Improved Solvation Models using Boundary Integral Equations

Matthew Knepley and Jaydeep Bardhan

Computational and Applied Mathematics
Rice University

SIAM Conference on the Life Sciences
Minneapolis, MN August 9, 2018

Collaborators

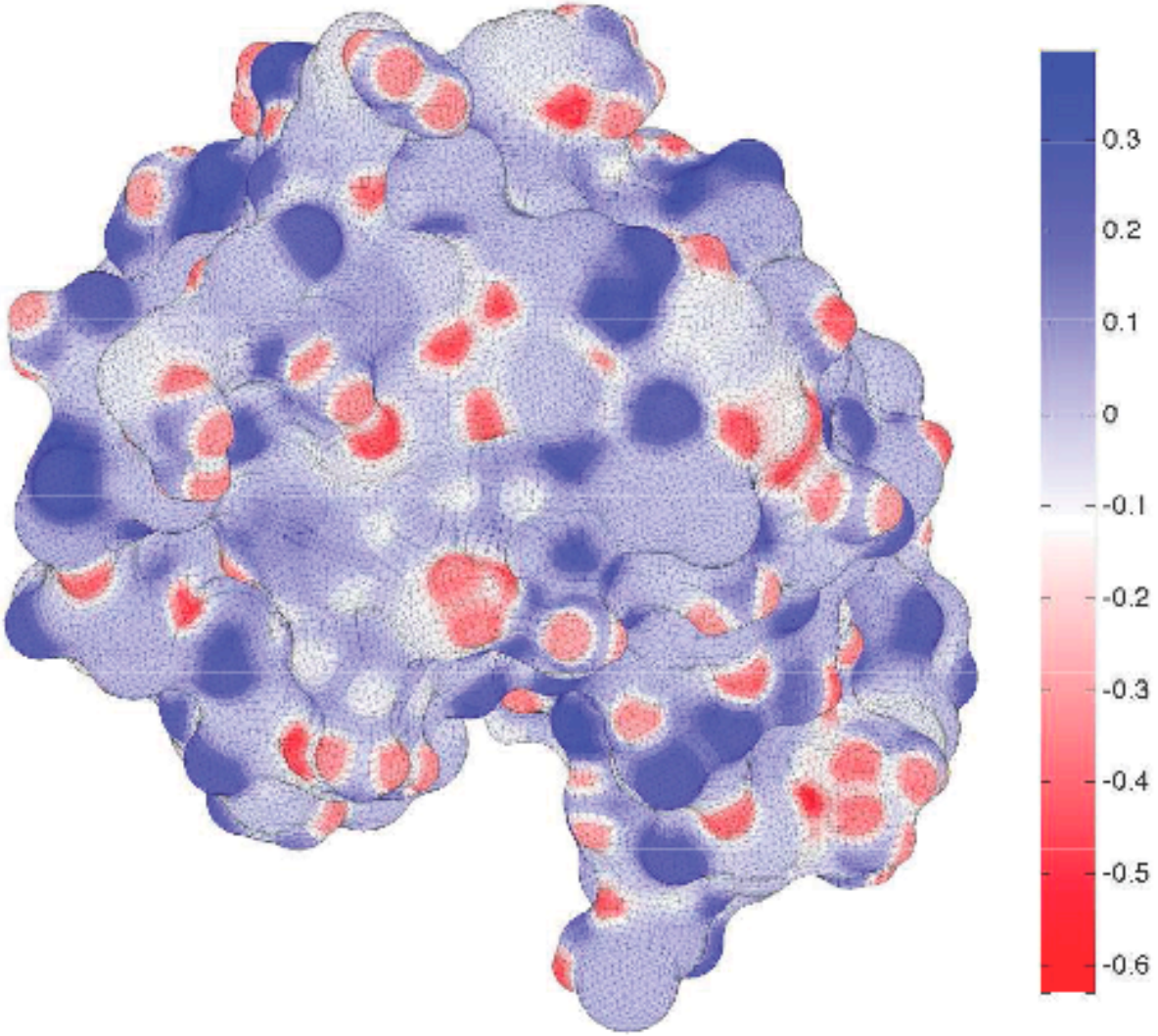
- Amir Molvai Tabrizi (postdoc, NE)
- Tom Klotz (grad student, Rice)
- Spencer Goossens (grad student, NE)
- Ali Rahimi (grad student, NE)

Main Point

Solvation computation
can benefit from
non-Poisson models.

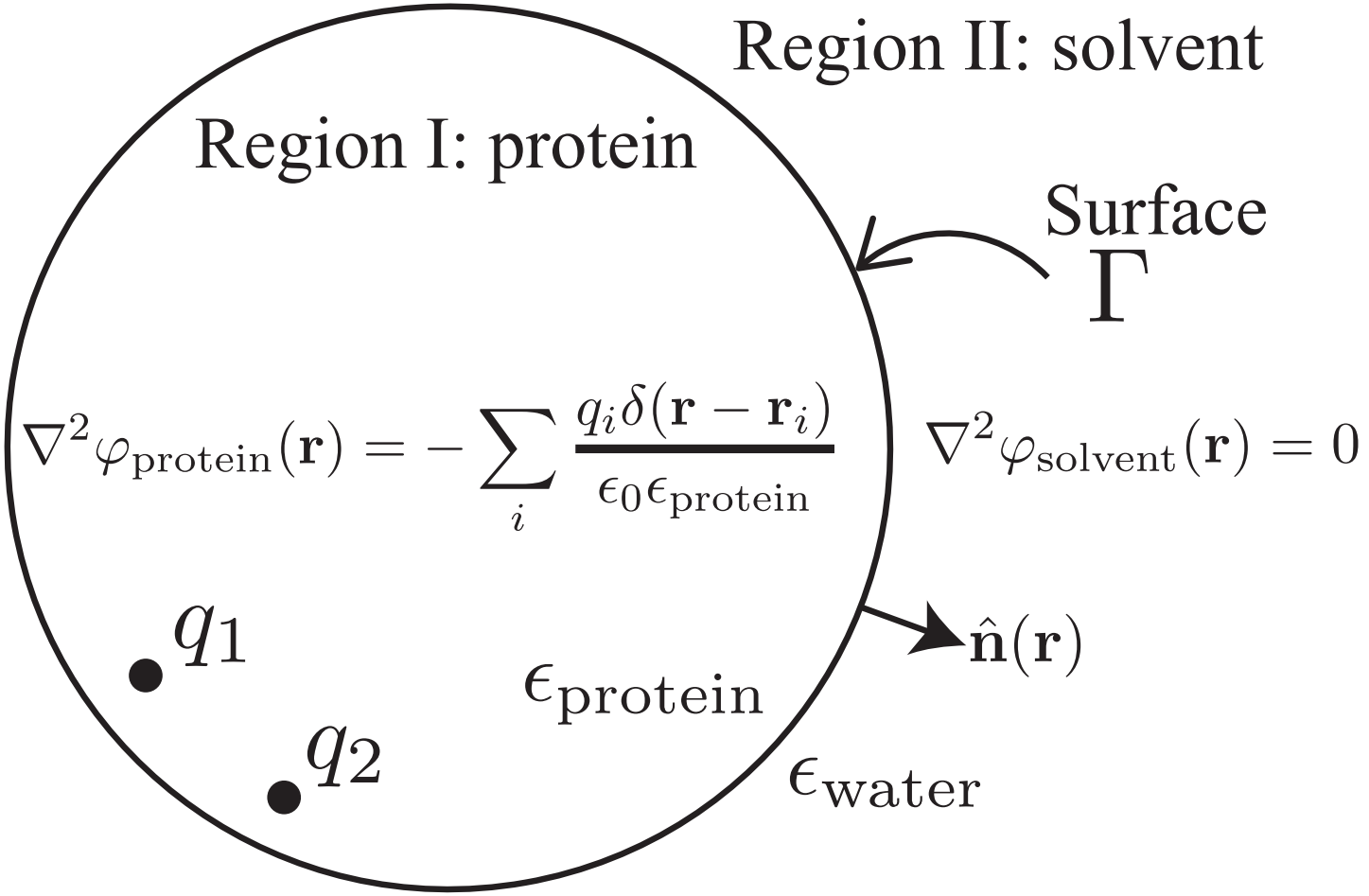
Bioelectrostatics

The Natural World



Induced Surface Charge on Lysozyme

Electrostatic Potential ϕ



We can write a Boundary Integral Equation (BIE) for the induced surface charge σ,

$$\sigma(\vec{r}) + \hat{\epsilon} \int_{\Gamma} \frac{\partial}{\partial n(\vec{r})} \frac{\sigma(\vec{r}')d^2\vec{r}'}{4\pi||\vec{r} - \vec{r}'||} = -\hat{\epsilon} \sum_{k=1}^{Q} \frac{\partial}{\partial n(\vec{r})} \frac{q_k}{4\pi||\vec{r} - \vec{r}_k||}$$

$$(\mathcal{I} + \hat{\epsilon}\mathcal{D}^*)\,\sigma(\vec{r}) =$$

where we define

$$\hat{\epsilon} = 2\frac{\epsilon_I - \epsilon_{II}}{\epsilon_I + \epsilon_{II}} < 0$$

This is equivalent to a PDE model for the potentials $\Phi_{I,II}$ in the two regions, and boundary conditions at the solute surface:

$$\epsilon_I \Delta \Phi_I = \sum_{k=1}^{Q} q_k \delta(\vec{r} - \vec{r}_k)$$
$$\epsilon_{II} \Delta \Phi_{II} = 0$$
$$\Phi_I|_{r=b} = \Phi_{II}|_{r=b}$$
$$\epsilon_I \frac{\partial \Phi_I}{\partial r}|_{r=b} = \epsilon_{II} \frac{\partial \Phi_{II}}{\partial r}|_{r=b}$$

Bioelectrostatics

Mathematical Model

The *reaction* potential is given by

$$\phi^R(\vec{r}) = \int_\Gamma \frac{\sigma(\vec{r}')d^2\vec{r}'}{4\pi\epsilon_1||\vec{r}-\vec{r}'||} = C\sigma$$

which defines G_{es}, the electrostatic part of the solvation free energy

$$\begin{aligned}\Delta G_{es} &= \frac{1}{2}\left\langle q, \phi^R \right\rangle \\ &= \frac{1}{2}\langle q, Lq\rangle \\ &= \frac{1}{2}\left\langle q, CA^{-1}Bq \right\rangle\end{aligned}$$

where

$$\begin{aligned}Bq &= -\hat{\epsilon}\int_\Omega \frac{\partial}{\partial n(\vec{r})}\frac{q(\vec{r}')d^3\vec{r}'}{4\pi||\vec{r}-\vec{r}'||} \\ A\sigma &= \mathcal{I} + \hat{\epsilon}\mathcal{D}^*\end{aligned}$$

Outline

Generalized Born Approximation

The pairwise energy between charges is defined by the *Still equation*:

$$G_{es}^{ij} = \frac{1}{8\pi}\left(\frac{1}{\epsilon_{II}} - \frac{1}{\epsilon_I}\right)\sum_{i,j}^{N}\frac{q_i q_j}{r_{ij}^2 + R_i R_j e^{-r_{ij}^2/4R_iR_j}}$$

where the *effective Born radius* is

$$R_i = \frac{1}{8\pi}\left(\frac{1}{\epsilon_{II}} - \frac{1}{\epsilon_I}\right)\frac{1}{E_i}$$

where E_i is the *self-energy* of the charge q_i, the electrostatic energy when atom i has unit charge and all others are neutral.

GB Problems

- No global potential solution, only energy
- No analysis of the error
  - For example, Salsbury 2006 consists of parameter tuning
- No path for systematic improvement
  - For example, Sigalov 2006 changes the model
- The same atoms have different radii in different
  - molecules,
  - solvents
  - temperatures
- LOTS of parameters
  - Nina, Beglov, Roux 1997

GB Problems

- No global potential
- No analysis of the
  - For example, Salsbury
- No path for system
  - For example, Sigalov 2
- The same atoms h rent
  - molecules,
  - solvents
  - temperatures
- LOTS of parameter
  - Nina, Beglov, Roux 19

TABLE 2: Atomic Born Radii Derived from Solvent Electrostatic Charge Distribution Tested with Free Energy Perturbation Methods in an Explicit Solvent[a]

atom	radius (Å)	
	Backbone	
C	2.04	carbonyl C, peptide backbone
O	1.52	carbonyl oxygen
	2.23	peptide nitrogen
CA	2.86	all CA except Gly
CA	2.38	Gly only
	Hydrogens	
H*	0.00	all hydrogens
	Side Chains	
CB	2.67	all residues
CG*	2.46	Val, Ile, Arg, Lys, Met, Phe, Thr, Trp, Gln, Glu
CD*	2.44	Ile, Leu, Arg, Lys
CD, CG	1.98	Asp, Glu, Asn, Gln
CB, CG, CD	1.98	Pro only
CE*, CD*, CZ,	2.00	Tyr, Phe rings
CE*, CD*, CZ*, CH2	1.78	Trp ring only
CE	2.10	Met only
CZ, CE	2.80	Arg, Lys
OE*, OD*	1.42	Glu, Asp, Asn, Gln
OG*	1.64	Ser, Thr
OH	1.85	Tyr only
NH*, NE, NZ	2.13	Arg, Lys
NE2, ND2	2.15	Gln, Asn
NE2, ND1	2.31	His only
NE1	2.40	Trp
S*	2.00	Met, Cys

[a] Patches N-term and C-term CAT, CAY: 2.06 Å. CY: 2.04 Å. OY: 1.52 Å. NT: 2.23 Å. * refers to a wild card character.

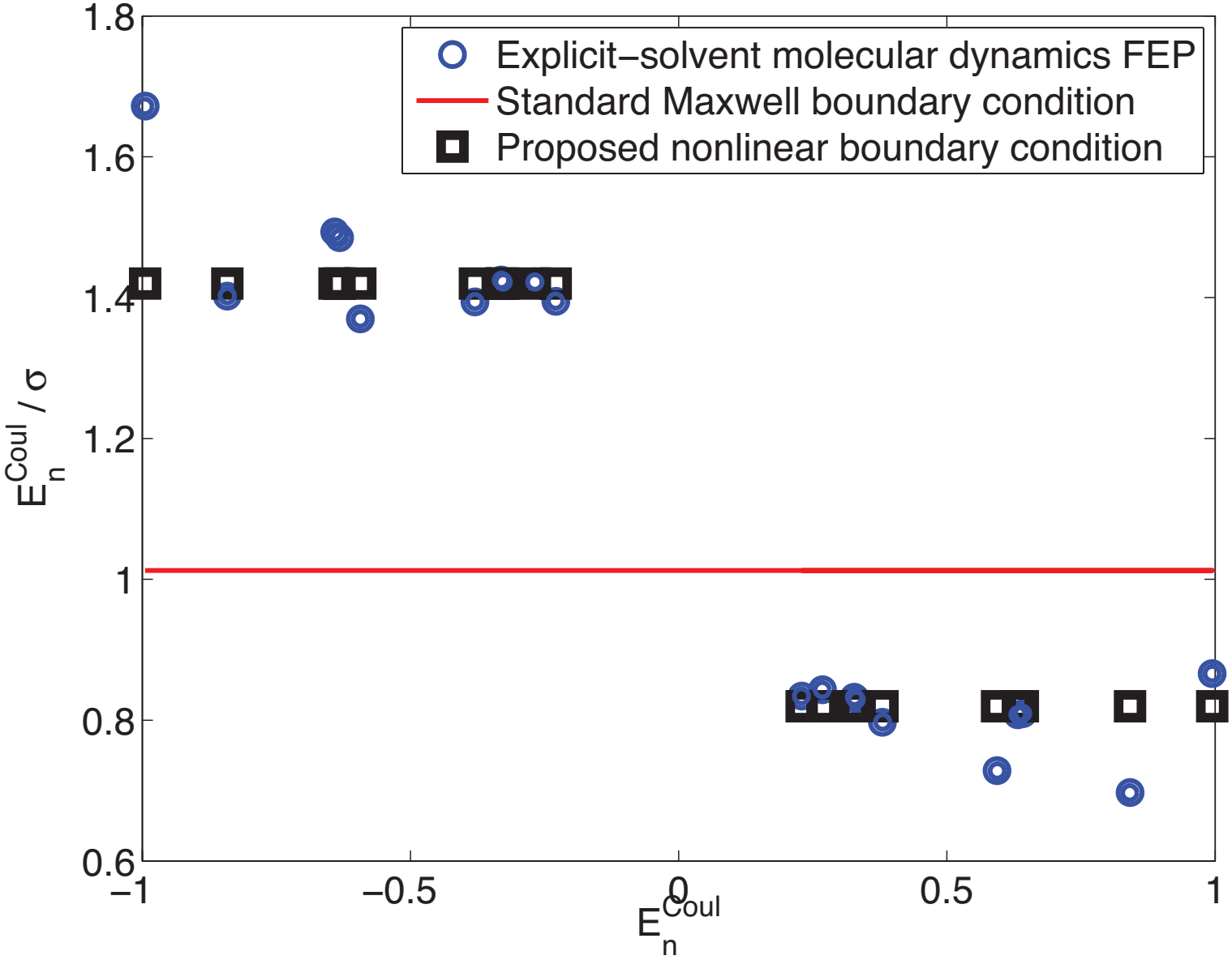
Implicit Solvent Models

State-of-the-art solvation models still use the same variation in radii

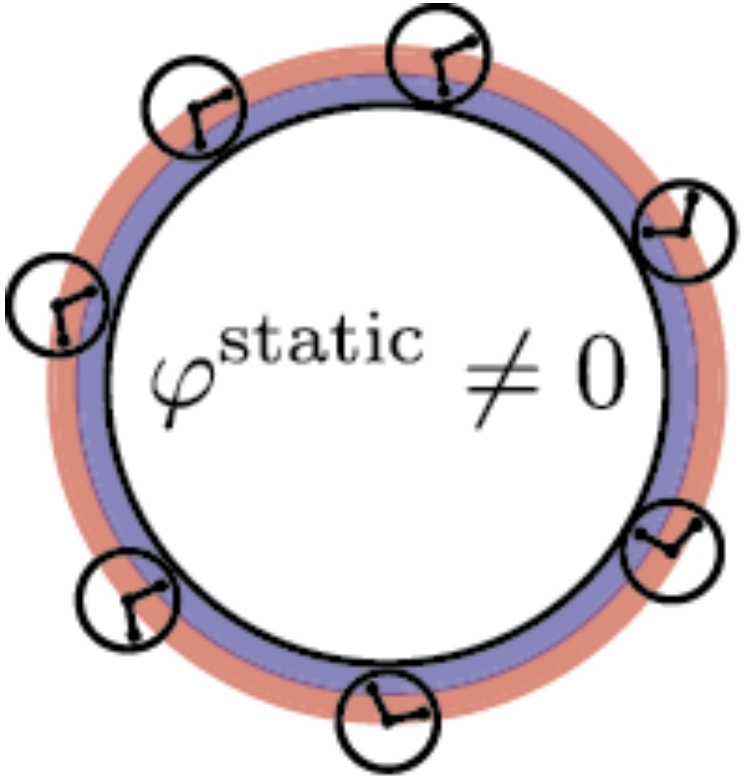
Biomolecular electrostatics — I want your solvation (model), J. Bardhan, Comp. Sci. & Disc., **5**(1), (2012)

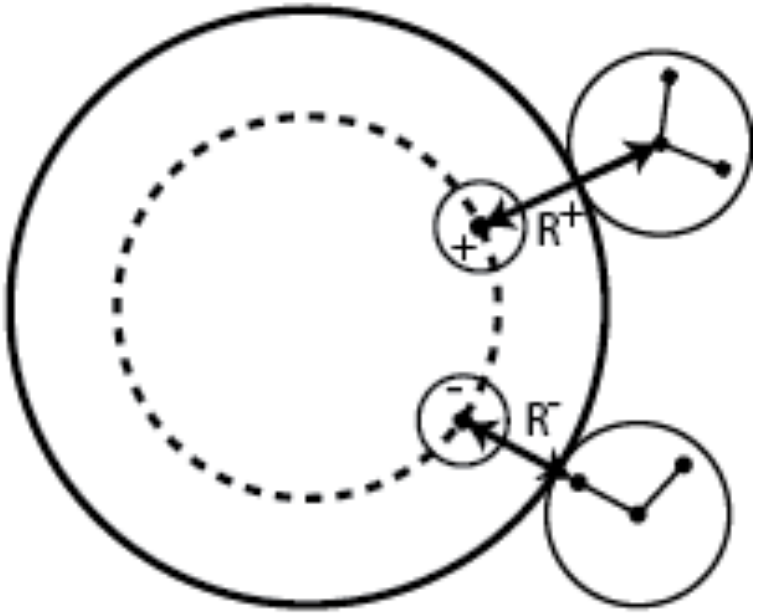
Outline

Origins of Electrostatic Asymmetry

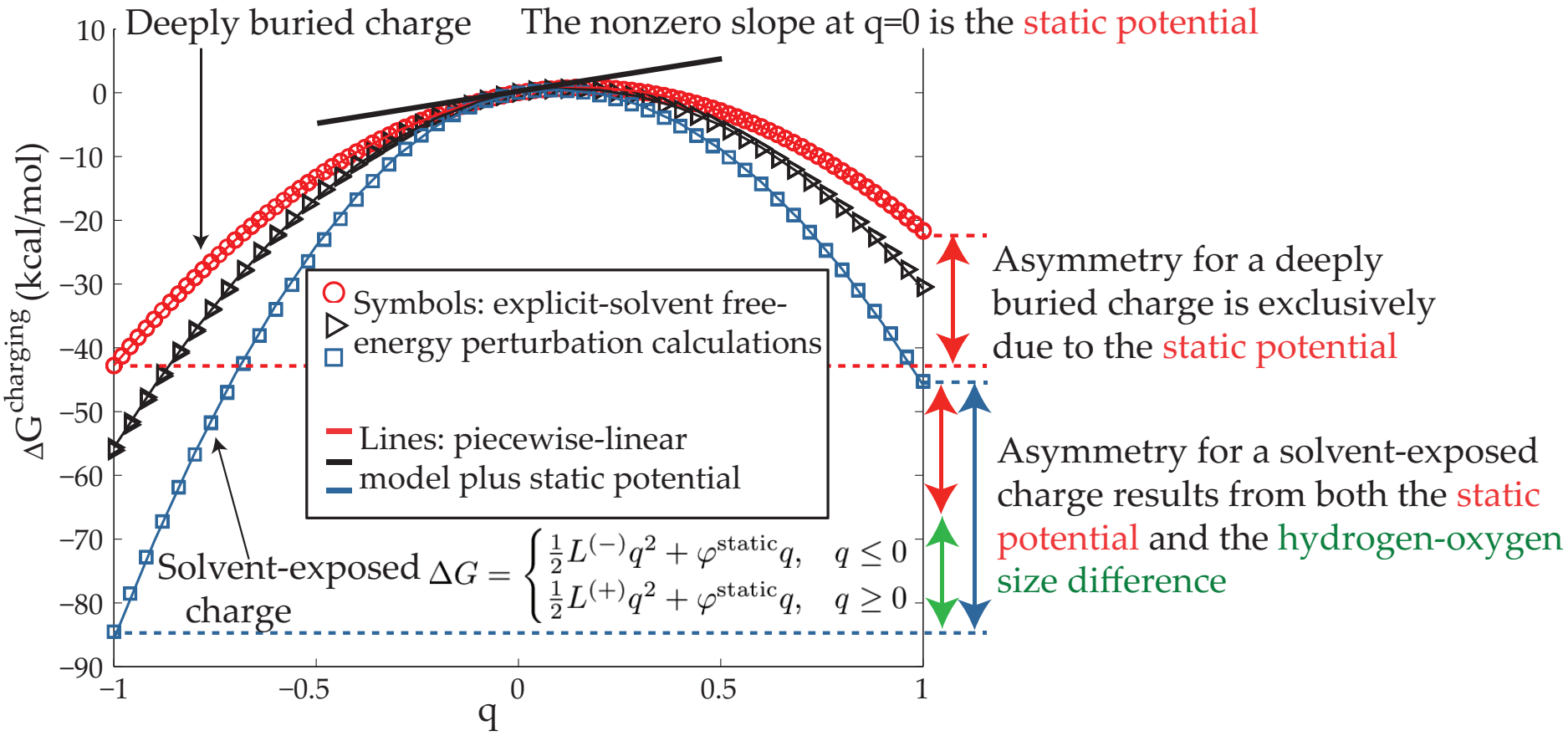


Origins of Electrostatic Asymmetry





Origins of Electrostatic Asymmetry



Main Idea

Maxwell Boundary Condition

Assume the model and energy, and derive the radii,

$$\epsilon_I \frac{\partial \Phi_I}{\partial n} = \epsilon_{II} \frac{\partial \Phi_{II}}{\partial n}$$

Main Idea

Solvation-Layer Interface Condition (SLIC)

Assume the energy and radii, and derive the model.

$$(\epsilon_I - \Delta\epsilon h(E_n))\frac{\partial \Phi_I}{\partial n} = (\epsilon_{II} - \Delta\epsilon h(E_n))\frac{\partial \Phi_{II}}{\partial n}$$

Main Idea

Using our correspondence with the BIE form,

$$\left(\mathcal{I} + h(E_n) + \hat{\epsilon}\left(-\frac{1}{2}\mathcal{I} + \mathcal{D}^*\right)\right)\sigma = \hat{\epsilon}\sum_{k=1}^{Q}\frac{\partial G}{\partial n}$$

where h is a diagonal nonlinear integral operator.

SLIC

Boundary Perturbation

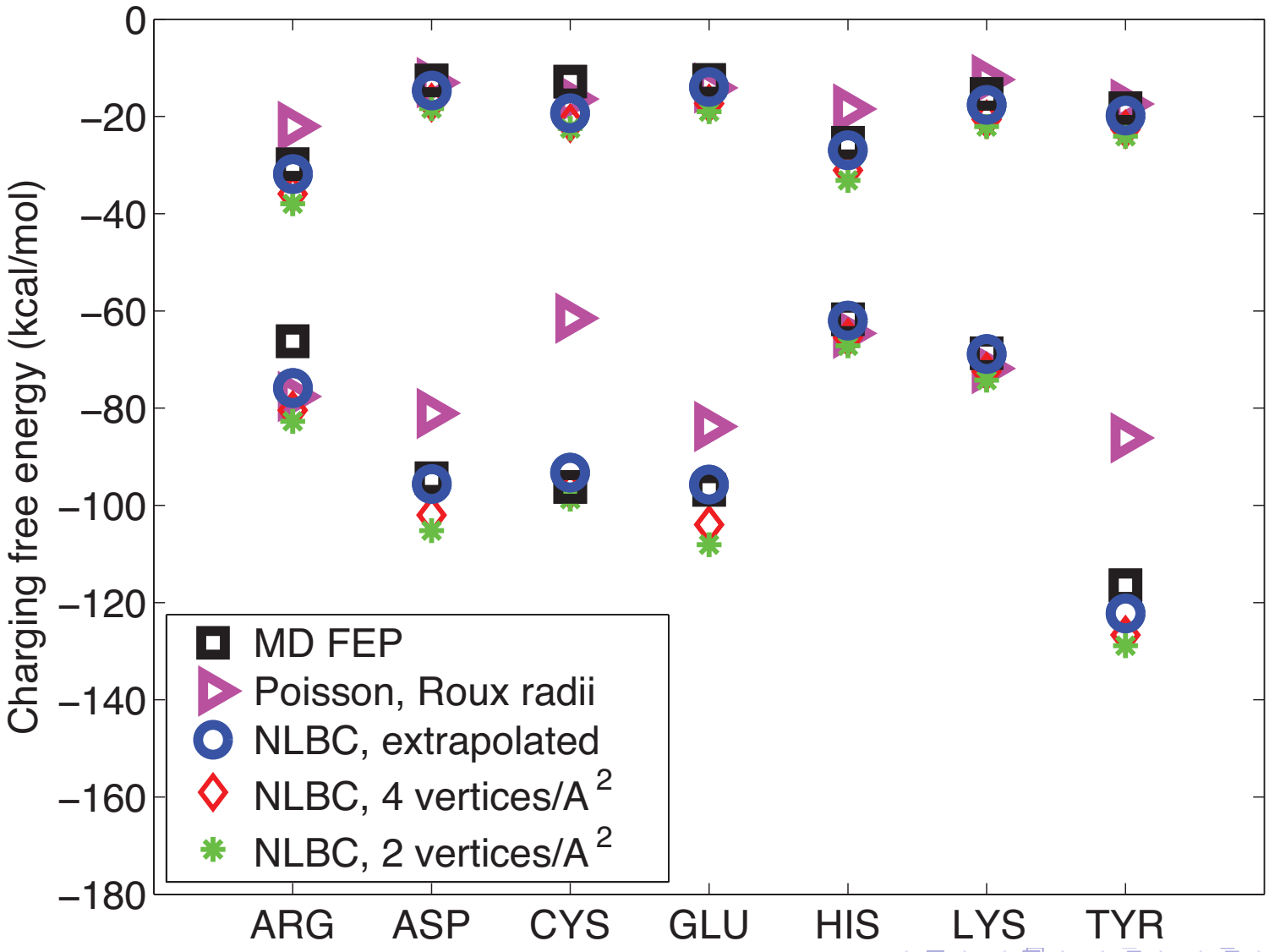
$$h(E_n) = \alpha \tanh\left(\beta E_n - \gamma\right) + \mu$$

where

- α Size of the asymmetry
- β Width of the transition region
- γ The transition field strength
- μ Assures $h(0) = 0$, so $\mu = -\alpha \tanh(-\gamma)$

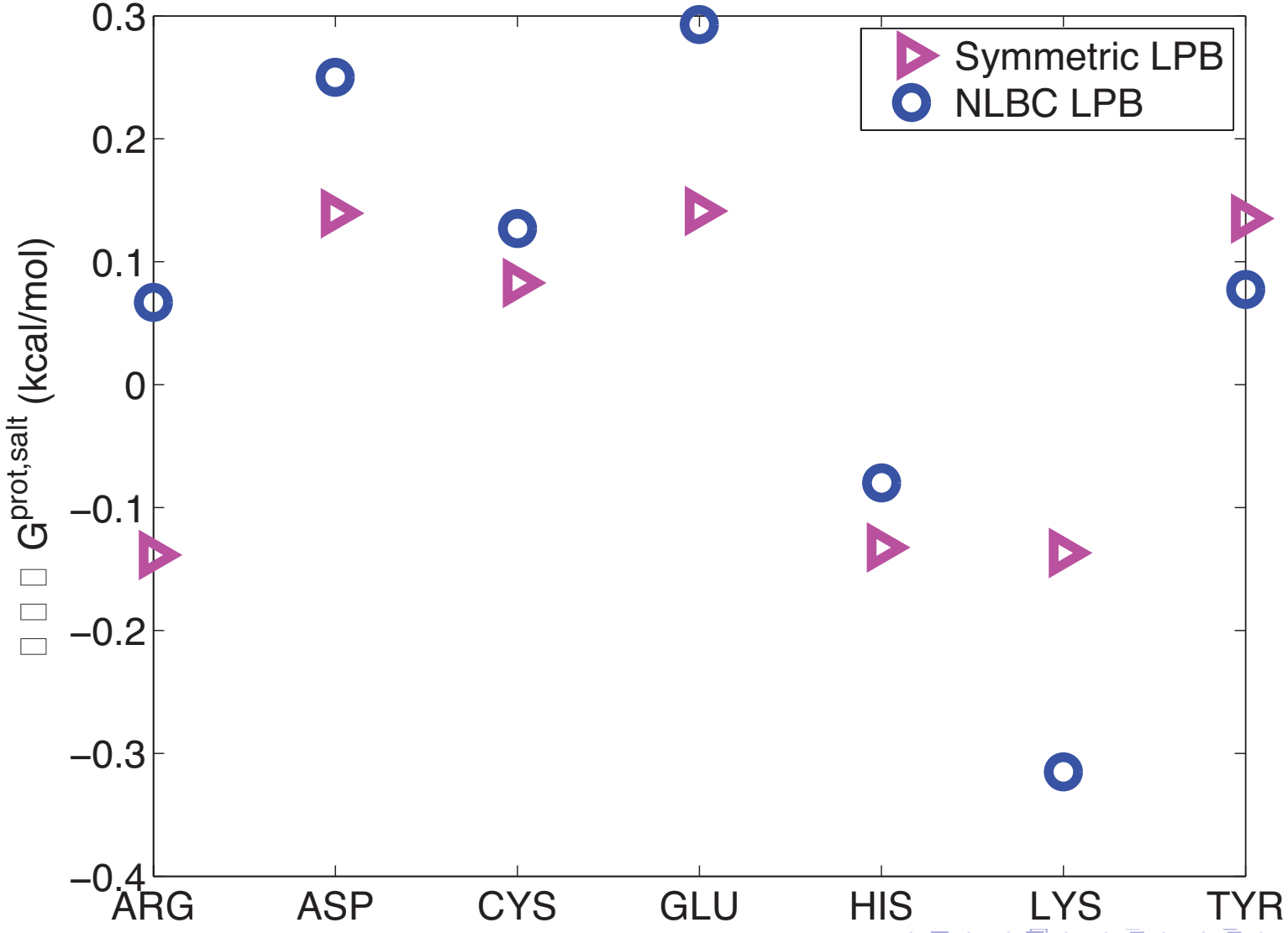
Accuracy of SLIC

Residues



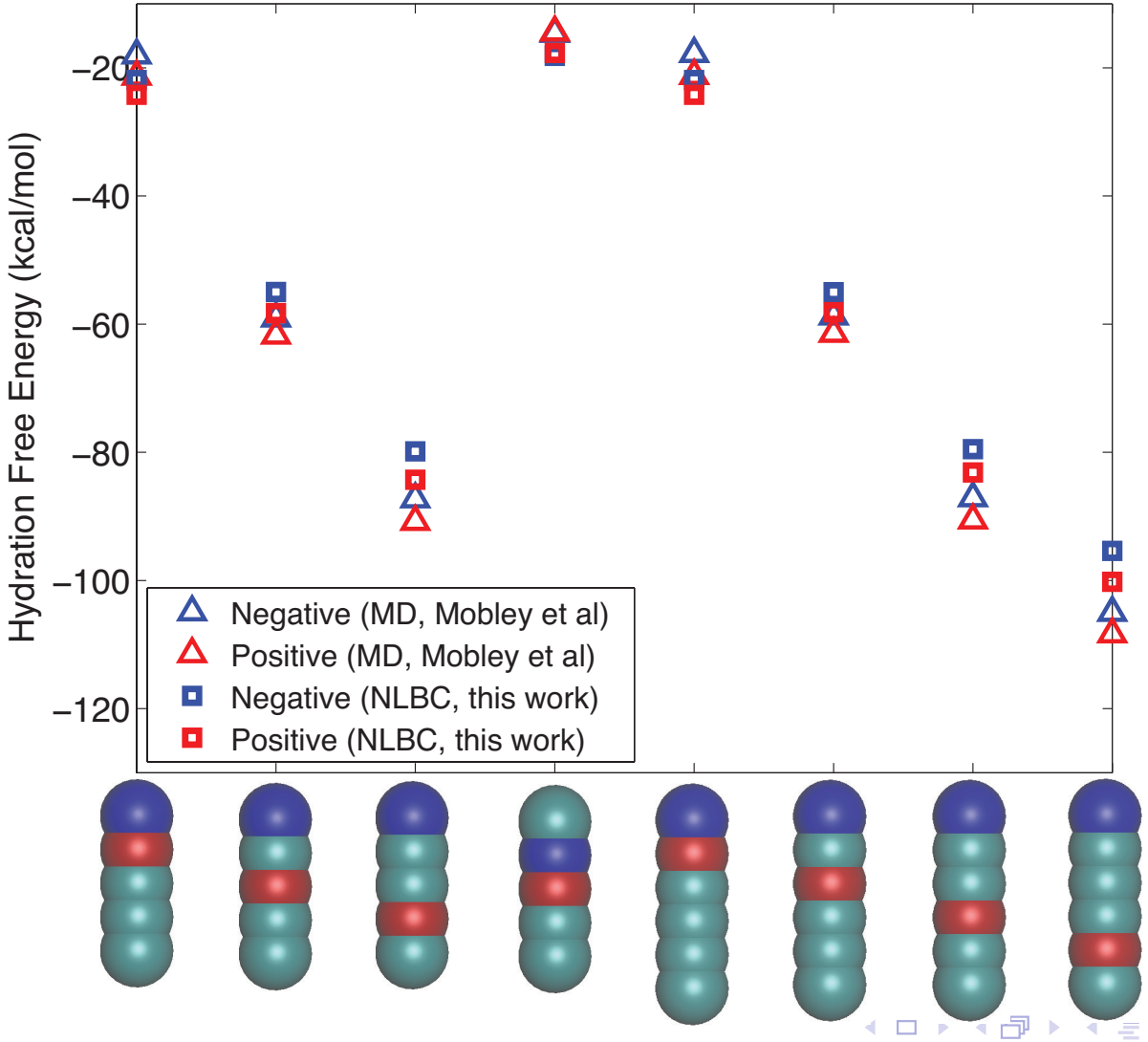
Accuracy of SLIC

Protonation



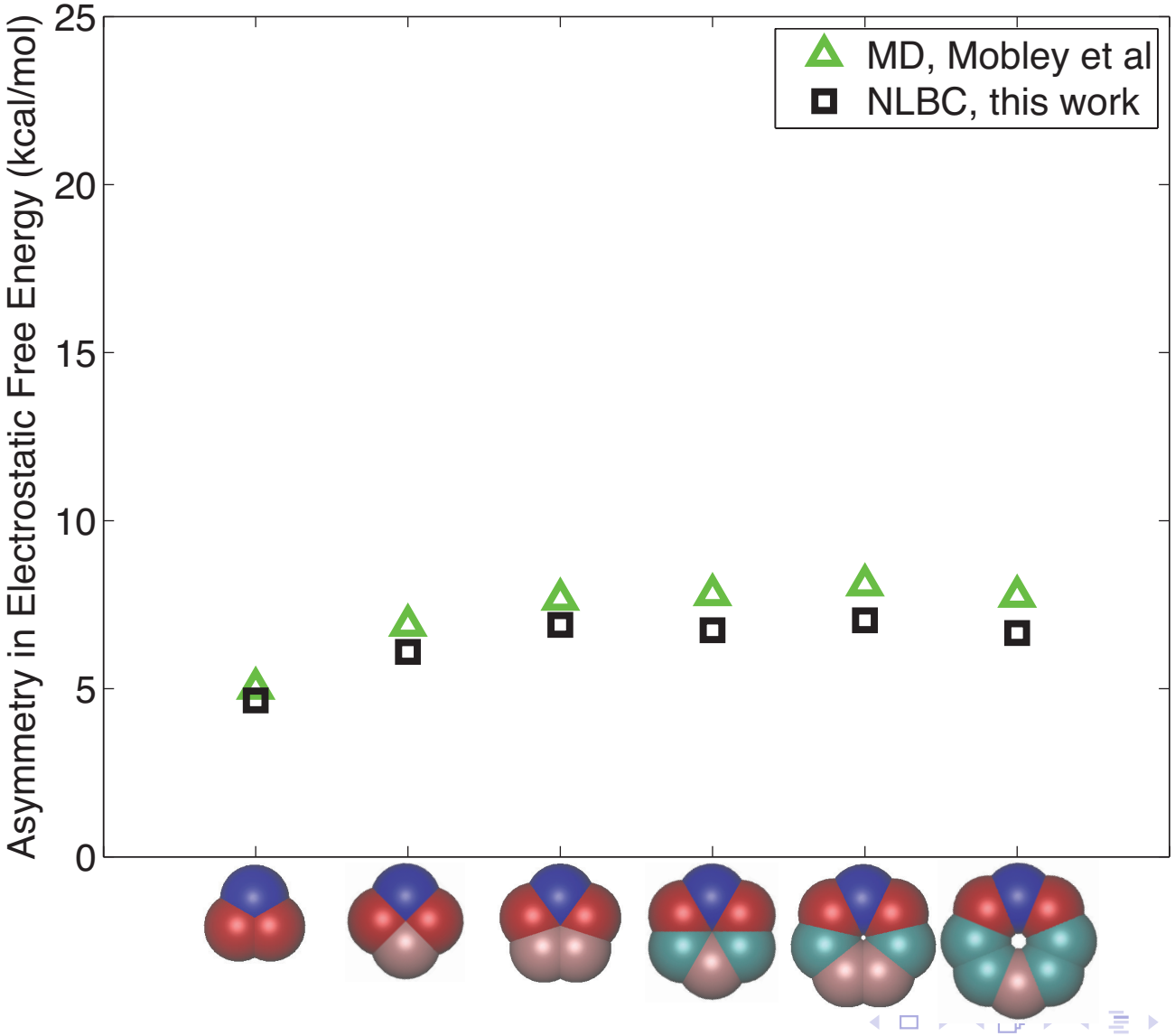
Accuracy of SLIC

Synthetic Molecules



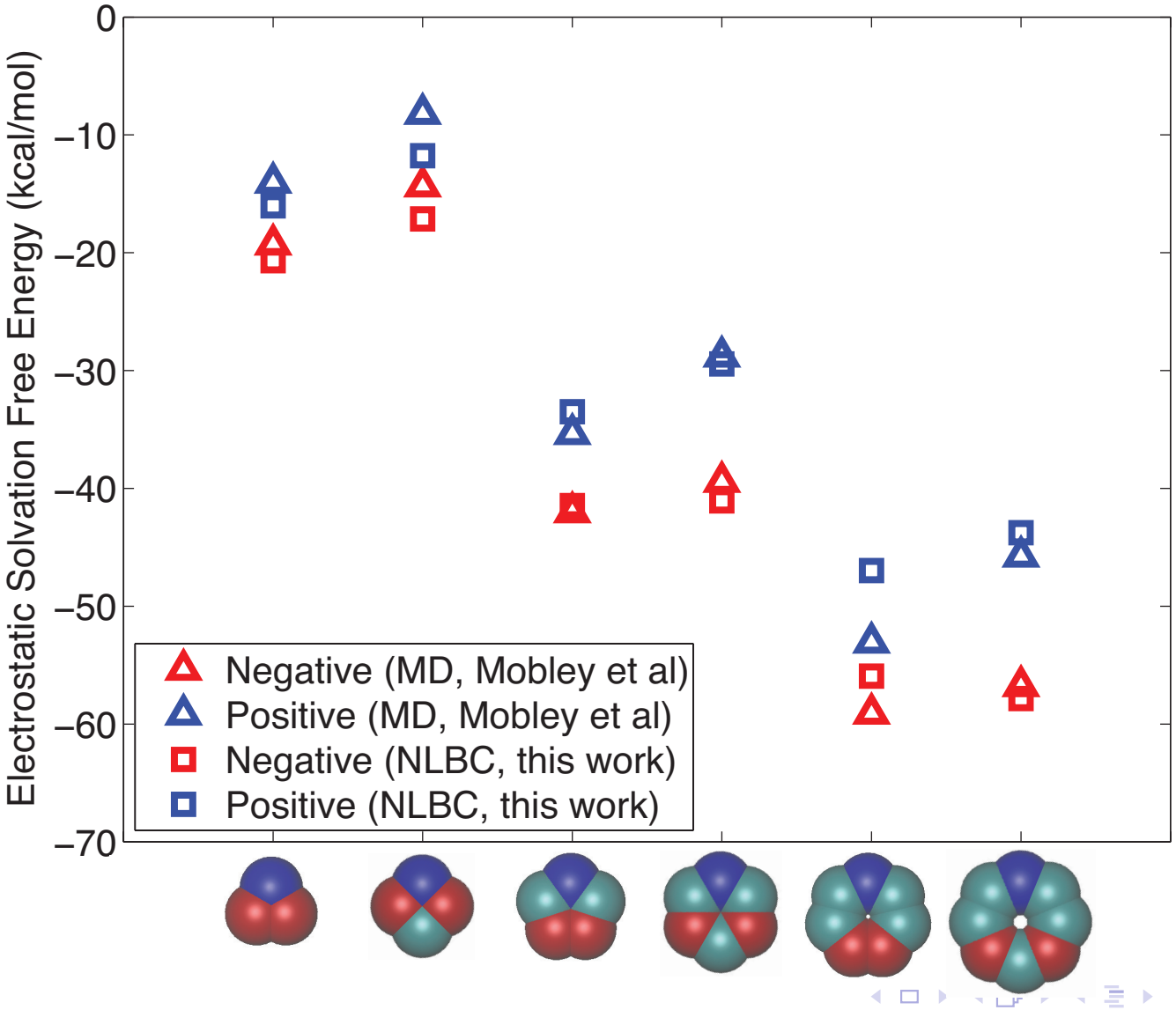
Accuracy of SLIC

Synthetic Molecules



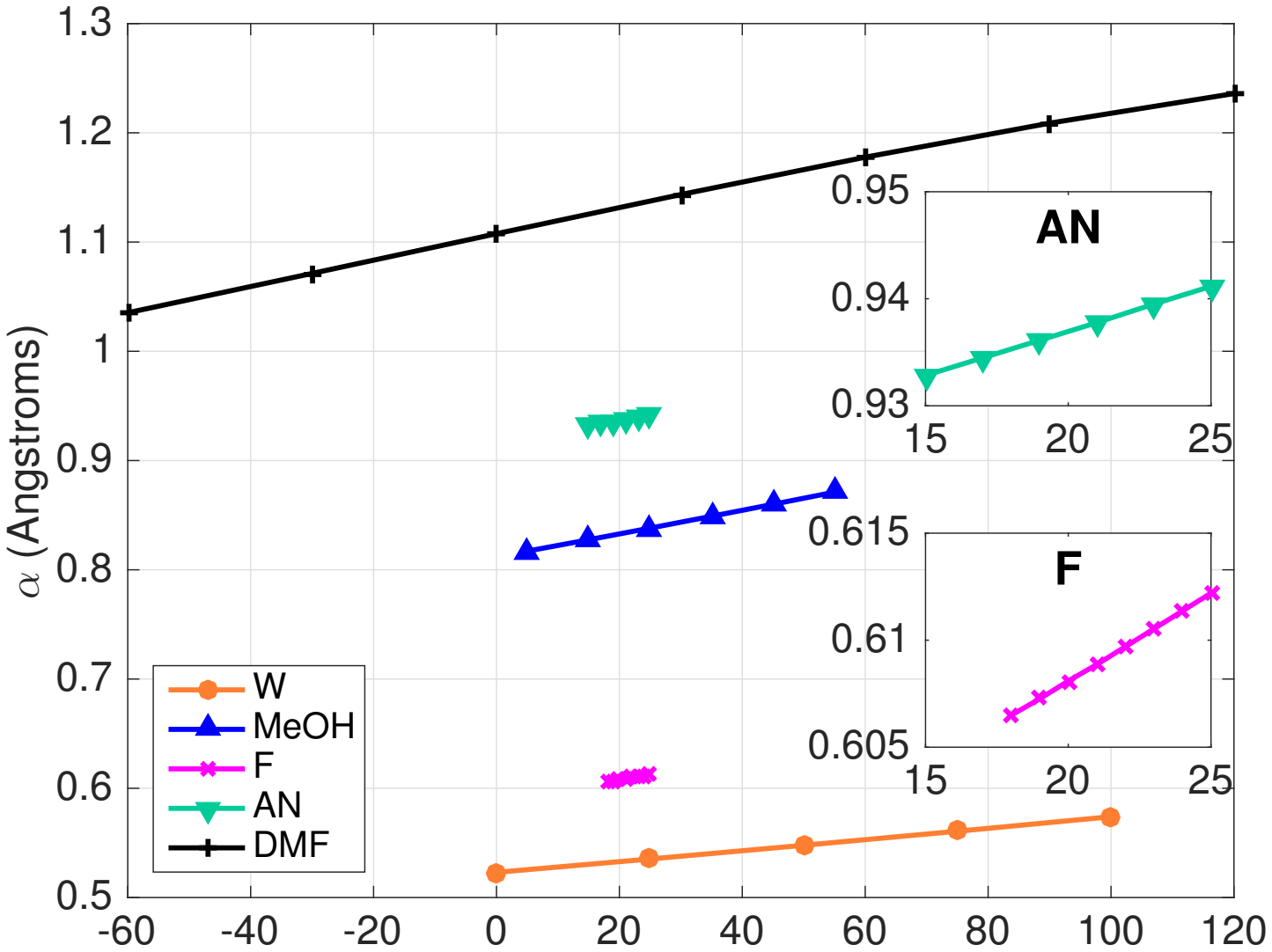
Accuracy of SLIC

Synthetic Molecules



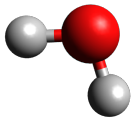
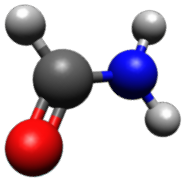
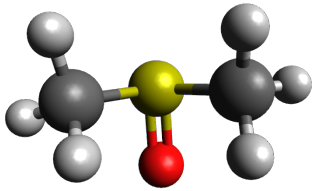
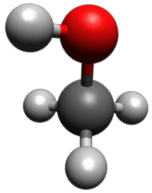
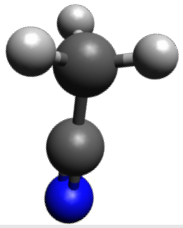
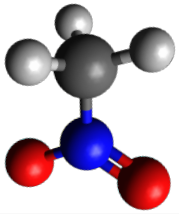
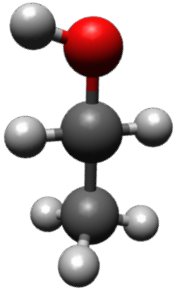
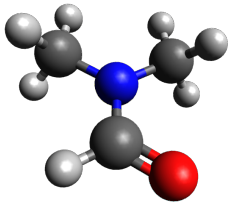
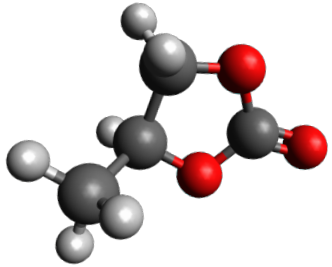
Thermodynamics

The parameters show linear temperature dependence



Model Validation

Courtesy A. Molvai Tabrizi

Water H_2O		Formamide CH_3NO		Dimethyl sulfoxide C_2H_6OS	
Methanol CH_3OH		Acetonitrile C_2H_3N		Nitromethane CH_3NO_2	
Ethanol C_2H_5OH		Dimethyl formamide C_3H_7NO		Propylene carbonate $CH_3C_2H_3O_2CO$	

Model Validation

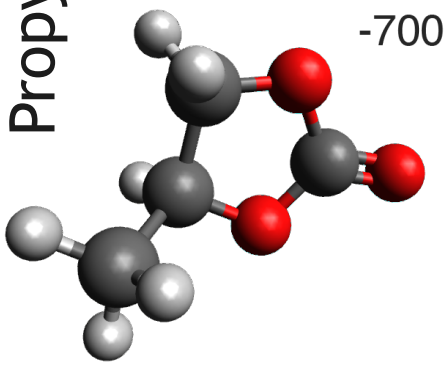
Courtesy A. Molvai Tabrizi

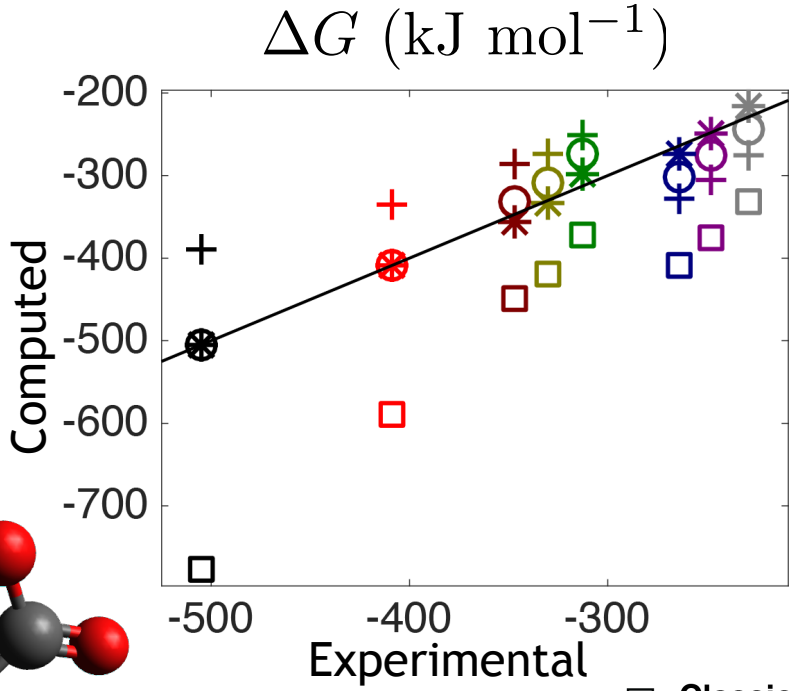
Solvent	r_s (Å)	$\epsilon_{out}(T)$	$\epsilon_{out}(25°\mathrm{C})$
W	1.370	$\epsilon_{out} = 87.740 - 4.0008\mathrm{e}{-1}\ T + 9.398\mathrm{e}{-4}\ T^2 - 1.410\mathrm{e}{-6}\ T^3$	78.3
MeOH	1.855	$\log_{10}\epsilon_{out} = \log_{10}(32.63) - 2.64\mathrm{e}{-3}(T-25)$	32.6
EtOH	2.180	$\log_{10}\epsilon_{out} = \log_{10}(24.30) - 02.70\mathrm{e}{-3}\ (T-25)$	24.3
F	1.725	$\epsilon_{out} = 109 - 7.2\mathrm{e}{-1}\ (T-20)$	105.4
AN	2.135	$\epsilon_{out} = 37.50 - 1.6\mathrm{e}{-1}\ (T-20)$	36.7
DMF	2.585	$\epsilon_{out} = 42.04569 - 2.204448\mathrm{e}{-1}\ T + 7.718531\mathrm{e}{-4}\ T^2 - 1.000389\mathrm{e}{-6}\ T^3$	37.0
DMSO	2.455	$\epsilon_{out} = -60.5 + (5.7\mathrm{e}4/(T+273.15)) - (7.5\mathrm{e}6/(T+273.15)^2)$	46.3
NM	2.155	$\log_{10}\epsilon_{out} = \log_{10}(35.8) - 1.89\mathrm{e}{-3}\ (T-30)$	36.6
PC	2.680	$\epsilon_{out} = 56.670738 + 2.58431\mathrm{e}{-1}\ T - 7.7143\mathrm{e}{-4}\ T^2$	62.6

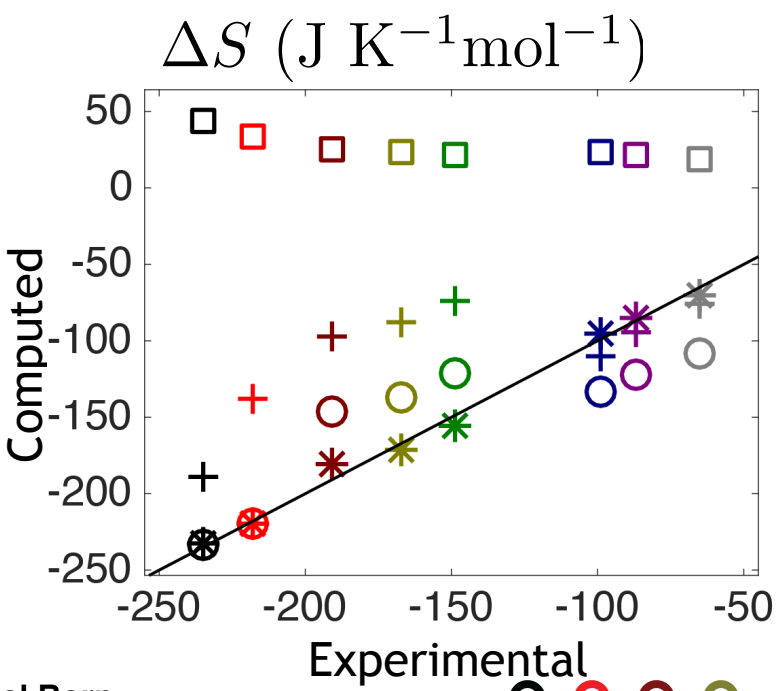
Model Validation

Courtesy A. Molvai Tabrizi

Propylene carbonate
$CH_3C_2H_3O_2CO$





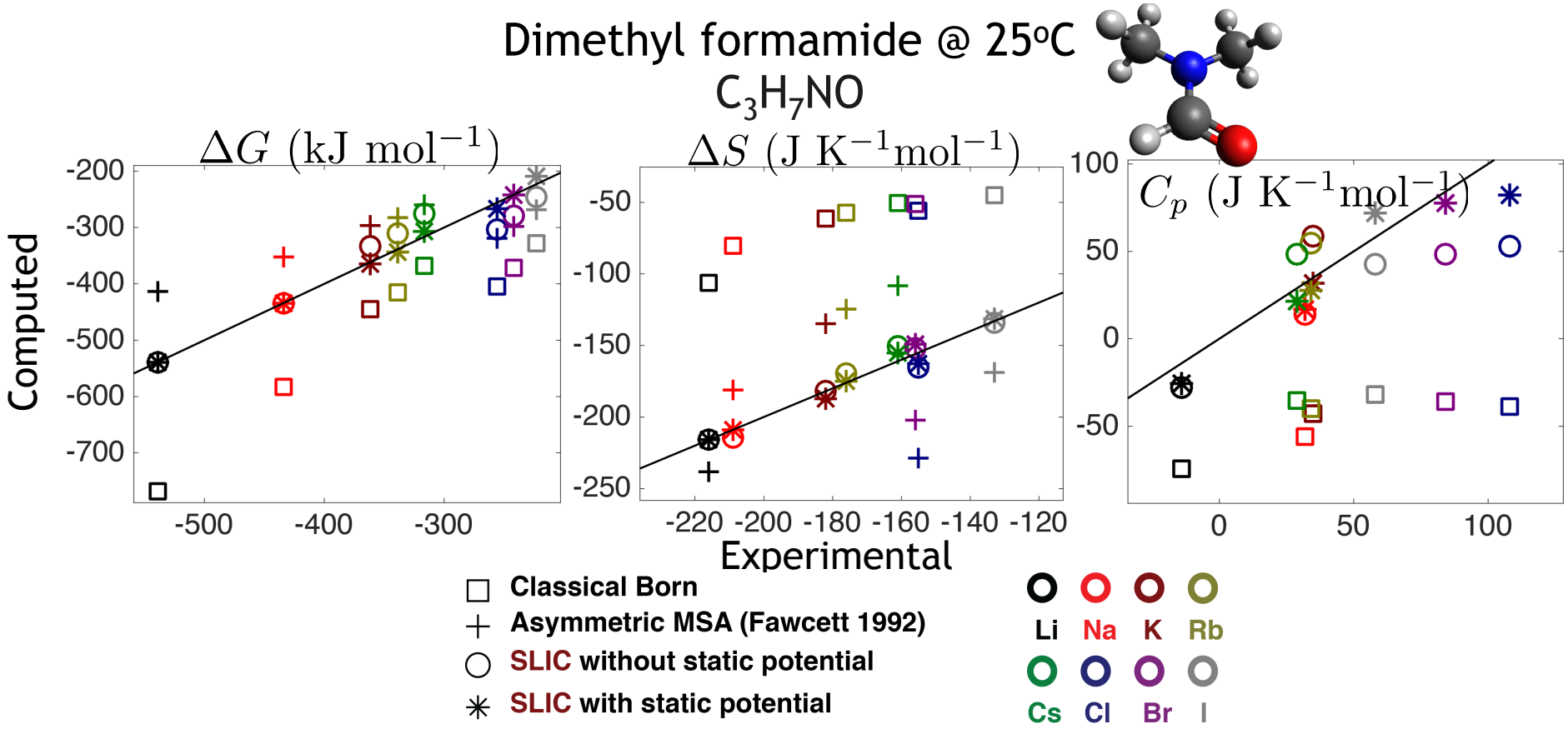


- □ Classical Born
- + Asymmetric MSA (Fawcett 1992)
- ○ SLIC without static potential
- ✳ SLIC with static potential

Li Na K Rb
Cs Cl Br I

Model Validation

Courtesy A. Molvai Tabrizi



Model Validation

Courtesy A. Molvai Tabrizi

A. Molavi Tabrizi, M.G. Knepley, and J.P. Bardhan, *Generalising the mean spherical approximation as a multiscale, nonlinear boundary condition at the solute-solvent interface*, Molecular Physics (2016).

Thermodynamic Predictions

Courtesy A. Molvai Tabrizi

Solvent	Ion	ΔG (kJ mol^{-1})	ΔS (J K^{-1} mol^{-1})	C_p (J K^{-1} mol^{-1})
W	F^-	-430 (-429)	-67 (-115)	-86 (-45)
MeOH	Rb^+	-326(-319)	-178 (-175)	55
	F^-	-415	-116	-79 (-131)
EtOH	Rb^+	-319 (-313)	-197 (-187)	128
	F^-	-405	-145	-153 (-194)
F	Rb^+	-340 (-334)	-135 (-130)	27
	F^-	-418	-128	36 (28)
AN	F^-	-390	-192	147
DMF	F^-	-389	-230	105
DMSO	Rb^+	-348 (-339)	-151 (-180)	32
	F^-	-400	-160	186(60)
NM	Rb^+	-324 (-318)	-186 (-183)	19
	F^-	-391	-182	95(71)
PC	F^-	-394	-149	67

Experimental Data in Parentheses

Thermodynamic Predictions

Courtesy A. Molvai Tabrizi

A. Molavi Tabrizi, S. Goossens, A. Rahimi, M.G. Knepley, and J.P. Bardhan,
Predicting solvation free energies and thermodynamics in polar solvents and mixtures using a solvation-layer interface condition.
Journal of Chemical Physical (2017).

Main Successes of SLIC

Accurate charging free energy

- using crystal radii (no fitting/temp dep)
- for (de-)protonation
- for individual atoms
- for mixtures

Main Successes of SLIC

Accurate charging free energy

- using crystal radii (no fitting/temp dep)
- for (de-)protonation
- for individual atoms
- for mixtures

Main Successes of SLIC

Accurate charging free energy

- using crystal radii (no fitting/temp dep)
- for (de-)protonation
- for individual atoms
- for mixtures

Main Successes of SLIC

Accurate charging free energy

- using crystal radii (no fitting/temp dep)
- for (de-)protonation
- for individual atoms
- for mixtures

Main Successes of SLIC

Accurate charging free energy

- using crystal radii (no fitting/temp dep)
- for (de-)protonation
- for individual atoms
- for mixtures

Main Successes of SLIC

Accurate transfer free energy

- for water-octanol system
- on par with explicit-solvent MD

Reinterpretation of
Mean Spherical Approximation

- Explains temperature dependence of model coefficients

Main Successes of SLIC

Accurate transfer free energy

- for water-octanol system
- on par with explicit-solvent MD

Reinterpretation of
Mean Spherical Approximation

- Explains temperature dependence of model coefficients

What is missing from SLIC?

- Large packing fraction
  - No charge oscillation or overcharging
  - Classical DFT?

    (Gillespie, Microfluidics and Nanofluidics, 2015)
- No dielectric saturation
  - Possible with different condition
- No long range correlations
  - Use nonlocal dielectric

Future Work

- More complex solutes
  - Better nonlinear solvers
- Mixtures
  - Preliminary work is accurate
- Integration into community code
  - Psi4, QChem, APBS

Thank You!

http://cse.buffalo.edu/~knepley

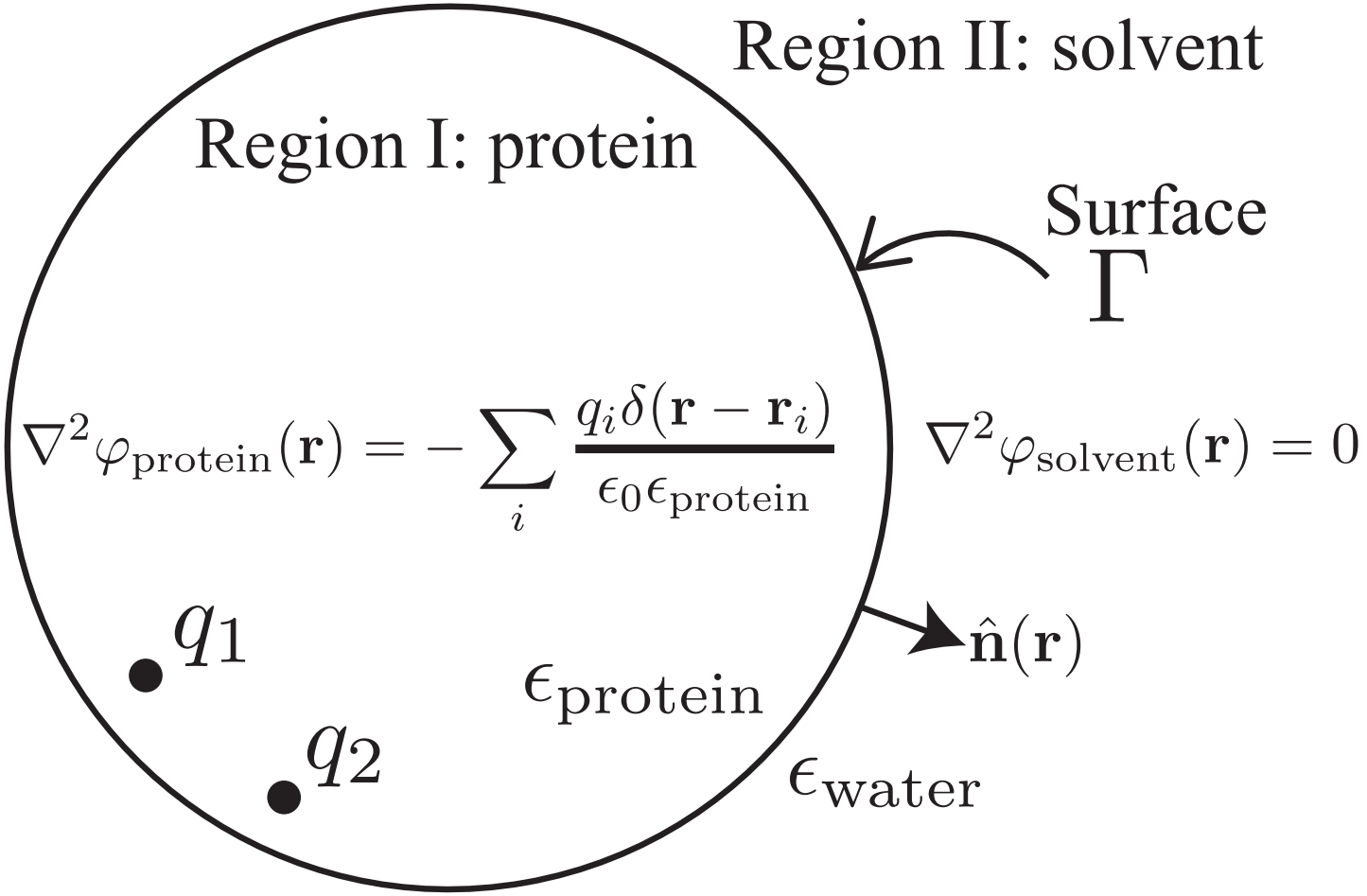
Outline

- Approximate Boundary Conditions

Bioelectrostatics

Physical Model

Electrostatic Potential ϕ



Kirkwood's Solution (1934)

The potential inside Region I is given by

$$\Phi_I = \sum_{k=1}^{Q} \frac{q_k}{\epsilon_1 \left|\vec{r} - \vec{r}_k\right|} + \psi,$$

and the potential in Region II is given by

$$\Phi_{II} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{C_{nm}}{r^{n+1}} P_n^m(\cos\theta) e^{im\phi}.$$

Kirkwood's Solution (1934)

The reaction potential ψ is expanded in a series

$$\psi = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} B_{nm} r^n P_n^m(\cos\theta) e^{im\phi}.$$

and the source distribution is also expanded

$$\sum_{k=1}^{Q} \frac{q_k}{\epsilon_1 \left|\vec{r} - \vec{r}_k\right|} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{E_{nm}}{\epsilon_1 r^{n+1}} P_n^m(\cos\theta) e^{im\phi}.$$

Kirkwood's Solution (1934)

By applying the boundary conditions, letting the sphere have radius b,

$$\Phi_I|_{r=b} = \Phi_{II}|_{r=b}$$
$$\epsilon_I \frac{\partial \Phi_I}{\partial r}|_{r=b} = \epsilon_{II} \frac{\partial \Phi_{II}}{\partial r}|_{r=b}$$

we can eliminate C_{nm}, and determine the reaction potential coefficients in terms of the source distribution,

$$B_{nm} = \frac{1}{\epsilon_I b^{2n+1}} \frac{(\epsilon_I - \epsilon_{II})(n+1)}{\epsilon_I n + \epsilon_{II}(n+1)} E_{nm}.$$

Approximate Boundary Conditions

Theorem: The BIBEE boundary integral operator approximations

$$A_{CFA} = \mathcal{I}\left(1 + \frac{\hat{\epsilon}}{2}\right)$$

$$A_P = \mathcal{I}$$

have an equivalent PDE formulation,

$$\epsilon_I \Delta \Phi_{CFA,P} = \sum_{k=1}^{Q} q_k \delta(\vec{r} - \vec{r}_k)$$

$$\epsilon_{II} \Delta \Phi_{CFA,P} = 0$$

$$\Phi_I|_{r=b} = \Phi_{II}|_{r=b}$$

$$\frac{\epsilon_I}{\epsilon_{II}} \frac{\partial \Phi_I^C}{\partial r}|_{r=b} = \frac{\partial \Phi_{II}}{\partial r} - \frac{\partial \psi_{CFA}}{\partial r}|_{r=b}$$

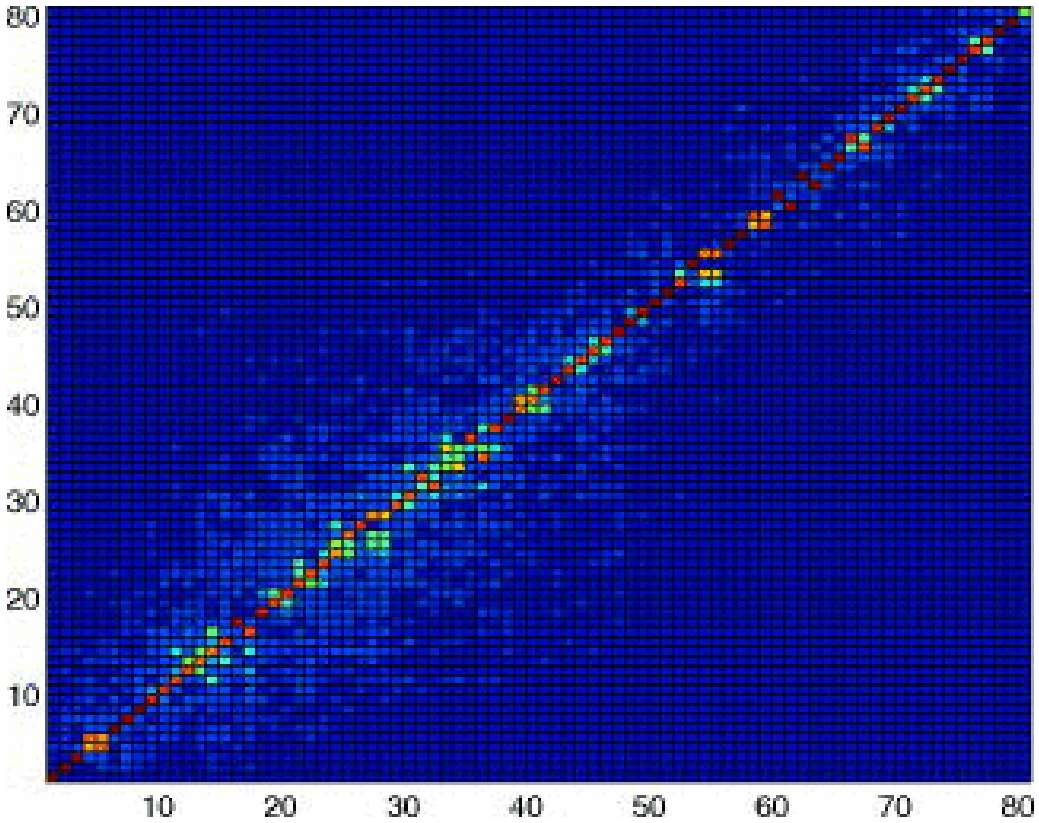
or

$$\frac{3\epsilon_I - \epsilon_{II}}{\epsilon_I + \epsilon_{II}} \frac{\partial \Phi_I^C}{\partial r}|_{r=b} = \frac{\partial \Phi_{II}}{\partial r} - \frac{\partial \psi_P}{\partial r}|_{r=b},$$

where Φ_1^C is the Coulomb field due to interior charges.

Approximate Boundary Conditions

Theorem: For spherical solute, the BIBEE boundary integral operator approximations have eigenspaces are identical to that of the original operator.



BEM eigenvector $e_i \cdot e_j$ BIBEE/P eigenvector

Proof of PDE Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these PDEs are equivalent to the original BIEs,

- Start with the fundamental solution to Laplace's equation $G(r, r')$
- Note that $\int_\Gamma G(r, r')\sigma(r')d\Gamma$ satisfies the bulk equation and decay at infinity
- Insertion into the approximate BC gives the BIBEE boundary integral approximation

Proof of PDE Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these PDEs are equivalent to the original BIEs,

- Start with the fundamental solution to Laplace's equation $G(r, r')$
- Note that $\int_\Gamma G(r, r')\sigma(r')d\Gamma$ satisfies the bulk equation and decay at infinity
- Insertion into the approximate BC gives the BIBEE boundary integral approximation

Proof of PDE Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these PDEs are equivalent to the original BIEs,

- Start with the fundamental solution to Laplace's equation $G(r, r')$
- Note that $\int_\Gamma G(r, r')\sigma(r')d\Gamma$ satisfies the bulk equation and decay at infinity
- Insertion into the approximate BC gives the BIBEE boundary integral approximation

Proof of PDE Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these PDEs are equivalent to the original BIEs,

- Start with the fundamental solution to Laplace's equation $G(r, r')$
- Note that $\int_\Gamma G(r, r')\sigma(r')d\Gamma$ satisfies the bulk equation and decay at infinity
- Insertion into the approximate BC gives the BIBEE boundary integral approximation

Proof of Eigenspace Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these integral operators share a common eigenbasis,

- Note that, for a spherical boundary, $\mathcal{D}^*$ is compact and has a pure point spectrum
- Examine the effect of the operator on a unit spherical harmonic charge distribution
- Use completeness of the spherical harmonic basis

Proof of Eigenspace Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these integral operators share a common eigenbasis,

- Note that, for a spherical boundary, $\mathcal{D}^*$ is compact and has a pure point spectrum
- Examine the effect of the operator on a unit spherical harmonic charge distribution
- Use completeness of the spherical harmonic basis

Proof of Eigenspace Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these integral operators share a common eigenbasis,

- Note that, for a spherical boundary, $\mathcal{D}^*$ is compact and has a pure point spectrum
- Examine the effect of the operator on a unit spherical harmonic charge distribution
- Use completeness of the spherical harmonic basis

Proof of Eigenspace Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these integral operators share a common eigenbasis,

- Note that, for a spherical boundary, $\mathcal{D}^*$ is compact and has a pure point spectrum
- Examine the effect of the operator on a unit spherical harmonic charge distribution
- Use completeness of the spherical harmonic basis

Proof of Eigenspace Equivalence

Proof: Bardhan and Knepley, JCP, **135**(12), 2011.

In order to show that these integral operators share a common eigenbasis,

- Note that, for a spherical boundary, $\mathcal{D}^*$ is compact and has a pure point spectrum
- Examine the effect of the operator on a unit spherical harmonic charge distribution
- Use completeness of the spherical harmonic basis

The result does not hold for general boundaries.

Series Solutions

Note that the approximate solutions are *separable*:

$$B_{nm} = \frac{1}{\epsilon_1 n + \epsilon_2(n+1)}\gamma_{nm}$$
$$B_{nm}^{CFA} = \frac{1}{\epsilon_2}\frac{1}{2n+1}\gamma_{nm}$$
$$B_{nm}^{P} = \frac{1}{\epsilon_1 + \epsilon_2}\frac{1}{n+\frac{1}{2}}\gamma_{nm}.$$

If $\epsilon_I = \epsilon_{II} = \epsilon$, both approximations are exact:

$$B_{nm} = B_{nm}^{CFA} = B_{nm}^{P} = \frac{1}{\epsilon(2n+1)}\gamma_{nm}.$$

Series Solutions

Note that the approximate solutions are *separable*:

$$B_{nm} = \frac{1}{\epsilon_1 n + \epsilon_2(n+1)}\gamma_{nm}$$
$$B_{nm}^{CFA} = \frac{1}{\epsilon_2}\frac{1}{2n+1}\gamma_{nm}$$
$$B_{nm}^{P} = \frac{1}{\epsilon_1 + \epsilon_2}\frac{1}{n+\frac{1}{2}}\gamma_{nm}.$$

If $\epsilon_I = \epsilon_{II} = \epsilon$, both approximations are exact:

$$B_{nm} = B_{nm}^{CFA} = B_{nm}^{P} = \frac{1}{\epsilon(2n+1)}\gamma_{nm}.$$

Asymptotics

BIBEE/CFA is exact for the $n = 0$ mode,

$$B_{00} = B_{00}^{CFA} = \frac{\gamma_{00}}{\epsilon_2},$$

whereas BIBEE/P approaches the exact response in the limit $n \to \infty$:

$$\lim_{n\to\infty} B_{nm} = \lim_{n\to\infty} B_{nm}^{P} = \frac{1}{(\epsilon_1 + \epsilon_2)n}\gamma_{nm}.$$

Asymptotics

BIBEE/CFA is exact for the $n = 0$ mode,

$$B_{00} = B_{00}^{CFA} = \frac{\gamma_{00}}{\epsilon_2},$$

whereas BIBEE/P approaches the exact response in the limit $n \to \infty$:

$$\lim_{n\to\infty} B_{nm} = \lim_{n\to\infty} B_{nm}^{P} = \frac{1}{(\epsilon_1 + \epsilon_2)n}\gamma_{nm}.$$

Asymptotics

In the limit $\epsilon_1/\epsilon_2 \to 0$,

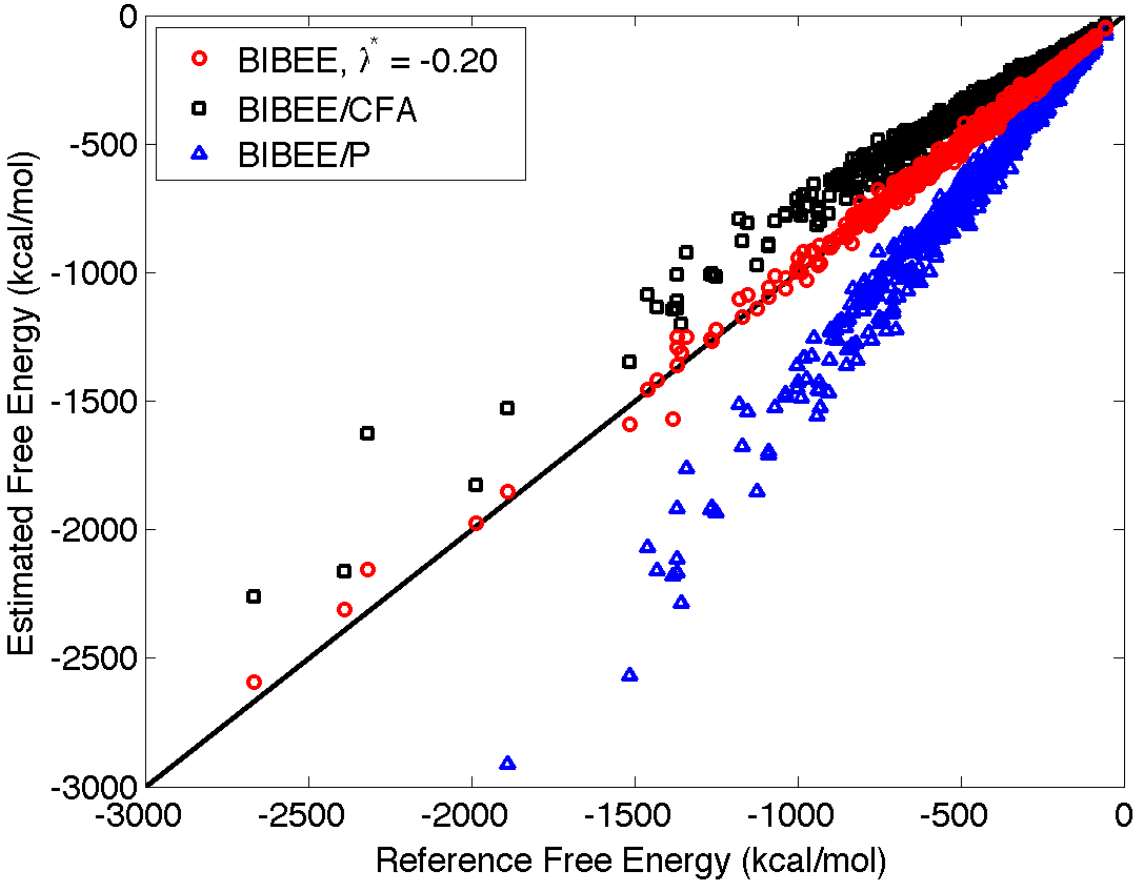
$$\lim_{\epsilon_1/\epsilon_2 \to 0} B_{nm} = \frac{\gamma_{nm}}{\epsilon_2(n+1)}$$
$$\lim_{\epsilon_1/\epsilon_2 \to 0} B_{nm}^{CFA} = \frac{\gamma_{nm}}{\epsilon_2(2n+1)},$$
$$\lim_{\epsilon_1/\epsilon_2 \to 0} B_{nm}^{P} = \frac{\gamma_{nm}}{\epsilon_2\left(n+\frac{1}{2}\right)},$$

so that the approximation ratios are given by

$$\frac{B_{nm}^{CFA}}{B_{nm}} = \frac{n+1}{2n+1}, \qquad \frac{B_{nm}^{P}}{B_{nm}} = \frac{n+1}{n+\frac{1}{2}}.$$

Improved Accuracy

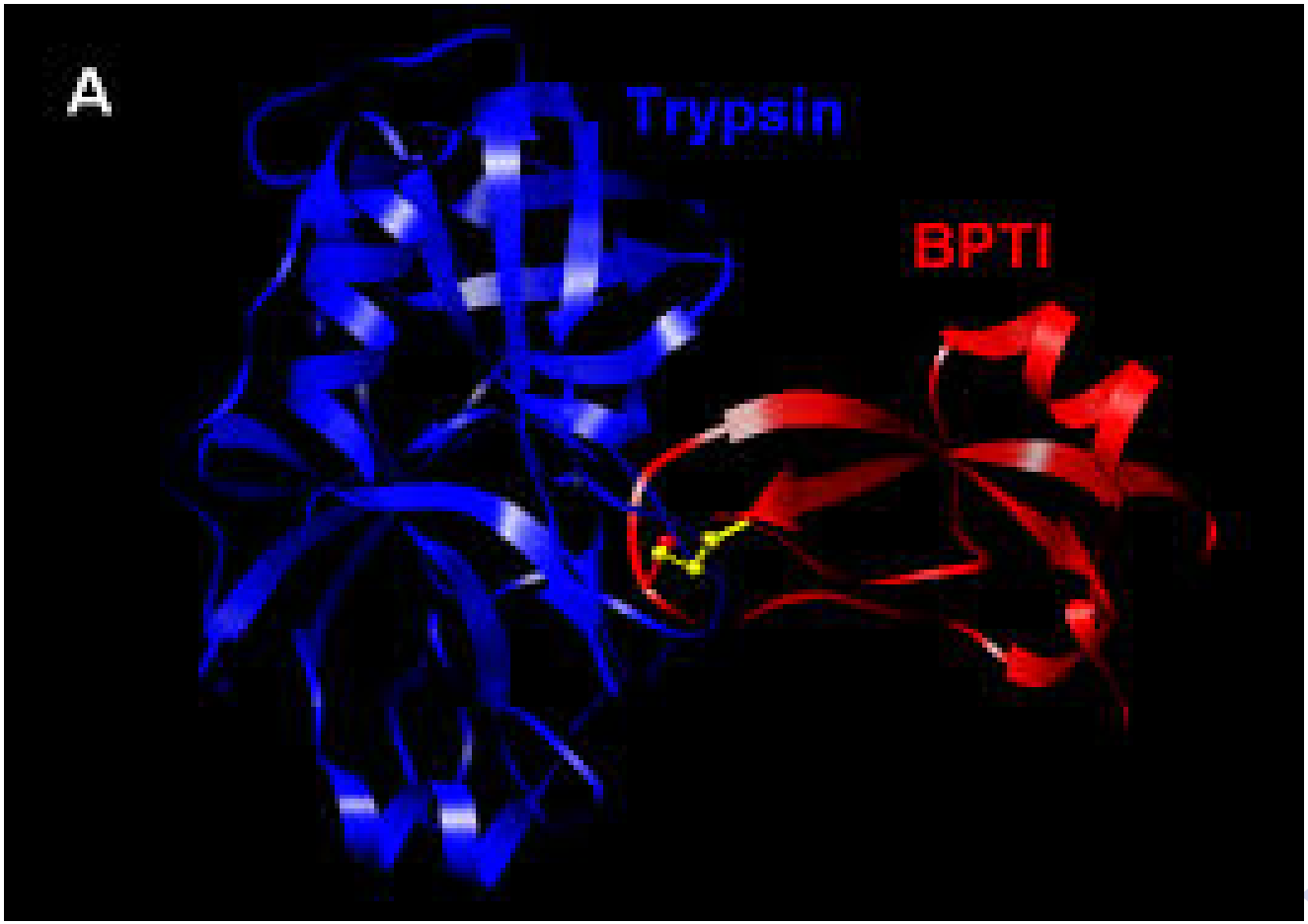
BIBEE/I interpolates between BIBEE/CFA and BIBEE/P



Bardhan, Knepley, JCP, 2011.

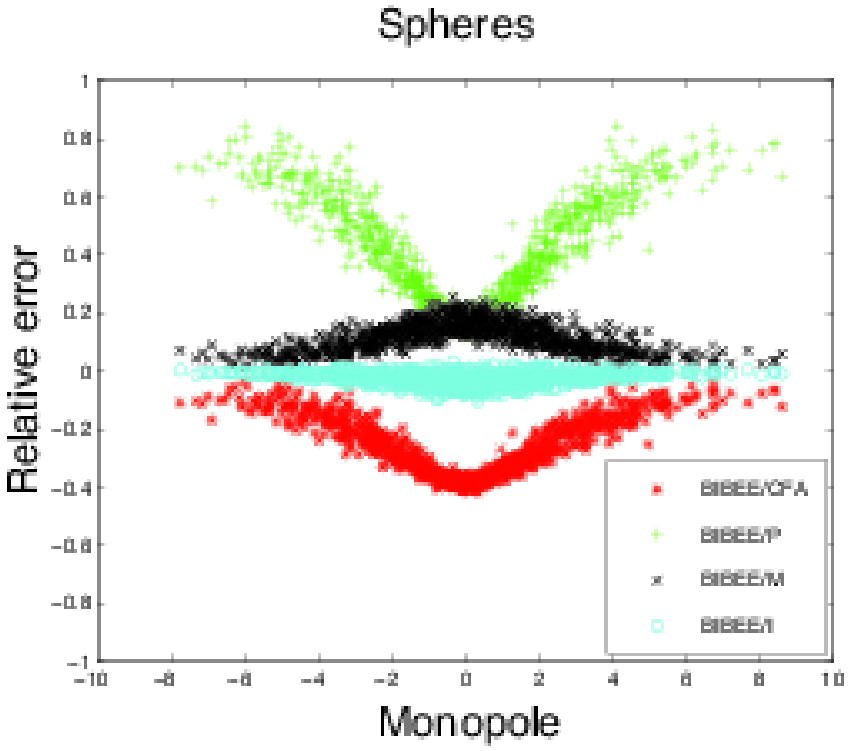
Basis Augmentation

We examined the more complex problem of protein-ligand binding using trypsin and bovine pancreatic trypsin inhibitor (BPTI), using *electrostatic component analysis* to identify residue contributions to binding and molecular recognition.

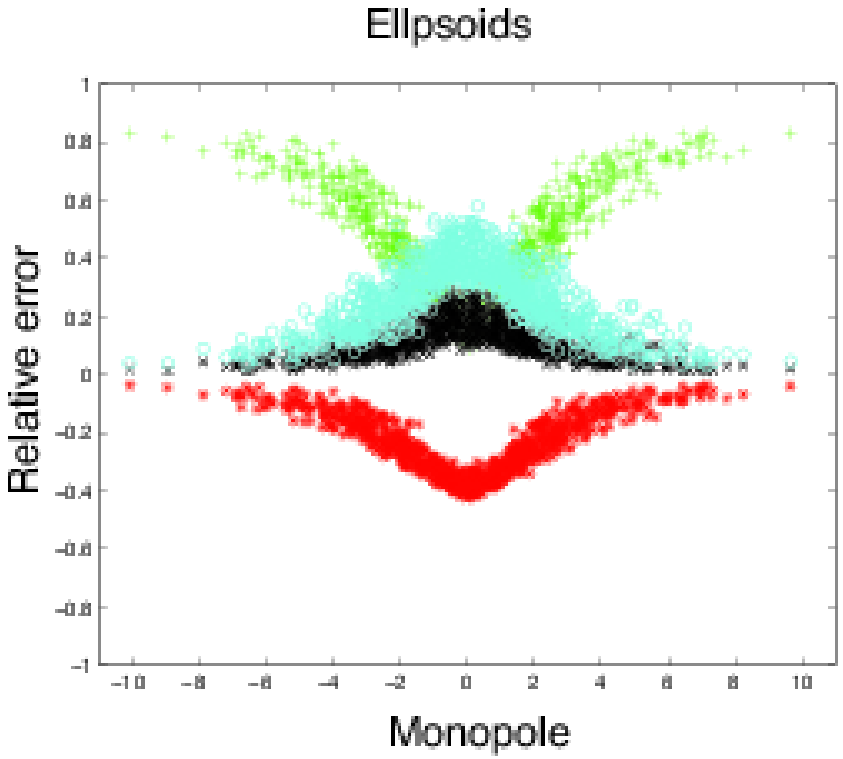


Basis Augmentation

Looking at an ensemble of synthetic proteins, we can see that BIBEE/CFA becomes more accurate as the monopole moment increases, and BIBEE/P more accurate as it decreases. BIBEE/I is accurate for spheres, but must be extended for ellipses.



(a)



(b)

Basis Augmentation

For ellipses, we add a few low order multipole moments, up to the octopole, to recover 5% accuracy for all synthetic proteins tested.