# Binding Free Energy of Protein-Ligand by Combining Docking and MD Simulation: A Comparison of Calculation Methods

MEIDY TRIANA PAKPAHAN[a,b], HIROAKI SAITO[a], KAZUTOMO KAWAGUCHI[a], HIDEMI NAGAO[a]

[a]Institute of Science and Engineering, Kanazawa University, Kakuma, Kanazawa 920-1192 Japan,
E-mail: meidy@wriron1.s.kanazawa-u.ac.jp, saito@wriron1.s.kanazawa-u.ac.jp,
kkawa@wriron1.s.kanazawa-u.ac.jp, nagao@wriron1.s.kanazawa-u.ac.jp
[b]Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jl. Ganesha 10,
Bandung 40132 Indonesia

**Abstract.** *Accurate methods of computing the affinity of ligand with protein target are strongly needed in the drug discovery process. Many attempts have been made and several algorithms have been developed for this purpose. We compared the protein-ligand binding free energies (ΔG) in various methods include docking score function, combining docking score function and molecular dynamics (MD) simulation with explicit and implicit solvent model, and molecular-mechanics Poisson Boltzmann surface area (MM-PBSA) approach with and without the inclusion of entropic contributions. We tested these various methods to human plasminogen kringle-3 domain protein with the ligand trans-(aminomethyl) cyclohexanecarboxylic acid (AMCHA). The results showed the comparison between these various methods and the experimental affinity value. We found that combining docking score function and MD simulation with explicit solvent model was more favorable and close to the experimental result. This indicated that combining docking score function and MD simulation with explicit solvent model could be more accurate and effective in the protein-ligand binding free energy calculation.*

**Keywords:** binding free energy, drug discovery, protein-ligand interactions, docking, molecular dynamics, human plasminogen, mmpbsa.

## 1 Introduction

Docking is a term used for computational schemes that attempt to find the "best" matching between two molecules: a receptor and a ligand [1]. In particular, protein-ligand docking occupies a very special place in the general field of docking, because of its application in medicine. Accurate methods of computing the affinity of ligand with protein target are strongly needed. Many attempts have been made and several algorithms have been developed for this purpose. Calculating the protein-ligand binding free energy ($\Delta G_{bind}$) is one of the methods to evaluate the affinity of ligand. Binding free energy is the free energy differences between the bound and unbound states of protein and ligand (see Fig. 1).

In order to estimate the binding free energy, several methods have been widely used such as docking score function and molecular dynamics (MD)-based computational techniques, for instance, the thermodynamic integration (TI), free energy perturbation (FEP), linear interaction energy (LIE), and molecular mechanics-Poisson-Boltzmann/Generalized Born and surface area (MM-PBSA/MM-GBSA) methods [2]. Docking score function is a fast approximation to find the correct conformation of a ligand and its receptor but performs rigid receptor-flexible ligand docking, assumes that the protein is a rigid object and attempts to dock the ligand to it. The water molecules are treated implicitly; hence, this method is not completely reliable. On the other hand, MD simulation can treat both protein and ligand in a flexible manner, allowing the relaxation of the structure of binding site

around the ligand, moreover the water molecules are treated explicitly. However, the main problem with the MD simulation is time consuming.

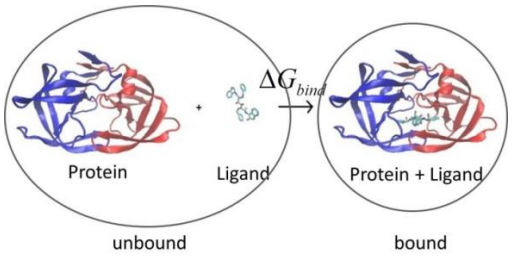


Figure 1: Binding free energy ($\Delta G_{bind}$) of protein and ligand.

To solve the problem of docking and MD simulation, we proposed an approach to combine docking score function with MD simulation. Previous research showed that the value of scoring function has been improved by using this approach with explicit solvent model [3]. Further, in this study, we want to compare the accuracy of our approach with MM-PBSA/MM-GBSA methods. MM-PBSA/MM-GBSA combines molecular mechanics and continuum/implicit solvent approaches to estimate the free energy of binding.

## 2 Materials

Human plasminogen is about 810 amino acid residues long containing different conserved domains such as ligand binding sites, putative domain interaction sites, active sites and cleavage sites [4]. In this study, we applied our approach to human plasminogen (hPgn) kringle-3 (K3) domain and ligand trans-(aminomethyl) cyclohexanecarboxylic acid (AMCHA). hPgn K3 domain contains 83 residues. We retrieved this NMR solution structure from the Protein Data Bank (PDB ID: *2L0S*). It has been shown that lysine-binding activity can be engineered via a Lys57→Asp mutation. The affinity of r(K57D)K3 for the lysine analogue trans-(aminomethyl)-cyclohexane carboxylic acid (AMCHA) was investigated from ligand-induced NMR chemical shift perturbations, which enabled for mapping the binding site on the mutated domain surface. Homology modeling combined with *in silico* docking of lysine-like zwitterionic ligands via AutoDock 4.0 supports functionality of the engineered (K57D)K3 LBS, whose electrostatic focal centers are defined by the Arg36/Arg71 cationic, Asp55/Asp57 anionic pairs and hydrophobic Trp62/Trp72 pairs. Experimental result has shown that the binding affinity of ligand AMCHA to hPgn K3-domain is -5.08±0.01 kcal/mol [5].

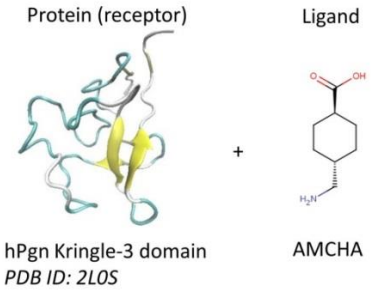


Figure 2: Human plasminogen (hPgn) kringle-3 (K3) domain as receptor, with ligand trans-(aminomethyl)-cyclohexane carboxylic acid (AMCHA).

# 3 Methods

## 3.1 Docking Score Function

Docking processes were performed using Autodock Vina [6], calculation of binding free energy is as follow,

$$\Delta G_{bind} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors} \quad (1)$$

where $\Delta G_{vdW}$ refers to the 12-6 Lennard-Jones potential, $\Delta G_{elec}$ refers to the Coulombic with Solmajer dielectric, $\Delta G_{hbond}$ corresponds to the 12-10 potential with Goodford directionality, $\Delta G_{desolv}$ refers to the stouten pairwise atomic solvation parameters and $\Delta G_{tors}$ is proportional to the number of rotatable bonds.

All the torsion angles in ligand were set free to perform flexible docking. The best ligand conformation with highest affinity value was further selected as the initial conformation for MD simulation.

## 3.2 MD Simulation

MD simulations of 6 ns were carried out using Amber 11 software package [7]. We employed both explicit and implicit water model to make a comparison. The parameters and charges for the ligand were determined using *antechamber* module by utilizing the general atom force field (GAFF) and the AM1-BCC charge method. The complex (ligand-bound protein) was immersed with TIP3P water molecules with the solvent box of 10 Å between any atoms. The temperature of the system was gradually increased to 300 K during the first 20 ps, and equilibration steps done in NPT ensemble with a Langevin thermostat. Periodic boundary condition with the cut-off 12 Å was employed in the simulation with the explicit solvent model. For the implicit solvent model, we adopted Generalized Born (GB) approach (Amber parameter: *igb=2*), and run the simulation without the periodic boundary condition. The temperature was maintained at 300 K, and the system was coupled to a temperature bath with coupling constants of 1.0 ps. Bonds to hydrogen atoms constrained using the SHAKE algorithm. Total number of atoms in the respective systems was approximately 14,000.

## 3.3 MM-PBSA/MM-GBSA Calculation

The MM-PBSA/GBSA method combines the molecular mechanical energies with the continuum solvent approaches. We performed MM-PBSA/MM-GBSA integrated in Amber. The snapshots for MM-PBSA/MM-GBSA analyses were taken every 10 ps of 6,000 ps MD production runs, resulting in a total of six hundred snapshots analyzed. The binding free energy is calculated following this equation,

$$\Delta G_{bind} = \overline{G}_{complex} - [\overline{G}_{protein} + \overline{G}_{ligand}] \quad (2)$$

where $\overline{G}$ is the average free energy of the complex, protein, and ligand, are calculated according to the equation,

$$\overline{G} = \overline{E}_{MM} + \overline{G}_{solvation} - T\overline{S} \quad (3)$$

where $\overline{E}_{MM}$ are determined with the *sander* program and represent the internal energy (bond, angle, and dihedral), van der Waals and electrostatic interactions (See equation (4)). $T\overline{S}$ is the entropy contribution estimated using normal mode (*nmode*) analysis.

$$\overline{E}_{MM} = \overline{E}_{\text{int}} + \overline{E}_{elec} + \overline{E}_{vdW} \quad (4)$$

The solvation free energy can be calculated as follow,

$$\overline{G}_{solvation} = \overline{G}_{polar} + \overline{G}_{non-polar} \quad (5)$$

where $\overline{G}_{polar}$ is the electrostatic contribution calculated with a numerical solver for the PB method as implemented in the *pbsa* program or by GB method implemented in *sander*. Energy estimation with GBSA were made with the Onufriev's GB [8] parameters (*igb*=2). The non-polar contribution has been determined with the solvent-accessible-surface-area (SASA) dependent terms using the equation,

$$\overline{G}_{non-polar} = \gamma\,\mathrm{SASA} + b \quad (6)$$

where SASA were determined with the Molsurf method using a probe radius 1.4 Å. Parameter of $\gamma$=0.0072 kcal/mol Å$^2$ and $b$=0.0 kcal/mol.

## 4 Results and Discussion

Table 1 shows the result of the docking score function. The best conformation obtained is with the binding affinity 4.6 kcal/mol. Further, we adopted this structure as the initial structure to perform the first MD simulation. We run MD simulation for 6 ns and calculated the root-mean-square-deviation (RMSD) of ligand and protein of the complex both in explicit and implicit water solvent models. In Fig.4, two RMSD of ligand and protein were shown for first and second MD simulation. When the RMSD shows the complex is not highly fluctuated, we took the most stable structures by assuming that the complexes are in the equilibrium state according to their RMSDs and then performed the second docking process. In the case of explicit solvent model, we selected complexes at time 4,290 ps from the first MD trajectory and 4,457 ps from the second MD trajectory. For the implicit solvent model, complexes at time 5,626 ps from first MD and 5,386 ps from second MD were adopted. We repeated this technique until the third docking process and obtained the best ligand conformations that close to the experimental result (See Table 2). In total we performed three docking processes and two MD simulations for each explicit and implicit solvent model.

In accordance with the result we attained in Table 2, it has been shown that using the explicit solvent model is closer to the experimental result. Hence, we attempted the calculation of free energy of binding using MM-PBSA/MM-GBSA for the explicit MD production. We also compared the result with and without the inclusion of entropic term. Table 3 shows the final comparison of the various methods of calculation that we have attempted. The results show that combining docking and MD simulation with explicit solvent model is more favorable and closer to the experimental result compared to the other methods we have performed. On the other hand, MM-PBSA/GBSA results are shown overestimated. These results suggest the limitations of MM-PBSA/GBSA method which need to be considered. Continuum electrostatics models ignore the molecular structure of the solvent; in some cases this might affect the results, particularly when complexes are bridged by water molecules. Furthermore, the value of the protein/ligand dielectric constant is chosen empirically, and takes into account. The inclusion of entropic contributions brings the MM-PBSA/GBSA values somewhat closer to experimental absolute affinities. However, such entropic terms are costly and contain large uncertainties. Force-field inconsistencies may also be an issue: PB and GB results depend strongly on adequate atomic charges and van der Waals radii, which are often optimized for MD simulations.

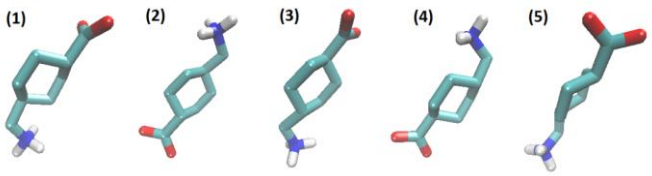


Figure 3: Five best ligand conformations obtained from first docking process (see Table 1 for details).

Table 1: Five best selected ligand conformations at first docking process [3].

| Mode | Affinity (kcal/mol) | RMSD lower bound (Å) | RMSD upper bound (Å) |
|---|---|---|---|
| **1** | **-4.6** | **0.000** | **0.000** |
| 2 | -4.5 | 4.047 | 5.242 |
| 3 | -4.5 | 0.769 | 1.676 |
| 4 | -4.4 | 3.857 | 4.880 |
| 5 | -4.4 | 1.349 | 1.430 |

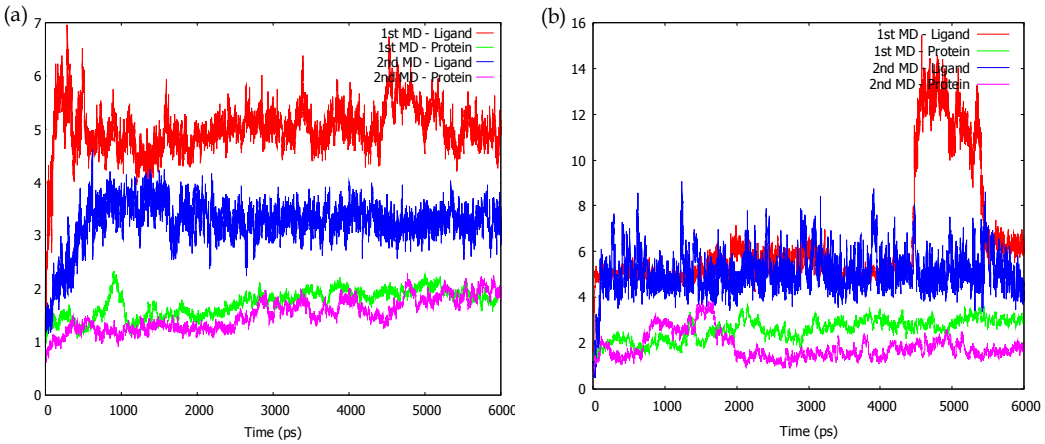


Figure 4: RMSD of ligand and protein of the complex as a function of MD time (a) Explicit solvent model, (b) Implicit solvent model.

Table 2: $\Delta G_{bind}$ results with explicit and implicit solvent model.

| Docking process | $\Delta G_{bind}$ (kcal/mol) | |
|---|---|---|
| | **Explicit solvent model** [3] | **Implicit solvent model** |
| Second | -4.9 | -5.6 |
| Third | -5.0 | -5.5 |
| Experimental result: -5.08±0.01 kcal/mol | | |

Table 3: Comparison of $\Delta G_{bind}$ results with various calculation methods.

| Method | $\Delta G_{bind}$ (kcal/mol) |
|---|---|
| Experimental [5] | -5.08±0.01 |
| Molecular docking (Autodock) | -4.6 |
| Combination of docking and MD simulation with explicit solvent [3] | -5.0 |
| Combination of docking and MD simulation with implicit solvent | -5.5 |
| AMBER PBSA without the inclusion of entropic term | -23.22±4.92 |
| AMBER GBSA without the inclusion of entropic term | -21.52±3.35 |
| AMBER PBSA with the inclusion of entropic term | -16.23±5.14 |
| AMBER GBSA with the inclusion of entropic term | -14.53±3.67 |

We also tried to confirm the presence of water molecules around the binding site and ligand during the simulations by calculating the number of water molecules within 3Å around the ligand, binding site and protein surface (See Fig. 5). The number of water molecules around the binding site and ligand has been decreased significantly. This implies that the binding site and ligand itself are hydrophobic causing the water molecules moved away from this region to further stabilize the conformation. Furthermore, we investigated the contribution of solvent in the MM-PBSA/MM-GBSA binding free energy calculation that we have performed. In the calculation, the free energy of solvent ($\overline{G}_{solvation}$) is equal to 96.44 kcal/mol for PB approach and 98.14 kcal/mol for GB approach. The positive value of $\overline{G}_{solvation}$ means that solvation energy does not contribute to the affinity of ligand to the protein. $\overline{E}_{MM}$ energy contributes to the binding energy. However, this large interaction energy (-119 kcal/mol) should not consist of only hydrophobic interaction energy, it should include the polar interactions like hydrogen bond energy.

Further analysis is also shown in Figure 6. This figure shows the structure of complex at the first docking process and final structure of the explicit solvent model with combining docking and MD simulation approach. The first docking structure shows that lysine binding site is engineered by cationic Arg36/Arg71, anionic Asp55/Asp57 and hydrophobic Trp62/Trp72 as also reported by Christen et al [5]. Note that this first docking structure we obtained by performing rigid receptor-flexible ligand docking using Autodock. The hydrophobic residues Trp62 and Trp72 are shown associated each other at the first docking structure, whereas in the final structure they are shown in distant, as the consequences of the interaction between these residues with water molecules.

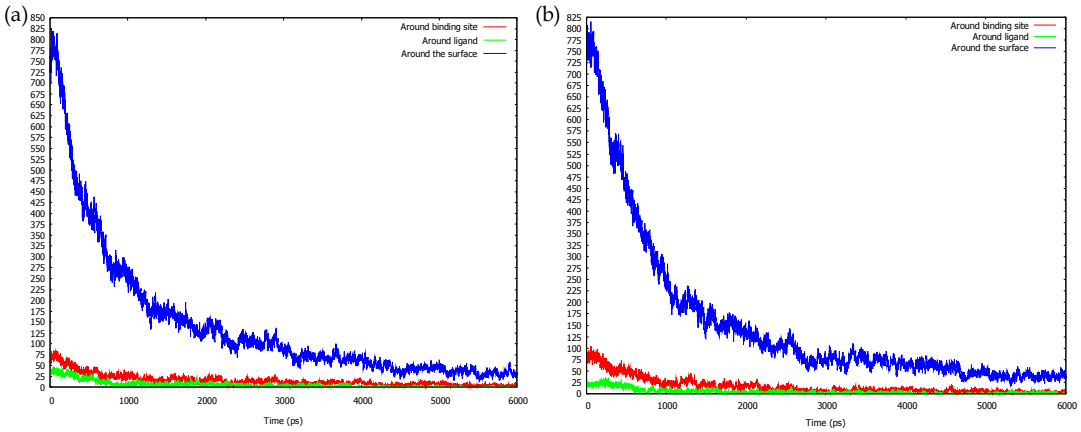


Figure 5: Number of water molecules around ligand (colored as green), binding site (colored as red), and around the protein surface (colored as blue) during the MD simulations. (a) First MD simulation with explicit solvent. (b) Second MD simulation with explicit solvent.

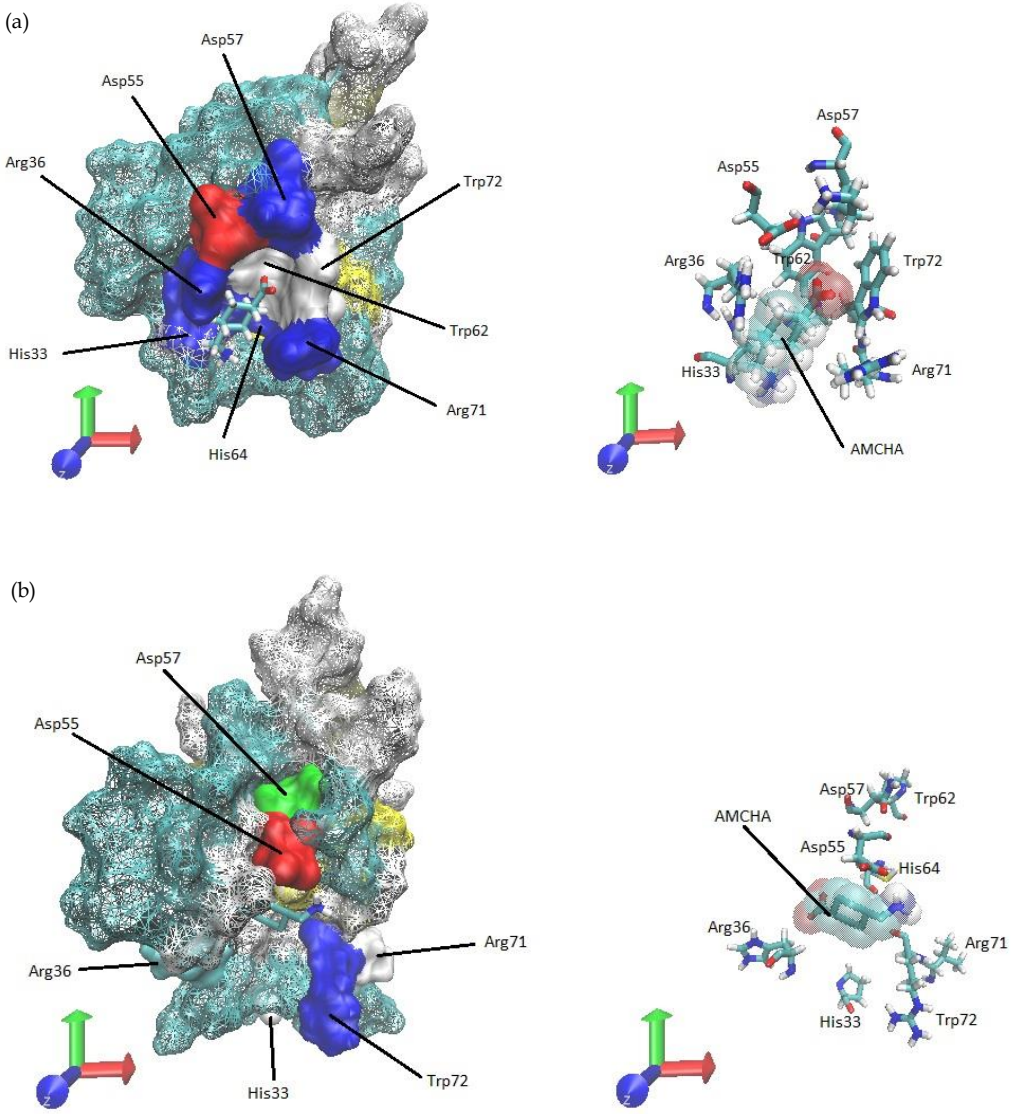


Figure 6: (a) First docking structure. (b) Final docking structure.

## 5 Conclusion

We compared the calculation of $\Delta G_{bind}$ between experimental and various methods and proposed an approach to combine docking score function and MD simulation. We denoted that the flexibility of protein and ligand and explicit water molecules are important factors for obtaining the most stable structure in docking simulation. The protein flexibility allows increased affinity to be achieved, and this can be done by performing the MD simulations. Thus, combining docking score function and MD simulation with explicit solvent model should be effective and accurate in obtaining the best affinity

between ligand and protein. Docking score function can be used to predict the binding orientation of ligand and evaluate binding free energy (affinity), and MD simulation is used to obtain the most stable structure of complexes.

## References

[1] I. Halperin, B. Ma, H. Wolfson, and R. Nussinov (2002), Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions, *PROTEINS: Structure, Function, and Genetics*, **47**, 409 – 443.

[2] H. Alonso, A. A. Bliznyuk, and J. E. Gready (2006), Combining Docking and Molecular Dynamic Simulations in Drug Design, *Wiley. Med. Res. Rev.*, **26**, 531-568.

[3] M. T. Pakpahan, M. Rusmerryani, K. Kawaguchi, H. Saito, and H. Nagao (2013), Evaluation of scoring functions for protein-ligand docking, *AIP Conf. Proc.*, **1518**, 645-648.

[4] D. S. Chauhan, S. Chandra, A. Gupta, and T. R. Singh (2012), Molecular modelling, docking and interaction studies of human-plasmogen and *salmonella*-enolase with enolase inhibitors, *Bioinformation 8*, **4**, 185-188.

[5] M. T. Christen, P. Frank, J. Schaller, and M. Llinás (2010), Human Plasminogen Kringle 3: Solution Structure, Functional Insights, Phylogenetic Landscape, *Biochemistry,* **49**, 7131-7150.

[6] O. Trott, and A. J. Olson (2010), AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *J. Comput. Chem.*, **31**, 455-461.

[7] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman (2010), AMBER 11, University of California, San Francisco.

[8] A. Onufriev, D. Bashford, and D. A. Case (2004), Docking score function can be used to predict the binding orientation of ligand and evaluate binding free energy (affinity), and MD simulation is used to obtain the most stable structure of complexes, *PROTEINS: Structure, Function, and Bioinformatics*, **55**, 383-394.