Theor Chem Acc (2012) 131:1252
DOI 10.1007/s00214-012-1252-5

FEATURE ARTICLE

Designing electrostatic interactions in biological systems via charge optimization or combinatorial approaches: insights and challenges with a continuum electrostatic framework

Mala L. Radhakrishnan

Received: 27 March 2012 / Accepted: 23 June 2012 / Published online: 22 July 2012

Abstract Electrostatic interactions between biological molecules are crucially influenced by their aqueous environment, with efficient and accurate models of solvent effects required for robust molecular design strategies. Continuum electrostatic models provide a reasonable balance between computational efficiency and accurate system representation. In this article, I review two specific molecular design strategies, charge optimization and combinatorial design, paying particular attention to how the continuum framework (also briefly described herein) successfully enables both theoretical insights and molecular designs and presents a challenge in design applications due to what I call "the isostericity constraint." Efforts to work around the isostericity constraint and other challenges are discussed. Additionally, particular emphasis is placed on using such models in the rational design of particularly tight, specific, or promiscuous interactions, in keeping with the increased sophistication of current molecular design applications.

Keywords Continuum electrostatics · Charge optimization · Combinatorial design · Specificity · Promiscuity · Implicit solvation

1 Introduction

Electrostatic interactions play a crucial role in mediating biomolecular recognition. Unlike short-range, shape-dependent dispersion forces or nonspecific "hydrophobic" interactions, electrostatic interactions have the flexibility to be both nonspecific and highly specific, due to their generating both long-range and short-range effects. Indeed, examples of electrostatic-mediated biomolecular interactions are far too numerous to describe here, but typical natural examples include electrostatic features that contribute to increased affinity [1–3], specificity [4–6], and promiscuity [7, 8]. In design applications, electrostatics are an effective way to influence molecular recognition, as specific patterns of hydrophobic and charged groups can be designed to exploit similarities and differences among desired and undesired potential binding partners [9]. Moreover, a single residue or functional group modification can greatly alter affinity [10–12], specificity [13, 14], or pH-dependent behavior [15]. The long-range nature of electrostatics is important for both thermodynamic affinity enhancements [16] and kinetic capturing, orienting, association, and (arguably) steering of potential partners [17–22].

Accordingly, great efforts have been made to understand electrostatic interactions in various biological contexts. In the case of protein systems, studies reveal that although the specific role of electrostatics in binding can vary among complexes [23], potentially playing a larger role for smaller interfaces [23, 24], polar residues tend to be more conserved in interfacial "hot spots" [25], suggesting their crucial roles in mediating recognition. Indeed, Coulombic electrostatic interactions are "optimized" in natural proteins and their complexes, in that they are significantly more favorable than those in random "decoys" [26], with the favorability likely stemming from orientation-dependent electrostatic features rather than overall monopoles (the total charges on species) [27]. Optimization of Coulombic interactions explains why binding between known partners is nearly always weakened in electrolytic solution, in which such interactions are screened (i.e., dampened) by the counterions in solution [28].

M. L. Radhakrishnan (✉)
Department of Chemistry, Wellesley College, 106 Central Street, Wellesley, MA 02481, USA
e-mail: mradhakr@wellesley.edu

Ironically, despite the importance of electrostatics in mediating recognition, they often contribute *unfavorably* overall to molecular folding and binding processes [23, 26, 29]. This idea may seem strange, as the simple adage that "opposites attract" can easily be satisfied through positively and negatively charged groups within the system; however, the reason for this phenomenon is that biomolecular interactions often occur in the aqueous phase. A biomolecule pays a large opportunity cost upon folding or binding, as it must give up favorable interactions with polar water molecules, becoming partially *desolvated*, to interact with anything else.

Unfortunately, solvation, which confounds the otherwise intuitive nature of electrostatics, is also difficult to quantitatively model. As an example, significant efforts to model the energetic contributions of salt bridges (i.e., the interaction between oppositely charged amino acid side chains) within proteins and their complexes show that they may contribute either favorably or unfavorably toward folding or binding [30–36], but throughout the years, it remains unclear how robust certain results are to the models used [37–41]. Even qualitative conclusions can depend crucially on the balance between desolvation and interaction energetics. The high interest in the accurate modeling of salt bridges over multiple decades demonstrates both the importance of biomolecular electrostatic interactions and the challenges in their quantitative modeling.

In this paper, I focus on one class of widely used electrostatic models—the continuum models—and describe their use in biomolecular and small molecule design. I review two specific design techniques that incorporate continuum models—electrostatic charge optimization and combinatorial molecular design—with the emphasis being on methodology rather than on applications. Section 2 provides a very brief overview of continuum electrostatic theory. Section 3 describes electrostatic charge optimization theory, its successful applications toward both analysis and design, and open problems within the field. Finally, Sect. 4 describes the incorporation of continuum electrostatic models into a combinatorial design framework and the challenges therein. As a whole, my goal is to introduce the reader to both the strengths and challenges of using a continuum model as part of a molecular design strategy.

2 Background: continuum electrostatic models

2.1 Overview

The modeling of electrostatic interactions in biological macromolecular systems has inherent challenges. First, the "real" electrostatic properties of a molecule depend on the constant adjustment of and uncertainty in electron density within all species and are best modeled through quantum mechanical approaches. Such computation is currently prohibitively expensive, especially when the dynamics of macromolecular processes occurring in solution are considered. Many models therefore approximate true charge distributions with fixed point charges at atom centers. The continuum model, which motivates this work and is described in detail below, implicitly accounts for solute charge polarization among other things, through assuming that the fixed point charges lie within a polarizable medium within the biological molecules. More explicit, "microscopic" models allow for charge polarization through modeling polarizable dipoles, drude oscillators, fluctuating charges, self-consistent multipoles, and other approaches, and are addressed within multiple reviews and recent work [42–50]. Despite much progress, the accurate yet efficient representation of biomolecular charge distributions remains a challenge.

The other, related major challenge in modeling biomolecular electrostatics is accounting for electrostatic interactions with the solvent. Two major approaches are generally used. The first approach, explicit solvent modeling, represents individual water molecules *explicitly*. The solvent polarization is captured through rearrangements and reorientations of individual water molecules as they interact with the biomolecular solutes via time-dependent molecular dynamics simulations. Assuming a point charge model of charge distributions, Coulomb's Law can be used to calculate the potential energy U resulting from each pair of charges q_1 and q_2, separated by distance r:

$$U = \frac{q_1 q_2}{4\pi\varepsilon_0 r}, \quad (\varepsilon_0 \text{ is the permittivity of free space constant, } 8.854 \times 10^{-12}\text{C}/(\text{Vm})) \tag{1}$$

Summing over all pairs of charges in the system can be expensive; distance cutoffs are sometimes used to increase the efficiency. Due to the long-range nature of electrostatics, such cutoffs may lead to inaccurate results [51, 52], though improved cutoff schemes, such as those that gradually drive the potential energy to zero as the cutoff distance is approached rather than abruptly cutting it off, can achieve better accuracy [53]. Many methods exist to treat long-range electrostatics within explicit solvent models; the most common assume periodic boundary conditions for the simulated system and employ numerical strategies such as Ewald summation, fast multipole methods, or others to facilitate computation [54–57]. Explicit solvent models are widely used for quantifying thermodynamic properties of biological systems and have been reviewed throughout the years [45, 58, 59]. They have the added advantage of also accounting for nonpolar solvent effects, such as dispersion interactions, on an atomistic level. Nevertheless, they are generally too computationally expensive for molecular design purposes, as in such cases, one wishes to compare

thermodynamic properties of a huge number of potential designs. With each potential design requiring extensive simulation in which the explicit solvent degrees of freedom are adequately sampled, such an application is currently infeasible in general.

To enable efficient modeling of solvent for use in design, therefore, alternative models are needed. These models are *implicit* models, which account for the effects of solvent polarization and reorganization without explicitly modeling actual solvent molecules. Such models allow for more rapid evaluation of electrostatic properties and have been shown to perform similarly to explicit models in identifying minimum-energy conformations of biological molecules [60] and quantifying solvation energies [61]. Nevertheless, specific implicit models have noted weaknesses, including energetic biases toward secondary structural elements [62], inaccurate conformational sampling due to erroneous quantification of salt bridge effects [63–66], and the inability to account for effects resulting from discrete water molecules [67, 68]. This paper focuses on only the electrostatic effects of solvent, although it is important to note that implicit electrostatic solvent models are generally used in conjunction with other implicit models to account for crucial nonpolar solvent effects, such as cavitation and dispersion, to obtain "total" solvation energies. Such nonpolar models have also been an active area of research over the past decade [69–73].

One major class of implicit electrostatic models, "continuum electrostatic" models, treats the solvent as a polarizable, macroscopic continuum (i.e., a dielectric medium). Such models motivate the rest of this work. Continuum models have seen *enormous* usage in the treatment of biomolecular systems over multiple decades, with far too many success stories to describe here. As my goal is to discuss continuum electrostatic models within a molecular design framework, I will only briefly summarize the theory behind continuum models and provide the reader here with representative (but not all) reviews over the previous 15 years for more detail on theory and applications [59, 74–81]. The vast number of reviews speaks to their great utility in representing biomolecular systems. They provide a balance between accuracy and computational expense, making them appropriate for use in molecular design applications.

2.2 Theoretical framework and the Poisson–Boltzmann equation

In continuum electrostatic models, the solvent is represented as a high-dielectric, polarizable continuum and the biological solutes are represented as lower-dielectric cavities with embedded charge distributions (Fig. 1). In aqueous solvent, the solvent dielectric constant (which quantifies the degree of polarizability) is roughly 80, the macroscopic dielectric constant of water, but the "best" value of the lower dielectric constant of the biomolecule has been sought after for multiple decades [82, 83], with it likely being system dependent [84–86], given its inherent crudeness as a physical model [45]; values typically range from 2 to 40, depending on whether the dielectric constant is to capture only electronic polarization or more substantial spatial rearrangements or reorientation of charges. The dielectric solvent model assumes a linear response of the solvent to the field generated by the solute charge distribution; the "reaction field" generated by the solvent to interact with a given solute charge is proportional to the solute charge's magnitude. The linear response model is valid for small to moderate inducing electric fields for a re-organizing solvent such as water; once the field is strong enough, the response due to solvent reorganization saturates [87].

Given a charge distribution $\rho(\vec{r})$ and a spatially varying dielectric constant $\varepsilon(\vec{r})$ (i.e., a solvent region and a solute region), the electrostatic potential everywhere in space ($\phi(\vec{r})$) can be obtained by solving the Poisson equation:

$$-\nabla \cdot (\varepsilon(\vec{r})\nabla\phi(\vec{r})) = \frac{\rho(\vec{r})}{\varepsilon_0} \tag{2}$$

The presence of monovalent, mobile ions within solvent is often implicitly incorporated into the model, as they are present in biological systems and screen electrostatic interactions. Through Debye–Huckel theory [88], the Poisson equation can be extended to implicitly account for the reorganization of monovalent, mobile ions within the solvent, resulting in the (nonlinear) Poisson–Boltzmann equation (PBE):

$$-\nabla \cdot (\varepsilon(\vec{r})\nabla\phi(\vec{r})) = \frac{\rho(\vec{r})}{\varepsilon_0} - \frac{2q_s c_s}{\varepsilon_0}\sinh(\beta q_s \phi(\vec{r})) \tag{3}$$

where c_s is the bulk salt concentration (and is nonzero only in regions accessible to ions), q_s is the unit charge (1.6×10^{-19} C in SI units), and β is the reciprocal of the Boltzmann constant multiplied by the temperature. While powerful, the implicit, mean-field treatment of mobile ions described by the PBE is unable to capture their fluctuations

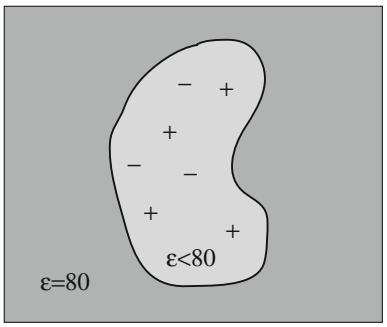

Fig. 1 Schematic showing a biological molecule (*light gray*) within solvent (*dark gray*) as modeled via the continuum framework

or correlations [81, 89] and cannot account for the finite volumes of each type of ion within solution [81, 90].

Assuming a small magnitude for potential (valid for small-magnitude charge distributions), the linearized approximation to the PBE is commonly used:

$$-\nabla \cdot (\varepsilon(\vec{r})\nabla\phi(\vec{r})) = \frac{\rho(\vec{r})}{\varepsilon_0} - \varepsilon(\vec{r})\kappa^2\phi(\vec{r}) \tag{4}$$

$$\kappa^2 = \frac{2\beta}{\varepsilon_0\varepsilon_w} c_s q_s^2 \tag{5}$$

Here, ε_w is the dielectric constant of water and κ is nonzero only in regions accessible by ions. Like the Poisson equation, the linearized Poisson–Boltzmann equation (LPBE) assumes linear response, although now, the reaction field is generated by both solvent polarization and mobile ion rearrangement. Linearity provides mathematical conveniences such as reciprocity (the potential at point i due to a source at point j equals the potential at j due to i) and superposition (the potential is a linear sum of those generated by individual sources) [91, 92].

The PBE and/or the linearized Poisson–Boltzmann equation can be solved numerically using well-established finite difference [93–95], boundary element [96–98], or finite element [99–101] techniques (reviewed elsewhere [102, 103]). Its solution provides the potential at desired locations. Specifically, when the linearized Poisson–Boltzmann equation (LPBE) is solved, linearity can be exploited such that one obtains, through superposition, the total electrostatic energy of the system by summing over the product of potentials and charges q_i at each location:

$$G = \sum_i \frac{1}{2} q_i \phi_i \tag{6}$$

The overall factor of ½ results from two different reasons [104]. First, it prevents double counting energetic contributions resulting from a charge feeling the potential generated by another charge (which, by reciprocity, is equal for a pair of interacting charges). Second, it accounts for the fact that the cost of generating a reaction field in response to a given charge is one-half of the resulting interaction energy with that charge [42]. Because this cost is in part an entropic contribution of reorienting solvent, the resulting energy in Eq. 6 is a free energy.

2.3 Generalized Born and other approaches

While less computationally demanding than explicit solvent simulation, the numerical solution of the LPBE is still computationally expensive, especially when high levels of quantitative accuracy are desired. Therefore, more efficient continuum models exist. The most common alternative is the generalized Born (GB) class of methods (reviewed in Ref. [105]). GB methods originate in the fact that the Poisson equation can be solved analytically for the simple case of a low-dielectric ($\varepsilon_{\text{solute}}$) sphere of radius a with a point charge q located at the center within a high-dielectric (ε_w) medium, yielding a closed-form expression for its solvation energy, modeled as the free energy change upon transferring it from low- to high-dielectric media:

$$\Delta G_{\text{solv}} = -\frac{kq^2}{2a}\left(\frac{1}{\varepsilon_{\text{solute}}} - \frac{1}{\varepsilon_w}\right) \tag{7}$$

where k is a constant to achieve appropriate units. Likewise, the difference in interaction energy of two isolated spheres i and j upon solvation can be analytically calculated in the limit where their radii are much smaller than their separation distance, r_{ij}:

$$\Delta G_{\text{inter}} = \frac{kq_iq_j}{r_{ij}}\left(\frac{1}{\varepsilon_w} - \frac{1}{\varepsilon_{\text{solute}}}\right) \tag{8}$$

Of course, biological molecules are nonspherical and have many, closely interacting charges, so these equations are inapplicable as written, although GB methods approximate solvation energies using analytical expressions based on the above equations. Each atom has an "effective Born radius," equal to the radius that a hypothetical spherical cavity containing only that atomic charge would have in order to generate the same self-polarization energy (i.e., the solvation energy due to that one charge interacting with its own solvent reaction field). The effective Born radius depends on the shape of the entire molecule and can be "rigorously" determined by charging up only one atom center in a biomolecule and calculating the solvent reaction potential via numerical solution of the Poisson equation, but this eliminates increased efficiency. Rather, effective Born radii are estimated by other means. Because the accuracy of the GB method depends crucially on the effective Born radii [106], rapid but accurate estimates are necessary and have been a huge area of research spanning multiple decades [105, 107–116].

Even if one were to obtain "exact" Born radii to evaluate self-polarization energies upon solvation via Eq. 7, one still needs to evaluate the change in interaction energies between each pair of charges upon solvation. By "interpolating" between Eqs. 7 and 8 above, Still et al. [117] devised the following, surprisingly accurate formula for calculating solvent reaction field energies, double-summing over all pairs of charges $i \geq j$:

$$\Delta G_{\text{solv}} = \sum_{i \geq j} \frac{kq_iq_j}{\sqrt{r_{ij}^2 + a_ia_j \exp\frac{-r_{ij}^2}{4a_ia_j}}}\left(\frac{1}{\varepsilon_w} - \frac{1}{\varepsilon_{\text{solute}}}\right) \tag{9}$$

Equation 9 reduces to Eq. 7 when $r = 0$ and to Eq. 8 when $r \gg a_i$ and a_j. Active efforts to refine this

interpolation exist [106, 118]. The GB-based methods are widely used as more efficient alternatives to solving the LPBE, and certain implementations have closely reproduced LPBE results [119]. However, common implementations may systematically underestimate self-polarization energies [120], may generate abnormally large fluctuations from native protein structure in simulation [121], and may incorrectly capture energetic breakdowns of residue-based contributions toward macromolecular binding [104], in addition to having other limitations of continuum models noted previously in this section. Thus, due to its relative generality when compared to more variable GB-based methods, rigorous solution of the LPBE is often considered the benchmark of continuum models [68, 106].

As part of their development, GB-based models have been extended to implicitly incorporate salt effects [122, 123]. In addition to Poisson- and GB-based methods, other continuum electrostatic models exist and are often rooted in PBE or GB models. As nonexhaustive examples, recent work has used the surface-based GB framework [124] as a starting point to approximate solvation energies without the need for obtaining effective Born radii [125, 126]. Closed-form, analytical approximations to the Poisson equation have been developed with the potential to be applied to arbitrarily shaped biological molecules [127, 128]. A variation in the PBE that explicitly models solvent dipoles has also been developed [129]. As a final example, "nonlocal" continuum models, in which the solvent reaction field at a point in space is also a function of nearby electric fields rather than only of the field at the given point (which is the assumption of the traditional Poisson equation), have been under development and analysis [130–132].

In summary, the primary assumption underlying continuum models is that solvent (and solute) polarization can be treated at the bulk level, without microscopic-level details. The effects of desolvation and the dampening of solute interactions by solvent are incorporated into the model but at far less computational expense than explicit simulation. As a result, continuum models have seen wide use in the analysis of biomolecular interactions. Nevertheless, the computational expense required for a continuum model still affords some challenges in design applications using charge optimization and combinatorial approaches, as described below.

3 Electrostatic charge optimization as a guide for design

3.1 Overview and theoretical framework

The continuum models introduced above allow for quantification of the electrostatic component of binding between biological molecules in aqueous phase. In a design application where one seeks to alter an existing molecule to change its binding affinity or specificity to potential targets in solution, one may wish not only to quantify binding, but also to *evaluate* the existing electrostatics relative to what "could be," if the molecule were optimally suited for the design goal. The method of charge optimization, first introduced by Lee, Kangas, and Tidor [133, 134], quantifies this spirit. Electrostatic charge optimization determines a provably optimal charge distribution for a molecule (referred to hereafter as the "ligand") of a given shape that will maximize affinity toward the desired target (termed the "receptor"). While this charge distribution is hypothetical, as it may not be achievable by a chemically possible molecule, the optimal charge distribution can be compared to the ligand's existing charge distribution to identify molecular regions that can be altered to more closely approximate optimality.

The method relies on the linear response assumption above, such that the solvent reaction field potential generated by a charge is directly proportional to the charge's magnitude. As a result, the solvation energy of a single charge q_i is quadratic in the charge magnitude, and the interaction energy of one charge q_i with the reaction field generated by another (q_j) is proportional to the product of the two charge magnitudes, as is its direct Coulombic interaction with q_j as well. Thus, Eq. 6 above can be written:

$$G = \sum_i \frac{1}{2} q_i \phi_i = \sum_i \frac{1}{2} q_i c_{ii} q_i + \sum_j \sum_i \frac{1}{2} q_i c_{ij} q_j = q^T M q \tag{10}$$

where q is a vector representing the charge distribution of the system, often expressed with each element corresponding to an atom-centered point charge. M is a symmetric matrix whose elements are the proportionality constants c_{ij} above (the factor of ½ is folded into the matrix M here). Specifically, M_{ij} is (one-half) the potential at atom center i due to the Coulombic and reaction field potentials created by a unit charge at atom center j.

For a binding reaction, the change in free energy is the difference in free energy between the bound and unbound states:

$$\Delta G_{\text{elec}} = q^T (M_{\text{bound}} - M_{\text{unbound}}) q = q^T (M_{\text{diff}}) q \tag{11}$$

We now assume rigid binding: Neither partner undergoes conformational rearrangement upon binding. Under this assumption, the Coulombic contributions within each binding partner cancel between the bound and unbound states, leaving only the reaction field piece in $\boldsymbol{M}_{\mathbf{diff}}$ for charges within the same partner. We split $\boldsymbol{q}$ into n- and m-dimensional subvectors $\boldsymbol{q}_{\boldsymbol{L}}$ and $\boldsymbol{q}_{\boldsymbol{R}}$, representing the ligand and receptor charge distributions, and we also subdivide the $(n + m)x(n + m)$-dimensional unit-charge potential difference matrix, $\boldsymbol{M}_{\mathbf{diff}}$, into submatrices:

$$\Delta G_{\text{elec}} = q^T (M_{\text{diff}}) q = \begin{bmatrix} q_L^T q_R^T \end{bmatrix} \begin{bmatrix} L & \frac{C}{2} \\ \frac{C^T}{2} & R \end{bmatrix} \begin{bmatrix} q_L \\ q_R \end{bmatrix} \quad (12)$$

$$= q_L^T L q_L + q_L^T C q_R + q_R^T R q_R \quad (13)$$

The first term in Eq. 13 involves charges on only the ligand and represents the ligand desolvation penalty—that is, the energetic penalty to replace a high-dielectric solvent with a low-dielectric cavity in the shape of the receptor. $\boldsymbol{L}$ is an *nxn* matrix whose *ij*th element equals one-half the reaction field potential at charge center *i* due to a unit charge at charge center *j* (the Coulombic portion has canceled out). Likewise, the last term involves only charges on the receptor and represents the receptor desolvation penalty. The middle term involves charges on both the ligand and receptor and accounts for their solvent-screened Coulombic interaction. These three terms are pictorially represented in Fig. 2.

The rigid binding assumption guarantees that $\boldsymbol{L}$ and $\boldsymbol{R}$ are both positive semi-definite for systems of biological interest because under this assumption, the binding process is equivalent to replacing high-dielectric solvent with a low-dielectric cavity in the shape of the binding partner and then allowing charges on each partner to interact. A charge distribution will interact more favorably with a higher dielectric than with a lower one, and thus, replacing high-dielectric solvent with a lower-dielectric region will be unfavorable (or neutral for a completely uncharged binding partner) [135].

Charge optimization exploits the fact that these matrices are positive semi-definite. Graphically, the electrostatic binding free energy as a function of either $\boldsymbol{q_L}$ or $\boldsymbol{q_R}$ is an upward-facing paraboloid of dimension *n* or *m*, respectively. Therefore, there exists a provable minimum as a function of either—a ligand charge distribution that minimizes binding free energy toward the receptor, and vice versa. This minimum may not be unique [135] as a paraboloid may be "flat" along at least one dimension. To find an optimal charge distribution, we differentiate Eq. 13 with respect to either $\boldsymbol{q_L}$ or $\boldsymbol{q_R}$ and set the resulting expression to zero. For example, the expression for the optimal ligand charge distribution is as follows:

$$q_{L,\text{opt}} = -\frac{1}{2} L^{-1} C q_R \quad (14)$$

$\boldsymbol{q_{L,\text{opt}}}$ is the set of ligand point charges that would bind more tightly to the target than any other set of point charges at these locations on an identically shaped ligand. Conceptually, such an optimum exists in a solvated system because a ligand with charge magnitudes that are too high would interact too much with solvent in the unbound state, while a ligand with charge magnitudes that are too low would not interact with the receptor enough in the bound state (Fig. 3). The optimal charge distribution represents the "sweet spot" where favorable interactions between ligand and target optimally overcome the unfavorable ligand desolvation penalty.

To compute $\boldsymbol{q_{L,\text{opt}}}$, one must compute the inverse of the $\boldsymbol{L}$ matrix. Generally, $\boldsymbol{L}$ is explicitly calculated and then inverted. Computation of $\boldsymbol{L}$ scales as *n*, the number of ligand atoms (it requires 2*n* potential calculations when using the LPBE described above or *n* calculations of Born radii when using GB-based methods). One also must compute the $\boldsymbol{C*qr}$ vector product, which requires one potential calculation. Constraints can be enforced on the optimal charge distribution via Lagrange multipliers and other methods such that the overall monopole equals a specified integer or individual charges are constrained within a given range.

3.2 Applications in analysis and design

After initial development using simple model geometries [133, 134, 136], electrostatic charge optimization has been applied to many biological systems, both to analyze existing interactions and as a design strategy, to suggest changes to alter the affinity. One of the first analyses showed that the protein barnase is electrostatically close to its hypothetical optimum for tight binding toward its natural partner barstar [137, 138], suggesting that a natural molecule that strongly binds to another may mimic this hypothetical optimum as a consequence of evolution. Another early analysis by Sulea et al. evaluated the optimality of metal ions in binding sites [139]. Because only one charge was being optimized, the one-dimensional free energy parabola could be directly visualized and the optimality of an intermediate value could be explicitly seen. For example, the authors confirmed that the monovalent K^+ ion is optimal for binding to the 18-crown-6 ether and that neutral or +2 species of similar radii would be significantly worse.

More complex analyses have been since undertaken. Several small molecule ligands have been evaluated in comparison with their optimal counterparts for binding their respective targets, including kinase inhibitors [140, 141], HIV-1 reverse transcriptase inhibitors [142], and substrates for *E. coli* glutaminyl tRNA synthetase [143]. Such analyses provide general insight into the electrostatic determinants of tight binding that can be indirectly applied to design novel inhibitors. In multiple cases, it was found that key hydrogen-bonding interactions across the interface are nearly optimal [141, 142], suggesting that electrostatics play a critical role in the recognition of targets and that optimality can be approached within the constraints of chemical space. Charge optimization can also help explain

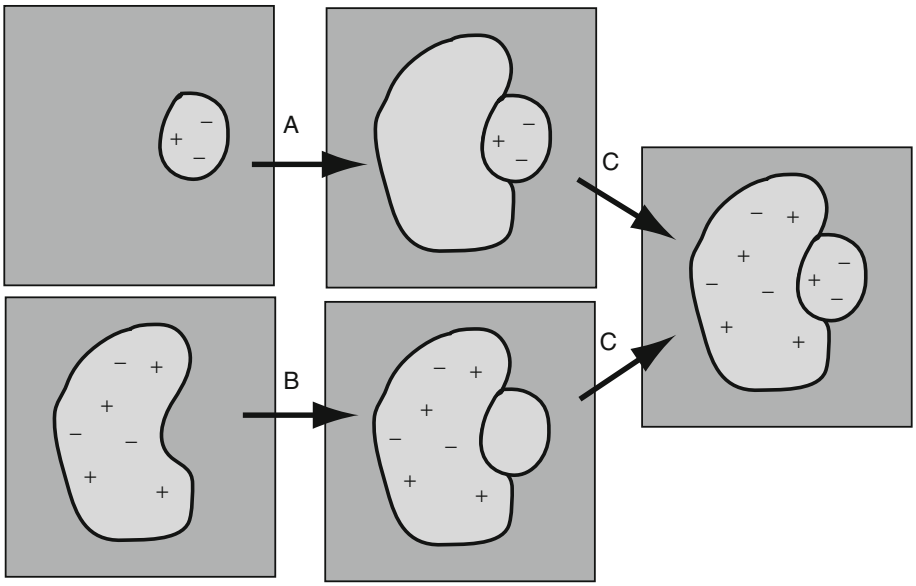

Fig. 2 Schematic showing the physical meaning of the ligand desolvation penalty (A), receptor desolvation penalty (B), and screened Coulombic interaction (C). In the thermodynamic cycle used to go from the unbound to the bound state, both partners are brought together to form the bound state and their charges are allowed to interact with solvent but not with each other, leading to a "fictitious" intermediate state in which each partner is "desolvated" by the low-dielectric cavity in the shape of its binding partner, but is unable to interact with the partner's charges (resulting in a desolvation penalty $= A + B$). From the fictitious intermediate state, the interaction between the binding partners is now turned on, resulting in the bound state, thus completing the cycle. In practice, the fictitious intermediate state is represented by two separate states, depicted above, providing separate values for A and B, and the bound state is reached by charging up the other binding partner in each case (providing C in either case). The total electrostatic component of the binding free energy is $A + B + C$

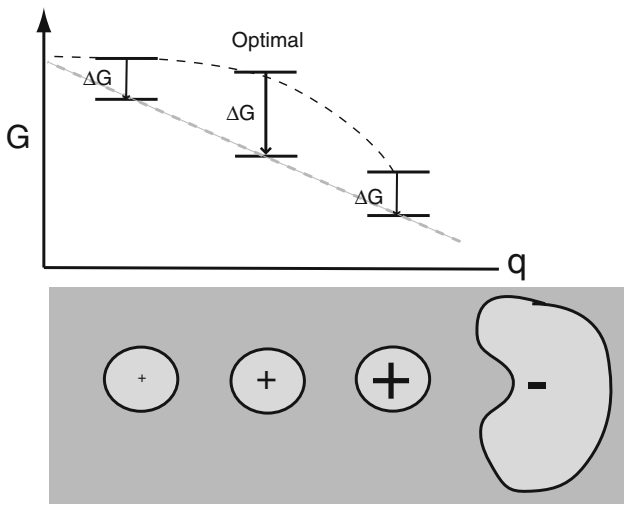

Fig. 3 Qualitative schematic showing the physical basis for the optimal charge distribution, using the one-dimensional case as an example. Three potential ligands with increasing charge magnitude are shown at *left*, with a receptor at *right*. Above each ligand, the *upper line* represents the relative stabilization of the unbound state due to desolvation effects (quadratic with respect to charge), and the *lower line* represents the relative stabilization of the bound state due to interaction (linear with respect to charge). The optimal binding free energy represents the intermediate balance between these two factors. Note that the binding free energies shown here exclude the constant receptor desolvation penalty

why target mutations can lead to decreased affinity, as drugs that are fairly optimal toward wild type may not be toward mutants [142].

The "true" test of charge optimization as a *design* strategy, however, is through predictions of chemical modifications to increase affinity that are subsequently validated experimentally. Charge optimization has produced these successes. Because the optimality of a ligand charge distribution is relative only to isosteric (i.e., shape-preserving) analogues with identical charge centers, the alterations tend to be relatively shape-preserving. For example, charge optimization revealed that a carboxylate group in a transition state analogue of *Bacillus subtilis* chorismate mutase was too negative for optimal binding, and so it was hypothesized that replacing this group with an essentially isosteric nitro group would increase the affinity [144]. This hypothesis was experimentally tested and validated, with the designed alteration leading to the most potent inhibitor of *Bacillus subtilis* chorismate mutase then known, binding $\sim$3 times more tightly than the original inhibitor [11]. Similarly, a fairly isosteric threonine to valine substitution was predicted in the design of peptides to bind HIV-1 protease more strongly than natural substrates [145]. This prediction was validated through explicit modeling of the substitution and ultimately by experiment,

which demonstrated a tenfold improvement in binding affinity for the designed peptide over any substrates. Experimental validation of predictions suggested by charge optimization has helped to establish its potential as a design strategy.

Charge optimization has also been used to suggest residue substitutions in an antibody system [146] and protein-based inhibitors of HIV-1 cell entry [147]. By providing the electrostatic signature of the "ideal" molecule, charge optimization allows for a somewhat tangible design goal. Indeed, using a series of neuraminidase inhibitors, Armstrong et al. showed that inhibitors that more closely resembled their optimal counterparts—both energetically and in their charge distributions—tended to bind more tightly to their targets [148]. Although this result was not a prediction, but rather an analysis of an existing series, it further demonstrates the utility of charge optimization as a prediction and design strategy.

3.3 Challenges and open problems in electrostatic charge optimization

3.3.1 The isostericity constraint

While charge optimization has been used to predict small changes in existing molecules to increase binding affinity, it has somewhat limited utility toward predicting more substantial changes in existing molecules or toward de novo design. Its limitations mainly stem from the assumptions upon which it rests. Under linear response, the electrostatic free energy is an analytical expression of the charge distribution only if the shapes of the ligand and complex and the locations of the charge centers are constant. This idea has been exploited not only for charge optimization but also for the rapid energetic evaluation of isosteric (identically shaped) analogues in binding sites [149]. However, if the dielectric boundary (i.e., the shapes of the ligand, receptor, or complex) changes, the elements of $\boldsymbol{L}$ and $\boldsymbol{Cq_r}$ also change, as all potentials must be recalculated (see Fig. 4). Consequently, the optimal charge distribution is optimal relative only to isosteric alternatives. This is perhaps the biggest limitation of using charge optimization for design, as few, if any, molecular alterations are *perfectly* isosteric.

One solution is to use charge optimization as a first step to reduce the design search space by identifying qualitative electrostatic alterations that might lead to tighter binding, thus eliminating alternatives. Chemical substitutions with these qualitative features can then be explicitly modeled using other methods. Armstrong et al. [148] showed that in lead optimization, where shapes and binding modes of compounds change along a series, changes that altered charges to qualitatively reproduce electrostatic optimality relative to a previous compound tended to improve affinity.

In completely de novo design, charge optimization as described above has more limited utility due to the fixed-shape restriction, as the general molecular shape is far less defined, if at all. However, charge optimization may be used to "map out" general characteristics of an optimal ligand. Sulea et al. [150] used a "probe-based" approach to characterize binding surfaces of proteins to aid in the design of optimal ligands. Specifically, small probes whose shapes were complementary to the local receptor surface were charge-optimized across the receptor surface, leading to a visualization of location-specific ligand optimal charges. The quantitative values achieved assume independence of the probes and would vary were the probes replaced with chemically realistic ligands, but the map nevertheless provides a convenient, ligand-independent visualization of qualitative ligand electrostatic properties to achieve tight binding at particular interfacial locations. Indeed, applying the probe-based approach to both trypsin and the PDZ domain resulted in optimal ligand properties

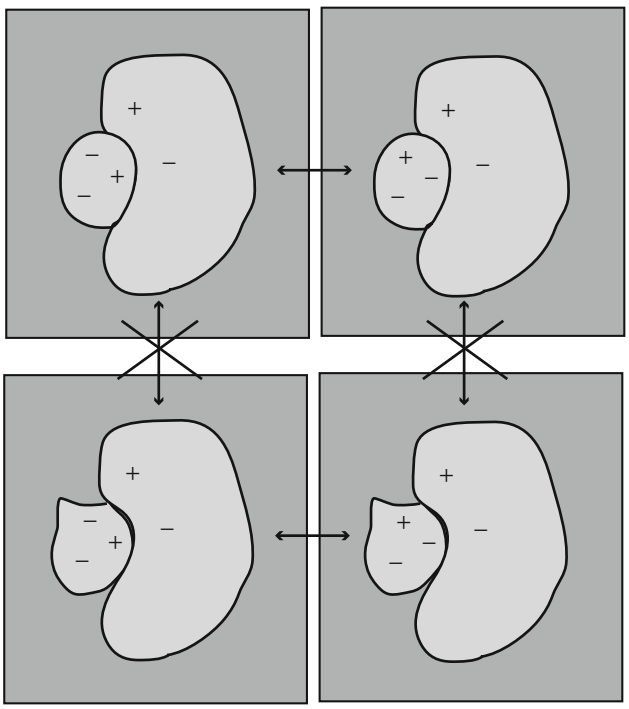

Fig. 4 Schematic depicting the isostericity constraint. *Horizontal arrows* depict changes in charge values in the molecular species involved, whose effect on free energy can be easily computed via an analytical formula (e.g., Eq. 13 in the case of binding free energy) assuming a one-time numerical pre-calculation of the relevant matrices. *Vertical arrows* depict changes in molecular shape, whose effect on free energy *cannot* be easily computed, as all matrix elements must be numerically recalculated for any shape change, even if the interacting point charges remain at the same absolute locations

that qualitatively matched experimentally known binders to each, providing some validation of this method. Another potential strategy for incorporating charge optimization in de novo design, especially for binding sites that are defined by a pocket or other spatial restriction, is to lay down a grid of points to cover the binding site and perform charge optimization on this grid. Then, one could design molecules that superimpose favorably with the resulting optimal grid charge distribution. This approach has yet to succeed, mainly because binding energetics may be highly sensitive to grid characteristics, and very different charge distributions could lead to nearly optimal energetics [M. Altman, personal communication]. Grid-based approaches to calculating potentials (rather than charges) have proven more useful for the rapid evaluation of molecular fragments for design and are discussed further below; nevertheless, successfully addressing the isostericity constraint to enable charge optimization to better tackle de novo design remains an open problem.

Another assumption upon which traditionally applied charge optimization rests is the rigid binding of the ligand. Most ligands undergo some conformational change upon binding, and so this assumption appears to be a substantial limitation. When the original charge optimization theory is applied to a flexible ligand, Coulombic contributions between ligand charges do not cancel out between the unbound and bound states, and the $\boldsymbol{L}$ matrix now accounts for both the (rigid) desolvation of the bound-state conformation and the screened, Coulombic "deformation" component to go from the unbound to the bound conformation (the latter component may be found using a cycle in which the ligand is completely desolvated, deformed, and resolvated, with the deformation penalty calculated directly with Coulomb's Law). While the desolvation component is never favorable for any charge distribution, the deformation component could be either favorable or unfavorable, and therefore, the overall $\boldsymbol{L}$ matrix need not be positive semidefinite. Hence, there need not exist a provable minimum in an unconstrained optimization.

By using the method of Lagrange multipliers to impose constraints on the total charge, Gilson [151] optimized the charge distribution of XK263 toward HIV-1 protease to "saddle points" on the free energy surface, assuming different ligand conformations in the unbound and bound states. While he found that "optimal" charge distributions and energetics were highly specific to the unbound-state geometry, the optimal charges obtained assuming rigid binding still produced desirable energetics when applied to the flexible systems. In other words, the optimal charge distribution obtained via the rigid binding assumption—while not strictly optimal in the flexible cases—appeared to be robustly "good" to changing ligand conformations. Therefore, while the rigid binding assumption may be limiting, the insights provided via traditional charge optimization may still be useful in systems that involve ligand conformational change. He also found that optimizing flexible ligands may produce charges that are optimized for the ligand bound-state conformation rather than for binding the target; the intramolecular Coulombic part of the energetics appeared to dominate over desolvation and interaction components, suggesting that a designed molecule with the "optimal" charge distribution would preferentially adopt the bound-state conformation over the proposed unbound-state conformation. In such cases, the "optimal binding free energy" is more an artifact of destabilizing the proposed unbound-state conformation rather than optimizing the "true" binding free energy. Recent promising work has overcome some of these issues to produce a charge optimization protocol that gives consistent results when applied in cases of ligand conformational change (Y. Shen, M.K. Gilson, and B. Tidor, unpublished data).

Nevertheless, the rigid binding assumption and isostericity constraints necessitate further study of the robustness of the optimal charge distribution to small perturbations in shape, conformation, or other properties. Such a rigorous analysis has yet to be done systematically, although preliminary efforts were conducted by using multiple independently refined binding sites within a crystal structure asymmetric unit [144], in which it was found that while absolute optimal binding free energies were quite conformationally dependent, the improvement in binding free energy upon optimization was fairly robust to conformation. Systematically optimizing a ligand charge distribution to multiple conformations obtained from a MD simulation, while it may be computationally costly, may lead to fundamental insights about which charge locations are most sensitive to perturbations to their local environment.

3.3.2 *Putting the results in context*

The optimal charge distribution can depend strongly on whether the entire molecule or a subset is being optimized. Sims et al. [140] demonstrated this idea with a fragment of a PKA inhibitor, PKI(5–24). When optimizing the entire molecule, the optimal charges of certain residues were predicted to *worsen* binding when other residues remained at wild-type values. They noted that if one wishes to alter parts of a molecule known in advance, one should optimize only those parts, as optimal charges are optimal only in the context of all other charges optimized.

Additionally, the optimal charge distribution should not be interpreted alone for design purposes but should be placed in context with the *sensitivities* of the binding free energies to each charge's value. An atom's charge might be quite far from its optimal value, but the binding free energy might be very insensitive to its value, such that altering it

will only minimally improve binding. Evaluating the results of a charge optimization along with accurate measures of sensitivity can allow for the prediction of changes to most greatly improve binding. Multiple strategies can be used to quantify the sensitivity. When a charge distribution is not optimal, the sensitivity of binding free energy to a charge is simply the derivative of the quadratic energy function with respect to it, which amounts to the potential difference between bound and unbound states at that charge location [151]. Such sensitivity analyses can be useful without optimization, as a way to understand how local perturbations of charge distributions can improve affinity [152] or affect stability [153]. However, in analyzing an optimal charge distribution, the gradient of the charge distribution upon (unconstrained) optimization is zero, by definition. The sensitivities are therefore captured by looking at convexity, or the "steepness" of the paraboloid in the various directions. Mathematically, independent sensitivities are given by the eigenvalues and corresponding eigenvectors of $\boldsymbol{L}$, with the larger eigenvalues and their eigenvectors denoting directions in charge space that have great impact on binding energetics. While mathematically the most "exact" measure of global sensitivity, the fully charge-coupled nature of the eigenvectors makes interpretation for practical applications difficult. At the other extreme, many have opted for an atom-by-atom approximate sensitivity that is easily interpretable but only valid assuming other charges remain at the globally optimal charge distribution [138, 141, 142]. In this approximation, the sensitivity of the binding free energy to an atom's charge is proportional to the atom's corresponding diagonal element of $\boldsymbol{L}$. Large diagonal values therefore represent that an atom is very "important" for determining the optimal binding free energy. This approximation becomes exact if only one charge is being optimized, as in the "probe-based" approach of Sulea noted earlier [150]. Here, sensitivity was found to depend highly on the concavity of the molecular surface interacting with the probe charge, and calculated sensitivity values for potential ligands interacting with trypsin and the PDZ domain were evaluated through comparison with experimental SAR data. Finally, as useful compromises between fully coupled and decoupled sensitivities, some have opted for metrics in which the sensitivity of an entire group of atoms on a molecule can be evaluated together [140], or when the other charges are allowed to "re-optimize" after one optimal charge changes by a charge unit [154]. In the latter case, such coupled sensitivity analysis identified a chemical moiety on celecoxib, an inhibitor of both COX2 and CAII, whose potential change in charge distribution was predicted to greatly influence binding to CAII but not to COX2; the authors validated this finding by comparing to existing experimental SAR data.

3.3.3 Numerical challenges and computational efficiency

Most commonly, the LPBE is numerically solved to estimate the potentials used to generate $\boldsymbol{L}$ and $\boldsymbol{Cq_r}$. Numerical imprecision may cause the smallest positive eigenvalues of $\boldsymbol{L}$ (corresponding to eigenmodes that have little impact on the binding free energy) to become negative, which drastically impact the optimization, as $\boldsymbol{L}$ is no longer positive semidefinite due to numerical error. These artifacts can be handled with "regularization" techniques such as singular value decomposition. Additionally, instead of explicitly constraining charge magnitudes, small positive eigenvalues may also be removed via conditioning techniques to avoid unreasonably large magnitudes along eigendirections toward which the binding free energy is insensitive.

Finally, electrostatic charge optimization can be computationally intensive, generally requiring $2n$ calculations of the potential when the LPBE is used, where n is the number of charge centers (or basis functions, more generally). Such costs can make optimization of multiple ligands or large ligands prohibitively expensive, and it also prohibits one from addressing the isostericity constraint by evaluating optimal charge distributions of systematically varied shapes. It also limits the size of the basis set used in the optimization. The limitations of a "fixed" atom-centered point charge model in the general modeling of electrostatic interactions have been noted [155, 156] with advances such as explicit modeling of lone pairs [155] and polarizable force fields (mentioned in Sect. 2) demonstrating the push toward a more "complete" representation of the charge distributions. With a fixed basis set (the incorporation of polarizable charge distributions into charge optimization theory is yet another potential future goal), the size and nature of the basis set used in charge optimization can greatly impact the obtained results [134]. It has been shown that in the limit of a complete basis and other physical constraints, the optimal electrostatic binding free energy is provably negative [135], while the computed optimal binding free energies using atom-centered point charges are often positive.

A more efficient charge optimization implementation can therefore allow not only for more molecules to be analyzed, but also for a given molecule to be studied in far more depth. To that end, Bardhan et al. developed a "Reverse-Schur" approach [157] that mathematically combines the optimization problem with the boundary-element formulation for calculating potentials, such that the optimization can be done as part of a single (large) solution of a linear system. Through clever matrix regularization techniques and preconditioning, the "co-optimization" was shown to be two orders of magnitude faster than the traditional, "one-row-at-a-time" assembling of the $\boldsymbol{L}$ matrix elements prior to the actual optimization.

Nevertheless, while this approach is clearly promising, additional speedups are certainly possible. It is worthwhile future work to evaluate the robustness of optimal charge distributions and their resulting energetics as a function of approximate (linear) models used to calculate the matrix elements. Robustness would require $\boldsymbol{L}$ matrix eigenvalue and eigenvector correspondence between methods. If a faster method can be shown to reproduce similar optimal characteristics as one using the LPBE, it could be another avenue by which the scope of problems addressed by charge optimization could increase. For example, using the GB method, $\boldsymbol{L}$ can be obtained by only two energetic evaluations [104]. While obtaining the Born radii needed for these energetic evaluations still scales as n, these computations are generally fast, suggesting a promising role for other approximate methods if they are shown to perform similarly to LPBE-based analyses.

3.4 Electrostatic specificity optimization for design

In keeping with the notion that either specific or promiscuous binding might be a more important design goal than affinity in certain cases, Kangas and Tidor extended the electrostatic charge optimization framework to attain optimal promiscuity (i.e., binding to multiple partners) toward desired targets and specificity against undesired targets [158]. Their framework was flexible to handle general design goals; a ligand has a set of desired targets, with each desired target potentially existing as multiple states (e.g., conformations, titration states). A ligand also has a set of undesired targets, or decoys. Here, the goal is to optimize a ligand such that it will bind as well as possible to *at least one* target state for a given target, while avoiding all possible decoys. Mathematically, they defined an objective function that uses Boltzmann-weighting of energies toward alternatives and, in the case of one target with multiple states, reduces to the free energy difference between the best binding desired state and the best binding decoy. Due to the Boltzmann weighting, the objective function is no longer generally convex, although it may become so in the zero temperature limit, under certain cases. Nevertheless, global optimality—a hallmark of traditional affinity optimization—does not generally apply here.

In the same work, Kangas and Tidor also introduced the "general specificity ligand"—a ligand that is (provably) optimally specific to its receptor when compared to all other isosteric receptors (under the rigid binding and fixed charge location assumptions). Such a design goal may be desirable when actual decoys are unknown but one still wishes to maximize specificity toward a receptor rather than affinity. The specificity-optimized ligand provides an interesting juxtaposition to the affinity-optimized ligand. While the latter is a ligand that will bind more tightly to a receptor than all other isosteric ligands, the former is a ligand that will bind more tightly to a receptor than it will to all other isosteric receptors. In the affinity case, the "competitors" are other ligands, while in the specificity case, the "competitors" are the other receptors. The specificity-optimized ligand is thus the ligand for whom the given receptor is optimized. One can write it in closed form by simply switching the ligand and receptor charge distributions in the expression for the affinity-optimized ligand, such that the receptor is now optimized toward the ligand, and then solving again for the ligand charge distribution:

$$q_{L,\mathrm{spec,opt}} = -2C^{-1}Rq_R \tag{15}$$

Interestingly, the ligand affinity-optimized and specificity-optimized charge distributions are always different (unless the receptor is completely uncharged), as it is impossible for both ligand and receptor to be mutually optimal for one another unless both are completely uncharged [134].

Electrostatic specificity optimization with known targets and decoys has been applied to optimize pepstatin for narrow specificity toward HIV-protease [159]; here, the "decoys" were other aspartyl proteases, pepsin and cathepsin D. The optimally specific charge distribution revealed that improvements in specificity could be achieved in multiple ways—modifications that improved binding to the target but worsened binding to the decoys, modifications that improved binding to both, but more so to the target, and modifications that worsened binding to both, but more so to the decoys. The optimal charge distribution was used to suggest specific (nonisosteric) chemical substitutions that were subsequently evaluated through explicitly constructing the shape-altering mutation and re-evaluating the new electrostatic properties through separate LPBE calculations, in addition to computing nonpolar components of the binding free energy. Additionally, HIV-protease inhibitors were optimized for broad specificity against a "panel" of HIV-1 protease mutants, and it was found that many inhibitors were fairly close to their broad-specificity optima, especially tipranavir, which experimentally is least impacted by the resistance mutations considered. Interestingly, when an inhibitor was optimized for broad specificity against multiple conformational variants of the wild-type protease, the resulting charge distribution was similar to the one optimized to bind broadly to explicit mutants. This suggests that optimization toward a wild-type ensemble may be a useful strategy to design inhibitors with broad recognition profiles when actual mutant targets are unknown.

In a different study, the HIV-1 reverse transcriptase (RT) inhibitor rilpivirine was simultaneously optimized to bind promiscuously to three variants of RT [142]. Here, the (convex) objective function was the sum of binding

affinities toward each variant, and the constraints, also convex, guaranteed that binding toward each variant was no worse than a threshold value above the affinity optimum for that variant. The resulting optimal ligand was far more hydrophobic than the original, likely because the individual targets required qualitatively different optimal charge distributions, and hydrophobicity represented the best "compromise."

The idea that an optimally promiscuous ligand tends to be hydrophobic is in line with anecdotal and experimental evidence that hydrophobic ligands tend to be more broadly recognizing, as are smaller ligands [160]. These ideas and other physical determinants of specificity and promiscuity were systematically explored by extending the theory of the general specificity ligand [161]. As the electrostatic free energy of binding is quadratic in receptor charges (assuming isostericity and linear response), the promiscuity of a ligand in this contrived case can be thought of as the "width" of this paraboloid, that is, the volume of receptor charge space that resides near the minimum point of the paraboloid. Shallow paraboloids represent promiscuous ligands, as many receptors will bind the ligand with similar binding free energies, while steep ones represent more specific ligands. Assumptions about isostericity, rigid binding, and other physical constraints were relaxed in turn from the contrived system in order to probe their effects on specificity. For example, it was found that hydrophobic ligands are more promiscuous because they are not as sensitive to shape differences as charged ligands, and their potential binding partners lie toward the center of biological "charge space," so there are more partners whose electrostatic properties are potentially complementary.

Electrostatic charge optimization has proven useful for gaining fundamental insights into the electrostatic determinants of molecular recognition, for analyzing specific systems, and for guiding design. Many studies are validated by experimental observations, either done prior to the analyses, or, in true design applications, afterward. Nevertheless, there are several challenges and open problems whose study can further increase its potential utility.

4 Combinatorial design using a continuum electrostatic framework

4.1 Overview: the need for pairwise decomposability and the return of the isostericity constraint

Here, I discuss the use of continuum solvent methods within a combinatorial design framework (the incorporation of various solvent models into other design strategies such as docking is addressed in reviews elsewhere [162–164]). Combinatorial approaches have been most commonly applied to protein design but have also been applied in the design of small molecules. Combinatorial design for proteins and small molecules is shown schematically in Fig. 5. The designed molecule consists of a "fixed" portion—often the protein backbone or a molecular scaffold used in a combinatorial synthetic scheme—and n variable parts—side chains or R groups. The design problem reduces to identifying which R groups are placed at each position in order to optimize some property (stability or binding energy, for example). As proteins naturally consist of a chemically invariable backbone and variable side chains, the combinatorial design strategies work particularly well here, although of course, in reality, the backbone is not rigid. Various strategies have been developed to generate or to account for multiple backbone conformations during the design [165–171]. Assuming a fixed region and n positions at which p possibilities are considered (p counts not only unique chemical moieties but also conformational variants of individual moieties), the size of the molecular and conformational space is p^n. As an example, to design a small 25-amino-acid peptide allowing an average of 10 conformations for each of the 20 amino acids, $p = 200$ and $n = 25$, for a total of $200^{25} \sim 10^{57}$ possibilities. In drug design applications, the number of positions is generally less, but the number of options at each position is generally much higher, leading again to a large search space. Such problems necessitate efficient algorithms to determine "good" or, more desirably, globally optimal solutions. Methods include dead-end-elimination [172–175], A* [176], and integer- and mixed-integer programming formulations [177–179] as well as stochastic methods, such as the ones used by Baker et al. [180]. We will not discuss them here except to state that often (and especially in the case of the global optimization), they rely on the expression for binding or folding free energy being *pairwise decomposable* in the molecular moieties being designed. In other words, efficient combinatorial optimization requires that the energy being optimized be expressed as follows:

$$E = E_0 + \sum_i E(i_r) + \sum_i \sum_{j>i} E(i_r, j_s) \tag{16}$$

Here, E_0 is the fixed portion of the energy, the single sum is a sum over positions of "*self-energies*" $E(i_r)$, where each term depends only on the individual option r chosen at a given position i, and the double sum is over pairs of positions, such that each term depends on the options chosen at a pair of positions and likely represents an interaction energy between the pair.

The algorithmic requirement for pairwise decomposability of the energy function is a primary challenge to incorporating many of the continuum electrostatic models into the combinatorial design process [181–185]. With the

exception of methods that assume a constant shape for the dielectric cavity encompassing the designed species, such as the spherical-cavity-based Tanford–Kirkwood methods [186, 187], Poisson equation–based methods are fundamentally nonpairwise decomposable. The reaction field generated by any charge within a molecule depends on the shape of the entire molecule, so calculating self- and pair-energies would require knowing the target molecular shape, which is unknown unless all R groups allowed at any given position were isosteric. This is the same premise that limits the utility of charge optimization to isosteric analogues. In other words, the use of continuum models in combinatorial design is challenged greatly by the isosteric constraint.

4.2 Electrostatic models within the pairwise-decomposable framework

Combinatorial design presents a general challenge in which the most efficient algorithms limit the functional form of the objective function. One can either globally optimize an "approximate" energy function, or one can very coarsely sample the enormous space using a more accurate energy function. One can either accurately score or sample, but not both. Throughout the years, various approaches have been taken to model solvent and other electrostatic effects within a combinatorial framework, many of which have been reviewed previously [69, 185] but will also be mentioned here. One approach is to use more empirical models that capture effects of solvent without necessarily motivating them entirely through physics-based formalisms, and representing them by pairwise additive functions. A historically used functional form for solvent screening of interactions that may not account for desolvation effects uses Coulomb's Law to model interactions between solute charges, but scales it by a distance-dependent dielectric "constant" in the denominator to account for solvent screening:

$$G_{\text{elec}} = \frac{q_1 q_2}{4\pi\varepsilon_0 r \varepsilon(r)} \tag{17}$$

The most common form of $\varepsilon(r)$ results in the electrostatic energy dropping off quadratically as a function of distance rather than linearly [188], but it can take on other pairwise functional forms [189, 190]. Ignoring or crudely modeling desolvation results in unrealistic designed molecules that have buried polar and hydrogen-bonding groups making no favorable interactions in the bound state. To work around this issue, scoring functions are often in place to ensure the satisfaction of hydrogen-bonding groups in the desired state, using donor and acceptor angles, distance, or functions based on experimental data to quantify "satisfaction" [191, 192]. Other, more parameterized empirical methods involve assuming the hydration energy of each atom is proportional to its exposed surface area (which can also be approximated in a pairwise-decomposable way [193–195]), with the proportionality constants determined by fitting to experimental data [196–201]; such models were still found to bury more polar groups relative to GB-based models [183], although with improved parameterization they perform reasonably well in protein design [202]. Still other models estimate desolvation by "excluded volume" rather than surface area [203]. Wisz and Hellinga developed a pairwise-decomposable empirical model that used varying dielectric constants between pairs of charges and in accounting for the desolvation of individual charges [204]. Cerutti et al. developed a pairwise-decomposable scheme based on approximations applied at various distance

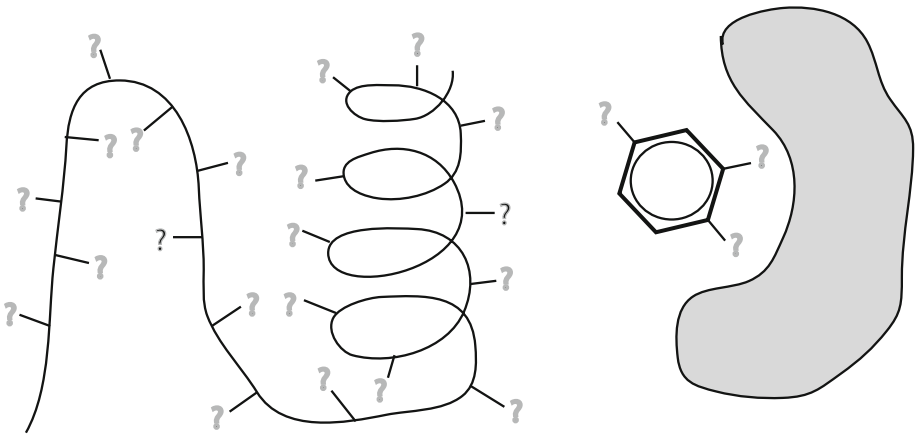

Fig. 5 Simplified schematic showing sample combinatorial frameworks in protein design (*left*) and drug design (*right*) toward a target (shown in *gray*). The *black lines* on the designed molecules represent "fixed" portions—in the example shown on the *left*, the protein backbone is held fixed, and at *right*, the benzene "scaffold" is fixed. The question marks indicate unknown positions at which multiple possibilities (unique chemical moieties and/or their varied conformations) may be considered, leading to a combinatorially large number of possible designs

regimes, some of which involved parameters fit to extensive data [205, 206]. In early years, empirical accounting for solvation has allowed for multiple design applications [194, 207], many of which were predictions that were subsequently experimentally validated [208–210]. In recent years, such solvation models have allowed for highly varied design applications that are far too numerous to list here, so the reader is referred to recent application-focused reviews [211–215]. Nevertheless, it has been found that the more accurate treatment of long-range electrostatics and polar interactions with solvent can be important in accurate structure prediction [180], thus suggesting the necessity of more quantitative prediction of solvent effects for robust design.

Marshall et al. provided an advance in the application of continuum models toward combinatorial design [216]. They developed pairwise-decomposable LPBE-based approximations for the desolvation and screened Coulombic interaction energies between components in protein design. They showed that certain one-body approximations were fairly accurate, such as assuming the desolvation of the backbone by all side chains upon folding was equal to the sum of the desolvations caused by each individual side chain in the absence of all others. For accurately approximating the desolvation of a side chain, however, it was necessary to consider the effect of not only the backbone but also other side chains as well. For computing side chain desolvation and screened interaction energies for a given side chain, a two-body framework was used, which added the contribution caused by "perturbing" the dielectric boundary to include every other side chain, one at a time. While this method showed reasonable accuracy in test cases, it systematically underestimated side chain desolvation and interaction energies, as it always assumed a smaller low-dielectric region than the designed protein would encompass. Parameterization was used to improve agreement initially. In later work [217], their method was improved at no additional computational cost by assuming a "generic side chain" of three connected spheres at each of the nonexplicitly modeled side chains for each pairwise term, to better approximate the shape and size of the ultimately designed protein cavity.

In addition to pairwise-decomposable LPBE-based methods, pairwise-decomposable generalized Born-based models have also been developed (note that many GB methods are pairwise at the *atomic* level, as Born radii are often calculated as a function involving sums over pairs of atoms, but they are not necessarily pairwise *decomposable*, due to the functional form of the energies). Pokala et al. [195] use "pseudoatoms," similar to the "generic side chains" above, to approximate the shape of the unknown side chains at remaining protein positions. Another model [218] results from the introduction of *residue* Born radii rather than the more traditional atomic Born radii. Interaction energies are defined as a fitted function of pairwise products of residue Born radii, where the fitting parameters depend only on the two residues in question. This method was recently applied to the computational redesign of asparaginyl-tRNA synthetase to bind aspartyl adenylate [219], although the designs did not show experimental activity. As another approach, the Sheffield solvation model [220] uses an alternative *ansatz* that somewhat resembles the Still equation, in which force-field atomic radii replace effective Born radii, and local and global shape features are captured in an average way using two parameters fit to PBE-acquired solvation energies. Because radii no longer explicitly depend on all atoms, the functional form is pairwise decomposable in a manner suited for protein design, although the model has thus far been designed for the analysis of small molecules. Nevertheless, pairwise-decomposable methods that are rooted in more traditional continuum electrostatic approaches may provide a promising balance between physically based accuracy and efficiency in combinatorial design in future applications.

In small molecule design, assuming the isostericity of designed molecules may be reasonable in certain cases, allowing one to more accurately approximate the solvation energy as pairwise additive. One may assume isostericity if the binding pocket shape is well defined and encompasses the designed molecule. In such a case, no water molecules likely enter the binding pocket in the bound state, such that the entire binding cavity, along with the receptor, can be modeled as a low dielectric in the bound state. Assuming a fairly tight cavity, this approximation is robust to changes in ligand shape when modeling the bound state. One turns the bound-state energy into a pairwise additive function by laying down a very fine grid over the binding site and precalculating desolvation and interaction potentials—essentially analogous to a (very laborious) $\boldsymbol{L}$ matrix and $\boldsymbol{Cq_r}$ vector calculation described in the context of charge optimization. After this set of calculations, however, the bound-state self and pairwise electrostatic energy of any functional group or pair of groups can be quickly estimated by extrapolating the relevant charges out onto the grid and performing the relevant matrix multiplication. Again, this technique exploits the idea that Poisson-based electrostatics *are* pairwise additive for the isosteric case. The issue arises in the modeling of the unbound state. To avoid computing separate potentials for each individual ligand, one must assume that the unbound shape of the ligand is still the "entire" binding cavity. This approximation is not robust to molecular shape and thus is the major weakness of such a model. Nevertheless, the method was found to work quite well in separating known binders from nonbinders in an engineering charged binding site [221] and in

designing novel HIV-1 protease inhibitors [222], when used in conjunction with hierarchical energy functions, described below; the latter design also required constraints enforcing the satisfaction of hydrogen-bonding groups.

Hierarchical strategies allow for the integration of continuum electrostatic models into the design process without directly coupling them to the scoring function used in the optimization. In such a strategy, one initially uses a more approximate energy function that is compatible with pairwise additive-based combinatorial optimization schemes. These schemes can efficiently determine the global optimum and enumerate all solutions within an energetic threshold of the global minimum. This "best subset" of structures found via the approximate energy function are then re-evaluated with a more accurate energy function or model that often includes continuum electrostatics. Assuming that there is reasonable correlation between the pairwise additive model and the continuum electrostatic one, one can have statistical confidence that the global energy minimum according to the more accurate method would be captured when enumerating the best n kcal/mol of the more approximate method. A framework has been developed that in part quantifies the statistical likelihood of capturing the global minimum of the more accurate energy function based on the value n of enumeration conducted at the initial stage [223]. Hierarchical energy functions have been applied to numerous designs, integrated with either docking [224–226] or combinatorial design [10, 222, 227] methodologies.

Finally, one must accept that the "benchmark" LPBE model, often used as a standard for accuracy, is itself an approximation and therefore may lead to erroneous prediction in design applications. In a critical study [228], Jaramillo et al. found that an LPBE-based model, along with multiple other continuum models, tended to favor the burial of polar or charged groups, which can lead to designs that are intuitively not physically realizable. Additionally, Morozov et al. found that empirical-based hydrogen-bond terms lead to better separation of native proteins from decoys than purely electrostatic models, and they identified other weaknesses of a purely Poisson-based approach [229]. Such findings suggest that even an LBPE-based approach that was perfectly compatible with algorithmic methods would not necessarily lead to robust design. With the limitations of current physics-based models, it may remain necessary to incorporate empirical-based constraints or aspects of the scoring function. Additionally, continuum models cannot capture effects of structurally ordered, discrete water molecules; such waters often play a role in stability or mediate interactions between binding partners [230–233]. In addition to modeling bulk solvent with a continuum model, it may be important to include discrete water molecules as part of the design. Various approaches to include discrete water molecules within combinatorial design exist, including using rotamers that are "pre-solvated" with discrete water molecules [234], considering a water molecule as a "position" within the overall design [10], and allowing waters to be placed at multiple positions around designed ligand [235].

4.3 Combinatorial design for specificity and promiscuity

Continuum electrostatic models have also been incorporated into more complicated designs in which the specificity or promiscuity was explicitly desired. Rather than thoroughly describe applications, I focus on methodology and problem formulation. First, one may incorporate specificity explicitly into the objective function, although it must remain pairwise decomposable in order to be compatible with many combinatorial optimization algorithms. For designing specificity over known decoys, one could minimize the difference between desired and undesired states, although naively optimizing this difference may result in an unrealistic conformation of the undesired states that would "conformationally relax" in reality to be far more stable. Therefore, one must allow for many possible conformations of the undesired states and either implicitly or explicitly incorporate an optimization of the undesired state conformations into the ultimate "fitness function" being optimized [236, 237]. Another strategy is to determine the best candidates for the desired state and select members from that subset that have the desired properties toward decoys through separate calculations or analyses. This approach, along with hierarchical energy functions ultimately using a LPBE framework, was used in the combinatorial, simultaneous design of calmodulin and target peptide mutants that specifically recognize each other over their wild-type variants [227]. Another application of negative design is the creation of proteins that prefer the monomeric state over aggregation. In protein design applications, surface positions are often poorly predicted, as models generally predict little energetic difference between placing apolar and polar residues on the surface. To prevent designs that may aggregate, pairwise-decomposable knowledge-based potentials that reflect the solvent exposures of each amino acid in native proteins can be incorporated [238]. Along those lines, it is important to note that in cases where a protein is being designed to bind tightly to a partner, such a design must also ensure that the protein fold is stable, either implicitly or explicitly creating a negative design problem in which the unfolded states are competitors. Modeling the unfolded state is indeed challenging, and electrostatics have been shown to affect the energetics and conformational biases of unfolded-state populations [239–242]. Finally, there are successful

designs incorporating continuum or empirical implicit solvent models in which increased specificity of the designed state over others was achieved simply by increasing affinity toward the desired states without explicitly considering competitors [243, 244]. Indeed, it has been argued that physical characteristics that optimize affinity to a desired target are unlikely to do so toward competitors to the same extent; however, if competitors are structurally similar, explicit negative design could be expedient [245], and there are certainly cases where explicit consideration of competing states is necessary, including the design of specifically interacting coiled coils [246–248] and homo- vs. hetero-dimer interactions [249]. Overall, due to the larger computational expense required when considering multiple states, there is still room for improvement in the incorporation of some of the more traditional continuum models directly into the negative design framework, as most successful applications have employed more empirical solvation and electrostatic models.

Combinatorial design of promiscuous binders or folders requires a different approach. Statistical-mechanical-based objective functions using Boltzmann weights allow for the most generality in design, but they are not pairwise decomposable and have limited utility in deterministic optimization, although they may be used in stochastic optimization [250]. Instead, one can minimize a linear sum of energies, although such a sum may not guarantee that each individual state has the desired level of favorability. In integer-programming-based formulations of combinatorial optimization, hard constraints can be applied to ensure that all designed energies are better than a certain threshold. Such a framework was developed for the combinatorial design of optimally small drug cocktails to collectively bind tightly to an ensemble of targets [251]. The method designs one molecule to bind toward a set of targets if such a molecule exists, and if not, it designs the smallest number of molecules that would. Like all globally optimal methods described here, it relies on pairwise decomposability of energies and, without isostericity, presents challenges to the incorporation of continuum electrostatic models.

Finally, the specificity or promiscuity of a designed molecule can be enforced through system-specific criteria, independently of the model used for electrostatics. For example, it has been proposed that HIV-1 protease inhibitors should be designed to fit within the region encompassed by its natural substrates such that target mutations that weaken drug binding will also weaken substrate binding, resulting in a less-fit virus [252, 253]. Therefore, one strategy for designing promiscuous drug molecules in such a system is to constrain groups to lie within the "substrate envelope." This strategy was successfully used in the combinatorial design of HIV-1 protease inhibitors using the substrate envelope as an isostericity constraint, thus enabling the use of an LPBE-based design framework [222]. Another strategy is to focus on those interactions that are maintained between the designed molecule and "constant" features of all potential binding partners, such as backbone moieties in the case of a mutating protein [254]. In existing protein–protein interactions, it was found that promiscuity could be achieved by a similar mechanism, whereby a small number of key "hotspot" residues may be optimized to maintain adequate promiscuity toward multiple partners, suggesting a shared mechanism of interaction toward all partners [255]. However, the same study also identified another natural mechanism of protein–protein promiscuity, in which different residues are used for each partner (perhaps due to structural diversity of the partners), with the overall interface representing a "compromise" between affinity toward each partner and overall promiscuity. These findings were supported by another redesign study in which the promiscuous protein calmodulin was found to have both "affinity" residues, which appeared to be optimized for all targets, and less-optimized "specificity"-determining residues that varied when the protein was designed for interaction with only one partner [256]. These studies suggest that knowing the structural similarities of intended targets can focus the design toward taking advantage of common interaction elements, if they exist [243], reducing the search space and potentially allowing for more computationally expensive electrostatic models to be used in the design process.

5 Concluding remarks

Continuum electrostatic models have many strengths when used to design biomolecular systems. First, unlike many other implicit models, they often account for both solvent screening and desolvation effects in a physically based way; this can allow for a more insightful breakdown of energetics when interpreting results. Additionally, shortcomings can often be attributed to physical, model-based assumptions rather than poor parameterization or ineffective scoring functions, leading to more robust improvement strategies and a more widely applicable model overall. Secondly, while continuum models are not as accurate in theory as a well-converged explicit solvent simulation, they are more computationally efficient, making them particularly suited for molecular design applications. As discussed in detail, a major challenge in incorporating continuum electrostatics into the design of molecules is the dependence of electrostatic solvation free energies on the entire shape of the molecule (the "isostericity" constraint). Nevertheless, great strides have been made in such

incorporation due to accurate approximation schemes, simplifying assumptions, and hierarchical models. Future progress may arise from a fundamental rethinking of the relationship between the algorithmic and scoring components of the optimization in such a way that more flexible energetic expressions can be efficiently optimized. With designs pushing the forefront of both the physical models and the algorithmic components, a creative and integrated expertise of both computer science and physical science will be crucial in developing such a paradigm shift.

Acknowledgments Mala Radhakrishnan wishes to acknowledge Michael Altman, Jaydeep Bardhan, David Green, David Huggins, Kathryn Loving, and Bruce Tidor for helpful discussions and critical feedback on the manuscript. She also acknowledges Wellesley College for support.

References

1. Sepulveda C, Poch A, Espinoza R, Cardemil E (2010) Electrostatic interactions play a significant role in the affinity of Saccharomyces cerevisiae phosphoenolpyruvate carboxykinase for Mn(2+). Biochimie 92(7):814–819. doi:10.1016/j.biochi.2010.02.032
2. Cortines JR, Weigele PR, Gilcrease EB, Casjens SR, Teschke CM (2011) Decoding bacteriophage P22 assembly: identification of two charged residues in scaffolding protein responsible for coat protein interaction. Virology 421(1):1–11. doi:10.1016/j.virol.2011.09.005
3. Busby B, Oashi T, Willis CD, Ackermann MA, Kontrogianni-Konstantopoulos A, MacKerell AD, Bloch RJ (2011) Electrostatic interactions mediate binding of obscurin to small ankyrin 1: biochemical and molecular modeling studies. J Mol Biol 408(2):321–334. doi:10.1016/j.jmb.2011.01.053
4. Thompson D, Simonson T (2006) Molecular dynamics simulations show that bound Mg2+ contributes to amino acid and aminoacyl adenylate binding specificity in aspartyl-tRNA synthetase through long range electrostatic interactions. J Biol Chem 281(33):23792–23803. doi:10.1074/jbc.M602870200
5. Brown CJ, Srinivasan D, Jun LH, Coomber D, Verma CS, Lane DP (2008) The electrostatic surface of MDM2 modulates the specificity of its interaction with phosphorylated and unphosphorylated p53 peptides. Cell Cycle 7(5):608–610
6. Thompson D, Plateau P, Simonson T (2006) Free-energy simulations and experiments reveal long-range electrostatic interactions and substrate-assisted specificity in an aminoacyl-tRNA synthetase. Chem Bio Chem 7(2):337–344. doi:10.1002/cbic.200500364
7. James LC, Tawfik DS (2003) The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. Protein Sci 12(10):2183–2193. doi:10.1110/ps.03172703
8. Friedler A, Veprintsev DB, Rutherford T, von Glos KI, Fersht AR (2005) Binding of Rad51 and other peptide sequences to a promiscuous, highly electrostatic binding site in p53. J Biol Chem 280(9):8051–8059. doi:10.1074/jbc.M411176200
9. Huggins DJ, Sherman W, Tidor B (2012) Rational approaches to improving selectivity in drug design. J Med Chem 55:1424–1444
10. Lippow SM, Wittrup KD, Tidor B (2007) Computational design of antibody-affinity improvement beyond in vivo maturation. Nat Biotechnol 25(10):1171–1176. doi:10.1038/nbt1336
11. Mandal A, Hilvert D (2003) Charge optimization increases the potency and selectivity of a chorismate mutase inhibitor. J Am Chem Soc 125(19):5598–5599. doi:10.1021/ja029447t
12. Wang YT, Wright JD, Doudeva LG, Jhang HC, Lim C, Yuan HS (2009) Redesign of high-affinity nonspecific nucleases with altered sequence preference. J Am Chem Soc 131(47):17345–17353. doi:10.1021/ja907160r
13. Caputo A, Parrish JC, James MNG, Powers JC, Bleackley RC (1999) Electrostatic reversal of serine proteinase substrate specificity. Proteins-Structure Function and Genetics 35(4):415–424. doi:10.1002/(sici)1097-0134(19990601)35:4<415:aid-prot5>3.0.co;2-7
14. Zhang YXL, Radhakrishnan ML, Lu XH, Gross AW, Tidor B, Lodish HF (2009) Symmetric signaling by an asymmetric 1 erythropoietin: 2 erythropoietin receptor complex. Mol Cell 33(2):266–274. doi:10.1016/j.molcel.2008.11.026
15. Sarkar CA, Lowenhaupt K, Horan T, Boone TC, Tidor B, Lauffenburger DA (2002) Rational cytokine design for increased lifetime and enhanced potency using pH-activated "histidine switching". Nat Biotechnol 20(9):908–913. doi:10.1038/nbt725
16. Joughin BA, Green DF, Tidor B (2005) Action-at-a-distance interactions enhance protein binding affinity. Protein Sci 14(5):1363–1369. doi:10.1110/ps.041283105
17. Zhou HX, Wong KY, Vijayakumar M (1997) Design of fast enzymes by optimizing interaction potential in active site. Proc Nat Acad Sci USA 94(23):12372–12377. doi:10.1073/pnas.94.23.12372
18. Antosiewicz J, Wlodek ST, McCammon JA (1996) Acetylcholinesterase: role of the enzyme's charge distribution in steering charged ligands toward the active site. Biopolymers 39(1):85–94. doi:10.1002/(sici)1097-0282(199607)39:1<85:aid-bip9>3.3.co;2-k
19. Shaul Y, Schreiber G (2005) Exploring the charge space of protein–protein association: a proteomic study. Proteins-Structure Function and Bioinformatics 60(3):341–352. doi:10.1002/prot.20489
20. Alsallaq R, Zhou HX (2007) Prediction of protein–protein association rates from a transition-state theory. Structure 15(2):215–224. doi:10.1016/j.str.2007.01.005
21. Schreiber G, Haran G, Zhou HX (2009) Fundamental aspects of protein–protein association kinetics. Chem Rev 109(3):839–860. doi:10.1021/cr800373w
22. Vijayakumar M, Wong KY, Schreiber G, Fersht AR, Szabo A, Zhou HX (1998) Electrostatic enhancement of diffusion-controlled protein–protein association: comparison of theory and experiment on barnase and barstar. J Mol Biol 278(5):1015–1024. doi:10.1006/jmbi.1998.1747
23. Sheinerman FB, Honig B (2002) On the role of electrostatic interactions in the design of protein–protein interfaces. J Mol Biol 318(1):161–177. doi:10.1016/s0022-2836(02)00030-x
24. Kundrotas PJ, Alexov E (2006) Electrostatic properties of protein–protein complexes. Biophys J 91(5):1724–1736. doi:10.1529/biophysj.106.086025
25. Hu ZJ, Ma BY, Wolfson H, Nussinov R (2000) Conservation of polar residues as hot spots at protein interfaces. Proteins-Structure Function and Genetics 39(4):331–342. doi:10.1002/(sici)1097-0134(20000601)39:4<331:aid-prot60>3.0.co;2-a
26. Brock K, Talley K, Coley K, Kundrotas P, Alexovy E (2007) Optimization of electrostatic interactions in protein–protein complexes. Biophys J 93(10):3340–3352. doi:10.1529/biophysj.107.112367
27. Das M, Basu G (2009) Coulomb energies of protein–protein complexes with monopole-free charge distributions. J Mol Graph Model 27(7):846–851. doi:10.1016/j.jmgm.2008.12.002
28. Zhang Z, Witham S, Alexov E (2011) On the role of electrostatics in protein–protein interactions. Phys Biol 8(3):10. doi:10.1088/1478-3975/8/3/035001
29. Talley K, Ng C, Shoppell M, Kundrotas P, Alexov E (2008) On the electrostatic component of protein–protein binding free energy. PMC Biophysics 1:2

30. Hendsch ZS, Tidor B (1994) Do salt bridges stabilize proteins—a continuum electrostatic analysis. Protein Sci 3(2):211–226
31. Lounnas V, Wade RC (1997) Exceptionally stable salt bridges in cytochrome P450cam have functional roles. Biochemistry 36(18):5402–5417. doi:10.1021/bi9622940
32. Elcock AH (1998) The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. J Mol Biol 284(2):489–502. doi:10.1006/jmbi.1998.2159
33. Xiao L, Honig B (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. J Mol Biol 289(5):1435–1444. doi:10.1006/jmbi.1999.2810
34. Kumar S, Nussinov R (1999) Salt bridge stability in monomeric proteins. J Mol Biol 293(5):1241–1255. doi:10.1006/jmbi.1999.3218
35. Vijayakumar M, Zhou HX (2001) Salt bridges stabilize the folded structure of barnase. J Phys Chem B 105(30):7334–7340. doi:10.1021/jp011214l
36. Dong F, Zhou HX (2002) Electrostatic contributions to T4 lysozyme stability: solvent-exposed charges versus semi-buried salt bridges. Biophys J 83(3):1341–1347
37. Bosshard HR, Marti DN, Jelesarov I (2004) Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. J Mol Recognit 17(1):1–16. doi:10.1002/jmr.657
38. Yu ZY, Jacobson MP, Josovitz J, Rapp CS, Friesner RA (2004) First-shell solvation of ion pairs: correction of systematic errors in implicit solvent models. J Phys Chem B 108(21):6643–6654. doi:10.1021/jp0378211
39. Pace CN, Grimsley GR, Scholtz JM (2009) Protein ionizable groups: pK Values and their contribution to protein stability and solubility. J Biol Chem 284(20):13285–13289. doi:10.1074/jbc.R800080200
40. Gong HP, Freed KF (2010) Electrostatic solvation energy for two oppositely charged ions in a solvated protein system: salt bridges can stabilize proteins. Biophys J 98(3):470–477. doi:10.1016/j.bpj.2009.10.031
41. Salari R, Chong LT (2010) Desolvation costs of salt bridges across protein binding interfaces: similarities and differences between implicit and explicit solvent models. J Phys Chem Lett 1(19):2844–2848. doi:10.1021/jz1010863
42. Warshel A, Kato M, Pisliakov AV (2007) Polarizable force fields: history, test cases, and prospects. J Chem Theory Comput 3(6):2034–2045. doi:10.1021/ct700127w
43. Cieplak P, Dupradeau FY, Duan Y, Wang JM (2009) Polarization effects in molecular mechanical force fields. J Phys-Condes Matter 21(33):21. doi:10.1088/0953-8984/21/33/333102
44. Lopes PEM, Roux B, MacKerell AD (2009) Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications. Theor Chem Acc 124(1–2):11–28. doi:10.1007/s00214-009-0617-x
45. Warshel A, Sharma PK, Kato M, Parson WW (2006) Modeling electrostatic effects in proteins. BBA-Proteins Proteomics 1764(11):1647–1676. doi:10.1016/j.bbapap.2006.08.007
46. Schnieders MJ, Baker NA, Ren PY, Ponder JW (2007) Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. J Chem Phys 126(12):21. doi:10.1063/1.2714528
47. Schnieders MJ, Ponder JW (2007) Polarizable atomic multipole solutes in a generalized Kirkwood continuum. J Chem Theory Comput 3(6):2083–2097. doi:10.1021/ct7001336
48. Tan YH, Tan CH, Wang J, Luo R (2008) Continuum polarizable force field within the Poisson-Boltzmann framework. J Phys Chem B 112(25):7675–7688. doi:10.1021/jp7110988
49. Wang JM, Cieplak P, Li J, Hou TJ, Luo R, Duan Y (2011) Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. J Phys Chem B 115(12):3091–3099. doi:10.1021/jp112133g
50. Wang JM, Cieplak P, Li J, Wang J, Cai Q, Hsieh MJ, Lei HX, Luo R, Duan Y (2011) Development of polarizable models for molecular mechanical calculations II: induced dipole models significantly improve accuracy of intermolecular interaction energies. J Phys Chem B 115(12):3100–3111. doi:10.1021/jp1121382
51. Schreiber H, Steinhauser O (1992) Molecular-dynamics studies of solvated polypeptides—Why the cutoff scheme does not work. Chem Phys 168(1):75–89. doi:10.1016/0301-0104(92)80111-8
52. Schreiber H, Steinhauser O (1992) Cutoff size does strongly influence molecular-dynamics results on solvated polypeptides. Biochemistry 31(25):5856–5860. doi:10.1021/bi00140a022
53. Beck DAC, Armen RS, Daggett V (2005) Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. Biochemistry 44(2):609–616. doi:10.1021/bi0486381
54. Darden T, York D, Pedersen L (1993) Particle mesh Ewald—an n.log(n) method for Ewald sums in large systems. J Chem Phys 98(12):10089–10092. doi:10.1063/1.464397
55. Shan YB, Klepeis JL, Eastwood MP, Dror RO, Shaw DE (2005) Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. J Chem Phys 122(5):054101. doi:10.1063/1.1839571
56. Wu XW, Brooks BR (2005) Isotropic periodic sum: a method for the calculation of long-range interactions. J Chem Phys 122(4):044107. doi:10.1063/1.1836733
57. Sagui C, Darden TA (1999) Molecular dynamics simulations of biomolecules: long-range electrostatic effects. Annu Rev Biophys Biomolec Struct 28:155–179. doi:10.1146/annurev.biophys.28.1.155
58. Levy RM, Gallicchio E (1998) Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. Annu Rev Phys Chem 49:531–567. doi:10.1146/annurev.physchem.49.1.531
59. Koehl P (2006) Electrostatics calculations: latest methodological advances. Curr Opin Struct Biol 16(2):142–151. doi:10.1016/j.sbi.2006.03.001
60. Huang A, Stultz CM (2007) Conformational sampling with implicit solvent models: application to the PHF6 peptide in tau protein. Biophys J 92(1):34–45. doi:10.1529/biophysj.106.091207
61. Zhang LY, Gallicchio E, Friesner RA, Levy RM (2001) Solvent models for protein-ligand binding: comparison of implicit solvent Poisson and surface generalized born models with explicit solvent simulations. J Comput Chem 22(6):591–607. doi:10.1002/jcc.1031
62. Roe DR, Okur A, Wickstrom L, Hornak V, Simmerling C (2007) Secondary structure bias in generalized born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. J Phys Chem B 111(7):1846–1857. doi:10.1021/jp066831u
63. Zhou RH, Berne BJ (2002) Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? Proc Nat Acad Sci USA 99(20):12777–12782. doi:10.1073/pnas.142430099
64. Zhou RH (2003) Free energy landscape of protein folding in water: explicit vs. implicit solvent. Proteins-Structure Function and Genetics 53(2):148–161. doi:10.1002/prot.10483
65. Stultz CM (2004) An assessment of potential of mean force calculations with implicit solvent models. J Phys Chem B 108(42):16525–16532. doi:10.1021/jp047126t
66. Tan CH, Yang LJ, Luo R (2006) How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. J Phys Chem B 110(37):18680–18687. doi:10.1021/jp063479b

67. Gouda H, Kuntz ID, Case DA, Kollman PA (2003) Free energy calculations for theophylline binding to an RNA aptamer: MM-PBSA and comparison of thermodynamic integration methods. Biopolymers 68(1):16–34. doi:10.1002/bip.10270
68. Baker NA (2005) Improving implicit solvent simulations: a Poisson-centric view. Curr Opin Struct Biol 15(2):137–143. doi:10.1016/j.sbi.2005.02.001
69. Boas FE, Harbury PB (2007) Potential energy functions for protein design. Curr Opin Struct Biol 17(2):199–204. doi:10.1016/j.sbi.2007.03.006
70. Levy RM, Zhang LY, Gallicchio E, Felts AK (2003) On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. J Am Chem Soc 125(31):9523–9530. doi:10.1021/ja029833a
71. Wagoner JA, Baker NA (2006) Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. Proc Nat Acad Sci USA 103(22):8331–8336. doi:10.1073/pnas.0600118103
72. Dzubiella J, Swanson JMJ, McCammon JA (2006) Coupling nonpolar and polar solvation free energies in implicit solvent models. J Chem Phys 124(8):12. doi:10.1063/1.2171192
73. Tan C, Tan YH, Luo R (2007) Implicit nonpolar solvent models. J Phys Chem B 111(42):12263–12274. doi:10.1021/jp073399n
74. Cramer CJ, Truhlar DG (1999) Implicit solvation models: equilibria, structure, spectra, and dynamics. Chem Rev 99(8):2161–2200. doi:10.1021/cr960149m
75. Roux B, Simonson T (1999) Implicit solvent models. Biophys Chem 78(1–2):1–20. doi:10.1016/s0301-4622(98)00226-9
76. Orozco M, Luque FJ (2000) Theoretical methods for the description of the solvent effect in biomolecular systems. Chem Rev 100(11):4187–4225. doi:10.1021/cr990052a
77. Feig M, Brooks CL (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr Opin Struct Biol 14(2):217–224. doi:10.1016/j.sbi.2004.03.009
78. Chen JH, Brooks CL, Khandogin J (2008) Recent advances in implicit solvent-based methods for biomolecular simulations. Curr Opin Struct Biol 18(2):140–148. doi:10.1016/j.sbi.2008.01.003
79. Kukic P, Nielsen JE (2010) Electrostatics in proteins and protein-ligand complexes. Future Med Chem 2(4):647–666. doi:10.4155/fmc.10.6
80. Simonson T (2001) Macromolecular electrostatics: continuum models and their growing pains. Curr Opin Struct Biol 11(2):243–252. doi:10.1016/s0959-440x(00)00197-4
81. Grochowski P, Trylska J (2008) Review: continuum molecular electrostatics, salt effects, and counterion binding-a review of the Poisson-Boltzmann theory and its modifications. Biopolymers 89(2):93–113. doi:10.1002/bip.20877
82. Gilson MK, Honig BH (1986) The dielectric-constant of a folded protein. Biopolymers 25(11):2097–2119. doi:10.1002/bip.360251106
83. Simonson T, Brooks CL (1996) Charge screening and the dielectric constant of proteins: insights from molecular dynamics. J Am Chem Soc 118(35):8452–8458. doi:10.1021/ja960884f
84. Schutz CN, Warshel A (2001) What axe the dielectric "constants" of proteins and how to validate electrostatic models? Proteins-Structure Function and Genetics 44(4):400–417. doi:10.1002/prot.1106
85. Karp DA, Gittis AG, Stahley MR, Fitch CA, Stites WE, Garcia-Moreno B (2007) High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. Biophys J 92(6):2041–2053. doi:10.1529/biophysj.106.090266
86. Chimenti MS, Castaneda CA, Majumdar A, Garcia-Moreno B (2011) Structural origins of high apparent dielectric constants experienced by ionizable groups in the hydrophobic core of a protein. J Mol Biol 405(2):361–377. doi:10.1016/j.jmb.2010.10.001
87. Gilson MK (2000–2006) Introduction to continuum electrostatics, with molecular applications pharmacy. ucsd.edu/labs/gilson/ce_www1a.pdf. 2011
88. Hill TL (1986) An introduction to statistical thermodynamics. Dover, New York
89. Vlachy V (1999) Ionic effects beyond Poisson-Boltzmann theory. Annu Rev Phys Chem 50:145–165. doi:10.1146/annurev.physchem.50.1.145
90. Borukhov I, Andelman D, Orland H (1997) Steric effects in electrolytes: a modified Poisson-Boltzmann equation. Phys Rev Lett 79(3):435–438. doi:10.1103/PhysRevLett.79.435
91. Sharp KA, Honig B (1990) Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. J Phys Chem 94(19):7684–7692. doi:10.1021/j100382a068
92. Fogolari F, Zuccato P, Esposito G, Viglino P (1999) Biomolecular electrostatics with the linearized Poisson-Boltzmann equation. Biophys J 76(1):1–16
93. Gilson MK, Sharp KA, Honig BH (1988) Calculating the electrostatic potential of molecules in solution—method and error assessment. J Comput Chem 9(4):327–335. doi:10.1002/jcc.540090407
94. Davis ME, McCammon JA (1989) solving the finite-difference linearized Poisson-Boltzmann equation—a comparison of relaxation and conjugate-gradient methods. J Comput Chem 10(3):386–391. doi:10.1002/jcc.540100313
95. Nicholls A, Honig B (1991) A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J Comput Chem 12(4):435–445. doi:10.1002/jcc.540120405
96. Liang J, Subramaniam S (1997) Computation of molecular electrostatics with boundary element methods. Biophys J 73(4):1830–1841
97. Altman MD, Bardhan JP, White JK, Tidor B (2009) Accurate solution of multi-region continuum biomolecule electrostatic problems using the linearized Poisson-Boltzmann equation with curved boundary elements. J Comput Chem 30(1):132–153. doi:10.1002/jcc.21027
98. Lu BZ, Cheng XL, Huang JF, McCammon JA (2009) An adaptive fast multipole boundary element method for Poisson-Boltzmann electrostatics. J Chem Theory Comput 5(6):1692–1699. doi:10.1021/ct900083k
99. Cortis CM, Friesner RA (1997) Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. J Comput Chem 18(13):1591–1608. doi:10.1002/(sici)1096-987x(199710)18:13<1591:aid-jcc3>3.0.co;2-m
100. Holst M, Baker N, Wang F (2000) Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples. J Comput Chem 21(15):1319–1342. doi:10.1002/1096-987x(20001130)21:15<1319:aid-jcc1>3.0.co;2-8
101. Chen L, Holst MJ, Xu JC (2007) The finite element approximation of the nonlinear Poisson-Boltzmann equation. Siam Journal on Numerical Analysis 45(6):2298–2320. doi:10.1137/060675514
102. Baker NA (2004) Poisson-Boltzmann methods for biomolecular electrostatics. Num Comput Methods Pt D 383:94
103. Lu BZ, Zhou YC, Holst MJ, McCammon JA (2008) Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. Commun Comput Phys 3(5):973–1009

104. Carrascal N, Green DF (2010) Energetic decomposition with the generalized-born and Poisson-Boltzmann solvent models: lessons from association of G-protein components. J Phys Chem B 114(15):5096–5116. doi:10.1021/jp910540z
105. Bashford D, Case DA (2000) Generalized born models of macromolecular solvation effects. Annu Rev Phys Chem 51:129–152. doi:10.1146/annurev.physchem.51.1.129
106. Onufriev A, Case DA, Bashford D (2002) Effective Born radii in the generalized Born approximation: the importance of being perfect. J Comput Chem 23(14):1297–1304. doi:10.1002/jcc.10126
107. Hawkins GD, Cramer CJ, Truhlar DG (1996) Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. J Phys Chem 100(51):19824–19839. doi:10.1021/jp961710n
108. Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A 101(16):3005–3014. doi:10.1021/jp961992r
109. Onufriev A, Bashford D, Case DA (2000) Modification of the generalized Born model suitable for macromolecules. J Phys Chem B 104(15):3712–3720. doi:10.1021/jp994072s
110. Lee MS, Salsbury FR, Brooks CL (2002) Novel generalized Born methods. J Chem Phys 116(24):10606–10614. doi:10.1063/1.1480013
111. Zhu J, Alexov E, Honig B (2005) Comparative study of generalized Born models: Born radii and peptide folding. J Phys Chem B 109(7):3008–3022. doi:10.1021/jp046307s
112. Chen JH, Im WP, Brooks CL (2006) Balancing solvation and intramolecular interactions: toward a consistent generalized born force field. J Am Chem Soc 128(11):3728–3736. doi:10.1021/ja05716r
113. Mongan J, Svrcek-Seiler WA, Onufriev A (2007) Analysis of integral expressions for effective Born radii. J Chem Phys 127(18):10. doi:10.1063/1.2783847
114. Cai W, Xu ZL, Baumketner A (2008) A new FFF-based algorithm to compute Born radii in the generalized Born theory of biomolecule solvation. J Comput Phys 227(24):10162–10177. doi:10.1016/j.jcp.2008.08.015
115. Lee MS, Feig M, Salsbury FR, Brooks CL (2003) New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. J Comput Chem 24(11):1348–1356. doi:10.1002/jcc.10272
116. Schaefer M, Karplus M (1996) A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 100(5):1578–1599. doi:10.1021/jp9521621
117. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 112(16):6127–6129. doi:10.1021/ja00172a038
118. Onufriev AV, Sigalov G (2011) A strategy for reducing gross errors in the generalized Born models of implicit solvation. J Chem Phys 134(16):15. doi:10.1063/1.3578686
119. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL (2004) Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 25(2):265–284. doi:10.1002/jcc.10378
120. Grycuk T (2003) Deficiency of the Coulomb-field approximation in the generalized Born model: an improved formula for Born radii evaluation. J Chem Phys 119(9):4817–4826. doi:10.1063/1.1595641
121. Wang T, Wade RC (2003) Implicit solvent models for flexible protein–protein docking by molecular dynamics simulation. Proteins-Structure Function and Genetics 50(1):158–169. doi:10.1002/prot.10248
122. Srinivasan J, Trevathan MW, Beroza P, Case DA (1999) Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. Theor Chem Acc 101(6):426–434. doi:10.1007/s002140050460
123. Tjong H, Zhou HX (2007) GBr(6)NL: a generalized Born method for accurately reproducing solvation energy of the nonlinear Poisson-Boltzmann equation. J Chem Phys 126(19):195102. doi:10.1063/1.2735322
124. Ghosh A, Rapp CS, Friesner RA (1998) Generalized born model based on a surface integral formulation. J Phys Chem B 102(52):10983–10990. doi:10.1021/jp982533o
125. Bardhan JP (2008) Interpreting the Coulomb-field approximation for generalized-Born electrostatics using boundary-integral equation theory. J Chem Phys 129(14):144105. doi:10.1063/1.2987409
126. Bardhan JP, Knepley MG (2011) Mathematical analysis of the boundary-integral based electrostatics estimation approximation for molecular solvation: Exact results for spherical inclusions. J Chem Phys 135(12). doi:10.1063/1.3641485
127. Gordon JC, Fenley AT, Onufriev A (2008) An analytical approach to computing biomolecular electrostatic potential. II. Validation and applications. J Chem Phys 129(7):075102. doi:10.1063/1.2956499
128. Fenley AT, Gordon JC, Onufriev A (2008) An analytical approach to computing biomolecular electrostatic potential. I. Derivation and analysis. J Chem Phys 129(7):075101. doi:10.1063/1.2956497
129. Abrashkin A, Andelman D, Orland H (2007) Dipolar Poisson-Boltzmann equation: ions and dipoles close to charge interfaces. Phys Rev Lett 99(7):077801. doi:10.1103/PhysRevLett.99.077801
130. Hildebrandt A, Blossey R, Rjasanow S, Kohlbacher O, Lenhof HP (2004) Novel formulation of nonlocal electrostatics. Phys Rev Lett 93(10):4. doi:108104
131. Hildebrandt A, Blossey R, Rjasanow S, Kohlbacher O, Lenhof HP (2007) Electrostatic potentials of proteins in water: a structured continuum approach. Bioinformatics 23(2):E99–E103. doi:10.1093/bioinformatics/btl312
132. Bardhan JP (2011) Nonlocal continuum electrostatic theory predicts surprisingly small energetic penalties for charge burial in proteins. J Chem Phys 135(10). doi:10.1063/1.3632995
133. Lee LP, Tidor B (1997) Optimization of electrostatic binding free energy. J Chem Phys 106(21):8681–8690
134. Kangas E, Tidor B (1998) Optimizing electrostatic affinity in ligand-receptor binding: theory, computation, and ligand properties. J Chem Phys 109(17):7522–7545
135. Kangas E, Tidor B (1999) Charge optimization leads to favorable electrostatic binding free energy. Phys Rev E 59(5):5958–5961
136. Chong LT, Dempster SE, Hendsch ZS, Lee LP, Tidor B (1998) Computation of electrostatic complements to proteins: a case of charge stabilized binding. Protein Sci 7(1):206–210
137. Lee LP, Tidor B (2001) Barstar is electrostatically optimized for tight binding to barnase. Nature Structural Biology 8(1):73–76
138. Lee LP, Tidor B (2001) Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. Protein Sci 10(2):362–377
139. Sulea T, Purisima EO (2001) Optimizing ligand charges for maximum binding affinity. a solvated interaction energy approach. J Phys Chem B 105(4):889–899. doi:10.1021/jp0038714
140. Sims PA, Wong CF, McCammon JA (2004) Charge optimization of the interface between protein kinases and their ligands. J Comput Chem 25(11):1416–1429. doi:10.1002/jcc.20067
141. Ahn JS, Radhakrishnan ML, Mapelli M, Choi S, Tidor B, Cuny GD, Musacchio A, Yeh LA, Kosik KS (2005) Defining Cdk5

ligand chemical space with small molecule inhibitors of Tau phosphorylation. Chem Biol 12(7):811–823
142. Minkara MS, Davis PH, Radhakrishnan ML (2012) Multiple drugs and multiple targets: an analysis of the electrostatic determinants of binding between non-nucleoside HIV-1 reverse transcriptase inhibitors and variants of HIV-1 RT. Proteins-Structure Function and Bioinformatics 80(2):573–590. doi:10.1002/prot.23221
143. Green DF, Tidor B (2004) Escherichia coli glutaminyl-tRNA synthetase is electrostatically optimized for binding of its cognate substrates. J Mol Biol 342(2):435–452. doi:10.1016/j.jmb.2004.06.087
144. Kangas E, Tidor B (2001) Electrostatic complementarity at ligand binding sites: application to chorismate mutase. J Phys Chem B 105(4):880–888. doi:10.1021/jp003449n
145. Altman MD, Nalivaika EA, Prabu-Jeyabalan M, Schiffer CA, Tidor B (2008) Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. Proteins-Structure Function and Bioinformatics 70(3):678–694. doi:10.1002/prot.21514
146. Clark LA, Boriack-Sjodin PA, Eldredge J, Fitch C, Friedman B, Hanf KJM, Jarpe M, Liparoto SF, Li Y, Lugovskoy A, Miller S, Rushe M, Sherman W, Simon K, Van Vlijmen H (2006) Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. Protein Sci 15(5):949–960. doi:10.1110/ps.052030506
147. Green DF, Tidor B (2005) Design of improved protein inhibitors of HIV-1 cell entry: optimization of electrostatic interactions at the binding interface. Proteins-Structure Function and Bioinformatics 60(4):644–657. doi:10.1002/prot.20540
148. Armstrong KA, Tidor B, Cheng AC (2006) Optimal charges in lead progression: a structure-based neuraminidase case study. J Med Chem 49(8):2470–2477. doi:10.1021/jm051105l
149. Sayle R, Nicholls A (2006) Electrostatic evaluation of isosteric analogues. J Comput-Aided Mol Des 20(4):191–208. doi:10.1007/s10822-006-9045-3
150. Sulea T, Purisima EO (2003) Profiling charge complementarity and selectivity for binding at the protein surface. Biophys J 84(5):2883–2896
151. Gilson MK (2006) Sensitivity analysis and charge-optimization for flexible ligands: applicability to lead optimization. J Chem Theory Comput 2(2):259–270. doi:10.1021/ct050226y
152. Sims PA, Wong CF, McCammon JA (2003) A computational model of binding thermodynamics: the design of cyclin-dependent kinase 2 inhibitors. J Med Chem 46(15):3314–3325. doi:10.1021/jm0205043
153. Zhang H, Wong CF, Thacher T, Rabitz H (1995) Parametric sensitivity analysis of avian pancreatic-polypeptide (APP). Proteins-Structure Function and Genetics 23(2):218–232. doi:10.1002/prot.340230211
154. Bhat S, Sulea T, Purisima EO (2006) Coupled atomic charge selectivity for optimal ligand-charge distributions at protein binding sites. J Comput Chem 27(16):1899–1907. doi:10.1002/jcc.20481
155. Harder E, Anisimov VM, Vorobyov IV, Lopes PEM, Noskov SY, MacKerell AD, Roux B (2006) Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator. J Chem Theory Comput 2(6):1587–1597. doi:10.1021/ct600180x
156. Kim BC, Young T, Harder E, Friesner RA, Berne BJ (2005) Structure and dynamics of the solvation of bovine pancreatic trypsin inhibitor in explicit water: a comparative study of the effects of solvent and protein polarizability. J Phys Chem B 109(34):16529–16538. doi:10.1021/jp051569v
157. Bardhan JP, Altman MD, Tidor B, White JK (2009) "Reverse-Schur" approach to optimization with linear PDE constraints: application to biomolecule analysis and design. J Chem Theory Comput 5(12):3260–3278. doi:10.1021/ct9001174
158. Kangas E, Tidor B (2000) Electrostatic specificity in molecular ligand design. J Chem Phys 112(20):9120–9131
159. Sherman W, Tidor B (2008) Novel method for probing the specificity binding profile of ligands: applications to HIV protease. Chem Biol Drug Des 71(5):387–407. doi:10.1111/j.1747-0285.2008.00659.x
160. Hopkins AL, Mason JS, Overington JP (2006) Can we rationally design promiscuous drugs? Curr Opin Struct Biol 16(1):127–136. doi:10.1016/j.sbi.2006.01.013
161. Radhakrishnan ML, Tidor B (2007) Specificity in molecular design: a physical framework for probing the determinants of binding specificity and promiscuity in a biological environment. J Phys Chem B 111(47):13419–13435. doi:10.1021/Jp074285e
162. Halperin I, Ma BY, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins-Structure Function and Genetics 47(4):409–443. doi:10.1002/prot.10115
163. Yuriev E, Agostino M, Ramsland PA (2010) Challenges and advances in computational docking: 2009 in review. J Mol Recognit 24(2):149–164. doi:10.1002/jmr.1077
164. Foloppe N, Hubbard R (2006) Towards predictive ligand design with free-energy based computational methods? Curr Med Chem 13(29):3583–3608. doi:10.2174/092986706779026165
165. Desjarlais JR, Handel TM (1999) Side-chain and backbone flexibility in protein core design. J Mol Biol 290(1):305–318. doi:10.1006/jmbi.1999.2866
166. Kono H, Saven JG (2001) Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. J Mol Biol 306(3):607–628. doi:10.1006/jmbi.2001.4422
167. Larson SM, England JL, Desjarlais JR, Pande VS (2002) Thoroughly sampling sequence space: large-scale protein design of structural ensembles. Protein Sci 11(12):2804–2813. doi:10.1110/ps.0203902
168. Saunders CT, Baker D (2005) Recapitulation of protein family divergence using flexible backbone protein design. J Mol Biol 346(2):631–644. doi:10.1016/j.jmb.2004.11.062
169. Fu XR, Apgar JR, Keating AE (2007) Modeling backbone flexibility to achieve sequence diversity: the design of novel a-helical ligands for Bcl-XL. J Mol Biol 371(4):1099–1117. doi:10.1016/j.jmb.2007.04.069
170. Humphris EL, Kortemme T (2008) Prediction of protein–protein interface sequence diversity using flexible backbone computational protein design. Structure 16(12):1777–1788. doi:10.1016/j.str.2008.09.012
171. Smith CA, Kortemme T (2011) Predicting the tolerated sequences for proteins and protein interfaces using RosettaBackrub flexible backbone design. PLoS One 6(7):11. doi:10.1371/journal.pone.0020451
172. Desmet J, Demaeyer M, Hazes B, Lasters I (1992) The dead-end elimination theorem and its use in protein side-chain positioning. Nature 356(6369):539–542. doi:10.1038/356539a0
173. Lasters I, Desmet J (1993) The fuzzy-end elimination theorem—correctly implementing the side-chain placement algorithm-based on the dead-end elimination theorem. Protein Eng 6(7):717–722. doi:10.1093/protein/6.7.717
174. Goldstein RF (1994) Efficient Rotamer elimination applied to protein side-chains and related spin-glasses. Biophys J 66(5):1335–1340
175. Looger LL, Hellinga HW (2001) Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. J Mol Biol 307(1):429–445. doi:10.1006/jmbi.2000.4424

176. Leach AR, Lemon AP (1998) Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. Proteins-Structure Function and Genetics 33(2):227–239. doi:10.1002/(sici)1097-0134(19981101)33:2<227:aid-prot7>3.0.co;2-f
177. Althaus E, Kohlbacher O, Lenhof HP, Muller P (2002) A combinatorial approach to protein docking with flexible side chains. J Comput Biol 9(4):597–612. doi:10.1089/10665270276027733 6
178. Kingsford CL, Chazelle B, Singh M (2005) Solving and analyzing side-chain positioning problems using linear and integer programming. Bioinformatics 21(7):1028–1036. doi:10.1093/bioinformatics/bti144
179. Zhu YS (2007) Mixed-integer linear programming algorithm for a computational protein design problem. Ind Eng Chem Res 46(3):839–845. doi:10.1021/ie0605985
180. Das R, Baker D (2008) Macromolecular modeling with Rosetta. In: Annual review of biochemistry, vol 77. Annual Reviews, Palo Alto, pp 363–382. doi:10.1146/annurev.biochem.77.062906.171838
181. Suarez M, Jaramillo A (2009) Challenges in the computational design of proteins. J R Soc Interface 6:15. doi:10.1098/rsif.2008.0508.focus
182. Lippow SM, Tidor B (2007) Progress in computational protein design. Curr Opin Biotechnol 18(4):305–311. doi:10.1016/j.copbio.2007.04.009
183. Lopes A, Alexandrov A, Bathelt C, Archontis G, Simonson T (2007) Computational sidechain placement and protein mutagenesis with implicit solvent models. Proteins-Structure Function and Bioinformatics 67(4):853–867. doi:10.1002/prot.21379
184. Kortemme T, Baker D (2004) Computational design of protein–protein interactions. Curr Opin Chem Biol 8(1):91–97. doi:10.1016/j.cpba.2003.12.008
185. Vizcarra CL, Mayo SL (2005) Electrostatics in computational protein design. Curr Opin Chem Biol 9(6):622–626. doi:10.1016/j.cbpa.2005.10.014
186. Tanford C, Kirkwood JG (1957) Theory of protein titration curves. 1. General equations for impenetrable spheres. J Am Chem Soc 79(20):5333–5339. doi:10.1021/ja01577a001
187. Havranek JJ, Harbury PB (1999) Tanford-Kirkwood electrostatics for protein modeling. Proc Nat Acad Sci USA 96(20):11145–11150. doi:10.1073/pnas.96.20.11145
188. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4(2):187–217. doi:10.1002/jcc.540040211
189. Gabb HA, Jackson RM, Sternberg MJE (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol 272(1):106–120. doi:10.1006/jmbi.1997.1203
190. Mallik B, Masunov A, Lazaridis T (2002) Distance and exposure dependent effective dielectric function. J Comput Chem 23(11):1090–1099. doi:10.1002/jcc.10104
191. Dahiyat BI, Gordon DB, Mayo SL (1997) Automated design of the surface positions of protein helices. Protein Sci 6(6):1333–1337
192. Kortemme T, Morozov AV, Baker D (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. J Mol Biol 326(4):1239–1259. doi:10.1016/s0022-2836(03)00021-4
193. Street AG, Mayo SL (1998) Pairwise calculation of protein solvent-accessible surface areas. Fold Des 3(4):253–258. doi:10.1016/s1359-0278(98)00036-4
194. Wernisch L, Hery S, Wodak SJ (2000) Automatic protein design with all atom force-fields by exact and heuristic optimization. J Mol Biol 301(3):713–736. doi:10.1006/jmbi.2000.3984
195. Pokala N, Handel TM (2004) Energy functions for protein design I: efficient and accurate continuum electrostatics and solvation. Protein Sci 13(4):925–936. doi:10.1110/ps.03486104
196. Eisenberg D, McLachlan AD (1986) Solvation energy in protein folding and binding. Nature 319(6050):199–203. doi:10.1038/319199a0
197. Ooi T, Oobatake M, Nemethy G, Scheraga HA (1987) Accessible surface-areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Nat Acad Sci USA 84(10):3086–3090. doi:10.1073/pnas.84.10.3086
198. Fraternali F, vanGunsteren WF (1996) An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 256(5):939–948. doi:10.1006/jmbi.1996.0139
199. Hou TJ, Qiao XB, Zhang W, Xu XJ (2002) Empirical aqueous solvation models based on accessible surface areas with implicit electrostatics. J Phys Chem B 106(43):11295–11304. doi:10.1021/jp025595u
200. Pei JF, Wang Q, Zhou JJ, Lai LH (2004) Estimating protein-ligand binding free energy: atomic solvation parameters for partition coefficient and solvation free energy calculation. Proteins-Structure Function and Bioinformatics 57(4):651–664. doi:10.1002/prot.20198
201. Dahiyat BI, Mayo SL (1996) Protein design automation. Protein Sci 5(5):895–903
202. Schmidt am Busch M, Lopes A, Amara N, Bathelt C, Simonson T (2008) Testing the Coulomb/accessible surface area solvent model for protein stability, ligand binding, and protein design. BMC Bioinformatics 9:16. doi:10.1186/1471-2105-9-148
203. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins-Structure Function and Genetics 35(2):133–152. doi:10.1002/(sici)1097-0134(19990501)35:2<133:aid-prot1>3.0.co;2-n
204. Wisz MS, Hellinga HW (2003) An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. Proteins-Structure Function and Genetics 51(3):360–377. doi:10.1002/prot.10332
205. Cerutti DS, Jain T, McCammon JA (2006) CIRSE: a solvation energy estimator compatible with flexible protein docking and design applications. Protein Sci 15(7):1579–1596. doi:10.1110/ps.051985106
206. Cerutti DS, Ten Eyck LF, McCammon JA (2005) Rapid estimation of solvation energy for simulations of protein–protein association. J Chem Theory Comput 1(1):143–152. doi:10.1021/ct049946f
207. Liang SD, Grishin NV (2004) Effective scoring function for protein sequence design. Proteins-Structure Function and Genetics 54(2):271–281. doi:10.1002/prot.10560
208. Dahiyat BI, Mayo SL (1997) De novo protein design: fully automated sequence selection. Science 278(5335):82–87. doi:10.1126/science.278.5335.82
209. Ogata K, Jaramillo A, Cohen W, Briand JP, Connan F, Choppin J, Muller S, Wodak SJ (2003) Automatic sequence design of major histocompatibility complex class I binding peptides impairing CD8(+) T cell recognition. J Biol Chem 278(2):1281–1290. doi:10.1074/jbcM.206853200
210. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. Proc Nat Acad Sci USA 98(25):14274–14279. doi:10.1073/pnas.251555398
211. Pantazes RJ, Grisewood MJ, Maranas CD (2011) Recent advances in computational protein design. Curr Opin Struct Biol 21(4):467–472. doi:10.1016/j.sbi.2011.04.005
212. Saven JG (2011) Computational protein design: engineering molecular diversity, nonnatural enzymes, nonbiological cofactor complexes, and membrane proteins. Curr Opin Chem Biol 15(3):452–457. doi:10.1016/j.cpa.2011.03.014

213. Mandell DJ, Kortemme T (2009) Computer-aided design of functional protein interactions. Nat Chem Biol 5(11):797–807. doi:10.1038/nchembio.251
214. Tian P (2010) Computational protein design, from single domain soluble proteins to membrane proteins. Chem Soc Rev 39(6):2071–2082. doi:10.1039/b810924a
215. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG (2011) Theoretical and computational protein design. In: Annual review of physical chemistry, vol 62. Annual Reviews, Palo Alto, pp 129–149. doi:10.1146/annurev-physchem-032210-103509
216. Marshall SA, Vizcarra CL, Mayo SL (2005) One- and two-body decomposable Poisson-Boltzmann methods for protein design calculations. Protein Sci 14(5):1293–1304. doi:10.1110/ps.041259105
217. Vizcarra CL, Zhang NG, Marshall SA, Wingreen NS, Zeng C, Mayo SL (2008) An improved pairwise decomposable finite-difference Poisson-Boltzmann method for computational protein design. J Comput Chem 29(7):1153–1162. doi:10.1002/jcc.20878
218. Archontis G, Simonson T (2005) A residue-pairwise generalized Born scheme suitable for protein design calculations. J Phys Chem B 109(47):22667–22673. doi:10.1021/jp055282+
219. Polydorides S, Amara N, Aubard C, Plateau P, Simonson T, Archontis G (2011) Computational protein design with a generalized Born solvent model: application to asparaginyl-tRNA synthetase. Proteins-Structure Function and Bioinformatics 79(12):3448–3468. doi:10.1002/prot.23042
220. Grant JA, Pickup BT, Sykes MJ, Kitchen CA, Nicholls A (2007) A simple formula for dielectric polarisation energies: the Sheffield Solvation model. Chem Phys Lett 441(1–3):163–166. doi:10.1016/j.cplett.2007.05.008
221. Huggins DJ, Altman MD, Tidor B (2009) Evaluation of an inverse molecular design algorithm in a model binding site. Proteins-Structure Function and Bioinformatics 75(1):168–186. doi:10.1002/prot.22226
222. Altman MD, Ali A, Reddy G, Nalam MNL, Anjum SG, Cao H, Chellappan S, Kairys V, Fernandes MX, Gilson MK, Schiffer CA, Rana TM, Tidor B (2008) HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. J Am Chem Soc 130(19):6099–6113. doi:10.1021/ja076558p
223. Green DF (2010) A statistical framework for hierarchical methods in molecular simulation and design. J Chem Theory Comput 6(5):1682–1697. doi:10.1021/ct9004504
224. Given JA, Gilson MK (1998) A hierarchical method for generating low-energy conformers of a protein-ligand complex. Proteins-Structure Function and Genetics 33(4):475–495. doi:10.1002/(sici)1097-0134(19981201)33:4<475:aid-prot3>3.0.co;2-b
225. Gruneberg S, Stubbs MT, Klebe G (2002) Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. J Med Chem 45(17):3588–3602. doi:10.1021/jm011112j
226. Floriano WB, Vaidehi N, Zamanakos G, Goddard WA (2004) HierVLS hierarchical docking protocol for virtual ligand screening of large-molecule databases. J Med Chem 47(1):56–71. doi:10.1021/jm030271v
227. Green DF, Dennis AT, Fam PS, Tidor B, Jasanoff A (2006) Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. Biochemistry 45(41):12547–12559. doi:10.1021/bi060857u
228. Jaramillo A, Wodak SJ (2005) Computational protein design is a challenge for implicit solvation models. Biophys J 88(1):156–171. doi:10.1529/biophysj.104.042044
229. Morozov AV, Kortemme T, Baker D (2003) Evaluation of models of electrostatic interactions in proteins. J Phys Chem B 107(9):2075–2090. doi:10.1021/jp0267555
230. Janin J (1999) Wet and dry interfaces: the role of solvent in protein–protein and protein-DNA recognition. Structure with Folding & Design 7(12):R277–R279. doi:10.1016/s0969-2126(00)88333-1
231. Li Z, Lazaridis T (2003) Thermodynamic contributions of the ordered water molecule in HIV-1 protease. J Am Chem Soc 125(22):6636–6637. doi:10.1021/ja0299203
232. Szep S, Park S, Boder ET, Van Duyne GD, Saven JG (2009) Structural coupling between FKBP12 and buried water. Proteins-Structure Function and Bioinformatics 74(3):603–611. doi:10.1002/prot.22176
233. Li Z, Lazaridis T (2007) Water at biomolecular binding interfaces. Phys Chem Chem Phys 9(5):573–581. doi:10.1039/b612449f
234. Jiang L, Kuhlman B, Kortemme TA, Baker D (2005) A "solvated rotainer" approach to modeling water-mediated hydrogen bonds at protein–protein interfaces. Proteins-Structure Function and Bioinformatics 58(4):893–904. doi:10.1002/prot.20347
235. Huggins DJ, Tidor B (2011) Systematic placement of structural water molecules for improved scoring of protein–ligand interactions. Protein Eng Des Sel 24(10):777–789. doi:10.1093/protein/gzr036
236. Leaver-Fay A, Jacak R, Stranges PB, Kuhlman B (2011) A generic program for multistate protein design. Plos One 6(7). doi:10.1371/journal.pone.0020937
237. Yanover C, Fromer M, Shifman JM (2007) Dead-end elimination for multistate protein design. J Comput Chem 28(13):2122–2129. doi:10.1002/jcc.20661
238. DeLuca S, Dorr B, Meiler J (2011) Design of native-like proteins through an exposure-dependent environment potential. Biochemistry 50(40):8521–8528. doi:10.1021/bi200664b
239. Cho JH, Raleigh DP (2005) Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. J Mol Biol 353(1):174–185. doi:10.1016/j.jmb.2005.08.019
240. Anil B, Craig-Schapiro R, Raleigh DP (2006) Design of a hyperstable protein by rational consideration of unfolded state interactions. J Am Chem Soc 128(10):3144–3145. doi:10.1021/ja057874b
241. Fitzkee NC, Bertrand Garcia-Moreno E (2008) Electrostatic effects in unfolded staphylococcal nuclease. Protein Sci 17(2):216–227. doi:10.1110/ps.073081708
242. Shen JK (2010) Uncovering specific electrostatic interactions in the denatured states of proteins. Biophys J 99(3):924–932. doi:10.1016/j.bpj.2010.05.009
243. Karanicolas J, Kuhlman B (2009) Computational design of affinity and specificity at protein–protein interfaces. Curr Opin Struct Biol 19(4):458–463. doi:10.1016/j.sbi.2009.07.005
244. Yosef E, Politi R, Choi MH, Shifman JM (2009) Computational design of calmodulin mutants with up to 900-fold increase in binding specificity. J Mol Biol 385(5):1470–1480. doi:10.1016/j.jmb.2008.09.053
245. Eaton BE, Gold L, Zichi DA (1995) Lets get specific—the relationship between specificity and affinity. Chem Biol 2(10):633–638. doi:10.1016/1074-5521(95)90023-3
246. Schreiber G, Keating AE (2011) Protein binding specificity versus promiscuity. Curr Opin Struct Biol 21(1):50–61. doi:10.1016/j.sbi.2010.10.002
247. Grigoryan G, Reinke AW, Keating AE (2009) Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature 458(7240):859–U2. doi:10.1038/nature07885
248. Havranek JJ, Harbury PB (2003) Automated design of specificity in molecular recognition. Nature Structural Biology 10(1):45–52. doi:10.1038/nsb877

249. Bolon DN, Grant RA, Baker TA, Sauer RT (2005) Specificity versus stability in computational protein design. Proc Nat Acad Sci USA 102(36):12724–12729. doi:10.1073/pnas.0506124102
250. Allen BD, Mayo SL (2009) An efficient algorithm for multistate protein design based on FASTER. J Comput Chem 31(5): 904–916. doi:10.1002/jcc.21375
251. Radhakrishnan ML, Tidor B (2008) Optimal drug cocktail design: methods for targeting molecular ensembles and insights from theoretical model systems. J Chem Inf Model 48(5): 1055–1073. doi:10.1021/ci700452r
252. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. Structure 10(3):369–381. doi:10.1016/s0969-2126(02)00720-7
253. King NM, Prabu-Jeyabalan M, Nalivaika EA, Wigerinck P, de Bethune MP, Schiffer CA (2004) Structural and thermodynamic basis for the binding of TMC114, a next-generation human immunodeficiency virus type 1 protease inhibitor. J Virol 78(21):12012–12021. doi:10.1128/jvi.78.21.12012-12021.2004
254. Ohtaka H, Freire E (2005) Adaptive inhibitors of the HIV-1 protease. Prog Biophys Mol Biol 88(2):193–208. doi:10.1016/j.pbiomolbio.2004.07.005
255. Humphris EL, Kortemme T (2007) Design of multi-specificity in protein interfaces. PLoS Comput Biol 3(8):1591–1604. doi:10.1371/journal.pcbi.0030164
256. Fromer M, Shifman JM (2009) Tradeoff between stability and multispecificity in the design of promiscuous proteins. PLoS Comput Biol 5(12):16. doi:10.1371/journal.pcbi.1000627