open
forcefield

@openforcefield

www.openforcefield.org

Tuning Potential Functions to Host-Guest Binding Data

A case study with implicit solvent

Jeffry Setiadi, Gilson Lab UCSD. 14th of May 2023

Outline

- Background and motivation
- Optimize parameters to Host-Guest ΔG_{b}
- Benchmark to Protein-Ligand ΔG_{b}
- Benchmark to Hydration free energies ΔG_{solv}
- Optimize to both Host-Guest ΔG_{b} and ΔG_{solv}

Tuning potential Functions

Background

- We want a force field that can describe protein-ligand binding accurately (CADD)
- Forcefield parameters are often optimized to fit QM calculations
- Recent efforts to optimize FF parameters to physical and chemical data:
  - Small molecule crystals
  - Liquid-state properties
- OpenFF v2.0.0 “Sage” – refitted LJ to condensed phase mixture data

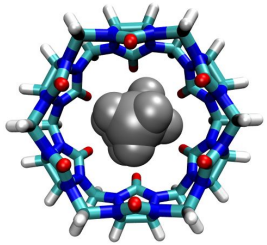
This project

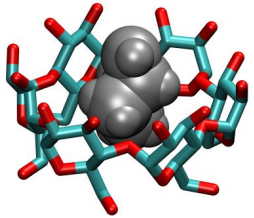
- Target binding data – host-guest systems as a surrogate for protein-ligand systems
- Optimize parameters of an implicit solvent model as a case study

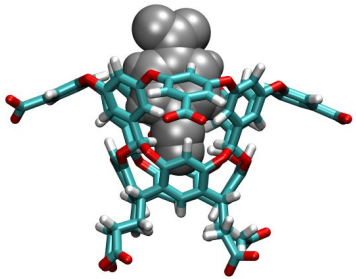
Host-Guest Binding Data Why are they useful?

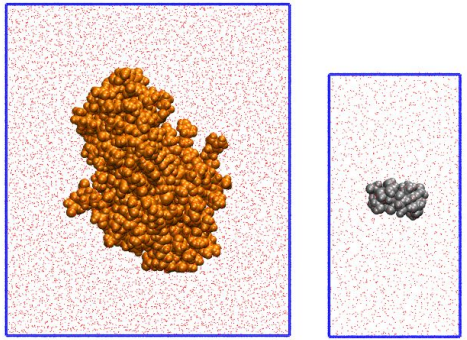
Background

- Host-guest systems
  - Pharmaceutical applications
  - conformational changes upon binding
  - binding affinities comparable to protein-ligand systems
  - predictable protonation states
- SAMPL blind challenges
  - test of force fields
  - test of methods
- Smaller system size – a good model to test and optimize force fields









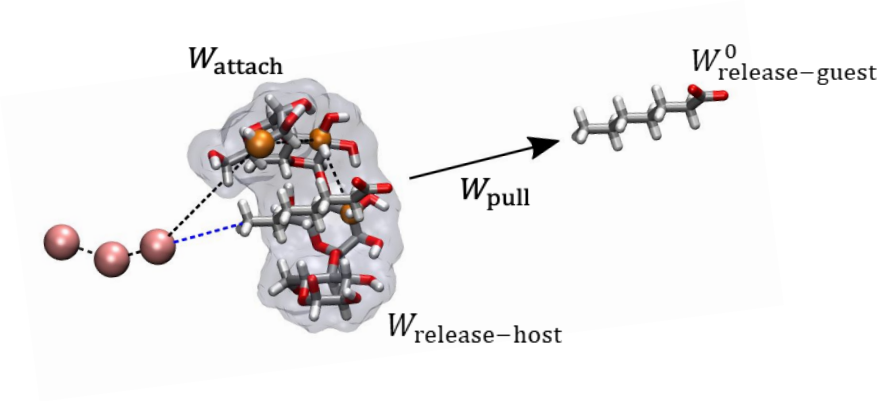
Infrastructure

Simon Boothroyd

https://github.com/openforcefield/openff-evaluator

pAPRika

An advanced toolkit for binding free energy calculations



$$\Delta G^0_{\text{b}} = -\left(W_{\text{attach}} + W_{\text{pull}} + W_{\text{release-host}} + W^0_{\text{release-guest}}\right)$$

https://github.com/GilsonLabUCSD/pAPRika

Calculations

- attach-pull-release (APR) method with pAPRika integrated in the OpenFF-Evaluator infrastructure
- APR Phases
  - *attach* – 15 windows
  - *pull* – 18 Å over 46 windows
  - *release* – 15 windows

30 ns production per window

JCTC Journal of Chemical Theory and Computation Article
Cite This: J. Chem. Theory Comput. 2019, 15, 6225−6242 pubs.acs.org/JCTC

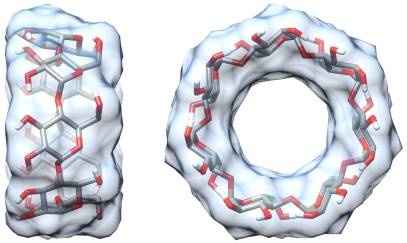
Binding Thermodynamics of Host−Guest Systems with SMIRNOFF99Frosst 1.0.5 from the Open Force Field Initiative

David R. Slochower,[†] Niel M. Henriksen,[†] Lee-Ping Wang,[‡] John D. Chodera,[§] David L. Mobley,[∥] and Michael K. Gilson*[,†]

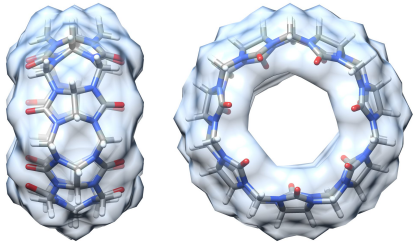
Data Set

- Host molecules
  - cyclodextrins – αCD, βCD
  - cucurbiturils – CB7, CB8
  - octa-acids – OA, TEMOA
- 126 host-guest complexes
  - SAMPL Challenges
  - amines, cyclic alcohols, carboxylates and drug-like molecules
- **Training set** – 36 complexes

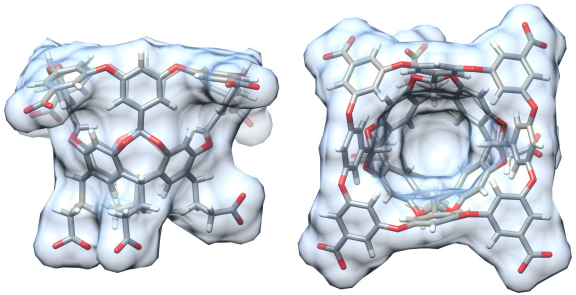
Cyclodextrin



Cucurbituril



Octa-acid



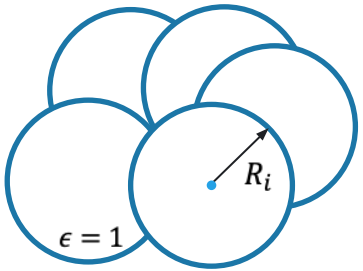
https://github.com/slochower/host-guest-benchmarks

Host-Guest Binding

Potential Function

Force Field

- Open Force Field v2.0.0 “Sage”
- Partial Charges – AM1-BCC ELF10
- Generalized Born Implicit Solvent – OBC2 model (
  Or



OpenMM - RTX 3090
- Implicit – 2.0 μs/day
- Explicit – 600 ns/day

Parameter Optimization

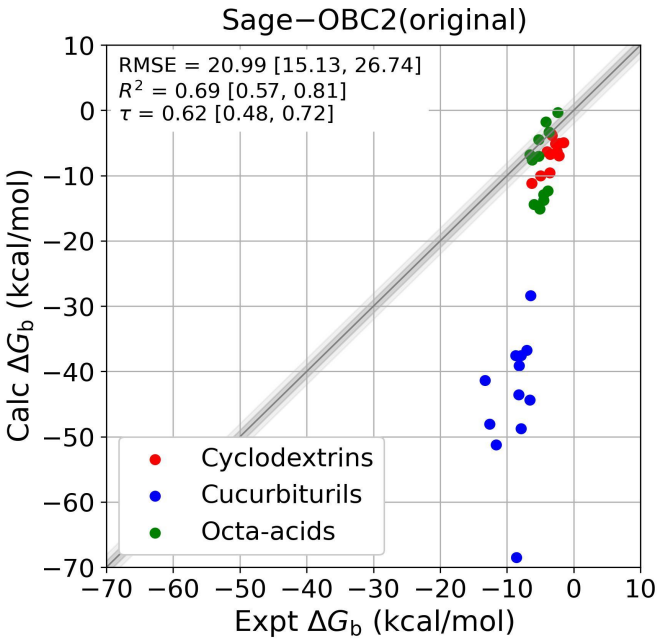
- *mBondi2* radii as the initial values
- Optimize all 5 relevant GB radii

Atom Type	SMIRKS	GB Radius (Å)
H	[#1:1]	1.20
H-N	[#1:1]-[#7]	1.30
C	[#6:1]	1.70
N	[#7:1]	1.55
O	[#8:1]	1.50

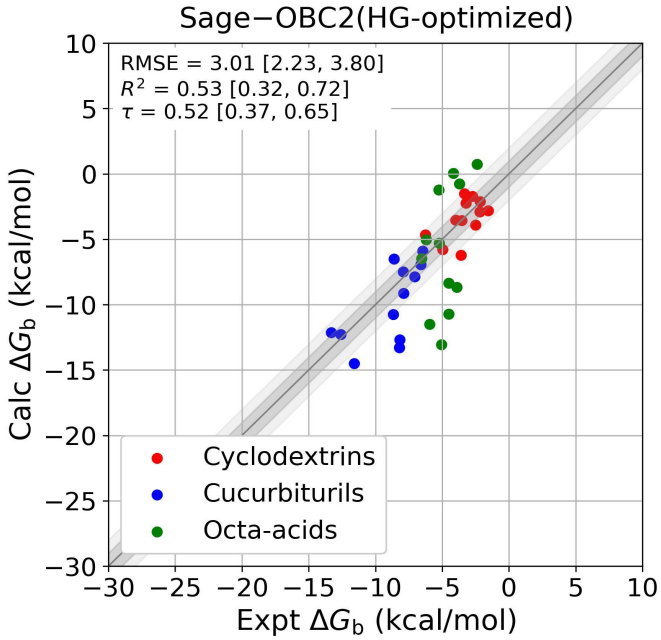
Host-Guest Binding

Host-Guest Training Set

before optimization



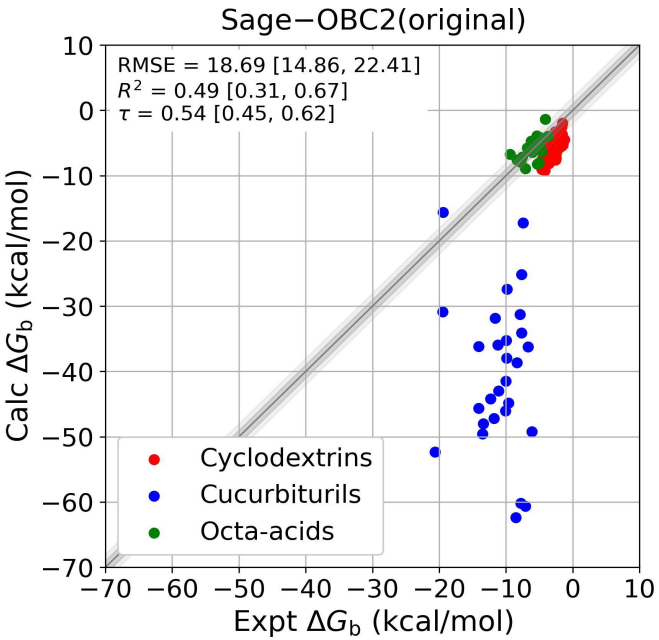
after optimization



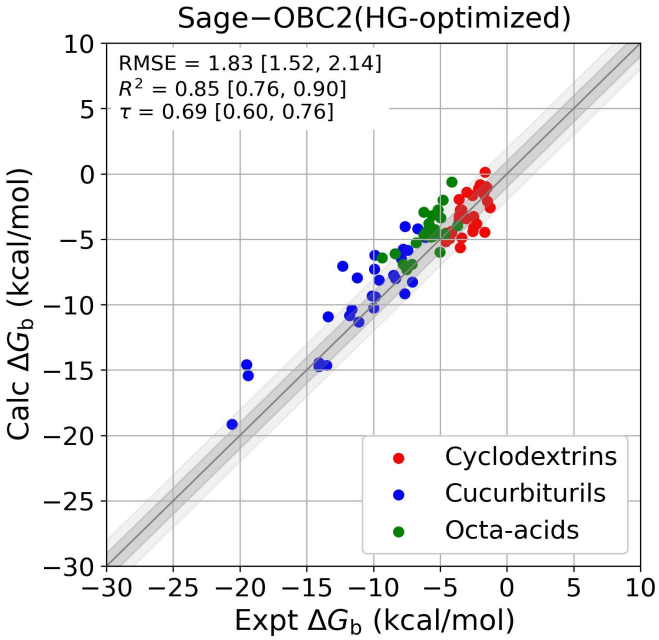
ForceBalance converged after 21 iterations

Host-Guest Binding Host-Guest Test Set

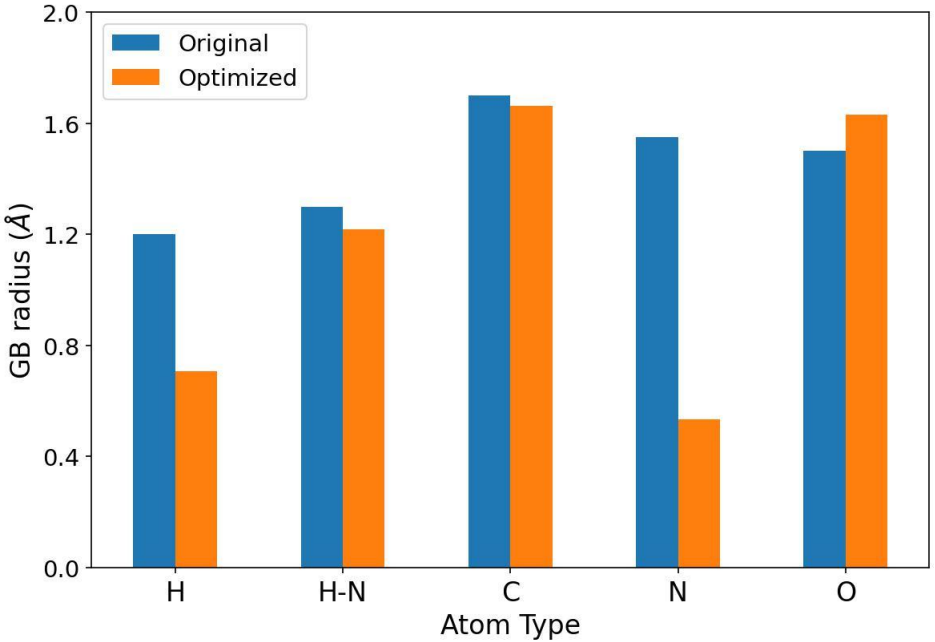
before optimization to the training set



after optimization to the training set

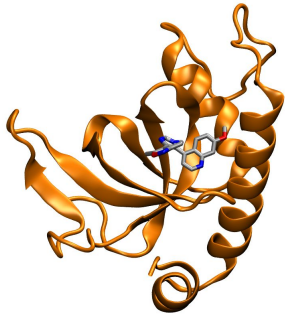


Host-Guest Binding GB radii

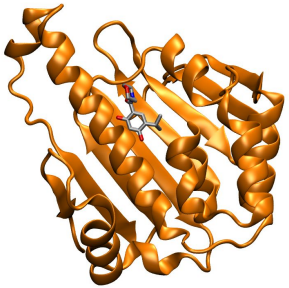


Atom Type	Original GB Radius (Å)	HG-Optimized GB Radius (Å)
H	1.20	0.71
H-N	1.30	1.22
C	1.70	1.66
N	1.55	0.53
O	1.50	1.63

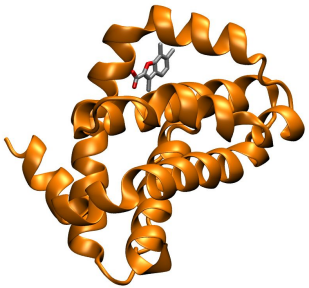
Test Systems



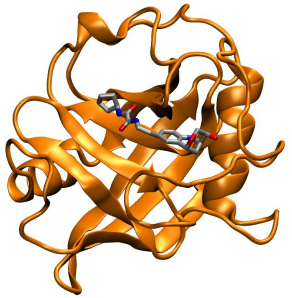
PWWP1
12 compounds



HSP90
18 compounds



MCL-1
19 compounds



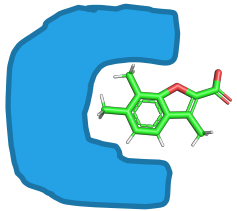
Cyclophilin D
10 compounds

- Total of 59 protein-ligand systems
- ΔG_b range between -2.7 and -12.6 kcal/mol
- Initial structures are freely available on GitHub

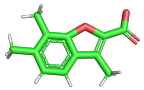
https://github.com/IAlibay/fragment-opt-abfe-benchmark

OpenMM - RTX 3090
- Implicit – 900 ns/day
- Explicit – 400 ns/day

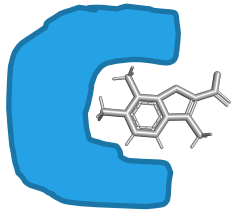
Alibay et. al. (2022)



ΔG^{site}_{Elec}



ΔG^{bulk}_{Elec}



ΔG^{site}_{LJ}

AMBER ff99SB-ILDN

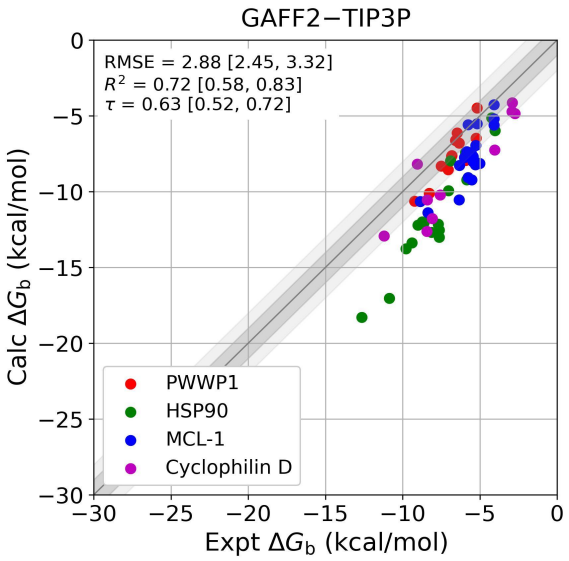
Calculations

- Double-decoupling method – *thermodynamic integration* (TI)
  - *OpenMM* and *OpenMMTools*
- Boresch-restraint
  - anchor atoms - *MDRestraintGenerator*
  - *pAPRika* for restraint setup
  - *site* – 15 windows, 30 ns
  - *bulk* – analytical
- Electrostatics – *annihilation*
  - Scale partial charges and GBSA
  - *site* – 11 windows, 30 ns
  - *bulk* – 11 windows, 5 ns
- Lennard-Jones – *decouple*

Protein-Ligand Binding

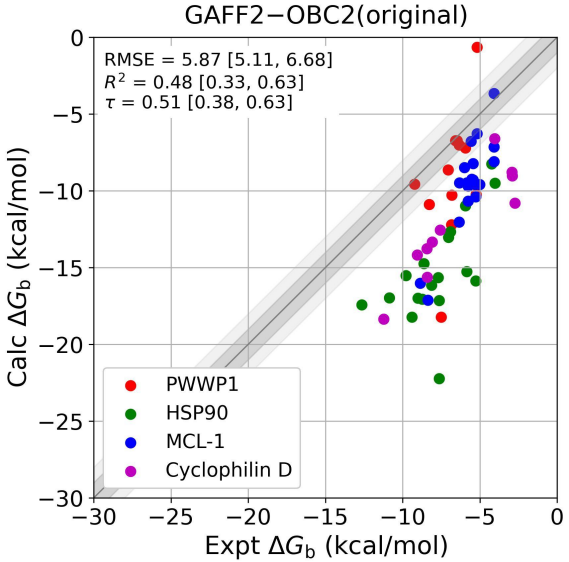
Original GB radii

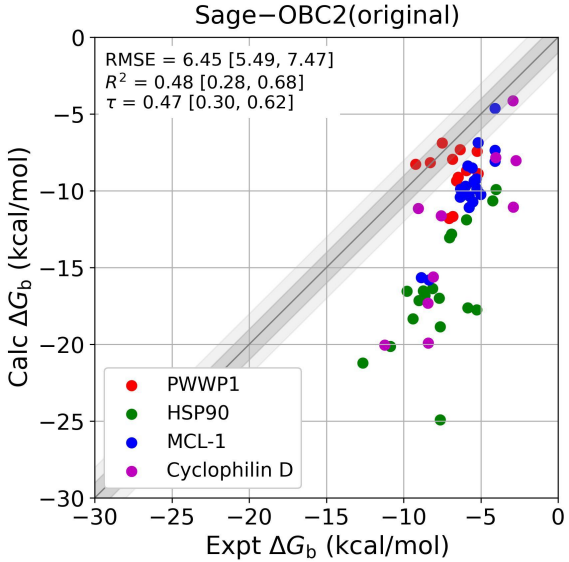
Explicit Solvent



Alibay et. al. (2022)

Implicit Solvent – before optimization to the H-G training set

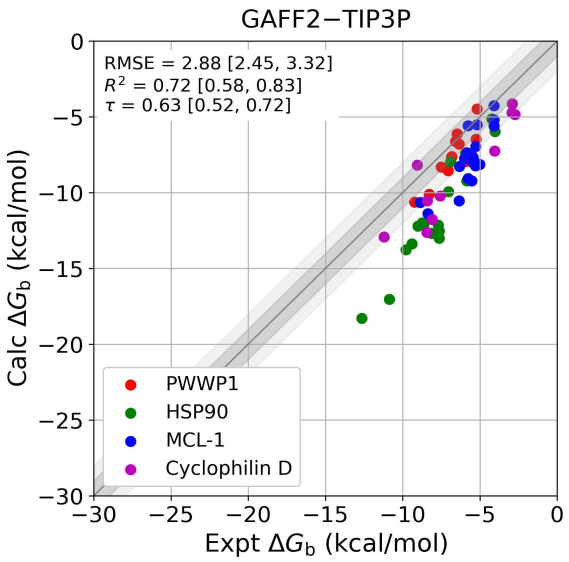




Protein-Ligand Binding

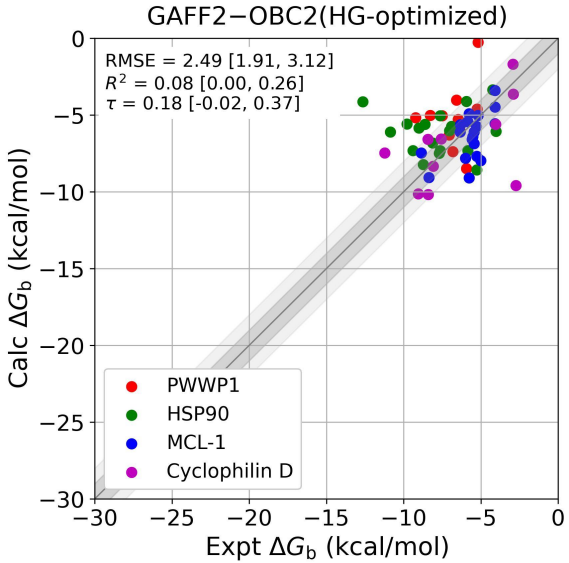
Host-Guest optimized GB radii

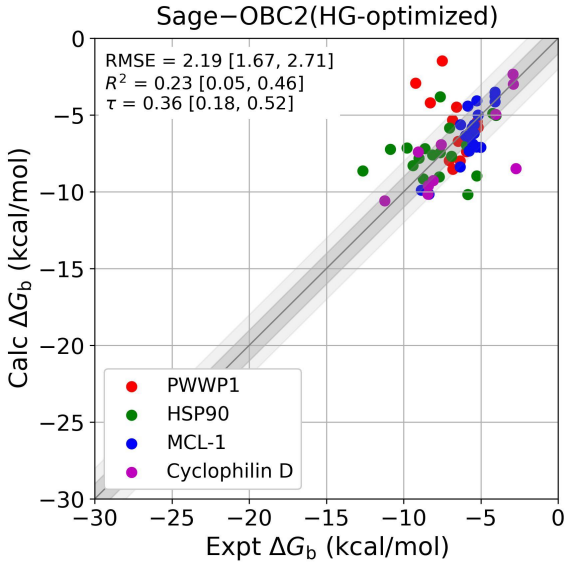
Explicit Solvent



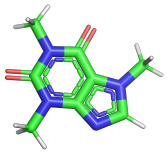
Alibay et. al. (2022)

Implicit Solvent – after optimization to the H-G training set

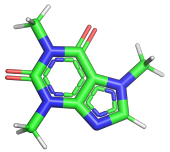




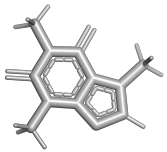
HFE Calculations



ΔG_{Elec}^{bulk}



ΔG_{Elec}^{vac}



ΔG_{LJ}^{bulk}

Calculations

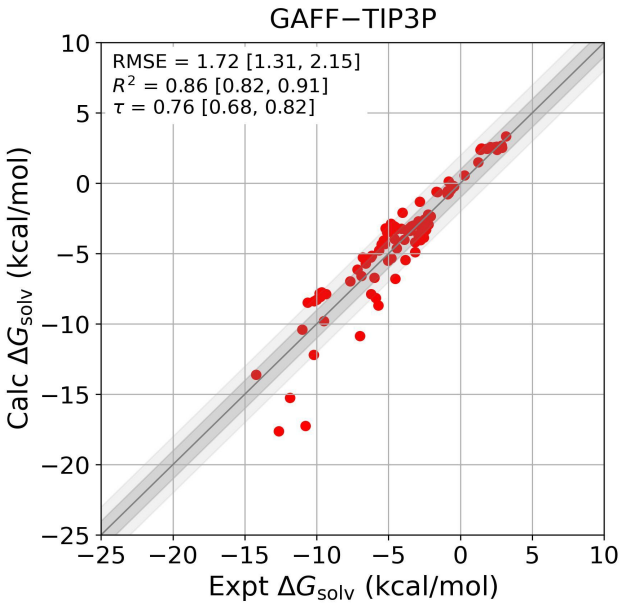
- 100 neutral small molecules from FreeSolv database – only molecules containing elements of H, C, N, and O.
- Yank through the OpenFF-Evaluator infrastructure
- Electrostatics – *annihilation*
  - *bulk* - scales partial charges and GBSA
  - *vacuum* - scales partial charges
- Lennard-Jones - *decouple*
  - softcore potential
- Number of λ - *trailblaze* algorithm
- 2 ns production run per window

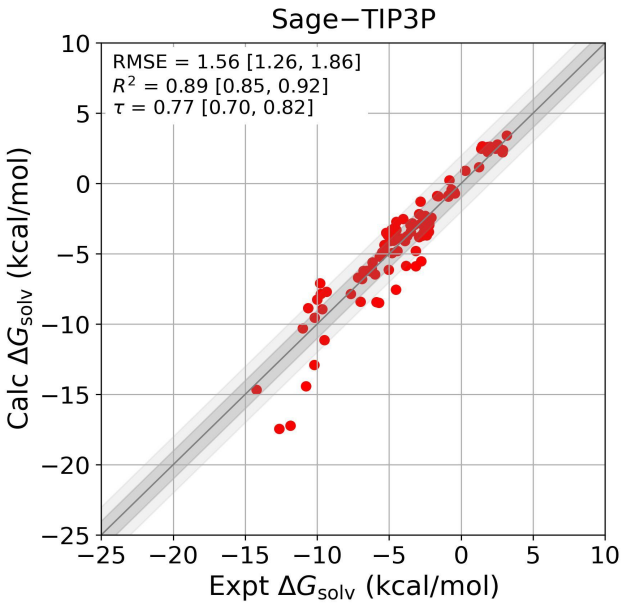
Hydration Free Energy

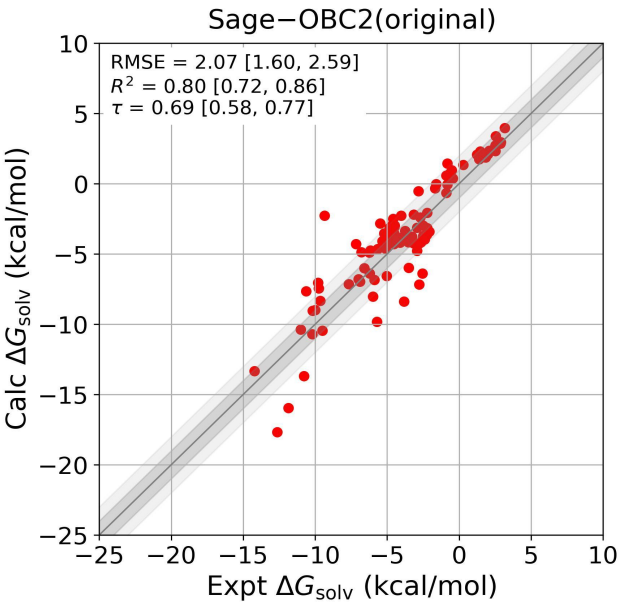
Comparison with Explicit Solvent

Explicit Solvent

Implicit Solvent





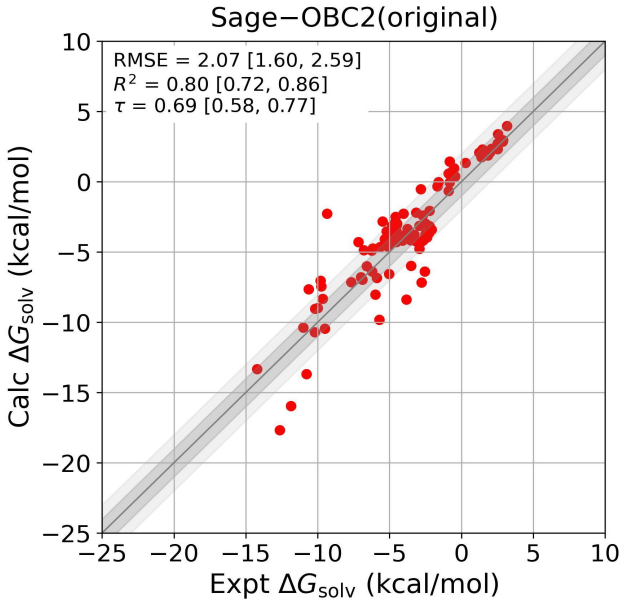


Mobley Lab (2017)

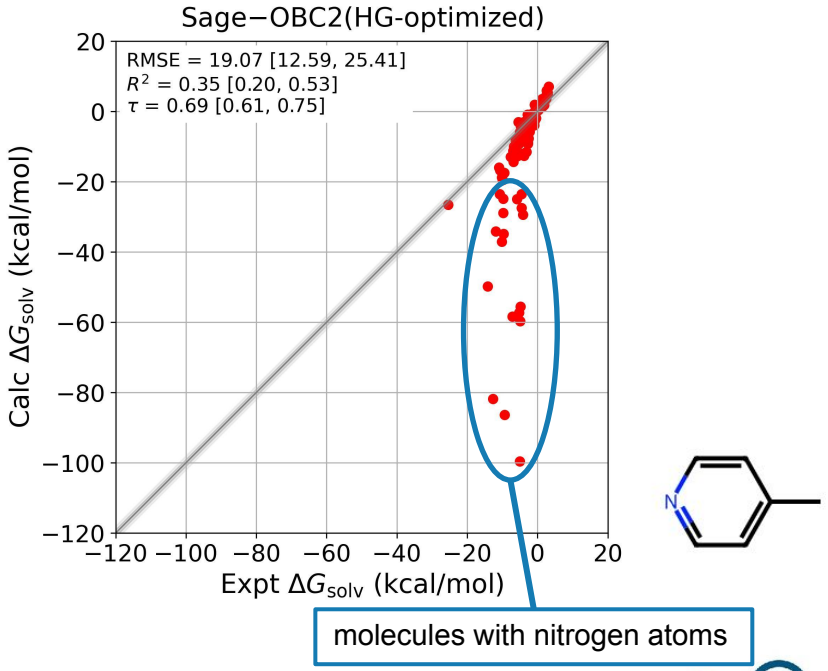
Hydration Free Energy

Implicit solvent

before optimization to the H-G training set



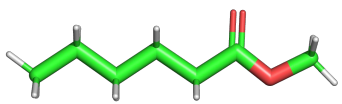
after optimization to the H-G training set



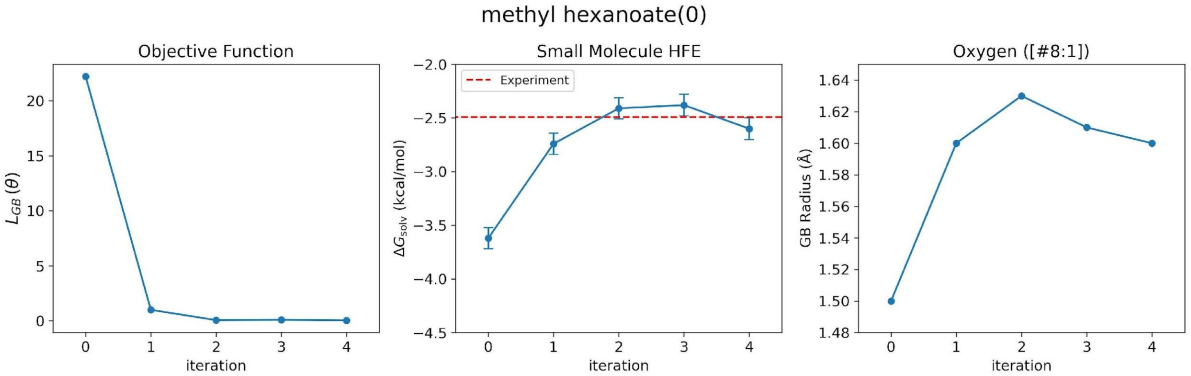
HFE and H-G Binding

Individual Optimization

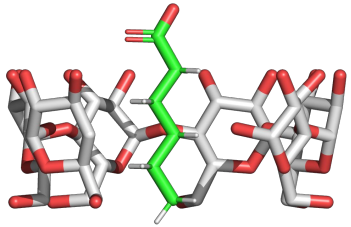
ΔG_{solv}



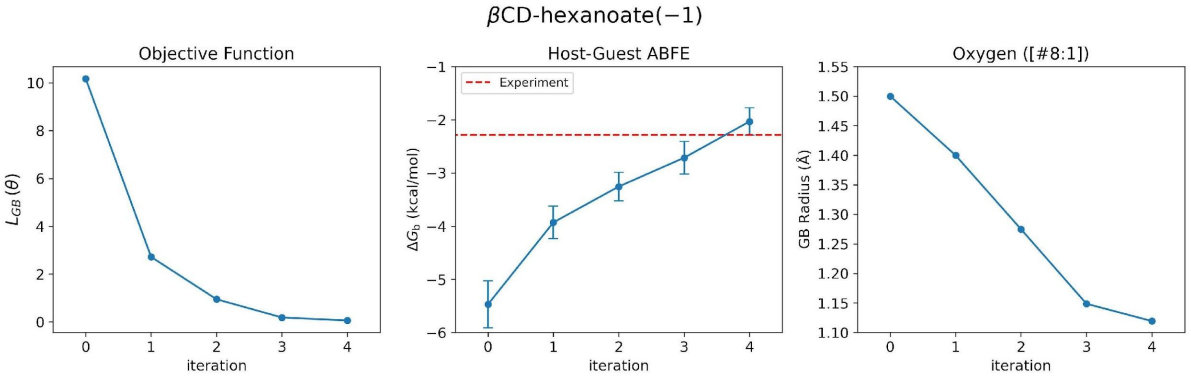
methyl hexanoate(0)



ΔG_b



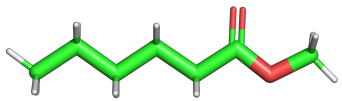
βCD-hexanoate(−1)



HFE and H-G Binding

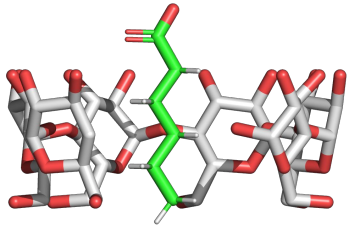
Optimize to both properties

ΔG_{solv}

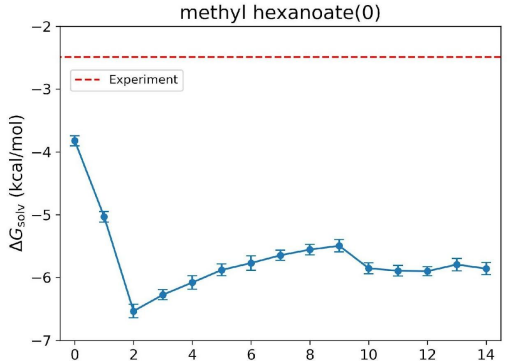


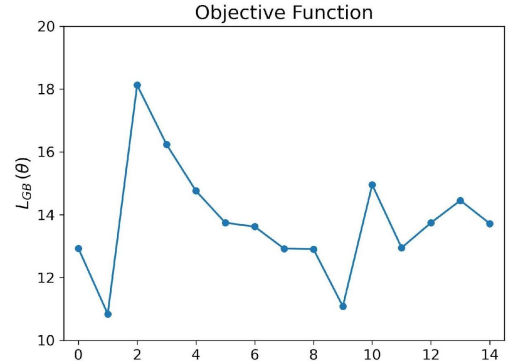
methyl hexanoate(0)

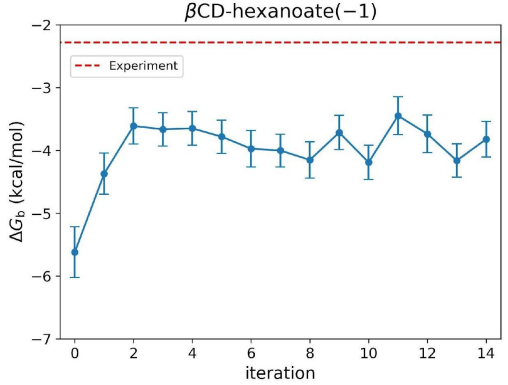
ΔG_b

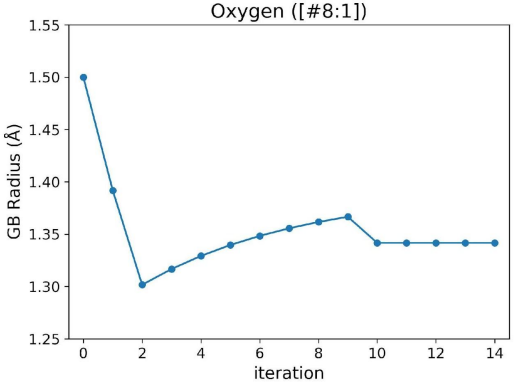


βCD-hexanoate(−1)









Conclusion

Summary

- **Host-guest systems –** optimized GB radii to host-guest ΔG_{b} with the infrastructure and give good results for the host-guest test set.
- **Protein-ligand systems -** The HG-optimized GB radii improves the RMSE but deteriorates R^2.
- **Hydration free energy -** Fitting GB radii to host-guest ΔG_b deteriorates the small molecule ΔG_{solv}.
- **Optimization to two properties -** The optimizer cannot find good parameters (GB radii) that can fit both host-guest ΔG_b and ΔG_{solv}

What's next?

Implicit solvent

- Test with other types of host molecule and diversify the training set.
- Optimize other parameters in the implicit solvent model.
- Investigate different number of atom types with the OBC2 model.
- Modify the OBC2 model or try a different implicit solvent model.

Explicit solvent – integrate host-guest ΔG_b into

Acknowledgements

- David Slochower (alumni)
- Niel Henriksen (alumni)
- Mike Gilson

NIH National Institutes of Health
Turning Discovery Into Health

open forcefield

- Simon Boothroyd (alumni)
- David Dotson
- Jeffrey Wagner
- Matt Thompson
- Lee-Ping Wang

SDSC SAN DIEGO SUPERCOMPUTER CENTER

Thank you for listening!

open
forcefield

www.openforcefield.org

open
forcefield

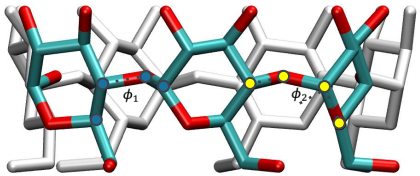
@openforcefield

www.openforcefield.org

Supplementary Slides

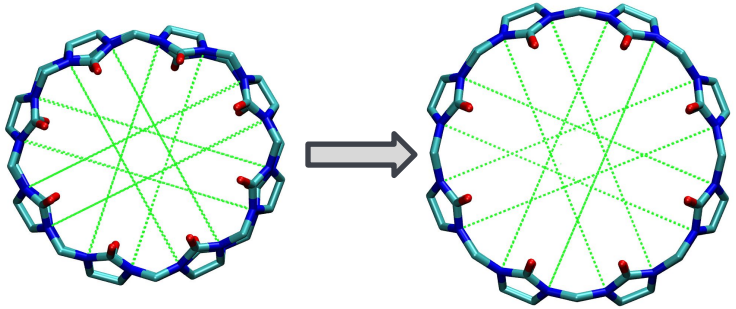
Host-Guest Binding Calculations

Host conformational restraints



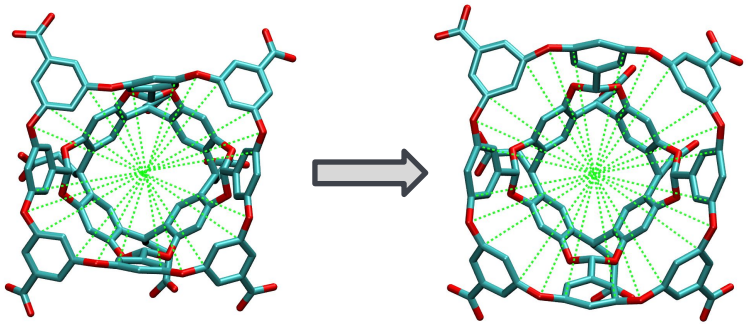
<u>Cyclodextrin</u>

- 2 dihedral restraints on glycosidic link
- $k_\phi = 6.0\ \text{kcal/mol/rad}^2$
- αCD – $[\phi_1, \phi_2] = [-108.8°, 104.3°]$
- βCD – $[\phi_1, \phi_2] = [-112.5°, 108.7°]$



<u>Cucurbiturils</u>

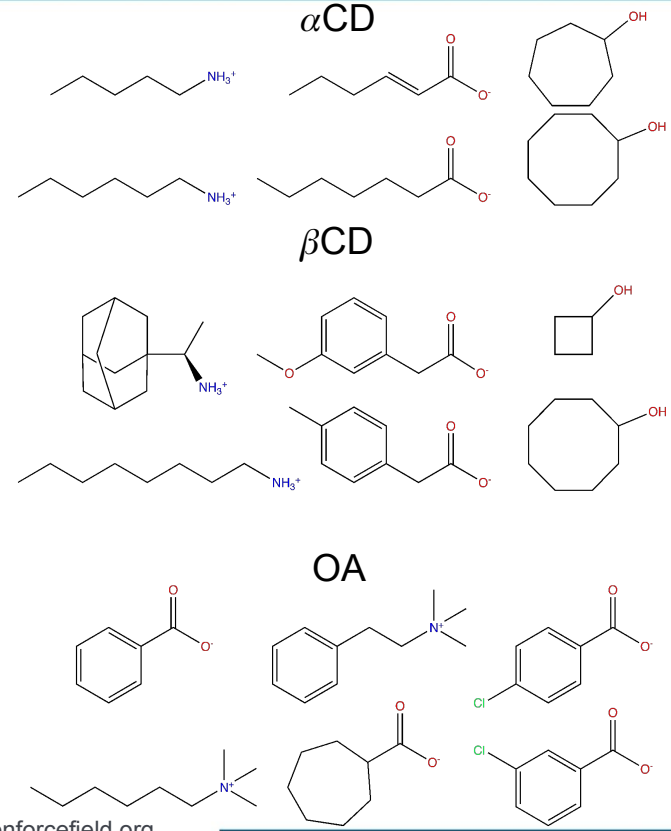
- CB7 – 14 jack restraints
- CB8 – 16 jack restraints
- $k_{jack} = 15.0\ \text{kcal/mol/Å}^2$

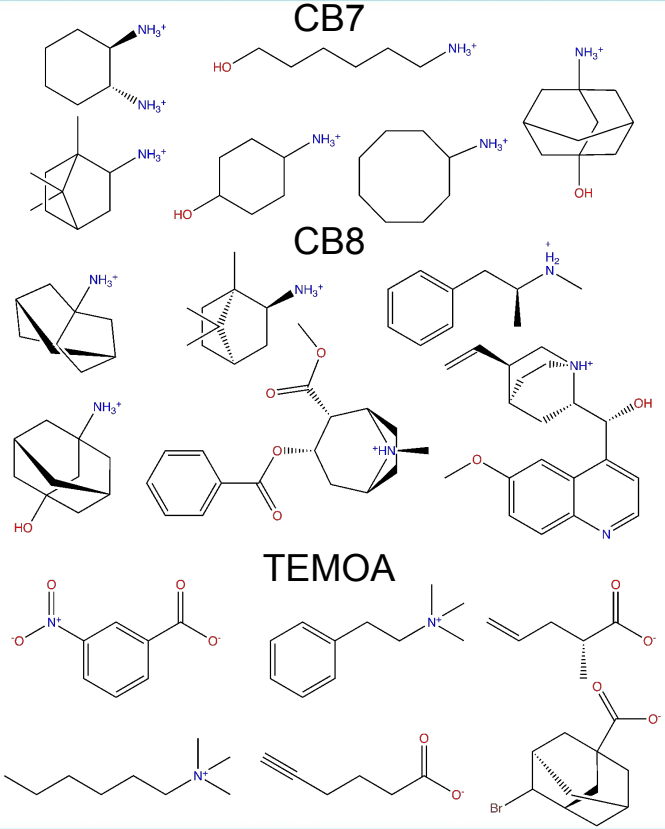


<u>Octa acids</u>

- OA – 12 jack restraints
- TEMOA – 12 jack restraints
- $k_{jack} = 25.0\ \text{kcal/mol/Å}^2$

Host-Guest Training Set





Implicit Solvent Model

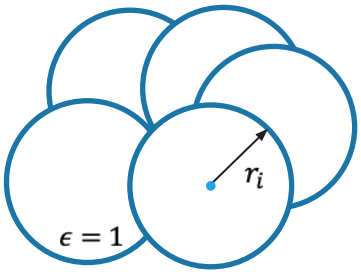
$$\Delta G_{GB} = -\frac{1}{2}\sum_{ij}\frac{q_i q_j}{f^{GB}}\left(1-\frac{1}{\epsilon_w}\right) \qquad \Delta G_{SA} = 4\pi E_{SA}\sum_i (r_i + r_{solvent})^2\left(\frac{r_i}{R_i}\right)^6$$

$$f^{GB} = \left[d_{ij}^2 + R_i R_j \exp\left(-\frac{d_{ij}^2}{4R_i R_j}\right)\right]^{1/2}$$

$$R_i^{-1} = \rho_i^{-1} - r_i^1 \tanh(\alpha\Psi - \beta\Psi^2 + \gamma\Psi^3) \qquad \rho_i = r_i - 0.09\ \text{Å}$$

$OBC1: \alpha = 0.8, \beta = 0.0, \gamma = 2.91$

$OBC2: \alpha = 1.0, \beta = 0.8, \gamma = 4.85$



$\epsilon_w = 80$

Alchemical Scaling of GBSA

$$s_i(\lambda, \eta_i) = \lambda\eta_i + (1 - \eta_i)$$
$$s_{ij}(\lambda, \eta_i, \eta_j) = s_i(\lambda, \eta_i)\, s_j(\lambda, \eta_j)$$

$$\Delta G_{GB}(\lambda, \eta) = -\frac{1}{2}\sum_{ij} s_{ij}(\lambda, \eta_i, \eta_j)\frac{q_i q_j}{f^{GB}}\left(1-\frac{1}{\epsilon_w}\right)$$

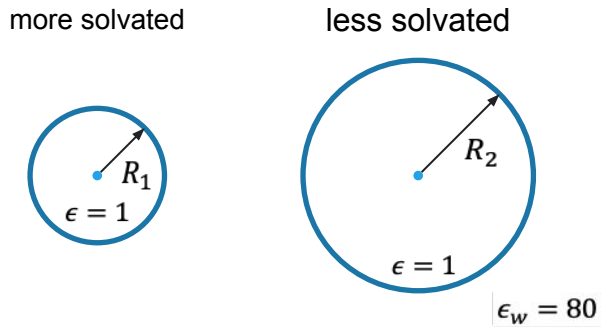
$$\Delta G_{SA}(\lambda, \eta) = 4\pi E_{SA}\sum_i s_i(\lambda, \eta_i)(r_i + r_{solvent})^2\left(\frac{r_i}{R_i}\right)^6$$

https://github.com/choderalab/openmmtools

http://getyank.org/latest/algorithms.html#generalized-gbsa-model

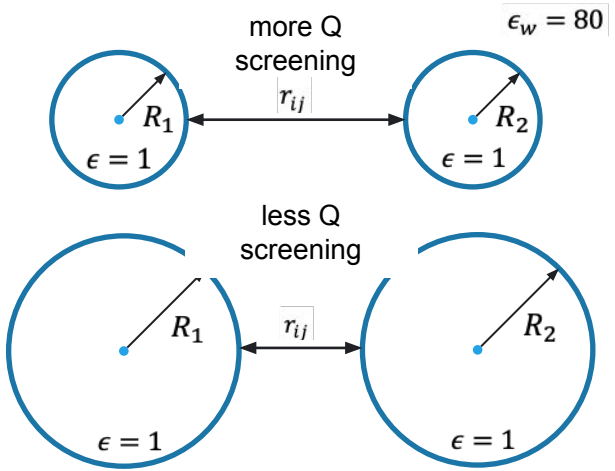
Hydration Free Energy

- A smaller GB radius increases solvation → ΔG_{solv} becomes more negative
- A larger GB radius decreases solvation → ΔG_{solv} becomes more positive



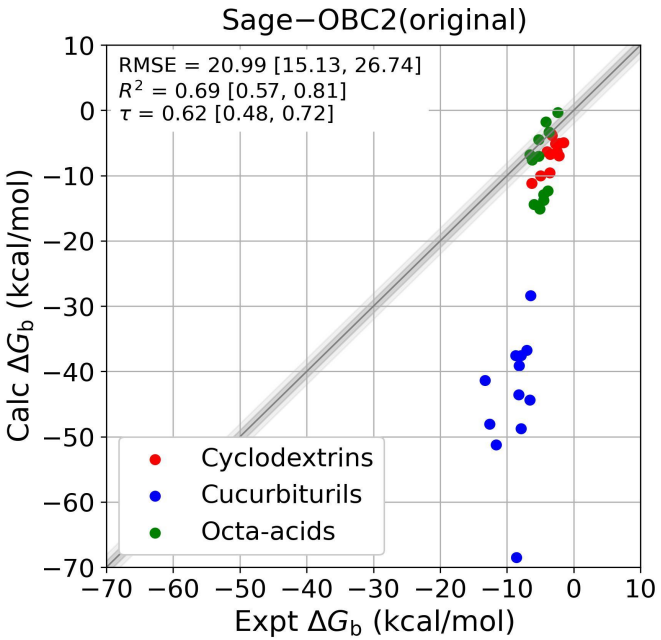
H-G Binding Free Energy

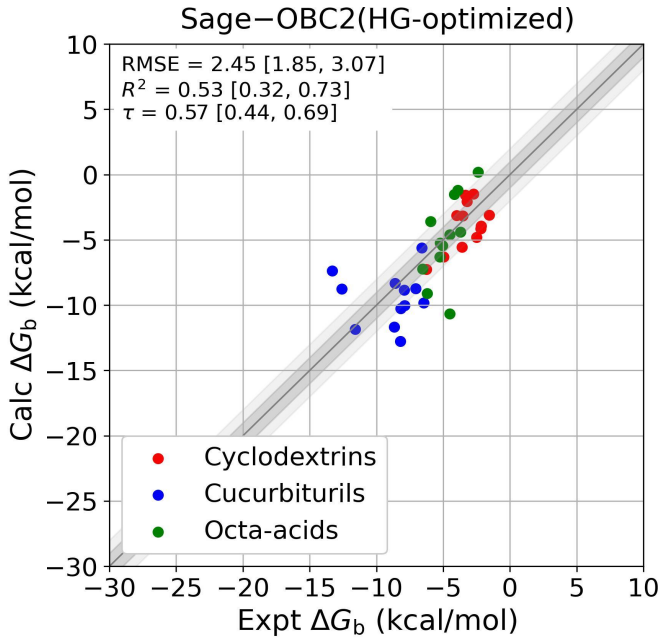
- A smaller GB radius increases solvation → more charge screening → ΔG_{b} becomes more positive
- A larger GB radius decreases solvation → less charge screening → ΔG_{b} becomes more negative



Optimization – 9 atom type

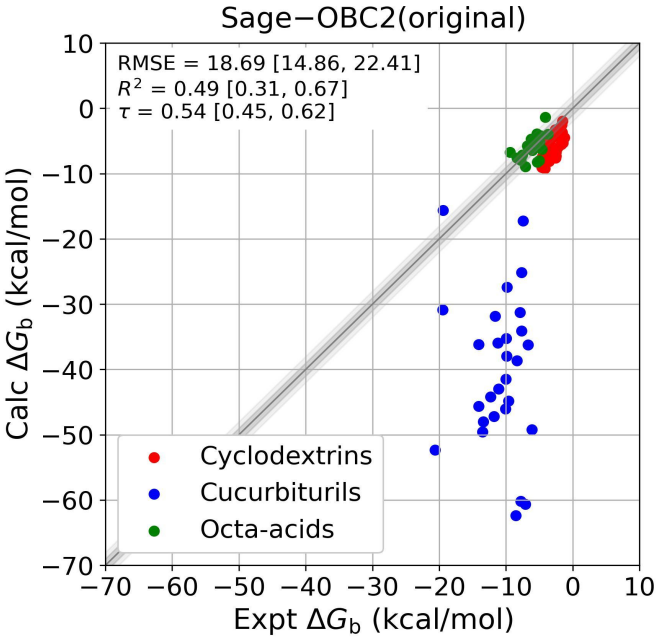
Host-Guest Training Set

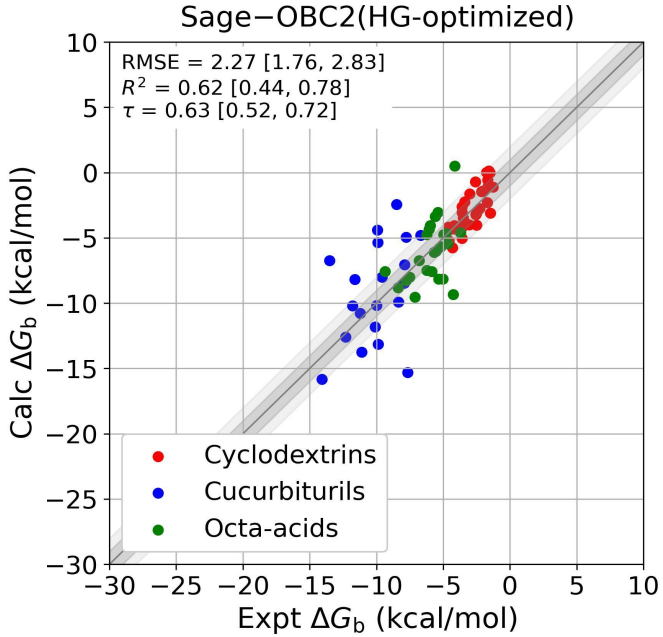




Optimization – 9 atom type

Host-Guest Test Set





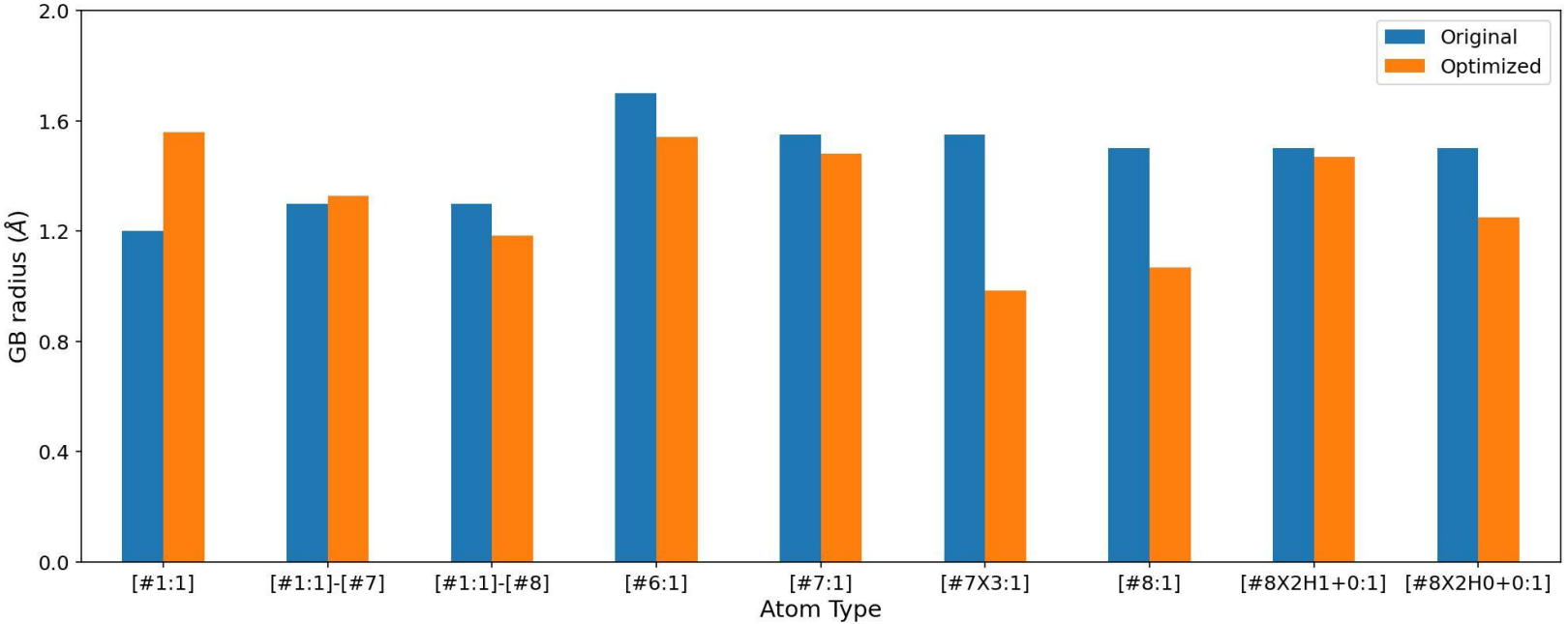
Optimization – 9 atom type

GB radii comparison – 9 atom type

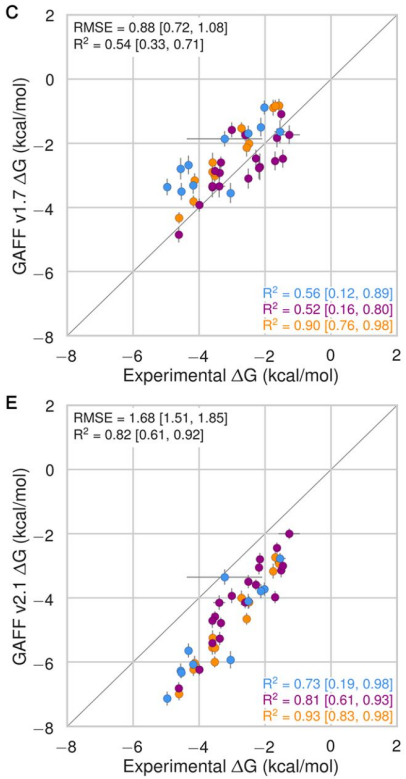
SMIRKS	Description	Original GB Radius (Å)	HG-Optimized GB Radius (Å)
[#1:1]	hydrogen	1.20	1.56
[#1:1]-[#7]	hydrogen bound to nitrogen	1.30	1.33
[#1:1]-[#8]	hydrogen bound to oxygen	1.30	1.18
[#6:1]	carbon	1.70	1.54
[#7:1]	nitrogen	1.55	1.48
[#7X3:1]	amine nitrogen	1.55	0.98
[#8:1]	carbonyl oxygen	1.50	1.07
[#8X2H1+0:1]	alcohol oxygen	1.50	1.47
[#8X2H0+0:1]	ether oxygen	1.50	1.25

Optimization – 9 atom type

GB radii comparison – 9 atom type



Host-Guest Stability



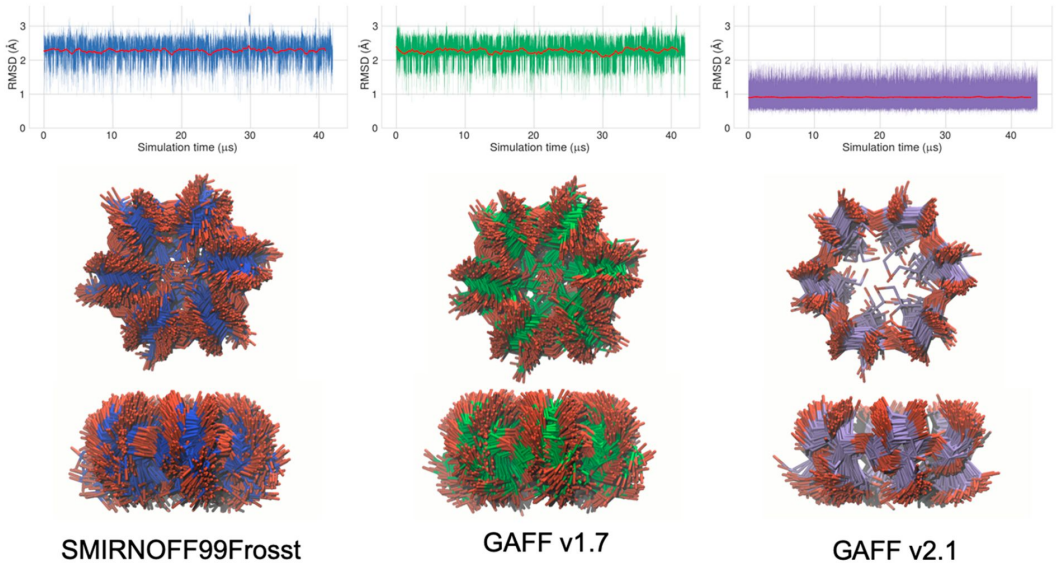
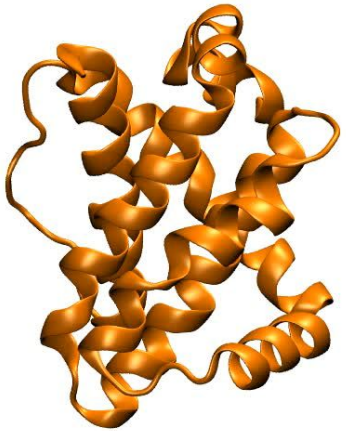


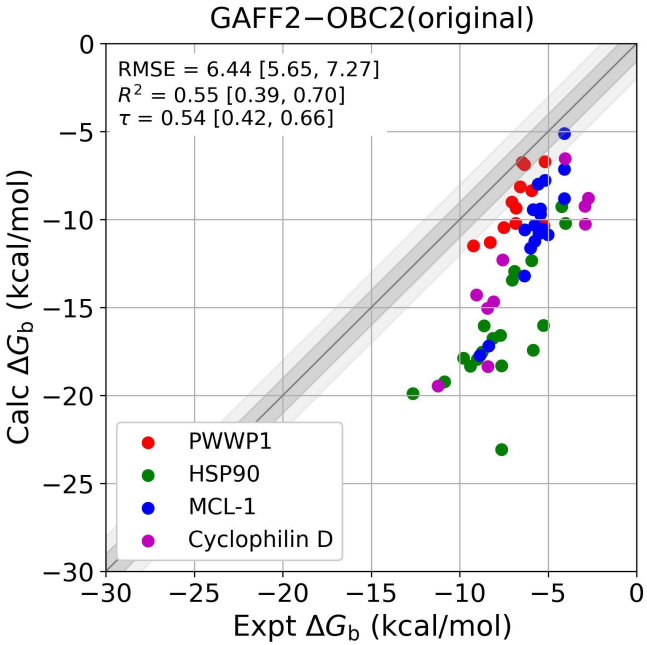
Figure 14. Top: Root mean square deviation (RMSD in Å) of free βCD in the three force fields. Each RMSD is calculated relative to the initial structure, a gas-phase minimization of βCD with GAFF v1.7. A 1000 frame moving average is plotted in red. Middle: top-view of the unoccupied cavity of βCD with no guest (200 snapshots over 1 μs). Bottom: side-view of the unoccupied 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.

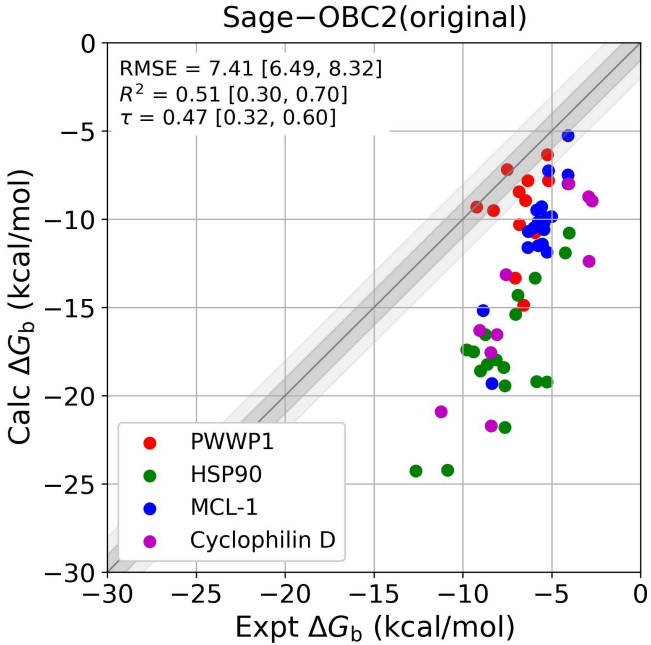
Protein stability

HG-optimized - 9 atom types



Implicit Solvent – before training





Protein-Ligand Binding

ff14SB Benchmark

Implicit Solvent – after training