# Calculation of Absolute Binding Free Energies for Charged Ligands and Effects of Long-Range Electrostatic Interactions

JOHAN ÅQVIST

*Department of Molecular Biology, Uppsala University, Biomedical Centre, Box 590, S-75124 Uppsala, Sweden*

*Received 20 October 1995; accepted 2 January 1996*

## ABSTRACT

A recently proposed molecular dynamics method for estimating binding free energies is applied to the complexation of two charged benzamidine inhibitors with trypsin. The difficulties with calculations of absolute binding energies for charged molecules, associated with long-range electrostatic contributions, are discussed and it is shown how to deal with these effectively. In particular, energetic effects caused by the trunction of dipole–dipole interactions in the medium surrounding the charged ligand are examined and found to be significant. Calculations of the absolute binding energy for benzamidine using the free energy perturbation approach are also reported. These simulations illustrate the typical problems associated with annihilation transformations of molecules bound inside proteins. © 1996 by John Wiley & Sons, Inc.

## Introduction

Computer simulation approaches, such as free energy perturbation (FEP), are now becoming widely used for evaluating binding affinities between host molecules and their ligands.[1–4] These types of methods are, however, much more difficult to apply to the problem of calculating *absolute* binding free energies,[5–7] as compared to relative ones. This is basically due to sampling and convergence problems associated with large changes in the configuration of the system that may accompany the binding process. These problems are also amplified if the ligand is charged, due to long-range electrostatic effects that will be discussed. We recently presented a semiempirical method[8] for the estimation of absolute binding free energies

* Author to whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 17, No. 14, 1587–1597 (1996)

CCC 0192-8651 / 96 / 141587-11

that is quite different in that it does not require simulation of any transformation processes. Instead, the binding energy is evaluated solely from the difference between the average ligand–solvent interaction energies in the bound and free states. (The term solvent will be used here for the entire medium surrounding the ligand, which in the bound state thus also includes the host[8].) This method, which we will refer to as the linear interaction energy (LIE) approximation, is based on linear reponse assumptions. It divides the ligand–solvent interaction into polar and nonpolar parts and the free energy of binding is estimated as

$$\Delta G_{\text{bind}} = \Delta G^{\text{el}}_{\text{bind}} + \Delta G^{\text{vdw}}_{\text{bind}} \simeq \frac{1}{2}\Delta\langle V^{\text{el}}_{l-s}\rangle + \alpha\Delta\langle^{\text{vdw}}_{l-s}\rangle \qquad (1)$$

where $\Delta$ denotes the difference between simulations of the bound and free states. The subscripts $l$ and $s$ denote ligand and solvent, respectively, while the superscripts el and vdw refer to the electrostatic and van der Waals (or Lennard–Jones) parts of the potential energy function. The factor of $1/2$ for the electrostatic term is derived from the assumption that the systems exhibit linear response with respect to electrostatic forces, while the coefficient $\alpha$ is an empirical parameter that relates the average van der Waals interaction energies to a corresponding nonpolar (or "hydrophobic") binding contribution.[8]

The assumption that the binding energy can be estimated from only the interactions between the ligand and the surrounding medium is clearly an approximation and may need some further motivation (see ref. 8). Let us therefore first consider the electrostatic contribution to the solvation free energy of a solute in a given medium. We will define this as the free energy associated with taking a system ($A$), corresponding to the solute "cavity" interacting with the medium through a Lennard–Jones (LJ) potential, all solvent–solvent interactions, and any possible intramolecular solute interactions, but without the electrostatic solute–solvent potential turned on and "transforming" it into a system ($B$) described by the full potential of the system. For any given configuration of the particles the potential energy difference between the two states, $B$ and $A$, then simply becomes $\Delta V = V_B - V_A = \Delta V^{\text{el}}_{l-s} = V^{\text{el}(B)}_{l-s}$. As shown in ref. 8, a series expansion of the general expression for the free energy difference between two states, described by potentials $V_B$ and $V_A$ leads to

$$\Delta G = \frac{1}{2}\{\langle\Delta V\rangle_A + \langle\Delta V\rangle_B\} - \frac{\beta}{4}\{\langle(\Delta V - \langle\Delta V\rangle_A)^2\rangle_A - \langle(\Delta V - \langle\Delta V\rangle_B)^2\rangle_B\} + \cdots \qquad (2)$$

where $\langle\ \rangle_A$ and $\langle\ \rangle_B$ denote averages evaluated on the corresponding potential energy surfaces and $\beta^{-1} = kT$. With the above choice of states $\Delta V$ is a purely electrostatic term, and for the type of "charging" process that we are considering it seems possible that one could assume a linear polarization response of the medium. The linear response assumption implies that the free energy functions of the two states are parabolas of equal curvature.[8–11] It is easily shown that this is equivalent to truncating eq. (2) after the second term (harmonic free energy functions) and equivalent to saying that the fluctuations of the energy gap is equal on the two surfaces (equal curvature). Furthermore, if the systems are considered to be completely relaxed, the average of $\Delta V$ evaluated on state $A$, $\langle\Delta V\rangle_A = \langle V^{\text{el}(B)}_{l-s}\rangle_A$, will tend to zero because the solvent in this case does not experience the charge distribution of the solute. We are then left with the term $1/2\langle V^{\text{el}}_{l-s}\rangle_B$, which constitutes the first part of eq. (1). We noted in ref. 8 that this result was obtained in slightly different ways by other authors as well,[12,13] and some test calculations on monopolar and dipolar molecules in water were reported and indicated that the linear response behavior was reasonably well fulfilled.[8] In the present work we also *confirm* the validity of this approximation for solvation of benzamidine (Bza) in water (see below).

The second term of eq. (1) depends on an empirical parameter ($\alpha$) that relates the difference in solute–solvent van der Waals interaction energy between the bound and free states of the ligand to a corresponding nonpolar binding contribution. The rationale for employing this type of approximation was that experimentally one observes a linear dependence of solvation free energy on size parameters, such as surface area, for nonpolar molecules in different solvents[14,15] and that the average solute–solvent (LJ) interaction energy from molecular dynamics (MD) simulations also was found to scale linearly with molecular size for typical nonpolar molecules.[8] These results thus suggest that it may be possible to relate the change in solute–solvent van der Waals energy upon

binding to a free energy contribution by a simple linear scaling. The empirical coefficient $\alpha$ will thus reflect the relationships between the dependencies of average van der Waals energies and nonpolar free energy contributions on molecular size in water and in the protein.[8] One can note that the dependence of $\Delta G$ on only the $V_{l-s}$ terms (and not on $V_{l-l}$ and $V_{s-s}$) is not really an approximation per se, because the underlying solvation energies, in the perturbation formulation, only depend on the remaining potential terms through the Boltzmann factor implicit in the sampling of an MD simulation.[8] The form of the dependence in eq. (1) is, however, a very simplifying approximation that can be justified if the linear reponse assumption holds and if solvation energies and average interaction energies scale linearly also for nonpolar interactions. As argued earlier, there are reasons to believe that these criteria can be fulfilled, at least in some cases. This work presents another test of these hypotheses and also addresses the general problem of how to deal with long-range electrostatics in binding calculations. For readers that might feel uncomfortable with the empirical or approximative nature of the above approach to obtain free energies, we recommend that they simply view it as an empirical "scoring method" (of which there are several other examples)[16-19] based on MD force field calculations. The optimal value of the parameter $\alpha$ found in ref. 8, by calibration on various endothiapepsin-inhibitor complexes, was $\alpha = 0.16$. An interesting question is naturally to what extent this value is system specific, or whether it is generally applicable (it is likely to be force field dependent, which is of minor concern).[8]

The original parametrization of eq. (1) has already been tested on some other systems. Åqvist and Mowbray[20] applied the LIE method to the problem of sugar recognition by a bacterial glucose/galactose receptor.[20] In that case the method was found to be rather successful in describing the energetics of sugar binding. Hansson and Åqvist recently reported calculations of absolute binding constants for some inhibitors to HIV-1 proteinase that also yielded quite reasonable results.[21] These inhibitors were not only challenging from the viewpoint that they are rather large molecules, but two of the compounds (pepstatin and acetylpepstatin) also carry a net negative charge. The LIE approach was also compared to FEP calculations of the relative binding energy between these molecules and gave similar results.[21] Furthermore, the structure of the complex of HIV-1 proteinase and the neutral inhibitor DMP-323 turned out to be very accurately predicted by the simulations (no experimental structure was available at the time these calculations were carried out).[21] More recent work on newly synthesized proteinase inhibitors for which crystal structures are underway also seems very promising (J. Hultén, N. Bonham, B. Classon, H. Danielsson, T. Hansson, A. Karlén, I. Kvarnström, U. Nillroth, B. Samuelsson, J. Åqvist, and A. Hallberg, in preparation).

In the work of ref. 21 the electrostatic problems associated with charged inhibitors were dealt with in a "brute force" type of manner, namely by using very large system sizes and cutoff radii. This, of course, made the calculations quite time consuming. In the present work, we examine a more efficient approach to calculate binding energies for charged molecules that makes use of the so-called local reaction field (LRF) method.[22] We will also demonstrate that not only long-range ligand–solvent interactions are of importance (note our definition of "solvent" above), but also that long-range solvent–solvent interactions have a pronounced effect on the calculated energetics. These results cast some shadows over typical MD simulation protocols that employ nonbonded cutoff radii of around 10 Å, insofar as they are used for studying charged compounds. As a comparison to the LIE results, we also calculate the absolute binding constant for one of the ligands with standard FEP procedures that involve annihilation of the ligand. Such a procedure is intrinsically difficult as indicated above, and this simulation turns out to illustrate the typical sampling difficulties encountered when large molecules are annihilated using FEP methods.

## Long-Range Electrostatic Interactions

In this communication we report calculations of the binding of the inhibitors Bza and *p*-carbethoxy-benzamidine (Bza—$COOC_2H_5$) to $\beta$-trypsin. At normal pH, Bzas are protonated and positively charged and this is also manifested in the crystal structure of the trypsin–Bza complex where an ion pair is formed between the amidinium group of the ligand and the negative side chain of Asp189.[23] The two main reasons for the difficulty of computing absolute binding energies for charged ligands both have to do with the long-range nature of monopolar electric fields. First, the electrostatic solvation energy (or self-energy) of a charged group that is calculated using a finite interaction range (due to a finite, usually spherical, system

and/or to the use of an interaction cutoff radius) will always lack the contribution from the medium outside of this range. This energy contribution is dominated by the Born term, $\Delta G_{Born} = -(Q^2/2r)(1 - 1/\varepsilon)$, that depends on the net charge within the "interaction sphere," its radius, and the dielectric constant of the surrounding. This means that in order for Born terms not to enter into the binding energy calculation, the simulation systems must have the same net charge and size (as well as the same $\varepsilon$ outside). Otherwise, the different free energy contributions from the outside can give large contributions to binding energy, which is usually not desirable because these are given by a continuum approximation that may not always be accurate enough. For instance, if we consider a ligand with net charge $-1$ and a protein with net charge $+5$, there will be a positive Born term of roughly 90 kcal/mol associated with bringing the ligand into the protein if the interaction sphere radius is 16 Å and includes these charges. Thus, in order to be able to compare the results from two simulations of a charged ligand (in solution and bound to the host) for evaluating the free energy of binding, it is necessary to ensure that the two systems have the same charge. How should this then be done? There are basically two options. Either one has to add extra ions to the water system or neutralize some (distant) charges in the protein, by simply turning them off or by adding appropriate counterions. We found earlier that the approach of neutralizing excess protein charges far away from the ligand seems to be preferable because it considerably improves convergence compared to the case with free counterions in the system.[21] If the simulation system is finite (e.g., a sphere) this *must*, in fact, be done for charges near the system boundary that otherwise will not be sufficiently solvated (screened) and therefore excert vacuumlike fields into the system. The error introduced by neglecting distant charges can easily be accounted for by Coulomb's law corrections using a high dielectric constant typical of long-range charge–charge interactions in solvated proteins. Note that the problem discussed above only really pertains to the case when the ligand is charged, because the Born terms are very large. For dipolar ligands, however, the contributions from the corresponding Onsager terms are quite small provided that the simulation systems are sufficiently large.

The second problem with charged ligands is that the use of a solvent–solvent cutoff causes an overpolarization of the solvent toward the solute charge. This effect was discussed in detail earlier[24-26] and is due to the fact that such a cutoff limits the number of solvent–solvent interactions, thereby reducing the resistance to polarization by the solute charge. Furthermore, as water dipoles are more polarizable/reorientable than protein ones, this overpolarization effect will be more pronounced in water than in a protein. The result will therefore be that the solvation energy of the charge is overestimated more in water than in the protein, resulting in an artificial "antibinding" contribution (see below). It is therefore essential to take into account, not only long-range ligand–solvent interactions, but also long-range solvent–solvent interactions. While this can be done by not using any cutoffs at all,[24-25] or very large ones,[21] such solutions are naturally rather expensive. Here, we instead choose to employ the recently published LRF method.[22] The solvent–solvent interactions within a normal cutoff of 10 Å are then calculated in each MD step as usual, while the potential at the centre of each "charge group" (i.e., a group of atoms making up a dipolar or charged entity) due to groups outside of the cutoff is expanded in a third-order Taylor series. This long-range expansion is updated with the same interval as the normal cutoff sphere, but the forces and energies associated with it are calculated at each step for all the particles. The LRF expansion was only used for solvent–solvent interactions while all interactions involving the ligand were explicitly evaluated, i.e., without any cutoff. The method was first tested by FEP calculations of the free energy of charging an $Na^+$ ion in water. In this case we obtained a value of $\Delta G = -99.7 \pm 0.3$ kcal/mol without any cutoffs[25] and $\Delta G = -99.6 \pm 0.2$ kcal/mol using the LRF method, which clearly demonstrates its efficiency. One can note here that using the same 10-Å cutoff without any long-range treatment gives a $-8$ kcal/mol error for $Na^+$ hydration due to the overpolarization effect.[25] This same effect will also be demonstrated in the binding calculations below. It should perhaps also be emphasized that both of the difficulties with charged ligands discussed above are quite general and not dependent on the method used for free energy calculation.

## Computational Procedures

The simulations were carried out starting from the crystal structure (3PTB) of the trypsin–Bza complex.[23] The initial structure of the Bza—

$COOC_2H_5$ complex was modeled from the same structure without any manipulation of the protein coordinates. In both the simulations of the complexes and the ligands in water, spherical systems of radius 16 Å with restrained boundaries were used. Water was represented by the rigid SPC model[27] with the surface molecules (of the sphere) subjected to radial and polarization restraints according to the Surface Constrained All-Atom Solvent (SCAAS) model.[28] Protein atoms within the outermost 2-Å layer of the 16-Å sphere were restrained to their crystallographic positions with harmonic force constants of 20 kcal/(mol Å$^2$). All protein net charges within 12 Å of the amindinium carbon atom were included while those beyond this distance were kept neutral. With this choice the protein is overall neutral and the problem with Born terms discussed above can be avoided, because both the protein and water systems then carry a net charge of +1. A continuum correction for the effect of neutralizing these charges is also evaluated (see below). Furthermore, as discussed above it is important *not* to include fully charged groups too near the simulation boundary in a finite system because there will not be enough "matter" around the charge for a realistic dielectric screening of it. The protein simulations with Bza included 206 water molecules of which 26 were crystallographic ones, and the water simulations included 573 waters (two water molecules less were used in the Bza—$COOC_2H_5$ simulations). The simulations were run at $T = 300$ K using an MD time step of 0.002 ps with a version[24,29] of the ENZYMIX program[30] that uses the Groningen molecular simulation (GROMOS) extended atom force field[31] with modifications that have been reported elsewhere.[8,29] (These pertain to the LJ parameters of water–aliphatic carbon interactions as well as those of charged carboxylate ions, which were calibrated to reproduce hydration free energies.) The force field parameters used for the Bza moiety were the standard GROMOS LJ parameters for the relevant atom types plus partial atomic charges derived from AM1/SM2 calculations using the program AMSOL.[32] These charges were found to be rather compatible with the regular values used in GROMOS[31] and only minor modifications were needed to merge them with the force field. (The charges used for Bza were: uncharged benzene ring and +0.36, −0.28, and +0.30 charge units on the C, N, and H atoms of the amidinium group, respectively.) The simulated trajectories each lasted 280 ps, of which 30–80 ps were discarded for equilibration. The only exception was the water simulation of Bza—$COOC_2H_5$ that required 150 ps for equilibration, presumably due to the initial conformation as well as the flexibility of its tail. This trajectory was therefore extended to 350 ps to verify that the calculated energy averages were stable.

For comparison with the LIE results, an attempt was also made to evaluate the absolute free energy of binding for Bza using FEP simulations. These involved the same amount of computer time as used for the normal MD trajectories with Bza. As noted above, calculations of this type where the ligand is, in principle, annihilated in its protein site are usually problematic when one is dealing with large inhibitors. One of the basic problems here is how to allow solvent to enter the binding site during the annihilation process. This is a well-known problem that mainly arises from the fact that as long as there are any remainders of the ligand present (that have LJ bumps or spikes associated with them), these can prevent solvent from "rushing" into the protein cavity (a kind of "pinball" effect is you wish). Thus, not until the very last perturbation step, when these bumps or spikes really vanish, does the cavity become fully accessible to solvent molecules. Another problem is that parts of the protein may have to move away in order to open a channel for solvent to enter. It should also be noted here that it does not help much if one, instead of using a normal LJ potential, designs a repulsive term that does not have any singularity, because it is rather the existence of repulsive centers than their singularities that is the problem. Here, we chose to try to circumvent this dilemma by transforming Bza (in solution and in the protein) into four water molecules rather than to completely annihilate it. The FEP procedure is described in more detail in Figure 1.

## Results and Discussion

To warrant a detailed discussion of binding energetics, it is essential to first examine the structural agreement between the simulations and experiment. Figure 2 shows the average MD structure, calculated over the last 50 ps, of the trypsin active site with bound Bza superimposed on the crystal conformation. It can be seen that the position of the Bza inhibitor is virtually identical to the X-ray result and that the deviations of the protein structure are small. The root mean square (rms) coordinate shift for all protein and inhibitor atoms within a 12-Å sphere centered on the branching

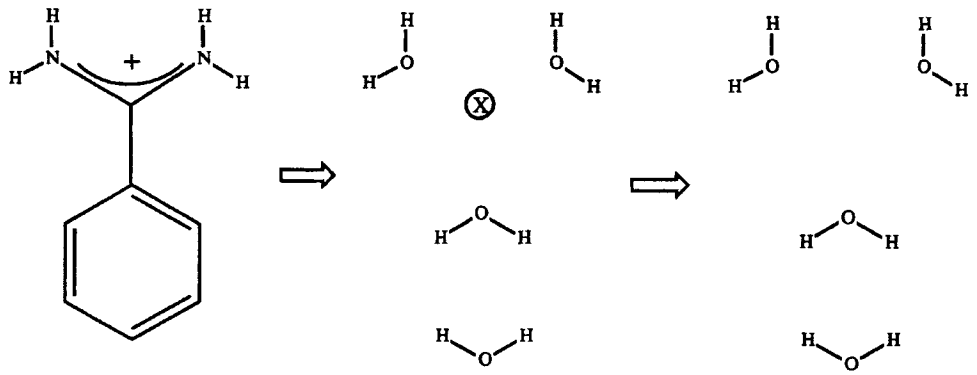

**FIGURE 1.** FEP scheme used to evaluate the binding energy of Bza. The ligand was perturbed in two steps into four water molecules. (In water as well as inside the protein, identical procedures were used in the calculations of the bound and unbound ligand.) The first step involves annihilating the net positive charge on the amidinium group and making two water molecules out of the two $NH_2$ groups and two other ones out of the benzene ring. These waters have their final charges but their geometries (bonds and angles) are kept fixed by the SHAKE algorithm[40] at their original values. The other bonds, angles, and torsions are scaled down with the FEP coupling parameter $\lambda$ so that they disappear in this process. The amidinium carbon atom is in this step perturbed to a state with very small LJ parameters $[(A_i, B_i) = (2.5, 0.05)$; $V_{LJ} = A_iA_j/r^{12} - B_iB_j/r^6$ measured in kcal/mol and Å]. In the second step, this atom is completely annihilated and the water geometries are restored to their normal values (again using SHAKE). The same number of MD integration steps as in the LIE simulations of Bza (14,000) was used for the entire perturbation process, distributed as follows: 14% for equilibration, 57% for step 1, and 29% for step 2. (A smaller time step of $\Delta t = 1$ fs had to be used in these calculations to avoid instabilities.) The number of $\lambda$ values used was 40 for step 1 and 20 for step 2 and at each $\lambda$ point the first 20% of the data was discarded. The errors associated with the FEP calculations are difficult to determine, because reverse trajectories are not possible to calculate, so they were in this case simply estimated from forward and backward FEP summation on the same trajectories.

carbon of the benzene ring is 0.81 Å. The corresponding value for an 8-Å sphere is 0.83 Å. Thus, we can conclude that the simulations, in spite of their relatively long duration, preserved the experimental conformation very well and that we therefore are sampling configurations that should indeed be relevant to the experimental binding measurements.

The average ligand–solvent interaction energies for the water and protein simulations with Bza are given in Table I (first entry) together with the corresponding LIE estimate for the absolute bind-

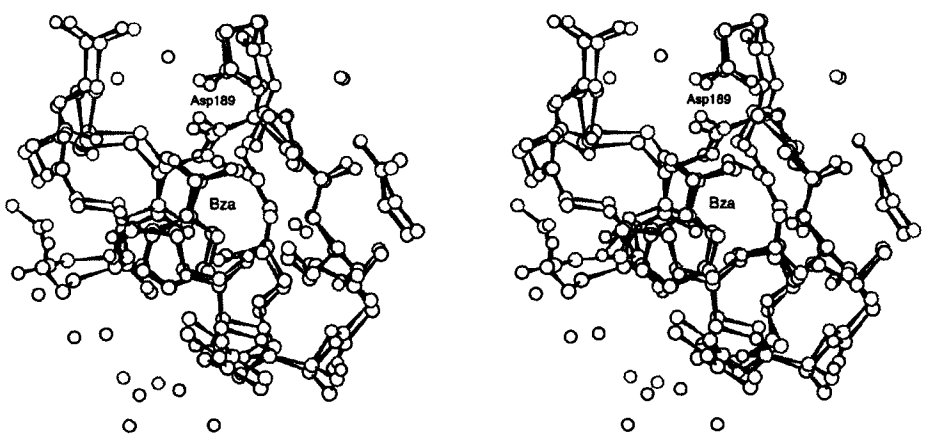


**FIGURE 2.** Stereo view of the average MD structure calculated from the last 50 ps of the trajectory (black) superimposed on the crystal structure (gray).[23] Water molecules are indicated by circles.

ing free energy. The convergence errors estimated by averaging over the first and last 100-ps segments of the trajectories are given in the table and these clearly indicate that the averages are stable. This can also be seen from Figure 3 that shows the behavior of 25-ps subaverages of the ligand–solvent interaction energies $V_{el}^{p}$, $V_{vdw}^{p}$, $V_{el}^{w}$, and $V_{vdw}^{w}$, during the 200-ps trajectory used for data collection (i.e., after the equilibration period). The relative rms fluctuations of these subaverages are 0.7, 2.2, 1.2, and 4.0%, respectively, which shows that the trajectories are well converged.

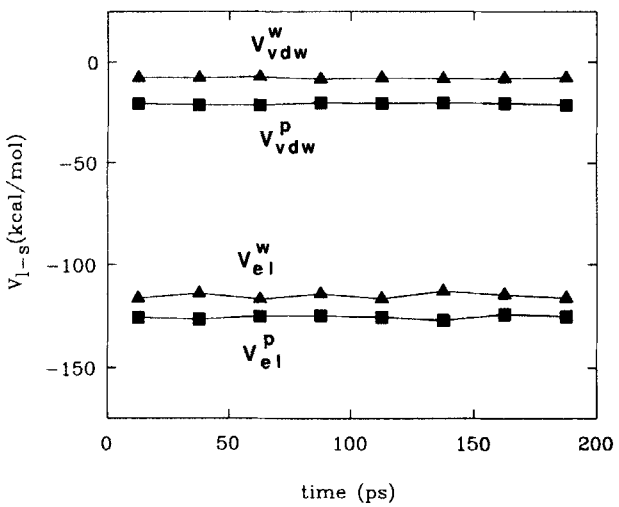


**FIGURE 3.** The variation of 25-ps subaverages of the ligand–solvent interaction energy components versus time, along the 200-ps trajectory (after equilibration) used to obtain the first entry of Table I. The two upper curves denote the Lennard–Jones interactions while the lower ones denote the electrostatic interactions (protein simulation, squares; water simulation, triangles).

We can see from Table I (first entry) that both the electrostatic and van der Waals interactions are more favorable in the protein than in water and their respective contributions to $\Delta G_{bind}^{calc}$ according to eq. (1) are −5.1 and −2.1 kcal/mol, yielding a total free energy of binding of −7.2 ± 0.1 kcal/mol. This value is in fairly good agreement with the experimental result of $\Delta G_{bind}^{obs} = -6.4$ kcal/mol.[33] Because we did not include any charged protein groups beyond 12 Å from the inhibitor, one should now ask how large the contribution from these groups can be. Such a correction can be estimated rather simply by Coulomb's law calculations using either a uniform dielectric constant or a distance dependent one as proposed by Mehler and Eichele,[34] or by continuum calculations of the Poisson–Boltzmann type.[35,36] If one uses an effective uniform dielectric constant of 80, which would be a typical value for long-range charge–charge interactions in a solvated system where most of the charges are exposed to water, this correction amounts to +1.0 kcal/mol. The Mehler and Eichele dielectric function yields a very similar value of +1.1 kcal/mol. This correction, which reflects the fact that the net charge of the protein is positive, would thus bring the total binding energy down to $\Delta G_{bind}^{corr} = -6.2$ kcal/mol.

The above results thus indicate that the LIE approximation works quite well for the trypsin–Bza case. It is fairly straightforward to examine the electrostatic linear response assumption for Bza in water, because FEP calculations of charging the

**TABLE I.**
**Average MD ligand–solvent interaction energies and calculated and observed[33] free energies of binding.**

| Inhibitor | Method | $\langle V_{el}^{p} \rangle$ | $\langle V_{vdw}^{p} \rangle$ | $\langle V_{el}^{w} \rangle$ | $\langle V_{vdw}^{w} \rangle$ | $\Delta G_{bind}^{calc}$ | $\Delta G_{bind}^{corr}$ |
|---|---|---|---|---|---|---|---|
| Bza | LRF/10 Å | −125.38 ± 0.02 | −20.87 ± 0.13 | −115.18 ± 0.12 | −7.94 ± 0.06 | −7.2 ± 0.1 | −6.2 |
| Bza | LRF/7 Å | −126.03 ± 1.46 | −20.64 ± 0.73 | −115.98 ± 0.05 | −8.01 ± 0.10 | −7.1 ± 0.9 | −6.1 |
| Bza | $R_c$ = 10 Å | −136.39 ± 1.00 | −21.02 ± 0.37 | −138.75 ± 0.36 | −7.52 ± 0.09 | −1.0 ± 0.8 | 0.0 |
| Bza—$COOC_2H_5$ | LRF/10 Å | −133.62 ± 1.02 | −29.46 ± 0.08 | −125.67 ± 0.39 | −15.86 ± 0.01 | −6.2 ± 0.7 | −5.2 |
| Bza | FEP | | | | | −17.7 ± 3 | −16.7 |
| Bza | Exptl | | | | | | −6.4 |
| Bza—$COOC_2H_5$ | Exptl | | | | | | −5.0 |

Energies are in kilocalories/mole. The superscripts *p* and *w* denote simulations with the benzamidine inhibitors bound to the protein site and free in solution respectively. The error bars are estimated from averaging of the first and second halves of the MD trajectories. The last column gives the estimated binding free energy corrected for the long-range Coulomb interactions between the ligand and protein charges not included in the simulations. The first and fourth entries as well as the FEP calculations were carried out with the LRF[22] procedure using a 10-Å solvent–solvent cutoff; the second entry denotes LRF with a 7-Å cutoff. The third entry corresponds to using a simple 10-Å cutoff for solvent–solvent interactions without any long-range correction for these. No cutoffs were used for ligand–solvent interactions in any of the simulations.

inhibitor in solution can easily be carried out. This was done as outlined above by gradually turning off $V_{l-s}^{el}$, i.e., the solute–solvent electrostatic interaction, during an additional 70-ps FEP simulation. The free energy obtained from this check was $\Delta G = +58.14$ kcal/mol excluding the Born correction (the error estimated from forward and backward integration of the same trajectory is $\pm 0.1$ kcal/mol). Dividing this $\Delta G$ value by $\langle V_{el}^{w} \rangle$ in Table I gives a ratio of 0.50, in agreement with the linear response expectation. It is more difficult to carry out such a calculation accurately for the protein system because large conformational changes associated with charging/discharging of the inhibitor might be expected. (This essentially involves the same problems as when calculating the absolute binding energy with FEP.) Therefore, the validity of electrostatic linear response in the enzyme in this case can only be inferred or hypothesized from the good agreement of the estimated binding energy plus the demonstration of its validity for solvation of Bza in water.

To examine the stability and limits of the LRF procedure, we also repeated the protein and water simulations for Bza using a smaller value of the solvent–solvent cutoff, $R_c = 7$ Å, which considerably increases the speed of the calculations. The results of these simulations are given as the second entry in Table I where it can be seen that the value of $\Delta G_{bind}^{calc} = -7.1 \pm 0.9$ kcal/mol does not change appreciably. However, the convergence becomes considerably worse for the protein simulation. This probably reflects the fact that there are several mobile net charges present in the system that can cause the long-range solvent–solvent potential (now beyond 7 Å) to fluctuate more than is compatible with the expansion updating interval (every 50 MD steps). Thus, the use of a smaller cutoff radius requires a more frequent updating of the expansion to maintain a given level of accuracy, which then again will increase the computational cost. However, we can say that within the error estimates of Table I the LRF method appears to be quite stable, although very small cutoffs are probably not advisable to use. It should be noted that for the test case of ion solvation in water, it has been shown here (see above) and elsewhere[22] that the LRF procedure reproduces the result with no cutoff at all, which is an important finding in that it shows that the method really provides a cheaper alternative to not using any cutoffs.

To clearly demonstrate the severe effect of neglecting long-range solvent–solvent interactions, we also carried out the protein and water simulations using a simple 10-Å cutoff (without LRF) for all interactions except those involving the ligand (which were treated as before). The corresponding results are given as the third entry of Table I. Here, it can be seen that the polarization of the medium toward the ligand now becomes drastically overestimated in both the protein and water. The average electrostatic ligand–solvent interaction energies become about 10 and 20 kcal/mol more negative in the enzyme and in water, respectively. This effect was also observed earlier for ionic hydration.[25] The larger overpolarization observed for the water system probably reflects the fact that the majority of dipolar groups are more rigid in the protein and the fact that there are several charged groups present in the latter case, toward which the overpolarization will become distributed. Due to this antibinding effect, the resulting uncorrected value of $\Delta G_{bind}^{calc}$ rises to $-1$ kcal/mol. It should again be stressed here that the cutoff problem for charged systems is quite general and does not depend on the method used for free energy calculations. The same also holds for the problems with possible Born terms discussed above. We argued earlier[24, 25] that the often used implementation of periodic boundary condition (PBC) simulations for liquid systems, with a simple cutoff of say 10 Å and without Ewald or reaction field treatments, is not really adequate for treating problems of the kind addressed here that involve charged solutes. With spherical models of the present type it appears, however, that the problems can be solved quite easily without a too severe computational cost. For uncharged ligands, on the other hand, the influence of a smaller solvent–solvent cutoff does not appear to be very pronounced. In the cases of the neutral HIV-1 proteinase inhibitor studied in ref. 21 and for glucose binding to the glucose/galactose receptor,[20] it was found that a 10-Å solvent–solvent cutoff yielded very similar energetics compared to simulations with $R_c = 15$ Å or the LRF procedure. In this context, it should also be mentioned that a recent study by Wood[37] indeed confirms our observation that a short solvent–solvent cutoff gives rise to an overpolarization effect. By introducing a continuum correction for the use of such a cutoff (not to be confused with the Born continuum term that is used to correct for the use of a solute–solvent cutoff),[24] it was shown in ref. 37 that the PBC simulations of Straatsma and Berendsen[26] as well as ours[25] yielded internally consistent results for absolute hydration energies of ions. However, while such a correction may be able to take care of

the *energetic* error associated with the cutoff-induced overpolarization, it does not remedy the fact that structure, e.g., the conformation of charged protein side chains, is affected by the use of short cutoffs.[21] Furthermore, when absolute solvation energies are evaluated from PBC simulations in combination with a Born continuum term, the results are found to differ from those obtained with a finite size system such as SCAAS with a similar Born correction.[25] This discrepancy does not, as hypothesized in ref. 37, arise from any extra surface polarization of the SCAAS model (J. Åqvist, in preparation), but rather from polarization outside the ion cutoff sphere in the PBC model. This polarization contributes to the calculated free energy by its presence in the Boltzmann factor that dictates thermal sampling and is thus to some extent counted twice: here and in the Born continuum correction (J. Åqvist, in preparation, and ref. 25). For the type of problem addressed in the present work, however, the aforementioned Born correction controversy does not appear because we only consider *differences* in solvation energies where Born terms cancel (see above).

For comparison, we also made an attempt to evaluate the absolute binding free energy of Bza using the FEP procedure described in Figure 1. The results from these calculations were not too encouraging, yielding a binding energy (including the long-range Coulomb correction) of $-16.7 \pm 3$ kcal/mol (cf. Table I). Although it might be possible to devise more efficient perturbation protocols for this type of calculation, we believe that this result is rather illustrative in that it highlights the main problem with calculations of absolute binding energies using FEP. That is, the free energy cost of annihilating the inhibitor (or, in our case, "reducing the volume" of it to that of four water molecules) inside the protein will always be overestimated if the generated cavity cannot be properly filled with solvent that has to come in from the outside. This will consequently lead to an *overestimation* of binding strength. In our case, penetration of solvent into the binding cavity turned out to be hampered by the motion of the Gln192 side chain that, in response to the perturbation process, partly shuts off the entrance into the active site. The use of restraints on the structure might in some cases offer a solution to this type of problem,[7] but such an approach is usually not desirable as it is somewhat artificial. Another strategy that has been proposed recently[38] is to use a fourth spatial dimension in the calculations that would allow solvent to enter cavities without having to displace intervening protein groups. Whether this method really works in practice for binding calculations, however, remains to be seen. It should be noted here that the above type of problems are rather associated with the artificial *transformation process* used in FEP than with the nature of the bound and unbound states themselves. That is to say that the main difficulty with annihilation is not in modeling the ligand-free state (apoprotein) itself, which can usually be done, but rather in ensuring that the transformation process really ends up in a reasonable model for the ligand-free state.

The results from the calculations on Bza—$COOC_2H_5$ are also summarized in Table I. There it can be seen that electrostatic and van der Waals energies, in the enzyme as well as in water, are more negative than for Bza. This simply reflects the fact that Bza—$COOC_2H_5$ has additional dipolar groups and LJ centers, i.e., more interactions in the protein and in water, compared to Bza. The estimated binding energy for Bza—$COOC_2H_5$ is $-5.2 \pm 0.7$ kcal/mol including the long-range Coulomb correction ($-6.2$ kcal/mol without this correction), indicating a slightly weaker binding than Bza. Also this value is in good agreement with the experimental result of $-5.0$ kcal/mol.[33] Due to slower equilibration, the trajectory for the water system was in this case extended to obtain roughly the same averaging length as for the other entries in Table I. (Averaging over only the last part of the initial 280-ps water trajectory, after 150-ps equilibration, yielded an uncorrected binding energy value of $-5.8 \pm 0.6$ kcal/mol.) As reported by Mares-Guia et al.,[33] the slightly weaker binding of Bza—$COOC_2H_5$ appears to be due to unfavorable dipole–dipole interactions between the enzyme and the tail of the inhibitor or, at least, less favorable interactions than the tail dipole(s) have in water. In comparison to Bza, for which the electrostatic and van der Waals contributions to $\Delta G^{corr}_{bind}$ are $-4.1$ and $-2.1$ kcal/mol, respectively, we find that the corresponding values for Bza—$COOC_2H_5$ are $-3.0$ and $-2.2$ kcal/mol. In other words, the difference in binding appears to be due to a 1 kcal/mol loss in the electrostatic contribution, which is caused by the tail dipole(s).

The present study demonstrated that the LIE approximation gives very reasonable estimates for the binding free energy of two charged trypsin inhibitors. We made no attempt here to refine the value of the empirical nonpolar (or hydrophobic) scaling parameter in eq. (1) and more simulation data is clearly needed to be able to judge its

generality. As noted earlier,[20] one may of course also consider the possibility of treating the electrostatic coefficient empirically as well, in cases where the linear response assumption does not hold. Another possibility, in case deviations from this assumption are found, is to try to actually evaluate them for different chemical groups by FEP simulations (J. Åqvist and T. Hansson, in preparation). We did, in fact, examine the accuracy of the electrostatic factor here for the water system by FEP calculations of charging the Bza inhibitor and this yielded the result expected from linear response. In a recent article,[39] Carlson and Jorgensen examined how well the *free energy of hydration* for some neutral molecules is described by different linear response equations, such as eq. (1) where both electrostatic and van der Waals coefficients are treated empirically and with an additional term proportional to the volume or surface area. Their results suggest that the electrostatic factor for neutral dipolar molecules in water is somewhat lower than 1/2, in agreement with our findings in ref. 8 for methanol and some simple dipoles. The inclusion of an additional volume or surface term in eq. (1) also seems to offer an improvement for $\Delta G_{hyd}$ estimates.[39] As discussed in ref. 8, because we calculate *binding energies* that are basically differences between two solvation free energies, the inclusion of such additional terms may not be as crucial as it seems to be for estimating an absolute solvation energy like $\Delta G_{hyd}$.[39] This could, e.g., be due to similar ligand volumes/surfaces or similar costs of "cavity creation" in the bound and unbound states. It should perhaps also be emphasized that the van der Waals coefficient should not be equal for binding and solvation energy calculations, because in the former case it reflects both protein and solution properties.[8]

In summary, we feel so far that the examples of endothiapepsin inhibitor complexes,[8] sugar-receptor interactions,[20] HIV-1 proteinase inhibitors,[21] and the binding of charged trypsin inhibitors reported here, lend support to the use of linear response methods for binding calculations. The simple LIE approximation of eq. (1) appears to give reasonable results, although the entire range of applicability of the method remains to be determined. We should therefore perhaps emphasize that the present study does not claim to prove that the approximation always holds, but rather that it seems to work well for the specific, albeit rather pertinent, test cases presented here. It thus seems worthwhile to pursue these ideas and explore various types of refinement along the lines discussed above. It should also be evident from the present work that great care has to be taken in binding calculations dealing with charged ligands, irrespective of the method used, but that the problems can be dealt with effectively. As far as actual drug design applications are concerned, where putative ligands as a rule have to be modeled (docked) into their binding sites, the main obstacle may rather be to find the "correct" conformations (both in the host complex and in water) than to score their binding strength. Therefore, one would, of course, ultimately like to combine an efficient conformational search algorithm with an accurate binding energy estimation method. A problem with our present approach is that it is rather time consuming. This is, however, due to features that we also feel are its major strengths, namely, the use of a microscopic force field that allows geometry fluctuations and thermal sampling, and the explicit inclusion of water molecules. It is difficult to say at present what the most effective computational strategy to structure-based drug design in the future will look like. Several promising approaches are currently under development and the main factors determining their success will be the energetic and structural accuracy together with computational cost.

## Acknowledgments

This work was supported by a grant from the Swedish Natural Science Research Council (NFR).

## References

1. D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Biophys. Chem.*, **18**, 431 (1989).
2. W. L. Jorgensen, *Acc. Chem. Res.*, **22**, 184 (1989).
3. P. A. Kollman and K. M. Merz, Jr., *Acc. Chem. Res.*, **23**, 246 (1990).
4. T. P. Straatsma and J. A. McCammon, *Annu. Rev. Phys. Chem.*, **43**, 407 (1992).
5. K. M. Merz, Jr., *J. Am. Chem. Soc.*, **113**, 406 (1991).
6. S. Miyamoto and P. A. Kollman, *J. Am. Chem. Soc.*, **114**, 3668 (1992).
7. S. Miyamoto and P. A. Kollman, *Proteins*, **16**, 226 (1993).
8. J. Åqvist, C. Medina, and J.-E. Samuelsson, *Protein Eng.*, **7**, 385 (1994).
9. R. A. Marcus, *Annu. Rev. Phys. Chem.*, **15**, 155 (1964).
10. E. A. Carter and J. T. Hynes, *J. Phys. Chem.*, **93**, 2184 (1989).
11. G. King and A. Warshel, *J. Chem. Phys.*, **93**, 8682 (1990).
12. A. Warshel and S. T. Russell, *Q. Rev. Biophys.*, **17**, 283 (1984).

13. B. Roux, H.-A. Yu, and M. Karplus, *J. Phys. Chem.*, **94**, 4683 (1990).
14. A. Ben-Naim and Y. Marcus, *J. Chem. Phys.*, **81**, 2016 (1984).
15. K. A. Sharp, A. Nicholls, R. F. Fine, and B. Honig, *Science*, **252**, 106 (1991).
16. N. Tomioka, A. Itai, and Y. Iitaka, *J. Comput. Aided Mol. Design*, **1**, 197 (1987).
17. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comput. Chem.*, **13**, 505 (1992).
18. S. H. Rotstein and M. A. Murcko, *J. Med. Chem.*, **36**, 1700 (1993).
19. H.-J. Bohm, *J. Comput. Aided Mol. Design*, **8**, 243 (1994).
20. J. Åqvist and S. L. Mowbray, *J. Biol. Chem.*, **270**, 9978 (1995).
21. T. Hansson and J. Åqvist, *Protein Eng.*, **8**, 1137 (1995).
22. F. S. Lee and A. Warshel, *J. Chem. Phys.*, **97**, 3100 (1992).
23. M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber, *Acta Crystallogr., Sect. B*, **39**, 480 (1983)
24. J. Åqvist, *J. Phys. Chem.*, **94**, 8021 (1990).
25. J. Åqvist, *J. Phys. Chem.*, **98**, 8253 (1994).
26. T. P. Straatsma and H. J. C. Berendsen, *J. Chem. Phys.*, **89**, 5876 (1988).
27. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, In *Intermolecular Forces*, B. Pullman, Ed., Reidel, Dordrecht, The Netherlands, 1981, p. 331.
28. G. King and A. Warshel, *J. Chem. Phys.*, **91**, 3647 (1989).
29. J. Åqvist, M. Fothergill, and A. Warshel, *J. Am. Chem. Soc.*, **115**, 631 (1993).
30. A. Warshel and S. Creighton, In *Computer Simulation of Biomolecular Systems*, W. F. van Gunsteren and P. K. Weiner, Eds., ESCOM, Leiden, The Netherlands, 1989, p. 120.
31. W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos BV, Groningen, The Netherlands, 1987.
32. C. J. Cramer and D. G. Truhlar, *Science*, **256**, 213 (1992).
33. M. Mares-Guia, D. L. Nelson, and E. Rogana, *J. Am. Chem. Soc.*, **99**, 2331 (1977).
34. E. L. Mehler and G. Eichele, *Biochemistry*, **23**, 3887 (1984).
35. J. Warwicker and H. C. Watson, *J. Mol. Biol.*, **157**, 671 (1982).
36. M. K. Gilson and B. H. Honig, *Nature*, **330**, 84 (1987).
37. R. H. Wood, *J. Chem. Phys.*, **103**, 6177 (1995).
38. T. C. Beutler and W. F. van Gunsteren, *J. Chem. Phys.*, **101**, 1417 (1994).
39. H. A. Carlson and W. L. Jorgensen, *J. Phys. Chem.*, **99**, 10667 (1995).
40. J. P. Ryckaert, G. Cicotti, and H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 327 (1977).