

Conformational assignment of gas phase peptides and their H-Bonded complexes using far-IR/THz: IR-UV ion dip experiment, DFT-MD spectroscopy, and Graph Theory for modes assignment.

Daria Ruth Galimberti,^a Sana Bougueroua,^a Jérôme Mahé,^a Matteo Tommasini,^b Anouk M. Rijs,^c and Marie-Pierre Gaigeot^{*a}

The combined approach of gas phase IR-UV ion dip spectroscopy experiments and DFT-based molecular dynamics simulations for theoretical spectroscopy reveals the 3D structures of (Ac-Phe-OMe)_{1,2} peptides using the far-IR/THz signatures. Both experimental and simulated IR spectra are well-resolved in the 100-800 cm⁻¹ domain, allowing an unambiguous assignment of the conformers, that could not be achieved in other more congested spectral domains. We also present and make proofs-of-principles for our newly developed theoretical method for the assignment of (anharmonic) vibrational modes from MD simulations based on Graph Theory coupled to APT weighted-internal coordinates velocities DOS spectra. The principles of the method are reviewed, applications to the simple gas phase water and NMA (N-Methyl-Acetamide) molecules are presented, and the application to the more complex (Ac-Phe-OMe)_{1,2} peptidic systems shows that the complexity in assigning vibrational modes from MD simulations is released with the graphs. Our newly developed graph-based methodology is furthermore shown to allow an easy comparison between the vibrational modes of isolated monomer(s) and their complexes, as illustrated by the (Ac-Phe-OMe)_{1,2} peptides.

1 Introduction

Characterizing the three dimensional conformational arrangement of molecules is one of the main goals in chemical-physics and analytical chemistry, where vibrational spectroscopy is one of the key methods. A vibrational spectrum reveals the fingerprints of molecular structures, hence containing all the information about the 3D organisation. The well-known paradigm is the unique one-to-one relationship in between the spectral signatures and the 3D structure. In the gas phase, vibrational spectroscopy is one of the very few methods capable of directly probing 3D structures, and action spectroscopies such as IR-MPD (Infra Red Multiphoton Dissociation), IR-PD (Infra Red Photon Dissociation) and IR-UV ion dip, have been largely applied within the past two decades to reveal gas phase structures of molecules and clusters¹⁻¹¹. Molecules and clusters in various domains of interest

have hence been characterized with these infrared techniques, typically in relation with proteomics and metabolics,^{2,9,11-20} glycoscience,^{11,21-25} DNA science,²⁶⁻²⁹ atmospheric science and astrophysics,³⁰⁻³² to name just a few examples. Most of these experiments to date have probed the fingerprint 1000-2000 cm⁻¹ and the 3000-4000 cm⁻¹ regions.

Congestion of the spectra is a well-known issue in gas phase spectroscopy (also well-known in liquid phase spectroscopy of course) as soon as the size of the molecular system increases (congestion however already appears in rather smallish systems, see for instance ref.^{33,34}), for complex molecular structures, and for molecular systems composed of too many similar oscillators. Congestion hides the conformational details that are contained in the spectral features, hence making a definitive structural assignment hard and sometimes almost impossible^{33,34}.

To go beyond spectral congestion, vibrational action spectroscopy can typically be coupled with ion mobility^{35,36} or isotopic substitution in a '2D-experiment' that can hence provide a 'second dimension' and extra information, that ultimately might remove ambiguities for the final structural assignment. See for instance refs.^{11,24,37-39} for reviews on spectroscopy/ion mobility couplings. Another option, that we follow in the present work,

^a LAMBE UMR8587, Univ Evry, Université Paris-Saclay, CNRS, 91025 Evry, France. E-mail: mgaigeot@univ-evry.fr

^b Department of Chemistry, Materials, Chemical Engineering "G. Natta" Politecnico di Milano, 20133 Milano, Italy.

^c Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld 7c, 6525 ED Nijmegen, The Netherlands. E-mail: a.rijs@science.ru.nl

* Corresponding author.

is to probe lower frequency modes and hence gain new and less congested spectral information, that will lead to the removal of structural ambiguities. The temptation for the far-IR/THz spectral features (10-800 cm^{-1} , 24-0.3 THz) is that larger amplitude motions and possibly more collective motions are probed in this spectral domain, such as backbone torsional motions and hydrogen bond dynamics, containing a wealth of structural information, beyond the more local view given by the 1000-4000 cm^{-1} signatures. Such motions have been to date mostly employed to characterize crystals, semiconductors, liquids and biomolecules⁴⁰⁻⁴³ in the condensed phase. Far-IR/THz gas phase spectroscopy has up-to-now been developed and applied by the groups of Asmis *et al.*^{32,44-47}, Fielicke *et al.*⁴⁸⁻⁵⁴, Rijs *et al.*⁵⁵⁻⁵⁷, using Free Electron Lasers (FEL). The experiments by the Rijs group has in particular shown that well-resolved and non congested far-IR/THz spectra can be measured for biologically relevant gas phase peptides and their complexes with water molecules.⁵⁵⁻⁵⁸ This far-IR/THz IR-UV ion dip gas phase action spectroscopy experiment is the basis of the present work.

Despite all the 3D structural information are contained within the spectral features, it is not so obvious to make the correspondence from IR signatures to the 3D structure by experiments alone. The gas phase IR action spectroscopy experimental community generally relies on calculations to obtain a definitive structural assignment. To that end, static harmonic spectra calculations are performed, where a search for molecular structures is conducted, minima on the potential energy surface are found (OK), providing an energy ranking in between the conformers and isomers (electronic energy ranking or finite temperature free energy ranking), and harmonic spectra are calculated for a selection of lower energy conformers. This strategy is usually achieved at the DFT level of electronic representation. Calculated spectra are compared to the experimental spectrum, and a match is searched for one given conformer. See ref.⁵⁹ for a recent review of this theoretical strategy.

We have shown in the past decade that finite temperature DFT-based molecular dynamics simulations can be a more relevant theoretical strategy in order to extract anharmonic theoretical spectra, to be matched to the experiment for structural assignment. See for instance our two reviews on the topic^{59,60} and a short selection of relevant papers for the gas phase in refs.⁶¹⁻⁶⁵. This is the theoretical strategy that we employ in the present work, combining finite temperature DFT-MD anharmonic spectroscopy with IR-UV ion dip experiments, for the exploration of the far-IR/THz spectral features. See refs.^{55,56,58,66} for our previous successful combined approaches for gas phase dipeptides, phenol derivative molecules, including interactions with water molecules.

Employing the combined strategy of IR-UV ion dip experiment and DFT-MD theoretical IR spectroscopy, here we characterize the Ac-Phe-OMe capped amino acid and its dimeric form (Ac-Phe-OMe)₂ in the far-IR/THz domain, for two reasons. The conformational assignment of the dimeric form appears to be particularly challenging by using the classical 1000-4000 cm^{-1} range, which gives only the information about the global family structure, *i.e.* a β -Sheet structure as shown by Gerhards's group⁶⁷⁻⁶⁹.

We show here that the conformation of the dimer can be unambiguously assigned by using the far infrared signatures. Probing this domain not only removes spectral congestion but it also provides specific vibrational features that allow to unambiguously assign one 3D conformation. Another reason to characterize the (Ac-Phe-OMe)₂ dimer is that it probably is the smallest peptidic system that adopts a β -sheet structure, a common structure for large size proteins and peptides. One of our general goals in coupling IR-UV ion dip experiment with DFT-MD spectroscopic simulations is to reveal the generic spectroscopic signatures of protein motifs in the far infrared/THz domain. We have already mapped such features for (di)peptides in ref.⁵⁶. Characterizing a model β -sheet peptidic motif such as (Ac-Phe-OMe)₂ is a necessary step into this direction. β -sheet structures are also known to be involved into neurodegenerative diseases like Alzheimer, Parkinson or Huntington, having a common molecular basis related to the misfolding of proteins^{70,71}. Characterizing the structure of small but relevant β -sheet models is an important first step in the more general scheme of revealing the key spectroscopic signatures of β -sheets.

It is one success to be able to calculate a precise anharmonic IR spectrum that can be matched with the experimental spectrum, and hence reveal the 3D structure responsible for the spectral features, it is an entirely other issue to be able to extract the precise knowledge of the molecular motions that give rise to each of the signatures. This aspect is mandatory to achieve if one's goal is to map vibrational motions in the far-IR/THz domain (of here peptidic assemblies) and be able to use such mapping to make systematic structural assignment, possibly without relying on spectroscopic calculations. Refs.^{56,59,60,72} have reviewed methods to extract vibrational motions from the gas phase dynamical trajectories and hence assign motions to vibrational modes. See also ref.⁷³ for the same strategies applied in the condensed phase. Briefly, methods are usually based on VDOS (Vibrational Density of States) or ICDOS (Internal Coordinate Density of States), through Fourier transforms of time-correlation functions of velocities (VDOS)/internal coordinates (ICDOS), see refs.^{56,60}. Methods to extract effective normal modes have also been developed in the literature, see for instance refs.^{72,74-80}. Although these latter are the dynamical equivalent to the vibrational harmonic representation, they are not so easy to apply in a systematic way, therefore their (still) limited use in the literature. Bowman *et al.*^{79,80} have also developed a strategy based on the vibrational excitation of the harmonic modes, followed in time in the dynamics.

Here, we present an entirely alternate theoretical method for assigning vibrational modes, based on Graph Theory. To our best knowledge, this is the first time Graph Theory is employed to that end. This paper presents a proof-of-principle of the methodology, with two demonstrations on the isolated water molecule spectrum and on the temperature effect on the IR active modes of the isolated NMA (N-Methyl-Acetamide) molecule. We then apply the methodology to the far-IR/THz modes of (Ac-Phe-OMe) and its dimeric (Ac-Phe-OMe)₂ β -sheet form, illustrating how the Graph Theory based assignment of vibrational modes is powerful for complicated molecular systems (although still smallish in

size).

The IR-UV ion dip experiment is described in section 2.1 while the description of the finite temperature DFT-MD simulations for anharmonic spectroscopy can be found in section 2.3. The far-IR/THz coupled experiment-theory spectroscopies for characterizing the (Ac-Phe-OMe) structure are presented in section 3, showing a definitive and unambiguous conformational assignment, same in section 4 for the β -sheet model (Ac-Phe-OMe)₂. We then present our newly developed approach for revealing vibrational modes based on Graph Theory in section 5. Proofs of principle can be found for the gas phase water molecule IR spectroscopy at the end of section 5, and in section 6.1 for gas phase NMA spectra at two different temperatures. Applications of this new method to the more complex (Ac-Phe-OMe)_{1,2} peptides are in section 6.2.

2 Methods

2.1 IR-UV ion dip spectroscopy

The IR-UV ion dip spectra are recorded using the Nijmegen pulsed molecular beam set-up coupled with a time-of-flight mass spectrometer, which is discussed in detail elsewhere.^{2,58} The peptide Ac-Phe-OMe (Bachem) was used without further purification. The sample was placed in a glass container and positioned inside a heatable molecular beam source. The sample was heated to 120°C and supersonically expanded in 2 bar Helium using a pulsed valve (Series 9, Iota One; General Valve Corporation). The nozzle temperature was maintained at 128°C to prevent condensation and clogging of the nozzle. The molecular beam was skimmed about 7 cm downstream from the nozzle and subsequently perpendicularly crossed with a UV and IR laser beam to record IR and UV spectra. The peptide monomer and dimer were electronically excited and ionized using the (1 + 1)-REMPI scheme and detected at the dual microchannel plate detector of the reflector time-of-flight spectrometer.

The IR spectra of the molecules and clusters were recorded using IR-UV ion dip spectroscopy with the UV laser set to 37578.5 and 37538.5 cm⁻¹ respectively, as was reported before by Schwing et al.⁶⁷ The IR and UV beams were spatially overlapped but the IR pulse precedes the UV pulse by about 200-500 ns. The IR radiation is produced by the Free Electron Laser for Infrared eXperiments (FELIX) located at the FELIX laboratory of the Radboud University. The FELIX pulse energy is typically 80 mJ in the mid-IR to 20 mJ in the far-IR region of the spectrum with a pulse duration of 8 to 10 microseconds. All IR spectra were corrected for the wavelength-dependent photon flux of the IR light sources. In order to compensate for long term fluctuations in sample density and UV power, alternating IR-on and IR-off ion signals were measured by operating the IR laser at a repetition rate of 10 Hz and the molecular beam and the UV laser at 20 Hz. Since the experiments are performed over a very wide frequency range, all IR spectra were corrected for the wavelength-dependent photon flux.

2.2 Conformational search and harmonic spectroscopy

An initial conformational search for the Ac-Phe-OMe monomer has been performed through a 100 ps molecular dynamics at the semi empirical PM6 representation⁸¹ (for fast enough samplings) at 1000 K (high enough temperature for a sufficiently exhaustive PES sampling, without possible issues in passing the barriers in between conformers/isomers). The time step was 0.4 fs. Hundred conformations have been extracted and optimized at the BLYP-D3/6-311+G(d,p) level using the Gaussian package⁸². 21 different structures of Ac-Phe-OMe remained after geometry optimizations. The electronic energies (ΔE) and Gibbs free energies (ΔG) of the five lower energy conformers have been calculated with the BLYP-D3/6-311+G(d,p) and B3LYP-D3/6-311+G(d,p) DFT electronic representations⁸², and will be presented in more details in Table 1 in section 3. A cut-off for the free energy difference in between the final optimized structures has been arbitrary fixed around 2 kcal.mol⁻¹ to select the conformations for further investigations, i.e. either for static harmonic spectra calculations or for the DFT-MD simulations of the anharmonic spectra. Static harmonic spectra of the five lower energy conformers have been obtained with the Gaussian package. These spectra are mainly used for the discussion on spectral congestion in section 3. The construction of the (Ac-Phe-OMe)₂ dimer follows a different strategy, described in section 4.

2.3 DFT-MD simulations and DFT-MD theoretical spectroscopy

The DFT-MD simulations performed here follow our general schemes, see e.g. refs.^{56,59} for more details. The trajectories are accumulated with the CP2K package⁸³, all analyzes are performed with our own codes. The trajectories consist in Born-Oppenheimer molecular dynamics in the NVE ensemble, with time-steps of 0.4 fs. The electronic representation is the BLYP-D3⁸⁴⁻⁸⁶, the dual basis set representation used in CP2K consists here in the 450 Ry energy cutoff plane wave basis set coupled with the aug-TZV2P gaussian basis set. Pseudopotentials of the GTH type (Goedecker-Teter-Hutter)⁸⁷ are used. The size of the cubic box is systematically optimised to ensure that there is no interaction between the system and its 3D replica (periodic boundary conditions are indeed systematically used in our simulations). Trajectories are done at 50 K, see for instance refs.^{55,56} for justification of this temperature with respect to the experimental conditions.

The initial structures for the dynamics are the ones from the static calculations reported in section 2.2 and re-optimised at the level of theory used in the dynamics. The DFT-MD infrared spectra of Ac-Phe-OMe and of its dimeric (Ac-Phe-OMe)₂ form are obtained as an average over three separate 20 ps trajectories, thus over 60 ps. The same initial geometry is used for the three trajectories (the optimized one), while a Boltzmann distribution ensures sampling of different initial atomic velocities within the three trajectories.

Within the well-known time-correlation function formalism in linear response theory^{88,89}, an IR absorption spectrum is calcu-

lated by:

$$\begin{aligned}
 I(\omega) &= \frac{2\pi\beta\omega^2}{3cV} \int_{-\infty}^{+\infty} dt e^{i\omega t} \langle \delta\mu(t) \cdot \delta\mu(0) \rangle \\
 &= \frac{2\pi\beta}{3cV} \int_{-\infty}^{+\infty} dt e^{i\omega t} \langle \delta\left(\frac{d\mu(t)}{dt}\right) \cdot \delta\left(\frac{d\mu(0)}{dt}\right) \rangle \quad (1)
 \end{aligned}$$

where $\beta = 1/kT$, ω is the frequency of the absorbed light, c is the speed of light in vacuum, V the volume of the system, $\mu(t)$ is the instantaneous dipole moment vector of the system at the time t , $\frac{d\mu(t)}{dt}$ its time derivative, $\delta\mu(t) = \mu(t) - \langle \mu \rangle$ is the fluctuation with respect to the mean value and $\langle \dots \rangle$ refers to the equilibrium time correlation function. The quantum correction factor $\beta\hbar/(1 - \exp(-\beta\hbar\omega))$ has been applied to correct the classical line shape⁹⁰ in equation 1.

We refer the reader to our reviews in refs.^{59,60} and to refs.^{91,92} for summaries of the advantages and limitations of dynamical anharmonic spectra calculations over static harmonic/anharmonic spectra calculations. This is not the purpose of this discussion.

3 Far IR/THz spectroscopy and conformational assignment of Ac-Phe-OMe

The chemical scheme for Ac-Phe-OMe is presented in figure 1, the caps 'Ac' and 'OMe' are respectively the acetyl $\text{CH}_3\text{-CO}$ and the O-CH_3 groups. The labeling of the dihedral angles in the figure will be used for the structural discussions in the text, and also for the modes assignment in section 6.2.

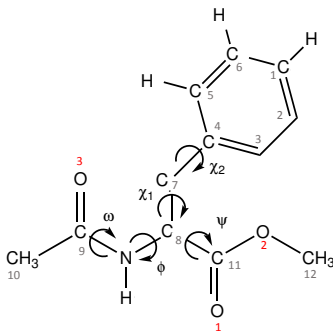


Fig. 1 Chemical scheme of the Ac-Phe-OMe monomer with the labels used in the text for the atoms and the definition of dihedral angles (ϕ , ψ , ω and χ) from the Ramachandran notation⁹³. The caps 'Ac' and 'OMe' correspond respectively to an acetyl group $\text{CH}_3\text{-CO}$ and to a O-CH_3 group.

To illustrate the conformational selection of the IR-UV ion dip experiment, figure 2 presents the REMPI spectrum of the gas phase Ac-Phe-OMe molecule. The spectrum is nearly identical to the one published by Gerhards and Unterberg⁶⁷, only the peaks relative intensities are different in our two works. The IR-UV ion dip infrared spectrum at the top of fig. 3 (blue line) has been measured via the $S_0 \rightarrow S_1$ transition at 37578.5 cm^{-1} , corresponding to the dominant conformation of the monomer in the UV spectrum (spectrum in purple) and using the same transition as in ref.⁶⁷. This is the only conformer of Ac-Phe-OMe

investigated in this work and in ref.⁶⁷.

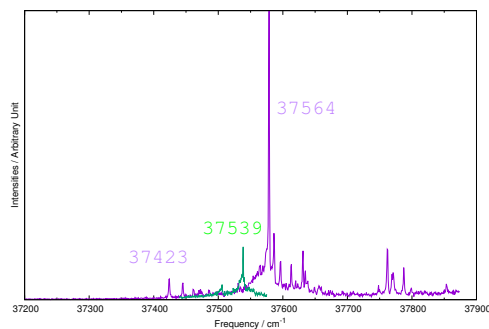


Fig. 2 REMPI spectra of the Ac-Phe-OMe monomer (in purple) and its dimeric form $(\text{Ac-Phe-OMe})_2$ (in green).

The far-IR/THz IR-UV ion dip experimental spectrum of the Ac-Phe-OMe gas phase peptide ($100\text{-}800 \text{ cm}^{-1}$) is presented at the top of figure 3 (in blue), together with dynamical DFT-MD spectra (averaged over three trajectories of 20 ps each, as explained in the method section) for three conformers ($\beta_L(g^+)$, $\beta_L(a)$, $\beta_L(g^-)$, see their 3D representations in fig. 5). Before discussing the dynamical spectra and the associated structural assignment in the far IR/THz domain, we briefly discuss the optimized conformations of Ac-Phe-OMe and the impossibility to assign a structure based on the $1000\text{-}2000 \text{ cm}^{-1}$ fingerprint region.

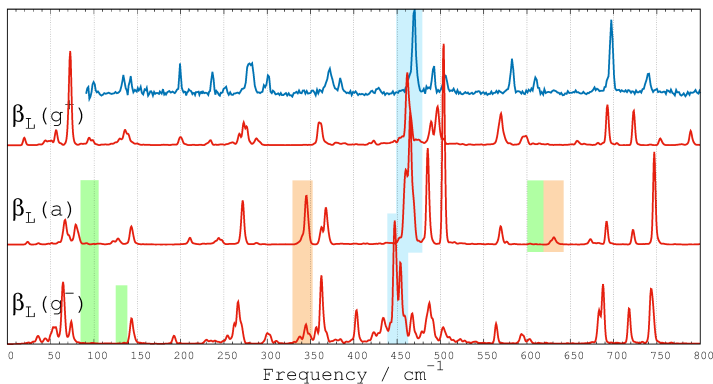


Fig. 3 Experimental IR-UV ion dip spectrum of the Ac-Phe-OMe gas phase monomer (in blue) compared with the calculated dynamical DFT-MD IR spectra (in red) for the three monomeric isomers $\beta_L(g^+)$, $\beta_L(a)$, and $\beta_L(g^-)$ in the far infrared/THz domain. The green boxes highlight missing spectral features in the theoretical spectra that are however present in the experimental one. Conversely, the orange boxes highlight extra spectral features in the theoretical spectrum that are absent in the experimental one. The blue boxes show the assignment of the $\omega(\text{N-H})_{\text{phe}}$ wagging motion.

Five conformers of lower energy have been identified at the BLYP and B3LYP levels of calculation, see Table 1 for the relative energies and names of the structures. See section 2 for the strategy employed for the structural exploration. We remind that a cut-off for the free energy difference has been arbitrary fixed around 2 kcal.mol^{-1} to select the conformations for further investigations in this work. Note that the next conformer in terms of

free energy lies at $3.07 \text{ kcal.mol}^{-1}$ at the BLYP-D3/6-311+G(d,p) level.

Ac-Phe-OMe				
BLYP-D3/6-311+G(d,p)		B3LYP-D3/6-311+G(d,p)		Structure
ΔE	ΔG 50K	ΔE	ΔG 50K	
0	0	0	0	$\beta_L(g^+)$
0.99	0.93	1.17	1.09	$\beta_L(g^a)$
1.13	1.10	1.31	1.25	$\delta_D(g^+)$
1.66	1.68	1.75	1.78	$\gamma_L(g^+)$
2.02	1.71	2.18	1.91	$\beta_L(g^-)$

Table 1 Relative electronic energies (ΔE) and free energies (ΔG) for the five lower energy conformers of Ac-Phe-OMe identified in this work (energies are in kcal.mol^{-1}). ΔG is calculated at 50 K, and the results are given for two levels of DFT electronic representations.

The two lower energy conformers are $\beta_L(g^+)$ and $\beta_L(a)$ with $\sim 1 \text{ kcal/mol}$ difference in energy (ΔG and ΔE , BLYP and B3LYP electronic representations). β_L refers to the backbone amide ϕ and ψ angles (see fig. 1) within the following ranges in the Ramachandran plot, $\phi \sim [-120^\circ, +120^\circ]$, $\psi \sim [-120^\circ, +120^\circ]$, corresponding to a β -sheet type organisation of the backbone. There is a C5 interaction between the N-H_{Phe} and the C=O amide backbone groups, i.e. a weak N-H_{Phe}...O=C hydrogen bond interaction leading to the formation of an internal five membered ring between these two groups. g^+ , a and g^- refer to the orientation of the Phenylalanine (Phe) aromatic ring with respect to the backbone, defined by the dihedral angle χ_1 (see fig. 1, $\chi_1 = \text{N-C}\alpha\text{-C}\beta\text{-C}$): $\chi_1 = 60^\circ$ for $\beta_L(g^+)$, 180° for $\beta_L(a)$ and -60° for $\beta_L(g^-)$. This latter $\beta_L(g^-)$ conformation is found roughly 2 kcal/mol (ΔG and ΔE) higher in energy than the $\beta_L(g^+)$ lowest energy conformer. Sandwiched inbetween the three β_L type conformers of Ac-Phe-OMe, one can find other symmetries of the δ_D and γ_L type, i.e. with ϕ and ψ angles in the ranges $[-120^\circ, +120^\circ]$ & $[-120^\circ, 0^\circ]$ for δ_D and $[-120^\circ, 0^\circ]$ & $[0^\circ, +120^\circ]$ for γ_L (γ -turn). These structures also have an internal N-H_{Phe}...O=C C5 interaction. In the previous work by Gerhards and coworkers⁶⁷, only the two lower energy conformers $\beta_L(g^+)$ and $\beta_L(a)$ have been further investigated for the theoretical spectroscopy (the same list of conformations as here has been found by these authors), under the assumption that only the lower energy conformers could be formed in the experiment. Note that the relative energies found here are similar to the ones in ref.⁶⁷

Previous spectroscopic results from Gerhard's group^{67,68} show that none of the signatures in the 3000-4000 and 1450-1800 cm^{-1} ranges are able to distinguish the two lower energy conformations $\beta_L(g^+)$ and $\beta_L(a)$, while their latest published results in the fingerprint region indicate that the 1000-1450 cm^{-1} might be conformer selective⁶⁹. To illustrate this issue, figure 4 shows the comparison between experiment (remeasured in the present work, black lines) and (scaled) static harmonic spectra (B3LYP-D3/6-311+G(d,p) level, colored lines, one color per conformer) in the 1000-2000 cm^{-1} fingerprint region for the five optimized structures of Ac-Phe-OMe (table 1). One can see that the five conformers provide very similar signatures in the 1400-1800 cm^{-1} while the 1100-1400 cm^{-1} signatures offer slightly more conformer selectivity. In particular, the experimental broad peak at

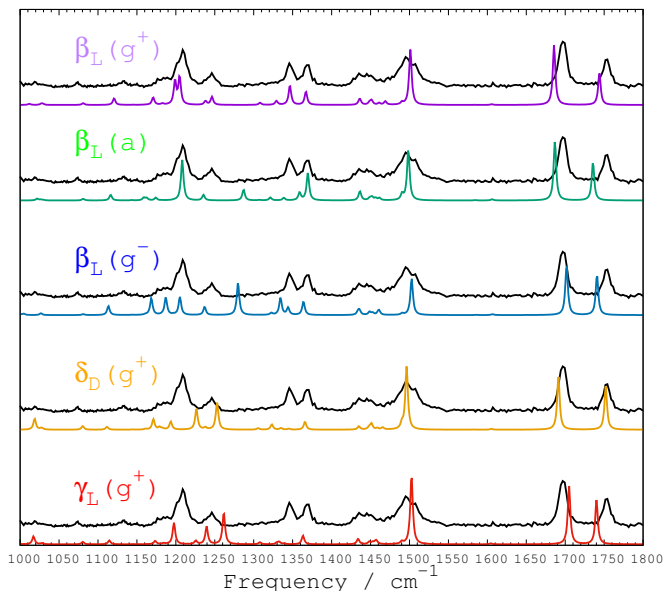


Fig. 4 Experimental IR-UV ion dip spectra of the Ac-Phe-OMe capped amino acid (in black) compared with the calculated static harmonic B3LYP-D3/6-311+G(d,p) spectra (in color). The theoretical spectra are scaled by 0.967 and a 1% gaussian convolution is applied.

1209 cm^{-1} and the smaller peak at 1246 cm^{-1} provide a clear marker in favour of assigning the two lower energy conformers $\beta_L(g^+)$ and $\beta_L(a)$, with $\beta_L(g^+)$ maybe providing a global better agreement with experiment. These bands are due to the Amide III vibrational motions, i.e. modes composed of CCH and CNH bendings (coupled to CC, CO and CN stretchings). All other conformers, including the $\beta_L(g^-)$ conformer from the same β -sheet family, have too many dispersed signatures in this domain, not matching the experimental patterns.

The three optimized β_L -type structures are presented in figure 5, where one can see that the C5 H-bonded N-H_{Phe} and C=O groups are located roughly above the Phe ring (but not interacting with it) in the $\beta_L(g^+)$ structure, that only the C=O is located above the Phe ring in $\beta_L(a)$, while both N-H_{Phe} and C=O groups are pointing outwards the Phe ring in $\beta_L(g^-)$.

According to the above spectroscopic discussion in the 1000-2000 cm^{-1} fingerprint region, one expects $\beta_L(g^+)$ and $\beta_L(a)$ to be good conformer candidates. For the DFT-MD theoretical spectroscopy in the far IR/THz domain, we hence chose to unravel the signatures of these two conformers, as well as for those of the third $\beta_L(g^-)$ beta-type conformer for the sake of comparison. See section 2 for all details related to the dynamics and to the calculation of the dynamical anharmonic IR spectra. See sections 2 and 5 for the details on the bands assignments using the newly developed Graph Theory approach, and please refer to section 6.2 for graphs for a limited selection of bands/modes. Hereafter, we do not detail the methodology, we only use the results on the modes and comment accordingly.

As shown in our previous works on gas phase far-IR/THz spectra, the dynamical DFT-MD spectra provide such agreement with experimental signatures that it is possible to unambiguously assign the 3D conformations^{55,56,66,94}. The comparison between

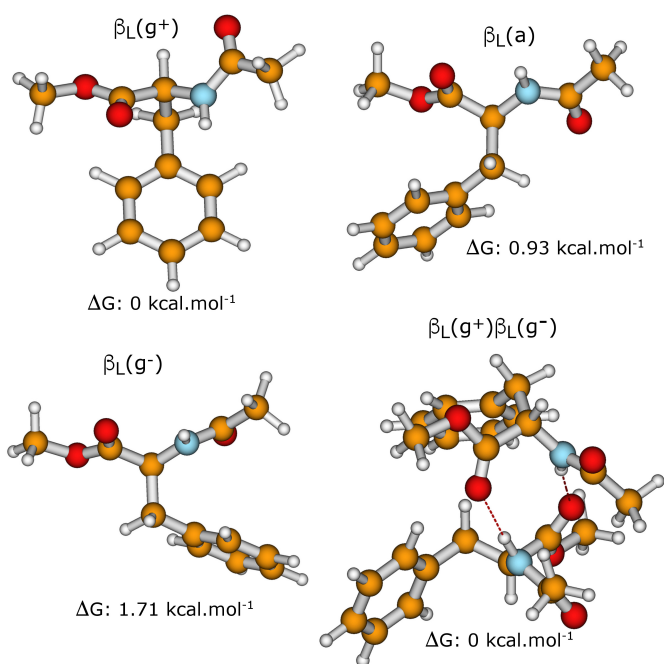


Fig. 5 Ac-Phe-OMe and (Ac-Phe-OMe)₂ peptides optimized geometries in the β_L general structural organisation. The " $\beta_L(g^+)$ " conformation corresponds to the assigned one for the monomer (in our experimental conditions). The " $\beta_L(g^+)$ - $\beta_L(g^-)$ " dimer structure corresponds to the assigned one (in our experimental conditions). Free energies have been calculated at 50K at the BLYP-D3/6-311+G(d,p) level, see Tables 1-2. See also fig. 1.

the IR-UV ioin dip experiment of Ac-Phe-OMe and the dynamical DFT-MD theoretical spectra calculated for the three β_L conformers is presented in figure 3 (100-800 cm^{-1} , experiment in blue, dynamical DFT-MD spectra in red). Based on the number and absolute positions of the theoretical bands (positions are shifted by 7 cm^{-1} on average from experiment), and on the shapes and relative intensities of the bands (both coming from the dynamics, no *a posteriori* model applied), one can conclude that $\beta_L(g^+)$ is the conformation formed in the experiment. Especially well given by the theoretical spectrum are the 680-750 cm^{-1} ($\omega(\text{C-H})_{\text{Ring}}$ waggings) and 550-620 cm^{-1} double bands (backbone and side-chain bending motions around the C_α link), the 450-510 cm^{-1} massif ($\omega(\text{N-H})_{\text{Phe}}$ wagging). Same very good agreement below 400 cm^{-1} , with however some theoretical bands missing some intensity. This latter is well-known at low temperature trajectories, where equipartition of energy is an issue, see related discussions in ref.⁵⁹. The 260-300 cm^{-1} double band massif (coupled bending motions along the backbone and around C_α - C_β -C side chain) is remarkably well given by the theoretical spectrum of $\beta_L(g^+)$ (the higher frequency band is slightly shifted in position), while the two other conformations display different active band-shapes in this domain. Note also that the theoretical band-widths provide as well defined and resolved bands as in the experiment in the whole far-IR domain.

Highlighted in green in figure 3, one can see missing peaks (or peaks too much shifted in position) in the theoretical spectra that are present in the experimental spectrum, in disfavour of $\beta_L(a)$

and $\beta_L(g^-)$ conformational assignment. We hence observe no counterpart for the experimental peak at 135 cm^{-1} in the $\beta_L(g^-)$ theoretical DFT-MD spectrum. For the experimental peak at 100 cm^{-1} neither $\beta_L(a)$ nor $\beta_L(g^-)$ provide counterpart. Highlighted in orange in figure 3 are extra peaks present in the theoretical spectra of $\beta_L(a)$ and $\beta_L(g^-)$ conformations that are absent in the experimental spectrum. For both $\beta_L(a)$ and $\beta_L(g^-)$ a theoretical peak at $\approx 345 \text{ cm}^{-1}$ finds no counterpart in the experiment. Also, the experimental peak at 710 cm^{-1} is shifted at $\approx 725 \text{ cm}^{-1}$ in the $\beta_L(a)$ theoretical spectrum. We do not observe the same drawbacks for $\beta_L(g^+)$ for which the DFT-MD spectrum definitely provide the best agreement with the experiment spectrum.

In blue in figure 3, we have highlighted the $\omega(\text{N-H})_{\text{Phe}}$ wagging motion band of the N-H_{Phe} amide. Position of this band is identical for $\beta_L(g^+)$ and $\beta_L(a)$ conformers (respectively located at 463 and 465 cm^{-1} , vs 471 cm^{-1} in the experiment), but the band shape for $\beta_L(g^+)$ is remarkably in even better agreement with the experiment. There is one band per NH group for small peptides (only one group for Ac-Phe-OMe), see ref.⁵⁶ where we discussed this point, and the position obtained here for the three conformers is compatible with the range found in ref.⁵⁶ for a N-H group involved in a C5 weak interaction. The lower frequency observed for $\beta_L(g^-)$ (451 cm^{-1}) is also in line with our previous work⁵⁶, i.e. lower frequencies for N-H groups being more free of intermolecular interactions. Here the $\beta_L(g^-)$ conformer is the one with the longer C5 H-bond, i.e. $2.51 \pm 0.12 \text{ \AA}$ to be compared to 2.30 - $2.40 \pm 0.07 \text{ \AA}$ for the two other β_L conformers, thus providing this lower frequency of the N-H_{Phe} wagging motion. The $\omega(\text{N-H})_{\text{Phe}}$ wagging is part of a more complex massif, which shape- and band-positions are in remarkable agreement with the experiment for $\beta_L(g^+)$, these other bands being due to backbone bending motions around the amide N-H_{Phe} group.

As already discussed in refs.^{55,56,94}, the 90-400 cm^{-1} range appears to be more conformer selective than the 400-800 cm^{-1} . In that respect, the theoretical spectrum of $\beta_L(g^+)$ provides a remarkable agreement with the experiment. All bands in this domain are due to bending motions delocalised along the backbone and along the side chain. Due to their intrinsic nature, these motions are sensitive to the 3D-structure of the peptide, thus providing the means to distinguish between different conformers. Even more delocalised and larger amplitude dihedral motions start to kick-in at roughly 150 cm^{-1} and these motions are systematically dominant in all peaks located below 100 cm^{-1} in our theoretical spectra. Of special interest is the 100 cm^{-1} peak located in experiment and in the theoretical spectrum of $\beta_L(g^+)$, absent in the spectrum of the two other conformers, that records rather highly delocalised bending motions along the backbone and the side chain of Ac-Phe-OMe (with some torsions also participating). This is the signature of the relative angular motion of the side-chain versus the backbone, corresponding to rather large-amplitude motions. Equivalent motions, but far less delocalised over the whole molecule, are seen for the two other conformers of Ac-Phe-OMe at higher frequencies. The 50-80 cm^{-1} bands obtained in the spectrum of $\beta_L(g^+)$ are due to delocalized torsional motions along the whole backbone while the $\sim 25 \text{ cm}^{-1}$ band also contains torsions of the side chain around χ_1 and χ_2 dihedrals.

These bands cannot be recorded yet at the FELIX beamline.

4 Far IR/THz spectroscopy and conformational assignment of (Ac-Phe-OMe)₂ β -sheet model

For the construction of the (Ac-Phe-OMe)₂ β -sheet structure, considering the results obtained for the Ac-Phe-OMe monomer, one would expect a conformation in the form of $\beta_L(g^+)$ - $\beta_L(g^+)$ under the assumption that the dimer is solely assembled from $\beta_L(g^+)$ monomers. Such assumption does not take into account that, while they assemble to form a dimer, the monomers could however isomerise into a different β_L orientation as a result of intermolecular interactions. Alternatively, all possible combinations of dimers might be produced during the initial stage of the experiment and only the lowest free energy dimer will remain once passing through the supersonic expansion, following thermodynamics rules. Looking at the relative electronic and free energies of the six possible (Ac-Phe-OMe)₂ β -sheet conformations built on β_L orientations of the monomers (Table 2), one can see that the $\beta_L(g^+)$ - $\beta_L(g^-)$ and $\beta_L(g^+)$ - $\beta_L(g^+)$ conformers are quasi-isoenergetic, while the $\beta_L(g^+)$ - $\beta_L(a)$ conformer is only ~ 1.1 - 1.5 kcal/mol higher in energy. All six possible conformers are found within less than 3 kcal/mol of energy. No search for energy barriers has been conducted here. Note that the pio-

BLYP-D3/6-311+G(d,p)		B3LYP-D3/6-311+G(d,p)		Structure
ΔE	ΔG 50K	ΔE	ΔG 50K	
0	0	0	0	$\beta_L(g^+)$ - $\beta_L(g^-)$
0.68	0.36	0.68	0.36	$\beta_L(g^+)$ - $\beta_L(g^+)$
1.54	1.11	1.49	1.14	$\beta_L(g^+)$ - $\beta_L(a)$
2.80	2.26	2.73	2.31	$\beta_L(g^-)$ - $\beta_L(a)$
3.07	2.45	2.87	2.39	$\beta_L(g^-)$ - $\beta_L(g^-)$
3.37	2.70	3.37	2.77	$\beta_L(a)$ - $\beta_L(a)$

Table 2 Relative electronic energies (ΔE) and free energies (ΔG) for six conformers of (Ac-Phe-OMe)₂ β -sheet built on β_L -types structures of the monomers (energies in kcal.mol⁻¹). ΔG is calculated at 50 K, and the results are given for two levels of DFT electronic representations.

neering work of Gerhards and coworkers⁶⁷⁻⁶⁹ took only the symmetrical structures $\beta_L(g^+)$ - $\beta_L(g^+)$ and $\beta_L(a)$ - $\beta_L(a)$ into account, because the $\beta_L(g^+)$ and $\beta_L(a)$ monomeric conformations were found lower in energy.

We here investigate all the 6 possible β_L - β_L structures through DFT-MD simulations and figure 6 compares the six dynamical far IR/THz spectra (red lines) with the IR-UV ion dip experiment (blue line). For simplicity of notation in this figure and in the following text and figures, we sometimes label the conformers without the β_L notation (e.g. g^+ for $\beta_L(g^+)$). As for Ac-Phe-OMe, each theoretical spectrum is generated as an average over three separate trajectories (each one with different initial conditions for the dynamics). We focus here only on the 100-400 cm⁻¹ spectral signatures, as the monomer spectroscopy showed that this is the domain the most conformational selective below 800 cm⁻¹.

Analysing the number of bands in the 100-400 cm⁻¹ domain, their absolute and relative positions, their band-shapes and in-

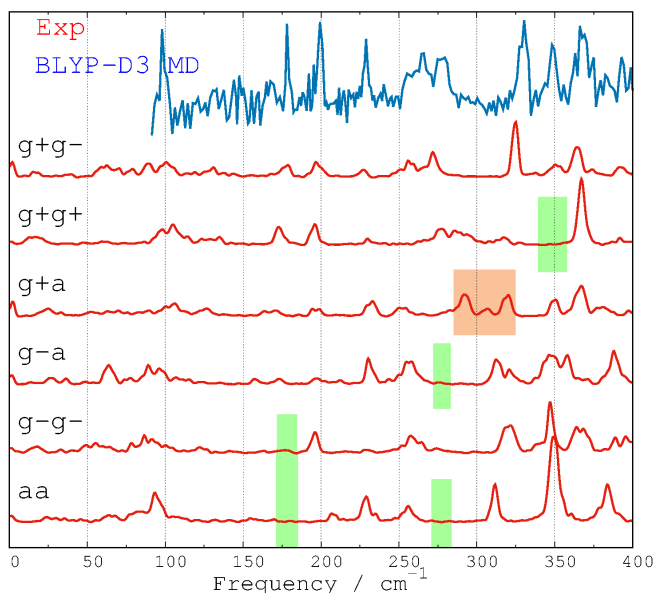


Fig. 6 Experimental IR-UV ion dip spectrum of the gas phase (Ac-Phe-OMe)₂ β -sheet (in blue) compared with the calculated dynamical DFT-MD IR spectra (in red) of six possible isomers of the gas phase (Ac-Phe-OMe)₂. The green boxes highlight missing spectral features in the theoretical spectra that are however present in the experimental spectrum. The orange boxes highlight extra spectral features that are present in the theoretical spectra but are absent in the experimental spectrum.

tensities with respect to the well resolved bands in the experimental spectrum, one can conclude that the DFT-MD theoretical spectrum of the $\beta_L(g^+)$ - $\beta_L(g^-)$ isomer of the gas phase (Ac-Phe-OMe)₂ β -sheet structure provides a precise account of the experimental features. The 300-400 cm⁻¹ triple-bands are in excellent agreement with the experiment (one however note that the middle-band intensity appears too low), the 250-300 cm⁻¹ double-bands is not present with such accuracy in any other isomer (band-positions, shapes and intensities), the 150-250 cm⁻¹ triple-bands are just located on-spot (although too low in absolute intensities), isomer $\beta_L(g^+)$ - $\beta_L(g^+)$ being the only other isomer displaying the same triple-bands. The 100 cm⁻¹ band seems less conformer-selective as all six isomers spectra display this particular band, by a few cm⁻¹ from one isomer to the other.

Similarly to fig. 3, we have highlighted the bands in fig. 6 that are missing in the theoretical spectra in order to match the experiment (green boxes) and the bands that are appearing in the theoretical spectra but are not present at these particular frequencies in the experiment (orange boxes). Strikingly, only the theoretical spectrum of $\beta_L(g^+)$ - $\beta_L(g^-)$ does not have any of these boxes highlighted. We hence unambiguously assign isomer $\beta_L(g^+)$ - $\beta_L(g^-)$ to the (Ac-Phe-OMe)₂ β -sheet, which is also the lowest energy conformer.

Once again, we found that the 100-400 cm⁻¹ range is the most conformational selective. This is due to the nature of the vibrational modes in this spectral range. We find delocalised modes over the whole backbone and side chain residue, mainly due to the combination of bendings and dihedrals motions. In particular, we found the dihedral angle χ_1 (fig. 1) involved in most of the

modes below 200 cm^{-1} . This internal coordinate directly probes the rotation of the phenyl ring (rotation that would induce conformational transitions in between g^+ , a and g^- orientations) and we hence expect the signatures of this internal coordinate to be indeed relevant for conformational identification.

5 Graph Theory for vibrational spectroscopy assignment

As already emphasized in the introduction of this paper, it is one achievement to calculate (anharmonic) vibrational spectra, it is another task to assign the vibrational bands to atomic motions and hence reveal the nature of the movements that give rise to active IR bands. We and others have in the past developed theoretical methods to that end^{72,74–80}. Here we develop a new theoretical strategy that combines our work⁹⁵ where we demonstrated that dynamical IR spectra can be calculated through velocities time correlations functions weighted by Atomic Polar Tensors (APT) with other works^{56,94} in which the ICDOS (Internal Coordinates Density of States) were used to assign the motions behind the IR bands. The disadvantage of the ICDOS is that there is no activity taken into account, i.e. the peaks intensities cannot be used for the final interpretation of any IR active band in terms of relative participation of the individual motions into the final motion responsible for the activity of the band. Therefore the need for another methodology, which we propose here to be based on the combination of ICDOS with the APTs, that we explain below.

By doing this, one is able to reconstruct any IR spectrum as a combination of internal coordinates (ICs), simultaneously extracting the knowledge of the intra- and inter-molecular motions involved in the modes and their percentage of participation into the activity of the modes. To achieve this, Graph Theory from theoretical informatics is used, automatically assigning the IR bands and providing the graph of connectivity in between the ICs responsible for the IR activity, including their percentage of participation into the final IR band. The different elements of the theory are described in the following.

5.1 Time-correlation formalism for IR spectroscopy in molecular dynamics

We refer the reader to section 2.3 of this paper for the general equation (eq. 1) from linear response theory for the calculation of IR dynamical anharmonic spectra from (DFT-)MD simulations. In ref.⁹⁵, we have shown how to go from eq. 1 to an equation based on the Fourier transform of the time-correlation function of the cartesian atomic velocities weighted by Atomic Polar Tensors (APT). These latter contain the activity of the vibrational modes. This was obtained by expanding the time derivative dipole moment vector into cartesian coordinates derivatives. Here we choose to expand the time derivative dipole moment vector into internal coordinates (ICs) labeled hereafter R_m (for any coordinate that belongs to the $3N - 6$ set of non redundant internal coordinates vector $R = [R_1, R_2, R_3, \dots, R_m, \dots, R_{3N-6}]^T$, N is the

total number of atoms in the molecular system):

$$\frac{d\mu(t)}{dt} = \sum_{m=1}^{3N-6} \frac{\partial \mu(t)}{\partial R_m} \cdot \frac{\partial R_m(t)}{\partial t} = \sum_{m=1}^{3N-6} \frac{\partial \mu(t)}{\partial R_m} \cdot \dot{R}_m(t) \quad (2)$$

With this, one can rewrite equation 1 into:

$$I(\omega) = \frac{2\pi\beta}{3cV} \sum_m \sum_l \int_{-\infty}^{+\infty} dt e^{i\omega t} \left\langle \frac{\partial \mu(t)}{\partial R_m} \cdot \dot{R}_m(t) \frac{\partial \mu(0)}{\partial R_l} \cdot \dot{R}_l(0) \right\rangle \quad (3)$$

where \dot{R}_l is the velocity of the l^{th} IC. This equation is the mirror to the one established in ref.⁹⁵ in terms of cartesian coordinates. The IR spectrum is calculated through the Fourier transform of the time-correlation function of the velocities of the internal coordinates modulated by derivatives of the dipole moment with respect to the internal coordinates, which are the equivalent of the APTs in cartesian coordinates within a transformation. The reason to use here ICs instead of cartesian coordinates is that our goal is to ultimately provide the decomposition of each IR active band in terms of intra- and inter-molecular coordinates, the ICs being the natural coordinates to use to achieve such goal.

For vibrational small displacements, the $\frac{\partial \mu(t)}{\partial R_m}$ elements can be considered constant with time, equation 3 can hence be rewritten as:

$$I(\omega) = \frac{2\pi\beta}{3cV} \sum_m \sum_l \frac{\partial \mu}{\partial R_m} \frac{\partial \mu}{\partial R_l} \int_{-\infty}^{+\infty} dt e^{i\omega t} \langle \dot{R}_m(t) \cdot \dot{R}_l(0) \rangle \quad (4)$$

One important issue in equation 4, already discussed in ref.⁹⁵ for the equivalent equation expressed in cartesian coordinates, is that the final IR intensity $I(\omega)$ includes all the self- and cross-correlations in between all internal coordinates. All motions in the dynamics are correlated, which is of course maintained in the calculation of the IR spectrum through the dipole moment in eq. 1, and is still maintained once the dipole moment is expanded into the ICs contribution in eq. 4.

One can hence rewrite eq. 4 into two sums, one related to the self-correlation contributions ($\langle \dot{R}_m(t) \cdot \dot{R}_m(0) \rangle$, $\forall m = 1, \dots, 3N - 6$), and one to the cross-correlation contributions ($\langle \dot{R}_m(t) \cdot \dot{R}_l(0) \rangle$, $\forall m \neq l$):

$$I(\omega) = \frac{2\pi\beta}{3cV} \sum_m \frac{\partial \mu}{\partial R_m} \frac{\partial \mu}{\partial R_m} \int_{-\infty}^{+\infty} dt e^{i\omega t} \langle \dot{R}_m(t) \cdot \dot{R}_m(0) \rangle + \frac{2\pi\beta}{3cV} \sum_m \sum_{l \neq m} \frac{\partial \mu}{\partial R_m} \frac{\partial \mu}{\partial R_l} \int_{-\infty}^{+\infty} dt e^{i\omega t} \langle \dot{R}_m(t) \cdot \dot{R}_l(0) \rangle \quad (5)$$

All these equations involve $\frac{\partial \mu}{\partial R_m}$ components, while the APTs (the equivalent derivative with respect to the atomic cartesian coordinates) are defined, by construction, in cartesian coordinates. APTs can be calculated from standard quantum chemical packages such as Gaussian⁸² largely used in the gas phase community, one has therefore to transform APTs into their equivalent in internal coordinates for eq. 5 to be used. We now introduce the vector $\xi = [x_1, y_1, z_1, x_2, \dots, z_N]^T$ that collects the $3N$ cartesian displacements of the N atoms of the system. It is always possible to find a linear transformation such as:

$$R = B \cdot \xi \quad (6)$$

$$\frac{\partial \mu_u}{\partial R_i} = \sum_j A_{ji} \frac{\partial \mu_u}{\partial \xi_j} \quad (7)$$

where $B_{ij} = \frac{\partial R_i}{\partial \xi_j}$ and $A_{ji} = \frac{\partial \xi_j}{\partial R_i}$, and μ_u is the u^{th} cartesian coordinate of the dipole moment vector μ . We adopted the Wilson definitions of internal coordinates⁹⁶ to describe the vibrational subspace and the transformation A and B matrices, defining the relationships between internal (R) and cartesian displacements coordinates (ξ).

With eq. 7 one can express APT components, defined in cartesian coordinates, into the internal coordinates needed for the bands assignments. Please also refer to ref.⁹⁵ for the schemes tested and the most robust adopted one for adapting the APTs to the change in conformations along the trajectory. Note that, for the sake of simplicity in this work, all applications shown in section 6 have no isomerisation occurring over the trajectories, therefore only one APT is needed, see ref.⁹⁵. In this work, APTs are calculated with the Gaussian code⁸², at the BLYP-D3/6-311++G** level for consistency with the DFT-MD electronic representation. We also refer the reader to ref.⁹⁷ for an automatic recognition of structures along the trajectories achieved through Graph Theory algorithms.

Velocities of the internal coordinates are obtained by numerical derivation with a five points central difference algorithm:

$$\dot{R}(t) = \frac{-R(t+2\delta t) + 8R(t+\delta t) - 8R(t-\delta t) + R(t-2\delta t)}{12\delta t} \quad (8)$$

This algorithm is a good compromise between numerical accuracy and computational time, with the rather short $\delta t=0.4$ fs time-scale from the DFT-MD trajectories.

5.2 Reconstruction of the IR spectrum from individual ICs active spectra

The basic idea used here is that the dynamical IR spectrum of any molecular system given by eq. 1 can equivalently be obtained through eq. 5, once the ICs, APTs and the transforms in eqs. 6-7 are calculated. Eq. 5 contains the IR activity of the bands thanks to the APTs. One can use a decomposition of the IR spectrum into components arising from each IC (self-part in eq. 5) and arising from its correlation with any other IC (cross-part in eq. 5), hence keeping the IR activity of each of these components into the decomposition of $I(\omega)$.

With these considerations in mind, in practice, the dynamical $I(\omega)$ IR spectrum obtained by eq. 1 can be reconstructed by the spectral components of ICs $I_{IC_{ml}}(\omega) = \frac{2\pi\beta}{3cV} \frac{\partial \mu}{\partial R_m} \frac{\partial \mu}{\partial R_l} \int_{-\infty}^{+\infty} dt e^{i\omega t} \langle \dot{R}_m(t) \cdot \dot{R}_l(0) \rangle$. For each band of the $I(\omega)$ spectrum, one has:

$$IR(\omega_x) = \sum_{m,l} I_{IC_{ml}}(\omega_x) = \sum_m I_{IC_m}(\omega_x) + \sum_{m,l \neq m} I_{IC_{ml}}(\omega_x) \quad (9)$$

where ω_x is the frequency at which the maximum amplitude of one given band is located. The first term in eq. 9 gives the contributions arising from the self-correlations of internal velocities, while the second term provides cross-correlation contributions from pairs of velocities. Cross-correlations $\langle \dot{R}_m(t) \cdot \dot{R}_l(0) \rangle$ in $I_{IC_{ml}}(\omega)$ can be positive or negative depending on the relative

phase of the motions of the two ICs, therefore $I_{IC_{ml}}(\omega)$ can likewise be positive or negative. The contribution of each spectral component $I_{IC_{ml}}$ into the active $IR(\omega_x)$ band is given by the normalised weight w_{ml} :

$$w_{ml} = \frac{I_{IC_{ml}}(\omega_x)}{IR(\omega_x)} \quad (10)$$

What we do in practice is to reconstruct the surface area of each band in the $IR(\omega)$ spectrum by the sum of the $I_{IC_{ml}}$'s surface areas, as these latter possess the IR activity. With this reconstruction, we not only know which $I_{IC_{ml}}$'s contribute to each IR band, we also know the percentage of participation of each $I_{IC_{ml}}$ into the final IR band (the w_{ml} weights in eq. 10 above), hence providing the individual contributions of the internal coordinates into that particular band. Such reconstruction not only includes the self-part contribution of the internal coordinates but also all cross-parts, which is the main issue in reconstruction the full IR activity of one given band. Note that the procedure is applied in between ω_x and $\omega_x \pm 10 \text{ cm}^{-1}$, as the bands in the current work are extremely thin.

We also adopt the following conventions. Any $I_{IC_{ml}}(\omega_x) < 1\%$ of $IR(\omega_x)$ is not taken into account for the final band assignment, and any contribution $I_{IC_{ml}}(\omega_x) > 10\%$ is always printed in the final graph. As soon as the spectral reconstruction procedure achieves 80% of the area of $IR(\omega_x)$, adding up $I_{IC_{ml}}(\omega_x)$ contributions from higher % contributions to lower % contributions, the fitting procedure is stopped for that particular $IR(\omega_x)$ band. From empirical practice on the peptides presented in this work in sections 6.1-6.2, delocalized modes in the far IR/THz domain ($< 800 \text{ cm}^{-1}$) can be made of multiple small contributions, each one can be $< 10\%$. These contributions are added up for the final band assignment until one reaches the 80% threshold mentioned above, and a criteriom at 4% is taken for printing these contributions in the graphs.

5.3 Graph Theory for the final assignment and representation of the modes

To have a direct and graphical view of the contributions of the ICs into the IR spectrum bands, graphs are used. In Graph Theory, the $I_{IC_{ml}}(\omega_x)$ self- ($m=l$) and cross- ($m \neq l$) contributions to $IR(\omega_x)$ can be seen as a coloured indirect graph labelled $G=(V, E, F_V, F_E)$, where:

- V is the set of vertices of the graph G . Each $I_{IC_m}(\omega_x)$ self-contribution represents one vertex of G .
- E is the set of edges of the graph G . Given two vertices a and b of V , $[a, b]$ belongs to E if and only if a (i.e. $IC_a = R_a$) and b (i.e. $IC_b = R_b$) are cross-correlated ($I_{IC_{ml}}(\omega_x) \neq 0$).
- $F_V: V \rightarrow [0,100]$ is the weight (here percentage of contribution of a given IC, self-term $I_{IC_m}(\omega_x)$) on the vertices of the graph G .
- $F_E: E \rightarrow [0,100]$ is the weight (here percentage of contribution of a pair of correlated ICs, $I_{IC_{ml}}(\omega_x)$) on the edges of the graph G .

For the graphical representation of the graph G , colored graphs are presented in order to distinguish the possible IC components in the vibrational modes. All conventions for the graphs representations are reported in fig. 7. Typically, there are four colors for the vertices of the graph G : red for stretching IC motions, light

blue for bending IC motions, gray for torsional IC motions, yellow for intermolecular hydrogen bond motions. The percentage of

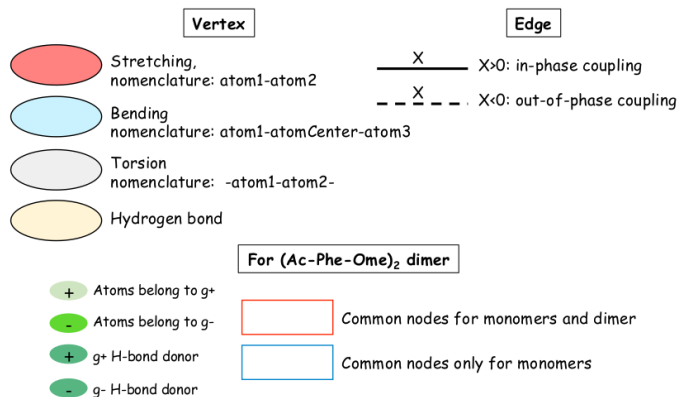


Fig. 7 Conventions adopted for the graphs representations.

participation of the IC component represented by the vertex into the mode assignment is written next to the vertex (F_V value). The connections in between the vertices, i.e. the edges in the graph G (cross-contributions of a pair of ICs in the underlying IR spectrum), are traced with a black line, either a continuous line for in-phase couplings (positive values for F_E cross-terms $I_{C_{ml}}(\omega_x)$) or a dashed line for out-of-phase couplings (negative values for cross-terms $I_{C_{ml}}(\omega_x)$).

One graph is obtained per mode/per active IR band ($I(\omega_x)$), with the following data contained in the graph:

- Vertex - mm :

$$I_{C_{mm}}(\omega_x) \rightarrow F_V \rightarrow \frac{2\pi\beta}{3cV} \left| \frac{\partial \mu}{\partial R_m} \right|^2 \int_{\omega_{x_{min}}}^{\omega_{x_{max}}} dt e^{i\omega t} \langle \dot{R}_m(t) \cdot \dot{R}_m(0) \rangle$$

- Edge - ml :

$$I_{C_{ml}}(\omega_x) \rightarrow F_E \rightarrow \frac{2\pi\beta}{3cV} \frac{\partial \mu}{\partial R_m} \frac{\partial \mu}{\partial R_l} \int_{\omega_{x_{min}}}^{\omega_{x_{max}}} dt e^{i\omega t} \langle \dot{R}_m(t) \cdot \dot{R}_l(0) \rangle$$

Various aspects of the Graph Theory are highly advantageous when using this theory for modes assignments. As soon as the graph is plotted, one can immediately get the self- and cross-term contributions of the internal coordinates into one given vibrational mode (band). One can hence immediately conclude from the graph whether a vibrational mode is made of coupled or uncoupled internal motions: one looks at whether the graph is made of connected elements and/or of disconnected elements. By eye, one can immediately see from the graph whether a mode is made of localized (one main single component in the graph) or (highly) delocalized/collective motions, simply by counting the number of connected elements in the graph. On the other hand, by using the weights displayed on the graph, one can get the information on the 'electronic' vs 'mechanic' couplings at play in the motions, in a very efficient way. As will be used in the discussions of the results in section 6, ICs can be mechanically correlated without participating to the final IR intensity (which arises from the 'electronic' coupling contained in the APTs in eqs. 5 & 7). In that case,

the edge on the graph would have a high value of the weight but low/even zero weight on one of the connected vertices.

The ultimate advantage of graphs is the natural capability for comparing graphs and extract similarities. This is exactly what we need for comparing vibrational modes in between molecular systems. To achieve similarity measurements, we apply the following working hypothesis that the two graphs to be compared have the same set and same order of the ICs internal coordinates. Then, comparison of adjacency matrices is performed, resulting in one of the following cases: 1) the two graphs are equal if the adjacency matrices are identical, 2) one graph is a sub-graph of the other graph if one matrix is included in the second one, 3) the two graphs are different when there is no match in the two matrices, and 4) the two graphs share only some nodes and/or some edges, the algorithm hence provides the common graph. One could go one step further and also include the weights into the similarity algorithm. Isomorphism would hence have to be applied, see for instance ref.⁹⁷ for more details on isomorphism and related algorithms. This has not been done in the present work.

We can readily illustrate this theory on the IR spectrum assignment of the prototype very simple gas phase water molecule. DFT-MD has been accumulated for 15 ps at 50 K (hence avoiding the rotational motion of the molecule), $\delta t = 0.4$ fs, BLYP functional for the electronic representation, and mixed gaussian aug-TZV2P-GTH and 300 Ry plane waves basis sets (CP2K code, see section 2 for more details on DFT-MD). The three graphs for the three IR active vibrational modes of the water are presented in figure 8. The two higher frequency modes are immediately assigned as the

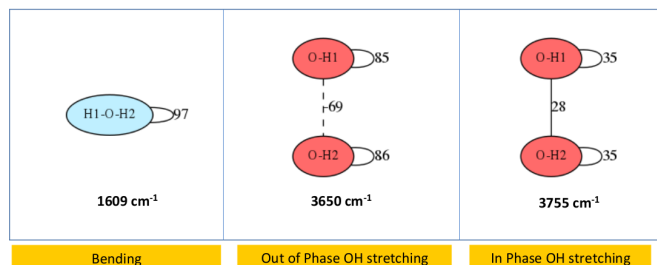


Fig. 8 Graphs of the three active IR modes for a water gas phase molecule at 50 K. See fig. 7 for the nomenclatures.

in-phase and out-of-phase O-H stretching motions of H_2O , while the 1609 cm^{-1} band is the bending motion. This latter is completely decoupled from the stretching motions, as expected, i.e. the bending IC is never showing up in the graphs of the 3650 and 3755 cm^{-1} stretching bands, and vice versa. One can see that hence the 1609 cm^{-1} bending mode has only one single vertex in the graph representation. For the two stretching modes, one can observe that the two O-H groups systematically equally participate to each vibrational mode, the percentage of contribution is readily seen on the vertices of the graphs. Very nicely, the in-phase/out-of-phase nature of the stretching modes is given by the sign of the weight on the edges, i.e. $I_{C_{ml}}(\omega_x)$ cross-term. The higher 3755 cm^{-1} mode has a positive (+28) value of the

weight for the cross-contribution in between the two O-H stretching motions, thus an in-phase motion of the two stretchings, while the lower 3650 cm^{-1} mode has a negative value (-69) of the cross-contribution weight, therefore the out-of-phase stretching motion. One can also see that the values of the weights on the vertices (self-contribution) and the ones on the edges (cross-contribution) are of the same order of magnitude, showing that both contributions are equally relevant for the final activity of the stretching bands. Restricting band assignments to the self-contribution only would hence not be correct, the correlation between the two stretching motions has to be included for the final comprehension.

6 Applications of Graph Theory for the IR spectrum assignment of gas phase peptides

6.1 IR spectroscopy of N-Methyl-Acetamide NMA at two temperatures

Before applying the Graph Theory method to the modes assignment of $(\text{Ac-Phe-OMe})_{1,2}$ peptides in the far-IR/THz spectral domain (section 6.2), we present one more proof-of-principle of the methodology, by taking the example of the gas phase NMA (N-Methyl-Acetamide) molecule with assignments of the vibrational modes in two dynamical spectra, one at low temperature (20 K) and one at room temperature ($\sim 300\text{ K}$). With the comparison of modes at these two rather different temperatures, one especially explores the (di)similarities in between the vibrational motions and the potential red-/blue-shifts of the vibrational peaks.

Figures 9-10 report the graphs for three selected vibrational modes in the IR spectrum of NMA, comparing the assignments in between 20 K and 300 K dynamics of gas phase NMA. The atomic labels used in the graphs are in fig. 9-left. All details on the coloring nomenclatures are in fig. 7. Details on these trajectories have

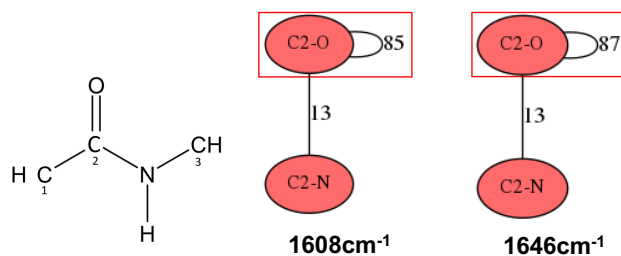


Fig. 9 Left: nomenclature adopted for the labeling of the NMA molecule atoms and used in the graphs of the modes for this molecule. Right: two graphs for the modes at 1608 cm^{-1} and 1646 cm^{-1} respectively observed for the 20 K and 300 K NMA gas phase molecule. See fig. 7 for the nomenclatures.

been published in ref.⁹⁸, here we use 3 ps and 10 ps trajectories respectively for the low and room temperature of NMA, the BLYP functional and the gaussian TZV2PX-MOLOPT-GHT plus 340 Ry plane waves basis sets (CP2K code, see section 2 for more details on DFT-MD). We show and comment three selected modes in the $400\text{--}2000\text{ cm}^{-1}$ spectral domain, at $\sim 1600\text{ cm}^{-1}$, $\sim 1400\text{ cm}^{-1}$ and $\sim 400\text{ cm}^{-1}$ to illustrate the diversity of motions and cou-

plings in these motions, and how the graphs capture these data. The $1608(20\text{ K})/1646(300\text{ K})\text{ cm}^{-1}$ modes of gas phase NMA are the localized $C=O$ Amide I stretching, see fig.9-right, with a slight blue-shift in frequency at the higher temperature. The graphs nicely show that although the $C=O$ stretching is mechanically coupled with the $C-N$ backbone stretching, see the values on the edges connecting the two vertices, the $C-N$ motion does not participate to the final IR activity of the mode ($C=O$ and $C-N$ are not electronically coupled, no value next to the $C-N$ vertex).

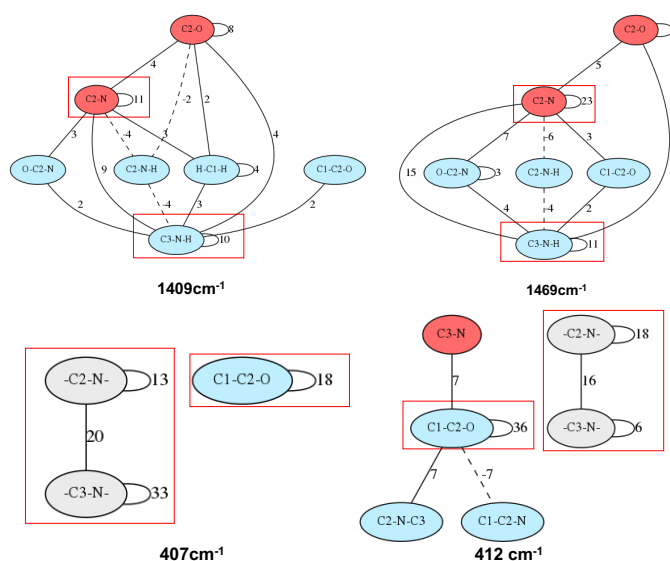


Fig. 10 Graphs for the modes at $1409/1469\text{ cm}^{-1}$ and $407/412\text{ cm}^{-1}$ observed for the 20 K and 300 K NMA gas phase molecule. See fig. 7 for the nomenclatures.

The graphs for the $\sim 1400\text{ cm}^{-1}$ and $\sim 400\text{ cm}^{-1}$ modes nicely illustrate the complexity of more delocalized/collective modes with either one single graph that contains all the coupled motions (this is for the $1409(20\text{ K})/1469(300\text{ K})\text{ cm}^{-1}$ modes) or with two disconnected graphs (this is for the $407(20\text{ K})/412(300\text{ K})\text{ cm}^{-1}$ modes). The $\sim 1400\text{ cm}^{-1}$ mode of NMA is a combination of the $C=O$ and backbone $C-N$ stretching motions coupled with the $C-N-H$ bending motion. Several other bendings are also present in the mode, showing rather delocalized motions over the whole molecule. Interestingly, while the graphs show that the $C=O$ stretch is coupled to the backbone $C-N$ and to the $N-H$ bending, the percentage of participation of both the $C=O$ vertex and of the associated edges are small. The similarities between the two modes at the two temperatures, highlighted by the rectangles in the figure, nicely illustrate that the nature of this mode has not changed with the increase in temperature. The blue-shift of this mode at higher temperature is hence due to the temperature effect rather than a change in the underlying motions.

The two graphs for the $\sim 400\text{ cm}^{-1}$ mode of isolated NMA at 20 K and 300 K show that the mechanical nature of this mode is not changed by temperature either, the two disconnected graphs (for each temperature) are present at both temperatures, and all motions are similar as highlighted by the rectangles in both

graphs. Interestingly, the graph at higher temperature shows that there are additional couplings to the $C-C-O$ bending motion probed within the motions, presumably due to larger amplitude motions that are accessible at the higher temperature. In-phase (positive weight on the edge) and out-of-phase (negative weight on the edge) motions are appearing in between the $C-C-O$ bending and respectively the $C-N-C$ and $C-C-N$ backbone bendings. Also at 300 K, despite the $C-N$ backbone stretching is mechanically coupled with the $C-C-O$ bending (weight = +7 on the edge), there is no weight on the $C-N$ vertex, indicating that this $C-N$ coordinate does not participate in the final (electronic) activity of the band at 412 cm^{-1} (it therefore does not participate into the change in the dipole moment of the molecule).

6.2 IR spectroscopy of Ac-Phe-OMe and of the β -sheet model (Ac-Phe-OMe)₂: effect of the hydrogen bonds on the signatures

The experimental IR-UV ion dip spectra and the matching theoretical spectrum from the $\beta_L(g^+)$ (monomer) and $\beta_L(g^+) - \beta_L(g^-)$ (dimer) conformations of the (Ac-Phe-OMe)_{1,2} peptides have been presented in sections 3-4 of this work, see also figures 3 and 6. For simplifying our text, in the following $\beta_L(g^+)$ and $\beta_L(g^-)$ are replaced by the short-notations g^+ and g^- . In the following, we comment only a few selected IR bands below 800 cm^{-1} and we systematically comment how the band assignment are/are not modified from the monomer(s) to the dimer. We comment bands going from 800 cm^{-1} down to 60 cm^{-1} .

- N-H wagging motion ($450\text{--}650\text{ cm}^{-1}$, fig. 11). With our current definitions of non-redundant internal coordinates, the out-of-phase torsion of ϕ (labeled $-C8-N-$ in the graphs) and ω (labeled $-C9-N-$ in the graphs), see also fig. 1, defines the $\omega(N-H)$ out-of-plane N-H wagging motion. This is the dominant motion that appears in the graphs for the IR bands located at 463 cm^{-1} for g^+ and 453 cm^{-1} for g^- : the large values over these two motions on the two graphs vertices show that these two modes are highly localized over these two movements, and there are only small mechanical couplings with the backbone at the N-Terminal ($-C9-C10-$ torsion, small values on the edges connecting the $-C8-N-$ torsion to the $-C9-C10-$ torsion). Very nicely, the out-of-phase motion in between the ϕ and ω torsions that defines the out-of-plane wagging motion of the N-H group appears with the large negative value at the edge connecting these two torsional motions. Note that there is one IR N-H wagging band per N-H group, one per monomer.

There are two main bands in the $g^+ - g^-$ dimer assigned to the $\omega(N-H)$ waggings, blue-shifted by $\sim 100\text{--}200\text{ cm}^{-1}$ from the bands in the monomers. The blue-shift is the result of the $N-H \cdots O=C$ hydrogen bonds forming the β -sheet structure in the dimer. See ref.⁵⁶ for a systematic characterization of the values of the blue-shifts in the wagging motions in the far-IR/THz domain obtained on dipeptides.

One can immediately see that the graphs for the $g^+ - g^-$ dimer are more complex and more busy than the ones for each monomer (same criteria used), nicely showing the higher complexity in inter-molecular hydrogen bonded motions. The notation in the

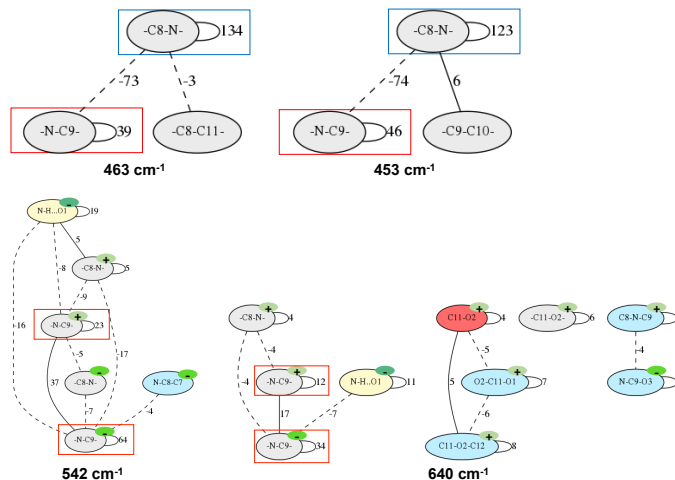


Fig. 11 Graphs for the modes at (top) $463/453\text{ cm}^{-1}$ respectively for g^+ and g^- monomers, and (bottom) $542/640\text{ cm}^{-1}$ for $g^+ - g^-$ dimer. See fig. 7 for the nomenclatures.

graphs with green circles carrying the sign $+/-$ refers to the strand g^+/g^- responsible for the motions. One can hence see that the graph for the 542 cm^{-1} mode is composed of one single graph with several interconnected vertices. The IR mode has both N-H waggings from both peptide strands as main contributors (follow the values on the $-C8-N-$ and $-N-C9-$ vertices in the graph), and that these two $\omega(N-H)$ are mechanically coupled (follow the negative values on the edges in the graph). The values on the vertices also show that despite both N-H contribute to the final IR mode, N-H on the g^- peptide strand dominates the motion. Very nicely, one also sees that the $g^- \omega(N-H)$ motion is coupled with the intermolecular hydrogen bond to g^+ , by following the edge in between the vertices labeled $-N-C9-$ and $N-H \cdots O1$. The value on the $N-H \cdots O1$ vertex also shows that the H-Bond is a large contributor to the final activity of this mode.

The graph for the 640 cm^{-1} IR mode shows even more complexity than the one for the 542 cm^{-1} mode, and it is especially composed of two disconnected sub-graphs. By following the red rectangles in the graphs of these two modes (the 542 and the 640 cm^{-1}), one can immediately observe the dominant similarities in between the two modes, and indeed one can see here that the same $\omega(N-H)$ waggings are responsible for this higher frequency mode, also with the same coupling of the wagging to the inter-molecular H-Bond. Beyond the wagging motions, the graph for the 640 cm^{-1} IR mode also shows that there are other motions participating to the final assignment of this band (typically $C=O$ stretching and several backbone torsions), however disconnected from the wagging motions. The two sub-graphs are indeed not connected to each other. These other motions participate to a far lesser extent to the final activity of the mode, as the small values both on the vertices and edges of the disconnected graph show.

- Delocalized motions at $\sim 250\text{--}280\text{ cm}^{-1}$, (fig. 12). The graphs of the 276 and 265 cm^{-1} IR bands, respectively for the g^+ and g^- monomers, show highly delocalized modes constructed on coupled bending motions over the backbones. Both graphs display dominant similarities (red rectangles) for the COC C-

terminal bending and for the CNC 'central' bending of the backbone, but also (not within red rectangles for the sake of clarity of the graphs) for the N-Terminal N-C=O and N-C-C bendings. The values on the edges of the graphs are positive for the 276 cm^{-1} band of g^+ , showing in-phase coupled motions, while they can be positive or negative for the 265 cm^{-1} mode of g^- , hence with in-phase and out-of-phase couplings. Once the two peptide strands form the $g^+ - g^-$ H-Bonded dimer, the two IR modes located at 271 and 257 cm^{-1} have the same general assignment, also identical to the 276 and 265 cm^{-1} IR bands for the individual monomers. The COC C-Terminal bending dominates these two modes, see the larger values on the $C11 - O2 - C12$ vertices in both graphs, one for g^+ (271 cm^{-1}), one for g^- (257 cm^{-1}), systematically coupled to the inter-molecular hydrogen bond motions, either through the intermolecular $O \cdots H$ stretching or the intermolecular $-O \cdots H-$ torsion. One nicely also sees the COC couplings to other backbone C-Terminal bendings, i.e. OC=O and CCO, either localized on the same strand or localized on the partner strand. Note the systematic positive values on the edges of the graphs, showing in-phase coupled motions. Note also that the 'central' backbone CNC bending is not participating anymore in the dimer modes, while its contribution to the monomers modes was important (compare the value at the vertex of $C8 - N - C9$ for the two monomers graphs with the value of the $C11 - O2 - C12$ vertex), this motion is hence inhibited by the intermolecular H-Bond in the dimeric peptide. One can also observe that the contributions to these modes are of both electronic and mechanical nature, as indicated by non-zero values of the weights on the vertices and edges.

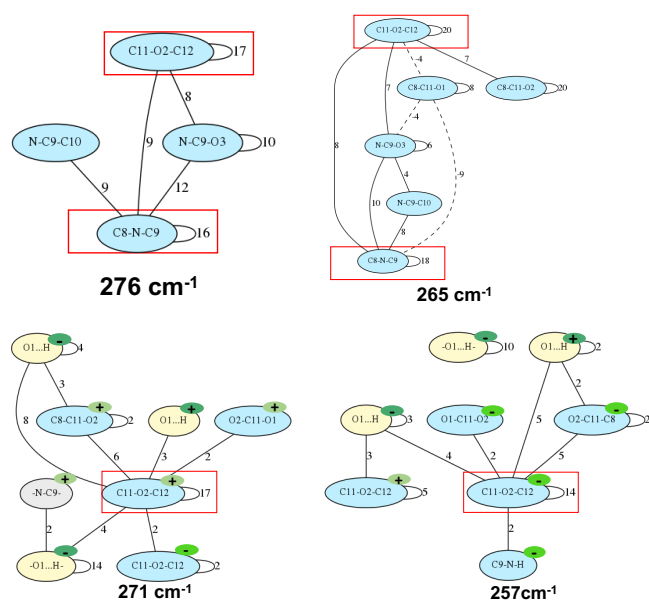


Fig. 12 Graphs for the modes at (top) $276/265\text{ cm}^{-1}$ respectively for g^+ and g^- monomers, and (bottom) $271/257\text{ cm}^{-1}$ for $g^+ - g^-$ dimer. See fig. 7 for the nomenclatures.

- **Torsional delocalized motions at $\sim 200\text{ cm}^{-1}$, (fig. 13).** We illustrate one more motion in the $\sim 200\text{ cm}^{-1}$ spectral domain in figure 13. The graph for the g^+ monomer nicely illustrates a

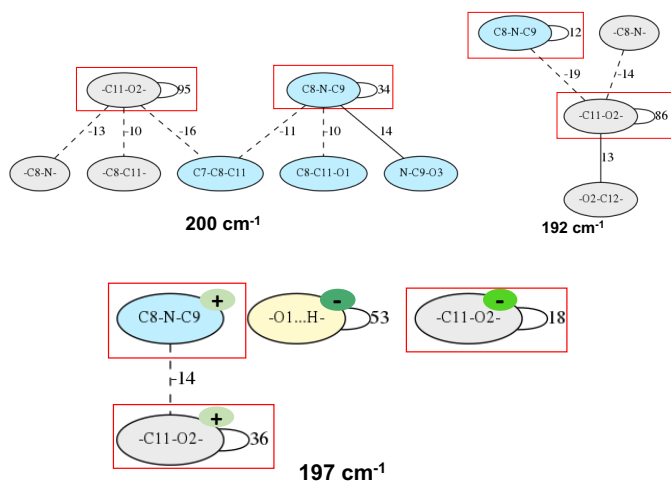


Fig. 13 Graphs for the modes at (top) $200/192\text{ cm}^{-1}$ respectively for g^+ and g^- monomers, and (bottom) 197 cm^{-1} for $g^+ - g^-$ dimer. See fig. 7 for the nomenclatures.

highly delocalized torsional 200 cm^{-1} mode, where two motions dominate, i.e. the torsion around the C-Terminal $-C11 - O2-$ backbone bond and the bending around the 'central' backbone CNC. Each of these two motions is mechanically coupled to several torsions and bendings, these coupled motions however do not systematically participate to the final IR activity of the mode (when there are no values on the associated vertices). Note in the graphs the systematic out-of-phase couplings to the C-O torsion, and the mixing of in-/out-of-phase couplings to the CNC bending. Nicely, the 200 cm^{-1} IR mode of g^+ shows the mechanical coupling of the backbone motions to the bending of the side-chain, with the χ_1 ($C7 - C8 - C11$ in the graph) torsion coupled to both the dominant CO torsion and CNC bending. The zero-value on the graph edge associated to this motion shows that it does not participate to the final IR activity. The same motions are responsible for the 192 cm^{-1} mode of g^- , but one can see that the graph of this mode is simpler than the one for g^+ , with only the two dominant CO torsion and CNC backbone bending motions (similarities in the red rectangles) dominating the whole mode and providing its final IR intensity.

The IR mode at 197 cm^{-1} in the dimer is nicely composed of the two CO torsions from each monomer strand, however uncoupled to each other as the disconnected graphs show. The CO torsion on the g^+ strand is mechanically coupled to the CNC bending on the same strand, this latter not participating to the final IR activity (no value on the associated vertex). Added to the overall IR activity of the dimer mode, the torsion around one of the intermolecular hydrogen bonds is a large contributor, as revealed by the largest value on the $-O1 \cdots H$ vertex.

7 Conclusions and perspectives

The work presented here has shown once more that the combined approach in between IR-UV ion dip spectroscopy experiments in the gas phase and DFT-based molecular dynamics simulations for theoretical spectroscopy is successful for revealing the 3D structures of (here) peptides and dimers of peptides in the far-IR/THz

spectral domain. Especially emphasized in the present work is the importance of this non-congested spectral region in order to unambiguously assign the structures. Remarkably, the theoretical and experimental spectra provide the same well-resolved features in this low frequency domain, making the experiment-calculation one-to-one matching powerful. Taking the example of Ac-Phe-OMe and of its (Ac-Phe-OMe)₂ dimeric form, we have shown that previous ambiguities in the conformational assignment of these peptides have been removed. As already emphasized in ref.⁵⁶ where a mapping of the vibrational motions of dipeptides was achieved, the vibrational signatures in the 100-400 cm⁻¹ region are especially relevant into the final conformational assignment.

This paper has also reported and demonstrated our proof-of-principle in using Graph Theory for vibrational modes assignments. To our best knowledge, it is the first time Graph Theory has been employed in the context of revealing the intra- and inter-molecular motions responsible for the vibrational modes. The whole methodology is based on the re-writing of the IR spectrum in terms of Fourier transforms of correlations of internal coordinates velocities modulated by Atomic Polar Tensors (transformed from cartesian coordinates where they are naturally defined to the internal coordinates space). This APT-based approach is powerful in that it conserves the strength of the Internal Coordinates Velocities-DOS in identifying the coordinates that vibrate at a given frequency, while simultaneously maintaining the knowledge on the IR-activity of such motions.

Our strategy for assigning the vibrational modes to the active IR bands is therefore organized over two steps: one step consists in reconstructing the exact theoretical IR spectrum (calculated through the Fourier transform of the time correlation function of the dipole moment of the system) in terms of the spectra of the APT weighted-internal coordinates velocities, simultaneously extracting the contributions of self- and cross-terms; the second step consists in applying Graph Theory algorithms to transform these spectral data into graphs where the vertices are the internal coordinates that participate to the modes (self-terms) and the edges represent the internal coordinates cross-terms in the spectrum reconstruction. The cross-terms are highly important contributors to the final spectrum reconstruction as they carry a wealth of information that the self-terms do not. The graphs hence obtained immediately show how many motions are participating to each given vibrational mode, whether the mode results in localized or delocalized/collective motions, how many coupled motions participate to the mode together with the percentage of participation of each of the motions into the final mode. Similarities between graphs can be easily obtained, hence providing similarities in between vibrational modes. This is especially useful when assigning the vibrational modes of dimers (here the (Ac-Phe-OMe)₂ β -sheet peptidic model), providing the similarities/disimilarities in the modes in going from the monomer to the dimer. Very interestingly, the graphs immediately carry the information on the electrical and mechanical couplings in the motions for the final IR intensity of the mode.

The applications of this methodology to the Ac-Phe-OMe monomer and the (Ac-Phe-OMe)₂ β -sheet peptidic model have shown that the complexity in assigning vibrational modes from

MD simulations is released with the graphs. Our newly developed graph-based methodology furthermore allows an easy comparison between the modes of isolated monomer(s) and their complexes (here illustrated with a H-Bonded dimer): the graphs allow an easy identification of the vibrational modes 'localized' on the monomer(s) and not being affected by the formation of the inter-molecular complex, while conversely it also allows an easy identification of the vibrational modes of the monomer(s) that are being affected by the formation of the inter-molecular complex.

This paper showed the proof-of-principle of this newly developed methodology for vibrational modes assignment based on Graph Theory, more developments are needed in order to make this strategy more broadly applicable. The essential element of the graph-based method is the reconstruction of the IR spectrum in terms of correlations of internal coordinates velocities modulated by Atomic Polar Tensors, which conserve the IR activity of the underlying vibrational motions.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was performed under Grant ANR DYNWIN ANR-14-CE35-0011-01 and using HPC resources from GENCI-France Grant 072484 (CINES/IDRIS/TGCC). SB and MPG also acknowledge funding from the LABEX CHARM₃AT (LABoratoire d'EXcellence CHARM₃AT CHimie des ARchitectures Moléculaires Multifonctionnelles et des MATériaux). Discussions with Prof Dominique Barth, Dr Franck Quessette, PhD Simone Pezzotti, Prof Alberto Milani and Prof Chiara Castiglioni are gratefully acknowledged.

Notes and references

- 1 A. M. Rijs and J. Oomens, *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, Springer International Publishing, ISBN 978-3-319-19204-8, 2015.
- 2 A. M. Rijs and J. Oomens, in *IR Spectroscopic Techniques to Study Isolated Biomolecules*, ed. A. M. Rijs and J. Oomens, Springer International Publishing, 2015, pp. 1–42.
- 3 E. Gloaguen and M. Mons, *Topics Curr. Chem.*, 2015, **364**, 225.
- 4 J. P. Schermann, *Spectroscopy and Modeling of Biomolecular Building Blocks*, Elsevier Science, Amsterdam, The Netherlands, 2007.
- 5 T. R. Rizzo, J. A. Stearns and O. V. Boyarkin, *Int. Rev. Phys. Chem.*, 2009, **28**, 481–515.
- 6 N. C. Polfer and J. Oomens, *Mass Spectrom. Rev.*, 2009, **28**, 468.
- 7 N. C. Polfer, *Chem. Soc. Rev.*, 2011, **40**, 2211.
- 8 C. N. Stedwell, J. F. Galindo, A. E. Roitberg and N. C. Polfer, *Annu. Rev. Anal. Chem.*, 2013, **6**, 267.
- 9 A. Cismesia, M. Bell, L. Tesler, M. Alves and N. Polfer, *Analyst*, 2018, **143**, 1615.
- 10 E. Mucha, M. Marianski, F. Xu, D. Thomas, G. Meijer, G. von Helden, P. Seeberger and K. Pagel, *Nat. Comm.*, **9**, 4174.
- 11 M. Kamrath and T. Rizzo, *Acc. Chem. Res.*, **51**, 1487.

- 12 C. Seaiby, A. Zabuga, A. Svendsen and T. Rizzo, *J. Chem. Phys.*, 2016, **144**, 014304.
- 13 R. Dunbar, J. Martens, G. Berden and J. Oomens, *J. Phys. Chem. A.*, **122**, 5589.
- 14 J. Martens, G. Berden, H. Bentlage, K. Coene, U. Engelke, D. Wishart, M. von Scherpenzeel, L. Kluijtmans, R. Wevers and J. Oomens, *J. Inherited Metabolic Disease*, **41**, 367.
- 15 E. Sinclair, K. Hollywood, C. Yan, R. Blankley, R. Breitling and P. Barran, *Analyt.*, **143**, 4783.
- 16 D. Stuchfield and P. Barran, *Curr. Op. Chem. Biology*, **42**, 177.
- 17 W. Sohn, S. Habka, E. Gloaguen and M. Mons, *Phys. Chem. Chem. Phys.*, **19**, 17128.
- 18 E. Gloaguen, B. Tardivel and M. Mons, *Struct. Chem.*, **27**, 225.
- 19 S. Ishiuchi, Y. Sasaki, J. Lisy and M. Fujii, *Phys. Chem. Chem. Phys.*, 2018, DOI:10.1039/c8cp05839c.
- 20 H. Ke and J. Lisy, *Phys. Chem. Chem. Phys.*, **17**, 25354.
- 21 E. Mucha, A. G. Florez, M. Marianski, D. Thomas, W. Hoffman, W. Struwe, H. Hahm, S. Gewinner, W. Schöllkopf, P. Seiberger, G. von Helden and K. Pagel, *Angew. Chemie. Int.*, **56**, 11248.
- 22 C. Masellis, N. K. M. Kamrath, D. Clemmer and T. Rizzo, *J. Am. Soc. Mass Spectrom.*, **28**, 2217.
- 23 A. D. Depland, G. Renois-Predelus, B. Schindler and I. Compagnon, *Int. J. Mass Spectrom.*, **434**, 65.
- 24 I. Compagnon, B. Schindler, G. Renois-Predelus and R. Daniel, *Curr. Op. Struct. Biology*, **50**, 171.
- 25 I. Usabiaga, A. Camiruaga, A. Insausti, P. Carcabal, E. Cocinero, I. Leon and J. Fernandez, *Frontiers in Phys.*, **6**, 1.
- 26 S. Boldissar and M. de Vries, *Phys. Chem. Chem. Phys.*, **20**, 9701.
- 27 I. Chen and M. de Vries, *Phys. Chem. Chem. Phys.*, **18**, 20005.
- 28 M. Ligare, A. Rijs, G. Berden, M. Kabelac, D. Nachtigallova, J. Oomens and M. de Vries, *J. Phys. Chem. B.*, **119**, 7894.
- 29 R. van Outersterp, J. Martens, G. Berden, J. Steill, J. Oomens and A. Rijs, *Phys. Chem. Chem. Phys.*, **20**, 28319.
- 30 J. Wagner, D. McDonald and M. Duncan, *J. Phys. Chem. Letters*, **9**, 4591.
- 31 J. Wagner, D. McDonald and M. Duncan, *J. Phys. Chem. A.*, **122**, 192.
- 32 N. Heine and K. Asmis, *Int. Rev. Phys. Chem.*, 2015, **34**, 1–34.
- 33 I. Hunig and K. Kleineremanns, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2650–2658.
- 34 J. M. Bakker, C. Plützer, I. Hünig, T. Häber, I. Compagnon, G. von Helden, G. Meijer and K. Kleineremanns, *ChemPhysChem*, 2005, **6**, 120–128.
- 35 M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz and M. F. Jarrold, *J. Phys. Chem.*, 1996, **100**, 16082–16086.
- 36 D. E. Clemmer and M. F. Jarrold, *J. Mass Spectrom.*, 1997, **32**, 577–592.
- 37 J. Seo, W. Hoffman, S. Warnke, X. Huang, S. Gewinner, W. Schöllkopf, M. Bowers, G. von Helden and K. Pagel, *Nat. Chem.*, **9**, 39.
- 38 W. Hoffman, G. von Helden and K. Pagel, *Curr. Op. Struct. Biology*, **46**, 7.
- 39 S. Daly, A. Kulesza, F. Poussigue, A. Simon, C. Choi, G. Knight, F. Chiro, L. MacAleese, R. Antoine and P. Dugourd, *Chem. Sci.*, **7**, 1609.
- 40 D. Grischkowsky, S. Keiding, M. van Exter and C. Fattinger, *J. Opt. Soc. Am. B*, 1990, **7**, 2006–2015.
- 41 C. T. Nemes, C. Koenigsmann and C. A. Schmuttenmaer, *J. Phys. Chem. Letters*, 2015, **6**, 3257–3262.
- 42 A. Bergner, U. Heugen, E. Bründermann, G. Schwaab, M. Havenith, D. R. Chamberlin and E. E. Haller, *Rev. Scient. Inst.*, 2005, **76**, 063110.
- 43 Y. Xu and M. Havenith, *J. Chem. Phys.*, 2015, **143**, 170901.
- 44 M. Weichman, S. Debnath, J. Kelly, S. Gewinner, W. Schöllkopf, D. Neumark and K. Asmis, *Top. Catal.*, **61**, 92.
- 45 T. Esser, H. Knorke, K. Asmis, W. Schöllkopf, Q. Yu, C. Qu and J. Bowman, *J. Phys. Chem. Letters*, **9**, 798.
- 46 M. R. Fagiani, H. Knorke, T. K. Esser, N. Heine, C. T. Wolke, S. Gewinner, W. Schöllkopf, M.-P. Gageot, R. Spezia, M. A. Johnson and K. R. Asmis, *Phys. Chem. Chem. Phys.*, 2016, **18**, 26743–26754.
- 47 J. A. Fournier, C. T. Wolke, C. J. Johnson, M. A. Johnson, N. Heine, S. Gewinner, W. Schöllkopf, T. K. Esser, M. R. Fagiani, H. Knorke and K. R. Asmis, *Proc. Natl. Acad. Sci.*, 2014, **111**, 18132–18137.
- 48 M. R. Fagiani, X. Song, P. Petkov, S. Debnath, S. Gewinner, W. Schöllkopf, T. Heine, A. Fielicke and K. R. Asmis, *Angew. Chemie. Int.*, 2017, **56**, 501–504.
- 49 X. Li, P. Claes, M. Haertelt, P. Lievens, E. Janssens and A. Fielicke, *Phys. Chem. Chem. Phys.*, 2016, **18**, 6291–6300.
- 50 A. Shayeghi, R. L. Johnston, D. M. Rayner, R. Schäfer and A. Fielicke, *Angew. Chemie. Int.*, 2015, **54**, 10675–10680.
- 51 A. Shayeghi, R. Schäfer, D. M. Rayner, R. L. Johnston and A. Fielicke, *J. Chem. Phys.*, 2015, **143**, 024310.
- 52 C. Kerpel, D. J. Harding, D. M. Rayner, J. T. Lyon and A. Fielicke, *J. Phys. Chem. C*, 2015, **119**, 10869–10875.
- 53 M. Savoca, J. Langer, D. J. Harding, D. Palagin, K. Reuter, O. Dopfer and A. Fielicke, *J. Chem. Phys.*, 2014, **141**, 104313.
- 54 V. J. F. Lapoutre, M. Haertelt, G. Meijer, A. Fielicke and J. M. Bakker, *J. Chem. Phys.*, 2013, **139**, 121101.
- 55 S. Jaeqx, J. Oomens, A. Cimas, M.-P. Gageot and A. M. Rijs, *Angew. Chemie. Int.*, 2014, **53**, 3663–3666.
- 56 J. Mahe, D. J. Bakker, S. Jaeqx, A. M. Rijs and M.-P. Gageot, *Phys. Chem. Chem. Phys.*, 2017, **19**, 13778–13787.
- 57 V. Yatsyna, D. J. Bakker, P. Salén, R. Feifel, A. M. Rijs and V. Zhaunerchyk, *Phys. Rev. Lett.*, 2016, **117**, 118101.
- 58 D. J. Bakker, A. Dey, D. P. Tabor, Q. Ong, J. Mahe, M.-P. Gageot, E. L. Sibert and A. M. Rijs, *Phys. Chem. Chem. Phys.*, 2017, **19**, 20343–20356.
- 59 M.-P. Gageot and R. Spezia, in *Theoretical Methods for Vibrational Spectroscopy and Collision Induced Dissociation in the Gas Phase*, ed. A. M. Rijs and J. Oomens, Springer International Publishing, 2015, pp. 99–151.
- 60 M.-P. Gageot, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3336–3359.

- 61 C. Marinica, G. Grégoire, C. Desfrancois, J. P. Schermann, D. Borgis and M. P. Gaigeot, *J. Phys. Chem. A.*, 2006, **110**, 8802.
- 62 A. Cimas, T. D. Vaden, T. S. J. A. de Boer, L. C. Snoek and M. P. Gaigeot, *J. Chem. Theor. Comput.*, 2009, **5**, 1068.
- 63 V. Brites, A. L. Nicely, N. Sieffert, M. P. Gaigeot and J. M. Lisy, *Phys. Chem. Chem. Phys.*, 2014, **16**, 13086.
- 64 J. P. Beck, M.-P. Gaigeot and J. M. Lisy, *Spectro. Chimica Acta A: Molecular and Biomolecular Spectroscopy*, 2014, **119**, 12.
- 65 V. Brites, J. M. Lisy and M. P. Gaigeot, *J. Phys. Chem. A.*, 2015, **119**, 2468.
- 66 D. J. Bakker, Q. Ong, A. Dey, J. Mahé, M.-P. Gaigeot and A. M. Rijs, *J. Mol. Spectros.*, 2017, **342**, 4 – 16.
- 67 M. Gerhards and C. Unterberg, *Phys. Chem. Chem. Phys.*, 2002, **4**, 1760–1765.
- 68 M. Gerhards, C. Unterberg and A. Gerlach, *Phys. Chem. Chem. Phys.*, 2002, **4**, 5563–5565.
- 69 H. Fricke, A. Gerlach and M. Gerhards, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1660–1662.
- 70 C. Soto, *Nat Rev Neurosci*, 2003, **4**, 49–60.
- 71 C. Bleiholder, N. F. Dupuis, T. Wyttenbach and M. T. Bowers, *Nat. Chem.*, 2011, **3**, 172–177.
- 72 M.-P. Gaigeot, M. Martinez and R. Vuilleumier, *Mol. Phys.*, 2007, **105**, 2857–2878.
- 73 M. Thomas, M. Brehm, R. Fligg, P. Vohringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–6622.
- 74 M. Martinez, M.-P. Gaigeot, D. Borgis and R. Vuilleumier, *J. Chem. Phys.*, 2006, **125**, 144106.
- 75 G. Mathias, S. D. Ivanov, A. Witt, M. D. Baer and D. Marx, *J. Chem. Theory. Comput.*, 2012, **8**, 224–234.
- 76 M. Schmitz and P. Tavan, *J. Chem. Phys.*, 2004, **121**, 12233–12246.
- 77 M. Schmitz and P. Tavan, *J. Chem. Phys.*, 2004, **121**, 12247–12258.
- 78 M. Thomas, M. Brehm and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2015, **17**, 3207–3213.
- 79 J. Bowman, X. Zhang and A. Brown, *J. Chem. Phys.*, 2003, **119**, 646.
- 80 M. Kaledin, A. Brown, A. Kaledin and J. Bowman, *J. Chem. Phys.*, 2004, **121**, 5646.
- 81 J. J. P. Stewart, *Journal of Molecular Modeling*, 2007, **13**, 1173–1213.
- 82 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuse-ria, 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, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox., *Gaussian 09 and Revision C.02 and Gaussian and Inc. and Wallingford CT and 2009*.
- 83 J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. H. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.
- 84 A. D. Becke, *Phys. Rev. A.*, 1988, **38**, 3098–3100.
- 85 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B.*, 1988, **37**, 785–789.
- 86 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 87 S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B.*, 1996, **54**, 1703.
- 88 D. A. McQuarrie, *Statistical Mechanics*, University Science Books, 2000.
- 89 R. Kubo, M. Toda and N. Hashitsume, *Statistical Physics II*, Springer Berlin Heidelberg, 1991, vol. 31.
- 90 M. P. Gaigeot and M. Sprik, *J. Phys. Chem. B.*, 2003, **107**, 10344–10358.
- 91 C. Qu and J. Bowman, *Phys. Chem. Chem. Phys.*, 2018, DOI: 10.1039/c8cp04990d.
- 92 T. Roy and R. Gerber, *Phys. Chem. Chem. Phys.*, **15**, 9468.
- 93 G. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biology*, 1963, **7**, 95 – 99.
- 94 J. Mahe, S. Jaesq, A. M. Rijs and M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2015, **17**, 25905–25914.
- 95 D. R. Galimberti, A. Milani, M. Tommasini, C. Castiglioni and M.-P. Gaigeot, *J. Chem. Theory Comput.*, 2017, **13**, 3802–3813.
- 96 J. E. Wilson, J. C. Decius and P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*, Dover Publications, New edition edition edn, 1980.
- 97 S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth and M.-P. Gaigeot, *J. Chem. Phys.*, 2018, **149**, 184102–15.
- 98 M. P. Gaigeot, R. Vuilleumier, M. Sprik and D. Borgis, *J. Chem. Theory Comput.*, 2005, **1**, 772.