How super-localization affects Vibrational Energy Exchange process in proteins

Luca Maggi*

Computational Biomedicine Section, Institute of Advanced Simulation IAS-5
and
Institute of Neuroscience and Medicine INM-9,
Forschungszentrum Jülich, Wilhelm-Johnen-Straße, 52425 Jülich, Germany

and

Department of Physics, RWTH Aachen University, 52078, Aachen, Germany

*) Author to whom the correspondence should be addressed: l.maggi@fz-juelich.de

Abstract. Recent experimental findings on a protein have shown the diffusion of vibrational energy occurs prevalently along non-bonded contacts instead through the backbone interaction, as it might be expected. These results are explained presenting a theoretical picture, supported by computational calculations, that accounts for these different behaviors in vibrational energy exchange process showing the collective motions on the backbone present a "superlocalized" nature as their asymptotic decay with the distance r is proportional to e^{-r^d} with $d\sim1.8$, whereas collective motions associated to non-bonded contacts result simply localized, i.e. $d\sim1$.

I. Introduction

The primary structure of a protein consists of a sequence of different monomers, called residues. Each residue is connected to the adjacent ones via covalent bonds and interacts with all the others through a broad range of weaker non-bonded interactions, which are usually modeled by means of Lennard-Jones and electrostatic potentials [1].
The residue sequence encodes the three-dimensional residues arrangement, the so-called secondary and tertiary structure, which we can refer to as the topology or conformation. The protein topology is a combination of short-range ordered (e.g. alpha helices or beta sheets) and disordered parts, whose different arrangement in the space results into a complex structure. The protein topology does not present any long-range ordering and shares common features with disordered solids[2-3] and fractals[4] [5-6]. The presence of interactions among residues implies the existence of a potential energy. This depends on the residues space arrangement; therefore, each protein topology is associated to a specific value of the potential energy. There exist topologies particularly relevant since they correspond to potential energy minima. In case the entropic contribution is much less relevant than the enthalpic one, which is directly associated to the protein potential energy, those topologies represent stable (or equilibrium) conformations, which proteins tend to maintain. If a protein is in thermodynamic equilibrium, it will visit close-to-equilibrium conformations. Its motion, thus, can be reasonably represented as a "vibration" around a stable conformation and the potential energy can be approximated as harmonic. In other words, if $\boldsymbol{r}_i(t)$ is the instantaneous position of the i-th residue and $\boldsymbol{r}_i^o$ is its equilibrium position, the residue displacement is $\boldsymbol{u}_i(t) = (\boldsymbol{r}_i(t) - \boldsymbol{r}_i^o)$ and, its potential energy (U_i) can be written as a quadratic form $U_i = \frac{1}{2}\sum_{j=1}^{N}(\boldsymbol{u}_i \boldsymbol{K}_{ij}) \cdot \boldsymbol{u}_j$, where N is the total number of residues and K_{ij} is the 3 by 3 matrix containing the coupling constants between the i and j-th residue degrees of freedom. Therefore, the total energy of a protein system, which we refer to as vibrational energy (E_{vib}), can be written as:

$$E_{vib} = \sum_{i=1}^{N} E_{vib}^i = \sum_{i=1}^{N} \frac{1}{2}\left[m_i \dot{\boldsymbol{u}}_i \cdot \dot{\boldsymbol{u}}_i + \sum_{j=1}^{N}(\boldsymbol{u}_i \boldsymbol{K}_{ij}) \cdot \boldsymbol{u}_j\right] \quad (1)$$

where m_i is the mass of the i-th residue. This work investigates the exchange of vibrational energy of i-th residue, E_{vib}^i, among protein residues. The large amount of experimental[7-9], theoretical/computational[10-12] (To cite some) studies addressing this phenomenon, highlights its biophysical relevance. Vibrational energy exchange can be associated and may accompany events related to protein function[13] as for instance conformational changes[14] or allosteric modulation[15-16]. Moreover, its inherent complexity that makes the employment of different theoretical and experimental techniques necessary to deeply comprehend this particular phenomenon. The goal of this study is, therefore, to contribute to this vast scientific effort trying to theoretically explain recently published experimental findings[17]. These studies have shown that the E_{vib}^i exchange ($Evib\,xc$) occurs prevalently among

residues connected by non-bonded interactions (named contacts). This might seem counter-intuitive since $E_{vib}\,xc$ between two residues implies a displacement transfer, as can be deduced from Eq. (1). This transfer should occur easier through stiffer interactions like the bonded ones, i.e. along the so-called backbone. However, we will show the bond strength does not play a central role in this process whereas protein geometrical features, instead, are crucial.

In order to investigate $E_{vib}\,xc$ process, we will analyze, thus, the displacement transfer between two residues. To this aim we will write an explicit expression for the correlation between the i-th and j-th residue displacement. This will help us to identify and recognize the relevance of all the "ingredients" contributing to the displacement transfer and consequently to $E_{vib}\,xc$ process.

II. Displacement Correlation and Normal Mode Localization

The displacement correlation between different residues is defined as $\langle \boldsymbol{u}_i(t)\,\boldsymbol{u}_j(t+\Delta t)\rangle$(where $\langle \dots \rangle$ is the ensemble average). To find an explicit expression for this quantity we start from the equation of motion of a single residue in contact with an external heat bath:

$$m_i\,\frac{d^2\boldsymbol{u}_i}{dt^2} = -\sum_j \boldsymbol{K}_{ij}\,\boldsymbol{u}_j - \mu\,\frac{d\,\boldsymbol{u}_i}{dt} + \boldsymbol{\xi}_i(t) \tag{2}$$

Which is equivalent to a damped oscillator equation of motion, where μ is the viscosity coefficient and $\boldsymbol{\xi}_i(t)$ is a stochastic external force statistically defined as a gaussian process, with a correlation function $\langle \boldsymbol{\xi}_i(t)\boldsymbol{\xi}_j(t')\rangle = C_o\delta_{i,j}\delta(t-t')$ (where C_o is a constant and δ's are delta-functions) and average zero, $\langle \boldsymbol{\xi}_i(t)\rangle = 0$. Including all the residues, the Eq. (2) can be written as:

$$\boldsymbol{M}\,\frac{d^2\boldsymbol{u}}{dt^2} = -\boldsymbol{K}\,\boldsymbol{u} - \mu\,\frac{d\boldsymbol{u}}{dt} + \boldsymbol{\xi}(t) \tag{3}$$

Where $\boldsymbol{M}$ a diagonal matrix containing the masses of all the residues and we have introduced the $3N$-vector $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots \boldsymbol{u}_N)$ and $3N$ by $3N$ matrix $\boldsymbol{K}$, also called Hessian matrix, composed by the submatrices $\boldsymbol{K}_{ij}$. We now consider the equation for an overdamped oscillator, disregarding the term $\boldsymbol{M}\,\frac{d^2\boldsymbol{u}}{dt^2}$. This will not modify the final solution features we are interested in, as it will be evident in the following, but it will greatly simplify its analytical calculation. Introducing also $\tilde{\boldsymbol{u}} = \boldsymbol{M}^{-\frac{1}{2}}\,\boldsymbol{u}$, Eq. (3) can be recast as:

$$\mu\,\boldsymbol{M}^{-1}\,\frac{d\tilde{\boldsymbol{u}}}{dt} = -\boldsymbol{D}\,\tilde{\boldsymbol{u}} + \boldsymbol{\xi}\,\boldsymbol{M}^{-\frac{1}{2}}. \tag{4}$$

Where $\boldsymbol{D} = \boldsymbol{M}^{-\frac{1}{2}}\boldsymbol{K}\,\boldsymbol{M}^{-\frac{1}{2}}$, is the so-called dynamical matrix . This is a real-valued symmetric matrix, its set of eigenvectors $\{\boldsymbol{v}_n\}$, the so-called normal modes, are orthonormal and it is convenient to expand $\tilde{\boldsymbol{u}}$ on this basis, $\tilde{\boldsymbol{u}}(t) = \sum_{n=1}^{3N} c_n(t)\,\boldsymbol{v}_n$. Introducing this expansion in Eq. (4) and multiplying by $\boldsymbol{v}_m$ both sides we have:

$$\sum_n \mu\,(\boldsymbol{M}^{-1}\,\boldsymbol{v}_n)\cdot\boldsymbol{v}_m\,\frac{dc_n}{dt} = \sum_n -(\boldsymbol{D}\,\boldsymbol{v}_n)\cdot\boldsymbol{v}_m\,c_n + (\boldsymbol{\xi}\,\boldsymbol{M}^{-\frac{1}{2}})\cdot\boldsymbol{v}_m\,. \tag{5}$$

The matrices $\boldsymbol{M}^{-1}$ and $\boldsymbol{M}^{-\frac{1}{2}}$ can be projected on the normal mode basis set $\{\boldsymbol{v}_n\}$. On this set these matrices can be approximated as diagonal unitary matrix multiplied by either $\bar{m}^{-1}$ or $\bar{m}^{-\frac{1}{2}}$ respectively, where $\bar{m}$ the average mass of a protein residue. This approximation is clarified and justified in Supplementary Materials. Therefore, Exploiting this simplification and the normal mode orthonormality we can write $3N$ equations for the coefficient $c_n(t)$:

$$\gamma\,\frac{dc_n}{dt} = -\omega_n^2\,c_n + \tilde{\xi}_n \tag{6}$$

Where ω_n^2 are the eigenvalues of $\boldsymbol{D}$, $\gamma = \frac{\mu}{\bar{m}}$ and $\tilde{\xi}_n = (\boldsymbol{\xi}\,\boldsymbol{M}^{-\frac{1}{2}})\cdot\boldsymbol{v}_n$. It should be noticed $\tilde{\xi}_n$ still preserve the same property as $\boldsymbol{\xi}_i(t)$, since $\langle\tilde{\xi}_n(t)\tilde{\xi}_m(t')\rangle = C_o'\delta_{n,m}\delta(t-t')$, where $C_o' = {C_o}/{\sqrt{\bar{m}}}$, and $\langle\tilde{\xi}_n(t)\rangle = 0$. Assuming $c_n(0) = c_n^0$ as initial condition the solution of Eq. (6) is:

$$c_n(t) = c_n^0\, e^{-\lambda_n t} + \int_0^t d\tau\, e^{-\lambda_n (t-\tau)}\, \tilde{\xi}_n(\tau) \tag{7}$$

In which $\lambda_n = \frac{\omega_n^2}{\gamma}$. The i-th residue displacement is simply, $\tilde{\boldsymbol{u}}_i(t) = \sum_{n=1}^{3N} c_n(t)\boldsymbol{v}_n^i$, where $\boldsymbol{v}_n^i$ is a vector including the entries of n-th normal mode corresponding to the three degrees of freedom of the i-th residue. Hence, the correlation between a residue i and j can be, thus, written as:

$$\langle \tilde{\boldsymbol{u}}_i(t)\tilde{\boldsymbol{u}}_j(t+\Delta t)\rangle = \sum_{n,m} \langle c_n(t)c_m(t+\Delta t)\rangle\ \boldsymbol{v}_n^i\ \boldsymbol{v}_m^j \tag{8}$$

Noticing that $\langle c_n^0\, \tilde{\xi}_m(\tau)\rangle = 0$, thus:

$$\langle c_n(t)c_m(t+\Delta t)\rangle = \left[\langle c_n^0\, c_m^0\rangle\ e^{-(\lambda_n+\lambda_m)\,t} + C_o'\delta_{n,m}\int_0^t d\tau\, e^{-(\lambda_n+\lambda_m)(t-\tau)}\right] e^{-\lambda_m \Delta t} \tag{9}$$

Introducing Eq. (9) in (8) and solving the integral we get:

$$\langle \tilde{\boldsymbol{u}}_i(t)\tilde{\boldsymbol{u}}_j(t+\Delta t)\rangle = \sum_{n,m}\left[\langle c_n^0\, c_m^0\rangle\ e^{-(\lambda_n+\lambda_m)\,t} - \frac{C_o'\delta_{n,m}}{(\lambda_n+\lambda_m)}\left(1 - e^{-(\lambda_n+\lambda_m)t}\right)\right] e^{-\lambda_m \Delta t}\ \ \boldsymbol{v}_n^i\ \boldsymbol{v}_m^j \tag{10}$$

Finally, since we are considering close-to-equilibrium systems we assume $t \gg \frac{1}{(\lambda_n+\lambda_m)}$, therefore Eq. (10) reads:

$$\langle \tilde{\boldsymbol{u}}_i(t)\tilde{\boldsymbol{u}}_j(t+\Delta t)\rangle = \sum_m \left[-\frac{C_o'}{2\lambda_m}\right] e^{-\lambda_m \Delta t}\ \ \boldsymbol{v}_m^i\ \boldsymbol{v}_m^j \tag{11}$$

The displacement correlation between the i-th and j-th residue is a sum of decaying exponential functions where λ_m's are the decaying coefficients and the inner product $\boldsymbol{v}_m^i\ \boldsymbol{v}_m^j$ "weights" these functions depending on the specific couple of residues i and j we are considering. This means the normal modes values corresponding to each residue affects the correlation and, thus, the displacement transfer. Since normal modes are normalized there exist two extreme cases. A normal mode can be zero everywhere but on one residue, in this case we talk about complete localization. Conversely, we refer to a complete delocalization when a normal mode assumes the same value for all the protein residue. Therefore, the displacement correlation can be different depending on the degree of normal mode localization, with a minimum value in the case of complete localization. Investigating the localization of normal modes, thus, turns out to be essential to gain information about the displacement transfer and, ultimately, on $E_{vib}\,xc$ process.

III. Backbone and Contacts normal modes localization Property

In the following we will study, hence, this particular normal mode property. However, since the main goal of this study is to investigate differences in $E_{vib}\,xc$ process between backbone and contacts, we aim at finding possible differences between them related to normal modes localization. Therefore, we split the dynamical matrix $\boldsymbol{D}$ as a sum of two matrices, one corresponding to backbone $\boldsymbol{D}_{back}$ and the other one to the contacts $\boldsymbol{D}_{con}$, $\boldsymbol{D} = \boldsymbol{D}_{back} + \boldsymbol{D}_{con}$. This can be accomplished as $\boldsymbol{D}$ includes only pairwise interactions coupling constants, thus, it is easy to distinguish from those associated to the backbone, which are, given a residue i, the $i-1$ and $i+1$-th. Whereas, instead, all the other constants belong to contacts (See Fig.1). Moreover, the absolute value between backbone and contacts also should be different, however, as we will show in the following this will not affect the localization property.

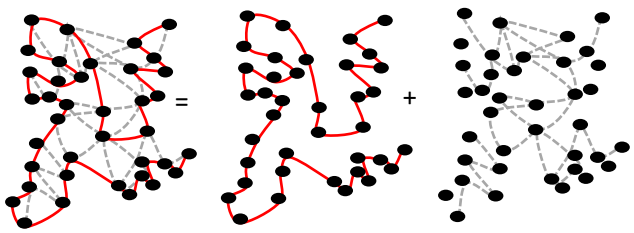

Figure 1: Protein representation in our model. The solid red and the dashed grey lines represents covalent backbone and non-sequential residue bonds, which connect residues (black spheres). The entire structure is made up by the sum of these two contributions

We will analyze separately, thus, the normal mode sets of $\boldsymbol{D}_{back}$, $\{\boldsymbol{v}_{n,back}\}$ and $\boldsymbol{D}_{con}$, $\{\boldsymbol{v}_{n,con}\}$. This is equivalent to study normal modes of particular systems in which either the backbone interactions or the contacts are neglected. Although, those systems are unphysical, they are instrumental to understand and highlight the differences between the backbone and contacts related to the normal modes localization and, thus, to the $Evib\,xc$ process.
In this work we computed the sets $\{\boldsymbol{v}_{n,con}\}$ and $\{\boldsymbol{v}_{n,back}\}$ for 15 different proteins (see Supplementary Materials) and quantify their global localization calculating the so-called participation ratio (see below).
Therefore, we need first to calculate $\boldsymbol{D}_{back(con)}$ for all the proteins considered. To accomplish this goal, in this work, we employed an approximated model called Anisotropic Network Model[18] (ANM). ANM is a further development of the Gaussian Network Model[19], which already has been shown to give reasonable results regarding normal mode localization[20] . Despite its approximated nature, thus, we expect to obtain feasible estimation of the quantities under investigation. Furthermore, its easy technical implementation allows us to apply this model to a large number of different proteins (see Supplementary Materials), since only the knowledge of stable protein conformations is required. In the following we will outline the main aspects of this method and its exploitation in this work.
In ANM the potential energy of the protein is defined as:

$$U(\boldsymbol{r_1}, \boldsymbol{r}_2 \dots \dots \boldsymbol{r}_N) = \sum_{ij} U_{ij}(|\boldsymbol{d}_{ij}|) = \sum_{ij} \eta_{ij}(|\boldsymbol{d}_{ij}| - |\boldsymbol{d}^0_{ij}|)^2 \tag{12}$$

where $\boldsymbol{d}_{ij}$ and $\boldsymbol{d}^0_{ij}$ are the distances between the i-th and j-th residue at each time and at the equilibrium, respectively. Contributions coming from residues distant more than 15 Å are assumed 0. η_{ij} is a constant coupling the i-th and the j-th residue with the following values:

$$\eta_{ij} = \begin{cases} \eta_{back} & if\ \ j = i \pm 1 \\ \frac{\eta_{back}}{10} & otherwise \end{cases} \tag{13}$$

where η_{back} is an arbitrary value. This choice of η_{ij} is the only slight difference from the ANM as presented by Atilgan et al.[18]. This comes from the idea to distinguish the backbone and the contacts contributions in the protein potential energy also according to their bond strength other than their sequence position. The difference in the bond energy between these two interactions can be assessed to be about one order of magnitude [1]. U is a sum of pairwise interactions and it can be split into two parts:

$$U = \underbrace{\sum_{i,j=i\pm1} U_{ij}}_{U_{back}} + \underbrace{\sum_{i,j\neq i\pm1} U_{ij}}_{U_{con}} \tag{14}$$

where all the elements U_{ii} are 0. $\boldsymbol{K}$ can be calculated as the potential energy hessian matrix (∇^2U) evaluated in the equilibrium positions ($\boldsymbol{r}^o_i$), for the backbone ($\boldsymbol{K}_{back}$) and the contacts ($\boldsymbol{K}_{con}$) separately. Finally knowing the mass of each protein residue and, thus $\boldsymbol{M}$, we can compute the corresponding dynamical matrices $\boldsymbol{D}_{back}$ and $\boldsymbol{D}_{con}$, and the corresponding sets of normal modes $\{\boldsymbol{v}_{n\,con}\}$ and $\{\boldsymbol{v}_{n\,back}\}$. We can now quantify the localization of $\{\boldsymbol{v}_{n\,back}\}$ and $\{\boldsymbol{v}_{n\,con}\}$ employing the participation ratio (P_n), defined as [21]:

$$P_n = \frac{1}{3N}\left(\sum_{\alpha=1}^{3N} |v_n^\alpha|^4\right)^{-1}. \tag{15}$$

Where α is a specific degree of freedom of a residue. P_n assumes well-distinguishable values in case of complete (de)localized normal modes. Indeed P_n = 1 if the n-th normal mode is totally delocalized, i.e. if all the degrees of freedom of the system contribute the same to that normal mode, which corresponds to identical values v_n^α for any α, whereas a normal mode is, instead, completely localized (all v_n^α equal to zero except for one α) $P_n = 1/3N \ll 1$. Therefore, the participation ratio, roughly, tells us the percentage of degrees of freedom significantly involved in a normal mode. P_n can be computed for each normal mode n. Here we focus on its probability distribution for $\{\boldsymbol{v}_{n\,back}\}$ and $\{\boldsymbol{v}_{n\,con}\}$ sets, disregarding the association of each P_n to a particular normal mode.

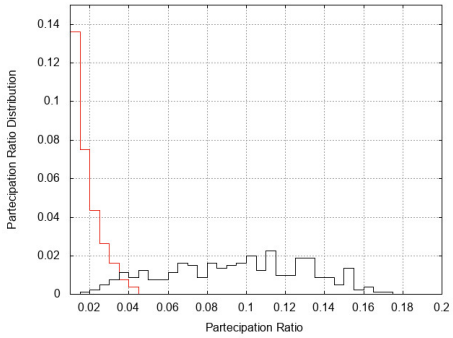

Figure 2: Participation Ratio Distribution vs the participation ratio for $\{\boldsymbol{v}_{n\,back}\}$ (red) and $\{\boldsymbol{v}_{n\,con}\}$ (black)

Fig. 2 shows the P_n distribution, as a result of the average over 15 different protein topologies. The P_n distributions point out two relevant features:

i. A significant difference between the localization of the two sets of normal modes is present. The most extended normal mode in $\{\boldsymbol{v}_{n\,back}\}$'s has a P_n about four time smaller than the most extended one in the $\{\boldsymbol{v}_{n\,con}\}$ set. This agrees with the experimental findings mentioned above about the difference in $Evib\,xc$. Normal modes associated with the backbone are more localized and, thus, less prone to exchange $Evib$. However, it does not clarify the reason of this difference.

ii. Both $\{\boldsymbol{v}_{n\,back}\}$ and $\{\boldsymbol{v}_{n\,con}\}$ turn out to be fairly localized, namely the largest values of P_n are below 20% and 5% for $\{\boldsymbol{v}_{n\,con}\}$ and $\{\boldsymbol{v}_{n\,back}\}$, respectively. This means the single normal modes involve a small only number of degrees of freedom.

It should be point out that varying the values of η_{ij} does not affect at all these features. This means they are totally related to the protein topology despite the nature of the chemical bond connecting residues. Regarding (ii), it is noteworthily that similar localization pattern of normal modes has been already reported to occur in fractals media[22]. Besides, it turns out to be not the only common feature these structures share with protein systems. Experimental[23-24] and computational[25] studies have shown that relevant protein descriptors present power-law scaling which can be found also in typical self-similar structure as fractals. For instance, it has been shown that the protein mass M, which is included in a sphere of radius L, scales as $M \sim L^{d_{mass}}$. Similarly, the scaling of the normal mode density of states $\mathcal{N}$ with respect to the frequency ω has been found to be $\mathcal{N} \sim \omega^{d_s-1}$. Reuveni et al. al.[6] brilliantly showed the analogy between the displacement transfer and a random walk on a protein structure, such a random walk presents peculiar features strengthening once again the similarity between protein and fractals as the mean square displacement (MSD) time dependence follows a power-law, $MSD \sim t^{2/d_w}$[6, 26]. The exponent dw is related to d_{mass} and d_s through the relation $d_w = \frac{2d_{mass}}{d_s}$[27]. This parameter can provide valuable information about the displacement transfer process. Indeed, when $d_{mass} = d_s, d_w = 2$, a normal diffusive process is recovered. This is the case of a perfectly ordered systems as, for instance, linear chain or square lattice, made by identical particles interacting harmonically, whose normal modes are completely delocalized[28]. Differently, protein systems present $d_{mass} = 2.5$ and $d_s = 1.6$[13], thus, $d_w = 3.3$. Therefore, the random walk process, analogues to a displacement transfer, is sub-diffusive, and the normal modes which drives the transfer result localized affecting the $Evib\,xc$ process as well.
This localization implies that only a small portion of degree of freedom is significantly involved in the normal modes as already shown by the participation ratio and, thus, the magnitude of the normal mode components v_n^α should decay within the protein structure and we expect this decay to be fast as the most extended normal mode includes less than 20% of the protein degree of freedom. We will show that the study of the normal modes decay will allow us to explain the differences between $\{\boldsymbol{v}_{n\,con}\}$ and $\{\boldsymbol{v}_{n\,back}\}$ pointed out in (i).

IV. Normal Modes Decay

Since v_n^α is associated to a particular residue degree of freedom, which can be located in the three-dimensional space, v_n^α can be thought as a function of the Euclidean space $v_n^\alpha(r)$. In the following, we will derive for a general protein topology an approximate expression for $v_n^\alpha(r)$ as a function of r. For the sake of simplicity, in our

reasoning we will assume a one-to-one correspondence between residues and degrees of freedom, without this affecting the final result. The starting point is the eigenvalue equation for the normal modes:

$$\omega_n^2 \, \boldsymbol{v}_n = \boldsymbol{D} \; \boldsymbol{v}_n \tag{16}$$

We split the dynamical matrix into two parts:

$$\omega_n^2 \, \boldsymbol{v}_n = (\boldsymbol{D_o} + \boldsymbol{\Delta}) \; \boldsymbol{v}_n \tag{17}$$

where $\boldsymbol{D_o}$ is a dynamical matrix of a three-dimensional square lattice of identical particles harmonically interacting through a single coupling constant δ, and $\boldsymbol{\Delta} = \boldsymbol{D} - \boldsymbol{D_o}$. The value for δ, has been chosen to be the average of all the entries of $\boldsymbol{D}$. These particular choices of $\boldsymbol{D_o}$ and δ are necessary to ensure that specific algebraic relations hold (see below). $\boldsymbol{D_o}$ is associated to a perfectly "ordered" system, whose all normal modes are completely delocalized[28]. The matrix $\boldsymbol{\Delta}$ instead includes all the deviation from this ordered system. Starting from Eq. (17) we have:

$$\boldsymbol{v}_n = \left[\boldsymbol{I} - \frac{\boldsymbol{D_o}}{\omega_n^2}\right]^{-1} \frac{\boldsymbol{\Delta}}{\omega_n^2} \boldsymbol{v}_n \tag{18}$$

with I the $3N \times 3N$ identity matrix. We can now define:

$$G(\omega^2{}_n) = \sum_{q=0}^{\infty} \left[\left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)^q\right] \frac{\boldsymbol{\Delta}}{\omega_n^2} \tag{19}$$

In case $\frac{\|\boldsymbol{D_o}\|}{\omega_n^2} < 1$, or equivalently $\|\boldsymbol{D_o}\| < \omega_n^2$, we can write:

$$\left[\boldsymbol{I} - \frac{\boldsymbol{D_o}}{\omega_n^2}\right]^{-1} = \sum_{q=0}^{\infty} \left[\left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)^q\right] \tag{20}$$

Verifying the above conditions requires computing $\|\boldsymbol{D_o}\|$. Since we have chosen $\boldsymbol{D_o}$ to be associated to a square lattice, $\|\boldsymbol{D_o}\| \sim \delta$ [29] Therefore, we need to prove

$$\delta < \omega_n^2 \tag{21}$$

Eq. (21) shows that the value of δ should be upper-bounded by ω_n^2 so that Eq. (20) holds. On the other hand, its value should be comparable with the $\boldsymbol{\Delta}$ entries' magnitude, thus, the decomposition of $\boldsymbol{D}$, as presented in Eq. (17), is plausible. Those conditions narrow the range of values that δ can assume. The best way to verify Eq. (21) is to measure δ and ω_n^2 experimentally, avoiding approximated computational estimates. However, whereas measurements of ω_n^2 can be easily performed and they have been done for a wide range of proteins[30], to the best of our knowledge, δ has never been measured. Therefore, we estimate this value starting from two experimentally known quantities: The average value of protein bond elastic constant k, and the average mass of a protein residue $< m >$. The former has been measured only for the bacteriorhodopsin[31] and its value is around 10^{-1}N/m, which we assume is representative order of magnitude for any protein; $< m >$ has been calculated to be around 111 Da (see Supplementary Materials), namely $1.78 \cdot 10^{-25}$ Kg. Then, the estimate for δ can be obtained as $k/< m >, \sqrt{\delta} \sim 0.75$ Thz, which is comparable with the lowest vibrational frequency experimentally measured in proteins[30]. Hence, according to this assessment Eq. (17) can be considered approximately satisfied for real proteins, and Eq. (18) can be rewritten as:

$$v_n^i = \sum_j \sum_k \left[\sum_{q=0}^{\infty} \left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)^q\right]_{ik} \frac{\Delta_{kj}}{\omega_n^2} \; v_n^j = \sum_j G_{ij} \; v_n^j \tag{22}$$

i.e the component v_n^i is given by the sum of all the components of the normal mode associated to all the other residues j times G_{ij}. Since v_n^i decays there should be a residue "0" associated with the maximum value. Furthermore, as pointed out above, we expect a fast decay as suggested by short participation ratio values. Therefore, we assume, as approximation, the residue 0 to be the only relevant term in the sum, and we rewrite Eq. (22) as:

$$v_n^i \sim G_{i0} \; v_n^0 \tag{23}$$

where:

$$G_{i0} = \sum_k \underbrace{\left[\sum_{q=0}^{\infty} \left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)^q\right]_{ik}}_{g_{ik}} \frac{\boldsymbol{\Delta}_{k0}}{\omega_n^2} \tag{24}$$

where the sum over k runs over all the neighbors of the residue 0, namely all of the residues interacting with it. If we consider i not interacting directly with 0, the element for $q = 0$ in g_{ik} is zero and g_{ik} can be rewritten as:

$$\sum_{q=1}^{\infty} \left\{ \delta_{q,1} \left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)_{ik} + (1 - \delta_{q,1}) \left[\sum_{l_1, l_2 \ldots l_{q-1}} \left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)_{il_1} \left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)_{l_1 l_2} \ldots \left(\frac{\boldsymbol{D_o}}{\omega_n^2}\right)_{l_{q-1} k} \right] \right\} \tag{25}$$

where $\delta_{q,1}$ is a Kroenecker delta. In Eq. (21), the entries of the matrix $\boldsymbol{D_o}/\omega_n^2$ assume the values of either 0 or $\left(\frac{\delta}{\omega_n^2}\right)$, depending on the presence of the interaction between the two corresponding residues. Each interaction, if present, "contributes" with an element $\left(\frac{\delta}{\omega_n^2}\right)$ in the inner sum of Eq. (25), and q is the number of interactions (i.e. residues) involved to connect the residue i to k. Thus, each collection of connecting residues, named "path", will contribute to the sum with an element $\left(\frac{\delta}{\omega_n^2}\right)^q$. Since $\left(\frac{\delta}{\omega_n^2}\right)$ is smaller than one, the dominant term in the sum is associated to the smaller value of q, i.e. the shortest residue path needed to connect i to k ($\mathcal{M}_{ik}$). This means:

$$g_{ik} \sim \left(\frac{\delta}{\omega_n^2}\right)^{\mathcal{M}_{ik}} \tag{26}$$

Therefore, G_{ij} can be recast as:

$$G_{ij} \sim \sum_k \left(\frac{\delta}{\omega_n^2}\right)^{\mathcal{M}_{ik}} \frac{\boldsymbol{\Delta}_{kj}}{\omega_n^2} \tag{27}$$

Multiplying and dividing by δ one gets:

$$G_{ij} \sim \sum_k \left(\frac{\delta}{\omega_n^2}\right)^{(\mathcal{M}_{ik}+1)} \frac{\boldsymbol{\Delta}_{kj}}{\delta} \tag{28}$$

The above sum is dominated by the element with the smallest exponent $(\mathcal{M}_{ik} + 1)$, which corresponds to the shortest path $\mathcal{M}_{i0}$ connecting i to 0:

$$|v_n^i| \sim \left(\frac{\delta}{\omega_n^2}\right)^{\mathcal{M}_{i\mathbf{0}}} |v_n^0| \tag{29}$$

Finally, Eq. (29) can be generalized to all the residues that are $\mathcal{M}$-residues "away" from the residue 0:

$$|v_n^{\mathcal{M}}| \sim \left(\frac{\delta}{\omega_n^2}\right)^{\mathcal{M}} |v_n^0| \tag{30}$$

We can now define a localization length $\xi(\omega_n)$ as:

$$-\frac{1}{\xi(\omega_n)} = \lim_{\mathcal{M}\to\infty} \frac{1}{\mathcal{M}\,\bar{L}} \frac{|v_n^{\mathcal{M}}|}{|v_n^0|} = \frac{1}{\bar{L}} \ln\left(\frac{\delta}{\omega_n^2}\right) \tag{31}$$

where $\bar{L}$ is the average distance between two residues in a protein and $\frac{\ln C}{\mathcal{M}\,\bar{L}}$ doesn't appear as it tends to zero as $\mathcal{M} \to \infty$. Inserting Eq. (31) in Eq. (30) we obtain:

$$|v_n^{\mathcal{M}}| \sim \exp\left(-\frac{\ell}{\xi(\omega_n)}\right) |v_n^0| \tag{32}$$

In the above equation $\ell = \mathcal{M}\,\bar{L}$ is the average minimum path connecting two residues $\mathcal{M}$-residues distant on a protein, also called chemical distance in the context of fractal structures[32] where ℓ is related to the Euclidean distance r through a power law $\ell \sim r^{d_{min}}$. We have calculated ℓ and the corresponding r for different protein topologies and we found this relation to hold, in line with other aforementioned similarities with fractal structure.

Fig. 3 shows the values of those quantities, averaged over 15 different proteins, corresponding to the backbone and the contacts.

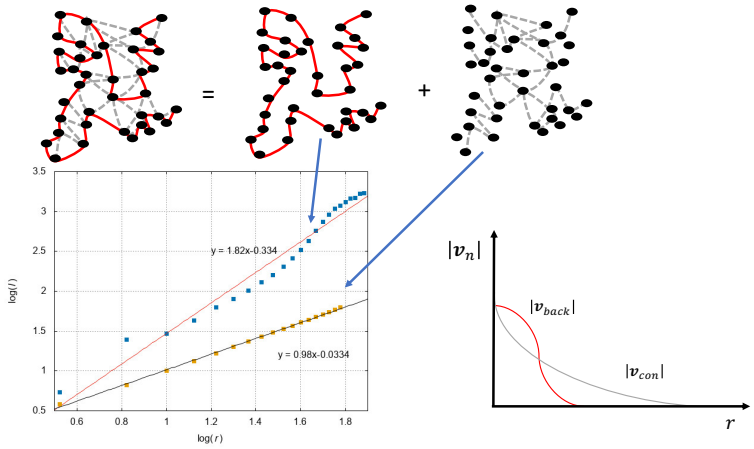

Figure 3: Different localization behavior between $\{\boldsymbol{v}_{back}\}$ and $\{\boldsymbol{v}_{con}\}$. On the left, a log-log plot between the chemical distance l and Euclidean distance r. Blue dots are associated with the backbone whereas yellow dots represent the contacts. Values are averaged over 15 different protein topologies (see Supplementary Materials). On the right, the representation of the different behavior of $|v_n|(r)$ associated to the backbone and to the contacts

V. Conclusion

The value of the exponent (d_{min}) governing the scaling assumes different values in the two cases. Indeed, the same pair of residues is connected differently depending if we consider either the backbone or contacts only (see Fig. 1), therefore ℓ and in turns d_{min} differ. In particular, we have found $d_{min} \sim 1$ for the contacts, whereas $d_{min} \sim 1.8$ for the backbone (Fig. 3). Recasting Eq. (32):

$$|v_n^{\mathcal{M}}| \sim \exp\left(-\frac{r^{d_{min}}}{\xi(\omega_n)}\right) |v_n^0| \tag{33}$$

it is evident that the distance decay of the normal modes associated with the backbone presents a qualitative different behavior with respect to the distance decay of contact-related normal modes (see Fig. 3). This decay law goes to zero faster than an exponential function and this behavior produces more localized normal modes, which are called *superlocalized*. The *superlocalization* of "waves" as molecular vibrations or electrons has been previously theoretically/computationally investigated[22, 33-34] and indirect experimentally evidence observed[23, 35]. According to our theoretical explanation the above mentioned recent experimental findings are the fingerprints of the *superlocalization* of normal modes in proteins.
In this work we have shown how the phenomenon of normal mode *superlocalization* affects the $E_{vib}\,xc$ process making covalent bonds less prone to transfer any kind of vibrations and, thus energy. This feature, which might seem counter-intuitive in the first place, is instead explained showing the corresponding normal modes are *superlocalized*.

VI. Supplementary Materials

See supplementary Materials for a Justification of the approximation employed in Eq. 5 and a complete list of proteins over which the participation ratio and d_{min} values have been averaged.

VII. Author Contributions

L.M. conceived the research. Perform all the calculations and wrote the paper.

VIII. Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 604102 (Human Brain Project) and from the European Unions Horizon 2020 Research and Innovation Programme under Grant Agreement No. 720270 (HBP SGA1). Co-design project 6 (CDP6) is a recipient of the Human Brain Project SGA1 grant. The author gratefully acknowledge the computing time granted through JARA-HPC on the supercomputer JURECA at Forschungszentrum Jülich (Project ID: jias59)

IX. References

1. Finkelstein, A. V.; Ptitsyn, O. *Protein physics: a course of lectures*. Elsevier: 2016.
2. Volkenstein, M. V. Physical Approaches to Biological Evolution. *Nato Adv Sci I B-Phy* **1991,** *263*, 301-315.
3. Ciliberti, S.; De Los Rios, P.; Piazza, F. Glasslike structure of globular proteins and the boson peak. *Phys Rev Lett* **2006,** *96* (19).
4. Dewey, T. G. *Fractals in molecular biophysics*. Oxford University Press: 1998.
5. Enright, M. B.; Leitner, D. M. Mass fractal dimension and the compactness of proteins. *Phys Rev E* **2005,** *71* (1).
6. Reuveni, S.; Granek, R.; Klafter, J. Anomalies in the vibrational dynamics of proteins are a consequence of fractal-like structure. *P Natl Acad Sci USA* **2010,** *107* (31), 13696-13700.
7. Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. Watching hydrogen-bond dynamics in a β-turn by transient two-dimensional infrared spectroscopy. *Nature* **2006,** *444* (7118), 469-472.
8. Ghosh, A.; Ostrander, J. S.; Zanni, M. T. Watching proteins wiggle: Mapping structures with two-dimensional infrared spectroscopy. *Chem Rev* **2017,** *117* (16), 10726-10759.
9. Baumann, T.; Hauf, M.; Schildhauer, F.; Eberl, K. B.; Durkin, P. M.; Deniz, E.; Löffler, J. G.; Acevedo-Rocha, C. G.; Jaric, J.; Martins, B. M. Site-Resolved Observation of Vibrational Energy Transfer Using a Genetically Encoded Ultrafast Heater. *Angewandte Chemie International Edition* **2019,** *58* (9), 2899-2903.
10. Moritsugu, K.; Miyashita, O.; Kidera, A. Vibrational energy transfer in a protein molecule. *Phys Rev Lett* **2000,** *85* (18), 3970.
11. Leitner, D. M.; Yamato, T. Recent developments in the computational study of protein structural and vibrational energy dynamics. *Biophysical Reviews* **2020**, 1-6.
12. Buchenberg, S.; Leitner, D. M.; Stock, G. Scaling Rules for Vibrational Energy Transport in Globular Proteins. *J Phys Chem Lett* **2016,** *7* (1), 25-30.
13. Leitner, D. M.; Straub, J. E. *Proteins: energy, heat and signal flow*. CRC Press: 2009.
14. Bastida, A.; Zuniga, J.; Requena, A.; Miguel, B.; Candela, M. E.; Soler, M. A. Conformational Changes of Trialanine in Water Induced by Vibrational Relaxation of the Amide I Mode. *J Phys Chem B* **2016,** *120* (2), 348-357.
15. Li, G. F.; Magana, D.; Dyer, R. B. Anisotropic energy flow and allosteric ligand binding in albumin. *Nat Commun* **2014,** *5*.
16. VanWart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. Exploring residue component contributions to dynamical network models of allostery. *J Chem Theory Comput* **2012,** *8* (8), 2949-2961.
17. Kondoh, M.; Mizuno, M.; Mizutani, Y. Importance of Atomic Contacts in Vibrational Energy Flow in Proteins. *J Phys Chem Lett* **2016,** *7* (11), 1950-1954.
18. Atilgan, A. R.; Durell, S.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* **2001,** *80* (1), 505-515.
19. Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design* **1997,** *2* (3), 173-181.
20. Wu, Y.; Yuan, X.; Gao, X.; Fang, H.; Zi, J. Universal behavior of localization of residue fluctuations in globular proteins. *Phys Rev E* **2003,** *67* (4), 041909.

21. Edwards, J. T.; Thouless, D. J. Numerical Studies of Localization in Disordered Systems. *J Phys Part C Solid* **1972,** *5* (8), 807-&.
22. Gefen, Y.; Thouless, D. J.; Imry, Y. Localization Effects near the Percolation-Threshold. *Phys Rev B* **1983,** *28* (12), 6677-6680.
23. Stapleton, H. J.; Allen, J. P.; Flynn, C. P.; Stinson, D. G.; Kurtz, S. R. Fractal Form of Proteins. *Phys Rev Lett* **1980,** *45* (17), 1456-1459.
24. Helman, J. S.; Coniglio, A.; Tsallis, C. Fractons and the Fractal Structure of Proteins. *Phys Rev Lett* **1984,** *53* (12), 1195-1197.
25. Leitner, D. M. Energy flow in proteins. *Annu Rev Phys Chem* **2008,** *59*, 233-259.
26. Alexander, S.; Orbach, R. Density of States on Fractals - Fractons. *J Phys Lett-Paris* **1982,** *43* (17), L625-L631.
27. Mosco, U. Invariant field metrics and dynamical scalings on fractals. *Phys Rev Lett* **1997,** *79* (21), 4067-4070.
28. Takeno, S.; Kashiwamura, S.; Teramoto, E. Frequencies of Localized Lattice Vibrations in One-and Three-Dimensional Lattices. *Progress of Theoretical Physics Supplement* **1962,** *23*, 124-140.
29. Montroll, E. W. Dynamics of a Square Lattice .1. Frequency Spectrum. *J Chem Phys* **1947,** *15* (8), 575-591.
30. Benavraham, D. Vibrational Normal-Mode Spectrum of Globular-Proteins. *Phys Rev B* **1993,** *47* (21), 14559-14560.
31. Rico, F.; Rigato, A.; Picas, L.; Scheuring, S. Mechanics of proteins with a focus on atomic force microscopy. *J Nanobiotechnol* **2013,** *11*.
32. Havlin, S.; Ben-Avraham, D. Diffusion in disordered media. *Advances in Physics* **1987,** *36* (6), 695-798.
33. Levy, Y.-E.; Souillard, B. Superlocalization of electrons and waves in fractal media. *EPL (Europhysics Letters)* **1987,** *4* (2), 233.
34. Nakayama, T.; Yakubo, K.; Orbach, R. L. Dynamical Properties of Fractal Networks - Scaling, Numerical Simulations, and Physical Realizations. *Reviews of Modern Physics* **1994,** *66* (2), 381-443.
35. Vanderputten, D.; Moonen, J. T.; Brom, H. B.; Brokkenzijp, J. C. M.; Michels, M. A. J. Evidence for Superlocalization on a Fractal Network in Conductive Carbon-Black Polymer Composites. *Phys Rev Lett* **1992,** *69* (3), 494-497.

X. Figure Captions

Figure 1: Protein representation in our model. The solid red and the dashed grey lines represents covalent backbone and non-sequential residue bonds, which connect residues (black spheres). The entire structure is made up by the sum of these two contributions

Figure 2: Participation Ratio Distribution vs the participation ratio for $\{\boldsymbol{v}_{n\,back}\}$ (red) and $\{\boldsymbol{v}_{n\,con}\}$ (black)

Figure 3: Different localization behavior between $\{\boldsymbol{v}_{back}\}$ and $\{\boldsymbol{v}_{con}\}$. On the left, a log-log plot between the chemical distance l and Euclidean distance r. Blue dots are associated with the backbone whereas yellow dots represent the contacts. Values are averaged over 15 different protein topologies (see Supplementary Materials). On the right, the representation of the different behavior of $|v_n|(r)$ associated to the backbone and to the contacts