Glassylike low-frequency dynamics of globular proteins

P. Etchegoin*
Centro Atómico Bariloche & Instituto Balseiro, Comisión Nacional de Energía Atómica & Universidad Nacional de Cuyo, 8400-San Carlos de Bariloche, Río Negro, Argentina
(Received 11 December 1997)

It is shown that globular proteins display a universal average structure factor accounting for their common aminoacid base. This property, linked to the construction of a random elastic network to study the low-frequency dynamics, leads to striking similarities between the physics of globular proteins and that of glassy or amorphous nanocrystals. A few consequences of this interpretation are inquired into, particularly, in connection with some of the universal properties of the low-frequency dynamics that were recently discovered for globular proteins. [S1063-651X(98)14606-8]

PACS number(s): 87.15.By, 87.15.He, 61.43.Bn, 61.43.Fs

Globular proteins constitute an important type of biological macromolecule with a myriad of functions that are essential to life. The dynamics of these macromolecules is of prime importance to the understanding of some of the functions that proteins undertake in living organisms. The latter has received a considerable amount of attention in the past, from both theoretical [1] and experimental [1,2] points of view.

Low-frequency vibrations ($\omega \leqslant 200\ \mathrm{cm}^{-1}$) in globular proteins are related to large-scale collective motions, and were observed experimentally by means of Raman and infrared spectroscopy several years ago [3]. The exact nature of these modes, however, remained unclear for many years [1]. A renewed interest emerged recently in the low energy dynamics of proteins, on account of the possible links between these molecular motions and some of their biological functions, in particular protein folding and catalytic functions of enzymes. Two recent breakthroughs contributed principally to our understanding of the low-frequency dynamics of proteins. First, ben-Avraham [4] made the very important observation that the low energy density of states of several proteins that range from 39 to 375 aminoacids fall into a *universal curve*. Moreover, it was argued [4] that the density of states $g(\omega)$ for the lowest modes below $\sim 10\ \mathrm{cm}^{-1}$ was very well approximated by $g(\omega) \sim \omega$, contrary to the expectations for a three-dimensional harmonic Debye solid. This peculiarity was attributed to the intrinsic fractal nature of protein structures, which make them behave as two-dimensional objects in this energy range. On the other hand, Tirion [5] showed recently that the low energy dynamics of globular proteins can be understood by assuming some sort of *random elastic network* (REN) shaped by the underlying structure of carbons and simple pairwise Hookean forces among cites. Some further progress in this direction was made recently in Ref. [6]. The basic idea underlying the REN model is that the exact structure of the chemical bond surrounding a given site is not important. It is well known from semiempirical molecular dynamics methods [7] that a given site in a macromolecule interacts not only with its near neighbors through the stretching of the chemical bond, bond angle, and dihedral angle torsions, but also with distant atoms of the molecular structure through the van der Waals and electrostatic interactions. The idea of the REN model is to replace all these interactions by Hookean forces with a single adjustable parameter, regardless of their origin, in the hope that for long wavelengths the detailed nature of the microscopic interactions will not be relevant, and that these interactions will be ruled by some sort of central limit theorem. The model assumes a global minimum for the total energy from the starting crystallographic data, and circumvents, in that manner, the costly energy minimization regularly carried through in molecular mechanics methods. Two sites within the structure are connected through a Hookean spring if and only if they are separated by a distance which is smaller than, or equal to, a prescribed cutoff. The random elastic network constructed in this manner is a disordered structure, and the connection with the physics of structurally disordered matter is self-evident. The low-frequency dynamics of these models seem to reproduce [5] the observed dependence $g(\omega) \sim \omega$ at low frequencies proposed by ben-Avraham [4], as well as the temperature factors derived from x rays for the different residues of the proteins.

In this paper, we shall show that there are striking analogies between the physics of globular proteins (seen as random elastic networks) and the physics of amorphous and glassy solids that can help us to interpret some of the features of the low-frequency dynamics. By assuming the REN model as correct, and taking into account that all low-frequency properties of globular proteins seem to fall into a universal behavior, we obtain the obvious conclusion that globular proteins ought to be structurally alike in the statistical sense. This conclusion, which is more or less intuitive, can be put on a more rigorous ground by calculating the average structure factors $S(\vec{q})$ for several proteins of different sizes. We are interested in the heavy atoms of the protein structure only, since none of the hydrogens participate in the low-frequency dynamics. The structure factor is defined as

$$S(\vec{q}) = \frac{1}{N}\left|\sum_{i=1}^{N} \exp(i\vec{q}\cdot\vec{R}_i)\right|^2, \qquad (1)$$

where N is the number of heavy atoms in the protein, and $\vec{R}_i$ their coordinates. The structure factor has the limits

*Electronic address: etchegoi@cab.cnea.edu.ar

 © 1998 The American Physical Society

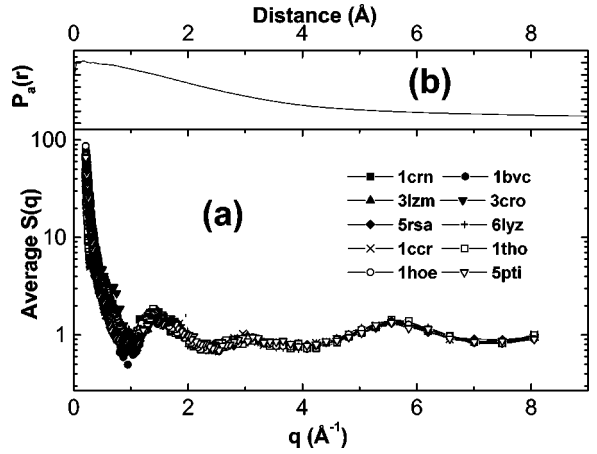

FIG. 1. (a) Average structure factors for ten different globular proteins. The labels correspond to the PDB codes of the proteins. Despite some small scatter, the curves look remarkably universal for this set of proteins. In (b), we display the atomic position autocorrelation function $P_a(r)$ calculated from the data in (a) using the average value of the ten proteins. Note that by $r\sim 6$ Å , $P_a(r)$ has fallen to a 25% of the value at $r\sim 0$.

$S(\vec{q}\rightarrow 0)\rightarrow N$ and $S(\vec{q}\rightarrow\infty)\rightarrow 1$. Equation (1) defines the structure factor as a function of a vector $\vec{q}$, i.e., it defines a scalar function $S(\vec{q})$ for each vector $\vec{q}$ in three dimensions. Inasmuch as we are interested in averaged quantities, however, we define the effective structure factor $\overline{S(\vec{q})}=S(|\vec{q}|)$, which is nothing but an average over the entire solid angle Ω of $S(\vec{q})$, with $|\vec{q}|=\text{const}$.

In Fig. 1(a), we show the averaged structure factors calculated for ten typical globular proteins. The structures of the proteins were taken from the Brookhaven Protein Data Bank [8] (PDB), and conveniently transformed for the calculation. The averaged structure factors were calculated for the following proteins (the code after the name corresponds to the PDB code): Crambin (1crn), bovine pancreatic trypsin inhibitor (5pti), α-amylase inhibitor (1hoe), *E. coli* thioredoxin (1tho), Rice ferrytochrome *c* (1ccr), ribonuclease A (5rsa), lysozyme (6lyz), biliverdin apomyoglobin complex (D) (1bvc), T4 lysozyme (3lzm), and the gene regulating protein (3cro). These globular proteins range from 652 (1crn) to 3530 (3cro) atoms. The calculation is performed by averaging Eq. (1) over a set of 10^4 randomly selected directions for each $|\vec{q}|$. As can be readily appreciated from Fig. 1(a), the average structure factors of these proteins look remarkably similar. From the average $S(|\vec{q}|)$ of the ten proteins displayed in Fig. 1(a), we obtain the autocorrelation function of the atomic positions [9] $P_a(r)$ in Fig. 1(b) which is the amplitude of the Fourier transform of $S(|\vec{q}|)$. The data in Fig. 1(a) show three clear peaks about $\vec{q}\sim$ 5.5, 3, and 1.5 Å^{-1}. The first one is a manifestation of short range order produced by the nearest neighbors among heavy atoms which range from 1.2 to 1.5 Å for typical C–C, N–C, C–O, etc., bonds in proteins. The other two peaks, accordingly, speak for intermediate range order in the range 2–6 Å . The fact that they appear very often in different globular proteins is not surprising if we take into account that proteins are built up from a common aminoacid base; i.e., the peaks in $S(|\vec{q}|)$ for $|\vec{q}|>1$ Å^{-1} tell us that the same types of structures with the same kind of coordination appear very often in all these proteins. Conversely, $S(|\vec{q}|)$ displays a continuous increase for $|\vec{q}|<1$ Å^{-1} which indicates that, on the average, globular proteins look like a *quasicontinuum* for distances $d>6-10$ Å ; a result which can also be readily observed in the atomic position autocorrelation function $P_a(r)$ in Fig. 1(b) which drops to ~ 25 % with respect to the value a short distances for $|\vec{q}|\sim 6$ Å.

We concentrate for the time being on the effect of the structural similarities among globular proteins, and their consequences for the REN model, whereupon we note that the universal behavior observed for the low-frequency excitations (where the REN model is valid) is nothing but a natural consequence of the universal nature of $S(|\vec{q}|)$ displayed in Fig. 1(a). Likewise, the fact that proteins display short and intermediate range orders makes the random elastic network constructed upon the coordinates of the heavy atoms of the protein similar to that of a glass or amorphous material.

Glasses form a subset of amorphous solids [9] in which the ground state properties are dominated by metastable configurations with tunneling modes [9]. The low temperature properties of glasses are, as a consequence, quite peculiar, and are dominated by these tunneling modes which bear a specific heat C of the form $C\sim T$ instead of the conventional $C\sim T^3$ expected for a Debye solid. This behavior is observed at very low temperatures on the order of ~ 1 K in insulating glasses, and, consequently, only phonon energies in the range ~ 1 cm^{-1} are relevant in this range. With these provisos in mind, glasses and amorphous solids look essentially identical in their short, intermediate, and long range order structural properties. In the range $\sim 20-200$ cm^{-1}, amorphous and glasses present an *excess* of modes as seen in the vibrational density of states (VDOS) obtained through Raman scattering, infrared absorption, and neutron scattering [10]. This peak was called the *boson peak* [9], and its origin remained unclear for many years until it became increasingly evident that it was related to the presence of intermediate range order in these structures. Several models have been suggested to understand this anomaly; among them we find the phonon-fracton model of Alexander and Orbach [11], and the log-normal distribution of frequencies [12]. In fact, Denisov and Rylev [12] proposed in 1990 that the low-frequency vibrational density of states of glasses follows a log-normal distribution of the form

$$g(\omega)\sim\frac{1}{\sigma\omega}\exp\left(-\frac{(\ln\omega-\mu)^2}{2\sigma^2}\right). \tag{2}$$

This conjecture was tested by several authors in subsequent years [13].

In view of the structural similarities between glasses, amorphous materials, and globular proteins (at least in the framework of the REN model, which seems to represent the low-frequency properties quite properly), we may wonder whether the universal curve proposed by ben-Avraham [4] is the equivalent of the boson peak for glasses. In Fig. 2(a), we show the data points used by ben-Avraham to propose the universal low-frequency density of states of globular proteins [4]. The data points are a collection for five different pro-

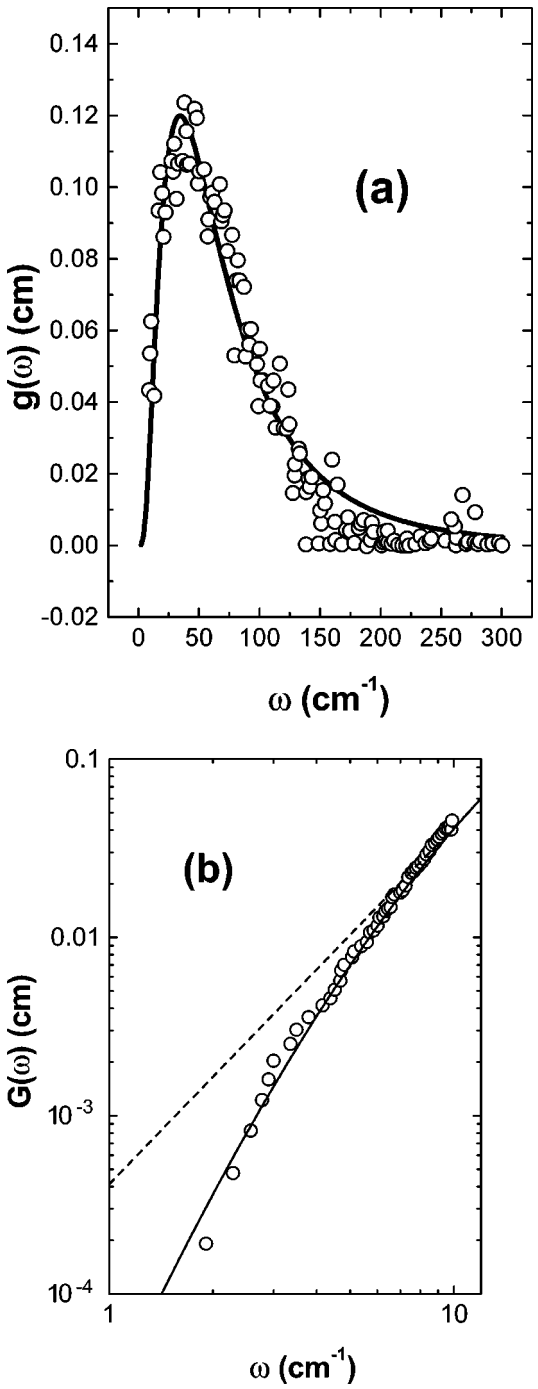

FIG. 2. (a) Density of vibrational normal modes from Ref. [4] taken from five proteins: crambin, bovine pancreatic trypsin inhibitor, ribonuclease A, lysozyme, and g-actin. The solid line is a fit with a log-normal distribution function [Eq. (2)]. (b) Log-log plot of the integrated number of vibrational states below ω as a function of ω for g-actin taken from Ref. [4]. The dashed line is a fit with pure quadratic function. The solid curve is a fit with an integrated log-normal distribution. See text for further details.

teins. The solid line is a fit with a log-normal distribution like Eq. (2). There is considerable scatter around 150 cm^{-1} in the data points, which is not surprising, taking into account that some of them belong to small globular proteins like Crambin. It is naturally expected that the smaller the size of the protein the larger the fluctuations with respect to universal or averaged properties. Notwithstanding, it is quite clear form Fig. 2(a) that this excess of modes or *universal peak* can be interpreted as represented by a log-normal distribution. Note that the fit involves only two parameters in Eq. (2).

Furthermore, in Fig. 2(b) we show the fraction of normal modes below frequency ω, $G(\omega)$, as a function of ω obtained from the slowest 70 modes of g-actin from Ref. [4]. In this region, ben-Avraham proposed $G(\omega)\sim\omega^2$, showing an anomalous dispersion. The dashed line in Fig. 2(b) is a fit with a function of the form $G(\omega)=A\omega^2$, where A is a constant. The fit tries to reproduce mainly the data points with the largest abscissas, to minimize most effectively the overall differences with the data. In addition, the fit is forced to satisfy a pure quadratic function which also implies the necessary condition $G(\omega=0)=0$. The result is that the fit follows reasonably well the data points above ~ 5 cm^{-1}, with the outlay that points below ~ 5 cm^{-1} are poorly reproduced. This difference is, in fact, not very dramatic when looked at on a linear plot, but it is greatly amplified when examined on a log-log plot as in Fig. 2(b). Conversely, the solid line in Fig. 2(b) is a fit with an integrated log-normal distribution. We observe that the low-frequency density of states is much better reproduced in this case and, in particular, the downward shift of the curve on a log-log plot for frequencies below ~ 4 cm^{-1} is closely followed.

In the energy range of vibrational modes where the log-normal distribution holds, it is, in fact, irrelevant whether the protein itself behaves like a glass or an amorphous material at low temperatures. The only relevant characteristic is the presence of short and intermediate range orders, and, accordingly, the absence of long range correlations. In the range 2–10 cm^{-1}, we showed that the log-normal distribution of modes still holds for g-actin. At even lower energies, however, the situation may change, and the real attributes of the total energy of the protein can be ascertained. An amorphous solid, which presents disorder but is in a global minimum of the total energy, should follow a Debye law at very low temperatures, i.e., for sufficiently long wavelengths the solid cannot discern the disorder, and a conventional behavior is recovered. If a globular protein were in a global minimum of the total energy, there would be a gap in the vibrational excitations which is related to a finite size effect; the minimum eigenvalue of the dynamical matrix of the protein should roughly scale with the size of the protein l as $\omega \sim 1/l$. This would be the prediction of the REN model, which assumes from the start a global minimum in the total energy. In any case, this is most unlikely to be the case in real proteins. The fact that the initial energy minimization is difficult in conventional molecular mechanics methods applied to proteins is nothing but a manifestation that the molecules very often have several possible configurations which are very close in energy and separated by moderate barriers. An exploration of all possible available configurations for a medium size protein is a formidable and time consuming task. We believe that, at sufficiently low wavelengths, the density of vibrational states in a globular protein should be governed by tunneling states among the possible local minima of the total energy and that the real *glassy* nature of the atomic arrangement should be revealed. There should be, accordingly, a crossover to a constant VDOS at adequately low energies.

In conclusion, we have shown that globular proteins have a universal average structure factor, and that they exhibit short and intermediate range orders accounting for their common aminoacid base. The construction of a random elastic network using these structures leads immediately to a model that resembles a nanocrystal (typical sizes in the range $\sim 30-150$ Å) of a glassy or amorphous material. The latter do show an anomaly in the vibrational density of states at low frequencies, known as a *boson peak*, which is reasonably well described by a log-normal distribution [12]. We propose the interpretation that the boson peak for proteins is, in fact, the *universal curve* proposed by ben-Avraham [4] for the low-frequency VDOS. We have shown that a two parameter fit with a log-normal distribution can successfully reproduce the overall shape of the *universal curve*, as well as the fine details of the integrated density of states $G(\omega)$ for low frequencies. Finally, we predict that, at sufficiently low energies, the REN model should break down, and the real density of tunneling states should take over. This would produce a VDOS of the form $g(\omega) \sim$ const, and the protein could be interpreted as a glassy nanocrystal. It would be very interesting, in our opinion, to perform specific-heat measurements of globular proteins at very low temperatures (<1 K) to clarify this particular aspect of the dynamics. These experiments have not been performed to date to the very best of our knowledge.

Thanks are due to A. Fainstein and R. G. Pregliasco for general support and stimulating discussions.

[1] Charles L. Brooks III, Martin Karplus, and B. Montgomery Pettitt, *Proteins, A Theoretical Perspective of Dynamics, Structure and Thermodynamics* (Wiley Interscience, New York, 1988); J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, Cambridge, 1987).

[2] An account of early experimental studies as well as the description of the experimental techniques can be found M. V. Volkenstein, *Molecular Biophysics* (Academic, New York, 1977).

[3] L. Genzel, F. Keilmann, T. P. Martin, G. Winterling, Y. Yacoby, H. Fröhlich, and W. Makinen, Biopolymers **15**, 213 (1976); M. Ataka and S. Tanaka, *ibid.* **18**, 507 (1979).

[4] Daniel ben-Avraham, Phys. Rev. B **47**, 14 559 (1993).

[5] Monique M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).

[6] Turkan Haliloglu, Ivet Bahar, and Burak Erman, Phys. Rev. Lett. **79**, 3090 (1997).

[7] M. Levitt, J. Mol. Biol. **168**, 595 (1983).

[8] The data can be retrieved from the ftp node *ftp.pdb.bnl.gov*

[9] N. E. Cusack, *The Physics of Structurally Disordered Matter* (Hilger, Bristol, 1987).

[10] *Amorphous Solids, Low Temperature Properties*, edited by W. A. Phillips (Springer, Berlin, 1981).

[11] S. A. Alexander and R. Orbach, J. Phys. (France) Lett. **43**, 625 (1982).

[12] Yu V. Denisov and A. P. Rylev, Pis'ma Zh. Eksp. Teor. Fiz. **52**, 1017 (1990) [JETP Lett. **52**, 411 (1990)]; V. K. Malinovsky, A. P. Sokolov, and V. N. Novikov, Phys. Lett. A **153**, 63 (1991).

[13] A. Feher, I. M. Yurkiin, L. I. Deich, M. Orendáč̃, and I. D. Turyanitsa, Physica B **194-196**, 395 (1994); R. T. Phillips and M. K. Ellis, J. Non-Cryst. Solids **164-166**, 135 (1993).