Fingerprints of polycyclic aromatic hydrocarbons (PAHs) in infrared absorption spectroscopy

Matteo Tommasini [a,*], Andrea Lucotti [a], Michela Alfè [b], Anna Ciajolo [b], Giuseppe Zerbi [a]

[a] Dipartimento di Chimica, Materiali e Ingegneria Chimica – Politecnico di Milano, Piazza Leonardo da Vinci, 32 – 20133 Milano, Italy
[b] Istituto di Ricerche sulla Combustione (IRC) – CNR, P.le Tecchio 80, Napoli, Italy

1. Introduction

Infrared (IR) spectroscopy certainly plays a role long since recognized among the techniques available for the characterization of aromatic compounds [1–3]. For instance (see Supplementary Information) the IR spectra of some derivatives of benzene analyzed in the CH out-of-plane bending region (900–600 cm^{-1} range) do clearly show markers depending on the molecular structures. As expected, the simulation of the IR spectra provided by Density Functional Theory (DFT) calculations (see Supplementary Information) reveals that the strong IR signals in this region are due to collective CH out-of-plane bending vibrations, delocalized over different regions of the molecule and slightly coupled with vibrations of the molecular skeleton. Based on empirical correlations, out-of-plane CH bending vibrations in polycyclic aromatic hydrocarbons (PAHs) with different chemical structures have been assigned [4,5] and put in relation with the topology of CH bonds collectively involved in the normal modes. The application of this spectroscopic assignment has proved useful in the study of carbon-based materials with controlled nanostructure obtained from pyrolysis of selected precursors [6].

Intrigued by the evident success of DFT calculations in reproducing the IR spectral features of variously substituted benzene
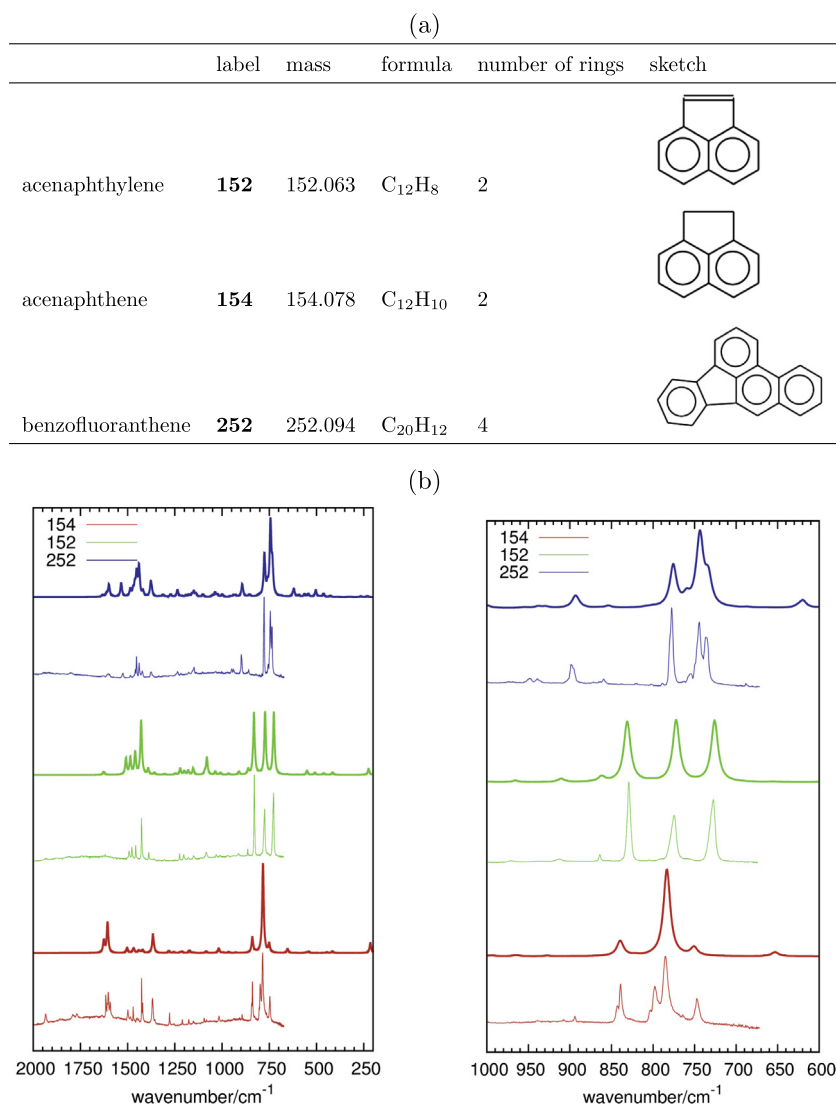
---

derivatives and even larger PAHs (see Fig. 1), we have decided to address the problem of describing the characteristic out-of-plane bending vibrations in more complex PAHs, which could be produced by combustion or obtained by direct synthesis. One of the aims of this investigation is to provide a theoretical background for the use of IR spectroscopy in the characterization of combustion products. In particular, we would like to explore the sensitivity of IR spectroscopy to different molecular shapes and assess the information that can be retrieved from IR spectra of polycyclic aromatic combustion products, which consist of mixtures of several compounds with similar nature. Finally, we notice that the potential applications of the present investigation may go beyond the characterization of combustion products since polyaromatic compounds are also of interest in Materials Science [7,8] and Astrophysics [9].

2. Experimental and computational details

The set of chemical formulae considered in this study has been determined from the analysis of mass spectrometry data collected on products originated from the incomplete combustion of gaseous fuel under well-controlled conditions. This has allowed to carry out the theoretical investigation on a realistic guess of structures, thus avoiding as far as possible bias effects caused by arbitrary choices of PAHs. A complex mixture of combustion products was isokinetically sampled from a laminar premixed flame burning ethylene under fuel-rich conditions (McKenna burner, ethylene/O_2, 50.0/50.0, C/O = 1, T_{max} = 1650 K, cold gas velocity = 4 cm/s). A stainless steel water-cooled probe was vertically inserted along the flame axis at 14 mm of height above the burner, corresponding to the end of soot formation region where flame-formed species (including soot, tar-like species and PAH) reach their maximum concentration [10]. Additional details on the combustion set-up and sampling procedure are given in Refs. [10,11]. The combustion products, recovered along the sampling line, were extracted with dichloromethane (DCM) to remove soluble organic species that include light PAH (C < 24). Following DCM extraction, the solid sample was re-extracted with the N-methylpyrrolidone (NMP) in order to detach strongly adsorbed species, containing large PAH (C > 24) [10]. Large PAH were identified in NMP-extract by Atmospheric Pressure Laser Desorption Mass Spectrometry (AP-LDI-MS). The mass spectrum consists of a sequence of high intensity signals spaced 24–26 u, identified as PAHs with an even number of carbon atoms (even-C-numbered PAHs), superimposed

(a)

	label	mass	formula	number of rings	sketch
acenaphthylene	**152**	152.063	$C_{12}H_8$	2	
acenaphthene	**154**	154.078	$C_{12}H_{10}$	2	
benzofluoranthene	**252**	252.094	$C_{20}H_{12}$	4	

(b)

Fig. 1. (a) The three PAHs used for benchmarking the computational procedure adopted for the simulation of IR absorption spectra. The informal naming scheme refers to the rounded molecular mass. (b) Comparison between experimental (thinner line, below) and simulated (thicker line, above) IR spectra of the three PAHs reported in panel (a).

on a sequence of ions spaced at 12 u relatively to the higher intensity signals, ascribed to odd-C-numbered PAHs (Fig. 2). The PAH pattern extends up to 600 u that is the higher mass detectable. The large PAHs selected for this study arise from the even-C-numbered sequence, which is the most abundant in the analyzed combustion products. For a given mass number many structural isomers do exist and their number increases with the number of carbon atoms, as the molecules get larger. PAHs differ-entiation by mass spectrometry is difficult since mass spectra of isomeric PAHs are virtually identical. In the case of PAH from combustion sources, their structures are deduced by considering specific addition pathways. A pool of structural isomers (Table 1) were hypothesized by taking into account two mechanisms mainly responsible for PAH formation and growing in combustion flames:(i) pure acetylene based pathway (hydrogen abstraction acetylene addition, HACA) accounting for a mass increment of 24 u [12,13]) and (ii) intramolecular rearrangements (acetylene addition followed by isomerization, resulting in a mass increment of 26 u [14–17]). The intramolecular rearrangement is responsible of the interconversion of the PAH isomers leading to more stable PAHs.

Micro FT-IR measurements reported in Fig. 1 were carried out with Nicolet Nexus equipment coupled with a Thermo-Nicolet Continuμm infrared microscope and a cooled MCT detector (77 K). The spectra of the samples (as powders) were acquired by using the diamond anvil cell technique with a 15× infrared objective (64 scans, 1 cm^{-1} resolution, 650–4000 cm^{-1} spectral range). Compared with the KBr pellet technique, the micro FT-IR setup allows recording spectra with a minimal sample amount.

DFT calculations on the set of PAHs reported in Table 1 have been carried out with Gaussian09 [18] with the B3LYP functional and 6-311G(d,p) basis set. Geometry optimization was followed by a frequency calculation aiming at simulating the infrared absorption. The main results of our DFT calculations can be found in Supplementary Information (optimized geometries, vibrational frequencies and IR intensities). The computed band positions (cm^{-1}) and intensities (km/mol) have been directly used to reproduce infrared absorption spectra of individual PAHs by summing Lorentzian line shapes (10 cm^{-1} FWHM) corresponding to the computed bands.

The "R" statistical package [19] has been used for hierarchical cluster analysis [20] of the simulated IR spectra in the 200-1800 cm^{-1} wavenumber range. For the sake of simplicity, we have adopted classical and well-established data clustering tools, choosing the popular Ward's hierarchical algorithm [20]. Details of this procedure are reported in Supplementary Information. The dendrogram produced by hierarchical cluster analysis has been represented by using the APE library available within R [21]. OpenBabel tools [22] have been employed to obtain ring counts

and help with data processing of the optimized molecular structures.

3. Data analysis

Detailed results of the clustering analysis carried out on our DFT simulations of the IR spectra of the selected PAHs are listed in Supplementary Information. Here below we highlight the main outcomes that emerged from the careful inspection of the correlations among the molecules, which we think are also relevant to IR spectroscopy. It is worth mentioning that this data mining work would have been prohibitive without the help coming from hierarchical cluster analysis and data representation, both offered by "R" environment [19]. As it is apparent from the inspection of Figs. 3 and 4, Ward's clustering algorithm proceeds to group the different PAHs in such a way that the strong IR signals (almost ever corresponding to CH-out-of-plane vibrations – see Fig. 5) of each cluster are very close, so that a representative wavenumber (i.e., a marker peak) can be associated with the majority of the clusters. This is for instance the case of cluster #1 (875 cm^{-1}), #2 (760 cm^{-1}), #3 (863 cm^{-1}), #4 (900 cm^{-1}), #5 (851 cm^{-1}), #7 (829 cm^{-1}), #9 (783 cm^{-1}), #10 (857 cm^{-1}) (in parenthesis the unscaled average vibrational wavenumber determined from DFT calculations). The situation for clusters #6, #8 and #11 appears somewhat more scattered and no clear marker can be found, even though the clustering algorithm correctly identifies groups with clearly similar IR spectra overall (see Supplementary Information for the cluster-wise detailed representation of the simulated IR spectra). A summary of the outcomes of the present cluster analysis is reported in tabular form in Table 2 as a collection of the stronger IR modes in the 200–1800 cm^{-1} region.

Encouraged by the good performance of Ward's hierarchical clustering in analyzing and correctly retrieving the similarities among the simulated IR spectra, we have decided to dwell more into the molecular origin of the IR signals hoping to find a connection between the marker bands found by the statistical analysis and the molecular structure of the many PAHs considered. For this reason we have also analyzed the vibrational nuclear displacements associated with the strong IR features reported in Table 2. In all cases these have been found to be CH-out-of-plane bend-ing vibrations, localized in different ways along the periphery (edge) of the PAHs. We report in Table 2 a sketch of these cluster-specific displacement patterns, while in Fig. 5 we report a selection of relevant out-of-plane normal modes for the 8 clusters (#1–#5, #7, #9, #10) where markers can be more easily found. The inspection of Table 2 and Fig. 5 suggests two observations:

(i) The strong IR peaks of the PAHs considered are usually associated with collective CH out-of-plane bending vibrations.
(ii) The nuclear displacements associated with intense IR peaks possess peculiar patterns that are dependent on the topology of the molecular edge and differ from cluster to cluster.

Point (i) is expected, based on the chemical structure of the molecules considered and well-established spectroscopic correlations [1–3,6]. Point (ii) is more interesting and it reveals information on molecular structure that could be retrieved from IR spectra.

Even though the markers found can be nicely traced back to IR signals well-known in the literature (named SOLO, DUO, TRIO and QUATRO according to the number of CH bonds involved in the vibration along the edge of the polyaromatic system [6]), the inspection of the results from DFT calculations (Fig. 5) reveals that for complex edge topologies the nuclear displacement patterns are more involved and usually imply the simultaneous motion of more than one (SOLO, DUO, TRIO, QUATRO) moiety. This happens for
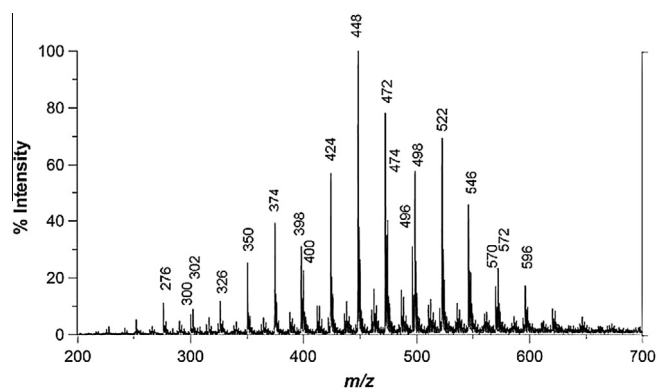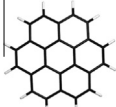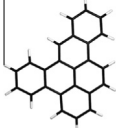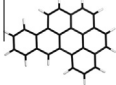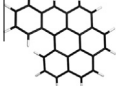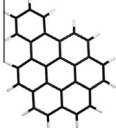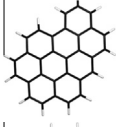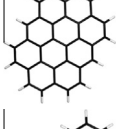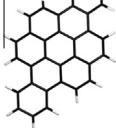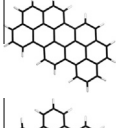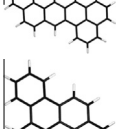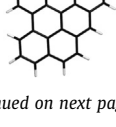

Fig. 2. AP-LDI-MS spectrum of large PAH (C > 24) sampled in ethylene flame (adapted from [10]).
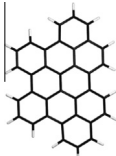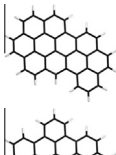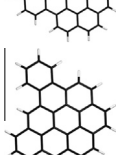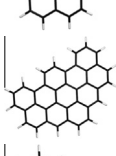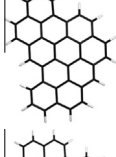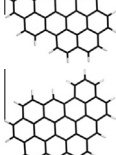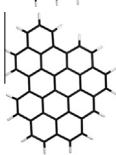
Table 1
The set of 51 PAHs considered in this work sorted by increasing molecular weight. The naming scheme (label) refers to the rounded molecular mass, followed by a one-letter code for distinguishing among different isomers.
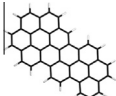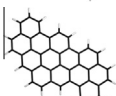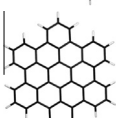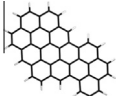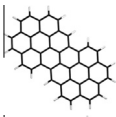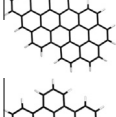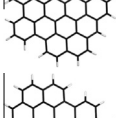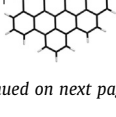
	Label	Mass	Formula	Number of rings	Sketch
1	**300**	300.352	$C_{24}H_{12}$	7	
2	**302a**	302.368	$C_{24}H_{14}$	6	
3	**302b**	302.368	$C_{24}H_{14}$	6	
4	**326a**	326.389	$C_{26}H_{14}$	7	
5	**326b**	326.389	$C_{26}H_{14}$	7	
6	**350a**	350.411	$C_{28}H_{14}$	8	
7	**374**	374.432	$C_{30}H_{14}$	9	
8	**398**	398.454	$C_{32}H_{14}$	10	
9	**400**	400.469	$C_{32}H_{16}$	9	
10	**424a**	424.491	$C_{34}H_{16}$	10	
11	**424b**	424.491	$C_{34}H_{16}$	10	
12	**424c**	424.491	$C_{34}H_{16}$	10	

Table 1 (*continued*)

	Label	Mass	Formula	Number of rings	Sketch
13	**424d**	424.491	$C_{34}H_{16}$	10	
14	**424e**	424.491	$C_{34}H_{16}$	10	
15	**448a**	448.512	$C_{36}H_{16}$	11	
16	**448b**	448.512	$C_{36}H_{16}$	11	
17	**448c**	448.512	$C_{36}H_{16}$	11	
18	**448d**	448.512	$C_{36}H_{16}$	11	
19	**448e**	448.512	$C_{36}H_{16}$	11	
20	**472a**	472.534	$C_{38}H_{16}$	12	
21	**472b**	472.534	$C_{38}H_{16}$	12	
22	**472c**	472.534	$C_{38}H_{16}$	12	
23	**472d**	472.534	$C_{38}H_{16}$	12	
24	**474**	474.55	$C_{38}H_{18}$	11	
25	**496a**	496.555	$C_{40}H_{16}$	13	

Table 1 (*continued*)

	Label	Mass	Formula	Number of rings	Sketch
26	**496b**	496.555	$C_{40}H_{16}$	13	
27	**498**	498.571	$C_{40}H_{18}$	12	
28	**522a**	522.592	$C_{42}H_{18}$	13	
29	**522b**	522.592	$C_{42}H_{18}$	13	
30	**522c**	522.592	$C_{42}H_{18}$	13	
31	**546a**	546.614	$C_{44}H_{18}$	14	
32	**546b**	546.614	$C_{44}H_{18}$	14	
33	**546c**	546.614	$C_{44}H_{18}$	14	
34	**546d**	546.614	$C_{44}H_{18}$	14	
35	**570a**	570.635	$C_{46}H_{18}$	15	
36	**570b**	570.635	$C_{46}H_{18}$	15	
37	**570c**	570.635	$C_{46}H_{18}$	15	
38	**570d**	570.635	$C_{46}H_{18}$	15	
39	**570e**	570.635	$C_{46}H_{18}$	15	

Table 1 (*continued*)

	Label	Mass	Formula	Number of rings	Sketch
40	**570f**	570.635	$C_{46}H_{18}$	15	
41	**572d**	572.651	$C_{46}H_{20}$	14	
42	**572e**	572.651	$C_{46}H_{20}$	14	
43	**572f**	572.651	$C_{46}H_{20}$	14	
44	**572a**	596.672	$C_{48}H_{20}$	15	
45	**572c**	596.672	$C_{48}H_{20}$	15	
46	**596b**	596.672	$C_{48}H_{20}$	15	
47	**596c**	596.672	$C_{48}H_{20}$	15	
48	**596e**	596.672	$C_{48}H_{20}$	15	
49	**596f**	596.672	$C_{48}H_{20}$	15	
50	**596g**	596.672	$C_{48}H_{20}$	15	
51	**596h**	596.672	$C_{48}H_{20}$	15	

instance in cluster #1. The vibrational motion of a CH solo bond placed within a (symmetric) bay topology of the edge is significantly coupled with the motion of the duo moieties along the molecular edge (see Fig. 5 and Table 2). The marker resulting from this coupling (duo + bay-solo) appears at a lower frequency (875 cm^{-1}) with respect to the solo marker of a straight edge, found in cluster #4 at 900 cm^{-1}. The latter cluster also features the case of molecule **472d** whose classification is difficult. Even though its simulated spectrum can be accepted within cluster #4 (see Fig. 4) the strongest IR peaks of **472d** are associated with

rather different displacements with respect to the rest of cluster #4 (see Fig. 5). This kind of situation, to a worse degree, is also found in those clusters for which a straightforward IR marker can-not be found (i.e. clusters #6, #8, #11). The simulated spectra can be certainly grouped based on overall similarity, but the strongest IR features of the molecules within each cluster appear rather scattered in frequency, thus making the search for common nuclear displacement patterns more difficult. Furthermore, the case of molecule **474** belonging to cluster #5 reveals that the common nuclear displacement pattern found within the cluster also couples
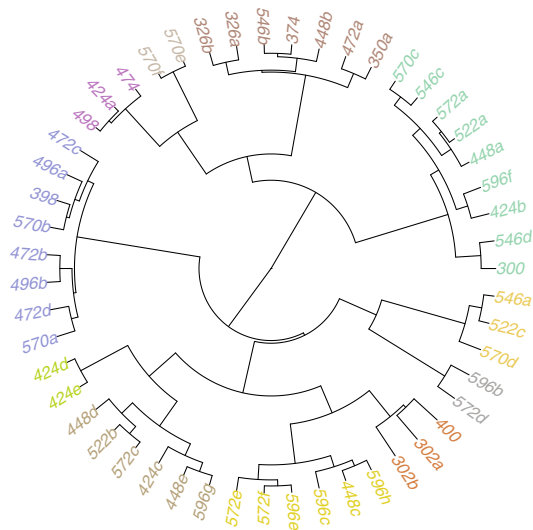
Fig. 3. Outcome of Ward's hierarchical cluster analysis carried out on the DFT simulated IR spectra of the set of 51 selected PAHs. This is shown as a phylogram [21] representing the clustering hierarchy. The color scheme of the phylogram is given for the optimal choice of 11 clusters (see Supplementary Information). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to other CH out-of-plane bending coordinates that are characteristic of the specific molecule. Even if a single IR marker cannot be found in clusters #6, #8 and #11, still Ward's algorithm performs well and it is able to group similar PAHs based on their similar IR spectra. The members of the three latter clusters are reported in Supplementary Information and the common structures which still can be identified are discussed there.

4. Analysis of PAH edges through CH bonds

The discussion above has shown that the IR signals in the out-of-plane CH bending region carry significant information about the molecular edge of PAHs. This was the direct result of the statistical analysis carried out on the spectral features simu-lated with DFT methods. We strive to address now a more funda-mental question, i.e. the connection existing between the wavenumber of the characteristic signals (SOLO, DUO, TRIO, QUATRO) and the edge topology. To this aim it is convenient to introduce two descriptors of the edge hydrogen atoms.

The first descriptor is the topological index τ_k which, for a given k-th hydrogen (bonded to carbon C_k), counts the number of hydrogens which are bonded to the two carbons directly bonded to C_k. In the set of PAHs here considered (all formed uniquely by sp^2-hybridized carbons) τ_k can assume just the values of 0, 1 and 2. In particular, $\tau = 0$ corresponds to SOLO hydrogens, $\tau = 1$ is found in DUO moieties, while $\tau = 2$ characterizes the central hydrogens of TRIO and QUATRO moieties.

The second descriptor is inspired by McKean's work on isolated CH stretching frequencies [23,24]. However, instead of considering CH stretching vibrations (which are not directly related to the present analysis of bending vibrations) we take into consideration isolated CH out-of-plane (opla) bending vibrations. The concept is the same as in McKean works: all hydrogens except one under consideration are isotopically substituted with deuterium. This effectively decouples hydrogen vibrations and one can analyze each CH bond individually, independently from the others. A summary of the two CH descriptors for the full set of PAHs considered here is graphically reported in the plots of Fig. 6. As expected [23,24] a linear correlation is found between the isolated CH stretching

wavenumber and the CH bond length (left panel of Fig. 6). More interestingly, the right panel of Fig. 6 clearly shows that the values of the isolated CH opla wavenumber group into six distinct clusters. The topological index τ is a fairly robust cluster marker. It is constant within the same cluster, even though for $\tau = 0$ we find two clusters and for $\tau = 1$ we find three clusters. To help rationalize these data, we have directly reported in Fig. 7 a representative selection of PAH structures with the classification of edge hydrogen atoms, as deduced from the right panel of Fig. 6 (see Supplementary Information for the representation over the full set of PAHs). Fig. 7 adopts the same color scheme of Fig. 6. Inspection of Fig. 7 straightforwardly reveals the description of the six clusters:

- (#1, $\tau = 0$) red: SOLO hydrogen atoms with no nearby hydrogens
- (#2, $\tau = 1$) green: DUO hydrogen atoms with no nearby hydrogens
- (#3, $\tau = 2$) blue: TRIO/QUATRO hydrogen atoms with no nearby hydrogens
- (#4, $\tau = 0$) magenta: SOLO hydrogen atoms with nearby hydrogens (bay)
- (#5, $\tau = 1$) dark green: DUO hydrogen atoms with nearby hydrogens
- (#6, $\tau = 1$) cyan: DUO hydrogen atoms in crowded non-planar geometry (helicene-like, found in **326b**; see also **572d** and **596b** in Supplementary Information)

Hence, we deduce that the topological index τ plays the main role in determining the isolated CH opla wavenumber. However, a secondary influence is caused by the hydrogen–hydrogen dis-tance along the PAH edge. We name "crowded" those edge hydro-gens characterized by short non-interacting H–H distances (see for instance the case of **326b** in Fig. 7). Crowded SOLO and DUO hydro-gen atoms find their own clusters (#4, #5, #6), well separated from that of free SOLO (#1) and free DUO (#2). Since there are no crowded TRIO/QUATRO ($\tau = 2$) hydrogen atoms in our PAH set, cluster #3 has no counterpart. These findings constitute the molecular basis of the empirical correlations previously described and available in the literature [4–6].

Isolated CH frequencies are not an easily measurable quantity, since they require selective isotopic substitutions that are difficult to achieve in a controlled way. However, they can be quickly evaluated based on available vibrational force fields (for instance, those computed in this work with DFT). Furthermore, they form a well-defined basis set of localized oscillators whose linear combi-nations could be used to describe the normal modes of interest in isotopically un-substituted PAHs, leading to an effective vibrational Hamiltonian for the description of opla modes. Hence, apart from the couplings existing among these localized oscillators, the fact that the diagonal part of this effective Hamiltonian (i.e., isolated CH OPLA frequencies) shows definite trends related to molecular topology is a clear indication of the fundamental connection existing between PAH edges and the associated signatures in the IR out-of-plane bending region. However, developing such an effective Hamiltonian based on present results is not a straight-forward task and it would require further efforts beyond the scope of this work. An alternative much simpler way to relate edge topol-ogy to opla modes in PAHs is to introduce the average topology index $\langle \tau_i \rangle$ for a given i-th normal mode of interest, as follows:

$$\langle \tau_i \rangle = \frac{\sum_k L_{ki}^2 \tau_k}{\sum_k L_{ki}^2} \tag{1}$$

where τ_k is the topological index of the k-th edge hydrogen and L_{ki} is the out-of-plane displacement of k-th hydrogen in a given
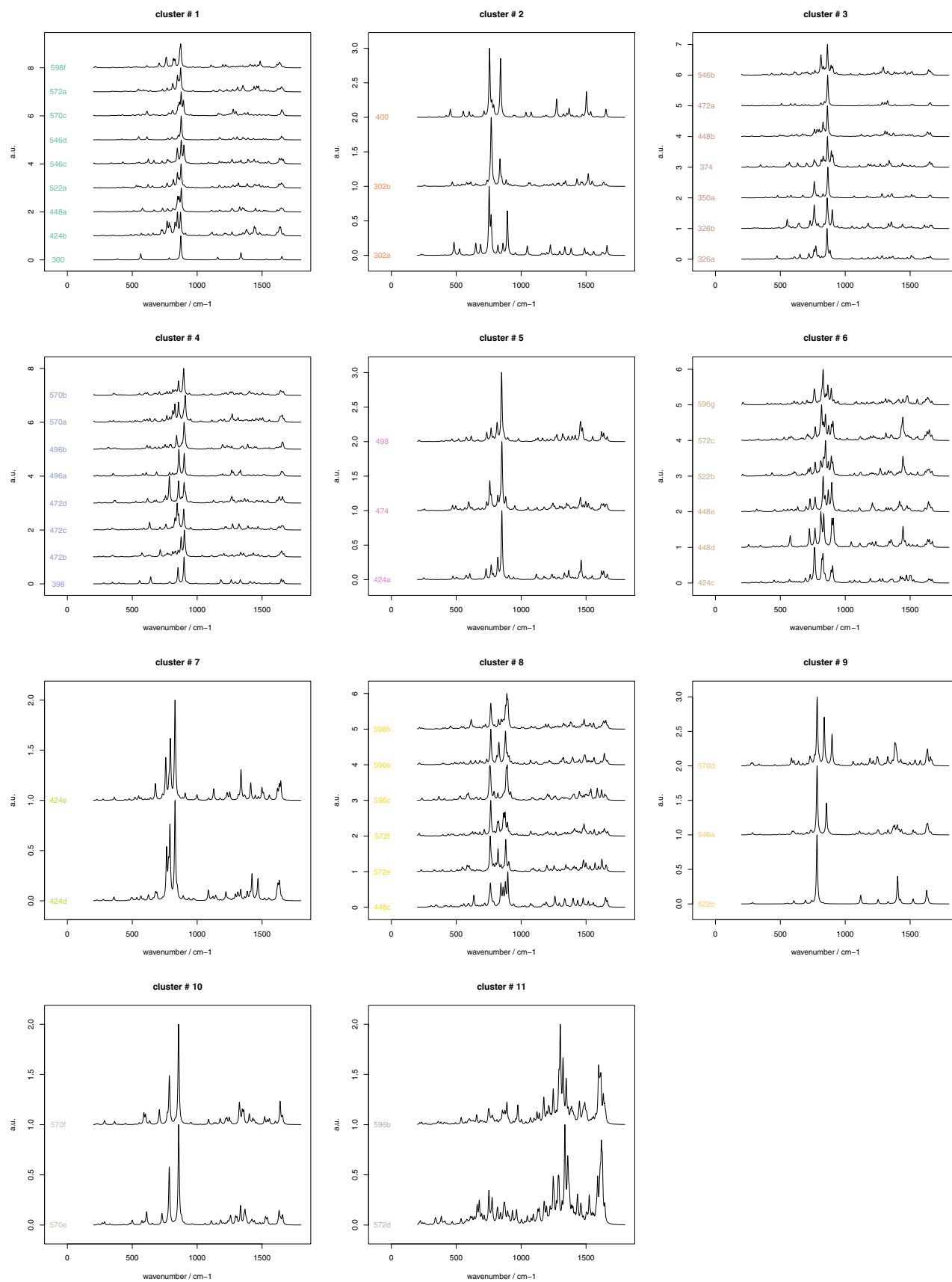
Fig. 4. The simulated IR spectra of the 51 PAHs considered (see Table 1) grouped according to Ward's hierarchical clustering algorithm (see text for details). The figure is meant to provide a quick overview of the IR spectra, for a more detailed representation the interested reader should consult the material in Supplementary Information. Data from unscaled DFT calculations.
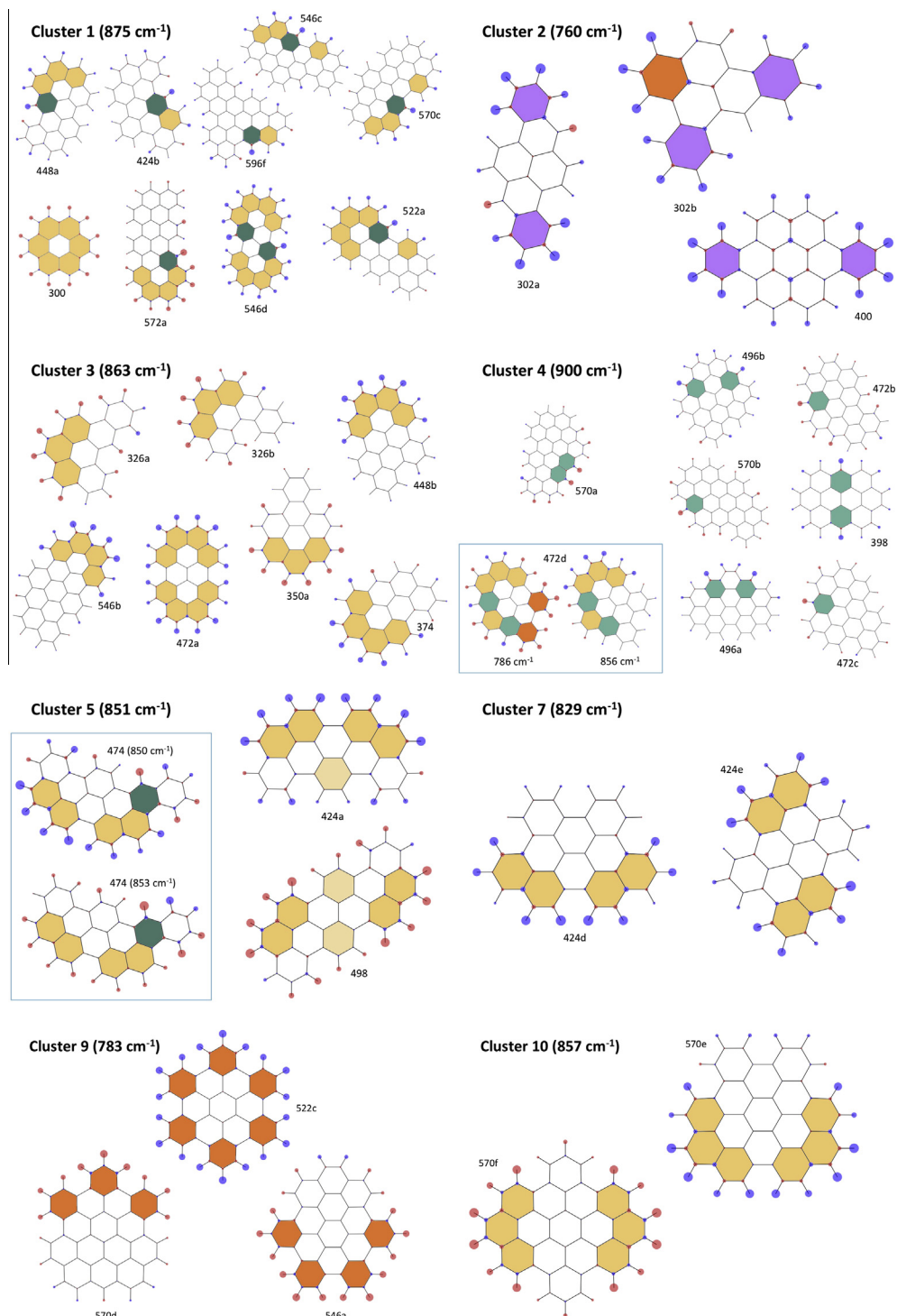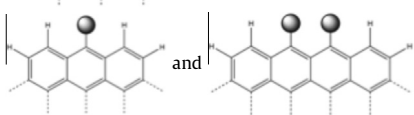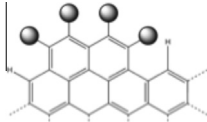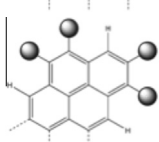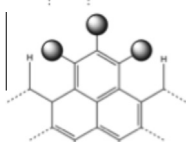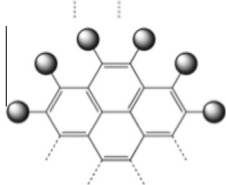
Fig. 5. The nuclear displacements associated with the intense IR signals of the clusters for which a marker band can be identified (see text). The average wavenumber of the marker band is given for each cluster. The displacements are all associated with CH out-of-plane bending. The size of the circle is proportional to the amplitude of the z-displacement of each atom within the normal mode; the color of the circle is indicative of the relative sign of the displacement. Data from unscaled DFT calculations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
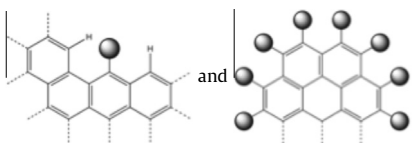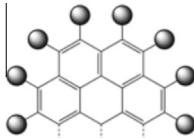
i-th normal mode. Hence, the average topology index provides the indication of which kind of edge hydrogen atoms participate to a given opla mode. The analysis which follows has been carried out on the selected CH out-of-plane modes with large IR absorbance within the full set of PAHs of Table 1. In this way, by making use of the results from DFT calculations, we can explore the relationship existing between the peak position of opla modes and the average topology of the hydrogen atoms involved in the associated normal modes. The summary of this evaluation is reported in Fig. 8. The correlation existing between $\langle \tau \rangle$ and opla peak position v_{op} (cm^{-1}) is evident. The quadratic fitting function obtained using wavenumber ($v_{op,i}$) and nuclear displacement data (L_{ki}) from DFT calculations has the following equation:

Table 2

Description of the local patterns of CH out-of-plane bending displacements associated with the unscaled IR marker bands identified by the hierarchical clustering algorithm (compare with Fig. 5). These findings of the clustering analysis are compatible with the known experimental markers of CH out-of-plane bending reported in the literature [6].

Cluster	Marker (cm^{-1})	Sketch	description
#1	875	and	SOLO (bay), DUO
#2	760		QUATRO
#3	863		DUO
#4	900	and	SOLO
#5	851		DUO
#7	829		DUO
#9	783		TRIO
#10	857		DUO

SOLO DUO TRIO QUATRO
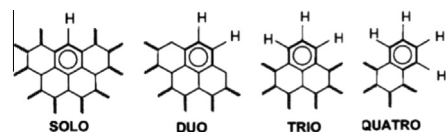
SOLO	860–910		
DUO	800–810	810–860	
TRIO	750–770	770–800	800–810
QUATRO	730–750	750–770	

$$\langle \tau \rangle = -4.483 \times 10^{-5} v_{op}^2 + 6.634 \times 10^{-2} v_{op} - 23.13 \qquad (2)$$

The data do show some spread, but almost all points are found within a distance of just 0.26 (units of $\langle \tau \rangle$) from the fitting function. SOLO, DUO, TRIO and QUATRO signals distribute along the trend, from low to high τ values. Nuclear displacements of selected modes are also reported in the top panel of Fig. 8 to help appreciating the link existing between the topological index, the peak location and the molecular structure. In the present analysis we also consider two additional model molecules with ribbon shape (denoted zig and arm, originally not included in the set of PAHs

reported in Table 1) which offer examples of extended armchair and zigzag edge structures. As expected, their characteristic opla signals follow the general trend. Additionally, the arm model also presents a well-defined QUATRO feature, due to a normal vibration localized at the end of the ribbon.

The horizontal spread of DUO signals found in Fig. 8 can be interpreted based on McKean isolated CH opla vibrations. In fact, while passing from (**572c**, arm) to **498** and finally **300** we observe a progressive increase of v_{op} that corresponds to a transition from crowded DUO hydrogens (**572c**, arm) to uncrowded ones, as in coronene (*i.e.*, molecule **300**). Consistently to the data reported in
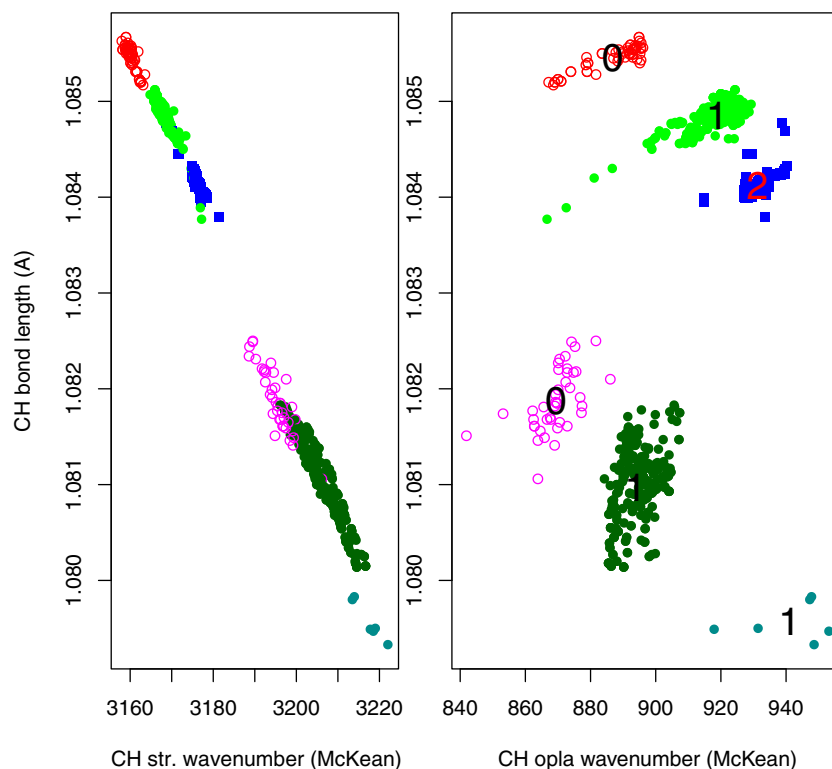
Fig. 6. Correlations between isolated CH normal modes (after McKean's selective deuteration scheme [23]) and CH bond length: the left panel reports the case of CH stretching vibrations; the right panel reports the case of CH out-of-plane vibrations. Data clustering has been carried out based on the data of the right panel, making use of the topological index τ, CH bond length and opla wavenumber. The τ-index is reported for each cluster. The data classification scheme worked out in the right panel is then adopted in the left panel. All reported data have been obtained from DFT calculations (B3LYP/6-311G(d,p)).
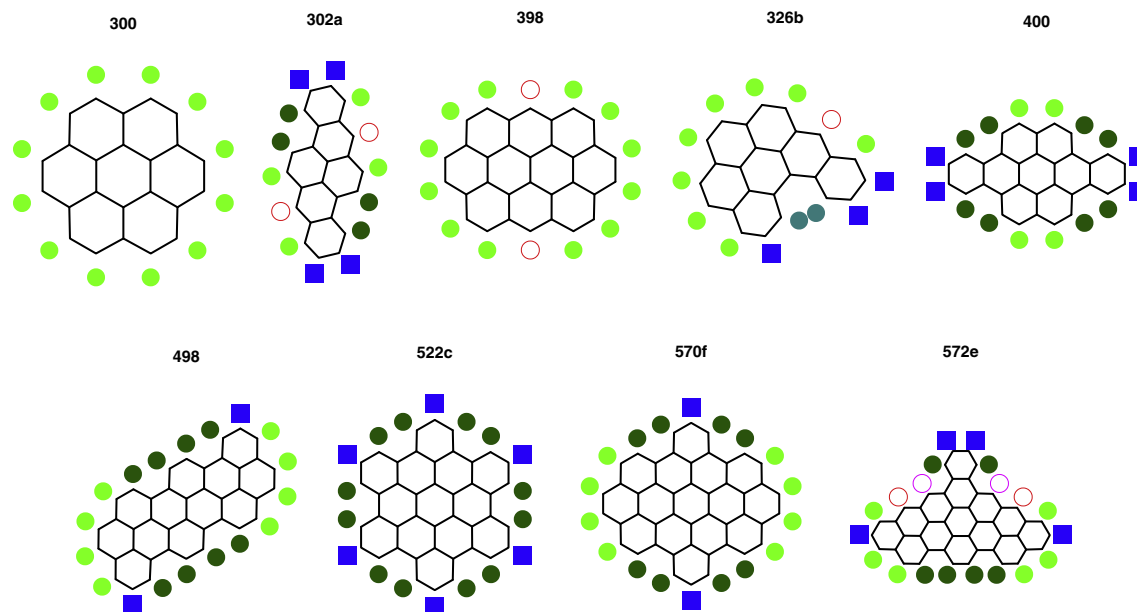
Fig. 7. Representation of the topological index for a selection of the PAHs considered in this work (see Supplementary Information for the representation of the complete set of PAHs). Open circles denote $\tau = 0$, full circles $\tau = 1$, full squares $\tau = 2$. The colors are given in accordance to the data clustering scheme of isolated opla wavenumber / CH bond length reported in Fig. 6.

Fig. 8, Fig. 6 shows that the majority of uncrowded DUO hydrogens (cluster #2) have a higher isolated v_{op} than crowded DUO hydrogens (cluster #5).

The fitting function (2) could be used to analyze experimental data in order to characterize the average edge topology of polyaromatic samples. Under this regard, it is useful to consider two issues:
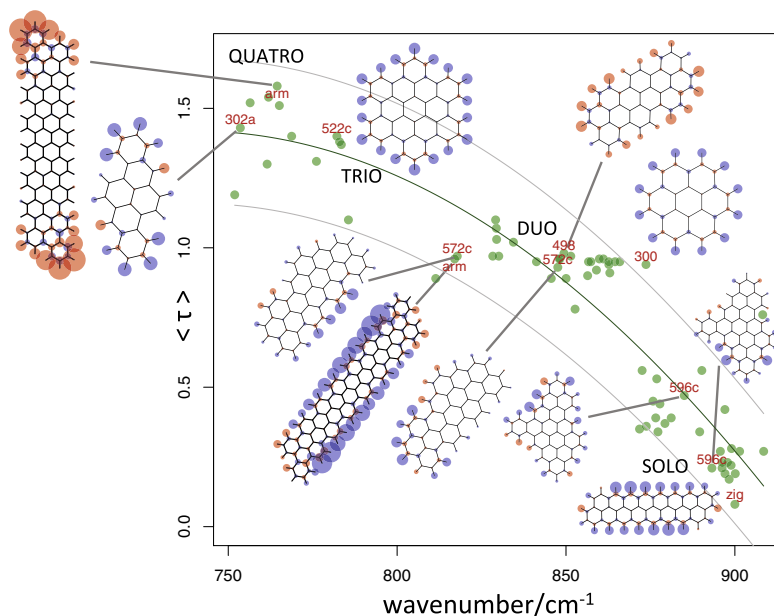
Fig. 8. Correlation between the average topological index $\langle\tau\rangle$ and the opla modes with largest IR absorption found in the range 400–1000 cm^{-1}. For each PAH we have considered only those opla modes whose intensity is at least 0.75 times that of the strongest mode in the given range. The modes of selected PAHs are reported to help illustrate the dispersion of values over the SOLO, DUO, TRIO and QUATRO regions.

(i) Vibrational wavenumber obtained from DFT calculations in this work (B3LYP/6-311G(d,p)) should be scaled [25] when making comparison with experimental data (hence the fitting coefficients of Eq. (2) should be changed accordingly). Our experience with opla modes of PAHs is that a factor of 0.98 provides a fair match against experimental data. For instance, this can be appreciated in Fig. 2 and comparing in Fig. 9 the wavenumber-scaled distribution of CH out-of-plane peaks (from DFT) with the experimental IR spectrum collected on a soot sample obtained from a pre-mixed ethylene/O$_2$ flame.

(ii) The average $\langle\tau\rangle$ depends on how the modes are localized on hydrogen atoms with different topological index. As a first approximation, which turns out to be reasonable based on the examples reported in Fig. 8, one could just assume uniform distribution of the opla mode of interest over the hydrogen atoms with the same topological index (SOLO, DUO) or within the triplet/quadruplet of hydrogen atoms forming a TRIO/QUATRO moiety.

This leads to the following "ideal" values of the average topological index:

– SOLO 0
– DUO 1
– TRIO (1 + 2 + 1)/3 = 1.33
– QUATRO (1 + 2 + 2 + 1)/4 = 1.5

It is worth realizing that these four ideal values nicely correspond to the correct regions of the plot reported in Fig. 8.

Conclusions

By considering the case of acenaphthylene, acenaphthene and benzofluoranthene, we have shown that B3LYP/6-311G(d,p) DFT calculations can reproduce with good accuracy the IR spectra. Secondary effects on peak positions and relative intensities, most
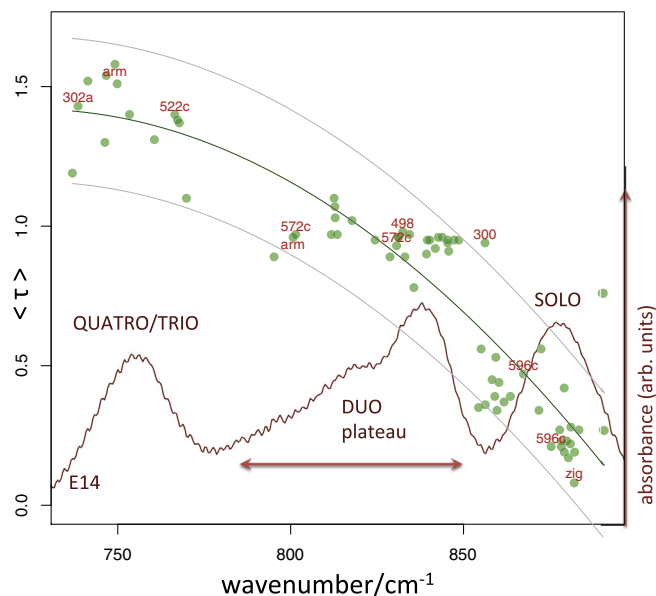
Fig. 9. Comparison between the experimental IR spectrum of a soot sample (dark red line) and the classification of CH bonds introduced in this work (wavenumbers from DFT have been scaled by 0.98 with respect to the data reported in Fig. 8). The soot sample was obtained from premixed ethylene/O$_2$ flame and was collected at 14 mm above the burner.

likely due to aggregation, of course cannot be modeled by calculations on isolated molecules. The systematic theoretical analysis of a set of 51 PAHs, whose chemical structures have been deduced from mass-spectrometry data on combustion products, has shown that IR spectra in the CH out-of-plane bending region (600–900 cm^{-1}) are dependent upon the structure of the molecules. This is in accordance with previous spectroscopic assignments [1–3,5,6]. In particular, complex molecular topologies show IR markers that are due to linear combinations of four fundamental

vibrational patterns which are named SOLO, DUO, TRIO, QUATRO in the literature [5]. From the point of view of data analysis approaches, we have shown that Ward's hierarchical cluster analysis [20] can nicely handle the set of IR spectra simulated by DFT and allows one quickly to group PAHs based on the similarity among their spectra. Furthermore, the optimal clustering balance concept [26], allows to unambiguously fixing the best number of clusters to be considered in the representation of the data (see Supplementary Information).

The link existing in PAHs between edge topology and opla IR signals has been further investigated by introducing the concept of isolated CH opla vibration (after McKean [23,24]) and a topological index (τ) which differentiates hydrogen atoms based on the connectivity of the carbon to which they are bonded. We have discovered a correlation involving CH bond length, τ-index and isolated opla wavenumber. Furthermore, by introducing the average τ-index of a collection of different CH oscillators, we have shown that this correlation is the molecular origin of the characteristic SOLO, DUO, TRIO, QUATRO signals.

Among the outcomes of the present investigation, it is worth mentioning that the IR features of PAHs in the 600–900 cm^{-1} range are highly dependent on the chemical structure and topology of the molecule at the edge and little influence from the bulk is observed. Hence, for modeling the IR response of graphene edges in this wavenumber range, one could consider rather small models that are computationally more appealing. For instance, this is the case of ribbon models denoted zig and arm in Fig. 8. This fact is remarkable because even if the system is π-conjugated (thus in principle affected by long range interactions) CH opla bending vibrations mainly involve atoms at the molecular edge and the polarization of the vibrational displacements is orthogonal to the plane where π-conjugation occurs. The situation for the simulation of Raman spectra is opposite, since the G and D features involve the whole molecular structure and are thus highly dependent on π-conjugation [27].

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version. These data include the Cartesian coordinates of the compounds investigated in this article and details about the data analysis.

References

[1] R.N. Jones, C. Sandorfy, The application of infrared and raman spectrometry to the elucidation of molecular structure, Chemical Applications of Spectroscopy – Techniques of Organic Chemistry, vol. IX, Wiley, 1956.

[2] L. Bellamy, The Infrared Spectra of Complex Molecules, Springer, Dordrecht, 1975.

[3] G. Socrates, Infrared and Raman Characteristic Group Frequencies, third ed., Wiley, Chichester, West Sussex, England, 2004.

[4] F.C. Witteborn, S.A. Sandford, J.D. Bregman, L.J. Allamandola, M. Cohen, D.H. Wooden, A.L. Graps, Astrophys. J. 341 (1989) 270.

[5] M. Von Zander, B.G. Teubner, Polycyclische aromaten (reihe: Teubner studienbücher chemie, stuttgart, 1995, isbn 3–519–03537–5), Angewandte Chemie 108 (19) (1996), http://dx.doi.org/10.1002/ange.19961081935. 2411-2411.

[6] A. Centrone, L. Brambilla, T. Renouard, L. Gherghel, C. Mathis, K. Mullen, G. Zerbi, Structure of new carbonaceous materials: the role of vibrational spectroscopy, Carbon 43 (2005) 1593–1609, http://dx.doi.org/10.1016/j.carbon.2005.01.040.

[7] K. Seto, Y. Furukawa, Study on solid structure of pentacene thin films using raman imaging, J. Raman Spectrosc. 43 (2012) 2015–2019, http://dx.doi.org/10.1002/jrs.4090.

[8] M.D. Watson, A. Fechtenkötter, K. Müllen, Big is beautiful – aromaticity revisited from the viewpoint of macromolecular and supramolecular benzene chemistry, Chem. Rev. 101 (5) (2001) 1267–1300.

[9] C.W. Bauschlicher Jr., C. Boersma, A. Ricca, A.L. Mattioda, J. Cami, E. Peeters, F.S. de Armas, G.P. Saborido, D.M. Hudgins, L.J. Allamandola, The nasa ames polycyclic aromatic hydrocarbon infrared spectroscopic database: the computed spectra, Astrophys. J. Suppl. Ser. 189 (2010) 341–351, http://dx.doi.org/10.1088/0067-0049/189/2/341.

[10] M. Alfe, B. Apicella, A. Tregrossi, A. Ciajolo, Identification of large polycyclic aromatic hydrocarbons in carbon particulates formed in a fuel-rich premixed ethylene flame, Carbon 46 (2008) 2059–2066, http://dx.doi.org/10.1016/j.carbon.2008.08.019.

[11] M. Alfe, B. Apicella, J.-N. Rouzaud, A. Tregrossi, A. Ciajolo, The effect of temperature on soot properties in premixed methane flames, Combust. Flame 157 (2010) 1959–1965.

[12] M. Frenklach, D. Clary, W. Gardiner, S. Stein, Detailed kinetic modeling of soot formation in shock-tube pyrolysis of acetylene, Proc. Combust. Inst. 20 (1984) 887–901.

[13] M. Frenklach, J. Warnatz, Detailed modeling of PAH profiles in a sooting low-pressure acetylene flame, Combust. Sci. Technol. 51 (1987) 265–283.

[14] J.B. Howard, J.P. Longwell, J.A. Marr, C.J. Pope, W.F. Busby Jr., A.L. Lafleur, K. Taghizadeh, Effects of PAH isomerizations on mutagenicity of combustion products, Combust. Flame 101 (3) (1995) 262–270, http://dx.doi.org/10.1016/0010-2180(94)00210-J. URL: http://www.sciencedirect.com/science/article/pii/001021809400210J.

[15] L.T. Scott, N.H. Roelofs, Thermal rearrangements of aromatic compounds. 11. benzene ring contractions at high temperatures. evidence from the thermal interconversions of aceanthrylene, acephenanthrylene, and fluoranthene, Journal of the American Chemical Society 109 (18) (1987) 5461–5465. arXiv:http://pubs.acs.org/doi/pdf/10.1021/ja00252a025, doi:10.1021/ja00252a025. URL http://pubs.acs.org/doi/abs/10.1021/ja00252a025.

[16] A. Stone, D. Wales, Theoretical studies of icosahedral c60 and some related species, Chem. Phys. Lett. 128 (5–6) (1986) 501–503, http://dx.doi.org/10.1016/0009-2614(86)80661-3. URL http://www.sciencedirect.com/science/article/pii/0009261486806613.

[17] L. Brouwer, J. Troe, Thermal isomerization of azulene to naphthalene in shock waves, Int. J. Chem. Kinet. 20 (5) (1988) 379–386, http://dx.doi.org/10.1002/kin.550200504.

[18] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery, Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09 Revision D.01, gaussian Inc., Wallingford CT, 2009.

[19] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013, URL <http://www.R-project.org>.

[20] J.H. Ward, Hierarchical grouping to optimise an objective function, J. Am. Statist. Assoc. 58 (1963) 236–244.

[21] E. Paradis, J. Claude, K. Strimmer, Ape: analyses of phylogenetics and evolution in r language, Bioinformatics 20 (2) (2004) 289–290, http://dx.doi.org/10.1093/bioinformatics/btg412. arXiv:http://bioinformatics.oxfordjournals.org/content/20/2/289.full.pdf+html, URL http://bioinformatics.oxfordjournals.org/content/20/2/289.abstract.

[22] N. O'Boyle, M. Banck, C. James, C. Morley, T. Vandermeersch, G. Hutchison, Open babel: An open chemical toolbox, J. Cheminformat. 3 (1) (2011) 33, http://dx.doi.org/10.1186/1758-2946-3-33. URL http://www.jcheminf.com/content/3/1/33.

[23] D.C. McKean, Individual CH bond strengths in simple organic compounds: effects of conformation and substitution, Chem. Soc. Rev. 7 (1978) 399–422, http://dx.doi.org/10.1039/CS9780700399.

[24] D.C. McKean, CH bond dissociation energies, isolated stretching frequencies, and radical stabilization energy, Int. J. Chem. Kinet. 21 (6) (1989) 445–464, http://dx.doi.org/10.1002/kin.550210608.

[25] J.P. Merrick, D. Moran, L. Radom, An evaluation of harmonic vibrational frequency scale factors, J. Phys. Chem. A 111 (45) (2007) 11683–11700, http://

dx.doi.org/10.1021/jp073974n. arXiv:http://pubs.acs.org/doi/pdf/10.1021/jp073974n.

[26] Y. Jung, H. Park, D.-Z. Du, B. Drake, A decision criterion for the optimal number of clusters in hierarchical clustering, J. Global Optimizat. 25 (1) (2003) 91–111, http://dx.doi.org/10.1023/A:1021394316112.

[27] C. Castiglioni, M. Tommasini, G. Zerbi, Raman spectroscopy of polyconjugated molecules and materials: confinement effect in one and two dimensions, Philos. Trans. Roy. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 362 (2004) 2425–2459, http://dx.doi.org/10.1098/rsta.2004.1448.