Foundations of biomaterials: Models of protein solvation

L. Ridgway Scott

The Institute for Biophysical Dynamics, The Computation Institute, and the Departments of Computer Science and Mathematics, The University of Chicago
Polar liquids play a singular role in:

- Protein interactions (water most important molecule in biology)
- Ionic solvation (ionic liquids)
- Environmental pollution (HCl hydrate)
- Energy production (methane hydrate)
- Batteries
- Photovoltaic cells

Dielectric effect critical for protein interactions.

Picoscale models needed.
Modeling issues for solvation

- Solvents are mobile, not fixed in orientation
- Nonlocal effects (frequency dependence): high dimension
- Nonlinear models: do they achieve same results as nonlocal ones?
- Ionic effects: can size effects be modeled via continuum equations? E.g., $NaCl$
- Algorithms to resolve nonlinear (e.g., ionic) interactions
- Need molecular scale models of solvation
Protein sidechains have large electrostatic gradients

Figure 1: Different models suggest different modes of interactions. Shown are two nonlocal dielectric models with (A) and without (B) ionic effects.
Water is complicated

Water network from a molecular dynamics simulation
(a) Polarity of water molecule

(b) Ability to rotate allows water to align with electric field $e$

Result: water screens electric field
Competing effects: why this is so hard

**Protein sidechains have large electrostatic gradients**

**Water is a strong dielectric**

**Hydrophobic groups modify the water structure**

---

Figure 2: Three competing effects that determine protein behavior. These conspire to weaken interactive forces, making biological relationships more tenuous and amenable to mutation.
Charges in a dielectric are like lights in a fog.
Consider two charge distributions $\rho$ (fixed charges) and $\gamma$ (polar groups free to rotate). Resulting electric potential $\phi$ satisfies

$$\Delta \phi = \rho + \gamma,$$

where the dielectric constant of free space is set to one.

Write $\phi = \phi_\rho + \phi_\gamma$, where $\Delta \phi_\gamma = \gamma$ and $\Delta \phi_\rho = \rho$.

Ansatz of Debye [3]: the electric field $e_\gamma = \nabla \phi_\gamma$ is parallel to (opposing) the resulting electric field $e = \nabla \phi$:

$$\nabla \phi_\gamma = (1 - \varepsilon) \nabla \phi.$$  \hspace{1cm} (2)

Thus $\nabla \phi_\rho = \nabla \phi - \nabla \phi_\gamma = \varepsilon \nabla \phi$ and

$$\nabla \cdot (\varepsilon \nabla \phi) = \rho.$$  \hspace{1cm} (3)
Polarization field and Debeye’s Ansatz as projection

Define \( \mathbf{p} = \nabla \phi \gamma \): called the polarization field.

Recall \( \mathbf{e} = \nabla \phi \).

Write \( \mathbf{p} = (\varepsilon - \varepsilon_0)\mathbf{e} + \zeta \mathbf{e}^\perp \), so that

\[
\varepsilon = \varepsilon_0 + \frac{\mathbf{p} \cdot \mathbf{e}}{\mathbf{e} \cdot \mathbf{e}},
\]

with the appropriate optimism that \( \mathbf{p} = 0 \) when \( \mathbf{e} = 0 \).

That is, \( \varepsilon - \varepsilon_0 \) reflects the correlation between \( \mathbf{p} \) and \( \mathbf{e} \).

As defined, \( \varepsilon \) is a function of \( r \) and \( t \), and potentially singular.

However, Debye postulated that a suitable average \( \tilde{\varepsilon} \) should be well behaved:

\[
\tilde{\varepsilon} = \varepsilon_0 + \left\langle \frac{\mathbf{p} \cdot \mathbf{e}}{\mathbf{e} \cdot \mathbf{e}} \right\rangle.
\]
In bulk water $\varepsilon$ is a (temperature-dependent) constant:

$$\varepsilon \approx 87.74 - 40.00 \tau + 9.398 \tau^2 - 1.410 \tau^3, \quad \tau \in [0, 1], \quad (4)$$

where $\tau = T/100$ and $T$ is temperature in Centigrade (for $T > 0$) [4].

$\varepsilon >> 1$: opposing field strength $E_\gamma = \nabla \phi_\gamma$ much greater than inducing field.

$\varepsilon$ increases with decreasing temperature; when water freezes, it increases further: for ice at zero degrees Centigrade, $\varepsilon \approx 92$.

Increased coherence yields increased dielectric
But model fails when the spatial frequencies of the electric field $e = \nabla \phi$ are commensurate with the size of a water molecule, since the water molecules cannot orient appropriately to align with the field.
Manipulations leading to (3) valid when $\varepsilon$ is an operator, even nonlinear.

Frequency-dependent versions of $\varepsilon$ have been proposed, and these are often called ‘nonlocal’ models.

The operator $\varepsilon$ must be represented either as a Fourier integral (in frequency space), or as an integral in physical space with a nonlocal kernel [1, 6].
Debye observed that the effective permittivity is frequency dependent:

\[ \epsilon(\nu) = \epsilon_0 + \frac{\epsilon_1 - \epsilon_0}{1 + \tau_D^2 \nu^2} \]  

(5)

where \( \tau_D \) is characteristic time associated with dielectric material and \( \nu \) is temporal wave wave number. Many experiments have verified this [5]:
Polar residues cause spatial high frequencies

Some polar sidechains: $\pm$ charges at distance 2.2Å
Charged sidechains form salt bridge networks

Arginine

\[
\begin{align*}
&\text{CH}_2 \\
&\text{CH}_2 \\
&\text{CH}_2 \\
&\text{NH} \\
&\text{C} \\
&\text{NH}_2^+ \\
&\text{NH}_2
\end{align*}
\]

Aspartic acid

\[
\begin{align*}
&\text{CH}_2 \\
&\text{CH}_2 \\
&\text{C} \\
&\text{O}^-
\end{align*}
\]

Lysine

\[
\begin{align*}
&\text{CH}_2 \\
&\text{CH}_2 \\
&\text{CH}_2 \\
&\text{NH}_3^+
\end{align*}
\]

Glutamic acid

\[
\begin{align*}
&\text{CH}_2 \\
&\text{CH}_2 \\
&\text{COO}^-
\end{align*}
\]
Survey of results from two papers:

Xie Dexuan, Yi Jiang, Peter Brune, and L. Ridgway Scott.

Xie Dexuan, Yi Jiang, and L. Ridgway Scott.
Efficient algorithms for solving a nonlocal dielectric continuum model for protein in ionic solvent *SISC*, to appear
Linearized Poisson-Boltzmann equation:

\[-\nabla \cdot (\epsilon \nabla \Phi(r)) = \rho\] (6)

In bulk dielectric, \(\epsilon\) is a constant.

But in general it depends on the frequency of \(\Phi\), e.g.

\[\epsilon f = f + K \ast f\] (7)

where

\[\hat{K}(\xi) = \frac{1}{1 + |\xi|^2}.\] (8)

So convolution with \(K\) is the inverse of a PDE.

Can solve using a system of two PDEs.
Figure 3: Comparisons of analytical free energy differences calculated from the nonlocal dielectric model with two values of $\lambda$ ($\lambda = 15\text{Å}$ and $\lambda = 30\text{Å}$) and the values from chemical experiments.
Step 0: Let $G$ be the solution to $\Delta G = \rho$ in all space:

$$G(x) = \sum \frac{q_i}{|x - x_i|}.$$  

Find $u_0 \in H^1_0(\Omega)$ such that

$$a_0(u_0, w) = \ell_0(w) \quad \forall w \in H^1_0(\Omega),$$  

where $\ell_0$ and $a_0$ are linear and bilinear forms as defined by ($\lambda > 0$ is a model parameter)

$$a_0(u_0, w) = \lambda^2 \int_{\Omega} \nabla u_0 \cdot \nabla w \, d\mathbf{r} + \int_{\Omega} u_0 w \, d\mathbf{r},$$

$$\ell_0(w) = \int_{\Omega} G(\mathbf{r})w(\mathbf{r}) \, d\mathbf{r}.$$
For $\phi = (\Phi, u)$ and $\nu = (v_1, v_2)$, define

$$a(\phi, \nu) = \int_{\Omega} \epsilon(r) \nabla \Phi(r) \cdot \nabla v_1(r) \, dr$$

$$+ (\epsilon_s - \epsilon_\infty) \int_{D_s} \nabla u(r) \cdot \nabla v_1(r) \, dr$$

$$+ \lambda^2 \int_{\Omega} \nabla u(r) \cdot \nabla v_2(r) \, dr + \int_{\Omega} (u(r) - \Phi(r)) v_2(r) \, dr,$$

where $\epsilon_p$, $\epsilon_s$, $\epsilon_\infty$ and $\lambda$ are constants, and

$$\epsilon(r) = \begin{cases} 
\epsilon_p, & r \in D_p \text{ (protein)}, \\
\epsilon_\infty, & r \in D_s \text{ (solvent)}, 
\end{cases}$$
Nonlocal variational equations

Find \( \varphi_1 = (\Psi, u_1), \varphi_2 = (\tilde{\Phi}, u_2) \in \mathcal{V} \) such that

\[
\begin{align*}
    a(\varphi_1, \tilde{v}) &= \ell_1(\tilde{v}) \quad \forall \tilde{v} \in \mathcal{V}, \\
    a(\varphi_2, \tilde{v}) &= \ell_2(\tilde{v}) \quad \forall \tilde{v} \in \mathcal{V},
\end{align*}
\]

where \( \ell_1(\tilde{v}) \) and \( \ell_2(\tilde{v}) \) are two linear forms as defined by

\[
\ell_1(\tilde{v}) = (\epsilon_{\infty} - \epsilon_s) \int_{D_p} \nabla u_0(r) \cdot \nabla v_1(r) \, dr + (\epsilon_p - \epsilon_{\infty}) \int_{D_s} \nabla G(r) \cdot \nabla v_1(r) \, dr,
\]

\[
\ell_2(\tilde{v}) = \frac{1}{\epsilon_0} \sum_{i=1}^{n} q_i \int_{D_s} c_i(r) v_1(r) \, dr.
\]
The nonlocal model solution $\Phi$ is given by

$$\Phi = \Psi + \tilde{\Psi} + G$$

Key point: singularity $G$ is added at the end. $\nabla G$ appears in $\ell_1$ only on the solvent domain $D_s$ where there are no fixed charges. The auxiliary variables satisfy

$$u = \Phi \ast Q_\lambda, \quad u_1 = \Psi \ast Q_\lambda, \quad u_2 = \tilde{\Psi} \ast Q_\lambda$$

$$u = u_0 + u_1 + u_2$$
Surface mesh of BPTI
Figure 4: Nonlocal model yields different potential at protein surface.
Polarization field $\nabla \phi_\gamma$ saturates for large fixed fields:

$$\lim_{|\nabla \phi| \to \infty} (1 - \varepsilon) \nabla \phi = \lim_{|\nabla \phi| \to \infty} \nabla \phi_\gamma = C,$$

(15)

One simple model that satisfies (15) is

$$\varepsilon(x) = \varepsilon_0 + \frac{\varepsilon_1}{1 + \lambda |\nabla \phi(x)|}$$

(16)

for some constants $\varepsilon_0$, $\varepsilon_1$, and $\lambda$.

Both the nonlocal and nonlinear models of the dielectric response have the effect of representing frequency dependence of the dielectric effect. $|\nabla \phi(x)|$ provides a proxy for frequency content, although it will not reflect accurately high-frequency, low-power electric fields.

Combination of nonlocal and nonlinear dielectric models may be needed.
Hydrophobic ($\text{CH}_n$) groups remove water locally.

This causes a reduction in $\varepsilon$ locally.

(Resulting increase in $\phi$ makes dehydrons sticky.)

This can be quantified and used to predict binding sites.

The placement of hydrophobic groups near an electrostatic bond is called wrapping.

Like putting insulation on an electrical wire.

(Wrapping modifies dielectric effect)

We can see this effect on a single hydrogen bond.
Wrapping protects hydrogen bond from water

Well wrapped hydrogen bond

Underwrapped hydrogen bond
Hydrogen bonds (B) that are not protected from water do not persist.

From De Simone, et al., PNAS 102 no 21 7535-7540 (2005)
Dynamics of dehydrons

Dynamics of hydrogen bonds and wrapping

Figure 5: Distribution of bond lengths for two hydrogen bonds formed in a structure of the sheep prion [2]. Horizontal axis measured in nanometers, vertical axis represents numbers of occurrences taken from a simulation with 20,000 data points with bin widths of 0.1 Ångstrom. Distribution for the well-wrapped hydrogen bond (H3) has smaller mean value but a longer (exponential) tail, whereas distribution for the underwrapped hydrogen bond (H1) has larger mean but Gaussian tail.
Ligand binding removes water

Binding of ligand changes underprotected hydrogen bond (high dielectric) to strong bond (low dielectric)

No intermolecular bonds needed!
Dehydrons


Well-wrapped hydrogen bonds are grey, and dehydrons are green.

The standard ribbon model of “structure” lacks indicators of electronic environment.
Mathematical explanation

Charges $\rho$ induce an electric field $e = \nabla \phi$ given by

$$\nabla \cdot (\epsilon \nabla \phi) = \nabla \cdot (\epsilon e) = \rho$$

Energy $= - \int \rho \phi \, dx$

Hydrophobicity affects the operator $\epsilon$: removing water reduces $\epsilon$.

When $\epsilon$ goes down, $\phi$ goes up.

Hydrophilic groups contribute to the right-hand side $\rho$.

Hydrophobicity and hydrophilic are orthogonal, not opposites.
The HIV protease has a dehydron at an antibody binding site.

When the antibody binds at the dehydron, it wraps it with hydrophobic groups.
A model for protein-protein interaction

Foot-and-mouth disease virus assembly from small proteins.
Dehydrons guide binding of component proteins VP1, VP2 and VP3 of foot-and-mouth disease virus.
Extreme interaction: amyloid formation

Standard application of bioinformatics: look at distribution tails. If some is good, more may be better, but too many may be bad. Too many dehydrons signals trouble: the human prion.

Genetic code minimizes changes of polarity due to single-letter codon mutations, but it facilitates changes in wrapping due to single-letter codon mutations.

<table>
<thead>
<tr>
<th>First Position</th>
<th>Second Position</th>
<th>Third Position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>u</strong>&lt;br&gt;uuc&lt;br&gt;uug&lt;br&gt;uuuaug</td>
<td><strong>c</strong>&lt;br&gt;ucu&lt;br&gt;ucc&lt;br&gt;uca&lt;br&gt;ucg</td>
<td><strong>a</strong>&lt;br&gt;uau&lt;br&gt;uac&lt;br&gt;uag&lt;br&gt;ugc</td>
</tr>
<tr>
<td>Leu 4</td>
<td></td>
<td>Ser 0 + −</td>
</tr>
<tr>
<td><strong>c</strong>&lt;br&gt;cuc&lt;br&gt;cug&lt;br&gt;cua&lt;br&gt;cug</td>
<td><strong>a</strong>&lt;br&gt;ccu&lt;br&gt;ccc&lt;br&gt;cca&lt;br&gt;ccg</td>
<td><strong>g</strong>&lt;br&gt;cau&lt;br&gt;cac&lt;br&gt;cag&lt;br&gt;cgc</td>
</tr>
<tr>
<td>Leu 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>a</strong>&lt;br&gt;auc&lt;br&gt;aug&lt;br&gt;auc&lt;br&gt;aug</td>
<td><strong>g</strong>&lt;br&gt;acu&lt;br&gt;acc&lt;br&gt;aca&lt;br&gt;acc</td>
<td><strong>c</strong>&lt;br&gt;aau&lt;br&gt;aac&lt;br&gt;aag&lt;br&gt;aag</td>
</tr>
<tr>
<td>Ile 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>g</strong>&lt;br&gt;guu&lt;br&gt;guc&lt;br&gt;gua&lt;br&gt;gug</td>
<td><strong>a</strong>&lt;br&gt;gcu&lt;br&gt;gcc&lt;br&gt;gca&lt;br&gt;gcc</td>
<td><strong>c</strong>&lt;br&gt;gau&lt;br&gt;gac&lt;br&gt;gag&lt;br&gt;gag</td>
</tr>
<tr>
<td>Val 3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

First digit after residue name is amount of wrapping. Second indicator is polarity; ||: nonpolar, + − : polar, − − : negatively charged, + + : positively charged.
Drug ligand wrapping

Drug ligand provides additional non-polar carbonaceous group(s) in the desolvation domain, enhancing the wrapping of a hydrogen bond.

- $\text{CH}_n$ (n= 1,2,3)
- Carbonyl O
- Amide N

intramolecular wrapping:

\[ \rho = 15 \]
Desolvation spheres for flap Gly-49–Gly-52 dehydron containing nonpolar groups of the wrapping inhibitor.
Aligned paralogs

Aligned backbones for two paralog kinases; dehydrons for Chk1 are marked in green and those for Pdk1 are in red.
Selective wrapper

Dehydron Cys673-Gly676 in C-Kit is not conserved in its paralogs Bcr-Abl, Lck, Chk1 and Pdk1. By methylating Gleevec at the para position (1), the inhibitor becomes a selective wrapper of the packing defect in C-Kit.
Phosphorylation rates from spectrophotometric assay on the five kinases Bcr-Abl (blue), C-Kit (green), Lck (red), Chk1 (purple), and Pdk1 (brown) with Gleevec (triangles) and modified Gleevec methylated at positions (1) and (2) (squares). Notice the selective and enhanced inhibition of C-Kit.
Conclusions

Some advances in solvation modeling

- Now tractable to compute nonlocal dielectric models
- Simple models (wrapping/dehydrons) give consistent predictions at picoscale
- Have been used to aid drug design
Thanks

This talk was based on joint work with

- Peter Brune (Argonne Nat. Lab.), Yi Jiang and Dexuan Xie (UWisconsin-Milwaukee)
- Steve Berry (U. Chicago), Ariel Fernández (Calderon Inst., Argentina), Chris Fraser (Bioanalytical Computing), Kristina Rogale Plazonic (Princeton), Harold Scheraga (Cornell)

We thank the Institute for Biophysical Dynamics at the University of Chicago and NSF for support.

We are also grateful to the developers of the PDB, Viper, DIP, and other biological data bases.
References


