# Supporting Information:
# Mechanism of action of HBV capsid assembly modulators can be predicted from binding to early assembly intermediates

Anna Pavlova,[†] Leda Bassit,[‡] Bryan D. Cox,[‡] Maksym Korablyov,[¶] Christophe Chipot,[§,∥] Dharmeshkumar Patel,[‡] Diane L. Lynch,[†] Franck Amblard,[‡] Raymond F. Schinazi,[‡] and James C. Gumbart[*,†]

†*School of Physics and School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, Georgia, 30332, USA*

‡*Center for AIDS Research, Laboratory of Biochemical Pharmacology, Department of Pediatrics, Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, Georgia, 30322 USA*

¶*MIT Media Lab, Massachusetts Institute of Technology, Boston, Massachusetts, 02139, USA*

§*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, USA*

∥*Laboratoire international associé CNRS-UIUC. UMR 7019. Université de Lorraine. B.P. 70239. 54506 Vandœuvre-lès-Nancy, France*

E-mail: gumbart@physics.gatech.edu

# Computational Details

## Structure Preparation

For structures containing mutations, WT residues were restored. The 5E0I crystal structure had missing residues at the spike tops of $\alpha 4$ and $\alpha 3$ helices for some of the dimers. These residues were modeled based on the other, fully resolved, dimers of these structure. Additionally, because the Y132A mutation is located at the inter-dimer interface, the Y132 residue was manually placed in the same position as in the WT 3J2V structure for both 5E0I and 5T2P structures. We used the Molfacture plugin in VMD to construct the different HAP compounds bound to 5E0I structure, SBAs and GLPs bound to the 5T2P structure, and AT61 bound to 4G93 structure. The energy of all built compounds was minimized in the binding site, while the protein structure was restrained before the addition of water and ions. All systems were solvated and ionized with 0.15 M NaCl, using the solvate and ionize plugins in VMD,[S1] respectively.

## MD protocol

All simulations employed the CHARMM36 force field[S2] for the capsid protein and the TIP3P model for water.[S3] For the drug molecules we use the CGenFF[S4] parameters obtained from the CGenFF webserver (https://cgenff.umaryland.edu/).[S5,S6] Rigid bonds were used for all covalent hydrogen bonds, allowing us to integrate the equations of motion with a 2-fs time step. Van der Waals interactions were cutoff at 12 Å and a smoothing function was applied from 10 to 12 Å to ensure a smooth decay to zero. Long-range electrostatic interactions were calculated using the particle-mesh Ewald method.[S7] The temperature and the pressure were kept constant at the biologically relevant values of 310 K and 1 bar, respectively. The Langevin thermostat was used to control the temperature in all simulations. For pressure control we used the Langevin piston[S8] in NAMD and the Berendsen barostat[S9] in AMBER. For the Langevin piston we used a period of 200 fs and a decay of 100 fs,[S8] while for the

Berendsen thermostat we used $\tau$=1.0 ps.[S9]

## Definition of base and spike angles

All geometric centers were determined based on the positions of backbone atoms. Base angles were calculated from the centers of $\alpha5$ helices in the tetramer (see Figure S2A). These centers were calculated for each of the two $\alpha5$ helices in a protein dimer, after which the vector going through both centers and pointing away from the tetramer interface was determined. Next, the procedure was repeated for the other dimer in the tetramer, and the base angle was defined as the angle between the two determined vectors. Because the upper parts of the protein spikes (helices $\alpha4$ and $\alpha3$) were highly flexible, we decided to define the spike angle based on the lower parts of these helices displayed in Figure S2B. Specifically, the lower spike part was divided into top and bottom segments; residues 49 to 65 and 103 to 110 defined the top segment, while residues 56 to 65 and 96 to 103 defined the bottom segment. For each dimer we calculated the vector going through the geometric centers of the bottom and top segments and pointing towards the spike top. The spike angle was defined as the angle between these two vectors. We investigated if the value of spike angle was sensitive to the exact residue selection by also calculating the same angle using residues 51 to 58 and 101 to 108 for the top segment and residues 58 to 65 and 94 to 101 for the bottom segment. Figure S3 shows that the two different spike angle definitions resulted in very minor differences in the value of this angle.

## FEP calculations

At each stage of the stratified reaction path, data collection was prefaced by thermalization in the amount of one fourth to one third of the sampling. To augment the accuracy of the free-energy calculation, the alchemical transformations have been performed bidirectionally, and the relative binding free energy was determined using the Bennett acceptance ratio method.[S10] Estimation of errors in free-energy calculations is notoriously difficult owing to sources of very different natures, from the finite length of the simulations to force-field inaccuracies,[S11] presupposing stringent underlying approximations. We have chosen to provide error bars based on the hysteresis between the forward and backward transformations of the bidirectional free-energy calculations, which has proven more realistic than a mere estimate of the statistical error (Table S4). The relative binding free energy between substrates A and B, $\Delta\Delta G(A\rightarrow B)$, is defined as $\Delta G_{\mathrm{bind}}(B) - \Delta G_{\mathrm{bind}}(A) = \Delta G_{\mathrm{bound}} - \Delta G_{\mathrm{unbound}}$, where $\Delta G_{\mathrm{bound}}$ and $\Delta G_{\mathrm{unbound}}$ are the FEP energies for bound and unbound state, respectively. Table S4 shows all calculated FEP energies.

## Theoretical p$K_a$ calculations

The p$K_a$s of HAP7 and HAP12 were calculated relative to triethylamine (TEA) employing Eq. 1 derived in Gangarapu et al.[S12]

$$pK_{\mathrm{a,HAP}} = pK_{\mathrm{a,TEA}} + 1/(2.303\mathrm{RT})(\Delta\Delta G_{\mathrm{solv,Neu}} - \Delta\Delta G_{\mathrm{solv,H^+}} + \Delta\Delta G_{\mathrm{gas}}) \tag{1}$$

$R$ and $T$ are the ideal gas constant and temperature, respectively. $\Delta\Delta G_{\mathrm{solv,Neu}}$ and $\Delta\Delta G_{\mathrm{solv,H^+}}$ are the differences in solvation free energies between HAP and TEA for the neutral and protonated forms, respectively. Finally, $\Delta\Delta G_{\mathrm{gas}}$ is the difference in gas-phase deprotonation energy between HAP and TEA. All $\Delta\Delta G$ are defined relative to HAP ($\Delta\Delta G = \Delta G_{\mathrm{HAP}} - \Delta G_{\mathrm{TEA}}$). The Gaussian16 program[S13] was used for calculations of all properties. All geometries were optimized at B3LYP/6-31+G** level of theory. Single point calculations on these geometries were performed with a larger basis set, B3LYP/6-311+G(2d,2p), in both gas phase and water. We used solvation model based on density (SMD) to calculate the solutes solvation free energies in water.[S14] Deprotonation energy ($\Delta G_{\mathrm{gas}}$) was calculated as the difference in total energies and thermal corrections in gas phase between the deprotonated and protonated states. The total energy was taken from the single point calculation at B3LYP/6-311+G(2d,2p) level of theory in gas phase. Unfortunately, the HAP compounds were too large for a frequency calculation with this basis set. Therefore, frequencies were calculated at the same level of theory as the geometry optimization and the computed thermal energies were scaled by a factor of 0.977, as suggested for the B3LYP/6-31G(d) level of theory.[S15] In a previous study a similar approach was shown to have a mean unsigned error of 0.33 p$K_a$ units.[S12]

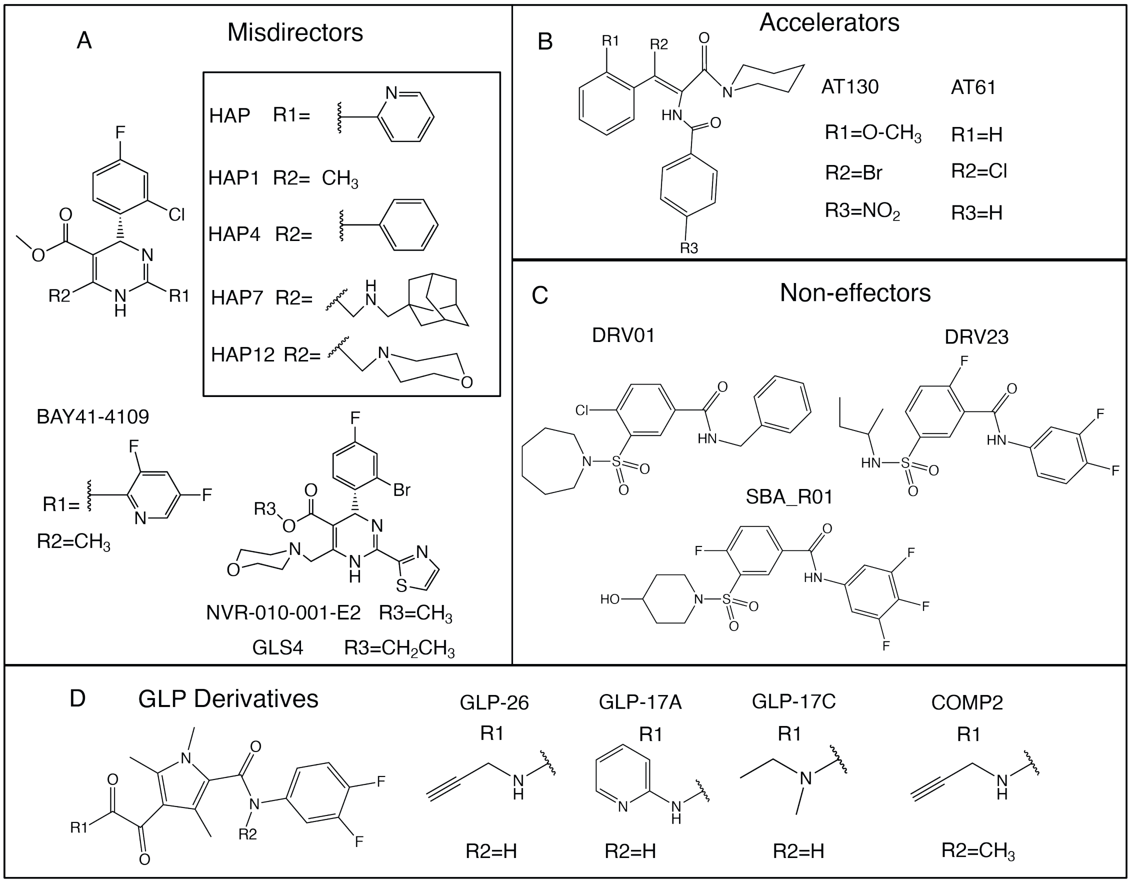

Figure S1: Structures of the known classes of CAMs. A) Heteroaryldihydropyrimidines (HAPs) that misdirect the assembly into non-capsid structures. B) Phenylpropenamides (ATs) that accelerate capsid assembly. C) Sulfamoyl benzamides (SBAs) that do not alter the assembly of empty capsids, yet prevent capsid incorporation of viral DNA. D) GLP derivatives studied here.

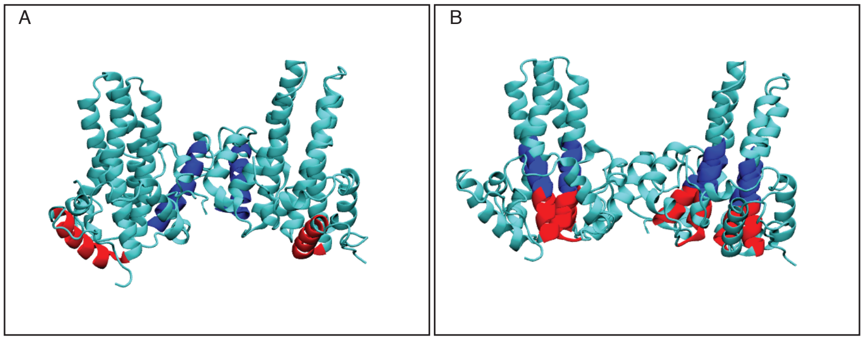

Figure S2: Definition of base and spike angles used in our structure analysis. A) $\alpha 5$ helices used for base angle calculations are shown (residues 111 to 127). The outer and inner $\alpha 5$ helices are shown in red and dark blue, respectively. B) The top and bottom segments of $\alpha 3$ and $\alpha 4$ helices used for spike angle calculation. The top segment is colored dark blue and the bottom segment is colored red.

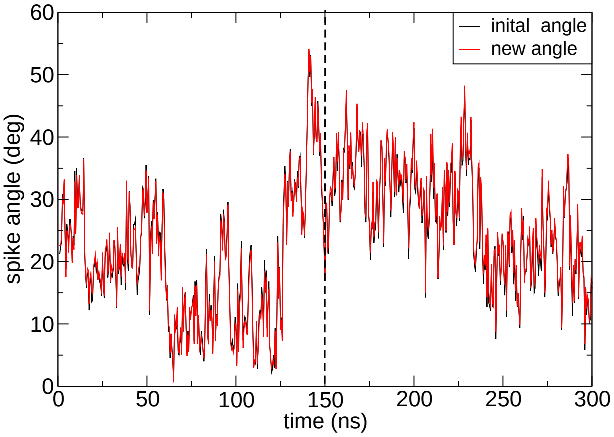

Figure S3: Comparison of spike angle values between the initial definition (black) and a slightly altered definition (red) for our apo-tetramer simulation. Dotted line separates the two 150-ns-long simulations.

# Quasi-Equivalent Structures in HBV capsid

The HBV virus capsid has an icosahedral geometry with triangulation number T = 4.[S16] Therefore, it is assembled from 240 copies of the same core protein, HBc, that can assume four different conformations in the shell (A, B, C, or D), depending on its position.[S17,S18] More specifically, the HBc dimers adopt either AB or CD conformations in the assembled capsid.[S19,S20] While the two dimers are structurally very similar, larger differences are observed in the quaternary structures for the four quasi-equivalent dimer-dimer contacts found in the capsid. In the assembled capsid there are four quasi-equivalent tetramers (ABCD, DCBA, BAAB, CDCD) and two quasi-equivalent hexamers (CDCDCD and ABCDBA). Note that the dimer interface is not symmetric with respect to the two dimers: the $\alpha5$ helix of one dimer is wedged between helices $\alpha5$ and $\alpha2$ of the other dimer; therefore, ABCD and DCBA are structurally different tetramers. For our tetramer simulations, we chose the ABCD structures, which contain all monomer structures found in the capsid. The structural differences between the four quasi-equivalent tetramers in the crystal structures are relatively small in comparison to the structural fluctuations observed in MD simulations (see Table S1). Additionally, the range of spike and base angles observed in capsid structures for quasi-equivalent tetramers is significantly narrower in comparison to MD simulations (Figure S4). Therefore, we argue that simulating one quasi-equivalent tetramer is sufficient.

For the hexamer, both CDCDCD and ABCDBA hexamers, referred to as symmetric hexamer (SH) and asymmetric hexamer (AH), respectively, were simulated. Three different overlapping tetrameric units are present in the hexamer, resulting in three base and spike angles for each analyzed structure. In AH the different tetramers often showed widely different base and spike angle distributions relative to each other in the same simulation and relative to different simulations (see Figure S5). In contrast, the SH simulations display only slight deviation from the symmetry for the same distributions. Our results suggest that SH is more structurally stable than AH and, therefore, it is a more realistic model for the assembly nucleus. Hence, we have used the results from the simulations of SH for comparison to other structures.

Table S1: Root mean square deviation (RMSD) in Å between the four quasi-equivalent tetramer structures found in assembled capsids. The MD column displays averaged RMSD from the two 150-ns-long MD simulations of the WT ABCD tetramer starting from structure 3J2V relative to each capsid structure.

| Structure | ABCD | DCBA | BAAB | CDCD | MD (300 ns) |
|---|---|---|---|---|---|
| ABCD | 0 | 2.8 | 2.9 | 2.0 | 4.3 |
| CDBA | | 0 | 2.9 | 1.9 | 4.9 |
| BAAB | | | 0 | 1.8 | 5.0 |
| CDCD | | | | 0 | 4.7 |

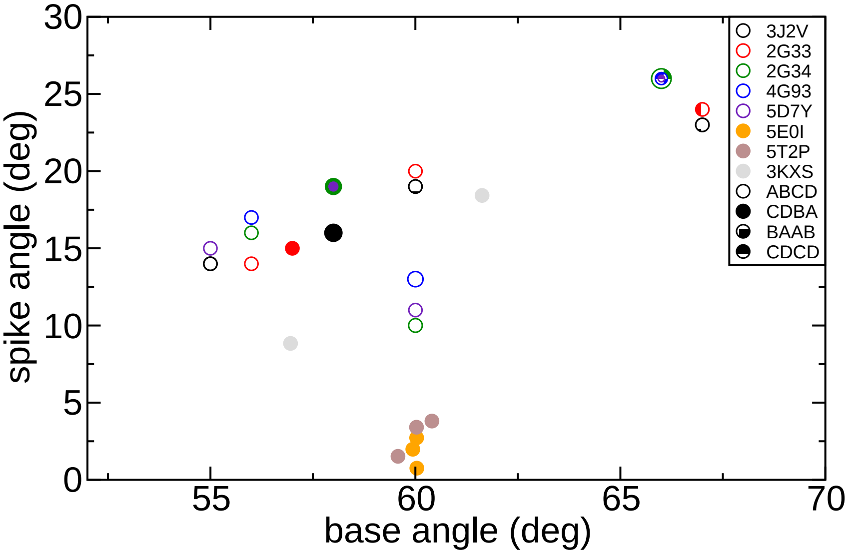

Figure S4: Calculated values for spike and base angles for the different quasi-equivalent tetramers in known HBV capsid and hexamer structures. Circle color indicates PDB code, while circle pattern indicates order of the monomers in the tetramer, as shown in the legends. Four quasi-equivalent tetramer structures are found in T4 capsids: ABCD, CDBA, BAAB and CDCD. For hexamer pdb structures (pdb codes 5E0I, 5T2P and 3KXS) the values were calculated for all possible tetramers and plotted in the same color.

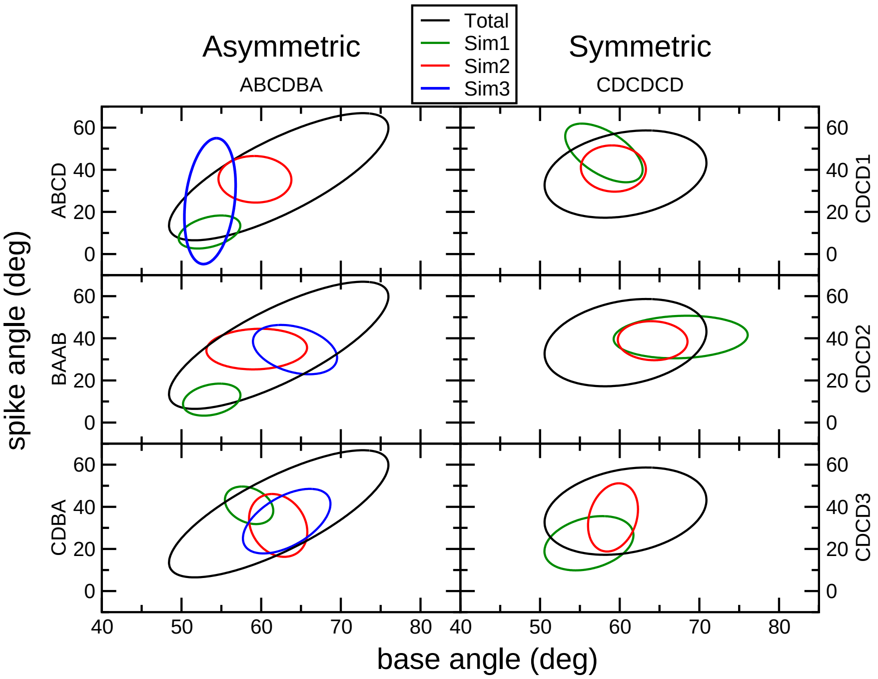

Figure S5: Comparison of standard deviation ellipses for different tetramers in the asymmetric (left graphs) and symmetric (right graphs) hexamer. Each graph shows the ellipses for a specific tetramer during different simulations and the ellipse resulting from averages of all tetramers and all simulations (black).

## Repeated simulations

To ensure that our results were not artifacts of the starting structures, we also performed WT tetramer and hexamer simulations using the structure of Y132A mutant hexamer with bound NVR-010-001-E2 (pdb code 5E0I). The drug molecule was removed and the mutation was reversed prior to the simulations. The resulting tetramer simulations showed a similar range of base angles and lower spike angles (Table 3) in comparison to the simulations of ACBD tetramer from the capsid. There is a close resemblance to the results of the Y132A mutant starting from the same structure, suggesting this mutation does not significantly alter tetramer conformations (Figure S6A). One of the two hexamer simulations approximated the results for the SH from the capsid, while the other simulation resembled the results from the AH in terms of base and spike angle sampling (Figure S7). Figure S6 shows that the same conclusions about tetramers, hexamers, mutants and CAMs could be drawn from simulations starting from a different crystal structure.

To further validate our results we also performed simulations of Y132A mutant starting from the capsid ABCD tetramer, and repeated the simulations of the WT tetramer starting from the capsid structure. Figure S8 shows that for both systems one of the simulations agreed with the previous results, while the second simulation sampled base and spike angles found in the hexamer simulations and not in previous apo tetramer simulations. Our simulations suggest that apo-tetramers have a greater flexibility than apo-hexamers or drug-bound tetramers.

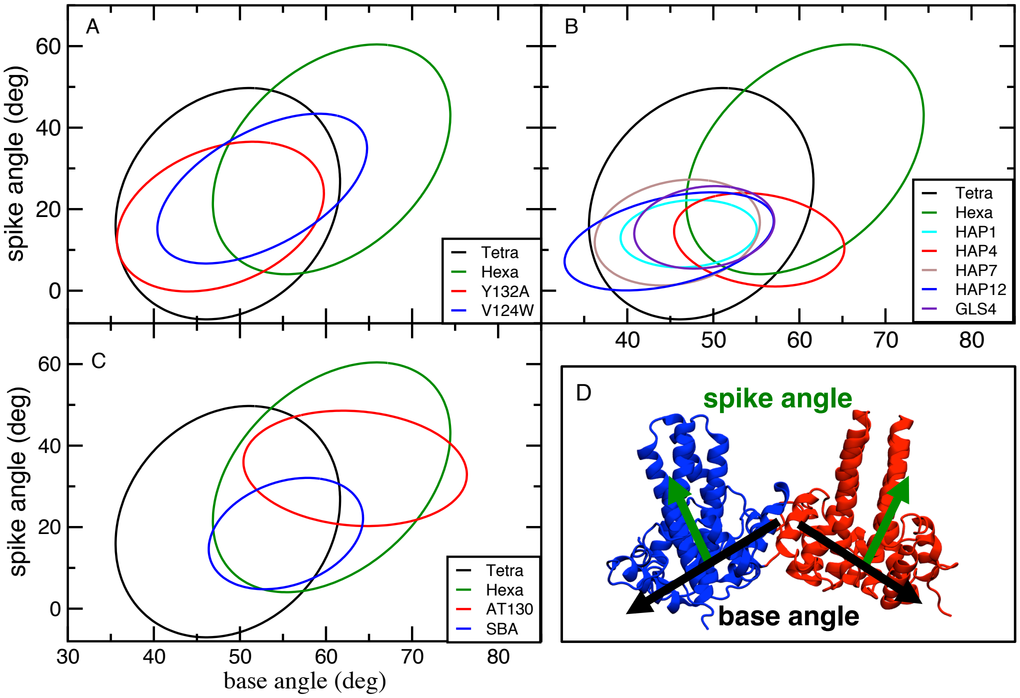

Figure S6: Comparison of standard deviation ellipses for apo tetramer and hexamer simulations starting from the hexamer structure (pdb code 5E0I) to the results of mutant and selected compound-bound tetramers.

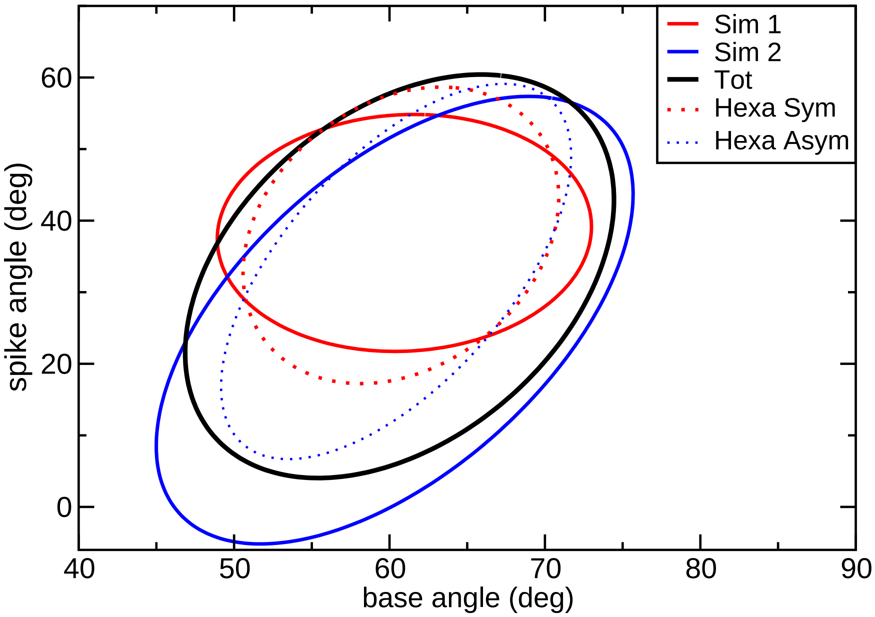

Figure S7: Comparison of standard deviation ellipses for the two simulations starting from the hexamer structure 5E0I to the corresponding structures of the symmetric and asymmetric hexamers from 3J2V.

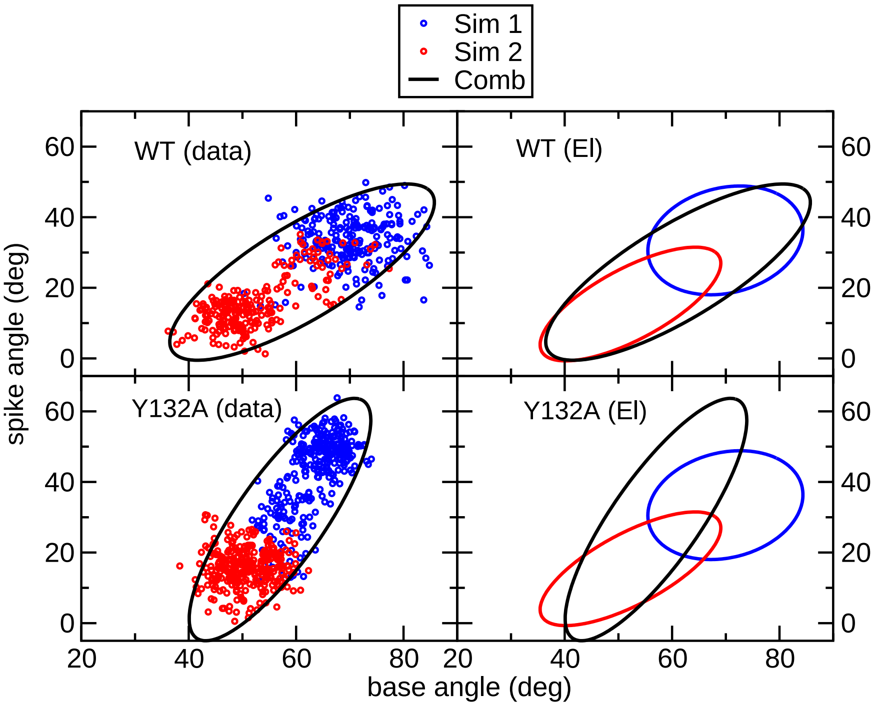

Figure S8: Top left: raw data and the resulting standard deviation ellipse for rerun of WT simulation. Top right: comparison of ellipses from the two simulations to the combined ellipse for the rerun of WT simulations. Bottom left: raw data and the resulting standard deviation ellipse for rerun of tetramer with Y132A mutation starting from the capsid state. Bottom right: comparison of ellipses from the two simulations to the combined ellipse for rerun of tetramer with Y132A mutation starting from the capsid state

# Additional simulation data

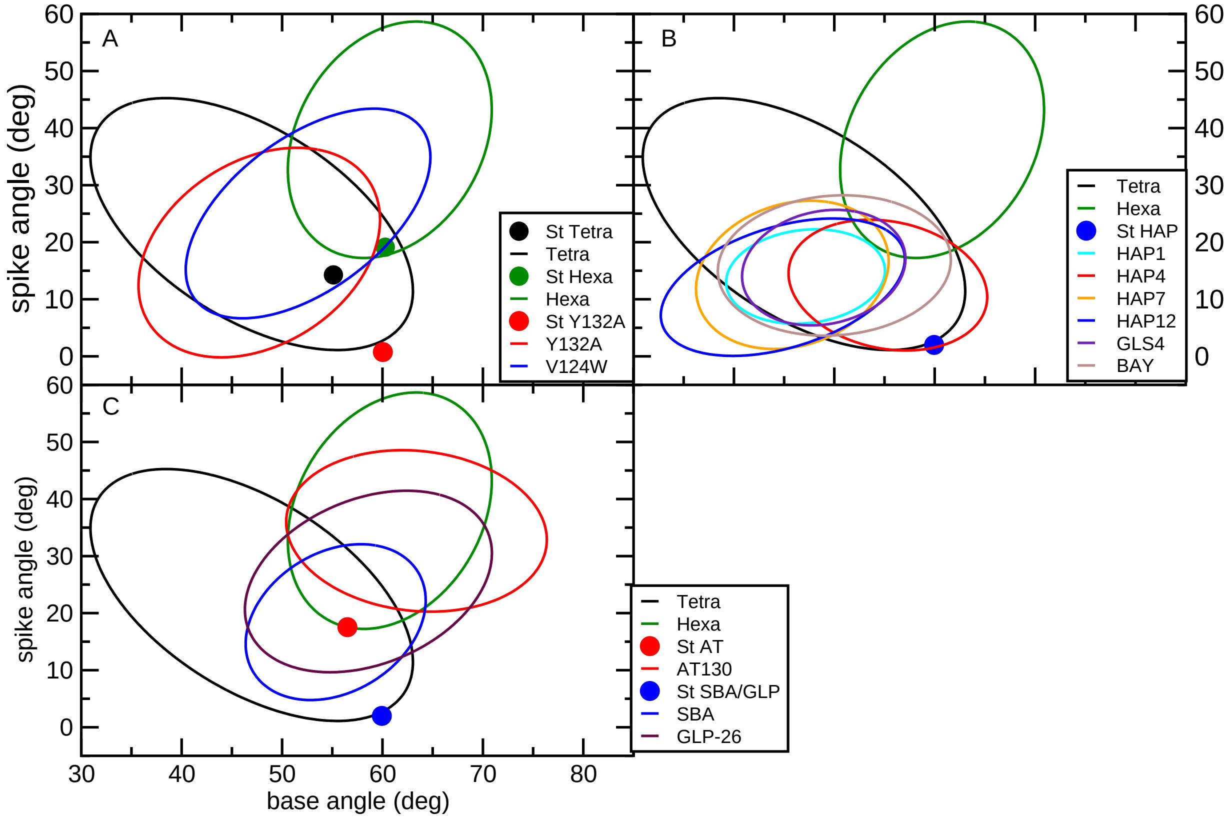

Figure S9: Standard deviation ellipses for selected simulated tetramers. Symmetric hexamer results are added for comparison. In addition, the spike and base values for starting structures are also shown.

Table S2: Average values (Av) and standard errors (SE) in deg for the studied systems. SEs were calculated from averages of the two independent simulations.

| System | Base Angle | | Spike Angle | |
|---|---|---|---|---|
| | Av | SE | Av | SE |
| Apo | | | | |
| WT Tetra (3J2V) | 47.0 | 4.4 | 23.2 | 3.5 |
| Hexa Asym | 62.2 | 1.6 | 36.7 | 2.6 |
| Hexa Sym | 60.7 | 0.1 | 37.9 | 0.6 |
| Y132A (3J2V) | 57.1 | 6.6 | 30.1 | 14.3 |
| V124W | 52.4 | 3.9 | 25.2 | 5.1 |
| WT Tetra (5E0I) | 48.6 | 1.2 | 21.3 | 7.3 |
| Y132A (5E0I) | 47.4 | 0.8 | 18.9 | 2.7 |
| WT Hexa Sym (5E0I) | 60.6 | 0.3 | 32.2 | 6.1 |
| HAP | | | | |
| HAP1 | 47.1 | 1.0 | 14.0 | 0.5 |
| HAP4 | 55.3 | 1.2 | 12.5 | 2.1 |
| HAP7 | 45.8 | 2.6 | 14.3 | 2.3 |
| HAP12 | 44.9 | 1.9 | 12.1 | 2.9 |
| BAY | 50.0 | 2.2 | 15.9 | 0.2 |
| GSL4 | 49.0 | 0.4 | 15.5 | 0.6 |
| 1 GLS4 Hexa | 60.1 | <0.1 | 10.4 | 1.8 |
| 3 GLS4 Hexa | 60.1 | <0.1 | 8.1 | 0.4 |
| AT | | | | |
| AT130 | 63.4 | 3.0 | 34.6 | 2.1 |
| AT61 | 73.4 | 2.2 | 34.2 | 3.2 |
| 1 AT130 Hexa | 60.7 | <0.1 | 37.6 | 0.2 |
| 3 AT130 Hexa | 60.7 | 0.1 | 34.4 | 0.3 |
| SBA/GLP | | | | |
| SBA_R01 | 55.3 | 0.4 | 18.4 | 0.7 |
| DRV01 | 58.2 | 3.1 | 34.4 | 4.5 |
| DRV23 | 57.3 | 0.6 | 28.9 | 1.3 |
| GLP-17A | 55.3 | 1.6 | 16.3 | 2.0 |
| GLP-17C | 54.7 | 1.1 | 22.0 | 5.2 |
| Comp2 | 57.0 | 3.4 | 23.8 | 6.7 |
| GLP-26 Tetra | 58.6 | 1.1 | 25.6 | 1.9 |
| 1 GLP-26 Hexa | 60.3 | 0.2 | 20.4 | 6.8 |

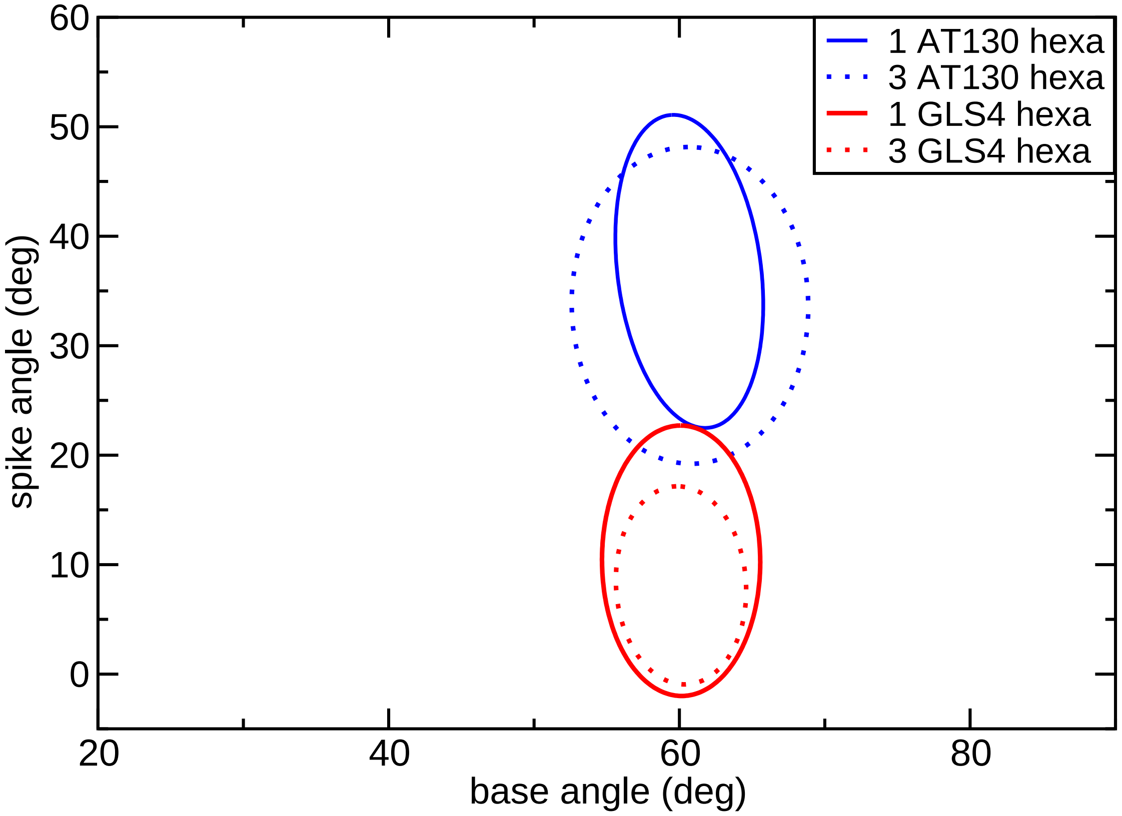

Figure S10: Standard deviation ellipses for AT130 and GLS4 with either one or three (one at each tetrameric interface) bound molecules. FOAs between one molecule and three molecule systems were 94% and 60%, for AT130 and GLS4, respectively.

Table S3: Average number of hydrogen bond present in the combined simulations between the specified capsid protein amino acid and CAM. The backbone of Leu140 is able to form two hydrogen bonds simultaneously and both were counted in the average. The errors were calculated from standard deviation.

| CAM | Trp102 | Thr128 | Leu140 |
|---|---|---|---|
| HAP | | | |
| HAP1 | 0.96±0.004 | 0.0 | 0.07±0.05 |
| HAP4 | 0.98±0.0 | 0.0 | 0.0 |
| HAP7 | 0.98±0.004 | 0.05±0.01 | 0.05±0.04 |
| HAP12 | 0.94±0.02 | 0.0 | 0.05±0.03 |
| BAY | 0.98±0.004 | 0.0 | 0.06±0.03 |
| GLS4 | 0.97±0.02 | 0.0 | 0.0 |
| AT | | | |
| AT130 | 0.48±0.1 | 0.18±0.05 | 0.0 |
| AT61 | 0.34±0.25 | 0.07±0.06 | 0.03±0.03 |
| SBA | | | |
| SBA_R01 | 0.53 ±0.15 | 0.9 ±0.1 | 0.0 |
| DRV01 | 0.40 ±0.17 | 0.43 ±0.15 | 0.0 |
| DRV23 | 0.53 ±0.002 | 0.95 ±0.004 | 0.94 ±0.04 |
| GLP | | | |
| GLP-17A | 0.86 ±0.03 | 0.89 ±0.01 | 1.50 ±0.05 |
| GLP-17C | 0.79 ±0.07 | 0.84 ±0.03 | 0.18 ±0.11 |
| Comp2 | 0.23±0.21 | 0.0 | 1.49±0.09 |
| GLP-26 | 0.88±0.005 | 0.8±0.04 | 1.1±0.43 |

Table S4: Summary of the alchemical free-energy calculations performed to estimate the relative binding free energies of HAP1, HAP4, HAP7 and GLS4 to the Cp149 tetramer. $\Delta G_{\text{forward}}$ and $\Delta G_{\text{backward}}$ represent the free-energy perturbation estimates for the annihilation and creation of the substrate in the bound and in the unbound states. $\Delta G_{\text{BAR}}$ is the Bennett acceptance ratio estimate of the binding free energy, $\Delta G_{\text{bind}}$, based on the bidirectional transformations.[S10] All free energies are given in kcal/mol.

| | time (ns) | $\Delta G_{\text{forward}}$ | $\Delta G_{\text{backward}}$ | $\Delta G_{\text{BAR}}$ | $\Delta\Delta G_{\text{exp}}$ |
|---|---|---|---|---|---|
| $\text{HAP7}_{\text{neu}} \rightarrow \text{HAP7}_{\text{char}}$ | | | | | |
| bound | 160 | +16.8 | -14.5 | +15.6±1.2 | |
| unbound | 40 | 23.2 | -22.8 | +23.0±0.2 | |
| ΔΔG | | | | +7.4±1.2 | > NA |
| $\text{HAP12} \rightarrow \text{HAP7}_{\text{neu}}$ | | | | | |
| bound | 80 | +68.1 | -64.6 | +66.0±1.7 | |
| unbound | 40 | +64.4 | -64.8 | +64.6±0.2 | |
| ΔΔG | | | | +1.4±1.9 | > 4 |
| HAP12 → HAP4 | | | | | |
| bound | 15 | -11.0 | +11.1 | -11.3±0.0 | |
| unbound | 15 | -13.7 | +14.1 | -13.9±0.2 | |
| ΔΔG | | | | +2.6±0.2 | +3 |
| HAP12 → GLS4 | | | | | |
| bound | 15 | -23.2 | +18.4 | -20.8±2.4 | |
| unbound | 15 | -19.4 | +20.0 | -19.7±0.3 | |
| ΔΔG | | | | -1.1±2.4 | -0.5 |
| HAP4 →HAP1 | | | | | |
| bound | 15 | -26.2 | +28.5 | -27.5±1.2 | |
| unbound | 15 | -26.3 | +26.3 | -26.5±0.0 | |
| ΔΔG | | | | -1.0±1.2 | -1.6 |

# Convergence of simulations

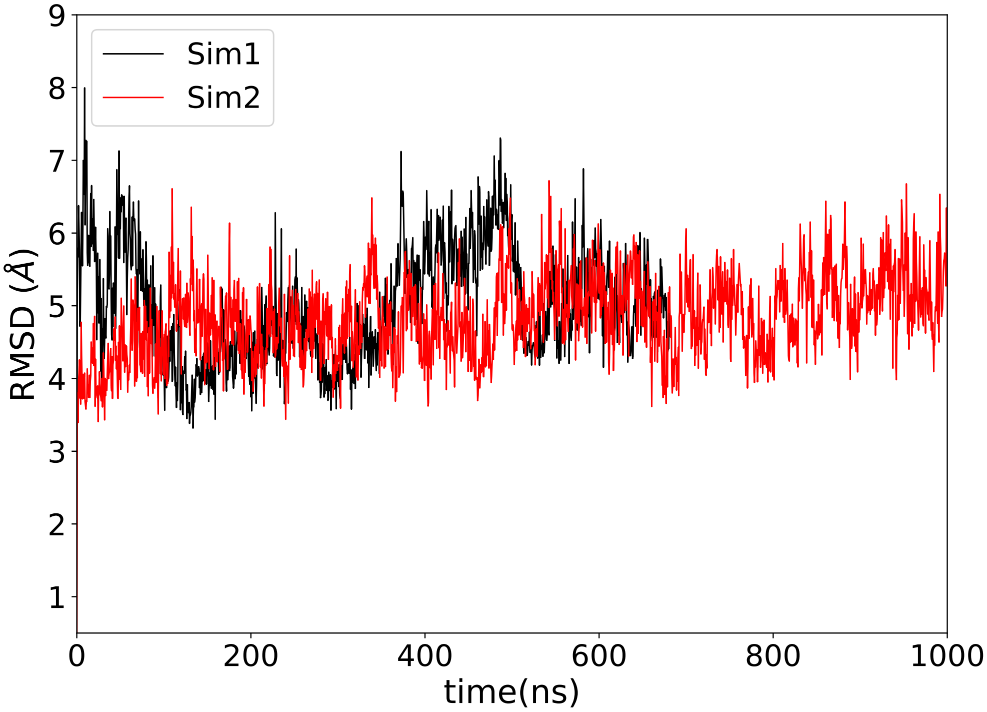

Figure S11: RMSD of the full Apo WT simulations.

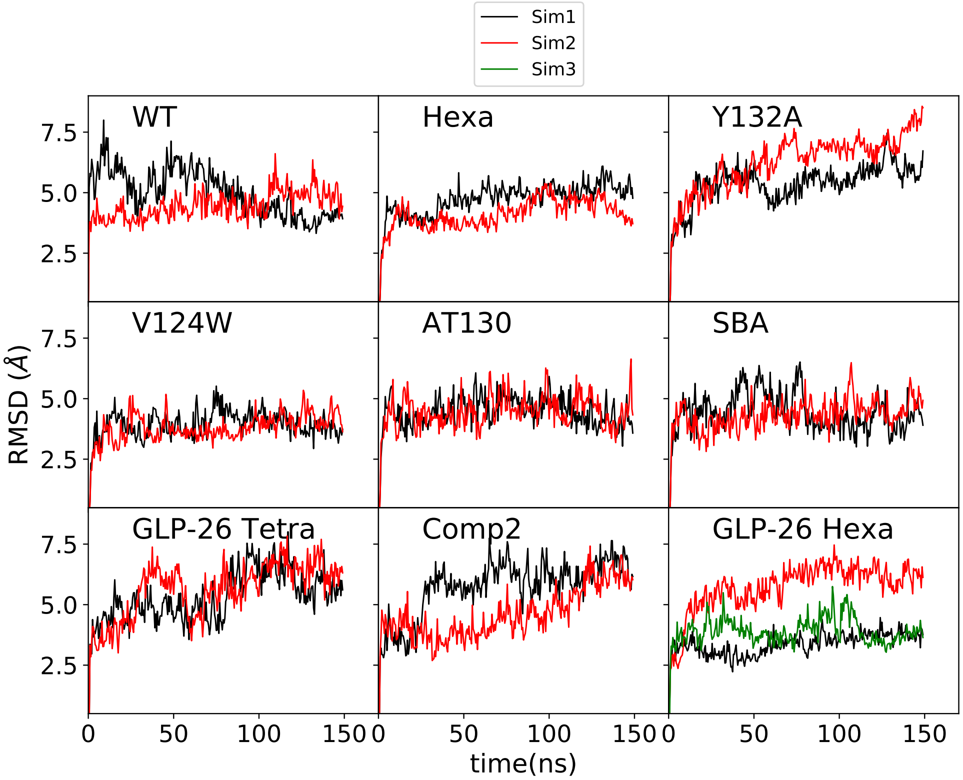

Figure S12: RMSD of the simulated systems. The WT Apo simulation is cut off at 150 ns for comparison on similar time scales. See Figure S11 for RMSD of the full system.

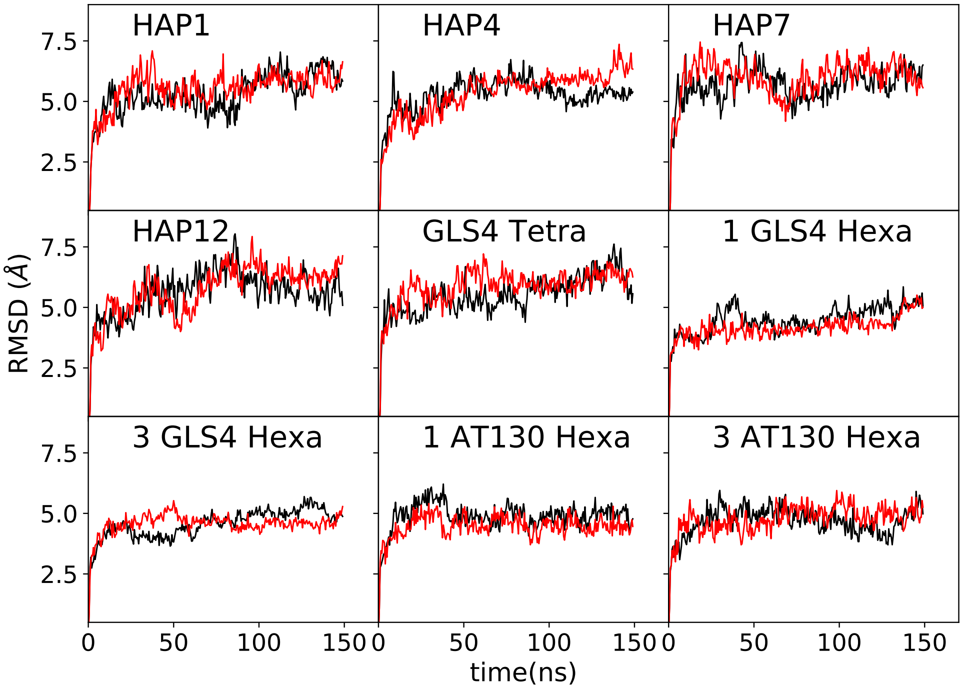

Figure S13: RMSD of additional simulated systems.

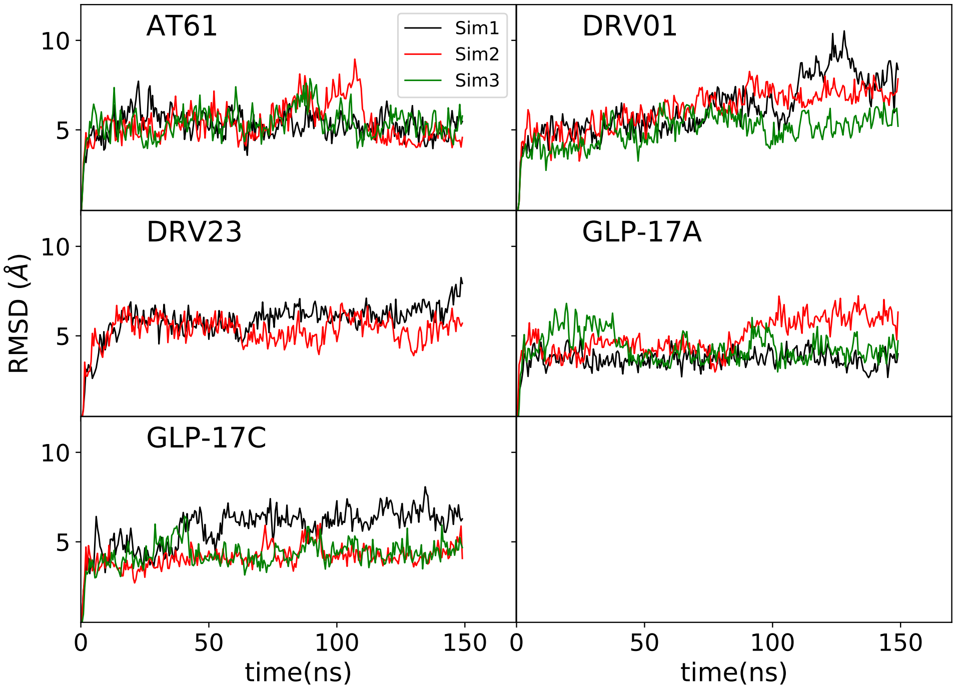

Figure S14: RMSD of additional simulated systems with bound tetramers.

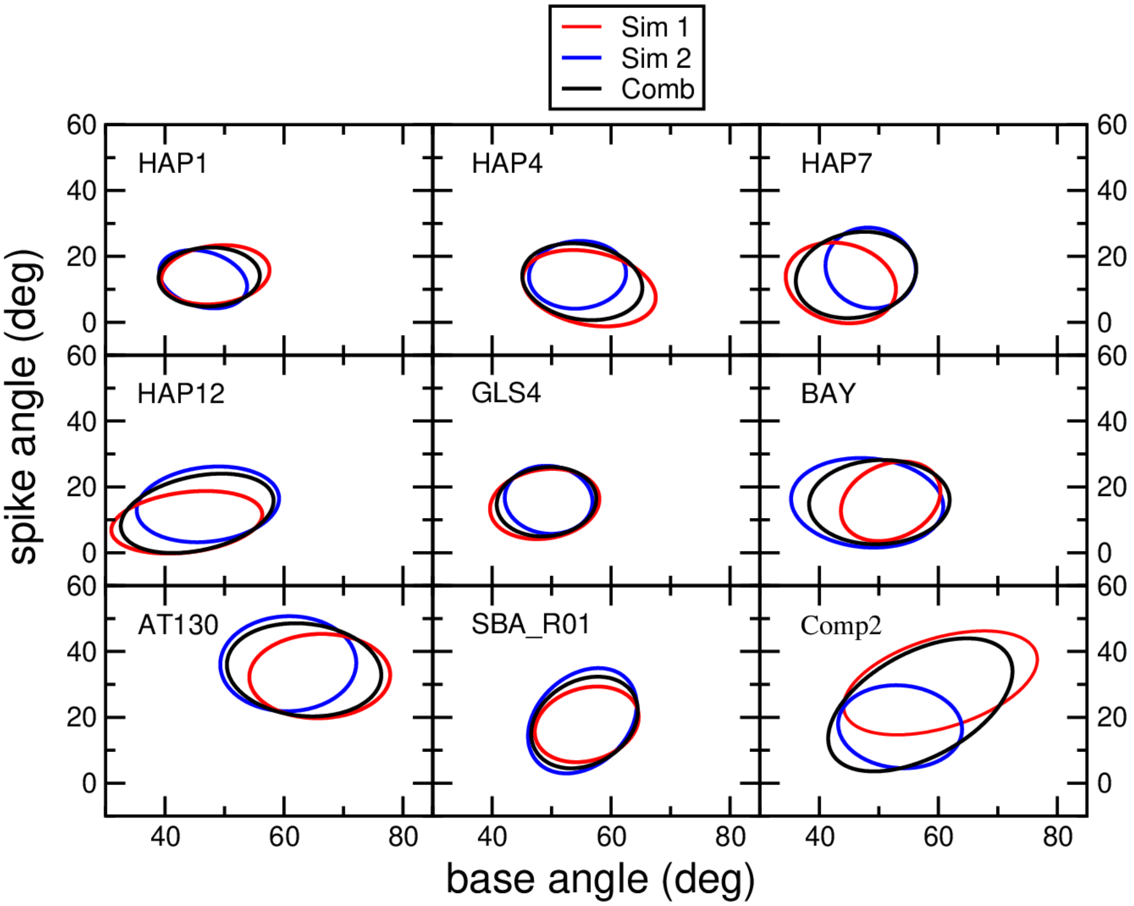

Figure S15: Comparison of standard deviation ellipses resulting from each of the two simulations to the one obtained from combining both simulations for the studied tetramers with bound drug compounds. Raw data is plotted in in Figure S19.

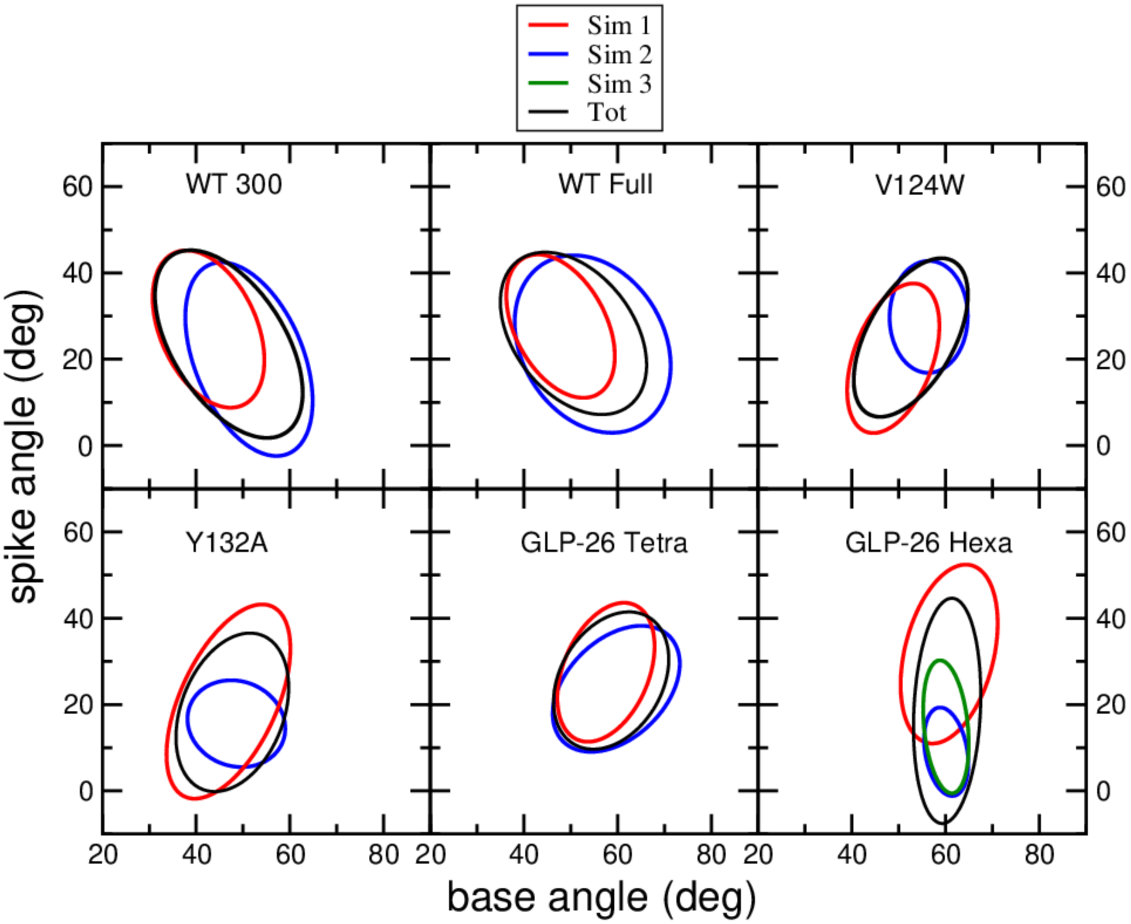

Figure S16: Comparison of standard deviation ellipses resulting from each of the two simulations to the one obtained from combining both simulations for the studied apo-tetramers and systems with bound compounds. The data is for tetramers unless specified otherwise. Raw data is plotted in in Figure S20.

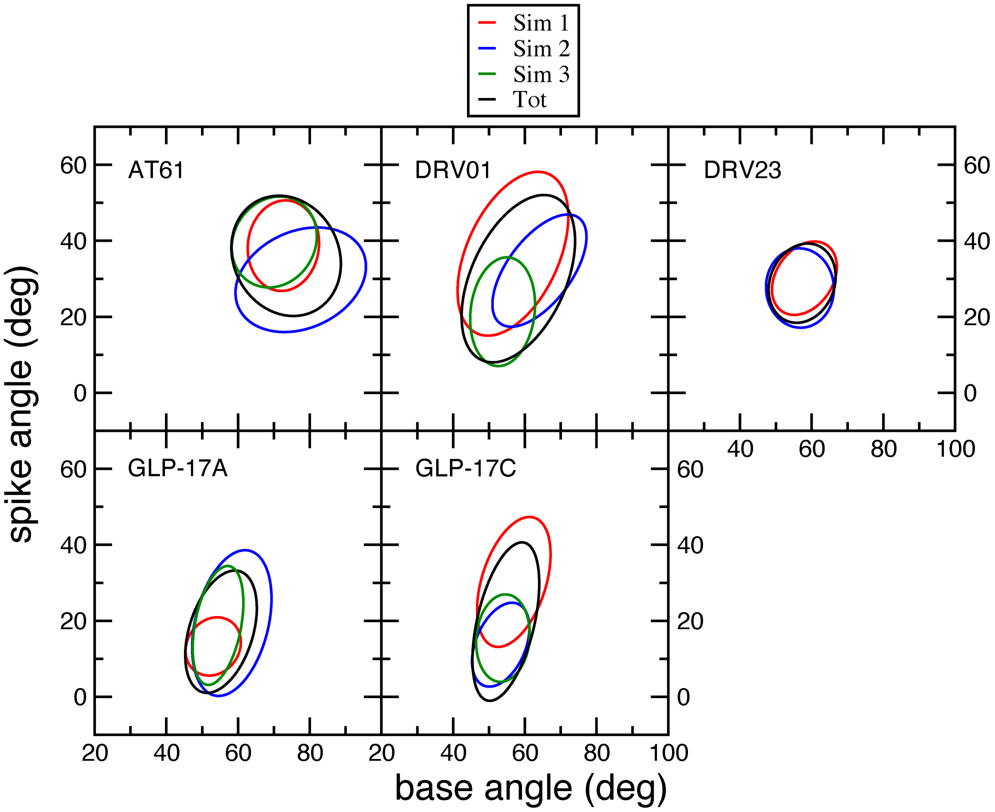

Figure S17: Comparison of standard deviation ellipses resulting from each of the two simulations to the one obtained from combining both simulations for the tetramers with bound novel compounds. Raw data is plotted in in Figure S21.

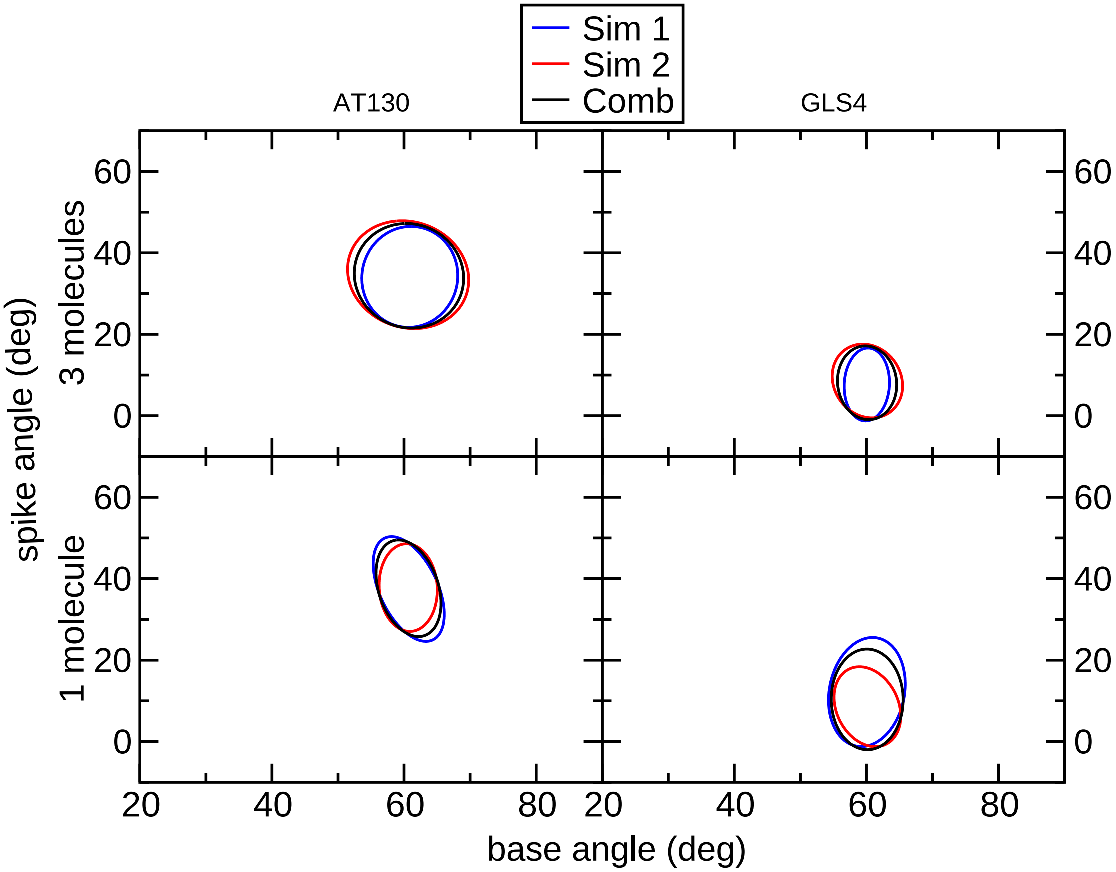

Figure S18: Comparison of standard deviation ellipses resulting from each of the two simulations to the one obtained from combining both simulations for hexamers with bound GLS4 and AT130. Top graphs display the results with three bound molecules. Bottom graphs show the results for hexamers with one bound molecule. Raw data is plotted in in Figure S22.

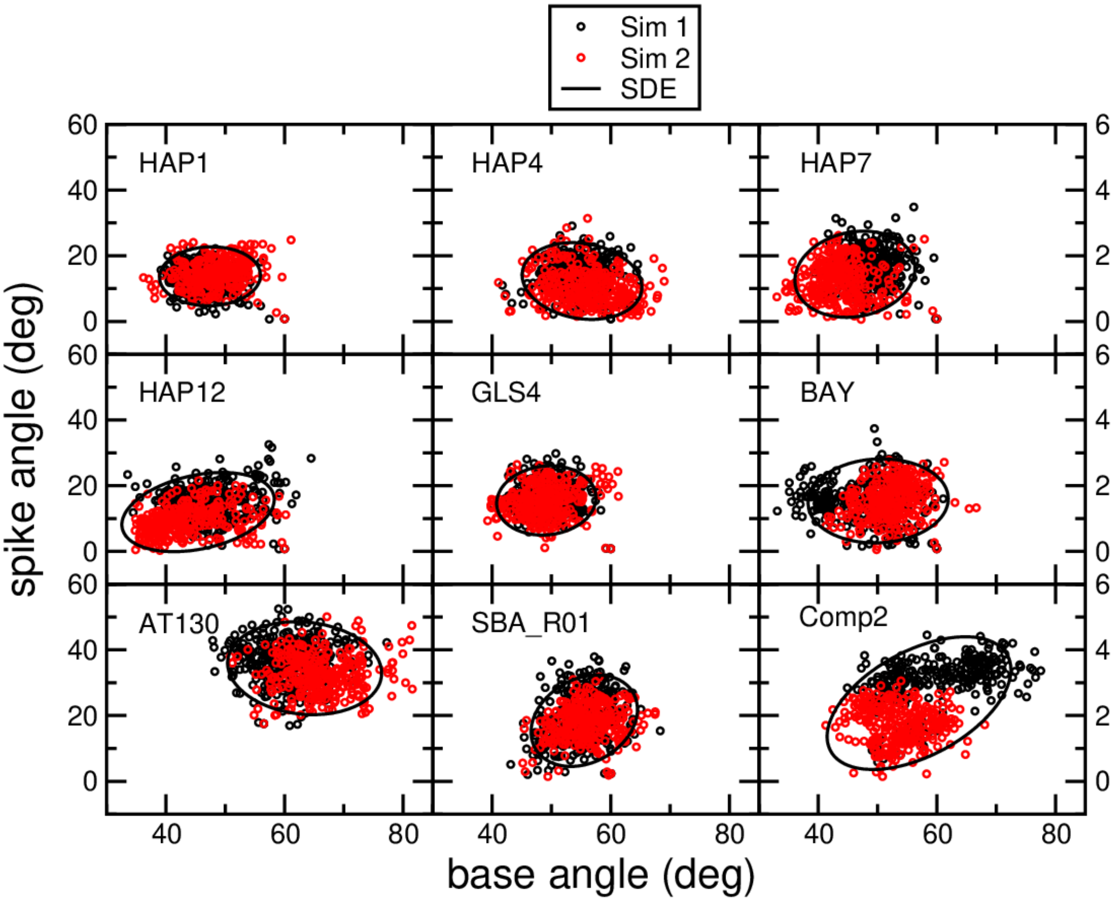

Figure S19: Distribution of spike and base angles and the resulting standard deviation ellipses for the studied tetramers with bound drug compounds.

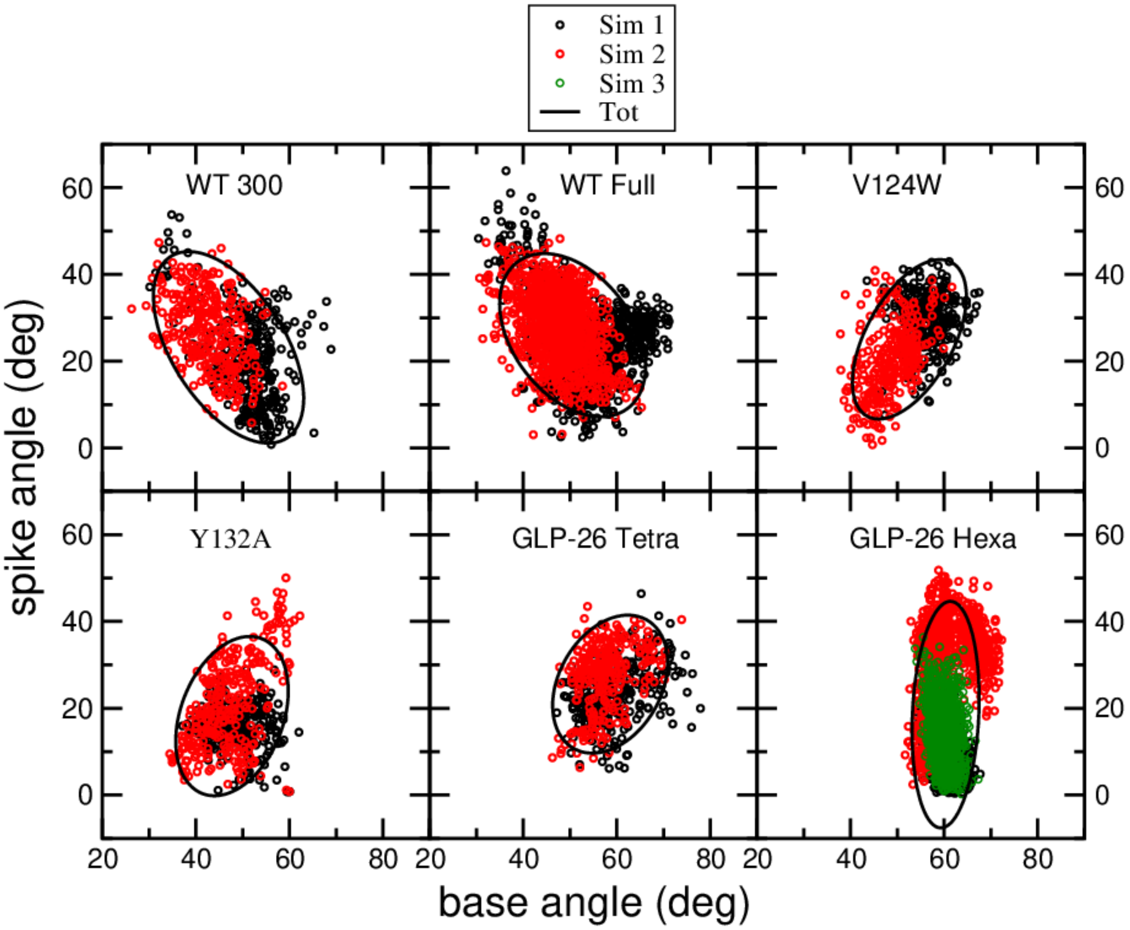

Figure S20: Distribution of spike and base angles and the resulting standard deviation ellipses for the studied apo-tetramers and systems with bound compounds. The data is for tetramers, unless specified otherwise.

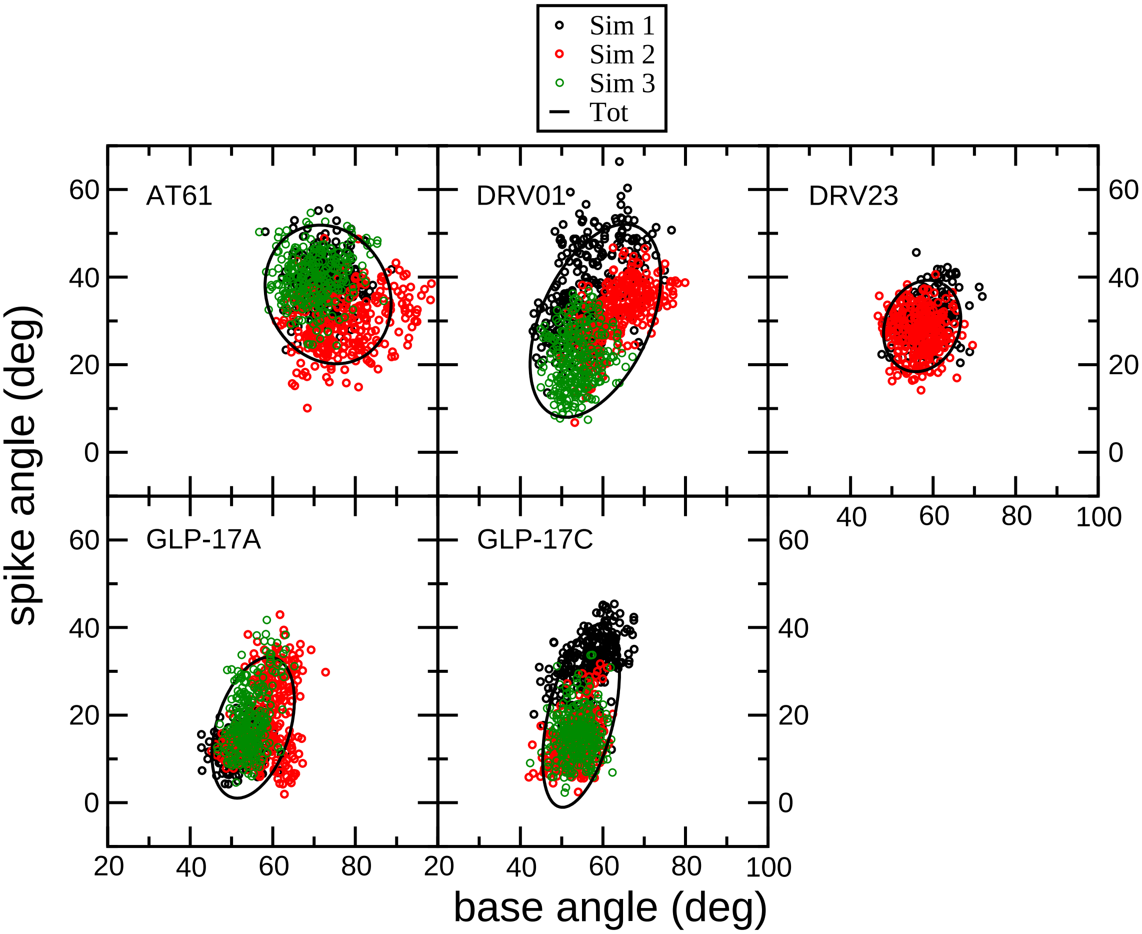

Figure S21: Distribution of spike and base angles and the resulting standard deviation ellipses for the studied tetramers with bound compounds.

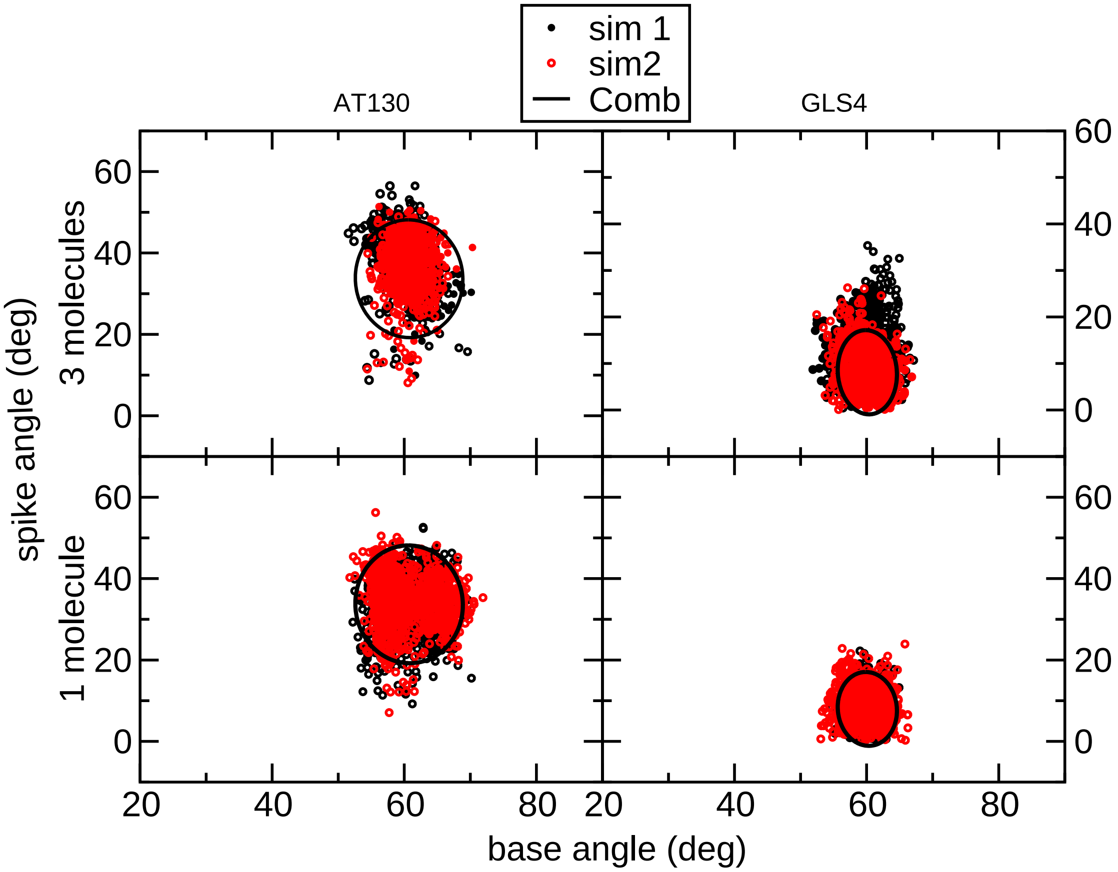

Figure S22: Distribution of spike and base angles and the resulting standard deviation ellipses for hexamers with bound GLS4 and AT130. Top graphs display the results with three bound molecules. Bottom graphs show the results for hexamers with one bound molecule.

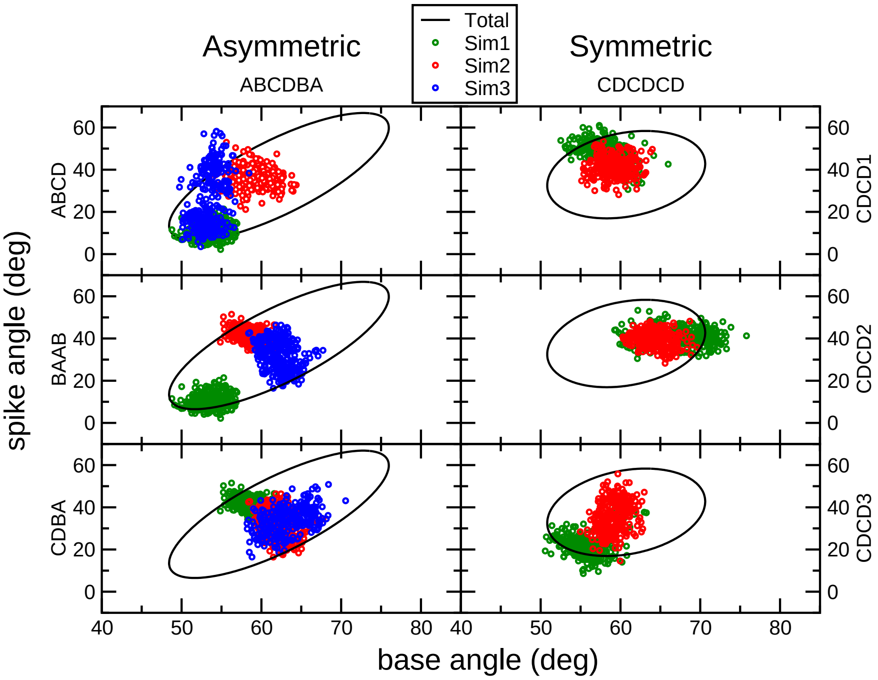

Figure S23: Distribution of spike and base angles for the different tetramers in the asymmetric (left graphs) and symmetric (right graphs) hexamer. Each graph shows the ellipses for a specific tetramer during different simulations and the ellipse resulting from averages of all tetramers and all simulations (black).

# References

(S1) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(S2) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell Jr., A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone $\phi$, $\psi$ and side-chain $\chi_1$ and $\chi_2$ dihedral angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(S3) 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.

(S4) 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.

(S5) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144–3154.

(S6) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM general gorce field (CGenFF) II: Assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155–3168.

(S7) 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.

(S8) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulations — The Langevin piston method. *J. Chem. Phys* **1995**, *103*, 4613–4621.

(S9) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(S10) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden thermodynamics of mutant proteins: a molecular dynamics analysis. *Science* **1989**, *244*, 1069–1072.

(S11) Dixit, S. B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin–streptavidin system revisited. *J. Phys. Chem. A* **2001**, *105*, 9795–9799.

(S12) Gangarapu, S.; Marcelis, A. T. M.; Zuilhof, H. Accurate p$K_a$ calculation of the conjugate acids of alkanolamines, alkaloids and nucleotide bases by quantum chemical methods. *ChemPhysChem* **2013**, *14*, 990–995.

(S13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.;

Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian˜16 Revision B.01. 2016; Gaussian Inc. Wallingford CT.

(S14) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

(S15) Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Computational thermochemistry: Scale factor databases and scale factors for vibrational frequencies obtained from electronic model chemistries. *J. Chem. Theory Comput.* **2010**, *6*, 2872–2887.

(S16) Venkatakrishnan, B.; Zlotnick, A. The structural biology of hepatitis B virus: Form and function. *Annu. Rev. Virol.* **2016**, *3*, 429–451.

(S17) Zandi, R.; van der Schoot, P.; Reguera, D.; Kegel, W.; Reiss, H. Classical nucleation theory of virus capsids. *Biophys. J.* **2006**, *90*, 1939–1948.

(S18) Hu, J.; Liu, K. Complete and incomplete hepatitis B virus particles: Formation, function, and application. *Viruses* **2017**, *9*, 56.

(S19) Wynne, S. A.; Crowther, R. A.; Leslie, A. G. W. The crystal structure of the human hepatitis B virus capsid. *Mol. Cell* **1999**, *3*, 771–780.

(S20) 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.