PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2022, **24**, 1183

Raman spectra-based structural classification analysis of quinoidal and derived molecular systems†

Arthur P. Pena, *[a] Renata G. Almeida,[b] João Luiz Campos,[a] Hélio F. Dos Santos,[c] Eufrânio N. da Silva Júnior[b] and Ado Jorio *[a]

Received 17th September 2021,
Accepted 1st December 2021

DOI: 10.1039/d1cp04261k

rsc.li/pccp

This work reports a classification analysis method based on the vibrational Raman spectra of 38 quinones and related structures, spectrally ordering and classifying the compounds. The molecular systems are relevant for chemical and biological processes, with applications in pharmacology, toxicology and medicine. The classification strategy uses a combination of principal component analysis with K-means clustering methods. Both theoretical simulations and experimental data are analysed, thus establishing their spectral characteristics, as related to their chemical structures and properties. The protocol introduced here should be broadly applicable in other molecular and solid state systems.

1 Introduction

Quinones are organic aromatic compounds that can be found in nature or synthesized. In nature, quinones can be found in chemical and biological processes, such as the respiratory chain and photosynthesis.[1–4] Structurally, in the most simple form, quinones show two carbonyl residues, separated by vinyl groups within a ring (Fig. 1, left (a)) or adjacent to each other (Fig. 1 (a), right). Quinone compounds can sustain benzene (benzoquinone), naphthalene (naphthoquinone), anthracene (anthraquinone) ring structures, and similar ones.[5,6] Quinones can also be used as a precursor for the synthesis of several derived molecular systems, such as phenazines. Phenazines are organic, heterocyclic, nitrogenous aromatic compounds, also called dibenzo[*b*,*e*]pyrazine.[7] Fig. 1(b) shows the most basic forms of a phenazine. The phenazines analysed in this work were synthesized from quinones.[8] It is possible to find these quinones and phenazines grouped with many other structures forming more complex molecules, as described in this work.

In the last few decades the study of the electronic[9] and chemical[10] properties of quinones has led to interesting results, especially in their applications in pharmacology, toxicology and medicine[1,11,12] with remarkably known antitumor,[13–15] antimalarial,[16,17] trypanocidal[18–20] and leishmanicidal[21] potential activity. Phenazines have also been widely explored in biology,[7,22] where we can mention Barry *et al.*'s[23] investigations of their potential against tuberculosis and Cezairliyan *et al.*'s[24] identification of phenazines being capable of killing nematodes. Most recently, Jardim *et al.*[8] reported on the synthesis of specific quinone and phenazine compounds for the development of new drugs against tuberculosis.

The vibrational modes of the *p*-benzoquinone molecule were firstly reported by Stammreich and Forneris, followed by the investigation of the polarization dependence of its Raman spectrum.[25] Durnick and Wait[26] published the investigation of the fundamental Raman active vibrations in phenazines using a He–Ne laser, along with some infrared active mode investigation. Stenman and Räsänen[27] investigated the symmetry as well as the Raman active modes of solid state 1,4-naphthoquinone. Delarmelina *et al.*[28] published a complete theoretical and experimental investigation of lapachol, α- and β-lapachone Raman and infrared spectra. In addition, studies using time-resolved resonance Raman spectroscopy,[29] characterization *via* resonance Raman spectroscopy of quinone co-factors in solution,[30,31] in enzymatic catalysis,[32] and surface-enhanced Raman spectroscopy (SERS) investigation[33] can be found in the literature.

In this work we analyse the Raman spectra (both theoretically and experimentally) of **38** quinones and derived structures, some, to our knowledge, never characterized before using Raman spectroscopy. The relevance of comparing both simulated and experimental data in this analysis is that, when it is established that these data properly correlate, one can perform the analysis and predictions according to the information provided by the simulated data, avoiding the influence of experimental artifacts.

[a] *Departamento de Física, ICEX, UFMG, Belo Horizonte, MG, 31270-901, Brazil. E-mail: adojorio@fisica.ufmg.br*
[b] *Departamento de Química, ICEX, UFMG, Belo Horizonte, MG, 31270-901, Brazil*
[c] *Departamento de Química, ICE, UFJF, Juíz de Fora, MG, 36036-900, Brazil*

† Electronic supplementary information (ESI) available: Details of the samples and experimental Raman data. See DOI: 10.1039/d1cp04261k

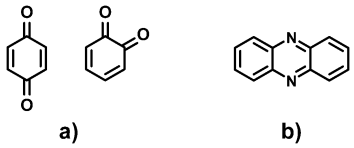

Fig. 1 Most basic forms of (a) quinone [*para*-benzoquinone (left) and *ortho*-benzoquinone (right)] and (b) phenazine chemical structures.

Considering the complex vibrational structure, we make use of Principal Component Analysis (PCA) and K-means clustering.[34–38] These methods have been widely used in the last few decades in materials science, biology and chemistry to improve the extraction of information from data analysis in broader, automatic, faster, and more efficient ways. The complexity of the data we analysed here is due to the number of analyzed compounds (**38**) and the number of vibrational Raman active modes, which goes up to 207 modes for the most complex analysed structure.

Therefore, here we bring an in depth study and the proposal of a classification analysis method using the combination of PCA with K-means clustering statistical learning methods, applied to the vibrational spectra of these **38** quinones and related structures from Raman spectroscopy. The analysis was initially performed for the simulation data, which are free from experimental artefacts, and further compared to related experimental data, showing compatible results. Our contribution is, therefore, twofold: (i) we present new data and analysis related to these relevant organic aromatic compounds, the quinones and phenazines; and (ii) we propose a methodology for Raman spectral analysis that might contribute to big data protocols such as the development of genome initiative of materials.[39]

2 Methodology

2.1 Raman spectra simulation

Molecular geometries were optimized *via* Density Functional Theory (DFT) using the m062x functional and 6-31+G(2d,p) basis-set. The Raman spectra were calculated within the harmonic approximation considering a single molecule for each compound. For the Raman intensities I_i^R:[40–43]

$$I_i^R = C(\nu_0 - \nu_i)^4 \nu_i^{-1} B_i^{-1} S_i, \quad (1)$$

in which ν_0 is the laser excitation frequency, and ν_i and S_i the calculated frequency (in cm^{-1}) and Raman scattering activity (in ref. 4 amu^{-1}) for each normal mode. C is a constant and since we do not do any absolute intensity analysis, it was not used here. B_i[40,41] is a temperature factor that accounts for the contribution from excited vibrational modes and was set as 1. The calculations were performed using Gaussian 09[44] software, and the outputs were visualized in the GaussView[45] software.

The structure optimization and vibrational analysis were carried out in the gas phase. In general, the calculated molecules are rigid; however, for those with a flexible side chain, the conformation was defined by rotating the side chain in order to minimize steric contacts.

2.2 Sample description

A table with the names, chemical formulae, structures and the references or the means by which the **38** compounds (see Table S1) were obtained are shown in the ESI,† followed by pictures of their morphology shown in Fig. S1 (ESI†), as obtained using a microscope utilized for Raman spectroscopy experiments. The compounds were in a solid, microscopic, powder-like state, varying between crystalline and amorphous aspects (in some cases, both aspects could be found in the same sample).

2.3 Raman spectra measurements

Raman spectra were obtained using a WiTec Alpha 300 RA confocal Raman spectroscope with a 633 nm laser as the excitation source. The 633 nm He–Ne laser is linearly polarized, and both the laser-to-microscope and microscope-to-spectrometer coupling are done with optical fibers. The optics (including spectrometer gratings) are polarization dependent, and the system configuration is chosen to maximize the optical efficiency of the system.

The backscattered Raman signal was collected by a 10X/0.25 NA Zeiss EC Epiplan objective lens with an accumulation time of 30 seconds, sent to a back-illuminated Charged-Coupled Device (CCD), located after a 600 g mm^{-1}, BLZ = 500 nm grating. The laser power was adjusted to 4.0 mW as measured at the sample location. In total, a set of **38** compounds were measured (see Fig. S1 in the ESI†), including quinones and derived compounds. Since these molecules have aromatic rings in their structures, it was possible to observe a wide line of luminescence in the spectra of most of the compounds, generating a baseline in the Raman spectrum, which was removed in the data treatment with the Project FIVE 5.0 WiTec software.

2.4 Statistical analysis

We applied the statistical learning methods based on Principal Component Analysis (PCA) combined with K-means clustering using the Scikit–Learn library (version 0.22.1),[46] from Python programming language, in both simulated and experimental Raman spectra of the compounds. For the theoretical simulation, the PCA considered a 38 × 3800 dimensional matrix, where **38** is the number of compounds and 3800 is the number of spectral Raman data, one point per cm^{-1}. For the experimental data, similarly we utilized a 38 × 977 dimensional matrix, where 977 is the number of experimental spectral Raman data one point per 2.1 cm^{-1} on average within the 40–1800 cm^{-1} spectral range. We checked that the different pitches utilized in the theoretical and experimental data do not interfere in the PCA results. PCA offers a way to evaluate the differences between the compounds involved taking as information only the variance in the input Raman data. The K-means clustering application enlightens the investigation about the grouping of similar spectral structures calculating the cluster centroids given in the PCA plot.

When running the PCA with all **38** spectra, we find sample **38** to be too far away from all the others. Therefore, we considered sample **38** to be most different and executed PCA

again excluding sample **38**. The spectral ordering is then calculated considering the three first principal components, weighted by their respective variance, as follows:

$$d = \sqrt{\mathrm{PC1}_{\mathrm{var}}\left(x_{37} - x_j\right)^2 + \mathrm{PC2}_{\mathrm{var}}\left(y_{37} - y_j\right)^2 + \mathrm{PC3}_{\mathrm{var}}\left(z_{37} - z_j\right)^2},$$
$$J = 1, 2, 3, \ldots, 36, \quad (2)$$

where x stands for the PC1 axis, y for the PC2 axis, z for the PC3 axis, and j stands for each of the samples in decreasing order, from 1 to 36, all calculated with respect to the most distant sample in this case, which is sample **37**.

3 Results and discussion

3.1 Spectral ordering

Fig. 2 shows the Raman spectra of the **38** compounds, both (a) the simulated and (b) the experimental data in a heat map (see Fig. S1 (ESI†) for each experimental datum separately). The spectra are ordered, from bottom to top, according to increased spectral complexity, as defined by the PCA ordering. Fig. 3(a) plots the PCA compound ordering of the simulated *versus* the experimental data, showing that the simulated data are a considerably consistent representation of the experimental data, so that analyses and predictions can be made here according to the information provided by the simulated data. The relevance of polarization configuration dependence is shown in Fig. 3(b), where the polarization scattering geometries of samples **1**, **4**, **5** and **24**, which are samples with macroscopic crystalline aspect, were modified (see the caption).

To order the spectra according to spectral complexity, we applied PCA for processing the similarity among the **38** simulated Raman spectra data, and the K-means clustering classification method in order to partition the clusters observed in the PCA. The first three principal components (PC1, PC2 and PC3) accounting for 70.25% of the total spectral variance (PC1: 46.88%; PC2: 15.55%; and PC3: 7.82%) (see the ESI,† Fig. S2), are shown in Fig. 4, each point representing one of the 37 spectra, collected according to the K-means clustering labeling output. Compound (**38**) was not considered in this analysis due to a significantly larger distance from the others, interfering in the understanding of the plot by grouping too closely all the other **37** data points. **38** appears further away along the same PC-direction as compound **37**. We found that **7** clusters (or **8**, including sample (**38**)) to better describe the similarities and differences among the samples.

3.2 Spectral compositions at the first principal component

Fig. 5 shows the compositions of the Raman spectra of some samples in PC1, using the methodology applied by Campos *et al.*[35] Except for compounds (**37**) and (**38**) that stand alone, the composition plot for **38** being the predicted data obtained from the parameterization from the PCA of other **37** spectra. From (a) to (f) three examples are displayed, representing the center and the two extremes of each cluster partition. For each partition, the Raman spectra of the selected samples (above), and their compositions in the PC1 plot (below) are shown. These composition plots give the weights of the Raman modes that mostly contributed to the PC1 variance (and, consequently, the distancing between the points in Fig. 4) individually for each sample.

Between the dashed red lines in each plot are the most characteristic modes of quinoidal compounds (top)[30–32] and the analogue PCA composition regions with the most expressive variations (bottom) between the samples: mostly, the C–H

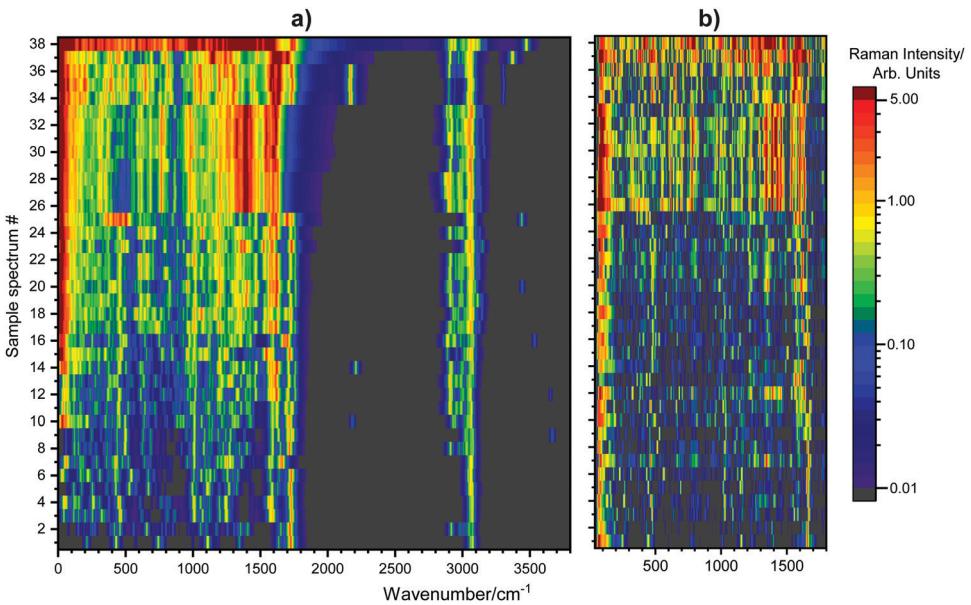

Fig. 2 Heat scale plot for the Raman spectra of the **38** compounds. Each horizontal line corresponds to one Raman spectrum. (a) Simulated Raman spectra in the region between 0 and 3800 cm^{-1}. (b) Experimental Raman spectra in the region between 40 and 1800 cm^{-1}. In (b), the region above 1800 cm^{-1} was removed due to the presence of Etalon fringes.

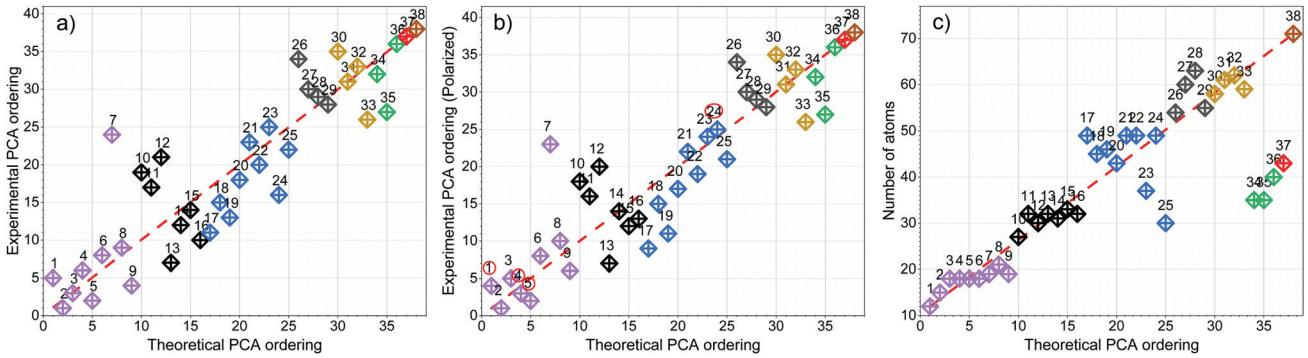

Fig. 3 (a) Plot of the PCA compound ordering of the simulated *versus* experimental data. Compound numbers on top of each data point and cluster colors indicating that the K-means partitioning is based on the simulated data analysis. The red dashed line represents a figurative complete match between theoretical and experimental orderings. (b) Plot of the PCA compound ordering of the simulated *versus* experimental data for 90° rotation of the samples. (c) Plot of the PCA-based theoretical spectral ordering *versus* the respective number of atoms *N*.

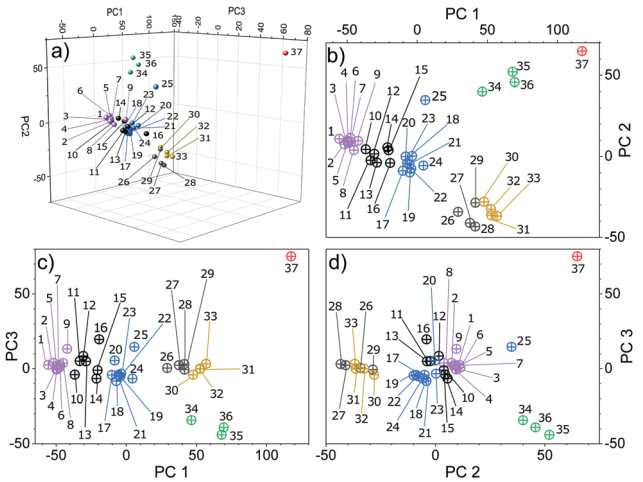

Fig. 4 PCA score plots relative to the theoretical spectra of compounds (**1**) to (**37**). (a) Three-dimensional (3D) scatter plot of the three first Principal Components (PCs) (70.25% of the total variance). The 2D plots are shown in (b), (c) and (d) to give a better notion about the relative distances between the compounds. The distances between points were calculated as a weighted norm (see eqn (2)) relative to the most isolated (in this case **37**).

bendings, C=C and/or C=O stretchings, and associated vibrational modes. The main vibrational modes are labeled in (a) and (c) in Fig. 5. From Fig. 5 it is possible to realize that the general variance of the molecular vibrations within one cluster partition is similar, changing most significantly from one partition to another.

3.3 Ordering and clustering interpretation

Fig. 6 shows the molecular structures for the **38** quinoidal and derived molecular systems, ordered according to the spectra-based PCA. The more complex the chemical structure is, the more complex will be the Raman spectrum (compare Fig. 2 and 6).

One important aspect defining the complexity of the Raman spectra is the number of atoms *N*, which defines the number of vibrational modes as $3N - 6$. This aspect is explored in Fig. 3(c), where we plotted the PCA-based theoretical spectral ordering on the *X* axis, and on the *Y* axis the respective number of atoms *N*. The data points follow roughly the diagonal (dashed line), indicating the relevance of *N* (or the equivalent $3N - 6$) on defining the spectral complexity, as expected. However, the data spread from the dashed line shows that the spectral PCA depends not only on the number of vibrational modes, but also on their specific Raman cross sections and frequencies, which depend on the type of element and their location in the molecular structure. For example, the spectral ordering within the cluster of spectra from 1 to 9, or the clustering of larger molecules, such as the ones related to spectra from **25** to **37**, cannot be explained by *N*.

The first cluster (purple box) is composed of the simplest structures, namely, *p*-benzoquinones and *p*-naphthoquinones, with single atoms or small substitutions (for instance Cl, Br, I, OH, ONa or CH_3) bonded to the main benzo- or naphthoquinone structure. In the second cluster (black box) are found the first *o*-quinones of the whole set of samples ((**12**), (**13**), and (**14**)), and the molecules have substitutions larger than the first cluster, with aromatic ring substituents or a long open chain, like for sample **16**. The third (blue box) cluster shows a set of quinones with longer and more complex patterns of substitutions, being mainly characterized by the presence of sequential aromatic substituents or by the presence of nitrogen atoms in the substitutions. Notice that the samples being "*ortho*-quinone" or "*para*-quinone" do not represent a determinant factor for the ordering/classification considering their vibrational characteristics. The fourth (gray box) and the fifth (yellow box) clusters are characterized by phenazines with more complex substituents. Open chains of aliphatic compounds (alkanes) or aromatic sequences are found. These two clusters are very close to each other in the PCA scores (see Fig. 4). Compound (**29**), for example, which contains a triazole ring and substituted phenyl as all compounds in group **5**, falls into group **4** according to the K-means analysis. From the mathematical point of view, the ordering is dictated by PC1, which has the highest variance (notice the PC1 ordering of samples **28** and **29**, for example). From the physical-chemistry point of

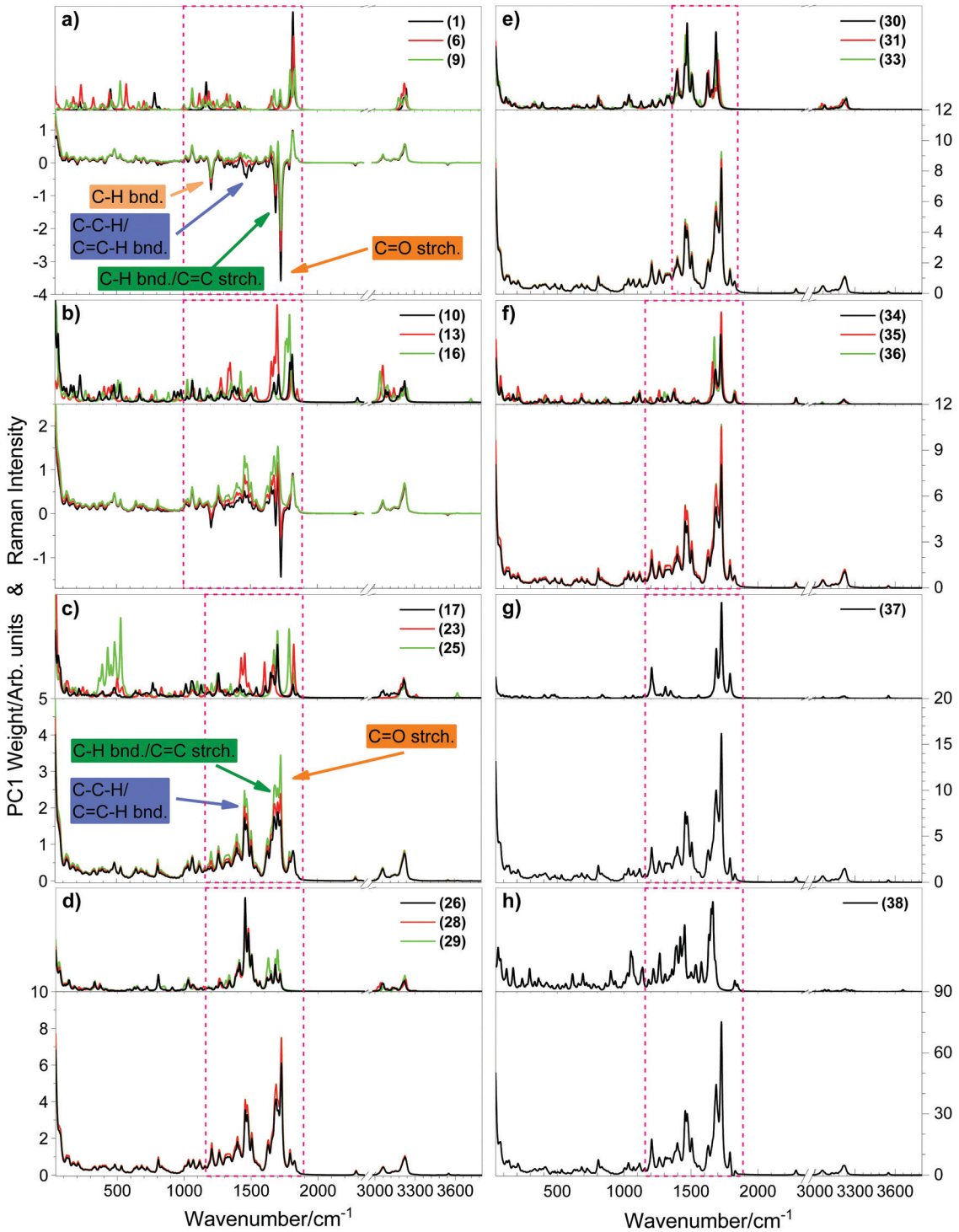

Fig. 5 (a)–(h): Raman spectra (top) and Raman spectra compositions in PC1 (bottom) of selected samples. Each curve stands for one sample, as displayed in the legends. At the bottom plot of (a) and (c) the main vibrational modes with larger variance are indicated (''bnd.'' stands for bending, and ''strch.'' means stretching). Partition (h) (bottom) shows the prediction of the spectral composition to compound (**38**) using the PCA parameterized to the other **37**.

view, the fifth cluster is characterized mainly by the presence of a bromine atom in the aromatic chain substituents and by an aromatic ring bonded in the triazole, and these structural aspects should be responsible for the actually obtained clustering. The sixth (green box) cluster is characterized by the group of alkyne substituents, with the complexity defined by the size of the structure that ends the bond of the aromatic ring, the last one being a carbonyl bonded in the aromatic ring.

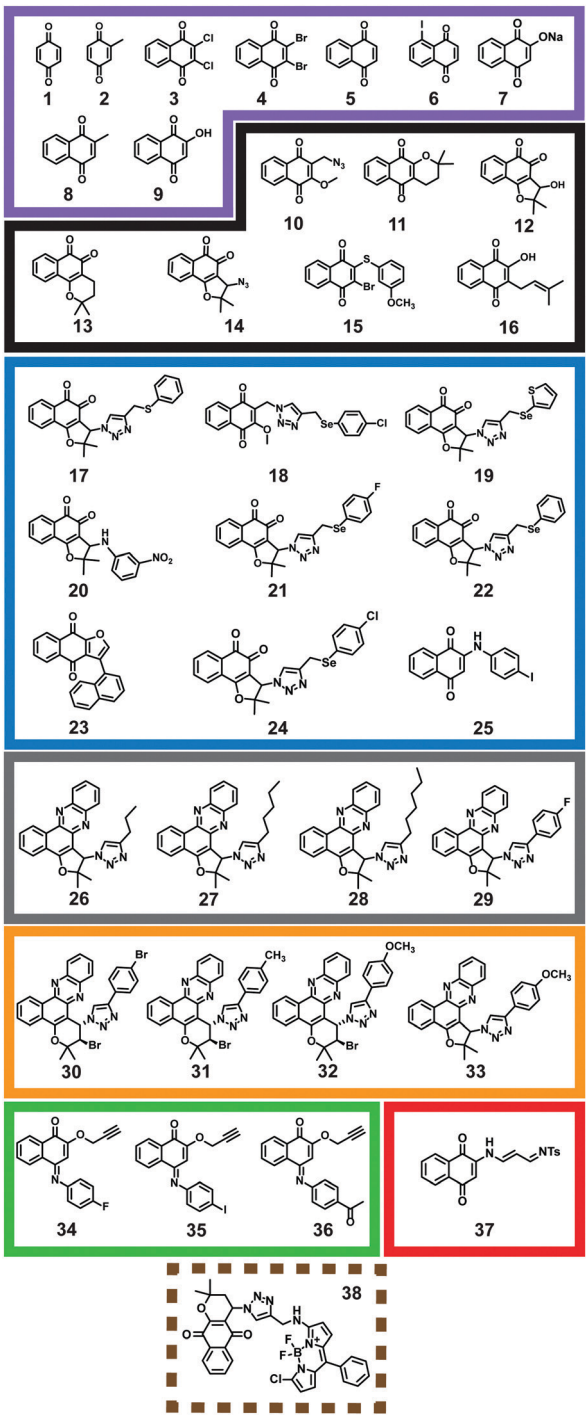

Fig. 6 Schematic organization of how the molecules grouped together in accordance with the PCA relative distances. The grouping boxes follow the same color-code used in Fig. 3. The dashed brown box refers to the compound number (**38**), disregarded in Fig. 4.

Sample **37**, characterized by the tosyl substituent, is by itself the seventh (red box) cluster. The bodipy substituent characterizes the eighth (brown box) cluster, with sample **38** being the most complex structure, with the larger bonded structure, relatively to the other molecules.

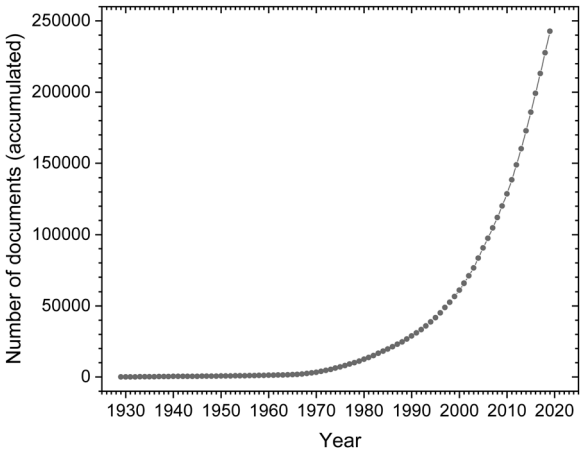

Fig. 7 Accumulative number of Raman papers in the literature. The data are built based on the Scopus database using the following search expressions in the "keyword, title, or abstract" fields (date of search, September 17, 2020): "Raman spectr*" OR "Raman microsc*" OR "Raman scat*".

4 Conclusions

In this study, **38** samples of quinoidal compounds and derived molecular systems were analysed *via* back scattering confocal Raman spectroscopy and simulated *via* DFT and molecular dynamics under harmonic approximation.

Our algorithm was able to compute the ordering of the Raman spectra (and so the structures) based on the variance in the regions related mostly to the C–H bending, and C–C and C═X stretching (X = C, O or N) vibrational modes, with the higher weight relative to the C–H bending and C═X stretching from the quinoidal or phenazinic nuclei structures (C–H and C═X modes) and substituents (C–H modes). The obtained ordering was found to be relative not only to the size of the chemical structure, but also to how the aromatic substitutions are bonded to the main structure. There is exception for the cases of the sixth-cluster (green box) samples (**34**), (**35**) and (**36**), where there was the presence of an alkyne (C≡C), which was not present in any other sample. The analysis of the first principal component (PC1) shows that the spectral distribution in the PC1 weights is similar within the same K-means partition, changing significantly when compared to the spectral composition distribution among clusters.

Therefore, we demonstrate that PCA and K-means clustering Raman-based analysis can be utilized to structurally order and classify molecular systems. Interestingly, we found in the literature information that indicates a link between the cluster divisions and biological/pharmacological aspects of some of the samples, like the antifungal activity of samples (**3**), (**4**) and (**5**)[47] from the purple cluster in Fig. 4 and 6 and HIV-1 inhibition activity of samples (**10**), (**11**) and (**15**)[48] (black cluster in Fig. 4 and 6), indicating that the method utilized here might be a way of grouping and/or selecting similar compounds not only by its physical/spectroscopic characteristics, but also biological/pharmacological applications.

Finally, the method discussed here should not be applicable only to molecules, but also to other amorphous or crystalline solids. In this sense, it is important to stress that with the advance of lasers and detectors, Raman spectroscopy is gaining importance very rapidly (see Fig. 7).[49] Furthermore, the development of theoretical techniques has triggered a new and large amount of theoretical Raman data, within the materials' genome initiative.[39] For example, Taghizadeh *et al.*[50] created the "Computational 2D Materials Database (C2DB)" based on calculated Raman spectra of 733 different two-dimensional systems. In this perspective, the method introduced here might be very helpful for the analysis of greater amounts of vibrational and spectral data in physical chemistry.

Author contributions

A. P. P. performed the experiments, the PCA and K-means clustering and wrote the paper. R. G. A. and E. N. S. J. provided the samples. J. L. C. provided guidance on the PCA. H. F. S. performed the theoretical vibrational simulations. A. J. idealized the project, guided the work and significantly revised the paper. All authors revised and agreed with the present for of the paper.

Conflicts of interest

The authors declare no conflicts of interest for this paper.

Acknowledgements

The authors acknowledge M. D. D. Costa for the data of Fig. 7. This work was supported by CNPq (302775/2018-8) and FAPEMIG (TEC - RED-00282-16).

Notes and references

1 J. L. Bolton, M. A. Trush, T. M. Penning, G. Dryhurst and T. J. Monks, *Chem. Res. Toxicol.*, 2000, **13**, 135–160.
2 O. A. Ghosheh, A. A. Houdi and P. A. Crooks, *J. Pharm. Biomed. Anal.*, 1999, **19**, 757–762.
3 F. A. V. Castro, D. Mariani, A. D. Panek, E. C. A. Eleutherio and M. D. Pereira, *PLoS One*, 2008, **3**, e3999.
4 E. T. Sousa, W. A. Lopes and J. B. D. Andrade, *Qumica Nova*, 2016, **39**, 486–495.
5 L. Dantas-Pereira, E. F. Cunha-Junior, V. V. Andrade-Neto, J. F. Bower, G. A. Jardim, E. N. da Silva Júnior, E. C. Torres-Santos and R. F. Menna-Barreto, *Curr. Pharm. Des.*, 2021, **27**, 1807–1824.
6 N. C. de Lucas, A. B. Ferreira and J. C. Netto-Ferreira, *Revista Virtual de Qumica*, 2015, **7**, 403–463.
7 A. Chaudhary and J. M. Khurana, *Res. Chem. Intermed.*, 2018, **44**, 1045–1083.
8 G. A. Jardim, E. H. Cruz, W. O. Valença, J. M. Resende, B. L. Rodrigues, D. F. Ramos, R. N. Oliveira, P. E. Silva and E. N. D. Silva Jr, *J. Braz. Chem. Soc.*, 2015, **26**, 1013–1027.
9 J. W. Sidman, *J. Am. Chem. Soc.*, 1956, **78**, 2363–2367.
10 T. S. Eckert and T. C. Bruice, *J. Am. Chem. Soc.*, 1983, **105**, 4431–4441.
11 M. N. d. Silva, V. F. Ferreira and M. C. B. Souza, *Qumica Nova*, 2003, **26**, 407–416.
12 A. Rötig, J. Mollet, M. Rio and A. Munnich, *Mitochondrion*, 2007, **7**, S112–S121.
13 M. R. Ribeiro, P. P. Souza, L. D. M. Ferreira, S. L. Pereira, I. da, S. Martins, R. de, A. Epifanio, L. V. Costa-Lotufo, P. C. Jimenez, C. Pessoa and M. O. de Moraes, *et al.*, *Lett. Org. Chem.*, 2011, **8**, 347–351.
14 L. V. Costa-Lotufo, R. C. Montenegro, A. P. N. Alves, S. V. F. Madeira, C. Pessoa, M. E. A. Moraes and M. O. Moraes, *Revista Virtual de Qumica*, 2010, **2**, 47–58.
15 J.-J. Lu, J.-L. Bao, G.-S. Wu, W.-S. Xu, M.-Q. Huang, X.-P. Chen and Y.-T. Wang, *Anti-Cancer Agents Med. Chem.*, 2013, **13**, 456–463.
16 Y.-P. Wan, T. H. Porter and K. Folkers, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 952–956.
17 G. Carr, E. R. Derbyshire, E. Caldera, C. R. Currie and J. Clardy, *J. Nat. Prod.*, 2012, **75**, 1806–1809.
18 A. Pinto, R. Menna-Barreto and S. De Castro, *et al.*, *Phytomedicines*, 2007, 109–127.
19 E. N. Silva Júnior, M. A. B. Moura, A. V. Pinto, M. D. C. F. Pinto, M. C. B. Souza, A. J. Araújo, C. Pessoa, L. V. Costa-Lotufo, R. C. Montenegro and M. O. D. Moraes, *et al.*, *J. Braz. Chem. Soc.*, 2009, **20**, 635–643.
20 E. N. Silva Junior, I. M. Melo, E. B. Diogo, V. A. Costa, J. D. Souza Filho, W. O. Valença, C. A. Camara, R. N. de Oliveira, A. S. de Araujo and F. S. Emery, *et al.*, *Eur. J. Med. Chem.*, 2012, **52**, 304–312.
21 T. T. Guimarães, F. Maria do Carmo, J. S. Lanza, M. N. Melo, L. Rubens, I. M. de Melo, E. B. Diogo, V. F. Ferreira, C. A. Camara and W. O. Valença, *et al.*, *Eur. J. Med. Chem.*, 2013, **63**, 523–530.
22 J. B. Laursen and J. Nielsen, *Chem. Rev.*, 2004, **104**, 1663–1686.
23 V. C. Barry, J. Belton, M. L. Conalty, J. M. Den-steny, D. W. Edward, J. O'sulli-van, D. Twomey and F. Winder, *et al.*, *Nature*, 1957, **179**, 1013–1015.
24 B. Cezairliyan, N. Vinayavekhin, D. Grenfell-Lee, G. J. Yuen, A. Saghatelian and F. M. Ausubel, *PLoS Pathog.*, 2013, **9**, e1003101.
25 H. Stammreich and T. T. Sans, *J. Chem. Phys.*, 1965, **42**, 920–931.
26 T. J. Durnick and S. C. Wait Jr, *J. Mol. Spectrosc.*, 1972, **42**, 211–226.
27 F. Stenman and J. Räsänen, *Spectrochim. Acta, Part A*, 1973, **29**, 405–410.
28 M. Delarmelina, G. B. Ferreira, V. F. Ferreira and J. W. D. M. Carneiro, *Vib. Spectrosc.*, 2016, **86**, 311–323.
29 S. K. Sahoo, S. Umapathy and A. W. Parker, *Appl. Spectrosc.*, 2011, **65**, 1087–1115.
30 S. X. Wang, N. Nakamura, M. Mure, J. P. Klinman and J. Sanders-Loehr, *J. Biol. Chem.*, 1997, **272**, 28841–28844.
31 G. Backes, V. L. Davidson, F. Huitema, J. A. Duine and J. Sanders-Loehr, *Biochemistry*, 1991, **30**, 9201–9210.

32 P. Moenne-Loccoz, N. Nakamura, V. Steinebach, J. A. Duine, M. Mure, J. P. Klinman and J. Sanders-Loehr, *Biochemistry*, 1995, **34**, 7020–7026.
33 M. Umadevi, A. Ramasubbu, P. Vanelle and V. Ramakrishnan, *J. Raman Spectrosc.*, 2003, **34**, 112–120.
34 F. Bonnier and H. Byrne, *Analyst*, 2012, **137**, 322–332.
35 J. L. E. Campos, H. Miranda, C. Rabelo, E. Sandoz-Rosado, S. Pandey, J. Riikonen, A. G. Cano-Marquez and A. Jorio, *J. Raman Spectrosc.*, 2018, **49**, 54–65.
36 Z. Ye, Y. Ye, H. Mohamadian, P. Bhattacharya and K. Kang, Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005., 2005, pp. 90–95.
37 J. D. Holliday, S. L. Rodgers, P. Willett, M.-Y. Chen, M. Mahfouf, K. Lawson and G. Mullier, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 894–902.
38 L. Junlin and F. Hongguang, *Pattern Recogn.*, 2011, **44**, 1721–1737.
39 J. J. de Pablo, B. Jones, C. L. Kovacs, V. Ozolins and A. P. Ramirez, *Curr. Opin. Solid State Mater. Sci.*, 2014, **18**, 99–117.
40 P. L. Polavarapu, *J. Phys. Chem.*, 1990, **94**, 8106–8112.
41 D. Michalska and R. Wysokiński, *Chem. Phys. Lett.*, 2005, **403**, 211–217.
42 R. Wysokiński, K. Hernik, R. Szostak and D. Michalska, *Chem. Phys.*, 2007, **333**, 37–48.
43 H. F. D. Santos, W. B. D. Almeida, A. M. Do Val and A. C. Guimarães, *Qumica Nova*, 1999, **22**, 732–736.
44 M. Frisch, F. Clemente, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino and G. Zhe, *Gaussian 09, revision A. 01*, Gaussian Inc.,2009.
45 T. A. M. J. M. Dennington and R. Keith, *GaussView, Version 6*, 2016.
46 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
47 V. K. Tandon, R. B. Chhor, R. V. Singh, S. Rai and D. B. Yadav, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 1079–1083.
48 T. Ilina, E. Semenova, T. Pronyaeva, A. Pokrovskii, I. Nechepurenko, E. Shults, O. Andreeva, S. Kochetkov and G. Tolstikov, *Dokl. Biochem. Biophys.*, 2002, **382**, 56–59.
49 M. D. D. Costa, *et al.* PhD Thesis, *Universidade Federal de Minas Gerais*, 2021.
50 A. Taghizadeh, U. Leffers, T. G. Pedersen and K. S. Thygesen, *Nat. Commun.*, 2020, **11**, 1–10.

