Exploring thz protein vibrations by means of modal analysis: All-atom vs coarse-grained model

AIMETA 2019
XXIV Conference
The Italian Association of Theoretical and Applied Mechanics
Rome, Italy, 15-19 September 2019

EXPLORING THZ PROTEIN VIBRATIONS BY MEANS OF MODAL ANALYSIS: ALL-ATOM VS COARSE-GRAINED MODEL

Giuseppe Lacidogna, Domenico Scaramozzino, Gianfranco Piana and Alberto Carpinteri

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

E-mail: giuseppe.lacidogna@polito.it, domenico.scaramozzino@polito.it, gianfranco.piana@polito.it, alberto.carpinteri@polito.it

Keywords: Protein Vibrations, Terahertz Range, Finite-Element Model, HIV-1 Protease.

Abstract. *It is strongly believed among the scientific community that protein dynamics constitutes the fundamental link existing between structure and function. Indeed, by means of modal analysis, it has been shown that low-frequency vibrational modes are strictly related to protein conformational transitions. In the last years, some of the authors made use of all-atom mechanical models, treating proteins like elastic lattice structures, in order to investigate the expansion-contraction vibrational modes. In particular, low-frequency modes were found to involve the whole protein structure and occur in the so-called terahertz (THz) range. Vibrations at THz frequencies, especially around or below 1 THz, were also detected experimentally on lysozyme and Na^+/K^+-ATPase powder samples by means of Raman spectroscopy using modern ultra-low frequency (ULF) filters. In order to focus mostly on low-frequency vibrations occurring in the THz range, different finite-element coarse-grained models, based on the coordinates of C_α atoms, are then proposed in this contribution. In particular, the case of HIV-1 protease subunit is investigated and its THz mechanical vibrations are analyzed.*

1 INTRODUCTION

It is pretty established that protein functionality is affected by the three-dimensional arrangement of the amino acids chain, which in turn seems to be itself correlated to the protein biological aim, thus leading to the belief that evolution played a crucial role in this self-adapting process [1]. Among the scientific community, it is also believed that protein activity is correlated with its intrinsic dynamic behavior [2,3]. For this reason, several researches were carried out focusing on protein dynamic features. For example, numerical calculations were performed by exploiting normal mode analysis, treating proteins like elastic networks, made up of point masses connected by linear springs [4-7]. Despite these models were very simplified, they were extremely effective in describing protein dynamics and assessing the strict relationship existing between low-frequency modes and protein conformational changes (i.e. the changes of the three-dimensional structure of the protein, generally occurring when proteins perform their biological activities) [3,8-10].

Among the different elastic models which can be found in the literature, some mechanical all-atom models were recently developed by some of the authors with an engineering approach, in order to investigate the expansion-contraction modes of proteins [11,12]. Based on the rigidity of covalent bonds, the numerical calculations confirmed that the low-frequency modes occur in the so-called terahertz (THz) range. It is then possible that resonance phenomena occurring at THz frequencies may, in a sense, be involved within conformational transitions (i.e. conformational changes) and, eventually, in the mechanisms of protein activity.

From an experimental perspective, the relationship between THz vibrations and protein biological functionality was investigated by several experimental studies based on spectroscopy techniques, such as Raman and THz time-domain spectroscopy (THz-TDS).

Brown *et al.* [13] applied Raman spectroscopy to investigate the low-frequency spectra of different samples of α-chymotrypsin and always found a pronounced peak at 29 cm^{-1} (~ 0.87 THz), except for the sample that was denatured with SDS (sodium dodecyl sulfate). This allowed to conclude that the THz Raman peak must arise from motions involving the overall structure of the molecule. Low-frequency Raman vibrations were also detected by Painter *et al.* [14] within the spectra of various proteins. However, the experimental detection of low-frequency peaks (< 1 THz) is hindered by the strong Rayleigh signal deriving from elastic scattering. To overcome this problem, Turton *et al.* [15] applied Femtosecond optical Kerr-effect (OKE) spectroscopy for measuring the depolarized Raman spectrum of lysozyme, showing that underdamped low-frequency collective motions may be responsible for directing biochemical reactions. Recently, some of the authors made use of Raman spectroscopy, coupled with ultra-low frequency (ULF) filters, to investigate the THz vibrations of lysozyme [11] and Na^+/K^+-ATPase [16] powder samples, obtaining some pronounced Raman peaks around and below 1 THz.

In addition to Raman spectroscopy, THz-TDS has emerged in the last decades as a powerful experimental tool to investigate protein behavior [17]. Terahertz spectroscopy was used by Castro-Camus and Johnston [18] to analyze the variation of the absorption coefficient in the 0.1 – 2 THz range upon the conformational change of PYP (photoactive yellow protein). Chen *et al.* [19] used THz-TDS to detect the reversible conformational transition of PsbO protein by analyzing the corresponding fluorescence emission spectra. Acbas *et al.* [20] showed the existence of underdamped modes in chicken egg white lysozyme (CEWL) crystals for frequencies > 10 cm^{-1} (> 0.3 THz) by means of the crystal anisotropy terahertz microscopy (CATM) method. CATM was also exploited to investigate possible allosteric functions associated to intramolecular vibrations [21], as well as the changes in vibrational directionality upon protein-ligand binding [22].

Given the biological relevance of low-frequency protein vibrations occurring in the THz range, here we show the numerical calculations, arising from modal analysis, aimed at exploring the lowest THz vibrations in HIV-1 protease, whose activity is pivotal for the life-cycle of HIV. In particular, the THz vibrations of HIV-1 protease subunit were investigated by means of different all-atom and coarse-grained finite-element models. From all the models, we show that frequencies in the low-THz range characterize the lowest motions of the protein; however, different modeling procedures provide different outcomes both in terms of frequency values and mode shapes.

2 METHODOLOGY

As mentioned in the Introduction, the lowest vibrations of HIV-1 protease subunit were analyzed by means of different all-atom and coarse-grained models, using the finite-element commercial code LUSAS [23]. Specifically, in Section 2.1 we recall the all-atom model developed by some of the authors in [11,12] taking into account only short-range interactions (covalent bonds). In Section 2.2, a coarse-grained model, based only on C_α atoms and backbone bonds, is presented which is consistent with the hypotheses of the previous all-atom one. These short-range models are able to provide information only on the expansion-contraction and rigid-block motions. To complete the analysis, in Section 2.3 we make use of a coarse-grained space truss model which considers also the influence of long-range interaction and is able to describe also other kinds of motion.

2.1 All-atom model based on covalent bonds

The all-atom model developed in [11,13] was used here to investigate the THz expansion-contraction vibrations of HIV-1 protease subunit. It was built by considering the coordinates of all the heavy atoms of the protein (no hydrogens), which were available from the Protein Data Bank [24] (pdb code: 1hhp). A cutoff value r_c equal to 2 Å was considered to generate the connections (simulating covalent bonds), which were modelled as 3D thin beams clamped to each other. In Fig. 1 the model geometry is displayed.

As for the mechanical parameters, the entire mass of the protein (10.8 kDa ~ 1.79×10^{-23} kg) was equally divided among all the atoms, whereas the average axial stiffness of the connections was set equal to 200 N/m [11,13]. Note that the numerical scaling reported in [11,13] was also used within the software environment, dealing with such small physical quantities.

Figure 1: HIV-1 protease subunit all-atom model based on covalent bonds (AA).

2.2 Coarse-grained model based on backbone bonds

Consistently with the assumptions of the all-atom model recalled in Section 2.1, a coarse-grained model was developed based on Cα atoms and on the protein backbone. Specifically, the coordinates of Cα atoms were considered and only the connections simulating the protein backbone were generated. The resulting model geometry is shown in Fig. 2. The mass of the protein was equally divided among each Cα atom and the stiffness of the backbone bonds was derived from the features of the backbone rigidity.

Figure 2: HIV-1 protease subunit coarse-grained model based on backbone bonds (CGB).

Specifically, considering for C-C and C-N bonds the stiffness [11] and equilibrium length reported in Tab. 1, the mean values (considering one C-C and two C-N bonds between two consecutive Cα atoms) could be computed, which could be associated to the mean backbone bond.

Covalent bond	Equilibrium length [Å]	Stiffness [N/m]
C-C	1.54	180
C-N	1.47	160
Mean - Backbone	1.49	166.7

Table 1: Parameters of the backbone bonds.

Assuming that the bond stiffness is inversely proportional to the length, one could finally compute the stiffness of the Cα-Cα link k through Eq. (1):

$$k = k_{mean} \frac{L_{mean}}{L}, \tag{1}$$

where k_{mean} and L_{mean} are reported in Tab. 1 and L is computed depending on the actual coordinates of the consecutive Cα atoms.

2.3 Coarse-grained space truss model with long-range interactions

Other coarse-grained models were then considered in order to take into account long-range interactions among Cα atoms, thus allowing for the description of distortional modes in addition to expansion-contraction ones. The model was made up of point masses at Cα atoms and an assembly of elastic connections, modelled as 3D bar elements. The bars simulated the interactions between residues and were connected together by means of spherical hinges. The connections were created using two different cutoff values, namely 10 and 12 Å, thus leading to two different models (Fig. 3). The axial bar stiffness was set *a posteriori* after the comparison with the experimental *B*-factors available from the PDB file ($k_{average,CG10}$ = 2.88 N/m and $k_{average,CG12}$ = 0.28 N/m). Note that this model, developed by a purely engineering approach, is

equivalent to the well-known Anisotropic Network Model (ANM) [25] when the distance-weight for the force constant is equal to 1, as recently shown in [26].

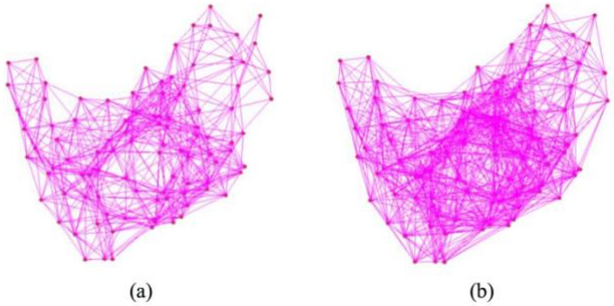

Figure 3: HIV-1 protease subunit coarse-grained space truss models: (a) 10 Å (CG10); (a) 12 Å (CG12).

Summarizing, four different finite-element models were used in this analysis to investigate the lowest THz vibrations of HIV-1 protease subunit, as shown in Tab. 2.

Model	Representation	Features
AA	All-atom	Covalent bonds
CGB	Coarse-grained (C_α atoms)	Backbone bonds
CG10	Coarse-grained (C_α atoms)	Long-range interactions (r_c = 10 Å)
CG12	Coarse-grained (C_α atoms)	Long-range interactions (r_c = 12 Å)

Table 2: Different finite-element models used in the present analysis.

3 RESULTS AND DISCUSSION

In Fig. 4, the lowest vibrational frequencies are reported, which arise from the models considered in Tab. 2. As can be seen, the lowest frequencies lie in the sub-THz frequency range (tens or hundreds of GHz) and their value depend on the specific features of the considered model.

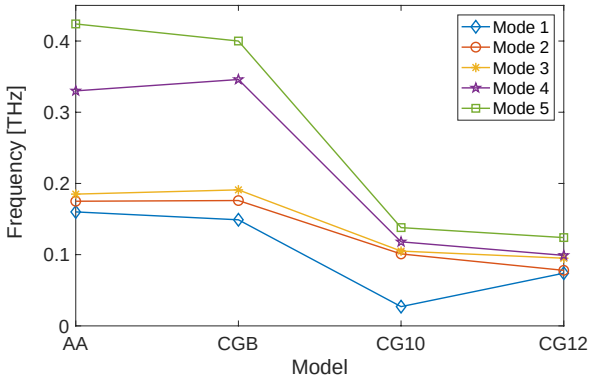

Figure 4: Lowest frequencies obtained from the different models.

In particular, models AA and CGB (which are able to describe only expansion-contraction and rigid-block motions) provide approximately the same frequency values. Moreover, a remarkable agreement was also found regarding the mode shapes arising from AA and CGB. This suggests that local details, e.g. amino acid side chains, are not necessary in order to

describe such kinds of motions, since they only depend on the vibrations of the protein backbone.

Contrariwise, considering models which take into account also long-range interactions, i.e. CG10 and CG12, leads to a different evaluation in terms of the low frequencies, which are found to lie in a lower region of the frequency spectrum. Moreover, since these kinds of models are able to describe also distortional motions, the eigenmodes are not necessarily related to purely expansion-contraction vibrations. For instance, the first natural modes associated to CG10 and CG12 were found to involve a hinge-bending motion at the C-terminus and a distortional movement in the flap region (residues 45-55), respectively. Note that both these motions are biologically relevant since they are related to the observed conformational change occurring in the open-to-close transition of HIV-1 protease, which implicates high flexibility in the flap region and at the C-terminus. The comparison between CG10 and CG12 also leads us to observe that, when considering models with long-range interactions, a small change in the cutoff value (from 10 to 12 Å) can have a not negligible impact on the results, especially in terms of vibrational motions. However, due to the simplicity and low computational cost of such models, they can be a powerful tool for investigating the low-frequency biologically-relevant motions.

4 CONCLUSIONS

In this contribution, we made use of modal analysis to investigate the vibrations of HIV-1 protease subunit in the THz frequency range. In particular, we found that, in order to describe the lowest expansion-contraction and rigid-block vibrations, the modelling of local details, such as amino acid side chains, is not necessary since AA and CGB provided approximately the same results both in terms of frequencies and mode shapes. However, when distortional motions had to be described, also non-local interactions had to be included within the model. These distortional vibrations were found to occur in a lower region of the frequency spectrum (down to tens of GHz) and are related to motions which are relevant for the protein activity. As a matter of fact, the lowest eigenmodes deriving from CG10 and CG12 were found to imply high-flexibility in the C-terminus and flap regions, which are the ones underlying the open-to-close conformational change. Therefore, it is possible that some resonance phenomena in the low-THz range may occur and govern these conformational transitions.

REFERENCES

[1] S. Ovchinnikov, H. Park, N. Varghese, P.S. Huang, G.A. Pavlopoulos, D.E. Kim, H. Kamisetty, N.C. Kyrpides, D. Baker, Protein structure determination using metagenome sequence data. *Science*, **355**, 294-298, 2017.

[2] L.Q. Yang, P. Sang, Y. Tao, Y.X. Fu, K.Q. Zhang, Y.H. Xie, S.Q. Liu, 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, 2013.

[3] S. Mahajan, Y.H. Sanejouand, On the relationship between low-frequency normal modes and large-scale conformational changes of proteins. *Archives of Biochemistry and Biophysics*, **567**, 59-65, 2015.

[4] M.M. Tirion, Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*, **77**, 1905-1908, 1996.

[5] I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, **2**, 173-181, 1997.

[6] K. Hinsen, Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins: Structure, Function, and Genetics*, **33**, 417-429, 1998.

[7] A.R. Atilgan, S.R. Durell, R.L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*, **80**, 505-515, 2001.

[8] F. Tama, Y.H. Sanejouand, Conformational change of proteins arising from normal mode calculations. *Protein Engineering*, **14**, 1-6, 2001.

[9] W. Zheng, B.R. Brooks, Normal-Mode-Based Prediction of Protein Conformational Changes Guided by Distance Constraints. *Biophysical Journal*, **88**, 3109-3117, 2005.

[10] P. Petrone, V.S. Pande, Can Conformational Change Be Described by Only a Few Normal Modes? *Biophysical Journal*, **90**, 1583-1593, 2006.

[11] A. Carpinteri, G. Lacidogna, G. Piana, A. Bassani, Terahertz mechanical vibrations in lysozyme: Raman spectroscopy vs modal analysis. *Journal of Molecular Structure*, **1139**, 222-230, 2017.

[12] A. Carpinteri, G. Piana, A. Bassani, G. Lacidogna, Terahertz vibrational modes in Na/K-ATPase, *Journal of Biomolecular Structure and Dynamics*, **37**, 256-264, 2019.

[13] K.G. Brown, S.C. Erfurth, E.W. Small, W.L. Peticolas, 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, 1972.

[14] P.C. Painter, L.E. Mosher, C. Rhoads, Low-Frequency Modes in the Raman Spectra of Proteins. *Biopolymers*, **21**, 1469-1472, 1982.

[15] D.A. Turton, H.M. Senn, T. Harwood, A.J. Lapthorn, E.M. Ellis, K. Wynne, Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nature Communications*, doi: 10.1038/ncomms4999, 2014.

[16] G. Lacidogna, G. Piana, A. Bassani, A. Carpinteri, Raman spectroscopy of Na/K-ATPase with special focus on low-frequency vibrations. *Vibrational Spectroscopy*, **92**, 298-301, 2017.

[17] L. Xie, Y. Yao, Y. Ying, The Application of Terahertz Spectroscopy to Protein Detection: A Review. *Applied Spectroscopy Reviews*, **49**, 448-461, 2014.

[18] E. Castro-Camus, M.B. Johnston, Conformational changes of photoactive yellow protein monitored by terahertz spectroscopy. *Chemical Physics Letters*, **455**, 289-292, 2008.

[19] H. Chen, G.Y. Chen, S.Q. Li, L. Wang, Reversible Conformational Changes of PsbO Protein Detected by Terahertz Time-Domain Spectroscopy. *Chinese Physics Letters*, **26**, 084204, 2009.

[20] G. Acbas, K.A. Niessen, E.H. Snell, A.G. Markelz, Optical measurements of long- range protein vibrations. *Nature Communications*, **5**, 3076, 2014.

[21] K.A. Niessen, M. Xu, A.G. Markelz, Terahertz optimal measurements of correlated motions with possible allosteric function. *Biophysical Reviews*, doi: 10.1007/s12551-015-0168-4, 2015.

[22] K.A. Niessen, M. Xu, A. Paciaroni, A. Orecchini, E.H. Snell, A.G. Markelz, Moving in the Right Direction: Protein Vibrations Steering Function. *Biophysical Journal*, **112**, 933-942, 2017.

[23] LUSAS, User Reference Manual (Version 15.1), FEA Ltd., Kingston upon Thames, Surrey, UK, 2015.

[24] https://www.rcsb.org

[25] E. Eyal, L.W. Yang, I. Bahar, Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*, **22**, 2619-2627, 2006.

[26] D. Scaramozzino, G. Lacidogna, G. Piana, A. Carpinteri, A finite-element-based coarse-grained model for global protein vibration. *Meccanica*. doi: 10.1007/s11012-019-01037-9.