Sensitivity of Molecular Dynamics Simulations to Equilibration Scheme: A Case Study of Bromodomain Protein BRD4-Ligand Complex System

Yunhui Ge,[†] Gaetano Calabró,[‡] Christopher I. Bayly,[‡] and David L. Mobley^{*,†,¶}

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

[‡]*OpenEye Scientific Software, 9 Bisbee Court, Suite D, Santa Fe, NM 87508 USA*

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

E-mail: dmobley@mobleylab.org

Keywords

Molecular Dynamics simulations, Equilibration scheme, Bromodomain-containing protein 4, Ligand unbinding, Binding site unfolding

ABSTRACT

In molecular dynamics (MD) simulations, an equilibration phase is used to bring the system to the desired conditions (e.g. temperature and pressure) before collecting data in more extensive production phases. Ideally, the equilibration phase would bring the system to the correct equilibrium ensemble appropriate for the target thermodynamic conditions. Many studies give relatively little attention to details of the equilibration protocol as long as the

system eventually reaches stabilization at the correct temperature, pressure, volume, etc. However, in a previous study we found a surprising instability of two ligands in a binding site. That led us to study the origin of this instability more, and in the present work we have traced it to details of the equilibration protocol. We found for the studied system, different equilibration schemes caused dramatic differences in sampled configurations of the system in production runs, highlighting the importance of careful consideration of equilibration schemes in MD simulations.

INTRODUCTION

Molecular dynamics (MD) simulations have proven a useful tool to probe conformational dynamics at atomic-level detail with good agreement with experimental measurements.¹⁻⁵ Recent advancements in computer hardware have increased the timescales accessible to MD simulations. Meanwhile, more accurate force fields and enhanced sampling techniques have been developed, allowing for accurate modeling of different biological systems (e.g., proteins, DNA, protein-ligand interactions, etc.). MD simulations have also been employed in drug discovery projects by mainstream pharmaceutical companies given the accuracy in favorable situations.^{6,7}

While MD has had great success, this success depends on preparing a simulation which occupies and explores the relevant states of interest, and the process of preparation receives varying levels of attention. Biomolecular simulations typically take place in a largely aqueous environment, but experimental structural information (such as crystal or NMR structures) does not provide an equilibrium structure of the full system in this environment. Thus, typically, available structures (such as X-ray crystal structures) for the molecules involved are taken and “solvated” – immersed in an aqueous environment (with suitable co-solvents, counterions, etc.), followed by a process of “equilibration” which seeks to bring the resulting system to the correct conditions of interest.

In MD simulations, we seek to simulate configurations of the systems we are interested in at some specific thermodynamic state (i.e. a specified system composition at a target temperature and pressure). After preparing an initial solvated structure, the first step is normally to relax the solvated system to a low-energy state (energy minimization). Then an equilibration phase is performed to bring the system to a specific thermodynamic state before collecting data for analysis in what is called a "production" phase or run. Depending on the desired conditions for the production run, for most biomolecules, equilibration simulations can be performed under a canonical ensemble (NVT) or an isothermal-isobaric ensemble (NPT), or an NVT equilibration may precede an NPT equilibration for practical reasons.⁸

The equilibration phase serves as the initial stage of a typical MD simulation, and thus seeks to provide a good starting point to collect meaningful data for analysis during the production phase. This is especially true when computational resources are limited in that simulation times are relatively short compared to the possible timescales in the system, as is usually the case. For example, a simulation might span nanoseconds to milliseconds, but typical protein folding timescales range from μs to s ,⁹ so equilibration of proteins seeks to ensure these stay stable in their folded state. If, after the equilibration phase, the system remains unstable during what is hoped to be the production phase, some of the production data may need to be discarded as additional equilibration.¹⁰ Worse, if this lack of equilibration is not detected, results from the production phase will be biased by this additional equilibration and the associated transient behavior as we will discuss later in this work. On the other hand, if equilibration does a good job ensuring the system is in an appropriate and representative state entering the production phase, this can save a great deal of computer time and prevent analysis difficulties. One difficulty, of course, is knowing what equilibration protocol(s) are best for this task.

Thus from one perspective, equilibration is critically important for efficiency. On the other hand, one argument is that this is simply a matter of efficiency; if production data is carefully examined to discard any unequilibrated data (e.g. via automatic equilibration

detection¹⁰), careful attention to the equilibration protocol might not be needed. While this is certainly true if the system being simulated has sufficiently short correlation times and fast motions, it may not be true in biological systems where timescales for conformational change could be far longer than typical simulation times. It would not be shocking, then, if some biological systems required careful attention to the equilibration protocol to ensure that it remains near (or reaches) its true equilibrium structure while undergoing heating and other initial transient influences.

In general, there are two categories of equilibration protocols: (1) A simple scheme that just equilibrates the system briefly so that it is at the desired temperature, pressure, etc., while taking minimal action to preserve the starting structure. In this scheme it is assumed that normal sampling will allow the structural degrees of freedom to reach (or remain at) equilibrium concurrently. (2) A gentle scheme which carefully and gently equilibrates the system so as to preserve as much of the starting structure(s) (e.g., crystal structures) as possible. At some level, the first approach could be characterized as an optimistic approach, assuming that the structural degrees of freedom will be minimally perturbed away from their preferred state(s) or will be able to rapidly recover to those states. The second approach is more pessimistic, treating any experimental inputs as privileged information that ought to be retained throughout equilibration as best as possible to avoid losing it due to transient perturbations. To our best knowledge, we are not aware of a study which has analyzed the differences in results stemming from these two approaches.

In another study¹¹ where water sampling was our main focus, we consistently observed ligand unbinding events for a series of ligands bound to one of the proteins studied, though experimentally the ligands ought to have remained stably bound. This provided the main motivation for the present work, which seeks to further explore the reasons behind the unexpected instability of the system. Our results show that it was the equilibration protocol that played an important role in the instability of ligands in the binding site in this system. The results from this work highlight the importance of thorough consideration of the choice

of equilibration scheme.

METHODS

Here, we studied two BRD4(1) protein-ligand complex systems (PDB IDs: 5I80, 5I88)¹² in this study (Figure 1). Both systems were included in another study¹¹ focusing on water sampling. However, we consistently observed ligand unbinding events in the simulations of both systems and thus decided to further explore these two systems in this work. BRD4 is a potential target for cancer therapy development^{13,14} and has been previously used in several studies focusing on free energy calculations.^{15,16} In the rest of the paper, we refer to these two protein-ligand complexes as 5I80 and 5I88, their PDB codes.

Figure 1: Two BRD4 systems were studied in this work.

We used two different schemes to prepare the system and equilibrate protein-ligand bound structures. In both schemes, the AMBER ff14SB force field¹⁷ was used for protein parameterization in conjunction with TIP3P water model.¹⁸ The ligand was parameterized using Open Force Field (codenamed “Parsley”¹⁹). For 5I80, the ligand was parameterized using Open Force Field version 1.2.1²⁰ for simulations using the simple scheme and version 1.3.0²¹ for simulations using the gentle scheme. For 5I88, the ligand was parameterized using Open Force Field version 1.2.1²⁰ for simulations using both the simple and gentle schemes. In both schemes, all crystallographic water molecules were retained when preparing the system for simulations. For each system (5I80, 5I88), we prepared 7 equilibrated structures using the

simple scheme and 1 equilibrated structure using the gentle scheme. For each equilibrated structure, 5 replicates were performed in the production phase for statistical analysis. This combination of eight equilibrated structures, with five replicates each, led to a total of 40 simulations for each system.

As mentioned, we employed two different equilibration schemes in this work. We will describe detailed settings for each scheme in the following sections. As discussed in the Introduction, one scheme is more optimistic and equilibrates in a simple manner; the other equilibrates gently to minimally perturb the input structure. The largest difference in practice between the two schemes is that in one scheme we applied a series of titrated position restraints on the protein and ligand and in the other no restraints were used. In the rest of this paper, we will call these two schemes the "simple scheme" (for the protocol which is a simple combination of NVT and NPT equilibration without using restraints) and the "gentle scheme" (for the protocol which gently releases restraints).

Simulation preparation and equilibration using the simple scheme.

We used `pdbfixer` 1.6 (<https://github.com/openmm/pdbfixer>) to add the missing heavy atoms to the receptor. Then, the PROPKA algorithm^{22,23} on PDB2PQR web server²⁴ 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 at pH of 6.8 which is the experimental pH condition listed on the PDB website (<https://www.rcsb.org>).

The simulations were performed using OpenMM²⁵ v7.4.2. All input files/scripts are available as part of the SI (https://github.com/MobleyLab/equilibration_brd4). The systems were first minimized until forces were below a tolerance of 10 kJ/mol, followed by 1 ns NVT equilibration (without warmup) and 10 ns NPT equilibration. 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)^{26,27} with nonbonded cutoffs of 10 Å.

Each system was simulated at 277K which is the experimental temperature listed on the PDB website (<https://www.rcsb.org>).

The last frame of the simulation was extracted and used as the starting point for longer simulations in the production phase. The total equilibration time was thus 11ns.

Simulation preparation and equilibration using the gentle scheme.

The system was prepared using Spruce TK toolkit (<https://www.eyesopen.com/spruce-tk>) on Orion platform for building missing side chains, capping chain breaks, optimization hydrogen placements, assigning partial charges. All crystallographic water molecules and excipients were also retained in the outcome structure.

One concern was that differences in the prepared structures (e.g., protonation states) before running any simulations caused the observed discrepancies. So we did a thorough check on the number of atoms for each residue in the prepared structures from different preparation schemes, and found that every residue was protonated identically. Another difference in the prepared structures using Spruce compared to our scheme was the retained excipients in the crystal structures. These excipients were organic molecules other than the ligands and helped to stabilize the system for crystallization. In the crystal structure of 5I80, there are two molecules of 1,2-ethanediol and one molecule of dimethyl sulfoxide. In the crystal structure of 5I88, there are one molecule of triethylene glycol, one molecule of glycerol, one molecule of dimethyl sulfoxide and two molecules of 1,2-ethanediol. We further tested if these excipients overstabilized the system in the simulations and we will discuss our results later in this paper. A quick summary is that we did not observe differences in the simulation in the presence or absence of the excipients.

Following preparation, we used the outcome structures in the protocol for protein-ligand MD simulation preparation on Orion platform:

Solvation The protein-ligand complex was solvated in water adding counter ions to neutralize the Complex and adding further 50 mM of Na^+ Cl^- ions.

Minimization The minimization was performed by restraining the ligand and protein heavy atoms with harmonic restraints with a force constant of $5 \text{ kcal}/(\text{mol}\text{\AA}^2)$.

WarmUp After the minimization, a NVT run of 0.01 ns ($T=300\text{K}$) was performed while restraining the ligand and protein heavy atoms using harmonic restraints with a force constant of $2.0 \text{ kcal}/(\text{mol}\text{\AA}^2)$.

After the WarmUp stage, four additional equilibration stages were conducted:

Equilibration I An NPT run of 0.01 ns ($T=300\text{K}$, $P=1\text{atm}$) was performed by restraining the ligand and protein heavy atoms by using harmonic restraints with a force constant of $1.0 \text{ kcal}/(\text{mol}\text{\AA}^2)$.

Equilibration II An NPT run of 0.02 ns ($T=300\text{K}$, $P=1\text{atm}$) was performed by restraining the ligand and protein heavy atoms by using harmonic restraints with a force constant of $0.5 \text{ kcal}/(\text{mol}\text{\AA}^2)$.

Equilibration III An NPT run of 0.1 ns ($T=300\text{K}$, $P=1\text{atm}$) was performed by restraining the ligand and protein heavy atoms by using harmonic restraints with a force constant of $0.2 \text{ kcal}/(\text{mol}\text{\AA}^2)$.

Equilibration IV An NPT run of 0.1 ns ($T=300\text{K}$, $P=1\text{atm}$) was performed by restraining the ligand heavy atoms and protein carbon alpha atoms by using harmonic restraints with a force constant of $0.1 \text{ kcal}/(\text{mol}\text{\AA}^2)$.

Unrestrained equilibration A short NPT run of 2 ns ($T=300\text{K}$, $P=1\text{atm}$) was performed as additional equilibration and no restraints were applied.

After the short equilibration run, an equilibrated structure was extracted and used for longer production runs. The total equilibration time was thus 2.24ns.

Longer production simulations.

For the simple scheme, we prepared 7 different equilibrated structures and used 2 of them for the longer production runs ($1 \mu\text{s}$ for 5I80 and $1.5 \mu\text{s}$ for 5I88). For the other 5 structures, we performed 500 ns production runs to validate what we observed in the two longer production

runs. Even though these are shorter simulation compared to the two longer simulations ($\geq 1 \mu s$), they are long enough to capture those motions we observed in longer simulations (more details later). We had one equilibrated structure using the gentle scheme and performed long production runs ($1 \mu s$ for 5I80 and $1.5 \mu s$ for 5I88). For each equilibrated structures, we simulated 5 separate trials in the production phase which yielded 80 simulations (8 equilibrated structures * 5 replicates * 2 systems) in total for 5I80 and 5I88 together.

RESULTS AND DISCUSSION

MD simulations are sensitive to equilibration schemes deployed for the BRD4-ligand complex systems studies in this work.

To monitor the motions of the receptor and ligand, we monitored the binding site RMSD and the distance of a key contact between the protein and ligand, both as a function of time. Particularly, we computed the pairwise distance between an oxygen atom on the ligand and a nitrogen atom on ASN140 (Figure 2A) which is a key interaction between the ligand and receptor, and the root mean squared deviation (RMSD) of selected receptor residues in the binding site (backbone + $C\beta$) in the binding site (Figure 2B). The crystal structures of 5I80 and 5I88 were used as reference structures in RMSD calculations. We can see in the simulations that when started from the gently equilibrated structures, the ligand was stable in the binding site and tightly bound to the receptor for all trials for both 5I80 and 5I88 (Figure 3C-D,4C-D). However, in the simulations that started from the equilibrated structures from the simple scheme, the ligand became relatively unstable in the binding site and largely lost its key contact with the protein (constituting unbinding events, or at least the beginning of such events), as reflected by dramatic changes in the calculated pairwise distance (Figure 3A-B,4A-B). The ligand of 5I88 was more unstable than that of 5I80 as we can see in both simulations, reflected by the majority of the trajectories having a distance value larger than 1 nm. We also observed faster unbinding events in 5I88 simulations than

those of 5I80 (Figure 3A-B,4A-B).

In simulations started from both the 5I80 and 5I88 structures, we also found that the binding site partially unfolded after ligand escaped from the site (Figure 3E-F,4E-F). In contrast, when the ligand was tightly bound, the binding site maintained the crystallographic conformation (Figure 3G-H,4G-H). We observed a modest RMSD change of 0.2 - 0.3 nm in one trajectory started from the gently equilibrated structure of 5I80 and 5I88, respectively (3G,4G). But no unbinding events were observed in those simulations (Figure 3C,4C).

Figure 2: We monitored pairwise distance and RMSD changes during simulations. (A) The oxygen atom on the ligand and ASN140 on the receptor forms a key interaction for binding. (B) Selected residues (shown in cyan) in the binding site for RMSD calculations: ALA80, TRP81, PRO82, PHE83, GLN84, GLN85, PRO86, VAL87, ASP88, ALA89, VAL90, LYS91, LEU92, ASN93, LEU94, PRO95, ASP96, TYR97, THR131, MET132, PHE133, THR134, ASN135, CYS136, TYR137, ILE138, TYR139, ASN140, LYS141, PRO142, GLY143, ASP144, ASP145, ILE146, VAL147, LEU148, MET149.)

The results shown in Figure 3 and 4 were surprising as we did not expect such dramatically different behaviors of the simulated system due to different equilibration schemes, nor had we observed such effects in other systems we studied previously using the same simple equilibration scheme.¹¹ Thus, to confirm that these effects were indeed due to equilibration differences, we performed additional simulations/analysis to validate our observations. One difference between the equilibrated structures prepared by our simple scheme and the gentle scheme was that all excipients (molecules other than the ligands and helped to stabilize the system for crystallization) in the crystal structures were retained in the gently equilibrated structures while they were removed in our scheme. So we performed an additional set of tests in which we checked whether these excipients, rather than the equilibration protocol,

Figure 3: The pairwise distance (blue) and RMSD (magenta) as a function of simulation time for simulations of the 5I80 ligand bound. Dark blue/magenta show the average value of the property (RMSD/distance) across all trajectories; shaded lines are values from the individual trajectories. (A,B,E,F) Simulations started from two different equilibrated structures from the simple scheme. (C,D,G,H) Simulations started from the gently equilibrated structure in the presence of excipients (C,G) and in the absence of excipients (D,H). The red line in panel A-D is the computed distance in the crystal structure. In each panel, the dark/solid line shows the average value for the computed distance or RMSD based on 5 separate trajectories for each starting structure, and shaded lines show individual values for each trajectory. Only a distance between 0 to 1 nm is shown.

were what changed the outcome of our simulations. Additionally we ran a series of additional equilibration and production simulations to ensure that our results were not simply the result of chance.

Excipients can be any of a number of different molecules, but here, in the crystal structure of 5I80, there are two molecules of 1,2-ethanediol and one molecule of dimethyl sulfoxide. In the crystal structure of 5I88, there are one molecule of triethylene glycol, one molecule of glycerol, one molecule of dimethyl sulfoxide and two molecules of 1,2-ethanediol.

To further test whether these excipients in the simulation box prepared by Spruce stabilize the receptor and ligand in the simulations, we performed simulations using gently equilibrated structures in the absence of these excipients. For each system (5I80, 5I88), 5 separate simulations were performed for 500 ns in addition to the 2 longer production sim-

Figure 4: The pairwise distance (blue) and RMSD (magenta) as a function of simulation time for simulations of the 5I88 ligand bound. Dark blue/magenta show the average value of the property (RMSD/distance) across all trajectories; shaded lines are values from the individual trajectories. (A,B,E,F) Simulations started from two different equilibrated structures from the simple scheme. (C,D,G,H) Simulations started from the gently equilibrated structure in the presence of excipients (C,G) and in the absence of excipients (D,H). The red line in panel A-D is the computed distance in the crystal structure. In each panel, the dark/solid line shows the average value for the computed distance or RMSD based on 5 separate trajectories for each starting structure, and shaded lines show individual values for each trajectory. Only a distance between 0 to 1 nm is shown.

ulations ($1 \mu\text{s}$ for 5I80 and $1.5 \mu\text{s}$ for 5I88) with the same settings. The results are shown in Figure 3D,H,4D,H. Similar to our observations in the simulations where excipients were present, we can see the ligand was tightly bound in the pocket of the receptor and maintained the crystallographic pose in these simulations. Thus, including excipients in the simulations is apparently not the reason why we observed different behaviors in two sets of simulations starting from different equilibration schemes, since our results seem independent of whether excipients are retained.

We performed additional 5 equilibration simulations using our simple scheme and then ran 5 separate simulations (500 ns) for each equilibrated structure. We sought to determine whether the effects observed above were simply random statistical fluctuations. As above, we monitored the pairwise distance and RMSD values for the binding site residues. The

results are summarized in Figure 5 and 6. In most trajectories of 5I80, our observations were similar to those simulations discussed above and the ligand and binding site were unstable in these simulations. In two simulations (Figure 5G-J), the ligand is more stable than the rest of the simulations but still the computed distance was centered between 0.4 and 0.6 nm whereas the simulations started from the gently equilibrated structures were centered at about 0.3 nm (Figure 3C-D, 4C-D). Simulations begun from the 5I88 structure all agreed with our observations in longer simulations (Figure 4) and the ligand consistently unbound very quickly in the production runs. Meanwhile, the binding site also became more unfolded (Figure 6).

Our results suggest that using the simple equilibration scheme for these systems results in structures which go on to become unstable in the production runs and sometimes even leads to unbinding events. We find that both the ligand and the protein are unstable. To assess this, we further checked the equilibration simulations and the corresponding equilibrated structures resulting from the simple scheme. In Figure 7 we can see that in the initial NVT equilibration simulations, both the 5I80 ligand and 5I88 ligand were still stable in the binding site (Figure 7A,C,E,G). But in the later NPT equilibration simulations, 5I88 became more unstable in the binding site, as reflected by increased distance between the two atoms that formed a key interaction between the ligand and receptor (Figure 7D). The RMSD value of the binding site residues also got larger and centered at 0.15-0.20 nm. The receptor of 5I80 was more stable than 5I88 (Figure 7F) but the ligand still became more flexible in the active site reflected by larger distance change seen in several trajectories (Figure 7B).

We were interested in determining how the geometry of the bound complex differs as a result of the equilibration protocol, so we checked equilibrated structures from all equilibration simulations. To do so, we computed how much each residue in the receptor deviates from its crystallographic geometry by computing a heavy atom RMSD for each residue relative to its crystallographic conformation. The resulting by-residue RMSDs are shown in Figure 8 with a color map. We found most structures from our simple equilibration scheme have

Figure 5: The pairwise distance (blue) and RMSD (magenta) as a function of simulation time of 5I80. Simulations were started from five different equilibrated structures from the simple scheme. The red line in panel A-C,G,H is the computed distance in the crystal structure. In each panel, the solid line is the averaged number of the computed distance or RMSD based on 5 separate trajectories for each starting structure. The shaded lines are computed distance and RMSD for each trajectory. Only a distance between 0 to 1 nm is shown.

larger RMSD values relative to the crystal structure, especially for binding site residues. In contrast, the equilibrated structures from the gentle equilibration scheme had much smaller RMSD values for those residues (≤ 0.1 nm). Apparently our simple equilibration scheme destabilizes the receptor and the ligand in the active site and correspondingly the simulations started from these structures showed unbinding and unfolding events of the ligand and receptor, respectively.

While in the limit of infinitely long simulations, the details of the equilibration protocol ought not to be important and protein-ligand complexes will relax to the state preferred by

Figure 6: The pairwise distance (blue) and RMSD (magenta) as a function of simulation time of 5I88. Simulations were started from five different equilibrated structures from the simple scheme. The red line in panel A-C,G,H is the computed distance in the crystal structure. In each panel, the solid line is the averaged number of the computed distance or RMSD based on 5 separate trajectories for each starting structure. The shaded lines are computed distance and RMSD for each trajectory. Only a distance between 0 to 1 nm is shown.

the force field, our simulations are not infinitely long, and the relaxation or equilibration time required may be highly system-dependent, as well as dependent on the equilibration protocol. Here, our results seem to indicate that the gentle equilibration scheme ensures that the present systems stay near the crystal structure during equilibration until their surroundings are well equilibrated. Once the surroundings are well equilibrated, the system appears to remain stable in its crystallographic conformation. In contrast, in our simple protocol (which we commonly used in the group until this study), the initial rapid rearrangement of solvent and other factors appears to introduce additional perturbations in the structure

Figure 7: The pairwise distance (blue) and RMSD (magenta) as a function of simulation time of for equilibration simulations of the 5I80 ligand bound (A,B,E,F) and 5I88 ligand bound (C,D,G,H). Dark blue/magenta show the average value of the property (RMSD/distance) across all trajectories; shaded lines are values from the individual trajectories. Panel A,E,C,G are results from NVT equilibration (1ns) and panel B,F,D,H are results from NPT equilibration simulations (10 ns). The red line in panel A-D is the computed distance in the crystal structure. In each panel, the dark/solid line shows the average value for the computed distance or RMSD based on 7 separate trajectories for each starting structure, and shaded lines show individual values for each trajectory. Only a distance between 0 to 1 nm is shown.

of the protein-ligand complex that drive it away from the crystal structure enough that this actually destabilizes the structure of the binding site. Once this destabilization occurs, we find that we never recover the bound structure, even after hundreds of nanoseconds of simulation. It is perhaps surprising that such relatively modest differences in equilibration protocol result in such lasting effects on the production simulation, and indeed we have not observed similar effects in other systems where the same simple equilibration scheme was used.¹¹ Still, this work indicates these effects occur at least occasionally and may profoundly affect the results of other studies of protein-ligand interactions.

How is the gentle equilibration scheme able to retain the canonical structure with such a short overall simulation time? This scheme runs through five distinct stages of restrained MD in a cumulative total of 0.24 ns. Since the following 2 ns unrestrained NPT simulation is identical in form to the 10 ns of the second stage of the simple scheme, the difference

in behavior must be due to this initial, brief, multi-stage protocol. There are two key assumptions behind it: the first is that, even after minimization resolves immediate short-range high gradients, the local movement in minimization leaves a lot of strain in local regions of the starting structure. Perhaps a gradual warmup of the system without restraints could allow this strain to dissipate, but that has not been explored here. The second assumption is that the starting configuration of protein and ligand is an important bias to be conserved at first; if lost due to brief fast motion as artificial strain is released, it may be difficult to return to it again in a timely fashion. Thus we use restraints to retain the starting structure for the protein and ligand, while allowing solvent and counterions (with fast relaxation times) to absorb the excess potential energy stemming from the initial high gradients. On the other hand, the restraints themselves can be artificial reservoirs of potential energy: thus the four successive equilibrations after warmup gradually decrease the restraint strength, as excess potential energy is drained out via the thermostat.

Equilibration is an important step in preparing MD simulations, but one which is not typically regarded as particularly important in terms of determining the outcome of a specific study of protein-ligand interactions. However, our findings indicate that details of the equilibration protocol may profoundly affect the results of MD-based studies, even after relatively long simulations. In our case, we were surprised to find that the equilibration scheme has such a dramatic effect on the resulting production runs to the extent that these runs differ even on whether the ligand remains stably bound to the protein. In theory, such effects will be alleviated if extremely long simulations are performed as all possible conformation of the system will be ultimately sampled correctly. However, the computational power to run such simulations may not be generally available in the real world, meaning that real-world equilibration protocols can dramatically effect conclusions from MD-based studies. While we have encountered this effect relatively seldom, and it is clearly system-dependent, we believe these results have significant implications for the field. In general, careful and gentle equilibration may be a larger and more important factor than is commonly realized, at least

Figure 8: Equilibrated structures of 5I80 (A) and 5I88 (B). The binding site is circled. The gently equilibrated structures are highlight in red and the other structures are from equilibration using our simple scheme. The averaged RMSD values for each residue are labeled using color maps (blue $\rightarrow$ white $\rightarrow$ red, 0.1 nm $\rightarrow$ 0.4 nm $\rightarrow$ 1.0 nm).

for systems like those we studied in this work.

CONCLUSION

In this work, we carefully studied the effect of equilibration schemes on the production simulations of two specific systems. Surprisingly, we consistently saw abnormal (i.e, unbinding and unfolding) events in simulations started from structures equilibrated with our simple scheme. However, when a more gentle equilibration scheme like the one used on Orion

platform was used, we observed more reasonable behavior of the system in the production simulations. We believe what we found in this work highlights the importance of careful and gentle equilibration schemes. While these effects are clearly system dependent, it is important to keep in mind that for some systems, the equilibration scheme will play a key role in determining the outcome of production simulations and thus deserves careful selection and testing.

ACKNOWLEDGEMENTS

DLM appreciates financial support from the National Institutes of Health (R01GM108889 and R01GM132386). We appreciate the Open Force Field Consortium for its support of the Open Force Field Initiative, which provided software infrastructure used in this work.

Declarations

Supplementary Information Input files for simulations, equilibrated structures and scripts for running simulations are freely available at https://github.com/MobleyLab/equilibration_brd4.

Data and Software Availability

Datasets. Input files for simulations, equilibrated structures and scripts for running simulations are freely available at https://github.com/MobleyLab/equilibration_brd4.

Software. 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>). Simulations using the gentle equilibration scheme was prepared and performed on OpenEye Scientific’s Orion molecular design platform (<https://www.eyesopen.com/orion>).

Analysis was performed using Mdtraj (v1.9.4, <https://github.com/mdtraj/mdtraj>).

Conflict of interest

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) van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. Molecular simulation as an aid to experimentalists. *Current opinion in structural biology* **2008**, *18*, 149–153.
- (2) LaConte, L. E. W.; Voelz, V.; Nelson, W.; Enz, M.; Thomas, D. D. Molecular dynamics simulation of site-directed spin labeling: experimental validation in muscle fibers. *Biophysical journal* **2002**, *83*, 1854–1866.
- (3) Peter, C.; Rueping, M.; Wörner, H. J.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. Molecular dynamics simulations of small peptides: can one derive conformational preferences from ROESY spectra? *Chemistry (Weinheim an der Bergstrasse, Germany)* **2003**, *9*, 5838–5849.
- (4) Showalter, S. A.; Brüschweiler, R. Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field. *Journal of chemical theory and computation* **2007**, *3*, 961–975.
- (5) Markwick, P. R. L.; Cervantes, C. F.; Abel, B. L.; Komives, E. A.; Blackledge, M.; McCammon, J. A. Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *Journal of the American Chemical Society* **2010**, *132*, 1220–1221.
- (6) Adams, D. *The Hitchhiker’s Guide to the Galaxy*; San Val, 1995.

- (7) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *Journal of Chemical Information and Modeling* **2017**, *57*, 2911–2937.
- (8) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* **2018**, *1*, 5957.
- (9) Zwier, M. C.; Chong, L. T. Reaching biological timescales with all-atom molecular dynamics simulations. *Current Opinion in Pharmacology* **2010**, *10*, 745–752.
- (10) Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* *12*, 1799–1805.
- (11) Ge, Y.; Wych, D. C.; Samways, M. L.; Wall, M. E.; Essex, J. W.; Mobley, D. L. Enhancing Sampling of Water Rearrangements on Ligand Binding: A Comparison of Techniques. *bioRxiv* **2021**,
- (12) 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. *Journal of Medicinal Chemistry* **2016**, *59*, 5391–5402.
- (13) Donati, B.; Lorenzini, E.; Ciarrocchi, A. BRD4 and Cancer: Going beyond transcriptional regulation. *Molecular Cancer* **2018**, *17*, 164–13.
- (14) Duan, Y.; Guan, Y.; Qin, W.; Zhai, X.; Yu, B.; Liu, H. Targeting Brd4 for cancer therapy: inhibitors and degraders. *MedChemComm* **2018**, *9*, 1779–1802.

- (15) Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *Journal of Chemical Theory and Computation* **2017**, *13*, 3260–3275.
- (16) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chemical Science* **2016**, *7*, 207–218.
- (17) 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.
- (18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.
- (19) 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**,
- (20) Wagner, J.; Thompson, M.; Dotson, D.; hyejang,; Rodríguez-Guerra, J. openforce-field/openforcefields: Version 1.2.1 "Parsley" Update. <https://doi.org/10.5281/zenodo.4021623>.
- (21) Wagner, J.; Thompson, M.; Dotson, D.; hyejang,; Rodríguez-Guerra, J. openforce-field/openforcefields: Version 1.3.0 "Parsley" Update. <https://doi.org/10.5281/zenodo.4118484>.
- (22) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treat-

- ment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
- (23) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (24) 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 Research* **2004**, *32*, W665–W667.
- (25) 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.
- (26) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \log(N)$ Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.
- (27) 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.