Numerical Evaluation of Protein Global Vibrations at Terahertz Frequencies by means of Elastic Lattice Models [†]

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

Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; giuseppe.lacidogna@polito.it (G.L.); gianfranco.piana@polito.it (G.P.); alberto.carpinteri@polito.it (A.C.)

* Correspondence: domenico.scaramozzino@polito.it

[†] Presented at 1st International Electronic Conference on Applied Sciences, 10–30 November 2020; Available online: https://asec2020.sciforum.net/.

Abstract: Proteins represent one of the most important building blocks for most biological processes. Their biological mechanisms have been found to correlate significantly with their dynamics, which is commonly investigated through Molecular Dynamics (MD) simulations. However, important insights on protein dynamics and biological mechanisms have also been obtained via much simpler and computationally efficient calculations based on Elastic Lattice Models (ELMs). The application of Structural Mechanics approaches, such as modal analysis, to the protein ELMs has allowed to find impressive results in terms of protein dynamics and vibrations. The low-frequency vibrations extracted from the protein ELM are usually found to occur within the terahertz (THz) frequency range and correlate fairly accurately with the observed functional motions. In this contribution, the global vibrations of lysozyme will be investigated by means of a Finite Element (FE) truss model and we will show that there exists complete consistency between the proposed FE approach and one of the more well-known ELMs for protein dynamics, the Anisotropic Network Model (ANM). The proposed truss model can consequently be seen as a simple method, easily accessible to the Structural Mechanics community members, to analyze protein vibrations and their connections with the biological activity.

Keywords: protein vibrations; elastic lattice model; terahertz frequency; lysozyme; B-factors

1. Introduction

The fundamental paradigm that has characterized most of the research in the field of protein computational biology relies on the sequence-structure-dynamics-function relationship [1]. The linear sequence of amino acids arranges itself into the tertiary structure of the protein through the complex chemo-physical phenomenon of protein folding [2,3]. The folded protein structure is not a static entity, but it is generally highly dynamic and flexible. The resulting dynamics of the protein obviously depends on the three-dimensional protein structure. Finally, it has been demonstrated that the dynamics and flexibility of the protein acts towards its functionality and biological mechanism [1,4,5]. Therefore, the sequence of amino acids generates the protein structure, which in turn drives the biological function through its dynamic and flexibility features.

Several computational approaches have been suggested to reproduce the protein dynamics and vibrations correctly. Molecular Dynamics (MD) and normal mode analysis (NMA) represent two of the most widely used methodologies [6,7]. The former looks at the complete trajectories of the atoms, whereas the latter investigates the small-amplitude normal modes of the system, which is characterized by multi-parameter harmonic potentials. Later on, it has been shown that simplified

models relying only on a single-parameter harmonic potential, i.e., an ensemble of Hookean springs, are sufficient to describe the slow dynamics of the protein with good details [8]. Then, it has also been found that even coarse-grained models, only based on the C^α atoms of the protein, are able to properly reproduce the protein slow dynamics as well as the fluctuations observed in the X-ray crystallographic experiments, i.e., the B-factors [9,10].

Here, we report the results that arise from the application of modal analysis to the Elastic Lattice Model (ELM) of hen egg-white lysozyme, a well-known globular protein. The structural model has been developed within a Finite Element (FE) framework and the outcomes have provided useful insights regarding the lysozyme flexibility, the biologically relevant motions as well as the vibrational frequencies. In particular, the latter have been found to lie in the (sub-)THz frequency range. More detailed information about this analysis and the obtained results can be found in [11].

2. Materials and Methods

The structural model was developed within the FE code LUSAS [12] and relies on modelling the protein structure as an ELM, i.e., a 3D spatial truss. The elastic bars of the truss were generated by taking into account only the C^α atoms of the lysozyme (PDB: 4YM8) and by connecting those nodes that lie closer than a selected cutoff value. Five different cutoffs were selected, i.e., 8, 10, 12, 15 and 20 Å, in order to generate various ELMs and investigate the influence of such a geometrical parameter on the outcomes. The elastic bars were assigned the same axial rigidity *EA*, where *E* is the Young's Modulus and *A* the cross-sectional area. The value of rigidity was defined based on the comparison between the experimental B-factors and the computed ones (see below). The masses of all the ELM nodes were set equal among them, and equal to the total mass of the protein divided by the number of nodes [11].

Based on the generated ELMs, modal analysis was carried out, by solving the multi-degree-of-freedom (MDOF) free-vibration problem within the FE environment. This implies solving the following eigenvalue-eigenvector equation:

$$(\boldsymbol{K} - \omega_n{}^2 \boldsymbol{M})\boldsymbol{\delta}_n = \boldsymbol{0}, \tag{1}$$

being $\boldsymbol{K}$ and $\boldsymbol{M}$ the global stiffness and mass matrix of the ELM, respectively, and ω_n and $\boldsymbol{\delta}_n$ the eigenvalue (angular frequency) and eigenvector (mode shape) related to the nth vibrational mode. Note that, for an ELM counting N nodes and which is not externally constrained, we obtain $3N$-6 non-zero eigenvalues and $3N$-6 non-rigid eigenvectors [11]. The stiffness and mass matrix were computed within the FE environment as [13]:

$$\boldsymbol{K} = \sum_{b=1}^{B} \boldsymbol{C}_b{}^T \boldsymbol{N}_b{}^T \boldsymbol{k}_b{}^* \boldsymbol{N}_b \boldsymbol{C}_b, \qquad \qquad \boldsymbol{M} = m\boldsymbol{I}, \tag{2}$$

where $k_b{}^*$ represents the 2 × 2 stiffness matrix of the bth elastic bar in the local reference system, $\boldsymbol{N}_b$ is the 2 × 6 rotation matrix for the bth bar between the local and global reference systems, $\boldsymbol{C}_b$ is the 6 × $3N$ expansion matrix which is used to expand the stiffness matrix of the bth bar to the global structural dimension, m is the mass of each ELM node and $\boldsymbol{I}$ is the $3N \times 3N$ identity matrix [11].

One of the most used methods for the extraction of the normal modes from the coarse-grained protein crystal structure is the Anisotropic Network Model (ANM) [10]. The FE-based ELM described above can be seen as the counterpart of the ANM by following a purely Structural Mechanics approach. The main difference is that the ELM defined above also includes the quantitative information about the mass of the system, which is commonly not included in ANM calculations [11]. Nevertheless, it can be easily demonstrated that the stiffness matrix of the ELM $\boldsymbol{K}$ corresponds to the Hessian matrix of the ANM $\boldsymbol{H}$. The latter can be represented as a $N \times N$ matrix, whose elements are 3 × 3 sub-matrices $\boldsymbol{H}_{ij}$. These contain the second partial derivatives of the harmonic potential of the linear spring connecting nodes i and j (see [10,11,14] for more details). In [11], we have shown that there exists complete consistency between the ELM $\boldsymbol{K}$ and the ANM $\boldsymbol{H}$. In particular, when the decay parameter of the ANM spring constant is set equal to 1 [14], the two matrices are exactly the same.

Based on the eigenvalues and eigenvectors obtained from Equation (1), the B-factors can be numerically computed as [11,15]:

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

being B_i the B-factor of the ith node, k_B the Boltzmann's constant, T the absolute temperature, $\delta_{i,n}$ the absolute displacement of the ith node according to the nth mode and ω_n the angular frequency of the nth mode shape. The B-factors obtained from Equation (3) can be directly compared with the experimental ones contained in the PDB file [16]. Moreover, by posing that the average value of the experimental B-factors should match the average value of the computed ones, one can fix the axial rigidity EA to be assigned to the elastic bars of the ELM.
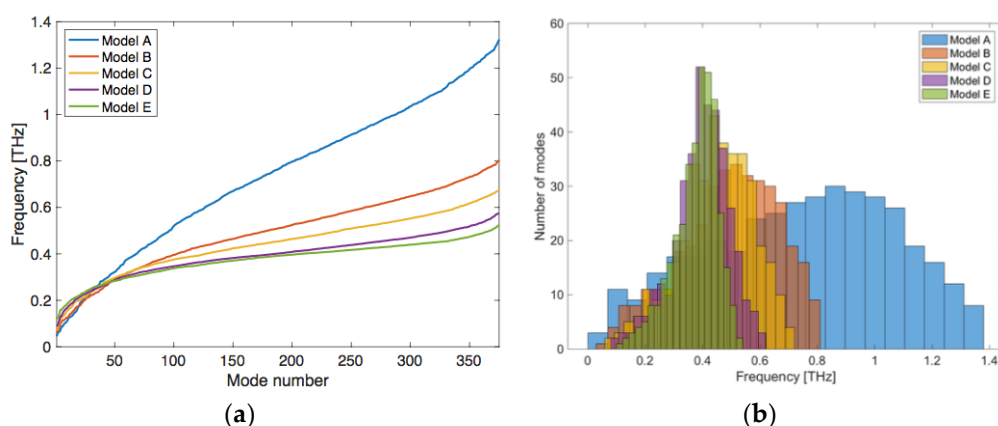
3. Results and Discussion

Table 1 reports the structural features of the five generated ELMs, according to the selected cutoff value. It can be seen that, increasing the cutoff value, the number of elastic connections within the ELM increases, therefore the mean length of the elastic bars is greater. However, in order to have consistency between the average value of the numerical and experimental B-factors, the axial rigidity EA assigned to the elastic bars needs to decrease as the cutoff increases. As a result, the axial stiffness of the mean elastic connection gets lower for higher cutoffs. Table 1 also shows the obtained Pearson correlation coefficients between the experimental and numerical B-factors. As can be seen, the correlation is around 60–70% and it tends to increase for higher cutoffs [11].

Table 1. The five generate ELMs for the lysozyme: structural features and correlation between the experimental and numerical B-factors.

Model	Cutoff (Å)	Mean Length of the Elastic Bars (Å)	EA (pN)	Stiffness of the Mean Connection (N/m)	Correlation B_{num} vs B_{exp}
A	8	5.71	831	1.455	0.57
B	10	7.21	235	0.326	0.67
C	12	8.61	124	0.144	0.66
D	15	10.59	71	0.067	0.69
E	20	13.46	45	0.033	0.72

From the analysis of the eigenvectors δ_n, i.e., the mode shapes, it was found that the low-frequency vibrational modes obtained from the ELMs of lysozyme reproduce fairly accurately the biologically relevant motion of the lysozyme [11]. Also, the ELM-based mode shapes were found to imply high flexibility in the same regions of the protein as previously obtained by Levitt et al. [7], who made use of NMA in internal coordinates. The cutoff parameter was found to have a certain influence on the obtained mode shapes. As extensively reported in [11], the mode shapes obtained from models with close cutoff values, i.e., 10, 12 and 15 Å, exhibit several common features. Conversely, using much lower and much higher values of the geometrical cutoff, i.e., 8 and 20 Å, leads to some differences in the displacement field of the mode shapes. Nevertheless, all models were efficiently able to detect the regions expected to have the highest flexibility.

From Equation (1) the angular frequencies ω_n associated to each mode shape were also obtained, which are linearly related to the vibrational frequencies f_n ($\omega_n = 2\pi f_n$). Figure 1 shows the frequencies for the lysozyme ELM, depending on the selected cutoff value. Figure 1a displays the entire set of frequencies according to the mode number, whereas Figure 1b reports the statistical distribution of these frequency values. As can be see, the vibrational frequencies lie in the (sub-)THz range. The lowest frequency ranges from 0.046 THz for Model A up to 0.117 THz for Model E [11]. It can be inferred that increasing the cutoff value leads to higher values of the frequency. This phenomenon occurs up to about the 50th mode. Afterwards, it can be seen that higher values of the frequency are obtained for the lower cutoff values. Finally, Figure 1b shows that the histograms of frequencies, that resemble Gaussian-like distributions. This seems to be a common feature to all globular proteins, as shown in [17,18]. It is also evident that, adopting when higher cutoff values for the protein ELM, the whole range of obtained frequencies becomes remarkably narrower.

Figure 1. Frequencies in the (sub-)THz frequency range for the lysozyme: (**a**) frequency for all obtained normal modes; (**b**) frequency distributions.

4. Conclusions

In this contribution, we have discussed the results of modal analysis on the lysozyme ELM, which were previously reported in [11]. In particular, we observed that the cutoff parameter has a certain influence on the assigned rigidity/stiffness values of the ELM members and on the correlation between the numerical and experimental B-factors. In this case, the correlation was found to increase as the geometrical cutoff increases. The obtained mode shapes were found to reproduce fairly accurately the biologically relevant motion of the enzyme, with some slight differences in the displacement fields due to the specific selected cutoff value. Finally, the frequency values were found to distribute according to Gaussian-like distributions in the (sub-)THz frequency range. The range of the complete set of frequencies was also found to shrink as the cutoff value of the ELM increases. Future studies will investigate more deeply such frequency distribution, which seems to be a common feature to all globular proteins.

References

1. Bahar, I.; Jernigan, R.L.; Dill, K.A. *Protein Actions: Principles & Modeling*; Garland Science: New York, NY, USA, 2017.
2. Levinthal, C. How to Fold Graciously. In *Mossbauer Spectroscopy in Biological Systems*; The University of Illinois Press: Champaign, IL, USA, 1969.
3. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230.
4. Karplus, M.; McCammon, J.A.; Gelin, B.R. Dynamics of folded proteins. *Nature* **1977**, *267*, 585–590.
5. Teilum, K.; Olsen, J.G.; Kragelund, B.B. Functional aspects of protein flexibility. *Cell. Mol. Life Sci.* **2009**, *66*, 2231.
6. Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6679–6685.
7. Levitt, M.; Sander, C.; Stern, P.S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* **1985**, *181*, 423–447.
8. Tirion, M.M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908.
9. 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.

10. 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.

11. Scaramozzino, D.; Lacidogna, G.; Piana, G.; Carpinteri, A. A finite-element-based coarse-grained model for global protein vibration. *Meccanica* **2019**, *54*, 1927–1940.

12. Available online: https://www.lusas.com (accessed on 6 November 2020).

13. Carpinteri, A. *Advanced Structural Mechanics*; CRC Press, Taylor & Francis Group: Boca Raton, FL, USA, 2017.

14. Eyal, E.; Lum, G.; Bahar, I. The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics* **2015**, *31*, 1487–1489.

15. Dykeman, E.C.; Sankey, O.F. Normal mode analysis and applications in biological physics. *J. Phys. Condens. Matter* **2010**, *22*, 423202.

16. 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.

17. Ben-Avraham, D. Vibrational normal-mode spectrum of globular proteins. *Phys. Rev. B* **1993**, *47*, 14559–14560.

18. Haliloglu, T.; Bahar, I.; Erman, B. Gaussian dynamics of folded proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090–3093.