JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

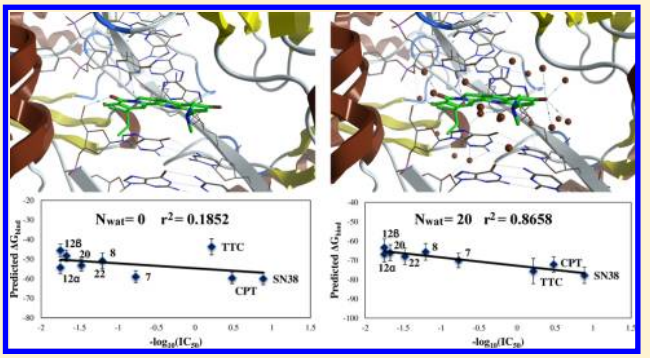
Explicit Ligand Hydration Shells Improve the Correlation between MM-PB/GBSA Binding Energies and Experimental Activities

Irene Maffucci and Alessandro Contini*

Dipartimento di Scienze Farmaceutiche—Sezione di Chimica Generale e Organica "Alessandro Marchesini", Università degli Studi di Milano, Via Venezian, 21 20133 Milano, Italy

Supporting Information

ABSTRACT: Molecular Mechanics Poisson−Boltzmann Surface Area (MM-PBSA) and Molecular Mechanics Generalized Born Surface Area (MM-GBSA) methods are widely used for drug design/discovery purposes. However, it is not clear if the correlation between predicted and experimental binding affinities can be improved by explicitly considering selected water molecules in the calculation of binding energies, since different and sometimes diverging opinions are found in the literature. In this work, we evaluated how variably populated hydration shells explicitly considered around the ligands may affect the correlation between MM-PB/GBSA computed binding energy and biological activities (IC_{50} and ΔG_{bind}, depending on the available experimental data). Four different systems—namely, the DNA-topoisomerase complex, α-thrombin, penicillopepsin, and avidin—were considered and ligand hydration shells populated by 10−70 water molecules were systematically evaluated. We found that the consideration of a hydration shell populated by a number of water residues (N_{wat}) between 30 and 70 provided, in all of the considered examples, a positive effect on correlation between MM-PB/GBSA calculated binding affinities and experimental activities, with a negligible increment of computational cost.



INTRODUCTION

Computational methods for binding energy predictions are gaining more and more importance in drug design.[1,2] Several methods are based on molecular dynamics (MD), which provide a statistically meaningful conformational ensemble for thermodynamic calculations at a reasonable computational cost. Among MD-based methods, the most popular are Molecular Mechanics Poisson−Boltzmann/Generalized Born Surface Area (MM-PB/GBSA),[3,4] linear interaction energy (LIE),[5] thermodynamic integration (TI),[1] and free energy perturbation (FEP).[6] However, the latter two are quite demanding, in terms of computational resources, and are more difficult to apply for drug design/discovery purposes, while MM-PB/GBSA methods, because of their computational efficiency, are getting used more and more in this field and their application within virtual screening protocols has also been reported.[7,8]

MM-PBSA and MM-GBSA methods stemmed from the intuition that a combination of molecular mechanics (MM) energies, polar and nonpolar solvation terms, and an entropy term can provide an approximation of the free energy of binding of a ligand to a receptor. The MM-PB/GBSA binding free energy is evaluated by factoring it as reported in eq 1:

$$\Delta G^0_{bind,solv} = \Delta G^0_{bind,vacuum} + \Delta G^0_{solv,complex} - (\Delta G^0_{solv,ligand} + \Delta G^0_{solv,receptor}) \quad (1)$$

$\Delta G^0_{bind,vacuum}$ is obtained by calculating the average interaction energy between receptor and ligand and by estimating the entropy change, generally with a normal-mode analysis. The solvated free energies are calculated by solving the linearized Poisson−Boltzmann (PB) equation[9] or the Generalized Born (GB)[10,11] model for each state. In both models, the solvent is described as a continuous medium with a certain dielectric constant (i.e., ε = 80, for water), while a low internal dielectric constant is assigned to the solute (generally, ε_{in} = 1 for proteins). From the solution of PB or GB equation is obtained the electrostatic contribution to the solvation free energy (ΔG_{el}), to which a nonelectrostatic contribute (ΔG_{nonel}) is then added. Conventionally, ΔG_{nonel} is considered proportional to the surface area that is accessible to the solvent (SA).[3]

The average of the interaction energies between receptor and ligand is obtained by considering a pre-established number of snapshots for the complex, receptor, and ligand, all taken from the MD trajectory of the solvated complex. This offers the possibility to calculate each free energy component of eq 1 by starting from a single MD trajectory.

However, the correlation of results with biological activity might depend on the chosen simulation protocols. Indeed, Hou et al. recently reported two very comprehensive articles on

Received: January 18, 2013

   dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

MM-PB/GBSA calculations evidencing a high sensitivity for both methods toward the preset internal dielectric constant (ε_{in}).[12,13] Unfortunately, the authors noticed that an universal ε_{in}, suitable for all the investigated protein systems, was not found, even if a thorough analysis of the binding pocket might aid the choice of the proper value of ε_{in}. Indeed, for α-thrombin ligands, the Spearman correlation coefficient (r_s) between experimental and predicted binding energy raised from 0.45 to 0.69 and 0.81 by setting ε_{in} = 1, 2, or 4, respectively. Similar results were also reported by Yang and co-workers,[14] who studied the binding to α-thrombin of 28 different ligands obtaining the best correlation with experiments, in terms of Pearson's correlation coefficient (r^2), by setting ε_{in} = 4 (r^2 = 0.73). Conversely, when penicillopepsin was considered, ε_{in} = 1, 2, or 4 led to r_s = 0.29, 0.41, and 0.32, respectively, while for avidin complexes the authors obtained r_s = 0.92, 0.67, and 0.51 by varying ε_{in} to 1, 2, and 4, respectively. Coherently, for the same system, Genheden and Ryde observed a variation in r^2 of 0.60, 0.33, 0.21, 0.15, and 0.13 by setting ε_{in} = 1, 2, 4, 10, and 25, respectively.[15] Thus, it appears that ε_{in} has a relevant effect on the quality of MM-PB/GB prediction, but properly choosing ε_{in} for drug screening/design purposes might be nontrivial, especially when a set of ligands with known binding activities is not available.

MM-PB/GBSA binding energy predictions might also be improved by explicitly considering selected water molecules in the calculation; however, this strategy is controversial.[16−20] For example, Wong and co-workers reported that weak and strong binders of TCR-SEC3 systems could be discriminated only by accounting key explicit water molecules in MM-PBSA calculations.[16] However, in a recent article where MM-GBSA was used to calculate the binding affinities on a large and diverse selection of protein−ligand complexes,[19] it was reported that, by including crystallographic water molecules within 3.5 Å from the ligand, the predictive quality was weakly reduced, as evidenced by the lowering of r^2 from 0.63 to 0.58. A worse correlation obtained when including explicit solvent was also observed by Checa et al. in a study involving the calculation of the relative binding affinities of trypsin inhibitors.[20] Although the authors recommended the use of a water cap to properly balance the solute−solvent interactions, this produced a correlation that was equivalent to that obtained using only implicit solvation. On the other hand, Wallnoefer et al. reported that MM-GB/SA calculation on factor Xa provided significant correlation with experiments only when one key water, selected from the crystallographic position, or a set of water molecules, selected by analyzing B-factors from MD trajectory, was accounted.[21]

Thus, it appears clear that there is the global effort to find a way to improve the predictivity of MM-PB/GBSA calculations by focusing on the solvation term, either by modifying the ε_{in} parameter or considering explicitly water in the calculation.

Nevertheless, to our knowledge, examples of systematic investigations evidencing how the inclusion of differently populated hydration shells around the ligand influences the correlation between experimental data and computed binding energies are not reported.

Concerning our experiences, MD simulations were applied to investigate different biological systems[22−26] and MM-PB/GBSA calculations adopted to compute binding free energies in protein−ligand,[22] DNA−ligand,[23] and protein−protein complexes.[25] In the past few years, our group has been involved in the design and synthesis of compounds targeting topoisomerases,[23,27] which are well-known targets for cancer therapy.[28] With an interest in designing new camptothecin derivatives, we evaluated MM-PB/GBSA approaches to select candidate compounds to be synthesized. After having chosen a suitable test set of recently reported compounds with available experimental activities,[29] we conducted MD simulations followed by MM-PBSA and MM-GBSA calculation and results were correlated to experiments. Unfortunately, preliminary results were disappointing, with correlation coefficients r^2 between predicted binding energies and $-\log_{10}(IC_{50})$ as low as 0.19 and 0.02 for GB and PB methods, respectively.

This failure drove us toward investigating the role of water networks surrounding the ligand, leading us to the definition of an excellent case study suggesting how the consideration of an optimal number of explicit water in the ligand hydration shell might improve the predictivity of MM-PB/GBSA calculations.

In order to prove the reliability and robustness of our protocol, we tested it on three other protein systems previously considered for assessing the performance of MM-PB/GBSA methods,[12] the α-thrombin, and penicillopepsin complexes, chosen because of their quite low correlation coefficient obtained by using dielectric solvation with the default ε_{in}, and avidin complexes, where good correlations were already obtained by using dielectric solvation.[12,15,30,31]

The obtained results confirmed that the protocol reported herein works well with different protein−ligand complexes, thus providing an interesting and easily applicable method that improve the correlation of MM-PB/GBSA computed binding energies with experimental activity data at almost no additional computational cost.

MATERIALS AND METHODS

1. Preparation of Complexes. *1.1. Topoisomerase.* All models were derived from the 1K4T crystal structure.[32] The considered system is made of human topoisomerase I B, 22 pairs of bases of double helix DNA, and Topotecan as the ligand (see Figure 1A). The DNA filament composed by

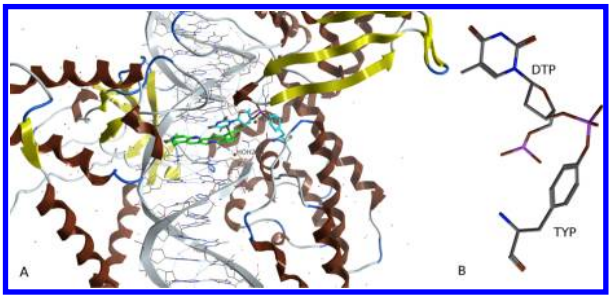


Figure 1. (A) The complex between topoisomerase, DNA, and topotecan (carbons colored in green); evidence on TYP and DTP residues (carbons colored in cyan). (B) The covalent bond between DTP and TYP.

nucleotides 1−22 is cleaved between thymine 10 and guanidine 11, and the latter is replaced by its 5′-thio derivative (TGP) because of technical reasons related to X-ray resolution.[32] Accordingly to the topoisomerase I cleaving mechanism,[33] the dangling 3′ phosphate group of thymine 10 is covalently bound to the Tyr 723 residue.

For those reasons, to prepare a suitable system for MD simulations, the TGP residue was replaced by guanosine monophosphate. A special attention has been given to the covalently bounded thymidine−phosphotyrosine system, for

dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

which the definition and charge parametrization of two nonstandard residues (DTP and TYP for thymidine and phosphotyrosine, respectively) was mandatory. Conformation and orientation independent partial charges for nonstandard residues were derived with the R.E.D. IV software,[34] accordingly to the *ff99SB* force field (RESP-A1 charge model),[35] using two conformations and four orientations for each structure. It is necessary to emphasize that, in the Amber *ff99SB* force field, a total charge of −1.0, −0.3079, and −0.6921 is attributed to internal bases, 5′ bases, and 3′ bases, respectively.[35] Moreover, when considering the binary complex between DNA and topoisomerase, there is an intact DNA strand, with an integer total charge of −21.0, and two cleaved strands: one is free and has an integer charge of −11.0, the other one has a free 5′ end, while the 3′ end is covalently bound to TYP 723. Since a net charge is requested for the protein−DNA system, the charge of the TYP residue has been restrained to −0.6921 while DTP has been treated as an internal base.

For DTP, an −OMe group has been used to cap the 5′ phosphate and the parametrization has been carried by restraining the final total charge of the uncapped DTP to −1.0. Conversely, to parametrize the TYP residue, three caps were added (see Figure S1 in the Supporting Information (SI)) and charge restraints were imposed as follows: a NH-methyl cap on C-terminus with a charge of 0.0, an acetyl cap on N-terminus with a charge of 0.0 and, on the phosphate group, a methoxyl cap with a charge of −0.3079, as this cap simulates the covalently bounded DNA strand. The backbone N and C atoms were restrained to the same charge value as reported in the force field for the standard tyrosine.[35] Finally, bending parameters for angles involving CA-OS-P and C-OS-P atom types have been calculated from the QM optimized structure using the *parmchk* tool of the Amber 11 suite (see SI for force field details).[36] Finally, the covalent bond between the phosphotyrosine and the thymidine residues has been created using the LEaP *bond* command (Figure 1B).

The complex between DNA and topoisomerase has been finally processed by H++ server by choosing default settings,[37] in order to establish the correct protonation state for each residue under physiological conditions. The resulting total charge of the DNA-topoisomerase complex was −12, HIS 22, 146, 167, 311, 376, and 342 were protonated on the N^{ε}, while HIS 66, 199, 206, and 315 were protonated on both N^{ε} and N^{δ}.

The ligand test set was made of camptothecin (CPT), topotecan (TTC), and seven other derivatives, for which an experimental IC_{50} value was available in the literature (see Figure S2 in the SI).[29] For ligands characterized by an additional stereocenter at position 5, experimentally evaluated as racemates,[29] both the α and β epimers were considered in MM-PB/GBSA calculation only for compound 12, as docking calculations previously reported evidenced a significant difference in predicted binding energies only for this compound,[29] while the β epimer was considered for compounds 7, 8, 20, and 22. Regarding this latter derivative, a keto−enol tautomerism might be possible through an intramolecular H-transfer between the C-5 ammonium and the amidic carbonyl (T1 and T2; see Figure S3 in the Supporting Information). Both tautomers were evaluated, but only the T2 tautomer, which is stabilized by resonance, provided significant correlation with experiments and was thus considered in the following discussion. Each derivative has been docked with MOE[38] (placement = Alpha Triangle, 800 000 minimum iterations and 5 000 000 maximum iterations; scoring = Affinity dG; the top 30 structures were subjected to force field refinement up to a gradient of 0.01 and rescored with Affinity dG). A simple pharmacophore model (see Figure S4 in the SI), consisting of an H-bond acceptor feature centered on the lactone moiety (F1, radius of 2.8 Å) and an aromatic feature centered on the quinolone moiety (F2, radius of 2.5 Å) was designed on the crystal structure of topotecan[32] and used as a pharmacophore restraint in all dockings. The top-ranked conformations (see the SI for coordinates of docked ligands) were then used to build the complexes. Partial charges were calculated by following the RESP-A1 model with R.E.D. IV,[34] using two conformations and four orientations.

1.2. α-Thrombin and Penicillopepsin. Since the crystallographic structures of the α-thrombin and penicillopepsin complexes with all the considered ligands were available (see Table S1 in the SI for Protein Databank (PDB) entries),[39−43] no further action, save the protonation of the two proteins with the H++ server, was necessary to build starting geometries for MD simulations. A total charge of −2 was obtained for α-thrombin. HIS 72, 95, 116, 145, and 271 of 1DWB, 1DWC and 1DWD structures and HIS 71, 94 115, 144, and 263 of 1D3D, 1D3P, 1D3Q and 1D3T were protonated at N^{ε}. A total charge of −22 was obtained for penicillopepsin, where HIS 54, 98, and 159 were protonated at N^{ε}. Partial charges for the α-thrombin ligands (see Figure S5 in the SI) were derived as previously described for camptothecin derivatives (see the SI for charge details).

Penicillopepsin ligands APT, APU, APV and APW (Figure S6 in the SI) were considered as non-natural tetrapeptides containing a divaline group substituted with isovaleric acid (IVA) at the N-terminus and with one of the nonstandard residues LTA, STA, DFI, or DFO at the C-terminus. Partial charges were thus derived for IVA (capped with a NH-methyl group), and LTA, STA, DFI, and DFO (all capped with acetyl groups) and corresponding force field libraries were generated. Conformations were generated through the low-mode molecular dynamics conformational search implemented in MOE, using default parameters, and charges computed with R.E.D. IV as explained earlier (see the SI for additional force field details). Conversely, charge parametrization for PP4, PP5, and PP6 ligands was done on the complete structures by restraining the atoms corresponding to the central valine residue to the same charge value as reported in the *ff99SB* force field for valine (see SI for details).[35]

1.3. Avidin. Since the crystallographic structure of avidin−biotin (BTN1) was the only one available (1AVD),[44] the starting geometries of the six biotin analogues (BTN2-BTN7) were manually generated from the avidin−biotin complex using MOE software. It has been shown that the neutral form of the guanidinium group in BTN2 and BTN5 is dominant when it is bound to the protein,[45] therefore, the neutral form of the guanidinium group for these ligands was used in our simulations. Partial charges for the biotin analogues were derived as previously described for camptothecin derivatives (see the SI for details).

The protein structure was protonated through the H++ server, obtaining a total charge of +9 with HIS 48 and 172 being protonated at N^{ε}.

In all cases, QM geometry optimization and electrostatic potential calculation were performed at the HF/6-31G* level, accordingly to the *ff99*SB force field,[35] by using the Gaussian09 software package.[46]

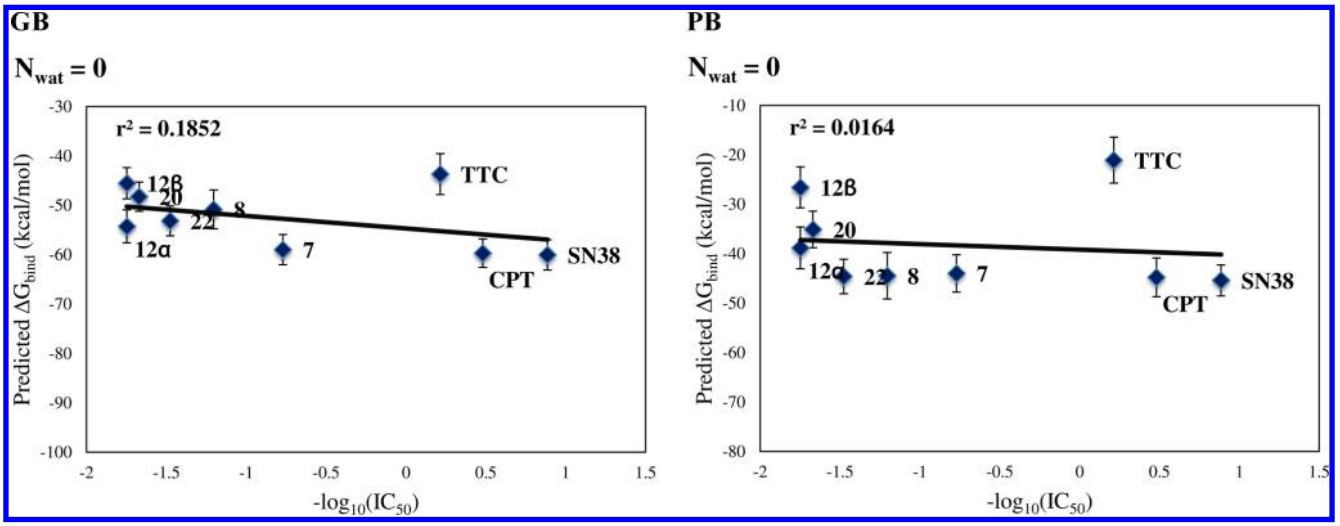


Figure 2. Correlation between $-\log_{10}(IC_{50})$ and the predicted ΔG_{bind} value obtained with the GB method (left) and the PB method (right).

2. Molecular Dynamics. MD simulations were carried out with the *pmemd* module of the Amber 11 package,[36] using the *ff99SB*[35] and *gaff* force fields.[47] In each case, the system total charge was neutralized by adding the proper number of Na^+/Cl^- ions and solvent, a cubic box of TIP3P water, has been added up to a distance of 10 Å from the solute. The systems were then relaxed by minimizing hydrogens, ions, and waters (1000 cycles of steepest descent and 5000 cycles of conjugated gradient). The solvent box was then equilibrated at 300 K by 45 ps of NVT and 45 ps of NPT simulation. This step was followed by a minimization involving side chains, water, and ions and by a total minimization (2500 cycles of steepest descent and 5000 cycles of conjugated gradient) with restraints applied on backbone atoms (10.0 kcal/mol) and on ligands (5.0 kcal/mol). The systems were then heated up to 300 K in 6 steps of 20 ps each (ΔT = 50 K), where backbone and ligand restraints were reduced from 10.0 to 5.0 and from 5.0 to 0.5 kcal/mol, respectively. Full equilibration was then performed in NVT ensemble (100 ps, with a restraint of 5.0 and 0.5 kcal/mol on the backbone and ligands, respectively) and in the NPT ensemble (4 steps of 100 ps each, reducing backbone and ligand restraints from 5.0 to 2.0 and from 0.5 to 0.2 kcal/mol, respectively, followed by a 1 ns NPT equilibration with backbone and ligand restraints of 1.0 and 0.1 kcal/mol, respectively). Finally, unrestrained production runs were performed at 300 K for 4 ns, which is a length that is considered adequate for similar calculations.[12] An electrostatic cutoff of 8.0 Å, a time step of 0.002 ps, and the SHAKE algorithm, constraining bonds involving hydrogens, were applied to all calculations.[48] Root-mean-square deviation (RMSD) analyses of receptor backbone and ligand atoms were made to assess the system stability. With regard to avidin complexes, the analyses have been conducted on each of the two monomers, because the avidin sites are independent of each other.[49]

In some cases (CPT, 7, 12*β*, BEN, MIT, PP4, and PP6), poor RMSD convergence was observed, so the NPT equilibration step was extended to 2 ns. Unrestrained production runs were made in the above-described conditions and most of the RMSD analyses showed convergence to values between 1.0 and 3.0 Å (see Figures S8–S12 in the SI). The complex between MIT and *α*-thrombin still did not show RMSD convergence within the adopted simulation time. Convergence was observed by prolonging the simulation by an additional 4 ns (see Figure S10 in the SI); however, MM-PB/GBSA analysis performed on the last ns of 4 ns or 8 ns trajectories did not provide significant differences (see Table S6 in the SI), and for this reason results obtained from the shorter run are reported in the following discussion, coherently with the protocol adopted for all other complexes.

Both MM-PBSA and MM-GBSA analyses were performed by using the MMPBSA.py python script implemented in the Amber 11 package.[50] Analyses were conducted on the last ns of 4 ns of production run trajectory by selecting 100 evenly spaced out snapshots. The atomic radii developed by Onufriev and coworkers (*igb* = 5)[51] was chosen for all GB calculation, and a salt molar concentration in solution was set at 0.15 M in both GB and PB calculations (*saltcon* and *istrng* parameters, respectively). The default PB solver implemented in the *sander* module was used for PB calculation and, unless differently specified, default parameters were adopted.

The entropy term in the binding energy calculations reported herein was neglected, considering that the benefits of including this term are controversial[12,14,21,52] and entropy estimations by normal-mode analysis are rather consuming, in terms of CPU time.[53]

When a ligand hydration shell had to be considered in MM-PB/GBSA analyses, corresponding trajectories were obtained using the *ptraj* keyword "*closest*", which allows one to retain only the requested number of those water molecules that, in each frame of the solvated MD trajectory, are the closest to the atoms specified in the mask (the ligand atoms, in our case). For performance reason, it has been found convenient to sample the requested snapshots with the *ptraj* "*offset*" keyword, and successively perform the "*closest*" analysis on the reduced trajectory. MM-PB/GBSA were then run by setting "*strip_mdcrd* = 0" (avoid the stripping of water molecules) and "*interval=1*" (consider all frames in the MD trajectory) in the input file. The water molecules (10, 20, 30, 40, 50, 60, or 70, depending from the chosen N_{wat}) defining the ligand hydration shell were then included in both the complex and receptor files and considered as part of the receptor atoms. Indeed, we noticed that considering water molecules as part of the ligand atoms provided worse correlation and higher standard deviations to the computed ΔG_{bind}.

dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

Unless differently specified, the square of Pearson's correlation coefficient (r^2) between computed binding energies and available experimental values such as the $-\log_{10}(IC_{50})$ (for topoisomerase) and ΔG_{bind} (for α-thrombin, penicillopepsin, and avidin) was used as an evaluation metric.

With regard to avidin complexes, MM-PB/GBSA analyses were performed on separate monomers and the results were averaged.

All hydrogen-bond analyses of ligand−water interactions were performed on the last ns of production run with VMD,[54] requesting a donor−acceptor distance of 4.0 Å and an angle cutoff of 30°. The same software has been used to visualize grid density maps, generated with a *ptraj* analysis of the entire production run by setting a cubic box (50 Å × 50 Å × 50 Å, mesh = 0.5 Å) centered on the ligand center of mass.

RESULTS AND DISCUSSION

1. Topoisomerase. Binding energies were initially evaluated for CPT derivatives by using MM/PBSA and MM/GBSA methods with default settings, i.e. in the absence of explicit water molecules. In both cases, results were discouraging with correlation coefficients between computed binding energies and $-\log_{10}(IC_{50})$ as low as 0.02 and 0.19 for the two methods, respectively (Figure 2). MM-PB/GBSA analyses were then repeated by including in the receptor mask the two crystallographic water molecules interacting with both the ligand and the binary topoisomerase−DNA complex (residues HOH 28 and 133 in the 1K4T PDB file),[32] but no improvement was observed. This failure could be attributed to the mobility of water residues during the MD simulations, since we observed that even hydrogen-bound waters can rearrange, with a specific residue being replaced by a neighboring one. The water bridge between the receptor and the ligand is thus maintained during the entire simulation, but not necessarily involving the same water molecule.

Hydrogen-bond analyses performed for TTC and SN38 complexes confirmed this assumption (see Tables S7 and S8 in the SI). For example, we observed that the C3═O of TTC is involved in hydrogen bonds with water molecules for 73% of the simulation time, but five different water residues contribute in determining this occupancy: WAT16384 (37.40%), WAT20971 (25.60%), WAT28324 (4.30%), WAT22562 (3.20%), and WAT20947 (2.50%). The same occurs for all the TTC atoms which can be involved in a hydrogen bond: C14═O, N6, O2, and OH bound to C9. In particular, the OH bound to C9 interacts with 10 different water residues for ∼20% of the MD simulation. A similar behavior is observable for SN38, whose atoms create a network of hydrogen bonds involving ∼20 water residues. Those findings do not conflict with results from X-ray analysis, as it should be remembered that crystallographic residues are identified as mean electron densities and those densities might be determined by multiple water residues which concur in occupying a well-defined position.[55] Coherently with this assumption, the presence of regions surrounding the ligand and showing a high water density during the MD simulation was verified by a grid analysis performed for water oxygen atoms (Figure 3).

Results show high-density regions matching crystallographic waters observed in 1K4T.pdb structure. Most of those regions are found in proximity of those TTC atoms involved in hydrogen bonds with solvent residues. Based on the above observations, we evaluated the possibility to perform MM-PB/GBSA calculations by explicitly considering hydration shells

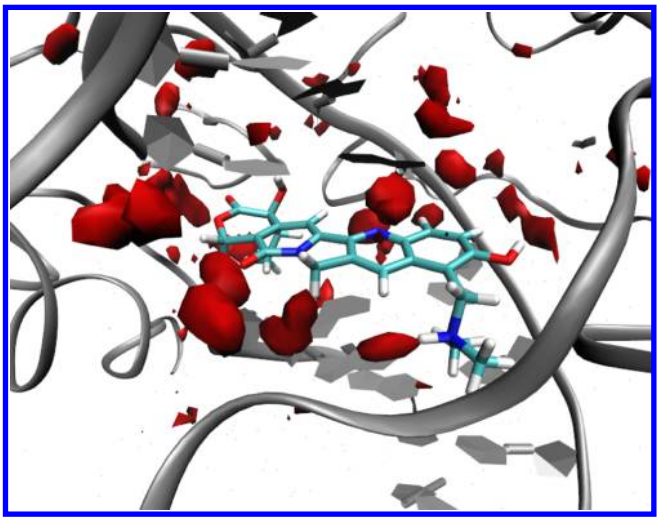


Figure 3. Water density plots obtained by grid analysis (*ptraj*; grid box = 50 Å × 50 Å × 50 Å, mesh = 0.5 Å; visualization with VMD specifying an isovalue of 45) of the topoisomerase−DNA−TTC complex.

around the ligand, formed by a constant number of water molecules selected from the MD trajectory (N_{wat}). This strategy should allow one to overcome questions related to the swapping of water residues during the MD run, since the solvent molecules of the hydration shell are selected frame-by-frame by the *ptraj* "*closest*" command to be those having the shortest distance between water and ligand center of masses. In the meanwhile, N_{wat} should not be chosen too high in order to avoid a loss in calculation performance in terms of both speed and quality, being this latter possibly affected by noises introduced by unnecessary water molecules.

For each topoisomerase complex, we thus generated five systems, including ligand hydration shells with N_{wat} = 10, 20, 30, 40, and 50, respectively. MM-GBSA and MM-PBSA analyses were then carried out for each system, and the resulting binding energies were correlated with the pharmacological data. Figure 4 reports the structure of the topoisomerase−DNA−topotecan ternary complex, where a 20-molecule hydration shell (N_{wat} = 20) is considered. Table S2 in the SI

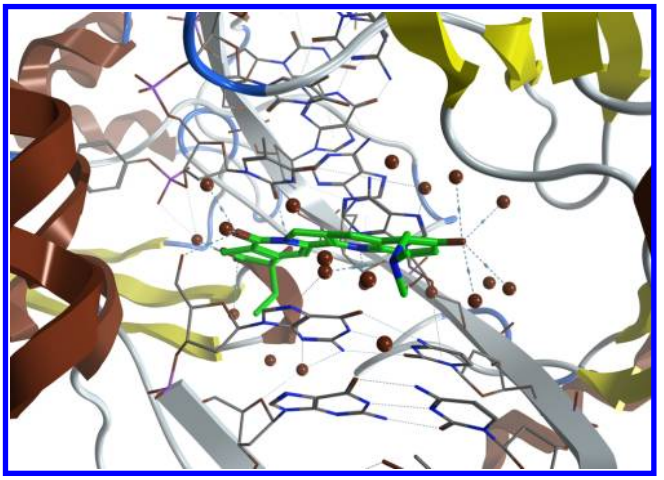


Figure 4. Complex between topoisomerase, DNA, and topotecan (carbon colored in green), with a ligand hydration shell populated by 20 water molecules (red).

dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

reports all MM-PB/GBSA computed binding energies for the topoisomerase I system, while graphs showing correlations with experimental data are reported in Figures S13 (GB) and S14 (PB) in the SI.

Decidedly better results were obtained by both the PB and GB methods for each of the considered hydration shells (Figure 5), with r^2 between the calculated binding energies and

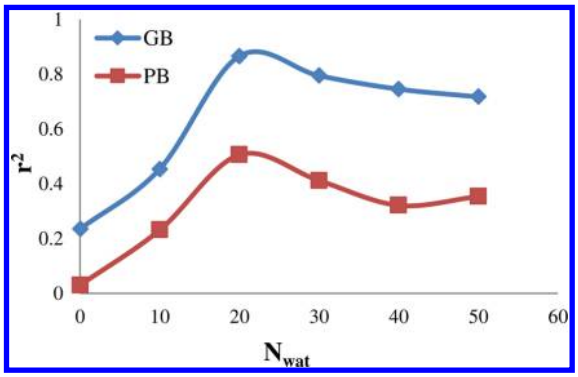


Figure 5. Variation of r^2 in dependency of N_{wat} for topoisomerase. Results of MM-PBSA and MM-GBSA calculations are reported in red and blue, respectively.

experimental $-\log_{10}(IC_{50})$ raising up to 0.51 for PB and to 0.87 for GB (Figure 5), both obtained with N_{wat} = 20 (see Figure 6).

By analyzing the dependency of r^2 on variations of N_{wat} (Figure 5), it is evident that the correlation sharply increases while switching from N_{wat} = 10 to N_{wat} = 20, while we assist to a small decrease of r^2 with an hydration shell of 30 or more residues, although the correlation remains acceptable (for GB, r^2 = 0.79, 0.74, and 0.71 for N_{wat} = 30, 40, and 50).

These results suggest that, by considering only 10 water residues around the ligand, all the explicit interactions between the examined system and the solvent are not correctly or sufficiently treated. At the same time, using a hydration shell that is too populated does not give further improvement, in this case being even detrimental. As suggested previously, this might happen due to background noises generated by the inclusion of a large number of solvent molecules around the ligand; this noise increases as the solvation sphere becomes wider, thus counteracting the benefits of the explicit consideration of solute−solvent interactions.

It appears that the increase in r^2 is, in large part, caused by the better correlation obtained for TTC. This might be explained by comparing the results of hydrogen-bond analyses conducted on stripped trajectories (N_{wat} = 10, 20, and 30) for TTC CPT and SN38 (see Tables S9−13 in the SI), the latter of those are chosen as examples scarcely affecting the variation of r^2. Hydrogen bonds with an occupancy above 5% were then analyzed and, for SN38 and CPT, we observed that the number of hydrogen bonds remained constant at the increase of N_{wat}. Conversely, an increment in the number of hydrogen bonds between TTC and solvent residues was observed when shifting from N_{wat} = 10 to N_{wat} = 20, while no further increase was found when shifting from N_{wat} = 20 and N_{wat} = 30 (see Tables S9−11 in the SI). These results justify the fact that the highest r^2 was obtained for N_{wat} = 20 and suggest that a better correlation might be obtained by explicitly considering not only those few water residues making stable hydrogen bonds bridging the ligand and the receptor, but also those water involved in transient interactions, but contributing in determining the water buffer between the ligand and the binding site residues.

Improvement in correlation when an hydration shell is considered can also be observed for the fluorinated derivative (12). By analyzing the IC_{50} values of CPT derivatives included in the dataset, it can be deduced that substitution at the C12 position of the C-ring is detrimental for the activity (see compounds 7, 8, 12, 20, and 22) and this effect is more pronounced when a hydrogen-bond acceptor is found at the C12 (compounds 12, 20, and 22). Although the same IC_{50} has been measured for 12α and 12β epimers,[29] MM-PB/GBSA calculations with N_{wat} = 0 notably overestimated the activity of the 12α epimer, while convergence between 12α and 12β ΔG_{bind} predictions was observed when a water shell is explicitly considered, mainly due to a better estimation for 12α. This may be explained by observing the water density plots resulting from grid analysis (see Figure S21 in the SI) on 12α and 12β MD trajectories, indicating that, in 12α, the F atom is more solvent-exposed.

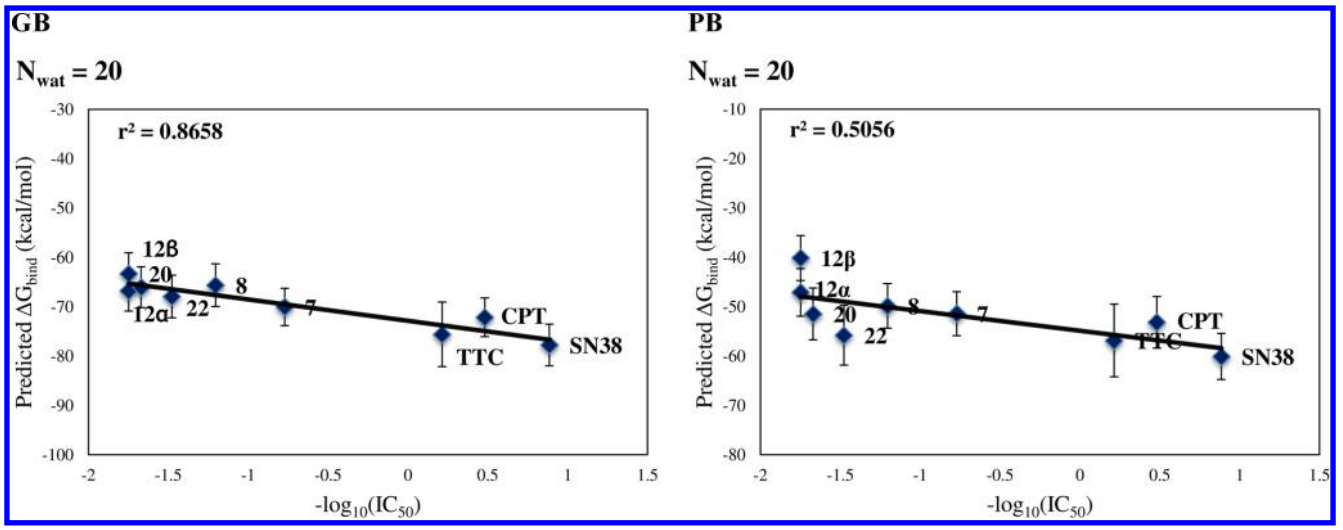


Figure 6. Correlation between $-\log(IC_{50})$ and predicted binding energies (kcal/mol) obtained with GB (left) and PB (right) methods considering a 20-molecule hydration sphere around the ligand.

dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

It must be noted that better correlations were generally observed for the GB method, coherently with results observed by other authors.[12,39] A comprehensive comparison of PB and GB methods goes beyond the scope of this article, but considering that GB calculations are performed at a fraction of the computational cost needed for PB calculations (in our examples, GB calculations, on average, required 6 less time units than PB), the former might be better suited for drug discovery/design applications, as also suggested by others.[12,31,56] Even if both MM-PBSA and MM-GBSA analyses were performed on all considered systems, in the following discussion, only the latter method was considered while results for MM-PBSA calculations are reported in the SI.

1.1. α-Thrombin. In this case, a good correlation between computed binding energies and experimental free energies of binding (ΔG_{bind}) was obtained even without considering an explicit hydration shell (r^2 = 0.67 (r_s = 0.82)). However, we observed a correlation that was higher than that reported for the same system by Hou et al. (r_s = 0.45) when the default dielectric constant was used.[12] Probably, this discrepancy can be attributed to the more-accurate charge parametrization protocol herein adopted, where ab initio HF/6-31G* instead of AM1[57] geometries were used and conformation and orientation independent RESP charges were derived by the R.E.D. procedure.[34] It should be considered that such an increase in accuracy is obtained at the expenses of computation time (depending on ligand dimension, the total charge parametrization procedure, considering two different conformations and four different spatial orientations, requires, on average, ∼3 h per ligand on a workstation equipped with a quad core 2.13 GHz Xeon processor). The choice between higher accuracy or faster calculations should, in any case, be up to the user's need, and it is not within the scope of this article to argue on this point.

By analyzing the dependency of r^2 to N_{wat} (Figure 7), it is evident that a worse correlation was found for N_{wat} = 10 when

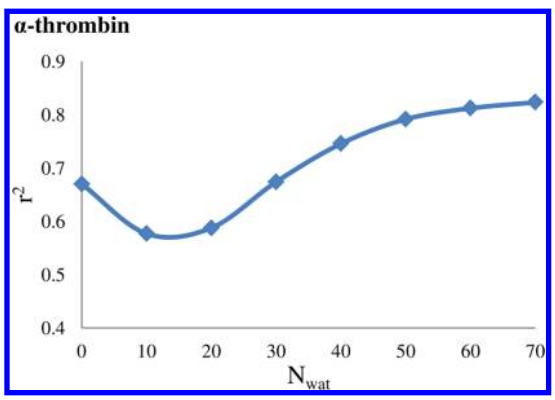


Figure 7. Variation of r^2 in the dependency of N_{wat} for α-thrombin.

compared to dielectric solvation only (r^2 = 0.58 and 0.67 for N_{wat} = 10 and 0, respectively) suggesting that, in this case, a small hydration shell only introduces noise, without improving the treatment of solute−solvent interactions. However, when increasing N_{wat}, r^2 consequently raised and reached, for N_{wat} = 50, the rather good value of 0.78. Since, for N_{wat} up to 50, a convergence of r^2 to a plateau value was not observed, the analysis was repeated for N_{wat}= 60 and 70, obtaining a small increase of r^2 up to 0.83.

Thus, for this example, we also assisted to an improvement in correlation between the predicted and experimental ΔG_{bind} values, but only when considering larger hydration shells, even if the increase in r^2 value is less significant (Figure 8) than that obtained for topoisomerase. Table S3 in the SI reports all MM-PB/GBSA computed binding energies for the α-thrombin system, while graphs showing correlations with experimental binding free energies are reported in Figures S15 (GB) and S16 (PB) in the SI.

The reason why the improvement in correlation between predicted and experimental data of penicillopepsin system is milder than that of topoisomerase might be attributable to a different role of solvent molecules in mediating ligand−receptor binding. Indeed, hydrogen-bond analyses conducted, for example, on compounds MID and BT2 only showed hydrogen bonds with occupancies between 0.10 and 3.30% (see Tables S14 and S15 in the SI). Another evidence supporting this hypothesis is provided by the evaluation of the water density around the ligand during the MD simulation time (Figure 9), decidedly lower than that observed for topoisomerase (Figure 3).

The mild improvement in r^2 obtained by including an explicit hydration shell around the ligands is probably due to the contribution given to the ligand−receptor interaction by those water residues involved in transient hydrogen bonds.

1.2. Penicillopepsin. For this system, a poor dependency on ε_{in} was evidenced by Hou et al. and generally a low correlation between theoretical and experimental ΔG_{bind} was observed (the obtained correlation coefficients, r_s, were 0.29, 0.41, and 0.32 for ε_{in} = 1, 2, and 4, respectively).[12]

In this case also, MM-PB/GBSA calculations where conducted by varying N_{wat} from 0 to 70. By correlating experimental and MM-GBSA ΔG_{bind} computed for N_{wat} = 0, we obtained an r^2 = 0.46 (r = 0.68), which is coherent with previously reported calculations.[12]

Conversely, the results obtained by including ligand hydration shells (N_{wat} = 10−70) showed, in this case also, an improvement in correlation between the predicted and experimental ΔG_{bind} values (see Figure 10).

In this case, a constant increase in r^2 was observed while rising N_{wat} from 0 to 70, with a sharp hike for 20 ≤ N_{wat} ≤ 30 (r^2 = 0.70) and a mild convergence to a plateau value of 0.79 obtained for N_{wat} = 70. The improvement in r^2 obtained by including a solvation shell around the ligands could be explained by the results of hydrogen-bond analyses. For example, in APU, the backbone C═O of VAL325 is involved for a total of 19.10% of simulation time in hydrogen bonds with either WAT1003 (10.90%), WAT9795 (8.10%), or WAT7021 (0.10%), while in APT, the N28 of LTA327 is involved in hydrogen bonds with water molecules (WAT7302, WAT7512, and WAT10356) for 54.80% of the considered simulation time (see Tables S16 and S17 in the SI). Furthermore, areas with a high water density were located by grid analysis around the ligands, as can be seen for APT in Figure 11.

An evident effect of considering a ligand hydration shell is the better estimate of the binding affinity of the APU ligand, which was severely underestimated when only dielectric solvation was used (see Figure 12).

The improvement of r^2 obtained when explicit hydration shells are considered is attributable largely to this ligand. Thus, coherently with analyses done for topoisomerase ligands, we compared results from hydrogen-bond analyses performed on the trajectories of APU, APT, and PP5 complexes obtained by

dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

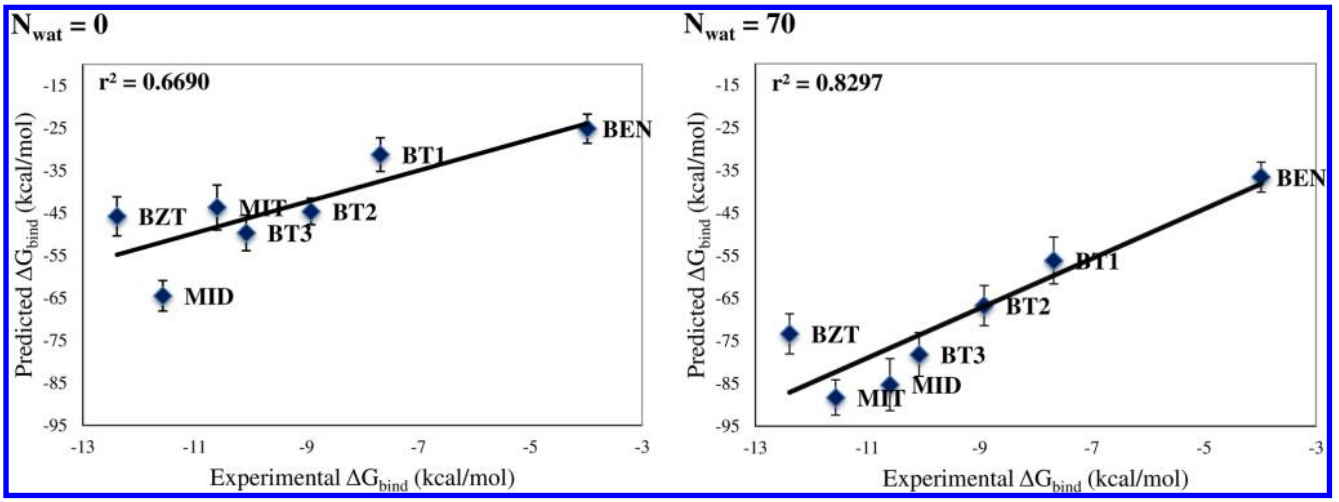


Figure 8. Correlation between experimental free energy of binding and predicted binding energies obtained for α-thrombin with N_{wat} = 0 (dielectric solvation only) and N_{wat} = 70.

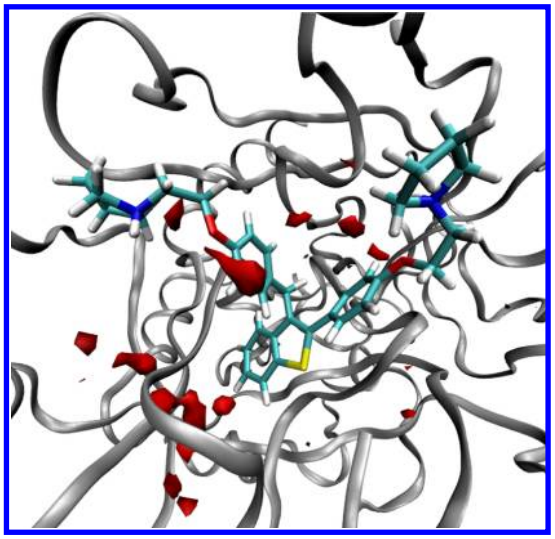


Figure 9. Water density plots obtained by grid analysis (*ptraj*; grid box = 50 Å × 50 Å × 50 Å, mesh = 0.5 Å; visualization with VMD specifying an isovalue of 45) of an α-thrombin−BT2 complex.

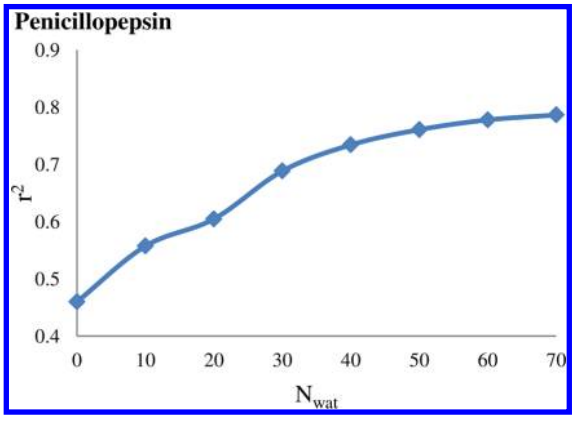


Figure 10. Variation of r^2 in dependency of N_{wat} for penicillopepsin.

stripping the solvent, except for the water molecules closest to the ligands' center of mass (N_{wat} = 10, 20, or 70; see Tables S18−S21 in the SI). Analogous to what was observed for TTC, we noticed that the number of hydrogen bonds with an

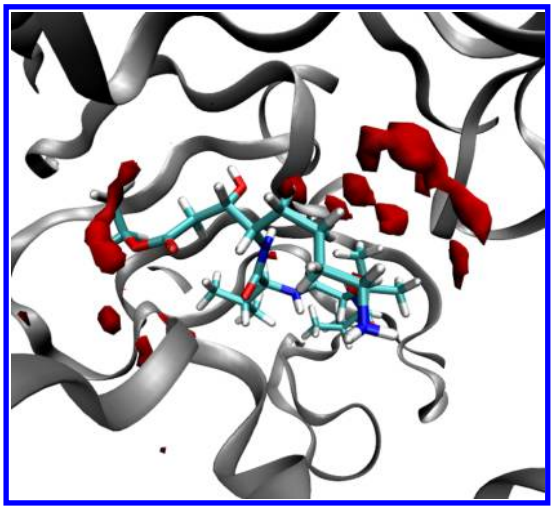


Figure 11. Water density plots obtained by grid analysis (*ptraj*; grid box = 50 Å × 50 Å × 50 Å, mesh = 0.5 Å; visualization with VMD specifying an isovalue of 45) of penicillopepsin−APT complex.

occupancy of ≥5% in which APU is involved increased from 41 to 70 when varying N_{wat} from 10 to 20, then remained constant up to N_{wat} = 70 (see Tables S18 and S19 in the SI). Conversely, the number of hydrogen bonds, with occupancies of >5%, found between water and APT or PP5, increased by only 6 for APT and 2 for PP5, while varying N_{wat} from 10 to 20−70 (see Tables S20 and S21 in the SI). The slight improvement in r^2 observable while varying N_{wat} from 30 to 70 is reasonably due to the contribution of hydrogen bonds between ligands and solvent having occupancies of <5%.

Table S4 in the SI reports all MM-PB/GBSA computed binding energies for the penicillopepsin system, while graphs showing correlations between predicted and experimental binding energies are reported in Figures S17 (GB) and S18 (PB) in the SI. In this case, it can be observed that the PB method provides better correlation than GB, although positive ΔG_{bind} values are obtained for the former.

1.3. Avidin. This system was chosen because of the high correlation between predicted and experimental binding energy obtained by Hou et al. (ε_{in} = 1, r_s = 0.92), as well as by other authors,[15,30,52] which reported r^2 values generally above 0.60.

dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

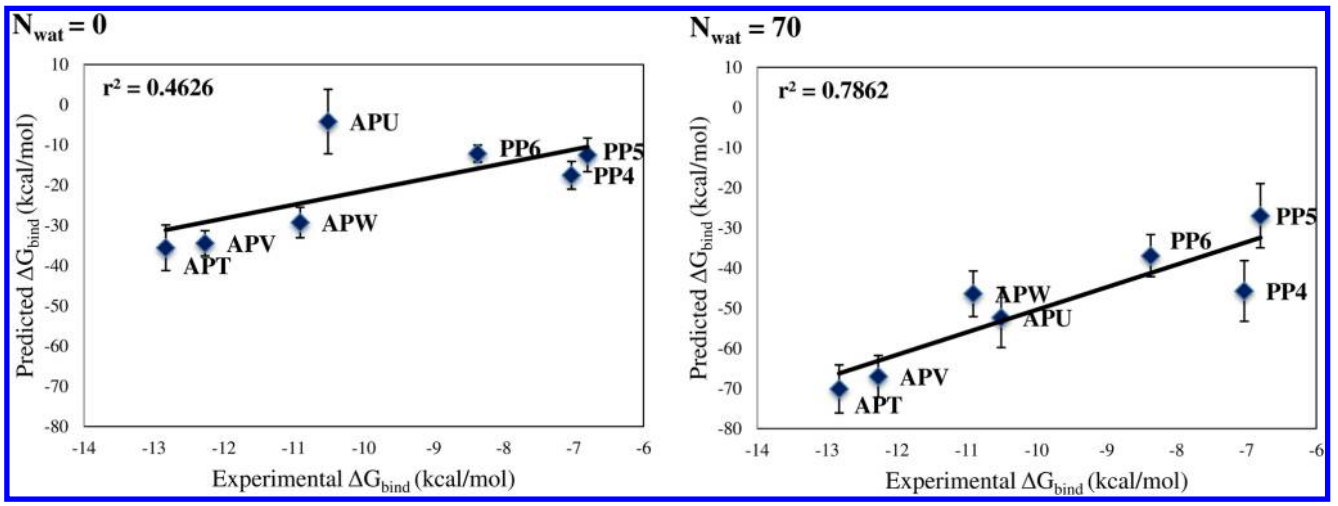


Figure 12. Correlation between experimental free energy of binding and predicted binding energies obtained for penicillopepsin with N_{wat} = 0 (dielectric solvation only) and N_{wat} = 70.

By only considering the dielectric solvation (N_{wat} = 0), we also obtained a good correlation with an r^2 = 0.72. Only minor improvements in correlation and immediate convergence are obtained using N_{wat} values from 10 to 70 (r^2 = 0.83 and 0.85, respectively, as shown in Figure 13).

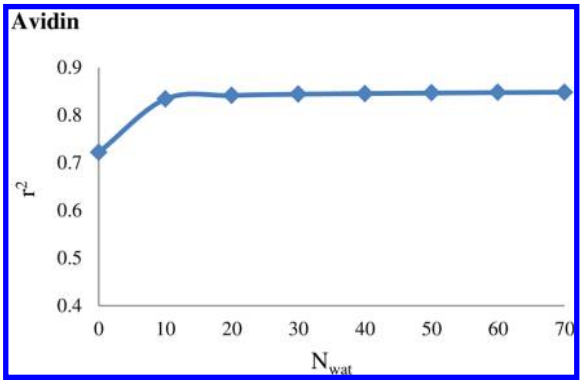


Figure 13. Variation of r^2 in dependency of N_{wat} for avidin.

The correlation between MM-GBSA predicted and experimental ΔG_{bind} for N_{wat} = 0 and 70 are shown in Figure 14, while Figures S19 and S20 in the SI show correlations between experimental and predicted binding energies obtained for all avidin ligands with GB and PB methods, respectively.

Detailed MM-PB/GBSA computed binding energies are reported in Table S5 in the SI.

As observed for other system herein considered, the small increase in r^2 might be due to the explicit inclusion of a few water molecules that, although not firmly bridging the ligand−receptor interactions, contribute in defining a water buffer between the ligands and the avidin binding site. Nevertheless, the weak dependence of this system from the inclusion of explicit water is probably due to the burial of the avidin binding site, as also evidenced by the analysis of the avidin−biotin crystal structure,[44] where only two water molecule can be selected within 4.5 Å from the biotin ligand atoms in chain A and one in chain B. A grid analysis of the avidin−BTN2 complex also confirmed this observation, since no areas of relevant water density were observed in the proximity of ligand atoms (see Figure S22 in the SI). Moreover, from the

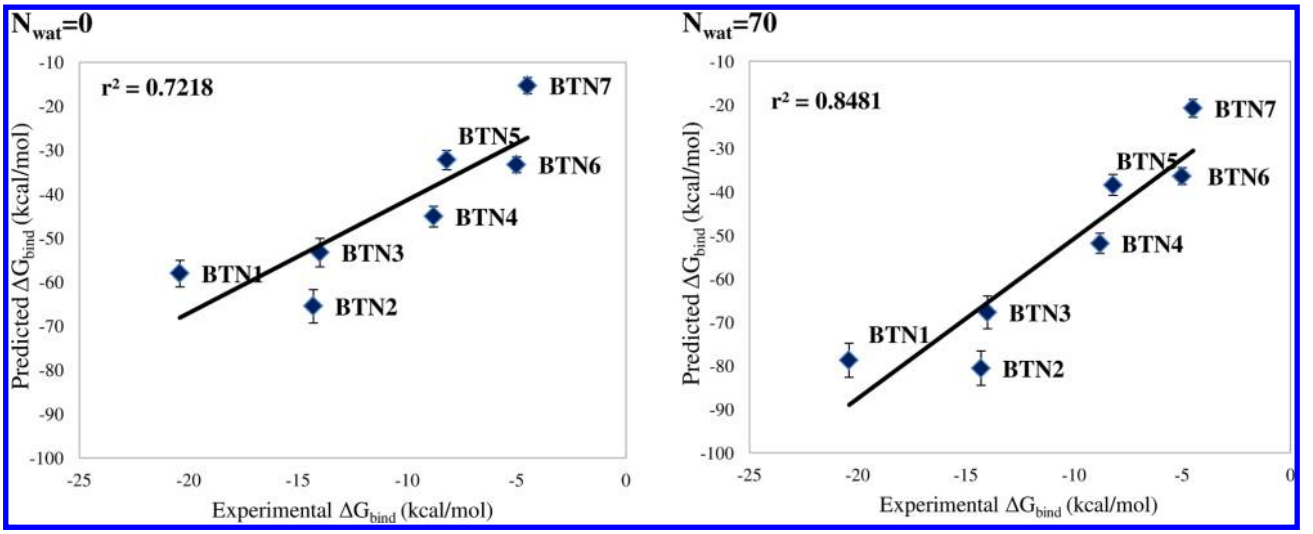


Figure 14. Correlation between experimental free energy of binding and predicted binding energies obtained for avidin with N_{wat} = 0 (dielectric solvation only) and N_{wat} = 70.

dx.doi.org/10.1021/ct400045d | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

hydrogen-bond analyses performed on BTN1 and BTN2, no hydrogen bonds with occupancies of ≥5% were observed between the ligand and water molecules (see Tables S22 and S23 in the SI).

CONCLUSIONS

In the present study, we evaluated how hydration shells explicitly considered around the ligand atoms may affect the correlation between MM-PB/GBSA binding energies and biological activities (IC_{50} and ΔG_{bind} in the examples considered herein). Four different case studies—namely, the DNA-topoisomerase bound to nine camptothecin derivatives, α-thrombin, penicillopepsin, and avidin, each bound to seven ligands having different structures, physicochemical properties, and ionization states—were considered.

The number of water molecules of the ligand hydration shell included in the receptor mask (N_{wat}) was set from 0 to 50 for topoisomerase and from 0 to 70 for α-thrombin, penicillopepsin, and avidin with a ΔN_{wat} = 10, an interval considered adequate to catch the influence on the prediction quality of the hydration shell dimension. In all four examples, the correlation coefficient r^2 increased from a minimum of 13 (avidin) to a maximum of 68 (topoisomerase) when hydration shells were considered (see Table 1).

Table 1. Correlations between the Experimental $-\log_{10}(IC_{50})$ (Topoisomerase) or ΔG_{bind} (α-Thrombin, Penicillopepsin, and Avidin) and MM-GBSA Binding Energies Obtained for N_{wat} = 0−70

	topoisomerase	α-thrombin	penicillopepsin	avidin
N_{wat} = 0	0.19	0.67	0.46	0.72
N_{wat} = 10	0.45	0.58	0.56	0.83
N_{wat} = 20	0.87	0.61	0.60	0.84
N_{wat} = 30	0.79	0.69	0.69	0.84
N_{wat} = 40	0.75	0.76	0.73	0.85
N_{wat} = 50	0.72	0.80	0.77	0.85
N_{wat} = 60	NC[a]	0.82	0.78	0.85
N_{wat} = 70	NC[a]	0.83	0.79	0.85

[a]NC = not calculated.

Therefore, this study suggests that the inclusion of explicit water actually has a crucial role in MM-PB/GBSA analysis, although the inclusion of only a few specific solvent molecules, as those selected from crystal structures, might have no effect or even worsen the estimate of relative binding energies, as observed by other authors.[19,20] Conversely, the consideration, as part of the receptor, of an hydration shell populated by a defined number of water residues (N_{wat}), which are those that, in each frame of the MD trajectory, are the closest to the ligand center of mass, provided in the considered examples a variable improvement in the correlation of MM-PB/GBSA computed binding energies with biological activities.

The optimum value of N_{wat} actually depended on the system under examination; thus, fine tuning of this parameter might still be necessary to reach the maximum performance. However, based on our data, a value of $N_{wat} \approx 30$ could be considered as a default in MM-PB/GBSA calculations. It should also be considered that the method reported herein can be applied at a negligible computational cost, with respect to standard dielectric solvation. The major expense is due to the selection of water molecules by the "*closest*" *ptraj* command (see the Materials and Methods section for the description), which can be quite consuming if performed on the entire trajectory (10−30 min for the 1 ns trajectory, depending on system size, on a standard workstation). However, by selecting those snapshots to be considered for MM-PB/GBSA calculations with *ptraj* prior to the "*closest*" command, the time needed was reduced by a factor of 10. Under those conditions, the total time-per-ligand when including a hydration shell was <10% higher than standard dielectric-solvation calculations.

Therefore, the protocol described herein is a valid alternative to the variation of the solute internal dielectric constant; however, it does present the advantage that the inclusion of a hydration shell comprised of 30−50 molecules always provided improvements in correlation, while the proper selection of ε_{in} might be nontrivial and a wrong choice might severely affect the quality of predicted relative binding affinities.[12]

It should also be emphasized that, although the correlation measured by the square Pearson's coefficient, r^2, improved in all the examples considered herein by explicitly considering ligand hydration shells, the estimate of absolute binding free-energy estimate was actually worsened. Indeed, enhanced correlation with experiments does not promise better absolute values of binding free energies.[14] Moreover, the evaluation of correlation suffers by uncertainties due to relevant standard deviations of MM-PB/GBSA computed ΔG_{bind} that, in the herein adopted conditions, are close to 10% of the mean value. Although methods to reduce the uncertainty in MM-PB/GBSA ΔG_{bind} calculations were reported,[31] it should also be considered that standard deviations of 10% or above, with respect to mean value, are often observed for experimental data obtained for complex biological systems,[22,23,27,58,59] thus making a statistically rigorous comparison among different computational protocols difficult. Nevertheless, improvements in the predictivity of theoretical methods should also be evaluated as trends, and we are confident that the protocol described herein, because of its favorable benefit−cost ratio, might be useful for drug-discovery applications.

ASSOCIATED CONTENT

Supporting Information

Supplementary figures and tables, coordinates of all considered ligands, and additional force field information for nonstandard residues. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel.: +39 02 503 14480. Fax. +39 02 503 14476. E-mail: alessandro.contini@unimi.it.

Funding

This work was supported by the MIUR through the "FIRB—Programma 'Futuro in Ricerca'" (Grant No. RBFR087YAY).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the CINECA through the "Progetto Di Migrazione" for the availability of high-performance computing resources and support.

REFERENCES

(1) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for

the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44* (21), 3417−3423.
(2) Srivastava, H. K.; Sastry, G. N. Molecular Dynamics Investigation on a Series of HIV Protease Inhibitors: Assessing the Performance of MM-PBSA and MM-GBSA Approaches. *J. Chem. Inf. Model.* **2012**, *52* (11), 3088−3098.
(3) Massova, I.; Kollman, P. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov.* **2000**, *18* (1), 113−135.
(4) 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.
(5) Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35* (6), 358−365.
(6) Price, D.; Jorgensen, W. Improved convergence of binding affinities with free energy perturbation: Application to nonpeptide ligands with pp60src SH2 domain. *J. Comput.-Aided Mol. Des.* **2001**, *15* (8), 681−695.
(7) Xu, D. *Autodock2MMGBSA, A multi-level virtual screening rescoring and refinement scheme that combines consensus scoring, simulated annealing and MM-GBSA binding free energy methods*, American Chemical Society: Washington, DC, 2012; p NORM-153.
(8) Xu, D.; Sawaya, N.; McCammon, J. A.; Li, W. W. In *Autodock2MMGBSA, A multi-level virtual screening rescoring and refinement scheme that combines consensus scoring and MM-GBSA binding free energy methods*, American Chemical Society: Washington, DC, 2010; p COMP-78.
(9) Jackson, J. D. *Classical Electrodynamics*; Wiley: New York, 1999.
(10) Constanciel, R.; Contreras, R. Self consistent field theory of solvent effects representation by continuum models: Introduction of desolvation contribution. *Theor. Chim. Acta* **1984**, *65* (1), 1−11.
(11) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112* (16), 6127−6129.
(12) 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*, 69−82.
(13) Hou, T.-J.; Wang, J.-M.; Li, Y.-Y.; Wang, W. 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.* **2011**, *32*, 866−877.
(14) Yang, T.; Wu, J. C.; Yan, C.; Wang, Y.; Luo, R.; Gonzales, M. B.; Dalby, K. N.; Ren, P. Virtual screening using molecular simulations. *Proteins: Struct., Funct., Bioinf.* **2011**, *79* (6), 1940−1951.
(15) Genheden, S.; Ryde, U. Comparison of end-point continuum-solvation methods for the calculation of protein−ligand binding free energies. *Proteins: Struct., Funct., Bioinf.* **2012**, *80* (5), 1326−1342.
(16) 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*, 422−429.
(17) Hayes, J. M.; Skamnaki, V. T.; Archontis, G.; Lamprakis, C.; Sarrou, J.; Bischler, N.; Skaltsounis, A.-L.; Zographos, S. E.; Oikonomakos, N. G. Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: The role of water molecules examined. *Proteins* **2011**, *79*, 703−719.
(18) Freedman, H.; Huynh, L. P.; Le, L.; Cheatham, I. I. I. T. E.; Tuszynski, J. A.; Truong, T. N. Explicitly Solvated Ligand Contribution to Continuum Solvation Models for Binding Free Energies: Selectivity of Theophylline Binding to an RNA Aptamer. *J. Phys. Chem. B* **2010**, *114*, 2227−2237.
(19) Greenidge, P. A.; Kramer, C.; Mozziconacci, J.-C.; Wolf, R. M. MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *J. Chem. Inf. Model.* **2012**, *53* (1), 201−209.
(20) Checa, A.; Ortiz, A. R.; de Pascual-Teresa, B.; Gago, F. Assessment of solvation effects on calculated binding affinity differences: Trypsin inhibition by flavonoids as a model system for congeneric series. *J. Med. Chem.* **1997**, *40* (25), 4136−4145.
(21) Wallnoefer, H. G.; Liedl, K. R.; Fox, T. A challenging system: Free energy prediction for factor Xa. *J. Comput. Chem.* **2011**, *32* (8), 1743−1752.
(22) Ferri, N.; Corsini, A.; Bottino, P.; Clerici, F.; Contini, A. Virtual screening approach for the identification of new Rac1 inhibitors. *J. Med. Chem.* **2009**, *52* (14), 4087−4090.
(23) Ferri, N.; Radice, T.; Antonino, M.; Beccalli, E. M.; Tinelli, S.; Zunino, F.; Corsini, A.; Pratesi, G.; Ragg, E. M.; Gelmi, M. L.; Contini, A. Synthesis, structural, and biological evaluation of bis-heteroarylmaleimides and bis-heterofused imides. *Bioorg. Med. Chem.* **2011**, *19* (18), 5291−5299.
(24) Pellegrino, S.; Contini, A.; Clerici, F.; Gori, A.; Nava, D.; Gelmi, M. L. 1H-Azepine-4-amino-4-carboxylic acid: A new alpha,alpha-disubstituted ornithine analogue capable of inducing helix conformations in short Ala-Aib pentapeptides. *Chem.—Eur. J.* **2012**, *18* (28), 8705−8715.
(25) Contini, A.; Cappelletti, G.; Cartelli, D.; Fontana, G.; Gelmi, M. L. Molecular dynamics and tubulin polymerization kinetics study on 1,14-heterofused taxanes: Evidence of stabilization of the tubulin head-to-tail dimer−dimer interaction. *Mol. BioSyst.* **2012**, *8* (12), 3254−3261.
(26) Casoni, A.; Clerici, F.; Contini, A. Molecular dynamic simulation of mGluR5 amino terminal domain: Essential dynamics analysis captures the agonist or antagonist behaviour of ligands. *J. Mol. Graph. Model.* **2013**, *41*, 72−78.
(27) Ferri, N.; Beccalli, E. M.; Contini, A.; Corsini, A.; Antonino, M.; Radice, T.; Pratesi, G.; Tinelli, S.; Zunino, F.; Gelmi, M. L. Antiproliferative effects on human tumor cells and rat aortic smooth muscular cells of 2,3-heteroarylmaleimides and heterofused imides. *Bioorg. Med. Chem.* **2008**, *16* (4), 1691−1701.
(28) Pommier, Y.; Leo, E.; Zhang, H.; Marchand, C. DNA Topoisomerases and Their Poisoning by Anticancer and Antibacterial Drugs. *Chem. Biol.* **2010**, *17* (5), 421−433.
(29) Samorì, C.; Guerrini, A.; Varchi, G.; Fontana, G.; Bombardelli, E.; Tinelli, S.; Beretta, G. L.; Basili, S.; Moro, S.; Zunino, F.; Battaglia, A. Semisynthesis, Biological Activity, and Molecular Modeling Studies of C-Ring-Modified Camptothecins. *J. Med. Chem.* **2009**, *52* (4), 1029−1039.
(30) Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. An MM/3D-RISM Approach for Ligand Binding Affinities. *J. Phys. Chem. B* **2010**, *114* (25), 8505−8516.
(31) Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* **2010**, *31* (4), 837−846.
(32) Staker, B. L.; Hjerrild, K.; Feese, M. D.; Behnke, C. A.; Burgin, A. B.; Stewart, L. The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc. Natl. Acad. Sci.* **2002**, *99* (24), 15387−15392.
(33) Chhatriwala, H.; Jafri, N.; Salgia, R. A review of topoisomerase inhibition in lung cancer. *Cancer Biol. Ther.* **2006**, *5* (12), 1600−1607.
(34) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821−7839.
(35) Hornak, V. A., R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.
(36) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26* (16), 1668−1688.
(37) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A server for estimating pK_as and adding missing

hydrogens to macromolecules. *Nucleic Acids Res.* **2005**, *33* (Suppl 2), W368−W371.
(38) *MOE, MOE v2010.10*; Chemical Computing Group, Inc.: Montreal, Canada, 2010.
(39) Chirgadze, N. Y.; Sall, D. J.; Briggs, S. L.; Clawson, D. K.; Zhang, M.; Smith, G. F.; Schevitz, R. W. The crystal structures of human alpha-thrombin complexed with active site-directed diamino benzo[b]thiophene derivatives: A binding mode for a structurally novel class of inhibitors. *Protein Sci.* **2000**, *9* (1), 29−36.
(40) Banner, D. W.; Hadvary, P. Crystallographic analysis at 3.0-Å resolution of the binding to human thrombin of four active site-directed inhibitors. *J. Biol. Chem.* **1991**, *266* (30), 20085−20093.
(41) James, M. N. G.; Sielecki, A. R.; Moult, J. Crystallographic Analysis of a Pepstatin Analogue Binding to the Aspartyl Proteinase Penicillopepsin at 1.8 Angstroms Resolution. In *Peptides: Structure and Function, Proceedings of the of the Eighth American Peptide Symposium*, 1983; Hruby, V., Rich, D., Eds.; Pierce Chemical Company: Rockford, IL, 1983; pp 521−530.
(42) James, M. N.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. Crystallographic analysis of transition state mimics bound to penicillopepsin: Difluorostatine- and difluorostatone-containing peptides. *Biochemistry* **1992**, *31* (15), 3872−3886.
(43) Ding, J.; Fraser, M. E.; Meyer, J. H.; Bartlett, P. A.; James, M. N. G. Macrocyclic Inhibitors of Penicillopepsin. 2. X-ray Crystallographic Analyses of Penicillopepsin Complexed with a P3−P1 Macrocyclic Peptidyl Inhibitor and with Its Two Acyclic Analogues. *J. Am. Chem. Soc.* **1998**, *120* (19), 4610−4621.
(44) Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. Three-dimensional structure of the tetragonal crystal form of egg-white avidin in its functional complex with biotin at 2.7 A resolution. *J. Mol. Biol.* **1993**, *231* (3), 698−710.
(45) Green, Thermodynamics of the binding of biotin and some analogues by avidin. *Biochem. J.* **1966**, *101* (3), 774-780.
(46) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *G09, Gaussian 09*, Revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
(47) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157−1174.
(48) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.
(49) 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.
(50) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314−3321.
(51) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55* (2), 383−394.
(52) 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.
(53) Case, D. A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **1994**, *4* (2), 285−290.
(54) Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(55) Schiffer, C.; Hermans, J. Promise of Advances in Simulation Methods for Protein Crystallography: Implicit Solvent Models, Time-Averaging Refinement, and Quantum Mechanical Modeling. In *Methods in Enzymology*; Carter, C. W., Jr., Sweet, R. M., Eds.; Academic Press: New York, 2003; Vol. *374*, pp 412−461.
(56) Kongsted, J.; Ryde, U. An improved method to predict the entropy term with the MM/PBSA approach. *J. Comput.-Aided Mol. Des.* **2009**, *23* (2), 63−71.
(57) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902−3909.
(58) Ferri, N.; Bernini, S. K.; Corsini, A.; Clerici, F.; Erba, E.; Stragliotto, S.; Contini, A. 3-Aryl-*N*-aminoylsulfonylphenyl-1H-pyrazole-5-carboxamides: A new class of selective Rac inhibitors. *Med. Chem. Comm.* **2013**, *4* (3), 537−541.
(59) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: A web-accessible database of experimentally determined protein−ligand binding affinities. *Nucleic Acids Res.* **2007**, *35* (Database issue), D198−D201.