The Glass-like Structure of Globular Proteins and the Boson Peak

Stefano Ciliberti, Paul De Los Rios, and Francesco Piazza

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

(Dated: February 8, 2020)

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.

PACS numbers: 87.15.-v, 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) [4, 5].

Interestingly, there exists a remarkable similarity of the Raman and neutron–scattering spectra of proteins with those of glasses and super-cooled liquids [6], 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 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 model [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 behaviour is viewed as a symptomatic effect of the topological phase transition which is conjectured to happen in glasses at low temperatures [12]. 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 [13], 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 [10, 20, 21, 22, 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 pair-wise harmonic potentials,

\[
\mathcal{V}(\{\vec{r}_i\}) = \sum_{i<j} V(\vec{r}_i, \vec{r}_j) = \sum_{i<j} \kappa_{ij} \left( |\vec{r}_{ij}| - |\vec{r}_{ij}^{(0)}| \right)^2 \tag{1}
\]

where \( \vec{r}_{ij} = \vec{r}_i - \vec{r}_j \), \( \vec{r}_i \) being the position of the \( i \)-th particle, \( \vec{r}_i^{(0)} \) its equilibrium position and \( k_{ij} \) the stiffness of the spring connecting particles \( i \) and \( j \). More precisely, the vector \( \vec{r}_i \) represents the instantaneous position of the \( \alpha \)-carbon of the \( i \)-th amino-acid, \( \vec{r}_i^{(0)} \) 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 \( k_{ij} = \kappa \theta(\vec{r}_c - |\vec{r}_i^{(0)} - \vec{r}_j^{(0)}|) \) (sharp cutoff model [20]) or \( k_{ij} = \kappa \exp(-|\vec{r}_i^{(0)} - \vec{r}_j^{(0)}|^2/\ell^2) \) (Gaussian model [21]), which is the one we adopt here.
functions are surely not random, an analysis of the pair correlation function (1) evaluated at the equilibrium structure. The comparison shows that the protein structure is characterized by two 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 pair-wise 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 $r_c \approx 3 \, \text{Å}$ in an all-atom representation [21], which coarse-grains to $r_c \approx \langle N_o \rangle^{1/3} \rho_c \approx 8 \, \text{Å}$ when the average number of atoms per amino-acid $\langle N_o \rangle \approx 18$ 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 $\langle c \rangle = \frac{1}{3N} \sum_{i=1}^{N} \sum_{\alpha=1}^{3} K_{i\alpha,i\beta}$. By decreasing the cutoff $r_c$, the average number of neighbors per residue diminishes accordingly. Thus, a local measure of compactness may be introduced that is proportional to $\langle c \rangle$. 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 $\mathcal{O}(r_c)$ 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 be 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 $\tilde{r}^{(0)} = r^{(0)} + \delta \tilde{r}$, then it is sufficient to take $\delta \tilde{r} = U \tilde{c}$, where $U$ 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
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: $r_c^* = 5.25$ Å (Serum Albumin) and $r_c^* = 3.5$ Å (Ubiquitin). The physical units for frequencies were obtained with $r_c = 8$ Å.

![FIG. 2: Boson peak analysis for two globular proteins of different size. Left panels: Serum Albumin (1AO6), $N = 578$ residues. Right panels: Ubiquitin (1U8I), $N = 76$ 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.](image1)

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 [31] 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, 27]. 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.

![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 describes totally uncorrelated spectra, are also shown for comparison. Upper inset: $J_0 \equiv \langle s^2 \rangle / 2$ is plotted versus $r_c$. 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.](image2)
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 anticipates such inherent instability before the critical cutoff is reached. Accordingly, the best-fit values of $r_c^{*}$ for all the analyzed structures does never exceed the first off-chain coordination shell (see Fig. 4). Keeping in mind that the optimal value of $r_c$ 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 $r_c^{*}$ is substantially anti-correlated 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 $d_0 \simeq 3.83$ Å is the inter-residue distance along the main chain. Moreover, we also find a positive correlation between $r_c^{*}$ 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 [20].

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 able of rapidly accessing different minima in the native energy landscape while keeping a relative mechanical rigidity.

We thank L. Casetti, O. Martin, M. Mérzard and G. Parisi for interesting discussions. S. C. also thanks the EPFL for its hospitality. S. C. is supported by ECHP Program, contract HPRN-CT-2002-00319 (STIPCO).

| Protein           | $N$ | $p$ | $(\alpha + \beta)$ | $r_c^{*}$ (Å) |
|-------------------|-----|-----|---------------------|--------------|
| Insulin           | 51  | 0.20| 0.53               | 4.57         |
| Protein G         | 56  | 0.21| 0.70               | 3.64         |
| Ubiquitin         | 71  | 0.20| 0.46               | 3.53         |
| PDZ binding domain| 85  | 0.21| 0.55               | 4.03         |
| Lysozyme          | 162 | 0.17| 0.74               | 4.27         |
| Adenylate Kinase  | 214 | 0.12| 0.64               | 7.85         |
| LAO               | 238 | 0.16| 0.60               | 5.44         |
| CYSB              | 260 | 0.17| 0.59               | 4.70         |
| PBGD              | 296 | 0.16| 0.60               | 3.70         |
| Thermolysin       | 316 | 0.18| 0.53               | 4.55         |
| HSP70 ATP-binding domain | 382 | 0.15| 0.66               | 5.28         |
| Fab-fragment      | 437 | 0.13| 0.48               | 5.70         |
| Serum Albumin     | 578 | 0.12| 0.70               | 5.70         |

TABLE I: Correlation of $r_c^{*}$ with structural parameters.

[1] H. Frauenfelder, F. Parak and R.D. Young Ann. Rev. Bioph. Bioph. Chem. 17, 451 (1988).
[2] M.V. Volkenstein, Physical Approaches to Biological Evolution, Springer, Berlin, 1994.
[3] F. Piazza et al., Phys. Rev. Lett. 94, 145502 (2005).
[4] J. L. Green et al., J. Phys. Chem. 98, 13780 (1994).
[5] I. E. T. Iben, et al., Phys. Rev. Lett. 62, 1916 (1989).
[6] A. Orecchini et al., J. Phys. Chem. B 105, 12150 (2001); H. Leyser et al., Phys. Rev. Lett. 82, 2987 (1999); W. Doster, S. Cusack, and W. Petry Phys. Rev. Lett. 65, 1080 (1990).
[7] S. A. Alexander and R. Orbach, J. Phys. (France) 43, 625 (1982).
[8] Yu. V. Denisov and A. P. Rylev, JETP Lett. 52, 411 (1990); M. Nollmann and P. Etchegoin, Physica A 294 (1-2), 44-50 (2001).
[9] N. G. C. Astrath et al., Phys. Rev. B 71, 214202 (2005); R. Inoue, et al., Phys. Rev. Lett. 95, 056102 (2005); E. Courtens et al., Solid State Commun. 117, 187 (2001); O. Pilla et al., Phys. Rev. Lett. 85, 2136 (2000); Y. Inamura et al., Physica B 284-288, 1157 (2000).
[10] V. Lubchenko and P. G. Wolynes, PNAS 100, 1515 (2003); W. A. Phillips, Rep. Prog. Phys. 50, 1657 (1987).
[11] M. Foret et al., Phys. Rev. Lett. 77, 3831 (1996).
[12] S. Taraskin et al., Phys. Rev. Lett. 86, 1255 (2001).
[13] G. Parisi, Europ. Phys. J. E9, 213 (2002).
[14] M. Wyart et al., arXiv:cond-mat/0409687.
[15] L. E. Silbert et al., Phys. Rev. Lett., 95, 098301 (2005)
[16] S. Ciliberti et al., J. Chem. Phys. 119, 8577 (2003).
[17] T. S. Grigera et al., Nature 422, 289 (2003).
[18] M. M. Tirion, Phys. Rev. Lett. 77, 1905 (1996).
[19] A. R. Atiţan et al., Biophys. J. 80, 505 (2001).
[20] T. Haliloglu et al., Phys. Rev. Lett. 79, 3090 (1997).
[21] K. Hinsen, Proteins 33, 417 (1998).
[22] F. Tama and Y. H. Sanejouand, J. Phys. Chem. B 103, 4104 (1999).
[23] M. J. Horbach et al., Phys. Rev. Lett. 80, 2733 (1998).
[24] G. Ruocco et al., Phys. Rev. Lett. 84, 5788 (2000); J. Horbach et al., J. Phys. Chem. B 103, 4104 (1999).
[25] W. A. Phillips, Rep. Prog. Phys. 50, 1657 (1987).
[26] See e.g. T. Guhr et al., Phys. Rev. 299, 189 (1989).
[27] A. I. Chumakov et al., Phys. Rev. Lett. 92, 245508
[28] P. De Los Rios, S. Ciliberti and F. Piazza, in preparation.

[29] R. Burioni et al., Proteins 55, 529 (2004).

[30] see e.g. S. Ciliberti and T. S. Grigera, Phys. Rev. E 70, 061502 (2004).