Enhancing Sampling of Water Rearrangements on Ligand Binding: A Comparison of Techniques

Yunhui Ge,[†] David C. Wych,[†,‡] Marley L. Samways,[¶] Michael E. Wall,[‡] Jonathan W. Essex,[¶] and David L. Mobley[*,†,§]

†*Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA*

‡ *Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

¶*School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom*

§*Department of Chemistry, University of California, Irvine, CA 92697, USA*

E-mail: dmobley@mobleylab.org

1 ABSTRACT

Water often plays a key role in protein structure, molecular recognition, and mediating protein-ligand interactions. Thus, free energy calculations must adequately sample water motions, which often proves challenging in typical MD simulation timescales. Thus, the accuracy of methods relying on MD simulations ends up limited by slow water sampling. Particularly, as a ligand is removed or modified, bulk water may not have time to fill or rearrange in the binding site. In this work, we focus on several molecular dynamics (MD) simulation-based methods attempting to help address water motions and occupancies: BLUES, using nonequilibrium candidate Monte Carlo (NCMC); *grand*, using grand canonical Monte Carlo (GCMC); and normal MD. We assess the accuracy and efficiency of these methods in sam-

pling water motions. We selected a range of systems with varying numbers of waters in the binding site, as well as those where water occupancy is coupled to the identity or binding mode of the ligand. We analyzed water motions and occupancies using both clustering of trajectories and direct analysis of electron density maps. Our results suggest both BLUES and *grand* enhance water sampling relative to normal MD and *grand* is more robust than BLUES, but also that water sampling remains a major challenge for all of the methods tested. The lessons we learned for these methods and systems are discussed.

2 INTRODUCTION

In their natural environment, proteins are surrounded by waters which critically affect their structure, function and dynamics.[1,2] Buried water molecules in the binding sites[3–5] also play important roles such as facilitating receptor-ligand recognition and stabilizing proteins.[2,6–9] A previous study done on 392 high-resolution protein-ligand crystal structures observed at least one water molecule bridging the protein and ligand in 85% of the systems.[10]

While typical MD simulations can be used to model interactions between proteins and water molecules, these often fail to adequately sample water exchange between bulk and buried hydration sites since water rearrangements in binding sites can often be extremely slow.[11,12] This poses significant challenges to binding free energy calculations[13–15] especially in relative binding free energy (RBFE) calculations which show promise in guiding experimental work in the lead optimization stage in real drug discovery projects.[13,16]

In a typical RBFE calculation, two structurally similar ligands are compared by simulating them in both the protein-ligand complex and in the solution state where one ligand is transformed into another via unphysical or alchemical pathway. However, even closely related ligands have differences in water placement in the binding site.[17–20] In RBFE calculations, when simulating the protein-ligand complex, the simulation timescale is normally too short (e.g., ns) to allow adequate sampling of water rearrangements when transforming

one ligand to another which impairs the accuracy of such calculations.

A variety of methods seek to advance the knowledge of optimal placement of water molecules and facilitate binding free energy calculations.[21–32] Among these methods, we are especially interested in two: nonequilibrium candidate Monte Carlo (NCMC)[33] which efficiently hops water molecules between energy basins, and grand canonical Monte Carlo (GCMC)[34–37] which allows the fluctuations in the number of water molecules in a simulation according to a specified chemical potential. Both methods show promise in improving water sampling in molecular simulations and GCMC has shown the ability to incorporate the thermodynamics of buried water in binding free energy calculations.[31,32,38–40] Ben-Shalom et al.[30] recently studied a Monte Carlo (MC)/MD hybrid approach and found robust sampling of buried hydration sites and improved accuracy in relative binding free energy calculations. However, this approach was implemented in a different simulation engine (AMBER package[41]) than the one used in this work (OpenMM[42]) for MD, NCMC and GCMC. So we didn't include this approach in this work (check Section 3 for more details).

In this work, we seek to compare the efficiency and accuracy between NCMC and GCMC methods in water sampling using a broad range of systems. We also compare with plain MD simulations as a point of reference. The results from a comprehensive comparison between these techniques provide valuable lessons regarding water sampling issues in MD simulations and may be helpful for applications such as binding free energy calculations.

3 METHODS

3.1 Force field limitations.

Before we move to simulation details, we address force field limitations, a key concern in any simulation study. Molecular simulations are conducted with an underlying energy model, or force field, which approximates the underlying physics – ideally reasonably well. Even though force fields for proteins, small molecules and solvents have been developed for several decades,

a perfect force field would still be an approximation, and present-day force fields still seem not to have reached the limitations of the functional form and thus are not perfect. Thus, even if all other aspects of simulations are correct (timescale, preparation, etc.) predictions from simulations still may differ from experimental measurements. In addition, other factors like the temperature at which the diffraction data was collected in experiments may also contribute to the discrepancies between simulations and experimental measurements.

In this work, when examining the efficiency of the different computational methods examined, we do not address the issue of any potential force field limitations. In general, a better sampling method ought to more efficiently yield results closer to the correct value given the chosen force field, but it won't address force field problems (i.e., the force field does not well represent the true system or the conditions in which experimental measurements were conducted). In principle, it is possible that a better sampling method might yield worse agreement with experiment, if the correct answer for the force field differs from reality. Thus, when comparing methods, a successful simulation is one which captures the true force field answer for the system. Ideally that would also agree with experiment. But if it doesn't, and we indeed have captured the correct force field answer for the system (which may be assessed by agreement among all of the methods examined, or with a gold standard approach, for example), we still consider such a simulation as success.

3.2 Selected targets.

We selected the targets from two recent studies focusing on using enhanced sampling of water motions to improve the accuracy of binding free energy calculations,[30,31] including several proteins: Protein Tyrosine Phosphatase 1B (PTP1B), Heat Shock Protein 90 (HSP90), Bruton's Tyrosine Kinase (BTK), transcription initiation factor TFIID subunit 2 (TAF1(2)), bromodomain-containing protein 4 (BRD4) and thrombin. In addition to being different receptors, these targets differ in binding site positions, number and occupancy of buried water sites. We aim to include enough diversity and cover a broad range of systems which

were also previously studied so that we may validate our results against prior work. Several targets studied in this work also differ in the occupancy of water sites between congeneric ligands which may pose challenges in relative binding free energy calculations. Figure 1 shows the binding sites of these systems, along with crystallographic water molecules and the relevant Protein Data Bank (PDB) IDs.

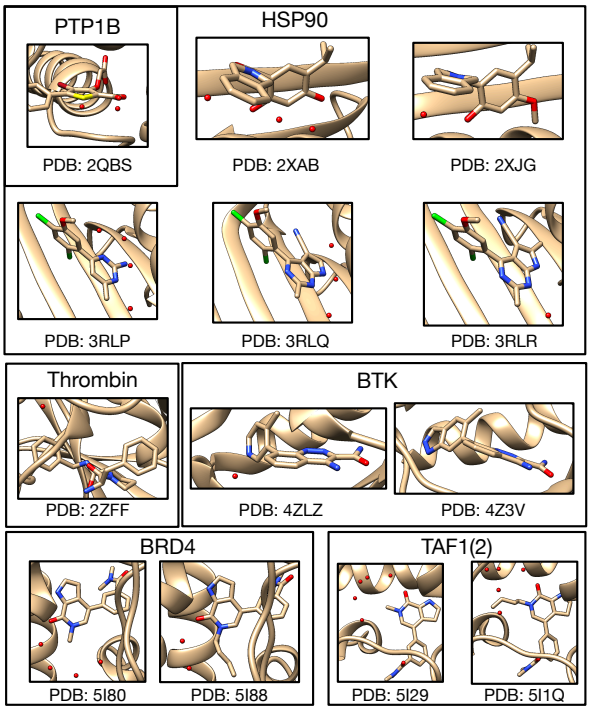


Figure 1: All 13 protein-ligand systems studied in this work and their hydration sites (red spheres) and their PDB IDs.

3.3 Molecular Dynamics Simulations.

The ligand was parameterized using Open Force Field version 1.2.1 (codenamed "Parsley").[43,44] The AMBER ff14SB force field[45] was used for protein parameterization in conjunction with TIP3P water model.[46] BLUES, *grand* and normal MD simulations were performed using OpenMM (version 7.4.2).[42] A time step of 2 fs, and a friction constant of 1 ps^{-1} were used in MD simulations. Long-range electrostatics were calculated using Particle Mesh Ewald (PME)[47,48] with nonbonded cutoffs of 10 Å. Each system was simulated at the

experimental temperature listed on the PDB website (`https://www.rcsb.org`). We used pdbfixer 1.6 (`https://github.com/openmm/pdbfixer`) to add the missing heavy atoms to the receptor. Then, the PROPKA algorithm[49,50] on PDB2PQR web server[51] was used to protonate the receptors residues at experimental pH values. The pK_a values of ligands were calculated using Chemicalize (ChemAxon, `https://www.chemaxon.com`) and then were used to determine protonation states of ligands based on the simulation pH conditions.

For each target, we performed two separate MD simulations with different starting velocities, one set (1) with ordered water molecules removed prior to simulation and another, (2) with ordered water molecules retained. Ideally, the two versions of simulations will converge to similar results (e.g., suggesting similar occupancies of target sites).

The systems were first minimized until forces were below a tolerance of 10 kJ/mol using the L-BFGS optimization algorithm[52] implemented in OpenMM, followed by 1 ns NVT equilibration and 10 ns NPT equilibration. The force evaluations for the two equilibration phases are 0.5M and 5M. The production run was performed in the NPT ensemble for 70 ns of a single simulation block (equivalent to 35M force evaluations) in consideration of our cluster's actual wallclock time limit and was extended 9 times to 700 ns in total. Each individual 70 ns unit in this 700ns production run constitutes a simulation block for the purposes of the analysis we present here.

Our previous work showed that restraining the protein and ligand to maintain the crystallographic pose was helpful in water insertion to the target sites in BLUES simulations since this helps keep protein cavities from collapsing; when they collapse, it can be difficult for simulations to re-fill them. It is also interesting to test this idea in plain MD simulations although we believe the success of the approach could be system dependent. To do so, position restraints of 10 kcal/mol $Å^{-2}$ were applied on the heavy atoms of the protein and all atoms of the ligand to maintain the crystallographic pose. The same simulation protocol was used as that for unbiased MD, except that we applied these restraints in both minimization/equilibration and production runs. Two separate production runs were performed for

100 ns both in the presence and the absence of crystallographic water molecules.

We performed restrained MD simulations both to explore the benefits of using restraints in water sampling and used these as a cross-validation of BLUES simulations where the same restraints were applied. As noted in Section 3.1, it is not guaranteed that the force field used in this work will produce structural results which agree with experimental structures. The restrained MD simulations can help to check if the buried hydration sites shown in the deposited crystal structures of the studied target systems are also favorable with the force field used here. We consider those hydration sites favorable in our simulations when they were always occupied in simulations or had an average occupancy of more than 70% with many transitions in simulations. This may or may not correspond to the hydration site being favorable in a thermodynamic sense, depending on whether sampling is adequate. Since both restrained MD and BLUES simulation restrained the protein and ligand to the crystallographic pose, we expect the same favorable hydration sites in both simulations. When discrepancies are observed in the results from both simulations compared to the crystal structures, it is possible that such disagreements are due to force field limitations.

3.4 BLUES Simulations.

BLUES combines nonequilibrium candidate Monte Carlo (NCMC)[53] with classical MD simulations to enhance the sampling of important degrees of freedom in ligand binding.[32,54–57] One advantage of using NCMC moves in water sampling is they can efficiently hop water molecules between energy basins and the likelihood of these moves is independent of the barrier heights which is normally a challenge in conventional MD simulations. The details of theory and implementation of BLUES in water sampling can be found in prior work.[32]

In BLUES, we defined a spherical region within which the water hops occur, using a heavy atom on the ligand which is close to the center of the ligand (selected visually) as the center of this sphere (see `https://github.com/MobleyLab/water_benchmark_paper` for a detailed list of selected atoms). This region should ideally cover the target water sites and

extend out to bulk water to allow bulk water to exchange with water in the binding site. Since the size of the region is a parameter that may affect the success rate of NCMC moves to rehydrate the target water sites, we tested several radii for the sphere, typically using 0.8, 1.0 and 1.5 nm for most target systems. For several systems, we only used 1.0 and 1.5 nm to cover all target hydration sites.

A BLUES simulation consists of a number of BLUES iterations, where each iteration of BLUES is composed of an NCMC moves and conventional MD. In each NCMC move, interactions between the selected water molecule and its environment are gradually turned off, then the water molecule is randomly proposed to be moved to a new position in the predefined region before its interactions are turned back on. This approach allows the environment to relax in response to the proposed water translation, improving acceptance of moves and thereby accelerating water exchange and sampling. Here, we used the same number of NCMC steps (5000 steps) and MD steps (1000 steps) for all of the systems. For a single simulation block, in consideration of our cluster's actual wallclock time limit, 3000 BLUES iterations were performed using hydrogen mass repartitioning scheme with 4 fs timesteps,[58] resulting in 12 ns simulation time and 18M force evaluations. In the analysis present here, we performed 10 simulation blocks in total (120 ns, 180M force evaluations). The same restraints on the protein and ligand were applied in BLUES simulations as used in the restrained MD simulations described earlier. Simulations were done both in the presence and absence of crystallographic water molecules.

3.5 *grand* Simulations.

Grand canonical Monte Carlo (GCMC)[34–37] shows particular promise for enhancing water sampling and facilitating binding free energy calculations.[31,38–40,59–62] In the grand canonical ensemble the chemical potential (μ) of the fluctuating species (here, water molecules), the volume and the temperature is constant. The water molecules can be inserted (transferred from) or removed (transferred to) from the system to enhance water sampling — judicious

choice of the chemical potential gives an equilibrium between the simulated system and bulk water.

In this work, the *grand* package[63] was used to perform GCMC moves with MD sampling using OpenMM simulation engine.[42] We used OpenMM so that all of simulation techniques (MD, NCMC, GCMC) studied in this work used the same engine (OpenMM) which provides an opportunity to conduct a relatively fair comparison between these techniques, avoiding scenarios where implementation differences in different engines might bias the results. This is also one reason that the Monte Carlo (MC)/MD hybrid approach recently presented by Ben-Shalom et al.[30] was not examined in this work since the hybrid approach used there was implemented in the AMBER simulation package[41] and is impractical here (because, with OpenMM, acceptance/rejection of MC moves must occur off-GPU, and OpenMM performance on CPUs is extremely poor).

In *grand*, as in BLUES, a GCMC region needs to be defined first. To do that, we selected two atoms (e.g., Cα) on the receptor so that the middle point between them is used as the center of a spherical GCMC region for enhanced water sampling (see `https://github.com/MobleyLab/water_benchmark_paper` for a detailed list of selected atoms). All target hydration sites are within this defined spherical region. The radius varies between systems and is dependent on the binding site size. Then the equilibration process was executed in three stages. The first GCMC/MD stage was to equilibrate the water distribution and involved an initial 10000 GCMC moves, followed by 1 ps of GCMC/MD (100 iterations, where each iteration includes 5 MD steps of 2 fs each, followed by 1000 GCMC moves). The second 500 ps NPT simulation was to equilibrate the system volume. The final GCMC/MD stage was to equilibrate the waters at the new system volume and involves 100k GCMC moves over 500 ps. The force evaluations for the three equilibration phases are: 0.1M,0.25M,0.35M. The production simulation involved 2.5 ns of GCMC/MD (50 GCMC moves carried out every 1 ps of MD) for each single simulation block (1.4M force evaluations) in consideration of our cluster's wallclock time limit and was extended to 12.5 ns (5 blocks, 7M force evaluations)

in total. Unlike conventional MD and BLUES simulations, enhanced sampling (GCMC) of water molecules was carried out even in the equilibration phase and we find that this type of equilibration outperforms that done for MD and BLUES simulation in several systems (more details later).

There are additional two key parameters used in *grand* simulations: the excess chemical potential (μ') of bulk water and the standard state volume of water (V^o). Both parameters affect the acceptance probabilities of GCMC moves as depicted in previous work.[63] For internal consistency in the GCMC/MD simulations, prior work suggested it was more appropriate to calculate the values of the excess chemical potential and standard state volume of water from simulations, rather than using the experimental values.[63] The former is calculated as the hydration free energy of water, and the latter as the average volume per water molecule. The details of these calculations can be found in prior work[63] and the calculated results at different temperatures used in this work can be found in Table S1.

In *grand* simulations, we only simulated the systems in the absence of crystallographic water molecules. Two separate runs were performed for each system. Based on our results, *grand* simulations were able to rehydrate all target water sites (check Section 3.7 for how we defined a success case) within five simulation blocks (12.5 ns, 7M force evaluations) in most simulations (exceptions will be discussed later). If the results show different occupancies in water sites or the protein/ligand blocked the successful insertion of water molecules by GCMC moves, the same restraints on the protein/ligand as used in BLUES and restrained MD simulations were applied in *grand* simulations to try and help the results converge faster (e.g., 12.5 ns).

3.6 Trajectory Analysis

The simulated trajectories were analyzed using different approaches: (1) clustering-based analysis and (2) electron density calculations. For both approaches, MDTraj 1.9.4[64] was used to align trajectories to the crystal structure.

3.6.1 Clustering-based Analysis.

The water sites present within the predefined GCMC region (described in Section 3.5) were subjected to a clustering analysis, using average-linkage hierarchical clustering as implemented in SciPy, with a distance cutoff of 2.4 Å. This clustering essentially groups waters from different simulation frames which are considered to be the same site. For each cluster, the occupancy is calculated as a percentage, based on the number of frames in which that site is occupied by a water molecule relative to the total number of simulation frames. Note that, prior to clustering, a distance matrix of all water observations from the simulation was built. The distances between waters from the same simulation frame were set to an arbitrarily high value ($\sim 10^8$ Å) in order to discourage the merging of distinct water sites. This helps to make sure that distinct water sites did not get clustered together. Otherwise, the sites might get merged and thus return occupancies greater than 100%. All of these operations were done using build-in functions in *grand* (v1.0.0/v1.0.1) package. An example script is available on `https://github.com/MobleyLab/water_benchmark_paper`. After we obtained these populated hydration sites in simulations, we performed a visual examination of these sites and compared them to the crystallographic waters to find the corresponding sites in the crystal structure.

For GCMC simulation data, extra steps were taken before clustering-based analysis. Particularly, as the GCMC implementation in *grand* makes use of non-interacting 'ghost' water molecules, which are used for insertion moves, these waters were first translated out of the simulation cell, such that they would not interfere with visualisation or structural analyses.

3.6.2 Electron Density Calculations.

Mean structure factors were computed from aligned MD trajectory snapshots. Structure factor calculations were performed using `xtraj.py`, a Python script distributed in the LUNUS open source software for processing, analysis, and modeling of diffuse scatter-

ing[65] (`https://github.com/lanl/lunus`). `xtraj.py` combines methods in the Computational Crystallography Toolbox (CCTBX)[66] and the MDTraj library for MD trajectory analysis[64] to compute the structure factor of each snapshot. The `xtraj.py` script is invoked at the Unix command line as `lunus.xtraj` when LUNUS is installed as a module in CCBTX. In `xtraj.py`, MDTraj I/O methods are used to read the trajectory in chunks that may be processed in parallel using MPI. A reference PDB structure is read in using the CCTBX I/O methods, and the atomic coordinates are replaced by those in a snapshot from the MD trajectory. Structure factors are computed from the modified structure using the `cctbx.xray.structure.structure_factor()` method and are accumulated within each MPI rank, along with a count of the number of frames processed. The global sums of the structure factors and frame counts are computed via MPI reduction, and the mean is computed as the aggregate sum of the structure factors divided by the frame count. Electron density maps were computed from the structure factors using CCP4 tools.[67] By default the maps were normalized to have units of the standard deviation and a zero mean. Maps were computed in absolute units (electrons per cubic Angstroms) as needed using the amplitude of the structure factor at Miller indices (0,0,0) (F000) and volume values reported by `mmtbx.utils.f_000()`, `cctbx.xray.structure.unit_cell().volume()`, respectively, within `xtraj.py` using the `fft` method[68–70] in CCP4. Example scripts to perform this analysis are available on `https://github.com/MobleyLab/water_benchmark_paper`.

To compare the experimental and calculated electron density maps, we visualized both maps using Coot molecular graphics[71,72] (v0.9.4) we used a contour level of 3 sigma for calculated water electron density maps and 1.5 sigma for experimental protein/water maps across all systems. However, making this quantitative also requires calculating a metric describing density agreement, such as the real space correlation coefficient (RSCC). Our previous experience with RSCC suggests it may still need improvement as a metric, so measuring quantitative agreement is a research topic we do not address in the present work.

3.7 Accuracy and Efficiency Comparison

Before we move on to Section 4, it is important to clarify our definition of a success case in this work. In the simulation where all crystallographic water molecules were removed in the starting structures, we checked if all target sites can be rehydrated. Correspondingly, to define a success case, a main question here is how much we should expect that the water site will be occupied in the successful simulation? The crystallographic water occupancy is not available from the experimental crystallography data (waters are universally deposited at 100% occupancy) which makes it more difficult to judge the simulation's performance. In a recent study by Ross et al.,[31] the average water occupancy of target water sites was checked over simulation times ranging from 30 ps to 1 ns, and simulations were considered successful when the average water occupancy achieved at least 80% (see bottom right panel of Figure 2, where occupancies are around 80%; these are cited as examples of successful rehydration). Inspired by that work, we consider simulations a success when all target water sites are occupied for at least 80% of the time in a single simulation block (Figure 2A).

Here, for each technique, multiple separate simulations were performed and we checked all of them to find any simulations that achieve success for all target sites (Table 1). The simulation length of a single simulation block is not the same in different techniques (BLUES: 12ns, *grand*: 2.5ns, MD: 35ns) so the definition here is not perfect. However, in practice our results are not sensitive to this simulation time because the performance difference between these techniques is very large (more details in Table 2). There might be other good or better definitions of success than the one employed here; however, we use this one because of the literature precedent, and hope the field will settle on a more universal definition of success in future work.

When analyzing electron density maps, we must use a different criterion of success. There, we consider a test successful if the averaged electron density map calculated from simulations overlaps well with the experimental $2F_o$-F_c map (Figure 2B) from visual inspection. In most systems, this analysis led us to the same conclusions as did the clustering based analysis.

We will talk about a few exceptions later in Section 4.

As BLUES and *grand* use both MD and NCMC or GCMC, we must account for the nontrivial cost of the NCMC/GCMC portion. Thus, we decided to use total force evaluations to compare efficiency between different simulation techniques. While this is not a perfect metric, it at least does a better job accounting of these differing costs than does a more traditional metric like total simulation time. Additionally, comparisons based on wallclock time do not account for differences in compute hardware or in how much optimization has gone into improving efficiency for the particular task at hand, whereas a comparison based on force evaluations places diverse methods on relatively equal footing.

Typically, a simulation will have a total cost, in force evaluations (FEs), of $(N + M) \times n$ where N is the number of MD steps per iteration, M is the number of NCMC/GCMC steps per iteration, and n is the number of total iterations. We check the time (force evaluations) required to achieve success as defined above (Figure 2). Since multiple separate simulations were performed for each technique, we reported force evaluations required to achieve success for each simulation (Table 2). The average force evaluations across separate simulations for each technique were used when comparing the efficiency of rehydrating all target sites between these techniques. We also observed some cases where one technique was able to rehydrate all target sites in only some simulations but not all of them. In such cases, we also report the failures in Table 1 (labeled as "F").

GCMC moves were applied in the equilibration phases of *grand* simulations. This may introduce biases in our efficiency comparison since no enhanced sampling was applied in the equilibration phase of BLUES and MD simulations. A better way to compare these techniques in the future study would be starting simulations from the same point for production runs to exclude potential bias from starting structures equilibrated differently. However, that approach was not employed here because we chose to equilibrate with protocols which had been recommended for each method in prior work. In this work, we considered force evaluations of equilibration simulations in our efficiency comparison by reporting the sum of

force evaluations of both equilibration and production phases for these methods (Table 2)

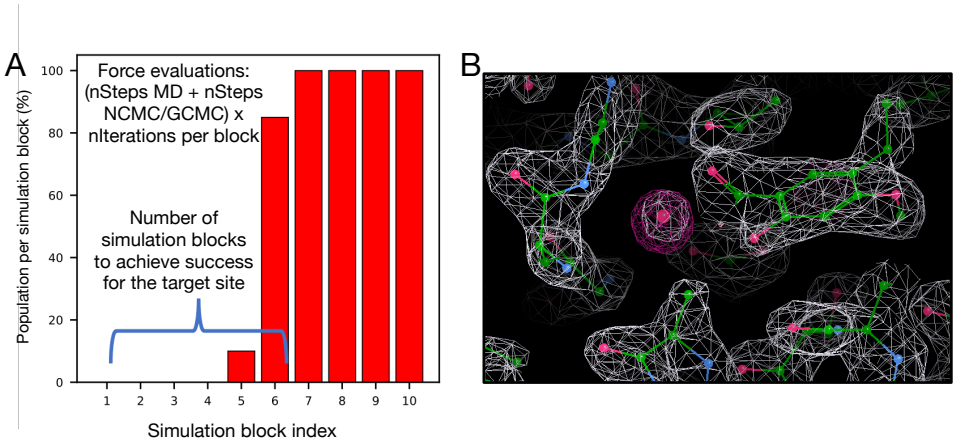


Figure 2: Examples of water occupancy and electron density maps for success cases. (A) Bar graphs show the water occupancy of a target hydration site in a single simulation. We consider it successful if the occupancy of a target water site is higher than 80% in a single simulation block. If successful, we check the force evaluations required to achieve this point including the equilibration phases. (B) The calculated electron density map of water molecules from simulation data (magenta) and the experimental determined density map ($2F_o$-F_c map) (white).

4 RESULTS AND DISCUSSION

4.1 Both BLUES and *grand* outperform normal MD simulations at sampling water motions and rearrangements.

Based on the definition described in Section 3.7, we calculated the overall success rate of each simulation technique. We defined success as achieving at least 80% occupancy of all target hydration sites in a single simulation block. Even though we performed BLUES and MD simulations in the presence and absence of crystallographic water molecules, we only compared performance between techniques based on simulations done without crystallographic waters. Based on our success definition, the water molecules need to escape from the target sites prior to rehydration in simulations begun with ordered water molecules retained. However, we observed those hydration sites were always occupied in simulations of almost all systems

studied in this work, when simulations were begun with crystallographic waters. Thus, we only focused our analysis on those simulations in which crystallographic water molecules were removed in advance. These simulations provided an opportunity to test if these water molecules are favorable with the force field used in this work.

BLUES and *grand* successfully rehydrated all target sites in 8/10 (80%) and 9/10 (90%) systems, respectively. They both improve water sampling relative to normal MD (70%) when applied to the systems studied in this work (Table 1), given the simulation lengths tested here. Although it only shows minor but not significant advantages of BLUES and *grand* than MD if we only consider the number of success case, in those systems where all simulation techniques were able to rehydrate all target sites, normal MD was much more expensive than BLUES and *grand* (Table 2).

We also find performance of these methods is highly dependent on the choice of protein target and other details. Remarkably, we find that *grand* is the only method that can rehydrate both target water sites in the PTP1B system (PDB: 2QBS), in which previous work also showed success using GCMC.[31] It is also interesting to see that in the case of a HSP90 system (PDB: 3RLQ), both *grand* and BLUES failed to rehydrate one specific hydration site out of three target sites in total but normal MD was able to do so. We will discuss more details of these findings in Section 4.3.2.

As shown in Figure 1, we studied 13 systems. One BTK-ligand system (PDBid: 4Z3V) does not have buried water site and serves as a control. Particularly, we included it in this work to check if these techniques put water molecules in the binding site where no ordered water molecules were placed in the deposited crystal structures. Based on our definition described in 3 section, we do not consider this system when comparing the overall success rate between these methods. The two BRD4 systems (PDBid: 5I80, 5I88) were also excluded since binding site unfolding and frequent ligand unbinding events were observed in the simulations. These observations are interesting and may be important for future work on BRD4, but are not our main focus here. We may further explore these systems and report

our findings in a separate study.

The ligand in the BTK-ligand system (PDBid: 4Z3V) observed experimentally (in crystal structures) displaces a crystallographic water molecule bridging the ligand and protein in another BTK-ligand system (PDB: 4ZLZ). In our simulations, none of the techniques employed here (MD/GCMC/NCMC) led to insertion of a water in the region from which the ligand had displaced it, confirming that all these techniques can distinguish hydration sites in the area of the binding site.

Table 1: Summary of performance of each technique in each system (PDBid listed) simulation shown as n/m where n is the number of successfully rehydrated water sites and m is the number of target water sites. A successfully rehydrated water site is defined to have an occupancy of at least 80% in a single simulation block. All ordered water molecules were removed prior to simulations.

	2QBS	2XAB	2XJG	3RLP	3RLQ	3RLR	4ZLZ	5I29	5I1Q	2ZFF
BLUES (restrained)	1/2	3/3	1/1	4/4	2/3	1/1	1/1	5/5	5/5*	1/1*
MD (unrestrained)	1/2	2/3	1/1	4/4	3/3*	1/1	0/1	5/5*	5/5*	1/1
MD (restrained)	0	0	0	0	0	0	0	5/5*	2/5*	0
grand (unrestrained)	2/2	3/3	1/1	4/4	2/3	1/1	1/1	5/5	5/5	1/1
grand (restrained)	2/2	N.D.	N.D.	4/4	N.D.	N.D.	N.D.	N.D.	5/5	0

*based on the electron density map analysis.

4.2 *grand* (GCMC/MD) is more efficient than BLUES and MD in rehydrating all target water sites.

We reported the force evaluations of multiple replicates for each simulation method in Table 2. There are several systems which pose challenges to *grand* because of protein/ligand motions (TAF1(2), PDB: 5I1Q) or where it is known that maintaining the crystallographic pose is critical for successful water insertion (PTP1B, PDB: 2QBS). In these systems, we used position restraints on the protein/ligand (see Section 3) and observed a better performance of *grand* (Table 2). We are able to do so in this work because the crystal structures are available for all of these systems we studied. However, it is important to highlight such

limitations in *grand* as we will discuss in more detail below, as these restrictions may impact *grand*'s utility in making predictions when structural information for the simulated systems might not be available.

We compared efficiency using the average force evaluations across multiple replicas for each technique. In some systems, only one replica of a technique was able to rehydrate all target sites. Failed simulations were reported in Table 2 as "F". Those systems where some simulations failed to rehydrate all target sites are indeed challenging for these techniques, as reflected by the fact that even in the successful simulations, it was expensive to do so (Table 2). For example, in the case of a HSP90 system (PDB: 3RLP), it cost the only successful BLUES simulation 126M force evaluations (equivalent to 84 ns simulation time). In MD simulations, it could be even more expensive in the only successful trial. For example, it took 175M force evaluations (equivalent to 350 ns) for MD simulation to rehydrate the single target site in the case of a HSP90 system (PDB: 2XJG).

In all these systems, *grand* simulation more efficiently populates the water sites than BLUES and MD (Table 2). We noticed that in *grand* simulations of several targets (HSP90 (PDB: 2XAB, 2XJG, 3RLR), TAF1(2) (PDB: 5I29)), all of the targets' water sites were occupied even before the production run. This is due to the fact that GCMC was used to equilibrate water in the equilibration phase, unlike BLUES and MD, where the equilibration was done in the normal NVT/NPT ensemble. These results highlight the benefits of using GCMC to equilibrate water molecules even without using GCMC in production runs; this approach has been shown to help obtain adequate water sampling for better binding free energy estimations even without applying it in production runs in a previous study.[31] One potential way to take advantage of GCMC sampling in BLUES/MD simulations is running GCMC to equilibrate water molecules in prior to production runs. In this work, we did not apply this strategy in our tests. However, if all water sites are rehydrated by GCMC moves in the equilibration simulations then the approach becomes equivalent to running BLUES/MD simulations begun with all crystallographic water molecules retained, an approach we also

tested in this work. Although our results are system dependent, we observed no transitions (removal/rehydration) of water molecules in the target sites in BLUES/MD simulations for most systems. That said, successful attempts of rehydrating water molecules are still done by GCMC moves and using GCMC in equilibration phase does not improve BLUES/MD performance in production runs.

In *grand* simulations of a PTP1B system (PDB: 2QBS), we performed simulations with restraints on heavy atoms in addition to the non-restrained simulations because (1) the success of water insertion is sensitive to the receptor conformation (as observed in a previous study[31] and elaborated personal communication with author Greg Ross) and (2) in several systems, we observed the protein/ligand motions impaired the performance of *grand*. The latter factor should not be a problem if we run simulations long enough to allow the protein/ligand to rearrange enough that water insertions from *grand* are more likely to be accepted. But in this work we do not have unlimited computational resources and in most systems *grand* is able to rehydrate all target sites no more than 5 simulation blocks (e.g., 12.5 ns, 7M force evaluations, see Section 3.5). Given these issues, we used restraints in *grand* simulations when some replicates failed to insert water molecules due to protein/ligand motions.

Restraints can be used to keep a protein/ligand in a specific conformation to accelerate the sampling of target water sites when the structural information of the system with target hydration sites being occupied is known (such as from crystal structures or other techniques), as is the case here. This is an important factor we considered when we selected these systems in this study since then we can investigate the performance of these techniques in placing water in known structures. Here, using restraints on the heavy atoms of both receptor and ligand significantly improve the performance (both efficiency and accuracy) of GCMC simulations of this system (all three replicates rehydrated both target water sites within 2.5 ns and 1.4M force evaluations).

As mentioned in Section 3, we used both clustering-based analysis and electron density

map-based analysis to gauge success (shown in Table 1 and 2). For most targets, we can obtain clear results in terms of whether all target sites were rehydrated or not using clustering-based analysis and shown in Figure S15. However, in TAF1(2) simulations (PDBs: 5I1Q, 5I29), no clear water patterns can be revealed from such analysis due to the flexibility of the protein side chains and ligand when no restraints were used. So we switched to electron density map-based analysis to check occupancy of the target sites. In BLUES simulations of one TAF1(2) complex system (PDB: 5I1Q) and one thrombin system (PDB: 2ZFF), we observed occupancy of all target sites but not all of them stayed occupied at least 80% time of a single simulation block, our success criterion (see Section 3). However, electron densities calculated from these trajectories showed that all target sites were sampled in simulation and calculated densities overlapped well with the experimental ($2F_o$-F_c) maps. We will discuss these results in more detail in the following section. Based on that, we still considered both of these as successes and used the simulation time taken to obtain the calculated densities in our efficiency comparison (Table 2).

Table 2: Summary of the efficiency of each technique (in force evaluations) in each system (PDBid listed) simulation. All ordered water molecules were removed prior to simulations. If any simulation failed to rehydrate each target site based on our defined criteria (with an occupancy of higher than 80% in a single simulation block, see Section 3), the result is shown as "F" in the table. We calculated force evaluations (in million evaluations) required to achieve the point where all hydration sites were successfully rehydrated in each simulation. The calculated force evaluations include both equilibration and production phases. See Section 3 for detailed force evaluation in equilibration phases for each method. "N.D." means no data since the simulation was not conducted.

	2QBS	2XAB	2XJG	3RLP	3RLQ	3RLR	4ZLZ	5I29	5I1Q	2ZFF
BLUES (restrained)	F/F/F	77.5/77.5/41.5	23.5/23.5	F/F/131.5	F/F/F	5.5/5.5/5.5*	23.5/23.5/23.5	5.5/5.5/5.5*	23.5/77.5**	F/F/23.5**
MD (unrestrained)	F/F	F/F	F/180.5	F/285.5**	F/285.5**	215.5/180.5**	F/F	F/320.5**	F/285.5**	320.5/355.5
MD (restrained)	F/F	F/F	F/F	F/F	F/F	F/F	F/F	5.5/5.5	F/F	F/F
grand (unrestrained)	F/3.5	2.1/0*	0.7/0.7*	3.5/9.1	F/F	0.7/0.7*	2.1/2.1	0/0*	11.9/F	F/4.9
grand (restrained)	2.1/2.1/2.1	N.D.	N.D.	F/0.7/F	N.D.	N.D.	N.D.	N.D.	F/3.5/3.5	F/F/F

*all water sites were already occupied before the production run and occupied in the production run.
**based on the electron density map analysis.

4.3 Lessons we learned from failures.

We found that none of these simulation techniques can rehydrate all target sites in all of the systems we studied. To understand the advantages and limitations and better develop these techniques in the future, we analyze the failures.

4.3.1 Failures of MD simulations.

Large energy barriers can impede water rearrangements, making it difficult for unbiased MD simulation to adequately sample rearrangements of buried water molecules. Given this, we were not surprised that MD failed to rehydrate each individual target site in this study. However, we noticed that using restraints on the receptor and ligand was helpful to achieve better performance in BLUES and *grand* simulations for several targets. Thus, we tested the same restraints in MD simulations to explore potential benefits in water sampling. The results showed that no significant performance differences were observed using restraints in MD simulations compared to normal MD. We only observed in one system (TAF1(2), PDB: 5I29) in which all target sites were rehydrated faster in the simulations where restraints were applied than in unrestrained MD. In conclusion, the benefits of using restraints on the protein and ligand in water sampling using MD simulation are system-dependent and are negligible in most of the systems studied in this work.

4.3.2 Failures of BLUES simulations.

The PTP1B system (PDB: 2QBS) is a challenging case for BLUES. Different strategies were tested but BLUES still cannot rehydrate Site 2 (Figure 3A-B). This binding site has two ordered waters crystallographically, so we further tested whether both water molecules are favorable in the binding site by checking their stability when simulations are begun from crystallographic poses. We set up these simulations in the presence of crystallographic water molecules. In MD simulations with restraints on the protein and ligand, both water molecules stay in the binding site for 100 ns. But in BLUES simulations,

Site 1 could be emptied and rehydrated in the simulation (Figure 3C) which agrees with the simulation started after removal of crystallographic water molecules (Figure 3B). However, Site 2 is so favorable that NCMC moves could not remove the water molecule from the site (Figure 3C). Taking these results together, we can conclude that when the protein-ligand complex maintains the crystallographic pose, (1) BLUES can rehydrate Site 1 but not Site 2, (2) BLUES can remove the water molecule from Site 1 but not Site 2, and (3) both Site 1 and Site 2 are favorable in the unbiased MD simulations (with restraints). So our hypothesis is the protein/ligand need to reorient/move to successfully insert the water molecule to Site 2. Thus, we did tests where we removed the restraints on the ligand but still kept the protein heavy atoms restrained to provide flexibility of the ligand in BLUES simulation. However, we found we still could not insert the water into Site 2 whereas Site 1 could be rehydrated. We also parametrized the protein with deprotonated CYS215 as suggested by the author of a previous study[31] where they successfully rehydrated both sites (personal communication) using GCMC simulation. But this did not help BLUES to rehydrate Site 2 either.

We then checked the protocol work distributions that are accumulated over the course of the NCMC move attempts for each site when we applied restraints on both the ligand and protein (Figure 3F). We found the mean value of the protocol work for Site 2 is higher than that for Site 1 and is far away from the favorable region (work values close to 0), making it less possible to accept these moves. From the work distribution (Figure 3F), we can see that several attempts for Site 1 are in the region where is close to 0 whereas none of the moves attempted for Site 2 are close to being there. This system, and Site 2 in particular, thus appears to be particularly challenging for BLUES; even given so much attention, we are still unable to rehydrate these sites reliably.

In unbiased MD simulation, Site 2 was rehydrated whereas Site 1 was not (Figure 3D). This differs from our BLUES results (Figure 3B) and is due to ligand flexibility and motion which were allowed here since no restraints were applied in normal MD. Particularly, we find that the ligand moved to partially occupy or block Site 1 in the binding site (Figure S1)

so that water molecules could not be inserted. This is not observed in BLUES simulations since we restrained the ligand to the crystallographic pose. We additionally performed five separate unbiased MD simulation (200 ns for each) in the presence of all water molecules in the crystal structure to test whether both hydration sites are favorable in the binding site with the chosen force field. We found in 2/5 of these simulations, Site 1 stayed occupied over the course of the simulation (200 ns) (Figure S2E-F). We observed the water molecule escaped from Site 1; when this happened, the ligand, which could move, took the water's space in Site 1 so that no rehydration happened in 2/5 simulations (Figure S2B-C). We also found in 1/5 simulations that the water molecule escaped from Site 1 first and came back before the ligand could take the position of Site 1 (Figure S2D). These results suggest that both Site 1 and 2 are favorable in the binding site but the successful rehydration of Site 1 in unbiased MD simulation is dependent on the ligand motion. If the ligand moves closer to the binding site, taking the space in Site 1, then the water site is not likely to be successfully rehydrated (Figure 3D, Figure S2C-D).

A previous study[31] where GCMC was used successfully rehydrated both hydration sites in this system and we expected the same success in this work using *grand*. However, when no restraints on the protein/ligand were applied, we found that it is challenging for *grand* simulations to rehydrate Site 2 (Figure S3B). In the presence of restraints, as in BLUES, both sites could be rehydrated (Figure 3E). These results suggest the crystallographic pose of both the protein/ligand is critical for successful insertion of water molecules to both sites, as also confirmed by the author of the previous study (personal communication) which used position restraints on the heavy atoms of the protein/ligand.[31]

Both BLUES and *grand* failed to rehydrate one target site in one HSP90 system (PDB: 3RLQ) but normal MD could do so. It is clear that both Site 2 and 3 (Figure 4B-C) could be rehydrated in both BLUES and *grand* simulations but not Site 1. In normal MD simulations, surprisingly, Site 1 could be rehydrated even though the occupancy is lower

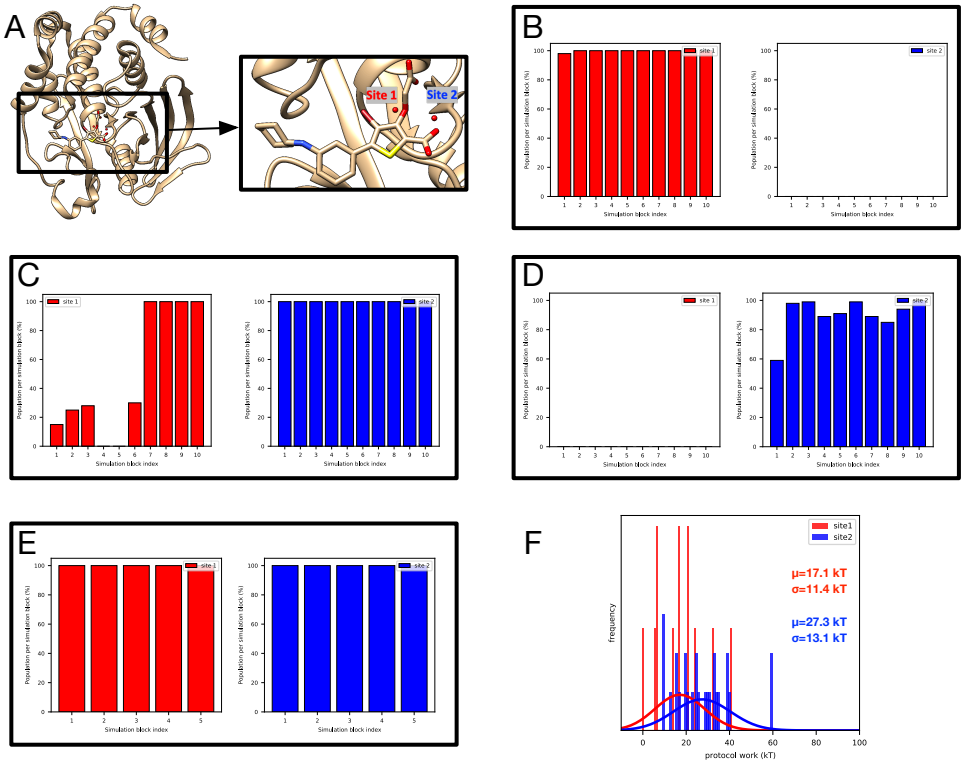


Figure 3: (A) PTP1B system with target hydration sites (red and blue). Bar graphs show the water occupancy of target sites in a single (B) BLUES simulation (with ordered water molecules removed prior to simulation), (C) BLUES simulation (with ordered water molecules retained), (D) normal MD simulation, and (E) *grand* simulation with restraints on the protein and ligand heavy atoms. (F) Protocol work distribution of NCMC with water hopping move attempts for Site 1 (red) and Site 2 (blue) fitted with a normal distribution. The mean and standard deviation of the work distribution are shown using the same color code.

than 80%, the standard we defined for success. However, the calculated electron density map from simulations agrees well with the experimental map (Figure 5C) so that we still consider it a success case. This is the only case we studied where MD appears superior to BLUES and *grand* simulations. We also performed both BLUES and normal MD simulations where all of the crystallographic water molecules were retained in the simulations. We saw that the water in Site 1 was removed in BLUES simulations very quickly (Figure 4D) and escaped from the site in normal MD simulations (Figure 5A), suggesting this site was not favorable in simulations with this force field. It is possible that the biased sampling in BLUES (NCMC) and *grand* (GCMC) just accelerated the escaping of the water molecule from Site 1, but that

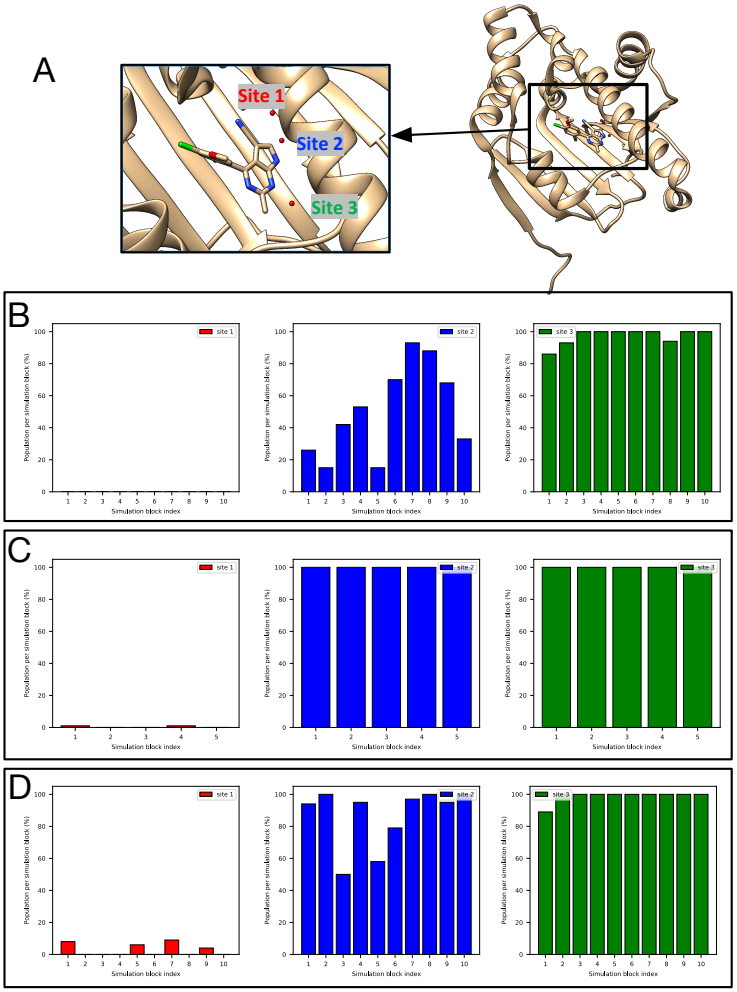


Figure 4: (A) The HSP90 system (PDB: 3RLQ) and target water sites. Bar graphs show the water occupancy of target sites in a single (B) BLUES simulation (with ordered water molecules removed prior to simulation), (C) *grand* simulation, and (D) BLUES simulation (with ordered water molecules retained).

this took much longer in normal MD. In fact, by checking the experimental electron density map we found Site 1 has a weaker peak than that of Site 2 and 3, suggesting the probability of observing a water molecule in this site perhaps ought to be lower (Figure S4). However, out of a desire to avoid overfitting, crystallographic water molecules are typically deposited at 100% occupancy even when density is relatively weak (as is the case for this water), complicating interpretation. It is also notable that there were two copies of the protein in the asymmetric unit in the crystal structure with this PDB code (3RLQ) and we used the first chain to prepare our simulations. But in the second chain, water molecules were

deposited in both Site 2 and 3 but not Site 1 (unlike in the first chain), further suggesting the uncertainty of the occupancy of Site 1 in the crystal structure. Thus, our simulation results here seem somewhat consistent with the relatively lower experimental electron density for this water, though we are skeptical that this particular site is favorable at all with the present force field.

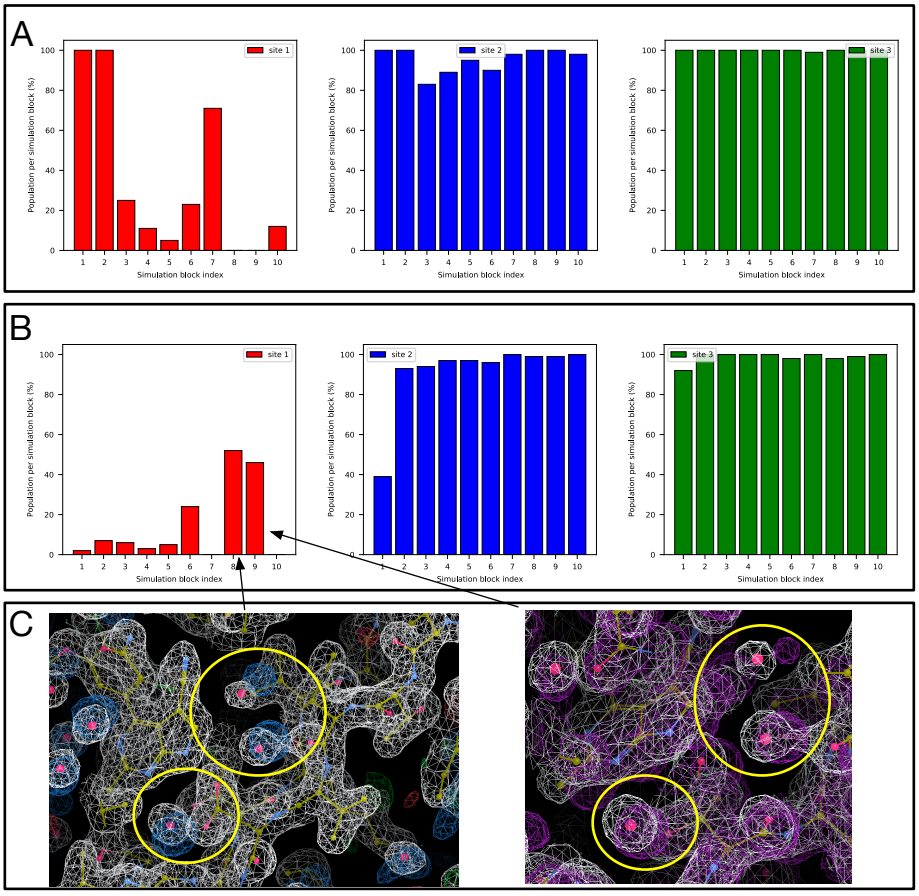


Figure 5: Bar graphs show the water occupancy of target sites of the HSP90 system (PDB: 3RLQ) in a single MD simulation (A) with ordered water molecules retained, (B) with ordered water molecules removed prior to simulation. (C) The calculated electron density map from two separate simulations (blue and purple) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The target hydration site is circled in yellow. The calculation is based on two MD simulation trajectories shown in (B).

4.4 Lessons we learned about the systems studied.

In the following sections, we will discuss what we learned about the systems we simulated, including insights beyond simple analysis of water sampling. We hope these results will aid future work on these systems.

4.4.1 HSP90 (PDB: 2XAB)

Both BLUES and *grand* can rehydrate three target sites in HSP90 with this ligand (Figure 6). But unbiased MD can only rehydrate 2/3 sites even with much longer simulation times (700 ns). All three sites were highly favorable and none of them could be removed whether simulations started with or without ordered waters.

4.4.2 HSP90 (PDB: 2XJG)

Relative to the HSP90 system just prior, the ligand in this case is modified in a way which displaces two water molecules in the binding site. MD/BLUES/*grand* can all rehydrate the only target site (Site 1 in Figure 7A) although it took MD much longer (350 ns in total, 175M force evaluations) to achieve this (Table 2).

Besides the target site, we found another favorable site (Site 2 in Figure 7A) near the binding site in BLUES/MD/*grand* simulations (Figure 7C-D). The electron density map from our simulations also confirms the existence of this extra water site (circled in cyan in Figure 7B). By checking snapshots extracted from the simulations, we found this water molecule forms a hydrogen bonding network that also involves SER52, ASP93, THR184 and the crystallographic water in Site 1. A previous study of this system also observes this site being occupied in their simulations but no ordered water is deposited in the crystal structure.[30] We did not see significant experimental electron density in this water site either. Both that work and our work used same solvent model (TIP3P) and force field for protein (AMBER ff14SB), suggesting this is a force field issue.

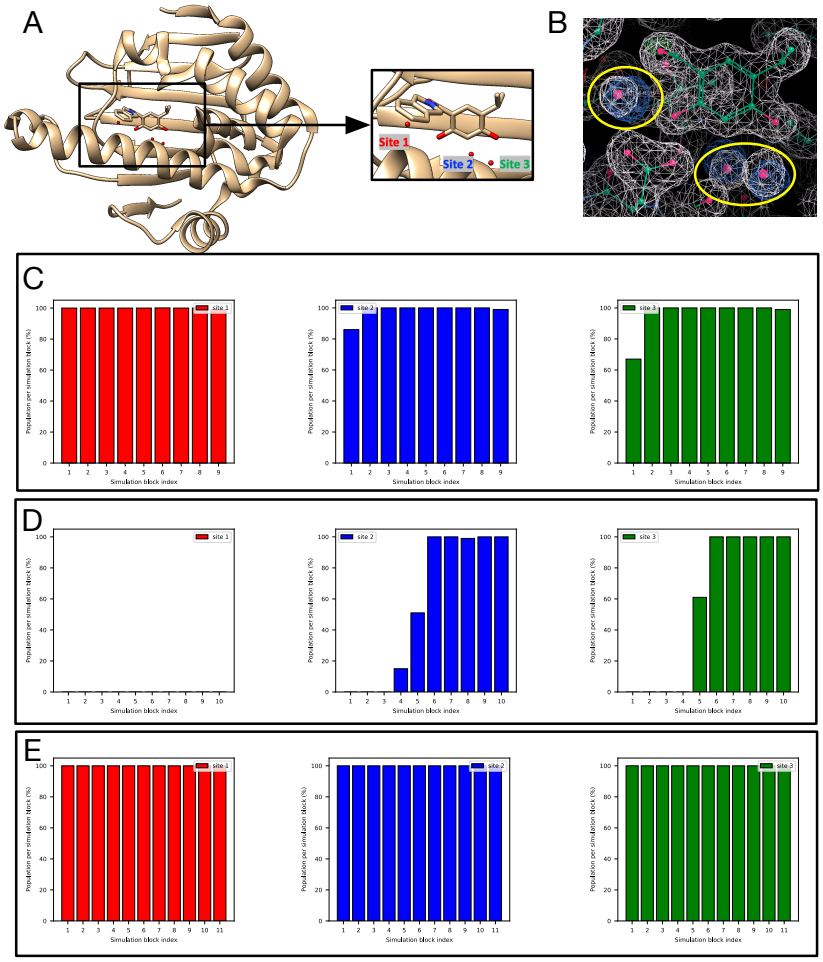


Figure 6: (A) The HSP90 system (PDB: 2XAB) and target water sites. (B) The calculated electron density map (blue) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The target hydration sites are circled. The calculation is based on a BLUES simulation trajectory. Bar graphs show the water occupancy of target sites in a single (C) BLUES simulation, (D) unbiased MD simulation, and (E) *grand* simulation.

4.4.3 HSP90 (PDB: 3RLP)

This additional HSP90 case focuses on a different ligand series (Figure 1) from those above (PDBs: 2XAB, 2XJG). This system has four target water sites (Figure 8A) and it is more challenging to insert a water molecule in to Site 1 than the other sites (Site 2-4) (Figure 8C-E). In unbiased MD simulation, the occupancy of Site 1 did not exceed 80% in any simulation block (Figure 8C) but the calculated electron density map overlaps well with the experimental $2F_o - F_c$ map (Figure 8B). So we consider it a success case for unbiased MD

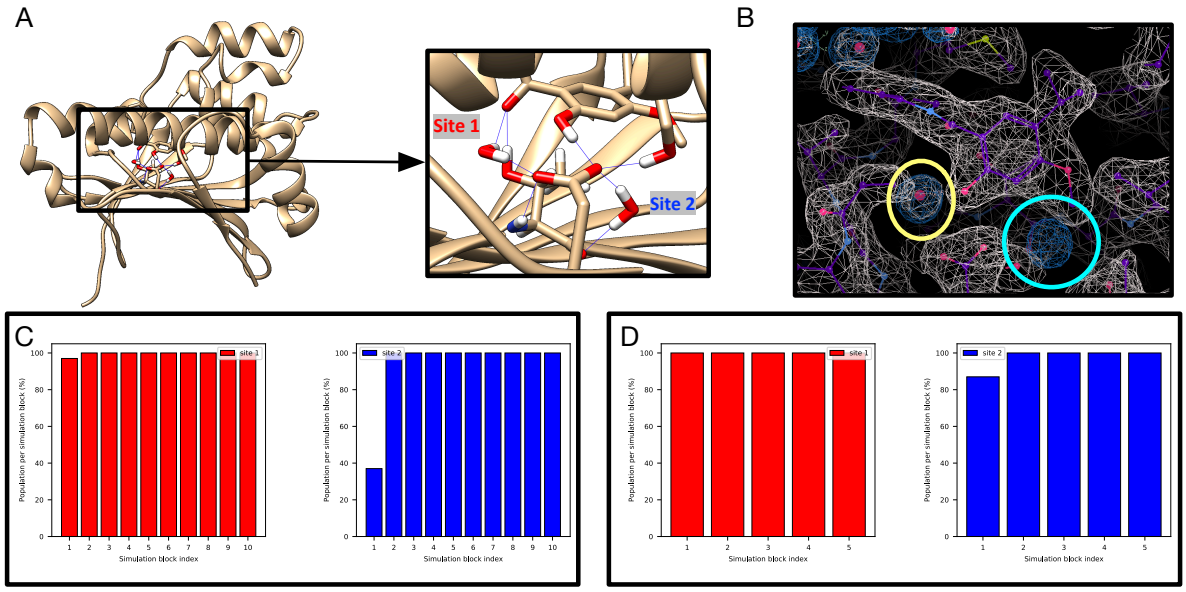


Figure 7: The HSP90 system (PDB: 2XJG) and target water sites. (B) The calculated electron density map (blue) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The target hydration site is circled in yellow and the extra site is circled in cyan. The calculation is based on a BLUES simulation trajectory. Bar graphs show the water occupancy of target sites in a single (C) BLUES simulation, and (D) *grand* simulation.

simulations. Here, all simulation methods including unbiased MD successfully rehydrated all four target sites. Still, MD requires more force evaluations for success: ∼2 and ∼34 times more than BLUES and *grand* simulations, respectively (Table 2).

Besides exploring water sampling issues, we also learned about the protonation state of the ligand. Based on pKa estimates from Chemicalize (a ChemAxon product, `https://www.chemaxon.com`), there are two possible protonation states for the ligand at the experimental conditions (pH=4.3) (Figure S5A-C). However, the ligand is not stable in the binding site with one of the protonation states and escaped quickly in both unbiased MD and *grand* simulations even at a timescale shorter than 2.5 ns (Figure S5D-E). We didn't observe such unbinding events in BLUES simulation because the ligand was restrained. The other protonation state of the ligand showed much better stability in the simulations (as long as 700 ns of unbiased MD). Thus, we believe for this system, the ligand protonation state as shown in Figure S5C dominates when the ligand is bound.

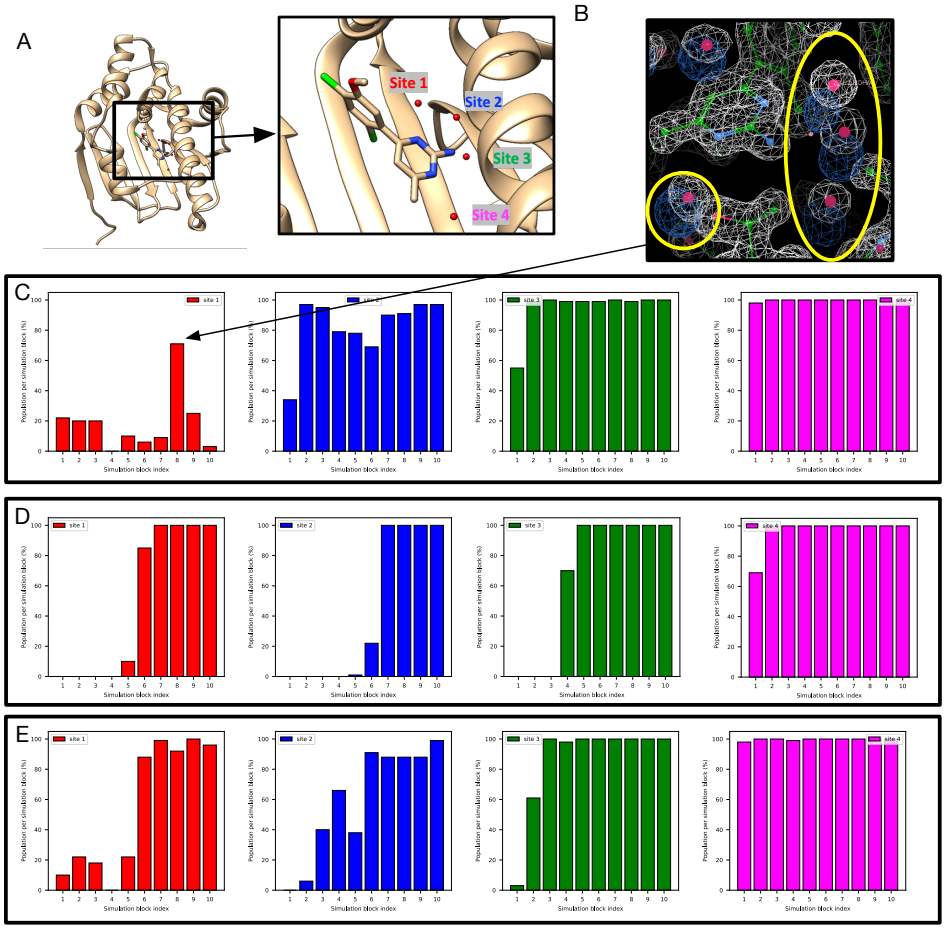


Figure 8: (A) The HSP90 system (PDB: 3RLP) and target water sites. (B) The calculated electron density map (blue) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The target hydration sites are circled. The calculation is based on a MD simulation trajectory shown in (C). Bar graphs show the water occupancy of target sites in a single (C) unbiased MD simulation, (D) BLUES simulation, and (E) *grand* simulation.

4.4.4 HSP90 (PDB: 3RLR)

The modified ligand in this system displaced additional two water molecules from the system mentioned above (HSP90, PDB: 3RLQ). The only target site (Figure S6A) can be rehydrated in the equilibration phase of BLUES/*grand* and in the production run of MD with a high cost (350 ns, 175M force evaluations). In both simulations the water site was rehydrated in the equilibration phase.

4.4.5 TAF1(2) (PDB: 5I29)

There are five hydration sites in the TAF1(2) system (Figure S7A). The binding site is close to bulk solvent and the ligand is flexible in the binding site which poses challenges in the analysis of the unbiased MD simulation using clustering-based methods. So we used our electron density-based analysis and found that all target sites were rehydrated in MD simulations (Figure S7B). But both BLUES and *grand* simulation were much more efficient since they both rehydrated all five sites during the equilibration phase. It is not surprising that *grand* is able to do so since GCMC moves were deployed in equilibration simulations. However, it was surprising to see all five sites rehydrated during the preparation for BLUES simulations given the fact that no biased sampling was applied in equilibration – in other words, these sites were rehydrated while equilibrating with standard MD. This may be because the protein and ligand heavy atoms were restrained to the crystallographic pose in equilibration for BLUES. To test if this fast equilibration for BLUES here was simply a random fluctuation, we performed four additional equilibration simulations. In three of these five additional equilibration simulations, all five sites were rehydrated. In the other two simulations, 3/5 and 4/5 sites were rehydrated. Thus, we find that these water sites rehydrate quickly when the simulation is restrained. Additionally, this helps make this system an easy case for BLUES and *grand*, or for any method which restrains the ligand and protein.

In this system, we also found both BLUES and *grand* simulations could remove the water molecules from the sites after they were occupied and then rehydrate them again, indicating we could converge population estimates. This ability to sample multiple water transitions into and out of the sites is not common for the systems studied here. One possible reason for the additional ease of sampling here could be that this binding site is large and more exposed to the bulk solvent than binding sites in the other systems examined.

4.4.6 TAF1(2) (PDB: 5I1Q)

A modification of the ligand in the TAF1(2) system discussed above changes the water network in the binding site (Figure 9A). In BLUES simulations, Site 5 is never occupied for at least 80% of the time in a single simulation block (Figure 9C). But the calculated electron density map agrees well with the experimental $2F_o - F_c$ map so we consider this a success case. Similarly, we found that unbiased MD simulation achieved good agreement between the calculated electron density map and the experimental map (Figure S8). Presumably this is because water moves around enough in the region of the water site to accumulate significant density there, even though it does not occupy the specific site a large fraction of the time. Similar to another TAF1(2) system (section 4.4.5), clustering-based analysis did not return clear water patterns in the binding site and electron density map-based analysis provides a better approach to check hydration site occupancy here.

Even though we showed that BLUES simulations can rehydrate all five sites in TAF1(2), the occupancy of Site 5 is not high in the simulations (Figure 9C). This observation agrees with the results of the simulations with ordered water molecules retained (Figure 9D). This is consistent with the results of the MD simulations done with restraints on the protein and ligand in the presence of crystallographic water molecules (Figure S8E). These results suggest Site 5 is not as favorable as other sites in the simulations when the protein and ligand maintain the crystallographic pose. In fact, we observed slightly different water network in that region in our simulations when compared to the crystal structure (Figure S9). We believe this is a limitation of the force field since different simulations converged to the same results.

This system poses challenges to *grand* simulations when the protein and ligand are not restrained; without restraints, *grand* has difficulty rehydrating each target site. The ligand moves in the binding site when it is not restrained and may occupy the space of one hydration site (Figure S10A-B), blocking the successful insertion of water molecules (Figure S10C). When the protein and ligand were restrained to maintain the crystallographic pose, all five

sites were rehydrated in *grand* simulations (Figure S10D), as in the case of the other TAF1(2) ligand examined above.

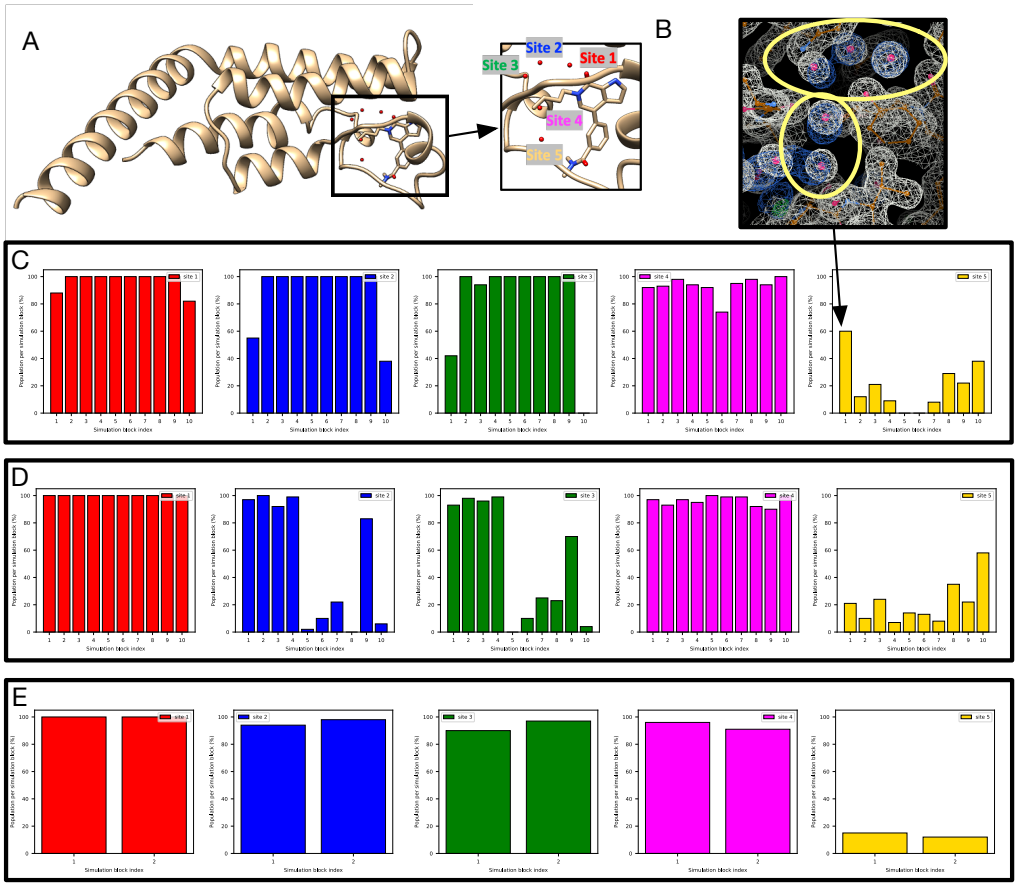


Figure 9: The TAF1(2) system (PDB: 5I1Q) and target water sites. (B) The calculated electron density map (blue) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The target hydration sites are circled. The calculation is based on a BLUES simulation trajectory shown in (C). Bar graphs show water occupancy in the target sites in a single (C) BLUES simulation (with ordered water molecules removed prior to simulation), and (D) BLUES simulation (with ordered water molecules retained), and (E) MD simulation with restraints.

4.4.7 BTK (PDB: 4ZLZ)

This BTK system has one target hydration site in the binding site, bridging the protein and ligand as shown in prior work.[73] Both BLUES and *grand* can rehydrate this site efficiently (Figure 10).

When we checked populated hydration sites in the simulations, we found another water

molecule in the crystal structure that is close to the target site (Figure S11). Although this is not the target site, we found a poor sampling of this site in BLUES simulations. Even in the simulation starting from a structure with all crystallographic water molecules (Figure S11), we still found only one populated hydration site centered between the two crystallographic water molecules (Figure S11D). This is consistent with the results of the simulations with ordered water molecules removed prior to simulations (Figure S11A). The calculated electron density map of this populated site revealed by BLUES simulation overlaps with both water sites in the crystal structure including the target site (Figure S11B). We also noticed this water site close to the target site in the crystal structure shows a weaker peak in experimental densities. A previous study[30] done on the same system did not sample this site either in their simulations (personal communication). Given the fact that both studies used same protein force field (AMBER ff14SB) and solvent model (TIP3P), it is possible that the discrepancy we observed is due to the force field limitation.

In our unbiased MD simulation, we found the ligand is very unstable in the binding site when the target crystallographic water is absent (Figure S12D). We checked two pairwise distances between the protein and ligand and found that the distances significantly change during the simulation (Figure S12). We find that when the crystallographic water is missing, the ligand is more flexible in the binding site (Figure S12D), reflected by increased distances between the protein and ligand (Figure S12B). When the hydration site is occupied, the ligand is more stable (Figure S12C). We also observe similar flexibility of the ligand in the binding site in unbiased MD simulations in the TAF1(2) (PDB: 5I1Q) system.

4.4.8 Thrombin (PDB: 2ZFF)

In the thrombin system (PDB: 2ZFF), both unbiased MD and *grand* simulations captured the target water site (Figure 11D-E). In BLUES simulations, we also observed an occupancy of about 60% (Figure 11C) which is lower than the standard we defined for a success case (>80%). However, we found a good agreement between the calculated electron density

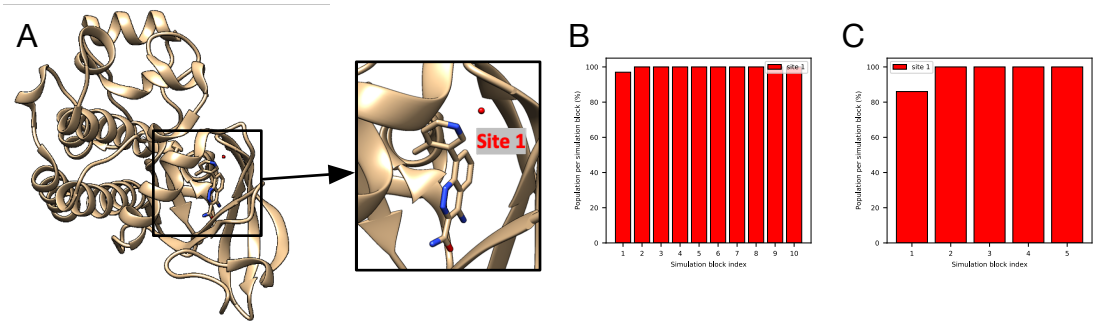


Figure 10: The BTK system (PDB: 4ZLZ) and target water sites. Bar graphs show the water occupancy of target sites in a single (B) BLUES simulation, and (C) *grand* simulation.

and experimental density map (Figure 11B), suggesting our simulations indeed recover the hydration site. So we still consider this a success case.

We noticed in BLUES simulations that this site was not occupied as frequently as we saw in unbiased MD and *grand* simulations. The main difference between BLUES and MD/*grand* simulation protocol other than the technique itself is that restraints on the protein and ligand heavy atoms were used in BLUES simulation. We then checked the distance between selected atoms between the protein and ligand (Figure 12B) and observe a correlation between the distance and the success of water insertion (Figure 12C-F). When the distance increases, it is more likely that the water can be inserted (Figure 12C,E). In contrast, when the distance drops, the likelihood of water insertion decreases (Figure 12D,F). These results suggest that additional space in the binding site is required to successfully insert the water. Thus, we find that the protein-ligand restraints used in BLUES simulations impair the performance of BLUES. We further tested this idea by using the same restraints in *grand* simulations and the probability of water insertion significantly dropped (Figure S13).

It is also important to mention that two possible protonation states of the ligand are suggested based on the pKa calculations using Chemicalize (ChemAxon product, `https://www.chemaxon.com`) at pH = 7.5 (Figure S14). In our simulations, both protonation states were stable in the binding site when no restraints were applied, meaning that we cannot tell from this data which is preferred or dominant. This is different from the case of HSP90 (PDB: 3RLP) in which only one ligand protonation state shows reasonable stability

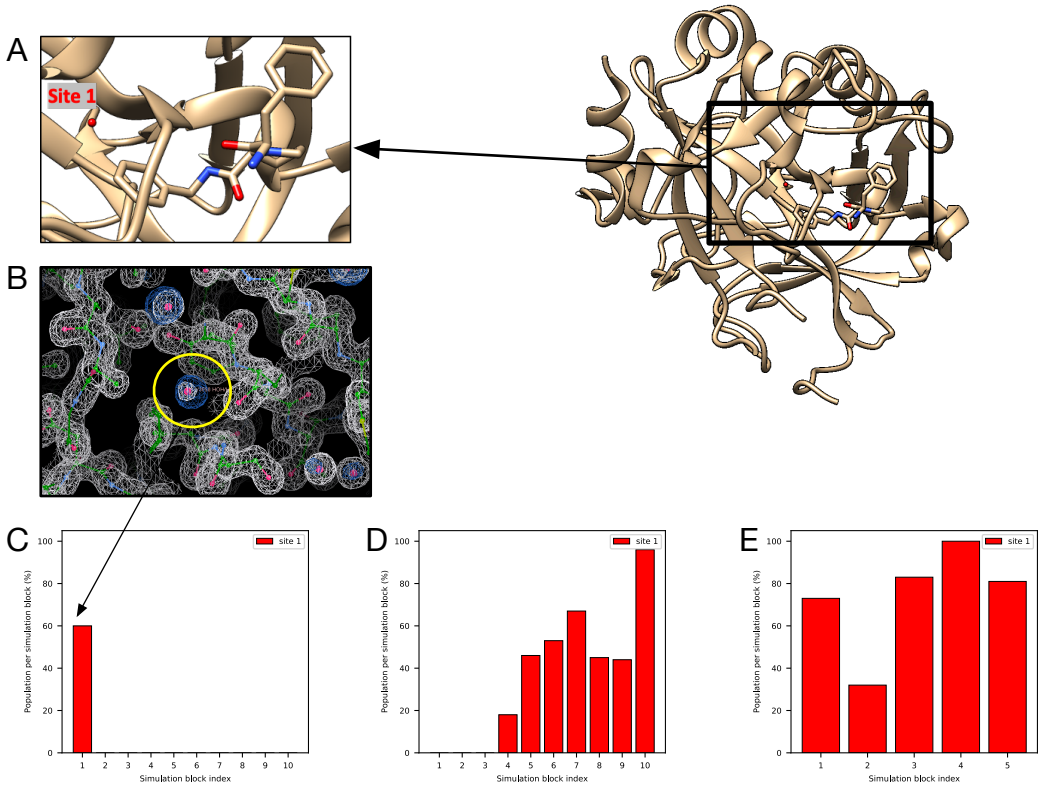


Figure 11: (A) Thrombin system (the target water site shown in red). (B) The calculated electron density map (blue) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The target hydration site is circled in yellow. The calculation is based on a BLUES simulation trajectory shown in (C). Bar graphs show the water occupancy of target sites in a single (C) BLUES simulation, (D) unbiased MD simulation, and (E) *grand* simulation.

of the ligand whereas the other one leads to unbinding of the ligand very quickly.

5 DISCUSSION

Although it is well known that water molecules can influence different biological processes (e.g., protein-ligand binding)[6,74–80] and computation is frequently used to explore such processes, systematic comparisons of water sampling techniques are infrequent. However, we believe such comparisons are important since we can only improve the methods after we learn where and how they fail.

In this work, we studied the occupancy of buried hydration sites sampled using several different simulation techniques. We studied a range of protein-ligand systems, most of which

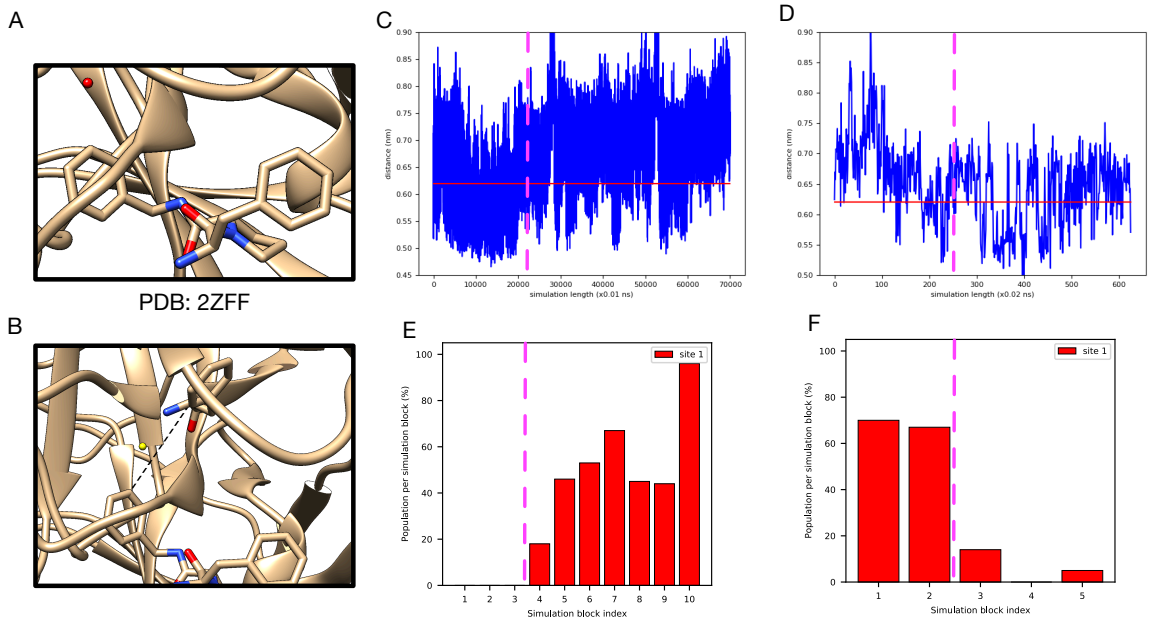


Figure 12: (A) Thrombin system (the target water site shown in red). (B) The atoms selected to compute distance between the protein and ligand. The distance change during (C) unbiased MD and (D) *grand* simulations (no restraints used). Bar graphs show the water occupancy of target hydration site in a single (E) unbiased MD and (F) *grand* simulation (no restraints used).

have hydration sites which vary their occupancy as different congeneric ligands bind.

One important lesson we learned from this work is that neither the clustering-based analysis nor electron density map analysis alone can adequately capture a complete picture of water occupancy and rearrangement in the full range of outcomes we encountered in our simulations. The use of clustering-based analysis provides actual water sites occupied by favorable water molecules in simulations which can be compared to the experimental crystal structures. In this case, occupancy information can also be obtained by calculating the frequency of favourable regions being occupied by water molecules in simulations, enabling more robust quantitative analysis.

This clustering-based approach compares the results from simulations with crystallographic water molecules which are deposited by crystallographers based on the electron density maps. However, those crystallographic waters are based on interpretations from crystallographers which cannot avoid human bias/errors.[81,82] Additionally, crystallographic

water molecules are typically deposited at 100% occupancy even if the experimental density is relatively weak and might suggest lower occupancy. This poses difficulties in directly comparing between simulation-predicted and experimental occupancy. Water occupancies are typically not refined in order to avoid overfitting, but still this limitation precludes direct comparison between simulations and experiments. This, however, is a limitation which cannot be addressed within the scope of the present work. Still it is important to keep in mind that the water molecules in the crystal structure are not always reliable, and thus differences between water patterns shown in the crystal structure and revealed in the simulation do not necessarily lead to the conclusion that something is wrong in the simulation.

Additionally, the clustering-based analysis works best when the protein/ligand are restrained. When they are not restrained, as we observed in this work, protein/ligand motions may interfere with the water patterns from clustering analysis as the water sites will change locations as the protein and ligand rearrange. Correspondingly, many water sites are returned, making it difficult to compare with crystallographic waters since it is not easy to assign water sites populated from simulations to the crystallographic sites for comparison (Figure S15). But this issue could be overcome by a better-designed analysis tool.

An alternative method is comparing calculated electron density of water molecules with the experimental electron density maps (we examine both $2F_o$-F_c and F_o-F_c). This analysis brings us one step closer to the original experimental data than does analyzing discrete water molecules in the provided structure deposited in the PDB. We find that this approach also helps with analysis of our simulation, since we are able to compare regions of significant water occupancy rather than limit our analysis to a single site with specific coordinates. However, compared to the clustering-based analysis, it requires extra work to calculate occupancies of water sites, making it less appropriate in quantitative analysis.

Another advantage of using electron density maps in the analysis is that doing so offers an opportunity to compare with the F_o-F_c map (difference map) so that differences between simulated and deposited water patterns in the crystal structure can be further analyzed.

As we mentioned earlier, both force fields and crystal structures are not perfect so it is not surprising to see water sites populated in simulations differ from those in crystal structures. Comparing the calculated electron density map with the difference map from experimental densities could help to assess simulation performance in recovering all hydration sites in the crystal structure. For example, if the water sites sampled in the simulation are in the region with positive peaks (shown as green in electron density maps), it is possible that the simulation captures the water molecules that are suggested by experimental electron density but have not been modelled by crystallographers. In contrast, if the water sites suggested by simulations are in a region with no peaks then it is possible that the force field is not accurate and placed water molecules in sites which should be devoid of water. It is notable that the complete interpretation of difference map can be more complicated and also relates to factors other than water molecules (e.g., ions, protonation states, co-solvents, etc.) but in any case, these maps provide information for additional consideration when examining discrepancies between simulations and crystal structures.

Based on our experience in this work, we suggest researchers use both approaches in the study of water sites, to allow the analysis of both specific, discrete, well-defined water sites and broader favorable regions that are occupied by water (sometimes even sporadically) in the simulation. In addition, applying two approaches allows for cross validation to ensure consistency.

The analysis we performed in this work did not consider potential biases introduced by using position restraints on the protein and ligand and compared the results directly with normal MD simulations where no restraints were used. Without restraints, our enhanced sampling methods often simply did not achieve adequate acceptance. Ideally, one would correct for the effect of these restraints such as with reweighting techniques.[83] However, our restraints here were relatively strong (10 kcal/mol $Å^{-2}$ on all heavy atoms of the protein and all atoms of the ligand). Reweighting techniques with such strong restraints would likely result in a very small number of effective samples contributing to final estimates, and thus

introduce substantial statistical uncertainty. Thus, reweighting was not employed here. But in future work with weaker restraints, reweighting techniques may be helpful to correct for the effects of restraints when computing properties like the hydration site occupancy.

Another issue making our analysis more difficult is that there is no well-established definition for successful water rehydration/sampling in simulations. The definition we used in this work is reasonable but definitely not the only possible definition. This definition is important since it may affect the assessment of different methods. Our results, however, did not show dramatic sensitivity to how we defined success (Table 2) for efficiency comparisons. We also tried to use both occupancy-based and electron density-based definitions for cross validation.

These challenges highlight that the challenging topic of water occupancy still requires more attention, both in terms of computational modeling and experimental interpretation (since crystallographic waters currently seem to be deposited only at full occupancy, even if the underlying density is weak). Based on our experience in this work, inspection of experimental crystal structures provides no clear indication as to which buried waters in the binding sites will be difficult to rehydrate in simulations. For example, the PTP1B system only has two water molecules in the binding site but poses challenges in all MD/BLUES/*grand* simulations. Both BLUES and MD simulations failed to rehydrate both water sites and *grand* is only robust when the protein and ligand are restrained to the crystallographic pose. Small modifications in a congeneric series of ligands could lower the chances for successful rehydration. Even with the same receptor (e.g., HSP90), the difficulty of rehydrating all target sites in the binding sites varies between ligands with minor structural differences (Table 2). On the experimental side, it would be more helpful if crystallographers would deposit more information on water molecules in crystal structures, such as including water occupancies in the refined model. We hope this work will draw more attention to these water-related issues so that we can improve our understanding of roles of water molecules in the active sites of the protein targets in future work.

None of the methods we studied in this work can handle water sampling perfectly although *grand* appears more robust than MD and BLUES. Even using GCMC in equilibration phase is helpful for adequate water sampling of target sites in several systems (Table 2). However, we also observed that protein/ligand motions may impair *grand* performance in water rehydration. These motions are expected in simulations when no restraints were applied to the protein/ligand but may take timescales beyond the typical free energy calculation simulation time in a single trial (e.g., > 50 ns). Restraining the protein/ligand avoided this issue in most cases we studied. Alternatively, applying restraints only in the equilibration runs and removing them in the production runs is also a way to alleviate this issue. Either way requires prior knowledge of the simulated structure with target hydration sites occupied (e.g., crystal structures, docked poses or homology models) which may not be always available in blind challenges or in a discovery setting. Moreover, our results on one thrombin system (PDB: 2ZFF) suggest protein/ligand flexibility is sometimes necessary for successful water rehydration attempts (Figure 12) to allow response to water insertion. Unfortunately, such information may not be available in advance, impairing *grand*'s predictive power.

BLUES enhances water sampling relative to normal MD but appears less efficient than *grand* (Table 1,2). In the BLUES protocol used in this work, we deployed 3000 iterations of NCMC moves in a single simulation block, accumulating 18M force evaluations (including both MD and NCMC steps) which is equivalent to 12 ns simulation time. In *grand*, a typical single run (1.4M force evaluations, 2.5 ns) performs 125000 GCMC moves in which each GCMC move attempts to insert/remove a water molecule in to the site. This is about 42 times more attempts than BLUES (3000 attempts) in a single run in this work. The difference between the protocols of BLUES and *grand* in this work is due to the fact that *grand* performs instantaneous water insertion/deletion through GCMC moves but BLUES alchemically turns off/on the interactions of the water molecule with its surrounding environment before and after translating it to a new location. Thus, for one water insertion attempt, BLUES is more expensive than GCMC which explains the performance differences between BLUES

and *grand* (Section 4.2, Table 2). Additionally, *grand* applies GCMC moves during the equilibration phase and can help water sampling in target sites (Table 2) whereas BLUES only runs normal MD.

In theory, BLUES has potential in rehydrating water sites in the binding site where protein sidechain reorientation is required for successful attempts where instantaneous insertion of water molecules by *grand* may fail due to atomic clashes. It is notable that current performance of BLUES relies on the use of restraints on the protein/ligand which keeps the protein cavities from the protein cavities from quickly collapsing. But in the future work we could extend BLUES to allow more complex moves, such as a combination of sidechain rearrangement and water hopping moves so that there is no need to restrain the whole protein/ligand but only regions where are not part of the binding/hydration target sites. But this is more appropriate when prior knowledge of the system (e.g., binding/hydration site location, sidechain/ligand motions) is available.

Normal MD simulations encountered difficulties in rehydrating each target site in most of the systems studied here. Even in those cases which were successful, MD simulations were more expensive than *grand* and BLUES. Using restraints in MD simulations does not typically help to insert water molecules to the buried sites. It is interesting, though, that MD simulations can rehydrate all target sites in a HSP90 system (PDB: 3RLQ) where both BLUES and *grand* failed (Figure 4). It is possible that running longer simulations using BLUES and *grand* will ultimately rehydrate this site given the fact that the same force field and simulation engine were used as normal MD simulations. However, this particular site (Site 1) among all target sites is still questionable as we discussed in Section 4.3.2 and its existence may not be consistent with the experimental electron density.

In most systems studied here, the hydration sites stayed occupied (100%) in the simulation after the water was successfully inserted, suggesting they are highly favorable with the force field. Ideally, we would obtain water site occupancy estimates from simulations with reversible transitions of water molecules into and out of such sites. However, for highly

favorable hydration sites, such transitions were not observed in either BLUES or *grand* simulations. One way to solve this issue could be to perform more selective move proposals so that more sampling can be focused on selected regions (e.g., target water sites) instead of a broadly defined spherical region as it is in the current BLUES settings. One way to test this idea is to combine the latest move type in BLUES, molecular darting moves[57] (moldarting), with current water hopping moves. That is, we can identify regions where water is favorable in simulations. Then, we can use moldarting to propose NCMC moves between these regions for enhanced sampling to obtain more reliable estimate of hydration site populations.

6 CONCLUSION

In this work we assessed MD/BLUES/*grand* performance in water sampling using a range of protein-ligand systems. Our results suggest both BLUES and *grand* enhance water sampling relative to normal MD, and *grand* is more robust than BLUES. The lessons we learned about these methods may help the broader community and point to further opportunities for improvement. We also discussed what we learned about each system studied in this work and hopefully these insights are useful for future work on these systems. We also highlighted issues in analyzing water sampling, and we hope that this work will draw more attention to this topic.

7 ACKNOWLEDGEMENTS

D.L.M. appreciates financial support from the National Institutes of Health (R01GM108889 and R01GM132386). DLM and YG also appreciate financial support from XtalPi. We appreciate the Open Force Field Consortium for its support of the Open Force Field Initiative, which provided software infrastructure used in this work. M.E.W. was supported by the Exascale Computing Project (No. 17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration, and

the University of California Laboratory Fees Research Program (No. LFR-17-476732). The author also appreciates the insightful discussions with Ido Ben-Shalom, Gregory A. Ross and Matteo Aldeghi for previous work done on some of the systems studied in this work. The author thanks help of BLUES package usage and simulation preparation from Samuel Gill, Teresa Danielle Bergazin, Léa El Khoury and Jeffrey Wagner.

8 ASSOCIATED CONTENT

Supporting Information Available

Supporting information is available free of charge via the Internet at `http://pubs.acs.org`.

Supporting tables of parameters used in *grand* simulations; supporting figures of water occupancy of target sites and experimental/calculated electron density maps.

Input files for simulations and scripts for analysis are freely available at `https://github.com/MobleyLab/water_benchmark_paper`.

Simulations were performed using the open-source package BLUES (v0.2.4, `https://github.com/MobleyLab/blues`), OpenMM (v7.4.2, `https://github.com/openmm/openmm`), and *grand* (v1.0.0 and v1.0.1, `https://github.com/essex-lab/grand`).

Analysis was performed using Mdtraj (v1.9.4, `https://github.com/mdtraj/mdtraj`), *grand* (v1.0.0 and v1.0.1, `https://github.com/essex-lab/grand`), CCTBX (v2021.1, `https://github.com/cctbx/cctbx_project`), LUNUS (`https://github.com/mewall/lunus`), Phenix (v1.91.1, `https://www.phenix-online.org`), Coot (v0.9.4, installed with Phenix), CCP4 (v7.1, `https://www.ccp4.ac.uk`).

9 Notes

D.L.M. is a member of the Scientific Advisory Board of OpenEye Scientific Software and an Open Science Fellow with Silicon Therapeutics.

References

(1) Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **2008**, *108*, 74–108.

(2) Levy, Y.; Onuchic, J. N. Water Mediation in Portein Folding and Molecular Recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415.

(3) Ernst, J. A.; Clubb, R. T.; Zhou, H.-X.; Gronenborn, A. M.; Clore, G. M. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* **1995**, *267*, 1813–1817.

(4) Meyer, E. Internal Water Molecules and H-Bonding in Biological Macromolecules: A Review of Structural Features with Functional Implications. *Protein Sci.* **1992**, *1*, 1543–1562.

(5) Nittinger, E.; Schneider, N.; Lange, G.; Rarey, M. Evidence of Water Molecules—A Statistical Evaluation of Water Molecules Based on Electron Density. *J. Chem. Inf. Model.* **2015**, *55*, 771–783.

(6) Baron, R.; Setny, P.; McCammon, J. A. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* **2010**, *132*, 12091–12097.

(7) Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016**, *116*, 7673–7697.

(8) Park, S.; Saven, J. G. Statistical and Molecular Dynamics Studies of Buried Waters in Globular Proteins. *Proteins* **2005**, *60*, 450–463.

(9) Takano, K.; Yamagata, Y.; Yutani, K. Buried Water Molecules Contribute to the Conformational Stability of a Protein. *Protein Eng. Des. Sel.* **2003**, *16*, 5–9.

(10) Lu, Y.; Wang, R.; Yang, C.-Y.; Wang, S. Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.

(11) Denisov, V. P.; Halle, B.; Peters, J.; Hoerlein, H. D. Residence Times of the Buried Water Molecules in Bovine Pancreatic Trypsin Inhibitor and Its G36S Mutant. *Biochemistry* **1995**, *34*, 9046–9051.

(12) Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017**, *117*, 10694–10725.

(13) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937.

(14) Maurer, M.; de Beer, S.; Oostenbrink, C. Calculation of Relative Binding Free Energy in the Water-Filled Active Site of Oligopeptide-Binding Protein A. *Molecules* **2016**, *21*, 499.

(15) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.

(16) Schindler, C. E. M.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M. K. I.; Follows, B.; Fuchß, T.; Grädler, U.; Gunera, J.; Johnson, T.; Jorand Lebrun, C.; Karra, S.; Klein, M.; Knehans, T.; Koetzner, L.; Krier, M.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Schulz, R.; Steinbrecher, T.; Tanzer, E.-M.; Unzue Lopez, A.; Viacava Follis, A.; Wegener, A.; Kuhn, D. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.

(17) Crawford, T. D.; Tsui, V.; Flynn, E. M.; Wang, S.; Taylor, A. M.; Côté, A.; Audia, J. E.; Beresini, M. H.; Burdick, D. J.; Cummings, R.; Dakin, L. A.; Duplessis, M.; Good, A. C.; Hewitt, M. C.; Huang, H.-R.; Jayaram, H.; Kiefer, J. R.; Jiang, Y.; Murray, J.; Nasveschuk, C. G.; Pardo, E.; Poy, F.; Romero, F. A.; Tang, Y.; Wang, J.; Xu, Z.; Zawadzke, L. E.; Zhu, X.; Albrecht, B. K.; Magnuson, S. R.; Bellon, S.; Cochran, A. G. Diving into the Water: Inducible Binding Conformations for BRD4, TAF1(2), BRD9, and CECR2 Bromodomains. *J. Med. Chem.* **2016**, *59*, 5391–5402.

(18) Woodhead, A. J.; Angove, H.; Carr, M. G.; Chessari, G.; Congreve, M.; Coyle, J. E.; Cosme, J.; Graham, B.; Day, P. J.; Downham, R.; Fazal, L.; Feltell, R.; Figueroa, E.; Frederickson, M.; Lewis, J.; McMenamin, R.; Murray, C. W.; O'Brien, M. A.; Parra, L.; Patel, S.; Phillips, T.; Rees, D. C.; Rich, S.; Smith, D.-M.; Trewartha, G.; Vinkovic, M.; Williams, B.; Woolford, A. J.-A. Discovery of (2,4-Dihydroxy-5-Isopropylphenyl)-[5-(4-Methylpiperazin-1-Ylmethyl)-1,3-Dihydroisoindol-2-Yl]Methanone (AT13387), a Novel Inhibitor of the Molecular Chaperone Hsp90 by Fragment Based Drug Design. *J. Med. Chem.* **2010**, *53*, 5956–5969.

(19) Kung, P.-P.; Sinnema, P.-J.; Richardson, P.; Hickey, M. J.; Gajiwala, K. S.; Wang, F.; Huang, B.; McClellan, G.; Wang, J.; Maegley, K.; Bergqvist, S.; Mehta, P. P.; Kania, R. Design Strategies to Target Crystallographic Waters Applied to the Hsp90 Molecular Chaperone. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3557–3562.

(20) Smith, C. R.; Dougan, D. R.; Komandla, M.; Kanouni, T.; Knight, B.; Lawson, J. D.; Sabat, M.; Taylor, E. R.; Vu, P.; Wyrick, C. Fragment-Based Discovery of a Small Molecule Inhibitor of Bruton's Tyrosine Kinase. *J. Med. Chem.* **2015**, *58*, 5437–5444.

(21) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation Thermodynamic Mapping of Molecular Surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37*, 2029–2037.

(22) Beglov, D.; Roux, B. An Integral Equation To Describe the Solvation of Polar Molecules in Liquid Water. *J. Phys. Chem. B* **1997**, *101*, 7821–7826.

(23) Grant, J. A.; Pickup, B. T.; Nicholls, A. A Smooth Permittivity Function for Poisson-Boltzmann Solvation Methods. *J. Comput. Chem.* **2001**, *22*, 608–640.

(24) 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.

(25) 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.

(26) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein–Ligand Binding. *Proc. Natl. Acad. Sci. U.S.A* **2007**, *104*, 808–813.

(27) Woods, C. J.; Malaisree, M.; Hannongbua, S.; Mulholland, A. J. A Water-Swap Reaction Coordinate for the Calculation of Absolute Protein–Ligand Binding Free Energies. *J. Chem. Phys.* **2011**, *134*, 054114.

(28) Hamelberg, D.; McCammon, J. A. Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method. *J. Am. Chem. Soc.* **2004**, *126*, 7683–7689.

(29) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of Water Molecules in Protein Binding Sites. *J. Am. Chem. Soc.* **2007**, *129*, 2577–2587.

(30) Ben-Shalom, I. Y.; Lin, Z.; Radak, B. K.; Lin, C.; Sherman, W.; Gilson, M. K. Accounting for the Central Role of Interfacial Water in Protein–Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 7883–7894.

(31) Ross, G. A.; Russell, E.; Deng, Y.; Lu, C.; Harder, E. D.; Abel, R.; Wang, L. Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo. *J. Chem. Theory Comput.* **2020**, *16*, 6061–6076.

(32) Bergazin, T. D.; Ben-Shalom, I. Y.; Lim, N. M.; Gill, S. C.; Gilson, M. K.; Mobley, D. L. Enhancing Water Sampling of Buried Binding Sites Using Nonequilibrium Candidate Monte Carlo. *J Comput Aided Mol Des* **2021**, *35*, 167–177.

(33) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. Nonequilibrium Candidate Monte Carlo Is an Efficient Tool for Equilibrium Simulation. *Proc. Natl. Acad. Sci. U.S.A* **2011**, *108*, E1009–E1018.

(34) Adams, D. Chemical Potential of Hard-Sphere Fluids by Monte Carlo Methods. *Mol. Phys.* **1974**, *28*, 1241–1252.

(35) Adams, D. Grand Canonical Ensemble Monte Carlo for a Lennard-Jones Fluid. *Mol. Phys.* **1975**, *29*, 307–311.

(36) Mezei, M. A Cavity-Biased (T, V, μ) Monte Carlo Method for the Computer Simulation of Fluids. *Mol. Phys.* **1980**, *40*, 901–906.

(37) Mezei, M. Grand-Canonical Ensemble Monte Carlo Study of Dense Liquid: Lennard-Jones, Soft Spheres and Water. *Mol. Phys.* **1987**, *61*, 565–582.

(38) Bruce Macdonald, H. E.; Cave-Ayland, C.; Ross, G. A.; Essex, J. W. Ligand Binding Free Energies with Adaptive Water Networks: Two-Dimensional Grand Canonical Alchemical Perturbations. *J. Chem. Theory Comput.* **2018**, *14*, 6586–6597.

(39) Ross, G. A.; Bodnarchuk, M. S.; Essex, J. W. Water Sites, Networks, And Free Energies with Grand Canonical Monte Carlo. *J. Am. Chem. Soc.* **2015**, *137*, 14930–14943.

(40) Deng, Y.; Roux, B. Computation of Binding Free Energy with Molecular Dynamics and Grand Canonical Monte Carlo Simulations. *J. Chem. Phys* **2008**, *128*, 115103.

(41) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; III, T. C.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman., P. *Amber18*; University of California: San Francisco, 2018.

(42) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput Biol* **2017**, *13*, e1005659.

(43) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Lucas, X.; Tjanaka, B.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0, the Parsley Small Molecule Force Field. *ChemRxiv* **2020**,

(44) Wagner, J.; Thompson, M.; Dotson, D.; hyejang,; Rodríguez-Guerra, J. openforcefield/openforcefields: Version 1.2.1 "Parsley" Update. `https://doi.org/10.5281/zenodo.4021623`.

(45) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(46) 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*, 926–935.

(47) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N -log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys* **1993**, *98*, 10089–10092.

(48) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(49) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K_{a} Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.

(50) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_{a} Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

(51) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

(52) Liu, D. C.; Nocedal, J. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Program. 45*, 503–528.

(53) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. Nonequilibrium Candidate Monte Carlo Is an Efficient Tool for Equilibrium Simulation. *Proc. Natl. Acad. Sci. U.S.A 108*, E1009–E1018.

(54) Gill, S. C.; Lim, N. M.; Grinaway, P. B.; Rustenburg, A. S.; Fass, J.; Ross, G. A.; Chodera, J. D.; Mobley, D. L. Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo. *J. Phys. Chem. B* **2018**, *122*, 5579–5598.

(55) Burley, K. H.; Gill, S. C.; Lim, N. M.; Mobley, D. L. Enhancing Side Chain Rotamer Sampling Using Nonequilibrium Candidate Monte Carlo. *J. Chem. Theory Comput.* **2019**, *15*, 1848–1862.

(56) Sasmal, S.; Gill, S. C.; Lim, N. M.; Mobley, D. L. Sampling Conformational Changes of Bound Ligands Using Nonequilibrium Candidate Monte Carlo and Molecular Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 1854–1865.

(57) Gill, S. C.; Mobley, D. L. Reversibly Sampling Conformations and Binding Modes Using Molecular Darting. *J. Chem. Theory Comput.* **2021**, *17*, 302–314.

(58) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.

(59) Woo, H.-J.; Dinner, A. R.; Roux, B. Grand Canonical Monte Carlo Simulations of Water in Protein Environments. *J. Chem. Phys* **2004**, *121*, 6392–6400.

(60) Ross, G. A.; Bruce Macdonald, H. E.; Cave-Ayland, C.; Cabedo Martinez, A. I.; Essex, J. W. Replica-Exchange and Standard State Binding Free Energies with Grand Canonical Monte Carlo. *J. Chem. Theory Comput.* **2017**, *13*, 6373–6381.

(61) Wahl, J.; Smieško, M. Assessing the Predictive Power of Relative Binding Free Energy Calculations for Test Cases Involving Displacement of Binding Site Water Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 754–765.

(62) Bodnarchuk, M. S.; Packer, M. J.; Haywood, A. Utilizing Grand Canonical Monte Carlo Methods in Drug Discovery. *ACS Med. Chem. Lett.* **2020**, *11*, 77–82.

(63) Samways, M. L.; Bruce Macdonald, H. E.; Essex, J. W. Grand: A Python Module for Grand Canonical Water Sampling in OpenMM. *J. Chem. Inf. Model.* **2020**, *60*, 4436–4441.

(64) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.

(65) Lee, J. W.; Foote, R. S. In *Micro and Nano Technologies in Bioanalysis. Methods in Molecular Biolog^{TM} (Methods and Protocols)*; Ballante, F., Ed.; Humana Press: Totowa, NJ., 2009; pp 269–279.

(66) Grosse-Kunstleve, R. W.; Sauter, N. K.; Moriarty, N. W.; Adams, P. D. The Computational Crystallography Toolbox: Crystallographic Algorithms in a Reusable Software Framework. *J. Appl. Crystallogr.* **2002**, *35*, 126–136.

(67) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the *CCP* 4 Suite and Current Developments. *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 235–242.

(68) Eyck, T. F. T. Crystallographic fast Fourier transforms. *Acta Crystallogr A* **1973**, *29*, 183–191.

(69) Read, R. J.; Schierbeek, A. J. A phased translation function. *J. Appl. Crystallogr.* **1988**, *21*, 490–495.

(70) A.Immirzi, In *Crystallographic Computing Techniques*; F.R.Ahmed,, Ed.; Copenhagen: Munksgaard, 1966; p 399.

(71) Emsley, P.; Cowtan, K. *Coot* : Model-Building Tools for Molecular Graphics. *Acta Crystallogr D Biol Crystallogr 60*, 2126–2132.

(72) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of *Coot*. *Acta Crystallogr D Biol Crystallogr 66*, 486–501.

(73) Smith, C. R.; Dougan, D. R.; Komandla, M.; Kanouni, T.; Knight, B.; Lawson, J. D.; Sabat, M.; Taylor, E. R.; Vu, P.; Wyrick, C. Fragment-Based Discovery of a Small Molecule Inhibitor of Bruton's Tyrosine Kinase. *J. Med. Chem.* **2015**, *58*, 5437–5444.

(74) Pearlstein, R. A.; Hu, Q.-Y.; Zhou, J.; Yowe, D.; Levell, J.; Dale, B.; Kaushik, V. K.; Daniels, D.; Hanrahan, S.; Sherman, W.; Abel, R. New Hypotheses about the Structure–Function of Proprotein Convertase Subtilisin/Kexin Type 9: Analysis of the Epidermal Growth Factor-like Repeat A Docking Site Using WaterMap. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2571–2586.

(75) Pearlstein, R. A.; Sherman, W.; Abel, R. Contributions of Water Transfer Energy to Protein-Ligand Association and Dissociation Barriers: Watermap Analysis of a Series of P38α MAP Kinase Inhibitors: Water Transfer in Structure-Kinetic Relationships. *Proteins* **2013**, *81*, 1509–1526.

(76) Bortolato, A.; Tehan, B. G.; Bodnarchuk, M. S.; Essex, J. W.; Mason, J. S. Water Network Perturbation in Ligand Binding: Adenosine A $_{2A}$ Antagonists as a Case Study. *J. Chem. Inf. Model.* **2013**, *53*, 1700–1713.

(77) Hummer, G. Molecular Binding: Under Water's Influence. *Nature Chem* **2010**, *2*, 906–907.

(78) Setny, P.; Baron, R.; McCammon, J. A. How Can Hydrophobic Association Be Enthalpy Driven? *J. Chem. Theory Comput.* **2010**, *6*, 2866–2871.

(79) Barillari, C.; Duncan, A. L.; Westwood, I. M.; Blagg, J.; van Montfort, R. L. M. Analysis of Water Patterns in Protein Kinase Binding Sites: Water Patterns in Protein Kinase Binding Sites. *Proteins* **2011**, *79*, 2109–2121.

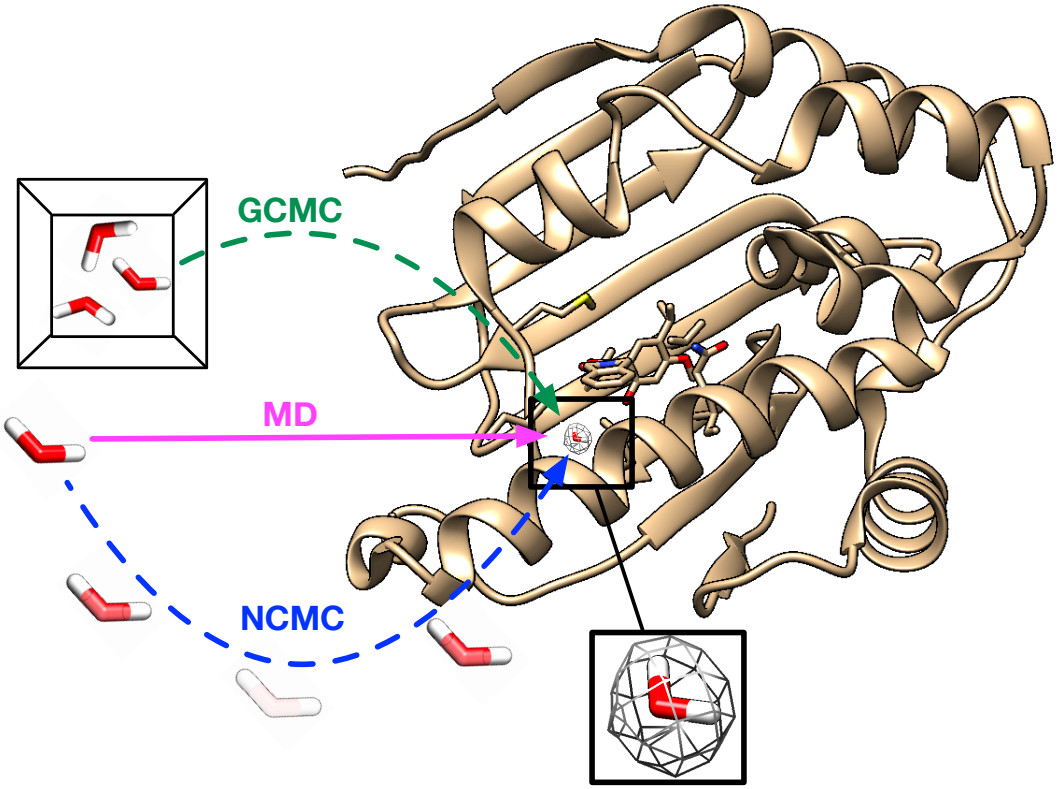
(80) Robinson, D. D.; Sherman, W.; Farid, R. Understanding Kinase Selectivity Through Energetic Analysis of Binding Site Waters. *ChemMedChem* **2010**, *5*, 618–627.

(81) Fields, B. A.; Bartsch, H. H.; Bartunik, H. D.; Cordes, F.; Guss, J. M.; Freeman, H. C. Accuracy and precision in protein crystal structure analysis: two independent refinements of the structure of poplar plastocyanin at 173 K. *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 709–730.

(82) Ohlendorf, D. H. Accuracy of refined protein structures. II.Comparison of four independently refined models of human interleukin 1beta. *Acta Crystallogr D Biol Crystallogr* **1994**, *50*, 808–812.

(83) Menzer, W. M.; Xie, B.; Minh, D. D. L. On Restraints in End-Point Protein–Ligand Binding Free Energy Calculations. *J Comput Chem 41*, 573–586.

10 For Table of Contents Only



11 Supplementary Information

Enhancing Sampling of Water Rearrangements on Ligand Binding: A Comparison of Techniques

Yunhui Ge[1], David C. Wych[1,2], Marley L. Samways[3], Michael E. Wall[2], Jonathan W. Essex[3] and David L. Mobley[1,4,*]

[1]Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA

[2]Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

[3]School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom

[4]Department of Chemistry, University of California, Irvine, CA 92697, USA

*Corresponding Author

Email: dmobley@mobleylab.org

12 Supporting Tables

Table S1: The excess chemical potential and standard state volume of water used in *grand* simulations at different temperatures for different systems. Both BLUES and MD simulations were performed at the same temperature for these systems as in *grand* simulations.

Temperature (K)	Excess chemical potential (kcal/mol)	Standard state volume (Å^3)	Systems and PDB IDs
278	-6.34	29.823	PTP1B (2QBS) HSP90 (2XAB, 2XJG) Thrombin (2ZFF) BTK (4ZLZ) TAF1(2) (5I1Q, 5I29)
286	-6.19	30.035	HSP90 (3RLP, 3RLQ, 3RLR)

13 Supporting Figures

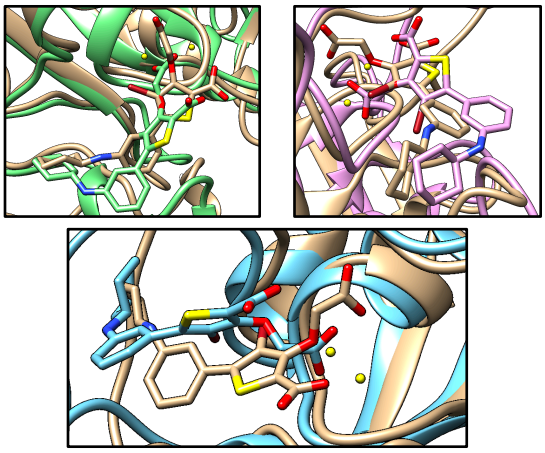


Figure S1: Snapshots extracted from unbiased MD simulations of the PTP1B system (PDB: 2QBS). The crystallographic pose is shown in tan and simulated structures are shown in green, pink and blue. The target hydration sites in the binding site are shown in yellow.

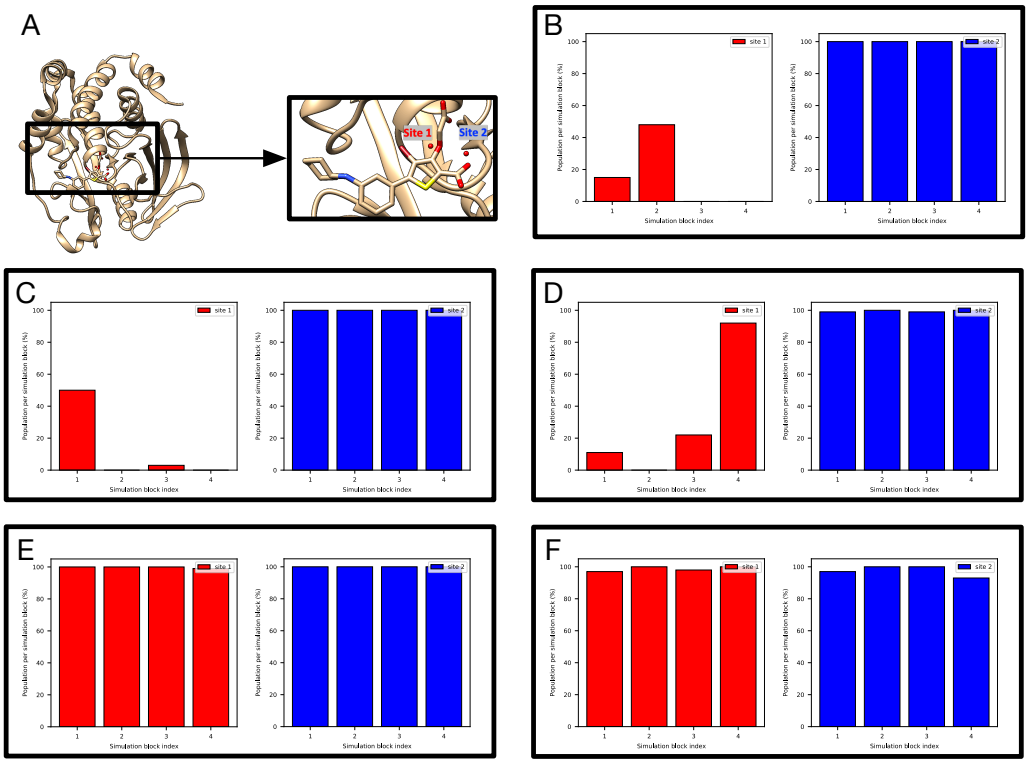


Figure S2: (A) The PTP1B system (PDB: 2QBS) and target water sites. (B-F) Bar graphs show the water occupancy of target sites in unbiased MD simulations (with ordered water molecules retained).

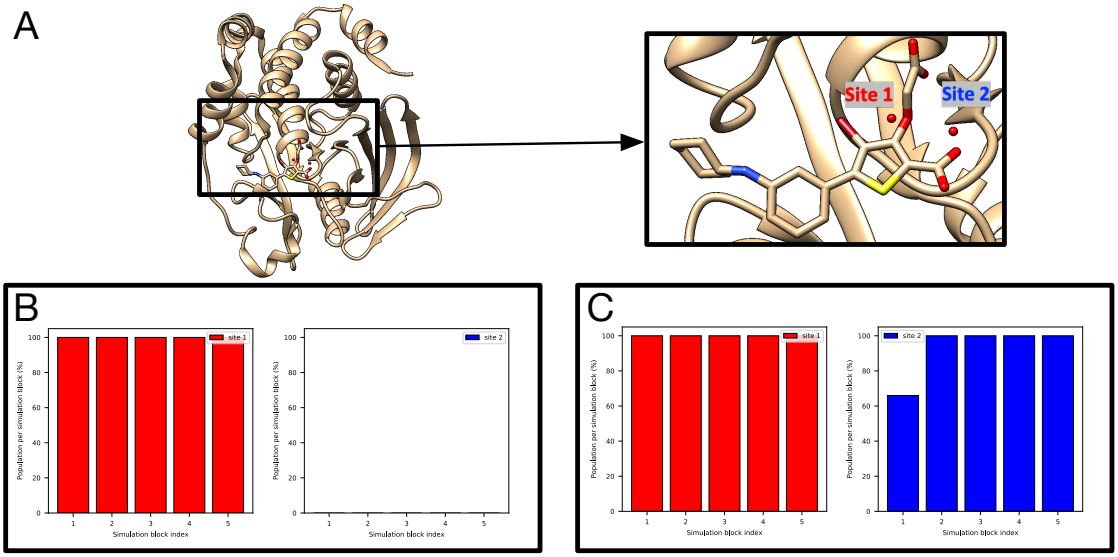


Figure S3: (A) The PTP1B system (PDB: 2QBS) and target water sites. (B-C) Bar graphs show the water occupancy of target sites in *grand* simulation (no restraints).

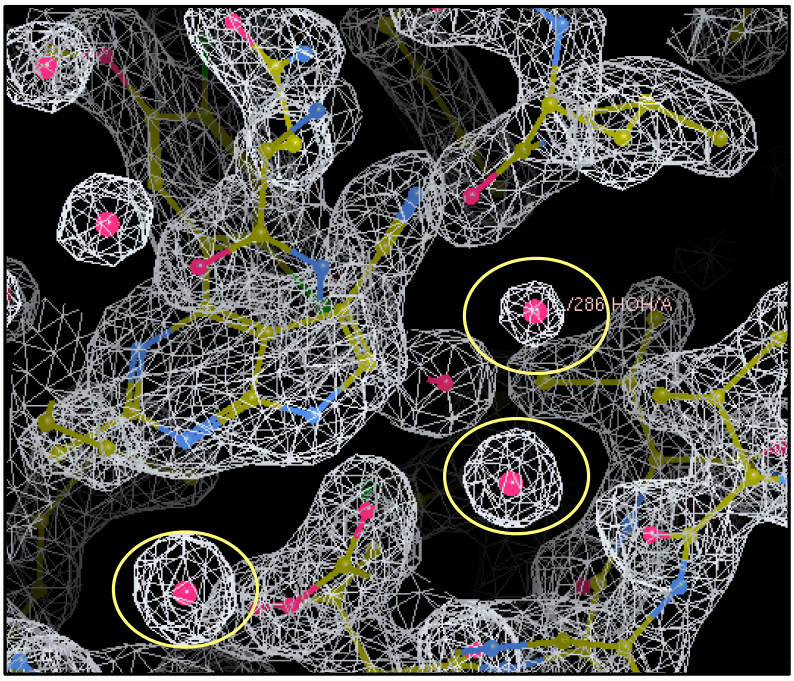


Figure S4: The experimental electron density ($2F_o - F_c$) map (white) of the HSP90 system (PDB: 3RLQ) obtained from `https://www.rcsb.org/structure/3RLQ`. The target hydration sites are circled in yellow

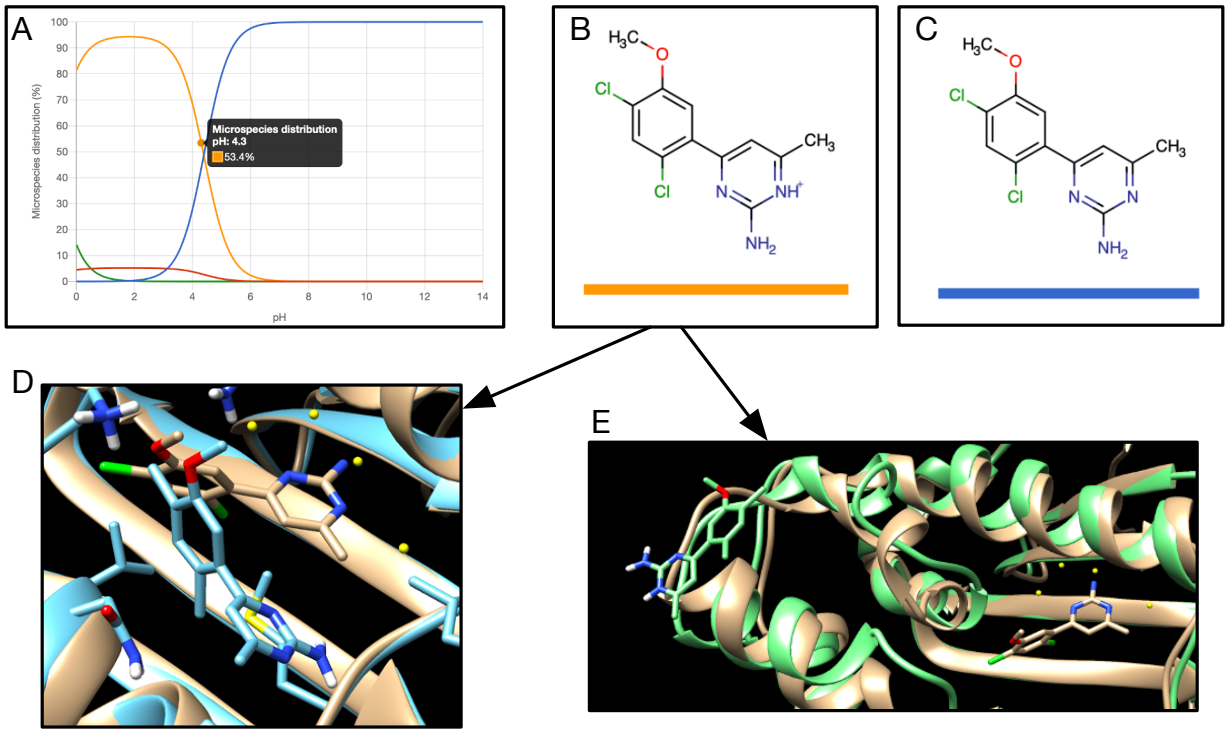


Figure S5: (A) Calculated pKa values and the microspecies distribution (in %) of the ligand in the HSP90 system (PDB: 3RLP). (B-C) Populated ligand with different protonation states at pH 4.3. (D-E) Snapshots (blue and green) extracted from simulations using the ligand with protonation state in (B) overlap with the crystallographic pose (tan). The ligand is not stable in the simulation and escapes from the binding site.

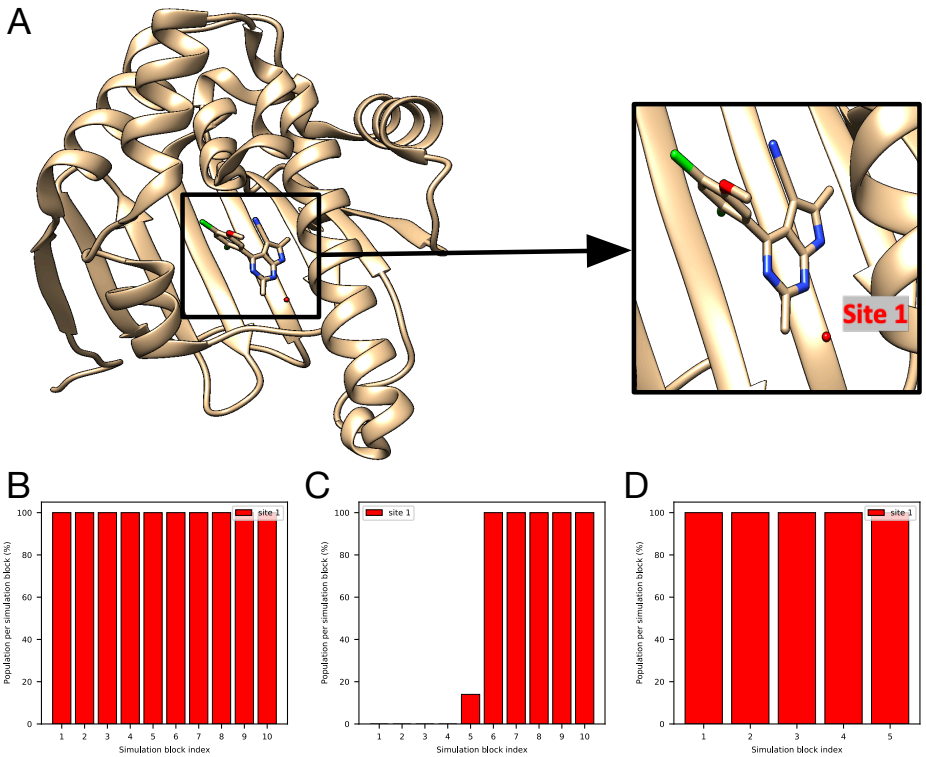


Figure S6: (A) The HSP90 system (PDB: 3RLR) and target water sites. Bar graphs show the water occupancy of target sites in a single (B) BLUES simulation, (C) unbiased MD simulation, and (D) *grand* simulation.

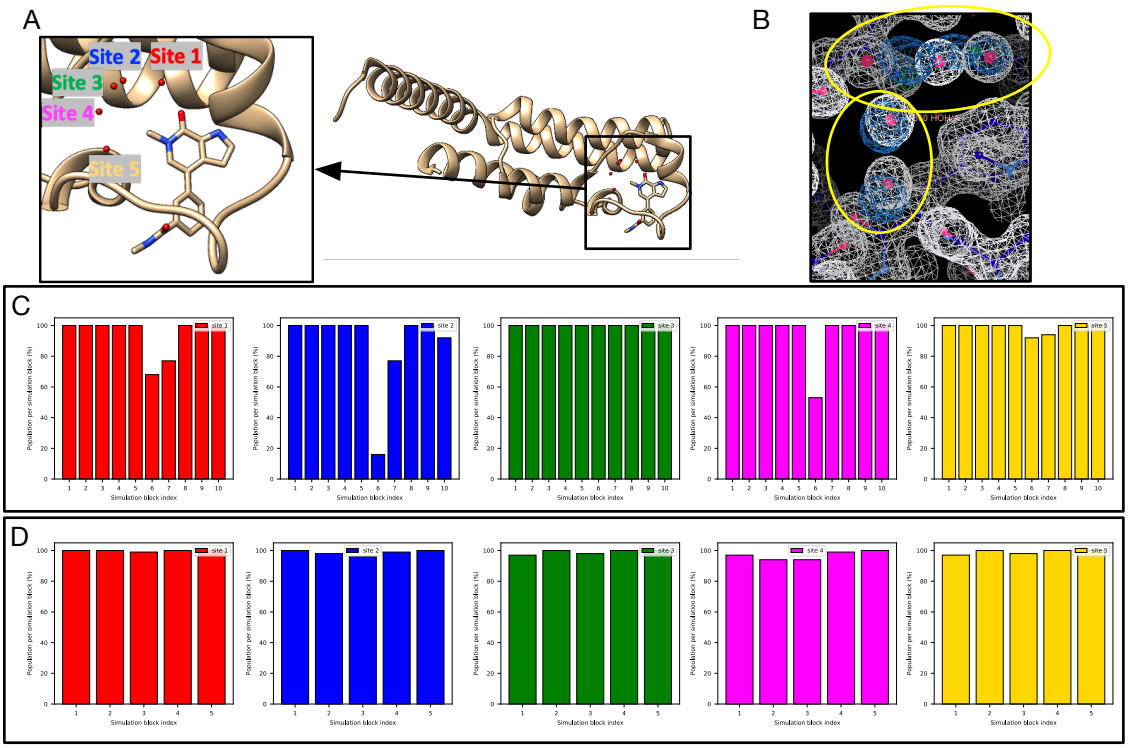


Figure S7: (A) The TAF1(2) system (PDB: 5I29) and target water sites. (B) The calculated electron density map (blue) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The target hydration sites are circled. The calculation is based on a MD simulation trajectory. Bar graphs show the water occupancy of target sites in a single (C) BLUES simulation, and (D) *grand* simulation.

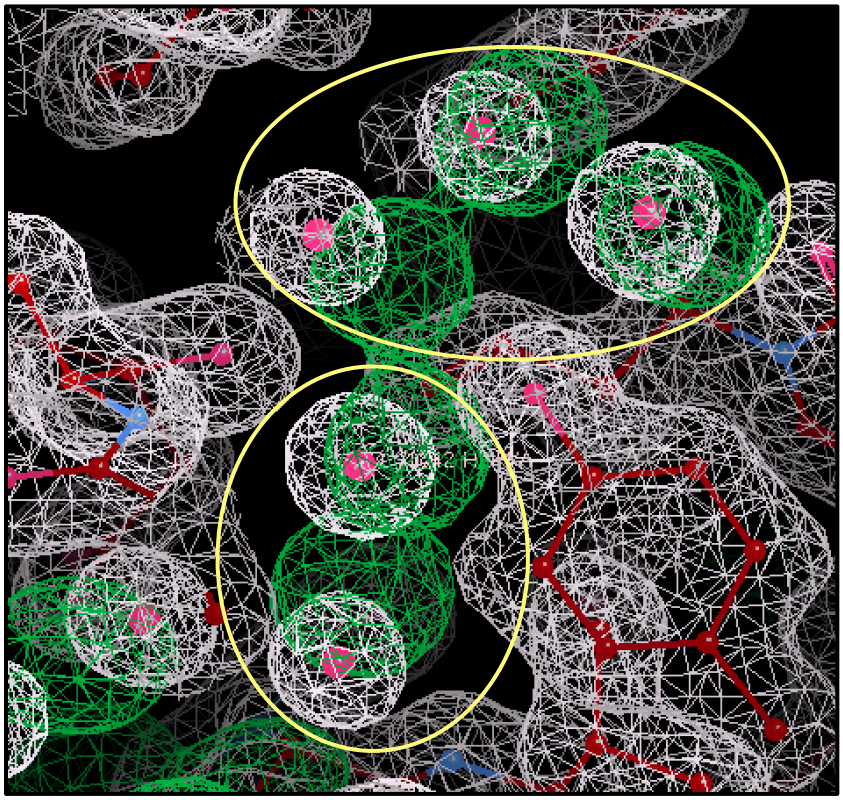


Figure S8: The calculated electron density map (green) overlaps with the experimental electron density ($2F_o - F_c$) map (white) of a TAF1(2) system (PDB: 5I1Q). The target hydration sites are circled. The calculation is based on a MD simulation trajectory.

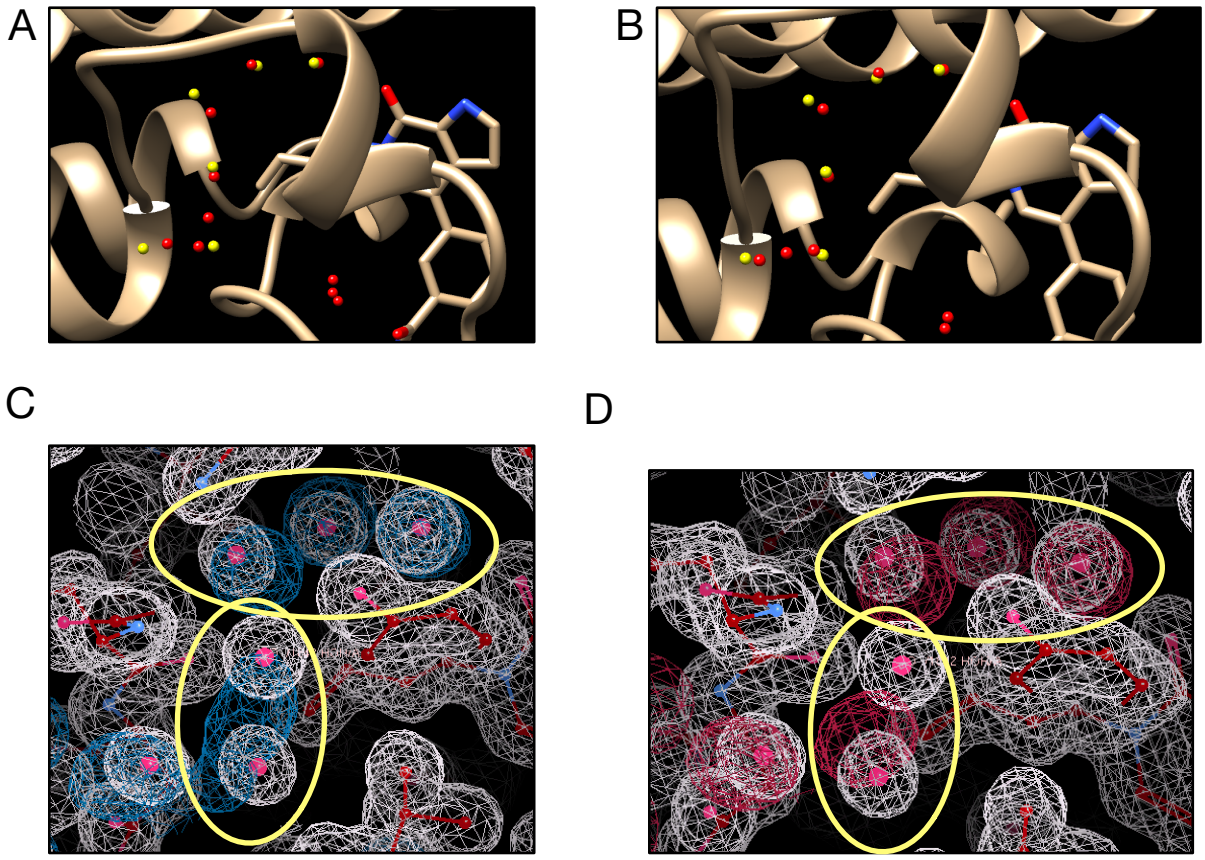


Figure S9: (A-B) Populated water sites revealed from simulations (red) of a TAF1(2) system (PDB: 5I1Q) and the crystallographic water sites (yellow). The crystallographic pose is shown in tan. (C-D) The calculated electron density map (blue, red) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The calculation is based on BLUES simulation trajectories.

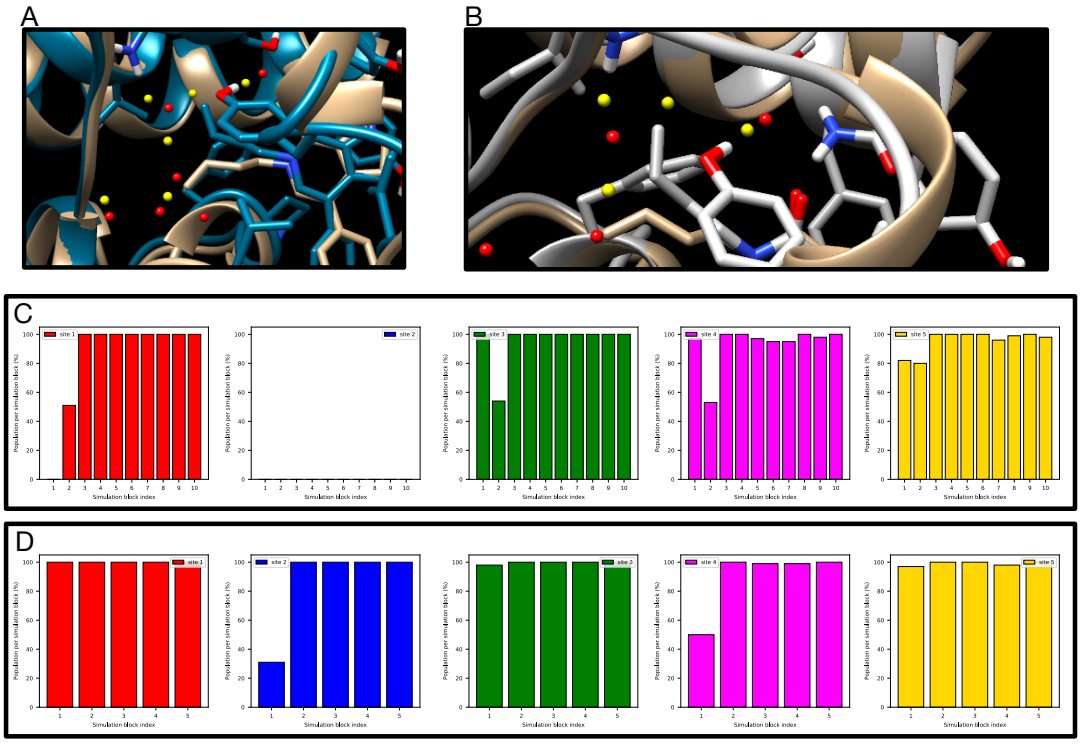


Figure S10: (A-B)Snapshots extracted from *grand* simulations of a TAF1(2) system (PDB: 5I1Q) in the absence of restraints. The crystallographic pose is shown in tan and simulation snapshots are shown in blues and white. Bar graphs show the water occupancy of target sites in a single (C)*grand* simulation, (D) *grand* simulation with restraints.

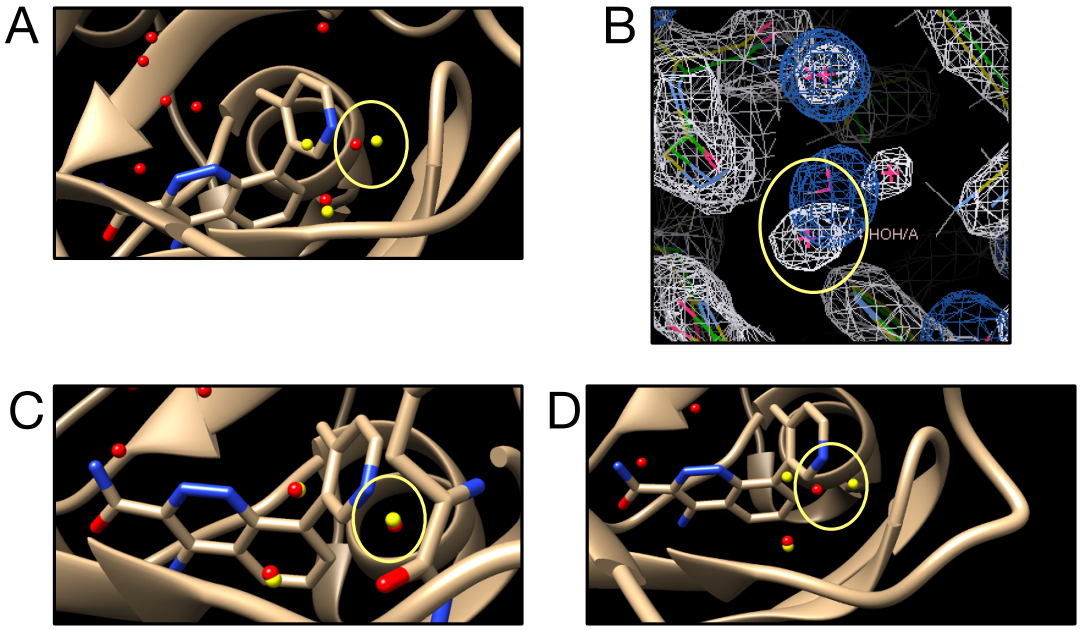


Figure S11: (A) Populated water sites (red) sampled in BLUES simulations when the crystallographic water is not present in the starting structure of the simulation of the BTK system (PDB: 4ZLZ). The crystallographic water is shown in yellow. The target site is circled. (B) The calculated electron density map (blue) overlaps with the experimental electron density ($2F_o - F_c$) map (white). The calculation is based on BLUES simulation trajectories. (C) The starting point of BLUES simulation when all crystallographic water molecules were present. (D) A snapshot extracted from simulations starting from the structure shown in panel (C). Both panel (C) and (D) use the same color code as panel (A).

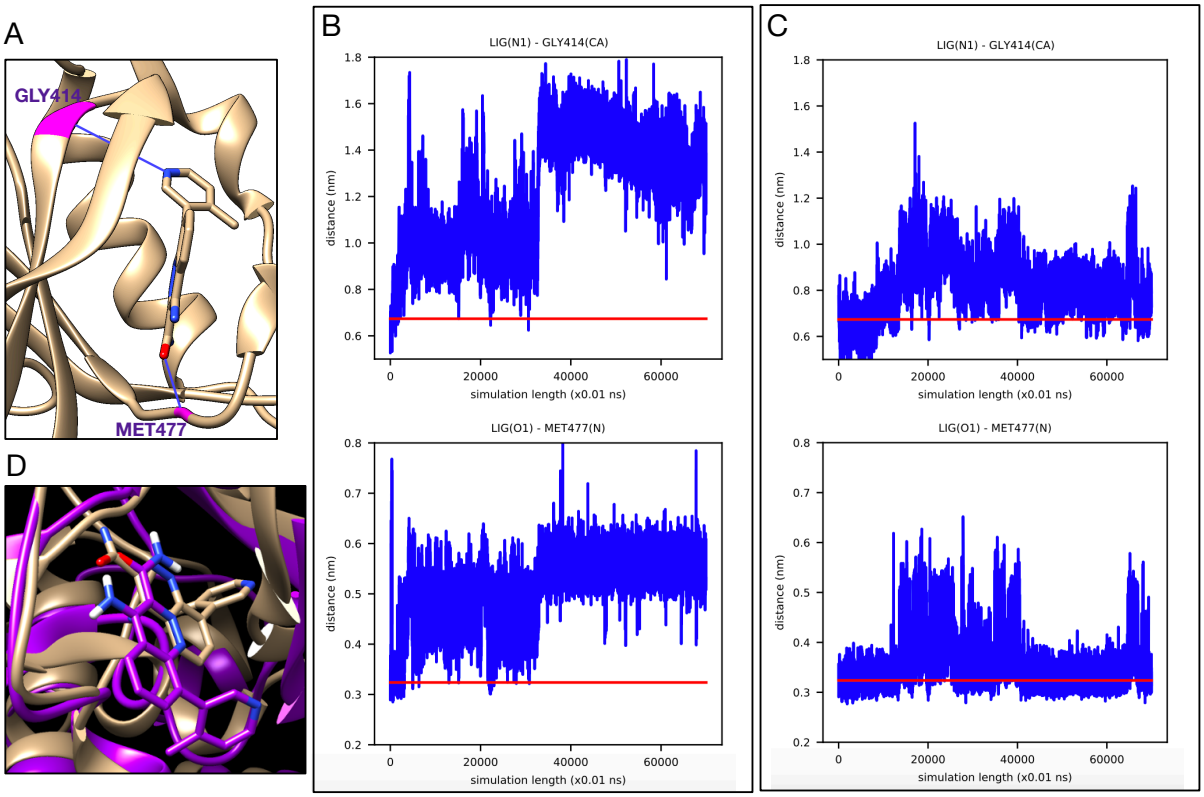


Figure S12: (A) Pairwise distances between atoms on the protein (BTK, PDB: 4ZLZ) and ligand (shown as blue lines). The distance change during MD simulations (B) in the absence of crystallographic water and (C) in the presence of crystallographic water in the starting structure. The distance measured in the crystal structure is in shown as the red line. (D) A snapshot (purple) extracted from MD simulation in the absence of crystallographic water. The crystallographic pose is shown in tan.

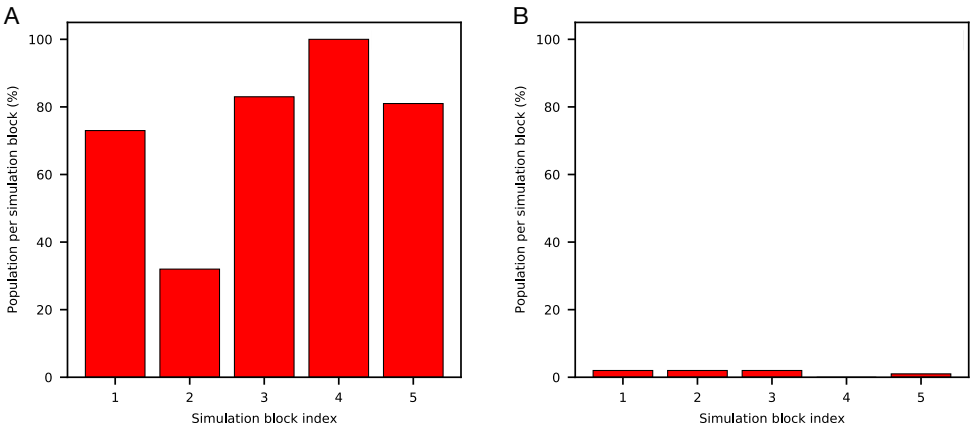


Figure S13: Bar graphs show the water occupancy of target hydration site of thrombin (PDB: 2ZFF) in a single *grand* simulation (A) with ordered water molecules removed prior to simulation (B) with ordered water molecules retained.

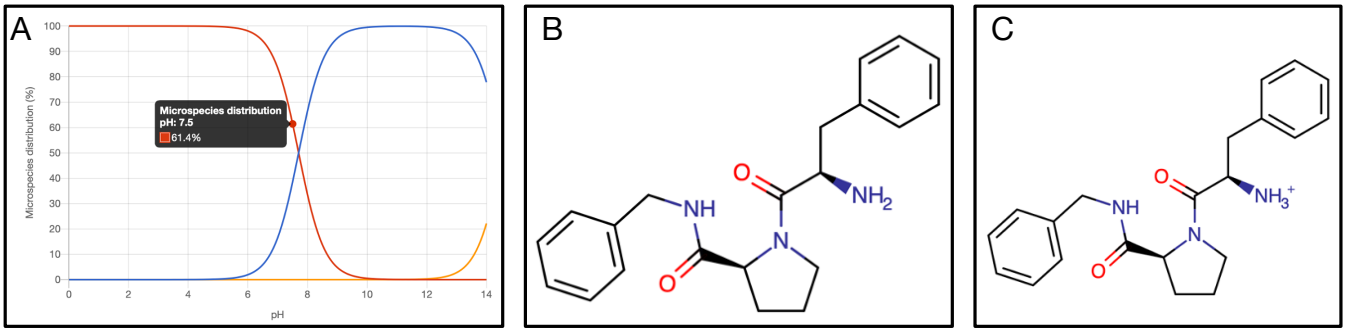


Figure S14: (A) Calculated pKa values and the microspecies distribution (in %) of the ligand in the thrombin system (PDB: 2ZFF). (B-C) Populated ligand with different protonation states at pH 7.5.

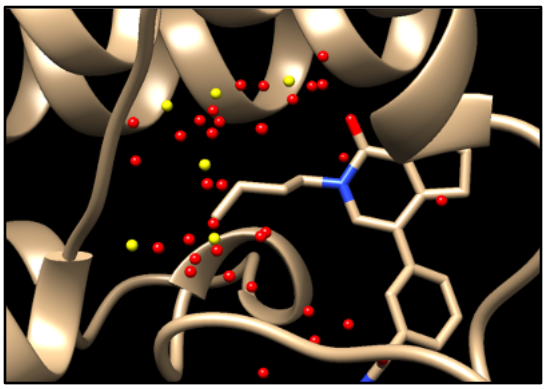


Figure S15: An example of overcrowded water sites from clustering analysis (red) using build-in functions in *grand* package. Crystallographic water molecules are shown in yellow. The crystal structure is shown in tan (TAF1(2), PDB: 5I1Q).