1 Supporting information

1.1 Setting up the systems and MD simulations

We employed QM calculations to aid the parametrization of neutral G3 and G5 guests. QM calculations were performed using Gaussian 09 [66] along with the M06-2X density functional[67, 68], the SMD implicit solvent model[69] and the G-31+G(d) basis set.[70] The resulting parameters are shown in Figure 1.

Fig. 1 Results from the QM calculation indicate that the distance between the hydrogen atoms from the hydroxide ion and the nitrogen atom in the tetramethylammonium group should be d = 2.962 Å, and that the O-N-C angle (between the oxygen atom in the hydroxide ion, N atom in the tetramethylammonium group and the connecting C atom to the rest of the structure) should be $\phi = 180°$.

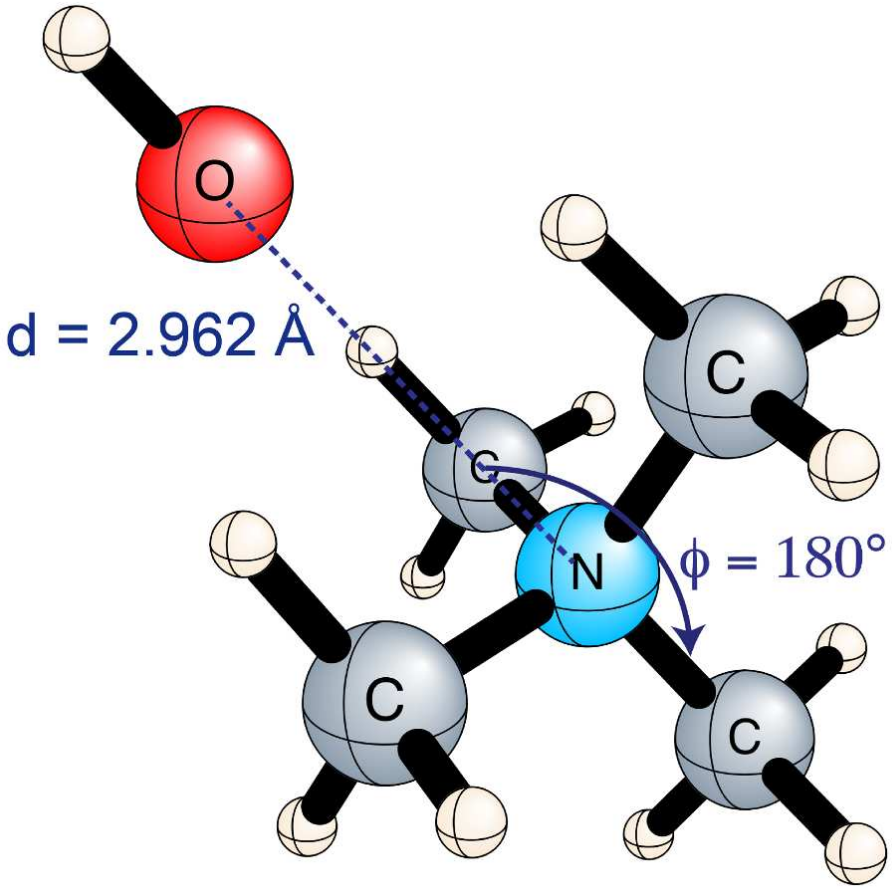


Table 1 The total number of atoms, and the number of Na ions and Cl ions used in each system, respectively. Systems containing G1 to G5, and G6 used 25 mM Nacl and 165 mM NaCl, corresponding to the ionic strength at pH 11.5 of 10 mM and 50 mM Na_3PO_4, respectively.

	OAH						OAMe					
		1		2		neutral		1		2		neutral
G1	11878	9/-	11894	11/2	11884	8/-	11899	9/-	11895	11/2	11902	8/-
G2	11884	9/-	11891	11/2	11884	8/-	11890	9/-	11882	11/2	11911	8/-
G3	11920	7/-	11882	9/2	11887	8/-	11908	7/-	11900	9/2	11893	8/-
G4	11900	9/-	11886	11/2	11897	8/-	11906	9/-	11898	11/2	11918	8/-
G5	11903	7/-	11898	9/2	11897	8/-	11894	7/-	11886	9/2	11903	8/-
G6	11873	9/-	11828	21/12	11875	12/4	11906	9/-	11898	21/12	11894	8/-

Table 2 Molecular dynamics simulations. All simulation lengths are in ns. '1' and '2' stand for neutralized systems and for systems with ionic concentrations corresponding to experimental ionic strength, respectively; "neutral" stands for systems with neutral guests.

	OAH			OAMe		
	1	2	neutral	1	2	neutral
G1	1000	1000	700	1000	400	400
G2	1000	1000	800	900	500	500
G3	2000	2500	2000	2000	2000	2000
G4	2500	2500	2000	2000	2000	2000
G5	2000	1700	1500	1900	1500	1500
G6	2000	1500	1000	1400	1500	1200
	10500	10200	8000	9200	7900	7600

1.2 GalaxyDock-HG

We randomly generated 50 initial binding poses, which were then optimized using CSA with the AutoDock4 scoring function. The latter comprises the van der Waals energy, directional hydrogen bond energy, Coulomb electrostatic energy, and desolvation free energy as follows:

$$E_{AutoDock} = w_{vdW}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + w_{hbond}\sum_{i,j}h(t_{ij})\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + w_{qq}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}} + w_{desolv}\sum_{i,j}\left(S_i V_j + S_j V_i\right)exp\left(-\frac{r_{ij}}{2\sigma^2}\right), \quad (1)$$

where:
- A_{ij}, B_{ij} and C_{ij}, D_{ij} are parameters for the van der Waals energy and for the hydrogen bond energy, respectively
- $h(t_{ij})$ is the weight factor to describe hydrogen bond directionality,
- q_i, q_j are partial charges,
- $\varepsilon(r_{ij})$ is a distance-dependent dielectric constant,
- S, V, σ are desolvation energy parameters.

Partial charges were taken from the CGENFF-generated parameters. Figure 2 describes the GalaxyDock-HG protocol for choosing binding poses.

Fig. 2 Protocol for GalaxyDock-HG

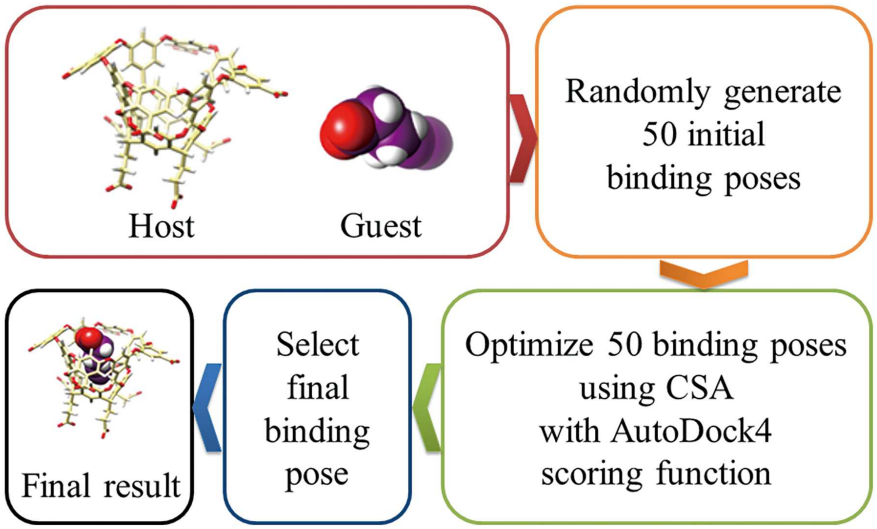


Table 3 Energy values for host-guest systems containing charged guests, provided by GalaxyDock-HG and the averaged experimental values. All values are in kcal/mol.

Guest	OAH Energy	OAH ΔG^{exp}_{bind}	OAMe Energy	OAMe ΔG^{exp}_{bind}
G1	-2.39	-5.22	-2.31	-5.36
G2	-3.59	-4.49	-3.49	-5.15
G3	-1.47	-4.78	-1.61	-5.85
G4	-5.95	-9.38	-3.87	-2.38
G5	-2.43	-4.12	-2.26	-3.91
G6	-3.88	-5.12	-3.45	-4.48

1.3 Free energy calculations

The free energy spent to turn the restraints off between the guest and the host was computed analytically with the following formula:

$$\Delta G^{complex}_{restr\ off} = -kTln\left[\frac{8\pi^2 V\left(K_r K_{\theta_A} K_{\theta_B} K_{\phi_A} K_{\phi_B} K_{\phi_C}\right)^{1/2}}{r^2 \sin\theta_A \sin\theta_B \left(2\pi kT\right)^3}\right], \quad (2)$$

where:
- k = Boltzmann constant,
- T = simulation temperature,
- V = volume of the simulation box,
- K_r = force constant of distance restraint,
- K_{θ_A}, K_{θ_A} = force constants of angle restraints,
- K_{ϕ_A},K_{ϕ_B},K_{ϕ_C} = force constants of dihedral restraints,
- r = distance between selected atom in guest and selected atom in protein,
- θ_A, θ_B = selected angles.

Table 4 TI simulations: λ points for each free energy component: restraints ($\Delta G^{complex}_{restr\ on}$), electrostatics ($\Delta G_{charge}$) and van der Waals interactions (ΔG_{vdW})

restraints	charge	vdW
0.0005	0.0	0.0
0.0020	0.1	0.1
0.0020	0.2	0.2
0.0040	0.3	0.3
0.00625	0.4	0.4
0.0100	0.5	0.5
0.01875	0.6	0.6
0.03	0.675	0.675
0.05	0.725	0.725
0.075	0.775	0.775
0.10	0.825	0.825
0.20	0.875	0.875
0.30	0.925	0.925
0.40	0.95	0.95
0.50	0.975	0.975
0.60	1.0	1.0
0.70	-	-
0.70	-	-
0.80	-	-
0.90	-	-
1.00	-	-

Table 5 HREM-BAR simulations: λ points for the electrostatics (ΔG_{charge}) and van der Waals (ΔG_{vdW}) free energy component

charge	vdW
1.0	1.00
0.916	0.995
0.833	0.985
0.75	0.960
0.666	0.930
0.583	0.900
0.5	0.850
0.416	0.800
0.333	0.750
0.25	0.700
0.166	0.650
0.083	0.600
-	0.550
-	0.500
-	0.450
-	0.390
-	0.330
-	0.270
-	0.220
-	0.180
-	0.00

1.4 QM Calculations

To properly understand the effect of pH = 11.5 upon the electrostatics of the guest molecules, we performed QM calculations to understand the anionic guests' proton affinity, and the cationic guests' affinity for hydroxide. For the anionic guests,

Table 6 Results from TI free energy calculations for each system, in kcal/mol.

	OAH			OAMe		
	1	2	neutral	1	2	neutral
G1	-0.9±1.3 ^{*,m}	-4.5±0.6 ^{*,m,d}	-7.6±0.7 ^{*,m}	-3.1±0.9 ^{d}	-4.2±0.8 ^{*,d}	-6.1±0.9 ^{*,m}
G2	2.1±1.8 ^{*,m}	-2.7±0.7 ^{*,d}	4.7±0.7 ^{m}	*-6.4±1.4* ^{m}	-7.1±0.6 ^{*,m}	5.8±0.7 ^{*,m}
G3	-4.8±1.1 ^{*,m}	-0.1±0.2 ^{*,d}	-4.5±1.2 ^{*,m}	-1.1±1.6 ^{d}	-2.6±0.7 ^{*,m}	-6.5±1.5 ^{*,m}
G4	-7.8±1.3 ^{*,d}	-5.1±0.7 ^{*,d}	-8.1±1.5 ^{*,m}	-9.9±2.4 ^{*,m}	-9.3±2.0 ^{m}	-1.9±1.2 ^{*,m}
G5	-1.7±0.1 ^{*,d}	-1.6±0.9 ^{*,d}	-1.6±0.5 ^{*,m}	-2.2±1.0 ^{d}	-2.6±0.9 ^{*,m}	-2.6±0.6 ^{*,m}
G6	-4.9±1.5 ^{*,m}	-4.4±1.6 ^{d}	5.9±1.4 ^{*,m}	-5.9±0.7 ^{*,d}	-7.0±0.3 ^{*,m}	9.3±2.5 ^{*,m}

* The structure on which we performed FES resulted from MD simulations with NOE restraints.
^{d} The ligand was placed in the host through GalaxyDock-HG.
^{m} The ligand was placed in the host manually.

Table 7 Results from HREM-BAR free energy calculations for each system, in kcal/mol.

	OAH			OAMe		
	1	2	neutral	1	2	neutral
G1	-4.2±1.0 ^{*,m}	-4.5±0.2 ^{m}	-6.4±0.7 ^{m}	-4.2±0.4 ^{d}	-4.2±0.0 ^{*,m}	-
G2	-2.2±0.2 ^{*,m}	-3.2±0.2 ^{d}	-6.2±0.3 ^{*,m}	-5.3±0.3 ^{d}	-4.7±0.3 ^{*,m}	-
G3	*-2.9±0.7* ^{*,m}	*-2.2±0.4* ^{*,d}	-3.4±0.0 ^{*,m}	*-4.0±0.5* ^{*,d}	*-3.21±0.4* ^{*,m}	*-7.6±0.5* ^{*,m}
G4	-5.0±0.1 ^{d}	-4.4±0.5 ^{d}	*-8.0±1.8* ^{*,m}	-10.5±1.0 ^{*,m}	-6.6±0.3 ^{*,m}	-
G5	*0.2±0.2* ^{d}	*-0.7±0.1* ^{*,d}	*-2.7±0.3* ^{*,m}	*-1.6±0.7* ^{d}	-2.9±0.7 ^{*,m}	*-3.0±0.4* ^{*,m}
G6	*-3.9±0.2* ^{*,m}	-4.1±0.1 ^{d}	-4.2±0.0 ^{*,m}	*-4.7±0.2* ^{*,d}	*-4.5±0.5* ^{*,m}	-

* The structure on which we performed FES resulted from MD simulations with NOE restraints. Italicized values were computed after the end of the competition and therefore were not part of the submitted sets.
^{d} The ligand was placed in the host through GalaxyDock-HG.
^{m} The ligand was placed in the host manually.

Table 8 Results* from TI simulations for OAH-G1 when increasing equilibration and production times from 20 ps and 200 ps per λ point to 50 ps and 500 ps per λ point, and the average experimental binding energy. All values are in kcal/mol.

	1	2	ΔG^{exp}_{bind}
20 ps/200 ps	-0.9±1.3	-4.5±0.6	-5.22
50 ps/500 ps	-2.7±0.2	-5.3±1.0	-5.22

* The standard deviations were obtained by repeating each simulation three times.

Table 9 Results (in kcal/mol) from TI and HREM-BAR calculations for the OAH-G4 and OAMe-G4 complexes when G4 is reparametrized with a lonepair.

	OAH-G4	OAMe-G4
TI	-4.6±1.1	0.0±0.4
HREM-BAR	-3.8±0.7	1.5±0.4

these are pK_a calculation, whereas for the cationic guests, these are the analog of pK_a calculations for hydroxide association. We will refer to them as "pK_a calculations" for the remainder of this discussion.
In this work we used a relative pK_a scheme.[71, 72] In the relative scheme (eq. 3) the free energy of binding a proton (or hydroxide) is computed for the guest molecule and an analogous ligand in the bulk aqueous phase. The difference be-

tween these binding free energies is then added to the experimentally determined pK_a (Table 11, column 4) value, to arrive at the calculated pK_a value (Table 10, column 3). Finally, this is readily converted into a free energy value by eq. 4 (see values in Table 10, column 5).

All QM calculations in this work were performed using Gaussian 09 [66] at the M06-2X/6-31+G(d) level of theory and the G-31+G(d) basis set.[70]. We utilized a "vertical" solvation scheme, where aqueous phase free energies were approximated by first performing a geometry optimization in the gas phase, and then performing a single-point energy calculation on this geometry in the bulk aqueous phase. All QM optimizations performed with "Tight" wavefunction and geometry convergence criteria. The calculations utilized the "UltraFine" numerical quadrature as required by the M06-2X functional.

$$\mathrm{p}K_\mathrm{a}^\mathrm{rel}(\mathrm{AH}^+) = \mathrm{p}K_\mathrm{a}^\mathrm{exp}(\mathrm{LH}^+) + \left[\Delta G^*_\mathrm{aq}(\mathrm{AH}^+) - \Delta G^*_\mathrm{aq}(\mathrm{LH}^+)\right] / \left[\ln(10) RT\right] \quad (3)$$

$$\Delta G_\mathrm{aq} = \mathrm{p}K_\mathrm{a} RT \ln(10) \quad (4)$$

Table 10 Computed pK_a values for Octa-Acid Guests in bulk aqueous phase using M06-2X density functional and SMD implicit solvent. Population of charged species or neutral species and the corresponding free energy of changing the protonation state in solution are calculated at pH = 11.5 and at 25 ℃. In some cases experimental pK_a data was available for the guest molecule, here we present the experimental pK_a value. Free energy values are from 100% neutral reference population to the equilibrium population, and are in kcal/mol. '−' denotes charged species.

Guest	Analog	pK_a/ pK_b	Pop. (%)	ΔG_corr (kcal/mol)
OA-G1	**L1**	5.20	100.00 (−)	−8.5
OA-G2	expt.	3.55	100.00 (−)	−10.8
OA-G3	**TMAO**	5.53	100.00 (*o*)	3.4
OA-G4	**L2**	4.84	100.00 (−)	−9.0
OA-G5	**TMAO**	5.48	100.00 (*o*)	3.3
OA-G6	expt.	3.46	100.00 (−)	−10.9

Table 11 Systematic names and experimental pK_a or pK_b values for octa-acid guests and/or analogs. All pK_a and pK_b measurements were taken at 25 ℃.

Molecule	Analog	Systematic Name	pK_a/ pK_b	Ref.
L1	**G1**	butanoic acid	4.83	73
OA-G2	–	4-cyanobenzoic acid	3.55	73
TMAO	**G3**	tetramethyl ammonium hydroxide	4.2	74
L2	**G4**	adamantane-1-carboxylic acid	5.1	75
TMAO	**G5**	tetramethyl ammonium hydroxide	4.2	74
OA-G6	–	3-nitrobenzoic acid	3.46	73

Fig. 3 Chemical structure of the analog molecules used to compute pK_a values for the octa-acid guests. **L1** (left) and **L2** (right).

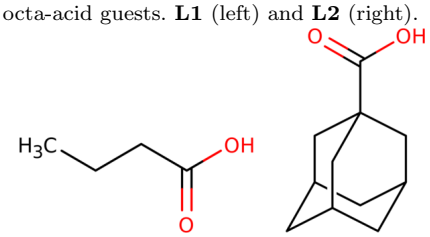
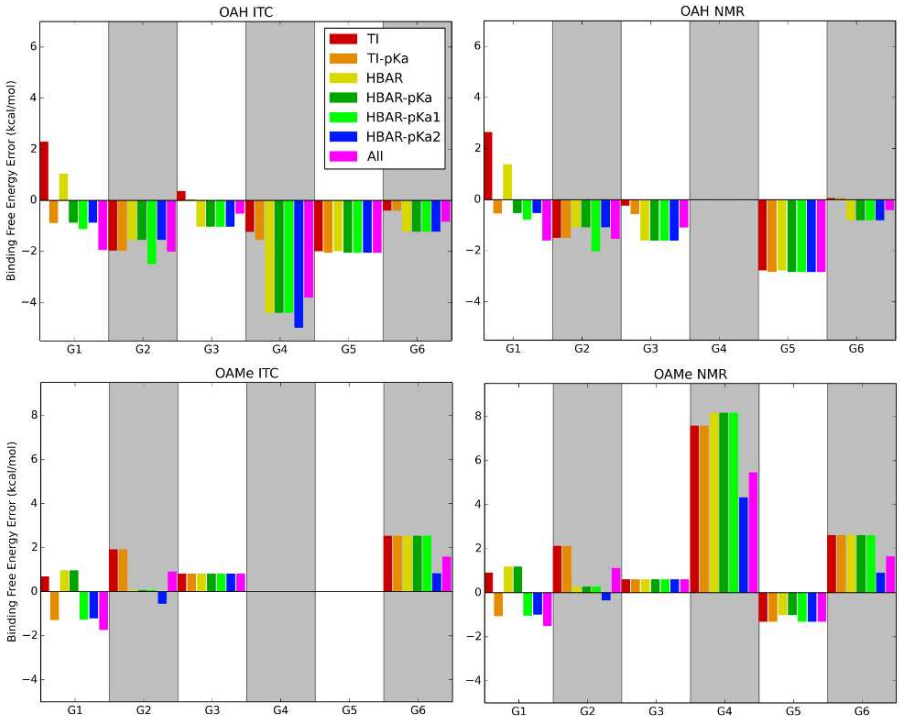


Fig. 4 Top row: $\Delta G^{exp}_{bind} - \Delta G^{comp}_{bind}$ for each OAH-guest complex. Bottom row: Difference between ITC and NMR values experimental and all submitted computed values for OAMe and each guest.



1.5 Analyzing the results from free energy calculations

1.6 Guest ranking

Guest ranking by binding free energy was not consistent among experimental (ITC, NMR) nor computational methods (TI, HREM-BAR, GalaxyDock-HG). But, as Table 12 shows, there was somewhat consensus as to what the strongest and weakest binders are. All computational methods indicated G4 as the strongest binder to both OAH and OAMe. ITC also indicated G4 as the strongest binder

Table 12 Ranking the guests by binding free energy by experimental methods, by the lowest-RMSD computational method and GalaxyDock-HG (GD-HG).

		Strongest	—	—	—	—>	Weakest
OAH	ITC	G4 >	G1 >	G6 >	G2 >	G3 >	G5
OAH	NMR	G3 >	G1 >	G6 >	G5 >	G2 >	—
OAH	TI	G4 >	G6 >	G1 >	G3 >	G2 >	G5
OAH	GD-HG	G4 >	G6 >	G2 >	G5 >	G1 >	G3
OAMe	ITC	G3 >	G1 >	G2 >	G6	—	—
OAMe	NMR	G3 >	G1 >	G2 >	G6 >	G5 >	G4
OAMe	HBAR-ps2	G4 >	G3 >	G6 >	G2 >	G1 >	G5
OAMe	GD-HG	G4 >	G2 >	G6 >	G5 >	G1 >	G3

to OAH. But for OAMe, G4 was ranked the weakest binder by NMR (no value available from ITC), and both experimental methods ranked G3 as the strongest binder. TI and HREM-BAR (and ITC for OAH) pointed towards G5 and G3 as the weakest binders. GalaxyDock-HG also ranked G3 as the weakest binder. This was probably due to the parametrization of neutral G3 and G5. GalaxyDock-HG had almost the same ranking for both OAH and OAMe systems. The only large fluctuation was between the reported energy for OAH-G4 and OAMe-G4, which increased by 2.08 kcal/mol, from -5.95 kcal/mol for OAH to -3.87 kcal/mol for OAMe (see Table 3). This indicates that even if the method is sensitive to the difference between the hosts, the offset was not enough to change guest ranking. The difference between OAH and OAMe for all other guests was between 0.08 kcal/mol and 0.43 kcal/mol.

Fig. 5 Top: ITC experimental data for OAH and OAMe systems versus computed $\Delta G_{\mathrm{bind}}^{\mathrm{TI-ps}}$. The TI-ps had the lowest RMSD values for OAH systems for the ITC full OAH experimental set. Bottom: NMR experimental data for OAH and OAMe systems versus computed $\Delta G_{\mathrm{bind}}^{\mathrm{HBAR-ps2}}$. The HBAR-ps2 had the lowest RMSD values for OAMe systems for the NMR full OAMe experimental set. Values for OAH are represented bu green symbols, and for OAMe by blue symbols.