Discrete breathers in nonlinear network models of proteins

B. Juanico and Y.–H. Sanejouand

Ecole Normale Supérieure, Laboratoire Joliot-Curie,
USR 3010 du CNRS, 46, allée d'Italie, 69364 Lyon Cedex 07, France

F. Piazza and P. De Los Rios

Ecole Polytechnique Fédérale de Lausanne, Laboratoire de Biophysique Statistique,
ITP–SB, BSP-720, CH-1015 Lausanne, Switzerland

We introduce a topology-based nonlinear network model of protein dynamics with the aim of investigating the interplay of spatial disorder and nonlinearity. We show that spontaneous localization of energy occurs generically and is a site-dependent process. Localized modes of nonlinear origin form spontaneously in the stiffest parts of the structure and display site-dependent activation energies. Our results provide a straightforward way for understanding the recently discovered link between protein local stiffness and enzymatic activity. They strongly suggest that nonlinear phenomena may play an important role in enzyme function, allowing for energy storage during the catalytic process.

The predictions of elastic network models (ENMs) of proteins [1–4] have proven useful in quantitatively describing amino-acid fluctuations at room temperature [1], often in good agreement with isotropic [2], as well as anisotropic measurements [5, 6]. Moreover, it has been shown that a few low-frequency normal modes can provide fair insight on the large amplitude motions of proteins upon ligand binding [7–9], as previously noticed when more detailed models were considered [10–12], also by virtue of the robust character of the collective functional motions [13].

However, low-frequency modes of proteins are known to be highly anharmonic [14, 15], a property which has to be taken into account in order to understand energy storage and transfer within their structure as a consequence of ligand binding, chemical reaction, *etc* [16, 17]. Indeed, there is growing experimental evidence that long-lived modes of nonlinear origin may exist in proteins [18, 19]. Likewise, many theoretical studies have appeared suggesting that localized vibrations may play an active role in, e.g., enzyme catalysis [20]. These include topological excitations such as solitons [21] as well as discrete breathers (DBs) [22, 23].

The latter are nonlinear modes that emerge in many contexts as a result of both nonlinearity and discreteness [24]. Although their existence and stability properties are well understood in systems with translational invariance, much less is known of the subtle effects arising from the interplay of spatial disorder and anharmonicity [25–27]. For this purpose, in the present work we introduce the nonlinear network model (NNM). Our aim is to extend the simple scheme of ENMs, known to capture the topology-based features of protein dynamics [1–3], by adding anharmonic terms. Within the NNM framework, we show that spontaneous localization of energy can occur in protein-like systems and that its properties may be intuitively rationalized in the context of specific bio-
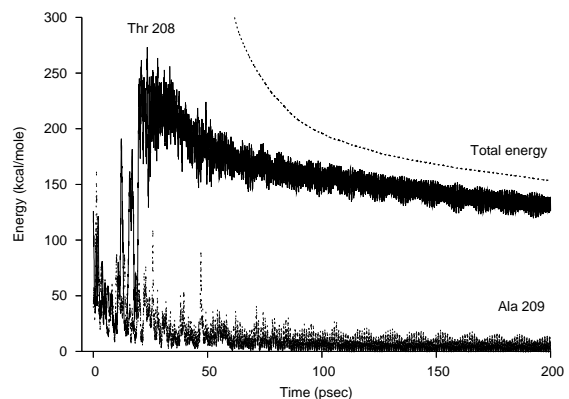

Figure 1: Energy as a function of time, when citrate synthase is cooled down as a consequence of surface friction. Dashed line: total energy. Solid line: energy of Threonine 208, the amino-acid the most involved in the DB. Dotted line: energy of Alanine 209, also involved in the DB. $k_B T_{eq} = 20$ kcal/mol.

logical functions. In our model, the potential energy of a protein, E_p, has the following form:

$$E_p = \sum_{d_{ij}^0 < R_c} \left[\frac{k_2}{2}(d_{ij} - d_{ij}^0)^2 + \frac{k_4}{4}(d_{ij} - d_{ij}^0)^4 \right] \quad (1)$$

where d_{ij} is the distance between atoms i and j, d_{ij}^0 their distance in the structure under examination (as e.g. solved through X-ray crystallography) and R_c is a cutoff that specifies the interacting pairs. As done in numerous studies, only C_α atoms are taken into account [4] and k_2 is assumed to be the same for all interacting atom pairs [1]. As in previous ENM studies [8, 28], we take $R_c = 10$ Å, and fix k_2 so that the low-frequency part of the linear spectrum match actual protein frequencies, as calculated through realistic force fields [10–12]. This gives $k_2 = 5$ kcal/mol/Å^2, with the mass of each C_α
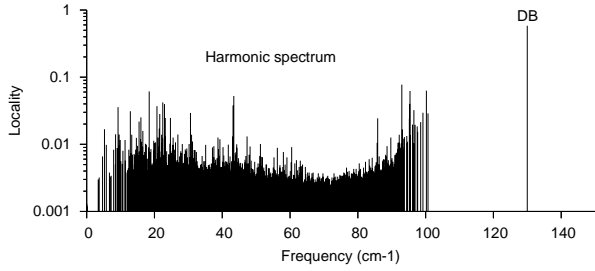
Figure 2: Locality of citrate synthase harmonic modes, as a function of their frequencies, together with the locality and frequency of a discrete breather (DB).

fixed to 110 a.m.u., that is, the average mass of amino-acid residues. Note that standard ENM corresponds to $k_4 = 0$, while in the present work $k_4 = 5$ kcal/mol/Å^4.

Proteins live and perform their functions immersed in water and exchange energy with the solvent through their sizable surface portion. In a previous paper we showed that complex energy relaxation patterns are observed as a result of the inhomogeneity of the coupling to the solvent of bulk and surface atoms [29]. In the presence of nonlinearity, boundary relaxation is known to drive a wide array of systems towards regions of phase space corresponding to localized modes that emerge spontaneously [30–33]. Thus, in order to study *typical* excitations of nonlinear origin in protein structures, it appears natural to perform a boundary cooling experiment. Our protocol is the following. After 50 psec of microcanonical molecular dynamics (MD) simulation performed at a temperature T_{eq}, the protein is cooled down by adding a linear dissipation term to the force acting on surface atoms, that is, those belonging to amino-acids with more than 25 Å^2 of solvent accessible surface area. This represents nearly 40% of the amino-acid residues, for all proteins considered in the present study. The viscous friction coefficient γ is set to 2 psec^{-1}, a typical value for protein atoms in a water environment [16]. Hereafter, the equilibration energies considered are in the range $k_B T_{eq} = 2 - 20$ kcal/mol, that is, of the order of, e.g., the energy release of ATP hydrolysis. With such initial conditions, energy in the system remains high for a period of time long enough so that localization can occur.

In Fig. 1, we show the energy of dimeric citrate synthase (PDB code 1IXE) as a function of time, as well as the energy of two amino-acids of monomer A, Thr 208 and Ala 209. After $t = 20$ psec and a few large fluctuations, a DB centered at Thr 208 forms. At $t = 200$ psec, more than 80% of the total energy is located there. Note the slow decay of the total energy after $t = 100$ psec and the periodic energy exchanges of Thr 208 with Ala 209, another among the few amino-acids involved in the DB. Note also that at $t = 20$ psec the energy of Thr 208 is higher than at $t = 0$, that is, when the friction was turned on, a clear-cut demonstration of the known tendency of DBs to harvest energy from lower-energy excitations [24]. In order to check that the phenomenon
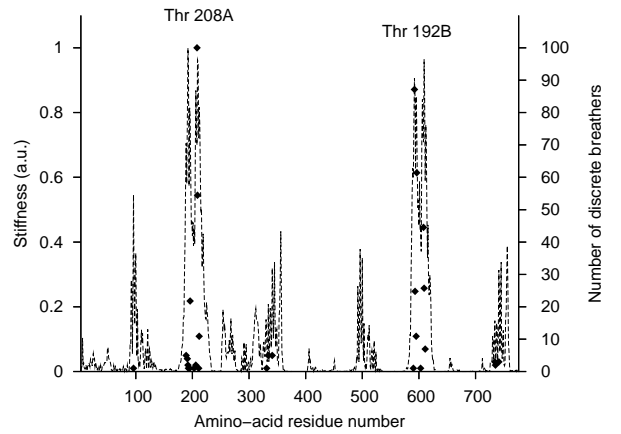

Figure 3: Stiffness of dimeric citrate synthase as a function of residue number (dashed line). The number of DBs found at a given site out of 500 instances is also reported (black diamonds, right y-axis).

shown in Fig. 1 is indeed the spontaneous localization of a DB, we switched off the friction at $t = 200$ psec and performed 100 more psec of microcanonical MD simulation. Then, we projected the latter trajectory on the first eigenvector of the corresponding velocity-covariance matrix, which gives the pattern of correlated atomic velocities involved in the DB. The Fourier transform of such a projection yields an accurate value for the DB frequency, while the spectral line-width provides information on the DB stability over the 100 psec analysis time-span.

In Fig. 2 we report the harmonic spectrum of citrate synthase as well as the DB frequency as functions of a locality measure. The latter is defined as $L_k = \sum_{i,\alpha} [\xi_{i\alpha}^k]^4 / [\sum_{i,\alpha} [\xi_{i\alpha}^k]^2]^2$, where $\xi_{i\alpha}^k$ is the α (x, y, z) coordinate of the i-th atom in the k-th displacement pattern (normalized eigenvector, DB). As expected, the DB frequency (130 cm^{-1}) lies above the highest frequency of the harmonic spectrum (101 cm^{-1}). Moreover, the corresponding spatial pattern is much more localized than any of the harmonic modes (note the logarithmic scale).

Starting from random initial conditions, we obtained 500 stable DBs following the above-outlined protocol. Although in many cases several DBs emerged, we decided to retain only the runs where a single DB catched most of the system energy, and more energy than the average amount per site at $t = 0$. In Fig. 3 we report the number of DBs found at each site. The largest fraction (20 %) of these highly energetic DBs formed at Thr 208 in monomer A, but we also observed DBs at 27 other sites, noteworthy at Thr 192 of monomer B (18%). Note also that, although the studied protein is a dimer, that is, with a an approximate but clear two-fold structural symmetry, the probability to observe a DB at a given site varies from one monomer to the other, indicating that the localization dynamics is rather sensitive to small changes in the local environment. As shown in Fig. 3, this probability is higher in the stiffest parts of the protein scaffold, as measured through an indicator of local stiffness s_i. For

amino-acid i, the latter is defined as:

$$s_i = \frac{1}{\mathcal{N}_i} \sum_{j,\alpha} \sum_{k \in \mathcal{S}} [\xi_{j\alpha}^k]^2 \theta(R_c - d_{ij}^0) \qquad (2)$$

where $\mathcal{N}_i = \sum_j \theta(R_c - d_{ij}^0)$ is the number of neighbors of the i-th residue and $\theta(x)$ is the Heaviside step function. The second sum is over the set $\mathcal{S}$ of the ten highest frequency harmonic modes. The averaging over the $\mathcal{N}_i$ neighbors slightly smoothes mode contributions and helps underlining the fact that in each monomer of citrate synthase there is a stretch of nearly fourty consecutive amino-acids (residues 185-225) with a remarkably stiff environment, deeply buried at the interface between the two monomers. This is obviously where most DBs tend to emerge. Note, however, that the relationship between high-frequency harmonic modes and spontaneous energy localization is not a straightforward one: for instance, DBs were observed only a couple of times at the site the most involved in the highest frequency normal mode, namely, Ser 213. As a matter of fact, as suggested by the large energy fluctuations observed at site Thr 208 before the DB shown in Fig. 1 springs up, a competition between *potential* DBs is likely to occur, with possible weak-to-strong energy transfers, before a given site is occupied by a stable mode.

In lattice systems sites are obviously equivalent. Here, as shown in Fig. 4, the energy-frequency relationship is site-dependent. Furthermore, the probability for a DB to localize at a given site depends in a non-obvious fashion upon the energy it needs to reach a given frequency at that location. While most DBs emerge at Thr 208,
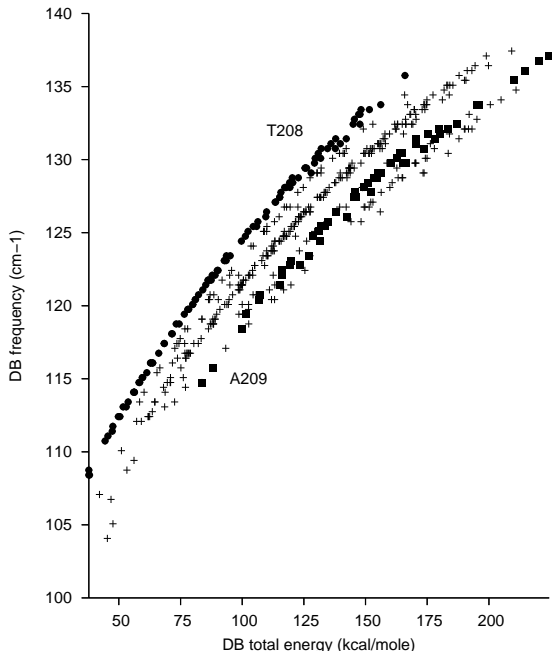

Figure 4: DB frequencies in citrate synthase as a function of their energy (pluses). The cases of Threonine 208 (filled circles) and Alanine 209 (filled squares) are highlighted. Using our protocol, no DB with an energy lower than 37.8 kcal/mole was observed, out of a total of 500 cases.
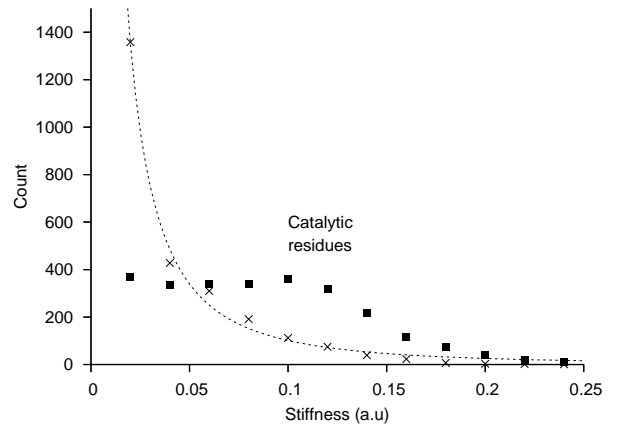

Figure 5: Stiffness of the environment of amino-acid residues involved in enzymatic activity (black squares), compared to that of amino-acids of same chemical type (crosses) randomly chosen within the same set of enzyme structures. The broken line is only a guide for the eye.

i.e. the site where the least energy is required for a given frequency, many DBs are also observed at Ala 209 in monomer A, one of the sites that demands more energy. In more than one dimension one expects DBs to appear only above a characteristic energy [24, 34]. Hence, our results hint at a strong site-dependence of such energy threshold, non-trivially related to local structural properties. To shed light on this intriguing feature, a detailed characterization of the small-ampitude side of the DB energy-amplitude curves at different sites is currently under way.

In the following step, we looked for DBs in other proteins, both dimeric and monomeric. For small proteins like HIV-1 protease (PDB code 1A30), a dimeric 2×99 amino-acids enzyme, no DB could be obtained. This is likely to be due to the fact that in small proteins too many amino-acids are in direct interaction with a site where energy dissipation occurs. This means that small proteins may require more detailed models, like all-atom schemes, where cutoff values of the order of 5 Å are customary [5, 13, 35].

In the case of aconitase (PDB code 1FGH), a monomeric 753 amino-acids enzyme, and alkaline phosphatase (PDB code 1ALK), a dimeric 2×499 amino-acids enzyme, DBs prove nearly as easy to generate than in the case of citrate synthase. However, for proteins of similar sizes, the probability of similar events turns out to vary significantly from a protein to another. For instance, in the cases of phospholipase D (PDB code 1V0Y), a monomeric 504 amino-acids enzyme, and isoamylase (PDB code 1BF2), a monomeric 750 amino-acids enzyme, out of 100 cooling MD simulations, only 8 and 5 DBs were obtained, respectively, in contrast to citrate-synthase, where our success rate is over 50 %. This points to the intriguing conclusion that not only DBs in proteins are site-selective, but also appear to be non-trivially fold-selective.

In all the analyzed structures, spontaneous localization of energy occurs in the stiffest parts of the structure. Thus, we turn now to examine the relationship between protein stiffness and function. Following the hypotheses that enzymatic activity may require some kind of energy storage and that DBs may play a role in the process, we computed high-frequency normal modes for a set of 833 enzymes from the 2.1.11 version of the Catalytic Site Atlas [36]. Then, we determined the stiffness of each amino-acid known to be involved in enzymatic activity according to (2). As a comparison, we also determined stiffnesses of amino-acids of the same chemical type, but picked at random among those not known to be involved in enzymatic activity. As shown in Fig. 5, catalytic amino-acids tend to be located in stiffer parts of enzyme structures, in agreement with our hypotheses. This is not an obvious result, since for the sake of catalytic activity amino-acids have to interact with enzyme substrates, that is, to be accessible to them. Such

a trend has already been noticed in other studies. Noteworthy, using the ease of displacing any given amino-acid residue with respect to the others as a stiffness measure, it was shown that roughly 80 % of the catalytic residues are located in stiff parts of enzyme structures [37]. In a more indirect way, it was also remarked that global hinge centers colocalize with catalytic sites in more than 70 % of enzymes [38]. So, stiff parts may play a role of pivot, allowing for accurate large-amplitude conformational changes of enzymes upon substrate binding.

What our results further suggest is that stiff parts of enzyme structures may also play another major role in enzyme function, namely by allowing for an active role of nonlinear localized modes in energy storage during the catalytic process.

[1] M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[2] I. Bahar, A. R. Atilgan, and B. Erman, Fold. Des. **2**, 173 (1997).
[3] K. Hinsen, Proteins **33**, 417 (1998).
[4] I. Bahar and Q. Cui, eds., *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. C&H/CRC Mathematical & Computational Biology Series, vol. 9* (CRC press, Boca Raton, 2005).
[5] D. Kondrashov, A. Van Wynsberghe, R. Bannen, Q. Cui, and G. Phillips, Structure **15**, 169 (2007).
[6] C. Micheletti, G. Lattanzi, and A. Maritan, J. Mol. Biol. **321**, 909 (2002).
[7] F. Tama and Y.-H. Sanejouand, Prot. Eng. **14**, 1 (2001).
[8] M. Delarue and Y.-H. Sanejouand, J. Mol. Biol. **320**, 1011 (2002).
[9] W. G. Krebs, V. Alexandrov, C. A. Wilson, N. Echols, H. Yu, and M. Gerstein, Proteins **48**, 682 (2002).
[10] B. R. Brooks and M. Karplus, Proc. Natl. Acad. Sci. USA **82**, 4995 (1985).
[11] O. Marques and Y.-H. Sanejouand, Proteins **23**, 557 (1995).
[12] D. Perahia and L. Mouawad, Comput. Chem. **19**, 241 (1995).
[13] S. Nicolay and Y.-H. Sanejouand, Phys. Rev. Lett. **96**, 078104 (2006).
[14] R. Levy, D. Perahia, and M. Karplus, Proc. Natl. Acad. Sci. USA **79**, 1346 (1982).
[15] S. Hayward, A. Kitao, and N. Go, Proteins **23**, 177 (1995).
[16] D. Sagnella, J. Straub, and D. Thirumalai, J. Chem. Phys. **113**, 7702 (2000).
[17] D. M. Leitner, Phys. Rev. Lett. **87**, 188102 (2001).
[18] J. Edler, R. Pfister, V. Pouthier, C. Falvo, and P. Hamm, Phys. Rev. Lett. **93**, 106405 (2004).
[19] A. H. Xie, L. van der Meer, W. Hoff, and R. H. Austin, Phys. Rev. Lett. **84**, 5435 (2000).
[20] A. Sitnitsky, Physica A **371**, 481 (2006).
[21] F. d'Ovidio, H. G. Bohr, and P.-A. Lindgard, Phys. Rev. E **71**, 026606 (2005).
[22] J. F. R. Archilla, Y. B. Gaididei, P. L. Christiansen, and J. Cuevas, J. Phys. A **35**, 8885 (2002).
[23] G. Kopidakis, S. Aubry, and G. P. Tsironis, Phys. Rev. Lett. **87**, 165501 (2001).
[24] S. Flach and C. R. Willis, Phys. Rep. **295**, 181 (1998).
[25] T. Dauxois, A. Litvak-Hinenzon, R. MacKay, and A. Spanoudaki, eds., *Energy localisation and transfer in crystals, biomolecules and josephson arrays. Advanced Series in Nonlinear Dynamics, vol.22* (World Scientific, Singapore, 2004).
[26] F. Abdullaev, O. Bang, and M. P. Sorensen, eds., *Nonlinearity and Disorder: Theory and Applications*, vol. 45 (Kluwer Academic Publishers, Dordrecht, 2001).
[27] K. O. Rasmussen, D. Cai, A. R. Bishop, and N. Gronbech-Jensen, Europhysics Letters **47**, 421 (1999).
[28] H. Valadie, J.-J. Lacapere, Y.-H. Sanejouand, and C. Etchebest, J. Mol. Biol. **332**, 657 (2003).
[29] F. Piazza, P. De Los Rios, and Y.-H. Sanejouand, Phys. Rev. Lett. **94**, 145502 (2005).
[30] G. P. Tsironis and S. Aubry, Phys. Rev. Lett. **77**, 5225 (1996).
[31] F. Piazza, S. Lepri, and R. Livi, Chaos **13**, 637 (2003).
[32] R. Reigada, A. Sarmiento, and K. Lindenberg, Chaos **13**, 646 (2003).
[33] R. Livi, R. Franzosi, and G.-L. Oppo, Phys. Rev. Lett. **97**, 060401 (2006).
[34] M. Kastner, Phys. Rev. Lett. **92**, 104301 (2004).
[35] K. Suhre and Y.-H. Sanejouand, Act. Cryst. D **60**, 796 (2004).
[36] C. T. Porter, G. J. Bartlett, and J. M. Thornton, Nucl. Ac. Res. **32**, D129 (2004).
[37] S. Sacquin-Mora, E. Laforet, and R. Lavery, Proteins **67**, 350 (2007).
[38] L. Yang and I. Bahar, Structure **13**, 893 (2005).