ELSEVIER

Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems

John Marelius, Karin Kolmodin, Isabella Feierberg, and Johan Åqvist

Department of Cell and Molecular Biology, Biomedical Centre, Uppsala University, Uppsala, Sweden

A new molecular dynamics program for free energy calculations in biomolecular systems is presented. It is principally designed for free energy perturbation simulations, empirical valence bond calculations, and binding affinity estimation by linear interaction energy methods. Evaluation of ligand-binding selectivity and free energy profiles for nucleophile activation in two protein tyrosine phosphatases as well as absolute binding affinity estimation for a lysine-binding protein are given as examples. © 1999 by Elsevier Science Inc.

Keywords: molecular dynamics, free energy perturbation, empirical valence bond method, protein tyrosine phosphatases

INTRODUCTION

Computer simulation and modelling methods have become important tools in many areas of chemistry and biology. With the rapid increase in computer power it is nowadays possible to carry out quantitatively meaningful calculations on large systems such as proteins. The focus of computer simulations in the biomolecular field has also shifted somewhat from the more descriptive type of earlier work to quantitative calculations of energetics for some of the key problems in biochemistry. Hence, we find topics such as protein folding, enzyme catalysis, ion channel permeation, and ligand binding at the top of the list of applications. It has also been widely realised that the calculation of free energies often provides the link between structure and function.

Molecular dynamics (MD) simulation can be used to sample the thermally accessible region of conformational space of a microscopic model of a molecular system. From the ensemble of sampled structures and their associated potential energies (given by the force field or molecular mechanics potential energy function) it is, in principle, possible to calculate free energies. Quantities such as binding free energies, solvation free energies, and activation free energies are particularly interesting to calculate because they are the direct result of thermodynamic or kinetic experiments. It is thus possible both to quantitatively verify calculated results against experimental data and to make predictions that can be tested experimentally.

Our motivation for writing a new molecular dynamics program is to have a tool tailored for some specific kinds of free energy calculations, namely, (1) free energy perturbation (FEP) simulations,[1,2] (2) empirical valence bond (EVB) calculations[3,4] of reaction free energies, and (3) linear interaction energy (LIE) calculations[5–7] of receptor–ligand binding affinities.

Free energy perturbation

The free energy associated with transforming a molecular system from one state i to another state j, described by the potentials V_i and V_j, respectively, can be calculated using the perturbation formula represented by Eq. (1)[1,2]:

$$\Delta G_{i\rightarrow j} = -RT \ln\langle \exp[-(V_j - V_i)/RT]\rangle_i \qquad (1)$$

where $\langle \cdot \rangle_i$ denotes the ensemble average calculated using the potential V_i. In practice this average is evaluated using MD or Monte Carlo methods. Although Eq. (1) is exact, it is only practically useful when the states are so similar that configurational sampling using V_i also samples relevant (i.e., low-energy) configurations on V_j. This condition is not fulfilled if

Color Plates for this article are on page 261.

Address reprint requests to: Dr. Johan Åqvist, Department of Cell and Molecular Biology, Uppsala University, Biomedical Centre, Box 596, SE-751 24 Uppsala, Sweden. phone: +46 18 174109, fax: +46 18 536971, e-mail: aqvist@xray.bmc.uu.se

Journal of Molecular Graphics and Modelling 16, 213–225, 1998

655 Avenue of the Americas, New York, NY 10010

1093-3263/99/$–see front matter
PII S1093-3263(99)00012-1

the two states differ significantly, e.g., in chemical bonding arrangements or in the number or type of atoms involved. By introducing a set of intermediate mapping potentials (unphysical states) as linear combinations between V_i and V_j [see Eq. (2)], sufficient sampling can be attained, provided that these intermediate potentials are sufficiently closely spaced.

$$V_m = \sum_{i=1}^{N} \lambda_i^m V_i \quad \text{and} \quad \sum_{i=1}^{N} \lambda_i^m = 1 \tag{2}$$

Here, V_m is the effective mapping potential formed as a linear combination over all the N states using the mapping vector $\boldsymbol{\lambda}_\mathrm{m} = (\lambda_1^\mathrm{m}, \ldots, \lambda_\mathrm{N}^\mathrm{m})$. The total free energy change is then calculated as a sum over all steps between the end-point potentials, as shown in Eq. (3):

$$\Delta G_{\mathrm{initial}\rightarrow\mathrm{final}} = \sum_{i=1}^{n-1} \Delta G_{i\rightarrow i+1} \tag{3}$$

where n is the total number of λ steps (mapping potentials) and each term of the sum is given by Eq. (1). This procedure can be used in combination with appropriate thermodynamic cycles (see below) to calculate, e.g., solvation free energies of small molecules and for predicting effects of mutations in proteins or ligands.

Empirical valence bond method

The empirical valence bond method is an efficient technique for modelling chemical reactions in solutions and enzymes.[3,4] It is, in fact, easily combined with FEP and thus provides one of the few efficient simulation approaches available for calculating free energy profiles (surfaces) of chemical reactions in solvated systems.

The EVB method describes a given reaction in terms of a number of resonance structures, or valence bond (VB) states, that represent different bonding arrangements and charge distributions. The energy of a particular VB state i, corresponding to the diagonal element H_{ii} of the EVB Hamiltonian, is given by Eq. (4):

$$H_{ii} = \epsilon_i = V_{\mathrm{bond}}^{(i)} + V_{\mathrm{angle}}^{(i)} + V_{\mathrm{torsion}}^{(i)} + V_{\mathrm{nonbond}}^{(i)} + V_{\mathrm{rs}}^{(i)} + V_{\mathrm{ss}}^{(i)} + \alpha^{(i)} \tag{4}$$

The first term denotes the sum of Morse potentials describing the bonds in the reacting molecular fragments. The second to fourth terms denote the bond angle bending contribution and nonbonded electrostatic and van der Waals interactions between the reacting groups. The fifth term is the interactions between the reactants and the surrounding protein and solvent, and the sixth term represents the internal potential energy of the solvent/protein system. The constant term $\alpha^{(i)}$ is the gas-phase energy of VB state i when all fragments are at infinite separation. For every geometric configuration of the system the actual ground-state energy, E_g, is obtained by mixing the VB states using off-diagonal elements (H_{ij}) in the EVB Hamiltonian and solving the corresponding secular equation.

The free energy on the ground-state potential surface can then be evaluated as a function of a reaction coordinate X from Eq. (5):

$$\Delta G(X) = \Delta G(\boldsymbol{\lambda}_\mathrm{m}) - RT \ln\langle \delta(X' - X) \cdot \exp(-[E_\mathrm{g}(X') - \epsilon_\mathrm{m}(\mathrm{X}')]/\mathrm{RT})\rangle_\mathrm{m} \tag{5}$$

where ϵ_m is an FEP mapping potential of the type described above, i.e., $\epsilon_\mathrm{m} = \Sigma\, \lambda_i^\mathrm{m}\ \epsilon_i$. The δ function in practice translates to a discretisation of the reaction coordinate X into finite intervals. Data from simulations at several different λ steps may contribute to a given interval of X. Likewise, a simulation using a given value of λ will usually sample a range of X values. It is common practice to take the energy gap between two VB states, $\Delta\epsilon = \epsilon_i - \epsilon_j$ as the reaction coordinate X in Eq. (5). The energies sampled during the MD/FEP simulation are used to evaluate the ground state reaction free energy profile, using this combination of free energy perturbation and umbrella sampling.

The keystone of the EVB method is the possibility to calibrate the Hamiltonian against experimental data or quantum chemical calculations for relevant reference reactions in solution or gas phase. In practice, running EVB simulations is conceptually similar to FEP calculations and poses no difficulties. The tricky part is often to find, compile, and interpret the data required for calibration of the method for the given system. Once this is done, the calibration involves determining gas-phase energy differences $\Delta\alpha_{ij}$ as well as off-diagonal matrix elements H_{ij} between pairs of VB states, so that the EVB potential surface reproduces the reaction free energies and barrier heights for the reference reaction. Calibration against solution data involves simulations of uncatalysed reactions (assuming the same mechanisms) with the reacting fragments in water and fitting the preceding parameters so that calculated and observed free energies coincide. After calibration of the potential surface for the reference solution reaction, the relevant parameters are used without change for studying the corresponding reaction in an enzyme. Starting from a mechanistic hypothesis, one can thus calculate the free energy profile (potential of mean force) and compare with data from reaction kinetics experiments to test the hypothesis. A mechanistic model validated in this way can then be used to predict, e.g., the catalytic properties of related enzymes or the selectivity for different substrates.

Linear interaction energy method

Since FEP calculations of absolute binding free energies are computationally very demanding and often subject to convergence problems, there is great interest in other methods to predict binding affinities for drug design applications. One approach is "scoring" of receptor–ligand interactions in a single conformation using empirical scoring functions calibrated on a set of experimental complex structures. Among others, Böhm,[8] Eldridge et al.,[9,10] Head et al.,[11] Takamatsu and Itai,[12] and Jain[13] have presented such methods. Another type of approach that has attracted considerable attention is to try to approximate free energies from force field simulations[5,14,15] without using rigorous FEP. Our so-called linear interaction energy (LIE) method, which has been described in a series of publications,[7,16–20] involves simulations only of the two physical states (bound and free ligand) and can thus reduce both convergence problems and computational expense associated with adequate sampling along the whole transformation pathway. The binding free energy is estimated using averages of force field energies.[5] Related procedures have been reported by Paulsen and Ornstein,[21] Jones-Hertzog and Jorgensen,[6] and Gorse and Gready.[22]

The main idea of the LIE method is to consider contributions

from polar and nonpolar (all nonelectrostatic) interactions to the total binding energy separately. The polar part can be treated using the electrostatic linear response approximation, while the nonpolar contribution is calculated using an empirical formula calibrated against a set of experimental binding data. The calculated binding free energy is given by Eq. (6).

$$\Delta G_{\text{bind}} = \alpha(\langle V_{\text{l–s}}^{\text{LJ}}\rangle_{\text{bound}} - \langle V_{\text{l–s}}^{\text{LJ}}\rangle_{\text{free}}) + \beta(\langle V_{\text{l–s}}^{\text{el}}\rangle_{\text{bound}} - \langle V_{\text{l–s}}^{\text{el}}\rangle_{\text{free}}) \quad (6)$$

where α is the empirical scaling factor for nonpolar interaction energies, the term $\langle V_{\text{l–s}}^{\text{LJ}}\rangle$ represents the average Lennard–Jones interaction energies between ligand and surrounding (water + protein in the "bound" case, only water in the term with subscript "free"), β is the scaling factor for electrostatics, and $\langle V_{\text{l–s}}^{\text{el}}\rangle$ represents the average ligand-surrounding Coulomb interaction energies from the simulations of bound and free ligand. The linear response approximation states that $\beta = 1/2$. While this value produces accurate results for ionic compounds, lower β values should be used for neutral dipolar compounds.[7,23]

Long-range electrostatics

In applications focused on quantitative calculations of energetics, the treatment of long-range electrostatic interactions is of paramount importance. Truncation of nonbonded interactions beyond a cutoff radius, typically 10 Å or so, has been shown to produce severe artefacts for electrostatic energies.[17,19,24] Nevertheless, many biomolecular MD simulations are still carried out using such cutoffs, probably because no attempts are made to assess critically the reliability of calculated electrostatic energies in such cases. However, a number of approaches for treating long-range electrostatics are available, e.g., reaction field, multipole expansion, and lattice summation methods.[25] In our program we have implemented the so-called local reaction field (LRF) method,[26] which is a particularly efficient multipole expansion method for simulations of finite spherical systems. With only about 50% extra computer time this technique yields close agreement with infinite cutoff results.[17,26] It involves a Taylor expansion to the third order of the long-range electrostatic potential on each charge group of the system. This expansion is then updated at regular intervals (typically 25–50 time steps). This is quite different from the so-called "double time-step" treatment of electrostatics, in which the long-range forces are considered constant during the updating interval.

DESCRIPTION OF THE PROGRAM

Molecular dynamics simulation involves numerical integration of the equations of motion for a system of atoms that interact through an analytical potential energy function (force field). Q uses the leap-frog version of Verlet's integration algorithm [Eq. (7)]:

$$\mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}_i(t)}{m_i}\Delta t$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right)\Delta t \quad (7)$$

where $\mathbf{r}_i$, $\mathbf{v}_i$, and $\mathbf{F}_i$ are the position and velocity of particle i and the resulting force acting on it, respectively. The temperature of the system can be controlled by scaling the velocities according to the algorithm of Berendsen et al.[27]

Q is intended for free energy calculations in biomolecular systems solvated in a spherical droplet of explicit water molecules. Using a spherical boundary[28–30] makes it possible to limit the size of the simulated system, i.e., to focus the simulation on a smaller region such as a binding site, and also makes accurate treatment of long-range electrostatics rather inexpensive. A large solute molecule does not have to be included in its entirety within the simulation sphere but parts of it can protrude out of the sphere, where no nonbonded interactions are evaluated. For comparison, a corresponding simulation with periodic boundary conditions (PBC) would require that the entire molecule be enclosed in the periodic box together with a sufficiently thick layer of surrounding solvent. This leads to severalfold more particles to simulate, with a significant increase in computational cost. As an example, the simulation of protein tyrosine phosphatase 1B described below would have required for a PBC simulation a cubic box with sides of about 85 Å, compared with the 16-Å sphere used. Another critical issue with PBC is that the treatment of long-range electrostatics further increases the computational effort. While PBC simulations are adequate for many types of applications, our aim here has been to design an efficient tool for the types of applications mentioned above rather than write another general-purpose MD program.

Boundary conditions

In the development of Q, much attention has been devoted to the boundary restraints applied to water molecules near the sphere surface. Ideally, water molecules in the boundary zone should behave like they were part of a macroscopic system. With no restraining forces, water molecules would eventually evaporate from the system. A simple half-harmonic radial potential gives both a decreased density and an exaggerated polarisation in the boundary zone. These problems were addressed recently by Essex and Jorgensen[31] by introducing a radial restraining potential that not only pushes water molecules from the outside back into the sphere, but also includes a "boundary attraction" term that pulls molecules from inside the sphere toward the boundary. While their potential is a polynomial in r (the distance from the sphere centre) with coefficients calibrated to reproduce the density of bulk water, Q uses a combination of a Morse-type boundary attraction potential and a half-harmonic potential outside the sphere. The radial restraining potential can be written as in Eq. (8):

$$V_{\text{water}}(r) = \begin{cases} \frac{1}{2}k_{\text{rad}}(r - r_0)^2 - D_{\text{e}} & \text{if} \quad r > r_0 \\ D_{\text{e}}(1 - e^{a(r-r_0)})^2 - D_{\text{e}} & \text{otherwise} \end{cases} \quad (8)$$

where r is the distance from the sphere centre, k_{rad} is the force constant of the half-harmonic potential, D_{e} is the depth ("dissociation energy") of the Morse potential, and a is the exponential coefficient of the Morse term. r_0 is the target water radius r_{w} minus the average deviation distance from the minimum of the half-harmonic potential at the current temperature T: $r_0 = r_{\text{w}} - [(k_{\text{B}}T)/k_{\text{rad}}]^{1/2}$. The appropriate values for D_{e} and a that give a correct and uniform density throughout the sphere must be calibrated for each value of r_{w} (e.g., with a 15-Å sphere, the values are $D_e = 1$ kcal/mol and $a = 0.5$ Å^{-1}). On the basis of such calibrations at radii ranging from 12 to 30 Å, empirical functions giving D_{e} and a as functions of r_{w} were

designed. The radial restraining forces are applied only at the oxygen atoms of each water molecule.

The polarisation of water near the boundary is controlled by restraining the distribution of angles between the radial axis and the molecular dipole vector to the ideal distribution, which is given by $p(\theta) = \frac{1}{2} \sin \theta$. If the solute is not neutral, the distribution can be corrected by the net surface polarisation induced by a central charge. This procedure is essentially the same as that of the SCAAS[32] model as implemented in the Enzymix[33] program. However, we have found it more efficient to divide the restrained boundary region into three subshells, each obeying the desired distribution. The outermost, middle, and innermost shells are 0.5, 1.0, and 1.5 Å thick; the restrained border zone is thus 3 Å in total. Within each shell, the water molecules are at each time step sorted by the angle θ between the molecular dipole vector and the radial vector from the centre of the sphere. A harmonic potential term is then applied to rotate each molecule i towards its target value θ_i in the ideal distribution.

The restraining of solute atoms outside the boundary of the simulation sphere is more straightforward. No pairwise nonbonded interactions are calculated outside the sphere and the atoms there are held fixed at their desired positions by strong harmonic restraints (200 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2}). To avoid excessive strain across the sphere boundary, atoms in a buffer zone inside the boundary may also be restrained by harmonic potentials. An overview of the boundary restraints is given in Color Plate 1.

When comparing simulations in finite spherical systems of, say, a solvated receptor–ligand complex with a reference simulation of a ligand in water, such as in FEP or LIE calculations of free energies, it is critical that the two systems be matched with each other so that constant free energy contributions from the outside cancel. In particular, the systems must be equivalent with respect to the free energy associated with polarisation of the medium outside the simulation spheres.[17] The leading term of this energy is given in the continuum approximation by Born's formula $\Delta G_{\text{Born}} = -(1/4\pi\epsilon_0)(Q^2/2r)(1-1/\epsilon)$, where Q is the net charge of the microscopic system enclosed in a sphere of radius r, outside of which the dielectric constant is ϵ. This means that the two simulated systems should be of equal size and have equal net charge when free energies of charged molecules are to be calculated. Either counterions must be added to one of the systems (which has been shown earlier to slow down convergence[19]), or excess charge on the protein must be removed.[17,19] In any case, it is necessary to neutralise charged groups near the boundary of the simulation sphere since these in reality would be screened by a high dielectric constant. To fulfill these requirements, charged residues far from the ligand and near the boundary of the simulation sphere can be modelled as corresponding neutral dipolar groups, which are available in most force fields. For charged ligands, this procedure means that Coulomb interactions between the ligand and the neutralised charges of the protein are neglected. Since all these groups should be far from the ligand, the energy can be approximated by Coulomb's law [Eq. (9)] with a high dielectric constant characteristic of long-range electrostatic interactions in solvated proteins.[17]

$$\Delta G^{\text{el}}_{\text{corr.}} = \frac{1}{4\pi\epsilon_0} \sum_{\substack{\text{p} \in \text{neglected ionic sites} \\ \text{l} \in \text{ligand atoms}}} \frac{q_\text{p}\, q_\text{l}}{\epsilon\, r_{\text{p–l}}} \qquad (9)$$

Here, q_p is the formal (integer) charge of the neglected ionic group, q_l is the partial charge of the ligand atom, $r_{\text{p–l}}$ is the distance between the ligand atom and a central atom in the ionic group (e.g., NZ in lysine), and ϵ is the dielectric constant, typically 80.

Force fields

Q is not associated with any particular force field. The force fields are defined in parameter files, separate from the program, and the choice of force field is thus simply a matter of which parameter file to use. Any force field could be used with the program, as long as it shares the common functional form of Eq. (10):

$$\begin{aligned} V_{\text{pot}} = & \sum_{\text{bonds}} \frac{1}{2} k_\text{b} (r - r_0)^2 + \sum_{\text{angles}} \frac{1}{2} k_\theta (\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} K_\varphi [1 + \cos(n\,\varphi - \delta)] \\ & + \sum_{\substack{\text{improper} \\ \text{dihedrals}}} \frac{1}{2} k_\xi (\xi - \xi_0)^2 + \sum_{\text{atom pairs } i,j} \frac{1}{4\,\pi\,\epsilon_0} \\ & \cdot q_i\, q_j\, r_{ij}^{-1} + A_{ij}\, r_{ij}^{-12} - B_{ij}\, r_{ij}^{-6} \end{aligned} \qquad (10)$$

where V_{pot} is the total potential energy, k_b is a bond-stretching force constant, r is the distance between two bonded atoms, r_0 is the reference bond length, k_θ is an angle-bending force constant, θ is the angle between two bonds, θ_0 is the reference angle, K_φ is a force constant for rotation around a dihedral angle, n is the multiplicity (number of minima per full turn) of the dihedral angle φ, δ is the phase shift (location of first maximum), k_ξ is an out-of-plane bending force constant for the improper dihedral angle ξ with reference angle ξ_0, and q_i and q_j are the partial charges of atoms i and j separated by the distance r_{ij}. A_{ij} and B_{ij} are the Lennard–Jones parameters for the interaction between atoms i and j. The Lennard–Jones parameters are defined per atom type as A_i and B_i and are combined using either of the two standard rules to determine the effective interaction parameters. The geometric rule is simply: $A_{ij} = A_i \cdot A_j$ and $B_{ij} = B_i \cdot B_j$, where $A_i = A_{ii}^{1/2}$ and $B_i = B_{ii}^{1/2}$. Some force fields use the form $\epsilon_{ij}[(R^*_{ij}/r_{ij})^{12} - 2(R^*_{ij}/r_{ij})^6]$ for the 6–12 Lennard–Jones potential. In this case the atom type parameters ϵ_i, ϵ_j, R_i^* and R_j^* are combined using the rules: $\epsilon_{ij} = (\epsilon_i \cdot \epsilon_j)^{1/2}$ and $R^*_{ij} = \frac{1}{2}(R^*_i + R^*_j)$. Several Fourier components of the dihedral terms, with different K_φ, n, and δ, can be added for the same dihedral angle to allow a more accurate modelling of the barriers for rotation. An alternative form of the improper dihedral potential using trigonometric functions just as for normal dihedrals is also implemented.

The molecular fragments, e.g., amino acid residues, defined in the force fields are divided into charge groups, which are groups of atoms whose partial charges add up to an integer. The cutoff of nonbonded interactions is then done on the basis of these groups, i.e., either all pairwise interactions between the two groups are evaluated, or none. The average size of the charge groups varies between force fields, from a few atoms to entire residues. Some force fields designate a "switching centre" in each charge group and performs a cutoff only on the basis of the distance between the switching centres, while

others include all interactions between two groups if any pair of atoms is within the cutoff radius.

The long-range electrostatic interactions can be taken into account in Q by the LRF procedure. The expansion coefficients are updated when the lists of nonbonded interaction pairs are regenerated, typically every 25–50 time steps.

We have currently translated GROMOS87,[34] GROMOS96,[35] Amber/OPLS,[36] Amber95,[37] and CHARMM[38] version 22 for use with Q. Some properties of our implementations of these force fields are given in Table 1.

Solvent

Solvation of solute molecules prior to simulation proceeds by inserting water molecules from a file containing coordinates for a box or sphere of water, in such a way that the volume of the simulation sphere not occupied by solute is filled with water, while avoiding too close contacts between solute atoms and water molecules. The water box or sphere may represent either an equilibrated configuration from a simulation or a lattice setup. A water box can be replicated in all directions and can thus be used to solvate a system of any size.

Solvent is at this stage limited to the SPC[39] and TIP3P[40] water models, but apart from the distinction between solute and solvent in regard to energy bookkeeping, any number of molecules of any kind can be simulated. Simulations in vacuum are also possible.

Restraints

Bond lengths, or any desired atom–atom distances, can be constrained with the SHAKE algorithm.[41] SHAKEing of bonds and angles in water is enabled by default. Various parts of the simulated system can be restrained using harmonic potentials. Either the initial coordinates or a separate set of coordinates, e.g., from an earlier simulation may be used as the reference in the restraining potential. A sequence of solute atoms can be restrained, e.g., to allow equilibration of solvent before letting the solute move. A molecule can also be restrained as one entity to its initial geometrical centre. An individual atom may be restrained to a point, a line, or a plane. Atom–atom distances may be restrained. A radial "soft wall" or half-harmonic potential can also be applied to a set of atoms, e.g., to restrict movement of the solutes or counterions. Restraints can be applied selectively to specific EVB or FEP states.

Free energy perturbation

An important feature of Q is the possibility to redefine easily the interaction parameters for a selected set of atoms, i.e., to modify the force field and the molecular topology. This set of atoms, which we call Q atoms, typically constitutes a ligand bound to a receptor, or the reacting fragments in an enzymatic reaction modelled by the EVB method. In the LIE case, no changes to the force field are necessary and defining a set of Q atoms is merely a convenient way of bookkeeping ligand-surrounding interaction energies separately.

The typical procedure in FEP calculations is to define the initial and final states of a transformation in terms of changes in connectivity, partial charges, hybridisation (atom types), and parameters for bonds and all kinds of angles. Bonds that are broken or formed can be redefined as Morse potentials to ensure correct asymptotic behaviour. For the same reason, cross-terms between angles, dihedrals, and improper dihedrals and bonds that are broken or formed can be introduced by scaling the relevant angle energy by the ratio of the instantaneous value of Morse the bond potential to its dissociation energy.[42] Alternatively, modification of bonds can be accomplished with ordinary harmonic potentials or by using SHAKE. It is also possible to replace the standard Lennard–Jones potential with an exponential function giving a softer repulsion, or to disable nonbonded interactions between pairs of atoms. SHAKE constraints can be imposed on atom-atom distances. Off-diagonal Hamiltonian functions required for EVB calculations can be defined.

The force field energies for an arbitrary number of EVB or FEP states are evaluated separately. The effective potential used to calculate forces is a sum over all states weighted by the mapping vector $\boldsymbol{\lambda}$ [Eq. (2)]. We have found no convincing argument to scale each force field parameter individually, as is done in some programs. On the contrary, the scaling of entire potential functions is intuitive and efficient for bookkeeping. Q can be used for both the "single topology" and "dual-topology"

Table 1. Force fields available in Q

Force field[a]	CH_n groups[b]	LJ type[c]	Impropers[d]	Charge groups[e]	Cutoff[f]	Atom types[g]	Ref.
GROMOS87	Extended	A_{ij},B_{ij}	Harmonic	≤10 atoms	Switching	28	34
GROMOS96	Extended	A_{ij},B_{ij}	Harmonic	≤10 atoms	Switching	28	35
Amber/OPLS	Extended	A_{ij},B_{ij}	Periodic	≤11 atoms	Switching	39	36
Amber95	All-atom	$\varepsilon_{ij},R^*_{ij}$	Periodic	Residues	Any	48	37
CHARMM v.22	All-atom	$\varepsilon_{ij},R^*_{ij}$	Harmonic	≤13 atoms	Switching	186	38

[a] Our implementation of the named force field.
[b] Hydrogen atoms on aliphatic carbons may either be explicitly treated (all atom) or modelled as an extended atom.
[c] The Lennard–Jones potential can be written either as $(A_{ij}/r^{12}) - (B_{ij}/r^6)$ or as $\epsilon_{ij}\,[(R^*_{ij}/r_{ij})^{12} - 2\,(R^*_{ij}/r_{ij})^6]$, using the geometric or arithmetic rules, respectively, to combine parameters for pairs of atom types. Treatment of 1–4 interactions (LJ and electrostatic) is specific for each force field.
[d] Improper dihedrals can be modelled either with harmonic potentials or with a periodic function like ordinary dihedrals.
[e] Typical number of atoms in a charge group.
[f] Cutoffs are always applied to whole charge groups and are based either on the distance between designated switching atoms or on the smallest distance between any pair of atoms in two charge groups.
[g] The number of different atom types defined in the Q implementation of the force field.

approaches to multistate simulations.[43] In the former, the same single set of coordinates is used for all states (with "dummy atoms" where atoms are removed) but their interaction parameters are transformed. In the latter method, which is said to be less efficient,[43] the initial and final states are defined as separate fragments present simultaneously and gradually turned off and on, respectively. We prefer to make the smallest possible change of the simulated system, i.e., to transform the smallest possible number of parameters, in order to minimise convergence problems, and thus use the single topology approach. However, more important than the choice of topological representation is, in our opinion, the definition of a smooth pathway that connects the initial and final states. This is particularly critical at an end point, where atoms disappear and may require very closely spaced λ points.

During the simulation, the potential energies for each state (interactions among the Q atoms and their interactions with the rest of the system) are stored periodically. A perturbation simulation consists of configurational sampling at a series of λ values that connect the initial and final states through many small steps. Each step of the simulation usually starts from the final coordinates of the preceding step. The free energy change for the transformation is calculated after the simulation using the free energy analysis program. In the case of standard FEP, where the "reaction coordinate" is identical to the mapping parameter λ, it is given by Eq. (3) above. In EVB calculations, the reaction coordinate is usually the energy gap between two VB states, $X = \Delta V = \epsilon_i - \epsilon_j$ [cf. Eq. (5)], and the reference state is the ground state energy given by the solution of the EVB secular determinant, as described earlier. The free energy for the reverse process $\Delta G_{\text{final} \rightarrow \text{initial}}$ is also calculated by the free energy analysis program. It is common to give the arithmetic mean of the forward and backward values as the final result, with the difference or hysteresis as a measure of the convergence error. In general, however, multiple independent simulations of both the forward and reverse processes provide a more meaningful error estimate.

THE PROCEDURE

The steps below outline the typical procedure for a molecular dynamics-based free energy calculation with Q.

1. Select a force field.
2. Prepare the coordinate input file. This step includes defining the protonation state of ionisable groups on the basis of the chosen size and centre of the simulation sphere and, for crystal structures, whether some or all crystallographic water molecules should be retained.
3. If necessary, define new fragments, molecules, or residues in a molecular library file.
4. Use the interactive topology handling program to generate a molecular topology, i.e., a file containing all information about the simulated system (coordinates, bonds, all kinds of angles, charges, Lennard–Jones parameters, etc.). At this stage it may become necessary to add parameters to the force field parameter file. The program will list all missing interaction parameters and let the user add parameters until all interactions are explicitly defined. It can also calculate force field energies to make it possible to verify that the input structure, parameters, and fragment libraries are reasonable already before the simulation has been set up.
5. Write an FEP input file. This is required if some atoms should be treated in a special way, e.g., for collecting ligand-surrounding interaction energies or for perturbations of any kind. The FEP file designates selected atoms as Q atoms and allows modifications of the interaction parameters where these atoms are involved. Specifically, one can change partial charges, atom types (Lennard–Jones parameters and masses), bonds, and bond parameters like force constants and angles of all kinds.
6. Prepare a series of MD input files to define a multistage simulation protocol with slow heating, equilibration of water, followed by data collection and stepwise perturbation. This could be done with various degrees of automation using our script tool or your own script.
7. Test your simulation setup by running the sequence of input files through the test program, which runs only one MD step.
8. Submit the simulation job. It is possible to extract coordinates from the running simulation to obtain snap-shots of the structure and to follow energies that appear in the log file.
9. Analyse the results by looking at snap-shot, average, or final structures, by calculating energy averages, or use the free energy analysis program to calculate changes in free energy during a perturbation.

RESULTS: SELECTED EXAMPLES

We present examples from the three categories of free energy calculation introduced above, which show how Q can be used in various biochemical applications. All of the following simulations were performed using the GROMOS87 force field, with modifications by Åqvist et al.[5,44] The nonbonded cutoff radius was 10 Å, beyond which electrostatic interactions were treated by the LRF approximation. Simulation spheres with 16-Å radius centred at the ligand or reacting fragment were used and the space not occupied by the solutes was filled with randomly oriented SPC water molecules from a grid. Water molecules were kept rigid with SHAKE.

Simulation of bulklike water in a small sphere

We first present some basic results from simulations of a 15-Å sphere of pure water and a sphere containing an Na^+ ion at its centre. Figure 1a shows the average water number density in 3-Å shells obtained with the restraining potential of Eq. (8). For comparison, the density obtained without the surface attraction term is also depicted in Figure 1a. It can be seen that the density is indeed well-behaved and is virtually uniform inside the sphere and close to the experimental value. The effect of the surface attraction is also obvious, namely, that it compensates for the otherwise reduced density near the surface. Figure 1b shows the polarisation distribution of water molecules in the outermost 3-Å shell of the sphere with and without the surface polarisation restraints. It is evident from the figure that without the polarisation restraints the distribution becomes too peaked, with an enhanced tendency for the surface water molecules to align their dipole moments tangential to the boundary. Finally, Figure 1c shows the average polarisation of water around a Na^+ ion as a function of distance, using the surface restraints described above. It can be seen that the behaviour closely matches the expected polarisation in a dielectric continuum.

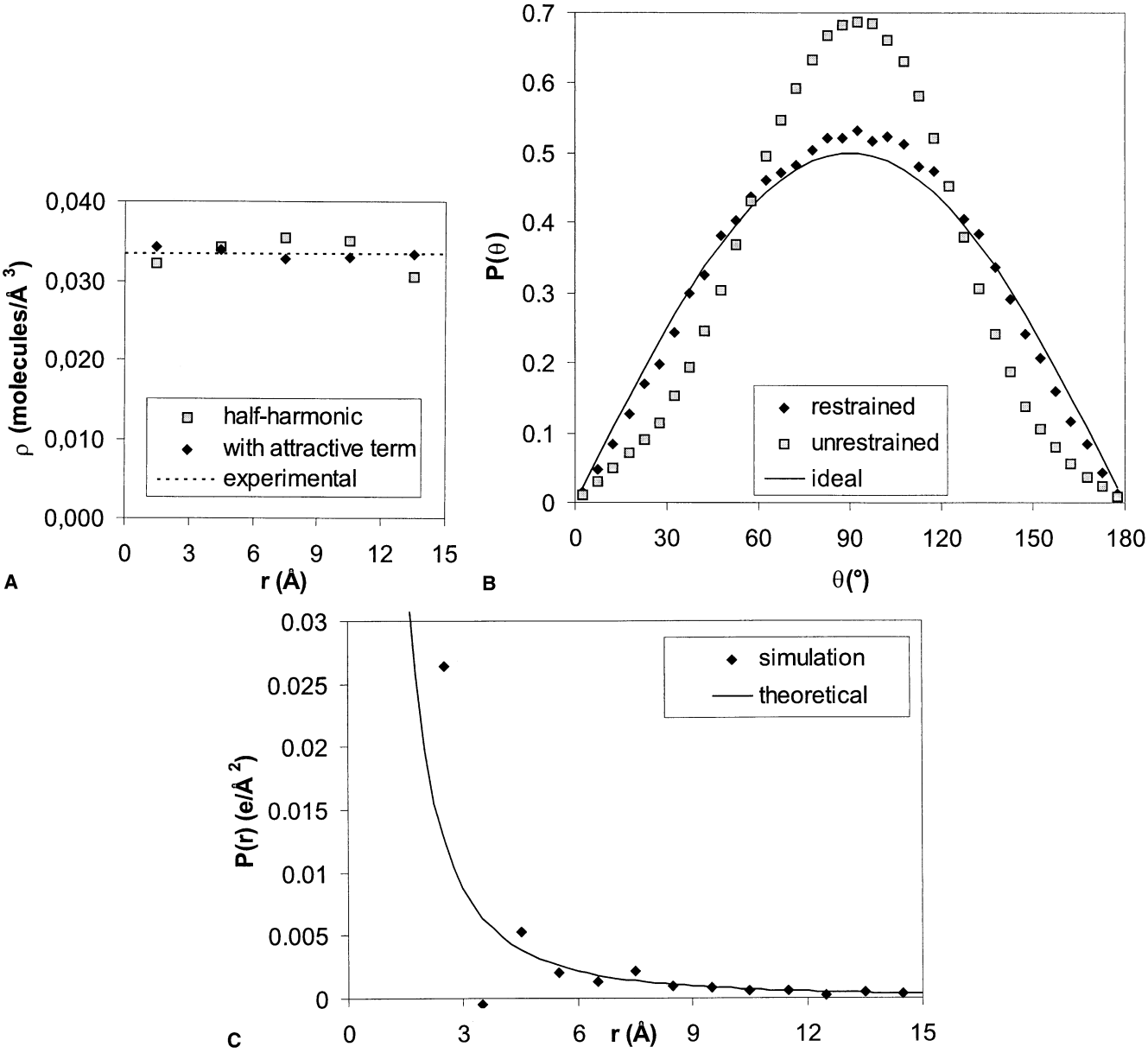

Figure 1. (a) Number–density plot in a 15-Å sphere of pure water with a boundary attraction term in the radial restraining potential (diamonds), with only a half-harmonic potential (squares) and the experimental number density (dotted line). (b) Probability distribution for the polarisation in a 3-Å surface layer of a 15-Å sphere of pure water. The distribution of angles between molecular dipole vector and radial axis for a simulation with polarisation restraints (diamonds) and without (squares) and the theoretical result for freely rotating molecules (solid line) is plotted. (c) Average polarisation of water around a Na^+ ion (diamonds), calculated in 1-Å spherical shells, and the corresponding expected polarisation in a continuum system $P(r) = Q/4\pi r^2 \cdot [1 - (1/\epsilon)]$, with $\epsilon = 80$. Note that the deviations for small r are due to solvation structure in the microscopic system.

FEP application: Difference in substrate affinity between wild-type and an active site mutant in two tyrosine phosphatases

Protein tyrosine phosphatases (PTPs) are ubiquitous signal transduction enzymes, which catalyse dephosphorylation of phosphotyrosine residues.[45,46] Several PTPs have been isolated and characterised, such as human PTP1B[47] and bovine low molecular weight PTP (lmPTP).[48] Although there is almost no structural or sequence homology between lmPTP and PTP1B, the active sites of both enzymes possess a "catalytic loop," consisting of the PTP common signature motif Cys-$(X)_5$-Arg.[49] The cysteine residue of the catalytic loop (Cys-215 in PTP1B, Cys-12 in lmPTP) is the essential nucleophile for PTP activity. Experiments show that cysteine-to-serine mutants are completely inactive.[47,49] To investigate the difference in the sub-

strate affinity of wild-type (wt) lmPTP and PTP1B and their Cys → Ser mutants, we performed relative binding free energy calculations using FEP in the thermodynamic cycle shown in Figure 2 with phenyl phosphate dianion bound in the active site. The catalytic cysteine was then slowly transformed to serine, and vice versa, in both enzymes.

The simulations of lmPTP were performed starting from a crystal structure of the enzyme complexed with a sulphate ion, PDB entry 1PHR.[48] The ion was removed and the phenyl phosphate ligand was docked into the protein structure using Insight II.[50] The PTP1B structure was a serine mutant with a phosphotyrosine ligand,[47] PDB entry 1PTV, which was replaced manually by phenyl phosphate. Each simulation was prepared by slow heating of the system from 1 to 300 K followed by 100-ps equilibration at 300 K. The perturbations were performed at this temperature using 51 λ steps, 1-fs time steps, and 5-ps sampling at each λ. Energies were collected every fifth femtosecond. The energy data from the first 2 ps of each λ step were discarded for equilibration. Forward and backward FEP simulations were done, starting from mutant PTP1B and wt lmPTP. The backward simulations, preceded by a 25- to 50-ps equilibration, were started from the end points of the forward runs.

The results shown in Table 2 show that the simulations forward and backward yielded similar energies with small standard deviations. The negative $\Delta\Delta G_{\text{bind}}$ values predict that the serine mutants of both lmPTP and PTP1B bind phenyl phosphate about 8 kcal/mol more strongly than the native enzymes (experimental binding measurements are underway).

EVB application: Energetics of nucleophile activation in two tyrosine phosphatases

Dephosphorylation of phosphotyrosine by protein tyrosine phosphatases proceeds via a displacement mechanism in which the catalytic cysteine attacks the phosphorus atom of the substrate, yielding a cysteinyl phosphate intermediate. Since a negatively charged thiolate group is a better nucleophile than the protonated thiol, the catalytic cysteine is believed to be deprotonated prior to the nucleophilic attack. Depending on its pK_a, the cysteine could be deprotonated already in the free enzyme, or it could be ionised after proton transfer to the substrate phosphate group in the Michaelis complex.[51]

We have employed the empirical valence bond (EVB) method[3] to study the energetics of nucleophile activation by proton transfer to a dianionic substrate in two PTPs, the bovine liver low molecular weight PTP (lmPTP) and human PTP1B. The two valence bond (VB) states Φ_1 and Φ_2 used in the calculations are shown in Figure 3. These states represent the reactants and products for the reaction in which a proton is transferred from the cysteine residue to the phenyl phosphate dianion. The charges and nonbonded parameters used to model the VB states were those developed by Hansson et al.[51] Bonds were represented by standard Morse potentials to allow realistic dissociation and bond formation. Starting coordinates for the protein simulations were the same as described above. The protein systems were equilibrated during a 20-ps stepwise heating scheme and thereafter 50-ps simulation at 300 K. The water system was equilibrated by simulation at 300 K for 50 ps. All equilibrations were performed with the mapping vector $\lambda_1 = \lambda_2 = 0.50$ (half of either VB state) and after equilibration the free energy perturbation proceeded toward reactants and products. In total, 83 λ points were used and each λ step was sampled for 5 ps with a time step of 1 fs. The energy data from the first 2 ps of each λ step were discarded for equilibration.

The EVB potential was calibrated to reproduce experimental data for the uncatalysed reference reaction in solution. In the case of proton transfer the difference in free energy between the two states, $\Delta G°$ can be obtained from the difference in pK_a between the donor and acceptor (cysteine pK_a = 8.3, phenyl

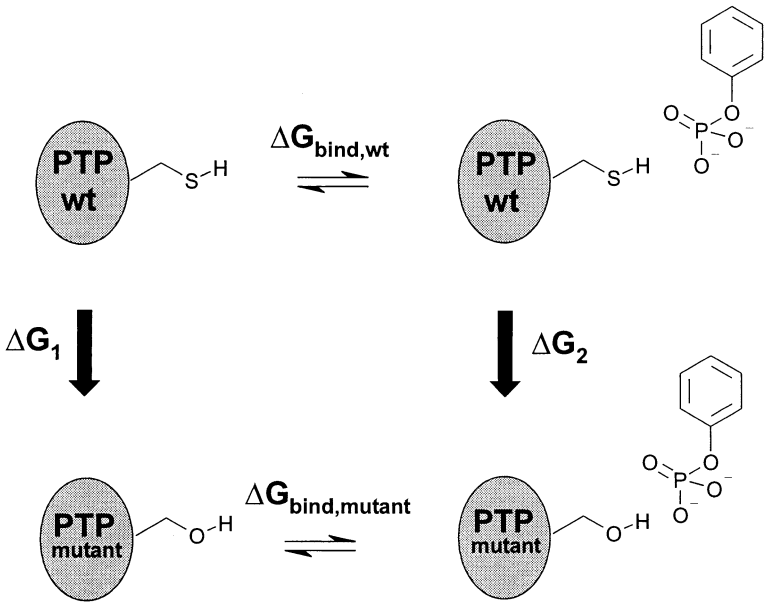

Figure 2. Thermodynamic cycle for relative substrate affinity of wild-type and mutant tyrosine phosphatases. The vertical arrows show the paths of the simulations. The difference in ligand affinity, $\Delta\Delta G_{bind}$, is obtained as $\Delta\Delta G_{bind} = \Delta G_{bind,\ mutant} - \Delta G_{bind,\ wt} = \Delta G_2 - \Delta G_1$.

Table 2. Results obtained from FEP simulations

Simulated system	ΔG (kcal/mol)[a]		ΔG_{ave} (kcal/mol) (Cys → Ser)[b]	$\Delta\Delta G_{bind}$ (kcal/mol)[c]
	Cys → Ser	Ser → Cys		
Free PTP1B	−7.6 ± 0.1	6.5 ± 0.1	−7.1 ± 0.4	−7.3 ± 0.8
PTP1B + ligand	−15.2 ± 0.1	13.5 ± 0.1	−14.3 ± 0.5	
Free lmPTP	−9.5 ± 0.1	9.0 ± 0.1	−9.2 ± 0.2	−8.0 ± 0.3
lmPTP + ligand	−17.2 ± 0.1	17.3 ± 0.1	−17.2 ± 0.1	

[a] The free energy difference obtained from the simulation. Ser → Cys indicates the forward mutation and Cys → Ser refers to the backward mutation. The error is the convergence error obtained from summation in the two directions on the same trajectory.
[b] The average free energy difference calculated from the two independent simulations of columns 2 and 3, with the standard error of the mean. The sign of this value corresponds to the Cys → Ser mutation.
[c] The difference in binding energy between the wt–ligand complex and the mutant–ligand complex. A negative value indicates that the serine mutant binds the ligand stronger than the wild type. The error is the sum of the errors of the terms.

phosphate pK_a = 5.7). Once $\Delta G°$ is known the activation energy $\Delta G^{‡}$ can be determined from a linear free energy relationship compiled by Eigen.[52] For the proton transfer described by $\Phi_1 \rightarrow \Phi_2$ (Figure 3) the resulting difference in free energy is 3.5 kcal/mol and the activation energy is 7.8 kcal/mol. This estimate of the barrier effectively includes zero-point energy and tunnelling effects since it is obtained from experimental data. In typical EVB studies of enzymatic reactions it is usually assumed that these quantum-mechanical effects do not differ significantly between the water and enzyme environments. However, it is straightforward to implement the path integral method within the EVB framework for assessing such approximations.[53–55] Thus, the reaction free energy profile obtained from the water simulation containing only the solvated reacting fragments was calculated using the umbrella sampling approach [Eq. (5)]. For every reaction coordinate interval the sample size weighted average was calculated. The EVB parameters $\Delta\alpha$ and H_{12} were adjusted until the calculated profile reproduced the experimental values. The resulting values ($\Delta\alpha$ = −105.6 kcal/mol, H_{12} = 24.2 kcal/mol) were then used when evaluating the corresponding simulations of the reaction in the two enzymes.

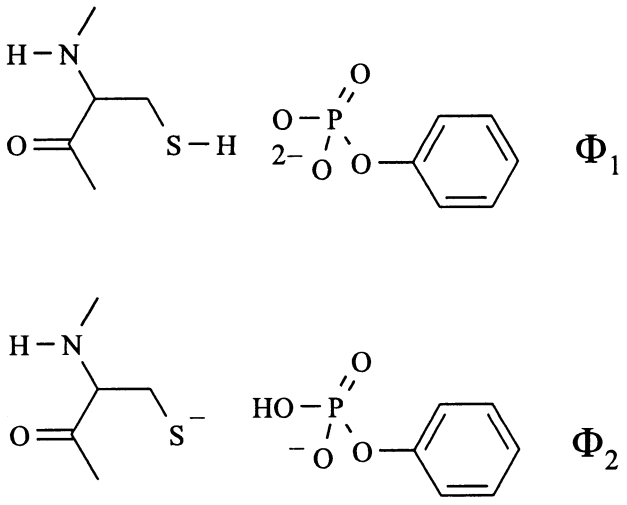

Figure 3. Valence bond states for nucleophile activation in tyrosine phosphatases. The reaction studied is $\Phi_1 \rightarrow \Phi_2$, where a proton is transferred from the cysteine residue to the dianionic phenyl phosphate.

The simulated reaction free energy profiles are shown in Figure 4. The upper curve is the calibrated profile of the reference reaction in solution. The other two curves show that both enzymes have a significant catalytic effect on the reaction, i.e., the activation energies and the free energy differences between the two states are lowered compared with the solution reaction. In comparing the two enzymes it appears that Φ_2 is energetically more stable in PTP1B than in lmPTP, which indicates that the catalytic cysteine has a lower pK_a in the former enzyme. This is consistent with experimental data[56,57] as well as computed pK_a values.[58]

LIE application: Absolute binding free energy for the lysine-binding protein–lysine complex

As an example of binding free energy calculations with the LIE method, we present results for lysine binding to the lysine/arginine/ornithine-binding protein, using the crystal structure of the *Salmonella typhimurium* protein in complex with lysine by Oh et al.,[59] PDB entry 1LST. This calculation is somewhat demanding owing to the three ionic groups of the ligand, and illustrates the important points in LIE calculations.

The protein was made net neutral by modelling charged residues far from the ligand as neutral dipolar groups. The solvated complex was equilibrated by an 8-ps slow heating protocol, during which the positions of all heavy solute atoms were restrained followed by another 14 ps of unrestrained equilibration at 300 K. The solvated ligand was equilibrated for 20 ps at 300 K. In both cases, sampling of ligand-surrounding interaction energies was done every 20 fs during a 500-ps simulation with 2-fs time steps.

Color Plate 2 shows that ligand-surrounding interaction energies fluctuate but that the 25-ps average values (plotted with thick white lines) attain more or less stable average values after about 50 ps. Both the Lennard–Jones and electrostatic energies are more negative in the bound simulation, i.e., they both contribute favourably to the calculated binding affinity. The average energies are given in Table 3, as well as the free energy contribution from the interactions between the ligand and the charged groups that were switched off during the simulation, according to Eq. (9). The errors in the average energies are estimated on the basis of the statistical inefficiency of the

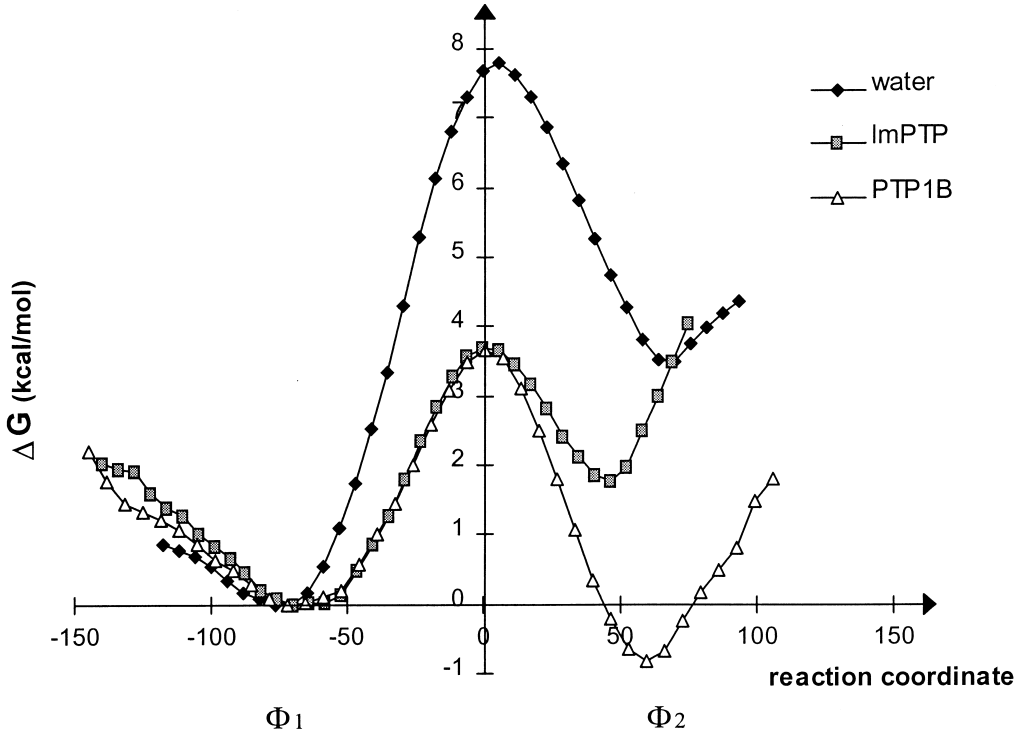

Figure 4. Free energy profile for proton transfer in tyrosine phosphatases. The upper curve is the calibrated profile for the uncatalysed reference reaction in water. Both enzymes stabilise the cysteine thiolate anion form significantly, compared with the solution reaction.

data[20] and the errors of the free energy contributions are root mean squares of the errors the involved averages. The calculated absolute binding affinity is in reasonable agreement with experiment, although somewhat overestimated.

SUMMARY

The program presented here is well suited for studying structure–function relationships in biomolecules, utilising free energy calculations, in applications such as

- Investigation of the effect of specific mutations on ligand binding affinity or activation free energy by FEP calculations
- Testing of hypotheses about enzyme reaction mechanisms by calculation of reaction free energy profiles using the EVB method
- Receptor–ligand binding free energy prediction with the LIE method

The program can be used with many popular force fields, is relatively easy to use, and runs on many common platforms.

SOFTWARE

The MD, topology preparation, and free energy analysis programs are written in Fortran90. There are no restrictions in the

Table 3. LIE binding free energy calculation for lysine–lysine-binding protein

Energy (kcal/mol)[a]	Solvated complex	Ligand in water	Scaling factor[b]	Binding affinity contribution[c]
$\langle V_{l-s}^{LJ} \rangle$	-10.6 ± 0.5	$+6.0 \pm 0.1$	0.181	-3.0 ± 0.1
$\langle V_{l-s}^{el} \rangle$	-282.8 ± 2.3	-261.1 ± 1.4	0.500	-10.8 ± 1.3
$\Delta G_{corr.}^{el}$				$+1.5$
ΔG_{bind}^{LIE}				**-12.3 ± 1.3**
$\Delta G_{bind}^{expt.}$				**-10.9**

[a] Average Lennard–Jones and Coulomb ligand-surrounding interaction energies from the simulations and the free energy correction for the charged residues that were not included in the simulation. The free energy of binding was calculated using Eq. (6). Errors were calculated using the procedure described in Ref. 20.
[b] LIE scaling factor [α and β in Eq. (6)].
[c] Net contribution to the free energy of binding from the different parts of Eq. (6). The errors are the square root of the sum of squared errors of each term.

programs on the number of particles or the number of force field parameters. The programs use only the minimum amount of memory required for each particular task (dynamic memory allocation). It is thus feasible to run simulations of moderate size on standard personal computers. A typical simulation of a solvated protein–ligand system in a sphere of radius 18 Å comprising about 3 000 particles requires 2 to 4 MB of memory.

Although developed under Windows NT™ on Digital Alpha™ workstations, the code is designed to be portable to any platform that has Fortran90. Specifically, the code also runs on Intel-based PCs with Windows NT or Windows95/98™, on Compaq/Digital workstations with Digital UNIX™ version 4 and on Silicon Graphics computers with IRIX™ version 6. A number of utility programs written in PERL and Fortran77 are also available and are expected to be portable without problems.

The topology handling program reads molecular structures in PDB format and can write PDB or Sybyl mol2 file formats. All other input files are free-format text files. Coordinate and trajectory files generated by the MD program are Fortran binary files.

Q is distributed free of charge for academic use and can be obtained from the Åqvist group web site at http://aqvist.bmc.uu.se, where source code, makefile, executable images for some platforms, manual, and force field files are available, as well as some examples. Details about obtaining a license to use Q will also be given at the web site, or by e-mail to Johan Åqvist (aqvist@xray.bmc.uu.se).

ACKNOWLEDGEMENTS

We gratefully acknowledge support from the Swedish Research Council for Engineering Sciences (TFR) and the Swedish Natural Science Research Council (NFR). The protein phosphatase research is funded by the EC (DGXII). I.F. is a graduate student of the Swedish Foundation for Strategic Research (SSF/SBNet).

We thank Dr. Peter Vagedes (Department of Chemistry, Freie Universität, Berlin, Germany) for translating the CHARMM force field.

We also thank agent 007's provider of fancy equipment, who lent his name to the program.

REFERENCES

1 Kollman, P.A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* 1993, **93,** 2395–2417
2 Beveridge, D.L., and DiCapua, F.M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18,** 1989, 431–492
3 Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions.* Wiley, New York, 1988
4 Åqvist, J., and Warshel, A. Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chem. Rev.* 1993, **93,** 2523–2544
5 Åqvist, J., Medina, C., and Samuelsson, J.E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* 1994, **7,** 385–391
6 Jones-Hertzog, D.K., and Jorgensen, W.L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem.* 1997, **40,** 1539–1550
7 Hansson, T., Marelius, J., and Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* 1998, **12,** 27–35
8 Böhm, H.J. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* 1998, **12,** 309–323
9 Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V., and Mee, R.P. Empirical scoring functions. I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* 1997, **11,** 425–445
10 Murray, C.W., Auton, T.R., and Eldridge, M.D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand–receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J. Comput.-Aided Mol. Des.* 1998, **12,** 503–519
11 Head, R.D., Smythe, M.L., Oprea, T.I., Waller, C.L., Greene, S., and Marshall, G.R. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* 1996, **118,** 3959–3969
12 Takamatsu, Y., and Itai, A. A new method for predicting binding free energy between receptor and ligand. *Proteins* 1998, **33,** 62–73
13 Jain, A.N. Scoring noncovalent protein–ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J. Comput.-Aided Mol. Des.* 1996, **10,** 427–440
14 Gerber, P.R., Mark, A.E., van Gunsteren, W.F., and Mark, A.E. An approximate but efficient method to calculate free energy trends by computer simulation: Application to dihydrofolate reductase-inhibitor complexes. *J. Comput.-Aided Mol. Des.* 1993, **7,** 305–323
15 Radmer, R.J., and Kollman, P.A. The application of three approximate free energy calculation methods to structure based ligand design: Trypsin and its complex with inhibitors. *J. Comput.-Aided Mol. Des.* 1998, **12,** 215–227
16 Åqvist, J., and Mowbray, S.L. Sugar recognition by a glucose/galactose receptor. Evaluation of binding energetics from molecular dynamics simulations. *J. Biol. Chem.* 1995, **270,** 9978–9981
17 Åqvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* 1996, **17,** 1587–1597
18 Hultén, J., Bonham, N.M., Nillroth, U., Hansson, T., Zuccarello, G., Bouzide, A., Åqvist, J., Classon, B., Danielsson, H., Karlén, A., Kvarnström, I., Samuelsson, B., and Hallberg, A. Cyclic HIV-1 protease inhibitors derived from mannitol: Synthesis, inhibitory potencies and computational predictions of binding affinities. *J. Med. Chem.* 1997, **40,** 885–897
19 Hansson, T., and Åqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* 1995, **8,** 1137–1144
20 Marelius, J., Graffner-Nordberg, M., Hansson, T., Hall-

berg, A., and Åqvist, J. Computation of affinity and selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases. *J. Comput.-Aided Mol. Des.* 1998, **12,** 119–131
21 Paulsen, M.D., and Ornstein, R. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* 1996, **9,** 567–571
22 Gorse, A.D., and Gready, J.E. Molecular dynamics simulations of the docking of substituted N5-deazepterins to dihydrofolate reductase. *Protein Eng.* 1997, **10,** 23–30
23 Åqvist, J., and Hansson, T. On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem.* 1996, **100,** 9512–9521
24 Åqvist, J. Comment on transferability of ion models. *J. Phys. Chem.* 1994, **98,** 8253–8255
25 Allen, M.P., and Tildesley, D.J. *Computer Simulation of Liquids.* Oxford University Press, Oxford, 1987
26 Lee, F.S., and Warshel, A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.* 1992, **97,** 3100–3107
27 Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., di Nola, A., and Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 1984, **81,** 3684–3690
28 Warshel, A. A microscopic model for calculations of chemical processes in aqueous solution, *Chem. Phys. Lett.* 1978, **55,** 454–458
29 Berkowitz, M., and McCammon, J.A. Molecular dynamics with stochastic boundary conditions. *Chem. Phys. Lett.* 1982, **90,** 215–217
30 Brünger, A.T., Brooks, C.L., III, and Karplus, M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* 1984, **105,** 495–500
31 Essex, J.W., and Jorgensen, W.L. An empirical boundary potential for water droplet simulations. *J. Comput. Chem.* 1995, **16,** 951–972
32 King, G., and Warshel, A. A surface contained all-atom solvent model for effective simulations of polar solutions. *J. Chem. Phys.* 1989, **91,** 3647–3661
33 Warshel, A., and Creighton, S. Microscopic free energy calculations in solvated macromolecules as a primary structure–function correlator and the MOLARIS program. In: *Computer Simulation of Biomolecular Systems* (van Gunsteren, W.F., and Wiener, P.K., eds.). ESCOM, Leiden, 1989, pp. 120–138
34 van Gunsteren, W.F., and Berendsen, H.J.C. *Groningen Molecular Simulation (GROMOS) Library Manual.* Biomos B.V., Groningen, The Netherlands, 1987
35 van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., and Tironi, I.G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide.* Hochschulverlag AG an der ETH Zürich, Switzerland, 1996
36 Jorgensen, W.L., and Tirado-Rives, J. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 1988, **110,** 1657–1666
37 Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 1995, **117,** 5179–5197
38 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 1993, **4,** 187–217
39 Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., and Hermans, J. Interaction models for water in relation to protein hydration. In: *Intermolcular Forces* (Pullman, B., ed.). Riedel, Dordrecht, The Netherlands, 1981, pp. 331–342
40 Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983, **79,** 926–935
41 Ryckaert, J.P., Ciccotti, G., and Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Physics* 1977, **23,** 327–341
42 Åqvist, J. Free energy perturbation study of metal ion catalyzed proton transfer in water. *J. Phys. Chem.* 1991, **95,** 4587–4590
43 Pearlman, D.A. A comparison of alternative approaches to free energy calculations. *J. Phys. Chem.* 1994, **98,** 1487–1493
44 Åqvist, J., Fothergill, M., and Warshel, A. Computer simulation of the carbon dioxide–hydrogen carbonate interconversion step in human carbonic anhydrase I. *J. Am. Chem. Soc.* 1993, **115,** 631–635
45 Fischer, E.H., Charbonneau, H., and Tonks, N.K. Protein tyrosine phosphatases: A diverse family of intracellular and transmembrane enzymes. *Science* 1991, **253,** 401–406
46 Barford, D., Jia, Z., and Tonks, N.K. Protein tyrosine phosphatases take off. *Nature Struct. Biol.* 1995, **2,** 1043–1053
47 Jia, Z., Barford, D., Flint, A.J., and Tonks, N.K. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* 1995, **268,** 1754–1758
48 Su, X.D., Taddei, N., Stefani, M., Ramponi, G., and Nordlund, P. The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. *Nature (London)* 1994, **370,** 575–578
49 Cirri, P., Chiarugi, P., Camici, G., Manao, G., Raugei, G., Cappugi, G., and Ramponi, G. The role of Cys12, Cys17 and Arg18 in the catalytic mechanism of low-M(r) cytosolic phosphotyrosine protein phosphatase. *Eur. J. Biochem.* 1993, **214,** 647–657
50 Biosym/MSI. *Insight II.* Biosym/MSI, San Diego, California, 1995
51 Hansson, T., Nordlund, P., and Åqvist, J. Energetics of nucleophile activation in a protein tyrosine phospatase. *J. Mol. Biol.* 1997, **265,** 118–127
52 Eigen, M. Proton transfer, acid-base catalysis and enzymatic hydrolysis. I. Elementary processes. *Angew. Chem. (Intl. Ed. Engl.)* 1964, **3,** 1–72
53 Hwang, J.K., Chu, Z.T., Yadav, A., and Warshel, A. Simulations of quantum mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions. *J. Phys. Chem.* 1991, **95,** 8445–8448
54 Lobaugh, J., and Voth, G.A. A path integral study of electronic polarisation and nonlinear coupling effects in condensed phase proton transfer reactions. *J. Chem. Phys.* 1994, **100,** 3039–3047

55 Hinsen, K., and Roux, B. Potential of mean force and reaction rates for proton transfer in acetylacetone. *J. Chem. Phys.* 1997, **106,** 3567–3577
56 Zhang, M., Zhou, M., Van Etten, R.L., and Stauffacher, C.V. Crystal structure of bovine low molecular weight phosphotyrosyl phosphatase complexed with the transition state analog vanadate. *Biochemistry* 1997, **36,** 15–23
57 Lohse, D.L., Denu, J.M., Santoro, N., and Dixon, J.E. Roles of aspartic acid-181 and serine-222 in intermediate formation and hydrolysis of the mammalian protein-tyrosine-phosphatase PTP1. *Biochemistry* 1997, **36,** 4568–4575
58 Peters, G.H., Frimurer, T.M., and Olsen, O.H. Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases. *Biochemistry* 1998, **37,** 5383–5393
59 Oh, B.H., Pandit, J., Kang, C.H., Nikaido, K., Gokcen, S., Ames, G.F., and Kim, S.H. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J. Biol. Chem.* 1993, **268,** 11348–11355