# Towards molecular dynamics simulation of large proteins with a hydration shell at constant pressure

Valère Lounnas, Susanna K. Lüdemann, Rebecca C. Wade*

*EMBL, Meyerhofstrasse 1, Postfach 10.2209, 69012 Heidelberg, Germany*

## Abstract

Molecular dynamics simulation of a large protein in explicit water with periodic boundary conditions is extremely demanding in terms of computation time. Consequently, we have sought approximations of the solvent environment that model its important features. Here, we describe our SAPHYR (Shell Approximation for Protein HYdRation) model in which the protein is surrounded by a shell of water molecules maintained at constant pressure. In addition to the usual pairwise interatomic interactions, these water molecules are subjected to forces approximating van der Waals and dipole–dipole interactions with the implicit surrounding bulk solvent. The SAPHYR model is tested for a system of one argon atom in water and for the protein ubiquitin, and then applied to cytochrome P450cam, a protein with over 400 residues. The results demonstrate that structural and dynamic properties of the simulated systems are improved by use of the SAPHYR model, and that this model provides a significant computational saving over simulations with periodic boundary conditions. © 1999 Elsevier Science B.V. All rights reserved.

*Keywords:* Molecular dynamics simulation of proteins; Protein hydration; Pressure (in simulations); Implicit solvent model; Ubiquitin; Cytochrome P450

## 1. Introduction

Molecular dynamics simulations (MDS) of large proteins challenge computational resources. At present, simulations of proteins of several hundred residues in an explicit aqueous environment pose problems of computational feasibility due to the large amounts of CPU time required. Such macromolecular systems, when properly immersed in minimal periodic boxes of solvent molecules, comprise a large number of atoms: typically approximately $10^5$ atoms for a protein of approximately 500 residues. For condensed systems of such size, integrating Newton's equations of motion over times of even several hundred picoseconds requires large amounts of computing resources. This computational problem is a barrier to the widespread use of MDS for large proteins in explicit aqueous medium and few such simulations have been reported in the literature.

* Corresponding author. Fax: +49 6221 387517; e-mail: wade@embl-heidelberg.de

To illustrate the situation, we here consider the protein, cytochrome P450cam, a member of the ubiquitous and important cytochrome P450 (P450) superfamily [1,2]. It has more than 400 residues and its mechanism of action has been extensively studied by many different experimental approaches [3,4]. However, owing to its size, published MDS of this system has been restricted to simulations in which the number of dynamic atoms is limited in order to achieve computational feasibility. To study properties dependent predominantly on local dynamics, the dynamics of a small region of the protein–solvent system have been simulated while keeping the rest of the protein fixed. For example, substrate binding modes and affinities have been successfully studied by simulating only the dynamics of a spherical region centred on the active site [5,6]. However, many protein properties depend on their global dynamics and that of water molecules. In P450cam, the active site is buried in the protein and dynamic motions of the protein are clearly required for the substrate to access the active site [7]. Binding and catalysis involve diffusion of water molecules to and from the active site via channels [8]. Binding of substrate at the active site is influenced by binding of a cation at the protein surface [9]. The use of MDS to obtain an atomic-detail understanding of any of these three aspects of P450cam requires a full treatment of the protein dynamics in its solvent environment. The goal of the work described in this paper is to develop a model for simulating proteins with a hydration shell that is sufficiently complete to address issues such as these in MDS and yet computationally significantly more efficient than simulation of proteins under periodic boundary conditions.

MDS of proteins with a hydration shell is quite common. The size and construction of the shell varies. Frequently, MDS is performed considering only crystallographically resolved waters or a thin layer of water molecules over the protein surface (see, e.g. in the case of P450cam, references [10–12]). However, such a reduction of the solvent phase can cause spurious artifacts in the protein dynamics which are principally due to, firstly, the loss of solvent-induced dielectric shielding, secondly, the introduction of artificial

surface tension forces resulting from the presence of an interface with vacuum, and thirdly, missing protein–solvent hydrogen-bonds.

The lack of proper dielectric screening at the surface of the protein and in its interior causes structural deformation and damping of protein dynamics due to overestimated electrostatic interactions among polar and charged groups which should be shielded by the external electrostatic field of the bulk solvent environment. Attempts to alleviate this problem include using a distance-dependent dielectric constant and/or neutralizing charged groups (see, e.g. in the case of P450cam, references [10–13]). These modifications improve computational efficiency but, as they are usually uniformly implemented for all intramolecular interactions, they fail to preserve the difference between the dielectric properties of the protein interior and its exterior, and thus may lead to biased molecular dynamics.

On the other hand, the lack of explicit water molecules in the bulk solvent leads to missing packing and hydrogen-bond interactions between protein and solvent. This causes unbalanced forces at the protein surface, inducing a surface tension effect that hinders protein motions, and is principally characterised by a reduced radius of gyration and frozen atomic fluctuations [14]. A detailed study has been carried out on myoglobin showing that the surface tension artifacts on the protein dynamics diminish as the number of explicit water molecules present around the protein increases and reaches higher degrees of hydration [14]. However, at hydration degrees at which surface tension artifacts completely disappear, the number of explicit water molecules substantially exceeds the number required to form the first few layers of primary hydration, approximately within 5 Å of the protein nuclei, and matches the amount necessary for MDS under periodic boundary conditions.

A number of methods of alleviating these problems in simulations with hydration shells have been proposed. One approach is stochastic boundary dynamics in which the water molecules located at the surface of a sphere consisting of part or all of a protein–solvent system are subjected to Langevin dynamics [15]. This permits the frictional and stochastic diffusive influences

of the bulk solvent to be accounted for and thus eliminates some of the spurious interfacial effects between the explicitly simulated system and the surrounding vacuum [16,17]. However, the non-collisional physical effects of the implicit solvent phase are reduced to including only the mean field of van der Waals forces, neglecting Coulombic interactions. Again, this can cause biased dynamical behaviour of the protein especially when the hydration layer is reduced to a coverage only a few angstroms thick.

One approach to account for the dielectric properties of the surroundings of an explicit spherical region is the SCAAS model [18] in which surface constraints are introduced to model surface polarisation effects. Another is the solvent boundary potential developed by Beglov and Roux [19] which accounts for the van der Waals and electrostatic interactions of an infinite bulk system surrounding simulated solute and water molecules in a sphere of variable radius. This has been applied to small systems including alanine dipeptide. While it is clearly advantageous for the purposes of controlling pressure and surface tension for the system simulated to be spherical [20], proteins are not spherical and thus, in a spherical system, additional explicit water molecules are simulated that are of little importance for the protein dynamics. Thus for computational efficiency, we require a model that includes only an explicit hydration layer around the solute.

Beglov and Roux [21] simulated a non-spherical system by using a half-harmonic restraining potential to prevent water molecules escaping from a system in which the water layer thickness around a solute was dynamically adjusted to account for fluctuations of the solute conformation during Langevin dynamics simulations. The dielectric properties of the bulk solvent were not accounted for. Juffer and Berendsen [22,23] developed a more detailed boundary model, called dynamic surface boundary conditions, in which a simulated cluster of atoms is surrounded by special boundary atoms that can move but are position-restrained by a harmonic potential. The dielectric properties of the bulk solvent were accounted for with a reaction field computed by the boundary element method. However, this model is computationally quite demanding.

Given the pros and cons of the above models, our SAPHYR (Shell Approximation for Protein HYdRation) model, which is described in the next section, is designed to provide a computationally efficient representation of the dominant packing and electrostatic effects of the implicit bulk solvent surrounding a protein with a hydration shell of approximately 5 Å thickness. For computational efficiency, simple analytical formulations are used to compute van der Waals and electrostatic contributions of the bulk solvent. The hydration shell can be irregular in shape and is able to adjust to changes in the solute structure. Furthermore, coupling of the shell thickness to the pressure bath maintains the system at constant pressure. Tests on a model system and a well-characterised protein (ubiquitin) show how our model compares with other hydration models. In the final section, we show that our hydration shell model can usefully be applied to large proteins by giving results for simulations of P450cam in aqueous solution.

## 2. Theory: SAPHYR hydration shell model

### 2.1. Definition of hydration shell model

The model of the hydration shell that we have developed in order to circumvent the artifactual effects that occur when simulating a protein with a hydration shell in vacuo is depicted schematically in Scheme 1. In this model, the hydration around the protein is represented by an approximately 5-Å thick layer of explicit water molecules covering the protein surface, Zone A, and an infinite region, Zone B, extending beyond Zone A. The thickness of Zone A, $R_{layer} = 5$ Å (see below), is defined with respect to the atomic loci at the protein surface. Thus, the distance to the protein surface of any explicit water molecule, $j$, in Zone A is given by the distance between the water oxygen centre and the closest (either hydrogen or non-hydrogen) protein atom, $i$, and denoted $r_j^{clos} = r_{ij}$, see Scheme 1. The closest atom $i$ and the distance $r_j^{clos}$ are calculated at very little computational cost every timestep of the simulation (see Appendix A).
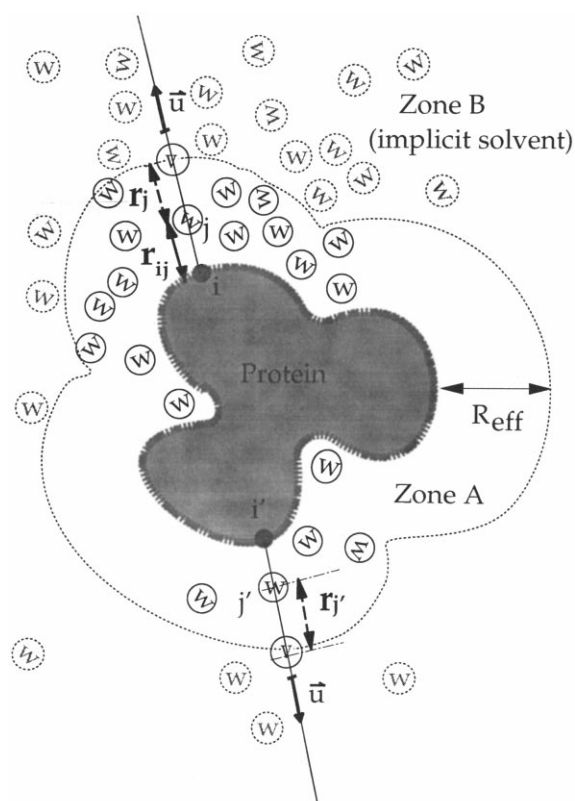
Scheme 1. The hydration shell model. The shaded region symbolizes the protein interior. The explicit water molecules in Zone A are represented by circles marked by the letter W, whereas the implicit water molecules of Zone B are represented by dotted circles marked with W. The position of the repulsive Lennard–Jones wall is symbolized by the dotted line between Zones A and B of the solvent. The solvent forces due to implicit Zone B act on each explicit water molecule $j$ of Zone A along the line connecting the closest protein atom $i$ to its oxygen atom; to compute the Lennard–Jones repulsion, a virtual water molecule $V$ is considered to be located on this line at the distance delimiting the position of the repulsive wall. The unit vector $\vec{\mathbf{u}}$ points along this line and away from the protein surface towards Zone B.

## 2.2. Forces due to implicit solvent (in Zone B) on explicit water molecules (in Zone A)

In Zone B, the solvent phase is treated implicitly and its effect on the explicit water molecules in Zone A is twofold.

### 2.2.1. Van der Waals forces due to implicit solvent on explicit water molecules

First, any water molecule of Zone A interacts at any instant, $t$, during the simulation with the

van der Waals core of a virtual water molecule located on the line connecting it to the closest protein atom at an effective distance $R_{\text{eff}} = R_{\text{layer}} + \sigma_w^T$ from the protein atom. The parameter $\sigma_w^T$ stands for the van der Waals thermal diameter for the explicit water model chosen for the simulation (in this work either SPC/E [24] or TIP3P [25]). The interaction with the virtual water molecule consists only of the repulsive term of the Lennard–Jones potential of the oxygen atom. The corresponding additional force on explicit water molecule $j$ in Zone A is

$$\vec{\mathbf{F}}_{\text{wall}}^{\text{LJ}-12}(r_j) = \left\{ -\frac{48\,\varepsilon_w\,\sigma_w^{12}}{r_j^{13}} \right\} \cdot \vec{\mathbf{u}} \qquad (1)$$

where $\sigma_w$ and $\varepsilon_w$ are the classical Lennard–Jones distance and energy parameters for the water/water interaction. $r_j$ is the distance from the explicit water, $j$, in Zone A to the centre of its corresponding virtual water in Zone B and is given by

$$r_j = R_{\text{eff}} - r_j^{\text{clos}} = R_{\text{layer}} + \sigma_w^T - r_j^{\text{clos}}. \qquad (2)$$

The unit vector $\vec{\mathbf{u}}$ is parallel to the line passing through atoms $i$ and $j$ and is oriented in the direction $i$ to $j$, i.e. pointing away from the protein surface (cf. Scheme 1). Similar techniques of confinement have been employed in other simulation methodologies [21].

### 2.2.2. Electrostatic forces due to implicit solvent on explicit water molecules

The second effect that the implicit solvent phase in Zone B exerts on the water molecules in Zone A takes account of the average electrostatic influence of the water molecules that are not explicitly modelled in Zone B. This arises from the forces that 'freely' rotating dipoles, like water molecules, experience between each other when they are in gaseous or condensed phase. Because of thermal fluctuations, any pair of permanent dipoles at respective locations $\vec{\mathbf{r}}_1$ and $\vec{\mathbf{r}}_2$ in a gas or a fluid can instantaneously adopt distinct relative orientations, see Scheme 2. The potential energy $E_p$ of such a pair of interacting dipoles can be evaluated according to the following clas-

sical expression:

$$E_p = E_p(\vec{\mathbf{r}}_{12}) = -\vec{\boldsymbol{\mu}}_2 . \vec{\mathbf{E}}_1(\vec{\mathbf{r}}_{12}) \qquad (3)$$

where $\vec{\mathbf{r}}_{12} = \vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1$ represents the vector joining the two dipoles, $\vec{\boldsymbol{\mu}}_2$ is the dipole moment of the second dipole, and $\vec{\mathbf{E}}_1(\vec{\mathbf{r}}_{12})$ is the electrostatic field vector that the first dipole creates at the location of the second one. $\vec{\mathbf{E}}_1(\vec{\mathbf{r}}_{12})$ can be further expressed as the negative gradient of the electrostatic potential $V_1(\vec{\mathbf{r}}_{12})$ created by the first dipole $\vec{\boldsymbol{\mu}}_1$:

$$\vec{\mathbf{E}}_1(\vec{\mathbf{r}}_{12}) = -\nabla V_1(\vec{\mathbf{r}}_{12}), \qquad (4)$$

and $V_1(\vec{\mathbf{r}}_{12})$ itself has a simple analytical expression:

$$V_1(\vec{\mathbf{r}}_{12}) = \frac{\vec{\boldsymbol{\mu}}_1 . \vec{\mathbf{r}}_{12}}{4\pi\varepsilon_0 r_{12}^3} \qquad (5)$$

provided that the distance between the two dipoles $|\vec{\mathbf{r}}_{12}|$ is substantially larger than the charge separation within the dipole itself. The charge separation in a water molecule can be considered to be $\sim 1$ Å and the minimum possible distance between two water molecules to be $\sim 3$ Å. Therefore, we assume that the above condition is valid for MDS of liquid water.
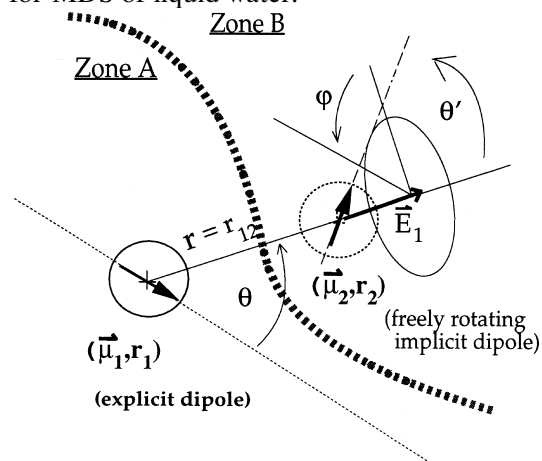


Scheme 2. Diagram showing parameters for defining the interaction between a fixed dipole of an explicit solvent molecule in Zone A with a freely rotating implicit dipole in Zone B. The interface between the two regions is marked by a thick dotted line.

From Eq. (5), analytical expressions for the net average interaction energies between a freely rotating dipole and a fixed dipole and, in addition, between two freely rotating dipoles can be derived using the following integration schemes. Assume that one of the two dipoles, say the first dipole located at $\vec{\mathbf{r}}_1$, is fixed and that the second one, at $\vec{\mathbf{r}}_2$, is free to rotate because of its thermal energy. Then, the relative occurrence of all the possible orientations of the second dipole with respect to the fixed dipole must obey a Boltzmann distribution. Therefore, the average potential energy $\xi(r,\theta)$ of this ensemble of configurations can be expressed as

$$\xi(r,\theta) = \langle \mathbf{E}_p \rangle_\Omega = \frac{\displaystyle\int_{\Omega(\theta',\varphi)} E_p e^{-\beta E_p} d\Omega}{\displaystyle\int_{\Omega(\theta',\varphi)} e^{-\beta E_p} d\Omega}, \qquad (6)$$

where $\Omega$ and $\beta$ are respectively the configurational space describing the relative position between the two dipoles and $(kT)^{-1}$. Choice of an appropriate system of spherical coordinates as illustrated in Scheme 2 allows a straightforward integration combining Eqs. (3)–(6) (see, e.g. Lorrain and Corson [26]) and leads to a simple analytical expression

$$\xi(r,\theta) = \langle E_p \rangle_\Omega = -\frac{\mu^4(3\cos^2\theta + 1)}{3kT(4\pi\varepsilon_0)^2 r^6}, \qquad (7)$$

where $\mu = \mu_1 = \mu_2$ in the case of identical permanent dipoles, $r = |\vec{\mathbf{r}}_{12}|$, and $\theta$ is the angle between the dipole moment vector $\vec{\boldsymbol{\mu}}_1$ of the fixed dipole and the field vector $\vec{\mathbf{r}}_{12}$ (cf. Scheme 2). The above expression is actually a mathematical approximation which is valid only when the permanent dipoles do not exceed a few debye at room temperature. Hence, this formula should be applicable for water molecules in condensed phase at 300 K and provides an adequate description of the average influence that any implicit water molecule located at $\vec{\mathbf{r}}_2$ in the continuum of Zone B would exert on any explicit water molecule at position $\vec{\mathbf{r}}_1$ at instant $t$ in Zone A. It should also be noted that, whatever the value of $\theta$, i.e. for any

possible orientation of the explicit water in Zone A, the average interaction potential energy with an implicit water in Zone B is always attractive and decreases as the sixth power of the separation distance.

Now, consider further that the explicit water of Zone A is indeed also rotating freely. Eq. (7) can then be further averaged over all possible relative orientations, i.e. by integrating it over all solid angle values defined by $\theta$. Note that in this case it is not necessary to introduce the Boltzmann weight factors again because they were already accounted for in the previous integration scheme. Eq. (7) thus further simplifies to

$$\bar{\xi}(r) = \langle \xi_p(r,\theta) \rangle_\theta = -\frac{2\mu^4}{3kT(4\pi\varepsilon_0)^2 r^6}. \quad (8)$$

For a water molecule with a dipole moment of $\sim 2.35$ debye, (as in the TIP3P and SPC/E models) at 298 K, Eq. (8) becomes

$$\bar{\xi}(r)_{(kJ/mol)} = -\frac{2.98 \times 10^4}{r^6_{(\text{Å})}}. \quad (9)$$

Hence, within the approximations stated above, the average attractive potential energy between two isolated near neighbour water molecules at a separation distance of 3 Å in liquid phase is $-40.8$ kJ/mol. This exceeds the upper bound of the energy value for a very strong hydrogen bond in water by approximately 30% (for TIP3P water, this is $-30$ kJ/mol at 2.7 Å separation [25]. This interaction energy drops rapidly to $-3.6$ kJ/mol for water molecules at a separation distance of 4.5 Å corresponding approximately to the second solvation shell. In a homogeneous bulk environment, the net effect of this attractive potential energy caused by the water surrounding any single water molecule cancels out because of the isotropy of the water pair distribution. However, near an interface with vacuum, the asymmetry of the pair distribution of water molecules across the interface results in unbalanced forces and strong surface tension effects.

In order to model the total induced effect of the implicit waters in Zone B, we use the above concepts in conjunction with the fact that Zone B can be represented as a continuum field, each point of which is characterised by a probability of water occupation. Neglecting the influence of the protein on the water structure in Zone B, it is possible to use the conditional pair correlation of water $g^{bulk}(r,\theta)$ together with Eq. (7) to carry out the following summation for any explicit water molecule $j$ in Zone A

$$E_{tot}^{dip}(r_j,\theta_j) = \int_{ZoneB} \xi(r,\theta) g^{bulk}(r,\theta) \rho^0 d\tau, \quad (10)$$

where $d\tau$ is the volume element describing the spatial integration over Zone B, and $(r,\theta)$ represent, as stated previously, the distance and orientation of the explicit water $j$ in Zone A relative to the field point in Zone B. $\theta_j$ is the orientation of water $j$ with respect to the interface. $\rho^0$ is the average density of bulk water and serves to normalise the summation. Note that this summation implicitly assumes that the contributions from all the continuum field points are purely additive, i.e. that the dipole/dipole interactions of two different water molecules with a third one are pairwise additive. Three body terms and higher correlations are neglected in this integration scheme. We will show in Section 4 how these non-additive non-negligible effects can be accounted for by an overall scaling factor.

The integral in Eq. (10) is difficult to evaluate when the interface between the explicit and the implicit region has a complicated geometry, see Scheme 3. At any instant $t$, a given water molecule $j$ in Zone A has an instantaneous position and orientation that is unique. Hence, evaluating this integral in its full detail by numerical means for many explicit water molecules at every step of an MD trajectory, though feasible in principle, would in practice be a cumbersome task. Since our purpose is to alleviate surface artifacts at low computational cost, we further simplify the integral in Eq. (10). First, inspection of Eq. (7) shows that it is reasonable to replace $\xi(r,\theta)$ by its average over $\theta$ values, i.e. $\bar{\xi}(r)$ of Eq. (8). Indeed, the interaction energies from Eq. (7) range only between values which are half and twice the value of $\bar{\xi}(r)$ from Eq. (8). Second, the interface

between Zone A and Zone B is assumed to be planar everywhere. The latter assumption transforms the domain of integration, Zone B, into a new domain $\Omega_1(r_j)$ which is half spherical and of cylindrical symmetry with respect to the line connecting the centre of explicit water $j$ to the centre of the closest protein atom (cf. Scheme 3). These two simplifications allow a summation which, while preserving the essential features of the integral, leads to the following expression

$$E_{\text{tot}}^{\text{dip}}(r_j) = \int_{\Omega_1(r_j)} \bar{\xi}(r) g^{\text{bulk}}(r) \rho^0 \text{d}\Omega_1. \quad (11)$$

The dependence of the summation reduces to only $r_j$ and consequently, the degree of burial of the explicit water $j$ in Zone A (see Scheme 3).

Alternatively, the resultant force acting on wa-
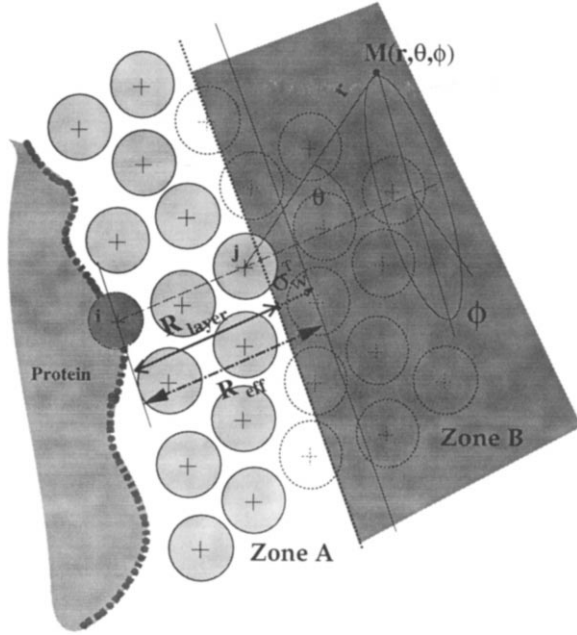


Scheme 3. Detail of the interface between the explicit hydration shell (Zone A) and the implicit continuum (Zone B) (cf. Scheme 1) showing how this interface is assumed to be planar in order to calculate the integral in Eq. (13). The protein surface is represented by the dotted line with the shaded circle representing protein atom $i$. Explicit water molecules are represented by light shaded circles and implicit water molecules are represented by dotted circles. The volume of integration in Eq. (13) is shown by the dark shaded area.

ter $j$ across the planar interface can be evaluated directly:

$$\vec{F}_{\text{tot}}^{\text{dip}}(r_j) = -\int_{\Omega_1(r_j)} \nabla[\bar{\xi}(r)] g^{\text{bulk}}(r) \rho^0 \text{d}\Omega_1. \quad (12)$$

Because of the cylindrical symmetry of domain $\Omega_1$, all the components of the forces parallel to the interface mutually cancel out resulting in an overall sum of forces which acts only along the axis perpendicular to the interfacial plane. Therefore, using the gradient of Eq. (8), Eq. (12) can be expressed in spherical coordinates as

$$\vec{F}_{\text{tot}}^{\text{dip}}(r_j).\vec{u} = \frac{8\pi\rho^0\mu^4}{kT(4\pi\varepsilon_0)^2} \int_{\theta=0}^{\frac{\pi}{2}} \int_{r_{\text{min}}}^{\infty}$$
$$\frac{g^{\text{bulk}}(r)}{r^7} \cos\theta\sin\theta\,\text{d}\theta\,\text{d}r. \quad (13)$$

If we further assume that $g^{\text{bulk}}(r)$ is a step function such that $g^{\text{bulk}}(r) = 0$ for $r < \sigma_w^T$ and $g^{\text{bulk}}(r) = 1$ for $r \geq \sigma_w^T$, then this integral can be solved analytically and gives:

$$\vec{F}_{\text{tot}}^{\text{dip}}(r_j) = \frac{A}{\left(r_j - \sigma_w^T\right)^4}\vec{u} \quad \text{if } |r_j - \sigma_w^T| \geq \sigma_w^T$$

and

$$\vec{F}_{\text{tot}}^{\text{dip}}(r_j) = \frac{A\left(3\sigma_w^{T^2} - 2\left(r_j - \sigma_w^T\right)^2\right)}{\sigma_w^{T^6}}\vec{u}$$
$$\text{if } |r_j - \sigma_w^T| < \sigma_w^T,$$

with

$$A = S_{\text{fit}} x\left(\frac{\pi\rho^0\mu^4}{3kT(4\pi\varepsilon_0)^2}\right). \quad (14)$$

The scaling factor $S_{\text{fit}}$ is introduced here to fit the analytically derived force expression to the one obtained by MDS. That is, the summation of Eq. (13) can also be evaluated either by numerical integration using the known $g^{\text{bulk}}(r)$ of bulk water, or from MDS by pairwise averaging all elec-

trostatic force contributions across an imaginary interface drawn in a periodic system containing explicit water molecules in both Zone A and Zone B according to

$$F_{MDS}^{elec}(r) = \langle \vec{\mathbf{F}}(r).\vec{\mathbf{u}} \rangle_{MDS}$$

$$= \frac{1}{N_{norm}} \left( \sum_{k=1}^{N_{step}} \sum_{j \epsilon \, ZoneA} \delta\left(r_j - \sigma_w^T - r\right) \right.$$

$$\left. \left[ \sum_{j' \epsilon \, ZoneB} \vec{\mathbf{f}}_{j \leftarrow j'}(t_k).\vec{\mathbf{u}} \right] \right) \qquad (15)$$

where

$$\vec{\mathbf{f}}_{j \leftarrow j'}(t_k) = \sum_{i' \epsilon j'} \sum_{i \epsilon j} \frac{q_{i'} q_i \vec{\mathbf{r}}_{i'i}}{4\pi\epsilon_0 r_{i'i}^3} \qquad (15a)$$

with $i$ and $i'$ being the atomic indices of the oxygen and hydrogen atoms for each pair of water molecules $(j, j')$ and the overall discrete summation being normalised by

$$N_{norm} = \sum_{k=1}^{N_{step}} \sum_{j \epsilon ZoneA} \delta\left(r_j - \sigma_w^T - r\right). \qquad (15b)$$

We used a simple model system composed of a single argon atom embedded in a periodic box of water molecules to establish the functional form of $F_{MDS}^{elec}(r)$ across virtual planar and curved interfaces located 10 Å from the argon atomic centre. The function thus obtained from the MD trajectory served to assess the quality of both analytical and numerical summations of Eq. (13), (see Fig. 1).

### 2.2.3. Empirical solvent implicit forces

In addition to the 'exact' force profile analytically derived from the dipole–dipole interaction formalism described above, we examined the properties of two other empirical force profiles based on the $r^{-4}$ dependence observed in Eq. (14). These two empirical force profiles are in fact
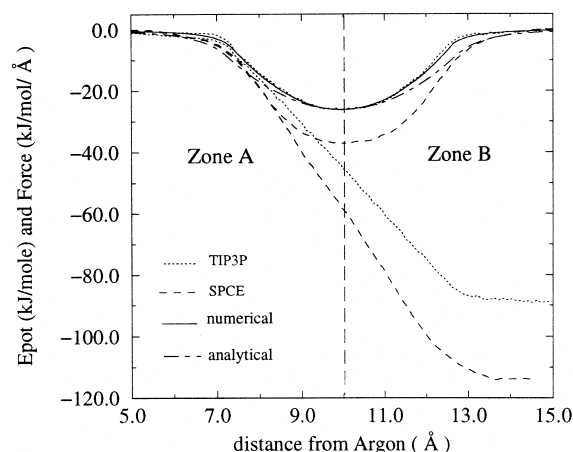


Fig. 1. Electrostatic component of the forces on an explicit water molecule due to solvent in Zone B computed from simulations of an argon atom in a periodic box of explicit water molecules. An imaginary planar interface (dashed vertical line) is positioned 10 Å from the argon nucleus separating the simulation periodic box into Zone A and Zone B. The electrostatic forces are attractive and thus oriented in the direction from Zone A to Zone B. The electrostatic mean force (inverted bell shape curve) and the total electrostatic potential energy (S-shaped curve) are shown for the TIP3P (dotted lines) and SPC/E models (dashed lines). The solid line results from numerical integration of Eq. (13) and fitting with a scaling factor of $1/6.67$ to the total electrostatic mean force for the TIP3P model. The dot-dashed line results from analytical integration of Eq. (13) using a step function for $g(r)$ and fitting with a scale factor of $1/3.15$ to the total electrostatic mean force for the TIP3P model.

identical to the 'exact' force profile in the region where $(r_j - \sigma_w^T) \geq \sigma_w^T$, i.e. when the van der Waals envelope of an explicit water molecule is not in contact with the continuum region. However, they differ both from the 'exact' profile and from each other in the interfacial region $(r_j - \sigma_w^T) < \sigma_w^T$ where the explicit and the continuum solvent descriptions overlap. These two additional force profiles will be referred to in this work as 'empirical 1' and 'empirical 2'.

The 'empirical 1' force profile is defined as follows:

$$\vec{\mathbf{F}}_{tot}^{emp-1}(r_j) = f_{scal}\left( \frac{\pi \rho^0 \mu^4}{3kT(4\pi\varepsilon_0)^2} \right) \frac{\vec{\mathbf{u}}}{r_j^4} \qquad (16)$$

for all $r_j > 0$, and where $f_{scal} = 2.0$ was chosen to obtain a minimum value of the corresponding

mean force potential of approximately $-50$ kJ/mol at $r_j = \sigma_w^T$. In contrast to the 'exact' profile, the 'empirical 1' profile contains a singularity at $r_j = 0$ which prevents an explicit water $j$ from crossing the interface by more than a distance equal to $\sigma_w^T$ during the simulation. On the other hand, this expression does not exclude the force from becoming dramatically large at small $r_j$ causing the system to explode. However, as shown later in this section (cf. Eq. (20)), this artifact can be avoided by implementing this force in combination with the Lennard−Jones repulsive force (LJ-12) described at the beginning of this section (cf. Eq. (1)). Indeed, in this case the LJ-12 repulsive term, with $r^{-13}$ dependence, always dominates the implicit dipole attractive term, with $r^{-4}$ dependence, as $r_j$ tends to 0, hindering explicit solvent molecules from reaching $r_j \sim 0$.

The 'empirical 2' force profile is defined as follows:

$$\vec{F}_{tot}^{emp-2}(r_j) = f_{switch}(r_j) \frac{A}{\left(r_j - \sigma_w^T\right)^4}$$
$$\left(1 - f_{switch}(r_j)\right) B\left(r_j - \sigma_w^T\right), \qquad (17)$$

for all $r_j$, and where the switching function

$$f_{switch}(r_j) = 1 - \exp\left\{ -\left(r_j - \frac{\sigma_w^T}{2\sigma_w^T}\right)^\alpha \right\}, \qquad (18)$$

allows a continuous transition between the inverse quartic ($r^{-4}$) and the linear ($r$) dependencies in Eq. (17) which correspond respectively to the two ranges $(r_j - \sigma_w^T) \geq \sigma_w^T$ and $(r_j - \sigma_w^T) < \sigma_w^T$ defined above. Because of the switching function, the singularity at $r_j = 0$ in 'empirical 1' is not present in 'empirical 2'. For the latter, parameter $A$ is identical to the one defined in Eq. (14), and parameter $B$ ($< 0$) is empirically chosen so that the maximum of the force around $r_j = 2\ \sigma_w^T$ is $-14$ and $-20$ kJ/mol/Å for, respectively, the TIP3 and SPC/E models. In this work, the values of B were thus $-5.65$ and $-7.23$.

Thus the three different force profiles are designed with a common characteristic which is to exert an attraction on the explicit water molecules

of Zone A along the unit vector $\vec{u}$ pointing away from the protein surface. The potential of mean force associated with each force profile can be derived by integrating the analytical expressions of Eqs. (14),(16),(17).

### 2.3. Control of the pressure of the hydration shell

Considering the Lennard−Jones repulsive forces described at the beginning of this section (cf. Eq. (1)) as well, the overall solvent-induced mean force acting on the explicit water molecules of Zone A and resulting from the modeled implicit continuum of Zone B can be summarized as

$$\vec{F}_{ZoneB}^{im}(r_j) = \left\{ \vec{F}_{tot}^{dip}(r_j) + \vec{F}_{wall}^{LJ-12}(r_j^*) \right\}\vec{u}, \qquad (20)$$

where $\vec{F}_{tot}^{dip}(r_j)$ can be selected from any of the three force profiles and $r_j$ is defined as in Eq. (2). $r_j^*$ is redefined as

$$r_j^* = R_{layer}^{dyn} + \sigma_w^T - r_j^{clos} \qquad (21)$$

in order to allow the position of the Lennard−Jones (LJ-12) confinement wall, defined by the parameter $R_{layer}^{dyn}$, to adapt dynamically as the pressure of the water in Zone A relaxes to an external (or reference) pressure $P_{ret}$ according to the following scheme

$$dR_{layer}^{dyn}(t) = C_{relax}(P_{water}(t) - P_{ref})\delta t, \qquad (22)$$

where $C_{relax}$ is an empirically adjusted relaxation coefficient and $\delta t$ is the simulation time step. The instantaneous pressure of water $P_{water}(t)$ at any simulation time step $t$, can be calculated from the kinetic energy of the centres of mass of the water molecules and the virial of the internal forces, $F_{ij}^{int}$, acting on the $N_w = N_w^{shell}$ number of water molecules forming the hydration layer around the protein:

$$P_{water}(t) = \frac{1}{3V_w}\left[ \sum_{j=1}^{N_w} \frac{p_i^2}{m_i} + \sum_{i=1}^{N_w-1} \sum_{j=i+1}^{N_w} \vec{r}_{ij} \cdot \vec{F}_{ij}^{int} \right].$$
$$(23)$$

In contrast to MDS with periodic boundary conditions, the volume $V_w$ of the explicit solvent layer around the protein surface must be evaluated. We chose to approximate this volume by

$$V_w = N_w^{\text{shell}} . v_w^0 . f_{\text{shell}}, \qquad (24)$$

where $v_w^0 \sim 29.9$ Å$^3$ is the bulk average volume occupied by a water molecule and $f_{\text{shell}} \sim 1$ is the density of the water in the hydration shell.

The coupling to a reference pressure that we report here is essentially similar to the one classically employed in MD simulations performed in the (N,P,T) ensemble with periodic boundary conditions. The main differences are that the location vectors of the water molecules are not scaled as the system dimensions are modified, and that the pressure is evaluated only for the water molecules in order to avoid the complication of determining the potentially highly fluctuating volume (due to partial folding or unfolding or to conformational changes) of the protein at every time step of the MD trajectory. Another point of importance is that the implicit forces of Eq. (20) are included in the forces $F_{ij}^{\text{int}}$ when computing the virial in Eq. (23). Indeed, in the evaluation of the pressure in a classical MD simulation using periodic boundary conditions, all the forces resulting from the interactions with the minimum image are included in the virial of internal forces and thus contribute to the pressure. In our scheme, the forces from the implicit solvent in Zone B correspond to the minimum image forces.

The SAPHYR hydration shell model was implemented in a molecular dynamics code as described in Appendix A.

## 3. Methods

### 3.1. Simulation protocols

Simulations were performed for three solutes either in a periodic box or with a hydration shell. The systems were set up as follows.

(1) One argon atom was immersed in a periodic cubic box of sides 29 Å containing 867 water molecules and all solvent molecules closer than 2.4 Å to the solute were eliminated. The box consisted of several pre-equilibrated boxes of 216 water molecules placed side by side. Starting coordinates for argon in a hydration sphere were obtained by removing water molecules beyond 10 Å from the argon atom.

(2) Coordinates of ubiquitin (including 58 water molecules) were taken from the crystal structure solved at 1.8 Å [27]. Polar hydrogens for the two proteins were built with the Whatif program [28] by optimizing the hydrogen bond network [29], and non-polar hydrogen atoms were added in standard locations. Two hundred additional water molecules were incrementally added using the GRID program [30,31]. This procedure was aimed at producing an optimised water coverage better equilibrated than that produced by the standard 'superimpose and delete' protocols. Finally, the system was superimposed on a $43.87 \times 48.5 \times 65.8$-Å$^3$ box of water molecules (pre-equilibrated as for the argon simulations) large enough that all protein non-hydrogen atoms were further than 10 Å from the box edge. Then, all solvent molecules closer than 2.4 Å to any non-hydrogen atom of the solute were eliminated and any water molecules added in protein cavities by the superposition procedure were removed leaving a total of 4481 water molecules. For simulations in a hydration shell, all water molecules further than 5 Å from any protein atom were removed resulting in a system with 491 water molecules.

(3) Coordinates of P450cam with camphor bound (including 216 water molecules) were taken from the crystal structure of the complex solved at 1.6 Å (2cpp in the Protein Data Bank) [32]. Protonation states and heme and camphor parameters were assigned as described elsewhere [33]. Sixteen positive counterions, 15 sodium and one potassium were added using the GRID program in order to ensure the electronegativity of the protein surface. Protons and further water molecules were added in the same way as for ubiquitin (approx. 1000 water molecules were added with the GRID program and the rest were added by superposition resulting in a 5-Å thick layer of approx. 1500 water molecules).

All simulations were carried out with the AR-GOS program [34]. For the argon and P450cam systems, the all-atom CHARMm22 [35] force field

was used, whereas for the ubiquitin systems, the all-atom Amber 4.0 force field [36] was used. Two solvent models, TIP3P [25] (as modified in CHARMm) and SPC/E [24] were used. Sodium and potassium cations were assigned parameters from reference [37]. Unless stated otherwise, an atom-based spherical cut-off of 12 Å was used in conjunction with a force shifting scheme for curtailing the long-range electrostatic interactions [38]. In the force shifting scheme, the power parameter $\beta$ was set to 4.

All systems were subjected to up to 100 steepest descent energy minimization steps before thermalisation. In all cases, the heating process was achieved in three successive simulations of 5 ps each at 100, 200 and 300 K with random velocity reassignment from a Maxwellian distribution every 100 timesteps. The equations of motion were integrated with the Verlet Leapfrog algorithm using a timestep of 2 fs. All bonds were constrained with the SHAKE algorithm [39] and the force pairlists were evaluated every 10 timesteps in all cases except for the computationally inexpensive system of argon in a spherical water droplet for which the force pairlists were updated every timestep.

After thermalisation, simulations were run at constant temperature with the solvent and the solute coupled separately to external temperature baths at 300 K using temperature relaxation times between 0.1 and 0.4 ps [40]. All simulations with periodic boundary conditions were carried out in the (N,P,T) ensemble with the system pressure maintained close to the reference pressure of 1 atm via a classical coupling utilizing positional scaling of the atomic coordinates [40]. For systems with a shell of water molecules, unless otherwise stated, the pressure of the solvent was maintained around the reference pressure of 1 atm by the special scheme described above. In all cases, the relaxation coefficient, $C_{\text{relax}}$, was set to the value of $10^{-9}$ nm/Pa/ps in order to match roughly a pressure relaxation time of 0.1 ps in the water shell.

The simulations were performed on Silicon Graphics Power Challenge computers with R8000 and R10 000 processors. The protein systems simulated with a hydration shell, force shifting

(with a 12-Å cutoff) and implicit solvent forces required approximately five- to sixfold less CPU time than the simulations in a periodic box without force shifting. However, it should be noted that neither the implementation of force shifting nor of the implicit solvent forces has been optimised. Computational efficiency is sensitive to the value of the non-bonded cutoff employed and the size of the system simulated. While it is possible to use shorter non-bonded cutoffs in periodic box simulations when long-range force treatments such as Ewald summation are employed, the shell model should still be more computationally efficient for large systems. This is because of better scaling of CPU time for force evaluation with system size, and because evaluation of the non-bonded pair-list is faster.

## 4. Results and discussion

### 4.1. Comparison of hydration shell model to MDS simulation under periodic boundary conditions

The forces due to implicit Zone B were first calibrated using simulations of one argon atom surrounded by water molecules. Two simulations of argon immersed in a cubic periodic system containing respectively TIP3P and SPC/E models were produced without force shifting and analysed in order to extract the quantity defined in Eq. (15). This quantity represents the pairwise averaging of all electrostatic force contributions across an imaginary planar interface drawn in the simulation box. We chose to position this virtual interface at a distance of 10 Å from the argon atom located in the centre of the periodic box. The dimensions of the simulation box were sufficiently large that the summation could be performed in a region where the solvent has reached or is close to bulk.

The profile thus obtained is shown in Fig. 1 and corresponds to the electrostatic attractive mean force that the molecules of Zone B exert on average on any other water molecule located either in Zone A or B, at a distance $r$ from the argon nucleus. At $r < 5$ Å, i.e. at a distance $> 5$ Å from the interface, the water molecules in Zone A do not on average experience any force

from the water molecules in Zone B. This is consistent with the $r^{-6}$ and therefore short-range dependence of the interaction potential between dipoles (see Section 2). As $r$ increases and the separation distance to the interface becomes shorter, the attractive mean force increases until it reaches a maximum value when the centre of the water molecule reaches the boundary between Zones A and B. Because the planar interface plays the role of an attractor for the water molecules in Zone A, the force is reported as negative in Fig. 1. In fact, its action with respect to the argon atom is positive and oriented in the direction of the unit vector $\vec{u}$ towards Zone B and perpendicular to the planar interface. When a water molecule starts to penetrate Zone B, the attractive force from the ensemble of other waters in Zone B necessarily tends to diminish until it vanishes. The reason is that the projection of the individual force contributions exerted on the penetrating water are no longer systematically applied in the direction pointing towards Zone B. The contributions from the water molecules closer to the interface than the penetrating water are indeed projected in the opposite direction towards Zone A. Sensitivity of the force profile to the choice of a planar interface can be assessed from the data in Fig. 2 in which the force profile derived with a planar interface is compared to force profiles derived with curved interfaces with radii of curvature of 5 and 10 Å. The shape and magnitude of the force profile is maintained but the position of the maximum force shifts slightly into Zone A by approximately 1 Å as the curvature changes from infinity to 5 Å.

As shown in Fig. 1, the mean attractive forces induced across the planar interface are enhanced by roughly one third for the simulation carried out with the SPC/E water model as compared to the TIP3P water model. This difference probably originates from the enhanced atomic partial charges on the oxygen and hydrogen sites of the SPC/E model with respect to those of the TIP3P model. Also shown in Fig. 1 are the average electrostatic potential energies for both TIP3P and SPC/E water as a function of $r$ across the interface. For $r > 13$ Å, these potential energies reach values of approximately $-89$ and $-115$
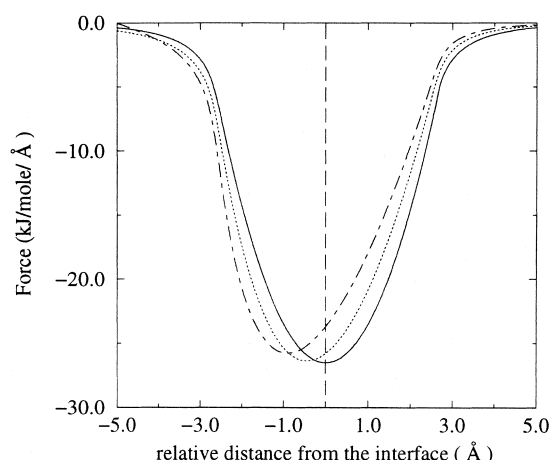


Fig. 2. Electrostatic component of the implicit solvent force obtained by numerical integration of Eq. (13) using interfaces of various curvatures. The solid, dotted and dashed lines represent respectively a planar interface, and interfaces with radii of curvature of 10 Å and 5 Å. The position of the interface is shown by the vertical dashed line. The magnitude of the force is fitted to the TIP3P water model (cf. Fig. 1).

kJ/mol, which correspond, after addition of the Lennard–Jones energy term and halving to derive average interaction energies, to total interaction energies of $-41.7$ and $-43.7$ kJ/mol, respectively. These are consistent with the reported values for these models [24,25].

### 4.2. Comparison of analytical forces due to implicit Zone B with forces from MDS under periodic boundary conditions

We observe exceptionally good agreement between the numerical integration of Eq. (12) [making use of the calculated $g(r)$ from periodic box simulations with TIP3P and SPC/E water models) and the MD average (cf. Fig. 1), provided that the numerical function is scaled by a factor of $1/6.7$ and $1/5.5$ for, respectively, the SPC/E and TIP3P water models. This scaling factor essentially accounts for the non-additive, higher correlation terms between water molecules in the liquid state which are neglected in the integral of Eq. (6). For example, in liquid water, hydrogen bonds prevent water molecules from freely rotating as they are assumed to do in the derivation of Eqs. (7),(8). In addition, at short separation dis-

tances (i.e. approx. 3 Å), where the dipole–dipole interaction dominates the electrostatic potential energy of water, the mathematical formulation of Eqs. (5),(7) reaches the limit of applicability. This effect is also accounted for in the overall scaling factor deduced from fitting the MDS trajectory average. It is thus quite remarkable to observe that the simple formalism from Eqs. (5)–(11) can, after fitting, provide a robust approximation of the MD trajectory average.

In practice, we have performed analytical integration using a step function for $g(r)$ (see Section 2). For this, the best scaling factors are $1/3.15$ and $1/2.3$ for the TIP3P and SPC/E models, respectively. The parameter $S_{\text{fit}}$ in the analytical expression of the forces in Eqs. (14),(17) was assigned the latter values to calculate the force parameters used in the simulations reported here. It should be noted that the agreement with the MD results is not as good for the analytical as for the numerical integrations. We have purposely chosen to match the minima of the curve (see Fig. 1) in order to preserve the magnitude of the force near the interface. This induces an overestimate of the magnitude of the forces further away from the interface where we assume that they are sufficiently weak that even a large relative difference is of little consequence for the dynamics of the simulated system.

Fig. 3a,b illustrates the differences between the three different force profiles and their corresponding potentials. The 'empirical 1' force profile exhibits a rather classical distance/shape dependence. In the region where the attractive term dominates, the force increases continuously as the water molecules approach the outer border of the hydration shell. A particularity of this force profile is that the maximum magnitude of the forces, at around $r = 5$ Å, will actually depend on the position, $R_{\text{layer}}^{\text{dyn}}$ of the repulsive Lennard–Jones wall during the course of the simulation (see Section 2). The 'empirical 2' force profile has a maximum which is at distance of approximately $\sigma_w^T$ from the hydration shell border. This allows the implicit mean forces to decrease linearly and vanish as water molecules reach the interface. Although the magnitude of the maximum in the profile of empirical 2 is

smaller than in empirical 1, it exerts significant effects on more water molecules. Finally, the exact profile has features of both the empirical profiles. It exerts significant forces on the water molecules from a distance of approximately 4 Å from the interface and the forces increase continuously until a maximum is reached at the boundary with the implicit solvent continuum. This latter profile is thus expected to be have the most influence on the dynamics of the hydration shell.

## 4.3. Analysis of simulations of an argon atom in water

We performed MD simulations for an argon atom surrounded only by water molecules (TIP3P and SPC/E) within 10 Å of the argon nucleus. We subjected the water molecules to each of the three different mean force profiles described above. The properties of the water in these simulations are shown in Table 1.

The pair correlation functions for the system of argon solvated in SPC/E water were computed as described in Appendix B and are displayed in Fig. 4a,b. They were also calculated for the system of argon in TIP3P water (data not shown) but we report only the plots for the SPC/E model because it is more structured than the TIP3P model and thus allows better discrimination between the three different force profiles. In all cases, the results demonstrate that applying implicit solvent forces and coupling the system to an external pressure bath has a fundamental effect on the solvent structure in the hydration shell. Similar effects were obtained for the oxygen–hydrogen pair correlation (not shown) as for the oxygen–oxygen pair correlation.

Fig. 4a shows that the oxygen–oxygen pair correlation function of the solvent in the hydration shell is substantially affected by the absence of implicit solvent (dotted lines). The lack of water molecules surrounding the spherical droplet creates a surface tension which tends to compress the system. In this case, the virial of the internal forces acting inside the system and the internal pressure are large and negative (cf. Table 1). There are in fact two competing effects: evapora-
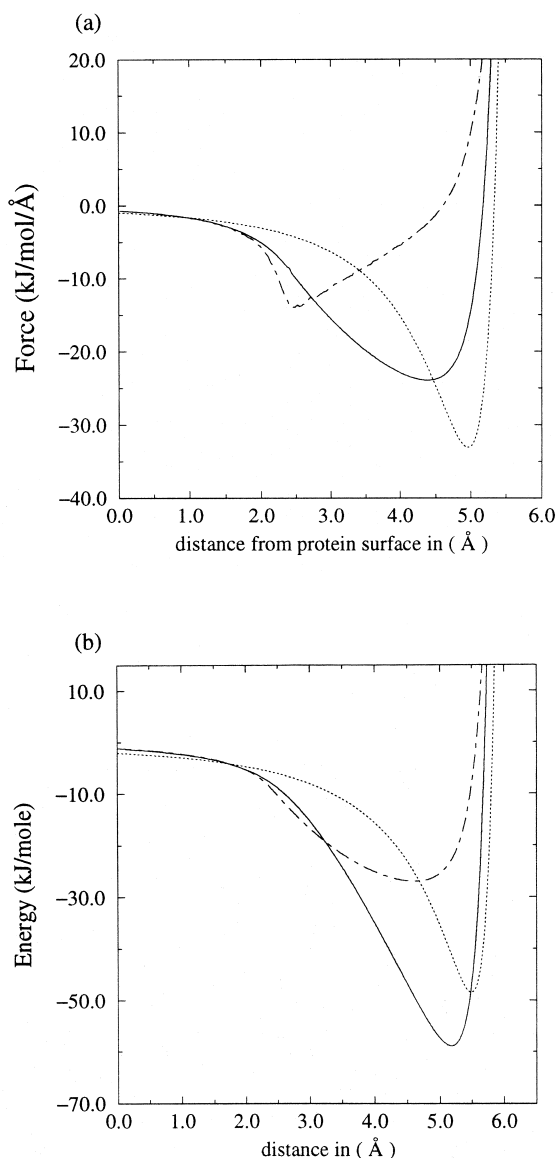
(a)



(b)



Fig. 3. (a) Distance dependence of the three profiles of implicit solvent forces including both the attractive term arising from the dipole–dipole interaction and the Lennard–Jones repulsive term. (b) The corresponding potentials of mean force deduced from the analytical expressions in Eqs. (1),(14),(16),(17). The values of the parameters used are those fitted for the TIP3P water model. The profiles for the exact, 'empirical 1' and 'empirical 2' forces are shown respectively by the solid, dotted and dot-dashed lines.

tion of water molecules and surface tension. In the absence of confinement forces, evaporation of

water molecules occurs during the heating and equilibration phases (see Fig. 4b). However, as the droplet evaporates, the density of the droplet of water becomes inhomogeneous and this is reflected in the oxygen–oxygen pair correlation function (Fig. 4a). It should be remembered here that in the normalisation procedure (cf. Appendix B), it is assumed that the droplet is always spherical with a radius of 10 Å. This is not true because such a small volume of water undergoes large fluctuations during its dynamics. Moreover, the water molecules that evaporated were excluded from the calculation of the pair correlations. Additionally, it should be noted that the periodic box reference simulations were performed without force shifting. If force shifting is applied to the periodic system, the water–water pair correlation functions become less structured. All of this prevents any rigorous quantitative analysis of our pair correlations. However, qualitatively it appears from Fig. 4a that the exact force, (Eq. (14)), and the exact force additionally scaled by one-half provide upper and lower bounds for the pair correlation function calculated from the periodic box of SPC/E water. The two empirical profiles appear to best reproduce the pair correlation function from the periodic box. In addition, the pressure can be controlled solely by a Lennard–Jones confinement term (cf. Table 1) for which the water–water pair correlation matches that obtained for the 'empirical 2' force profile. From these data, we conclude that good pressure coupling can be achieved with all the implicit confinement force profiles. This leaves the question of what the physical effect of the implicit dipole interaction term is. This question can be better addressed by examining Fig. 4b.

The argon–water pair correlations in Fig. 4b show the density of the solvent in concentric incremental spheres around the solute. The simulation with only the repulsive Lennard–Jones force (labelled '3') shows the second worst density level after the simulation with no implicit solvent force at all (dotted line), and the worst density for the second peak at $r = 6.3$ Å. In contrast, the exact force profile (bold) is the one that is best able to restore the correct level of density near the outside of the spherical droplet both around

Table 1
Comparison between the different force profiles for the MD simulations of an argon atom in water performed with the SPC/E and TIP3P water models[a]

| Force profile | Pressure (atm) | $T$ (K) | $-$Virial[b] ($10^{-3}$ kJ/mol) | D* ($10^{-9}$ m$^2$/s) | $\tau$ (ps) | E$^{int}$ (kJ/mol) |
|---|---|---|---|---|---|---|
| Hydration shell simulations | | | | | | |
| Exact | 3.6 | 326 | $-9.0$ | 2.84 | 3.83 | $-51.2$ |
| | *0.9* | *307* | *$-11.5$* | *3.60* | *2.56* | *$-41.7$* |
| One-half exact | $-2.3$ | 315 | 6.0 | 2.26 | 4.49 | $-52.0$ |
| | *1.5* | *313* | *$-4.0$* | *3.85* | *2.49* | *$-41.4$* |
| Empirical 1 | 1.9 | 320 | $-4.8$ | 2.69 | 4.02 | $-51.7$ |
| | *0.9* | *320* | *2.5* | *3.78* | *2.34* | *$-41.1$* |
| Empirical 2 | 0.9 | 311 | $-2.5$ | 2.37 | 5.07 | $-52.4$ |
| | *$-4.2$* | *309* | *11.3* | *3.73* | *2.62* | *$-41.7$* |
| None | $-1263$ | 301 | 3410 | 2.86 | 6.01 | $-51.2$ |
| | *$-1201$* | *303* | *3240* | *5.68* | *2.93* | *$-40.0$* |
| LJ repulsion | 7.8 | 307 | 21.0 | 2.36 | 5.07 | $-52.7$ |
| | *0.2* | *307* | *$-0.5$* | *3.60* | *2.49* | *$-41.7$* |
| Periodic box simulations | | | | | | |
| Plain cutoff | 1.0 | 314 | 3510 | 1.75 | 8.95 | $-58.6$ |
| | *$-8.3$* | *315* | *3570* | *5.13* | *4.28* | *$-45.6$* |
| Force shifting | $-5.6$ | 296 | 4700 | 25.6 | 1.55 | $-58.4$ |
| | *$-3.8$* | *311* | *4700* | *38.0* | *0.91* | *$-46.7$* |

[a]Upper values are for the SPC/E model and lower values in italics are for the TIP3P model.
[b]Virials for simulations in a hydration shell and in a periodic box cannot be compared directly because of the different number of atoms in each system.

$r = 9$ Å and at $r = 6.3$ Å. This is because, in this case, the interface acts as an attractor to the water molecules. Concomitantly, a pronounced depletion in the region around $r = 8$ Å is induced by this attraction. As the magnitude of the implicit force decreases, this depletion reduces. The 'empirical 2' force has the flattest profile. Thus, the density of the solvent in the crucial region up to 3 Å from the interface shows extreme sensitivity to the boundary force profile utilized.

In order to evaluate the effects of the different force profiles on the dynamic properties of the simulated systems, we analysed the orientational correlation of the dipole moments of the water molecules. For the reference periodic system simulated without force shifting, the relaxation time $\tau = 8.95$ ps is in good agreement with results usually obtained with the SPC/E model [41] (see

Table 1). However, the reference simulation using force shifting gives a relaxation time that is six times faster: $\tau = 1.55$ ps. Relative to this value, the effect of performing simulations with a hydration shell is to increase the relaxation time. One explanation is provided by analysis of the dipole relaxation in 5-Å thick shells around the argon atom. This showed that, in hydration shell and periodic box (without force shifting) systems, the water molecules in the 5-Å shell closest to argon relaxed slower than in bulk and those further away [5–10 Å] relaxed slightly faster than in bulk. However, for the SPC/E water model, there is sensitivity to the implicit force profile used, and the shell system simulated with the exact implicit force behaves most similarly to the reference periodic system simulated with force shifting. With the TIP3P model, the dynamic properties are
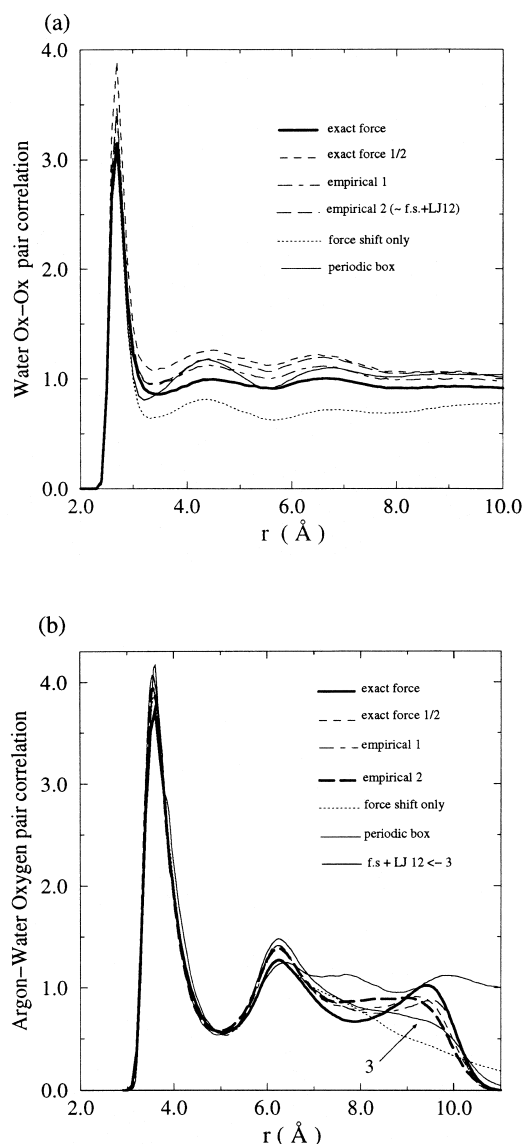
Fig. 4. Radial pair-correlation functions calculated from the trajectories of the test system of an argon atom surrounded by SPC/E water molecules. The pair correlation functions are normalized as described in Appendix B. (a) Water oxygen–oxygen; and (b) Argon–water oxygen pair correlation functions. Key: thin solid line: reference simulation of argon in a periodic box with a plain cutoff (no force shifting). The other lines are for argon with a hydration shell and force shifting: bold solid line: exact force; dashed line: exact force scaled by one-half; dot-dashed line: 'empirical 1' force; bold long-dashed line: 'empirical 2'; and dotted line: no implicit solvent force. In (a), the functions computed with the

much less sensitive to the use of force shifting and the implicit solvent force profile.

## 4.4. Analysis of simulations of ubiquitin in water

The different implicit force profiles were then tested in simulations of a well-characterised, small, globular protein, ubiquitin. It has 76 residues (residues 73–76 form a flexible tail extending into the solvent region), no disulphide bridges and is rather stable. In addition, ubiquitin has been the subject of extensive MDS by other authors [42,43] and these studies provide a baseline for evaluating the validity of our approach.

We performed six simulations of ubiquitin with a hydration shell using the three different force profiles and the SPC/E and TIP3P water models. In addition, we performed two control simulations with the SPC/E water model, one with the system immersed in a rectangular box of water molecules with periodic boundary conditions, and the second of the system with a hydration shell but without implicit solvent forces. All the simulations were carried out for a minimum of 400 ps. For each simulation, properties that can be directly related to experimental observables such as the root mean square deviation (RMSD), the radius of gyration (Rgyr), and the temperature factors (Bfac) are summarized in Table 2. Examples of the behaviour of these quantities are shown in Fig. 5a–d.

### 4.4.1. Reference simulation in periodic box

The RMSD, Rgyr, and B-factors calculated for the periodic box reference simulation match remarkably well the results obtained for an almost identical system simulated [43] with the particle mesh Ewald technique to compute the long-range electrostatic interactions. In particular, the value obtained for the backbone non-hydrogen atom

Lennard–Jones repulsive component of the implicit solvent only almost exactly superimpose on those for 'empirical 2' and are not shown for clarity. In (b), the function computed with the Lennard–Jones repulsive component only is shown by the thin solid line labelled '3'.

Table 2
Summary of all simulations of ubiquitin in water

| Simulation type | $\langle RMSD\rangle$[a] (Å) | Rgyr[b] (Å) | $\langle Bfac\rangle$[c] (Å$^2$) | $\langle Max(\Delta Bfac)\rangle$[d] (%) |
|---|---|---|---|---|
| Crystal structure | − | 11.16 *10.72* | 12.3 | − |
| **Hydration shell simulations:** | | | | |
| Exact (SPC/E) | 1.38 (0.18) *0.91 (0.13)* | 11.15 (0.05) *10.60 (0.06)* | 15.8 | 59.9 |
| 1/2 exact (SPC/E) | 1.30 (0.13) *0.80 (0.08)* | 11.11 (0.04) *10.66 (0.05)* | 12.0 | 77.7 |
| Exact (TIP3P) | 1.41 (0.15) *0.92 (0.10)* | 11.11 (0.05) *10.64 (0.05)* | 17.9 | 92.1 |
| 1/2 exact (TIP3P) | 1.24 (0.12) *0.80 (0.09)* | 11.10 (0.04) *10.62 (0.05)* | 14.6 | 78.4 |
| Empirical 1 (SPC/E) | 1.54 (0.18) *1.03 (0.12)* | 11.16 (0.05) *10.60 (0.07)* | 18.8 | 77.4 |
| Empirical 2 (SPC/E) | 1.29 (0.09) *0.85 (0.06)* | 11.10 (0.04) *10.60 (0.05)* | 14.8 | 97.3 |
| None (SPC/E) | 1.09 (0.06) *0.80 (0.07)* | 11.05 (0.04) *10.57 (0.04)* | 3.2 | 352.9 |
| **Periodic box simulation:** | | | | |
| Plain cutoff (SPC/E) | 1.36 (0.15) *0.79 (0.05)* | 11.25 (0.04) *10.72 (0.05)* | 13.0 | 77.2 |

[a]Average RMSD from the crystallographic coordinates calculated after the first 50 ps for all non-hydrogen atoms (upper value) and for the backbone non-hydrogen atoms (lower value in italics). Standard deviations are given in parentheses.
[b]Average radius of gyration calculated after the first 50 ps and excluding residues 73−76 for all non-hydrogen atoms (upper value) and for the backbone non-hydrogen atoms (lower value in italics). Standard deviations are given in parentheses.
[c]Average temperature factors calculated as the positional fluctuation $4\pi^2/3\langle\Delta r^2\rangle$ over the last 200 ps of each trajectory.
[d]Maximized relative difference between the calculated and the crystallographic B-factors. The percentages are calculated as the averaged sum over all non-hydrogen atoms $i$ of $N^{-1}\sum_{i=1}^{N}|Bf_{X\text{-ray}}(i) - Bf_{MDS}(i)|/\min(Bf_{X\text{-ray}}(i);Bf_{MDS}(i))$.

RMSD of 0.79 Å is fully compatible with the values Fox and Kollman reported of between 0.7 and 0.8 Å for the same simulation time interval. The same high quality of agreement holds for the B-factors. In all other reported simulations of ubiquitin, the B-factors are abnormally high and sometimes overshoot crystallographic values by a factor of two or three. The B-factors we report here are the lowest ones and they are also those that best match the crystallographic ones (see Fig. 5c and Fox and Kollman [43]). One difference between the simulations concerns the radius of gyration of the protein which is the experimental quantity of crucial interest in our work. Fox and Kollman report the radius of gyration for the full protein including the tail residues 73−76. They observed that the tail folds towards a position close to the rest of the protein. We observe the same event not only in our simulation in a periodic box but also in simulations with the hydration shell. In order to alleviate complications due to this phenomemon, we report all values of the radius of gyration excluding the tail residues. The radius of gyration computed for the backbone non-hydrogen atoms fluctuates along the simulation in a periodic box with an average value
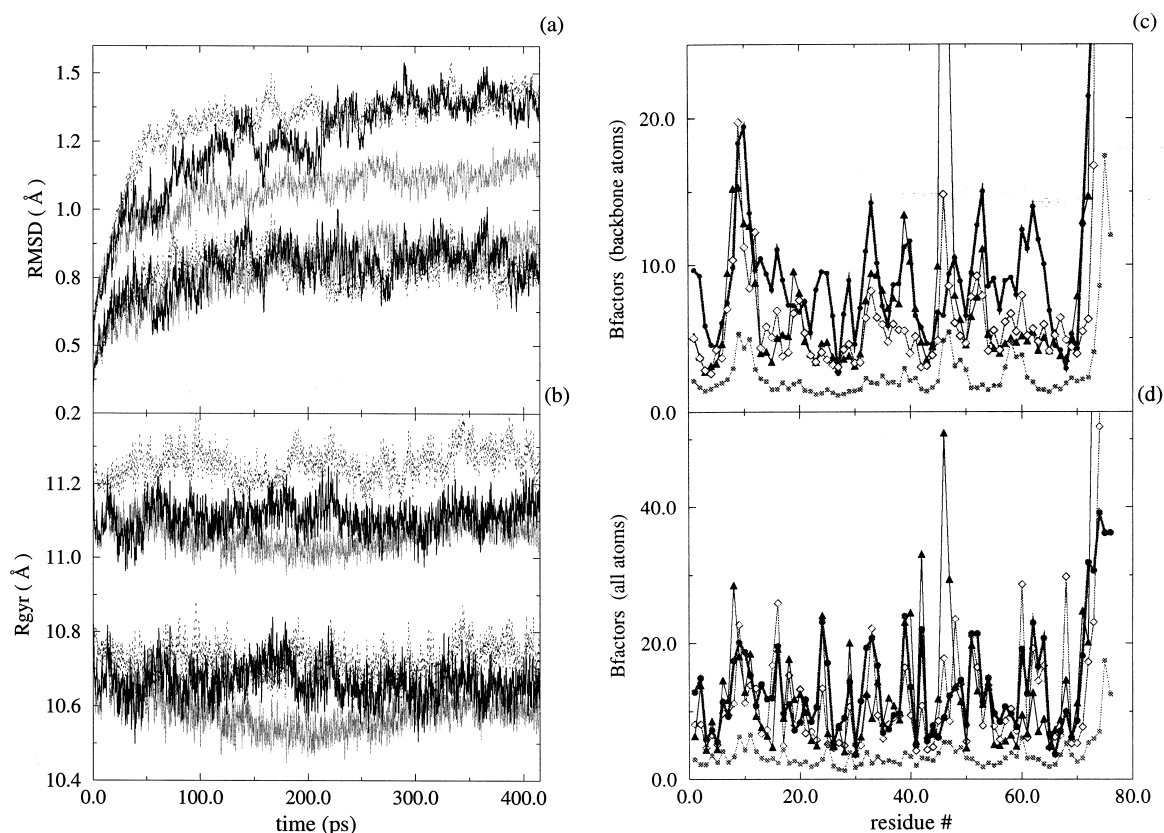
Fig. 5. Properties of three simulations of ubiquitin with the SPC/E water model. The three simulations are for the periodic box (dark grey and diamonds), hydration shell with no implicit force (light grey and squares), and hydration shell with exact force scaled by one-half (black and triangles). (a) Time evolution of the RMSD from the crystallographic coordinates. (b) Time evolution of the radius of gyration $R_{gyr}$, calculated as $\sum_{i=1}^{N} m_i(\vec{r}(i) - \vec{r}_{CM})^2 / \sum_{i=1}^{N} m_i$. In (a) and (b), the upper curves are for all non-hydrogen atoms and the lower curves are for backbone non-hydrogen atoms only. (c,d) Temperature factors $(8/3\pi^2 \langle \Delta r^2 \rangle)$ computed over the last 200 ps of the trajectories for each residue are shown for backbone non-hydrogen atoms (c) and all non-hydrogen atoms (d). Crystallographic B-factors are shown with a bold grey line with filled circles.

exactly equal to the corresponding experimental quantity (cf. Table 2 and Fig. 5b). Thus, the properties of the crystal structure calculated from our reference simulation in full explicit solvent can be viewed as being those attainable by state-of-the-art computer simulation methods.

### 4.4.2. Reference simulation with hydration shell and no implicit solvent force

It is generally recognized that the low RMS deviation of simulations performed in reduced explicit solvent environments (or in the complete absence of solvent) results from an increased

pressure at the system surface [14]. This second reference simulation shows that indeed, when no implicit solvent forces are applied in a hydration shell system, the RMSD for both the backbone and protein side chains are the lowest among all simulations. Concomitantly, the radii of gyration for the backbone atoms and all non-hydrogen atoms, as well as the calculated B-factors, are also the lowest among all simulations. All these measured quantities reflect the fact that the system is compressed and that its dynamics is dramatically hindered by surface tension effects. The radius of gyration is a very sensitive measure of the spatial
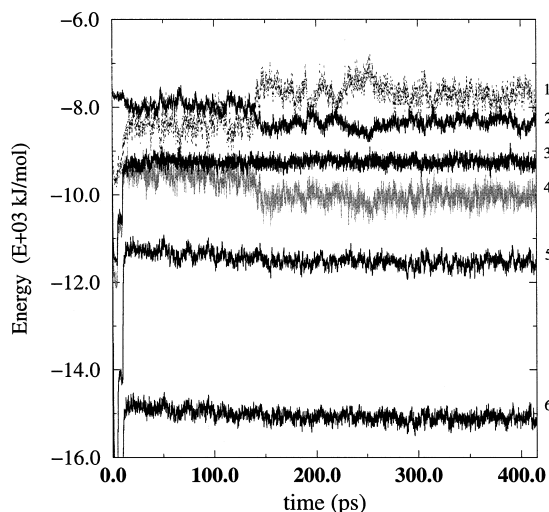
Fig. 6. Evolution of energy components along the simulation of ubiquitin with a 5-Å thick hydration shell of TIP3P water molecules with exact implicit solvent forces: (1) solute–solvent intermolecular Coulombic interaction; (2) solute intramolecular Coulombic interaction; (3) solute total potential energy; (4) solvent intermolecular Coulombic interaction; (5) solvent total energy; and (6) solvent total potential energy. The transition observed in the energy contributions 1, 2 and 4 around 140 ps corresponds to the rearrangement of the flexible C-terminal tail region of ubiquitin. Two ionic pairs not present in the crystal structure are formed; they connect the flexible C-terminal tail Gly76 and Arg74 residues, and Arg72 to Asp39 and contribute to the formation of a network of five charged residues that includes Asp52.

extent of the protein. A small contraction of 0.15 Å induces a fourfold reduction of the thermal flexibility of the protein (cf. Table 2).
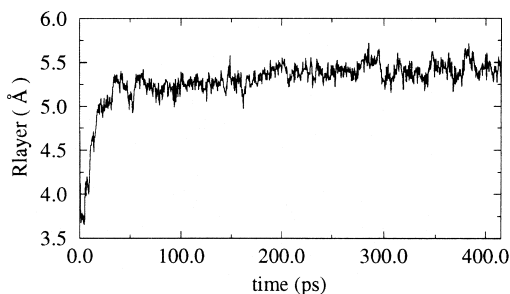


Fig. 7. Time evolution (after energy minimization) of the positional parameter for the repulsive Lennard–Jones wall during the MD trajectory of ubiquitin with the exact implicit solvent force and the TIP3P water model.
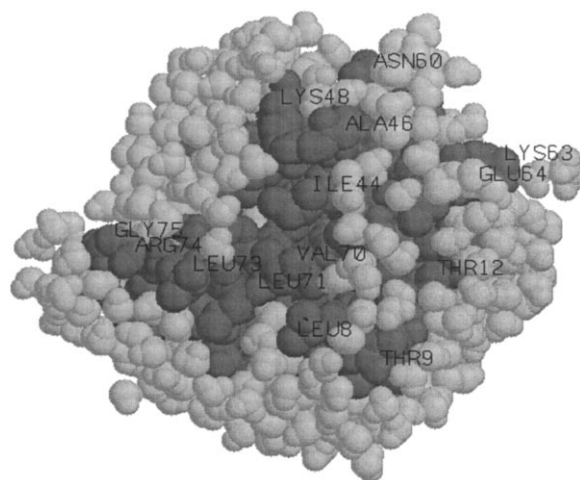


Fig. 8. Space filling model showing the desolvated region on the surface of ubiquitin (dark shading) at the end of a trajectory of 400 ps using the exact implicit solvent force profile with the TIP3P water model. The rest of the protein remains almost completely solvated by explicit water molecules (light shading).

### 4.4.3. Simulations with a hydration shell and implicit solvent forces

The simulations carried out with implicit solvent forces all exhibit a partial improvement of the radius of gyration, and consequently, increased protein flexibility. The RMSD of the side chains becomes comparable to the RMSD obtained in the periodic box simulation indicating that the implicit solvent force counterbalances the surface tension effect and restores normal thermal flexibility to the protein surface. The B-factors also reach a level equivalent to or higher than those in the periodic box reference simulation. Because it is not possible to report the full B-factor plots for all simulations, we calculated a coefficient of maximum relative difference with respect to the crystallographic data (cf. Table 2 footnotes). This coefficient can take the value of zero only when all individual B-factors are identical to the crystallographic ones. For instance, for simulations with the SPC/E water model, the lowest score obtained with the exact implicit solvent force indicates the smallest difference from the crystallographic B-factors in individual atom-to-atom comparison, although the overall average B-factor differs more from the crystallographic
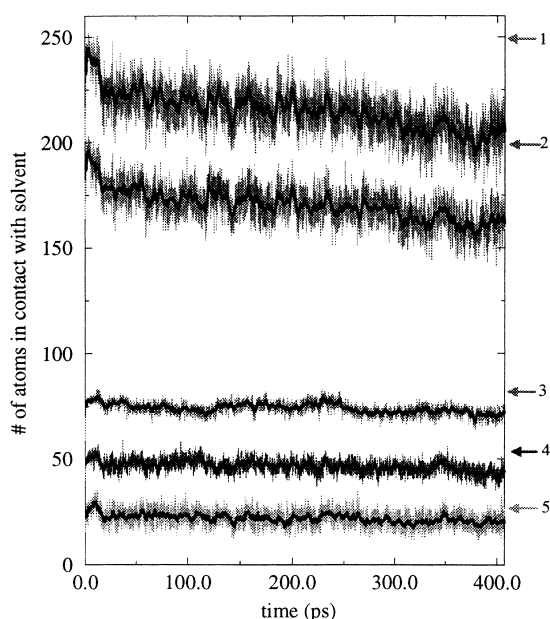
Fig. 9. Evolution of protein desolvation by atom-type for the simulation of ubiquitin along the trajectory performed with the exact force profile scaled by one-half and the SPC/E water model. For each atom type, the number of atoms in contact with the solvent was calculated at each time step of the trajectory using the condition that at least one water molecule is found within a distance criterion of, respectively, 3.2, 3.5 and 4.0 Å for oxygen, nitrogen and carbon atoms. The arrows indicate the corresponding level of solvent contacts for each atom-type at the end of the 400-ps trajectory performed with periodic boundary conditions. (1) all carbon atoms; (2) side chain carbon atoms; (3) oxygen; (4) nitrogen; (5) backbone carbon atoms.

value than in the simulations in full explicit solvent or with the exact implicit solvent force scaled by one-half.

Other properties of both the solute and the solvent phase of the simulated system were monitored to detect possible artifacts in the dynamics caused by the implementation of the implicit solvent forces. In all cases, the average pressure of the system was properly driven to the external reference pressure, and the temperatures for both the solute and the solvent were appropriately maintained around 300 K. The various energy contributions to the total energy terms of the solute and the solvent exhibited small constant fluctuations around constant average values (see

Fig. 6) during the simulations indicating that the systems were at equilibrium.

In all cases, the parameter $R_{\mathrm{layer}}^{\mathrm{dyn}}$ controlling the system pressure, which was initially set to 5 Å, shrank during heating before recovering to an equilibrium value close to 5 Å (see Fig. 7). The initial shrinkage is due to the large negative pressure of the solvent around the protein surface at low temperature. As the system temperature increases during thermalisation, the position of the repulsive wall recovers its original value around 5 Å. There is a systematic small but constant positive drift of $R_{\mathrm{layer}}^{\mathrm{dyn}}$ for all simulations which originates from solvent reordering in the hydration shell. Indeed, in all simulations with a hydration shell, partial desolvation of the protein surface occurs systematically in the vicinity of the same residues or groups of residues (see Fig. 8). As a consequence, the extent of the hydration layer in some other parts of the protein is slightly larger than initially, and the pressure control parameter $R_{\mathrm{layer}}^{\mathrm{dyn}}$ tends to adjust to this newer, inhomogenous distribution of the hydration layer.

To analyse the origin of the desolvation, we plotted the evolution of the protein solvent exposure according to atom-type as a function of the simulation time, see Fig. 9. A pronounced differential desolvation of the carbon atoms of the protein side chains occurs from the beginning of the trajectory, with the greatest desolvation taking place during equilibration. In addition, among the residues involved in this process are those presenting the largest crystallographic and calculated B-factors, namely residues 6–11, 44–49 and 70–76; and also 60–66. As shown in Fig. 6, a major cause of solvent reorganization is the folding of the flexible C-terminal tail onto the protein resulting in a more compact structure of the protein–solvent system.

Some of the calculated protein–solvent and solvent–solvent pair correlation functions around the protein surface are displayed in Figs. 10 and 11. Here again, the effect of the lack of explicit solvent beyond the hydration shell is reflected in a higher density of the first peak in the correlation functions for the different atom types. The implicit solvent forces contribute, by releasing the
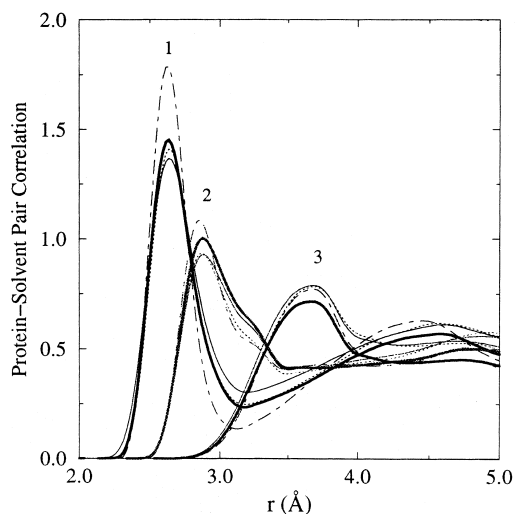
Fig. 10. Protein/solvent–oxygen pair correlation functions calculated for all solvent exposed oxygens (1), nitrogens (2), and side chain carbons (3) for the ubiquitin system for the two reference simulations with periodic boundary conditions (dotted line) and with a hydration shell with no implicit forces (dot-dashed line), and two simulations using implicit solvent forces (exact: bold line; 'empirical 1': plain line) with a hydration shell. For each frame along the trajectory, any oxygen, nitrogen, or side chain carbon atom for which a water molecule is closer than respectively 3.2, 3.5, and 4.0 Å was considered to be solvent exposed and included in the calculation of the pair correlation function. For the system with periodic boundary conditions, the pair correlation functions were calculated for a subset of water molecules within 5 Å of a protein atom.
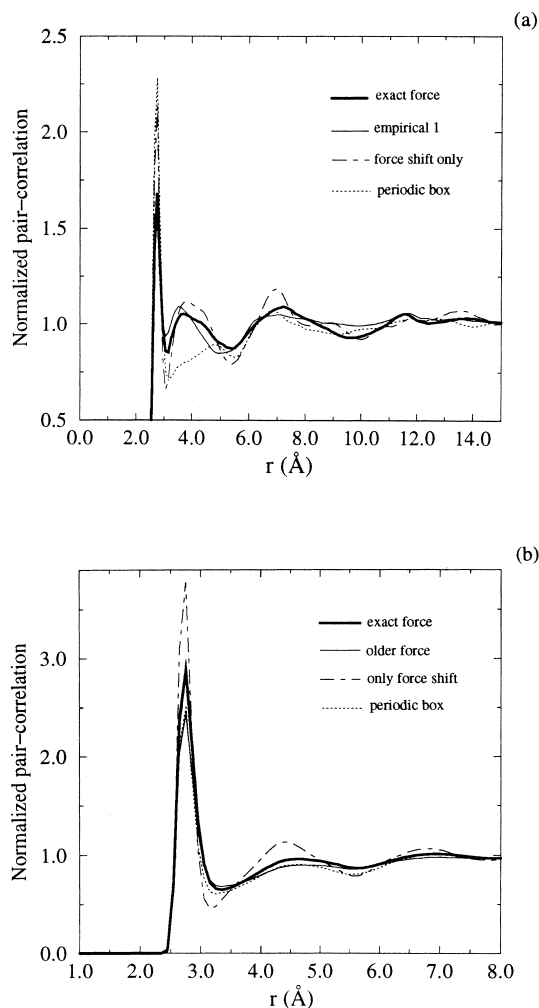


Fig. 11. Water oxygen–oxygen pair correlation functions calculated for the distribution of solvent around the surface of ubiquitin for two different degrees of hydration, $h = 0.27$ (a) and $h = 0.83$ (b), for four different systems: the two reference systems with periodic boundary conditions (dotted line) and with a hydration shell without implicit solvent forces (dot-dashed line), and two others with a hydration shell employing implicit solvent forces with the exact (bold line) profile and the 'empirical 1' (plain line) profile. The pair correlations for $h = 0.27$ are obtained from a subset of water molecules for which the distance to the closest protein atom is within 3 Å. For $h = 0.83$, all water molecules present in the systems were included in the pair correlation calculation, except for the periodic system where again a subset of water within 5 Å of the protein surface was used.

pressure in the primary layer of hydration, to restoring the pair correlation profile observed for the simulation in a periodic box. Using the normalisation procedure described in Appendix B, we extracted the pair correlation functions in Fig. 11a,b. For the reported simulations there are 491 water molecules in the hydration shell. This number corresponds to an hydration degree $h = 0.83$; where $h$ is the mass ratio of added water to dehydrated protein, a definition used in experiments on hydrated protein powders. It is generally observed that enzymatic activity and other dynamic properties of dehydrated proteins are recovered at $h \sim 0.4$ [14,44]. So we examined the solvent structural properties at hydration levels below ($h = 0.27$, Fig. 11a) and above ($h = 0.83$, Fig. 11b) the critical value of $h = 0.4$. The pair correlation at $h = 0.27$ is actually a mock of the

true one because we used the same simulation as for $h = 0.83$, and simply excluded all water
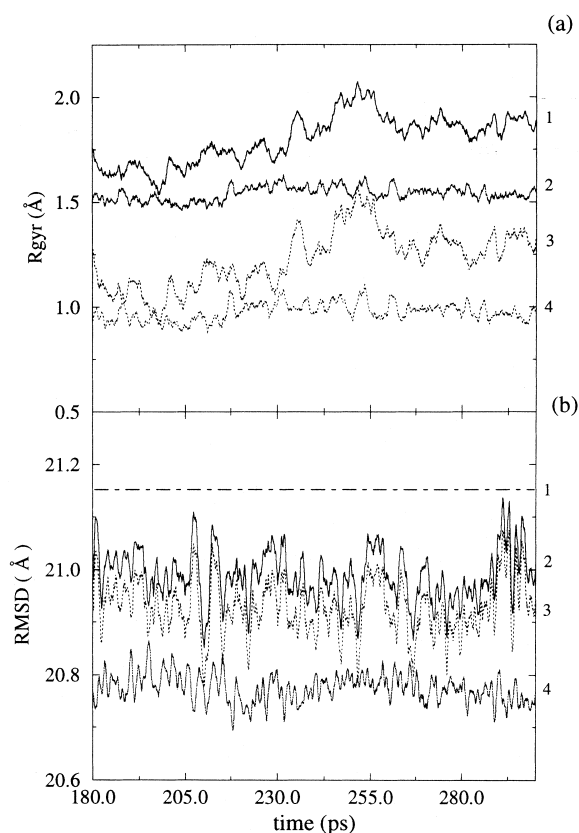
Fig. 12. (a) RMSDs for the last 120 ps of two MD simulations of cytochrome P450cam with a 5-Å thick hydration shell, one with the 'empirical 1' implicit solvent forces (lines 1 and 3) and one without implicit forces (lines 2 and 4). Lines 1 and 2 are for all non-hydrogen atoms whereas lines 3 and 4 are for only the backbone non-hydrogen atoms. (b) the radii of gyration calculated for the same trajectories: lines 2 and 3 are for, respectively, all atoms and backbone atoms only for the trajectory with implicit solvent forces, line 4 is the all-atom radius of gyration of the simulation without implicit solvent forces, and line 1 show the value, at 21.16 Å, of the all-atom radius of gyration of the crystal structure.
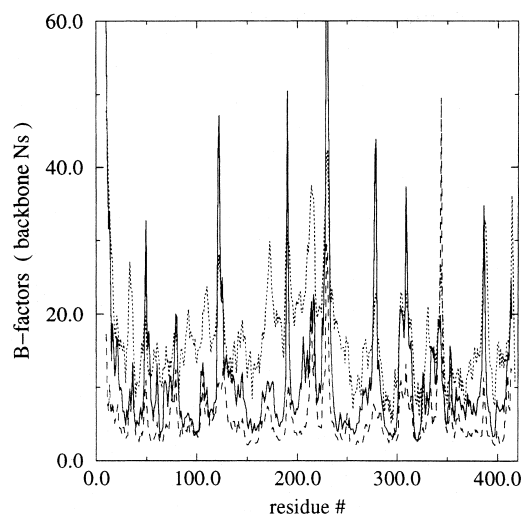


Fig. 13. B-factors of the backbone nitrogens along the sequence of P450cam for simulations with a hydration shell and with the 'empirical 1' implicit solvent force (solid line) or without implicit solvent forces (dashed line). B-factors from the crystal structure are shown by the dotted line.

molecules beyond 3 Å from the protein surface for its computation. In addition, we compared the pair correlations with the equivalent ones calculated in the periodic box system using the same exclusion procedure. We observe that the most closely bound water molecules (for $h = 0.27$) have structural properties differing from bulk and that, in addition, the structure of these water molecules when part of a reduced hydration shell does not match that exhibited for the periodic box system. However, at larger $h$ (Fig. 11b), we observe very

good agreement between the pair correlation from the hydration shell with implicit solvent forces and the pair correlation obtained for a region within 5 Å of the protein surface in the periodic box system. At both levels of $h$, we observe a considerable improvement with respect to the pair correlation from the reference hydration shell system with no implicit solvent force.

### 4.5. Analysis of simulations of cytochrome P450cam in water

We further tested the effects of the implicit solvent forces on cytochrome P450cam, a large protein of more than 400 residues. The simulation of this system in full explicit solvent would require the use of a periodic box with minimum dimensions $95.2 \times 86.9 \times 74.0$ Å$^3$ to ensure that every part of the protein is at least 10 Å from the closest box side. It would contain 17 553 explicit water molecules in addition to the 6489 atoms of the protein. In practice, it is not currently feasible to carry out MDS for systems of such size without using large amounts of CPU time on a parallel supercomputer (such as done in Wlodek et al. [45]). Therefore, we have simulated this system only in a 5-Å thick hydration shell. Simulations of
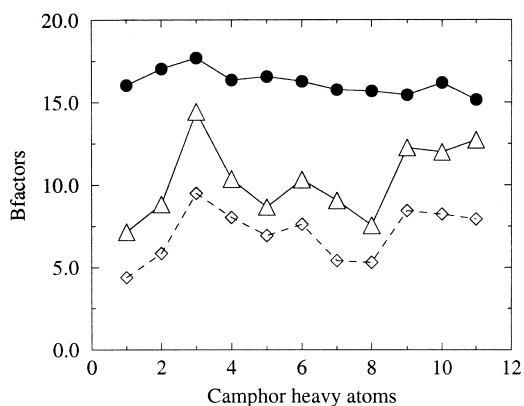
Fig. 14. B-factors for the non-hydrogen atoms of the substrate, camphor, in the buried active site pocket of cytochrome P450cam for simulations with a hydration shell and with the 'empirical 1' implicit solvent force (solid line and triangles) or without implicit solvent forces (dashed line and diamonds). B-factors from the crystal structure are shown by the filled circles.

300 ps were performed with the 'empirical 1' implicit solvent forces, and for reference, without any implicit solvent forces.

The RMSDs from the initial coordinates, the radius of gyration, and the temperature factors for this simulation and the reference simulation of the same duration are reported in Figs. 12–14. The RMSDs for the reference simulation are extremely low, approximately 0.7 Å for the backbone atoms, consistent with the effects of surface tension whereas they reach higher values of 1–1.5 and 1.5–2 Å for the protein backbone and all non-hydrogen atoms, respectively, with implicit solvent forces. These latter values are consistent with those expected in MDS of a large system. Although it does not reach the experimental value, the radius of gyration of P450cam is more than 0.2 Å greater than the value for the reference simulation which is itself 0.4 Å smaller than the radius of gyration of the crystal structure. The shrinkage and recovery of the radius of gyration is more striking for the simulation of P450cam than for the ubiquitin simulation due to the difference in protein size. As a consequence of the partial recovery of its native spatial extent, P450cam exhibits enhanced atomic motions which are exemplified by the B-factors reported in Fig. 13. This figure shows that the calculated B-factors for the simulation using implicit solvent forces are

not only markedly higher than those of the reference simulation but also in very good agreement with the crystallographic distribution of B-factors. Furthermore, we probed the effect of the implicit solvent force on the interior of the protein as P450cam has an active site that is completely buried inside the protein. The plot of the B-factors for its substrate camphor bound in the active site, (cf. Fig. 14), show that the positional fluctuations of the substrate in the active site are also enhanced by the action of the implicit solvent forces on the protein hydration shell.

## 5. Conclusions

The SAPHYR hydration shell model was designed to provide a reasonable description of hydration effects for use in MDS of large proteins that is more computationally feasible than the use of periodic boundary conditions. The rationale for this model is as follows:

(1) The atomic nature of the solvent is important for many features of protein function and therefore models are required in which at least the first hydration layer is modelled explicitly. (2) The water molecules close to the protein surface display structural and dynamic properties differing from bulk solvent (see Markarov et al. [46] and references therein) which are essential for protein flexibility. It is thus reasonable to model the first hydration layer and the bulk solvent differently. (3) The usual approximation of the bulk solvent surrounding the hydration layer as vacuum leads to artifacts such as surface tension effects, escape of water molecules or reduced dynamics of the protein. (4) The most important features of the bulk solvent can be modelled at little computational cost as additional forces on the shell water molecules representing van der Waals and dipole–dipole interactions with the implicit solvent.

We have described three functions to model the permanent dipole interactions between explicit and implicit water. Although these three functions exhibit different dependencies in the crucial region within 3 Å of the interface with the continuum, they appear to function as well as each other for many of the reported quantities. Addition of any of these forces between explicit

and implicit water in hydration shell simulations gives a clear improvement over neglecting them. Clearly, further improvements could be made in the model, e.g. by accounting for surface water polarisation and the frictional and diffusive effects of the bulk solvent. Although differences between the results for the three force profiles are small, we conclude from Table 2 that in practice, when implemented with force shifting it is overall preferential to use the '1/2-exact' force for both the SPC/E and the TIP3P water models.

## 6. Nomenclature

| Bfac | Temperature factor; |
| MDS | Molecular dynamics simulation; |
| P450 | Cytochrome P450 |
| Rgyr | Radius of gyration |
| RMSD | Root mean square deviation |
| $\sigma_w^T$ | Thermal diameter of a water molecule |

## Appendix A

*Implementation of the SAPHYR hydration shell model*

It should be possible to implement the SAPHYR model at very low additional CPU cost in all classical MDS codes. We implemented the SAPHYR model in the ARGOS molecular dynamics code [34], by modifying only two subroutines. They are the solute–solvent forces evaluation subroutine FORSW.f, and the subroutine NEWTON.f, in which the pressure of the system is evaluated.

In FORSW.f, at each time step $t$, the forces from implicit Zone B (cf. Eq. (20)) were added to the solvent force accumulator-array for each solute–solvent interaction pair $(i,j)$ satisfying $i = i_j^{\text{clos}}$ $(t - \Delta t)$. The time step retardation, $-\Delta t$, allows efficient evaluation and updating of $r_j^{\text{clos}}$ which is calculated as:

$$r_j^{\text{clos}}(t - \Delta t) = \min\{r_{ij}(t - \Delta t)\}_i. \qquad (25)$$

To this end, a conditional statement was introduced in the innermost loop of FORSW.f as follows to dynamically reassign the array $\{r_j^{\text{clos}}, i_j^{\text{clos}}\}_j$ for each water molecule $j$, where $i_j^{\text{clos}}$ is the index of the closest protein atom:

```
…
IF(RWI2(IAX).GT.RI2CLOS(IDWRK))THEN
    RI2CLOS(IDWRK) = RWI2(IAX)
    IDSAW(IDWRK) = ISAWRK
ENDIF
…
```

where IAX is the index of the solute–solvent interaction pair $(i,j)$, RWI2 $= r_{ij}^{-2}$, IDWRK $=$ IDWL(IAX) the water molecule index from the solvent indexed array, ISAWRK $=$ ISAL(IAX) the protein atom index from the solute indexed array, and RI2CLOS and IDSAW are respectively the arrays storing $(r_j^{\text{clos}})^{-2}$ and the associated closest protein atom $i_j^{\text{clos}}$. At the initial step $t = 0$ of the MD trajectory, no implicit forces are evaluated and the array $\{r_j^{\text{clos}}, i_j^{\text{clos}}\}_j$ is initialised through the first pass in FORSW.f and updated in subsequent calls of FORSW.f.

In addition, we implemented a simple scheme in FORSW.f to distinguish between water molecules on the protein surface and internal water molecules which should not be subjected to the forces from Zone B. A special array is attached to the list of water molecules which, for each of them, stores the number of solute atomic neighbours found within a chosen input distance. Then, for each water molecule, the additional forces from Zone B are set to zero if the number of solute atomic neigbours exceeds a predefined input number. The identity of water molecules subjected to forces from Zone B is dynamically updated every time step thus allowing the force computation for water molecules to adapt as they move between internal and external locations during a simulation. This scheme was tested in the MDS of P450cam with a water shell. This protein contains crystallographically observed internal water molecules and these were prevented from feeling forces from Zone B by the criterion that they have more than 15 protein atom neighbours within 4.0 Å of their oxygen atom.

## Appendix B

*Computation of solvent–solvent pair correlation functions*

Special volume normalisation procedures were used to calculate solvent–solvent pair correlation

functions in the hydration shell systems. For the simulations of argon in a droplet of water molecules of 10 Å radius, the accessible volume of any water molecule 2 at a distance $r$ from any other water molecule 1 located at a distance $r'$ from the argon nucleus was computed as the volume intercept between the spherical volume elements $\Omega(r) = 4\pi r^2 \delta r (\delta r = 0.1$ Å$)$ and the 10-Å sphere centred respectively at water molecule 1 and the argon nucleus as depicted in Scheme 4a. For the solvent pair correlation in the vicinity of the protein surface, Scheme 4b, the normalisation volume was evaluated assuming the protein surface to be flat. This volume was thus calculated as the intercept between two planes separated by the solvent layer thickness (5 Å) and the spherical volume element $\Omega(r)$ centred at water molecule 1.
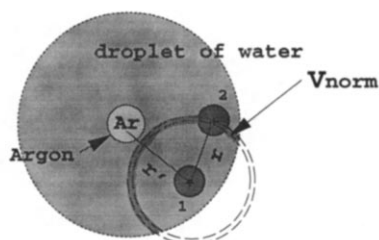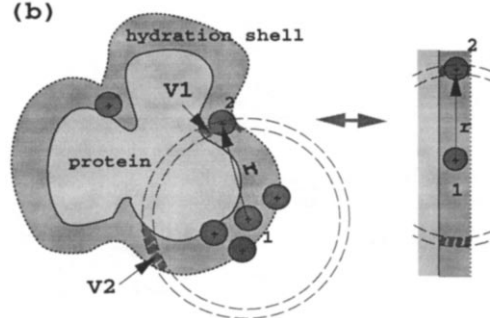


Scheme 4. Diagram to show how the volume of normalization is calculated for computing the water oxygen–oxygen pair correlation functions in (a) a spherical droplet of explicit water (dark shaded circles) centred on an argon atom (light shaded circle), and (b) a hydration shell around a protein surface. See text for details.

## References

[1]  FASEB J., 6 (1992).

[2]  FASEB J., 10 (1996).

[3]  T.L. Poulos, Curr. Opin. Struct. Biol. 5 (1995) 767–774.

[4]  S. Sligar, in: P.R. Ortiz de Montellano (Eds.), Cytochrome P450. Structure, Mechanism, and Biochemistry, Plenum Press, New York, London, 1995.

[5]  V. Helms, E. Deprez, E. Gill, C. Barret, G.H.B. Hoa, R.C. Wade, Biochemistry 35 (1996) 1485–1499.

[6]  V. Helms, R.C. Wade, J. Am. Chem. Soc. 120 (1998) 2710–2713.

[7]  S.K. Lüdemann, O. Carugo, R.C. Wade, J. Mol. Mod. 3 (1997) 369–374.

[8]  T.I. Oprea, G. Hummer, A.E. Garcia, Proc. Natl. Acad. Sci. USA 94 (1997) 2133–2138.

[9]  E. Deprez, N.C. Gerber, C. Di Primo, P. Douzou, S.G. Sligar, G. Hui Bon Hoa, Biochemistry 33 (1994) 14464–14468.

[10] D.L. Harris, G.H. Loew, J. Am. Chem. Soc. 116 (1994) 11671–11674.

[11] M.D. Paulsen, R.L. Ornstein, Proteins 11 (1991) 184–204.

[12] M.D. Paulsen, R.L. Ornstein, J. Comput. Aided Mol. Des. 8 (1994) 389–404.

[13] D. Harris, G. Loew, J. Am. Chem. Soc. 117 (1995) 2738–2746.

[14] P.J. Steinbach, B.R. Brooks, Proc. Natl. Acad. Sci. USA 90 (1993) 9135–9139.

[15] C.L. Brooks III, A. Brünger, M. Karplus, Biopolymers 24 (1985) 843–865.

[16] C.L. Brooks III, M. Karplus, J. Chem. Phys. 79 (1983) 6312–6325.

[17] C.L. Brooks III, M. Karplus, J. Mol. Biol. 208 (1989) 159–181.

[18] G. King, A. Warshel, J. Chem. Phys. 91 (1989) 3647–3661.

[19] D. Beglov, R. Benoît, J. Chem. Phys. 100 (1994) 9050–9063.

[20] Mezei, J. Comput. Chem. (1998).

[21] D. Beglov, B. Roux, Biopolymers 35 (1995) 171–178.

[22] A.H. Juffer, Groningen (1993).

[23] A.H. Juffer, H.J.C. Berendsen, Mol. Phys. 79 (1993) 623–644.

[24] H.J.C. Berendsen, J.R. Grigera, T.P. Straatsma, J. Phys. Chem. 91 (1987) 6269–6271.

[25] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926–935.

[26] P. Lorrain, D. Corson, Electromagnetic Fields and Waves, Freeman, San Francisco, 1970.

[27] S. Vijay-Kumar, C.E. Bugg, W.J. Cook, J. Mol. Biol. 194 (1987) 531–544.

[28] G. Vriend, J. Mol. Graph. 8 (1990) 52–56.

[29] R.W.W. Hooft, C. Sander, G. Vriend, Proteins: Struct. Func. Genet. 26 (1996) 525–529.

[30] P.J. Goodford, J. Med. Chem. 28 (1985) 849–857.

[31] R.C. Wade, P.J. Goodford, J. Med. Chem. 36 (1993) 148–156.

[32] T.L. Poulos, B.C. Finzel, A.J. Howard, J. Mol. Biol. 195 (1987) 687–700.

[33] V. Lounnas, R.C. Wade, Biochemistry 36 (1997) 5402–5417.

[34] T.P. Straatsma, J.A. McCammon, J. Comput. Chem. 11 (1990) 943–951.

[35] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, J. Comput. Chem. 4 (1983) 187–217.

[36] W.D. Cornell, et al., J. Am. Chem. Soc. 117 (1995) 5179–5197.

[37] A. Wallqvist, J. Phys. Chem. 94 (1990) 8021–8024.

[38] P.J. Steinbach, B.R. Brooks, J. Comput. Chem. 15 (1993) 667–683.

[39] J.P. Ryckaert, G. Cicotti, H.J.C. Berendsen, J. Comput. Phys. 23 (1977) 327–341.

[40] H.J.C. Berendsen, J.P.M. Postma, W.V. vanGunsteren, A. DiNola, J.R. Haak, J. Chem. Phys. 81 (1984) 3684–3690.

[41] K.S.B. Bagchi, M.L. Klein, J. Chem. Phys. 107 (1997) 8561–8567.

[42] R. Abseher, S. Lüdemann, H. Schreiber, O. Steinhauser, J. Mol. Biol. 249 (1995) 604–624.

[43] T. Fox, P.A. Kollman, Proteins: Struct. Func. Genet. 25 (1996) 315–334.

[44] J.A. Rupley, G. Careri, Adv. Protein Chem. 41 (1991) 37–172.

[45] S.T. Wlodek, T.W. Clark, L.R. Scott, J.A. McCammon, J. Am. Chem. Soc. 119 (1997) 9513–9522.

[46] V.A. Makarov, M. Feig, B.K. Andrews, B.M. Pettitt, Biophys. J. 75 (1998) 150–158.