Journal of Computer-Aided Molecular Design, **12:** 215–227, 1998.
KLUWER/ESCOM

The application of three approximate free energy calculations methods to structure based ligand design: Trypsin and its complex with inhibitors

Randall J. Radmer & Peter A. Kollman*
Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, U.S.A.

Received 27 May 1997; Accepted 29 July 1997

Key words: free energy calculations, ligand design, molecular dynamics, Pictorial Representation of Free Energy Changes (PROFEC)

Summary

Three approximate free energy calculation methods are examined and applied to an example ligand design problem. The first of the methods uses a single simulation to estimate the relative binding free energies for related ligands that are not simulated. The second method is similar, except that it uses only first derivatives of free energy with respect to atomic parameters (most often charge, van der Waals equilibrium distance, and van der Waals well depth) to calculate free energy differences. The last method PROFEC (Pictorial Representation of Free Energy Components), generates contour maps that show how binding free energy changes when additional particles are added near the ligand. These three methods are applied to a benzamidine/trypsin complex. They each reproduce the general trends in the binding free energies, indicating that they might be useful for suggesting how ligands could be modified to improve binding and, consequently, useful in structure-based drug design.

Introduction

Molecular dynamics (MD)- and Monte Carlo (MC)-based free energy calculations have been used to determine relative binding free energies for many molecular systems (see, for example, Refs. 1–5). One of the major advantages of these methods is that they can be used to calculate binding free energies for typical ligand/macromolecule systems, making them potentially useful for designing drugs that interact with biological targets such as enzymes and nucleic acids. The major drawback of these methods is the amount of computer time required for reasonably accurate estimates [6, 7]. Although available computer resources are constantly increasing, this is likely to remain a severe limitation of free energy calculations for some time to come.

An alternative approach is to use less accurate but faster methods which would allow the binding free energies of more ligands to be examined. We expect that this would still be useful for identifying good ligands, even though the decrease in accuracy would likely result in some errors (disregarding some useful ligands and excessive interest in some poor ligands). In a previous paper [7], we examined the possibility of using a single simulation to calculate relative free energies of solvation for many related solutes, and found that reasonable estimates could be obtained. In this paper, we will examine this method in greater detail and evaluate it for a benzamidine/trypsin system.

The other two methods examined in this paper give more general indications of how ligand modifications change the binding free energy, and should be useful for suggesting specific changes that improve binding. The first of these uses derivatives of binding free energy with respect to atomic parameters. The second is a new method that gives a pictorial representation of how ligand changes affect binding free energy. Because these methods are based on MD and MC simulations, they take into account binding site flexibility and ligand solvation, which are not considered by many structure-based design methods (for example, the ligand docking program DOCK [8]) but may be critical to assess tight binding ligands.

* To whom correspondence should be addressed.

Theoretical background

This section gives the theory behind the three methods examined. In each, changes in binding free energy resulting from changes in the ligand are estimated based only on trajectories generated using MD or MC simulations of a reference ligand; no modified ligands are simulated. Although we will not examine the possibility of modifying receptors in this paper, it is worth noting that these methodologies could also be used to optimize a receptor to bind a specific ligand.

Simulations of the ligand/receptor complex are performed and their trajectories saved, and, independently, simulations of the solvated ligand are performed and their trajectories saved (because we are generally interested in the free energy of the bound ligand/receptor system relative to solvated ligand and solvated receptor). Using a thermodynamic cycle (see Ref. 9) it can be shown that the change in the free energy of binding resulting from a change in the ligand is given by

$$\Delta\Delta G_{Binding} = \Delta G_{Bound} - \Delta G_{Solvated} \tag{1}$$

where ΔG_{Bound} and $\Delta G_{Solvated}$ are the change in free energy resulting from a change in ligand, using the bound ligand/receptor simulation and the solvated ligand simulation, respectively. The calculation of ΔG_{Bound} and $\Delta G_{Solvated}$ is dependent on the method used, and will be described below. The kinetic (ideal gas) component of the free energy will not be included in any of the calculations because we are assuming it can be calculated analytically [10] or that it can be made to cancel if thermodynamic cycles are used [9].

Single simulation, multiple perturbation (SSMP)

Traditional free energy perturbation methods [11] can be used to calculate free energy differences for many ligand modifications based only on simulations of the reference ligand and the following equation:

$$\Delta G_{k0} = -RT \ln\langle e^{-\Delta\nu_{k0}/RT}\rangle_0 \tag{2}$$

where $\Delta G_{k0} = G_k - G_0$ is the free energy difference between state k (ligand k) and state 0 (the reference ligand), $\Delta\nu_{k0} = \nu_k - \nu_0$ is the potential energy difference, R is the gas constant, T is the temperature of the surroundings, and $< \ldots >_0$ indicates an ensemble average at state 0.

In practice, the ensemble average is estimated using coordinates ($\mathbf{r}_i$) from the saved trajectories to calculate ν_0 and ν_k (after some modification to the coordinates, if necessary). The free energy dilterence is given by the following equation:

$$\Delta G_{k0} \approx -RT \ln \frac{1}{n_0} \sum_{i=1}^{n_0} e^{-\Delta\nu_{k_0}(\mathbf{r}_i)/RT} \tag{3}$$

It is worth noting that the only approximation in this method is the use of a finite number of samples (from a simulation of finite length) to estimate the ensemble average. For a detailed discussion of the limitations of this method, see Ref. 7.

Free energy estimates using free energy derivatives

The change in free energy as a function of a single parameter (α) can be given by finding its Taylor series expansion:

$$\begin{aligned}\Delta G(\alpha_0, \Delta\alpha) &= G(\alpha) - G(\alpha_0) \\ &= \sum_{i=1}^{\infty} \frac{1}{i!} \frac{d^i G(\alpha_0)}{d\alpha_0^i} (\Delta\alpha)^i\end{aligned} \tag{4}$$

where α_0 is the reference value, α is the perturbed value, and $\Delta\alpha = \alpha - \alpha_0$. For small values of $\Delta\alpha$, this can be truncated to a single term:

$$\Delta G(\alpha_0, \Delta\alpha) \approx \frac{dG(\alpha_0)}{d\alpha_0} \Delta\alpha \tag{5}$$

Equations 4 and 5 can be generalized to include many parameters ($\vec{\alpha}$). For simplicity, we show only the generalized form of Equation 5:

$$\Delta G(\alpha_0, \Delta\alpha) \approx \nabla_{\vec{\alpha_0}} G(\vec{\alpha_0}) \cdot \vec{\Delta\alpha} \tag{6}$$

where $\vec{\alpha}$ and $\vec{\alpha_0}$ are the sets of all adjustable parameters, $\vec{\Delta\alpha} = \vec{\alpha} - \vec{\alpha_0}$, and $\nabla_{\vec{\alpha_0}} G(\vec{\alpha_0})$ is the gradient with respect to $\vec{\alpha_0}$.

In Equation 6, the gradient of the free energy is given by (see, for example, Ref. 3)

$$\nabla_{\vec{\alpha_0}} G(\vec{\alpha_0}) = \langle \nabla_{\vec{\alpha_0}} \nu(\mathbf{r}, \vec{\alpha_0}) \rangle_{\vec{\alpha_0}} \tag{7}$$

where ν is the potential energy and $< \ldots >_{\vec{\alpha_0}}$ indicates an ensemble average at the state defined by $\vec{\alpha_0}$.

As above, the ensemble average is estimated using coordinates from the saved trajectories. Combining Equation 6 and 7 and indicating the finite time average explicitly gives

$$\Delta G(\alpha_0, \Delta\alpha) \approx \frac{1}{n} \sum_{i=1}^{n} \nabla_{\vec{\alpha_0}} \nu(\mathbf{r}_i, \vec{\alpha_0}) \cdot \vec{\Delta\alpha} \tag{8}$$

Examples of the use of this relationship are given by Cieplak et al. [12, 13] and by Smith and van Gunsteren [14].

Two approximations have been used in this section. The first is truncation of the Taylor series expansion of $\Delta G(\alpha, \alpha_0)$ to give Equation 6. The second is the use of a finite number of samples to estimate the ensemble average in Equation 7. We will examine the validity of the first approximation later in this paper.

Pictorial representation of free energy changes (PROFEC)

The free energy cost of adding a Lennard-Jones particle at a specific point in space can be calculated using Equation 2 (in a manner similar to the particle insertion method [15, 16]. If this is done for each point from a set of grid points defined with respect to the ligand, the free energy of adding a particle at a set of points around the ligand can be calculated as follows:

$$\Delta G(i, j, k) = -RT \ln\langle e^{-\Delta v(i,j,k)/RT}\rangle_0 \qquad (9)$$

where i, j, and k indicate a point in space relative to three atoms on the ligand, $\Delta G(i, j, k)$ is the change in free energy resulting from the addition of a Lennard-Jones particle at this point, and $v(i, j, k)$ is the van der Waals interaction energy between the particle and surrounding atoms. In general, ligand atoms are not included in the calculation because we are primarily interested in effects from the binding site and solvent. However, some regions of the ligand can be included if it is felt that this would better characterize the binding. In general, we include regions of the ligand that are not of immediate interest, especially if they are not rigid with respect to the region of interest. For example, we are interested in how changes in benzamidine's phenyl ring affect binding to trypsin (described in detail below), so interactions with the ring atoms are not included in the calculations but interactions with atoms of the amidine group are included.

The position and orientation of the grid is defined using one ligand atom to indicate the origin of the grid, a second atom to define the x-axes, and a third atom to define the xy-plane (the grid can also be translated so that it is centered on other regions of the ligand). If the ligand is not rigid, then grids fixed with respect to each rigid subgroup (for example, phenyl rings, amides, methylenes, etc.) should be used.

The 'cost' field calculated using Equation 9 can then be displayed as contour surfaces around the ligand (using a molecular graphics program) to suggest where new atoms could be added to improve binding

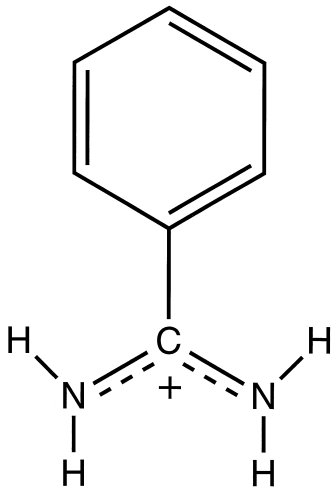

Figure 1. Benzamidine.

free energy. The electrostatic properties of the binding site can be examined by calculating the derivative of binding free energy with respect to charge at each grid point (see Equation 5). However, it is more useful to calculate the derivative assuming that a Lennard-Jones particle is first placed at that grid point. This weights the calculated free energy derivatives by how favorable (sterically) it is to have an atom at that point. To do this, umbrella sampling (see Ref. 17) can be used, giving the following equation:

$$\left[\frac{dG(i, j, k)}{dq}\right]_{LJ(i,j,k)} = \frac{\langle (dv(i, j, k)/dq) e^{-\Delta v(i,j,k)/RT}\rangle_0}{\langle e^{-\Delta v(i,j,k)/RT}\rangle_0} = \frac{\langle \phi(i, j, k) e^{-\Delta v(i,j,k)/RT}\rangle_0}{\langle e^{-\Delta v(i,j,k)/RT}\rangle_0} \qquad (10)$$

where $\phi(i, j, k)$ is the electrostatic potential at each grid point, $\Delta v(i, j, k)$ is the change in potential energy resulting from the addition of the Lennard-Jones particle (as used in Equation 9), and $[\ldots]_{LK(i,j,k)}$ indicates that the derivative is calculated assuming a Lennard-Jones particle is added at point i, j, and k. This derivative can be displayed by coloring the contour plot and can be used to suggest how a ligand charge distribution should be changed to improve binding. For example, this could show where hydrogen bonding groups could be added.

Although not shown explicitly, the ensemble averages in Equations 9 and 10 are approximated using a finite number of samples, as with Equation 3. They are therefore subject to the same sampling errors as the other methods.

This method is primarily intended to show changes in binding free energies, but it can also be used to show how the specificities of different receptors change as ligands are changed. This could aid in the design of a ligand that is intended to bind one receptor but not another. By analogy to Equation 1, this can be calculated from

$$\Delta\Delta G_{\text{Specificity}} = \Delta G_{\text{bound}+} - \Delta G_{\text{Bound}-} \qquad (11)$$

where $\Delta G_{\text{Bound}+}$ is the free energy for the receptor one wishes to optimize binding for and $\Delta G_{\text{Bound}-}$ is the free energy for the receptor one wishes to select against. An example of this would be designing a ligand such as trimethoprim (an antibacterial) that binds to bacterial dihydrofolate reductase much more tightly than it binds to the related human dihydrofolate reductase.

Methodology

All free energy calculations are based on MD simulations of benzamidine (see Figure 1) bound to trypsin and simulations of solvated benzamidine. The simulations were done using the SANDER module of AMBER 4.1 [18], with a time step of 1.0 fs. The Berendsen temperature coupling method was used to keep the average temperature constant at 300 K using a coupling constant of 0.2 ps [19]. The TIP3P [20] water model was used for all water molecules, and the SHAKE [21] algorithm was used to constrain all bond-lengths to their equilibrium values. An 8 Å cutoff was used for the non-bonded interactions. Bond, angle, torsional, and van der Waals parameters, as well as atomic charges, are from the Cornell et al. [23] force field. This force field does not have atomic charges for benzamidine or torsion parameters for benzamidine's phenyl/amidine bond; these were developed as follows.

Benzamidine atomic charges were developed using a RESP based [23] method. This consisted of fitting atomic charges to a Hartree–Fock 6-31G* electrostatic potential using an AM1 optimized structure. These charges were used to calculate a molecular mechanical structure (where the phenyl/amidine bond dihedral was constrained to its Hartree–Fock 6-31G* optimum value of 40.2°, and the final charges* were fit to the Hartree–Fock 6-31G* electrostatic potential calculated for this structure. Torsion parameters** for the phenyl/amidine bond in benzamidine were parametrized to reproduce the Hartree–Fock 6-31G* optimum dihedral angle and the barriers to rotation at 0.0° and 90.0°. All semiempirical and ab initio calculations were done using Gaussian 94 [24].

Atomic charges for the substituted benzamidines were generated using the RESP procedure by fitting to Hartree–Fock 6-31G* electrostatic potentials using AMI optimized structures. To determine if the benzamidine phenyl/amidine torsion parameter could also be used for the substituted benzamidines, we calculated 6-31G* optimized structures for *p*-amino, *p*-nitro, and *p*-carboxy benzamidines and compared them with unsubstituted benzamidine. For the *p*-amino and *p*-carboxy benzamidines, the amidine group was rotated 31° out of the plane of the phenyl ring. For the *p*-nitro benzamidine, the amidine group was 45° out of the plane. Since these are all within less than 10° of unsubstituted benzamidine, a common torsion parameter is appropriate for all the benzamidines.

The starting structure for the bound benzamidine simulations was built by adding hydrogens to the heavy-atom positions of a trypsin–benzamidine crystal structure (1BTY [25]). All crystal waters were maintained and additional waters were added to form a shell with a 16 Å radius around the benzamidine ligand. The net charge on the system was neutralized by adding nine chloride ions. Each of these were restrained to be near one of nine lysines on the protein surface. During the simulation, only protein residues within 10 Å of the benzamidine were allowed to move. These also had their backbone atoms restrained to the crystal structure positions with a force constant of 1 kcal/Å^2. All waters within 16 Å of the benzamidine were also allowed to move. These were restrained to the benzamidine with a force constant of 0.01 kcal/Å^2.

Equilibration of the bound system consisted of minimizing all atoms, followed by 200 ps of MD. Data collection consisted of doing 10 identical simulations, each 50 ps in length. Every 0.1 ps the coordinates were saved on disk. All free energy calculations involving the bound system are done using these saved coordinate sets.

The solvated benzamidine was simulated using periodic boundary conditions, in a cubic box with 1532 waters (about 36 Å on a side). The Berendsen pres-

* Benzamidine atomic charges: q(amidine nitrogen) = −0.800; q(amidine hydrogen) = 0.439; q(amidine carbon) = 0.635; q(Carbon 1) = −0.097; q(carbon 2) = −0.094; q(hydrogen 2) = 0.146; q(carbon 3) = −0.135; q(hydrogen 3) = 0.173; q(carbon 4) = −0.044; q(hydrogen 4) = 0.170.

** Benzamidine torsion parametrization: A = −3.25 kcal/mol, with a periodicity 2; A = 0.85 kcal/mol, with periodicity 4; A = 0.25 kcal/mol, with periodicity 6.

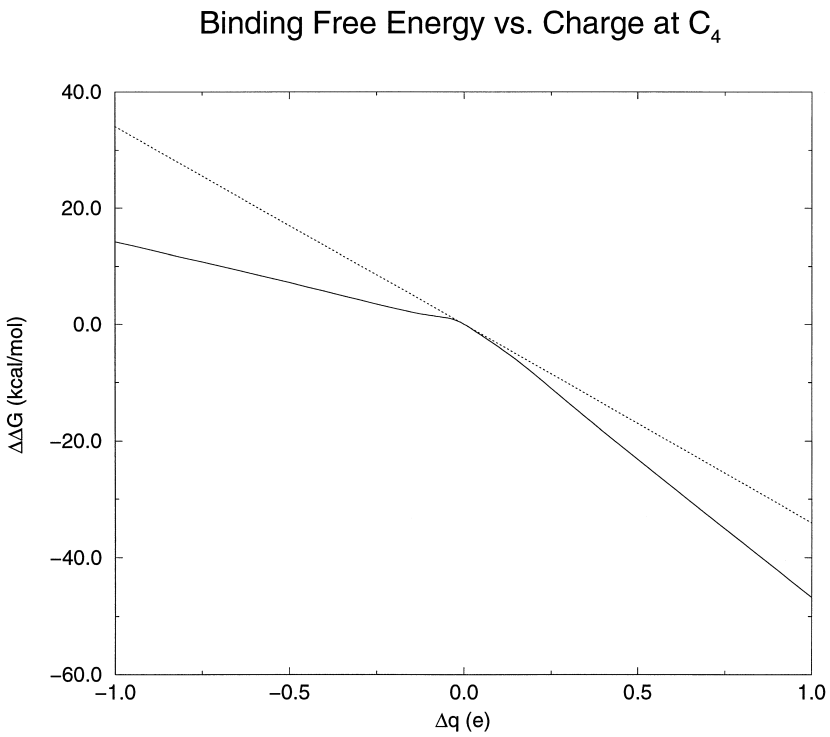

Figure 2. Change in binding free energy versus charge at C4 for Benzamidine bound to trypsin. The solid line is calculated using the SSMP method (Equations 3 and 1) and the dotted line is calculated using the linear approximation (Equations 8 and 1).

sure coupling method was used to keep the pressure constant at 1 bar, using a coupling constant of 0.2 ps [19]. This system was equilibrated for 200 ps. Data collection was the same as for the bound system (coordinates are saved every 0.1 ps during 10 simulations, each 50 ps long). All free energy calculations involving the solvated system are done using these saved coordinates.

The three methods described in the previous section were applied to the coordinates from the 10 benzamidine/trypsin simulations and the 10 solvated benzamidine simulations. It should be noted that the free energy derivatives with respect to parameters could have been found during the simulations, but we chose to use the same set of coordinates for each calculation to simplify comparison of the methods. For the PROFEC method, the Lennard-Jones test particle used corresponds to a united-atom methyl group ($R^* = 2.00$Å, $\epsilon = 0.15$ kcal/mol), and interaction energies with carbons and hydrogens of the phenyl ring are not included in the calculation.

Results and discussion

Coordinates from the 10-benzamidine/trypsin simulations and the 10 solvated benzamidine simulations were used with Equation 3 and 1 to estimate the change in binding free energy resulting from the change in para substituents on benzamidine. Table 1 gives the resulting free energy estimates for each of the 10 pairs of simulations (one bound simulation and one solvated simulation per pair) as well as the best free energy estimate calculated by combining all of the data. From this, the accuracy of the SSMP method can be evaluated for the nine benzamidine derivatives examined.

In general, the results are not highly accurate for this system and these simulation conditions; errors in excess of 2 kcal/mol are common for substitutions involving one heavy atom, and larger errors typically occur for larger substituents. However, we are primarily interested in using this method to identify better ligands, so in Table 2 we show only the rank order of binding free energy. From this, it can be seen that the substituent that gives the best experimental binding free energy (NH_2) is also predicted to give

Table 1. Experimental [31] and estimated relative binding free energy (kcal/mol) of para-substituted benzamidines to trypsin

Substituent	Exp.	Best	Simulation number									
			1	2	3	4	5	6	7	8	9	10
NH_2	−0.40	−2.57	−3.93	−2.59	−2.48	−2.50	−2.27	−1.83	−3.68	−2.09	−2.23	−2.30
OH	−0.10	−1.28	−1.37	−1.78	−1.73	−0.87	−1.09	−1.14	−1.28	−0.18	−1.20	−0.59
H	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CH_3	0.27	−0.27	−0.42	−1.37	−0.22	−0.73	−0.11	1.75	−0.74	−0.47	2.29	3.10
F	0.54	1.18	1.39	0.46	0.99	1.06	1.35	1.61	1.44	1.30	1.47	0.87
Cl	0.68	0.96	3.29	0.10	1.05	2.12	0.29	5.36	−0.44	2.12	3.00	2.89
OCH_3	0.75	2.93	6.69	−1.17	11.00	8.54	9.51	23.37	4.60	2.45	3.86	12.59
CO_2^-	1.43	49.29	71.55	44.68	70.34	78.22	34.38	68.82	84.30	71.33	61.65	45.12
NO_2	1.50	9.21	14.82	1.89	9.19	11.39	12.59	15.05	31.82	16.28	10.83	18.58

the best binding by each of the 10 pairs of simulations, and the second best binder (OH) is correctly predicted by nine of the 10 simulations. The rankings for substitutions that do not bind as well as benzamidine are not as impressive, but are still often within one place of the correct result. This is consistent with results from an earlier study [7] that suggested that if perturbations involve the addition of one or two heavy atoms, the relative binding free energy can be estimated qualitatively.

It is interesting to contrast the SSMP method with the linear interaction energy (LIE) method recently proposed by Aqvist et al. [26] (see also Ref. 27), a semiempirical, simulation-based free energy calculation method that estimates binding free energies. A strength of the LIE method is that it gives a ligand's absolute binding free energy from a single simulation of the ligand bound to the receptor and a single simulation of the solvated ligand (using average van der Waals and electrostatic interaction energies). From absolute binding free energies, relative free energies of unrelated ligands can be found. The SSMP method (and, to a lesser extent, all free energy calculation methods based on Equation 2) can only accurately calculate free energy differences for related ligands. On the other hand, the SSMP method only requires one bound simulation and one solvated simulation of a reference ligand to estimate the binding free energy of a large number of related ligands. In principle, this makes the two methods nicely complementary, because the LIE method could be used to estimate free energies for a set of unrelated ligands (by explicitly simulating them), and the SSMP method can be used to examine how binding free energy changes as a result of small changes in these ligands (by reusing coordinates from the LIE simulations).

If more accurate calculations are desired, additional simulations designed to sample the perturbed states more completely should be performed. One way this could be done is by performing a simulation using 'soft atoms' at positions where substituents are to be added (see Ref. 28). The rationale for this is that the simulation will sample conformations appropriate for a (hard) Lennard-Jones atom at the soft atom position and will also sample conformations appropriate for no atom (or a hydrogen) at the soft atom position. The difficulty with this is it is not clear how to parametrize the soft atom such that it does generate coordinates corresponding to all of the states of interest. This may be especially problematic for highly charged atoms like those in charged function groups or even hydroxyl and amine groups.

Another way to improve sampling of the perturbed states is by using the composite reference state (CRS) method described elsewhere [7]. This could be done for the benzamidine/trypsin system by performing a simulation of benzamidine with a small Lennard-Jones particle (possibly fluorine) at the para position. Coordinates from this simulation and the benzamidine simulation would be combined, giving a set of coordinates that could have been generated from a single simulation of benzamidine with a soft atom at the paraposition. Additional simulations with different substituents can also be performed and combined with the others. The advantage of this method is that more than one substituent can be included in the reference coordinate set and each substituent will be sampled (because they were simulated independently). The disadvantage is that the energy function of the reference

Table 2. Experimental [31] and estimated ranking of binding of para-substituted benzamidines to trypsin

Substituent	Exp.	Best	Simulation number									
			1	2	3	4	5	6	7	8	9	10
NH_2	1	1	1	1	1	1	1	1	1	1	1	1
OH	2	2	2	2	2	2	2	2	2	3	2	2
H	3	4	4	5	4	4	4	3	5	4	3	3
CH_3	4	3	3	3	3	3	3	5	3	2	5	6
F	5	6	5	7	5	5	6	4	6	5	4	4
Cl	6	5	6	6	6	6	5	6	4	6	6	5
OCH_3	7	7	7	4	8	7	7	8	7	7	7	7
CO_2^-	8	9	9	9	9	9	9	9	9	9	9	9
NO_2	9	8	8	8	7	8	8	7	8	8	8	8

state is not known exactly. It is calculated after the simulations are completed and is a function of the free energy difference between the different states explicitly simulated. Thus, inaccuracies in estimated free energy differences between the simulated states will affect the determination of the reference state energy.

We would have liked to evaluate the SSMP method for cases where ligand atoms are removed (such as changing a benzamidine aromatic carbon and hydrogen to a nitrogen), but were unable to locate appropriate experimental data. This would be of interest to us because we expect estimates of this type of perturbation to give larger errors than estimates involving the addition of atoms, and we would like to characterize the magnitude of the error. Larger errors are expected because simulations of the reference state (in this case, benzamidine) must be able to sample all accessible conformations of the perturbed state (in this case, all the substituted benzamidines). This requirement is met when atoms are added because they could appear in cavities that spontaneously form in the receptor or solvent. To adequately sample perturbed states that result from the removal of atoms, the simulation would have to sample conformations where the atom that will be removed is superimposed on atoms of the receptor or solvent. This is not likely to happen, so larger errors can be expected (as above, sampling could be improved by using soft atoms or the CRS method).

Coordinates from the 10 benzamidine/trypsin simulations and the 10 solvated benzamidine simulations were used to estimate the change in binding free energy as a function of changes in the charge (q), van der Waals equilibrium distance (R^*), and van der Waals well depth (ϵ) of the carbon and the hydrogen at the para positions of benzamidine (position 4). The changes in binding free energy are calculated using two methods: the SSMP using Equations 3 and 1 and linearized free energy changes using Equations 8 and 1. Since the results from both methods are based on identical sets of conformations, the differences result exclusively from ignoring higher order deriva- tives. This allows us to examine the magnitude of the errors resulting from the use of the linear approximation. The results are shown in Figures 2–8 (the results for the meta positions are shown in Figures 10–23 as supplementary material).

Figures 2–4 plot the free energy as a function of change in charge at benzamidine's para position, showing that the linear approximation gives the correct general trend in binding free energies for this case. Even when the charge on two atoms is changed simultaneously, the change in free energy is still predicted reasonably well by the linear approximation (see Figure 4, which shows binding free energy as a function of charge on the hydrogen, where the charge on the adjacent carbon is adjusted to maintain a net charge of zero). Plots showing the change in binding free energy as a function of change in charge at the meta positions (Figures 10–23 as supplementary material) show essentially the same thing. Figures 5 and 6 give the change in binding free energy resulting from changes in the van der Waals equilibrium distance (R^*) at the para position, and Figures 7 and 8 show the change in binding free energy resulting from changes in the van der Waals equilibrium well depth (ϵ) at the same position. As with changes in charge, the linear approximation does give general trends in the free energy.

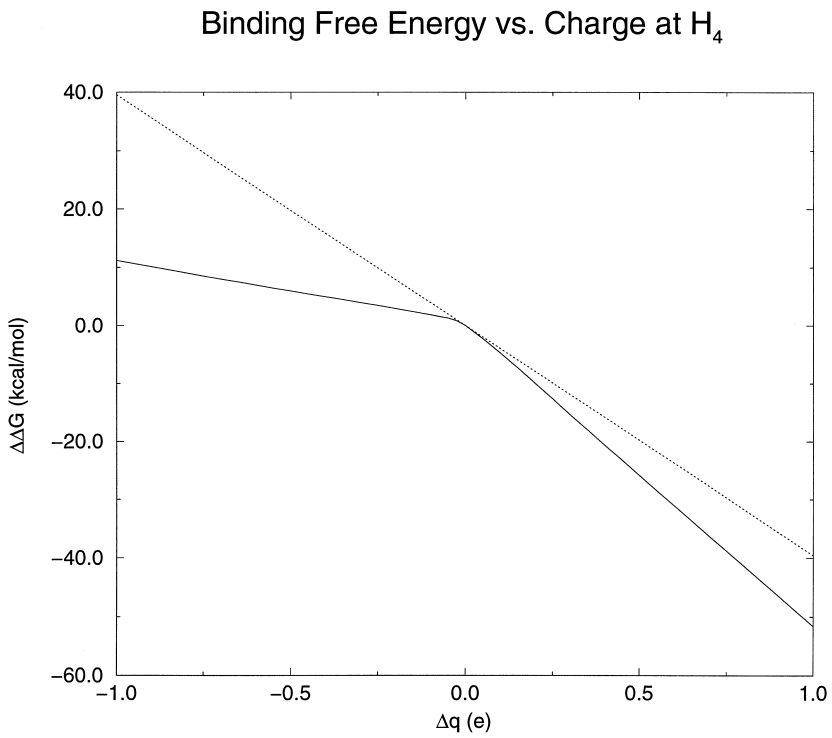

Figure 3. Change in binding free energy versus charge at H4 for benzamidine bound to trypsin.

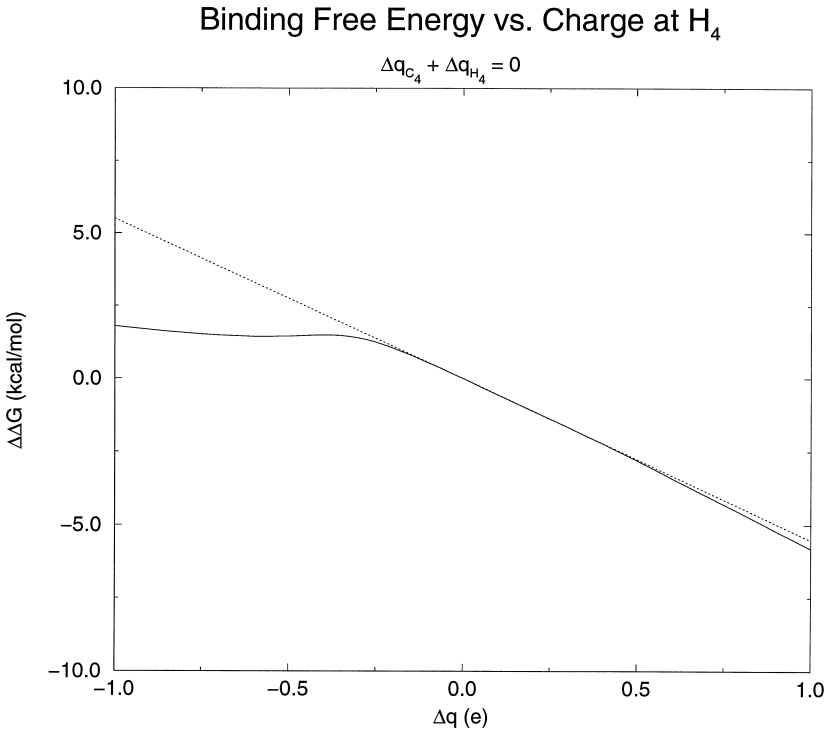

Figure 4. Change in binding free energy versus charge at H4 ($\Delta\nu_{C4} + \Delta\nu_{H4} = 0$) for benzamidine bound to trypsin.

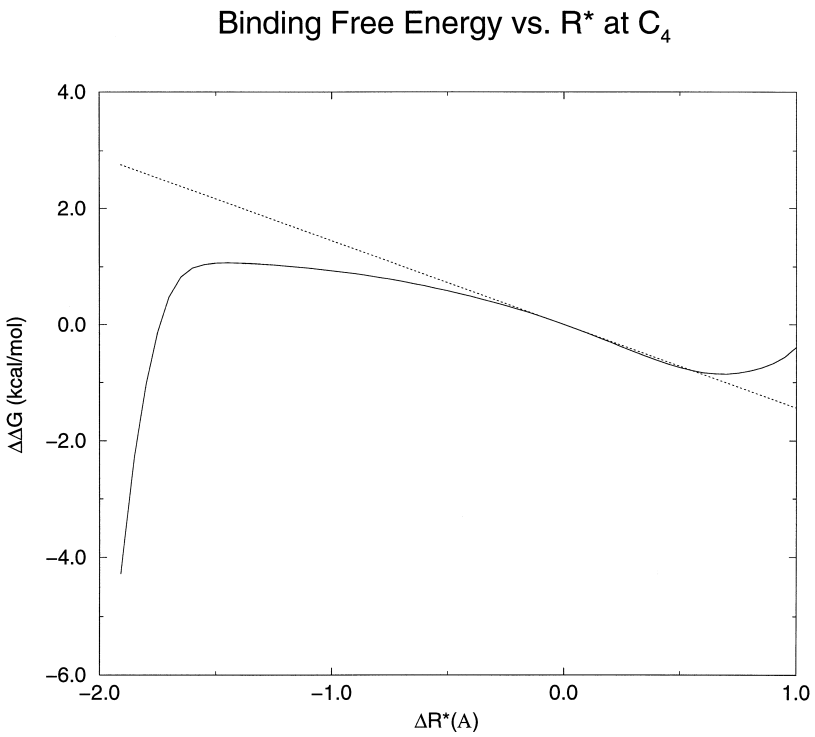

Figure 5. Change in binding free energy versus R* at C4 for benzamidine bound to trypsin.

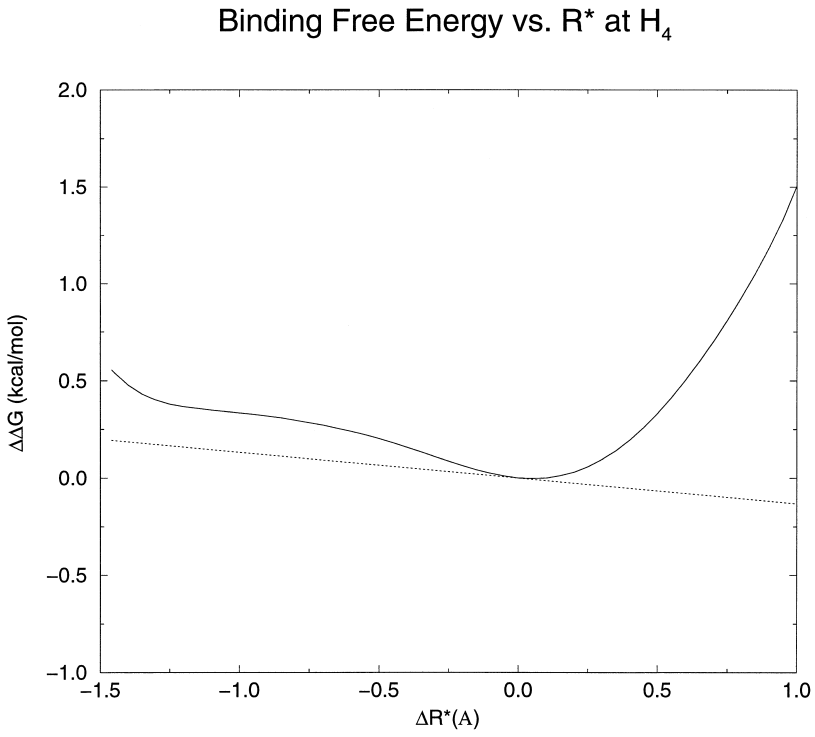

Figure 6. Change in binding free energy versus R* at H4 for benzamidine bound to trypsin.

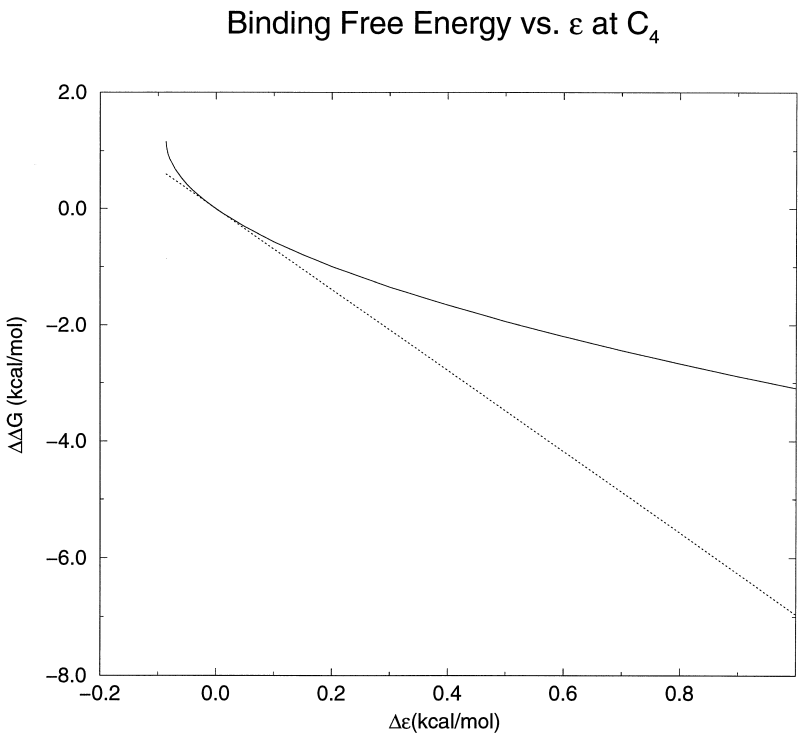

Figure 7. Change in binding free energy versus ϵ at C4 for benzamidine bound to trypsin.

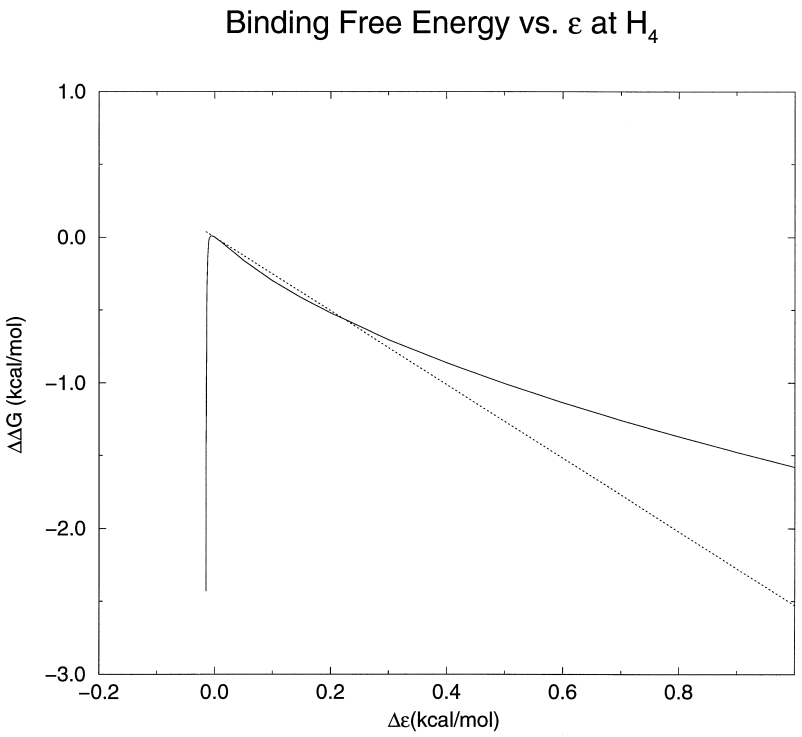

Figure 8. Change in binding free energy versus ϵ at H4 for benzamidine bound to trypsin.

In all cases, the linear approximation is adequate to indicate whether the value of parameters should be increased or decreased to improve binding free energies. Thus, it can be useful for suggesting new substituents that give better binding ligands. However, it will not give accurate free energy estimates. Our results seem consistent with those published by Smith and van Gunsteren [14], who examined the use of the first terms in a free energy expansion for a system with one adjustable parameter, and found that free energy estimates based on only the linear term are not accurate. (Unfortunately, they also found that free energies estimated using the higher order derivatives converge more slowly than estimates using only the linear terms.)

A major limitation in the use of free energy derivatives is that all the changes are calculated for the existing atom centers (possibly including additional 'dummy' atoms). Thus, it is not useful for suggesting changes that involve adding new substituents distant from existing atoms. The PROFEC method is intended to address this problem by giving a visual representation of how binding free energy can be changed for a continuous region around the ligand. Contour plots of this free energy can then be used to identify regions around the ligand where additional groups can be added (presumably filling permanent or transient cavities in the receptor). When visualized in three dimensions, this shows the general shape of the binding site. However, its usefulness also comes from the fact that it includes effects from changes in solvation free energy.

Coordinates from the benzamidine/trypsin and the solvated benzamidine simulations were thus used with Equations 9 and 1 to calculate the change in binding free energy resulting from adding a Lennard-Jones particle at points around the ligand. The resulting three-dimensional field was then used to make contour maps of the cost of adding a particle and displayed using MidasPlus [29, 30] and a MidasPlus delegate (written by ourselves) that displays variable colored contour lines. Figure 9 shows the surface corresponding to a free energy change of zero ($\Delta\Delta G = 0$), dividing the space around the benzamidine into two regions. Particles added inside the surface (close to benzamidine) will improve binding free energy, and particles outside the surface will make binding worse. From this it can be seen that there may be enough space to add a small substituent (or atom) at positions 3 and 4, but it will not help much. This is reasonably consistent with the experimental data (see Table 1).

Equations 10 and 1 were used to find the derivative of binding free energy with respect to charge, as a function of position around the ligand. This was used to color the contour lines in Figure 9, with red indicating a negative derivative, and blue indicating a positive derivative. The results in the area around the para position being red (presumably because of the nearby hydroxyl group of Ser^{195}), suggesting that a particle with a positive charge (or a dipole with its positive end pointing out, such as a hydrogen bond donating group) be added at this point. The same conclusion can be reached by examining Figures 3 and 4. This is also consistent with the experimental results shown in Table 1, since hydrogen bond donor substituents are the best binders and hydrogen bond acceptor substituents are the worst.

Conclusions

Based on the results shown here, all three methods (SSMP, free energy derivatives with respect to atomic parameters, and PROFEC) can be useful, if their limitations are kept in mind. For example, the calculations give the rank order of substituents at the para position of benzamidine reasonably accurately, but the calculated free energies are in error by many kilocalo- ries. For reasons discussed in the text, one can be most confident of modifications involving changes in atom type (e.g. hydrogen to fluorine) or the addition of one or two heavy atoms (removing atoms is likely to give larger errors because of the need to sample states with overlapping atoms). Thus, we would only advise using this method to identify ligand modifications that are likely to give improved binding. We do not suggest that this method be used for quantitative estimates of binding free energies.

For this system, the second method (free energy derivatives) gives generally correct indications of how the binding will change as parameters are varied, and should be useful for suggesting substituents that could improve binding. However, this is also not likely to give accurate estimates of free energy changes.

The main drawback with the use of free energy derivatives is that, in practice, it can be difficult to use them to suggest real changes. This is because most real changes involve changing many parameters (not just one) and changes often happen at points in space distinct from existing atoms. The third method, PROFEC, was developed to address these problems by repeatedly placing a real atom over a continuous region of

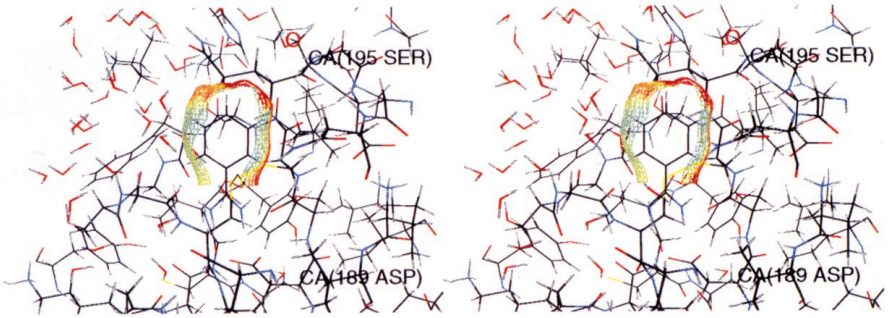

Figure 9. Stereoview of PROFEC contour for benzamidine bound to trypsin.

space around the ligand. This can be used to give a visual representation of where atoms should be added to improve binding.

Based on our results, we would propose that ligand design could be done as follows. Perform a simulation of a known ligand bound to the target receptor and a simulation of the solvated ligand. Based on saved coordinates, use PROFEC to identify regions around the ligand where substituents should be added, and to suggest the size, shape, and charge distribution that could improve binding. Free energy derivatives can be used to give additional information at points corresponding to existing atom centers. SSMP methods could then be used to evaluate the possibilities suggested by PROFEC and the free energy derivatives. The most promising ligands could then be evaluated further using more accurate free energy calculation methods, or tested experimentally.

We are currently evaluating the possibility of using the PROFEC cost field to screen compounds from a molecular data base using the molecular docking program DOCK (for an overview of DOCK, see Ref. 8). Although the field is not completely appropriate for this purpose, we feel that it could be useful because it is based on a dynamic structure of the receptor and because solvation information can be included if deemed appropriate (we have not yet examined this issue).

We are also applying PROFEC to study protein–ligand binding (R.W. Dixon, R.J. Radmer and P.A. Kollman, manuscript in preparation; J.W. Pitera, R.J. Radmer and P.A. Kollman, unpublished results) and protein stability [32]. Preliminary results are encouraging in each of these areas.

Acknowledgements

Molecular graphics images were produced using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081, T. Ferrin, P.I.). We would like to thank Jed Pitera for helpful discussions. We would like to acknowledge support from NIH Grants GM-39552 (C.S. Craik, P.I.) and GM-29072 (P.A. Kollman, P.I.).

References

1. McCammon, J.A. and Harvey, S.C., Dynamics of Proteins and Nucleic Acids, Cambridge University Press, New York, NY, 1987.
2. Beveridge, D.L. and DiCapua, F.M., Annu. Rev. Biophys. Biophys. Chem., 18 (1989) 431.
3. Van Gunsteren, W.F., Protein Eng., 2 (1988) 5.
4. Straatsma, T.P. and McCammon, J.A., Annu. Rev. Phys. Chem., 43 (1992) 407.
5. Kollman, P.A., Chem. Rev., 93 (1993) 2395.
6. Pearlman, D.A., J. Phys. Chem., 98 (1994) 1487.
7. Radmer, R.J. and Kollman, P.A., J. Comput. Chem., 18 (1997) 902.
8. Kuntz, I.D., Meng, E.C. and Shoichet, B.K., Acc. Chem. Res., 27 (1994) 117.
9. Tembe, B.L. and McCammon, J.A., Comput. Chem., 8 (1984) 281.
10. Allen, M.P. and Tildesley, D.J., Computer Simulation of Liquids, Oxford University Press, New York, NY, 1987.
11. Zwanzig, R.W., J. Chem. Phys., 22 (1954) 1420.
12. Cieplak, P. and Kollman, P.A., J. Comput.-aided Mol. Design, 7 (1993) 291.
13. Cieplak, P., Pearlman, D.A. and Kollman, P.A., J. Chem. Phys., 101 (1994) 627.
14. Smith, P.E. and van Gunsteren, W.F., J. Chem. Phys., 100 (1994) 577.
15. Widom, B., J. Chem. Phys., 39 (1963) 2808.
16. Shing, K.S. and Chung, S.T., J. Phys. Chem., 91 (1987) 1674.
17. Torrie, G.M. and Valleau, J.P., J. Comput. Phys., 23 (1977) 187.
18. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham III, T.E., Ferguson, D.M., Seibel, G.L., Singh, U.C.,

Weiner, P. and Kollman, P.A., AMBER v. 4.1, University of California, San Francisco, CA, U.S.A., 1995.
19. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R., J. Chem. Phys., 81 (1984) 3684.
20. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L., J. Chem. Phys., 79 (1983) 926.
21. Van Gunsteren, W.F. and Berendsen, H.J.C., Mol. Phys., 34 (1977) 1311.
22. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A., J. Am. Chem. Soc., 117 (1995) 5179.
23. Bayly, C.I., Cieplak, P., Cornell, W.D. and Kollman P.A., J. Phys. Chem., 97 (1993) 10269.
24. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P.M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T., Petersson, G.A., Montgomery, J.A., Raghavachari, K., AI-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Replogle, E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C. and Pople, J.A., Gaussian 94, Revision B.3, Gaussian Inc., Pittsburgh, PA, U.S.A., 1995.
25. Katz, B.A., Finer-Moore, J., Mortezaei, R., Rich, D.H. and Stroud, R.M., Biochemistry, 34 (1995) 8264.
26. Aqvist, J., Medina, C. and Samuelsson, J.-E., Protein Eng., 7 (1994) 385.
27. Hansson, T. and Aqvist, J., Protein Eng., 8 (1995) 1137.
28. Beutler, T.C., Mark, A.E., van Schaik, R.C., Gerber, P.R. and van Gunsteren, W.F., Chem. Phys. Lett., 222 (1994) 529.
29. Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R., J. Mol. Graph., 6 (1988) 13.
30. Huang, C.C., Pettersen, E.F., Klein, T.E., Ferrin, T.E. and Langridge, R., J. Mol. Graph., 9 (1991) 230.
31. Mares-Guia, M., Nelson, D.L. and Rogana, E., J. Am. Chem. Soc., 99 (1977) 2331.
32. Wang, L., Veenstra, D.L., Radmer, R.J. and Kollman, P.A., “Can One Predict Protein Stability? An Attempt to do so for Residue 133 in T4 Lysozyme Using a Combination of Free Energy Derivatives, PROFEC and Free Energy Perturbation Methods”, Proteins: Structure, Function and Genetics, accepted for publication.