Terahertz Protein Vibrations: The Usefulness of Coarse-Grained Numerical Models

Terahertz Protein Vibrations: The Usefulness of Coarse-Grained Numerical Models

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

Abstract Understanding the way in which proteins vibrate in their folded state is pivotal for a broad comprehension of their biological activity. In particular, vibrations in the terahertz range are indicated in the current literature as being involved in protein conformational changes. Nowadays, frequencies around or below 1 THz can be detected for example by Raman spectroscopy using proper ultra-low frequency filters. In previous studies, some of the authors performed modal analysis of all-atom lattice models to investigate the expansion-contraction mode shapes associated to low-frequency Raman peaks detected experimentally on lysozyme and Na^+/K^+-ATPase powder samples. In this contribution, all-atom calculations are compared to new ones derived from a simplified coarse-grained mechanical model; the latter was built-up considering only C_α atoms, i.e., the protein backbone. The efficacy of the coarse-grained model in describing delocalized and global expansion-contraction protein vibrations as well as its limitations are discussed.

Keywords Protein vibrations · Modal analysis · Coarse-grained model · Raman spectroscopy · THz range

Introduction

Protein three-dimensional structure is known to be strictly related to biological functionality, which is performed in a dynamic fashion. Therefore, analyzing the way in which proteins vibrate around the native state is of utmost importance, in order to comprehend the complex mechanisms hidden behind protein activity. Vibrational motions can occur at different scales, e.g., involving amino acid side chains, peptide bonds, secondary structures, as well as the whole protein. They are completely defined by the corresponding displacement field and frequency of vibration. Generally, the smaller the scale, the higher the frequency. Among all the vibrational modes, the low-frequency ones have received increasing attention by the scientific community, as they have been indicated to be involved in protein conformational changes [1]. These motions affect large portions of the protein and are believed to play a crucial role in controlling binding activity, which is expressed through the opening and closing of specific clefts.

Protein dynamics can be effectively investigated by means of both numerical and experimental tools. As for the former, molecular dynamics simulations and normal mode analysis [2–5] are generally employed, by using all-atom and coarse-grained models. Each modeling procedure has its own advantages and drawbacks, and leads to different levels of approximation. Anyway, it has been shown that even simplified models, such as coarse-grained ones based on native topology, are able to capture the essential protein behavior [6, 7].

From an experimental viewpoint, protein vibrations can be detected by means of spectroscopy techniques, such as THz-TDS (terahertz time-domain spectroscopy) [8] and Raman spectroscopy. In particular, the latter has proven to be a powerful tool in detecting vibrations around or below 1 THz (~30 cm^{-1}), by using specific ULF (ultra-low frequency) filters. In previous studies, some of the authors made use of ULF-Raman spectroscopy on lysozyme [9] and Na^+/K^+-ATPase [10] powder samples, obtaining some peaks around 0.8 THz. Modal analysis (i.e., linear normal mode calculations) was performed to investigate the expansion-contraction dynamics of all-atom lattice models [9, 11]. The mode shapes associated to the Raman peaks were found to involve the whole lysozyme structure and a large portion of Na^+/K^+-ATPase.

In this contribution, a simplified coarse-grained mechanical model is proposed, which is based only on C_α atoms, aimed at investigating the low-frequency expansion-contraction protein vibrations. The effectiveness of the model is shown by comparing the results with those deriving from previous all-atom simulations, both regarding vibrational frequencies and mode shapes. Its limitations are discussed as well.

G. Lacidogna · D. Scaramozzino (✉) · G. Piana · A. Carpinteri
Politecnico di Torino, Department of Structural, Geotechnical and Building Engineering, Torino, Italy
e-mail: domenico.scaramozzino@polito.it

M. E. Grady (ed.), *Mechanics of Biological Systems and Materials & Micro-and Nanomechanics, Volume 4*,
Conference Proceedings of the Society for Experimental Mechanics Series, https://doi.org/10.1007/978-3-030-30013-5_1

Coarse-Grained Mechanical Model and Modal Analysis

The proposed coarse-grained mechanical model aims at focusing on low-frequency expansion-contraction modes around protein native state. As mentioned in the Introduction, these motions generally involve large portions of the protein, therefore it seems unnecessary to take into account local details such as, for example, amino acid side chains. For this reason, based on PDB (Protein Data Bank) information [12], only C_α atoms positions have been used for the construction of the model.

In order to analyze the dynamic behavior of the protein backbone, a cut-off distance of 4 Å has been set (i.e., C_α atoms are connected only if their distance is lower than the imposed cut-off value), since generally the distance between two consecutive C_α atoms is 3.8 Å. The protein structure turns out to be a lattice model, made up of nodes, corresponding to C_α atoms, and links, corresponding to C_α–C_α bonds. Two main parameters are still needed to completely define the mechanical model: node mass and interatomic bond stiffness. As for the former, the total mass of the protein has been equally divided among C_α atoms; as for the latter, it has been derived from the rigidity of covalent bonds of the protein backbone (–C_α–C–N–C_α–).

As shown in Fig. 1, each C_α–C_α link must in a sense simulate the effect of three covalent bonds. Considering for C–C and C–N bonds the stiffness and equilibrium length reported in Table 1, one can compute the mean values, which are supposed to belong to an "ideal" backbone bond.

Assuming that the stiffness of an "ideal" bond is inversely proportional to the bond length, one can finally compute the stiffness of the C_α–C_α link through Eq. (1):

$$k_{C_\alpha - C_\alpha} = k_{backbone} \frac{L_{backbone}}{L_{C_\alpha - C_\alpha}}, \tag{1}$$

where $k_{backbone}$ and $L_{backbone}$ are reported in Table 1 and $L_{C\alpha-C\alpha}$ is computed depending on the actual positions of the C_α atoms. In accordance with [9, 11], Eq. (1) provides the axial stiffness of interatomic bonds, modeled as elastic bars. This stiffness value is different for different links, based on their length, contrarily from single-parameter approaches adopted in other works [2–4]. Built the coarse-grained FE (Finite Element) model and assigned the mechanical properties, Lusas software [13] is used to perform modal analysis. It basically consists in solving the following equation:

$$\left(\boldsymbol{K} - \omega_n^2 \boldsymbol{M}\right) \cdot \boldsymbol{\delta_n} = \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{K}$ and $\boldsymbol{M}$ are the global stiffness and mass matrices, respectively; ${\omega_n}^2$ represents the n^{th} eigenvalue, which is related to the n^{th} frequency of vibration f_n by the following formula:

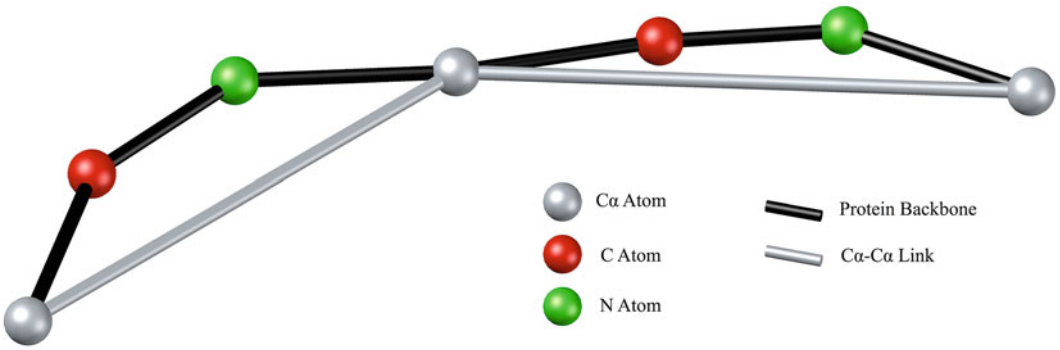

Fig. 1 Scheme for the derivation of C_α–C_α bond stiffness

Table 1 Covalent bond parameters

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

[a]Considering 1 C–C and 2 C–N bonds between two consecutive C_α atoms

$$f_n = \frac{\omega_n}{2\pi}, \tag{3}$$

and $\boldsymbol{\delta_n}$ is the corresponding eigenvector, i.e., the displacement field of the nth vibrational mode. Note that, since protein structure is not externally constrained, the first six vibrational modes are rigid motions with zero-frequency. In the following section, the results deriving from modal analysis are shown, with respect to lysozyme and Na^+/K^+-ATPase.

Results and Discussion

In previous works [9–11] some of the authors performed Raman spectroscopy measurements on lysozyme and Na^+/K^+-ATPase samples, with special ULF filters, and detected strong peaks around 0.8 THz. The obtained spectra, with focus on the low-frequency region, are shown in Fig. 2. By means of modal analysis applied to all-atom models, it was found that in the frequency range near 0.8 THz lysozyme exhibits very collective motions [9], whereas some delocalized vibrations, involving the protein ends, are identified as regards Na^+/K^+-ATPase [11].

Here, we performed modal analysis by means of the mechanical coarse-grained model proposed in the previous section, and compared the results with those arising from all-atom calculations. In Fig. 3, the all-atom and coarse-grained lattice structures are reported for both lysozyme (pdb code: 4ym8) and Na^+/K^+-ATPase (2zxe). It is noteworthy that, by using the coarse-grained model, there was approximately an eight-fold reduction in the number of nodes with respect to the all-atom representation, thus leading to a remarkable saving in terms of computational cost: the reduction lies from 1000 to 129 nodes for lysozyme, and from 10,133 to 1296 for Na^+/K^+-ATPase. Moreover, an easier interpretation of the results in terms of vibration modes was achieved at the same time.

By applying modal analysis on both all-atom and coarse-grained models, we obtained all the vibrational frequencies and the mode shapes. In Fig. 4, the comparison between the results is shown in terms of eigenfrequencies (expressed in THz). As can be seen, in the range 0.1–4.5 THz, the obtained eigenfrequencies almost coincide, i.e., on average both models provided approximately the same frequency value for each eigenmode. Therefore, in order to investigate the vibrational modes lying in a frequency range which can be associated to an experimental Raman peak (Fig. 1), one can use more conveniently the coarse-grained model, if the peak is found up to 4.5 THz (~150 cm^{-1}). Contrariwise, for higher frequencies (above 5 THz) the correlation between the all-atom and coarse-grained results gets lost since local motions dominate, such as vibration of amino acid side chains, which in turn cannot be captured by the coarse-grained model.

It is interesting to observe also the comparison in terms of modal displacements. This can be achieved by means of Modal Assurance Criterion (MAC) [14], which is defined according to the following equation:

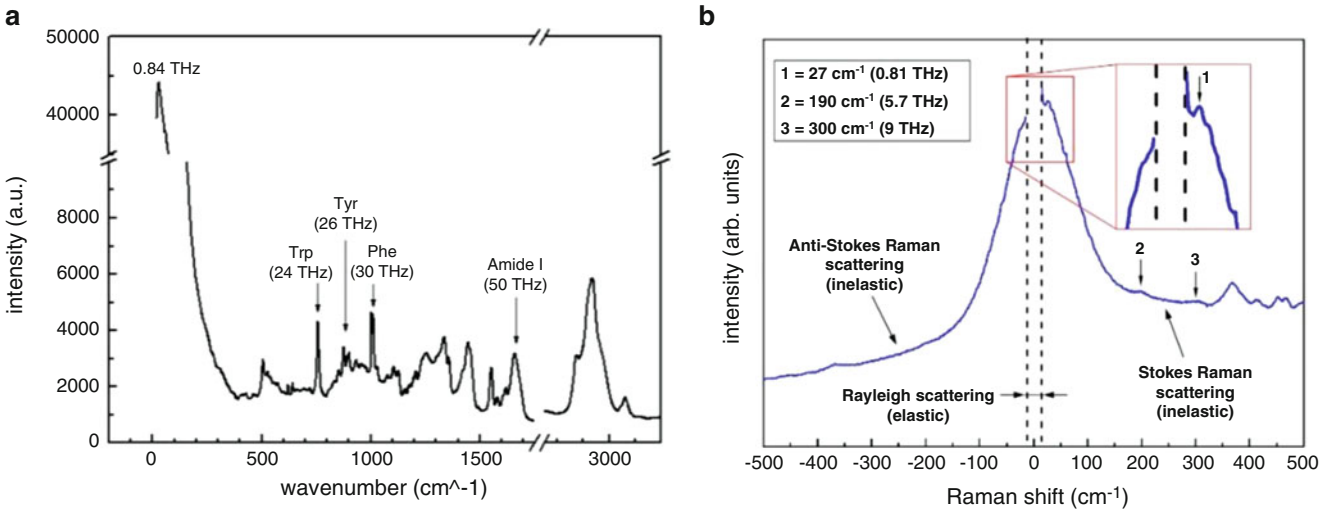

Fig. 2 Raman spectra of (**a**) lysozyme [9] and (**b**) Na^+/K^+-ATPase [11]

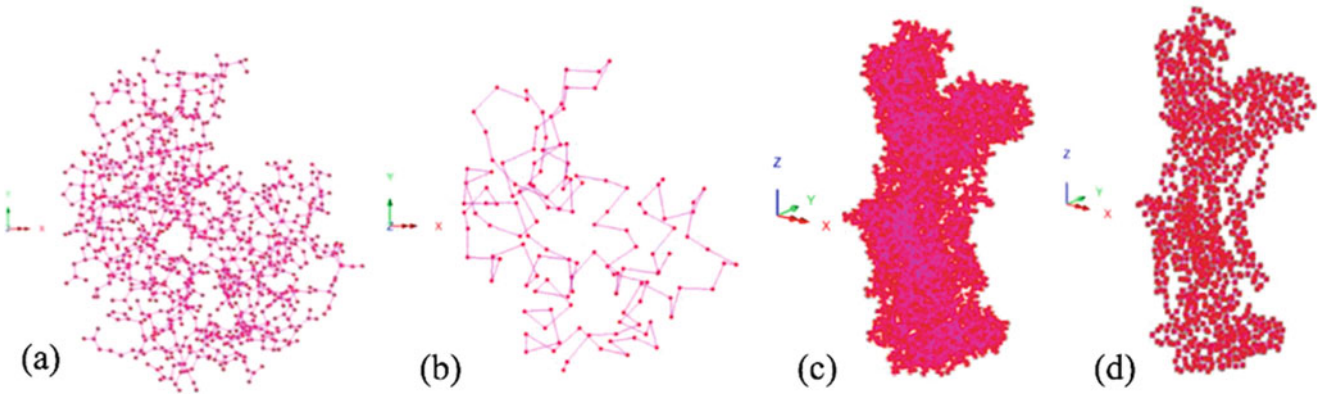

Fig. 3 All-atom model of (**a**) lysozyme and (**c**) Na^+/K^+-ATPase; coarse-grained model of (**b**) lysozyme and (**d**) Na^+/K^+-ATPase

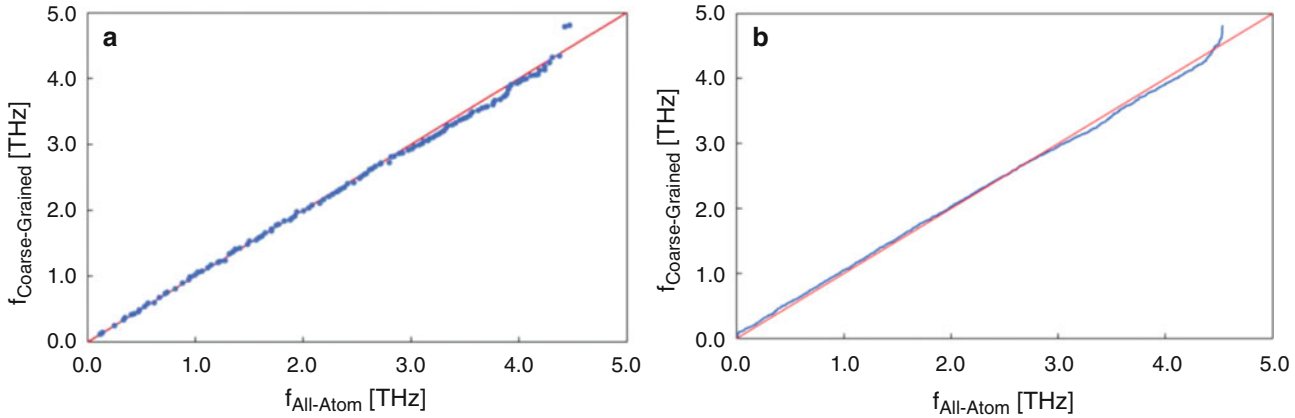

Fig. 4 Frequency comparison between coarse-grained and all-atom model for (**a**) lysozyme and (**b**) Na^+/K^+-ATPase

$$MAC_{i,j} = \frac{\left(\boldsymbol{\delta}_{i,CG}{}^T \boldsymbol{\delta}_{j,AA}\right)^2}{\left(\boldsymbol{\delta}_{i,CG}{}^T \boldsymbol{\delta}_{i,CG}\right)\left(\boldsymbol{\delta}_{j,AA}{}^T \boldsymbol{\delta}_{j,AA}\right)}, \tag{4}$$

where $\boldsymbol{\delta}_{i,CG}$ and $\boldsymbol{\delta}_{j,AA}$ represent the i^{th} and j^{th} modal shape obtained by coarse-grained and all-atom calculations, respectively. Obviously, as far as $\boldsymbol{\delta}_{j,AA}$ is concerned, only the displacements of C_α atoms are considered within the vector. For example, regarding the seventh vibrational modes of lysozyme ($f_{7,CG} = 0.114$ THz, $f_{7,AA} = 0.118$ THz), one obtains $MAC_{7,7} = 0.985$, which is also evident from Fig. 5, which shows the normalized displacements for each residue, i.e. C_α atom.

In Fig. 6, the MAC matrix deriving from the calculation based on Eq. (4) is shown for the first 30 pairs of modes (not including the six rigid motions). Each grid of the matrix, of coordinates *i* and *j*, shows the MAC value (in color scale) between the i^{th} coarse-grained and the j^{th} all-atom eigenmode. As can be seen, high MAC values are obtained within the matrix diagonal, especially for the low-frequency eigenmodes, where values higher than 0.9 are generally obtained. Otherwise, for higher mode numbers, MAC values decrease, as well as they spread around the diagonal, since some interaction appears between different modes. As far as Na^+/K^+-ATPase is concerned, the results deriving from the MAC analysis based on Eq. (4) cannot be easily interpreted, since the transmembrane pump is made up of three different amino acid chains. Therefore, although the range of the obtained all-atom and coarse-grained eigenfrequencies is almost the same (Fig. 4b), the exact correspondence between the eigenmodes cannot be achieved, since different chains can vibrate alternatively at similar frequencies.

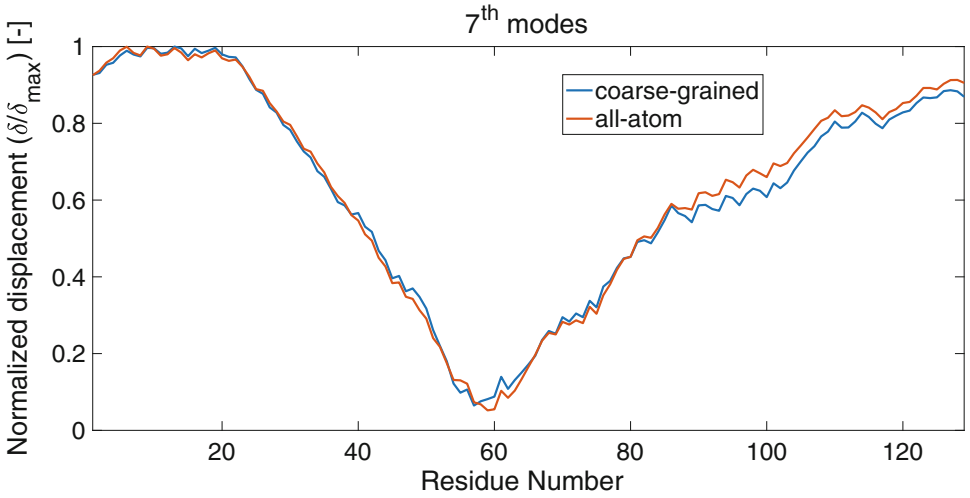

Fig. 5 Modal shapes comparison for lysozyme: seventh modes ($MAC_{7,7} = 0.985$)

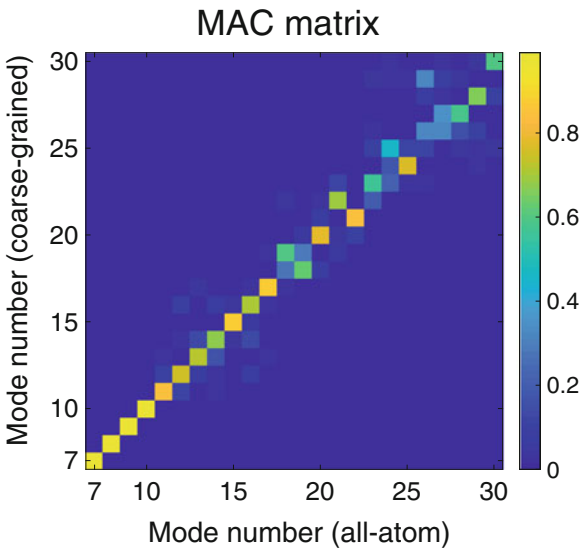

Fig. 6 MAC matrix for the first 30 pairs of modes of lysozyme

Conclusions

In this contribution, we presented a simplified coarse-grained model, based only on C_α atoms, aimed at focusing on focusing on low-frequency expansion-contraction protein dynamics. This model was developed to analyze the vibrations of protein backbone, neglecting long-range interactions. Lysozyme and Na^+/K^+-ATPase were modeled in LUSAS finite element code: the comparison with all-atom calculations confirmed that, when global or delocalized motions occur, frequencies in the terahertz range are involved and the coarse-grained model is sufficient to capture the essential expansion-contraction dynamics. In particular, the results arising from all-atom and coarse-grained calculations show an impressive agreement both in terms of eigenfrequencies and, at least for the low-frequency modes, modal shapes as far as single-chain proteins are involved, like in the case of lysozyme. However, when it comes to more complex structures, made up of several amino acid chains, such as Na^+/K^+-ATPase, the two models provide approximately the same values in terms of eigenfrequencies, but the exact correspondence between the mode shapes cannot be found.

References

1. S. Mahajan, Y.H. Sanejouand, On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. Arch. Biochem. Biophys. **567**, 59–65 (2015)
2. M.M. Tirion, Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys. Rev. Lett. **77**, 1905–1908 (1996)
3. I. Bahar, A.R. Atilgan, B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold. Des. **2**, 173–181 (1997)
4. 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. Biophys. J. **80**, 505–515 (2001)
5. A. Nicolai, P. Delarue, P. Senet, Low-frequency, functional, modes of proteins all-atom and coarse-grained normal mode analysis, in *Computational methods to study the structure and dynamics of biomolecules and biomolecular processes. Springer series in bio-/neuroinformatics*, ed. by A. Liwo, (Springer, Berlin, Heidelberg, 2014)
6. V. Tozzini, Coarse-grained models for proteins. Curr. Opin. Struct. Biol. **15**, 144–150 (2005)
7. I. Bahar, A.J. Rader, Coarse-grained normal mode analysis in structural biology. Curr. Opin. Struct. Biol. **15**, 586–592 (2005)
8. G. Acbas, K.A. Niessen, E.H. Snell, A.G. Markelz, Optical measurements of long-range protein vibrations. Nat. Commun. **5**, 3076 (2014)
9. A. Carpinteri, G. Lacidogna, G. Piana, A. Bassani, Terahertz mechanical vibrations in lysozyme: Raman spectroscopy vs modal analysis. J. Mol. Struct. **1139**, 222–230 (2017)
10. G. Lacidogna, G. Piana, A. Bassani, A. Carpinteri, Raman spectroscopy of Na/K-ATPase with special focus on low-frequency vibrations. Vib. Spectrosc. **92**, 298–301 (2017)
11. A. Carpinteri, G. Piana, A. Bassani, G. Lacidogna, Terahertz vibrational modes in Na/K-ATPase. J. Biomol. Struct. Dyn. **37**, 256–264 (2019)
12. Protein Data Bank. https://www.rcsb.org
13. http://www.lusas.com
14. R.J. Allemang, D.L. Brown, A correlation coefficient for modal vector analysis. *Proceedings of the 1st IMAC*, Orlando, FL, USA, 1982