Thermodynamics of Water Mediating Protein-Ligand Interactions in Cytochrome P450cam: A Molecular Dynamics Study

Volkhard Helms and Rebecca C. Wade
European Molecular Biology Laboratory, 69012 Heidelberg, Germany

ABSTRACT Ordered water molecules are observed by crystallography and nuclear magnetic resonance to mediate protein-ligand interactions. Here, we examine the energetics of hydrating cavities formed at protein-ligand interfaces using molecular dynamics simulations. The free energies of hydrating two cavities in the active site of two liganded complexes of cytochrome P450cam were calculated by multiconfigurational thermodynamic integration. The complex of cytochrome P450cam with 2-phenyl-imidazole contains a crystallographically well defined water molecule mediating hydrogen bonds between the protein and the inhibitor. We calculate that this water molecule is stabilized by a binding free energy of −11.6 ± 6.6 kJ/mol. The complex of cytochrome P450cam with its natural substrate, camphor, contains a cavity that is empty in the crystal structure although a water molecule in it could make a hydrogen bond to camphor. Here, solvation of this cavity is calculated to be unfavorable by +15.8 ± 5.0 kJ/mol. The molecular dynamics simulations can thus distinguish a hydrated interfacial cavity from an empty one. They also provide support for the notion that protein-ligand complexes can accommodate empty interfacial cavities and that such cavities are likely to be unhydrated unless more than one hydrogen bond can be made to a water molecule in the cavity.

INTRODUCTION

Water plays a fundamental role in the interactions of proteins with their ligands. The binding process generally involves an entropically favored displacement of solvent molecules from the protein and ligand surfaces and an enthalpically favored reorganization of these solvent molecules (Chervenak and Toone, 1994). For example, approximately 65 water molecules appear to be released on binding of glucose to hexokinase (Rand et al., 1993). Some solvent molecules may, however, be trapped at the protein-ligand interface. These may make an enthalpic contribution to the ligand binding free energy by, for example, mediating hydrogen bond bridges between the ligand and the protein. Because of their mobility relative to the protein and their ability to both accept and donate hydrogen bonds, water molecules are adaptable liganding partners that are able to fill empty space, modulate the binding specificity of the protein, and play a role in its function. Examples of protein structures in which water molecules have been observed to mediate protein-small molecule interactions include those of the complexes of cholesterol oxidase with a steroid substrate (Li et al., 1993), retinol-binding protein with retinol (Cowan et al., 1990), adipocyte lipid-binding protein with arachidonic acid (LaLondo et al., 1994b), adipocyte lipid-binding protein with palmitate and with hexadecanesulfonic acid (LaLondo et al., 1994a), and L-arabinose-binding protein with L-arabinose, D-fucose, and D-galactose (Quiocho et al., 1989; Zacharias et al., 1993). Water molecules have also been observed at protein-protein interfaces, e.g., in the lysozyme D1.3 antibody-antigen complex (Bhat et al., 1994) and in the hirudin-thrombin complex (Rydel et al., 1991), and at protein-DNA interfaces, e.g., in the lambda (Beamer and Pabo, 1992) and *trp* (Shakked et al., 1994) repressor-operator complexes and in the complex of erythroid transcription factor GATA-1 with DNA (Clore et al., 1994).

In a protein-ligand complex, cavities may be present at the intermolecular interface. For example, statistical analysis shows that cavities are frequently present at protein-protein interfaces (Hubbard and Argos, 1994). Such interfacial cavities may or may not be hydrated. We here calculate free energies for the occupation of interfacial cavities by water using a molecular dynamics (MD) simulation method, thus providing a means to distinguish empty interfacial cavities from hydrated ones. We consider two possible solvation sites in two complexes of the well characterized protein, cytochrome P450cam, which has an internal substrate binding site.

Cytochrome P450cam: a well characterized model system

Members of the cytochrome P450 heme protein family are ubiquitous in biological systems where they play important roles in the synthesis of steroids and fatty acids and in the metabolism of xenobiotics (Porter and Coon, 1991). Until recently, only the crystallographic structure of P450cam from the bacterium *Pseudomonas putida* was known and it has provided the structural basis for understanding possible reaction mechanisms. P450cam catalyzes the 5-*exo*-hydroxylation of camphor, which binds in a buried active site that is not directly accessible from the solvent. Crystallographic structures have been solved for the substrate-free

Received for publication 21 March 1995 and in final form 8 June 1995.
Address reprint requests to Dr. Rebecca C. Wade, European Molecular Biology Laboratory, Meyerhofstr. 1, 69012 Heidelberg, Germany. Tel.: 49-6221-387553; Fax: 49-6221-387306; E-mail: wade@emblheidelberg.de.

0006-3495/95/09/810/15 $2.00

form of the enzyme (Poulos et al., 1986) and for complexes with a number of different substrates and inhibitors (Poulos et al., 1987; Poulos and Howard, 1987; Raag and Poulos, 1989; Raag and Poulos, 1991; Raag et al., 1993). A number of MD simulations of P450cam with different ligands have been reported by the group of Ornstein (see, e.g., Paulsen and Ornstein, 1991), which have been directed mostly to structural aspects of ligand binding. The different spin states of the heme iron were analyzed by Loew and co-workers (1993), and Jones et al. (1993) have calculated the free energy difference for the binding of two isomers of nicotine to P450cam.

The substrate-free structure indicates the presence of approximately six water molecules in the active site (Poulos et al., 1986). When a ligand binds, some or all of this solvent is displaced, making a favorable contribution to the free energy of ligand binding. Any solvent remaining when a ligand is bound may assist ligand binding by filling empty space and bridging hydrogen bonds. Two water molecules were located in the active site of the crystal structure of the complex of P450cam with 2-phenyl-imidazole (2pheIm; Fig. 1 *A*). One acts as the sixth ligand to the heme iron whereas the other, water 802, is a well defined solvent molecule (with 95% occupancy and a B-factor of 0.25 nm^2)

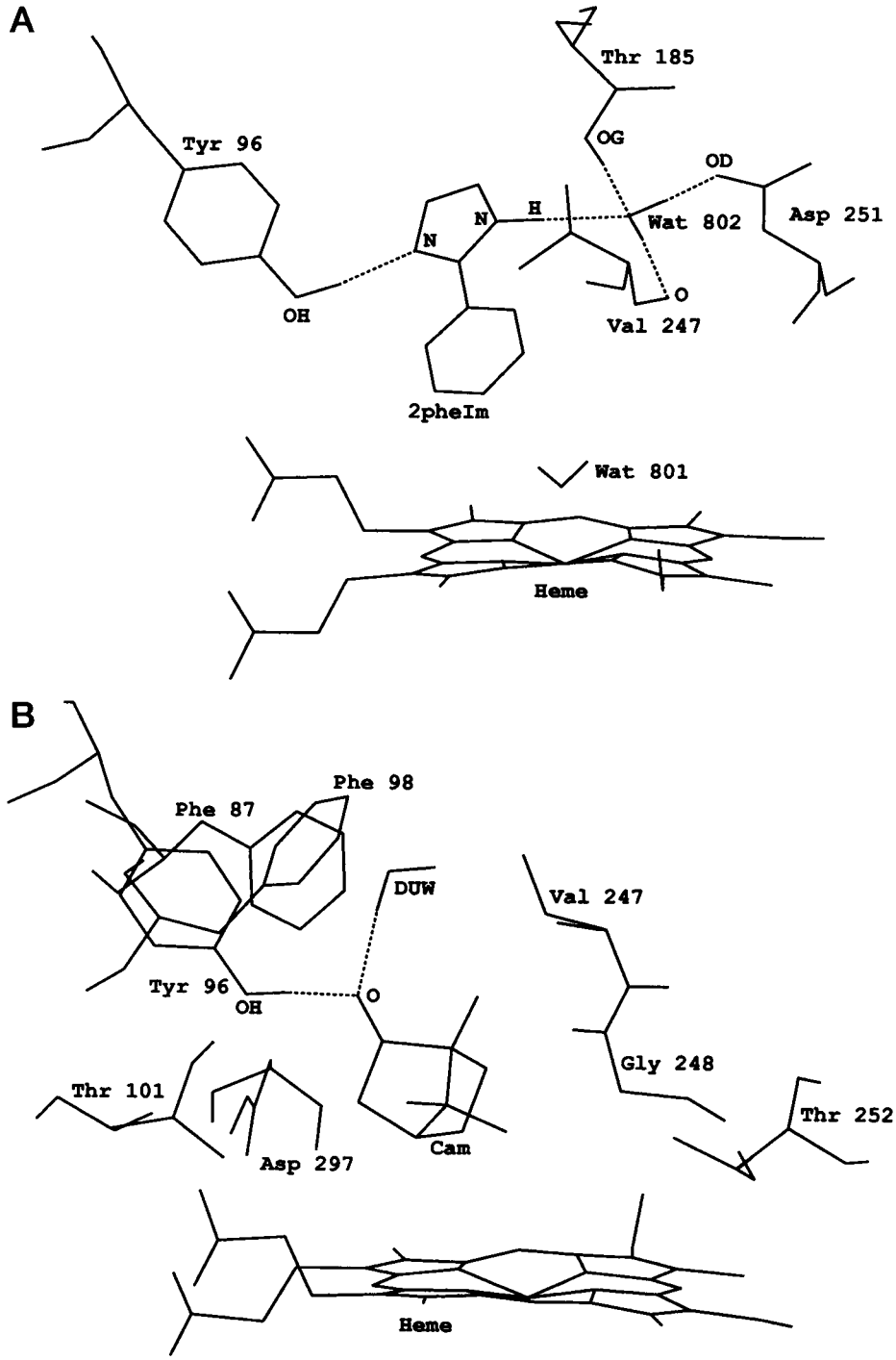

FIGURE 1 The active site of cytochrome P450cam in the crystal structures of the enzyme complexed with (*A*) 2-phenyl-imidazole, with the crystallographically well defined water molecule WAT802 shown, bridging hydrogen bonds (dotted lines) between the ligand and the protein, and (*B*) camphor, with the additional water molecule DUW suggested by the GRID calculations, but not observed in the crystal structure, modeled into an optimal geometry.

participating in hydrogen bonds with 2pheIm and adjacent protein residues. This water molecule is clearly mediating the ligand-protein interaction and influencing the ligand binding mode and orientation. Indeed, an MD simulation of P450cam complexed with 2pheIm by Harris et al. (1995) indicates the importance of water 802 for the orientation of the ligand. In contrast, no water molecules were assigned in the active site of the complex of P450cam with camphor (Fig. 1 *B*) despite the presence of interfacial cavities large enough to accommodate water molecules.

Water binding site proposed in previous GRID calculations

In an earlier study (Wade, 1990), the interaction energy of a probe water molecule in the active site of P450cam with the protein was calculated with the GRID program (Goodford, 1985; Boobbyer et al., 1989; Wade et al., 1993; Wade and Goodford, 1993). This program implements a computationally fast method of determining energetically favorable ligand binding sites on molecules of known structure. The probe molecule is moved through the protein matrix on a grid, and at each point, energies are calculated as the sum of Lennard-Jones, electrostatic, and hydrogen bond terms with explicit modeling of hydrogen bond geometries. For the camphor-bound form of P450cam, two energy minima below −24 kJ/mol were located for a water molecule inside the active site. One of them, which was labeled site D, is located in a hydrophobic region of the active site but is within hydrogen bonding distance of the camphor oxygen. This is not a favorable site for a water probe in the absence of camphor. In the presence of camphor, however, a water molecule in this position could donate a hydrogen bond of optimal geometry to the camphor oxygen. Thus, the hydrogen-bonding capacity of camphor would be fully satisfied and the water molecule would contribute, along with Tyr 96, to stabilizing the binding of camphor in the active site in the correct orientation for the regiospecific reaction. Fig. 1 *B* shows the arrangement of protein residues around an optimally positioned water molecule in the proposed site D.

It was proposed that this cavity could be filled by camphor analogues with a hydrophobic moiety added at the camphor carbonyl position, which could make favorable van der Waals contacts with the surrounding residues (Wade, 1990). With these additional interactions, such compounds were predicted to bind to P450cam with greater affinity than camphor itself. In addition, any water in this cavity would be pushed into the bulk solvent upon binding of such ligands, with a resultant gain in entropy. In agreement with this proposal, compounds designed to occupy this region have been found to bind more tightly than camphor (V. Helms, E. Deprez, E. Gill, C. Barret, G. Hui Bon Hoa, and R. C. Wade, unpublished data).

As stated in the previous section, a water molecule was not observed at site D in the crystal structure. This discrepancy between the GRID and the crystallographic studies indicates that the determination of the hydration state of an interfacial cavity such as site D is not simple. Thus, we here calculate the free energy of hydrating site D with a MD simulation method. We also calculate the free energy of placing the water 802, which provides a good example of a water molecule mediating a protein-ligand interaction, at its observed position in the complex of P450cam with 2pheIm.

Calculation of free energies of cavity hydration by MD simulations

The process of hydration of a cavity in a protein with a water molecule can be considered as two steps: first, the removal of one water molecule from bulk aqueous solvent and, second, the insertion of one water molecule into the cavity. The free energy changes associated with these steps can be calculated by using MD simulations together with the methods of thermodynamic perturbation and thermodynamic integration (TI) (Beveridge and DiCapua, 1989; Reynolds et al., 1992; Straatsma and McCammon, 1992; van Gunsteren and Mark, 1992; Kollman, 1993; van Gunsteren et al., 1993), which allow the calculation of nonphysical, so-called alchemical, processes during a MD simulation. We demonstrated this previously by calculating the free energies of hydrating two cavities in sulfate-binding protein, one of which was observed crystallographically to be solvated whereas the other was not (Wade et al., 1991). In agreement with experiment, we calculated the free energy of cavity hydration to be unfavorable in the case of the empty cavity for which no strong hydrogen bonds could be made and favorable in the case of the occupied cavity for which four hydrogen bonds to the water molecule could be made.

For the cavities studied here at protein-ligand interfaces in P450cam, we have followed the same basic approach. These cavities are, however, larger than those studied previously and this has necessitated the development and use of a more accurate and generally applicable protocol for calculating the free energy differences. This involves (1) the use of the multiconfigurational thermodynamic integration (MCTI) method (Straatsma and McCammon, 1991) and (2) the application of a flat-bottomed harmonic well (FBHW) potential to the solvent molecule removed or inserted during the simulations. These two features are outlined here.

The MCTI method

During a simulation to calculate a free energy difference, system A is transformed smoothly into system B by coupling the system Hamiltonian H to a coupling parameter λ that changes from 0 to 1.

$$H(\lambda) = (1 - \lambda)H(A) + \lambda H(B)$$

In the MCTI method (Straatsma and McCammon, 1991), the change in λ is carried out in discrete steps, which are referred to as windows. For each λ value or window, a MD

simulation consisting of an initial equilibration stage and a subsequent data collection phase is performed. Derivatives of $H(\lambda)$ with respect to λ are calculated for each window as the average from a representative statistical ensemble. The total free energy difference between the two systems, A and B, is obtained as the sum of the contributions of the single windows. The statistical error in $\partial H/\partial\lambda$ accumulated during a MCTI simulation can be calculated from the autocorrelation of $\partial H/\partial\lambda$ during data collection in each of the windows (Straatsma et al., 1986).

The FBHW potential

When a system is simulated without periodic boundary conditions, as is the case for our protein simulations, and the interactions of one particle with the rest of the system are completely removed, it is possible for the particle to escape from the simulated system. One method to prevent the particle from escaping is to assign it a very high mass as was done in our previous work (Wade et al., 1991). This method seemed suitable for small cavities in which the water molecule has little freedom to move, but tests with a mass of 9000 amu showed that it was not appropriate for the larger cavities studied here in which the sampling of the cavity space by the perturbed water was severely slowed by the artificial mass.

Hermans and Shankar (1986) used a harmonic restraint $U(r) = 0.5(1 - \lambda)K(r - r_0)^2$ in their study of xenon binding to myoglobin. The problem with this is that it restricts the space sampled by the perturbed water during the simulations. Therefore, we constructed a FBHW potential, which prevents the water from escaping from a specified spherical region but does not artificially perturb its motion when it is inside the spherical region.

$$U(r) = K(r - r_0)^2 \qquad \text{if} \quad r > r_0$$

$$U(r) = 0 \qquad \text{otherwise}$$

where r is the distance of the particle from the center of the FBHW, r_0 is the bottom radius, and K is a force constant. This potential was used for the simulations described here with a range of r_0 values between 0.02 and 0.8 nm. It permits the calculation of the free energy of hydrating a cavity by a water molecule that has considerable freedom to move.

In the next section, we present the parameterization and set-up of the simulated systems and the protocols for the MD simulations. The results of the free energy calculations for the complexes of P450cam with camphor and with 2pheIm and for simulations of bulk water systems are given in the subsequent section. Finally, the results of the simulations and possible implications for the design of ligands are discussed.

METHODS

Coordinates

The crystallographic coordinates of cytochrome P450cam, complexed with camphor at 0.163 nm resolution (Poulos et al., 1987) and P450cam complexed with 2pheIm at 0.21 nm resolution (Poulos and Howard, 1987), were taken from the files 2cpp and 1phe, respectively, of the Brookhaven Protein Data Bank (Bernstein et al., 1977). The structures consist of residues 10–414 and do not contain coordinates for the first 9 amino acids. An amino-terminal blocking group was added to residue 10 and the first 9 residues were omitted for the simulations. The coordinates of the missing atoms of Lys216 were modeled with the QUANTA software package (Molecular Simulation, Inc., Waltham, MA). If, as for 1phe, two alternative positions for residue side chains were given in the crystal structure, we used set A. Crystallographic water 515 was replaced by a potassium ion because it is octahedrally coordinated by four protein backbone oxygens and two water molecules and specific binding of K^+ to P450cam has been demonstrated experimentally (Di Primo et al., 1990).

The environment of all aspartate, glutamate, and histidine residues was analyzed graphically to determine their protonation states. Three histidines were assigned as doubly protonated: His62, His270, and His355. The side chain of Asp297 lies 0.27 nm from one of the heme propionate group oxygens (Poulos et al., 1987). To carry out the simulations, either the propionate group or the aspartate side chain must be protonated, although in reality both functional groups could share a common proton. We tried both possibilities in test simulations and found that protonation of Asp297 led to geometries closer to the crystal structure. Hydrogen atoms were added to the polar protein atoms and the 203 crystallographic waters with the ARGOS program (Straatsma and McCammon, 1990). The positions of Tyr, Thr, Ser, and water hydrogens were adjusted by graphical inspection to optimize hydrogen bond formation.

Parameterization

The standard 37D4 GROMOS force field (van Gunsteren and Berendsen, 1987) was used with an SPC/E water model (Berendsen and Grigera, 1987). Polar hydrogens were modeled explicitly; all aliphatic carbons were modeled as united atoms.

The heme group and the camphor and 2pheIm ligands are shown in Fig. 2. The additional parameters that were constructed for them are listed in Table 1. The bonded parameters of the heme moiety and the axial cysteine ligand were taken from the 2cpp crystal structure. Improper dihedrals were added to maintain the planarity of the heme during the simulation. The Lennard-Jones nonbonded parameters for Fe were taken from Collins and co-workers (1991). MD simulations (20 ps) of the substrate-free P450cam structure with a sixth ligand water molecule were performed for five different combinations of Fe and N charges and the remaining charge was distributed over the porphyrin ring. The set of partial atomic charges that resulted in the closest agreement of the average Fe-water oxygen distance (0.201 nm) with the experimental crystal structure value (0.228 nm) was chosen.

For camphor, missing bonded parameters were modeled based on the 2cpp crystal structure and partial atomic charges were assigned by analogy to those for amino acid residues in the GROMOS force field, following the common procedure for the use of this force field. The partial atomic charges used result in a dipole moment of 2.24 D, which is lower than the experimental value of 2.97 D measured in cyclohexane (Crossley et al., 1968).

Bond and bond angle parameters for 2pheIm were taken as average values from the following three x-ray structures in the Cambridge Structural Database (Allen et al., 1979) that contain 2pheIm: #SORBAC (1), #ACMPIM10 (2), and #GAFJOM (3). The dihedral angle between the two rings could not be taken from these three crystal structures, as the dihedral angle has a range of values (121.4° and −135.0° for the two units of (1), −164.4° (2), and 151.4° (3)), whereas the two rings are almost planar (−176.3°) in the 1phe crystal structure. A calculation of the torsional

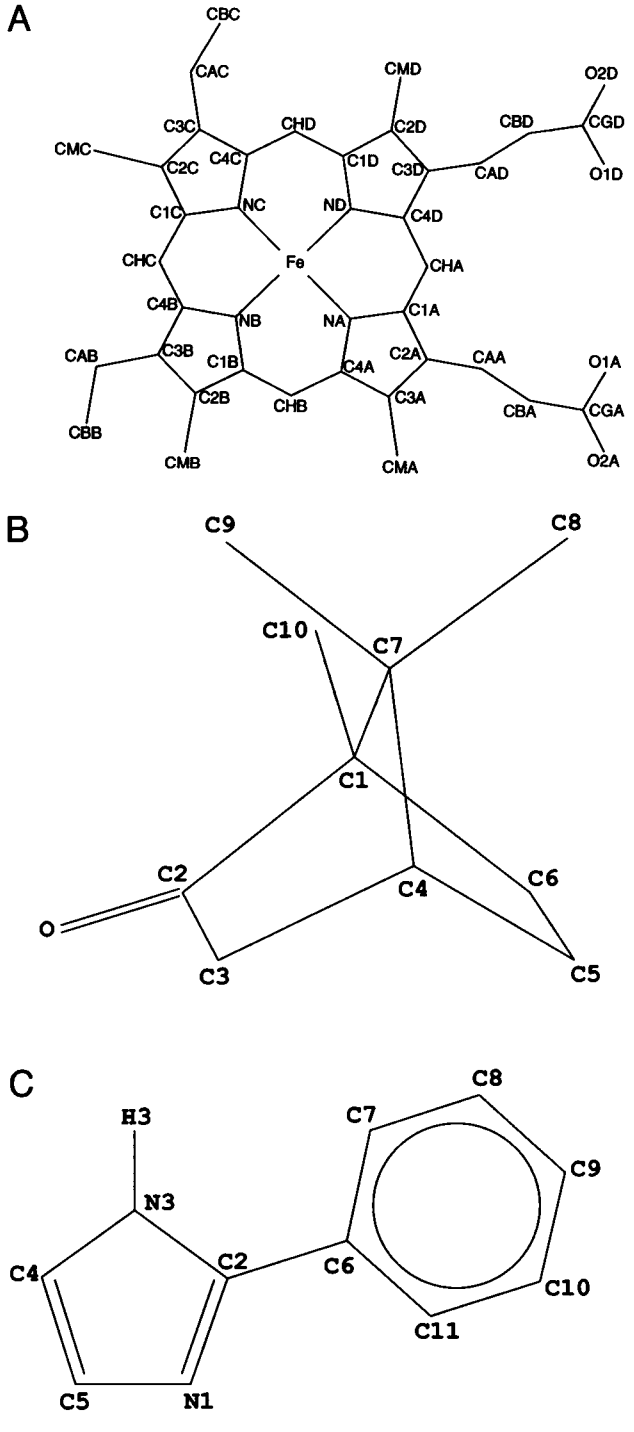

FIGURE 2 Sketch of (*A*) the heme, (*B*) camphor, and (*C*) 2pheIm.

barrier around the bond between the two rings with the MNDO semi-empirical method implemented in MOPAC (Dewar and Thiel, 1977) resulted in two energy minima at 90° and 270° with a barrier height of 8.8 kJ/mol. This is in good agreement with an MM2 study for 5-phenylimidazole that gave a torsional barrier of 7.7 kJ/mol, also with minima at 90° and 270° (Garduno-Juarez and Ocampo-Garcia, 1989). As there is no analogous chemical fragment in the GROMOS parameter set, partial atomic charges for 2pheIm were assigned from calculations using the MNDO/ESP method (Besler et al., 1990).

Set-up of simulations

The MD simulations were done with versions 5.811 and 6.0 of the ARGOS program (Straatsma and McCammon, 1990).

The solvent-solvent nonbonded pair list was updated every 10 steps, the solvent-solute list every 15 steps, and the solute-solute list every 20 steps. A twin-range nonbonded cutoff was applied in all simulations; all nonbonded forces within 0.8 nm were calculated every step whereas interactions between 0.8 and 1.0 nm were updated only every fifth step. All bond lengths were constrained by the SHAKE algorithm (Ryckaert et al., 1977) and a time step of 2 fs was used.

Protein systems

For simulations of both the camphor and the 2pheIm complexes, a 3.2-nm sphere of solvent (constructed from a preequilibrated box of SPC/E water (Berendsen and Grigera, 1987)) was superimposed on the protein and centered on the oxygen of the water molecule to be perturbed. In the P450cam-camphor complex, this water molecule was positioned at the energy minimum (site D) located by the GRID program (Wade, 1990) (see Fig. 1 *B*). During equilibration, this water molecule was treated as a dummy atom without electrostatic and Lennard-Jones interactions. It was named DUW (for dummy water). All added water molecules closer than 0.26 nm to any nonhydrogen atom of the protein were removed. Four additional water molecules that were superimposed on a hydrophobic region of the active site were also excluded from the simulations. In the case of 2pheIm, this resulted in a system of 11,457 atoms including 2,500 solvent molecules.

The same heating and equilibration protocol was used for both complexes. The added hydrogens, Lys216 atoms and the amino-terminal blocking group were subjected to 100 steps of steepest descent energy minimization (SD-EM) whereas all other atoms were held fixed. Then, all water molecules between 2.2 and 3.2 nm from the central water molecule were restrained to their positions with a force constant of 2000 kJ/nm^2 × mol and the added water molecules within 2.2 nm were subjected to 100 steps of SD-EM along with the hydrogens and modeled atoms. This minimization was followed by an additional 100 steps of SD-EM in which the crystallographic waters were restrained to their crystallographic positions with a force constant of 2000 kJ/mol × nm^2 rather than being fixed. The same atoms were allowed to move during three successive MD simulations under constant particle number, volume, and temperature (NVT) conditions at 100 K, 200 K, and 300 K, after which they were again energy minimized with 100 steps SD-EM.

The protein and solvent atoms were then heated at the same time. First, all atoms within 1.8 nm of the center were allowed to move freely and the remaining atoms were fixed, and 100 steps of SD-EM were performed. Then all atoms within 1.2 nm of the center were assigned to be free and atoms between 1.2 and 1.4 nm were restrained to their crystallographic positions by a weak restraining force of 250 kJ/nm^{-2} mol^{-1}. The rest of the system was held fixed. If any atom of a charge group was lying outside a restraint distance range, all atoms of the charge group were either restrained or fixed. All water molecules beyond 2.8 nm from the center were removed. This led to a system of 5574 atoms for the 2pheIm system. A total of 20 water molecules and 394 solute atoms were completely free to move and 19 water molecules and 236 solute atoms were restrained. The heating phase consisted of 5 ps of MD each at 100 K, 200 K, and 300 K under constant particle number, volume, and energy (NVE) conditions. New atomic velocities were assigned every 0.2 ps acccording to a Maxwell-Boltzmann distribution, and 100 ps of simulation followed at 300 K under NVT conditions. The solute and solvent parts of the system were separately coupled to a Berendsen thermostat (Berendsen et al., 1984) at 300 K with coupling times of 0.1 and 0.4 ps, respectively. Equilibration was monitored by inspection of the individual energy components (data not shown).

To check the effect of limiting the moving part of the system to a sphere of 1.2-nm radius, we also constructed a bigger protein system for the 2pheIm case. Starting from the system obtained after the solvent-only

TABLE 1 Force field parameters used in MD simulations of the complexes of cytochrome P450cam with camphor and with 2pheIm

A: Parameters for the heme with a cysteine ligand

Bond parameters	r_e (nm)	k (kJ/nm^{-2} mol^{-1})
Fe-N	0.209	4.18×10^5
Fe-SG	0.220	4.18×10^5

Angle parameters	θ	k (kJ/rad^{-2} mol^{-1})
SG-Fe-N	102°	418
CB-SG-Fe	109.5°	251
N_i-Fe-N_{i+1}	90°	418
N_i-Fe-N_{i+2}	155.6°	418
C1-N-C4	108°	376
N-C1-C2, C1-C2-C3, N-C4-C3, C2-C3-C4	108°	418
N-C1-CH, C2/C4-C3-CM, CH-C1-C2, C1/C3-C2-CA, N_i/$C3_i$-$C4_i$-CH_{i+1}, $C4_i$-CH_{i+1}-$C1_{i+1}$, Fe-N-C1/C4	126°	418
C2-CA-CB, CA-CB-CG	111°	418
CB-CG-O1/2	117°	502
O1-CG-O2	126°	502

Proper dihedral parameters	n	k (kJ/rad^{-2} mol^{-1})
C2-C3-CA-CB	2	0.42

Improper dihedral parameters	ϕ	k (kJ/rad^{-2} mol^{-1})
Fe-C1-C4-N	8.6°	167

Improper dihedrals on pyrrole rings as for histidine, improper dihedrals on propionate groups as for aspartate in the GROMOS87 force field

Lennard-Jones parameters	A (kJ nm^{-12} mol^{-1})	B (kJ nm^{-6} mol^{-1})
Fe[¶]	4.30115×10^{-5}	8.86909×10^{-3}
K[‡]	1.06892×10^{-3}	8.8958×10^{-3}

Partial atomic charges	q (e)
Fe	1.0
N	−0.4
CH	0.1
CM	0.2
CG	0.27[§]
O1, O2	−0.635[§]
SG	−0.5
CB	−0.1

B: Parameters for camphor

Atom label	*GROMOS atom types*
C1, C2, C7	C
C4	CH1
C3, C5, C6	CH2
C8, C9, C10	CH3
O	O

Bond parameters	r_e (nm)	k (kJ nm^{-2} mol^{-1})
C1-C2, C1-C6, C1-C7, C4-C7, C7-C8, C7-C9, C1-C10	0.153	3.34×10^5

B: Parameters for camphor—*Continued*

Angle parameters	θ	k (kJ rad^{-2} mol^{-1})
C1-C2-C3, C6-C1-C2, C3/C5-C4-C7	109.5°	460
C1-C7-C4	90°	460
C1-C2-O	121°	502
C1-C6-C5, C6-C1-C7, C1/C4-C7-C8/C9, C8-C7-C9, C2-C1-C7, C2/C6/C7-C1-C10	111°	460

Improper dihedral parameters	ϕ	k (kJ rad^{-2} mol^{-1})
C2-C1-C3-O	0°	502

Partial atomic charges[‖]	q (e)
C2	0.38
O	−0.38
Other atoms	0

C: Parameters for 2pheIm

Atom label	*GROMOS atom types*
N1, N3	NR5*
C2	CB
H3	H
C4, C5	CR51
C6-C11	CR61

Bond parameters	r_e (nm)	k (kJ nm^{-2} mol^{-1})
N1-C2	0.131	4.18×10^5
C2-N3	0.138	4.18×10^5
N3-C4	0.138	4.18×10^5
C4-C5	0.134	4.18×10^5
C5-N1	0.139	4.18×10^5
N1-C6	0.147	4.18×10^5

Angle parameters	θ	k (kJ rad^{-2} mol^{-1})
N1-C2-N3, C4-C5-N1	111°	418
N1/N3-C2-C7	124.5°	418
C2-N3-C4, N3-C4-C5	108°	418
C2-C6-C7/C11	120°	418

Proper dihedral parameters	n	k (kJ rad^{-2} mol^{-1})
N1-C2-C6-C7	2	2.09

Improper dihedral parameters

On imidazole ring as for histidine, on phenyl ring as for phenylalanine in the GROMOS87 force field

Partial atomic charges	q (e)
N1, N3	−0.4
C2, H3	0.3
C4, C5	0.1
Others	0

[¶] Taken from Collins et al., 1991.

[‡] Taken from Åquist, 1990.

[§] Taken from the aspartate residue of the GROMOS87 force field.

[‖] Assigned with the same charges as equivalent groups in proteins in the GROMOS87 force field; they are also the same as used by (Paulsen and Ornstein, 1991).

equilibration stage, the protein and solvent were energy minimized with all atoms within 2.2 nm of the center allowed to move. The heating stages were done with all atoms within 2.0 nm free, atoms between 2.0 and 2.2 nm restrained with a force constant of 250 kJ/nm^{-2} mol^{-1}, and all remaining atoms fixed. As for the small system, 100-ps MD equilibration were then performed at 300 K under NVT conditions. The number of freely moving atoms was 2317 for the big system (204 water molecules and 1704 solute atoms free and 111 water molecules and 359 solute atoms restrained). The amount of computing time necessary for the big system was three times as great as that for the small system. However, analysis of the individual components of the system energy (data not given) showed that, in contrast to the small system, the big protein system was not fully equilibrated after 100 ps of MD simulation at 300 K. Therefore, a detailed comparison of atomic mobilities and torsional flexibilities was not justified. In general, the atomic mobilities were quite similar: for the big system, the root mean square (RMS) fluctuations of most active site residues had a broad uniform distribution of 0.04–0.05 nm, whereas for the small system the differences between single residues were more pronounced with values between 0.029 and 0.070 nm. The RMS fluctuations for 2pheIm and water 802, e.g., were 0.088 and 0.072 nm for the small system and 0.071 and 0.088 nm for the big system. This indicates that the restriction of the moving system to a sphere of 1.2 nm gives a reasonable description of the atomic mobilities of the active site residues.

In addition, the fluctuations of the torsional dihedrals on the protein backbone for the smaller system were approximately 8–10°. This can be compared with a MD simulation of the camphor-bound P450cam complex in vacuo (Paulsen et al., 1991) where only crystal waters were included, but all atoms were treated as dynamic, which showed average RMS fluctuations of 12.9° for the backbone dihedral angles. Moreover, Harris et al. (1995) found that similar hydroxylation ratios could be predicted from MD simulations of the full protein and from simulations in which only the active site region was dynamic.

Large changes in pressure during the simulations could influence the free energies calculated. To ensure that such pressure abnormalities were absent, we monitored the energy due to the harmonic restraint potential applied to atoms in the restrained shell region. If, for example, the pressure in the inner dynamic region was too high, one would expect that the removal of a water molecule would be followed by a relaxation of the atoms in the restrained shell. In our simulations, no drift in the restraint energy was observed, either for the removal of a water molecule from the 2pheIm complex or for the introduction of a water molecule into the camphor complex.

Bulk water systems

Two systems were constructed by replication of a preequilibrated box of 216 water molecules (298 K, constant particle number, pressure, and temperature (NPT) conditions):

(1) A cubic box of 461 water molecules of 2.4-nm box dimensions, treated with periodic boundary conditions; 200 ps MD of equilibration were performed under NVT conditions at 300 K.

(2) A 2.6-nm radius sphere of water molecules. Molecules in the inner 1.2 nm were allowed to move freely, molecules in the region between 1.2 and 1.4 nm from the center were restrained to their starting positions by a harmonic potential with a force constant of 2000 kJ nm^{-2} mol^{-1}, and the remaining waters were fixed. The system contained 2459 water molecules in total; 238 molecules were in the inner region and 46 molecules were in the restrained region. Although the spherical system has a smaller number of freely moving atoms, the amount of computing time necessary was larger than for the cubic box by a factor of 2.1 because a large number of nonbonded interactions with the fixed atoms had to be calculated.

The spherical system with extended wall boundary conditions is equivalent in dimensions to the systems set up for the protein calculations. One characteristic of extended wall boundaries is that, during processes of particle insertion or removal, work is done against the restraining potential on atoms in the restraint region (van Gunsteren et al., 1993). This undesirable contribution to a calculated free energy can, in principle, be reduced by cancellation of errors on combining results for a spherical water system with the results for the spherical protein systems. Yet in a spherical water system, it is possible for the water molecule that is mutated to move close to the border region and this could introduce an artifact because of the anisotropic motion of the water molecule's surroundings in such configurations. To separate the effect of the extended wall boundary from that of the FBHW potential, all simulations to study the dependence of the calculated free energy difference on the size of the FBHW bottom radius were done for the cubic water box. The comparison of free energy differences calculated in the spherical and cubic solvent systems allows the effects of any artifacts due to the spherical boundary region to be detected. For the cubic box, two different starting configurations for the MCTI calculations were used, both starting from the coordinates and velocities obtained after 200-ps NVT equilibration. The first configuration, A, was with the water molecule to be removed, named WAT, 0.41 nm from the edge of the box. For bottom radii up to 0.3 nm, the FBHW was centered on WAT, for a bottom radius of 0.3 nm it was shifted by 0.1 nm toward the center of the box and for a bottom radius of 0.5 nm it was shifted by 0.2 nm. The second configuration, B, was with the water molecule closest to the center of the box assigned as WAT and the FBHW centered on WAT. Both starting configurations can be considered as equally valid.

Free energy calculations

During the perturbation simulations, the interaction of the perturbed water molecule with its surroundings was gradually switched on or off by coupling it linearly to variable λ during each simulation. Lennard-Jones and electrostatic nonbonded interactions were coupled to λ simultaneously. Electrostatic decoupling, as used in our previous work (Wade et al., 1991) to prevent the problems associated with simulating perturbed charged atoms when they have small radii, was not necessary. Instead, we used a recently developed separation-shifted scaling method (Zacharias et al., 1994) for the bulk water and 2pheIm simulations that enables singularities to be avoided when the interactions of the perturbed atoms become very small. Shifting of the interactions was done with the adjustable parameter set to 0.05 nm.

MCTI calculations (Straatsma and McCammon, 1991) were performed for 21 windows with λ values of 0.0, 0.05, $\cdots$, 1.0. For the protein systems, each window consisted of 1000 MD steps of equilibration followed by 1000 steps of data collection. Dynamic windows were used for data collection for the bulk water systems with the number of data collection steps varying between 2000 and 5000 steps. Data collection for a window was stopped when both of the following criteria were met: statistical error from data autocorrelation of $\partial H/\partial\lambda < 5$ kJ mol^{-1} and drift in $\partial H < 5$ kJ mol^{-1} during the window length. The convergence of the simulations was monitored by following the free energy change versus the length of the data collection window (Zacharias et al., 1994). For the protein simulations, coordinate sets were recorded every 100 steps during data collection to analyze the extent of spatial sampling of the perturbed water molecule.

In parallel with the MCTI analysis, a multi-step thermodynamic perturbation analysis was performed. The results of both analyses showed no significant differences (data not shown).

Control over the mobility of the perturbed water

The force constant K for the FBHW potential was taken as 2000 kJ nm^{-2} mol^{-1}, which is a typical restraining force in MD simulations. The influence of the size of the bottom radius on the calculated free energy was investigated by varying the bottom radius from 0.02 to 0.5 nm in the bulk water periodic box system.

Application of the FBHW in the spherical bulk water system was also tested. A 60-ps equilibration at 300 K under NVT conditions was performed. Then the closest water to the center of the sphere was assigned as WAT. During two independent MCTI calculations, WAT was removed by applying the FBHW potential with a bottom radius of 0.8 nm (21 $\times$

(1000 + 3000) steps) and with a bottom radius of 0.4 nm (21 × (1000 + 2000) steps).

Energy calculations with GRID and GROMOS energy functions

The interaction energies for a water probe in the 2cpp-DUW position and the 1phe-WAT802 position were calculated with version 11 of the GRID program. Interaction energies for the GROMOS force field were calculated for these positions in the crystal structures of the proteins after minimizing and equilibrating the added hydrogens and the added solvent with the ARGOS program. The position of the water molecule was minimized, keeping all other atoms fixed, and the nonbonded energy between the water molecule and the rest of the system was calculated.

Calculation of accessible volume for water molecules during MD simulations

The spread in the positions of a water molecule in superimposed snapshots taken at 0.5-ps intervals during a MD simulation was considered as a representative measure of the accessible conformational space of the water molecule during the simulation. A van der Waals surface was constructed around the resultant water blob (water oxygen positions) and its volume was calculated with the GRASP program (Nicholls and Honig, 1992).

RESULTS

Bulk water system

The excess free energy (before correction for polarization) of an SPC/E water molecule allowed to move without any restraints in a periodic box of water was calculated to be -30.4 ± 0.9 kJ mol^{-1} in an MCTI calculation of 21 × (1000 + 5000) steps (total simulation time of 252 ps). This result was taken as a reference point for the following simulations in which a FBHW restraint potential was applied to the perturbed water molecule.

Simulations with the FBHW potential

Fig. 3 shows the results for a simulation under periodic boundary conditions which was started from starting structure A with a 0.5-nm bottom radius. In this simulation, the calculated ΔF was 30.1 ± 1.2 kJ mol^{-1} after 21 × (1000 + 2000) steps (total of 126 ps of simulation) and 31.3 ± 1.1 kJ mol^{-1} after 21 × (1000 + [2000 − 5000]) steps (160.4 ps) (see Fig. 3 *A*). This example demonstrates the improvement in the accuracy of the statistical error and the convergence of the free energy as sampling is increased. Application of dynamic windowing successfully prevented unnecessary data collection as only windows 1–5, 7, 11, 13, and 15 were extended to the maximal number of 5000 data collection time steps. This particular simulation therefore lasted only 160.4 ps instead of 252 ps for a simulation with 21 × (1000 + 5000) steps.

Fig. 4 summarizes the results for all simulations with a FBHW potential and shows the dependence of the calculated total ΔF value upon the size of the bottom radius for

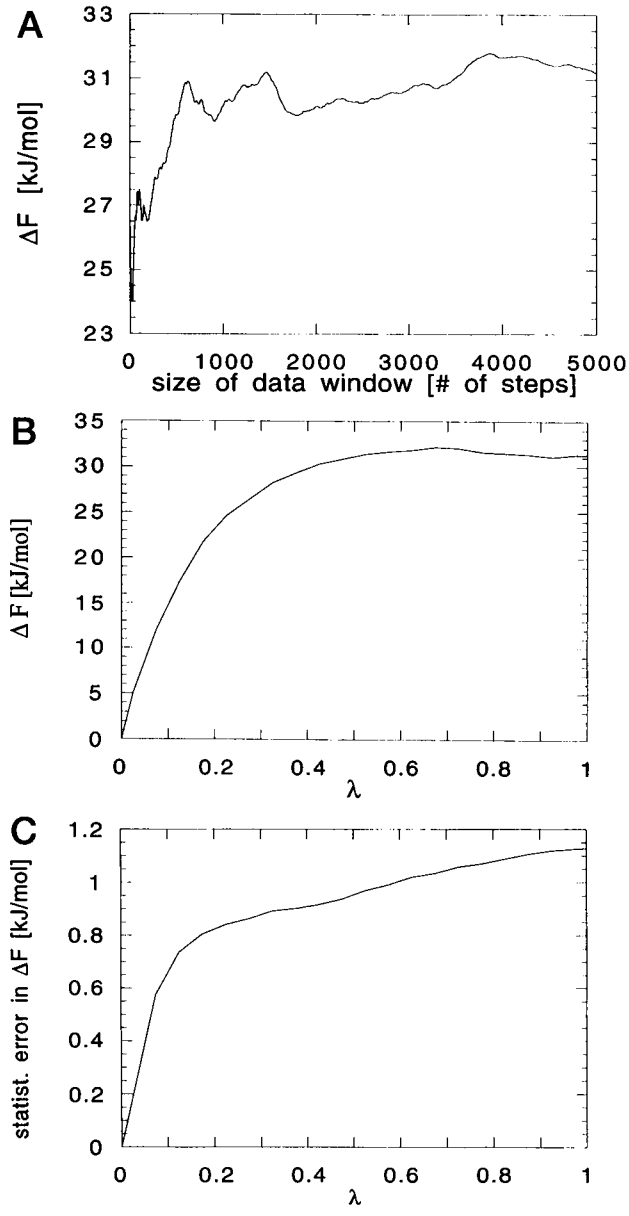

FIGURE 3 Calculated free energy changes (ΔF) during a MCTI simulation in which a water molecule was removed from a box of 460 water molecules under periodic boundary conditions with a FBHW potential with a bottom radius of 0.5 nm. (*A*) ΔF versus the number of time steps in the data-gathering window, (*B*) cumulative ΔF versus λ, and (*C*) cumulative statistical error in ΔF versus λ.

the A and B starting configurations. The results for A and B differ by approximately 3 kJ mol^{-1} for bottom radii of 0.1 and 0.5 nm, which is far more than the statistical error of the single simulations (approximately 1 kJ mol^{-1}). To account for this, a mean value for ΔF was calculated from all of the simulations with different bottom radii as ΔF = 30.3 kJ mol^{-1} with a standard deviation of 1.7 kJ mol^{-1}. The average statistical error during the simulations was 1.1 kJ mol^{-1}. The two sources of error were then simply added, as they can be considered to be independent of each other. One

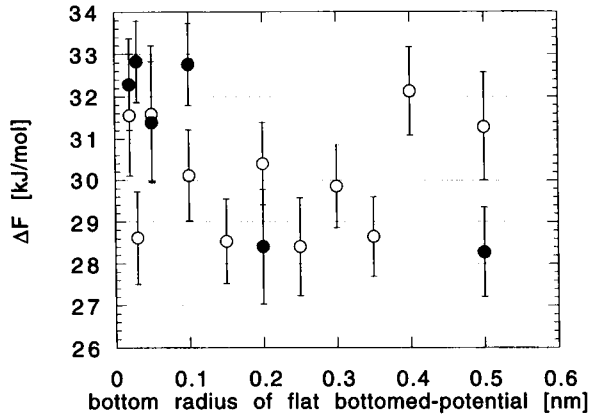

FIGURE 4 The excess free energy ΔF calculated for removing one water molecule, WAT, from a periodic box of 460 water molecules versus the bottom radius of the flat-bottomed harmonic well FBHW potential applied to WAT. Two different starting coordinates and velocities were used: In A, WAT was located at its position after 200-ps NVT equilibration (○); in B, WAT was exchanged with the closest water molecule to the center of the box after equilibration (●).

describes the inherent statistics of each simulation; the other describes differences between separate simulations due to different starting conditions (A or B) and different simulation conditions (size of FBHW radius). This gave an overall ΔF for SPC/E water of 30.3 ± 2.7 kJ mol^{-1} for a broad range of bottom radii. These results are summarized in Table 2.

Two simulations were carried out with zero bottom radius ($r_0 = 0$), one in which the water oxygen was restrained to its starting position by a 200-kJ nm^{-2} mol^{-1} harmonic restraint and the other in which it was fixed in this position. The resultant free energies were ΔF = 32.8 ± 1.1 kJ mol^{-1} and 32.3 ± 1.1 kJ mol^{-1}, respectively. Comparable results were obtained from simulations with an unphysically heavy mass on the perturbed water oxygen. Results for 21 × (1000 + [2000 < x < 5000]) window simulations are 31.4 ± 1.1 kJ mol^{-1} for a 64 amu mass on the oxygen, 31.2 ± 1.1 kJ mol^{-1} for 500 amu, 33.8 ± 1.0 kJ mol^{-1} for 1000 amu, and 33.7 ± 1.1 kJ mol^{-1} for 9000 amu.

The results for the removal of a water molecule from a 2.6-nm sphere of water molecules were ΔF = 28.2 ± 1.3 kJ mol^{-1} with a 0.8-nm bottom radius and 29.8 ± 1.5 kJ mol^{-1} with a 0.4-nm bottom radius. They are within the error range of the results for the cubic box.

Complex of P450cam with camphor

First, test simulations were performed starting from the crystal structure of P450cam with the addition of a water molecule at the GRID energy minimum site D. When the water molecule was not restrained, we observed that, after the 15-ps NVE heating stage, the hydrogen bond between Tyr96 and camphor was broken after 3 ps of the NVT equilibration. For approximately 80 ps, two intermediate hydrogen bonds were formed; the water molecule accepted a hydrogen bond from Tyr96 and donated a hydrogen bond to the camphor carbonyl oxygen. But these hydrogen bonds were present only 83% (Tyr-DUW) and 64% of the time (DUW-Cam) and finally the camphor keto group turned away from these hydrogen bond partners. This is in clear contrast to the experimental finding of a strong hydrogen bond between camphor and Tyr96. Within our parameterization, camphor and the SPC/E water molecule have approximately the same total dipole moment, but the dipole along one of the water O-H bonds is approximately 1.5 times that along the camphor-carbonyl bond. This makes water an energetically more favorable hydrogen bond partner for Tyr96 than camphor. Transient occupation of region D by a crystallographic water molecule has also been observed in a MD simulation of P450 cam with l-norcamphor (Harris and Loew, 1995).

The next step was to introduce a water molecule gradually into site D with the MCTI procedure. Fig. 5 summarizes the results of the different simulations and shows the dependence of the calculated total ΔF for the insertion of a water molecule into site D of the camphor-bound P450cam upon the size of the bottom radius. Graphical inspection of the recorded coordinate sets of the simulations with the larger bottom radii (in the range 0.2–0.5 nm) showed that the water molecule drifted away from its starting position once it had picked up some kinetic energy (it was inserted with zero velocity) and moved away from camphor into the hydrophobic region of the cavity.

We divided the calculated values into two classes (small and large) according to whether the bottom radius was less

TABLE 2 Free energy ΔF in kJ mol^{-1} for the insertion of a water molecule into solvent and protein environments

	SD*		Statistical error‡		ΔF§		ΔF_{hydr}‖	
Bulk water	1.7		1.1		−30.3 ± 2.7			
Size of bottom radius of FBHW potential	Large	Small	Large	Small	Large	Small	Large	Small
2cpp: site D water	0.6	1.2	1.7	0.6	−14.5 ± 2.3	−24.9 ± 1.8	15.8 ± 5.0	5.4 ± 4.5
1phe: water 802	2.4	0.6	1.5	0.9	−41.9 ± 3.9	−53.0 ± 1.5	−11.6 ± 6.6	−22.7 ± 4.2

* SD of the ΔF values in the separate simulations from the mean value for ΔF.

‡ RMS value of the statistical errors in each of the simulations.

§ mean of the ΔF values calculated in the separate simulations. Errors given are the sum of SD and the statistical error.

‖ $\Delta F_{hydr} = \Delta F_{protein\ site} - \Delta F_{water}$ is the free energy of hydration of the site in the protein.

than or greater than 0.09 nm, choosing this criterion from the grouping of the energy values apparent in Fig. 5. Errors were evaluated as for bulk water. The results are shown in Table 2.

Complex of cytochrome P450cam with 2phelm

When comparing the average structure during equilibration with the crystallographic starting structure, the most obvious difference was the relative orientation of the two rings of 2pheIm, which are almost planar in the crystal structure. As we had modeled the torsional dihedral of 2pheIm with minima at 90° and 270° and as there were no obvious steric constraints, 2pheIm assumed a twisted structure with the two rings at 90° in the simulation. This is in contrast to the crystal structure, but the phenyl ring is neither in contact with the hydration site nor is it making direct contacts with any active site residue except the heme in either of the two conformations.

The strong hydrogen bond coordination of water 802 observed in the x-ray crystal structure weakened slightly during the MD equilibration. Two hydrogen bonds were well preserved; the hydrogen bond with Asp251 became a bit longer (average water oxygen-Oδ2 distance during the 100-ps NVT simulation at 300 K was 0.30 nm compared with 0.26 nm in the x-ray structure) and the one with 2pheIm a bit shorter (average N3-water oxygen distance of 0.30 nm compared with 0.31 nm). But the hydrogen bond to Thr185 was present in only 20% of the 200 saved coordinate sets (average Oγ-water oxygen distance of 0.39 nm compared with 0.30 nm). As in the simulation of the camphor complex, Thr185 turned slightly and donated a hydrogen bond to Asp251. Also, the weak hydrogen bond between Val247 and water 802 in the crystal structure was present only 28% of the time during the simulation (average carbonyl O-water oxygen distance of 0.40 nm compared with 0.34 nm).

The calculated atomic fluctuations for water 802 of 0.072 nm during the MD equilibration correspond to a B-factor of 0.14 nm^2, which compares reasonably with the experimental B-factor of 0.25 nm^2.

Fig. 6 summarizes the results for the MCTI calculations with the FBHW potential and shows the dependence of the calculated final ΔF difference for the removal of water 802 upon the size of the bottom radius. Analysis of ΔF versus λ (data not given) showed that the free energy results for different bottom radii started to diverge in the first few windows when the interactions of the water molecule were only slightly modified.

As in the case of camphor, we divided the calculated ΔF values into two classes (small and large). In this case, the boundary between the two classes, chosen on the basis of the grouping of the calculated energies, was a bottom radius of 0.11 nm. The results are shown in Table 2.

Interaction energies for ideal water geometries in the crystal structures

The interaction energies of the DUW at site D in the crystal structure of the camphor complex and the crystallographic water 802 in the crystal structure of the 2pheIm complex were calculated for the GROMOS and GRID energy functions. The energies for DUW were −19.6 kJ mol^{-1} (GROMOS) and −31.8 kJ mol^{-1} (GRID). The energies for WAT802 were −45.6 kJ mol^{-1} (GROMOS) and −47.2 kJ/mol (GRID). In both cases, the GRID interaction energy was more favorable than the GROMOS energy.

Accessible volumes of conformational space for water molecules

During the MD equilibration of the 2pheIm complex, which was done without any restraining FBHW potential, the volume inside the van der Waals surface around WAT802 (calculated from snapshots from the 30- to 100-ps interval

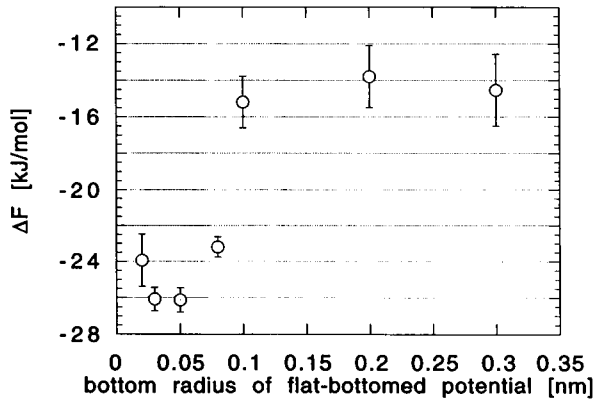

FIGURE 5 The excess free energy ΔF calculated for inserting a water molecule into site D in the active site of cytochrome P450cam complexed with camphor versus the bottom radius of the FBHW potential.

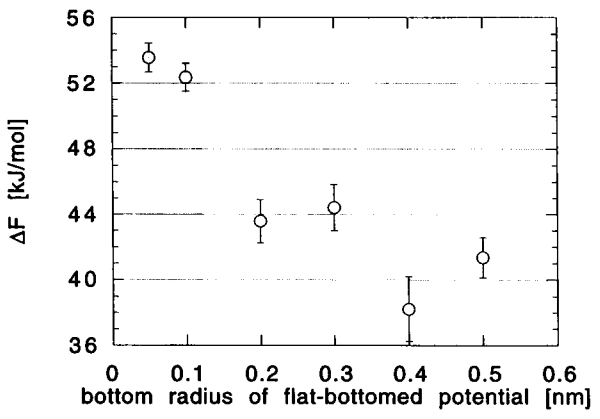

FIGURE 6 The excess free energy ΔF calculated for removing crystallographic water 802 from the cytochrome P450cam-2pheIm complex versus the bottom radius of the FBHW potential.

of the 300 K NVT trajectory), was calculated as 0.070 nm^3. The corresponding volumes during MCTI calculations, where a restraining FBHW potential was applied, were 0.059 nm^3 for the 0.05-nm bottom radius FBHW potential and 0.080 nm^3 for the 0.1-nm bottom radius FBHW potential. The values for larger bottom radii increased rapidly up to 0.353 nm^2 for a 0.5-nm bottom radius.

For the camphor system, values could be calculated reasonably only for the MCTI calculations. For the 0.05-nm bottom radius FBHW potential, the volume was 0.063 nm^3, which is similar to that for the same bottom radius in the 2pheIm complex.

DISCUSSION

Bulk water system

Although many simulations have been reported previously in which the excess chemical potential of water was calculated, the simulations of the bulk system are interesting for several reasons. First, they provide a good calibration point because the experimental value of the excess chemical potential of water is known. Second, our methodology can be compared with other methods of calculating the excess chemical potential of water. Third, the bulk water system provides a very good test of the properties and usefulness of the FBHW potential, because simulations with and without the FBHW potential can be compared. In simulations of the insertion or removal of a water molecule in a spherical system, special devices, such as the FBHW potential, have to be employed to prevent the escape of the water molecule from the system, whereas simulations for a periodic system can be performed both with and without such devices.

The free energy change on removing a SPC/E water molecule from a periodic box was calculated as -30.4 ± 0.9 kJ mol^{-1} from a single MCTI calculation without a FBHW potential. Use of the separation-shifted potential (Zacharias et al., 1994) guaranteed a smooth accumulation of ΔF as λ increased for all the bulk water simulations. The SPC/E water model includes implicit polarization (Berendsen and Grigera, 1987) and therefore the above value has to be corrected by a polarization contribution of +5.2 kJ mol^{-1} (Berendsen and Grigera, 1987) to yield the excess chemical potential of SPC/E water. The corrected value for SPC/E water of -25.2 ± 0.9 kJ mol^{-1} agrees well with the experimental value of the excess chemical potential of water of -26.4 kJ mol^{-1} (Ben-Naim and Marcus, 1986).

The above result may be compared with previous calculations. The Gibbs free energy of TIP4P water was calculated as -25.5 ± 1.3 kJ mol^{-1} by Monte Carlo simulations (Jorgensen et al., 1989). The free energy of hydration of SPC water was calculated by Hermans et al. (1988) as -23.0 kJ mol^{-1}, by Heiner (1992) as -24.1 kJ mol^{-1} by free energy simulations and as -26.4 kJ mol^{-1} by the Widom particle insertion method, and by Quintana and Haymet (1992) as -23.8 kJ mol^{-1} by Ewald summation. Wade et al. (1991) calculated the free energy of hydration of SPC/E water as -26.8 ± 1.7 kJ mol^{-1}.

The results for simulations of the spherical system were in good agreement with the simulations with periodic boundary conditions (both with a FBHW potential). All additional testing was therefore done for the periodic box that was computationally more efficient by a factor of 2.

Simulations of the removal of a water molecule from a periodic box, applying a FBHW potential of varying bottom radius produced ΔF values that agree well with the above result obtained without a FBHW potential. Indeed, the differences in the ΔF values obtained in simulations starting with different initial positions of the perturbed water molecule were as large as the differences obtained by applying different bottom radii (Fig. 4). We observed a weak tendency to higher ΔF values for the very small bottom radii (0.02 nm). A sphere of 0.02-nm radius is less than the thermally accessible volume of a water molecule at 300 K; a typical temperature factor of 0.30 nm^2 for water molecules bound to protein corresponds to RMS fluctuations of approximately 0.11 nm. A similar trend toward higher ΔF values was also observed when the water mobility was reduced by restraining the water oxygen to its starting position or increasing the oxygen mass. According to statistical mechanics, the thermodynamic properties of shapeless point particles are independent of their mass. However, this might not be the case for water molecules, which form dense hydrogen bond networks with strong directionality.

Ignoring the trend toward higher ΔF for small bottom radii, we calculated an average value of ΔF from the simulations using the FBHW potential with different bottom radii between 0.02 and 0.5 nm as 30.3 ± 2.7 kJ mol^{-1}. It has been previously recognized (see, e.g., Chipot et al., 1994) that statistical errors for single calculations underestimate the actual error of free energy calculations and that separate calculations from different starting structures can give a more complete idea of the involved error. To provide a cautious error estimate, we have assumed that the two sources of errors (error within a simulation and error between simulations) are independent.

Protein systems

The free energies of hydration of the protein sites are given in Table 2. Hydration of the water 802 site in the 2pheIm structure is favored by -11.6 ± 6.6 kJ mol^{-1} (large FBHW radius). Hydration of site D in the camphor structure is unfavorable; the hydration free energy is $+15.8 \pm 5.0$ kJ mol^{-1}.

Computational aspects

Comparison of the protein simulations with the bulk water simulations shows a marked difference in the influence of the bottom radius of the FBHW potential on the calculated free energies; it is of little importance in the case of the

isotropic water system but has a considerable effect in the nonuniform protein system. For both the hydration sites considered, the calculated ΔF is shifted by approximately 10 kJ mol^{-1} to a more negative value on going from large to small FBHW bottom radii (see Table 2). This indicates that the region immediately surrounding the initial positions that the water molecules were inserted at, i.e., the crystallographically observed location of water 802 and the GRID energy minimum in the camphor complex, are more energetically favorable than other regions in the cavities surrounding these positions. This is of course to be expected as the hydrogen bond geometries are optimized at the insertion positions. Comparison of the accessible volume for the water molecule in the 2pheIm complex during MD equilibration (0.070 nm^3) and during MCTI calculation shows that the water mobility was restricted only by the flat-bottomed potential when its bottom radius was equal to or less than 0.1 nm. This radius corresponds to that for distinguishing between small and large bottom radii on the basis of calculated free energy.

Conformational changes seen during the standard MD simulations are consistent with the calculated hydration free energies. Water 802 remained bound in a stable conformation in the 2pheIm complex, whereas the additional water molecule in the camphor complex caused deviations from the crystallographic structure by pushing the camphor molecule away from Tyr96.

The interaction energies calculated with the GRID and GROMOS energy functions are negative (favorable) for both site D and water 802. The GRID and GROMOS interaction energies are almost identical for the water 802 site, whereas for site D, the GRID energy is far more favorable than the GROMOS energy. Thus, the GRID and GROMOS energy scales are not simply shifted with respect to each other. Comparing the GRID energy to the excess chemical potential of water (−26.4 kJ mol^{-1}), the water 802 site should be occupied, but in contrast to the simulation results, site D would also be expected to be hydrated, albeit by a less strongly bound water molecule. The situation is different for the GROMOS interaction energy, for which occupation of site D is unfavorable relative to bulk solvent. The GROMOS interaction energies are more negative than the ΔF calculated from the simulations for large bottom radii by 5.1 kJ mol^{-1} (site D) and 3.7 kJ mol^{-1} (water 802). These differences may be partly attributed to the loss of entropy of the water molecule and the protein on water binding. They are of similar magnitude to the entropic cost of transferring a water molecule from solution to protein estimated to be between 0 and 8 kJ mol^{-1} from the standard entropies of anhydrous salts and their corresponding hydrates (Dunitz, 1994).

Although the free energies calculated here with the GROMOS energy function are consistent with experimental observations, possible problems with its modeling of water-hydrophobic and water-protein interactions have recently been raised. Smith and van Gunsteren (1994) calculated average water residence times on the surface of a protein from MD simulations and found them to be much shorter than expected. The authors suggest that the strength of the interaction between protein and solvent may be underestimated in the force field. On the other hand, a recent MD study of a decane-water interface using the SPC/E model (van Buuren et al., 1993) has shown that, in simulations with surfactant/oil/water systems, it appeared that the solubility of decane in water was far too high. The Lennard-Jones part of the intermolecular potential parameters was then optimized in a set of simulations. A modification of the original GROMOS force field was also used in a MD study of flavodoxin (Leenders et al., 1994) in which the van der Waals repulsion properties of polar atoms were increased with respect to the normal repulsion parameters. It is difficult to estimate the importance of these two force field corrections for this work. However, they have approximately opposite effects and would partly cancel each other out. In addition, it should be noted that in the MD simulations no account is taken of induced electronic polarization on insertion of the water molecule into the protein and electrostatic interactions with aromatic protons are neglected.

Comparison with experimental data

The solvation of the active site of P450cam has been studied by high pressure techniques (Di Primo et al., 1992) and the inactivation volume change upon application of hydrostatic pressure has been found to be directly related to the initial degree of hydration of the heme pocket. Camphor has a larger activation volume than camphor analogues designed to fill site D (E. Deprez, V. Helms, C. Barret, E. Gill, G. Hui Bon Hoa, and R. C. Wade, unpublished data), suggesting that the analogues do not displace more water and thus that site D is unoccupied in the presence of camphor.

The results from the MD simulations are in agreement with the crystallographic structures determined by Poulos and co-workers (1987) for the camphor and 2pheIm complexes. No water molecules were detected in the active site of the camphor complex. However, recent time-resolved crystallography of the complex of P450cam with its product seems to indicate electron density for water molecules within the active site including one near site D (I. Schlichting, unpublished data). Thus this site may be transiently occupied.

Implications for protein-ligand interactions and protein design

The interfacial cavity at site D in the camphor complex raises two questions. Why is it present and why is it preferential for it to be empty rather than hydrated?

The presence of a cavity means that the packing of the surrounding atoms is not optimal and its formation therefore has an energetic cost (which we have not attempted to calculate). In the case of P450cam, this cost may be toler-

ated because the cavity may play a functional role as it is near a proposed substrate access channel between Thr185, Phe87, and Ile395 (Poulos et al., 1986).

Statistical studies (Hubbard et al., 1994) of protein structures show that empty cavities rarely exceed 0.05 nm^3. However, they also indicate the importance of polarity in determining cavity solvation. Although the number of hydrogen bonds formed by water molecules in protein cavities varies from zero to four, three is the most common (Williams et al., 1994). Thus it can be anticipated that the enthalpic contribution of favorable hydrogen bonds to a water molecule in a cavity can compensate the energetic cost of cavity formation. Therefore, as argued by Wolfenden and Radzicka (1994) on the basis of experiments with cyclohexane as a model for a nonpolar protein environment, the occupation of nonpolar cavities in proteins by water molecules is expected to be highly unlikely. However, contrary to this expectation, water molecules with long residence times were recently detected in a completely nonpolar, buried protein cavity in a recent nuclear magnetic resonance study of interleukin 1β (Ernst et al., 1995). In the cytochrome P450 complexes examined here, water 802 makes three hydrogen bonds in the crystal structure whereas a water at site D could make only one. Thus, for site D, the satisfaction of camphor's hydrogen-bonding capacity by a water molecule is apparently outweighed by its own hydrogen-bonding capacity being unfulfilled.

It has been shown that both empty and hydrated cavities can be engineered into proteins. For example, Fitzgerald et al. (1994) created a cavity in cytochrome *c* peroxidase that was observed to be occupied by five ordered water molecules. Eriksson et al. (1992a) introduced a hydrophobic cavity into lysozyme with the ability to bind benzene and showed that cavities can be introduced into lysozyme with only minor conformational rearrangements of the protein (Eriksson et al., 1992b). However, a more general understanding of the hydration properties of engineered cavities is necessary so that their effects on protein stability and their ability to bind diverse ligands can be predicted. In the case of P450cam, the ability to predict ligand-binding affinities requires a good understanding of the solvation structure and thermodynamics of the whole unliganded active site (a subject we are addressing in ongoing studies.)

In conclusion, the MD simulation method described here enables the probability that a water molecule mediates a protein-ligand interaction by occupying an interfacial cavity to be calculated. The protein-ligand complexes studied here provide examples of empty and hydrated interfacial cavities that are distinguished in the calculations. The calculations show that an empty interfacial cavity can be tolerated and suggest that such cavities are likely to be unhydrated unless the surrounding protein and ligand atoms have the capacity to make more than one hydrogen bond to a water molecule in the cavity.

We thank T. P. Straatsma for the provision of the ARGOS program, for great help and much patience while guiding us to its features, and for interesting and stimulating discussions; S. Hubbard and P. J. Goodford for critical reading of the manuscript; D. Harris for the provision of a reprint and helpful comments; J. P. Jones for the provision of P450cam heme parameters; I. Schlichting, E. Deprez, and G. Hui Bon Hoa for sharing unpublished results; and T. Poulos for discussion of crystallographic data. We thank HLRZ, Jülich, for access to computing facilities.

This work was supported in part by the EU Biotechnology Programme (BIO2-CT94–2060).

REFERENCES

Allen, F. H., S. Bellard, M. . Price, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink, and B. G. Watson. 1979. The Cambridge Crystal Data Center: computer-based search, retrieval, analysis, and display of information. *Acta Cryst. B.* B35:2331–2339.

Åquist, J. 1990. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* 94:8021–8024.

Beamer, L. J., and C. O. Pabo. 1992. Refined 1.8 Å crystal structure of the lambda repressor-operator complex. *J. Mol. Biol.* 227:177–196.

Ben-Naim, A., and Y. Marcus. 1986. Solvation thermodynamics of nonionic solutions. *J. Chem. Phys.* 81:2016–2027.

Berendsen, H. J. C., and J. R. Grigera. 1987. The missing term in effective pair potentials. *J. Phys. Chem.* 91:6269–6271.

Berendsen, H. J. C., J. P. M. Postma, W. V. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. M. J. Smith, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and T. Tasumi. 1977. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Besler, B. H., K. M. Merz, and P. A. Kollman. 1990. Atomic charges derived from semiempirical methods. *J. Comp. Chem.* 11:431–439.

Beveridge, D. L., and F. M. DiCapua. 1989. Free energy via molecular simulations. *Annu. Rev. Biophys. Biophys. Chem.* 18:431–492.

Bhat, T. N., G. A. Bentley, G. Boulot, M. I. Greene, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza, and R. J. Poljak. 1994. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA.* 91:1089–1093.

Boobbyer, D. N. A., P. J. Goodford, P. M. McWhinnie, and R. C. Wade. 1989. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* 32:1083–1094.

Chervenak, M. C., and E. J. Toone. 1994. A direct measure of the contribution of solvent reorganization to the enthalpy of ligand binding. *J. Am. Chem. Soc.* 116:10533–10539.

Chipot, C., C. Millot, B. Maigret, and P. A. Kollman. 1994. Molecular dynamics free energy simulations: influence of the truncation of long-range nonbonded electrostatic interactions on free energy calculations of polar molecules. *J. Chem. Phys.* 101:7953–7962.

Clore, G. M., A. Bax, J. G. Omichinski, and A. M. Gronenborn. 1994. Localization of bound water in the solution structure of a complex of the erythroid transcription factor GATA-1 with DNA. *Structure.* 2:89–94.

Collins, J. R., D. L. Camper, and G. H. Loew. 1991. Valproic acid metabolism by cytochrome P450: a theoretical study of stereoelectronic modulators of product distribution. *J. Am. Chem. Soc.* 113:2736–2743.

Cowan, S. W., M. E. Newcomer, and T. A. Jones. 1990. Crystallographic refinement of human serum retinol binding protein at 2 Å resolution. *Proteins.* 8:44–61.

Crossley, J., W. F. Hassell, and S. Walker. 1968. Dielectric studies. XIX. Molecular relaxation of some rigid molecules. *Can. J. Chem.* 46: 2181–2185.

Dewar, M. J. S., and W. Thiel. 1977. Ground states of molecules. 38. The MNDO method: approximations and parameters. *J. Am. Chem. Soc.* 99:4899–4907.

Di Primo, C., G. H. B. Hoa, P. Douzou, and S. G. Sligar. 1990. Mutagenesis of a single hydrogen bond in cytochrome P450 alters cation binding and heme solvation. *J. Biol. Chem.* 265:5361–5363.

Di Primo, C., G. H. B. Hoa, P. Douzou, and S. G. Sligar. 1992. Heme-pocket-hydration change during the inactivation of cytochrome P450cam by hydrostatic pressure. *Eur. J. Biochem.* 209:583–588.

Dunitz, J. D. 1994. The entropic cost of bound water in crystals and biomolecules. *Science.* 264:670.

Eriksson, A. E., W. A. Baase, J. A. Wozniak, and B. W. Matthews. 1992a. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature.* 355:371–373.

Eriksson, A. E., W. A. Baase, X. J. Zhang, D. W. Heinz, M. Blaber, E. P. Baldwin, and B. W. Matthes. 1992b. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science.* 255:178–183.

Ernst, J. A., R. T. Clubb, H. X. Zhou, A. M. Gronenborn, and G. M. Clore. 1995. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science.* 267:1813–1817.

Fitzgerald, M. M., M. J. Churchill, D. E. McRee, and D. B. Goodin. 1994. Small molecule binding to an artificially created cavity at the active site of cytochrome *c* peroxidase. *Biochemistry.* 33:3807–3818.

Garduno-Juarez, R., and E. Ocampo-Garcia. 1989. Conformational studies by molecular mechanics and molecular orbital methods on the antiamoebic drug 1-(4-imidazolylsulfonyl)-4-phenylimidazole. *Computers Chem.* 13:117–122.

Goodford, P. J. 1985. Computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* 28:849–857.

Harris, D. L., Chang Y. T., and G. H. Loew. 1995. Molecular dynamics simulations of phenylimidazole inhibitor complexes of cytochrome P450 cam. *In* Modelling of Biomolecular Structures and Mechanisms. Zwolle, Kluwer. 189–202.

Harris, D. L., and G. H. Loew. 1993. Determinants of the spin state of the resting state of cytochrome P450 cam. *J. Am. Chem. Soc.* 115: 8775–8779.

Harris, D. L., and G. H. Loew. 1995. Prediction of regiospecific hydroxylation of camphor analogs by cytochrome P450 cam. *J. Am. Chem. Soc.* 117:2738–2746.

Heiner, A. P. 1992. Predictive aspects of molecular dynamics simulations for proteins. Ph.D. thesis. Groningen University, Groningen, The Netherlands 161 pp.

Hermans, J., A. Pathiaseril, and A. Anderson. 1988. Excess free energy of liquids from molecular simulations: application to water models. *J. Am. Chem. Soc.* 110:5982–5986.

Hermans, J., and S. Shankar. 1986. The free energy of xenon binding to myoglobin from molecular dynamics simulations. *Isr. J. Chem.* 27: 225–227.

Hubbard, S. J., and P. ARGOS. 1994. Cavities and packing at protein interfaces. *Protein Sci.* 3:2194–2206.

Hubbard, S. J., K. H. Gross, and P. Argos. 1994. Intramolecular cavities in globular proteins. *Protein Eng.* 7:613–626.

Jones, J. P., W. F. Trager, and T. J. Carlson. 1993. The Binding and regioselectivity of reaction of (R)- and (S)-nicotine with cytochrome P450 cam: parallel experimental and theoretical studies. *J. Am. Chem. Soc.* 115:382–387.

Jorgensen, W. L., J. F. Blake, and J. K. Buckner. 1989. Free energy of TIP4P water and the free energies of hydration of CH_4 and CL^- from statistical perturbation theory. *Chem. Phys.* 129:193–203.

Kollman, P. A. 1993. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* 93:2395–2417.

LaLondo, J. M., D. A. Bernlohr, and L. J. Banaszak. 1994a. X-ray crystallographic structures of adipocyte lipid-binding protein complexed with palmitate and hexadecanesulfonic acid: properties of cavity binding sites. *Biochemistry.* 33:4885–4895.

LaLondo, J. M., J. A. Levenson, J. J. Rose, D. A. Bernlohr, and L. J. Banaszak. 1994b. Adipocyte lipid-binding protein complexed with arachidonic acid. *J. Biol. Chem.* 41:25339–25347.

Leenders, R., W. F. van Gunsteren, H. J. C. Berendsen, and A. J. W. G. Visser. 1994. Molecular dynamics simulations of oxidized and reduced *Clostridium beijerinckii* flavodoxin. *Biophys. J.* 66:634–645.

Li, J., A. Vrielink, P. Brick, and D. M. Blow. 1993. Crystal structure of cholesterol oxidase complexed with a steroid substrate: implications for flavin adenine dinucleotide dependent alcohol oxidases. *Biochemistry.* 32:11507–11515.

Nicholls, A., and B. Honig. 1992. GRASP, v.1.1. Columbia University, New York.

Paulsen, M. D., M. B. Bass, and R. L. Ornstein. 1991. Analysis of active site motions from a 175 psec molecular dynamics simulation of camphor-bound cytochrome P450 cam. *J. Biomol. Struct. Dyn.* 9:187–203.

Paulsen, M. D., and R. L. Ornstein. 1991. A 175 psec molecular dynamics simulation of camphor-bound cytochrome P450 cam. *Proteins.* 11: 184–204.

Porter, T. D., and M. J. Coon. 1991. Cytochrome P-450: multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem.* 266:13469–13472.

Poulos, T. L., B. C. Finzel, and A. J. Howard. 1986. Crystal structure of substrate-free *Pseudomonas putida* cytochrome P450. *Biochemistry.* 25: 5314–5322.

Poulos, T. L., B. C. Finzel, and A. J. Howard. 1987. High-resolution crystal structure of cytochrome P450 cam. *J. Mol. Biol.* 195:687–700.

Poulos, T. L., and A. J. Howard. 1987. Crystal structures of metapyrone- and phenylimidazole-inhibited complexes of cytochrome P450 cam. *Biochemistry.* 26:8165–8174.

Quintana, J., and A. D. J. Haymet. 1992. The chemical potential of water: molecular dynamics computer simulation of the CF and SPC models. *Chem. Phys. Lett.* 189:273–277.

Quiocho, F. A., D. K. Wilson, and N. K. Vyas. 1989. Substrate specificity and affinity of a protein modulated by bound water molecules. *Nature.* 340:404–407.

Raag, R. H., H. Li, B. C. Jones, and T. L. Poulos. 1993. Inhibitor-induced conformational change in cytochrome P450. *Biochemistry.* 32: 4571–4578.

Raag, R., and T. L. Poulos. 1989. The structural basis for substrate-induced changes in redox potential and spin equilibrium in cytochrome P450cam. *Biochemistry.* 28:917–922.

Raag, R., and T. L. Poulos. 1991. Crystal structures of cytochrome P450 cam complexed with camphane, thiocamphor, and adamantane: factors controlling P450 substrate hydroxylation. *Biochemistry.* 30:2674–2684.

Rand, R. P., N. L. Fuller, P. Butko, G. Francis, and P. Nicholls. 1993. Measured change in protein solvation with substrate binding and turnover. *Biochemistry.* 32:5925–5929.

Reynolds, C. A., P. M. King, and W. G. Richards. 1992. Free energy calculations in molecular biophysics. *Mol. Phys.* 76:251–275.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion in a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* 23: 327–341.

Rydel, T. J., A. Tulinsky, W. Bode, and R. Huber. 1991. Refined structure of the hirudin-thrombin complex. *J. Mol. Biol.* 221:583–601.

Shakked, Z., G. Guzikevich-Guerstein, F. Frolow, D. Rabinovich, A. Joachimiak, and P. B. Sigler. 1994. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature.* 368:469–473.

Smith, P. E., and W. F. van Gunsteren. 1994. Translational and rotational diffusion of proteins. *J. Mol. Biol.* 236:629–636.

Straatsma, T. P., H. J. C. Berendsen, and A. J. Stam. 1986. Estimation of statistical errors in molecular simulation calculations. *Mol. Phys.* 57: 89–95.

Straatsma, T. P., and J. A. McCammon. 1990. ARGOS, a general vectorized molecular dynamics program. *J. Comp. Chem.* 11:943–951.

Straatsma, T. P., and J. A. McCammon. 1991. Multiconfigurational thermodynamic integration. *J. Chem. Phys.* 91:3631–3637.

Straatsma, T. P., and J. A. McCammon. 1992. Computational alchemy. *Annu. Rev. Phys. Chem.* 43:407–435.

van Buuren, A. R., S. J. Marrink, and H. J. C. Berendsen. 1993. A molecular dynamics study of the decane/water interface. *J. Phys. Chem.* 97:9206–9212.

van Gunsteren, W. F., and H. J. C. Berendsen. 1987. GROMOS. Groningen Molecular Simulation Library, Groningen, The Netherlands.

van Gunsteren, W. F., T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, and P. E. Smith. 1993. Computation of free energy in practice: choice of

approximations and accuracy limiting factors. *In* Computer Simulation of Biomolecular Systems. Leiden, Escom. 315–348.

van Gunsteren, W. F., and A. E. Mark. 1992. On the interpretation of biochemical data by molecular dynamics computer simulations. *Eur. J. Biochem.* 204:947–961.

van Gunsteren, W. F., P. K. Weiner and A. J. Wilkinson, Ed. (1993). Computer Simulation of Biomolecular Systems. Escom, Leiden, The Netherlands.

Varadarajan, R., and F. M. Richards. 1992. Crystallographic structures of ribonuclease S variants with nonpolar substitution at position 13: packing and cavities. *Biochemistry.* 31:12315–12327.

Wade, R. C. 1990. Solvation of the active site of cytochrome P450 cam. *J. Comp. Aid. Mol. Des.* 4:199–204.

Wade, R. C., K. J. Clark, and P. J. Goodford. 1993. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. I. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* 36:140–147.

Wade, R. C., and P. J. Goodford. 1993. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. II. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* 36:148–156.

Wade, R. C., M. H. Mazor, and J. A. McCammon. 1991. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers.* 31:919–931.

Williams, M. A., J. M. Goodfellow, and J. M. Thornton. 1994. Buried waters and internal cavities in monomeric proteins. *Protein Sci.* 3:1224–1235.

Wolfenden, R., and A. Radzicka. 1994. On the probability of finding a water molecule in a nonpolar cavity. *Science.* 265:936–937.

Zacharias, M., T. P. Straatsma, and J. A. McCammon. 1993. Inversion of receptor binding preferences by mutagenesis: free energy thermodynamic integration studies on sugar binding to L-arabinose binding protein. *Biochemistry.* 32:7428–7434.

Zacharias, M., T. P. Straatsma, and J. A. McCammon. 1994. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* 100:9025–9031.