

Research Article

Predicting Protein Interactions by Brownian Dynamics Simulations

Xuan-Yu Meng,^{1,2} Yu Xu,^{1,2} Hong-Xing Zhang,² Mihaly Mezei,³ and Meng Cui¹

¹ *Department of Physiology and Biophysics, Virginia Commonwealth University, 1220 East Broad Street, P.O. Box 980551, Richmond, VA 23298, USA*

² *State Key Laboratory of Theoretical and Computational Chemistry, Institute of Theoretical Chemistry, Jilin University, Changchun 130023, China*

³ *Department of Structural and Chemical Biology, Mount Sinai School of Medicine, New York, NY 10029, USA*

Correspondence should be addressed to Meng Cui, mcui@vcu.edu

Received 15 July 2011; Accepted 19 October 2011

Academic Editor: Sergio Pantano

Copyright © 2012 Xuan-Yu Meng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a newly adapted Brownian-Dynamics (BD)-based protein docking method for predicting native protein complexes. The approach includes global BD conformational sampling, compact complex selection, and local energy minimization. In order to reduce the computational costs for energy evaluations, a shell-based grid force field was developed to represent the receptor protein and solvation effects. The performance of this BD protein docking approach has been evaluated on a test set of 24 crystal protein complexes. Reproduction of experimental structures in the test set indicates the adequate conformational sampling and accurate scoring of this BD protein docking approach. Furthermore, we have developed an approach to account for the flexibility of proteins, which has been successfully applied to reproduce the experimental complex structure from the structure of two unbounded proteins. These results indicate that this adapted BD protein docking approach can be useful for the prediction of protein-protein interactions.

1. Background

Protein-protein interactions are involved in most cellular process. In order to gain a better understanding of the function of a protein, it is necessary to know how it interacts with other proteins at the molecular level. Both experimental and computational methodologies are used to achieve this goal. Experimental techniques, such as X-ray crystallography and nuclear magnetic resonance (NMR), have been used to determine an increasing number of protein structures that are deposited into the Protein Data Bank (PDB). However, determining the structures of protein complexes is more difficult by using these techniques. In the PDB, the number of protein complexes as biologically functional units is relatively small (11,331 (PDB search criteria: chain type: protein; entity type: protein; number of entities >1; October 14, 2011)) compared to the total number of protein structures (70,852) [1, 2]. As for the computational tools, a successful protein docking method can be used to predict the three-

dimensional structures of protein complexes and provide theoretical understanding of protein-protein interactions at the atomic level. The most challenging problems for the computational approaches are extensive enough conformational sampling, accurate scoring, and inclusion of the flexibility of proteins [3, 4].

A variety of search algorithms were developed and implemented in protein docking programs, such as fast Fourier transforms (FFT), geometric matching, and the Monte Carlo technique [3]. FFT-based methods, which are used in ZDOCK, FTDOCK [5, 6], and so forth, use correlations to search the sampling space and evaluate the putative complexes with grid representation [7]. These methods initially assess the geometrical surface complementarity and further expand to include electrostatics and hydrophobic complementarity. The Monte Carlo methods, which are used in RosettaDock, ICM [8, 9], and so forth, conduct stochastic rotation/translation searches from random initial structures around the known or hypothetical binding site

and thus explore only certain regions of the surface [10]. The Brownian dynamics (BD) method, which is similar to the force-biased Monte Carlo approach [11–13], has been used in the past to predict protein-protein interactions [14–20]. In earlier work, we used BD to successfully simulate the recognition between scorpion toxins and potassium channels [21, 22]. Recently, our previous BD prediction of the interaction between scorpion toxin Lq2 and KcsA potassium channel has been verified by the potassium channel-charybdotoxin complex structure (PDB: 2A9H), which was determined by NMR studies [21, 23]. The sequence identity between the KcsA of our model and the experimental structure is 97%. Scorpion toxins Lq2 and charybdotoxin (CTX) share 78% sequence identity, and the RMSD of C_α atoms between Lq2 and CTX is 1.62 Å. The RMSD of our predicted KcsA-Lq2 model and experimental KcsA-CTX structure is 2.85 Å (based on the backbones of Lq2 and CTX only, two KcsA structures are superimposed), which shows excellent agreement between the BD predicted and experimentally solved complex structures. Our BD results indicate that the strong electrostatic interactions between scorpion toxins and potassium channels are the main driving force for their recognition and association.

In the original version of the BD program (MacroDox), only electrostatic interactions between two proteins are included [24]. However, for protein docking simulations it is critical to include the short-range interaction energy terms to rank the final complexes correctly. Accordingly, we have introduced additional local energy minimizations after BD simulations by using a force field, which includes electrostatics, van der Waals, and desolvation energy terms. We used the Coulomb potential with a sigmoidal distance-dependent dielectric permittivity to model the electrostatic energy term, a 12-6 Lennard-Jones potential to model the van der Waals energy term, and the general approach of Wesson and Eisenberg [25] to model the desolvation term. The 12-6 Lennard-Jones potential parameters of four united atom types (C, N, O, and S) were obtained from the AUTODOCK program. The atomic solvation parameters were calculated based on the absolute partial charge on the atom. These energy terms were introduced by the corresponding potential maps. We previously used a similar grid-based force field for protein loop predictions [26].

In order to systematically evaluate the performance of our newly adapted BD program for protein-protein docking simulations, we applied the BD approach on a test set of 24 crystal protein complexes from PDB, which was used by other groups for protein docking approach evaluations [27]. The results from redocking experiments showed that the root mean square deviation (RMSD) between the predicted structures and the crystal structures is within 2 Å. We have further developed the approach to account for the flexibility of proteins, which has been successfully applied to reproduce the experimental complex structure from two unbounded proteins. These results indicate that this adapted BD protein docking approach can be useful for prediction of protein-protein interactions.

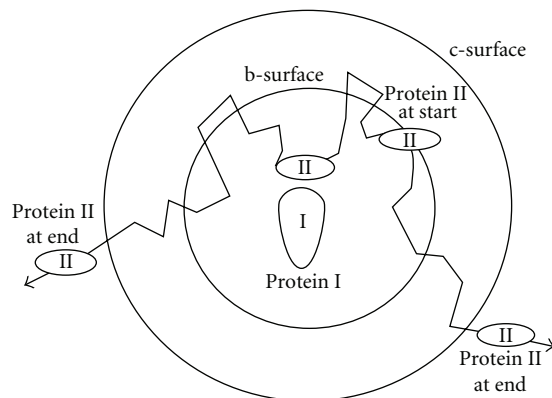


FIGURE 1: A schematic representation of the Brownian dynamics simulations of the association between two proteins. Simulations are conducted in coordinates defined relative to the position of the center of the protein, protein I.

2. Methods

2.1. Global BD Sampling. The program package MacroDox, version 3.2.2 [24], was used to assign charges on proteins, solve the linearized Poisson-Boltzmann equation, and run the BD simulations. The BD algorithm for this program has been detailed by Northrup et al. [28, 29].

For the BD simulations of protein-protein interactions, there are only two solute particles, the receptor and ligand proteins treated as rigid bodies. In this study, the receptor of the complex was defined as protein I and was kept fixed with its center of mass (COM) at the origin while the ligand was defined as protein II of which translational and rotational motions were simulated (Figure 1). In each trajectory, protein II starts with a random orientation and position on the b-surface, which was defined as the sum of the maximum radii of the two proteins plus 2 Å. Protein II was subjected to three forces: electrostatic, the random Brownian force, and the frictional force due to solvent viscosity. During the simulation, the motion of the complex(es) satisfying the reaction criteria for encounter complex formation is retained as an effective trajectory [30]. Within such an effective trajectory, two configurations which have lowest electrostatics interaction energy and shortest distance between protein I and protein II, respectively, are recorded as two independent sampled conformations. The trajectory was terminated when protein II either escaped out of the c-surface (5 Å outside of the b surface) or was running longer than 20 ns.

2.2. Structure Refinement by Local Energy Minimization. Local energy minimizations were conducted for all putative protein complexes obtained from the BD simulations. We developed a shell-based grid force field based on our earlier work [26], which includes van der Waals (vdW), electrostatics, and hydrophobic potentials to represent protein I (united atom types) and solvation effect. Figure 2 displays the two-shell grid force field generated from protein I of 1AY7. The inner shell includes 948,240 grids with 0.5 Å

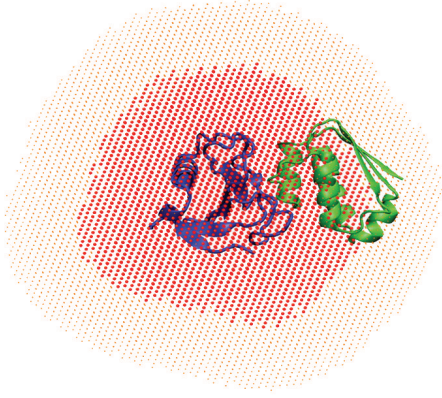


FIGURE 2: The shell-based grid force field of protein I of 1AY7. Protein I is shown in blue and protein II in green. Each point represents a force field grid with a 3 Å resolution. Accordingly, for the inner shell (in red) one point corresponds to 6 grids with 0.5 Å resolution; for the outer shell (in orange) one point corresponds to 4 grids with 0.75 Å resolution.

resolution representing electrostatics, vdW, and hydrophobic potentials, 15 Å width from the surface of protein I. The outer shell includes 962,432 grids with 0.75 Å resolution representing electrostatics potentials, 30 Å width from the surface of protein I.

The van der Waals interactions are described by the 12-6 Lennard-Jones potentials:

$$\phi(r) = \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6}, \quad (1)$$

where the C_{12} and C_6 Lennard-Jones parameters of four atom types (united atoms C, N, O, and S) are from the AUTODOCK program [31]. Since the real vdW potential ("real" stands for the standard vdW potential as defined in (1)) is extremely sensitive to the distance r , we also used a soft potential which replaces the energy value calculated at r by the lowest value calculated from r , $r - 0.5$ Å, and $r + 0.5$ Å using (1). As shown in Figure 3, this soft potential has the effect of smoothing the potential energy surface and thus helping protein II to cross local energy barriers during the energy minimization.

2.2.1. Electrostatics. Coulomb potential $\phi(r)$, with a sigmoidal distance-dependent dielectric function was used to model solvent screening, based on the work of Mehler and Solmajer [32]:

$$\begin{aligned} \phi(r) &= \frac{Q}{4\pi\epsilon(r) \cdot r}, \\ \epsilon(r) &= A + \frac{B}{1 + ke^{-\lambda Br}}, \end{aligned} \quad (2)$$

where Q is partial charge and $B = \epsilon_0 - A$, $\epsilon_0 = 78.4$ (the dielectric constant of bulk water at 25°C), $A = 6.02944$, $\lambda = 0.018733345$, and $k = 213.5782$ are parameters. The original parameter set has been modified to produce better results when comparing with the GB model [33]. When the

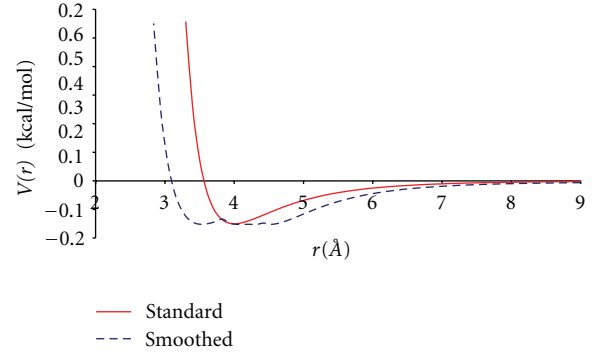


FIGURE 3: Representation of the real (standard) and the soft (smoothed) van der Waals potential used in this work. Soft potential has the effect of widening the region of maximum affinity at ϵ and also reduces the potential energy at $r = 0$ to a finite value.

distance between two charges is <1.32 Å, a dielectric constant of 8 is used [26].

2.2.2. Desolvation. The general approach of Wesson and Eisenberg was used [25], and the atomic solvation parameters were calculated based on the absolute partial charge on the atom:

$$\begin{aligned} \Delta G_{\text{desolv}} &= W_{\text{desolv}} \sum_{i,j} (S_i V_j + S_j V_i) e^{-r_{ij}^2/2\sigma^2}, \\ S_i &= a_i + k |q_i|, \end{aligned} \quad (3)$$

where i is the index of atoms in the ligand, j the index of atoms in the receptor, $W_{\text{desolv}} = 1$, linear regression coefficient or weight for the desolvation free energy term, S_i the solvation term for atom i , a_i atomic solvation parameter, k charge-based atomic solvation parameter, q_i partial atomic charge on atom i , V_i the atomic fragmental volume of atom i , r_{ij} the distance between atom i and atom j (in Å), and σ the Gaussian distance constant = 3.5 Å. The parameters are obtained from the AUTODOCK4 program [34]. Only nonpolar carbon atoms (the absolute value of charge is <0.2) were used for desolvation energy calculations.

The rigid energy minimizations of the protein complexes are based on the downhill simplex method [35] using the newly developed force field [26, 30]. Six variables (three translation values and three rotation values of protein II relative to protein I) are allowed to change during the energy minimizations.

2.3. Accounting for Protein Flexibility by MD Simulations. To account for protein flexibility, we conducted molecular dynamics simulations on each individual protein complex obtained from the BD docking with low interaction energies. The complexes were subjected to 100 steps of the steepest descent (SD) energy minimizations followed by 2ps simulated annealing (SA) from 300 K to 50 K, and additional 100 steps of SD energy minimizations by using the generalized Born (GB) solvation module in the CHARMM program [36]. The optimized complex structures were rescored by the same force field we developed here, but in the explicit form

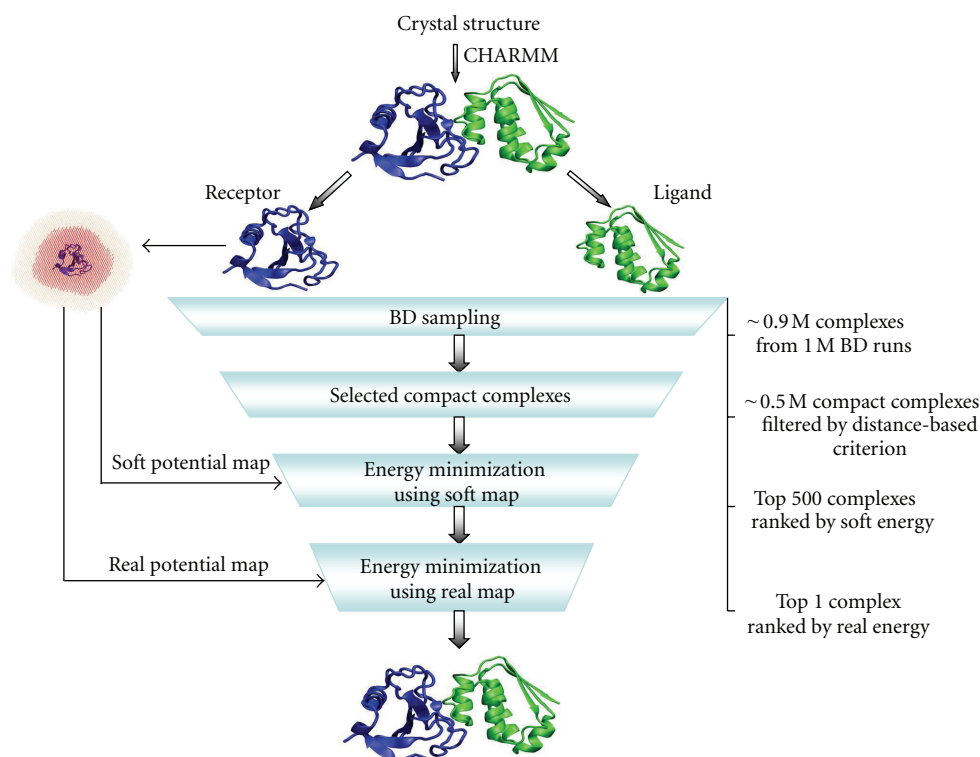


FIGURE 4: Flowchart of the docking protocol used in this work.

(i.e., without resorting to the grids and without smoothing the LJ potential) with cutoff values of 8 Å and 15 Å for van der Waals and electrostatic interactions, respectively. The top 200 low energy complex structures obtained from rigid body energy minimization by soft energy potential were selected to conduct the flexible protein MD simulations in this study.

2.4. Computational Costs. All simulations were conducted on a Sun Linux Beowulf cluster (2.6 GHz Opteron and 1TB RAM (4 GB–32 GB per node)) at the Center for High Performance Computing (CHiPC) at Virginia Commonwealth University. Computational costs are closely related to the size of both the protein receptor and ligand. For BD sampling, one million runs can be completed within 20 h/CPU for this test set (average residue number is about 500 in the protein complexes). For the soft optimization, it usually takes 12 h using 40 CPUs by splitting the trajectory into 40 subtrajectories. For the real optimization, only the top 500 complexes, ranked by soft interaction energies, are selected to be further calculated, which usually takes 1 h per CPU. For each complex the post-MD-simulated annealing refinement process takes about 30 minutes per CPU.

3. Results

3.1. Docking Bound Complexes. To evaluate the performance of our newly developed BD protein docking approach, we first applied it for a redocking experiment on a test set of 24 protein-protein complex crystal structures which was

previously used by several other groups for protein docking method evaluations [27]. The computational protocol of the BD protein docking method including conformational sampling, compact complex selection, and energy minimization is presented in Figure 4. Firstly, the experimental crystal complex structures were subjected to 200 steps of the steepest descent (SD) followed by 2000 steps of the adopted basis Newton-Raphson Method (ABNR) energy minimizations using the generalized born (GB) solvation module in the CHARMM program [36] to remove steric overlap that produces bad contacts. Then BD simulations were conducted to explore the configuration space of the protein ligand relative to the protein receptor by translational and rotational motions to form the protein complexes. The intermolecular forces and torques between proteins were given by the sum of electrostatic and excluded volume forces [24].

For each system, one million independent BD simulations were conducted that usually generated about 0.9 million protein complexes. To reduce computational costs, a distance-based criterion was introduced to filter the complexes from the BD sampling. The protein complexes whose distance between the surface of protein I and the COM of the protein II is smaller than the maximum radius of protein II were retained. This filtering step usually yielded around 0.5 million of these “compact complexes,” which were further optimized using precalculated shell-based grid soft potentials (see Figure 3). Then, the optimized protein complexes were ranked by the soft protein interaction energies. The top 500 complexes with the lowest interaction energies were selected and followed by structural optimization using the shell-based

grid real potentials. The optimized complex structures were reranked and the one, with the lowest interaction energies were selected as the final predicted protein complexes.

The predicted results from the newly developed BD protein docking approach are listed in Table 1. The RMSDs from the corresponding crystal structures of all the redocked complexes are within 2 Å, ranging from 0.18 to 1.98 Å (the RMSDs are based on the C_{α} atoms of the ligands since the receptors are always fixed). With the single exception of 2PCB (see Section 4), in each system studied, the near-native conformation was ranked first based on our scoring function.

To further verify the shell-based grid force field, we also optimized the crystal structures with the real potential map (unsmoothed vdW potential) (Table 1). One can see that the interaction energies of crystal complexes are comparable to those of our redocked complexes. On the other hand, we analyzed the correlation between RMSD and interaction energy for each complex. Taking 1CHO as an example, the plot of interaction energies versus RMSD is shown in Figure 5. After the optimization with the grid-based real potentials the complex with the lowest interaction energy has the smallest RMSD to the crystal structure. Both aspects indicate that this force-field-based scoring function can discriminate between the crystal conformation and other decoy conformations.

3.2. Test BD Protein Docking Approach on Unbound Proteins.

To further evaluate BD docking approach for real protein-protein interactions, we also conducted docking studies using the unbound structures of the two proteins in a complex. Crystal structures for proteins, Trypsin and BPTI, are available in both bound (Trypsin/BPTI, PDB: 2PTC) and unbound (Trypsin, PDB: 5PTP; BPTI, PDB: 1BPI) forms, which provide us an opportunity to further evaluate our BD approach. 90,079 complexes were obtained from BD docking simulations and followed by rigid body local energy minimization by the soft potential. The predicted complex with the lowest RMSD (2.9 Å) from the crystal structure complex is ranked as the 142nd by interaction energy. It suggests that even though the BD approach is able to sample the native configuration of the complex correctly, it cannot be selected based on the energy ranking. It is not surprising since the flexibility of both proteins was not considered during the BD docking. We conducted MD-simulated annealing simulations for the top 200 complexes with the lowest interaction energies obtained from the BD docking simulations and followed by rescoring with the same force field we developed here (see Section 2). The energy ranking for the structure with the lowest RMSD has been significantly improved to the 4th with the RMSD of 3.2 Å from the crystal structure. The optimized complex structure of Trypsin/BPTI superimposed with the crystal complex structure is shown in Figure 6.

4. Discussion

4.1. BD Sampling and Electrostatics Correlation. Sampling of the complex conformation is the critical step in protein-

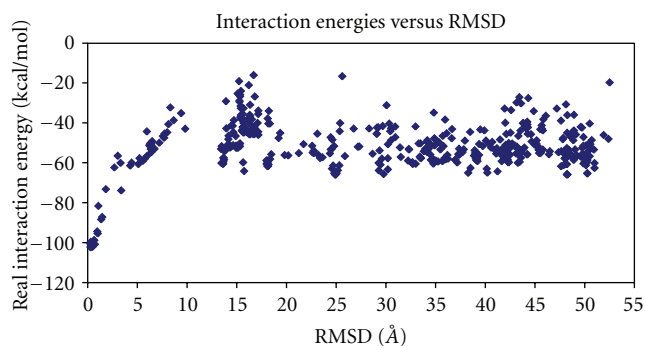


FIGURE 5: Interaction energies versus RMSD of the 500 representative conformations of 1CHO from real potential optimization. The RMSD calculations are based on the C_{α} atoms of the crystal and optimized structures.

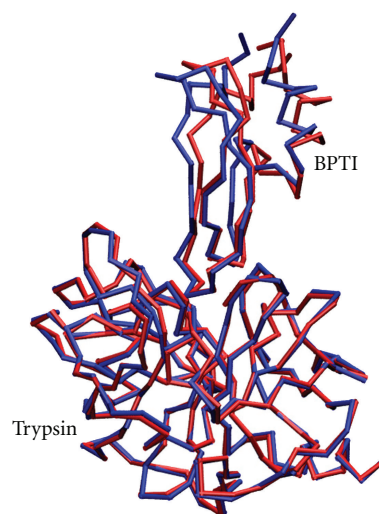


FIGURE 6: The superposition of the predicted Trypsin/BPTI complex (blue) with the crystal complex (red) structures. The RMSD between the predicted and crystal BPTI structures is 3.2 Å.

protein docking process, since neither scoring function evaluation nor following optimization refinement would be enough if correct or near-native conformations cannot be produced by sampling. In our docking procedure, several significant modifications of sampling parameters were made compared to the traditional BD simulations. As described in Section 2 (Figure 1), the radius of b-surface is 2 Å larger than the sum of the maximum radii of the two proteins; c-surface is 5 Å from the b-surface. Both are much smaller than default values in the MacroDox (default values: the radius of b-surface is 20 Å larger than the sum of the maximum radii of the two proteins and the radius of c-surface is 200 Å). Also, 20 ns was set as the longest simulation time for each BD run. This choice of parameters significantly reduced BD simulation time per run making it possible to perform 1 million BD runs for each complex in reasonable computation time.

TABLE 1: Summary of redocking results.

| Complex PDB | Res (Å) | Receptormame | Ligandname | Docking results | | Crystal structures | |
|--------------------|---------|----------------------|---------------|-----------------------|--|--------------------|--|
| | | | | RMSD ^a (Å) | Interaction energy (kcal mol ⁻¹) | RMSD (Å) | Interaction energy (kcal mol ⁻¹) |
| Protease-inhibitor | | | | | | | |
| 1CA0 | 2.10 | Chymotrypsin | APPI | 1.27 | -93.70 | 0.84 | -93.22 |
| 1CBW | 2.60 | Chymotrypsin | BPTI | 0.54 | -83.51 | 0.19 | -85.02 |
| 1ACB | 2.00 | Chymotrypsin | Eglin C | 1.00 | -103.41 | 0.70 | -103.49 |
| 1CHO | 1.80 | Chymotrypsin | Ovomuciod | 0.30 | -102.35 | 0.38 | -102.42 |
| 1CGI | 2.30 | Chymotrypsinogen | HPTI | 0.18 | -147.17 | 0.15 | -147.10 |
| 2KAI | 2.50 | Kallikrein A | BPTI | 1.10 | -114.68 | 0.21 | -113.85 |
| 2SNI | 2.10 | Subtilisin BPN | CI-2 | 0.27 | -108.81 | 0.28 | -108.86 |
| 2SIC | 1.80 | Subtilisin BPN | SSI | 0.89 | -94.41 | 0.20 | -104.20 |
| 1CSE | 1.20 | Subtilisin Carlsberg | Eglin C | 1.24 | -98.28 | 0.088 | -103.19 |
| 2TEC | 1.98 | Thermitase | Eglin C | 0.44 | -108.55 | 0.68 | -109.99 |
| 1TAW | 1.80 | Trypsin (bovine) | APPI | 1.10 | -97.13 | 0.86 | -97.14 |
| 2PTC | 1.90 | Trypsin (bovine) | BPTI | 0.98 | -96.22 | 0.36 | -96.25 |
| 3TGI | 1.80 | Trypsin (rat) | BPTI | 0.39 | -102.43 | 0.52 | -102.44 |
| 1BRC | 2.50 | Trypsin (rat) | APPI | 1.43 | -90.55 | 0.55 | -90.04 |
| Enzyme-inhibitor | | | | | | | |
| 1FSS | 3.00 | Acetylcholinesterase | Fasciculin II | 0.17 | -137.88 | 0.20 | -137.88 |
| 1BVN | 2.50 | α -Amylase | Tendamistat | 0.25 | -142.67 | 0.24 | -142.51 |
| 1BGS | 2.60 | Barnase | Barstar | 0.48 | -112.46 | 0.57 | -112.37 |
| 1AY7 | 1.70 | Ribonuclease sa | Barstar | 0.49 | -77.41 | 0.51 | -77.37 |
| 2B5R | 1.70 | TEM-1 lactamase | BLIP | 0.78 | -154.04 | 0.37 | -158.39 |
| 1UGH | 1.90 | UDG | UGI | 0.54 | -135.58 | 0.51 | -135.43 |
| Electron transport | | | | | | | |
| 2PCB ^c | 2.80 | Cyt c Peroxidase | Cytochrome c | 1.98 | -87.18 | 0.22 | -81.79 |
| 2PCF | NMR | Cytochrome f | Plastocyanin | 0.28 | -118.96 | 0.23 | -119.32 |
| Antibody-antigen | | | | | | | |
| 1MLC | 2.10 | Fab D44.1 | Lysozyme | 1.59 | -93.99 | 0.44 | -95.80 |
| 1VFB ^d | 1.80 | Fv D1.3 | Lysozyme | 0.53 | -84.78 | 0.43 | -84.59 |

^aRMSDs are calculated for the C α atoms of the ligand protein since the receptor proteins are always fixed during the simulations. ^bSequence number denotes which sequence is the predicted conformation belonging to the BD sampling. Each conformation has unique sequence number as a label. Total number denotes how many compact complexes were obtained. From this column one can see where the final predicted complex(es) come from and how many compact complexes can be obtained from one million BD runs. These data are collected to demonstrate the necessary and sufficient condition for one million BD runs. ^cThe conformation presented here is the 9th conformation by energy ranking. The top 8 conformations are regarded as one cluster RMSD around 30 Å from the crystal structure (see Section 4). ^dThree million BD runs were conducted for this complex. Usually one million is enough for most complexes; however, due to the positive electrostatics 1.61 kcal mol⁻¹ for this complex, the sampled near-native conformation appeared at around 1.4 million runs.

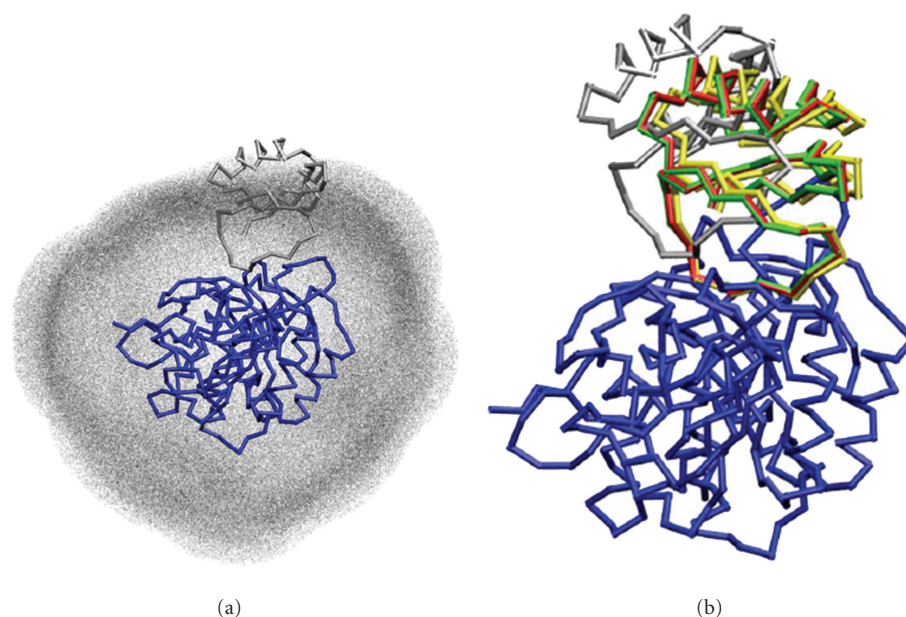


FIGURE 7: BD sampling and structure refinement of 2SNI. (a) 428,140 compact complexes obtained from BD global sampling. Each gray point represents the center of mass (COM) of one sampled conformation. The receptor of crystal complex is in blue, and near-native conformation obtained after BD sampling is in silver. (b) Representation of near-native conformations obtained from BD sampling (silver, RMSD 11.47 Å), soft potential optimization (yellow, RMSD 2.45 Å) and real potential optimization (red, RMSD 0.27 Å). The receptor and the ligand of crystal complex are in blue and green, respectively.

BD approach has been used to predict the protein-protein interactions using electrostatics forces as the biasing guide in addition to Brownian stochastic motions. Adding electrostatics forces to the docking/diffusion process has the advantage of speeding up the encounter of the receptor and the complex formation. When the attractive electrostatics is the dominant interaction at the binding site, it can even produce the near-native conformations with only a few thousand BD runs [21, 22, 37, 38]. On the other hand, when the vdW interaction or repulsive electrostatics at the binding site is the dominant force, one million BD runs would be a reasonable number based on our calculations. It could cover the entire sampling space around the receptor (Figure 7(a)). From the column of Sequence number/total number in Table 1 one can see where the best solution comes from. It also indicates that one million BD runs are sufficient for sampling in most cases. One exception is 1VFB where three million BD runs were needed due to the repulsive electrostatics interaction of $1.61 \text{ kcal mol}^{-1}$ between the receptor and ligand based on our calculations.

Cluster analysis was also conducted to illustrate the electrostatics effect on BD sampling within this test set. As shown in Figures 8(a) and 8(b), the more negative (stronger) the electrostatics interaction between the receptor and ligand is, the more of the conformations will fall into the cluster of the near native; and the near-native conformations are produced increasingly early as well. Generally when the electrostatic interaction energy is lower than -5 kcal mol^{-1} , the BD sampling procedure produces the near-native conformations faster and more easily than in other cases.

4.2. Shell-Based Grid Force Field. The grid potential maps generated from the receptor were defined in a shell around the surface of the molecule (see Section 2, Figure 2). This shell-based grid map has the advantage of largely reducing the number of grids compared to the 3D box grid map. Taking 1AY7 as an example, the box-shape potential map would include 1,739,781 and 1,737,808 grids for the inner and outer boxes, respectively; while in the same condition the shell-based map only takes half the number of grids, 948,240 and 962,432 for the inner and outer shells, respectively. In practice, the computer memory limits the number of grids involved in the calculation. For example, in earlier work, only part of the protein surface could be covered for sampling with the box-shape map [27]. Therefore, the binding site had to be identified with the aid of other programs or experimental data. Our shell-based potential map is able to cover the whole surface area of the receptor for conducting a global search, which is necessary when clear information is unavailable for the binding sites of protein receptors.

We used a force-field-based scoring function, which includes vdW, electrostatics, and hydrophobic potentials. We found that when the smoothed vdW energy term (soft potential) was used for the energy optimization, it was much easier for a ligand to find the global energy minima due to its smooth potential energy surface. As seen from Figure 7(b), the RMSD of the ligands between the BD sampling and the crystal structure was 11.5 Å , which was reduced to 2.5 Å after the energy minimization by using the soft potential. Moreover, although the interaction energies obtained from the soft potential function are not as accurate as the real

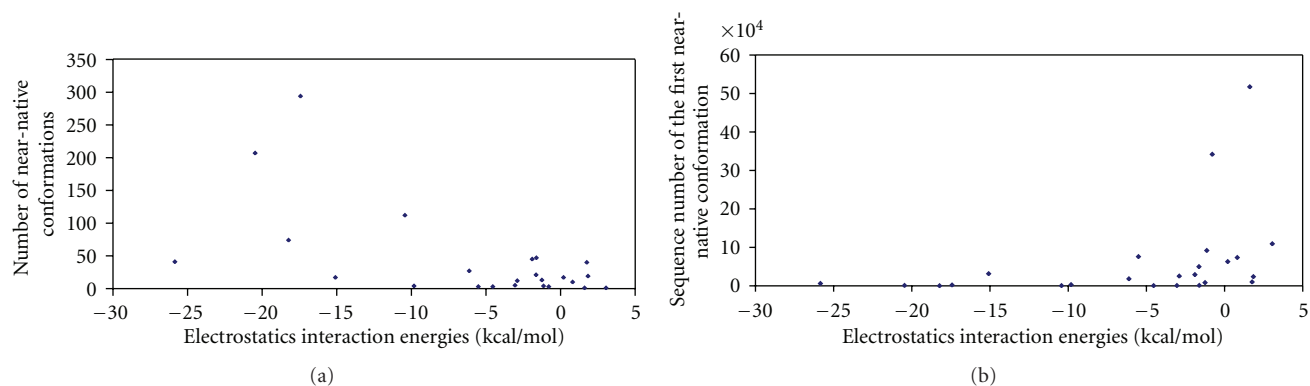


FIGURE 8: (a) Number of predicted near-native conformations versus electrostatics interaction energies within the test set. Conformations of which C_{α} RMSDs $< 5 \text{ \AA}$ compared to the crystal structure are regarded as predicted near-native conformations. (b) Sequence number of the first produced near-native conformations versus electrostatics interaction energies within the test set.

ones, they can still be used for preranking of the predicted complexes. The top 500 preranked complexes were selected and followed by a further energy minimization using the real potential function, which was found to be adequate and effective for the predictions of native protein complexes.

The scoring function exhibits excellent performance in the redocking test set. Only exception was in finding the near-native conformation for the 2PCB. It was ranked the 9th, but the top 8 conformations formed a single cluster with RMSD around 30 \AA from the crystal structure. A similar problem existed in Fernández-Recio's result [27]. We have checked the crystal structure of 2PCB and found an abnormal phenomenon that OD2 of Asp34 of Cyt c Peroxidase and OE2 of Glu90 of Cytochrome c is apart only by 2 \AA . It was speculated that occasional crosslinking between these two proteins might account for this phenomenon [39]. Such a short distance between two negatively charged oxygen atoms may lead to higher interaction energy, and thus other decoy conformations appear to have more favorable interaction energy.

In addition to the RMSDs, which were calculated based on all the C_{α} atoms of ligands, we also calculated the RMSDs for the ligands based on the interface C_{α} atoms of the predicted structures and compared them with the results from other groups (Table 2). The ligand interface residues were defined as having at least one atom within 5 \AA of an atom on the receptor molecule. RMSDs based on interface C_{α} atoms in this test set are within 1 \AA except for 2PCB (1.18 \AA). 18 of 24 complex RMSDs are smaller than those from the ICM method [27].

4.3. Crystal Packing. Considering that protein structures in the test set were determined by X-ray crystallography, it is possible that the protein complexes could be affected by crystal packing. We reconstructed the crystal packing protein structures for all the proteins in the test set by using the Swiss PDB Viewer program [42]. We found that in 12 out of the 24 crystal structures, ligands were located at the interface of multiple subunits in the crystal packing, and thus native conformations could be affected by the crystal packing. To test this possibility, we conducted a comparison study in

which crystal packing information was included by adding additional related subunits. However, 2 complexes out of 12 packing structures had to be excluded because the additional subunits block the entrances of protein ligands to access the binding sites of protein receptors. Thus, 10 complexes were completely redocked under the condition of crystal packing. The results (Table 3) show that RMSDs of global C_{α} atoms range from 0.26 to 1.44 \AA . Comparing the results without crystal packing, there are no significant differences in RMSDs, which indicates that the crystal packing could be ignored for protein docking method evaluations based on the current test set.

4.4. Prediction of Unbound Proteins. When applying BD sampling approach to predict the complex structure starting from the structures of the unbound proteins Trypsin and BPTI, we found that BD approach was able to sample the near-native conformations, which indicates that the current BD approach is efficient and feasible for real protein complex sampling. However, the near-native poses often have unfavorable interaction energies due to minor steric clashes and therefore cannot be selected based on the energy ranking. This is almost an inevitable consequence arising from rigid body docking process. Although the soft potential used in the current BD approach includes implicitly some partial protein flexibility, it is still not enough to rank the sampled conformations correctly. To achieve high-accuracy prediction, protein flexibility, including flexible side-chains and backbones, has to be considered in either the sampling stage or the postrefinement process. Our new approach based on MD-simulated annealing is a postrefinement process, which accounts for the flexibility of both proteins, and shows very promising results.

Prediction of a correct protein-protein complex relies on sampling the near-native conformation and discriminating it from other decoy conformations. Our BD sampling is validated to sample the near-native conformation within 1 million runs in most cases and works best for the electrostatic attractive protein complexes. We recommend performing 3 million runs for electrostatically unfavorable systems or for any system where electrostatic information is unavailable.

TABLE 2: Redocking results compared with other groups.

| Complex PDB | RMSD (Å) ^a | ICM ^b | Other docking methods | | |
|--------------------|-----------------------|------------------|-----------------------|---------------------|--------------------|
| | | | Nussinov ^c | FTDOCK ^d | BiGER ^e |
| Protease-inhibitor | | | | | |
| 1CA0 | 0.44 | 0.4 | — | — | — |
| 1CBW | 0.24 | 0.5 | — | — | — |
| 1ACB | 0.58 | 0.5 | 0.9 | — | 0.6 |
| 1CHO | 0.26 | 0.3 | 0.5 | 0.8 | — |
| 1CGI | 0.15 | 0.4 | — | 1.0 | — |
| 2KAI | 0.30 | 0.8 | 1.2 | 0.4 | — |
| 2SNI | 0.16 | 0.3 | 1.1 | 0.6 | — |
| 2SIC | 0.58 | 0.4 | 1.1 | 0.8 | 3.8 |
| 1CSE | 0.53 | 0.3 | 1.3 | — | — |
| 2TEC | 0.18 | 0.3 | 1.2 | — | 3.6 |
| 1TAW | 0.39 | 0.7 | — | — | — |
| 2PTC | 0.72 | 0.4 | 0.6 | 0.7 | — |
| 3TGI | 0.21 | 0.3 | — | — | — |
| 1BRC | 0.55 | 0.7 | — | — | — |
| Enzyme-inhibitor | | | | | |
| 1FSS | 0.13 | 0.4 | — | — | — |
| 1BVN | 0.23 | 0.4 | — | — | — |
| 1BGS | 0.33 | 0.6 | — | — | — |
| 1AY7 | 0.30 | 0.7 | — | — | — |
| 2B5R | 0.60 | 1.3 | — | — | — |
| 1UGH | 0.39 | 0.4 | — | — | — |
| Electron transport | | | | | |
| 2PCB | 1.18 | 1.2 | — | — | — |
| 2PCF | 0.19 | 1.1 | — | — | — |
| Antibody-antigen | | | | | |
| 1MLC | 0.57 | 0.4 | — | 0.8 | — |
| 1VFB | 0.24 | 0.5 | 1.5 | 0.7 | — |

^aRMSDs are calculated for the ligand interface C_α atoms in this work. ^bDocking with known binding site followed by refinement of interface side-chain conformations [27]; RMSD (Å) calculated for the ligand interface C_α atoms. ^cGlobal docking followed by hydrophobicity and connectivity filters [40]; RMSD (Å) calculated for the ligand heavy atoms when only receptor is superimposed onto the real structure. ^dGlobal docking and filtering with distance restraints [6]; RMSD (Å) when both receptor and ligand C_α atoms are superimposed onto the real complex. ^eGlobal docking [41]; RMSD (Å) calculated for the ligand C_α atoms when only the receptor C_α atoms are superimposed onto the real structure.

TABLE 3: Comparison between docking with crystal packing and without crystal packing.

| Complex PDB | Crystal packing | | Without packing | |
|-------------|-----------------|--|-----------------|--|
| | RMSD* (Å) | Interaction energy (kcal mol ⁻¹) | RMSD* (Å) | Interaction energy (kcal mol ⁻¹) |
| 1CBW | 0.72 | −161.60 | 0.54 | −83.51 |
| 1ACB | 0.49 | −131.61 | 1.00 | −103.41 |
| 1CHO | 0.23 | −140.39 | 0.30 | −102.35 |
| 1CGI | 1.07 | −138.82 | 0.18 | −147.17 |
| 2KAI | 0.98 | −140.16 | 1.10 | −114.68 |
| 2SNI | 0.58 | −120.52 | 0.27 | −108.81 |
| 2SIC | 1.44 | −103.29 | 0.89 | −94.41 |
| 2TEC | 0.76 | −127.74 | 0.44 | −108.55 |
| 1TAW | 0.44 | −103.90 | 1.10 | −97.13 |
| 1FSS | 0.26 | −175.90 | 0.17 | −137.88 |

* RMSDs are calculated for the C_α atoms of the ligand protein.

Another critical step in our study is using MD-simulated annealing method to account the protein flexibility, which can significantly improve the ranking of near native conformations. In a real case of protein-protein study, the top 500 or more conformations from the soft energy ranking list may need to be included for flexible protein refinement by the MD-simulated annealing simulations.

5. Conclusions

We have developed an adapted Brownian dynamics method to predict the structures of protein complexes. It has three steps. In the first step, global BD sampling, one million independent BD simulations are conducted to explore the entire possible conformational spaces of the protein complexes. In step two, the compact complexes selection, a distance-based criterion is used to select compact complexes. In step three, local energy minimization, all compact complexes are optimized using grid-based soft potentials followed by grid-based real potentials. The complexes with the lowest interaction energies are selected as final predicted protein complexes. To reduce the computational costs for energy evaluations, a grid-based force field was developed to represent protein receptors and solvation effect. The performance of this newly developed BD protein docking approach was evaluated on a test set of 24 crystal protein complexes which was previously used by other groups. The results show that the RMSDs between the predicted and the crystal structures are within 2 Å, which are close to the accuracy of the ICM results from Fernández-Recio et al. [9]. However, compared with the ICM approach, the advantage of our newly adapted BD method is its global sampling, which does not require knowledge of the binding sites of the protein receptors before conducting the BD simulations. Shell-based potential maps can significantly reduce the computer memory requirements compared to rectangular box-based potential maps, which make the global sampling possible. The predicted near-native conformations have the lowest interaction energies for all complexes in the test set except the 2PCB. The results indicate that the grid-based force field scoring function we developed can discriminate the crystal conformation from other sampled conformations. One million BD runs are usually required for the global sampling. However, in the case of positive electrostatics interaction between the protein receptor and ligand, more BD runs are needed. On the other hand, BD sampling has a distinct advantage when stronger electrostatic attraction ($< -5 \text{ kcal mol}^{-1}$) exists between the ligand and receptor proteins.

In the current adapted BD protein docking method, both the protein receptor and ligand are treated as rigid bodies. The flexibility during docking is partially considered by using the grid-based soft potentials. We found that a postrefinement process based on the MD-simulated annealing to account for the flexibility of proteins shows us very promising results and can be also useful if combined with other rigid docking approaches to account for protein conformational changes after the rigid docking process. In our future work, more elegant techniques, such as local move

Monte Carlo (LMMC) approach for loop sampling, [26] could also be implemented to account for the flexibility of proteins for mimicking the ligand-induced effect.

Acknowledgments

The authors thank Professor S. H. Northrup for his permission to adapt the MacroDox3.2.2 program and for his helpful discussions. The computations were supported by the Center for High Performance Computing (CHiPC) at Virginia Commonwealth University (VCU). This work was supported by National Institute Health Grants (DC008996 and S10RR027411 to M. Cui) and Teragrid Supercomputing Grant (TG-MCB090019 to M. Cui).

References

- [1] S. Dutta and H. M. Berman, "Large macromolecular complexes in the protein data bank: a status report," *Structure*, vol. 13, no. 3, pp. 381–388, 2005.
- [2] H. M. Berman, T. Battistuz, T. N. Bhat et al., "The protein data bank," *Acta Crystallographica Section D*, vol. 58, no. 6 I, pp. 899–907, 2002.
- [3] D. W. Ritchie, "Recent progress and future directions in protein-protein docking," *Current Protein and Peptide Science*, vol. 9, no. 1, pp. 1–15, 2008.
- [4] R. Méndez, R. Lepae, M. F. Lensink, and S. J. Wodak, "Assessment of CAPRI predictions in Rounds 3-5 shows progress in docking procedures," *Proteins: Structure, Function and Genetics*, vol. 60, no. 2, pp. 150–169, 2005.
- [5] R. Chen, L. Li, and Z. Weng, "ZDOCK: an initial-stage protein-docking algorithm," *Proteins: Structure, Function and Genetics*, vol. 52, no. 1, pp. 80–87, 2003.
- [6] H. A. Gabb, R. M. Jackson, and M. J. E. Sternberg, "Modelling protein docking using shape complementarity, electrostatics and biochemical information," *Journal of Molecular Biology*, vol. 272, no. 1, pp. 106–120, 1997.
- [7] M. Eisenstein and E. Katchalski-Katzir, "On proteins, grids, correlations, and docking," *Comptes Rendus Biologies*, vol. 327, no. 5, pp. 409–420, 2004.
- [8] J. J. Gray, S. Moughon, C. Wang et al., "Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations," *Journal of Molecular Biology*, vol. 331, no. 1, pp. 281–299, 2003.
- [9] J. Fernández-Recio, M. Totrov, and R. Abagyan, "ICM-DISCO docking by global energy optimization with fully flexible side-chains," *Proteins: Structure, Function and Genetics*, vol. 52, no. 1, pp. 113–117, 2003.
- [10] S. Vajda and D. Kozakov, "Convergence and combination of methods in protein-protein docking," *Current Opinion in Structural Biology*, vol. 19, no. 2, pp. 164–170, 2009.
- [11] C. S. Pangali, M. Rao, and B. J. Berne, "Monte Carlo simulation of water," *Abstracts of Papers of the American Chemical Society*, vol. 175, p. 1, 1978.
- [12] C. S. Pangali, M. Rao, and B. J. Berne, "Determination of potential of mean force of 2 noble-gas atoms dissolved in water," *Abstracts of Papers of the American Chemical Society*, vol. 175, p. 4, 1978.
- [13] P. J. Rossky, J. D. Doll, and H. L. Friedman, "Brownian dynamics as smart Monte Carlo simulation," *The Journal of Chemical Physics*, vol. 69, no. 10, pp. 4628–4633, 1978.

- [14] I. V. Ouporov, H. R. Knull, and K. A. Thomasson, "Brownian dynamics simulations of interactions between aldolase and G- or F-actin," *Biophysical Journal*, vol. 76, no. 1 I, pp. 17–27, 1999.
- [15] D. C. Pearson and E. L. Gross, "Brownian dynamics study of the interaction between plastocyanin and cytochrome f," *Biophysical Journal*, vol. 75, no. 6, pp. 2698–2711, 1998.
- [16] S. L. Lowe, C. Adrian, I. V. Ouporov, V. F. Waingeh, and K. A. Thomasson, "Brownian dynamics simulations of glycolytic enzyme subsets with F-Actin," *Biopolymers*, vol. 70, no. 4, pp. 456–470, 2003.
- [17] D. Motiejunas, R. Gabdoulline, T. Wang et al., "Protein-protein docking by simulating the process of association subject to biochemical constraints," *Proteins: Structure, Function and Genetics*, vol. 71, no. 4, pp. 1955–1969, 2008.
- [18] I. V. Ouporov, H. R. Knull, A. Huber, and K. A. Thomasson, "Brownian dynamics simulations of aldolase binding glyceraldehyde 3-phosphate dehydrogenase and the possibility of substrate channeling," *Biophysical Journal*, vol. 80, no. 6, pp. 2527–2535, 2001.
- [19] A. H. Elcock, D. Sept, and J. A. McCammon, "Computer simulation of protein-protein interactions," *Journal of Physical Chemistry B*, vol. 105, no. 8, pp. 1504–1508, 2001.
- [20] E. J. Haddadian and E. L. Gross, "Brownian dynamics study of cytochrome f interactions with cytochrome c 6 and plastocyanin in *Chlamydomonas reinhardtii* plastocyanin, and cytochrome c6 mutants," *Biophysical Journal*, vol. 88, no. 3, pp. 2323–2339, 2005.
- [21] M. Cui, J. Shen, J. M. Briggs et al., "Brownian dynamics simulations of interaction between scorpion toxin Lq2 and potassium ion channel," *Biophysical Journal*, vol. 80, no. 4, pp. 1659–1669, 2001.
- [22] M. Cui, J. Shen, J. M. Briggs et al., "Brownian dynamics simulations of the recognition of the scorpion toxin P05 with the small-conductance calcium-activated potassium channels," *Journal of Molecular Biology*, vol. 318, no. 2, pp. 417–428, 2002.
- [23] L. Yu, C. Sun, D. Song et al., "Nuclear magnetic resonance structural studies of a potassium channel-charybdotoxin complex," *Biochemistry*, vol. 44, no. 48, pp. 15834–15841, 2005.
- [24] S. H. Northrup, T. Laughner, and G. Stevenson, "Macrodox macromolecular simulation program," 1999.
- [25] L. Wesson and D. Eisenberg, "Atomic solvation parameters applied to molecular dynamics of proteins in solution," *Protein Science*, vol. 1, no. 2, pp. 227–235, 1992.
- [26] M. Cui, M. Mezei, and R. Osman, "Prediction of protein loop structures using a local move Monte Carlo approach and a grid-based force field," *Protein Engineering, Design and Selection*, vol. 21, no. 12, pp. 729–735, 2008.
- [27] J. Fernández-Recio, M. Totrov, and R. Abagyan, "Soft protein-protein docking in internal coordinates," *Protein Science*, vol. 11, no. 2, pp. 280–291, 2002.
- [28] S. H. Northrup, J. O. Boles, and J. C. L. Reynolds, "Electrostatic effects in the Brownian Dynamics of association and orientation of heme proteins," *Journal of Physical Chemistry*, vol. 91, no. 23, pp. 5991–5998, 1987.
- [29] S. H. Northrup, K. A. Thomasson, C. M. Miller et al., "Effects of charged amino acid mutations on the bimolecular kinetics of reduction of yeast iso-1-ferricytochrome c by bovine ferrocycytochrome b5," *Biochemistry*, vol. 32, no. 26, pp. 6613–6623, 1993.
- [30] M. Cui, M. Mezei, and R. Osman, "Modeling dimerizations of transmembrane proteins using Brownian dynamics simulations," *Journal of Computer-Aided Molecular Design*, vol. 22, no. 8, pp. 553–561, 2008.
- [31] G. M. Morris, D. S. Goodsell, R. S. Halliday et al., "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *Journal of Computational Chemistry*, vol. 19, no. 14, pp. 1639–1662, 1998.
- [32] E. L. Mehler and T. Solmajer, "Electrostatic effects in proteins: comparison of dielectric and charge models," *Protein Engineering*, vol. 4, no. 8, pp. 903–910, 1991.
- [33] V. Tsui and D. A. Case, "Theory and applications of the Generalized Born solvation model in macromolecular simulations," *Biopolymers*, vol. 56, no. 4, pp. 275–291, 2000.
- [34] R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell, "Software news and update a semiempirical free energy force field with charge-based desolvation," *Journal of Computational Chemistry*, vol. 28, no. 6, pp. 1145–1152, 2007.
- [35] J. A. Nelder and R. Mead, "A simplex-method for function minimization," *Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [36] W. Im, M. S. Lee, and C. L. Brooks, "Generalized born Model with a simple smoothing function," *Journal of Computational Chemistry*, vol. 24, no. 14, pp. 1691–1702, 2003.
- [37] W. Fu, M. Cui, J. M. Briggs et al., "Brownian dynamics simulations of the recognition of the scorpion toxin maurotoxin with the voltage-gated potassium ion channels," *Biophysical Journal*, vol. 83, no. 5, pp. 2370–2385, 2002.
- [38] X. Huang, H. Liu, M. Cui et al., "Simulating the interactions of toxins with K+ channels," *Current Pharmaceutical Design*, vol. 10, no. 9, pp. 1057–1067, 2004.
- [39] H. Pelletier and J. Kraut, "Crystal structure of a complex between electron transfer partners, cytochrome c peroxidase and cytochrome c," *Science*, vol. 258, no. 5089, pp. 1748–1755, 1992.
- [40] R. Norel, D. Petrey, H. J. Wolfson, and R. Nussinov, "Examination of shape complementarity in docking of unbound proteins," *Proteins: Structure, Function and Genetics*, vol. 36, no. 3, pp. 307–317, 1999.
- [41] P. N. Palma, L. Krippahl, J. E. Wampler, and J. J. G. Moura, "Bigger: a new (soft) docking algorithm for predicting protein interactions," *Proteins: Structure, Function and Genetics*, vol. 39, no. 4, pp. 372–384, 2000.
- [42] N. Guex and M. C. Peitsch, "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling," *Electrophoresis*, vol. 18, no. 15, pp. 2714–2723, 1997.