Universal low-frequency vibrations of proteins from a simple interaction potential

M. Nöllmann[1] and P. Etchegoin[1,2]
[1]*Instituto Balseiro, Universidad Nacional de Cuyo, 8400 San Carlos de Bariloche, Río Negro, Argentina*
[2]*Centro Atómico Bariloche, Comisión Nacional de Energía Atómica, 8400 San Carlos de Bariloche, Río Negro, Argentina*
(Recieved 4 January 1999; revised manuscript received 4 June 1999)

A pairwise Born potential connecting the heavy atom sites within a prescribed cutoff, and the equation of motion method (EOM), reproduce the existence of a universal singularity in the low-frequency vibrational density of states of typical globular proteins. This is due to quasilocalization of acoustic waves and an analogy with a similar feature found in glasses is stressed. We explain the dependence of this anomaly with the effective dimensionality of the protein. The EOM method allows for the study of even the largest proteins with a simple personal computer. [S1063-651X(99)04910-7]

PACS number(s): 87.15.By, 87.15.He, 87.15.Kg

The collective low-frequency dynamics of globular proteins is essential for a basic understanding of some of the functions they undertake in living matter. The glassylike nature of protein dynamics at low frequencies is a well established phenomenon [1]. Specific heat measurements of both, globular proteins and isolated secondary structures of polyaminoacids [2], reveal a characteristic glass anomaly [3] below $\sim 3-5$ K (i.e., below ~ 5 cm^{-1}), which can be accounted for by the presence of vibrational two-level systems (TLS's). The multiple conformational states that lead to a glassylike behavior have been confirmed by both molecular dynamics analysis [4] and optical spectroscopy [5]. The presence of many low-lying states not related by symmetry and with nearly the same energy seems to be the hallmark of several complex systems including glasses, proteins, neural networks, and spin glasses. Particular attention was given to the apparent *fractal dimension* of the low-frequency phonons and its influence on experimentally measured quantities such as electron-spin relaxations [6]. At slightly higher frequencies ($\sim 5-200$ cm^{-1}) the dynamics of proteins is harmonic and vibrations are collective molecular motions. These modes would be characterized in solid-state theory as *acoustic phonons*. The importance of collective ''*acoustic*'' motions triggered a considerable effort to develop normal-mode analysis methods that isolate and individualize these modes [7]. These methods confirmed that low-frequency phonons are properly described as backbone vibrations in dihedral-angle space; an approximation widely used in the literature [7].

Normal mode analysis [7,8], molecular dynamics simulations [7], effective medium theories [6], and experiments [9] agree in that the low-frequency dynamics of globular proteins is characterized by a peak in the vibrational density of states (DOS) around ~ 50 cm^{-1}. This energy corresponds to collective harmonic modes not related to the TLS's, but rather to the intrinsic disorder and effective dimensionality of these molecular assemblies. In this paper, we shall show how a simple calculation with realistic structures for the globular proteins and simplified interaction potentials reproduces this anomaly and gives further insight into its origin.

ben-Abraham [8] argued that the singularity in the DOS is *universal* and one of us [10] proposed an interpretation based on a similar anomaly (*boson-peak*) known for dielectric glasses. Moreover, a lognormal distribution was suggested for this feature [10], according to a conjecture originally put forward by Denisov *et al.* [11] to fit the boson-peak anomaly in glasses. In fact, a lognormal distribution fits the low frequency DOS of globular protein remarkably well in a wide variety of cases as shown in Fig. 1, where a collection of previous results are gathered. Besides, Tirion [12] pointed out that a simple random elastic network (REN) with pairwise Hookean interactions connecting sites within a prescribed cutoff can be used to model the low frequency harmonic dynamics of proteins below 10 cm^{-1}. Through a normal mode analysis, this method reproduced quite well the calculated temperature factors of the different residues in G-actin [12]. A similar approach has been used by Haliloglu *et al.* [13], who extended the normal mode analysis to intermediate energies ($\sim 2-100$ cm^{-1}) and obtained a result resembling the universal boson-peak of ben-Abraham [8], but with a nonvanishing DOS at $\omega \to 0$. So far, a clear-cut theoretical explanation of the physical origin of the boson peak was not given. Moreover, full cartesian normal-mode analysis is currently limited to proteins no larger than ~ 150 residues, since large Hessian matrices of order N^2 ($N=$ number of atoms) are needed. The simplified potential of Ref. [12] renders a substantial reduction in the stringent CPU memory requirements to solve the generalized eigenvalue problem.

We report calculations of the vibrational DOS based on a different implementation of the REN model. Rather than applying a normal-mode analysis, we use the equation of motion (EOM) method. Starting with an initial perturbation $[\vec{r}_i(0)]$ at a randomly chosen site i, the time evolution of the system is numerically evaluated. From $\vec{r}_i(t)$, the Green function of the problem is calculated, and from the imaginary part of the Laplace transform the local DOS is evaluated [3,14–16]. Frequencies below ~ 5 cm^{-1}, are poorly represented with the EOM method. An average of 50 local DOS's at randomly chosen sites in the interior of the protein structure is taken as representative of the full DOS. The EOM method requires the storage of positions and velocities for the N atomic sites in the computation. Thus, the CPU memory requirements scale linearly with N, allowing for the calculation of very large proteins on relatively modest computers. The price to be paid is the total time needed for a

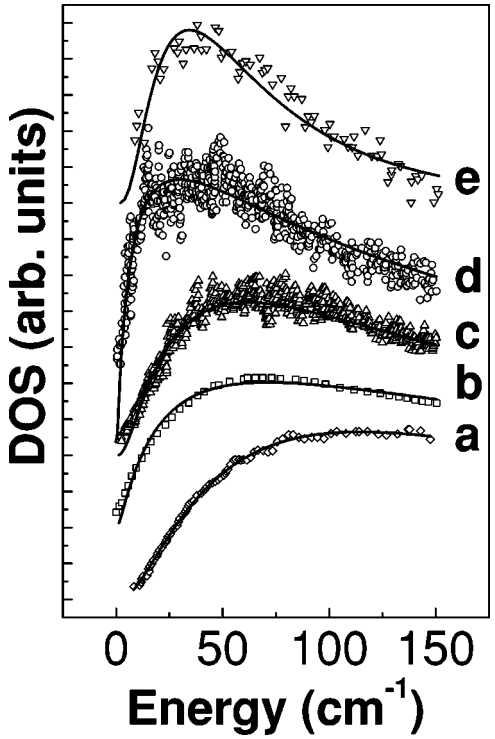

FIG. 1. Low-frequency DOS obtained from (a) neutron scattering in bovine pancreatic trypsin inhibitor [9] (5pti); (b) molecular dynamics of 5pti including damping of the modes [6]; (c) 100-ps molecular dynamics simulation of 5pti in water [6]; (d) 100-ps molecular dynamics of 5pti in vacuum [6]; and (e) normal mode analysis of lysozyme (6lyz), ribonuclease (5rsa), 5pti, and crambin (1crn) [8]. Curves are vertically displaced for clarity. Solid lines are lognormal fits to the data.

reliable representation of the lowest frequencies. Typical protein structures with ~ 2000 atoms may easily take $\sim 15-20$ h on a Pentium II processor. The method does not give information on the eigenvectors of the system [3,14–16].

The important results we report are (i) the EOM method together with a pairwise Born potential leads to a singularity in the low-frequency DOS of typical globular proteins with all the characteristics of the boson-peak anomaly; (ii) the energy position of the boson peak depends on the connectivity and, accordingly, on the effective dimensionality (d_s) of the REN-model; (iii) the different DOS's in the limit $\omega \to 0$ reported in the literature can be easily explained in terms of the particular connectivities of the different networks; and (iv) a simple model can be used to explain the dependence of the peak with d_s.

The calculation is implemented by retrieving the structures of several typical globular proteins from the Brookheaven protein data bank (PDB). Hydrogens are ignored for the computation and, besides, the average mass of the atoms in the protein is assigned to each atomic position. The collective elastic properties depend fundamentally on the topology and connectivity of the underlying network of atomic positions and, therefore, mass differences among sites are irrelevant. Each site is connected with its nearest neighbors up to a distance defined by a prescribed cutoff (hereafter R_C). Once the connectivity has been established, the interaction among sites is modeled by a pairwise Born potential of the form

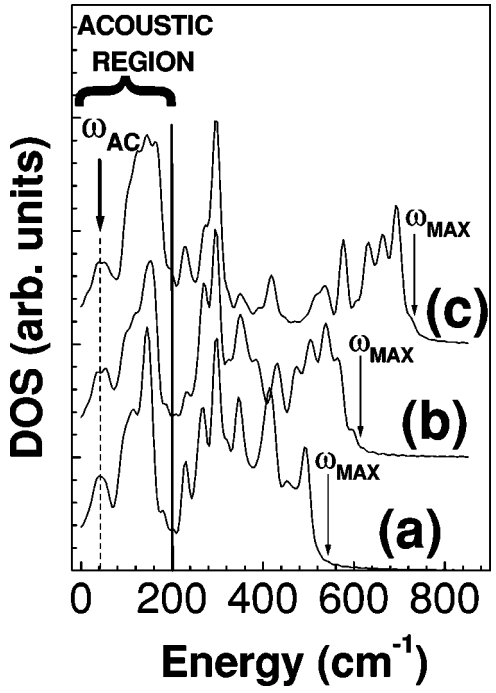

FIG. 2. DOS in crambin (1crm) for different values of A and B [see Eq. (1)] and $R_C = 1.9$ Å. The curves were vertically shifted for clarity. The different results correspond to (a) $A/B=1$, (b) $A/B=2$, and (c) $A/B=4$, while B is chosen to produce ω_{AC} at ~ 30 cm^{-1}. The ω_{AC} peak is the low-energy boson-peak anomaly present in this structure.

$$U_{ij} = A\left[(\vec{\mu}_i - \vec{\mu}_j)\cdot\left(\frac{\vec{r}_{ij}}{|\vec{r}_{ij}|}\right)\right]^2 + B|\vec{\mu}_i - \vec{\mu}_j|^2, \qquad (1)$$

where $\vec{r}_{ij}$ is the distance between atoms i and j, $\vec{\mu}_i$ ($\vec{\mu}_j$) is the displacement from equilibrium of the atom $i(j)$, and A,B are the bond stretching and bending potential parameters, respectively. The presence of two parameters in Eq. (1) gives more flexibility and allows for a clear separation between *acousticlike* and high frequency localized modes. This can be appreciated in Fig. 2, where different calculations of the DOS versus energy are shown for crambin (1crn) with a fixed cutoff (R_C) and different potential parameters. By changing the ratio A/B we note that (i) an *acoustic region* where the DOS is essentially the same exists; (ii) the maximum attainable frequency (ω_{MAX}) shifts with increasing A/B; and, finally, (iii) there is a low-frequency peak (ω_{AC}) at ~ 30 cm^{-1}. If A/B is increased, the low-frequency DOS is dominated by acoustic-bond-bending-like phonons, while higher frequency modes (in particular, ω_{MAX}) depend on A and are essentially bond stretching in character. By having two parameters in Eq. (1), the spectral weight of high-frequency localized modes can be pushed away from the *acoustic region* through a pertinent choice. The peak labeled as ω_{AC} in Fig. 2 consistently appears in all studied protein structures. The position of this peak (as well as its low-frequency tail towards $\omega \to 0$) depends on the chosen R_C. In addition, the peak is fairly well reproduced by a lognormal distribution and has all the characteristics of the boson-peak anomaly. The dependence of ω_{AC} with d_s will be studied later but, before that, we evaluate the average number of

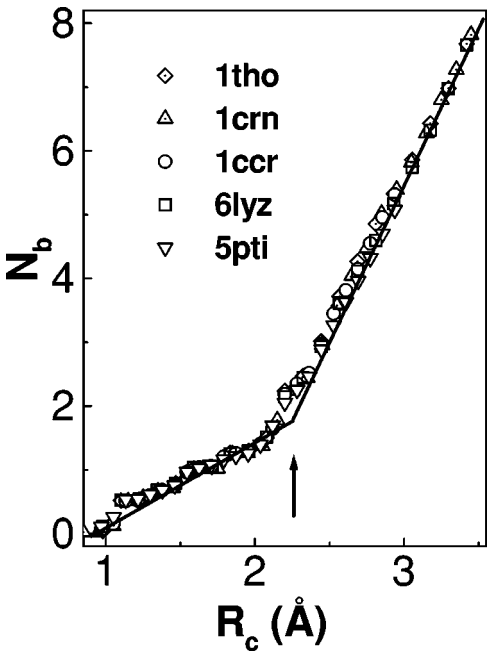

FIG. 3. Average number of connections per atom (N_b) vs the cutoff distance (R_C) as calculated for thioredoxin mutant (1tho) (diamonds), crambin (1crn) (triangles), rice ferrytochrome C (1ccr) (circles), lysozyme (6lyz) (squares), and bobine pancreatic trypsin inhibitor (5pti) (inverted triangles). The solid line is a guide to the eye and shows a clear crossover between two distinct regions at $R_C \sim 2.25$ Å.

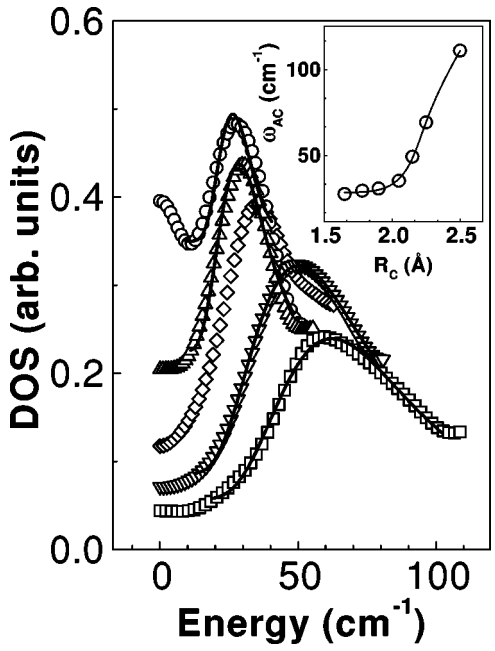

FIG. 4. DOS's in thioredoxin mutant (1tho) at low frequencies for different values of R_C and $A/B=2$. Symbols pertain to $R_C=1.65$ (circles), 1.9 (triangles), 2.05 (diamonds), 2.15 (inverted triangles), and 2.25 (squares) Å. Solid lines correspond to lognormal fits performed on ω_{AC}. The inset shows the shift of ω_{AC} vs R_C with the analytic model proposed in the text (solid line). See text for further details.

connections per atom in the network (N_b) versus R_C. This is relevant to define the connectivity and the d_s of the network. The results for several typical globular proteins are displayed in Fig. 3. The main results can be summarized as follows: (i) the curve seems to be universal, (ii) two clear regions below and above $R_C \sim 2.25$ Å can be observed and, (iii) there are small substructures for $R_C < 2.25$ Å. These results are not surprising since they render the same information of the average universal structure factor $S(q)$ in Ref. [10]. Their interpretation is as follows: below $R_C \sim 2.25$ Å, the main backbone connections along the polypeptide chain are established and the network behaves essentially as a one-dimensional (1D) object. The importance of typical chemical bond distances with nearest neighbors is observed as small substructures in both N_b and $S(q)$ [10]. Slightly above 2.25 Å, connections among the different secondary structures of the protein appear. There is a continuos crossover in d_s between one and two dimensions (2D) at $R_C \sim 2.25$ Å. We expect the best representation of the dynamics of real proteins with this simplified potential to occur for $R_C \sim 2.25$ Å, since real proteins behave approximately as objects with $d_s \sim 2$ [8,12]. At much larger distances ($R_c \gg 2.25$ Å), connections among tertiary structures would eventually be formed, creating a three-dimensional (3D) network. In Fig. 4 we now turn to the calculation of different low-frequency DOS's for fixed R_C's around 2.25 Å. Note that (i) the boson peak (ω_{AC}) appears in all cases and it shifts for different R_C's, and (ii) the DOS's show a low energy tail towards $\omega \to 0$, which disappears for increasing R_C's. The latter can be explained in terms of the residual 1D characer of the network. Below $R_C = 2.25$ Å, the network retains large portions of the original backbone which behave as a 1D object. Since the DOS for *d*-dimensional harmonic solid is $\propto \omega^{(d-1)}$, any rest of a 1D structure will produce a nonvanishing DOS for $\omega \to 0$. Note that the tail in Fig. 4 is abruptly reduced when the network passes through the dimensional crossover at 2.25 Å. It is important to realize that the effective dimension d_s is an average *topological* property. In a real fractal, an object with $1 < d_s < 2$ has neither pure 1D nor 2D objects inside; the DOS is in this case $\propto \omega^{(d_s-1)}$ for $\omega \to 0$. A real protein, however, is better described as a collection of coupled 1D objects forming a network with slight 2D character. In this case, d_s can also be $1 < d_s < 2$, but there is a partial separation of the intrinsic dynamics of the different substructures and the DOS will be dominated at $\omega \to 0$ by the objects with the lowest dimension. Figure 4 also shows lognormal fits to the boson-peak anomalies. Albeit without demonstration, the lognormal distribution conjecture [10,11] seems to be extremely successful in reproducing the boson-peak anomaly in calculations and experiments, even when the peak is less pronounced (see Fig. 1).

Finally, we briefly show how the R_C dependence of the boson peak in Fig. 4 can be understood. For *acoustic* waves we expect $\omega_{AC} \propto k \propto 1/\lambda$. If $\lambda \sim L$, where L is the *localization length* due to disorder, a singularity in the DOS can be expected. From the theory of fractons [17], L depends on ω_{AC} through $L(\omega_{AC}) \propto 1/\omega_{AC}^{d_s/3}$. Thus, $\omega_{AC} = \omega_0 \exp[-C/(d_s/3 - 1)]$, where ω_0 and C are constants.

Up until now, we have made some approximations as to obtain a reliable relation between the acoustic peak frequency and the effective network dimensionality d_s. Addi-

tionally, in order to understand the relation between ω_{AC} and the cutoff distance we must model the interdependence between the latter and the effective dimensionality, which is less trivial and depends on the topological aspects of the network. In order to explicitly demonstrate this, consider a linear chain with fixed interatomic distance a. The effective dimensionality of the chain is zero until $R_c=a$, where full connectivity is achieved and a steplike change to $d_s=1$ is accomplished. This abrupt transition is smeared out if disorder is introduced in the chain. Similarly, a perfect rectangular two-dimensional network with lattice parameters a and b will show two steplike effective dimensionality changes as a function of R_c. Once again, these transitions are smeared out when moderate disorder is introduced in the lattice. In a general case, numerical simulations must be used to evaluate the relation between d_s and R_c. For the particular case of a linear disordered chain, numerical simulations [18] show that an excellent analytic ansatz for the crossover between two given dimensions is given by $d_s(R_C)=d_1\tanh[(R_C-R_0)/\Delta]+d_2$, where R_0 is the *average cutoff distance* where the transition takes place, (d_2-d_1) and (d_2+d_1) are the upper and lower dimensions, respectively, and Δ a measure of the disorder. We believe that this ansatz is a good approximation in general for arbitrary lattices. In this manner, the function $\omega_{AC}[d_s(R_C)]$ can be compared with the calculation. The inset of Fig. 4 shows ω_{AC} as a function of R_C together with a fit where we used Δ, C, and ω_0 as fitting parameters, and assumed a crossover from 1D to 2D at $R_0=2.25$ Å. It is quite clear that the a simple model of this sort can very well account for the dependence of ω_{AC} on R_C. A more complete theory, however, should also render the presumed lognormal lineshape of the peak.

In closing, we showed that the boson-peak anomaly in real protein structures with simplified interactions (Born potential + REN model) can be easily reproduced. The EOM method has modest memory requirements which scale linearly with N, allowing for the study of large proteins in modest computers. The boson peak is interpreted as quasilocalization of acoustic waves in a fractallike structure and depends on d_s. A lognormal distribution seems to repeatedly produce excellent fits for the boson-peak, supporting the conjecture put forward in Refs. [10,11]. The nonvanishing DOS at $\omega\rightarrow 0$ in Ref. [13] is easily explained by the residual number of 1D-like structures for the specific chosen connectivity. Our results are similar to those of [13] for $R_C<2.25$ Å, which stands as a crossover between 1D- and 2D-like structures. The boson-peak details do depend on the specific R_C chosen, unlike the lowest frequencies (<10 cm^{-1}), which were shown in Ref. [12] to be somewhat independent for sufficiently large R_C's. The universality of the anomaly in Ref. [8] is naturally explained by the statistical structural similarities of different proteins [10]. Finally, inasmuch as the protein is treated as a disordered nanocrystal, this model is also relevant for the low frequency dynamics of glasses and, in particular, the analogy with the boson-peak anomaly is put on a specific formal ground.

We are indebted to A. A. Aligia for a critical reading of the manuscript.

[1] I.E.T. Iben *et al.*, Phys. Rev. Lett. **62**, 1916 (1989); F.H. Stillinger and T.A. Weber, Phys. Rev. A **25**, 978 (1982); D.L. Stein, Proc. Natl. Acad. Sci. USA **82**, 3670 (1985).
[2] G.P. Singh *et al.*, Z. Phys. B **55**, 23 (1984); L. Finegold and J. Cude, Nature (London) **238**, 38 (1972).
[3] N.E. Cusack, *The Physics of Structurally Disordered Matter* (Hilger, Bristol, 1987).
[4] R. Elber and M. Karplus, Science **235**, 318 (1987).
[5] C.W. Rella *et al.*, Phys. Rev. Lett. **77**, 1648 (1996); D. Thorn Leeson and D.A. Wiersma, *ibid.* **74**, 2138 (1995); D. Thorn Leeson *et al.*, J. Phys. Chem. **98**, 3913 (1994); W. Köhler and J. Friedrich, J. Chem. Phys. **90**, 1270 (1989).
[6] R. Elber and M. Karplus, Phys. Rev. Lett. **56**, 394 (1986).
[7] S. Hayward and N. Go, Annu. Rev. Phys. Chem. **46**, 223 (1995).
[8] D. ben-Avraham, Phys. Rev. B **47**, 14 559 (1993).
[9] J. Smith *et al.*, J. Chem. Phys. **93**, 2974 (1990).
[10] P. Etchegoin, Phys. Rev. E **58**, 845 (1998).
[11] Yu.V. Denisov and A.P. Rylev, Pis'ma Zh. Éksp. Teor. Fiz. **52**, 1017 (1990) [JETP Lett. **52**, 411 (1990)].
[12] M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[13] T. Haliloglu *et al.*, Phys. Rev. Lett. **79**, 3090 (1997).
[14] S. Alexander and R. Orbach, J. Phys. (France) Lett. **43**, 625 (1982).
[15] S. Alexander *et al.*, Rev. Mod. Phys. **53**, 175 (1981).
[16] A. MacKinnon, in *The Recursion Method and Its Applications*, edited by D.G. Pettifor and D. Weaire (Springer, Berlin, 1984).
[17] D. Staufer and A. Aharony, *Introduction to Percolation Theory*, 2nd ed. (Taylor & Francis, London, 1994), p. 126.
[18] M. Nöllmann and P. Etchegoin (unpublished).