Hydration dependence of the mass fractal dimension and anomalous diffusion of vibrational energy in proteins

Matthew B. Enright, Xin Yu, and David M. Leitner*
Department of Chemistry and Chemical Physics Program, University of Nevada, Reno, Nevada 89557, USA
(Received 9 December 2005; revised manuscript received 10 March 2006; published 8 May 2006)

Vibrational dynamics of proteins and energy flow depend on protein geometry as well as interactions of a protein molecule with the surrounding solvent. We compute the mass fractal dimension D of proteins ranging from 100 to over 10 000 amino acids comparing values for the bare protein with those computed when buried and hydration waters are included in the calculation. Including water in the calculation increases D by about 0.3 to 2.87 on average above D computed for the dehydrated protein. The mass fractal dimension of proteins that are partially unfolded by molecular dynamics (MD) simulation is also computed and found to vary little when the radius of gyration changes within about 10% of that for the Protein Data Bank structure. MD simulations of vibrational energy diffusion in proteins reveal that the exponent characterizing anomalous diffusion of vibrational energy does not change much with hydration, which is seen to be due to an increase in the spectral dimension with hydration by a factor similar to the increase in D.

DOI: 10.1103/PhysRevE.73.051905 PACS number(s): 87.10.+e, 87.15.−v, 82.35.Lr

I. INTRODUCTION

Protein relaxation dynamics depends on interactions within a protein molecule, between the protein and the surrounding solvent molecules [1–7], and on the geometry [8,9] of the protein itself. Single molecule experiments are providing a close look at the rich complexity of protein motion including structural fluctuations between different parts of a protein [10–12]. These fluctuations can be described, as recently demonstrated by Granek and Klafter [13], by exploiting connections between protein geometry and dynamics, properties that are related by the "fraction" models of Alexander and Orbach [14], who explored the vibrational properties of fractal objects, motivated by the possibility of fracton vibrations of protein molecules [15–18]. The vibrational dynamics of a fractal object depends on two characteristic dimensions, the mass fractal dimension and the spectral dimension [14,19]. Values for the spectral dimension [20] and mass fractal dimension [21] have been computed in recent years for proteins ranging in size from about 100 to several thousand amino acids. These calculations have been carried out for the bare protein, using structures available from the Protein Data Bank (PDB). Because the low-frequency vibrations of proteins are coupled to that of the solvent, it is of interest to compute protein dimensions including buried and hydration water molecules in the calculation. Also, a protein in water is not static and its structure is not precisely that captured in the PDB. In this paper, we examine the influence of buried and hydration water molecules on the mass fractal dimension, and how they influence the flow of vibrational energy in a protein molecule. We also address how modest change in protein structure by partial unfolding from the crystallographic structure influences the mass fractal dimension.

We recently computed the mass fractal dimension for 200 proteins from the PDB excluding all water from the calculation [21]. For this set we found the dimension to range from about 2.3 for proteins with about 100 amino acids to about 2.6 for proteins with more than 1000. The former value is consistent with the diffusion of vibrational energy computed for several proteins ranging in size from about 100 to 230 amino acids [22]. The presence of hydration water can influence a calculation of the mass fractal dimension even near the core, as the "surface" of a protein typically contains many cavities, which may approach the core. It is thus of interest to include the water molecules in the calculation of the mass fractal dimension explicitly.

We present below results of calculations of the mass fractal dimension for proteins with solvent water molecules present. Using the same set of 200 proteins studied in Ref. [21], we now include in our calculation of the mass fractal dimension D the water molecules that were captured in the crystallographic structure. We find that inclusion of these buried water molecules and some hydration water molecules only slightly increase the computed value of D. To address the influence of the solvation water on D we select 20 of the 200 proteins ranging in size from 100 to 4000 amino acids and add water to the protein by molecular dynamics (MD) simulation. We find that the mass fractal dimension of protein molecules is on average 2.8–2.9 when both buried and hydration water molecules are included in the calculation, about 0.2–0.3 higher than when water is not included.

It is also of interest to address how sensitive the fractal dimension computed for a protein is to the crystallographic structure. To begin addressing this question, we partially unfold ten proteins ranging in size from about 100 to about 400 amino acids by heating them in a MD simulation. The extent of unfolding is quantified by changes in the radius of gyration R_G from that of the native structure. We have carried out the simulations to address modest structural changes rather than unraveling of the collapsed polymer, for which the mass fractal dimension D clearly becomes smaller. For changes of R_G by about 10% we find below little change in the value of D.

An important motivation for investigating the mass fractal dimension of proteins is its influence on energy flow. We

*Corresponding author. Email address: dml@chem.unr.edu

TABLE I. A list of the 20 proteins for which MD simulations were carried out. Listed after the PDB code is, from left to right, the number of residues (N); the rms deviation per atom from the PDB structure after the 50-ps simulation (rmsd); $D_{10\%}$ computed for the PDB structure ($D_{10\%}$ PDB); $D_{10\%}$ computed for the protein after the 50-ps simulation and without water ($D_{10\%}$ dehy); $D_{10\%}$ computed for the protein after the simulation including water ($D_{10\%}$ sol); D computed for the PDB structure (D PDB); D computed for the protein after the 50-ps simulation and without water (D dehy); D computed for the protein after the simulation including water (D sol).

Name	N	rmsd	$D_{10\%}$ PDB	$D_{10\%}$ dehy	$D_{10\%}$ sol	D PDB	D dehy	D sol
9RNT	104	1.401	2.631	2.544	2.938	2.315	2.258	2.815
1RP7	159	0.992	2.753	2.668	2.923	2.352	2.341	2.837
2AYH	214	0.887	2.768	2.770	2.942	2.423	2.417	2.846
1EMB	225	1.065	2.730	2.739	2.918	2.419	2.427	2.854
1RAY	258	0.866	2.769	2.778	2.913	2.443	2.453	2.855
1SBP	309	0.961	2.761	2.780	2.925	2.452	2.454	2.853
1R66	322	0.905	2.805	2.813	2.909	2.480	2.486	2.866
1RGY	360	1.069	2.807	2.811	2.917	2.508	2.513	2.867
1A39	401	1.032	2.766	2.789	2.925	2.463	2.481	2.860
1RJP	474	1.139	2.844	2.867	2.927	2.523	2.539	2.872
1CPU	495	0.982	2.834	2.841	2.949	2.512	2.515	2.861
1CRL	534	0.876	2.859	2.878	2.947	2.526	2.542	2.863
1EX1	602	1.059	2.823	2.850	2.939	2.547	2.560	2.876
1A47	683	1.198	2.833	2.850	2.941	2.534	2.553	2.862
1RGN	823	1.318	2.683	2.702	2.905	2.463	2.468	2.854
1HTY	988	1.631	2.872	2.887	2.954	2.592	2.592	2.885
1KEV	1404	1.185	2.823	2.830	2.934	2.597	2.605	2.899
1KY5	1720	1.248	2.853	2.874	2.949	2.614	2.630	2.890
1KEK	2462	1.257	2.928	2.922	2.958	2.653	2.656	2.901
1CW3	3952	1.299	2.769	2.804	2.940	2.668	2.679	2.908

therefore compare vibrational energy diffusion in proteins both with and without water. We shall see that at least for the modest-sized proteins we consider for this study, with about 200–250 amino acids, there is little change in the exponent that characterizes anomalous diffusion ν when the protein is hydrated. We find that this is due to increases by similar percentages upon hydration of both the mass fractal dimension and the spectral dimension, the ratio of which determines ν.

In the following section we describe the MD simulations we use to solvate the proteins and partially unfold them, and discuss the method for computing the mass fractal dimension D for each protein in our sample set. We also present the method that we use for computing the flow of vibrational energy in protein molecules. In Sec. III we present results for D for bulk water, hydrated proteins, and for denatured proteins. We also present results for the diffusion of vibrational energy in dehydrated and hydrated proteins and discuss these results in terms of the differences in D and the spectral dimension for these systems. Concluding remarks are given in Sec. IV.

II. COMPUTATIONAL METHODS

A. Mass fractal dimension

Structures for the 200 proteins in our data set were obtained from the PDB. The complete list of the 200 proteins is provided in Table I of Ref. [21]. The 20 proteins for which MD simulations have been run before computing D are listed in Table I above.

The MD simulations were performed using the GROMACS 3.2 package [23,24], with the GROMO96 force field [25]. For each simulation, the protein molecule was embedded in a box containing simple point charge (SPC) model water [26] that extended to 12 Å between the protein and the edge of the box.

Before the MD simulation, internal constraints were relaxed by energy minimization using the steepest descent method with no constraints. The minimized structure was supplied to a position-restrained dynamics for 50 ps, where lengths of all bonds in the model system were constrained with the LINCS algorithm [27]. The time step used was 1 fs with a neighbor list update every ten steps. The simulations were run as an NTV ensemble, using Berendsen's coupling algorithm [28] to keep the temperature at 300 K. Periodic boundary conditions were applied in all three dimensions. Electrostatic interactions were evaluated using the particle-mesh Ewald (PME) [29,30] method with a grid width of 1.2 Å and a fourth-order spline interpolation; van der Waals interactions were truncated at 9 Å. The coordinates were saved every 10 ps.

A similar procedure to the one described above was repeated again to obtain unfolded proteins. In this case, MD simulations were carried out for ten proteins at 1000 K for 120 ps. MD simulations have also been performed for pure

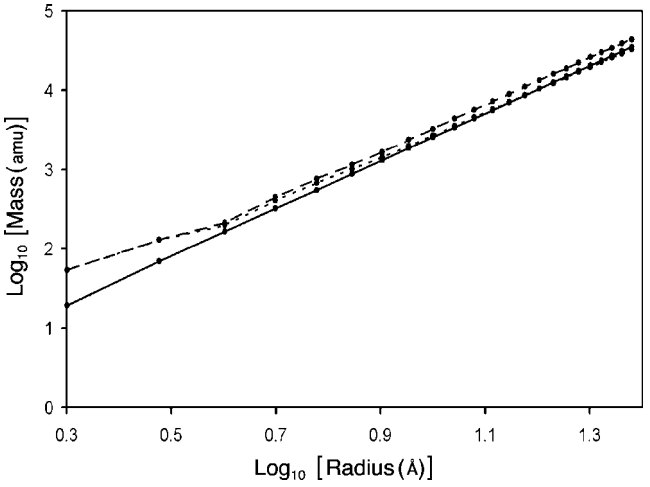

FIG. 1. Plot of $\log_{10} M$ vs $\log_{10} R$ for a box of water molecules (points connected by solid line) and the protein 1 CW3 (two sets of points connected by dashed lines), where values of M are the masses enclosed by concentric spheres of radius R centered at an O atom for water, or a C_α for the protein. The two sets of points for the protein correspond to using the closest C_α to the center of mass of the protein. Points connected by the long-dashed line correspond to a calculation of M including water molecules, whereas those connected by the short-dashed line correspond to a calculation of M where only protein atoms are included.

water in the box containing 17 122 water molecules. The total simulation time was 50 ps and the temperature was 300 K.

The mass fractal dimension D is defined by

$$M \sim R^D, \tag{1}$$

where M is mass and R is a length scale. The dimension D can be computed for a single protein by plotting the mass of all atoms contained inside concentric spheres of radius R on a log-log scale. The slope gives D. We recently [21] carried out this calculation for 200 proteins ranging from $N \approx 100$ to 11 000 amino acids including the 58 analyzed in Ref. [20]. We neglected any water molecules in that calculation, and here report results including these water molecules, in addition to calculations with the added water after MD simulations.

Describing how we calculate D in practice is easiest by example. Figure 1 provides an example for a calculation on the protein 1CW3 with 3952 amino acids, and for bulk water. Figure 1 presents a log-log plot of the enclosed mass M of all protein atoms and water molecules inside a sphere as a function of its radius R. The three sets of points, where each set appears to fall on a line when R is at least 4 Å, have been computed for concentric spheres centered at an α carbon for the protein calculation and an oxygen atom for the water calculation. Protein results are shown when the calculation is centered at the closest α carbon to the center of mass of the protein, and another farther away. Data are shown for R ranging from 2 to 20 Å. The length scale of this particular protein is significantly larger than 20 Å, but we nevertheless only calculate M for R from 5 Å up to 20 Å to avoid finite-size effects when computing D for the interior of the protein. This range strikes a balance between having enough points to meaningfully compute D, and keeping those over a range where D does not change much with small changes in the number of points used in the calculation. To avoid as much as possible finite-size effects when computing D for the interior of the smallest proteins in our sample set, we have computed M for R up to 16 Å for proteins with up to 200 amino acids, and up to 18 Å for proteins with from 200 to 400 amino acids. Nevertheless, the results that we report below are very similar to those that we obtain when we calculate M as a function of R up to 20 Å for all proteins. The lower value of $R=5$ Å was chosen after considering 3–8 Å as a lower limit, and fitting lines to these. The largest correlation coefficient was found with 5 Å, since significant deviations from the best-fit line were typically found for points with smaller R. The average value of the slopes of the lines in Fig. 1 gives us an estimate for D for the protein 1CW3, which we calculate by averaging slopes obtained for such plots using all of the C_α's of the protein backbone as centers.

We shall address how partial unfolding of the protein influences the mass fractal dimension. One measure of unfolding of a protein is radius of gyration, R_G. For a given protein configuration we compute the radius of gyration as

$$R_G = \sqrt{\sum_i m_i \mathbf{r}_i^2 \Big/ \sum_i m_i}, \tag{2}$$

where the sum is over each atom i of mass m_i and distance $\mathbf{r}_i$ from the center of mass.

B. Vibrational energy diffusion

We examine how the mean square displacement of kinetic energy, $\langle R^2 \rangle$, in a protein varies with time. For a diffusive process,

$$\langle R^2 \rangle \sim t^{2\nu}, \tag{3}$$

where $\nu = 1/2$ corresponds to normal diffusion. We see below that in proteins $\nu < 1/2$ so that the process is subdiffusive.

Anomalous subdiffusion occurs on percolation clusters or on fractal objects generally. Connections between ν, the fractal dimension of the cluster D, and the spectral dimension $\bar{d}$, have been established, relations that were originally derived by Alexander and Orbach [14], who developed a theory of vibrational excitations on fractal objects which they called fractons. They found that

$$\nu = \frac{\bar{d}}{2D}, \tag{4}$$

where D is the mass fractal dimension and $\bar{d}$ is the spectral dimension, which describes the scaling of the density of vibrational states, $\rho(\omega)$, with frequency,

$$\rho_L(\omega) \propto \omega^{\bar{d}-1}. \tag{5}$$

We shall calculate for several proteins the time dependence of the variance of a vibrational energy wave packet [22]. We consider the kinetic energy $E_n(t)$ of atom n at time

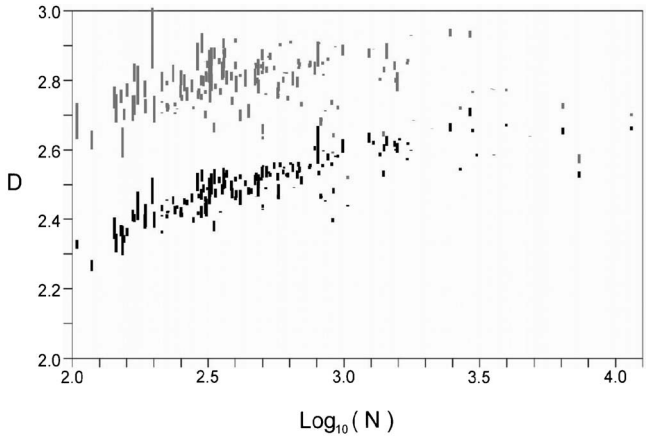

FIG. 2. Values of $D_{10\%}$ (gray) and D (black) computed for 200 proteins in the PDB, each containing N residues. Calculations were carried out by excluding water molecules, in which case the dimension is the value at the bottom of the vertical line in the plot, and including the water in the PDB, in which case the dimension is the value at the top of the vertical line.

t. We start with a cold protein and introduce a spatially localized excitation in the form of a wave packet placed near a backbone atom in the interior of the molecule. The center of kinetic energy, $\mathbf{R}_0(t)$, of the wave packet is found from

$$\mathbf{R}_0(t) = \frac{\sum_n \mathbf{R}_n E_n(t)}{\sum_n E_n(t)}. \tag{6}$$

The variance is

$$\langle \mathbf{R}^2(t) \rangle = \frac{\sum_n [\mathbf{R}_n - \mathbf{R}_0(t)]^2 E_n(t)}{\sum_n E_n(t)}. \tag{7}$$

Our focus is on the propagation of a wave packet by the normal modes of the protein. We idealize the initial wave packet as a traveling wave and take the displacement of atom n to initially have the form

$$\mathbf{U}_n(t) = \frac{\mathbf{A}_n}{b^2} \exp\left(- \frac{(\mathbf{R}_n - \mathbf{R}' - \mathbf{v}_0 t)^2}{2b^2} \right) e^{i(\mathbf{Q}_0 \cdot \mathbf{R}_n - \omega_0 t)}, \tag{8}$$

from which displacements and velocities at $t=0$ are determined. Displacements and velocities for $t>0$ are then computed in terms of normal modes as

$$\mathbf{U}_n(t) = \sum_\alpha \mathbf{e}_n^\alpha \cos(\omega_\alpha t) \sum_{n'} \mathbf{e}_{n'}^\alpha \cdot \mathbf{U}_{n'}(0) + \sum_\alpha \mathbf{e}_n^\alpha \frac{\sin(\omega_\alpha t)}{\omega_\alpha} \sum_{n'} \mathbf{e}_{n'}^\alpha \cdot \mathbf{V}_{n'}(0), \tag{9a}$$

$$\mathbf{V}_n(t) = \sum_\alpha \mathbf{e}_n^\alpha \cos(\omega_\alpha t) \sum_{n'} \mathbf{e}_{n'}^\alpha \cdot \mathbf{V}_{n'}(0) - \sum_\alpha \mathbf{e}_n^\alpha \omega_\alpha \sin(\omega_\alpha t) \sum_{n'} \mathbf{e}_{n'}^\alpha \cdot \mathbf{U}_{n'}(0). \tag{9b}$$

In the computational work that follows, the variance of the wave packet as a function of time that we report is an average over results for ten wave packets, centered initially about an atom of the protein backbone whose position $\mathbf{R}'$ lies near the center of mass of the protein. The width of the initial wave packet is $b=3$ Å. The magnitude of the wave vector of the initial excitation, Q_0, is 0.63 Å^{-1} and it points +45° from the x, y, and z axis; we take $\omega_0=9.4$ ps^{-1} for wave packets that include all normal modes, and $\mathbf{v}_0=0$ for simplicity. All components of $\mathbf{A}_n$ for all atoms are taken to be the same, and the magnitude is unimportant as it cancels out when we compute the center of energy and its variance. Displacements and velocities are then propagated with Eq. (9), which give the positions $\mathbf{R}_n(t)$ and kinetic energies $E_n(t)$ used in Eqs. (6) and (7) to locate the center of energy and variance in the protein.

III. RESULTS AND DISCUSSION

A. Fractal dimension of folded proteins

Figure 1 shows a plot of $\log_{10} M$ vs $\log_{10} R$ for bulk water, where the center of the calculation is an oxygen atom centered in a box of water molecules at 1 g cm^{-3}. The average value of D that we find for 20 such calculations is 2.995 ± 0.051, where the error represents two standard deviations. We thus find for bulk water the expected value of $D=3$. This result stands in contrast to the smaller values we find for proteins. Figure 1 also provides a similar plot for a particular protein, 1CW3, discussed above.

Figure 2 presents results for the mass fractal dimension computed for 200 proteins ranging in size from 100 to 11 000 amino acids. The full list of proteins is provided in Table I of Ref. [21]. Values of D shown in Fig. 2 were computed both with and without the water molecules included in the PDB. The value of D is found to be greater for each protein when the water molecules are included. In Fig. 2 we show results for $D_{10\%}$, which are the values of D computed using as centers the innermost 10% of the α carbons of each protein. This is calculated as illustrated in Fig. 1 for the protein 1CW3, where $\log_{10} M$ vs $\log_{10} R$ is plotted using as a center the α carbon closest to the center of mass. In this case we find a slope that gives a value of 2.94 for the mass fractal dimension when water is included in the calculation, and 2.80 when water molecules are excluded.

Computing the slopes using the nearest 10% of all α carbons to the center of mass and averaging, we compute $D_{10\%}$ for all proteins in the set both with and without the inclusion of water molecules. These results are shown as gray lines, where the top of the vertical line is the value of $D_{10\%}$ computed when all water molecules in the PDB are included, and the bottom of the vertical line is the value computed when water is removed. The average value of $D_{10\%}$ is found to be 2.76 when water is absent in the calculation and 2.79 when it

is present. The black lines in Fig. 2 represent values of D, computed using as centers all α carbons of the protein. Again, the top of the vertical line is the value of D computed when water is present in the calculation, and the bottom when water is absent. The average value of D for the set of 200 proteins is found to be 2.49 when water is absent in the calculation and 2.52 when it is present. The overall difference for both $D_{10\%}$ and D is thus rather small. For proteins with less than 1000 amino acids we find D for the dehydrated proteins to be on average 2.46, and is on average 2.58 for proteins with at least 1000 amino acids. With the water molecules included in the PDB we find values of 2.49 and 2.60, respectively. We see also that values of $D_{10\%}$ and D appear to converge to similar values, about 2.6 or 2.7, for the larger proteins in the set.

The PDB contains water molecules buried inside the protein and some water molecules surrounding the protein. To address the influence of solvent water molecules on our calculation of D, we need to surround the protein with more water. We have thus taken the coordinates of the protein molecules and transformed them to the center of a box in which we have surrounded the protein with water. To allow the water molecules to adjust to the protein, we have run MD simulations of the solvated proteins, as described in Sec. II. We have carried out this procedure and calculated D for the hydrated protein for 20 proteins out of the 200 in the set, ranging in size from 100 to 4000 amino acids. The 20 proteins and results from the calculations are listed in Table I.

The protein structure of course changes during the 50-ps MD simulation with water, and is different from the PDB structure. We list in Table I the rms deviation per protein atom of the structure after the MD simulation compared to the PDB structure, where we see that it is about 1 Å per atom. We also list the values of $D_{10\%}$ and D for each of the 20 proteins before and after the 50-ps MD simulation, the latter computed without water molecules to compare with the values using the PDB structure without water present. We see that there is little change in these values. For the 20 proteins in the table, $D_{10\%}$ is 2.800 ± 0.170 using the PDB structure, where the error is two standard deviations. Using the protein structure after the simulation we find $D_{10\%}$ to be 2.796 ± 0.132. Similarly, D, computed using all α carbons of the protein as centers for the calculation, does not change much in going from the PDB structure to the structure after the simulation. For the 20 proteins in the table, D is 2.508 ± 0.202 using the PDB structure and is 2.504 ± 0.202 using the protein structure after the simulation.

Addition of water to the protein increases the values of $D_{10\%}$ and D, as seen both by comparing their values listed in Table I, and by the comparison provided in Fig. 3, where we plot these values for the 20 proteins. As in Fig. 2, we plot values of $D_{10\%}$ as gray vertical lines; the top of the line is the value of $D_{10\%}$ computed when water is present, and the bottom of the line when water is absent. On average we find that $D_{10\%}$ is 2.796 ± 0.132 for the 20 proteins when water is absent, and is 2.933 ± 0.030 with water. We see in Fig. 2 that the difference appears to diminish for the larger proteins in the set. The value of $D_{10\%}$ does not appear to depend much on protein size when water is included in the calculation. The value of D computed using all α carbons as centers for the calculation also increases when accounting for water molecules. We find D to be 2.504 ± 0.202 for the 20 proteins when water is absent, and is 2.866 ± 0.044 with water. Again, the difference appears to become smaller as the size of the protein increases. We see also that values of $D_{10\%}$ and D both approach 2.8–2.9 when water is included in the calculation for proteins of any size.

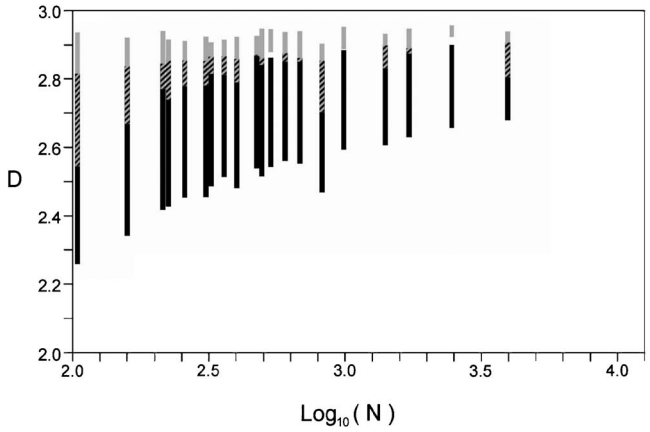

FIG. 3. Same as Fig. 2, but only for the 20 proteins listed in Table I. Values of D were computed after 50-ps MD simulations. The value at the top of the vertical bars represents the dimension computed with inclusion of the surrounding water molecules.

Figure 3 indicates that calculation of $D_{10\%}$, a representative dimension for the "core" of the protein, is influenced by the presence of water, even for larger proteins. This is because water molecules approach the core through numerous cavities in the "surface" of the protein. It is of interest to examine how the relative mass of water that is included in the calculation varies as the size of the concentric spheres increases. We provide this information in Fig. 4, which shows the mass of water relative to the mass of all atoms contained in a sphere of radius R for R up to 24 Å. We note that $D_{10\%}$ is calculated using R up to 20 Å, and less for the smaller proteins, as described in Sec. II. We see in Fig. 4 that for a number of proteins the mass of water is 10% of the total mass in the sphere when its radius is 12 Å around α carbons near the core, and is 10% of the total mass in the sphere for most of the proteins when the radius is 14 Å and beyond. With the exception of the two smallest proteins in the set, the relative amount of water never exceeds 20% in any of the spheres used for the calculation of $D_{10\%}$. For the others, we see that the content of water varies from about 5% to 10% for R ranging from 5 to 12 Å in the calculation of $D_{10\%}$, and from 10% to about 20% for the larger R. While for the larger proteins there is typically less water in the calculation, there are clearly exceptions, as Fig. 4 reveals. For example, the fraction of water is about twice as high in 1KY5, with 1720 amino acids as in 1A47, with 683 amino acids, for R up to about 12 Å.

B. Fractal dimension of partially unfolded proteins

The mass fractal dimension for the core of folded proteins is about 2.8 for an isolated protein, and about 2.9 when water

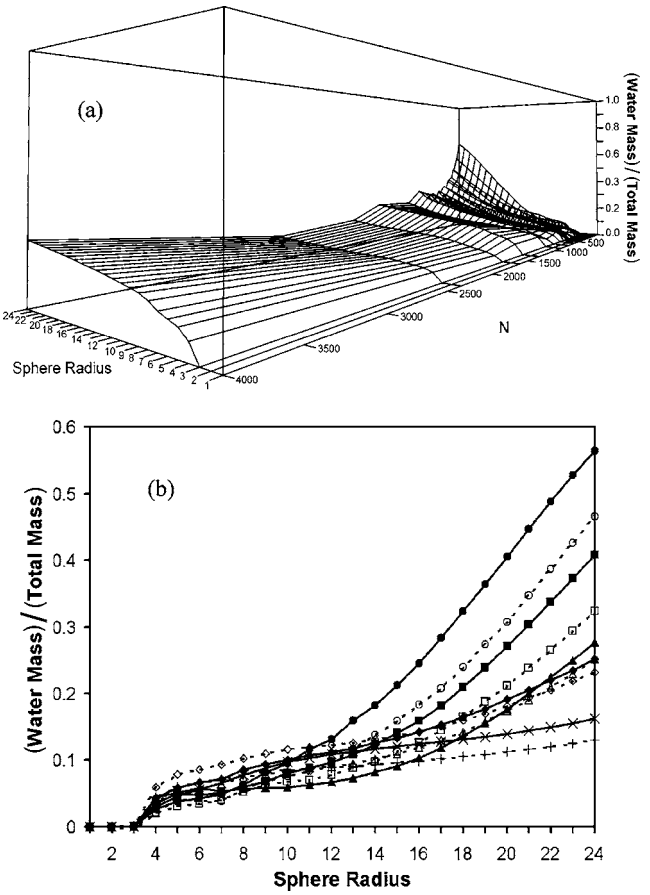

FIG. 4. (a) The fraction of water mass to total mass contained in spheres of radius R (Å), around the 10% of C_α's closest to the center of mass of the proteins listed in Table I is plotted against protein size N and radius R; (b) shows this fraction for some of these proteins as a function of R. Values of N plotted are 159 (filled circles), 249 (open circles), 258 (filled squares), 360 (open squares), 474 (filled triangles), 534 (open triangles), 683 (filled diamonds), 988 (open diamonds), 1720 (×), and 2462 (+).

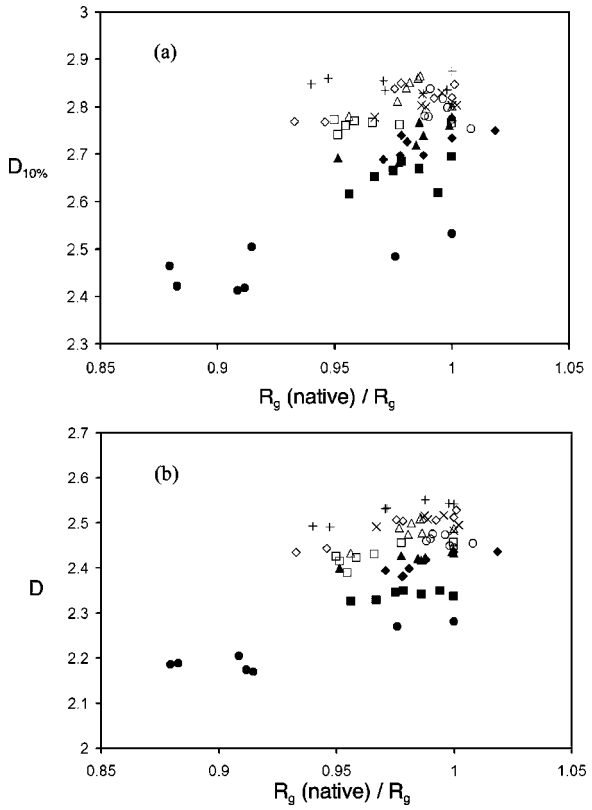

FIG. 5. Plot of (a) $D_{10\%}$ and (b) D for the ten smallest proteins listed in Table I, as a function of R_G for the native structure, relative to R_G of a particular structure obtained by partial unfolding during MD simulation, $R_G(\text{native})/R_G$. N is 104 (filled circles), 159 (filled squares), 214 (filled triangles), 225 (filled diamonds), 258 (open circles), 309 (open squares), 322 (open triangles), 360 (open diamonds), 401 (×), and 474 (+).

molecules lying towards the core are included in the calculation. In this subsection we consider how D and $D_{10\%}$, where the latter represents the core of the protein, change when the protein undergoes partial unfolding from its native structure. For this comparison we calculate D and $D_{10\%}$ for the protein without including surrounding water molecules. Partial unfolding of the ten smallest proteins listed in Table I was carried out by MD simulation at 1000 K for 120 ps. We then computed D and $D_{10\%}$ for the six structures obtained in 20-ps intervals for each of these ten proteins. Partial unfolding during these simulations could lead to a larger or sometimes smaller radius of gyration, R_G, of about 10% from that for the native structure. The more dramatic unfolding corresponding to unraveling of the collapsed globule clearly must yield smaller D and is not considered here.

Figure 5 shows how $D_{10\%}$ and D vary as $R_G(\text{native})/R_G$ varies. Values of $D_{10\%}$ and D at $R_G(\text{native})/R_G=1$ are just those listed in Table I for the ten smallest proteins when no water is included in the simulation. We see in Fig. 5 both $D_{10\%}$ and D change little with variation of $R_G(\text{native})/R_G$ in the range from about 0.85 to 1.05. For the larger proteins in this set there is little noticeable change in either $D_{10\%}$ or D with change in $R_G(\text{native})/R_G$, while there appears to be a slight decrease in D with decreasing $R_G(\text{native})/R_G$ for the smaller proteins with about 225 or fewer amino acids. Most of the changes during a simulated unfolding begin at the surface [31], which influences the calculation of the mass fractal dimension more for the smaller proteins than larger ones. Nevertheless, changes in R_G by about 10% appear to barely influence the mass fractal dimension of the protein.

C. Vibrational energy diffusion

We carry out calculations of vibrational energy diffusion in two proteins, green fluorescent protein, 1 emb, a protein with 225 amino acids; and 1 ray, a protein with 258 amino acids. These proteins, while not the smallest that we have studied above, are small enough so that we can compute the flow of vibrational energy by the vibrational modes including a layer of water molecules, too. We compare for both proteins the diffusion of vibrational energy without water and including all water molecules that contain any atom within 3 Å of each protein. The hydrated proteins, 1 emb and 1 ray, are then surrounded by 600 and 602 water molecules, respectively.

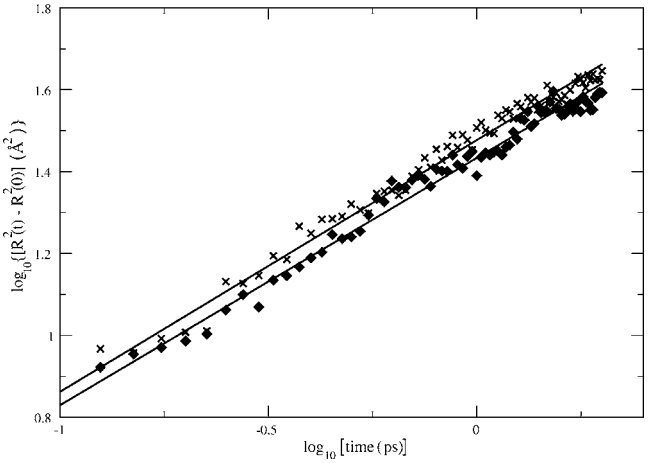

FIG. 6. Plot of $\log[\langle R^2(t)\rangle - \langle R^2(0)\rangle]$ vs $\log(t)$ for green fluorescent protein (1 emb), without water (diamonds) and with the 600 water molecules ($\times$) within 3 Å of the protein. The slopes of the plotted lines fit to the results from 0.1 to 2.0 ps are 0.60 and 0.61 for the dehydrated and hydrated protein, respectively, indicating anomalous subdiffusion with exponent $\nu=0.30$ for both systems.

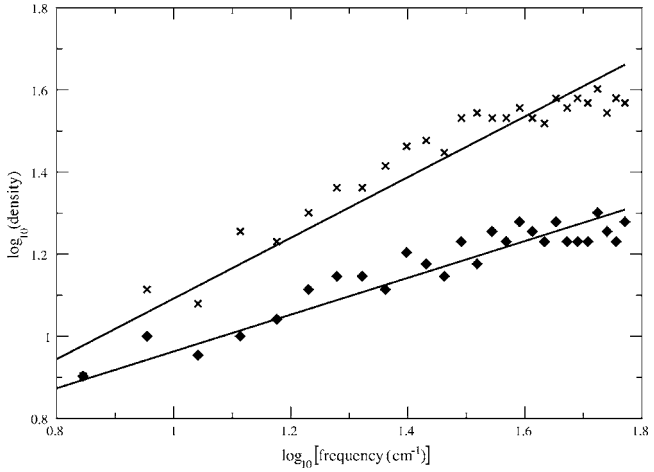

FIG. 7. The log of the vibrational density, $\log_{10}\rho$, is plotted vs the log of the vibrational frequency, $\log_{10}\omega$, for 1 emb from frequencies of 5–60 cm^{-1}. The slope of the line through the data is $\bar{d}-1$, where $\bar{d}$ is the spectral dimension. A best fit through the points for 1 emb without (diamonds) and with ($\times$) hydration water gives spectral dimensions of 1.45 and 1.74, respectively.

For this first layer of water, the dynamics of the protein and water are strongly coupled [1–7]. As one indication of coupling between protein and hydration water vibrations, we have calculated the projection of the water molecules onto the vibrational modes up to 200 cm^{-1}. The number of protein atoms explicitly used in the calculation of the vibrational modes of 1 emb is 2290, compared to 1800 atoms of the 600 water molecules, or about 44% of the atoms belong to water. Only the low-frequency normal modes of a protein, up to 150 or 200 cm^{-1}, are delocalized and carry vibrational energy through it [22]. We find the projection of the vibrational eigenmodes onto the water to be similar to this percentage; on average 50% of each eigenmode up to 100 cm^{-1} projects onto water molecules, while on average 44% of each eigenmode projects onto water molecules between 100 and 200 cm^{-1}. We thus find as expected that the low-frequency vibrational modes for the protein and hydration water in the first shell around the protein are strongly coupled and cannot be assigned to either the protein or the water.

Results for the average variance of ten wave packets in both 1 emb and hydrated 1 emb are presented as a function of time in Fig. 6. A best fit to a log-log plot of the data for 1 emb and hydrated 1 emb from a time of 0.1–2.0 ps gives an exponent of 0.60 ± 0.02 and 0.61 ± 0.02, respectively, where the error is twice the standard error in the slope, so that $\nu=0.30$ for both the dehydrated and hydrated protein. The presence of solvent thus does not appear to influence the anomalous diffusion of vibrational energy in the protein. Similarly, a best fit to the data for dehydrated and hydrated 1 ray over the same time interval gives an exponent of 0.57 ± 0.02 and 0.60 ± 0.02, respectively, so that $\nu=0.29$ or 0.30 for both the dehydrated and hydrated protein. Again, the presence of solvent does not appear to influence the anomalous diffusion of vibrational energy in the protein. We note that we reached a similar conclusion with the calculation of vibrational energy diffusion in cytochrome c, a protein with 103 amino acids, for which ν was found to be 0.26 for both the dehydrated protein and the protein hydrated with 400 water molecules [22]. The reason why ν is the same for hydrated and dehydrated cytochrome c was not investigated.

The exponent ν is related to the fractal dimension D and the spectral dimension $\bar{d}$ by $\nu=\bar{d}/2D$ [Eq. (4)]. We have seen that for proteins the size of 1 emb the fractal dimension increases from about 2.4 to about 2.8, or by 15–20% (Table I). The spectral dimension would then have to increase by this percentage for ν to remain unchanged. In Fig. 7 we plot for 1 emb the log of the vibrational density, $\log_{10}\rho$ versus the log of the vibrational frequency, $\log_{10}\omega$, for frequencies from 5 to 60 cm^{-1}. At higher frequencies, the plot appears less linear and the vibrational density starts to decrease with vibrational frequency. A best fit to the data in Fig. 7 gives a slope of 0.45 ± 0.05 and 0.74 ± 0.05 for dehydrated and hydrated 1 emb, respectively. Since the slope is $\bar{d}-1$, this gives spectral dimensions of 1.45 and 1.74, respectively, for 1 emb and hydrated 1 emb. The fractal dimension of 1 emb is 2.43 when calculated without water and 2.85 when calculated with water (Table I). With Eq. (4), we then expect that the exponent ν characterizing the spreading of vibrational energy in 1 emb to be 0.30 for both dehydrated and hydrated 1 emb, respectively, the same value as found by fitting to the time dependence of the variance of vibrational wave packets in these systems. Similarly, we find for dehydrated and hydrated 1 ray spectral dimensions of 1.50 ± 0.05 and 1.78 ± 0.05. The mass fractal dimension for dehydrated and hydrated 1 ray was found to be 2.45 and 2.86, respectively, so that we would expect to find the exponent ν characterizing the time dependence of the spreading of vibrational wave packets to be 0.31 for 1 ray, either hydrated or not, which lies within statistical error of the value 0.29–0.30 found by direct calculation of vibrational energy flow in dehydrated and hydrated 1 ray.

IV. CONCLUDING REMARKS

We have examined here the influence of hydration on the anomalous diffusion of vibrational energy in protein molecules. Anomalous diffusion depends on two characteristic dimensions, the spectral dimension and the mass fractal dimension. We have compared here values for the mass fractal dimension of proteins computed with and without water molecules. Protein dynamics and relaxation processes depend on protein geometry, and since water buried inside the protein and near the protein surface couples dynamically to the protein [1–7], it is of interest to determine protein geometry with water present. We have compared the mass fractal dimension for proteins in a set of 200 molecules from the PDB ranging in size from 100 to 11 000 amino acids. The PDB contains the coordinates for water buried inside the protein and some water molecules in the first layer surrounding it. Including the water molecules from the PDB structure in the calculation of the mass fractal dimension only slightly increases its value from on average 2.49 when no water is included to 2.52 when it is.

Since the PDB contains less than one solvation shell, we surrounded the protein with water to carry out a calculation with hydration water present. We computed the mass fractal dimension of 20 protein molecules ranging in size from 100 to 4000 amino acids after running MD simulations of the protein in water. We found that by including water in the calculation the average mass fractal dimension for the 20 proteins studied is 2.87, a value that does not depend much on the size of the protein, and similar to what we find for large proteins, with more than 1000 amino acids, with or without water. Changes in protein structure due to partial unfolding leading to changes in R_G within 10% of that for the native structure does not change these dimensions much.

The influence of hydration on vibrational energy flow in proteins was also studied. Earlier simulations by us [22] on vibrational energy flow in dehydrated and hydrated cytochrome c revealed ν, the exponent characterizing anomalous diffusion, to be the same for both, but the origin of the similarity was not explored. Anomalous diffusion of vibrational energy depends on the ratio of the spectral dimension to the mass fractal dimension. The mass fractal dimension of a protein increases when the hydration water to which it is dynamically coupled is included in the calculation. The spectral dimension also increases, as we have seen for modest-sized proteins with about 200 to 250 amino acids, by about the same percentage as the mass fractal dimension, so that ν is essentially the same whether or not the protein is hydrated.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF Grant No. CHE-0512145) and by a Research Innovation Award from the Research Corporation.

[1] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, and F. G. Parak, Proc. Natl. Acad. Sci. U.S.A. **99**, 16047 (2002).
[2] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[3] F. Piazza, P. De Los Rios, and Y.-H. Sanejouand, Phys. Rev. Lett. **94**, 145502 (2002).
[4] A. R. Bizzarri and S. Cannistraro, J. Phys. Chem. B **106**, 6617 (2002).
[5] D. Russo, G. Hura, and T. Head-Gordon, Biophys. J. **86**, 1852 (2004).
[6] M. Tarek and D. J. Tobias, Phys. Rev. Lett. **88**, 138101 (2002).
[7] J. M. Gruschus and J. A. Ferretti, J. Biomol. NMR **20**, 111 (2001).
[8] T. G. Dewey, *Fractals in Molecular Biophysics* (Oxford University Press, New York, 1997).
[9] T. Nakayama, K. Yakubo, and R. L. Orbach, Rev. Mod. Phys. **66**, 381 (1994).
[10] H. Yang *et al.*, Science **302**, 262 (2003).
[11] S. C. Kuo and S. Xie, Phys. Rev. Lett. **93**, 180603 (2004).
[12] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Phys. Rev. Lett. **94**, 198302 (2005).
[13] R. Granek and J. Klafter, Phys. Rev. Lett. **95**, 098106 (2005).
[14] S. Alexander and R. Orbach, J. Phys. (France) Lett. **43**, L-625 (1982).
[15] H. J. Stapleton, J. P. Allen, C. P. Flynn, D. G. Stinson, and S. R. Kurtz, Phys. Rev. Lett. **45**, 1456 (1980).
[16] A. R. Drews, B. D. Thayer, H. J. Stapleton, G. C. Wagner, G. Giugliarelli, and S. Cannistraro, Biophys. J. **57**, 157 (1990).
[17] R. Elber, in *The Fractal Approach to Heterogeneous Chemistry*, edited by D. Avnir (John Wiley and Sons, New York, 1989), p. 345.
[18] R. Elber and M. Karplus, Phys. Rev. Lett. **56**, 394 (1986).
[19] R. Rammal and G. Toulouse, J. Phys. (France) Lett. **44**, L-13 (1983).
[20] R. Burioni, D. Cassi, F. Cecconi, and A. Vulpiani, Proteins: Struct., Funct., Bioinf. **55**, 529 (2004).
[21] M. B. Enright and D. M. Leitner, Phys. Rev. E **71**, 011912 (2005).
[22] X. Yu and D. M. Leitner, J. Chem. Phys. **119**, 12673 (2003).
[23] E. Lindahl, B. Hess, and D. van der Spoel, J. Mol. Model. **7**, 306 (2001).
[24] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. **91**, 43 (1995).
[25] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996).
[26] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Intermolecular Forces*, edited by B. Pullman (Reidel Publishing, Dordrecht, The Netherlands, 1981), pp. 331–342.
[27] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).

[28] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[29] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[30] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[31] V. Daggett and M. Levitt, Proc. Natl. Acad. Sci. U.S.A. **89**, 5142 (1992).