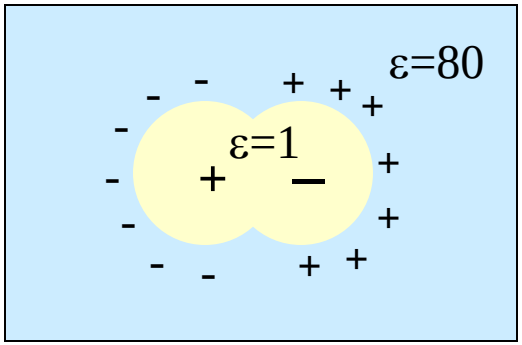
Implicit Models of Solvation

- Goal: Approximate (conformation-dependent) solvation free energies without explicit solvent
- Caution: This field is fraught with controversy and various groups of researchers with dogmatic points of view. On the other hand, implicit solvent models have proven very useful for a wide range of applications to biomolecular systems.
- Focus will be on models of solvation based on continuum electrostatics; nonpolar aspects covered previously.



Hierarchy of Solvent Models

Computational expense

Polarizable explicit solvent

Fixed charge explicit solvent

Nonlinear Poisson-Boltzmann

Linear Poisson-Boltzmann

Generalized Born

Distance-dependent dielectric

Surface area based models

Physical reality

Introduction to PB

- Basic idea: Treat solvent as a dielectric continuum. Essentially exact within the constraints of this approximation.
- Often treated as the “gold standard” for implicit solvent models. However, the continuum approximation simply cannot account for a number of important “first-shell” solvation effects related to the size of a water molecule. [There is absolutely no information about the size of a water molecule with PB treatments.]
- Numerical solution of the continuum electrostatics introduces additional approximations.
- Barry Honig was one of the first people to recognize the usefulness of PB for macromolecules, and developed the most important program, Delphi.
- There are other methods that are formally similar, e.g., the Langevin dipole method of Warshel. But PB has had the largest impact by far.

The Poisson Part of PB

$$\nabla \cdot \varepsilon(\vec{r})\nabla \varphi(\vec{r}) = -4\pi\rho(\vec{r})$$

$\varphi(\vec{r}) =$ electrostatic potential

$\varepsilon(\vec{r}) =$ dielectric constant (small inside protein; 80 outside)

$\rho(\vec{r}) =$ charge density (partial charges inside; ions outside)

This is one of the fundamental equations of classical electrostatics. In fact, Coulomb's law can be derived as a special case where the dielectric is constant.

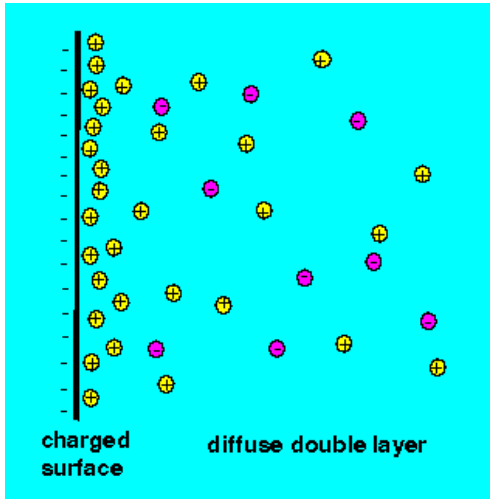
The treatment of charge density due to ions in solution needs special consideration. Can build a model based on the assumptions of Debye-Huckel theory; that's how we get the Boltzmann part ...

Ionic Contributions to Charge Density

Debye-Huckel theory gives the density of ions as

$$\rho_i(\vec{r}) = \rho_i^0 e^{-q_i\varphi(\vec{r})/kT}$$

Ionic density in bulk solution



This just gives a model for the enrichment of, e.g., negative ions in places where the potential is positive. So, for a 1:1 salt solution, we have

$$\rho_{ionic}(\vec{r}) = \rho_+(\vec{r}) + \rho_-(\vec{r}) = \rho^0 e^{-\varphi(\vec{r})/kT} - \rho^0 e^{+\varphi(\vec{r})/kT} = 2\rho^0 \sinh\left(\frac{\varphi(\vec{r})}{kT}\right)$$

The general result is written as ...

The Boltzmann Part of PB

$$\nabla \cdot \varepsilon(\vec{r})\nabla \varphi(\vec{r}) \boxed{- \kappa \sinh[\varphi(\vec{r})]} = -4\pi\rho(\vec{r})$$

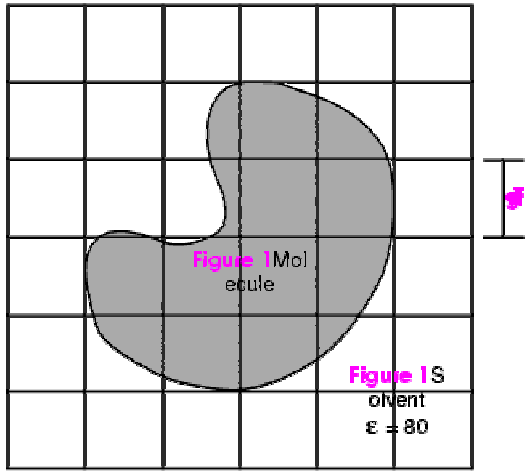
This is the nonlinear PB equation. Note that the charge density on the right side is now just the partial charges in the cavity. In many applications, the ionic strength term is “linearized”, by expanding it in a Taylor series and taking the first term:

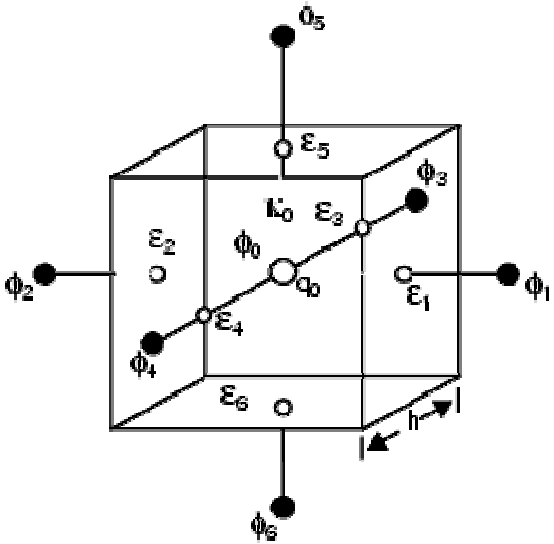
$$\sinh[\varphi(\vec{r})] = \varphi(\vec{r}) + \frac{\varphi(\vec{r})^3}{6} + \cdots \approx \varphi(\vec{r})$$

$$\nabla \cdot \varepsilon(\vec{r})\nabla \varphi(\vec{r}) \boxed{- \kappa\varphi(\vec{r})} = -4\pi\rho(\vec{r})$$

This is reasonable when the ionic strength of the solution is not too high. The linear PB equation is computationally simpler to solve, but the nonlinear PB equation is probably safer for things like DNA/RNA where ionic strength effects are known to be important.

Finite Difference Solutions for PB





- Delphi uses this method.
- Lay down a 3D grid around the protein. The PB equation must be satisfied at every grid point. Derivatives in the PB equation replaced by finite differences over this cube.
- Partial charges of protein must be distributed onto grid (similar to PME). Dielectric constant simple to assign.
- The value of the potential at each grid point depends on the neighboring grid points. Thus you need an iterative solution.
- The accuracy of this method depends critically on the fineness of the grid. Can perform a series of calculations with increasing grid size (“focusing”).

The PB equation can also be solved by a boundary element method. The two approaches are formally equivalent. My understanding is that FD is more efficient for nonlinear PB, but BE may be more efficient for linear PB.

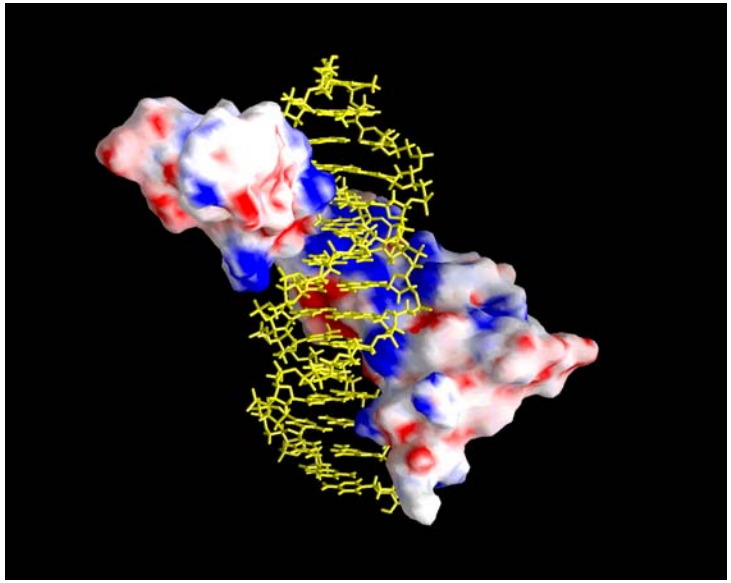
What is the internal dielectric constant of a protein?

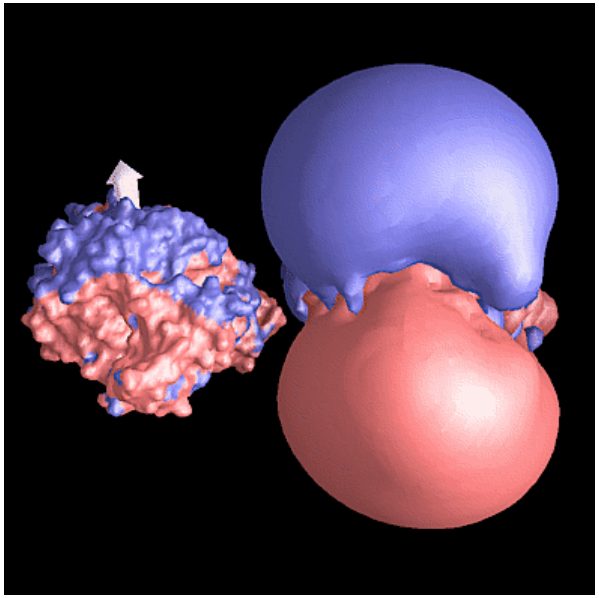
- Nonpolar organic molecules have dielectric constant of ~2-4. This is due primarily to electronic polarizability. This is a reasonable estimate for the internal dielectric of a protein.
- Some computational work, e.g., predicting the pKa's of groups inside proteins, has shown better agreement with experiment using larger values of the internal dielectric, such as 20. Some experiments have also suggested large internal dielectrics.
- However, these large values may reflect effects that should not be considered as part of the dielectric constant, specifically
  1) Conformational relaxation. When the charge state of a group changes, the rest of the macromolecule can respond to help stabilize it. However, most PB calculations are done with a fixed structure; using a high dielectric introduces an artificial screening that can crudely mimic the effects of relaxation.
  2) Protonation state changes. Changing the charge state of a group can induce other protonation states to change. If this is not explicitly taken into account, the effective dielectric again appears to be bigger.

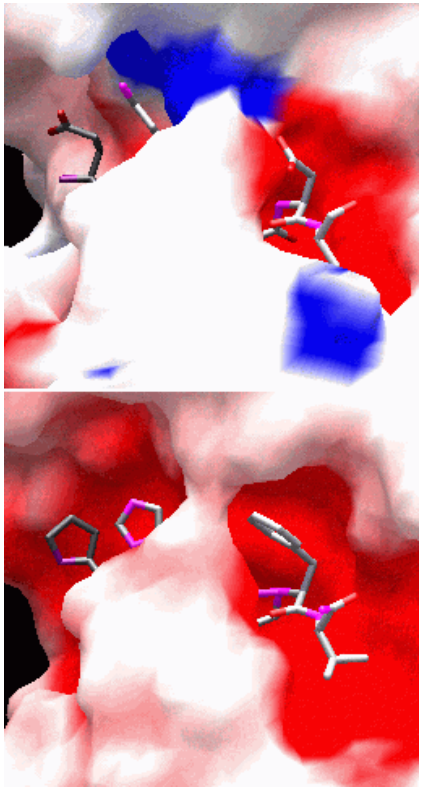
Grasp

This is a general program for visualizing various properties on the surface of a macromolecule (developed by Nicholls and Honig), but the most important application is visualizing the electrostatic field (i.e., as the field lines intersect the surface).

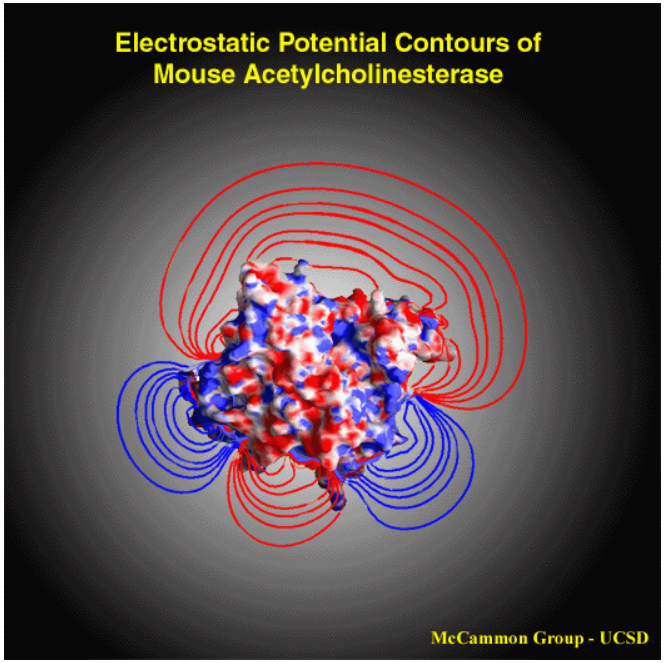
Red = negative
Blue = positive



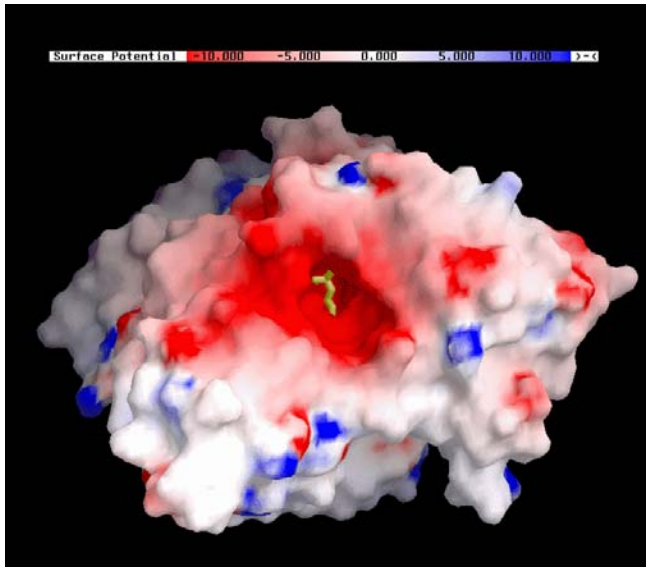




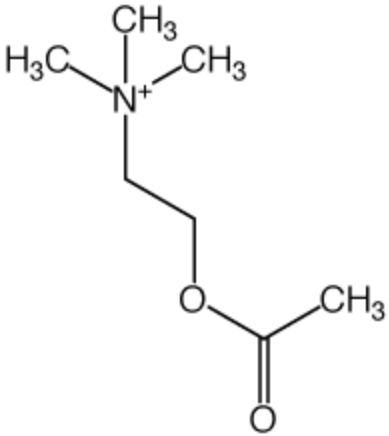
Applications of PB. 1. Role of Electrostatics in Ligand Binding



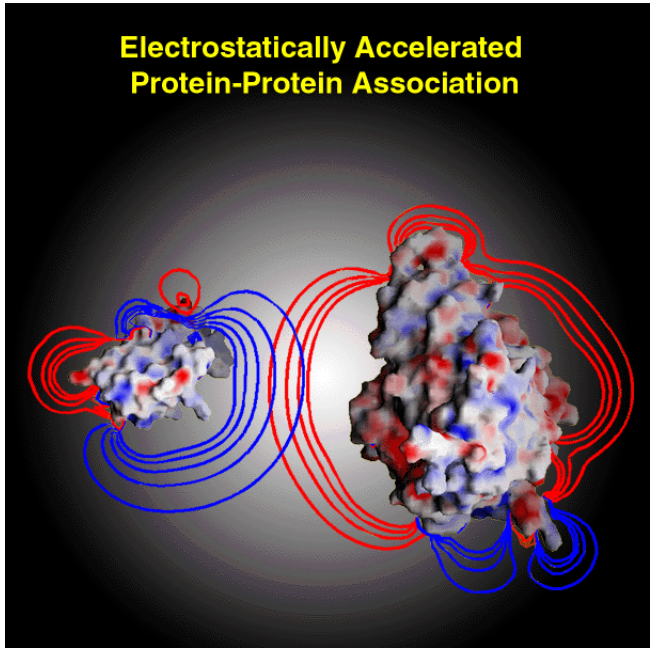
Electrostatic forces speed the binding of the positively-charged substrate to acetylcholinesterase by a factor of more than 100.



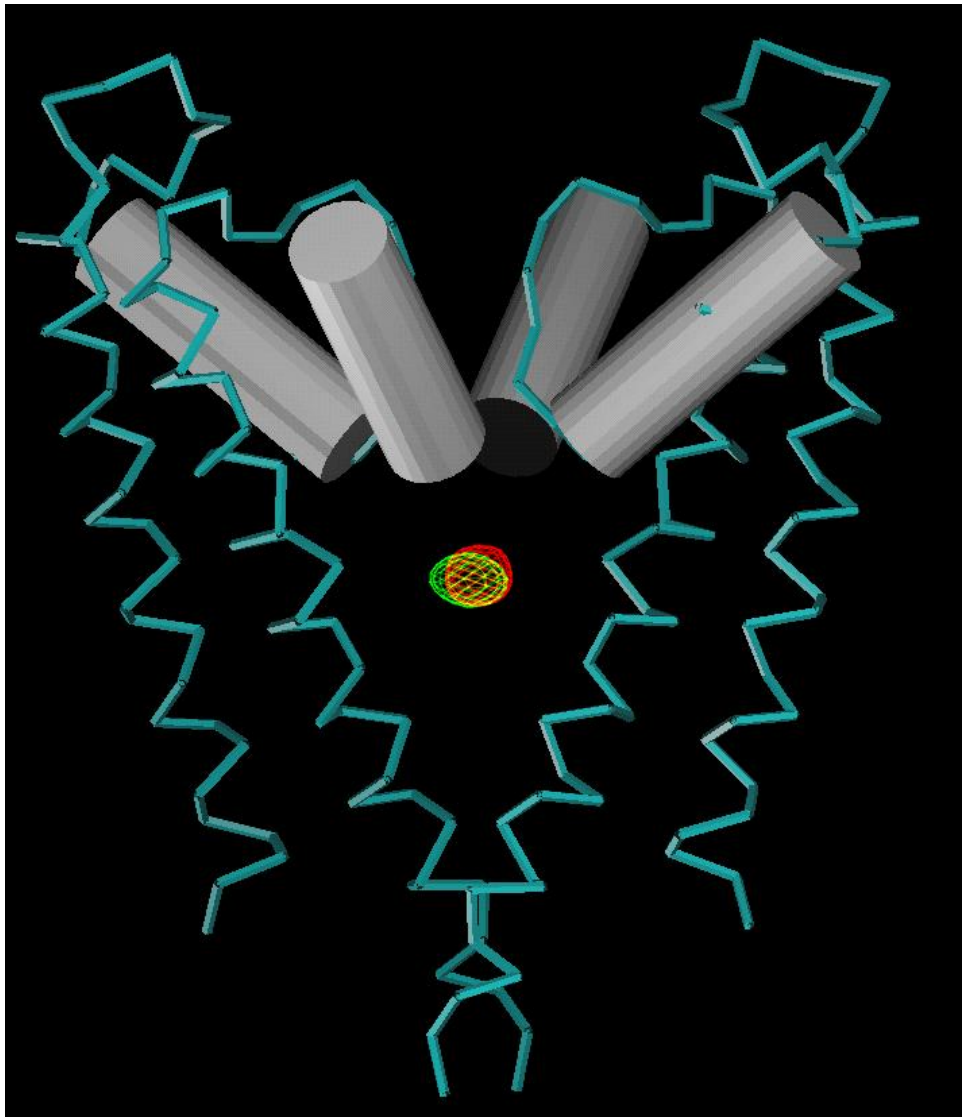
Honig group, Columbia U.



AChE and Fasciculin 2 bind with electrostatically-steered, diffusion-controlled kinetics.



Applications of PB. 2. Role of electrostatics in ion channels



Continuum electrostatics calculations based on the finite-difference Poisson-Boltzman equation are used to show that the cavity and the pore helices of the KcsA channel are "tuned" to be preferably occupied by a monovalent cation. Note that the role of helix dipole in other proteins is controversial, but in membranes (low dielectric), seems to be fairly strong.

B. Roux and R. MacKinnon, "The cavity and pores helices in the KcsA K+ channel: electrostatic stabilization of monovalent cations", *Science* **285**, 100-102 (1999).

Applications of PB. 3. pKa calculations

The effective pKa of a group (e.g., carboxylate) can be strongly perturbed by its environment in a macromolecule. For example, a carboxylate that forms a salt bridge with Arg or Lys is more likely to be negatively charged, while a carboxylate buried in a hydrophobic cavity is more likely to be neutral (desolvation of a unit charge costs about 60 kcal/mol!).

PB calculations can quantify this effect, and predict how the effective pKa (i.e., the pH at which the group is 50% protonated) shifts from the nominal value.

model system (neutral) ——— in macromolecule (neutral)

e.g., acetic acid | | **e.g., Asp**

model system (charged) ——— in macromolecule (charged)

$$pKa_i^{macro} = pKa_i^{model} - \frac{\left(\Delta G_i^{macro} - \Delta G_i^{model}\right)}{2.303RT}$$

The tricky part is that the pKa of the group will depend on protonation states of all other titratable side chains; need some sort of optimization strategy.

Other Applications of PB

- Molecular dynamics. It's tough to get derivatives of the solvation free energy, which are formally needed for integrating Newton's equations of motion. But it can be done.
- Scoring function for modeling or ligand docking. We'll have a lecture later on MM-PBSA, which can provide reasonably accurate estimates of binding free energies.
- Identifying residues critical for binding, or explain mutagenesis results. It's straightforward to quantify the electrostatic consequences of mutations, e.g., in a protein-protein interface.

Introduction to GB

- Originally developed by Clark Still. Original paper is virtually impossible to understand; GB approximation remains a bit mysterious.
- My own viewpoint: GB can be understood as an environment-specific distance dependent dielectric model. In other words, it correctly predicts (at least qualitatively) differences in effective electrostatic interactions in protein interior vs. solvent-exposed exterior.
- Easier to obtain derivatives of solvation free energies than in PB methods; very useful for MD and minimization.

Basic Idea of GB

The Born model of ion solvation: PB equation for a single charge at the center of a sphere has an analytical solution ...

$$\Delta G_{solv} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon_{ext}}\right)\frac{q^2}{R}$$

Now consider a charge inside an arbitrary cavity (e.g., macromolecule). Assuming we can compute the associated solvation free energy, equate this with the Born equation to determine the sphere size that would give the same result. This "effective Born radius" is usually represented by the symbol α

$$\Delta G_{solv} = -\frac{1}{2}\left(1 - \frac{1}{\varepsilon_{ext}}\right)\frac{q^2}{\alpha}$$

If the atom is solvent exposed, the Born radius will be close to (but larger than) the atomic radius. For a buried atom, the Born radius can become quite large.

Now, consider several charges inside an arbitrary cavity. Each one has its own Born radius. The GB method provides a method for estimating the total free energy of solvation, based on these Born radii.

The GB Pair Term

$$\Delta G^{solv} = \sum_i \sum_j \Delta G_{ij}^{pair} = \sum_i \Delta G_i^{self} + 2\sum_i \sum_{j>i} \Delta G_{ij}^{pair}$$

$$\Delta G_{ij}^{pair} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{\text{int}}} - \frac{1}{\varepsilon_{ext}}\right)\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp\left(-\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right)}}$$

Q: How did Still come up with this term?
A: No one knows. However, notice the following ...

1. If i=j, then get back the Born equation: $\Delta G_i^{self} = \Delta G_{ii}^{pair} = -\frac{1}{2}\left(\frac{1}{\varepsilon_{\text{int}}} - \frac{1}{\varepsilon_{ext}}\right)\frac{q_i^2}{\alpha_i}$
2. In the limit of large internuclear distances, get complete screening. That is, the pair term almost exactly cancels the Coulomb's law attraction:

$$\Delta G_{ij}^{pair} + \Delta G_{ji}^{pair} = -\left(\frac{1}{\varepsilon_{\text{int}}} - \frac{1}{\varepsilon_{ext}}\right)\frac{q_i q_j}{r_{ij}}$$

3. You can also show that the pair term leads to correct behavior in the limit of a dipole inside a sphere, which is also analytical (i.e., using PB equation).
4. Other than the limiting behaviors, I know of no physical basis for the eqn.

How do we get Born α's?

<u>From solving PB equation</u>:

In practice, it would be enormously inefficient to solve the PB equation for each partial charge. However, an interesting recent paper ...

> ***Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect. ALEXEY ONUFRIEV, DAVID A. CASE, DONALD BASHFORD. J. Comp. Chem. 23 (2002) 1297.***

<u>Original volume integral formulation</u>:

$$E_k = \frac{3}{8\pi}\left(\frac{1}{\varepsilon_{\text{int}}} - \frac{1}{\varepsilon_{ext}}\right)\int_V \frac{q_k^2}{\left|\vec{R} - \vec{r}_k\right|^4} d^3\vec{R}$$

$\vec{r}_k =$ position of the charge

$\vec{R} =$ vector of integration

This is not an exact solution for the solvation free energy of the charge in the cavity, but can be shown to be a first-order approximation. Specifically, in the boundary element formulation of PB, it represents the interaction of the point charge with the surface charge it induces, but neglects higher-order effects due to the electric field induced by the surface charge.

How do we get Born α's?

<u>Surface integral formulation</u>:

Friesner pointed out that, by Green's theorem, volume integral can be transformed to a surface integral (better computational scaling)

$$E_k = -\frac{1}{8\pi}\left(\frac{1}{\varepsilon_{\text{int}}} - \frac{1}{\varepsilon_{ext}}\right)\int_S \frac{q_k^2}{\left|\vec{R}-\vec{r}_k\right|^4}\left(\vec{R}-\vec{r}_k\right)\cdot\vec{n}\ \ d^2\vec{R}$$

$\vec{r}_k =$ position of the charge

$\vec{R} =$ vector of integration

$\vec{n} =$ surface normal

<u>Analytical approximations</u>:

Calculation of the Born radii by surface or volume integration can be slow. These methods basically assign Born radii based on local environment of atom, using analytical, differentiable formulas. In my opinion, these methods do not capture differences in burial nearly as well as integration methods, but clearly they are much faster.

A Note About Surfaces

- All continuum electrostatics calculations must define where the surface of the macromolecule somehow, i.e., where does the high dielectric end and the low dielectric begin. There are multiple possible definitions, all of which have their own pathologies.
- “Molecular surface”: This is the most commonly used surface for PB. Imagine rolling a sphere (size of a water) on the surface of the macromolecule. All points of contact between the sphere and the macromolecule define the surface. An analytical definition has been provided by Connolly. People sometimes refer to this as the “correct” surface, but that it meaningless. There are problems with sharp “cusps”, and it is virtually impossible to get derivatives. Difficult to find internal cavities.
- “Van der Waals surface”: Simply take the intersection of the VDW spheres defined by each atom. This surface is problematic because it leads to large amounts of high dielectric in the interior of macromolecules, due largely to small cavities (too small to actually accommodate waters).
- “Gaussian surface”: Center a 3D Gaussian around each atom (something like an “s” orbital. Add together all the Gaussians. Can then either 1) treat the surface as a continuum, or 2) draw an isocontour. This has the advantage of much easier derivatives.

PB vs. GB

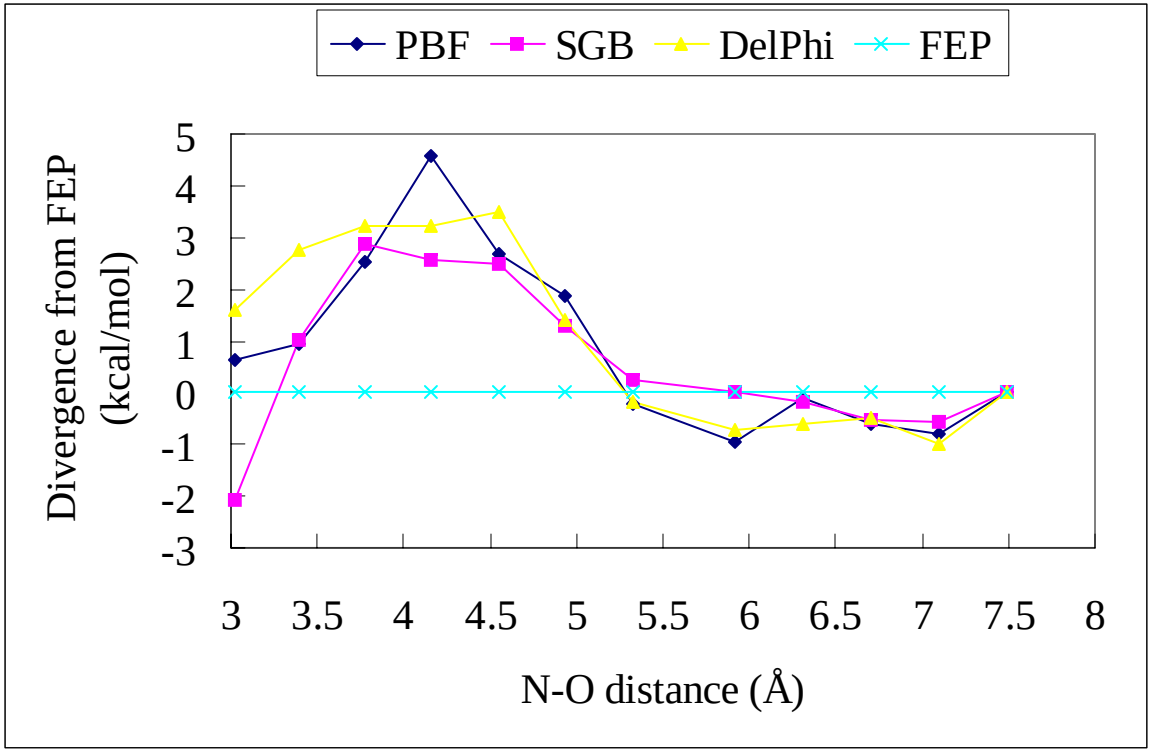
- There are some rather dogmatic viewpoints out there on this issue. Anthony Nicholls gave an infamous talk at ACS called “GB and other bad solvation models”.
- Two types of studies have been performed:
  1) Treat PB as the correct answer (dangerous!) and see how well GB (and other simplified models) reproduce it.
  2) Compare PB and GB at reproducing solvation free energies of small molecules. This is fine, but small molecules (all atoms solvent exposed) are very different than macromolecules (buried core; more complicated surfaces).
- Computational expense of energies and derivatives needs to be considered.

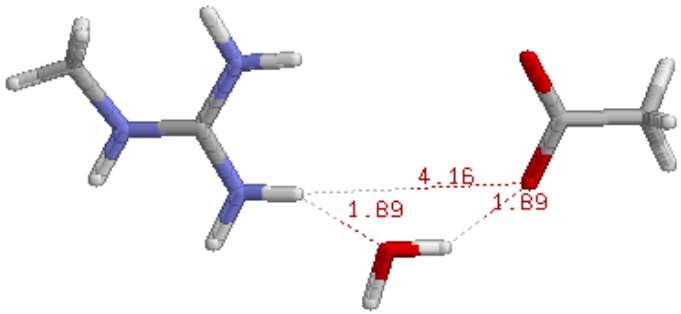
Implicit Solvent vs. Explicit Solvent

- Implicit solvent generally cannot be used when trying to learn about dynamics or kinetics, which are obviously influenced by the granularity of water. Dynamics are generally much faster in implicit solvent.
- Keep in mind that the key issue is not the cost of calculating an energy of interaction between the protein and explicit solvent, but getting a solvation free energy (i.e., ensemble averaging over solvent configurations).
- Explicit solvent models, and periodic boundary conditions, of course are not free from artifacts.
- From my research and others, it is clear that implicit solvent models can have a tough time with charged systems. Water in confined spaces is another tricky issue.

When does implicit solvent break down?

An example from my own work.





FEP: Explicit solvent
PBF/Delphi: 2 different PB solvers

Conclusions:

- Deviations between implicit and explicit solvent are reproducible across several implicit models, including both GB and PB.
- Peak deviation at 4-5 Å readily explained in terms of first shell solvation effect that we refer to as “bridging waters”.

Heuristic Models

Distance Dependent Dielectric

Use Coulomb's law, but make dielectric constant vary with distance between charges. Most commonly, dielectric increases linearly with distance.

$$\varepsilon(r) \propto r$$

This was used very frequently in the early days of molecular dynamics. It's better than vacuum (ε=1), but ε=r still grossly overestimates electrostatic interactions. ε=4r works a little better. However, whatever the details, note that DDD treats all pairs of charges in a macromolecule identically, regardless of whether they are on the surface or in the interior. [This is the big advantage of GB; although the screening term is probably not perfect, it does treat core and surface atoms differently.]

Surface Area Based Methods

$$\Delta G_{solv} = \sum_i \sigma_i A_i$$

Not used much in MD (derivatives are tricky), but shows up in other contexts in combination with force fields, e.g., homology modeling. No real physical basis.