

Binding thermodynamics of host-guest systems with SMIRNOFF99Frosst from the Open Force Field Group

This manuscript ([permalink](#)) was automatically generated from slochower/smirnoff-host-guest-manuscript@81da787 on April 18, 2019.

Authors

- **David R. Slochower**

 [0000-0003-3928-5050](#) ·  [slochower](#) ·  [drslochower](#)

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA

- **Niel M. Henriksen**

 [0000-0002-7916-0757](#) ·  [nhenriksen](#)

Atomwise, Inc., San Francisco, CA 94105, USA

- **John D. Chodera**

 [0000-0003-0542-119X](#) ·  [jchodera](#) ·  [jchodera](#)

Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065

- **Michael K. Gilson**

 [0000-0002-3375-1738](#) ·  [michaelkgilson](#)

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA

Abstract

Designing ligands that bind their target with high affinity and specificity is a key step in small-molecule drug discovery. Yet accurate predictions of protein-ligand binding free energies are difficult and errors in the calculations can be traced to challenges adequately sampling conformational space, ambiguous protonation states, and other causes. Noncovalent complexes between a cavity-containing host molecule and drug-like guest molecules have emerged as powerful tools for modeling protein-ligand binding. Due to their small size and extensive experimental characterization, calculations of host-guest binding free energies, enthalpies, and entropies offer an opportunity to directly probe, and ultimately optimize, force fields.

The Open Force Field Initiative aims to create a modern, open software infrastructure for automatically generating and validating force fields using high-quality data sets. The first force field to arise out of this effort, named SMIRNOFF99Frosst, has one tenth the number of parameters of a typical general small molecule force field, such as GAFF, yet predicts binding thermodynamics that are on average, at least as accurate. Here, we report the results of free energy calculations on 43 α and β -cyclodextrin host-guest pairs for which experimental data are available. Our calculations were performed using the attach-pull-release method as implemented in the open source package, `pAPRika`. On binding free energies, the root mean square error of the predictions relative to experiment is 0.82 kcal/mol for SMIRNOFF99Frosst and 1.58 kcal/mol for GAFF version 2.1. These results suggest significant room for improvement in force fields, and will help create a transparent and robust method of evaluating future candidate parameter sets from the Open Force Field Group. Improving the performance of force fields for predicting binding affinities will help reduce the timescale and cost required to generate drug candidates.

Introduction

Methods

Choice of host-guest system

In this study, we report the binding thermodynamics of 43 host-guest complexes (Figure 1 and Table 1) computed using three different force fields. The complexes consist of either α - or β -cyclodextrin as host molecules and a series of small molecule guests containing ammonium, carboxylate, or cyclic alcohol functional groups. Cyclodextrins are cyclic polymers consisting of six (α CD) or seven (β CD) glucose monomers in the shape of a truncated cone. The equilibrium constants and standard molar enthalpies of binding for these 43 complexes have been measured using isothermal titration calorimetry [1] and computationally in [2]. As in Henriksen, et al. [2], only a single structural isomer was considered for the 1-methylammonium guests.

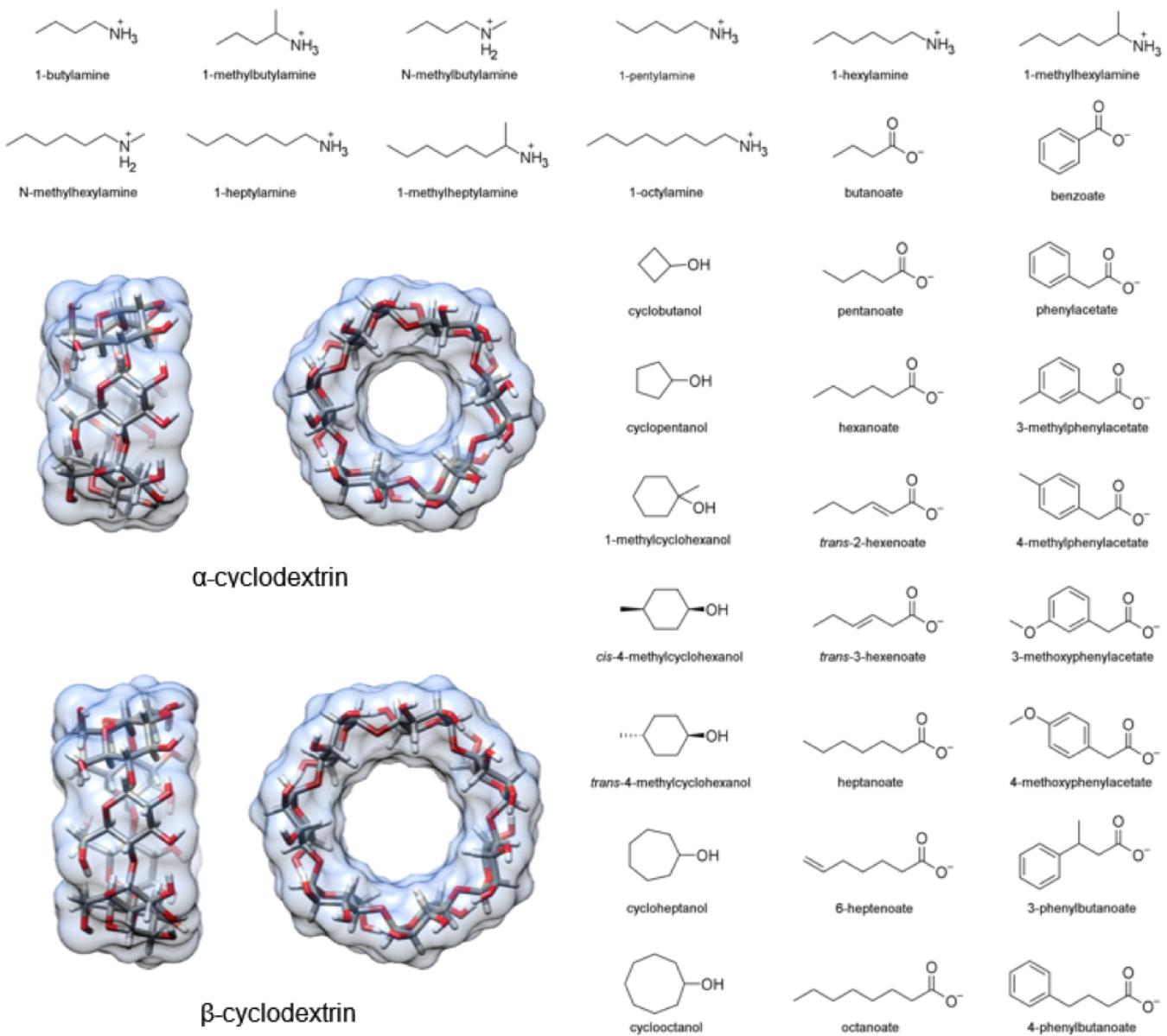


Figure 1: Structures of the two cyclodextrin hosts and 33 guest molecules in this study which together comprise 43 unique host-guest pairs.

Table 1: The 43 unique host-guest combinations used in this study. The formal charge of each guest is listed in brackets. The guest names correspond to Tables 1 and 2 in Rekharsky et al. [1]. ^a Only the R enantiomer was considered. ^b Only the S enantiomer was considered.

Host-guest ID	Host	Guest[charge]
a-bam	α CD	1-butylamine[+1]
a-nmb	α CD	n-methylbutylamine[+1]
a-mba	α CD	1-methylbutylamine[+1] ^a
a-pam	α CD	1-pentylamine[+1]
a-ham	α CD	1-hexylamine[+1]
a-nmh	α CD	n-methylhexylamine[+1]
a-mha	α CD	1-methylhexylamine[+1] ^a
a-hpa	α CD	1-heptylamine[+1]
a-mhp	α CD	1-methylheptylamine[+1] ^b
a-oam	α CD	1-octylamine[+1]

Host-guest ID	Host	Guest[charge]
b-ham	β CD	1-hexylamine[+1]
b-mha	β CD	1-methylhexylamine[+1] ^a
b-oam	β CD	1-octylamine[+1]
a-cbu	α CD	cyclobutanol[0]
a-cpe	α CD	cyclopentanol[0]
a-chp	α CD	cycloheptanol[0]
a-coc	α CD	cyclooctanol[0]
b-cbu	β CD	cyclobutanol[0]
b-cpe	β CD	cyclopentanol[0]
b-mch	β CD	1-methylcyclohexanol[0]
b-m4c	β CD	cis-4-methylcyclohexanol[0]
b-m4t	β CD	trans-4-methylcyclohexanol[0]
b-chp	β CD	cycloheptanol[0]
b-coc	β CD	cyclooctanol[0]
a-but	α CD	butanoate[-1]
a-pnt	α CD	pentanoate[-1]
a-hex	α CD	hexanoate[-1]
a-hx2	α CD	trans-2-hexenoate[-1]
a-hx3	α CD	trans-3-hexenoate[-1]
a-hep	α CD	heptanoate[-1]
a-hp6	α CD	6-heptenoate[-1]
a-oct	α CD	octanoate[-1]
b-pnt	β CD	pentanoate[-1]
b-hex	β CD	hexanoate[-1]
b-hep	β CD	heptanoate[-1]
b-ben	β CD	benzoate[-1]
b-ph _a	β CD	phenylacetate[-1]
b-mp3	β CD	3-methylphenylacetate[-1]
b-mp4	β CD	4-methylphenylacetate[-1]
b-mo3	β CD	3-methoxyphenylacetate[-1]
b-mo4	β CD	4-methoxyphenylacetate[-1]
b-pb3	β CD	3-phenylbutanoate[-1]
b-pb4	β CD	4-phenylbutanoate[-1]

Application of force field parameters

We sought to compare force fields directly, and as such, attempted to minimize additional differences between the simulations with each force field. In all simulations, we applied AM1-BCC [3,4] partial atomic charges to both the host and guest molecules using the antechamber program. The host charges were calculated using a single glucose molecule with methoxy caps on the O1 and O4 alcohols (Figure 2); each

glucose monomer in the cyclodextrin polymer has identical charges. We used TIP3P water [5] and Joung-Cheatham monovalent ion parameters [6] in each simulation set.

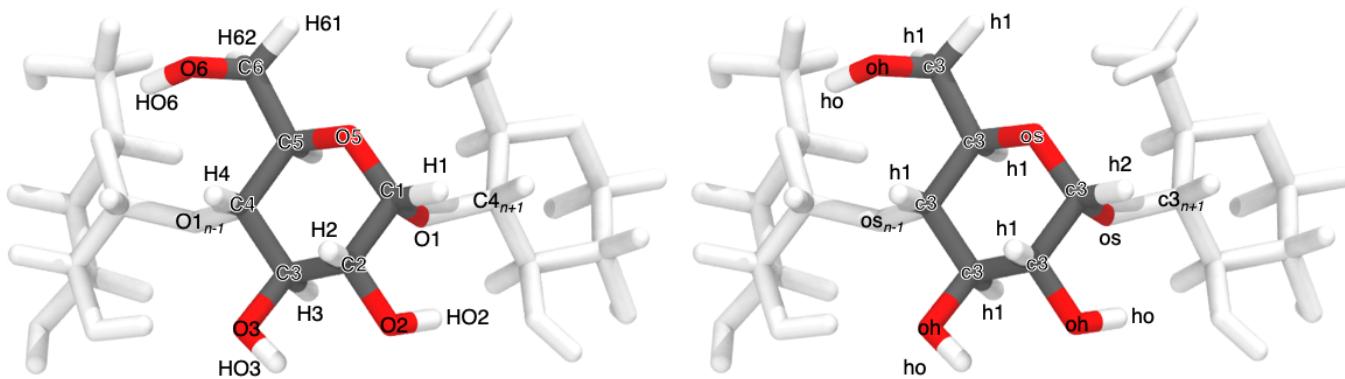


Figure 2: Atom names (left) and GAFF atom types (right) for a glucose monomer in α CD shown with two flanking monomers. The remaining three glucose monomers are hidden for clarity.

GAFF v1.7 bond, angle, torsion, and Lennard-Jones parameters were applied using the `tleap` program distributed with AmberTools16. These simulations were performed as part of Henriksen, et al. [2] and are described in additional detail therein.

GAFF v2.1 parameters were applied in an identical manner to the GAFF v1.7 parameters, using the `tleap` program distributed with AmberTools18 and substituting `leaprc.gaff` for `leaprc.gaff2` in the `tleap` input file. In GAFF v2.1, the bond and angle parameters have been updated to reproduce small molecule geometries obtained from high-level quantum mechanical calculations. The force constants for the bond and angle parameters were tuned to reproduce the vibrational spectra of over 600 molecules. The torsion parameters were optimized to reproduce the rotational potential energy surface of 400 model compounds. Finally, the Lennard-Jones coefficients were redeveloped to reproduce interaction energies and pure liquid properties. See #7.

To apply SMIRNOFF99Frosst parameters, we followed a [multistep process](#), beginning with the prepared GAFF v1.7 files. The host and guest molecules were parameterized with the Open Force Field Toolkit version 0.0.3, SMIRNOFF version 1.0, and SMIRNOFF99Frosst version 1.0.5. Once parameterized with SMIRNOFF99Frosst, ParmEd [7] was used to combine the host and guest with the solvent and ions, which retained their TIP3P water parameters and Joung-Cheatham ion parameters, respectively. See #8.

Thermodynamic calculation

We used the attach-pull-release (APR) method as implemented in the open source package pAPRika version 0.0.3, to calculate absolute binding free energies. A complete description of the APR method has been characterized in the literature [10,11,8,9]. The attachment and release phases consisted of 15 independent windows. During the attachment phase, the force constants on the host and guest are scaled by a λ parameter that goes from $\lambda = 0$, at which point all restraints are turned off, to $\lambda = 1$, at which point all restraints are at their maximum force constant. The λ windows are more densely spaced where the force constant is smaller to improve sampling along highly curved regions of the potential of mean force. Conformational restraints were applied between neighboring glucose units of the cyclodextrin to limit the incursion of monomers into the host cavity. These restraints were applied along the pseudodihedrals $O5_n - C1_n - O1_n - C4_{n+1}$ and $C1_n - O1_n - C4_{n+1} - C5_{n+1}$ to improve convergence and sampling of the bound state (Figure 2 for atom names). To further improve sampling, we applied a hard wall restraint that confined the guest molecule to within a sphere of 12.3 and 13.5 Å of α CD and β CD, respectively, during the bound state. The release phase is the conceptual reverse of the attach phase, in which the conformational restraints on the host are gradually turned off ($\lambda = 1 \rightarrow 0$) in the absence of the guest. This explicit release phase is performed once for α CD and once for β CD. Finally, an analytic correction is performed to

compute the work of moving the guest from the restricted volume enforced by the APR restraints to standard state at 1 M concentration.

The pulling phase consisted of 45 independent, equally spaced windows. During the pulling phase, the λ parameter represents the target value of a distance restraint with constant force constant. This target distance is increased uniformly in 45 increments of 0.4 Å, yielding windows that separate the host and guest by 18 Å over the course of the calculation.

Due to the asymmetry of the primary and secondary alcohols of cyclodextrin (Figure 17), and the lack of symmetry of the small molecule guests, there are generally two distinct binding poses that do not interconvert during the simulation timescale. To account for this effect, we separately compute the binding free energy and enthalpy for each orientation [8] and combine the results to produce a single value for each host-guest combination. In this manuscript, we refer to calculations whereby the guest is pulled out of the primary face of cyclodextrin with a `-p` suffix and calculations whereby the guest is pulled out of the secondary orientation with a `-s` suffix.

Thermodynamic integration was used to compute the binding free energy (ΔG). The binding enthalpy (ΔH) was computed as the difference in mean potential energy of the bound state (in the absence of any restraints) and the unbound state (where the guest is held far away from the host, but the conformational restraints on the host are disabled). The binding entropy (ΔS) was computed by subtraction using ΔG and ΔH .

Simulations

Simulations were performed with the `pmemd.cuda` module of AMBER 16 (calculations of the GAFF v1.7 force field) and AMBER 18 (calculations of the GAFF v2.1 and SMIRNOFF99Frosst force fields) molecular dynamics software [12]. Each window for each system was independently solvated and simulated.

Solvation consisted of 2000 TIP3P waters for the α CD systems and 2210 waters for the β CD systems in an orthorhombic box. The host and guest were oriented via non-interacting dummy atoms along the simulation box z axis to minimize the amount of solvent required. Each simulation contained enough Na^+ or Cl^- ions to neutralize the host-guest complex and an additional 50 mM NaCl to match the experimental conditions in [1]. In the GAFF simulations, hydrogen mass repartitioning [13] was used to adjust the mass of hydrogen atoms to 3.024 Da by transferring mass from bonded heavy atoms, enabling a simulation timestep of 4 fs. Equilibration consisted of 50,000 steps of energy minimization, 100 ps of heating from 0 to 300 K, followed by 2000 ps of additional NPT simulations. A Langevin thermostat, the Monte Carlo barostat, a nonbonded cutoff of 9 Å and default PME parameters, were used for the NPT simulations. Production NPT simulations were run for a minimum of 5 ns and maximum of 50 ns per window, except for the windows used to calculate the enthalpy, which were simulated for 1 μ s. In the GAFF v1.7 and GAFF v2.1 simulations, the exact length of each window's simulation was determined by the restraint coordinate uncertainty. For restraint energy U in λ window i , each window (except for the windows used to calculate ΔH) was simulated until the value $w(\lambda_i)$:

$$w(\lambda_i) = \begin{cases} \left\langle \frac{\partial U}{\partial \lambda_i} \right\rangle_{\text{SEM}} \cdot \frac{\lambda_{i+1}}{2} & i = 0 \\ \left\langle \frac{\partial U}{\partial \lambda_i} \right\rangle_{\text{SEM}} \cdot \frac{\lambda_{i+1} - \lambda_{i-1}}{2} & i \in [1, N-1] \\ \left\langle \frac{\partial U}{\partial \lambda_i} \right\rangle_{\text{SEM}} \cdot \frac{1 - \lambda_{i-1}}{2} & i = N \end{cases}$$

fell below a threshold of 0.02 kcal/mol during the attach phase and 0.1 kcal/mol during the pull phase. In contrast, SMIRNOFF99Frosst production NPT simulations were run for 10 ns per window, except for the first and last window which ran for 1 μ s to calculate ΔH .

Statistics

The uncertainty in each simulation window was estimated using blocking analysis [14] and propagated into an uncertainty for ΔG after integrating 100,000 curves and an uncertainty for ΔH by adding in quadrature.

For each force field, we computed the root mean squared error (RMSE), mean signed error (MSE), the coefficient of determination (R^2), Kendall's rank correlation coefficient (τ), and the slope and intercept of the computed properties relative to the experimental values. We also test how the force fields compare to each other using these statistics. Comparisons with experiment have 43 data points, for the 43 unique host-guest complexes listed in Table 1; comparisons between force fields have 86 data points, representing the two separate calculations performed for each host-guest pair.

The overall RMSE and R^2 values for each comparison are reported as the mean with the 95% confidence interval in brackets. These values are estimated by using the uncertainties assigned to each data point to create 100,000 hypothetical graphs through resampling with replacement. The R^2 values for each functional group subset is also reported in the bottom right corner in each graph.

Results and discussion

This results section is organized as follows. We first present a comparison of SMIRNOFF99Frosst and two iterations of the General AMBER Force Field (GAFF [15]) on predicting binding free energies (ΔG) and binding enthalpies (ΔH) of small molecule guests to α -cyclodextrin (α CD) and β -cyclodextrin (β CD). We then detail how the behaviors of guest molecules changes between force fields and finally we summarize the parameter differences between SMIRNOFF99Frosst and GAFF along with the effects of the parameter differences.

Binding free energies and enthalpies

SMIRNOFF99Frosst does about as well as, or even better than GAFF v1.7, despite have far fewer numerical parameters on predicted ΔG and ΔH compared to the values measured with isothermal titration calorimetry. SMIRNOFF99Frosst has an overall agreement of under 1 kcal/mol on binding free energies and under 2 kcal/mol on binding enthalpies across the 43 host-guest systems (Figure 3). Overall, the correlation between SMIRNOFF99Frosst and experiment is mediocre, ranging from an R^2 value of 0.34 to 0.54 for ΔG . The results are more varied for ΔH with R^2 ranging from 0.09 to 0.54. In particular, SMIRNOFF99Frosst underestimates the binding free energy and binding enthalpy of cyclic alcohols and mostly underestimates the binding entropy (Figure 18). Guests containing ammonium or carboxylate functional groups appear to be equally over- and under-estimated for binding free energy.

Notably, SMIRNOFF99Frosst does better on α CD than β CD (Figure 22).

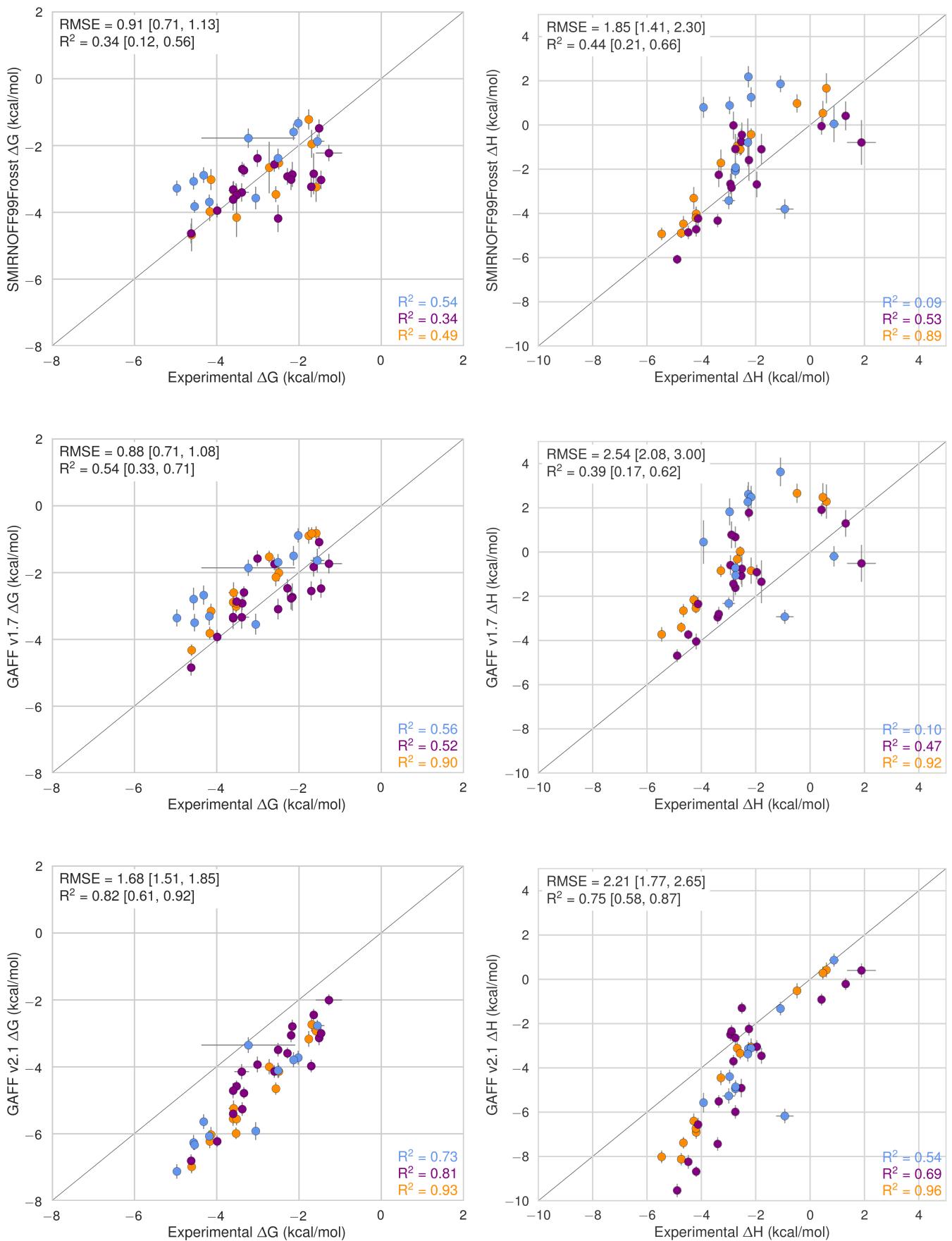


Figure 3: Comparison of calculated absolute binding free energies (ΔG) and binding enthalpies (ΔH) with experiment with SMIRNOFF99Frosst parameters (top), GAFF v1.7 parameters (middle), or GAFF v2.1 parameters (bottom) applied to both host and guest. The orange, blue, and purple coloring distinguish the functional group of the guest as an ammonium, alcohol, or carboxylate, respectively.

Likewise, GAFF v1.7 has very similar properties with an RMSE for ΔG of 0.88 kcal/mol and RMSE for ΔH of 2.54 kcal/mol. Both SMIRNOFF99Frosst and GAFF v1.7 systematically underestimate the binding free energy for cyclic alcohols except for beta-cyclodextrin with cyclopentanol (b-cpe).

Overall, GAFF v2.1 has a larger RMSE than either SMIRNOFF99Frosst or GAFF v1.7 on binding free energies, but strikingly strong correlations with the experimental values across all three functional group classes. Compared to SMIRNOFF99Frosst and GAFF v1.7, GAFF v2.1 uniformly overestimates both the binding free energy and binding enthalpy.

SMIRNOFF99Frosst does a great job predicting the binding free energy of α CD with alcohols, with four of the five points nearly falling on the line of identity.

Show highlights here

Likewise, SMIRNOFF99Frosst does a great job with amines (?) with three of the four points falling nearly on the line of identity.

With β CD, on teh other hand, things are a little different: two of the three carboxylates are overestimated, ammoniums are great, but the alcohols are all over the place (two overestimated, two underestimated.)

This difference is likely because... (?) of flexibility? (Is alpha more flexible than beta?)

Guest preferences for primary and secondary binding

There are two orientations for binding: one where the primary functional group is interacting with the more narrow opening primary alcohols and one with the polar group oriented out of the wider, secondary opening.

The experimental data is a Boltzmann weighted ensemble of these two orientations.

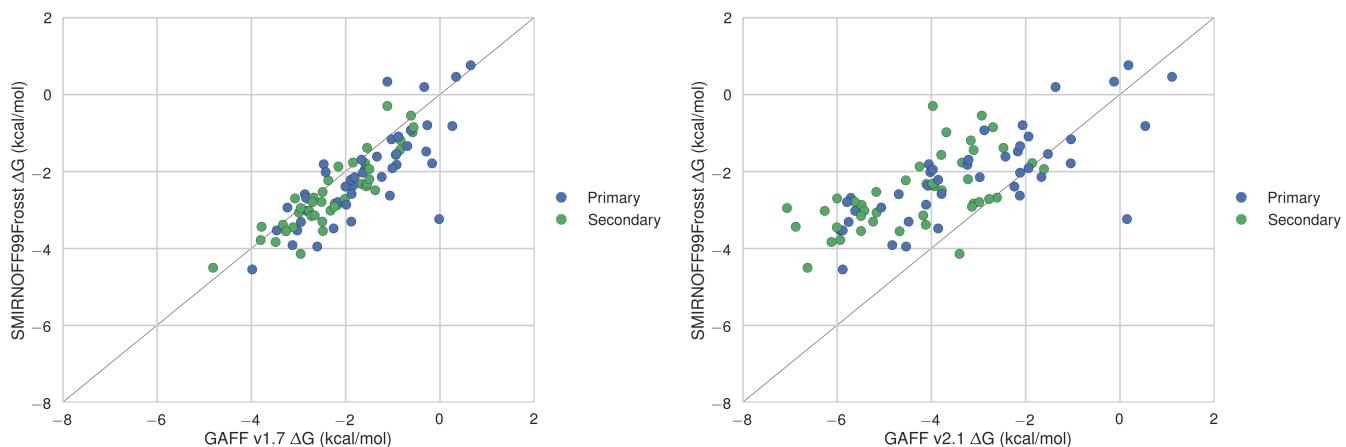


Figure 4: Binding free energies (ΔG) with the primary orientation results colored in blue and secondary orientation results colored in green.

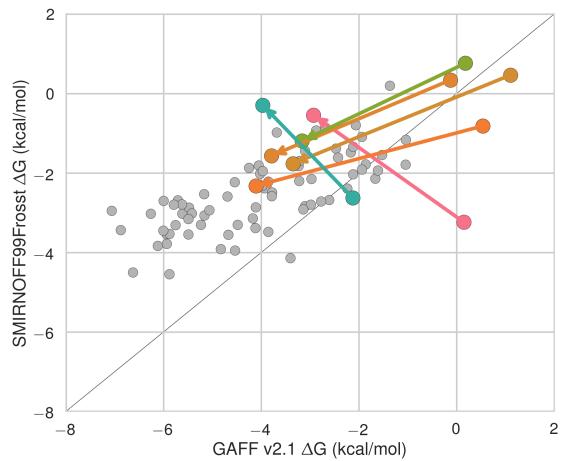
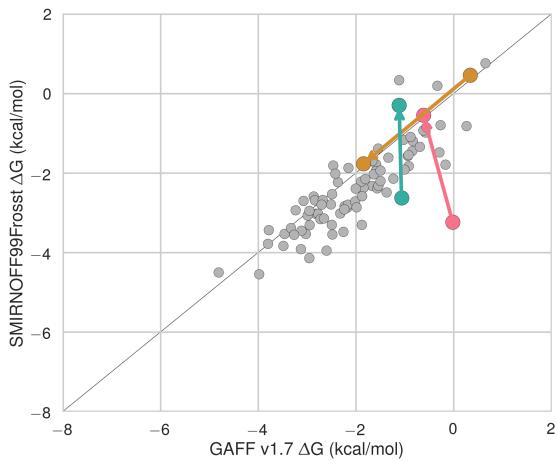


Figure 5: Binding free energies (ΔG) replotted from Figure 4 with points whose difference in binding free energy along either axis is greater than 2 kcal/mol shown in color. Arrows point from primary to secondary. This is backwards of the next plot.

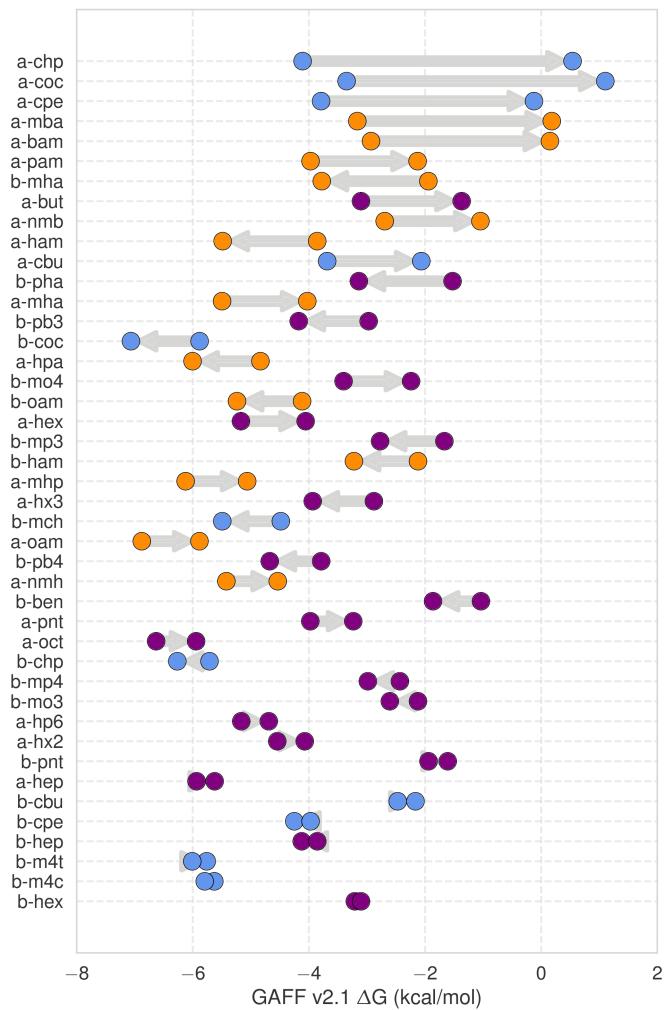
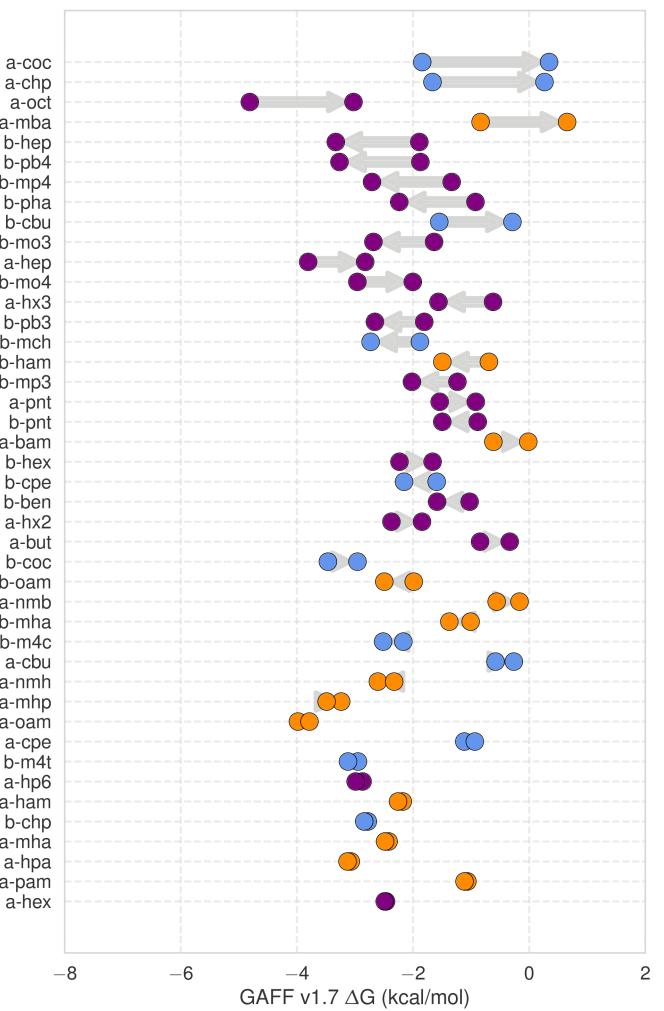
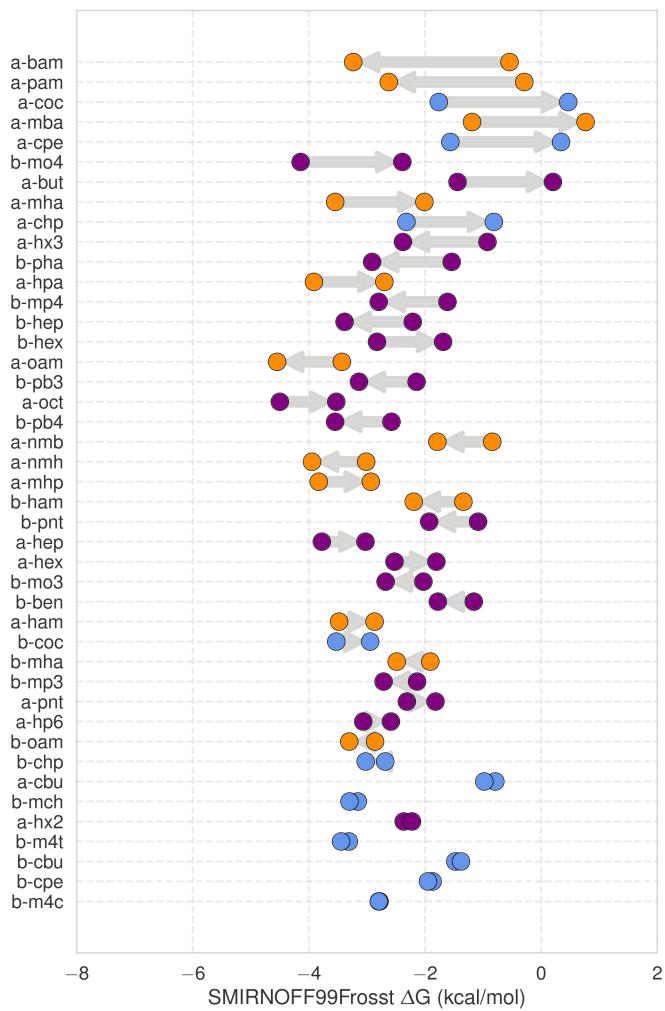


Figure 6: The differences in binding free energy (ΔG) between guests leaving through either the primary or secondary face of α CD or β CD. Arrows point from the binding affinity for the secondary to the binding affinity for the primary cavity. The systems with the largest ΔG difference are placed at the top.

Guest preferences for α CD and β CD

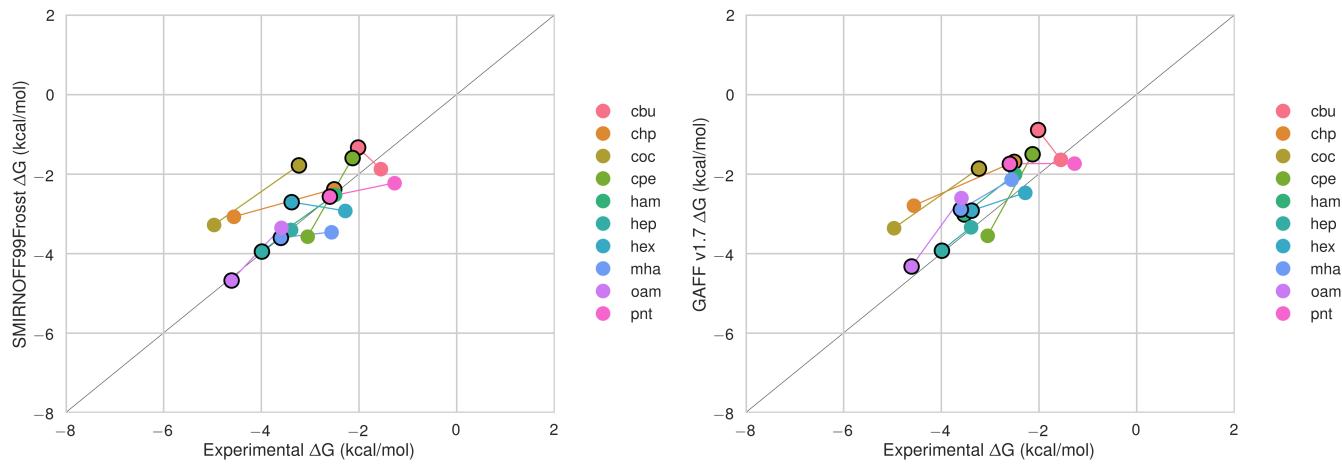


Figure 7C

Figure 7: The differences in binding free energy (ΔG) between the same guest binding to either α CD or β CD. The binding affinity for α CD is circled in black.

Functional group interactions

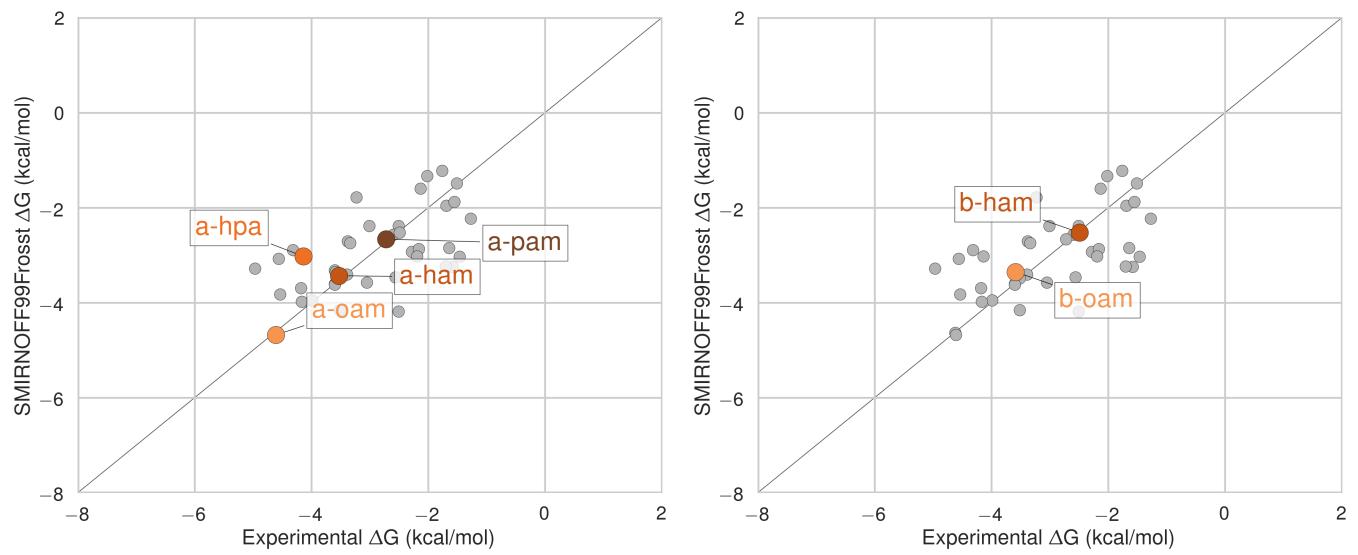


Figure 8: Binding free energy (ΔG) comparisons showing ammonium guests in color and highlighted. Darker colors indicate...

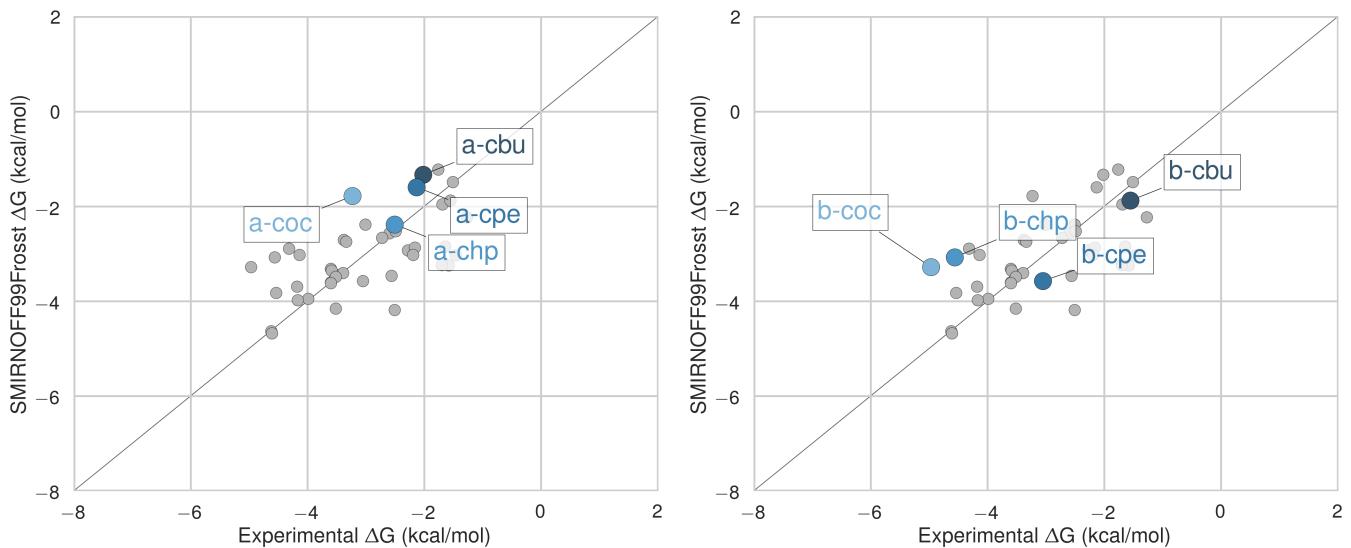


Figure 9: Binding free energy (ΔG) comparisons showing alcohols guests in color and highlighted. Darker colors indicate...

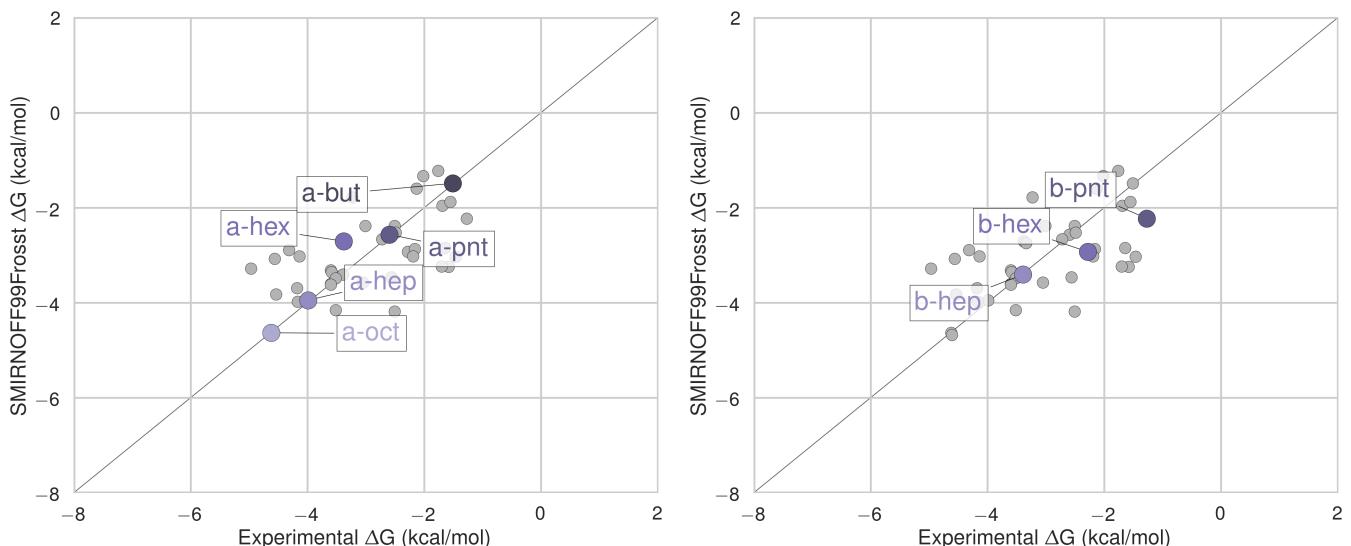


Figure 10: Binding free energy (ΔG) comparisons showing alcohols guests in color and highlighted. Darker colors indicate...

Differences in force field parameters between SMIRNOFF99Frosst and GAFF

Next, we summarize the parameter differences between SMIRNOFF99Frosst, a decendent of `parm@Frosst` and GAFF v1.7 (released circa March 2015 according to `gaff.dat` distributed with AMBER16) and GAFF v2.1 (under active development).

The σ and ϵ parameters are identical between SMIRNOFF99Frosst and GAFF v1.7. Compared to GAFF v2.1, SMIRNOFF99Frosst has deeper well depths for oxygens and decreased σ values for the hydroxyl hydrogens.

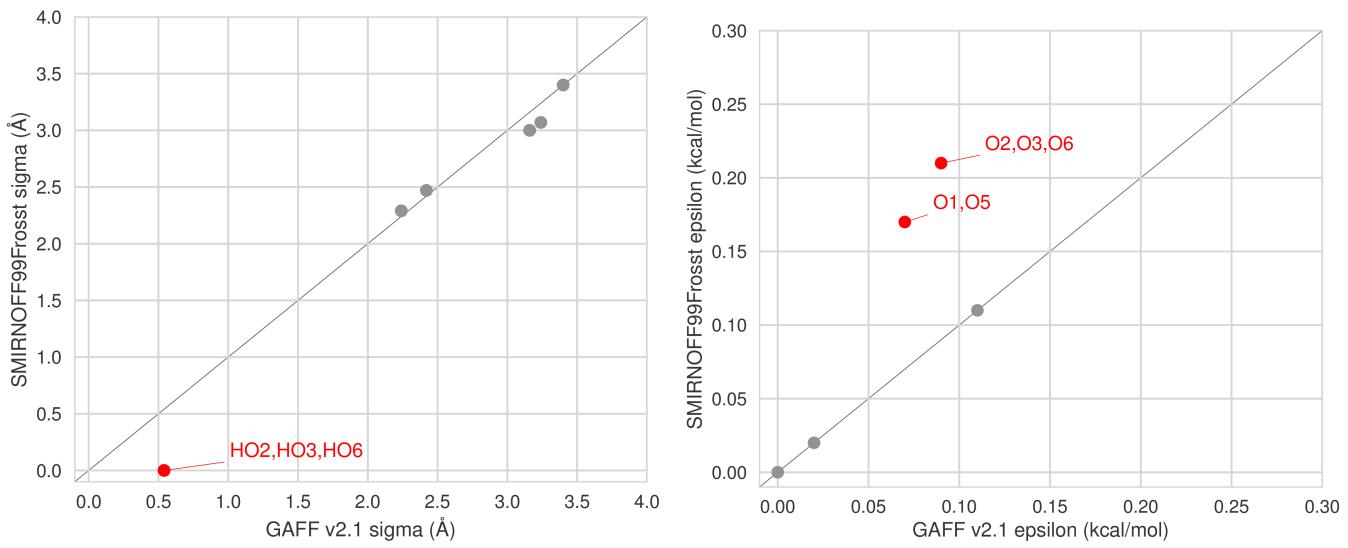


Figure 11: A comparison of Lennard-Jones nonbonded parameters for SMIRNOFF99Frosst and GAFF v2.1. Values that differ by more than 10% are labeled in red. Atom names refer to Figure 2.

Compared to GAFF v1.7, SMIRNOFF99Frosst tends to have slightly larger bond force constants, except for the O–H hydroxyl bond force constant, which is much stronger. In GAFF v2.1, the O–H hydroxyl bond force constant is consistent with SMIRNOFF99Frosst, but the carbon-oxygen bond constants are weaker. Equilibrium bond lengths are very similar (Figure 24).

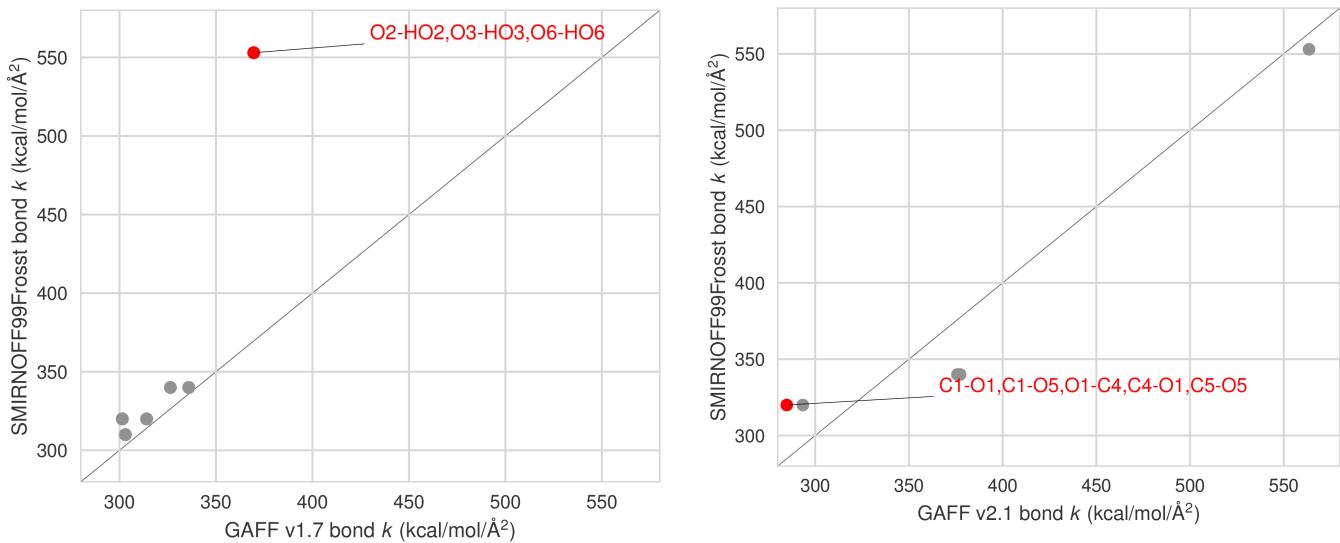


Figure 12: A comparison of bonded parameters for SMIRNOFF99Frosst, GAFF v1.7, and GAFF v2.1. Values that differ by more than 10% are labeled in red. Atom names refer to Figure 2.

Relative to GAFF v1.7 and GAFF v2.1, SMIRNOFF99Frosst has fewer unique angle parameters applied to αCD; several distinct parameters appear to be compressed into a single force constant, around 50 kcal/mol/rad². These parameters correspond to C–C–C, C–O–C, and O–C–O angles. The C–C–C angles are primarily around the ring of the glucose monomer. The C–O–C angles are both around the ring and between monomers (e.g., C₁–O₁–C₄ and C₁–O₅–C₅). Weaker force constants for these parameters may lead to increased flexibility.

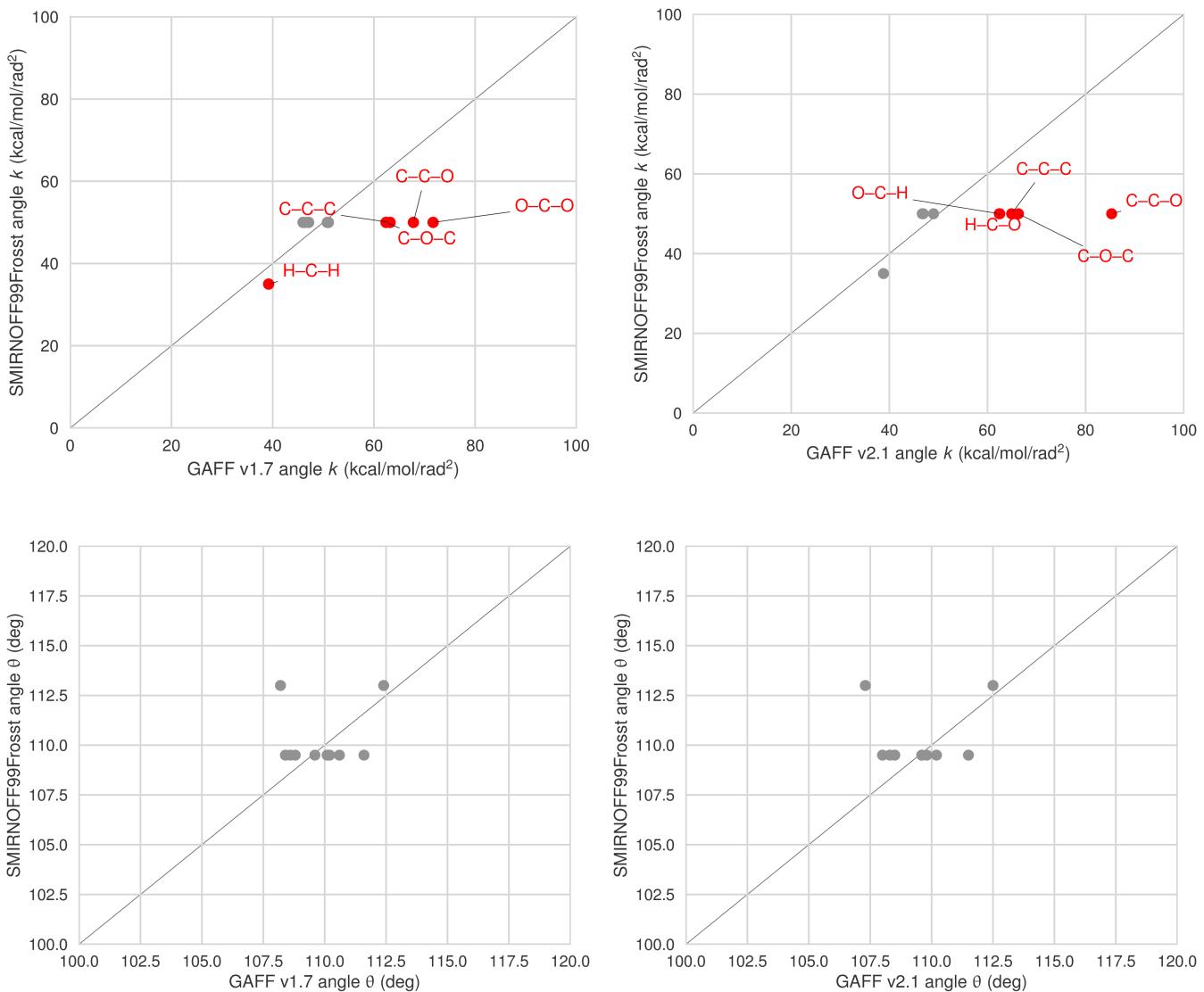


Figure 13: A comparison of angle parameters for SMIRNOFF99Frosst, GAFF v1.7, and GAFF v2.1. Values that differ by more than 10% are labeled in red. Precise atom names have been omitted to compress multiple angles with the same parameter values into a single label.

Dihedrals

Include the table of dihedral differences here.

The dihedral parameters between SMIRNOFF99Frosst and GAFF v1.7 are very similar, with the exception of... (Table 2).

Table 2: Dihedral parameter differences between SMIRNOFF99Frosst and GAFF v1.7.

							SMIRNOFF99Frosst	GAFF v1.7
Atom 1	Atom 2	Atom 3	Atom 4	Per	Phase	Height (kcal/mol)	Height (kcal/mol)	
H1	C1	C2	O2	1	0	0.25	–	
H1	C1	C2	O2	3	0	0.00	0.16	

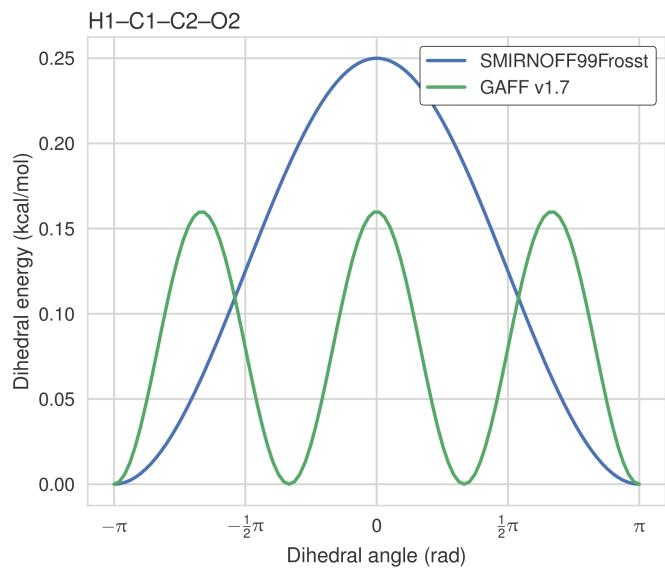


Figure 14: The dihedral energy term applied to H1-C1-C2-O2 in SMIRNOFF99Frosst and GAFF v1.7. Atom names refer to Figure 2.

The dihedral parameters in GAFF v2.1 differ significantly from those in SMIRNOFF99Frosst.

Table 3: Dihedral parameter differences between SMIRNOFF99Frosst and GAFF v2.1, where one dihedral has fewer or more periodicity terms than the corresponding term in the other force field.

						SMIRNOFF99Frosst	GAFF v2.1
Atom 1	Atom 2	Atom 3	Atom 4	Per	Phase	Height (kcal/mol)	Height (kcal/mol)
C1	C2	O2	HO2	1	0	0.25	–
C1	C2	O2	HO2	3	0	0.16	0.00
C1	O5	C5	C4	1	0	–	0.00
C1	O5	C5	C4	2	0	0.10	0.16
C1	O5	C5	C4	3	0	0.38	0.24
C1	O5	C5	C6	1	0	–	0.00
C1	O5	C5	C6	2	0	0.10	0.16
C1	O5	C5	C6	3	0	0.38	0.24
C2	C1	O5	C5	1	0	–	0.00
C2	C1	O5	C5	2	0	0.10	0.16
C2	C1	O5	C5	3	0	0.38	0.24
C2	C3	O3	HO3	1	0	0.25	–
C2	C3	O3	HO3	3	0	0.16	0.00
C5	C6	O6	HO6	1	0	0.25	–
C5	C6	O6	HO6	3	0	0.16	0.00
H1	C1	C2	O2	1	0	0.25	–
H1	C1	C2	O2	3	0	0.00	0.16
O1	C1	C2	O2	1	0	–	0.02
O1	C1	C2	O2	2	0	1.18	0.00
O1	C1	C2	O2	3	0	0.14	1.01
O2	C2	C1	O5	1	0	–	0.02

						SMIRNOFF99Frosst	GAFF v2.1
O2	C2	C1	O5	2	0	1.18	0.00
O2	C2	C1	O5	3	0	0.14	1.01
O5	C5	C6	O6	1	0	–	0.02
O5	C5	C6	O6	2	0	1.18	0.00
O5	C5	C6	O6	3	0	0.14	1.01
HO2	O2	C2	C3	1	0	0.25	–
HO2	O2	C2	C3	3	0	0.16	0.00
HO3	O3	C3	C4	1	0	0.25	–
HO3	O3	C3	C4	3	0	0.16	0.00

Table 4: Dihedral parameter differences between SMIRNOFF99Frosst and GAFF v2.1, only height differences.

						SMIRNOFF99Frosst	GAFF v2.1
Atom 1	Atom 2	Atom 3	Atom 4	Per	Phase	Height (kcal/mol)	Height (kcal/mol)
C1	C2	C3	C4	1	0	0.20	0.11
C1	C2	C3	C4	2	0	0.25	0.29
C1	C2	C3	C4	3	0	0.18	0.13
C1	C2	C3	O3	3	0	0.16	0.21
C1	O5	C5	H5	3	0	0.38	0.34
C2	C3	C4	C5	1	0	0.20	0.11
C2	C3	C4	C5	2	0	0.25	0.29
C2	C3	C4	C5	3	0	0.18	0.13
C3	C4	C5	C6	1	0	0.20	0.11
C3	C4	C5	C6	2	0	0.25	0.29
C3	C4	C5	C6	3	0	0.18	0.13
C4	C5	C6	O6	3	0	0.16	0.21
H1	C1	C2	H2	3	0	0.15	0.16
H2	C2	C3	H3	3	0	0.15	0.16
H2	C2	O2	HO2	3	0	0.17	0.11
H3	C3	C4	H4	3	0	0.15	0.16
H3	C3	O3	HO3	3	0	0.17	0.11
H4	C4	C5	H5	3	0	0.15	0.16
H5	C5	C6	H61	3	0	0.15	0.16
H5	C5	C6	H62	3	0	0.15	0.16
O1	C1	O5	C5	1	0	1.35	0.97
O1	C1	O5	C5	2	0	0.85	1.24
O1	C1	O5	C5	3	0	0.10	0.00
O2	C2	C3	C4	3	0	0.16	0.21
O2	C2	C3	O3	2	0	1.18	1.13
O2	C2	C3	O3	3	0	0.14	0.90
O3	C3	C4	C5	3	0	0.16	0.21

						SMIRNOFF99Frosst	GAFF v2.1
H61	C6	O6	HO6	3	0	0.17	0.11
H62	C6	O6	HO6	3	0	0.17	0.11

...Inter-residue dihedrals...

Table 5: Inter-residue dihedral parameter differences between SMIRNOFF99Frosst and GAFF v2.1.

ID	Atom 1	Res 1	Atom 2	Res 2	Atom 3	Res 3	Atom 4	Res 4	Per	Phase	SMIRNOFF99Frosst	GAFF v2.1
1	C1	<i>n</i>	O1	<i>n</i>	C4	<i>n+1</i>	C3	<i>n+1</i>	1	0	–	0.00
	C1	<i>n</i>	O1	<i>n</i>	C4	<i>n+1</i>	C3	<i>n+1</i>	2	0	0.10	0.16
	C1	<i>n</i>	O1	<i>n</i>	C4	<i>n+1</i>	C3	<i>n+1</i>	3	0	0.38	0.24
2	C1	<i>n</i>	O1	<i>n</i>	C4	<i>n+1</i>	C5	<i>n+1</i>	1	0	–	0.00
	C1	<i>n</i>	O1	<i>n</i>	C4	<i>n+1</i>	C5	<i>n+1</i>	2	0	0.10	0.16
	C1	<i>n</i>	O1	<i>n</i>	C4	<i>n+1</i>	C5	<i>n+1</i>	3	0	0.38	0.24
3	C2	<i>n</i>	C1	<i>n+1</i>	O1	<i>n+1</i>	C4	<i>n+1</i>	1	0	–	0.00
	C2	<i>n</i>	C1	<i>n+1</i>	O1	<i>n+1</i>	C4	<i>n+1</i>	2	0	0.10	0.16
	C2	<i>n</i>	C1	<i>n+1</i>	O1	<i>n+1</i>	C4	<i>n+1</i>	3	0	0.38	0.24
4	O1	<i>n</i>	C4	<i>n+1</i>	C3	<i>n+1</i>	O3	<i>n+1</i>	1	0	–	0.02
	O1	<i>n</i>	C4	<i>n+1</i>	C3	<i>n+1</i>	O3	<i>n+1</i>	2	0	1.18	0.00
	O1	<i>n</i>	C4	<i>n+1</i>	C3	<i>n+1</i>	O3	<i>n+1</i>	3	0	0.14	1.01
5	O1	<i>n</i>	C4	<i>n+1</i>	C5	<i>n+1</i>	O5	<i>n+1</i>	1	0	–	0.17
	O1	<i>n</i>	C4	<i>n+1</i>	C5	<i>n+1</i>	O5	<i>n+1</i>	2	0	1.18	0.00
	O1	<i>n</i>	C4	<i>n+1</i>	C5	<i>n+1</i>	O5	<i>n+1</i>	3	0	0.14	0.00

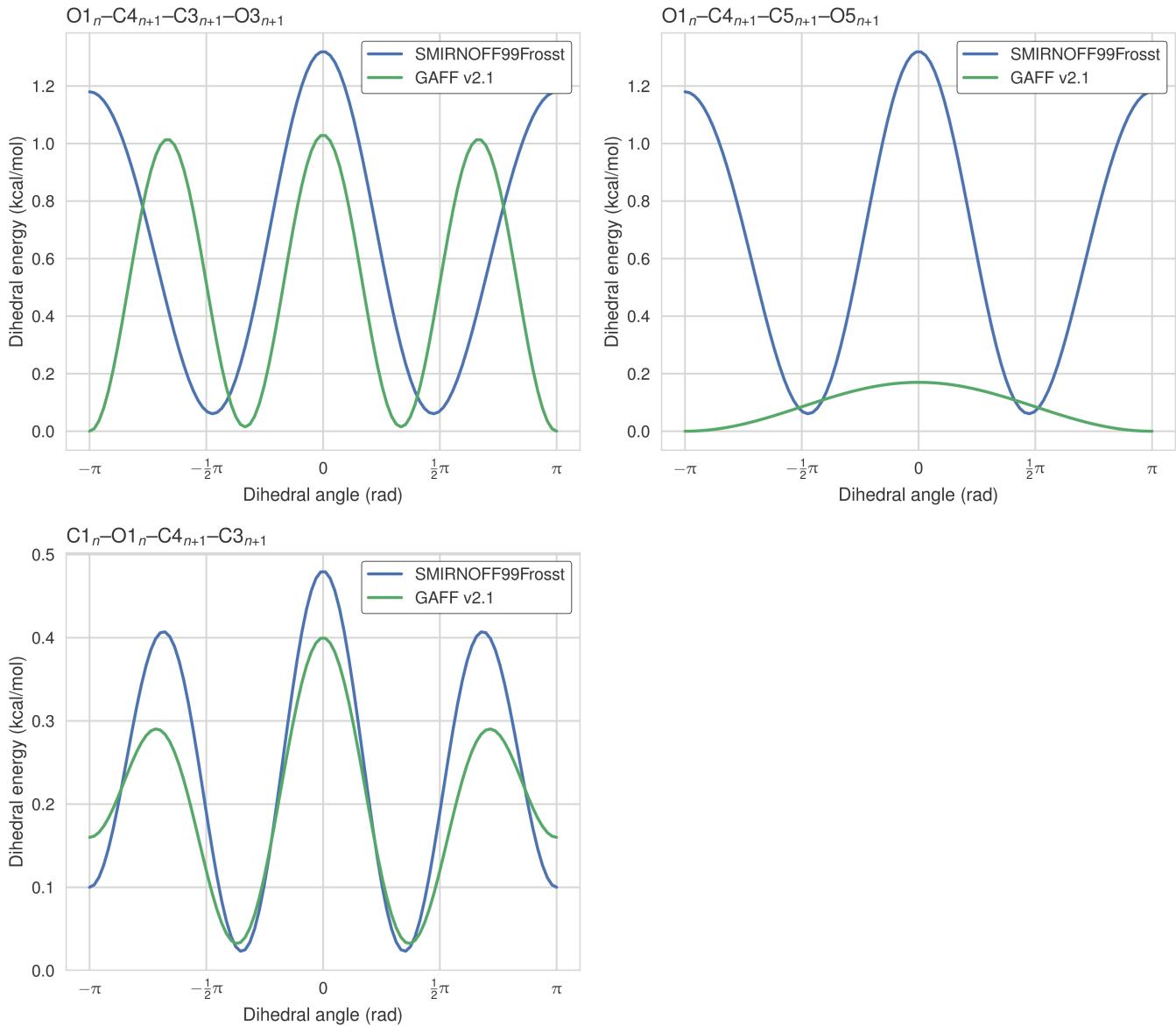


Figure 15: The dihedral energy term applied to three inter-residue dihedrals in SMIRNOFF99Frosst and GAFF v2.1. The third dihedral, affecting atoms O1_n-C4_{n+1}-C5_{n+1}-O5_{n+1} is quite significantly different, with multiple minima and barrier heights. This dihedral partially controls the rotation of glucose monomers towards or away from the interior of the cyclodextrin cavity. Surprisingly, glucose monomers in GAFF v2.1 penetrate the open cavity much less frequently than in SMIRNOFF99Frosst, despite the lower and broader dihedral energy in GAFF v2.1. Atom names refer to Figure 2.

Structural consequences of the force field parameter differences

In both SMIRNOFF99Frosst and GAFF v1.7, the average RMSD of βCD is between 2 and 2.5 Å over 43 μs of simulation. GAFF v2.1 is significantly more rigid, with an average RMSD less than 1.0 Å from the initial structure (Figure 16).

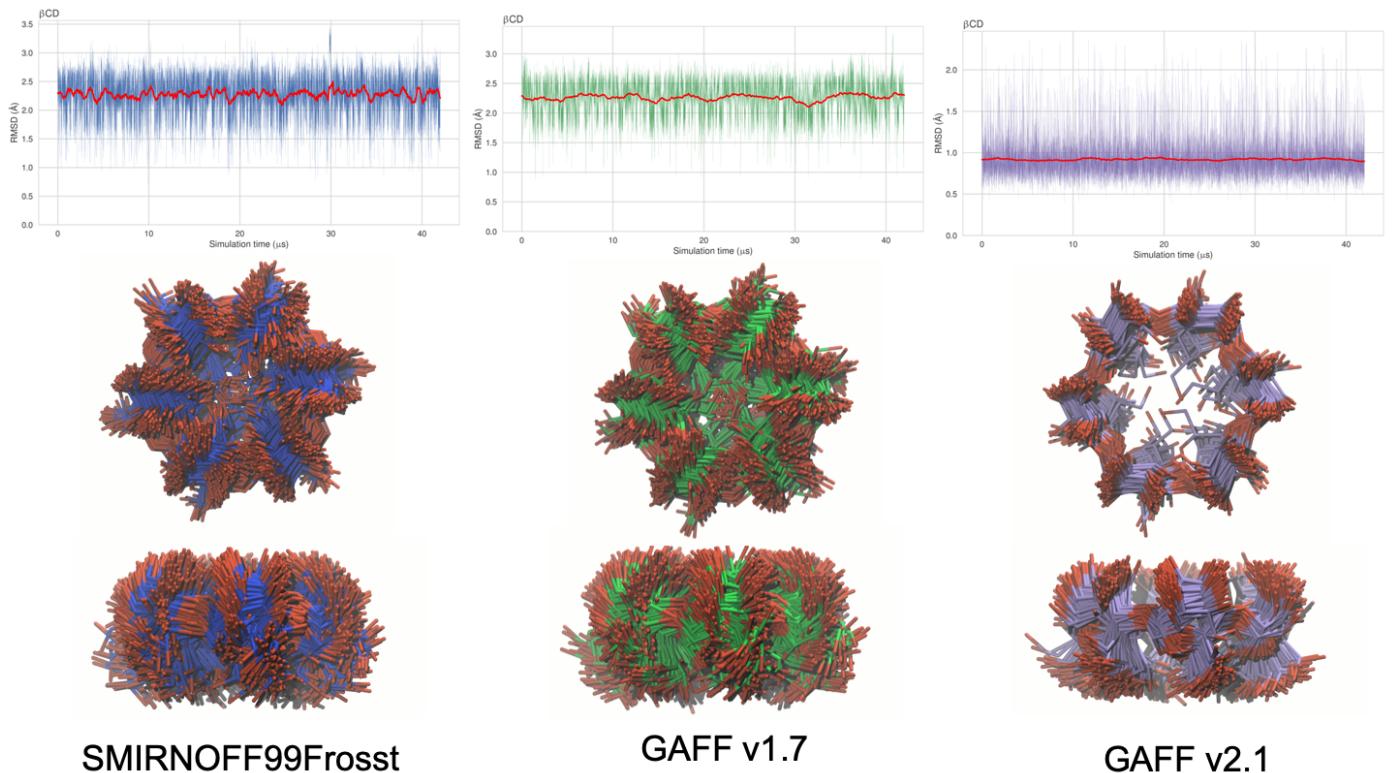


Figure 16: Top: Root mean square deviation (RMSD) of free β CD in the three force fields, all relative to the same initial structure. A 1000 frame moving average is plotted in red. Middle: to-view of the open cavity of β CD with no guest (200 snapshots over 1 μ s). Bottom: side-view of the open cavity. The carbons are colored blue in SMIRNOFF99Frosst, green in GAFF v1.7, and purple in GAFF v2.1. Hydrogen atoms have been hidden for clarity.

★ Update font size in RMSD plots.

Supporting Information

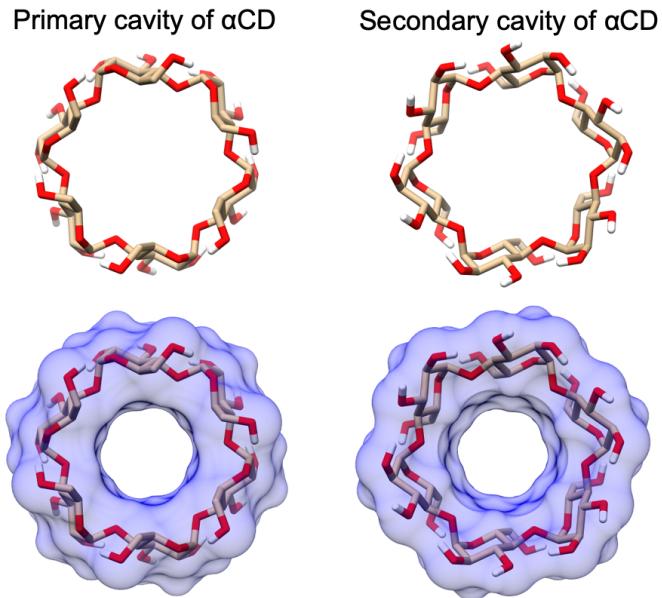


Figure 17: The primary (left) and secondary (right) cavity of α CD.

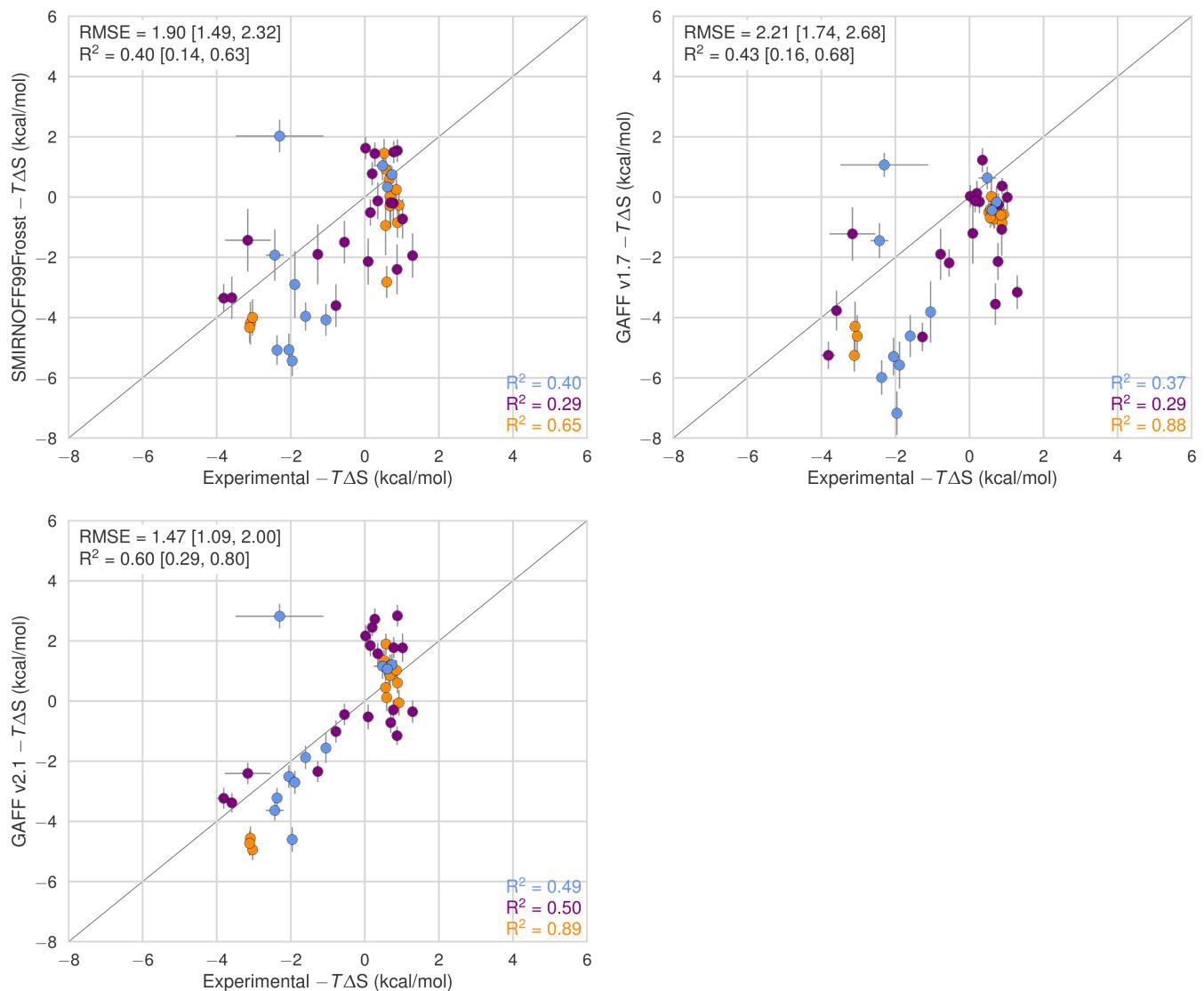


Figure 18: Comparison of calculated absolute binding entropies ($-T\Delta S$) with experiment with SMIRNOFF99Frosst parameters (top), GAFF v1.7 parameters (middle), or GAFF v2.1 parameters (bottom)

applied to both host and guest. The orange, blue, and purple coloring distinguish the functional group of the guest as an ammonium, alcohol, or carboxylate, respectively.

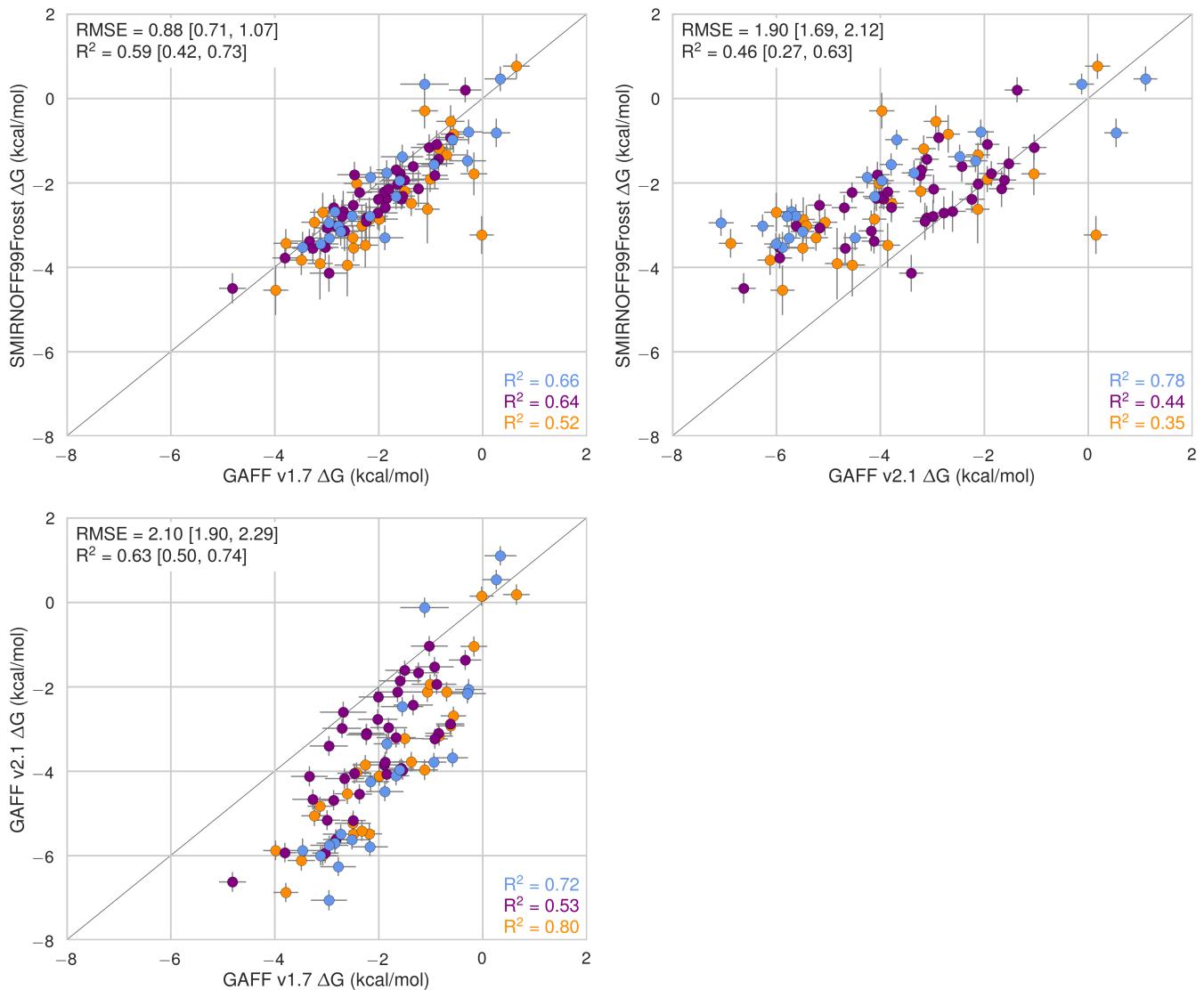


Figure 19: Comparison of calculated absolute binding free energies (ΔG) between force field combinations. The orange, blue, and purple coloring distinguish the functional group of the guest as an ammonium, alcohol, or carboxylate, respectively.

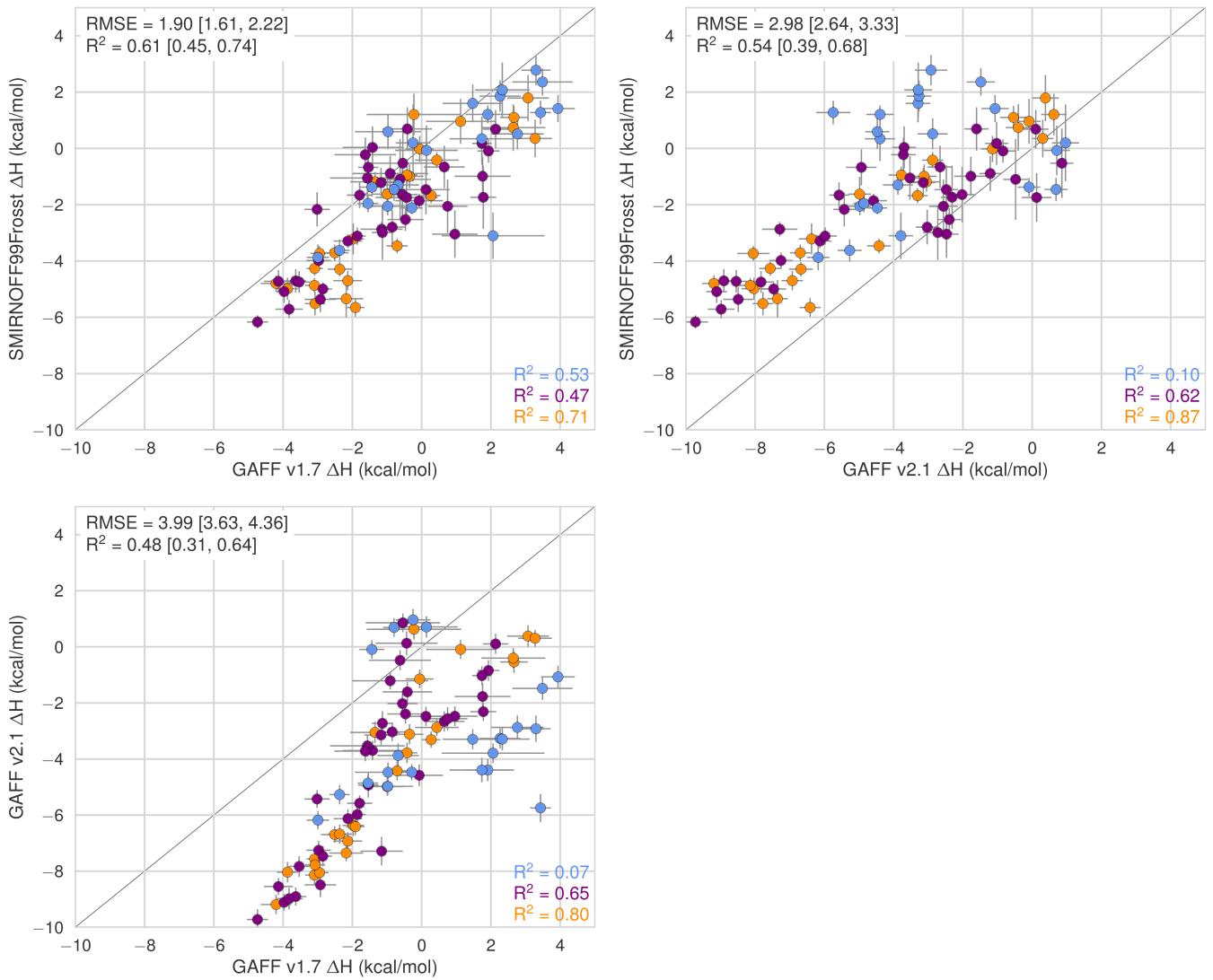


Figure 20: Comparison of calculated absolute binding free enthalpies (ΔH) between force field combinations. The orange, blue, and purple coloring distinguish the functional group of the guest as an ammonium, alcohol, or carboxylate, respectively.

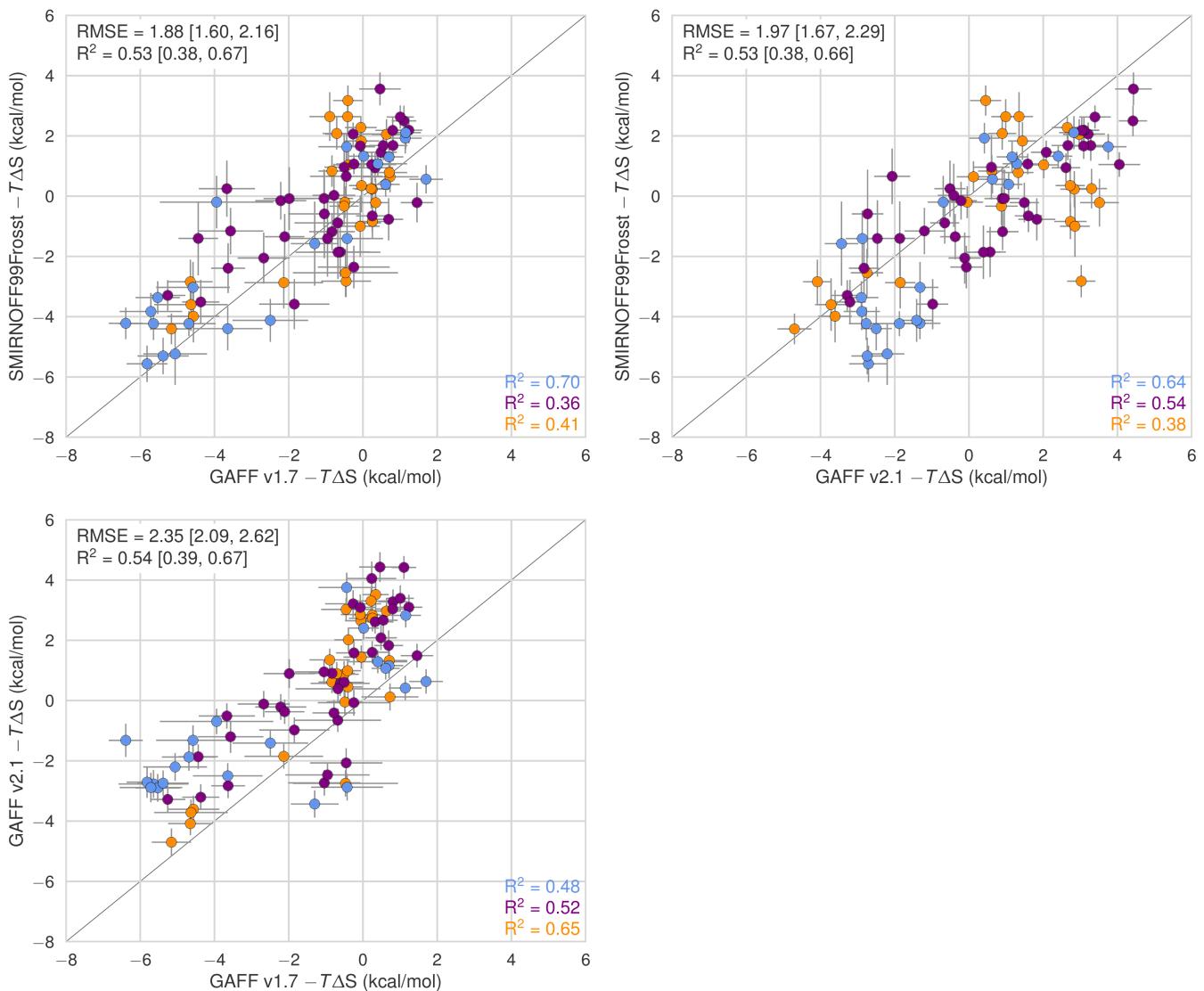


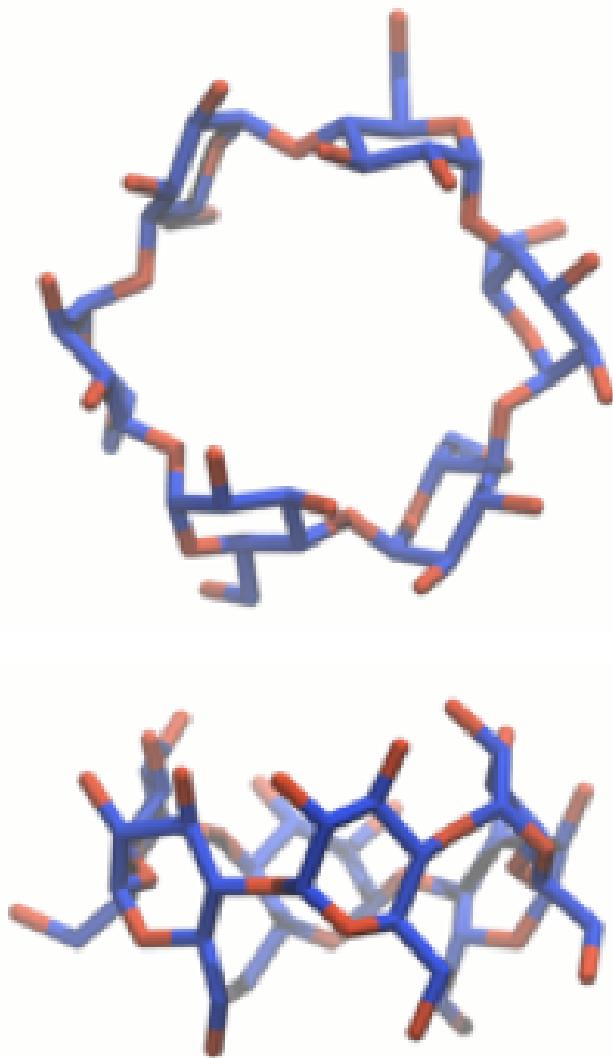
Figure 21: Comparison of calculated absolute binding free entropies ($-T\Delta S$) between force field combinations. The orange, blue, and purple coloring distinguish the functional group of the guest as an ammonium, alcohol, or carboxylate, respectively.

a b

c d

e f

Figure 22: Binding free energies (ΔG) replotted from Figure 3, with α CD points colored in blue and β CD points in grey (left) or α CD points in grey with β CD points colored in green (right).



SMIRNOFF99Frosst

Figure 23: A minimized structure of α CD in the SMIRNOFF99Frosst force field.

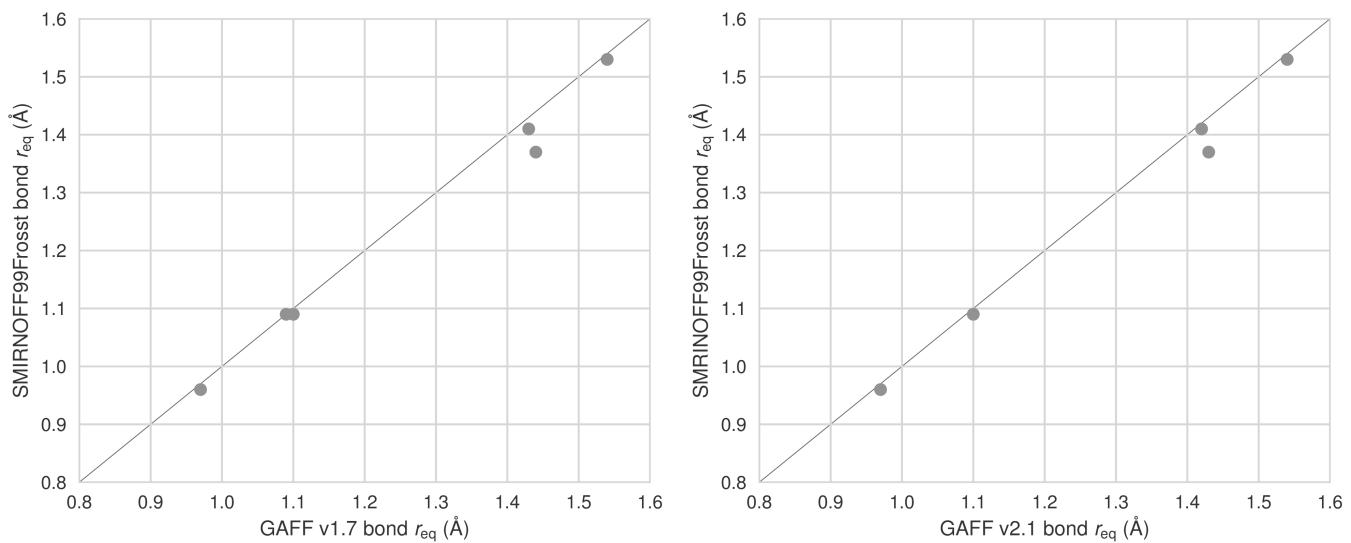


Figure 24: A comparison of bond equilibrium lengths for SMIRNOFF99Frosst, GAFF v1.7, and GAFF v2.1. Atom names refer to Figure 2.

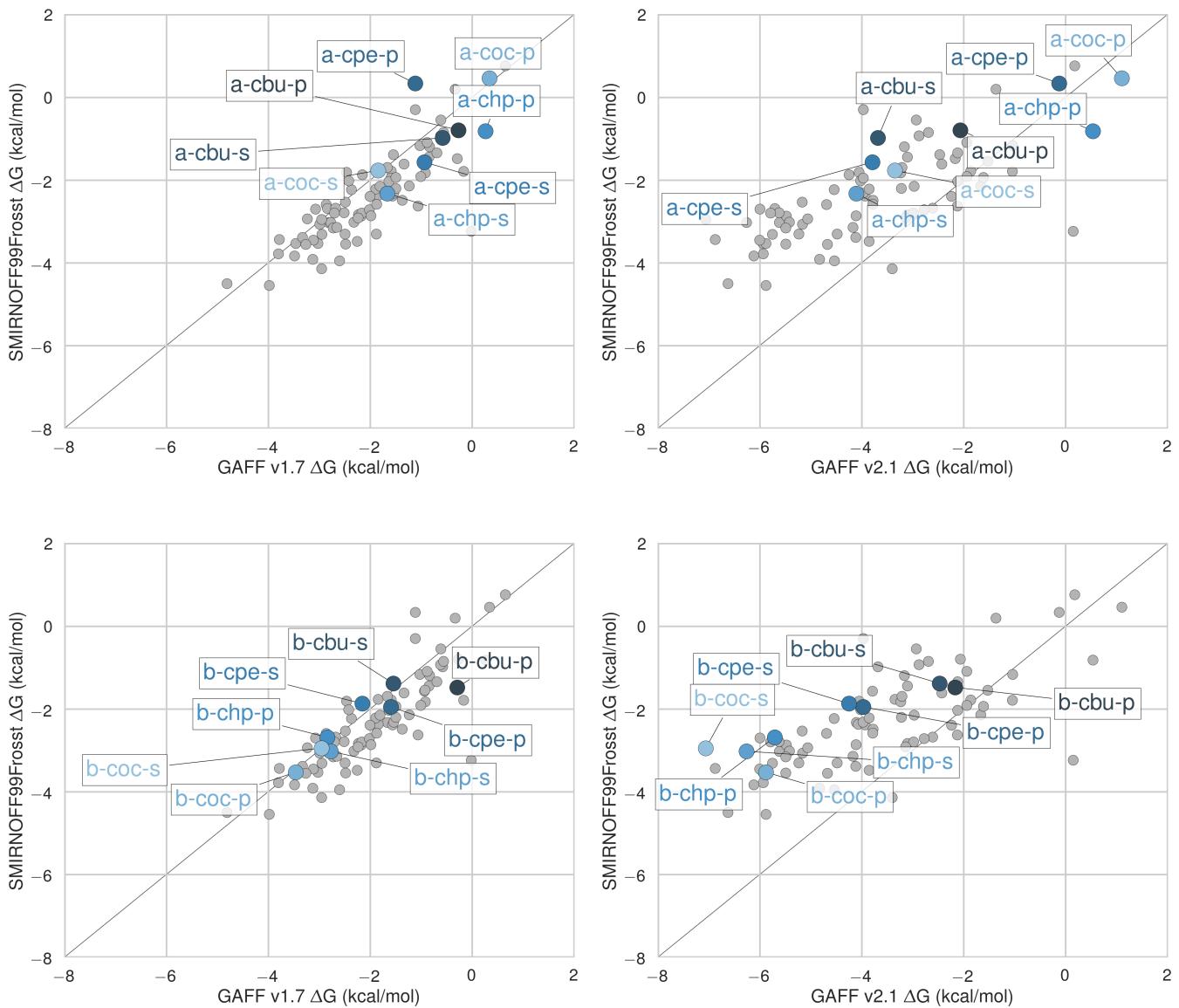
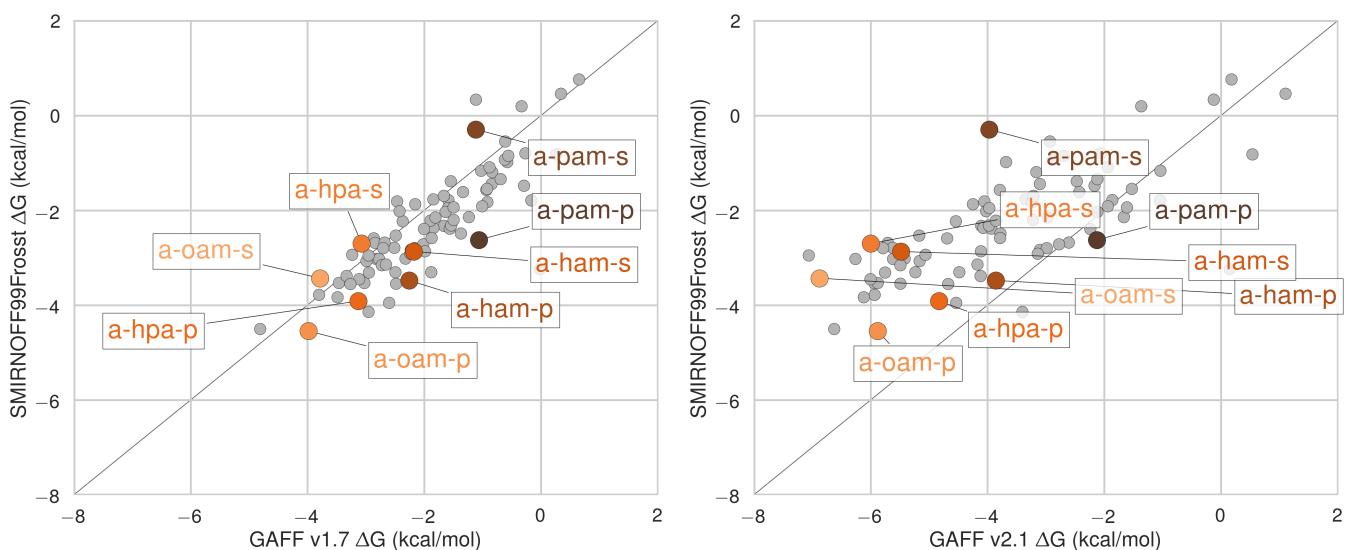


Figure 25: Binding free energy (ΔG) comparisons showing alcohols guests in color and highlighted.



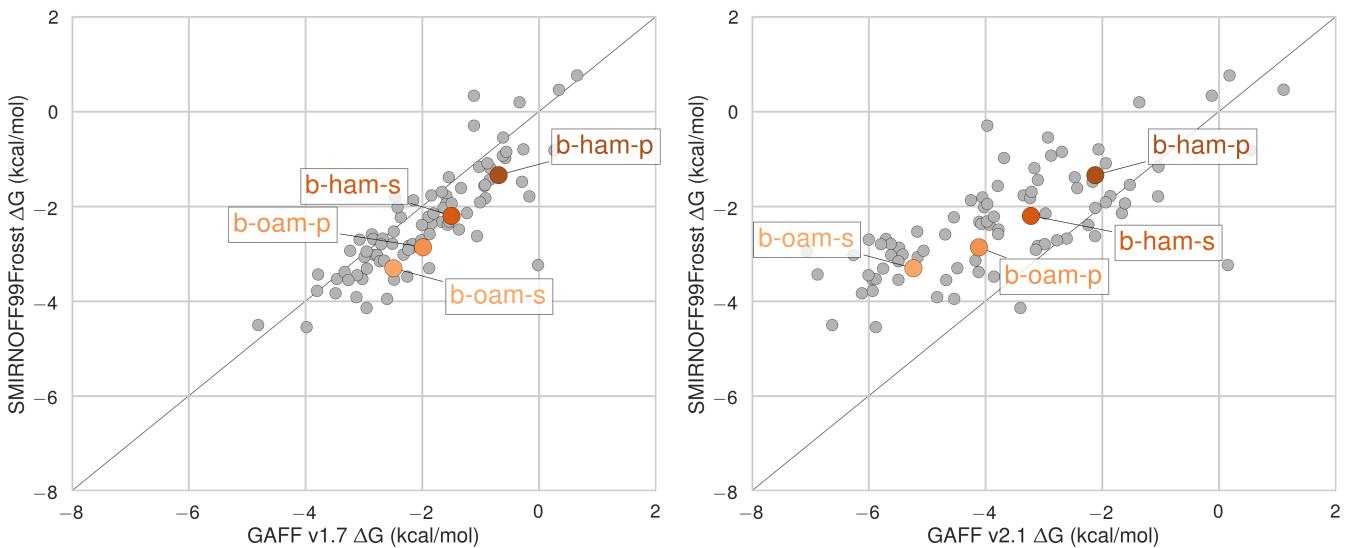
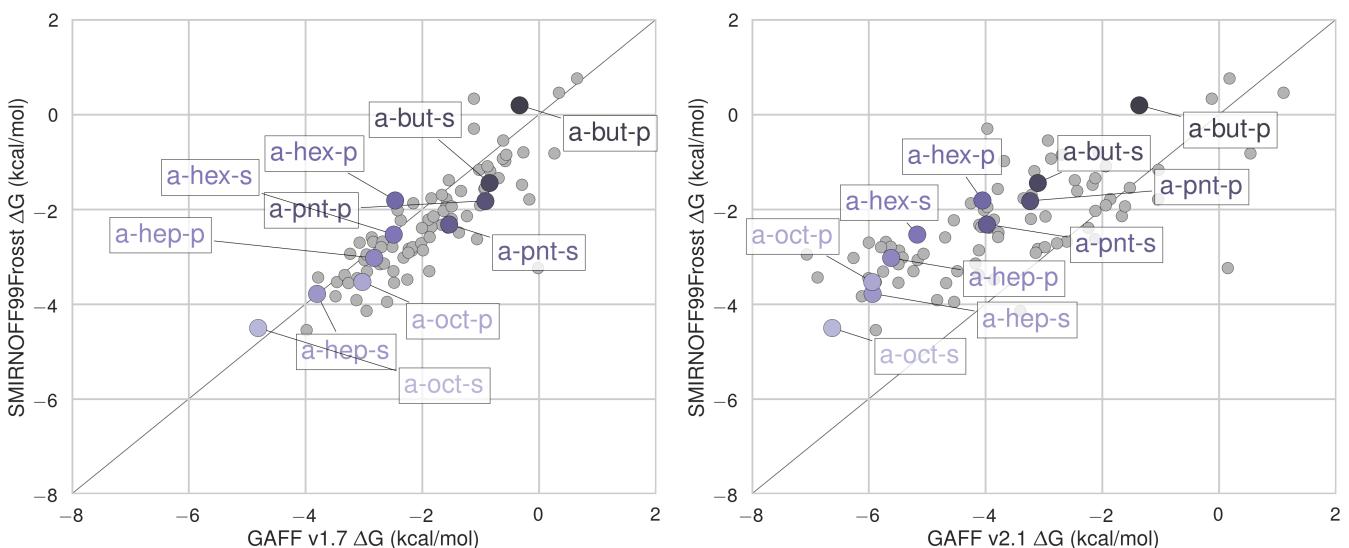


Figure 26: Binding free energy (ΔG) comparisons showing ammonium guests in color and highlighted.



The figure consists of two side-by-side scatter plots. Both plots have a y-axis labeled "SMIRNOFF99Frosst ΔG (kcal/mol)" ranging from -8 to 2 and an x-axis labeled "GAFF v1.7 ΔG (kcal/mol)" or "GAFF v2.1 ΔG (kcal/mol)" ranging from -8 to 2. A diagonal line $y = x$ is drawn in both plots. Data points are represented by grey circles. Several crystal structures are highlighted with blue boxes and labels: "b-hex-p", "b-hep-s", "b-hex-s", "b-hep-p", "b-pnt-p", and "b-pnt-s". In the left plot (v1.7), most points are clustered between -4 and 1 kcal/mol on both axes. In the right plot (v2.1), points are more widely distributed, with some reaching up to 1.5 kcal/mol on the y-axis.

Figure 27: Binding free energy (ΔG) comparisons showing carboxylates guests in color and highlighted.

References

- 1. Thermodynamic and Nuclear Magnetic Resonance Study of the Reactions of α - and β -Cyclodextrin with Acids, Aliphatic Amines, and Cyclic Alcohols**
Mikhail V. Rekharsky, Martin P. Mayhew, Robert N. Goldberg, Philip D. Ross, Yuko Yamashoji, Yoshihisa Inoue
The Journal of Physical Chemistry B (1997-01) <https://doi.org/cmwfn>
DOI: [10.1021/jp962715n](https://doi.org/10.1021/jp962715n)
- 2. Evaluating Force Field Performance in Thermodynamic Calculations of Cyclodextrin Host–Guest Binding: Water Models, Partial Charges, and Host Force Field Parameters**
Niel M. Henriksen, Michael K. Gilson
Journal of Chemical Theory and Computation (2017-07-11) <https://doi.org/gd2z2t>
DOI: [10.1021/acs.jctc.7b00359](https://doi.org/10.1021/acs.jctc.7b00359) · PMID: [28696692](https://pubmed.ncbi.nlm.nih.gov/28696692/) · PMCID: [PMC5606194](https://pubmed.ncbi.nlm.nih.gov/PMC5606194/)
- 3. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation**
Araz Jakalian, David B. Jack, Christopher I. Bayly
Journal of Computational Chemistry (2002-10-18) <https://doi.org/cktk6g>
DOI: [10.1002/jcc.10128](https://doi.org/10.1002/jcc.10128) · PMID: [12395429](https://pubmed.ncbi.nlm.nih.gov/12395429/)
- 4. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method**
Araz Jakalian, Bruce L. Bush, David B. Jack, Christopher I. Bayly
Journal of Computational Chemistry (2000-01-30) <https://doi.org/cvvpkv>
DOI: [10.1002/\(sici\)1096-987x\(20000130\)21:2<132::aid-jcc5>3.0.co;2-p](https://doi.org/10.1002/(sici)1096-987x(20000130)21:2<132::aid-jcc5>3.0.co;2-p)
- 5. Comparison of simple potential functions for simulating liquid water**
William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, Michael L. Klein
The Journal of Chemical Physics (1983-07-15) <https://doi.org/dg9sq8>
DOI: [10.1063/1.445869](https://doi.org/10.1063/1.445869)
- 6. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations**
In Suk Joung, Thomas E. Cheatham III
The Journal of Physical Chemistry B (2008-07) <https://doi.org/cgwnj7>
DOI: [10.1021/jp8001614](https://doi.org/10.1021/jp8001614) · PMID: [18593145](https://pubmed.ncbi.nlm.nih.gov/18593145/) · PMCID: [PMC2652252](https://pubmed.ncbi.nlm.nih.gov/PMC2652252/)
- 7. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset**
Michael R. Shirts, Christoph Klein, Jason M. Swails, Jian Yin, Michael K. Gilson, David L. Mobley, David A. Case, Ellen D. Zhong
Journal of Computer-Aided Molecular Design (2016-10-27) <https://doi.org/f9m3wn>
DOI: [10.1007/s10822-016-9977-1](https://doi.org/10.1007/s10822-016-9977-1) · PMID: [27787702](https://pubmed.ncbi.nlm.nih.gov/27787702/) · PMCID: [PMC5581938](https://pubmed.ncbi.nlm.nih.gov/PMC5581938/)
- 8. Computational Calorimetry: High-Precision Calculation of Host–Guest Binding Thermodynamics**
Niel M. Henriksen, Andrew T. Fenley, Michael K. Gilson
Journal of Chemical Theory and Computation (2015-08-26) <https://doi.org/f7q3mj>
DOI: [10.1021/acs.jctc.5b00405](https://doi.org/10.1021/acs.jctc.5b00405) · PMID: [26523125](https://pubmed.ncbi.nlm.nih.gov/26523125/) · PMCID: [PMC4614838](https://pubmed.ncbi.nlm.nih.gov/PMC4614838/)
- 9. Overcoming dissipation in the calculation of standard binding free energies by ligand extraction**
Camilo Velez-Vega, Michael K. Gilson
Journal of Computational Chemistry (2013-08) <https://doi.org/gbdv9w>
DOI: [10.1002/jcc.23398](https://doi.org/10.1002/jcc.23398) · PMID: [24038118](https://pubmed.ncbi.nlm.nih.gov/24038118/) · PMCID: [PMC3932244](https://pubmed.ncbi.nlm.nih.gov/PMC3932244/)

10. The SAMPL4 host–guest blind prediction challenge: an overview

Hari S. Muddana, Andrew T. Fenley, David L. Mobley, Michael K. Gilson

Journal of Computer-Aided Molecular Design (2014-03-06) <https://doi.org/f5585s>

DOI: [10.1007/s10822-014-9735-1](https://doi.org/10.1007/s10822-014-9735-1) · PMID: [24599514](https://pubmed.ncbi.nlm.nih.gov/24599514/) · PMCID: [PMC4053502](https://pubmed.ncbi.nlm.nih.gov/PMC4053502/)

11. Bridging Calorimetry and Simulation through Precise Calculations of Cucurbituril–Guest Binding Enthalpies

Andrew T. Fenley, Niel M. Henriksen, Hari S. Muddana, Michael K. Gilson

Journal of Chemical Theory and Computation (2014-08) <https://doi.org/f6hv4x>

DOI: [10.1021/ct5004109](https://doi.org/10.1021/ct5004109) · PMID: [25221445](https://pubmed.ncbi.nlm.nih.gov/25221445/) · PMCID: [PMC4159218](https://pubmed.ncbi.nlm.nih.gov/PMC4159218/)

12. The Amber Molecular Dynamics Package<http://ambermd.org/>

13. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning

Chad W. Hopkins, Scott Le Grand, Ross C. Walker, Adrian E. Roitberg

Journal of Chemical Theory and Computation (2015-03-30) <https://doi.org/f697mk>

DOI: [10.1021/ct5010406](https://doi.org/10.1021/ct5010406) · PMID: [26574392](https://pubmed.ncbi.nlm.nih.gov/26574392/)

14. Error estimates on averages of correlated data

H. Flyvbjerg, H. G. Petersen

The Journal of Chemical Physics (1989-07) <https://doi.org/bjpm5j>

DOI: [10.1063/1.457480](https://doi.org/10.1063/1.457480)

15. Development and testing of a general amber force field

Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, David A. Case

Journal of Computational Chemistry (2004) <https://doi.org/cdmcnb>

DOI: [10.1002/jcc.20035](https://doi.org/10.1002/jcc.20035) · PMID: [15116359](https://pubmed.ncbi.nlm.nih.gov/15116359/)