Implicit Solvent Electrostatics in Biomolecular Simulation

Nathan A. Baker[1], Donald Bashford[2] and David A. Case[3]

[1] Dept. of Biochemistry and Molecular Biophysics, Washington University, St. Louis, MO 63110, USA
[2] Dept. of Molecular Biotechnology, St. Jude Childrens Research Hospital, Memphis, TN 38105, USA
[3] Dept. of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Abstract. We give an overview of how implicit solvent models are currently used in protein simulations. The emphasis is on numerical algorithms and approximations: since even folded proteins sample many distinct configurations, it is of considerable importance to be both accurate and efficient in estimating the energetic consequences of this dynamical behavior. Particular attention is paid to calculations of pH-dependent behavior, as a paradigm for the analysis of electrostatic interactions in complex systems.

Key words: Electrostatics, biomolecular simulation, implicit solvent, continuum solvent, Poisson-Boltzmann, Generalized Born

1 Introduction

Computer simulations have become an important method for understanding structure, dynamics, and function of proteins. Models of inter- and intra-molecular energetics are important components of such simulations. These energetic properties are determined by a combination of short- and long-range interactions. Short-range energetics include a variety of contributions such as van der Waals, bonding, angular, and torsional interactions. On the other hand, long-range energetics are predominately governed by electrostatic interactions. Due to their slow decay over distance, electrostatics cannot be neglected in biomolecular modeling as these forces are important at all length scales. Therefore, models which accelerate the evaluation of these interactions can provide important benefits to molecular simulation.

Due to the importance of electrostatics in biomolecular systems, a variety of computational methods have been developed for better understanding these interactions (see [1–6] and references therein). Popular methods for understanding electrostatic interactions in biomolecular systems can be classified as “explicit" methods which treat the solvent in full atomic detail and “implicit" methods which model solvent

influences in a pre-averaged continuum fashion. Explicit solvent methods, by definition, offer a more detailed description of biomolecular solvation; however, they also require integration over numerous solvent degrees of freedom. These extra degrees of freedom dramatically increase computational requirements and can limit the ability to use explicit solvent methods to generate converged estimates of thermodynamic and kinetic observables from a biomolecular simulation.

The sampling issues associated with explicit solvent treatments have necessitated the development of robust implicit solvent approaches. As their name implies, implicit solvent techniques are derived by pre-averaging over the solvent and counterion coordinates. The result of this averaging is a linear and local continuum dielectric model for solvent response and a mean field charge continuum for the counterion distribution. Despite some artifacts in the implicit solvent approach (e.g., see Roux [3] or Holm et al. [7] and references therein), these methods have enjoyed widespread use over recent years due to their significant reduction of complexity in solvated biomolecular systems.

In this chapter, we will focus on implicit solvent methods for biomolecules in water, but the logical first proving ground for such applications is small molecules in water. In the most straightforward approach, the solute is described as a quantum-mechanical system in a region of dielectric dielectric constant 1 (where water cannot penetrate), surrounded by the high-dielectric solvent. This demands self-consistency between the electronic structure and the reaction potential arising from the partial orientation of solute dipoles around the solute. The dependence of the reaction potential on the solute charge distribution and hence, its electronic structure, gives rise to the self-consistency requirement. The earliest applications of this idea date back to Onsager [8] who used a simplified spherical model. Application to more complex molecular shapes has been pioneered by Tomasi and co-workers [9, 10], and comprehensive reviews of this approach are available [10, 11]. The terms, polarized continuum model or self-consistent reaction field are sometimes used to describe such calculations.

The application of continuum model to the *entire* solvent region (including water molecules in direct contact with the solute) seems like a severe approximation, but actually is known to give quite a good account of solvation free energies and pK_a behavior in small, fairly rigid molecules [11–16]. This is undoubtedly due in part to a careful parameterization of the boundary between the solvent and solute, so that the average energetic consequences of even first-shell waters are incorporated into the continuum model. Models for this dividing surface can be made that appear to be transferable, and not overly dependent on the detailed chemical nature of the solute. (This effective boundary varies with temperature in a way that is not easy to model; for this reason, continuum models are much less successful in predicting quantities like solvation enthalpies and entropies.)

A simpler, molecular mechanics, model describes the electrostatic properties of the solute with a set of fixed partial charges, usually centered on the atoms. The essential energetic idea, illustrated in Fig. 1, is to break the process of bringing a molecule from vacuum to solvent into three hypothetical steps: reduction to zero of the molecule's charges in vacuum; solvation of the purely non-polar molecule; and

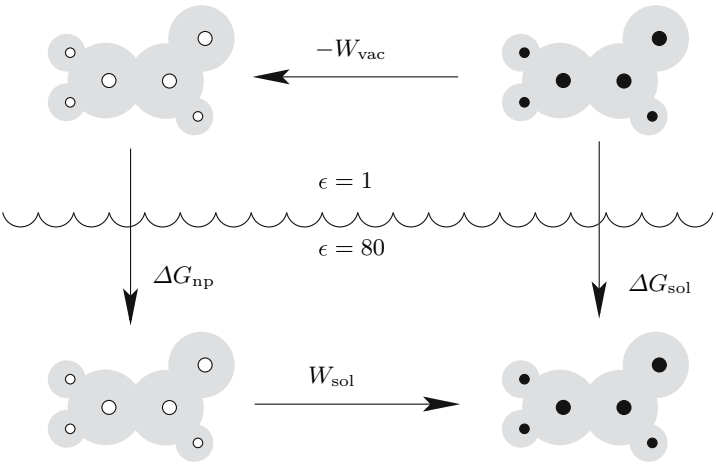

Figure 1. Thermodynamic cycle for computing solvation free energies. Filled circles represent atoms having partial charges, whereas empty circles represent a hypothetical molecule in which the partial charges have been set to zero

restoration of the original partial charges in the solvent environment. The overall solvation free energy is then,

$$\Delta G_{\mathrm{sol}} = \Delta G_{\mathrm{np}} - W_{\mathrm{vac}} + W_{\mathrm{sol}} \,, \tag{1}$$

where W_{vac} and W_{sol} are the work of charging in the vacuum and solvent environment, respectively, and ΔG_{np} is the free energy of solvation of the hypothetical molecule with all partial charges set to zero. For some small molecules, the ΔG_{np} term can be estimated as the solvation energy of a sterically equivalent alkane. More often, empirical formulas relating ΔG_{np} to surface area and/or volume are used [17–20]. Our focus here, however, will be on the electrostatic, or "polarization" component ΔG_{pol} defined by,

$$\Delta G_{\mathrm{pol}} = W_{\mathrm{sol}} - W_{\mathrm{vac}} \,. \tag{2}$$

Implicit solvent models have been used to study biomolecular function for over 80 years, starting with work by Born on ion solvation [21], Linderström-Lang [22] and Tanford and Kirkwood [23] on protein titration, and Flanagan et al on hemoglobin dimer assembly [24]. However, the applicability of these models was dramatically improved in the 1980's with increasing computer power and the availability of new numerical implicit solvent models which accurately described biomolecular geometries. Numerical methods for biomolecular implicit solvent electrostatics were first introduced by Warwicker and Watson [25], who described the numerical solution of the Poisson-Boltzmann (PB) equation to obtain the electrostatic potential in the active site of an enzyme.

In addition to PB approaches, which we will discuss in the next section, further speed advantages may be had by adopting additional approximations. Currently, the most popular of these fall under the category of "generalized Born" models, and

these will be discussed in Sect. 3. It is worth noting, however, that computational efficiency, although an important consideration, is not the only, or even the primary reason, for an interest in implicit solvent models. Here are some additional considerations that make many investigators willing to put up with the inevitable loss of physical realism that arises from replacing explicit solvent with a continuum:

1. There is no need for the lengthy equilibration of water that is typically necessary in explicit water simulations; implicit solvent models correspond to instantaneous solvent dielectric response.
2. Continuum simulations generally give improved sampling, due to the absence of viscosity associated with the explicit water environment; hence, the macromolecule can more quickly explore the available conformational space.
3. There are no artifacts of periodic boundary conditions; the continuum model corresponds to solvation in an infinite volume of solvent.
4. New (and simpler) ways to estimate free energies become feasible; since solvent degrees of freedom are taken into account implicitly, estimating free energies of solvated structures is much more straightforward than with explicit water models [26–28].
5. Implicit models provide a higher degree of algorithm flexibility. For instance, a Monte-Carlo move involving a solvent exposed side-chain would require non-trivial re-arrangement of the nearby water molecules if they were treated explicitly. With an implicit solvent model this complication does not arise.
6. Most ideas about "landscape characterization" (that is, analysis of local energy minima and the pathways between them) make good physical sense only with an implicit solvent model. Trying to find a minimum energy structure of a system that includes a large number of explicit solvent molecules is both difficult and generally pointless: the enormous number of potential minima that differ only in the arrangement of water molecules might easily overwhelm an attempt to understand the important nature of the protein solute.

We now turn to an analysis of computational algorithms that are prominent in models employing continuum solvent ideas.

2 Poisson-Boltzmann Methods

2.1 Derivation

The Poisson-Boltzmann (PB) equation is derived from a continuum model of the solvent and aqueous counterion medium around the biomolecules of interest [1, 3–5, 7]. There are numerous derivation of the PB equation based on statistical mechanics treatments (see Holm et al [7] for a review). However, the simplest derivation begins with Poisson's equation [29–31]:

$$-\nabla \cdot \epsilon(\mathbf{x}) \nabla \phi(\mathbf{x}) = 4\pi\rho(\mathbf{x}) \text{ for } \mathbf{x} \in \Omega \tag{3}$$

$$\phi(\mathbf{x}) = \phi_0(\mathbf{x}) \text{ for } \mathbf{x} \in \partial\Omega \tag{4}$$

which describes the electrostatic potential $\phi(\mathbf{x})$ in an inhomogeneous dielectric medium of relative permittivity $\epsilon(\mathbf{x})$ due to a charge distribution $\rho(\mathbf{x})$. The equation is usually solved in a domain Ω with a Dirichlet condition $\phi_0(\mathbf{x})$ on the domain boundary $\partial\Omega$ that describes the asymptotic behavior of the solution.

For a system without mobile aqueous counterions, the charge distribution includes only "fixed" biomolecular charges, usually modeled as point charges centered on the atoms:

$$\rho_f(\mathbf{x}) = \sum_{i=1}^{M} Q_i \delta\left(\mathbf{x} - \mathbf{x}_i\right) \ , \tag{5}$$

where Q_i are the charge magnitudes and $\mathbf{x}_i$ are the charge positions. For a system with mobile counterions, the charge distribution also includes a term which describes the mean-field distribution of m different species of mobile ions:

$$\rho_m(\mathbf{x}) = e_c \sum_{j=1}^{m} \overline{n}_j z_j \exp\left[-e_c z_j \phi(\mathbf{x})/kT - V_j(\mathbf{x})/kT\right] \tag{6}$$

where $\overline{n}_j$ is the number density of ions of species j (in the absence of an external field), z_j is the valence of ions of species j, e_c is the elementary charge, and V_j is the steric interaction between the biomolecule and ions of species j which prevents overlap between the biomolecular and mobile counterion charge distributions. It is assumed that the macromolecule is at infinite dilution and that ϕ differs from zero only in a region near the macromolecule so that the $\overline{n}_j$ are adequate normalization constants.

Substitution of these two charge densities into the Poisson equation gives the "full" or nonlinear PB equation:

$$\begin{aligned} -\nabla \cdot \epsilon(\mathbf{x}) \nabla \phi(\mathbf{x}) - 4\pi e_c \sum_{j=1}^{m} \overline{n}_j z_j \exp\left[-e_c z_j \phi(\mathbf{x})/kT - V_j(\mathbf{x})/kT\right] \\ = 4\pi \sum_{i=1}^{M} Q_i \delta\left(\mathbf{x} - \mathbf{x}_i\right) \text{ for } \mathbf{x} \in \Omega \end{aligned} \tag{7}$$

$$\phi(\mathbf{x}) = \phi_0(\mathbf{x}) \text{ for } \mathbf{x} \in \partial\Omega \ . \tag{8}$$

This equation can be simplified somewhat for a 1:1 electrolyte such as NaCl where $q_1 = 1$, $q_2 = -1$. Assuming the steric interactions with the biomolecule are the same for all ion species ($V_1 = V_2 = V$), 7 reduces to

$$\begin{aligned} -\nabla \cdot \epsilon(\mathbf{x}) \nabla \phi(\mathbf{x}) + 8\pi e_c \overline{n} e^{-V(\mathbf{x})/kT} \sinh[e_c \phi(\mathbf{x})/kT] \\ = 4\pi \sum_{i=1}^{M} Q_i \delta\left(\mathbf{x} - \mathbf{x}_i\right) \text{ for } \mathbf{x} \in \Omega \end{aligned} \tag{9}$$

This is the canonical form of the nonlinear PB equation, and it is useful to examine how the biomolecular structure enters into each term. The dielectric function $\epsilon(\mathbf{x})$ is related to the shape of the biomolecule; assuming lower values of $2 - 20$ in

the biomolecular interior and higher values proportional to bulk solvent outside the biomolecule. This dielectric function has been represented by a number of models, including discontinuous descriptions [9, 25] based on unions of atomic spheres, or a more complex molecular surface definition [32, 33] as well as smoother spline-based [34] and Gaussian [35] description. Not surprisingly, the results of PB calculations have been shown to be sensitive to the choice of surface definition; different surfaces appear to require different parameter sets [36–38]. The ion accessibility parameter $\overline{\kappa}^2(\mathbf{x})$ (see 2.12) usually varies from a value proportional to the bulk ionic strength outside the biomolecule to a value of zero at distances between the biomolecule and ions that are less than the ionic radii. Like the dielectric function, this accessibility function $\overline{\kappa}^2(\mathbf{x})$ is often implemented as either a smooth or discontinuous function. Finally, the biomolecular data enters explicitly into the fixed charge distribution on the right hand side of the equation via the partial atomic charge magnitudes $\{Q_i\}$ and locations $\{\mathbf{x}_i\}$.

For many systems where the mean-field linear dielectric approximations implicit in the PB equation are valid [7], the nonlinear PB equation can be reduced to a linear equation by assuming $e_c\phi(\mathbf{x})/kT \ll 1$ in the ion-accessible solvent region. In this case, the exponential functions (whether in the $\sinh$ of 9 or the $\exp$ of 7) can be truncated at first order in the Taylor series. By also assuming all steric factors are the same $V_j = V$, this linearization yields the linearized PB equation:

$$-\nabla \cdot \epsilon(\mathbf{x})\nabla\phi(\mathbf{x}) + \overline{\kappa}^2(\mathbf{x})\phi(\mathbf{x}) = 4\pi \sum_{i=1}^{M} Q_i \delta\left(\mathbf{x} - \mathbf{x}_i\right) \text{ for } \mathbf{x} \in \Omega \tag{10}$$

$$\phi(\mathbf{x}) = \phi_0(\mathbf{x}) \text{ for } \mathbf{x} \in \partial\Omega \; . \tag{11}$$

where

$$\overline{\kappa}^2(\mathbf{x}) = e^{-V(\mathbf{x})/kT} \frac{4\pi e_c^2}{kT} \sum_s \overline{n}_s z_s^2 = e^{-V(\mathbf{x})/kT} \frac{8\pi e_c^2 I}{kT} \; , \tag{12}$$

and I is the ionic strength $(1/2)\sum_s \overline{n}_s z_s^2$. Note that the above differs somewhat from the usual expression for κ^2 in Debye–Hueckel theory; in regions where $V(\mathbf{x}) = 0$, and the dielectric constant is that of the solvent, they are related by $\kappa^2 = \overline{\kappa}^2/\epsilon$.

Some of the primary uses of the PB equations are the calculation of electrostatic energies and forces for use in biomolecular simulations. The most intuitive way to calculate energy is through a charging process. The incremental work, δG of adding an increment of charge density, $\delta\rho$ to an already-existing charge distribution that is producing a potential $\phi(\mathbf{x};\rho)$ is $\int \delta\rho(\mathbf{x})\phi(\mathbf{x};\rho)\; d\mathbf{x}$. If the potential is linear in the charge, as in the Poisson or Linearized Poisson–Boltzmann equations, the integration from zero to the full charge distribution gives,

$$G[\phi] = \frac{1}{2}\int_{\Omega} \rho_f \phi d\mathbf{x} \; . \tag{13}$$

This formula suggests a practical procedure for calculating molecular energitics: solve the relevant linear equation for ϕ, and then multiply times the charge distribution through 13 or (more commonly) the analogous sum over point charges. The application of this idea is discussed further in Sect. 4.2.

A deeper approach which is useful for analyzing the mathematical problem of determining the potential, developing approximation schemes, and for the calculation of quantities such as forces is based on the fact that the PB equation is related to an energy functional in ϕ. The PB equation then arises from the requirement that ϕ minimize this functional while satisfying the boundary conditions. In this sense, the PB equation directly defines a free energy functional [39–41]. For the nonlinear PB equation in 9, this functional is:

$$G[\phi] = \int_\Omega \left[\rho_f \phi - \frac{\epsilon}{8\pi} \left(\nabla\phi\right)^2 - 2kT\overline{n}e^{-V/kT} \left(\cosh\left(\frac{e_c\phi}{kT}\right) - 1 \right) \right] d\mathbf{x} \,. \quad (14)$$

This functional involves several physically-intuitive terms, including (in order) the charge-potential interaction energy, the dielectric polarization energy, and the energy required to assemble the counterion distribution. Like the nonlinear PB equation, this can be linearized for small $\phi(\mathbf{x})$ to give the linearized PB equation free energy:

$$G[\phi] = \int_\Omega \left[\rho_f \phi - \frac{\epsilon}{8\pi} \left(\nabla\phi\right)^2 - \frac{\overline{\kappa}^2}{2}\phi^2 \right] d\mathbf{x} \,. \quad (15)$$

One can confirm that this expression is consistent with the work-of-charging expression, 13 through integration by parts and substitution of the original PB equation (10).

The free energy expressions in (14) and (10) can be differentiated with respect to atomic displacements to give electrostatic forces [34, 42]. The saddle-point approximation used to derive the PB equation implies $\delta G[\phi]/\delta\phi = 0$, therefore the force on atom i derived from the nonlinear PB equation is given by differentiation with respect to the atomic coordinate $\mathbf{y}_i$

$$\mathbf{F}_i[\phi] = -\int_\Omega \left[\phi \left(\frac{\partial \rho_f}{\partial \mathbf{y}_i} \right) - \frac{\left(\nabla\phi\right)^2}{8\pi} \left(\frac{\partial \epsilon}{\partial \mathbf{y}_i} \right) \right.$$
$$\left. + 2\overline{n}e^{-V/kT} \left(\cosh(e_c\phi/kT) - 1\right) \left(\frac{\partial V}{\partial \mathbf{y}_i} \right) \right] d\mathbf{x} \,. \quad (16)$$

The corresponding force equation for the linearized PB is

$$\mathbf{F}_i[\phi] = -\int_\Omega \left[\phi \left(\frac{\partial \rho_f}{\partial \mathbf{y}_i} \right) - \frac{\left(\nabla\phi\right)^2}{8\pi} \left(\frac{\partial \epsilon}{\partial \mathbf{y}_i} \right) - \frac{\left(\phi\right)^2}{2} \left(\frac{\partial \overline{\kappa}^2}{\partial \mathbf{y}_i} \right) \right] d\mathbf{x} \,. \quad (17)$$

The papers cited above present a number of different numerical methods for evaluating the force and energy integrals based on stability considerations and different dielectric and ion accessibility surface definitions.

As mentioned above, it is important to note that both the nonlinear and linearized PB equations are approximate theories which should not be applied blindly to biomolecular systems — particularly those with high charge densities. The PB equation is based on a mean field approximation of the statistical mechanics of the counterion system; as such it neglects counterion correlations and fluctuations which become

important at high ion concentrations and valencies [7]. Furthermore, the Poisson equation is based on the assumption of linear and local polarization of the solvent with respect to an applied field [43]. This assumption can break down under high fields or in highly-ordered systems of water. In short, the PB equation (and other implicit solvent models described in this chapter) works best for describing the electrostatic properties of biomolecules with low linear charge density in solutions of monovalent ions at low concentration.

2.2 Numerical Methods

Analytical solutions of the PB equations are not available for realistic biomolecular geometries. Therefore, this equation is generally solved numerically by a variety of computational methods. There are a few methods which try to solve the PB equations using atom-centered basis functions [44, 45]. However, the most common numerical techniques are based on discretization of the problem domain into simplices which support locally-defined basis functions. These methods include finite difference, finite element, and boundary element methods (see Fig.2). By projecting the solution into this discrete basis, these methods transform the continuous partial differential equation into a set of (possibly nonlinear) coupled algebraic equations. In particular, Newton methods [46, 47] are typically used used to transform the nonlinear equations into linear problems which can be solved using a variety of matrix algebra methods [48].

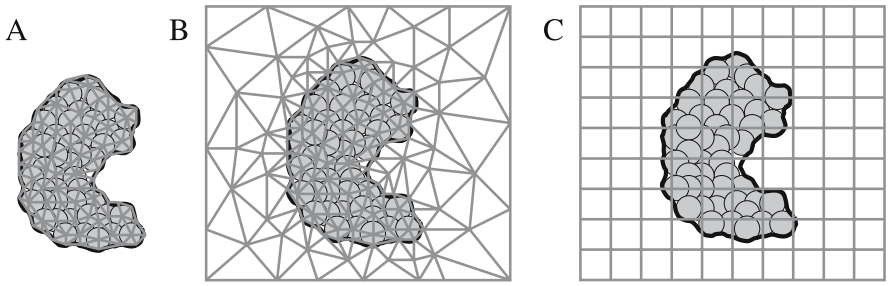

Figure 2. Discretization schemes used in solution of the PB equation: A. boundary element, B. finite element, C. finite difference

Finite element (FE) methods [48, 49] use an adaptive discretization of the problem domain, often into triangles (2D) or tetrahedra (3D). These methods offer a significant degree of adaptivity, allowing the problem unknowns to be distributed to minimize the error in the calculated solution (see Fig. 2A,B). When implemented correctly, FE methods can leverage very efficient algebraic multigrid solvers and thereby allow optimal numerical performance: the solution of an equation with N unknowns in $\mathcal{O}N$ time [50]. FE methods have been successfully applied to both the

linear and nonlinear versions of the PB equation [51–55]. Furthermore, using parallel refinement methods developed by Bank and Holst [56], FE methods have been ported to massively parallel supercomputers and thereby enabled the solution of the PB equation for large microtubule structures [57].

Boundary element (BE) methods (see Fig. 2A) represent a particular implementation of finite element methods for surface representations of Poisson's equation [58, 59] and related implementations for linearized [60, 61] and nonlinear [62] versions of the PB equation. These methods rely on a discontinuous description of the solute-solvent dielectric interface; for the Poisson equation, this interface serves as the surface upon which a 2D finite element discretization is performed. For the PB equation, a number of BE methods are available. The linearized PB equation can be solved by extending the surface mesh a short distance from the biomolecule [60, 61]. The nonlinear version of the PB equation is solved by a boundary element formulation in conjunction with a coarse volume discretization in a manner similar to the finite difference Green function-based solution of the PB equation proposed by Levitt and co-workers [63]. The advantage of all BE methods is a much smaller numerical system to be solved (due to the lower dimensionality of the surface discretization). However, BE methods generate a much denser matrix problem than FE or finite difference methods and must be accelerated by wavelet or multigrid techniques [61, 64].

Finite difference (FD) methods rely on a general tensor product (typically Cartesian) mesh discretization of the system (see Fig. 2C). These discretizations include traditional finite difference [65, 66] and finite volume or multigrid [46, 47, 67] methods. All of these methods use various orders of Taylor expansion to generate "stencils" which represent the derivative operators in the PB equation. When applied to the PB equation, these stencils form a series of sparse algebraic equations which are solved with a variety of techniques. In particular, large FD problems benefit from a multigrid solution [46, 47] of the PB equation which allows $\mathcal{O}(N)$ solution of equations involving N grid points. Although FD methods do not permit adaptive refinement in the manner of BE and FE discretizations, they can employ a unique method known as electrostatic focusing [68, 69] to emulate limited adaptivity. As illustrated in Fig. 3B, focusing allows users to choose coarse FD grids for global calculations which are then used to define the boundary conditions on finer grids at regions of interest (binding and active sites, titratable residues, etc.). The result of successive levels of focusing are highly accurate local solutions to the PB equation with reduced levels of computational effort (compared to a global solution at high resolution). Focusing methods methods have also been parallelized [67] following the parallel FE ideas of Bank and Holst [56] and later implemented in a similar manner by Balls and Colella [70]. See Fig. 3 for a schematic of the FD and FE implementations of these parallel approaches. The resulting methods were used to solve the PB equation for a number of large biomolecular systems, including the ribosome and a million-atom microtubule [67].

Table 1 presents a partial list of the more popular PB solvers available at the time of publication.

Table 1. Popular PB equation software packages

Software	*Description*	*Availability*
APBS[67]	Solves the PB equation in parallel using the PMG FD multigrid package[46, 47] and the FEtk finite element software[50]	`http://agave.wustl.edu/apbs/`
Delphi[71]	Solves the PB equation sequentially with a highly optimized FD Gauss-Seidel solver	`http://trantor.bioc.columbia.edu/delphi/`
GRASP[72]	Solves the PB equation sequentially to obtain solutions for visualization	`http://trantor.bioc.columbia.edu/grasp/`
MEAD[73]	Solves the PB equation sequentially with an FD successive over-relaxation (SOR) solver	`http://www.scripps.edu/mb/bashford/`
ZAP[35]	Solves the PB equation very rapidly with a FD discretization and a smooth dielectric function	`http://www.eyesopen.com/products/toolkits/zap.html`
UHBD[74]	Multipurpose program with emphasis on Brownian dynamics; solves PB equation sequentially with an FD preconditioned conjugate gradient solver	`http://mccammon.ucsd.edu/uhbd.html`
MacroDox	Multipurpose program with emphasis on Brownian dynamics; solves PB equation sequentially with an FD SOR solver	`http://prin.chem.tntech.edu/`
Jaguar[53, 54]	Multipurpose program with emphasis on quantum chemistry; solves PB equation sequentially with an FE solver	`http://www.schrodinger.com/Products/jaguar.html`
CHARMM[75]	Multipurpose program with emphasis on molecular dynamics; solves PB equation sequentially with a FD multigrid solver and can be linked to APBS	`http://yuri.harvard.edu`
Amber[76]	Multipurpose program with emphasis on molecular dynamics; solves PB equation sequentially with a preconditioned conjugate gradient solver	`http://amber.scripps.edu`

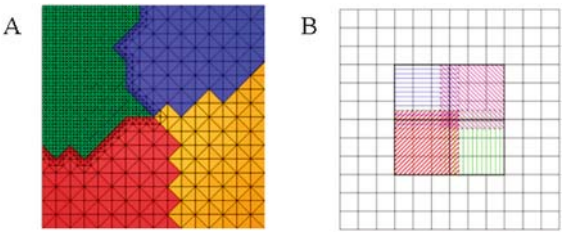

Figure 3. Parallel finite element and finite difference approaches to the PB equation. A. Depiction of parallel adaptive refinement in a finite element solution of the PB equation [56, 57]. The global domain is subdivided among the available processors; the problem is solved on the entire (coarse) domain. However, subsequent adaptive refinement is constrained to the appropriate subdomain specific to each processor. B. Demonstration of sequential [68] and parallel [67] focusing. In sequential focusing, a coarse grid (finer black lines) is used to obtain the global solution and set the boundary conditions on a finer grid (boundary denoted by heavy black line) surrounding the region of interest. Parallel focusing proceeds along similar lines except that the finer grid is further subdivided into "focusing subdomains" which are assigned to specific processors. The focusing proceeds as usual, except the final subdomain assigned to each processor is much smaller than with sequential focusing

3 Generalized Born and Related Approximations

The underlying physical picture on which the generalized Born approximation is based is the same as for the Poisson-Boltzmann calculations discussed above. In the case of simple spherical ion of radius a and charge eQ, the potentials can be found analytically and the result is the well-known Born formula [21],

$$\Delta G_{pol} = -\frac{Q^2}{2a}\left(1 - \frac{1}{\epsilon_{sol}}\right) \tag{18}$$

If we imagine a "molecule" consisting of charges $Q_1 \ldots Q_N$ embedded in spheres of radii, $a_1 \ldots a_N$, and if the separation r_{ij} between any two spheres is sufficiently large in comparison to the radii, then the solvation free energy can be given by a sum of individual Born terms, and pairwise Coulombic terms:

$$\Delta G_{pol} \simeq \sum_{i}^{N} -\frac{Q_i^2}{2a_i}\left(1 - \frac{1}{\epsilon_{sol}}\right) + \frac{1}{2}\sum_{i}^{N}\sum_{j\neq i}^{N}\frac{Q_i Q_j}{r_{ij}}\left(\frac{1}{\epsilon_{sol}} - 1\right) \tag{19}$$

where the factor $(1/\epsilon_{sol} - 1)$ appears in the pairwise terms because the Coulombic interactions are re-scaled by the change of dielectric constant upon going from vacuum to solvent.

The project of generalized Born theory can be thought of as an effort to find a relatively simple analytical formula, resembling 19, which for real molecular geometries will capture as much as possible, the physics of the Poisson equation. The linearity of the Poisson equation (or the linearized Poisson–Boltzmann equation) assures that ΔG_{pol} will indeed be quadratic in the charges, as both (13) and (19) assume.

However, in calculations of ΔG_{pol} based on direct solution of the Poisson equation, the effect of the dielectric constant is not generally restricted to the form of a prefactor, $(1/\epsilon_{sol} - 1)$, nor is it a general result that the interior dielectric constant, $\epsilon_{\rm in}$ has no effect. With these caveats in mind, we seek a function f^{GB}, to be used as follows:

$$\Delta G_{pol} \simeq -\left(1 - \frac{1}{\epsilon_{sol}}\right) \frac{1}{2} \sum_{ij} \frac{Q_i Q_j}{f_{ij}^{GB}} \tag{20}$$

Here the self ($i = j$) f^{GB} terms can be thought of as "effective Born radii," while in the off-diagonal terms, it becomes an effective interaction distance. The most common form chosen [77] is

$$f_{ij}^{GB}(r_{ij}) = \left[r_{ij}^2 + R_i R_j \exp(-r_{ij}^2/4R_i R_j)\right]^{1/2} \ , \tag{21}$$

in which the R_i are the effective Born radii of the atoms, which generally depend not only on a_i, the radius of atom i, but on on the radii and relative positions of all other atoms. In principle, R_i could be chosen so that if one were to solve the Poisson equation for a charge Q_i placed at the position of atom i, and no other charges, and a dielectric boundary determined by all of the molecule's atoms and their radii, then the self energy of charge i in its reaction field would be equal to $-(Q^2/2R_i)(1 - 1/\epsilon_{sol})$. These values of R_i have been called "perfect" radii, and are known to give a reasonably good approximation to Poisson theory when used in conjunction with 21 [78]. Obviously, this procedure would have no practical advantage over a direct calculation of ΔG_{pol} using a numerical solution of the Poisson equation. To find a more rapidly calculable approximation for the effective Born radii, we turn to a formulation of electrostatics in terms of integration over energy density.

3.1 Derivation in Terms of Energy Densities

In the classical electrostatics of a linearly polarizable media [79] the work required to assemble a charge distribution can be formulated either in terms of a product of the charge distribution with the electric potential (as in 13), or in terms of the scalar product of the electric field $\mathbf{E}$ and the electric displacement $\mathbf{D}$:

$$G = \frac{1}{2} \int_\Omega \rho_f(\mathbf{x}) \phi(\mathbf{x}) d\mathbf{x} \tag{22}$$

$$= \frac{1}{8\pi} \int_\Omega \mathbf{E} \cdot \mathbf{D} d\mathbf{x} \tag{23}$$

We now introduce the essential approximation used in most forms of generalized Born theory: that the electric displacement is Coulombic in form, and remains so even as the exterior dielectric is altered from 1 to ϵ_{sol} in the solvation process. In other words, the displacement due to the charge of atom i (which for convenience is here presumed to lie on the origin) is,

$$\mathbf{D}_i \approx \frac{Q_i \mathbf{r}}{r^3} \ . \tag{24}$$

This is termed, the Coulomb field approximation. In the spherically symmetric case (as in the Born formula), this approximation is exact, but in more complex geometries, there may be substantial deviations. The work of placing a charge Q_i at the origin within a "molecule" whose interior dielectric constant is $\epsilon_{\rm in}$, surrounded by a medium of dielectric constant $\epsilon_{\rm ex}$ and in which no other charges have yet been placed is then,

$$G_i = \frac{1}{8\pi} \int (\mathbf{D}/\epsilon) \cdot \mathbf{D} d\mathbf{x} \approx \frac{1}{8\pi} \int_{\rm in} \frac{Q_i^2}{r^4 \epsilon_{\rm in}} d\mathbf{x} + \frac{1}{8\pi} \int_{\rm ex} \frac{Q_i^2}{r^4 \epsilon_{\rm ex}} d\mathbf{x} \ . \tag{25}$$

The electrostatic component of the solvation energy is found by taking the difference in W_i when $\epsilon_{\rm ex}$ is changed from 1.0 to ϵ_{sol},

$$\Delta G_{pol,i} = \frac{1}{8\pi} \left(\frac{1}{\epsilon_{sol}} - 1 \right) \int_{\rm ex} \frac{Q_i}{r^4} d\mathbf{x} \tag{26}$$

where the contribution due to the interior region has canceled in the subtraction. Comparing 26 to the Born formula (18) or to (20) or (19), we conclude that the effective Born radius should be,

$$R_i^{-1} = \frac{1}{4\pi} \int_{\rm ex} \frac{1}{r^4} d\mathbf{x} \tag{27}$$

It is convenient to re-write this in terms of integration over the interior region, excluding a radius a_i around the origin,

$$R_i^{-1} = a_i^{-1} - \frac{1}{4\pi} \int_{{\rm in},\, r>a_i} \frac{1}{r^4} d\mathbf{x} \ . \tag{28}$$

Note that in the case of a monatomic ion, where the molecular boundary is simply the sphere of radius a_i, this equation becomes $R_i = a_i$ and the Born formula is recovered exactly.

The integrals in (27) or (28) can be calculated numerically by a variety of quadrature schemes [77, 80–82], which have most of the usual tradeoffs between accuracy and computational efficiency. Many investigators have chosen to accept additional approximation in order to obtain pairwise analytical formulas for the effective radii. If the molecule consisted of a set of non-overlapping spheres of radius a_j at positions, $\mathbf{r}_{ij}$ relative to atom i, then 28 could be written as a sum of integrals over spherical volumes (see Fig. 4),

$$R_i^{-1} = a_i^{-1} - \frac{1}{4\pi} \sum_j \int_{|\mathbf{r}-\mathbf{r}_{ij}|<a_j} \frac{1}{r^4} d\mathbf{x} \ . \tag{29}$$

The integrals over spheres can then be calculated analytically [83] leading to,

$$R_i^{-1} = a_i^{-1} - \sum_j \frac{a_j}{2(r_{ij}^2 - a_j^2)} - \frac{1}{4 r_{ij}} \log \frac{r_{ij} - a_j}{r_{ij} + a_j} \ . \tag{30}$$

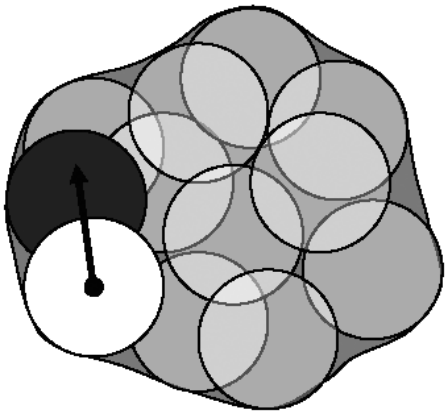

Figure 4. Representation of a molecule as a set of overlapping spheres. The integral needed in 29 can be approximately written as a sum over all of the spheres except for the white one, as in 29

An analytical expression is also available for the case of atom j overlapping with the central atom, i, provided j does not overlap any other atom j' [83]. Although in practice, the atoms j do overlap one another to some extent, these overlaps can be neglected to a first approximation, and empirical corrections can be introduced to compensate for neglect of overlap. This is referred to as the pairwise descreening approximation [84]. Hawkins et al. [84, 85] have introduced such a scheme based on re-scaling the van der Waals radii by factors S_j. The expression for the generalized Born radii takes the form,

$$R_i^{-1} = a_i^{-1} - \sum_j H(r_{ij}, S_j a_j) \; , \tag{31}$$

where H is a rather complex expression which, apart from rescaling, is essentially 30 if i and j do not overlap, and having different functional forms in overlapping cases [85].

3.2 Limitations and Variations of the GB Model

The crux of the generalized Born model is the Coulomb field approximation, 24, and this is also the main source of its deviation from solvation energies calculated using solutions of the Poisson equation. In spherically symmetric cases, such as the case analyzed in the original Born theory, $\mathbf{D}$ is given exactly by the Coulomb field, and GB solvation goes over into Poisson-equation solvation. Several studies [81–83] have analyzed deviations from the Coulomb-field approximation for the case of charges at arbitrary positions within a spherical dielectric boundary, a case for which analytical solutions of the the Poisson equation are available [86]. The Coulomb field approximation leads to significant over-estimation of self-energies, and under-estimation of the screening of charge–charge interactions. Some additional quantitative sense of the limitations of the Coulomb field approximation can be gained

by considering the case of a charge near a planar dielectric boundary, a situation analyzed in [87]. In the usual case where $\epsilon_{sol} \gg \epsilon_{\rm in}$, the magnitude of ΔG_{pol} is underestimated by a factor of almost 2 compared to the exact expression. This suggests that for charges buried somewhat below the surface of large macromolecules, the solvation energy may be underestimated, and thus the effective radii overestimated due to the Coulomb-field approximation.

It is not yet clear to what extent these limitations in the generalized Born approach might be ameliorated by clever parameterization, or by explicit attempts to go beyond the Coulomb field approximation. There are a number of "flavors" of GB theory in current use; these have recently been reviewed [88, 89], and we only give a brief account of these here.

1. Schaefer and Karplus [90] presented a general formalism for decomposing energy functions based on integration of the Coulomb-field energy density into pairwise atomic terms. The integration over the solute interior in 28 can be rewritten as an integral over all space with the integrand multiplied by a step function $P(\mathbf{r})$ whose value is 1 in the molecular interior and zero elsewhere. This function can then be written as a sum of atomic terms,

$$P(\mathbf{r}) = \sum_j P_j(\mathbf{r}) \ , \tag{32}$$

   where, for example, the P_j might be step functions corresponding to Vornoi volumes of the atoms, but in principle, any function satisfying 32 is admissible. Schaefer and Karplus [90] proposed a Gaussian form for the P_k, normalized according to the "effective volume" of each atom, which is a parameter that characterizes its contribution to the total solvent-inaccessible volume of the solute. Based on this idea, they developed an analytical, continuous and differentiable pairwise-atomic expression for the electrostatic energy of the solute, which they term ACE (Analytical treatment of Continuum Electrostatics). The model compares reasonably well with the results of numerical solution of the Poisson equation for a test set of small molecules and several proteins, and is suitable for use in molecular mechanics calculations.
2. Recently, Onufriev et al. have developed a different approach to the problem of overlapping spheres versus realistic molecular surfaces in generalized Born theory [91, 92]. They noted that pairwise GB theories using functional forms such as 31 have been parameterized to perform quite well for the calculation of small molecule solvation free energies, but do not do as well at reproducing Poisson calculations for macromolecules, particularly the energy terms pertaining to deeply buried atoms. The dielectric shape represented by overlapping spheres still includes numerous small voids in the molecular interior (shown in red in Fig. 4). This is essentially the difference between defining the dielectric interface as the van der Waals surface, which is the assumption of conventional GB theory or as the molecular (or solvent-excluded) surface [93], which is the usual definition in Poisson theory, and which is illustrated as the outer extent of color in Fig. 4. For small molecules, the distinction makes little difference, but for

macromolecules, use of the van der Waals surface leads to internal cavities of solvent dielectric that are unrealistic because they are too small for a solvent-sized probe to fit into. To compensate for this "missed volume," Onufriev et al. introduce *ad hoc* correction factors that effectively increase the volumes of deeply buried atomic spheres. Solvation energy contributions from atoms near the surface are not so strongly affected. This means that the parameterizations that have been developed for small molecules with considerable effort can be carried over into macromolecular calculations with little or no change. This is now the standard generalized Born model in the Amber and NAB program packages.

3. Ghosh et al.[94] have proposed an alternative approach, in which the Coulomb field is still used in place of the correct field, and Green's theorem is used to convert the volume integral in 28 to a surface integral. There are potential computational advantages in the surface integral approach, especially for large systems, where the number of points required to discretize the surface should grow less rapidly than in a volume representation. In practice, empirical short-range and long-range corrections are added to improve agreement with numerical Poisson theory.
4. Other investigators have returned to numerical quadrature to evaluate 28 over the proper molecular volume[81, 82]. These methods (which may also include terms designed to correct deficiencies in the Coulomb field approximation) are generally slower than pairwise methods, but can also more faithfully mimic numerical Poisson results[89].

4 Applications

PB and GB methods have been used in numerous biomedical and biophysical applications; see [1, 3–6, 69] for more information. We have only space here to illustrate a few of these.

4.1 Electrostatic Potential Analysis

One of the most recognizable products of computational electrostatic calculations are the colorful images produced by coloring biomolecular surfaces by electrostatic potentials or by rendering electrostatic isocontours around a biomolecule. As mentioned above, Table 1 presents several popular software packages for performing PB calculations. This form of analysis has proven useful for giving some insight into the role of electrostatics in biomolecular function, although over-analysis with these qualitative methods can be inappropriate.

In addition to visual analysis, electrostatic potentials have been used in a more quantitative manner to identify active and ligand-binding sites [95–98], to predict protein-protein [99–106] and protein-membrane [107–109] membrane interfaces at and around the biomolecular surfaces.

4.2 Energetic Analysis

Energy analysis is one of the primary quantitative applications of PB methods. This analysis is driven by the functionals 14 and 15 for the nonlinear and linearized PB equations, respectively. Numerical solutions of the PB equation using FE and FD methods (with a few exceptions, see [63]) typically include self-energies which represent the interaction of an atomic charge distribution with itself. These self-energies are artifacts; the true self-energy of a point charge is infinite. Furthermore, these self-energies are extremely sensitive to the discretization and can give rise to large variations in the total energy as the mesh position and spacing is changed. For these reasons, energies obtained from a single PB calculation are meaningless; they must be performed as a series of calculations to calculate a free energy difference. If performed correctly, these differences should be free from self-energy artifacts.

In general, self-energies are removed through solvation energy calculations. These calculations determine the change in energy for transferring the protein charges from a homogeneous dielectric (the same as the protein interior) to an inhomogeneous dielectric with the appropriate permittivity values:

$$\Delta_{\mathrm{solv}} G = G_{\mathrm{sys}} - G_{\mathrm{ref}} \tag{33}$$

where G_{sys} is the energy of the system with the inhomogeneous dielectric and G_{ref} is the energy of the system with the homogeneous dielectric. Both calculations use exactly the same conformations of the molecule and the same discretization. This procedure removes the self-energies because the discontinuities due to the rapidly-varying Green function near the atomic centers are canceled by the two calculations with the same (interior) dielectric coefficient, discretization, and biomolecular conformation.

Total energies can be easily obtained from solvation energies by adding the interaction energies between the biomolecular charge distributions using a uniform dielectric equal to that of the protein interior. As these charge interactions are usually available as analytical expressions, they can be easily calculated *without* the self-interaction terms. For example, Coulomb's law energies describe the interaction between point charges and can be calculated according to

$$G = \sum_{i}^{M} \sum_{j>i}^{M} \frac{Q_i Q_j}{\epsilon_p r_{ij}} \,, \tag{34}$$

where ϵ_p is the dielectric of the protein interior. When this Coulomb's law expression is added to the solvation energy as calculated with point charges, the total electrostatic energy is obtained. As an application of this procedure, consider the binding energy calculation depicted in Fig. 5. In this case, both the protein and ligand molecules change conformation upon binding; removal of self-energies is very important to the accuracy of the calculation.

The electrostatic energies and forces calculated in this fashion have a number of other applications. Recent work has focused on the use of PB in several "high-throughput" applications such as Brownian [110–112] and Langevin [76, 113–115]

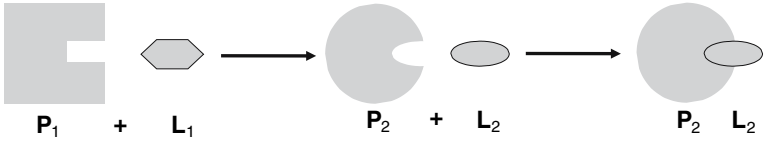

Figure 5. A schematic of ligand binding to a biomolecule with conformational change

dynamics, free energy calculations [116–118], and pK_a analysis [119–121]. These types of applications require repeated solution of the PB equation millions of times during the course of a simulation. As such, current research in PB methodology has worked to accelerate solvers for such demanding applications. Due to this "need for speed", FD methods have gained the most widespread use and demonstrated the best speed and efficiency of the methods described in the previous section. The lack of adaptivity in FD methods actually contributes to their efficiency by providing very simple problem setup and calculation of observables such as forces and energies. It should be noted that adaptive finite element methods are often the optimal choice for very large and strongly nonlinear problems, including the PB equation under certain circumstances. For the high-throughput methods mentioned above, adaptive methods do not *currently* provide the necessary level of efficiency. However, it seems likely that, with development of rapid discretization and error estimation methods, these approaches will eventually become the methods of choice for rapid solution of the PB equation.

4.3 Studying Protonation Equilibria

Solution acidity is an important thermodynamic variable that can affect biochemical function in a way that is often as profound as that of temperature or of the concentration of other allosteric effectors such as cofactors or phosphates. Many *in vitro* experiments mimic cellular compartments by regulating pH closely, commonly with buffering agents. The experimental study of titration behavior and the response of biomolecules to changes in pH has a long history, and there is a large amount known about the thermodynamics of proton binding [122, 123]. Structural correlations are less well-developed, but are becoming of increasing interest as methods for monitoring site-specific proton binding (particular by NMR) become more routine.

There is also a long history of theoretical and computational approaches to study of behavior of proteins and nucleic acids as a function of solution acidity [124–127]. This is known to be a difficult problem, since almost all biomolecules have multiple sites that can bind or release protons, and these are coupled to one another in complex ways. The computational study of the this problem has recently been reviewed [126], and we can only give a brief overview of the field here.

In principle, the most rigorous way to estimate an individual pK_a value for a protein side-chain would involve a free energy simulation connecting the protonated and de-protonated forms of the molecule:

$$\mathrm{p}K_a = -\log_{10} K_a = \frac{\Delta G}{kT \ln 10} \tag{35}$$

In a molecular mechanics approach (where covalent bonds cannot be broken) this in practice would involve parallel, explicit solvent simulations on the protein of interest and on a model compound with the same functionality and with a known $\mathrm{p}K_a$. The computed $\mathrm{p}K_a$ difference can then be added to the known model compound value to estimate the macromolecular result. This model effectively assumes that energetic contributions outside the molecular mechanics model (such as the strength of the O–H chemical bond) are the same in the protein as in the model compound. Calculations of this general kind were introduced nearly 80 years ago by Linderström-Lang [22] who modeled the protein as a sphere with the charges of the ionizable groups spread uniformly over its surface. As it become understood that proteins were not fluid globules but contained more specific structures, Tanford and Kirkwood [128, 129] developed a model of protein ionizable sites as point charges uniformly spaced at a short fixed distance beneath the surface of a sphere. As actual protein structures became known, the Tanford–Kirkwood model was adjusted to use charge placements derived from the actual structures [130], and empirical corrections for differential solvent exposure were introduced [131, 132].

The modern models, which have nearly displaced the sphere-based models, can be thought of as the natural extension of Tanford and Kirkwood's ideas to more detailed and realistic molecular surface shapes and charge distributions (see Fig. 6). The fundamental assumptions are as follows: (1) The free energy of ionization can be divided into an internal part that includes bond breaking and other electronic structure changes and is confined to a relatively small number of atoms within the ionizing functional group, and an external part that includes interactions with the larger surroundings. (2) The internal part is the same for both the ionizing group in the protein and the corresponding model compound, but the external part may differ. (3) Since the steric changes in protonation/deprotonation are subtle, and similar in both the protein and model compound, the steric contribution to the difference of the external part can be neglected. (4) The remaining difference in the external part is purely electrostatic. (5) A implicit solvent model is adequate to describe this electrostatic difference. As a practical matter, other simplifying approximations are often introduced as well, such as neglect of conformational change and simplification of charge models; but these approximations can be lifted without changing the basic ideas.

4.4 Multiple Interacting Sites

Our discussion so far has considered the thermodynamics of protonation of a single titrating site, where it makes sense to define a $\mathrm{p}K_a$ value as in 35, and the fraction of sites that have a proton bound can be trivially calculated once K_a is known. Proteins, of course, have multiple titrating sites, and one must consider the relative energetics of different protonation possibilities, and then the statistics of thermal ensembles over the possible states. A rough "$\mathrm{p}K_a$" can then be identified as the pH value at which the populations of the protonated and deprotonated forms

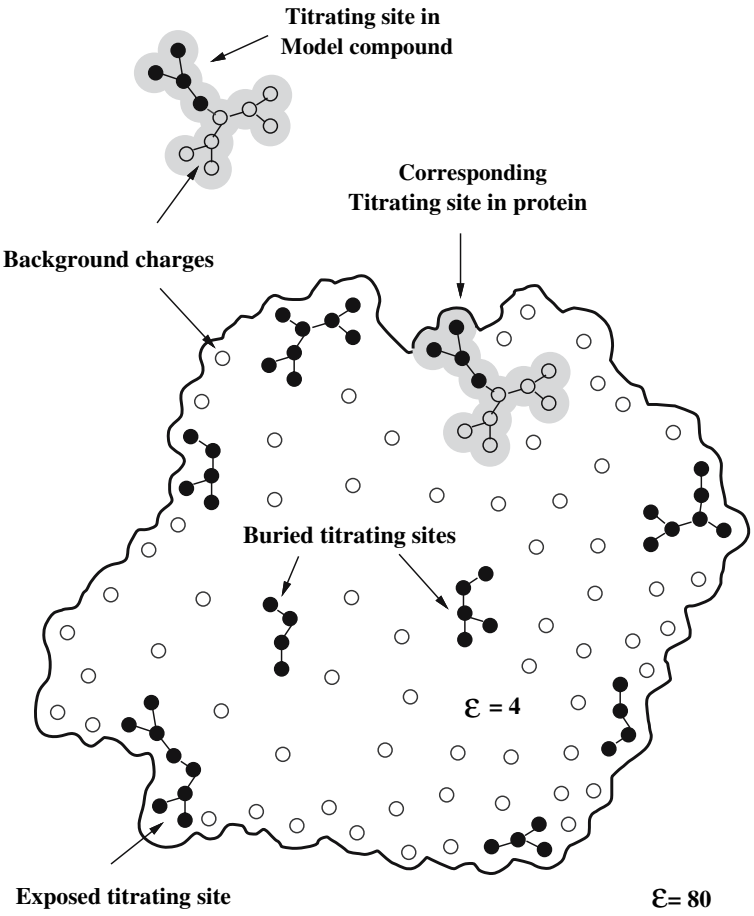

Figure 6. Continuum solvent model for computing pKa values in proteins

are equal. In extreme cases, however, this may not be well-defined or unique, and it is really the properties of the entire titration curve that are of greatest interest [122, 127, 133, 134]. Generally, proton binding is slightly anti-cooperative (because repulsive charge-charge interactions grow as more protons bind), but significant excursions from "normal" behavior can occur, either as the result of strong coupled interactions between nearby sites, or as a result of pH-dependent changes, such as global unfolding at low or high pH.

By "protonation state" of the protein, we mean a specification of which sites are protonated and which are deprotonated. If the protein has N sites, each with two possible protonation states, there are 2^N possible protonation states of the protein. Let them be denoted by the N-element vector, x whose elements x_i can each take on values representing "protonated" or "deprotonated." Let $\mathbf{0}$ be a particular value of this vector chosen as a reference state, for example, the state with all sites in

their neutral state, as in the definition of $\mathrm{p}K_{\mathrm{intr}}$ (see below). Consider the chemical equilibrium between some state x and the reference state:

$$\mathrm{M}(0) + \nu(x)\mathrm{H}^+ \rightleftharpoons \mathrm{M}(x) \tag{36}$$

where $\nu(x)$ is the number of protons that would be released on going from state x to the reference state. The free energy change for going to the right in this reaction at some fixed pH is,

$$\begin{aligned} \Delta G(x, \mathrm{pH}) &= -RT \ln \frac{[\mathrm{M}(x)]}{[\mathrm{M}(0)]} \\ &= +\mu^\circ_{\mathrm{M}}(x) - \mu^\circ_{\mathrm{M}}(0) - \nu(x)(\mu^\circ_{\mathrm{H}^+} - 2.303RT\mathrm{pH}) \ , \end{aligned} \tag{37}$$

where the μ° are standard chemical potentials.

Suppose $P_i(x_i)$ is the change in the protein's chemical potential for changing site i from its reference state to state x_i, while all other sites remain in the reference state. The expression for chemical potential will then contain a sum of these P_i, but more terms are needed for the site–site interactions. Since the interactions are presumed to be governed by linear equations (the Poisson or linearized Poisson-Boltzmann equations) these terms will be pairwise additive. Therefore,

$$\mu^\circ_M(x) = \mu^\circ_M(0) + \sum_i^N P_i(x_i) + \frac{1}{2}\sum_{ij, i\neq j}^N W_{ij}(x_i, x_j) \ , \tag{38}$$

where W_{ij} is the electrostatic interaction between sites i and j, relative to the reference state. Inserting this into 37 and writing $\nu(x)$ as a sum over sites, $\sum_i \nu_i(x_i)$, gives,

$$\Delta G(x, \mathrm{pH}) = \sum_i [P_i(x_i) - \nu_i(x_i)(\mu^\circ_{\mathrm{H}^+} - 2.303RT\mathrm{pH})] + \frac{1}{2}\sum_{ij, i\neq j}^N W_{ij}(x_i, x_j) \tag{39}$$

It is convenient to define a quantity, the intrinsic $\mathrm{p}K$ or $\mathrm{p}K_{\mathrm{intr}}$ of a site which is the $\mathrm{p}K_a$ that the site would have if all other sites were held in their neutral states. Assuming that the reference state is this neutral state, one finds, $P_i(x_i) = \nu_i(x_i)(\mu^\circ_{\mathrm{H}^+} - 2.303RT\mathrm{p}K_{\mathrm{intr},i})$, so that,

$$\Delta G(x, \mathrm{pH}) = \sum_i \nu_i(x_i) 2.303RT(\mathrm{pH} - \mathrm{p}K_{\mathrm{intr},i}) + \frac{1}{2}\sum_{ij, i\neq j}^N W_{ij}(x_i, x_j) \ . \tag{40}$$

The intrinsic $\mathrm{p}K_a$values of each site, $\mathrm{p}K_{\mathrm{intr},i}$, can be calculated by the methods of Sec. 4.3. The $W_{ij}(x_i, x_j)$ can be calculated by considering the additional electrostatic work to change site i's charges from the reference state to x_i if site j is in state x_j instead of its reference state.

Computation of experimentally meaningful quantites typically requires a partition function or average over all possible protonation states. For example, the protonation fraction of site i at a particular pH is

$$\theta_i(\mathrm{pH}) = \frac{\sum_x^{2^N} \nu_i(x_i) \exp[-\Delta G(x, \mathrm{pH})/RT]}{\sum_x^{2^N} \exp[-\Delta G(x, \mathrm{pH})/RT]} \ . \tag{41}$$

Since the sums grow exponentially in the number of sites, the explicit use of such formula is only feasible if the number of sites is not too large.

There are two main ways in which the multiple-site problem can be addressed. One approach makes a mean-field (or Tanford-Roxby [130]) approximation, in which sites interacts not with the particular protonation states of other sites, but with their (pH-dependent) average protonation; in effect, sidechains with a pK_a value near the solution pH will be represented with charge distributions that are intermediate between the protonated and de-protonated forms [126, 130]. The other is Monte Carlo sampling over protonation states to estimate thermal averages [126, 135].

4.5 Constant pH Simulations

Combining the above methods with molecular dynamics should, in principle, give a methodology that accounts for conformational flexibility in the pK_a prediction problem, and allows for simulations of pH effects on macromolecule conformation.

Some workers use a continuous titration coordinate, so that partially protonated sites are represented by charge distributions that interpolate between the protonated and de-protonated endpoints [136–139]. This is similar in spirit to the mean field approach to the multi-site titration problem. Recent work by Lee et al. [140] describes a novel approach for avoiding mean field approximations in a continuous protonation state model, drawing on the ideas of λ-dynamics [141]. In this constant pH model, a potential is constructed along a coordinate interpolating between the protonated and deprotonated states, and equations of motion are used to propagate the λ coordinate. Convergence to an intermediate charge state is avoided by introducing an energetic barrier centered at $\lambda = 1/2$, which forces λ toward values representing full protonation or deprotonation. The barrier height is a tunable parameter, trading off between protonation state transition rate and fraction of simulation time spent in intermediate protonation states. Convergence problems are apparent in application to proteins, but this is a nearly universal result: trying to average over both conformational states and protonation states is a daunting task for all but the simplest situations. Despite the convergence issues, the model produces predicted pKa values that are in good agreement with experimental values. The average absolute error is 1.6 for hen egg white lysozyme, 1.2 for turkey ovomucoid, and 0.9 for bovine pancreatic trypsin inhibitor (BPTI).

A variety of constant pH methods have been proposed that avoid the use of intermediate charge states. All of these employ MD simulations to sample conformations, and periodically update the force field parameters (principally the charges) based on energetic analyses of the current configuration, using a Monte Carlo method to choose the new parameters. The methods differ in the frequency with which protonation state updates are made, and in the methods of estimating the energies of protonation or deprotonation events [142–144]. Due to the computational expense

of PB calculations, MC steps are relatively infrequent (every 1 to 5 ps) in these PB based methods.

Greater efficiency can be obtained by using the generalized Born (GB) model for both the dynamics and the protonation state sampling [145]. This permits much more frequent Monte Carlo protonation steps, which appears to improve convergence. These frequent changes of the protonation state do not appear to adversely affect the stability of the simulation. Unlike the PB based methods, which sample over all relevant protonation states at each MC step, the current implementation changes the state of at most one titratable group on each MC step. The method was applied to lysozyme, where 1 ns trajectories at a range of pH values are less well converged but nevertheless yield pKa values with only 0.82 RMS error with respect to experimental data.

5 Conclusions

Implicit solvent models significantly reduce the degrees of freedom in simulations by approximate treatment of solvent behavior. However, there are several types of implicit solvent models and it is often difficult to determine which model offers the efficiency and accuracy needed for a particular application.

PB theory is usually the standard for assessing implicit solvent methodologies [78, 89, 146]. Most other implicit solvent methods, such as GB, attempt to approximate the solution to the Poisson or Poisson-Boltzmann equation with varying descriptions of dielectric coefficients and ionic accessibilities. Several studies have shown that, with appropriate parameterization, PB methods can provide polar solvation energies and forces for proteins and small molecules that compare very well with results from explicit solvent simulations and experiment [17, 147, 148]. Unfortunately, PB methods are often slow. Although there have been numerous attempts to remedy this problem, the historical problems with PB efficiency have opened the door for GB models and other simpler approximations. As described above, GB methods can be very efficient at evaluating solvation energies and forces for large biomolecular systems.

However, GB methods also have their weaknesses. Recently, Onufriev and coworkers demonstrated that GB methods can, in principle, offer levels of accuracy on par to PB solvers [78]. However, despite recent advances in GB methods, it is not clear that the same levels of accuracy are achieved in the routine application of GB models [89]. While a number of groups have developed methods to address some of these accuracy issues, it appears that there is ample room for improvement in GB as well as PB methods.

As mentioned above, implicit solvent methods offer the prospect of improved simulation efficiency at the expense of *approximate* treatment of solvation. The choice of implicit solvent model for a simulation must ultimately be made by careful consideration of the benefits and caveats outlined above. Simulators should use caution and determine, perhaps a priori or by comparison of results for short model

simulations, whether the acceleration offered by GB or PB methods provides the appropriate level of accuracy for the biological system of interest.

Acknowledgments

NAB thanks the NIH and Alfred P. Sloan Foundation for financial support. DAC and DB were supported by NIH grant GM-57513.

References

[1] N. A. Baker and J. A. McCammon. Electrostatic interactions. In P. Bourne and H. Weissig, editors, *Structural Bioinformatics*, pages 427–440. John Wiley and Sons, Inc., New York, 2003.

[2] M. Gilson. Introduction to continuum electrostatics. In D. A. Beard, editor, *Biophysics Textbook Online*, volume Computational Biology. Biophysical Society, Bethesda, MD, 2000.

[3] B. Roux. Implicit solvent models. In O. M. Becker, Jr. MacKerell, A. D., B. Roux, and M. Watanabe, editors, *Computational Biochemistry and Biophysics*, pages 133–152. Marcel Dekker, New York, 2001.

[4] M. E. Davis and J. A. McCammon. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.*, 94:7684–7692, 1990.

[5] B. Honig and A. Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268(5214):1144–9, 1995.

[6] N. A. Baker. Poisson-boltzmann methods for biomolecular electrostatics. *Meth. Enzymol.*, 383:94–118, 2004.

[7] Christian Holm, Patrick Kekicheff, and Rudolf Podgornik, editors. *Electrostatic effects in soft matter and biophysics*, volume 46 of *NATO Science Series*. Kluwer Academic Publishers, Boston, 2001.

[8] L. Onsager. Electric moments of molecules in liquids. *J. Amer. Chem. Soc.*, 58:1486–1493, 1936.

[9] S. Miertus, E. Scrocco, and J. Tomasi. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.*, 55:117–129, 1981.

[10] J. Tomasi and M. Persico. Molecular interactions in solution: an overview of methods based on continuous distributions of solvent. *Chem. Rev.*, 94:2027–2094, 1994.

[11] J. Tomasi. Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theor. Chem. Acc.*, 112:184–203, 2004.

[12] C. Lim, D. Bashford, and M. Karplus. Absolute pKa calculations with continuum dielectric methods. *J. Phys. Chem.*, 95:5610–5620, 1991.

[13] W.H. Richardson, C. Peng, D. Bashford, L. Noodleman, and D.A. Case. Incorporating solvation effects into density functional theory: Calculation of absolute acidities. *Int. J. Quantum Chem.*, 61:207–217, 1997.

[14] J.J. Klicic, R.A. Friesner, S.Y. Liu, and W.C. Guida. Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. *J. Phys. Chem. A*, 106:1327–1335, 2002.

[15] D.M. Chipman. Computation of pKa from Dielectric Continuum Theory. *J. Phys. Chem. A*, 106:7413–7422, 2002.

[16] M.S. Busch and E.W. Knapp. Accurate pKa Determination for a Heterogeneous Group of Organic Molecules. *Chemphyschem*, 5:1513–1522, 2004.

[17] D. Sitkoff, K.A. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.*, 98:1978–1988, 1994.

[18] M. Nina, D. Beglov, and B. Roux. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B*, 101:5239–5248, 1997.

[19] E. Gallicchio, L.Y. Zhang, and R.M. Levy. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Nonpolar Hydration Free Energy Estimators. *J. Comput. Chem.*, 23:517–529, 2002.

[20] E. Gallicchio and R.M. Levy. AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling. *J. Comput. Chem.*, 25:479–499, 2004.

[21] M. Born. Volumen und hydratationswärme der ionen. *Z. Phys.*, 1:45–48, 1920.

[22] K. Linderstrøm-Lang. On the ionisation of proteins. *Comptes-rend Lab. Carlsberg*, 15(7):1–29, 1924.

[23] C. Tanford and J. G. Kirkwood. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.*, 79:5333–5339, 1957.

[24] M. A. Flanagan, G. K. Ackers, J. B. Matthew, G. I. H. Hanania, and F. R. N. Gurd. Electrostatic contributions to the energetics of dimer-tetramer assembly in human hemoglobin: pH dependence and effect of specifically bound chloride ions. *Biochemistry*, 20:7439–7449, 1981.

[25] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. *J. Mol. Biol.*, 157(4):671–9, 1982.

[26] E. Demchuk, D. Bashford, G. Gippert, and D.A. Case. Thermodynamics of a reverse turn motif. Solvent effects and side-chain packing. *J. Mol. Biol.*, 270:305–317, 1997.

[27] J. Srinivasan, T.E. Cheatham, III, P. Kollman, and D.A. Case. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.*, 120:9401–9409, 1998.

[28] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, and T.E. Cheatham, III. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accts. Chem. Res.*, 33:889–897, 2000.

[29] J. O. Bockris and A. K. N. Reddy. *Modern Electrochemistry: Ionics*. Plenum Press, New York, 1998.

[30] John David Jackson. *Classical Electrodynamics*. John Wiley and Sons, New York, 2nd edition, 1975.

[31] C. J. Bötcher. *Theory of Electric Polarization*. Elsevier Science, New York, 1973.

[32] B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, 55(3):379–400, 1971.

[33] M. L. Connolly. The molecular surface package. *J. Mol. Graph.*, 11(2):139–41, 1993.

[34] W. Im, D. Beglov, and B. Roux. Continuum solvation model: electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comp. Phys. Commun.*, 111(1–3):59–75, 1998.

[35] J. Andrew Grant, Barry T. Pickup, and Anthony Nicholls. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.*, 22(6):608–640, 2001.

[36] M. Nina, W. Im, and B. Roux. Optimized atomic radii for protein continuum electrostatics solvation forces. *Biophys. Chem.*, 78(1-2):89–96, 1999.

[37] F. Dong, M. Vijaykumar, and H. X. Zhou. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. *Biophys. J.*, 85(1):49–60, 2003.

[38] J. Wagoner and N. A. Baker. Solvation forces on biomolecular structures: A comparison of explicit solvent and Poisson-Boltzmann models. *J. Comput. Chem.*, 25(13):1623–9, 2004.

[39] K. A. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.*, 94(19):7684–7692, 1990.

[40] F. Fogolari and J. M. Briggs. On the variational approach to Poisson-Boltzmann free energies. *Chemical Physics Letters*, 281(1–3):135–139, 1997.

[41] A. M. Micu, B. Bagheri, A. V. Ilin, L. R. Scott, and B. M. Pettitt. Numerical considerations in the computation of the electrostatic free energy of interaction within the Poisson-Boltzmann theory. *J. Comput. Phys.*, 136(2):263–271, 1997.

[42] M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *J. Phys. Chem.*, 97(14):3591–3600, 1993.

[43] D. Beglov and B. Roux. Solvation of complex molecules in a polar liquid: an integral equation theory. *J. Chem. Phys.*, 104(21):8678–8689, 1996.

[44] B. Egwolf and P. Tavan. Continuum description of solvent dielectrics in molecular-dynamics simulations of proteins. *J. Chem. Phys.*, 118(5):2039–56, 2003.

[45] B. Egwolf and P. Tavan. Continuum description of ionic and dielectric shielding for molecular-dynamics simulations of proteins in solution. *J. Chem. Phys.*, 120(4):2056–2068, 2004.

[46] M. Holst and F. Saied. Multigrid solution of the Poisson-Boltzmann equation. *J. Comput. Chem.*, 14(1):105–13, 1993.

[47] M. J. Holst and F. Saied. Numerical solution of the nonlinear Poisson-Boltzmann equation: developing more robust and efficient methods. *J. Comput. Chem.*, 16(3):337–64, 1995.

[48] Dietrich Braess. *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics.* Cambridge University Press, Cambridge, 1997.

[49] O. Axelsson and V. A. Barker. *Finite Element Solution of Boundary Value Problems. Theory and Computation.* Academic Press, Inc., San Diego, 1984.

[50] M. Holst. Adaptive numerical treatment of elliptic systems on manifolds. *Advances in Computational Mathematics*, 15(1-4):139–191, 2001.

[51] M. Holst, N. Baker, and F. Wang. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation. I. Algorithms and examples. *J. Comput. Chem.*, 21(15):1319–1342, 2000.

[52] N. Baker, M. Holst, and F. Wang. Adaptive multilevel finite element solution of the Poisson-Boltzmann equation II. Refinement at solvent-accessible surfaces in biomolecular systems. *J. Comput. Chem.*, 21(15):1343–1352, 2000.

[53] C. M. Cortis and R. A. Friesner. An automatic three-dimensional finite element mesh generation system for the Poisson-Boltzmann equation. *J. Comput. Chem.*, 18:1570–1590, 1997.

[54] C. M. Cortis and R. A. Friesner. Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. *J. Comput. Chem.*, 18:1591–1608, 1997.

[55] P E Dyshlovenko. Adaptive numerical method for Poisson-Boltzmann equation and its application. *Comp. Phys. Commun.*, 147:335–338, 2002.

[56] R. E. Bank and M. Holst. A new paradigm for parallel adaptive meshing algorithms. *SIAM Journal on Scientific Computing*, 22(4):1411–1443, 2000.

[57] N. A. Baker, D. Sept, M. J. Holst, and J. A. McCammon. The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers. *IBM Journal of Research and Development*, 45(3–4):427–438, 2001.

[58] R. J. Zauhar and R. S. Morgan. The rigorous computation of the molecular electric potential. *J. Comput. Chem.*, 9(2):171–187, 1988.

[59] A. H. Juffer, E. F. F. Botta, B. A. M. van Keulen, A. van der Ploeg, and H. J. C. Berendsen. The electric potential of a macromolecule in a solvent: a fundamental approach. *J. Comput. Phys.*, 97:144–171, 1991.

[60] S. A. Allison and V. T. Tran. Modeling the electrophoresis of rigid polyions: application to lysozyme. *Biophys. J.*, 68(6):2261–70, 1995.

[61] A. J. Bordner and G. A. Huber. Boundary element solution of the linear Poisson-Boltzmann equation and a multipole method for the rapid calculation of forces on macromolecules in solution. *J. Comput. Chem.*, 24:353–367, 2003.

[62] A. H. Boschitsch and M. O. Fenley. Hybrid boundary element and finite difference method for solving the nonlinear Poisson-Boltzmann equation. *J. Comput. Chem.*, 25(7):935–955, 2004.

[63] Z. Zhou, P. Payne, M. Vasquez, N. Kuhn, and M. Levitt. Finite-difference solution of the Poisson-Boltzmann equation: complete elimination of self-energy. *J. Comput. Chem.*, 17:1344–1351, 1996.

[64] Y. N. Vorobjev and H. A. Scheraga. A fast adaptive multigrid boundary element method for macromolecular electrostatic computations in a solvent. *J. Comput. Chem.*, 18(4):569–583, 1997.

[65] A. Nicholls and B. Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.*, 12(4):435–445, 1991.

[66] M. E. Davis and J. A. McCammon. Solving the finite difference linearized Poisson-Poltzmann equation: a comparison of relaxation and conjugate gradient methods. *J. Comput. Chem.*, 10:386–391, 1989.

[67] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl. Acad. Sci. USA*, 98(18):10037–10041, 2001.

[68] M. K. Gilson and B. H. Honig. Calculation of electrostatic potentials in an enzyme active site. *Nature*, 330(6143):84–6, 1987.

[69] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules – theory and applications. *Annu. Rev. Biophys. Biophys. Chem.*, 19:301–332, 1990.

[70] G. T. Balls and P. Colella. A finite difference domain decomposition method using local corrections for the solution of Poisson's equation. *J. Comput. Phys.*, 180(1):25–53, 2002.

[71] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.*, 23(1):128–37, 2002.

[72] A. Nicholls, K. A. Sharp, and B. Honig. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11(4):281–96, 1991.

[73] D. Bashford. An object-oriented programming suite for electrostatic effects in biological molecules. In Y. Ishikawa, R. R. Oldehoeft, J. V. W. Reynders, and M. Tholburn, editors, *Scientific Computing in Object-Oriented Parallel Environments*, volume 1343 of *Lecture Notes in Computer Science*, pages 233–240. Springer, Berlin, 1997.

[74] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. Electrostatics and diffusion of molecules in solution - simulations with the university of houston brownian dynamics program. *Comp. Phys. Commun.*, 91(1-3):57–95, 1995.

[75] Jr. MacKerell, A. D., B. Brooks, III Brooks, C. L., L. Nilsson, B. Roux, Y. Won, and M. Karplus. Charmm: the energy function and its parameterization with an overview of the program. In P. v. R. Schleyer, editor, *The Encyclopedia of Computational Chemistry*, volume 1, pages 271–277. John Wiley and Sons, Chichester, 1998.

[76] R. Luo, L. David, and M. K. Gilson. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem.*, 23(13):1244–53, 2002.

[77] W.C. Still, A. Tempczyk, R.C. Hawley, and T. Hendrickson. Semianalytical treatement of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[78] A. Onufriev, D.A. Case, and D. Bashford. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comput. Chem.*, 23:1297–1304, 2002.

[79] J.D. Jackson. *Classical Electrodynamics*. Wiley and Sons, New York, 1975.

[80] M. Scarsi, J. Apostolakis, and A. Caflisch. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A*, 101:8098–8106, 1997.

[81] M.S. Lee, F.R. Salsbury, Jr., and C.L. Brooks, III. Novel generalized Born methods. *J. Chem. Phys.*, 116:10606–10614, 2002.

[82] M.S. Lee, M. Feig, F.R. Salsbury, and C.L. Brooks. New Analytic Approximation to the Standard Molecular Volume Definition and Its Application to Generalized Born Calculations. *J. Comput. Chem.*, 24:1348–1356, 2003.

[83] M. Schaefer and C. Froemmel. A precise analytical method for calculating the electrostatic energy of macromolecules in aqueous solution. *J. Mol. Biol.*, 216:1045–1066, 1990.

[84] G.D. Hawkins, C.J. Cramer, and D.G. Truhlar. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.*, 246:122–129, 1995.

[85] G.D. Hawkins, C.J. Cramer, and D.G. Truhlar. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, 100:19824–19839, 1996.

[86] J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special applications to zwitterions. *J. Chem. Phys.*, 2:351–361, 1934.

[87] D. Bashford and D.A. Case. Generalized Born Models of Macromolecular Solvation Effects. *Annu. Rev. Phys. Chem.*, 51:129–152, 2000.

[88] M. Feig and C.L. Brooks, III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struct. Biol.*, 14:217–224, 2004.

[89] M. Feig, A. Onufriev, M.S. Lee, W. Im, D.A. Case, and C.L. Brooks. Performance Comparison of Generalized Born and Poisson Methods in the Calculation of Electrostatic Solvation Energies for Protein Structures. *J. Comput. Chem.*, 25:265–284, 2004.

[90] M. Schaefer and M. Karplus. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.*, 100:1578–1599, 1996.

[91] A. Onufriev, D. Bashford, and D.A. Case. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B*, 104:3712–3720, 2000.

[92] A. Onufriev, D. Bashford, and D.A. Case. Exploring Protein Native States and Large-Scale Conformational Changes With a Modified Generalized Born Model. *Proteins*, 55:383–394, 2004.

[93] M.L. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221:709–713, 1983.

[94] A. Ghosh, C.S. Rapp, and R.A. Friesner. Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B*, 102:10983–10990, 1998.

[95] A. H. Elcock. Prediction of functionally important residues based solely on the computed energetics of protein structure. *J. Mol. Biol.*, 312(4):885–896, 2001.

[96] M. J. Ondrechen, J. G. Clifton, and D. Ringe. Thematics: a simple computational predictor of enzyme function from structure. *Proc. Natl. Acad. Sci. USA*, 98(22):12473–8, 2001.

[97] Z. Y. Zhu and S. Karlin. Clusters of charged residues in protein three-dimensional structures. *Proc. Natl. Acad. Sci. USA*, 93(16):8350–5, 1996.

[98] A. M. Richard. Quantitative comparison of molecular electrostatic potentials for structure-activity studies. *J. Comput. Chem.*, 12(8):959–69, 1991.

[99] R. Norel, F. Sheinerman, D. Petrey, and B. Honig. Electrostatic contributions to protein-protein interactions: Fast energetic filters for docking and their physical basis. *Prot. Sci.*, 10(11):2147–2161, 2001.

[100] S. M. de Freitas, L. V. de Mello, M. C. da Silva, G. Vriend, G. Neshich, and M. M. Ventura. Analysis of the black-eyed pea trypsin and chymotrypsin inhibitor-alpha-chymotrypsin complex. *FEBS Lett*, 409(2):121–7, 1997.

[101] L. Lo Conte, C. Chothia, and J. Janin. The atomic structure of protein-protein recognition sites. *J Mol Biol*, 285(5):2177–98, 1999.

[102] J. Janin and C. Chothia. The structure of protein-protein recognition sites. *J. Biol. Chem.*, 265(27):16027–30, 1990.

[103] V. A. Roberts, H. C. Freeman, A. J. Olson, J. A. Tainer, and E. D. Getzoff. Electrostatic orientation of the electron-transfer complex between plastocyanin and cytochrome c. *J. Biol. Chem.*, 266(20):13431–41, 1991.

[104] R. C. Wade, R. R. Gabdoulline, and F. De Rienzo. Protein interaction property similarity analysis. *International Journal of Quantum Chemistry*, 83(3–4):122–127, 2001.

[105] J. Novotny and K. Sharp. Electrostatic fields in antibodies and antibody/antigen complexes. *Prog. Biophys. Mol. Biol.*, 58(3):203–24, 1992.

[106] A. J. McCoy, V. Chandana Epa, and P. M. Colman. Electrostatic complementarity at protein/protein interfaces. *J. Mol. Biol.*, 268(2):570–84, 1997.

[107] A. Arbuzova, L. B. Wang, J. Y. Wang, G. Hangyas-Mihalyne, D. Murray, B. Honig, and S. McLaughlin. Membrane binding of peptides containing both basic and aromatic residues. experimental studies with peptides corresponding to the scaffolding region of caveolin and the effector region of marcks. *Biochemistry*, 39(33):10330–10339, 2000.

[108] Jung-Hsin Lin, Nathan Andrew Baker, and J. Andrew McCammon. Bridging the implicit and explicit solvent approaches for membrane electrostatics. *Biophys. J.*, 83(3):1374–1379, 2002.

[109] D. Murray and B. Honig. Electrostatic control of the membrane targeting of C2 domains. *Molecular Cell*, 9(1):145–154, 2002.

[110] R. R. Gabdoulline and R. C. Wade. Simulation of the diffusional association of barnase and barstar. *Biophys J*, 72(5):1917–29, 1997.

[111] S. H. Northrup, S. A. Allison, and J. A. McCammon. Brownian dynamics simulation of diffusion-influenced biomolecular reactions. *J. Chem. Phys.*, 80:1517–1524, 1984.

[112] J. L. Smart, T. J. Marrone, and J. A. McCammon. Conformational sampling with Poisson-Boltzmann forces and a stochastic dynamics monte carlo method: Application to alanine dipeptide. *J. Comput. Chem.*, 18(14):1750–1759, 1997.

[113] N. V. Prabhu, P. Zhu, and K. A. Sharp. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *J. Comput. Chem.*, 25(16):2049–2064, 2004.

[114] Q. Lu and R. Luo. A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.*, 119(21):11035–11047, 2003.

[115] B.Z. Lu, W.Z. Chen, C.X. Wang, and X. Xu. Protein Molecular Dynamics With Electrostatic Force Entirely Determined by a Single Poisson-Boltzmann Calculation. *Proteins*, 48:497–504, 2002.

[116] F. Fogolari, A. Brigo, and H. Molinari. Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophys. J.*, 85(1):159–166, 2003.

[117] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and III Cheatham, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research*, 33(12):889–97, 2000.

[118] J. M. J. Swanson, R. H. Henchman, and J. A. McCammon. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.*, 86(1):67–74, 2004.

[119] J. E. Nielsen and J. A. McCammon. On the evaluation and optimization of protein x-ray structures for pK_a calculations. *Prot. Sci.*, 12(2):313–26, 2003.

[120] R. E. Georgescu, E. G. Alexov, and M. R. Gunner. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophys. J.*, 83(4):1731–1748, 2002.

[121] J. Warwicker. Improved pK_a calculations through flexibility based sampling of a water-dominated interaction scheme. *Prot. Sci.*, 13(10):2793–805, 2004.

[122] J. Wyman and S.J. Gill. *Binding and linkage*. University Science Books, Mill Valley, CA, 1990.

[123] R.A. Alberty. *Thermodynamics of Biochemical Reactions*. John Wiley, New York, 2003.

[124] J.B. Matthew, F.R.N. Gurd, B. Garcia-Moreno E., M.A. Flanagan, K.L. March, and S.J. Shire. pH-dependent processes in proteins. *CRC Crit. Rev. Biochem.*, 18:91–197, 1985.

[125] P. Beroza and D.A. Case. Calculations of proton-binding thermodynamics in proteins. *Meth. Enzymol.*, 295:170–189, 1998.

[126] D. Bashford. Macroscopic electrostatic models for protonation states in proteins. *Frontiers Biosci.*, 9:1082–1099, 2004.

[127] B. García-Moreno and C.A. Fitch. Structural interpretation of pH and salt-dependent processes in proteins with computaional methods. *Meth. Enzymol.*, 380:20–51, 2004.

[128] C. Tanford and J.G. Kirkwood. Theory of titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.*, 79:5333–5339, 1957.

[129] C. Tanford. Theory of protein titration curves. II. Calculations for simple models at low ionic strength. *J. Am. Chem. Soc.*, 79:5340–5347, 1957.

[130] C. Tanford and R. Roxby. Interpretation of protein titration curves. *Biochemistry*, 11:2192–2198, 1972.

[131] S.J. Shire, G.I.H. Hanania, and F.R.N. Gurd. Electrostatic effects in myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin. *Biochemistry*, 13:2967–2974, 1974.

[132] J.B. Matthew and F.R.N. Gurd. Calculation of electrostatic interactions in proteins. *Meth. Enzymol.*, 130:413–436, 1986.

[133] A. Onufriev, D.A. Case, and G.M. Ullmann. A novel view of pH titration in biomolecules. *Biochemistry*, 40:3413–3419, 2001.

[134] D. Poland. Free energy of proton binding in proteins. *Biopolymers*, 69:60–71, 2003.

[135] P. Beroza, D.R. Fredkin, M.Y. Okamura, and G. Feher. Protonation of interacting residues in a protein by Monte Carlo method: Application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. USA*, 88:5804–5808, 1991.

[136] A.M. Baptista, P.J. Martel, and S.B. Petersen. Simulation of protein conformational freedom as a function of pH: Constant-pH molecular dynamics using implicit titration. *Proteins*, 27:523–544, 1997.

[137] U. Börjesson and P.H. Hünenberger. Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. *J. Chem. Phys.*, 114:9706–9719, 2001.

[138] A.M. Baptista. Comment on "Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines". *J. Chem. Phys.*, 116:7766–7768, 2002.

[139] U. Börjesson and P.H. Hünenberger. pH-dependent stability of a decalysine α-helix studied by explicit-solvent molecular dynamics simulations at constant pH. *J. Phys. Chem. B*, 108:13551–13559, 2004.

[140] M.S. Lee, F.R. Salsbury, and C.L. Brooks. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins*, 56:738–752, 2004.

[141] X. Kong and C.L. Brooks, III. λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys.*, 105:2414–2423, 1996.

[142] A.M. Baptista, V.H. Teixeira, and C.M. Soares. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.*, 117:4184–4200, 2002.

[143] M. Dlugosz and J.M. Antosiewicz. Constant-pH molecular dynamics simulations: a test case for succinic acid. *Chem. Phys.*, 302:161–170, 2004.

[144] M. Dlugosz, J.M. Antosiewicz, and A.D. Robertson. Constant-pH molecular dynamics study of protonation-structure relationship in a hexapeptide derived from ovomucoid third domain. *Phys. Rev. E*, 69:021915, 2004.

[145] J. Mongan, D.A. Case, and J.A. McCammon. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.*, 25:2038–2048, 2004.

[146] R. Zhou, G. Krilov, and B. J. Berne. Comment on "can a continuum solvent model reproduce the free energy landscape of a -hairpin folding in water?" the poisson-boltzmann equation. *J. Phys. Chem. B*, 108(22):7528–30, 2004.

[147] J. Wagoner and N.A. Baker. Solvation Forces on Biomolecular Structures: A Comparison of Explicit Solvent and Poisson-Boltzmann Models. *J. Comput. Chem.*, 25:1623–1629, 2004.

[148] M. Nina, W. Im, and B. Roux. Optimized atomic radii for protein continuum electrostatics solvations forces. *Biophys. Chem.*, 78:89–96, 1999.