

FEP/GBSA
and the
Approximated Generalized
Born Potential

John Terhorst

Department of Chemistry
Yale University

Monday, September 26, 2011

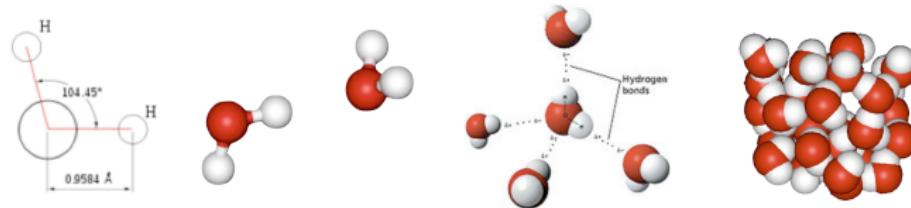


Overview

- Background
 - Explicit versus implicit water models.
 - The generalized Born/ surface area (GB/SA) solvation model.
 - GB/SA in protein simulations.
- Implementing free-energy perturbation (FEP) with GB/SA
 - Challenges with FEP/GBSA.
 - Energy trajectories and free energies of hydration.
- Making it better
 - The approximated generalized Born potential.
 - Statistical significance of the approximation.
 - Chlorine scan.

Explicit solvent models

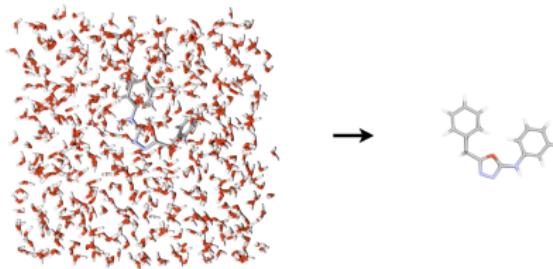
- The classical approach to modeling solvents.
 - Explicit treatment of all atoms.
 - Each atom is defined with its own coordinates, van der Waals radius, partial charge, and polarizability (if applicable).
 - All interactions between all atoms within their solvation shells or within a pre-defined cutoff are considered.
- Rely on molecular mechanics, rather than quantum mechanics.
 - Parameters are usually based on quantum mechanical calculations.



- A change in the liquid structure as a result of a perturbation to the system necessitates that all interactions be relaxed, or equilibrated, to once again achieve the system's low-energy state.

Implicit solvent models

- Designed to curtail some problems of the explicit approach to representing solvent, by abstracting solvent molecules into a dielectric continuum with the average properties of the liquid.



- The average influence of the solvent is determined by direct estimation of the free energy of solvation.
- Solvent is always equilibrated, by definition.
 - No sampling of solvent is required.
 - Avoids errors arising from incomplete sampling of solvent configurations.
- Dedicated sampling of solutes.
- Simplification of system setup.

The GB/SA solvation model

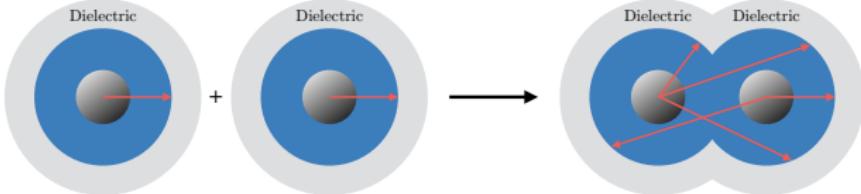
- Details: Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. (1990)¹

$$G_{\text{sol}} = G_{\text{cav}} + G_{\text{vdW}} + G_{\text{pol}}$$

$$G_{\text{cav}} + G_{\text{vdW}} = G_{\text{np}} = \sum_i \sigma_i S A_i$$

$$G_{\text{pol}} = -166.0 \left(1 - \frac{1}{\varepsilon}\right) \frac{q^2}{\alpha}$$

$$G_{\text{pol}} = -166.0 \left(1 - \frac{1}{\varepsilon}\right) \sum_i \sum_j \frac{q_i q_j}{r_{ij}^2 + \alpha_{ij}^2 \exp(-r_{ij}^2/2\alpha_{ij}^2)^{1/2}}$$

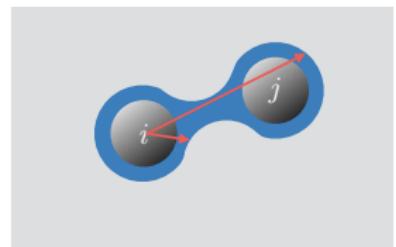
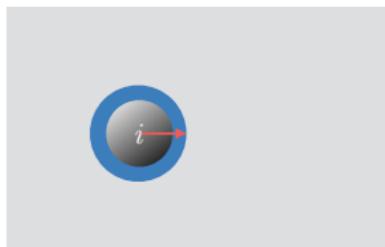
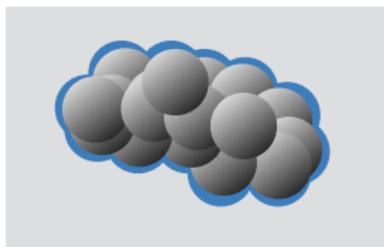


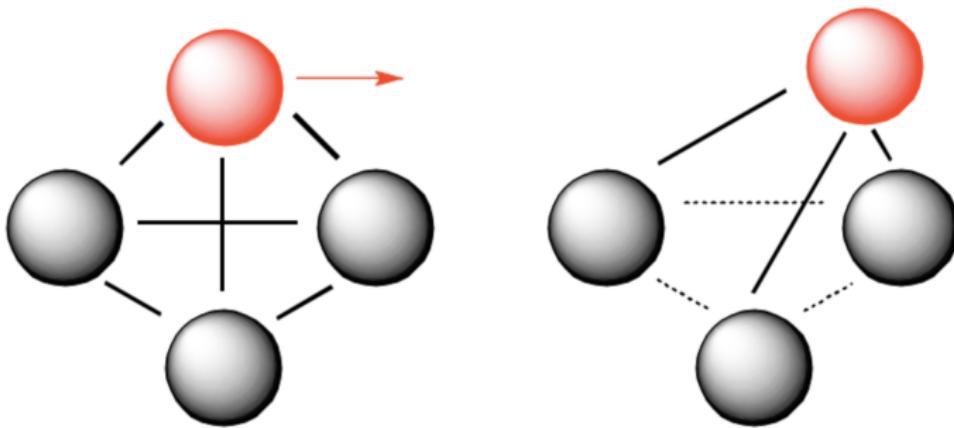
¹ J. Am. Chem. Soc. 1990, 112, 6127.

The GB/SA solvation model

- Details: Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. (1997)²

$$G'_{\text{pol},i} = \frac{-166.0}{R_{\text{vdW},i} + \phi + P_1} + \sum^{1,2} \frac{P_2 V_j}{r_{ij}^4} + \sum^{1,3} \frac{P_3 V_j}{r_{ij}^4} + \sum^{1,\geq 4} \frac{P_4 V_j CCF}{r_{ij}^4}$$





- The generalized Born energy is not pairwise decomposable:
 - The Born radius of any given atom depends on the volume and position of every other atom in the system.
 - All atom-pair energies are affected by every move in a Monte Carlo simulation, due to changes in Born radii of the pair's constituent atoms.
- Therefore, all atom-pair energies must be calculated after every move in a Monte Carlo simulation.

There is great appeal in employing implicit solvents in simulations of large GB/SA solvation in protein simulations biomolecular systems, and many have done so in:

- RNA hairpin unfolding³ (MD)
- Nucleic acid conformational dynamics⁴ (MD)
- Determination of free-energy surfaces in peptides⁵ (MD)
- Potential application of Monte Carlo sampling algorithms:
 - Flexible docking⁶
 - Concerted rotation with angles⁷

Few have exploited GB/SA in rigorous Monte Carlo free-energy calculations of binding affinities. Why not?

³ *J. Mol. Biol.* **2002**, *317*, 493.

⁴ *Biophys. J.* **2003**, *85*, 790.

⁵ *Proteins* **2004**, *56*, 310

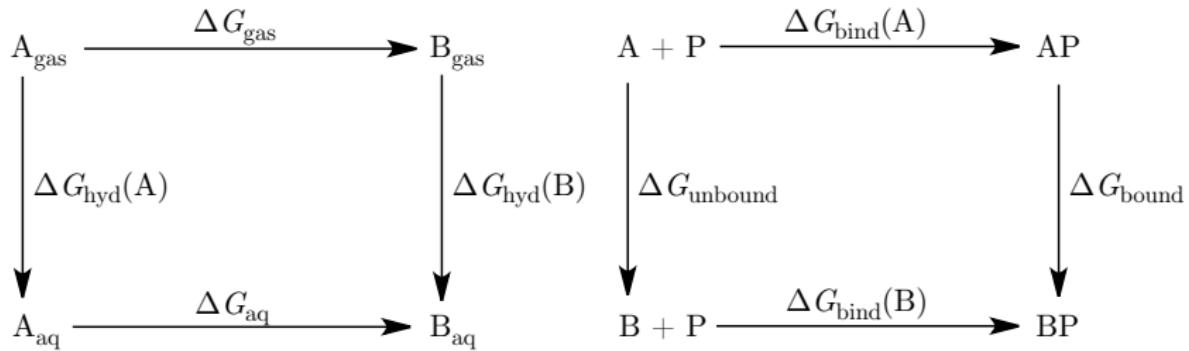
⁶ *J. Comput. Chem.* **2004**, *24*, 1637.

⁷ *J. Chem. Phys.* **2003**, *118*, 4261.

Monte Carlo free-energy perturbation

From perturbation theory, FEP is the methodology of choice for computation of relative free energies of binding for different ligands to a protein. Zwanzig⁸ in 1954:

$$\Delta G(A \rightarrow B) = G_B - G_A = -k_B T \ln \left\langle \exp \left(-\frac{E_B - E_A}{k_B T} \right) \right\rangle_A$$



Perturbations are split into small increments.

3 energy calculations are performed for each accepted move...

⁸ J. Chem. Phys. 1954, 22, 1420–1426

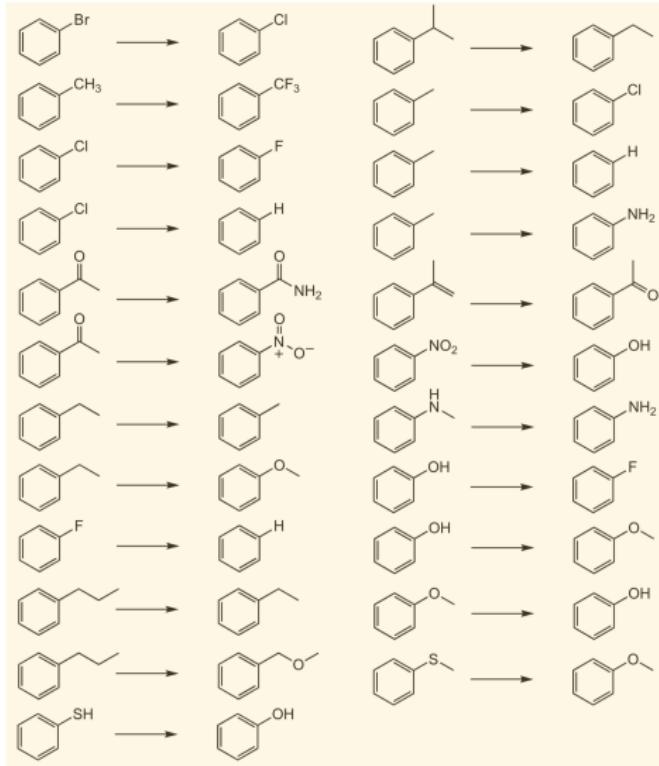
Our GB/SA implementation uses united-atom (UA) charges on saturated alkyl groups, but perturbations often require UA charges → AA charges or AA charges → UA charges.



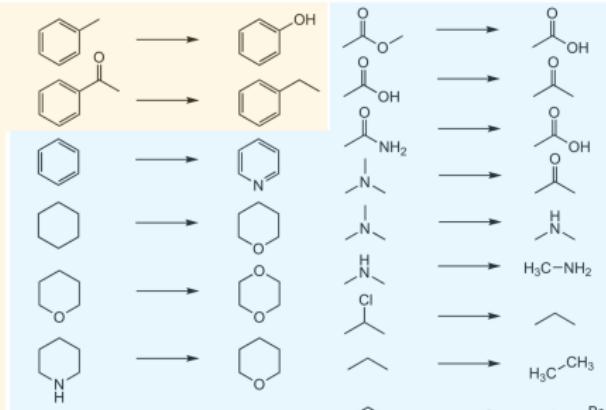
- Dummy atoms (SASA and Coulombic)
- 1,2 and 1,3 terms are ignored for hydrogens on saturated carbons
- Adjustment of volume due to atomic overlap can be problematic
- Non-bonded terms are ignored for all hydrogens
- Corrections to the initially assigned charges are made for N—O compounds

Test cases: typical perturbation

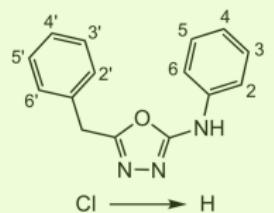
PhX to PhY



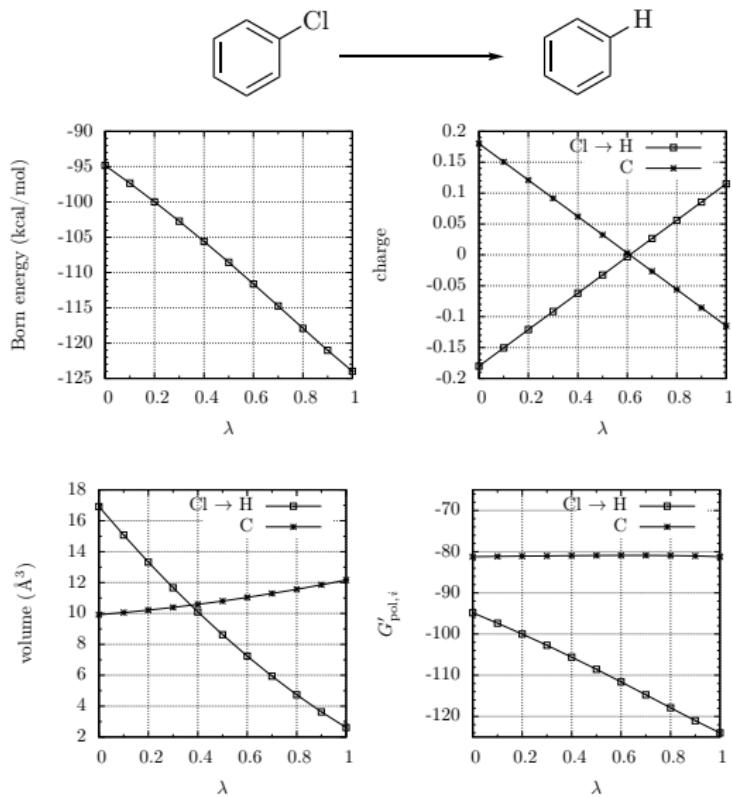
Small



Cl Scan



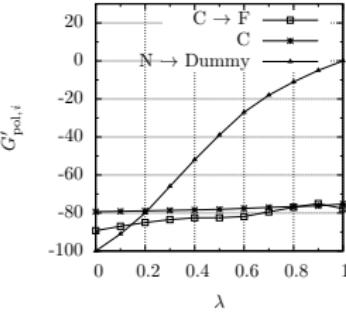
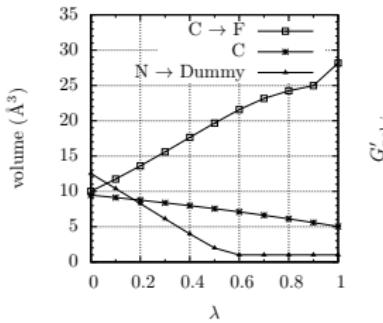
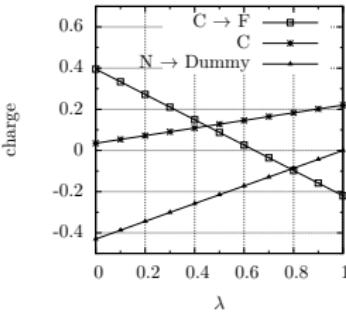
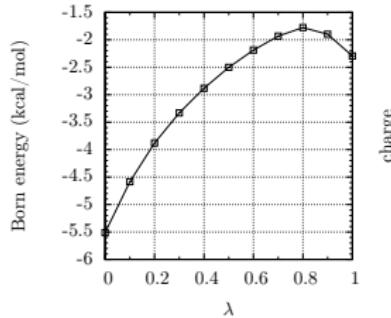
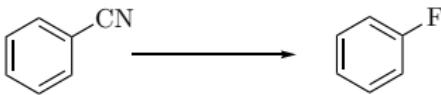
Energy components: chlorobenzene to benzene



The challenge: large volume change; significant overlap; reversal of charge distribution.

The results: smooth perturbation of charge and volume; volume of C increases due to loss of overlap from Cl/H; charge reversal yields correct charges on benzene.

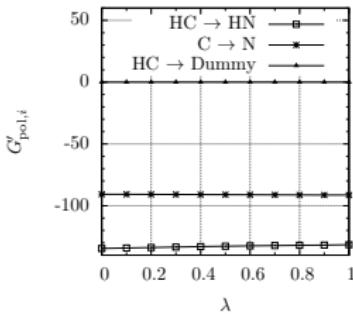
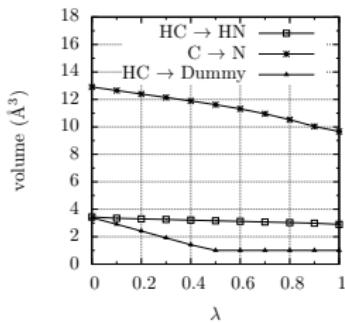
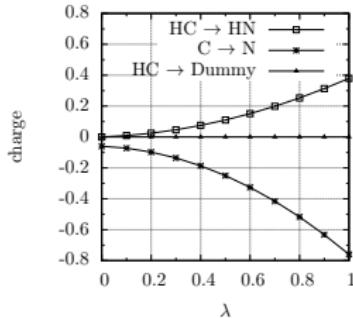
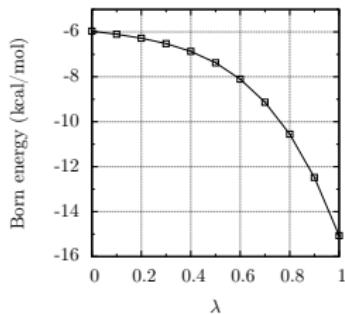
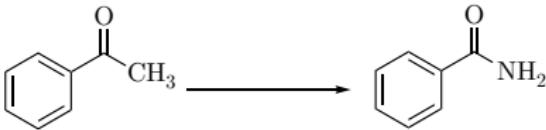
Energy components: cyanobenzene to fluorozene



The challenge: one dummy atom; significant overlap; reversal of charge distribution.

The results: smooth perturbation of charge; volume of C decreases due to emergence of F; dummy charge scales to zero, volume to zero by $\lambda = 0.6$ due to overlap.

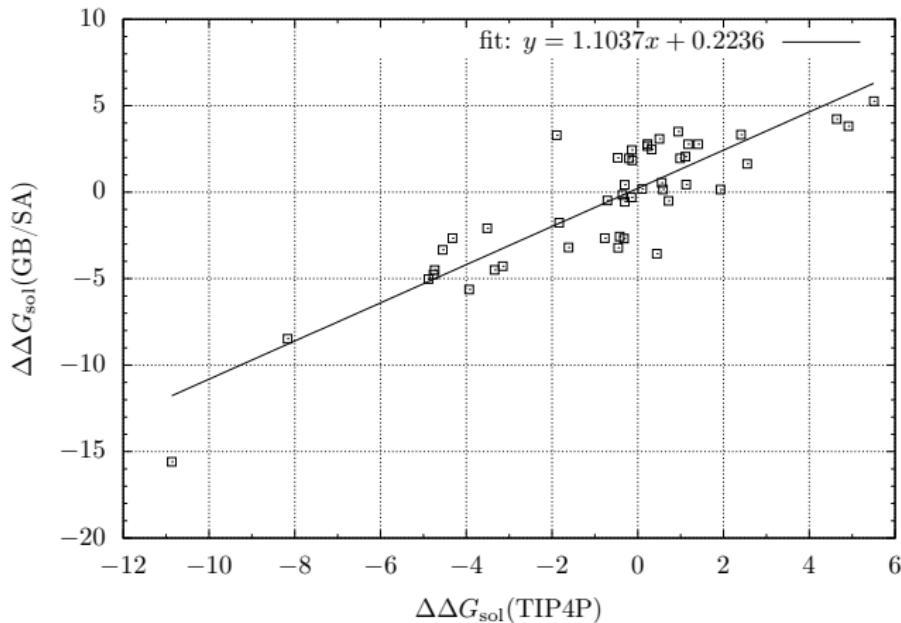
Energy components: acetophenone to benzamide



The challenge: three perturbations, one dummy atom; dispersion of charge; UA/AA conversion; slight geometry change.

The results: smooth dispersion of charge; dummy volume scales to zero by $\lambda = 0.5$ due to overlap, electrostatic contribution constant at zero.

Free energies of solvation: TIP4P and GB/SA



GB/SA trajectories behaved more like gas-phase trajectories than TIP4P trajectories; GB/SA and gas-phase trajectories typically followed a more predictable path, whereas TIP4P trajectories often fluctuated, owing to variations in water molecule configurations.

Approximated Generalized Born Potential

How can we overcome the limitations of a standard GB/SA implementation in MC simulations?

$$G'_{\text{pol},i} = \frac{-166.0}{R_{\text{vdW},i} + \phi + P_1} + \sum \frac{P_2 V_j}{r_{ij}^4} + \sum \frac{P_3 V_j}{r_{ij}^4} + \sum \frac{P_4 V_j \text{CCF}}{r_{ij}^4}$$

Gives α_i
(Born radius)

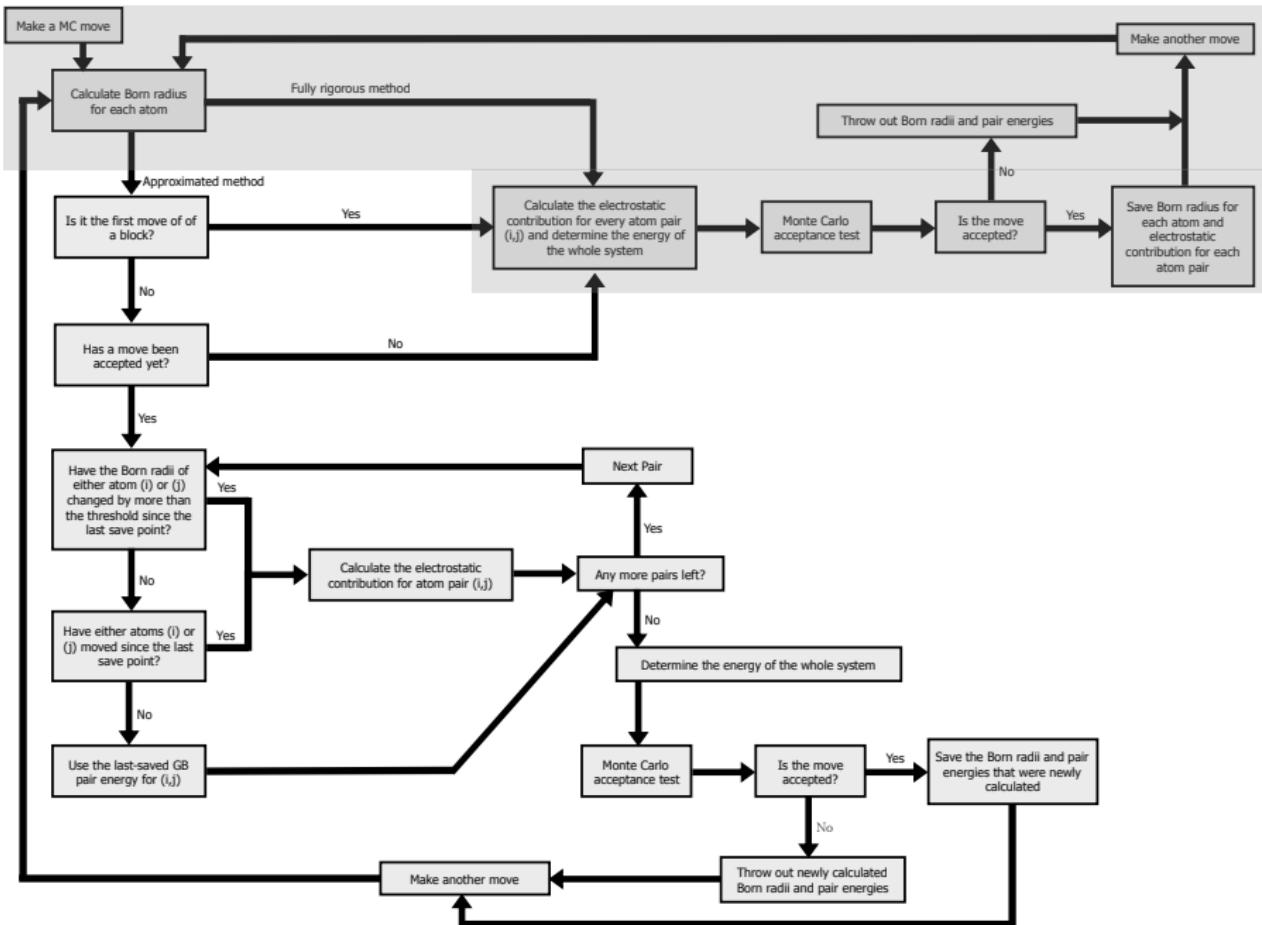
Calculated once in GETSIG and
GBSASETUP – only updated at the
start of a new block.

Calculated in CALCEGB
– performed for every move.

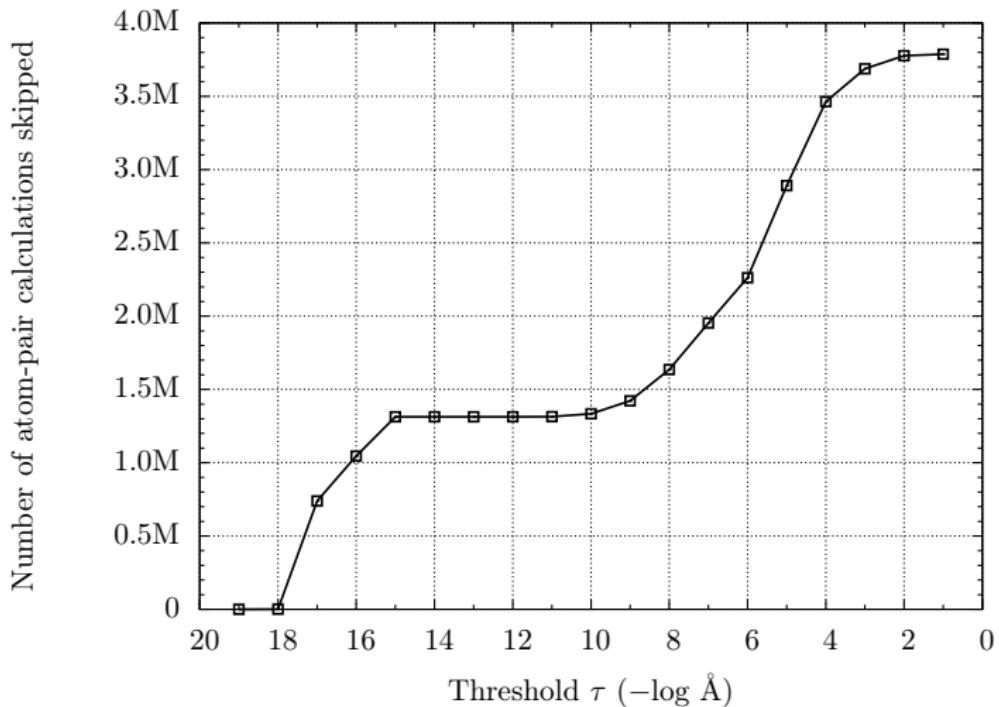
$$G_{\text{pol}} = -166.0 \left(1 - \frac{1}{\epsilon}\right) \sum_i \sum_j \frac{q_i q_j}{r_{ij}^2 + \alpha_{ij}^2 \exp(-r_{ij}^2/2\alpha_{ij}^2)^{1/2}}$$

How can we overcome the limitations of a standard GB/SA implementation in MC simulations?

- Assumption: the impact of a moving atom on the Born radius of a distant atom is small.
- Don't recalculate atom-pair energy after a move if:
 - ① neither atom moved; and
 - ② the Born radius of neither atom changed by more than a pre-specified amount (threshold) since the last accepted move.
- A large number of atom-pair energy calculations would be skipped
- Should have a minimal effect on the resulting ensemble of energies.

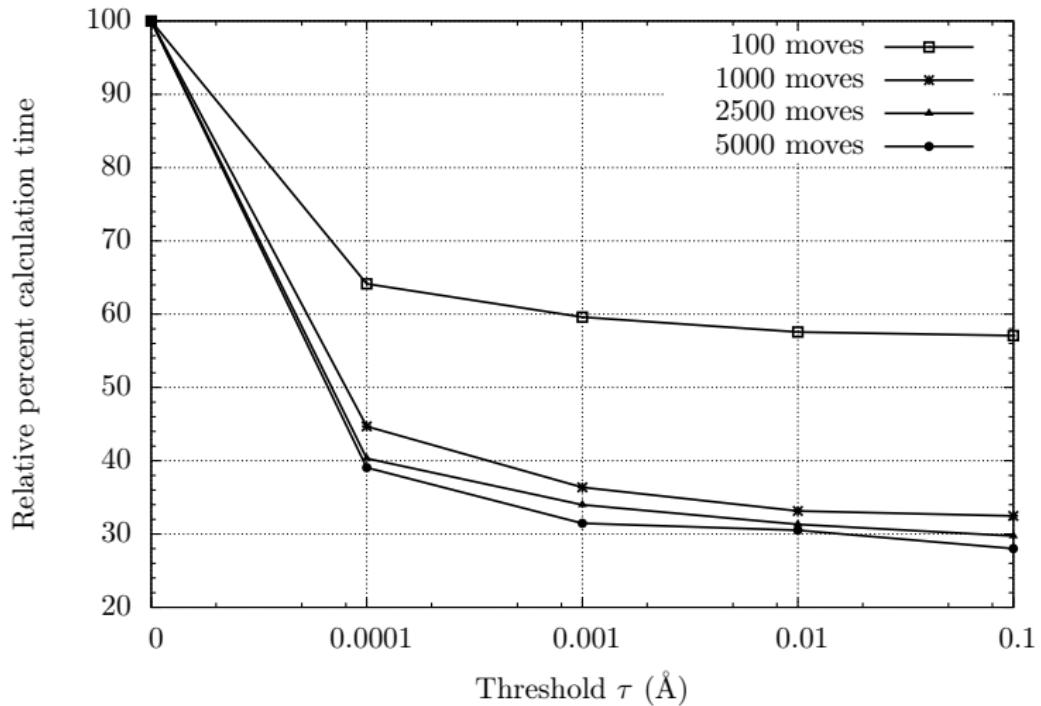


The impact of the threshold



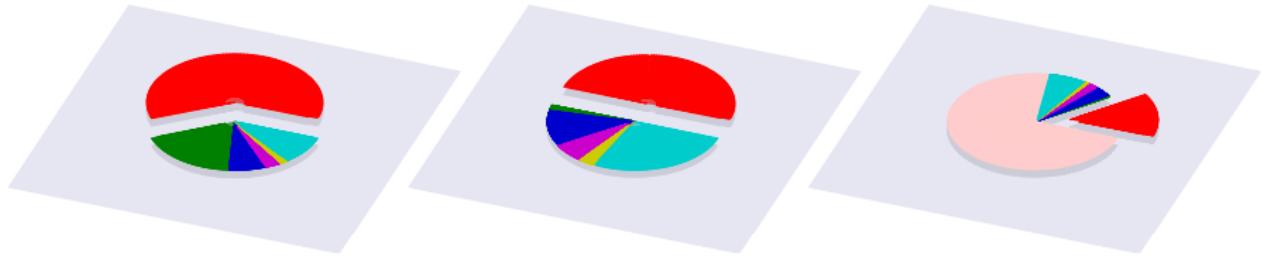
- Response curve is system specific but trivially generated.
- By $\tau = 0.005 \text{ \AA}$, most pairs (95%+) were being skipped.

Improved efficiency



- Most speed-up is achieved between threshold values of 0.001 and 0.01 Å
- Little additional benefit is afforded at larger values.

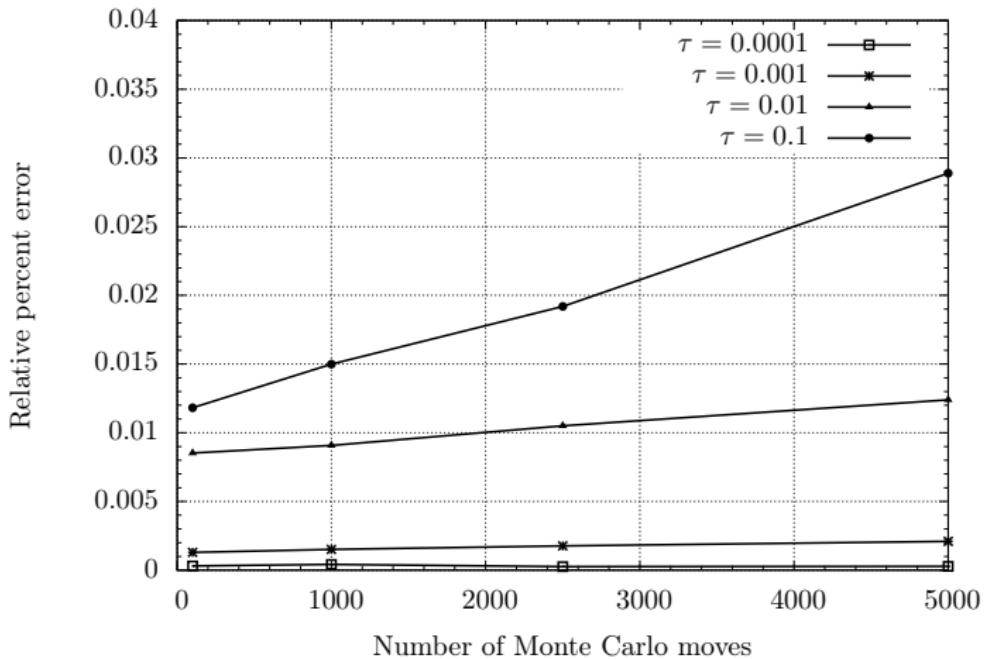
Improved efficiency



No threshold With threshold Relative
red = **CALCEGB**, green = **EXPJ**, cyan = **READ**, pink = time saved

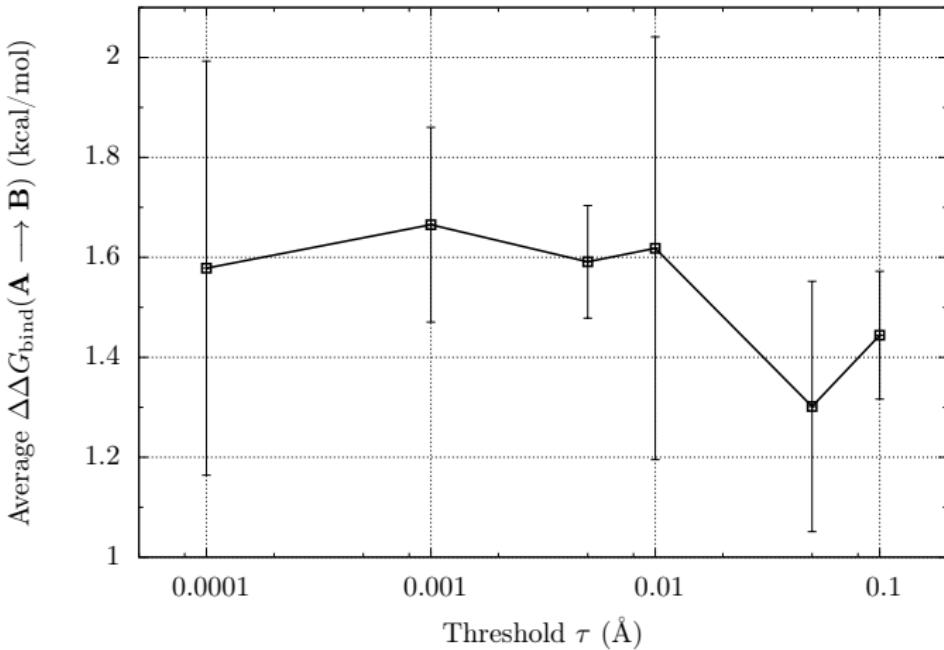
- With $\tau = 0.1 \text{ \AA}$
- Throughout the process of implementing the new code, we monitored performance statistics in order to gauge our progress and to identify bottlenecks in the code.
- This also proved valuable in determining which coding methods were most efficient in cases where several possible coding methods were available.

Drift: accumulation of error



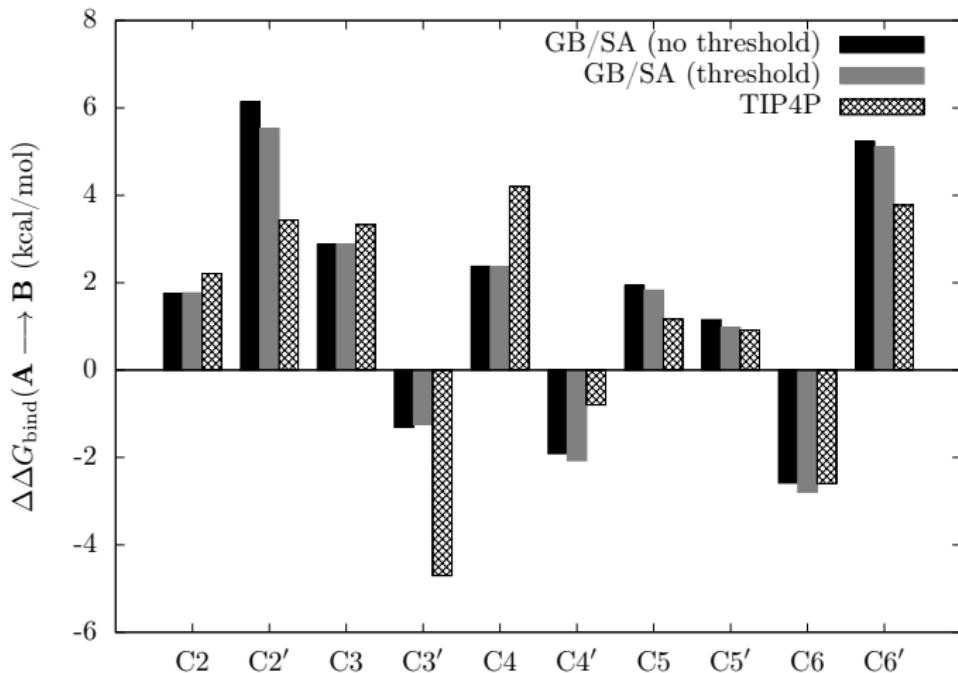
- Drift is well controlled at threshold values below 0.01 \AA
- Suggests values of 0.001 or 0.005 \AA to be the best compromise between speed and accuracy.

Statistical impact of the approximation



- Any error introduced by the approximation is below the statistical noise of the MC procedure.
- No statistical dependence was found on the threshold.

Chlorine scan



- Qualitative agreement between GB/SA and TIP4P
- GB/SA predicted the same substitution pattern as TIP4P, which produced a highly active compound. (Leung et al, 2010).

Summary

- It was of interest to be able to employ GB/SA solvation in our efforts towards computer-aided drug design using FEP while maintaining computational efficiency.
- FEP was successfully implemented within the GB/SA protocol: test perturbations revealed smooth trajectories of energies and energy components; free energies of solvation with TIP4P were well reproduced with GB/SA, but long simulations of large systems were too slow for practical use.
- An approximated generalized Born potential was implemented wherein atom pair energies are only updated after a Monte Carlo move if their Born radii have changed by more than a specified threshold since the last accepted move. By skipping a large number of energy pair calculations, significant speedup was achieved with minimal loss of accuracy.
- A chlorine scan used in our search for NNRTIs was performed using our FEP/GBSA protocol and yield results qualitatively similar to those with TIP4P, correctly predicting an active substitution pattern.

Acknowledgments



Professor William
Jorgensen



Dr. Julian
Tirado-Rives



Paty
Morales de Tirado



Dr. Julien Michel



Dr. Laura DeFeo



Dr. Sara Nichols



Jorgensen Lab Members

Alissa Hare
Jakub Kostal
Cheryl Leung
Ricardo Gallardo-Macias
Dr. Jose Cisneros Trigo
Dr. Markus Dahlgren
Dr. Anil Ekkati
Dr. Mariela Bollini
Dr. Patric Schyman

Family & Friends

Phil & Lynda Terhorst
Erin & Kevin McAnoy
Paul & Vicki Terhorst
Dr. George Gardenier
Jeremi Szaniawski
Prof. Teresa Longin
Prof. David Soulsby
...and so many more

Committee Members

Prof. John Tully
Prof. Martin Saunders

