

# Weighted Ensemble Simulations Reveal Novel Conformations and Modulator Effects in Hepatitis B Virus Capsid Assembly

Diane L. Lynch,<sup>†</sup> Zixing Fan,<sup>‡</sup> Anna Pavlova,<sup>†</sup> and James C. Gumbart<sup>\*,†,¶</sup>

<sup>†</sup>*School of Physics, Georgia Institute of Technology, Atlanta, Georgia, 30332, USA*

<sup>‡</sup>*Interdisciplinary Bioengineering Graduate Program, Georgia Institute of Technology, Atlanta, Georgia, 30332, USA*

<sup>¶</sup>*School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, Georgia, 30332, USA*

E-mail: [gumbart@physics.gatech.edu](mailto:gumbart@physics.gatech.edu)

## Abstract

Molecular dynamics (MD) simulations provide a detailed description of biophysical processes allowing mechanistic questions to be addressed at the atomic level. The promise of such approaches is partly hampered by well known sampling issues of typical simulations, where time scales available are significantly shorter than the process of interest. For the system of interest here, the binding of modulators of Hepatitis B virus capsid self-assembly, the binding site is at a flexible protein-protein interface. Characterization of the conformational landscape and how it is altered upon ligand binding is thus a prerequisite for a complete mechanistic description of capsid assembly modulation. However, such a description can be difficult due to the aforementioned sampling issues of standard MD, and enhanced sampling strategies are required. Here we employ the Weighted Ensemble methodology to characterize the free-energy landscape of our earlier determined functionally relevant progress coordinates. It is shown that this approach provides conformations outside those sampled by standard MD, as well as an increased number of structures with correspondingly enlarged binding pockets conducive to ligand binding, illustrating the utility of Weighted Ensemble for computational drug development.

# Introduction

Significant progress has been made in experimental approaches to resolve the underlying dynamics for proteins, for example solid-state nuclear magnetic resonance (NMR),<sup>1</sup> HDX-MS,<sup>2</sup> and time-resolved cryo-EM<sup>3</sup> among others; however, transient states often remain experimentally inaccessible. While molecular dynamics (MD) simulations in principle can fully describe the dynamics, typical all-atom MD often can not reach physiologically relevant timescales. For example, it is only with specialized hardware that Ayaz et al.<sup>4</sup> observe both ligand binding and resulting conformational changes MD simulations employing hundreds of  $\mu$ s of simulation.

Hepatitis B virus (HBV) infection leads to severe liver damage and is the leading cause of liver disease<sup>5</sup> with approximately a million deaths per year.<sup>6</sup> HBV is an enveloped nucleocapsid, including an icosahedral shell structure comprising 120 core protein (Cp) homodimers (Fig. 1) containing the viral genome, which is surrounded by a membrane envelope.<sup>7</sup> The capsid contains, protects, and delivers the viral genome during the HBV life cycle. Recent work has shown that a promising approach for reducing viral replication is to target the HBV nucleocapsid with ligands that disrupt capsid self-assembly (capsid assembly modulators or CAMs) producing non-infectious particles.<sup>8,9</sup> The continued development of CAMs is an active area of research.<sup>6</sup>

The building blocks of HBV capsids are core protein (Cp) dimers, which oligomerize into early-assembly intermediates, e.g., tetramers and hexamers, and ultimately the full capsid.<sup>10,11</sup> Formation of the hexamer via the tetramer is the rate-limiting step of assembly.<sup>10-12</sup> These earlier studies,<sup>10,11</sup> supported by MD simulations,<sup>13,14</sup> posit that the dimer's conformation must undergo a shift to an "assembly active" state for productive assembly, as such capsid formation is described by a conformational selection mechanism. Environmental and other factors, such as CAMs, certain mutations, or ion concentration can alter the conformational landscape, resulting in capsid assembly modulation.<sup>10-12,15-19</sup>

Currently CAMs are distinguished based on their effects on capsid formation, where class

II compounds are those that accelerate the production of normal, yet empty capsids<sup>18,19</sup> (phenylpropenamides, e.g., AT-130), while class I CAMs generate aberrant structures such as tubes and sheets<sup>17,20</sup> (e.g., heteroaryldihydropyrimidines like GLS4). Although assembly outcomes are significantly different, surprisingly CAMs share the same binding site at the tetrameric inter-dimer interface (Fig. 1b, c). Mechanistically this points to allosteric effects,<sup>21</sup> with CAM-specific alterations in the tetramer conformational ensemble, although rational control of such effects remain elusive.

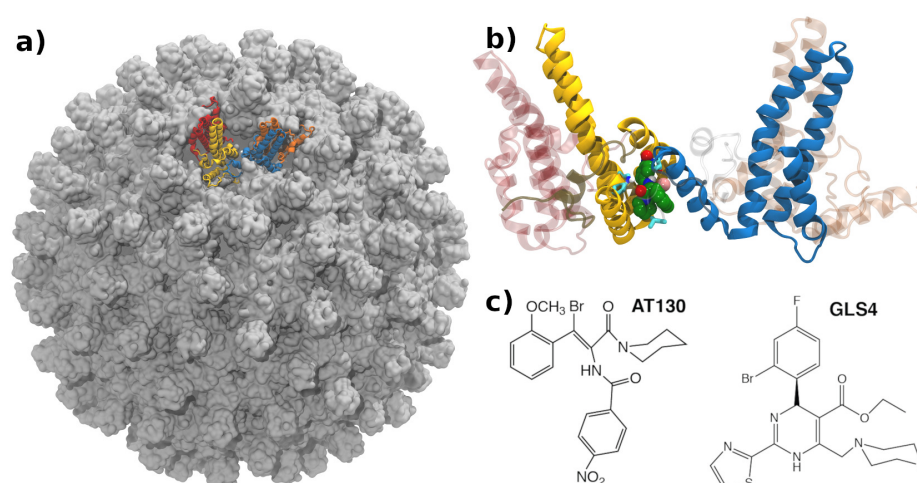


Figure 1: HBV capsid structure (PDB: 6HU4). a) The 240-unit capsid structure contains four quasi-equivalent tetramers. The ABCD tetramer, used for earlier MD studies,<sup>14</sup> is rendered in ribbon with the A, B, C, D subunits colored in red, gold, blue, and dark orange, respectively. b) The tetrameric unit with a bound CAM shown in a space-filling representation. c) Chemical structures of AT-130 and GLS4.

Given the allosteric nature of HBV capsid assembly modulation, a detailed mechanistic description for capsid assembly will require a comprehensive evaluation of CAM specific modulations of the conformational ensemble. Moreover, comprehensive docking efforts in structure based drug design (SBDD) programs commonly employ ensemble docking, where protein flexibility is approximated by using a collection of diverse protein structures.<sup>22</sup> However, the success of such endeavors is dependent on the structures comprising these collections, which can be limited in the conformational space explored in standard MD simulations.

Various approaches to more fully explore the free energy landscape have been employed

in MD applications, such as metadynamics, replica-exchange molecular dynamics, or other enhanced sampling approaches.<sup>23</sup> In addition, a straightforward approach to accelerate sampling is to employ hydrogen mass repartitioning (HMR),<sup>24</sup> allowing a larger (4-fs) time step for MD simulations. Earlier reports have established the validity of HMR in the equilibrium sampling of proteins<sup>24</sup> as well as membrane systems,<sup>25</sup> although the use of HMR for protein-ligand kinetic quantities has recently been questioned.<sup>26</sup> Intriguingly, the weighted ensemble (WE) simulation approach<sup>27</sup> has been developed for rare-event processes and rigorously describes kinetic events, although it has also been employed to effectively generate an equilibrium ensemble.<sup>28,29</sup> WE simulations offer the benefit of both enhanced sampling as well as unbiased dynamics and has been applied to a variety of systems, including  $k_{\text{on}}$  and  $k_{\text{off}}$  rates of protein-protein or protein-ligand binding,<sup>30–32</sup> protein folding events,<sup>33</sup> conformational changes,<sup>34,35</sup> drug-membrane partitioning,<sup>36</sup> as well as phase separation<sup>37</sup> and continues to be developed with an open source python-based implementation readily available.<sup>38</sup> Moreover, Xu et al.<sup>39</sup> and Hellemann and Durrant<sup>40</sup> have recently applied WE simulations for the study of ligand binding site properties in the context of ensemble docking studies.

Here, we have employed WE simulations, coupled with HMR, to explore the conformational space of apo, as well as CAM-bound, HBV tetramers. We demonstrate that WE-generated apo structures are capable of binding large HBV CAM ligands, with an increase in the number of conformations with large-volume ligand binding sites. Moreover, given the recent report of difficulties of HMR with ligand-based systems,<sup>26</sup> we report a comparison of standard time-step MD with HMR for apo and CAM bound (AT-130 and GLS4; Fig. 1c) tetrameric systems. Overall, we find that the use of HMR in equilibrium simulations does not significantly perturb these protein-ligand systems and that WE simulations provide an effective approach for producing an expanded representation of the conformational ensemble, including a collection of unique structures that are difficult to access via standard MD.

## Results and Discussion

### HMR produces only modest effects on the conformational sampling of the HBV tetramer

In the course of our work, Sahil et al.<sup>26</sup> reported sensitivity of ligand unbinding kinetics to the use of HMR. Their study illustrated that the use of HMR produces slowed ligand binding events, attributed to the effects of the longer timestep on both ligand diffusion, an effect observed earlier,<sup>25</sup> and protein fluctuations. Here we have studied apo, AT-130, and GLS4 bound HBV tetramers using both a standard 2-fs time-step as well as HMR (4-fs time step) in order to characterize the distributions observed and point to any potential artifacts introduced via the application of the longer HMR time step. Using 12 replicas for each system, each simulation lasting 500 ns, root mean square deviation (RMSD) to the starting configuration, as well as the pairwise (all to all) RMSDs, which characterize the conformational space explored, were computed. In addition for both ligand-bound simulations after protein superimposition the RMSD, using the CAM heavy atoms, were calculated.

On average, the 2-fs and HMR results display similar behavior for the systems studied, i.e., apo, AT-130 bound, and GLS4 bound. The average RMSD computed using the 2-fs and 4-fs trajectories compared to the starting structure for protein, CAM binding pocket, and ligand are within standard error (Figure S1). Although the RMSD distributions for individual replicas appear noisy (Figure S2), the overall protein distributions are quite similar (Figure S2a), while the analogous distributions for the CAM binding site using the 4-fs data appears shifted and somewhat wider than the 2-fs results (Figure S2d). Pairwise RMSDs (Figures S3 and S4) for the protein and CAM binding site were also computed, with the pairwise computation measuring RMSD over the entire set of frames, rather than deviation from just the initial structure. Again, although the individual replicas appear noisy, the overall protein distributions are quite similar (Figure S4a), while the CAM binding pocket in the 4-fs trajectories appears variable and slightly shifted (Figure S4d-f). Analogous analysis was

performed for the ligand bound AT-130 and GLS4 simulations (Figures S5-S10) with similar results to those observed for the apo simulations, i.e., the introduction of HMR produces very small alterations in the protein structures, while the CAM binding site appears more susceptible to noise. However, it is important to note that convergence of such simulations can be difficult, with multiple replicas recommended with the number dependent on the length of the trajectory as well as the system under study.<sup>41</sup> Given the goal in the present study is the exploration of the equilibrium conformational ensemble, rather than kinetic properties<sup>26</sup> of the HBV tetramer, we have used HMR in the following WE simulations.

## Effects of CAM binding on tetrameric structures

Our biophysics-based approach<sup>14,42</sup> for exploring the molecular details of CAM specific capsid assembly modulation has revealed CAM-class specific altered tetrameric conformations, which are associated with accelerated, yet normal (Class II), or aberrant capsid (Class I) assembly. Upon CAM binding, ligand-specific structural changes of the HBV dimer occur, which generate intermediates that ultimately lead to altered capsid assembly. As such CAM binding is characterized by both its binding affinity as well as the degree to which it alters the underlying tetrameric conformational ensemble.

Pavlova et al.<sup>14</sup> established that angles defining the inter-dimer orientation, i.e. the base and spike angles (Fig. 2a), are effective at discriminating the conformational preferences of the apo, as well as CAM-bound tetramers. The base and spike angles describe the opening and closing or bending of the tetramer respectively, with the base angle defined by the  $\alpha 5$  helices of the Cp monomers and the spike angle defined by the relative direction of the  $\alpha 3/\alpha 4$  helices in each dimer. From these earlier simulations it was apparent that the combination of these two angles characterized the observed ligand-dependent conformational variation in apo and CAM-bound simulations.<sup>14,42</sup> Importantly these angles provide a mechanistically relevant reduced representation for describing the CAM specific modulation of the dynamics of the tetrameric inter-dimer orientations. Additionally, the use of HMR does not alter the

average value of the base and spike angles (Figure S11).

## **Standard MD simulations of the apo HBV tetramer produce limited exploration of base and spike angles**

The HMR-based standard MD replicas described above, whose RMSD have all nearly reached a plateau (Figure S12), were used to evaluate the apo and CAM-specific base and spike angle landscape (Fig. 2b), with the angle data represented by a boundary encompassing all the data points. The class II misdirector (GLS4) base and spike angle pair are shifted to lower values and explore a more restricted region of inter-dimer orientations relative to those of the apo and AT-130 simulations, particularly the spike angle (average values and standard errors are reported in Figure S11). The apo and AT-130 results are clearly similar, and although noisy, the AT-130 simulations appear wider and reach larger base/spike values, reflecting its mechanism of action, namely assembly acceleration.

The binding site for CAM compounds, regardless of functional outcome, lies at the interface of the B and C chains in the tetrameric unit of the capsid (Fig. 1b). We have evaluated the CAM-binding-site volumes (see Methods for residues used) from the apo, AT-130, and GLS4 simulations (illustrated in Figures 2d-f) using Epock.<sup>43</sup> The distributions observed in the apo simulations are spread over a large range and reflect the flexibility of the apo binding site (Fig. 2c), while the pocket is well defined in the AT-130- and GLS4-bound simulations. Relative to AT-130, the GLS4 ligand displaying a somewhat larger available volume. The secondary maximum at lower volume in the GLS4 simulations corresponds to a set of structures where the C-terminus extends around the ligand limiting available space (see Figure S13 for an illustration). Of particular note, very few structures from the apo simulation could accommodate either ligand, highlighting a known issue with using structures from apo simulations for docking.<sup>44</sup>



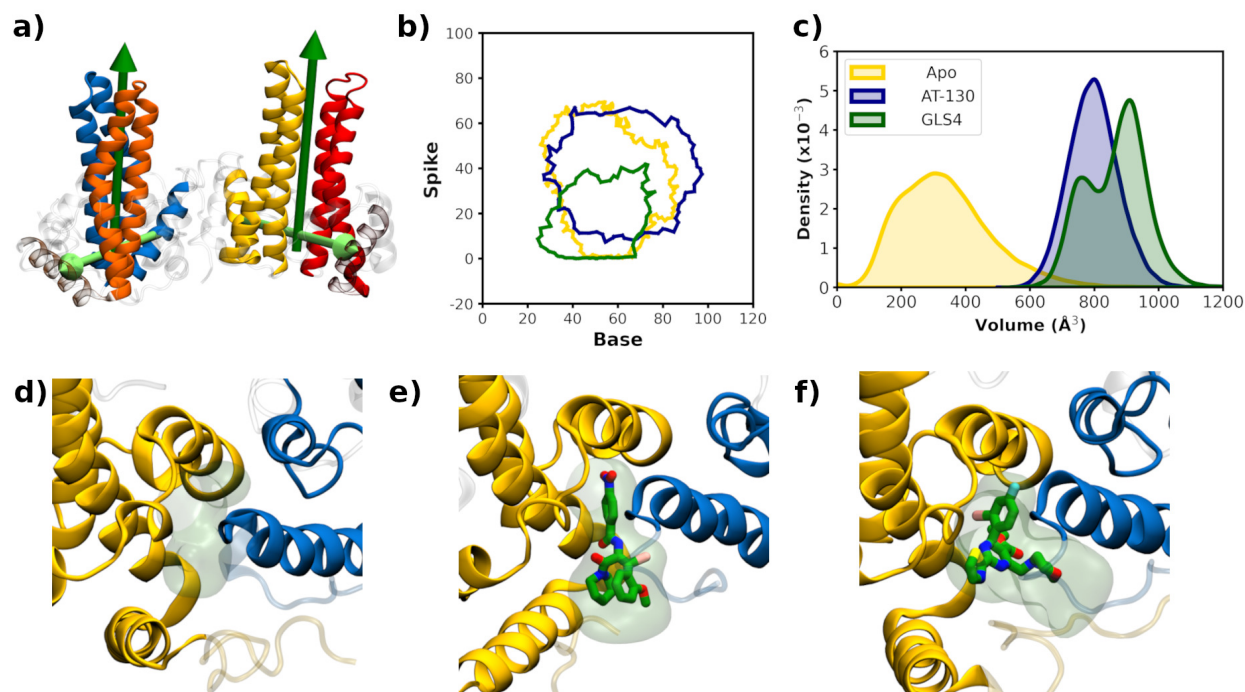


Figure 2: a) Illustration of the base and spike angles, with the vectors describing the base/spike angles rendered a light green/dark green, respectively. b) The base and spike angles for the apo (gold), AT-130 (dark blue), and GLS4 (dark green) rendered as a boundary for our standard MD simulations. c) Distribution of binding pocket volumes computed with Epack for the apo, AT-130, and GLS4 bound standard MD simulations. d)-f) Structures of the binding pocket volumes for c) apo (volume =  $287\text{\AA}^2$ ), d) AT-130 (volume =  $785\text{\AA}^2$ ), and e) GLS4 (volume =  $955\text{\AA}^2$ ). Chain B is rendered in gold ribbon, chain C in dark blue ribbon, ligand is rendered in licorice with the carbons, oxygen, sulfur, or bromine atoms green, red, yellow, or pink. The outline of the Epack-derived volume is a transparent green surface.

## WE simulations more effectively explore base and spike angles

Weighted ensemble simulations are a resampling strategy designed to produce rare-event transitions, without the addition of applied forces, along an appropriately chosen progress coordinate or coordinates<sup>27</sup> and belong to the path-sampling class of enhanced sampling approaches. In addition to steady-state conditions, where an initial state is transitioned to a predefined target state, equilibrium conditions are also possible.<sup>28,45</sup> An advantage of the latter is that a target state definition is not necessary; as such, these simulations explore the conformational space of the system in an unbiased fashion.

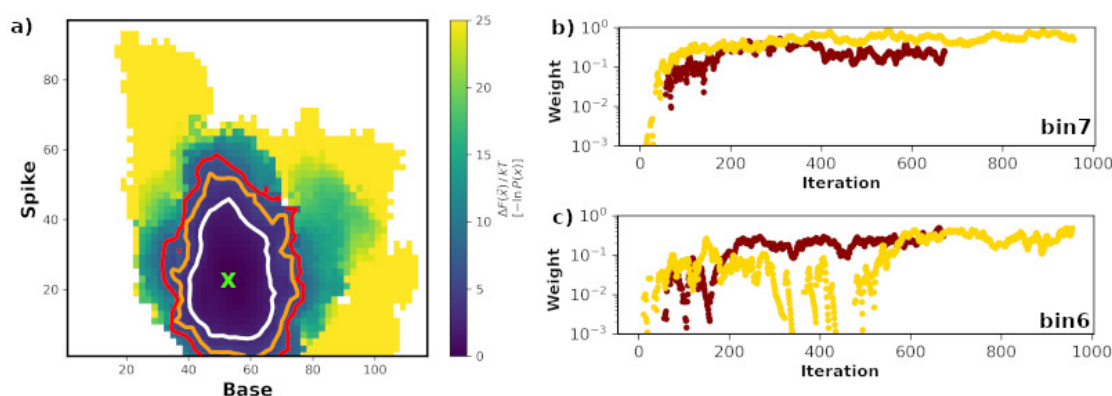


Figure 3: Apo WE simulations. a) 2D free energy landscape after combining the two replicas. Contoured at  $2.5kT$  (white),  $5.0kT$  (orange), and  $7.5kT$  (red) and removing the first 250 iterations. The initial base/spike angle pair is indicated with a green X. b) and c) Bin weights as function of WE iteration for bins 6 (base=35°-55°, spike=30°-50°) and bin 7 (base=35°-55°, spike=30°-50°).

Two replicas of WE simulations for the apo HBV tetramers were performed (6  $\mu$ s total) employing 2D progress coordinates consisting of the base and spike angles discussed above. As detailed in the methods, each angle was partitioned into 20° bins. The evolution of the base and spike angles are provided in Figure S14, while the 2D landscape generated from the combination of the two replicas is illustrated in Fig. 3a. It is apparent that although a single structure was employed to start the WE simulations, the equilibrium distribution spreads out such that a wide range of angles are sampled. In Fig. 3b and c, the weights as a function of iteration for two highly populated bins for the individual replicas are reported.

Similar to the distribution of rate constants reported by Adhikari et al.,<sup>33</sup> each replica behaves somewhat differently, although reasonable convergence of the highly populated bins is achieved. However, even for runs of 3  $\mu$ s convergence of the smaller weights appears more problematic, e.g., the magnitude of the smaller weights in bins oscillates dramatically with small changes in population (Figure S15).

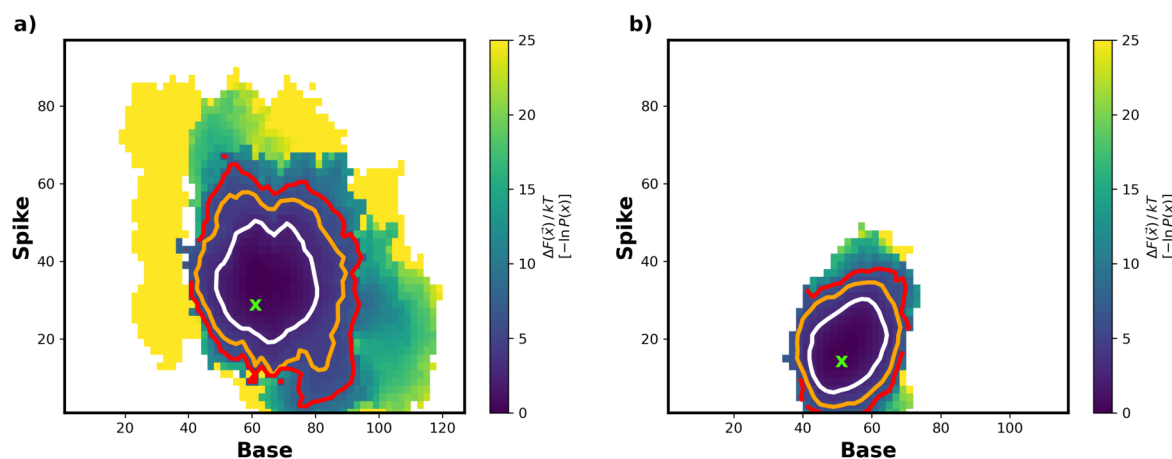


Figure 4: AT-130 and GLS4 WE 2D energy landscapes. a) AT-130 and b) GLS4 results after combining the two replicas (3  $\mu$ s for each replica). Contoured at 2.5kT (white), 5.0kT (orange), and 7.5kT (red) and removing the first 250 iterations. The initial base and spike angle pair is indicated with a green X.

Analogous WE simulations, two replicas of 3  $\mu$ s each, were performed for AT-130- and GLS4-bound tetramers, and, similar to the apo results, the major bins appear reasonably well converged. Figures S16-S19 display progress coordinate evolution, the 2D landscapes, and the weights as a function of iteration for these simulations. The resulting combined 2D landscapes for the CAM-bound simulations (Fig. 4) and progress coordinate evolution (Figures S16 and S18) reveal an altered exploration pattern for the GLS4-bound simulations relative to those for the apo and AT-130 bound systems. Similar to the apo WE results, the AT-130 bound tetramer explores a wide range of base and spike angles and is consistent with the shifting of these angles to larger values relative to the apo simulations (Fig. 3a) as observed in standard MD (Fig. 2b). Despite the fact that the current WE settings provide robust sampling along the base/spike progress coordinates for both apo and AT-130

simulations, in the case of GLS4, both replicas explore a much narrower range (Figure S18). Although these results are consistent with the observations that GLS4, a misdirector, does not readily sample the inter-dimer orientations necessary for normal capsid assembly,<sup>14,42</sup> it is likely a more aggressive re-binning during the WE simulations, or the use of the adaptive binning procedure<sup>46</sup> could provide further exploration of these inter-dimer orientations.

The WE conformational ensemble successfully produces additional conformations not observed in standard MD. To illustrate this point the base/spike region observed in the apo WE simulations is compared to standard MD in Figure 5a (both a collective 6  $\mu$ s) where, as in (Fig. 2b) the standard MD results are represented via a boundary. Again, similar results are obtained for the AT-130 WE simulations, i.e., the WE results explore regions outside those sampled in standard MD (Figure S20a), while in contrast, the GLS4-bound tetramers in the WE simulations, relative to the standard MD results, sample a comparable region of the base/spike angle space (Figure S20c).

## **Apo WE simulations explore enlarged ligand binding sites**

Holo structures are often preferred to their apo counterparts for drug discovery purposes as their ligand binding sites are already well-formed,<sup>47</sup> although such structures may be unavailable for the protein of interest. In such cases, methods to address this shortcoming include template based approaches to binding site refinement,<sup>48</sup> metadynamics to explore an enhanced sampling of the binding site,<sup>44</sup> or induced-fit procedures.<sup>49</sup> However, WE-based simulations are an attractive alternative to these approaches since the underlying dynamics remains unbiased.

Although we are focused on the CAM-specific allosteric modulation of the HBV tetramer, initial binding is still required. Therefore Epock<sup>43</sup> was used to estimate the CAM pocket volumes for the WE-generated apo conformations. In Figure 5a, the unweighted WE data is colored by pocket volume; it is noteworthy that the larger volumes correspond to the more extreme values of the progress coordinates, whereas standard MD produces few, if any,

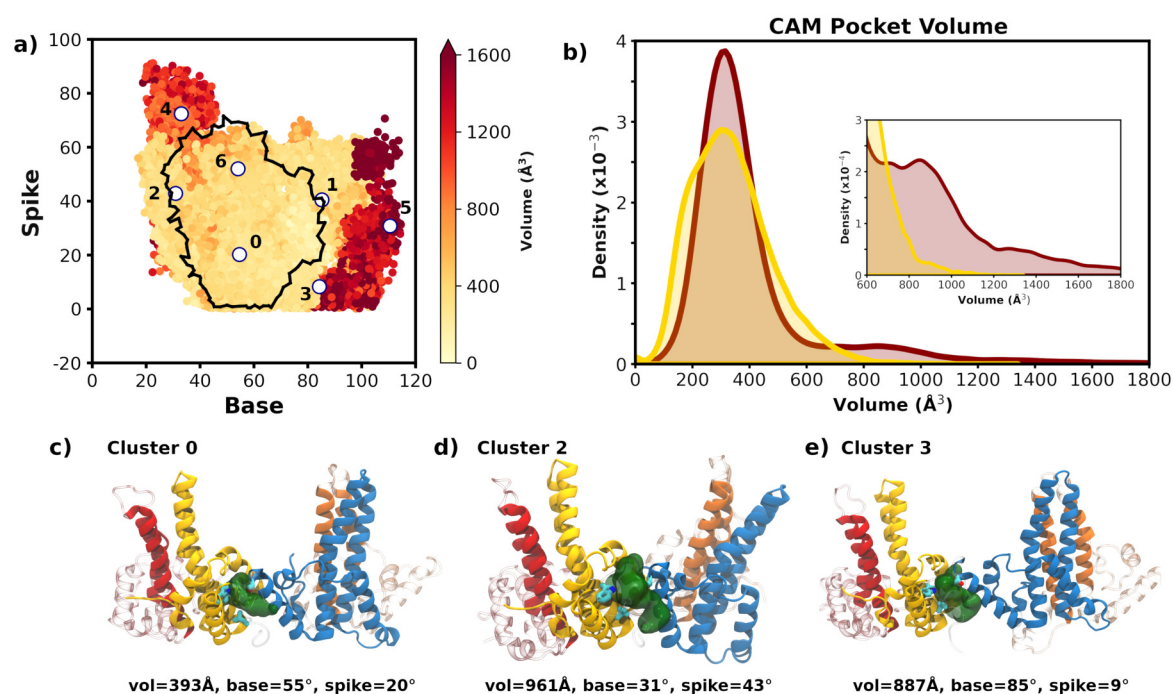


Figure 5: WE samples a wider distribution of base and spike angles. a) Scatter plot of the base and spike angle progress coordinates (unweighted). Collective data for the combined replicas is reported, with the first 250 iterations of WE simulation and the first 10 ns of standard MD data removed. For clarity, the standard MD data is rendered as a boundary (black line). The WE data is colored by the CAM pocket volume. Cluster centers are marked with white circles and are numbered. b) Histogram of the CAM pocket volumes for WE and standard MD, with the inset for the larger volumes. c) - e) Representative conformations of the tetramer structures from selected clusters showing the Epock pocket volume. W102, T128, and L140 are rendered in cyan licorice, and the volume rendered as a dark green surface.

conformations with larger volumes. Figure 5b displays the CAM pocket volume distribution for the WE replicas compared with pocket volumes from the standard MD simulations. The distributions are centered about a volume of  $\sim 350 \text{ \AA}^3$  for both the WE and standard MD runs. Although the fraction of the total number of conformations with large pocket volumes is small relative to the standard MD results, WE simulations generate an increased number of conformations with larger-than-average pocket volume (Figure 5b, inset). These structures are likely to be amenable to docking larger CAMs than those with an average ligand binding site volume. We emphasize that the progress coordinates currently used do not explicitly include information concerning the ligand binding site; larger binding site volumes are naturally produced in the course of sampling the functionally relevant base and spike angles. Given that the number of conformations needed to generate a successful ensemble docking protocol can be quite large,<sup>50</sup> our results suggest that the use of WE simulations will enhance the success rate in ensemble docking efforts. In addition to exploring the CAM binding pocket volume in the apo simulations, the AT-130 and GLS4 simulations were also analyzed (Figure S20b and c). As expected the presence of the ligand reduces the variation in the binding site volumes for both AT-130 as well as GLS4.

Clustering is often employed in ensemble docking in order to extract a small representative subset of protein conformations and is performed on global protein RMSD<sup>22</sup> or, alternatively, RMSD of the ligand binding site residues directly.<sup>51</sup> In the present case, although the binding pocket volume is required to be of sufficient size to bind a ligand, it is the functional, i.e., allosteric response of the tetramer, that is thought to drive modulation of HBV capsid assembly.

Both the WE (unweighted data) and standard MD data were clustered using hierarchical clustering, employing the base and spike angles. An average euclidean linkage<sup>52</sup> was used. Both the WE and standard MD results were downsampled, with  $\sim 25000$  structures retained at a 200-ps interval for each. As discussed by Hellemann and Durrant,<sup>40</sup> such downsampling may skew the distribution due to the merging and splitting steps in the WE protocol. How-



ever, we find that such effects are modest in the base and spike angle distributions used here (Figure S22). The WE data generated seven clusters, whose centers are labeled in Figure 5a, and span the range of base/spike angles. Representative conformations extracted from each cluster produce structures with a range of binding pocket volumes. Several of these conformations are illustrated in Figure 5c-e, labeled with their volumes, base, and spike angles, while structures from all seven clusters, along with the hierarchical dendrogram, are displayed in (Figure S23). It is noteworthy that some of the conformations with very extreme values for the base and spike angles have extremely large volumes (e.g. cluster center 5) and appear nearly dissociated. Clustering of the apo standard MD produced four clusters (Figure S24).

### Apo WE simulations provide suitable conformations for CAM docking

In order to validate that the conformations obtained from the apo simulations are conducive to docking, we extracted structures from clusters 0, 1, 2, 3, and 6 of the WE simulations and conformations from the four standard MD clusters and docked GLS4 to each. The ligand binding site volumes for the extracted conformations are reported in Table 1. GLS4 was chosen for these docking studies since it requires a significantly larger volume than that observed in our standard MD simulations (Figure 2b), thus posing a significant challenge for these apo-derived structures. Clusters 4 and 5 from the WE simulations were eliminated from this initial docking calculation as they have extreme values for the base and spike angle pairs.

Table 1: Ligand Binding Pocket Volumes ( $\text{\AA}^3$ ) for representative conformations.

Cluster Number	standard MD	WE	standard MD (extreme)
0	244	393	860
1	452	471	897
2	160	961	1099
3	295	887	702
6	-	469	-

The GLS4 ligand molecule was docked into a volume centered around the CAM binding pocket of the protein using AutoDock Vina. Not unexpectedly, due to their small binding

pocket volume, all four structures taken from the standard MD clusters failed to support ligand binding, with the ligand docked outside the pocket in all poses. However, in the case of the structures taken from the WE simulations, several produced ligand bound states. An example from cluster 2 illustrates that the pose reproduces essential interactions, i.e., the Trp102 sidechain interacts with the pyrimidine group, the ethyl-ester moiety is sandwiched between Thr109 on chain B and Val124 on chain C, and the morpholino group is proximal to Thr128.

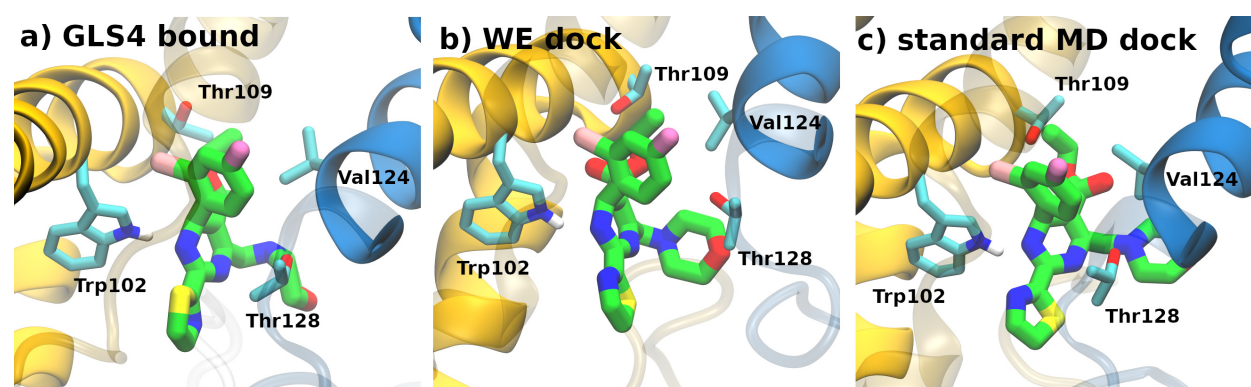


Figure 6: Docking of GLS4 to the tetramer. a) Initial conformation of the GLS4-bound structure. b) sample conformations taken from docking GLS4 into WE cluster 2 (docking score -9.3 kcal/mol) and c) structure taken from docking GLS4 into an extreme example from standard MD cluster 2 (docking score -8.3 kcal/mol).

Although we observed an enhancement of large volume structures in the WE simulations (Figure 5b inset), standard MD does produce some structures that generate expanded ligand binding sites. Structures were extracted from each of the four standard MD apo clusters that have base and spike angles significantly different from their cluster average as well as enlarged pocket volumes. Figure S25 presents snapshots of these conformations and details of the selection criterion used. Volumes for these structures are also included in Table 1. Docking of GLS4 to one of the clusters produced a pose that supports GLS4 binding (Figure 6c), similar to both the initial GLS4-bound conformation and the WE-derived docked one (Figure 6a, b). After superposition of the binding site residues, GLS4 heavy atom RMSDs to the initial conformation for the WE/standard-MD-derived structures are 2.9 Å and 2.2 Å, respectively. Given that the docking scores for these poses are both favorable ( $< -8$  kcal/mol), these



results highlight that although rare, conformations produced by WE simulations and even standard MD can accommodate large ligands such as GLS4. However, it has been noted that the number of structures necessary to generate a successful docking campaign can be considerable, while only one suitable structure was obtained with standard MD. Therefore, we contend that WE offers a reliable way to maximize the number of conformations that can be successfully employed.<sup>50</sup>

## Conclusions

CAM efficacy is not only characterized by binding affinity but also by the resulting conformational changes in early-assembly intermediates such as the tetramer. Our study demonstrates that standard MD simulations for the apo HBV Cp tetramer rarely sample conformations that display ligand binding site volumes of sufficient size for CAM binding. Furthermore these standard MD simulations produce conformations which span a restricted region of the functionally relevant HBV tetramer inter-dimer orientations as characterized by the base and spike angles. However, both of these deficiencies are alleviated by the use of the Weighted Ensemble approach. The use of WE-based simulations not only produce conformations characterized by base and spike angles outside those sampled in standard MD, but also generate an enhanced number of larger volume CAM binding pockets. Furthermore, we show that these WE produced apo structures can be successfully used in docking GLS4, a reasonably large, well-studied CAM. These results open the way for further virtual screening studies targeting the allosteric modulation of HBV capsid assembly. Additionally, similar to results observed for the apo system, standard and WE-based MD simulations for the AT-130 bound case produce enhanced sampling of the HBV tetramer. In contrast, the exploration of the base and spike angle regions is comparable for the GLS4 bound state, suggesting a ligand-dependent effect. However, additional simulations are needed to confirm this. Finally, the use of HMR is shown to have only a modest effect on the protein and ligand equilibrium distributions. The present study, along with recent work by Xu et al.<sup>39</sup> and Hellemann and

Durrant,<sup>40</sup> has established that the WE approach is an effective enhanced sampling method for generating and characterizing ligand binding alterations in the protein conformational landscape and produces viable conformations for further docking studies.

## Methods

### Molecular Dynamics

The starting structures for the apo, AT-130, and GLS4 simulations, based on the 3J2V,<sup>53</sup> 4G93,<sup>54</sup> and 5E0I<sup>55</sup> (where the bound ligand was changed to GLS4) structures respectively, were taken from the previous equilibrated simulations of Pavlova et al.<sup>14</sup> and minimized for 2000 steps prior to simulation. For standard molecular dynamics each system was run for 500 ns employing 12 replicas for an aggregate of 6  $\mu$ s of simulation. All MD simulations were carried out with NAMD3<sup>56</sup> using the CHARMM36m force field for proteins<sup>57</sup> along with the TIP3P water model.<sup>58</sup> CHARMM force field parameters for GLS4 were taken from Pavlova et al.,<sup>14</sup> where the CGenFF<sup>59</sup> parameters via the webserver was used, while parameters for AT-130 were taken from Pang et al.<sup>60</sup> For the standard MD simulations, either a 2-fs or a hydrogen mass repartitioning (HMR) scheme was employed, with the latter allowing the use of a 4-fs time step.<sup>24,25</sup> In all simulations, the van der Waals cutoff was set to 1.2 nm, with a smoothing function applied from 1.0 to 1.2 nm. Long-range electrostatics were calculated using the particle mesh Ewald method.<sup>61</sup> A constant temperature of 310 K was maintained using Langevin dynamics and a constant pressure of 1 atm was maintained using the Langevin piston method available in NAMD.<sup>56</sup>

Molecular dynamics settings for the evolution of the replicas in the WE simulations were identical to those of the standard HMR MD described above, while the initial configuration was taken after 40 ns of standard MD. WE specific parameters are described further below. Unless otherwise noted, visualization and analysis were done using VMD version 1.9.4.<sup>62</sup> Pair-wise RMSD calculations were performed with MDTraj<sup>63</sup> and delaunay triangulation,

available in SciPy,<sup>64</sup> was used for generating progress coordinate boundaries.

## Weighted Ensemble

In the current WE simulations a 2-D progress coordinate was chosen based on the results of earlier work<sup>14</sup> which identified specific dimer-dimer orientations of the HBV tetramer, described by the base and spike angles of (Figure 2a), sensitive to the presence and/or type of CAM. The base angle reflects an opening or closing of the Cp dimers in the tetramer and is calculated based on the geometry of the constituent  $\alpha$ -5 helices, defined using the backbone atoms of residues 111 to 127, specifically for each dimer the vector between the geometric centers of the two  $\alpha$ -5 helices computed, and the angle between these 2 intra-dimer vectors is the base angle. The geometry of the  $\alpha$ -3 and  $\alpha$ -4 helices are used in the spike angle definition. Dividing these helices into an upper and lower portion about a flexible hinge generates two segments: an upper segment using residues 49 to 56 and residues 103 to 110 and a lower segment using residues 56 to 65 and 96 to 103. Given this definition the vector from the lower to upper center is evaluated for each dimer and the angle between these resultant vectors is the spike angle.

Although WE simulations are often run in order to provided unbiased kinetic data using steady state conditions, equilibrium conditions can also be applied. In fact WE has been used in order to effectively sample such equilibrium populations.<sup>28,30</sup> Analogous to the initial/preparatory equilibrium simulations for peptide-protein and protein-protein interactions<sup>29,33</sup> we employed equilibrium WE simulation using WESTPA version 1<sup>65</sup> (release. 2020.06) using 2-D progress coordinates specified by the base and spike angles discussed above, until reasonable convergence was obtained for the larger population regions. The simulations were performed using equilibrium conditions, with a  $\tau = 50$ ps, 8 trajectories per bin, and the bins subdivided at 20° boundaries for both the base and spike angles, initially set at [15,35,55,75,95] and [10,30,50,70,90] respectively. Given the variance observed in multiple WE runs<sup>33</sup> and as recommended<sup>66</sup> two independent WE simulations were per-

formed, each run for a collective  $\sim 3\mu\text{s}$  each. In the case of the apo simulations this resulted in 650/960 iterations for each replica respectively, while for AT-130 and GLS4 bound simulations the iterations per replica are 610/500 and 1350/1365 respectively. Note, when the progress coordinate exceeded the extreme bin boundaries these were subsequently extended. The evolution of each progress coordinate and the 2-D landscape for each replica were generated using the WESTPA tools *w\_pdist* and *plothist*, with the replicas combined using *w\_multi\_west*.

## Epock Analysis

CAM pocket volumes were computed using Epock.<sup>43</sup> The pocket was identified as those residues within a close contact ( $4.5 \text{ \AA}$ ) between heavy atoms of the ligand in the AT-130 and GLS4 structures from the starting structures and include residues 121, 124, 125, 127, 128, 129, 132, 133, 134, 136, and 138 from chain C and residues 23, 25, 29, 30, 33, 37, 102, 105, 106, 109, 110, 114, 115, 118, 138, 139, and 140 from chain B. The volume was obtained using a spherical probe ( $10 \text{ \AA}$  radii) along with non-contiguous points removal, which removes free space points outside a given radius here taken as  $3 \text{ \AA}$ . Both are centered on the geometrical center of the binding site residues listed above. Additionally an exclusion region, defined by a  $3 \text{ \AA}$  sphere centered at the COM of the sidechains of chain B residue 29 and chain C and residue 127, 129, and 133 was used. Binding pocket volumes using these choices are illustrated in Figure S26, where ligands under study are effectively encompassed.

## Clustering Analysis

The structures used for clustering were taken from the collection of all WE or standard MD conformations, the number of conformations used was 25412/27000 for the WE/standard MD respectively, nearly the same in order to cluster approximately the same amount of data at  $\sim 0.2\text{ns}$  intervals. These differences result from the WE algorithm, which generates data after each iteration. In both cases initial data was dropped prior to analysis, 10ns in the

case of standard MD and where the highly populated bin weights were approximately leveled off for the WE simulations, at 250 iterations. Xu et al.<sup>39</sup> employed a different approach; with the WE simulations using ligand binding site information (SASA) to sample alternate conformations, with the clustering performed using a representative pathway taken from the WE simulation. In our case we have focused on the functionally important motion for HBV tetrameric assembly modulation, i.e. extreme base and spike angle coordinates. In this case a representative trajectory may in fact miss the relevant apo conformations. Moreover, clustering WE simulations has been discussed by Hellemann and Durrant, where they employ a 2-tier approach to avoid striding over the branching points in a WE simulation.

The clustering of the base and spike angles was performed with SciPy<sup>64</sup> hierarchical clustering, using average euclidean linkage, with the data preprocessed prior to clustering such that the data has zero mean and unit standard deviation using sklearn.<sup>67</sup> The merging distance at each clustering step was used to determine the number of clusters, with a jump in the distance as another cluster pair is merged, indicating a separation in the data. Unless otherwise noted, i.e. for the apo standard MD described in Figure S25, the structure whose base and spike angles were closest to the center for each cluster was extracted for illustration and the docking calculations.

## Docking

The GLS4 molecule was docked into the capsid assembly modulator (CAM) binding pocket of the target protein using AutoDock Vina (version 1.1.2).<sup>68,69</sup> To account for potential conformational adjustments during docking, the side chains of residues within a 6 Å radius of the GLS4 molecule were set as flexible. This flexibility ensures accurate modeling of protein-ligand interactions in the binding pocket.

The docking simulation explored a search space defined as a 30 Å × 30 Å × 30 Å box, centered on the original position of GLS4 within its bound conformation ((PDB code 5E0I). The center of this box was determined after structural alignment of the protein conforma-

tions to the GLS4-bound structure. To achieve robust and consistent docking results, the exhaustiveness parameter was set to its maximum allowable value of 32.

## Acknowledgement

This work was supported by the National Institutes of Health (R01-AI148740). We also used the Hive cluster in this work, which is supported by the National Science Foundation (MRI-1828187) and is managed by the Partnership for an Advanced Computing Environment at Georgia Tech.

## References

- (1) Porat-Dahlerbruch, G.; Goldbourn, A.; Polenova, T. Virus Structures and Dynamics by Magic-Angle Spinning NMR. *Annu. Rev. Virol.* **2021**, *8*, 219–237.
- (2) James, E. I.; Murphree, T. A.; Vorauer, C.; Engen, J. R.; Guttman, M. Advances in Hydrogen/Deuterium Exchange Mass Spectrometry and the Pursuit of Challenging Biological Systems. *Chem. Rev.* **2021**, *122*, 7562–7623.
- (3) Papasergi-Scott, M. M.; ndez, G.; Batebi, H.; Gao, Y.; Eskici, G.; Seven, A. B.; Panova, O.; Hilger, D.; Casiraghi, M.; He, F.; Maul, L.; Gmeiner, P.; Kobilka, B. K.; Hildebrand, P. W.; Skiniotis, G. Time-resolved cryo-EM of G-protein activation by a GPCR. *Nature* **2024**, *629*, 1182–1191.
- (4) Ayaz, P.; Lyczek, A.; Paung, Y.; Mingione, V. R.; Iacob, R. E.; de Waal, P. W.; Engen, J. R.; Seeliger, M. A.; Shan, Y.; Shaw, D. E. Structural mechanism of a drug-binding process involving a large conformational change of the protein target. *Nat. Commun.* **2023**, *14*, 1885.

- (5) Tang, L. S. Y.; Covert, E.; Wilson, E.; Kottlil, S. Chronic Hepatitis B Infection: A Review. *JAMA* **2018**, *319*, 1802.
- (6) Bassit, L.; Amblard, F.; Patel, D.; Biteau, N.; Chen, Z.; Kasthuri, M.; Zhou, S.; Schinazi, R. F. The premise of capsid assembly modulators towards eliminating HBV persistence. *Expert Opin. Drug Discov.* **2023**, *18*, 1031–1041.
- (7) Inoue, J.; Sato, K.; Ninomiya, M.; Masamune, A. Envelope Proteins of Hepatitis B Virus: Molecular Biology and Involvement in Carcinogenesis. *Viruses* **2021**, *13*, 1124.
- (8) Cole, A. G. Modulators of HBV Capsid Assembly as an Approach to Treating Hepatitis B Virus Infection. *Curr. Opin. Pharmacol.* **2016**, *30*, 131–137.
- (9) Yang, L.; Liu, F.; Tong, X.; Hoffmann, D.; Zuo, J.; Lu, M. Treatment of Chronic Hepatitis B Virus Infection Using Small Molecule Modulators of Nucleocapsid Assembly: Recent Advances and Perspectives. *ACS Infect. Dis.* **2019**, *5*, 713–724.
- (10) Zlotnick, A.; Johnson, J. M.; Wingfield, P. W.; Stahl, S. J.; Endres, D. A Theoretical Model Successfully Identifies Features of Hepatitis B Virus Capsid Assembly. *Biochemistry* **1999**, *38*, 14644–14652.
- (11) Ceres, P.; Zlotnick, A. Weak Protein-protein Interactions are Sufficient to Drive Assembly of Hepatitis B Virus Capsids. *Biochemistry* **2002**, *41*, 11525–11531.
- (12) Stray, S. J.; Ceres, P.; Zlotnick, A. Zinc Ions Trigger Conformational Change and Oligomerization of Hepatitis B Virus Capsid Protein. *Biochemistry* **2004**, *43*, 9989–9998.
- (13) Hadden, J. A.; Perilla, J. R.; Schlicksup, C. J.; Venkatakrishnan, B.; Zlotnick, A.; Schulten, K. All-atom Molecular Dynamics of the HBV Capsid Reveals Insights Into Biological Function and Cryo-EM Resolution Limits. *eLife* **2018**, *7*, e32478.

- (14) Pavlova, A.; Bassit, L.; Cox, B. D.; Korablyov, M.; Chipot, C.; Patel, D.; Lynch, D. L.; Amblard, F.; Schinazi, R. F.; Gumbart, J. C. The Mechanism of Action of Hepatitis B Virus Capsid Assembly Modulators Can Be Predicted from Binding to Early Assembly Intermediates. *J. Med. Chem.* **2022**, *65*, 4854–4864.
- (15) Bourne, C. R.; Katen, S. P.; Fulz, M. R.; Packianathan, C.; Zlotnick, A. A Mutant Hepatitis B Virus Core Protein Mimics Inhibitors of Icosahedral Capsid Self-assembly. *Biochemistry* **2009**, *48*, 1736–1742.
- (16) Tan, Z.; Maguire, M. L.; Loeb, D. D.; Zlotnick, A. Genetically Altering the Thermodynamics and Kinetics of Hepatitis B Virus Capsid Assembly Has Profound Effects On Virus Replication in Cell Culture. *J. Virol.* **2013**, *87*, 3208–3216.
- (17) Bourne, C.; Lee, S.; Venkataiah, B.; Lee, A.; Korba, B.; Finn, M. G.; Zlotnick, A. Small-molecule Effectors of Hepatitis B Virus Capsid Assembly Give Insight Into Virus Life Cycle. *J. Virol.* **2008**, *82*, 10262–10270.
- (18) Delaney, W. E.; Edwards, R.; Colledge, D.; Shaw, T.; Furman, P.; Painter, G.; Locarnini, S. Phenylpropenamide Derivatives AT-61 and AT-130 Inhibit Replication of Wild-Type and Lamivudine-Resistant Strains of Hepatitis B Virus in Vitro. *Antimicrob. Agents Chemother.* **2002**, *46*, 3057–3060.
- (19) Feld, J. J.; Colledge, D.; Sozzi, V.; Edwards, R.; Littlejohn, M.; Locarnini, S. A. The Phenylpropenamide Derivative AT-130 Blocks HBV Replication at the Level of Viral RNA Packaging. *Antiviral Res.* **2007**, *76*, 168–177.
- (20) Deres, K. et al. Inhibition of Hepatitis B Virus Replication by Drug-induced Depletion of Nucleocapsids. *Science* **2003**, *299*, 893–896.
- (21) Packianathan, C.; Katen, S. P.; Dann, C. E.; Zlotnick, A. Conformational Changes in the Hepatitis B Virus Core Protein Are Consistent with a Role for Allostery in Virus Assembly. *J. Virol.* **2010**, *84*, 1607–1615.



- (22) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys J* **2018**, *114*, 2271–2278.
- (23) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2022**, *4*, 1583.
- (24) 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.
- (25) Balusek, C.; Hwang, H.; Lau, C. H.; Lundquist, K.; Hazel, A.; Pavlova, A.; Lynch, D. L.; Reggio, P. H.; Wang, Y.; Gumbart, J. C. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2019**, *15*, 4673–4686.
- (26) Sahil, M.; Sarkar, S.; Mondal, J. Long-time-step molecular dynamics can retard simulation of protein-ligand recognition process. *Biophys. J.* **2023**, *122*, 802–816.
- (27) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43–57.
- (28) Suárez, E.; Lettieri, S.; Zwier, M. C.; Stringer, C. A.; Subramanian, S. R.; Chong, L. T.; Zuckerman, D. M. Simultaneous Computation of Dynamical and Equilibrium Information Using a Weighted Ensemble of Trajectories. *J. Chem. Theory Comput.* **2014**, *10*, 2658–2667.
- (29) Zwier, M. C.; Pratt, A. J.; Adelman, J. L.; Kaus, J. W.; Zuckerman, D. M.; Chong, L. T. Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein–Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered p53 Peptide. *J. Phys. Chem. Lett.* **2016**, *7*, 3440–3445.

- (30) Saglam, A. S.; Chong, L. T. Protein–protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chem. Sci.* **2019**, *10*, 2360–2372.
- (31) Nunes-Alves, A.; Zuckerman, D. M.; Arantes, G. M. Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. *Biophys. J.* **2018**, *114*, 1058–1066.
- (32) Dixon, T.; Lotz, S. D.; Dickson, A. *Computational Methods for Estimating the Kinetic Parameters of Biological Systems*; Springer US, 2021; p 325–334.
- (33) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *J. Am. Chem. Soc.* **2019**, *141*, 6519–6526.
- (34) Sztain, T. et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* **2021**, *13*, 963–968.
- (35) Dommer, A. et al. #COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol. *Int. J. High Perform. Comput. Appl.* **2023**, *37*, 28–44.
- (36) Zhang, S.; Thompson, J. P.; Xia, J.; Bogetti, A. T.; York, F.; Skillman, A. G.; Chong, L. T.; LeBard, D. N. Mechanistic Insights into Passive Membrane Permeability of Drug-like Molecules from a Weighted Ensemble of Trajectories. *J. Chem. Inf. Model.* **2022**, *62*, 1891–1904.
- (37) Poruthoor, A. J.; Sharma, A.; Grossfield, A. Understanding the free-energy landscape of phase separation in lipid bilayers using molecular dynamics. *Biophys. J.* **2023**, *122*, 4144–4159.
- (38) Russo, J. D. et al. WESTPA 2.0: High-Performance Upgrades for Weighted Ensemble

- Simulations and Analysis of Longer-Timescale Applications. *J. Chem. Theory Comput.* **2022**, *18*, 638–649.
- (39) Xu, X.; Closson, J. D.; Marcelino, L. P.; Favaro, D. C.; Silvestrini, M. L.; Solazzo, R.; Chong, L. T.; Gardner, K. H. Identification of small-molecule ligand-binding sites on and in the ARNT PAS-B domain. *J. Biol. Chem.* **2024**, *300*, 107606.
- (40) Hellemann, E.; Durrant, J. D. Worth the Weight: Sub-Pocket EXplorer (SubPEX), a Weighted Ensemble Method to Enhance Binding-Pocket Conformational Sampling. *J. Chem. Theory Comput.* **2023**, *19*, 5677–5689.
- (41) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding False Positive Conclusions in Molecular Simulation: The Importance of Replicas. *J. Chem. Theory Comput.* **2018**, *14*, 6127–6138.
- (42) Fan, A.; Pavlova, A.; Jenkins, M. C.; Bassit, L.; Salman, M.; Lynch, D. L.; Patel, D.; Korablyov, M.; Finn, M. G.; Schinazi, R. F.; Gumbart, J. C. Biophysics-Guided Lead Discovery of HBV Capsid Assembly Modifiers. *ACS Infect. Dis.* **2024**, *10*, 1162–1173.
- (43) Laurent, B.; Chavent, M.; Cragolini, T.; Dahl, A. C. E.; Pasquali, S.; Derreumaux, P.; Sansom, M. S.; Baaden, M. Epock: Rapid analysis of protein pocket dynamics. *Bioinformatics* **2014**, *31*, 1478–1480.
- (44) Basciu, A.; Mallocci, G.; Pietrucci, F.; Bonvin, A. M. J. J.; Vargiu, A. V. Holo-like and Druggable Protein Conformations from Enhanced Sampling of Binding Pocket Volume and Shape. *J. Chem. Inf. Model.* **2019**, *59*, 1515–1528.
- (45) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The “weighted ensemble” path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *J. Chem. Phys.* **2010**, *132*, 054107.

- (46) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *J. Phys. Chem. A* **2021**, *125*, 1642–1649.
- (47) McGovern, S. L.; Shoichet, B. K. Information Decay in Molecular Docking Screens against Holo, Apo, and Modeled Conformations of Enzymes. *J. Med. Chem.* **2003**, *46*, 2895–2907.
- (48) Guterres, H.; Park, S.-J.; Jiang, W.; Im, W. Ligand-Binding-Site Refinement to Generate Reliable Holo Protein Structure Conformations from Apo Structures. *J. Chem. Inf. Model.* **2020**, *61*, 535–546.
- (49) Miller, E. B. et al. Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein–Ligand Binding. *J. Chem. Theory Comput.* **2021**, *17*, 2630–2639.
- (50) Evangelista Falcon, W.; Ellingson, S. R.; Smith, J. C.; Baudry, J. Ensemble Docking in Drug Discovery: How Many Protein Configurations from Molecular Dynamics Simulations are Needed To Reproduce Known Ligand Binding? *J. Phys. Chem. B* **2019**, *123*, 5189–5195.
- (51) Acharya, A. et al. Supercomputer-Based Ensemble Docking Drug Discovery Pipeline with Application to Covid-19. *J. Chem. Inf. Model.* **2020**, *60*, 5832–5852.
- (52) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.
- (53) Yu, X.; Jin, L.; Jih, J.; Shih, C.; Hong Zhou, Z. 3.5 Å cryo-EM Structure of Hepatitis B Virus Core Assembled From Full-Length Core Protein. *PLoS One* **2013**, *8*, e69729.
- (54) Katen, S. P.; Tan, Z.; Chirapu, S. R.; Finn, M.; Zlotnick, A. Assembly-Directed Antivirals Differentially Bind Quasiequivalent Pockets to Modify Hepatitis B Virus Capsid Tertiary and Quaternary Structure. *Structure* **2013**, *21*, 1406–1416.

- (55) Klumpp, K.; Lam, A. M.; Lukacs, C.; Vogel, R.; Ren, S.; Espiritu, C.; Baydo, R.; Atkins, K.; Abendroth, J.; Liao, G.; Efimov, A.; Hartman, G.; Flores, O. A. High-resolution Crystal Structure of a Hepatitis B Virus Replication Inhibitor Bound to the Viral Core Protein. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 15196–15201.
- (56) Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
- (57) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D. CHARMM36m: an Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73.
- (58) 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.
- (59) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell Jr., A. D. CHARMM General Force Field: a Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (60) Pang, Y. T.; Pavlova, A.; Tajkhorshid, E.; Gumbart, J. C. Parameterization of a drug molecule with a halogen  $\sigma$ -hole particle using ffTK: Implementation, testing, and comparison. *J. Chem. Phys.* **2020**, *153*, 164104.
- (61) Darden, T. A.; York, D. M.; Pedersen, L. G. Particle Mesh Ewald: an  $N \log N$  Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (62) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

- (63) 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. MD-Traj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532.
- (64) Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
- (65) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *J. Chem. Theory Comput.* **2015**, *11*, 800–809.
- (66) Bogetti, A. T.; Leung, J. M. G.; Russo, J. D.; Zhang, S.; Thompson, J. P.; Saglam, A. S.; Ray, D.; Abraham, R. C.; Faeder, J. R.; Andricioaei, I.; Adelman, J. L.; Zwier, M. C.; LeBard, D. N.; Zuckerman, D. M.; Chong, L. T. A Suite of Advanced Tutorials for the WESTPA 2.0 Rare-Events Sampling Software [Article v2.0]. *Living J. Comput. Mol. Sci.* **2022**, *5*, 1655.
- (67) Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (68) Eberhardt, T.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898.
- (69) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.

## TOC Graphic

