Enzyme-Inhibitor Association Thermodynamics: Explicit and Continuum Solvent Studies

Haluk Resat, Tami J. Marrone, and J. Andrew McCammon
Department of Chemistry and Biochemistry, and Department of Pharmacology, University of California at San Diego, La Jolla, California 92093-0365 USA

ABSTRACT Studying the thermodynamics of biochemical association reactions at the microscopic level requires efficient sampling of the configurations of the reactants and solvent as a function of the reaction pathways. In most cases, the associating ligand and receptor have complementary interlocking shapes. Upon association, loosely connected or disconnected solvent cavities at and around the binding site are formed. Disconnected solvent regions lead to severe statistical sampling problems when simulations are performed with explicit solvent. It was recently proposed that, when such limitations are encountered, they might be overcome by the use of the grand canonical ensemble. Here we investigate one such case and report the association free energy profile (potential of mean force) between trypsin and benzamidine along a chosen reaction coordinate as calculated using the grand canonical Monte Carlo method. The free energy profile is also calculated for a continuum solvent model using the Poisson equation, and the results are compared to the explicit water simulations. The comparison shows that the continuum solvent approach is surprisingly successful in reproducing the explicit solvent simulation results. The Monte Carlo results are analyzed in detail with respect to solvation structure. In the binding site channel there are waters bridging the carbonyl oxygen groups of Asp189 with the NH_2 groups of benzamidine, which are displaced upon inhibitor binding. A similar solvent-bridging configuration has been seen in the crystal structure of trypsin complexed with bovine pancreatic trypsin inhibitor. The predicted locations of other internal waters are in very good agreement with the positions found in the crystal structures, which supports the accuracy of the simulations.

INTRODUCTION

Understanding how reactant molecules associate, and the relative stabilities of their different arrangements, has long been an important goal of the chemical sciences. In biochemical applications, knowledge of the free energy profile (i.e., the potential of mean force between the solutes) can aid in the design of biologically active agents, and it is helpful in the analysis of binding kinetics (Lybrand, 1995; Marrone et al., 1997). Although knowledge of the potential of mean force (pmf) is valuable in the understanding of molecular associations, calculations of such quantities using simulations with explicit solvent models have mostly been limited to small solutes because of computational expense (Pratt et al., 1994).

Appropriate sampling of the solvent degrees of freedom is essential for an accurate determination of free energies of binding. However, configurational sampling is especially difficult—and the calculations are correspondingly more expensive—when binding of the solutes involves the trapping of solvent molecules within concave binding surfaces. For example, in ligand-receptor binding, water molecules may have to be emptied out of a binding channel as the ligand starts to penetrate. The capping of the channel by the ligand may form a confined space, i.e., a water pocket, and this pocket gets smaller as the ligand moves down the channel. Using the usual canonical or microcanonical simulation methods, it would be very hard to represent the equilibration of the water molecules that must be expelled. However, as discussed in earlier reports (Resat and Mezei, 1994, 1996; Resat et al., 1996), the sampling difficulties due to the changes in the solvation pattern during docking can be relieved in principle by utilizing grand canonical rather than the more conventional canonical or microcanonical ensemble simulation methods. The grand canonical ensemble Monte Carlo (GCMC) simulation method has been shown to be quite efficient in obtaining good statistical convergence for predicting the solvation patterns in the crystal hydrates of a polydisaccharide (Resat and Mezei, 1994) and of a nucleic acid/drug system (Resat and Mezei, 1996). The equivalence of the grand canonical and canonical ensemble in pmf calculations was shown by computing the pmf between oppositely charged ions and between small hydrophobic particles (Resat et al., 1996). Following up on the earlier successes in overcoming the confined space effects, we employ the GCMC simulation method in the present work to calculate the pmf along a one-dimensional curvilinear reaction coordinate between the enzyme trypsin and its inhibitor benzamidine, a prototypical ligand-receptor system.

Although explicit solvent simulations are the preferred choice whenever feasible, their computational expense makes pmf calculations difficult for large biomolecular as

Received for publication 26 August 1996 and in final form 5 November 1996.

Address reprint requests to Dr. Tami J. Marrone, Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, Department 0365, La Jolla, CA 92093-0365. Tel.: 619-822-0255; Fax: 619-534-7042; E-mail: tmarrone@ucsd.edu.

Dr. Resat's present address is Department of Physics, Koc University, Istinye, Istanbul 80860, Turkey. E-mail: hresat@ku.edu.tr.

Dr. McCammon's E-mail address is jmccammon@ucsd.edu.

0006-3495/97/02/522/11 $2.00

semblies. The use of simplified but less expensive theoretical approaches such as continuum dielectric representations of solvent (Madura et al., 1994; Honig and Nicholls, 1995; Gilson, 1995) or integral equation methods (Friedman, 1985; Pettitt et al., 1986) can be helpful for such cases. For this reason, another aim of the present work is to compare the results of the explicit solvent GCMC simulations with those of continuum solvent Poisson equation calculations. The benzamidine-trypsin system should be a particularly difficult test case for continuum methods, because individual water molecules bridge charged solute groups during binding. Nevertheless, our investigation with different sets of parameters shows that the continuum solvent approach performs surprisingly well.

POTENTIAL OF MEAN FORCE

We have recently presented the details of the grand canonical ensemble pmf calculations. Therefore, the underlying theory will be repeated here only briefly; further details may be found in Resat et al. (1996) and in Friedman (1985). The pmf, $W(\mathbf{R})$, is the free energy of a system as a function of a selected set of coordinates (reaction coordinates, $\mathbf{R}$). It is related to the equilibrium probability distribution ρ of states as

$$\beta W(\mathbf{R}) = -\ln \rho(\mathbf{R}) + \mathcal{C}, \qquad (1)$$

where $\beta = 1/k_B T$, T is the temperature, and C is a constant. In the above equation, the reaction coordinate $\mathbf{R}$ may be multidimensional.

In many instances, calculating the probability distribution function $\rho(\mathbf{R})$ using a direct Boltzmann sampling in a computer simulation is not practical (Mezei and Beveridge, 1986; Straatsma and McCammon, 1992). Therefore, to obtain adequate sampling, the pmf calculations are generally done utilizing non-Boltzmann or biased (umbrella) sampling (Patey and Valleau, 1975). Even though the use of a biasing potential alters the Hamiltonian $\mathcal{H}$ used in the simulations, it can be shown that the effects of the biasing can be eliminated from the calculated thermodynamic quantities. If the biasing potential is $U_b(\mathbf{R})$, then the pmf is given as

$$\beta W(\mathbf{R}) = -\ln \rho_b(\mathbf{R}) - \beta U_b(\mathbf{R}) + \ln\langle e^{\beta U_b(\mathbf{R})}\rangle_b + \mathcal{C}, \qquad (2)$$

where $\rho_b(\mathbf{R})$ and $\langle\cdots\rangle_b$, respectively, are the probability distribution and the ensemble average of the enclosed quantity calculated using the biased Hamiltonian, $\mathcal{H} + U_b(\mathbf{R})$. As Eq. 2 shows, a uniform sampling of the distribution of states, i.e., ρ_b = constant, would be obtained if an optimal choice for the biasing potential, $U_b(\mathbf{R}) = -W(\mathbf{R})$, could be made. In most cases, the pmf does not have a simple form, and it is difficult to choose a near-optimal biasing potential. To avoid such problems and to obtain the optimal sampling, Mezei (1987) developed the adaptive umbrella sampling scheme, which is used in this study. In the adaptive umbrella sampling algorithm, the optimal umbrella sampling potential is self-consistently determined and refined during the simulation, and the simulation is run until an acceptable convergence is obtained. In other words, the biasing potential is updated and adapted at regular intervals during the simulation.

COMPUTATIONS

Crystal structure

The crystal structure for the trypsin-benzamidine complex was taken from the Protein Data Bank (PDB) archives (name: 3ptb.pdb); it had been deposited by Bode and Schwager (1975). It is the structure for bovine β-trypsin refined to 1.8-Å resolution with an R factor of 0.23. The estimated standard deviation of the atomic coordinates is less than 0.1 Å. The amidinium group of the inhibitor benzamidine is twisted out of the benzene plane by about 7°. Coordinates of the waters used in the analysis of the bridging waters at the binding site, discussed in the Results/Internal waters section, were taken from the crystal structure of the free trypsin (PDB: 2ptn.pdb) deposited by Walter et al. (1982). For this, the free trypsin crystal structure was superimposed on the trypsin structure in the enzyme-inhibitor complex. The two structures are almost identical. An extensive study of solvent structure in crystals of trypsin may be found in Finer-Moore et al. (1992).

Reaction coordinate assignment

Even for small flexible molecules, the reaction coordinate in Eq. 1 will typically have many degrees of freedom. This large dimensionality makes the computational approach difficult and expensive. For this reason, computations are usually feasible only when various approximations are employed. For example, the free energy calculations can be performed along a certain reaction coordinate of reduced dimensionality, assuming that the predetermined reaction coordinate appropriately represents the physical process. For this initial study of a biomolecular system, we simply use rigid geometries for the enzyme and the inhibitor, together with a prescribed path of relative displacement. Another major reason for using rigid molecules was to be able to appropriately compare with the continuum solvent (Poisson equation) calculations.

The reaction coordinate was defined as a curvilinear one-dimensional path, along which the inhibitor slightly rotates as it moves down the binding site channel. Although detailed realism is not a primary goal of the current study, the use of rigid molecules is perhaps not a drastic approximation, because the crystal structures of free trypsin and trypsin inhibited by benzamidine are very similar (Bode and Schwager, 1975; Bode et al., 1976; Marquart et al., 1983). Earlier simulation studies also suggest that the structural fluctuations in and around the binding site channel are quite small, irrespective of the inhibitor position (Wong and McCammon, 1986a,b; Luty et al., 1995).

The reaction coordinate (Fig. 1) was determined by visual analysis of the binding site channel. For simplicity, the enzyme was fixed during the simulation, and only the inhibitor was moved along the reaction coordinate as a function of the coupling parameter. The motion of benzamidine was defined, using analytical formulas, as a curved line that lies approximately at the centerline of the binding site channel. The curvature of the linear path was small, and the inhibitor was also slightly rotated around its central axis as it was translated. The maximum span of the path was such that benzamidine does not go beyond the entrance to the binding site channel, because it is difficult to define a unique path outside this channel. It has been shown in earlier studies that once the benzamidine enters the channel, it will bind with almost absolute certainty. Docking studies (Luty et al., 1995) further suggest that benzamidine does not tumble inside the channel. These observations and the large association constant (Mares-Guia et al., 1977; Bode and Schwager, 1975) help to motivate the rigid geometry assumption and the empirical selection of the reaction coordinate.

Protonation state assignment

The protonation states of the trypsin residues were determined using the algorithm developed by Antosiewitz et al. (1994). The pK_a calculations were performed for the isolated protein, and all polar residues were explicitly included in the calculations. The pH was 7. The computed protonation states were equivalent to the charge states calculated using the standard model at this pH; Lys and Arg residues were positively charged, and Glu and Asp residues were negatively charged. The N- and C-termini were in their charged states. For His57 and Tyr39, the Antosiewitz et al. algorithm predicted charges of +0.289 and −0.145, respectively; however, we chose to assign total charges of zero for those two residues to be able to use standard force-field parameters. Thus, the total charge of trypsin was +6 (= 14 Lys + 2 Arg − 4 Glu − 6 Asp + Ile16 − Asn245).

The site charges for the inhibitor benzamidine were derived from the electrostatic potentials determined using quantum calculations employing the 6–31g** basis set (Fig. 2). Starting with the crystal coordinates, the Gaussian-92

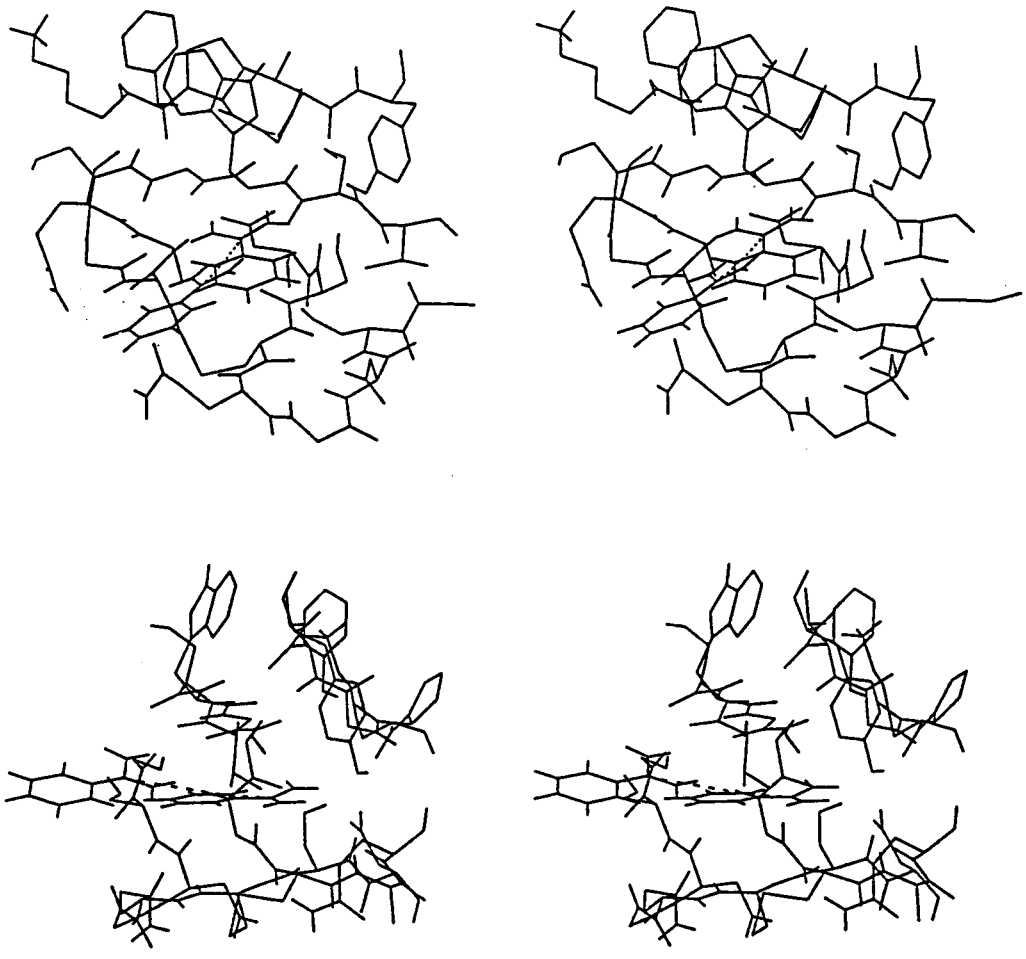

FIGURE 1 Reaction coordinate utilized in the calculations. Stereo pictures from two different angles show the path traveled by benzamidine. For clarity, only a portion of trypsin is shown, and the dashed line connects the locations of C7:Ben along the path.

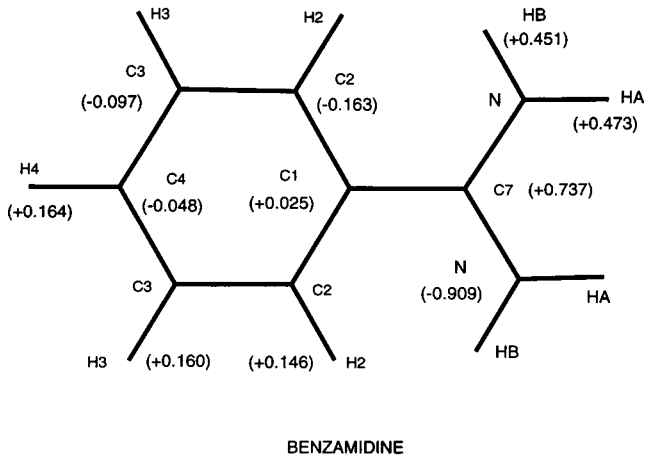

FIGURE 2 Benzamidine site labels and charges used in the simulation.

program (Frisch et al., 1992) was used to calculate the electrostatic potential distribution, which was then fitted to derive the site charges, using Merz-Kollman algorithm as implemented in the same program. All hydrogens were explicitly included, and the total charge of the inhibitor was +1.

Grand canonical Monte Carlo simulation setup

Starting with the published crystal structure (Bode and Schwager, 1975), the trypsin-benzamidine complex was rotated in the simulation unit cell until all atoms were separated from any of the closest image atoms by more than 12 Å. This optimization also placed the binding site region in such an orientation that the movement of the inhibitor along the defined reaction pathway was almost along a principal diagonal of the unit cell. This optimized orientation minimized the replication effects due to the applied periodic boundary conditions. The resulting rectangular unit cell box size was 51.56 × 51.26 × 52.46 $Å^3$. Using the volume available to the solvent, the unit cell should contain approximately 3664 waters. The excess chemical potential was adjusted appropriately to obtain the estimated solvent density.

As discussed in the previous section, pK_a studies resulted in the standard residue protonation states at pH 7. For the enzyme, only polar hydrogens and the capping protons of the N-terminus residue were included as hydrogen interaction sites, and united atom models were employed for nonpolar protons. Site charges and short-range interaction parameters for the protein were taken from the OPLS force field library (Jorgensen and Tirado-Rives, 1988). As described above, inhibitor site charges (Fig. 2) were determined using ab initio methods, and all hydrogens were included as interaction sites. The short-range parameters for the ring part were assigned using the OPLS parameters for benzene, and analogous arginine parameters were assigned to the amidinium group. The water molecules were modeled with the TIP3P model (Jorgensen et al., 1983).

The solute-solvent and solvent-solvent interactions were truncated, using a spherical cutoff at 11 Å. The solute-solvent cutoff criteria were based on the distance from a water to the nearest site of the solute groups, and each protein residue and the benzamidine were treated as a separate group. Solute and solvent move step sizes were chosen such that the average move acceptance rate was approximately 50%. To accelerate the sampling rate, distance-dependent preferential sampling was employed both for the regular moves and for the insertion/deletion attempts. The distance dependence of the preferential sampling was implemented using the carboxylate group of Asp189 as the center. The preferential sampling adopted a $1/r^3$ form for waters at distances greater than 8 Å from the Asp 189 carboxylate group, and the preferential sampling weight function was uniform for distances smaller than 8 Å. This allowed for an enhanced sampling of the waters in or around the binding site channel.

The reaction coordinate was divided into four windows to reduce the size of the reaction coordinate region covered in each simulation for more efficient sampling. The regions were $R < 5.84$ Å, 5.36 Å $< R <$ 6.78 Å, 6.31 Å $< R <$ 8.78 Å, and 8.43 Å $< R$ for windows 1 to 4, respectively. The constant (μ,V,T) ensemble grand canonical Monte Carlo (GCMC) molecular simulations (Adams, 1975; Resat et al., 1996) for each region were performed using adaptive umbrella sampling (Mezei, 1987). Simulations were run for 30, 60, 50, and 48 million steps (one step = one regular move + one insertion/deletion step) for windows 1 to 4, respectively, at 298 K. The convergence of the simulations was determined by the changes in the calculated pmf, and the estimated statistical sampling error was approximately ±1 kcal/mol.

Poisson equation calculations

To determine how well they compare with the explicit solvent simulation results, the pmf utilizing a continuum solvent was also computed using the Poisson equation (Honig and Nicholls, 1995; Gilson, 1995; Madura et al., 1994; Pratt et al., 1994; Soman et al., 1989). The Poisson equation was solved using a finite-difference algorithm and focusing within the UHBD program (Davis et al., 1991). Calculations utilized a 110^3 grid with focusing to a final grid spacing of 0.25 Å. Dielectric constants of 1 and 78 were assigned for the trypsin-benzamidine complex and for the continuum solvent, respectively. The dielectric boundary between the solvent and the biomolecules was defined using a molecular surface with different surface probe radii (further details are given in the next section). To ensure proper focusing, the focusing volume was defined after investigating how the electrostatic potentials on protein sites change as the reaction coordinate is varied. This was done by comparing the electrostatic potentials when the inhibitor was at its bound position V_B and when the inhibitor was at the entrance to the channel V_U, i.e., at the opposite ends of the investigated pathway. Fig. 3 reports the difference $\Delta V =$

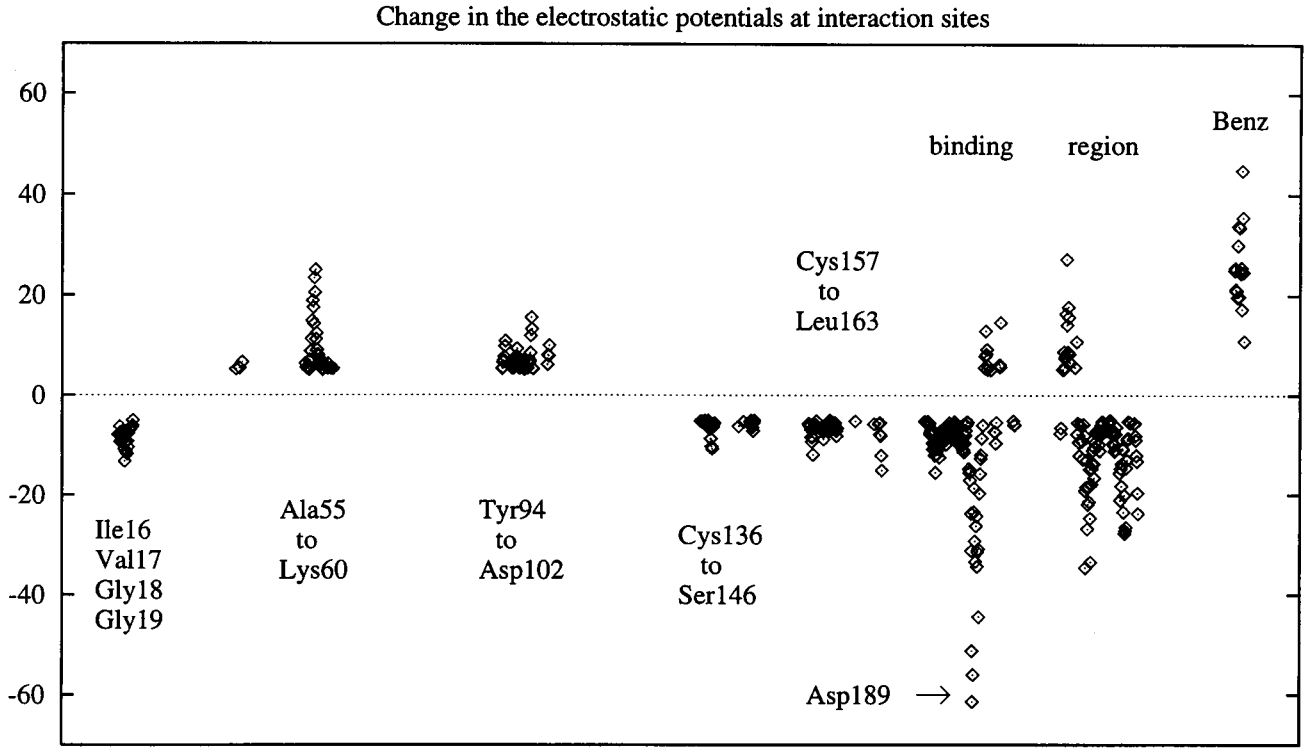

FIGURE 3 Changes in the local electrostatic potentials as calculated using the Poisson equation, where the solvent is represented as a continuum dielectric. The figure reports the difference in the electrostatic potentials calculated when the inhibitor is at its bound position and is at the entrance of the binding site channel (R = 3.89 Å and 9.79 Å in Fig. 4). Binding region is defined by residues 172, 183 to 195, and 214 to 228.

$V_U - V_B$. Note that the sites with large ΔV's are expected to be sensitive to the electrostatic calculations, and therefore it was assured that those sites and their vicinity were explicitly included in the computations during the focusing stage. Also notice that the reported ΔV is relevant to the reaction-coordinate-dependent protonation state determination, which, however, was not pursued in this work. Using various error estimation methods, such as a half-grid spacing shift, the error in the Poisson calculations was estimated to be ±2 kcal/mol.

RESULTS

Explicit solvent GCMC simulations

Because the molecular geometries are kept rigid, the total pmf $W(\mathbf{R})$ can be separated into the direct intermolecular interaction potential energy $U(\mathbf{R})$ between the enzyme and the inhibitor (Fig. 4 *a*) and the solvation parts $W_s(\mathbf{R}) = W(\mathbf{R}) - U(\mathbf{R})$ (Fig. 4 *b*). The interaction potential energy highly favors the complex formation, which partly explains the strong inhibitory behavior of benzamidine. The abscissa in Fig. 4 (and in Figs. 7 and 8) gives the reaction coordinate values as the distance between the carbon of the amidinium group of benzamidine C7:Ben and the carboxyl carbon CG of Asp189. Fig. 4, *b* and *c*, respectively, reports the solvation contribution to the pmf and the total pmf. As Fig. 4 *b* shows, there are three distinct regions in the solvation contribution to the pmf: two plateaus at small and large distances, and a region connecting them.

Fig. 5 shows how the region between Asp189 and benzamidine is occupied by the waters. As benzamidine starts to separate from Asp189 (Fig. 5, *a* and *b*), one or two waters begin to form linkages between the two charged groups. When benzamidine is pulled farther away (Fig. 5 *c*), one of the bridging waters, as well as the water 416, change their location to keep the hydrogen bond chain intact. Because the linking waters in Fig. 5 *b* are strongly hydrogen bonded to two ionic groups, they have a very favorable free energy, and their removal from the bridging configuration contributes to the free energy barrier seen in Fig. 4 *b*. However, the association thermodynamics of the enzyme and the inhibitor represent the competition of two opposing effects. The strong direct interaction energy (Fig. 4 *a*) between the enzyme and the inhibitor dominates the unfavorable solvent contribution (Fig. 4 *b*), and the inhibitor bound complex is the highly preferred configuration (Fig. 4 *c*). Thus, as experimentally measured, the association constant is very high and benzamidine is a good inhibitor of trypsin (Mares-Guia et al., 1977). Bridging waters as seen in our results have also been implied by crystallography experiments; one water strongly hydrogen bound to Asp189 has actually been detected in the crystal structure of trypsin with the bovine pancreatic trypsin inhibitor (Huber et al., 1974; Marquart et al., 1983). The same bound water, and a second one (Fig.

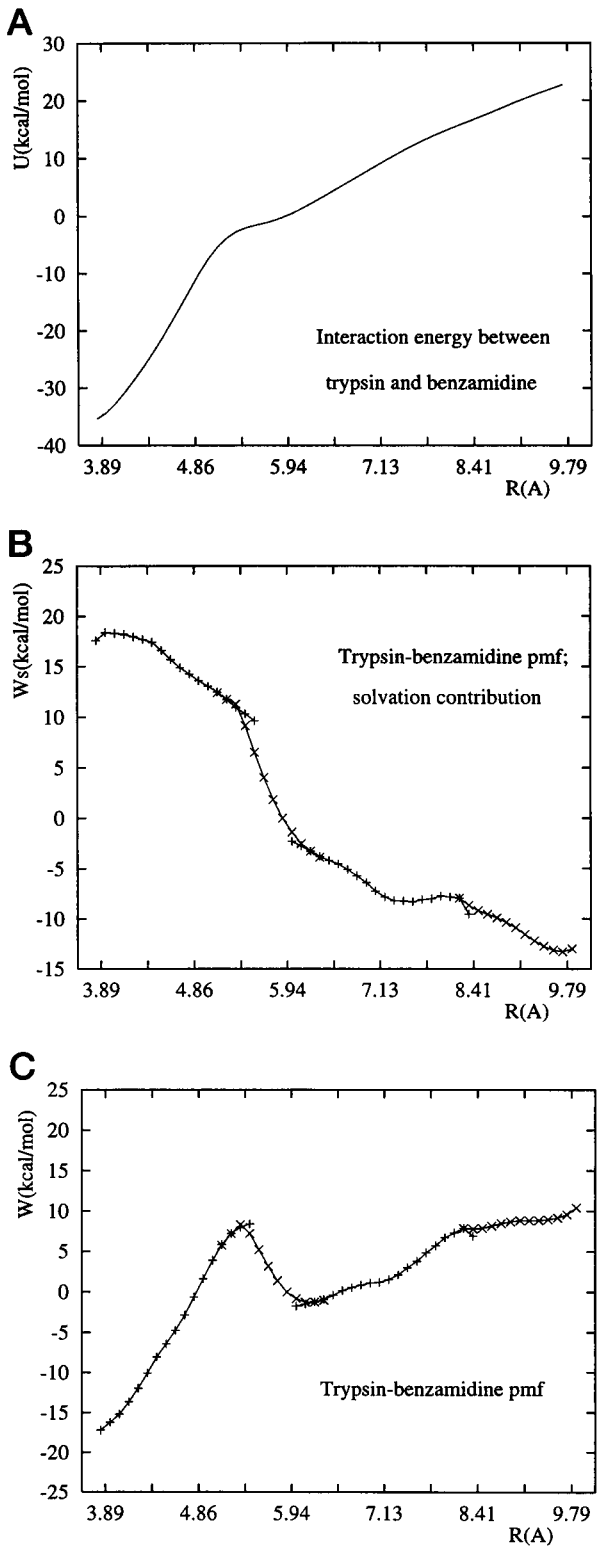

FIGURE 4 (*a*) Potential energy of interaction between trypsin and benzamidine as a function of the reaction coordinate, $U(\mathbf{R})$. (*b*) Solvation contribution, $W_s(\mathbf{R})$. (*c*) Total potential of mean force between trypsin and benzamidine as calculated in the GCMC simulations. The abscissa represents the distance between the C7 site of benzamidine and CG of Asp189 and is in angstroms.

5 *c*), has also been detected in the crystal structure of free trypsin (Bode et al., 1976).

The distribution of the number of waters during the GCMC simulation run (Fig. 6) closely resembles a Gaussian with a halfwidth at half-maximum of approximately 20 waters. The average excess chemical potential of the waters in the simulation cell was −6.16 kcal/mol. As expected, this value is very close to the bulk excess chemical potential of the same water model, TIP3P: −6.4 ± 0.5 kcal/mol (Beglov and Roux, 1994).

Poisson equation calculations using continuum dielectric solvent

As discussed above, simpler and computationally cheaper approaches are of interest for calculations of the thermodynamic properties of biochemical systems. For example, solvation free energies are frequently calculated using empirical formulas in terms of the solvent-accessible surfaces/volumes (Eisenberg and McLachlan, 1986), and using the Poisson equation in which the solvent is typically treated as a homogeneous dielectric continuum (Madura et al., 1994), or a combination of both (Sitkoff et al., 1994). Even though such approaches are somewhat inferior to the explicit solvent models in terms of physical representation, they can at least be very useful in studying the general trends. In particular, the Poisson equation has been shown to be quite successful in predicting solvation free energies (Gilson, 1995; Honig and Nicholls, 1995). Although the Poisson equation has been applied to a wide range of systems, its use in pmf studies has been somewhat limited (Rashin and Namboodiri, 1987; Pratt et al., 1994).

Although it is an analytical formula, the Poisson equation includes parameters that are determined empirically. Definition of the dielectric boundary between solutes and the dielectric continuum, the dielectric constants of the interior and the exterior of the solutes, and the solvent exclusion radius of polar hydrogens all contribute to the uncertainties in the calculations. Here we use a couple of values for each parameter and compare the results with the explicit solvent GCMC simulation results.

The dielectric boundary in this work is based on the so-called molecular surface (Connolly, 1983). The molecular surface is defined as the surface formed by the points touched by the surface of a spherical probe as it is rolled over the van der Waals volume of the solute. Use of the molecular surface thus avoids the sharp grooves that are often found in van der Waals surfaces and that are typically assigned solvent dielectric values, even though the grooves may be too narrow to be occupied by the solvent. The probe representing an aqueous environment is generally chosen as a sphere of radius 1.4 Å (Madura et al., 1994). Recently, Rashin (1989) advocated a much smaller probe radius of 0.8 Å when molecular association mechanisms are investigated. Horvath et al. (1996) used an even smaller probe radius, 0.5 Å. Rashin argued that behavior of the waters bridging the two solute molecules would be quite different from that of

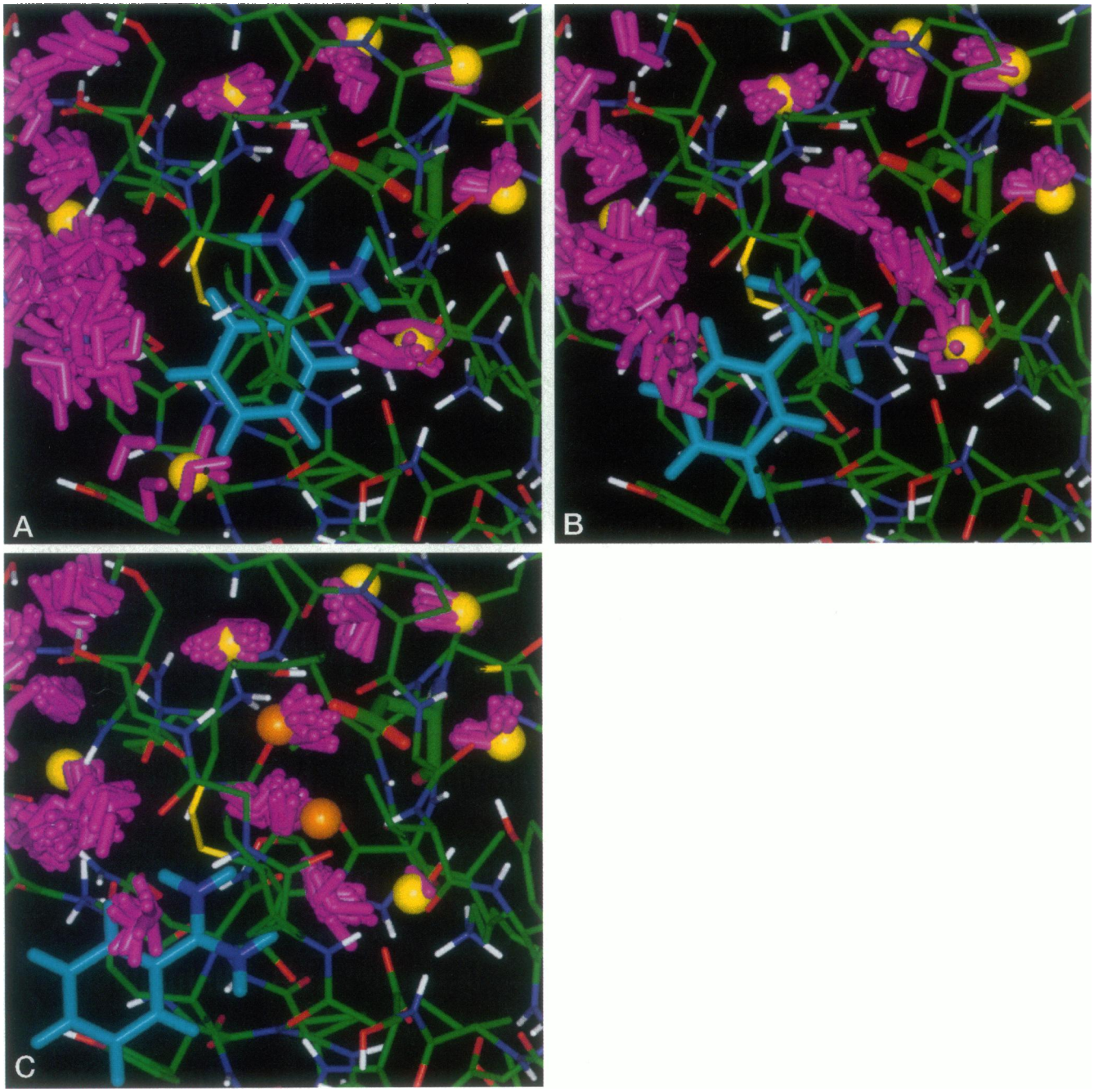

FIGURE 5 Typical snapshots from the GCMC simulations from three different ranges of the reaction coordinate R. (*a*) $R \simeq 3.89$ Å. (*b*) $R \simeq 6.06$ Å. (*c*) $R \simeq 8.21$ Å. Trypsin atoms are color coded according to the atom type: red, oxygen; green, carbon; dark blue, nitrogen; white, hydrogen; yellow, sulfur. Anchoring Asp189 residue is pictured with thicker lines than the other residues. Benzamidine is shown as light and dark blue. Yellow spheres represent the water oxygens detected in the experiments. The labels for the waters in *a*, in clockwise order starting from the leftmost molecule, are w802, w415, w562, w705, w704, w416, and w710. Because it overlaps with the inhibitor, w710 is not shown in *b* and *c*. In *c*, orange spheres show the additional waters detected in the crystal structure of free trypsin. Water molecules of the GCMC simulations are coded in magenta. For clarity, only a part of the enzyme is shown.

bulk waters. Because of their limited mobility, particularly along the bridging direction, the radius of the bridging waters would be effectively reduced—hence a smaller probe radius. To test his hypothesis, we repeated the Poisson equation pmf calculations with a probe radius of 0.8 Å.

Because they carry large partial charges, polar hydrogens that lie close to the molecular surface pose numerical problems by giving rise to artificially large polarization energies. To avoid such problems, in the Poisson equation calculations the dielectric boundaries are defined by adding a small repulsion core to the polar hydrogens. Accurate treatment of polar hydrogens is particularly important in our case. This follows from the fact that the amidinium group has four exposed hydrogens, which are highly polar. Therefore, we used two

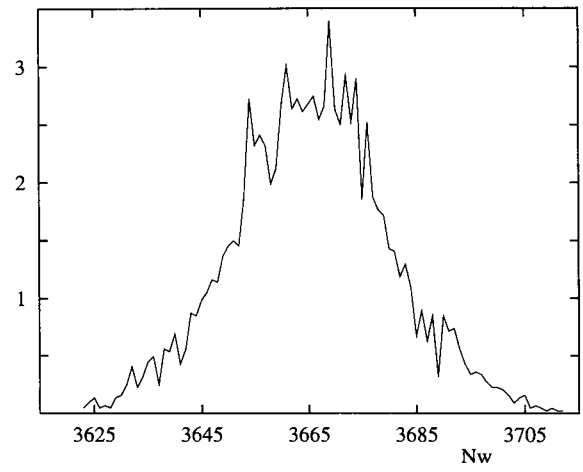

FIGURE 6 Distribution of the number of waters in the GCMC simulations. The plot shows the percentage of simulation configurations having a certain number of waters *Nw*. It is roughly a Gaussian distribution, with a halfwidth at half-maximum of approximately 20 waters.

different polar hydrogen radius values, 1.0 and 1.2 Å, to gauge the sensitivity of the results to variations of this parameter.

Recently there have been attempts to add the effects of nonelectrostatic interactions to Poisson equation estimates of solvation free energies (Sitkoff et al., 1994; Simonson and Brunger, 1994). These approaches generally use an empirical formula based on the molecular surface area. Adding this term, with a coefficient of 6 cal/mol/Å^2, did not influence the results of the present work. This is due to the fact that the total surface area of the enzyme-inhibitor complex changes very little as a function of the reaction coordinate.

Fig. 7 reports the results for three different sets of Poisson equation calculations employing dielectric constants of 1 and 78 for the interior of the biomolecular complex and for the continuum solvent regions, respectively. Comparison of curves *a* and *b* shows that the effect of the polar hydrogen radius assignment can be quantitatively significant, ~4 kcal/mol. However, the effect of the probe radius selection is much more important, especially at small distances. This can be explained using arguments similar to Rashin's. With a probe radius of 1.4 Å, the solvent cavity starts to be formed at around R = 6.5 Å as the inhibitor moves away from the enzyme. Up to this value of R, the dielectric map is left unchanged around the amidinium group and the carboxylate of Asp189, so that the solvation free energy of the enzyme-inhibitor complex stays approximately constant. Changes in the free energy start at smaller distances for smaller probe radius. The best agreement with the GCMC simulations is obtained for the probe radius advocated by Rashin. The results that give the best agreement, for probe radius = 0.8 Å and polar hydrogen radius = 1.2 Å, are compared with the GCMC simulations in Fig. 8. Because there is a strongly hydrogen-bonded, bridging water molecule with structural behavior very different from that of the bulk water, one expects that the continuum solvent model would not perform well, at least at smaller distances. Our results show, however, that contrary to our expectations, the agreement with the explicit solvent simulation results is surprisingly good. The comparison is limited, however, by the fact that the two curves are only determined to within a constant. We have chosen to match the curves at smaller separations, both because this is the best characterized by experiment, and because it better matches the features in the curves. Note also that the electrostatic cutoff effects are expected to be more important at larger separations, and matching at the bound structure minimizes such errors.

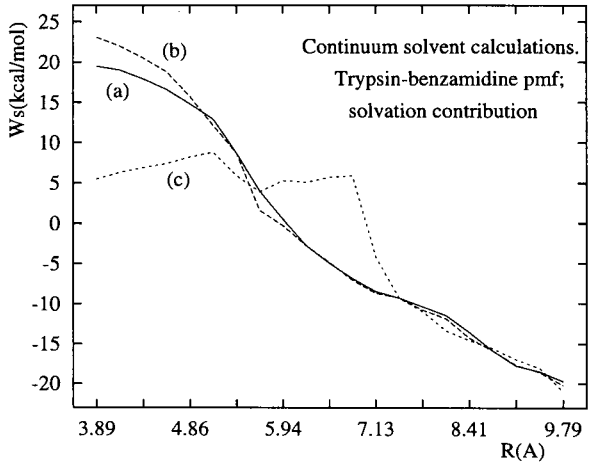

FIGURE 7 Solvation contribution to the potential of mean force, $W_s(\mathbf{R})$, between trypsin and benzamidine, calculated using the Poisson equation. As explained in the text, the three curves represent the results of different parameter sets: (*a*) Probe radius *P* defining the molecular surface = 0.8 Å and the radius of polar hydrogens R_h = 1.2 Å. (*b*) p = 0.8 Å and R_h = 1.0 Å. (*c*) p = 1.4 Å and R_h = 1.2 Å. The abscissa represents the distance between the C7 site of benzamidine and CG of Asp189 and is in angstroms.

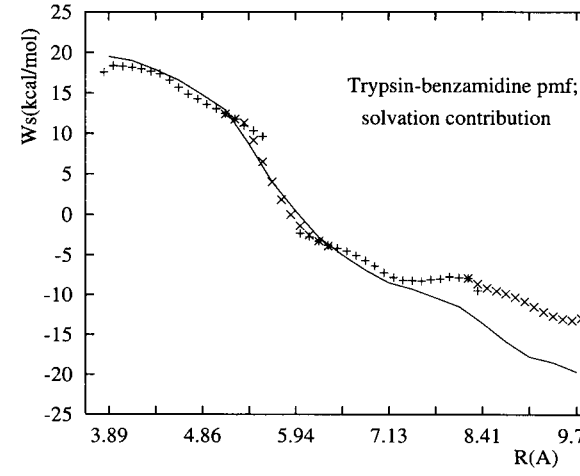

FIGURE 8 Solvation contribution to the potential of mean force. Comparison of the continuum solvent Poisson equation and explicit solvent GCMC simulations. The abscissa represents the distance between the C7 site of benzamidine and CG of Asp189 and is in angstroms.

Internal waters

Another way of evaluating the success of the GCMC simulations is by analysis of the buried waters in the protein structure. Reproduction of the internal waters observed in experiments would show that the solvation patterns sampled in the simulations appropriately represented the actual system, and that the force fields were satisfactory. For this purpose, we computed the predicted locations for the internal waters. To mimic the crystallographic experimental conditions in which the inhibitor is at its bound position (R = 3.89 Å), only the configurations for which R was smaller than 4.12 Å were used in the internal solvent site analysis.

Internal solvation pattern calculations were done by first calculating the solvent density distribution on a grid, and then performing a generic solvent site analysis (Mezei and Beveridge, 1984) or using the Lounnas-Pettitt algorithm (Lounnas and Pettitt, 1994), which averages the grid densities recursively. These methods have been successfuly utilized earlier, and the details of these calculations with references may be found in Resat and Mezei (1996). As in the earlier study, these two analysis methods gave almost identical results. Therefore, the results reported below refer to the common outcome of both analysis methods.

The overall agreement of the GCMC simulation results with experiments is exceptional; the molecular simulations predict 20 of the 22 internal waters to within 0.8 Å. Most of the waters are predicted to within 0.4 Å (Table 1). For the waters where there is a disagreement, water 802 is predicted at a distance of 1.43 Å, and water 430 does not exist in our results. Occupancy analysis (Table 1) showed that almost all of the predicted water sites are fully occupied. Even though the overall agreement with experiments is very satisfactory, there are a few interesting discrepancies that deserve discussion. Bode and Schwager report that internal water 416 forms a weak hydrogen bond with nitrogen of benzamidine (Bode and Schwager, 1975). Our analysis predicts that there may be a water at that location, but with about one-fifth occupation probability. This water resides in a side pocket of the binding channel, and its location changes as the inhibitor moves away from its bound position (Fig. 5).

More interestingly, the simulations do not reproduce the internal water 430, which is supposed to form hydrogen bonds with Ile16:N, Gly140:O, and Gly142:O. Note that Ile16 is the N-terminus, which forms an internal salt bridge with Asp194 in trypsin, and that it was modeled in its ionic form in the simulations. Capping hydrogens explicitly included in the simulations were assigned by molecular modeling template matching. Because of the possibly large electrostatic interaction effects, a slight misplacement of the protons might have caused the discrepancy. Although it probably is unrelated, it is still interesting enough that the water 430 was not seen in the bovine trypsinogen-pancreatic trypsin inhibitor structure (Bode et al., 1978) where, because of its turned position, Asp194 does not form a salt bridge with Ile16. Because water 430 is quite far away from the binding site channel (the distance to Asp189:CG is 11.1 Å), the effects of any error in its occupancy are expected to be small.

In addition to the experimentally detected waters as reported by Bode and Schwager (1975), the molecular simulations predicted three more internal waters whose coordinates are reported in Table 2. One of these waters has only a partial occupancy. Such information could be used as an aid in analyzing the experimental data by pointing out not-so-well-resolved solvent molecules.

TABLE 1 Internal waters: comparison with experiments

Water	d(MC-Expt)*	ρ_{occ}#	σ§
406	0.30	1.00	0.37
408	0.25	1.00	0.77
410	0.22	1.00	0.25
412	0.09	1.00	0.23
415	0.14	1.00	0.25
416	0.26	0.22	0.32
429	0.36	1.00	0.46
430	—	N.E.¶	—
457	0.32	1.00	0.49
473	0.15	1.00	0.58
562	0.79	1.00	0.27
604	0.15	1.00	0.30
701	0.61	0.55	0.34
703	0.34	1.00	0.73
704	0.26	1.00	0.27
705	0.11	1.00	0.30
707	0.32	1.00	0.30
710	0.73	1.00	0.68
714	0.64	1.00	0.27
717	0.27	1.00	0.26
722	0.39	1.00	0.24
802	1.43	0.98	0.52

*Distance from MC water to experimental water in angstroms.
#Occupancy probability of a particular site.
§RMS fluctuation of waters in angstroms.
¶No equivalent site was detected in the simulations; see the text.

SUMMARY AND DISCUSSION

We calculated the pmf between the enzyme trypsin and its inhibitor benzamidine along a predetermined curvilinear path by explicit solvent simulations and by continuum solvent calculations based on the Poisson equation. Our aim was 1) to show that the grand canonical ensemble simulations can be helpful in accurately sampling the solvent degrees of freedom, and 2) to investigate whether the continuum solvent calculations can approximately reproduce the explicit solvent simulation results for association thermodynamics. Based on the very good agreement with ex

TABLE 2 Additionally predicted waters

Water	x	y	z	ρ_{occ}*	σ#
A	9.90	3.04	11.60	1.00	0.48
B	9.72	2.34	14.21	1.00	0.66
C	−7.00	13.52	26.77	0.55	0.51

*Occupancy probability of a particular site.
#RMS fluctuation of waters in angstroms.

periments in reproducing the internal water positions and the structural properties around the binding site residue Asp189, our results clearly show that the grand canonical ensemble simulations can be very useful in such systems. It was also found that the Poisson equation calculations can give results that are in reasonable agreement with the explicit solvent simulations. However, the quality of the agreement strongly depends on the choice of empirical parameters. Further studies that investigate a variety of systems and employ a wider range of parameters are certainly needed.

We would like to thank to Chung Wong for his help in the initial stages of the project and Jan Antosiewicz for his help with the pK_a calculations.

Financial support for this research was provided by NSF and the NSF Supercomputer Centers MetaCenter Program. TJM is a National Institutes of Health postdoctoral fellow.

REFERENCES

Adams, D. J. 1975. Grand canonical ensemble (Monte Carlo) for a Lennard-Jones fluid. *Mol. Phys.* 29:307–311.

Antosiewicz, J., J. A. McCammon, and M. K. Gilson. 1994. Prediction of pH-dependent properties of proteins *J. Mol. Biol.* 238:415–436.

Beglov, D., and B. Roux. 1994. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* 100:9050–9063.

Bode, W., and P. Schwager. 1975. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. II. Crystallographic refinement, calcium binding site, benzamidine binding site and active site at pH 7.0. *J. Mol. Biol.* 98:693–717.

Bode, W., P. Schwager, and R. Huber. 1976. Structural studies of the pancreatic trypsin inhibitor-trypsin complex and its free components: structure and function relationship in serine protease inhibition and catalysis. *In* Proteolysis and Physiological Regulation, Miami Winter Symposia Series, Vol. 11. D. W. Ribbons and K. Brew, editors. Academic Press, New York. 43–76.

Bode, W., P. Schwager, and R. Huber. 1978. The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. The refined crystal structures of the bovine trypsinogen-pancreatic trypsin inhibitor and of its ternary complex with Ile-Val at 1.9 Å resolution. *J. Mol. Biol.* 118:99–112.

Connolly, M. L. 1983. Analytical molecular surface calculation. *J. Appl. Crystall.* 16:548–558.

Davis, M. E., J. D. Madura, B. A. Luty, and J. A. McCammon. 1991. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comput. Phys. Commun.* 62:187–197.

Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.

Finer-Moore, J. S., A. A. Kossiakoff, J. H. Hurley, T. Earnest, and R. M. Stroud. 1992. Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. *Proteins Struct. Funct. Genet.* 12:203–222.

Friedman, H. L. 1985. A Course in Statistical Mechanics. Prentice-Hall, Englewood Cliffs, NJ.

Frisch, M. J., G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wongand, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople. 1992. Baussian 92. Gaussian, Inc., Pittsburgh, PA.

Gilson, M. 1995. Theory of electrostatic interactions in macromolecules *Curr. Opin. Struct. Biol.* 5:216–223.

Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

Horvath, D., D. van Belle, G. Lippens, and S. J. Wodak. 1996. Development and parametrization of continuum solvent models. I. Models based on the boundary element method. *J. Chem. Phys.* 104:6679–6695.

Huber, R., D. Kukla, W. Bode, P. Schwager, K. Bartels, J. Deisenhofer, and W. Steigemann. 1974. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. *J. Mol. Biol.* 89:73–101.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.

Lounnas, V., and B. M. Pettitt. 1994. A connected-cluster of hydration around myoglobin: correlation between molecular dynamics simulations and experiment. *Proteins Struct. Funct. Genet.* 18:133–147.

Luty, B. A., Z. R. Wasserman, P. F. W. Stouten, C. N. Hodge, M. Zacharias, and J. A. McCammon. 1995. A molecular mechanics/grid method for evaluation of ligand-receptor interactions. *J. Comp. Chem.* 16:454–464.

Lybrand, T. P. 1995. Ligand-protein docking and rational drug design. *Curr. Opin. Struct. Biol.* 5:224–228.

Madura, J. D., M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon. 1994. Biological applications of electrostatic calculations and brownian dynamics simulations. *Rev. Comp. Chem.* 5:229–267.

Mares-Guia, M., D. L. Nelson, and E. Rogana. 1977. Electronic effects in the interaction of para-substituted benzamidines with trypsin: the involvement of the π-electronic density at the central atom of the substituent in binding. *J. Am. Chem. Soc.* 99:2331–2336.

Marquart, M., J. Walter, J. Deisenhofer, W. Bode, and R. Huber. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. B.* 39:480–490.

Marrone, T. J., J. M. Briggs, and J. A. McCammon. 1997. Structure based drug design: computational advances. *Annu. Rev. Pharmacol. Toxicol.* 37: in presss.

Mezei, M. 1987. Adaptive umbrella sampling: self-consistent determination of the non-Boltzmann bias. *J. Comput. Phys.* 68:237–248.

Mezei, M., and D. L. Beveridge. 1984. Generic solvent sites in a crystal. *J. Comput. Chem.* 5:523–527.

Mezei, M., and D. L. Beveridge. 1986. Free energy simulations. *Ann. N.Y. Acad. Sci.* 482:1–23.

Patey, G. N., and J. P. Valleau. 1975. A Monte Carlo method for obtaining the interionic potential of mean force inionic solution. *J. Chem. Phys.* 63:2334–2339.

Pettitt, B. M., M. Karplus, and P. J. Rossky. 1986. Integral equation model for aqueous solvation of polyatomic solutes: application to the determination of the free energy surface for the internal motion of biomolecules. *J. Phys. Chem.* 90:6335–6345.

Pratt, L. R., G. Hummer, and A. E. Garcia. 1994. Ion pair potentials of mean force in water. *Biophys. Chem.* 51:147–165.

Rashin, A. A. 1989. Electrostatics of ion-ion interactions in solution. *J. Phys. Chem.* 93:4664–4669.

Rashin, A. A., and K. Namboodiri. 1987. A simple method for the calculation of hydration enthalpies of polar molecules with arbitrary shapes. *J. Phys. Chem.* 91:6003–6012.

Resat, H., and M. Mezei. 1994. Grand canonical Monte Carlo simulation of water positions in crystal hydrates. *J. Am. Chem. Soc.* 116:7451–7452.

Resat, H., and M. Mezei. 1996. Grand canonical ensemble Monte Carlo simulation of the dCpG/proflavine crystal hydrate. *Biophys. J.* 71: 1179–1190.

Resat, H., M. Mezei, and J. A. McCammon. 1996. Use of the grand canonical ensemble in potential of mean force calculations. *J. Phys. Chem.* 100:1426–1433.

Simonson, T., and A. T. Brunger. 1994. Solvation free energies estimated from macroscopic continuum theory: an accuracy assessment. *J. Phys. Chem.* 98:4683–4694.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Soman, K., A.-S. Yang, B. Honig, and R. Fletterick. 1989. Electrical potentials in trypsin isozymes. *Biochemistry.* 28:9918–9926.

Straatsma, T. P., and J. A. McCammon. 1992. Computational alchemy. *Annu. Rev. Phys. Chem.* 43:407–35.

Walter, J., W. Steigemann, T. P. Singh, H. Bartunik, W. Bode, and R. Huber. 1982. On the disordered activation domain in trypsinogen. Chemical labelling and low-temperature crystallography. *Acta Crystallogr., Sect. B.* 38:1462.

Wong, C. F., and J. A. McCammon. 1986a. Computer simulation and the design of new biological molecules. *Isr. J. Chem.* 27:211–215.

Wong, C. F., and J. A. McCammon. 1986b. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* 108:3830–3832.