

Energy controlled insertion of polar molecules in dense fluids

Gianni De Fabritiis,^{a)} Rafael Delgado-Buscalioni,^{b)} and Peter V. Coveney^{c)}

Centre for Computational Science, Department of Chemistry, University College London, 20 Gordon Street, WC1H 0AJ London, United Kingdom

(Received 29 September 2004; accepted 1 November 2004)

We present a method to search low energy configurations of polar molecules in the complex potential energy surfaces associated with dense fluids. The search is done in the configurational space of the translational and rotational degrees of freedom of the molecule, combining steepest-descent and Newton–Raphson steps which embed information on the average sizes of the potential energy wells obtained from prior inspection of the liquid structure. We perform a molecular dynamics simulation of a liquid water shell which demonstrates that the method enables fast and energy-controlled water molecule insertion in aqueous environments. The algorithm finds low energy configurations of incoming water molecules around three orders of magnitude faster than direct random insertion. This method represents an important step towards dynamic simulations of open systems and it may also prove useful for energy-biased ensemble average calculations of the chemical potential. © 2004 American Institute of Physics. [DOI: 10.1063/1.1835957]

Many processes of physical, chemical, and biological interest involve open systems which exchange matter with their surroundings. Molecular dynamics (MD) and Monte Carlo (MC) simulations of these systems often require a method for molecule insertion and, therefore, a method for searching configurations with prescribed (low) potential energy. Indeed, a randomly placed molecule is likely to overlap with pre-existing atoms, releasing into the system a very high amount of energy.

The most natural setting for these systems is the grand canonical (GC) ensemble. Several methods for GC simulations require the location of energy cavities for insertion (such as cavity-biased methods for GCMC^{1–3}) or careful control of the solvent insertion energy in the case of GCMD.^{4,5} Mass, momentum, and energy transfer are also a key feature of a class of hybrid methods for nonequilibrium simulations which couple an open MD region with an interfacing continuum-fluid-dynamics domain.^{6,7} Open boundaries in such hybrid schemes can avoid finite size effects in small MD simulation boxes,⁸ thereby saving on computational time. These sort of open boundaries could also be used to improve the closed “water shells” widely used to hydrate restricted subdomains⁹ in many MD simulations of biological systems.

Water insertion is also particularly important in protein simulations. For instance, it is possible to study protein unfolding via gradual water insertion in the protein’s cavities.^{10,11} On the other hand, water molecules buried in protein cavities at very low energies are essential for protein structure and function.^{12–14} Indeed, some tools for MD simulations (such as DOWSER¹²) are specialized for water insertion in hydrophilic cavities, leaving empty, however, the larger hydrophobic cavities which frequently contain stable

yet disordered water molecules relevant to protein function.^{13,15}

Several methods for the calculation of ensemble averages require sampling the potential energy released to the system upon insertion of a test molecule.^{1,16–18} Examples include calculation of the chemical potential, hydration energies, and pair distribution functions.¹⁹ The applicability of these methods can be expanded to dense fluids using techniques that bias the sampling towards low energy configurations. Some of these techniques, such as cavity-biased^{20,21} or excluded volume map²² sampling, are, however, hampered by the considerable amount of time needed to find “cavities” where the test molecule could be inserted without overlapping with others. In fact, these cavities are just proxies to search low energy configurations which could better be identified by an energy controlled insertion method.

The algorithms for water insertion proposed in the literature usually involve rather lengthy steps which comprise three separate parts: location of a suitable “cavity,” normally using an expensive grid search with $O(10^6)$ different cells;^{3,4,21} random insertion in the cavity, followed by a large number of energy minimization steps (either of the inserted molecule^{4,12} or of the entire system¹⁰) and, finally, thermostating the whole system over a one to ten picoseconds period to extract the extra energy released upon insertion. In this communication, we present a method to locate low energy configurations of dense liquids that allows insertion of solvent molecules *on-the-fly*: avoiding expensive grid search, nonlocal energy minimization and thermostating steps.

On the potential energy surface, low energies are located inside energy wells whose local minima span a relatively large range of energy values. The main idea of the present method is to reconstruct the energy landscape with a limited number of probes by constraining the search to be *inside* the energy wells. In fact, any excursion outside the explored well implies the loss of all the information accumulated on the current well which is effectively equivalent to a random re-

^{a)}Electronic mail: g.defabritiis@ucl.ac.uk

^{b)}Electronic mail: r.delgado-buscalioni@ucl.ac.uk

^{c)}Electronic mail: p.v.coveney@ucl.ac.uk

start. Efficiency is obtained by minimizing both the number of probes needed to determine if the target energy is found within the well and the number of explored wells per successful insertion. The present minimization algorithm generalizes nontrivially to multiple degrees of freedom the USHER algorithm for insertion of Lennard-Jones atoms.²³ It shares with some other global minimization methods the recipe of applying in turns random moves and local energy minimization.^{24–26} However, it is distinguished from these others in the way the minimization is performed via a combined steepest-descent and Newton–Raphson iterator which is tailored adaptively to the structure of the potential energy landscape being searched.

The method uses local information on the gradient and the average size of the potential wells, which are dependent on the molecule's location and the thermodynamic state, respectively. The input parameters specify the maximum distance ΔR and rotation angle $\Delta\Theta$ that the incoming molecule can jump without exiting the current well together with a measure of the roughness of the potential energy surface ΔE_R . The insertion algorithm starts by selecting a random location for the center of mass of the molecule and placing the atoms at the equilibrium bond and angle positions in a random orientation. The nonbonded potential energy of an incoming molecule is given by

$$U = \frac{1}{2} \sum_{i \neq j} V_{LJ}(r_{ij}) + \frac{1}{2} \sum_{i \neq j} V_C(r_{ij}), \quad (1)$$

where V_{LJ} and V_C are the Lennard-Jones and Coulomb pair potentials, respectively¹ and the index i runs over the atoms of the molecule and j over all other atoms, which remain fixed while inserting. The energy $E = 2U$ released to the system upon insertion is computed and compared with the target energy E_T . The insertion succeeds once the energy difference $\Delta E = E - E_T$ is less than a certain prescribed tolerance set here at 10^{-3} Kcal/mol.

It is likely that for the random starting configuration ΔE will be a large positive value because there is a high chance that the inserted molecule will overlap with others. Then, the force $\mathbf{F} = \sum_i \mathbf{F}_i$ applied to the center of mass \mathbf{r}_{cm} and torque $\boldsymbol{\tau} = \sum_i \mathbf{r}_{cm,i} \times \mathbf{F}_i$ are used to compute the next displacement and rotation. Here, the index i runs over the atoms of the inserted molecule and $\mathbf{r}_{cm,i} = \mathbf{r}_i - \mathbf{r}_{cm}$. The molecule is translated by $\delta r = \min(\Delta E/F, \Delta R)$ where F is the magnitude of the force on the center of mass and ΔR is the maximum displacement. With the reference system fixed to the molecule, we then compute the rotation angle around the torque axis $\delta\theta = \min(\Delta E/\tau, \Delta\Theta)$ and rotate the molecule around the center of mass. The resulting update rule is finally given by

$$\begin{aligned} \mathbf{r}_{cm}^{n+1} &= \mathbf{r}_{cm}^n + \frac{\mathbf{F}^n}{F^n} \delta r, \\ \mathbf{r}_{cm,i}^{n+1} &= \mathcal{R} \cdot \mathbf{r}_{cm,i}^n, \end{aligned} \quad (2)$$

where \mathcal{R} is the rotation matrix around the axis of torque of angle $\delta\theta$. This is equivalent to a first order steepest descent procedure for large energy differences and a second order Newton method for energy close to the target energy.²³ The angular minimization is stopped when the angle $\delta\theta$ is less

than 1° to avoid oscillations due to the coupling of rotational and translational degrees of freedom. If during the iterations ΔE increases by more than ΔE_R then the current attempt is abandoned and a new random configuration is generated. This provides a threshold to control the amount of time spent searching in the well and the number of wells explored.

The insertion algorithm in Eq. (2) does not require a baroque implementation and indeed can be easily included in any molecular dynamics program. The code used here is based on the serial version of a well established parallel molecular dynamics code NAMD with the CHARMM 27 force field,²⁷ but it has been designed to interface easily with any other serial or parallel MD code. The search algorithm applies in general to small polar molecules but given its importance we focus on controlled insertion of water molecules in aqueous environments. We use the TIP3P model for water, widely utilized in biological simulations.²⁸ This water model is based on three interaction sites, bonds (O–H) and angle (H–O–H) being constrained rigidly or, in its flexible version (used here), by a harmonic potential with equilibrium configurations of 0.96 \AA and 104.52° , respectively.

As stated, the restriction on the maximum displacement and rotation has the effect of limiting the search to the current potential well. For water, the maximum displacement can be extracted from the oxygen–hydrogen pair distribution function g_{OH} .²⁸ We found that an optimum value for the maximum displacement $\Delta R = 1 \text{ \AA}$ is half of the first peak in g_{OH} which is around 2 \AA . Exploring the potential energy landscape provides another simple way of obtaining the input parameters. In Fig. 1(a), we show a cross-section of the potential energy surface for a displacement of up to 5 \AA around an equilibrated water molecule in the direction of the axes \mathbf{i} and \mathbf{j} . The unit vectors \mathbf{i} , \mathbf{j} , \mathbf{k} form a reference system fixed rigidly to the water molecule with the axis \mathbf{i} being in the direction of the dipole. As shown in Fig. 1(a) the optimum value of ΔR is approximately the radius of the potential energy well, corroborating information furnished from the pair distribution function. It is more difficult to obtain structural information for the angular degrees of freedom. However, a simple inspection of Fig. 1(b) provides a gross estimate of potential energy wells in the rotational degrees of freedom as being between $90\text{--}100^\circ$ wide; therefore the maximum rotation can be fixed at $\Delta\Theta = 45^\circ$. The value of ΔE_R , which sets the maximum uphill energy jump allowed in one move, is important to reduce the number of unsuccessful wells explored. We found that an optimal value is near $\Delta E_R = 3 \text{ Kcal/mol}$.

It is well known that the local structure of liquid water at equilibrium consists of a hydrogen bond network formed by oxygen and hydrogen atoms from neighboring water molecules. This structure makes it very hard for an incoming water molecule to find low energy configurations by forming hydrogen bonds with pre-existing molecules. However, the insertion algorithm needs only to control the thermodynamics by inputting into the system a specified amount of energy which depends on the ensemble considered. We performed an MD simulation of bulk water using a simple spherical water shell to show that it is possible to insert water molecules on-the-fly while precisely controlling the energy re-

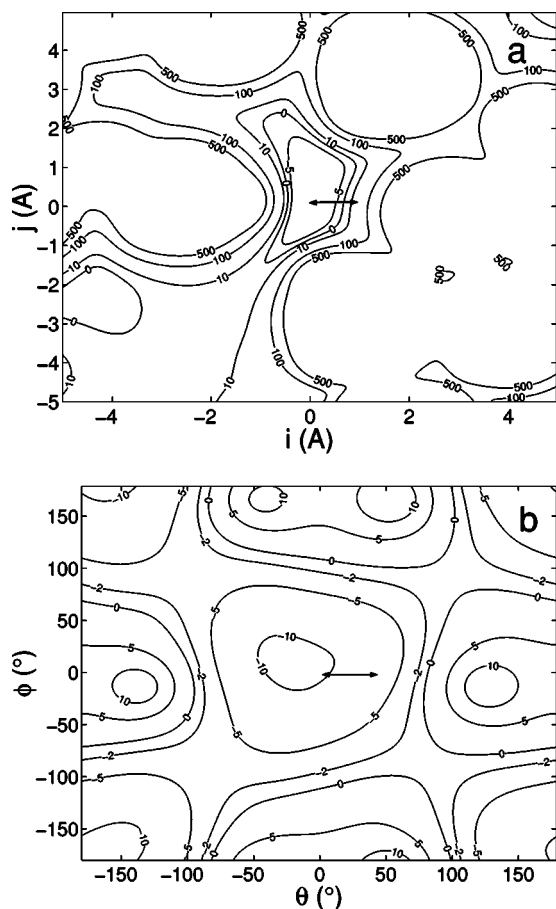


FIG. 1. Contour plot of the potential energy landscape in Kcal/mol: (a) for translation relative to the axes i and j fixed to the water molecule; (b) for a rotation θ about the axis j and ϕ about the axis k for an equilibrated periodic liquid water system at 300 K and density 0.96 g/cm^3 . The maximum translational displacement $\Delta R = 1.0 \text{ \AA}$ and maximum rotational angles $\Delta\theta = \Delta\phi = 45^\circ$ are indicated by double-headed arrows. For visual convenience angles smaller than -90° and larger than 90° in θ are plotted although being redundant.

leased to the system. In a previous work²³ considering Lennard-Jones atoms, it was shown that this procedure ensures thermodynamic consistency after a relaxation time of the order of the collision time. We set up an equilibrated TIP3P bulk water system within a sphere of radius 37.5 \AA at 300 K and a pressure of 1 atm. The simulations were run with a 12 \AA cutoff radius and without corrections to the long ranged electrostatic forces.²⁷ The water molecules in the outer shell of length $d = 12.5 \text{ \AA}$ play the role of a reservoir confined in the sphere by a simple constant radial force field specified by an acceleration \mathbf{g} acting only within the outer shell. The effect of this force is a linear decay of the pressure in the water shell according to the usual formula for the hydrostatic pressure in an incompressible fluid $P_1 = P_0 - \rho g d$, where P_1 is the pressure at the surface of the water sphere and P_0 is the pressure of the bulk that we want to maintain. We impose $P_1 = 0$ by setting $g = P_0 / (\rho d)$.

In the present set up, the flow rate of molecules to the inner shell is controlled by the applied pressure force, while the number of reservoir molecules in the outer shell is fixed at the bulk density. This implies that molecules which, due to fluctuations or sudden pressure waves, move outside the

sphere are removed and reinserted using the insertion method at a random location in the outer shell, with a velocity given by the Maxwell–Boltzmann distribution at 300 K. We note that the present setting can be generalized to avoid finite-size effects due to periodic boundary conditions in a hydrodynamically consistent way.⁶ The total energy of the system can be fixed by setting the amount of energy released upon insertion equal to the energy lost when a molecule moves out through the open boundary.⁶ On average, the exchanged potential energy per molecule is equal to the mean energy per molecule: By inserting at this energy target we kept the total energy under control (without drift) with *no thermostat* at all. In other situations, such as at constant temperature, it is sufficient to release a moderately greater energy, for example equal to the excess chemical potential, which can be thermalized dynamically by the thermostat.

An estimate of the efficiency of this insertion method can be obtained by determining the average number of energy evaluations, including failed well searches, needed to insert a single water molecule at the specified energy. Each iteration of the insertion algorithm corresponds to one energy evaluation on the solvent molecule, which is a three atom-force calculation for TIP3P water. In particular, it takes an average of 206 iterations, exploring 34 wells, to insert at the reference energy of the mean energy per molecule (-11 Kcal/mol), and 36 iterations (only 6 wells) at the energy of the excess chemical potential (-5.8 Kcal/mol , calculated using the Bennett method¹⁶). We note that the computational cost required by the insertion method in a typical MD simulation is quite small. For instance, in the simulation of the open water shell mentioned above, incoming water molecules were inserted at a target energy of $E_T = -11 \text{ Kcal/mol}$ within a volume of 155.4 nm^3 at a rate of 141 per picosecond. The amount of CPU time devoted to insertion was only 3% of the grand total of the simulation.

Interestingly, the mean number of iterations to explore a well which leads to the correct target energy is only around 12, independent of the target energy. The method may be improved further by reducing the total number of searched wells but it is already optimal in the sense that the number of iterations to explore a single well does not depend on the target energy. Future applications may require searching many more degrees of freedom, e.g., conformational searches, for which it is impractical to fix each maximum displacement *a priori*. In this case, it would be useful to set up an adaptive rule to infer the input parameters from the efficacy of the search itself.

It is useful to compare our insertion algorithm with a direct random insertion. To this end, the probability distribution $f(E)$ of releasing a total energy E upon random insertion was estimated by computing a histogram from 10^7 random insertion trials. The number of trials required to obtain an energy smaller than E is given by the reciprocal of the cumulative distribution $1/F(E)$ where $F(E) = \int_{-\infty}^E f(E') dE'$. This number is compared with the number of iterations (energy evaluations) required by the insertion algorithm in Fig. (2). The insertion algorithm is around three orders of magnitude faster than a random insertion for energies lower than the chemical potential and so may provide an

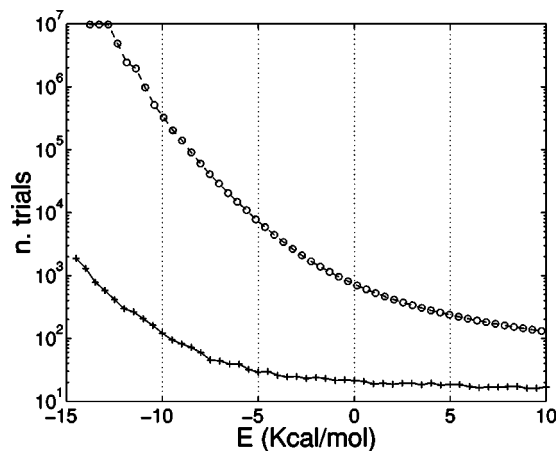


FIG. 2. Number of energy evaluations per molecule required to insert a water molecule while releasing an energy less than E (Kcal/mol) to the system. The proposed insertion algorithm (crosses) is around three orders of magnitude faster than random insertion (circles) at low energies. The histogram for random insertion is computed from 10^7 trials.

efficient alternative to biased methods, such as cavity-biased sampling,^{20,21} to reconstruct the probability distribution $f(E)$. Indeed, the present algorithm enables one to identify the important low energy regions very accurately where an un-biased sampling can be performed. This appealing approach enables fast computation of the chemical potential from the probability distribution $f(E)$ at low energies.²⁹

In summary, we have reported a new method for the insertion of polar molecules in dense fluids by a generalization of the USHER protocol.²³ The energy minimization is applied concurrently to all degrees of freedom (translational and rotational for water) and is independent of the specific potential used. Indeed, the method is even more general. It may be applied to other problems related to conformational searches and minima of potential energy surfaces with many more degrees freedom. Given its importance for computational biology, we focused on water and demonstrated that it is possible to efficiently insert water molecules in aqueous environments while controlling the thermodynamic state. This task is commonly considered to be very time consuming, but we are able to achieve it at negligible computational cost thanks to a very efficient configurational search algorithm. The present algorithm is an essential tool for perform-

ing hybrid MD-continuum simulations^{6,8} of biological interest. Indeed, it represents an important step towards a general method for performing MD simulations of open systems, for which a dynamic calculation of the chemical potential^{18,29} could be used to control the insertion rate so as to maintain constant the solvent chemical potential.

This research was supported by the EPSRC Integrative Biology project GR/S72023 and by the EPSRC RealityGrid project GR/R67699.

- ¹M. Allen and D. Tildesley, *Computer Simulations of Liquids* (Oxford University Press, London, 1987).
- ²D. Adams, *Mol. Phys.* **29**, 307 (1975).
- ³M. Mezei, *Mol. Phys.* **61**, 565 (1987).
- ⁴T. C. Jie Ji and B. M. Pettitt, *J. Chem. Phys.* **96**, 1333 (1992).
- ⁵B. G. C. Lynch, *Chem. Phys.* **258**, 405 (2000).
- ⁶R. Delgado-Buscalioni and P. V. Coveney, *Phys. Rev. E* **67**, 046704 (2003).
- ⁷E. G. Flekkøy, G. Wagner, and J. Feder, *Europhys. Lett.* **52**, 271 (2000).
- ⁸S. Barsky, R. Delgado-Buscalioni, and P. V. Coveney, *J. Chem. Phys.* **121**, 2403 (2004).
- ⁹G. King and A. Warshel, *J. Chem. Phys.* **91**, 3647 (1989).
- ¹⁰J. M. Goodfellow, M. Knaggs, M. A. Williams, and J. M. Thornton, *Faraday Discuss.* **103**, 339 (1996).
- ¹¹M. A. Williams, J. M. Thornton, and J. M. Goodfellow, *Protein Eng.* **10**, 895 (1997).
- ¹²L. Zhang and J. Hermans, *Proteins: Struct., Funct., Genet.* **24**, 433 (1996).
- ¹³I. Hofacker and K. Schulten, *Proteins: Struct., Funct., Genet.* **30**, 100 (1998).
- ¹⁴M. Jensen, E. Tajkorsheid, and K. Schulten, *Biophys. J.* **85**, 2884 (2003).
- ¹⁵S. Cai, S. Stevens, A. Budor, and E. Zuiderweg, *Biochemistry* **42**, 9 (2003).
- ¹⁶C. H. Bennett, *J. Comput. Phys.* **22**, 245 (1976).
- ¹⁷K. S. Shing and K. E. Gubbins, *Mol. Phys.* **46**, 1109 (1982).
- ¹⁸N. Lu, J. K. Singh, and D. A. Kofke, *J. Chem. Phys.* **118**, 2977 (2003).
- ¹⁹B. Guillot, Y. Guissani, and S. Bratos, *J. Chem. Phys.* **95**, 3643 (1991).
- ²⁰P. Jedlovsky and M. Mezei, *J. Am. Chem. Soc.* **122**, 5125 (2000).
- ²¹A. Pohorille and M. A. Wilson, *J. Chem. Phys.* **104**, 3760 (1996).
- ²²G. L. Deitrick, L. Scriven, and H. Davis, *J. Chem. Phys.* **90**, 2370 (1989).
- ²³R. Delgado-Buscalioni and P. V. Coveney, *J. Chem. Phys.* **119**, 978 (2003).
- ²⁴M. Levitt and A. Warshel, *Nature (London)* **253**, 694 (1975).
- ²⁵Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6611 (1987).
- ²⁶M. Saunders, *J. Am. Chem. Soc.* **109**, 3150 (1987).
- ²⁷L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten, *J. Comput. Phys.* **151**, 283 (1999).
- ²⁸W. L. Jorgensen, J. Chandrasekhar, and D. Madura, *J. Chem. Phys.* **79**, 926 (1983).
- ²⁹R. Delgado-Buscalioni, G. De Fabritiis, and P. V. Coveney (preprint).