Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.


**Supplementary Information** for "Calculating binding free energies of host-guest systems using AMOEBA polarizable force field"

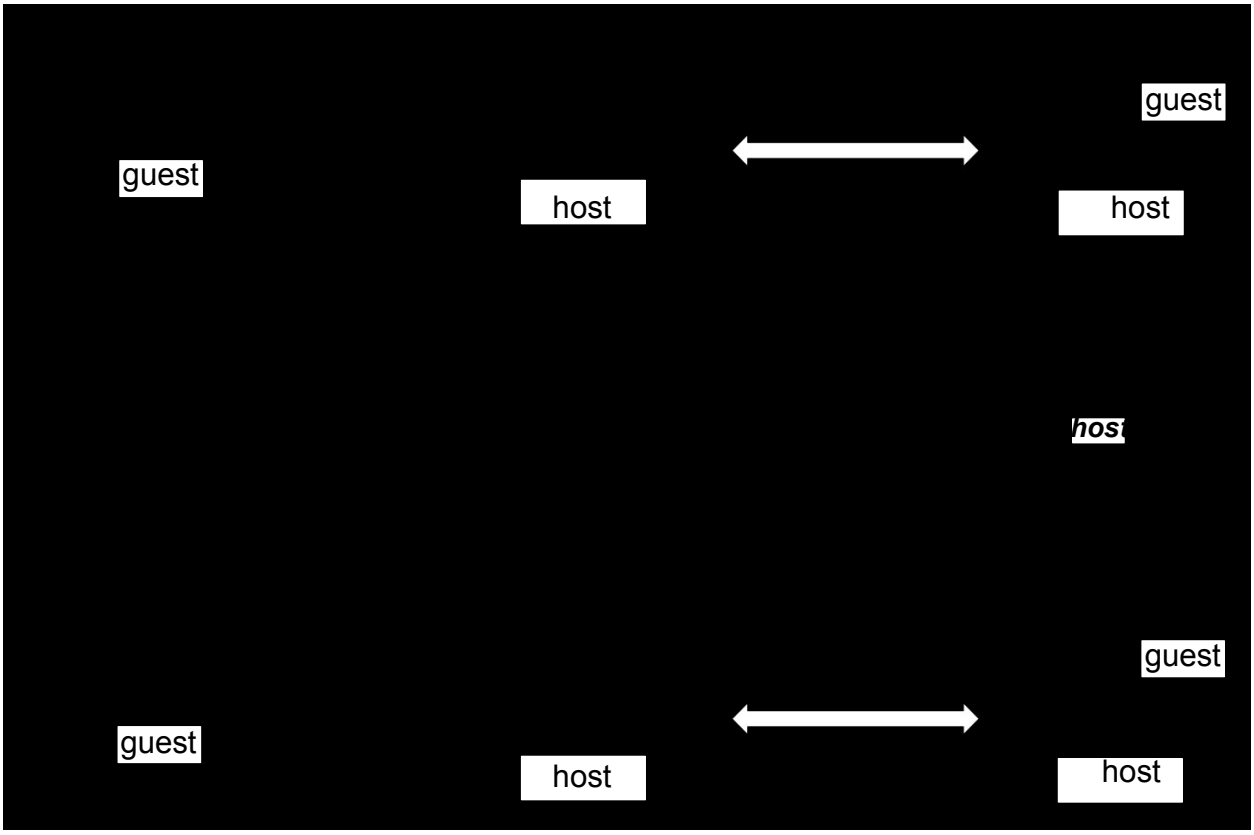


**Figure S1.** Thermodynamic cycle for calculating the binding free energy of the host-guest system. The binding free energy ($^{\Delta G_{bind}}$) is defined as the difference between the decoupling free energies from both solvent and solvated protein complex. $^{\Delta G_{host}}$ indicates that the ligand is decoupled from its protein environment, and $^{\Delta G_{hyd}}$ indicates that the ligand is removed from a water environment.

**Table S1.** Free energy composition of host-guest systems. All guests bind to the cucurbit[7]uril host. All free energies are in kcal/mol. For BAR and OSRW, $\Delta G_{bind} = \Delta G_{host-guest} - \Delta G_{hydration} + G_{correction}$, where $G_{correction} = 6.245$ kcal/mol.

| Guest | Host-guest Free Energy ($\Delta G_{host-guest}$) | | | Hydration Free Energy ($\Delta G_{hydration}$) | | | Binding Free Energy ($\Delta G_{bind}$) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | OSRW 1 | OSRW 2 | BAR | OSRW 1 | OSRW 2 | BAR | OSRW 1 | OSRW 2 | BAR | Experiment[f] |
| C1 | -189.58 | -188.11 | -189.19 | -170.42 | -170.59 | -170.67 | -12.91 | -11.18 | -12.27 | -9.90 |
| C1_LGH[a,b] | -187.87 | -188.20 | – | – | – | – | -11.21 | -11.26 | – | |
| C2 | -70.21 | -70.57 | -70.20 | -56.87 | -56.84 | -57.50 | -7.09 | -7.48 | -6.46 | -9.60 |
| C3 | -65.43 | -64.33 | -71.29 | -52.42 | -52.71 | -58.45 | -6.76 | -5.38 | -6.59 | -6.60 |
| C3_2[c] | -63.87 | -65.42 | – | -52.23 | -52.93 | – | -5.39 | -6.25 | – | |
| C3_LGH[a] | -67.78 | -64.75 | – | -51.28 | -52.28 | – | -10.26 | -6.22 | – | |
| C4 | -180.21 | -180.12 | -181.27 | -164.19 | -163.61 | -163.68 | -9.77 | -10.26 | -11.34 | -8.40 |
| C5[d] | -68.64 | -67.61 | -68.01 | -57.64 | -58.10 | -58.39 | -4.75[g] | -3.26[g] | -3.37 | -8.50 |
| C5b[d] | -208.76 | -207.18 | -213.19 | -199.57 | -200.50 | -203.55 | -2.94[g] | -0.43[g] | -3.39 | |
| C5b_LGH[a,d] | -211.84 | -210.32 | – | -200.48 | -199.18 | – | -5.11 | -4.89 | – | |
| C6 | -77.00 | -76.25 | -76.04 | -63.39 | -63.34 | -63.62 | -7.36 | -6.66 | -6.18 | -7.90 |
| C7 | -74.25 | -74.54 | -74.69 | -58.04 | -58.15 | -57.96 | -9.96 | -10.14 | -10.49 | -10.10 |
| C8 | -73.59 | -73.48 | -73.77 | -55.81 | -55.89 | -55.68 | -11.53 | -11.35 | -11.84 | -11.80 |
| C9 | -77.98 | -77.60 | -77.98 | -55.86 | -56.78 | -56.32 | -15.87 | -14.57 | -15.42 | -12.60 |
| C9_LGH[a,e] | – | – | – | -56.07 | -56.05 | – | -15.66 | -15.31 | – | |
| C10 | -182.05 | -186.04 | -186.17 | -175.04 | -175.11 | -174.86 | -0.76[g] | -4.68 | -5.06 | -7.90 |
| C10_2[c] | -184.24 | -183.61 | – | -173.67 | -174.09 | – | -4.32 | -3.27 | – | |
| C10_LGH[a] | -183.17 | -182.70 | – | -173.48 | -172.69 | – | -3.45 | -3.76 | – | |
| C11 | -71.68 | -71.19 | -71.69 | -55.12 | -55.14 | -54.97 | -10.31 | -9.81 | -10.48 | -11.10 |
| C12 | -69.23 | -69.89 | -68.91 | -50.39 | -51.10 | -50.56 | -12.59 | -12.54 | -12.11 | -13.30 |
| C13 | -71.08 | -70.59 | -71.56 | -51.58 | -51.34 | -51.39 | -13.25 | -13.00 | -13.92 | -14.10 |
| C14 | -79.05 | -77.97 | -78.05 | -59.47 | -59.46 | -59.39 | -13.33 | -12.16 | -12.41 | -11.60 |

[a] For all LGH ligand simulations, the height of the Gaussian bias (preventing OSRW from resampling that point) was lowered in order to promote more stable sampling. Not applicable to BAR simulations. [b] For C1_LGH, the host-guest free energies were coupled with C1 hydration free energies to compute the binding free energy. [c] C3_2 and C10_2 were independent OSRW simulations with the same

system and parameters as C3 and C10. [d] C5 and C5b, are different protonation states of the C5 ligand. [e] For C9_LGH, the hydration free energies were coupled with C9 host-guest free energies to compute the binding free energy. [f] All experiment values hold absolute uncertainties of 0.1 kcal/mol. [g] These values were not included in the average for the reported results.

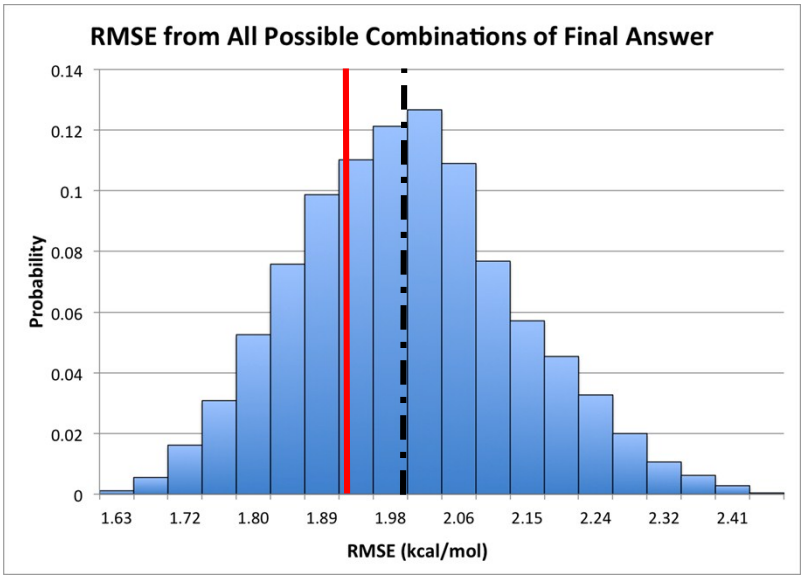


**Figure S2.** Probability distribution of RMSE between calculated and experimental results from all possible OSRW answer combinations across all ligands. Solid red line represents the reported value in Table 2 of the main text, while the black dot-dashed line represents the mean value.

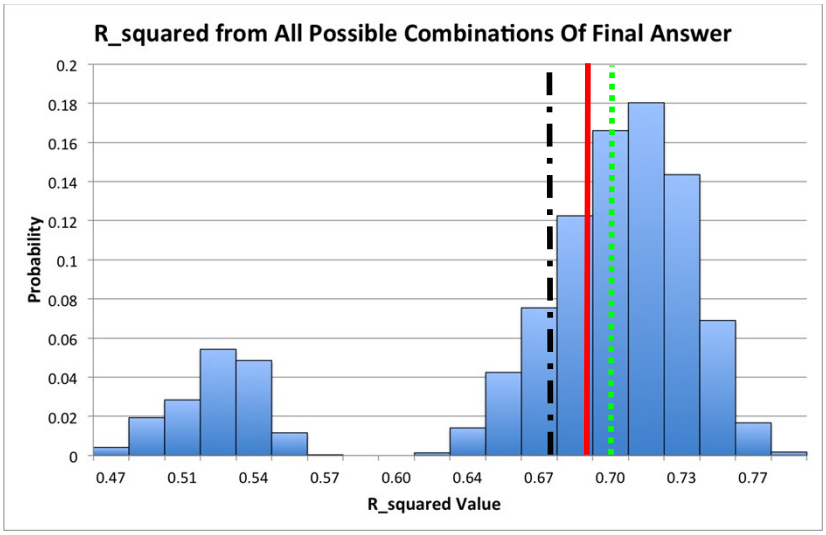


**Figure S3.** Probability distribution of $R^2$ correlation coefficient from all possible OSRW answer combinations. Solid red line shows reported $R^2$ value. The black dot-dashed line represents the location of the mean while the green dotted line represents the median.

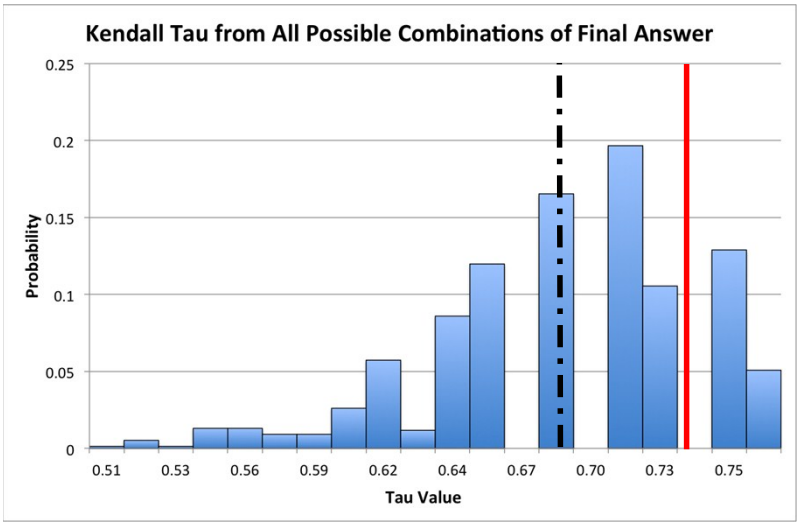


**Figure S4.** Probability distribution of Kendall τ correlation coefficient from all possible OSRW answer combinations. Solid red line shows the approximate location of the reported τ value while the black dot-dashed line represents the mean.

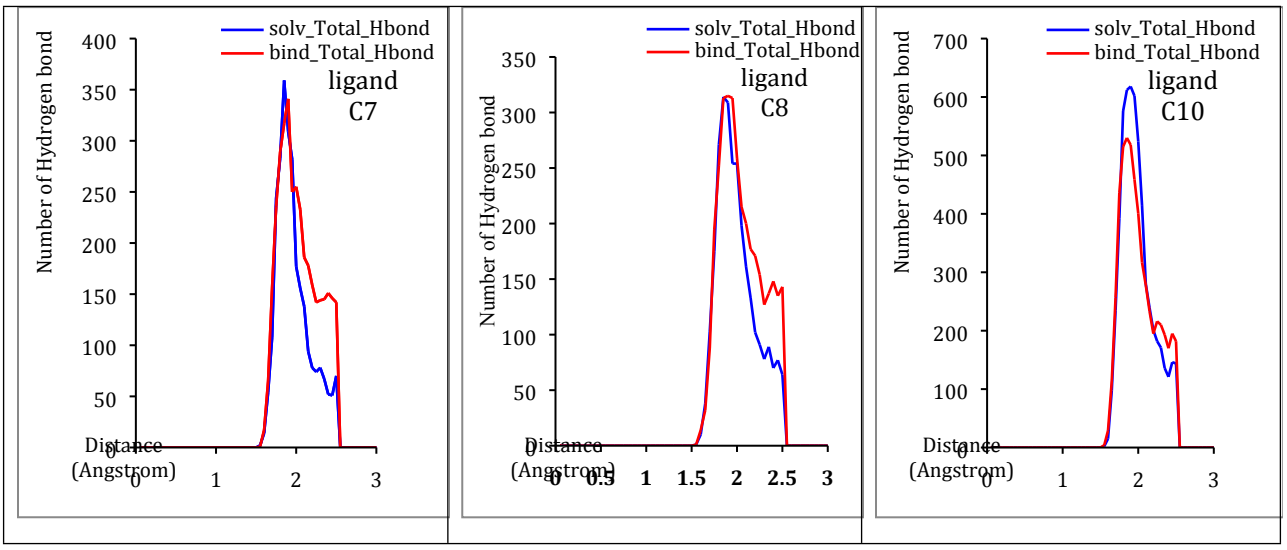


**Figure S5.** Plots of the total number of hydrogen bonds for ligand C7, C8 and C10 between guest and water in solution (solv_Total_Hbond) and between guest and host/water in host-guest complex (bind_Total_Hbond).

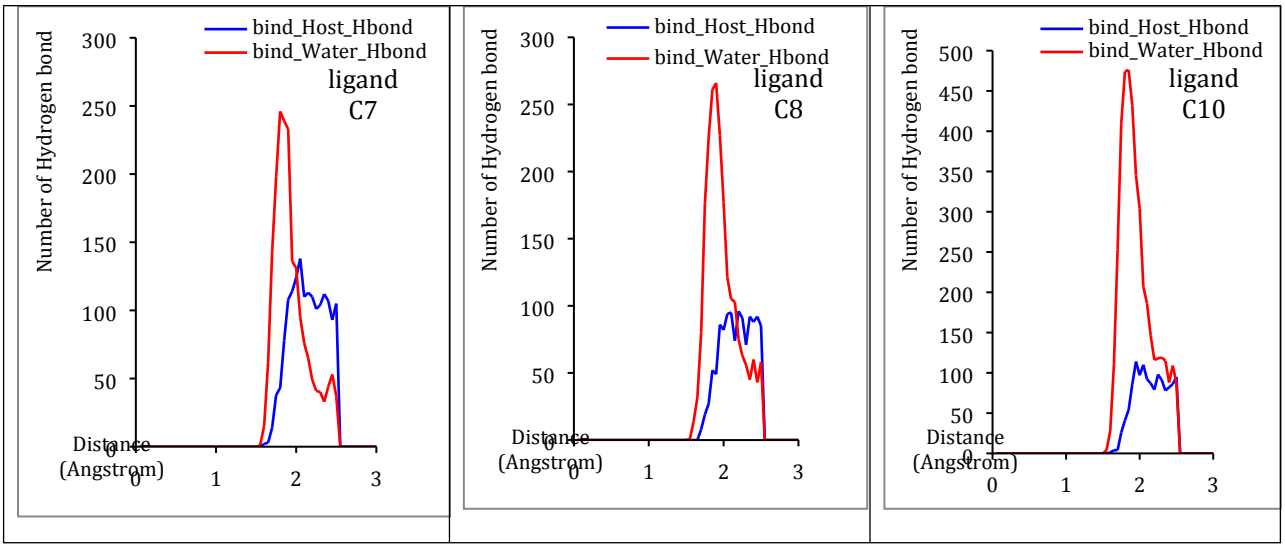


**Figure S6.** Plots of hydrogen bond numbers between guest and host (bind_Host_Hbond), and between guest and water (bind_Water_Hbond) in host-guest complex.

**Table S2.** Rotational entropy computed from PMF curves.

| Guest ligand | $-T\Delta S$ (kcal/mol)[a] |
|---|---|
| C7 | 0.042 |
| C8 | 0.053 |
| C10 | 0.078 |

[a] $\Delta S = -k_B \Sigma p \ln(p) - S_{reference}$, where $p$ is the probability taken from the PMF distribution and $k_B$ is Boltzmann's constant. $S_{reference}$ is the entropy of the unbound state, taken as a uniform PMF distribution over the rotation angle bins, $S_{reference} = -k_B \sum \frac{1}{n} \ln\left(\frac{1}{n}\right)$, where $n$ is the number of bins. $T = 300$K.

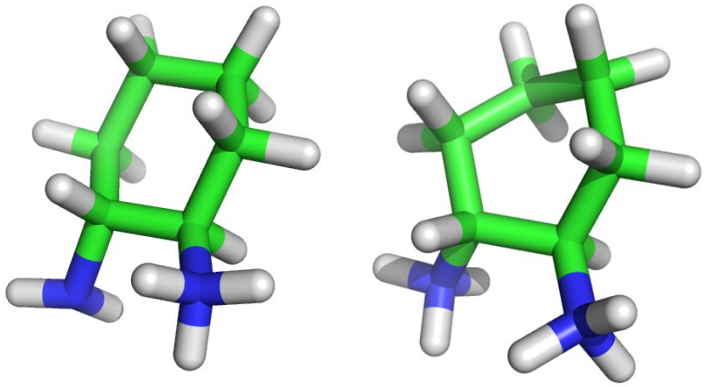


**Figure S7.** Structures of different protonation states of guest C5: C5 and C5b.