First-principles calculations of Raman vibrational modes in the fingerprint region for connective tissue

E. T. Sato and H. Martinho*
Centro de Ciências Naturais e Humanas,
Universidade Federal do ABC, Av. dos Estados 5001,
Santo André-SP, 09210-580, Brazil

Vibrational spectroscopy has been widely employed to unravel physical-chemical properties of biological systems. Due to its high sensitivity to monitor real time "in situ" changes, Raman spectroscopy has been successfully employed, e.g., in biomedicine, metabolomics, and biomedical engineering. The grounds of interpretation of Raman spectra in these cases is the isolated macromolecules constituent vibrational assignment. Due to this, probe the anharmonic interactions or the mutual interactions among specific moieties/side chains to name but a few is a challenge. We present a complete vibrational modes calculation for connective tissue in the fingerprint region (800 – 1800 cm\(^{-1}\)) using first-principles Density Functional Theory. Our results indicated that important spectral features correlated to molecular characteristics have been ignored within the usual tissue spectral bands assignments. In particular, we found that the presence of confined water is the main responsible for the observed spectral complexity. Our calculations accounted for the inherent complexity of the spectral features in this region and useful spectral markers for biological processes were unambiguously identified.

The rapid, noninvasive and high spatial resolution capabilities of Raman spectroscopy technique have been employed to obtain biochemical and structural pieces of information of biological samples. Biomedicine[1], metabolomics [2], biomedical engineering[3, 4] are examples of fields where this tool have been successfully used to acquire high-quality data. In particular, several optical-biopsy studies have shown that molecular interaction features in cells and tissues which cannot be accessed by conventional histopathology can be probed by this technique[5]. Raman is of special interest due to their high sensitivity to detect biochemical and molecular variations in tissues[5].

To a first approximation, spectrum of biological tissue is a convolution of isolated biological macromolecules (e.g., carbohydrates, proteins, lipids, deoxyribonucleic acid, ribonucleic acid) spectra. Hence, the tissue vibrational bands assignment is usually made based on isolated macromolecules assignment. There are several literature compilations (see, e.g., refs. [6–9] ) listing the vibrational bands of biological tissues constituents. These compilations eventually are used to perform a qualitative interpretation of the spectra. However, a large amount of relevant pieces of information are absent to one using this qualitative vibrational assignment. Anharmonic interactions which gives rise to coupling among harmonic vibrational modes is an example[10, 11]. Moreover, mutual interactions among specific moieties could be analyzed only in comparative basis. The usage of computational simulations of small specific parts and short time intervals of the macromolecule is another method to understand this interactions. These approaches usually obscure relevant physical-chemical data from the environment. At scale of real biology they are in fact only a small part of the overall picture[12].

Computer simulations could be a suitable tool to interpret experimental data aiming understand biochemical changes translating structural changes that lead to macroscopic biological processes. Vibrational spectra of macromolecules and tissues are a important class of experimental data addressing this issue. The atomistic models based on quantum mechanical calculations have better prediction of the materials properties. However, due to its inherent complexity, atomistic modeling of biological systems are still in the early stages. In a previous[13] we presented a computational model for skin (STmod). The model consisted of a collagen peptide cutout including confined water submitted to periodic boundary conditions. The model was able to successfully explain important experimental structural and general biochemical trends of normal and inflammatory tissues.

In the present work a detailed vibrational modes assignment of a connective tissue based on the STmod is presented. To the best of our knowledge this is the first report on literature concerning complete vibrational assignment of a tissue. The vibrational calculations were performed on \(C_n\) \((n – 8)\), \(C_{1s}\), \(D_0\), and \(D_1\) unit cells of STmod. The numeric subscript indicates the number of water molecules inside the unit cell. The "s" subscript related to the presence of external water solvating the \(C_1\) model. Starting from a hydrated collagen peptide each unit cel was obtained and calculations performed on periodic boundary conditions. More details concerning the obtainment of these structures and previous characterizations as well could be found in ref. [13]. Figure 1 shown the unit cell for \(C_6\), \(C_{1s}\), \(C_2\), and \(D_0\) structures.

Density Functional Theory (DFT)[14, 15] was used in order to obtain the equilibrium geometries and har-

* herculano.martinho@ufabc.edu.br
FIG. 1. Unit cells for \( C_0, C_1, C_2, \) and \( D_0 \) unit cells. Structural pieces of information could be found on Table I of ref. [13].

monic frequencies. The simulations were implemented in the CPMD program[16] using the BLYP functional[17] augmented with dispersion corrections for the proper description of van der Waals interactions[18, 19]. The cutoff energy was considered up to 100 Ry. The wave functions were optimized and then the vibrational modes were obtained using the Hessian matrix. Finally the linear response for the values of polarization and polar tensors of each atom in the system was calculated to evaluate the eigenvectors of each vibrational mode. The Raman-active modes were obtained from the atomic polar tensors for each atom in the system and the corresponding eigenvectors of Hessian[20]. Harmonic frequencies were compared to experimental Raman data of normal(NM) oral mucosa tissue (see ref. [3] for experimental details). The spectra were simulated as a convolution of Gaussian lineshape peaks centered on the calculated frequencies using the Fityk [21] program. The linewidth was chosen to be 20 cm\(^{-1}\). Figure 2a) presents the results for fingerprint region. The experimental Raman spectrum of NM is also shown.

From direct inspection, we found that \( C_2 \) and \( C_3 \) models presented a large set of represented bands (thirteen). The worst model was \( C_7 \) one, which has only four peaks according to the experimental results, followed by \( C_0 \), consistent with only five peaks. The \( C_3 \) model presented a set of negative high frequency (\( \lesssim 940 \) cm\(^{-1}\)) modes indicating some degree of mechanical instability. Thus, we concluded that \( C_2 \) is the suitable model to represent the connective tissue from the vibrational modes point of view.

Connective tissue vibrational modes based on \( C_2 \) model are shown in Table I. The comparison between the present work and literature bands assignment are also shown. We notice that our results presented extra pieces of information. The most striking feature is the activation of methyl, methylene, and amine side chains vibrations along the fingerprint spectral window. Figure 2b)-2e) shows the percentage contribution to the total eigenvectors projected on the wavelength axis for some of these vibrations. Amine vibrations (Fig. 2 b) are present on almost the entire region weighting around 10% of contribution. It is possible to observe that methyl groups contributions (Fig. 2c) are \( \gtrsim 10\% \) below 1300 cm\(^{-1}\).
TABLE I. Connective tissue Raman bands assignment based in C2 model. Additional vibrations compared to usual assignment in the fingerprint region are shown in red. It is important to notice that torsion and stretching at least one of the rings, C-C, CH, CN, NH, C=O, CH3, OH stretching, CH2 symmetrical and asymmetrical stretching, Amide III, and H2O scissoring vibrations appeared in all bands in some degree.

| Experimental (cm⁻¹) | Assignment |
|---------------------|-------------|
| 800 – 880           | CH3 symmetrical stretching; C-C stretching; and C-C-H deformation usually of proline, hydroxyproline, and collagen backbone |
| 880 – 980           | CH3 and confined H2O symmetrical stretching; C=O wagging; CH2 twisting; CH2, CH3 and confined H2O rocking; C-C stretching; and C-C wagging usually of proline, hydroxyproline, and collagen backbone |
| 980 – 1,050         | CH3 symmetrical and asymmetrical stretching; CH and CH2 twisting; CH2 and CH3 rocking; CH3 scissoring, C-C aromatic ring stretching usually of proline and collagen |
| 1,050 – 1,150       | CH3 and NH2 symmetrical and asymmetrical stretching; CH2 wagging; CH2 twisting; CH3 rocking; C-C stretching; C-N stretching vibration; usually of collagen, several bands of moderate intensity, belonging to Amide III and other groups (proteins) |
| 1,150 – 1,230       | CH3 and NH2 symmetrical and asymmetrical stretching; CH wagging; CH2 twisting; CH2, CH3 rocking; C-C aromatic; CH2 wagging vibrations from proline side chains; C-C/C-N stretching (proteins); several bands of moderate intensity, belonging to Amide III and other groups (proteins), differences in collagen content, hydroxyproline, proteins, including collagen I |
| 1,230 – 1,280       | CH3 asymmetrical stretching; CH wagging; CH2 twisting; CH2 wagging; C-N stretching; C-N in plane stretching; CH2 rocking; second Amide; several bands of moderate intensity, belonging to Amide III and other groups (proteins); differences in collagen content, protein band |
| 1,280 – 1,320       | CH3 asymmetrical stretching; CH wagging; CH2 and CH3 rocking, CH2; CH3 twisting; CH2 bending; CH2 wagging; CH2 deformation; several bands of moderate intensity, belonging to Amide III and other groups (proteins); differences in collagen content, bending mode of collagen |
| 1,320 – 1,400       | NH3 symmetrical and asymmetrical stretching; NH2 rocking, CH and CH2 deformation; CH2 bending; CH3; CH2 twisting; CH rocking; C=O symmetric stretch; CH2 wagging in collagen; CH3 and CH2 deforming modes of collagen, several bands of moderate intensity, belonging to Amide III and other groups (proteins) |
| 1,400 – 1,480       | CH3 symmetrical stretching; NH2 symmetrical and asymmetrical stretching; CH2 twisting; CH2, CH3 and NH2 rocking; CH2 stretching/CH3 asymmetric deformation; CH2 wagging; CH, CH2, CH3 deformation; CH3 scissoring; C-H and CH2 bending mode of proteins, collagen |
| 1,480 – 1,540       | NH3 asymmetrical stretching; NH2 rocking; C-N stretching vibration C=C stretching in benzenoid ring, Amide II, collagen |
| 1,540 – 1,620       | CH3 and confined H2O symmetrical stretching; CH3 and NH2 asymmetrical stretching; CH, CH2 and OH wagging; CH2 and NH2 twisting; CH2 and NH2 rocking; C-C stretching; Amide I; Amide II |
| 1,620 – 1,700       | CH3, NH2 and H2O symmetrical stretching; CH3 and NH2 asymmetrical stretching; CH and CH2 wagging; CH2 and NH2 twisting; CH2 and CH3 rocking; CH and CH2 scissoring; C-C stretching band; C=O stretching of collagen, Amide I, differences in collagen content, intermolecular bending mode of water, proteins, including collagen I |
| 1,700 – 1,800       | CH3, NH2 and H2O symmetrical stretching; CH3 asymmetrical stretching; Amide II; CH2 and OH wagging; CH2 and NH2 twisting; CH2, CH3 and NH2 rocking; CH2 scissoring; Amide I; C=O stretching vibrations |

and around 1450, 1550, and 1700 cm⁻¹. The side chains are key factors determining the properties and reactivity of molecules. Thus, one expect that molecular transformations under, e.g., pathological processes, will display overall changes in the fingerprint vibrational region. It is important to notice that our calculations indicated that the side chains vibrational activation occurs only in the presence of confined water. The anhydrous C2 model did not display this characteristic. The water dimer itself presents spectral features around 900 and 1600 – 1750 cm⁻¹ (Fig. 2 d) which usually are not described. Amide II and III vibrations are also present on the overall spectra weighting around 5% (Fig. 2e).

In fact, experimentally observed spectral changes in fingerprint region for tissues have been qualitatively reported to correlated to water content (see, e.g., refs. [22–25]). Elderly and diabetis[22, 23], oral cancer[24], cervical cancer[25] to name but a few are examples of physiological situations where spectral complexity emerges beyond the isolated molecule vibrational bands assignment. Protein-water interactions are known to play a critical role in the function of several biological systems and macromolecules including collagen in tissues[26]. Small changes in structure and dynamical behavior of water molecules at the peptidewater interface can effectively change both the structure and dynamics of the protein function[27]. Our model indicates that the main source of this complexity is the presence of confined water enabling distant and isolated side chains coupling. A large set of wagging, scissoring, twisting and rocking vibrations
of side chains usually assigned in the high-wavenumber region ($\gtrsim 2000$ cm$^{-1}$) appeared in the fingerprint region damped to usually assigned vibrations (see Table I). Coexistence of symmetrical and asymmetric stretchings are also present in a more complicated fashion that go beyond the protein, collagen, proline and hydroxyproline usual bands assignment. Interestingly, the region $880 - 940$ cm$^{-1}$ appeared to retain information about the confined water content (Fig. 2 e). In fact, the contributions to the eigenvectors in this region goes from 30% confined water and 70% of CH$_3$,CH,C-C,C-C-H vibrations.

From the qualitative and quantitative pieces of information generated by our first principle calculations one can found useful spectral markers which could be correlated to biological process of interest. We will comment two examples.

Confined Water. Since the 800 – 880 region is dominated by C,H vibrations (Table I) the difference between the integrated areas of these two regions will be a suitable qualitative quantifier for the confined water content,

$$[H_2O]_{confined} \propto I_{880-940} - I_{800-880}$$ (1)

a key parameter for describing important processes, as just commented in the previous discussion.

Methylation. The 800 – 880 cm$^{-1}$ region could also be used to probe the protein methylation process. Protein methylation is the process through which methyl groups are added to proteins under the action of specific enzymes, the methyltransferases. Usually it occurs on nitrogen atoms in N-terminals and cannot be reversed creating new amino acid residues.[28]. The above-cited spectral region does not present amine nor hydroxil contributions being exclusive of methyl and protein backbone vibrations. Thus computing their ratio to N-H stretching appearing in the high-wavenumber region [3],

$$\text{Protein}_{CH_3} = \frac{I_{800-880}}{I_{N-H}}$$ (2)

will give a useful protein methyl quantifier. Usually the methyl band in the $\sim 2940$ cm$^{-1}$ high-wavenumber region is considered to evaluate the methylation. However, it includes contributions from lipids and proteins[3]. Protein methylation modulates cellular and biological processes including transcription, RNA processing, protein interactions and protein dynamics[29]. Methyl-binding protein domains and improved antibodies with broad specificity for methylated proteins are being used to characterize the so-called $\text{protein methylene}$. They also have the potential to be used in high-throughput assays for inhibitor screens and drug development[30].

In summary, our vibrational modes calculation for connective tissue in the fingerprint region indicated that important spectral features correlated to molecular characteristics have been ignored within the usual tissue spectral bands assignments. Our results indicated that the presence of confined water is the the main responsible for the observed spectral complexity being a factor that cannot be ignored. The inherent complexity of the spectral features in this region could be rationalized by our calculations and useful spectral markers for biological processes could be identified.

Acknowledgements The authors would like to thank the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - 311146/2015-5) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - 2011/19924-2) for the financial support. We would also thank the computational resources provided by Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-UNICAMP) and Sistema de Computação Petaflopica (Tier 0) (Santos Dumont-LNCC) under Sistema Nacional de Processamento de Alto Desempenho (SINAPAD) of the Ministério da Ciência, Tecnologia e Inovação (MCTI).

[1] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N. J. Fullwood, B. Gardner, P. L. Martin-Hirsch, M. J. Walsh, M. R. McAlinsh, N. Stone, and F. L. Martin, Nat. Protoc. 11, 664 (2016).
[2] E. C. Santos, H. S. Martinho, K. Annes, R. F. Leite, and M. P. Milazzotto, Biomed. Opt. Express 6, 2830 (2015).
[3] L. F. C. S. Carvalho, E. T. Sato, J. D. Almeida, and H. S. da Martinho, Theor. Chem. Acc. 130, 1221 (2011).
[4] L. F. C. S. Carvalho, T. M. Pereira, T. D. Magrini, A. S. R. Cavalcante, H. S. Martinho, and J. D. Almeida, Photodiagnosis Potodyn. Ther. 16, 27 (2016).
[5] H. Martinho, “Cancer – cares, treatments and prevention,” (iConcept Press Ltd., 2013) Chap. Advances in Raman-based optical biopsy.
[6] Z. Movasaghi, S. Rehman, and I. U. Rehman, Appl. Spectrosc. Rev. 42, 493 (2007).
[7] Z. Movasaghi, S. Rehman, and I. U. Rehman, Appl. Spectrosc. Rev. 43, 134 (2008).
[8] J. D. Gelder, K. D. Gussem, P. Vandenabeele, and L. Moens, J. Raman Spectrosc. 38, 1133 (2007).
[9] R. Tuma, J. Raman Spectrosc. 36, 307 (2005).
[10] T. Siebert, B. Guchhait, Y. Liu, R. Costard, and T. Elsasser, J. Phys. Chem. B 119, 9670 (2015).
[11] E. Shikhovtseva and V. Nazarov, Biophys. Chem. 214–215, 47 (2016).
[12] M. G. Saunders and G. A. Voth, Annu. Rev. Biophys. 42, 73 (2013).
[13] E. T. Sato, A. R. Rocha, L. F. C. S. Carvalho, J. D. Almeida, and H. Martinho, Phys. Rev. E 91, 063310 (2015).
[14] P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
[15] W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
[16] “Cpmd, http://www.cpmd.org/,copyright ibm corp
1990-2008, copyright mpi für festkörperforschung stuttgart." (1997-2001).

[17] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B 37, 785 (1988).

[18] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani, Phys. Rev. B 71, 195119 (2005).

[19] I.-C. Lin, M. D. Coutinho-Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli, and U. Rothlisberger, Phys. Rev. B 75, 205131 (2007).

[20] A. R. Leach, “Molecular modelling: Principles and applications,” (Pearson Education, 2001) Chap. Empirical Force Field Models: Molecular Mechanics.

[21] “Fityk: y(x) curve fitting and data analysis,” available on http://www.unipress.waw.pl/fityk/ (2009).

[22] L. Pereira, L. dos Santos, P. Fávero, A. A. Martin, et al., Spectrochim. Acta A Mol. Biomol. Spectrosc. 149, 1009 (2015).

[23] M. B. Lopes, R. Rajasekaran, L. Cançado, A. C. Figueira, and A. A. Martin, Photochem. Photobiol. 93, 613 (2017).

[24] E. Barroso, R. Smits, T. Bakker Schut, I. Ten Hove, J. A. Hardillo, E. B. Wolvius, R. J. Baatenburg de Jong, S. Koljenovic, and G. J. Puppels, Anal. Chem. 87, 2419 (2015).

[25] H. S. Martinho, M. C. B. M. Yassoyama, P. O. Andrade, R. A. Bitar, A. M. E. Santo, E. A. L. Arisawa, and A. A. Martin, J. Biom. Opt. 13, 054029 (2008).

[26] N. N. Fathima, M. Baias, B. Blumich, and T. Ramasami, Int. J. Biol. Macromol. 47, 590 (2010).

[27] T. A. Lima, E. T. Sato, E. T. Martins, P. Homem-de Mello, A. F. Lago, M. D. Coutinho-Neto, F. F. Ferreira, C. Giles, M. O. C. Pires, and H. Martinho, J. Phys.-Condens. Mat. 24, 195104 (2012).

[28] S. Clarke, Curr. Opin. Cell Biol. 5, 977 (1993).

[29] L. Afjehi-Sadat and B. A. Garcia, Curr. Opin. Cell Biol. 17, 12 (2013), omics.

[30] S. Carlson and O. Gozani, J. Mol. Biol. 426, 3350 (2014).