# Solvent diffusion outside macromolecular surfaces

Erik Lindahl and Olle Edholm
*Theoretical Physics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden*
(Received 18 July 1997)

The effect of the inhomogeneous environment upon solvent molecules close to a macromolecular surface is evaluated from a molecular-dynamics simulation of a protein, myoglobin, in water solution. The simulation is analyzed in terms of a mean-field potential from the protein upon the water molecules and spatially varying translational diffusion coefficients for solvent molecules in directions parallel and perpendicular to the protein surface. The diffusion coefficients can be obtained from the slope of the average-square displacements vs time, as well as from the integral of the velocity autocorrelation functions. It is shown that the former procedure gives a lot of ambiguities due to the variation of the slope of the curve with time. The latter, however, after analytic correction for the contribution from algebraic long-time tails, furnish a much more reliable alternative. [S1063-651X(98)01401-9]

PACS number(s): 82.20.Wt, 61.20.Ja

## I. INTRODUCTION

The mobility of water outside macromolecules such as proteins is difficult to assess through experimental methods. Nevertheless, there seems to be agreement about a reduced mobility close to the protein. Polnaszek and Bryant [1] came up with a factor 5–10, as did Kimmich *et al.* [2] with a different method. Halle and Piculell [3] reported a factor of about 100 in the radial and about 10 in the lateral direction to the protein surface. Computer simulations offer a simple method to resolve such variations of the diffusion coefficient $D$ as a function of the distance to the surface. This has traditionally been done by fitting the square displacement of solvent molecules to a linear expression in time [4–7],

$$\langle(\mathbf{r}(t)-\mathbf{r}(0))^2\rangle = 6Dt, \quad t\to\infty. \tag{1.1}$$

These studies all end up with a reduced translational diffusion coefficient close to the protein, but one reduced by a factor 2–4 rather than 10–100. Ahlström *et al.* [8] determined the variation of the diffusion coefficient by fitting the Green's function of the diffusion equation in a space bounded on the inner side by a reflecting sphere, with results similar to those of [4–7].

Recently, it has been observed that the fitting by Eq. (1.1) has its difficulties since there are systematic deviations from an expression of this type [9,10]. This was suggested to be due to a fractal dimensionality of the protein surface. We want to show that there are several simpler reasons why a result like that in Eq. (1.1) is not valid. In addition, the observed nonlinearity of a plot of the mean-square displacement (MSD) against time makes it difficult to evaluate the diffusion coefficient from Eq. (1.1). We therefore propose the use of an alternative method to define the diffusion coefficient from the velocity autocorrelation function (VAC) using the Kubo formula [11]

$$D = \frac{1}{3}\int_0^\infty \langle \mathbf{v}(t)\cdot\mathbf{v}(0)\rangle dt. \tag{1.2}$$

This is formally equivalent to Eq. (1.1) and follows from pure mathematics and the fact that $\mathbf{v}$ is the time derivative of $\mathbf{r}$. For a homogeneous system in an infinite space, it is just a matter of taste and maybe statistical accuracy which method to chose. In a system with boundaries, spatial variation of the diffusion coefficient, or deterministic forces, however, this is not so.

First, Eq. (1.1) is obtained from the solution of the diffusion equation in an infinite space. The corresponding formula for the process outside a complicated (maybe fractal) surface looks differently. In Sec. IV we show that the analytic solution to the diffusion equation outside an infinite planar reflecting surface gives MSDs deviating from Eq. (1.1) in a way similar to the MSDs in our simulations and those of [9,10].

Second, the diffusion equation (1.1) and the analytic solution of the diffusion equation outside a planar surface are strictly valid only for very long times. This is usually not a problem in a homogeneous system since one may then evaluate the diffusion coefficient from the slope of the square displacement against time at very long times. For a system with a surface, however, molecules once close to the surface will after long times on the average be far away from the surface. Going to long times will thus yield only the bulk diffusion coefficient and a MSD linear in time. One therefore has to make a compromise between the necessity to go into times long enough to avoid nondiffusive short-time effects and the desire to resolve spatial variations of the diffusion coefficient. Such a compromise can be found if there exist times

$$t \ll \frac{D}{|\nabla D|^2} \tag{1.3}$$

that are still large enough for the nondiffusive short-time effects to die out. For such times, Eq. (1.1) or the appropriate alternative for diffusion outside some surface can be used.

Third, even if the first two problems can be resolved, there might be nondiffusive effects also at long times since the macromolecule may induce a mean potential acting on the solvent molecules. Usually, one would expect this deter-

 

ministic force to be attractive. In an analysis using Eq. (1.1) or an equivalent relation, such a mean attraction would show up as a slower diffusion close to the macromolecular surface. One could then adopt a more general equation such as the Smoluchowski or Fokker-Planck equation [12] to simultaneously try to evaluate a spatially varying diffusion coefficient and a spatially varying mean force. This, however, is far from trivial.

To solve these problems, one needs to be careful when defining the diffusion coefficient. We assume self-diffusion in an infinite homogeneous space to be characterized by a MSD at long times proportional to time with the proportionality constant being six times the diffusion coefficient. This self-diffusion coefficient is a property of the fluid that may be defined locally also in a finite nonhomogeneous space, although in this case it cannot be derived from the MSD at long times. It can, however, be obtained from local short-time data through the Kubo formula (1.2) since this involves much shorter time scales. The result then defines the slope the MSD would have at very long times in an infinite space for a fluid with the same global properties as ours has locally.

Equation (1.2) is valid as long as the diffusion coefficient and mean potential do not vary appreciably over distances that a solvent molecule travels in the time it takes for the VAC to decay to zero. For Brownian motion, this happens on a time scale of the inverse damping $\beta^{-1}$. We get an upper estimate of the distance by multiplying with the free flight velocity $\sqrt{3k_BT/m}$:

$$\frac{1}{\beta}\sqrt{\frac{3k_BT}{m}}=\frac{k_BT}{m\beta}\sqrt{\frac{3m}{k_BT}}=D\sqrt{\frac{3m}{k_BT}}. \tag{1.4}$$

For water, this evaluates to less than 0.01 nm.

We have here applied both Eqs. (1.1) and (1.2) to an analysis of a simulation of myoglobin in water solution. In addition to the diffusion coefficients, we have also found and evaluated a mean potential by which the first shell of water molecules is bound to the protein surface with a binding energy of 2–3 kJ/mol.

## II. MOLECULAR-DYNAMICS SIMULATION

Myoglobin was simulated in a periodic box containing 5763 water molecules. As the initial structure we used the entry 1MBA [13] from the Brookhaven Protein Data Bank. For the simulations a standard molecular-dynamics program GROMOS [14] was used. All potential parameters were the standard ones of GROMOS and the water model was the simple-point-charge (SPC) one [15]. United atoms were used for the aliphatic hydrocarbons, while polar hydrogens were represented explicitly. All bond lengths were kept fixed using the algorithm SHAKE [16]. The step length for the integration of the equations of motion was chosen to 0.002 ps, and a cutoff at 1.0 nm was used. The system was initially equilibrated for 200 ps. Pressure scaling was used during part of the equilibration to obtain a volume of the system corresponding to the volume at atmospheric pressure. Temperature scaling [17] to a heat bath with temperature 300 K using a time constant of 0.1 ps was used during the entire simulation. Production runs to obtain MSDs were typically 200–400 ps, storing coordinates every 0.2 ps. To calculate VACs shorter runs (20–40 ps) were used, but data were stored more often (typically every 0.01 ps) due to the fast decay of the VAC.

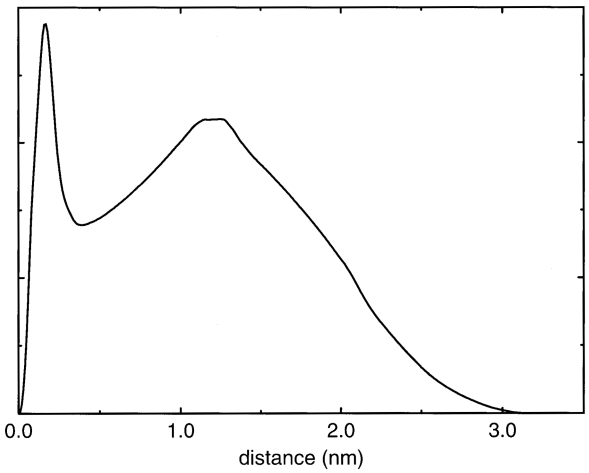


FIG. 1. Number of water molecules vs closest distance to a protein atom (relative scale).

## III. RESULTS

### A. The mean potential

A count of the average number of water molecules at different distances from the protein gives a result like that in Fig. 1. The drop at big distances is due to the finite size of the periodic box, which also means every water molecule has a periodic copy of the protein within 3 nm. This result was converted to a mean potential $U_{\text{mean}}(r)$ using the Boltzmann distribution:

$$N(r)dr \propto A(r)e^{-U_{\text{mean}}(r)/k_BT}dr \tag{3.1}$$

or

$$U_{\text{mean}}(r)=C-k_BT\ln[N(r)/A(r)]. \tag{3.2}$$

Here $N(r)dr$ is the number of water molecules in a shell of thickness $dr$ at distance $r$ from the protein and $A(r)dr$ is the volume of that shell in space. The volume was calculated by counting the number of points on a fine grid covering the entire periodic box having different distances to the closest protein atom. The constant $C$ was chosen to give the mean potential zero at long distances from the protein. In this way the solid curve in Fig. 2 was generated. This shows a minimum with a depth of about 2 kJ/mol at the distance 0.33 nm, a weak local maximum at 0.45 nm, and a leveling off to the constant value further out. There are about 700 water molecules inside this maximum. The protein surface area can be calculated using, e.g., the method and program of Lee and Richards [18] and is then found to be 75 $\text{nm}^2$. Since a water molecule typically covers an area of about 0.1 $\text{nm}^2$, this means that essentially a whole first shell of water lies in the first minimum of the mean potential.

As a control, the mean potential was evaluated independently by averaging the mean force perpendicular to the protein on the water molecules as a function of distance from the protein. The result was subsequently integrated to give

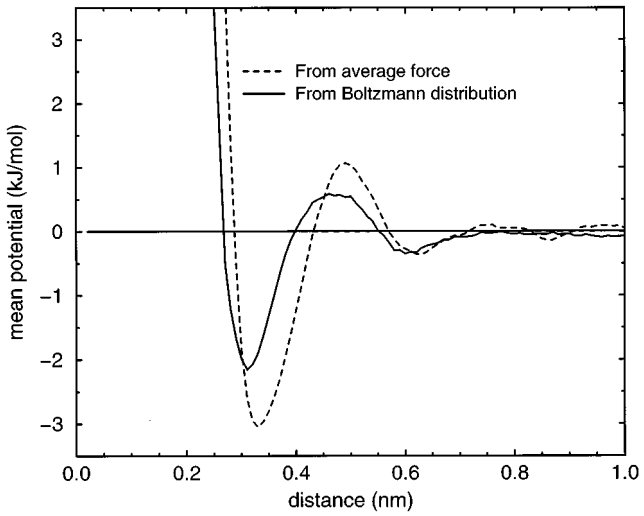


FIG. 2. Mean potential vs distance from protein. The solid line is from the Boltzmann distribution and the dotted line is from integrated average forces.

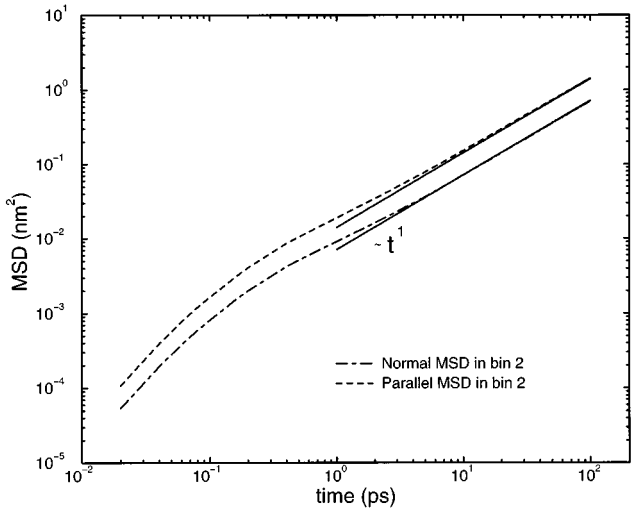


FIG. 3. Mean square displacement vs time in a log-log plot. The solid lines represent a linear increase with time.

the dotted curve in Fig. 2. This curve is similar to the solid one obtained from Eq. (3.2), but the minimum is deeper (3 kJ/mol) and positioned slightly further out from the protein surface.

From this we conclude that the first layer of water molecules outside a protein surface will be bound to the protein by a weak mean potential (a couple of kJ/mol). Since the protein is an atomic surface enabling weaker or stronger bonding at different places this will not be uniform; calculations of the residence times for waters close to a protein [7] show a variation of a factor 20. For a Boltzmann factor, the corresponding difference in energy would be $k_BT\ln 20 \approx 7.5$. Thus the binding energy probably varies between zero and the order of 10 kJ/mol, with an average of 2–3 kJ/mol.

### B. Diffusion coefficients from mean-square displacements

For each water molecule, we approximate the normal to the protein surface by the direction from the closest protein atom to the water's center of mass in each step. The change in distance along this direction was taken as the normal displacement $\Delta\mathbf{r}_\perp$ of the water molecule. The square displacement parallel to the surface $\Delta\mathbf{r}_\parallel^2$ follows from the total square displacement and $\Delta\mathbf{r}^2 = \Delta\mathbf{r}_\perp^2 + \Delta\mathbf{r}_\parallel^2$.

Waters were divided into five bins based on their average distance to the surface during the simulation. The bin limits were 0.45, 0.9, 1.2, and 1.5 nm.

For each such bin, square displacements were averaged over both molecules and the time origin [$t=0$ in Eq. (1.1)]. The resulting MSDs are plotted vs time in Fig. 3 for bin 2, with an appearance similar to the other bins. The curves show a free flight behavior (i.e., the molecules do not ''feel'' the presence of their neighbors) up to approximately 0.5 ps. For intermediate times (0.5–10 ps) the increase is less than linear for both curves, although for long times it seems to tend to linearity. Since the normal motion has one degree of freedom and the parallel two, the *bulk* MSD at large times should be expected to increase as $2Dt$ and $4Dt$, respectively. However, this is not the case for the time period studied in our simulations. Fitting the MSDs to a power law $Ct^\alpha$ yields exponents $\alpha$ in the range 0.85–0.96, i.e., a less-than-linear increase, as observed by [9,10]. We also calculated standard deviations of the fits by dividing the data into ten subsets and repeating the process on each of these. The standard deviations of the results were then scaled to a sample 10 times larger.

The exponent being less than unity means that values of $D$ obtained from a linear fit will depend on the lower boundary of the fitting region, as can be seen in Fig. 4, albeit with large standard deviations. This is a severe problem with the MSD method since it leads to systematic errors in the value of the diffusion coefficient, errors that will therefore not show up in the standard deviations obtained from subsets.

The normal and parallel diffusion coefficients obtained by fits from 20 ps to 100 ps are plotted as functions of distance to the protein surface in Fig. 5. For comparison, the values from the VAC can be found in the same figure.

### C. Diffusion coefficients from velocity autocorrelation functions

The projection of a water molecule's velocity on the previously mentioned direction normal to the protein surface is

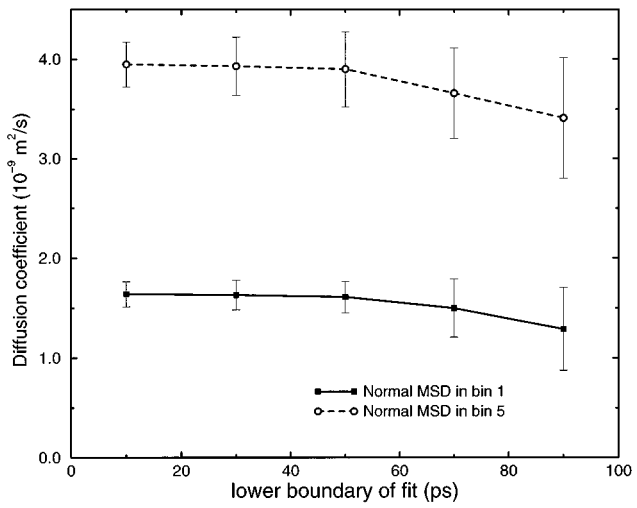


FIG. 4. Diffusion coefficient from the mean-square displacement as a function of the time window used for the fit. The upper boundary was 100 ps.

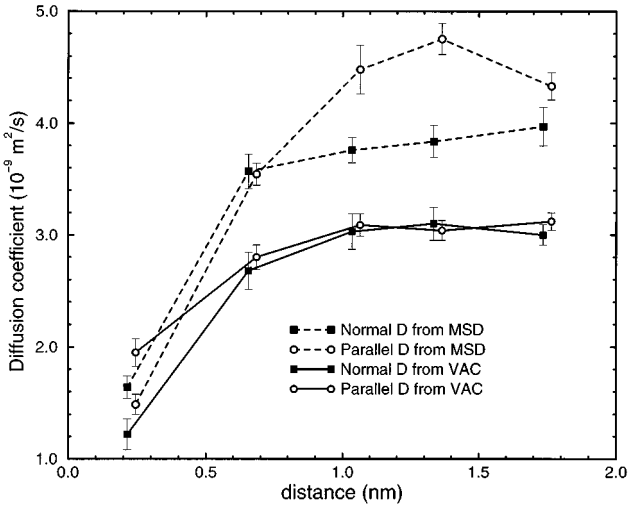


FIG. 5. Diffusion coefficients perpendicular and parallel to the surface calculated with the two methods vs distance from the protein.

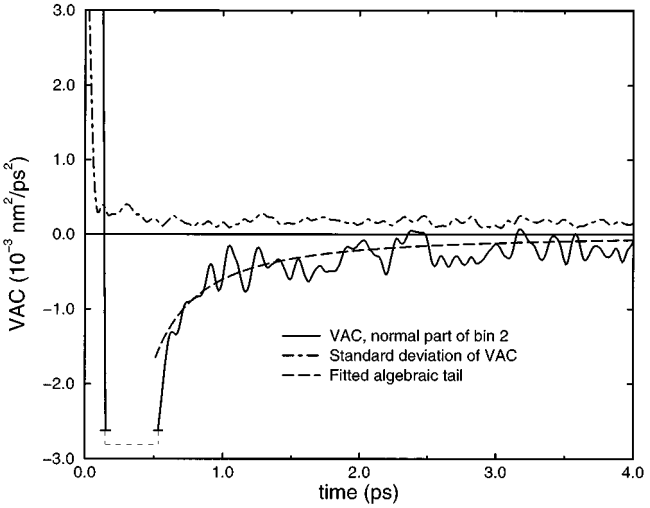


FIG. 7. Algebraic tail fitted to an enlarged part of the velocity autocorrelation function.

taken as the normal velocity. The difference between the total velocity and this is used as the parallel velocity.

The autocorrelation $\rho(t)$ for each of these vectors was calculated according to

$$\rho(t)=\langle \mathbf{v}(t)\cdot\mathbf{v}(0)\rangle, \tag{3.3}$$

where the average is over time origin ($t=0$) and molecules. The water molecules were placed into bins based on their distance to the surface at the time origin, instead of their average distance. This is a very good approximation since a molecule does not travel far before the VAC has decayed to zero and it helps us avoid averaging out small motions, which is a risk in the MSD case. Figure 6 shows the resulting VAC for the normal velocity in bin 2. Since the motion is not perfectly Brownian, the VAC will have an algebraic long-time tail, which we need to include in the integral. This causes a problem since the tail is of the same order as the noise after a few picosecond. We solve it by calculating this contribution analytically. Following Alder and Wainwright [19], we assume a time dependence $Ct^{-3/2}$ and vary $C$ to fit the VAC in the range 0.5–3.0 ps; see Fig. 7. As above, standard deviations were obtained by dividing the data into subsets. It is by no means obvious where we should cut the numerical integration and use the analytical result to infinity, but the sum of the two parts does not change by more than a few percent depending on where we do the cut, further justifying our choice of fitting function. In the results that follow, we have taken the average of cuts at 0.5–3.0 ps in steps of 0.5 ps. This gives another standard deviation, which we add quadratically to the earlier one. As an example, the integral of the tail from 3.0 ps would contribute about 15% to the result.

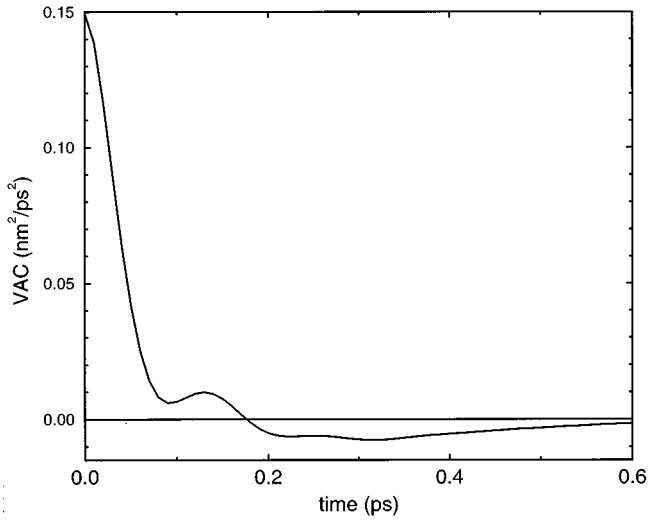


FIG. 6. Velocity autocorrelation function for the normal part of bin 2.

The resulting diffusion coefficients as functions of distance to the surface can be found in Fig. 5. The values from the VAC are in general lower and smoother than the corresponding data from the MSD. Note that the MSD values would be lower if one had chosen a lower boundary larger than 20 ps for the fits in Sec. III B. Both types show a clear decrease close to the protein surface. The decrease is much larger for the MSD data, and this does not show any difference between normal and parallel diffusion, whereas the VAC data show a decrease in normal $D$ almost twice the one in parallel. The most probable reason for the decrease being smaller with the VAC data is that this does not include the effect from the mean potential. The diffusion coefficient for bulk SPC water is $3.6\times10^{-9}$ m$^2$/s [15], which agrees reasonably well with our values far off from the protein, but is slightly larger than the experimental value $2.3\times10^{-9}$ m$^2$/s [20].

## IV. THE PRESENCE OF A SURFACE

Our results above confirm the observations of Bizzarri and Cannistraro [9] that water molecules close to a protein surface do not follow the ordinary diffusive relation (1.1) valid in an infinite space but rather an equation of the type

$$\langle(\mathbf{r}(t)-\mathbf{r}(0))^2\rangle=6Dt^{\alpha}. \tag{4.1}$$

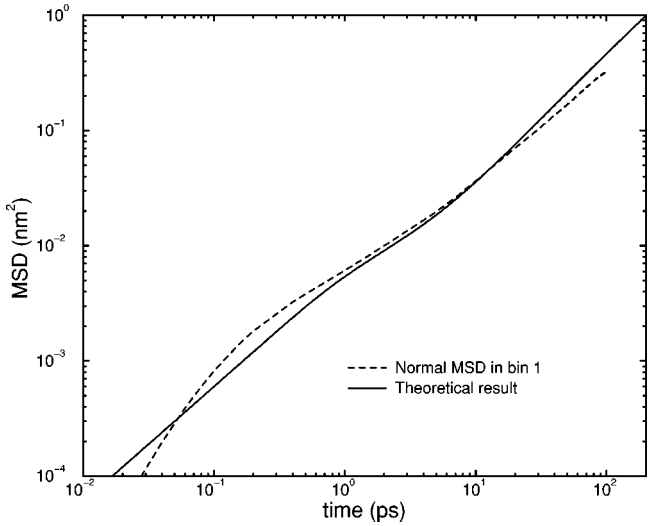


FIG. 8. Mean-square displacement of the solvent molecules in bin 1 vs time from the simulation and from the analytical solution of the diffusion equation outside a planar reflecting surface. $D=3\times10^{-9}$ m$^2$/s and $z(0)=0.14$ nm are used.

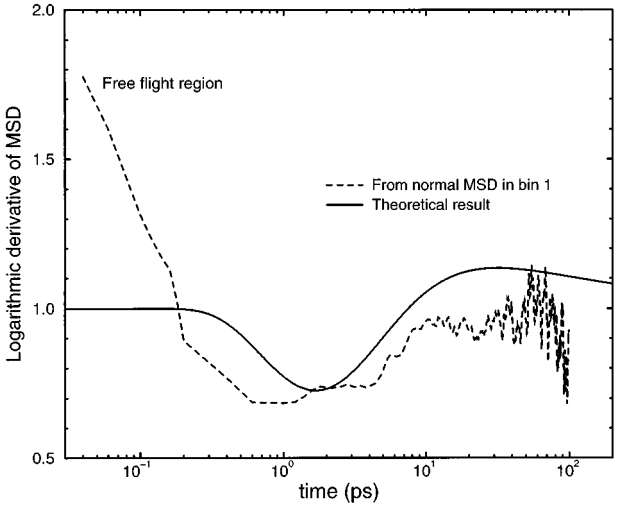


FIG. 9. Logarithmic derivatives of the curves in Fig. 8. This is approximately the exponent $\alpha$.

Reference [9] reports the exponent $\alpha$ having smaller values close to the protein (down to values as low as 0.6). This behavior was observed for times up to 10 ps, corresponding to displacements up to 0.3 nm.

We will try to explain this behavior by introducing the protein surface in the diffusion equation. The simplest model of the effect of a protein on surrounding solvent molecules is to treat the protein as a rigid reflecting boundary. Even with the water molecules undergoing perfect diffusion, this boundary condition will invalidate Eq. (1.1). If the protein surface is modeled as an infinite planar surface, the problem can be solved analytically in terms of the Green's function for the infinite problem and a mirror source. For diffusion in the two directions parallel to the surface the classical result

$$\langle[x(t)-x(0)]^2\rangle=\langle[y(t)-y(0)]^2\rangle=2Dt \tag{4.2}$$

is regained, while one obtains

$$\langle[z(t)-z(0)]^2\rangle=2Dt\left[1-\frac{2z(0)}{\sqrt{\pi Dt}}e^{-z(0)^2/4Dt}+\frac{z(0)^2}{Dt}\mathrm{erfc}[z(0)/\sqrt{4Dt}]\right] \tag{4.3}$$

in the perpendicular $z$ direction. Here the complementary error function erfc is defined as

$$\mathrm{erfc}(y)=1-\frac{2}{\sqrt{\pi}}\int_0^y e^{-x^2}dx. \tag{4.4}$$

The integral can be evaluated numerically or the function looked up in a table. Equation (4.3) is compared with the normal MSD from bin 1 in Fig. 8. For the constants in Eq. (4.3) we have chosen values reasonable in bin 1 [$D=3.0\times10^{-9}$ m$^2$/s, $z(0)=0.14$ nm], although the correct expression would be an average over $z(0)$ values in the bin. This approximation at least in part accounts for the simulation curve being more stretched out than the theoretical one.

A time-varying exponent $\alpha(t)$ in $\langle[z(t)-z(0)]^2\rangle=2Dt^{\alpha}$ may be estimated from the derivative of the logarithm of Eq. (4.3) with respect to ln$t$. Introducing $s=4Dt/z(0)^2$ to save space, the derivative becomes

$$\alpha(s)=1+\frac{\frac{2}{\sqrt{\pi s}}e^{-1/s}-\frac{4}{s}\mathrm{erfc}(1/\sqrt{s})}{1-\frac{4}{\sqrt{\pi s}}e^{-1/s}+\frac{4}{s}\mathrm{erfc}(1/\sqrt{s})}. \tag{4.5}$$

Clearly we get the classical result $\alpha=1$ in the limit of very long or very short times. This and the corresponding logarithmic derivative of the data from bin 1 are plotted in Fig. 9, with the same constants as above. Further away from the surface, $z(0)$ will be larger, and the kink will appear later. The physical reason for the variation of the exponent is the available space being smaller close to a protein due to the presence of the rigid wall. A solvent molecule in this region will therefore show a smaller MSD than a molecule far from the surface where the entire three-dimensional space is available. Thus, using Eq. (1.1) to evaluate the diffusion coefficient, one will get the local diffusion coefficient for times short enough for the solvent molecule not to feel the presence of the surface. For intermediate times the MSD will be too small due to the surface and again for very long times the molecules once close to the surface will on the average be quite far away, so the MSD approaches the bulk value. Apart from the free-flight region, we will thus get an exponent that begins at 1 and decreases, then increases above 1, and finally asymptotically approaches unity. At a curved surface, the exact solution above is no longer valid, but the qualitative effect still exists for the same physical reasons. One would expect it to be smaller at a convex surface and larger at a concave one.

## V. DISCUSSION

The two procedures to evaluate diffusion coefficients, from the slope of the MSD vs time and from the integral of the VAC, are formally equivalent. One formula can be derived from the other if the slope is taken at the limit of infinite time and the integral is taken up to infinite time. Neither, however, is possible to do using data from a computer simulation of limited length. In practice, one finds that in the latter case the integral can be cut at a fairly small finite time, especially if we make an analytic correction for the long-time algebraic tail. This makes it possible to reliably evaluate diffusion coefficients in the presence of macromolecular surfaces, mean potentials, and a strongly spatially varying diffusion.

When the MSD is used, we find with the present variation of the diffusion coefficients (Fig. 5) from Eq. (1.3) that we have to use times $t \ll 400$ ps. Since the nondiffusive effects at short times have died out after times $t \gg 1$ ps, there exists a time window at 10–100 ps for which diffusion coefficients should be possible to evaluate from the MSD. The presence of the protein surface may be more of a problem as seen from Fig. 9. The exponent $\alpha$ in the relation $2Dt^{\alpha}$ is smaller than one for short times and is still slightly varying when we approach 100 ps, which is in the entire suitable time window. These problems are avoided by using the Kubo formula (1.2) to calculate diffusion coefficients from the VAC instead.

The diffusion coefficients obtained from the VAC do not show an as large reduction close to the surface as the ones calculated from MSD or NMR experiments. The VAC describes very local events on time scales of picosecond, while the MSD involves motions over hundreds of picosecond. For these times a mean potential will result in a decreased mobility and thus a lower diffusion coefficient close to the surface. NMR data, finally, are typically obtained from correlation functions decaying over even larger time scales, say, 10–100 ns. The potential can be expected to have the same effect on NMR results and one must also keep in mind that the interpretation of relaxation data in terms of diffusion coefficients is far from trivial and relies upon modeling where the coefficients are fitted to yield either a correct frequency or concentration dependence of the relaxation.

The reduction of mobility from the potential cannot easily be distinguished from a reduced diffusion coefficient in a single NMR experiment. However, measurements as a function of temperature yield the same activation energy ($20 \pm 5$ kJ/mol) in bulk and at the surface [3]. This does exclude a very high energy barrier, but not one of 2–3 kJ/mol as in our case. Taking the potential into account, we altogether get a reduction of water mobility close to the protein by a factor 5–10. This is consistent with [1,2], but lower than the factor 10–100 of [3].

## ACKNOWLEDGMENT

We thank the Göran Gustafsson Foundation for grants supporting the workstations making the present simulations possible.

[1] C. F. Polnaszek and R. G. Bryant, J. Am. Chem. Soc. **106**, 429 (1984).
[2] R. Kimmich, T. Gneiting, K. Kotitschke, and G. Schnur, Biophys. J. **58**, 1183 (1990).
[3] B. Halle and L. J. Piculell, J. Chem. Soc. Faraday Trans. 1 **82**, 415 (1986).
[4] M. Levitt and R. Sharon, Proc. Natl. Acad. Sci. USA **85**, 7557 (1988).
[5] C. L. Brooks III and M. Karplus, J. Mol. Biol. **208**, 159 (1989).
[6] M. Norin, F. Haeffner, K. Hult, and O. Edholm, Biophys. J. **67**, 548 (1994).
[7] I. Muegge and E. W. Knapp, J. Phys. Chem. **99**, 1371 (1995).
[8] P. Ahlström, O. Teleman, and B. Jönsson, J. Am. Chem. Soc. **110**, 4198 (1988).
[9] A. R. Bizzarri and S. Cannistraro, Phys. Rev. E **53**, 3040 (1996).
[10] A. R. Bizzarri and S. Cannistraro, Europhys. Lett. **37**, 201 (1997).
[11] See, e.g., J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic, London, 1976), pp. 239–241.
[12] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[13] M. Bolognesi, S. Onesti, G. Gatti, A. Coda, P. Ascenzi, and M. Brunori, J. Mol. Biol. **205**, 529 (1989).
[14] W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual* (Biomos, Groningen, 1987).
[15] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), p. 331.
[16] J. P. Ryckaert, G. Cicotti, and H. J. C. Berendsen, Comput. Phys. **23**, 327 (1977).
[17] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 2684 (1984).
[18] B. Lee and F. M. Richards, J. Mol. Biol. **55**, 379 (1971).
[19] B. J. Alder and T. E. Wainwright, Phys. Rev. A **1**, 18 (1970).
[20] G. D. Zeiss and W. J. Meath, Mol. Phys. **30**, 161 (1975).