PROTEINS: Structure, Function, and Genetics 45:428–437 (2001)

Classical Molecular Interaction Potentials: Improved Setup Procedure in Molecular Dynamics Simulations of Proteins

Josep Lluis Gelpí,[1] Susana G. Kalko,[1,2] Xavier Barril,[3] Jordi Cirera,[1] Xavier de la Cruz,[1] F. Javier Luque,[3*] and Modesto Orozco[1*]

[1]*Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Barcelona, Spain*
[2]*Instituto de Biología Molecular de Barcelona, CSIC, Barcelona, Spain*
[3]*Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain*

ABSTRACT **The latest version of the classical molecular interaction potential (CMIP) has the ability to predict the position of crystallographic waters in several proteins with great accuracy. This article analyzes the ability of the CMIP functional to improve the setup procedure of the molecular system in molecular dynamics (MD) simulations of proteins. To this end, the CMIP strategy is used to include both water molecules and counterions in different protein systems. The structural details of the configurations sampled from trajectories obtained using the CMIP setup procedure are compared with those obtained from trajectories derived from a standard equilibration process. The results show that standard MD simulations can lead to artifactual results, which are avoided using the CMIP setup procedure. Because the CMIP is easy to implement at a low computational cost, it can be very useful in obtaining reliable MD trajectories. Proteins 2001;45:428–437.** © 2001 Wiley-Liss, Inc.

Key words: CMIP; MD simulations; counterions; water molecules

INTRODUCTION

Molecular dynamics (MD) is one of the most useful theoretical methods with which to study the structure, flexibility, and reactivity of biomolecules.[1–8] MD relies on the integration of Newtonian motion laws from the forces determined by using classical force fields. It permits the simulation of biological and bioorganic systems in many experimental conditions, including the physiologically similar isothermic/isobaric ensemble.[1–8]

Early MD simulations of proteins were performed in vacuo, limiting their validity in describing the structural details of biomolecules. Current MD simulations treat explicitly the solvent and, in conjunction with methods to represent long-range effects,[8–10] are very valuable in gaining insight into the structure and flexibility of proteins under physiological conditions.[1–8] In addition to the problems derived from the accuracy of classical force fields, the strongest limitation of MD arises from the small time step (~1–2 fs) required to integrate Newtonian equations. Thus, ~5 10^5–10^6 energy and force calculations are required to obtain just 1 ns of trajectory. This severely limits the length of MD simulations to values [currently within the range of several nanoseconds (ns)], clearly below the time scale of a large variety of biological processes.

The reduced length of the MD trajectories reinforces the importance of making a precise setup of the system to obtain reliable sampling of the configurational space. Generally, MD simulations of proteins exploit crystal or nuclear magnetic resonance (NMR) structural data to define the starting conformation of the system. The starting structure is then immersed in large pre-equilibrated boxes of solvent (typically water in biomolecular studies). The solvent molecules closer to the protein than a predefined distance cutoff are then removed to avoid steric clashes and to obtain systems of suitable density. Eventually counterions (typically Na^+ and Cl^-) are added to neutralize the system or to mimic a given ionic strength. Finally, the system is partially energy-minimized to relax unfavorable contacts, random velocities are assigned to atoms in both the protein and the solvent, and the system is equilibrated for few hundred picoseconds (ps) at a given temperature and pressure (typically 298 K and 1 atm) before data collection.

Complex multistep equilibration protocols are typically used to facilitate relaxation of the biomolecular systems.[11,12] Accordingly, before a global equilibration of the system, the solvent is relaxed for several picoseconds in the presence of a protein, which is partially or totally frozen. This type of procedure increases the cost of the simulation but helps avoid artifactual structural distortions that can occur due to a bad initial arrangement of solvent molecules. However, even multistep equilibration protocols can fail when solvent relaxation is coupled to protein movements (e.g., water molecules entering buried cavities in proteins) or when it requires large time scales (e.g., water diffusion through channels). Under these circumstances, artifactual results can be obtained, even

Grant sponsor; Fundació (La Marató TV3); Grant number: 3004/97; Grant sponsor; Dirección General de Investigación Científica y Técnica; Grant number: PB98-1222; Grant number: PM99-0046; Grant sponsor; European Union; Grant number: BIO4-CT98-0354; Grant sponsor; Centre de Supercomputació de Catalunya (CESCA).

*Correspondence to: Modesto Orozco, Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain. E-mail: modesto@luz.bq.ub.es; or F. Javier Luque. E-mail: javier@fars.far.ub.es

Received 8 December 2000; Accepted 12 June 2001

TABLE I. Details of the Standard Equilibration Process Used in MD Simulations of Catalase, Thymidine Kinase, and Acetylcholinesterase*

	Catalase	Thymidine kinase	Acethylcholinesterase
No. of waters	3,734	980	670
Radii of the cap	35	25	25
Radii mobile region	15	15	20
Nonbonded cutoff	10	10	11
Equilibration procedure	Solvent optimization[a]	Solvent optimization	Solvent optimization
	System optimization[b]	System optimization	System optimization
	200-ps heating solvent (100 K)	100-ps heating solvent (100 K)	50-ps heating solvent (150 K)
	200-ps heating protein + ions (100 K)	100-ps heating protein + ions (100 K)	50-ps heating system (150 K)
	Mixing velocities + 200-ps heating system (300 K)	Mixing velocities + 100-ps heating system (300 K)	100-ps equilibration
	200-ps equilibration	200-ps equilibration	
Length trajectory	5 ns	1 ns	1 ns

CMIP, classical molecular interaction potentials; MD, molecular dynamics.
*The CMIP-based equilibration process is based on a CMIP pre-equilibration followed by the standard equilibration shown here.
[a]Solvent optimization 3000 cycles steepest descent + 5000 cycles conjugated gradient.
[b]System optimization for 2000 cycles steepest descent + 3000 cycles conjugated gradient.

for very long trajectories, from the inadequate equilibration of the biomolecule + solvent system.

In this article, we define the classical molecular interaction potential (CMIP) and its application to the setup of the solvent atmosphere in MD simulations of biomolecules. The method implies a very small increase in the total cost of the MD simulation, but it represents a significant improvement in the quality of the trajectories compared with the results obtained from MD simulations run using standard equilibration protocols for different protein systems.

METHODS

Overview of the CMIP

The CMIP functional defines the potential around a macromolecule as the addition of steric (V_{vW}) and electrostatic (V_{ele}) contributions:

$$V(\mathbf{r}) = V_{ele}(\mathbf{r}) + V_{vW}(\mathbf{r}) \quad (1)$$

The CMIP allows us to compute the electrostatic potential using four formalisms[13]: the Coulombic expression with (1) a constant dielectric model (see eq. 2), (2) a linear distance-dependent dielectric model[13,14] (see eq. 3), (3) the Mehler–Solmajer sigmoidal distance-dependent dielectric model[13,15] (see eq. 4), and (4) the potential obtained by solving numerically the linear Poisson equation[13,16] (see eq. 5). Finally, the van der Waals potential is computed using a Lennard–Jones expression (eqs. 6 and 7):

$$V_{ele}(\mathbf{r_i}) = \sum_{j=1}^{J} \frac{Q_j}{\epsilon|\mathbf{r_j} - \mathbf{r_i}|} \quad (2)$$

where ϵ is the dielectric constant and Q_j are atomic charges taken from classical force fields,

$$V_{ele}(\mathbf{r_i}) = \sum_{j=1}^{J} \frac{Q_j}{\mathrm{EPS}|\mathbf{r_j} - \mathbf{r_i}|^2} \quad (3)$$

where EPS is a scaling factor (typically 1 or 4),

$$V_{ele}(\mathbf{r_i}) = \sum_{j=1}^{J} \frac{Q_j}{\varepsilon(\mathbf{r_i})|\mathbf{r_j} - \mathbf{r_i}|} \quad (4)$$

where

$$\varepsilon(\mathbf{r_i}) = -8.5525 + \frac{\varepsilon_0 + 8.5525}{1 + 7.7839\exp[-0.003627(\varepsilon_0 + 8.5525)|\mathbf{r_j} - \mathbf{r_i}|]}$$

and ϵ_0 is the dielectric constant of vacuum.

$$\nabla \cdot [\varepsilon(\mathbf{r_i}) \cdot \nabla \cdot V_{ele}(\mathbf{r_i})] = -4\pi\rho(\mathbf{r_i}) \quad (5)$$

where ρ is the charge density,

$$V_{vW}(\mathbf{r_i}) = \sum_{z=1}^{Z} V_{vW}^{z}(\mathbf{r_i}) \quad (6)$$

where z are the different van der Waals atom types (possible probe atoms), and V_{vW}^{z} is computed as:

$$V_{vW}^{z}(\mathbf{r_i}) = \sum_{l=1}^{L} (\zeta_l\zeta_z)^{1/2}\left[\left(\frac{R_z + R_l}{\mathbf{r_l} - \mathbf{r_i}}\right)^{12} - 2\left(\frac{R_z + R_l}{\mathbf{r_l} - \mathbf{r_i}}\right)^{6}\right] \quad (7)$$

where the sum for each grid point runs for all the van der Waals particles (l) of the protein (typically $l \leq j$); ζ and R are the van der Waals parameters.

In order to speed up calculations, the space around the protein is divided into small cubic elements, each characterized by a point located in the center of the cube. The interaction energy (E_i) of the protein with a probe placed in a fixed orientation at a grid point i is computed using eq. 8. The program uses a trilinear interpolation routine to smooth the potential function between grid elements. Note that for each probe the protein potential is computed only once, which enhances the efficiency of the calculation:

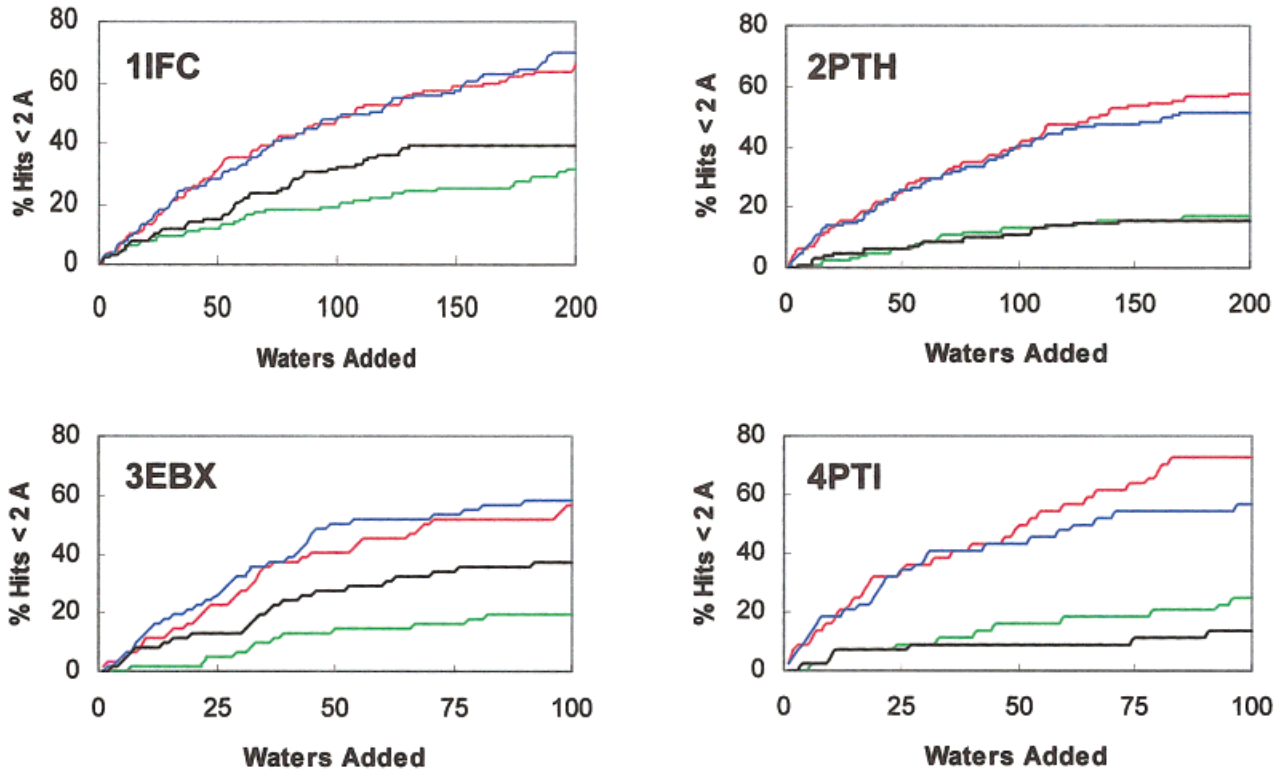

Fig. 1. Percentage of hits (i.e., waters located at <2 Å from crystal coordinates) in iterative classical molecular interaction potential (CMIP) calculations for increasing number of waters added for selected proteins. Total number of bound crystal waters is 62 (3EBX), 44 (4PTI), 98 (1IFC), and 101 (2PTH). Green, constant dielectric (eq. 2); black, linear-dependent dielectric constant (eq. 3); red, sigmoidal dielectric constant (eq. 4); blue, Poisson–Boltzman (eq. 5). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$$E_i = q_i V_{\text{ele}}(i) + \sum_z \delta_i V^z_{\text{vw}}(i) \tag{8}$$

where q_i is the atomic (probe) charge at grid point i, and δ_i is a integer equal to 1 when the van der Waals particle of type z is at the grid point i, and 0 otherwise.

For polyatomic probe molecules, all possible external rotations are considered. Moreover, for flexible molecules, a library of accessible conformers is stored for each probe molecule and the total interaction energy is then computed as the addition of intra- and intermolecular terms (eq. 9). This does not increase significantly the cost of the CMIP calculation because the intramolecular term is already defined in the library:

$$E_i^{\text{TOT}} = E_i^{\text{inter}} + E_i^{\text{intra}} \tag{9}$$

A desolvation term can also be added. To this end, the free energy of hydration of the probe molecule determined experimentally or from self-consistent reaction field (SCRF) data[13] is also stored in the library. Finally, an entropy correction is introduced to account for the loss of degrees of freedom of the probe molecule when moving from solution (typically a reference state of 1 M is used) to the binding site. Considering all these corrections, the final "free energy" of binding is defined as

$$G_i = -\text{RT} \ln \left[\frac{\sum_{\text{bound}} \exp(-E_i^{\text{TOT}}/\text{RT})}{(V_{\text{ref}}/V_i) N_{\text{rot}} \sum_{\text{conf}} \exp(-E_i^{\text{intra}}/\text{RT})} \right] \tag{10}$$

where $V_{\text{ref}}/V_{\text{i}}$ is the ratio between the volume available to the probe molecule at 1 M concentration in aqueous solution and the volume of the grid element i, and N_{rot} is the number of external rotations considered. The sum "*bound*" runs over all the internal rotations and conformers of the probe molecule placed at point i, and the sum *conf* runs only over all the conformers of the unbound probe.

To speed up calculations, those orientations leading to steric collapse are discarded from the exponential average. If required, the best orientation (and conformer) for binding at grid element i can be optimized using a gradient technique, which yields an optimum orientation and optimized interaction energy for this grid point.

The CMIP uses an iterative procedure to locate the best positions for singular waters and counterions. Accordingly, every time a new probe molecule (water or counterion) is added to the system, the total potential of the system (protein + water or counterion) is recomputed. The process finishes when (1) the desired number of waters or

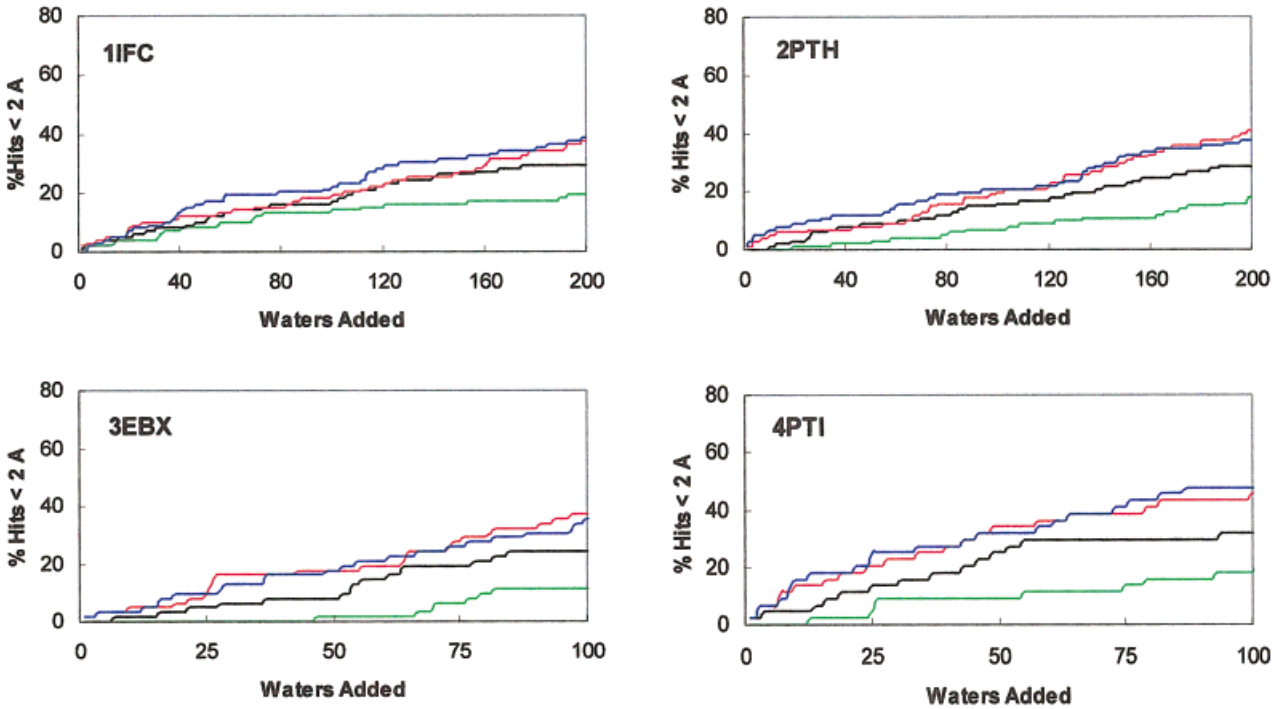

Fig. 2. Percentage of hits (i.e., waters located at <2 Å from crystal coordinates) in noniterative classical molecular interaction potential (CMIP) calculations for increasing number of waters added for selected proteins. For color code, see Fig. 1. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

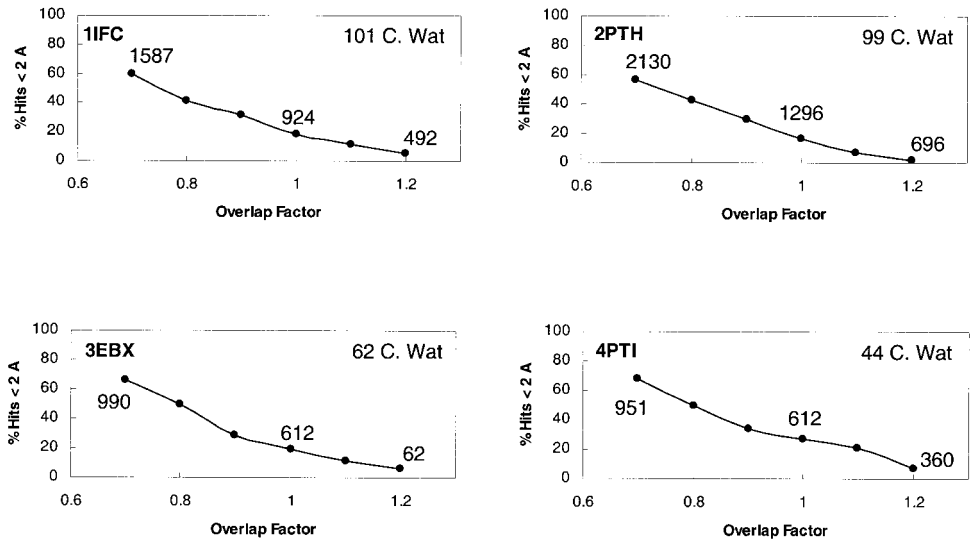

Fig. 3. Percentage of hits (i.e., waters located at <2 Å from crystal coordinates) when using Shell solvation procedure in the LEAP module of AMBER is used in conjunction with different overlap factors. The total number of water added for each overlap factor is indicated.

counterions is placed, or (2) the free energy of binding (eq. 10) is below (in absolute numbers) a given cutoff.

Selected Protein Systems

We first examined the ability of the CMIP to locate crystallographic water molecules. For this purpose, we selected four proteins from the Protein Data Bank (PDB). These proteins were chosen to have (1) small size, (2) high resolution (<1.5 Å), (3) no ligands bound, and (4) >30 and <100 water molecules in the crystal. The selected proteins are lipid-binding protein (code 1IFC), peptidyl-tRNA hydrolase from *Escherichia coli* (code 2PTH), erabutoxin B (code 3EBX), and trypsin inhibitor (code 4PTI). The CMIP calculations were performed using a grid spacing of 0.7 Å,

i.e., $\sim 10^6$ grid points (similar results were obtained in test calculations for a grid spacing of 0.5 Å, suggesting a small dependence of the results on grid size). Different electrostatic formalisms were considered with $\epsilon = 2$ in eqs. 2, EPS = 2 in eq. 3, and $\epsilon = 2$ (interior), and 80 (exterior) for eq. 5.

We also explored the dynamics of three large protein systems: thymidine kinase,[17] acetylcholinesterase,[18,19] and catalase.[20] The starting structures were taken from the crystallographic structure. MD simulations were run using a standard equilibration process (Table I), and after having placed up to 20 water molecules and 4 counterions with the CMIP (the Poisson expression was used together with a grid spacing of 0.5 Å) in the region of interest (the substrate channel in catalase, the inhibitor binding site in acetylcholinesterase, and the ATP and thymidine-binding sites in thymidine kinase) before equilibration. The placement of water/ions for these systems was based exclusively on energetic consideration; i.e., no entropy corrections were taken into account, since desolvation/entropic contributions are expected to be not too different for a common probe placed in different grid elements.

MD simulations were performed using SHAKE,[21] which allowed us to use a 2-fs time step for integration of Newtonian equations. Protein residues close to the region of interest (Table I), counterions near the mobile residues, and all water molecules were allowed to move during the trajectories, while residues far from the region of interest where kept frozen. The AMBER-95[22] and TIP3P[23] force fields were used. When necessary, ligand molecules were parametrized using the PAPQMD[24] and RESP[25] methodologies, using HF/6-31G(d) data as the reference. All MD simulations were carried out using the AMBER computer program.[26] CMIP calculations were performed using a locally developed code.[27]

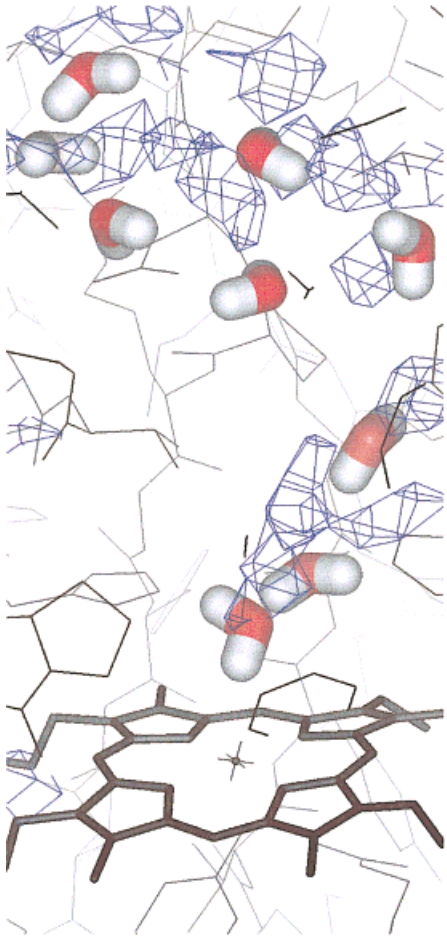

Fig. 4. Regions of high apparent (2 g/ml) water density in the main channel of catalase obtained in molecular dynamics (MD) simulations, and optimum location of waters according to the classical molecular interaction potential (CMIP) procedure (eq.5).

RESULTS AND DISCUSSION

Crystal Waters

We first examined the ability of the CMIP to locate crystallographic waters bound to small proteins (see Methods). The results displayed in Figure 1 demonstrate that the CMIP is able to locate water molecules properly if accurate electrostatic potentials (i.e., the sigmoidal model and the Poisson potential; eqs. 4 or 5) are used. Thus, ~70–80% of the bound crystallographic waters in 1IFC (98 crystallographic waters) and 2PTH (101 crystallographic waters) are located when 200 waters are positioned with the CMIP. Similarly, ~60–70% of the bound crystallographic waters are located for 3EBX (62 crystallographic waters) and 4PTI (44 crystallographic waters) when 100 waters are positioned with the CMIP. The performance of constant and distance-dependent dielectric models (eqs. 2 and 3) is clearly poorer (10–30% of fitted molecules) for all the protein systems (Fig. 1).

These results suggest that the CMIP locates correctly the waters tightly bound to proteins in crystals, but only if suitable electrostatic potentials are used. At this point, the quality of the sigmoidal function is quite remarkable, since the computational cost of the grid generation is similar to that of the simpler Coulombic models, and clearly less expensive than solving the Poisson equation.

Calculations in Figure 1 were performed using an iterative procedure, where after a new probe is added the perturbation induced in the grid by its presence is evaluated, which yields to a new grid potential. This method is more expensive than noniterative calculations like those in docking programs, in algorithms for the determination of interaction profiles in macromolecules,[28] or in programs for the setup of molecular dynamics.[26] However, it is expected to produce better results when large or very polar probes (e.g., ions) are added, but the need for the iterative procedure is not obvious for the location of bound water molecules. To investigate this, we repeated calculations in Figure 1, using a noniterative procedure, for the same definitions of the electrostatic potential (Fig. 2). As a reference, the results shown in Figure 2 (eq. 4) should be close to those obtained with GRID, since the image method used in that program to represent charge screening leads to dielectric profiles similar to those derived by the Mehler–Solmajer model. Comparison of Figures 1 and 2 shows the superior performance of the iterative CMIP procedure over

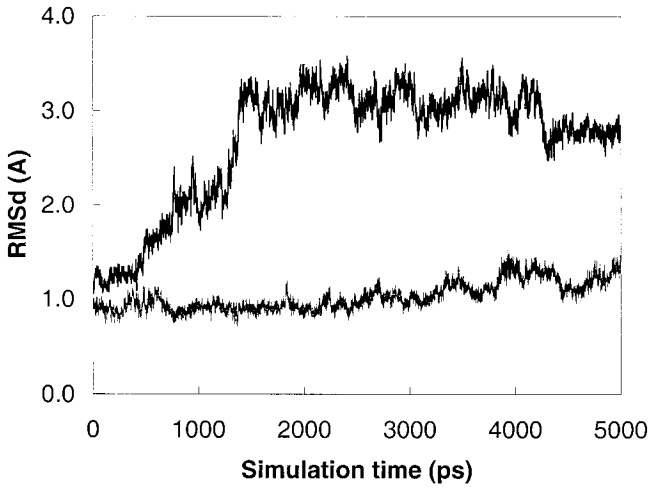

Fig. 5. Root mean square deviation from crystal structure (in Å) of the external loop region in simulations of catalase performed using a standard equilibration procedure (top line in black), and classical molecular interaction potential (CMIP)-based method (bottom line in gray).

noniterative protocols. Even larger differences are expected for bigger systems (e.g., catalase or acetylcholinesterase) in which water forms networks of H-bonded molecules.

An additional test of the quality of the CMIP procedure with respect to the standard procedures to locate waters in MD codes was to compare results in Figure 1 with those obtained when the solvation procedure (Shell program in the LEAP module in AMBER[26] is used). The Shell program centers the protein in a box of pre-equilibrated waters and defines a first solvation shell by elimination of those molecules that are too close or too far from the protein. The proximity criterion for elimination of solvent molecules is based on an overlap factor, which controls the degree of interpenetration between solute and solvent van der Waals spheres. The default value for the overlap factor is 1,[26] which means that no overlap of van der Waals spheres is allowed, but smaller values can be used to reinforce solvation, at the expense of the increase in the number of repulsive contacts.

Figure 3 shows the percentage of crystal water molecules found when the Shell code is used with different overlap factors ranging from 0.7 to 1.2. The use of small overlap factors allows us to recover more crystal waters. However, this increase in the number of hits has two undesirable effects: (1) there are a large number of unfavorable (i.e., atom–atom repulsion of >10 kcal/mol) interactions between solvent and protein, which can yield unstable MD trajectories; and (2) a very large number of solvent molecules are added. Thus, to recover ~60% of the crystal waters in the proteins studied, 1,000–2,000 molecules need to be added, i.e., ~10–20 times more water molecules than in the crystal. The CMIP allows a better fit of the crystal waters, introducing only a very small fraction of the molecules included using the Shell procedure.

Finally, the solvent atmosphere determined by the CMIP in the main channel of catalase (see below) was compared with the regions of large apparent density obtained by integrating water density along the MD trajectory[12,29] of catalase (see below for details). The results in Figure 4 demonstrate the similarity between the regions where more water is detected during the trajectories, and those where the CMIP locates water molecules. This confirms the ability of the CMIP to provide a reasonable picture of solvation atmosphere around proteins with a reduced computational effort.

MD Calculations

In order to determine the impact of the pre-equilibration of water and ions around the protein in the trajectory, we have performed MD simulations of three large proteins. In all cases, owing to the size of the proteins, solvent was defined using a drop of water around the region of interest, and nonbonded interactions were evaluated using standard cutoff criteria (Table I). However, to verify that our conclusions were not biased by the use of these simple simulation protocols, we run 1-ns trajectories for acetylcholinesterase (the smaller system considered in this study) periodic boundary conditions ($90 \times 80 \times 80$ Å^3 box containing 16,320 water molecules) and the Particle Mesh Ewald (PME) method[9] to account for long-range effects. These high-level trajectories were fully comparable to those obtained using the standard protocol in Table I (data for the PBC–PME simulation to be published elsewhere), which confirms that the main conclusions described above are not biased by the use of nonperiodic systems.

MD simulations of catalase

Catalases are key enzymes for the elimination of hydrogen peroxide in eukaryotes and prokaryotes. Among the three families of catalases, heme-catalases are the most ubiquitous.[30–32] These enzymes are catalytically very efficient, and in fact diffusion seems to be the rate-limiting step of the reaction.[31,33,34] This feature has been related to the existence of a narrow channel connecting the exterior of the protein with the catalytic heme group.[20,35] The channel is structurally well characterized in few mutants of peroxysomal catalase A from *Saccharomyces cerevisiae*, where a well-ordered string of waters is located in the crystal,[20,35] especially in the vicinity of the heme group (Fig. 4).

MD simulations of catalase (~20,000 atoms) performed using a very conservative standard equilibration process (800 ps; see Table I) and very long trajectories (5 ns) fail to reproduce the structure of the channel (Fig. 4). This obeys a displacement of the external loop during the trajectory, as noted in the root-mean-square deviation (RMSD) from the crystallographic structure (Fig. 5). Such a displacement determines the collapse of the hydrophobic channel, which is now not wide enough to allow the pass of a small molecule like water or hydrogen peroxide, in clear disagreement with available experimental data.[20,35] These results clearly demonstrate that the standard protocol cannot avoid the occurrence of artifactual results, even when very large equilibration/simulation periods are used. On the contrary, when the CMIP is used to solvate the channel region, the MD results are in much better agreement with

TABLE II. Selected Structural (Å) and Energetic (kcal/mol) Details Corresponding to the Interaction of the Huprine X in the Binding Pocket of Acetylcholinesterase Obtained from MD Simulations Carried Out With Standard Equilibration and the CMIP Pre-equilibration Protocols

Distance[a]	Standard	CMIP
Drug–Trp84	3.8 (0.5)	4.0 (0.5)
Drug–Phe330	4.2 (0.5)	4.0 (0.5)
NH—OC(His440)	2.9 (0.2)	2.9 (0.1)
NH_2—Oδ1(Asp72)	2.8 (0.1)	4.6 (0.4)
NH_2—Oδ2(Asp72)	4.8 (0.3)	5.5 (0.5)

CMIP, classical molecular interaction potentials; MD, molecular dynamics.
[a]Distance (Å) between the drug and key residues of the enzyme. Values were obtained by averaging during the last 500-ps MD simulations. Standard deviations are given in parentheses.

the experimental data. Thus, the external loop does not collapse and remains in a configuration similar to that observed in the crystal (Fig. 6), as noted in the RMSD plot (Fig. 5). The result is that the integrity of the channel and, accordingly, its ability to interact with H_2O/H_2O_2, remain intact (Fig. 6).

In summary, despite a long, conservative standard equilibration protocol, the 5-ns MD simulation provides incorrect sampling for catalase. The problem apparently originates from an inadequate hydration of deeply buried regions, such as the substrate channel in catalase. The diffusion of water molecules to these regions seems longer than the time required for the hydrophobic collapse. As a result, the trajectory does not recover the correct three-dimensional (3D) channel structure in a ≥5-ns time scale.

MD simulations of thymidine kinase

Thymidine kinase catalyzes the phosphorylation of thymidine before its incorporation into DNA. The viral thymidine kinase has a wide specificity for substrate, with the ability to 5′-phosphorylate nucleosides not closely related to thymidine, such as guanosine derivatives.[36–38] Thymidine kinase is a very large protein, whose active site contains two connected sites, which bind ATP and the substrate.[17] The exact orientation of the two binding sites is crucial for correct arrangement of the phosphate group of ATP and the O5′ group of the nucleoside.

Extended MD simulations (1 s) were started from the crystal conformation[17] using the standard equilibration procedure (Table I) fail to reproduce a suitable structure for the enzyme. Thus, the RMSD plot (Fig. 7) shows a dramatic displacement of a loop (the "lid" region, which is a key structural element of the ATP binding site) with respect to crystal data. This structural alteration is accompanied by a displacement of thymidine at the nucleoside binding site (Fig. 8). As a result, the MD simulation predicts a disruption in the active site geometry.

There are several explanations for such structural change in the active site of thymidine kinase. There is no negative counterion in the space occupied by ATP during the standard equilibration. Moreover, no counterion occupies that position along the 1-ns simulation period. This generates a dramatic unbalance of charge at the ATP binding site, which results in a strong displacement of the lid region from its initial position, leading to a loss of structure at the active site. By contrast, the nucleoside binding site is larger than the volume of thymidine. This extra volume is partially occupied by water molecules linking the nucleoside to residues of the enzyme. These waters are deeply inserted inside in the protein, and it was not possible to detect their appearance in MD simulations, at least during the 1-ns simulation. Once more, the time scale of the active site reorganization is shorter than that of water diffusion to the active site.

In another MD simulation, the CMIP was used to locate Cl^- and Na^+ and water molecules (see Methods) around the binding sites. The two water molecules bridging thymidine and the protein in the crystal structure[17] are properly predicted, and two Cl^- ions are placed in the ATP binding site (similar results where found when SO_4^{2-} was used as a negative counterion). Trajectories started using these pre-equilibrated configurations are stable. No structural alteration is observed in the lid region and the nucleotide binding site retains the configuration found in the crystal (Fig. 5). Clearly, the CMIP helps the MD simulation in sampling meaningful regions of the configurational space.

MD simulations of acetylcholinesterase

Acetylcholinesterase is one of the major targets for the treatment of Alzheimer disease.[39] Tacrine and huperzine A are among the few drugs available to treat this disease.[40] The fact that the binding sites of tacrine[18] and huperzine A[19] partially overlap can be exploited to design new drugs structurally related to those compounds. This is the case of the recently developed huprines (also denoted tacrine–huperzine A hybrids), which in some cases exhibit much larger inhibitory potency compared with either tacrine or huperzine A.[41–44]

MD simulations are valuable for exploring the binding mode of new drugs provided reliable structural information is obtained from the trajectory. At this point, the crystallographic structures of the acetylcholinesterase complexed with tacrine[18] or huperzine A[19] demonstrate the important role of water-mediated bridges for correctly locating the inhibitors within the binding site. Keeping in mind the close structural similarity of huprines to either tacrine or huperzine A, the description of those water-mediated contacts is expected to be crucial for predicting correctly the binding mode of huprines.[41–44]

For the purposes of this study, we performed two independent MD simulations. using the standard and CMIP equilibration processes. In the two cases the starting conformation was built up by placing manually the huprine X (the 9-ethyl-3-chloro-huprine derivative) in the binding site of the *Torpedo californica* enzyme. To this end, the location of the drug was performed in an effort to maintain the structural elements that link the drug to specific residues in the binding site. Thus, the quinoline ring was placed between the rings of Trp84 and Phe300; the quinoline NH group in tacrine (H-bonded to His440) and huprine was superimposed; the NH_3 group in tacrine

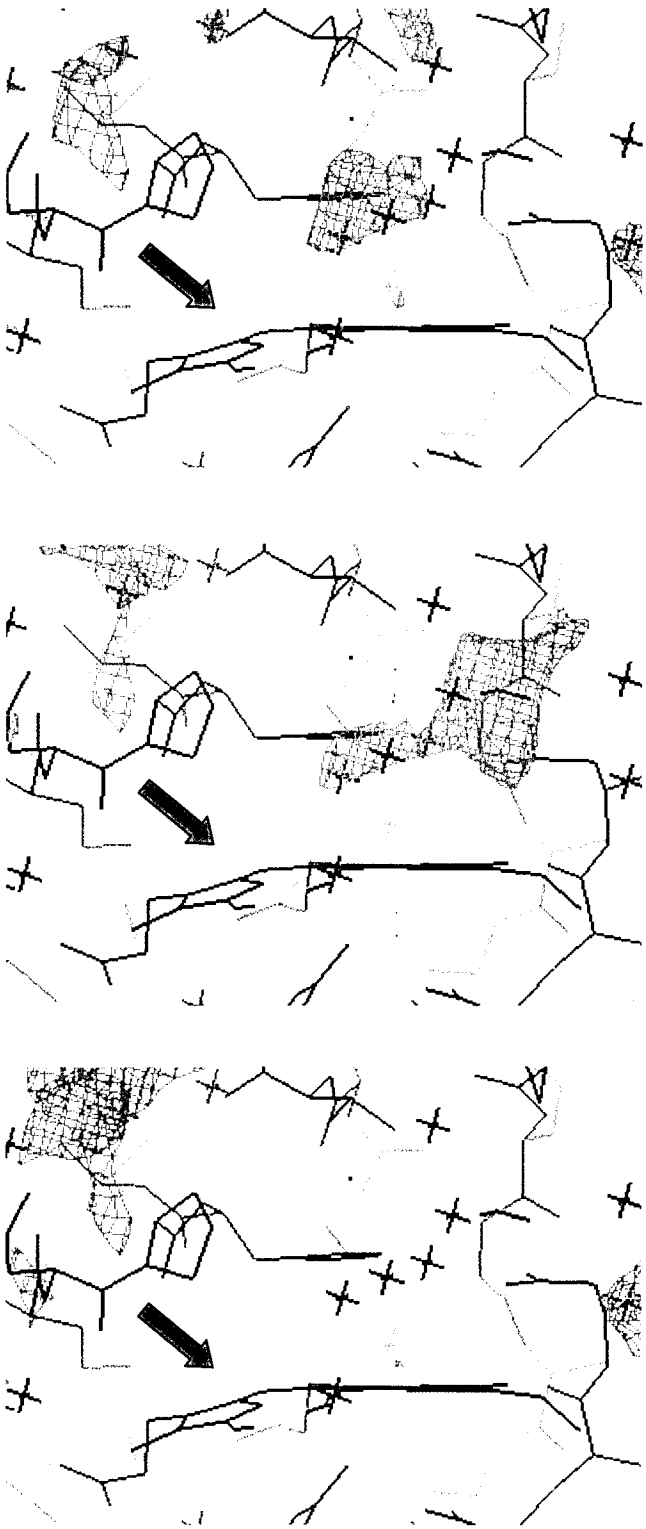

Fig. 6. Representation of the substrate channel in the three catalase structures. Top, crystal conformation; middle, average molecular dynamics (MD) structure obtained following the classical molecular interaction potential (CMIP) pre-equilibration process; bottom, average MD structure obtained following the standard equilibration process. Contour level represented here corresponds to the 0 kcal/mol interaction energy between the protein and a neutral oxygen molecule. For the sake of clarity, a common structure of the protein containing crystal waters is represented in all the plots. The position of the heme group is pointed with an arrow in all the structures.

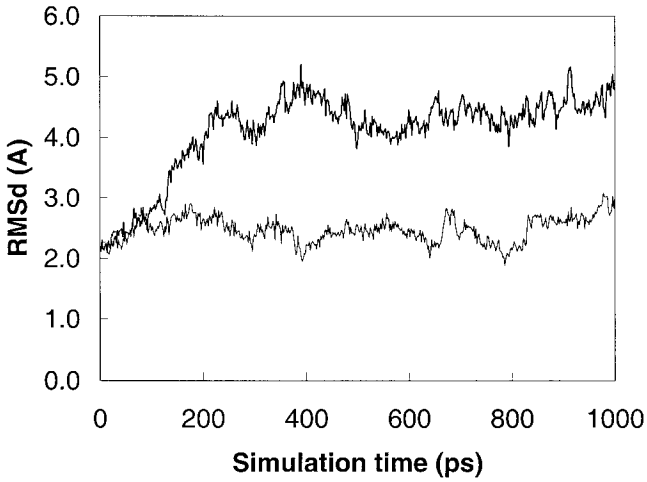

Fig. 7. Root-mean-square deviation (RMSD) from crystal structure (in Å) of the external loop region in simulations of thymidine kinase performed using a standard equilibration procedure (top line in black), and classical molecular interaction potential (CMIP)-based method (bottom line in gray).

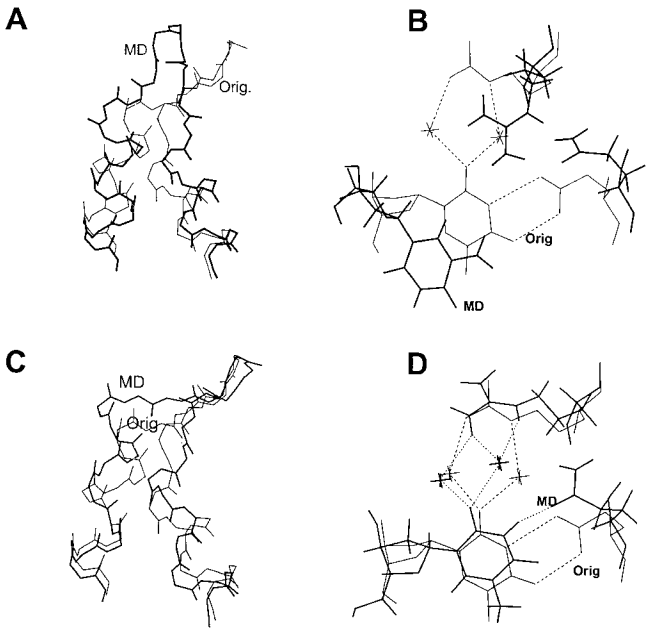

Fig. 8. Details of the final structures of molecular dynamics (MD) trajectories (1 ns) of thymidine kinase obtained using the standard equilibration protocol (**A,B**), and the classical molecular interaction potential (CMIP) pre-equilibration method (**C,D**). A and B show a detail of the lid region; B and D illustrate a detail of the nucleoside binding site. In all cases, the crystal (Orig) structure is displayed for comparison.

and huprine was superimposed with the NH_3 group in huperzine A, which forms a cation–π interaction with Trp84; and finally the carbobicyclic rings of huperzine A and huprine were also matched. The acetylcholinesterase–huprine complex was then hydrated using the standard protocol or the CMIP method and was equilibrated using the procedure shown in Table I. Two 1-ns trajectories were performed from these equilibrated structures.

Analysis of the binding pocket in the two trajectories shows the maintenance of common structural features, but

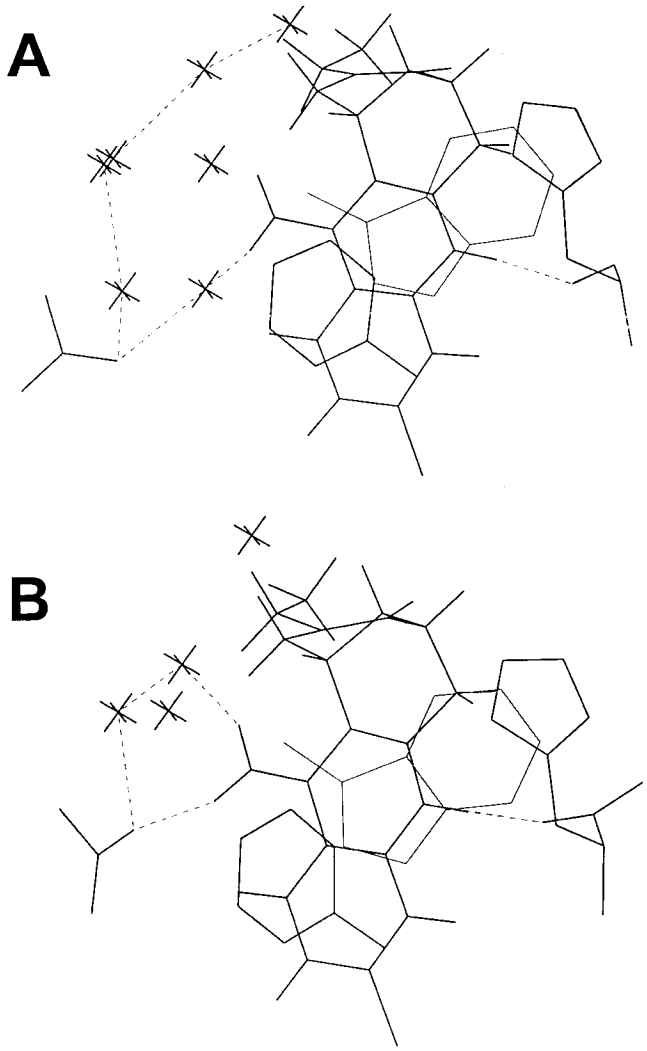

Fig. 9. Structural details of huprine at the binding site of acetylcholinesterase. The rings of Trp84 and Phe300 lie above and below the plane of the quinoline unit. The —NH group is hydrogen-bonded to the carbonyl group of His440. The —NH_2 group interacts through water-mediated contacts (**A**) or directly (**B**) with Asp72. The former situation corresponds to the simulation performed using the standard equilibration protocol, and the latter is obtained using the classical molecular interaction potential (CMIP) pre-equilibration procedure.

also of significant differences in the binding mode (Table II and Fig. 9). In particular, in the simulation performed using the standard protocol the huprine form a direct interaction between the NH_2 group and the carboxylic group of Asp72 (Fig. 9). However, in the CMIP pre-equilibrated simulation the huprine interacts with Asp72 through water-mediated contacts. Clearly, the water molecules positioned using the standard protocol do not reorient properly, to avoid approach of the positively charge quinoline NH_2 group with the negatively charged Asp72. On the contrary, the CMIP properly orients the water molecules in the starting conformation and the water-mediated contacts remain stable along the trajectory.

It is worth noting that free-energy calculations performed using the binding mode derived from CMIP calculations were able to reproduce relative binding free energies for a series of huprine derivatives while the same calculations performed using the alternative binding mode provide erroneous results.[41,42,44] The recently solved crystallographic structure (J.L. Sussman, personal communication) of the acetylcholinesterase–huprine X complex confirms the existence of a well-defined network of hydrogen–bond water that links the quinoline NH_2 group to Asp72 and that also encompasses other residues in the binding site, such as Tyr121 and Ser122.

CONCLUDING REMARKS

The correct placement of solvent molecules around proteins is crucial to obtain reliable trajectories from current MD simulations. The standard equilibration protocols, based on immersing the protein in a box of pure water, partial equilibration of solvent, and global equilibration of the system, do not warrant the reliability of the structures sampled along the trajectory, even when extremely large (≤800 ps) equilibration protocols are used. The reason is related to the intrinsic time scale of protein and solvent movements, which makes difficult the correct positioning of solvent molecules in hindered sites of the macromolecule. Without these key structural solvent (water or counterions) molecules, structural artifacts can occur along the trajectory.

In some cases, the use of crystallographic water molecules is valuable in alleviating this problem. However, in most cases of interest, the trajectories start from structures where the positions of solvent molecules are unknown. In these cases, the CMIP strategy constitutes a very fast and efficient alternative for pre-equilibration of the system by the introduction of a limited set of solvent molecules around the protein. The CMIP is able to capture most of the crystal water molecules reported in several protein structures. More importantly, the pre-equilibration of the structure of very large proteins using the CMIP leads to a dramatic improvement in the quality of the MD trajectories, with a very small increase in the computational effort.

ACKNOWLEDGMENTS

The authors are grateful to Professor J.L. Sussman for giving us the details of the crystallographic data of the acetylcholinesterase–huprine X complex before publication, and to Professor I. Fita for many helpful comments on catalase simulations. A fellowship from the Ministerio de Educación y Cultura (to X.B.) is gratefully acknowledged.

REFERENCES

1. Brooks CL, Karplus M, Pettitt BM. Proteins. A theoretical perspective of dynamics, structure and thermodynamics. New York: John Wiley & Sons; 1988.
2. Kollman P. Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 1993;93:2395–2417.
3. Kollman P. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. Acc Chem Res 1996;29:461–469.
4. Kollman P, Merz K. Computer modeling of the interactions of complex molecules. Acc Chem Res 1990;23:246–252.
5. Karplus M, Petsko GA. Molecular dynamics simulations in biology. Nature 1990;347:631–639.

6. Van Gunsteren WF, Mark AE. On the interpretation of biochemical data by molecular dynamics computer simulation. Eur J Biochem 1992;204:947–961,
7. Van Gunsteren WF, Berendsen HJC. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. Angew Chem Int Ed Engl 1990;29:992–1023.
8. Scott WR, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF. The GROMOS Biomolecular simulation package. J Phys Chem A 1999;103:3596–3607.
9. Darden TE, York D, Pedersen L. Particle mesh Ewald: an $n-\log(N)$ method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10009.
10. Van Gunsteren WF, Luque FJ, Timms D, Torda AE. Molecular mechanics in biology: from structure to function, taking account of solvation. Annu Rev Biophys Biomol Struct 1994;23:847–863.
11. Tirado-Rives J, Orozco M, Jorgensen WL. Molecular dynamics simulations of the unfolding of barnase in water and 8M aqueous urea. Biochemistry 1997;36:7313,7329.
12. Shields G, Laughton CA, Orozco M. Molecular dynamics simulation of the $d(A \cdot T–T)$ triple helix. J Am Chem Soc 1997;119:7463–7469.
13. Orozco M, Luque FJ. Theoretical methods for the description of the solvent effect in biomolecular systems. Chem Rev 2000;100: 4187–4225.
14. Blaney JM, Weiner PK, Dearing A, Kollman P, Jorgensen EC, Oatley SJ, Burridge JM, Blake CCF. Molecular mechanics simulation of protein–ligand interactions: binding of thyroid hormone analogues to prealbumin J Am Chem Soc 1982;104:6424–6434.
15. Mehler EL, Solmajer T. Electrostatic effects in proteins: comparison of dielectric and charge models. Protein Eng 1991;4:903–910.
16. Gilson MK, Honig B. Calculations of the total electrostatic energy of a macromolecular system: solvation energies, binding energies and conformational analysis. Proteins 1988;4:7–18.
17. Champness JN, Bennett MS, Wien F, Visse R, Summers WC, Herdewijn P, Declercq E, Ostrowski T, Jarvest RL, Sanderson MR. Exploration of the active site of herpes simplex virus thymidine kinase by x-ray crystallography of the enzyme in complex with its substrate, antiviral drugs and other nucleoside analogues. Proteins 1998;32:350–361.
18. Harel M, Schalk I, Ehret-Sabatier L, Bouet F, Goeldner M, Hirth C, Axelsen P, Silman I, Sussman JL. Quaternary ligand binding to aromatic residues in the active-site gorge of acetylcholinesterase. Proc Natl Acad Sci USA 1993;90:9031–9035.
19. Raves ML, Harel M, Pang YP, Silman I, Kozikowski AP, Sussman JL. Structure of acetylcholinesterase complexes with the nootropic alkaloid (−)–huperzine A. Nature Struct Biol 1997;4:57–63.
20. Mate MJ, Zamocky M, Nykyri LM, Hertzog C, Alzari PM, Betzel C, Kolle F, Fita I. Structure of catalase-A from *Saccharamoyces cerevisiae*. J Mol Biol 1999;268:135–149.
21. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes J Comput Phys 1977;23:327–341.
22. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz K, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. An all-atom empirical energy function for the simulation of nucleic acids. J Am Chem Soc 1995;117:11946–11975.
23. Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
24. Alemán C, Canela EI, Franco R, Orozco M. PAPQMD—a new strategy of the evaluation of force parameters from quantum mechanical calculations. J Comp Chem 1991;12:664–674.
25. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restrains from deriving atomic charges: the RESP model. J Phys Chem 1993;97: 10269–10280.
26. Case DA, Pearlman DA, Caldwell JW, Cheatham TE, Ross WS, Simmerling CL, Darden TA, Merz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Tsui V, Radmer RJ, Duan Y, Pitera J, Massova GL, Seibel UC, Singh PK, Weiner PK, Kollman PA. AMBER 6. San Francisco: University of California; 2000.
27. Gelpi JL, Luque FJ, Orozco M. CMIP computer program. University of Barcelona. 2000. Program is available upon request to the authors.
28. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 1985;28:849–857.
29. Feig M, Pettitt BM. Crystallographic water sites from a theoretical perspective. Structure 1998;6:1351–1354.
30. Izawa S, Inoue Y, Kimura A. Importance of catalase in the adaptative response to hydrogen peroxide: analysis of acatalasaemic *Saccharomyces cerevisiae*. Biochem J 1996;320:61–67.
31. Schonbaum GR, Chance B. Catalase. In: Boyer PD, editor. The enzymes. Vol 13. New York: Academic Press; 1976. p 363–408.
32. Klotz MG, Klassen GR, Loewen PC. Phylogenetic relationships among prokaryotic and eukaryotic catalases. Mol Biol Evol 1997; 14:951–958.
33. Chance B, Herbert D. The enzyme–substrate compounds of bacterial catalase and peroxides. Biochem J 1950;46:402–414.
34. Chance B. Effect of pH upon the reaction kinetics of the enzyme–substrate compounds of catalase. J Biol Chem 1952;194 471–481.
35. Fita I, Rossmann MG. The active center of catalase. J Mol Biol 1985;185:21–37.
36. Chen MS, Wlaker J, Prusoff WH. Kinetic studies of herpes simplex virus type 1-encoded thymidine and thymidylate kinase, a multifunctional enzyme. J Biol Chem 1979;254:10747–10753.
37. Waldman SS, Haeusslein E, Milman G. Purification and characterization of herpes simplex virus (type 1) thymidine kinase produced in *Escherichia coli* by a high efficiency expression plasmid utilizing a lambda PL promoter and cI857 temperature-sensitive repressor J Biol Chem 1983;258:11571–11575.
38. Cheng YC, Huang ES, Lin JC, Mar EC, Pagano JS, Dutschman GE, Grill SP. Unique spectrum of activity of 9-[(1,3-dihydroxy-2-propoxy)methyl]-guanine against herpesviruses in vitro and its mode of action against herpes simplex virus type 1. Proc Natl Acad Sci USA 1983;80:2767–2770.
39. Benzi G, Moretti A. Is there a rational for the use of acetylcholinesterase inhibitors in the therapy of Alzheimer's disease. Eur J Pharmacol 1998;346:1–13.
40. Brufani M, Filocamo L, Lappa S, Maggi A. New acetylcholinesterase inhibitors. Drugs Future 1997;22:397.
41. Camps P, El Achab R, Görbig DM, Morral J, Muñoz-Torrero D, Badia A, Baños JE, Vivas NM, Barril X, Orozco M, Luque FJ. Synthesis, in vitro pharmacology, and molecular modeling of very potent tacrine–huperzine A hybrids as acetylcholinesterase inhibitors of potential interest for the treatment of Alzheimer's disease. J Med Chem 1999;42:3227.
42. Barril X, Orozco M, Luque FJ. Predicting relative binding free energies of tacrine–huperzine A hybrids as inhibitors of acetylcholinesterase. J Med Chem 1999;42:5110–5119.
43. Camps P, Cusack B, Mallender WD, El Achab R, Morral J, Muñoz-Torrero D, Rosenberry TL. Huprine X is a novel high-affinity inhibitor of acetylcholinesterase that is of interest for the treatment of Alzheimer's disease. Mol Pharmacol 2000;57:409.
44. Camps P, El Achab R, Morral J, Muñoz-Torrero D, Badia A, Baños JE, Vivas NM, Barril X, Orozco M, Luque FJ. New tacrine–huperzine A hybrids (huprines): highly potent tight-binding acetylcholinesterase inhibitors of interest for the treatment of Alzheimer's disease. J Med Chem 2000;43:4657–4666.