Chapter 3

New Free Energy Calculation Methods for Structure-Based Drug Design and Prediction of Protein Stability

Lu Wang, Mats A. L. Eriksson, Jed Pitera, and Peter A. Kollman

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143

We summarize some aspects of our recent studies of protein stability and protein-ligand interactions by a combination of several newly developed approximate free energy calculation methods and rigorous free energy calculation methods. The approximate free energy simulation methods employed in our studies include free energy derivatives (FED), pictorial representation of free energy changes (PROFEC), chemical Monte Carlo/molecular dynamics simulation (CMC/MD), Poisson-Boltzmann continuum electrostatics/solvent accessible area (PB/SA) and generalized Born approximation/solvent accessible area (GB/SA) methods. The thermostability of T4 lysozyme and the binding free energies of HIV-1 reverse transcriptase inhibitors were analyzed as test cases. It is shown that these different approaches are complementary to each other and when combined, can make predictions efficient, comprehensive and insightful. Potential applications of this strategy in structure-based drug design and protein engineering are discussed.

Calculations of free energies of ligand binding or protein stability can be very useful in drug design and protein engineering (1). The most rigorous approaches for calculating free energy changes, i.e., the free energy perturbation (FEP) or thermodynamic integration (TI) methods, have been limited in practical applications, due to their computationally intensive nature. Recently, there has been considerable interest in developing approximate, yet efficient free energy simulation methods. Notably, Åqvist *et al* (2) proposed the linear interaction energy approximation that correlates the binding free energies of ligands with their average interaction energies in protein and solvent. Radmer & Kollman (3) developed the pictorial representation of free energy changes (PROFEC) which can suggest modifications of a ligand to increase its binding affinity. To quickly rank the binding affinities of a series of ligands, Pitera & Kollman (4) introduced the chemical Monte Carlo/molecular dynamics (CMC/MD) technique and Kong & Brooks (5) have developed the λ dynamics simulation method. The potential of using free energy derivatives (FED) in estimation of free energy changes (6,7) and analog design (8) has been explored. In

In Rational Drug Design; Parrill, Abby L., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 1999.

addition, Shen *et al* (9,10) advocated the Poisson-Boltzmann electrostatics/solvent accessible area (PB/SA) scheme for estimation of binding free energies of ligands. In general, these methods are much faster than FEP or TI and therefore more suited to practical applications. Due to their approximate nature, they may be less accurate. In this work, we tested some of these approximate methods in two cases. One is the thermostability problem of T4 lysozyme (11). Another is the binding free energies of HIV-1 reverse transcriptase (RT) TIBO inhibitors (12). For T4 lysozyme, we first used FED and PROFEC to suggest candidate modifications that may improve the stability of the protein and then analyzed an interesting qualitative prediction with a TI calculation. For the HIV-1 RT inhibitors, we first analyzed their binding affinities with the CMC/MD and PB/SA methods and later supported one of the results by TI calculations. As a comparison, the solvation free energies of the ligands were calculated with both PB/SA method and the generalized Born approximation/solvent accessible area (GB/SA) method (13). For both cases, we found that the approximate methods gave quite reasonable results and therefore may be useful in practical structure-based drug design and protein engineering applications.

Theory

Thermodynamic cycles. The following thermodynamic cycle was used to assess the effect of a modification on the stability of T4 lysozyme:

$$\begin{array}{ccc} & \Delta G_{u\rightarrow f} & \\ T4\,(f) & \Leftarrow & T4\,(u) \\ \Delta G_{f\rightarrow f^*} \Downarrow & & \Downarrow \Delta G_{u\rightarrow u^*} \\ T4^*(f) & \Leftarrow & T4^*(u) \\ & \Delta G^*_{u\rightarrow f} & \end{array} \qquad (1)$$

where T4 and T4* stand for T4 lysozyme and a further modified enzyme. "f" and "u" stand for the folded and unfolded states of the enzyme. Because free energy is a state function, we have

$$\Delta\Delta G_{fold} = \Delta G^*_{u\rightarrow f} - \Delta G_{u\rightarrow f} = \Delta G_{f\rightarrow f^*} - \Delta G_{u\rightarrow u^*} \qquad (2)$$

where $\Delta\Delta G_{fold}$ is the folding free energy difference or stability difference between T4* and T4. Direct calculations of $\Delta G^*_{u\rightarrow f}$ and $\Delta G_{u\rightarrow f}$ are difficult due to the large conformational changes involved. Fortunately, $\Delta G_{f\rightarrow f^*}$ and $\Delta G_{u\rightarrow u^*}$ can be readily determined by free energy calculation methods.

The following thermodynamic cycle was used to calculate the relative binding affinity of two ligands, L_1 and L_2, to the same protein, P:

$$\begin{array}{ccc} & \Delta G_{L_1,b} & \\ L_1 + P & \Rightarrow & L_1{\bullet}P \\ \Delta G_{L_1\rightarrow L_2,s} \Downarrow & & \Downarrow \Delta G_{L_1\rightarrow L_2,p} \\ L_2 + P & \Rightarrow & L_2{\bullet}P \\ & \Delta G_{L_2,b} & \end{array} \qquad (3)$$

where $\Delta G_{L_1,b}$ and $\Delta G_{L_2,b}$ are the binding free energies of L_1 and L_2. $\Delta G_{L_1\rightarrow L_2,s}$ and $\Delta G_{L_1\rightarrow L_2,p}$ are the mutational free energies of $L_1\rightarrow L_2$ in the solvent and in the protein, respectively. The difference in binding free energy between L_1 and L_2,

$\Delta\Delta G_b$, can be expressed as

$$\Delta\Delta G_b = \Delta G_{L2,b} - \Delta G_{L1,b} = \Delta G_{L1\rightarrow L2,p} - \Delta G_{L1\rightarrow L2,s} \quad (4)$$

Similarly, direct calculations of $\Delta G_{L1,b}$ and $\Delta G_{L2,b}$ are difficult, but $\Delta G_{L1\rightarrow L2,s}$ and $\Delta G_{L1\rightarrow L2,p}$ can be calculated from simulations.

Free energy calculation with thermodynamic integration (TI) method. In this study, TI (1) was used to calculate the free energy changes. For the transformation of one state into another, a coupling parameter λ is introduced and the Hamiltonians of the two states are defined as H_0 (λ=0) and H_1 (λ=1). The free energy change of the transformation is expressed as the following integral (1)

$$\Delta G = \int_0^1 < \partial H(\lambda) / \partial\lambda >_\lambda \, d\lambda \quad (5)$$

where $< \partial H(\lambda) / \partial\lambda >_\lambda$ is an ensemble average at λ. In practice, a number of evenly-spaced windows with different λ values ranging from 0 to 1 are chosen and at each window, $< \partial H(\lambda) / \partial\lambda >_\lambda$ is calculated by averaging over molecular dynamics trajectories. The kinetic energy contribution can be neglected and ΔG is estimated by

$$\Delta G \approx \sum_i < \partial V(\lambda) / \partial\lambda >_{\lambda_i} \Delta\lambda \quad (6)$$

where λ_i is the λ value of the *i*th window and $\Delta\lambda$ is the interval between successive windows. According to the trapezoidal approximation, corrections are made to the contributions of the first and last windows (multiplying by 0.5). V is the potential function that describes the atomic interactions in the system.

Free energy derivatives (FED). The partial derivatives of free energy with respect to the nonbonded interaction parameters, q_i, ε_i and R_i^*, were calculated by the following equations (6,7):

$$\partial G / \partial q_i = \left\langle \sum_j q_j / \left(\varepsilon R_{ij}\right) \right\rangle = 1 / q_i \langle V_{coul}(i) \rangle \quad (7)$$

$$\partial G / \partial \varepsilon_i = \left\langle \sum_j \left[1 / (2\varepsilon_i)\right] \varepsilon_{ij} \left[\left(R_{ij}^* / R_{ij} \right)^{12} - 2 \left(R_{ij}^* / R_{ij} \right)^6 \right] \right\rangle = \left[1 / (2\varepsilon_i)\right] \langle V_{L-J}(i) \rangle \quad (8)$$

$$\partial G / \partial R_i^* = \left\langle \sum_j \varepsilon_{ij} \left(12 / R_{ij}^*\right) \left[\left(R_{ij}^* / R_{ij} \right)^{12} - \left(R_{ij}^* / R_{ij} \right)^6 \right] \right\rangle \quad (9)$$

where $<V_{coul}(i)>$ and $<V_{L-J}(i)>$ are the mean Coulombic and Lennard-Jones interaction energies of the *i*th atom with the rest of the system. To analyze the change of a protein's stability due to modifications of the properties of its *i*th atom, according to equation 2, the free energy derivatives need to be calculated for both the folded and the unfolded states of the protein and their difference indicates the stability change.

Pictorial representation of free energy changes (PROFEC). The PROFEC contour maps can be used to visualize how a protein's stability or a ligand's binding affinity changes when additional particles are added to a residue of the protein or the ligand (3). The contour map is generated by evaluating the insertion free energy of a test particle at various grid points near the residue of interest, using coordinates from molecular dynamics simulations (3). To analyze a protein's stability, according to equation 2, two contour maps for the folded and the unfolded states of the protein have to be generated and their difference map is used to indicate the stability change of the protein upon modification of the residue of interest. The free energy cost of adding a test particle at a grid point is calculated by (3)

$$\Delta G\,(i, j, k) = -RT \ln \langle \exp(-\Delta V(i, j, k) / RT \rangle_0 \qquad (10)$$

where i, j and k are the coordinates of a grid point, ΔG (i, j, k) is the insertion free energy and ΔV(i, j, k) is the interaction energy between the test particle and the surrounding atoms.

Chemical Monte Carlo/molecular dynamics (CMC/MD). The CMC/MD method (4) has recently been developed for determination of relative binding free energies of a series of ligands to a common receptor. The method (described in detail in ref. 4) employs the MD method for generating a set of coordinates for one distinct chemical system and the MC method to sample the *chemical* space of the system. Applied to a protein-inhibitor system, the chemical space can be 5-10 different derivatives of the inhibitor. Each derivative is included in the simulated system but the potential function is "masked" so that only one ligand interacts with the protein and solvent at a time. The Monte Carlo steps consist of changes to the "masking" function, effectively changing the ligand being simulated. MD is used to propagate the coordinates of all the ligands. During the course of the MC/MD run, the probability (P_i) of each inhibitor 'i' is accumulated according to the Metropolis (14) criteria for accepting an inhibitor in an MC step:

$$\text{if } \Delta E_i \leq 0 \Rightarrow P_{acc} = 1, \qquad \text{if } \Delta E_i > 0 \Rightarrow P_{acc} = \exp(-\Delta E_i/RT) \qquad (11)$$

where ΔE_i is the difference in protein-inhibitor interaction energy between a randomly chosen ligand and the old ligand and P_{acc} is the acceptance probability. Prior to each MC step, the "Boltzmann" probabilities of each ligand 'i' is calculated by:

$$P_i = \exp(-\Delta E_i/RT)/\Sigma\exp(-\Delta E_i/RT) \qquad (12)$$

It can be shown that if an infinite number of MC steps were performed on a given Cartesian conformation, the resulting probability distribution would coincide with that calculated from equation 12. In the TIBO-HIV-1 RT systems studied here, we used the averaged P_i's from equation 12, since they also allow for estimations of the relative free energies of poorly sampled inhibitors. The relative free energy of the bound state for inhibitors 'j' and 'i' is then related to their ratio of probabilities according to:

$$\Delta G_j - \Delta G_i = -RT \ln P_j/P_i \qquad (13)$$

Solvent effects, i.e. differences in free energy of binding due to different free energies of solvation (ΔG_{solv}), can be taken into account by testing the acceptance against $\Delta E_i - \Delta G_{solv,i}$, rather than ΔE_i in equation 11. Using ΔG_{solv} as a biasing potential in the MC step, the sampling mirrors the binding free energy - which is the relevant property when ranking inhibitors - rather than the free energy of the bound state.

The CMC/MD, as outlined above, was found to converge very slowly when applied to a series of HIV-1 RT TIBO inhibitors. In order to increase the convergence rate, a variant of this method was developed - herein called the "adaptive CMC/MD" method (J. Pitera, unpublished). Rather than sampling the chemical space according to the relative free energies of the inhibitors, the goal of the adaptive CMC/MD is instead to sample this space *evenly*. This can be achieved by introducing biasing offsets, $\Delta G_{offs,i}$, that for each ligand 'i' reflects its relative free energy in the bound state. An MC sampling by testing the acceptance against ΔE_i-$\Delta G_{offs,i}$, rather than ΔE_i in equation 11, would then result in an even sampling of all ligands, since all ΔE_i-$\Delta G_{offs,i}$ would be equal to zero. The offsets are solved iteratively. Starting with all $\Delta G_{offs,i}$=0, the probabilities of each ligand are calculated according to equation 12, averaged over a certain number of CMC/MD-cycles (a CMC/MD run). A first set of $\Delta G_{offs,i}$'s, relative to some arbitrarily chosen ligand, is estimated from equation 13, and is then used as biasing offsets in the next MC/MD run. The offsets are then adjusted after each CMC/MD run, by averaging the P_i's from equation 12, and add the adjusted offsets obtained from equation 13 to $\Delta G_{offs,i}$. When this procedure has converged, all P_i's are equal and the relative free energies of the bound state ($\Delta G_{bound,i}$) equals to -$\Delta G_{offs,i}$. $\Delta\Delta G_b$ can then be calculated by subtracting ΔG_{solv} from ΔG_{bound}. We estimated ΔG_{solv} by the GB/SA method discussed below.

Poisson-Boltzmann continuum electrostatics/solvent accessible area (PB/SA) and Generalized Born approximation/solvent accessible area (GB/SA) methods. In both the PB/SA and the GB/SA methods, a solvated protein or a small solute molecule is represented as a low dielectric cavity containing fixed charges and dipoles. The solvent water is represented as a medium of dielectric constant 80 which may contain ions. For the PB/SA method, the electrostatic field around a protein or a small solute molecule in the presence of salt is estimated by the solution to the linearized Poisson-Boltzmann equation (15)

$$\vec{\nabla}(\varepsilon(\mathbf{x})\vec{\nabla}\phi(\mathbf{x})) - \kappa^2\phi(\mathbf{x}) + 4\pi\rho(\mathbf{x}) = 0 \qquad (14)$$

where ϕ is the electric potential, ε is the dielectric constant, ρ is the fixed charge density and κ is the modified Debye-Huckel parameter which depends on the ionic strength and temperature of the solution. The Poisson-Boltzmann equation can be solved numerically by the finite difference method, in which the continuous functions are approximated by distinct values at points on a cubic grid (15). With the electrostatic potential obtained from solving the PB equation, the electrostatic interaction between a protein or a small solute with the solvent is expressed as (15)

$$\Delta G_{pol} = 1/2 \sum \phi_i q_i \qquad (15)$$

where ϕ_i is the potential on charge q_i and the sum is over the fixed charges. To estimate the electrostatic contribution to the hydration free energy of a molecule, two calculations, one for the molecule *in vacuo* and the other for the molecule in aqueous solution, should be performed and their difference in ΔG_{pol} gives the electrostatic contribution to the solvation free energy. $\Delta G_{L_1 \to L_2,s}$ and $\Delta G_{L_1 \to L_2,p}$ in equation 4 may be approximately estimated by the PB/SA or the GB/SA methods. For $\Delta G_{L_1 \to L_2,s}$, this involves the calculation of the solvation free energy difference of L_1 or L_2. According to equation 15, the calculation of the electrostatic contribution to $\Delta G_{L_1 \to L_2,s}$ is straightforward. The nonpolar contribution to $\Delta G_{L_1 \to L_2,s}$ can be calculated according to the following empirical linear relation which correlates the

solvation free energies of nonpolar solutes with their solvent accessible area (16,17)

$$\Delta G_{npol} = \sum \sigma A \qquad (16)$$

where A is the solvent accessible area (in $Å^2$). σ is the empirical solvation parameter. We used σ = 5 cal/mol^{-1}Å^{-2} in this work. The solvent accessible areas were calculated with Connolly's MS program (18). For $\Delta G_{L_1 \rightarrow L_2,p}$, this involves the calculation of the solvation free energy difference between L_1•P and L_2•P. Direct and accurate calculation of this difference is difficult because the solvation free energies of the two complexes are large numbers and the estimated difference may have a large error. As a first order approximation, we estimated $\Delta G_{L_1 \rightarrow L_2,p}$ as the difference in interaction energy between L_1 and L_2 with the protein and the solvent in energy minimized conformations of the complexes. The polar contribution to $\Delta G_{L_1 \rightarrow L_2,p}$ was estimated as the difference in electrostatic interaction energies between the two ligands with the protein and the solvent, which were calculated with equation 15 by summing over the atoms of L_1 or L_2. The nonpolar contribution to $\Delta G_{L_1 \rightarrow L_2,p}$ was estimated as the difference in van der Waals interaction energy between the two ligands with the protein and the solvent.

The GB/SA method uses a similar relation for calculation of the nonpolar contribution to the solvation free energy. For the electrostatic part, it uses the so-called generalized Born approximation which express the interactions between the fixed charges and their interactions with the solvent as a sum of pairwise interactions (13). The generalized Born approximation is only valid for the Poisson equation which correspond to κ=0 in equation 14. Since it has not been tested extensively on protein systems, we only used it to calculate the solvation free energies of the HIV-1 RT inhibitors.

Methods

1. T4 lysozyme

The models of the folded and unfolded states. The T4 lysozyme we analyzed is a mutant that has an unnatural amino acid at position 133, S-2-amino-3-cyclopentylpropanoic acid (Cpe) (19), which differs from an alanine residue in that it has a cyclopentyl group attached at C_β on the side chain. Our previous simulation study indicated that replacement of the original Leu at 133 with Cpe will better stabilize the enzyme than with 19 other natural amino acids, a prediction which was confirmed experimentally (19).

The structure of the mutant T4 lysozyme with a Cpe was obtained by model building using standard geometries based on the structure of the wild type T4 lysozyme (11) from the Brookhaven Protein Data Bank and was subjected to energy minimizations before the molecular dynamics simulations. An 18 Å of cap TIP3P water (20) molecules centered around the $C_{\varepsilon 2}$ atom of Cpe133 was used in the simulations on the enzyme. 17 counterions (Na^+ or Cl^-) were added to keep the whole system neutral. Only residues within the sphere and the cap water molecules, which consists of about 1600 protein and counterion atoms and 260 water molecules, were allowed to move in the molecular dynamics simulations. The cap water molecules were kept from escaping by a weak repulsive potential (1.5 kcal/mol) at the surface of the sphere.

The unfolded state of the enzyme was represented by a terminally blocked solvated dipeptide, Ace-X-NMe, in which Ace and NMe are the acetyl and N-methylamide groups respectively and X is the unnatural amino acid residue. The backbone of the dipeptide was chosen to be in the extended state ($-180° < \phi < 0°$,

$0°<\psi<180°$). The side chain dihedral χ (N-C_α-C_β-C_γ) was chosen to be around 180°. The dipeptide was placed at the center of a box of 30 × 30 × 30 $Å^3$ filled with TIP3P water molecules under standard conditions. The number of water molecules is about 760. It should be noted that although different boundary conditions were used for the folded and unfolded states, because only $\Delta\Delta G$ contribute to the stability difference of the protein (equation 2), the error due to different boundary conditions is most likely canceled. Especially, the mutations we studied are nonpolar, in which the major contributions are short-range interactions, and we used the same cutoff radius (see below) in the simulations of the folded and unfolded states.

Force field parameters. The all atom force field developed by Cornell *et al* (21) were used. The atomic charges of the unnatural amino acid were obtained by fitting the electrostatic potential around the dipeptide model using the RESP method (22). The electrostatic potential was obtained by a single point *ab initio* quantum mechanical calculation using Gaussian94 (23) with 6-31G* basis set on a geometry generated by energy minimization with the AM1 method.

MD simulations. The MD simulations of PROFEC and CMC/MD were performed with the SANDER module of the AMBER4.1 program (24). The MD simulations of FED and TI were performed with the GIBBS module. Each simulation was performed with 2 fs time step, 8 Å cutoff radius and restrained temperature around 300K (25). The bond lengths were constrained by the SHAKE algorithm (26). For the simulations in water, periodic boundary conditions were applied and the pressure was controlled at 1 atm (25). The SETTLE algorithm was used to speed up the calculations on water molecules (27).

FED and PROFEC calculations. The free energy derivatives (FED) and the PROFEC contours were calculated from 300 ps and 100 ps MD simulations in the enzyme and in water, respectively. The $C_{\varepsilon 2}$ of Cpe133 was used as the origin of the grid. The two hydrogens attached to $C_{\varepsilon 2}$ were used to define the x-axis and the xy-plane. The parameters of the Lennard-Jones test particle are R*=2.0 Å and ε=0.15 kcal/mol, which are close to the van der Waals parameters of a tetrahedral carbon atom. The PROFEC results were visualized with UCSF MidasPlus (28) through a special delegate program written by R. J. Radmer.

Free energy calculations (TI). Each calculation was performed from λ=0→1 (the forward change) and λ→0 (the backward change). The average of the two results and their absolute difference were taken as the estimated ΔG and the hysteresis. The simulation time for each change ranges from 164 ps to 504 ps and the number of windows are from 41 to 126 (see Table II). For each window, the first 2 ps simulation was used as equilibration and the following 2 ps was used as sampling.

2. HIV-1 RT and its TIBO inhibitors

Force field parameters for the TIBO derivatives. Van der Waals (VDW) parameters of the chlorine atoms were taken from parameters used for chloroform (29) and the parameters for the sulfur atom (VDW, bond, angles, dihedrals and improper dihedrals) were adopted from a parameterization of thiobiotin (30). We used both the conformation of 8Cl-TIBO (R86183, see Table III) in complex with HIV-1 RT (31) as well as the A-form of the crystal structure of 9Cl-TIBO (R82913) (32) to estimate the partial atomic charges of 8Cl-TIBO, 9Cl-TIBO and unchlorinated TIBO (R82150). The two respective conformers were geometrically optimized using Gaussian94 (23) at the STO-3G level, each followed by a calculation of the electrostatic potential with the 6-31G* basis set. Atomic partial charges of the TIBO derivatives were fitted to the electrostatic potentials around the two structures using

the RESP method (22). Comparison of the partial charges evaluated from the two conformers individually showed a very small difference and we therefore evaluated the partial charges of the remaining TIBO derivatives (Table III) using only the TIBO conformer of 8Cl-TIBO in HIV-1 RT.

Setup and equilibration of HIV-1 RT in complex with 8Cl-TIBO. Unresolved residues (modeled as alanines) as well as hydrogens were added to the 3.0 Å resolution crystal structure of 8Cl-TIBO in HIV-1 RT (31). After a short minimization of the hydrogens *in vacuo*, the complex was hydrated by immersing it in a 55 Å radius sphere of TIP3P-water (20). The solvent sphere and the protein-inhibitor complex were minimized to let the protein relax in an aqueous environment. All water molecules beyond the first hydration shell (i.e. at a distance > 3.5 Å from any protein atom) were then removed and counterions (11 Cl^-) were added to achieve electroneutrality. Protein residues with any atom closer than 12 Å from 8Cl-TIBO were chosen to be flexible in the simulations and all protein residues, water molecules and counterions further than 15 Å from any flexible residue were deleted. A 20 Å radius spherical cap of TIP3P-water, including the hydrating water molecules within the sphere from the previous step, was centered on TIBO and equilibrated for 50 ps at 300 K. The protein, 8Cl-TIBO and the hydrating water molecules outside the water cap were then kept rigid. Thereafter, the flexible residues (as defined above) and 8Cl-TIBO together with the cap of water molecules were then heated (50 ps) and equilibrated for 300 ps at 300 K. The simulations were carried out with the SANDER module of AMBER 4.1 (24) using the Cornell *et al.* force field (21). We applied a dual cutoff of 9 and 13 Å, respectively, where energies and forces due to interactions between 9 and 13 Å were updated with the same frequency as the non-bonded list, i.e., every 20 time steps. A time step of 2 fs was used and all bonds were constrained with the SHAKE algorithm (26). The temperature was maintained using the Berendsen method (25), with separate couplings of the solute and solvent to the heat bath because the relaxation times of the solute and solvent may be different (33).

Setup and equilibration of 8Cl-TIBO in solution. As starting conformer we chose the A-form from the crystal structure of 9Cl-TIBO (32), with a substitution of the atoms at positions 8 and 9. 8Cl-TIBO was then immersed in a box of TIP3P water with dimensions 34 × 33 × 29 $Å^3$. Keeping the inhibitor rigid, the water molecules were equilibrated at constant pressure for 100 ps. The TIBO atoms were then released and the system was equilibrated for 200 ps, using the same dual cutoff and time step as for 8Cl-TIBO in HIV-1 RT.

Adaptive CMC/MD. The method was applied to 8 different TIBO derivatives, shown in Table III. Each inhibitor was positioned in the equilibrated HIV-1 RT - 8Cl-TIBO complex (see above), by substituting and/or deleting atoms in 8Cl-TIBO. The inhibitors were then allowed to relax in the binding pocket by individually minimizing them, keeping everything but the inhibitor rigid. Due to problems with the SHAKE algorithm during the MD steps, the time step was reduced to 1.5 fs and one MC step was performed every 20 MD time steps. We applied the adaptive CMC/MD method for two sets of inhibitors, the $\Delta G_{off,i}$'s (see the "Theory" section) were iteratively adjusted every 500 MC steps for set 1 and we shortened that interval to every 125 MC steps for set 2. The free energies of solvation for the TIBO derivatives were estimated from GB/SA (13) calculations, using the program MacroModel/BatchMin, version 4.5 (34). For these calculations, we used our RESP derived charges on the derivatives, which were minimized *in vacuo* prior to the calculations.

PB calculations. The 8 different TIBO - HIV-1 RT systems were further minimized, now with flexible residues, water molecules and counterions as in the MD

simulations (see above). All water molecules and counterions were then removed and the PB calculations were carried out with the latest Delphi package (35,36), using a dielectric constant ε=2 for both the protein and the inhibitors and with a ionic strength of 0.13 M. For estimations of $\Delta G_{L1 \rightarrow L2,s}$ (equation 3), the same structures as for the GB/SA calculations (above) was used, with ε=2 and an ionic strength of 0.13 M.

Free energy calculations (TI). For these calculations, which were carried out with the GIBBS module of AMBER4.1, we applied the same parameters and protocol as for the MD simulations (above). Starting with the equilibrated systems of 8Cl-TIBO in HIV-1 RT and in solution, respectively, the 8-chloro atom was perturbed into a hydrogen (R82150), using a window size ($\Delta\lambda$) of 0.02 (i.e. 51 windows in the λ-interval [0,1]). For TIBO in solution each window was equilibrated for 2 ps prior to a data collection time of 5 ps per window. The equilibration/data collection times for TIBO in HIV-1 RT were 3 ps and 8 ps, respectively.

Results

1. The stability of T4 lysozyme

Free energy derivatives and PROFEC. The free energy derivatives with respect to VDW radius (R*) of eight hydrogen atoms on the cyclopentyl ring of Cpe were calculated (Table I). The configurations of the hydrogens are defined as either pro-α or pro-β. The pro-α hydrogen is on the opposite face of the cyclopentyl ring from the C_β atom; the pro-β hydrogen is on the same face as the C_β atom. This definition of configurations is similar to that of the anomers of sugars. From Table 1, one sees that the free energy derivatives of HD12 (pro-α), HE12 (pro-β), HE21 (pro-α) and HD22 (pro-β) are negative, indicating that introducing some VDW group larger than hydrogen on either one of these sites may stabilize the protein. Because the free energy derivative of HE21 is the lowest, we focused our analyses around CE2 where HE21 is attached.

Table I. The free energy derivatives of Cpe (kcal/mol)[a]

Atom	dG/dR*,prot	dG/dR*,soln	Δprot-soln	configuration
HD11	7.7	6.1	1.6	pro-β
HD12	3.9	6.4	-2.5	pro-α
HE11	6.7	2.5	4.2	pro-α
HE12	4.5	6.3	-1.8	pro-β
HE21	2.4	6.3	-3.9	pro-α
HE22	7.1	2.5	4.6	pro-β
HD21	6.8	3.7	3.1	pro-α
HD22	6.4	8.9	-2.5	pro-β

[a] The free energy derivatives were obtained by 300 ps MD simulations in the enzyme and in water.

The PROFEC contour of zero VDW potential with $C_{\varepsilon 2}$ of Cpe as the origin is shown in Figure 1. Interestingly, the contour has the shape of a vase; its mouth faces the cavity and its neck embraces $C_{\varepsilon 2}$. The contour agrees with the free energy derivatives in that there is much more space for introducing a group at HE21 (negative derivatives) than at HE22 (positive derivatives). A natural proposal is the introduction of a methyl group at HE21 in the α configuration. In the following, we refer to this modification as Cpe→α-Mcpe and refer to the introduction of the methyl group at HE22 in the β configuration as Cpe→β-Mcpe. Figure 2 shows the superimposition of a methyl group at HE21 and at HE22. Obviously, the methyl

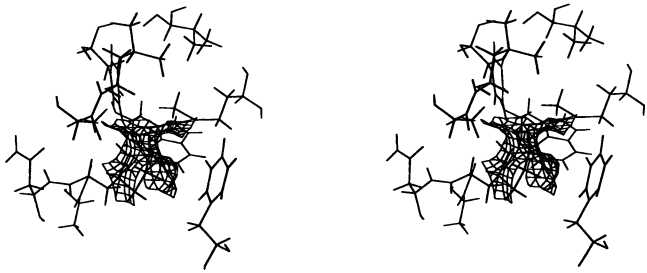

Figure 1. The PROFEC contour centered at $C_{\varepsilon 2}$ of Cpe133.

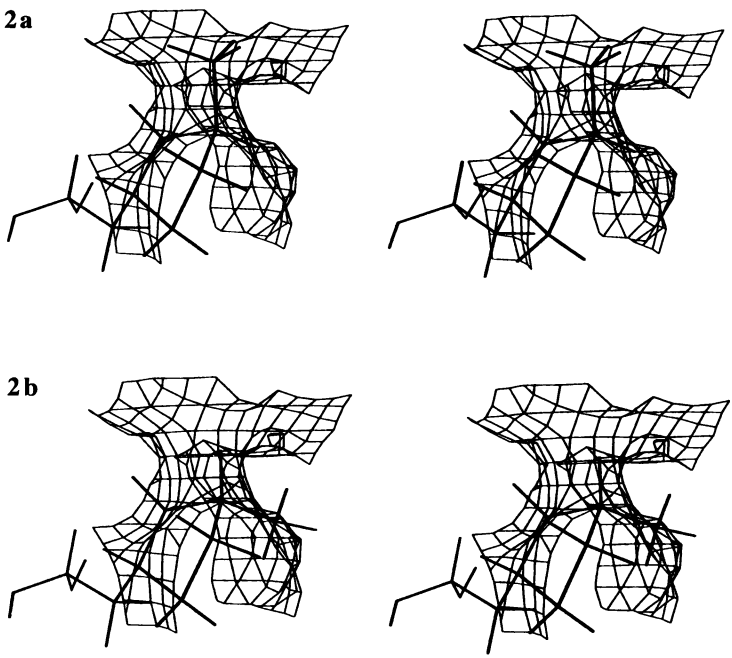

Figure 2. The superimposed structures of a methyl group at HE21 in the α configuration (a), and at HE22 in the β configuration (b).

group introduced at HE21 can fit very well in the cavity while the methyl group introduced at HE22 will collide with the wall of the vase belly. This suggests that adding a methyl group at HE21 in α configuration will improve the stability of the enzyme with a Cpe at position 133.

Table II. The ΔG of Cpe→Mcpe in water and in the enzyme (kcal/mol)[a]

medium	config.	time	forward ΔG	backward ΔG	average ΔG
protein	α	164 ps	-1.92	-2.20	-2.06
protein	α	324 ps	-2.70	-2.59	-2.65
protein	α	488 ps	-2.64	-2.56	-2.60
protein	β	164 ps	-0.24	0.30	0.03
protein	β	324 ps	-0.66	-0.22	-0.44
water	α	164 ps	-1.18	-0.94	-1.06
water	α	324 ps	-0.51	-0.36	-0.44
water	α	504 ps	-0.86	-0.85	-0.86
water	β	324 ps	-0.66	-0.22	-0.44

[a] Each window has 2 ps / 2 ps for equilibration and sampling. The 164, 324 and 504 ps simulations have 41, 81 and 126 windows respectively. The 488 ps simulation: 41 windows for the electrostatic contribution and 81 windows for the VDW contribution.

Free energy calculations (TI). We have calculated the free energy changes for introducing a methyl group in the α (at HE21) and β (at HE22) configurations in water and in the T4 lysozyme (Table 2). One sees that the calculated average ΔGs are not sensitive to the length of simulation time and the hystereses are ≤ 0.5 kcal/mol. For the simulations in the enzyme, we used the results of 324 ps for forward or backward change as the estimated free energy changes: ΔG (Cpe→α-Mcpe) = -2.65 kcal/mol and (Cpe→β-Mcpe)= -0.44 kcal/mol. For the simulations in water, Cpe→α-Mcpe and Cpe→β-Mcpe should give similar results as is seen from the results of 324 ps for both configurations. Since the average ΔG of Cpe→α-Mcpe fluctuates around -0.8 kcal/mol for different lengths of simulation time, -0.8 kcal/mol is taken as the estimated average ΔG in water. Based on these results, Cpe→α-Mcpe will stabilize the enzyme by 1.8 kcal/mol while Cpe→β-Mcpe will destabilize the enzyme by 0.4 kcal/mol. Therefore, the free energy calculations support the predictions of PROFEC.

2. Relative binding affinity to HIV-1 RT for a series of TIBO inhibitors

Adaptive CMC/MD and PB calculations. In Table IV, we present the relative ranking of the TIBO derivatives found in the CMC/MD and PB calculations. The numerical values of the relative free energies and full details will be presented elsewhere. The three best binding TIBO derivatives were also ranked as the best with the adaptive CMC/MD method. In this context, it should be noted that deviations in the rank order from the experimental results also might be due to an imperfect agreement between HIV-1 RT activity and binding affinity, caused by differences in cell penetration ability and metabolic stability between the TIBO derivatives. The experimental EC_{50} values (Table III) of the three next inhibitors, ranked 4 to 6, are very close to each other and they are also ranked 4 to 6 with the CMC/MD method. In the PB calculations, $\Delta\Delta G_b$ for R84914 has significantly been overestimated and its rank order is thus too favorable, which shifts the rank order of these three derivatives downwards. Omitting R84914 in the ranking according to PB calculations, the agreement with experiment is very good and the order is reversed for only two inhibitors (R87027 and R84674). Finally, the two inhibitors with the highest $\Delta\Delta G_b$'s

(R84914 and R80902) are also correctly ranked with both methods. The poor binding of R80902, which has an oxygen instead of a sulfur atom at position 2 (Table III), is mainly due to its favorable free energy of solvation (ΔG_{solv}). According to the GB/SA method, we estimated ΔG_{solv} for R80902 around 2 kcal/mole lower than for the other TIBO derivatives. Both the adaptive CMC/MD and the PB calculations have very reasonable average errors in the $\Delta\Delta G_b$ compared to experiment (0.71 kcal/mole and 1.12 kcal/mole, respectively).

Table III. The selected set of TIBO derivatives (see also Figure 3).

Compound	R_1	R_2	R_3	R_4	EC_{50} [a]
R86183	8-Cl	S	H	$-CH_2-CH=C(CH_3)_2$	4.6
R82913	9-Cl	S	H	$-CH_2-CH=C(CH_3)_2$	33
R82150	H	S	H	$-CH_2-CH=C(CH_3)_2$	44
R80902	H	O	H	$-CH_2-CH=C(CH_3)_2$	4200
R84674	8-CH_3	S	H	$-CH_2-CH=C(CH_3)_2$	14
R84963	H	S	-CH_3 (*trans*)[b]	$-CH_2-CH=C(CH_3)_2$	39
R84914	H	S	-CH_3 (*cis*)[b]	$-CH_2-CH=C(CH_3)_2$	790
R87027	8-Cl	S	H	$-CH_2-CH=C(CH_2CH_3)_2$	5.1

[a] ref. 12.
[b] relative stereochemistry of the methyl groups at positions 5 and 7.

Table IV. Rank order of $\Delta\Delta G_{b,i}$ [$=\Delta G_{b,i}-\Delta G_{b,R86183} - (\Delta G_{solv,j}-\Delta G_{solv,R86183})$] for the TIBO derivatives from adaptive CMC/MD, PB calculations and experiments (12).

Compound	adaptive CMC/MD rank	PB-calculations rank	experimental $\Delta\Delta G_{b,i}$ [a]	experimental rank
R86183	**3**	**1** (1)[b]	0	**1** (1)
R87027	**1**	**4** (3)	0.06	**2** (2)
R84674	**2**	**2** (2)	0.66	**3** (3)
R82913	**6**[c]	**5** (4)	1.17	**4** (4)
R84963	**4**[c]	**6** (5)	1.27	**5** (5)
R82150	**5**[c]	**7** (6)	1.34	**6** (6)
R84914	**7**[c]	**3**	3.05	**7**
R80902	**8**	**8** (7)	4.04	**8** (7)

[a] calculated from $\Delta\Delta G_{b,i}=-RT\ln(EC_{50,i}/EC_{50,R86183})$, T=298 K, R=1.986 cal/K/mole.
[b] The ranking in parenthesis is when excluding R84914.
[c] Average rank order from two sets of MC/MD-runs. Set 1 consists of the following derivatives with known experimental binding affinities: R86183, R81913, R84963, R82150, R84914 and R80902 and was run for 450 ps. Set 2 consists of R86183, R87027, R84674, R81913, R84963, R82150 and R84914 and was run for 560 ps.

Free energy calculations (TI). Perturbing 8Cl-TIBO (R86183) into 8-H TIBO (R82150), yields a relative binding free energy (-1.9±0.5) in close agreement with experiment (12) (-1.34). The perturbation in water is well converged as seen from the close values from the forward and reverse runs (Table V). However, in spite of the extended equilibration/data collection times (see "Methods") for the inhibitor in HIV-1 RT, these perturbations show a considerable hysteresis. Decomposition of the free energies shows that the van der Waals contribution is responsible for the slow convergence, differing by 1.4 kcal/mole between the forward and reverse runs. The same decomposition also reveals that the better binding of R86183 compared with R82150 is almost entirely due to a more favorable van der Waals contribution to the

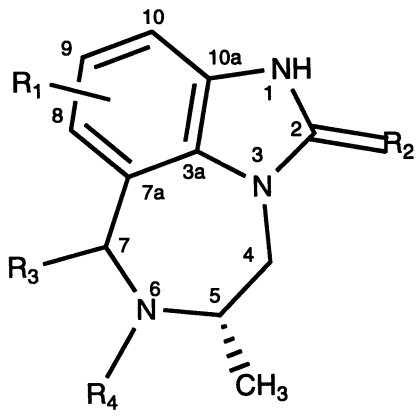

Figure 3. TIBO derivatives. The substitutions are explained in Table III.

free energy for the former inhibitor. Considering that the relatively non-polar TIBO is positioned in a hydrophobic and aromatic binding pocket, it is not surprising that van der Waals forces dominate the binding interaction.

Table V. Free energy calculations, R82150→R86183. The numbers are in kcal/mole and error estimates are within parenthesis.

	$\Delta G_{L1\rightarrow L2}$ [a]			$\Delta G_{L1\rightarrow L2,P}$ [a]			$\Delta\Delta G_b$ [a]
	forward	reverse	average	forward	reverse	average	
ΔG_{tot} [c]	1.53	1.69	1.61	-0.78	0.13	-0.32	-1.9 (0.5)[b]
ΔG_{elec}	2.52	2.70	2.61	2.71	2.20	2.45	-0.16 (0.26)
ΔG_{vdw}	0.70	0.70	0.70	-1.65	-0.27	-0.96	-1.7 (0.7)
ΔG_{pmf}	-1.69	-1.71	-1.70	-1.83	-1.73	-1.73	-0.03 (0.11)

[a] see equations 3-4.
[b] experimental value: -1.34 kcal/mole (12).
[c] ΔG_{tot} is the total free energy, ΔG_{elec} and ΔG_{vdw} are the electrostatic and van der Waals contributions, respectively, and ΔG_{pmf} is the bond potential of mean force contribution.

Conclusions

We have studied the stability of T4 lysozyme and the binding free affinities of the RT TIBO inhibitors by several approximate and efficient free energy calculation methods and the rigorous TI method. For T4 lysozyme, the FED and PROFEC are useful for suggesting promising sites and candidate modifications to improve the stability of the protein. The results were supported by the TI calculations. The combination of FED and PROFEC appears to be very efficient and effective in making predictions. By using FED, one can quickly find the promising sites and then PROFEC can be used on these sites to suggest candidate modifications. Compared with TI, FED and PROFEC calculations are much faster: only a few hundred ps of MD simulation is required to obtain reasonable results. The information generated by FED and PROFEC can be quite comprehensive because in many cases it is appropriate to calculate the free energy derivatives for many atoms in a single MD simulation and use the same trajectories to construct the PROFEC contours. For the RT inhibitors, both the chemical MC/MD method and the PB calculations are able to rank TIBO derivatives in good agreement with experiment. Since these methods are quite different in nature, each with their own set of approximations, they serve as a good complement to each other. That is, if both methods predict the same rank order, the reliability of this prediction will significantly increase. Considerable simulation times (1.1 ns) were required for a reasonable estimate of the relative free energies of just *two* TIBO derivatives bound to RT in our TI calculations. In that perspective, the two more approximate methods worked surprisingly well and produced valuable information with substantially less effort and time. However, one sees that by decomposing of free energy differences into components, the TI calculations can provide important insights into the nature of TIBO-RT interactions.

In summary, the approximate methods used here are able to make quite reasonable predictions with much less computational cost than the rigorous TI calculations. Therefore, they are valuable at least as fast screening tools in the last stages of structure-based drug refinement and protein engineering. The rigorous free energy calculations (TI, or FEP) are computationally expensive, but can be used to support the predictions of the approximate methods and help to gain insights into the nature of molecular interactions. We envision a hierarchy of computational methods that can be applied to these sorts of problems. First, FED and PROFEC can be used to

suggest lead plausible modifications for lead optimizations. Second, CMC/MD or PB calculations can be used to rank the binding affinities or stabilities of many ligands or mutants in a short time. Finally, traditional free energy methods (TI, or FEP) can be used to analyze a few particularly interesting cases.

Acknowledgment

L. Wang and M. A. L. Eriksson contributed equally to this work. This work was supported by the NIH (GM-29072, P. A. Kollman; GM-56531, P. Ortize de Montellano; GM-56609, E. Arnold). We acknowledge the use of the facilities of UCSF computer graphics laboratory, supported by NIH P41-RR01081 (T. Ferrin, P.I.). Mats Eriksson acknowledges a postdoctoral grant from the Swedish Natural Research Council (NFR).

References

(1) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395-2417.
(2) Åqvist, J.; Medina, C.; Samuelsson, J.-E. *Protein Eng.* **1994**, *7*, 385-391.
(3) Radmer, R. J.; Kollman, P. A. *J. Comput.-Aided Mol. Design.* **1997**, accepted for publication.
(4) Pitera, J.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, submitted for publication.
(5) Kong, X.; Brooks III, C. L. *J. Chem. Phys.* **1996**, *105*, 2414-2423.
(6) Gerber, P. R.; Mark, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Design* **1993**, *7*, 305-323.
(7) Cieplak, P.; Pearlman, D. A.; Kollman, P. A. *J. Chem. Phys.* **1994**, *101*, 627-633.
(8) Pang, Y. P.; Kollman, P. A. *Perspect. Drug Discov. Design* **1995**, *3*, 106-122.
(9) Shen, J.; Quiocho, F. A. *J. Comput. Chem.* **1995**, *16*, 445-448.
(10) Shen, J.; Wendoloski, J. *J. Comput. Chem.* **1996**, *17*, 350-357.
(11) Weaver, L. H.; Matthews, B. M. *J. Mol. Biol.* **1987**, *193*, 189-199.
(12) Pauwels, R.; Andries, K.; Debyser, Z.; Kukla, M. J.; Schols, D.; Breslin, H. J.; Woestenborghs, R.; Desmyter, J.; Janssen, M. A. C.; de Clercq, E.; Janssen, P. A. J. *Antimicrob. Agents. Chemother.* **1994**, *38*, 2863-2870.
(13) Still, C. W.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.
(14) Metropolis, N. R.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087-1092.
(15) Sharp, K.; Honig, B. *Ann. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301.
(16) Hermann, R. B. *J. Phys. Chem.* **1971**, *76*, 2754.
(17) Nozaki, Y.; Tanford, C. H. *J. Biol. Chem.* **1971**, *246*, 2211.
(18) Connolly, M. L. *Science* **1983**, *221*, 709-713.
(19) Mendel, D.; Ellman, J. A.; Chang, Z.; Veenstra, D. L.; Kollman, P. A.; Schultz, P. G. *Science* **1992**, *256*, 1798-1802.
(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.
(21) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Mertz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-1597.
(22) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
(23) 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.; Montegomery, J. A.; Rahavachari, K.; Al-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.; Pople, J. A. *Gaussian 94, Revision B.3* **1995**, Gaussian Inc., Pittsburg, PA.

(24) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; III, T. E. C.; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER, version 4.1* **1995**, University of California at San Francisco.
(25) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
(26) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327-341.
(27) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952-962.
(28) Ferrin, T. E.; Huang, C. C.; Jarvis, L. E.; Langridge, R. *J. Mol. Graph.* **1988**, *6*, 13-27.
(29) Fox, T.; Thomas, B. E.; McCarrick, M.; Kollman, P. A. *J. Phys. Chem.* **1996**, *100*, 10779-10783.
(30) Miyamoto, S.; Kollman, P. A. *Proteins* **1993**, *16*, 226-245.
(31) Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A. J.; Hughes, S. H.; Arnold, E. *Struct. Biol.* **1995**, *2*, 407-415.
(32) Liaw, Y. C.; Gao, Y. G.; Robinson, H.; Wang, A. H. J. *J. Am. Chem. Soc.* **1991**, *113*, 1857-1859.
(33) Guenot, J.; Kollman, P. A. *Prot. Sci.* **1992**, *1*, 1185-1205.
(34) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Am. Chem. Soc.* **1990**, *11*, 440-467.
(35) Gilson, M. K.; Sharp, K. A.; H., H. B. *J. Phys. Chem.* **1988**, *9*, 327-335.
(36) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144-1149.