Supporting Information

A Linear Interaction Energy Model for Cavitand Host-Guest Binding Affinities

Joel José Montalvo-Acosta, Paulina Pacak, Diego Enri Barreto Gomes and Marco Cecchini*

Laboratoire d'Ingénierie des Fonctions Moléculaires
UMR7177 CNRS, Université de Strasbourg,
F-67083 Strasbourg Cedex, France

June 1, 2018

1 Computational details

1.1 Preparation of the systems

The initial atomic coordinates for hosts CB7, OAH and OAM and their guests were obtained from free-available online publications of the blind challenges SAMPL4,[1] SAMPL5[2] and HYDROPHOBE.[3] Initial geometries for the BCD and CB8 hosts were obtained from the Cambridge Structural Database[4] and their guests where modeled from SMILES. The protonation state of the hosts and guests were assigned using the Marvin suite software[5] at the experimental pH; the CB7, CB8, and BCD hosts were modeled as neutral, the OAH and OAM hosts had a net charge of -8, the charge of the guests is shown in Figures S1, S3, S4, S5, and S8. Initial atomic coordinates for the host-guest complexes were extracted from the top ranking binding mode predicted by docking using the CHEMPLP scoring function implemented in PLANTS.[6] Force field parameters for hosts and guests were generated using the General Amber Force Field (GAFF)[7] with AM1-bcc charges.[8] Also, to assess the impact of the force field on the binding affinity predictions, MD simulations were performed using the CHARMM general Force Field (CGenFF).[9]

*Corresponding author: mcecchini@unistra.fr

1.2 Molecular dynamics simulations

All Molecular Dynamics (MD) simulations were carried out using GROMACS 5.1.2[10] using periodic boundary conditions and a time step of 2 fs. Each molecular system was solvated in a cubic box with a minimum distance of 1.4 nm between the solute and the edge of the box. The TIP3P model was used to represent water molecules and counterions were added to grant neutrality of the simulation box. Electrostatic and van der Waals interactions were computed using particle mesh Ewald (PME) with a real-space cut-off of 1.2 nm, a grid spacing of 0.12 nm, a spline order of 4 and a relative tolerance of 1×10^{-6} for the reciprocal space. The LINCS algorithm was used to constrain all covalent bonds involving hydrogens. The simulation protocol started with an energy minimization of 10000 steps of steepest descent until a maximum force of $10\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ was attained. The system was then slowly heated to the target temperature (i.e., 298 K) using a modified Berendsen thermostat[11] in 6 steps with increments of 50 K every 50 ps. The system was equilibrated for 1 ns at constant volume and for another 1 ns at the constant pressure of 1 bar. The first 500 ps of simulation were carried out using the Berendsen barostat,[12] the remaining 500 ps using the Parinello-Rahman barostat,[13] which grants correct sampling of the NPT canonical ensemble. In both cases, a barostat coupling parameter of 1 ps and a isothermal compressibility of $4.5 \times 10^{-5}\,\mathrm{bar^{-1}}$ were used. Finally, a production phase of 20 ns was performed at constant temperature and pressure. Molecular configurations were saved every 5 ps for further analysis. The MD simulations with CGenFF parameters were performed using a soft harmonic restrain on the distance between the centers of mass (COMs) of the host and the guest using PLUMED 2.3.0[14] with a force constant of $10\,\mathrm{kcal\,mol^{-1}\,\AA^{-2}}$ to avoid spontaneous unbinding, which was observed for several host-guest complexes in unbiased MD.

1.3 Linear interaction energy model

In the linear interaction energy (LIE) method,[15] the ligand binding affinity is evaluated as

$$\Delta G_b^\circ = \beta \left[\left\langle U_{L-s}^{elec} \right\rangle_b - \left\langle U_{L-s}^{elec} \right\rangle_{ub} \right] + \alpha \left[\left\langle U_{L-s}^{vdw} \right\rangle_b - \left\langle U_{L-s}^{vdw} \right\rangle_{ub} \right] \tag{1}$$

where $\langle U_{L/s}^{elec} \rangle_b$ and $\langle U_{L/s}^{vdw} \rangle_b$ are ensemble averages of the electrostatic and van der Waals contributions to the interaction energy of the ligand with the surrounding in the bound (b) and the unbound (ub) states, and α and β two empirical parameters which are typically determined by linear fitting to experimental binding affinities. Since our simulations were performed using PME to treat both the electrostatic and van der Waals (LJ-PME[16]) interactions, which does not allow direct atomic pair-wise decomposition, a post-processing step was introduced to evaluate the ensemble averages in Equation 1. Starting with the bound state, the trajectory of the solvated complex was split in two trajectories, one containing the receptor and the solvent molecules (and also counter-ions if the system was not neutral), the other containing the ligand alone. From these two trajectories along with the one of the solvated complex, ensemble averages of the electrostatic and van der Waals (PME) energies for the complex in solution ($\langle U_{RL+solv}^{elec} \rangle_b$ and $\langle U_{RL+solv}^{vdw} \rangle_b$, respectively), the receptor in solution ($\langle U_{R+solv}^{elec} \rangle_b$ and $\langle U_{R+solv}^{vdw} \rangle_b$, respectively) and the ligand

in vacuum ($\langle U_L^{elec}\rangle_b$ and $\langle U_{R+solv}^{vdw}\rangle_b$, respectively) were extracted. Then, the electrostatics and van der Waals contributions to the ligand interaction energy were evaluated as

$$\langle U_{L/s}^{elec}\rangle_b = \langle U_{RL+solv}^{elec}\rangle_b - \langle U_{R+solv}^{elec}\rangle_b - \langle U_L^{elec}\rangle_b \quad (2)$$

$$\langle U_{L/s}^{vdw}\rangle_b = \langle U_{RL+solv}^{vdw}\rangle_b - \langle U_{R+solv}^{vdw}\rangle_b - \langle U_L^{vdw}\rangle_b \quad (3)$$

Similarly, upon separating the MD trajectory of the ligand in three, i.e. free ligand in solution, free ligand in vacuum, and solvent alone, the electrostatics ($\langle U_{L/s}^{elec}\rangle_{ub}$) and the van der Waals ($\langle U_{L/s}^{vdw}\rangle_{ub}$) contributions to the ligand interaction energy in the unbound state were evaluated as

$$\langle U_{L/s}^{elec}\rangle_{ub} = \langle U_{L+solv}^{elec}\rangle_{ub} - \langle U_{solv}^{elec}\rangle_{ub} - \langle U_L^{elec}\rangle_{ub} \quad (4)$$

$$\langle U_{L/s}^{vdw}\rangle_{ub} = \langle U_{L+solv}^{vdw}\rangle_{ub} - \langle U_{solv}^{vdw}\rangle_{ub} - \langle U_L^{vdw}\rangle_{ub} \quad (5)$$

Finally, statistical errors associated with the numerical determination of ΔG_b° by LIE were estimated as in Baron *et al.*[17] For this purpose, the MD trajectories of the bound and unbound states were split in two chunks, named A and B, and ensemble averages of the electrostatic and van der Waals contributions to the ligand interaction energy were computed per chunk. Then, the statistical error associated with each ensemble average was estimated as

$$\langle E_{L-s}\rangle = \frac{1}{2}\left|\langle U_{L-s}^{A}\rangle - \langle U_{L-s}^{B}\rangle\right| \quad (6)$$

and that on ΔG_b° by a LIE-like equation as

$$Error_b = \beta\left[\left\langle E_{L-s}^{elec}\right\rangle_b + \left\langle E_{L-s}^{elec}\right\rangle_{ub}\right] + \alpha\left[\left\langle E_{L-s}^{vdw}\right\rangle_b + \left\langle E_{L-s}^{vdw}\right\rangle_{ub}\right] \quad (7)$$

1.4 Assessing the impact of the training set on the performance of the LIE model

To test the robustness of our LIE model for cavitand host-guest systems with respect to the selection of the training set, two analyses were carried out. First, the entire data set of 61 host-guest complexes (training plus test) was randomly split in two new sets of 14 and 47 host-guest complexes, which were used as new training and test sets, respectively. New LIE parameters were obtained and the RMSE for the test set evaluated. By repeating the procedure 1×10^5 times, the frequency distributions for the LIE parameters α and β and the RMSE for the test set shown in Figure S6 were obtained. The results indicate that the most populated values of α and β are essentially equivalent to those used in the *Main Text*, which were based on an arbitrary selection of the training set. Moreover, the distribution of the RMSE shows that free energy results within 1.5 kcal/mol from experiments are obtained almost independently of the training set. Based on this analysis, we conclude that the empirical parameters of our LIE model for cavitand host-guests as well as the accuracy of the binding affinity predictions are essentially independent of the training set.

In a second analysis, the size of the initial training set, which was composed of n=14 CB7-guest complexes (see *Main Text*), was systematically reduced by random elimination

of $k = 1, 2, 3 \ldots 11$ complexes. For each value of k, $\frac{n!}{k!(n-k)!}$ unique training sets of $n - k$ members were generated. New LIE parameters were obtained and the RMSE for the test set evaluated; note that the k complexes removed from the training set were considered as part of the test set to preserve the size of the full data set (61 host-guest complexes). Average values and associated errors for α, β and RMSE for the test set as a function of the size of the training set are presented in Figure S7. The results clearly show that reliable LIE models yielding RMSE <1.5 kcal/mol can be obtained using a training set including as little as 7 experimental determinations of the host-guest binding affinity.

1.5 Convergence analysis

Convergence analysis of the binding affinity predictions by LIE was performed to explore the efficiency of the methodology and its suitability for virtual screening campaigns. For this purpose, the simulation time (t_{min}) required to obtain predictions with a deviation of <0.5 kcal/mol from the ΔG_b° at full sampling (20 ns) was used as a convergence metric. The frequency distribution of t_{min} for all complexes of the test set was evaluated and fitted with an exponential function of the form $\exp(-t/\tau)$, whose characteristic time τ was of 1.1 ns. Based on this analysis, we conclude that the simulation time required to obtain converged binding affinity results in most complexes of the test set was <2 ns independently of the force field, which demonstrates the remarkable efficiency of our LIE model; see Figure S9 (top). Based on these results we conclude that accurate binding-affinity predictions in host-guests can be obtained by LIE with a few nanosecond MD. Finally, Figure S9 (bottom) shows the rapid convergence of the LIE parameters (α and β) as a function of MD sampling.

1.6 Computing the Strain Energy of the Host (ΔE_{str})

The LIE model in Equation 2 of the *Main Text* evaluates the binding affinity of the guest including the strain energy of the host on complexation, ΔE_{str}. Here, the computation of the strain energy was done using `sander`[18] from the AmberTools17 suite package[19] as follows

1. Initially, a cluster analysis (using the GROMACS tool "`gmx cluster`") on the production trajectory of a host-guest complex (e.g. a CB[7-8]-steroid complex) was performed in order to isolate the structure of the most populated cluster. The clustering procedure was based on the "single linkage" method where one conformation of the complex is added to a particular cluster if its all-atom RMSD (after mass-weight fitting) was less than 0.25 nm.

2. The central conformation of the most populated cluster was submitted to energy minimization by performing 50000 steps of conjugate gradient, prior to four cycles of local minimization using steepest descent (`maxcyc=50000` and `ntmin=0` in `sander`) until the root-mean-square of the energy gradient was less than 1×10^{-5} kJ mol^{-1} Å^{-1} (`drms=`1×10^{-5} in `sander`). All geometry optimizations were done *in vacuum* for neutral system, whereas the GBSA implicit model by Hawkins *et al*[20] (`igb=1` in `sander`) was used for formally charged systems.

3. The atoms of the host were extracted from the optimized structure of the complex and its intramolecular energy evaluated, which yields the configurational energy of the host at the minimum in the bound state, E_b^{host}.

4. Then, the conformation of the host previously optimized in the presence of guest (i.e. the bound state) was energy minimized with the guest removed using the same procedure. Evaluation of the energy of the host yields the configurational energy of the host at the minimum in the unbound state, E_{ub}^{host}.

5. Finally, the strain energy of the host was determined as $\Delta E_{str} = E_{ub}^{host} - E_b^{host}$.

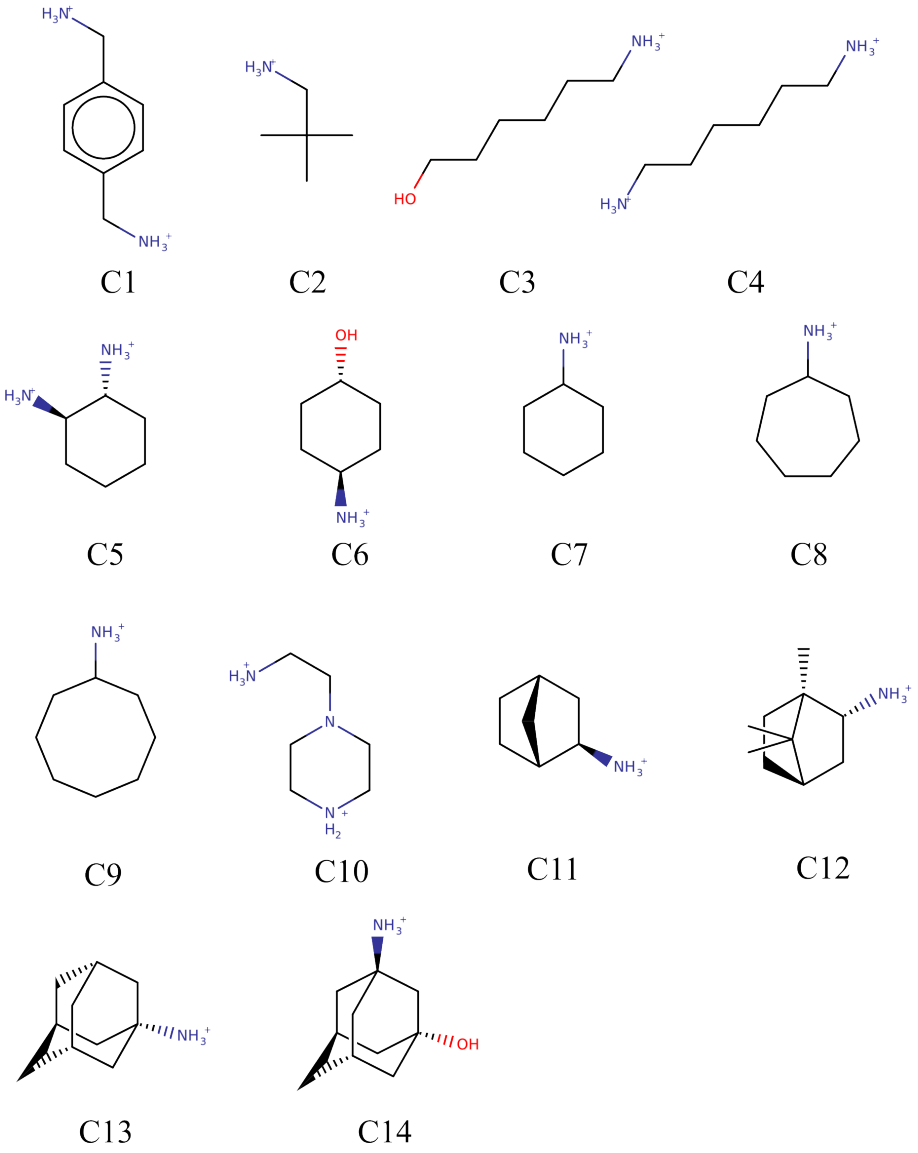

Figure S1: Chemical structures of CB7 guests used as a **training** set, shown in their protonated states used in the computations to produce the LIE models.

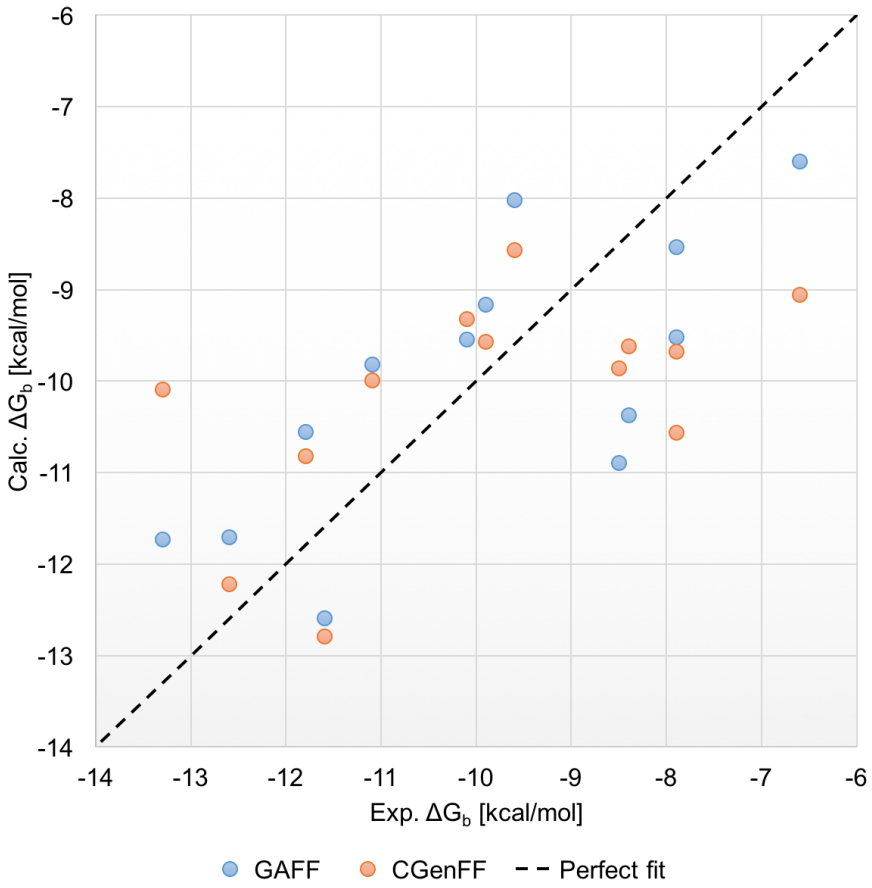

Figure S2: Experimental vs calculated binding free energy values in aqueous solution for the **training** set used to build the GAFF (R=0.79 and RMSE=1.35 kcal/mol) and CGenFF (R=0.63 and RMSE=1.70 kcal/mol) LIE models.

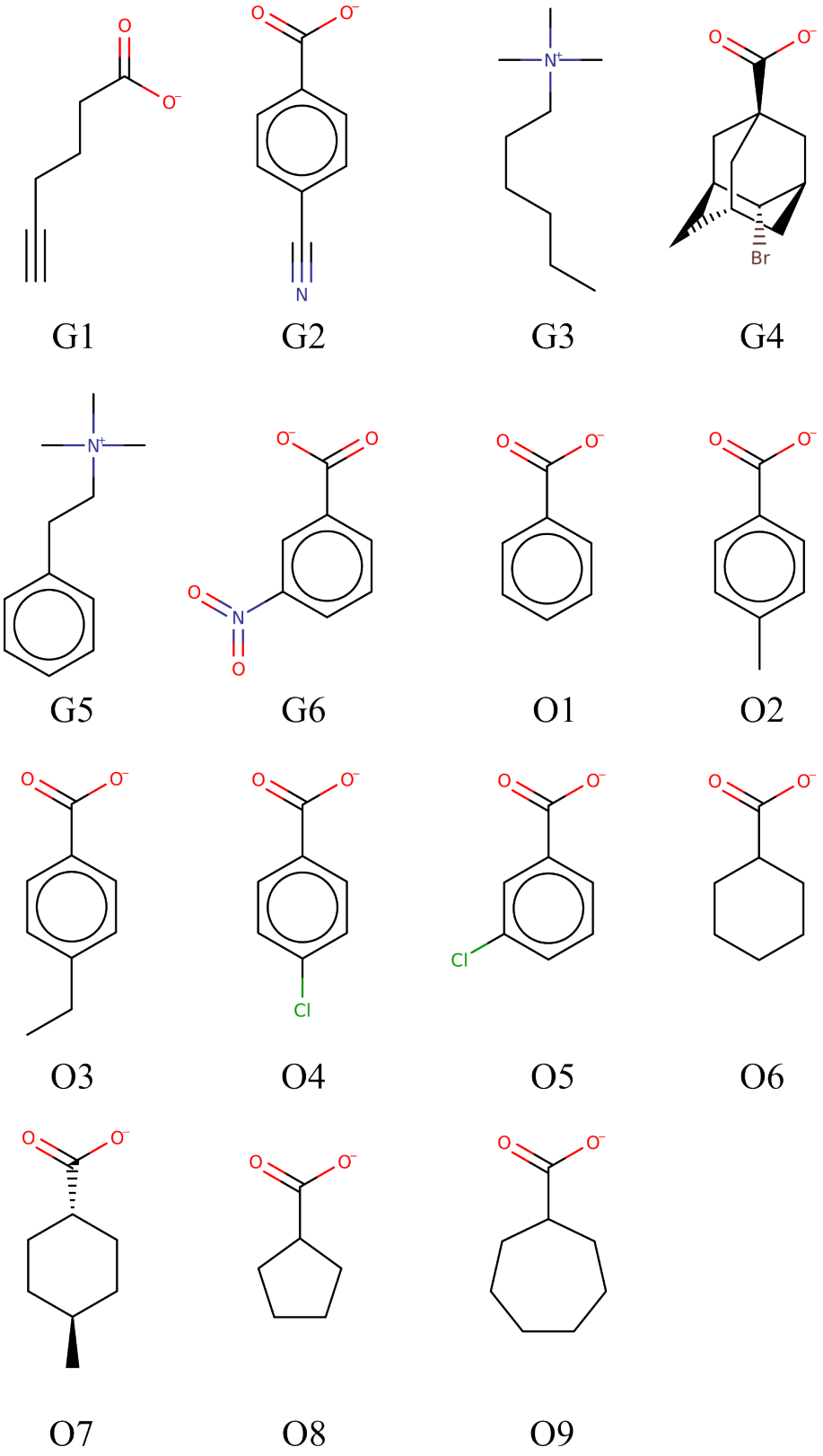

Figure S3: Chemical structures of the 15 ligands for octa-acid, OAH, (Gxx and Oxx ligand families reported in references Yin *et al*[2] and Muddana *et al*,[1] respectively) and the 6 ligands for the tetramethylated octa-acid, OAM, (Gxx[2] ligand family), which were part of the **test** set. The guests are shown in the protonation form used in the calculations.

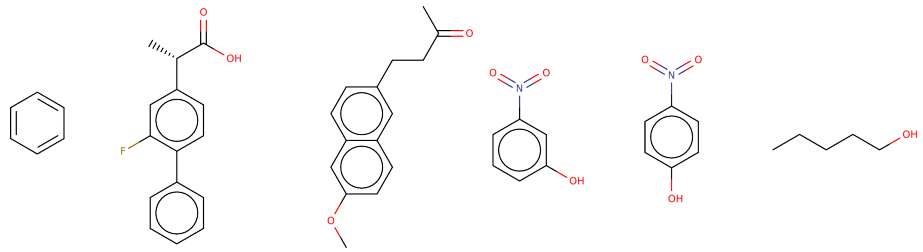

Figure S4: Chemical structures of the six β-cyclodextrin guests that were part of the **test** set. The guests are shown in the protonation form used in the calculations.

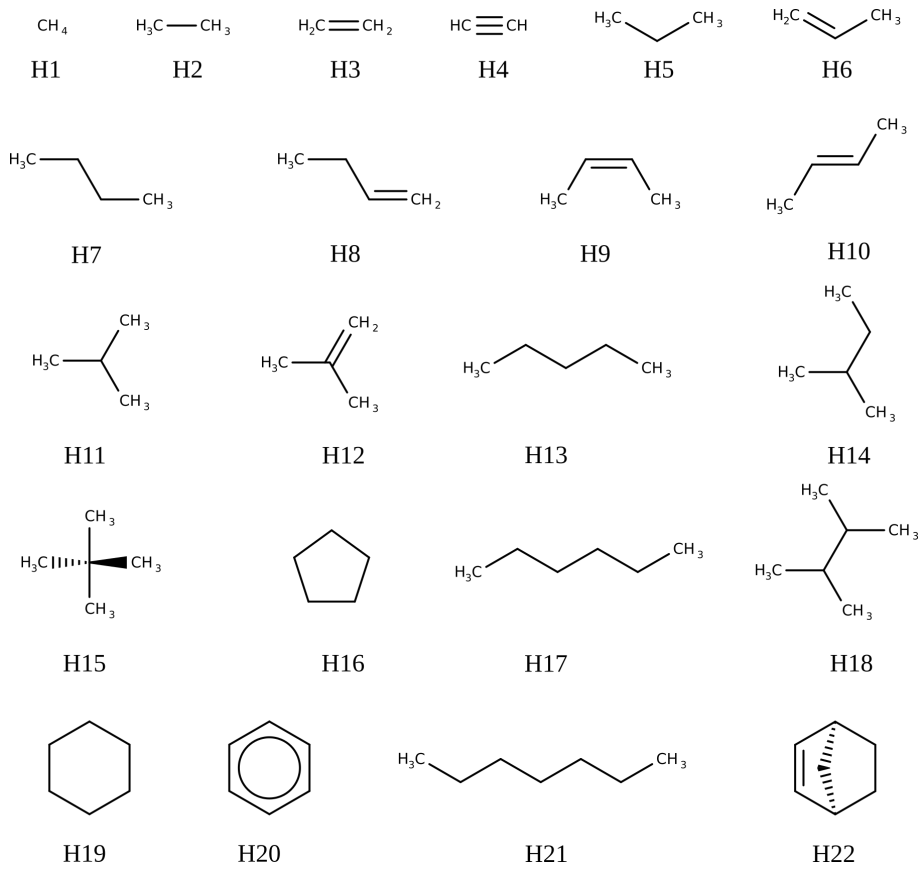

Figure S5: Chemical structures the 22 cucurbit-7-uril (CB7) guests that were part of the **test** set. All these guests are neutral and belong to the HYDROPHOBE challenge.[3]

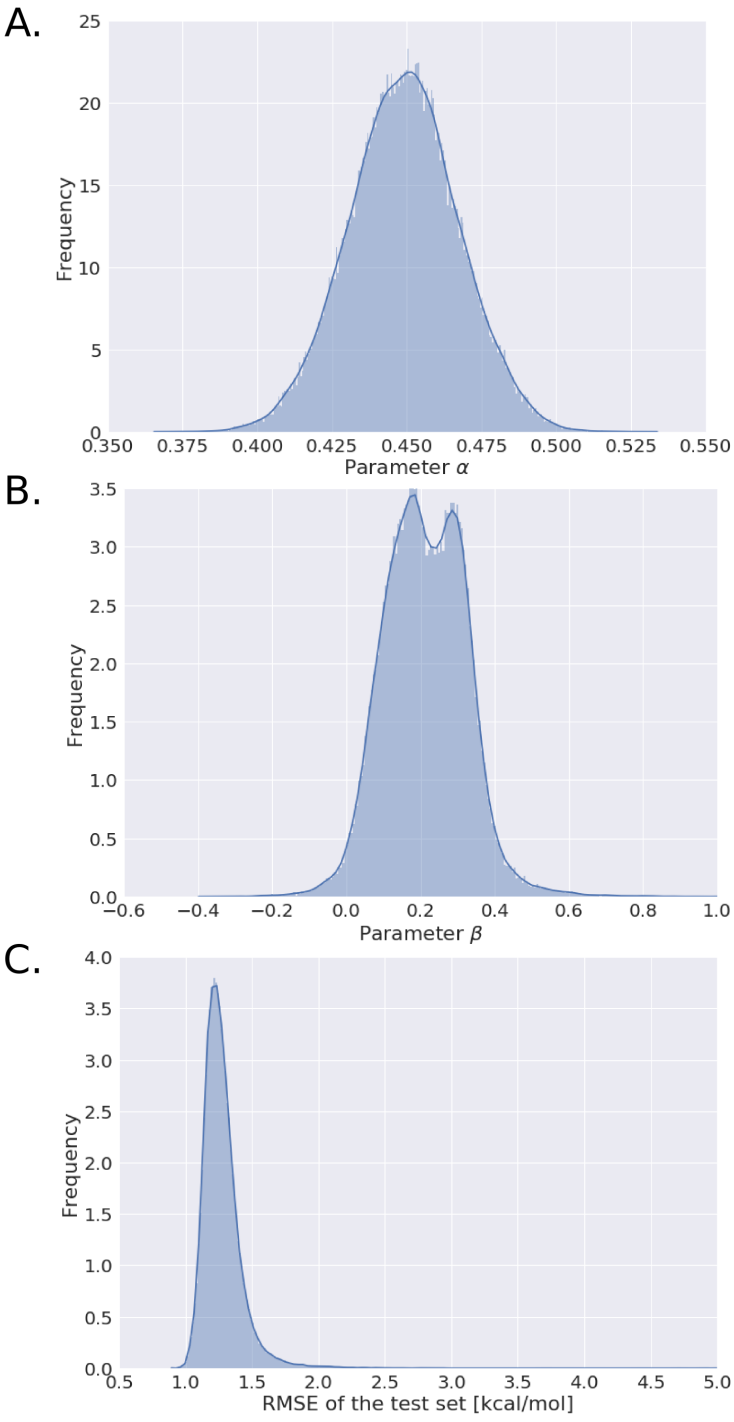

Figure S6: Frequency distributions of the GAFF/LIE parameters α (A), β (B), and the RMSE for the test set (C) upon splitting of the full data set into 1×10^5 random training/test sets.

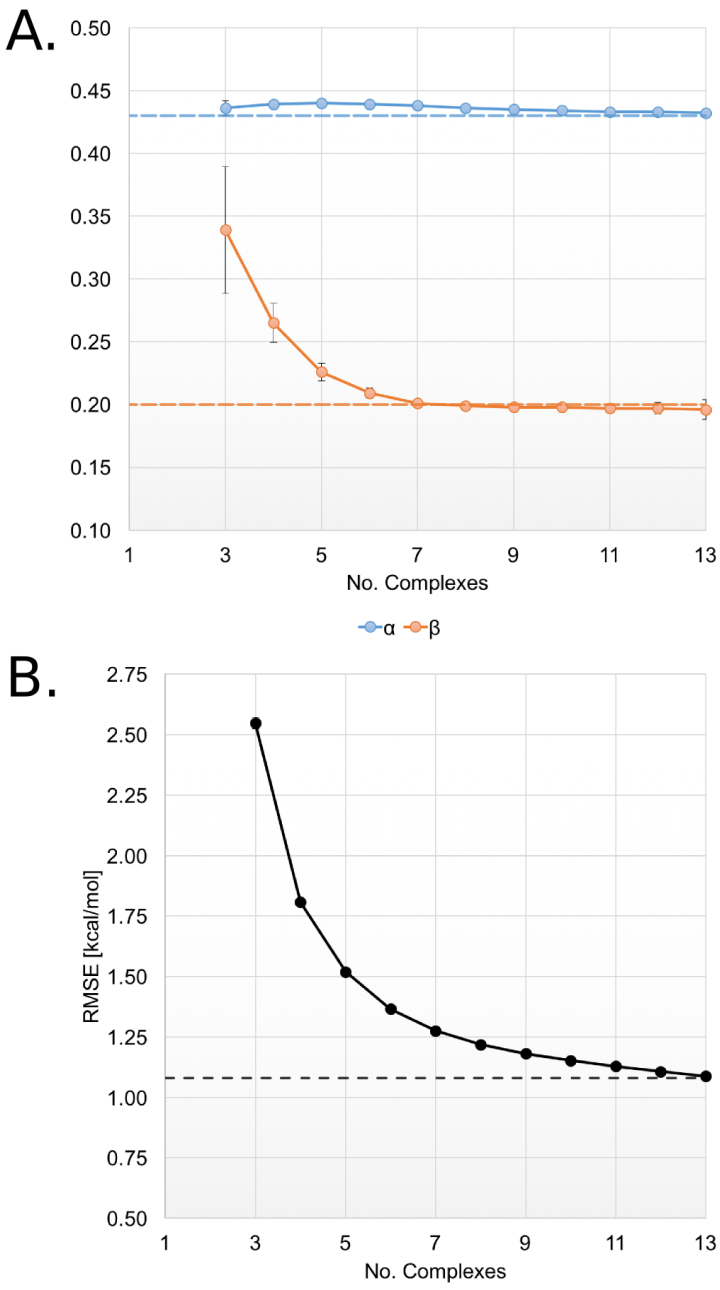

Figure S7: Average values of the LIE parameters (A) and the RMSE for the test set (B) after removing $k = 1, 2, 3 \ldots 11$ members from the training set. The dashed lines represent the values for α, β and RMSE for the test set obtained using the original (n=14) training set.

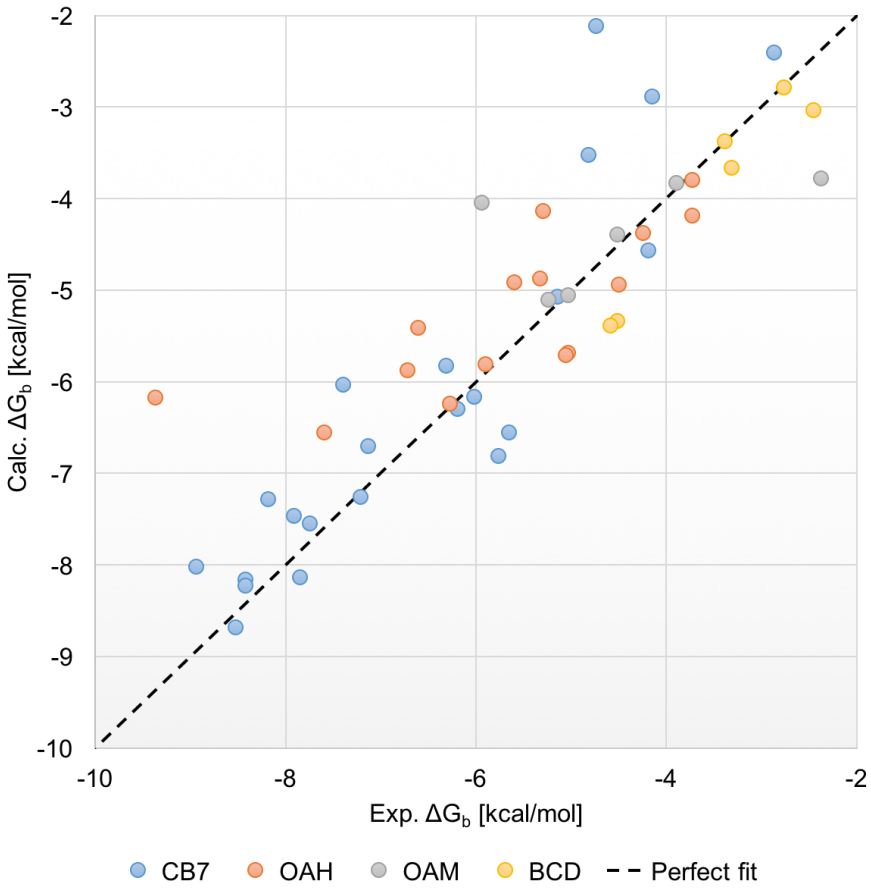

Figure S8: Experimental vs calculated binding free energy values in aqueous solution for the **test** set using the CGenFF LIE model (overall R=0.87 and RMSE=0.92 kcal/mol).

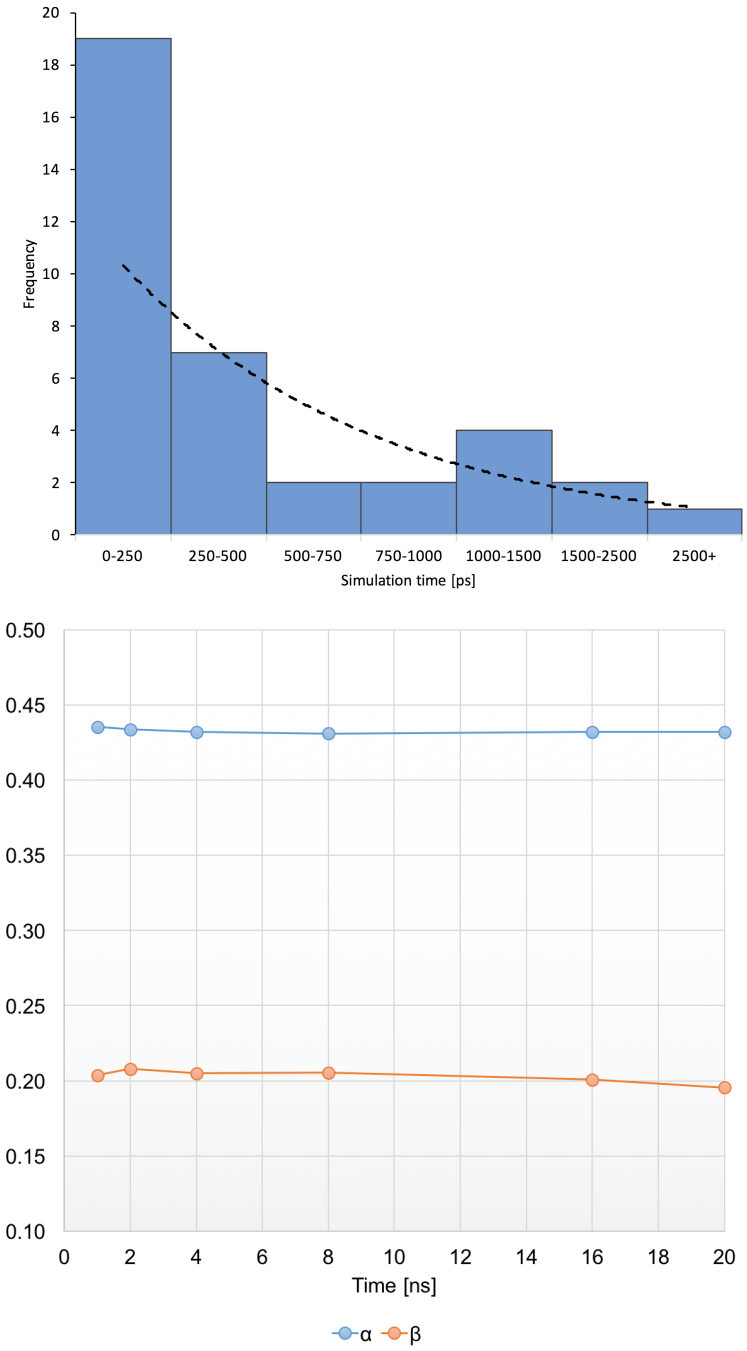

Figure S9: On top, the frequency distribution for t_{min} values computed for the **test** set using GAFF. On bottom, convergence analysis for parameters of the GAFF/LIE model.

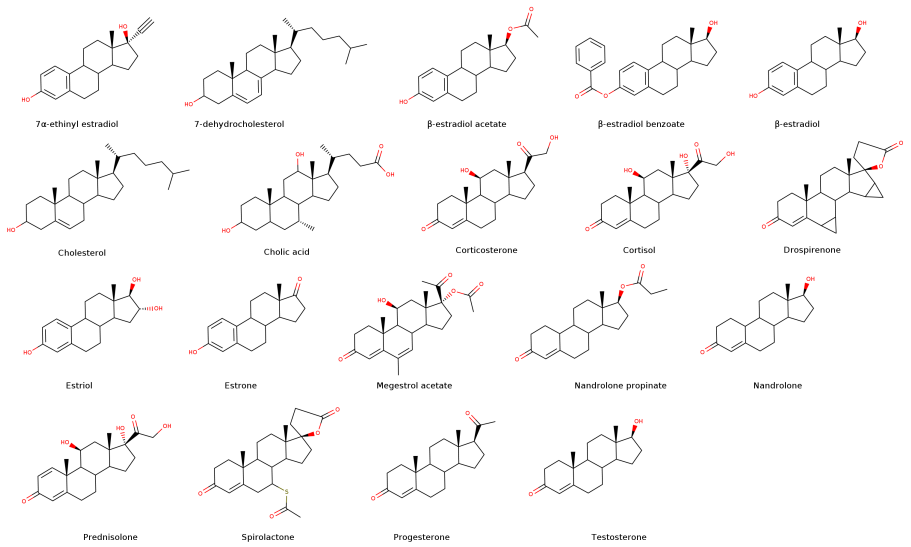

Figure S10: Chemical structures of **steroid compounds** that bind to CB[7/8] used in this study.

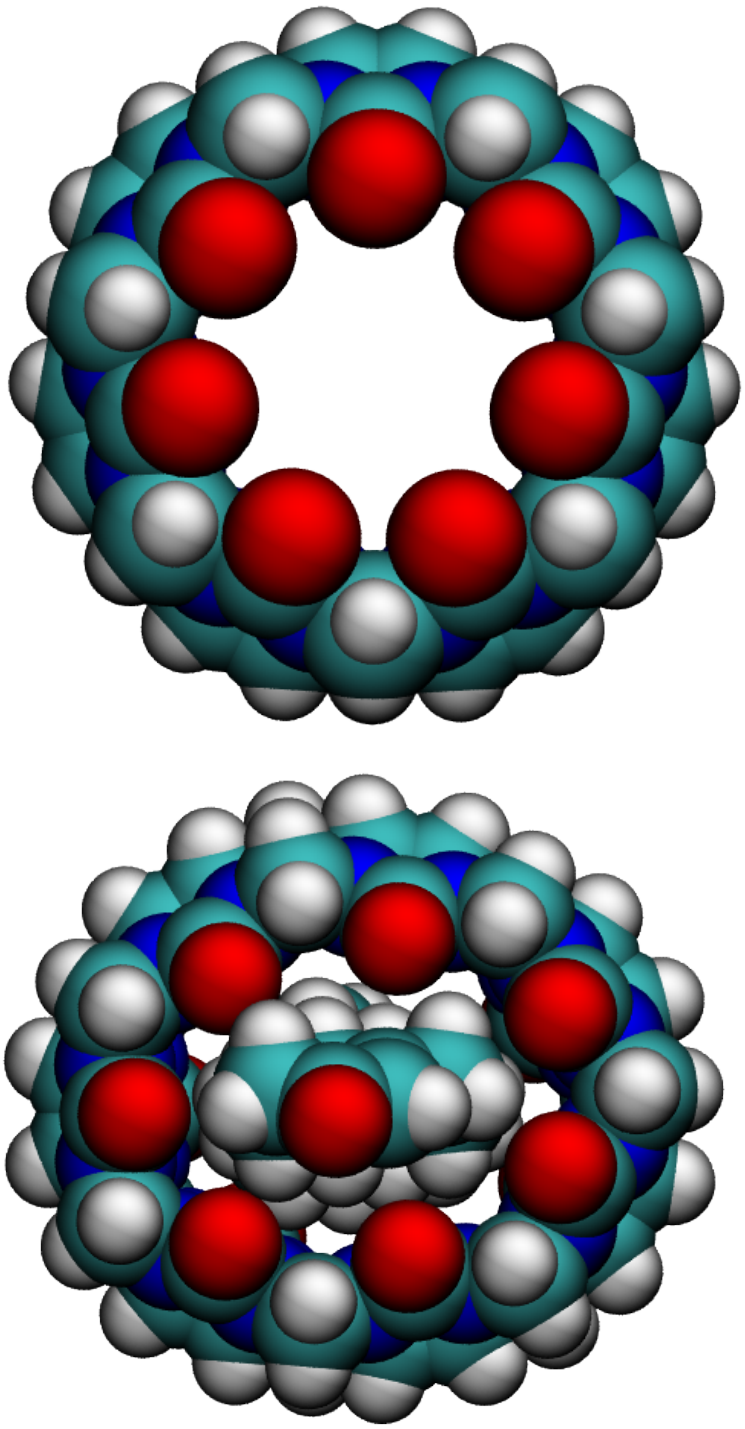

Figure S11: Geometry of the most populated cluster of CB7 in the unbound state (top) and the bound state to nandrolone (bottom) from MD simulations in explicit solvent. The structural deformation of the host in the bound state is striking.

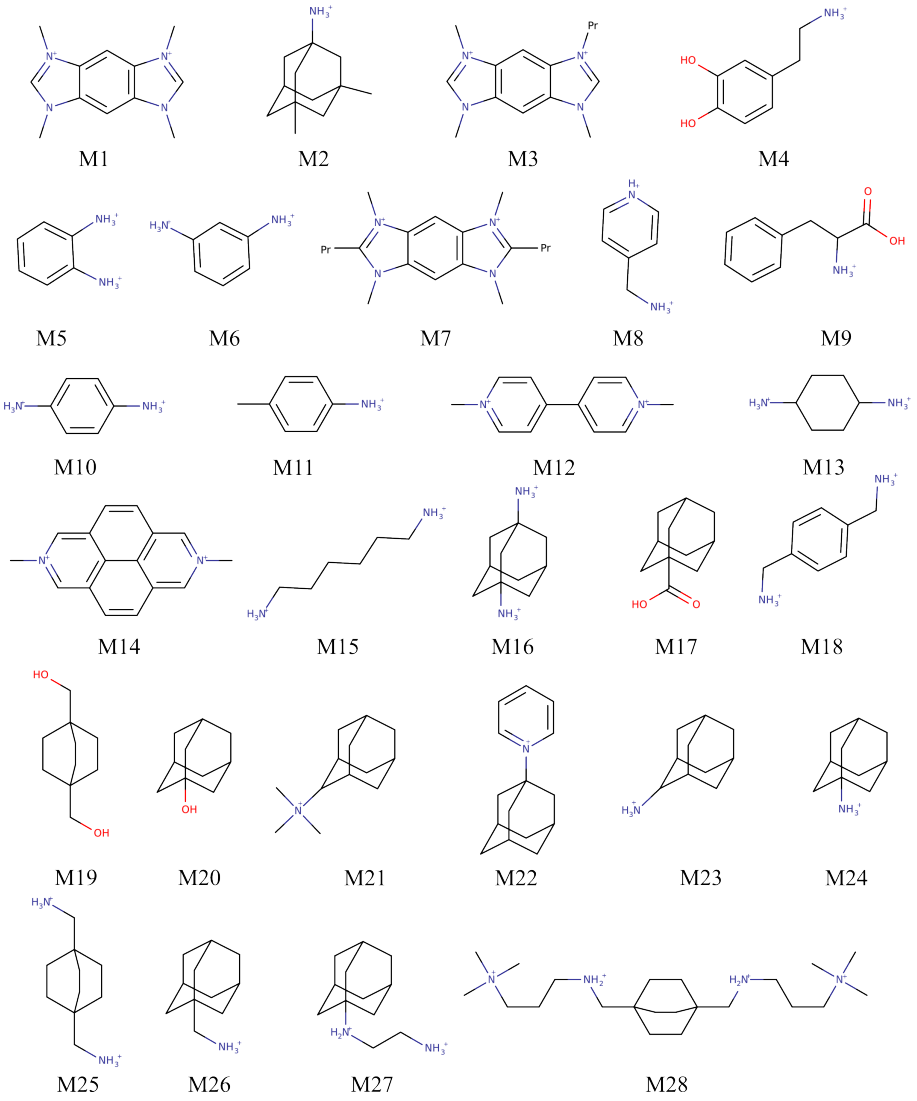

Figure S12: Chemical structures of the CB7 guests from the **Muddana** set, shown in their protonated states used in the computations.

Table S1: Experimental and predicted binding free energy values (in kcal/mol), using the LIE model based on GAFF, for all host-guest systems in the training set. The van der Waals and the electrostatic contributions to the ligand interaction energies in the bound and the unbound states were used to obtain the LIE parameters by linear fitting of experimental binding free energies using Eq. 1 in main text.

Host	Guest	$Exp.\Delta G_b^\circ$	$\langle U_{L-s}^{vdw}\rangle_b - \langle U_{L-s}^{vdw}\rangle_{ub}$	$\langle U_{L-s}^{elec}\rangle_b - \langle U_{L-s}^{elec}\rangle_{ub}$	$Calc.\ \Delta G_b^{LIE}$	ΔE_{str}	$Calc.\ \Delta G_b^{LIE} + \Delta E_{str}$
CB7	C1	-9.90	-18.42 ± 0.1	-6.21 ± 0.2	-9.16 ± 0.1	2.7	-6.5 ± 0.1
	C2	-9.60	-17.65 ± 0.1	-2.13 ± 0.3	-8.02 ± 0.1	1.4	-6.59 ± 0.1
	C3	-6.60	-22.72 ± 0.5	10.89 ± 2.2	-7.59 ± 0.7	0.9	-6.72 ± 0.7
	C4	-8.40	-17.84 ± 0.3	-13.51 ± 0.4	-10.37 ± 0.2	1.7	-8.65 ± 0.2
	C5	-8.50	-21.86 ± 0.1	-7.47 ± 0.4	-10.89 ± 0.1	3.6	-7.33 ± 0.1
	C6	-7.90	-21.61 ± 0.1	3.82 ± 0.1	-8.53 ± 0.1	0.7	-7.79 ± 0.1
	C7	-10.10	-21.07 ± 0	-2.41 ± 0.2	-9.54 ± 0	0.8	-8.75 ± 0
	C8	-11.80	-23.22 ± 0	-2.85 ± 0.2	-10.55 ± 0.1	0.8	-9.8 ± 0.1
	C9	-12.60	-25.65 ± 0.1	-3.35 ± 0.3	-11.7 ± 0.1	2.1	-9.61 ± 0.1
	C10	-7.90	-21.63 ± 0.2	-1.09 ± 1.3	-9.52 ± 0.3	2.4	-7.09 ± 0.3
	C11	-11.10	-21.62 ± 0	-2.57 ± 0.2	-9.81 ± 0	0.8	-8.97 ± 0
	C12	-13.30	-25.95 ± 0.5	-2.84 ± 0.7	-11.73 ± 0.3	2.7	-9.02 ± 0.3
	C13	-14.10	-28.97 ± 0.1	-3.18 ± 0.2	-13.09 ± 0.1	0.9	-12.24 ± 0.1
	C14	-11.60	-31.72 ± 0.1	5.27 ± 0.1	-12.59 ± 0.1	1.8	-10.8 ± 0.1

Table S2: Experimental and predicted binding free energy values (in kcal/mol), using the LIE model based on CGenFF, for all host-guest systems in the training set. Also, the van der Waals and electrostatic contributions to the ligand interaction energies used to get the LIE parameters from linear fitting with experimental binding free energy values, are presented.

Host	Guest	$Exp.\Delta G_b^{\circ}$	$\langle U_{L-S}^{vdw}\rangle_b - \langle U_{L-S}^{vdw}\rangle_{ub}$	$\langle U_{L-S}^{elec}\rangle_b - \langle U_{L-S}^{elec}\rangle_{ub}$	$Calc.\ \Delta G_b^{LIE}$
CB7	C1	-9.90	-14.86 ± 0.1	3 ± 0.3	-9.57 ± 0.1
	C2	-9.60	-12.76 ± 0.1	-1.8 ± 0.1	-8.56 ± 0.1
	C3	-6.60	-13.21 ± 0.1	-4.18 ± 0.1	-9.05 ± 0.1
	C4	-8.40	-13.2 ± 0.1	-11.36 ± 0.3	-9.62 ± 0.1
	C5	-8.50	-14.62 ± 0	-2.59 ± 0.4	-9.86 ± 0
	C6	-7.90	-15.13 ± 0.2	3.94 ± 0.6	-9.67 ± 0.2
	C7	-10.10	-14.02 ± 0.1	-0.75 ± 0.3	-9.32 ± 0.1
	C8	-11.80	-16.07 ± 0.1	-2.56 ± 0.4	-10.81 ± 0.1
	C9	-12.60	-18.03 ± 0.1	-3.99 ± 0.3	-12.22 ± 0.1
	C10	-7.90	-15.52 ± 0.1	-3.91 ± 0.5	-10.56 ± 0.1
	C11	-11.10	-14.82 ± 0.1	-2.57 ± 0.4	-9.99 ± 0.1
	C12	-13.30	-14.9 ± 0.1	-3.09 ± 0.5	-10.08 ± 0.1
	C13	-14.10	-17.84 ± 0	-1.05 ± 0.3	-11.86 ± 0.1
	C14	-11.60	-20.29 ± 0.2	7.53 ± 0.1	-12.79 ± 0.1

Table S3: Statistical metrics used to assess the accuracy of the computed ΔG_b^{LIE} values for the training set using the LIE models based on GAFF and CGenFF LIE models.

Metric	GAFF	CGenFF
RMSE	1.35	1.70
MAE	1.25	1.48
R	0.79	0.63
R^2	0.62	0.39
Slope	0.68	0.34
Intercept	-4.30	-6.75

Table S4: Experimental and predicted binding free energy values (in kcal/mol), using the GAFF LIE model with and without ΔE_{str} values, for all host-guest systems in the test set. Also, the van der Waals and electrostatic contributions to the ligand interaction energies used to compute binding free energy values are presented.

Host	Guest	$Exp.\Delta G_b^\circ$	$\langle U_{L-S}^{vdw}\rangle_b - \langle U_{L-S}^{vdw}\rangle_{ub}$	$\langle U_{L-S}^{elec}\rangle_b - \langle U_{L-S}^{elec}\rangle_{ub}$	$Calc.\ \Delta G_b^{LIE}$	ΔE_{str}	$Calc.\ \Delta G_b^{LIE} + \Delta E_{str}$
OAH	G1	-5.04	-11.79 ± 0.1	1.19 ± 0.2	-4.83 ± 0.1	0.1	-4.74 ± 0.1
	G2	-4.25	-13.88 ± 0	6.84 ± 0.3	-4.6 ± 0.1	0.1	-4.52 ± 0.1
	G3	-5.06	-11.17 ± 0.1	-2.83 ± 0.2	-5.37 ± 0.1	0.1	-5.27 ± 0.1
	G4	-9.37	-13.97 ± 0.2	-1.75 ± 0.1	-6.36 ± 0.1	0.3	-6.09 ± 0.1
	G5	-4.50	-11.54 ± 0.2	2.25 ± 0.2	-4.51 ± 0.1	0.4	-4.1 ± 0.1
	G6	-5.33	-12.03 ± 0	0.76 ± 0.3	-5.02 ± 0.1	0.2	-4.8 ± 0.1
	O1	-3.73	-9.34 ± 0	0.11 ± 0.1	-3.99 ± 0	0.1	-3.94 ± 0
	O2	-5.90	-12.17 ± 0.1	-0.51 ± 0.6	-5.34 ± 0.2	0.1	-5.28 ± 0.2
	O3	-6.28	-12.9 ± 0.1	0.18 ± 0.2	-5.51 ± 0.1	0.2	-5.29 ± 0.1
	O4	-6.72	-12.15 ± 0.1	-0.43 ± 0.6	-5.31 ± 0.1	0.1	-5.23 ± 0.1
	O5	-5.30	-9.27 ± 0	-2.7 ± 0.4	-4.53 ± 0.1	0.1	-4.42 ± 0.1
	O6	-5.60	-9.26 ± 0.1	-5.83 ± 0.5	-5.15 ± 0.1	0.1	-5.06 ± 0.1
	O7	-7.60	-12.61 ± 0.1	-7.97 ± 0.6	-7.02 ± 0.1	0.1	-6.93 ± 0.1
	O8	-3.73	-8.82 ± 0.1	-4.55 ± 0.1	-4.7 ± 0.1	0.1	-4.62 ± 0.1
	O9	-6.61	-10.59 ± 0.3	-6.4 ± 0.4	-5.83 ± 0.2	0.2	-5.59 ± 0.2
OAM	G1	-5.24	-11.11 ± 0.1	3.52 ± 1	-4.07 ± 0.3	0.1	-3.98 ± 0.3
	G2	-5.04	-13.15 ± 0.2	8.16 ± 0.3	-4.02 ± 0.1	0.2	-3.85 ± 0.1
	G3	-5.94	-10.61 ± 0	1.52 ± 0.1	-4.26 ± 0	0.2	-4.06 ± 0
	G4	-2.38	-11.54 ± 0	-2.69 ± 0.9	-5.5 ± 0.2	5.7	0.21 ± 0.2
	G5	-3.90	-10.46 ± 0	3.25 ± 0.1	-3.85 ± 0	0.6	-3.29 ± 0
	G6	-4.52	-11.29 ± 0.1	4.68 ± 0.4	-3.92 ± 0.1	0.5	-3.41 ± 0.1
BCD	B1	-2.77	-7.52 ± 0.1	0.88 ± 0.2	-3.06 ± 0.1	0.2	-2.9 ± 0.1
	B2	-4.52	-15.58 ± 0.1	1.03 ± 0.6	-6.49 ± 0.2	0.9	-5.62 ± 0.2
	B3	-4.59	-15.71 ± 0.1	2.23 ± 0.2	-6.31 ± 0.1	1.8	-4.48 ± 0.1
	B4	-3.32	-11.9 ± 0.1	0.96 ± 0.1	-4.93 ± 0.1	0.2	-4.7 ± 0.1

	B5	-3.39	-11.4 ± 0.1	0.75 ± 0.2	-4.75 ± 0.1	0.8	-3.98 ± 0.1
	B6	-2.46	-8.98 ± 0.2	0.5 ± 0.1	-3.76 ± 0.1	0.1	-3.63 ± 0.1
CB7	H01	-4.74	-4.01 ± 0	0 ± 0	-1.72 ± 0	0.08	-1.64 ± 0
	H02	-4.82	-6.87 ± 0	0.05 ± 0.1	-2.94 ± 0	0.05	-2.89 ± 0
	H03	-4.15	-5.72 ± 0	0.24 ± 0	-2.41 ± 0	0.04	-2.37 ± 0
	H04	-2.87	-7.26 ± 0	0.31 ± 0	-3.06 ± 0	0.03	-3.03 ± 0
	H05	-5.15	-9.71 ± 0	-0.02 ± 0	-4.18 ± 0	0.05	-4.13 ± 0
	H06	-4.19	-8.35 ± 0	-0.19 ± 0	-3.63 ± 0	0.02	-3.61 ± 0
	H07	-7.14	-13.64 ± 0.1	0.02 ± 0	-5.86 ± 0	0.04	-5.82 ± 0
	H08	-6.02	-12.06 ± 0	-0.37 ± 0.1	-5.26 ± 0	0.03	-5.23 ± 0
	H09	-6.2	-12.3 ± 0	-0.29 ± 0	-5.35 ± 0	0.16	-5.19 ± 0
	H10	-5.66	-12.45 ± 0	-0.31 ± 0	-5.41 ± 0	0.80	-4.61 ± 0
	H11	-7.4	-13.04 ± 0	-0.05 ± 0.1	-5.62 ± 0	0.10	-5.52 ± 0
	H12	-6.32	-12.15 ± 0	-0.34 ± 0.1	-5.29 ± 0	0.10	-5.19 ± 0
	H13	-7.75	-16.31 ± 0.1	-0.04 ± 0	-7.02 ± 0	0.40	-6.62 ± 0
	H14	-7.92	-16.15 ± 0	-0.09 ± 0	-6.96 ± 0	0.22	-6.74 ± 0
	H15	-8.19	-15.77 ± 0	0 ± 0	-6.78 ± 0	0.16	-6.62 ± 0
	H16	-7.22	-15.73 ± 0	-0.08 ± 0	-6.78 ± 0	0.19	-6.59 ± 0
	H17	-8.43	-18.12 ± 0.2	-0.18 ± 0.1	-7.83 ± 0.1	1.37	-6.46 ± 0.1
	H18	-8.53	-18.3 ± 0	-0.16 ± 0.1	-7.9 ± 0	0.81	-7.09 ± 0
	H19	-8.43	-18.95 ± 0	-0.1 ± 0.1	-8.17 ± 0	0.13	-8.04 ± 0
	H20	-5.77	-14.48 ± 0	-0.5 ± 0	-6.33 ± 0	0.07	-6.26 ± 0
	H21	-8.94	-18.39 ± 0.2	-0.23 ± 0	-7.95 ± 0.1	0.03	-7.92 ± 0.1
	H22	-7.85	-18.27 ± 0	0.15 ± 0	-7.83 ± 0	0.12	-7.71 ± 0

Table S5: Experimental and predicted binding free energy values (in kcal/mol), using the CGenFF LIE model, for all host-guest systems in the test set. Also, the van der Waals and electrostatic contributions to the ligand interaction energies used to compute binding free energy values are presented.

Host	Guest	$Exp.\Delta G_b^\circ$	$\langle U_{L-S}^{vdw}\rangle_b - \langle U_{L-S}^{vdw}\rangle_{ub}$	$\langle U_{L-S}^{elec}\rangle_b - \langle U_{L-S}^{elec}\rangle_{ub}$	$Calc.\ \Delta G_b^{LIE}$
OAH	G1	-5.04	-9.11 ± 0.2	4.19 ± 0.5	-5.68 ± 0.1
	G2	-4.25	-6.37 ± 0.1	-2.04 ± 0.3	-4.37 ± 0.1
	G3	-5.06	-8.02 ± 0.1	-5.12 ± 0.3	-5.7 ± 0.1
	G4	-9.37	-9.39 ± 0.1	0.35 ± 0.4	-6.17 ± 0.1
	G5	-4.50	-7.21 ± 0.1	-2.23 ± 0.4	-4.94 ± 0.1
	G6	-5.33	-7.22 ± 0.1	-1.35 ± 0.2	-4.87 ± 0.1
	O1	-3.73	-6.37 ± 0.1	0.3 ± 0.2	-4.18 ± 0.1
	O2	-5.90	-8.79 ± 0.1	-0.05 ± 0.3	-5.8 ± 0.1
	O3	-6.28	-9.57 ± 0.1	1.03 ± 0.6	-6.23 ± 0.1
	O4	-6.72	-8.67 ± 0.1	-1.8 ± 0.4	-5.87 ± 0.1
	O5	-5.30	-5.88 ± 0.1	-3.12 ± 0.3	-4.13 ± 0.1
	O6	-5.60	-7.14 ± 0.1	-2.52 ± 0.3	-4.91 ± 0.1
	O7	-7.60	-9.35 ± 0.1	-4.69 ± 0.1	-6.55 ± 0.1
	O8	-3.73	-5.05 ± 0.1	-5.72 ± 0.4	-3.79 ± 0.1
	O9	-6.61	-7.91 ± 0.1	-2.37 ± 0.2	-5.41 ± 0
OAM	G1	-5.24	-8.97 ± 0.1	10.26 ± 0.2	-5.1 ± 0.1
	G2	-5.04	-7.89 ± 0.1	1.95 ± 0.3	-5.05 ± 0.1
	G3	-5.94	-5.7 ± 0	-3.48 ± 0	-4.04 ± 0
	G4	-2.38	-5.91 ± 0.8	1.59 ± 1.9	-3.77 ± 0.7
	G5	-3.90	-5.63 ± 0	-1.38 ± 0	-3.83 ± 0
	G6	-4.52	-6.96 ± 0.1	2.6 ± 0	-4.39 ± 0.1
BCD	B1	-2.77	-4.45 ± 0.1	1.96 ± 0	-2.78 ± 0
	B2	-4.52	-8.37 ± 0.2	2.42 ± 0.3	-5.33 ± 0.1
	B3	-4.59	-8.75 ± 0.1	4.91 ± 0.3	-5.38 ± 0.1
	B4	-3.32	-5.98 ± 0.2	3.58 ± 0.2	-3.66 ± 0.2

	B5	-3.39	-5.51 ± 0.4	3.35 ± 0.4	-3.37 ± 0.3
	B6	-2.46	-4.86 ± 0.2	2.21 ± 0.3	-3.03 ± 0.1
CB7	H01	-4.74	-3.2 ± 0	0.07 ± 0	-2.11 ± 0
	H02	-4.82	-5.32 ± 0	0 ± 0	-3.51 ± 0
	H03	-4.15	-4.49 ± 0.1	1.06 ± 0.2	-2.88 ± 0.1
	H04	-2.87	-4.02 ± 0.1	3.18 ± 0.2	-2.4 ± 0.1
	H05	-5.15	-7.67 ± 0.1	-0.06 ± 0.1	-5.06 ± 0.1
	H06	-4.19	-6.99 ± 0.1	0.66 ± 0.1	-4.56 ± 0.1
	H07	-7.14	-10.13 ± 0.1	-0.15 ± 0.1	-6.7 ± 0
	H08	-6.02	-9.43 ± 0.1	0.8 ± 0.1	-6.16 ± 0
	H09	-6.2	-9.54 ± 0	0.08 ± 0	-6.29 ± 0
	H10	-5.66	-9.94 ± 0.1	0.09 ± 0	-6.55 ± 0.1
	H11	-7.4	-9.13 ± 0.1	-0.04 ± 0.1	-6.03 ± 0.1
	H12	-6.32	-8.87 ± 0	0.38 ± 0.1	-5.82 ± 0
	H13	-7.75	-11.42 ± 0	-0.07 ± 0.1	-7.54 ± 0
	H14	-7.92	-11.34 ± 0.1	0.24 ± 0	-7.46 ± 0.1
	H15	-8.19	-11.02 ± 0	-0.09 ± 0	-7.28 ± 0
	H16	-7.22	-10.97 ± 0	-0.22 ± 0.1	-7.26 ± 0
	H17	-8.43	-12.39 ± 0.1	0.27 ± 0	-8.16 ± 0.1
	H18	-8.53	-13.17 ± 0.1	0.18 ± 0.1	-8.68 ± 0.1
	H19	-8.43	-12.44 ± 0	-0.13 ± 0	-8.22 ± 0
	H20	-5.77	-10.38 ± 0	0.57 ± 0	-6.81 ± 0
	H21	-8.94	-12.18 ± 0.2	0.33 ± 0	-8.01 ± 0.1
	H22	-7.85	-12.43 ± 0	0.91 ± 0	-8.13 ± 0

Table S6: Statistical metrics used to assess the accuracy of the computed ΔG_b^{LIE} values for the test set.

Force Field	Metric	[a]Overall	[a]OAH	[a]OAM	BCD	CB7
[a]GAFF	RMSE	1.08	0.66	1.06	1.48	1.17
	MAE	0.88	0.55	0.90	1.37	0.95
	R	0.81	0.88	0.98	0.95	0.92
	R^2	0.66	0.77	0.97	0.91	0.84
	Slope	0.72	0.56	0.20	1.47	1.05
	Intercept	-1.09	-2.10	-3.03	0.26	1.20
CGenFF	RMSE	0.92	1.06	0.97	0.54	0.88
	MAE	0.64	0.74	0.61	0.42	0.64
	R	0.87	0.77	0.55	0.97	0.91
	R^2	0.76	0.60	0.30	0.94	0.83
	Slope	0.80	0.44	0.26	1.26	1.05
	Intercept	-0.86	-2.75	-3.19	0.49	0.70

[a]Values obtained without considering the outlier complexes OAH/G4 and OAM/G4 simulated with GAFF parameters.

Table S7: Experimental and predicted binding free energy values (in kcal/mol), using the GAFF LIE model with and without ΔE_{str} values, for all CB[7-8]/steroids complexes. Also, the van der Waals and the electrostatic contributions to the ligand interaction energies used to compute the final binding free energy values are presented.

Host	Guest	$Exp.\Delta G_b^{\circ}$	$\langle U_{L-S}^{vdw}\rangle_b - \langle U_{L-S}^{vdw}\rangle_{ub}$	$\langle U_{L-S}^{elec}\rangle_b - \langle U_{L-S}^{elec}\rangle_{ub}$	$Calc.\ \Delta G_b^{LIE}$	ΔE_{str}	$Calc.\ \Delta G_b^{LIE} + \Delta E_{str}$
CB7	Nandrolone	-9.61	-30.18 ± 0.2	-0.69 ± 0.1	-13.12 ± 0.1	3.20	-9.92 ± 0.1
	Testosterone	>-4.09	-30.39 ± 0.1	0.07 ± 0.1	-12.79 ± 0.1	4.44	-8.35 ± 0.1
CB8	β-estradiol benzoate	-8.53	-25.46 ± 0.1	1.62 ± 0	-10.62 ± 0.1	2.65	-7.97 ± 0.1
	β-estradiol	-8.56	-25.65 ± 0.2	1.42 ± 0.3	-10.74 ± 0.1	3.16	-7.58 ± 0.1
	β-estradiol acetate	-9.29	-24.96 ± 0.3	0.39 ± 0.3	-10.66 ± 0.2	1.76	-8.9 ± 0.2
	Cholic acid	-7.23	-33.24 ± 0.1	15.58 ± 0.3	-11.18 ± 0.1	3.53	-7.65 ± 0.1
	Cortisol	-9.00	-30.73 ± 0.3	5.37 ± 0.2	-12.14 ± 0.2	4.01	-8.13 ± 0.2
	Corticosterone	-9.70	-29.73 ± 0.1	4.15 ± 0.3	-11.95 ± 0.1	3.00	-8.95 ± 0.1
	Drospirenone	-10.99	-29.12 ± 0.2	-3.11 ± 0.1	-13.14 ± 0.1	2.18	-10.97 ± 0.1
	17-α-Ethinyl estradiol	-8.53	-25.48 ± 0.1	2.26 ± 0.3	-10.5 ± 0.1	2.27	-8.24 ± 0.1
	Estriol	-8.05	-26.53 ± 0.2	-0.06 ± 0.3	-11.42 ± 0.1	3.47	-7.95 ± 0.1
	Estrone	-8.72	-23.04 ± 0.2	0.43 ± 0	-9.82 ± 0.1	1.84	-7.98 ± 0.1
	Megestrol	-9.82	-28.1 ± 0.0	4.76 ± 0.1	-11.13 ± 0	3.25	-7.88 ± 0
	Nandrolone propionate	-9.12	-26.35 ± 0.4	-1.91 ± 0.1	-11.72 ± 0.2	2.24	-9.48 ± 0.2
	Nandrolone	-9.98	-26.17 ± 0.0	-1.42 ± 0.2	-11.54 ± 0.1	2.71	-8.83 ± 0.1
	Progesterone	-10.87	-27.59 ± 0.1	-2.46 ± 0.3	-12.35 ± 0.1	2.35	-10 ± 0.1
	Prednisolone	-8.18	-29.91 ± 0.1	5.22 ± 0.4	-11.82 ± 0.1	4.07	-7.75 ± 0.1
	Testosterone	-10.96	-26.82 ± 0.3	-1.84 ± 0.1	-11.9 ± 0.1	2.12	-9.78 ± 0.1
	[a]Spironolactone	-8.18	-33.84 ± 0.2	-1.82 ± 0.8	-14.92 ± 0.3	2.99	-11.92 ± 0.3
	Cholesterol	>-6.82	-29.01 ± 0.6	-0.22 ± 0.1	-12.5 ± 0.3	1.73	-10.76 ± 0.3
	7-Dehydrocholesterol	>-6.82	-22.96 ± 2.8	-0.02 ± 0.1	-9.7 ± 1.2	2.11	-7.59 ± 1.2

[a]Compound considered as an outlier and not included to compute statistical metrics.

Table S8: Statistical metrics used to assess the accuracy of the computed ΔG_b^{LIE} values for the CB[7-8]/steroids complexes using the GAFF LIE model with and without ΔE_{str} values.

Metric	[a]ΔG_b^{LIE}	[a]$\Delta G_b^{LIE} + \Delta E_{str}$
RMSE	2.45	0.81
MAE	2.27	0.67
R	0.55	0.82
R^2	0.30	0.67
Slope	0.46	0.77
Intercept	-7.24	-1.55

[a]Spironolactone was considered as an outlier and it was not included to compute statistical metrics.

Table S9: Statistical metrics used to assess the accuracy of the computed ΔG_b^{LIE} values for the test set using the GAFF LIE model with ΔE_{str} values, as presented in Equation 2 from the main text.

Metric	[a]Overall	[a]OAH	[a]OAM	BCD	CB7
RMSE	1.14	0.74	1.28	0.90	1.35
MAE	0.96	0.65	1.21	0.75	1.17
R	0.84	0.86	0.95	0.78	0.90
R^2	0.71	0.74	0.89	0.61	0.82
Slope	0.72	0.56	0.43	0.83	0.97
Intercept	-0.87	-1.98	-1.61	-1.30	0.93

[a]Values obtained without considering the outlier complexes OAH/G4 and OAM/G4 simulated with GAFF parameters.

Table S10: Experimental and predicted binding free energy values (in kcal/mol), using the GAFF LIE model with and without ΔE_{str} values, for all host-guest systems in the Muddana set. Also, the van der Waals and electrostatic contributions to the ligand interaction energies used to compute binding free energy values are presented.

Host	Guest	$Exp.\Delta G_b^{\circ}$	$\langle U_{L-s}^{vdw}\rangle_b - \langle U_{L-s}^{vdw}\rangle_{ub}$	$\langle U_{L-s}^{elec}\rangle_b - \langle U_{L-s}^{elec}\rangle_{ub}$	$Calc.\ \Delta G_b^{LIE}$	ΔE_{str}	$Calc.\ \Delta G_b^{LIE} + \Delta E_{str}$
CB7	M1	-5.30	-26.26 ± 0.1	-5.29 ± 0.1	-12.35 ± 0.1	7.04	-5.31 ± 0.1
	M2	-6.10	-19.77 ± 0.1	-1.36 ± 0.1	-8.77 ± 0.1	6.95	-1.82 ± 0.1
	M3	-6.20	-25.48 ± 0.1	-3.39 ± 0.2	-11.63 ± 0.1	7.83	-3.8 ± 0.1
	M4	-6.40	-19.33 ± 0.1	3.07 ± 0	-7.7 ± 0.1	2.01	-5.69 ± 0.1
	M5	-6.80	-17.97 ± 0.2	-10.71 ± 0.2	-9.87 ± 0.1	2.32	-7.55 ± 0.1
	M6	-6.80	-18.22 ± 0	-3.08 ± 0.1	-8.45 ± 0	3.99	-4.46 ± 0
	M7	-7.40	-29.25 ± 0.1	-0.41 ± 0	-12.66 ± 0	5.58	-7.08 ± 0
	M8	-7.70	-15.23 ± 0	-8.44 ± 0.7	-8.24 ± 0.2	2.32	-5.92 ± 0.2
	M9	-7.70	-18.2 ± 0.8	-9.37 ± 0.1	-9.7 ± 0.4	1.35	-8.35 ± 0.4
	M10	-8.70	-15.7 ± 0.2	-11.03 ± 0.3	-8.96 ± 0.1	3.38	-5.58 ± 0.1
	M11	-9.60	-16.63 ± 0.1	-2.45 ± 0.1	-7.64 ± 0.1	0.84	-6.8 ± 0.1
	M12	-9.80	-21.17 ± 0.1	-2.25 ± 0.1	-9.55 ± 0.1	2.11	-7.44 ± 0.1
	M13	-10.20	-22.85 ± 0.1	-3.09 ± 0.5	-10.44 ± 0.1	2.46	-7.98 ± 0.1
	M14	-10.50	-28.54 ± 0	-7.6 ± 0.2	-13.79 ± 0.1	10.48	-3.31 ± 0.1
	M15	-11.00	-17.77 ± 0.3	-14.57 ± 0.3	-10.56 ± 0.2	2.76	-7.8 ± 0.2
	M16	-11.50	-31.87 ± 0.1	4.13 ± 1	-12.88 ± 0.3	5.01	-7.87 ± 0.3
	M17	-11.80	-28.48 ± 0.9	-10.68 ± 1.3	-14.38 ± 0.6	0.83	-13.55 ± 0.6
	M18	-12.80	-18.33 ± 0.1	-6.19 ± 0.2	-9.12 ± 0.1	1.34	-7.78 ± 0.1
	M19	-13.40	-27.84 ± 0.3	1.72 ± 0.4	-11.63 ± 0.2	0.55	-11.08 ± 0.2
	M20	-14.10	-28.31 ± 0.3	1.75 ± 0.4	-11.82 ± 0.2	0.56	-11.26 ± 0.2
	M21	-16.90	-29.35 ± 0.1	-11.3 ± 0.1	-14.88 ± 0	0.82	-14.06 ± 0
	M22	-17.00	-28.43 ± 0.1	-8.41 ± 0	-13.91 ± 0	0.89	-13.02 ± 0
	M23	-19.10	-28.69 ± 0.1	-4.76 ± 0.3	-13.29 ± 0.1	0.71	-12.58 ± 0.1
	M24	-19.40	-28.9 ± 0.1	-3.08 ± 0.1	-13.04 ± 0	0.95	-12.09 ± 0
	M25	-19.50	-27.82 ± 0.1	-9.22 ± 0.2	-13.81 ± 0.1	1.73	-12.08 ± 0.1

M26	-20.30	-29.03 ± 0	-6.69 ± 0.3	-13.82 ± 0.1	0.73	-13.09 ± 0.1
M27	-20.60	-32.48 ± 0.1	-9 ± 0.8	-15.76 ± 0.2	2.12	-13.64 ± 0.2
M28	-21.50	-30.33 ± 0.1	-12.47 ± 0.1	-15.54 ± 0.1	1.38	-14.16 ± 0.1

Table S11: Experimental and predicted binding free energy values (in kcal/mol), using the CGenFF LIE model, for all host-guest systems in the Muddana set. Also, the van der Waals and electrostatic contributions to the ligand interaction energies used to compute binding free energy values are presented.

Host	Guest	$Exp.\Delta G_b^\circ$	$\langle U_{L-s}^{vdw}\rangle_b - \langle U_{L-s}^{vdw}\rangle_{ub}$	$\langle U_{L-s}^{elec}\rangle_b - \langle U_{L-s}^{elec}\rangle_{ub}$	$Calc.\ \Delta G_b^{LIE}$
CB7	M1	-5.30	-18.69 ± 0	-4.5 ± 0.8	-12.7 ± 0.1
	M2	-6.10	-10.43 ± 0	-0.07 ± 0.1	-6.89 ± 0
	M3	-6.20	-10.79 ± 0.2	-9.24 ± 0.4	-7.86 ± 0.1
	M4	-6.40	-16.02 ± 0.1	-7.07 ± 0.7	-11.14 ± 0.1
	M5	-6.80	-15.09 ± 0.2	-9.78 ± 0.4	-10.74 ± 0.1
	M6	-6.80	-14.67 ± 0.1	4.93 ± 0.5	-9.29 ± 0.1
	M7	-7.40	-20.42 ± 0.1	3.92 ± 0.2	-13.17 ± 0.1
	M8	-7.70	-7.97 ± 0.1	-15.95 ± 0.2	-6.53 ± 0.1
	M9	-7.70	-15.19 ± 0.1	1.86 ± 0.3	-9.88 ± 0.1
	M10	-8.70	-13.67 ± 0.1	-10.69 ± 0.2	-9.88 ± 0.1
	M11	-9.60	-13.97 ± 0	-1.91 ± 0.1	-9.37 ± 0
	M12	-9.80	-16.67 ± 0.1	7.46 ± 0.1	-10.4 ± 0.1
	M13	-10.20	-14.63 ± 0.1	-1.97 ± 0.2	-9.81 ± 0.1
	M14	-10.50	-22.6 ± 0.1	-0.92 ± 0.4	-14.99 ± 0.1
	M15	-11.00	-13.21 ± 0.1	-14.74 ± 0.7	-9.9 ± 0.1
	M16	-11.50	-21.07 ± 0	15.65 ± 0.4	-12.65 ± 0.1
	M17	-11.80	-17.45 ± 0.8	-5.91 ± 1.4	-11.99 ± 0.6
	M18	-12.80	-14.8 ± 0.1	-2.29 ± 0.5	-9.95 ± 0.1
	M19	-13.40	-18.87 ± 1.5	0.85 ± 2.6	-12.39 ± 1.2
	M20	-14.10	-17.47 ± 1.8	1.43 ± 3.1	-11.42 ± 1.4
	M21	-16.90	-17.77 ± 0.1	-8.87 ± 0.1	-12.44 ± 0
	M22	-17.00	-18.74 ± 0	-6.85 ± 0.1	-12.91 ± 0
	M23	-19.10	-17.7 ± 0.1	-2.51 ± 0.3	-11.88 ± 0.1
	M24	-19.40	-17.81 ± 0.1	-1.96 ± 0.1	-11.91 ± 0.1
	M25	-19.50	-19.03 ± 0.1	-8.25 ± 0.5	-13.22 ± 0.1

M26	-20.30	-17.97 ± 0.1	-4.84 ± 0.4	-12.24 ± 0.1
M27	-20.60	-21.65 ± 0.4	-0.49 ± 6	-14.33 ± 0.7
M28	-21.50	-18.43 ± 0.1	-10.83 ± 0.4	-13.03 ± 0.1

Table S12: Statistical metrics used to assess the accuracy of the computed ΔG_b^{LIE} values for the Muddana set using the GAFF (with and without ΔE_{str} values) and the CGenFF LIE models.

Metric	GAFF		CGenFF
	ΔG_b^{LIE}	$\Delta G_b^{LIE} + \Delta E_{str}$	ΔG_b^{LIE}
RMSE	3.77	4.26	4.40
MAE	3.12	3.55	3.51
R	0.73	0.87	0.57
R^2	0.53	0.76	0.32
Slope	0.35	0.61	0.23
Intercept	-7.40	-1.38	-8.45

References

(1) Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K. The SAMPL4 host–guest blind prediction challenge: an overview. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 305–317.

(2) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 host–guest challenge: Are we doing better? *J. Comput.-Aided Mol. Des.* **2017**, *31*, 1–19.

(3) Assaf, K. I.; Florea, M.; Antony, J.; Henriksen, N. M.; Yin, J.; Hansen, A.; Qu, Z.-w.; Sure, R.; Klapstein, D.; Gilson, M. K.; Grimme, S.; Nau, W. M. HYDROPHOBE Challenge: A Joint Experimental and Computational Study on the HostGuest Binding of Hydrocarbons to Cucurbiturils, Allowing Explicit Evaluation of Guest Hydration Free-Energy Contributions. *J. Phys. Chem. B* **2017**, *121*, 11144–11162.

(4) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge structural database. *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.* **2016**, *72*, 171–179.

(5) ChemAxon, MarvinSketch. 2014; `http://www.chemaxon.com/products/marvin/marvinsketch/`, Accessed Feb 8, 2017.

(6) Korb, O.; Stützle, T.; Exner, T. E. PLANTS: Application of ant colony optimization to structure-based drug design. International Workshop on Ant Colony Optimization and Swarm Intelligence. 2006; pp 247–258.

(7) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(8) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(9) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for druglike molecules compatible with the CHARMM allatom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.

(10) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(11) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(12) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.

(13) Parrinello, M.; Rahman, A. Strain fluctuations and elastic constants. *J. Chem. Phys.* **1982**, *76*, 2662–2666.

(14) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(15) Åqvist, J.; Medina, C.; Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng., Des. Sel.* **1994**, *7*, 385–391.

(16) Wennberg, C. L.; Murtola, T.; Pall, S.; Abraham, M. J.; Hess, B.; Lindahl, E. Direct-Space corrections enable fast and accurate Lorentz–Berthelot combination rule Lennard-Jones lattice summation. *J. Chem. Theory Comput.* **2015**, *11*, 5737–5746.

(17) Baron, R. *Computational drug discovery and design*; Humana Press, 2012.

(18) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

(19) Case, D.; Cerutti, D.; Cheatham, T.; Darden, T. AMBER 2017. 2017.

(20) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824–19839.