

July 8, 2004

**Protein electrostatics and solvation:
a preliminary sketch**

Thomas Simonson

Laboratoire de Biochimie (CNRS UMR 7654)

Departement of Biology

Ecole Polytechnique

91128 Palaiseau, France.

thomas.simonson@polytechnique.fr

<http://bioinfo.polytechnique.fr>

1 Some elementary electrostatics

1.1 Charges in vacuum

Force between two charges Coulomb showed experimentally that the force exerted by one charge on another (Fig. 1.1) is:

$$\mathbf{F} = \frac{qq'}{|\mathbf{r} - \mathbf{r}'|^2} \mathbf{u}$$

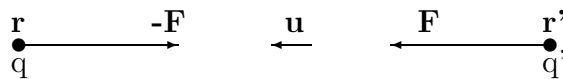


Figure 1: The charge q at \mathbf{r} exerts a force \mathbf{F} on q' located at \mathbf{r}' , and vice versa.

Electric field due to a charge By definition, the field \mathbf{E} at \mathbf{r}' is the “force per unit charge”:

$$\mathbf{E} = \mathbf{F}/q' = \frac{q}{|\mathbf{r} - \mathbf{r}'|^2} \mathbf{u}$$

The field “radiates” in all directions from q :

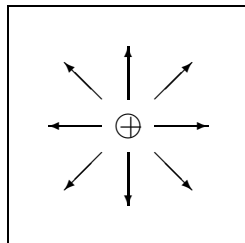


Figure 2: The field lines diverging from a positive point charge *in vacuo*.

We notice that the field \mathbf{E} obeys the equation

$$\operatorname{div} \mathbf{E}(\mathbf{r}') = 4\pi\rho(\mathbf{r}'),$$

where ρ is the charge density corresponding to the source q , namely the Dirac delta function:

$$\rho(\mathbf{r}') = q\delta(\mathbf{r}' - \mathbf{r})$$

and “div” is the divergence operator: $\operatorname{div} = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$

Energy and electrostatic potential From Newtonian mechanics, the force derives from a potential energy U :

$$\mathbf{F} = -\operatorname{grad} U$$

By analogy, we define the electrostatic potential ϕ :

$$\mathbf{E} = -\operatorname{grad} \phi$$

The energy has the form

$$U = \frac{qq'}{|\mathbf{r}' - \mathbf{r}|}$$

The potential produced by q at \mathbf{r}' is therefore:

$$\phi = \frac{q}{|\mathbf{r}' - \mathbf{r}|}$$

The case of multiple charges: the superposition principle A single charge q_1 at \mathbf{r}_1 produces at \mathbf{r}_0 the potential $\phi_1(\mathbf{r}_0) = q/|\mathbf{r}_0 - \mathbf{r}_1|$. If there are two charges, q_1 and q_2 , the potential is the sum of the individual potentials: $\phi = \phi_1 + \phi_2$: The potentials **add up**. For n charges,

$$\phi(\mathbf{r}_0) = \sum_{i=1}^n \frac{q_i}{|\mathbf{r}_0 - \mathbf{r}_i|}$$

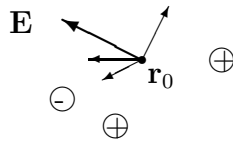


Figure 3: The field at \mathbf{r}_0 is the superposition of three individual fields.

$$\mathbf{E}(\mathbf{r}_0) = \sum_{i=1}^n \frac{q_i}{|\mathbf{r}_0 - \mathbf{r}_i|^2} \mathbf{u}_i$$

This is the **superposition principle**.

The field obeys the same equation as before:

$$\operatorname{div} \mathbf{E} = 4\pi\rho,$$

known as **Poisson's equation**.

The **energy** takes the form

$$U = \frac{1}{2} \sum_{ij} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}$$

1.2 Uniform dielectric medium

Dielectric screening An example of a dielectric medium is a simple ionic crystal (Fig. 3). There are charged groups, but they are held in place by crystal packing and cannot move very much (unlike a conductor). Suppose we apply an external field. The electronic cloud around each ion becomes deformed, or ‘stretched’:

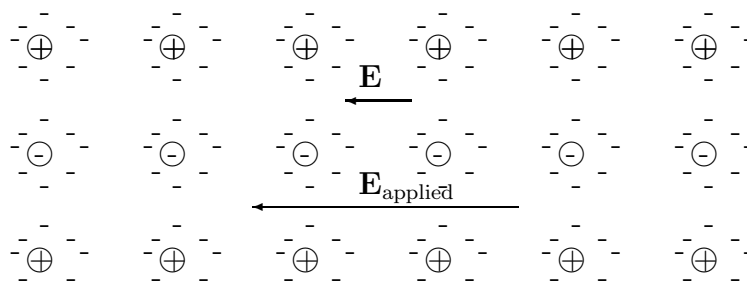


Figure 4: Deformation of the electronic cloud by an applied electric field.

The total field is reduced, or “screened”: $\mathbf{E} = \alpha \mathbf{E}_{\text{applied}}$, $\alpha < 1$.

This situation can be described approximately by an **induced dipole density \mathbf{P}** (Fig. 4):

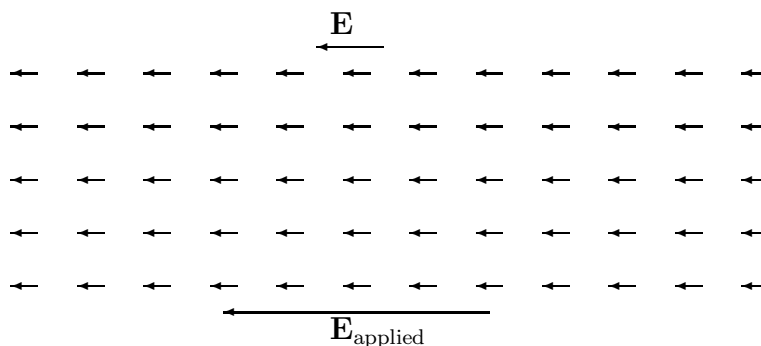


Figure 5: A simplified, dipolar view of the above.

Dielectric constant For small applied fields, the total field \mathbf{E} and the induced dipole density \mathbf{P} are both proportional to the applied field, and therefore to each other:

$$\mathbf{P} = \frac{\epsilon - 1}{4\pi} \mathbf{E}$$

This defines the **dielectric constant**, ϵ . It measures the induced polarization and dielectric screening.

The field from a uniform dipole density is $-4\pi\mathbf{P}$, so that $\mathbf{E} = \mathbf{E}_{\text{applied}} - 4\pi\mathbf{P}$ and

$$\mathbf{E}_{\text{applied}} = \epsilon\mathbf{E}$$

Not only the electron cloud, but also the permanent dipoles of small molecular groups can be polarized by external fields or charges:

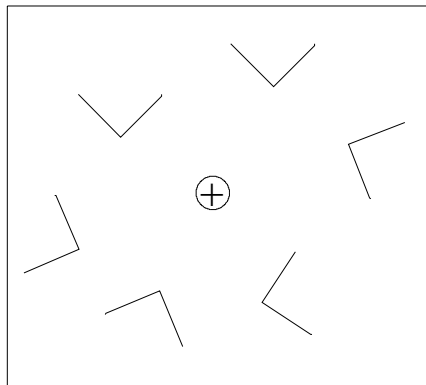


Figure 6: A point charge screened by water-like molecular dipoles.

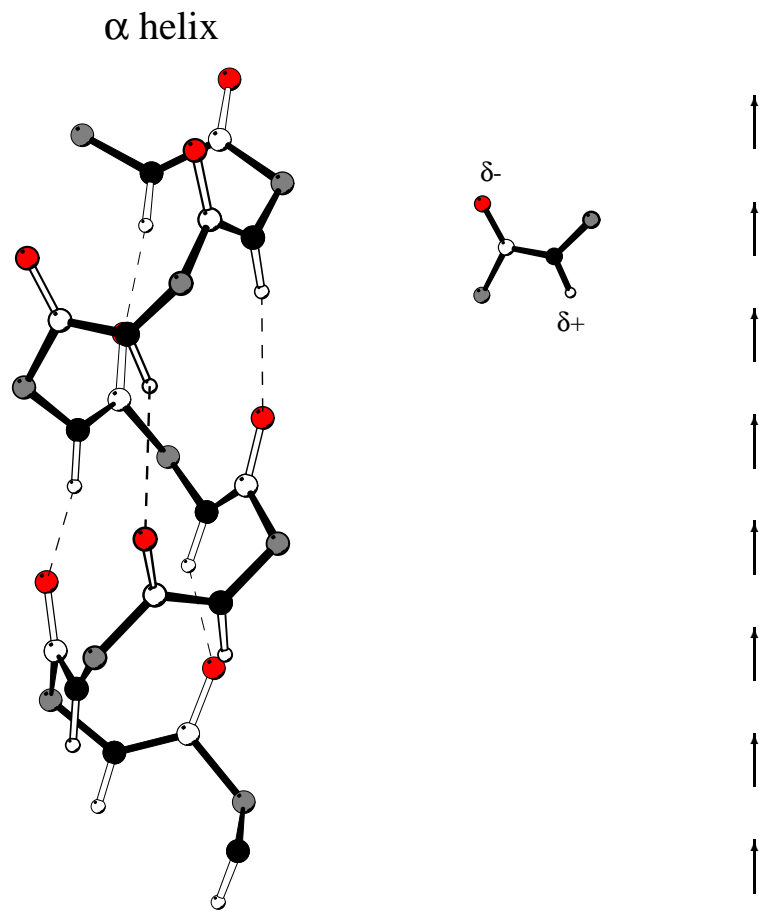


Figure 7: The peptide group idealized as a dipole.

Table 1: Some typical dielectric constants

molecule	dipole moment (Debyes) of a single molecule	dielectric constant ϵ of the liquid at 20° C
water	1.9	80
ethanol	1.7	24
acetic acid	1.7	4
acetamide	3.8	41

Poisson's equation Suppose the applied field $\mathbf{E}_{\text{applied}}$ is produced by a charge density ρ_f . Because of the induced polarization, the total field is reduced, or screened: $\mathbf{E} = \beta\mathbf{E}_{\text{applied}}$, $\beta < 1$. We saw above that $\beta = 1/\epsilon$:

$$\mathbf{E} = \mathbf{E}_{\text{applied}}/\epsilon$$

As a result, \mathbf{E} obeys a modified **Poisson's equation**:

$$\text{div } \mathbf{E} = 4\pi\rho_f/\epsilon$$

The work of polarization A charged particle q polarizes the dielectric (Fig. 4), so that there is a reaction potential ϕ_r on q . ϕ_r is proportional to q :

$$\phi_r = Cq$$

To calculate the work W to charge up the particle from zero to q , we introduce a fictitious intermediate state, with charge λq , $0 \leq \lambda \leq 1$.

$$W = \int_0^1 \phi_r(\lambda)q d\lambda = \int_0^1 \lambda Cq^2 d\lambda = \frac{1}{2}Cq^2 = \frac{1}{2}q\phi_r.$$

We have only included the interactions of q with the dielectric. There is also a direct interaction between the charge increment, $q d\lambda$ and the charge already present, λq . For a point charge, this interaction is infinite. For a very narrow charge, it is large but finite. It is a constant, which does not depend on the environment of the charge. Therefore, it can be ignored for most problems of interest.

The work to assemble a group of n charges can be shown to be

$$W = \frac{1}{2} \sum_{i=1}^n q_i \phi_i,$$

where ϕ_i is the total potential at site i . W is the **free energy** of the assembly.

2 Systems with dielectric boundaries

Poisson's equation The general form of Poisson's equation is:

$$\operatorname{div} \epsilon \mathbf{E} = 4\pi \rho_f \quad (1)$$

The charge density ρ_f is the “fixed” density that gives rise to the “applied” field. This field polarizes the dielectric, so that the total field \mathbf{E} is reduced, or screened.

Planar boundary with a uniform applied field Consider a dielectric, bounded by a plane, in a uniform applied field:

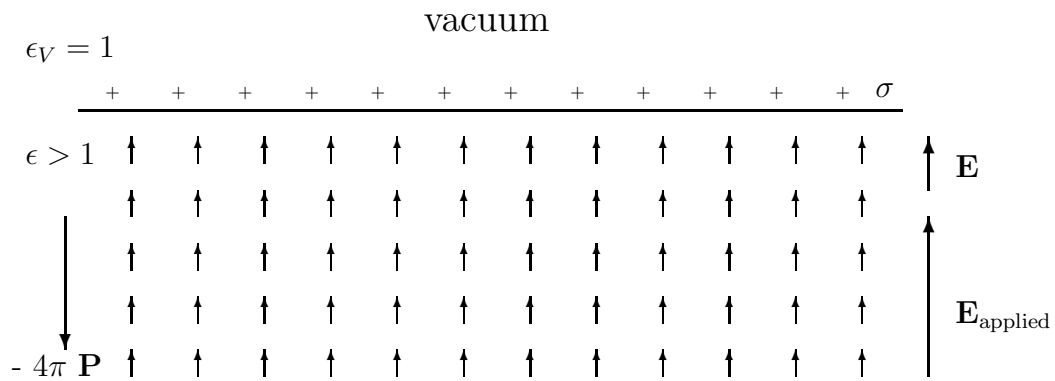


Figure 8: Planar dielectric interface.

The dipole density is terminated at the boundary, leaving an excess surface charge density σ . The total field in the dielectric is uniform: $\mathbf{E} = \mathbf{E}_{\text{applied}}/\epsilon$.

Planar boundary with a single charge Again, there is induced charge at the boundary, but it is no longer uniform. It has the same effect as an “image” charge, a single charge placed symmetrically on the other side of the boundary.

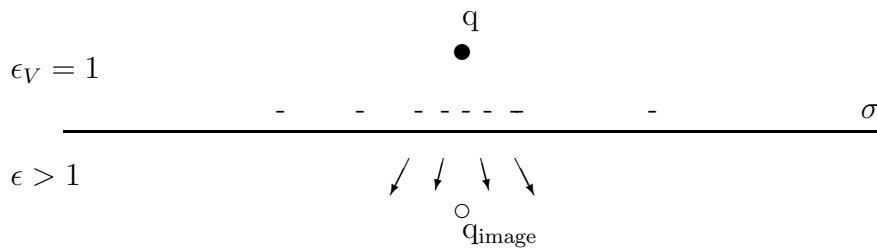


Figure 9: A single charge near a planar dielectric interface.

A spherical charge in a dielectric Outside the sphere (Fig. 10),

$$\operatorname{div} \epsilon \mathbf{E} = \frac{1}{r^2} \frac{\partial r^2 \epsilon E_r}{\partial r} = 4\pi \delta(r)$$

$$\implies E_r = \frac{C}{\epsilon r^2} \quad \text{outside}; \quad = \frac{C}{r^2} \quad \text{inside}$$

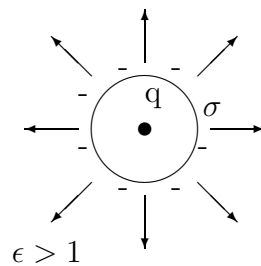


Figure 10: A single charge in a spherical cavity.

Induced charge We have seen that applied fields can give rise to induced polarization \mathbf{P} and induced charge ρ_i (eg, at the dielectric boundary, Fig. 8). In general,

$$\begin{aligned}\operatorname{div} \epsilon \mathbf{E} &= 4\pi \rho_f \\ \operatorname{div} \mathbf{E} &= 4\pi \rho = 4\pi(\rho_f + \rho_i)\end{aligned}$$

Using $\operatorname{div} \epsilon \mathbf{E} = \epsilon \operatorname{div} \mathbf{E} + \mathbf{E} \cdot \operatorname{grad} \epsilon$, we obtain

$$\rho_i = \left(\frac{1 - \epsilon}{\epsilon} \right) \rho_f + \frac{1}{4\pi \epsilon} \mathbf{E} \cdot \operatorname{grad} \epsilon$$

There is induced charge “on top” of ρ_f and wherever ϵ varies (at boundaries):

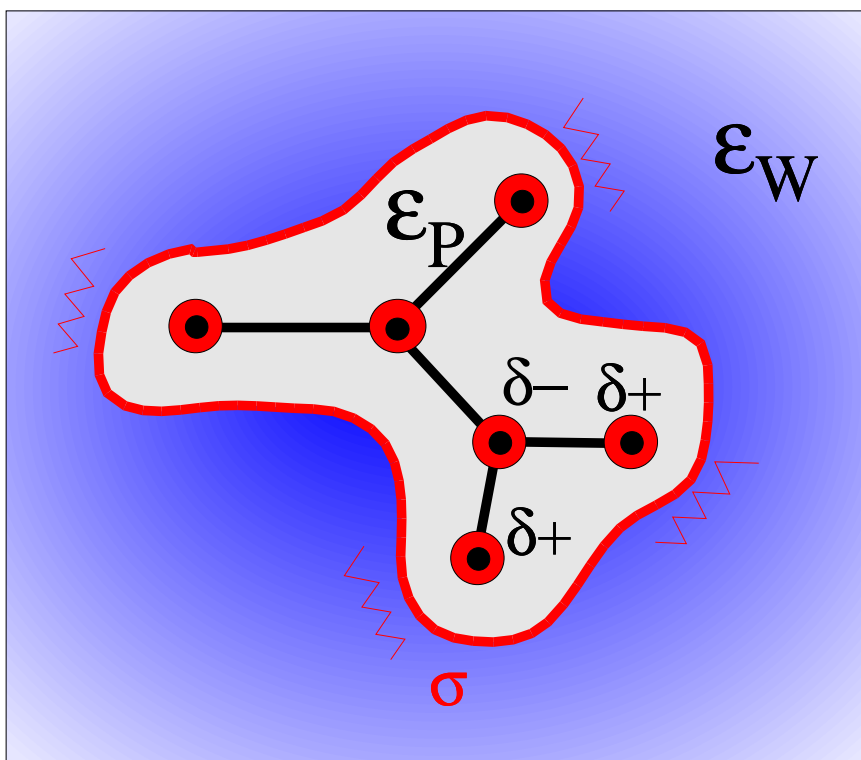


Figure 11: A (very!) small protein idealized as a dielectric body, containing partial charges on each atom (ρ_f). There is induced charge on top of each atom and smeared over the solute/solvent boundary. The surface charge (σ) is shown as “squiggles”.

Free energy The earlier expression is valid in the general case:

$$W = \frac{1}{2} \sum_{i=1}^n q_i \phi_i,$$

3 Solving Poisson's equation numerically

Taking the potential as the main variable, Poisson's equation has the form

$$\text{div } \epsilon \text{ grad } \phi = -4\pi \rho_f$$

The functions ϕ , ρ_f and ϵ can be discretized using a large cubic lattice (Fig. 12). Writing the div and grad operators as finite-differences, we obtain a system of equations, one for each grid node i :

$$\phi_i = \frac{\sum_j \epsilon_{ij} \phi_j + 4\pi q_i}{\sum_j \epsilon_{ij}}, \quad (2)$$

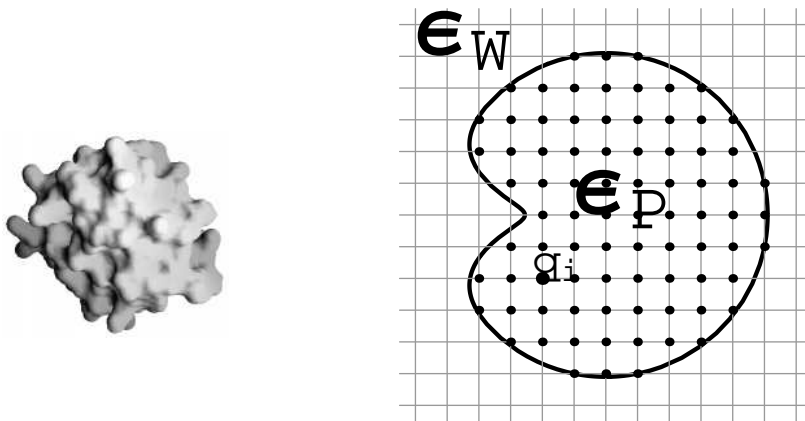


Figure 12: The solute (left) is bounded by its molecular surface. The space surrounding it is paved with a cubic lattice (right).

4 Applications

Innumerable applications fill the literature; see below; a sprinkling of topics includes:

- ligand binding
- pH-dependent stability
- enzyme activity
- protein–protein interactions
- electron transfer

5 Some references

- T Simonson (2004) Electrostatics and dynamics of proteins. *Reports on Progress in Physics*, 66:737–787.
- T Simonson, G Archontis, M Karplus (2002) Protein-ligand recognition: free energy simulations come of age. *Accounts of Chemical Research*, 35:430–435.
- B Roux, T Simonson, Eds. (1999) Implicit solvent models. *Biophys. Chem.*, 1-278.
- D Bashford, D Case (2000) Generalized Born models of macromolecular solvation effects. *Ann. Rev. Phys. Chem.*, 51:129
- M Schaefer, H van Vlijmen, M Karplus (1998) Electrostatic contributions to molecular free energies in solution. *Adv. Prot. Chem.*, 51:1-57.
- SY Noskov, C Lim (2001) Free energy decomposition of protein-protein interactions. *Biophys. J.*, 81:737.
- Michael Gilson. Introduction to continuum electrostatics. <http://gilsonlab.umbi.umd.edu>

6 Implicit solvent models

6.1 The concept of potential of mean force

A protein in solution occupies a conformation X with the Boltzmann probability:

$$P(X, Y) = \frac{e^{-U(X, Y)/kT}}{\int dX dY e^{-U(X, Y)/kT}} \quad (3)$$

where: X = coordinates of the solute (protein); Y = coordinates of the solvent.

The potential energy can be split into three terms:

$$U(X, Y) = U_u(X) + U_v(Y) + U_{uv}(X, Y) \quad (4)$$

where:

$U_u(X)$ = solute–solute interactions

$U_v(Y)$ = solvent–solvent interactions

$U_{uv}(X, Y)$ = solute–solvent interactions.

For any physical quantity $Q(X, Y)$, the mean value is:

$$\langle Q \rangle = \int dX dY Q(X, Y) P(X, Y) \quad (5)$$

We are primarily interested in the protein's behavior, not the solvent's. The probability distribution for the protein is:

$$\bar{P}(X) = \int dY P(X, Y) \quad (6)$$

By analogy to (3), we write:

$$\bar{P}(X) \stackrel{\text{def}}{=} \frac{e^{-W(X)/kT}}{\int dX e^{-W(X)/kT}} \quad (7)$$

$W(X)$ is called the **potential of mean force**. Its formal expression follows from:

$$e^{-W(X)/kT} = \int dY e^{-[U_u(X) + U_v(Y) + U_{uv}(X, Y)]/kT}. \quad (8)$$

We can also write

$$W(X) \stackrel{\text{def}}{=} U_{uu}(X) + \Delta W(X). \quad (9)$$

ΔW accounts **implicitly but exactly** for the solvent's effect on the protein. The main goal of an implicit solvent model is to approximate ΔW accurately and efficiently.

6.2 Surface area model (Eisenberg & McLachlan, 1986)

Eisenberg and McLachlan proposed:

$$\Delta W(X) = \sum_i \gamma_i A_i$$

The sum is over the protein atoms. γ_i is an empirical atomic 'surface tension'
 A_i is the solvent-accessible surface area of atom i

6.3 'Gaussian' solvent models EEF1, EEF2 of Lazaridis and Karplus

See T. Lazaridis, M. Karplus (1999) Effective energy function for proteins in solution, *Proteins*, 35, 133–152.

6.4 Dielectric continuum solvent model

The most promising implicit solvent models treat the solvent as a dielectric continuum. The protein atoms play the role of source charges, corresponding to the charge density ρ_f (Eq. 1). Poisson's equation has the form:

$$\nabla \epsilon E = 4\pi \rho_f \quad \text{Poisson}$$

The potential of mean force has the form:

$$W_{\text{elec}} = \frac{1}{2} \sum_i q_i \Phi_i \quad (10)$$

From the superposition principle, Φ_i can be written

$$\Phi_i = \sum_j q_j \Phi_{j \rightarrow i}$$

where $q_j \Phi_{j \rightarrow i}$ is the potential on i produced by q_j . One can easily show that

$$\Phi_{j \rightarrow i} = \Phi_{i \rightarrow j} \stackrel{\text{def}}{=} F(i, j)$$

so that

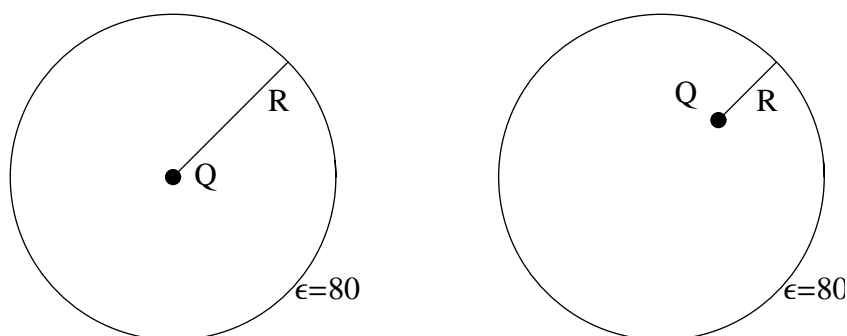
$$W_{\text{elec}} = \frac{1}{2} \sum_{ij} q_i q_j F(i, j), \quad (11)$$

with $F(i, j) = F(j, i)$. F is entirely determined by the geometry of the dielectric boundaries (eg the protein/solvent boundary; see Fig. 11). It can be obtained by numerically solving Poisson's equation (see above), but this is expensive, and the goal of most models is to replace the solution F of Poisson's equation by a more approximate, semi-empirical function.

6.5 The case of a single charge: self energy and Born radius

Consider a cavity with the shape of a protein, embedded in a dielectric medium representing the solvent, of dielectric constant ϵ . Suppose the cavity contains a single source charge Q . This charge polarizes the solvent; the polarization charge interacts with Q . The free energy is referred to as a ‘self-energy’. If the cavity is spherical and Q is at the center (left of figure), it takes the form:

$$W_{\text{self}} = \left(\frac{1}{\epsilon} - 1\right) \frac{Q^2}{2R}$$



If Q is off-center (right of figure), by analogy, we write the self-energy:

$$W_{\text{self}} \stackrel{\text{def}}{=} \left(\frac{1}{\epsilon} - 1\right) \frac{Q^2}{2R_{\text{Born}}} \approx \left(\frac{1}{\epsilon} - 1\right) \frac{Q^2}{2R} \quad (12)$$

The ‘Born radius’ R_{Born} characterizes the position of Q within the cavity (primarily its depth). If the cavity is not spherical (Fig. 13, below), we continue to define R_{Born} by equation (12).

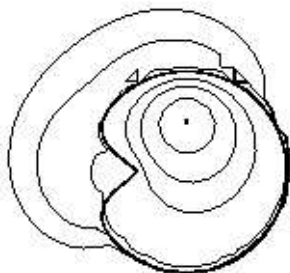


Figure 13: The cavity boundary is bold; the lines are equipotentials; Q is the dot.

6.6 Generalized Born model

The Generalized Born model (GB) approximates F by:

$$F(i, j) \approx F_{GB}(i, j) = \frac{q_i q_j}{r_{ij}} + \left(\frac{1}{\epsilon} - 1\right) q_i q_j f_{GB}(r_{ij}) \quad (13)$$

$$f_{GB}(r_{ij}) = \left(r_{ij}^2 + R_{Born,i} R_{Born,j} \exp[-r_{ij}^2/4R_{Born,i} R_{Born,j}]\right)^{-1/2} \quad (14)$$

The empirical F_{GB} (Fig. 14, below) gives the correct continuum dielectric result when $r_{ij} \rightarrow 0$ and $r_{ij} \rightarrow \infty$.

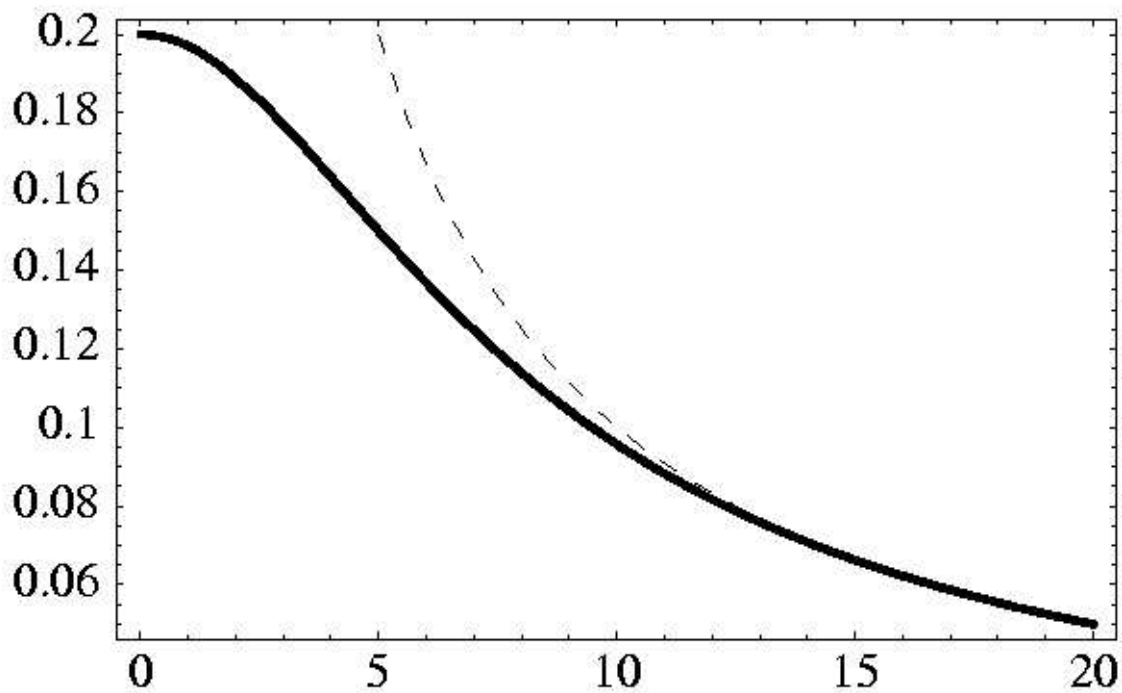


Figure 14: Bold: f_{GB} (with $R_{Born,i} = R_{Born,j} = 5$); dashed: $(\frac{1}{\epsilon} - 1) \frac{1}{r_{ij}}$.

To complete the solvent model, we still need an efficient approximation for the Born radii, $R_{Born,i}$, $R_{Born,j}$.

6.7 Approximation for the Born radius

The idea is to start from the self-energy of Q and to assume that *within* the protein cavity, the field is the same ‘radial’, or ‘Coulombic’ field that would exist in vacuum:

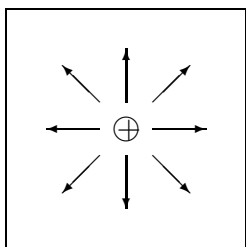
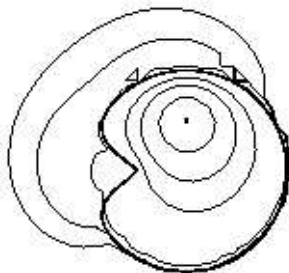


Figure 15: Field lines diverging radially from a positive point charge *in vacuo*.

No approximation is made concerning the field outside the protein. The approximation is equivalent to assuming the equipotential contours around Q in the figure below are circles:



With this approximation, an exact expression for W_{self} is obtained; see eg Schaefer & Karplus, J Phys Chem, 1996, 100:1578. The (cumbersome) exact expression can be fitted by a simpler function:

$$\frac{1}{R_{Born,i}} \approx \frac{1}{b_i} + \sum_{j \neq i} f(r_{ij}, V_j)$$

$$f(r_{ij}, V_j) = \frac{1}{\omega_{ij}} \exp(-r_{ij}^2/\sigma_{ij}^2) + \frac{V_j}{8\pi} \left(\frac{r_{ij}^3}{r_{ij}^4 + \mu_{ij}^4} \right)^4$$

where b_i , ω_{ij} , σ_{ij} , μ_{ij} , V_j are empirical parameters depending on the atomic size.

6.8 Bibliography

Applications are too numerous to begin to list here. Recently, GB has been used successfully to fold several small proteins from scratch. The original GB papers, more details on implicit solvent models, as well as recent applications are reviewed in:

- M Feig and CL Brooks III (2004) Recent Advances in the Development and Application of Implicit Solvent Models in Biomolecule Simulations. *Current Opinion in Structural Biology*, 14:217.
- T Simonson (2001) Macromolecular electrostatics: continuum models and their growing pains. *Current Opinion in Structural Biology*, 11:243.
- B Roux, T Simonson, Eds. (1999) Implicit solvent models. *Biophys. Chem.*, 1–278.
- D Bashford, D Case (2000) Generalized Born models of macromolecular solvation effects. *Ann. Rev. Phys. Chem.*, 51:129
- T Simonson (2004) Electrostatics and dynamics of proteins. *Reports on Progress in Physics*, 66:737–787.