Molecular Based Mathematical Biology
Research Article • DOI: 10.2478/mlbmb-2013-0007 • MBMB • 2013 • 124-150

VERSITA

Analysis of fast boundary-integral approximations for modeling electrostatic contributions of molecular binding

Abstract

We analyze and suggest improvements to a recently developed approximate continuum-electrostatic model for proteins. The model, called BIBEE/I (boundary-integral based electrostatics estimation with interpolation), was able to estimate electrostatic solvation free energies to within a mean unsigned error of 4% on a test set of more than 600 proteins—a significant improvement over previous BIBEE models. In this work, we tested the BIBEE/I model for its capability to predict residue-by-residue interactions in protein–protein binding, using the widely studied model system of trypsin and bovine pancreatic trypsin inhibitor (BPTI). Finding that the BIBEE/I model performs surprisingly less well in this task than simpler BIBEE models, we seek to explain this behavior in terms of the models' differing spectral approximations of the exact boundary-integral operator. Calculations of analytically solvable systems (spheres and tri-axial ellipsoids) suggest two possibilities for improvement. The first is a modified BIBEE/I approach that captures the asymptotic eigenvalue limit correctly, and the second involves the dipole and quadrupole modes for ellipsoidal approximations of protein geometries. Our analysis suggests that fast, rigorous approximate models derived from reduced-basis approximation of boundary-integral equations might reach unprecedented accuracy, if the dipole and quadrupole modes can be captured quickly for general shapes.

Keywords

Component analysis • boundary-element methods • BIBEE • molecular electrostatics • continuum solvation • implicit-solvent models

MSC: 92C05, 92C40, 65N38, 65N80

Amelia B. Kreienkamp[1†], Lucy Y. Liu[1†], Mona S. Minkara[1], Matthew G. Knepley[2], Jaydeep P. Bardhan[3], Mala L. Radhakrishnan[1*]

1 *Department of Chemistry, Wellesley College, Wellesley MA 02481, USA*

2 *Computation Institute, U. Chicago, Chicago IL 60637, USA*

3 *Department of Electrical and Computer Engineering, Northeastern University, Boston MA 02115, USA*

Received 2012-10-8
Accepted 2013-04-29

1. Introduction

Electrostatic interactions play central roles in molecular biophysics, mediating both the affinity and specificity of interactions between biological molecules. The forces that charged and polar chemical groups exert on one another extend over very long distances, thus crucially regulating biological processes through the influence on molecular structure and recognition of potential binding partners. Though important for understanding and predicting biomolecular behavior, electrostatics are difficult to quantify due to the fundamental uncertainty involved in characterizing charge distributions

† *These authors contributed equally to this work.*
* *E-mail: mradhakr@wellesley.edu*

as well as the enormous number of degrees of freedom in the aqueous solvent. Another complication for modeling arises in the large range of scales, from systems with a few dozen atoms (the waters surrounding biological ions such as sodium and potassium [83, 86]) to those with millions of atoms, e.g. [94].

Accordingly, a considerable range of physical models exist to study biomolecule electrostatics within solvent, from quantum-mechanical models involving the Schrödinger equation to empirical (i.e., non-physical) models based on statistical analysis of experimental data sets. Two of the most popular models sit between these extremes. First, all-atom molecular dynamics (MD) simulations in explicit solvent [47, 51, 84] provide one approach that balances computational time and accuracy, although the computational effort required to simulate the reorganization of water on an atomistic level in response to changes in the solute configuration makes them currently infeasible for use in many design applications.

The other popular approach between the two extremes treats solvent as a polarizable continuum, i.e., without accounting for the microscopic details of individual water molecules. Continuum models approximately account for the solvent response, but in a more computationally efficient (faster) way than explicit solvent models, making them more appropriate for use in high-throughput design and analysis applications. Such models have been reviewed extensively [25, 27, 31, 36, 47, 48, 63, 65, 72, 80], and the most common are based on the Poisson equation. Although less computationally intense than explicit solvent simulation, accurate numerical solution of the Poisson equation is still relatively costly, thus making accurate, efficient approximations of the Poisson equation an attractive alternative. Numerous approximate models that ultimately originate from the Poisson continuum framework have been developed, including the widely-used class of Generalized Born models [16, 62, 77, 82]; such models often require additional parameterization. Recently, a class of Poisson-based models known as the Boundary Integral-Based Electrostatic Estimation (BIBEE) methods have been developed that may provide both reasonable accuracy and computational efficiency without the need for extensive parameterization [9, 13, 15].

In this paper we investigate how several BIBEE variants perform for estimating the electrostatic contributions to three quantities that are often used in the analysis of biomolecular systems: solvation free energies, binding free energies, and relative binding free energies. The electrostatic component of the solvation free energy is a common standard for assessing the accuracy of electrostatic models. The electrostatic contribution to the binding free energy of two molecules involves taking the difference between (electrostatic) solvation free energies, and a relative binding free energy involves a difference between binding free energies. Relative binding free energies enable a valuable approach for biomolecular analysis called *electrostatic component analysis* [23], in which the contributions of individual protein residues or molecular moieties can be quantified, providing a systematic identification of residues or chemical groups that are critical for molecular recognition of binding partners. This approach has been used to identify crucial determinants of binding in protein–protein and drug–target [23, 35, 39, 50, 59, 61] systems. *Relative* binding free energies are also crucial in molecular design, when one wishes to quantify the effect of altering a biological molecule's existing residues on its binding properties in order to design mutants with tighter or more specific binding [55]. They also enable direct comparisons, or rankings, among potential binding partners. Because biological processes are crucially mediated by both absolute and relative molecular binding free energetics, the differences of solvation free energies are usually more important than the solvation free energies themselves [53].

Model approximations make repeated subtractions even more dangerous than usual: in most areas of computational modeling, of course, practitioners put significant effort to reformulate calculations so as to avoid computing small differences between large numbers. To our knowledge, however, no such reformulation exists for estimating molecular binding free energies, and therefore several groups have developed higher-order (more accurate) numerical techniques to compute these differences accurately using the actual Poisson model [5, 7, 11, 18, 19, 24, 71, 90, 93]. We have been developing the BIBEE approach to translate the physical insights underlying Generalized-Born (GB) theory into mathematical notions of boundary-integral operator approximations, so that future development of fast models may be conducted via mathematical insights in addition to physical ones.

The recently proposed BIBEE/I model [13] is able to predict protein solvation free energies to within 4% mean unsigned relative error over a large test set [32]. This represented a substantial improvement in accuracy over the original BIBEE models, which motivated us to test the viability of fast BIBEE models for component analysis. We adopted as a model system the widely studied protein–protein binders of trypsin and bovine pancreatic trypsin inhibitor (BPTI) [20, 21, 64]; binding involves numerous interactions between both charged chemical groups and polar ones, especially near the specificity pocket. We find that several BIBEE methods provide qualitative but not quantitative agreement with full Poisson calculations, and that component analysis applications demand substantially better accuracy than

even the latest variant offers. To explain these results, we performed component analysis on simpler, analytically solvable geometries; this work led us to obtain a new interpretation of the BIBEE model as a type of reduced-basis approximation [26, 29]. Our results suggest that such a strategy might offer more accurate Poisson approximations in the future.

The present study represents the first application of BIBEE models to component analysis, as a case study in thoroughly testing the performance of fast Poisson approximations using the same types of calculations as are used in practical applications across biological science and engineering. Our paper provides two modeling frameworks to guide the future development and application of fast mathematical models for electrostatics. First, we suggest that *component analysis* provides a mathematically meaningful and application-driven approach to checking model accuracy. Detailed descriptions of component analysis may be found elsewhere [23], but the essential idea is to characterize the contributions of individual chemical groups (such as side chains on a protein, or functional groups on a drug-like molecule) to binding affinity and specificity. Therefore, other researchers building fast approximation theories [54, 87] and other electrostatic theories should find component analysis useful as well, though it only applies to linear-response theories and not more complex models [42]. Second, we identify a key area for improvement in the BIBEE approximation of boundary-integral formulations for molecular electrostatics. The approximation has evolved significantly already, from its original exploration of a Generalized Born theory [9] through careful mathematical analysis showing bounding properties of multiple variants [15], and methods with improved accuracy [13]. Here, we suggest that what is really needed are fast algorithms to estimate the dominant modes of the boundary-integral operators.

In the following section, we describe our model for the electrostatic contribution to molecular binding free energies, and the Poisson and BIBEE continuum models. Section 3 presents details of the numerical calculations and electrostatic component analysis, as well as how we prepared the atomistic protein geometries and simplified model geometries for simulations. Section 4 presents the results of our study of BIBEE component analysis in the trypsin/BPTI system and model geometries, illustrating that accurate approximations of solvation free energies are not sufficient for accurate component analysis. We do find, however, that model performance can predictably depend on system features, and we provide model analysis to support the observations. Section 5 introduces the new reduced-basis interpretation of the BIBEE model, and results for model geometries indicate the new framework's potential as a highly accurate approximation for Poisson calculations. Section 6 concludes the paper with a discussion of our results, and a possible strategy to extend reduced-basis BIBEE to general shapes.

2. Theory

We first outline a simple, widely used model of molecular binding and the electrostatic contributions to binding, and how the continuum electrostatic model can be used to perform component analysis and investigate interactions between chemical groups. We then present the four approximate models based on the BIBEE (boundary-integral based electrostatics estimation) approach [9, 13, 15], and conclude by briefly describing analytical solutions to the Poisson problem in spherical and ellipsoidal harmonics [14, 45].

2.1. A Simple Model for Molecular Binding

Figure 1 is a schematic diagram of the process of (non-covalent) molecular binding; we refer to the binding partners as *ligand* and *receptor*. The thermodynamic cycle decomposes the binding free energy into two types of steps: the transfer of a molecule between the solvent and a reference medium, and the binding of two molecules in the reference medium. In particular, the unbound molecules are first desolvated (transferred out of the solvent and into the reference medium), brought together in the reference medium, and then finally the ligand–receptor complex is re-solvated. The free energy change associated with transferring a molecule from the reference medium to the solvent is called its *solvation free energy* $\Delta G^{\text{solv,total}}$, and it is commonly decomposed into a sum of electrostatic and non-electrostatic components,

$$\Delta G^{\text{solv,total}} = \Delta G^{\text{solv}} + \Delta G^{\text{solv,non-es}}, \tag{1}$$

where ΔG^{solv} is the electrostatic solvation free energy. Using the thermodynamic cycle of Figure 1, the electrostatic component of the binding free energy in the solvent can be written

$$\Delta G^{\text{bind}} = \Delta G^{\text{solv,L-R}} + \Delta G^{\text{bind}}_{\text{(reference)}} - \Delta G^{\text{solv,R}} - \Delta G^{\text{solv,L}} \tag{2}$$

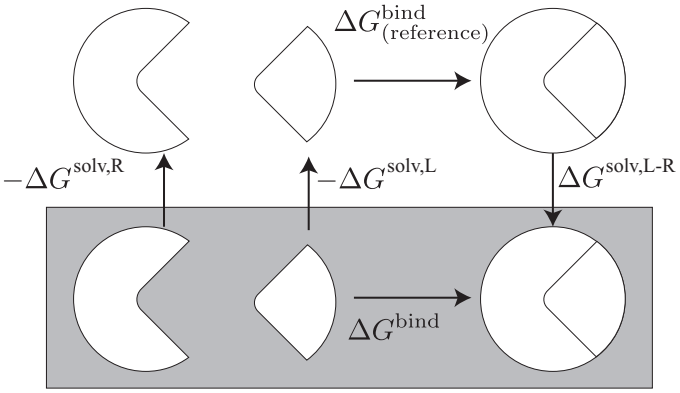

Fig 1. A simple thermodynamic cycle for estimating binding affinities. The shaded region in the bottom denotes the aqueous solution environment, and the unshaded region in the top denotes a reference medium in which estimating binding affinities is performed easily, e.g., a vacuum or homogeneous low-dielectric medium.

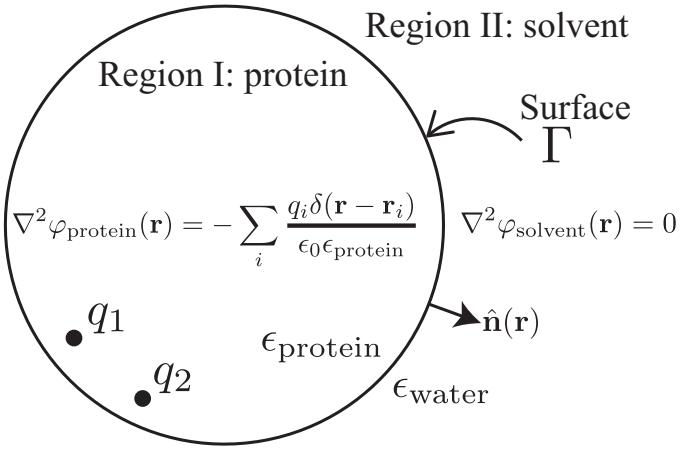

Fig 2. Schematic illustrating the continuum electrostatic model for molecular solvation used in this work.

where $\Delta G^{\text{bind}}_{(\text{reference})}$ is the free energy of binding in the reference medium. In the present study we follow the common approximation of assuming that the molecules are rigid, i.e. we do not account for conformational fluctuations and binding-induced conformational changes. Although more accurate predictions of molecular binding affinities require an account of these important considerations, the rigid-binding approximation has proved useful in numerous circumstances to provide insight into binding affinity and specificity. The electrostatic component of each term of Eq. (2) then gives the overall electrostatic contribution to the binding affinity.

2.2. Continuum Electrostatic Model for Solvation Free Energies

Figure 2 is a simplified diagram of the continuum electrostatic model we employ to estimate electrostatic solvation free energies. The solute interior, denoted as Region I in the figure, is modeled as a homogeneous dielectric with low dielectric constant $\epsilon_{\text{protein}} = 4$. The protein charge distribution, written $\rho(\mathbf{r})$, consists of a set of N_q discrete point charges, with values q_i and positions $\mathbf{r}_i$. The electrostatic potential in this region obeys the Poisson equation

$$\nabla^2\varphi_{\text{protein}}(\mathbf{r}) = -\rho(\mathbf{r})/\epsilon_0\epsilon_{\text{protein}}. \tag{3}$$

The solvent (exterior) region, denoted as Region II, is also modeled as a homogeneous dielectric, but with higher dielectric constant $\epsilon_{\text{water}} = 80$, approximately that of bulk water, and the potential is governed by the Laplace equation

$$\nabla^2 \varphi_{\text{solvent}}(\mathbf{r}) = 0. \tag{4}$$

We use this model, which describes a non-ionic aqueous solution, rather than the more biologically relevant Poisson–Boltzmann equation or its linearized form, because one of the purposes of this work is to understand the strengths and weaknesses of the BIBEE approach for component analysis, and to date the approximation applies only to the (non-ionic) Poisson problem.
The boundary between the dielectric regions is denoted Γ, and across it the potential is continuous, as is the normal component of the displacement field:

$$\varphi_{\text{protein}}(\mathbf{r}_\Gamma) = \varphi_{\text{solvent}}(\mathbf{r}_\Gamma) \tag{5}$$

$$\epsilon_{\text{protein}} \frac{\partial \varphi_{\text{protein}}(\mathbf{r}_\Gamma)}{\partial n} = \epsilon_{\text{water}} \frac{\partial \varphi_{\text{solvent}}(\mathbf{r}_\Gamma)}{\partial n}. \tag{6}$$

Here $\mathbf{r}_\Gamma$ denotes a point on the boundary and the normal direction is defined to point outward from Region I to Region II. The total electrostatic potential in the protein, $\varphi_{\text{protein}}(\mathbf{r})$, is the sum of the Coulomb potential from the charge $\rho(\mathbf{r})$, which we label as $\varphi_{\text{Coul}}(\mathbf{r})$, and another component that arises due to the difference in polarizability between the solute and solvent. This second component, denoted as $\varphi_{\text{reac}}(\mathbf{r})$, is often called the *reaction potential*. By defining the reference medium of the thermodynamic cycle in Figure 1 as a dielectric with the same permittivity as the protein, i.e. $\epsilon_{\text{protein}}$, the electrostatic component of the solvation free energy, ΔG^{solv}, is just the energy of interaction between the charge distribution and the reaction field:

$$\Delta G^{\text{solv}} = \frac{1}{2} \sum_{i=1}^{N_q} q_i \varphi_{\text{reac}}(\mathbf{r}_i). \tag{7}$$

The factor of 1/2 arises because the charges are interacting with their own reaction field in a linear-response theory, and the sum is over the N_q discrete charges because they represent the only fixed charges in the system. By defining the reference medium as one with permittivity $\epsilon_{\text{protein}}$ and assuming rigid binding, the electrostatic binding free energy in the reference medium (labeled as $\Delta G^{\text{bind}}_{(\text{reference})}$ in Figure 1) is equal to the Coulomb interaction energy between the molecular charge distributions.
Linear response also allows the reaction potential to be written as a scaled sum of the potentials induced by the individual charges; defining q as the N_q-length vector of charge values, we can write the N_q-length vector of reaction potentials at the charge locations as

$$\varphi_{\text{reac}} = Aq \tag{8}$$

where the symmetric, negative-definite *reaction-potential matrix* A has N_q rows and N_q columns. One can calculate A by performing N_q independent electrostatic calculations such that in the ith simulation, one sets $q_i = +1e$ and all the others to zero, and then computes the reaction potential at all of the charge locations.
We denote the vector of N_q^{ligand} ligand charges by q_L, the vector of the N_q^{receptor} receptor charges by q_R, the vector of $N_q^{\text{ligand}} + N_q^{\text{receptor}}$ charges for the ligand-receptor complex by $q_C = [q_L, q_R]$, and also denote by C_{RL} the Coulomb potential matrix for the interactions between the ligand and receptor charges. The Coulomb field at the ligand charges due to the receptor charges is then just $\varphi^{\text{Coul,receptor}}_{\text{complex}} = C_{RL} q_R$; by reciprocity, $\varphi^{\text{Coul,ligand}}_{\text{complex}} = C_{RL}^T q_L$. Eq. (2) is then written

$$\Delta G^{\text{bind}} = +\frac{1}{2} \begin{bmatrix} q_L^T & q_R^T \end{bmatrix} \begin{bmatrix} A^{complex}_{\text{LL}} & A^{complex}_{\text{LR}} \\ A^{complex}_{\text{RL}} & A^{complex}_{\text{RR}} \end{bmatrix} \begin{bmatrix} q_L \\ q_R \end{bmatrix} + q_R^T C_{RL} q_L - \frac{1}{2} q_L^T A^{ligand} q_L - \frac{1}{2} q_R^T A^{receptor} q_R, \tag{9}$$

where the superscripts on the matrices A denote the particular electrostatic problem; the reaction potential matrix for the complex, $A^{complex}$, has been partitioned into four block matrices that correspond to the reaction potentials that the

two charge distributions generate. Eq. (9) can be re-arranged into terms that depend on the ligand charge distribution only, on the receptor charge distribution only, and on both charge distributions:

$$\Delta G^{\text{bind}} = \underbrace{\frac{1}{2} q_L^T \left(A_{\text{LL}}^{complex} - A^{ligand} \right) q_L}_{\text{Only dependent on } q_L} + \underbrace{q_R^T \left(C_{\text{RL}} + A_{\text{RL}}^{complex} \right) q_L}_{\text{Dependent on both } q_L \text{ and } q_R} + \underbrace{\frac{1}{2} q_R^T \left(A_{RR}^{complex} - A^{receptor} \right) q_R}_{\text{Only dependent on } q_R}. \tag{10}$$

The first and last terms are the desolvation penalties paid by the ligand and receptor on binding, respectively; the second term, known as the interaction component, includes the direct Coulomb interaction between the two charge distributions as well as their solvent-screeened interaction on solvation.

2.3. Matrix Formalism and Component Analysis

To describe how we analyze the interactions between different chemical groups, it is helpful to write Eq. (10) in a simpler form as

$$\Delta G^{\text{bind}} = q_L^T L q_L + q_R^T R q_R + q_R^T C q_L \tag{11}$$

where

$$L = \frac{1}{2} \left(A_{\text{LL}}^{complex} - A^{ligand} \right) \tag{12}$$

$$R = \frac{1}{2} \left(A_{\text{RR}}^{complex} - A^{receptor} \right) \tag{13}$$

$$C = C_{\text{RL}} + A_{\text{RL}}^{complex}. \tag{14}$$

As noted previously, L and R are matrices representing the potential differences between the bound and unbound states assuming unit charges at each charge location in turn, on the ligand and receptor, respectively. For example, the (i, j) element of the L matrix is one-half the potential difference between the bound and unbound states at charge center i due to a unit charge at charge center j, with both charges belonging to the ligand charge distribution. Accordingly, $q_L^T L q_L$ represents the ligand desolvation penalty and $q_R^T R q_R$ represents the receptor desolvation penalty; $q_R^T C q_L$ represents the solvent-screened Coulombic interaction between the ligand and the receptor, with C_{ij} equaling the bound-state potential of charge i on the ligand due to a unit charge j on the receptor. Because the high dielectric solvent interacts favorably with the solutes, the ligand and receptor desolvation energies, $q_L^T L q_L$ and $q_R^T R q_R$ respectively, are always nonnegative under the rigid binding assumption; accordingly, L and R are positive semidefinite. The above matrix representation is independent of the method used to solve for the potentials to create the matrix elements, although of course, the matrix elements themselves will depend on the model used.

Once the matrix elements of L, C, and R are known, the electrostatic free energy of binding can be computed for any arbitrary charge distribution on either binding partner. Specifically, the charges on a particular group—an amino acid, for example—can be set to zero (approximating mutation to a hydrophobic isostere) and the electrostatic binding free energy trivially re-evaluated. The computational challenge is therefore in calculating the necessary matrix elements, and the goal of this study is to compare the accuracies of approximate models for doing this quickly.

If the binding free energy increases upon zeroing out the charges on a particular residue or chemical group, then that group's charge distribution favorably contributes to binding; conversely, if the binding free energy improves (decreases) upon removal of all charges, then the residue's charge distribution unfavorably contributes to binding. For each residue, we define

$$\Delta\Delta G_{\text{group,mut}}^{\text{bind}} = \Delta G_{\text{group}=0}^{\text{bind}} - \Delta G_{\text{original}}^{\text{bind}} \tag{15}$$

where $\Delta G_{\text{group}=0}^{\text{bind}}$ is the computed binding free energy when the charges on a given group are set to zero, $\Delta G_{\text{original}}^{\text{bind}}$ is the original binding free energy between the binding partners, and $\Delta\Delta G_{\text{group,mut}}^{\text{bind}}$ quantifies the change in binding free energy upon mutating a group to its hypothetical, hydrophobic isostere. We will usually denote this by $\Delta\Delta G^{\text{bind}}$. In this work, we systematically zero out the charges of entire residues to investigate the role of residues' charge distributions in mediating binding.

We note that if one wishes to analyze the residues of only one of the two binding partners, arbitrarily the ligand, then one need not calculate the full R and C matrices. For example, if only the ligand charge distribution is varied, one needs only the product Cq_R and the constant receptor desolvation penalty $q_R^T R q_R$.

2.4. Numerical Calculation of the Poisson Model

In this paper, we assess the accuracy of the approximate electrostatic models using complementary numerical methods for solving the Poisson continuum model: the finite-difference method (FDM) and the boundary element method (BEM). The finite-difference method has been a popular way to solve the Poisson and Poisson–Boltzmann problems, due in part to the wide availability of thoroughly tested software, e.g. DelPhi, UHBD, and APBS [8, 34, 46, 57, 70, 71, 85]. These methods solve an approximation to the spatially varying dielectric problem $\nabla \cdot (\epsilon(\mathbf{r})\nabla\varphi(\mathbf{r})) = -\rho(\mathbf{r})$ on a Cartesian grid. In contrast, boundary-element methods solve boundary-integral equation (BIE) re-formulations of the PDE problem [5, 6, 10, 44, 52, 56, 78, 89, 92]. For the mixed-dielectric Poisson problem here, one may use the well-known BIE

$$\sigma(\mathbf{r}) + \hat{\epsilon}\mathbf{n}(\mathbf{r}) \cdot \int_\Gamma \frac{\mathbf{r}-\mathbf{r}'}{4\pi|\mathbf{r}-\mathbf{r}'|^3}\sigma(\mathbf{r}')dA' = -\hat{\epsilon}\mathbf{n}(\mathbf{r}) \cdot \sum_k \frac{q_k}{\epsilon(\mathbf{r}_k)}\frac{\mathbf{r}-\mathbf{r}_k}{4\pi|\mathbf{r}-\mathbf{r}_k|^3} \tag{16}$$

where $\sigma(\mathbf{r})$ denotes the distribution of induced charge that develops at the dielectric boundary (the solute–solvent interface), $\hat{\epsilon} = 2(\epsilon_{\text{protein}} - \epsilon_{\text{water}})/(\epsilon_{\text{protein}} + \epsilon_{\text{water}})$ and the integral is assumed to be the principal value integral. Viewing the right-hand side as a linear map from the vector of point charges q to the normal component of the electric field we can write Eq. (16) in operator form as $(I + \hat{\epsilon}K)\sigma = Bq$; φ_{reac}, the vector of reaction potentials at the charge locations, is just the Coulomb potential induced by the induced surface charge $\sigma(\mathbf{r})$. Denoting the operator that maps from the surface charge to reaction potentials as S, the reaction potential matrix can be written

$$A = S(I + \hat{\epsilon}K)^{-1}B. \tag{17}$$

Readers interested in more details are referred to the extensive literature on numerical simulation of this BIE and others for solvation problems [7, 10, 12, 18, 19, 60, 73, 78, 91, 92].

2.5. Approximate Electrostatic Models

The BIBEE (boundary-integral-based electrostatics estimation) model was originally obtained by mathematical analysis of the surface-generalized Born (SGB) model of Ghosh et al. [33], who observed that the Coulomb-field approximation (CFA) used in many Generalized-Born (GB) theories could be related to the boundary-integral equation of Eq. (16). In particular, Ghosh et al. noted that in the CFA, the solvation free energy of a single charge is equal to the solvation free energy obtained by approximating Eq. (16) as

$$\left(1 - \frac{\hat{\epsilon}}{2}\right)\tilde{\sigma}_i^{\text{CFA,GB}}(\mathbf{r}) = -\hat{\epsilon}\mathbf{n}(\mathbf{r}) \cdot \frac{q_i}{\epsilon(\mathbf{r}_i)}\frac{\mathbf{r}-\mathbf{r}_i}{4\pi|\mathbf{r}-\mathbf{r}_i|^3}, \tag{18}$$

(seemingly neglecting the boundary-integral operator of Eq. (16) completely) and then calculating the effective Born radius R_i from the approximate reaction potential induced at $\mathbf{r}_i$ by $\tilde{\sigma}_i^{\text{CFA,GB}}(\mathbf{r})$. Surprisingly, this approximation gives the exact solution for a sphere with central charge, and this fact was noted by Ghosh et al. as a justification for the CFA. Bardhan showed that this exactness holds much more generally; in fact, this approximation is exact for any surface and charge distribution such that the charges generate a uniform normal electric field at the surface [9]. The BIBEE model generalized this approximation from the calculation of single-charge solvation free energies to *arbitrary* charge distributions

$$\left(1 - \frac{\hat{\epsilon}}{2}\right)\tilde{\sigma}^{\text{BIBEE/CFA}}(\mathbf{r}) = -\hat{\epsilon}\mathbf{n}(\mathbf{r}) \cdot \sum_k \frac{q_k}{\epsilon(\mathbf{r}_k)}\frac{\mathbf{r}-\mathbf{r}_k}{4\pi|\mathbf{r}-\mathbf{r}_k|^3}, \tag{19}$$

and here, in contrast to the GB/CFA approach, the *total* approximate surface charge is used to compute the reaction potential at all charge locations. A similar approach was later derived by Fedichev et al. [30]. The CFA essentially entails assuming $K = -\frac{1}{2}I$, in other words that all the eigenvalues of the electric field operator K are equal to $-\frac{1}{2}$. A complementary approximation can be found, in which one assumes that $K = 0$; because the BEM form of this approximation can be used as a preconditioner for Krylov-subspace iterative methods [9], we call this $K = 0$ approximation BIBEE/P.

Recently, two new variants on the BIBEE model were introduced [13], based on the Onufriev group's insightful analysis of GB theory [79]. The new BIBEE models, called BIBEE/M and BIBEE/I, exploit a simple but remarkable property

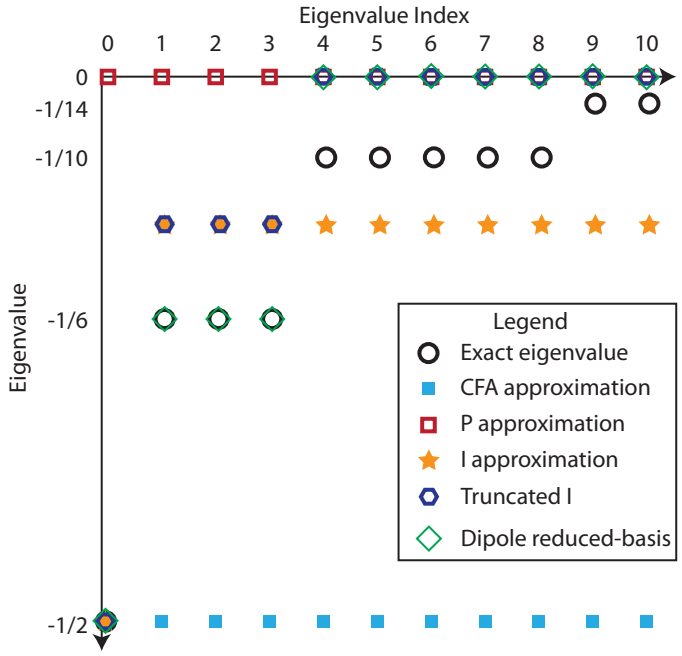

Fig 3. The exact eigenvalues of the integral operator K in Eq. (16) for a sphere [41], the approximate eigenvalues employed in previous BIBEE models [9, 13], and the approximate eigenvalues in a reduced-basis approach to BIBEE.

of the electric-field integral operator: the dominant eigenvalue of K is always $-1/2$, regardless of the surface (subject to certain restrictions on its smoothness), and the left and right eigenvectors are constant functions [15]. This mode can be treated analytically without approximation, leaving the other modes to be approximated. The BIBEE/M variant assumes all the other integral-operator eigenvalues to be 0, whereas BIBEE/I uses a single effective parameter λ^* for all other eigenvalues (to capture dominant modes such as the dipole and quadrupole response [13]). By fitting λ^* to match solvation free energies for a small set of proteins, one obtains with $\lambda^* = -0.2$ a surprisingly accurate model that exhibited 4% accuracy over the Feig et al. test set of more than 600 proteins [32]. A schematic of the various BIBEE approximations is shown in Figure 3. We note the various BIBEE approximations are essentially equally fast, approximately an order of magnitude faster than full BEM simulation; the various BIBEE models differ only in the least expensive step of the computation (application of the approximate inverse).

2.6. Analytical Solutions For Spherical and Ellipsoidal Solutes

Kirkwood's analytical solution for the reaction potential due to charges in a spherical solute begins by expanding the Coulomb potential from the charge distribution, as well as the reaction and solvent potentials, in spherical harmonics [45]. This is possible because spherical harmonics form a complete basis for the space of possible potentials, much like Fourier modes. We then enforce the boundary conditions, Eqs. (5) and (6), in order to determine the coefficient of each term in the series. This task is greatly simplified because spherical harmonics of different orders are orthogonal to each other, and therefore each harmonic can be treated individually. After this simplification, we obtain the equations below for a sphere of radius b,

$$\frac{C_{nm}}{\epsilon_{\text{protein}}} + b^{2n+1} R_{nm} = S_{nm} \tag{20}$$

$$\frac{C_{nm}}{\epsilon_{\text{water}}} - \frac{\epsilon_{\text{protein}}}{\epsilon_{\text{water}}} \frac{n}{n+1} b^{2n+1} R_{nm} = S_{nm}. \tag{21}$$

where the expansion coefficients for the Coulomb potential are denoted by C_{nm}, those for the reaction field by R_{nm}, and those for the potential in the solvent by S_{nm}; here (nm) is the index of a particular spherical harmonic $Y_n^m(\theta, \phi)$.

Because the coefficients C_{nm} can be computed directly from the solute charge distribution, we can solve these two linear equations in two unknowns to determine the coefficients; summing the series, we arrive at the full solvent and reaction fields.

Surprisingly, exactly the same formalism may be used to derive the BIBEE/CFA and BIBEE/P approximations. The key step is to view the approximate surface charges as satisfying a modified set of boundary conditions (see Eq. (41) in [13]), so that for BIBEE/CFA

$$\frac{\epsilon_{\text{protein}}}{\epsilon_{\text{water}}}\frac{\partial\varphi_{\text{Coul}}(\mathbf{r}_\Gamma)}{\partial n} = \frac{\partial\varphi_{\text{solvent}}(\mathbf{r}_\Gamma)}{\partial n} - \frac{\partial\varphi_{\text{reac}}(\mathbf{r}_\Gamma)}{\partial n}, \tag{22}$$

replaces Eq. (6), and we have decomposed φ_{protein} into direct Coulomb term from the solute charges, and the reaction field term:

$$\varphi_{\text{protein}} = \varphi_{\text{Coul}} + \varphi_{\text{reac}}. \tag{23}$$

Similarly, for BIBEE/P we replace Eq. (6) with

$$\frac{3\epsilon_{\text{protein}} - \epsilon_{\text{water}}}{\epsilon_{\text{protein}} + \epsilon_{\text{water}}}\frac{\partial\varphi_{\text{Coul}}(\mathbf{r}_\Gamma)}{\partial n} = \frac{\partial\varphi_{\text{solvent}}(\mathbf{r}_\Gamma)}{\partial n} - \frac{\partial\varphi_{\text{reaction}}(\mathbf{r}_\Gamma)}{\partial n}. \tag{24}$$

With these modified boundary conditions, we are able to derive a series solution for the approximate BIBEE potentials exactly as we did in the case of the full electrostatic operator above. Many consequences of this approximation are derived in [13].

Note that we have not used any specific properties of the spherical harmonics, other than the fact that they constitute a complete, orthogonal basis for the potential in space. Therefore, replacing the spherical harmonics with another complete, orthogonal basis will not change our methodology, but only the specific answers. The most general set of harmonic, orthogonal functions are the ellipsoidal harmonics [28, 40], which match more accurately the shape of many proteins. We can repeat the procedure above for ellipsoids, so that the Coulomb and reaction potentials are expanded in the internal ellipsoidal harmonics $\mathbb{E}_n^m(\mathbf{r})$, and the solvent potential is expanded in the external ellipsoidal harmonics $\mathbb{F}_n^m(\mathbf{r})$. The exact boundary conditions are [88]:

$$\frac{C_n^m}{\epsilon_{\text{protein}}} + R_n^m\frac{E_n^m(a)}{F_n^m(a)} = S_n^m \tag{25}$$

$$\frac{C_n^m}{\epsilon_{\text{water}}} + R_n^m\frac{\epsilon_{\text{protein}}}{\epsilon_{\text{water}}}\frac{\frac{\partial E_n^m(\lambda)}{\partial\lambda}\big|_a}{\frac{\partial F_n^m(\lambda)}{\partial\lambda}\big|_a} = S_n^m \tag{26}$$

where the integer pair (n, m) identifies a single harmonic just as for spherical harmonics, the functions $E_n^m(\lambda)$ and $F_n^m(\lambda)$ are the first-kind and second-kind Lamé functions for that harmonic, and the argument a is the ellipsoid's longest semi-axis. To derive a BIBEE model in ellipsoidal harmonics, we perform a similar procedure of considering the modified boundary condition as was done for the sphere.

3. Methods

3.1. Structure Preparation

Component analysis studies were carried out using a 1.85-Angstrom crystal structure of BPTI complexed with trypsin (PDB ID 3BTK [37]); Figure 4 is a rendering of the complex, with key residues highlighted. Resolved sulfate ions were removed from the structure, and water molecules with fewer than 3 potential protein hydrogen-bonding contacts were eliminated. Retained crystallographic waters were assigned to either BPTI or trypsin in the unbound state by manual examination of potential hydrogen bonding contacts with either partner. The amide groups of asparagine and glutamine side chains and the imidazole groups of histidine side chains were visually examined for potential hydrogen bonding contacts and flipped if a clear improvement in hydrogen bonding interactions would be obtained. Standard protonation states were assumed for all residues, and the δ-tautomer of Histidine-57 was modeled to preserve the hydrogen bonding network within the catalytic triad.

Hydrogen atoms were added using the HBUILD facility within CHARMM [22], using the all-atom CHARMM22 parameter set and force field [43]. Side chain or backbone atoms that were not resolved in the crystallographic experiment were

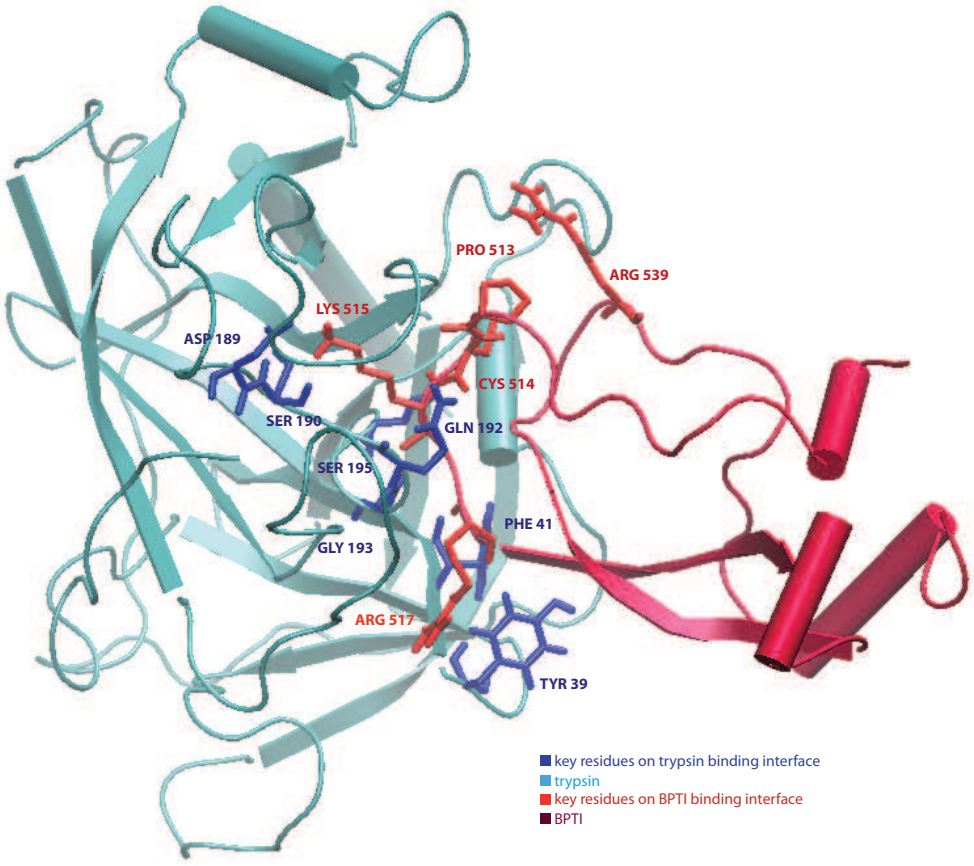

Fig 4. Rendering of the trypsin–BPTI complex, with trypsin in light blue and BPTI in dark red, with key residues for binding highlighted.

added and energy-minimized using CHARMM. The first two residues of BPTI were unresolved, and therefore Asp503 was computationally patched with an acetyl group. For all continuum electrostatic calculations, PARSE radii and charges were used [81]. Only atoms with nonzero PARSE radii were considered as charge centers in assembling the necessary matrices used in component analyses.

3.2. FDM calculations

A multigrid finite-difference numerical solver of the Poisson equation [2] was used to solve for the bound and unbound state potentials needed to assemble the relevant matrices used for component analyses. A 1.4-Å probe radius was used to define the molecular surface for the dielectric boundary. Potentials were solved on a $257 \times 257 \times 257$ uniform rectilinear grid, using a three-stage focusing procedure in which the structure occupied first 23 percent, then 92 percent, and finally 184 percent of the grid along the longest dimension, centering on the charged atom in the final case. The grid resolution at the highest focusing was 8.7 grids/Å. To account for the sensitivity of the calculated potentials to the placement of the grid, potentials were computed for three slight translations of the grid (translated approximately 0.5 Å relative to each other), and their average values were used.

3.3. BEM and BIBEE calculations

All BEM and BIBEE calculations were performed using the FFTSVD fast solver [3, 5]. Piecewise-constant basis functions were used to approximate the induced surface charge, and the qualocation discretization approach was used to define the BEM linear system [4, 10, 12, 17]. The required molecular-surface discretizations (surface meshes) consisted of planar triangles, which were generated using MSMS [74, 75] with a specified vertex density of 6.0 vertices/Å^2. We verified that this resolution was adequate by comparing the reaction-potential matrices for BPTI and trypsin computed at vertex densities of 4.0 and 8.0 vertices/Å^2.

3.4. Model system calculations

To systematically analyze model performance, spherical and ellipsoidal model systems were used (see Figure 2 for the sphere geometry) for which the complex solvation energy ΔG^{solv}, the ligand desolvation matrix L, and the interaction vector Cq_R could be computed analytically using the appropriate harmonics. The latter two terms allow for the quantification of the ligand-charge-dependent portion of binding free energies as well as for complete component analysis of ligand charge groups. We define the "ligand-dependent binding free energy", $\Delta G^{*,\text{bind}}$, to be the sum of the ligand desolvation penalty and the interaction terms:

$$\Delta G^{*,\text{bind}} = q_L^T L q_L + q_L^T C q_R \tag{27}$$

$\Delta G^{*,\text{bind}}$ is the total electrostatic binding free energy excluding the receptor desolvation penalty. The efficient calculation of $\Delta G^{*,\text{bind}}$ using model geometries of ligand and complex enables the charge locations and values on either binding partner to be studied thoroughly via random sampling, in order to understand the relationship between the electrostatic properties of the binding partners and their accurate modeling using the different BIBEE methods. For spheres, multipoles up to order 35 were used; for ellipsoids, calculations were truncated at order 10, because available algorithms for calculating ellipsoidal harmonics suffer numerical accuracy issues for higher orders [14]. We note that for a charge r away from the center of a sphere of radius a, truncation at order N leads to an error proportional to $(r/a)^N$; consequently, the minimum order needed for a given tolerance is problem specific.

Charge distributions were randomly generated within each binding partner. In the receptor cavity, 20 charges were used. Within each ligand, we defined four randomly-generated "residues," which were each assigned one of eight simple geometries: a line consisting of 3 charges in each of the three ordinal directions, a square, a triangle in each of 3 different planes, or a cube. Each residue consisted of charges that were allowed to vary from 0.85e to -0.85e. For generality in the analyses, residues were not constrained to have an overall integral charge. For spherical systems, all charges within either ligand or receptor were at least 1.4 Å away from the cavity boundary and from each other; for ellipsoidal geometries, all charges were constrained to lie within the dielectric cavity and to be at least 1.4 Å away from each other. While quantitative values sometimes differed, important trends were found to be robust to the specific details of the model systems (such as the number of charges on either partner, the minimum distance allowed between charges, and the anisotropy of the ellipsoidal geometries), the order of expansion of the analytical "exact" answer, and the number of trials used to compute relative RMSEs.

3.5. Matrix manipulations, figure generation, and calculation of error

Matlab [1] was used to perform all matrix calculations for component analysis, for generating all plots, for generating model systems, and for calculations on model systems using the analytically exact Kirkwood method [45] and analytical versions of the approximate BIBEE model [13]. To avoid magnifying small absolute differences between the methods, we excluded all data points whose magnitudes were less than 1 kcal/mol (via the most accurate method) in calculating relative RMSEs, for both the model and the biological systems, unless otherwise noted. Relative RMSE values for a vector quantity x were defined as follows:

$$\text{rel. RMSE} = \sqrt{\frac{\sum_n \left(\frac{x_{approx} - x_{exact}}{x_{exact}}\right)^2}{n}} \tag{28}$$

and all reported r values are standard Pearson's correlation coefficients.

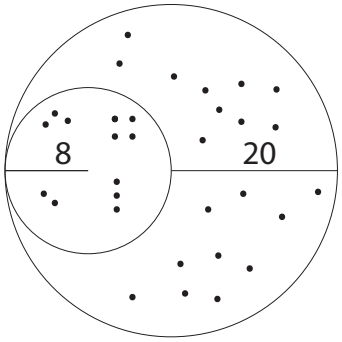

Fig 5. Schematic depicting a spherical ligand binding to a receptor to create a spherical complex. Radii are given in Angstroms. Charge distributions were randomly generated, with those on the ligand being grouped into four model "residues." An analogous model system was generated for a tri-axial ellipsoidal ligand binding to a receptor to form an ellipsoidal complex. For the analyses in which geometry was held fixed and charge distributions were randomly varied, we used ligand axes of 7.6, 6.5, and 4.4 Å, and receptor axes of 19, 18, and 13 Å. Relative to the complex center, the ligand is translated by [4.75 1.48 2.98] and rotated by [0.839 -0.535 0.098; -0.441 -0.775 -0.451; 0.318 0.336 -0.887].

4. Component Analysis for Protein–Protein Binding

4.1. Prediction of residue-based contributions in the trypsin–BPTI system

Figure 6 summarizes the performance of the BIBEE methods in a component-analysis application for the trypsin–BPTI system. All comparisons are made to high-resolution finite-difference method (FDM) numerical solutions of the Poisson equation. Figure 1 in the Supplemental Information compares the numerical convergence of both the FDM grid and the panel discretization used for the BIBEE methods, plotting $\Delta\Delta G^{\text{bind}}$ for all residues within BPTI and trypsin when calculated by BEM and FDM solutions to the Poisson equation. The comparison shows generally very good agreement, although the agreement is better for trypsin than for BPTI. Recall that the $\Delta\Delta G^{\text{bind}}$ is a relative binding free energy between the two partners when a given residue has all of its atomic charges set to zero and when its charges are at their original values, and therefore it quantifies the importance of a residue's charge distribution toward binding. Both methods clearly identify "outliers" for BPTI and trypsin — residues that contribute far more favorably toward binding than any other residues, as zeroing out their charge distributions greatly worsens binding. The residues with the highest $\Delta\Delta G^{\text{bind}}$ when calculated via either method are the canonical "specificity-determining residues" of Lys515 on BPTI and Asp189 on trypsin, demonstrating their importance in this system and in the utility of component analysis in identifying key "hot spot" residues in binding interactions. Points corresponding to these residues are labeled in Figure 6(a) and (b). As seen in Figure 6, none of the approximate methods shows highly quantitative agreement, although BIBEE/CFA shows good qualitative and modest quantitative agreement with the FDM values for both trypsin and BPTI. BIBEE/P incorrectly predicts the sign of $\Delta\Delta G$ for charged residues (red and green points in Figure 6(b) and (f)), meaning that it predicts residues to contribute favorably when the actual Poisson model predicts the opposite. However, BIBEE/P is quantitatively accurate for polar and hydrophobic residues (RMSE=79.74%, even when using a more stringent cutoff of 0.1 kcal/mol, rather than 1 kcal/mol used elsewhere, for inclusion in the rel. RMSE calculation (SI Figure 2)). These results suggest that BIBEE/P is better suited for modeling systems that do not involve changes in overall monopole. Finally, of particular note is BIBEE/I, which, although it had previously shown excellent average accuracy for ΔG^{solv} on numerous biological systems [13], does not perform accurately in this stringent application, which combines both the compounding errors of $\Delta\Delta G^{\text{bind}}$ and the irregular molecular shapes inherent to actual biological systems.

Taken together, these mixed results on a biological system led us to look more deeply into the approximate methods to better understand both the types of systems (charged, polar, etc.) in which each performed well or poorly, as well as the potential assumptions that contributed to their varying performances on this test system. For example, it was intriguing that BIBEE/P performed so poorly for predicting the contributions of charged residues but appeared to offer semi-quantitative accuracy for polar but uncharged groups. Below, we describe the analyses that followed on simplified geometries, and the insights we gained about the existing models that may lead to substantially improved models that are more suitable for application to biological systems.

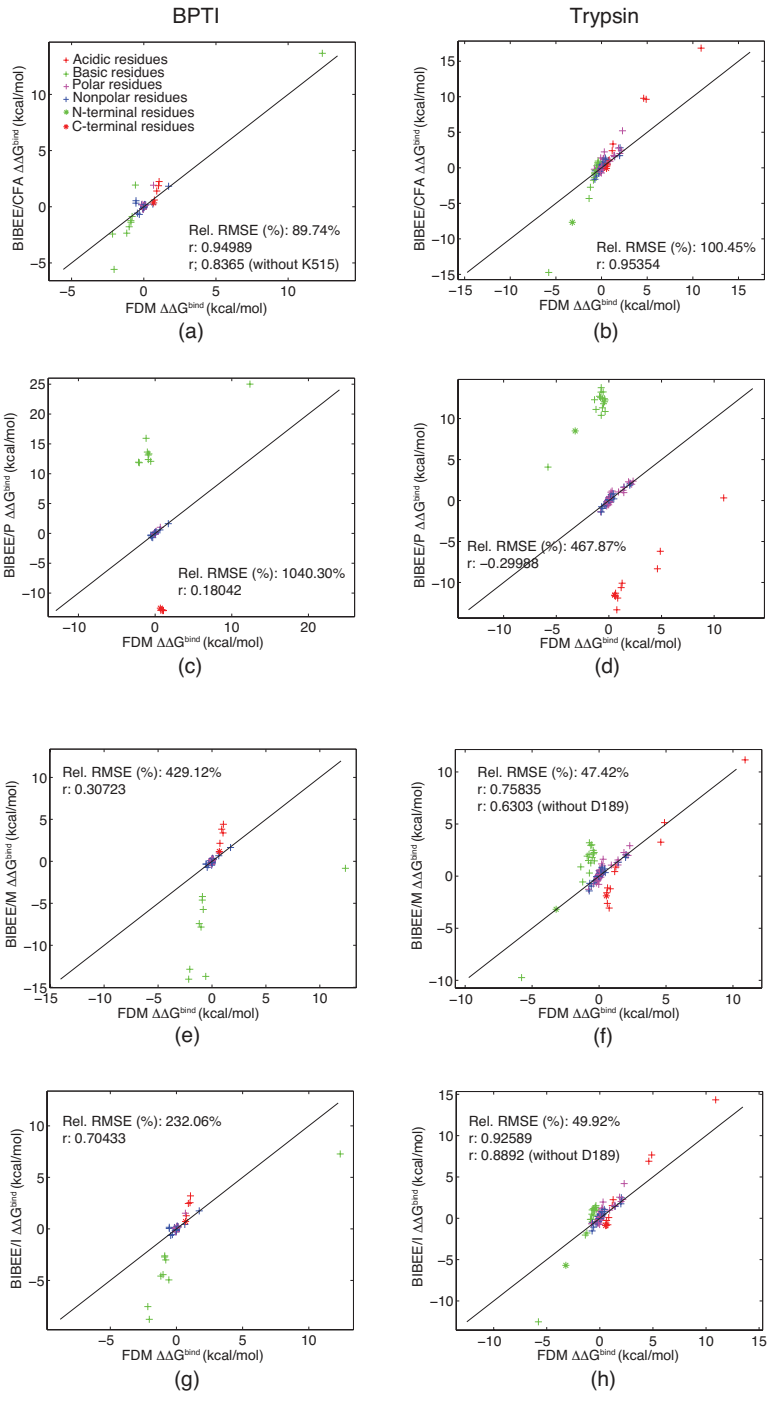

Fig 6. Comparison of residue-based contributions ($\Delta\Delta G^{\text{bind}}$) obtained using more approximate models to those obtained using the finite-difference method (FDM) solution of the Poisson Equation. Positive values of $\Delta\Delta G^{\text{bind}}$ indicate a residue that contributes favorably to binding, because setting all charges on the residue to zero worsens binding. Points on each graph are colored according to the legend in panel (a). Residues known to be crucial for molecular recognition in this system are labeled in panels (a) and (b). Approximate methods shown here are BIBEE/CFA (panels a and b), BIBEE/P (panels (c) and (d)), BIBEE/M (panels (e) and (f)), and BIBEE/I using an effective eigenvalue $\lambda^* = -0.2$ (panels (g) and (h)). Results for all residues on BPTI (left) and trypsin (right) are shown separately. Relative RMSE's exclude points with a magnitude of less than 1 kcal/mol when calculated using FDM; additional r values are given for cases in which outliers significantly perturb the quality of approximation.

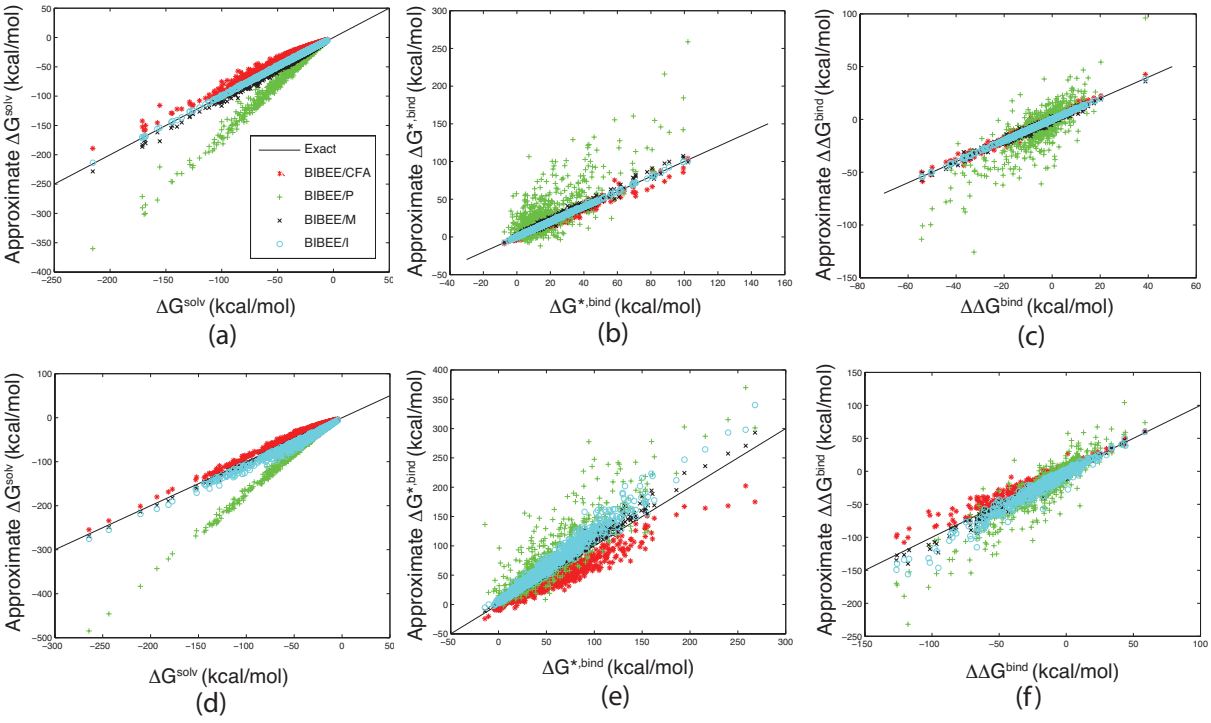

Fig 7. Comparison of approximate BIBEE methods in predicting complex ΔG^{solv} (Figs. 7(a) and (d)), the ligand-dependent $\Delta G^{*,\text{bind}}$ ((b) and (e)), and the $\Delta\Delta G^{\text{bind}}$ upon setting the charges within a random ligand residue to zero ((c) and (f)) for randomly generated instances of the spherical model system ((a)-(c)) and the ellipsoidal model system ((d)-(f)) described in the Methods. One thousand randomly generated sets of charge locations and values within the model system geometry were used as a set of model binding complexes. In all cases, approximate energies are compared with analytically calculated "exact" values using a high-order multipole expansion. The legend shown in (a) is applicable to all panels.

4.2. Prediction of residue-based contributions in model geometries

Figure 7 is a comparison of each of the BIBEE methods in predicting particular energetic quantities for randomly generated instances of the spherical model system ((a), (b), and (c)) and the ellipsoidal model system ((d)-(f)) described in the Methods. For each geometry, the quantities compared are the calculated complex ΔG^{solv} (Figs. 7(a) and (d)), the ligand-dependent $\Delta G^{*,\text{bind}}$ ((b) and (e), described in Methods), and the $\Delta\Delta G^{\text{bind}}$ ((c) and (f)) for determining the contribution of a randomly-selected model "residue" component within the ligand, by comparing the binding free energy when the residue is charged to when all charges on it are set to zero (see Methods). One thousand randomly generated sets of charge locations and values within the model system geometry were used as a set of model binding complexes. The spherical and ellipsoidal ligand and complex geometries remained constant throughout all trials; only charges varied. In all cases, approximate energies are compared with analytically calculated values using a high-order multipole expansion.

For all methods and both geometries, there is good correlation for ΔG^{solv} (r values ≥ 0.99 in all cases). In agreement with theory [15], BIBEE/CFA systematically overestimates ΔG^{solv} because the CFA entails approximating all of the integral-operator eigenvalues by the known extremal eigenvalue (Figure 3). Conversely, BIBEE/P systematically underestimates ΔG^{solv} because the approximation employs the other limit for a sphere [15]. By construction, BIBEE/M provides a more accurate approximation than BIBEE/P while still underestimating the exact answer, because it exactly treats the monopole response (i.e. the zero mode) while underestimating the other modes as BIBEE/P does. Finally, BIBEE/I very accurately predicts ΔG^{solv} for the spherical system, with an rel. RMSE of only 3% (Table 1), by assuming an approximate eigenvalue for modes beyond the monopole. However, the BIBEE/I model performed far worse for the ellipsoidal model system (rel. RMSE of 29%); we note that for the sphere we employed $\lambda^* = -0.12$ in the BIBEE/I model, which is

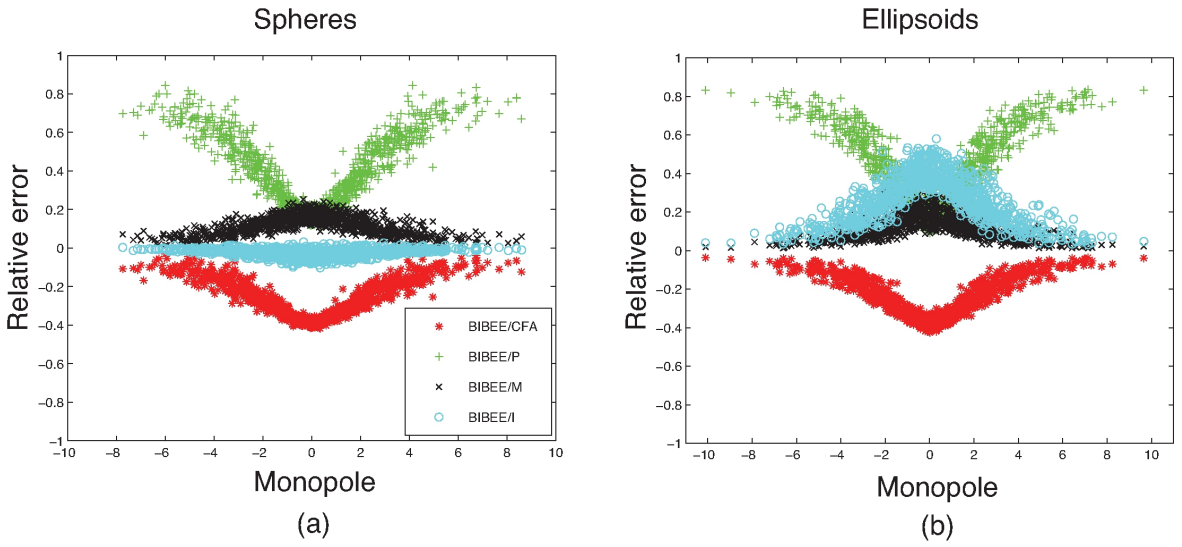

Fig 8. The accuracy of the original BIBEE approximations depends in predictable ways on the overall system monopole and the details of the BIBEE approximation. Relative errors in solvation free energies are plotted for 1000 random charge distributions in (a) spheres and (b) ellipsoids, with the legend in (a) applying to both panels.

in between the dipole eigenvalue $-1/6$ and the quadrupole eigenvalue $-1/10$, whereas for the ellipsoidal system, we employed $\lambda^* = -0.20$, because this value gave the best overall accuracy in earlier work on (non-spherical) protein systems [13] and was the value used in our modeling of trypsin-BPTI. As a control to compare directly to the spherical system results, we also ran a separate series of 1000 trials for the ellipsoidal model system in which $\lambda^* = -0.12$ for BIBEE/I. In this case, the performance was actually slightly better (rel. RMSE = 23% for solvation free energies) but still robustly worse than its performance for spherical systems.

Both ΔG^{bind} and $\Delta\Delta G^{\text{bind}}$ require taking differences of solvation energies, and so any systematic bounds seen in the spherical system need not hold (Figure 7, panels (b) and (c)), although BIBEE/CFA generally appears to underestimate $\Delta G^{*,\text{bind}}$, while BIBEE/P and BIBEE/M generally overestimate it for both spherical and ellipsoidal systems. This "reversal" in the bias of the methods in going from solvation free energies to binding free energies is interesting, and a potential subject of future analysis. Nevertheless, Figure 7 demonstrates that systematic bounds that hold for solvation free energies in physics-based approximate methods may not hold when calculating other quantities of biological interest. As shown in Table 1, BIBEE/I consistently outperforms the other original BIBEE methods (CFA, P, and M) for the spherical model system, while BIBEE/P consistently shows the poorest relative accuracy. In the ellipsoidal system, BIBEE/M shows the best performance when λ^* was set to either -0.12 (data not shown) or -0.20 in the BIBEE/I model. However, in all cases, the relative RMSE increases as one proceeds from evaluating ΔG^{solv} to $\Delta G^{*,\text{bind}}$ to $\Delta\Delta G^{\text{bind}}$, regardless of the precise BIBEE approximation. These results, even with simplified geometries, demonstrate the importance of evaluating a model's accuracy using multiple benchmarks in order to understand the model-specific propagation of error in obtaining quantities of biological importance that involve taking the differences of approximate values. We note that these trends observed for RMSE are likely to reverse if one considers absolute error and not relative error. Relative error is a more reasonable choice here because the magnitudes of solvation, binding, and mutational energies are generally quite different.

4.3. Influence of molecular monopole on BIBEE approximations

Figure 8 shows the percent error in the computed complex ΔG^{solv} values for each of the thousand randomly generated charge distributions in the sphere and ellipsoid model geometries analyzed above, as a function of the monopole (net charge) of the complex. In these calculations, monopoles were not constrained to be integral. For ease of visualization, points were excluded if $|\Delta G^{\text{solv}}|$ was less than 1 kcal/mol, because small absolute deviations in ΔG values create very large relative errors.

Figure 8 shows that BIBEE/CFA, BIBEE/M, and BIBEE/I become increasingly accurate as the magnitude of the monopole increases, with large overestimation of ΔG^{solv} for zero-monopole complexes and much smaller errors for complexes with large magnitude monopoles. Such a result is likely a consequence of BIBEE/CFA's exact modeling of the highest-magnitude eigenvalue of the solvation matrices for spherical systems, i.e. the component of solvent response that is dictated solely by the monopoles of the species involved. Conversely, BIBEE/P shows complementary behavior; it is most accurate when the monopole is zero or near zero and becomes extremely inaccurate for larger magnitude monopoles. This phenomenon results from BIBEE/P's increasingly accurate modeling of the smaller eigenmodes of solvation, which are associated with higher-order multipole contributions toward the solvation energy [13]. As before, while BIBEE/CFA tends to overestimate the ΔG^{solv} for zero-monopole species, BIBEE/M tends to underestimate. The different performance of BIBEE/I between the systems may be attributed to the use of different optimized λ^* (-0.12 for the sphere and -0.20 for the ellipsoid [13]), but the overall poorer accuracy of BIBEE/I for component analysis of trypsin/BPTI and for ellipsoidal systems (using either value of λ^* in the latter case) suggests that detailed analysis of this point is of secondary importance.

The observed solvation free energy results for a sphere may be analyzed as follows. We write the eigendecomposition of the exact reaction-potential matrix as $A = V\Lambda V^T$, where of course V is orthonormal and Λ is diagonal. For the sphere, the BIBEE approximate reaction-potential matrices have the same eigenbasis V as A, as shown previously [13], and therefore we may write $\hat{A} = V(\Lambda + E)V^T$ for any BIBEE model, where E is also diagonal and represents the error in the reaction-potential eigenvalue. The exact solvation free energy can be written as a sum over modes

$$\Delta G^{solv} = \sum_i \lambda_i (V_i^T q)^2 \tag{29}$$

where V_i represents the ith eigenvector, and an approximate solvation free energy is similarly written

$$\Delta G^{solv}_{approx} = \sum_i (\lambda_i + e_i)(V_i^T q)^2. \tag{30}$$

Writing $q_i = V_i^T q$ so that q_1 is the net monopole, the relative error is then

$$\text{Rel. error} = \frac{e_1 q_1^2 + e_2 q_2^2 + e_3 q_3^2 + \ldots}{\lambda_1 q_1^2 + \lambda_2 q_2^2 + \lambda_3 q_3^2 + \ldots}. \tag{31}$$

For the BIBEE/CFA model, $e_1 = 0$ and so one expects that the relative error decays to zero as the monopole magnitude increases, as seen in Figure 7. For BIBEE/P, $e_1 = \hat{\epsilon}/(2 - \hat{\epsilon})$, and with the given dielectric constants the relative error should approach 90% with increasing monopole, which it does; we have also verified that the relative error asymptotically approaches correct values for other values of the dielectric constants. In the BIBEE/M model, $e_1 = 0$ and all the other terms e_i are equal to those from BIBEE/P. As in the CFA model, zero error in the monopole term implies an inverse relationship between the net monopole and relative error, which can also be observed. Last, in BIBEE/I we have $e_1 = 0$ and also the parameter λ^* has been set to approximately zero out the other error terms [13]. As a result, the method has very small relative error even for small net monopoles, and again must approach zero with increasing monopole.

As a whole, the preceding results allow us to better place into context the results of our component analyses on the trypsin/BPTI system. For example, as we have seen that BIBEE/P systematically becomes less accurate as the monopole magnitude of the species under consideration increases, we can understand why it performed so poorly for charged residues on trypsin and BPTI but did reasonably well for polar uncharged trypsin residues. Likewise, given the poor performance of BIBEE/I on ellipsoidal systems relative to spherical systems, especially in computing relative binding free energies on the model systems (rel. RMSE for $\Delta\Delta G_{bind}$ is 100% for ellipsoids with but only 9% for spheres), we can understand the relatively poor performance of BIBEE/I in a component analysis application on an irregularly shaped molecule. To control for the different values of λ^* used for BIBEE/I for the sphere and ellipsoidal systems as mentioned above, we also carried out separate analyses on 1000 trials of the ellipsoid using the spherical system $\lambda^* = -0.12$, and we found the rel. RMSE for $\Delta\Delta G_{bind}$ still to be significantly higher for the ellipsoid (71%, data not shown).

5. A Reduced-Basis Interpretation Approach to Improving BIBEE Models

As Figure 3 illustrates, the first BIBEE models (BIBEE/CFA and BIBEE/P) replaced all of the integral-operator eigenvalues with one extremal value or another [9]; this allowed, after some analysis, the proof that BIBEE/CFA is always an upper bound, and BIBEE/P is a lower bound for some boundaries such as spheres [15]. These models are purely diagonal approximations to the boundary-integral operator (in the discretized form, the diagonal entries of the BEM matrix). The simplicity and speed advantages associated with inverting diagonal matrices is merely a red herring: a more fruitful interpretation for these models is that they approximate the eigendecomposition of the integral operator,

$$\mathcal{D}^* = V\Lambda_{\mathcal{D}^*}V^{-1} \tag{32}$$

with $\Lambda_{\mathcal{D}^*} = \lambda I$, so that $\mathcal{D}^* = \lambda I$. In general, the electric-field operator $\mathcal{D}^*$ is not symmetric and $V^{-1} \neq V^T$; we do, however, have the Calderon identity $\mathcal{G}\mathcal{D}^* = \mathcal{D}\mathcal{G}$, where $\mathcal{G}$ is the (symmetric) single-layer operator.
We begin our discussion using the special case of a spherical boundary so that $\mathcal{D}^*$ actually *is* symmetric, with $\mathcal{D}^* = -\frac{1}{2R}\mathcal{G}$ where R is the sphere radius [41]. Then we can write

$$\mathcal{D}^*_{\text{sphere}} = V\Lambda V^T, \tag{33}$$

where V are the surface spherical harmonics, and the diagonal entries of Λ are the corresponding eigenvalues of the integral operator: for $n \geq 0$,

$$\lambda_{nm} = -\frac{1}{2(2n+1)}. \tag{34}$$

Writing Eq. (16) as $(I + \hat{\epsilon}\mathcal{D}^*)\sigma(\mathbf{r}) = f(r)$, we can express the exact surface charge as

$$\sigma(\mathbf{r}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{+n} Y_n^m(\mathbf{r})\left[(1+\hat{\epsilon}\lambda_{nm})^{-1}\int Y_n^{m,*}(\mathbf{r})f(\mathbf{r})dA\right] \tag{35}$$

so the expression in square brackets represents the expansion coefficient for $\sigma(\mathbf{r})$ in the eigenfunctions of $\mathcal{D}^*$.
From this viewpoint, it is simple to test how approximating the operator spectrum affects the overall approximation accuracy. The BIBEE/I variant, for example, attains good accuracy for solvation free energies but does so at the expense of exhibiting the wrong asymptotic behavior for small eigenvalues (Figure 3). With Eq. (35), we may investigate whether a "three eigenvalue" approximation would be more accurate still, employing the exact operator eigenvalue $\lambda_{00} = -1/2$ for the monopole, one approximate eigenvalue λ^* for some dominant modes (say, dipoles and quadrupoles), and approximating all remaining modes using the exact asymptotic value 0. The surface charge for this variant, which we term truncated BIBEE/I, is

$$\hat{\sigma}(\mathbf{r}) = \underbrace{Y_0^0(\mathbf{r})\left[(1+\hat{\epsilon}\lambda_{00})^{-1}\int Y_0^{0,*}(\mathbf{r})f(\mathbf{r})dA\right]}_{\text{exact monopole}} + \underbrace{\sum_{1\leq n\leq k}\sum_{m=-n}^{n} Y_n^m(\mathbf{r})\left[(1+\hat{\epsilon}\lambda^*)^{-1}\int Y_n^{m,*}(\mathbf{r})f(\mathbf{r})dA\right]}_{\text{approximate dipole, quadrupole terms with } \lambda_{nm}\approx\lambda^*} + \underbrace{\sum_{n>k}\sum_{m=-n}^{n} Y_n^m(\mathbf{r})\left[(1+\hat{\epsilon}\cdot 0)^{-1}\int Y_n^{m,*}(\mathbf{r})f(\mathbf{r})dA\right]}_{\text{approximate higher-order terms with } \lambda_{nm}\approx 0}. \tag{36}$$

Naturally, a more accurate approach would be to use, for each of some number of dominant modes, the actual corresponding eigenvalue, and then approximating the rest by 0, so that

$$\hat{\sigma}(\mathbf{r}) = \underbrace{\sum_{n\leq k}\sum_{m=-n}^{n} Y_n^m(\mathbf{r})\left[(1+\hat{\epsilon}\lambda_{nm})^{-1}\int Y_n^{m,*}(\mathbf{r})f(\mathbf{r})dA\right]}_{\text{All dominant modes treated exactly}} + \underbrace{\sum_{n>k}\sum_{m=-n}^{n} Y_n^m(\mathbf{r})\left[(1+\hat{\epsilon}\cdot 0)^{-1}\int Y_n^{m,*}(\mathbf{r})f(\mathbf{r})dA\right]}_{\text{Higher-order terms approximated with } \lambda_{nm}\approx 0}. \tag{37}$$

Figure 3 provides a schematic of the reduced-basis BIBEE ideas. Also, it is important to note the difference between the above approximation and a mere truncation of the infinite sum over multipoles, which would give only the first summation in Eq. (37), i.e.

$$\hat{\sigma}(\mathbf{r}) = \underbrace{\sum_{n\leq k}\sum_{m=-n}^{n} Y_n^m(\mathbf{r})\left[(1+\hat{\epsilon}\lambda_{nm})^{-1}\int Y_n^{m,*}(\mathbf{r})f(\mathbf{r})dA\right]}_{\text{All dominant modes treated exactly}} + \sum_{n>k}\sum_{m=-n}^{n} Y_n^m(\mathbf{r})\left[\underbrace{(1+\hat{\epsilon}\lambda^*)^{-1}}_{=0}\int Y_n^{m,*}(\mathbf{r})f(\mathbf{r})dA\right]. \qquad (38)$$

A different way to think about the errors due to truncation is to consider that the higher modes would be modeled as if $\lambda^* = \infty$, when in fact it is known that $|\lambda_{nm}| \leq 1/2$.

As noted above, however, in general $\mathcal{D}^*$ is not symmetric, which greatly complicates calculating the expansion of the normal electric field at the boundary into the operator eigenfunctions. Currently, we can only test the reduced-basis BIBEE idea using ellipsoids, building on work by Ritter [67], who has shown that the surface ellipsoidal harmonics, which obey important orthogonality relations, are in fact eigenfunctions of $\mathcal{D}$, and closely related functions are eigenfunctions of $\mathcal{D}^*$. Then, similarly to the sphere, we may calculate the BIBEE approximation by expanding the potential in harmonics and employing the appropriate λ_{nm} or λ^* for each term as desired. In this way, we can conduct a preliminary test of the accuracy of the two reduced-basis BIBEE methods without the computational expense of actually computing the truncated SVDs of the dense matrices. Ellipsoids offer an important test not available in spheres, which is the fact that for a given expansion order n, the operator eigenvalues λ_{nm} are not all equal (contrast Eq. (34)). It is unfortunate that our tests of this model are limited presently to the sphere and ellipsoid, but ongoing work is focused on testing reduced-basis BIBEE for atomistic protein surfaces.

5.1. Application to Reaction-Potential Matrices

To estimate the performance of reduced-basis BIBEE models for computing reaction-potential matrices, we created an ensemble of ten random ellipsoids, each of which contained 100 random charges. The three semi-axes for each ellipsoid were assigned random values from a uniform distribution with mean 16 Å and width 6.4 Å, corresponding to a variation up to ± 20%; allowing larger relative variations did not qualitatively affect the results. In each ellipsoid, the 100 charges were picked at random to be at Cartesian lattice sites where the lattice spacing was 1.4 Å, and charges were required to lie within the ellipsoidal surface.

The plots in Figure 9 are of the mean absolute errors between the eigenvalues of the reaction-potential matrix, and the corresponding eigenvalues of three BIBEE approximations, for spheres and ellipsoids; error bars denote the standard deviations. Figure 9(a) contains the results for BIBEE/M and BIBEE/I for the sphere, as well as the reduced-basis truncated form of BIBEE/I; Figure 9(b) is a plot of the same results but zoomed in on the dominant modes. Figures 9(c) and (d) are plots of the same calculations for the random ellipsoids, showing two similar qualitative behaviors, and one noticeable deviation from the sphere results. The first similarity is that for smaller-magnitude eigenvalues, the BIBEE/I eigenvalues exhibit larger deviations from the exact ones than do the BIBEE/M or the truncated BIBEE/I. We attribute this phenomenon to the fact that BIBEE/I uses inaccurate values for the smallest-magnitude eigenvalues of the integral operator, whereas BIBEE/M and truncated BIBEE/I use the correct asymptote (Figure 3). Second, we obtain the expected result that the truncated BIBEE/I produces results similar to BIBEE/I for the dominant modes, and results similar to BIBEE/M for the smaller modes. The noticeable deviation between the sphere and ellipsoid results is that in the sphere case, BIBEE/M is less accurate for the dipole and quadrupole modes than are the BIBEE/I and truncated BIBEE/I modes; however, in the ellipsoid case BIBEE/M is more accurate.

Figures 10(a) and (b) include plots of the BIBEE/CFA, BIBEE/P, and the other reduced-basis approaches (up to dipole, quadrupole, and octopole terms; the monopole approach is BIBEE/M, which is plotted in Figure 9) for the sphere, employing the actual eigenvalues of the ellipsoid integral operator [67–69]. As in the previous Figure, in (a) are plotted the errors across the entire spectrum, and (b) is a zoomed-in plot on the dominant modes. Figures 10(c) and (d) are plots of the same models applied to random ellipsoids. Unsurprisingly, treating more modes exactly produces more accurate answers. What is surprising, however, is that only quadrupole modes are needed to estimate all reaction-potential eigenvalues to better than 2 kcal/mol/e^2 accuracy, which corresponds to a relative accuracy of better than 0.1% for the monopole, better than 1% for the dipole modes, a few percent for the quadrupole modes, approximately 10% for all other modes. This represents a reduced-basis approximation of $\mathcal{D}^*$ with 9 basis vectors. Adding the octopole modes

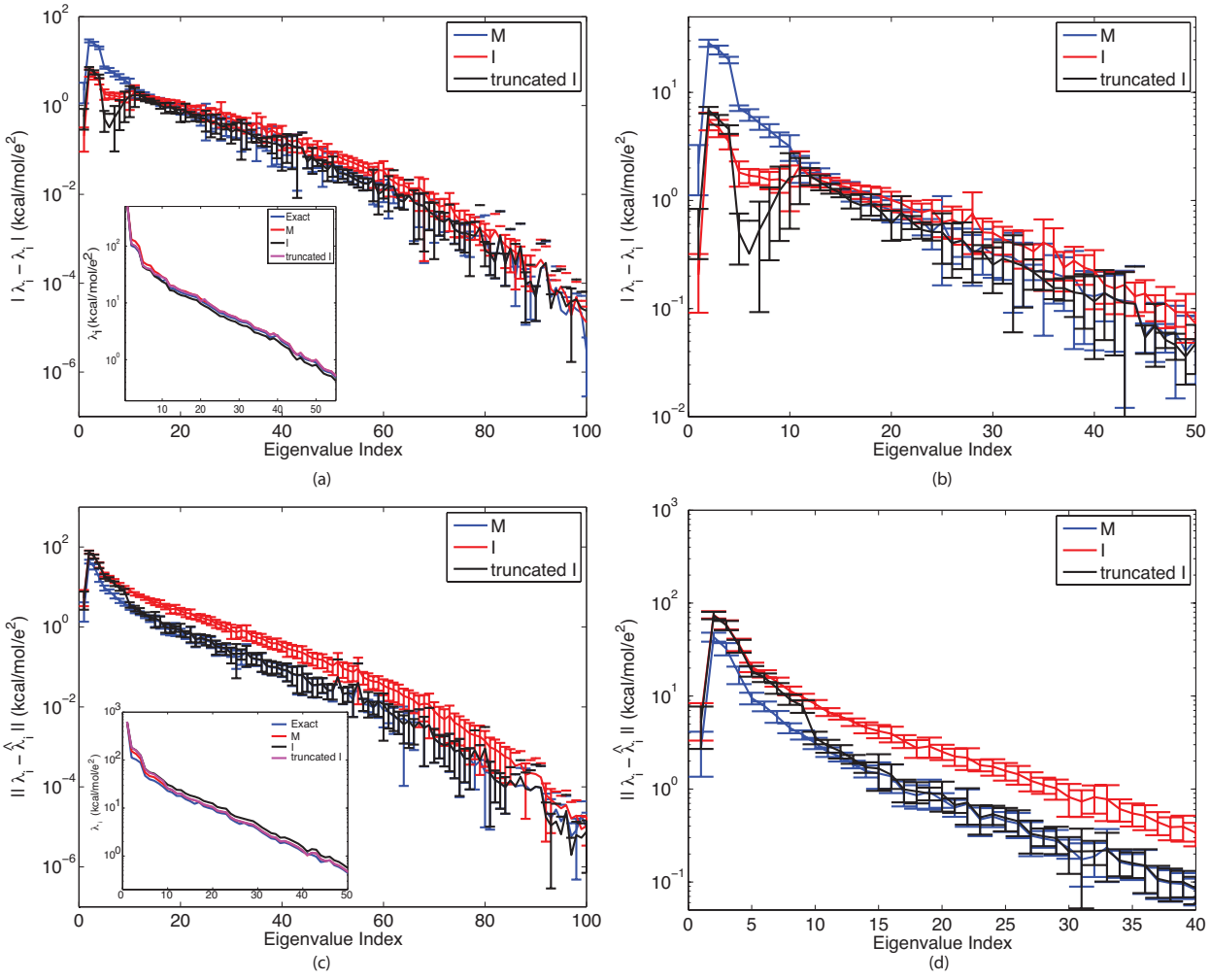

Fig 9. Average absolute errors in the eigenvalues of reaction-potential matrices when approximated with the BIBEE/M, BIBEE/I, and truncated BIBEE/I variants, on an ensemble of 10 random charge distributions in a 16-Å sphere ((a) and (b)) and in random ellipsoids with mean axis lengths of 16 Å ((c) and (d)).

leads to estimates that are accurate to within 1 kcal/mol/e^2, with the monopole and dipole modes estimated to within 0.4 kcal/mol/e^2 (the relative accuracy approaches 0.01% for the monopole). We also note the expected result that adding more modes leads more modes to be estimated accurately; notice the distinct increases in errors for BIBEE/M after the monopole mode (index 1), for BIBEE/dipole after the dipole modes (index 4). SI Figure 3 is a plot of the projections of the approximate reaction-potential eigenvectors for BIBEE/dipole onto the exact reaction-potential eigenvectors. The results show almost perfect agreement, and we note that some noise is inevitable due to the previously mentioned difficulties in computing the harmonics to high numerical accuracy. The combination of accurate reproduction of eigenvalues as well as eigenvectors suggests that the reduced-basis BIBEE methods should be able to perform well on component analysis. To further test the accuracy of the interpolation BIBEE and reduced-basis methods on both spherical and ellipsoidal model systems, we performed component analysis on the same 1000 randomly-generated spherical and ellipsoidal systems depicted schematically in Figure 7. The comparison of these approximate models for estimating ΔG^{solv}, $\Delta G^{*,\text{bind}}$, and $\Delta\Delta G^{\text{bind}}$ against the analytical results are shown in Figure 11 and Figure 12. The relative RMSEs are listed in Table 1.

The interpolation BIBEE approaches perform substantially better for the sphere geometries than for the ellipsoids for both $\lambda^* = -0.2$ (data shown) and $\lambda = -0.12$ (data not shown) in the case of ellipsoids, which suggests that the influence

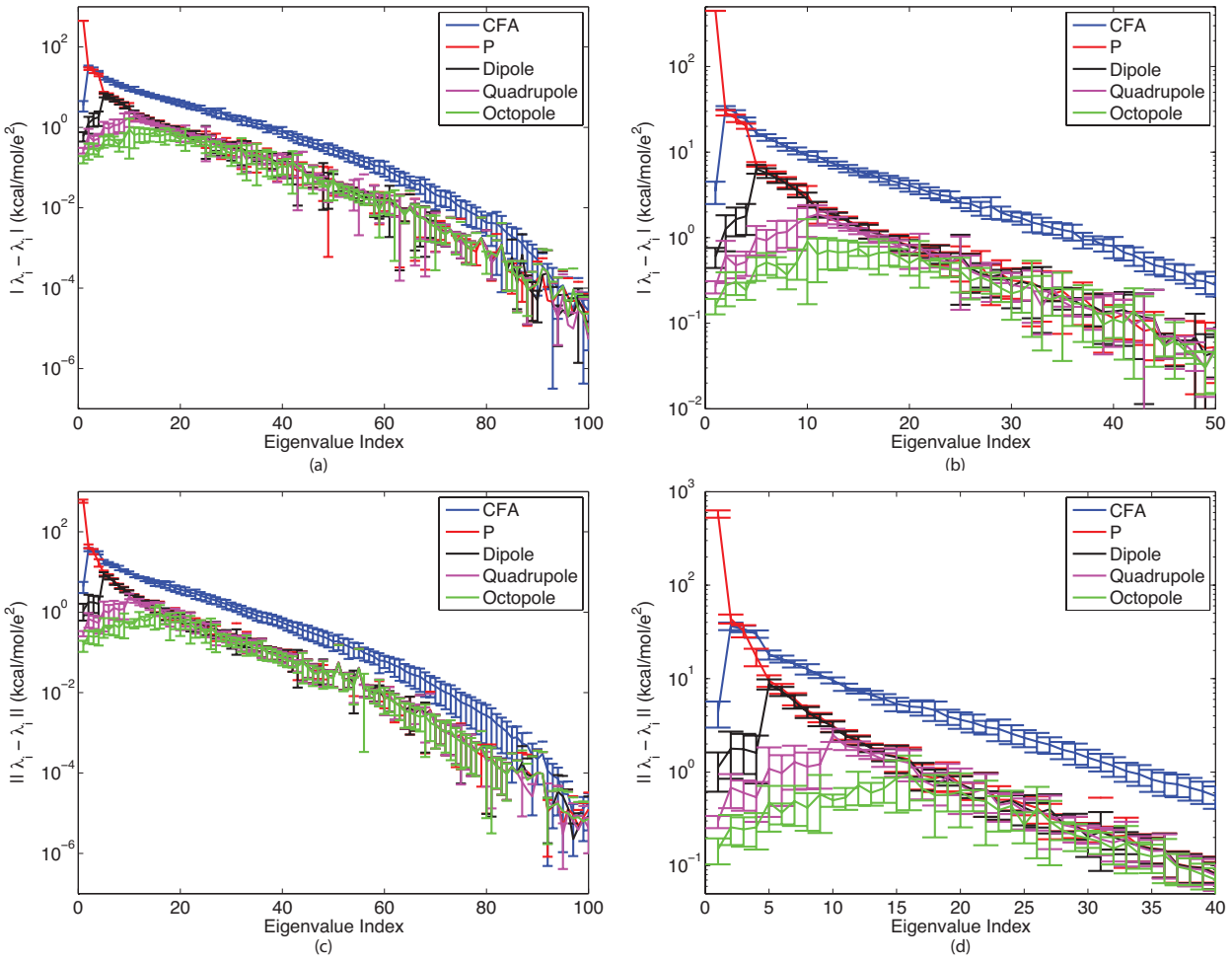

Fig 10. Average absolute errors in the eigenvalues of reaction-potential matrices when approximated with several BIBEE variants, on an ensemble of 10 random charge distributions in a 16-Å sphere ((a) and (b)) and in random ellipsoids with mean axis lengths of 16 Å ((c) and (d)).

of boundary anisotropy should be included wherever possible. Substantial advances in Generalized Born theories have been made previously by analysis of spheres [79], and ellipsoids may offer a new path to further improvements. Truncated BIBEE/I appears to offer minimal improvement. However, the reduced-basis approach provides better accuracy for both geometries, with the accuracy unsurprisingly improving as more modes are included, with the octopole approximation having a relative error of only $2-4\%$ (Table 1). It is interesting that the reduced-basis approximations also suffer in relative accuracy as one proceeds from solvation to binding to relative binding free energies. The persistence of this sensitivity, even when using an "optimal" approximation of the integral operator (these geometries allow us to use the exact reduced singular-value decomposition), supports our claim that component analysis offers a stringent, and therefore useful, test for the accuracy of fast approximate Poisson models.

We may also analyze the dependence of the reduced-basis approximations on the overall system monopole, as in Figure 7. These results, calculated using the same charge distributions, are plotted in SI Figure 4. For spheres, the reduced-basis approaches (in SI Figure 4 (c) and (d)) based on the actual operator eigenvalues are comparable to, or outperform, BIBEE/I and the truncated version ((a) and (b)), and for ellipsoids the reduced-basis approaches are all significantly more accurate. These results therefore indicate that the BIBEE/I approach introduces significant approximation error by assuming that the dominant operator eigenvalues are identical. In other words, even simple shape anisotropy (as the

shape deviates from a sphere to an ellipsoid) can sharply limit the accuracy of approximations that assume a spherical boundary.

We also note that for both the sphere and ellipsoid case, the truncated form of BIBEE/I offers minimal accuracy improvements over the original BIBEE/I. These results contrast with the incorrect asymptotic behavior of the spectrum for the BIBEE/I reaction-potential matrix, versus the corrected behavior for truncated BIBEE/I (Figure 9). Taken together, the results imply that for binding and component-analysis applications, the modes corresponding to smaller eigenvalues where the BIBEE/I and truncated form differ (indices greater than about 15 in Figure 9) contribute negligibly to the error.

The overall accuracy of BIBEE/I and its truncated form is superior for the sphere than the ellipsoid, which we attribute to the fact that in the sphere, the dipole and quadrupole terms contribute little to the overall error because the λ^* can be placed *a priori*. For the ellipsoids, however, a single λ^* fit for a set of proteins does not necessarily capture the anisotropy of the given geometry, and the dipole and quadrupole eigenvalues can be significantly different from the corresponding eigenvalues in the sphere geometry. Therefore, the errors in the ellipsoid BIBEE/I and truncated BIBEE/I are dominated by poor approximation of the dominant multipoles beyond the monopole (which, recall, is treated exactly)—meaning that in practice each ellipsoid requires a different optimal λ^*. On the other hand, it is encouraging that actual reduced-basis examples exhibit about the same approximation properties for both the sphere and ellipsoid examples, because this indicates that shape variation and anisotropy will be captured reasonably well for a range of geometries. In fact, the reduced based approach still showed excellent accuracy on preliminary trials in which ellipsoidal geometries were varied in addition to the charge distribution (data not shown), suggesting that the reduced-basis approach could show good performance across a range of geometries.

6. Discussion

In this paper, we have used electrostatic component analysis, a popular approach for biomolecular analysis and design, to analyze the performance and specific characteristics of recent variants of the BIBEE approach to approximating continuum electrostatics. Component analysis is widely used to identify protein residues that contribute significantly to molecular binding affinity [23], but requires many detailed electrostatic calculations. The atom-by-atom calculation of the Poisson or Poisson–Boltzmann equation is computationally expensive, especially for protein–protein binding, which motivates simple, faster approximations such as Generalized Born methods [23]. In this work, we sought to test whether the latest BIBEE model, known as BIBEE/I, provided accuracy suitable for component analysis, given that it was shown to estimate solvation energies to within a few percent using a test set of hundreds of proteins [13]. However, accuracy for a given problem can be substantially improved by tuning the single parameter in BIBEE/I for the molecular shape of interest, so it remains to be seen how well BIBEE may perform for other classes of molecules, such as nucleic acids. We were disappointed to find that for protein–protein binding, BIBEE/I did not robustly perform better than even the BIBEE/CFA model on the trypsin/BPTI model system (Figure 6). This failure motivated us to analyze the approximation properties of the BIBEE models in more detail, using analytically solvable model problems (spherical and ellipsoidal

Table 1. Relative root mean square error (rel. RMSE), %, for approximately calculated complex solvation free energies (ΔG^{solv}), ligand-dependent binding free energies ($\Delta G^{*,\text{bind}}$) and relative binding free energies $\Delta\Delta G^{\text{bind}}$ in spherical and ellipsoidal model systems, when compared to high-order "exact" multipole expansions. Approximate methods included here are BIBEE/CFA (CFA), BIBEE/P (P), BIBEE/M (M), BIBEE/I (I), the truncated BIBEE/I (TI), and the reduced basis approximations, using up through the dipole (D), quadrupole (Q) and octopole (O). 1000 randomly-generated charge distributions were used for the spherical geometry and ellipsoidal geometry. For robustness, points were excluded from the relative RMSE calculation if their "exact" values had magnitudes of less than 1 kcal/mol.

	Sphere			Ellipsoid		
	ΔG^{solv}	$\Delta G^{*,\text{bind}}$	$\Delta\Delta G^{\text{bind}}$	ΔG^{solv}	$\Delta G^{*,\text{bind}}$	$\Delta\Delta G^{\text{bind}}$
CFA	28	34	46	26	49	87
P	44	248	344	48	164	238
M	13	26	37	14	21	49
I	3	7	9	29	47	100
TI	4	7	9	23	34	86
D	6	8	10	6	10	20
Q	3	3	4	3	5	10
O	2	2	2	2	3	4

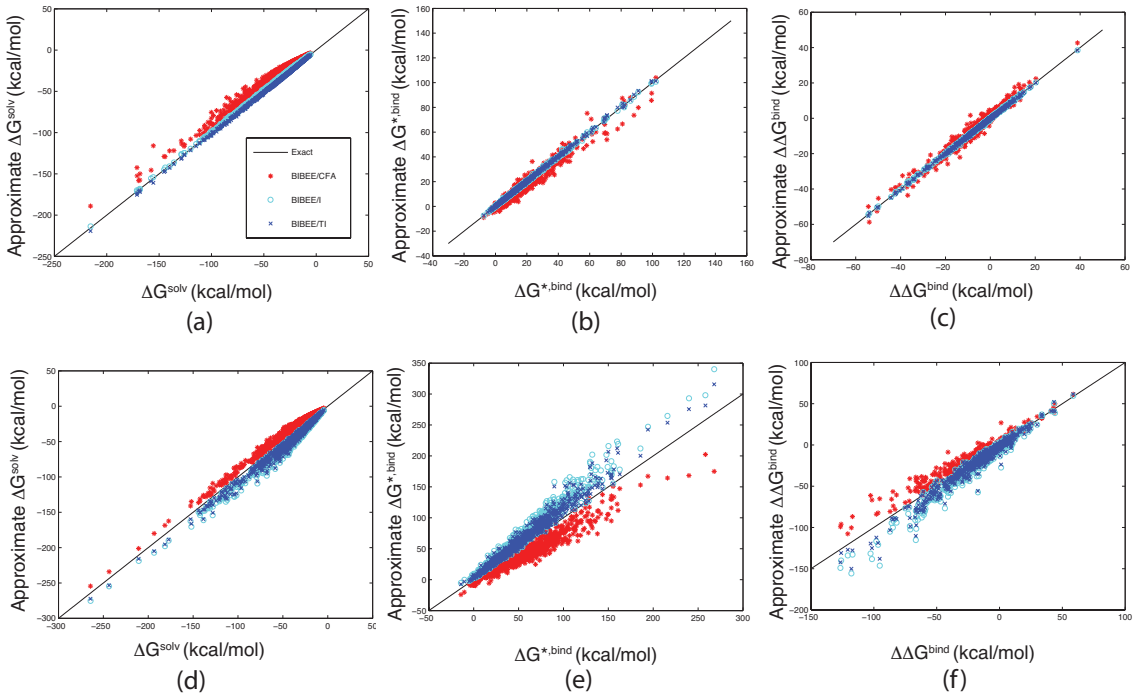

Fig 11. Comparison of interpolation BIBEE-based methods in predicting complex ΔG^{solv} (panels (a) and (d)), the ligand-dependent $\Delta G^{*,\text{bind}}$ ((b) and (e)), and the $\Delta\Delta G^{\text{bind}}$ upon setting the charges within a random ligand residue to zero ((c) and (f)) for randomly generated instances of the spherical model system ((a)-(c)) and the ellipsoidal model system ((d)-(f)) described in the Methods. One thousand randomly generated sets of charge locations and values within the model system geometry were used as a set of model binding complexes. In all cases, approximate energies are compared with analytically calculated "exact" values using a high-order multipole expansion. For comparison, the results using BIBEE/CFA are also shown. The legend in panel (a) is applicable in all panels.

geometries). We were able to explain why the BIBEE/P approximation, which gave very inaccurate results overall, offered reasonable accuracy for neutral residues; the same analysis indicated that the Coulomb-field approximation, and methods built to correct its errors, become increasingly accurate as the system's net charge increases in magnitude. Most biological macromolecules such as proteins and DNA tend to be highly charged, so our finding may offer a partial explanation why such simple electrostatic models can sometimes offer useful accuracy in screening studies and binding analysis, even though numerous critical studies have demonstrated the dangers of assuming that accurate solvation free energies lead to accurate binding free energies [38, 58, 66, 76].

By considering the more recent BIBEE variants, BIBEE/M and BIBEE/I, as approximations to the reduced SVD/eigendecomposition of the boundary-element method matrix, we were able to suggest a possible strategy for improving accuracy. Results on the analytically solvable model systems suggest that reduced-basis BIBEE methods might offer excellent accuracy, but the major challenge is to rapidly estimate this reduced basis. Early work on BIBEE models noted that the dominant eigenvectors of the reaction-potential matrix resembled low-order harmonics [9], even for realistic molecular shapes such as peptides. This suggests that shape approximations such as those used in spectral boundary-element methods [49] could provide a way to estimate reduced-basis approximations rapidly.

Acknowledgments

It is a pleasure to thank R. S. Eisenberg for his support of this collaboration, and to thank Michael Altman, David Green, and Bruce Tidor for useful software. MGK acknowledges partial support from the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357 and the U.S. Army Research Laboratory and the U.S. Army Research Office under contract/grant number W911NF-09-0488. The work of JPB was supported in part by a New Investigator award from Rush University and by the National Institute Of

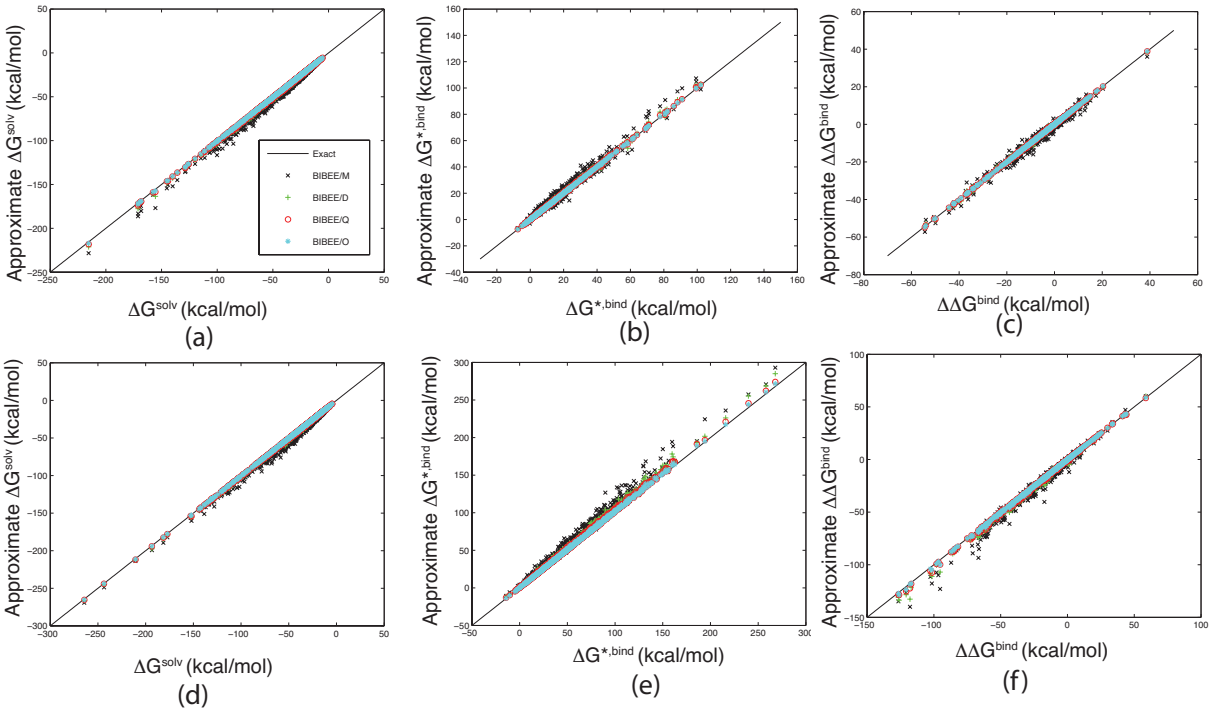

Fig 12. Comparison of reduced-basis approximations in predicting complex ΔG^{solv} (panels (a) and (d)), the ligand-dependent $\Delta G^{*,\text{bind}}$ ((b) and (e)), and the $\Delta\Delta G^{\text{bind}}$ upon setting the charges within a random ligand residue to zero ((c) and (f)) for randomly generated instances of the spherical model system ((a)-(c)) and the ellipsoidal model system ((d)-(f)) described in the Methods. One thousand randomly generated sets of charge locations and values within the model system geometry were used as a set of model binding complexes. In all cases, approximate energies are compared with analytically calculated "exact" values using a high-order multipole expansion. The legend in panel (a) is applicable in all panels.

General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award number R21GM102642. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. LYL and ABK were partially supported by a National Science Foundation Research Experience for Undergraduates Award. LYL, ABK, and MLR were supported by Wellesley College, in part by a Brachman Hoffman Award, and MSM was supported by the Howard Hughes Medical Institute and Wellesley College.

References

[1] Matlab v.6 and R2012b, The Mathworks, Inc., Natick, MA

[2] M. D. Altman. Computational ligand design and analysis in protein complexes using inverse methods, combinatorial search, and accurate solvation modeling. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A., 2006.

[3] M. D. Altman, J. P. Bardhan, B. Tidor, and J. K. White. FFTSVD: A fast multiscale boundary-element method solver suitable for BioMEMS and biomolecule simulation. IEEE T. Comput.-Aid. D., 25 (2006), 274–284.

[4] M. D. Altman, J. P. Bardhan, J. K. White, and B. Tidor. An efficient and accurate surface formulation for biomolecule electrostatics in non-ionic solution. Engineering in Medicine and Biology Conference (EMBC), 7591-7595, 2005.

[5] M. D. Altman, J. P. Bardhan, J. K. White, and B. Tidor. Accurate solution of multi-region continuum electrostatic problems using the linearized Poisson–Boltzmann equation and curved boundary elements. J. Comput. Chem., 30 (2009), 132–153.

[6] K. E. Atkinson. The Numerical Solution of Integral Equations of the Second Kind. Cambridge University Press, 1997.

[7] C. Bajaj, S.-C. Chen, and A. Rand. An efficient higher-order fast multipole boundary element solution for Poisson–Boltzmann-based molecular electrostatics. SIAM J. Sci. Comput., 33 (2011), 826–848.
[8] N. A. Baker, D. Sept, M. J. Holst, and J. A. McCammon. Electrostatics of nanoysystems: Application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA, 98 (2001), 10037–10041.
[9] J. P. Bardhan. Interpreting the Coulomb-field approximation for Generalized-Born electrostatics using boundary-integral equation theory. J. Chem. Phys., 129 (2008), Art. ID 144105.
[10] J. P. Bardhan. Numerical solution of boundary-integral equations for molecular electrostatics. J. Chem. Phys., 130 (2009), Art. ID 094102.
[11] J. P. Bardhan, M. D. Altman, J. K. White, and B. Tidor. Numerical integration techniques for curved-panel discretizations of molecule–solvent interfaces. J. Chem. Phys., 127 (2007), Art. ID 014701.
[12] J. P. Bardhan, R. S. Eisenberg, and D. Gillespie. Discretization of the induced-charge boundary integral equation. Phys. Rev. E, 80 (2009), Art. ID 011906.
[13] J. P. Bardhan and M. G. Knepley. Mathematical analysis of the boundary-integral based electrostatics estimation approximation for molecular solvation: Exact results for spherical inclusions. J. Chem. Phys., 135 (2011), Art. ID 124107.
[14] J. P. Bardhan and M. G. Knepley. Computational science and re-discovery: open-source implementation of ellipsoidal harmonics for problems in potential theory. Computational Science and Discovery, 5 (2012), Art. ID 014006.
[15] J. P. Bardhan, M. G. Knepley, and M. Anitescu. Bounding the electrostatic free energies associated with linear continuum models of molecular solvation. J. Chem. Phys., 130 (2009), Art. ID 104108.
[16] D. Bashford and D. A. Case. Generalized Born models of macromolecular solvation effects. Annu. Rev. Phys. Chem., 51 (2000), 129–152.
[17] C. Berti, D. Gillespie, J. P. Bardhan, R. S. Eisenberg, and C. Fiegna. Comparison of three-dimensional poisson solution methods for particle-based simulation and inhomogeneous dielectrics. Phys. Rev. E, 86 (2012), Art. ID 011912.
[18] A. J. Bordner and G. A. Huber. Boundary element solution of the linear Poisson–Boltzmann equation and a multipole method for the rapid calculation of forces on macromolecules in solution. J. Comput. Chem., 24 (2003), 353–367.
[19] A. H. Boschitsch, M. O. Fenley, and H.-X. Zhou. Fast boundary element method for the linear Poisson–Boltzmann equation. J. Phys. Chem. B, 106 (2002), 2741–2754.
[20] B. O. Brandsdal, J. Åqvist, and A. O. Smalås. Computational analysis of binding of P1 variants to trypsin. Protein Science, 10 (2001), 1584–1595.
[21] B. O. Brandsdal, A. O. Smalås, and J. Åqvist. free energy calculations show that acidic P1 variants undergo large pKa shifts upon binding to trypsin. Proteins: Structure, Function, and Bioinformatics, 64 (2006), 740–748.
[22] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem., 4 (1983), 187–217.
[23] N. Carrascal and D. F. Green. Energetic decomposition with the Generalized-Born and Poisson–Boltzmann solvent models: Lessons from association of G-protein components. J. Phys. Chem. B, 114 (2010), 5096–5116.
[24] D. Chen, Z. Chen, C. Chen, W. Geng, and G.-W. Wei. MIBPB: A software package for electrostatic analysis. J. Comput. Chem., 32 (2011), 756–770.
[25] J. H. Chen, C. L. Brooks, and J. Khandogin. Recent advances in implicit solvent-based methods for biomolecular simulations. Curr. Opin. Struc. Biol., 18 (2008), 140–148.
[26] Y. Chen, J. S. Hesthaven, Y. Maday, and J. Rodriguez. Certified reduced basis methods and output bounds for the harmonic maxwell's problems. SIAM J. Sci. Comput., 32 (2010), 970–996.
[27] C. J. Cramer and D. G. Truhlar. Implicit solvation models: Equilibria, structure, spectra, and dynamics. Chem. Rev., 99 (1999), 2161–2200.
[28] G. Dassios. Ellipsoidal harmonics: theory and applications. Cambridge University Press, 2012.
[29] B. Fares, J. S. Hesthaven, Y. Maday, and B. Stamm. The reduced basis method for the electric field integral equation. J. Comput. Phys., 230 (2011), 5532–5555.
[30] P. O. Fedichev, E. G. Getmantsev, and L. I. Menshikov. O(n log n) continuous electrostatics solvation energies calculation method for biomolecu les simulations. J. Comput. Chem., 32 (2011), 1368–1376.
[31] M. Feig and C. L. Brooks III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr. Opin. Struc. Biol., 14 (2004), 217–224.
[32] M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III. Performance comparison of generalized

Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J. Comput. Chem., 25 (2004), 265–284.
[33] A. Ghosh, C. S. Rapp, and R. A. Friesner. Generalized Born model based on a surface integral formulation. J. Phys. Chem. B, 102 (1998), 10983–10990.
[34] M. K. Gilson and B. Honig. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. Proteins, 4 (1988), 7–18..
[35] D. F. Green and B. Tidor. Design of improved protein inhibitors of HIV-1 cell entry: Optimization of electrostatic interactions at the binding interface. Proteins: Structure, Function, and Bioinformatics, 60 (2005), 644–657.
[36] P. Grochowski and J. Trylska. Continuum molecular electrostatics, salt effects, and counterion binding—a review of the Poisson–Boltzmann theory and its modifications. Biopolymers, 89 (2008), 93–113.
[37] R. Helland, J.Otlewski, O. Sundheim, M. Dadlez, and A. O. Smalås. The crystal structures of the complexes between bovine beta-trypsin and ten P1 variants of BPTI. J. Mol. Biol., 287 (1999), 923–942.
[38] Z. S. Hendsch, C. V. Sindelar, and B. Tidor. Parameter dependence on continuum electrostatic calculations: A study using protein salt bridges. J. Phys. Chem. B, 102 (1998), 4404–4410.
[39] Z. S. Hendsch and B. Tidor. Electrostatic interactions in the GCN4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding. Protein Science, 8 (1999), 1381–1392.
[40] E. W. Hobson. The theory of spherical and ellipsoidal harmonics. Chelsea Pub Co., 1931.
[41] G. C. Hsiao and R. E. Kleinman. Error analysis in numerical solution of acoustic integral equations. International Journal for Numerical Methods in Engineering, 37 (1994), 2921–2933.
[42] L. Hu and G.-W. Wei. Nonlinear Poisson equation for heterogeneous media. Biophys. J., 103 (2012), 758–766.
[43] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph–McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz–Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B, 102 (1998), 3586–3616.
[44] A. H. Juffer, E. F. F. Botta, B. A. M. van Keulen, A. van der Ploeg, and H. J. C. Berendsen. The electric potential of a macromolecule in a solvent: A fundamental approach. J. Comput. Phys., 97 (1991), 144–171.
[45] J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. J. Chem. Phys., 2 (1934), 351.
[46] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. Proteins, 1 (1986), 47–59.
[47] P. Koehl. Electrostatics calculations: latest methodological advances. Curr. Opin. Struc. Biol., 16 (2006), 142–151.
[48] P. Kukic and J. E. Nielsen. Electrostatics in proteins and protein-ligand complexes. Future Med Chem., 2 (2010), 647–666.
[49] S. Kuo, B. Tidor, and J. White. A meshless, spectrally accurate, integral equation solver for molecular surface electrostatics. ACM Journal on Emerging Technologies in Computing Systems, 4 (2008), 6.
[50] L.-P. Lee and B. Tidor. Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex. Protein Science, 10 (2001), 362–377.
[51] R. M. Levy and E. Gallicchio. Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. Ann. Rev. Phys. Chem., 49 (1998), 531–567.
[52] J. Liang and S. Subramaniam. Computation of molecular electrostatics with boundary element methods. Biophys. J., 73 (1997), 1830–1841.
[53] B. Lin and B. M. Pettitt. Electrostatic solvation free energy of amino acid side chain analogs: implications for the validity of electrostatic linear response in water. J. Comput. Chem., 32 (2010), 878–885.
[54] Y. Lin, A. Baumketner, W. Song, S. Deng, D. Jacobs, and W. Cai. Ionic solvation studied by image-charge reaction field method. J. Chem. Phys., 134 (2011), Art. ID 044105.
[55] S. M. Lippow, K. D. Wittrup, and B. Tidor. Computational design of antibody-affinity improvement beyond in vivo maturation. Nature Biotechnology, 25 (2007), 1171–1176.
[56] B. Z. Lu, X. L. Cheng, J. Huang, and J. A. McCammon. Order N algorithm for computation of electrostatic interactions in biomolecular systems. Proc. Natl. Acad. Sci. USA, 103 (2006), 19314–19319.
[57] J. D. Madura, J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. Ridgway-Scott, and J. A. McCammon. Electrostatics and diffusion of molecules in solution: Simulations with the

University of Houston Brownian Dynamics program. Comput. Phys. Comm., 91 (1995), 57–95.
[58] J. Michel, R. D. Taylor, and J. W. Essex. The parameterization and validation of generalized Born models using the pairwise descreening approximation. J. Comput. Chem., 25 (2004), 1760–1770.
[59] K. S. Midelfort, H. H. Hernandez, S. M. Lippow, B. Tidor, C. L. Drennan, and K. D. Wittrup. Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. J. Mol. Biol., 343 (2004), 685–701.
[60] S. Miertus, E. Scrocco, and J. Tomasi. Electrostatic interactions of a solute with a continuum – a direct utilization of *ab initio* molecular potentials for the prevision of solvent effects. Chem. Phys., 55 (1981), 117–129.
[61] M. S. Minkara, P. H. Davis, and M. L. Radhakrishnan. 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 (2012), 573–590.
[62] A. Onufriev, D. A. Case, and D. Bashford. Effective Born radii in the generalized Born approximation: The importance of being perfect. J. Comput. Chem., 23 (2002), 1297–1304.
[63] M. Orozco and F. J. Luque. Theoretical methods for the description of the solvent effect in biomolecular systems. Chem. Rev., 100 (2000), 4187–4225.
[64] J. J. Perona, C. A. Tsu, M. E. McGrath, C. S. Craik, and R. J. Fletterick. Relocating a negative charge in the binding pocket of trypsin. J. Mol. Biol., 230 (1993), 934–949.
[65] M. L. Radhakrishnan. Designing electrostatic interactions in biological systems via charge optimization or combinatorial approaches: insights and challenges with a continuum electrostatic framework. Theor. Chem. Acc., 131 (2012).
[66] K. N. Rankin, T. Sulea, and E. O. Purisima. On the transferability of hydration-parametrized continuum electrostatics models to solvated binding calculations. J. Comput. Chem., 24 (2003), 954–962.
[67] S. Ritter. The spectrum of the electrostatic integral operator for an ellipsoid. In R. E. Kleinman, R. Kress, and E. Martensen, editors, Inverse scattering and potential problems in mathematical physics, pages 157–167, Frankfurt/Bern, 1995.
[68] S. Ritter. On the magnetostatic integral operator for ellipsoids. J. Math. Anal. Appl., 207 (1997), 12–28.
[69] S. Ritter. On the computation of Lamé functions, of eigenvalues and eigenfunctions of some potential operators. Z. Angew. Math. Mech., 78 (1998), 66–72.
[70] W. Rocchia, E. Alexov, and B. Honig. Extending the applicability of the nonlinear Poisson–Boltzmann equation: Multiple dielectric constants and multivalent ions. J. Phys. Chem. B, 105 (2001), 6507–6514.
[71] W. Rocchia, S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. J. Comput. Chem., 23 (2002), 128–137.
[72] B. Roux and T. Simonson. Implicit solvent models. Biophys. Chem., 78 (1999), 1–20.
[73] S. Rush, A. H. Turner, and A. H. Cherin. Computer solution for time-invariant electric fields. J. Appl. Phys., 37 (1966), 2211–2217.
[74] M. Sanner, A. J. Olson, and J. C. Spehner. Reduced surface: An efficient way to compute molecular surfaces. Biopolymers, 38 (1996), 305–320.
[75] M. F. Sanner. Molecular surface computation home page. http://www.scripps.edu/ sanner/html/msms_home.html, 1996.
[76] M. Scarsi and A. Caflisch. Comment on the validation of continuum electrostatics models. J. Comput. Chem., 20 (1999), 1533–1536.
[77] M. Schaefer and M. Karplus. A comprehensive analytical treatment of continuum electrostatics. J. Phys. Chem., 100 (1996), 1578–1599.
[78] P. B. Shaw. Theory of the Poisson Green's-function for discontinuous dielectric media with an application to protein biophysics. Phys. Rev. A, 32 (1985), 2476–2487.
[79] G. Sigalov, P. Scheffel, and A. Onufriev. Incorporating variable dielectric environments into the generalized Born model. J. Chem. Phys., 122 (2005), Art. ID 094511.
[80] T. Simonson. Macromolecular electrostatics: Continuum models and their growing pains. Curr. Opin. Struc. Biol., 11 (2001), 243–252.
[81] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate calculation of hydration free-energies using macroscopic solvent models. J. Phys. Chem., 98 (1994), 1978–1988.

[82] W.C. Still, A. Tempczyk, R. C. Hawley, and T. F. Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. J. Am. Chem. Soc., 112 (1990), 6127–6129.
[83] S. Varma and S. B. Rempe. Coordination numbers of alkali metal ions in aqueous solutions. Biophys. Chem., 124 (2006), 192–199.
[84] A. Warshel, P. K. Sharma, M. Kato, and M. W. Parson. Modeling electrostatic effects in proteins. Biochimica et Biophysica Acta, 1764 (2006), 1647–1676.
[85] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to alpha-helix dipoles. J. Mol. Biol., 157 (1982), 671–679.
[86] T. W. Whitfield, S. Varma, E. Harder, G. Lamoureux, S. B. Rempe, and B. Roux. Theoretical study of aqueous solvation of K+ comparing ab initio, polarizable, and fixed-charge models. J. Chem. Theory Comput., 3 (2007), 2068–2082.
[87] Z. Xu and W. Cai. Fast analytical methods for macroscopic electrostatic models in biomolecular simulations. SIAM Review, 53 (2011), 683–720.
[88] C. Xue and S. Deng. Three-layer dielectric models for generalized coulomb potential calculation in ellipsoidal geometry. Phys. Rev. E, 83 (2011), Art. ID 056709.
[89] B. J. Yoon and A. M. Lenhoff. A boundary element method for molecular electrostatics with electrolyte effects. J. Comput. Chem., 11 (1990), 1080–1086.
[90] S. N. Yu, Y. C. Zhou, and G. W. Wei. Matched interface and boundary (MIB) method for elliptic problems with sharp-edged interfaces. J. Comput. Phys., 224 (2007), 729–756.
[91] R. J. Zauhar and R. S. Morgan. A new method for computing the macromolecular electric-potential. J. Mol. Biol., 186 (1985), 815–820.
[92] R. J. Zauhar and R. S. Morgan. The rigorous computation of the molecular electric potential. J. Comput. Chem., 9 (1988), 171–187.
[93] Y. C. Zhou, M. Feig, and G. W. Wei. Highly accurate biomolecular electrostatics in continuum dielectric environments. J. Comput. Chem., 29 (2008), 87–97.
[94] M. Zink and H. Grubmüller. Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study. Biophys. J., 96 (2009), 1350–1363.