

Implicit Solvent Models for Protein-Ligand Binding: Insights Based on Explicit Solvent Simulations

Linda Yu Zhang, Emilio Gallicchio and Ronald M. Levy*
Chemistry Department
Rutgers, The State University of New Jersey,
610 Taylor Road, Piscataway, NJ 08855

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 as 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 protein flexibility and thermal motions on charging the peptide in the solvated complex are also considered.

1 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 and accurate 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, and this was a major topic of discussion at the previous Workshop on Electrostatic Interactions in 1980 [1], 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 [2, 3, 4, 5, 6, 7]. Even with much faster algorithms for carrying out simulations with explicit solvent [8, 9, 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, e.g. 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, 12, 13, 14, 15, 16, 17, 18, 19, 20]. The standard approach to parameterizing continuum solvent electrostatic models has been to adjust non-bonded parameters to fit experimental solvation free energies for a database of small organic molecules [21, 22] and/or amino acids and nucleic acids [23]. Several groups have reported very good results fitting the experimental solvation data [21, 24]. 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 [5]. A better appreciation of the complementary strengths and weaknesses of the different solvation

models is likely to be obtained from studies of this kind.

Our workshop presentation summarized recent work in our laboratory 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 [25]. 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 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 [26]. 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 binding. 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 [27, 28].

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 section III 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 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.

2 Computational methods

All the calculations with explicit and implicit solvent were carried out with the IMPACT molecular modeling program package [6]. 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 [24, 29] also implemented

with IMPACT, and with Free Energy Perturbation (FEP) results using explicit solvent. The OPLS all atom force field was used for all the calculations [30]. 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 forty organic molecules in water to the results of FEP simulations with explicit solvent [25]. No additional parameterization was performed for the PBF calculations with OPLS.

2.1 Continuum solvent calculations

There are many levels for treating solvation. The most detailed is a solvent model which includes the solvent molecules (in our case, water) explicitly. Because of the enormous number of degrees of freedom, an explicit solvent calculation will be impractical for many problems in structural biology. A way to remove these degrees of freedom is to treat the solvent implicitly by representing the solvent as a dielectric continuum. The Poisson–Boltzmann (PB) equation is used to find the resulting electrostatic potential, $\phi(\mathbf{r})$, at point \mathbf{r} from the charge density, $\rho(\mathbf{r})$, of the system:

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

where $\epsilon(\mathbf{r})$ is the dielectric constant. While much 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 too slow for database surveys. Faster methods are needed for our 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 will be described in more detail below.

Here we briefly review the Surface Generalized Born (SGB) model. SGB is an approximation to the boundary element formulation of the Poisson–Boltzmann (PB) equation in which the solvent is treated as a continuum with dielectric constant of ϵ and the solute molecule is represented as point charges (q_k is the charge on atom k) inside a cavity formed by the molecular surface. The electrostatic potential, $\phi(\mathbf{r})$, is given by

$$\phi(\mathbf{r}) = \sum_k \frac{q_k}{|\mathbf{r} - \mathbf{r}_k|} + \int_S \frac{\sigma(\mathbf{R})d^2\mathbf{R}}{|\mathbf{r} - \mathbf{R}|}, \quad (2)$$

where the integration of the induced polarization charge density, $\sigma(\mathbf{r})$, is carried out over the surface of the dielectric boundary between solute and solvent. The

induced polarization charge density is given by

$$\sigma(\mathbf{r}) = \frac{1}{4\pi} \left(1 - \frac{1}{\epsilon}\right) \mathbf{E}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}). \quad (3)$$

where $\mathbf{E}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})$ is the normal component of the electric field at the surface of the boundary on the solute side. The normal component of the electric field can be split into two terms: a Coulombic term dependent on the internal charges of the solute and another dependent on the induced polarization charge distribution on the surface. Using the induced polarization charge density, we can find the electrostatic free energy G_{el} with

$$G_{el} = -\frac{1}{2} \int_S \sigma(\mathbf{R}) \phi_0(\mathbf{R}) d^2\mathbf{R}, \quad (4)$$

where the factor of 1/2 is due to the work required to polarize the dielectric and $\phi_0(\mathbf{R})$ is the bare Coulombic potential of the internal charges given by the first term in the right-hand side of equation 2.

The approximation in the GB (and SGB) models is related to the treatment of the induced polarization charge density in the free energy calculations. The spirit of the approximation is the Born sphere, a model where a single point charge, q , is placed in the center of a spherical dielectric cavity with a radius of α . The PB equation can be solved exactly for this case and the free energy of the charge is given by

$$G_{Born} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \frac{q^2}{\alpha}. \quad (5)$$

The important result from this derivation is that, according to the Gauss's Law, the normal component of the electric field due to the induced polarization charge density goes to zero because of the symmetry of the problem leaving only the Coulombic component:

$$\mathbf{E}_{Coul}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = \sum_k q_k \frac{(\mathbf{r} - \mathbf{r}_k) \cdot \mathbf{n}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_k|^3} \quad (6)$$

Even though most molecules are not spherical, in regions that are locally convex, the above Coulombic component may provide a reasonable approximation to the normal component of the electric field at the surface. Any deviations from this could be treated empirically later. So using only the Coulombic term for the induced polarization, the free energy of a single charge, q_k , at point \mathbf{r}_k (known as the self energy) for any *general* surface is given by the surface integral in the SGB model by:

$$G_{se}(q_k, \mathbf{r}_k) = -\frac{1}{8\pi} \left(1 - \frac{1}{\epsilon}\right) \int_S \frac{q_k^2}{|\mathbf{R} - \mathbf{r}_k|^4} (\mathbf{R} - \mathbf{r}_k) \cdot \mathbf{n}(\mathbf{R}) d^2\mathbf{R} \quad (7)$$

The total electrostatic free energy can be rewritten as a sum of self plus pairwise terms:

$$G_{el} = \sum_i G_{se}(q_i, \mathbf{r}_i) + \sum_{i \neq j} G_{pr}(q_i, q_j, \mathbf{r}_i, \mathbf{r}_j), \quad (8)$$

where G_{se} is given by equation 7 and G_{pr} is the free energy component due to the screened PB interaction between a pair of charges. In Still’s formulation of G_{pr} , the central variables to the approximation are the “Born α ” parameters. They appear in the screened energy term for a pair of atoms, i and j , as

$$G_{pr}(q_i, q_j, \mathbf{r}_i, \mathbf{r}_j) = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp(-D)}}, \quad (9)$$

where r_{ij} is the distance between the atom pair, α_i is the “Born α ” parameter for atom i , and

$$D = \frac{r_{ij}^2}{4\alpha_i \alpha_j}. \quad (10)$$

The “Born α ” parameter for each atom represents an effective radius for that atom’s charge as if it was in a Born sphere but with the energy as if that charge was alone in the solute cavity. In other words, α for atom k is found by equating G_{se} (Eq. 7) with G_{Born} (Eq. 5). It should be noted that when r_{ij} becomes sufficiently large, G_{pr} reduces to a Coulombic pair term (first term in Eq. 2) and when the charges overlap, it reduces to the Born equation for superimposing charges.

Ghosh and coworkers have added empirical corrections to the above formalism of the SGB model to accommodate for errors introduced by the the approximations [20]. They were able to achieve close agreement with numerical calculations of the Poisson–Boltzmann equation.

2.2 Free energy perturbation (FEP) calculations

Molecular dynamics (MD) simulations were used to carry out the FEP sampling [7] with explicit solvent. The perturbation 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 host is the murine MHC class I protein H-2KB (PDB name: 1VAC). The ligand was considered to be rigid in these simulations, while some calculations were carried out with the MHC protein rigid, in others protein relaxation was considered as described below. 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 [31]. The calculations included 2116 or 12554 TIP4P [32, 33] 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 [6]. The initial and target temperatures were set to be 298.15 K.

2.3 Ligand Electrostatic Charging Process

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 slowly and evenly for each atom in each window. The double-wide window method [6] was applied, with a total number of 9 windows for the peptide free in water and 10 windows for the peptide bound to the protein performed.

The charging process is divided into ten 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 itself with a full OPLS charge complement. For GLU-6 and LYS-7, the neutral form (denoted by a “n” in the residue name in the accompanying tables) of the residues were obtained by removing the extra net charge. 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 (from SIINFEⁿKⁿL to SIINFEKL). In the tenth 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 configuration 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 set 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 were unconstrained.

For the charging process of the free peptide in water, the uncharged peptide and 2116 TIP4P water molecules were inserted in a box of dimensions $39.8 \times 39.8 \times 39.8 \text{ \AA}^3$. Electrostatic intermolecular interactions were evaluated using the Ewald formula and Lennard-Jones intermolecular dispersion interactions were truncated at 13.5 \AA 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 12554 TIP4P water molecules were inserted in a box of dimensions $74.5 \times 74.5 \times 74.5 \text{ \AA}^3$. Electrostatic and Lennard-Jones intermolecular interactions were truncated at 13.5 \AA implementing a residue based cut-off scheme for the peptide-protein complex and a center of mass based cut-off scheme for the water molecules. Test calculations using the periodic Coulomb potential (Ewald formula) instead of residue based cut-offs 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 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 6 ps of data collection were carried out using a 3 fs time step and the RESPA multiple time step algorithm [8, 34].

3 Results and discussion

3.1 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 as well as the PBF method, which is a rigorous solution to the Poisson-Boltzmann equation using the finite element method [24, 29]. The coordinates of the complex were obtained from the Brookhaven Protein Data Bank (PDB) [35] (PDB name: 1VAC[36].)

In this work the electrostatic binding free energy ΔG_{ebind} is defined according to the thermodynamic cycle I shown in Figure 1,

$$\Delta G_{\text{ebind}} = \Delta G_1 - \Delta G_2 \quad (11)$$

$$= \Delta G_b - \Delta G_a \quad (12)$$

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 Figure 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[26] 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

$$\begin{aligned} \Delta G'_{\text{ebind}} &= \Delta G_1 - \Delta G_3 \\ &= \Delta G_{\text{complex}} - (\Delta G_{\text{protein}} + \Delta G_{\text{ligand}}) \end{aligned} \quad (13)$$

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

$$\begin{aligned} \Delta G'_{\text{ebind}} - \Delta G_{\text{ebind}} &= (\Delta G_b + \Delta G_d - \Delta G_c - \Delta G_a) - (\Delta G_b - \Delta G_a) \\ &= \Delta G_d - \Delta G_c \end{aligned} \quad (14)$$

Thus, to compare our results with results reported according to the alternative definition[26], it is necessary to estimate the charging free energy of the 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 or 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.

3.2 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 1 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 seconds 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 that the OPLS parameters were not explicitly readjusted for the PBF calculations. 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.[37]

The corresponding results for charging the peptide bound to the rigid MHC protein in the solvated complex are shown in Table 2 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 much worse for the peptide in the complex (Table 2) compared with the free peptide in solution (Table 1), 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 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 [25], the charge distribution is in close

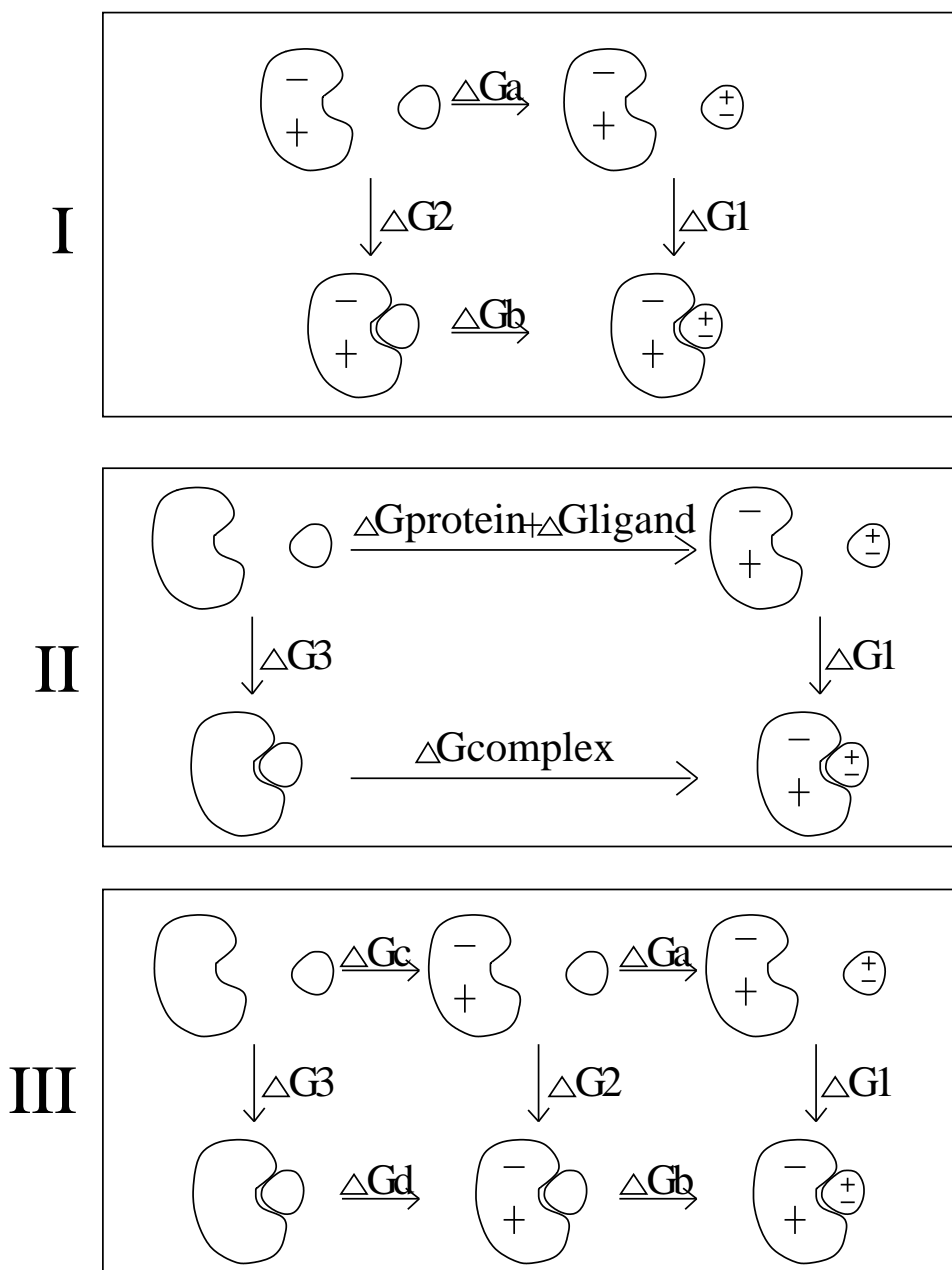


Figure 1: Thermodynamic cycles for calculation of the electrostatic contribution to the binding free energy of a free ligand binding to a free protein in aqueous solution

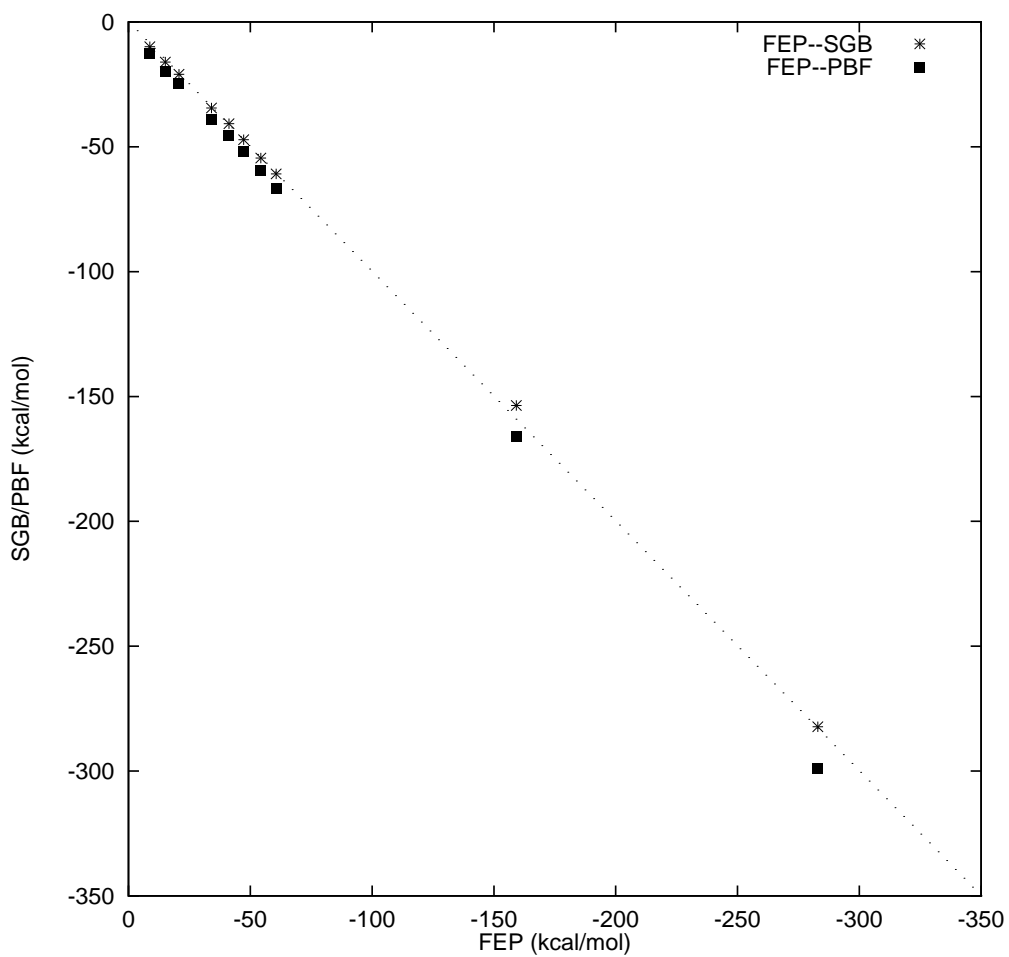


Figure 2: The comparison of the electrostatic solvation free energies of charging the octapeptide OVA-8 in water (kcal/mol) by FEP, SGB and PBF methods. The dashed line has unit slope. The stars indicate the correlation between the FEP and the SGB results, the squares indicate the correlation between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

Table 1: The cumulative solvation free energies of charging the free octapeptide OVA-8 with the sequence SIINFEKL in water (kcal/mol). (For SGB and PBF, $\epsilon_{\text{in}} = 1$, $\epsilon_{\text{out}} = 80$)

Charging steps	$\Delta G_{\text{solv}}^{\text{free}}$		
	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 (x 10)	9 days	7 secs	24 secs

proximity to the solute-solvent surface as it is for the free octapeptide, while in the complex the peptide charges are more buried. We are testing this by additional comparisons of model charging simulations, for which charge burial is an explicit parameter varied.

3.3 Electrostatic Contribution to Protein-Ligand Binding Free Energies

We first consider the electrostatic contribution to the binding of the OVA-8 octapeptide to the rigid MHC protein calculated by FEP simulations with explicit solvent. The results are shown in Table 3 for each charging step individually; the cumulative results are shown in Table 4. 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-2 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 (SIINFEⁿKⁿL), (including the two additional charging steps to create the charged residues and the zwitterionic groups, the direct Coulomb interaction energy is -140 kcal/mol).

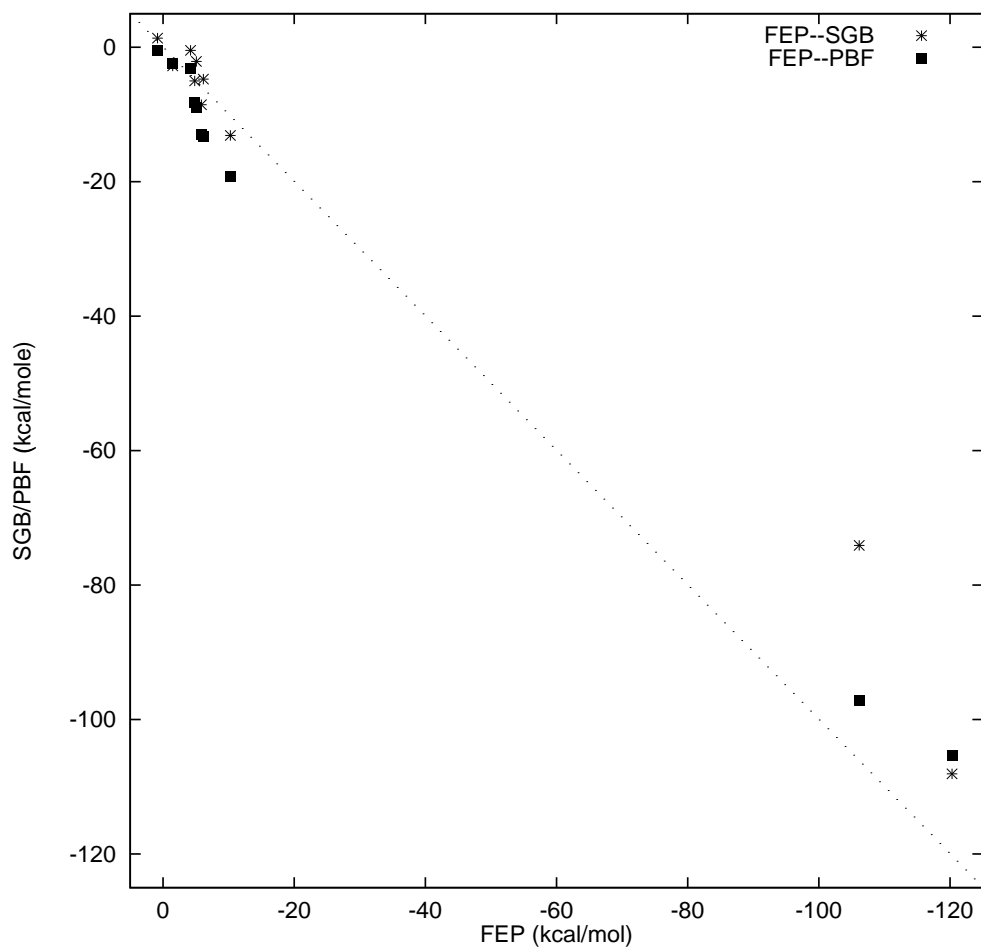


Figure 3: The comparison of the electrostatic solvation free energies ($\Delta G_{\text{solv}}^{\text{bound}}$) of charging the octapeptide OVA-8 bound to the rigid protein H-2KB in water (kcal/mol) by FEP, SGB and PBF methods. The dashed line has unit slope. The stars indicate the correlation between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

Table 2: 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, $\epsilon_{\text{in}} = 1$, $\epsilon_{\text{out}} = 80$)

Charging steps	$\Delta G_{\text{solv}}^{\text{bound}}$		
	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 (x 10)	19 days	2 mins	27 mins

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 4). Ionizing the charged residues (step 9 in Table 3) also favors binding, but creating the zwitterionic form of the peptide changes the equilibrium; however peptide ligands are not normally considered to be in their zwitterionic form in calculations of this kind [26]. 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.[26] 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. In order to compare the results of Froloff et al. to ours, the term given by Eq. (14) 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

Table 3: The FEP calculated, for each charging step, desolvation penalty ($\Delta G_{\text{desolv}} = \Delta G_{\text{solv}}^{\text{bound}} - \Delta G_{\text{solv}}^{\text{free}}$), the direct Coulomb interaction (ΔG_{coul}) and the electrostatic binding free energy ($\Delta G_{\text{ebind}} = \Delta G_{\text{desolv}} + \Delta G_{\text{coul}}$) of the octapeptide OVA-8 with sequence SIINFEKL bound to the rigid protein H-2KB in water (kcal/mol).

Charging step	ΔG_{desolv}	ΔG_{coul}	ΔG_{ebind}
1. S_1	7.24 ± 0.29	-4.78	2.46 ± 0.29
2. I_2	8.69 ± 0.18	-18.84	-10.15 ± 0.18
3. I_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 ⁿ K ⁿ →EK	-1.38 ± 1.04	-13.60	-14.98 ± 1.04
10. SL→S ⁺ L ⁻	109.58 ± 0.88	-46.28	63.30 ± 0.88

“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 “SIINFEⁿKⁿL” , 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 5. 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 8 and 9, 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 solvation free energy of the ionized residues in the complex as calculated with explicit solvent compared with the implicit solvent result. The results suggest that more extensive parameterization of the SGB model for ligands containing functional groups with net

Table 4: The cumulative FEP calculated desolvation penalty ($\Delta G_{\text{desolv}} = \Delta G_{\text{solv}}^{\text{bound}} - \Delta G_{\text{solv}}^{\text{free}}$), the direct Coulomb interaction (ΔG_{coul}) and the electrostatic binding free energies ($\Delta G_{\text{ebind}} = \Delta G_{\text{desolv}} + \Delta G_{\text{coul}}$) 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 ⁿ K ⁿ	44.09 ± 0.65	-59.68	-15.59 ± 0.65
8. SIINFE ⁿ K ⁿ L	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

charge is required in order to model binding phenomena involving such groups.

3.4 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 [38, 39, 40, 41]. 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 protein flexibility can have on the modeling of protein-ligand thermodynamics, we compare in Table 6 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

Table 5: The cumulative FEP, SGB and PBF calculated desolvation penalty ($\Delta G_{\text{desolv}} = \Delta G_{\text{solv}}^{\text{bound}} - \Delta G_{\text{solv}}^{\text{free}}$), the direct Coulomb interaction (ΔG_{coul}) and electrostatic 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, $\epsilon_{in} = 1$, $\epsilon_{out} = 80$)

Charging steps	ΔG_{desolv}			ΔG_{coul}	ΔG_{ebind}		
	FEP	SGB	PBF		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

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 -183 kcal/mol in the FEP simulations of solvated complex when all the protein degrees of freedom are included; the corresponding value is -201 kcal/mol extracted from FEP simulations where the protein has been treated as rigid.

From a preliminary analysis of the FEP trajectories, a physical picture that underlies this result has emerged which involves a combination of two effects. Firstly the thermal motions of the protein screen the favorable protein-ligand electrostatic interactions decreasing the charging energy of the ligand in the bound complex. Secondly, the final geometry of the ligand is somewhat more buried when the protein geometry is allowed to relax. This further decreases the charging free energy of the ligand in the solvated complex relative to the rigid protein result. A more complete analysis of these effects is in progress. One approach proposed to account for protein flexibility and motions in the context of continuum solvent calculations of protein ligand interactions is to run MD simulations with explicit solvent and to then carry out continuum solvent cal-

Table 6: 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).

Charging steps	ΔG^{bound}	
	Rigid	Flexible
1. S	-6.27 ± 0.07	-2.28 ± 0.17
2. SI	-22.80 ± 0.11	-6.35 ± 0.24
3. SII	-24.96 ± 0.11	-11.25 ± 0.28
4. SIIN	-49.42 ± 0.22	-28.74 ± 0.38
5. SIINF	-59.59 ± 0.27	-36.78 ± 0.43
6. SIINFE ⁿ	-55.31 ± 0.28	-34.65 ± 0.44
7. SIINFE ⁿ K ⁿ	-69.94 ± 0.30	-46.59 ± 0.49
8. SIINFE ⁿ K ⁿ L	-87.18 ± 0.33	-57.43 ± 0.52
9. SIINFEKL	-200.77 ± 0.74	-183.29 ± 1.31
10. S ⁺ IINFEKL ⁻	-261.21 ± 0.80	-328.77 ± 1.78

culations on coordinates sampled from the trajectories [42]. Alternatively, one can consider carrying out FEP simulations on an effective potential energy surface which includes solvent effects implicitly; for example by running FEP/MD trajectories including the SGB solvation term in the calculation of the forces.

4 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 realistic and accurate approach to the modeling of solvent effects. However, these simulations are very computationally time consuming, and this has led to extensive efforts to develop implicit solvent models which 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 as 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 This work was supported by a grant (GM30580) from the NIH. We thank Gabriella DelBuono for many helpful discussions and Anthony Felts for assistance with the manuscript.

References

- [1] In *The problem of long-range forces in the computer simulation of condensed media*. National resource for computation in chemistry, January 1980.
- [2] R. M. Levy and E. Gallicchio. *Annu. Rev. Phys. Chem.*, 49:531–567, 1998.
- [3] G. Hummer, L. R. Prutt, and A. Garua. *J. Phys. Chem.*, 102:7885–7895, 1998.
- [4] J. A. Darden, D. M. York, and L. G. Pederson. *J. Chem. Phys.*, 98:10089–10092, 1993.
- [5] F. Figueirido, G. S. Del Buono, and R. M. Levy. *J. Phys. Chem.*, 1996.
- [6] D. B. Kitchen, F. Hiratu, D. A. Kofke, J. D. Westbrook, M. Yormush, and R. M. Levy. *J. Comput. Chem.*, 11:1169–1180, 1990.
- [7] D. L. Beveridge. *Ann. Rev. Biophys. Chem.*, 18:431–464, 1989.
- [8] F. Figueirido, R. M. Levy, R. Zhou, and B. J. Berne. *J. Chem. Phys.*, 106:9835–9849, 1997.
- [9] J. A. Board, J. W. Causey, J. F. Leathrum, A. Windenuth, and K. Schulten. *Chem. Phys. Lett.*, 198:89–97, 1992.
- [10] H. Q. Ding, N. Karasawa, and W. A. Goddard. *J. Chem. Phys.*, 97:4309–4318, 1992.
- [11] J. Warwicker and H. C. Watson. *J. Mol. Biol.*, 157:671, 1982.
- [12] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson. *J. Am. Chem. Soc.*, 112:6129–6129, 1990.
- [13] D. Bashford and M. Karplus. *Biochemistry*, 29:10219–10225, 1990.
- [14] K. Sharp, A. Nicholls, R. Fine, and B. Honig. *Science*, 252:106, 1991.
- [15] M. K. Gilson. *Curr. Opin. Struct. Biol.*, 5:216–223, 1995.
- [16] J. Tomasi and M. Persico. *Chem. Rev.*, 94:2027–2094, 1994.
- [17] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar. *J. Phys. Chem.*, 100:19824–19834, 1996.
- [18] B. Honig and A. Nicholls. *Science*, 268:1144–1149, 1995.
- [19] B. Jayaram, Y. Liu, and D. L. Beveridge. *J. Chem. Phys.*, 109:1465–1471, 1998.
- [20] A. Ghosh, C. S. Rapp, and R. A. Friesner. *J. Phys. Chem. B*, 102:10983–10990, 1998.
- [21] D. Sitkoff, K. A. Sharp, and B. Honig. *J. Phys. Chem.*, 98:1978–1988, 1994.
- [22] B. Jayaram, D. Sprous, and D. L. Beveridge. *J. Phys. Chem. B*, 102:9571–9576, 1998.
- [23] M. Nina, D. Beglov, and B. Roux. *J. Phys. Chem. B*, 101:5239–5248, 1997.
- [24] B. Marten, K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnald, S. Sitkoff, and B. Honig. *J. Phys. Chem.*, 100:11775–11788, 1996.

- [25] L. Y. Zhang, E. Gallicchio, R. M. Levy, A. Ghosh, and R. A. Friesner. in preparation.
- [26] N. Froloff, A. Windemuth, and B. Honig. *Protein Science*, 6:1293–1301, 1997.
- [27] E. Kangas and B. Tidor. *J. Chem. Phys.*, 109:7522–7545, 1998.
- [28] L. T. Chong and S. E. Dempster, Z. S. Hendsch, L. P. Lee, and B. Tidor. *Protein Sci.*, 7:206, 1998.
- [29] C. Cortis and R. A. Friesner. *J. Comput. Chem.*, 18:1570–1590, 1997.
- [30] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. *J. Am. Chem. Soc.*, 118:11225, 1996.
- [31] J. P. Ryckaert, G. Ciccotti, and H. J. Berendsen. *J. Comp. Chem.*, 23:327, 1977.
- [32] W. J. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Imprey, and M. L. Klein. *J. Chem. Phys.*, 79:926, 1983.
- [33] W. L. Jorgesen and J. D. Madura. *Mol. Phys.*, 56:1381, 1985.
- [34] A. M. Tuckerman, B. J. Berne, and G. J. Martyna. *J. Chem. Phys.*, 97:1990, 1992.
- [35] F. C. Bernsteinn, T. F. Koetzle, G. J. Williams, E. E. Meyer Jr, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. *J Mol Biol*, 112:535, 1977.
- [36] D. H. Fremont, M. Matsumura, E. A. Stura, P. A. Peterson, and I. A. Wilson. *Proc. Natl. Acad. Sci. USA*, 92:2479–2483, 1995.
- [37] F. Hirata, P. Rejfern, and R. M. Levy. *J. Quantum Chem.*, 15:179–188, 1988.
- [38] J. Antosiewicz, J. A. McCammon, and M. K. Gilson. *Biochemistry*, 35:7819–7833, 1996.
- [39] P. E. Smith, R. Brunne, A. E. Mark, and W. F. van Gunsteren. *J. Phys. Chem.*, 97:2009–2014, 1993.
- [40] T. Simonsen and D. Perahia. *J. Am. Chem. Soc.*, 117:7987–8000, 1995.
- [41] G. Loftler, H. Schreiber, and O. Steinhauser. *J. Mol. Biol.*, 270:520, 1997.
- [42] T. Simonson, G. Archontis, and M. Karplus. *J. Phys. Chem. B*, 101:8349, 1997.