Solvent Models for Protein–Ligand Binding: Comparison of Implicit Solvent Poisson and Surface Generalized Born Models with Explicit Solvent Simulations

LINDA YU ZHANG,[1] EMILIO GALLICCHIO,[1] RICHARD A. FRIESNER,[2] RONALD M. LEVY[1]

[1]*Chemistry Department, Rutgers, The State University of New Jersey, 610 Taylor Road, Piscataway, New Jersey 08854*
[2]*Department of Chemistry, Columbia University, New York, New York 10027*

Received 13 July 2000; accepted 26 October 2000

ABSTRACT: Solvent effects play a crucial role in mediating the interactions between proteins and their ligands. Implicit solvent models offer some advantages for modeling these interactions, but they have not been parameterized on such complex problems, and therefore, it is not clear how reliable they are. We have studied the binding of an octapeptide ligand to the murine MHC class I protein using both explicit solvent and implicit solvent models. The solvation free energy calculations are more than 10^3 faster using the Surface Generalized Born implicit solvent model compared to FEP simulations with explicit solvent. For some of the electrostatic calculations needed to estimate the binding free energy, there is near quantitative agreement between the explicit and implicit solvent model results; overall, the qualitative trends in the binding predicted by the explicit solvent FEP simulations are reproduced by the implicit solvent model. With an appropriate choice of reference system based on the binding of the discharged ligand, electrostatic interactions are found to enhance the binding affinity because the favorable Coulomb interaction energy between the ligand and protein more than compensates for the unfavorable free energy cost of partially desolvating the ligand upon binding. Some of the effects of

Correspondence to: R. M. Levy; e-mail: ronlevy@lutecee.rutgers.edu

Contract/grant sponsor: NIH; contract/grant numbers: GM30580, GM52019, and RR06892

Computation time was obtained at NCSA via the NPACI program supported by the National Science Foundation

Contract/grant sponsor: Center for Biomolecular Simulations at Columbia University

protein flexibility and thermal motions on charging the peptide in the solvated complex are also considered. J Comput Chem 22: 591–607, 2001

Keywords: implicit solvent models; protein–ligand binding; protein electrostatics; generalized Born model; free energy perturbation

Introduction

Computer simulations have provided the basis for much of our molecular level understanding of solvation thermodynamics. At the most detailed level, the simulations include explicit molecular representations of the solvent. This approach is, in principle, the most realistic model for studying the physical chemistry of solvation. The treatment of electrostatic properties using fully atomistic simulations is difficult, because of the long-range character of the Coulomb interactions. Although it is well known that truncating Coulomb interactions can result in large errors in estimates of thermodynamic parameters, only recently have alternative methods been widely adopted in biomolecular simulations. In particular, there has been significant progress in theoretical and numerical developments associated with the use of the periodic Coulomb potential.[1–7] Even with much faster algorithms for carrying out simulations with explicit solvent,[8–10] the computational expense is too great for many simulations suggested by current problems in structural biology, which are poised to take advantage of large databases of macromolecular structures, for example, problems in drug design and structural genomics. Hence, the extensive effort to develop implicit solvation models for biomolecules, which are physically reasonable, but are still very fast.

A great deal of attention has been given in recent years to the use of continuum solvent models for biomolecular simulations. In these models, based on the Poisson equation or approximations to it, the solute is described in atomic detail, but the solvent is replaced by a dielectric continuum.[11–22] The standard approach to parameterizing continuum solvent electrostatic models has been to adjust nonbonded parameters to fit experimental solvation free energies for a database of small organic molecules and/or amino acids and nucleic acids. Several groups have reported very good results fitting the experimental solvation data.[23–30] Such comparisons between continuum model predictions and experiment have been much more extensive than has been possible for more complex solvation problems, such as those involving conformational equilibria and molecular association. This is due in part to the relative paucity of experimental data to benchmark problems of this kind on small systems and the difficulty in isolating the "pure" electrostatic component of the experimental measurements. For these more complex solvation problems, it is very informative to treat the results of Free Energy Perturbation (FEP) simulations with explicit solvent as the experimental data; then the explicit/implicit solvent approaches can be analyzed and compared. There are relatively few systems for which the results of continuum solvent and explicit solvent simulations have been directly compared.[4] A better appreciation of the complementary strengths and weaknesses of the different solvation models is likely to be obtained from studies of this kind.

We have been working on the comparison of continuum with explicit solvent simulations of a series of benchmark problems; including the solvation free energy of a large number of organic solutes, the conformational preferences of peptides, peptide hydrogen bonding in solution, and ligand binding to a protein. We focus on the later problem in this report. We have studied the binding of the OVA-8 peptide to the murine MHC class I protein H-2KB. MHC proteins recognize antigenic peptides as part of the cellular immune response system. We analyze the electrostatic contribution to binding of peptide to MHC protein from both continuum and explicit solvent perspectives. The results of continuum electrostatic calculations for MHC protein–peptide complexes were recently reported by Honig and collaborators.[31] These authors described a method to calculate the binding free energy (ΔG) of a protein–ligand complex using a continuum model of the solvent. They found generally that electrostatic interactions oppose binding. As discussed below, the conclusions concerning the contribution of electrostatics to ligand affinity depend on the reference states chosen for constructing the thermodynamic cycles associated with the binding. With respect to the binding of the completely discharged ligand as the reference state, we find that electrostatic terms generally enhance the bind-

ing. We note that this result is consistent with the recent analysis of Tidor, in which it is found that electrostatics can enhance both affinity and specificity simultaneously.[32, 33]

In the following section we review the computational methodology. For the continuum solvent calculations, both finite difference Poisson–Boltzmann and the Generalized Born models were employed; we review the connection between these models. In the Results and Discussion, we compare the results of the continuum solvent calculations of the solvation free energy of the peptide and the protein–ligand complex with the corresponding results obtained from FEP simulations with explicit solvent. Remarkably, we find that there is near quantitative agreement between the continuum and explicit solvent estimates of the solvation free energy of the peptide–ligand complex in water; however, the agreement is not as good for the protein–ligand complex. Issues related to the effects of protein flexibility are also discussed.

Computational Methods

All the calculations with explicit and implicit solvent were carried out with the IMPACT molecular modeling program package.[2] Most of the continuum solvent calculations were performed using the Surface Generalized Born (SGB) approximation,[20] as implemented within IMPACT; the results were compared with finite element solutions to the Poisson equation,[26, 34] also implemented within IMPACT, and with Free Energy Perturbation (FEP) results using explicit solvent. The OPLS all atom force field was used for all the calculations.[35] The OPLS Lennard–Jones radii were reparameterized for use with the SGB model by fitting the electrostatic component of the solvation free energies of a series of 40 organic molecules in water to the results of FEP simulations with explicit solvent.[36] No additional parameterization was performed for the PBF calculations with OPLS.

CONTINUUM SOLVENT CALCULATIONS

There are many models for treating solvation. The most detailed is a solvent model that includes the solvent molecules explicitly. Explicit solvent models[37–40] employing hundreds or thousands of solvent molecules have been the most widely used method for carrying out free energy simulations in liquid environments. However, this is computationally very expensive, and therefore, not suitable for "high throughput"-type modeling. Implicit Solvent models[16] treat the solvent as a continuous medium having the average properties of the real solvent. Generally, a solute is defined as a cavity with the shape of the solute and low dielectric constant that is embedded in a dielectric continuum.

The most rigorous continuum model is the Poisson–Boltzmann (PB) equation, which is used to find the resulting electrostatic potential, $\phi(\mathbf{r})$, at point $\mathbf{r}$ from the charge density, $\rho(\mathbf{r})$, and the dielectric constant, $\epsilon(\mathbf{r})$, of the system:

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \tag{1}$$

The PBF method used in this report is a PB solver by the finite difference method.[34] Although must faster than an explicit solvation calculation, solving the PB equation can be at least an order of magnitude slower than just calculating the molecular mechanics energies: the van der Waals and bare Coulombic interactions. This is still too slow for high throughput problems—for the docking of a large database of ligands to a protein receptor, for example. Faster methods are needed for these purposes. The next steps are approximations to the PB equation such as the Generalized Born (GB) model by Still and coworkers[12] and a modification called the surface generalized Born (SGB) method by Ghosh and coworkers.[20] These methods are able to calculate the solvation energies for a complex of a ligand and a receptor in a matter of seconds. They are described in more detail below. The simplest case, which has an analytical solution, is the Born model, in which the electrostatic solvation free energy of a charged ion is modeled by the transfer energy of a charged spherical shell from gas phase to solution. The dielectric constants of the solute and the solvent are denoted as ϵ_i and ϵ_o in this report.

The Born Model

Based on classical electrostatic theory,[41, 42] the total electrostatic energy in dielectric media is defined as,

$$G = \frac{1}{8\pi} \int \mathbf{E}(\mathbf{r}) \cdot \mathbf{D}(\mathbf{r})\, d^3\mathbf{r} \tag{2}$$

$$\mathbf{D}(\mathbf{r}) = \epsilon \mathbf{E}(\mathbf{r}) \tag{3}$$

where $\mathbf{E}$ and $\mathbf{D}$ are the electric field and electric displacement, ϵ is the dielectric constant of the dielectric medium. The $\mathbf{E}$ and $\mathbf{D}$ can be obtained by

Gauss's law, which is

$$\int_S \mathbf{D}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})\, d^2\mathbf{r} = \int_V 4\pi\rho\, d^3\mathbf{r} \tag{4}$$

or

$$\int_S \mathbf{E}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})\, d^2\mathbf{r} = \int_V 4\pi \frac{\rho}{\epsilon}\, d^3\mathbf{r} \tag{5}$$

where the left integral is the area integral over any closed surface, the right integral is the volume integral over the space inclosed by the surface, $\mathbf{n}(\mathbf{r})$ is the norm of the surface, and ρ is the free or macroscopic charge density (not including the induced or polarized charge density).

For the system of a uniformly charged spherical shell with dielectric constant ϵ_i inside and outside, considering the symmetry of the system, it is easy to obtain the electric field and electric displacement inside and outside the sphere, which are

$$\mathbf{E}_{\text{in}} = 0, \qquad \mathbf{D}_{\text{in}} = 0, \tag{6}$$

$$\mathbf{E}_{\text{out}} = \frac{q}{\epsilon_i r^3}\mathbf{r}, \qquad \mathbf{D}_{\text{out}} = \frac{q}{r^3}\mathbf{r} \tag{7}$$

where q is the total charge of the sphere, and we set the center of the coordinate at the center of the sphere.

The total electrostatic energy of the system is

$$G_1 = \frac{1}{8\pi}\int \mathbf{E}\cdot\mathbf{D}\, d^3\mathbf{r} \tag{8}$$

$$= \frac{1}{8\pi}\left[\int_{\text{in}} \mathbf{E}_{\text{in}}\cdot\mathbf{D}_{\text{in}}\, d^3\mathbf{r} + \int_{\text{out}} \mathbf{E}_{\text{out}}\cdot\mathbf{D}_{\text{out}}\, d^3\mathbf{r}\right] \tag{9}$$

$$= \frac{1}{8\pi\epsilon_i}\int \mathbf{E}_{\text{out}}\frac{q^2}{r^4}\, d^3\mathbf{r} \tag{10}$$

$$= \frac{q^2}{2\epsilon_i\alpha} \tag{11}$$

where α is the radius of the sphere.

For the system of a uniformly charged sphere with dielectric constant inside ϵ_i and outside ϵ_o, similar to the above derivation, the electric field and electric displacement inside and outside the sphere are

$$\mathbf{E}_{\text{in}} = 0, \qquad \mathbf{D}_{\text{in}} = 0, \tag{12}$$

$$\mathbf{E}_{\text{out}} = \frac{q}{\epsilon_o r^3}\mathbf{r}, \qquad \mathbf{D}_{\text{out}} = \frac{q}{r^3}\mathbf{r} \tag{13}$$

The total electrostatic energy of the system is

$$G_2 = \frac{q^2}{2\epsilon_o\alpha} \tag{14}$$

Hence, we can write the electrostatic free energy of transferring a spherical charged ion from a medium of dielectric constant ϵ_i to a medium of dielectric constant ϵ_o as the energy difference of the above two systems,

$$\Delta G_{\text{Born}} = \left(\frac{1}{\epsilon_o} - \frac{1}{\epsilon_i}\right)\frac{q^2}{2\alpha} \tag{15}$$

$$= -\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\frac{q^2}{2\alpha} \tag{16}$$

This equation is the famous Born equation,[43] which was first derived by setting ϵ_i equal to the vacuum dielectric constant 1. It is easy to prove that the Born equation is also valid for the case of a point charge q located in the center of a sphere with radius α.

Generalized Born Equation

We can generalize the Born model to any solute molecule with arbitrary cavity shape, where the atoms of the molecule are modeled as a set of small charged spheres with charge q_i and radius α_i or point charges in the center of the spheres with dielectric constant ϵ_i inside the spheres. First, we assume that theses spheres are so far away from each other that they look to each other like point charges. Then the total electrostatic free energy G_{tot} of this system in a medium of dielectric constant ϵ_o is given by the sum of the Coulombic interaction energy and the self-energies of the spheres, which can be decomposed into the total electrostatic energy (ΔG^0) of a set of spheres embedded in a uniform dielectric medium with dielectric constant ϵ_i, and the transfer free energy or the electrostatic free energy of solvation ΔG_{solv}.

$$G_{\text{tot}} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1\, j\neq i}^{n}\frac{q_i q_j}{\epsilon_o r_{ij}} + \frac{1}{2}\sum_{i}^{n}\frac{q_i^2}{\epsilon_o\alpha_i} \tag{17}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1\, j\neq i}^{n}\frac{q_i q_j}{\epsilon_i r_{ij}} - \frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i=1}^{n}\sum_{j=1\, j\neq i}^{n}\frac{q_i q_j}{r_{ij}} + \frac{1}{2}\sum_{i}^{n}\frac{q_i^2}{\epsilon_i\alpha_i} - \frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i}^{n}\frac{q_i^2}{\alpha_i} \tag{18}$$

$$= \left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1\, j\neq i}^{n}\frac{q_i q_j}{\epsilon_i r_{ij}} + \frac{1}{2}\sum_{i}^{n}\frac{q_i^2}{\epsilon_i\alpha_i}\right] \times \left[-\frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i=1}^{n}\sum_{j=1\, j\neq i}^{n}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i}^{n}\frac{q_i^2}{\alpha_i}\right] \tag{19}$$

$$\Delta G^0 = \left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1\,j\neq i}^{n}\frac{q_i q_j}{\epsilon_i r_{ij}} + \frac{1}{2}\sum_{i}^{n}\frac{q_i^2}{\epsilon_i \alpha_i}\right] \quad (20)$$

$$\Delta G_{\text{solv}} = -\frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i=1}^{n}\sum_{j=1\,j\neq i}^{n}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i}^{n}\frac{q_i^2}{\alpha_i} \quad (21)$$

The above expression for ΔG_{solv} is also valid for a set of uncharged spheres with point charges located in the center; however, in that case, ΔG_0 will only include the first term, because the self-energy of a point charge is infinite, and not included in the total electrostatic energy. For a real molecule with an arbitrary-shaped cavity formed by the molecular surface, the above expression for ΔG_{solv} needs to be modified. The approximation proposed by Still and coworkers,[12] which is called the Generalized Born (GB) equation, is given as

$$\Delta G_{\text{GB}} = -\frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-D}}} \quad (22)$$

$$= -\frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i=1}^{n}\sum_{j=1\,j\neq i}^{n}\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-D}}} - \frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i}^{n}\frac{q_i^2}{\alpha_i} \quad (23)$$

$$= \Delta G_{\text{pair}} + \Delta G_{\text{single}} \quad (24)$$

$$\Delta G_{\text{pair}} = -\frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i=1}^{n}\sum_{j=1\,j\neq i}^{n}\frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_{ij}^2 e^{-D}}} \quad (25)$$

$$\Delta G_{\text{single}} = -\frac{1}{2}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\sum_{i}^{n}\frac{q_i^2}{\alpha_i} \quad (26)$$

where $\alpha_{ij} = \sqrt{\alpha_i \alpha_j}$ and $D = r_{ij}^2/(2\alpha_{ij})^2$. ΔG_{single} is the self-energy, which has exactly the same form as the Born equation [eq. (16)]. The expression for the pair energy ΔG_{pair} provides the right limits; it gives Coulomb energy as $r_{ij} \to \infty$, and reduces to the self-energy or the Born equation as $r_{ij} \to 0$.

In the application of the GB equation to obtain the electrostatic solvation free energy, the major effort is the calculation of the effective Born radius α_i. The idea is to calculate the solvation free energy of a single point charge representing an atom in the cavity of the whole solute molecule, assuming that all other atoms are neutral and just displace the dielectric, then use the self-energy expression [eq. (26)] to *define* the effective Born radius. This can be carried out by calculating the surface integral of the interaction of the fixed point charge and the induced surface charge,[20] based on the boundary element formulation of the Poisson–Boltzmann (PB) equation.[44]

The Surface Generalized Born (SGB) Model

The SGB model is an approximation to the boundary element formulation of the Poisson–Boltzmann (PB) equation, in which the polarization effects throughout the entire volume of the system can be exactly reproduced by an appropriate distribution of induced polarization charge at the dielectric boundary; the electric field generated by the induced charge is called the reaction field.[44, 45] For a solute molecule, modeled by a set of point charges q_k inside a dielectric cavity with dielectric constants ϵ_i, embedded in solvent with dielectric constant ϵ_o, the electrostatic potential Φ at any position $\mathbf{r}$ is given by

$$\Phi(\mathbf{r}) = \frac{1}{\epsilon_i}\sum_{k}\frac{q_k}{|\mathbf{r} - \mathbf{r}_k|} + \int_S \frac{\sigma(\mathbf{R})\, d^2\mathbf{R}}{|\mathbf{r} - \mathbf{R}|} \quad (27)$$

$$= \Phi_f(\mathbf{r}) + \Phi_\sigma(\mathbf{r}) \quad (28)$$

$$\Phi_f(\mathbf{r}) = \frac{1}{\epsilon_i}\sum_{k}\frac{q_k}{|\mathbf{r} - \mathbf{r}_k|} \quad (29)$$

$$\Phi_\sigma(\mathbf{r}) = \int_S \frac{\sigma(\mathbf{R})\, d^2\mathbf{R}}{|\mathbf{r} - \mathbf{R}|} \quad (30)$$

where q_k is the charge on atom k, $\mathbf{r}_k$ is its coordinate, and $\sigma(\mathbf{R})$ is the induced polarization charge density on the dielectric boundary at point $\mathbf{R}$, where $\mathbf{R}$ represents the vector of integration over the surface of the molecule. $\Phi_f(\mathbf{r})$ is the potential of the free charges, and $\Phi_\sigma(\mathbf{r})$ is the potential of the induced polarization charge or the potential of the reaction field.

The total electrostatic energy G_{tot} of such a system can be written as

$$G_{\text{tot}} = \frac{1}{2}\int \rho_f(\mathbf{r})\Phi(\mathbf{r})\, d^3\mathbf{r} \quad (31)$$

$$= \frac{1}{2}\int \rho_f\big[\Phi_f(\mathbf{r}) + \Phi_\sigma(\mathbf{r})\big]\, d^3\mathbf{r} \quad (32)$$

$$= \frac{1}{2}\int \rho_f \Phi_f(\mathbf{r})\, d^3\mathbf{r} + \frac{1}{2}\int \rho_f \Phi_\sigma(\mathbf{r})\, d^3\mathbf{r} \quad (33)$$

$$\rho_f = \sum_{k} q_k \delta(\mathbf{r} - \mathbf{r}_k) \quad (34)$$

where ρ_f is the free charge density, not including the induced charge density. The first term of eq. (33) cor-

responds to the energy of the molecule embedded in the dielectric with the same dielectric constant, where there is no boundary and no induced charge; the second term of eq. (33) corresponds to the energy of the interaction of the set of point charges with the reaction field. The electrostatic free energy of solvation of the solute molecule is the energy of transfer from an environment with dielectric constant of the solute interior ϵ_i to the solvent ϵ_o, which is exactly the second term of eq. (33). So using eqs. (33), (34), and (30), we can write the solvation free energy or the reaction field energy as,

$$\Delta G_{\text{solv}} = \frac{1}{2}\int \rho_f \Phi_\sigma(\mathbf{r})\, d^3\mathbf{r} \tag{35}$$

$$= \frac{1}{2}\sum_k \int_S \frac{q_k \sigma(\mathbf{R})\, d^2\mathbf{R}}{|\mathbf{r}_k - \mathbf{R}|} \tag{36}$$

From Gauss's law, we can get the induced charge density on the dielectric boundary, which is given by,

$$\sigma(\mathbf{R}) = \frac{1}{4\pi}\left[\mathbf{E}_o(\mathbf{R})\cdot\mathbf{n}(\mathbf{R}) - \mathbf{E}_i(\mathbf{R})\cdot\mathbf{n}(\mathbf{R})\right] \tag{37}$$

where $\mathbf{E}_o(\mathbf{R})\cdot\mathbf{n}(\mathbf{R})$ and $\mathbf{E}_i(\mathbf{R})\cdot\mathbf{n}(\mathbf{R})$ are the normal component of the electric fields on the solvent and solute sides of the boundary, respectively.

Continuity of the normal component of the electric displacement across the dielectric boundary requires that,

$$\epsilon_o \mathbf{E}_o(\mathbf{R})\cdot\mathbf{n}(\mathbf{R}) = \epsilon_i \mathbf{E}_i(\mathbf{R})\cdot\mathbf{n}(\mathbf{R}) \tag{38}$$

Combining this boundary condition [eq. (38) and eq. (37)], the induced polarization charge density on the boundary can be written in terms of the normal component of the electric field at the surface, approached from inside the cavity, by the relation

$$\sigma(\mathbf{R}) = \frac{1}{4\pi}\left(\frac{\epsilon_i}{\epsilon_o} - 1\right)\mathbf{E}_i(\mathbf{R})\cdot\mathbf{n}(\mathbf{R}) \tag{39}$$

What we are really interested in is the electrostatic solvation free energy of a single point charge q_k inside an arbitrary molecular cavity. In the special case of a single point charge q_k located at the center of a spherical cavity with dielectric constant ϵ_i, $\mathbf{E}_i(\mathbf{R})\cdot\mathbf{n}(\mathbf{R})$ of eq. (39) can be obtained by Gauss's law, which is given by

$$\mathbf{E}_i(\mathbf{R})\cdot\mathbf{n}(\mathbf{R}) = q_k \frac{(\mathbf{R} - \mathbf{r}_k)\cdot\mathbf{n}(\mathbf{R})}{\epsilon_i |\mathbf{R} - \mathbf{r}_k|^3} \tag{40}$$

Although perfect spherical symmetry will not be manifested in a realistic molecular problem, for systems in which the surface is locally convex, eq. (40) can be used as a reasonable approximation to the normal component of the electric field at the boundary surface. Specifically, by using only the Coulombic contribution to the electric field [eq. (40)] as a basis for the induced polarization charge [eq. (39)], and integrating using eq. (36), we arrive at the approximation for the solvation free energy of a single point charge q_k at $\mathbf{r}_k$ in a cavity of arbitrary shape, given by

$$\Delta G_{\text{single}} = -\frac{1}{8\pi}\left(\frac{1}{\epsilon_i} - \frac{1}{\epsilon_o}\right)\int_S \frac{q_k^2}{|\mathbf{R} - \mathbf{r}_k|^4} \times (\mathbf{R} - \mathbf{r}_k)\cdot\mathbf{n}(\mathbf{R})\, d^2\mathbf{R} \tag{41}$$

One can easily verify that eq. (41) reduces to eq. (16) for the case of a spherical cavity; this indicates that the Born equation can also be derived by reaction field theory. Once ΔG_{single} is obtained numerically by carrying out the integration indicated in eq. (41), we can use the single energy expression/the Born equation [eq. (26)] to get the effective Born radius α_k. Then the total electrostatic free energy of solvation can be obtained by the generalized Born equation [eq. (22)] with the effective Born radius for every atom of the solute molecule. Ghosh and coworkers have added empirical corrections to the above formalism of the SGB model to accommodate errors introduced by the approximations.[20] They were able to achieve close agreement with numerical calculations of the Poisson–Boltzmann equation.

FREE ENERGY PERTURBATION (FEP) CALCULATIONS

The explicit solvent simulations were carried out by the free energy perturbation (FEP) technique and molecular dynamics sampling method.[1] The free energy difference between two related systems/states is given by,

$$\Delta G = -kT \ln\left\langle \exp\left(-[H_2 - H_1]/kT\right)\right\rangle_1 \tag{42}$$

$$= -kT \ln\left\langle \exp\left(-[V_2 - V_1]/kT\right)\right\rangle_1 \tag{43}$$

where H_1 and H_2 are the Hamiltonians, V_1 and V_2 are the total potential energy of the two systems/states, respectively. Because the kinetic energy contributions to the two Hamiltonians are equal at each point in phase space, the difference of the Hamiltonians is replaced by the difference of the potential functions in the above equations. $\langle \ldots \rangle_1$ denotes an ensemble average corresponding to the Hamiltonian H_1. Equation (43) is exact, and is the fundamental equation of the free energy perturbation technique; however, it is practically only directly useful for states 1 and 2, which are not

too far apart. The difficulty arises because with finite computer time, a simulation of the ensemble corresponding to H_1 will predominantly sample microstates for which H_1 is small, which are not necessarily the same as those for which H_2 are small. Therefore, the evaluation of eq. (43) is divided into windows, each one involving a small enough interval characterized by a parameter λ, to allow the free energy difference between the adjacent intermediate states to be calculated accurately. Usually $\lambda = 0$ is set to correspond to state 1 and $\lambda = 1$ to state 2. ΔG is then evaluated as the sum of these free energy differences ΔG_i between the nearby windows corresponding to λ_{i-1} and λ_i:

$$\Delta G = \sum_{i=1}^{N} \Delta G_i \tag{44}$$

$$= \sum_{i=1}^{N} -kT \ln\left\langle \exp\left(-[V(\lambda_i) - V(\lambda_{i-1})]/kT\right)\right\rangle_{\lambda_{i-1}} \tag{45}$$

PARAMETERIZATION OF OPLS LENNARD-JONES RADII FOR SGB

For continuum solvation models, one of the most important issues is how to define the dielectric boundary. In the SGB model, the dielectric medium was set to begin at the van der Waals surface of the solute molecule. However, the van der Waals parameters are force field dependent; for the OPLS force field they have been optimized for liquid state simulations. To use the OPLS van der Waals radii in the SGB calculations, it was found that a scaling factor as described in this section has to be applied. The atomic radii r^{SGB}, used by SGB to generate the solute cavity, were parameterized against FEP simulation results with the OPLS force field on a database of roughly 40 small organic molecules. The geometries and atomic partial charges were obtained, either from published literature[46, 47] as marked in Table I, or by *ab initio* quantum chemistry geometry optimization and electrostatic potential charge fitting performed at the HF/631g** level using the JAGUAR program.[26] The same geometry and atomic charges were used in both the FEP and SGB calculations.

The FEP calculations were carried out through the IMPACT program using molecular dynamics (MD) sampling. For computational efficiency, and to be consistent with the SGB calculations, the solute molecules were set to be rigid. The calculations generally included 216 to 1080 TIP3P water molecules in a cubic or rectangular box with periodic boundary condition applied. Spherical cutoffs for nonbonded interactions were used, ranging from 9.3 to 18.5 Å. The volume and temperature were kept constant during the simulation, with the relaxation time of 0.2 ps for velocity scaling. The initial and target temperatures were set to be 298.15 K. The FEP calculations were performed starting with the neutral solute molecule containing no atomic partial charges. Then the charges were grown slowly at the same scale for each atom in windows; between 20 and 100 windows was used, depending on the polarity of the solute. Double wide sampling was employed. Within each window the system was equilibrated for 2 ps, and data was collected for 2 ps.

The SGB calculations were first performed using the same OPLS Lennard-Jones radii $\sigma/2$ as used in the FEP calculations. The charging free energies obtained by the SGB method were much more negative than that of the FEP calculations. Then with fixed atomic partial charges, a set of SGB calculations were performed by systematically varying the OPLS Lennard-Jones radii $\sigma/2$ with two scaling factor as $r^{SGB} = a*\sigma/2+b$. The best agreement between SGB and FEP results were obtained by setting the SGB van der Waals radii r^{SGB} as,

$$r^{SGB} = 1.1 * \sigma/2 + 0.05 \tag{46}$$

The best fitting results are presented in Table I. The average unsigned error is 0.6 kcal/mol. Atomic radii set by eq. (46) were used for all the following SGB calculations on the protein–ligand binding problems.

LIGAND ELECTROSTATIC CHARGING PROCESS

The ligand charging process consists of electrically charging and discharging the OVA-8 peptide ligand. The octapeptide OVA-8, with amino acid sequence SER-ILE-ILE-ASN-PHE-GLU-LYS-LEU (SIINFEKL), has a net charge of 0, even though it has two oppositely charged side chains GLU-6 and LYS-7 in addition to the ionic end caps of the zwitterion form considered here. The receptor is the murine MHC class I protein H-2KB (PDB name: 1VAC). The ligand was considered to be rigid in all calculations, while some calculations were carried out with the MHC protein rigid (FEP, SGB, and PBF); in others, protein relaxation was considered as described below (FEP only). In the free energy perturbation (FEP) simulations with explicit solvent, the TIP4P water model was used throughout this study. The bond lengths and bond angles of the solvent molecules were constrained using the SHAKE algorithm.[48] The calculations included 2116

TABLE I.
The Calculated Charging Free Energies (kcal/mol) for Small Organic Molecules by the SGB and FEP Methods.

Solute	FEP	SGB	Δ (FEP-SGB)
Hexane	0.0	−0.5	0.5
Cyclohexane	0.0	0.1	−0.1
Methanol[a]	−5.8	−5.2	−0.6
Ethanol	−5.3	−5.3	0.0
2-Propanol	−5.6	−6.4	0.8
Acetone[a]	−6.2	−7.7	1.5
Methyl acetate[a]	−5.7	−7.8	2.1
Acetic acid[a]	−9.1	−10.0	0.9
Dimethyl ether	−2.6	−2.8	0.2
Methylamine	−5.2	−5.1	−0.1
Ethylamine	−5.7	−6.5	0.8
Dimethylamine	−3.9	−3.9	0.0
Trimethylamine	−1.6	−0.6	−1.0
Butylamine	−3.3	−4.1	0.8
N,*N*′-Dimethylethylenediamine	−7.3	−7.5	0.2
Acetonitrile	−5.9	−6.6	0.7
Methanethiol	−2.5	−2.8	0.3
Ethanethiol	−2.5	−2.6	0.1
Acetamide[a]	−13.0	−13.8	0.8
cis-*N*-Methylacetamide	−10.6	−10.4	−0.2
trans-*N*-Methylacetamide	−10.3	−10.5	0.2
N,*N*-Dimethylacetamide	−9.2	−8.1	−1.1
Benzene[a]	−1.9	−2.2	0.3
Toluene	−2.1	−2.9	0.8
Phenol[a]	−6.4	−6.7	0.3
Pyridine[a]	−3.6	−3.3	−0.3
Piperidine (eq)	−3.7	−4.2	0.5
Piperidine (ax)	−4.5	−4.9	0.4
Piperazine (eq)	−7.0	−8.4	1.4
Piperazine (ax)	−8.4	−9.3	0.9
N-Methylpiperazine (eq)	−5.3	−5.1	−0.2
N,*N*′-Dimethylpiperazine (eq)	−2.9	−0.8	−2.1
ala dipeptide ($C7_{eq}$)	−16.4	−17.2	0.8
ala dipeptide ($C7_{ax}$)	−16.5	−17.2	0.7
ala dipeptide (*C*5)	−16.9	−17.6	0.7
ala dipeptide (α')	−17.9	−18.2	0.3
ala dipeptide (β_2)	−18.1	−18.3	0.2
ala dipeptide (α_L)	−19.9	−19.4	0.5
Average unsigned error is			0.6

[a] The geometries and parameters are from refs. 46 and 47.

or 12,554 TIP4P [49, 50] water molecules in a cubic box with periodic boundary conditions applied for the peptide and protein–peptide complex, respectively. The pressure and temperature were kept constant during the simulation, with the relaxation time of 0.2 ps for velocity scaling.[2] The initial and target temperatures were set to be 298.15 K.

The free energy perturbation (FEP) calculations were performed starting with the neutral peptide containing no atomic partial charges, either free in water or bound to the charged protein. Then the charges were grown in for each atom in each window. The double-wide window method[2] was applied, with a total number of nine windows performed for the peptide free in water and 10 windows for the peptide bound to the protein.

The charging process is divided into 10 intermediate steps. In the first eight steps each amino acid residue is charged one at a time to its corresponding neutral form starting from the first residue

SER-1, and charging each residue in turn. The neutral form of a neutral amino acid residue is the residue with a full OPLS charge complement. For GLU-6 and LYS-7, the neutral forms (denoted by a "n" in the residue name in the accompanying tables) of the residues were obtained by neutralizing the net charges on the charged functional group (—COO^- and —NH_3^+). In the ninth step, the charging of the two oppositely charged side chains GLU-6 and LYS-7 was accomplished by restoring the final OPLS charges for the ionized residues (transformation from SIINFE^{n}K^{n}L to SIINFEKL). In the 10th and final step, two opposite unit charges were grown at the amino and carboxy ends of the two terminal residues to form the zwitterion (from SIINFEKL to S^+IINFEKL$^-$). The ligand is considered rigid, and the same peptide conformation is used for the bound and free charging processes in solution.

The protein H-2KB is divided into three domains. Preliminary calculations indicated that the removal of the third domain, which is located away from the binding pocket, did not substantially affect the calculated electrostatic binding free energy. The calculations reported below have been performed on the H-2KB protein with the third domain removed.

Two sets of calculations were performed for the bound ligand. In one set the protein is kept rigid in a energy minimized structure; in the other set the protein internal degrees of freedom were unconstrained.

For the charging process of the free peptide in water, the uncharged peptide and 2,116 TIP4P water molecules were inserted in a box of dimensions $39.8 \times 39.8 \times 39.8$ Å^3. Electrostatic intermolecular interactions were evaluated using the Ewald formula, and Lennard-Jones intermolecular dispersion interactions were truncated at 13.5 Å according to the atomic positions of the peptide and according to the water center of mass for the solvent. The system was equilibrated for 6 ps. Starting with this equilibrated sample, the charging of each residue was accomplished by first equilibrating for an additional 2 ps and then by performing a FEP calculation with nine windows. For each window the system was equilibrated for 2 ps, followed by 6 ps of data collection. Care was taken to ensure that overall charge neutrality of the peptide was conserved at all the intermediate charge states along the charging path.

For the charging process of the peptide bound to the protein in water, the protein-uncharged ligand complex and 12,554 TIP4P water molecules were inserted in a box of dimensions $74.5 \times 74.5 \times 74.5$ Å^3. Electrostatic and Lennard-Jones intermolecular interactions were truncated at 13.5 Å implementing a residue based cutoff scheme for the peptide–protein complex and a center of mass based cutoff scheme for the water molecules. Test calculations using the periodic Coulomb potential (Ewald formula) instead of residue based cutoffs gave the same results for the charging free energies of the neutral residues. The rigid protein calculation was carried out by first energy minimizing with the uncharged peptide and then equilibrating for 16 ps constraining the protein internal degrees of freedom. The FEP calculation was then performed with ten windows. In each window the system was equilibrated for an additional 8 ps, followed by 12 ps of data collection with a 2 fs time step.

A number of time-saving techniques were implemented to reduce the computational complexity of these FEP simulations. The free energy changes were calculated only in the forward direction by performing double-wide sampling at every other FEP step. The mutations from one charge state to the next were performed all at the same time and, in addition, the calculations at each window were performed in parallel on several processors starting with a properly equilibrated initial configuration. This reduced the required CPU time from about 200 days to 8 days by using 25 processors on the Origin 2000 system at the NCSA supercomputer center.

For the flexible protein calculation, the complex was first equilibrated for 24 ps gradually increasing the temperature to 298 K. The FEP calculation was then performed as above, except that within each window, 3 ps of additional equilibration and 12 ps of data collection were carried out using a 3-fs time step and the RESPA multiple time step algorithm.[8, 51]

Results and discussion

REFERENCE STATES FOR ELECTROSTATIC BINDING FREE ENERGIES

In this section, we study the electrostatic contribution to the binding of the murine MHC class I protein H-2KB with the octapeptide OVA-8 using FEP MD simulations with explicit solvent comparing the results with the SGB continuum solvent model and with PBF, which is a solution to the Poisson–Boltzmann equation using the finite element method.[26, 34] The coordinates of the complex were obtained from the Brookhaven Protein Data Bank (PDB)[52] (PDB name: 1VAC[53]).

In this work the electrostatic binding free energy ΔG_{ebind} is defined according to the thermody-

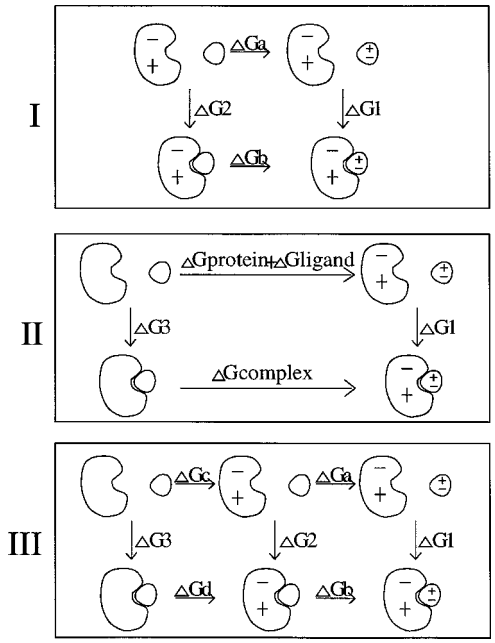

FIGURE 1. Thermodynamic cycles for calculation of the electrostatic contribution to the binding free energy of a free ligand binding to a free protein in aqueous solution.

namic cycle I shown in Figure 1,

$$\Delta G_{\text{ebind}} = \Delta G_1 - \Delta G_2 \quad (47)$$

$$= \Delta G_b - \Delta G_a \quad (48)$$

where ΔG_1 is the binding free energy of the ligand to the protein, ΔG_2 is the binding free energy of the uncharged ligand to the protein, ΔG_a and ΔG_b are, respectively, the free energies of charging the free ligand in water and bound to the protein in water (later referred to as the free and bound states of the ligand). According to this definition, the electrostatic binding free energy is a measure of the difference of binding affinity between the charged and uncharged forms of the ligand, or, in other words, of the influence of the charges of the ligand on the binding free energy. Note that the reference states for the charging processes (the states on the left in cycle I of Fig. 1) are the uncharged free ligand in water and the uncharged bound ligand in water.

The definition of the electrostatic component of the binding free energy sometimes used in studies using implicit solvation models[31] may differ from ours with respect to the assignment of the reference states. As shown by cycle II in Figure 1, in this alternative definition, the reference states are the unbound ligand and protein and the ligand–protein complex in their uncharged forms. By referring to cycle II of Figure 1, this alternative definition of ΔG_{ebind} is expressed as

$$\Delta G'_{\text{ebind}} = \Delta G_1 - \Delta G_3$$

$$= \Delta G_{\text{complex}} - (\Delta G_{\text{protein}} + \Delta G_{\text{ligand}}) \quad (49)$$

where $\Delta G_{\text{complex}}$, $\Delta G_{\text{protein}}$, and ΔG_{ligand} are the charging free energies in water of, respectively, the ligand–protein complex, the free protein and the free ligand, ΔG_1 is, as above, the binding free energy of the ligand to the protein, and ΔG_3 is the binding free energy of the uncharged ligand to the uncharged protein.

From cycle III in Figure 1, it is apparent that the reference states in our definition are intermediate states in the alternative definition. The relation between the two definitions of the electrostatic component of the free energy of binding is, therefore

$$\Delta G'_{\text{ebind}} - \Delta G_{\text{ebind}} = (\Delta G_b + \Delta G_d - \Delta G_c - \Delta G_a) - (\Delta G_b - \Delta G_d)$$

$$= \Delta G_d - \Delta G_c \quad (50)$$

Thus, to compare our results with results reported according to the alternative definition,[31] it is necessary to estimate the charging free energy of the entire protein with (ΔG_c) or without (ΔG_d) the bound uncharged ligand. Given the extent of the required charge mutation, this calculation is not feasible for FEP simulations with explicit solvent. We have, therefore, estimated $\Delta G_d - \Delta G_c$ using an implicit solvent model (SGB and PBF).

In this work we present the computation of ΔG_a and ΔG_b by FEP, SGB, and PBF, and we compare the predictions of the models and performance of these methods.

SOLVATION FREE ENERGIES OF THE FREE AND COMPLEXED LIGAND

The solvation free energies for charging the free octapeptide ligand in water calculated by FEP simulations are compared with the corresponding SGB and PBF results in Table II and Figure 2. The quantitative agreement between the implicit (SGB) and explicit solvent simulations for charging the neutral octapeptide is truly remarkable, especially considering the huge difference in the times required for the calculations—9 days for the FEP results compared with 7 s for the SGB results. Even for the last two charging steps, creating the ionized side chains and the zwitterionic peptide, the agreement is close to quantitative. The PBF results display a small systematic shift with respect to the SGB and FEP results, but this is understandable considering the fact

TABLE II.
The Cumulative Solvation Free Energies of Charging the Free Octapeptide OVA-8 with the Sequence SIINFEKL in Water (kcal/mol) (for SGB and PBF, $\epsilon_{in} = 1$, $\epsilon_{out} = 80$).

	ΔG^{free}_{solv}		
Charging Steps	FEP	SGB	PBF
1. S	-8.73 ± 0.28	-9.90	-12.78
2. SI	-15.11 ± 0.32	-16.09	-19.73
3. SII	-20.76 ± 0.37	-20.95	-24.79
4. SIIN	-34.10 ± 0.47	-34.46	-38.89
5. SIINF	-41.27 ± 0.51	-40.75	-45.67
6. SIINFE^n	-47.24 ± 0.53	-47.15	-51.97
7. SIINFE^{n}K^n	-54.35 ± 0.58	-54.51	-59.42
8. SIINFE^{n}K^{n}L	-60.62 ± 0.60	-60.89	-66.87
9. SIINFEKL	-159.23 ± 1.00	-153.59	-166.08
10. S$^+$IINFEKL$^-$	-282.97 ± 1.29	-282.28	-299.09
Timing/step	9 days	7 s	24 s

that the OPLS parameters were not explicitly readjusted to optimize the PBF calculations, as they were to optimize the SGB results on a small molecule database as described earlier. It is apparent that for the free peptide in solution the implicit solvent calculations capture the solvation thermodynamics as described by the explicit solvent simulations. This is true even though a detailed structural analysis of the solvent structured around the protein would reveal rich details that suggest a rather different picture from the continuum viewpoint.

The corresponding results for charging the peptide bound to the rigid MHC protein in the solvated complex are shown in Table III and Figure 3. The FEP solvation free energies for each charging step of the peptide in the complex are smaller than for the free peptide in solution; this reflects the displacement of many water molecules solvating the peptide by the protein. The agreement between the explicit and implicit solvent simulations of the charging free energies is worse for the peptide in

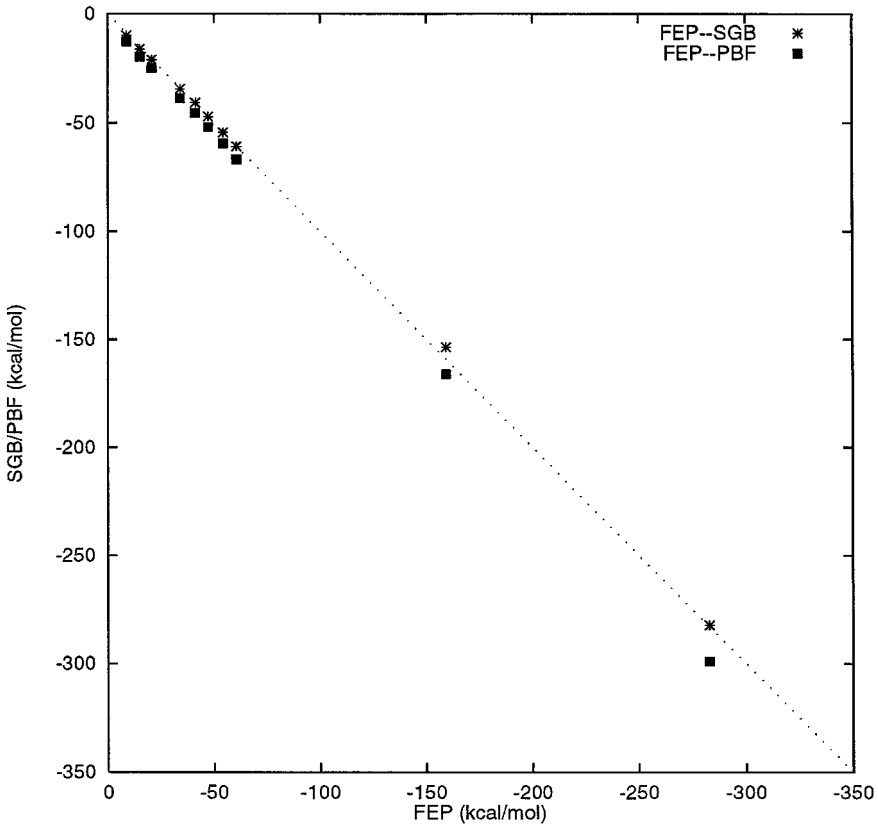

FIGURE 2. The comparison of the electrostatic solvation free energies of charging the octapeptide OVA-8 in water (kcal/mol) by FEP, SGB, and PBF methods. The dashed line has unit slope. The stars indicate the correlation between the FEP and the SGB results, the squares indicate the correlation between the FEP and the PBF results.

TABLE III.
The Cumulative Solvation Free Energies of Charging the OVA-8 with Sequence SIINFEKL Bound to the Rigid Protein in Water (kcal/mol) (for SGB and PBF, $\epsilon_{in} = 1$, $\epsilon_{out} = 80$).

	ΔG^{bound}_{solv}		
Charging Steps	FEP	SGB	PBF
1. S	-1.49 ± 0.07	-2.77	-2.42
2. SI	0.82 ± 0.11	1.36	-0.44
3. SII	-4.21 ± 0.11	-0.45	-3.11
4. SIIN	-4.84 ± 0.22	-4.99	-8.15
5. SIINF	-5.09 ± 0.27	-2.11	-8.97
6. SIINFE^n	-5.84 ± 0.28	-8.54	-12.96
7. SIINFE^{n}K^n	-10.26 ± 0.30	-13.12	-19.27
8. SIINFE^{n}K^{n}L	-6.15 ± 0.33	-4.74	-13.32
9. SIINFEKL	-106.14 ± 0.74	-74.10	-97.16
10. S$^+$IINFEKL$^-$	-120.30 ± 0.80	-108.07	-105.29
Timing/step	19 days	2 min	27 min

the complex (Table III) compared with the free peptide in solution (Table II), but the absolute value of the solvation free energies are also much smaller in magnitude; this serves to decrease the effects of discrepancies on estimates of the binding free energy (see below). It is unclear why there is a larger disagreement between explicit and implicit solvent models for charging the octapeptide in the complex than the free peptide. One source may be related to the fact that for the set of organic solutes on which the SGB model has been parameterized, the charge distribution is in close proximity to the solute–solvent surface as it is for the free octapeptide, while in the complex, the peptide charges are more buried.

ELECTROSTATIC CONTRIBUTION TO PROTEIN–LIGAND BINDING FREE ENERGIES

We first consider the electrostatic contribution to the binding of the OVA-8 octapeptide to the

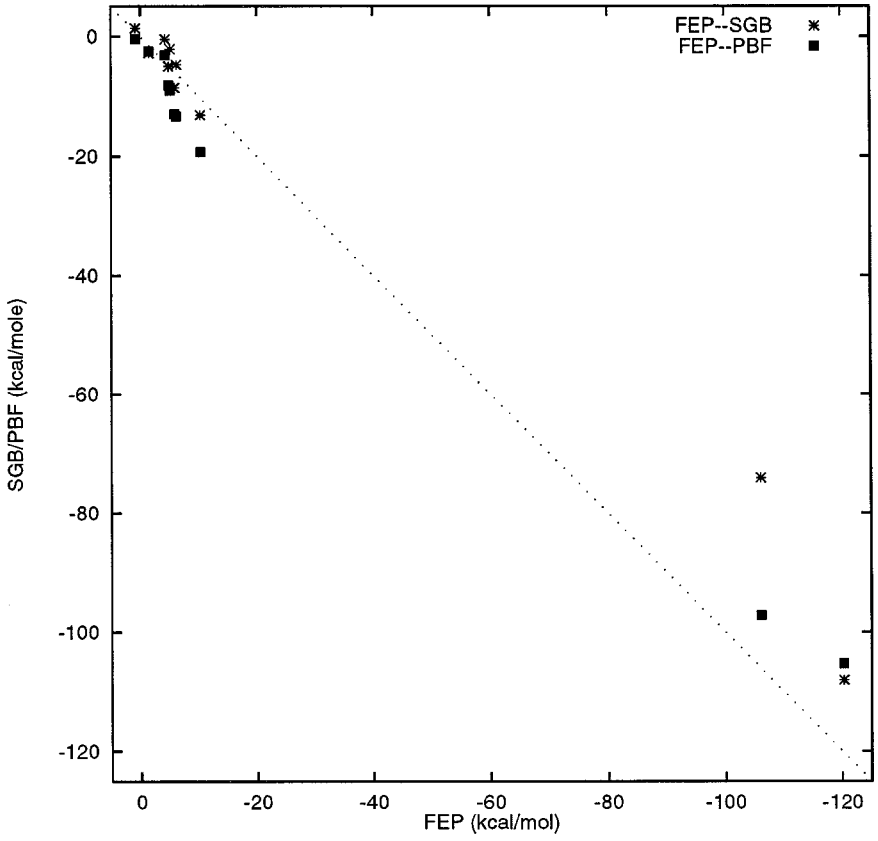

FIGURE 3. The comparison of the electrostatic solvation free energies (ΔG^{bound}_{solv}) of charging the octapeptide OVA-8 bound to the rigid protein H-2KB in water (kcal/mol) by FEP, SGB, and PBF methods. The dashed line has unit slope. The stars indicate the correlation between the FEP and the SGB results, the cubics indicate the correlation between the FEP and the PBF results.

TABLE IV.
The FEP Calculated, for Each Charging Step, Desolvation Penalty ($\Delta G_{desolv} = \Delta G_{solv}^{bound} - \Delta G_{solv}^{free}$), the Direct Coulomb Interaction (ΔG_{coul}) and the Electrostatic Binding Free Energy ($\Delta G_{ebind} = \Delta G_{desolv} + \Delta G_{coul}$) of the Octapeptide OVA-8 with Sequence SIINFEKL Bound to the Rigid Protein H-2KB in Water (kcal/mol).

Charging Steps	ΔG_{desolv}	ΔG_{coul}	ΔG_{ebind}
1. S_1	7.24 ± 0.29	-4.78	2.46 ± 0.29
2. 1_2	8.69 ± 0.18	-18.84	-10.15 ± 0.18
3. 1_3	0.62 ± 0.19	2.87	3.49 ± 0.19
4. N_4	12.72 ± 0.35	-23.83	-11.11 ± 0.35
5. F_5	6.92 ± 0.25	-9.92	-3.00 ± 0.25
6. E^n_6	5.22 ± 0.17	5.03	10.25 ± 0.17
7. K^n_7	2.68 ± 0.24	-10.21	-7.53 ± 0.24
8. L_8	10.38 ± 0.23	-21.35	-10.97 ± 0.23
9. $E^nK^n \rightarrow EK$	-1.38 ± 1.04	-13.60	-14.98 ± 1.04
10. $SL \rightarrow S^+ \rightarrow L^-$	109.58 ± 0.88	-46.28	63.30 ± 0.88

rigid MHC protein calculated by FEP simulations with explicit solvent. The results are shown in Table IV for each charging step individually; the cumulative results are shown in Table V. Two terms, which oppose each other, contribute to the binding. The direct Coulomb interactions ΔG_{coul} between the ligand and the protein favor binding—except for the process of creating the charge distribution on ILE-3 and the neutral form of GLU-6, both of which oppose binding. The total Coulomb interaction between the ligand and protein contributes -81 kcal/mol to the binding of the neutral peptide ($SIINFE^nK^nL$) (including the two additional charging steps to create the charged residues and the zwitterionic groups, the direct Coulomb interaction energy is -141 kcal/mol). The desolvation term ΔG_{desolv}—the difference between the solvation free energy of the bound and free ligand—opposes binding, but this term is generally smaller than the direct Coulomb interaction between the ligand and protein. The total electrostatic contribution to the binding energy ΔG_{ebind} (sum of the Coulomb interaction plus the desolvation free energy) is calculated to be -26.56 kcal/mol for the neutral peptide (Table V). Ionizing the charged residues (step 9 in Table IV) also favors binding, but creating the zwitterionic form of the peptide changes the equilibrium. The large difference in the desolvation penalties be-

TABLE V.
The Cumulative FEP Calculated Desolvation Penalty ($\Delta G_{desolv} = \Delta G_{solv}^{bound} - \Delta G_{solv}^{free}$), the Direct Coulomb Interaction (ΔG_{coul}) and the Electrostatic Binding Free Energies ($\Delta G_{ebind} = \Delta G_{desolv} + \Delta G_{coul}$) of the Octapeptide OVA-8 with Sequence SIINFEKL Bound to the Rigid Protein H-2KB in Water (kcal/mol).

Charging Steps	ΔG_{desolv}	ΔG_{coul}	ΔG_{ebind}
1. S	7.24 ± 0.29	-4.78	2.46 ± 0.29
2. SI	15.93 ± 0.34	-23.62	-7.69 ± 0.34
3. SII	16.55 ± 0.39	-20.75	-4.20 ± 0.39
4. SIIN	29.27 ± 0.52	-44.58	-15.31 ± 0.52
5. SIINF	36.19 ± 0.58	-54.50	-18.31 ± 0.58
6. $SIINFE^n$	41.41 ± 0.60	-49.47	-8.06 ± 0.60
7. $SIINFE^nK^n$	44.09 ± 0.65	-59.68	-15.59 ± 0.65
8. $SIINFE^nK^nL$	54.47 ± 0.68	-81.03	-26.56 ± 0.68
9. SIINFEKL	53.09 ± 1.24	-94.63	-41.54 ± 1.24
10. $S^+IINFEKL^-$	162.67 ± 1.52	-140.91	21.76 ± 1.52

tween charging the ionized residues (step 9) and the terminal groups (step 10) has a clear structural explanation. The amino and carboxyl terminal groups are completely buried in the complex; removing the groups in their zwitterionic form from the solvent leads to a very large desolvation penalty (110 kcal/mol, Table IV). In contrast, the ionized side chains E-6 and K-7 are completely exposed to solvent in the complex, so there is no desolvation penalty. Based on the geometry of the complex, it is very unlikely that the amino and carboxyl terminal groups are charged in the bound complex, and in fact, these groups were considered to be unionized in previous modeling studies.[31] In summary, the results for the rigid protein binding calculations indicate that the electrostatic contribution ΔG_{ebind} favors binding of the neutral peptide SIINFEKL.

Froloff et al.[31] studied the same system by Poisson–Boltzmann calculations and concluded that, according to the reference state they used for the calculation of ΔG_{ebind} discussed above, the electrostatic component always opposed binding. To compare the results of Froloff et al. to ours, the term given by eq. (50) has to be added to our FEP numerical results. This term was calculated by SGB to be 81 kcal/mol. By adding this term to the FEP estimates of ΔG_{ebind} values at each intermediate step we see that the values of ΔG_{ebind} from step 1 "S" to step 10 "S^+IINFEKL$^-$" indeed become positive and oppose binding. It is of interest to note that the FEP cumulative electrostatic binding free energy of state "SIINFE^{n}K^{n}L," when expressed with respect to Froloff et al. reference states, is about 54 kcal/mol, close to the corresponding value obtained by Froloff et al.

The cumulative electrostatic free energies for the binding of OVA-8 to the MHC protein calculated with explicit and implicit solvent are compared in Table VI. The explicit and implicit solvent predictions for the electrostatic component of the binding of the neutral peptide (charging steps 1 through 8) are in reasonably good agreement. This is because the binding free energy is dominated by the sum of the Coulomb interaction term and the solvation free energy of the free peptide for which the FEP and SGB estimates are in good agreement. In contrast, the agreement is poor for the last two charging steps involving the ionized residues. For steps 9 and 10, involving charging the ionized residues, the FEP estimates of the binding are much more favorable (e.g., −41.5 kcal/mol for SIINFEKL bound to the protein calculated by FEP compared with −15.14 kcal/mol calculated by SGB), due to the more favorable electrostatic charging free energy of the ionized residues in the bound peptide as calculated with explicit solvent compared with the implicit solvent result. In contrast, there is good agreement between the FEP and SGB electrostatic charging free energies of the ionized residues of the peptide in solution. These results suggest that additional parameterization of the current implementation of the SGB model for proteins may be necessary to correctly model binding phenomena involving ionized groups.

TABLE VI.
The Cumulative FEP, SGB, and PBF Calculated Desolvation Penalty ($\Delta G_{\text{desolv}} = \Delta G^{\text{bound}}_{\text{solv}} - \Delta G^{\text{free}}_{\text{solv}}$), the Direct Coulomb Interaction (ΔG_{coul}) and Elecrostatic Binding Free Energy (ΔG_{ebind}) of the Octapeptide OVA-8 with Sequence SIINFEKL Bound to the Rigid Protein H-2KB in Water (kcal/mol) (for SGB and PBF calculations, $\epsilon_{\text{in}} = 1$, $\epsilon_{\text{out}} = 80$).

	ΔG_{desolv}				ΔG_{ebind}		
Charging Steps	FEP	SGB	PBF	ΔG_{coul}	FEP	SGB	PBF
1. S	7.24	7.13	10.36	−4.78	2.46	2.35	5.58
2. SI	15.93	17.45	19.29	−23.62	−7.69	−6.17	−4.33
3. SII	16.55	20.50	21.68	−20.75	−4.20	−0.25	0.93
4. SIIN	29.27	29.47	30.74	−44.58	−15.31	−15.11	−13.84
5. SIINF	36.19	38.64	36.70	−54.50	−18.31	−15.86	−17.80
6. SIINFE^n	41.41	38.61	39.01	−49.47	−8.06	−10.86	−10.46
7. SIINFE^{n}K^n	44.09	41.39	40.15	−59.68	−15.59	−18.29	−19.53
8. SIINFE^{n}K^{n}L	54.47	56.15	53.55	−81.03	−26.56	−24.88	−27.48
9. SIINFEKL	53.09	79.49	68.92	−94.63	−41.54	−15.14	−25.71
10. S^+IINFEKL$^-$	162.67	174.21	193.80	−140.91	21.76	33.30	52.89

EFFECT OF PROTEIN FLEXIBILITY

To facilitate the comparison of implicit with explicit solvent model predictions for protein–ligand electrostatic interactions, the protein has been considered to be rigid in the explicit solvent simulations considered above. In this way, the protein dielectric constant could be unambiguously assigned a value of unity in the corresponding implicit solvent modeling. In fact, there is considerable uncertainty as to the appropriate value to assign for the dielectric constant of a protein.[54–57] The common assignment of protein dielectric constants greater than two in biomolecular modeling is an attempt to partially account for protein flexibility, which, in the context of a charge perturbation, leads to dielectric shielding and protein reorganization. The consistent treatment of protein flexibility and motions in the context of implicit solvent models of protein electrostatic effects is a current problem of considerable importance.

As a demonstration of the potentially large effects that protein flexibility can have on the modeling of protein–ligand thermodynamics, we compare in Table VII estimates of the charging free energy (sum of Coulomb interaction plus solvation free energy) of the OVA-8 peptide ligand in the solvated protein complex, calculated from FEP simulations in which the solvated protein is constrained to be rigid with corresponding simulations for which the protein is unconstrained, and therefore, experiences thermal fluctuations. The charging free energy of the ligand in the bound complex is consistently less negative when the protein motions are included explicitly in the FEP simulations. For example, to fully charge the neutral peptide SIINFEKL the change in free energy is estimated to be −184 kcal/mol in the FEP simulations of solvated complex when all the protein degrees of freedom are included (Table VII); the corresponding value is −201 kcal/mol extracted from FEP simulations for the rigid protein.

TABLE VII.
The Cumulative FEP Calculated Total Charging Free Energies of the OVA-8 with Sequence SIINFEKL Bound to the Rigid and Flexible Protein H-2KB in Water (kcal/mol).

	$\Delta G^{\text{bound}} = \Delta G^{\text{bound}}_{\text{solv}} + \Delta G_{\text{coul}}$	
Charging Steps	Rigid	Flexible
1. S	−6.27 ± 0.07	−2.63 ± 0.11
2. SI	−22.80 ± 0.11	−6.55 ± 0.15
3. SII	−24.96 ± 0.11	−11.67 ± 0.18
4. SIIN	−49.42 ± 0.22	−29.19 ± 0.25
5. SIINF	−59.59 ± 0.27	−37.84 ± 0.29
6. SIINFE^n	−55.31 ± 0.28	−36.18 ± 0.31
7. SIINFE^{n}K^n	−69.94 ± 0.30	−47.93 ± 0.35
8. SIINFE^{n}K^{n}L	−87.18 ± 0.33	−58.78 ± 0.38
9. SIINFEKL	−200.77 ± 0.74	−184.10 ± 0.86
10. S$^+$IINFEKL$^-$	−261.21 ± 0.80	−311.77 ± 1.46

The analysis of the simulation data reveals that in the flexible protein simulation the peptide becomes progressively more exposed to the solvent as the charge of the peptide is turned on. The uncharged peptide is buried deeply in the binding pocket. This behavior can be understood in terms of the hydrophobicity of the uncharged peptide. The uncharged peptide is repelled by water and, therefore, tends to be more buried in the binding pocket. As the electrostatic interactions between the peptide and the environment are turned on, the hydration of the peptide is partially reestablished.

The peptide–protein complex conformation used for the rigid receptor simulations is obtained by equilibrating the complex structure with the uncharged peptide. In this conformation, the peptide is deeply buried in the binding pocket. Moreover, because the complex is constrained to be rigid, the exposure of the peptide to the solvent does not increase with increasing peptide polarity, as is observed in the flexible receptor simulation.

We have shown (see Tables II and III) that the electrostatic charging free energy of the peptide in the rigid receptor is more negative than the charging free energy of the peptide in solution. Given that in the flexible receptor the peptide is more hydrated than in the rigid receptor, it is not surprising that the electrostatic charging free energy obtained from the flexible receptor simulation is less negative than the one obtained from the rigid receptor simulation (see Table VII).

The less negative electrostatic charging free energy observed in the flexible receptor FEP simulations are, therefore, interpreted in terms of the smaller electrostatic affinity of the charged peptide for the flexible protein. This is due to the competition between the solvent and the protein for the ligand. When the protein is rigid, however, the electrostatic charging free energies are biased towards more negative values because the solvent is prevented from reestablishing to the same extent the electrostatic interactions with the charged peptide. The results presented in Table VII suggest that the flexibility of the receptor plays an important role in determining electrostatic binding free energies.

A more detailed analysis of these effects is underway.

Summary and Conclusions

Solvent effects play a crucial role in mediating the interactions between proteins and their ligands. In principle, explicit solvent simulations provide the most detailed approach to the modeling of solvent effects. However, these simulations are very computationally time consuming, and this has lead to extensive efforts to develop implicit solvent models that are now very popular in biomolecular modeling. While good agreement between solvation free energies calculated by implicit and explicit solvent models has been reported for relatively small and rigid organic solutes in solution, comparisons for larger biomolecules on problems closer to the applications of most interest have not been made. We have studied the binding of an octapeptide ligand to the murine MHC class I protein using both explicit solvent and implicit solvent models. The solvation free energy calculations are more than 10^3 faster using the surface Generalized Born model compared to FEP simulations with explicit solvent. Surprisingly, there is near quantitative agreement between the explicit and implicit solvent model estimates for the electrostatic component of the solvation free energy of the rigid octapeptide. Although the agreement is not as good for the predicted charging free energy of the peptide in the solvated complex, the qualitative trends in the binding predicted by the explicit solvent FEP simulations are reproduced by the implicit solvent model.

We consider as a reference state for the analysis of the electrostatic component of ligand binding to a protein, the binding of the corresponding ligand without any partial charges. With respect to the binding of this species as the zero of energy, the addition of partial charges to the ligand results in a favorable binding free energy. This is because the favorable Coulomb interaction energy between the ligand and protein, more than compensates for the unfavorable free energy cost of partially desolvating the ligand upon binding to the protein.

The effects of protein flexibility and thermal motions on charging the peptide in the solvated complex were also considered. FEP charging free energy simulations of the peptide bound to the rigid protein were compared with corresponding simulations where the protein was unconstrained. The binding free energy is reduced relative to the rigid protein, both because of the additional screening of the favorable Coulomb interactions between the ligand and the protein, and because of the rearrangement of the ligand in the binding site pocket. It is important to include the conformational freedom of the protein when modeling protein–ligand interactions. Currently, implicit solvent models are not well adapted to include the effects of protein conformational flexibility. Efforts along these lines are underway.

Acknowledgments

We thank Gabriella DelBuono for many helpful discussions, and Anthony Felts for assistance with the manuscript.

References

1. Beveridge, D. L. Annu Rev Biophys Chem 1989, 18, 431.
2. Kitchen, D. B.; Hiratu, F.; Kofke, D. A.; Westbrook, J. D.; Yormush, M.; Levy, R. M. J Comput Chem 1990, 11, 1169.
3. Darden, J. A.; York, D. M.; Pederson, L. G. J Chem Phys 1993, 98, 10089.
4. Figueirido, F.; Del Buono, G. S.; Levy, R. M. J Phys Chem 1996, 100, 6389.
5. Levy, R. M.; Gallicchio, E. Annu Rev Phys Chem 1998, 49, 531.
6. Hummer, G.; Prutt, L. R.; Garua, A. J Phys Chem 1998, 102, 7885.
7. Hunenberger, P. H.; McCammon, J. A. Biophys Chem 1999, 78, 69.
8. Figueirido, F.; Levy, R. M.; Zhou, R.; Berne, B. J. J Chem Phys 1997, 106, 9835.
9. Board, J. A.; Causey, J. W.; Leathrum, J. F.; Windenuth, A.; Schulten, K. Chem Phys Lett 1992, 198, 89.
10. Ding, H. Q.; Karasawa, N.; Goddard, W. A. J Chem Phys 1992, 97, 4309.
11. Warwicker, J.; Watson, H. C. J Mol Biol 1982, 157, 671.
12. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6129.
13. Bashford, D.; Karplus, M. Biochemistry 1990, 29, 10219.
14. Sharp, K.; Nicholls, A.; Fine, R.; Honig, B. Science 1991, 252, 106.
15. Gilson, M. K. Curr Opin Struct Biol 1995, 5, 216.
16. Tomasi, J.; Persico, M. Chem Rev 1994, 94, 2027.
17. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem 1996, 100, 19824.
18. Honig, B.; Nicholls, A. Science 1995, 268, 1144.
19. Jayaram, B.; Liu, Y.; Beveridge, D. L. J Chem Phys 1998, 109, 1465.
20. Ghosh, A.; Rapp, C. S.; Friesner, R. A. J Phys Chem B 1998, 102, 10983.
21. Zacharias, M.; Luty, B. A.; Davis, M. E.; McCammon, J. A. Biophys J 1992, 63, 1280.

22. Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. J Phys Chem 1993, 97, 3591.
23. Eisenberg, D.; Mclachlan, A. D. Nature 1986, 319, 199.
24. Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. Proc Natl Acad Sci USA 1987, 84, 3086.
25. Sitkoff, D.; Sharp, K. A.; Honig, B. J Phys Chem 1994, 98, 1978.
26. Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, S.; Honig, B. J Phys Chem 1996, 100, 11775.
27. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. J Phys Chem B 101, 7147.
28. Nina, M.; Beglov, D.; Roux, B. J Phys Chem B 1997, 101, 5239.
29. Jayaram, B.; Sprous, D.; Beveridge, D. L. J Phys Chem B 1998, 102, 9571.
30. Dominy, B. N.; Brooks, C. L., I. J Phys Chem B 1999, 103, 3765.
31. Froloff, N.; Windemuth, A.; Honig, B. Protein Sci 1997, 6, 1293.
32. Kangas, E.; Tidor, B. J Chem Phys 1998, 109, 7522.
33. HELP, S. E.; Dempster, L. T. C.; Hendsch, Z. S.; Lee, L. P.; Tidor, B. Protein Sci 1998, 7, 206.
34. Cortis, C.; Friesner, R. A. J Comput Chem 1997, 18, 1570.
35. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. J Am Chem Soc 1996, 118, 11225.
36. Gallicchio, E.; Zhang, L. Y.; Friesner, R.; Levy, R. to appear.
37. Rossky, P. J.; Karplus, M. J Am Chem Soc 1979, 101, 1913.
38. Lybrand, T. P.; Ghosh, I.; McCammon, J. A. J Am Chem Soc 1985, 107, 7793.
39. Jorgensen, W. L.; Gao, J. J Am Chem Soc 1988, 110, 4212.
40. Kollman, P. Chem Rev 1993, 93, 2395.
41. Jackson, J. D. Classical Electrodynamics; John Wiley & Sons, Inc.: New York, 1975, 2nd ed.
42. Tamm, I. E. Fundamentals of the Theory of Electricity; Mir Publishers: Moscow, 1979.
43. Born, M. Z Phys 1920, 1, 45.
44. Zauha, R. J.; Morgan, R. S. J Comput Chem 1988, 9, 171.
45. Zauha, R. J.; Morgan, R. S. J Mol Biol 1985, 186, 815.
46. Jorgensen, W. L.; Nguyen, T. B. J Comput Chem 1993, 14, 195.
47. Carlson, H. A.; Nguyen, T. B.; Orozco, M.; Jorgensen, W. L. J Comput Chem 1993, 14, 1240.
48. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. J Comput Chem 1977, 23, 327.
49. Jorgensen, W. J.; Chandrasekhar, J.; Madura, J. D.; Imprey, R. W.; Klein, M. L. J Chem Phys 1983, 79, 926.
50. Jorgesen, W. L.; Madura, J. D. Mol Phys 1985, 56, 1381.
51. Tuckerman, A. M.; Berne, B. J.; Martyna, G. J. J Chem Phys 1992, 97, 1990.
52. Bernsteinn, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, Jr., E. E.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. J Mol Biol 1977, 112, 535.
53. Fremont, D. H.; Matsumura, M.; Stura, E. A.; Peterson, P. A.; Wilson, I. A. Proc Natl Acad Sci USA 1995, 92, 2479.
54. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Biochemistry 1996, 35, 7819.
55. Smith, P. E.; Brunne, R.; Mark, A. E.; van Gunsteren, W. F. J Phys Chem 1993, 97, 2009.
56. Simonsen, T.; Perahia, D. J Am Chem Soc 1995, 117, 7987.
57. Loftler, G.; Schreiber, H.; Steinhauser, O. J Mol Biol 1997, 270, 520.