

## Electrostatic Models for Calcium Binding Proteins

Robert Penfold<sup>\*,†</sup> and James Warwicker<sup>‡</sup>

*Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road,  
Reading RG6 6BZ, United Kingdom*

Bo Jönsson<sup>§</sup>

*Physical Chemistry 2, Chemical Centre, University of Lund, P.O. Box 124, Lund S-221 00, Sweden*

*Received: October 22, 1997; In Final Form: July 14, 1998*

The binding of calcium ions to biomolecules has been investigated by comparing three continuum models, all presupposing the dominance of Coulombic interactions. To reflect disparity in polarization response of protein and solvent, the “refined” model (RM) incorporates an inhomogeneous static dielectric constant but admits only approximate analysis within the mean field ansatz. Conversely, the “primitive” model (PM), with a uniform dielectric response, yields to essentially exact solutions of the statistical mechanical problem by Monte Carlo simulation techniques. The “elementary” model (EM) further assumes linear thermal response and ignores steric constraints altogether to obtain a trivially evaluated, analytic result. Two contrasting biomolecules are considered: the large, positively charged, extracellular serine protease subtilisin exhibiting a low affinity binding site and the smaller, negatively charged, intracellular, nonenzymatic protein buffer calbindin that binds calcium tenaciously. As functions of protein (fixed) charge mutations and solution ionic strength, the EM shows consistently good agreement with observed shifts in equilibrium binding constants ( $DpK^a$ ). Predictions from simulation and finite difference solutions of both Poisson–Boltzmann and Debye–Hückel equations in the PM are essentially identical. They also conform closely to experiment, though with systematic deviations at high ionic strength. Evidently both long and short range correlations are suppressed or mitigated in these systems, while nonlinear thermal effects appear to be unimportant even up to 1.0 M monovalent salt. With standard parameters, RM calculations generally overestimate  $DpK^a$ 's, while the differences between RM and PM results are independent of salt concentration. Large electrostatic potentials arising in the RM possibly indicate strained protein conformations, though structural relaxation is not accounted for in any of the models studied. Motivated by exact analysis of salt free cases in planar and spherical geometries, alternative mechanisms explaining the theoretical deviations are discussed, in particular self-image correlation of mobile ions and the location of an *effective* dielectric interface.

### 1. Introduction

It is universally assumed that Coulomb interactions, especially charge–charge, charge–dipole, and dipole–dipole, are central determinants for the structure of proteins and their stability against temperature, pH, ionic strength, etc. as well as their function and kinetic performance in a variety of biochemical processes.<sup>1–3</sup> With a long and venerable pedigree,<sup>4</sup> this notion is fundamental to our understanding of structure/function relationships in biomolecules.<sup>5</sup> Despite this widespread agreement, however, there remains considerable debate among theoreticians concerning not only the mathematical treatment of some electrostatic models, evidenced by the diversity of methods available, but also the consistency, generality, and efficacy of different models in various applications. Much of the controversy stems from attempts to engender routine computational tractability by invoking an *effective* pair interaction potential that, although necessarily dependent on the thermodynamic state of the system, can dramatically reduce the number of degrees of freedom to be considered explicitly.<sup>6</sup> In the theory of simple electrolyte solutions, this approach (gener-

ally formulated by McMillan and Mayer<sup>7</sup>) gives rise to both the “elementary” model (EM) and the so-called “primitive” model (PM).<sup>8,9</sup> Here a (polar) solvent averaged ion–ion potential of mean force is generated that corresponds, in leading order, to the Coulomb interaction in a uniform dielectric material but where the static dielectric constant differs from that of pure solvent reflecting a structural perturbation due to the free charges.<sup>10</sup> Only idealized fixed point charges remain explicitly in the EM to represent ionized moieties of the biomolecule. It is a strictly electrostatic model where the statistical mechanics of mobile salt ions is subsumed under an effective “screened” Coulomb interaction, and all other forces are absorbed into the dielectric background. The PM, however, recognizes short range repulsive interactions (usually described by a hard, purely elastic potential) that reflect the size and shape of all constituent charged particles.

On the other hand, for large, complex particles like proteins and enzymes it has long been argued that the dielectric response of the solute interior should differ significantly from that characterizing the highly polarizable solvent medium<sup>11</sup> leading to a “refined” model (RM)<sup>12</sup> incorporating a nonuniform dielectric continuum and, consequently, a dielectric interface. While this extension broadly addresses a physical mechanism in a plausible manner, to the authors' knowledge, no formal

<sup>†</sup> E-mail: robert.penfold@bbsrc.ac.uk. (fax) +44 (0)118 926 7917.

<sup>‡</sup> E-mail: james.warwicker@bbsrc.ac.uk. (fax) +44 (0)118 926 7917.

<sup>§</sup> E-mail: fk2boj@grosz.fkem2.lth.se. (fax) +46 (0) 46 10 4543.

justification has been provided in the context of an effective charge–charge interaction free energy, mediated by an immiscible solvent *mixture* (i.e., water and hydrocarbon), that has been generated as a suitable partial ensemble average over the generalized mixture degrees of freedom. Furthermore, in biomolecular applications the RM is almost never solved exactly<sup>13</sup> but first reduced, by invoking the mean field approximation, to a nonlinear partial differential equation (the Poisson–Boltzmann equation) that is amenable to numerical integrators. Pioneering work<sup>14</sup> employed classical finite difference (FD) techniques (mapping the gradient operator onto a discrete spatial grid) that remain popular in this area and, therefore, are often referred to as the FDPB model.<sup>15</sup> Although smooth spatial variation of the surface dielectric has been examined within the FDPB<sup>16</sup> (based only on polarization saturation due to the solute electric field and not including effects of disturbing the bulk solvent structure), more commonly a discontinuous step interface is assumed, between homogeneous dielectric media. The RM introduces two additional undetermined parameters:

1. the ratio of solvent/solute dielectric constants and
2. the location of the dielectric boundary

and the model behavior is quite sensitive to both. Rashin and co-workers<sup>17</sup> have attempted to introduce a “... self-consistent, physically reasonable and objective way to define the boundary position ...” by systematically adjusting the radii of atoms comprising the protein based on arguments concerning the extent of electron density that can be associated with each atom together with fits to experimental solvation enthalpies for simple ionic salts. The dielectric interface, however, is constructed in the conventional manner by rolling a spherical probe uniformly on the surface of protein “atoms” and is therefore related only to the structural aspects of the polarizable electron density without self-consistently accounting for energetic contributions to the position of an *effective* boundary. Different, but constant, probe radii are considered<sup>17</sup> reflecting alternative solvent orientations at the interface with solute, but the consequences for calculated hydration enthalpies are reported to be insignificant. On the contrary, we find the determination of metal ion binding constants is highly sensitive to probe radius. In a recent detailed study of acid dissociation constants for titratable residues in proteins using the FDPB approximation, Demchuk and Wade<sup>18</sup> have systematically examined the effects of varying the RM parameters in an effort to construct a measure for judging the necessity of the RM and provide *a priori* estimates for the extra information required. They conclude that the RM need only be invoked for a relatively small number of “buried” residues and produce a criterion, independent of data external to the model, that reasonably distinguishes this class. It should be clear, however, that their categorization relies upon an approximate solution of the RM behavior. In an alternative empirical approach, Warwicker<sup>19</sup> has suggested the RM is also appropriate for exposed titration sites but that the large desolvation free energy (i.e., the Born energy) encountered should be mitigated by increased solvent entropy upon neutralization of the solute residue where the ordering field of a free (i.e., an electric monopole), fixed (with respect to the protein reference frame) charge is removed. The question of a suitable “low” dielectric constant remains open, however, and seems crucially dependent on the property of interest. For example, a solvent/solute dielectric ratio around 40 is indicated to describe the formation of enzyme–inhibitor complexes with low and moderate surface charge densities (trypsin deactivation by bovine pancreatic trypsin inhibitor<sup>1</sup> or by benzamidine<sup>20,21</sup>), while the binding of nucleotides with high surface charge density requires

a greater screening.<sup>22</sup> A dielectric ratio between 4 and 8 is recommended for the general prediction of acid–base titration curves of biomolecules<sup>23,24</sup> though, in the case of large  $pK_a$  shifts, driven by charge burial, associated with proteins of natural redox function,<sup>25</sup> substantially higher ratios are required.<sup>26</sup> Moreover, the PM (i.e., a dielectric constant ratio of unity) has been found successful in the analysis of metal ion binding to proteins, in particular the binding of calcium to calbindin D<sub>9k</sub><sup>27,28</sup> as well as calmodulin and its tryptic fragments.<sup>29</sup> Warshel and co-workers<sup>30–32</sup> have repeatedly emphasized the fundamental inconsistency of defining a macroscopic protein dielectric “constant” that has any local meaning when the underlying microscopic structure is emphatically inhomogeneous. Various definitions have been suggested (Table 1 in ref 31) depending on the nature of the interactions that contribute most to the property of interest. It is argued that calculations of redox potentials,<sup>33</sup> ionizable group  $pK_a$ 's,<sup>34</sup> and hydration free energies<sup>35</sup> are all sensitive to microscopic details of the charge/dipole environment not described by global quantities, particularly where these are treated as fitting parameters rather than truly representative averages.

An important feature of the PM is the relative ease with which long range Coulomb correlations can be investigated. Essentially exact Monte Carlo methods<sup>36</sup> and approximate integral equation theories,<sup>37</sup> that nonetheless reach beyond the mean field ansatz, have elucidated the ionic strength dependence and the effects of multivalent salts on metal binding phenomena. To satisfactorily evaluate the statistical mechanical partition function of the RM requires a tractable statement of the corresponding effective Hamiltonian. Attempts to construct such an energy functional, for use in computer simulation or many-body theory, have been made<sup>38–40</sup> though the results appear either too cumbersome or rely on inappropriate geometrical assumptions and have yet to find routine acceptance. It would seem a weakness in the face of overwhelming application of the RM to biological problems that it is still comparatively untested in the modern paradigm of liquid state physics where statistical simulation plays such an important role in both the justification of models and the validation of approximate theories.<sup>41</sup>

A recent review<sup>42</sup> has highlighted the significance of metal ion binding, governed by charge–charge interactions, for overall protein stability, where higher order electric multipole contributions (“hydrophilic” interactions) are typically too weak and hydrophobic interactions are insufficiently specific. Since continuum electrostatic models most appropriately describe the physics of coupled charge distributions, metal ion binding phenomena provide a natural test bed to evaluate their relative merit in the context of biomolecular science. In the present work, the study of calcium ion binding is extended to consider the serine protease subtilisin that is both structurally and functionally very different from calbindin. As the focus is on Coulomb forces only, other contributions to the binding free energy are ignored (e.g., variations in protein conformation, entropy changes associated with imposition or relaxation of dynamical constraints, electronic dispersion energies, etc.). To make contact with experiment then, only *changes* in binding free energy, resulting from structure/charge perturbations of the protein (typically by site-directed mutagenesis<sup>43</sup>), are compared where the effects of interactions other than direct charge–charge terms are assumed to be minor and thus neglected. Furthermore, the comparison with experiment of predictions from the primitive and refined models is discussed. For ease of reference, abbreviations used throughout this paper are collected in a brief glossary (Table 1).

**Calcium Binding Proteins.** The subtilisin family of bacterial proteases<sup>44</sup> (along with the trypsin group and the  $\alpha/\beta$ -hydrolases)

**TABLE 1: Glossary of Abbreviations Used throughout This Paper**

abbreviation	description
EM	elementary model—point protein charges, screened Coulomb interaction, uniform, and homogeneous dielectric medium
PM <sup>a</sup>	primitive model—point protein charges, mobile ions, hard short range exclusion, uniform, and homogeneous dielectric medium
RM <sup>b</sup>	refined model—point protein charges, mobile ions, hard short range exclusion, uniform but inhomogeneous dielectric media
DH	Debye—Hückel (linear response) approximation
PB	Poisson—Boltzmann (nonlinear response) approximation
MC	Monte Carlo canonical ensemble simulation
FD	finite difference numerical integration method

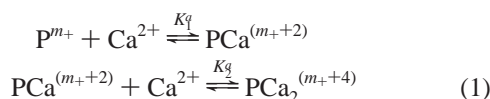
<sup>a</sup> In the primitive model (PM), MC simulations explicitly treat mobile ions as charged hard spheres (diameter 4.25 Å), while the finite difference solutions of the DH and PB equations approximately account for small ion excluded volume effects through a 2.125 Å thick Stern layer around the biomolecule. <sup>b</sup> In the refined model (RM), only finite difference calculations based on the DH and PB elliptic PDEs have been carried out though the excluded volume of the protein has been incorporated.

are characterized by the presence of a catalytic triad of residues (active serine, neighboring histidine, and buried aspartic acid) responsible for the specific cleavage of peptide bonds.<sup>45</sup> The class of serine proteases (as they are collectively known) are involved in a diverse range of cellular functions including digestion, blood clotting, hormone production, and complement activation.<sup>46</sup> Like all extracellular proteases, it is a robust enzyme though the kinetic barrier to thermal folding/unfolding is uncommonly large for a monomeric protein. The binding of metal ions (e.g., Ca<sup>2+</sup>), that contribute significantly to the overall free energy of folding, is an effective strategy for survival in the relatively harsh extracellular environment. Wild type subtilisin BPN' (Novo) from *Bacillus amyloliquefaciens* consists of 275 amino acids and binds two Ca<sup>2+</sup> ions: one specifically with high affinity ( $\log(K_1^a/M^{-1}) \approx 8$  at 338 K) and the second much less tenaciously ( $\log(K_2^a/M^{-1}) \approx 1.5$  at 338 K) at a site that can also be occupied by other cations in the absence of calcium.<sup>47</sup> For thermostability of the corresponding enzyme subtilisin J, from the strain *Bacillus stearothermophilus*, Jang et al.<sup>48</sup> noted the importance of negatively charged side chains accumulated in the region of the weak Ca<sup>2+</sup> binding site. Recent studies<sup>49,50</sup> have reported direct evidence for dramatically enhanced thermal stability conferred by occupation of the low affinity site and that remote mutations (up to 12 Å distant) can have a substantial effect despite minimal structural changes overall. This is a compelling example of the importance of Coulomb forces for the structure and function of biomolecules. Apo subtilisin is formally charged +5 at physiological pH (19 positive and 14 negative residues with N terminal alanine and C terminal glutamine both charged).

Conversely, bovine calbindin D<sub>9k</sub> is a globular intracellular, nonenzymatic protein, from the structurally related calmodulin superfamily, with a calcium regulatory function<sup>51</sup> that binds two Ca<sup>2+</sup> ions with almost equally high affinity ( $\log(K_1^a/M^{-1}) \approx 8.3$  and  $\log(K_2^a/M^{-1}) \approx 8.6$  at 298 K).<sup>52</sup> Both binding sites are characterized by a so-called EF-hand helix-loop-helix motif,<sup>53</sup> absent in the serine proteases.<sup>54</sup> Also in contrast to subtilisin, calbindin is relatively small consisting of just 75 amino acid residues, while the net charge of the apo form of the recombinant native protein is −8 and pH 7.5, assuming the N-terminal methionine is acetylated (10 positive and 18 negative residues).<sup>27</sup>

## 2. Methods

**2.1. Binding Constants.** Consider the following equilibria describing the consecutive uptake of two calcium ions by a generalized unloaded (apo) protein P<sup>m+</sup> with overall charge, say m<sub>+</sub> positive, to finally yield the loaded (holo) form PCa<sub>2</sub><sup>(m<sub>+</sub>+4)</sup>



The observable association or binding constants  $K_n^a$  (with Arabic subscript  $n$  indicating the binding sequence) are expressed in terms of the microscopic equilibrium constants (i.e., the  $\kappa$ 's)<sup>27</sup>

$$K_1^a = \kappa_1^a + \kappa_{II}^a \quad (2)$$

$$K_2^a = \frac{\kappa_1^a \kappa_{II,I}^a}{\kappa_1^a + \kappa_{II}^a} \quad (3)$$

$$K^a = K_1^a K_2^a = \kappa_1^a \kappa_{II,I}^a \quad (4)$$

$$\text{p}K^a = -\log K^a \quad (5)$$

where Roman subscript X refers to site X regardless of the state of other sites, while Y:X denotes the equilibrium at site Y given that site X is occupied. In turn, the microscopic equilibrium constants (i.e., the ratios of forward and reverse reaction rates at each site) are related to the *change* in chemical potential  $\Delta\mu$  for association or equivalently the *difference* in Gibbs free energy change per particle  $\Delta G/N$  for insertion from a reservoir at some reference state to the product and reactant states

$$\begin{aligned} \Delta\mu &= \mu_{\text{products}} - \mu_{\text{reactants}} \\ &= (G_{\text{products}} - G_{\text{reference}})/N - (G_{\text{reactants}} - G_{\text{reference}})/N \\ &= (G_{\text{products}} - G_{\text{reactants}})/N \\ &= \Delta G/N \\ &= -kT \ln \kappa \\ &= -kT \ln (\kappa^{\text{elec}} \kappa^{\text{rest}}) \\ &= \Delta\mu^{\text{elec}} + \Delta\mu^{\text{rest}} \end{aligned} \quad (6)$$

where we have partitioned  $\Delta\mu$  into a contribution arising solely from electrostatic interactions and a remainder term accounting for all other effects (solvent ordering, conformation changes of protein, etc.). Define, now, the *shift*  $DQ$  in some observable  $Q$  due to mutation ( $v$  = variant) of the wild type ( $n$  = native) protein

$$DQ = Q(v) - Q(n) \quad (7)$$

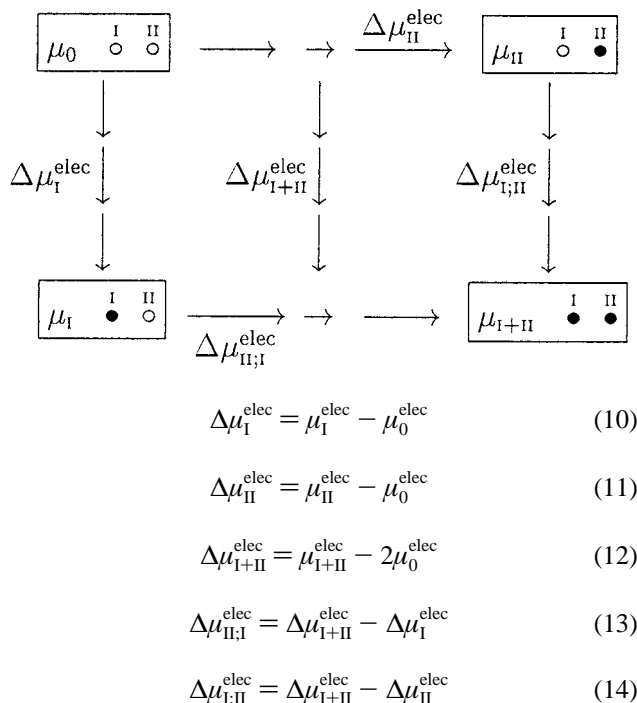
and assume that nonelectrostatic contributions to observed shifts in chemical potential are negligible so that

$$D\Delta\mu^{\text{rest}} = 0 \quad (8)$$

The *total* macroscopic equilibrium constants can be then reduced to a sum of electrostatic shifts only, i.e.

$$\begin{aligned}
DpK^a &= -\log \left( \frac{\kappa_I^a(v)\kappa_{II,I}^a(v)}{\kappa_I^a(n)\kappa_{II,I}^a(n)} \right) \\
&= \frac{1}{\ln 10} (\beta(\Delta\mu_I^{\text{elec}}(v) + \Delta\mu_{II,I}^{\text{elec}}(v) - \Delta\mu_I^{\text{elec}}(n) - \\
&\quad \Delta\mu_{II,I}^{\text{elec}}(n) + \Delta\mu_I^{\text{rest}}(v) + \Delta\mu_{II,I}^{\text{rest}}(v) - \Delta\mu_I^{\text{rest}}(n) - \\
&\quad \Delta\mu_{II,I}^{\text{rest}}(n))) \\
&= \frac{1}{\ln 10} (D(\beta\Delta\mu_{I+II}^{\text{elec}}) + D(\beta\Delta\mu_{I+II}^{\text{rest}})) \\
&= \frac{D(\beta\Delta\mu_{I+II}^{\text{elec}})}{\ln 10} \\
&= \frac{D(\beta\mu_{I+II}^{\text{elec}})}{\ln 10} - \frac{2D(\beta\mu_0^{\text{elec}})}{\ln 10} \\
&\approx \frac{D(\beta\mu_{I+II}^{\text{elec}})}{\ln 10} \\
&= \frac{D(\beta\Delta G_{I+II}^{\text{elec}}/N)}{\ln 10} \quad (9)
\end{aligned}$$

where the penultimate approximation follows since the protein (in micromolar concentration) mutations have only a small effect on the solvation chemical potential  $\mu_0$  of the metal ion in bulk. Furthermore, we have made use of relations from the thermodynamic cycle



Note that if, say, site I binds much more tenaciously than site II (i.e.,  $\kappa_I^a \gg \kappa_{II}^a$ ), then

$$K_1^a = \kappa_I^a + \kappa_{II}^a \approx \kappa_I^a \quad (15)$$

$$K_2^a = \frac{\kappa_I^a \kappa_{II,I}^a}{\kappa_I^a + \kappa_{II}^a} = \frac{\kappa_{II,I}^a}{1 + \kappa_{II}^a / \kappa_I^a} \approx \kappa_{II,I}^a \quad (16)$$

Thus, under these conditions, the site specific microscopic equilibrium constants are well approximated by the experimentally measured binding constants, as we should expect.

**2.2. Protein Structures.** The native protein structures are described in atomic detail taking X-ray crystallographic coordinates available from the Brookhaven Protein Data Bank (subtilisin Novo, 2SNI, at 2.1 Å resolution;<sup>55</sup> calbindin D<sub>9k</sub>, 3ICB, at 2.3 Å resolution<sup>56</sup>). For both proteins a simple fixed charge distribution is used with all negatively charged carboxylic oxygen atoms of glutamate (E) and aspartate (D) residues as well as the C-terminal, given a partial charge of −0.5 units. Lysine (K) residues carry a +1.0 charge on their ζ-nitrogen atom, while arginine (R) and histidine (H) residues share the positive unit charge equally over the tautomeric nitrogen atoms.

X-ray structures for the singly substituted subtilisin variants P172D and G131D are reported by Pantoliano et al.<sup>47</sup> and indicate that the main-chain protein fold is not significantly different from that of the wild type, though they have not been deposited in the data bank. Using the QUANTA package (Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA 02154, U.S.A.), these structures were constructed by adjustment of main chain torsion angles localized to the mutant residue and its adjoining neighbors while constraining the separation between the midpoint of the carboxyl oxygens of the aspartate (D) and the low affinity binding site to the experimental values<sup>47</sup> (6.8 Å for P172D and 13.2 Å for G131D). The destabilizing mutation E195Q was assumed to be without significant structural consequences so that the carboxyl group of the glutamate was transformed to an “amide group” by simply setting the partial oxygen atom charges to zero. Coordinates for the asparagine residue of the N76D mutant were imported from the crystal structure 1SO1 of the 6-fold substituted variant determined by Pantoliano et al.<sup>57</sup> to 1.8 Å resolution. For both simulation and finite difference calculations, explicit protons in subtilisin are ignored, while all heavy atoms are assigned a common diameter of 4 Å.

As previously described,<sup>27,28</sup> all the calbindin mutations (single: E17Q, D19N; double: E17Q+D19N; triple: E17Q+D19N+E26Q) were assumed sufficiently conservative that the partial charge of each carboxylic oxygen atom was simply removed to leave an effective amide group with no conformational change. Again, in accord with earlier simulation studies<sup>27,28</sup> of calbindin, the protein atoms were assigned hard core diameters of 3.6, 3.2, 3.3, and 2.0 Å for carbon, oxygen, nitrogen atoms, and polar protons, respectively.

**2.3. Monte Carlo Simulations.** As benchmarks, essentially exact statistical mechanical solutions for the PM have been generated by Monte Carlo simulation. The RM (i.e., the PM but with a nonuniform dielectric background) presents a difficulty in that the corresponding Hamiltonian cannot be expressed in a compact way, thereby precluding the application of routine simulation methods.

Within the cell model, the rigid protein remains fixed during the simulation at the center of a spherical hard boundary which confines the otherwise mobile, charged hard spheres (representing counterions and salt) to a constant volume determined, in principle, by the bulk concentration of protein. By using the conventional Metropolis algorithm,<sup>58</sup> configurations were sampled from the canonical ( $N, V, T$ ) ensemble with a pairwise additive interaction potential

$$u_{ij}(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} z_i z_j q^2 / (4\pi\epsilon_0 \epsilon_r r_{ij}) & r_{ij} \geq (\sigma_i + \sigma_j)/2 \\ \infty & r_{ij} < (\sigma_i + \sigma_j)/2 \end{cases} \quad (17)$$



**TABLE 2: Canonical Monte Carlo Simulation Details for Subtilisin in KCl/CaCl<sub>2</sub> Solution of 100 mM Ionic Strength at 338 K<sup>a</sup>**

run	N <sub>Ca<sup>2+</sup></sub>	N <sub>K<sup>+</sup></sub>	N <sub>Cl<sup>-</sup></sub>	[Ca <sup>2+</sup> ] (mM)	[K <sup>+</sup> ] (mM)	[Cl <sup>-</sup> ] (mM)
1	0	200	204	0	100	102
2	10	170	194	5	85	97
3	20	140	184	10	70	92
4	50	50	154	25	25	77

<sup>a</sup> Particles numbers  $N_X$  and corresponding cell concentrations  $[X]$  are listed. Statistical averages were accumulated over  $6.16 \times 10^6$  configurations after discarding  $3.08 \times 10^6$  configurations to equilibrate the system from a random initial state.

where  $z_i$  is the partial valency of atom  $i$  in units of the elementary charge  $q$  (of a proton),  $\sigma_i$  is the corresponding atomic diameter, and  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$  is the center separation of atoms  $i$  and  $j$ . The permittivity of free space is denoted  $\epsilon_0$ , and the relative static dielectric constant  $\epsilon_r$  is taken to be uniform throughout the cell at a value appropriate for solvent (water), i.e., the electrostatic PM. Excess chemical potentials ( $\mu_I^{\text{ex}}$ ,  $\mu_{\text{II}}^{\text{ex}}$  and  $\mu_{\text{I+II}}^{\text{ex}}$ ), comprising electrostatic and hard core contributions, were obtained by the test particle method<sup>59</sup> modified to account for insertion correlations required by the electroneutrality condition.<sup>60</sup>

To test the concentration independence of the low affinity binding constant in subtilisin and also mirror the experimental conditions,<sup>47</sup> simulations were carried out at 338 K ( $\epsilon_r = 65.3$ ) over a range of bulk free calcium ion concentrations ( $0 \leq [\text{Ca}^{2+}] \leq 25$  mM) with the total ionic strength of the solution maintained at 100 mM by addition of KCl. No systematic dependence of  $DpK^a$  on  $[\text{Ca}^{2+}]$  was observed. All mobile ions were assigned a diameter of 4.25 Å in accord with previous simulation studies of simple electrolyte solutions.<sup>61</sup> At the very low enzyme concentration studied experimentally (1.8 μM), the corresponding simulation cell size would require well over  $10^5$  salt ions at this ionic strength! Consequently, the cell diameter has been reduced to 185.11 Å giving an apparent enzyme concentration of 0.5 mM with several hundred mobile ions. The charge screening, however, based on a simple Debye–Hückel analysis, indicates that no sensible error is incurred. Particle numbers and ion concentrations for each simulation run are listed in Table 2. All runs started from a pseudorandom configuration and at least 5000 MC cycles (configurations per particle) were discarded in equilibration before accumulating averages over the subsequent 10000 MC cycles. Details of the calbindin simulations have been reported previously.<sup>27</sup>

**2.4. Poisson–Boltzmann and Debye–Hückel Approximations.** Consider the  $i$ th particle of type  $k$  moving under the influence of an inhomogeneous external field  $v_k(\mathbf{r}_i)$  and interacting through a pair potential  $u_{kl}(\mathbf{r}_i, \mathbf{r}_j)$  with the remaining  $(\sum_{k=1}^v N_k) - 1 = N - 1$  particles comprising a  $v$ -component system in volume  $V$  at thermal equilibrium with temperature  $T = 1/(k\beta)$  where  $k$  is Boltzmann's constant. Suppose now that this particle is fixed at position  $\mathbf{R}_i$  in  $V$  so that it exercises a total external potential on the other particles, depending on their configuration, of the form

$$u_k^{(1)}(\mathbf{R}_i) = \sum_{l \neq k}^v \sum_{j=1}^{N_l} (u_{kl}(\mathbf{R}_i, \mathbf{r}_j) + v_k(\mathbf{r}_j)) + \sum_{j \neq i}^{N_k} (u_{kk}(\mathbf{R}_i, \mathbf{r}_j) + v_k(\mathbf{r}_j)) \quad (18)$$

The distribution  $n_k(\mathbf{r})$  of  $k$ -type particles is then

$$n_k(\mathbf{r}) \propto \exp(-\beta v_k(\mathbf{r})) \langle \exp(-\beta u_k^{(1)}(\mathbf{r})) \rangle_{N-1} \quad (19)$$

where  $\langle \dots \rangle_M$  denotes a canonical configurational average in an  $M$ -particle system. Introducing now the mean internal potential at  $\mathbf{r}$

$$\hat{u}_k^{(1)}(\mathbf{r}) = \sum_{l=1}^v \int_V d\mathbf{s} n_l(\mathbf{s}) u_{kl}(\mathbf{r}, \mathbf{s}) \quad (20)$$

as an average over the equilibrium distribution of all particles, we obtain

$$n_k(\mathbf{r}) \propto \exp(-\beta(\hat{u}_k^{(1)}(\mathbf{r}) + v_k(\mathbf{r}))) \langle \exp(-\beta(u_k^{(1)}(\mathbf{r}) - \hat{u}_k^{(1)}(\mathbf{r}))) \rangle_{N-1} \quad (21)$$

In the case of Coulombic pair interactions (eq 17) in an external electric field, this becomes

$$n_k(\mathbf{r}) \propto \exp(-\beta z_k q \varphi(\mathbf{r})) \langle \exp(-\beta \Delta u_k^{(1)}(\mathbf{r})) \rangle_{N-1} \quad (22)$$

where  $\varphi(\mathbf{r})$  is the scalar mean field electrostatic potential. For a continuum dielectric model it is a straightforward application of electrostatic theory to write down Poisson's equation relating the divergence of Maxwell's displacement field to a fixed charge distribution

$$\nabla \cdot (\epsilon_r(\mathbf{r}) \nabla \varphi(\mathbf{r})) = - \frac{4\pi q}{\epsilon_0} \sum_k z_k n_k(\mathbf{r}) \quad (23)$$

where the sum ranges over all charge types in the system. To the extent that the sufficient (but not necessary) condition

$$\Delta u_k^{(1)}(\mathbf{r}) = u_k^{(1)}(\mathbf{r}) - \hat{u}_k^{(1)}(\mathbf{r}) = 0; \quad \forall \mathbf{r}; \quad \forall k \quad (24)$$

is satisfied everywhere for each component, then the mean field approximation is valid; eqs 22 and 23 can be coupled to eliminate the distributions  $n_k(\mathbf{r})$  leaving a partial differential equation in the mean electrostatic field which can be economically solved to high precision using a variety of numerical techniques.<sup>14,62–68</sup> In case the mean field is everywhere small compared to the average thermal energy, further computational efficiency can be achieved by linearizing the exponential response in eq 22 giving rise to the Debye–Hückel theory. It should not be overlooked that, though linearization of the mean field contribution may be mathematically valid, the Debye–Hückel distributions are unphysical (negative!) at short range<sup>36</sup> and will be in serious error at intermediate/long range in the presence of significant real correlations that have been artificially eliminated by the mean field approximation. Despite this defect, however, thermodynamic properties, computed as integrals over the distributions, for systems with only monovalent charges are often quite accurate.<sup>69</sup>

By definition, the EM excludes entropic contributions since the protein charges are assumed static, thus thermal effects enter only through the effective Debye–Hückel Hamiltonian. All the approximate Poisson–Boltzmann results reported here for the PM and RM have been obtained in a finite difference representation (FDPB) using a successive over-relaxation algorithm described previously.<sup>16,14</sup> In the refined model, a low<sup>16</sup> relative dielectric constant  $\epsilon_r = 3$  was generally chosen for the protein interior (and assumed independent of temperature), though for a few subtilisin calculations (Table 3) intermediate values<sup>23</sup>  $\epsilon_r = 10$  and  $\epsilon_r = 20$  were also specified. A typical solvent probe diameter of 1.4 Å was used to determine the position of

**TABLE 3: Comparison of Total  $\text{Ca}^{2+}$  Binding Constant Shifts Due to Point Mutation  $\text{DpK}^a$  in Subtilisin between Theoretical Estimates and Experiment<sup>47,48,57 d</sup>**

	$\text{DpK}^a$			
	P172D	G131D	E195Q	N76D
$T$ (K)	338.15	338.15	333.15	332.15
$\epsilon_r$ ( $\text{H}_2\text{O}$ )	65.3	65.3	66.7	67.2
$I$ (mM)	100	100	100	50
expt	$-0.53(05)^a$	$-0.33(05)^a$	$\approx +1.0^b$	$\approx -0.30^c$
DH – EM*	-0.56	-0.15		
DH – EM	-0.47	-0.14	+0.41	-0.58
MC – PM	$-0.66(10)$	$-0.28(10)$	$+0.61(10)$	$-0.73(10)$
DH – PM	-0.69	-0.29	+0.69	-0.76
PB – PM	-0.65	-0.26	+0.61	-0.69
DH – RM (20.0)	-0.95	-0.34	+1.15	-1.17
DH – RM (10.0)	-1.11	-0.39	+1.54	-1.62
CH – RM (3.0)	-1.53	-0.56	+2.84	-3.43
PB – RM (3.0)	-1.35	-0.52	+2.71	-3.19

<sup>a</sup> Reference 47. <sup>b</sup> Reference 48. <sup>c</sup> Reference 57. <sup>d</sup> Also included is an estimate (EM\*) reported by Pantoliano et al.<sup>47</sup> in the spirit of the elementary model, but with an inconsistent treatment of temperature effects on the screening length giving an effective uniform dielectric constant  $\epsilon_r = 128$ . The absolute temperature  $T$ , bulk aqueous static dielectric constant  $\epsilon_r(\text{H}_2\text{O})$ , and solution ionic strength  $I$  are listed. For the refined model, protein relative dielectric constants are indicated in parentheses. Standard errors on experimental and simulation results are indicated in parentheses.

the dielectric interface. Unless stated otherwise, each calculation comprised an initial run with a grid spacing of 3 Å extending 90 Å from the protein surface where zero potential boundary conditions were applied, followed by a focus run<sup>70</sup> with a 0.6 Å spatial division over either a  $60 \times 60 \times 60$  grid centered on each binding site for linear mobile ion response (Debye–Hückel) or a  $100 \times 100 \times 100$  grid centered on the midpoint of the vector joining the binding sites for exponential response (Poisson–Boltzmann). Electrostatic free energies  $G^{\text{elec}}$  (with respect to a free energy zero corresponding to the uncharged system) were obtained in the latter case by performing the “charging” integration with a simple trapezoidal quadrature over ten increments, and then the shift in binding constant due to protein charge mutation becomes

$$\begin{aligned} D(\Delta G_{\text{I+II}}^{\text{elec}}) &= D(G_{\text{I+II}}^{\text{elec}}(\text{holo}) - (G_{\text{I+II}}^{\text{elec}}(\text{apo}) + 2G_0^{\text{elec}})) \\ &= D(G_{\text{I+II}}^{\text{elec}}(\text{holo}) - G_{\text{I+II}}^{\text{elec}}(\text{apo})) \end{aligned} \quad (25)$$

while in linear Debye–Hückel response,  $G^{\text{elec}}$  is directly proportional to the Coulomb energy leading to further simplification

$$\begin{aligned} D(G_{\text{I+II}}^{\text{elec}}(\text{holo}) - G_{\text{I+II}}^{\text{elec}}(\text{apo})) &= D((G_{\text{I}}^{\text{elec}}(\text{holo}) - G_{\text{I}}^{\text{elec}}(\text{apo})) + (G_{\text{II}}^{\text{elec}}(\text{holo}) - G_{\text{II}}^{\text{elec}}(\text{apo}))) \\ &= D\left(\frac{1}{2}z_{\text{Ca}}^2 q(\varphi_{\text{I}}(\text{holo}) + \varphi_{\text{II}}(\text{holo}))\right) \\ &= 2D(q\varphi_{\text{I}}(\text{holo}) + q\varphi_{\text{II}}(\text{holo})) \end{aligned} \quad (26)$$

Thus, for the linear calculation, only electrostatic potentials are required in eq 26 and are computed independently at each site by linear interpolation of the converged potential map over neighboring grid points. To mimic the finite size effects of mobile ions, explicitly included in the MC simulations, a 2.125 Å thick Stern layer was imposed that effectively increases the

size of the protein without displacing the dielectric boundary. In the elementary model, all short range interaction effects (i.e., predominantly steric repulsions) are ignored, a uniform solvent dielectric response is imposed, and Debye–Hückel screening is assumed so that the electrostatic potential at binding site X is

$$\varphi_{\text{X}} = \varphi(\mathbf{r}_{\text{X}}) = \frac{q}{4\pi\epsilon_r\epsilon_0} \sum_i z_i \frac{\exp(-\kappa|\mathbf{r}_i - \mathbf{r}_{\text{X}}|)}{|\mathbf{r}_i - \mathbf{r}_{\text{X}}|} \quad (27)$$

where the sum ranges over all other charges  $z_i q$  on the protein and  $\kappa$  is the inverse screening length related to the ionic strength  $I$  of the solution with solvent mass density  $\rho$  by ( $L$  is Avogadro’s constant)

$$\kappa^2 = \frac{2q^2\rho LI}{\epsilon_r\epsilon_0 kT} \quad (28)$$

Owing to the cancellation of most terms by the mutation difference operator  $D$ , only the variant charges are considered in the calculation of  $\text{DpK}^a$ .

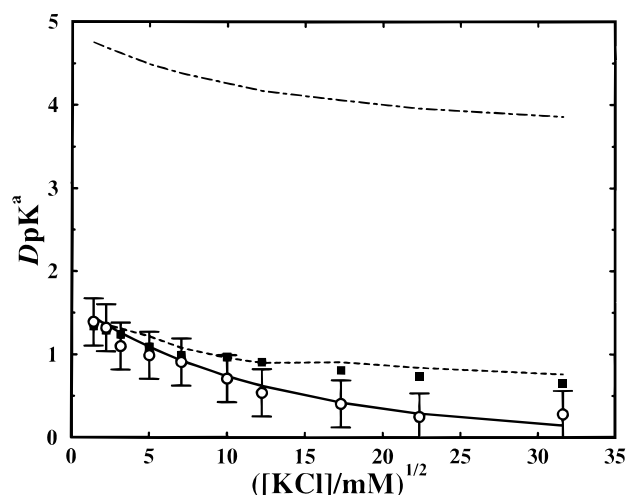
### 3. Results and Discussion

Table 3 summarizes the calculated binding constant shifts due to single point mutations of subtilisin from each of the elementary (EM), primitive (PM), and refined (RM) electrostatic models as well as listing the corresponding experimental values. The mutated residues of the stabilized variants P172D and G131D (relative to native) both lie within 15 Å of the low affinity site where the shifts in association constant were determined to high precision by measuring the decrease in the thermal deactivation rate of the enzyme under titration with calcium.<sup>47</sup> By using similar techniques, the destabilizing mutation E195Q, lying within 8 Å of the weak binding site, was reported by Jang et al.<sup>48</sup> to lower the calcium affinity by about an order of magnitude relative to wild-type, though full details are not given. Comparison of the electrostatic calculations with this result should also recognize the different bacterial origin of subtilisin J studied by Jang et al.<sup>48</sup> and the likelihood of minor structural variations from subtilisin BPN’ that have not been considered here. In contrast, the N76D mutation lies in the neighborhood of the high affinity site, some 5 Å distant, but the binding constant shift has only been indirectly estimated from the change in unfolding free energy obtained by differential scanning calorimetry (DSC)<sup>57</sup> rather than explicit titration with calcium. Comparing these quantities tacitly assumes that all contributions to the free energy of thermal denaturation, other than the free energy change associated with calcium binding, are unaffected by the single localized charge mutation. The shift in melting temperature determined by the DSC experiment yields no unequivocal information, however, about the effect of mutation on the unfolded form of the enzyme. Poor agreement may not, therefore, be necessarily indicative of a failure in the electrostatic models but rather signals a breakdown of this assumption. Moreover, since the thermal denaturation of subtilisin BPN’ is known to be irreversible<sup>71</sup> (and presumably that of subtilisin J also), all the native and variant enzymes are subject to possible deamidation of asparagine and glutamine residues<sup>72</sup> that would alter the partial charge array during the process. The electrostatic calculations do not account for such effects.

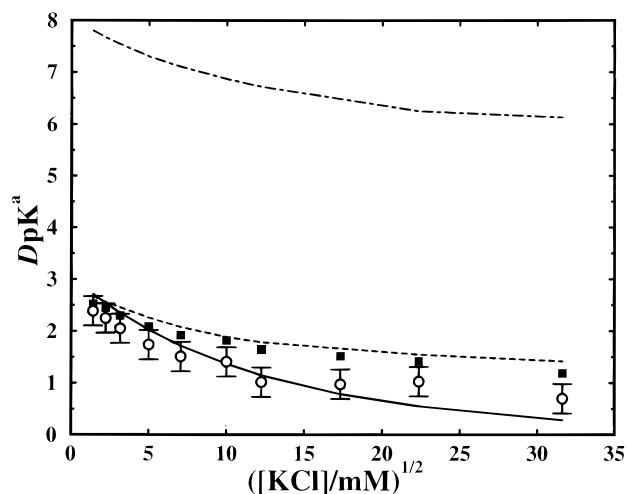
All the electrostatic models correctly predict the sign of  $\text{DpK}^a$ , successfully differentiating between stabilizing and destabilizing charge mutations and thereby confirming the dominant influence

of Coulomb forces on ion binding. A consistent trend is observed where the magnitude of  $DpK^a$  increases with each additional constraint on passing from the least to the most sophisticated models. For mutations around the low affinity site (P172D, G131D, and E195Q), where we stress that the experimental measurements are most direct and reliable, the EM underestimates the shifts, while the PM recovers much of this error, particularly for G131D variant where the discrepancy appears to be accounted for by exclusion of mobile ions only and not through protein mediated (low dielectric) interactions as supposed by Pantoliano et al.<sup>47</sup> Comparison of the PM results verify that the effects of correlated motion among salt ions are negligible at solution ionic strength up to 0.1 M and, furthermore, indicate that the linear response approximation is also valid. Overall, the RM with a conventional hydrocarbon estimate  $\epsilon_r = 3.0$  for the protein dielectric predicts substantially larger shifts than observed, though treated as fitting parameter good agreement is obtained for G131D and E195Q with  $\epsilon_r > 20$ . Similar general remarks are also applicable for the N76D mutation predominantly effecting the high affinity binding equilibrium, though all the theoretical estimates significantly exceed the experimental value suggesting mechanisms other than metal binding are important contributors to the free energy of unfolding.

Kesvatera et al.<sup>28</sup> have determined the macroscopic calcium binding constants of calbindin spectrophotometrically by direct titration against  $\text{Ca}^{2+}$  in the presence of a chromophoric chelator with known binding characteristics. Moreover, the total equilibrium constants are quoted as a function of ionic strength in the range 2–1000 mM for native protein, together with a single (D19N), double (E17Q/D19N), and triple (E17Q/D19N/E26Q) mutant. In Figures 1–3 the experimental results for  $DpK^a$  are compared with predictions from the electrostatic continuum models. For each variant, the EM accurately reflects the ionic strength dependence of the binding constant shifts, while the PM yields satisfactory agreement up to about 100 mM KCl though underestimating the charge screening at higher salt concentrations. The general close accord of Debye–Hückel and Monte Carlo results in the PM confirms, however, that mobile ion correlations and nonlinear effects on their distribution are insignificant for monovalent salts and contribute little to the calcium binding free energy. In the finite difference calculations, the contribution to the change in chemical potential of a free calcium ion in the presence of the charged protein is not accounted for. The free energy change for binding is therefore referenced to a state where the isolated ion does not interact with the protein at all and is thus independent of its mutation. For weakly charged proteins and/or in the presence of salt, this error is probably not large.<sup>29</sup> Further, only electrostatic contributions are considered explicitly, while mobile ion excluded volume effects are only crudely approximated by invoking a Stern layer around the protein. Both these effects may be attributed to the small deviations observed in PM results reflecting, in this case, negligible differences between the statistical mechanical behavior and the electrostatic problem. With a low protein dielectric ( $\epsilon_r = 3$ ), however, accounting only for induced polarization (electronic and nuclear) and, to some extent, the thermal fluctuations of constrained permanent dipoles,<sup>73</sup> the RM systematically exceeds the PM shifts in total binding constant between 3 and 5 pK units. It should be noted that the polarization of assimilated water and the substantial effects of coupling with the solvent reaction field<sup>31</sup> are neglected, while atomistic simulation studies<sup>33,34,74</sup> have indicated perma-



**Figure 1.** Comparison of measured (○) and predicted shifts in total calcium binding constant  $DpK^a$  due to the single mutation D19N in calbindin  $D_{9k}$  as a function of monovalent salt concentration at 298.15 K. Continuum electrostatic calculations in the elementary model (—), the primitive model (---), and the refined model (- · -) are included together with statistical Monte Carlo simulation results in the primitive model (■).



**Figure 2.** As Figure 1 but for the double calbindin mutant E17Q/D19N.

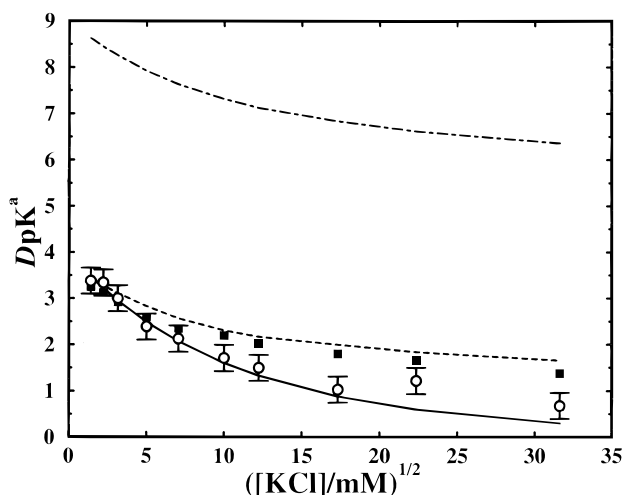
nent dipole reorientation due to conformational relaxation during the binding process implies larger values of  $\epsilon_r$ .

Taken together, the subtilisin and calbindin results exhibit two notable features: first, the apparently enviable success of the EM and, second, the substantial discrepancy between PM and RM predictions. As defined in eq 7,  $DpK^a$  is the shift in  $pK^a$  due to mutation at a *given* ionic strength. Instead, for a *given protein* (i.e., native or variant), consider the shift in  $pK^a$  due to a change in salt concentration and define

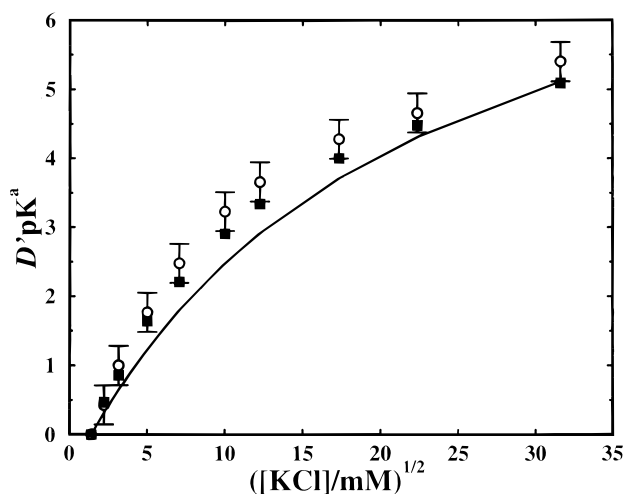
$$D'Q = Q(I) - Q(0) \quad (29)$$

for some observable  $Q = Q(I)$  dependent on ionic strength  $I$ . Figure 4 illustrates the  $D'pK^a$  shift for wild-type calbindin and is consistent with earlier conclusions<sup>28</sup> that the EM, in fact, overestimates the screening effect of salt (producing too small a pK shift) due to the increased mobile ion accessibility in the absence of an exclusion boundary around the protein. Simulation results in the PM, however, agree closely with experiment. Comparison of Figures 1 and 4 indicates a qualitative difference in the PM description of mobile solvated ions and fixed protein





**Figure 3.** As Figure 1 but for the triple calbindin mutant E17Q/D19N/E26Q.



**Figure 4.** Comparison of measured (○) and predicted shifts in total calcium binding constant  $D'pK^a$  due to ionic strength for wild-type calbindin  $D_{9k}$  as a function of monovalent salt concentration at 298.15 K. Continuum electrostatic calculations in the elementary model (—) are included together with statistical Monte Carlo simulation results in the primitive model (■).

charges. In Figure 4, showing the shift in binding constant due to salt concentration, the simulation predictions lie within the experimental error over the full range studied (up to 1.0 M). Conversely, from Figure 1 where the shift is with respect to charge mutation of the protein, the simulation results diverge from experiment with increasing ionic strength. Although this difference appears to be systematic and general, the evidence is inconclusive—the possibility of fortuitous error cancellation cannot be excluded.

It has been previously suggested<sup>27,75</sup> that the RM *per se* may not be so much at fault, but that the supposed “solutions” of it are only approximate descriptions of its true behavior. Evidently there is an asymmetry in the treatment of image charge effects. For the *fixed* protein charges, the electrostatic equations are certainly appropriate. Conversely, because the salt and counterion charges are *mobile*, a statistical mechanical problem arises. At thermal equilibrium, the motion is such that the average distribution of particles will follow the Boltzmann law (eq 19) with the potential of mean force (pmf). Ordinarily, in the PM say, where the surface charge density is small and salt concentrations/valencies are low, the pmf is well approximated

**TABLE 4: Shift in Total Calcium Binding Constant Due to Point Mutation  $DpK^a$  in Subtilisin Predicted by the Refined Model in Linear Response with Standard Parameters<sup>a</sup> Is Listed as a Function of Mobile Ion Exclusion (Stern) Layer Thickness  $\sigma$**

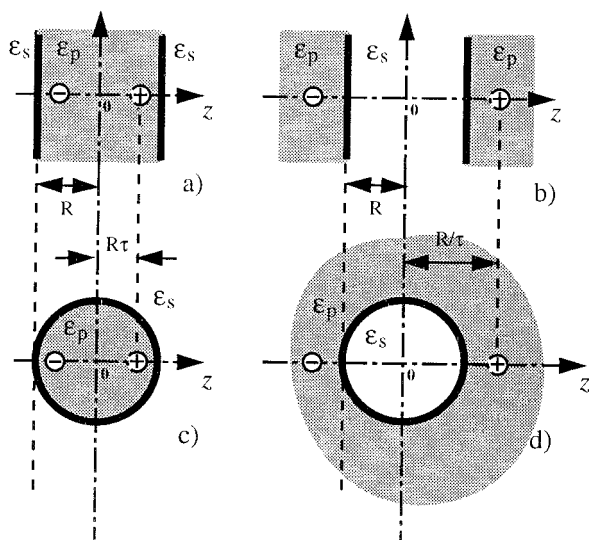
$\sigma$ (Å)	$DpK^a$			
	P172D	G131D	E195Q	N76D
1.400	−1.49	−0.53	+2.70	−3.40
2.125	−1.53	−0.56	+2.84	−3.43
3.400	−1.61	−0.59	+2.88	−3.47
5.000	−1.67	−0.61	+2.94	−3.51
7.400	−1.75	−0.65	+2.97	−3.56
10.000	−1.78	−0.67	+3.03	−3.61

<sup>a</sup> 1.4 Å solvent probe radius and low protein relative dielectric constant  $\epsilon_r = 3.0$ .

by the mean electrostatic potential. On the other hand, with a dielectric discontinuity the mobile ions will see strong correlations—at least with their own “images”, just as the fixed charges do. This suggests that the pmf could be very different from the electrostatic potential, even if all the mean field conditions appropriate to the PM are satisfied. That is, when setting up the equations describing the RM, it is *not* generally sufficient to write the mobile ion distributions in terms of the electrostatic potentials, even though these do contain effects from the surface polarization induced by the fixed charges. There is the possibility of a sizable correlation term (i.e., the self-image correlation) that should be included, or its neglect needs to be justified. The effect of this correlation, a desolvation penalty analogous to that considered by Warwicker,<sup>16</sup> would be to remove mobile ions from the neighborhood of the dielectric boundary. Indeed, for a planar geometry where the image potential can be rendered explicitly in closed form, both grand canonical<sup>76</sup> and canonical Monte Carlo simulation<sup>77</sup> as well as approximate potential analysis<sup>78</sup> and highly accurate integral equation theory (using the anisotropic hypernetted chain closure)<sup>79</sup> have demonstrated a substantial depletion of divalent mobile ions at the electrode surface, particularly at surface charge densities (averaging about  $-0.027 \text{ Cm}^{-2}$  for calbindin and  $+0.0064 \text{ Cm}^{-2}$  for subtilisin) and dielectric constant ratios ( $\epsilon_p/\epsilon_s = 1/16$  at 300 K) appropriate for proteins. In the absence of a straightforward route to directly estimate the magnitude of this effect for interfaces of arbitrary geometry, we simply consider the change in  $DpK^a$  on increasing the Stern layer thickness to mimic the repulsive image correlation as summarized in Table 4 for subtilisin. Evidently, depleting charge near the protein surface effectively lowers the solution ionic strength, reduces the local screening of the mutation charge/bound calcium ion interaction, and increases the magnitude of  $DpK^a$ . For mutations around the weak binding site (G131D, P172D, E195Q), an average 19% change in binding constant shift is observed on increasing the apparent mobile ion diameter from 0 to 8.6 Å, while for the high affinity site variant (N76D) the overall change drops to only 6%. Although not insignificant, even at the low bulk salt concentrations of physiological interest, the effect is too small to account for the PM/RM discrepancy and, in any case, would always lead to further divergence of the two predictions.

Instead of an ionic desolvation analysis (requiring an explicit study of mobile ion distributions near the protein surface), we will argue that the preponderance of the difference between PM and RM estimates of  $DpK^a$  lies in locating the dielectric boundary. Conventionally, the position of the dielectric interface is determined by the contact surface formed when a solvent probe sphere (commonly of radius 1.4 Å) is rolled over the





**Figure 5.** Geometries for (a) a planar protein slab; (b) a planar solvent pore; (c) a spherical protein; and (d) a spherical solvent cavity. The uniform relative dielectric permittivity of protein and solvent regions are  $\epsilon_p$  and  $\epsilon_s$ , respectively. An ion binding site is located at  $z_b = +R\tau$  for the convex boundaries (a and c) and at  $z_b = +R/\tau$  for the concave boundaries (b and d), where the relevant separation variable  $\tau$  lies in the range  $0 < \tau < 1$ . Similarly, the single, discrete charge of a mutable residue is located at  $z_m = -R\tau$  and  $z_m = -R/\tau$ .

macromolecule.<sup>80</sup> While it is clear that short range excluded volume interactions will contribute to the true effective potential arising from the continuum average (over both solvent degrees of freedom and those of internal protein motions) and therefore play a role in fixing the boundary, it is not obvious how these largely entropic forces enter, and we might suspect that longer ranged charge–charge interactions are likely to dominate them. Within the confines of the RM with a steplike dielectric function, there is no *a priori* reason to suppose that the optimal position of the step is related only to the protein structure and independent of any energetic criteria that are, otherwise, critically important features of electrostatic models. To examine this point in more detail, consider the idealized planar and spherical geometries (Figure 5) where the RM electrostatic problem surrenders to exact analysis. For clarity, the solvent is assumed pure with no mobile ions so nonlinear thermal effects do not enter. Only two fixed “protein charges” are included to nominally represent an ion (of charge  $q_b$ ) binding site located at  $z_b$  and a charged residue  $q_m$  at  $z_m$  that can be removed or replaced by site-directed-mutagenesis without structural perturbation. By using a multipole expansion in polar<sup>81</sup> or rectangular<sup>82</sup> coordinates, it is straightforward to show that the difference between the PM and RM results for the shift in binding free energy change due to mutation has a common form in all cases

$$\begin{aligned} \delta D\Delta G &= D\Delta G(\text{RM}) - D\Delta G(\text{PM}) \\ &= \frac{-q_b q_m}{\epsilon_0 |z_b - z_m|} \left( \frac{1}{\epsilon_p} - \frac{1}{\epsilon_s} \right) (\mathcal{A}\tau, \epsilon_p/\epsilon_s) - 1 \end{aligned} \quad (30)$$

where  $\tau$  characterizes the position of the charges relative to the dielectric interface between protein  $\epsilon_p$  and solvent  $\epsilon_s$ . Explicitly, the geometric term  $\mathcal{G}$  for each case is

$$\mathcal{A}\tau, \epsilon_p/\epsilon_s(\text{sphere}) = 2\tau \sum_{n=0}^{\infty} \frac{(-\tau^2)^n}{1 + \left(\frac{n}{n+1}\right)\frac{\epsilon_p}{\epsilon_s}} \quad (31)$$

$$\mathcal{A}\tau, \epsilon_p/\epsilon_s(\text{cavity}) = 2\tau \sum_{n=1}^{\infty} \frac{(-\tau^2)^n}{1 + \left(\frac{n+1}{n}\right)\frac{\epsilon_p}{\epsilon_s}} \quad (32)$$

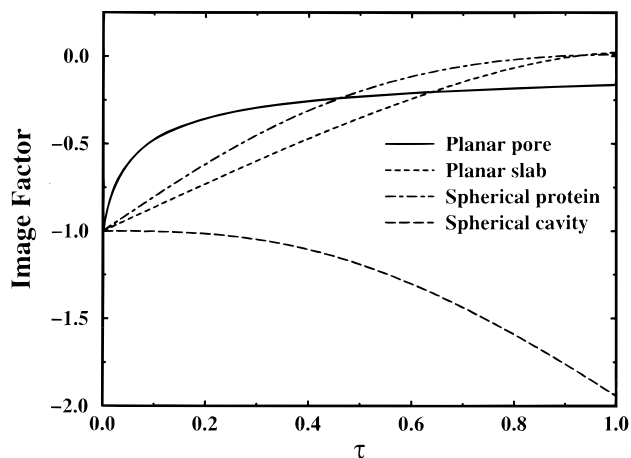
$$\mathcal{A}\tau, \epsilon_p/\epsilon_s(\text{slab}) = \left( \frac{2\tau}{1 - \frac{\epsilon_p}{\epsilon_s}} \right) \sum_{n=0}^{\infty} \zeta^{2n} \left( \frac{\zeta}{2n+1} - \frac{2n}{(2n)^2 - \tau^2} \right) \quad (33)$$

$$\mathcal{A}\tau, \epsilon_p/\epsilon_s(\text{pore}) = \left( \frac{2\tau}{1 - \frac{\epsilon_p}{\epsilon_s}} \right) \sum_{n=1}^{\infty} \frac{\zeta^{2n}}{(1 + 2n\tau)(1 + 2(n-1)\tau)} \quad (34)$$

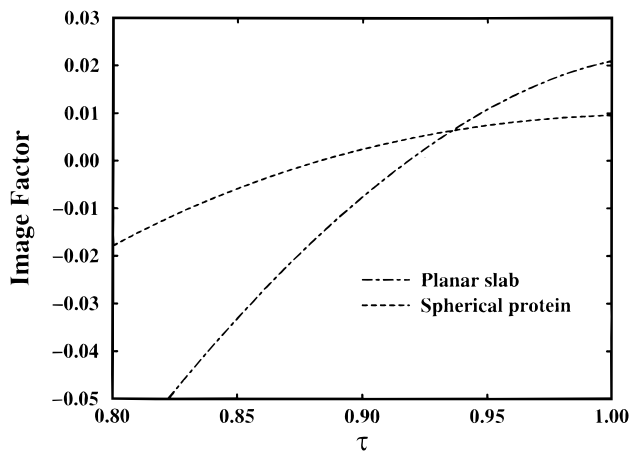
with

$$\zeta = \frac{1 - \frac{\epsilon_p}{\epsilon_s}}{1 + \frac{\epsilon_p}{\epsilon_s}} \quad (35)$$

The three factors in eq 30 can each be assigned a clear physical significance: the first is just the Coulomb energy associated with the charges in vacuum; the second accounts for embedding the charges in dielectric media (initially in protein, followed by encapsulation in solvent); while the third contains only geometry specific information. It is tempting to postulate that a similar partition of distinct mechanisms also holds, at least locally, for a general boundary geometry. As the charges recede from the interface,  $\mathcal{G}$  tends to zero and we recover the Born free energy for inserting a low dielectric protein into a high dielectric solvent continuum. This term obviously vanishes in the PM where there is no cost in polarization energy to solvate the protein. The direct Coulomb interaction energy in vacuum between a divalent ion and a monovalent mutation corresponds to a colossal shift in binding constant ( $DpK^a$  of order  $10^3$  for a charge separation in the range 10–100 Å). To recover experimentally observed  $DpK^a$ 's (of order unity) clearly requires the remaining factors in eqn 30 be of order  $10^{-3}$ . Given a typical dielectric ratio  $\epsilon_p/\epsilon_s = 1/40$  the last “image” factor in eq 30 (i.e.,  $\mathcal{G} - 1$ ) is plotted in Figure 6 for each example as a function of the dimensionless geometric variable  $\tau$  where we observe that the cases with “line-of-sight” charge interaction mainly through solvent (cavity and pore) have a strictly negative image factor for all values of  $\tau$ . On the other hand, for largely through protein interactions the image factor changes sign as the charges approach the dielectric boundary (i.e.,  $\tau \rightarrow 1$ ) as detailed in Figure 7. For essentially convex surfaces, then, it should be possible to eliminate the difference between PM and RM  $DpK^a$ 's (i.e.,  $\delta D\Delta G = 0$ ) by a suitable location of the dielectric interface, regardless of the relative permittivities assigned to protein and solvent. This is demonstrated for the  $DpK^a$  of calbindin in Figure 8 where a uniform decrease in solvent probe radius (including water penetration of the protein



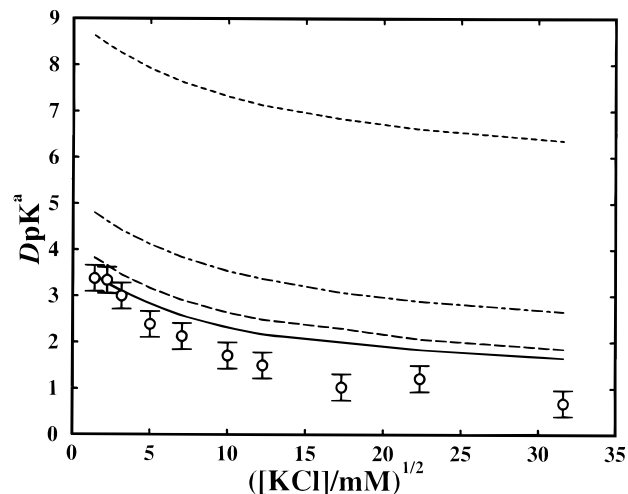
**Figure 6.** The image factor ( $\mathcal{A}(\tau, \epsilon_p/\epsilon_s) - 1$ ) contribution to the difference between primitive and refined continuum model predictions for the shift in binding free energy due to mutation, defined by eq 30, as a function of separation variable  $\tau$  for the dielectric boundary geometries illustrated in Figure 5; a planar solvent pore (—); a planar protein slab (---); a protein sphere (- · -); and a spherical solvent cavity (···). The ratio of relative dielectric permittivities for protein to solvent is  $\epsilon_p/\epsilon_s = 1/40$ .



**Figure 7.** Detail from Figure 6 for the convex boundaries, planar protein slab, and protein sphere.

surface corresponding to a negative probe radius) yields a uniform decrease in  $\delta D\Delta G$ .

Evidence for a second low affinity calcium binding site in subtilisin was first observed serendipitously by Pantoliano et al.<sup>83</sup> in a study primarily aimed at engineering enhanced thermostability of the enzyme by introducing a disulfide linkage through cysteine mutations T22C and S87C. At  $\text{Ca}^{2+}$  concentrations above 0.1 mM the low affinity site is occupied leading to a decreased relative stability of the mutant and implying that binding at this site causes sufficient conformational change to strain the disulfide bond. Subsequent work,<sup>49,50</sup> on the contrary, has shown that the presence of weakly bound calcium substantially increases the melting temperature, leading to a dramatic stability enhancement overall. A two-step process is implicated<sup>49</sup> where the ion binding first gives rise to a partially unfolded, though still highly structured, intermediate form that is more heat resistant than native enzyme in the absence of extraneous calcium. Detailed solution NMR studies<sup>84,85</sup> have demonstrated that, in accord with its putative function as an intracellular buffer with an efficient calcium binding mechanism, calbindin undergoes only minor structural reorganization to accommodate a  $\text{Ca}^{2+}$  ion at the C-terminal site and essentially



**Figure 8.** As Figure 3 but comparing finite difference calculations in the primitive model (—) and refined model with varying solvent probe radii of 1.4 Å (---), 0.0 Å (- · -), and -0.5 Å (···). Experimental results are denoted by the symbols (O).

no conformational change at the N-terminal site where the EF-hand loop is preformed in the apo protein. Compared to regulatory proteins (e.g., calmodulin and troponin C) which require extensive rearrangement and subsequent hydrophobic exposure for activity, the structural consequences of calcium binding in calbindin are greatly attenuated. For both subtilisin and calbindin (indeed, for any real biomolecule), the impact of structural perturbations in solution on binding or ionization events mediated by Coulomb forces cannot be discounted out of hand. In the context of a continuous protein medium without detailed covalent structure, the very substantial forces associated with charge interactions through a low dielectric material may drive conformational change, therefore requiring an electrostatic mechanism beyond the PM (e.g., reorganization energy of permanent protein dipoles<sup>86</sup>). The RM analysis here indicates that the relative displacement of protein charges with respect to each other is less significant than the position of the effective dielectric interface relative to the charges. Poor agreement between standard RM predictions and experiment do not, therefore, necessarily imply substantial structural relaxation of the protein.

#### 4. Concluding Remarks

It seems probable that the apparent predictive accuracy of the EM (at least for the calcium  $\text{DpK}^a$  in calbindin) results, in large measure, from a fortuitous cancellation of error. Rather like adding higher order terms in a nonvariational perturbation expansion, incorporating other mechanisms (i.e., mobile ion exclusion by protein in the PM and solvent exclusion in the RM) *ad hoc* does not necessarily guarantee improved predictions in general, particularly if the corresponding theory is not exact. Accepting this caveat, however, two mechanisms have been identified, not addressed in any of the models studied here but originating from charge interactions, that may subtly contribute to the binding of ions to proteins:

1. The “self-image” correlation associated with a dielectric interface giving rise to a desolvation free energy penalty that lowers the local ionic strength of the solution in the protein neighborhood, thereby unscreening the protein charges.
2. The location of a simple dielectric boundary that stems from an effective charge interaction mediated by solvent and protein therefore determined by energetic criteria in addition

to entropic excluded volume considerations. In these terms, it is not obvious that the *effective* surface of polarization (i.e., the dielectric interface) should necessarily coincide with the solvent exclusion boundary of the solute.

We might speculate that accounting for these effects in an extension of the RM could resolve the remaining discrepancy between theory and experiment, quite possibly gaining no further accuracy over the EM but certainly providing a deeper insight on protein behavior.

**Acknowledgment.** We are grateful to Bo Svensson for making the Monte Carlo simulation program available to us. Enlightening discussions with Richard Pickersgill, John Jenkins, and Eva Thulin have been of benefit. The U.K. Biotechnology and Biological Sciences Research Council provided financial support.

## References and Notes

- Honig, B.; Nicholls, A. *Science* **1995**, 268, 1144.
- Gilson, M. K. *Curr. Opin. Struct. Biol.* **1995**, 5, 216.
- Linse, S.; Jönsson, B.; Chazin, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 4748.
- Perutz, M. F. *Science* **1978**, 201, 1187.
- Warshel, A. *Acc. Chem. Res.* **1981**, 14, 284.
- Penfold, R.; Abbas, S.; Nordholm, S. *Fluid Phase Equilib.* **1996**, 120, 39.
- McMillan, W. G.; Mayer, J. E. *J. Chem. Phys.* **1945**, 13, 276.
- Friedman, H. L.; Dale, W. D. T. In *Modern Theoretical Chemistry*, Berne, B. J., Ed.; Plenum Press: New York, 1975; Vol. 5, *Statistical Mechanics*; p 85.
- Friedman, H. L. *Ann. Rev. Phys. Chem.* **1981**, 32, 179.
- Friedman, H. L. *J. Chem. Phys.* **1982**, 76, 1092.
- Tanford, C.; Kirkwood, J. G. *J. Am. Chem. Soc.* **1957**, 79, 5333.
- To motivate the appellation "refined" model, note that the term "civilized" has been applied (Carnie, S. L.; Chan, D. Y. C. *J. Chem. Phys.* **1980**, 73, 2949; Blum, L.; Henderson, D. *J. Chem. Phys.* **1981**, 74, 1902; Torrie, G. M.; Kusalik, P. G.; Patey, G. N. *J. Chem. Phys.* **1988**, 88, 7826.) to models including an explicit atomic solvent of hard spheres with embedded point dipoles. Apparently coined by R. H. Stokes (see: Carnie, S. L.; Chan, D. Y. C. *J. Chem. Soc., Faraday Tras. 2* **1982**, 78, 69.5), a civilized model electrolyte treats ions and solvent on an equal basis. Following this line, we suggest that accounting in some way for the different molecular nature of solvent and solute is an advance on the primitive model, but retaining a continuum picture remains, in some sense, unsophisticated.
- For a planar dielectric boundary, it is relatively trivial to compute the surface polarization charge (the so-called "image" charge term) and a number of studies have analyzed this situation using simulation methods and sophisticated integral equation theories (e.g., see refs 76–79). Some simulation work has addressed the case of a spherical boundary, where exact solutions of the Poisson equation exist but are expensive to compute (e.g., ref 75 and Linse, P. *J. Phys. Chem.* **1986**, 90, 6821).
- Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, 157, 671.
- In this work, the term "model" is synonymous with a Hamiltonian, that is a definition of particles (the short range interaction potential) and the forces between them (the long range part). The FDPB is not a model in this sense since metaphysical approximations (i.e., the mathematical mean-field ansatz) have been raised to the level of axiom.
- Warwicker, J. *J. Mol. Biol.* **1994**, 236, 887.
- Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, 89, 5588. Rashin, A. A.; Nambodiri, K. *J. Phys. Chem.* **1987**, 91, 6003.
- Demchuk, E.; Wade, R. C. *J. Phys. Chem.* **1996**, 100, 17373.
- Warwicker, J. *Protein Eng.* **1997**, 10, 809.
- Soman, K.; Yang, A.-S.; Honig, B.; Fletterick, R. *Biochemistry* **1989**, 28, 9918.
- Resat, H.; Marrone, T. J.; McCammon, J. A. *Biophys. J.* **1997**, 72, 522.
- Muegge, I.; Schweins, T.; Langen, R.; Warshel, A. *Structure* **1996**, 4, 475.
- Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol.* **1994**, 238, 415. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, 35, 7819.
- Juffer, A. H.; Argos, P.; Vogel, H. J. *J. Phys. Chem. B* **1997**, 101, 7664.
- Warwicker, J.; Gane, P. J. *FEBS Lett.* **1996**, 385, 105.
- Gane, P. J.; Freedman, R. B.; Warwicker, J. *J. Mol. Biol.* **1995**, 249, 376.
- Svensson, B.; Jönsson, B.; Woodward, C. E.; Linse, S. *Biochemistry* **1991**, 30, 5209.
- Kesvatera, T.; Jönsson, B.; Thulin, E.; Linse, S. *Biochemistry* **1994**, 33, 14170.
- Svensson, B.; Jönsson, B.; Thulin, E.; Woodward, C. E. *Biochemistry* **1993**, 32, 2828.
- Warshel, A.; Russell, S. T. *Q. Rev. Biophys.* **1984**, 17, 283.
- Warshel, A.; Åqvist, J. *Ann. Rev. Biophys. Biophys. Chem.* **1991**, 20, 267.
- Warshel, A.; Chu, Z. T. In *ACS Symp. Ser.: Structure and Reactivity in Aqueous Solution*; Cramer, D. G., Ed.; American Chemical Society: Washington, DC, 1994; p 71.
- Muegge, I.; Qi, P. X.; Wand, A. J.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, 101, 825.
- Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, 101, 4458.
- Florián, J.; Warshel, A. *J. Phys. Chem. B* **1997**, 101, 5583.
- Fushiki, M.; Svensson, B.; Jönsson, B.; Woodward, C. E. *Biopolymers* **1991**, 31, 1149.
- Svensson, B.; Woodward, C. E. *J. Phys. Chem.* **1995**, 99, 1614.
- Shaw, P. B. *Phys. Rev. A*, **1985**, 32, 2476; **1987**, 35, 2254.
- Juffer, A. H.; Botta, E. F. F.; van Keulen, B. A. M.; van der Ploeg, A.; Berendsen, H. J. C. *J. Comp. Phys.* **1991**, 97, 144.
- Wang, L.; Hermans, J. *J. Phys. Chem.* **1995**, 99, 12001.
- Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, CA, 1996; Chapter 1.
- Jernigan, R.; Raghunathan, R.; Bahar, I. *Curr. Opin. Struct. Biol.* **1994**, 4, 256.
- Russell, A. J.; Fersht, A. R. *Nature* **1987**, 328, 496.
- Liljas, A.; Rossmann, M. G. *Ann. Rev. Biochem.* **1974**, 43, 475.
- Warshel, A.; Naray-Szabo, G.; Sussman, F.; Hwang, J.-K. *Biochemistry* **1989**, 28, 3629.
- Siezen, R. J.; de Vos, W. M.; Leunissen, J. A. M.; Dijkstra, B. W. *Protein Eng.* **1991**, 4, 719.
- Pantoliano, M. W.; Whitlow, M.; Wood, J. F.; Rollence, M. L.; Finzel, B. C.; Gilliland, G. L.; Poulos, T. L.; Bryan, P. N. *Biochemistry* **1988**, 27, 8311.
- Jang, J. S.; Bae, K. H.; Byun, S. M. *Biochem. Biophys. Res. Comm.* **1992**, 188, 184.
- Genov, N.; Filippi, B.; Dolashka, P.; Wilson, K. S.; Betzel, C. *Int. J. Peptide Protein Res.* **1995**, 45, 391.
- Kidd, R. D.; Yennawar, H. P.; Sears, P.; Wong, C.-H.; Farber, G. K. *J. Am. Chem. Soc.* **1996**, 118, 1645.
- Rasmussen, H. *N. Eng. J. Med.* **1986**, 314, 1094; **1986**, 314, 1164.
- Linse, S.; Brodin, P.; Johansson, C.; Thulin, E.; Grundström, T.; Forsén, S. *Nature* **1988**, 335, 651.
- Moews, P. C.; Kretsinger, R. H. *J. Mol. Biol.* **1975**, 91, 201.
- Bryan, P. N.; Alexander, P.; Strausberg, S.; Schwarz, F.; Lan, W.; Gilliland, G. L.; Gallagher, D. T. *Biochemistry* **1992**, 31, 4937.
- McPhalen, C. A.; James, M. N. G. *Biochemistry* **1988**, 27, 6583.
- Szebenyi, D. M. E.; Moffat, K. J. *Biol. Chem.* **1986**, 261, 8761.
- Pantoliano, M. W.; Whitlow, M.; Wood, J. F.; Dodd, S. W.; Hardman, K. D.; Rollence, M. L.; Bryan, P. N. *Biochemistry* **1989**, 28, 7205.
- Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, 21, 1087.
- Widom, B. *J. Chem. Phys.* **1963**, 39, 2808.
- Svensson, B.; Woodward, C. E. *Mol. Phys.* **1988**, 64, 247.
- Card, D. N.; Valleau, J. P. *J. Chem. Phys.* **1970**, 52, 6232.
- Zauhar, R. J.; Morgan, R. S. *J. Comput. Chem.* **1990**, 11, 603.
- Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, 12, 435.
- Yoon, B. J.; Lenhoff, A. M. *J. Phys. Chem.* **1992**, 96, 3130.
- Holst, M.; Saied, F. *J. Comput. Chem.* **1993**, 14, 105.
- Zhou, H.-X. *Biophys. J.* **1993**, 65, 955.
- Harrison, R. W.; Kourinov, I. V.; Andrews, L. C. *Protein Eng.* **1994**, 7, 359.
- Alvarez-Ramirez, J.; Martinez, R.; Diaz-Herrera, E. *Chem. Phys. Lett.* **1997**, 266, 375.
- Penfold, R.; Jönsson, B.; Nordholm, S. *J. Chem. Phys.* **1993**, 99, 497.
- Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, 9, 327.
- Takahashi, T.; Sturtevant, J. M. *Biochemistry* **1981**, 20, 6185.
- Volkin, D. B.; Klibanov, A. M. In *Protein Function: A Practical Approach*; Creighton, T. E., Ed.; IRL Press: Oxford, 1989; p 1.
- Gilson, M. K.; Honig, B. *Biopolymers* **1986**, 25, 2097.
- Sham, Y. Y.; Muegge, I.; Warshel, A. *Biophys. J.* **1998**, 74, 1744.
- Woodward, C. E.; Svensson, B. *J. Phys. Chem.* **1991**, 95, 7471.
- Torrie, G. M.; Valleau, J. P.; Patey, G. N. *J. Chem. Phys.* **1982**, 76, 4615.



- (77) Bratko, D.; Jönsson, B.; Wennerström, H. *Chem. Phys. Lett.* **1986**, 128, 449.
- (78) Outhwaite, C. W.; Bhuiyan, L. B. *J. Chem. Soc., Faraday Trans. 2* **1982**, 78, 775; **1983**, 79, 707.
- (79) Kjellander, R.; Marčelja, S. *Chem. Phys. Lett.* **1984**, 112, 49.
- (80) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, 55, 379.
- (81) Kirkwood, J. G. *J. Chem. Phys.* **1934**, 2, 351.
- (82) Barrera, R. G.; Guzman, O.; Balaguer, B. *Am. J. Phys.* **1978**, 46, 1172.
- (83) Pantoliano, M. W.; Ladner, R. C.; Bryan, P. N.; Rollence, M. L.; Wood, J. F.; Poulos, T. L. *Biochemistry* **1987**, 26, 2077.
- (84) Skelton, N. J.; Kördel, J.; Akke, M.; Forsén, S.; Chazin, W. J. *Struct. Biol.* **1994**, 1, 239.
- (85) Skelton, N. J.; Kördel, J.; Chazin, W. J. *J. Mol. Biol.* **1995**, 249, 441.
- (86) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, 14, 161.