A combined hydration site analysis and implicit solvation alchemical approach for the modeling of host-guest binding

Emilio Gallicchio, Tom Kurtzman, Lauren Wickstrom, Ronald M Levy

Rajat K Pal, Kamran Haider, Junchao Xia, William Flynn

egallicchio@brooklyn.cuny.edu

http://compmolbiophysbc.org

History: 2013, SAMPL4, Octa-Acid

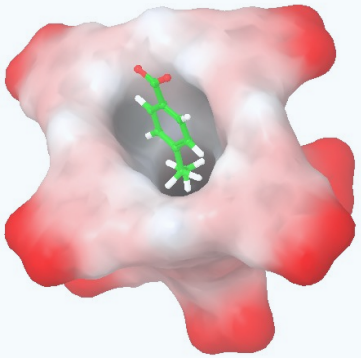
Spring 2013 CHEM525 Statistical Thermodynamics

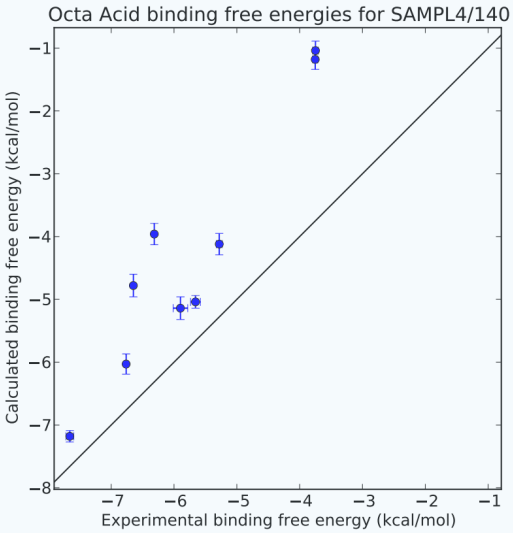
Instructors: Ronald Levy, Emilio Gallicchio

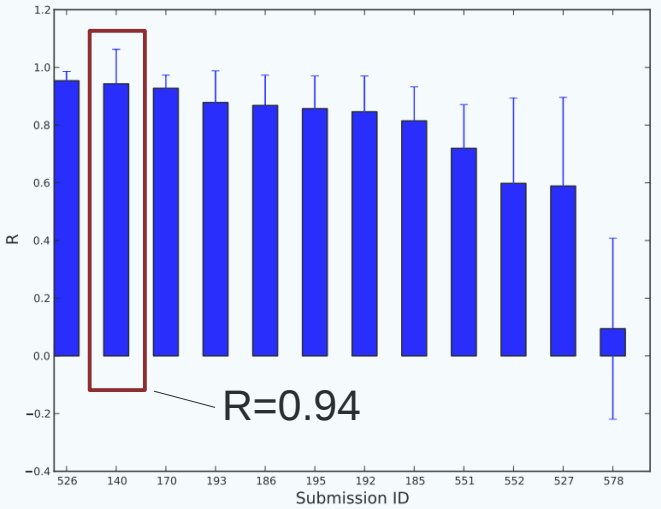
- He Chen
- Michael Fitzgerald
- Yang Gao
- Peng He, Peng
- Malathi Kalyanikar
- Chuan Kao
- Beidi Lu
- Yijie Niu
- Manasi Pethe
- Jie Zhu
- Haoyuan Chen

Students conducted all aspects of the experiment collaboratively, from building the molecules to the BEDAM calculations.

Very good results!







- BEDAM + AGBNP2 with empirical enclosed hydration corrections trained on available experimental data (carboxylic acid guests)

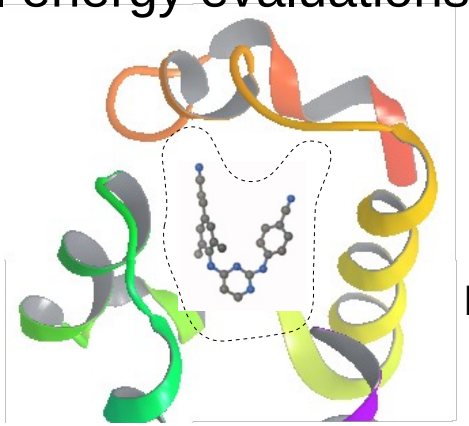
Free Energy Model: Single-Decoupling with Implicit Solvation

- With implicit solvation (AGBNP2) the adiabatic free energy of transfer is only two single-point potential energy evaluations

Hybrid potential:

$$U_\lambda(x) = U_0(x) + \lambda u(x)$$

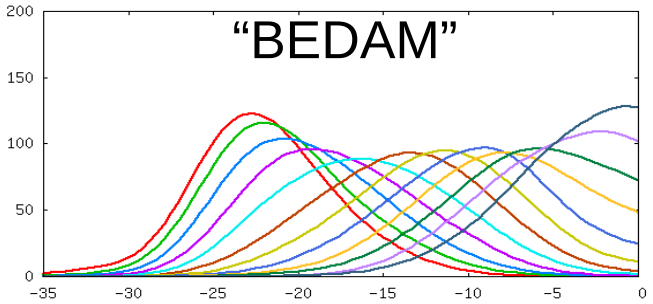
λ=0: uncoupled state
λ=1: coupled state



Binding energy $= u$

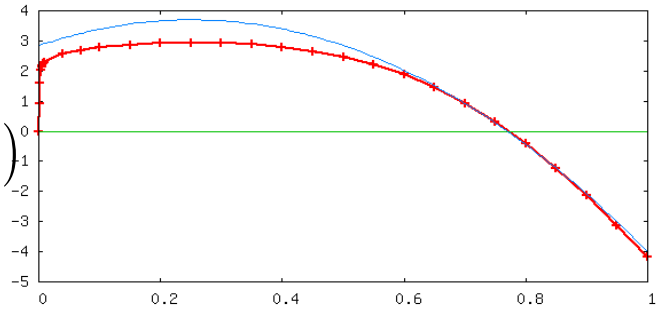
Ensembles of conformations of the complex as a function of λ are generated

$p_\lambda(u)$



u

$\Delta G_b^\circ(\lambda)$



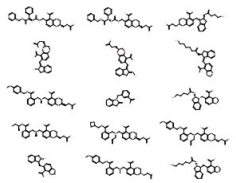
λ

The method is suitable for absolute binding free energies of relatively large ligands.

Gallicchio, Lapelosa, Levy, *JCTC* **(2010)** · Gallicchio & Levy, *Curr. Op. Struct. Biol.* **(2011)** · Gallicchio & Levy, *Adv. Prot. Chem.* **(2011)** · Lapelosa, Gallicchio, Levy, *JCTC* **(2012)** · Gallicchio, Levy *J. Comp. Aid. Mol. Design* **(2012)** · Tan, Gallicchio, Lapelosa, Levy *JCP* **(2012)** · Gallicchio, *Mol. Biosc* **(2012)** · Wickstrom, He, Gallicchio, Levy *JCTC* **(2013)** · Gallicchio et. al JCAMD **(2014)** · Gallicchio et al. *JCAMD* **(2015)** · Deng, Gallicchio et al. *JPCB* **(2015)** · Wickstrom et al. **(2015)**

Mostly Used in Free Energy-Based Virtual Screening

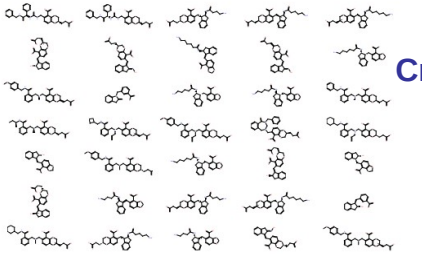
Ligand Library



Protonation/ tautomerization expansion

`LigPrep/Epik`

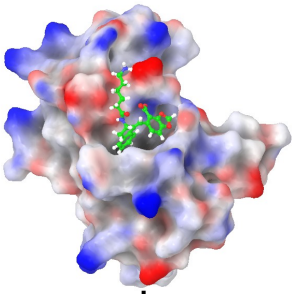
Expanded Library



Crystal Structures Analysis + Ensemble Docking

`AutoDock/Vina`
`Glide`

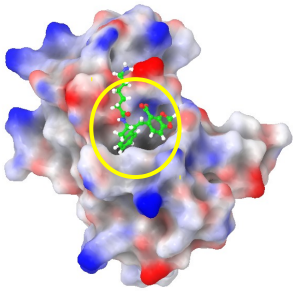
Docked Complexes



Filtering/Prioritization Structural Predictions

BEDAM Setup T-RE Reservoirs

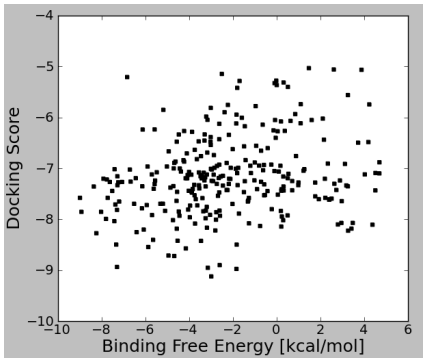
Prepped Complexes



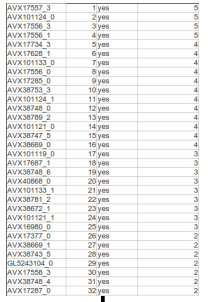
BEDAM parallel H-RE Calculations

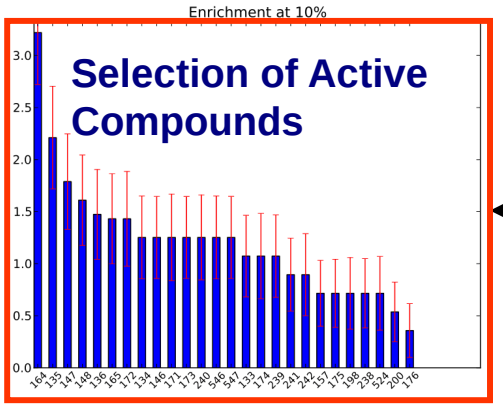
`IMPACT/OPLS/AGBNP2`

Binding Free Energy Estimates



Ranking

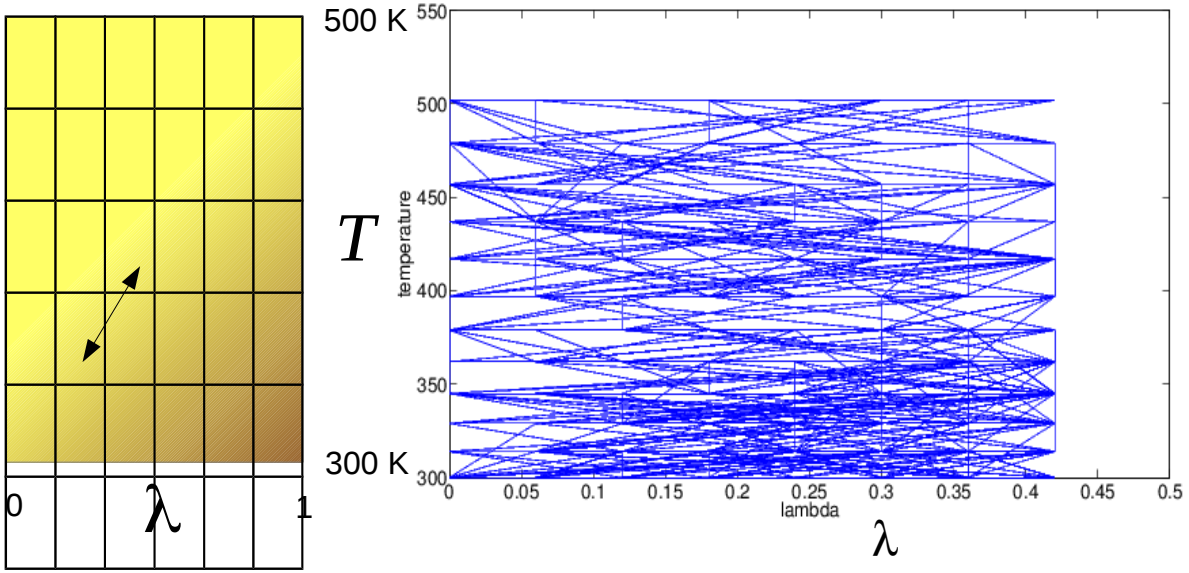


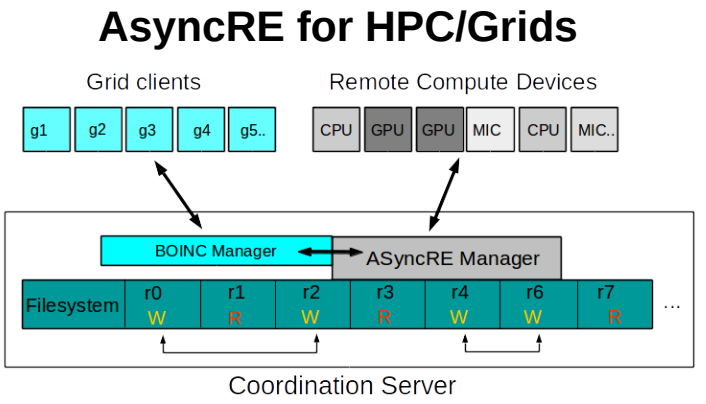


High performance, automated setup & unsupervised computation key to handling large datas

Multidimensional Replica Exchange

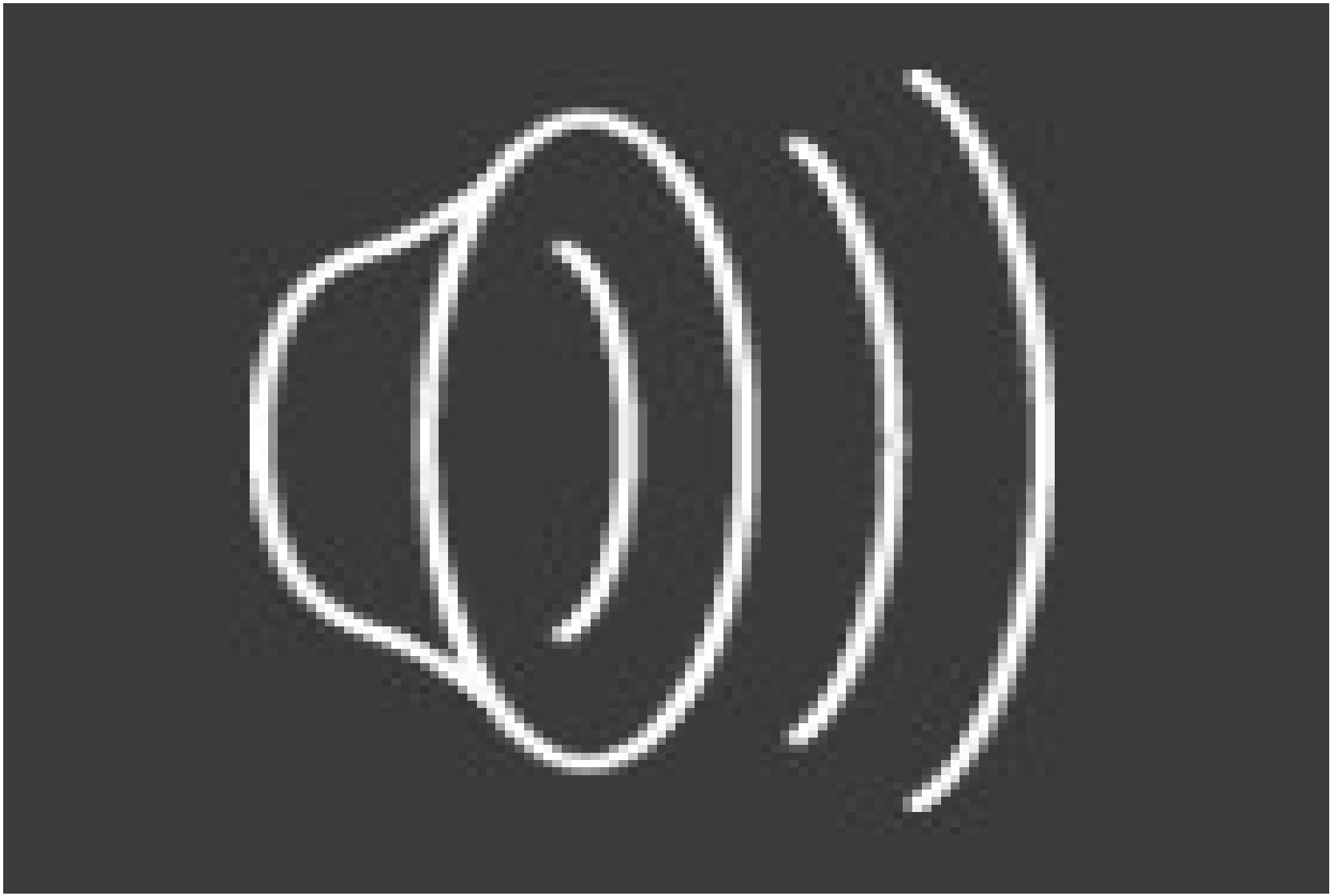
Alchemical and Temperature Spaces





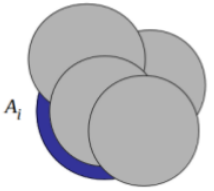
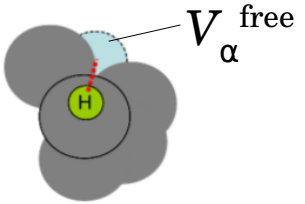
- Implicit solvation allows for RE in both alchemical and temperature dimensions
- λ-exchanges activate inter-molecular degrees of freedom
- T-exchanges activate intra-molecular degrees of freedom
- Data collected at all temperatures is reweighted at room temperature.

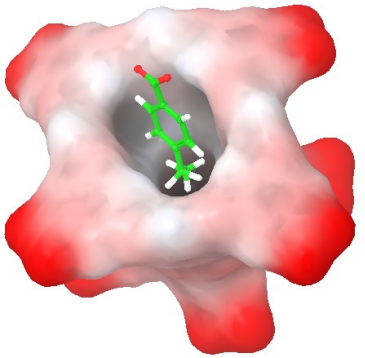
Trying to be Exhaustive ...



Potential Energy Model

OPLS/AGBNP2

	Electrostatic Model	Non-Polar Model	First-Shell Model
	Volume Scaling Factors	Surface Areas	Hydration Spheres Weights
1	$s_i = \frac{V_i^{\text{self}}}{V_i} = 1 - \frac{1}{2}\sum_j \frac{V_{ij}}{V_i} + \frac{1}{3}\sum_{j<k} \frac{V_{ijk}}{V_i} + \dots$	$A_i = \frac{\partial V}{\partial R_i}$	$w_\alpha = \frac{V_\alpha^{\text{free}}}{V_\alpha} = 1 - \sum_i \frac{V_{\alpha i}}{V_\alpha} + \sum_{i<j} \frac{V_{\alpha ij}}{V_\alpha} - \dots$
	Born Radii		
2	$\frac{1}{B_i} = \frac{1}{R_i} - \frac{1}{4\pi}\sum_{j\neq i} s_{ji} Q_{ji}$		
	Generalized Born Energy	Non-Polar Energy	Hydration Spheres Energy
3	$\Delta G_{\text{el}} = -\frac{1}{2} u_\epsilon \left[\sum_i \frac{q_i^2}{B_i} + 2\sum_{i<j} \frac{q_i q_j}{f_{ij}} \right]$	$\Delta G_{\text{np}} = \sum_i [\gamma_i A_i + \alpha_i W(B_i)]$	$\Delta G_{\text{HS}} = \sum_\alpha h_\alpha S(w_\alpha)$



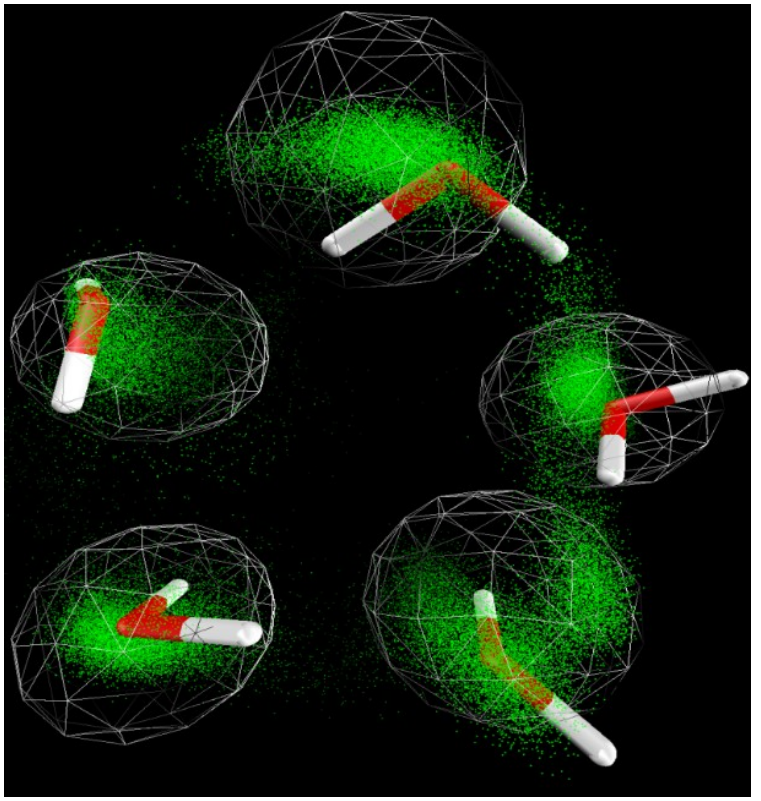
Empirically adjusted to treat enclosed hydration effects for SAMPL4

A Better Way: Hydration Site Analysis

Tom Kurtzman, Kamran Haider

Explicit Solvent MD Simulation (Restrained solute)

Clustering



- A water density distribution is generated in MD simulations.
- A clustering algorithm identifies high density hydration sites.
- For each hydration site, solvation energy and entropy are calculated using inhomogeneous solvation theory.

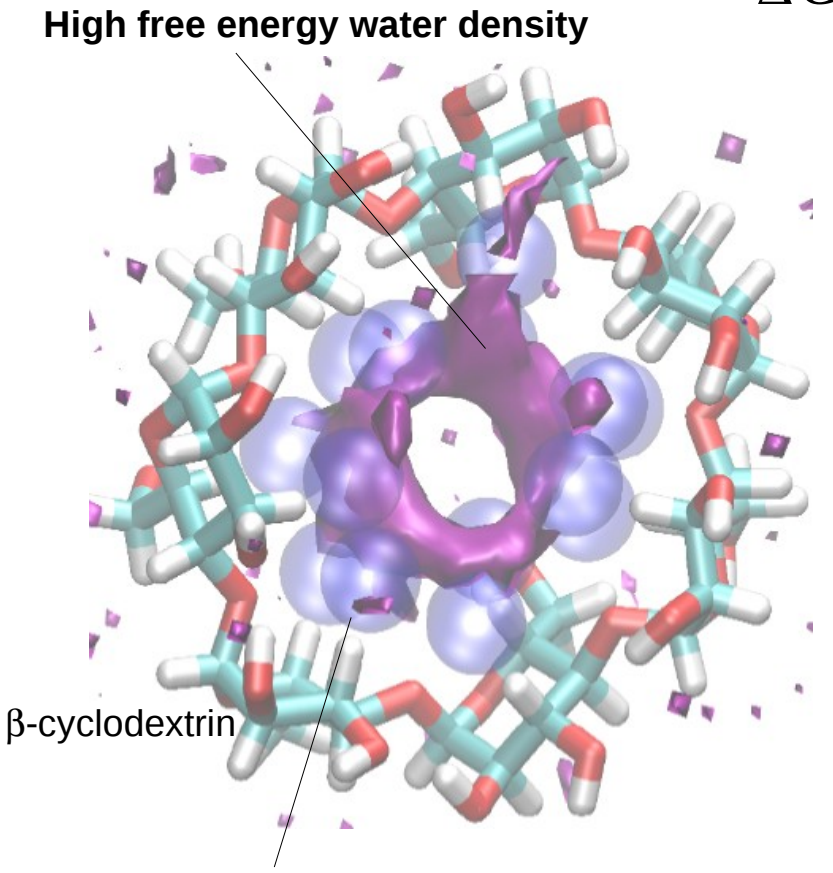
Young et al., 2007; Lazaridis, T. 1998

Training AGBNP2 on HSA Data?

Lauren Wickstrom

$$\Delta G_{\mathrm{HS}} \sim \sum_w h_w V_w(\mathrm{free})$$

$h_w > 0$ aids binding (water expulsion)

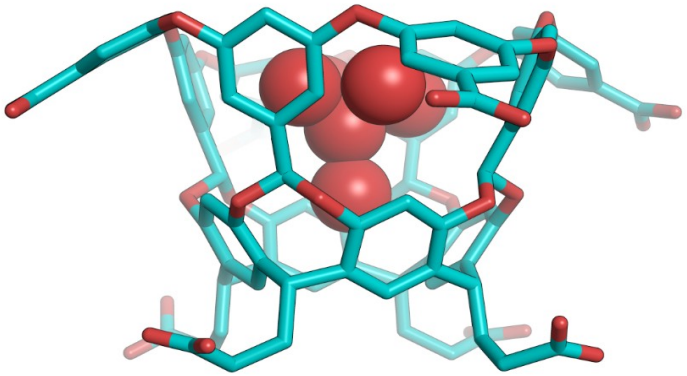


- Enclosed hydration effects can be incorporated into implicit solvation model
- Location and strength of hydration sites
- Combine accuracy of explicit solvation with versatility of implicit solvation

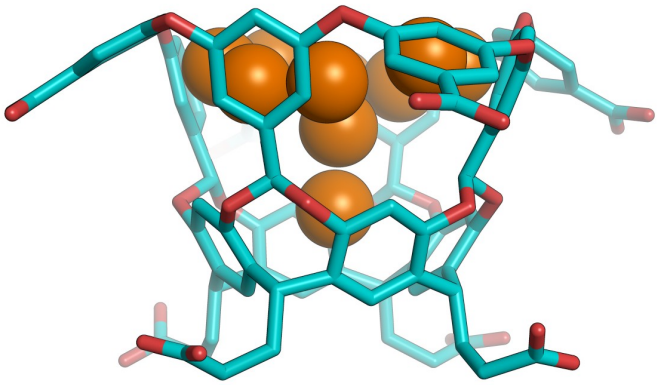
Lauren Wickstrom, Nanjie Deng, Peng He, Ahmet Mentes ,Crystal Nguyen, Michael K. Gilson, Tom Kurtzman, Emilio Gallicchio, and Ronald M. Levy. Parameterization of an effective potential for protein-ligand binding from host-guest affinity data. *J. Molecular Recognition* (2015).

Locating High Free Energy Sites

Octa-acid



Hydration Sites

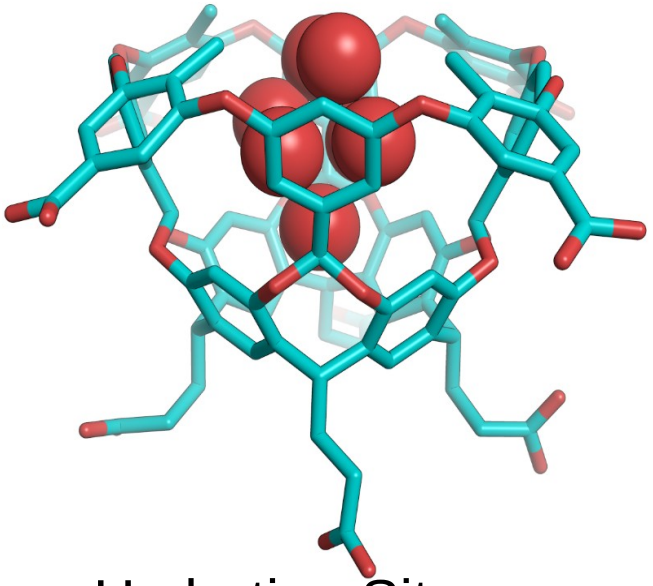


AGBNP2 Spheres

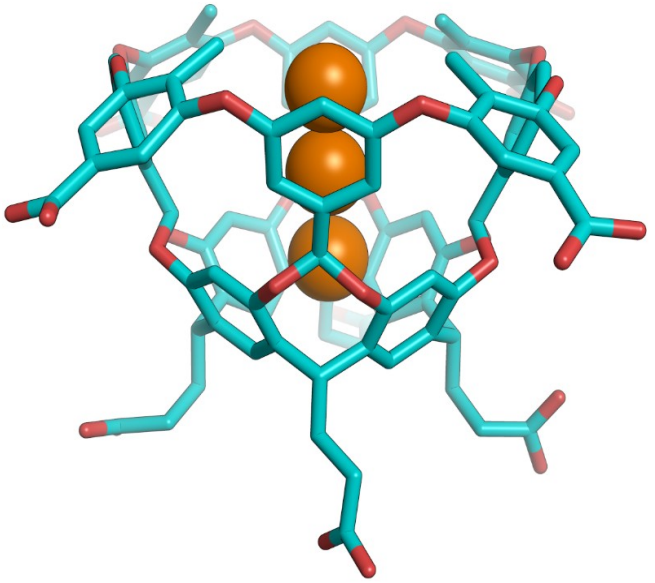
- AGBNP2 Spheres are “attached” to solute atoms and move accordingly
- Their free volumes also change during MD

Locating High Free Energy Sites

Methyl octa-acid



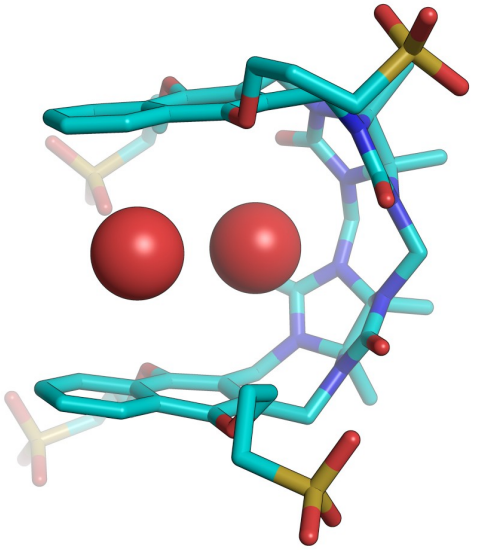
Hydration Sites



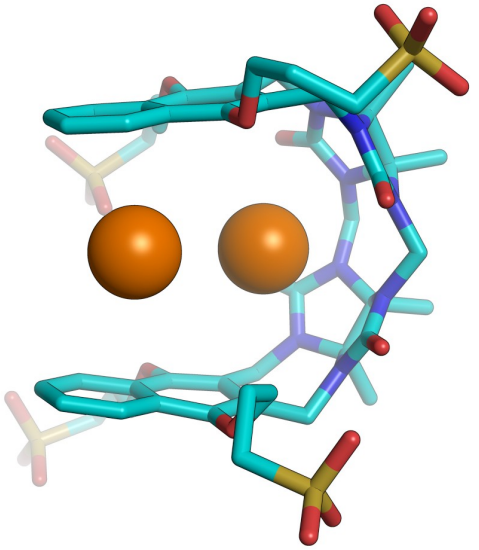
AGBNP2 Spheres

Locating High Free Energy Sites

CB Clip



Hydration Sites



AGBNP2 Spheres

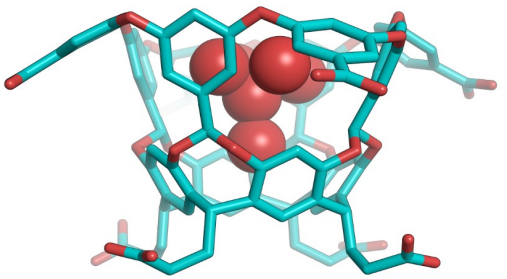
Scoring High Free Energy Sites

Preferred functional form to assign scores to AGBNP2 spheres:

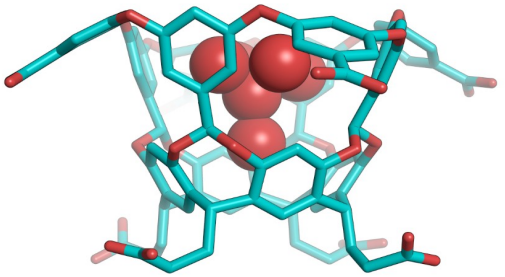
$$h_w = \text{Occupancy}_w \times [E_w - E_w(\text{bulk})]$$

Octa-acid

index	x	y	z	occ	Etot	score
0	-0.33	0.35	-0.17	0.62	-6.99	1.58
1	2.85	0.04	0.83	0.40	-8.14	0.56
2	1.42	-1.20	2.69	0.40	-8.14	0.56
3	1.02	-2.76	0.61	0.40	-8.14	0.56
4	2.54	-1.76	-1.00	0.40	-8.14	0.56
5	1.01	-0.95	0.46	0.29	-8.98	0.16
				2.52		3.97



Scoring High Free Energy Sites: Octa-Acid Hosts



- Validation on SAMPL4 data gave poor agreement
- Model overemphasized displacement of bottom cavity
- Turned to HSA scoring scheme used previously for docking scoring function

$$h_w = \text{constant}$$

- Such the overall AGBNP2 correction matched that in SAMPL4 (-7.3 kcal/mol on average)
- A mysterious additional -4 kcal/mol boost was needed to reproduce SAMPL4 data (more on this later)

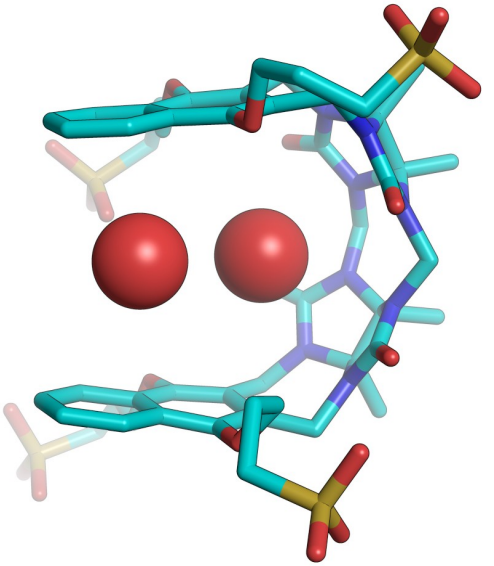
Scoring High Free Energy Sites: CB Clip

- Kept preferred parameterization scheme
- no adjustments

$$h_w = \text{Occupancy}_w \times [E_w - E_w(\text{bulk})]$$

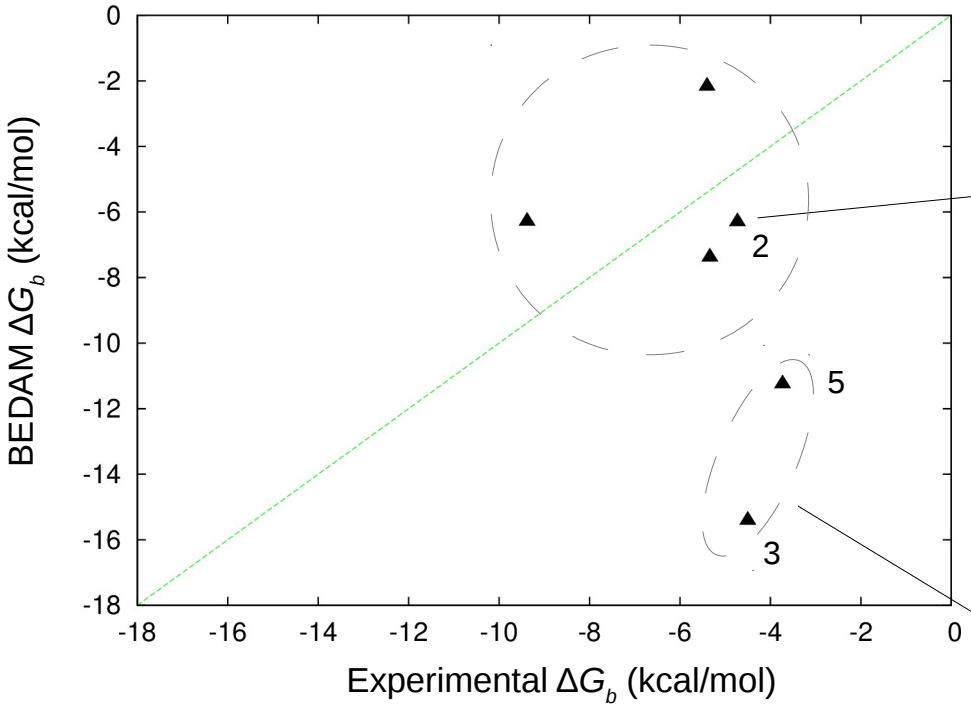
CB-Clip

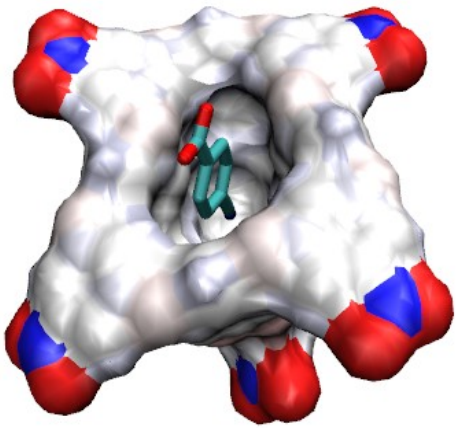
index	x	y	z	occ	Etot	score
0	-0.14	-0.19	0.08	0.92	-8.03	1.38
1	-0.8	1.33	2.6	0.53	-9.07	0.24
2	1.87	0.81	-2.62	0.46	-9.08	0.21
3	1.28	2.56	0.31	0.42	-8.52	0.42
4	1.72	1.97	-0.7	0.39	-8.71	0.32
6	0.17	2.44	1.46	0.32	-8.82	0.22
				3.04		2.80

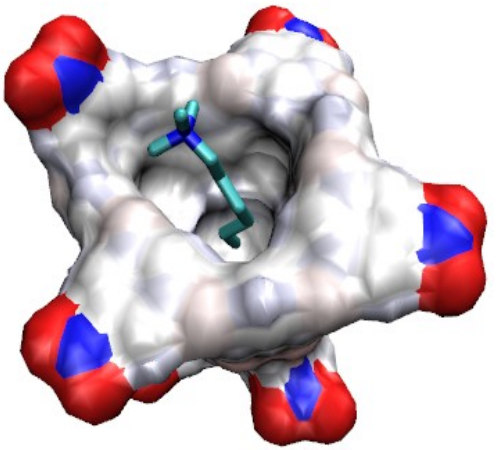


Results: Octa-Acid

- Disastrous



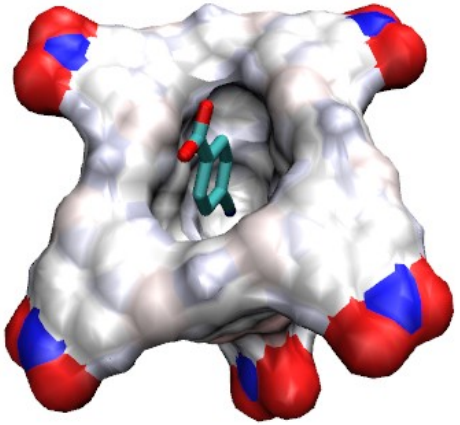


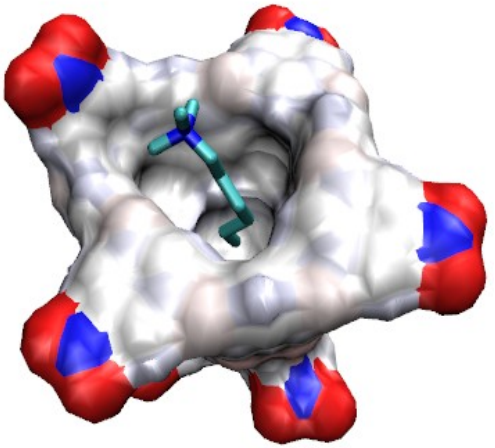


- Positively-charged guests are overly favorable with respect to negatively-charged guests
- Related to large negative charge of the host?

Hypothesis

- The overall charge of the host repels negatively-charged guests and attract positively charged ones
- Electrostatic model (Generalized Born) does not sufficiently screen charges
- Possible explanation for need of -4 kcal/mol boost for carboxylates and overly large affinities of alkylammonium guests





- Must take into account counter-ion screening: Debye-Huckel

$$E_{\mathrm{GB}} = -\left(1 - \frac{1}{\epsilon}\right)\frac{q_i q_j}{f_{GB}(r_{ij})} \longrightarrow E_{\mathrm{GB}} = -\left(1 - \frac{e^{-kr}}{\epsilon}\right)\frac{q_i q_j}{f_{GB}(r_{ij})}$$

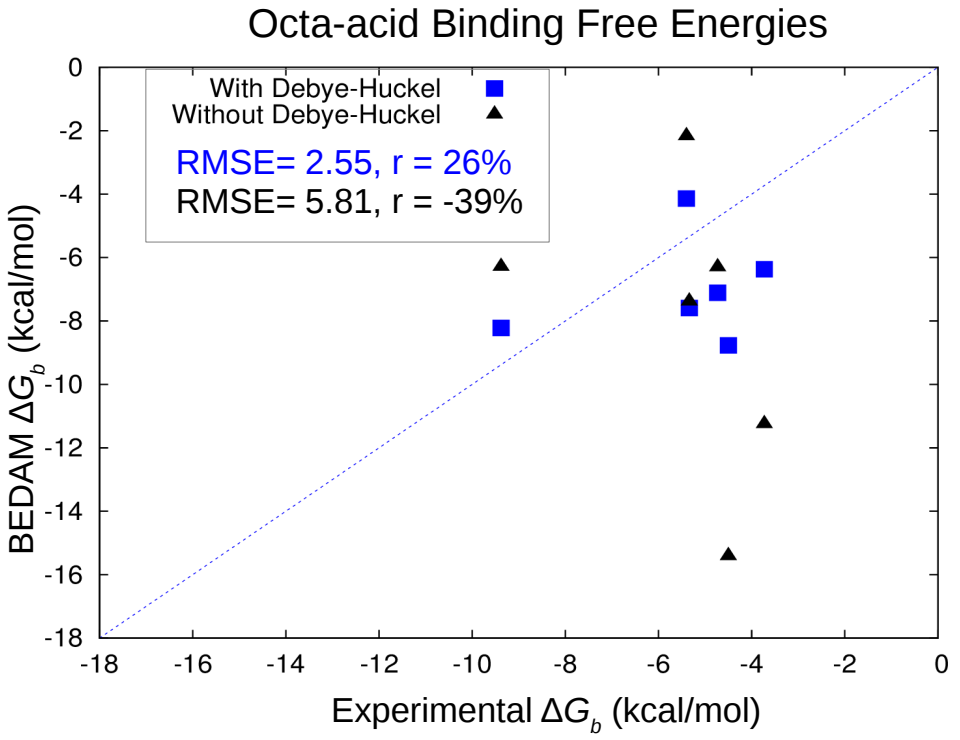
D. Case et al. ca. 1999

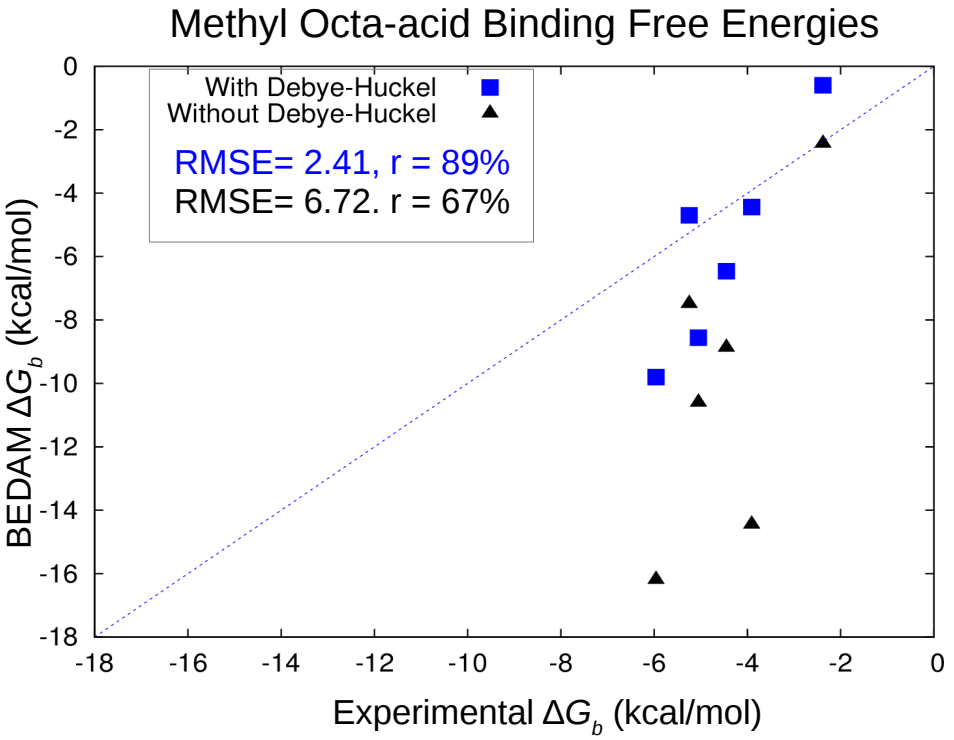
$$k = \left(8\pi\lambda_{\mathrm{B}} I\right)^{1/2}$$

I from 10 mM Na_3PO_4

Confirmation

- Recalculate with Debye-Huckel and remove the -4 kcal/mol (maybe no longer mysterious) empirical boost

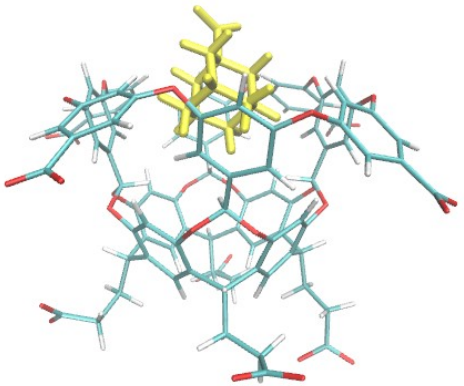




- The “charge bias” is almost eliminated
- Mediocre correlation coefficient (26%) for octa-acid is reasonable given the implicit solvation model and the small range of experimental data

Confirmation

Large decrease in affinity of guest 4 is nicely reproduced

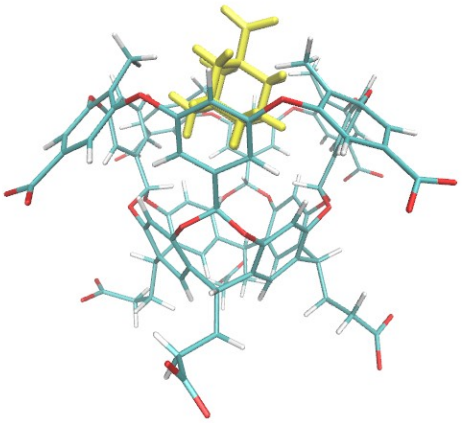


Expt

$\Delta\Delta G_b^\circ = +7.0$

$\Delta\Delta G_b^\circ = +8.0$

Calc.

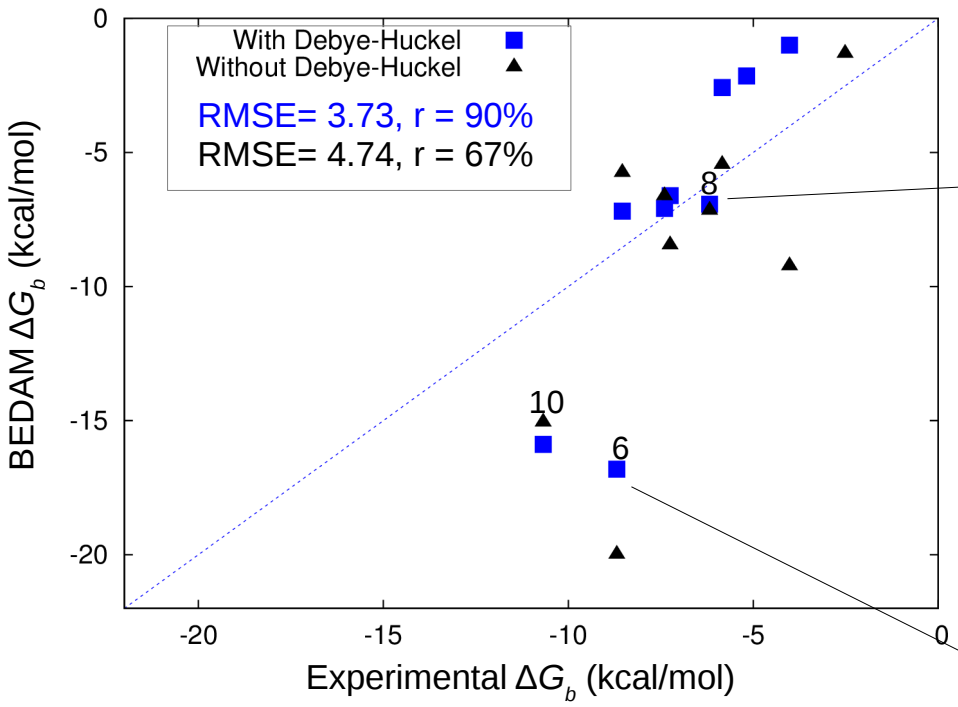


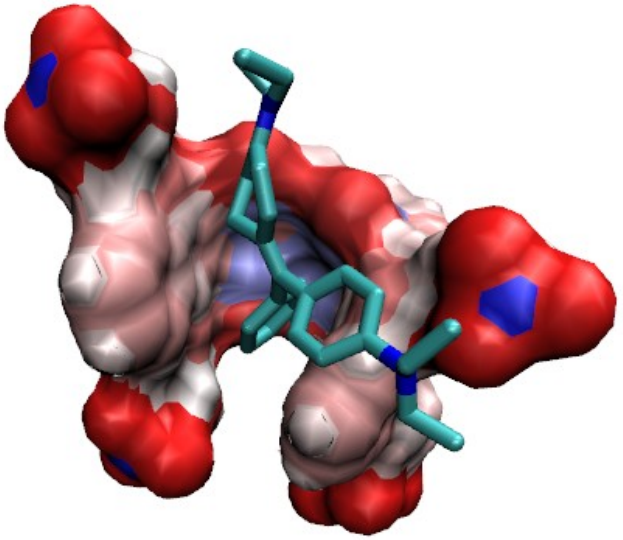
octa-acid

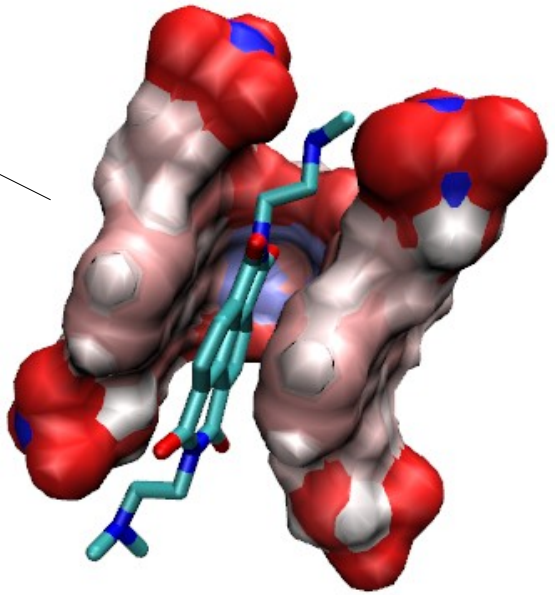
methyl-octa-acid

Results: CB Clip

- Good correlation







- Maybe issues with large and flat guests

Conclusions

- The results confirm the applicability of Hydration Site Analysis to alchemical binding free energy applications
- The first round of predictions was affected by insufficient Generalized Born electrostatic treatment of salt effects (a defect amplified by the large charge of the host).
- HSA hydration data and better physics (Debye-Huckel) reproduces experimental data without the need of empirical corrections.
- Implicit solvation not suitable to detect small binding free energy differences in octa-acid dataset. The resolution limit appears to be ~2 kcal/mol (sufficient for virtual screening never the less).
- The usefulness of SAMPL experiments to drive progress in the field is once again confirmed.

Thank you!

Poster by Rajat Pal et al. tonight