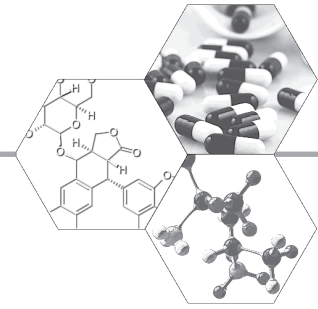

Electrostatics in proteins and protein–ligand complexes

Accurate computational methods for predicting electrostatic energies are of major importance for our understanding of protein energetics in general for computer-aided drug design as well as for the design of novel biocatalysts and protein therapeutics. Electrostatic energies are of particular importance in such applications as virtual screening, drug design and protein–protein docking due to the high charge density of protein ligands and small-molecule drugs, and the frequent protonation state changes observed when drugs bind to their protein targets. Therefore, the development of a reliable and fast algorithm for the evaluation of electrostatic free energies, as an important contributor to the overall protein energy function, has been the focus for many scientists over the past three decades. In this review we describe the current state-of-the-art in modeling electrostatic effects in proteins and protein–ligand complexes. We focus mainly on the merits and drawbacks of the continuum methodology, and speculate on future directions in refining algorithms for calculating electrostatic energies in proteins using experimental data.

Electrostatic interactions are among the most important factors to be considered when analyzing the function of biological molecules. Since the pioneering work of Linderstrøm-Lang [1], significant progress has been made in the qualitative understanding and quantitative evaluation of electrostatic interactions. It is now generally recognized that one must analyze the electrostatic forces in a biomolecule to properly understand its function. Structure-based calculations of the electrostatic potential in and around a biomolecule, therefore, play a large role in virtually any protein-related research effort, as witnessed by the omnipresence of graphical representations of electrostatic potential surface maps in publications and oral presentations.

A variety of computational methods and algorithms for studying proteins and their interactions with other molecules has been constructed. Methods such as molecular dynamics (MD) simulations, Brownian dynamics simulations, protein pKa calculations, protein-design algorithms and protein–drug and protein–protein docking algorithms are widely used in modern biological research. All these algorithms use 3D structures of protein molecules to predict and analyze protein characteristics, such as catalytic activity, folding pathway, stability, solubility and ligand/drug binding specificity. Electrostatic energies play a large role in determining all of the aforementioned characteristics, and the full potential of methods for predicting physical/biochemical characteristics will, therefore, only be realized once we are able to accurately calculate electrostatic energies and forces in and around proteins. The calculation of electrostatic interactions in biomolecules arguably presents one of the largest obstacles in improving the accuracy and usefulness of structure-based energy-calculation algorithms.

Unfortunately, electrostatic phenomena in biomolecular systems are very complex and much research is needed in this field to understand the factors that are important for determining their properties. The comparatively long-range nature of electrostatic forces presents a significant obstacle to progress, since one must consider interactions between a large number of solvent and solute atoms in order to solve the problem correctly. The many charged groups in proteins and the protein-induced variation in amino acid pKa values further complicate the calculation, since the protein charge distribution, which itself is dependent on the electrostatics in the protein, has to be calculated. The heterogeneous dielectric properties of the protein interior complicates the evaluation of electrostatic interactions even further by requiring energy-calculation methods to model protein and solvent reorganization in response to the variations in electrostatic field. Calculating the electrostatic field in a protein molecule correctly is thus akin to hitting a moving target using a shotgun with a bent barrel while being in the middle of an earthquake.

Despite the somewhat bleak analogy and the aforementioned difficulties, current algorithms have made significant headway in accurately predicting electrostatic effects in proteins. In this

Predrag Kukić[1,2] & Jens Erik Nielsen[1†]
[†]Author for correspondence
[1]School of Biomolecular and Biomedical Science, Centre for Synthesis and Chemical Biology, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland
[2]School of Computer Science and Informatics, Complex and Adaptive Systems Laboratory, University College, Dublin, Belfield, Dublin 4, Ireland
Tel.: +35 317 166 724
Fax: +35 317 166 898
E-mail: jens.nielsen@ucd.ie

FUTURE SCIENCE part of fsg

KEY TERMS

Continuum/macroscopic model: Model that mimics dielectric properties of a solvent using the homogenous dielectric medium of a high dielectric constant

Microscopic model: Model that explicitly represents all atoms in the protein–solvent system

Poisson–Boltzmann equation: Partial differential equation, which, for a known distribution of charges, the dielectric constant and ionic strength gives the electrostatic potential in and around the protein

review we focus on the methods that have proven effective and accurate in modeling electrostatic effects in proteins and protein–ligand complexes. We begin our review by introducing the concepts of electrostatic interactions in proteins and the basic laws and models necessary for electrostatic energy calculations, such as Coulomb's law, the Poisson equation, the Boltzmann distribution of ions in solution and the Born formula. Later, we provide an overview of the specific computational approaches used for modeling electrostatic interactions in proteins and protein–ligand complexes. We describe the widely used continuum models; the Poisson–Boltzmann equation (PBE) solvers and the models based on the Generalized Born (GB) framework, as well as highlight recent progress in improving the accuracy of these methods. We discuss the role of electrostatic energies in the association of proteins and ligands. We concentrate on the electrostatic contribution to protein–ligand binding affinities, the pH-dependence of the free energy of the binding process, and on the role of electrostatics in determining the association rates for diffusion-limited enzymes. Finally, we comment on current problems and discuss possible directions for future advances in the modeling of protein electrostatics.

This review gives an overview of basic methods for predicting protein electrostatic forces and their application to protein–ligand binding problems. Many other exciting areas of research also present formidable challenges to current models for predicting electrostatic energies, with computational protein design, the engineering of the pH-dependence of enzyme catalysis, prediction of protein-folding pathways and prediction of mutation-induced changes in protein stability presenting major challenges to computational biologists. It is outside the scope of this review to address all of these areas, and the interested reader is referred to a selection of recent reviews [2–11].

Electrostatic interactions in proteins

Before embarking on a detailed account of protein electrostatic modeling, we describe the basic equations that allow a qualitative understanding of the factors that influence the strength of electrostatic interactions in and around proteins. We begin our discussion with Coulomb's law, the Poisson equation, the Boltzmann distribution of ions in the solution and the Born model. We will proceed to explain the significance of permanent dipoles, mobile ions and electronic polarization in modulating the electrostatic field in the protein–solvent system.

■ Coulomb's law

The fundamental problem in classical electrostatic theory is to find the electrostatic potential at every point in space for a given distribution of charges. In the simplest scenario, when point charges are placed in a homogenous medium, such as a vacuum, the electrostatic potential $\varphi(r)$ at a particular position in the space can be calculated with Coulomb's law:

$$\varphi(r) = \frac{1}{4\pi\varepsilon_0\varepsilon}\sum_i \frac{q_i}{r_i}$$

EQUATION 1

where φ represents the electrostatic potential, q_i is the charge of that point charge and r_i is the distance from the point charge i. ε represents the dielectric constant of the medium, and a value of $\varepsilon_0 \sim 8.85 \times 10^{-12}$ F/m quantifies the propagation of the electric field in vacuum.

Another form of Coulomb's law frequently used in protein applications evaluates the total electrostatic interaction energy of the system consisting of N point charges immersed in the homogenous medium. The equation can be written as:

$$\Delta G_{el} = \frac{332\,kcal/mol}{\varepsilon_r}\sum_{i=1}^{N}\sum_{\substack{j=1,\\ i<j}}^{N}\frac{q_i q_j}{r_{ij}}$$

EQUATION 2

where ΔG_{el} designates the electrostatic interaction energy (kcal/mol at room temperature) relative to the energy between charges positioned at infinite separation. q_i and q_j are point charges given in units of electron charge ($e \sim 1.6 \times 10^{-19}$ C) and divided by the distance r_{ij} (Å). ε_r represents the relative dielectric constant of the system and is expressed by the value relative to that of a vacuum. This form of Coulomb's law has been frequently used in the microscopic modeling of electrostatic interactions in proteins, where the electrostatic free energies between all charges in the protein–solvent system are evaluated using **EQUATION 2** and low dielectric constant [4].

■ Poisson equation

Evaluation of the electrostatic energy in systems containing a vast amount of point charges, such as a protein–solvent system, can be a very complex and time-consuming process. Therefore, instead of explicitly representing all point charges in the system and their abilities to reorient, these effects are approximated in continuum models using the dielectric constant. The electrostatic potential in continuum models can be calculated using the Poisson equation:

$$\nabla\cdot\varepsilon(r)\nabla\varphi(r) + 4\pi\rho(r) = 0$$

EQUATION 3

where the dielectric constant, $\varepsilon(r)$, is a function of the position r, and $\rho(r)$ similarly represents the charge density in the medium. The charge density is simply the distribution of charges throughout the system, whereas $\varepsilon(r)$ reflects all effects that are not explicitly modeled, such as induced dipoles and/or the relaxation of charges. Thus, in continuum models the protein is usually represented by a dielectric constant in the range of $\varepsilon \sim 2$–20 and the solvent is modeled as a continuum of high dielectric ($\varepsilon \sim 80$).

The value of the dielectric constant in continuum models that accounts for the protein interior will be called 'the protein dielectric constant' ε_p in the following discussion. It is very important to emphasize here that the term 'protein dielectric constant' in continuum modeling does not take a constant value, since it depends on the way the constant is used in the calculations and which continuum model is used. This protein dielectric constant in continuum models has little to do with the fluctuations of dipole moments, as defined in microscopic modeling, but rather characterizes the parameters and effects that are not described explicitly in the specific continuum model in question. For a detailed account of the nature of the protein dielectric constant see the later section 'Modulators of electrostatic interactions in proteins' and discussions elsewhere [12].

Boltzmann distribution

Construction of the charge density $\rho(r)$ function in the Poisson equation is a straightforward process when positions of all charges are known. Permanent dipoles representing the backbone C=O bonds in proteins are not likely to reorient and the dipoles on side chains can only reorient within accepted geometries if the protein does not undergo large conformational changes. By contrast, ions in solution (Na^+, Cl^-, K^+ and Mg^{2+}) constantly change their position according to the local electrostatic potential and their interactions with surrounding water molecules. Thus, describing mobile ions in a solution explicitly in the Poisson equation is a very complicated process. Instead, their positions around a protein are described using a probability distribution function, based on the Boltzmann distribution:

$$n(r) = Ne^{-\frac{q\varphi(r)}{kT}}$$

EQUATION 4

where $n(r)$ and $\varphi(r)$ represent the concentration of ions (positive or negative) and mean potential at particular location r in the solution. N is the bulk concentration of ions, k is Boltzmann's constant ($k \sim 1.38 \times 10^{-23}$ J/K) and q represents the charge of the ion considered.

Given the expression for the concentration of ions in the solution **(EQUATION 4)**, the charge density of mobile ions $\rho_{ion}(r)$ in the solution can be derived using the simple equation:

$$\rho_{ion}(r) = qn(r)$$

EQUATION 5

EQUATIONS 4 & 5 for all ion types present in the solution are usually incorporated in the Poisson equation yielding the well-known PBE.

The Poisson-Boltzmann equation

The PBE is the most frequently used equation in continuum modeling of the protein–solvent system and can be written as:

$$\nabla \cdot \varepsilon(r)\nabla\varphi(r) - \varepsilon(r)\kappa(r)^2 \sinh(\varphi(r)) + 4\pi\rho(r)/kT = 0$$

EQUATION 6

where κ is the parameter dependent on the ionic strength of the solution, and is called the Debye–Hückel inverse length. It is related to the ionic strength (I) by the equation:

$$\kappa^2 = \frac{8\pi N_A e^2 I}{1000kT}$$

EQUATION 7

where N_A, e and k represent Avogadro's number, the electronic charge and Boltzmann's constant, respectively. T is the absolute temperature of the solution. The ionic strength (I) of the solution has a profound effect on electrostatic interactions in protein solutions. By changing the ionic strength of the solution one can increase or decrease electrostatic attractions/repulsions between charges in the solution and, thus, obtain an experimental estimate of the importance of electrostatic interactions for the quantity observed.

EQUATION 6 represents the nonlinear form of the PBE, and it is often linearized in protein applications by assuming $\sinh\varphi(r) \sim \varphi(r)$, which results in:

$$\nabla \cdot \varepsilon(r)\nabla\varphi(r) - \varepsilon(r)\kappa(r)^2 \varphi(r) + 4\pi\rho(r)/kT = 0$$

EQUATION 8

Analytic solutions of the PBE are only possible for simple geometric objects and, therefore, the PBE is typically solved for biomolecules using an iterative finite-difference approach [13]. Briefly, the protein and the solvent are mapped onto a cubic lattice with the parameters of the PBE [$\varepsilon(r)$, $\rho(r)$ and $\kappa(r)$] assigned to each point in the grid. Using these position-dependent parameters one

can account for not only the heterogeneous nature of the protein dielectric constant in continuum models, but also the variable concentration of ions with different valence in solution [14,15]. Once the parameters have been assigned, the PBE relates the flux of the electric field out of each grid cube to the charge inside that cube and the potentials on all grid points are iterated until the flux and potentials of neighboring cubes match.

Typically the PBE is solved on a 65 × 65 × 65 cubic grid, although larger and smaller grids can be used. For most proteins, a 65-cubed grid provides an adequate resolution for producing qualitative images of protein surface electrostatic potentials, whereas for quantitative calculations a resolution of 0.25 Å per grid point is required. In these cases, successive 'focusing' runs are typically performed to avoid the computational resource problems associated with solving the PBE on a very large grid **(Figure 1)**.

■ Solvation energy & the Born model

The solvation free energy (ΔG_{sol}) is an integral component of all free energy calculations in biomolecules, defined as a free energy change associated with transferring a molecule from vacuum to its position in the solvent. The solvation free energy consists of polar and nonpolar contributions. The polar contributions are typically dominant for charged molecules and consist of the so-called 'self-energy' and interaction energies with charges and dipoles in solution. The nonpolar contributions consist of the energy needed to create the cavity against the solvent pressure, the Van der Waals interactions with the solvent and the entropic penalty incurred by reorganizing the solvent around the molecule. In the following discussion we will only consider the polar part of the solvation free energy.

Born derived a very useful approximation of the solvation free energy associated with moving a charge from vacuum to a spherical solvent cavity, which is known as the Born model [16]:

$$\Delta G_{sol}^{pol} = -166 kcal/mol \frac{q^2}{a}\left(\frac{1}{\varepsilon_v} - \frac{1}{\varepsilon}\right), \varepsilon_v = 1$$

Equation 9

where q refers to the amount of charge (units of e), ε is the relative dielectric constant of the solvent and a is the radius of the cavity (Å). The Born model has found application in almost all models of electrostatic interactions in proteins. The Born equation **(Equation 9)** in combination with Coulomb's law **(Equation 2)** can be applied not only in moving a charge from a vacuum, but also in moving a charge between any two homogenous media, such as when calculating the free energy change of moving a drug from solution to its position in the target protein (see the later dicussion on GB formalism).

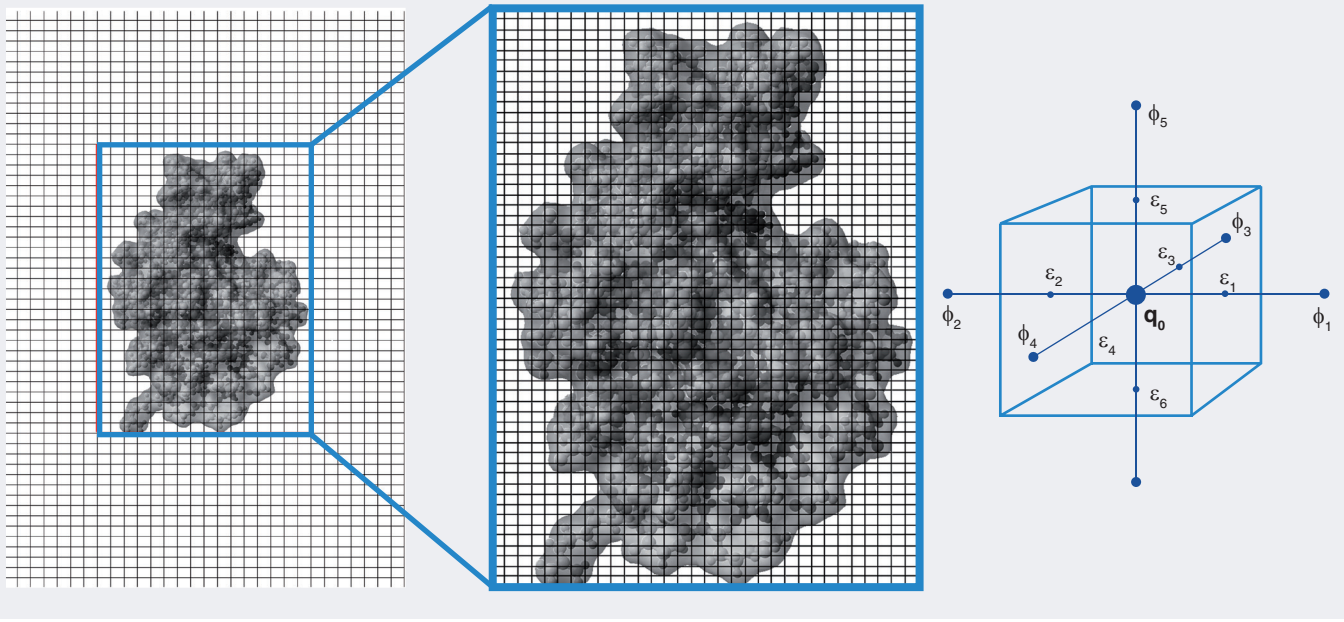

Figure 1. 'Sequential focusing' algorithm in solving the finite difference Poisson–Boltzmann equation. The coarse grid is used for the protein and the solvent, whereas the fine grid focuses only on the region of interest. Dielectric constant, charge density and ionic strength are assigned to each point of the grid in the iterative procedure. Electrostatic potential values calculated in the current step are used as the initial values in the next step.

■ Modulators of electrostatic interactions in proteins

Protein x-ray and NMR structures are used extensively in free energy calculations and in the following discussion we describe the relevant factors that should be considered when one applies classical electrostatic theory to protein 3D structures. We pay particular attention to the role of permanent dipoles and induced dipoles in calculations of the electrostatic potential (electrostatic field) inside and outside proteins.

Permanent dipoles originate from the tendency of electrons to concentrate around atoms with large electronegativities, forming an excess of negative charge in these atoms. The relative absence of electrons similarly creates a partial positive charge on atoms with small electronegativities. In electrostatic calculations, permanent dipoles are represented by partial charges in molecular force fields and these are usually derived from experimental data or from *ab initio* calculations [17–21].

A permanent dipole in the solute molecule can interact with other permanent and induced dipoles in the solute and with surrounding water molecules. These interactions are very complex and can produce the reorientation of permanent protein dipoles. However, the degree of reorientation is limited by covalent bonds and other steric effects in proteins (e.g., hydrogen bonds and Van der Waals interactions), such that the dipoles on side chains can only reorient within the accepted geometries and the backbone dipoles generally remain fixed. Therefore, knowing the protein 3D structure, we are able to assign the partial charges to specific atoms and bonds in the protein structure, and use them explicitly in the PBE **(EQUATION 8)** as an integral part of the charge distribution, the value $\rho(r)$. The reorientation of protein partial charges due to a structural perturbation, such as the introduction of a ligand or the titration of an ionizable group, can be estimated from MD simulations or directly from a known 3D protein structure. If the response of the protein structure is modeled correctly and the resulting structure(s) are used for PBE calculations, then all other polarization effects are included implicitly in the value of the protein dielectric constant, ε_p, when applying the PBE.

Permanent dipoles of water molecules show different behavior depending on their position in the solution. Water dipoles far away from the protein surface, so-called bulk water, have higher degrees of freedom than interfacial water molecules and permanent protein dipoles. Thus, the complex translational and rotational motions of bulk water molecules are treated implicitly in the PBE and are modeled as a continuum with a large dielectric constant. The dielectric constant of bulk water, ε_w, can be measured accurately, and a value of 78.46 is used in calculations of the phenomena at 25°C (0.1-MPa pressure) [22].

The reorientation of interfacial water molecules due to an imposed electrostatic field is considerably limited by the network of hydrogen bonds with solute atoms and other interfacial water molecules. Water molecules bound to the charged surface residues have recently been suggested to behave more like extensions of the solute rather than mobile bulk waters [23,24]. This is particularly the case with internal water molecules, which can bridge electrostatic interactions between solute atoms and, thus, play a prominent role in enzyme catalysis and ligand binding [25,26]. Although theoretical studies have provided insights into the role of these molecules, our knowledge of their importance is still quite limited. For reasons of simplicity, interfacial and internal water molecules are often treated as bulk water in PBE calculations. This simplification is likely to be one of the main reasons for inaccurate predictions of the electrostatic potential in proteins in which such discrete water molecules play an important role [27].

Induced dipoles exist in proteins as a consequence of the phenomenon known as electron polarization. Electron polarization accounts for the reorientation of the electronic cloud around a nucleus in the presence of the local electrostatic field $E(r_i)$, and is given by:

$$(\mu_i)^{n+1} = \alpha E(r_i)^n$$

EQUATION 10

where μ_i represents the dipole moment of the induced dipole i, defined by $\boldsymbol{\mu}_i = \boldsymbol{l}_i q_i$ ($\boldsymbol{l}_i$ is the vector between two opposing induced charges q_i and $-q_i$), and α is the proportionality constant called the molecular polarizability. $E(r_i)$ designates the sum of electrostatic fields originating not only from partial charges and ions in solute and solution, but also from reoriented electronic clouds existing on other atoms in and around proteins. The values of μ_i for all induced dipoles in the protein are solved using the iterative procedure, where μ_i^{n+1} values are determined by the field $E(r_i)^n$ from the previous iteration. Iterations of this complex many-body problem converge only with difficulty (sometimes leading to a so-called 'polarization catastrophe') and represent one of the main obstacles in accurately

evaluating electrostatic interactions in proteins at the atomic level. In order to avoid the polarization catastrophe, the damping parameter is usually introduced and, thus, the particularly long-range effect of electrostatic interactions in some proteins can be underestimated [KUKIC *ET AL.*, MANUSCRIPT IN PREPARATION].

In order to avoid the explicit representation of induced dipoles in the solute by including them in the charge distribution $\rho(r)$ **(EQUATION 8)**, the induced dipoles are instead replaced by $\varepsilon_p = 2$ if all other effects (permanent dipoles and protein relaxation) are included explicitly in the PBE [12]. Induced dipoles of water molecules are also modeled implicitly in PBE using a dielectric constant of water of $\varepsilon_w = 80$.

By using the spatially invariable protein dielectric constant (ε_p) in the PBE to account for all polarizability phenomena, the heterogeneous dielectric nature of the protein interior is ignored. Therefore, efforts are currently being made to develop polarizable force fields that explicitly account for the effect of electronic polarization in the PBE. Although polarizable force fields such as AMOEBA [28], CHARMM [20,21] and PFF [18] have been developed in the last few years, parameter sets for these polarizable force fields are still under development and their behavior in free energy calculations is not yet completely understood. Readers interested in using polarizable force fields to evaluate protein–ligand binding free energies are encouraged to read the review by Warshel *et al.* [19] and recent article by Maple *et al.* 18].

Modeling electrostatic interactions in proteins

The importance of predicting protein characteristics, such as catalytic activity, stability and ligand/drug binding affinities, from protein 3D structures has led to the development of a wide range of electrostatic models in the last few decades. These models have been divided by Warshel and co-workers into three groups: microscopic all-atom models, simplified microscopic models and continuum (macroscopic) models [4] (see Figure 1 as presented by Warshel *et al.* [4]).

Microscopic all-atom models use detailed all-atom representations of both the protein and the solvent and can be further divided into classical mechanical and quantum mechanical models. Formally rigorous all-atom models based on statistical mechanical approaches or molecular mechanics (MD; free energy perturbation [FEP] models and others) are very slow and often use inappropriate treatment of long-range electrostatic interactions [27]. These models are less attractive for calculating protein electrostatic energies because of the long convergence/calculation times. However, less computationally expensive all-atom methods, such as the linear response approximation (LRA) and the linear interaction energy (LIE) methods, have emerged as useful tools in structure-based drug design. These methods, together with more approximate and significantly faster approaches (PBE, Poisson–Boltzmann surface area [PBSA] and GB), which have proved to be quite effective, will be the main focus of this section.

■ Simplified microscopic models

It is a time-consuming and complicated process to correctly treat the large number of water molecules in microscopic all-atom models when calculating the electrostatic field [4]. Therefore, many simplified models have been developed. The protein dipoles langevin dipoles (PDLD) model represents a simplified microscopic model where water molecules close to the catalytic/binding site are accounted for by dipole moments μ_i^L of so-called Langevin dipoles, whereas distant water molecules are treated as a continuum with an ε_w of 80. Langevin dipoles are point dipoles whose time-averaged polarization is represented by a Langevin-type function rather than linear function given by **EQUATION 10**. An example of Langevin dipoles positioned on a cubic lattice around a protein–ligand complex is depicted in **FIGURE 2**.

Induced dipoles of solutes are included explicitly in the PDLD model, by assigning dipole moments μ_i to all atoms, according to **EQUATION 10**. The values of the solute point charges are obtained from force fields, whereas the distribution of mobile ions in the solution is approximated using the Boltzmann distribution **(EQUATION 4)**. The reorganization of the solute due to changes in the charge distribution is accounted for by MD simulations [29].

Since the charges in the PDLD model are not treated implicitly by a dielectric constant and since the relaxation of the protein structure is approximated using MD simulations, all relevant electrostatic interaction energies between solute charges are calculated using Coulomb's law **(EQUATION 2)** and a dielectric constant of 1. The solvation energy of the solute due to the surrounding Langevin dipoles is also estimated using Coulomb's law, whereas the solvation

energy of the solute due to the surrounding distant water, represented as a continuum in the PDLD model, is calculated using the GB solvent model. The electrostatic energy between ions and solute is estimated using Coulomb's law with a value of the dielectric constant being ε_p approximately 40–80 (for more details on the PDLD model see elsewhere [4]).

The PDLD model, like all microscopic all-atom models, deals with large energy contributions (i.e., electrostatic energies are not scaled by the dielectric constant in Coulomb's law). Therefore, even small errors in calculating individual electrostatic energies can lead to large absolute errors when treating systems with thousands of atoms. In order to scale large energy contributions in the PDLD model and achieve the precision of the current semi-macroscopic models, Warshel and co-workers further introduced the 'scaled microscopic' PDLD/S model [30]. The novelty in the PDLD/S model is represented by the fact that the solute-induced dipoles (and the reorganization of solute permanent dipoles) are accounted for implicitly by using a protein dielectric constant value of ε_p approximately 2–6. The scaled PDLD model defined in this way gives results that are more stable that its microscopic counterpart [29,30].

The accuracy of the PDLD and PDLD/S models in reproducing electrostatic free energies has been assessed with a diverse set of problems in recent years. The problems range from the examination of solvation energies of small molecules in solutions, the prediction of solvation energies of ionizable groups in solvated proteins and electrostatic interactions between them, to the evaluation of electrostatic free energies in enzyme catalysis and ligand binding [4,27,30–33]. Calculations with the standard PDLD model showed good agreement between calculated and experimental solvation energies of a variety of molecules in solution [30]. Furthermore, both the PDLD and the PDLD/S models were demonstrated as being capable of reproducing self-energies of acidic residues in proteins, charge–charge interaction energies between surface titratable groups and charge–charge interaction energies between buried titratable groups in the selected protein systems. The relevant benchmark energies in these studies were derived using experimental pH-dependent titration curves of nuclei and mutagenesis studies of bovine pancreatic trypsin inhibitor, subtilisin and cytochrome C protein systems. Compared with the PDLD, the PDLD/S model increases the precision of the energy calculations, mainly by scaling the relevant electrostatic energy terms (solvation and charge–charge interaction energy).

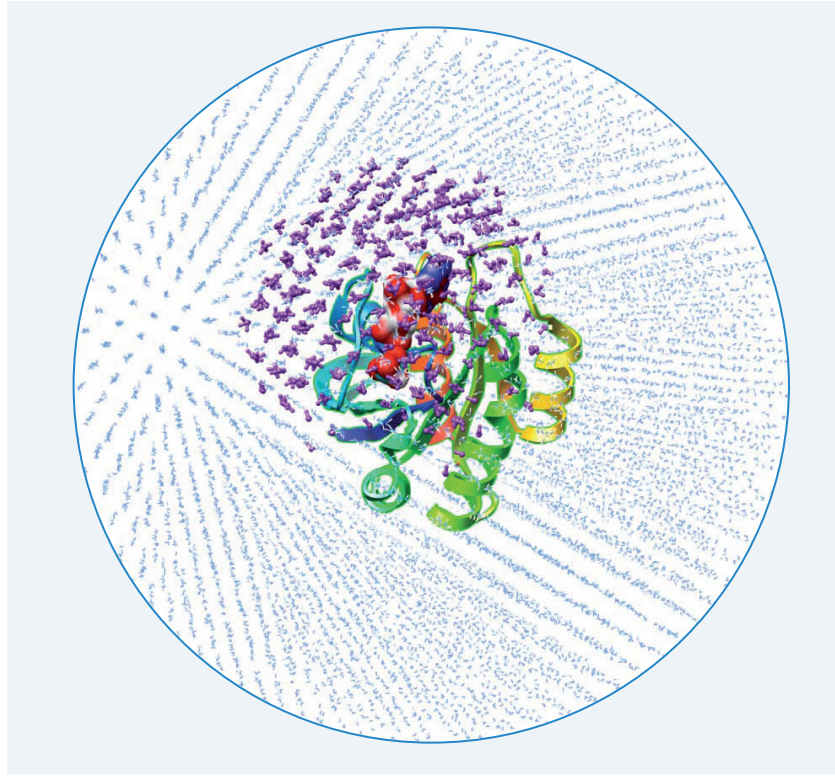

Figure 2. Simulation set-up for efficient free energy ligand-binding calculation. The protein (given by the secondary structure) and substrate (red) are represented explicitly, as is a sphere (radius 20 Å) of water (purple) centered on the binding site. The electrostatic effect of distant water on the binding site is approximated by Langevin dipoles (blue). The atoms in the volume defined by the explicit solvation sphere are simulated freely while those outside the sphere are restrained.

The PDLD and PDLD/S models have been further tested in enzyme catalysis and protein–ligand binding applications, and have produced reasonable results [4,27,30–33]. The PDLD/S model with $\varepsilon_p = 6$ gave best results in reproducing absolute reaction free energies of the trypsin and lysozome-catalyzed reactions.

■ Continuum models

Continuum models for modeling electrostatic interactions in proteins achieve a high level of computational efficiency by significantly reducing the level of detail used to represent the solvent surrounding the protein. All continuum models treat the solute(s) as low-dielectric media with embedded atomic charges that are immersed in a high-dielectric ionic solution. These models predict the electrostatic potential at every point in and around a protein for a given distribution of atomic charges and a given ionic strength.

KEY TERMS

NMR titration experiment: Experiment that measures chemical shift perturbations of nuclei when changing the pH of the solution (i.e., changing the charges on titratable groups in a protein)

GB model: Fast approximation to the PBE model in evaluating electrostatic free energy

Graphical representations of electrostatic potential maps produced with continuum models have revealed the importance of electrostatic interactions in many biomolecular complexes in the last three decades. In particular, the areas of ligand-binding specificity and diffusion-controlled ligand binding have been treated extensively with continuum methods. Here, we give an overview of the main characteristics of current continuum models. We concentrate on models based on the PBE and the GB framework.

■ PBE solvers & the protein dielectric constant

Poisson–Boltzmann equation solvers are able to calculate the electrostatic potential of complex protein molecules by using a spatial variation of the dielectric constants (ε), the ionic strength (*I*), and the charge distribution (ρ) **(EQUATION 8)**. The PBE can furthermore be solved quickly by employing an iterative procedure **(FIGURE 1)** [34]. Permanent solute dipoles are modeled explicitly in the PBE as point charges located at atom centers, whereas mobile ions in solution are accounted for using the Boltzmann distribution **(EQUATION 4)**. The dielectric constant models all contributions that are not explicitly included in the PBE and its value differs for the protein and the solvent. The dielectric constant used for the solvent (ε_w = 80) describes both induced and permanent dipoles in the solvent, as well as the relaxation of the solvent around the protein. The protein dielectric constant ε_p usually has a value between 2–20 depending on the exact type of calculation and the protein structure modeling scheme. When ε_p reflects only the effect of electronic polarizability, values of approximately 2 are used. When the reorganization of charges together with electronic polarizability are represented implicitly in the PBE model, the optimal value of ε_p can sometimes increase to values of more than 20 [35,36].

The main problem in employing ε_p is that the value used for one application cannot be transferred to any other. Typical examples of calculations that use different ε_p are those of charge–charge interactions and/or calculations of charge-solvation energies. When calculating charge–charge interactions, assuming the rigid structure of the protein, the interaction energies calculated depend on the shape of the macromolecule and the exact position of the charges in the protein. Therefore, different values of ε_p in PBE are necessary in order to mimic electrostatic interaction energies between different charged groups in proteins obtained by NMR titration experiment. Values as large as ε_p = 40 can be optimal for the calculation of charge–charge interactions in bovine βLG and HEWL [KUKIC *ET AL.*, MANUSCRIPT IN PREPARATION]. Furthermore, calculations of solvation penalties for moving ionizable groups from water to their position in the protein have been suggested to be more accurate if a smaller protein dielectric constant is used (ε_p ~ 2–6) [4,27]. Values of the protein dielectric constant are dependent on the parameters that are not described explicitly and, therefore, should be optimized separately for each particular application in order to achieve the desired level of accuracy.

A variety of PBE solvers are readily available, ranging from multipurpose software packages (UHBD [37] and CHARMM [38], for example) to stand-alone PBE solvers (APBS [13], DelPhi [39] and ZAP [201], for example). These PBE solvers have been used successfully in recent years for predicting solvation energies and electrostatic energy contributions to a wide range of processes. The role of solvation and electrostatic interaction energies on ligand/drug-binding processes and, therefore, the importance of PBE models in this phenomenon will be considered in detail later in this review. For more information on using PBE to calculate protein pKa values of titratable groups as well as to predict the pH-dependence of catalytic activity, protein stability and ligand-binding energies, see the review by Nielsen and the references therein [6].

■ PBE models in conjunction with solvent-accessible surface area

Predictions of solvation free energies using PBE models are often augmented with a term that describes the nonpolar contribution to solvation. The additional nonpolar term is usually estimated from the solvent-accessible surface area (SA), and PBE models used in conjunction with SA are referred to as PBSA models [40,41]. The nonpolar solvation free energy $\Delta G_{sol}^{non-pol}$ that accounts for the cost of cavity formation and Van der Waals interactions with the solvent can be approximated using an equation of the following form:

$$\Delta G_{sol}^{non-pol} = \gamma A + b$$

EQUATION 11

with *A* representing the SA, and γ and *b* being the adjustable parameters. The values of γ and *b* are usually calibrated using experimentally determined solvation energies of small molecules.

With recent improvements in calculating the polar contributions of solvation free energies using MD simulations in conjunction with PBE/GB models (see 'Macroscopic models turn microscopic), the limitations of this simple SA model **(Equation 11)** are becoming more obvious [42–44]. New efforts to develop an extended treatment of nonpolar solvation free energies, especially by separating the costs of Van der Waals interactions with the solvent and costs of cavity formations, have been constructed [45], and will hopefully increase the level of accuracy of current continuum models.

■ GB formalism

Binding free energies of protein–ligand complexes can be evaluated using the PBE (or PBSA) models in minutes of computer time. However, the screening of thousands of potential drugs usually requires higher computational efficiency than the PBE model can deliver. Therefore, very fast analytical methods based on the Born equation **(Equation 9)**, GB models, have been widely used in the calculation of free energies of ligand binding, in computer-aided drug design (CADD) and in conformational analysis of proteins.

The GB models for ligand binding are based on the assumption that the screening of electrostatic interactions between two charges can be estimated by the degree to which the charges interact with the solvent [46]. The solutes in the GB models are described by a collection of atoms, where each atom is represented by a sphere of a radius a_i, referred to as 'effective Born solvation radius'. The point charge of each atom q_i is positioned at the sphere's centre. As in the PBE models, the molecule in the GB model is described as a low-dielectric volume, ε_p, surrounded by a high-dielectric aqueous environment ε_w. Electrostatic interaction energies between point charges in GB models are calculated using Coulomb interactions *in vacuo* **(Equation 2)**. The electrostatic solvation energy ('polarization energy') of transferring a molecule from the protein to the aqueous environment is given by [40]:

$$\Delta G_{pol} = -166 kcal/mol\left(\frac{1}{\varepsilon_p}-\frac{1}{\varepsilon_w}\right)\sum_i^N\sum_j^N \frac{q_i q_j}{f_{GB}(r_{ij}, a_i, a_j)}$$

Equation 12

where N is the number of atoms in the system, r_{ij} is the distance between atoms i and j, and f_{GB} is the function that involves several empirical parameters. The effective Born solvation radius a_i of the atom i can be thought of as the distance between the charge and protein–solvent boundary. This parameter is adjustable and its value is usually determined using the solvation free energy from the PBE by:

$$a_i = -166\text{Å}\frac{q^2}{a}\left(\frac{1}{\varepsilon_p}-\frac{1}{\varepsilon_w}\right)\frac{1}{\Delta G^i_{pol}}$$

Equation 13

where ΔG^i_{pol} designates the solvation free energy of the unit charge positioned at the centre of the atom i in the uncharged solute, whereas ε_p and ε_w represent relative dielectric constants of the solute and the solution in the PBE.

Generalised Born-based models are unable to calculate electrostatic potential maps of proteins and their parameters are usually optimized using the results obtained with PBE solvers [47–51]. Optimized and calibrated GB models are usually augmented by the hydrophobic contribution to binding and referred to as GBSA models. With improvements in recent years, current GBSA models are capable of achieving the similar level of accuracy as numerical PBSA methods [42,52]. Their high computational efficiency makes them suitable for drug-design applications and different versions of GBSA models are now implemented in many ligand-docking programs [53–56].

■ Macroscopic models turn microscopic

The precision of macroscopic models that use rigid protein structures, as emphasized many times in the previous discussion, is highly connected with the value used for the protein dielectric constant ε_p. Therefore, recent efforts in improving the existing macroscopic models have mainly focused on limiting the number of effects that are treated implicitly and 'hidden' in the protein dielectric constant. In other words, there is a drive towards making macroscopic models more microscopic. The number of effects treated implicitly by the ε_p value can be limited by treating these effects explicitly, for example by modeling electronic polarization or protein conformational flexibility explicitly.

Several polarizable force fields have been reported recently, and it is only a matter of time before continuum models will be able to explicitly account for the induced dipoles in protein interiors. Particularly promising is the work of Schnieders and co-workers in developing the Polarizable Multipole Poisson–Boltzmann (PMPB) continuum electrostatics model, which is built on the AMOEBA polarizable force

field [28]. The PMPB model has been tested on predicting the change of the protein dipole moment when the protein is moved from vacuum to solution. The model proved able to mimic results obtained from the MD simulations with explicit water.

Another direction in improving the continuum paradigm includes the explicit incorporation of structural flexibility in the models. In classical PBE models, the response of the solute structure to changes in charge distribution (e.g., the pH-dependent titration of a titratable group) is accounted for by an average isotropic ε_p value. Therefore, PBE models are not able to account for the local anisotropic structural rearrangements in the protein.

The first step in treating the structural relaxation microscopically in PBE models includes the optimization of hydrogen bonds and/or side chains of individual residues [57,58]. For example, selected side chains in the protein were assigned a few predetermined rotamers depending on the charge distribution. These models have proven very successful in pKa predictions and they are still widely used. However, they still approximate the relaxation of the rest of the protein implicitly through the poorly defined protein dielectric constant.

A more radical step in including protein conformational flexibility explicitly in continuum models is to treat the solute using MD simulations. With this approach, the dielectric relaxation of a solute is modeled by structural rearrangements during the MD simulations. On the contrary, the effects of a high-dielectric solvent are still approximated by a continuum, and, therefore, the computationally expensive sampling over water degrees of freedom is still avoided. The first MD simulations of a protein using both the PBE and the GB models were performed by Gilson *et al.* for HIV-1 protease [59]. Ever since, MD simulations with GB and PBE models have been applied to study structural and dynamic properties of proteins [60], protein stability [61] and to predict pKa values [62,63].

The successful implementation of continuum models in MD simulations influenced the development of constant pH MD. Unlike the classical MD simulations, where protonation states of all titratable groups are kept fixed, the constant pH MD simulations allow the change of protonation states of titratable groups during the simulations. The sampling of protonation states in the protein is usually implemented using the PBE [64–66] or GB model [62], combined with Monte Carlo sampling. The constant pH MD simulations account for the aqueous environment in a realistic way and, therefore, their main contribution is likely to be in pH-dependent conformational changes and the pH-dependence of protein stability [42]. The accuracy of constant pH MD simulations have been assessed in reproducing experimental pKa values, and results have generally been encouraging [63,66,67].

The recently developed MD-based models are also prevalent in studies of ligand binding and computational drug design. The LRA and the LIE methods have been developed by Warshel and Åquist as an alternative to more time-consuming FEP-based approaches. Both the LRA and the LIE methods proved to be more useful in computationally demanding problems, as they employ conformational averaging only in the associated and dissociated states of a protein and a ligand.

The free binding energy in the LRA method is decomposed in electrostatic and nonelectrostatic components. The nonelectrostatic component is hard to estimate and consists of hydrophobic, Van der Waals, conformational entropy and water-penetration terms. The electrostatic part in the LRA model is evaluated from MD simulations in charged and noncharged forms of the ligand, by averaging electrostatic energies using the following equation (see the thermodynamical cycle in Figure 2 in the paper by Sham *et al.* [68]):

$$\Delta G^{el}_{bind} = \frac{1}{2}\left[\left\langle U^{el}_{P\text{-}L}\right\rangle_{on} + \left\langle U^{el}_{P\text{-}L}\right\rangle_{off} + \left\langle U^{el}_{W\text{-}L}\right\rangle_{on} + \left\langle U^{el}_{W\text{-}L}\right\rangle_{off}\right]$$

Equation 14

where U^{el} represents the electrostatic energy between the ligand (L) and its surroundings, protein (P) and water (W), and is calculated using Coulomb's law **(Equation 2)**. Brackets designate MD averages over trajectories obtained in both charged (*on*) and noncharged (*off*) states of the ligand.

The LRA method has usually been tested in combination with PDLD (PDLD–LRA) and PDLD/S (PDLD/S–LRA) models and performed particularly well in discriminating ligands with different binding affinities for the given receptor [27]. The exceptionally good correlation between PDLD/S–LRA predictions with a low ε_p value and observed binding free energies of HIV protease inhibitors has been recently reported [27,68].

The LIE approach adopts the electrostatic contribution to binding free energy given by **Equation 14**, but neglects the $(U^{el})_{off}$ terms in both protein and water environments. It has been shown that this approximation is valid in

ligand–water interactions, but can be important in protein–ligand interactions, when the protein provides a preorganized environment in other to stabilize the residual ligand charges [68]. The nonelectrostatic component in the LIE method is approximated using an empirical scaling of Van der Waals interactions for the bound and free ligand with its surroundings, protein and water.

The reliability of the LIE model in reproducing experimentally determined free binding energies has been extensively tested in recent years [69–71], and results showed the average unsigned error of 0.8 kcal/mol in predicting the binding free energy of binding a set of non-nucleoside inhibitors to HIV-1 reverse transcriptase [72]. The LIE model was successfully used in combination with docking algorithms and proved to be able to correctly predict binding modes for a set of HIV-1 reverse transcriptase inhibitors [70,72]. Furthermore, the LIE approach in combination with two continuum models (PBE and GB) has been tested to predict ligand-binding free energy for a set of inhibitors against aspartic protease plasmepsin II [73]. The continuum models in this hybrid approach have been used in the evaluation of ligand–water interaction energies, and close agreement between experimental binding free energies and predicted results has been reported [73].

Another group of hybrid free-energy models extensively used in protein–ligand binding consists of: molecular mechanics (MM)/PBSA and MM/GBSA models. Both these models use MD simulations of the complex, the unbound protein and the unbound ligand in order to calculate the average potential and solvation energies. Potential energy is calculated using empirical force fields, with electrostatic contribution of charge–charge interactions evaluated by the Coulomb's law. Polar and nonpolar solvation energies are predicted using either the PBSA or the GBSA model. The free binding energy is given by [74]:

$$\Delta G_{bind} = \langle U_{PL} \rangle - \langle U_P \rangle - \langle U_L \rangle + \langle G^{sol}_{PL} \rangle - \langle G^{sol}_{P} \rangle - \langle G^{sol}_{L} \rangle - T\Delta S$$

EQUATION 15

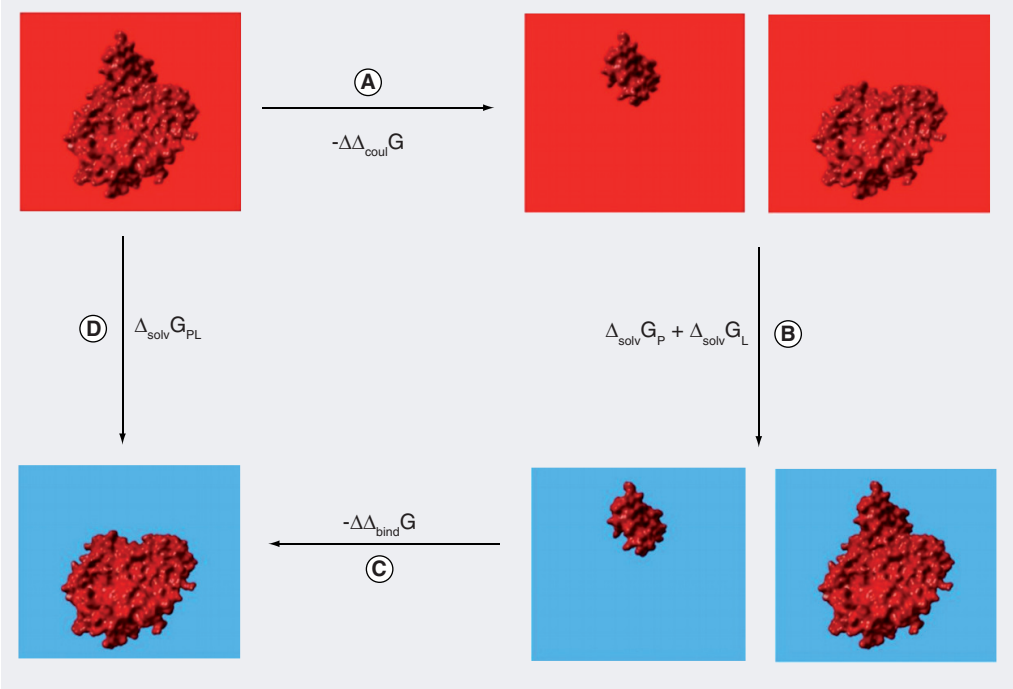

Figure 3. Thermodynamic cycle that illustrates the numerical procedure for calculating the binding free energy from PBE model. (A) Complex dissociation in a homogeneous dielectric environment with solute dielectric constant; **(B)** transfer of binding partners to an inhomogeneous dielectric environment with solute and solvent dielectric constants; **(D)** transfer of protein–ligand complex from homogeneous to inhomogeneous dielectric environment. The binding free energy of protein–ligand complex is depicted in **(C)**. It is important to stress here that many of the contributions (van der Waals energy, configurational entropy and mechanical effects) to binding free energies are neglected here.
Adapted with permission from [202].

where U and G designate potential energy and total solvation energy, respectively, in the protein, ligand and protein–ligand complex. ΔS represents the entropy change upon binding, T is the temperature, and brackets denote the Boltzmann averaging of energies over bound and unbound conformations.

The MM/GBSA has prevailed in ligand-binding applications due to higher computational efficiency comparing to the MM/PBSA model [42]. Several variations of MM/GBSA models have been tested in recent years and compared with experimental data and microscopic models [75–80]. For the majority of applications, MM/GBSA models showed the strong correlation between experimental and computational data.

Electrostatics in protein–ligand binding

Electrostatic interactions play a major role in determining protein–ligand binding specificity and the rate of protein–ligand association. An accurate and fast evaluation of electrostatic free energies of protein–ligand complexes is likely to become a key factor in future CADD studies [27]. In the following discussion we review the use of continuum electrostatic models in protein–ligand binding studies. We concentrate on the role of electrostatics in determining the association rates of diffusion-limited enzymes, the electrostatic contribution to protein–ligand binding affinities and the dominating role that electrostatics play in determining the pH-dependence of the binding constant.

■ Ligand-binding affinity

The binding affinity of a ligand to a protein controls the ligand concentration needed to achieve a high saturation of the protein. In the case of drug design, it is essential to achieve a tight association (high association constant K_A) in order to minimize the drug concentration needed during treatment. A lower drug concentration during treatment lowers the risk of side effects caused by nonspecific binding or binding to nontarget proteins, and is generally beneficial for issues related to administration, toxicology, metabolism and cost, for example. The prediction of the binding affinity of a drug typically relies on structure-based energy calculations performed according to the following thermodynamic cycle:

$$\Delta G_{bind} = \Delta G_{PL} - \Delta G_P - \Delta G_L = RT\ln K_A$$

Equation 16

where K_A represents the association constant, R is the gas constant and T refers to the absolute temperature. **Equation 16** defines the free energy of binding ΔG_{bind} as the difference between the free energy of the solvated protein–ligand complex ΔG_{PL} and the free energy of the solvated protein ΔG_p and solvated ligand ΔG_L individually. Evaluating the ΔG_{bind} value with acceptable reliability and accuracy is crucial for ranking candidate ligands in CADD. Unfortunately, it is very difficult to predict ΔG_{bind} accurately using current structure-based energy calculation methods, since issues such as conformational changes, ligand and water entropy and protein electrostatics are poorly understood and only modeled approximately using current methods.

The electrostatic free energy ΔG_{el} of binding represents an integral part of ΔG_{bind}, and the accurate evaluation of this term is of utmost importance for drug discovery. Computational models in ligand-binding applications usually decompose ΔG_{el} into a favorable charge–charge interaction energy (often called the Coulumbic energy) and an unfavorable desolvation penalty:

$$\Delta G_{el} = \Delta G_{coul} + \Delta G_{desolv}$$

Equation 17

where ΔG_{coul} refers to free energy changes in charge–charge interactions upon binding, and ΔG_{desolv} represents the energy penalty associated with moving the ligand from bulk water into the binding site of the protein. One example of the thermodynamic cycle used in the PBE solver is depicted and explained in **Figure 3**.

The prediction of ΔG_{el} is very sensitive to the particular computational model used and evaluation of ΔG_{el} has been the subject of many recent computational studies [81–84]. The computational efficiency represents a bottleneck of microscopic models in screening the potential drug candidates (i.e., these models are much too slow), whereas the problematic concept of the protein dielectric constant defines the major problem for the accuracy of macroscopic models (see previous discussion).

■ Electrostatic enhancement of diffusion-controlled protein–ligand association

Apart from the role of electrostatics in determining the binding specificity of a biological macromolecule, long-range electrostatic interactions also play a prominent role in steering charged ligands into the protein binding site.

This phenomenon was originally demonstrated for acetylcholine esterase and superoxide dismutase using Brownian dynamics simulations and protein electrostatic potential maps [85,86]. Electrostatic steering has been shown to be of particular importance in biological processes where speed is of the essence, such as the clearance of free radicals and the degradation of neurotransmitters. Therefore when designing drugs that target proteins whose function is diffusion-limited it is important to consider both the affinity of binding and association rates.

The electrostatic enhancement of binding association rate has been studied extensively in recent years by Zhou, Schreiber and co-workers [82,87–90]. In order to better understand the role of long-range electrostatics in protein–ligand association kinetics, they used the hypothetical two-step kinetic scheme:

$$P + L \underset{k_{-D}}{\overset{k_D}{\rightleftarrows}} P * L \xrightarrow{k_C} C$$

EQUATION 18

where k_D and k_{-D} represent the diffusion-controlled association and disassociation rate constants and k_C is the reaction-controlled rate constant. The loose complex (P * L) is referred to as an 'encounter complex' and is formed before the ligand is docked into its final position in the binding pocket. In comparison with the final complex C, which is stabilized by favorable interactions of many forces (hydrophobic interactions, hydrogen bonds, aromatic stacking, electrostatic and Van der Waals interactions), the encounter complex P * L is predominantly stabilized by long-range electrostatic interactions between the protein and the ligand. Favorable electrostatic interactions are produced by complementary charge distribution between binding partners, and the steering force enhances the overall association rate according to the equation:

$$k_{on} = k^0_{on} e^{-\frac{\Delta G_{el}}{RT}}$$

EQUATION 19

with k_{on} and k^0_{on} being the overall association constants in the presence and absence of electrostatic forces, and ΔG_{el} being the electrostatic free energy of the encounter complex. The accurate evaluation of the electrostatic energies represented in the ΔG_{el} term in **EQUATION 19** is necessary to accurately predict the effect of long-range electrostatics in increasing protein–ligand association rates.

The validity of Zhou and Schreiber's model and the decisive role of electrostatic forces in regulating the association rates of many protein–ligand complexes have been demonstrated by many experimental studies in recent years [89,91–96]. The model is now generally accepted and widely used in protein engineering studies [88,97,98], where one is able to predict mutation candidates inside and outside the protein binding pocket in order to achieve faster and tighter binding of protein–ligand complexes.

■ pH dependence of protein–ligand binding

The pH dependence of protein–ligand binding arises from changes in the pKa values of ionizable groups upon complex formation. It is well known that the pKa values of some ligand-titratable groups are perturbed from their solution values when the ligand binds to its receptor. The change in pKa values causes protons to be released or consumed upon binding [99–105]. The changes in pKa values of the ligand and the protein upon complexation are, therefore, of importance to the affinity of the drug [106], and pKa calculations of protein and ligand-titratable groups should thus be integrated into docking and screening algorithms [107,108].

The pH-dependent correction to a binding free energy $\Delta G_{bind,pH}$ is usually predicted using the following equation for various values of pH:

$$\Delta G_{bind,pH} = \int_{pH_{ref}}^{pH} \ln(10)\, kT \Delta Q_{bind}$$

EQUATION 20

where pH_{ref} represents the reference pH value when $\Delta G_{bind,pH}$ is assumed to be zero, k is the Boltzmann constant and T is the absolute temperature. The value ΔQ_{bind} refers to the amount of charges released into solution or consumed from it upon the binding ($\Delta Q_{bind} = Q_{PL} - Q_P - Q_L$). For more information on the pH dependence of the ligand-binding process, readers should consult the recent review by Jensen [109] and the references therein.

Several computational studies that reported changes of protein pKa values upon protein–protein binding have been reported recently, among which the studies by Klebe *et al.* [99] and Jensen *et al.* [103] are especially noteworthy. Experimental evidence of proton uptake/release induced by ligand binding was reported for a

KEY TERM

Vibrational stark effect spectroscopy: Infrared spectroscopy technique that detects changes in a protein's electrostatic field by measuring changes in the vibrational frequency of a calibrated probe inserted into the protein

series of ligands bound to the serine proteases trypsin and thrombin [110]. In the first study, authors were able to reveal receptor residues with altered protonation states using PBE model-based pKa calculations and to correctly predict the trend in ΔQ_{bind} after complex formation. In the latter study, the empirical method for pKa predictions PROPKA [111] has been used for calculating pKa values of 75 protein–protein complexes and their corresponding free forms. The results showed considerable net changes in protonation states of complexes ΔQ_{bind} relative to free forms. Furthermore, the reported pH-dependent corrections to a binding free energy $\Delta Q_{bind,pH}$ were able to induce the changes of the binding constant as high as three orders of magnitude.

Calculations of pKa values in drugs and drug–receptor complexes are time-critical since hundreds of thousands of such pKa predictions must be carried out during a virtual-screening exercise. Recently, reported models for determining pKa values of ionizable groups in drug candidates are based on quantum mechanics [112,113] and empirical predictions [114,115]. Readers interested in various approaches for predicting pKa values of ionizable groups in ligand molecules are encouraged to read the excellent review by Alexov and co-workers [106] and the references therein. On the other hand, calculations of pKa values of ionizable groups in receptor proteins are typically performed as part of a general *in silico* characterization of a protein. Therefore, these pKa values are calculated using standard microscopic and macroscopic models (as discussed previously) or even empirical models [111] of electrostatic interactions in proteins and their proper evaluation is necessary in ligand-binding applications.

Future perspective

In this review we have drawn attention to the importance of accurate modeling of electrostatic effects in proteins and protein–ligand complexes. We have mainly focused on continuum models to predict the electrostatic potential in and around proteins. Despite the fact that continuum models introduce some drastic assumptions into the calculation of electrostatic interactions in proteins, they have enjoyed widespread quantitative success in treating the wide range of electrostatic-mediated processes, including enzyme catalysis, protein stability, ligand binding and protonation equilibria. The need to adjust the value of the protein dielectric constant for different applications is the main problem connected with using continuum models, since it is hard to know *a priori* which dielectric constant will work well for a given problem. However, we believe that continuum methods can be improved significantly by modeling specific effects explicitly. In particular, we argue that improvements will come from the proper treatment of internal water molecules, and from using NMR titration experiments and vibrational Stark effect (VSE) spectroscopy to provide us with 'real-world' measurements of the electric field in protein interiors.

As emphasized in previous sections, continuum models treat water by replacing it with the medium of a dielectric constant of 80. For this reason, continuum models are able to achieve a high level of efficiency by predicting the average solvation properties of water rather than averaging over the electrostatic interactions of thousands of explicit water molecules. However, sometimes water molecules are not 'average' and their effect is consequently reproduced inaccurately by a continuum model. This is the case for conserved water molecules in protein–ligand complexes that bridge interactions between ligand and protein [116], tightly bound water molecules in active sites and other cavities; in these cases it is therefore important to treat the water molecules explicitly. Warshel and co-workers proposed a thermodynamic cycle that accounts for a 'water penetration' effect in the binding process by mutating the ligand to water molecules in both the protein binding site and in solution (see Figure 15.8 in the referenced article by Kato *et al.* [27]). We believe that explicit treatment of water molecules near an area of interest (an active site or a binding site), as implemented by Warshel *et al.* in the PDLD model [117], can be essential in further improvements to continuum methods. The preliminary research in this area has focused on so-called 'hybrid models' (models that incorporate the details of solvent in the simulation as they are needed [118,119]). However, developing a criterion for choosing the particular water molecules for explicit treatment and the determination of their position around the protein molecule will present a formidable challenge.

A continuum model represents a valuable tool in many applications, but only when the protein dielectric constant is calibrated properly for the particular system using the relevant experimental data. NMR pH-dependent

titration curves induced by 'through-space' electric effects, measured on many nuclei in a protein, present a unique type of data that can provide us with the desirable resolution of electrostatic field strength in protein interiors. By predicting the effects of ionizable groups on measured pH-dependent NMR titration curves with Buckingham's formula [120] and comparing these with measured 'ghost titrations', we can extract the '3D maps' of ε_p with almost atomistic resolution [120–122]. Furthermore, establishing a correlation between local ε_p values and properties of their immediate vicinity (charge distribution, flexibility of the local environment, the distance to protein–solvent boundary, packing density of atoms, or even atom and amino acid types) can make this model generally. A similar method for determining the position-dependent value of dielectric constant in lipid bilayers, using MD simulations and Fröhlich–Kirkwood theory, have been proposed recently by Stern *et al.* [123] and Nymeyer *et al.* [124]. Their 'dielectric profile' of the lipid bilayer represents a good attempt in understanding the heterogeneous nature of macromolecules. Furthermore, we argue that macroscopic theories can be proved reliable enough only when the consistent 3D ε_p maps can be applied for different types of applications with similar success, for example identical atomic ε_p values should be used in different electrostatic energy calculations and show a reasonable agreement with experiments. Of course, the employment of 3D ε_p maps must be accompanied by the necessary adjustments in the force field parameters (atomic radii and atomic charges) and the characteristics of the protein–solvent boundary.

Another powerful method for probing electrostatic fields in protein interiors is Stark effect spectroscopy. The Stark effect has now been studied for nearly 100 years, and represents the spectral changes in the presence of electrostatic field. Stark shifts of suitable probes can be used to measure not only the changes in magnitude of electrostatic field, but also the changes in the direction of electrostatic field in the protein. Changes in the electrostatic field can be detected when residues are mutated, the protein changes conformation and when a ligand binds [125,126]. In a recent study by Webb and Boxer [127], mutation-induced changes in the electrostatic field of the active site of human aldose reductase (hALR2) were measured using VSE spectroscopy. Different mutations of the hALR2 residues were constructed and the changes in the local electrostatic field were observed using a nitrile probe attached to the hALR2 inhibitor. The experimental results were further compared with the classical PBE calculations and revealed a significant discrepancy between predicted and experimental data. Therefore, we argue that optimization of current electrostatic models in proteins using experimental data obtained from NMR and Stark spectroscopy simultaneously constitutes a possible route to achieve an improved understanding of protein electrostatics.

Despite the modest advances in accurate predictions of protein–ligand binding energies in recent years, there is no doubt that CADD will play a prominent role in pharmaceutical industry in the future. CADD is expected to be the main tool in discovering new, more effective drugs, but also in anticipating the mutations that will be developed by a resistant strain [128–130] and creating the relevant drugs to treat them. Electrostatic free energies play a prominent role in predictions of ligand-binding free energy, and the full potential of CADD will, therefore, only be realized once we are able to accurately calculate electrostatic energies and forces in and around proteins. Thus, the calculation of electrostatic interactions in proteins arguably presents one of the largest obstacles in improving the accuracy and usefulness of structure-based energy calculation algorithms and major improvements on multiple fronts in electrostatic modeling is expected in the near future.

Acknowledgements

The authors wish to thank Michael Johnston and John Bradley for insightful discussions.

Financial & competing interests disclosure

This work was supported by The Irish Research Council for Science Engineering and Technology (IRCSET) Graduate Research Education Programme (GREP) scholarship, Science Foundation Ireland President of Ireland Young Researcher award (04/YI1/M537) and Science Foundation Ireland RFP award (08/RFP/BIC1140) to Jens Erik Nielsen. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- Permanent dipoles, induced dipoles and the ionic strength of a solution modulate electrostatic interactions inside and outside proteins. Permanent protein dipoles are described explicitly in continuum models of electrostatic interactions, whereas induced protein dipoles and water molecules are represented by the protein and solvent dielectric constants. Mobile ions are treated statistically using the Boltzmann distribution.
- Convergence and low computational efficiency are the main shortcomings of rigorous all-atom models of electrostatic interactions. Simplified microscopic models developed by Warshel and co-workers treat water molecules close to the active/binding site by Langevin dipoles and bulk water as a continuum. The protein dipoles langevin dipoles and protein dipoles langevin dipoles/S models perform better than microscopic models, but are still slower than macroscopic models.
- Macroscopic continuum models of electrostatic interactions (the Poisson–Boltzmann equation and generalized Born-based models) have been remarkably successful in explaining a wide range of phenomena in the last 20 years. However, the accuracy of these models is highly dependent on the value used for the protein dielectric constant.
- Molecular dynamics-based macroscopic models represent protein conformational flexibility explicitly by solving Newton's equations for motion of the solute. Molecular mechanics (MM)/PBSA, MM/GBSA, linear response approximation and linear interaction energy-based models, along with constant pH MD simulations have gained popularity for applications in ligand binding and studies of protein conformational changes.
- Electrostatic interactions between charged and polar groups in proteins and ligands play a major role in protein–ligand binding specificity, the pH dependence of the binding constant and in determining the rate of ligand binding. Therefore, an accurate and fast evaluation of electrostatic free energies in ligand/drug binding is likely to become a key factor in future computer-aided drug design studies.
- Validation of current electrostatic models against multiple types of experimental data (e.g., NMR and stark spectroscopy measurements) provides a novel route for constructing fully consistent models of electrostatic interactions in proteins.

Bibliography

Papers of special note have been highlighted as:
▪ of interest
▪▪ of considerable interest

1 Linderstrøm-Lang K. On the ionization of proteins. *C.r. Trav. Lab. Carlsberg* 15, 1–29 (1924).

2 Lippow SM, Tidor B. Progress in computational protein design. *Curr. Opin. Biotechnol.* 18, 305–311 (2007).

3 Honig B. Protein structure space is much more than the sum of its folds. *Nat. Struc. Mol. Biol.* 14, 458 (2007).

4 Warshel A, Sharma PK, Kato M, Parson WW. Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta* 1764(11), 1647–1676 (2006).

▪ **Focuses on microscopic models of electrostatic interactions in proteins and their application in studying pKa values, redox potentials, ion and proton channels, enzyme catalysis, ligand binding and protein stability.**

5 Kundrotas PJ, Alexov E. Electrostatic properties of protein–protein complexes. *Biophysical J.* 91(5), 1724–1736 (2006).

6 Nielsen JE. Analyzing enzymatic pH-activity profiles and protein titration curves using structure-based pKa calculations and titration curve fitting. *Methods Enzymol.* 454, 233–258 (2009).

7 Nielsen JE. Analyzing protein NMR pH-titration curves. *Ann. Reports Computational Chem.* 4, 89 (2008).

8 Nielsen JE. Rational (re)engineering of enzymes. *Enzyme Functionality* 113, 105–120 (2004).

9 Alexov EG. Protein–protein interactions. *Curr. Pharm. Biotechnol.* 9(2), 55–56 (2008).

10 Roca M, Messer B, Warshel A. Electrostatic contributions to protein stability and folding energy. *FEBS Lett.* 581(10), 2065–2071 (2007).

11 Vizcarra CL, Mayo SL. Electrostatics in computational protein design. *Curr. Opin. Chem. Biol.* 9, 622–626 (2005).

12 Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? *Structure Function Genetics* 44, 400–417 (2001).

13 Baker NA, Sept D, Holst MJ, McCammon JA. The adaptive multilevel finite element solution of the Poisson–Boltzmann equation on massively parallel computers. *IBM J. Res. Develop.* 45(3/4), 427–438 (2001).

14 Baker N. Poisson-Boltzmann methods for biomolecular electrostatics. *Methods Enzymol.* 383, 94–118 (2004).

15 Rocchia W, Alexov E, Honig B. Extending the applicability of the nonlinear Poisson–Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Phys. Chem. B* 105(28), 6507–6514 (2001).

16 Born VM. Volumen und hydratationswärme der ionen. *Zeitschrift für Physik A Hadrons and Nuclei* 1(1), 45–48 (1920).

17 Czodrowski P, Dramburg I, Sotriffer CA, Klebe G. Development, validation, and application of adapted PEOE charges to estimate pKa values of functional groups in protein–ligand complexes. *Proteins: Structure, Function, Bioinformatics* 65(2), 424–437 (2006).

18 Maple JR, Cao Y, Damm W *et al.* A polarizable force field and continuum solvation methodology for modeling of protein–ligand interactions. *J. Chem. Theory Comput.* 1, 694–715 (2005).

19 Warshel A, Kato M, Pisliakov AV. Polarizable force fields. History, test cases, and prospects. *J. Chem. Theory Comput.* 3(6), 2034–2045 (2007).

20 Patel S, Mackerell AD, Brooks III CL. CHARMM fluctuating charge force field for proteins. II Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* 25(12), 1504–1514 (2004).

21 Patel S, Brooks III CL. CHARMM fluctuating charge force field for proteins. I parameterization and application to bulk organic liquid simulations. *J. Computational Chem.* 25(1), 1–16 (2004).

22 Uematsu M, Franck E. Static dielectric constant of water and steam. *J. Phys. Chem. Ref. Data* 9(4), 1291–1306 (1980).

23 Cerutti DS, Baker NA, McCammon JA. Solvent reaction field potential inside an uncharged globular protein: a bridge between implicit and explicit solvent models? *J. Chem. Phys.* 127(15), 155101(2007).

24 LeBard DN, Matyushov DV. Redox entropy of plastocyanin. Developing a microscopic view of mesoscopic polar solvation. *J. Chem. Phys.* 128(15), 155106 (2008).

▪ **This article describes the Poisson–Boltzmann equation (PBE) model, which is able to account for protein induced dipoles using atomic multipole optimized energetics for biomolecular applications (AMOEBA) polarizable force field.**

25 Li Z, Lazaridis T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* 9(5), 573–581 (2007).

26 Li Z, Lazaridis T. Thermodynamics of buried water clusters at a protein–ligand binding interface. *J. Phys. Chem. B* 110(3), 1464–1475 (2005).

27 Kato M, Braun-Sand S, Warshel A. Challenges and progresses in calculations of binding free energies. What does it take to quantify electrostatic contributions to protein–ligand interactions? In: *Computational and Structural Approaches to Drug Discovery.* Stroud R, Finer-Moore J (Eds). RSC Publishing, UK, 268–292 (2008).

28 Schnieders MJ, Baker NA, Ren P, Ponder JW. Polarizable atomic multipole solutes in a Poisson–Boltzmann continuum. *J. Chem. Phys.* 126(12), 124114 (2007).

29 Sham YY, Chu ZT, Warshel A. Consistent calculations of pKa's of ionizable residues in proteins. Semi-microscopic and microscopic approaches. *J. Phys. Chem. B* 101(22), 4458–4472 (1997).

30 Lee FS, Chu ZT, Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. *J. Computational Chem.* 14(2), 161–185 (1993).

31 Ishikita H, Warshel A. Predicting drug-resistant mutations of HIV protease 13. *Angewandte Chemie International Edition* 47(4), 697–700 (2008).

▪▪ **Gives an overview of the PBE and explains numerical methods to solve it. Furthermore, the applications of the PBE model in predicting solvation free energy, ligand-binding free energy and pKa values are given.**

32 Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM. Electrostatic basis for enzyme catalysis. *Chem. Rev.* 106(8), 3210–3235 (2006).

33 Warshel A, Sharma PK, Chu ZT, Aqvist J. Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry* 46(6), 1466–1476 (2007).

34 Baker N. Biomolecular applications of Poisson–Boltzmann methods. *Rev. Computational Chem.* 5, 377–392 (2005).

35 Tynan-Connolly BM, Nielsen JE. Redesigning protein pKa values. *Protein Sci.* 16(2), 239–249 (2007).

36 Baran KL, Chimenti MS, Schlessman JL, Fitch CA, Herbst KJ, Garcia-Moreno BE. Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. *J. Mol. Biol.* 379(5), 1045–1062 (2008).

37 Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solution – simulations with the University of Houston Brownian Dynamics *Comput. Phys. Commun.* 91, 57–95 (1995).

38 MacKerell AD, Brooks B, Brooks CLI, Nilsson L, Roux YW, Karplus M. CHARMM. The energy function and its parameterization with an overview of the program. *Encyclopedia of Comp. Chem.* 1, 271–277 (1998).

39 Nicholls A, Honig B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.* 12, 435–445 (1991).

40 Feig M, Brooks III CL. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Structural Biol.* 14(2), 217–224 (2004).

41 Wagoner J, Baker N. Solvation forces on biomolecular structures. a comparison of explicit solvent and Poisson–Boltzmann models. *J. Comput. Chem.* 25(13), 1623–1629 (2004).

42 Chen J, Brooks III CL, Khandogin J. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Structural Biol.* 18(2), 140–148 (2008).

43 Wagoner JA, Baker NA. Assessing implicit models for nonpolar mean solvation forces. The importance of dispersion and volume terms. *Proc. Natl Acad. Sci. USA* 103(22), 8331–8336 (2006).

44 Dzubiella J, Swanson JMJ, McCammon JA. Coupling hydrophobicity, dispersion, and electrostatics in continuum solvent models. *Physical Review Letters* 96(8), 087802–087804 (2006).

45 Gallicchio E, Levy RM. AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* 25(4), 479–499 (2004).

46 Gilson MK, Zhou H-X. Calculation of protein–ligand binding affinities. *Ann. Rev. Biophysics Biomol. Struct.* 36, 21–42 (2007).

47 Liu H-Y, Zou X. Electrostatics of ligand binding. Parametrization of the generalized born model and comparison with the Poisson–Boltzmann approach. *J. Phys. Chem. B* 110, 9304–9313 (2006).

48 Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks III CL. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Computational Chem.* 25(2), 265–284 (2004).

49 Im W, Lee MS, Brooks III CL. Generalized born model with a simple smoothing function. *J. Computational Chem.* 24(14), 1691–1702 (2003).

50 Feig M, Im W, Brooks III CL. Implicit solvation based on generalized Born theory in different dielectric environments. *J. Chem. Phys.* 120(2), 903–911 (2004).

51 Chen J, Im W, Brooks III CL. Balancing solvation and intramolecular interactions. Toward a consistent generalized born force field. *J. Am. Chem. Soc.* 128(11), 3728–3736 (2006).

52 Jorgensen WL, Ulmschneider JP, Tirado-Rives J. Free energies of hydration from a generalized born model and an all-atom force field. *J. Phys. Chem. B* 108(41), 16264–16270 (2004).

53 Liu H-Y, Kuntz ID, Zou X. Pairwise GB/SA scoring function for structure-based drug design. *J. Phys. Chem. B* 108(17), 5453–5462 (2004).

54 Cecchini M, Kolb P, Majeux N, Caflisch A. Automated docking of highly flexible ligands by genetic algorithms. A critical assessment. *J. Computat. Chem.* 25, 412–422 (2004).

55 Cho AE, Wendel JA, Vaidehi N *et al.* The MPSim-dock hierarchical docking algorithm. application to the eight trypsin inhibitor cocrystals. *J. Comput. Chem.* 26, 48–71 (2005).

56 Friesner RA, Banks JL, Murphy RB *et al.* Glide. A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* 47(7), 1739–1749 (2004).

57 Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pKas in proteins. *Biophysical J.* 83(4), 1731–1748 (2002).

58 Nielsen JE, Vriend G. Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pKa calculations. *Proteins. Structure, Function, and Genetics* 43(4), 403–412 (2001).

59 David L, Luo R, Gilson MK. Comparison of generalized born and poisson models. Energetics and dynamics of HIV protease. *J. Computational Chem.* 21(4), 295–309 (2000).

60 Fan H, Mark AE, Zhu J, Honig B. Comparative study of generalized Born models. Protein dynamics. *PNAS* 102, 6760–6764 (2005).

61 Dominy BN, Minoux H, Brooks CL. An electrostatic basis for the stability of thermophilic proteins. *Proteins*: *Structure, Function, Bioinformatics* 57(1), 128–141 (2004).

62 Mongan J, Case DA, McCammon JA. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* 25(16), 2038–2048 (2004).

63 Machuqueiro M, Baptista AM. Acidic range titration of HEWL using a constant-pH molecular dynamics method. *Proteins*: *Structure, Function, Bioinformatics* 72(1), 289–298 (2008).

64 Baptista AM, Teixeira VH, Soares CM. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* 117(9), 4184–4200 (2002).

65 Dlugosz M, Antosiewicz JM. Constant-pH molecular dynamics simulations: a test case of succinic acid *Chem. Phys.* 302(1–3), 161–170 (2004).

66 Machuqueiro M, Baptista AM. Constant-pH molecular dynamics with ionic strength effects. Protonation-conformation coupling in decalysine. *J. Phys. Chem. B* 110(6), 2927–2933 (2006).

67 Długosz M, Antosiewicz JM. Constant-pH molecular dynamics study of protonation – structure relationship in a heptapeptide derived from ovomucoid third domain. *Phys. Rev. E* 69, 2 (2004).

68 Sham YY, Chu ZT, Tao H, Warshel A. Examining methods for calculations of binding free energies. LRA, LIE, PDLD-LRA, and PDLD/S-LRA calculations of ligands binding to an HIV protease. *Proteins: Structure, Function, Bioinformatics* 39(4), 393 - 407 (2000).

69 Warshel A, Sharma P, Chu Z, Aqvist J. Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis. Phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry* 46(6), 1466–1476 (2007).

70 Nervall M, Hanspers P, Carlsson J, Boukharta L, Åqvist J. Predicting binding modes from free energy calculations. *J. Med. Chem.* 51(9), 2657–2667 (2008).

71 Andéra M, Luzhkova VB, Åqvist J. Ligand binding to the voltage-gated kv1.5 potassium channel in the open state – docking and computer simulations of a homology model. *Biophys. J.* 94(3), 820–831 (2008).

72 Carlsson J, Boukharta L, Åqvist J. Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. *J. Med. Chem.* 51(9), 2648–2656 (2008).

73 Carlsson J, Andér M, Nervall M, Åqvist J. Continuum solvation models in the linear interaction energy method. *J. Phys. Chem. B* 110(24), 12034–12041 (2006).

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

75 Lyne PD, Lamb ML, Saeh JC. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* 49(16), 4805–4808 (2006).

76 Lee MS, Olson MA. Calculation of absolute protein–ligand binding affinity using path and endpoint approaches. *Biophys. J.* 90(3), 864–877 (2006).

77 Swanson JMJ, Henchman RH, McCammon JA. 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).

78 Zoete V, Michielin O. Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein–protein association using MM-GBSA. Application to the TCR-p-MHC complex. *Proteins: Structure, Function, Bioinformatics* 67(4), 1026–1047 (2007).

79 Rizzo RC, Toba S, Kuntz ID. A molecular basis for the selectivity of thiadiazole urea inhibitors with stromelysin-1 and gelatinase-A from generalized Born molecular dynamics simulations. *J. Med. Chem.* 47, 3065–3074 (2004).

80 Strockbine B, Rizzo RC. Binding of antifusion peptides with HIVgp41 from molecular dynamics simulations. Quantitative correlation with experiment. *Proteins: Structure, Function, Bioinformatics* 67(3), 630–642 (2007).

81 Brock K, Talley K, Coley K, Kundrotas P, Alexov E. Optimization of electrostatic interactions in protein–protein complexes. *Biophys. J.* 83, 3340–3352 (2007).

▪▪ **Builds on previous work by Zhou and Schreiber and describes the role of electrostatic interactions in the association rate of binding partners. The authors achieve a good agreement between experimental and calculated association rate data using the PBE and discuss the sensitivity of the model on the dielectric boundary definition.**

82 Tjong H, Zhou H-X. Accurate calculations of binding, folding, and transfer free energies by a scaled generalized born method. *J. Chem. Theory Comput.* 4, 1733–1744 (2008).

83 Bertonati C, Honig B, Alexov E. Poisson–Boltzmann calculations of nonspecific salt effects on protein–protein binding free energies. *Biophys. J.* 92, 1891–1899 (2007).

84 Schmidt M, Lopes A, Amara N, Bathelt C, Simonson T. Testing the Coulomb/accessible surface area solvent model for protein stability, ligand binding, and protein design. *BMC Bioinformatics* 9, 148 (2008).

85 Allison SA, Northrup SH, McCammon JA. Simulation of biomolecular diffusion and complex formation. *Biophys. J.* 49(1), 167–175 (1986).

86 Northrup SH, Boles JO, Reynolds JCL. Electrostatic effects in the Brownian dynamics of association and orientation of heme proteins. *J. Phys. Chem.* 91(23), 5991–5998 (1987).

87 Schreiber G, Haran G, Zhou HX. Fundamental aspects of protein–protein association kinetics. *Chem. Rev.* 109(3), 839–860 (2009).

88 Shaul Y, Schreiber G. Exploring the charge space of protein–protein association: a proteomic study. *Proteins: Structure, Function, Bioinformatics* 60, 341–352 (2005).

89 Alsallaq R, Zhou H-X. Electrostatic rate enhancement and transient complex of protein–protein association. *Proteins*: *Structure, Function, Bioinformatics* 71, 320–335 (2008).

90 Dong F, Zhou H-X. Electrostatic contribution to the binding stability of protein–protein complexes. *Proteins*: *Structure, Function, Bioinformatics* 65, 87–102 (2006).

91 Uter NT, Gruic-Sovulj I, Perona JJ. Amino acid-dependent transfer RNA affinity in a class I aminoacyl-tRNA synthetase. *J. Biol. Chem.* 280(25), 23966–23977 (2005).

92 Jaitin DA, Roisman LC, Jaks E *et al.* Inquiring into the differential action of interferons (IFNs): an IFN-α2 mutant with enhanced affinity to ifnar1 is functionally similar to IFN-β. *Mol. Cell. Biol.* 26(5), 1888–1897 (2006).

93 Kiel C, Serrano L, Herrmann C. A detailed thermodynamic analysis of ras/effector complex interfaces. *J. Mol. Biol.* 340(5), 1039–1058 (2004).

94 Johnson RJ, McCoy JG, Bingman CA, Phillips GN, Raines RT. Inhibition of human pancreatic ribonuclease by the human ribonuclease inhibitor protein. *J. Mol. Biol.* 368, 434–449 (2007).

95 Korennykh AV, Piccirilli JA, Correll CC. The electrostatic character of the ribosomal surface enables extraordinarily rapid target location by ribotoxins. *Nat. Struct. Mol. Biol.* 13(5), 436–443 (2006).

96 Stewart RC, Van Bruggen R. Association and dissociation kinetics for chey interacting with the P2 domain of CheA. *J. Mol. Biol.* 336(1), 287–301 (2004).

97 Li P, Martins IRS, Amarasinghe GK, Rosen MK. Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nat. Struct. Mol. Biol* 15(6), 613–618 (2008).

98 Trosset J-Y, Dalvit C, Knapp S *et al.* Inhibition of protein–protein interactions. The discovery of druglike β-catenin inhibitors by combining virtual and biophysical screening. *Proteins: Structure, Function, and Bioinformatics* 64(1), 60–67 (2006).

99 Czodrowski P, Sotriffer CA, Klebe G. Protonation changes upon ligand binding to trypsin and thrombin. Structural interpretation based on pKa calculations and ITC experiments. *J. Mol. Biol.* 367(5), 1347–1356 (2007).

100 Djurdjevic-Pahl A, Hewage C, Malthouse JPG. Ionisations within a subtilisin-glyoxal inhibitor complex. *Biochimica et Biophysica Acta* 1749(1), 33–41 (2005).

101 Blundell CD, Mahoney DJ, Cordell MR *et al.* Determining the molecular basis for the pH-dependent interaction between the link module of human TSG-6 and hyaluronan. *J. Biol. Chem.* 282(17), 12976–12988 (2007).

102 Lu Y, Yang C-Y, Wang S. Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. *J. Am. Chem. Soc.* 128, 11830–11839 (2006).

103 Mason AC, Jensen JH. Protein–protein binding is often associated with changes in protonation state. *Proteins* 71, 81–91 (2008).

104 Archontis G, Simonson T. Proton binding to proteins. A free-energy component analysis using a dielectric continuum model. *Biophys. J.* 88, 3888–3904 (2005).

105 Alexov EG. Calculating proton uptake/release and binding free energy taking into account ionization and conformation changes induced by protein–inhibitor association. Application to plasmepsin, cathepsin D and endothiapepsin–pepstatin complexes. *Proteins: Structure, Function, Bioinformatics* 56(3), 572–584 (2004).

106 Mitra R, Shyam R, Mitra I, Miteva M, Alexov E. Calculating the protonation states of proteins and small molecules. Implications to ligand–receptor interactions. *Current Computer-Aided Drug Design* 4(3), 169–179 (2008).

107 Polgar T, Keseru GM. Virtual screening for β-secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.* 48(11), 3749–3755 (2005).

108 Polgar T, Magyar C, Simon I, Keseru GM. Impact of ligand protonation on virtual screening against β-secretase (BACE1). *J. Chem. Information Modeling* 47(6), 2366–2373 (2007).

109 Jensen JH. Calculating pH and salt dependence of protein–protein binding. *Curr. Pharm. Biotechnol.* 9(2), 96–102 (2008).

110 Dullweber F, Stubbs MT, Musil D, Stürzebecher J, Klebe G. Factorising ligand affinity. a combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J. Mol. Biol.* 313, 593–614 (2001).

111 Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*: *Structure, Function, Bioinformatics* 61(4), 704–721 (2005).

▪ **Predicts NMR pH-dependent titration curves of amide nuclei in β-lactoglobulin using Coulomb's law and Buckingham's formula. A good correlation between predicted and experimental chemical shift can be used in validations of PBE models using the NMR spectroscopy.**

112 Lu H, Chen X, Zhan C-G. First-principles calculation of pka for cocaine, nicotine, neurotransmitters, and anilines in aqueous solution. *J. Phys. Chem. B* 111(35), 10599–10605 (2007).

113 Bryantsev V, Diallo M, Goddard W. pKa calculations of aliphatic amines, diamines, and aminoamides via density functional theory with a Poisson–Boltzmann continuum solvent model. *J. Phys. Chem. A* 111, 4422–4430 (2007).

114 Milletti F, Storchi L, Sforna G, Cruciani G. New and original pKa prediction method using grid molecular interaction fields. *J. Chem. Information Modeling* 47(6), 2172–2181 (2007).

115 Shelley J, Cholleti A, Frye L, Greenwood J, Timlin M, Uchimaya M. Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. *J. Computer-Aided Mol. Design* 21(12), 681–691 (2007).

▪▪ **Suggests the use of vibrational Stark effect spectroscopy with the nitrile probe as a general tool for measuring electrostatic fields in proteins. Measured experimental electric field values revealed substantial disagreement with the values predicted using the PBE model.**

116 Gouda H, Kuntz ID, Case DA, Kollman PA. Free energy calculations for theophylline binding to an RNA aptamer. Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers* 68, 16–34 (2003).

117 Warshel A. Calculations of electrostatic interactions in biological systems and in solutions. *Q Rev Biophys.* 17(3), 283–422 (1984).

118 Lee MS, Salsbury FRJ, Olson MA. An efficient hybrid explicit/implicit solvent method for biomolecular simulations. *J. Comput. Chem.* 25(16), 1967–1978 (2004).

119 Yu Z, Jacobson MP, Josovitz J, Rapp CS, Friesner RA. First-shell solvation of ion pairs. correction of systematic errors in implicit solvent models. *J. Phys. Chem. B* 108(21), 6643–6654 (2004).

120 Hass MA, Jensen MR, Led JJ. Probing electric fields in proteins in solution by NMR spectroscopy. *Proteins* 72(1), 333–343 (2008).

121 Farrell D, Sa-Miranda E, Webb H, Georgi N, Nielsen JE. Storage and analysis of protein NMR titration curves. *Proteins: Structure, Function, Bioinformatics.* 78 (4) 843–957 (2010).

122 Kukić P, Farrell D, Søndergaard C *et al.* Improving the analysis of NMR spectra tracking pH-induced conformational changes. Removing artefacts of the electric field on the NMR chemical shift. *Proteins: Structure, Function, Bioinformatics.* 78(4), 971–984 (2010).

▪ **Predicts NMR pH-dependent titration curves of amide nuclei in β-lactoglobulin using Coulomb's law and Buckingham's formula. A good correlation between predicted and experimental chemical shift can be used in validations of PBE models using the NMR spectroscopy.**

123 Stern HA, Feller SE. Calculation of the dielectric permittivity profile for a nonuniform system. Application to a lipid bilayer simulation. *J. Chem. Phys.* 118(7), 3401–3412 (2003).

124 Nymeyer H, Zhou H-X. A method to determine dielectric constants in nonhomogeneous systems. Application to biological membranes. *Biophys. J.* 94, 1185–1193 (2008).

125 Suydam IT, Snow CD, Pande VS, Boxer SG. Electric fields at the active site of an enzyme. Direct comparison of experiment with theory. *Science* 313(5784), 200–204 (2006).

126 Boxer SG. Stark realities. *J. Phys. Chem. B* 113, 2972–2983 (2009).

127 Webb LJ, Boxer SG. Electrostatic fields near the active site of human aldose reductase. 1. New inhibitors and vibrational stark effect measurements. *Biochemistry* 47(6), 1588–1598 (2008).

128 Ohtaka H, Freire E. Adaptive inhibitors of the HIV-1 protease. *Progress in Biophysics and Molecular Biology* 88(2), 193–208 (2005).

129 Ode H, Neya S, Hata M, Sugiura W, Hoshino T. Computational simulations of HIV-1 proteases multi-drug resistance due to nonactive site mutation L90M. *J. Am. Chem. Soc.* 128(24), 7887–7895 (2006).

130 Hou T, Yu R. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors. Mechanism for binding and drug resistance. *J. Med. Chem.* 50(6), 1177–1188 (2007).

■ Websites

201 Zap electrostatics software www.edusoft-lc.com/zap/

202 APBS software documentation http.//apbs.wustl.edu