Electrostatics and Solvation in Biomolecular Systems
Biomolecular electrostatics: proteins

Mean-squared charge per residue ($e^2$)

Number of chains (PDBAA)

Total charge ($e$)

Number of chains (PDBAA)
Biomolecular electrostatics: other molecules

- dsDNA
  - Approx. linear form
  - Close phosphate spacing
  - 2 e⁻ per 3.4 Å
- RNA
  - Structural diversity
  - Dense phosphate spacing
- Sugars
- Lipids

Dermatan sulfate picture from Alberts et al
How solvent interacts with biomolecules

- Water properties
  - Dipolar solvent (1.8 D)
  - Hydrogen bond donor and acceptor
  - Polarizable

- Functional behavior:
  - Bulk polarization
  - Site binding or specific solvation
  - Preferential hydration
  - Acid/base chemistry
  - Spine of hydration in DNA minor groove (Kollman, et al.)
  - Carbonic anhydrase reaction mechanism (Stryer, et al.)
How ions interact with biomolecules

- Non-specific screening effects
  - Depends only on ionic strength (not species)
  - Results of damped electrostatic potential
  - Described by Debye-Hückel or Poisson-Boltzmann theories for low ionic strengths

- Functional behavior:
  - Described throughout lectures
  - Binding constants
  - Rates

Electrostatic potential of AChE at 0 mM and 150 mM NaCl. Rate and binding affinity decrease with [NaCl] has been attributed to screening effects... although species-dependent influences have been observed. Radic Z, et al. 1997. J Biol Chem 272 (37): 23265-77.
How ions interact with biomolecules

- Site-specific binding
- Ion-specific
- Site geometry, electrostatics, coordination, etc. enables favorable binding
- Functional behavior: co-factors, allosteric activation, folding, etc.


Figure 1  Ions and the RNA folding problem. Top panel: tRNA folding can be visualized as successive formation of secondary structure (intermediate, I) and tertiary structure (folded, F) from unfolded (U) RNA. Counterions associated with the RNAs are illustrated as black dots (monovalent) and open circles (divalent); Mg²⁺ is particularly effective in promoting tertiary structure folding. Bottom panel: Environments that might be seen by an RNA-associated ion differ in the extent of direct ion-RNA contacts, water-mediated ion-RNA contacts, and hydration. The ion is represented by a small circle; water molecules are outlined, and shaded spheres represent RNA backbone phosphates.
How ions interact with biomolecules

- Hofmeister effects (preferential hydration), ca 1888
- How much salt is required to precipitate a protein? *It depends on the salt...*
- Partitioning of ions between water and nonspecific sites on biomolecule
- Dependent on ion type (solvation energy, etc.)
- Dominate at high salt concentrations
- Functional behavior: protein stability, membrane structure and surface potentials, protein-protein interactions

\[\text{most stabilizing (strongly solvated anions)}\]
\[
citrate^{2-} > SO_4^{2-} > PO_4^{2-} > F^- > Cl^- > Br^- > I^- > NO_3^- > ClO_4^-\]
\[\text{most destabilizing (weakly solvated anions)}\]
\[N(Me)_4^+ > NH_4^+ > Cs^+ > Rb^+ > K^+ > Na^+ > H^+ > Ca^{2+} > Mg^{2+} > Al^{3+}\]
Electrostatics and solvation in biomolecular simulation
Binding energies

Absolute

Relative
pK$_a$ calculations
Potentials of mean force

Structural analysis

Other applications

- Kinetics
  - Forces for molecular and Brownian dynamics
- Minimization and structure refinement
- Investigation importance of “molecular detail” in solvation and ion interactions
Computational methods for biomolecular electrostatics and solvation
Modeling biomolecule-solvent interactions

- **Solvent models**
  - Quantum
  - Explicit
    - Polarizable
    - Fixed charge
  - Integral equation
    - RISM
    - 3D methods
    - DFT
  - Primitive
    - Poisson equation
  - Phenomenological
    - Generalized Born, et al
    - Modified Coulomb’s law

- **Ion models**
  - Quantum
  - Explicit
    - Polarizable
    - Fixed charge
  - Integral equation
    - RISM
    - 3D methods
    - DFT
  - Field-theoretic
    - Extended models
    - Poisson-Boltzmann equation
  - Phenomenological
    - Generalized Born, et al
    - Modified Debye-Hückel

Increasing detail, cost
Explicit solvent simulations

- Sample the configuration space of the system: ions, atomically-detailed water, solute
- Sample with respect to a particular ensemble: NpT, NVT, NVE, etc.
- Molecular dynamics or Monte Carlo
- Advantages:
  - High levels of detail
  - Additional degrees of freedom readily included
  - All interactions are explicit
- Disadvantages
  - Slow and uncertain convergence
  - Boundary effects
  - Poor scaling
  - Some effects still not considered in many force fields...
Implicit solvent models

- Solute typically only accounts for 5-10% of atoms in explicit solvent simulation...
- ...so treat solvent effects implicitly:
  - Solvent as polarization density
  - Ions as “mobile” charge density
- Linear and local solvent response
- “Mean field” ion behavior
- Uncertain treatment of “apolar” effects
Implicit solvent issues

- Where is the molecular detail of solvent and ions important?
- Where does nonlinear solvent/ion response matter?
- What is the correct description of nonpolar solvation?
- What is the correct interface between implicit/explicit solvent methods?
“Potentials of mean force” (PMF) and solvation free energies
- Function of conformation
- Integration over explicit degrees of freedom yields free energy
  - *Global information*
- Mean forces
  - Derivatives of PMFs for atom positions
  - Integration yields PMFs
  - *Local information*
Polar solvation (implicit)

- Charging free energies
- Solvent: dielectric effects through Poisson equation
- Ions: mean-field screening effects through Poisson-Boltzmann equation
Electrostatics in a homogeneous dielectric

- An *isotropic* dielectric continuum exhibits the same response in all directions.
- The dielectric tensor can be reduced to a scalar.
- For a homogeneous isotropic dielectric, electrostatic energies are still governed by Coulomb’s law (with a dielectric coefficient).

\[
U = \frac{q_1 q_2}{4\pi \varepsilon_0 \varepsilon_r r}
\]

\[
F = \frac{q_1 q_2}{4\pi \varepsilon_0 \varepsilon_r r}
\]

Dielectric constant
Several contributions to polarizability
- Electronic polarizability
- Intramolecular rearrangement
- Reorientation of permanent dipole moment
- Hydrogen bonding networks
Molecular dielectric coefficients

- A heterogeneous molecule like a biomolecule shouldn’t really be represented by a continuum dielectric...
- ...however, that doesn’t keep people from trying
- Multiple dielectric values:
  - 1 = vacuum
  - 2-4 = atomic polarizability (solid)
  - 4-10 = some libration, minor sidechain rearrangement
  - 10-20 = significant internal rearrangement
- Multiple surface definitions:
  - van der Waals
  - Splines
  - Molecular surface
Gauss’ law, Gauss’ theorem, and Poisson equation

- Gauss’ law: the integral of the electric field over a surface equals the enclosed charge (general conservation relation)

\[ \int_{\partial \Omega} \epsilon(s) \mathbf{E}(s) \cdot ds = \int_{\Omega} \frac{\rho(x)}{\epsilon_0} \, dx \]

- Gauss’ theorem: the integral of a flux over a closed surface equals the enclosed divergence

\[ \int_{\partial \Omega} \mathbf{v}(s) \cdot ds = \int_{\Omega} \nabla \cdot \mathbf{v}(x) \, dx \]
\[ \int_{\partial \Omega} \epsilon(s) \mathbf{E}(s) \cdot ds = \int_{\Omega} \nabla \cdot (\epsilon(x) \mathbf{E}(x)) \, dx \]

- Poisson’s equation: divergence of the electric field equals the charge density

\[ \int_{\Omega} \nabla \cdot (\epsilon(x) \mathbf{E}(x)) \, dx = \int_{\Omega} \frac{\rho(x)}{\epsilon_0} \, dx \]
\[ \nabla \cdot (\epsilon(x) \mathbf{E}(x)) = \frac{\rho(x)}{\epsilon_0} \]
Poisson equation: structural elements

- Charge distribution & boundary conditions: solute atom positions and charges
- Dielectric function: solute atom radii, positions; solvent radius; polarizabilities
- Assumptions: linear and local response; no mobile ions

\[-\nabla \cdot \epsilon(\mathbf{x}) \nabla \phi(\mathbf{x}) = \rho(\mathbf{x}) \quad \text{for} \; \mathbf{x} \in \Omega\]

\[\phi(\mathbf{x}) = \phi_0(\mathbf{x}) \quad \text{for} \; \mathbf{x} \in \partial \Omega\]
Poisson equation: energies

- Total energies obtained from:
  - Integral of polarization energy
  - Sum of charge-potential interactions

\[ G[\phi] = \frac{1}{4\pi} \int \left\{ \rho(x)\phi(x) - \frac{\epsilon(x)}{2} [\nabla \phi(x)]^2 \right\} dx \]

\[ = -\frac{1}{8\pi} \int \epsilon(x) [\nabla \phi(x)]^2 dx \]

\[ = -\frac{1}{8\pi} \int \rho(x)\phi(x)dx = -\frac{1}{8\pi} \sum_i q_i \phi(x_i) \]
The Born ion

- What is the energy of transferring a non-polarizable ion from between two dielectrics?
- Consider as a “charging process”
  - Free energy for charging a sphere in solvent and vacuum
  - No polar energy for transferring the uncharged sphere to solvent

Image from Dill textbook.
Another route:
- Integrate polarization for dielectric media
- Assume ion is non-polarizable
- Subtract energies between media

\[
G_i = \frac{\varepsilon_0}{2} \int_{\text{solvent}} \varepsilon_i [\nabla \phi_i(x)]^2 \, dx
\]

\[
= \frac{\varepsilon_0}{2} \int_0^\infty \varepsilon_i \left( -\frac{q}{4\pi \varepsilon_0 \varepsilon_i r^2} \right)^2 4\pi r^2 \, dr
\]

\[
= \frac{q^2}{8\pi \varepsilon_0 \varepsilon_i a}
\]

\[
\Delta G = G_2 - G_1
\]

\[
= \frac{q^2}{8\pi \varepsilon_0 a} \left( \frac{1}{\varepsilon_2} - \frac{1}{\varepsilon_1} \right)
\]
A continuum description of ion

- Two nonpolarizable ions
  - Solve for polar energy as a function of separation
  - Poison equation
- Increase in energy as water is “squeezed” out
  - Desolvation effect
  - Smaller volume of polarized water
- Important points
  - Non-superposition of ion potentials
  - Reaction field causes repulsion at short distances
  - Dielectric medium “focuses” field
Simple treatment of ionic interactions

It is rather difficult to assess the magnitude of the contribution made by ionic interactions to the stability of the native structure. Let us consider the free energy change involved in bringing into contact two oppositely charged ions (see Kauzmann, 1959). We will calculate the difference in free energy $\Delta G$ between a final state, in which the charges have approached each other to within the sum of their ionic radii (when they make contact), and an initial state, in which the charges are far enough separated that their interaction energy is negligible. This free energy change is merely the work done on the system in bringing the charges together. For opposite charges this work is negative and is given by

$$\Delta G = - \int_x^a (z_1 z_2 e^2 / e r^2) \, dr$$  \hspace{1cm} (5-15)

$$z_1 z_2 e^2 / ea$$  \hspace{1cm} (5-16)

where $r$ is the variable distance of separation, $z_1$ and $z_2$ are the ionic valencies of the two ions, $e$ is the electronic charge, and $\varepsilon$ is the dielectric constant. The parameter $a$ is the contact distance and is equal to the sum of the ionic radii.

Equation 5-16 is only an approximate expression, of course. One shortcoming, among others, is that it was derived on the assumption of constant $\varepsilon$, but it is very probable that $\varepsilon$ varies considerably with $r$, because the microscopic environment between the two charges changes as they approach each other. If we assume $z_1 = -z_2 = 1$, and $\varepsilon = 80$ (which is the dielectric constant of water), then $\Delta G$ is approximately $-1$ kcal mole$^{-1}$ for $a = 4$ Å. As the charges approach each other near the surface of a protein, however, some lines of force will pass through the protein itself rather than through the solvent. This will tend to lower the effective dielectric constant and will thus increase the magnitude of the net free energy change. On the other hand, the initial separation of charges in a protein in an unfolded state might not be so great as to give rise to even a negligible electrostatic interaction energy. The change in free energy for forming the salt bridge starting from this unfolded state would therefore be correspondingly reduced.

The enthalpy change associated with ion-pair formation can be calculated from Equation 5-16, assuming that $a$ is temperature insensitive.

$$\Delta H = - T^2 [d(\Delta G/T)/dT] = (z_1 z_2 e^2 / ea) + T(z_1 z_2 e^2 / e^2 a)(d\varepsilon/dT)$$

$$= (\Delta G/e)[d(\varepsilon T)/dT]$$  \hspace{1cm} (5-17)

The entropy change is given by

$$\Delta S = - d\Delta G/dT = (z_1 z_2 e^2 / e^2 a)(d\varepsilon/dT)$$

$$= (\Delta G/e)(d\varepsilon/dT)$$  \hspace{1cm} (5-18)
Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics

W. Clark Still,* Anna Tempczyk, Ronald C. Hawley, and Thomas Hendrickson

Department of Chemistry, Columbia University
New York, New York 10027

Received February 12, 1990

Dealing with solvent has been a perpetual problem for molecular modeling. While using explicit solvent molecules provides one solution to the problem, such calculations require a major computational effort if converged energies are required. Here we describe a more practical alternative in which solvent is treated as a statistical continuum. The treatment provides both energies and derivatives analytically and thus may be used in a molecular mechanics force field. As we will show, it gives small-molecule hydration energies of comparable accuracy to those obtained from contemporary free energy perturbation methods but at only a fraction of the computational expense.

Method. We consider solvation free energy ($G_{sol}$) traditionally as consisting of a solvent–solvent cavity term ($G_{cav}$), a solute–solvent van der Waals term ($G_{vdW}$), and a solute–solvent electrostatic polarization term ($G_{pol}$):

$$G_{sol} = G_{cav} + G_{vdW} + G_{pol} \quad (1)$$
Noting that $G_{\text{sol}}$ for the saturated hydrocarbons in water is linearly related\textsuperscript{2} to solvent-accessible surface area (SA), we follow precedent\textsuperscript{1a,b} by setting

$$G_{\text{cav}} + G_{\text{vdW}} = \sum \sigma_k S A_k$$

(2)

where $SA_k$ is the total solvent-accessible surface area of atoms of type $k$ and $\sigma_k$ is an empirical atomic solvation parameter. As a preliminary value for $\sigma_k$, we use +7.2 cal/(mol Å\textsuperscript{2}) for all atom types to reproduce hydration energies of simple hydrocarbons using our recently described analytical surface area calculation.\textsuperscript{3}

For $G_{\text{pol}}$, the total electrostatic free energy ($G_{\text{es}}$, kcal/mol) of a system of widely separated particles (separations $r$ (Å), charges $q$ (electrons), radii $\alpha$ (Å)) in a medium of dielectric constant $\epsilon$ is given (eq 3) by the sum of Coulomb's law in a dielectric (term 1) and the Born equation (term 2). Term 1 can be expanded algebraically (eq 4) to give Coulomb's law in vacuo and a new second term which accounts for the effect of the dielectric medium on the pairwise interactions of charged particles. The sum of terms 2 and 3 in eq 4 is equal to $G_{\text{pol}}$ and has been termed the generalized Born (GB) equation.\textsuperscript{4} The similar form of terms 2 and 3 in eq

$$G_{\text{es}} = 332\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q_i q_j}{r_{ij} \epsilon} - 166 \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^{n} \frac{q_i^2}{\alpha_i}$$

(3)

$$= 332\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q_i q_j}{r_{ij}} - 332 \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^{n-1} \sum_{i=i+1}^{n} \frac{q_i q_j}{r_{ij}} - 166 \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^{n} \frac{q_i^2}{\alpha_i}$$

(4)

$$G_{\text{pol}} = -166 \left(1 - \frac{1}{\epsilon}\right) \sum_{i=1}^{n} \sum_{j=1}^{n} f_{\text{GB}}$$

(5)

.. 4 prompts us to combine them into a single expression (eq 5) where we define $f_{\text{GB}}$ as a function of $r_{ij}$ and $\alpha_i$ which makes eq 5 mimic the relevant equations of classical electrostatics. While we have not defined $f_{\text{GB}}$ uniquely, one simple but effective expression is $f_{\text{GB}} = (r_{ij}^2 + \alpha_{ij}^2 e^{-D})^{0.5}$ where $\alpha_{ij} = (\alpha_i \alpha_j)^{0.5}$ and $D = r_{ij}^2/(2\alpha_{ij})^2$.\textsuperscript{4}
<table>
<thead>
<tr>
<th>solute</th>
<th>$G_{pol}$ kcal/mol</th>
<th>eq 5</th>
<th>$G_{sol}$ kcal/mol</th>
<th>eqs 2 + 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>methanol</td>
<td>$-7.1 \pm 0.3$</td>
<td></td>
<td>$-5.1^{b,d}$</td>
<td>$-6.2$</td>
</tr>
<tr>
<td>ethanol</td>
<td>$-7.4 \pm 0.2$</td>
<td></td>
<td>$-5.0^{b}$</td>
<td>$-5.2$</td>
</tr>
<tr>
<td>2-propanol</td>
<td>$-7.1 \pm 0.3$</td>
<td></td>
<td>$-4.8^{b}$</td>
<td>$-4.3$</td>
</tr>
<tr>
<td>acetone</td>
<td>$-4.5 \pm 0.2$</td>
<td></td>
<td>$-3.9^{b} - 3.8^{d}$</td>
<td>$-3.2$</td>
</tr>
<tr>
<td>dimethyl ether</td>
<td></td>
<td></td>
<td>$-1.9^{b}$</td>
<td>$-2.0$</td>
</tr>
<tr>
<td>methyl acetate</td>
<td>$-3.1 \pm 0.2$</td>
<td></td>
<td>$-3.3^{b}$</td>
<td>$-2.2$</td>
</tr>
<tr>
<td>acetic acid</td>
<td>$-6.8 \pm 0.3$</td>
<td></td>
<td>$-6.7^{d,d}$</td>
<td>$-6.5$</td>
</tr>
<tr>
<td>N,N-dimethylacetamide</td>
<td>$-7.5 \pm 0.5$</td>
<td></td>
<td>$-9.7^{b}$</td>
<td>$-10.6$</td>
</tr>
<tr>
<td>acetamide</td>
<td>$-10.8 \pm 0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Z)-N-methylacetamide</td>
<td>$-10.1 \pm 0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E)-N-methylacetamide</td>
<td>$-7.4 \pm 0.6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(E)-N-methylacetamide dimer</td>
<td>$-4.6 \pm 0.2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alanine dipeptide (C7$_{eq}$)</td>
<td>$-13.1 \pm 0.4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alanine dipeptide (C5)</td>
<td>$-12.3 \pm 0.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alanine dipeptide ($\alpha_R$)</td>
<td>$-20.9 \pm 0.7$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetonitrile</td>
<td></td>
<td></td>
<td>$-3.9^{d}$</td>
<td>$-4.8$</td>
</tr>
<tr>
<td>methyl mercaptan</td>
<td>$-1.9 \pm 0.3$</td>
<td></td>
<td>$-1.2^{b} - 1.3^{d}$</td>
<td>$-3.3$</td>
</tr>
<tr>
<td>benzene</td>
<td>$-1.9 \pm 0.3$</td>
<td></td>
<td>$-0.9^{b} 0.8^{c}$</td>
<td>$-1.0$</td>
</tr>
<tr>
<td>toluene</td>
<td>$-0.7 \pm 0.2$</td>
<td></td>
<td>$-0.9^{b} - 0.8^{c}$</td>
<td>$-0.1$</td>
</tr>
<tr>
<td>phenol</td>
<td>$-8.0 \pm 0.7$</td>
<td></td>
<td>$-6.6^{b}$</td>
<td>$-6.3$</td>
</tr>
<tr>
<td>pyridine</td>
<td></td>
<td></td>
<td>$-4.7^{d}$</td>
<td>$-4.4$</td>
</tr>
<tr>
<td>imidazole</td>
<td></td>
<td></td>
<td>$-5.9^{d,8}$</td>
<td>$-6.5$</td>
</tr>
<tr>
<td>ammonium ion</td>
<td>$-103.9^{d} \pm 0.7$</td>
<td></td>
<td>$-79^{d}$</td>
<td>$-90.8$</td>
</tr>
<tr>
<td>methylammonium ion</td>
<td>$-71^{b} - 70^{d}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trimethylammonium ion</td>
<td>$-59^{d}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acetate ion</td>
<td>$-91.3^{d} \pm 1.4$</td>
<td></td>
<td>$-80^{b} - 77^{d}$</td>
<td>$-82.9$</td>
</tr>
<tr>
<td>ammonium acetate</td>
<td>$-47.7 \pm 0.9$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethane</td>
<td>0.0</td>
<td></td>
<td>$+1.8^{c}$ $\pm 2.1^{d}$</td>
<td>$+1.3$</td>
</tr>
<tr>
<td>$n$-butane (anti)</td>
<td>0.0</td>
<td></td>
<td>$+2.3^{c}$</td>
<td>$+1.9$</td>
</tr>
<tr>
<td>$n$-hexane (all anti)</td>
<td>0.0</td>
<td></td>
<td>$+2.6^{c}$</td>
<td>$+2.4$</td>
</tr>
<tr>
<td>cyclohexane (chair)</td>
<td>0.0</td>
<td></td>
<td>$+1.2^{c}$</td>
<td>$+1.8$</td>
</tr>
<tr>
<td>$N$-octane (all anti)</td>
<td>0.0</td>
<td></td>
<td>$+2.9^{c}$</td>
<td>$+2.9$</td>
</tr>
</tbody>
</table>

A continuum description of ion desolvation
Poisson-Boltzmann

- Simplifies to Debye-Hückel theory
- Continuum dielectric (Poisson equation)
- Non-correlated implicit ions (mean field theory)
- Limitations:
  - Low ion concentration
  - Low ion valency
  - No specific interactions: ion-solute, ion-ion, ion-solvent, solute-solvent, ...
Poisson-Boltzmann derivation: Step 1

- Start with Poisson equation to describe solvation and electrostatics
- Supplement biomolecular charge distribution with mobile ion term

\[-\nabla \cdot \epsilon(x) \nabla \phi(x) = \rho_f(x) + \rho_m(x)\]
Choose mobile ion distribution form
- Boltzmann distribution implies no ion-ion correlation
- Apparent lack of normalization implies grand canonical ensemble
- No detailed structure for ion desolvation

Result: nonlinear partial differential equation

Don’t forget boundary conditions!

\[-\nabla \cdot \epsilon(x) \nabla \phi(x) = \rho_f(x) + \sum_m q_m c_m e^{-\beta[q_m \phi(x) + V_m(x)]}\]
Equation coefficients: “fixed” charge distribution

\[-\nabla \cdot \epsilon(x) \nabla \phi(x) = \rho_f(x) + \sum_m q_m c_m e^{-\beta[q_m \phi(x) + V_m(x)]}\]

- Charges are modeled as delta functions: hard to represent
- Often discretized as splines to “smooth” the problem
- Higher-order charge distributions also possible
Equation coefficients: mobile ion distribution

\[-\nabla \cdot \epsilon(x) \nabla \phi(x) = \rho_f(x) + \sum_{m} q_m c_m e^{-\beta[q_m \phi(x) + V_m(x)]}\]

- Usually assume a single exclusion function for all ions
- Generally based on inflated van der Waals radii
Equation coefficients: dielectric function

\[-\nabla \cdot \epsilon(x) \nabla \phi(x) = \rho_f(x) + \sum_m q_m c_m e^{-\beta[q_m \phi(x) + V_m(x)]}\]

- Describes change in local polarizability
  - Low dielectric interior (2-20)
  - High dielectric exterior (80)
- Many definitions
  - Molecular
  - Solvent-accessible
  - van der Waals
  - Smoothed (Gaussian, spline)
- Results can be very sensitive to surface definition!
PB special cases: symmetric electrolyte

- Assume similar steric interactions for each species with solute
- Simplify two-term exponential series to hyperbolic sine

\[
\rho_m(x) = qce^{-\beta[q\phi(x)+V(x)]} - qce^{-\beta[-q\phi(x)+V(x)]} \\
= qce^{-\beta V(x)} \left[ e^{-\beta q\phi(x)} - e^{\beta q\phi(x)} \right] \\
= -2qce^{-\beta V(x)} \sinh [\beta q\phi(x)] \\
= -\kappa^2(x) \sinh [\beta q\phi(x)]
\]

\[-\nabla \cdot \epsilon(x) \nabla \phi + \kappa^2(x) \sinh [\beta q\phi(x)] = \rho_f(x)\]
PB special cases: linearization

- Assume similar steric interactions for each species with solute
- Assume very small local electrostatic energies
- Taylor series expansion of exponential
- Bulk solution electroneutrality

\[ \rho_m(x) = \sum_m q_m c_m e^{-\beta[q_m \phi(x) + V_m(x)]} \]

\[ \approx e^{-\beta V(x)} \sum_m q_m c_m \left[ 1 - \beta q_m \phi(x) \right] \]

\[ = - \left[ \beta e^{-\beta V(x)} \sum_m q_m^2 c_m \right] \phi(x) \]

\[ = - \kappa^2(x) \phi(x) \]

\[-\nabla \cdot \epsilon(x) \nabla \phi + \kappa^2(x) \phi(x) = \rho_f(x)\]
Poisson-Boltzmann energies

- Similar to Poisson equation
- Functional: integral of solution over domain
- Solution extremizes energy
- Basis for calculating forces: charge-field, dielectric boundary, osmotic pressure

\[
G[\phi] = \frac{1}{4\pi} \int_{\Omega} \left\{ \rho_f(x) \phi(x) - \frac{\epsilon(x)}{2} [\nabla \phi(x)]^2 + \sum_m c_m \left[ e^{-\beta[q_m \phi(x) + V_m(x)]} - 1 \right] \right\} dx \\
\approx \frac{1}{4\pi} \int_{\Omega} \left\{ \rho_f(x) \phi(x) - \frac{\epsilon(x)}{2} [\nabla \phi(x)]^2 + \frac{\kappa^2(x)}{2} [\phi(x)]^2 \right\} dx \\
= -\frac{1}{8\pi} \int_{\Omega} \rho_f(x) \phi(x) dx
\]
Poisson-Boltzmann equation

\[-\nabla \cdot \epsilon(x) \nabla \phi(x) = \rho_f(x) + \sum_m q_m c_m e^{-\beta [q_m \phi(x) + V_m(x)]}\]

\[G[\phi] = \frac{1}{4\pi} \int_\Omega \left\{ \rho_f(x) \phi(x) - \frac{\epsilon(x)}{2} [\nabla \phi(x)]^2 + \sum_m c_m e^{-\beta V_m(x)} \left[ e^{-\beta q_m \phi(x)} - 1 \right] \right\} dx\]

\[F_i[\phi] = -\frac{\partial G[\phi]}{\partial r_i} = -\frac{1}{4\pi} \int_\Omega \left\{ \frac{\partial \rho_f(x)}{\partial r_i} \phi(x) - \frac{1}{2} \frac{\partial \epsilon(x)}{\partial r_i} [\nabla \phi(x)]^2 + \sum_m c_m \frac{\partial e^{-\beta V_m(x)}}{\partial r_i} \left[ e^{-\beta q_m \phi(x)} - 1 \right] \right\} dx\]

- Reaction field
- Dielectric boundary
- “Osmotic”
Polar solvation (implicit)

- Charging free energies
  - Solvent: dielectric effects through Poisson equation
  - Ions: mean-field screening effects through Poisson-Boltzmann equation
Nonpolar solvation (implicit)

- It’s not just surface area!
- WCA formalism:
  - Cavity creation
  - Small length scales: proportional to volume (pressure) and area (surface tension)
  - Large length scales: proportional to area (surface tension)
- Dispersive interactions
  - Modeled by WCA formalism
  - Integral of potential over solvent-accessible volume

Nonpolar solvation: implementation

\[ W^{(np)}(x) = \gamma A(x; \sigma) + pV(x; \sigma) + \bar{\rho} \int_{\Omega} g_0(x, y; \sigma) U^{(np)}_{att}(x, y; \sigma) dy \]

\[ F^{(np)}_i(x) = -\gamma \frac{\partial A(x; \sigma)}{\partial x_i} - p \frac{\partial V(x; \sigma)}{\partial x_i} - \bar{\rho} \int_{\Omega} g_0(x, y; \sigma) \frac{\partial U^{(np)}_{att}(x, y; \sigma)}{\partial x_i} dy \]

Nonpolar solvation: implementation

\[ W^{(np)}(x) = \gamma A(x; \sigma) + pV(x; \sigma) + \rho \int_{\Omega} g_0(x, y; \sigma) U^{(np)}_{att}(x, y; \sigma) dy \]

\[ F_i^{(np)}(x) = -\gamma \frac{\partial A(x; \sigma)}{\partial x_i} - p \frac{\partial V(x; \sigma)}{\partial x_i} - \rho \int_{\Omega} g_0(x, y; \sigma) \frac{\partial U^{(np)}_{att}(x, y; \sigma)}{\partial x_i} dy \]

Putting it all back together

Applications of continuum electrostatics
Electrostatics software

- APBS (http://apbs.sf.net/)
  - PB electrostatics calculations
  - Fast finite element (FEtk) and multigrid (PMG) solvers from Holst group
  - Web-based interface (Gemstone, Baldridge group & NBCR)
  - Works with most popular visualization software (VMD, PMV, PyMOL)
  - Links with CHARMM, AMBER, TINKER
- PDB2PQR (http://pdb2pqr.sf.net/)
  - Prepares PDB files for other calculations
  - Assigns titration states (PROPKA) and optimizes hydrogen positions
  - “Repairs” missing heavy atoms
  - Web-based and command-line

[Image of a molecular structure with a software interface overlay]
Preparing for an electrostatics calculation

- PDB2PQR (http://pdb2pqr.sf.net/)
  - Parameter assignment
  - Heavy atom “repair”
  - Hydrogen bond optimization
  - Titration state calculation (PROPKA)
  - Hydrogen addition
- Also available as standalone tool
Visualization and analysis of electrostatic potentials
The interaction of AChE (structure: center, blue; electrostatic potential: left) with its inhibitor FAS2 (structure: center, red; electrostatic potential: right) is electrostatically driven. Blue surfaces denote positive potential isocontours; red surfaces denote negative.

The electrostatic potential and structure of two cAMP-dependent kinases: 1FOT:A (left) and 1JBP:E (right). The two proteins share only 51% sequence identity and adopt different conformations but still exhibit the same electrostatic potential motifs, share the same fold, and perform the same basic biological function.
Non-specific screening effects

mAChE at 0 mM NaCl

mAChE at 150 mM NaCl
Inspection of ligand binding sites

- NikR Ni(II) and DNA binding
Quantitative comparison of electrostatic potentials

- Do electrostatic potentials tell us anything about biomolecular function?
  - Ligand binding
  - Active sites or shifted pKₐs?
  - Structural (de)stabilization?


Multiresolution contour trees
Thermodynamics
Free energy cycles

- At the heart of most calculations...
- ...because we can’t usually directly calculate the quantity of interest
- Most important principle:
  - Energy is a state function
  - Integral of energy changes over a closed cycle is zero

$$\Delta G_{A\rightarrow B} + \Delta G_{B\rightarrow C} + \Delta G_{C\rightarrow D} + \Delta G_{D\rightarrow A} = 0$$
\[ \Delta G_{\text{sol}} - \Delta G_{\text{elec},s} - \Delta G_{\text{vdw},s} - \Delta G_{\text{cav}} + \Delta G_{\text{vdw},v} + \Delta G_{\text{elec},v} = 0 \]

\[ \Delta G_{\text{sol}} = \Delta G_{\text{elec},s} - \Delta G_{\text{elec},v} + \Delta G_{\text{vdw},s} - \Delta G_{\text{vdw},v} + \Delta G_{\text{cav}} \]
Solvation energies

- Absolute energies are generally not useful: inaccurate
- Solvation: an excellent way to remove “self energies”
- Building block for most electrostatics calculations
- “Total energies” can be recovered by adding in vacuum polar and nonpolar contributions

\[-\epsilon_h \nabla^2 \phi_h(x) = \rho(x)\]
\[-\nabla \cdot \epsilon_{ih} \nabla \phi_{ih}(x) = \rho(x)\]

\[\Delta_{\text{solv}} G = G[\phi_{ih}] - G[\phi_h]\]
\[= -\frac{1}{8\pi} \sum_i [\phi_{ih}(x_i) - \phi_h(x_i)]\]

\[\epsilon(x) = \epsilon_{\text{solv}}\]
\[\epsilon(x) = \epsilon_{\text{mol}}\]
Absolute binding free energy cycle

\[ \Delta G_1 + \Delta G_2 - \Delta G_3 - \Delta G_4 = 0 \]

\[ \Delta G_3 = \Delta G_1 + \Delta G_2 - \Delta G_4 \]
Separate calculation into two steps:
- Calculate electrostatic interaction for homogeneous dielectric (Coulomb’s law)
- Calculate solvation energy change upon binding

Self-interactions are removed in solvation energy calculation

Absolute binding energies are tricky...

$$\Delta G_3 = \Delta G_1 + (\Delta G_2 - \Delta G_4)$$

$$= \Delta_{\text{solv}} G + \Delta_{\text{coul}} G$$
Polar binding energy (PMF): two ions

- Water dielectric
- Two ions: 3 Å radii, non-polarizable, opposite charges
- Basic calculation:
  - Calculate solvation energies of isolated ions
  - Calculate solvation energy of “complex”
  - Subtract solvation energies
  - Add vacuum Coulomb’s law
Polar binding energy: how-to

- **Method #1** (allows for conformational change)
  - Calculate solvation energies for complex and isolated components. Use focusing as needed.
  - Subtract to calculate solvation energy change upon binding.
  - Calculate Coulombic energies for complex and isolated components *using same internal dielectric constant!* Subtract to calculate Coulombic energy change upon binding.
  - Add solvation and Coulombic energy changes.

- **Method #2** (fast but dangerous!)
  - Calculate absolute energies for complex and isolated components. Using focusing as needed. *Use the same grid, dielectric, etc. parameters for all calculations***
  - Subtract.
Relative binding free energy cycle

- Usually better accuracy
- Cancellation of numerical errors
- Cancellation of hard-to-quantify terms
- Useful for predicting mutations, changes in functional groups, etc.

\[
\Delta G_1 + \Delta G_2 - \Delta G_3 - \Delta G_4 = 0
\]

\[
\Delta \Delta G = \Delta G_1 - \Delta G_3 = \Delta G_4 - \Delta G_2
\]
Binding energy example

- Protein kinase A inhibition by balanol
- Continuum electrostatics analysis of protein mutations and functional group changes on binding affinity
Application to ribosomes

- Ribosome central to protein synthesis machinery
- Target for several pharmaceuticals
- Nucleoprotein composition make it computationally challenging
- Composed of two subunits (large and small):
  - 30S consists of 88,000 atoms and roughly 200 Å cube
  - 50S consists of more than 95,000 atoms and roughly 200 Å cube
- Function involves several interesting features:
  - Protein-nucleic acid association
  - Protein-protein association
  - Conformational changes
  - Salt dependence (type and quantity)
- Solved on 343 processors of Blue Horizon to 0.41 Å (30S) and 0.43 Å (50S) resolution

Ribosome-antibiotic binding

- Determine binding energies between 30S ribosomal subunit and aminoglycoside antibiotics
- Good agreement for experimental and computational relative binding free energies: 0.78 ± 0.13 slope with small molecules, 0.95 ± 0.19 slope without
- Suggests importance of basic groups on Ring IV

Application to microtubules

- Important cytoskeletal components: structure, transport, motility, division
- Typically 250-300 Å in diameter and up to millimeters in length
- Computationally difficult due to size (1,500 atoms Å\(^{-1}\)) and charge (-4.5 e Å\(^{-1}\))
- Solved LPBE at 150 mM ionic strength on 686 processors for 600 Å-long, 1.2-million-atom microtubule
- Resolution to 0.54 Å for largest calculation: quantitative accuracy

Microtubule stability and assembly

- Collaboration with Andy McCammon (UCSD) and Dave Sept (Wash U BME)
- Performed series of calculations on tubulin dimers and protofilament pairs
- Poisson-Boltzmann electrostatics and SASA apolar energies
- Observed 7 kcal/mol stronger interactions between protofilaments than within
- Determined energetics for helix properties; predict correct minimum for experimentally-observed A (52 Å) and B (8-9 Å) lattices

pKₐ calculations

- Want acid dissociation constant for residues in a particular structural context
- Use “model” pKₐs for amino acids
- Calculate “intrinsic” pKₐ from two calculations:
  - Binding of unprotonated residue
  - Binding of protonated residue
- Calculate titration state and actual from sampling of coupled pKₐs
- Conformational distributions can matter

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>α-carboxylic acid</th>
<th>α-amino</th>
<th>Side chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>2.35</td>
<td>9.87</td>
<td></td>
</tr>
<tr>
<td>Arginine</td>
<td>2.01</td>
<td>9.04</td>
<td>12.48</td>
</tr>
<tr>
<td>Asparagine</td>
<td>2.02</td>
<td>8.80</td>
<td></td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>2.10</td>
<td>9.82</td>
<td>3.86</td>
</tr>
<tr>
<td>Cysteine</td>
<td>2.05</td>
<td>10.25</td>
<td>8.00</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>2.10</td>
<td>9.47</td>
<td>4.07</td>
</tr>
<tr>
<td>Glycine</td>
<td>2.35</td>
<td>9.78</td>
<td></td>
</tr>
<tr>
<td>Histidine</td>
<td>1.77</td>
<td>9.18</td>
<td>6.10</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>2.23</td>
<td>9.76</td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td>2.33</td>
<td>9.74</td>
<td></td>
</tr>
<tr>
<td>Lysine</td>
<td>2.18</td>
<td>8.95</td>
<td>10.53</td>
</tr>
<tr>
<td>Methionine</td>
<td>2.28</td>
<td>9.21</td>
<td></td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>2.58</td>
<td>9.24</td>
<td></td>
</tr>
<tr>
<td>Proline</td>
<td>2.00</td>
<td>10.60</td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>2.21</td>
<td>9.15</td>
<td></td>
</tr>
<tr>
<td>Threonine</td>
<td>2.09</td>
<td>9.10</td>
<td></td>
</tr>
<tr>
<td>Tryptophan</td>
<td>2.38</td>
<td>9.39</td>
<td></td>
</tr>
<tr>
<td>Tyrosine</td>
<td>2.20</td>
<td>9.11</td>
<td>10.07</td>
</tr>
<tr>
<td>Valine</td>
<td>2.29</td>
<td>9.72</td>
<td></td>
</tr>
</tbody>
</table>
pK$_a$ calculations
Conformational changes: two conformations

- Same idea as binding free energies
- Calculate polar energy change due to conformational change in homogeneous dielectric (Coulomb’s law)
- Calculate polar solvation energy change due to conformation change in inhomogeneous dielectric
- Subtract.
Conformational change: multiple conformations

- **MM/PBSA**: include contribution from multiple conformations to energy
- Typically used for binding energy
- Accounts for conformational distribution effects on
  - Intra- and intermolecular energy (mechanics)
  - Solvation (Poisson-Boltzmann and apolar)
  - Entropy (quasi-harmonic)

\[
Z_i = \int e^{-\beta[U(x_i) + W(x_i)]} \, dx_i \approx z^\text{int}_i e^{-\beta\langle U(x_i) + W(x_i) \rangle}
\]

\[
\Delta G = -\frac{1}{\beta} \log \left( \frac{Z_{AB}}{Z_A Z_B} \right) \approx -\frac{1}{\beta} \log \left( \frac{z^\text{int}_{AB}}{z^\text{int}_A z^\text{int}_B} \right) + \langle U_{AB}(x_i) + W_{AB}(x_i) \rangle
\]

\[
- \langle U_A(x_i) + W_A(x_i) \rangle - \langle U_B(x_i) + W_B(x_i) \rangle
\]
MM-PBSA: computational alanine scanning

- Examine the interface of oncprotein MDM2 with N-terminus of tumor suppressor p53
- Apply MM-PBSA methods with normal mode entropies
- Surprisingly good results!
MM-PBSA: RNA-ligand interactions

- Calculate binding free energy of theophylline to RNA 33-mer
- Use normal mode entropy calculation
- Compare with thermodynamic integration
- Reasonable agreement between computational (-7.5 kcal/mol) and experimental (-9.0 kcal/mol) binding energies
- Pretty good relative binding free energies