JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

A GRID-Derived Water Network Stabilizes Molecular Dynamics Computer Simulations of a Protease

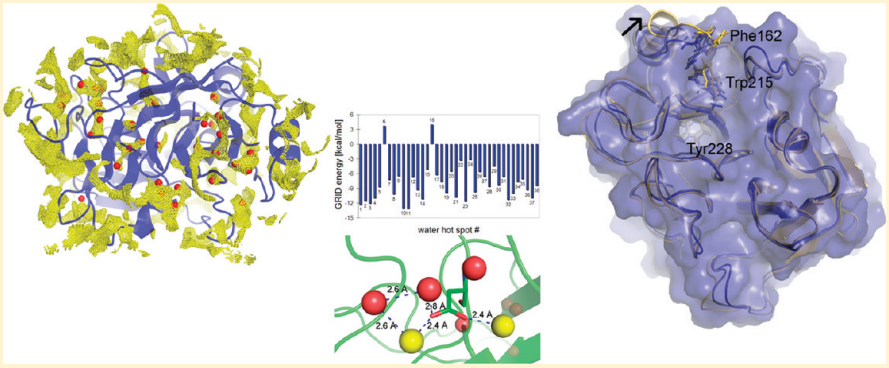
Hannes G. Wallnoefer,[†,‡] Klaus R. Liedl,[*,‡] and Thomas Fox[†]

[†]Computational Chemistry, Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co., KG, 88397 Biberach, Germany

[‡]Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 52a, 6020 Innsbruck, Austria

Supporting Information

ABSTRACT:



Structural water molecules are crucial for the stability and function of proteins. Recently, we presented a molecular dynamics (MD) study on blood coagulation factor Xa (fXa) to investigate the effect of water molecules on the flexibility of the protein structure. We showed that neglecting important water positions at the outset of the simulation leads to severe structural distortions during the MD simulations: A stable trajectory was obtained with a water set that was derived from all 73 X-ray structures of the protein. However, for many proteins of interest, only limited structural data is available, which precludes the merging of information from many X-ray structures. Here, we show that an in silico assembled water network, derived from molecular interaction fields generated with the GRID program, is a viable alternative to X-ray data. MD simulations with the GRID water set show a significantly improved stability over alternative setups without water or the X-ray resolved water molecules in the starting structure. The performance is comparable to a water setup derived from a recently presented clustering approach.

INTRODUCTION

Molecular dynamics (MD) simulations have become an essential tool for studying the flexible behavior of biomolecules.[1−5] Many examples in the literature have shown that (given a reasonable force field) they can be used to investigate both structural features and thermodynamic quantities, e.g., loop movements,[6−10] domain motions,[11,12] stability of proteins,[13−17] and free energies of binding of ligands.[18−23] Often, a standard setup procedure is employed which starts from an experimentally determined X-ray structure of the protein. Then the crystallographic water molecules are stripped, followed by soaking of the protein by a periodic box or a shell of water molecules. Afterward, a careful equilibration procedure is performed to not only remove close contacts and any hot spots which may have been created during the setup but also to allow the bulk water structure to adjust to the interactions presented and imposed by the protein.

Unfortunately this does not work for the serine protease factor Xa (fXa), which is very sensitive to the initial placement of water molecules. This is especially important for deeply buried hydration sites, which are unoccupied at the beginning of the simulation, as they are usually not filled by bulk water molecules during common simulation time scales in the order of nanoseconds. In some cases even empty water sites close to the protein surface are not occupied by water molecules from the bulk. This induces local distortions of the protein structure, unphysical movements, and the sampling of unreasonable or irrelevant parts of the phase space. As a prominent example in factor Xa, we showed that an empty water site close to the active site subpocket S4 leads to a fast conformational change of the Met168 side chain filling the empty site and, thus, changes the binding site geometry. This motion was irreversible during 12.5 ns of MD simulation.

A similar behavior was also observed in other serine proteases. Recently, over 20 water binding sites, which are conserved over

Received: March 23, 2011
Published: October 04, 2011

the whole protein family, were identified,[24−27] and it could be shown that several clusters of water molecules contribute significantly to the stability of the protein or were important for the function of the enzyme. Thus, it was proposed that water molecules should be considered as an important structural characteristic in serine proteases. However, the importance of water molecules was reported not only for serine proteases but also for many other protein systems. T4 lysozyme, one of the best studied proteins, was reported to have 62% of all water molecules resolved in 18 high-resolution X-ray structures to be conserved.[28] Nevertheless, not a single water position is occupied throughout all of the structures.

Another experimental indication that a water molecule plays an important structural role is the frequent occurrence in several members of a protein family at a comparable position.[29−31] Computer simulations of Subtilisin Carlsberg, which employed increasing concentrations of acetonitrile in the simulation box, showed that with increasing acetonitrile content important water molecules were replaced, resulting in significant changes in the secondary structure of the protein.[32]

As an example for the functional role of water molecules, we want to point to a channel of water molecules leading from the S1 subpocket of serine proteases to the protein surface. This channel provides the water molecules necessary for the peptide bond cleavage of the substrate, and consequently, this channel is populated with water molecules in all experimental structures with sufficient resolution.[33−36] (If some of these waters are not resolved this is probably due to low resolution of the experiment.)[37] Other studies showed that missing or badly placed water molecules in fXa may lead to computational artifacts.[38−42]

The serine protease fXa plays a prominent role in the blood coagulation cascade.[43−54] Therefore, it has been for a long time the target of drug discovery efforts, as inhibiting fXa may alleviate syndromes which are connected to an increased tendency of blood clot formation, e.g., thrombosis after surgery, or as a result of atrial fibrillation.[45,55,56]

Published MD simulations of fXa report fairly high deviations from the starting structure, well beyond the usually accepted 1.5 Å backbone root-mean-square deviation (RMSD) from the X-ray structure.[57−61] Recently, we showed that stable fXa simulations could be obtained if a well-chosen set of water molecules is placed in and around the protein prior to the usual MD simulation setup.[14] In particular, we found that an accurately predefined water network leads to faster converged simulations and that they stay closer to the starting structure.

For fXa, a large number of public X-ray structures are available,[62] which allowed us to pool all information about water molecules visible in the X-ray structures. To this end, we developed a clustering approach that merges all available X-ray water positions and clusters them based on an exclusion sphere algorithm[63] obtaining a set of 38 water molecules, which were added prior to the usual simulation setup.

However, this approach is only feasible if there are a large number of ideally well-resolved X-ray structures available. However, for many proteins of interest this prerequisite is not given. Therefore, we also evaluated an in silico approach; the underlying hypothesis is that structurally important water molecules should be located at energetically favorable positions.

One possibility to obtain these positions is the calculation of molecular interaction fields (MIFs), e.g., with the program GRID.[64,65] Here, the MIFs are generated by mapping the three-dimensional (3D) space around molecular targets with probes mimicking the chemical properties atoms or functional groups. As the probes are parametrized to match experimental data, not only enthalpic but also some entropic contributions are included (most parametrization is done to structural data and the distribution within the structures accounts for disordering effects). From the MIF generated with the "water" probe, the positions of important water molecules can be deduced.

GRID has been used previously to obtain water positions for computer simulations, e.g., for a cytochrome P450cam−ligand system[66,67] or acetylcholinesterase.[68] Another study determined not less than 755 water positions with GRID for cytochrome c oxidase.[69] However, all these studies concentrate on the in silico results, none compared the generated water positions to experimental data. To the knowledge of the authors, only Olkhova[69] et al. report that all of the 88 waters, which are resolved in the X-ray structure, are reproduced by GRID calculations.

However, GRID is by far not the only method applied for protein hydration probing. Similarly to GRID, Fold-X is a force field-based approach.[70,71] It is fitted to 74 X-ray structures and predicts positions for ions and water molecules within biomolecule structures with success rates of 76% for water and up to 97% for ions, respectively. Another empiric approach for the identification of protein interaction sites, which is applicable to protein hydration, was reported from Verdonk et al.[72,73] Superstar is based on small-molecule crystals. Interactions of different chemical groups (e.g., hydroxyl for water interaction) are analyzed toward their 3D distribution. Subsequently, the fragments are superposed, and interaction maps are generated. Based on these maps, interacting groups for structures are predicted. Successful applications were demonstrated for, e.g., L-arabinose-binding protein and dihydrofolate reductase.

MD simulations also have been used to evaluate protein hydration before. Henchman and McCammon for instance investigated the structure and dynamics of waters around acetylcholineesterase from a simulation trajectory.[74] They applied the well-known strategy to reduce water occupancy to hydration sites.[75] By averaging the water information for individual hydration[76] sites, conclusions on the number of neighboring atoms and buriedness are presented. Another simulation-based methodology is WaterMap.[57] It samples the hydration of a rigid binding site during time by MD, obtains hydration sites by clustering, and applies the inhomogeneous fluid theory[77−80] for calculation of enthalpic and entropic energies of the individual hydration sites. The main purpose is to estimate whether it is favorable for a potential ligand to displace water molecules from this hydration site or not. Another application was introduced recently by substituting the generalized Born surface area (GBSA) desolvation part of molecular mechanics/generalized Born surface area (MM/GBSA) by WaterMap energies for free energy calculations.[81,82]

A combination of simulation techniques and higher sophisticated water placement and energy evaluation is just add water molecules (JAWS).[83] In short, it places a grid onto the biomolecule and uses water molecules, which can appear and disappear. Free energy is then determined with an approximated Zwanzig equation approach.[84] Conformational sampling with Monte Carlo simulations using a grand canonical ensemble allowing fluctuations in the number of water molecules. Successful application is demonstrated for five protein targets in comparison with X-ray data.[83] An efficient inclusion of the JAWS methodology into binding affinity calculations was demonstrated for P38α MAP kinase inhibitors. Only an inclusion of important water

interactions with JAWS-placed water molecules yielded satisfying free energy calculations.[85] An approach designed for the analysis of water networks in protein−protein interfaces is WATGEN. It was fitted with the analysis of polar interactions of 126 protein−peptide complex structures with AMBER8 energies.[86] Experimental water sites are reproduced in 72% within 1.5 Å and in 88% within 2 Å. Furthermore, hydration sites not reported from experimental data could be identified. A follow-up study showed WATGEN applicability for 224 protein−RNA complexes.[87]

The above discussion is far from being comprehensive for methods to analyze protein hydration and water placement, but it illustrates that many approaches of differing complexity were proposed. In the present study, we want to evaluate if GRID is capable of generating a reasonable internal water network for fXa. The applicability in MD simulations is tested and compared to previously published results of experimentally derived water networks.

METHODS

Protein Structure. For the GRID calculations and the subsequent MD simulations, the structure with the PDB code 2boh was chosen. It has a comparatively good resolution of 2.2 Å and a large number of resolved X-ray waters (221), and in contrast to the apo-structures of factor Xa, all amino acid residues are resolved. The ligand and the light chain were removed. One calcium and one sodium ion, which are known to be structurally important,[88,89] were either kept or modeled according to other X-ray structures. The structures were protonated with the 3D-Protonate tool of MOE 2008.10[90] and then manually reviewed (especially for the correct protonation states of the histidine residues).

GRID Calculations. The program GRID[64,65] places a three-dimensional grid around the structure, and at every grid point the interaction energy between the protein and a predefined probe is calculated as a sum of an electrostatic (similar to coulomb with dielectrics and geometric parameters), a van der Waals (6−12 potential), and a hydrogen bonding term (6−8 potential).

The prepared protein structure was imported into Greater, which is the graphical user interface of GRID, and the H2O probe was used to obtain the MIF. 'MINIM', a utility program of the GRID package, was employed to determine the minima in the energy surface generated with the water probe. A minimum is defined as a grid point which is completely surrounded by points at which the GRID energy is higher (more positive). Thus, each minimum will be at the center of a block of 3 × 3 × 3 grid points. Subsequently, the 10 most favorable minima are populated with water molecules applying the 'FILMAP' program included in GRID. Then the protein structure plus the introduced water molecules were re-imported into Greater, followed by calculation of the MIF and analysis with MINIM and FILMAP.

This procedure was repeated until all additionally introduced waters had a solvent accessible surface area (SASA) of 0 and formed no direct interaction to protein atoms, yielding 110 water molecules in total. So the GRID force field energy is decisive for the order of waters placed in the structure, but there is no energy cutoff for placement. Then water molecules with a SASA > 0 were removed. In this way an optimal number of buried water molecules forming an intramolecular water network was obtained. The resulting water set of 31 molecules was included into the protein structure for the simulation setup.

Clustering of Experimental Water Molecules. From the PDB, 73 fXa structures were extracted. After alignment of the protein structures and removal of all water molecules with a calculated SASA > 0, a set of 1097 water positions was obtained.

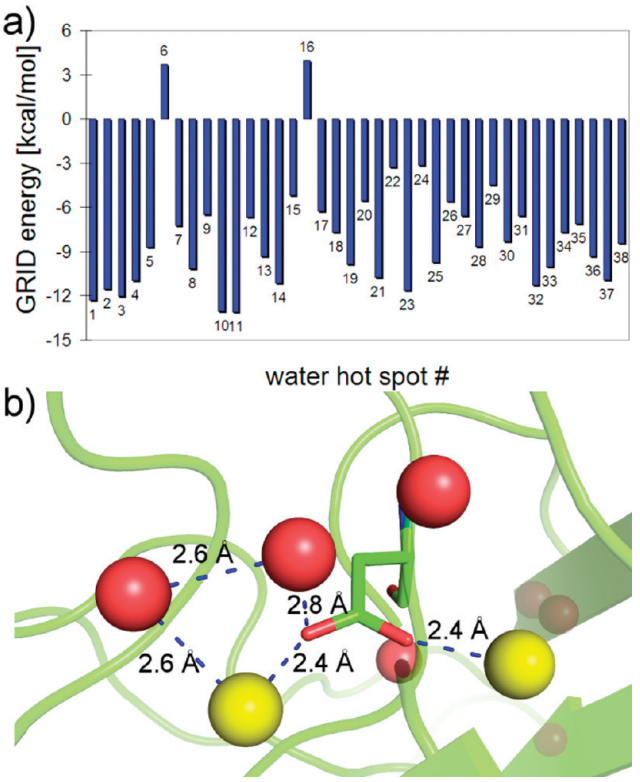


Figure 1. (a) GRID interaction energies (obtained with the water probe and the factor Xa structure 2boh) at the positions of the 38 water hot spots from the clustering of the PDB structures (top). With the exception of two water hot spots, the placement of water molecules at these hot spots is favorable. (b) Close-up view of the two unfavorable hot spots in 2boh which result from a short distance to Asp179 in the chosen X-ray structure (bottom, yellow spheres).

This set of water positions was subsequently clustered with an exclusion sphere algorithm,[63] resulting in 69 clusters. The cluster populations are depicted in Figure 1: The first 22 clusters are well populated, i.e., in more than 50% of the X-ray structures a water molecule is seen at the corresponding location, and 29 of the clusters are singletons, e.g., this water position is observed only in one single crystal structure. To further reduce the number of water sites, all singly occupied clusters were removed. For the remaining clusters, water positions were selected where optimal distances to hydrogen-bond partners are available. In two cases, the shape of the cluster suggested that it contained two water positions in hydrogen-bond distance. Consequently, two water molecules were placed accordingly. In the following, we will call the placed water molecules as "water hot spots". For more details, see Wallnoefer et al.[14]

MD Setup and Simulations. After addition of the GRID-derived water set, the subsequent preparation of the structure was performed in the LEAP tool of AMBER10.[91] The parameters for the Ca^{2+} ion were taken from Bradbrook et al.[92] Finally, a truncated octahedral box of TIP3P water molecules[93] with 10 Å distance from the protein to the outer wall of the box (about 6100 water molecules) was placed around the protein, resulting in a system of about 22 000 atoms. The pre-positioned water molecules were treated as TIP3P as well. Bulk water molecules overlapping with prepositioned ones were removed in the same

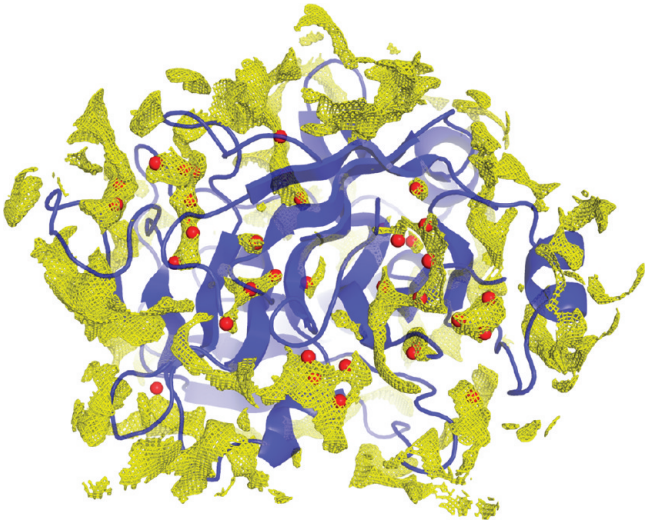


Figure 2. Experimental structure of factor Xa (PDB entry 2boh, blue) with the water positions obtained from the clustering approach (red). Also shown is the MIF calculated in GRID using the water probe and contoured at −6 kcal/mol (yellow).

way as if they would overlap with protein atoms. Vacuum voids were removed during equilibration steps in a NPT ensemble.

The SANDER module of AMBER10[91] with the ff99SB force field[94] was used. After an extensive equilibration as described before,[18] 12.5 ns of NPT MD simulation without any additional restraints was performed. The resulting trajectories were analyzed for convergence in the usual MD simulation parameters (energies, density, temperature, etc.). Further analyses were performed with the program PTRAJ of the AMBER toolkit.

RESULTS

Our basic hypothesis is that structurally important water molecules sit at energetically favorable positions and that the GRID program is capable of identifying these sites. Therefore, a necessary condition is that the 38 water hot spots obtained from clustering all crystallographic water positions coincide with regions of favorable GRID interaction energy obtained from a single X-ray structure.

Previous studies suggest GRID energy thresholds for water positions that should be taken into account at −6,[48] −8,[51] or −11 kcal/mol,[50] respectively. In Figure 1a we show the calculated GRID energies for the 38 water hot spots obtained in the clustering of the experimental data. Using a threshold of −11 kcal/mol, only 6 of the 38 X-ray waters would have been identified; a threshold of −8 kcal/mol would include 19 of the water positions, and a threshold of −6 kcal/mol would miss 10 water sites. Therefore, to minimize the risk of missing important water molecules, we chose a value of −6 kcal/mol as a threshold for favorable GRID energy for the comparison with the experimental data in Figure 2.

As seen in Figure 1a, the occupation of all but two water hot spots is favorable according to the GRID H2O probe at a contour obtained with −6 kcal/mol. These two exhibit a positive energy —meaning that placing a water molecule there would be unfavorable. This is caused by repulsion by the amino acid Asp179 (Figure 1b): For both water sites, Asp179 is the main interaction partner. As in many of the structures used for the clustering the local environment is slightly different, the resulting

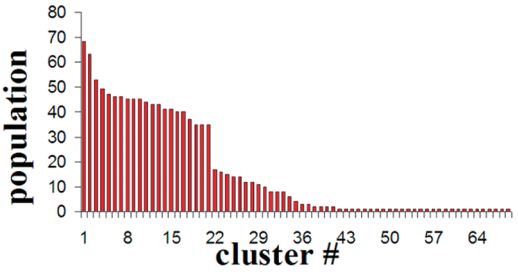


Figure 3. Population of the 69 clusters obtained from clustering water positions in all factor Xa heavy chain structures available from the PDB. Twenty-nine of the 69 clusters contain only one entry, meaning that this site is only occupied exactly once in the 73 protein structures. The largest cluster (cluster #1) represents a site which is occupied by a water molecule in 68 of the 73 protein structures.

water sites are too close to Asp179 when the clustered water molecules are placed into the X-ray structure 2boh. Presumably, this close contacts are artifacts due to the multiple alignment during water position generation or the cluster representative selection during the cluster algorithm. The other 8 water hot spots, which are outside the −6 kcal/mol contour, are just slightly less favorable than −6 kcal/mol. A minor increase in the cutoff would lead to an inclusion of them as well. Therefore, not only the exact interaction energy is decisive but also the structural environment. Figure 2 shows the water positions of the clustering as well as the GRID contour at −6 kcal/mol. It is noteworthy that, e.g., water channels or hydration clusters are very well reflected by the contour, although not each individual position is within the mesh. In total, this is a very satisfying result, as all but two clustered water hot spots are characterized by GRID as being energetically favorable, and most of these lead to a significant stabilization of the system. Moreover, the clustering algorithm seems to be reasonably well-behaved, in that only in two cases the addition of the water spots to the actual protein structure leads to (slight) repulsion with protein atoms—and this repulsion is relieved after a few steps of minimization forming a stable hydrogen-bond network.

The placement of a new explicit water network according to GRID with the iterative water placement described in the Methods Section leads to 31 water positions with a SASA of 0. These water positions have a high overlap with the 38 hot spots from the clustering (26 of the 38 positions are identically occupied applying the 2 Å criterion of Bui et al.;[86] with 25 below 1.5 Å and 17 with lower than 1 Å) . The main differences are that GRID does not place water molecules in the S2 and next to the S4 subpockets, which have SASAs of 0. Some water sites are less populated than in the clustering results (for clustering populations see Figure 3). Two extra water molecules are placed near the doubly populated site next to the S4 pocket, and one additional water is connecting the backbone of Cys43 and Ala47.

The rather similar water networks indicate that a comparable contribution of the hydrogen-bond network to structural stability can be expected. To this end, we tested whether placing the GRID water molecules into the starting structure leads to a comparable stabilization of MD simulations as shown for the clustered water set. The setup of the system was performed analogously to the clustered water simulations to ensure comparability.

In Figure 4, the 2D RMSD plots of the Cα atoms are shown. The stability of the MD simulations is rather similar to the simulations with the water on the hot spots of the clustering (Figure 4).

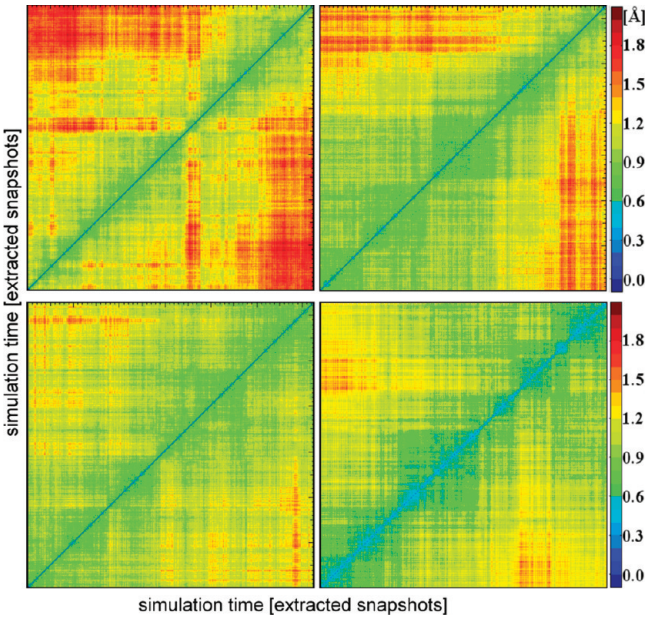


Figure 4. 2D RMSD plots of simulations of fXa containing four different water set-ups in the starting structure (2boh). On the upper left, no water was included in the system before soaking. On the upper right, all resolved water molecules from the X-ray structure were retained. On the lower left, the water set yielded from the clustering as described in Wallnoefer et al.[5] On the lower right, the simulation with the GRID-generated water positions is displayed.

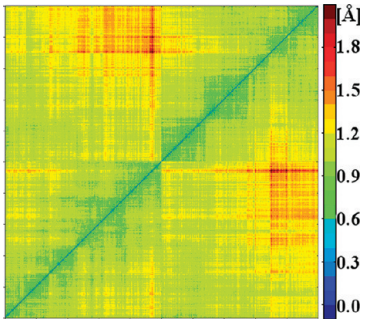


Figure 5. Combined 2D RMSD plot for the simulations with the GRID water set (upper right quadrant) and the clustered water (lower left quadrant). In the upper left and lower right quadrants, RMSD values below 2 Å point to a high conformational overlap.

Apart from a short conformational change at about 8 ns (Figure 4), all snapshots are within 1.3 Å from the starting structure.

Figure 5 presents a combined 2D RMSD plot for the simulations with the clustered water set and the GRID-generated water set. It is noteworthy that particularly the first half of the simulation with the GRID set is rather similar to the whole simulation with the clustered water set. The mutual C_α RMSDs are around 1 Å. After the distinct conformational transition after about 8 ns of simulation time, backbone RMSDs of 1.5 Å and higher are observed. Nevertheless, overall a sampling of similar conformational space is obtained.

The observed conformational transition is illustrated in Figure 6. A pronounced bend of the helix, which is connected to the Phe162 is responsible for the RMSD change. The helix is part

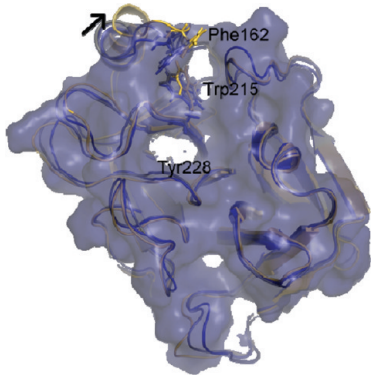


Figure 6. Conformational transition observed during the simulation with the GRID water set. The helix next to the Phe162 bends to the right. Blue: X-ray structure; gray: after 5 ns of simulation; and yellow: after 10 ns of simulation.

of the unoccupied binding channel of fXa. Therefore, this movement could be induced by the absence of the ligand or the substrate and corresponds to an opening event of the binding site. Presumably, this is a inherent function of unligated fXa, however, without further investigation no well-grounded conclusions can be drawn.

DISCUSSION AND CONCLUSION

Water molecules have started to gain increased interest in drug design, and their crucial contribution to stability and function of biomolecules is widely accepted. However, their correct and realistic treatment in structural models is far from trivial. Factor Xa is an example of a globular protein with a high percentage of buried and structurally important water molecules. This provided a perfect example for the validation of GRID as a 'water scanner' inside the protein. Subsequent to the previously presented clustering approach to generate a representative intraprotein water network for factor Xa on the basis of vast structural information available for this protein, we used the computer program GRID as an alternative for the analysis of the water structure in an experimental crystal structure. Having access to such a procedure is especially important if there is less data available than for factor Xa. A careful treatment of the structure of fXa was already shown to yield convincing results. Therefore, the important 'wet' regions inside of the protein body are already known and were used as validation data. In a first step, this existing data were compared to the GRID energy contour. It includes regions where the GRID force field assumes water molecules to interact energetically favorable with the surrounding. This analysis showed a remarkable agreement between the experimental results and the contours. Therefore, we concluded that GRID in principle is a viable approach to investigate the hydration of fXa.

However, to extend the picture from an abstract 'hydration density' to concrete water positions is far from trivial. This is highlighted by the failed approach to use the default water positioning of GRID to obtain a water network. Nevertheless, we improved the method by using an iterative procedure to place the water molecules in and around factor Xa. Thus, we yielded a water network that reproduced the experimental data surprisingly well; the same singly occupied voids were identified as with the clustering approach. Additionally identified water sites were found multiple times in occupied vaults, which have probably

several ways of being filled with water molecules and are therefore difficult to be detected with X-ray crystallography. Using this water network in a MD simulation resulted in a conformational ensemble, which is remarkably similar to the one obtained with the water network from the experimental data; both simulations cover a very similar conformational space and sample intermittently the same phase space except the above-described transition illustrated in Figure 6.

Clearly, water molecules play an important role in MD simulations, they strongly influence which conformations can be observed. Still, it is often treated rather simplistically in the simulation setup. As we have shown, a careful setup of the hydration either by merging experimental data or from analysis with GRID improves the convergence and the stability of the system significantly. Of course, not all unreasonable structures occurring during simulations are caused by a 'wrong' hydration; force field, equilibration, and convergence issues are important other reasons, and often it is difficult to pinpoint the exact reason why a MD trajectory has unexpectedly resulted in snapshots with high RMSD to the (experimental) starting structure. Nevertheless, the presented method provides a fast and reliable method to improve the initial hydration of the protein, thus minimizing potential problems arising from this source.

■ ASSOCIATED CONTENT

Supporting Information. PDB structure 2boh: One including the water set obtained by the clustering approach and one containing the water positions from the GRID approach. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: klaus.liedl@uibk.ac.at. Telephone: 0043 (0)512 507 5164.

■ REFERENCES

(1) Wallnoefer, H. G.; Fox, T.; Liedl, K. R. Challenges for Computer Simulations in Drug Design. In *Challenges and Advances in Computational Chemistry and Physics*; Kinetics and Dynamics; Springer: New York, 2010; Vol. 12, pp 431−463.

(2) Jacobs, D. J. Ensemble-based methods for describing protein dynamics. *Curr. Opin. Pharmacol.* **2010**, *10*, 760–769.

(3) McGeagh, J. D.; Ranaghan, K. E.; Mulholland, A. J. Protein dynamics and enzyme catalysis: Insights from simulations. *Biochim. Biophys. Acta, Proteins Proteomics* **2011**, *1814*, 1077–1092.

(4) Gallicchio, E.; Levy, R. M. Advances in all atom sampling methods for modeling protein-ligand binding affinities. *Curr. Opin. Struct. Biol.* **2011**, *21*, 161–166.

(5) Lin, J.-H. Accommodating protein flexibility for structure-based drug design. *Curr. Top. Med. Chem.* **2011**, *11*, 171–178.

(6) Grienke, U.; Schmidtke, M.; Kirchmair, J.; Pfarr, K.; Wutzler, P.; Dürrwald, R.; Wolber, G.; Liedl, K. R.; Stuppner, H.; Rollinger, J. M. Antiviral Potential and Molecular Insight into Neuraminidase Inhibiting Diarylheptanoids from *Alpinia katsumadai*. *J. Med. Chem.* **2010**, *53*, 778–786.

(7) Wallnoefer, H. G.; Lingott, T.; Gutiérrez, J. M.; Merfort, I.; Liedl, K. R. Backbone Flexibility Controls the Activity and Specificity of a Protein−Protein Interface: Specificity in Snake Venom Metalloproteases. *J. Am. Chem. Soc.* **2010**, *132*, 10330–10337.

(8) Reboul, C. F.; Andrews, D. A.; Nahar, M. F.; Buckle, A. M.; Roujeinikova, A. Crystallographic and molecular dynamics analysis of loop motions unmasking the peptidoglycan-binding site in stator protein MotB of flagellar motor. *PLoS ONE* **2011**, *6*, e18981.

(9) Schmidt, R. K.; Gready, J. E. Molecular Dynamics Simulations of L-Lactate Dehydrogenase: Conformation of a Mobile Loop and Influence of the Tetrameric Protein Environment. *J. Mol. Model.* **1999**, *5*, 153–168.

(10) Kamerlin, S. C. L.; Rucker, R.; Boresch, S. A targeted molecular dynamics study of WPD loop movement in PTP1B. *Biochem. Biophys. Res. Commun.* **2006**, *345*, 1161–1166.

(11) Hu, C.; Fang, J.; Borchardt, R. T.; Schowen, R. L.; Kuczera, K. Molecular dynamics simulations of domain motions of substrate-free S-adenosyl- L-homocysteine hydrolase in solution. *Proteins* **2008**, *71*, 131–143.

(12) Kleinekathöfer, U.; Israelewitz, B.; Dittrich, M.; Schulten, K. Domain Motion of Individual F1-ATPase β-Subunits during Unbiased Molecular Dynamics Simulations. *J. Phys. Chem. A* **2011**, *115*, 7267–7274.

(13) Wallnoefer, H. G.; Fox, T.; Liedl, K. R.; Tautermann, C. S. Dispersion dominated halogen−π interactions: energies and locations of minima. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14941.

(14) Wallnoefer, H. G.; Handschuh, S.; Liedl, K. R.; Fox, T. Stabilizing of a Globular Protein by a Highly Complex Water Network: A Molecular Dynamics Simulation Study on Factor Xa. *J. Phys. Chem. B* **2010**, *114*, 7405–7412.

(15) Pikkemaat, M. G.; Linssen, A. B. M.; Berendsen, H. J. C.; Janssen, D. B. Molecular dynamics simulations as a tool for improving protein stability. *Protein Eng.* **2002**, *15*, 185–192.

(16) Palma, R.; Curmi, P. M. Computational studies on mutant protein stability: The correlation between surface thermal expansion and protein stability. *Protein Sci.* **1999**, *8*, 913–920.

(17) Winger, M.; Van Gunsteren, W. F. Use of Molecular-Dynamics Simulation for Optimizing Protein Stability: Consensus-Designed Ankyrin Repeat Proteins. *Helv. Chim. Acta* **2008**, *91*, 1605–1613.

(18) Wallnoefer, H. G.; Liedl, K. R.; Fox, T. A challenging system: Free energy prediction for factor Xa. *J. Comput. Chem.* **2011**, *32*, 1743–1752.

(19) Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

(20) Lee, M. S.; Olson, M. A. Calculation of Absolute Protein-Ligand Binding Affinity Using Path and Endpoint Approaches. *Biophys. J.* **2006**, *90*, 864–877.

(21) Huang, N.; Jacobson, M. P. Physics-based methods for studying protein-ligand interactions. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 325–331.

(22) Gilson, M. K.; Zhou, H. X. Calculation of Protein-Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21.

(23) Foloppe, N.; Hubbard, R. E. Towards predictive ligand design with free-energy based computational methods. *Curr. Med. Chem.* **2006**, *13*, 3583–3608.

(24) Sreenivasan, U.; Axelsen, P. H. Buried water in homologous serine proteases. *Biochemistry* **1992**, *31*, 12785–12791.

(25) Guvench, O.; Price, D. J.; Brooks, C. L., III Receptor rigidity and ligand mobility in trypsin-ligand complexes. *Proteins* **2005**, *58*, 407–417.

(26) Lesk, A. M.; Fordham, W. D. Conservation and Variability in the Structures of Serine Proteinases of the Chymotrypsin Family. *J. Mol. Biol.* **1996**, *258*, 501–537.

(27) 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*, 2054–2064.

(28) Zhang, X. J.; Matthews, B. W. Conservation of solvent-binding sites in 10 crystal forms of T4 lysozyme. *Protein Sci.* **1994**, *3*, 1031–1039.

(29) Reichmann, D.; Phillip, Y.; Carmi, A.; Schreiber, G. On the Contribution of Water-Mediated Interactions to Protein-Complex Stability. *Biochemistry* **2008**, *47*, 1051–1060.

(30) Nilsson, M.; Hämäläinen, M.; Ivarsson, M.; Gottfries, J.; Xue, Y.; Hansson, S.; Isaksson, R.; Fex, T. Compounds binding to the S2-S3 pockets of thrombin. *J. Med. Chem.* **2009**, *52*, 2708–2715.

(31) Edsall, J. T.; McKenzie, H. A. Water and proteins. II. The location and dynamics of water in protein systems and its relation to their stability and properties. *Adv. Biophys.* **1983**, *16*, 53–183.
(32) Cruz, A.; Ramirez, E.; Santana, A.; Barletta, G.; López, G. E. Molecular dynamic study of subtilisin Carlsberg in aqueous and non-aqueous solvents. *Mol. Simul.* **2009**, *35*, 205–212.
(33) Maxwell, M. K.; Enrico, D. C. Conserved water molecules in the specificity pocket of serine proteases and the molecular mechanism of Na+ binding. *Proteins* **1998**, *30*, 34–42.
(34) Meyer, E.; Cole, G.; Radhakrishman, R.; Epp, O. Structure of native porcine pancreatic elastase at 1.65 angstrom resolution. *Acta Cryst., Sect. B* **1988**, *44*, 26–38.
(35) Silva, J.; Antunes, O. A. C.; de Alencastro, R. B.; De Simone, S. G. The Na+ binding channel of human coagulation proteases: Novel insights on the structure and allosteric modulation revealed by molecular surface analysis. *Biophys. Chem.* **2006**, *119*, 282–294.
(36) Caldwell, R. A.; Boucher, R. C.; Stutts, M. J. Serine protease activation of near-silent epithelial Na+ channels. *Am. J. Physiol. Cell. Physiol.* **2004**, *286*, C190–C194.
(37) Carugo, O.; Bordo, D. How many water molecules can be detected by protein crystallography?. *Acta Cryst., Sect. D* **1999**, *55*, 479–483.
(38) Rao, M. S.; Olson, A. J. Modelling of Factor Xa-inhibitor complexes: A computational flexible docking approach. *Proteins* **1999**, *34*, 173–183.
(39) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling Water Molecules in Protein-Ligand Docking Using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.
(40) Doan, B. T.; Fraternali, F.; Do, Q. T.; Atkinson, R. A.; Palmas, P.; Sklenar, V.; Wildgoose, P.; Strop, P.; Saudek, V. A NMR and MD study of the active site of factor Xa by selective inhibitors. *J. Chim. Phys. Phys.-Chim. Biol.* **1998**, *95*, 443–446.
(41) Fraternali, F.; Do, Q. T.; Doan, B. T.; Atkinson, R. A.; Palmas, P.; Sklenar, V.; Safar, P.; Wildgoose, P.; Strop, P.; Saudek, V. Mapping the active site of factor Xa by selective inhibitors: An NMR and MD study. *Proteins* **1998**, *30*, 264–274.
(42) Shokhen, M.; Khazanov, N.; Albeck, A. Screening of the active site from water by the incoming ligand triggers catalysis and inhibition in serine proteases. *Proteins* **2008**, *70*, 1578–1587.
(43) Borensztajn, K.; Peppelenbosch, M. P.; Spek, C. A. Factor Xa: at the crossroads between coagulation and signaling in physiology and disease. *Trends Mol. Med.* **2008**, *14*, 429–440.
(44) Bounameaux, H. The novel anticoagulants: Entering a new era. *Swiss Med. Wkly.* **2009**, *139*, 60–64.
(45) Ieko, M.; Tarumi, T.; Nakabayashi, T.; Yoshida, M.; Naito, S.; Koike, T. Factor Xa inhibitors: New anti-thrombotic agents and their characteristics. *Front. Biosci.* **2006**, *11*, 232–248.
(46) McVey, J. H. Tissue factor pathway. *Bailliere's Clin. Haem.* **1994**, *7*, 469–484.
(47) Broze, J. Tissue factor pathway inhibitor and the current concept of blood coagulation. *Blood Coagulation Fibrinolysis* **1995**, *6*, S1.
(48) Davie, E. W.; Fujikawa, K.; Kisiel, W. The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* **1991**, *30*, 10363–10370.
(49) Hauptmann, J.; Markwardt, F.; Walsmann, P. Synthetic inhibitors of serine proteinases. XIV. Influence of 3- and 4-amidinobenzyl derivatives on the formation and action of thrombin. *Thromb. Res.* **1978**, *12*, 735–744.
(50) Hertzberg, M. Biochemistry of factor X. *Blood Rev.* **1994**, *8*, 56–62.
(51) Stubbs, M. T.; Bode, W. Coagulation factors and their inhibitors. *Curr. Opin. Struct. Biol.* **1994**, *4*, 823–832.
(52) Tidwell, R. R.; Webster, W. P.; Shaver, S. R.; Geratz, J. D. Strategies for anticoagulation with synthetic protease inhibitors. Xa inhibitors versus thrombin inhibitors. *Thromb. Res.* **1980**, *19*, 339–349.
(53) Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J. K. How do serine proteases really work? *Biochemistry* **2011**, *28*, 3629–3637.
(54) Hedstrom, L. How do serine proteases really work? *Chem. Rev.* **2002**, *102*, 4501–4524.
(55) Kaiser, B. DX-9065a, a direct inhibitor of factor Xa. *Cardiovasc. Drug Rev.* **2003**, *21*, 91–104.
(56) Nazaré, M.; Will, D. W.; Matter, H.; Schreuder, H.; Ritter, K.; Urmann, M.; Essrich, M.; Bauer, A.; Wagner, M.; Czech, J.; Lorenz, M.; Laux, V.; Wehner, V. Probing the Subpockets of Factor Xa Reveals Two Binding Modes for Inhibitors Based on a 2-Carboxyindole Scaffold: A Study Combining Structure-Activity Relationship and X-ray Crystallography. *J. Med. Chem.* **2005**, *48*, 4511–4525.
(57) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.
(58) Daura, X.; Haaksma, E.; van Gunsteren, W. F. Factor Xa: simulation studies with an eye to inhibitor design. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 507–529.
(59) Ostrovsky, D.; Udier-Blagovic, M.; Jorgensen, W. L. Analyses of Activity for Factor Xa Inhibitors Based on Monte Carlo Simulations. *J. Med. Chem.* **2003**, *46*, 5691–5699.
(60) Singh, N.; Briggs, J. M. Molecular Dynamics Simulations of Factor Xa: Insights into Conformational Transition of Its Binding Subsites. *Biopolymers* **2008**, *89*, 1104–1113.
(61) Venkateswarlu, D.; Perera, L.; Darden, T.; Pedersen, L. G. Structure and dynamics of zymogen human blood coagulation factor X. *Biophys. J.* **2002**, *82*, 1190–1206.
(62) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(63) Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 357–369.
(64) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(65) Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field. *J. Med. Chem.* **2004**, *47*, 5114–5125.
(66) Helms, V.; Wade, R. C. Thermodynamics of a Water Mediating Protein-Ligand Interactions in Cytochrome P450cam: A Molecular Dynamics Study. *Biophys. J.* **1995**, *69*, 810–824.
(67) Wade, R. C. Solvation of the active site of cytochrome P450-cam. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 199–204.
(68) Henchman, R. H.; McCammon, J. A. Structural and Dynamic Properties of Water around Acetylcholinesterase. *Protein Sci.* **2002**, *11*, 2080–2090.
(69) Olkhova, E.; Hutter, M. C.; Lill, M. A.; Helms, V.; Michel, H. Dynamic Water Networks in Cytochrome c Oxidase from Paracoccus denitrificans Investigated by Molecular Dynamics Simulations. *Biophys. J.* **2004**, *86*, 1873–1889.
(70) Schymkowitz, J. W. H.; Rousseau, F.; Martins, I. C.; Ferkinghoff-Borg, J.; Stricher, F.; Serrano, L. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Soc. U.S.A.* **2005**, *102*, 10147–10152.
(71) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX web server: an online force field. *Nucleic Acids Res.* **2005**, *33*, W382–388.
(72) Verdonk, M. L.; Cole, J. C.; Watson, P.; Gillet, V.; Willett, P. SuperStar: Improved knowledge-based interaction fields for protein binding sites. *J. Mol. Biol.* **2001**, *307*, 841–859.
(73) Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: A Knowledge-based Approach for Identifying Interaction Sites in Proteins. *J. Mol. Biol.* **1999**, *289*, 1093–1108.
(74) Henchman, R. H.; Tai, K. S.; Shen, T. Y.; McCammon, J. A. Properties of water molecules in the active site gorge of acetylcholinesterase from computer simulation. *Biophys. J.* **2002**, *82*, 2671–2682.

(75) Lounnas, V.; Pettitt, B. M. A connected cluster of hydration around Myoglobin-correlation between molecular dynamics simulations and experiment. *Proteins* **1994**, *18*, 133–147.
(76) Henchman, R. H.; McCammon, J. A. Extracting hydration sites around proteins from explicit water simulations. *J. Comput. Chem.* **2002**, *23*, 861–869.
(77) Li, Z.; Lazaridis, T. Thermodynamics of buried water clusters at a protein-ligand binding interface. *J. Phys. Chem. B* **2006**, *110*, 1464–1475.
(78) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.
(79) Li, Z.; Lazaridis, T. Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease. *J. Am. Chem. Soc.* **2003**, *125*, 6636–6637.
(80) Li, Z.; Lazaridis, T. The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A. *J. Phys. Chem. B* **2004**, *109*, 662–670.
(81) Guimarães, C. R. W.; Mathiowetz, A. M. Addressing Limitations with the MM-GB/SA Scoring Procedure using the WaterMap Method and Free Energy Perturbation Calculations. *J. Chem. Inf. Model.* **2011**, *50*, 547–559.
(82) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of Explicit Solvent Effects to the Binding Affinity of Small-Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *ChemMedChem* **2011**, *6*, 1049–1066.
(83) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the Water Content in Protein Binding Sites. *J. Phys. Chem. B* **2009**, *113*, 13337–13346.
(84) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
(85) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Effects of Water Placement on Predictions of Binding Affinities for p38α MAP Kinase Inhibitors. *J. Chem. Theor. Comput.* **2011**, *6*, 3850–3856.
(86) Bui, H. H.; Schiewe, A. J.; Haworth, I. S. WATGEN: An algorithm for modeling water networks at protein-protein interfaces. *J. Comput. Chem.* **2007**, *28*, 2241–2251.
(87) Li, Y.; Sutch, B. T.; Bui, H.-H.; Gallaher, T. K.; Haworth, I. S. Modeling of the Water Network at Protein−RNA Interfaces. *J. Chem. Inf. Model.* **2011**, *51*, 1347–1352.
(88) Underwood, M. C.; Zhong, D.; Mathur, A.; Heyduk, T.; Bajaj, S. P. Thermodynamic Linkage between the S1 Site, the Na+ Site, and the Ca2+ Site in the Protease Domain of Human Coagulation Factor Xa. *J. Biol. Chem.* **2000**, *275*, 36876–36884.
(89) Griffon, N.; Di Stasio, E. Thermodynamics of Na+ binding to coagulation serine proteases. *Biophys. Chem.* **2001**, *90*, 89–96.
(90) Labute, P. Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009**, *75*, 187–205.
(91) Case, D. A.; Darden, T. E.; Cheatham, T. E.; 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.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Mongan, J.; Hornak, G.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, CA, 2008.
(92) Bradbook, G. M.; Gleichmann, T.; Harrop, S. J.; Habash, J.; Raftery, J.; Kalb, J.; Yariv, J.; Hillier, I. H.; Helliwell, J. R. X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes relating structure to thermodynamics of binding. *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 1603–1611.
(93) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
(94) Hornak, V.; Abel, 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.