Phys. Biol. **5** (2008) 026001 (14pp) doi:10.1088/1478-3975/5/2/026001

Discrete breathers in protein structures

Francesco Piazza[1] **and Yves-Henri Sanejouand**[2]

[1] Ecole Polytechnique Fédérale de Lausanne, Laboratoire de Biophysique Statistique, ITP-SB, BSP-722, CH-1015 Lausanne, Switzerland
[2] Ecole Normale Supérieure, Laboratoire Joliot-Curie, CNRS-USR 3010, 46 allée d'Italie, 69364 Lyon Cedex 07, France

E-mail: Francesco.Piazza@epfl.ch and Yves-Henri.Sanejouand@ens-lyon.fr

Received 23 February 2008
Accepted for publication 9 April 2008
Published 1 May 2008
Online at stacks.iop.org/PhysBio/5/026001

Abstract
Recently, using a numerical surface cooling approach, we have shown that highly energetic discrete breathers (DBs) can form in the stiffest parts of nonlinear network models of large protein structures. In the present study, using an analytical approach, we extend our previous results to low-energy discrete breathers as well as to smaller proteins. We confirm and further scrutinize the striking site selectiveness of energy localization in the presence of spatial disorder. In particular, we find that, as a sheer consequence of disorder, a non-zero energy gap for exciting a DB at a given site either exists or not. Remarkably, in the former case, the gaps arise as a result of the impossibility of exciting small-amplitude modes in the first place. In contrast, in the latter case, a small subset of linear edge modes acts as accumulation points, whereby DBs can be continued to arbitrary small energies, while unavoidably approaching one of such normal modes. In particular, the case of the edge mode seems peculiar, its dispersion relation being simple and little system dependent. Concerning the structure–dynamics relationship, we find that the regions of protein structures where DBs form easily (zero or small gaps) are unfailingly the most highly connected ones, also characterized by weak local clustering. Remarkably, a systematic analysis on a large database of enzyme structures reveals that amino-acid residues involved in catalysis tend to be located in such regions. This finding reinforces the idea that localized modes of nonlinear origin may play an important biological role, e.g., by providing a ready channel for energy storage and/or contributing to lower energy barriers of chemical reactions.

1. Introduction

Proteins are molecular machines whose functional motions are strongly related to, if not encoded within, their three-dimensional structure [1–3]. As a matter of fact, useful information on their functional motions can be obtained from the mere knowledge of their equilibrium structure, as solved for example through x-ray crystallography or NMR spectroscopy, even at the level of the harmonic approximation of the system potential energy [4–6]. Remarkably, this structure–dynamics–function relationship can be captured even when substantial amounts of structural details are missing. In particular, elastic network models (ENMs) of proteins [7–10] have been used for describing quantitatively amino-acid fluctuations at room temperature [7], often in very good agreement with isotropic [8], as well as with anisotropic measurements [11, 12]. They have also allowed us to show that a few low-frequency normal modes can often provide fair insight into the large amplitude motions of proteins upon ligand binding [13–15], demonstrating the robust character of these collective motions [16–18]. Taken together, such results highlight the important role of the peculiar equilibrium scaffolds of proteins [10, 13, 19], at the same time providing a rationale for the coarse graining of amino-acid assemblies.

Recently, the interest for problems potentially involving nonlinear effects in bio-molecules, such as localization and storage of energy, has increased in the community at the interface between physics and biology [20–26]. A hot case concerns enzymatic catalysis and, more specifically, the following question: how does an enzyme store and use the energy released at substrate binding or when a chemical bond is broken? It is noteworthy that this energy may be deployed on much longer time scales (microseconds to milliseconds) than those characteristic of the energetic process, and at very

distant places with respect to the catalytic site of the enzyme [27], so that it is highly unlikely that a normal-mode-assisted mechanism may prove enough for explaining the phenomenon.

Indeed, protein dynamics has been long known to be highly anharmonic [28, 29], a property which is certainly important in order to understand energy storage and transfer as a consequence of ligand binding, chemical reaction, etc [30, 31]. Among nonlinear effects, the possibility that localized vibrational modes of nonlinear origin may play a role in biological processes has recently been put forward by many authors [21], based on engrossing experimental studies reporting numerous subtleties of protein dynamics, notably through infra-red spectroscopy [31–34]. For example, the excitation of localized vibrations in α-helices has been proposed to be a way for enzymes to store energy during catalysis [20, 35]. Within this framework, energy transfer across helices would occur predominantly by hoppings of localized vibrations along the chain resulting from nonlinear coupling of spatially overlapping localized modes in resonance [36, 37].

Nonlinear excitations proposed to play an active role in protein functional dynamics include topological excitations such as solitons [38, 39] as well as discrete breathers (DBs) [40, 41]. The latter are nonlinear modes that emerge in many contexts as a result of both nonlinearity and spatial discreteness [42]. Their existence and stability properties are well understood in systems with translational invariance at zero temperature [43], and are also intensively investigated for nonlinear dynamical systems in the presence of a thermal environment [44–47]. However, not much is known regarding the subtle effects arising from the interplay of spatial disorder and anharmonicity [48–50], either in general or in the context of the functional dynamics of biological macro-molecules.

In a recent paper [51], we have introduced the nonlinear network model (NNM), with the aim of studying the simplest model that would take into account both the topology of protein structures and the anharmonicity of interparticle potentials. Building upon the many successes of ENMs [10], in the NNM framework, a protein is mapped onto a coarse-grained network of oscillators, whose equilibrium positions reflect the spatial arrangement of amino acids in the protein fold. It is noteworthy that, in the linear regime, at low temperature for instance, the dynamical behaviour of NNMs and that of ENMs are identical. However, in the nonlinear regime, the behaviour of NNMs becomes much more complex. For instance, by applying the technique of surface cooling, we have demonstrated that discrete breathers form spontaneously at a small subset of specific sites in a given structure, invariably in the stiffest regions. Remarkably, an interpretation of this finding can be proposed in terms of enzyme functional dynamics since, by studying stiffness patterns across a large data set of enzymes, we have shown that catalytic residues tend to sit in the most rigid portions of their structure. It is thus tempting to speculate that enzymes may take advantage of the well-known ability of discrete breathers to harvest and retain for long periods of times amounts of energy much larger than what is normally available at a given site at a given temperature, in order to achieve their function, which requires crossing energy barriers.

As it often happens, the preliminary findings reported in [51] have raised numerous questions. In particular, they have revealed several striking features of DBs in NNMs of proteins which deserve further detailed analysis, like the subtle, ubiquitous spatial modulation of their properties, such as their dispersion relation, connection with the edge normal modes, degree of localization and presence of an energy gap in the excitation spectrum. On the other hand, the numerical simulations analysed in [51] allowed us to study highly energetic DBs *only*, as a consequence of the lack of a simple and clear-cut criterium for recognizing low-energy ones. Moreover, the protein surface cooling procedure allowed us to observe DBs in large proteins *only*, like citrate synthase, a 2×371 amino-acids dimeric enzyme (PDB code 1ixe), probably as a consequence of a too large surface–volume ratio of the smaller proteins considered, like HIV-1 protease, a 2×99 amino-acids dimeric enzyme (PDB code 1a30).

It is the purpose of this paper to present a first analytical study of discrete breathers in NNMs of proteins in order to study low-energy DBs and to show that such DBs also exist in smaller proteins. Moreover, this is an opportunity to start answering, *en passant*, a fair number of interesting questions raised by our previous findings. The paper is organized as follows. In section 2, we discuss the main features of NNMs of proteins. In section 2.1, we show how it is possible, in the case of such models, to obtain approximate breather solutions for the equations of motion, based on a simple argument invoking a separation of timescales. In particular, we discuss how we solved the crucial problem of determining a suitable initial guess, a problem common to the majority of analytical methods for finding DB orbits, which is here further non-trivially faceted by the presence of spatial disorder. The localized solutions that we obtain, the spatial modulation of their basic properties and their connection with the static and dynamical features of the protein folds are analysed in depth in section 3. Finally, in section 4 we summarize our results and discuss the relevance of our findings in the context of enzyme function, while, in section 5, work still under way is rapidly evoked.

2. Nonlinear network models of proteins

In this section, we describe the nonlinear network model. Like in elastic network models, a given protein is modelled as an ensemble of N fictitious particles occupying the equilibrium positions of α-carbons, as found in the experimental structure. All particles have the same mass[3], which we set equal to the average amino-acid mass $M = 120$ Da, and each particle interacts with its neighbours, as specified by a cutoff distance R_c. More precisely, residues i and j interact if $|\boldsymbol{R}_i - \boldsymbol{R}_j| \leqslant R_c$, where $\boldsymbol{R}_i$ denotes the position vector of the ith residue in the equilibrium structure.

Let $\boldsymbol{u}_i = \boldsymbol{r}_i - \boldsymbol{R}_i$ denote the displacement vector of the ith residue, $\boldsymbol{r}_i$ being its instantaneous position. In the central

[3] As our fictitious particles occupy the equilibrium positions of amino acids, i.e. are located on the corresponding α-carbons, we will use the words particles and (amino-acid) residues interchangeably.

force approximation, the inter-particle potential energies may be expanded in power series as follows:

$$U(\boldsymbol{u}_i, \boldsymbol{u}_j) = \sum_{p=2}^{4} \frac{k_p}{p}(r_{ij} - R_{ij})^p + \mathcal{O}(u^5)$$
$$= \sum_{p=2}^{4} \frac{k_p}{p}(|\boldsymbol{u}_{ij} + \boldsymbol{R}_{ij}| - R_{ij})^p + \mathcal{O}(u^5), \quad (1)$$

where $\boldsymbol{u}_{ij} = \boldsymbol{u}_i - \boldsymbol{u}_j$ and $r_{ij} = |\boldsymbol{r}_{ij}| = |\boldsymbol{r}_i - \boldsymbol{r}_j|$, $R_{ij} = |\boldsymbol{R}_{ij}| = |\boldsymbol{R}_i - \boldsymbol{R}_j|$ are the inter-particle instantaneous and equilibrium distances, respectively. The total potential energy can then be written as

$$\mathcal{U}(\{\boldsymbol{u}\}) = \sum_i \epsilon_i, \quad (2)$$

where we have introduced the site energies ϵ_i:

$$\epsilon_i = \frac{1}{2}\sum_{j\neq i} c_{ij} U(\boldsymbol{u}_i, \boldsymbol{u}_j) \quad (3)$$

specified through the contact matrix

$$c_{ij} = \begin{cases} \theta(R_c - R_{ij}) & i \neq j \\ 0 & i = j \end{cases}, \quad (4)$$

$\theta(x)$ denoting the Heaviside step function. Setting $k_3 = k_4 = 0$ in equation (1) amounts to building a network of Hookean springs joining pairs of atoms separated by a distance smaller than R_c, that is, an ENM [7, 8, 10]. Here, we wish to study the simplest model capturing the combined effects of spatial disorder and nonlinearity. Hence, we restrict ourselves to symmetric potentials, by setting $k_3 = 0$. This choice allows us to get rid of distinct nonlinear features associated with asymmetric terms of the potentials, such as kinks and more complicated dc components of the localized modes [42], that are likely to interact with topological disorder in peculiar ways and, as such, deserve special attention in their own right.

The equations of motion for the mth residue then read

$$\ddot{\boldsymbol{u}}_m = \omega_0^2 \sum_{j\neq m} c_{mj} \times \left[\frac{(|\boldsymbol{u}_{jm} + \boldsymbol{R}_{jm}| - R_{jm}) + \beta(|\boldsymbol{u}_{jm} + \boldsymbol{R}_{jm}| - R_{jm})^3}{|\boldsymbol{u}_{jm} + \boldsymbol{R}_{jm}|}\right] \times (\boldsymbol{u}_{jm} + \boldsymbol{R}_{jm}), \quad (5)$$

where we have introduced the natural frequency $\omega_0 = \sqrt{k_2/M}$ and the parameter $\beta = k_4/k_2$.

We now want to look for solutions of the equations of motion in the form of localized, time-periodic modes with the angular frequency lying above the linear spectrum[4], i.e. the spectrum of the Hessian matrix of the potential energy, as given by equation (2). A sufficiently general ansatz has the form of a periodic sinusoidal oscillation modulated by a function of time that varies slowly on the timescale defined by the inverse frequency ω^{-1},

$$\boldsymbol{u}_m(t) = A\boldsymbol{\xi}_m(t)\cos\omega t. \quad (6)$$

[4] We observe that, in principle, the connectivity matrix c may be such that a sufficiently extended zone of forbidden frequencies in the linear spectrum might exist, in which case localized solutions with their frequency located in such a gap might be possible.

We assume that the envelope functions $\boldsymbol{\xi}_m(t)$ are bounded and such that $\max \boldsymbol{\xi}_m(t) \simeq \mathcal{O}(1)$, so that the amplitude A sets the physical scale for the oscillation amplitude.

In order for a physically sensible solution with its largest displacement at a given site k to exist, we must require $A \ll \min_j R_{jk}$, that is, the maximum vibration amplitude of the mode must be much smaller than the shortest bond between the kth particle and its neighbours[5]. Therefore, we can substitute the ansatz (6) into equation (5) and expand each addendum in the sum over j in series of A/R_{jm}. After some lengthy but straightforward algebra, we get

$$\left(\frac{\omega}{\omega_0}\right)^2 A\boldsymbol{\xi}_m C_\omega + \left(\frac{2\omega}{\omega_0^2}\right)A\dot{\boldsymbol{\xi}}_m S_\omega - \frac{A\ddot{\boldsymbol{\xi}}_m}{\omega_0^2} C_\omega$$
$$= -\sum_{j\neq m} c_{mj}\left\{\boldsymbol{R}_{jm}\left[(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})\epsilon_{jm}C_\omega + \left(\frac{\Delta\xi_{jm}^2}{2} - \frac{3(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})^2}{2}\right)\epsilon_{jm}^2 C_\omega^2 + \left(\frac{5(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})^3}{2} - \frac{3\Delta\xi_{jm}^2(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})}{2} + \beta R_{jm}^2(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})^3\right)\epsilon_{jm}^3 C_\omega^3\right] + R_{jm}\Delta\boldsymbol{\xi}_{jm}\left[(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})\epsilon_{jm}^2 C_\omega^2 + \left(\frac{\Delta\xi_{jm}^2}{2} - \frac{3(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})^2}{2}\right)\epsilon_{jm}^3 C_\omega^3\right]\right\} + \mathcal{O}(\epsilon^4), \quad (7)$$

where $C_\omega = \cos\omega t, S_\omega = \sin\omega t, \epsilon_{jm} = A/R_{jm}$ are the expansion parameters, $\Delta\boldsymbol{\xi}_{jm} = \boldsymbol{\xi}_j - \boldsymbol{\xi}_m$, the relative displacement patterns, and where we have introduced the unit distance vectors $\hat{\boldsymbol{R}}_{jm} = \boldsymbol{R}_{jm}/R_{jm}$.

Since we are assuming that the envelope functions are slowly varying during one breather oscillation, we can multiply equation (7) by $\cos\omega t$ and average over one period $2\pi/\omega$. By the same token, we neglect the second time derivatives of the envelope functions. By doing this, we finally obtain a nonlinear algebraic system of $3N$ equations whose $3N + 1$ unknowns are the time-averaged envelope patterns and the breather frequency:

$$\left(\frac{\omega}{\omega_0}\right)^2 \boldsymbol{\xi}_m = -\sum_{j\neq m} c_{mj}\left\{\hat{\boldsymbol{R}}_{jm}(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm}) + \frac{3A^2}{8R_{jm}^2} \times \left[\hat{\boldsymbol{R}}_{jm}\left((5 + 2\beta R_{jm}^2)(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})^3 - 3\Delta\xi_{jm}^2(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})\right) + \Delta\boldsymbol{\xi}_{jm}\left(\Delta\xi_{jm}^2 - 3(\hat{\boldsymbol{R}}_{jm}\cdot\Delta\boldsymbol{\xi}_{jm})^2\right)\right]\right\}. \quad (8)$$

We note that the system of equation (8) is only apparently underdetermined. In fact, one can normalize the displacement pattern $\boldsymbol{\xi}$ by taking any of its $3N$ components as the reference unit length. In practice, we shall keep the DB

[5] To be more quantitative, the Lindemann criterium [52] for avoiding (local) melting in classical systems prescribes $A < A_{\text{melt}} \simeq 0.17 \min_j R_{jk}$.

amplitude $|\boldsymbol{\xi}_m| = A_B$ fixed for a guess mode centred at site m, which means solving for the $3N$ variables $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \varphi_m, \vartheta_m, \ldots, \boldsymbol{\xi}_N, \omega_B\}$, where φ_m and ϑ_m are the azimuthal and polar angles, respectively, of the vector $\boldsymbol{\xi}_m$ and where ω_B is the DB frequency.

Equation (8) constitutes a generalization to arbitrary topology of well-known equations for approximate breather solutions in periodic lattices, which have been shown to produce accurate results in that context [53]. We note that, in analogy to lattice systems, we would have obtained the same set of equations by applying the so-called rotating-wave approximation (RWA), which amounts to truncating the expansion of $\cos^3 \omega t$ to the first harmonic [42]. However, RWA does not set restrictions on the time variation of the envelope functions. Hence, we prefer to follow the time-averaging approach since it reflects more correctly the physical situation. For vanishing amplitude A, we recover the eigenvalue problem of the elastic network—in the limit $A \to 0$, equation (8) gives the normal modes (NM) and the linear spectrum of the system. In expanded notation, the corresponding eigenproblem reads

$$\left(\frac{\omega}{\omega_0}\right)^2 \xi_m^\alpha = -\sum_{\beta=x,y,z} \sum_{j \neq m} \mathbb{K}_{mj}^{\alpha\beta} \xi_j^\beta, \tag{9}$$

where Greek apices indicate spatial directions. The matrix $\mathbb{K}$ is the Hessian of the potential energy:

$$\mathbb{K}_{mj}^{\alpha\beta} = c_{mj} \hat{\boldsymbol{R}}_{mj}^\alpha \hat{\boldsymbol{R}}_{mj}^\beta - \delta_{jm} \sum_{k \neq m} c_{mk} \hat{\boldsymbol{R}}_{mk}^\alpha \hat{\boldsymbol{R}}_{mk}^\beta, \tag{10}$$

δ_{jm} being the Kronecker symbol.

2.1. Solving the equations: the initial guess problem

We now wish to devise a general procedure for determining nonlinear localized solutions of equation (8). We note that, at variance with periodic media, sites in a spatially disordered network are not equivalent. Indeed, the connectivity, i.e. the number of neighbours, as specified by the cutoff distance R_c, varies from site to site and so do the directions of the outgoing bonds, so that each position has a unique local neighbourhood. Consequently, it is natural to expect that the existence of localized solutions and their properties will be subject to unknown positional constraints. In fact, the breathers obtained with the cooling procedure turned out to be site-selective, revealing that only a small fraction of the available sites would *support* a breather spontaneously self-exciting out of the energy drainage process at the surface. In particular, DBs single out special regions of the structure, namely those that are among both the most connected and the most buried ones. Moreover, we also found a high degree of site dependence in frequency–energy dispersion relations, suggesting that distinctive low-energy properties, such as the occurrence and the nature of an energy threshold for localization, might also be spatially modulated.

In line with the above observations, we assume that equation (8) will converge to a breather solution centred at a given site m provided (i) the site m *admits* a breather solution *at all* the given point in parameter space (the parameters of the model being R_c, β and either the DB frequency or its amplitude[6]) and (ii) the initial guess is close enough to a nonlinear periodic solution. Let us stress that we are explicitly assuming that, as a consequence of the non-equivalence of the sites, due to topological disorder, DB families are also indexed by the site at which they are localized. Such an assumption is strongly supported by our previous numerical results, since all DBs obtained proved to be highly localized ones, in fact more localized than any of the harmonic modes [51]. In practice, condition (i) should be regarded as a criterium for convergence within a specified tolerance for the sum of residuals. Importantly, such a criterium must be restricted to a homogeneous, site-independent protocol for identifying the initial guess. In other words, a protocol able to identify magnitudes and direction cosines at each site for the guess displacement field automatically, without arbitrariness, e.g. as what concerns the number of displaced particles, the directions of displacements, etc.

An analysis of the mode patterns obtained in [51] from a principal component analysis of the system trajectories in the quasi-stationary state revealed that such DBs typically concentrate the largest part of their energy on a single residue, leaving only a fraction of the order $1/\mathcal{N}$ on each of its $\mathcal{N}$ neighbours. Furthermore, a closer inspection allowed us to show that the displacement vector of the central residue and those of its neighbours seem to be oriented so as to maximize the associated distortion of the network. All other residues being virtually at rest for a typical DB eigenvector, this also corresponds to maximizing the system potential energy (2). The above observations suggest the following protocol for computing the initial guess to be fed to system (8).

Suppose we wish to solve equation (8) for a DB of given amplitude A_B at site m. The $3N$ unknowns are then $3N - 1$ time-averaged displacement coordinates and the frequency ω. Alternatively, we could as well fix the frequency and solve for the pattern and the amplitude. We start with the network in its equilibrium configuration. The first step consists in drawing at random the azimuthal and polar angles of the displacement vector $\boldsymbol{\xi}_m$, with the constraint $|\boldsymbol{\xi}_m| = A_B$, so as to uniformly sample the ensemble of all vectors centred at site m with a fixed modulus A_B. Typically, looping a number $\mathcal{O}(10^3)$ of times, we get a satisfactory convergence of the maximum strain direction at site m with a relative error of 1–2%. We then repeat the same operation sequentially at all the $\mathcal{N}(m)$ neighbouring sites, in an ascending order with respect to the bond distances $R_{jm} (j = 1, 2, \ldots, \mathcal{N}(m))$, with the only difference that the magnitude of the displacements is also varied. Following the numerical results of [51], we take as an initial guess for the magnitudes $|\boldsymbol{\xi}_j| = A_B/(\mathcal{N}(m) + 1)$ for all $\mathcal{N}(m)$ neighbours. The calculation proceeds site by site in such a way that, when optimizing the displacement at site j, the *shells* closer to the central residue are kept fixed in the previously determined optimal configurations. We coin this procedure the sequential-maximum-strain (SMS) method.

An illustration of how the SMS protocol performs is given in figure 2, where we show a scatter plot of the site

[6] By that, we shall always mean the maximum of the displacement field $A|\boldsymbol{\xi}_m|, m = 1, 2, \ldots, N$, where $\{\boldsymbol{\xi}_m\}$ is the calculated time-average DB displacement pattern.

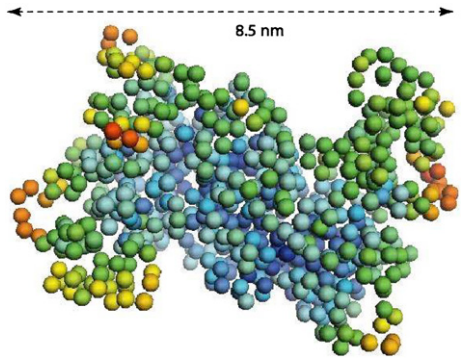

Figure 1. Structure of dimeric citrate synthase (PDB code 1IXE). Only α-carbons are shown, as spheres in a color scale corresponding to the crystallographic *B*-factors, from smaller (blue) to larger (yellow) fluctuations.

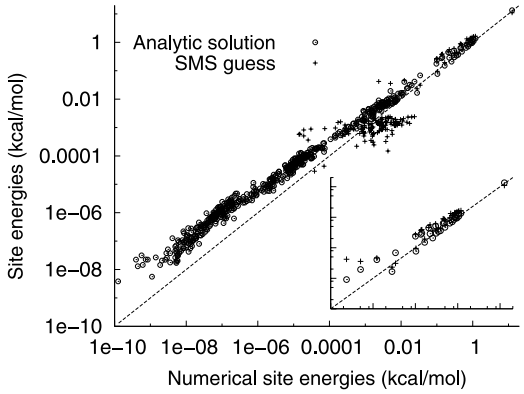

Figure 2. Discrete breathers at site THR 208 A of citrate synthase. A comparison of the site energies of the analytic breather solution and of the SMS pattern with a numerical DB obtained through the cooling process with amplitude $A_B = 0.929$ Å; see [51]. The dashed line is the plane bisector. The inset displays the site energies of all 32 neighbours of the central residue. Parameters of the NNM are $R_c = 10$ Å, $\beta = 1$ Å^{-2}.

potential energies of a numerical solution from [51] versus the corresponding SMS guess vector as well as the analytic solution for the case of the site associated with Threonine 208, in monomer A of citrate synthase (see figure 1). Note the wide separation of energies (logarithmic scale) between the central site and the neighbouring sites, as a consequence of the localized character of highly energetic DBs. As a further quantitative confirmation that SMS-based analytical breathers reproduce well those obtained during our previous cooling simulations, we note that the frequency of the analytic solution (equation (8)), also shown in figure 2, is $\omega_B = 115.3$ cm^{-1}, as compared to that determined by the numerics, namely $\omega_B = 114 \pm 1$ cm^{-1}. However, despite the excellent predictive power of the SMS-fed protocol, the lowest site energies of the numeric solution tend to be smaller with respect to the analytic solution. Since the breather is exponentially localized, these correspond to sites located far from the DB centre, which are likely to be close to, if not on, the protein surface. Although the effect is a small-amplitude one, it seems to be an interesting consequence of the surface cooling protocol itself, whereby breathers would

Table 1. Statistics of DB self-localization in dimeric Citrate Synthase, as found in our previous cooling simulations; see [51]. The probability of appearance p is the fraction of localization events out of a pool of 500 cases. Note that the slight asymmetry of citrate synthase has significant effects on the statistics of events in the two monomers.

	Monomer A		Monomer B	
Rank	Residue	p (%)	Residue	p (%)
1	THR 208	20.4	THR 192	17.8
2	ALA 209	11.1	ALA 196	12.5
3	ALA 196	4.5	THR 208	9.1
4	ALA 212	2.5	ALA 209	5.3

appear further *squeezed* into the interior of the structure so as to minimize its amplitude in the surface regions where the protein is in contact with the zero-temperature bath.

3. Results: analytical discrete breather solutions in NNMs of proteins

In the previous section we have discussed the general features of NNMs of proteins, along with a strategy for obtaining approximate analytic solutions. Since we are interested in computing localized modes whose properties are expected to be site dependent, we have devised a homogenous protocol for obtaining a well-defined guess for a DB at a given site. To be more explicit, the SMS method clarifies ambiguous points related to computing an initial guess for the DB pattern such as the number of displaced residues, the magnitude and directions of their initial displacements and so forth, the algorithm automatically taking care of these choices in compliance with a general requirement. Furthermore, we have shown that the pattern computed through the SMS procedure is remarkably close to the numerical DB obtained by surface cooling. In this section, we shall discuss the properties of the localized vibrations calculated from equation (8), starting from SMS guesses.

3.1. Insight from the cooling simulations: high-energy DBs

We have seen that SMS guesses approximate well the patterns of the self-localizing DBs obtained by surface cooling in [51]. It turns out that the analytical solutions that we compute starting from an SMS guess preserve, and even improve, such agreement throughout the whole energy range spanned by the numerical solutions. This point is illustrated in the case of citrate synthase in figure 3. The dispersion relations for the four DBs with the largest probabilities of occurrence in monomer A, as observed in cooling simulations (see table 1), are extremely well reproduced by our analytical approach. The spatial patterns are well reproduced too. They can be measured through the locality index

$$L = \frac{\sum_{i\alpha} \xi_{i\alpha}^4}{\left[\sum_{i\alpha} \xi_{i\alpha}^2\right]^2}, \tag{11}$$

where $\xi_{i\alpha}$ is the $\alpha(x, y, z)$ coordinate of the ith atom in the given displacement field $\boldsymbol{\xi}$. So, our theoretical calculations

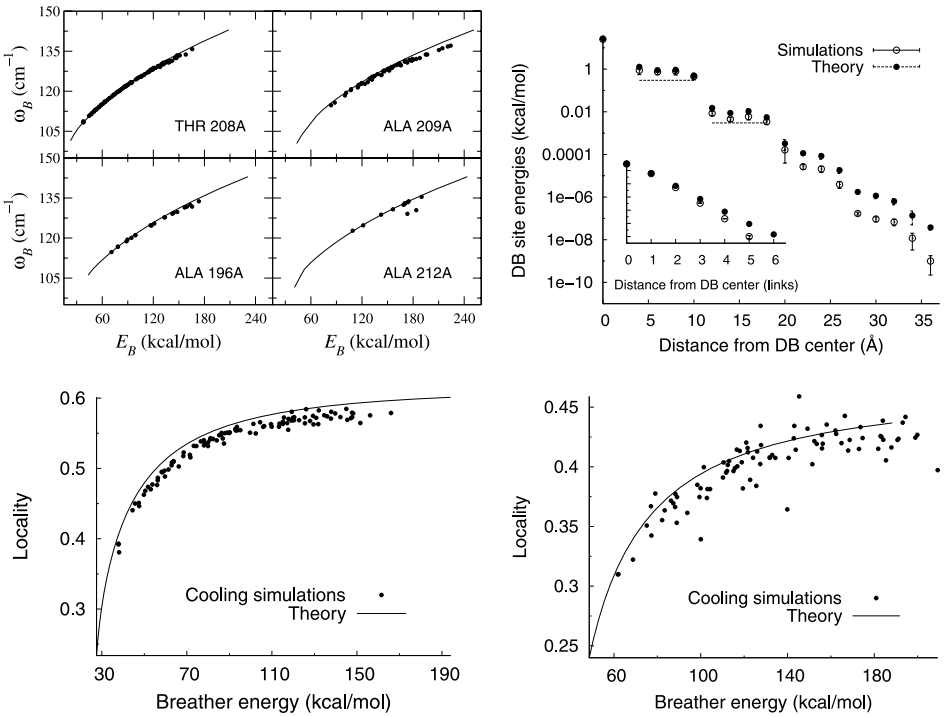

Figure 3. Comparison between analytical and numerical DBs, as obtained in cooling experiments with citrate synthase. Upper panels: numerical (circles) and analytical (solid lines) dispersion relations for DBs self-localizing on monomer A (left) and site energies versus distance from the central site (right) for the DB at THR 208 A with amplitude $A_B = 0.929$ Å. The inset shows the same energies versus average distance, now expressed in terms of the number of links from the central site. The y-axis units are the same as in the main plot. Lower panels: locality measures as defined in equation (11) for DBs localized at THR 208 A (left) and THR 192 B (right). Note that in the left panel, for example, the theory is only about 2% off the numerical results. Parameters of the NNM are $R_c = 10$ Å, $\beta = 1$ Å^{-2}.

confirm the strong site-to-site variability of the DB dispersion relation, already spotlighted in [51].

As a further instructive comparison between the analytical solutions computed from SMS guesses and those obtained through cooling simulations, figure 3 also shows how site energies vary as a function of the distance from the central site of the DB (upper right panel). In particular, it shows that local energies decrease exponentially for sites farther and farther away from the central site, as expected for a discrete breather. However, a closer inspection of the curves reveals that site energies, at least those in the first coordination shells, are sub-organized in plateaus. In other words, there are relatively ample (about 5 Å) intervals of distances from the central site where sites share more or less the same local energies. This behaviour suggests that it is the number of links (bonds) between two sites rather than their distance that matters. Indeed, a plot of site energies as a function of their separation from the central site, expressed in units of *links*, i.e. the distance in the sense of graph theory, confirms the validity of this conjecture (see the inset of figure 3). Here, the graph is the one defined by the connectivity matrix c_{ij} and the distance between two nodes (the sites) is the smallest number of links that have to be followed in order to go from one node to the other. As a matter of fact, by measuring distances in such units, the exponential decay of site energies is recovered.

The numerical breathers analysed in [51] correspond to simulation runs (the majority) where a single object gathered nearly all the system energy. Moreover, as a criterium for selecting mere DBs, we explicitly required that the energy collected be larger than the initial energy per particle in the thermalized state. In other words, we selected and studied high-energy DBs *only*. Hence, the comparison illustrated in figure 3 shows that the SMS scheme is surely a good procedure for computing guess modes for analytical DBs of relatively high energy. As shown hereafter, the situation gets more complex when we turn to the low-energy portions of the dispersion relation.

3.2. Low-energy breathers: does an energy gap exist?

In the context of regular lattices, interesting predictions have been formulated as to the presence of an energy gap in the excitation spectrum of discrete breathers, i.e. a finite energy threshold that has to be overcome in order to create a DB. It has been shown that, depending on the spatial dimension and type of nonlinearity in the inter-particle potentials, a finite threshold may, or may not, exist (for systems of an infinite size) [54, 55]. As shown hereafter, in the context of spatially disordered systems, such a question has to be formulated not only in terms of the *existence* of a gap, but also in terms of its *nature*.

In general, for Hamiltonian systems, discrete breathers occur in one-parameter families, indexed e.g. by their energy E_B, their frequency ω_B or alternatively their amplitude A_B (in

the language of ansatz (6)). In practice, in the presence of hard-type nonlinearities, when the DB frequency is lowered towards the linear band edge frequency ω_E, the breather amplitude may or may not approach zero.

In an ordered medium, if DBs of arbitrary small amplitudes exist at all, they can only appear in the limit $\omega_B \to \omega_E$, as a direct consequence of the constraint of non-resonance with any of the linear modes, ω_E being the edge frequency. In this case, a detuning exponent z can be introduced [54] and standard perturbation theory used to determine both the exponent and the coefficient in the relation

$$|\omega_B - \omega_E| = cA_B^z. \tag{12}$$

In general, it is possible to speculate that if DBs of arbitrary low amplitude exist in a discrete system, then they emerge from linear edge modes by means of a tangent bifurcation [55]. Moreover, this hypothesis can be used to calculate explicitly the energy at which such a bifurcation occurs. For example, in the case of periodic lattices, it has been shown that the bifurcation energy either vanishes or is asymptotically finite in the limit of a large system size, depending on the spatial dimension and on the nature of the potentials [54, 55]. Hence the conclusion that breather excitation spectra may, or may not, display a finite energy gap, depending on the nature of the studied system. It is important to stress that in an ordered system the bifurcation of the nonlinear edge mode marks a symmetry breaking, the emerging DB mode being exponentially localized, while the edge linear mode has an extended pattern. The question is, then, how this picture changes when spatial disorder enters the game.

In the context of weakly coupled chains of oscillators with on-site disordered potentials, it has been shown that low-amplitude breathers exist, approaching linear eigenvectors for vanishing amplitudes. Remarkably, the latter patterns are localized due to disorder and, therefore, there is no symmetry breaking and the DBs approach the linear modes *continuously*. Moreover, there are DBs originating not only from the edge mode, but also from modes at frequencies below the band edge [56, 57]. These DBs have been coined *intra-band breathers* and their frequencies have been shown to be dense within the gaps separating the frequencies of two consecutive linear modes. Under the general hypothesis that such features do not depend on the type of disorder, as long as (at least) the edge modes are localized, we may conjecture that a similar picture would be observed in the presence of spatial disorder. In topologically disordered systems, such as protein structures, sites are not equivalent and edge modes are intrinsically localized, different modes having their largest displacement in different regions of the structure. Hence, we may expect that different families of DBs may exist, localized at different sites and approaching different edge normal modes for vanishing amplitudes. In particular, we may expect that both the DB frequency and the corresponding pattern smoothly approach asymptotically the values of the corresponding linear modes with no symmetry breaking at all. In any case, such DBs would not show an energy gap in their spectrum. However, linear modes are characterized by a variable degree of localization, only a small fraction of them being strongly localized. In the case of citrate synthase, for example, equation (11) predicts a localization parameter greater than the average value plus two standard deviations for a mere 4% of the modes. Therefore, also in accordance with [56, 57], we can guess that the sites on which localized normal modes are centred will host zero-gap DBs, continuously approaching the corresponding linear patterns as their amplitude is reduced towards zero, but will it be the case for a generic site?

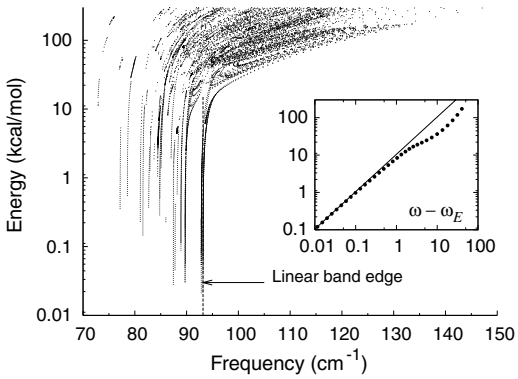

Figure 4. Dispersion relations for analytical DBs at all sites in HIV-1-protease (PDB 1A30). For the sake of clarity, only energies below 300 kcal mol^{-1} are shown. The inset shows a plot of energy versus frequency detuning from the edge normal mode, for the DB centred at site ILE 85A, the site with the largest displacement in the edge normal mode. The solid line in the inset is the first-order perturbative prediction obtained through the Lindstedt–Poincaré method. Parameters of the NNM are $R_c = 10$ Å, $\beta = 1$ Å^{-2}.

In order to answer the above questions and test our conjectures, we shall proceed as follows. First, we look for localized solutions centred at a given site and parametrized by the amplitude in the sense of ansatz (6)[7]. We fix $A_B = A_0$ at the site of choice and use the SMS pattern of amplitude A_0 as an initial guess. Then, we progressively decrease the value of A_B in small steps. If the sum of residuals of the optimal solution found for equation (8) is lower than a specified small tolerance (close to machine precision), we record a DB solution and calculate its physical properties. Then such a solution is used as a guess for calculating the DB for the following value of A_B. If the chosen amplitudes are sufficiently closely spaced, this algorithm enables us to follow a given DB family centred at the site of choice [55]. Besides the DB (potential) energy E_B and frequency ω_B, we also calculate the localization parameter, as defined in equation (11).

The results of the above-described calculations performed for all sites of a small dimeric enzyme (HIV-1 protease) are shown in figure 4. The plot shows the dispersion relation for all DBs found, that is, all points in the parameter space where a solution of system (8) could be found (to machine precision). The figure shows that intra-band breathers can exist

[7] In principle, in order to solve equation (8) for a guess centred at a given site, we could either fix the parameter A and solve for the DB pattern $\{\xi\}$ and frequency ω_B, or fix the latter quantity and determine the pattern and amplitude parameter. For exploratory purposes, fixing A appears the more natural choice also in view of the SMS strategy for computing the initial guess, whereby an SMS pattern is found at a given value of A. However, we have also checked that the same zeros of equation (8) would be found for identical choices of parameters through the alternative strategy.

within the framework of protein NNMs, i.e. localized modes of nonlinear origin with frequencies lying within the gaps of the linear spectrum. More precisely, we see that a limited number of linear frequencies act as *accumulation points* for the frequencies of DBs, which approach asymptotically the corresponding normal mode in the limit of zero amplitude. The inset of figure 4 provides a clear-cut illustration of this phenomenon, in the case of the DB approaching the edge mode. In real space, the DB pattern also continuously approaches the NM pattern. We repeated similar calculations for different proteins, obtaining analogous results. The rationale behind such observations can be resumed in the following two conjectures.

Conjecture 3.1. *Let us consider the site the most involved in a given normal mode, i.e. the one whose displacement is the largest in the NM pattern (hereafter simply the NM site). If a discrete breather centred at a given site m can be found at arbitrary small values of its amplitude* A_B*, its pattern will asymptotically tend to* $\boldsymbol{\xi}^{[m]}$*, that is, the normal mode for which site m is the NM site. Correspondingly, its frequency will also approach the corresponding linear frequency* $\omega^{[m]}$ *in the limit* $A_B \to 0$*. If the linear frequency* $\omega^{[m]}$ *lies below the edge, the breather surely exists in the frequency interval* $[\omega^{[m]}, \omega^{[m+1]}]$*, and may or may not exist for frequencies above* $\omega^{[m+1]}$*.*

As a matter of fact, there are sites that are *NM sites* of more than one normal mode. In this case, starting from one of such NM sites, we find that DBs asymptotically approach the normal mode with the highest frequency.

The above conjecture provides a solid interpretative framework for what we shall call zero-gap breathers, i.e. DBs that may be excited at arbitrary small energies. The opposite inference, however, is not always true, which means that not all DBs centred at a given NM site approach the corresponding normal mode in the low-amplitude limit. There may simply be NO low-amplitude limit at a given site. In that case, at a given value of the amplitude, the DB solution shifts from the selected site (that is, the site we started from at amplitude A_0) to another site. Note that, in general, the destination site need not necessarily be one of the neighbours of the original site. In the following, we will refer to phenomena of the like as *jumps*. This brings us to formulate our second conjecture.

Conjecture 3.2. *Let us suppose that a value of the amplitude at a given site m has been reached where a jump has occurred at another site n, i.e. the largest amplitude in the DB patterns has shifted from site m to n. If, after the whole structure has been scrutinized, low-amplitude breathers are never recovered at site m as the destination site of jumps observed from other sites during the whole analysis (covering all sites in the protein), the DBs centred at site m feature an energy gap in the spectrum, i.e. they exist but for energies higher than a finite threshold. In this case, an energy gap exists as a direct consequence of the impossibility of exciting small-amplitude breathers in the first place.*

Note that, in the case of zero-gap breathers, the possibility still exists that an energy gap may open up at higher energies. In these cases, however, despite being centred at the same site, a change of symmetry occurs in the patterns of the DBs lying above and below the gap. In this case, we shall speak of two different families of breathers.

The above considerations are illustrated in figure 5, for the case of the HIV-1 protease dimer. The upper panels refer to the breather at the NM site of the edge normal mode. No jump is observed, and the DB approaches continuously the edge normal mode with a detuning exponent $z = 2$ [8]. At variance with other properties of DBs in NNMs, we find that all DBs that exist up to arbitrary small amplitudes are characterized by the same detuning exponent $z = 2$. Thus, the value of z seems to be determined by the choice of the force field and not by the topology of the structure. On the other hand, the normalized scalar product p between the DB and the linear mode pattern approaches unity asymptotically when $E_B \to 0$ following a power law, such that $1 - p \propto E_B^2$ (see the upper right panel). Note that this DB is able to harvest a substantial amount of energy even at relatively small amplitudes, thus providing an easily accessible nonlinear energy storage channel. For example, typical energies involved in ATP hydrolysis, the most common fuelling mechanism for molecular machineries, are of the order of 10 kcal mol^{-1}. A DB at site ILE 85A would raise a similar amount of energy at a mere 0.2 Å of amplitude.

The lower plots in figure 5 show a case where jumps are observed. Starting at large amplitudes with an SMS guess centred at site ASN 83B, a DB solution can be followed until an amplitude $A = 0.669$ Å. Beyond that point, no DB solution is found, until the amplitude reaches the value $A = 0.462$ Å. At that point a DB solution is recovered, centred at the same site. However, the lower right panel shows that a change in symmetry has occurred, as revealed by taking the normalized projection onto the third mode from the edge as reference. Overall, there is a 4 kcal mol^{-1} wide interval where no DB seems to exist. Lowering the amplitude further, a new jump is observed at $A = 0.261$ Å. At this stage, the DB jumps to a neighbouring site (ILE 85B, which is about 7 Å away) and from this point it approaches asymptotically the corresponding normal mode (i.e. the one for which ILE 85B is the NM site). Therefore, we conclude that the DB centred at ASN 83B has a second energy gap (which marks the lower bound of its dispersion relation) of 19.03 kcal mol^{-1}, as explicitly reported in the lower left graph. As is clear from the lower right panel in figure 5, the solutions that we find in the frequency range below ω_2 are intra-band breathers. Starting from the SMS guess for ASN 83B at high energies, we are able to continue the solution until a frequency slightly above the band edge. Decreasing the amplitude further beyond this point, we no longer find any DB solution, until the intra-band DB centred at site ILE 85B is recovered, just below the second linear mode. This solution can then be further continued to asymptotically approach the third mode.

3.3. Zero-gap intra-band breathers: crossing harmonic levels

In general, zero-gap DBs approach a given normal mode in the low-energy limit, their frequency tending asymptotically

[8] At small energies, $E_B \propto \omega - \omega_E \propto A_B^z$.

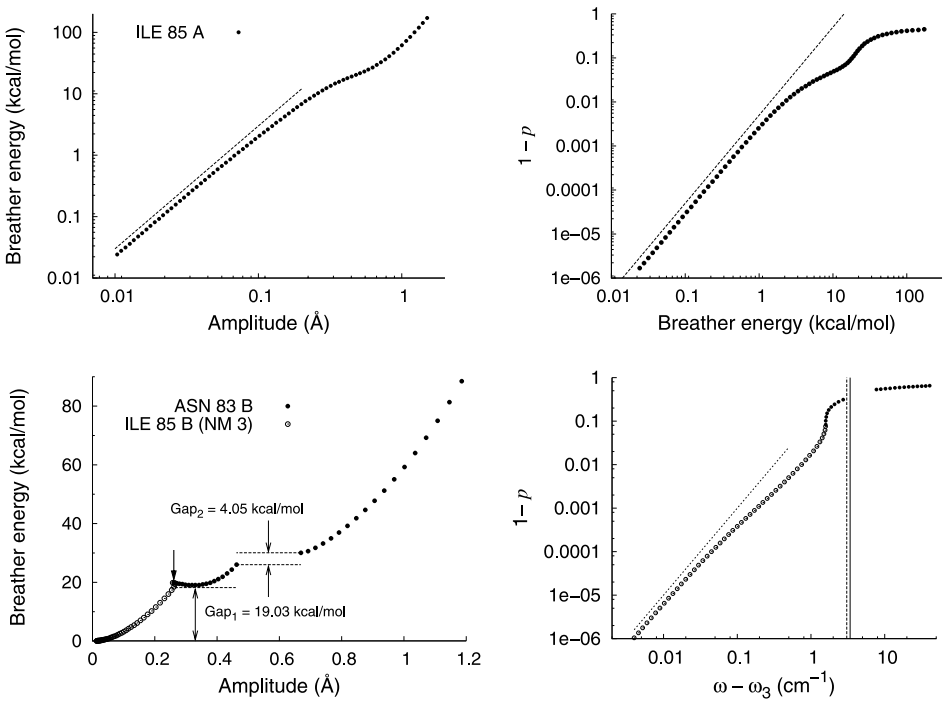

Figure 5. Examples of DB properties, in the case of HIV-1 protease (PDB 1A30). Left panels: plots of DB energy versus amplitude. Right panels: (complementary) projection of the DB pattern on the corresponding asymptotic normal mode as a function of the DB energy (first mode, upper plot) and of the frequency detuning from the mode (third mode, lower plot). The dashed line in the upper left plot is a power law with the detuning exponent $z = 2$. This in the upper right graph is a power law such that $1 - p \propto E_B^2$. The lower graphs portrait a case where a jump occurs, from the original site ASN 83B (filled circles) to the new site ILE 85B (empty circles). In the right graph, the dotted line is a power law such that $1 - p \propto (\omega - \omega_3)^2$, while the two vertical lines mark the edge mode and the second mode of frequency ω_2. Parameters of the NNM are $R_c = 10$ Å, $\beta = 1$ Å^{-2}.

to the corresponding eigenfrequency. However, we found that the behaviour of such solutions may vary to a substantial extent from mode to mode, that is, from site to site. The only exception is the case of DBs tending to the edge normal mode, which show a great regularity. In fact, we find that the same behaviour illustrated in figure 5 is reproduced for all DBs approaching the highest frequency mode in all analysed structures. In view of their robustness, that is, the reproducibility of their characteristics across different structures, such breathers may well happen to play a special role in NNMs of proteins. Concerning DBs originating from modes lying below the edge, we find a much more faceted situation. A deep and thorough understanding of the regularities, if any, displayed by such DBs in relation with their spatial environment and with the edge normal mode patterns and frequencies, surely an instructive study, extends beyond the purpose of the present communication. Here, we limit ourselves to two other demonstrative examples, found when studying the case of dimeric citrate synthase.

If the *target* normal mode corresponds to a frequency below the edge, the DB encounters a number of harmonic levels as its energy decreases. In such cases, the DB can be lost at one, several, or even at all such crossings, as we proceed from large amplitudes and a solution correspondingly recovered in the frequency interval between one mode and the adjacent one. As a matter of fact, we speak of intra-band DBs, as these are solutions that exist *only* within inter-mode gaps. Rigorously, a DB cannot exist at the exact frequency of a normal mode. Hence, adjacent NMs may be thought as boundaries separating different DB families. However, it has been shown in another context that breather frequencies are dense within successive harmonic levels, which thus act as accumulation points [56, 57].

In general, discontinuities seem to occur in the majority of cases as we continue a DB solution through the harmonic spectrum. One such example is reported in the upper graphs of figure 6. Here, the DB eventually approaches the eighth highest frequency mode and the crossings with the seven harmonic levels above can be simply guessed by looking at the dispersion relation. In fact, an inspection of the projection of the DB on the corresponding normal mode pattern allows one to spotlight the crossings, and suggests that these are accompanied by changes in the symmetry of the solution. Thus, we may speculate that such intra-band DBs belong to different families, each of them existing only within the gap between two given frequencies. Remarkably, however, this is not the only behaviour of breathers as they encounter linear frequencies. In fact, some DBs do not seem to be perturbed at all by crossing one or more linear modes. For example, the dispersion relation reported in the lower left graph of figure 6 shows no discontinuities as the DB crosses the first (edge) and second highest modes, as it approaches the third highest mode. The projection of the DB mode on its asymptotic pattern confirms that the breather is able to travel across two harmonic

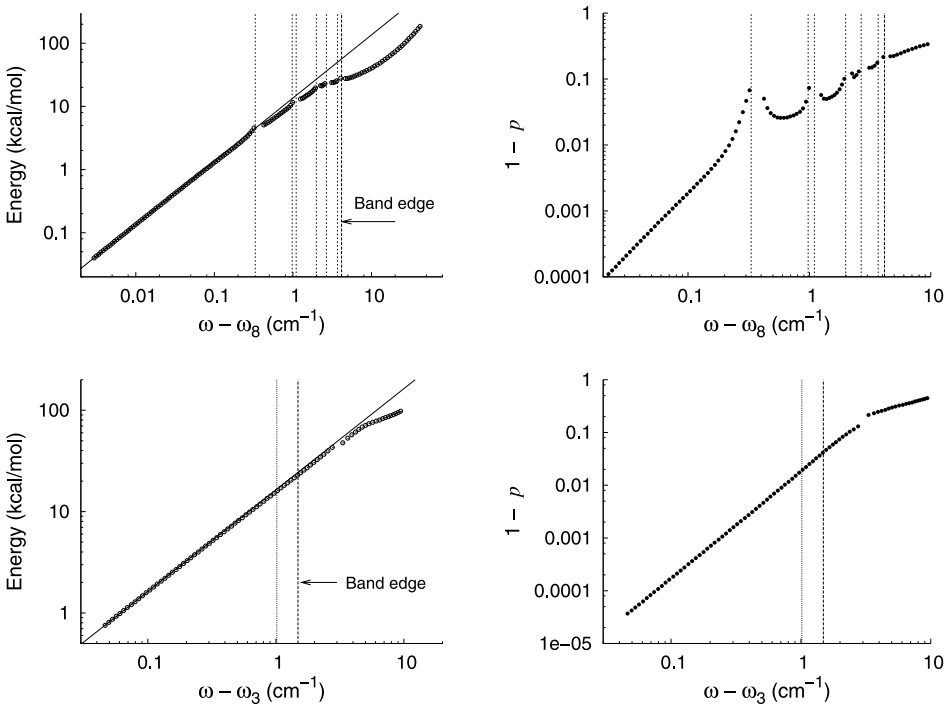

Figure 6. Energy (left) and normalized projection on the corresponding asymptotic normal mode (right) versus frequency detuning for DBs at sites GLY 255 A (upper panels) and ALA 196 B (lower panels) of citrate synthase, approaching the eight and third highest normal modes, respectively. The solid lines in the left panels are the result of the first-order Lindstedt–Poincarè perturbation theory. The vertical lines mark the highest normal modes.

levels virtually undisturbed, with no appreciable perturbation of its pattern.

The observation of an instability when a DB crosses a normal mode close to the band edge can be rationalized by examining the spatial overlap between the two patterns. Since both modes are exponentially localized, if the NM is localized far from the DB core, an instability will not be detectable numerically within machine precision. Conversely, if the DB is centred in a region where the edge mode is also confined, our algorithm will signal the instability. In this sense, our results point to the existence of *spatial selection rules* that govern the range of existence/stability of a DB located at a given site. It should be stressed that a more rigorous analysis based on numerical continuation algorithms such as the one employed in [56, 57] would allow a closer inspection of the bifurcations arising at the unstable level crossings, through e.g. a Floquet analysis of the DB stability. Work in this direction in the framework of NNMs of proteins is currently under way.

Finally, we wish to draw attention to another interesting feature displayed by our DB solutions that was also reported in [56], namely the presence of *tongues* in the dispersion relations of some DBs (see figure 7). These solutions display a distinctive turning at a specific point in the energy–frequency plane, marking a finite energy gap. Such tongues signal the occurrence in the vicinity of the turning point of a change of symmetry in the DB pattern, accompanied by a jump of the DB centre to a neighbouring site. In analogy to what was reported in the case of the systems analysed in [56], we conjecture that such jumps also mark a change in the stability of DBs

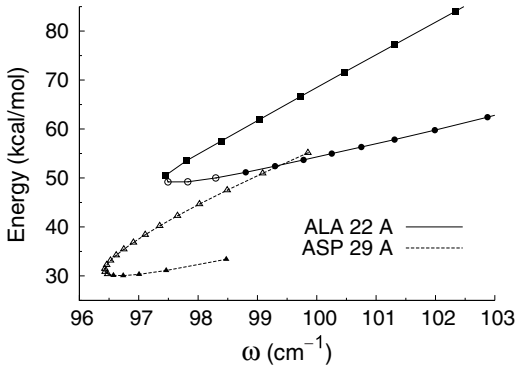

Figure 7. Discrete breathers in HIV-1-protease (PDB 1A30). Magnification of the dispersion relations of two DBs highlighting the occurrence of the bifurcation tongues. The change of symbols along the curves marks a change in the DB symmetry, i.e. the shift of the DB centre on a different site. Parameters of the NNM are $R_c = 10$ Å, $\beta = 1$ Å^{-2}.

through the appearance of a tangential bifurcation. However, our approximate method does not allow us to perform an accurate linear stability analysis of our solutions, and hence we shall defer further statements on that matter to our subsequent studies. We wish to stress, however, that the possibility of following the (putatively) unstable branches is here afforded by our choice to fix the amplitude at a given site and solve for the DB pattern along with its frequency. Had we followed one

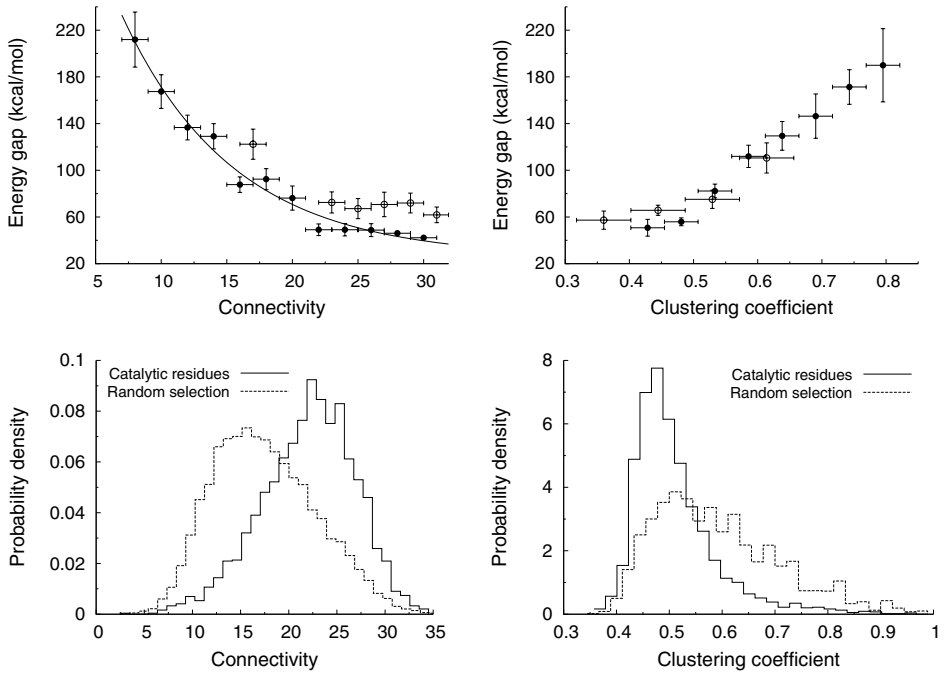

Figure 8. Upper panels. DB energy gaps versus connectivity (left) and clustering coefficient (right) in HIV-1-protease (PDB 1A30, filled circles) and Citrate Synthase (PDB 1IXE, empty circles). The solid line in the left panel is just to guide the eye. The error bars are the statistical errors associated with the binning of energy, connectivity and clustering axes. Parameters of the NNM are $R_c = 10$ Å, $\beta = 1$ Å^{-2}. Lower panels. Distribution of connectivities (left) and clustering coefficients (right) of amino-acid residues involved in enzymatic activity, compared to amino acids of same chemical type, randomly chosen within the same set of enzyme structures.

such breather by fixing its frequency, we would have lost the solution at the turning point.

4. Discussion

The complete analysis of HIV-1 protease reveals that 75.6% of sites are characterized by a non-zero energy gap, while 14.7% have a vanishing gap and 9.7% of the sites do not support any DB at all, in the sense that large amplitude solutions are found to shift to other sites. Of course, we speak here of the lowest energy gap among all possible *holes* that a single DB dispersion relation may display (see again the case of ASN 83B in figure 5), that is, the particular non-zero value of the DB energy below which no solution centred at the studied site is ever recovered. An analysis of a selection (10%) of sites in citrate synthase yields similar figures: 80% of non-zero gap DBs, 17% of zero-gap DBs and 3% of sites that do not seem to allow for any DB at all. It should be again stressed that the presence of sites that do not allow for DBs might be regarded as a consequence of the upper bound chosen for the amplitude, namely $A_0 = 1.5$ Å. It is likely that DBs could be recovered, at some of these sites, using larger values for the amplitude. However, the energy gap of such DBs is expected to be quite large, well over values that are expected to be sustainable for chemically bound molecules (i.e. 100–200 kcal mol^{-1}). In this sense, the presence in a given structure of regions where DBs are not found has to be regarded as a direct consequence of the requirement of chemical meaningfulness.

Since DBs with a finite energy gap seem to be the largest majority, we may ask what is the variability of energy gaps at all such sites, for a given structure, and investigate the relationship between gap magnitudes and the structural properties of local neighbourhoods. In general, gaps turn out to cover a rather wide energy range, between ∼40 and ∼200 kcal mol^{-1}. The lower bound is consistent with the lowest energy DBs that we were able to excite by surface cooling [51]. On the other end, as shown in the upper plots of figure 8, the largest gaps found are clearly associated with regions that are both poorly connected and surrounded by a highly connected neighbourhood. The latter fact can be quantified for a given site i by the fraction of its neighbours that are also interacting, namely its *clustering coefficient*,

$$C_3(i) = \frac{2}{\mathcal{N}(i)[\mathcal{N}(i)-1]} \sum_{j>k} c_{ij} c_{jk} c_{ki}, \qquad (13)$$

where $\mathcal{N}(i)$ is the number of neighbours of the ith residue (that is, its connectivity) and c_{ij} is the (ij) element of the connectivity matrix. On the other side, the most easily excitable DBs are those centred at highly connected sites with poorly connected neighbourhoods. It should be noted that the two above-illustrated correlations are not entirely independent, as there is a general average tendency for highly connected sites to be surrounded by less connected neighbours. However, as illustrated in figure 8, this fact provides a non-trivial reading frame for the gap–structure relationship.

Intuitively, a non-zero gap for DB formation may arise, as the amplitude is lowered, as a consequence of the interplay

between two competing mechanisms. First, obviously, the DB energy has to decrease when its amplitude is decreasing. However, as its energy drops, the DB becomes less localized. Consequently, the connectivities of the most distant regions from the DB centre become crucial, as it may happen that the *integral* over highly connected regions of many exponentially small terms results in a finite energy value, even for vanishingly small DB amplitudes[9]. Incidentally, this is the mechanism by which the spatial dimension enters the game in determining the presence of a finite threshold in a periodic medium [55], since in such cases the spatial dimension determines the degree of connectivity of each site. In the present case, we have seen that a zero energy gap marks DB solutions that exist for arbitrary small amplitudes, their energy also approaching zero as the amplitude vanishes. However, the above argument can still be employed in order to rationalize the variability of energy gaps associated with different structural properties of the DB centre location. Small gaps single out highly connected regions, whose neighbourhood tends to be less tightly connected. By virtue of this correlation (also clearly illustrated by the plot of the energy gaps versus clustering coefficients; see figure 8), it is clear that the balance between the two above-mentioned mechanisms is different for DBs centred at different sites. In particular, if the site is a highly connected one, the DB will harvest less energy in its tail, as compared to a DB localized in a poorly connected region, thus less effectively compensating the drop caused at its centre by the reduction in amplitude. Overall, this may yield a small energy gap. Conversely, a DB located in a loosely connected region will be more effective in counterbalancing the same reduction in energy by appealing to the contribution of its more connected tail regions, hence a higher energy gap.

The above argument provides a qualitative interpretation for the variability displayed by the energy gaps. Interestingly, when the structural properties of a large set of enzyme structures are investigated, such a rationalization also yields a coherent, as well as intriguing, biological picture. Indeed, the cooling simulations reported in [51] very clearly showed that DBs tend to form spontaneously in the stiffest regions of a given enzyme, as identified using a simple indicator of local rigidity. Remarkably, catalytic sites, that is, the vital spots for the initiation of an enzymatic activity, are also often found in such regions [58, 51]. As an extension of such results, the lower panels in figure 8 report the distribution of connectivities and clustering coefficients for known catalytic sites in a set of 833 enzymes from the 2.1.11 version of the catalytic site atlas [59]. As a comparison, we calculated the same distributions for a random selection of residues of the same chemical type from the same data set of structures. Manifestly, catalytic sites tend to be highly connected and, accordingly, also show a marked tendency to have less inter-connected environments. In this sense, these conclusions fit within the picture drawn in [51], as there is an obvious positive correlation between the degree of connectivity and stiffness; the more connected a region, clearly the less easily deformable.

From the biological point of view, this observation suggests a straightforward interpretation for the association of small gaps with highly connected sub-domains. Indeed, discrete breathers are more easily excitable in those regions where enzymes perform their activity. Thus, the latter may use DBs in order to lock down for relatively long periods of time the energy released by ongoing chemical reactions, such as ATP hydrolysis. While it is customary to appeal to energy storage mechanisms of chemical origin, we here provide evidences that an additional *mechanical* channel, based on localized vibrations of nonlinear origin, may exist. It is indeed tempting to imagine that the energy stored in a discrete breather could then be used to lower the barrier of a chemical reaction involved in a catalytic process. Because such a mechanism would drastically increase the efficiency of the enzyme[10], it is likely that, if it is possible to implement it in an actual protein structure and to make it work in a cellular environment, evolution has found the way to do it. *En passant*, this may provide an answer to another, long-standing question, namely: why are most enzymes so big, their active site often occupying a single, tiny region of the whole structure? Our results suggest that this might be so in order for them to have highly connected parts where DBs can easily form and store high amounts of energy for relatively long times, far enough away from the solvent and its dissipative effects.

5. Conclusion and outlook

Using an analytical approach, we have corroborated our previous numerical results—confirming, in particular, that the properties of discrete breathers in nonlinear network models of proteins are site dependent. Moreover, we have shown that DBs of arbitrary low amplitude cannot be excited anywhere in the structure. For a few sites, namely those associated with the largest displacements in the edge normal modes, the DB energy goes to zero as its amplitude approaches zero. However, for the majority of sites a lower bound exists for the allowed DB amplitudes. This, in turn, implies that the majority of sites host DBs that exist but for energies higher than a certain site-dependent gap. Remarkably, we have shown that non-zero gaps invariably arise as a consequence of the impossibility of exciting low-amplitude breathers in the first place.

While the profound origin of this puzzling phenomenology is still unclear, it is instructive to recall that similar phenomena arise in one-dimensional systems in the presence of cubic plus quartic nonlinearities, essentially from the requirement of strict convexness of the interaction potentials [61, 62]. In our case, despite the fact that the interaction potential between residues has the same functional form for all pairs, the equations of motion for a given residue in interaction with its neighbours contain quadratic, but also cubic nonlinearities, with coefficients that depend on the topology and on the spatial arrangement of its set of neighbours. Thus, it is tempting to speculate that spatial

[9] Of course, this argument is rigorous only in the limit of an infinite system size. However, it can still be employed to *rank* putative energy gaps in different situations depending on the corresponding connectivity patterns in the DB tail regions.

[10] Note that DB excitation over an energy gap would be a thermally activated process in the presence of a thermal environment [60], and hence even a small reduction factor in the threshold would magnify exponentially the excitation rate.

disorder and nonlinearity might team together, so as to produce a hierarchy of energy gaps in the DB dispersion relations. Further work in this direction is currently under way.

Although interesting *per se*, that is, within the context of complex nonlinear networks, our results may also prove to have a profound biological significance. Indeed, we have shown that DBs can form more easily (small energy gaps) in parts of the structure where connectivities are high and clustering coefficients low. Reciprocally, we have shown that catalytic residues tend to be highly connected and have low clustering coefficients. However, in order to establish a link between these two facts on firmer grounds, it remains necessary to show that, like elastic network models, nonlinear network models of proteins are able to capture, at the coarse-grained level, such key dynamical features of actual protein structures. Work is in progress along these lines.

Glossary

Discrete breathers. Spatially localized, time-periodic solutions of the equations of motion arising in many spatially extended discrete nonlinear systems. The frequency of the internal vibrational degree(s) of freedom lies outside the linear spectrum (within gaps, if any) so that it does not resonate with higher harmonics of the linear modes. More loosely, the same denomination is employed to indicate generic long-lived, localized vibrations with nonlinear frequencies, as arising for example in periodic media from modulational instability of linear edge modes or from surface cooling. Despite strong similarities, the question as to what is rigorously the link between the latter modes of vibrations (often also dubbed *chaotic* breathers) and known exact solutions of the equations of motion remains open.

Surface cooling. Numerical method allowing for spontaneous localization of energy in the form of chaotic breathers in finite nonlinear discrete dynamical systems with free ends. Importantly, the method does not require any preliminary assumptions on the nature of the asymptotic localized modes. In practice, the system is first thermalized and then cooled down by adding a friction term in the equations of motion of the particles sitting at the boundaries. In the case of proteins, friction is put on solvent-accessible amino-acid residues. Provided the initial energy density is above a given threshold (that may well be zero), the system evolves towards a quasi-stationary state where all the residual energy of the system is stored within a handful of sites and exponentially localized far from the boundaries.

Enzyme catalysis. Enzymes are able to catalyze specific chemical reactions, that is, to speed up these reactions by, typically, a factor of 10^9. To do so, chemical reactants bind in a pocket of the structure, the enzyme active site, where dedicated amino-acid residues, the so-called essential residues, are used so as to lower the energy barriers involved in the chemical reaction.

References

[1] Merchant K A, Best R B, Louis J M, Gopich I V and Eaton W A 2007 Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations *Biophys. J.* 534A

[2] Torchia D A and Ishima R 2003 Molecular structure and dynamics of proteins in solution: Insights derived from high-resolution nmr approaches *Pure Appl. Chem.* **75** 1371–81

[3] Anfinrud P A, Lim M and Jackson T A 1996 Structure, dynamics, and function of proteins: New insights from time-resolved ir spectroscopy *Prog. Biophys. Mol. Biol.* **65** SA302

[4] Brooks B R and Karplus M 1985 Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme *Proc. Natl Acad. Sci.* **82** 4995–9

[5] Marques O and Sanejouand Y H 1995 Hinge-bending motion in citrate synthase arising from normal mode calculations *Proteins* **23** 557–60

[6] Perahia D and Mouawad L 1995 Computation of low-frequency normal modes in macromolecules: improvements to the method of diagonalization in a mixed basis and application to hemoglobin *Comput. Chem.* **19** 241–6

[7] Tirion M M 1996 Large amplitude elastic motions in proteins from a single-parameter, atomic analysis *Phys. Rev. Lett.* **77** 1905–8

[8] Bahar I, Atilgan A R and Erman B 1997 Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential *Fold. Des.* **2** 173–81

[9] Hinsen K 1998 Analysis of domain motions by approximate normal mode calculations *Proteins* **33** 417–29

[10] Bahar I and Cui Q (ed) 2005 *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems (CandH/CRC Mathematical and Computational Biology Series* vol 9*)* (Boca Raton, FL: CRC Press)

[11] Kondrashov D, Van Wynsberghe A, Bannen R, Cui Q and Phillips G 2007 Protein structural variation in computational models and crystallographic data *Structure* **15** 169–77

[12] Micheletti C, Lattanzi G and Maritan A 2002 Elastic properties of proteins: insight on the folding process and evolutionary selection of native structures *J. Mol. Biol.* **231** 909–21

[13] Tama F and Sanejouand Y H 2001 Conformational change of proteins arising from normal mode calculations *Protein Eng. Des. Sel.* **14** 1–6

[14] Delarue M and Sanejouand Y H 2002 Simplified normal modes analysis of conformational transitions in dna-dependant polymerases: the elastic network model *J. Mol. Biol.* **320** 1011–24

[15] Krebs W G, Alexandrov V, Wilson C A, Echols N, Yu H and Gerstein M 2002 Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic *Proteins* **48** 682–95

[16] Lu M and Ma J 2005 The role of shape in determining molecular motions *Biophys. J.* **89** 2395–401

[17] Tama F and Brooks C 2006 Symmetry, form, and shape: Guiding principles for robustness in macromolecular machines *Annu. Rev. Biophys. Biomol. Struct.* **35** 115–33

[18] Nicolay S and Sanejouand Y H 2006 Functional modes of proteins are among the most robust *Phys. Rev. Lett.* **96** 078104

[19] Rueda M, Chacon P and Orozco M 2007 Thorough validation of protein normal mode analysis: A comparative study with essential dynamics *Structure* **15** 565–75

[20] Columbus F (ed) 2005 *Soft Condensed Matter. New Research Proc. Natl. Acad. Sci. USA vol 104* pp 1528–33 (New York: Nova Science)
[21] Peyrard M (ed) 1995 *Nonlinear Excitations in Biomolecules* (Berlin: Springer)
[22] Yuriy A K, Manevitch L I and Savin A V 2007 Wandering breathers and self-trapping in weakly coupled nonlinear chains: classical counterpart of macroscopic tunneling quantum dynamics *Preprint* arXiv:0705.1957
[23] Yakushevich L V, Savin A V and Manevitch L I 2002 Nonlinear dynamics of topological solitons in dna *Phys. Rev.* E **66** 016614
[24] Savin A V and Manevitch L I 2003 Discrete breathers in a polyethylene chain *Phys. Rev.* B **67** 144302
[25] Hennig D 2002 Energy transport in alpha-helical protein models: One-strand versus three-strand systems *Phys. Rev.* B **65** 174302
[26] Hennig D 2002 Moving electron-vibron breathers in random protein models *Physica* A **309** 243–67
[27] Falke J J 2002 Enzymology: A moving story *Science* **295** 1480–1
[28] Levy R, Perahia D and Karplus M 1982 Molecular dynamics of an alpha-helical polypeptide: temperature dependance and deviation from harmonic behavior *Proc. Natl Acad. Sci.* **79** 1346–50
[29] Hayward S, Kitao A and Go N 1995 Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis *Proteins* **23** 177–86
[30] Sagnella D, Straub J and Thirumalai D 2000 Time scales and pathways for kinetic energy relaxation in solvated proteins: Application to carbonmonoxy myoglobin *J. Chem. Phys.* **113** 7702–11
[31] Yu X and Leitner D 2003 Vibrational energy transfer and heat conduction in a protein *J. Phys. Chem.* B **107** 1698–707
[32] Woutersen S and Hamm P 2002 Nonlinear two-dimensional vibrational spectroscopy of peptides *J. Phys.: Condens. Matter* **14** R1035–62
[33] Xie A, van der Meer L, Hoff W and Austin R H 2000 Long-lived amide i vibrational modes in myoglobin *Phys. Rev. Lett.* **84** 5435–8
[34] Xie A, van der Meer A F G and Austin R H 2001 Excited-state lifetimes of far-infrared collective modes in proteins *Phys. Rev. Lett.* **88**
[35] d'Ovidio F, Bohr H G and Lindgård P A 2003 Solitons on h bonds in proteins *J. Phys.: Condens. Matter* **15** S1699–707
[36] Leitner D M 2002 Anharmonic decay of vibrational states in helical peptides, coils, and one-dimensional glasses *J. Phys. Chem.* A **106** 10870–6
[37] Leitner D M 2001 Vibrational energy transfer in helices *Phys. Rev. Lett.* **87** 188102
[38] d'Ovidio F, Bohr H G and Lindgård P A 2005 Analytical tools for solitons and periodic waves corresponding to phonons on lennard-jones lattices in helical proteins *Phys. Rev.* E **71** 026606–9
[39] Scott A 1992 Davydov's soliton *Phys. Rep.* **217** 1–67
[40] Archilla J F R, Gaididei Y B, Christiansen P L and Cuevas J 2002 Stationary and moving breathers in a simplified model of curved alpha–helix proteins *J. Phys. A: Math. Gen.* **35** 8885–902
[41] Kopidakis G, Aubry S and Tsironis G P 2001 Targeted energy transfer through discrete breathers in nonlinear systems *Phys. Rev. Lett.* **87** 165501
[42] Flach S and Willis C R 1998 Discrete breathers *Phys. Rep.* **295** 181–264 ISSN 0370-1573
[43] Aubry S 2006 Discrete breathers: Localization and transfer of energy in discrete hamiltonian nonlinear systems *Physica* D **216** 1–30
[44] Rumpf B 2007 Growth and erosion of a discrete breather interacting with rayleigh-jeans distributed phonons *Europhys. Lett.* **78** 26001
[45] Peyrard M 1998 The pathway to energy localization in nonlinear lattices *Physica* D **119** 184–99
[46] Flach S and Mutschke G 1994 Slow relaxation and phase-space properties of a conservative system with many degrees of freedom *Phys. Rev.* E **49** 5018–24
[47] Burlakov V M, Kiselev S A and Pyrkov V N 1990 Computer-simulation of intrinsic localized modes in one-dimensional and 2-dimensional anharmonic lattices *Phys. Rev.* B **42** 4921–7
[48] Dauxois T, Litvak-Hinenzon A, MacKay R and Spanoudaki A (ed) 2004 *Energy Localisation and Transfer in Crystals, Biomolecules and Josephson Arrays (Advanced Series in Nonlinear Dynamics* vol 22*)* (Singapore: World Scientific)
[49] Abdullaev F, Bang O and Sorensen M P (ed) 2001 *Nonlinearity and Disorder: Theory and Applications* vol 45 (Dordrecht: Kluwer)
[50] Rasmussen K O, Cai D, Bishop A R and Gronbech-Jensen N 1999 Localization in a nonlinear disordered system *Europhys. Lett.* **47** 421–7 ISSN 0295-5075
[51] Juanico B, Sanejouand Y H, Piazza F and De Los Rios P 2007 Discrete breathers in nonlinear network models of proteins *Phys. Rev. Lett.* **99** 238104
[52] Hansen J P and R M I 1986 *Theory of Simple Liquids* 2nd edn (London: Academic)
[53] Sandusky K W, Page J B and Schmidt K E 1992 Stability and motion of intrinsic localized modes in nonlinear periodic lattices *Phys. Rev.* B **46** 6161–8
[54] Flach S, Kladko K and MacKay R S 1997 Energy thresholds for discrete breathers in one-, two-, and three-dimensional lattices *Phys. Rev. Lett.* **78** 1207–10
[55] Kastner M 2004 Dimension dependent energy thresholds for discrete breathers *Nonlinearity* **17** 1923–39
[56] Kopidakis G and Aubry S 1999 Intraband discrete breathers in disordered nonlinear systems. i. delocalization *Physica* D **130** 155–86
[57] Kopidakis G and Aubry S 2000 Intraband discrete breathers in disordered nonlinear systems. ii. localization *Physica* D **139** 247–75
[58] Sacquin-Mora S, Laforet E and Lavery R 2007 Locating the active sites of enzymes using mechanical properties *Proteins* **67** 350–9
[59] Porter C T, Bartlett G J and Thornton J M 2004 *Nucleic Acids Res.* **32** D129
[60] Piazza F, Lepri S and Livi R 2003 Cooling nonlinear lattices toward energy localization *Chaos* **13** 637
[61] Aubry S, Kopidakis G and Kadelburg V 2001 Variational proof for hard discrete breathers in some classes of hamiltonian dynamical systems *Discrete Continuum Dyn. Syst.* B **1** 271–98
[62] James G 2001 Existence of breathers on fpu lattices *C. R. Acad. Sci., Paris* I **332** 581–6