Low-frequency harmonic perturbations drive protein conformational changes

International Journal of *Molecular Sciences*

MDPI

*Article*

# Low-frequency harmonic perturbations drive protein conformational changes

**Domenico Scaramozzino [1,*], Gianfranco Piana [1,2], Giuseppe Lacidogna [1] and Alberto Carpinteri [1,3]**

[1] Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; domenico.scaramozzino@polito.it (D.S.), gianfranco.piana@polito.it (G.P.), giuseppe.lacidogna@polito.it (G.L.), alberto.carpinteri@polito.it (A.C.)
[2] Department of Bridge Engineering, Tongji University, 1239 Siping Road, Shanghai, China
[3] Department of Civil and Environmental Engineering, Shantou University, Shantou, China
* Correspondence: domenico.scaramozzino@polito.it; Tel.: +39 011 0904854 (D.S.)

**Abstract:** Protein dynamics has been investigated since almost half a century, as it is believed to constitute the fundamental connection between structure and function. Elastic Network Models (ENMs) have been widely used to predict protein dynamics, flexibility and biological mechanism, from which remarkable results have been found regarding the prediction of protein conformational changes. Starting from the knowledge of the reference structure only, these conformational changes have been usually predicted either by looking at the individual mode shapes of vibrations, i.e. by considering the free vibrations of the ENM, or by applying static perturbations to the protein network, i.e. by considering a linear response theory. In this paper, we put together the two previous approaches and evaluate the complete protein response under the application of dynamic perturbations. Harmonic forces with random directions are applied to the protein ENM, which are meant to simulate the single frequency-dependent components of the collisions of the surrounding particles, and the protein response is computed by solving the dynamic equations in the underdamped regime, where mass, viscous damping and elastic stiffness contributions are explicitly taken into account. The obtained motion is investigated both in the coordinate space and in the sub-space of Principal Components (PCs). The results show that the application of perturbations in the low-frequency range is able to drive the protein conformational change, leading to remarkably high values of direction similarity. Eventually, this suggests that protein conformational change might be triggered by external collisions and favored by the inherent low-frequency dynamics of the protein structure.

**Keywords:** protein dynamics; low-frequency vibrations; modal analysis; elastic network model; harmonic perturbation; conformational change; principal component analysis.

**Citation:** Scaramozzino, D.; Piana, G.; Lacidogna, G.; Carpinteri, A. Low-frequency harmonic perturbations on elastic network models drive protein conformational changes. *Int. J. Mol. Sci.* **2021**, *22*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date



## 1. Introduction

Proteins affect virtually every biological process occurring in the human body [1]. Their correct functioning is pivotal for a variety of tasks, such as delivery of nutrients throughout and across cells, recognition and neutralization of pathogenic bacteria and viruses, providing of suitable strength and rigidity to tissues, activation of signaling pathways and catalytic reactions, etc. [2]. All these activities are performed within the physiological environment and in a highly dynamic fashion. This explains why so much research has been carried out in the last decades regarding protein dynamics and its relationship with the biological functionality. One of the main computational approaches used to investigate protein dynamics is Molecular Dynamics (MD) [3,4]. MD is based on the numerical integration of Newton's laws of motion of the molecular system under scrutiny, subjected to the forces arising from the gradients of the interatomic potentials [5]. Despite the high potential of MD simulations, its applicability to large peptide chains

and protein complexes, especially for the investigation of the large-scale slow dynamics, remains quite elusive and requires cautious analysis of the results.

While trying to overcome the limitations of MD simulations and come up with more simplified approaches which could be of value for a general understanding of protein functionality, it was found that elastic models based on single-parameter Hookean potentials are still able to describe the slow protein dynamics in good detail [6]. These models treat the protein as a network of elastic springs, connecting the atoms whose positions in the reference structure are assumed to be at the equilibrium, around which the thermal fluctuations take place [7–9]. Despite the simplicity of this model, the predicted fluctuations as well as the obtained vibrational frequencies were found in good agreement with those obtained by considering more complex semi-empirical potentials [6]. This discovery paved the way for the development of the coarse-grained elastic network models (ENMs), such as the Gaussian Network Model (GNM) [10–16] and the Anisotropic Network Model (ANM) [17]. The GNM assumes that the protein structure undergoes isotropic fluctuations around the equilibrium position, therefore it predicts the amplitude of these fluctuations and hence it can be identified as a unidimensional model. Conversely, the ANM takes also into consideration the directionality of the expected motion, the protein structure being modelled as an actual three-dimensional network.

The ANM was extensively used for the investigation of protein dynamics for three main reasons. Firstly, the computed fluctuations are found to exhibit a good agreement with the B-factors obtained from crystallographic experiments, therefore providing good estimates of the protein flexibility [17–23]. Secondly, the ANM low-frequency motions are found to describe fairly accurately the directionality of the protein conformational change [20,24–34]. These conformational changes usually occur when the protein switches its three-dimensional shape while performing its biological activity, e.g. during ligand-binding or phosphorylation phenomena, and therefore they are informative of the protein biological mechanism [1]. Thirdly, this model allows to obtain useful insights on the low-frequency dynamics with small computational burden, especially if compared with the more time-consuming MD simulations.

From a Structural Mechanics viewpoint, we have recently shown that the ANM can be seen as a spatial truss elastic model, where the atoms of the protein network can be replaced by frictionless spherical hinges and the Hookean connections by linear elastic bars [35,36]. In the traditional formulation of the ANM, the Hessian matrix of the network is computed and diagonalized to obtain the eigenvalues and eigenvectors. The former are associated with the vibrational frequencies, while the latter identify the mode shapes of vibration [17]. However, since the mass of the protein is not explicitly taken into account in the classical ANM, the eigenvalues are only qualitatively related to the vibrational frequencies. In our previous works, we have also added the explicit mass information, therefore obtaining more quantitative information about the frequencies of vibration via a classical free-vibration modal analysis [35,36]. In particular, in the case of lysozyme we observed that the lowest-frequency modes lie in the sub-THz frequency range, with frequency values of the order of few tens of GHz, in agreement with previous studies [6,37–40].

Another powerful application of the ANM, which has been developed since the last decade, is based on application of perturbations on the protein elastic network, both to probe protein flexibility and conformational changes. Eyal and Bahar [41] developed a methodology that made use of the ANM normal modes to assess the anisotropic mechanical resistance of proteins under external pulling forces. This analysis was able to explain the anisotropy of the mechanical resistance observed from single-molecule manipulation techniques, such as Atomic Force Microscopy (AFM). More recently, we made use of two structural metrics, which are well-known in the field of Structural Mechanics, i.e. compliance and stiffness, to study the flexibility of protein structures under pairwise force application [42]. These metrics enabled to predict the distribution of protein flexibility and rigidity throughout the protein chain and were verified against the

experimental B-factors. Referring to protein conformational changes, Ikeguchi et al. [43] observed that protein transitions can be numerically simulated by evaluating the linear response of the protein reference structure subjected to external forces applied at specific locations. Based on this finding, the Atilgan's group developed the perturbation-response scanning (PRS) technique, where directed forces are applied to the protein structure at single residues and the protein response is calculated and compared with the conformational change observed experimentally [44]. The method was shown to work well for the prediction of a variety of protein conformational changes [45], as well as for the detection of allosteric sites [46]. More recently, the PRS method was used by Liu et al. [47], coupled with an energy-based Metropolis Monte Carlo (MMC) algorithm, in order to simulate the complete closed-to-open transition of the GroEL subunit, induced by directional forces applied at the ATP-binding site.

From what we have reported above, it is evident that, when starting from the knowledge of the reference structure only, protein conformational changes have been analyzed with the ANM by following two separate approaches: (1) evaluating the normal modes of vibration of the reference elastic model, with subsequent comparison between each individual mode shape and the conformational change; (2) applying forces to the protein reference structure and evaluating the response of the network in terms of displacements, with subsequent comparison with the observed conformational change. Fundamentally, approach (1) considers the free-vibration dynamics of the protein, whereas approach (2) focuses on the static response of the protein structure under external forces. In this work, we put the two approaches together, therefore applying forces to the protein structure in a dynamic fashion. In this way, we exploit the main ideas behind both approaches: (1) conformational changes might be favored by the intrinsic protein dynamics along its low-frequency modes of vibration [24,29]; (2) conformational changes might be triggered by external perturbations [43,45].

In particular, we apply external harmonic perturbations, randomly distributed in the space-domain but with a well-defined frequency content in the time-domain, to the protein ANM. The equations of motions are numerically solved, by considering mass, viscous damping and elastic stiffness contributions, in order to assess the complete time-dependent dynamic response of the protein. The time-history of nodal displacements is then evaluated both in the coordinate space, as well as in the sub-space of principal components (PCs) via the application of Principal Component Analysis (PCA). The obtained time-dependent displacements are then compared to the observed conformational change, in order to find in which conditions these external perturbations are able to drive the conformational change. Results are shown here for lysine-arginine-ornithine(LAO)-binding protein, considering different perturbation patterns, damping coefficients and frequencies of excitation. The results of the analysis reveal that, when the external perturbation is applied in the low-frequency range, the protein structure undergoes a displacement field closely aligned with the observed conformational change, with a remarkably high overlap score (up to 0.95).

## 2. Methodology

In this section we describe the fundamentals of the ANM, starting from the calculation of the natural modes of vibration (Section 2.1), the evaluation of the protein response under external perturbations applied in a static fashion (Section 2.2), and finally how the two approaches can be put together in order to retrieve the complete protein response under time-dependent external perturbations (Section 2.3).

### *2.1. ANM fundamentals and free-vibration analysis*

The Anisotropic Network Model (ANM) treats the protein structure as a network of atoms connected by Hookean connections, which are meant to simulate the interatomic interactions in a simplified manner. In its traditional coarse-grained representation, $C_\alpha$

atoms are usually taken as the reference nodes for each amino acid [17]. The positions of these nodes can be taken from the coordinates of the crystal structure, which is available from the Protein Data Bank [48] and assumed to be the equilibrium state of the protein. Each pair of nodes $i$ and $j$ lying at a distance $r_{i,j}$ lower than a certain cutoff threshold $r_c$ are connected by a linear elastic spring, having a spring constant equal to $\gamma_{i,j}$. Commonly employed values of $r_c$ are around 12–15 Å [17]. For each connection among nodes $i$ and $j$, the Hessian (stiffness) matrix $[H_{i,j}]$ can be computed based on the values of the spring constant $\gamma_{i,j}$ and coordinates of atoms $i$ and $j$ [17,35]:

$$[H_{i,j}] = -\frac{\gamma_{i,j}}{{r_{i,j}}^2}\begin{bmatrix} (x_j - x_i)^2 & (x_j - x_i)(y_j - y_i) & (x_j - x_i)(z_j - z_i) \\ (x_j - x_i)(y_j - y_i) & (y_j - y_i)^2 & (y_j - y_i)(z_j - z_i) \\ (x_j - x_i)(z_j - z_i) & (y_j - y_i)(z_j - z_i) & (z_j - z_i)^2 \end{bmatrix}. \quad (1)$$

For a protein network made up of N nodes, the complete 3N × 3N Hessian matrix $[H]$ can be computed as a Boolean combination of the sub-matrices $[H_{i,j}]$, i.e.:

$$[H] = \begin{bmatrix} [H_{1,1}] & [H_{1,2}] & \dots & [H_{1,i}] & \dots & [H_{1,N}] \\ [H_{2,1}] & [H_{2,2}] & \dots & [H_{2,i}] & \dots & [H_{2,N}] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ [H_{i,1}] & [H_{i,2}] & \dots & [H_{i,i}] & \dots & [H_{i,N}] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ [H_{N,1}] & [H_{N,2}] & \dots & [H_{N,i}] & \dots & [H_{N,N}] \end{bmatrix}, \quad (2)$$

where:

$$[H_{i,i}] = -\sum_{j=1, j\neq i}^{N} [H_{i,j}]. \quad (3)$$

At this point, the ANM looks for the eigenvalues and eigenvectors of $[H]$. The former are qualitatively associated with the fundamental frequencies of vibration, while the latter reflect the natural mode shapes of the protein network [17]. Notice that, since the protein is not externally constrained, the first six mode shapes are associated with rigid-body motions (translations + rotations) of the entire molecule at zero-frequency. Thus, the mode shapes accounting for the internal deformability of the protein are related to the subsequent 3N–6 eigenvectors. These vibrational modes have been extensively used to analyze and predict protein conformational changes, and good agreement has often been found, especially for the most collective conformational transitions [24,27–30].

In order to have more quantitative information about the frequencies of vibration, the mass of the structure needs to be included into the calculations [35,36,49–51]. This can be easily done by considering the 3N × 3N mass matrix $[M]$ of the network:

$$[M] = \begin{bmatrix} [M_{1,1}] & [0] & \dots & [0] & \dots & [0] \\ [0] & [M_{2,2}] & \dots & [0] & \dots & [0] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ [0] & [0] & \dots & [M_{i,i}] & \dots & [0] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ [0] & [0] & \dots & [0] & \dots & [M_{N,N}] \end{bmatrix}, \quad (4)$$

where $[M_{i,i}] = m_i[I]$, $m_i$ being the mass of node $i$ (e.g. equal to the actual mass of the $i^{\text{th}}$ amino acid for a $C_\alpha$-only coarse-grained model), and $[I]$ represents a 3 × 3 unitary matrix. The Hessian and mass matrices reported above can then be diagonalized together, following the well-known approach of modal analysis to retrieve the fundamental modes of vibration. This yields the fundamental eigenvalue-eigenvector equation [52]:

$$([H] - \omega_n{}^2[M])\{\delta_n\} = \{0\}, \quad (5)$$

where $\omega_n$ is the angular frequency of vibration associated with mode $n$, and $\{\delta_n\}$ the 3N × 1 vector containing the mass-weighted displacements associated to the $n^{th}$ mode shape. Notice that $\omega_n$ is related to the vibrational frequency $f_n$ by $f_n = \omega_n /2\pi$. As in the traditional ANM, the mode shapes evaluated from Eq. (5) can be used to characterize the low-frequency dynamics of the protein [35,36] and investigate its conformational changes.

In order to obtain fully quantitative information about the natural vibrational frequencies $f_n$, we do not only need to include mass into the model but also to fix properly the values of the spring constants $\gamma_{i,j}$. Multiple choices can be made in this regard: the traditional ANM sets $\gamma_{i,j}$ as a constant value for all connections, i.e. $\gamma_{i,j} = \gamma$ [17], while other ANM-based approaches make these values dependent on the distance between the nodes, i.e. $\gamma_{i,j} \propto r_{i,j}{}^{-p}$ ($p$ = 1 in [35,36], 2 in [20], in the range 0–2.8 in [18]). In this work, we use the traditional ANM convention, i.e. all springs have a unique spring constant equal to $\gamma$. This value can be quantified by comparing the computed fluctuations arising from the normal modes to the experimental ones, which are known as the B-factors [17,35]. B-factors constitute a fingerprint of protein flexibility and can be experimentally retrieved from crystallographic experiments. These experimental values can be compared to the ones obtained from the normal mode calculations, based on the following expression [35,53]:

$$B_i = \frac{8}{3}\pi^2 k_B T \sum_{n=7}^{3N} \frac{\delta_{n,i}{}^2}{\omega_n{}^2}, \quad (6)$$

where $B_i$ represents the B-factor calculated for node $i$ of the network, $k_B$ is the Boltzmann's constant (1.38 × $10^{-23}$ J/K), $T$ is the absolute temperature, $\delta_{n,i}$ is the mass-weighted displacement of node $i$ associated with mode $n$, and $\omega_n$ is the angular frequency of mode $n$. By posing that the mean value of the calculated B-factors matches the mean value of the experimental B-factors, one is able to obtain the value of the spring constant $\gamma$, finally being able to obtain quantitative information about the vibrational frequencies [35,36].

### *2.2. Time-independent response under external perturbations*

As mentioned in the Introduction, the ANM has also been used to predict protein flexibility and conformational changes upon the application of external perturbations to the protein network [41,42,44,45,47]. This is usually done in a static fashion, meaning that dynamic effects are neglected. In this case, the protein response is obtained by solving the following matrix equation:

$$\{F\} = [H]\{u\}, \quad (7)$$

where $[H]$ is the ANM Hessian matrix evaluated from Eq. (2), $\{F\}$ is a 3N × 1 vector of external forces and $\{u\}$ is the 3N × 1 vector of nodal displacements accounting for the protein response. The choice of the force vector $\{F\}$ depends on the specific application. For example, in [42] pairwise pulling forces are applied for each couple of residues $i$ and $j$, whereas in the PRS technique a point force is usually applied at a single node [44,45] or in a localized region [46,47] with a random direction. In any case, the protein response $\{u\}$ is computed by inverting Eq. (7) as follows:

$$\{u\} = [H^{-1}]\{F\}, \tag{8}$$

where $[H^{-1}]$ is the pseudo-inverse Hessian matrix, which can be computed from the 3N–6 eigenvalues $\lambda_n$ and eigenvectors $\{\delta_n\}$ of the Hessian matrix, as follows:

$$[H^{-1}] = \sum_{n=7}^{3N} \frac{1}{\lambda_n}\{\delta_n\}\{\delta_n\}^{\mathrm{T}}. \tag{9}$$

The protein response $\{u\}$ computed through Eqs. (7-9) has been often compared to the observed conformational change, and good agreement has been found in certain cases [44,45,47].

*2.3. Time-dependent response under external harmonic perturbations*

The method developed in the present work can be seen as the generalization of the techniques reported in the previous subsections, in the sense that we consider the dynamics of the system while applying external perturbations, i.e. we apply forces to the protein ANM in a dynamic fashion. In this case, we can write the full equilibrium equations as follows [52]:

$$[M]\frac{\mathrm{d}^2}{\mathrm{d}t^2}\{u(t)\} + [C]\frac{\mathrm{d}}{\mathrm{d}t}\{u(t)\} + [H]\{u(t)\} = \{F(t)\}, \tag{10}$$

where $[M]$, $[C]$ and $[H]$ are the 3N × 3N mass, damping and Hessian matrix of the system, respectively, $\{u(t)\}$ is 3N × 1 displacement vector representing the time-dependent protein response, and $\{F(t)\}$ is the 3N × 1 vector of external time-dependent perturbations. Notice that when no forces act on the system and damping effects are neglected, i.e. $\{F(t)\} = \{0\}$ and $[C] = [0]$, the problem reduces to the free-vibration analysis reported in Section 2.1. Conversely, if inertia and damping forces are neglected and the external perturbations are time-independent, i.e. $[M] = [0]$, $[C] = [0]$ and $\{F(t)\} = \{F\}$, the problem reduces to the calculation of the protein response under external static perturbation described in Section 2.2. If all the terms are taken into account, i.e. $[M] \neq [0]$, $[C] \neq [0]$ and $\mathrm{d}\{F(t)\}/\mathrm{d}t \neq \{0\}$, Eq. (10) allows to obtain the complete time-dependent protein response under external dynamic perturbations.

In this work, we apply an external pattern of forces which is random in the space-domain, but it has a harmonic content in the time-domain. Therefore, the time-dependent force vector can be written as $\{F(t)\} = \{F\}\sin(\omega_F t)$, where $\{F\}$ is a 3N × 1 vector of force components randomly extracted from the uniform distribution $U \sim (-1,1) \times 10^{-10}$ N, and $\omega_F$ is the angular frequency associated to the harmonic excitation, with frequency $f_F = \omega_F/2\pi$. The protein response $\{u(t)\}$ can be decoupled in the space- and time-domain through the well-known relation $\{u(t)\} = [\Delta]\{p(t)\}$, where $[\Delta]$ is the 3N × 3N matrix containing the eigenvectors $\{\delta_n\}$ obtained via modal analysis from Eq. (5), and $\{p(t)\}$ is the 3N × 1 vector of principal coordinates associated with each normal mode $\{\delta_n\}$. In this way, Eq. (10) can be rewritten as [52]:

$$[M][\Delta]\frac{\mathrm{d}^2}{\mathrm{d}t^2}\{p(t)\} + [C][\Delta]\frac{\mathrm{d}}{\mathrm{d}t}\{p(t)\} + [H][\Delta]\{p(t)\} = \{F\}\sin(\omega_F t). \tag{11}$$

Pre-multiplying both sides of Eq. (11) by $[\Delta]^{\mathrm{T}}$, we observe that $[\Delta]^{\mathrm{T}}[M][\Delta] = [I]$ and $[\Delta]^{\mathrm{T}}[H][\Delta] = [\Omega]$, where $[I]$ is the 3N × 3N identity matrix and $[\Omega]$ is the 3N × 3N diagonal matrix containing the 3N natural angular frequencies $\omega_n^2$. At this point, by assuming that

the matrix product involving the damping matrix, $[\Delta]^T[C][\Delta]$, yields a diagonal matrix itself, the 3N-degree-of-freedom problem reported in Eq. (11) can be fully decoupled in a set of 3N single-degree-of-freedom equations, as follows [52]:

$$\frac{d^2}{dt^2}p_n(t) + 2\xi_n\omega_n\frac{d}{dt}p_n(t) + {\omega_n}^2 p_n(t) = \{\delta_n\}^T\{F\}\sin(\omega_F t), \qquad n = 1, \dots, 3N, \qquad (12)$$

where $\xi_n$ is the dimensionless damping coefficient associated with the $n^{th}$ mode of vibration $\{\delta_n\}$. Since we are only interested in the internal deformation of the protein, we can focus on the 3N–6 set of equations related to the non-rigid motions, i.e. $n = 7, \dots, 3N$. Considering underdamped conditions, i.e. $\xi_n < 1$, the general solution of Eq. (12) is:

$$p_n(t) = e^{-\xi_n\omega_n t}\left[A_n\cos(\omega_{d,n}t) + B_n\sin(\omega_{d,n}t)\right] + C_n\sin(\omega_F t + \phi_n), \qquad (13)$$

where the first term in the right-hand side represents the solution of the general integral, which is associated with the decaying damped response, while the latter represents the solution of the particular integral, which is related to the steady-state response due to the external harmonic excitation at frequency $\omega_F$ [52], and $\omega_{d,n}$ is the reduced frequency of mode $n$, i.e. ${\omega_{d,n}}^2 = {\omega_n}^2(1-{\xi_n}^2)$. The amplitude $C_n$ and phase $\phi_n$ of the steady-state response can be written as [52]:

$$C_n = \frac{\{\delta_n\}^T\{F\}}{{\omega_n}^2\sqrt{\left(1-{\beta_n}^2\right)^2 + (2\xi_n\beta_n)^2}}, \qquad \phi_n = -\arctan\left(\frac{2\xi_n\beta_n}{1-{\beta_n}^2}\right), \qquad (14)$$

where $\beta_n$ is the ratio between the frequency of applied excitation $\omega_F$ and the one of the $n^{th}$ normal mode $\omega_n$, i.e. $\beta_n = \omega_F/\omega_n$. The integration constants $A_n$ and $B_n$ depend on the initial conditions of the system. Assuming an initial resting condition, i.e. $p(t=0) = 0$ and $dp/dt(t=0) = 0$, one obtains:

$$A_n = -C_n\sin\phi_n, \qquad B_n = \frac{\xi_n\omega_n A_n - C_n\omega_n\cos\phi_n}{\omega_{d,n}}. \qquad (15)$$

The numerical computation of Eqs. (13-15) allows to obtain the complete temporal evolution of the vector of principal coordinates $\{p(t)\}$, and therefore the vector of time-dependent displacements of the protein nodes $\{u(t)\} = [\Delta]\{p(t)\}$.

### *2.4. Model parameters, comparison between the protein response and the observed conformational change and Principal Component Analysis (PCA)*

Besides the mass, damping and stiffness features of the protein network, the time-dependent response $\{u(t)\}$ depends on the applied force vector $\{F\}$ and frequency of excitation $\omega_F$. As mentioned above, the force vector has been chosen as a random vector, whose components are picked up from a uniform distribution in the range (-1,1) × $10^{-10}$ N. The simulation has been run for 100 different random force patterns, in order to assess the influence of the specific force pattern on the protein response. As for the value of the frequency $f_F$, 500 different values have been used in the range 0.001 THz – 0.5 THz, in order to investigate the influence of the excitation frequency on the response of the elastic network. The mass values have been set based on the actual amino acid atomic weights. The elastic network has been built from the protein crystal coordinates obtained from the PDB, and by adopting a cutoff value $r_c$ of 15 Å. All springs have been set the same stiffness value $\gamma$, which has been defined based on the quantitative comparison between the experimental and numerical B-factors, as explained in Section 2.1. Finally, since the

damping characteristics of the system are not easy to understand, as a preliminary testing condition we worked in the underdamped limit, by carrying out the analysis with three different values of the dimensionless damping coefficients $\xi_n$, i.e. $\xi_n$ = 0.001, 0.01 and 0.1. This value has also been kept equal for all modes. Discussions about the effect of damping and the reasonableness of these choices are addressed below. In conclusion, for a certain protein structure, 150,000 different simulations have been run, by considering 100 random force patterns, 500 different excitation frequencies and 3 different damping coefficients.

After the evaluation of the induced time-dependent displacement field $\{u(t)\}$, this has been compared with the experimentally-observed conformational change. The conformational change can be characterized by a 3N × 1 vector of nodal displacements $\{CC\}$, which is evaluated from the two crystal protein conformations available on the PDB (usually referred to as the "open" and "closed" form of the protein) after superposition [24]. The comparison can be quantitatively assessed by calculating the overlap between the two vectors $\{u(t)\}$ and $\{CC\}$, which is defined as [24]:

$$O(t) = \frac{|\{u(t)\}^{\mathrm{T}}\{CC\}|}{\sqrt{\{u(t)\}^{\mathrm{T}}\{u(t)\}} \cdot \sqrt{\{CC\}^{\mathrm{T}}\{CC\}}}. \quad (16)$$

The overlap defined in Eq. (16) provides a numerical estimate of the alignment between the calculated displacements $\{u(t)\}$ and the conformational change $\{CC\}$. If the two vectors are perfectly aligned, the overlap reaches the maximum value of 1, whereas it provides a value of 0 if the two vectors are orthogonal. In the previous literature, the overlap has been extensively used to assess the directionality correlation between the conformational change and the individual normal modes of vibrations [24,28,29]. In that case, since each normal mode of vibration has a fixed direction over time, the overlap is time-independent. Conversely, in this case, we are comparing the observed conformational change to the time-dependent response of the protein network $\{u(t)\}$, therefore the overlap changes over time as the displacement field evolves in the time-domain. From the analysis of the obtained overlap values, we can infer whether the induced time-dependent protein response is in agreement with the experimentally-observed conformational change.

Principal Component Analysis (PCA) is also used here to investigate the generated ensemble of protein conformations. PCA is a numerical technique widely adopted for dimensionality reduction [54], and it has also been used to evaluate the apparent motions of proteins from a set of experimental crystal structures [33,55]. The input of PCA is a matrix of coordinates $[X]$, with dimension s × 3N, being s the number of available structures and N the number of protein residues. In our case, s is equal to the number of generated protein conformations upon harmonic perturbations. From $[X]$, the covariance matrix $[\Sigma]$ can be calculated as [33]:

$$\Sigma_{i,j} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle, \quad (17)$$

where $r_i$ and $r_j$ represent the X-, Y- and Z-coordinates associated with conformation $i$ and $j$, respectively, and < > stands for the average over all the conformations. The covariance matrix $[\Sigma]$ is then decomposed as:

$$[\Sigma] = [P][\Lambda][P]^{\mathrm{T}}, \quad (18)$$

where $[\Lambda]$ is the 3N × 3N diagonal matrix of eigenvalues and $[P]$ is the 3N × 3N matrix of eigenvectors. Each column of $[P]$ represents a Principal Component (PC), ordered for descending order of its total variance, which is directly proportional to the corresponding

eigenvalue included in [$\Lambda$]. PCA is applied here to reduce the problem dimensionality and study the protein trajectory in the PC sub-space.

## 3. Results and Discussion

In this section, the results are reported for the case of LAO-binding protein, a widely studied protein, known to exhibit two different conformations, i.e. an open form (PDB code: 2lao) and a closed form upon ligand-binding (PDB code: 1lst) [24]. In the Supplementary Material, the results for other three proteins are reported, i.e. maltodextrin-binding protein (PDB codes: 1omp, 1anf), lactoferrin (PDB codes: 1lfh, 1lfg) and triglyceride lipase (PDB codes: 3tgl, 4tgl). The coordinates of the open form are used to build the elastic network model, with a cutoff of 15 Å (Figure 1a). Free-vibration modal analysis is run first, in order to obtain the theoretical B-factors from Eq. (6) and the value of the spring constant $\gamma$, which is found to be equal to 0.10 N/m (~ 0.15 kcal/molÅ$^2$). As a result, the frequency spectrum of the 3N–6 (N = 238) non-rigid modes related to the coarse-grained elastic network is found to lie in the range 0.05 THz – 0.8 THz (Figure 1b), the lowest frequency being equal to 50.9 GHz.

Figure 1c shows the displacement field involved in the open-to-closed conformational change (continuous line). By carrying out the traditional overlap comparison between the displacement field $\{CC\}$ and each individual normal mode $\{\delta_n\}$, it is found that the first non-rigid normal mode, i.e. $\{\delta_7\}$, is the one exhibiting the highest overlap value (0.76), as shown in Figures 1c and 1d. The second low-frequency mode $\{\delta_8\}$ agrees with the conformational change with an overlap of 0.55, while all the higher-frequency modes have lower overlap scores (Figure 1d). From these results, it is clear that the low-frequency modes are strictly related to the observed conformational change, as already reported in the previous literature [24,29].

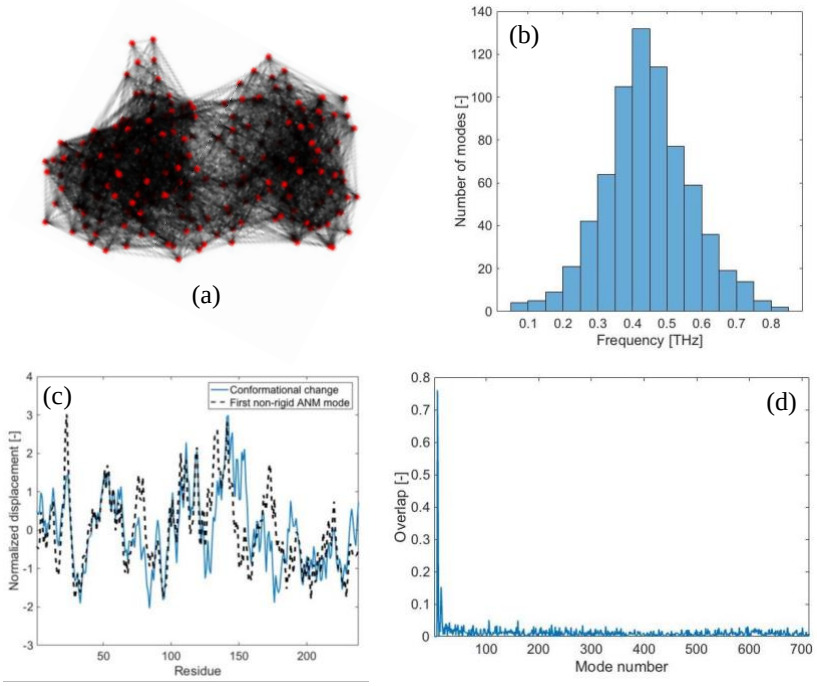


**Figure 1.** LAO-binding protein normal modes: **(a)** elastic network model of the open conformation (PDB code: 2lao), obtained with $r_c$ = 15 Å; **(b)** distribution of vibrational frequencies obtained from free-vibration modal analysis; **(c)** normalized values of the displacements of the open-to-closed conformational change (continuous line) and displacements associated with the first non-rigid normal mode (dashed line); **(d)** overlap values obtained from the comparison of the open-to-closed conformational change to each normal mode of vibration (maximum overlap 0.76).

The results reported above are based on the traditional analysis aimed at evaluating the similarity between the individual mode shapes of the protein structure and its conformational change [24]. What happens if we look at the complete time-dependent protein response upon harmonic random perturbations, as described in Section 2.3? Figure 2 shows the time-dependent response of LAO-binding protein, subjected to the random force pattern reported in Figure 2a, with an exciting frequency of 0.05 THz and a damping coefficient $\xi$ of 0.01. The response is reported in Figure 2b in terms of the global root-mean-squared-deviation (RMSD). The RMSD is a measure of the average displacements of the atoms from the initial position. It can be simply computed as:

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} u_i(t)^2}, \quad (19)$$

being $u_i(t)$ the absolute displacement of the $i^{th}$ node at instant $t$. As can be noticed from Figure 2b, the response of the protein network exhibits a transitory response at the beginning, and then enters a steady-state oscillation approximately from 400 ps onwards. Note that, being the frequency of the external oscillation of 0.05 THz, its period is equal to 20 ps. Moreover, since this frequency value is very close to the natural frequency of the first mode ($f_7$ = 0.051 THz), high amplifications in the response occur, leading to RMSD values of about 20 Å (Figure 2b). On the other hand, if we apply the same force pattern in a static way, i.e. by following the approach reported in Section 2.2, we would obtain a total RMSD of about 2.3 Å. This leads to a dynamic amplification value, evaluated as the ratio between the dynamic RMSD and static RMSD, of about 8.4. Such dynamic amplification factors might also explain why protein vibrations and responses under external forces, which are supposed to be theoretically valid only in the small-amplitude regime, might actually trigger large-scale conformational changes.

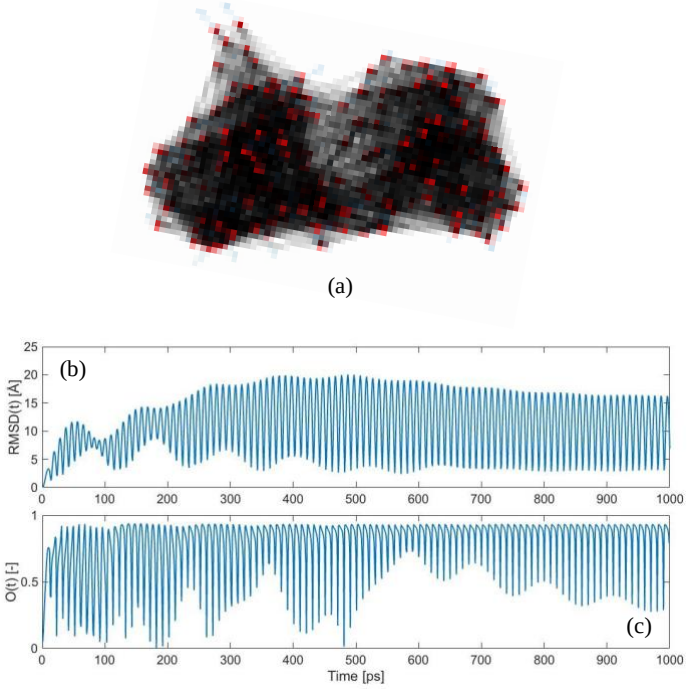


**Figure 2.** Dynamic force application on LAO-binding protein: **(a)** elastic network model of the open conformation, with applied external forces with random directions (shown as blue arrows). Exciting frequency of the external perturbation equal to 0.05 THz and damping coefficient $\xi$ equal to 0.01; **(b)** evolution of RMSD in the time-domain; **(c)** overlap values between the conformational change $\{CC\}$ and the time-dependent displacement vector $\{u(t)\}$ at each instant $t$ (maximum overlap 0.94).

Figure 2c shows the time-dependent overlap, obtained by comparing the calculated displacement field $\{u(t)\}$ with the conformational change $\{CC\}$, as described in Section 2.4. The results are astonishing, since values as high as 0.94 are frequently met. This unequivocally suggests that, even if the applied force pattern is completely random in the space-domain (Figure 2a), its dynamic application is able to drive the protein structure towards the known closed conformation, with remarkably high levels of agreement.

Also, it is interesting to observe how the overlap score is not maintained to these high values constantly, but it keeps oscillating between low and high values. This suggests that the direction of the protein motion $\{u(t)\}$ from the open towards the closed conformation is not linear – see Eq. (16). As a matter of fact, if this motion were linear, we would find a roughly constant value of the overlap throughout the entire simulation. The fact that this does not happen suggests that, while jiggling around its equilibrium position, the protein is sampling a variety of different conformations, among which lies the known closed form. These dynamic jumps between conformations happen in a continuous fashion and involve curvilinear pathways, as suggested here from our overlap calculations and already reported by previous authors [28,56–60].

The complete trajectory of LAO-binding protein upon the force pattern shown in Figure 2a is represented in the Supplementary Movie 1, which is available in the Supplementary Material. In the movie the blue structure refers to the protein conformation generated at each instant $t$ starting from the open form. Conversely, the red structure refers to the known closed form of the protein and it is kept fixed in all frames to help the visualization of the conformational change. As can be seen, after the motion enters in the steady-state regime, the perturbed protein structure keeps oscillating between open and closed conformations. Note that several times the known closed conformation (in red) is reached with high accuracy during the motion. The instants at which this occurs are the ones where high levels of overlap values have been obtained and reported in Figure 2c. As an example, Figure 3a shows the snapshot of the dynamic displacements evaluated at $t$ = 513.5 ps, compared with the displacements of the known open-to-closed conformational change. The overlap value and correlation coefficient between the two displacement fields are 0.935 and 0.904, respectively, showing high level of agreement. Higher than that found by following only the first normal mode of vibration (compare Figure 3a with Figure 1c).

In order to describe more quantitatively the ensemble of generated conformations, PCA has been applied to the set of structures obtained during the trajectory according to Eqs. (17-18). Figure 3b reports the PC score plot of all conformations in the PC1-PC2 sub-space. Note that PC1 and PC2 account for 93.2% and 6.7% of the total variance, therefore they account for 99.9% of the total variance. The black point refers to the open form (pdb: 2lao), the red point to the closed form (pdb: 1lst), while all the points associated with the generated conformations are in blue. A dynamical representation of Figure 3b can be observed in the Supplementary Movie 2, where the time-dependent evolution of the conformations in the PC1-PC2 plot is reported. From the movie and Figure 3b we see that, after a transitory, the steady-state trajectory implies a harmonic motion of the protein around the open form, mostly along the first PC (green arrows in Figure 3b). The information contained in Supplementary Movies 1 and 2 also suggest that the direction of PC1 involves mostly an opening-closing mechanism of the protein. During this harmonic oscillation, the closed conformation (red point in the PC score plot) is closely approached several times throughout the motion.

It is also interesting to observe that, even though we are applying forces at a frequency very close to the first natural mode ($f_F$ = 0.05 THz and $f_7$ = 0.051 THz), the time-dependent displacement field contains the information about the complete dynamics of the system, and not only that of the first natural mode. This can be immediately understood if one looks at the overlap values. By considering the trajectory which would be induced by the first natural mode alone, we would get a 0.74 overlap with the conformational change for the entire trajectory (Figure 1d). On the other hand, applying

forces dynamically excites all modes and eventually leads to a much higher agreement with the conformational change, with $O_{max}$ = 0.94 (Figure 2c and Figure 3a).

Here, it is also important to notice that the methodology developed in this work does not require any *a priori* knowledge of the target conformation. The closed conformation is only used to assess whether the conformations generated by perturbing dynamically the reference structure overlap properly with the target. Other methods have been developed in the existing literature based on the ANM, in order to find intermediate conformations given the two end structures [47,56,57,59,61,62]. Conversely, the methodology presented here relies only on the knowledge of the reference structure and aims at evaluating its dynamic response upon external harmonic perturbations. As a result, the generated conformations do not depend on the target form, but only on the intrinsic dynamics of the reference structure and how it responses to external perturbations. Nevertheless, the conformations generated by following this approach are able to reach the other form of the protein known experimentally with high levels of agreement (see Figure 3a).

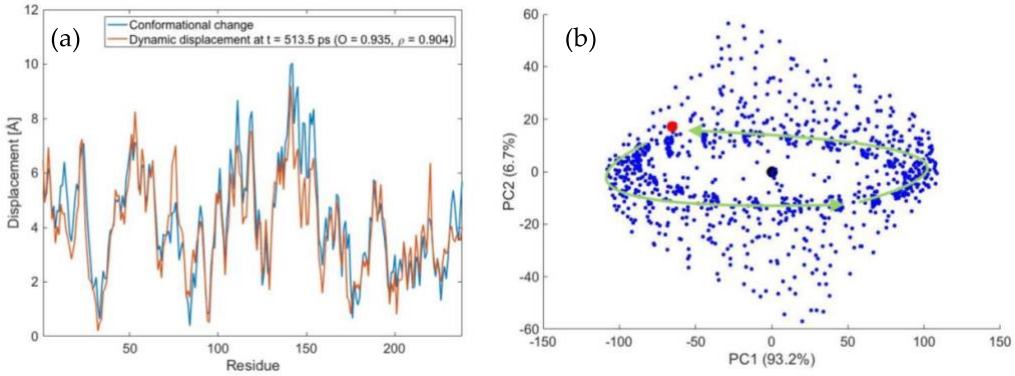


**Figure 3.** Trajectory of LAO-binding protein upon harmonic perturbations to the open form, with $f_F$ = 0.05 THz, $\xi$ = 0.01 and force pattern depicted in Figure 2a: **(a)** comparison between the displacements of the experimental conformational change (blue line) and dynamic displacements evaluated at $t$ = 513.5 ps (orange line); **(b)** PC score plot of all conformations. The blue points represent the ensemble of generated conformations, whereas the black and red point represent the open (pdb: 2lao) and closed (pdb: 1lst) conformation, respectively. The green arrows are associated with the trajectory in the steady-state regime. A scale factor of 0.35 has been applied to the force pattern in order to have comparable values of absolute displacements. See Supplementary Movies 1 and 2 for more details about the protein trajectory in the coordinate and PC space, respectively.

We have briefly mentioned above that the oscillating trend of the overlap values is a fingerprint of the non-linearity of the protein motion. This can also be assessed by a geometrical evaluation, as reported in Figure 4a. For each residue $i$, the coordinate difference between two subsequent conformations at time $t$ and $t + \Delta t$ provides the direction of the instantaneous motion at each time frame $\{\Delta u_i(t)\}$ [59]. By calculating the normalized cosine between vector $\{\Delta u_i(t)\}$ and the direction of motion at time $t$ = 0 ps, i.e. $\{\Delta u_i(0)\}$, as:

$$\cos\left[\theta_i(t)\right] = \frac{\{\Delta u_i(t)\}^{\mathrm{T}}\{\Delta u_i(0)\}}{\sqrt{\{\Delta u_i(t)\}^{\mathrm{T}}\{\Delta u_i(t)\}} \cdot \sqrt{\{\Delta u_i(0)\}^{\mathrm{T}}\{\Delta u_i(0)\}}}, \quad (20)$$

we can geometrically evaluate the non-linearity of the trajectory. If the motion were completely linear throughout the entire simulation, the cosine would only assume values +1 and -1, the former when the protein moves in the positive direction and the latter when it comes back (Figure 4a). Conversely, non-linear motions imply cosine values different from unity, which are also supposed to change during the simulation (Figure 4a). The more frequent the change, the stronger the non-linearity of the motion. Figure 4b shows the values of the normalized cosine for all 238 residues of LAO-binding protein for the entire simulation. As can be seen, the cosine values assume all possible values in the range

between -1 (dark blue) and +1 (bright yellow), suggesting that the motion is non-linear. Moreover, this variation appears to be cyclical, confirming what already observed visually from the Supplementary Movie 1, namely that the motion presents a strong harmonic nature. Figure 4c reports an enlargement of Figure 4b for residues 17-27 (the highly flexible flap of LAO-binding protein in the first domain) in the time range between 480 and 540 ps. This figure shows more clearly that each residue experiences a wide range of cosine values between -1 and +1, therefore the motion is non-linear.

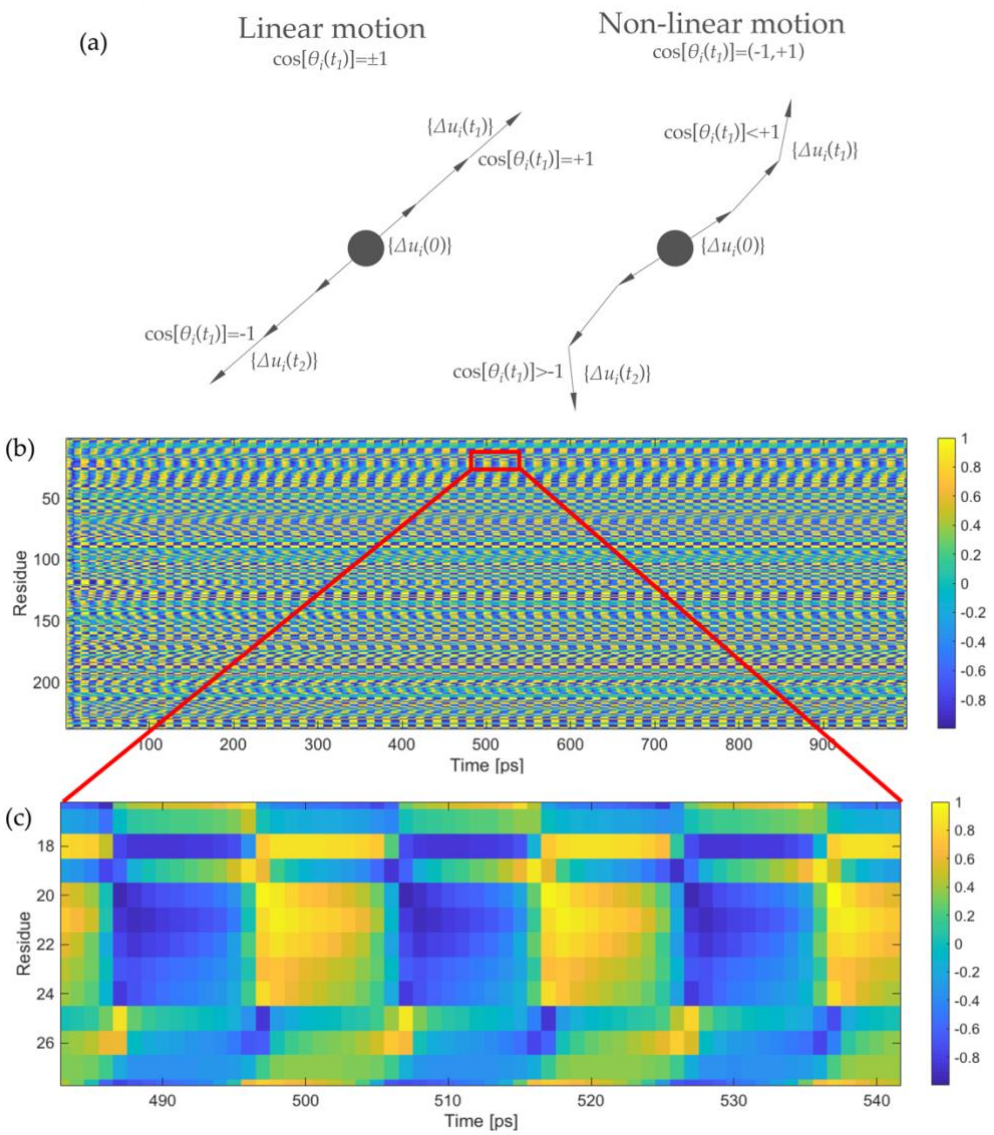


**Figure 4.** Evaluation of the motion non-linearity: **(a)** sketch of the vectors for the numerical evaluation of motion non-linearity; **(b)** values of the normalized cosine reported in Eq. (20) for the 238 residues of LAO-binding protein for the entire simulation; **(c)** enlarged portion of Figure 4b for the segment 17-27 and in the time frame between 480 and 540 ps. The value of the normalized cosines is reported in color scale, from dark blue (cosine equal to -1) to bright yellow (cosine equal to +1).

In the analysis reported above, the protein was perturbed with a specific random force pattern (Figure 2), pulsing at a selected frequency ($f_F$ = 0.05 THz) and with a defined damping coefficient ($\xi$ = 0.01). What happens if these three variables are modified? Figure 5 shows the obtained RMSD dynamic amplification, computed as the ratio between the maximum dynamic RMSD and the RMSD obtained under the application of the

perturbation in a static fashion, for the whole investigated frequency range (0.001 THz – 0.5 THz) and for the three selected damping coefficients $\xi$ (0.001, 0.01 and 0.1). The different colored curves are associated with each of the 100 different random force patterns applied to the protein structure.

Clear peaks are recognizable in the low-frequency range, around 0.05 – 0.15 THz, where the low-frequency protein modes are found to occur (Figure 1b). The intensity of the peaks is highly dependent on the adopted value of the damping coefficient. Very low values of $\xi$, such as 0.001 and 0.01, lead to amplifications of the order of 20–30. Conversely, amplification coefficients lower than 5 are found for higher damping coefficients. It is also evident that the most intense peaks are associated with the first low-frequency modes, in the region 0.05 THz – 0.06 THz. Other pronounced peaks are also found for higher modes, especially in the region between 0.08 THz and 0.15 THz. In the higher region of the spectrum, the dynamic amplification gets lower, eventually leading to de-amplified responses, i.e. with an RMSD amplification factor lower than 1, especially in presence of higher damping coefficients ($\xi$ = 0.1). Moreover, it can be seen that the specific force pattern has an influence on the overall system amplification. Nevertheless, highly amplified responses are always found in the low-frequency range (Figure 5). As briefly mentioned above, this dynamic amplification might be one underlying reason which enables the protein to achieve the large-scale conformational changes when it gets triggered in the low-frequency range, despite the theory behind all these calculations is strictly valid in the small-amplitude regime.

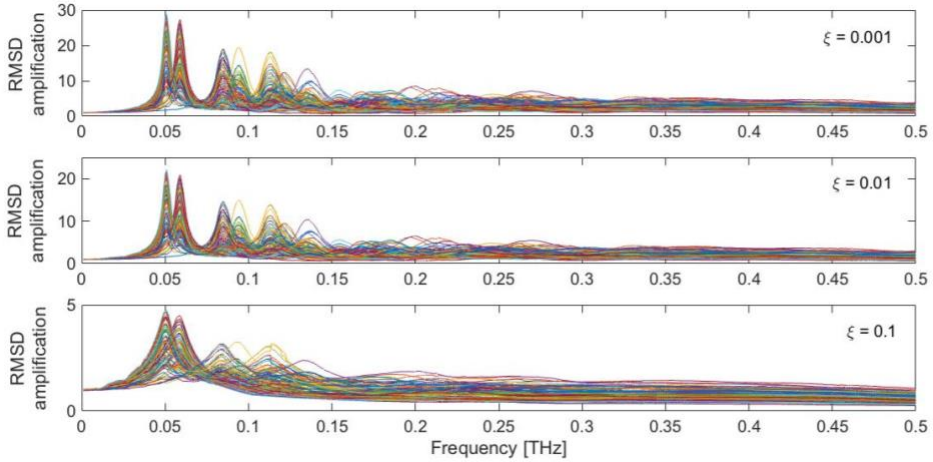


**Figure 5.** RMSD dynamic amplification for the LAO-binding protein response, as a function of damping ($\xi$ = 0.001, 0.01, 0.1) and different random force patterns. Each colored curve represents one of the 100 different random force patterns.

Figure 6 shows the maximum overlap values obtained during ten cycles of dynamic perturbation by comparing the calculated displacement field $\{u(t)\}$ with the observed conformational change $\{CC\}$, as a function of the exciting frequency, damping coefficient and specific force pattern. As can be seen, remarkably high values up to 0.95 are found in the low-frequency range, especially between 0.02 THz and 0.08 THz. It is interesting to observe how in this low-frequency range the maximum overlap score is always very high, despite the specific force pattern. The upper panel of Figure 7 shows the maximum overlap scores obtained for each of the 100 different force patterns, for each selected damping coefficient, while the lower panel reports the exciting frequency in correspondence of which the best overlap is met. As can be seen, despite the specific random force pattern, very high values of the overlap are always obtained (up to 0.95) with applied frequencies in the range 0.02 THz – 0.08 THz, which corresponds to the low-frequency end of the spectrum (Figure 1b). From Figure 5, it can also be seen that if the protein is excited at higher frequencies, say with frequencies higher than 0.1 THz, the

obtained overlap scores become lower, suggesting that the closed conformation cannot be sampled by applying harmonic excitations in this frequency range. Also, it can be noticed that the overlaps are higher when the damping coefficients are relatively low ($\xi$ = 0.001 and 0.01).

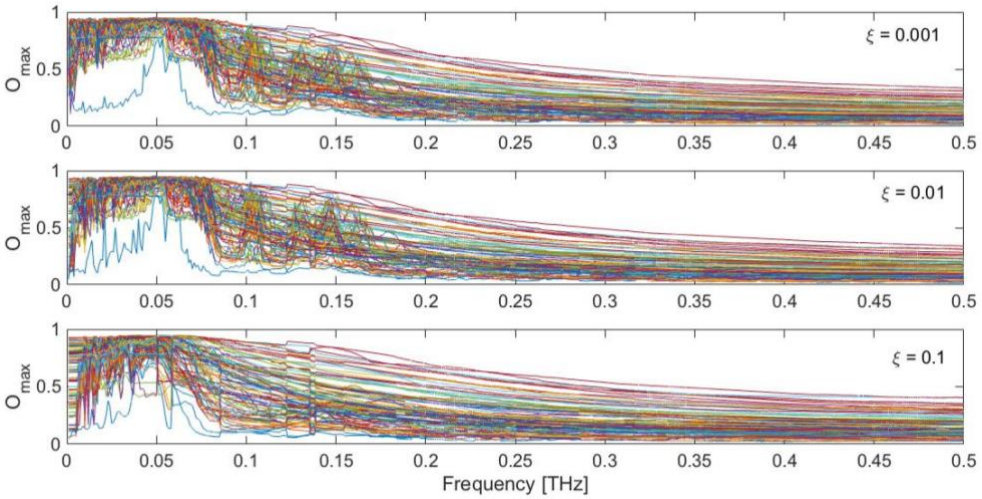


**Figure 6.** Maximum obtained overlap score for the LAO-binding protein response with respect to the observed conformational change, as a function of damping ($\xi$ = 0.001, 0.01, 0.1) and different random force patterns. Each colored curve represents one of the 100 different random force patterns.

Putting together the results obtained above, we can conclude that by exciting the open structure with external dynamic perturbations in the low-frequency range we can sample the closed conformation with remarkably high values of directionality correlations. In case of low damping coefficients ($\xi$ = 0.001 and 0.01), one also obtains high dynamic amplification factors in that frequency range, therefore potentially allowing to reach the closed conformation even with a small amount of force involved. Finally, it also seems that the specific force pattern, which is completely random in the space-domain (Figure 2a), has not a huge influence on the results, almost always leading to high overlap scores, as long as the forces are applied with an exciting frequency in the lower part of the mode spectrum (Figure 7).

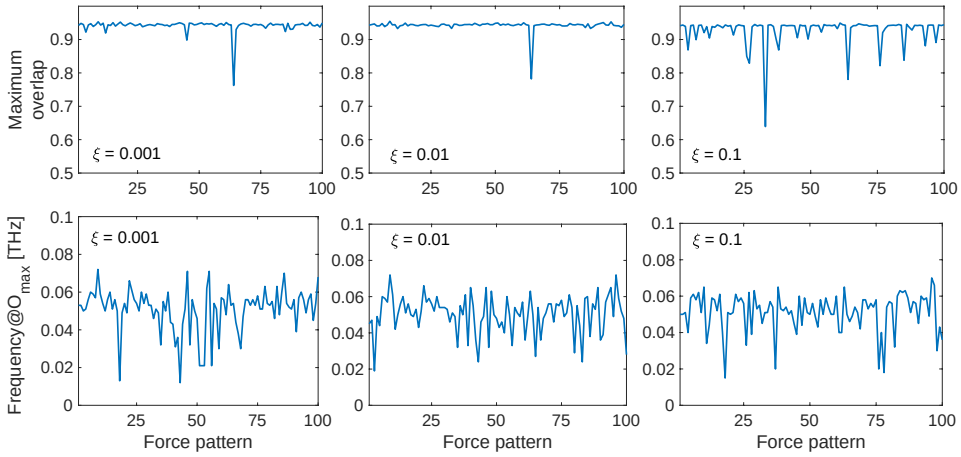


**Figure 7.** Maximum overlap score and corresponding applied frequency $f_F$ for the LAO-binding protein conformational change, as a function of damping ($\xi$ = 0.001, 0.01, 0.1) and specific random force pattern. The maximum overlap values, shown in the upper panels, are defined as the maximum values obtained over all the applied frequencies in the range 0.001 – 0.5 THz, and the corresponding optimal frequencies are reported in the lower panels depending on each of the 100 random force patterns.

The analysis reported above focused on LAO-binding protein. However, the same analysis was carried out with other proteins, i.e. maltodextrin-binding protein, lactoferrin and triglyceride lipase (see Supplementary Material). For maltodextrin-binding protein (PDB code of the open form: 1omp, PDB code of the closed form: 1anf, N = 370), one obtains a maximum overlap of 0.81 when comparing the second normal mode with the open-to-closed conformational change (Figure S1). However, overlap values as high as 0.95 can be found again when applying dynamic perturbations in the low-frequency range (Figures S3 and S4). As in the case of LAO-binding protein, the overlaps are higher for lower damping coefficients ($\xi$ = 0.001 and 0.01), which also lead to higher dynamic amplifications (Figure S2). In the case of lactoferrin (PDB code of the open form: 1lfh, PDB code of the closed form: 1lfg, N = 691), a maximum overlap of 0.46 is obtained between the third ANM mode and the conformational change (Figure S5). However, if the structure is perturbed dynamically in the low-frequency range, overlap values up to 0.88 can be obtained (Figures S7 and S8). From the comparison between lactoferrin and the two previous proteins, we understand that, when the individual modes have a higher agreement with the conformational change, the full dynamic response can sample the closed conformation better. Nevertheless, even when individual modes have lower similarities ($O_{max}$ = 0.46 for lactoferrin), the perturbation-based dynamic response allows to achieve a better agreement with the closed conformation ($O_{max}$ = 0.88). Finally, in the case of triglyceride lipase (PDB code of the open form: 3tgl, PDB code of the open form: 4tgl, N = 265), one obtains a really low value of the overlap when comparing individual ANM modes to the conformational change ($O_{max}$ = 0.27 for the fourteenth mode, Figure S9). As a result, the maximum overlap found by applying dynamic perturbation to the protein ANM is only 0.44 (Figures S11 and S12), showing that in this case the closed conformation cannot be sampled with high accuracy by the proposed method. This shows that the method proposed here always leads to higher overlaps than those obtained through the classic individual mode comparison. However, the method works better when the low-frequency modes have already a relevant similarity with the conformational change. As Tama and Sanejouand showed in their seminal work [24], this is a direct consequence of the degree of collectivity of the conformational change. Collective transitions are usually better captured by the low-frequency modes, whereas localized conformational changes usually are not. For the four proteins investigated here, the collectivity degree of their conformational transitions are 0.68 (LAO-binding protein), 0.67 (maltodextrin-binding protein), 0.48 (lactoferrin) and 0.07 (triglyceride lipase). As a result, LAO-binding protein and maltodextrin-binding protein reach very high values of the overlap from the full dynamic response (0.95), lactoferrin reaches a high value (0.88), while triglyceride lipase reaches a quite low value (0.44). This leads us to conclude that, with the proposed methodology, starting from the open conformation of the protein and without any *a priori* knowledge of the closed form, we are able to capture the closed conformations accurately as long as the conformational transition is quite collective in nature.

In all previous examples, we have investigated the conformational change from the open to the closed conformation. Figures S13 and S14 show the results of the analysis for LAO-binding protein, this time considering the closed conformation (pdb: 1lst) as reference and the open conformation (pdb: 2lao) as target. In agreement with what said above for the open-to-closed conformational transitions, the time-dependent force application generally allows to reach higher overlap values than those obtained by comparison with individual modes. As a matter of fact, when only looking at individual modes, the third ANM mode is the one showing the highest overlap, with $O_{max}$ = 0.56 (Figure S13). Conversely, applying dynamic forces can enhance this maximum overlap, reaching values of $O_{max}$ = 0.75 (Figure S14). Again, this suggests that the target conformation can generally be captured better by considering the full dynamic response of the protein, rather than looking at the trajectory generated with individual modes. However, despite the improvement, this value is lower than the one obtained when

looking at the open-to-closed conformational change (compare Figure S14 with Figure 6). This suggests that, as already noticed by Tama and Sanejouand [24] and subsequent researchers, the closed-to-open transition is generally more difficult to generate than the open-to-closed one when working with ENMs.

From the lower panels of Figures 7, S4, S8 and S12, one can also observe that there is not a unique value of the exciting frequency leading to the maximum overlap score. In fact, there exists a range of low-frequency values, where each specific force pattern is able to sample the target conformation with the highest directionality correlation. This suggests that, although we are often in the range of the first fundamental frequency, i.e. the frequency associated to the first non-rigid mode shape, the optimal excitation frequency might be slightly different than the fundamental frequency, mostly due to the not negligible involvement of higher-order modes in the definition of the complete protein dynamic response. Moreover, the exact values of these frequencies must be treated carefully, as they are strongly dependent on the model parameters, such as the cutoff and the definition of spring network [35,36]. The absolute values of these frequencies strongly depend on the value of the adopted spring constant $\gamma$, which in turn is defined upon comparison between the numerical and experimental B-factors (Section 2.1). In doing such direct comparison, we are implicitly assuming that the experimental B-factors are dominated by the protein fluctuations, i.e. they only depend on the internal protein dynamics. Unfortunately, this is not always the case, since studies have shown that B-factors might also include other contributions coming from rigid motions, crystal disorder, refinement effects, etc. [63–66]. Therefore, for all the above mentioned reasons, when using ENMs for the understating of protein motions and their corresponding frequencies of vibration, we can only have some insights on the expected frequency range, and not on the individual frequency values.

Additional considerations need to be done regarding the damping coefficients adopted in the present analysis. In this work, values of $\xi$ = 0.001, 0.01 and 0.1 have been used, meaning that the problem is treated in the underdamped regime ($\xi < 1$). Moreover, the value of $\xi$ has been kept constant for all the vibrational modes, i.e. $\xi_n = \xi$. However, one should be careful about such choices. As a matter of fact, in the framework of NMA and ENMs, the dynamic response of proteins is studied with no damping at all, i.e. $\xi = 0$. Conversely, few studies using Langevin Network Models (LNMs) have shown that the dynamics of macromolecules and proteins might be strongly overdamped, i.e. $\xi >> 1$ at least for the lowest-frequency modes [67,68]. We might also reasonably expect that the damping characteristics change for the different modes of vibration, the lowest-frequency ones being the more damped, the highest-frequency ones the less damped [68]. By using the LNM, Miller et al. [68] showed that at normal water viscosities most of the protein modes should be overdamped. Nevertheless, recent studies based on optical Kerr-effect (OKE) spectroscopy revealed the existence of underdamped global protein vibrations in the THz frequency range [69]. So, it is evident that there is still little consensus nowadays about the damping nature of these functional protein vibrations. Previous numerical results showing that undamped vibrational modes correlate well with protein conformational changes [24,28,29], as well as the outcomes of our calculations based on the underdamped assumption, seem to suggest that these functional conformational transitions can indeed be retrieved by working in the underdamped limit. Yet, extensive research efforts still need to be carried out in the future to address this open issue.

Remarks need to be given also regarding the physical meaning associated with the adopted perturbation scheme, i.e. harmonic excitations with a random direction in the space-domain but with a well-defined frequency content in the time-domain. In the previous literature, different force application patterns have been applied in a static fashion to probe protein flexibility [41,42] and protein conformational changes [44,45,47]. The approach proposed here, i.e. applying random forces to the protein ANM in a dynamic fashion, is supposed to simulate the external perturbations to the protein structure mainly due to Brownian motions of the surrounding particles [70]. These

collisions can be numerically simulated as random forces both in the space- and time-domain [68]. However, we know that every time-dependent variable can be decomposed according to their frequency component, e.g. via the Fourier Transform (FT) or other signal transformation techniques. In this way, a random excitation in the time-domain can always be represented as a sum of harmonic excitations at specific frequencies, weighted by their FT amplitude. Based on these considerations, the application of harmonic forces with random directions can be seen as the attempt to investigate the response of the protein structure under the different frequency-based components of the complex time-dependent excitations due to the particle collisions. From the results of the present analysis, we obtained that the low-frequency components of these excitations are able to drive the protein conformational change. Conversely, high-frequency components are not particularly relevant for the conformational transition (see Figure 6).

## 4. Conclusions

In this paper, the ANM has been used for the first time in order to investigate the complete dynamic response of protein structures under external dynamic perturbations. In particular, by considering the mass, viscous damping and elastic stiffness features of the protein ENM, the equations of motions have been numerically solved, in order to retrieve the protein response under the effect of harmonic forces applied to the protein structure. From the results, it has been observed that the application of dynamic forces to the open protein conformation in the low-frequency range is able to drive the protein towards the closed form with a high directionality alignment (overlap of 0.95). Such correlation is even higher than that usually obtained when comparing individual modes to the conformational change. This is mostly due to the fact that, when the external perturbations are applied harmonically at selected frequencies, the full dynamics of the system depends on the combination of the most involved modes, and not only on a single mode.

By analyzing the time-dependent overlap values, it has been possible to recognize how the closed form of the protein is not reached through a straight line pathway, since the overlap score usually oscillates between low and high values. This suggests that the protein samples the target conformation among a variety of other conformations, and it does so by following curvilinear pathways, as already suggested in the existing literature [58–60]. This has also been confirmed by geometrical calculations shown in this paper, where we have looked at the evolution of displacement vectors throughout the motion. The trajectory of the protein upon harmonic perturbation has also been shown and PCA has been used to reduce the problem dimensionality and investigate the ensemble of generated conformations in the sub-space of the PCs. Furthermore, taking into account the dynamic response of the protein network under these external forces, high dynamic amplification values are often found, especially when the perturbation is applied in the low-frequency range and for low values of the damping ratio (resonance effect). These high dynamic amplifications might explain why a combination of the normal modes of vibration, which theoretically are only valid in the small-amplitude regime, allows to reach the target form of the protein even in the case of large-scale conformational transitions. Moreover, the dynamic nature (frequency) of the excitation seems to be the central parameter driving the conformational transition, the specific force pattern having a smaller influence on the capability of the protein to sample its target conformation.

The well-known case of LAO-binding protein has been addressed in the main text of this paper. The results related to maltodextrin-binding protein, lactoferrin and triglyceride lipase have been briefly discussed and are available in the Supplementary Material. From all the cases, we have observed that external harmonic excitations in the low-frequency range always lead to higher overlaps with the observed conformational change rather than those obtained by comparing individual modes. These overlap values are generally higher, the more collective the conformational transition is. Moreover, as expected, we observed that the open-to-closed conformational change is easier to be

captured than the closed-to-open one. Future research efforts will be dedicated to investigate the effect of the ENM parameters, such as the cutoff, spring constant distribution, etc., the damping ratio, as well as the nature of protein conformational change on the obtained outcomes. Notwithstanding, the preliminary results shown here seem to suggest that the low-frequency components of the random external perturbations to the protein structure due to the surrounding environment might play a key role in driving the biologically-relevant conformational transition.

**Supplementary Materials:** The following are available online at www.mdpi.com/xxx/s1, Figures S1-S14, Supplementary Movies 1-2.

**Author Contributions:** Conceptualization, D.S. and G.P.; methodology, D.S.; software, D.S.; validation, D.S.; formal analysis, D.S.; investigation, D.S. and G.P.; resources, A.C.; data curation, D.S.; writing—original draft preparation, D.S.; writing—review and editing, G.P., G.L., A.C.; supervision, A.C. and G.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bahar, I.; Jernigan, R.L.; Dill, K.A. *Protein Actions: Principles & Modeling*; Garland Science: New York, 2017;
2. Alberts, B.; Johnson, A.; Lewis, J.; Morgan, D.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell, 6th Edition*; Garland Science: New York, 2002;
3. McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nature* **1977**, *267*, 585–590.
4. Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6679–6685.
5. Hospital, A.; Goñi, J.R.; Orozco, M.; Gelpí, J.L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinforma. Chem.* **2015**, *8*, 37–47.
6. Tirion, M.M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
7. Fuglebakk, E.; Tiwari, S.P.; Reuter, N. Comparing the intrinsic dynamics of multiple protein structures using elastic network models. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850*, 911–922.
8. Togashi, Y.; Flechsig, H. Coarse-grained protein dynamics studies using elastic network models. *Int. J. Mol. Sci.* **2018**, *19*.
9. Lezon, T.R.; Shrivastava, I.H.; Yang, Z.; Bahar, I. Elastic Network Models For Biomolecular Dynamics: Theory and Application to Membrane Proteins and Viruses. In; 2009; pp. 129–158.
10. Haliloglu, T.; Bahar, I.; Erman, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090–3093.
11. Bahar, I.; Atilgan, A.R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **1997**, *2*, 173–181.
12. Micheletti, C.; Carloni, P.; Maritan, A. Accurate and Efficient Description of Protein Vibrational Dynamics: Comparing Molecular Dynamics and Gaussian Models. *Proteins Struct. Funct. Genet.* **2004**, *55*, 635–645.
13. Bahar, I.; Erman, B.; Jernigan, R.L.; Atilgan, A.R.; Covell, D.G. Collective motions in HIV-1 reverse transcriptase: Examination of flexibility and enzyme function. *J. Mol. Biol.* **1999**, *285*, 1023–1037.
14. Bahar, I.; Jernigan, R.L. Cooperative fluctuations and subunit communication in tryptophan synthase. *Biochemistry* **1999**, *38*, 3478–3490.
15. Bahar, I.; Jernigan, R.L. Vibrational dynamics of transfer RNAs: Comparison of the free and synthetase-bound forms. *J. Mol. Biol.* **1998**, *281*, 871–884.

16. Bahar, I.; Atilgan, A.R.; Demirel, M.C.; Erman, B. Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* **1998**, *80*, 2733–2736.
17. Atilgan, A.R.; Durell, S.R.; 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*, 505–515.
18. Eyal, E.; Yang, L.W.; Bahar, I. Anisotropic network model: Systematic evaluation and a new web interface. *Bioinformatics* **2006**, *22*, 2619–2627.
19. Yang, L.; Song, G.; Jernigan, R.L. Comparisons of Experimental and Computed Protein Anisotropic Temperature Factors. *Proteins* **2008**, *76*, 164–175.
20. Yang, L.; Song, G.; Jernigan, R.L. Protein elastic network models and the ranges of cooperativity. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12347–12352.
21. Kurkcuoglu, O.; Jernigan, R.L.; Doruker, P. Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions. *Polymer (Guildf).* **2004**.
22. Kurkcuoglu, O.; Jernigan, R.L.; Doruker, P. Collective dynamics of large proteins from mixed coarse-grained elastic network model. In Proceedings of the QSAR and Combinatorial Science; 2005.
23. Koehl, P.; Orland, H.; Delarue, M. Parameterizing elastic network models to capture the dynamics of proteins. *J. Comput. Chem.* **2021**, 1–19.
24. Tama, F.; Sanejouand, Y.H. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* **2001**, *14*, 1–6.
25. Al-Bluwi, I.; Vaisset, M.; Siméon, T.; Cortés, J. Modeling protein conformational transitions by a combination of coarse-grained normal mode analysis and robotics-inspired methods. *BMC Struct. Biol.* **2013**, *13*, 1–14.
26. Sanejouand, Y.-H. Normal-mode driven exploration of protein domain motions. *arXiv* **2021**, *2103*, 11959.
27. Nicolay, S.; Sanejouand, Y.H. Functional modes of proteins are among the most robust. *Phys. Rev. Lett.* **2006**, *96*, 1–4.
28. Yang, L.; Song, G.; Jernigan, R.L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920–929.
29. Mahajan, S.; Sanejouand, Y.H. On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Arch. Biochem. Biophys.* **2015**, *567*, 59–65.
30. Mahajan, S.; Sanejouand, Y.H. Jumping between protein conformers using normal modes. *J. Comput. Chem.* **2017**, *38*, 1622–1630.
31. Delarue, M.; Sanejouand, Y.H. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: The Elastic Network Model. *J. Mol. Biol.* **2002**, *320*, 1011–1024.
32. Zheng, W.; Doniach, S. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13253–13258.
33. Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R.L. Close Correspondence between the Essential Protein Motions from Principal Component Analysis of Multiple HIV-1 Protease Structures and Elastic Network Modes. *Structure* **2008**, *16*, 321–330.
34. Dobbins, S.E.; Lesk, V.I.; Sternberg, M.J.E. Insights into protein flexibility: The relationship between normal modes and conformational change upon protein-protein docking. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10390–10395.
35. Scaramozzino, D.; Lacidogna, G.; Piana, G.; Carpinteri, A. A finite-element-based coarse-grained model for global protein vibration. *Meccanica* **2019**, *54*, 1927–1940.
36. Giordani, G.; Scaramozzino, D.; Iturrioz, I.; Lacidogna, G.; Carpinteri, A. Modal analysis of the lysozyme protein considering all-atom and coarse-grained finite element models. *Appl. Sci.* **2021**, *11*, 547.
37. Levitt, M.; Sander, C.; Stern, P.S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **1985**, *181*, 423–447.

38. Ben-Avraham, D. Vibrational normal-mode spectrum of globular proteins. *Phys. Rev. B* **1993**, *47*, 14559–14560.
39. Brooks, B.; Karplus, M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 4995–4999.
40. Markelz, A.; Whitmire, S.; Hillebrecht, J.; Birge, R. THz time domain spectroscopy of biomolecular conformational modes. *Phys. Med. Biol.* **2002**, *47*, 3797–3805.
41. Eyal, E.; Bahar, I. Toward a molecular understanding of the anisotropic response of proteins to external forces: Insights from elastic network models. *Biophys. J.* **2008**, *94*, 3424–3435.
42. Scaramozzino, D.; Khade, P.M.; Jernigan, R.L.; Lacidogna, G.; Carpinteri, A. Structural Compliance: A New Metric for Protein Flexibility. *Proteins Struct. Funct. Bioinforma.* **2020**, *88*, 1482–1492.
43. Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. Protein structural change upon ligand binding: Linear response theory. *Phys. Rev. Lett.* **2005**, *94*, 1–4.
44. Atilgan, C.; Atilgan, A.R. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. *PLoS Comput. Biol.* **2009**, *5*.
45. Atilgan, C.; Gerek, Z.N.; Ozkan, S.B.; Atilgan, A.R. Manipulation of conformational change in proteins by single-residue perturbations. *Biophys. J.* **2010**, *99*, 933–943.
46. Gerek, Z.N.; Ozkan, S.B. Change in allosteric network affects binding affinities of PDZ domains: Analysis through perturbation response scanning. *PLoS Comput. Biol.* **2011**, *7*, 18–25.
47. Liu, J.; Sankar, K.; Wang, Y.; Jia, K.; Jernigan, R.L. Directional Force Originating from ATP Hydrolysis Drives the GroEL Conformational Change. *Biophys. J.* **2017**, *112*, 1561–1570.
48. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
49. Carpinteri, A.; Lacidogna, G.; Piana, G.; Bassani, A. Terahertz mechanical vibrations in lysozyme: Raman spectroscopy vs modal analysis. *J. Mol. Struct.* **2017**, *1139*, 222–230.
50. Carpinteri, A.; Piana, G.; Bassani, A.; Lacidogna, G. Terahertz vibration modes in Na/K-ATPase. *J. Biomol. Struct. Dyn.* **2019**, *37*, 256–264.
51. Lacidogna, G.; Piana, G.; Bassani, A.; Carpinteri, A. Raman spectroscopy of Na/K-ATPase with special focus on low-frequency vibrations. *Vib. Spectrosc.* **2017**, *92*, 298–301.
52. Clough, R.W.; Penzien, J. *Dynamics of Structures*; Computers & Structures, Inc., 2003;
53. Dykeman, E.C.; Sankey, O.F. Normal mode analysis and applications in biological physics. *J. Phys. Condens. Matter* **2010**, *22*, 423202.
54. Teodoro, M.L.; J, G.N.P.; Kavraki, L.E. A Dimensionality Reduction Approach to Modeling Protein Flexibility. In Proceedings of the Int. Conf. Comput. Mole. Biol; 2002; pp. 299–308.
55. Sankar, K.; Mishra, S.K.; Jernigan, R.L. Comparisons of Protein Dynamics from Experimental Structure Ensembles, Molecular Dynamics Ensembles, and Coarse-Grained Elastic Network Models. *J. Phys. Chem. B* **2018**.
56. Kim, M.K.; Chirikjian, G.S.; Jernigan, R.L. Elastic models of conformational transitions in macromolecules. *J. Mol. Graph. Model.* **2002**, *21*, 151–160.
57. Kim, M.K.; Jernigan, R.L.; Chirikjian, G.S. Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J.* **2002**, *83*, 1620–1630.
58. Song, G.; Jernigan, R.L. An enhanced elastic network model to represent the motions of domain-swapped proteins. *Proteins Struct. Funct. Genet.* **2006**.
59. Maragakis, P.; Karplus, M. Large amplitude conformational change in proteins explored with a plastic network model: Adenylate kinase. *J. Mol. Biol.* **2005**, *352*, 807–822.

60. Hoffmann, A.; Grudinin, S. NOLB: Nonlinear Rigid Block Normal-Mode Analysis Method. *J. Chem. Theory Comput.* **2017**, *13*, 2123–2134.
61. Das, A.; Gur, M.; Cheng, M.H.; Jo, S.; Bahar, I.; Roux, B. Exploring the Conformational Transitions of Biomolecular Systems Using a Simple Two-State Anisotropic Network Model. *PLoS Comput. Biol.* **2014**, *10*.
62. Chandrasekaran, S.N.; Das, J.; Dokholyan, N. V.; Carter, C.W. A modified PATH algorithm rapidly generates transition states comparable to those found by other well established algorithms. *Struct. Dyn.* **2016**, *3*.
63. Carugo, O. Maximal B-factors in protein crystal structures. *BMC Bioinformatics* **2018**, *19*, 61.
64. Kuzmanic, A.; Pannu, N.S.; Zagrovic, B. X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nat. Commun.* **2014**, *5*, 3220.
65. Bastolla, U.; Dehouck, Y. The maximum penalty criterion for ridge regression: Application to the calibration of the force constant in elastic network models. *Integr. Biol. (United Kingdom)* **2017**, *9*, 627–641.
66. Na, H.; Hinsen, K.; Song, G. The Amounts of Thermal Vibrations and Static Disorder in Protein X-ray Crystallographic B-factors. *Proteins Struct. Funct. Bioinforma.* **2021**.
67. Lamm, G.; Szabo, A. Langevin modes of macromolecules. *J. Chem. Phys.* **1986**, *85*, 7334–7348.
68. Miller, B.T.; Zheng, W.; Venable, R.M.; Pastor, R.W.; Brooks, B.R. Langevin network model of myosin. *J. Phys. Chem. B* **2008**, *112*, 6274–6281.
69. Turton, D.A.; Senn, H.M.; Harwood, T.; Lapthorn, A.J.; Ellis, E.M.; Wynne, K. Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nat. Commun.* **2014**, *5*, 2–7.
70. Ermak, D.L.; McCammon, J.A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* **1978**, *69*, 1352–1360.