JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

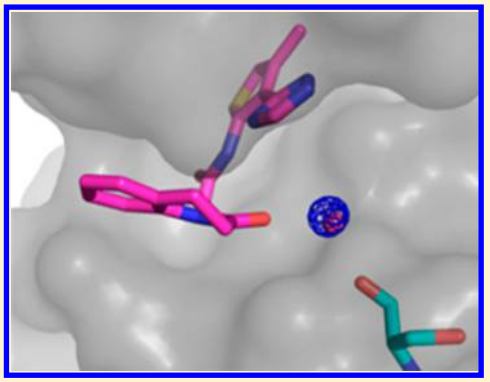
Including Explicit Water Molecules as Part of the Protein Structure in MM/PBSA Calculations

Yong-Liang Zhu,* Paul Beroza,† and Dean R. Artis

Department of Molecular Design, Elan Pharmaceuticals, 180 Oyster Point Boulevard, South San Francisco, California 94080, United States

Supporting Information

ABSTRACT: Water is the natural medium of molecules in the cell and plays an important role in protein structure, function and interaction with small molecule ligands. However, the widely used molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) method for binding energy calculation does not explicitly take account of water molecules that mediate key protein–ligand interactions. We have developed a protocol to include water molecules that mediate ligand–protein interactions as part of the protein structure in calculation of MM/PBSA binding energies (a method we refer to as water–MM/PBSA) for a series of JNK3 kinase inhibitors. Improved correlation between water–MM/PBSA binding energies and experimental IC_{50} values was obtained compared to that obtained from classical MM/PBSA binding energy. This improved correlation was further validated using sets of neuraminidase and avidin inhibitors. The observed improvement, however, appears to be limited to systems in which there are water-mediated ligand–protein hydrogen bond interactions. We conclude that the water–MM/PBSA method performs better than classical MM/PBSA in predicting binding affinities when water molecules play a direct role in mediating ligand–protein hydrogen bond interactions.



INTRODUCTION

The prediction of binding affinities is one of the most important tasks in the study of protein–ligand interactions in structure-based drug design. Free energy methods, such as molecular mechanics Poisson–Boltzmann surface area (MM/PBSA), have received much attention in recent literature references. These methods benefit from computational efficiency as only the initial (unbound) and final (bound) states of the system are evaluated. Once ligands are parametrized, the MM/PBSA model can provide estimates of protein–ligand binding energies that are in good agreement with experimentally measured binding affinities.[1−3]

The classical MM/PBSA method uses a continuum solvation model that represents solvent as a continuous medium instead of an ensemble of individual "explicit" solvent molecules. However, this method may be complicated by the presence of water molecules that directly mediate key interactions between ligand and protein. Water can act both as a hydrogen bond donor and acceptor at the interface of biomolecular complexes. Practical implementations of binding energy calculation strategies currently largely ignore the effects of explicit water, because the structural and thermodynamic effects of water molecules at binding interfaces are hard to determine and hard to model. Nevertheless, there are several studies in which discrete water molecules are included in binding energy calculations, as water plays an important role in mediating the interactions between molecules in these systems. For example, the experimental binding affinity of nevirapine to wild-type HIV-1 reverse transcriptase is much higher than to the Y181C mutant. Treesuman and Hannongbua[4] found that the MM/PBSA binding energy calculation discriminated the binding affinity difference between wild-type and mutant much better when key water molecules were included as part of protein. It has also been reported that incorporation of water molecules into a three-dimensional quantitative structure–activity relationship analysis improved the predictive ability of the models.[5] Nonetheless, a generalized method to include explicit water in MM/PBSA calculations and an evaluation of its success in reproducing experimental binding affinities across multiple protein targets have not been reported.

We have applied molecular dynamic simulations and MM/PBSA binding energy calculations to predict compound binding affinity for lead optimization in a number of projects at Elan. The correlation between calculated binding energies and experimental bioactivities, however, has not always been satisfactory, particularly when water molecules mediate ligand–protein hydrogen bond interactions, which are often important considerations in guiding medicinal chemistry. For example, our crystallography studies found the presence of a water molecule that appeared to provide a bridging interaction between certain inhibitors and their target protein JNK3. Our MM/PBSA calculations failed to predict the relative binding energies of these inhibitors, and we reasoned that including this

Received: March 26, 2013
Published: December 28, 2013

dx.doi.org/10.1021/ci4001794 | J. Chem. Inf. Model. 2014, 54, 462–469

water molecule explicitly in the MM/PBSA calculations might improve the accuracy of the results. Furthermore, we wished to automate the process of selecting and incorporating water molecules that were structurally tightly coupled to both ligand and protein into the process of binding energy computation. Therefore, we developed a method that takes into account water molecules that directly mediate protein−ligand interactions in calculating the binding energy from MM/PPBSA (water−MM/PBSA). Here, we evaluate the impact of the water−MM/PBSA method on the correlation between calculated binding free energies and experimental IC_{50}s in a series of thirty-three inhibitors of JNK3 kinase. The method is further evaluated in the context of similar studies for the neuraminidase, avidin and p38 systems.

METHODS

Molecular Dynamics Simulations. The Amber99 force field and Amber10 were used throughout this work.[6] The partial charges for the atoms of small molecule ligands were produced using the Antechamber tool with option AM1-BCC.

For any given protein target, a high-resolution crystal structure of the protein was used as a template structure to simulate all the ligands in the series. To prepare the template structure, the protonation states of the histidine residues in the protein were manually inspected and tautomer states were chosen that were most consistent with the local chemical environment. All crystallographically resolved water molecules were retained in the systems while ions were removed. A sphere of TIP3P water molecules was added with a diameter of 29 Å for each ligand−protein complex. Then a computer program was run to add additional water molecules to fill the gaps between ligand and protein. Briefly, a cubic grid with 0.25 Å spacing was superimposed on the binding site. Grid points that were beyond 3 Å from any heavy atom in the system were identified. These points were then clustered using a complete-linkage hierarchical clustering method provided in WatCH.[7] Oxygen atoms for the TIP3P water molecules were placed in the resulting cluster centroids. There was no consideration of energetics to place water molecules between the cavity and ligand atoms. However, the water placement was manually inspected to make sure water molecules were not forced into small voids in the protein where they would be energetically unfavorable.[8] Each complex eventually contained approximately 10 000 atoms and 1400 water molecules.

The resulting solvated ligand−protein complex was then energy minimized and equilibrated at 300 K for 100 ps. This was followed by 250 ps of production data collection of MD simulations that were carried out with belly type dynamics, i.e., only a subset of protein atoms at the ligand binding site was allowed to move, and the coordinates of the rest of protein were frozen. The ligand and all water molecules in the system were unconstrained. A primary cutoff of 12 Å was applied for nonbonded interactions, and a small harmonic potential was applied to constrain protein α-carbon atoms in place. A semiharmonic potential was applied to the solvated sphere to keep the solvation layer intact. Key parameters used in each step of the molecular dynamics protocol are listed in Table S1 (Supporting Information).

The temperature of the production run was maintained at 300 K, and a total of 250 snapshots were saved during the data collection period, one snapshot per 1 ps of MD simulation. A visual inspection of the trajectory was conducted to make sure that the protein−ligand complex was stable during the entire simulation period.

Classical MM/PBSA Calculations. We calculated the energy difference between the complex and the unbound ligand and protein from the following formula:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{protein}} - G_{\text{ligand}} \quad (1)$$

The free energy calculation was carried out by using the MM/PBSA tool provided in Amber10, which combines the molecular mechanical energies with the continuum solvent reaction field energy.[9] The molecular mechanical energies include the internal energy (bond, angle and dihedral), van der Waals and electrostatic interactions, although the internal energies in eq 1 cancel, leaving only the van der Waals and electrostatic terms. The dielectric constant was set to 1.0, and the interaction cutoff distance was 999 Å.

The hydrophobic contribution to the solvation energy is given by the product of the surface area and an effective surface tension term, which is proportional to the size of the ligand and largely cancels for similar size analogs in this study. The polar electrostatic contribution was calculated by integrating the Poisson−Boltzmann equation for the system.[10] The solvent dielectric was set to 80.0, the protein dielectric was set to 1.0 and the rest of the parameters were default values.

MM/PBSA energies were calculated for each of the 250 frames, and then averaged to obtain the overall calculated binding energy.

Selection of Interface Water Molecules for Water−MM/PBSA Calculation. For the water−MM/PBSA method, we included the effect of interfacial water molecules between protein and its binding inhibitors. Specifically, for each MD snapshot, we identified all water molecules whose oxygen atoms were within 3.5 Å of a protein heavy atom and a ligand heavy atom, i.e., were close to both protein and ligand. Usually one to seven water molecules for each snapshot met that criterion, and these water molecules were considered as a part of the protein structure for the purposes of the MM/PBSA calculations. So, the binding free energy is approximated by the difference of the following formula:

$$\Delta G_{\text{bind}} = G_{\text{complex+interfacial waters}} - G_{\text{protein+interfacial waters}} - G_{\text{ligand}} \quad (2)$$

Once the interfacial water molecules were added to the protein structure, the energy calculations followed the classical MM/PBSA approach as described above. The prediction index (PI) was calculated based on Pearlman.[11]

RESULTS

JNK3 Inhibitors. All 33 compounds from Bowers, et al.[12] were used for MD simulations and binding energy calculations (see Table S2A, Supporting Information, for compound structures and experimental IC_{50}s). Compound numbers in this paper follow those in the original paper. The X-ray cocrystal structure of compound 34 was solved at 2.4 Å (PDB code 3PTG) and was used as template for molecular dynamics simulations of all compounds.

Root-mean-square devation of atom coordinates stabilized quickly, and total energy and calculated ligand−protein interaction converged at about 150 ps from the start of simulations (data not shown). We ran replicate simulations, each with slightly perturbed initial coordinates,[13] for each

dx.doi.org/10.1021/ci4001794 | J. Chem. Inf. Model. 2014, 54, 462−469

JNK3−ligand complex. All 33 ligands were included in the last three runs (Run 3, 4 and 5 in Table S2B, Supporting Information) whereas some ligands were not in the first two runs (Run 1 and 2) because, originally, the scope of the investigation was limited to selected ligands and was expanded to all ligands later.

The correlation between calculated binding energies (average of three or five replicate simulations) and experimentally measured affinities is shown in Figure 1 (see Table S2B, Supporting Information, for calculated binding energy of each run of each compound). For the classical MM/PBSA method (Figure 1, left panel), the calculated energies differentiated potent double digit nanomolar compounds from less potent double digit micromolar ones. However, the overall correlation between calculated binding energy and IC_{50} is poor, with an r^2 of 0.55 (Figure 1, left panel).

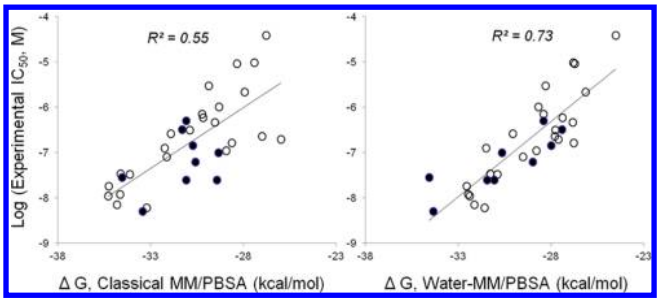


Figure 1. Correlations of MM/PBSA binding energies and IC_{50}s of JNK3 inhibitors. Each calculated energy is the average of calculations based on three or five replicate dynamic simulations. (Left: classical MM/PBSA; right: water−MM/PBSA. Black circles: charged amine or quinolinone containing compounds.)

Using the same simulation trajectories as above, we calculated ligand−protein binding energies by the water−MM/PBSA method (Figure 1, right panel). The water−MM/PBSA calculation showed improved correlation with experimental IC_{50}, with an r^2 of 0.73 (Figure 1, right panel). Moreover, the PI improved from 0.75 to 0.88.

Inspection of the costructure of compound 34 shows a water molecule (W47) that mediates a hydrogen bond interaction between the carbonyl of the quinolinone moiety and the backbone carbonyl Ser193 (Figure 2A). This water is conserved in the MD simulation (Figure 2B) and therefore was included as part of the protein structure in most of the MD frames used in the water−MM/PBSA calculation. Note that in the classical MM/PBSA approach the unfavorable interaction between these carbonyls would be shielded only by the polarization of the bulk solvent.

Other analogs in the series show a similar pattern of water-mediated interactions between ligand and protein. Indeed, the improvement of predictive power of water−MM/PBSA seems mainly the result of better estimates of binding energy for such compounds. These compounds (black circles in Figure 1) either have a charged amine (three compounds) or quinolinone group (six compounds). For all quinolinone containing compounds, water−MM/PBSA calculations included a water molecule that mediated a hydrogen bond interaction between the carbonyl of the quinolinone moiety and the carbonyl of S193 of protein; and the water molecule moved in a well-confined and narrow space during the whole simulation time, as shown in Figure 2B. For all three amine containing compounds, the water−MM/PBSA method included at least one water molecule that bridged hydrogen bond interactions between ligand and protein as shown for compound 18 in Figure 2C. The binding energies of these amines and quinolinone containing compounds are underestimated compared to the rest of the compounds, and they are mostly outliers in classical MM/PBSA−IC_{50} correlation (Figure 1, left panel). These compounds are much closer to the regression line in the water−MM/PBSA−IC_{50} correlation (Figure 1, right panel).

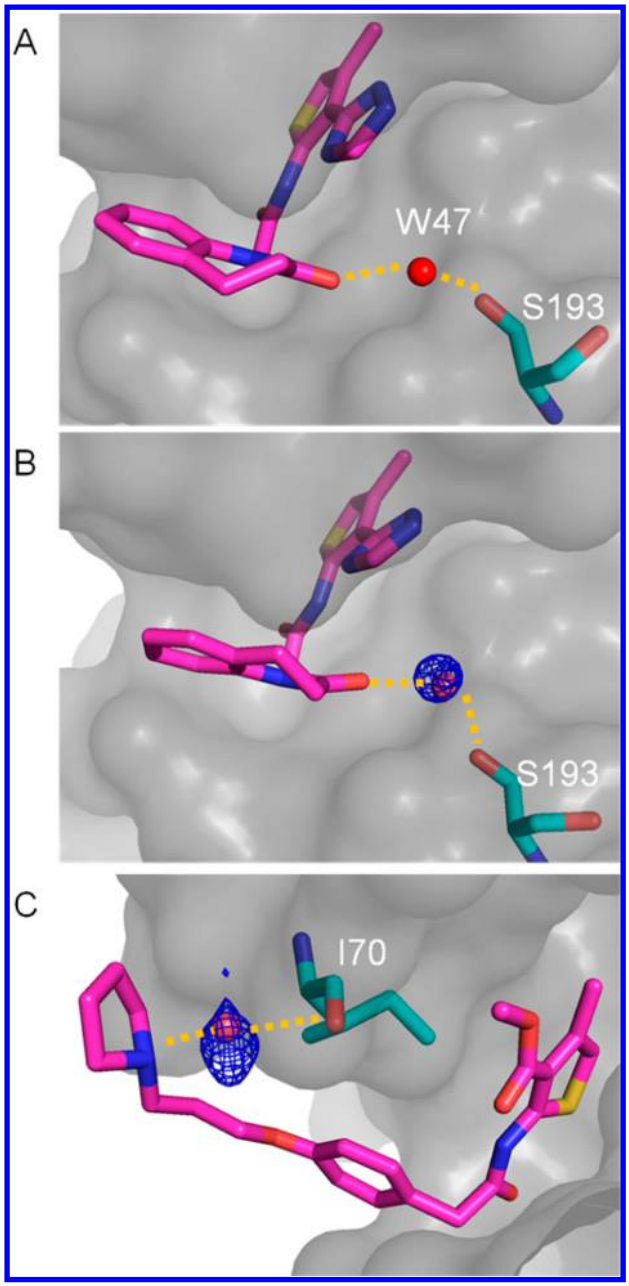


Figure 2. A. Cocrystal structure of compound 34 with JNK3 (PDB 3PTG) highlighting a water molecule, W47, that mediates the ligand−protein interaction. B. A snapshot from the compound 34 simulation highlighting the water equivalent to W47 in the experimental structure. C. A snapshot from the compound 18 simulation highlighting a water molecule that mediates a ligand−protein interaction. Ligands are in violet stick, key residues are in cyan stick, waters shown as red spheres and protein in gray surface. Orange dashed lines indicate hydrogen bonding interactions. The blue mesh indicates the water density isosurface for the simulation, i.e., regions where a water molecule is present at this location for most of the MD frames.

dx.doi.org/10.1021/ci4001794 | J. Chem. Inf. Model. 2014, 54, 462−469

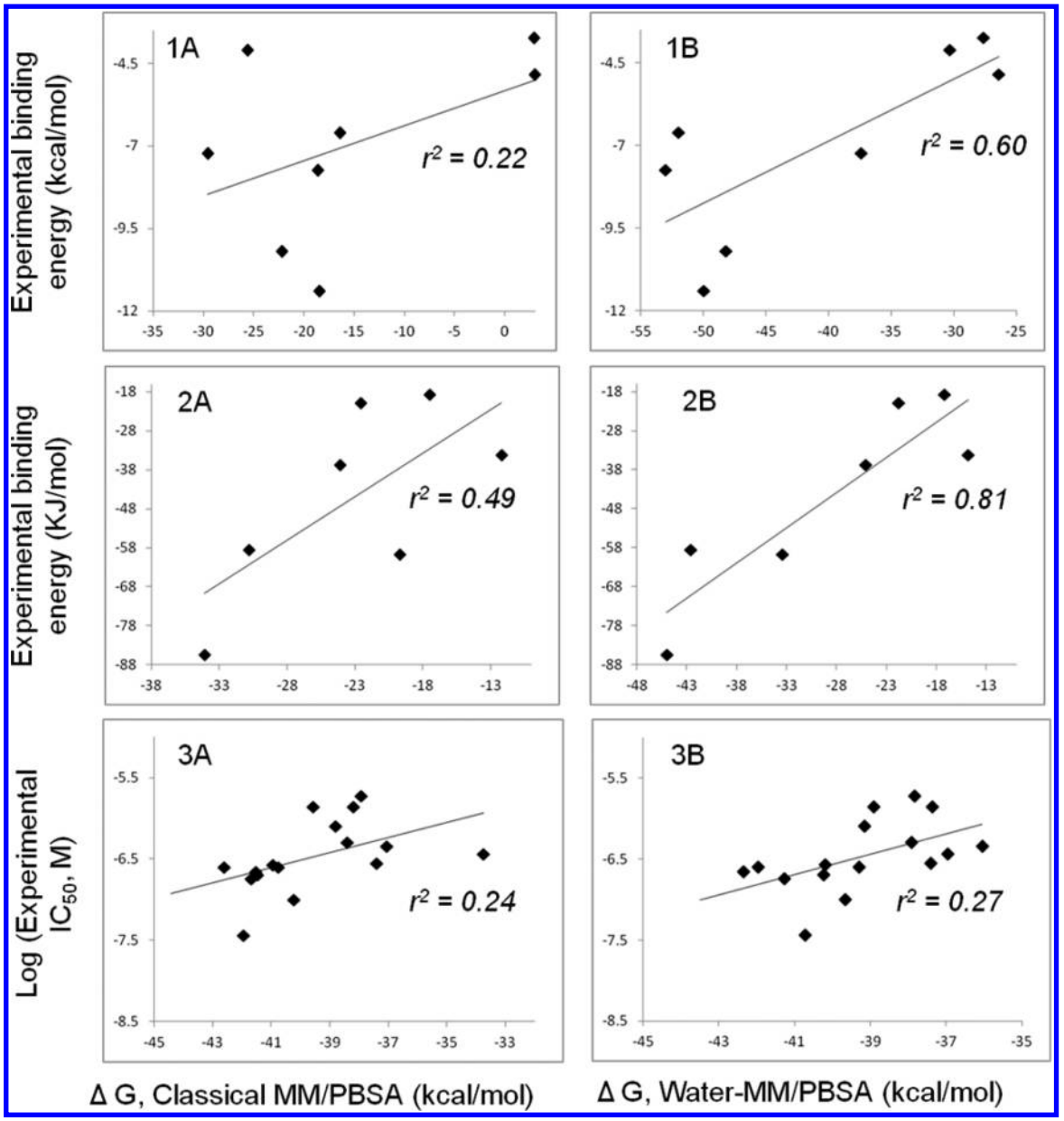


Figure 3. Correlation of calculated binding energy with IC_{50} of inhibitors for neuraminidase (1A and 1B), avidin (2A and 2B) and p38 (3A and 3B). For each enzyme system, left panel is classical MM/PBSA energy and right is water−MM/PBSA.

Generalization to Other Protein Systems. To further assess the efficacy of this approach to improve binding energy estimates of protein inhibitors, we examined published data sets for three proteins: neuramidase,[14] avidin,[15] and p38[11] (see Tables S3, S4 and S5, Supporting Information, for the structures, experimental binding activities and calculated MM/PBSA scores of these compounds). All MM/PBSA calculations for these three proteins are based on a single simulations for any protein−ligand complex to save CPU time. The comparison of the correlations between computation and experiment is shown in Figure 3.

Neuraminidase Inhibitors. Hou, et al.[14] reported that the binding energies calculated by MM/PBSA for eight inhibitors correlated poorly with their experimental activity (r^2 = 0.46). The calculated binding energy by classical MM/PBSA in our hands correlates with K_i even more poorly (r^2 = 0.22, Figure 3, panel 1A). This discrepancy is probably caused by different MD methodologies. Hou et al. used each individual costructure to model that inhibitor only. In our case, the first two ligands, d1 and d2, were modeled in their target influenza neuraminidase (PDB code 1NSD[16]) and the remaining ligands, d3−d8, were modeled in their target, an R292K variant of influenza neuraminidase (PDB code 2QWG).[17] The binding energy of the water−MM/PBSA calculation correlates to experimental activity significantly better (r^2 = 0.60, Figure 3, panel 1B) than that of classical MM/PBSA. The PI was improved from 0.27 for classical MM/PBSA to 0.64. The water−MM/PBSA method usually included 4−7 water molecules that mediate hydrogen bond interactions between ligand and protein in both neuraminidase or R292K variant (Figure 4). Most of these water molecules are also present in the cocrystal structures of neuraminidase or the R292K variant.

Avidin−Biotin Complexes. For the seven biotin analogs in complex with avidin cited by Weis, et al.,[15] we chose the PDB 2C1S structure,[18] as this costructure has a relatively high resolution. Water−MM/PBSA binding energy correlates

dx.doi.org/10.1021/ci4001794 | J. Chem. Inf. Model. 2014, 54, 462−469

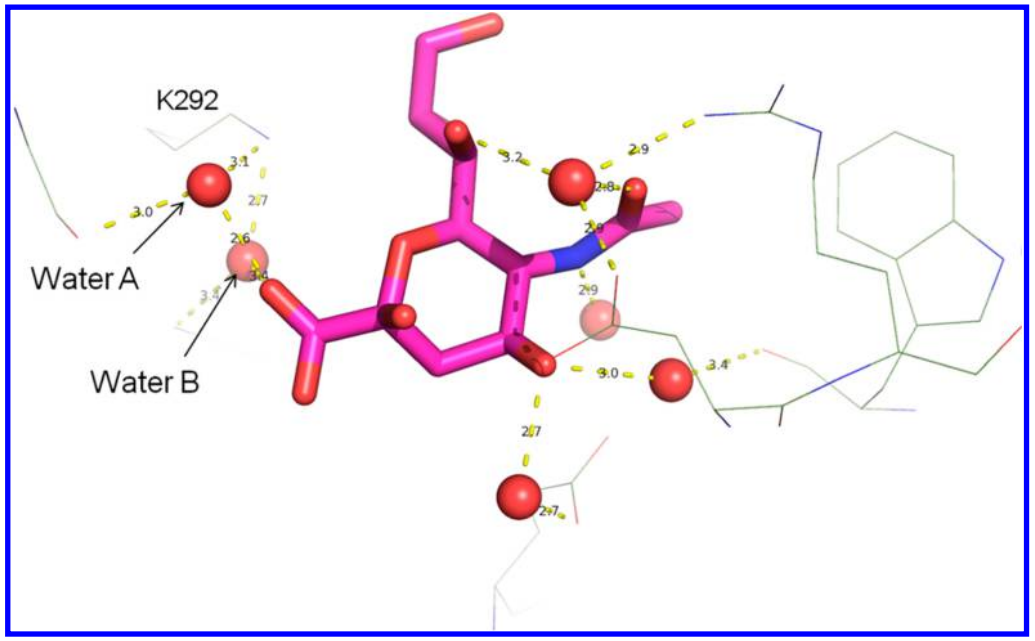


Figure 4. A snapshot of MD simulations of a ligand in neuraminidase R292K variant highlighting multiple water molecules that mediate ligand−protein interactions. Water molecules in red spheres, ligand in violet stick and protein in line. Yellow dashed lines indicate hydrogen bond interactions.

dramatically better (r^2 = 0.81, PI = 0.90) to K_i (calculated from the experimental binding energy) than classical MM/PBSA (r^2 = 0.49, PI = 0.70.) (Figure 3, panels 2A and 2B). Inspection of water−MM/PBSA calculations indicates that a water molecule is included as a part of protein for the three most potent analogs, which all have a carboxylate group. This water molecule mediates an interaction between the carboxylate group and R115 of avidin (Figure 5). For the same system and

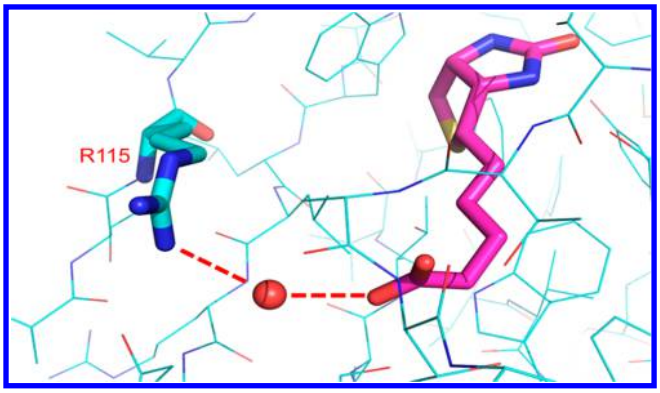


Figure 5. A snapshot of a biotin−avidin MD simulation. Biotin is shown in violet stick, protein in thin blue line except R115. Dashed red lines indicate hydrogen bonds between biotin/R115 and the water molecule (red ball).

ligands, Weis et al.[15] reported an r^2 of about 0.65, and Genheden and Ryde[19] reported an an r^2 of 0.79. Genheden and Ryde ran 20 separate MD simulations for each protein−ligand complex, which may explain why they obtained a better r^2. Furthermore, they used MM/GBSA and included the entropy term in the binding energy calculation. Kuhn and Kollman[20] reported an even better r^2 of 0.92 for this avidin system with the same seven ligands plus an additional one. The high r^2 from Kuhn and Kollman's work could be partially due to the inclusion of the entropy term as the r^2 would be 0.88 without consideration of the entropy term. The entropy term was ignored in this work due to its heavy CPU consumption.[21] Combination of water−MM/PBSA with an entropy calculation may yield still better results, and this will be the topic of further study.

p38 Inhibitors. Pearlman reported an MM/PBSA method for calculating free energies for a series of sixteen ligands of p38 MAP kinase whose binding constants span approximately 2 orders of magnitude and found that the MM/PBSA performed relatively poorly.[11] We simulated these 16 ligands in a p38 structure (PDB code 3HP5[22]). This structure was chosen as a template because it has a better resolution (2.3 Å) and a more similar compound bound in the protein compared to a PDB 3FC1, which was the structure deposited by Pearlman et al.[11] Residues that have direct contact with the ligand are well-defined in the 3PH5 costructure, although three proximal residues (33−35) in the P-loop are disordered. The classical MM/PBSA calculation based on our protocol correlates poorly with experiment IC_{50} (r^2 = 0.24, Figure 3, panel 3A), which is in agreement with Pearlman's result. However, unlike the other systems we examined, the water−MM/PBSA method does not appreciably improve the correlation (r^2 = 0.27, Figure 3, panel 3B). The PI was also unimproved (0.62 for classical MM/PBSA and 0.61 for water−MMPBSA. Even though some water molecules were included in water−MM/PBSA calculation, there were no water molecules that directly mediate ligand−protein hydrogen bond interactions. Lucarelli et al.[23] reported an improvement of the PI from 0.41 to 0.62 upon using a JAWS-optimized water distribution for the MC/FEP simulations. However, the authors did not give the value of r^2 and the improved PI is almost identical to what we reported here, 0.62 for classical MM/PBSA and 0.61 for water−MM/PBSA. Ligands 14 and 15 are the only ones to have a variable polar group, OH or NH, respectively, in the R1 position (see Table S4, Supporting Information). A water molecule mediating hydrogen bond interaction between the OH or NH_2 group was observed in the JAWS-optimized water distribution for the

dx.doi.org/10.1021/ci4001794 | J. Chem. Inf. Model. 2014, 54, 462−469

MC/FEP simulations. However, we did not observe any water molecule that interfaces between OH or NH_2 and the protein in our simulations. This can be due to the difference of protein structures: 1OUY used in Lucarelli study and 3PH5 in our study. 3PH5 has a resolution of 2.3 Å and its ligand is more similar to the 16 ligands studied in this work whereas 1OUY has a resolution of 2.5 and its ligand has an extra substitution group from the 3-position of the pyrimidopyridazinone core. Residue Asp168 is much closer to OH or NH_2 in 3PH5 than in 1OUY and does not have a room for a water molecule to stay between OH or NH_2 and Asp168. This highlights the importance of choosing the starting protein structure in any quantitative MD approach. Another plausible explanation is that the protocol used here (short simulations, harmonic restraints on Ca atoms, initial water placement protocol) prevent reorganization of Asp168. Thus, the problem could be due to limited sampling.

DISCUSSION

Although the inclusion of water molecules in MM/PBSA calculations of binding energies has been reported for a small number of ligand complexes in a few studies,[4,5] we have presented here a more complete, systematic methodology to include explicit water molecules as part of the protein in MM/PBSA calculation. For most systems, we found an improved correlation of calculated binding energies with experimentally measured IC_{50} values.

The improvement appears to result from the increased affinity to the protein for polar moieties relative to apolar ones. In the JNK3 system, for instance, the difference between water−MM/PBSA and classical MM/PBSA is about −5 to +5 kcal/mol. Seven of all nine amine or quinolinone containing compounds have equal or stronger binding energies calculated by water−MM/PBSA method (more negative values, up to −2.6 kcal/mol) than by the classical MM/PBSA method. Compounds that do not contain an amine or quinolinone group have an opposite trend: weaker binding energy (0.5 to 3.5 kcal) by water−MM/PBSA than by classical MM/PBSA. These two opposite trends put amine and quinolinone containing compounds closer to the regression line in the water−MM/PBSA binding energy IC_{50} correlation (Figure 1, right panel).

In both neuraminidase and avidin systems, water−MM/PBSA calculated binding energies differ more significantly from classical MM/PBSA calculated energies for ligands that interact with the protein through water-mediated hydrogen bonds. For example, the difference between water−MM/PBSA and classical MM/PBSA is −4.7 and −7.8 kcal/mol for d1 and d2 neuraminidase complexes, respectively. This difference is as large as −26 to −36 kcal/mol for the six neuraminidase R292K variant complexes (d3−d8) (Table 1). The bigger differences between water−MM/PBSA energies and classical MM/PBSA energies for R292K variant complexes (d3−d8) than for wild-type complexes (d1 and d2) may be explained by the observation that more water molecules were included in water−MM/PBSA calculation than in classical MM/PBSA calculation. In the case of water−MM/PBSA, two key water molecules, A and B in Figure 4, in the R292K variant system were replaced by R292 in the d1 or d2 neuraminidase systems. These results indicate that the more water molecules that are involved in ligand−protein hydrogen bond interactions, the more important it is to include them explicitly as part of the protein structure in the MM/PBSA calculation.

Table 1. Calculated binding energy of ligands with neuraminidase or its R292K variant

ligand	target	template structure	classical MM/PBSA	water−MM/PBSA	diff[a]	binding energy[b]
d1	wild type[c]	1NSD	−25.6	−30.3	−4.7	−4.09
d2	wild type	1NSD	−29.6	−37.4	−7.8	−7.23
d3	Mutant[d]	2QWG	2.89	−27.6	−30.49	−3.74
d4	Mutant	2QWG	2.96	−26.4	−29.36	−4.84
d5	Mutant	2QWG	−16.39	−52	−35.61	−6.61
d6	Mutant	2QWG	−22.2	−48.2	−26	−10.2
d7	Mutant	2QWG	−18.6	−53	−34.4	−7.73
d8	Mutant	2QWG	−18.5	−50	−31.5	−11.4

[a]Difference, i.e., Water−MM/PBSA − classical MM/PBSA. [b]Experimental binding energy in kcal/mol from Hou, et al.[14] [c]Neuraminidase. [d]Neuraminidase R292K variant.

The biotin−avidin system had the most significant improvement in correlation between water−MM/PBSA energies and experimental binding energies (Figure 2, panels 2A and 2B). Here, we observed that the water−MM/PBSA method lowers the calculated binding energies for three potent, carboxylate containing compounds (a, b and c) by more than −10 kcal/mol whereas there is only minimal difference between water−MM/PBSA and classical MM/PBSA for less potent, noncarboxylate analogs (d, e, f and g, Figure 6). The carboxylate group of

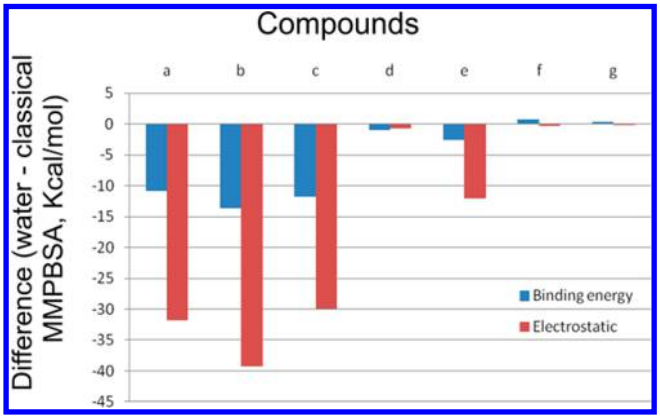


Figure 6. Difference of overall binding energy and the electrostatic term of water−MM/PBSA from classical MM/PBSA of biotin−avidin complexes.

compounds a, b and c engages a water-mediated hydrogen bond interaction with R115 of the protein (Figure 5). The difference of the electrostatic interaction term between water−MM/PBSA and classical MM/PBSA is the most significant among all energy terms and reflects the difference of overall binding energy very well (Figure 6), indicating that inclusion of key water molecules in MM/PBSA calculation is important for calculation of the electrostatic term. In other words, classical MM/PBSA's continuum model of water can seriously underestimate the electrostatic term, and the water−MMPBSA calculation can improve the estimate of the electrostatic interaction term when the system has water-mediated polar interactions.

Thus, the improved agreement with experiment we observe in these three systems can be rationalized structurally by the identification of waters that are tightly coupled to polar groups on both the protein and ligand. This improvement of the

dx.doi.org/10.1021/ci4001794 | J. Chem. Inf. Model. 2014, 54, 462−469

correlation between calculated binding energies with measured bioactivities, however, seems to be limited to complexes that involve water-mediated ligand−protein hydrogen bond interactions. None of the 16 ligands of p38 inhibitors have a polar group that interacts with protein through water-mediated hydrogen bonds, and there is no improvement in the calculated binding energy−activity correlation by water−MM/PBSA. It is important to note, however, that inclusion of explicit waters does not make the correlation worse.

It is important to identify which water molecules will be treated explicitly in the water−MM/PBSA calculation. In a T-cell receptor and staphylococcal enterotoxin 3 binding study, Wong et al. found that the calculated binding energies were able to discriminate between weak (wild type of a T-cell receptor) and strong binders (two mutant of the T-cell receptor) only when a few key explicit water molecules were included in the MM/PBSA analysis. When more (about 200) closest water molecules were included in calculation, the results gave an incorrect energetic trend as compared to experimental binding affinities.[4] We have conducted preliminary studies that agree with that report. If we increase the cutoff distance for the inclusion of explicit waters to 4.5 Å from both ligand and protein, poorer correlation with measured IC_{50} values results (data not shown). This observation also indirectly supports the conclusion that water−MM/PBSA calculation improves prediction only when water molecules mediate hydrogen bond ligand−portein interactions for some ligands in the series. It seems only closely coupled water molecules warrant the special treatment and that inclusion of more distant water molecules may well introduce additional errors. A more thorough investigation to confirm this preliminary observation and to optimize the cutoff parameter will be the subject of future work.

CONCLUSION

The main finding of this work is that an automated procedure facilitating modification of the MM/PBSA methodology to treat water molecules that are in close proximity to both ligand and protein as part of the protein structure can improve the correlation between calculated and experimental binding affinities. The method improves discrimination between strong- and weak-binding ligands when there are polar interactions that are mediated by water molecules. When such polar interactions are absent, the results do not improve, although this phenomenon was only observed in one system. It appears that a hybrid approach to the treatment of solvent in which waters that are tightly coupled to the protein/ligand system are treated as individual molecules and less tightly coupled molecules are treated as continuum dielectric gives a more accurate model of the protein/ligand binding energetics, at least in the context of the MM/PBSA methodology.

ASSOCIATED CONTENT

Supporting Information

Step details of molecular dynamic simulations, ligand structures and calculated binding energies for each proten system (Tables S1−S5). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Y.-L. Zhu. Address: NeuPharma Inc, 1175 Chess Drive, Suite 206, Foster City, CA 94404, USA. Tel: 510-299-0769. E-mail: yzhu@neupharma.com.

Present Address

[†]Quantisci, 1011 Muir Way, Belmont, CA 94002

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

JNK3, c-Jun N-terminal kinases 3; MD, molecular dynamics; MM/PBSA, molecular mechanics Poisson−Boltzmann surface area; p38, p38 mitogen-activated protein kinase; PDB, Protein Data Bank

REFERENCES

(1) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889−897.

(2) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48* (12), 4040−4048.

(3) Ferrari, A. M.; Degliesposti, G.; Sgobba, M.; Rastelli, G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. *Bioorg. Med. Chem.* **2007**, *15* (24), 7865−7877.

(4) Treesuwan, W.; Hannongbua, S. Bridge water mediates nevirapine binding to wild type and Y181C HIV-1 reverse transcriptase–evidence from molecular dynamics simulations and MM/PBSA calculations. *J. Mol. Graphics Modell.* **2009**, *27* (8), 921−929.

(5) Wong, S.; Amaro, R. E.; McCammon, J. A. MM/PBSA Captures Key Role of Intercalating Water Molecules at a Protein-Protein Interface. *J. Chem. Theory Comput.* **2009**, *5* (2), 422−429.

(6) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. In *AMBER 10*; University of California, San Francisco: San Francisco, 2008.

(7) Sanschagrin, P. C.; Kuhn, L. A. Cluster Analysis of Consensus Water Sites in Thrombin and Trypsin Shows Conservation between Serine Proteases and Contributions to Ligand Specificity. *Protein Sci.* **1998**, *7* (10), 2054−2064.

(8) Matthews, B. W.; Liu, L. A review about nothing: are apolar cavities in proteins really empty? *Protein Sci.* **2009**, *18* (3), 494−502.

(9) Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.* **2002**, *23* (13), 1244−1253.

(10) Wang, J.; Luo, R. Assessment of linear finite-difference Poisson-Boltzmann solvers. *J. Comput. Chem.* **2010**, *31* (8), 1689−1698.

(11) Pearlman, D. A. Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48* (24), 7796−7807.

(12) Bowers, S.; Truong, A. P.; Neitz, R. J.; Neitzel, M. L.; Probst, G. D.; Hom, R. K.; Konradi, A. W.; Sham, H. L.; Tóth, G.; Wu, J.; Pan, H.; Yao, N.; Artis, D. R.; Brigham, E. F.; Quinn, K.; Sauer, J. M.; Powell, K.; Ruslim, L.; Bard, F.; Yednock, T. A.; Griswold-Prenner, I. Design and synthesis of a novel, oraaly effectious tri-sunstituted

thiophene based JNK inhibitor. *Bioorg. Med. Chem. Lett.* **2011**, *21* (6), 1838−1843.
(13) Adler, M.; Beroza, P. Improved Ligand Binding Energies Derived from Molecular Dynamics: Replicate Sampling Enhances the Search of Conformational Space. *J. Chem. Inf. Model.* **2013**, *53* (8), 2065−2072.
(14) Hou, T.; Wang, J.; Li, Y.; Wang, W. 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.* **2011**, *51* (1), 69−82.
(15) Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **2006**, *49* (22), 6596−6606.
(16) Burmeister, W. P.; Henrissat, B.; Bosso, C.; Cusack, S.; Ruigrok, R. W. Influenza B virus neuraminidase can synthesize its own inhibitor. *Structure* **1993**, *1* (1), 19−26.
(17) Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **1998**, *6* (6), 735−746.
(18) Hytönen, V. P.; Määttä, J. A.; Niskanen, E. A.; Huuskonen, J.; Helttunen, K. J.; Halling, K. K.; Nordlund, H. R.; Rissanen, K.; Johnson, M. S.; Salminen, T. A.; Kulomaa, M. S.; Laitinen, O. H.; Airenne, T. T. Structure and characterization of a novel chicken biotin-binding protein A (BBP-A). *BMC Struct. Biol.* **2007**, *7* (8), 1−20.
(19) Genheden, S.; Ryde, U. Comparison of the Efficiency of the LIE and MM/GBSA Methods to Calculate Ligand-Binding Energies. *J. Chem. Theory. Comput.* **2011**, *7* (11), 3768−3778.
(20) Kuhn, B.; Kollman, P. A. Binding of a Diverse Set of Ligands to Avidin and Streptavidin: An Accurate Quantitative Prediction of Their Relative Affinities by a Combination of Molecular Mechanics and Continuum Solvent Models. *J. Med. Chem.* **2000**, *43* (20), 3786−3791.
(21) Hot, T.; Wang, J.; Li, Y.; Wang, W. 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.* **2011**, *51* (1), 69−82.
(22) Selness, S. R.; Devraj, R. V.; Monahan, J. B.; Boehm, T. L.; Walker, J. K.; Devadas, B.; Durley, R. C.; Kurumbail, R.; Shieh, H.; Xing, L.; Hepperle, M.; Rucker, P. V.; Jerome, K. D.; Benson, A. G.; Marrufo, L. D.; Madsen, H. M.; Hitchcock, J.; Owen, T. J.; Christie, L.; Promo, M. A; Hickory, B. S.; Alvira, E.; Naing, W.; Blevis-Bal, R. Discovery of N-substituted pyridinones as potent and selective inhibitors of p38 kinase. *Bioorg. Med. Chem. Lett.* **2009**, *19* (20), 5851−5856.
(23) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. J. Effects of Water Placement on Predictions of Binding Affinities for p38α MAP Kinase Inhibitors. *J. Chem. Theory Comput.* **2010**, *6* (12), 3850−3856.