits of near-native structures. This mainly originates from the difference of the way the two methods generate the decoy set of protein complexes. ZDOCK performs the grid-based search of conformations of protein-protein complex and keeps decoys with a certain number of top ranked energies in the final decoy set. Some decoys in the decoy set can be very similar to each other. However, the CSA docking carries out a rigorous conformational search maintaining structural diversity to give a collection of structurally dissimilar and locally minimized conformers. If some decoys are structurally close to each other within a d_{cut}, one with lower energy remains in the updated bank while the other is discarded. The rank of near-native structures given by the CSA is higher for three cases (1A0O, 1BRS and 1TAB). The ranks for 1A0O are 116th (CSA) and 619th (ZDOCK). For 1BRS, the ranks are 2nd by the CSA and 1019th only available from the top-2000 ZDOCK decoys. For 1TAB, the rank from the CSA is 40th where the ranks from either the top-1000 decoys or the-2000 decoys are not available. For the 1PPE case (1PPE) which is the easiest application, the near-native structures obtained from the two methods are energetically ranked as the first. For the other complexes, the ranks given by ZDOCK are higher.

4 CONCLUSION AND DISCUSSIONS

We have developed a new protein-protein docking method using the CSA coupled with statistical pairwise potentials. The CSA is able to carry out a rigorous conformational search of two interacting proteins to eventually produce a given number of distinct families of low-energy conformations of docked complexes. The energy function built with the statistical pairwise potentials, i.e. database-derived score function, consists of the DFIRE-base all-atom pairwise statistical potential, residue-based attractive and repulsive contact energies. The DFIRE-based energy helps make an accurate description of all-atom interactions between the two rigid bodies while the last two residue-based energy terms favor the residue-residue interactions within a certain range of contact distance. We have tried to give softness to the unbound rigid-body docking by combining all energy components since the interacting proteins may undergo conformational changes upon their association. 12 relatively small and dimeric unbound-proteins were taken from the benchmark0.0. We have carried out the CSA docking over the 12 test protein-pairs and subsequently compared the performance between the CSA and the ZDOCK methods. The results show that the CSA has a higher probability of having a near-native structure in the final decoy set than the ZDOCK because the CSA operation maintains a structural diversity in the bank. On the other hand, the ranks of near-native structures given by ZDOCK are generally higher than those given by the CSA docking. It is understood that these observations largely results from the differences of the conformational search methods and the energy functions used by the two docking tools.

The unbound rigid docking has an essential limitation that it cannot model the flexible motions of the interacting proteins. Modeling the protein's flexibility in protein-protein docking by torsional changes is almost impossible in reality because the number of degree of freedoms considered becomes tremendously large. However, the limitation can be alleviated partially by introducing the softness provided by the energy function into the interaction of the protein pair. Any energy functions used for these docking problems can not describe the protein-protein interaction perfectly. Therefore, one possible way to approach the solution is to generate a decoy set with structural diversity, in which each structure can be a representative of a local conformational space well-screened by the given energy function. Our major efforts into this work is not only providing a structural diversity to the generated decoy set using CSA, but also introducing softness given by the statistical pairwise potential into the rigid-body docking, in order to locate at least one near-native structure in the decoy set. The next step that we need to move on is to improve the docking accuracy to select the right near-native structure from the decoy set.

Acknowledgments

We gratefully thank Professor Yaoqi Zhao for providing us the DFIRE1.0 program. This work was supported by the grant No. R01-2003-000-11595-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

REFERENCES

[1] A. H. Elock, D. Sept and J. A. McCammon. Computer simulation of protein-protein interactions. J Phys Chem B, 105(8):1504-1518, 2001.

[2] S. Vajda, I. A. Vakser, M. J. E. Sternberg and J. Janin. Modeling of protein interactions in genomes. Proteins, 47(4):444-446, 2001.

[3] G. R. Smith and M. J. E. Sternberg. Prediction of protein-protein interactions by docking methods. Curr Opin Struct Biol, 12(1):28-35, 2002.

[4] C. J. Camacho and S. Vajda. Protein-protein association kinetics and protein docking. Curr Opin Struct Biol, 12(1):36-40, 2002.

[5] R. Mendez, R. Leplae, L. D. Maria and S. J. Wodak. Assessment of blind predictions of protein-protein interactions: current status of docking methods. Proteins, 52(1):51-67, 2003.

[6] J. Janin and B. Seraphin. Genome-wide studies of protein-protein interaction. Curr Opin Struct Biol, 13(3):383-388, 2003.

[7] R. B. Russell, F. Alber, P. Aloy, F. P. Davis, D. Korkin, M. Pichaud, M. Topf and A. Sali. A Structural perspective on protein-protein interactions. Curr Opin Struct Biol, 14(3):313-324, 2004.

[8] J. Lee, H. A. Scheraga and S. Rackovsky. New optimization method for conformational energy calculations on polypeptides: conformational space annealing. J Comput Chem, 18(9):1222-1232, 1997.

[9] J. Lee and H. A. Scheraga. Conformational space annealing by parallel computations: extensive onformational search of met-enkephalin and of the 20-residue membrane bound portion of melittin. Int J Quant Chem, 75(3):255-265, 1999.

[10] J. Lee, A. Liwo, D. R. Ripoll, J. Pillardy and H. A. Scheraga. Calculation of protein conformation by global optimization of a potential energy function. Proteins Suppl, 3:204-208, 1999.

[11] J. Lee, A. Liwo, D. R. Ripoll, J. Pillardy, J. A. Saunders, K. D. Gibson and H. A. Scheraga. Hierarchical energy-based approach to protein-structure prediction: blind-test evaluation with CASP3 targets. Int J Quant Chem, 77(1):90-117, 2000.

[12] J. Lee, S.-Y. Kim, K. Joo, I. Kim and J. Lee. Prediction of protein tertiary structure using PROFESY, a novel method based on fragment assembly and conformational space annealing. Proteins, 56(4):704-714, 2004.

[13] J. A. Saunders and H. A. Scheraga. Ab initio structure prediction of two-helical oligomers with a multiple-chain united-residue force field and global search. Biopolymers, 68(3):300-317, 2003.

[14] J. A. Saunders and H. A. Scheraga. Challenges in structure prediction of oligomeric proteins at the united-residue level: searching the multiple-chain energy landscape with CSA and CFMC. Biopolymers, 68(3):318-332, 2003.

[15] K. Lee, C. Czaplewski, S.-Y. Kim and J. Lee. An efficient molecular docking using conformational space annealing. J Comput Chem, 26(1):78-87, 2005.

[16] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. M. 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. J Am Chem Soc, 117(19):5179-5197, 1995.

[17] J. Janin, K. Hendrick, J. Moult, L. T. Eyck, M. J. E. Sternberg, S. Vajda, I. Vakser and S. J. Wodak. CAPRI: A Critical Assessment of PRedicted Interactions. Proteins. Proteins, 52(1):2-9, 2003.

[18] K. Lee, J. Sim and J. Lee. Study of protein-protein interaction using conformational space annealing. Proteins, 60(2):257, 2005.

[19] H. Zhou and Y. Zhou. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Proteins Science, 11(11):2714-2726, 2002.

[20] C. Zhang, S. Liu and Y. Zhou. Accurate and efficient loop selections using DFIRE-based all-atom statistical potential. Protein Science, 13(2):391--399, 2004.

[21] R. Chen, J. Minsteris, J. Janin and Z. Weng. A Protein Docking Benchmark. Proteins, 52(1):88-91, 2003.

[22] R. Chen, L. Li and Z. Weng. ZDOCK: An intial-stage protein-docking algorithm. Proteins, 52(1):80-87, 2003.

[23] H. L. Friedman. A course in statistical mechanics. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1985.

[24] U. Hobohm, M. Scharf, R. Schneider and C. Sander. Selection of a representative set of structures from the Brookhaven Protein Data Bank. Protein Science, 1(3):409-417, 1992.

[25] U. Hobohm and C. Sander. Enlarged representative set of protein structures. Protein Science, 3(2):522-524, 1994.