Alternatives for the simulation of the lysozyme protein structure

João M. Regauer[1], Gustavo Giordani[1], Ignacio Iturrioz[1]

[1]*Dept. of Mechanical Engineering, Federal University of Rio Grande do Sul*
Rua Sarmento Leite, 425, ZIP:90050-170, Porto Alegre, Rio Grande do Sul, Brazil
joaoregauer@gmail.com, gustavo.giordani@ufrgs.br, ignacio@mecanica.ufrgs.br

Abstract. The application of solid mechanics has recently expanded to various new fields, which were unthinkable a few years ago. Studies focusing on structural characteristics of microscopic entities have helped other areas such as biochemistry to enhance our understanding of complex phenomena. Special attention has been given to proteins, where the dynamic, flexibility, and vibration analysis are fundamental to comprehend their biological activities. This work seeks to evaluate the protein's structure through a frame representation where the links simulate the chemical bonds, using finite element method and optimization procedures to verify the feasibility of solid mechanics techniques. The numerical simulation of three conformational states of the lysozyme protein is performed (PDB code: 1DPX, 1DPW, and 4YM8), where modal analysis is used to investigate the protein dynamics, obtaining vibrational frequencies, modes shapes, and a local measure of flexibility. Special attention is given to the influence of the secondary bonds on the protein's behavior. The results are compared using Pearson's correlation of the temperature factor distribution obtained experimentally for each atom. The application of optimization tools utilizing the temperature factor to define the links' stiffness increased the correlation with experimental results. The method brought acceptable outputs with low computational cost, with potential for improvements, indicating the technique's feasibility.

Keywords: Proteins, Lysozyme, Vibration modes, Finite element, Optimization.

1 Introduction

The biological activity of proteins occurs dynamically, so their flexibility and vibration are fundamental to understanding their complex mechanisms of action. Thus, the dynamics of structures are generally investigated both experimentally and numerically. Generally, the most used numerical technique is molecular dynamics (MD) [1], studying the movement of atoms and molecules applying equations associating all existing atomic interaction forces in the system. Escobedo et al. [2] use this method to study a polyglutamine helix related to spinobulbar muscle atrophy disease; however, this method requires significant computational power, making it impractical in most scenarios.

The movements associated with the lowest vibration modes are the most relevant from a biological point of view, and they can be simulated with simplified schemes such as Coarse-Grained (CG) [3] and All-Atom (AA) [4], where the structural model is defined through unifiliar finite elements that unite the interacting atoms. Giordani et al. [5] proposed improvements to the CG and AA models using the lysozyme protein (PDB code: 4YM8), the same protein used in this work. According to Ma J. [6], simplified methods like these show great potential in studying complex biological structures, especially when combined with experimental data used for their calibration. Another possible analysis is used by Hu & Buehler [8], who uses a CG model to analyze the vibration modes of different strains and mutations of spike proteins of coronavirus-2, comparing the flexibility of some regions with the lethality and infectivity of the virus

In experimental terms, the Raman Spectroscopy technique, used by Lacidogna et al. [7], is one way to bring information about vibration frequencies.

Main Goal: Explore the simulation of protein structure using numerical models of different levels of complexity.

Specific objectives: (a) Create models employing unifiliar finite elements to evaluate three conformational states of the Lysozyme Protein, using structural systems typically used in macrostructure simulation. (b) Adjust the parameters of these models using experimentally measured parameters. (c) Discuss the influence of secondary links in the proposed models.

2 Materials and methods

Proteins are between 1-100 nm in size, allowing only large aggregates of proteins to be seen under a microscope. Therefore, indirect methods are needed to visualize the structure, such as X-ray crystallography [9]. In this method, purified protein crystals are arranged in an x-ray beam, and the ray deflection pattern is used to predict the positions of atoms present in the protein crystal. Experimental data can be accessed through the Protein Data Bank (PDB) [10], where it is possible to define the coordinates of each protein atom, the chemical element, the amino acid it belongs to, and the B-factor. The B-factor is a measure of confidence regarding the atom's position around its coordinates, being an experimental parameter related to the atom's flexibility.

The Chimera software [11] allows the reading of the file exported by the PDB and provides the chemical links existing in the protein's primary structure. The Protein Contact Atlas (PCA) [12] is a database of non-covalent bonds, which allows the evaluation of existing structures in the PDB and provides details about their secondary and tertiary bonds. In this work, the PCA is used to obtain hydrogen bonds and weak hydrogen bonds.

The calculated B-factor from a modal analysis simulation can be obtained using Eq. 1, where B_i is the B-factor calculated for each atom i as the result of the square sum of the x, y, and z displacements, divided by the inverse of the angular frequency ω_n, for each mode of vibration and multiplied by a constant. The term k_b in Eq. 1 indicates the Boltzmann constant, and T is the absolute temperature. The first six vibration modes are disregarded as they contain the free movement of the structure. The numerical and experimental results are normalized to compare the deviation around the mean value locally, and Pearson's correlation coefficient is applied to compare both approaches.

$$B_i = \frac{8\pi^2}{3} k_b T \sum_{n=7}^{3N} \frac{\left(a_{ix}^2 + a_{iy}^2 + a_{iz}^2\right)_n}{\omega_n^2}. \tag{1}$$

Three different configurations of the lysozyme protein were chosen for this study, obtained by x-ray crystallography and with data available in the RCSB Protein Data Bank (PDB) with codes 4YM8 (Sugahara et al. [13]), 1DPX, and 1DPW (Weiss et al. [14]). The last two were obtained in a slightly alkaline environment, and in the 1DPW case, 2-methyl-2,4-pentanediol (MPD) was used as precipitant, which changes their conformation states.

The mass values, the rigidity of the chemical bond, the distances, and the type of bonding of the atoms are used to define the structural elements. The stiffness values are estimated from the chemical bond dissociation energy, given in kcal/mol, which is linearly converted to N/m using the simple bond between carbons as a reference, simplifying and approximating the bond dynamics. A similar model is used by Carpinteri et al. [15], which uses different stiffness values for each chemical bond in the simulation of amino acids and Lysozyme protein, but in this work, the mass of each atom and the section area of the elements is calculated individually. The stiffness of weak hydrogen and hydrogen bonds are estimated in the same way as for primary bonds, using an average of typical values of dissociation energy found in Jeffrey and Saenger [16].

In this study, the commercial finite element software ANSYS-APDL [17] is used for the simulations. Point mass elements are defined for each atom of the protein, and beam elements are defined to represent the interactions between the atoms. Furthermore, the constants of the moment of inertia and torsional modulus are set sufficiently high allowing the elements to have only axial movements. For each conformation state evaluated, four different models are built and simulated, namely:

- Primary bonds (P)
- Primary bonds with the weak hydrogen and hydrogen bonds (H)
- Primary bonds with Optimization (PO)
- Primary bonds with weak hydrogen and hydrogen bonds and Optimization (HO).

Experimental data are obtained from the PDB, the primary bonds are obtained from the Chimera software, and the weak hydrogen and hydrogen bonds are extracted from the PCA. The code using the APDL language

performs the assembly of the elements and the modal analysis of the structure.

An optimization process seeks to approximate the numerical model's dynamic behavior to the experimental results where the modulus of elasticity is defined by Eq. 2. Then, the equation is applied for each connection, where C is a constant to be defined, E is the modulus of elasticity used for the other simulations, and βm is the average of the normalized experimental B-factor between the atoms of each bond.

$$E(\boldsymbol{\beta m}) = E * C^{-\boldsymbol{\beta m}} \tag{2}$$

Parameter C defines how the stiffness of the elements is changed by the experimental B-factor, generating different structures with different dynamics. Simulations are performed iteratively to find C-value with the best correlation with the experimental B-factor, using the bisection method to converge to an optimal point. A stop criterion based on the number of iterations is used; 20 iterations are performed to test the model and verify the behavior of the correlation concerning parameter C.

The three-dimensional representation of the bonds and atoms for the 4YM8 structure can be seen in Fig. 2.1, taken from ANSYS APDL, where the lines in blue represent the primary bonds. Red lines represent weak hydrogen and hydrogen bonds, and blue asterisks highlight the point masses of atoms.

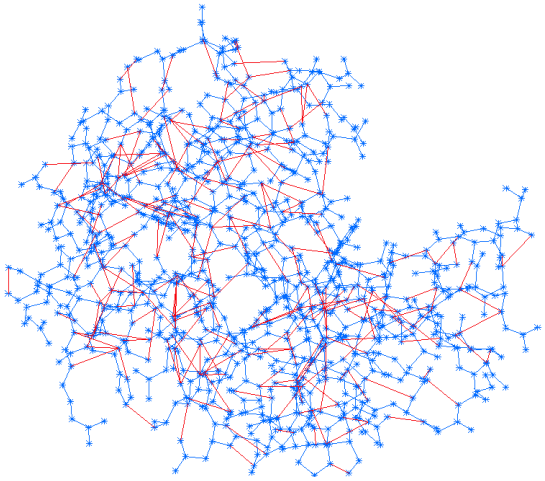


Figure 2.1 – Elements of the 4YM8 structure inserted into the ANSYS software, with lines representing bonds and asterisks representing the point masses of the atoms. In red bonds taken from the PCA, in blue bonds taken from the Chimera Software.

3 Results

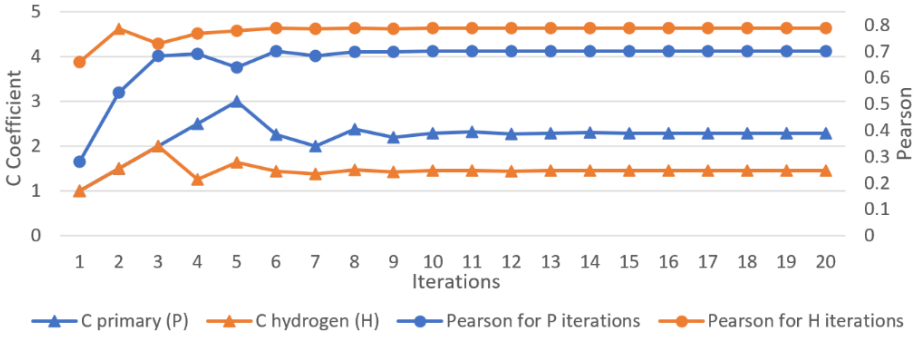
3.1 Analysis of optimization procedures

Two optimization processes for each conformation state are performed for the primary structure (P) and the primary structure with hydrogen bonds (H). In Figure 3.1(a), we have the results of the coefficient C (used in Eq. 2 to vary the modulus of elasticity of the elements) and the obtained Pearson coefficient with respect to the experimental B-factor for the simulations carried out concerning the 4YM8 state. In Figure 3.1(b), the values of the Pearson coefficients for each simulation performed are shown in the vertical axis and the C coefficient in the horizontal axis. In Fig. 3.1 (a), the hydrogen bonds (H) case shows a reduction in the Pearson correlation, using C = 2 in iteration 3, indicating that the best value of C is between 1 and 2. For primary bonds (P), this occurs in iteration 5, using C = 3, indicating an optimal value between 2 and 3. The other iterations reduce the search interval, converging to the values found and shown in Table 3.1. For the 1DPX and 1DPW conformation states, the behavior of iterations and convergence are very similar.

The bisection method, despite its simplicity, displays in the eighth iteration a difference of at most 0.2% to the best result found in the proposed cases. The results presented in Fig. 3.1 (b) show in both cases an increase in the Pearson value with increasing C up to an apparent point of global maximum, where the slope and sensitivity to the C parameter are low. The addition of hydrogen bonds (H) resulted in a lower C coefficient and a higher correlation, presenting a similar profile. This behavior also occurred in the other conformation states, indicating

the possibility of using a method with a higher degree of convergence or investigating the existence of a standard behavior in the analysis of primary structures using the proposed optimization method.

(a)



(b)

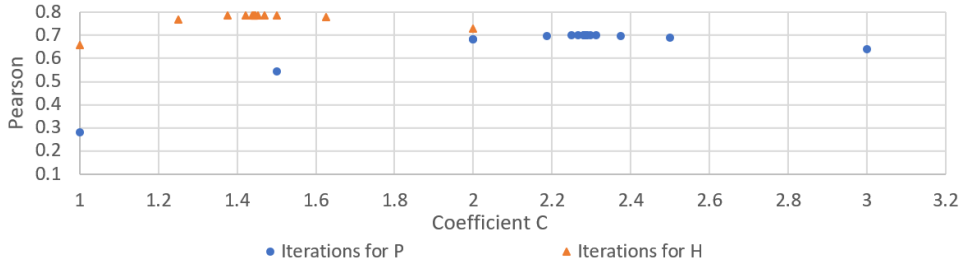


Figure 3.1 (a) Graph of the simulated C coefficient (left axis) and Pearson's coefficient in relation to the experimental one (right axis) for each iteration, using the conformation state 4YM8. (b) Graph of Pearson's coefficient in relation to the experimental for each simulated coefficient, using the 4YM8 configuration.

The C values with the best result found for the Pearson coefficient are shown in Table 3.1. The coefficients found for the different conformation states are similar for optimized primary structures (PO) and optimized hydrogen bonds (HO), with the most significant difference between 4YM8 and 1DPW around 15% considering the primary structure.

Table 3.1 – Optimal values found for coefficient C.

	4YM8	1DPX	1DPW
PO	2.282	1.997	1.979
HO	1.443	1.507	1.416

3.2 Experimental and simulated B-factor for primary structures

Figure 3.2 shows the experimental B-factors for 4YM8, normalized by Eq. 2 (in black). The vertical axis shows the deviations from the mean value, while the horizontal axis displays the atom number. The normalized B-factor for primary links (green) and the best result with optimized primary links (yellow) using Eq. 1 are compared to the experimental results. In the simulation with primary bonds (green), it is possible to observe several local peaks along the main chain, most of which are found in the amino acid side chain, which is connected to the carbon atom of the main chain and free at the other end, thus having greater flexibility. Many peaks coincide with the experimental peaks, even if there is a certain distance in the values caused by the primary chain vibration; however, in the experimental data, they do not always occur or are more accentuated in some atoms, as indicated by the green arrows in Fig. 3.2. We can also see in Fig. 3.2 an increased flexibility at the ends, especially near atom 1 indicated by the blue arrow and near atom 1000 indicated by the purple arrow. This increase does not occur in the experimental data for atom 1, indicating restrictions for this type of simulation. The simulation results with optimized primary links (PO) are indicated in yellow in Figure 3.2. At the extremities, where the model was inaccurate in reproducing flexibility, the result approximates the experimental. In some peaks that already matched experimental data, the peak value approximated the experimental with some exaggerations, such as the peak around atom 578, indicated by the red arrow.

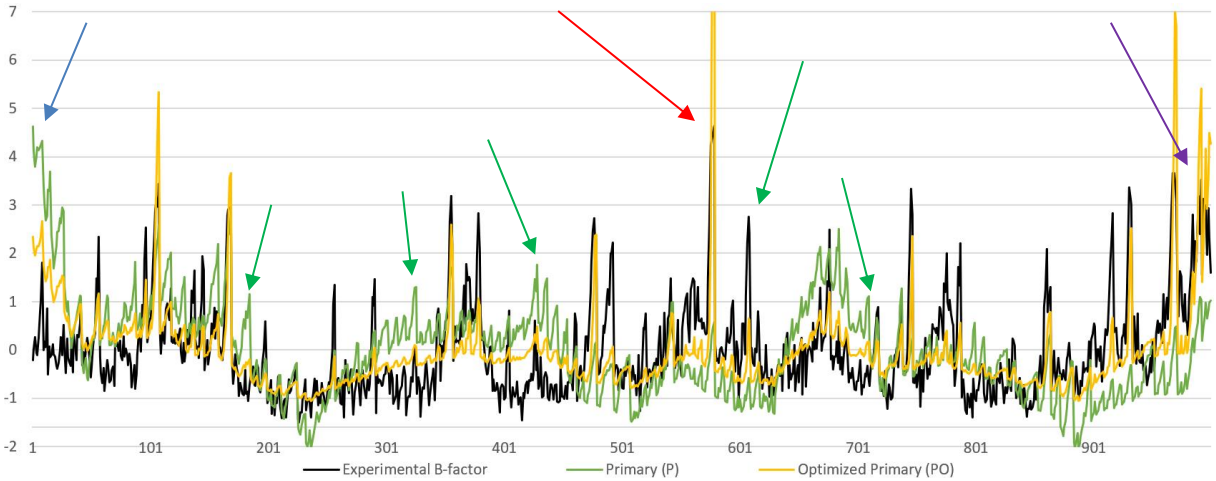


Figure 3.2 – Experimental and simulated B-factor for each atom using primary (P) and optimized primary (PO) bonds for the 4YM8 conformation state.

The optimization of the modulus of elasticity through the experimental B-factor generally attenuates local variations in regions of low flexibility and evidences some peaks in regions of high flexibility. The best result for the Pearson correlation caused at least one exaggerated peak, but it coincides with the highest experimental value.

3.3 Primary structures with added hydrogen bonds

The experimental and calculated B-factors are again obtained and compared with the addition of hydrogen bonds. The results can be seen in Fig. 3.3 for the 4YM8 configuration. Below the graph, the red triangles represent the added hydrogen bonds, and the green ones are related to the weak hydrogen bonds. The atoms that form alpha helices (H1 to H8) and beta-sheets (SA1 to SA3) are indicated by a purple line.

A visible improvement over the experimental result is seen with the addition of hydrogen bonds, especially at the ends and in the region containing SA1, SA2, and SA3, indicated in Fig. 3.3 by green arrows. However, the peaks that coincide with the side chains are accentuated in regions without hydrogen bonds; some coincide while others do not match the experimental values (yellow arrows). Like the model with only primary links, the optimization process approximates the result compared to the experimental results, generally intensifying some peaks that coincide and reducing some others.

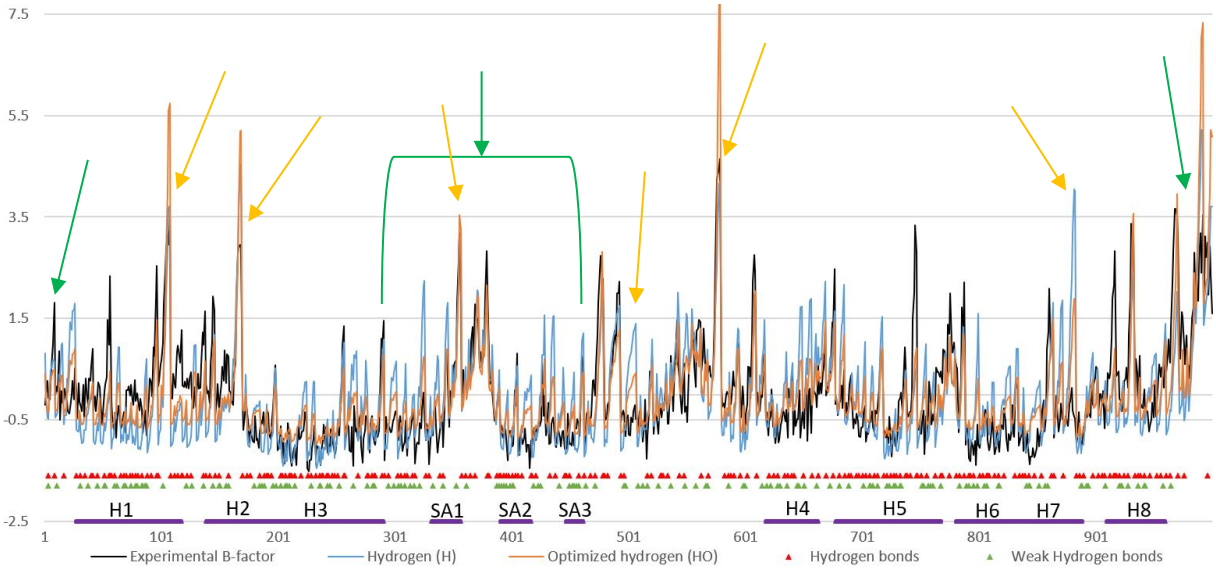


Figure 3.3 – Experimental and simulated B-factors for each atom using hydrogen bonds (H) and optimized (HO) for the 4YM8 configuration.

3.4 Results of the Pearson correlation coefficient between the experimental and numerical results

The B-factor results for the different approaches are compared in Figure 3.4, where the Pearson correlation coefficient between samples is calculated. The color gradient agrees with the result, going from blue, correlation 1, to red, correlation 0. BE solution is the abbreviation of the experimental B-factor.

The results of P versus BE are 0.28 for 4YM8, 0.18 for 1DPX, and 0.1 for 1DPW (bold and black borders Fig.3.4). The better output for 4YM8 and the worse output for 1DPW can indicate that the dynamic behavior of 4YM8 depends more on its primary structure when compared to 1DPW. Comparing the results between the three scenarios (4YM8, 1DPX, and 1DPW) for the P cases, a correlation close to 1 is found (green circles Fig.3.4). It indicates that the simulation of the primary structure of the lysozyme protein is unable to provide information about the different conformational states.

The BE cases comparison of the three conformations states shows a correlation between 0.8 and 0.85 (yellow circles Fig.3.4), indicating the experimental differences between the dynamics of the structures. The same comparison can be applied to the other simulated scenarios, resulting in correlations between 0.81 and 0.95 (purple circles Fig.3.4). The addition of H bonds resulted in slightest changes between conformation states (correlations between 0.91 and 0.95) with consistent improvement in the result versus BE, between 0.6 and 0.66. The positions of the hydrogen bonds obtained by PCA have slight variation between the conformation states, adding little information capable of differentiating these states but resulting in a significant improvement in the protein's dynamics.

The results for PO concerning BE are between 0.66 and 0.7 (bold and black borders Fig.3.4), showing consistent improvement in the results with optimization. Furthermore, the comparison of PO results shows a value similar to the difference found by crossing BE (0.84), indicating that the best result using the BE for optimization produces similar differences compared to the optimized models.

The association between HO and BE is around 0.79 - 0.78 (bold and black borders Fig.3.4), which is the best result obtained. The comparison between the H and HO models shows correlations between 0.9 and 0.85. Thus, the optimization in a model with some correlation (H) generates more minor changes in the structure caused by a smaller optimal value for the coefficient C, but it can still improve the results. This indicates that optimization can also be helpful in more accurate models, using experimental data for minor adjustments.

Despite the promising results for H and HO for the 1DPW state, the results show a better correlation than the BE of other lysozyme proteins states (orange circles Fig.3.4). This may indicate that the differences found in 1DPW compared to the other conformational states cannot capture additional information by considering hydrogen bonds or HO. These observations indicate that even the best result does not reproduce complex phenomena related to conformational states.

		4YM8 BE	4YM8 P	4YM8 PO	4YM8 H	4YM8 HO	1DPX BE	1DPX P	1DPX PO	1DPX H	1DPX HO	1DPW BE	1DPW P	1DPW PO	1DPW H	1DPW HO
4YM8	BE	1	**0.28**	**0.70**	**0.66**	**0.79**	0.80	0.28	0.56	0.66	0.70	0.83	0.28	0.58	0.69	0.74
4YM8	P		1	0.58	0.40	0.33	0.18	1	0.49	0.43	0.30	0.10	1	0.47	0.41	0.25
4YM8	PO			1	0.63	0.82	0.62	0.58	0.84	0.66	0.77	0.56	0.58	0.81	0.66	0.77
4YM8	H				1	0.90	0.63	0.40	0.56	0.91	0.80	0.55	0.40	0.48	0.95	0.75
4YM8	HO					1	0.73	0.33	0.70	0.86	0.91	0.64	0.33	0.59	0.87	0.83
1DPX	BE						1	**0.18**	**0.69**	**0.60**	**0.78**	0.85	0.18	0.64	0.67	0.80
1DPX	P							1	0.49	0.43	0.30	0.11	1	0.47	0.41	0.25
1DPX	PO								1	0.56	0.79	0.53	0.49	0.88	0.63	0.82
1DPX	H									1	0.85	0.54	0.43	0.49	0.92	0.73
1DPX	HO										1	0.61	0.29	0.61	0.81	0.81
1DPW	BE											1	**0.10**	**0.66**	**0.61**	**0.78**
1DPW	P												1	0.47	0.41	0.25
1DPW	PO													1	0.58	0.85
1DPW	H														1	0.85
1DPW	HO															1

Figure 3.4 – Correlation factors obtained by crossing all results.

4 Conclusions

The results for the primary structure model with optimization of the 4YM8 conformation state achieved results close to Giordani et al. [5], who used the CG method for 4YM8. The best proposal considers the addition of hydrogen bonds and alters the bond stiffness using an optimization procedure. The application of only primary structures proved to be useless in differentiating the dynamics between the conformational states. Although the hydrogen bonds have a small fraction of the rigidity of the carbon-carbon bonds, their presence results in significant improvements, indicating a possibility to differentiate the conformational states. Therefore, a more significant number of non-covalent bonds can improve the outputs using this methodology.

The optimization method, although simple, have good convergence. The behavior of Pearson's correlation in relation to the proposed function is similar between the proposed models and indicates a global maximum point, showing the possibility of using other better optimization methods. The increased correlation in the optimized structures indicates an approximation of the global dynamics of the structure, with local peaks and distortions, as well as possible changes in the direction of movement of the atoms in the evaluated vibration modes. Changes to the optimized function that minimize local errors are a possibility for future works.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Alberts, B., Bray D., Hopkin, K. et al. Essential Cell Biology. Third Edition. (2009). ISBN: 978-0-8153-4129-1.
[2] Escobedo, A., Topal, B., Kunze, M.B.A. et al. Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat Commun 10, 2034 (2019). https://doi.org/10.1038/s41467-019-09923-2
[3] Scaramozzino, D., Lacidogna, G., Piana, G. et al. A finite-element-based coarse-grained model for global protein vibration. Meccanica 54, 1927–1940 (2019). https://doi.org/10.1007/s11012-019-01037-9
[4] A. Carpinteri, G. Piana, A. Bassani & G. Lacidogna (2018): Terahertz vibration modes in Na/K-ATPase, Journal of Biomolecular Structure and Dynamics, DOI: 10.1080/07391102.2018.1425638
[5] 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.
[6] Ma, Jianpeng. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes Structure, Vol. 13, 373–380, March, 2005, Elsevier Ltd. DOI 10.1016/j.str.2005.02.002
[7] 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.
[8] Hu X., Buehler X., "Comparative Analysis of Nanomechanical Features of Coronavirus Spike Proteins and Correlation with Lethality and Infection Rate", Matter 4, 265–275 January 6, 2021 ª 2020 Elsevier Inc.
[9] Hajdu, J; Neutze, R; Sjögren, T; Edman, K; Szöke, A; Wilmouth, RC; Wilmot, CM (2000). "Analyzing protein functions in four dimensions". Nature Structural Biology. 7 (11): 1006–12. doi:10.1038/80911
[10] Berman H.M., J. Westbrook, Z. Feng, et al. (2000) The Protein Data Bank Nucleic Acids Research, 28: 235-242.
[11] Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - a visualization system for exploratory research and analysis. J Comput Chem. 2004 Oct.
[12] Kayikci, M. Venkatakrishnan A, Scott-Brown J,. Ravarani C. N. J, Flock T, Babu M.- Protein contacts atlas: visualization and analysis of non-covalent contacts in biomolecules - Nat Struct Mol Biol. 2018 Feb; 25(2): 185–194
[13] Sugahara, M., Nakane, T., Suzuki, M., Nango, E. (2015). Crystal structure of hen egg-white lysozyme
[14] Weiss, M.S., Palm, G.J., Hilgenfeld, R. (2000) Crystallization, structure solution and refinement of hen egg-white lysozyme at pH 8.0 in the presence of MPD Acta Crystallogr D Biol Crystallogr 56: 952-958
[15] Carpinteri A, Lacidogna G , Piana G, Bassani A; Terahertz mechanical vibrations in lysozyme: Raman spectroscopy vs modal analysis http://dx.doi.org/10.1016/j.molstruc.2017.02.099 0022-2860/© 2017 Elsevier
[16] Jeffrey, G. A.; Saenger, W. Hydrogen Bonding in Biological Structures, 1st ed.; Springer, Berlin, Heidelberg, 1991
[17] ANSYS, INC. ANSYS Verification Manual. Ansys Release 9.0. Canonsburg, 2004