

Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations

Kai Liu¹ · Etsuro Watanabe¹ · Hironori Kokubo¹

Received: 19 August 2016 / Accepted: 22 December 2016 / Published online: 10 January 2017
© Springer International Publishing Switzerland 2017

Abstract The binding mode prediction is of great importance to structure-based drug design. The discrimination of various binding poses of ligand generated by docking is a great challenge not only to docking score functions but also to the relatively expensive free energy calculation methods. Here we systematically analyzed the stability of various ligand poses under molecular dynamics (MD) simulation. First, a data set of 120 complexes was built based on the typical physicochemical properties of drug-like ligands. Three potential binding poses (one correct pose and two decoys) were selected for each ligand from self-docking in addition to the experimental pose. Then, five independent MD simulations for each pose were performed with different initial velocities for the statistical analysis. Finally, the stabilities of ligand poses under MD were evaluated and compared with the native one from crystal structure. We found that about 94% of the native poses were maintained stable during the simulations, which suggests that MD simulations are accurate enough to judge most experimental binding poses as stable properly. Interestingly, incorrect decoy poses were maintained much less and 38–44% of decoys could be excluded just by performing equilibrium MD simulations, though 56–62% of decoys were stable. The computationally-heavy binding free energy calculation can be performed only for these survived poses.

Keywords Molecular dynamics simulation · Protein–ligand binding prediction · Docking · Structural stability · Computer-aided drug design

Introduction

Drug discovery is known as a highly time-consuming and expensive process. Facing the grand challenge in front of pharmaceutical industry, Paul et al. demonstrate that target selection is one of the two key approaches to improve R&D efficiency and reduce the huge costs, especially those of Phase II and Phase III [1]. The target selection, which is at the early stage of drug discovery, highlights the importance of the predication and validation of druggable targets. During this lead compound generation period, *in silico* technologies, especially docking and scoring in virtual screening, play an essential role for their relatively higher efficiency and lower costs as compared with traditional experiment [2–4].

Although the docking and scoring are widely used, they still fall short of the robust reliability requirement from drug discovery [5–7]. Consequently, a tremendous amount of efforts have been made to improve the performance of both the docking power and scoring functions. In order to evaluate the various docking software, many benchmarking databases have been developed, such as the Community Structure Activity Resource (CSAR) [8] and Directory of Useful Decoys (DUD) [9] and Comparative Assessment of Scoring Functions (CASF) [10]. Recently, Lagarde et al. have well reviewed the evaluation of these data sets [11]. In addition, many researches have examined the various docking software and scoring functions [5, 6, 12–14]. The scoring functions can be largely responsible for the failure of pose prediction, especially

Electronic supplementary material The online version of this article (doi:10.1007/s10822-016-0005-2) contains supplementary material, which is available to authorized users.

✉ Hironori Kokubo
hironori.kokubo@takeda.com

¹ Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, 26-1, Muraoka-Higashi 2-chome, Fujisawa, Kanagawa 251-8555, Japan

for the self-docking case, in which the native ligand is docked back to its cognate protein [15, 16]. However, the sampling problem (including solvation and entropic effects), which has been highlighted by MM/GB(PB)SA calculations based on simulation snapshots [17–21], can also be one reason of the incorrect pose predictions. Unfortunately, the sampling problem of MM/GB(PB)SA calculation is not clearly illustrated, and the necessary number of snapshots, for example, is still a controversial issue. Although the statistical precision of MM/GB(PB)SA can be achieved by a large number of independent snapshots obtained from a long time molecular dynamics (MD) simulation, some studies have shown that calculation based on single minimized structures can give as good or even better result as MD in some cases [22–24]. The single snapshot calculation has emphasized the importance of the initial structure, especially for the poses from docking software. Therefore, it is necessary to examine the stability of initial structure of complex, particularly for the structure from docking.

The examination of stability by MD simulation strongly relies on the accuracy of the force-field parameters. However, to the best of our knowledge, it has not been fully investigated how much modern MD simulations can maintain experimental protein–ligand binding modes properly, although some attempts have been made [25, 26]. In this article we have systematically investigated the stability of ligand binding poses on a set of complexes with drug-like ligands by analyzing the ligand conformation and positional changes under MD simulation. In other words, our current interest focuses on the relative stability of various correct and incorrect ligand poses under MD simulation. As the first step, we focus on self-docking, which means the native ligand is docked back to its cognate protein. We then evaluate the relative stability of various ligand binding poses (correct/decoy poses) and examine the possibility of filtering out the unfavorable poses by equilibrium MD simulations. The calculation based on cross-docking would be discussed elsewhere.

The rest of this article is organized as follows. In “[Materials and methods](#)”, the procedures to build a drug-like ligand data set and various docking poses are explained. The details of MD simulations are also described there. In “[Result and discussion](#)”, the stabilities of ligand poses under MD are evaluated by comparing the root mean square deviation (RMSD) values with the corresponding poses from docking. It is shown that MD simulations are accurate enough to judge most experimental binding poses as stable properly, while decoy poses are much less stable. The possible reasons of some inconsistencies with the experimental data are also discussed. Summary and conclusion is devoted to the summary and the implications of our results for drug design.

Materials and methods

Collection of complexes

The drug-like ligands were initially selected from sc-PDB database (release 2013) [27, 28], which provided convenient options based on physicochemical properties of ligand. We set our criteria for drug-like molecule as listed in Table 1 according to Refs [29–33]. Based on these criteria, the ligand data set, which consists of 2345 complex structures out of 9283 entries from sc-PDB database, was selected for further refinement. Meanwhile, the binding affinity of the targets was an indispensable property of our interests. Therefore, those PDB files, which lack binding information (from the Binding MOAD database, release 2014 [34, 35]), were discarded. After that, 1039 PDB files were obtained.

Considering the expensive cost of our calculations on this large number of PDB files, the number of PDB files was further refined. We built a prior data set based on the cross-checking of our ligand data set with three available and widely used data sets for docking: Iridium [36], CSAR, and CASF data sets, in which the structure qualities of the complexes had been carefully examined. We then obtained 85 overlapped PDB files, 16 from Iridium data set (highly and moderately trustworthy set, release Feb. 13, 2014, 224 PDB files), 48 from CSAR data set (NRC-HiQ, release Sept. 24, 2010 and 2009–2011 update, 466 PDB files), and 21 from CASF data set (the core set, release 2013, 195 PDB files), respectively. The reasons for the low overlap (about 10%) among data sets were mainly attributed to: (1) The interest of each data set was not the same. For example, many small molecules (MW less than 200) were included in both CSAR and CASF data set. (2) The completeness of each data set was different. Although all the PDB files were from Protein Data Bank (PDB) [37], the number of entries included in each database was much different.

In addition to the prior data set of these 85 PDB files, some candidates from the remaining PDBs were rescued to enrich the ligand data set. Along with the general rule of thumb for high quality crystal structure selection [8, 10,

Table 1 The criterion of physicochemical properties of ligand [29–33]

Descriptor	Values
Molecular weight (MW)	$200 \leq MW \leq 460$
AlogP	$-1 \leq AlogP \leq 4.2$
Rotatable bonds	≤ 7
Number of rings	≤ 4
Number of H-bond donor (HBD)	≤ 4
Number of H-bond acceptor (HBA)	≤ 8

36], we further refined the ligand data set based on protein structural qualities, such as resolution, completeness of residue chains and lower B factor of proteins. Considering the efficiency of the calculations by MD, in addition, the complexes with large number of amino acids (>500) were also discarded. Consequently, 35 PDB files were added to our test set, and 120 PDB files were finally collected for the following calculation (Supporting Information, Table S1).

Complex preparation and initial poses generation

The structure of each complex, downloaded from PDB, was prepared by LigX module of MOE (release 2014. 09) [38]. The potential structural problems of amino acid residues in complex were automatically corrected and capped by ACE and NME for two terminals separately. Then the complex was protonated and optimized with constraint on whole complex by default settings in QuickPrep module with AMBER10 EHT force field. In addition, the atomic type of protonated ligand was carefully compared with that from sc-PDB and PDB. Only cofactor and ions within 6 Å from ligand were maintained, and others were removed.

Except the native pose from PDB files, other three docking poses were generated by Glide docking (version 67011, Schrödinger) [39]. All crystal waters were removed for docking since it is difficult to know whether these waters are displaceable without pre-knowledge. On the other hand, all the crystal water molecules were kept for the simulations of the native poses.

Self-docking of Glide standard precision (GlideSP) was used to obtain three docking poses for each protein structure. Except the default settings of GlideSP docking, the value of RMSD deviation was set to 2.0 Å. This was set as a threshold to enrich various possible binding poses [20]. After poses were ranked by Glide gscore, the RMSDs (heavy atom only) between generated poses and the native pose of ligand from crystal structure were calculated. According to the values of RMSD and gscore, three kinds of poses were defined (pose0: the closest to the native pose and its value of RMSD change is less than 2.0 Å, pose1: the smallest gscore pose with the RMSD larger than 2.0 Å from both pose0 and the native pose, and pose2: the smallest gscore pose with the RMSD larger than 2.0 Å from all of native, pose0, and pose1). The pose1 and pose2 can be considered as two reasonable decoys. We consider that 2.0 Å is a reasonable threshold to judge the similarity of the ligand binding poses.

MD preparation

We here have four different ligand binding poses for each complex, named the native pose and other three docking poses (pose0, pose1 and pose2). The ligand was

parametrized by RESP charge from Gaussian 09 [40] calculation with HF/6-31G(d) basis set, while the parameters (AM1-BCC charges) of cofactor were generated by default *sqm* program [41]. The AMBER14 ff14SB force field [42] was employed for amino acid residues. The parameters for Mg^{2+} and Zn^{2+} were from AMBER parm99.dat, while those for Mn^{2+} and Ca^{2+} came from data of Bradbrook et al. [43]. These parameters have been widely used and tested by AMBER user community. For other ions (for example, Ni^{2+} , Co^{2+} and Fe^{3+}), the parameters of 12-6-4 potential [44] were used. The complex was finally solvated in a TIP3P cuboid water box with at least 8 Å away from any protein atoms. Note that the crystal water existed only in native poses.

We used the same protocol for all MD calculations as follows. The particle mesh Ewald molecular dynamics simulation (PMEMD) was used for the electrostatic calculation and the SHAKE algorithm was employed for the constraint of hydrogen atoms. A default cutoff (8 Å) of intermolecular interaction is used to limit direct space sum in PMEMD. After 1000 steps of optimization (500 cycles of steepest descent and 500 cycles of conjugate gradient minimization) with weak constraints (1.0 kcal/(mol Å²)) on the heavy atoms of a protein/ligand complex, the whole box was again optimized for 1000 cycles without any constraint (500 cycles of steepest descent and 500 cycles of conjugate gradient minimization). Then, the complex was gradually heated to 300 K by Langevin thermostat for 200 ps in NVT ensemble with a weak constraint (0.5 kcal/(mol Å²)) on the whole complex.

It is important to consider the dependency on initial velocities when we judge the stability of the simulations [45]. We thus performed five independent simulations with different initial velocities since the heating process. For each MD run, the box was further equilibrated with the same weak constraint (0.5 kcal/(mol Å²)) on the complex for 500 ps under 1 atm, controlled by the Berendsen coupling algorithm. Finally, 10 ns production in NPT ensemble was performed without any constraint. During the production run, the time step and collection interval was set to 2 fs and 10 ps, respectively. In order to minimize the human fault, all the settings of MD simulations were prepared and generated by the script automatically.

Result and discussion

Basic features of the ligand data set

Figure 1 shows the selected properties of ligand molecules from the collected 120 PDB files. The wide distributions of these physicochemical properties imply the generality of our ligand data set. The diversity of ligand

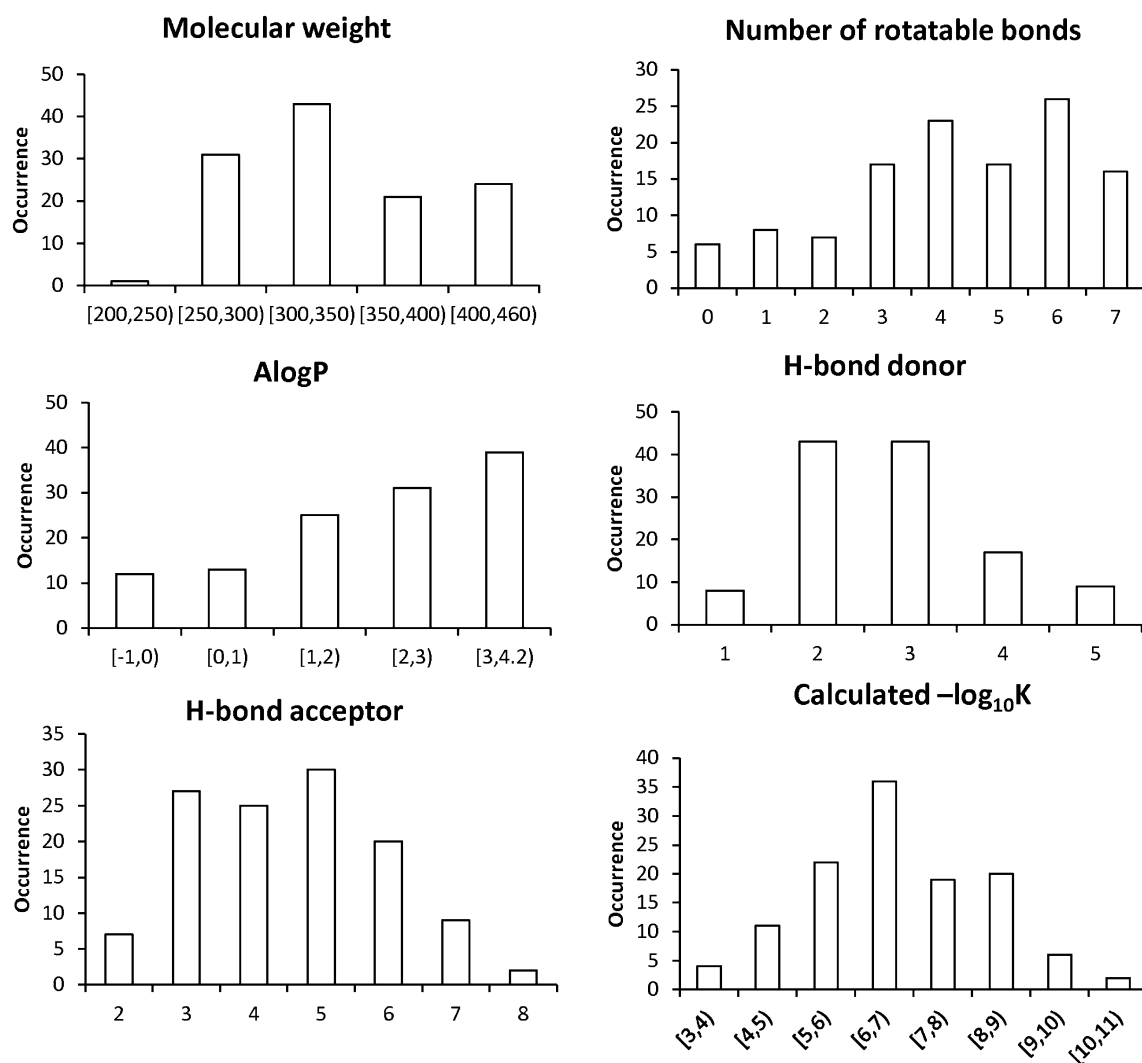


Fig. 1 Selected physicochemical properties of the data set

molecules was additionally calculated by Tanimoto similarity, which was only 0.04 for all ligands. The proteins in our data set consisted of kinase, enzyme and non-enzyme and so on. By examining the information from DrugBank [46], we found that 49 proteins (41%) were known as drug targets, for which small-molecule drugs were already on the market, such as the Tyrosine-protein kinase JAK2, Heat shock protein HSP 90-alpha, and Estrogen receptor I. 39 proteins (32%) were of potential pharmaceutical interest and some of them (11 targets) were under clinical trials. No record was found for the remaining 32 proteins (27%) (Supporting Information Table S1). We also accessed the binding affinity of complexes in our data set. The average $-\log_{10}K$ of the data set was 6.7, which corresponds to the typical activity level of lead compound (Because of the different activity

detection in experimental methods, here the binding affinity is roughly evaluated by $-\log_{10}K$ ($K = K_d, K_i$ or IC_{50})).

Initial poses obtained from self-docking

We obtained three different docking poses for each complex as explained in Materials and Methods. Table 2 shows the average RMSD of all docking poses. Compared to the native pose, the average RMSD values (docking power) of pose0, pose1 and pose2 were 0.7 ± 0.4 , 4.2 ± 1.9 and 5.0 ± 2.0 Å, respectively. The small RMSD deviation of pose0 shows that it was actually very close to the experimental pose.

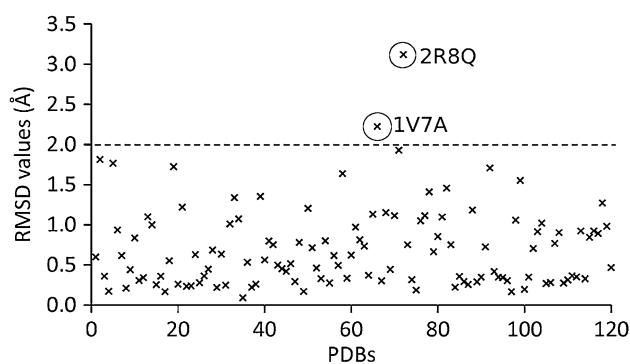
Table 2 also lists the coverage ratio, which was defined as the ratio that the closest pose to the native one was included in the top N docking poses. We see that only

Table 2 RMSD and rank of pose0, pose1 and pose2 from Glide docking

Pose	Rank of Glide score ^a	Coverage ratio ^b	Average RMSD of all PDBs (Å)	Standard deviation of RMSD (Å)
Pose0	TOP 1	42% (50/118)	0.7	0.4
	TOP 5	74% (87/118)		
	TOP 10	90% (106/118)		
Pose1	TOP 1	24% (29/120)	4.2	1.9
	TOP 5	87% (104/120)		
	TOP 10	94% (113/120)		
Pose2	TOP 1	0% (0/117)	5.0	2.0
	TOP 5	62% (73/117)		
	TOP 10	85% (99/117)		

^aTOP N: means the top N poses, ranked by Glide score

^bThe coverage ratio is the ratio of the closest pose to the native one that is included in the TOP N candidates

**Fig. 2** Distribution of RMSD values of pose0 from Glide self-docking

50/118 complexes (42%) were correctly ranked as the best for pose0 (pose0: the closest to the native pose). If taking the top 10 poses into consideration, 106/118 cases (90%) had the pose0 within these top 10 poses. In contrast, 29/120 complexes (24%) of pose1 were incorrectly ranked as the best pose based on Glide gscore function. Note that RMSDs of pose1 were more than 2.0 Å from the native pose. The coverage of pose1 and pose2 in top 10 was 113/120 (94%) and 99/117 (85%), respectively. These high ratios of pose1 and pose2 in top 10 suggest that these poses were comparable with pose0 in terms of the docking score. Therefore, we consider that these poses are reasonable and challenging decoys.

Figure 2 presents the RMSD value distribution of pose0 for all ligands. Two cases (The RMSD of 2R8Q and 1V7A were 3.1 and 2.2 Å, respectively) were excluded for the following analyses on pose0 because their RMSDs were larger than our criterion (2.0 Å) for the closeness. The RMSDs of pose0 of the remaining 118 PDBs were all less than 2.0 Å.

It suggests that the correct binding mode was sufficiently covered by GlideSP self-docking. For the detailed information of all the calculated RMSDs, see Table S3 in Supporting Information. Note that there were no pose2 for three complexes (2B1Z, 3V49 and 4GS9), because the RMSD values of them did not satisfy our criterion mentioned above.

Results from MD simulation

We now analyze the stability of complex by MD simulation statistically. In order to measure the statistical stability of the system, five independent simulations with different initial velocities are performed for each pose, starting from the same energy-minimized structure. Because the simulation trajectory depends on the initial velocities [42], it is difficult and risky to judge the stability of the ligand binding mode only from one simulation. Therefore, the binding mode is considered to be unstable during the simulation only when all five independent simulations fail to keep the initial binding modes.

We thus performed 2375 MD simulations in total for the four kinds of poses (120 complexes with native pose, 118 complexes with pose0, 120 complexes with pose1 and 117 complexes with pose2) with five different velocities. Among these 2375 MD simulations, more than 97% (2318/2375) MD runs have been normally finished. This high completion rate indicates the generality and reliability of the current protocol for MD simulation. The failed cases were attributed to the shake error during the simulations. The nature of this error would be due to the unfavorable initial geometry of the ligand inside, because all the calculations on the complexes with the native poses were normally terminated.

We simply measured the convergences after the 10 ns production simulations by the RMSD deviations of protein backbone (BB), shown in Fig. 3. We see that after the 10 ns production step the complexes were all equilibrated, except the only one case of 3QPN with pose2 (the average deviation of its five simulations is 5.1 ± 0.1 Å).

Figure 4 shows the percentage of acceptable MD runs for all four kinds of poses by the MD simulations. We define a pose as stable (“Acceptable”) by simulation when any simulation among the five parallel MD runs has the RMSD less than 2.0 Å, where the RMSD of each simulation is the average value of the last 500 snapshots (these snapshots come from the last 5 ns, the collection interval is 10 ps). The reference structure of the RMSD calculation is the first snapshot, which comes from the docking pose after the energy-minimization with constraint as described in “MD preparation”. Considering even the correct pose can experience the unbinding event depending on the initial velocity assignment, the RMSD value can be apart more

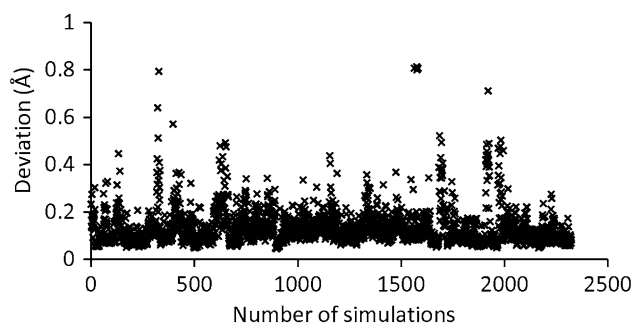


Fig. 3 Distribution of the deviation of RMSD for the protein backbone (BB) of complexes in all simulations, except pose2 of 3QPN, the deviation of which is 5.1 ± 0.1 Å among five parallel runs. For the 10 ns production step, the first half (5 ns) are discarded. Then the time-scale RMSDs of the backbone (BB) are calculated based on the final 500 snapshots (the collection interval is 10 ps). Finally, the convergence of the MD simulation is evaluated by the deviation of RMSDs of backbone (BB)

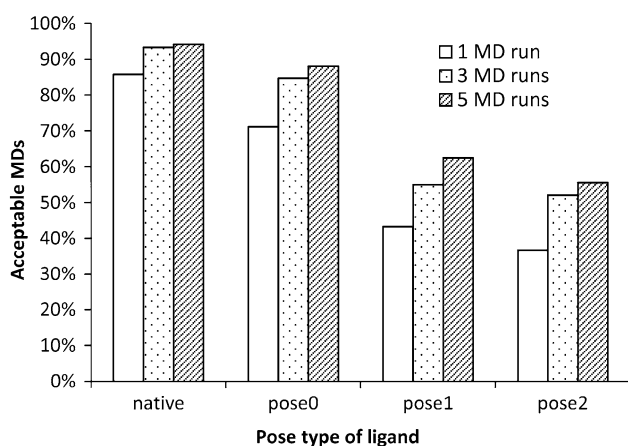


Fig. 4 Percentages of acceptable MD runs for all the pose types. The acceptable ratio is defined as (the number of poses that was considered as stable)/(the number of the poses). Detailed values are also presented in Table S4 in Supporting Information

than 2 Å in that case. Therefore, a pose is judged as unstable only when all five independent runs have the RMSD larger than 2.0 Å.

The analyses based on the first, first three and all five MD trajectories were initially compared. As the number of MD run increased, the number of acceptable MD also increased as it should be. The difference in the acceptable number by performing one, three or five simulations should be attributed to the overestimation in the protocol. The incorrect pose was considered to be stable, for example, if the RMSD change of any one among the parallel MD runs was accidentally less than 2.0 Å. The reason of allowing this overestimation was that we would like to include the correct binding pose (pose0) as much as possible. Instead, when only one simulation was performed, more decoys

were excluded, but the risk of excluding correct binding pose (pose0) was also increased. At the current step, we think that keeping the correct binding poses is more important than excluding more number of decoys because the remaining decoys can be further discriminated by other methods, for example, the more accurate but expensive binding free energy methods [47].

In the case of the native poses, about 94% (113/120) ligand poses were judged as stable by simulations. We also see that about 88% (104/118) of pose0, 63% (75/120) of pose1, and 56% (65/117) of pose2 were stable. It is understandable that the simulation could maintain the binding modes with high probability for the correct binding poses (native and pose0). The relatively low stability for the incorrect binding modes (pose1 and pose2) is also reasonable. This finding is interesting because it indicates the possibility that we can exclude some of wrong binding poses only by performing simple equilibrium simulations, while keeping the correct poses.

More detail information is available in Tables S3 and S4 in Supporting Information. Table S3 lists the stability of complexes by MD simulations. We see that the backbone and core residues were very stable for most cases, while the stability of the ligands differs very much among native, pose0, pose1, and pose2 as we have seen in Fig. 4 above. This result confirms that protein geometries from the high resolution crystal structures are reasonable and the simulations can maintain these structures.

We next examine the failed cases for native poses, where the ligand poses were judged as unstable by all five independent MD runs. There were only 7 failed cases among the 120 native poses by the current defined threshold (RMSD < 2.0 Å): 4BFZ (2.0 Å, which is the minimum RMSD value of ligand under 5 parallel MD runs), 4JP9 (2.1 Å), 4ANP (2.1 Å), 2BYH (2.4 Å), 4KP5 (3.0 Å), 3IOF (3.2 Å) and 1IA1 (4.1 Å). The slightly large RMSD values of 4BFZ and 3IOF might be attributed to the flexibility of protein side chain, because the RMSDs of core residues of these systems were also relatively large (1.4 and 1.7 Å for 4BFZ and 3IOF, respectively, see Table S4 in Supporting Information for details). For the case of 4ANP complex, the large fluctuation of ligand may be because of the weak strength of binding ($-\log_{10}K = 3.7$).

The intrinsically disordered fragment of a ligand might be also one reason. We observed that the flexible pyrimidine group of 4KP5 resulted in the large RMSD change of ligand under MD (see Fig. 5a). After excluding the pyrimidine group from RMSD calculation, the RMSD value was reduced to 0.9 Å. Because of this flexible pyrimidine group, the large RMSD change (1.9 Å) of ligand was also found in the corresponding calculation on pose0.

Another reason for the failed cases may be the unsuitable parameter of force field. Figure 5 b, c shows the ligand

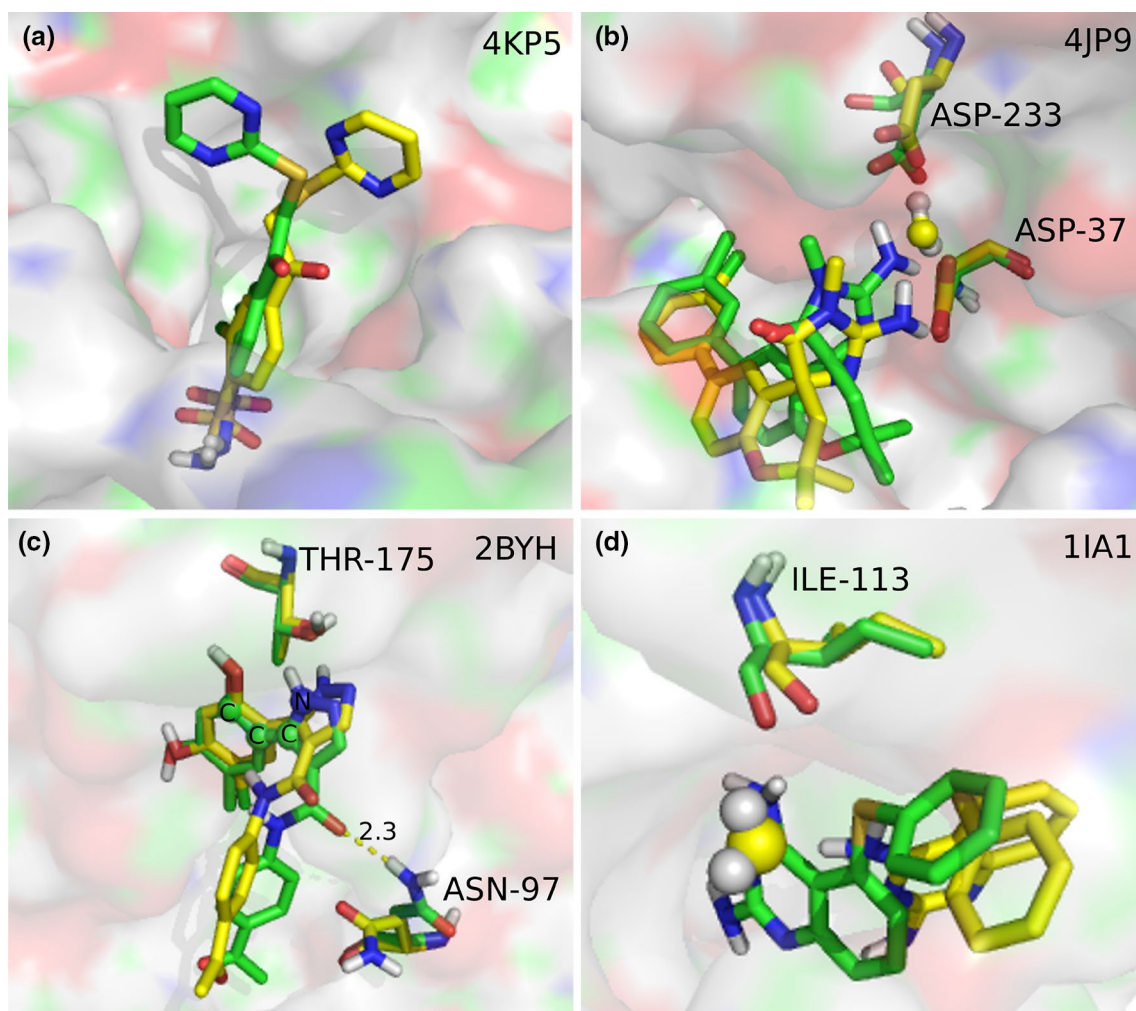


Fig. 5 **a** Complex of 4KP5 before MD (green) and after MD (final snapshot, yellow), **b** complex of 4JP9 before MD (green) and after MD (final snapshot, yellow). The H-bond between ligand and Asp233 before MD is replaced by the H-bond through bridging water after MD. **c** The H-bond between ligand and Asn97 in 2BYH complex (green) is lost after MD, and a new H-bond is formed between ligand

and THR175 (final snapshot, yellow). The dihedral angle of N–C–C–C is changed from -27.4° before MD (in crystal structure) to -70.5° after MD. **d** Complex of 1IA1 before MD (green) and after MD (final snapshot, yellow). The bridging water (yellow sphere) was trapped in pocket after MD. Complex name is shown at upper right of each figure, and unit of bond length in (c) is angstrom

pose changes before and after MD simulation of 4JP9 and 2BYH, respectively. The differences of poses were due to the breaking/forming of the hydrogen bond (H-bond) between the ligand and the protein. Especially for the case of 2BYH, the dihedral angle N–C–C–C of ligand was changed from -27.4° in crystal structure to -70.5° after MD simulation. In the corresponding calculations on the complex with pose0, the initial structures of them were very close to their native poses (only 0.3 and 0.2 Å for 4JP9 and 2BYH). But the similar H-bond breaking/forming behaviors were again found after MD simulations.

Compared to the result of 1IA1 with pose0, the behavior of 1IA1 with native pose was strange. The initial geometry of ligand was almost the same between the native pose and the pose0 from docking (RMSD=0.2 Å). After the MD

simulation, however, the values of RMSD changes were 4.1 and 1.5 Å for native pose and pose0, respectively. The difference in simulation between the native pose and pose0 was that the crystal water only existed in the native pose. In addition, a bridging water molecule between the ligand and Ile-112 was only observed in the case of the native pose after MD, shown in Fig. 5d. These results suggest that the initial position of water molecules (crystal water or water generated by AMBER) may be inappropriate.

Because pose0 is the closest to native one in definition, similar behavior of pose0 to that of native pose under MD should be generally expected. Although, this expectation was satisfied for most cases, there were 13 cases, where the RMSDs were larger than 2.0 Å. Table 3 lists the RMSD values of these cases.

Table 3 RMSD value changes (Å) of selected pose0 and the corresponding native pose

PDB ID	Pose0			Native pose	
	Ligand after docking ^a	Ligand after MD ^b	Core ^c	Ligand after MD ^b	Core ^c
2AIA	1.2	2.0/2.5	2.0	1.9	0.9
1L2S	1.8	2.0/0.7	1.5	0.7	0.7
3SFF	0.7	2.0/2.0	1.1	0.8	0.9
1NR6	1.1	2.0/1.7	1.2	1.3	1.4
3DD8	0.9	2.1/1.7	1.1	1.7	0.7
1DY4	1.0	2.2/1.5	0.6	1.9	0.7
2HD6	1.2	2.2/2.4	0.7	1.6	0.6
1TOW	1.8	2.3/1.3	0.8	1.0	0.7
1H46	1.1	2.3/2.2	0.7	1.7	0.6
3QPN	1.2	2.4/0.6	0.7	0.8	0.6
1JJE	0.4	3.4/3.6	1.1	1.9	1.1
3EKR	1.7	3.5/3.2	1.1	0.8	0.6
2QBS	0.6	5.4/NC	0.7	1.5	0.7

^aRMSD value is compared to the native pose

^bLigand means the RMSD of ligand is compared to the pose before MD (pose0 from docking or native pose from crystal). The value of A/B means that the reference of A after MD is pose0 from docking, while that of B is native pose. NC means not calculated

^cCore means the residues are within 6 Å away from ligand

In the cases of 1L2S, 1TOW, and 3EKR, the RMSDs of pose0 obtained from the docking were larger than 1.7 Å. The fact that these initial poses were not sufficiently close to the native poses may have caused the large RMSDs (>2.0 Å) after MD runs for these complexes. For the complex of 2AIA, 1DY4, 2HD6, 1H46, 1JJE and 3DD8, the larger RMSD changes (>1.6 Å) were observed also in the native poses. This suggests that even the native poses tended to be unstable under simulations and these behaviors may be due to the instability of the binding modes itself and/or the insufficient parameters of force field and simulation time.

We also compared the RMSDs not only with the initial poses of pose0 but also with the native poses in Table 3. It is interesting that the RMSDs to the native poses were less than 2.0 Å after MD runs in the six systems (1L2S, 1NR6, 3DD8, 1DY4, 1TOW, and 3QPN) though the corresponding RMSDs to the pose0 after docking were large than 2.0 Å. This result suggests that the MD simulation improved the binding poses obtained from docking, and the large RMSDs to pose0 after MD runs were obtained because the pose0s were not sufficiently similar to the native poses. This result implies that MD simulation can improve the docking poses in some cases.

Another possible reason for the failed cases of pose0 should be the influences of crystal water molecules because

the crystal waters were kept in the native pose calculations, while they were absent for the calculation of pose0 in our study (see Method section for detail protocols). We investigated the system of 2QBS as an example, where the RMSD of pose0 after MD [2QBS, RMSD=0.6 Å before MD (from docking), RMSD=5.4 Å after MD] was the largest. We found that one important bridging crystal water molecule was replaced after MD, shown in Fig. 6a, which caused the large RMSD changes after MD. The direct H-bond interaction between ligand and arginine residue resulted in the large change in the position of the hydrophobic motifs of the ligand. To confirm the effect of this water molecule we merged the crystal water into the complex structure of pose0, and performed the MD simulations similarly. These simulations confirmed the importance of this crystal water in 2QBS. The carboxylic group maintained at the similar position of the crystal structure through the H-bond with the bridge water in all 5 MD runs (Fig. 6b). This result emphasized the bias to remove the important crystal water molecule from the MD for pose0 calculation. The important water molecules can be placed before MD simulation by the analysis based on the three dimensional reference interaction site model (3D-RISM) [48] or grand canonical Monte Carlo simulation (GCMC) [49, 50], though the difference in the acceptable MD runs between native pose and pose0 was small in the current analyses as shown in Fig. 4.

The percentages of RMSD values larger than 2.0 Å for decoy poses (pose1 and pose2) were 38% (45/120) and 44% (52/117), respectively. These percentages are clearly larger than 6% (7/120) and 11% (13/118) for native and pose0, respectively. This observation is reasonable because pose1 and pose2 are incorrect binding modes. It indicates that MD simulations have the capability to distinguish incorrect binding modes from correct binding modes. Our current investigations suggest that if we use 2.0 Å as the threshold, 38–44% incorrect binding poses can be excluded just by performing equilibrium MD simulations. Although 6–11% correct poses are excluded improperly, this percentage can be further decreased by preparing binding modes closer to the native ones and by placing presumed initial water molecules by 3D-RISM or GCMC before MD (as shown in the case of 2QBS above).

In addition, if one uses 2.5 Å as threshold value, the percentage of the wrongly excluded correct binding poses is decreased to only 3%, though the percentage of properly excluding incorrect poses is also decreased to 19–24%. Such a criterion (2.0 Å) is actually adopted in our in-house protein–ligand binding free energy prediction tool to discriminate improper binding poses before the heavy free energy calculation and the detail will be described elsewhere.

In our current work, only self-docking was employed. Thus, we investigated the differences of the stability of

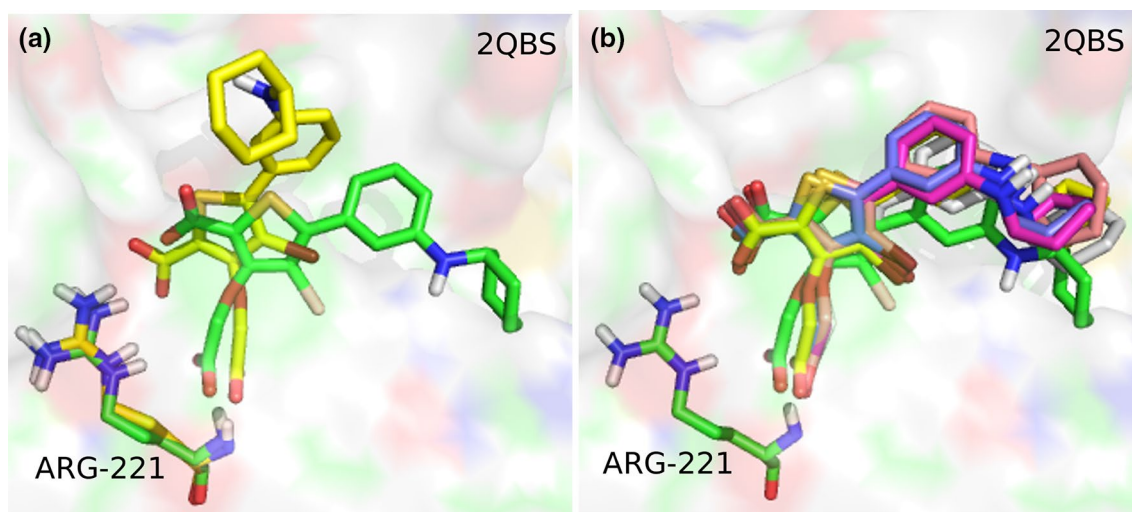


Fig. 6 Influence of crystal water molecule to the pose0. **a** Position of pose0 in 2QBS before (green) and after MD simulation (final snapshot, yellow) without including a crystal water molecule. **b** After including the crystal water molecule in simulations, the pose0s (final

snapshots of 5 independent MD runs) are stable under MD runs and close to docking pose0 in all five parallel MD runs. Complex name is shown at upper right of each figure

various correct and decoy poses by MD simulation. Our strategy is also extendable to the cases, when the complex belongs to the same local minimum around the crystal structure and sampled as one of the fluctuating structures during MD simulations. It would be also interesting to investigate cross-docking case, though it may be more difficult for the systematic study, considering the influences from the various structural differences (e.g. side-chain orientation, DFG-in/out, different folds) among template structures that we can select from the crystal structures or simulations. These selections may be somewhat arbitrary for the systematic investigation and can be strongly system-dependent.

Although our results were based on self-docking, some useful suggestion in practical usage may be obtained. The docking result showed that it was risky to select only the best scored pose as the correct binding mode, because the coverage ratio of the best scored poses was less than 50% in the self-docking. Therefore, more docking poses should be taken into consideration. If the top 10 scored poses were included here, for example, the coverage ratio was dramatically improved to 90%. Accordingly, the exclusion of incorrect binding poses will become crucial and complicated. Different stabilities of these docking poses were observed under MD, as shown in Fig. 4. The different performances between correct poses (pose0) and incorrect poses (pose1 and pose2) under MD indicated that simple equilibrium MD simulation is useful to exclude some incorrect poses, though many incorrect poses survived stably by the current procedure. Therefore, the further discrimination by more accurate methods such as the binding free energy

calculation will be required to determine the correct poses eventually.

Summary and conclusion

In this paper, we investigated how much the correct ligand binding poses were maintained as stable poses and how much the incorrect poses were judged as unstable by the modern MD simulations. The possibility to distinguish the correct and incorrect docking poses only from the independent parallel MD simulations was also investigated.

First, the high-quality data set of 120 complexes was created, which satisfied the typical physicochemical properties of the drug-like ligands. In addition to the native pose from the crystal structure, three binding poses (pose0, pose1, and pose2) were generated by docking software for each complex. Here, pose0 was the closest pose to the native one, and pose1 and pose2 were decoys which differed substantially from pose0. The high ranks of pose1 and pose2 in top 10 candidate poses from docking score suggested that these were challenging decoys.

Five parallel 10 ns MD simulations with different initial velocities were then performed for each complex. We found that about 94% (113/120) native poses and 88% (104/118) pose0 were recognized as stable by the simulations. It is understandable that the simulations maintained the correct binding modes (native and pose0) with high probability. This result suggests that the modern molecular dynamics method can properly perform the stable protein–ligand equilibrium simulations for most cases.

On the other hand, the percentages of RMSD values larger than 2.0 Å for decoy poses (pose1 and pose2) were 38% (45/120) and 44% (52/117), respectively. These percentages are clearly larger than 6% (7/120) and 11% (13/118) for native and pose0, respectively. It suggests that MD simulations have the ability to distinguish incorrect binding modes from correct binding modes.

The failure cases of pose0 may be attributed to the inappropriate parameters of force field. Currently, the parameter of ligand is generated by RESP charge fitting derived from quantum calculation results, which is a common and widely used procedure. However, different fitting method may result in large difference during MD simulation [51]. For those special cases, detailed investigation on the force field parameter generation of ligand would be of importance to the improvement of our protocol.

In addition, the analysis of 2QBS with pose0 highlighted the importance of crystal water. This information of the important water molecules is rarely known in practice. However, this problem may be alleviated by placing some presumed water molecules into complex from theoretical prediction, such as the method based on reference interaction site model or grand canonical Monte Carlo simulation.

In the current analysis, the stability of incorrect poses by the statistical analyses based on the 5 parallel runs may be overestimated. This is because the incorrect pose is judged as stable, when the RMSD change of any one among the 5 parallel MD runs is less than 2.0 Å. Thus, we can maintain the correct binding pose (pose0) as much as possible. Further discrimination between the correct poses and those survived decoys may be handled by the absolute binding free energy calculation based on perturbation theory.

Our investigations suggest that if we use 2.0 Å as the threshold 38–44% incorrect binding poses can be excluded just by performing equilibrium MD simulations. If one uses 2.5 Å instead, the percentage of the wrongly excluded correct binding poses can be decreased to only 3%, though the percentage of properly excluding incorrect poses is also decreased to 19–24%. We have actually adopted such a criterion (2.0 Å) in our in-house protein–ligand binding free energy prediction tool for the compromise between accuracy and efficiency. Because the computational cost of the binding free energy calculations is expensive, it is very useful if we can some exclude putative wrong poses by the simple equilibrium simulations. The detailed protocol of our binding free energy prediction tool and its application to drug design will be discussed elsewhere.

Acknowledgements The authors are grateful to Dr. Satoshi Endo and Dr. Toshimasa Tanaka for careful reading of the manuscript and valuable suggestions and comments. The computational resources are supported by the K computer of the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (Project ID: hp120013) and the TSUBAME Grid Cluster at the

Global Scientific Information and Computing Center of Tokyo Institute of Technology, supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Open Advanced Research Facilities Initiative (Project ID: 15IBD).

References

- Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 9:203–214
- Bajorath J (2002) Integration of virtual and high-throughput screening. *Nat Rev Drug Discov* 1:882–894
- Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3:935–949
- Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN (2007) Virtual screening in drug discovery—a computational perspective. *Curr Protein Pept Sci* 8:329–351
- Leach AR, Shoichet BK, Peishoff CE (2006) Prediction of protein–ligand interactions. Docking and scoring: successes and gaps. *J Med Chem* 49:5851–5855
- Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912–5931
- Kroemer RT (2007) Structure-based drug design: docking and scoring. *Curr Protein Pept Sci* 8:312–328
- Dunbar JB, Smith RD, Damm-Ganamet KL, Ahmed A, Esposito EX, Delproposito J, Chinnaswamy K, Kang Y-N, Kubish G, Gestwicki JE, Stuckey JA, Carlson HA (2013) CSAR data set release 2012: ligands, affinities, complexes, and docking decoys. *J Chem Inf Model* 53:1842–1852
- Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. *J Med Chem* 49:6789–6801
- Li Y, Liu Z, Li J, Han L, Liu J, Zhao Z, Wang R (2014) Comparative assessment of scoring functions on an updated benchmark: 1. compilation of the test set. *J Chem Inf Model* 54:1700–1716
- Lagarde N, Zagury J-F, Montes M (2015) Benchmarking data sets for the evaluation of virtual ligand screening methods: review and perspectives. *J Chem Inf Model* 55:1297–1307
- Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins* 47:409–443
- Cheng T, Li X, Li Y, Liu Z, Wang R (2009) Comparative assessment of scoring functions on a diverse test set. *J Chem Inf Model* 49:1079–1093
- Li Y, Han L, Liu Z, Wang R (2014) Comparative assessment of scoring functions on an updated benchmark: 2. evaluation methods and general results. *J Chem Inf Model* 54:1717–1736
- Hamza A, Wei N-N, Zhan C-G (2012) Ligand-based virtual screening approach using a new scoring function. *J Chem Inf Model* 52:963–974
- Nabuurs SB, Wagener M, de Vlieg J (2007) A flexible approach to induced fit docking. *J Med Chem* 50:6507–6518
- Sun H, Li Y, Shen M, Tian S, Xu L, Pan P, Guan Y, Hou T (2014) Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/GBSA and MM/PBSA rescoring. *Phys Chem Chem Phys* 16:22035–22045
- Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the MM/PBSA and MM/GBSA Methods. 1. The

- accuracy of binding free energy calculations based on molecular dynamics simulations. *J Chem Inf Model* 51:69–82
19. Greenidge PA, Kramer C, Mozziconacci J-C, Sherman W (2014) Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. *J Chem Inf Model* 54:2697–2717
 20. Thompson DC, Humblet C, Joseph-McCarthy D (2008) Investigation of MM-PBSA rescoring of docking poses. *J Chem Inf Model* 48:1081–1091
 21. Cao R, Huang N, Wang Y (2014) Evaluation and application of MD-PB/SA in structure-based hierarchical virtual screening. *J Chem Inf Model* 54:1987–1996
 22. Rastelli G, Del Rio A, Degliesposti G, Sgobba M (2010) Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J Comput Chem* 31:797–810
 23. Hou T, Wang J, Li Y, Wang W (2011) Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J Comput Chem* 32:866–877
 24. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10:449–461
 25. Mobley DL, Dill KA (2009) Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. *Structure* 17:489–498
 26. Clark AJ, Tiwary P, Borrelli K, Feng S, Miller EB, Abel R, Friesner RA, Berne BJ (2016) Prediction of protein–ligand binding poses via a combination of induced fit docking and metadynamics simulations. *J Chem Theory Comput* 12:2990–2998
 27. Kellenberger E, Muller P, Schalon C, Bret G, Foata N, Rognan D (2006) sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank. *J Chem Inf Model* 46:717–727
 28. Desaphy J, Bret G, Rognan D, Kellenberger E (2015) sc-PDB: a 3D-database of ligandable binding sites—10 years on. *Nucleic Acids Res* 43:D399–D404
 29. Hu L, Benson ML, Smith RD, Lerner MG, Carlson HA (2005) Binding MOAD (mother of all databases). *Proteins Struct Funct Bioinform* 60:333–340
 30. Ahmed A, Smith RD, Clark JJ, Dunbar JB, Carlson HA (2015) Recent improvements to Binding MOAD: a resource for protein–ligand binding affinities and structures. *Nucleic Acids Res* 43:D465–D469
 31. Kerns EH, Di L (2008) Drug-like properties concepts, structure design and methods: from ADME to toxicity optimization. Academic Press, Boston
 32. Michalsky E, Dunkel M, Goede A, Preissner R (2005) SuperLigands—a database of ligand structures derived from the Protein Data Bank. *BMC Bioinform* 6:122
 33. Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* 6:881–890
 34. Leeson PD, Young RJ (2015) Molecular property design: does everyone get it? *ACS Med Chem Lett* 6:722–725
 35. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL (2012) Quantifying the chemical beauty of drugs. *Nat Chem* 4:90–98
 36. Warren GL, Do TD, Kelley BP, Nicholls A, Warren SD (2012) Essential considerations for using protein–ligand structures in drug discovery. *Drug Discov Today* 17:1270–1281
 37. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Dyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
 38. Chemical Computing Group Inc. (2014) Molecular Operating Environment (MOE), 2014.09. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7
 39. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749
 40. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09 Revision, vol C 01. Gaussian Inc, Wallingford CT
 41. Walker RC, Crowley MF, Case DA (2008) The implementation of a fast and accurate QM/MM potential method in Amber. *J Comput Chem* 29:1019–1031
 42. Case DA, Berryman JT, Betz RM, Cerutti DS, Cheatham, III TE, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Homeyer N, Izadi S, Janowski P, Kaus J, Kovalenko A, Lee TS, LeGrand S, Li P, Luchko T, Luo R, Madej B, Merz KM, Monard G, Needham P, Nguyen H, Nguyen HT, Omelyan I, Onufriev A, Roe DR, Roitberg A, Salomon-Ferrer R, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM (2015) AMBER 2015. University of California, San Francisco
 43. Bradbrook GM, Gleichmann T, Harrop SJ, Habash J, Raftery J, Kalb (Gilboa) J, Yariv J, Hillier IH, Helliwell JR (1998) X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding. *J Chem Soc Faraday Trans* 94:1603–1611
 44. Li P, Merz KM (2014) Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J Chem Theory Comput* 10:289–297
 45. Sindhikara DJ, Kim S, Voter AF, Roitberg AE (2009) Bad seeds sprout perilous dynamics: stochastic thermostat induced trajectory synchronization in biomolecules. *J Chem Theory Comput* 5:1624–1631
 46. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res* 34:D668–D672
 47. Chipot C, Pohorille A (2007) Free energy calculations: theory and applications in chemistry and biology. Springer, Berlin
 48. Kovalenko A, Hirata F (1999) Self-consistent description of a metal-water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model. *J Chem Phys* 110:10095
 49. Essex JW, Jorgensen WL (1995) An empirical boundary potential for water droplet simulations. *J Comput Chem* 16:951–997
 50. Woo H-J, Dinner AR, Roux B (2004) Grand canonical Monte Carlo simulations of water in protein environments. *J Chem Phys* 121:6392
 51. Ahmed A, Sandler SI (2013) Hydration free energies of multi-functional nitroaromatic compounds. *J Chem Theory Comput* 9:2774–2785