J Comput Aided Mol Des (2010) 24:699–712
DOI 10.1007/s10822-010-9367-z

Dependency of ligand free energy landscapes on charge parameters and solvent models

Yuko Okamoto · Toshimasa Tanaka · Hironori Kokubo

Received: 22 February 2010 / Accepted: 6 May 2010 / Published online: 22 May 2010

Abstract We performed replica-exchange molecular dynamics (REMD) simulations of six ligands to examine the dependency of their free energy landscapes on charge parameters and solvent models. Six different charge parameter sets for each ligand were first generated by RESP and AM1-BCC methods using three different conformations independently. RESP charges showed some conformational dependency. On the other hand, AM1-BCC charges did not show conformational dependency and well reproduced the overall trend of RESP charges. The free energy landscapes obtained from the REMD simulations of ligands in vacuum, Generalized-Born (GB), and TIP3P solutions were then analyzed. We found that even small charge differences can produce qualitatively different landscapes in vacuum condition, but the differences tend to be much smaller under GB and TIP3P conditions. The simulations in the GB model well reproduced the landscapes in the TIP3P model using only a fraction of the computational cost. The protein-bound ligand conformations were rarely the global minimum states, but similar conformations were found to exist in aqueous solution without proteins in regions close to the global minimum, local minimum or intermediate states.

Keywords Free energy landscape · Replica-exchange molecular dynamics · RESP · AM1-BCC

Introduction

Exploring the global minimum free energy states of ligand molecules is essential, because they are initial states for various physicochemical processes such as protein–ligand binding and membrane permeation, where free energy differences between these initial states and the final states govern these phenomena. (In this article we use the term *ligands* as almost equivalent to *small molecules*.) Molecular simulations are powerful tools to elucidate conformational distributions under various conditions and then estimate global and local minimum free energy states.

It is presumed that ligands rarely bind with proteins in their global-minimum free energy conformations [1, 2]. Thus, most ligands are considered to have conformational strain energy upon binding with a protein, which is one of the important components of the total free energy differences. This conformational strain energy is often estimated by subtracting the global-minimum potential energy of a ligand after energy minimization from the potential energy of a ligand in a protein-bound state [1–6]. Here, the global-minimum potential energy is usually estimated by minimizing the energy of the conformations generated by systematic or stochastic searches. However, it is known that the global minimum potential energy thus obtained strongly depends on initial conformations and minimization procedures. More importantly, conformational strain energy of a ligand should correspond to an ensemble average of the ligand potential energy in a protein-bound state minus that of the global minimum free energy state in water at a temperature of 300 K and a pressure of 1 atm.

Yuko Okamoto and Hironori Kokubo contributed equally to this work.

Y. Okamoto
Department of Physics, School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

T. Tanaka · H. Kokubo (✉)
Pharmaceutical Research Division, Takeda Pharmaceutical Co., Ltd, 2-17-85 Juso-honmachi, Yodogawa-ku, Osaka 532-8686, Japan
e-mail: kokubo_hironori@takeda.co.jp

Ensemble average properties can be estimated by molecular dynamics simulations, which are based on empirical force-field functions and parameters. In 2004 Wang et al. developed a general Amber force field (GAFF [7, 8]) for most organic and pharmaceutical molecules composed of H, C, N, O, S, P, and halogens. RESP [9, 10] and AM1-BCC [11] are well established methods for the charge determination when GAFF model is applied. A protein-bound ligand conformation is usually used as the input conformation in the process of charge determination when it is available from Protein Data Bank [12] (PDB). However, it is not clear how much free energy landscapes of ligands are influenced by differences of charge parameters and solution models. Although the AM1-BCC model has been developed to reproduce the overall trend of RESP charges, Fujitani et al. claimed that the RESP model produced the absolute binding free energy in protein–ligand binding more accurately than the AM1-BCC model [13]. However, AM1-BCC charges have been used successfully in other cases [14–17]. One of us studied two urea models with different charge parameters and showed that these two models produced different thermodynamic solution properties [18, 19].

In this work, we examine in detail the dependency of free energy landscapes on different charge parameters and solvent models by using six ligands as a test set. The charge parameters obtained from different ligand conformations by AM1-BCC and RESP models are compared, and the free energy landscapes obtained by using these charge parameters are examined in detail. Three different solvent conditions, vacuum, GB implicit solvent model [20, 21], and TIP3P explicit solvent model [22], are applied to explore the dependency on solvent conditions. We also examine where bioactive conformations, defined as the conformations of protein-bound ligands, are on free energy landscapes. Although our studies are limited by the test set employed, we think that our results provide some useful insights.

This Article is organized as follows. "Methods" summarizes the details of our method for generating charge parameters, performing replica-exchange molecular dynamics simulations and analyzing free energy landscapes. "Results and discussion" presents the results and discussion about free energy landscapes generated by different charge parameters and solution models. The last section is devoted to conclusions.

Methods

We first selected six pharmaceutical ligand molecules: imatinib, atazanavir, rivaroxaban, sildenafil, topiramate, and biotin, for all of which crystal structures of protein complexes have been reported. Table 1 lists basic properties of the six ligand molecules. Although only six molecules cannot represent the vast chemical space of all pharmaceutical molecules, we see that these molecules generally have typical drug-like properties and obey the well-known rule-of-five [24]; that the most orally administered drugs have a molecular weight of 500 or less, an octanol–water partition coefficient c log P no higher than 5, five or fewer hydrogen-bond donor sites, and 10 or fewer hydrogen-bond acceptor sites. Only atazanavir has a larger molecular weight than the criterion and it is the most complicated molecule among the six molecules. Figure 1 shows the structures of the six ligand molecules complexed to proteins, as registered in the PDB.

We first extracted ligand conformations from the PDB files and determined the RESP charges, which were calculated by ab initio molecular orbital programs at the HF/6-31G* level and RESP utilities in the Antechamber program suite [8]. The molecular mechanics parameters by GAFF [7, 8] were employed for bond, angle, torsion, and van der Waals (vdW) parameters. Missing force-field parameters were assigned with the *parmcheck* utility in AMBER [26]. After performing energy minimization by steepest-descent and conjugate gradient methods (in 500 steps for each minimization), we carried out an equilibration simulation for 1.0 ns at a temperature of 300 K. Temperature was regulated by Langevin dynamics with the collision frequency 2.0 ps^{-1}. The SHAKE [27] method was employed to constrain bond lengths involving hydrogen atoms and

Table 1 Properties of the six ligand molecules examined in these studies

	NumrotB	Numheavy	Weight	HBacc	HBdon	*c*log *P*
Imatinib	8	37	495	5	2	2.8
Atazanavir	22	51	705	7	5	4.7
Rivaroxaban	6	29	436	4	1	1.6
Sildenafil	7	33	476	6	1	0.29
Topiramate	3	22	361	10	1	−0.39
Biotin	5	16	243	1	2	0.44

NumrotB: number of rotatable bonds, Numheavy: number of heavy atoms, Weight: molecular weight, HBacc: number of hydrogen-bond acceptor sites, HBdon: number of hydrogen-bond donner sites, and *c*log *P*: calculated Log octanol/water partition coefficient by MOE [23]

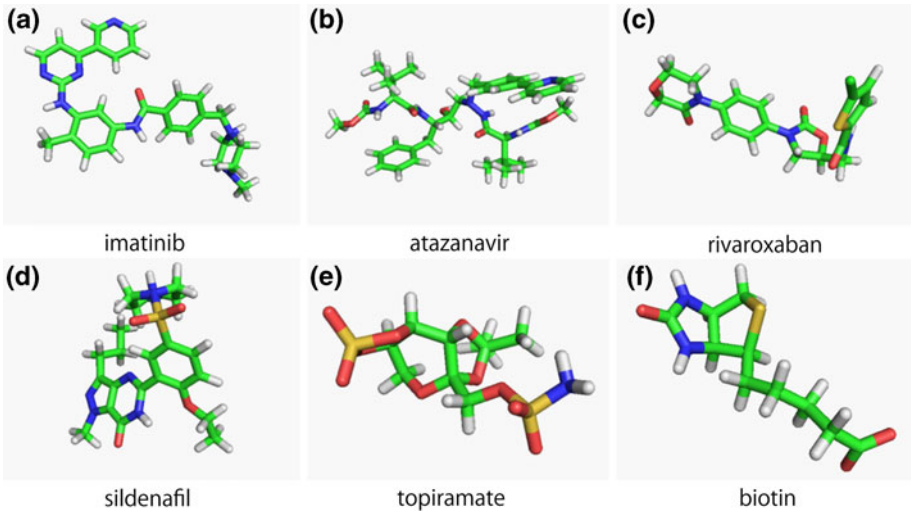

Fig. 1 Ligand molecule conformations complexed to proteins (from PDB). PDB IDs are **a** 1XBB, **b** 2AQU, **c** 2W26, **d** 2H42, **e** 1EOU, and **f** 1STP. The figures were created with PyMOL [25]

the time step was 2.0 fs. We then performed constant temperature simulation using the GB solvation model [20, 21] at 1,000 K for 50 ps in order to obtain various conformations without dependency on the initial conformations. Starting from the conformations extracted from the 1,000 K simulations at time steps of 10, 20, 30, 40, and 50 ps, we performed five independent simulated annealing simulations [28] from 1,000 to 200 K over 10 ns. The temperature was decreased exponentially every step to avoid abrupt temperature changes. We then performed equilibration simulations at 200 K for 300 ps and the final conformations thus obtained are referred to as the simulated annealing conformations. Thus we obtained five independent simulated annealing conformations for each ligand. The two ligand conformations out of five were selected. The first one was the conformation with the largest root-mean-square deviations (RMSD) from the PDB conformation. The second one was the conformation which was judged to be dissimilar to both the PDB and the first conformations by visual inspection after superposition. Thus, we ultimately prepared three different conformations for each ligand: the native conformation from the PDB, simulated annealing conformation 1, and simulated annealing conformation 2.

Using these three conformations for each ligand, we again determined the atomic charge parameters by two different methods: AM1-BCC and RESP to examine the conformational dependency of charge parameters. Thus, six different charge parameters were obtained for each ligand. Figure 2 shows the obtained atomic charges for three ligand conformations and the two charge models. In the case of the AM1-BCC model, we show only the results for the PDB conformations in Fig. 2 because AM1-BCC charges were almost identical for the different conformations. Note that we employed GAFF parameters for bond, angle, torsion, and vdW terms for all the molecular dynamics simulations.

We then studied the systems under three different solvent conditions: vacuum, GB, and TIP3P. Thus, 18 independent simulations were performed for each ligand, corresponding to six charge parameters in three solvent conditions. For TIP3P conditions we immersed each ligand in a TIP3P water cube with at least 8 Å solvation from the ligand surface, and the systems were neutralized by adding one sodium ion (for biotin) or one chloride ion (for imatinib and sildenafil) for charged ligands. The Berendsen thermostat was employed for simulations in vacuum and in TIP3P, and Langevin dynamics was employed for simulations in GB for temperature control at 300 K. The pressure was controlled by the weak-coupling method [29] at 1.0 atm for TIP3P systems. Before production runs, we performed energy minimization for 1,000 steps and equilibration simulations for 1.0 ns.

We then performed replica-exchange molecular dynamics [30–32] (REMD) simulations to obtain free energy landscapes accurately. REMD simulations can avoid getting trapped in local-minimum free energy states, because the temperature of each replica goes up and down by temperature exchange. Therefore, this method is useful for simulating systems with multiple-minimum states. We can identify not only the global-minimum free energy state but also the local-minimum free energy states.

The following eight temperatures (eight replicas) were used for the simulations in vacuum and GB solution: 260.5, 300.0, 345.5, 397.9, 458.3, 527.7, 607.8, and 700.0 K. On the other hand, 32 temperatures (32 replicas) were used for the simulations in TIP3P water model: 300.0. 304.9, 310.2, 315.8, 321.7 327.9, 334.3, 340.8, 347.6, 354.6, 361.7, 369.1, 376.7, 384.6, 392.7, 401.3, 410.0, 419.2, 428.8, 438.8, 449.3, 460.3, 471.8, 483.8, 496.4, 509.6, 523.3, 537.6, 552.5, 567.9, 583.7, and 600.0 K. This temperature distribution was chosen to achieve sufficient and uniform replica exchange from the preliminary runs. Before production runs, we performed

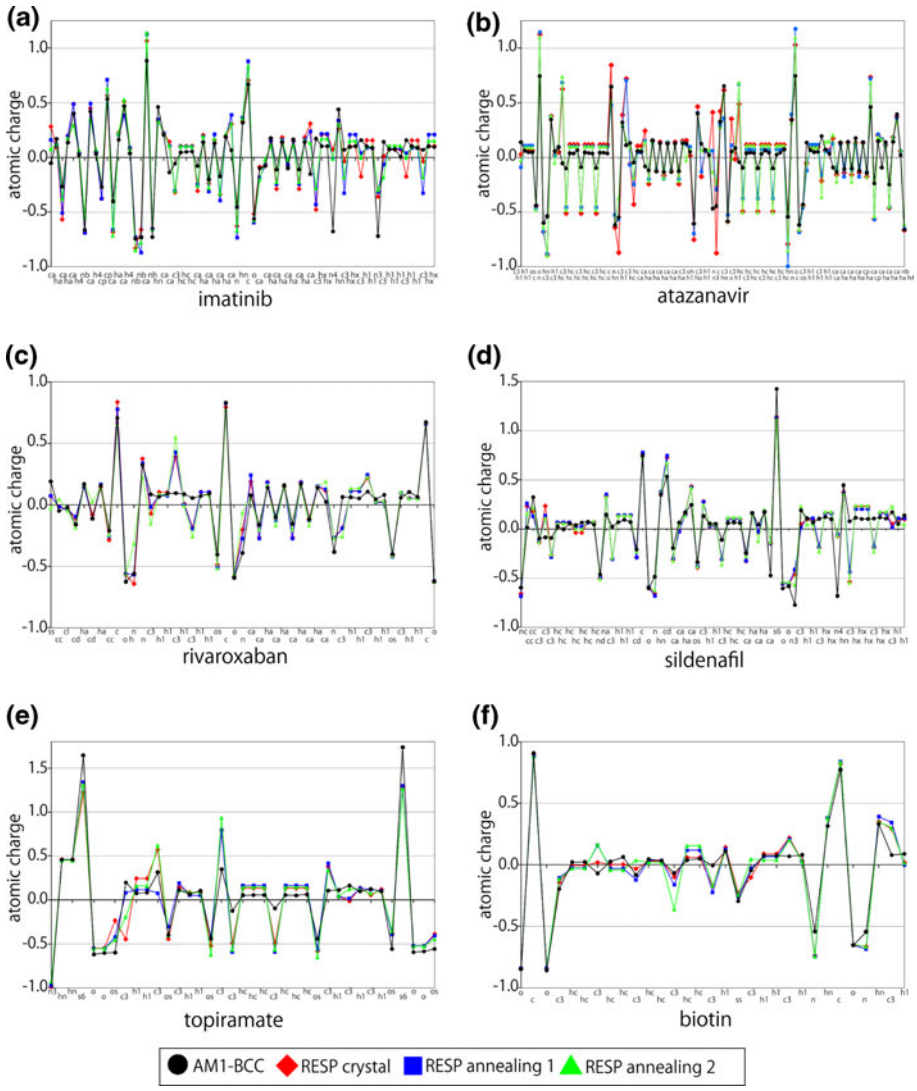

Fig. 2 Atomic charges of six ligand molecules. The *black circle* shows the AM1-BCC charge and other three symbols show the RESP charge on each atom based on the PDB conformations (*red diamond*, see Fig. 1) and conformations based on simulated annealing (*blue square* and *green triangle*). The two simulated annealing conformations are those with the largest RMSD values with respect to the PDB conformations out of five independent simulated annealing runs. The AM1-BCC charge model did not show conformational dependency, thus only the charges for the PDB conformations are shown

equilibration runs for 1.0 ns. We then performed 20 ns REMD simulations for all the systems and collected 20,000 snapshots at even intervals for each replica. The conformations at a temperature of 300 K were collected, and those 20,000 conformations were used for conformational analyses on free energy landscapes.

In order to characterize the structural properties and express the free energy landscapes, we employed the reaction coordinates of radius of gyration, surface area, RMSD, and principal component axes.

The radius of gyration R_g represents the deviation of the atoms in a molecule from its center, thus describing the extent of collapse or extension of the molecule:

$$R_g = \sqrt{\sum_{i=1}^{N} \frac{(\vec{r}_i - \vec{r}_G)^2}{N}},$$

where N is the total number of atoms in the molecule, $\vec{r}_i$ is the position of the i-th atom and $\vec{r}_G$ is the center of geometry (=center of mass without mass weighted) for the molecule. A low radius of gyration value corresponds to a compact state, whereas a high radius of gyration is indicative of a broader distribution of the atoms of the molecule from the center of geometry. Surface area was estimated by using a 1.4 Å sphere probe, and RMSDs from the PDB conformations were also calculated.

Principal component analysis [33–38] is an effective method to classify conformations of a molecule. First, N conformations at 300 K were extracted from the replica-exchange simulations at even intervals. Each structure was superimposed by translation and rotation on the reference structure. In this work we chose the PDB structures as the reference structures. This superposition was performed in order to remove the freedom of translations and rotations since our purpose is to analyze differences in structures. We then calculated the following variance–covariance matrix C_{ij}:

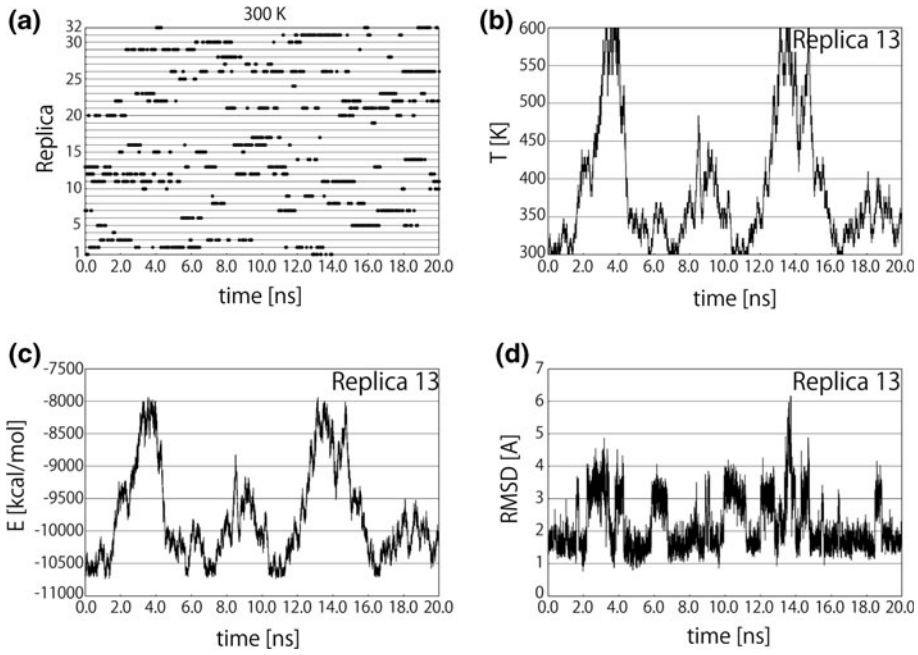

Fig. 3 Time series of **a** replica exchange (at 300 K), **b** temperature exchange, **c** total potential energy (in kcal/mol), and RMSD (in Å) from the global minimum free energy structure (in Fig. 12b) obtained from the simulation of atazanavir in TIP3P solution. Time series of Replica 13 is shown for **b**, **c**, and **d** as a representative

$$C_{ij} = \left\langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \right\rangle$$

where $\vec{q} = (q_1, q_2, q_3, \ldots, q_{3n-2}, q_{3n-1}, q_{3n}) = (x_1, y_1, z_1, \ldots, x_n, y_n, z_n)$, $\langle q_i \rangle = \sum_{i=1}^{N} q_i/N$, x_i, y_i, z_i are Cartesian coordinates of the ith atom, and n is the total number of atoms. This symmetric matrix was diagonalized and the eigenvector and eigenvalues were obtained. The first and second principal component axes were defined as the eigenvectors with the largest and second-largest eigenvalues, respectively. The ith principal component of each sampled structure μ_i is defined by the inner product:

$$\mu_i = \vec{v}_i \cdot \left(\vec{q} - \left\langle \vec{q} \right\rangle \right)$$

where $\vec{v}_i$ is the ith eigenvector.

We analyzed the free energy landscapes for three different pairs of the reaction coordinates: radius of gyration vs. surface area, radius of gyration vs. RMSD, and the first and second principal component axes. We used the following equation to calculate the free energy as a function of two reaction coordinates x_1 and x_2.

$$F(x_1, x_2) = -k_B T \ln P(x_1, x_2)$$

where k_B is the Boltzmann constant, T is absolute temperature, and $P(x_1, x_2)$ is the probability of finding the structure with the reaction coordinate values x_1 and x_2.

Results and discussion

Figure 1 shows the structures of six molecules in complex with proteins registered in the PDB, and Table 1 lists their basic properties. Most molecules have typical drug-like properties, though atazanavir has an exceptionally large molecular weight for an approved drug and is the most complicated among these molecules.

Figure 2 shows the atomic charges for three different ligand conformations and the two charge models for each ligand. We only show the PDB conformations for AM1-BCC because AM1-BCC charges were almost identical for different conformations. On the other hand, we see that RESP charges show some conformational dependency. AM1-BCC was designed to yield charge sets of comparable quality to RESP charges and we confirmed that AM1-BCC well reproduces the overall trend of the RESP model in each case. However, when we look in detail, large deviations between AM1-BCC and RESP charges are observed for some atoms, namely that even the signs of atomic charges are the opposite in some cases. We also confirmed that equivalent hydrogen atoms, such as those in a methyl group, have different charges in AM1-BCC. On the other hand, such chemically equivalent hydrogen atoms have the same charges in RESP due to the restraints of RESP method. We also remark that RESP treatment of equivalent hydrogen atoms is consistent with the AMBER force field parameters of amino acids.

In order to examine how much these charge differences affect the free energy landscapes, we performed REMD simulations. We used three different solvent conditions: vacuum, GB, and TIP3P. Thus, 18 independent simulations for each ligand were performed, corresponding to six charge parameter sets for each of three solvent conditions.

In Fig. 3 we show the time series of various quantities that we obtained from the REMD simulation of atazanavir in TIP3P water solution as a representative. RESP charge was

determined using the PDB conformation for atazanavir. Figure 3a is the time series of replica exchange. The time series of replica number at a temperature of 300 K is shown. We see that most replicas experienced 300 K many times. Figure 3b shows the time series of temperature exchange of one of the replicas (Replica 13). We observe that the temperature went up and down between the lowest value and the highest one and that sufficient sampling has been achieved. Other replicas behaved similarly. The acceptance ratios of replica exchange between all pairs of neighboring temperatures were almost uniform and large enough (ranging between 19 and 35%). Figure 3c is the time series of the total potential energy corresponding to Fig. 3b. When the temperature is high the total potential energy is high, and when the former is low, so is the latter. There is a strong correlation between Fig. 3b and c, as is to be expected. These figures show that the replica exchange simulations have been properly performed. We now examine how widely the conformational space was sampled. Figure 3d is the time series of the global minimum free energy structure (in Fig. 12b) of atazanavir. We see that the RMSD goes up and down between the lowest and highest values found, and that wide conformational space has indeed been sampled during the simulations. In addition, we further performed additional 20 ns REMD simulations (40 ns in total) and confirmed that free energy landscape was almost the same (data not shown). Comparing Fig. 3b and d, we see that the small RMSD conformations tend to be sampled more often at low temperatures. Thus far we examined how REMD simulations worked well in the case of the most complicated system in our work: atazanavir in TIP3P water. We also confirmed that REMD simulations have also been performed properly in all other systems.

Figure 4 shows the average radius of gyration at 300 K obtained from the REMD simulations by using six different charge parameters and three different solvent conditions for each ligand. We see that conformations with large radius of gyration tend to be more stable in GB and TIP3P solvent than in vacuum. It is worth noting that, while Agrafiotis and other researchers [5] pointed out that bioactive conformations defined as protein-bound ligand ones tend to be extended compared to global-minimum states, we did not observe such a tendency in GB and TIP3P conditions even in the case of the same ligand (biotin in Fig. 4f) that they have examined. In Fig. 4 we see no indication that the radii of gyration of the PDB conformations (black line) tend to be larger than the average radii of gyration in GB and TIP3P conditions (black circle). This

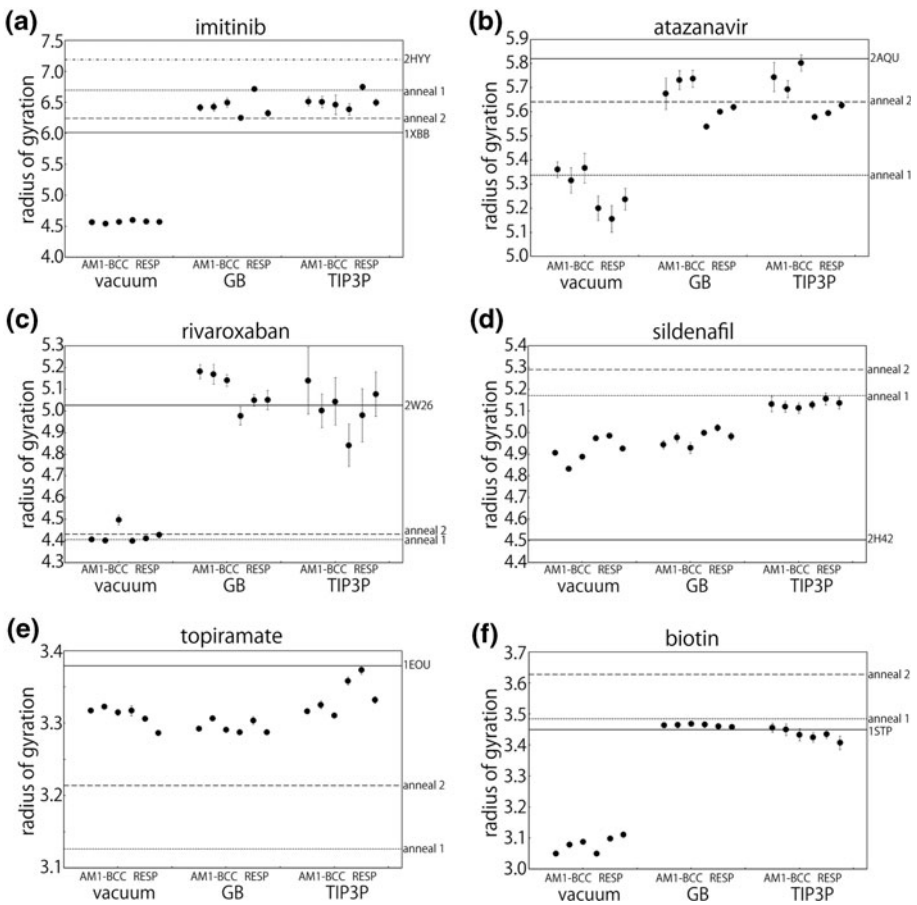

Fig. 4 Average radius of gyration (in Å) at 300 K for the following ligands: **a** imatinib, **b** atazanavir, **c** rivaroxaban, **d** sildenafil, **e** topiramate, and **f** biotin. A total of 18 simulations with six different charge parameters (AM1-BCC for PDB, annealing 1, and annealing 2 conformations, and RESP for PDB, annealing 1, and annealing 2 conformations; shown from *left* to *right*, respectively) under three conditions (in vacuum, GB, and TIP3P) were performed for each ligand. The *horizontal lines* "PDB ID", "anneal 1", and "anneal 2" show the radius of gyration of the three structures which were used in the charge determination. *Error bars* were estimated by dividing the data into 10 blocks and calculating the standard deviation from the average values of each block

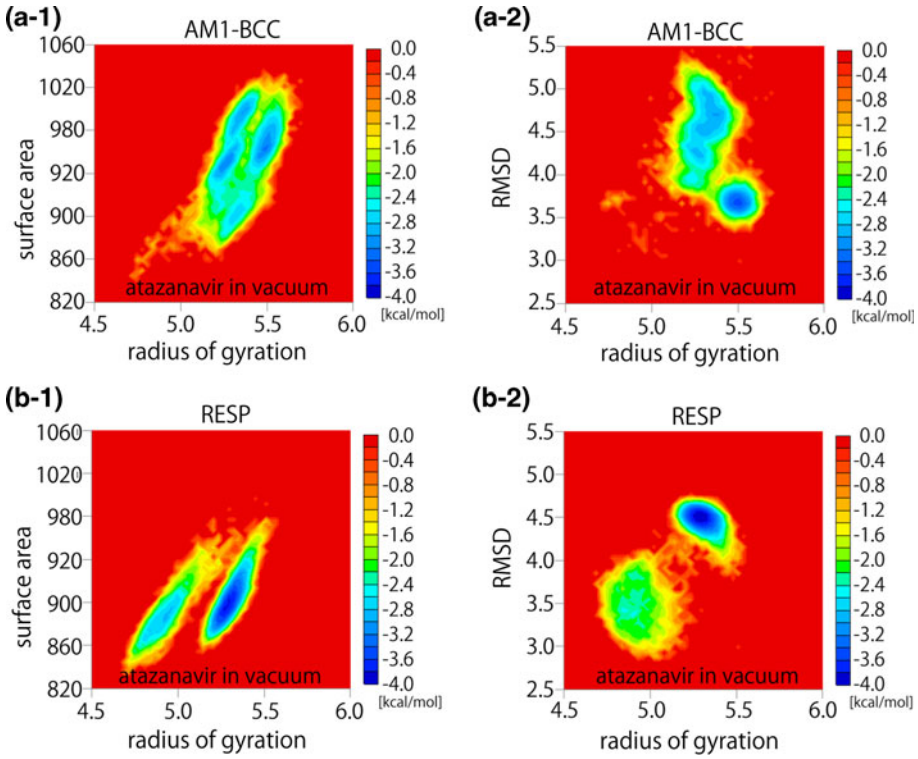

Fig. 5 Free energy landscapes of atazanavir in vacuum with **a** AM1-BCC and **b** RESP charges. (*a-1*) and (*b-1*) show the landscapes for the radius of gyration (in Å) and surface area (in Å^2), and (*a-2*) and (*b-2*) show those for the radius of gyration (in Å) and RMSD (in Å). RMSD was calculated with respect to the PDB conformation (PDB ID: 2AQU) and only heavy atoms were taken into account. The *color* shows the free energy (kcal/mol). Conformations at 300 K from the REMD simulations in vacuum were analyzed to draw the free energy landscape. The same conformation extracted from the PDB was used in both AM1-BCC and RESP charge determination procedures

discrepancy is considered due to the difference of the definition of global minimum state. We defined the global minimum state as the lowest free energy state in aqueous solution at 300 K (and 1 atm in TIP3P solution) as it should be, though their definition was based on the lowest potential energy after minimization procedures in less accurate solvent conditions. Note that smaller radius of gyration tended to be stable in vacuum condition.

Another observation in Fig. 4 is that we do not see any indication that the conformations used in the charge determination process have become more stable. Thus, the differences in charge parameters shown in Fig. 3 seem not to have a strong effect on the average radius of gyration.

Although the average radius of gyration can provide global trends, it is worth exploring free energy landscapes to examine the dependency on charge parameters and solvent conditions in detail.

In Figs. 5 and 6 we compare the free energy landscapes for the cases of AM1-BCC and RESP charges. Both methods used the same ligand conformation (from PDB) as input in the charge determination process. Figure 5 shows the case of atazanavir in vacuum. We see that the free energy landscape for the AM1-BCC charge model is qualitatively different from that for the RESP charge model. Comparing (a-2) with (b-2), conformations with a radius of gyration of 5.5 Å and RMSD of 3.7 Å are stable in AM1-BCC, but these were rarely sampled and were unstable in RESP. Both the conformational distribution and the most stable regions found are different. This implies that although AM1-BCC well reproduces the overall trends of the RESP model (see Fig. 2b) it does not necessarily mean that similar free energy landscapes are obtained.

On the other hand, Fig. 6 shows the cases of rivaroxaban in GB model. We see that the free energy landscapes for AM1-BCC and RESP are almost the same both qualitatively and quantitatively. Among the six ligand molecules that we studied here, we observed qualitative differences between AM1-BCC and RESP for atazanavir and topiramate, while for the other four ligand molecules the free energy landscapes were essentially identical under the same solvent conditions.

We next examine the influence on free energy landscapes by the conformational differences in charge determination. Figures 7 and 8 show the free energy landscapes for RESP charges determined from the PDB conformation and the simulated annealing conformations. Figure 7 shows the case of atazanavir in vacuum. Comparing (a-2) with (b-2), the conformations with a radius of gyration of 5.25 Å and RMSD of 4.0 Å are stable in (b-2), but these were rarely sampled and were unstable in (a-2). This implies that stable conformations can change greatly due to conformational differences in charge determination, even when the same RESP method is used.

Figure 8 shows the case of TIP3P solution instead of vacuum with the same molecule and charge parameters as in Fig. 7. We see that the free energy landscapes are very similar to each other even though the same charge parameters were used as in Fig. 7. We also observed a

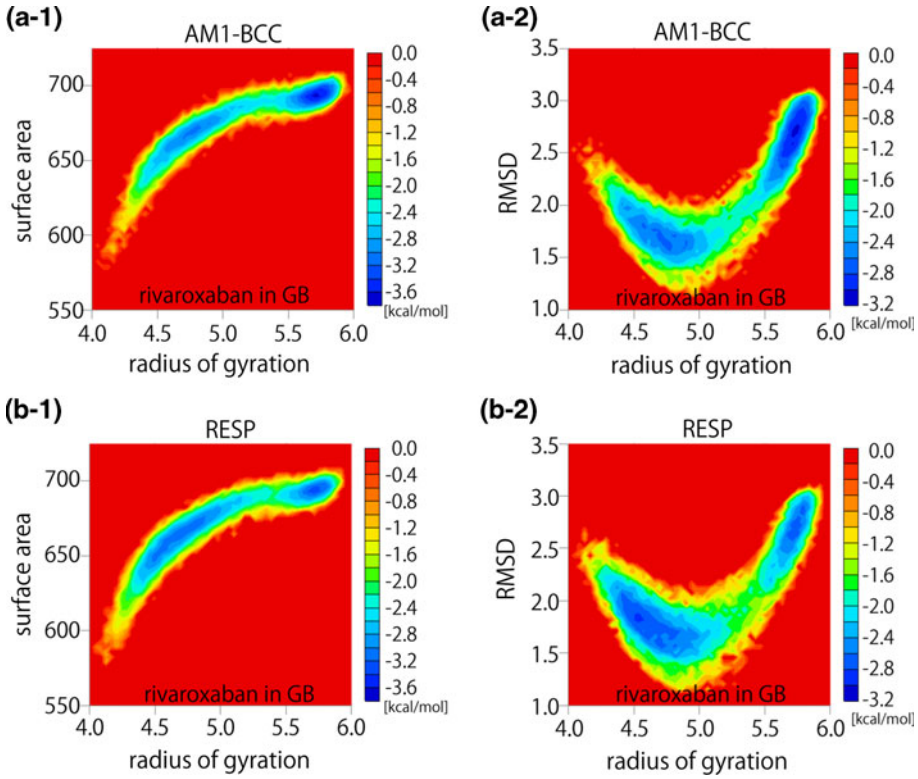

Fig. 6 Free energy landscapes of rivaroxaban in GB solution with **a** AM1-BCC and **b** RESP charges. (*a-1*) and (*b-1*) show the landscapes for the radius of gyration (in Å) and surface area (in Å^2), and (*a-2*) and (*b-2*) show those for the radius of gyration (in Å) and RMSD (in Å). Conformations at 300 K from the REMD simulations in GB solution were analyzed to draw the free energy landscape. The same conformation extracted from the PDB (PDB ID: 2W26) was used in both AM1-BCC and RESP charge determination procedures

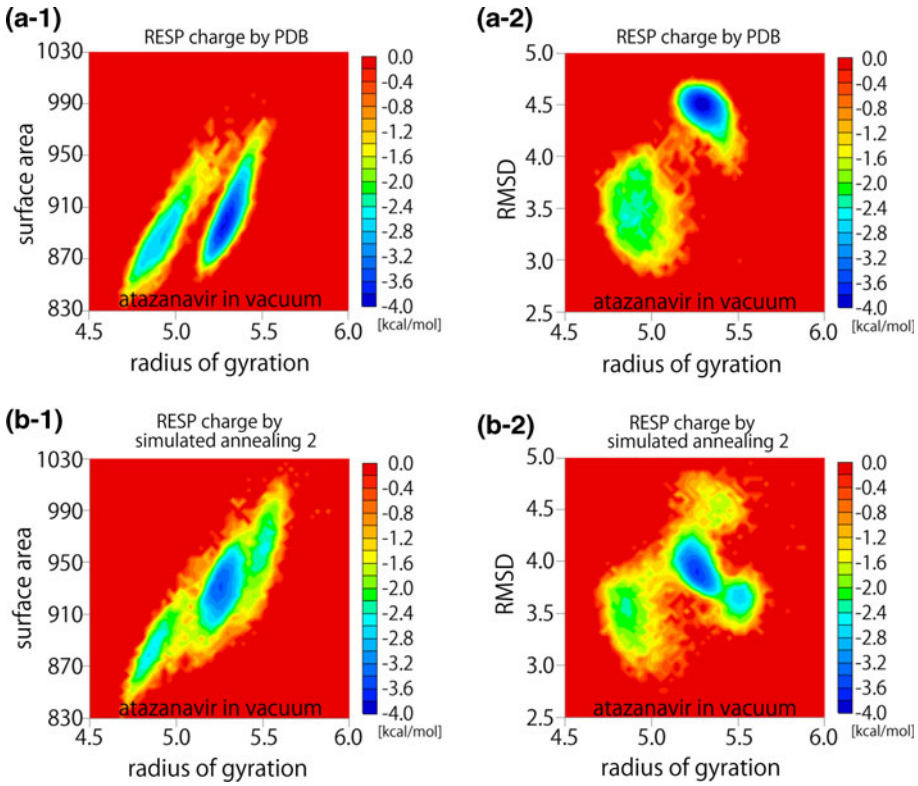

Fig. 7 Free energy landscapes of atazanavir in vacuum with RESP charge parameters determined using **a** the PDB conformation (PDB ID: 2AQU) and **b** the conformation obtained by simulated annealing. (*a-1*) and (*b-1*) show the landscapes for the radius of gyration (in Å) and surface area (in Å^2), and (*a-2*) and (*b-2*) show those for the radius of gyration (in Å) and RMSD (in Å). Conformations at 300 K from the REMD simulations in vacuum were analyzed to draw the free energy landscape

similar trend in other ligand molecules, namely that the influence of differences in charge parameters become small or disappear in GB and TIP3P solution compared to the results in vacuum. These observations are reasonable because both the GB and TIP3P models have some screening effect on internal ligand interactions.

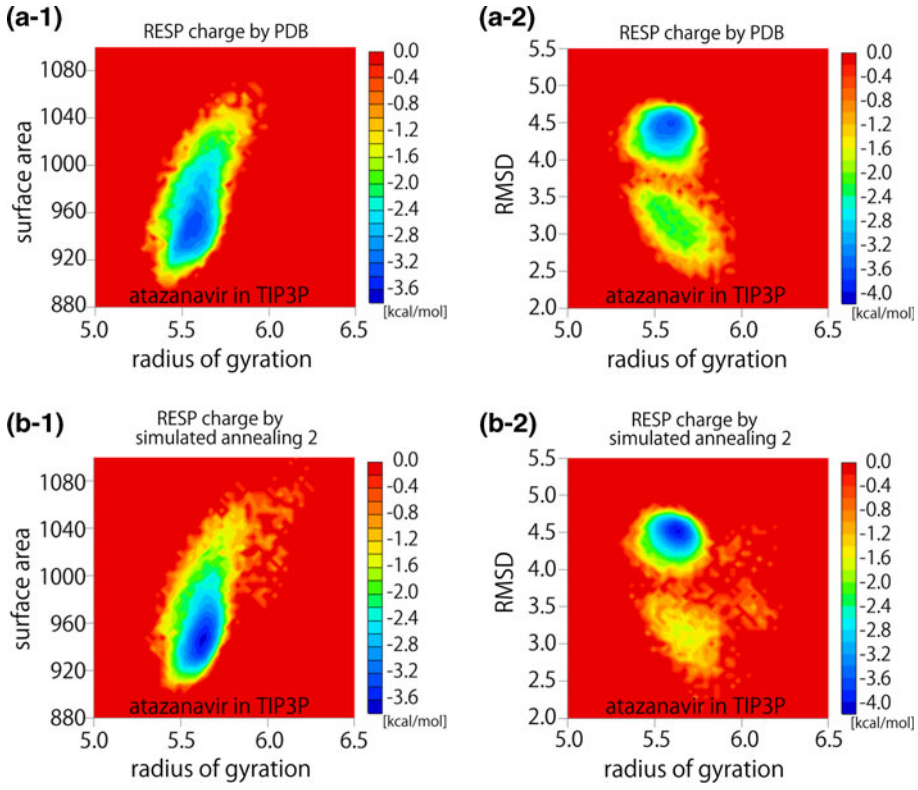

Fig. 8 Free energy landscapes of atazanavir in TIP3P solution with RESP charge parameters determined using **a** the PDB conformation (PDB ID: 2AQU) and **b** the conformation obtained by simulated annealing. (*a-1*) and (*b-1*) show the landscapes for the radius of gyration (in Å) and surface area (in Å^2), and (*a-2*) and (*b-2*) show those for the radius of gyration (in Å) and RMSD (in Å). Conformations at 300 K from the REMD simulations in TIP3P water were analyzed to draw the free energy landscape

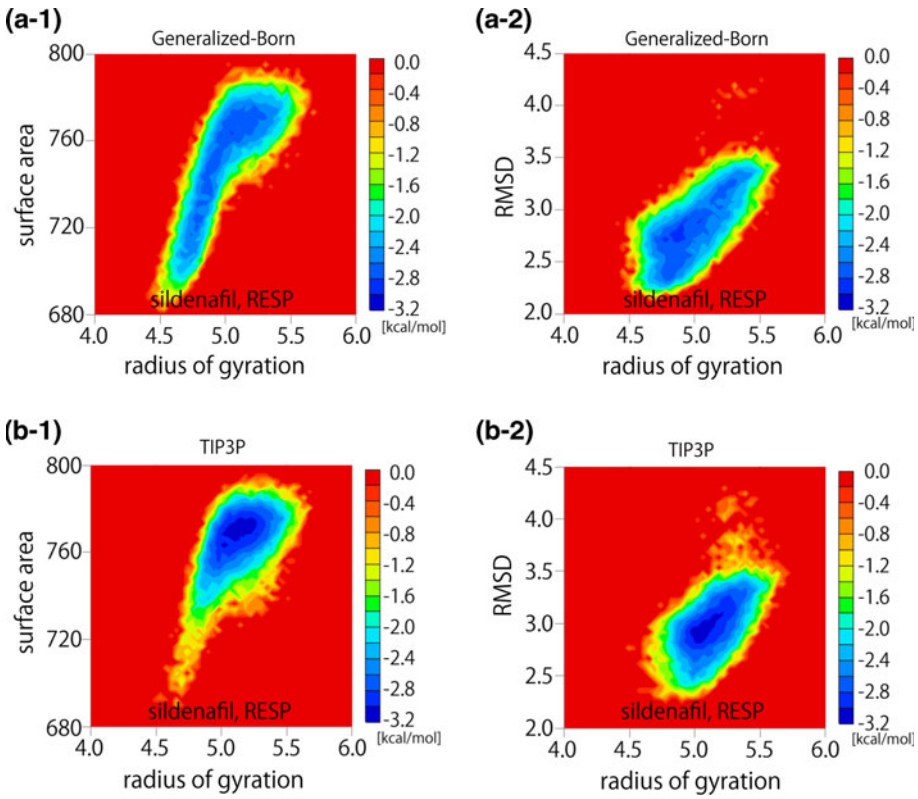

Fig. 9 Free energy landscapes of sildenafil in **a** GB solution and **b** TIP3P solution. (*a-1*) and (*b-1*) show the landscapes for the radius of gyration (in Å) and surface area (in Å^2), and (*a-2*) and (*b-2*) show those for the radius of gyration (in Å) and RMSD (in Å). The RESP charges were determined using the PDB conformation (PDB ID: 2H42). Conformations at 300 K from the REMD simulations were analyzed to draw the free energy landscape

Although Fig. 4 showed that stable conformations in vacuum were significantly different from those in aqueous conditions from the viewpoint of the average radius of gyration, the differences between the GB and TIP3P conditions were not clear. In Figs. 9 and 10 we compare the GB model with the TIP3P model. Figure 9 shows the free

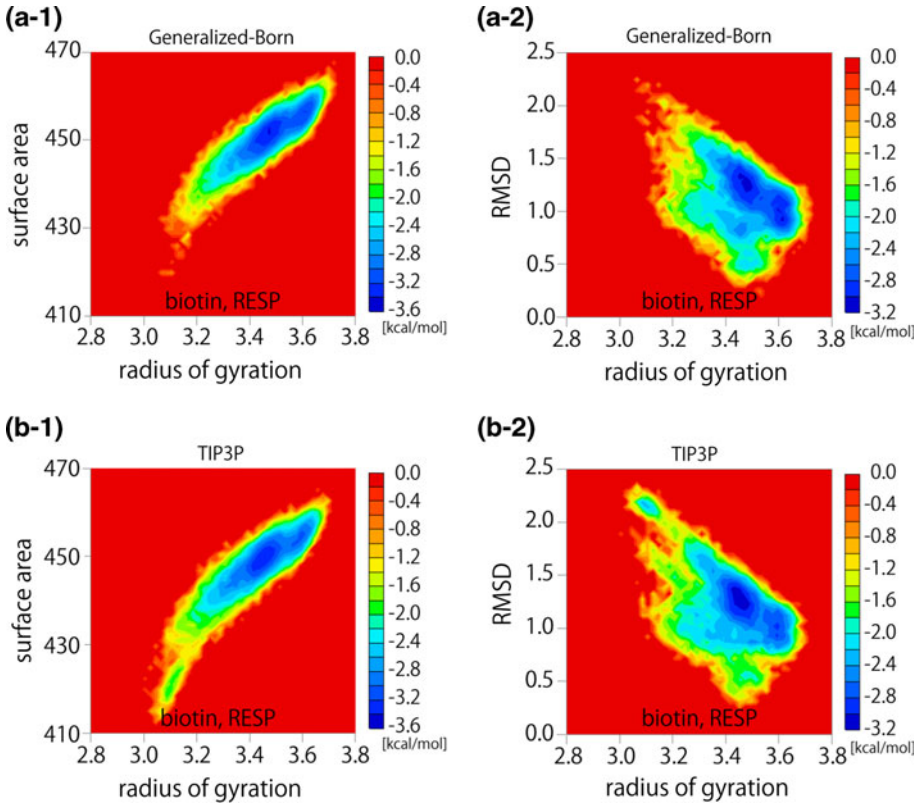

Fig. 10 Free energy landscapes of biotin in **a** GB solution and **b** TIP3P solution. (*a-1*) and (*b-1*) show the landscapes for the radius of gyration (in Å) and surface area (in Å^2), and (*a-2*) and (*b-2*) show those for the radius of gyration (in Å) and RMSD (in Å). The RESP charges were determined using the PDB conformation (PDB ID: 1STP). Conformations at 300 K from the REMD simulations were analyzed to draw the free energy landscape

energy landscapes for sildenafil using the RESP charge determined from the PDB conformation. We see that the overall landscapes are similar, but when we look in detail, Fig. 9(a-1) and (b-1) show that the region with a small radius of gyration and surface area tends to be sampled more often in the TIP3P solution than in the GB solution. Figure 10 shows the case of biotin. We also see that the GB and TIP3P models produced similar free energy landscapes. The only difference is that TIP3P simulations allow conformations with a smaller radius of gyration and surface area more often. We also examined other ligand molecules and found that the GB model well reproduced the landscapes obtained from TIP3P simulations for both AM1-BCC and RESP charges. Although more extensive studies are necessary to generalize this observation, our results suggest that the GB model can well mimic the TIP3P model for ligand free energy landscapes in water. Considering that computational cost of GB calculations is only a fraction of that of TIP3P ones, using the GB model seems an attractive option to identify ligand global and local minimum free energy states in water quickly, especially when a large number of compounds are analyzed.

We also performed GBSA simulations to include non polar solvation effects. LCPO model [39] was applied to calculate surface area. We found that the free energy landscapes obtained by GBSA simulations were almost identical to those by GB simulations for all six ligand systems. Our results indicate that non polar solvation effects estimated by LCPO model are relatively small enough not to affect ligand conformations sampled at 300 K.

Mapping the free energy landscapes for radius of gyration, surface area, and RMSD is an appropriate way to understand the sampled regions intuitively. However, it is not necessarily valid for distinguishing conformations, since different conformations can have the same radius of gyration, surface area, and RMSD. For example, biotin has a radius of gyration of 3.5 Å and surface area of 450 Å^2 when complexed to streptavidin (PDB ID: 1STP). If we only look at Fig. 10(a-1) or (b-1), it would seem that the PDB conformation is the global minimum free energy state. However, Fig. 10(a-2) and (b-2) show that a conformation with an RMSD of about 0.5 Å, which is very close to the PDB conformation, is not actually the global minimum state. Principal component analysis (PCA) is one method for distinguishing conformations by considering vectors appropriate to express large fluctuations of molecules.

Figure 11 shows the free energy landscape for the first and second principal component axes of imatinib in TIP3P water as determined by the RESP charge method from the PDB conformation. Imatinib has two different complex conformations with two proteins. The conformations shown in Fig. 11a and b are those of the complexes with

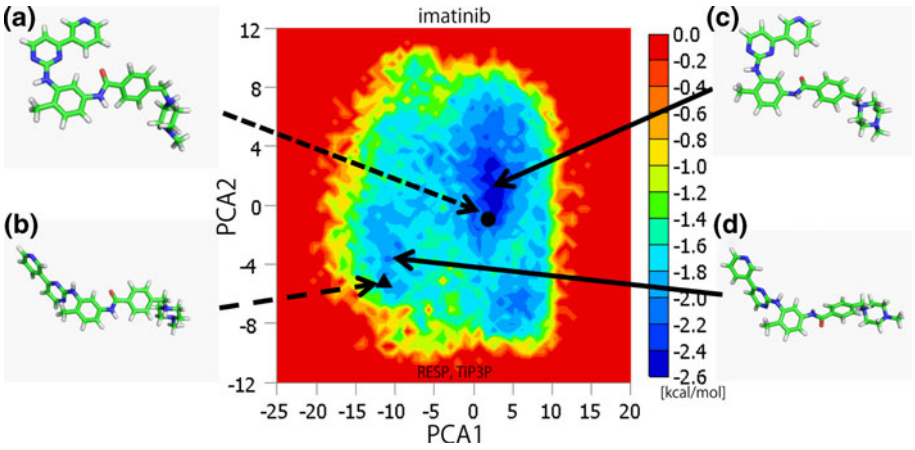

Fig. 11 Free energy landscape of imatinib in TIP3P solution with respect to the two major principal component axes, with representative conformations highlighted: **a** conformation in complex with SYK tyrosine kinase (PDB ID: 1XBB), **b** conformation in complex with ABL tyrosine kinase (PDB ID: 2HYY), **c** the global minimum free energy conformation, and **d** one of the local minimum free energy conformations. The RESP charges were determined using the PDB conformation (PDB ID: 1XBB). Conformations at 300 K from the REMD simulation were analyzed to draw the free energy landscape

SYK and ABL tyrosine kinases, respectively. The dotted arrows show where the PDB conformations exist on the free energy landscape, which were obtained by calculating the inner product with the first and second eigenvectors of the PCA. We see that the bound conformation of imatinib with SYK (a) exists in proximity to the global minimum free energy state (c). In contrast, the bound conformation with ABL (b) is in proximity to one of the local minimum free energy states (d), which is about 0.5 kcal/mol higher than the global minimum state. Thus, we see that the free energy penalty of imatinib in complex with both proteins is very small, and that these two PDB conformations were often sampled in water at 300 K without the presence of protein. It is reasonable because ligand's small free energy penalty upon binding should be preferable for high affinity for proteins.

Figure 12 shows the free energy landscapes along the first and second principal component axes for all of the ligand molecules studied here. The black circles show where the bioactive conformations complexed to proteins (from the PDB) exist on these landscapes. We found that none of the protein-bound conformations exists in the global-minimum free energy states except the case of imatinib, but that they exist near local-minimum states (Fig. 12b, e, f) or as intermediates between stable states (Fig. 12c, d). Although we see that ligands rarely bind in their lowest energy conformations [1, 2], protein-bound ligand conformations were sampled in water in the absence

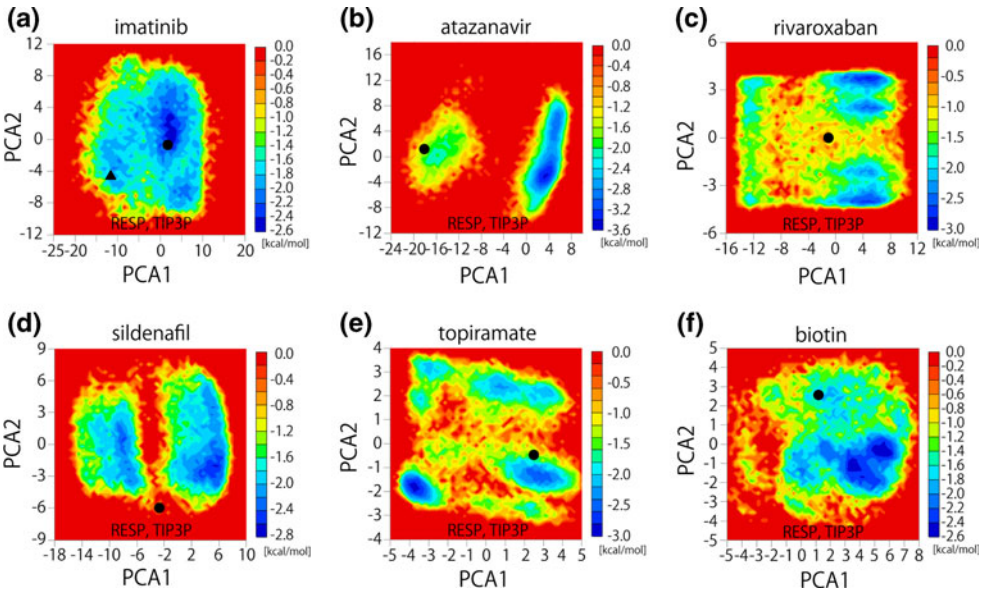

Fig. 12 Free energy landscapes with respect to the two major principal component axes for the six ligand molecules studied. The *black circles* (and *triangle* for imatinib) show where the protein-bound conformations (from PDB) lie on these landscapes. Conformations at 300 K from the REMD simulations were analyzed to draw the free energy landscape

of protein with small free energy penalties (less than 2.5 kcal/mol for all six ligand molecules). However, these protein-bound ligand conformations were not found in the free energy landscapes under vacuum conditions for most ligands (data not shown). This can be concluded from the average radius of gyration in vacuum shown in Fig. 4. The PDB conformations on these free energy landscapes are found to be similar to the conformations that exist close to the PDB conformations, as expected (compare Fig. 11a with c and Fig. 11b with d). We also confirmed the same trend for other ligands.

Figure 12 also shows that some ligands have distinct stable regions divided by relatively high free energy barriers. It suggests that ligand free energy landscapes are complicated so that conventional constant temperature molecular dynamics simulations at 300 K fail to yield accurate free energy landscapes. In order to confirm this hypothesis, we performed the constant temperature simulation of atazanavir in TIP3P water at 300 K for 20 ns, starting from the global minimum conformation, and confirmed that the local minimum state in the left region in Fig. 12b was never sampled until 12 ns, which made it difficult to obtain reliable free energy landscapes. Klimovich and Mobley also recently found this sampling problem for hydration free energy calculations [40]. On the other hand, REMD simulations sampled the local minimum state within several hundred picoseconds and did not remain in the same state for a long time. Therefore, generalize-ensemble algorithms such as REMD are necessary to obtain reliable results from molecular simulations of ligands with free energy barriers between their global and local minimum states.

Conclusions

Although the AM1-BCC model has been developed to reproduce the general results of RESP charges, the difference in ligand free energy landscapes between these two charge models had not been examined in detail. We have investigated how different charge assignments and solvent conditions affect the free energy landscapes of six ligands, all of which have the crystal structures with protein complexes registered in PDB.

First, we determined charge parameters of three different conformations for each ligand by AM1-BCC and RESP methods, and confirmed that the AM1-BCC method actually reproduced the overall trends of RESP charges in all cases. AM1-BCC charges were almost identical without dependency on ligand conformations, while the RESP method showed some conformational dependency as expected.

Second, we performed replica-exchange molecular dynamics simulations of these six ligand molecules to examine how much their free energy landscapes changed due to charge parameter differences. The simulations were performed in three solution models: vacuum, GB, and TIP3P. We found that subtle charge differences produced different free energy landscapes, especially very different in vacuum, for some ligands. However, these differences tended to be smaller in GB and TIP3P conditions. Mobley et al. also examined various charge models for hydration free energy calculations and reported that AM1-BCC method works almost as well as more computationally expensive ab initio methods [41]. Compact conformations tended to be more stable in vacuum conditions, while extended ones tended to be stable in aqueous solutions using GB and TIP3P models. The GB model reproduced the TIP3P model well in the sense that differences were observed only in the weighting of different conformational distributions and similar conformations tended to be stably sampled in both conditions. Note that the differences between the TIP3P and GB models may be important in some cases. For example, Mobley et al. recently showed that the TIP3P model produced more accurate hydration free energies than the GB model (a correlation coefficient R^2: 0.89 vs. 0.69–0.74) [42, 43]. However, considering that GB simulations can be performed with only a fraction of the computational cost of TIP3P simulations, our results suggest that GB is an efficient alternative to using the more rigorous TIP3P model. This is especially true when we need to evaluate a large number of compounds.

Third, we have examined where the protein-bound ligand conformations exist in the free energy landscapes. We found that these conformations were all included in the regions sampled under aqueous conditions. Although five ligands (excluding imatinib) did not bind in their global minimum conformations, the estimated free energy penalty for a ligand in the protein-bound state was less than 2.5 kcal/mol in all cases. On the other hand, these protein-bound ligand conformations were not found under vacuum conditions for most ligands, which implies that the free energy penalty tended to be higher than in aqueous conditions.

We conclude that a ligand's conformational strain energy, defined as the difference in internal potential energy between the protein-bound and free states, should be estimated using ensemble-averages obtained from molecular simulations in aqueous condition at around 300 K and 1.0 atm rather than by one point calculation after energy minimization or ensemble averages obtained from simulations in vacuum. In fact, this is a closer model of reality, since the free energy difference between these two thermodynamic states in aqueous solution determines the free energy of binding.

From these analyses of free energy landscapes of ligands we can roughly estimate the free energy penalties of ligand conformations by judging where conformations similar to

protein-bound states are on free energy landscapes. We hope that such information will contribute to accurate evaluations of ligand conformations obtained from docking results or conformational searches in structure-based or ligand-based drug design, though further investigations are necessary to explore these possibilities.

Acknowledgments We are grateful for the insight, discussions and collegiality provided by Dr. Masaki Tomimoto. The authors also thank Dr. Douglas Cary for reading the manuscript. The simulations and computations were performed on the TSUBAME Grid Cluster at Global Scientific Information and Computing Center of Tokyo Institute of Technology supported by the MEXT Open Advanced Research Facilities Initiative. This work was supported, in part, by Grants-in-Aid for Scientific Research on Innovative Areas ("Fluctuations and Biological Functions") and for the Next-Generation Super Computing Project, Nanoscience Program from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

1. Boström J (2001) Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools. J Comput Aided Mol Des 15:1137–1152
2. Perola E, Charifson PS (2004) Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. J Med Chem 47:2499–2510
3. Diller DJ, Merz KM (2002) Can we separate active from inactive conformations. J Comput Aided Mol Des 16:105–112
4. Boström J, Greenwood JR, Gottfries J (2003) Assessing the performance of OMEGA with respect to retrieving bioactive conformations. J Mol Graph Model 21:449–462
5. Agrafiotis DK, Gibbs AC, Zhu F, Izrailev S, Martin E (2007) Conformational sampling of bioactive molecules: a comparative study. J Chem Inf Model 47:1067–1086
6. Chen I, Foloppe N (2008) Conformational sampling of druglike molecules with MOE and catalyst: implications for pharmacophore modeling and virtual screening. J Chem Inf Model 48:1773–1791
7. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174
8. Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 25:247–260
9. Bayly CI, Cieplak P, Cornell W, Kollman PA (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 97:10269–10280
10. Cieplak P, Cornell WD, Bayly C, Kollman PA (1995) Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA and proteins. J Comput Chem 16:1357–1377
11. Jakalian A, Bush BL, Jack DB, Bayly CI (2000) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. J Comput Chem 21:132–146
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
13. Fujitani H, Tanida Y, Matsuura A (2009) Massively parallel computation of absolute binding free energy with well-equilibrated states. Phys Rev E 79:021914
14. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA (2007) Predicting absolute ligand binding free energies to a simple model site. J Mol Biol 371:1118–1134
15. Boyce SE, Mobley DL, Rockline GJ, Graves AP, Dill KA, Shoichet BK (2009) Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. J Mol Biol 394:747–763
16. Deng Y, Roux B (2008) Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. J Chem Phys 128:115103
17. Gilson MK, Zhou HX (2007) Calculation of protein-ligand binding affinities. Annu Rev Biophys Biomol Struct 36:21–42
18. Kokubo H, Pettitt MB (2007) Preferential solvation in urea solutions at different concentrations: properties from simulation studies. J Phys Chem B 111:5233–5242
19. Kokubo H, Rosgen J, Bolen D, Pettitt MB (2007) Molecular basis of the apparent near ideality of urea solutions. Biophys J 93:3392–3407
20. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 55:383–394
21. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL III (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25:265–284
22. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
23. Chemical Computing Group Inc., Montreal, Canada (2008) MOE (Molecular Operating Environment) version 2008.10
24. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
25. DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto
26. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688
27. Ryckaert J, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
28. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680
29. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
30. Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. J Phys Soc Jpn 65:1604–1608
31. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151
32. Mitsutake A, Sugita Y, Okamoto Y (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers. Peptide Sci 60:96–123
33. Teeter MM, Case DA (1990) Harmonic and quasiharmonic descriptions of crambin. J Phys Chem 94:8091–8097
34. Kitao A, Hirata F, Go N (1991) The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. Chem Phys 158:447–472
35. Garcia AE (1992) Large-amplitude nonlinear motions in proteins. Phys Rev Lett 68:2696–2699

36. Abagyan R, Argos P (1992) Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. J Mol Biol 225:519–532
37. Amadei A, Linssen ABM, Berendsen HJC (1993) Essential dynamics of proteins. Proteins 17:412–425
38. Kitao A, Go N (1999) Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 9:164–169
39. Weiser J, Shenkin PS, Still WC (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J Comput Chem 20:217–230
40. Klimovich PV, Mobley DL (2010) Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. J Comput Aided Mol Des 24 (in press)
41. Mobley DL, Dumont E, Chodera JD, Dill KA (2007) Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. J Phys Chem B 111:2242–2254
42. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA (2009) Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations. J Chem Theory Comput 5:350–358
43. Mobley DL, Dill KA, Chodera JD (2008) Treating entropy and conformational changes in implicit solvent simulations of small molecules. J Phys Chem B 112:938–946