Supporting information for: Binding Modes of Ligands Using Enhanced Sampling (BLUES): Rapid Decorrelation of Ligand Binding Modes via Nonequilibrium Candidate Monte Carlo

Samuel C. Gill,[†] Nathan M. Lim,[‡] Patrick B. Grinaway,[¶,§] Ariën S. Rustenburg,[¶,§] Josh Fass,[‖,§] Gregory A. Ross,[§] John D. Chodera,[⊥] and David L. Mobley[*,‡,†]

†*Department of Chemistry, University of California, Irvine*
‡*Department of Pharmaceutical Sciences, University of California, Irvine*
¶*Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065*
§*Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065*
‖*Tri-Institutional PhD Program in Computational Biology and Medicine, New York, NY 10065*
⊥*Memorial Sloan Kettering Cancer Center, New York NY 10065*

E-mail: dmobley@mobleylab.org
Phone: 949-824-6383

This document contains supporting figures (SI Figures 1–5) referenced in the main text of this work. Specifically, Figure S1 shows work distributions for rotating toluene in lysozyme as a function of the amount of NCMC relaxation. Figure S2 shows the work standard deviation for toluene in lysozyme as a function of the amount of switching. Figure S3 shows the dihedral progress coordinate used for 3-iodotoluene. Figure S4 shows the estimated MSM transition matrix for toluene in lysozyme. Figure S5 shows acceptance of NCMC vs standard MC move proposals as a function of dihedral angle/binding mode, given a fixed ensemble of MD snapshots.

A separate supporting `.tar.gz` file is available, containing an extensive set of input files, scripts, and code which can be used to reproduce the calculations described in this work.

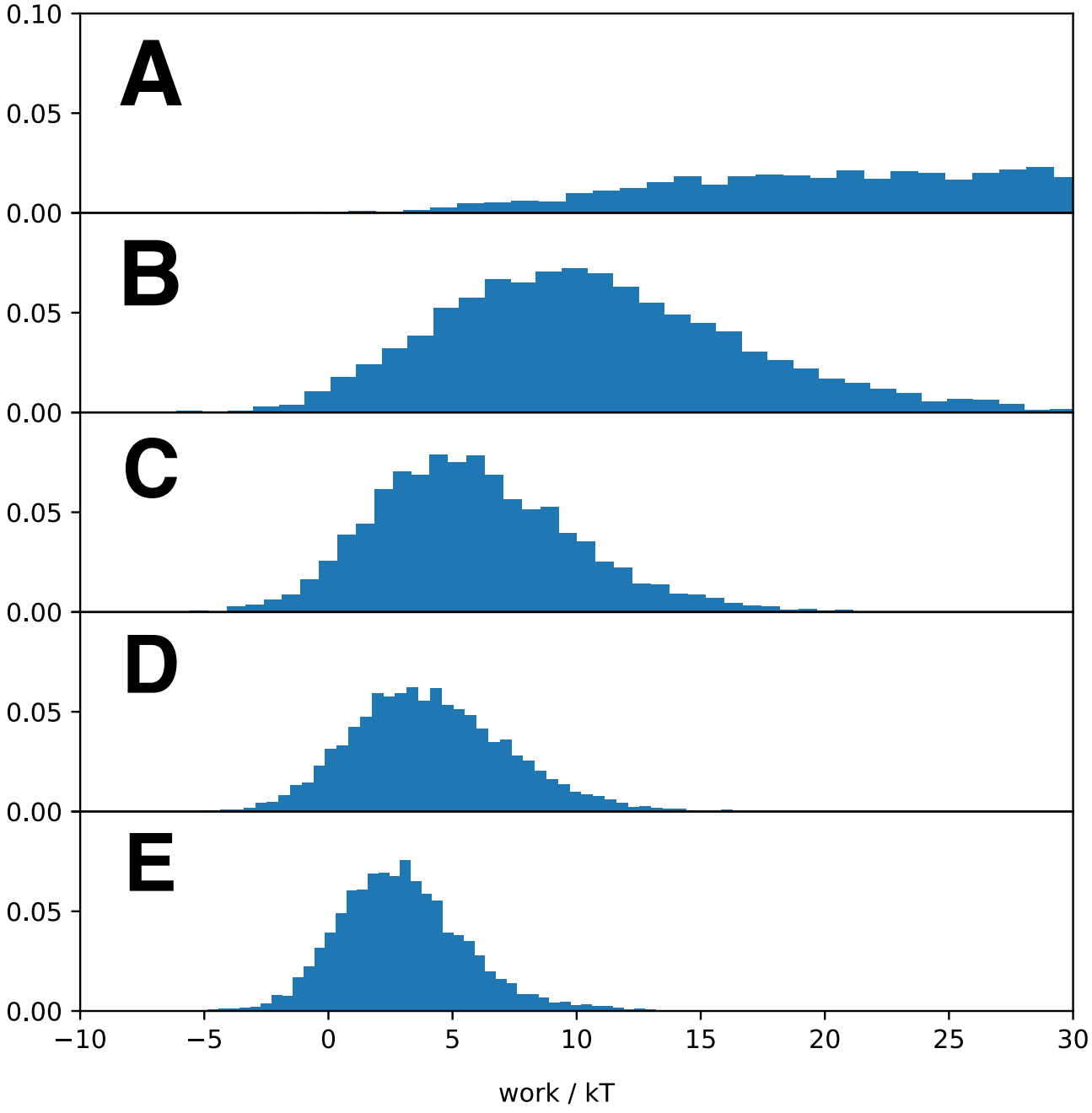
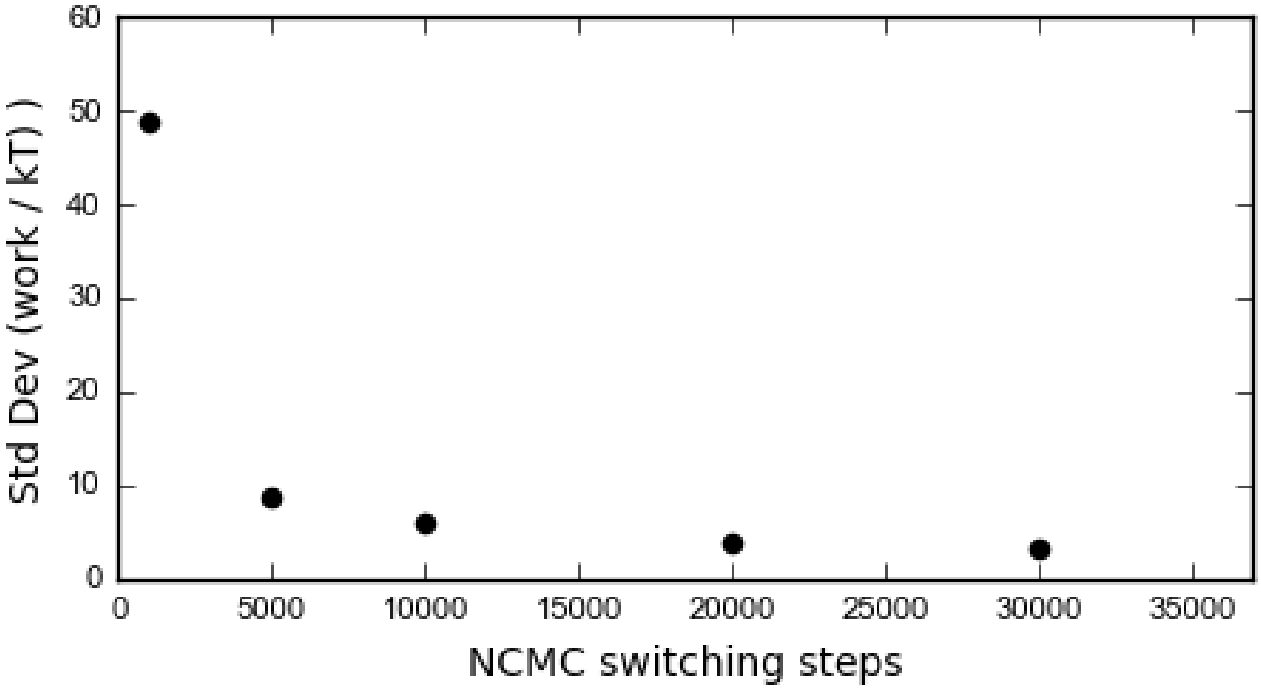


Figure S1: **Work distributions from increasing NCMC relaxation for rotation of toluene in T4 lysozyme.** The work distributions from 5000 NCMC+MD iterations of varying NCMC relaxation steps are plotted as a histogram over the range `[-10,30]`. A given histogram is over all the counts from that protocol. (*A*) Work distribution from 1000 NCMC steps. (*B*) Work distribution from 5000 NCMC steps. (*C*) Work distribution from 10000 NCMC steps. (*D*) Work distribution from 20000 NCMC steps. (*E*) Work distribution from 30000 NCMC steps. Increasing the number of relaxation steps increases the likelihood that a move will be accepted.



(a) MD/MC

Figure S2: **Work standard deviations from increasing NCMC relaxation for rotation of toluene in T4 lysozyme.** The standard deviation of the work distributions from 1000 NCMC+MD iterations of varying NCMC relaxation steps. As the number of relaxation steps increase the standard deviation also decreases, which is correlated with the probability of NCMC move acceptance.

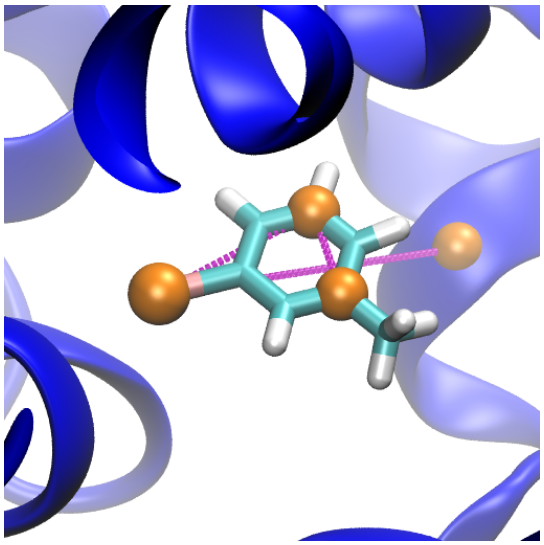


Figure S3: **Order parameters used for identifying binding modes for 3-iodotoluene in T4 lysozyme.** Shown is a depiction of the dihedral order parameter used to differentiate toluene's binding modes. The dihedral which we monitor is defined by the C1, C5, and I8 atoms of 3-iodotoluene and the alpha carbon of VAL111, shown in orange in CPK representation in orange. In the image, the atoms involved in the dihedral are connected by a purple line, and the dihedral angle measures rotation around the central dashed purple line. The protein is shown in a blue cartoon representation, and 3-iodotoluene is shown in cyan.

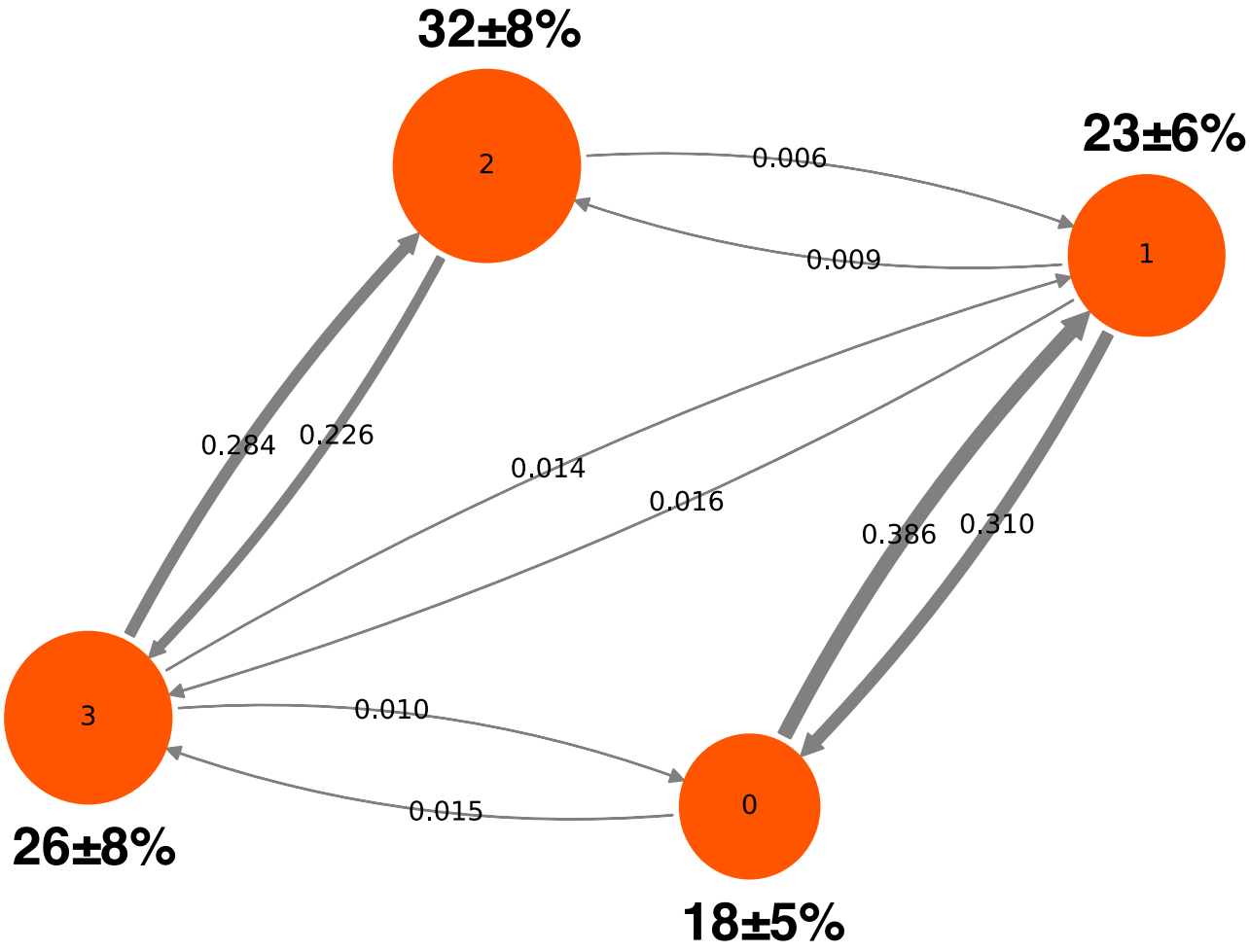
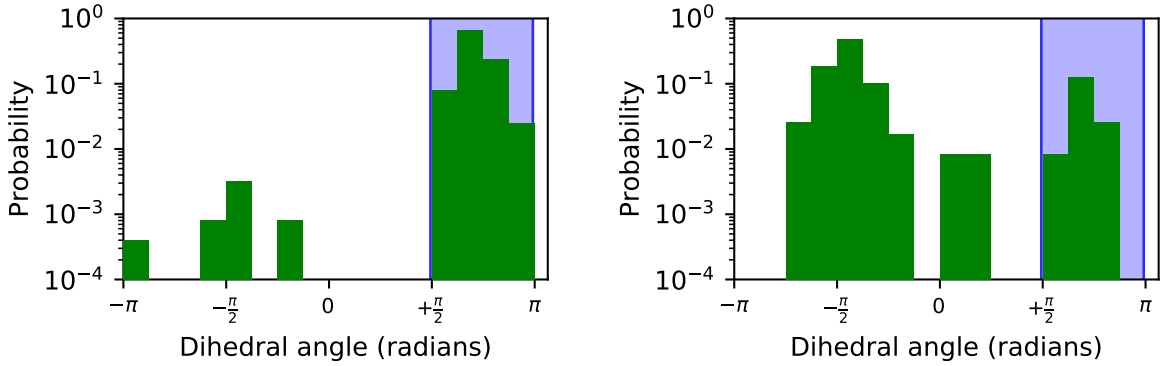


Figure S4: **Populations and transitions between MSM macrostates for toluene in lysozyme.** Visual representation of the MSM transition matrix and populations generated from Section 2.4 of the main text. The circles labeled with numbers represent separate macrostates, with the populations of each state given by the percentages above and below each circle. The arrows between them represent transition probabilities. Representative binding modes from a macrostate are pictured next to that macrostate.



(a) 3-iodotoluene rotations via MC

(b) 3-iodotoluene rotations via NCMC

Figure S5: **Acceptance of conventional MC move proposals versus NCMC move proposals proposed from a fixed set of configurations of 3-iodotoluene in T4 lysozyme L99A.** Shown is a comparison of the probability of accepting moves sampling a given ligand binding mode (monitored by a dihedral angle progress coordinate) for 3-iodotoluene in T4 lysozyme L99A for MC move proposals (left plot) and NCMC move proposals (right plot), giving the same ensemble of MD snapshots as a starting point for move proposals. Starting MD snapshots have the ligand in a binding mode in the blue region, but accepted moves involve random rotations of the ligand and thus can be to any binding mode/dihedral angle. In other words, each MC or NCMC trial starts from a selected MD snapshot from within the blue region and, if the move is accepted, the final dihedral angle is computed and a counter is incremented which is used to compute the probabilities on the vertical axis. The MC panel (a) shows data from ten trials of 2,000,000 MC attempts with an overall acceptance rate of $(1.2 \pm 0.2) \times 10^{-2}$ %, but significant moves (larger than 20 degrees) accepted at a rate of only ($(5 \pm 2) \times 10^{-5}$ %). The NCMC panel (b) shows data from seven trials (denoted by dashed vertical lines) of 2000 move attempts, with each move attempt consisting of 6500 NCMC switching steps. Here, the overall acceptance rate is 0.8 ± 0.1%, with moves larger than 20 degrees accepted at a rate of 0.68 ± 0.07 %. In (a), for MC, we observed a total of 13 significant rotations, whereas in (b), in the equivalent number of force evaluations we observed 24 (though the data shown here represents a much larger number of force evaluations for better statistics). In the MC case, because moves are instantaneous, very few significant moves outside the blue region are accepted, giving an (apparently false) impression that the binding mode in the initial green region is by far the more favorable binding mode. In contrast, in the NCMC case, because NCMC allows relaxation of the environment, *most* of the accepted moves are significant and outside the green region, indicating that the alternate binding mode is in fact likely to be more favorable.

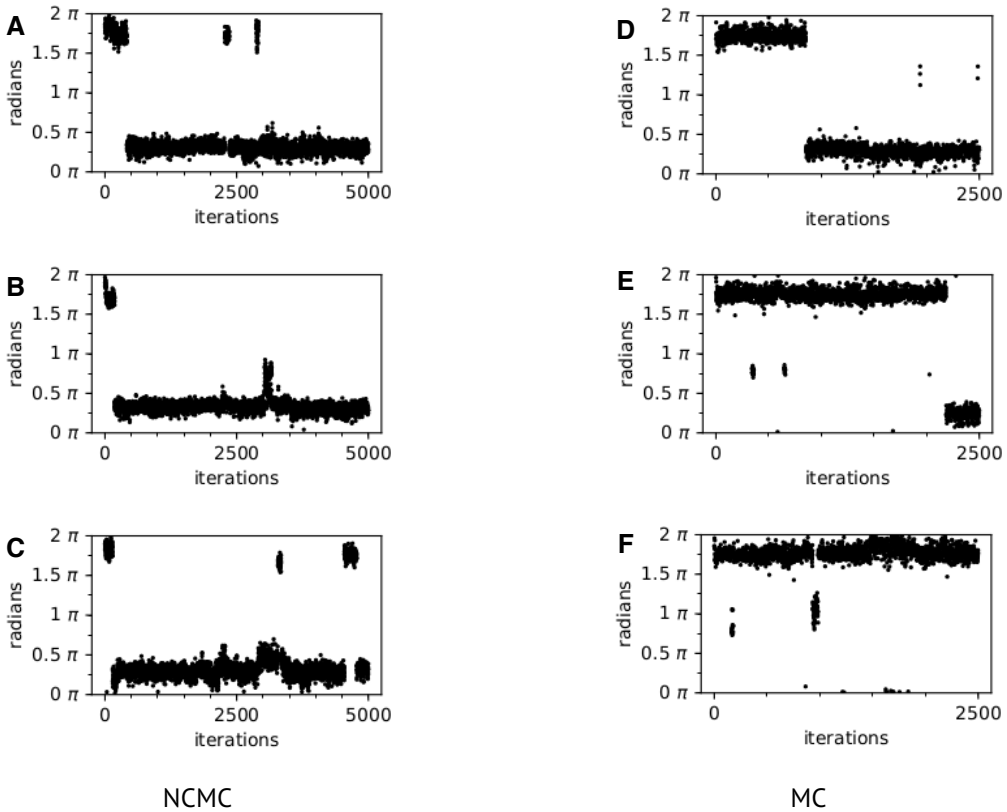


Figure S6: **3-iodotoluene trajectory comparisons between NCMC and MC as a function of a dihedral progress coordinate.** (*a,b,c*) Dihedral angle (corresponding to binding modes) observed with NCMC as a function of simulation time. (*d,e,f*) Dihedral angle (corresponding to binding modes) observed in the MC as a function of simulation time. Each iteration consists of the same number of energy evaluations, with either 6500 NCMC switching steps or 6500 MC attempts, followed by 10,000 steps of MD. For MC we simulated for 2500 iterations, while for NCMC we simulated for 5000 iterations The periodic dihedral plotted here spans from [0,2π]. The NCMC simulations show more consistent behavior between simulations compared with MC; each NCMC simulation transitions to another stable binding mode (around 1 rad) within 1000 iterations, while the MC simulations sometimes fail to transition to this binding mode at all within 2500 iterations. The overall success rate (per energy evaluation) for transitioning to alternate binding modes appears roughly comparable between the two cases.

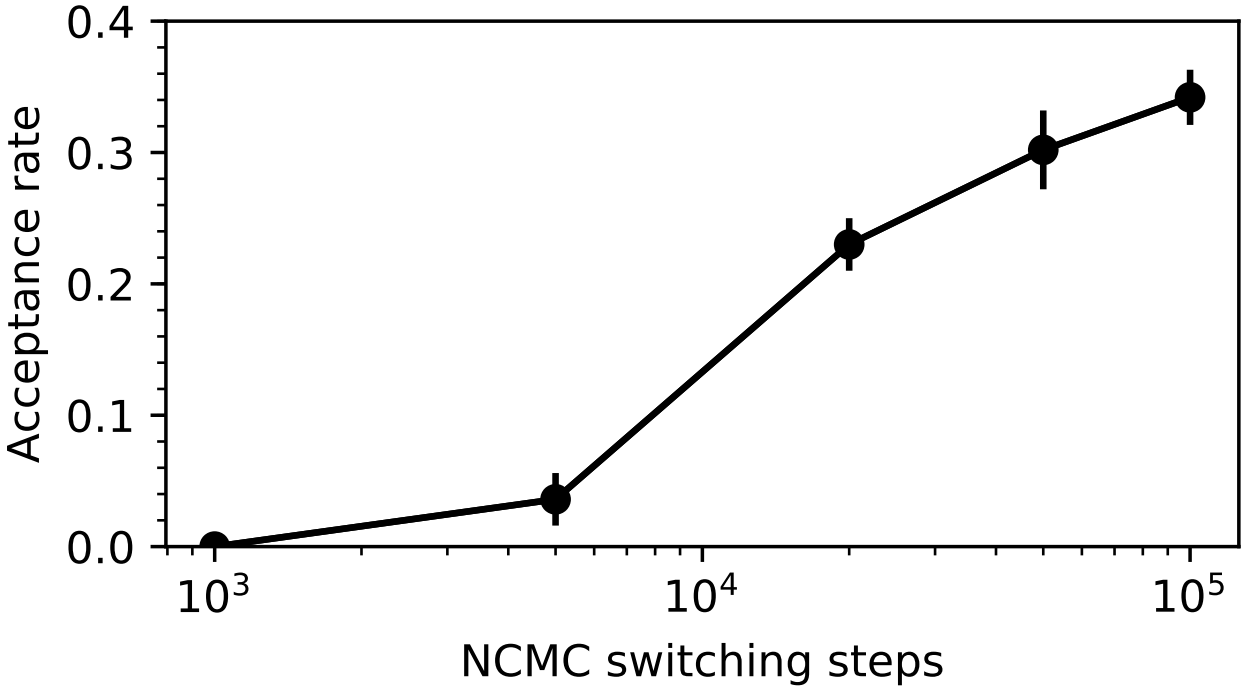


Figure S7: **Acceptance probability for toluene in solution as a function of the amount of NCMC relaxation.** The acceptance probability of turning toluene's steric and electrostatics off, followed by a random rotation and turning back on toluene's interactions is shown here as a function of the NCMC switching steps. At 1000 switching steps these moves show no acceptance, but the acceptance rates increase with increasing switching steps, up to 32%$\pm$2% for 100,000 steps. For each switching step, the uncertainty was calculated based on blocking of 500 BLUES iterations, consisting of the number of switching steps and 100 steps of MD. The number of blocks used was the amount that maximized the standard deviations of the acceptance rate across blocks.