

A finite-element-based coarse-grained model for global protein vibration

Original

A finite-element-based coarse-grained model for global protein vibration / Scaramozzino, D., Lacidogna, G., Piana, G., Carpinteri, A.. - In: MECCANICA. - ISSN 0025-6455. - STAMPA. - 54:13(2019), pp. 1927-1940. [10.1007/s11012-019-01037-9]

Availability:

This version is available at: 11583/2770755 since: 2019-12-02T17:18:51Z

Publisher:

Springer Netherlands

Published

DOI:10.1007/s11012-019-01037-9

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

Springer postprint/Author's Accepted Manuscript

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <http://dx.doi.org/10.1007/s11012-019-01037-9>

(Article begins on next page)

A Finite-Element-Based Coarse-Grained Model for Global Protein Vibration

Domenico Scaramozzino, Giuseppe Lacidogna*, Gianfranco Piana, Alberto Carpinteri

*Politecnico di Torino, Department of Structural, Geotechnical and Building Engineering,
Corso Duca degli Abruzzi 24 – 10129 Torino, Italy*

*Corresponding author: giuseppe.lacidogna@polito.it

Abstract

Protein mechanical vibrations play a pivotal role in biological activity. In particular, low-frequency (terahertz) modes are related to protein conformational changes, which represent the foundations for a correct protein functionality. Relying on the fact that such low-frequency motions involve large protein portions, thus modeling of local details is not necessary, coarse-grained models have proven their efficacy in capturing the essential dynamic behavior. In this paper, we show that a coarse-grained Finite Element (FE) space truss model is suitable for investigating protein vibrations. Hen egg-white lysozyme is selected as a case study and modal analysis is performed in order to investigate the protein dynamics; the influence of interaction cutoff values on optimal force constant, obtained vibrational frequencies and mode shapes is also explored. The validity of the structural model is demonstrated by comparing the calculated B -factors with the experimental ones. Moreover, from the methodology framework the truss model is shown to be consistent with the well-known Anisotropic Network Model (ANM) and this has been confirmed by the obtained results. The proposed truss model is then believed to be a simple yet powerful tool to investigate protein dynamics, and it could also be used to analyze conformational changes and protein stability from a Structural Mechanics viewpoint.

Keywords: Space Truss Structure, Protein Vibration, Modal Analysis, Coarse-Grained Model, Terahertz Range, Lysozyme, Modal Assurance Criterion, B -factor.

1. Introduction

Proteins constitute one of the main building blocks of biological systems and are responsible for many fundamental tasks. They are in charge of carrying molecules and nutrients along the body (hemoglobin) and inside the cell (kinesin, myosin), sustaining contraction processes in muscles (actin), conferring elasticity to tissues (elastin), catalyzing chemical reactions (lysozyme), activating signaling pathways (mitogen-activated protein kinase), etc. It is well known that biological functionality is strictly related to protein three-dimensional structure. In this sense, evolution may have played a fundamental role in determining the optimal folded shape [1].

Recently, it has also been suggested that the investigation of the mechanical behavior of such nanostructures may lead us to a broader understanding of their biochemical functioning [2], as well as to a proper description of the mechanical response of macroscopic systems [3]. For example, Structural Mechanics approaches have proven their efficacy in order to investigate complex biomechanical systems, such as collagenous tissues and fibrils [4-6]. Additionally, the application of mechanics and statistical mechanics methodologies has led to interesting insights regarding the behavior of biological systems at the cell scale [7-10].

Since biological activity is performed in a dynamic manner, a deeper investigation of protein vibrations around their native state is pivotal in order to understand the complex mechanisms

hidden behind their functionality [11]. For this purpose, the dynamics of protein structures is usually explored by means of both numerical and experimental techniques. As for the former, molecular dynamics (MD) simulations are generally employed. However, due to demanding computational cost of MD modeling, normal mode analysis (NMA) has emerged as a powerful numerical tool to investigate protein dynamics [12-19], by using both all-atom and coarse-grained representations. Each modeling procedure has its own advantages and drawbacks; however, even simplified models, like coarse-grained ones, have proven to be able to capture the essential protein dynamic behavior [20-23]. The reason may lie in the aforesaid strict relationship between the overall three-dimensional shape and biological activity.

Protein vibrations can arise at various scales, e.g. involving single amino acids, primary (covalent) bonds, secondary structures, as well as the whole protein three-dimensional structure. These motions are completely defined by the corresponding mode shape, which is identified by the displacement field, and their frequency of vibration. Generally, the smaller the scale of motion, the higher the frequency. Within the entire mode spectrum, the low-frequency modes, mostly occurring in the terahertz (THz) range, have received increasing attention among researchers, since they have been indicated to be strictly related to protein conformational changes [24-29]. These motions affect large portions of the protein and play a key role in controlling binding activity. It is then possible that some resonance phenomena in the THz range occur, which may control the protein shape change. Indeed, on a larger scale, it has already been suggested that thermodynamic resonance phenomena can affect cell biological activity [30-32]. Hence, a deeper understanding of the dynamic behavior may lead to a broader comprehension of the mechanisms of life at the protein level.

From an experimental perspective, protein vibrations can be detected by means of spectroscopy techniques, such as IR (infrared) [33], THz-TDS (terahertz time-domain spectroscopy) [34,35] and Raman spectroscopy [36,37]. In particular, the latter has emerged as a powerful tool in detecting vibrations in the THz range, by using modern ultra-low frequency (ULF) filters. In previous studies some of the authors performed low-frequency Raman measurements on lysozyme [38] and Na/K-ATPase [39] samples, obtaining some notable peaks around 0.8 THz. Modal analysis was then performed to investigate the vibrations arising from FE-based all-atom lattice models, and the modal displacements which were associated to the obtained Raman peaks were found to involve the entire lysozyme structure and large portions of Na/K-ATPase [38,40]. Such all-atom FE models, made of point-masses connected by elastic beams simulating local covalent bonds through their axial stiffness, were originally built for describing breathing-like (i.e. expansion-contraction) modes. In this paper, we make use of a coarse-grained structural model for investigating protein vibrations still by an engineering approach, but this time focusing on all possible global motions. Specifically, referring to the coordinates of C_α atoms only, a Finite Element (FE) space truss structural model was constructed in LUSAS FE software [41]. In this case, distortional modes become possible, and non-local elastic connections must be included in the model to simulate long-range interactions, which are important for global protein dynamics. After that mass and stiffness parameters were assigned to the model, modal analysis was performed in order to extract the dynamical features (frequencies and mode shapes). The case of hen egg-white lysozyme was selected as a benchmark: the results were validated by calculating the thermal B -factors and were also compared with those arising from the well-known Anisotropic Network Model (ANM) [17,42,43]. Finally, we show that the proposed structural model is consistent with ANM and can be used to investigate protein vibrational properties. Moreover, it is also suitable in perspective for studying protein conformational changes and stability from a Structural Mechanics viewpoint.

2. Methodology

In this section, the details for the construction of the space truss structural model are given. Moreover, the fundamentals of modal analysis and Anisotropic Network Model (ANM) theory are recalled.

2.1 FE Space Truss Model and Modal Analysis

We considered a space truss structural model for the investigation of the vibrational motions of lysozyme, by means of the FE commercial code LUSAS [41]. The model was made up of point masses, which represent C_α atoms, and an assembly of linear-elastic bars, which simulate the interactions between residues. The coordinates of C_α atoms were obtained from the PDB file (pdb code: 4ym8) [44], and the connections were created using different cutoff values (i.e. only the nodes whose distance was lower than the imposed cutoff were connected). Specifically, in this analysis, the cutoff parameters were set equal to 8, 10, 12, 15 and 20 Å, thus leading to five different models (Fig. 1a-e). Finally, from a structural viewpoint, the links were modelled as 3D bar elements and they were connected together by means of spherical hinges located at the nodes. In Fig. 1f, a zoomed picture of the truss model is shown, which highlights the single bar structural element, together with the associated internal boundary conditions and the nomenclature used in the remaining of Section 2.1.

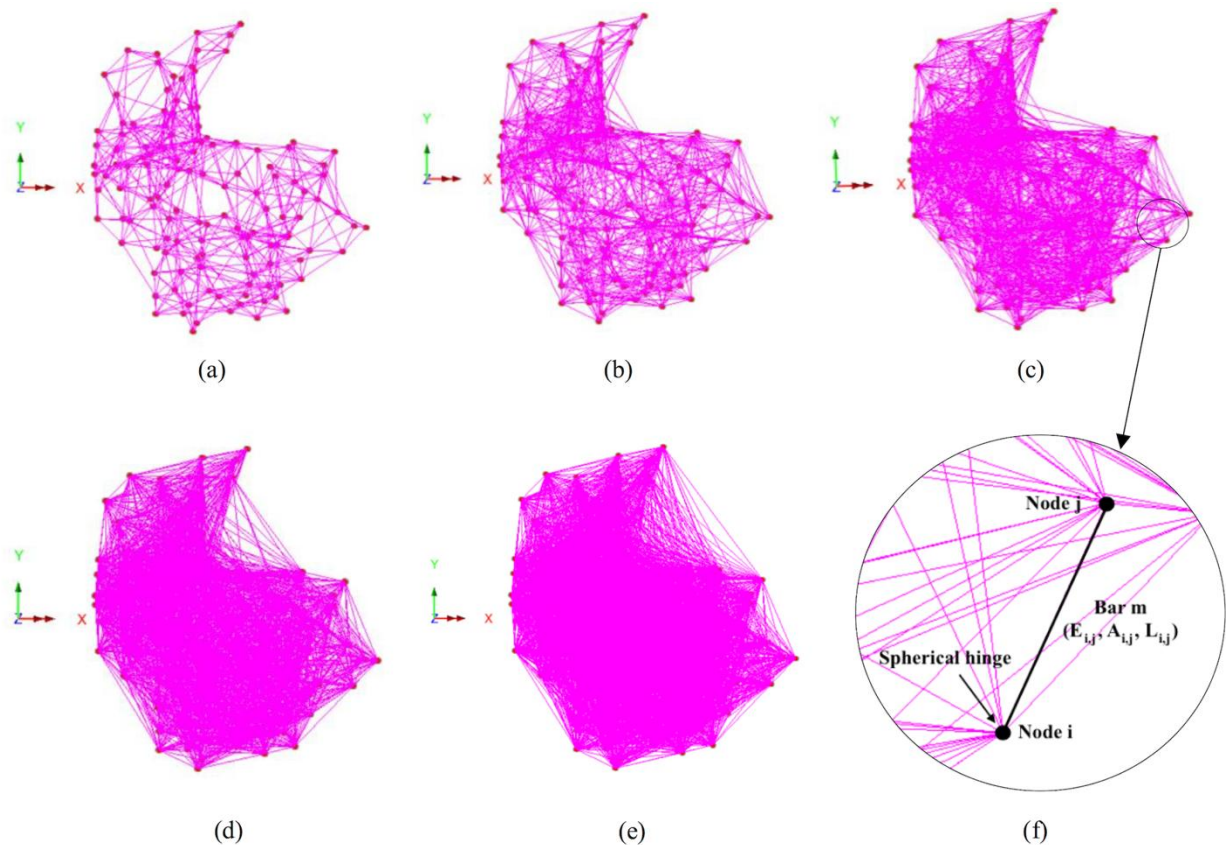


Fig. 1 Lysozyme models with varying cutoff values: **(a)** 8 Å; **(b)** 10 Å; **(c)** 12 Å; **(d)** 15 Å; **(e)** 20 Å; **(f)** zoomed picture of the truss model highlighting the single bar element

According to the FE theory, for a structure made up of N nodes and M connections, one can compute the $3N \times 3N$ global stiffness matrix \mathbf{K} starting from the 2×2 local bar element matrix \mathbf{k}_m^* , which is defined as follows [45]:

$$\mathbf{k}_m^* = \frac{E_{i,j}A_{i,j}}{L_{i,j}} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad (1)$$

$E_{i,j}$, $A_{i,j}$ and $L_{i,j}$ standing for the Young's Modulus, cross-sectional area and length of the m^{th} bar connecting nodes i and j (Fig. 1f). Note that the bars must not be intended as real physical elements, but just as fictitious connections simulating the interactions among residues. Since in the present analysis the Young's Modulus and the cross-sectional area were set equal for all the connections (i.e. $E_{i,j}A_{i,j} = EA$), the stiffness of each bar turned out to be inversely proportional to its length ($k_m \propto L_{i,j}^{-1}$). However, note that by assigning different rigidities to different bars with specific criteria based on their length, other stiffness variation laws could be obtained.

Taking into account the different orientation between the local reference system of each bar and the global reference system, the 2×2 local element matrix \mathbf{k}_m^* can be referred to the global reference system. This can be achieved by means of the 2×6 rotation matrix \mathbf{N}_m , which contains the direction cosines between the local and global reference systems and can be written as follows:

$$\mathbf{N}_m = \begin{bmatrix} \frac{(x_j - x_i)}{L_{ij}} & \frac{(y_j - y_i)}{L_{ij}} & \frac{(z_j - z_i)}{L_{ij}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{(x_j - x_i)}{L_{ij}} & \frac{(y_j - y_i)}{L_{ij}} & \frac{(z_j - z_i)}{L_{ij}} \end{bmatrix}, \quad (2)$$

being x_i , y_i and z_i the three-dimensional coordinates of node i in the global reference system, and x_j , y_j and z_j the coordinates of node j (Fig. 1f).

By pre- and post-multiplying the element stiffness matrix \mathbf{k}_m^* (Eq. 1) by the rotation matrix \mathbf{N}_m (Eq. 2), one obtains the 6×6 element stiffness matrix \mathbf{k}_m referred to the global reference system, which takes the following form:

$$\mathbf{k}_m = \mathbf{N}_m^T \mathbf{k}_m^* \mathbf{N}_m = \begin{bmatrix} \alpha_{i,j} & -\alpha_{i,j} \\ -\alpha_{i,j} & \alpha_{i,j} \end{bmatrix}, \quad (3)$$

being $\alpha_{i,j}$ a 3×3 matrix containing the stiffness values and the direction cosines:

$$\alpha_{i,j} = \frac{E_{i,j}A_{i,j}}{L_{i,j}} \begin{bmatrix} \frac{(x_j - x_i)^2}{L_{ij}^2} & \frac{(x_j - x_i)(y_j - y_i)}{L_{ij}^2} & \frac{(x_j - x_i)(z_j - z_i)}{L_{ij}^2} \\ \frac{(x_j - x_i)(y_j - y_i)}{L_{ij}^2} & \frac{(y_j - y_i)^2}{L_{ij}^2} & \frac{(y_j - y_i)(z_j - z_i)}{L_{ij}^2} \\ \frac{(x_j - x_i)(z_j - z_i)}{L_{ij}^2} & \frac{(y_j - y_i)(z_j - z_i)}{L_{ij}^2} & \frac{(z_j - z_i)^2}{L_{ij}^2} \end{bmatrix}. \quad (4)$$

Finally, the global stiffness matrix \mathbf{K} can then be computed by the FE code, as a sum over all the bar elements, as follows:

$$\mathbf{K} = \sum_{m=1}^M \mathbf{C}_m^T \mathbf{k}_m \mathbf{C}_m = \sum_{m=1}^M \mathbf{C}_m^T \mathbf{N}_m^T \mathbf{k}_m^* \mathbf{N}_m \mathbf{C}_m, \quad (5)$$

where \mathbf{C}_m is the $6 \times 3N$ expansion matrix [45]. The specific values of E and A , which define completely the stiffness of the structure, were set *a posteriori* in order to have the best match with

the experimental B -factors available from the PDB file (see below). Note that this affects the values of the vibrational frequencies, but not the mode shapes. Finally, the mass of the protein was equally divided among the nodes and, once the mass values were assigned to the C_α atoms, the FE code could also compute the $3N \times 3N$ global mass matrix \mathbf{M} . The assumption of a homogeneous mass distribution along the protein chain is a simplification, but it is justified by the fact that, in general, the mass un-homogeneities in proteins are not large, therefore a homogeneous mass distribution can give relatively accurate results both for frequencies and eigenmodes. Some exceptions for specific proteins exist, in which case the actual mass distribution should be considered.

Built the coarse-grained FE model and assigned the mechanical properties, a free vibration eigenvalue analysis (i.e. modal analysis) was performed. For the sake of clarity, it basically consists in solving the equations of motion for an MDOF (multi degree-of-freedom) undamped oscillator [45], which leads to the following generalized eigenvalue problem:

$$(\mathbf{K} - \omega_n^2 \mathbf{M}) \cdot \boldsymbol{\delta}_n = \mathbf{0}, \quad (6)$$

where ω_n^2 represents the n^{th} eigenvalue (square of the n^{th} natural angular frequency) and $\boldsymbol{\delta}_n$ is the corresponding eigenvector, i.e. the displacement field of the n^{th} mode shape. Ignoring the trivial solution $\boldsymbol{\delta}_n = \mathbf{0}$, which lacks of physical meaning, Eq. (6) can be solved by the following equation:

$$\det(\mathbf{K} - \omega_n^2 \mathbf{M}) = 0, \quad (7)$$

which allows to compute the n^{th} eigenvalue, which in turn is related to the n^{th} natural frequency of vibration f_n by the following formula:

$$f_n = \frac{\omega_n}{2\pi}. \quad (8)$$

Once obtained ω_n , the n^{th} eigenmode can be easily computed via Eq. (6). Note that, since the protein structure is not externally constrained, the first six eigenvalues are equal to zero and the corresponding eigenvectors correspond to rigid motions. For this reason, only the $3N-6$ non-rigid vibrational modes were taken into account.

From an operative viewpoint, working with SI units we have interatomic distance $\sim \text{\AA} = 10^{-10}$ m and atomic mass $\sim 10^{-26}$ kg; therefore, the scaling for distance and mass values already used by some of the authors in [38,40] was adopted to avoid numerical errors (this must not be intended as a physical scaling, but just as a numerical artifice to avoid computational problems).

Once the complete set of vibrational frequencies and eigenmodes were obtained from FE calculations, the estimated B -factors, which provide information about the flexibility/rigidity of the protein structure, could be computed according to the following equation [12]:

$$B_i = \frac{8}{3} \pi^2 k_B T \sum_{n=7}^{3N} \frac{\delta_{i,n}^2}{\omega_n^2}, \quad (9)$$

where B_i represents the calculated B -factor for the i^{th} residue, k_B is the Boltzmann constant ($\sim 1.38 \times 10^{-23} \text{ JK}^{-1}$), T is the absolute temperature, ω_n the angular frequency of the n^{th} vibrational mode and $\delta_{i,n}$ represents the mass-weighted displacement of the i^{th} residue for the n^{th} mode. Note that, since $\delta_{i,n}$ is computed by normalization through the mass matrix, the numerical scaling for mass parameters [38,40] had also to be taken into account for the calculation of B -factors via Eq. (9).

By comparing the calculated B -factors to the experimental ones available from the PDB file, the bar rigidity (EA) could be set and the absolute values of the vibrational frequencies could be obtained. Finally, normalized B -factors were also calculated, which are defined as follows:

$$B_{i,norm} = \frac{B_i - \mu(B)}{\sigma(B)}, \quad (10)$$

where $\mu(B)$ and $\sigma(B)$ represent the mean value and the standard deviation of the B -factor distribution along the protein chain, respectively.

2.2 Anisotropic Network Model (ANM)

The main purpose of the Anisotropic Network Model is the investigation of the dynamical features of proteins based on the assumption that these complex systems can be modeled as elastic networks at the residue level, i.e. considering only C_α atoms [17]. Vibrational motions can be obtained with relatively low computational cost by means of the extraction of eigenvalues and eigenvectors from the Hessian matrix, which contains the information regarding the connectivity between C_α atoms. In particular, for a system counting N residues, the $3N \times 3N$ Hessian matrix is defined as follows:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,i} & \dots & \mathbf{H}_{1,N} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{i,1} & \dots & \mathbf{H}_{i,i} & \dots & \mathbf{H}_{i,N} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{H}_{N,1} & \dots & \mathbf{H}_{N,i} & \dots & \mathbf{H}_{N,N} \end{bmatrix}, \quad (11)$$

where each 3×3 $\mathbf{H}_{i,j}$ submatrix contains the interaction information between residues i and j included within a certain cutoff distance. Submatrices $\mathbf{H}_{i,j}$ can be obtained by computing the second partial derivatives of the potential V with respect to the three directions of the global reference system:

$$\mathbf{H}_{i,j} = \begin{bmatrix} \frac{\partial^2 V}{\partial x_i \partial x_j} & \frac{\partial^2 V}{\partial x_i \partial y_j} & \frac{\partial^2 V}{\partial x_i \partial z_j} \\ \frac{\partial^2 V}{\partial y_i \partial x_j} & \frac{\partial^2 V}{\partial y_i \partial y_j} & \frac{\partial^2 V}{\partial y_i \partial z_j} \\ \frac{\partial^2 V}{\partial z_i \partial x_j} & \frac{\partial^2 V}{\partial z_i \partial y_j} & \frac{\partial^2 V}{\partial z_i \partial z_j} \end{bmatrix}. \quad (12)$$

Considering the residues i and j , whose initial distance in the equilibrium condition is $r_{i,j}^0$, the potential $V_{i,j}$ takes the form of the simple harmonic potential associated to an elastic spring connecting the two nodes:

$$V_{i,j} = \frac{\gamma}{2} (r_{i,j} - r_{i,j}^0)^2, \quad (13)$$

where γ is the elastic force constant (stiffness) of the spring, and $r_{i,j}$ represents the distance between the residues after deformation. Note that, γ being the stiffness of the fictitious connection between two residues, it has the same meaning of the bar stiffness considered in the FE space truss model

shown above (k_m , Eq. 1). By calculating the partial derivatives at the equilibrium position (i.e. $r_{i,j} = r_{i,j}^0$), the ANM yields the following expression for the off-diagonal $\mathbf{H}_{i,j}$ elements:

$$\mathbf{H}_{i,j} = -\gamma \begin{bmatrix} \frac{(x_j - x_i)^2}{r_{i,j}^2} & \frac{(x_j - x_i)(y_j - y_i)}{r_{i,j}^2} & \frac{(x_j - x_i)(z_j - z_i)}{r_{i,j}^2} \\ \frac{(y_j - y_i)(x_j - x_i)}{r_{i,j}^2} & \frac{(y_j - y_i)^2}{r_{i,j}^2} & \frac{(y_j - y_i)(z_j - z_i)}{r_{i,j}^2} \\ \frac{(z_j - z_i)(x_j - x_i)}{r_{i,j}^2} & \frac{(z_j - z_i)(y_j - y_i)}{r_{i,j}^2} & \frac{(z_j - z_i)^2}{r_{i,j}^2} \end{bmatrix}. \quad (14)$$

The diagonal terms $\mathbf{H}_{i,i}$ can be then calculated as a summation involving all the residues connected to the i^{th} node as follows:

$$\mathbf{H}_{i,i} = - \sum_{j=1, j \neq i}^N \mathbf{H}_{i,j}. \quad (15)$$

Eqs. (13-14) do not account for the interaction specificity among different residues since they consider the same force constant γ for all interacting pairs. An improvement of this has been achieved by considering also a distance-dependent force constant [42]:

$$\gamma_{i,j} \propto \frac{1}{r_{i,j}^p}, \quad (16)$$

where p is an empirical parameter which allows to consider lower force constants for longer interaction distances. In this case, Eq. (14) can be rewritten as follows:

$$\mathbf{H}_{i,j} \propto - \begin{bmatrix} \frac{(x_j - x_i)^2}{r_{i,j}^{2+p}} & \frac{(x_j - x_i)(y_j - y_i)}{r_{i,j}^{2+p}} & \frac{(x_j - x_i)(z_j - z_i)}{r_{i,j}^{2+p}} \\ \frac{(y_j - y_i)(x_j - x_i)}{r_{i,j}^{2+p}} & \frac{(y_j - y_i)^2}{r_{i,j}^{2+p}} & \frac{(y_j - y_i)(z_j - z_i)}{r_{i,j}^{2+p}} \\ \frac{(z_j - z_i)(x_j - x_i)}{r_{i,j}^{2+p}} & \frac{(z_j - z_i)(y_j - y_i)}{r_{i,j}^{2+p}} & \frac{(z_j - z_i)^2}{r_{i,j}^{2+p}} \end{bmatrix}. \quad (17)$$

By comparing Eq. (17) to Eqs. (3-4), it is clear that, when p is equal to 1, there is the complete consistency between the ANM and the proposed FE truss structural model. Moreover, the bar rigidity in the truss model (i.e. the product $E_{i,j}A_{i,j}$) represents the proportionality factor hidden in the ANM in Eqs. (16-17).

Once the ANM Hessian matrix is computed based on the residues coordinates from the PDB file [44], it can be decomposed in order to find the 3N-6 non-zero eigenvalues and 3N-6 non-rigid mode shapes. The decomposition takes the following form:

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad (18)$$

where \mathbf{U} represents the matrix containing the 3N-6 eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix containing the 3N-6 non-zero eigenvalues λ_n of the Hessian matrix. Moreover, from the calculated set of eigenvalues and eigenvectors, the pseudoinverse of \mathbf{H} can be computed:

$$\tilde{\mathbf{H}}^{-1} = \sum_{n=7}^{3N} \frac{1}{\lambda_n} \mathbf{U}_n \mathbf{U}_n^T, \quad (19)$$

which contains the information about the fluctuations of residues around the equilibrium position.

3. Results and Discussion

Hen egg-white lysozyme was analyzed here (pdb code: 4ym8). As reported above, the total mass of the protein (14.4 kDa $\sim 2.37 \times 10^{-23}$ kg) was equally divided among the 129 residues and the numerical scaling reported in [38,40] was adopted for inserting the input parameters into the LUSAS FE code. This led to a numerical mass value assigned to each C_α atom equal to 18.37 kg. Five different structural models were then created by varying the cutoff values in the range 8-20 Å (Fig. 1). In order to obtain the best match between the calculated and experimental B -factors, the numerical rigidity values (EA) shown in Tab. 1 were adopted for the different structural models. The table shows also the mean length of the bonds included into the model.

Tab. 1 Numerical rigidity values (EA) adopted for the five structural models

Model	Cutoff [Å]	Optimal EA [N]	Mean length [Å]
A	8	8.31×10^{-10}	5.71
B	10	2.35×10^{-10}	7.21
C	12	1.24×10^{-10}	8.61
D	15	0.71×10^{-10}	10.59
E	20	0.45×10^{-10}	13.46

As can be noted from Tab. 1, higher cutoff values imply lower bar rigidities. By dividing the optimal bar rigidity by the mean bond length from Tab. 1, one can obtain a measure of the mean stiffness k_{mean} associated to the different structural models (Fig. 2). Lower stiffness values need to be used when increasing cutoff values, and a remarkable reduction can be observed especially when passing from model A to model B ($\sim -78\%$).

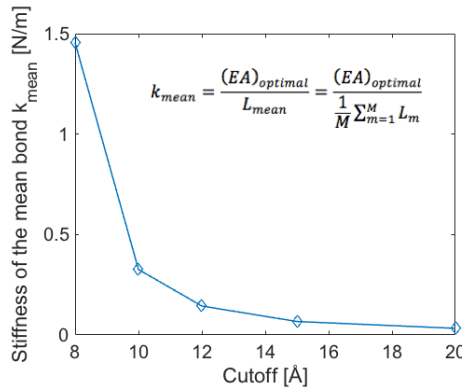


Fig. 2 Cutoff vs. mean bar stiffness values

By using the parameters reported above, the B -factors shown in Fig. 3 were obtained. In particular: Figs. 3a-e show the comparisons between the calculated B -factors from the space truss model (Eq. 9), the ones deriving from ANM [42] and the experimental ones available on the PDB file; in the same way, Figs. 3f-j show the comparisons between the normalized B -factors from (Eq. 10), the normalized ones deriving from ANM [42] and the normalized experimental ones. In particular, the black curves refer to the experimental values, the gray ones to the ANM results, whereas five different colors were used for the results from the five different models (blue: model A, magenta: model B, orange: model C, purple: model D, green: model E).

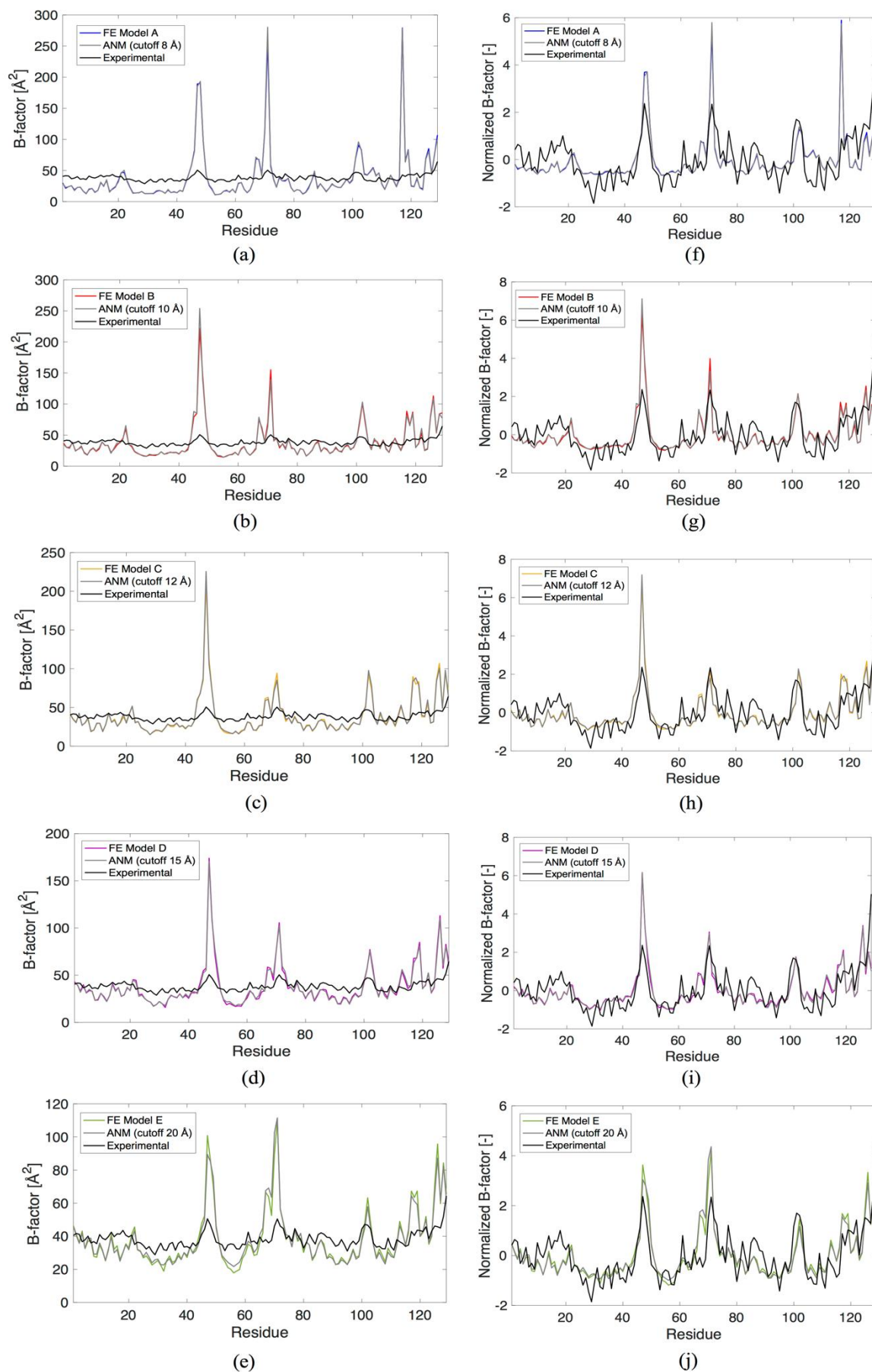


Fig. 3 Numerical and experimental B -factors: (a-e) absolute B -factors; (f-j) normalized B -factors

By observing Figs. 3a-e (absolute B -factors) and particularly Figs. 3f-j (normalized B -factors), it can be noted that the correlation between numerical and experimental results is generally acceptable, and it increases for higher cutoff values. In fact, by calculating the Pearson correlation coefficient between the calculated and experimental B -factors, one obtains the results shown in Tab. 2 and Fig. 4. The correlation coefficients arising from the proposed FE model vary from 0.57 (model A) up to 0.72 (model E), which are in agreement with the correlation values from ANM (Tab. 2 and Fig. 4). It should be noted that experimental B -factors from X-ray diffraction are not error-free and may not reflect only thermal fluctuations: this could indeed prevent a clear and ultimate comparison between experimental and numerical values, and this is also the reason why it is generally difficult to obtain very high correlation coefficients (e.g. ~ 1). However, the remarkable agreement which has been found in the literature [17,42] between the experimental B -factors and the ones calculated via simple mechanical models (e.g. NMA on coarse-grained models) suggests that these models are anyway able to identify the key parameters governing protein dynamics (e.g. the overall protein topology).

Tab. 2 Correlation coefficients between experimental and numerical B -factors from FE model (present study) and ANM [42]

Model	FE model	ANM [42]
A	0.57	0.56
B	0.67	0.64
C	0.66	0.64
D	0.69	0.68
E	0.72	0.69

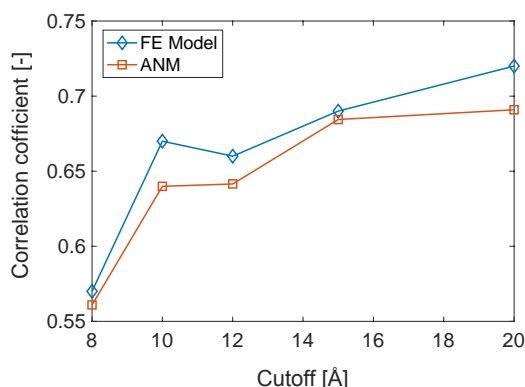


Fig. 4 Correlation coefficients vs. cutoff values from FE model (present study) and ANM [42]

Finally, from Fig. 3 and as has already been remarked in Section 2, it is evident how the FE structural model is consistent with the ANM and provides basically the same results, since the colored curves are always entirely overlapped to the gray ones: in fact, the B -factor profiles are almost identical, showing just minor differences at specific residues. Therefore, the space truss model, developed by a purely engineering approach, can be seen as the counterpart of ANM by following a purely Structural Mechanics approach and can be a valid candidate to investigate protein dynamical behavior.

In Fig. 5, the results from the five different truss models are shown together, both in terms of absolute B -factors (Fig. 5a) and normalized B -factors (Fig. 5b), in order to highlight the differences. As can be observed, the main differences arise in the portions of the protein chain which are predicted to be the most flexible ones, i.e. the regions in which the B -factor distributions exhibit higher peaks. Specifically these are found to involve mainly the 46-49 portion, as well as the 70-72, 101-103, 117-119 and 124-126 regions. As a general tendency, increasing the cutoff value (i.e. from model A to model E) leads to lower values of the calculated B -factors in these high-flexibility

regions, leading in turn to a better match with the experimental B -factor distribution. This tendency is reasonable, since increasing the cutoff value implies a more uniform distribution of the connections throughout the three-dimensional structure of the protein, and this in turn should be reflected in a more uniform distribution of the calculated B -factors along the protein chain, especially when dealing with globular proteins, such as lysozyme. Finally, it is interesting to observe that, in the other portions of the protein chain (i.e. outside of the most flexible regions), the five different models provide essentially the same values and distribution of the calculated B -factors.

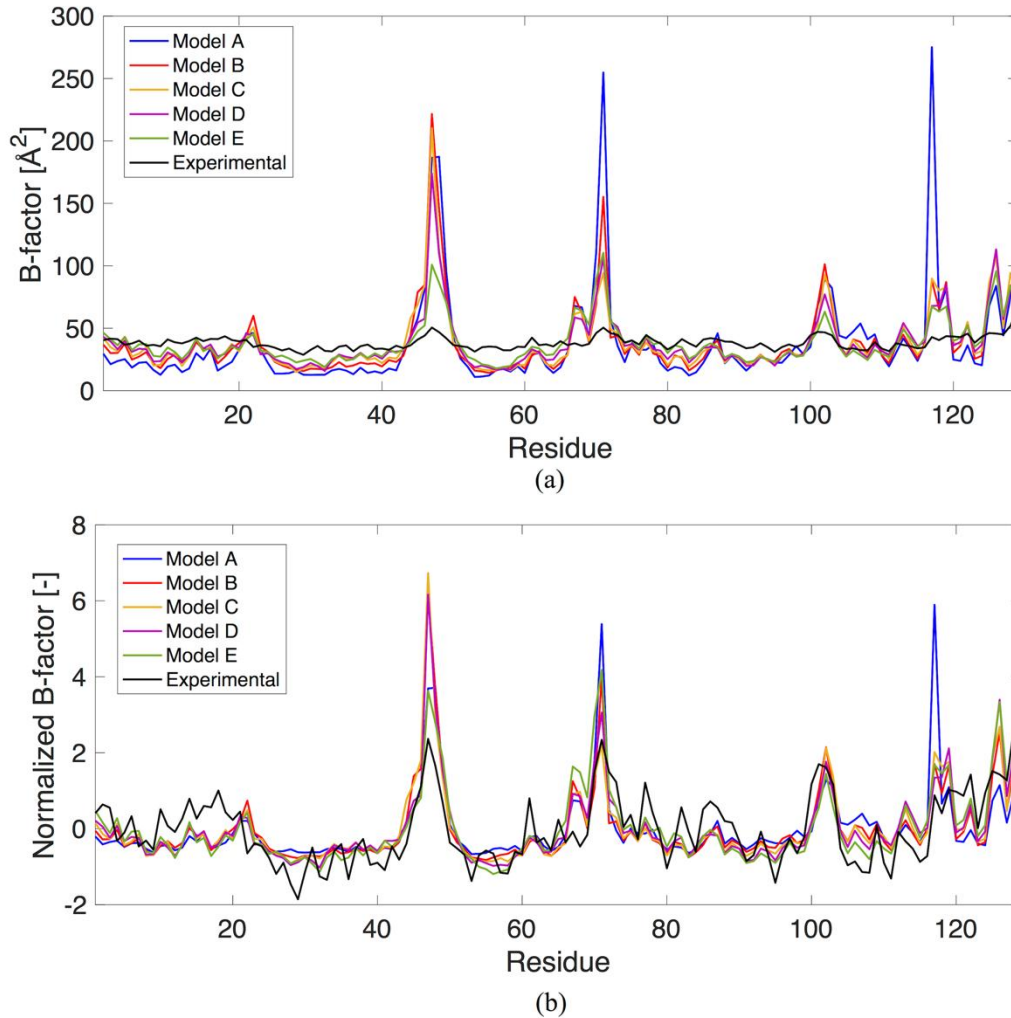


Fig. 5 Comparison between B -factors arising from the five different models: **(a)** absolute B -factors; **(b)** normalized B -factors

Since B -factors and in turn protein dynamics are dominated by the lowest vibrational modes, a detailed analysis of the characteristics of the first two vibrational modes is presented below. In particular, Fig. 6a shows the normalized modal displacements found for the lowest mode, which corresponds to the hinge-bending motion found by other authors [13], from the five considered models. As can be seen, although the models differ for their mechanical properties (total number of bars and stiffness parameters), they provide approximately the same modal displacements along the protein chain. The most flexible region is identified as being the 46-49 portion by all the models. However, model A (blue) and model E (green) exhibit some specific features: in particular, model E highlights another high-flexibility region in the 70-72 portion.

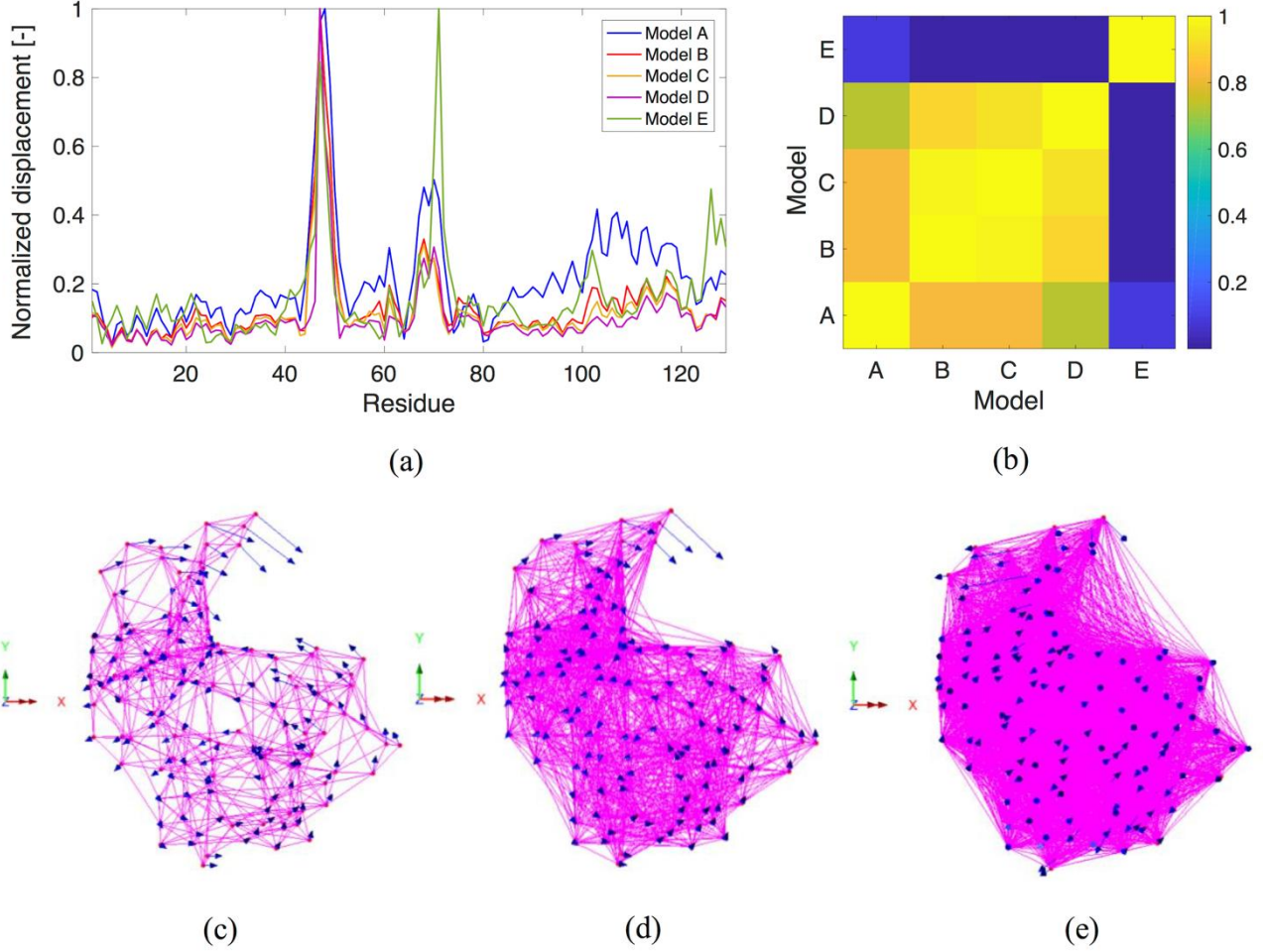


Fig. 6 First non-trivial vibrational modes: **(a)** normalized displacements; **(b)** MAC matrix; **(c)** 3D-shape from model A; **(d)** 3D-shape from model C; **(e)** 3D-shape from model E

A more quantitative comparison in terms of modal displacements obtained from the five models can be achieved by means of MAC (Modal Assurance Criterion) [46,47], which is defined as follows:

$$MAC_{m,n} = \frac{(\boldsymbol{\delta}_m^T \cdot \boldsymbol{\delta}_n)^2}{(\boldsymbol{\delta}_m^T \cdot \boldsymbol{\delta}_m)(\boldsymbol{\delta}_n^T \cdot \boldsymbol{\delta}_n)}, \quad (20)$$

where $\boldsymbol{\delta}_m$ and $\boldsymbol{\delta}_n$ represent the vectors of the m^{th} and n^{th} mode displacements, respectively. This index is a powerful tool to investigate modal analysis results, as it allows to explore the consistency between different mode shapes (or sets of mode shapes). Its value ranges from 0 to 1: 0 means completely uncorrelated or orthogonal vectors, while 1 indicates a perfect match. Moreover, it has been widely used in Structural Mechanics for studying the correspondence between modes deriving from numerical calculations and those arising from experimental investigations, detected for example by piezoelectric transducers [48].

Applying MAC operator to the first vibrational modes obtained from the five models (Fig. 6a), one obtains the results depicted in Fig. 6b, by means of a 2D matrix representation. Each grid square, of coordinates i and j , stands for the combination of the first non-trivial modes obtained from model i and model j . MAC values are shown by means of colors, ranging from dark blue (MAC = 0) to light yellow (MAC = 1). High MAC values can be found for the comparison between models A, B, C and D, whereas MAC values associated to the comparison between model E and the other ones are

much lower, which reflects the peculiarity of the mode shape from model E reported in Fig. 6a (green curve).

Note that MAC operator is not only sensitive to the displacement distribution along the protein chain (Fig. 6a), but also to the directionality of such displacements at each node. For sake of clarity, in Figs. 6c, 6d and 6e the three-dimensional representation of the first non-trivial modes is shown for model A, C and E, respectively. As can be seen, high similarity can be observed between the vector directions provided by model A and model C ($MAC_{1A,1C} \sim 0.82$), which depict an hinge-bending motion implying the opening-closing of the binding cleft. Otherwise, significant differences arise in terms of displacement directionality when it comes to model E (Fig. 6e). This is confirmed by the low MAC values found when comparing the first mode from model E to the first one from model A ($MAC_{1A,1E} \sim 0.08$) and model C ($MAC_{1C,1E} \sim 0.002$).

In Fig. 7, the same analysis is shown regarding the second eigenmodes. In particular, Fig. 7a shows the comparison in terms of modal displacements and Fig. 7b reports the calculated MAC values for the comparison of the second eigenmodes. As can be seen, the second modes are related to high flexibility in the 46-49 and 70-72 protein regions, from all the models. However, also in this case, model E exhibits specific features since the relative displacements in the 46-49 portion are much lower than those arising from the other models. As already found for the MAC comparison between the first modes, Fig. 7b shows that the second modes are found to be consistent for the different models (high MAC values), except for model E (lower MAC values). Again, these low MAC values are caused by the difference in terms of displacement directionalities, as can be seen by comparing the 3D-shape of model E (Fig. 7e) to the mode shapes from model A (Fig. 7c) and model C (Fig. 7d).

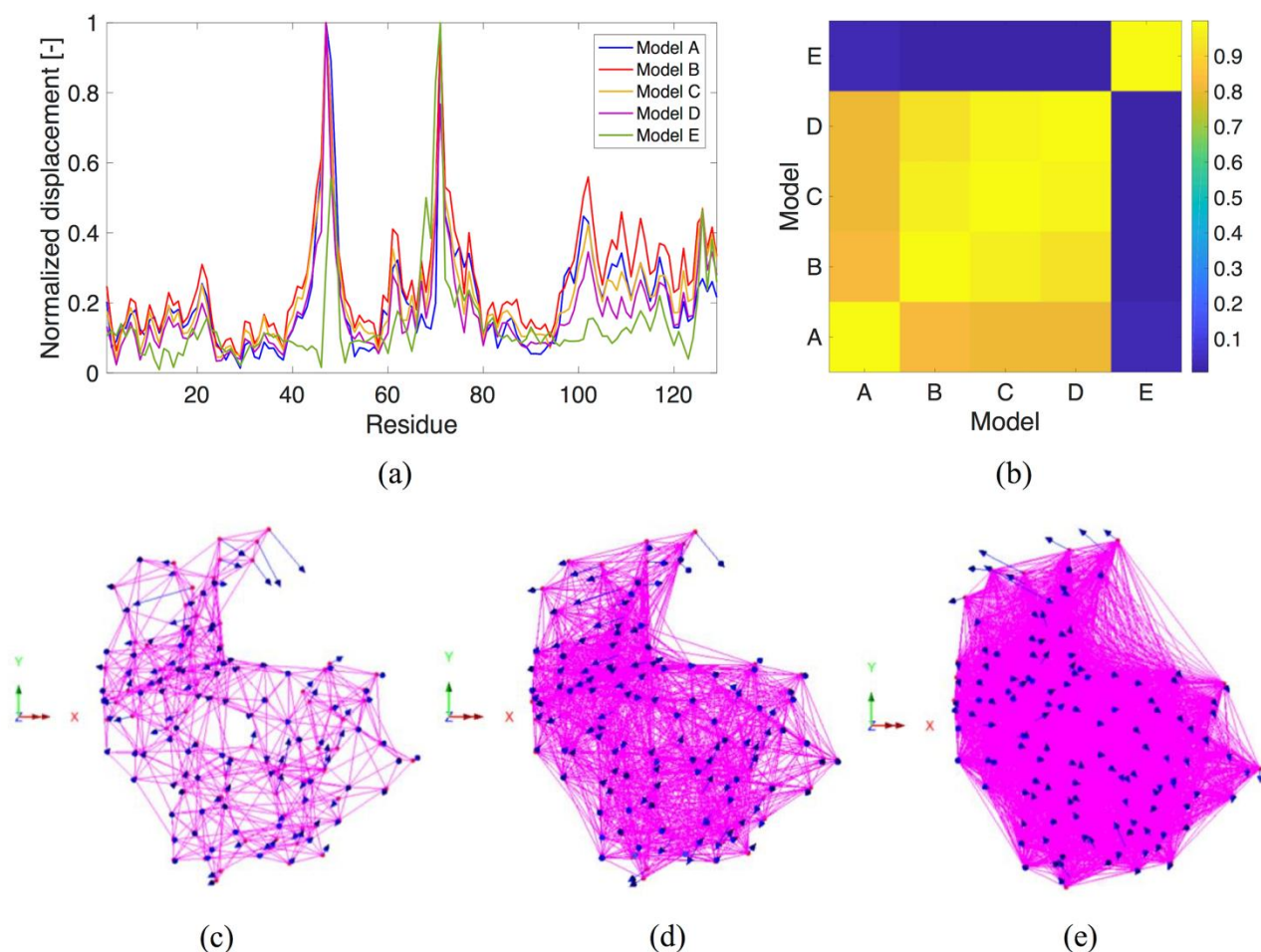


Fig. 7 Second non-trivial vibrational modes: (a) normalized displacements; (b) MAC matrix; (c) 3D-shape from model A; (d) 3D-shape from model C; (e) 3D-shape from model E

Since the first two modes deriving from model E seem to be uncorrelated to the respective first two mode shapes arising from the other models, it has been conjectured that increasing the cutoff value might have created an exchange between the lowest modes. Indeed, this has been confirmed by calculating MAC values between the first mode from model E (Fig. 6a, green curve) and the second one from the other models (Fig. 7a, blue, magenta, orange and purple curves). The results are shown in Fig. 8, and confirm that the first mode from model E, which implies high flexibility in the 46-49 and 70-72 regions, exhibit high similarity with the second modes from the other models ($\text{MAC} > 0.72$). This can also be verified by comparing the 3D-shape related to the first mode from model E (Fig. 6e) to the 3D-shapes related to the second eigenmodes from model A (Fig. 7c) and model C (Fig. 7d) and by observing that there exist several common features in the displacement field. This is confirmed by the obtained cross MAC values (Fig. 8), which are found to be equal to 0.72 ($\text{MAC}_{1E,2A}$) and 0.91 respectively ($\text{MAC}_{1E,2C}$).

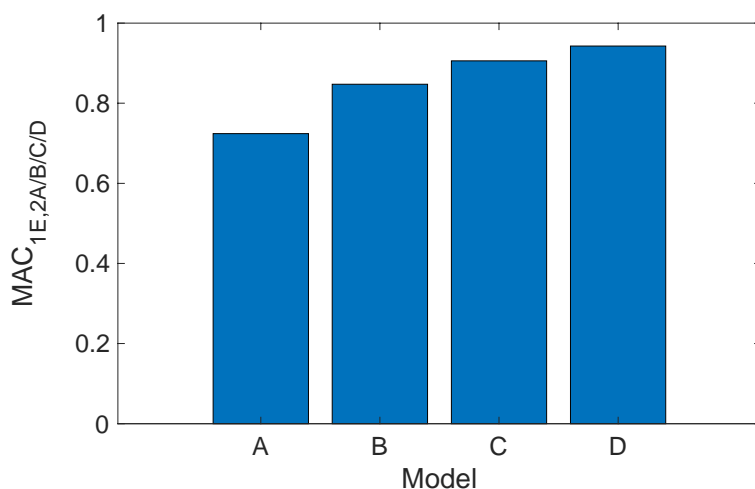


Fig. 8 MAC values between the 2nd mode from model E and the 1st modes from other models

Finally, in Fig. 9 the values of the first non-zero vibrational frequencies are shown depending on the mechanical parameters reported above (Tab. 1), which are associated to the low-frequency motions. As can be seen, the values of the vibrational frequencies strongly depend on the considered cutoff value. The fundamental frequency (i.e. the frequency associated to the first eigenmode) varies from 0.046 THz ($\sim 1.5 \text{ cm}^{-1}$) for model A to 0.117 THz ($\sim 3.9 \text{ cm}^{-1}$) for model E. Note that, despite the variation, this frequency range is in line with the results arising from the calculations performed by Markelz *et al.* [49] using CHARMM (Chemistry at HARvard Molecular Mechanics) simulation, as well as the ones by Levitt *et al.* [13]. Note also that the lowest frequency mode obtained by Levitt *et al.* [13], which occurs at 2.98 cm^{-1} , was found to involve the highest displacements near residues 47, 70 and 103 and to constitute a global hinge-bending motion, which is very similar to the one found via the present truss model (Fig. 6). In Fig. 9 the frequencies associated up to the five eigenmode are also shown.

However, it should be noted that, from these kinds of models (i.e. coarse-grained mechanical models), we are able to obtain an estimate of the frequency range in which the slowest protein motions occur, but not an accurate evaluation of the exact frequencies; in fact, as mentioned above and shown in Fig. 9, changing the model parameters leads to a not negligible change in the obtained frequency values. Exact estimates of such frequency values may be prevented by the simplified nature of these models, which condense all the degrees of freedom at the C_α atoms and do not take into account the effects of external conditions, e.g. damping phenomena caused by the specific environment around the protein. However, these models are believed to constitute a useful and low-computation-cost tool to get information about the frequency range of the slowest protein vibrations.

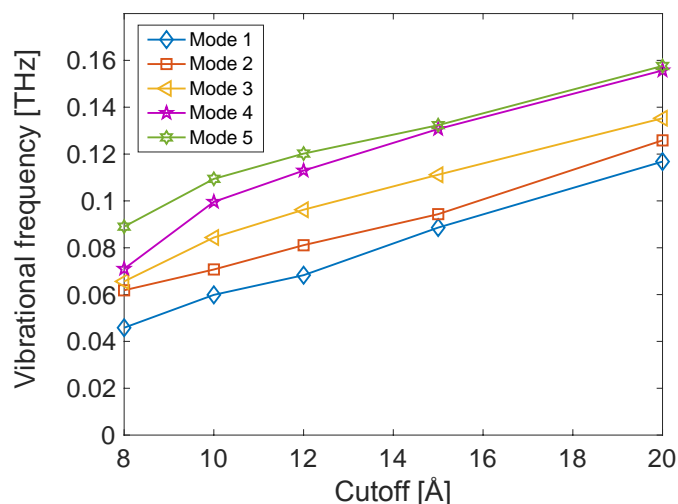


Fig. 9 Lowest vibrational frequencies

4. Conclusions

In this paper, a finite-element-based space truss model, which was developed by adopting a purely mechanical approach, was presented for the investigation of low-frequency protein vibrations. The model consists of point masses, concentrated in correspondence to the protein residues, connected by linear-elastic bar elements simulating both short- and long-range interactions (non-local model). Spherical hinges connect the bars in correspondence to the point masses. The total protein mass is equally distributed among the residues, whereas the elastic stiffness of each bar simulating the interaction between residues is inversely proportional to the connection distance. The case of hen egg-white lysozyme was selected as a benchmark and its mechanical vibrations were analyzed by using five different models, with changing interaction cutoff values (8, 10, 12, 15 and 20 Å). The validity of the model was assessed by comparing the estimated B -factors, arising from the calculated eigenmodes and eigenfrequencies, to the experimental ones available on the PDB file. Moreover, the influence of cutoff values on the optimal stiffness was investigated, and it was shown that the presented approach is consistent with the Anisotropic Network Model (ANM). In a sense, the proposed truss model can be seen as the counterpart of the ANM by following a purely Structural Mechanics approach. Then, a detailed description of the lowest motions of lysozyme was provided, highlighting the most flexible regions, and Modal Assurance Criterion (MAC) was used in order to calculate the similarity among eigenmodes deriving from the different models. Finally, the dependence of the lowest vibrational frequencies on the interaction cutoff was investigated, thus identifying the THz frequency range in which the fundamental lysozyme vibrations occur.

The proposed model represents an engineering approach for analyzing small-amplitude low-frequency (i.e., global) vibrations of protein structures, in which some approximations are introduced by looking at both experimental and modelling outcomes. The proposed methodology is in line with several other contributions in the Biomechanics and Mechanobiology communities, aiming at embedding crucial features of complex biomechanical systems via a simplified structural approach. At the same time, the main limitation of the model lies in the fact that it is rather simplified and it is now not able to take into account important factors, other than the intrinsic protein topology, which may indeed have a strong influence on protein behavior, e.g. the actual external environment, pH level, interaction with water and other molecules, etc. To overcome some of such limitations, one possibility is to provide the model some fictitious parameters (e.g. internal damping) which could take into account, at least in part, some external factors (e.g. water). Nevertheless, in line with other elastic network models present in the literature, it has proven to be an effective, yet simple, tool for the investigation of protein dynamical features.

The proposed truss model can thus be seen as a simple method, accessible to the Structural Mechanics scientific community members, for analyzing protein vibrations and their connections with biological activity. For example, some of the authors made use of this approach for studying the similarity between low-frequency vibration modes and conformational changes: this problem will be dealt with in future contributions. Moreover, the truss model constitutes a powerful tool for investigating not only protein dynamics but, more in general, protein mechanics, since it also allows for investigating the response of proteins when subjected to external forces, applied both in a static and dynamic way. Lastly, it is also suitable in perspective for analyzing, from a Structural Mechanics viewpoint, the stability of proteins when they undergo large conformational transitions, by making use of geometric nonlinear analyses.

References

- [1] Ovchinnikov, S., Park, H., Varghese, N., Huang, P. S., Pavlopoulos, G. A., Kim, D. E., Kamisetty, H., Kyrpides, N. C., Baker, D. (2017) Protein structure determination using metagenome sequence data. *Science*. 355, 294-298
- [2] Mofrad, M. R. K., Kamm, R. D. (2009) *Cellular Mechanotransduction: Diverse Perspectives from Molecules to Tissues*. Cambridge University Press, Cambridge
- [3] Puglisi, G., De Tommasi, D., Pantano, M. F., Pugno, N. M., Saccomandi, G. (2017) Micromechanical model for protein materials: From macromolecules to macroscopic fibers. *Physical Reviews*. 96, 042407
- [4] Maceri, F., Marino, M., Vairo, G. (2010) A unified multiscale mechanical model for soft collagenous tissues with regular fiber arrangement. *Journal of Biomechanics*. 43, 355-363
- [5] Marino, M., Vairo, G. (2014) Influence of inter-molecular interactions on the elasto-damage mechanics of collagen fibrils: A bottom-up approach towards macroscopic tissue modeling. *Journal of the Mechanics and Physics of Solids*. 73, 38-54
- [6] Pandolfi, A., Gizzi, A., Vasta, M. (2019) A microstructural model of cross-link interaction between collagen fibrils in the human cornea. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 377, 20180079
- [7] Deshpande, V. S., McMeeking, R. M., Evans A. G. (2006) A bio-chemo-mechanical model for cell contractility. *Proceeding of the National Academy of Sciences of the United States of America*. 103, 14015-14020
- [8] Chen, K., Vigliotti, A., Bacca, M., McMeeking, R., Deshpande, V. S., Holmes, J. W. (2018) Role of boundary conditions in determining cell alignment in response to stretch. *Proceeding of the National Academy of Sciences of the United States of America*. 115, 986-991
- [9] Shishvan, S. S., Vigliotti, A., Deshpande, V. S. (2018) The homeostatic ensemble for cells. *Biomechanics and Modeling in Mechanobiology*. 17, 1631-1662
- [10] Buskermolen, A. B. C., Suresh, H., Shishvan, S. S., Vigliotti, A., DeSimone, A., Kurniawan, N. A., Bouten, C. V. C., Deshpande, V. S. (2019) Entropic Forces Drive Cellular Contact Guidance. *Biophysical Journal*. 116, 1994-2008
- [11] Yang, L. Q., Sang, P., Tao, Y., Fu, Y. X., Zhang, K. Q., Xie, Y. H., Liu, S. Q. (2014) Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *Journal of Biomolecular Structure and Dynamics*. 32, 372-393
- [12] Dykeman, E. C., Sankey, O. F. (2010) Normal mode analysis and applications in biological physics. *Journal of Physics: Condensed Matter*. 22, 423202.
- [13] Levitt, M., Sander, C., Stern, P. S. (1985) Protein Normal-mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme. *Journal of Molecular Biology*. 181, 423-447.
- [14] Tirion, M. M. (1996) Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*. 77, 1905-1908
- [15] Bahar, I., Atilgan, A. R., Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*. 2, 173-181

- [16] Hinsen, K. (1998) Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins: Structure, Function and Genetics*. 33, 417-429
- [17] Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., Bahar, I. (2001) Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*. 80, 505-515
- [18] Nicolai, A., Delarue, P., Senet, P. (2014) Low-Frequency, Functional, Modes of Proteins: All-Atom and Coarse-Grained Normal Mode Analysis. In: Liwo, A. (ed.) *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*. Springer Series in Bio-/Neuroinformatics. Springer, Berlin, Heidelberg
- [19] Wako, H., Endo, S. (2017) Normal mode analysis as a method to derive protein dynamics information from the Protein Data Bank. *Biophysical Reviews*. 9, 877-893
- [20] Tozzini, V. (2005) Coarse-grained models for proteins. *Current Opinion in Structural Biology*. 15, 144-150
- [21] Bahar, I., Rader, A. J. (2005) Coarse-grained normal mode analysis in structural biology. *Current Opinion in Structural Biology*. 15, 586-592
- [22] Rader, A. J. (2010) Coarse-grained models: getting more with less. *Current Opinion in Pharmacology*. 10, 753-759
- [23] Takada, S. (2012) Coarse-grained molecular simulations of large biomolecules. *Current Opinion in Structural Biology*. 22, 130-137
- [24] Chou, K. C., Chen, N. Y (1977) The biological functions of low-frequency phonons. *Scientia Sinica*. 20, 447-457
- [25] Tama, F., Sanejouand, Y. H. (2001) Conformational change of proteins arising from normal mode calculations. *Protein Engineering*. 14, 1-6
- [26] Zheng, W., Brooks, B. R. (2005) Normal-Mode-Based Prediction of Protein Conformational Changes Guided by Distance Constraints. *Biophysical Journal*. 88, 3109-3117
- [27] Petrone, P., Pande, V. S. (2006) Can Conformational Change Be Described by Only a Few Normal Modes? *Biophysical Journal*. 90, 1583-1593
- [28] Mahajan, S., Sanejouand, Y. H. (2015) On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Archives of Biochemistry and Biophysics*. 567, 59-65
- [29] Nicolai, A., Barakat, F., Delarue, P., Senet, P. (2016) Fingerprints of Conformational States of Human Hsp70 at Sub-THz Frequencies. *ACS Omega*. 1, 1067-1074
- [30] Lucia, U. (2016) Electromagnetic waves and living cells: A kinetic thermodynamic approach. *Physica A: Statistical Mechanics and its Applications*. 461, 577-585
- [31] Lucia, U., Grisolia, G., Ponzetto, A., Silvagno, F. (2017) An engineering thermodynamic approach to select the electromagnetic wave effective on cell growth. *Journal of Theoretical Biology*. 429, 181-189
- [32] Lucia, U., Ponzetto, A. (2017) Some thermodynamic considerations on low frequency electromagnetic waves effects on cancer invasion and metastasis. *Physica A: Statistical Mechanics and its Applications*. 467, 289-295
- [33] Barth, A. (2007) Infrared spectroscopy of proteins. *Biochimica et Biophysica Acta*. 1767, 1073-1101
- [34] Acbas, G., Niessen, K. A., Snell, E. H., Markelz, A. G. (2014) Optical measurements of long-range protein vibrations. *Nature Communications*. 5, 3076
- [35] Xie, L., Yao, Y., Ying, Y. (2014) The Application of Terahertz Spectroscopy to Protein Detection: A Review. *Applied Spectroscopy Reviews*. 49, 448-461
- [36] Brown, K. G., Erfurth, S. C., Small, E. W., Peticolas, W. L. (1972) Conformationally Dependent Low-Frequency Motions of Proteins by Laser Raman Spectroscopy. *Proceeding of the National Academy of Sciences of the United States of America*. 69, 1467-1469
- [37] Painter, P. C., Mosher, L. E., Rhoads, C. (1982) Low-Frequency Modes in the Raman Spectra of Proteins. *Biopolymers*. 21, 1469-1472

- [38] Carpinteri, A., Lacidogna, G., Piana, G., Bassani, A. (2017) Terahertz mechanical vibrations in lysozyme: Raman spectroscopy vs modal analysis. *Journal of Molecular Structure*. 1139, 222-230
- [39] Lacidogna, G., Piana, G., Bassani, A., Carpinteri, A. (2017) Raman spectroscopy of Na/K-ATPase with special focus on low-frequency vibrations. *Vibrational Spectroscopy*. 92, 298-301
- [40] Carpinteri, A., Piana, G., Bassani, A., Lacidogna, G. (2019) Terahertz vibration modes in Na/K-ATPase. *Journal of Biomolecular Structure and Dynamics*. 37, 256-264
- [41] <http://www.lusas.com/>
- [42] Eyal, E., Yang, L. W., Bahar, I. (2006) Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*. 22, 2619-2627.
- [43] Eyal, E., Lum, G., Bahar, I. (2015) The anisotropic Network Model web server at 2015 (ANM 2.0). *Bioinformatics*. 31, 1487-9
- [44] <http://www.rcsb.org/>
- [45] Carpinteri, A. (2017) *Advanced Structural Mechanics*. CRC Press, Taylor & Francis Group, Boca Raton
- [46] Allemang, R. J., Brown, D. L. (1982) A correlation coefficient for modal vector analysis. *Proceedings of the 1st IMAC*, Orlando, Florida
- [47] Pastor, M., Binda, M., Harcarik, T. (2012) Modal Assurance Criterion. *Procedia Engineering*. 48, 543-548
- [48] Piana, G., Lofrano, E., Carpinteri, A., Paolone, A., Ruta, G. (2016) Experimental modal analysis of straight and curved slender beams by piezoelectric transducers. *Meccanica*. 51, 2797-2811
- [49] Markelz, A., Whitmire, S., Hillebrecht, J., Birge, R. (2002) THz time domain spectroscopy of biomolecular conformational modes. *Physics in Medicine and Biology*. 47, 3797-3805.