Glasslike Structure of Globular Proteins and the Boson Peak

Stefano Ciliberti,1 Paolo De Los Rios,2 and Francesco Piazza2

1Laboratoire de Physique Théorique et Modèles Statistiques, Université de Paris-Sud, bâtiment 100, 91405, Orsay Cedex, France
2Laboratoire de Biophysique Statistique-IPT, EPFL, CH-1015 Lausanne, Switzerland

(Received 23 November 2005; published 19 May 2006)

Vibrational spectra of proteins and topologically disordered solids display a common anomaly at low frequencies, known as boson peak. We show that such feature in globular proteins can be deciphered in terms of an energy landscape picture, as it is for glassy systems. Exploiting the tools of Euclidean random matrix theory, we clarify the physical origin of such anomaly in terms of a mechanical instability of the system. As a natural explanation, we argue that such instability is relevant for proteins in order for their molecular functions to be optimally rooted in their structures.

DOI: 10.1103/PhysRevLett.96.198103 PACS numbers: 87.15.He, 63.50.+x, 64.70.Pf

Proteins are characterized by mechanically stable, unique native structures that bear a precise relation with their biological functions. Yet, in most cases, specific functionality is accompanied by large-amplitude dynamical conformational changes that require high flexibility [1]. Protein structures are complex, hierarchical ones, characterized by short-range order and overall spatial correlations that bear strong similarities with those of glassy materials [2]. In actual fact, proteins and glasses share many physical properties, such as peculiar relaxation processes [3] and the occurrence of a dynamical transition as revealed by the temperature dependence of the atomic mean square displacements (MSD) [1,4,5].

Interestingly, there exists a remarkable similarity of the Raman and neutron-scattering spectra of proteins with those of glasses and supercooled liquids [4], i.e., a peak that develops at low temperatures in the low-frequency regions. Such anomaly, known as boson peak (BP), also shows up in the experimentally determined density of states when divided by the Debye law, i.e., $g(\omega)/\omega^2$ [6]. Several models have been proposed for the explanation of the BP in proteins, among which the phonon-fracton model [7], and the log-normal distribution model [8].

The BP is, on the other hand, a universal feature of many glassy systems [9]. In this context, several possible explanations have been proposed, from the two-level system scenario [10] to localized modes arising from a strong scattering of the phonons by the disorder [11], from “glassy” van Hove singularities [12] to a mechanical instability [13]. Recently, the possibility that a BP may be a general feature of weakly connected systems has also been investigated [14,15].

In a different analytical framework [16], the excess of low-energy modes with respect to the Debye behavior is viewed as a symptomatic effect of the topological phase transition which is conjectured to happen in glasses at low temperatures [13]. Recently, a quantitative description of the BP phenomenology has been given within the formalism of the Euclidean Random Matrix (ERM) theory [16], whose predictions have been confirmed by numerical simulations on realistic glass-forming systems, emphasizing its universal character [17].

In this Letter, we show that the emergence of a BP in globular proteins is the signature of a structural instability of the saddle-phonon kind akin to that predicted within the ERM theory of glasses. Remarkably, our explanation allows for a natural interpretation of such instability in proteins in terms of the mutual relations among their structure, dynamics, and biological function.

To investigate the vibrational properties of a given globular protein, we coarse grain its structure at the amino-acid level and build the associated elastic network (EN). The application of EN models to proteins is relatively recent [18], since it has commonly been assumed that little structural detail could be given up in order to model their complex energy landscapes. However, there is now strong evidence that most features of the large- and medium-scale dynamics of proteins’ fluctuations around their native states, related to function and stability, can be successfully reproduced by simple harmonic interactions between amino acids [19–23]. In view of the BP phenomenology, it is important to mention the growing consensus that an explanation in glasses could be found within a purely harmonic context [24].

In the framework of EN models, the potential energy is written as a sum of pairwise harmonic potentials,

$$V (\tilde{r}_i) = \sum_{i<j} V(\tilde{r}_i, \tilde{r}_j) = \sum_{i<j} \frac{k_{ij}}{2} (|\tilde{r}_{ij}| - |\tilde{r}_{ij}^{(0)}|)^2,$$

where $\tilde{r}_{ij} = \tilde{r}_i - \tilde{r}_j$, $\tilde{r}_i$ being the position of the $i$th particle, $\tilde{r}_i^{(0)}$ its equilibrium position, and $k_{ij}$ the stiffness of the spring connecting particles $i$ and $j$. More precisely, the vector $\tilde{r}_i$ represents the instantaneous position of the $\alpha$ carbon of the $i$th amino acid, $\tilde{r}_i^{(0)}$ its position in the native state as determined from x-ray crystallography or nuclear magnetic resonance, and $k_{ij}$ can take different functional forms, such as $k_{ij} = \kappa \theta(r_c - |\tilde{r}_i^{(0)} - \tilde{r}_j^{(0)}|)$ [sharp cutoff model [20]] or $k_{ij} = \kappa \exp(-|\tilde{r}_i^{(0)} - \tilde{r}_j^{(0)}|^2/r_c^2)$ [Gaussian...
the position vectors in the native structure function (1) evaluated at the equilibrium structure. Were $K$ contact matrix specified density.

scale structural properties are involved, proteins are well chain, while beyond such range spatial correlations are lost. We repeated this analysis for several loops. After a third, less resolved shell all pairwise spatial neighbors at fixed distance along the chain and the next-well-defined coordination shells, namely, the nearest neighbors at fixed distance along the chain and the next-nearest off-chain neighbors, including the pairs belonging to alpha helices and those lying at turning regions, such as loops. After a third, less resolved shell all pairwise spatial correlations are lost. We repeated this analysis for several proteins and always found that the second and the third peaks are always related to the presence of secondary motifs as well as to the intrinsic flexibility of the peptide chain, while beyond such range spatial correlations are absent. This fact is a clear indication that, as far as large-scale structural properties are involved, proteins are well approximated by random assemblies of amino acids with specified density.

The analogy between protein structures and disordered systems with no long-range order suggests that a common mechanism might be responsible for the emergence of the BP in both cases. In topologically disordered solids, this anomaly appears upon increasing the temperature or, as observed, for example, in silica, upon lowering the density. In the present case, we are dealing with proteins, i.e., objects whose equilibrium structure is fixed by the biological function. However, changes in the particle density may still be simulated by resorting to the free parameter $r_c$. In the framework of EN models, $r_c$ sets the range of inter-particle interactions and should in principle be tuned by fitting the low-frequency portion of experimental spectra at temperatures below the dynamical transition, where the protein vibrates harmonically within a local minimum. The usual alternative is to compare with spectra as determined by all-atom force fields [21]. By doing this, one obtains $\rho_c = 3 \, \text{Å}$ in an all-atom representation [21], which coarse grains to $r_c = \langle N_p \rangle^{1/3} \rho_c = 8 \, \text{Å}$ when the average number of atoms per amino acid $\langle N_p \rangle = 18$ is introduced. Interestingly, by its very definition, the parameter $r_c$ also allows to regulate an effective local density of the system by tuning the average connectivity $c = \frac{1}{3N} \sum^{N}_{i=1} \sum_{j=1}^{3} k_{ia,jb}$, where $k_{ia,jb}$. By decreasing the cutoff $r_c$, the average number of neighbors per residue diminishes accordingly. Thus, a measure of compactness may be introduced that is proportional to $c$. It can be shown that varying $r_c$ induces a change in the connectivity that scales with the interaction volume $r_c^3$ up to finite-size $O(r_c)$ corrections (see left inset in Fig. 1). This means that we can study the spectral features of a given protein structure with the additional degree of freedom of varying density by simply changing the interaction cutoff $r_c$, which thus plays in this context the role of a control parameter.

The vibrational spectrum of a protein for a certain value of the parameter $r_c$ is obtained by diagonalizing the contact matrix. However, especially for small proteins, the finite number of residues makes it difficult to analyze the low-frequency features of the spectra. In order to circumvent this problem, we generated a number of different conformers for each of the analyzed structures such that all of them are by construction compatible with the atomic MSDs as specified by the native contact matrices. More precisely, if we write the coordinates of a given conformer as $\mathbf{r}^{(0)} = \overline{\mathbf{r}}^{(0)} + \delta \overline{\mathbf{r}}$, then it is sufficient to take $\delta \overline{\mathbf{r}} = \mathbf{U} \overline{\mathbf{c}}$, where $\mathbf{U}$ is the matrix of eigenvectors of $K$ and the $3N - 6$ coefficients $c_k$ are drawn from as many one-dimensional Gaussian distributions with zero mean and standard deviations $\sigma_k = \sqrt{-k_B T/\lambda_k}$, $\lambda_k = -\omega_k^2$ being the eigenvalues of the contact matrix $K$. This procedure provides a simple means to construct an arbitrary number of conformations that are dynamically equivalent to the native one in the harmonic approximation.

In Fig. 2 we plot $g(\omega)$ and $g(\omega)/\omega^2$ for several values of the cutoff $r_c$ for two representative proteins of different size. Similar results were obtained for a choice of other proteins. A shoulder manifestly appears in the low-
the Gaussian model in glasses, the BP should be interpreted
forming liquids [17]. More rigorously, as it is the case for
utterly analogous to the one found in glasses and glass-
can be interpreted in terms of a topological instability
suggests that the BP in protein structures at low densities
is a precursor of the transition within a model that by
definition becomes meaningless at the critical point. This
is precisely what happens in our case, at an interaction
range below which protein structures start unfolding. We
also stress that the shift of \( \omega_{BP} \) towards zero frequency and
the divergence of the BP height as the systems lose rigidity
as a precursor of the transition within a model that by
definition becomes meaningless at the critical point. This
is precisely what happens in our case, at an interaction
range below which protein structures start unfolding. We
also stress that the shift of \( \omega_{BP} \) towards zero frequency and
the divergence of the BP height as the systems lose rigidity
is a spectral feature equally unveiled within different theo-
retical approaches [12–15].

It is also instructive to study the localization properties of
typical ensembles of spectra through the level-spacing
statistics \( P(s) \) [26]. As an example, we plot the results
obtained for ubiquitin in Fig. 3. Overall, the distribution
is very well described by a Wigner law, which holds for
fully extended spectra. As we decrease the cutoff, a small
contribution from localized modes is observed, as the
measure of \( J_0 \equiv \langle s^2 \rangle/2 \) shows (upper inset of Fig. 3).
Otherwise, \( J_0 \) should be close to 1 in the case of a localized
spectrum, which is never the case. A more refined analysis
[27] performed on several proteins clearly shows that the
only localized modes are due to the tail of the spectrum at
large frequencies, much alike structural glasses [12,28].
This conclusion, further confirmed by the level-spacing
statistics from the low-frequency portion of the spectra
(lower inset of Fig. 3), rules out the presence of localized
modes in the BP region.

The origin of a precursory feature of a topological
instability in proteins can be formally understood by recalling
that their structures are those of folded polymers. If the
interaction cutoff \( r_c \) is lowered below the first off-chain
coordination shell, native conformations lose their folded
nature and become more and more akin to liquids. In fact,
we argue that the appearance of the BP precisely antici-

FIG. 2. Boson peak analysis for two globular proteins of differ-
ent size. Left panels: serum albumin (1A06), \( N = 578 \) residues.
Right panels: ubiquitin (1UBI), \( N = 76 \) residues. The four upper
panels show the density of states for different values of \( r_c \) (for
1000 thermal replicas). In the four lower ones, we show the fits
to the BP frequency and height with the mean-field expressions
(2). The best-fit results are: \( r_c^* = 5.7 \) Å (serum albumin) and
\( r_c^* = 3.5 \) Å (ubiquitin). The physical units for frequencies were
obtained with \( r_c = 8 \) Å.

frequency region as \( r_c \) is reduced (see upper panels in
Fig. 2), and eventually a divergence develops if \( r_c \) is
decreased below a critical value. The origin of such peak
can be uncovered by tracking its position \( \omega_{BP} \) and height
\( h_{BP} \), i.e., our effective density, decreases. From the
lower panels of Fig. 2 one can clearly appreciate that the
scaling followed by \( \omega_{BP} \) and \( h_{BP} \) is very well interpolated
by the analytical functional forms predicted by the ERM
theory in the mean-field approximation [16], i.e.,

\[
\omega_{BP} \sim (r_c - r_c^*)^\alpha, \quad h_{BP} \sim (r_c - r_c^*)^{-\beta}
\]

with \( \alpha = 1 \) and \( \beta = 1/2 \). Therefore, our analysis strongly
suggests that the BP in protein structures at low densities
can be interpreted in terms of a topological instability
utterly analogous to the one found in glasses and glass-
forming liquids [17]. More rigorously, as it is the case for
the Gaussian model in glasses, the BP should be interpreted

FIG. 3. Plot of the level-spacing statistics of ubiquitin for
different values of the cutoff \( r_c \). The Wigner-Dyson (thick solid
line) and Poisson (dashed line) statistics, which describe totally
uncorrelated spectra, are also shown for comparison. Upper
inset: \( J_0 = \langle s^2 \rangle/2 \) is plotted vs \( r_c \). The dashed line represents
the value expected for a fully extended spectrum. Lower inset:
level-spacing statistics for frequencies \( \omega < 2.5 \) meV. The solid
line is a plot of the Wigner surmise.
pates such inherent instability before the critical cutoff is reached. Accordingly, the best-fit values of $r_0^*$ for all the analyzed structures never does exceed the first off-chain coordination shell (see Fig. 2). Keeping in mind that the optimal value of $r_0^*$ is around 8 Å, i.e., above its critical value, our results suggest that protein structures express an inherent trade off between spatial properties of liquids, i.e., increased degree of mobility, and the necessity of maintaining a certain structural stability. Interestingly, from an extensive analysis on a selection of 13 proteins, we find that $r_0^*$ is substantially anticorrelated with the packing fraction $p = 4/3(N/V)(d_0/2)^3$, i.e., a measure of global compactness, whereas weak correlation is found with indicators of local stability, such as the content of $\alpha$ helices and $\beta$ sheets. Here $N$ and $V$ are the number of residues and the volume, while $d_0 \approx 3.83$ Å is the inter-residue distance along the main chain. Moreover, we also find a positive correlation between $r_0^*$ and $N$, which may signal the larger mechanical stability of smaller proteins (see Table I).

The above conclusions may be interpreted by regarding proteins as molecular machines bound to keep a specified geometry in order to perform their biological function, yet preserving a high degree of structural flexibility in order to efficiently explore different conformational states. In this sense, the mechanical instability underlying the emergence of a BP appears to be a universal signature of their engineered ability to easily travel between adjacent local minima in their native states. We note that our results agree with recent estimates of the spectral dimension of globular proteins, whose non-Debye behavior has been interpreted in terms of a vibrational instability of the Peierls-Landau type [29].

Summarizing, in this Letter we have provided compelling evidence of the equivalence of the boson peak phenomenon in globular proteins and glasses. Our analysis suggests that a topological instability of the saddle-phonon type in proteins reflects the balance imprinted in their structures between being capable of rapidly accessing different minima in the native energy landscape while keeping a relative mechanical rigidity.

We thank L. Caserti, O. Martin, M. Mézard, and G. Parisi for interesting discussions. S. C. also thanks the EPFL for its hospitality. S.C. is supported by ECHP Program, Contract No. HPRN-CT-2002-00319 (STIPCO).