Some opinions on MD-based vibrational spectroscopy of gas phase molecules and their assembly: An overview of what has been achieved and where to go

Marie-Pierre Gaigeot

Version of Record: https://www.sciencedirect.com/science/article/pii/S1386142521004406
Manuscript_2f64dea9f86146abfa9d9e4ad4b059a8

Some opinions on MD-based vibrational spectroscopy of gas phase molecules and their assembly: an overview of what has been achieved and where to go.

Marie-Pierre Gaigeot,[a,*]

[a] *Université Paris-Saclay, Univ Evry, CNRS, LAMBE UMR8587, 91025 Evry-Courcouronnes, France*

Abstract

We hereby review molecular dynamics simulations for anharmonic gas phase spectroscopy and provide some of our opinions of where the field is heading. With these new directions, the theoretical IR/Raman spectroscopy of large (bio)-molecular systems will be more easily achievable over longer time-scale MD trajectories for an increase in accuracy of the MD-IR and MD-Raman calculated spectra. With the new directions presented here, the high throughput 'decoding' of experimental IR/Raman spectra into 3D-structures should thus be possible, hence advancing e.g. the field of MS-IR for structural characterization by spectroscopy. We also review the assignment of vibrational spectra in terms of anharmonic molecular modes from the MD trajectories, and especially introduce our recent developments based on Graph Theory algorithms. Graph Theory algorithmic is also introduced in this review for the identification of the molecular 3D-structures sampled over MD trajectories.

Keywords: DFT-MD, FF-MD, vibrational spectroscopy, anharmonicities, IR, Raman, SFG, VDOS, ICDOS, Machine Learning, APT, Raman tensor, Graph Theory

1. Introduction

The characterization of the three dimensional conformational arrangement of molecules, clusters, assemblies of (bio-)molecules, and more broadly of complex materials, is one of the main goals in chemical-physics, analytical chemistry and material design.

*Corresponding author
Email address: mgaigeot@univ-evry.fr ()

In the gas phase community, modern analysis of analytical and bio-analytical chemistry is based on multidimensional workflows where e.g. mass spectrometry (MS) is orthogonally coupled to techniques such as chromatography and electrophoretic separation, Ion Mobility Spectrometry (IMS), activation methods, and spectroscopies. The goal is to sequence (bio-)molecules and ultimately reveal their 3D microscopic structures. The crucial advantage of MS-based analytical chemistry is the ability to perform on-line separation and analysis of individual oligomers, while most of the condensed phase analytical chemistry tools (e.g. NMR, circular dichroism) only access ensemble-averaged information, or require preliminary off-line separation. Environment control is also possible in MS, with ligand/solvent molecules that can be added one at a time.

Among the orthogonal couplings to MS, infrared (IR) action spectroscopy is one of the very few techniques providing direct information into atomistic 3D-structures, with the paradigm of 1-to-1 spectral fingerprints $\leftrightarrow$ 3D-structure identification. In the gas phase where we restrict this opinion paper, action spectroscopies such as IR-MPD (Infra Red Multiphoton Dissociation), IR-PD (Infra Red Photon Dissociation) and IR-UV ion dip, developed and applied with various flavors and at diverse temperatures, provide the route for the 3D structural characterization of molecules. See a selection among the large literature of reviews on that topic [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19], and some applications in e.g. proteomics and metabolics, [11, 20, 2, 9, 21, 22, 23, 24, 25, 26, 27, 28] glycoscience, [10, 11, 12, 16, 18, 29, 11, 30, 31, 18, 32] DNA science, [33, 34, 35, 36] atmospheric science and astrophysics [37, 38, 39].

Most of these experiments probe the 1000-2000 cm^{-1} and the 3000-4000 cm^{-1} fingerprint regions and hence mostly provide local structural information on the molecular systems such as covalent and hydrogen bond organizations, angular organizations. These spectral domains are however rather congested as soon as the size of the molecular system increases (congestion already appears in rather smallish systems, see for instance ref. [40, 41]), for complex organizations of molecular systems, and more generally, 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 possibly ambiguous, and sometimes even al-

most impossible [40, 41]. Spectral congestion is also the 'plague' of condensed phases (liquid phase and inhomogeneous interfaces). One solution against spectral congestion is to probe lower frequency modes in the far-IR/THz spectral domain (10-800 cm^{-1}, 24-0.3 THz, currently probed in gas phase spectroscopy). The supplementary temptation for the far-IR/THz is that larger amplitude and more collective motions are the ones giving rise to spectroscopic signals in this low frequency region, such as backbone torsional motions and hydrogen bond dynamics. They thus contain a wealth of structural information directly related to the 3D spatial organization that the 1000-4000 cm^{-1} local view lacks. Such motions are well-known and well-used to characterize crystals, semiconductors, liquids and biomolecules [42, 43, 44, 45, 46, 47, 48] in the condensed phase.

There are a few experimental set-ups for far-IR/THz gas phase spectroscopy using Free Electron Laser lights (FEL), see e.g. the following literature [39, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63], showing e.g. well-resolved and non congested far-IR/THz spectra for e.g. biologically relevant gas phase peptides and their complexes with water molecules to take examples from our own works in this domain [64, 60, 61, 62, 63, 65]. Our group has participated to the emergence of THz gas phase spectroscopy by theoretically assigning the experimental signatures and extract the 3D-structures of gas phase molecules using solely the THz fingerprints. These works have in particular shown the wealth of structural information contained in the THz signatures, how much essential the THz signatures are for a definitive 3D-structure assignment, and that DFT-based Molecular Dynamics simulations (DFT-MD) are crucial tools for revealing and deciphering the anharmonic large amplitude motions together with the coupled motions that exist in this domain.

The past 5 years have seen the MS-IR community embark on the challenge of high throughput analytical chemistry of bio-oligomers of relevance in biological processes. See e.g. [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. The highthroughput automatic application of MS-IR is however still blocked by limitations in the IR decoding stage. Decoding IR signatures (i.e. band- positions, intensities and shapes) into a microscopic 3D-structure is indeed non-trivial and cannot be accomplished from experiments alone for complex flexible (bio)-molecules, see e.g. reviews in [66, 67, 1, 4]. Theoretical

chemistry calculations have to be employed. Screening the right 3D-structure to which the theoretical IR spectrum matches the experiment is very much like 'finding a needle in a haystack'. The current computational procedure for 'finding the best spectroscopic match' is to: find an ensemble of energetically stable conformers, calculate the IR spectra of selected lower energy conformers, and find the best spectroscopic match to the experiment. See e.g. refs [68, 69, 70, 71] for general presentations. The calculation of IR spectra comes in a variety of approaches, [68, 69, 70, 71] i.e. static harmonic/anharmonic, dynamical anharmonic through e.g. DFT-based molecular dynamics simulations that we pioneered for action spectroscopies in the gas phase community 15 years ago. [72, 73, 68, 63] Whatever the chosen flavor and representation of the interactions for the PES calculation (electronic: ab initio, DFT, semi-empirical ; classical force fields ; a combination of these methods), such strategy is so computational- and time-consuming that this theoretical methodology still cannot be applied as it is for the high-throughput screening of 3D structures required in bio-analytical chemistry. We will discuss strategies and routes in section 3 to possibly reduce computational burdens and hence achieve possible high throughput spectra calculations.

In this opinion paper we focus solely on finite temperature molecular dynamics MD simulations for the calculation of anharmonic vibrational spectra and discuss some advances required to make MD-IR both more spectroscopically accurate and less computationally costly, so that MD-IR could be applied at large scale for the high-throughput screening discussed above. We refer the reader to our reviews in refs. [74, 75] and to refs [69, 71] for discussions over the advantages and limitations of MD-based dynamical anharmonic spectra calculations over static harmonic/anharmonic spectra calculations.

We and others have shown MD-IR, and especially DFT-MD-IR where the DFT (Density Functional Theory) electronic representation is used for the calculation of the interactions, to be a relevant theoretical strategy in order to include vibrational anharmonicities into theoretical spectra. DFT-MD dynamical spectra have been shown to provide remarkable match to experiments and hence allow definitive structural assignments. See for instance our two reviews on the topic [74, 75] and a selection of relevant papers for gas phase molecules and clusters from our group in [72, 76, 77, 78, 79, 61,

80, 64, 60, 81, 63, 65]. DFT-MD simulations for theoretical anharmonic vibrational spectroscopy is therefore the method of choice put forward in this opinion paper.

We review this method, make an assessment and overview of where we stand, and where to go in order to apply more generally MD-based theoretical spectroscopy to characterize large organic and bio-molecules, assemblies of these molecules, clusters with e.g. water, also to sample large time-scales that, all put together, will allow a better statistical theoretical approach for vibrational spectroscopy.

It is one task to calculate a precise anharmonic IR spectrum that can be matched to the experimental spectrum, and hence reveal the 3D structure(s) that are responsible for the spectral features, it is another issue to be able to extract the precise knowledge of the molecular motions/modes that give rise to each of the signatures. Decoding IR spectra into the active molecular motions and modes is essential to go beyond the 'simple' theory/experiment spectra matching for the recognition of the 3D-structure(s). The atomistic knowledge of the vibrational modes provides a detailed understanding of the motions and the energy associated to them, and their possible recognition/transferrability from one conformer to others. This knowledge can be used to create maps of the vibrational motions for systematic structural assignment, possibly even without relying on spectroscopic calculations. Refs. [75, 74, 61, 82, 83, 84, 85] have reviewed methods to extract vibrational modes from the gas phase dynamical trajectories and asssign the vibrational motions. See also ref. [86, 75, 87, 88] for the same strategies applied in the condensed phase. Briefly, methods are usually based on VDOS (Vibrational Density of States) or ICDOS (Internal Coordinates Density of States), through Fourier transforms of time-correlation functions of velocities (VDOS)/internal coordinates (ICDOS), see e.g. refs. [75, 61]. More ambitious methods to extract finite temperature 'effective normal modes' have also been developed, see for instance refs. [89, 90, 91, 92, 93, 94, 95, 96]. Although these latter are the (anharmonic) dynamical equivalent to the vibrational harmonic static normal modes, they are not so easy to apply in a systematic way, therefore they are (still) not very much in wide-spread use. Bowman *et al.* [95, 96] have also developed a strategy based on the vibrational excitation of the harmonic modes, followed in time with the dynamics, leading to the

assignment of (anharmonic)) vibrational modes.

Our group has recently developed an entirely different theoretical method for assigning vibrational modes based on Graph Theory algorithms.[65] To our best knowledge, this is the first time Graph Theory has been employed to that end. This method will be reviewed in section 4.

Our paper reviews the MD formalism for vibrational spectroscopy in section 2.1, including a discussion of possible issues 2.2, and briefly highlights nuclei quantum MD in section 2.3. Section 2.4 reviews our simplification for the MD-IR signal calculation for a better convergence of the signal, that also allows a mixed representation of the MD trajectory and IR calculation, which brings us to the next discussion on the representation of the interactions in MD spectroscopic simulations in section 2.5.

Section 3 discusses some directions that we believe are very promising for improving MD-based spectroscopy calculations and extend their application to large molecular systems, and to large time- and conformational-sampling. We have made choices for the topics and discussions, thus this section is by no means exhaustive.

Section 4 deals with the assignment of the vibrational modes, where we especially put forward our new theory based on graph theory algorithms. Graph theory will also be present as the main algorithmic means in section 5 where we present our own work for the automatic identification of 3D-structures along MD trajectories.

In all these discussions, while citing the literature we also take the liberty to highlight and discuss in more details some of our own works.

2. MD-based dynamical anharmonic spectroscopy: a short review of the formalism

We and others have reviewed (DFT-)MD simulations for vibrational spectroscopy, see e.g. [75, 74, 86, 97]. Here we give a brief (biased) selection of papers applying finite temperature MD gas phase IR, Raman and VCD spectroscopies [86, 97, 73, 68], IR/Raman spectrocopies of liquids, [98, 86, 99] SFG (Sum Frequency Generation) spectroscopy of (aqueous) interfaces [87, 88, 100, 101, 102], that show the versatility of MD-based theoretical spectroscopy. We take advantage of this opinion paper to

review a few key issues related to MD-based vibrational spectroscopy and discuss some possible issues and caveats.

2.1. Theoretical background

Within the well-known time-correlation function formalism from linear response theory [103, 104], an IR absorption spectrum is calculated by:

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

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 $\langle\ldots\rangle$ refers to the equilibrium time correlation function. This formula is for a classical nuclei representation, the quantum correction factor $\beta\hbar/(1-exp(-\beta\hbar\omega))$ has thus been applied to correct the classical line shape[105] in equation 1.

The advantage of the linear response theory for theoretical spectroscopy is that there are no (harmonic) assumptions on the calculation of the dipole moment. Coupled with MD simulations at finite temperature, the fluctuations of the dipole moment over time reflect the subtle changes in the electronic distribution over the molecule(s) as the 3D-conformations evolve with time and with the surrounding interactions. Therefore, both the exploration of the potential energy surface (PES) for the sampling of the conformations over time and of the dipole fluctuations at finite temperature take into account anharmonicities in the final vibrational calculated spectrum. As will be rediscussed in section 2.2 the choice of the temperature for the MD simulation is one central element. The shape and broadening of the IR bands result from the underlying conformational dynamics at a given temperature and from the anharmonicities in the mode-couplings. Whenever isomerization and/or proton transfers occur over time, these events together with all intermediate/transient conformations explored over time are naturally taken into account into the final calculation of the IR spectrum in equa-

tion 1. This participates to the final shaping of the IR bands.

Equivalent formula based on the linear response theory can be written for a Raman spectrum, for a VCD, for SFG, see refs. [86, 97, 73, 106, 88, 107], where polarizability tensors (Raman, SFG), magnetic moments (VCD), now enter into the correlation function. Cross-correlations involving some of these properties are found for the SFG and VCD signals.

2.2. Discussions on some possible issues on MD-based vibrational spectra

Within the linear response formalism any dynamical vibrational spectrum is calculated through a Fourier transform of a time-correlation function of a given property (i.e. dipole, polarizability, ...). MD is, by construction, the modeling tool to accumulate the time evolution of the properties entering into the time-correlation function. One central element is thus the length of the trajectory over which the time-correlation function is calculated in order to ensure convergence of the calculated spectrum. By construction, a time-correlation function has to tend to zero at 'long time scales' which can be hard to achieve over rather short time-scale DFT-MD trajectories for an isolated gas phase molecule. The zero-limit of any time-correlation function of a property $\mathscr{A}(t)$ measures the time needed for the loss of the information in $\mathscr{A}(0)$ known at the initial time t=0 of the dynamics. This decorrelation time is dependent on the $\mathscr{A}(t)$ property and on each molecular system. It is easy to get this zero limit for condensed phase systems simply because the time-correlation is averaged (sampled) over all the molecules of e.g. the liquid. To recover such statistical sampling for an isolated molecule one has to either accumulate a very long trajectory (beyond 100 ps time-scale), which is unreasonable in DFT-MD, or simulate a sufficient number of trajectories that differ for their initial conditions (positions and velocities of the atoms) and average the time-correlation functions or their fourier transforms over all these trajectories. This is typically what we have done for DFT-MD-IR of the THz spectroscopy of small peptides in refs. [81, 108, 109, 65]. See also a related discussion in ref. [88] for aqueous solid interfaces and convergence of the DFT-MD-SFG spectra over the length of the trajectories. In this convergence issue, the most critical point is the convergence of the

band-intensities and band-shapes. Band-positions are usually converged within 2-5 ps trajectories, even for isolated molecules, but the difficulty in reaching equipartition for gas phase molecules necessitates typically at least 30-50 ps trajectories together with statistics over several trajectories for the intensities and shapes of the bands to be converged (i.e. not changing anymore with the increase in trajectory time-length and/or statistical sampling). The fewer number of degrees of freedom in the molecular system the longer these trajectories have to be accumulated; some small isolated molecular systems unfortunately never reach equipartition. The difficulties in converging band-intensities and band-shapes can hinder the theory-experimental match. Such difficulties are released in condensed phase systems thanks to the statistical sampling of properties that are averaged over a large number of molecules.

The choice of the temperature in the MD trajectory has an influence on the positions of the vibrational bands. To make a 1-to-1 comparison to an experiment, the temperature in the MD simulation has to match the experimental one. In our past investigations, we have either made room-temperature DFT-MD trajectories to calculate IR spectra for the interpretation of room-temperature IR-MPD and IR-PD experiments,[73, 68, 79, 110, 78, 72, 111] or 50 K temperature DFT-MD trajectories for the interpretation of low-temperature IR-UV ion dip experiments.[81, 108, 109, 65] These dynamical spectra have shown rather good to excellent agreements with the experimental band-positions. The low 50 K temperature in some of our DFT-MD trajectories was in particular necessary for the assignment of low temperature spectroscopic experiments, not only for the band-positions but also for the band-shapes that are very narrow in the experiments. Increasing the temperature in the DFT-MD would broaden the theoretical bands, which could impede the theory-experiment assignment. The increase in temperature in the DFT-MD could also induce conformational isomerizations that are not probed at the lower temperature in the experiment.

For a given 3D-structure of a molecule, as temperature is increased each vibrational oscillator in the molecule will explore/sample more anharmonic motions, which has as immediate consequence to red-shift the frequency of that oscillator. The choice of the temperature thus controls the extent of vibrational anharmonicity probed in each

oscillator, it controls the absolute position of the oscillator's band, therefore the careful choice of the temperature to be made for the MD. However, mode couplings is the other source of anharmonicities, these couplings are less affected by the choice of temperature in the (classical nuclei) MD simulations. Mode couplings being the essence of the far-IR/THz vibrational modes, our 50 K trajectories of gas phase peptides have typically shown that these anharmonicities are very well accounted for, see e.g.[81, 108, 109, 65].

An issue that is however remaining with classical nuclei (DFT-)MD simulations is the 'classical temperature' versus the 'quantum temperature' of the nuclei, i.e. the ZPE-Zero Point Energy contained in the oscillators. The classical treatment of the nuclei at e.g. room temperature will not lead to the same sampling of the anharmonicities in the motions of the X-H oscillators as in the quantum treatment of the nuclei, to the extent that the amount of 'classical vibrational red-shift' will be insufficient in comparison to the 'quantum red-shift'. The difference between classical and quantum vibrational red-shift will depend on each X-H vibrator within the molecule of interest, and is therefore hard to quantify *a priori*. At low classical MD temperatures, such as the 50 K trajectories accumulated by us for THz-IR spectroscopy of isolated peptides,[81, 108, 109, 65] classical and quantum temperatures of the oscillators are now rather comparable. The advantage of the (classical) MD-based vibrational spectrocospy at low temperature is that the couplings between the modes are still naturally included, therefore not only the large amplitude anharmonic motions are sampled but also the anharmonic couplings between the modes, which allows DFT-MD-IR in the THz/far-IR domain to be quantitatively accurate for the band-positions. We have found the dynamical vibrational bands to be within less than 10 cm^{-1} from their experimental counterparts.[81, 108, 109, 65]

See e.g. refs [112, 113] for similar discussions on gas phase molecules IR spectroscopy. See also Bowman al. [69] for some comparisons of anharmonic VSCF, VCI and dynamical MD spectra using the same potential energy and dipole surfaces for small molecules, and a discussion on band-positions from the MD-based trajectories. See also ref. [65] for discussions on temperature effects in classical MD trajectories

and their influence on band-positions.

One last comment. Beyond the temperature and sampling/convergence issues discussed above, the accuracy of a MD dynamical vibrational spectrum is primarily due to the quality of the representation used for the calculation of the interactions. The accuracy of the MD vibrational band-positions is arising from the internal motions of the (anharmonic) oscillators and to the interactions/couplings between these oscillators. These can be roughly speaking qualified as the intra-molecular interactions. In contrast, intensities of the bands are reflecting the intermolecular interactions, or in other words the charge fluxes between the atoms. The choice of ab initio/semi-empirical/force field representations for the calculations of these intra- and inter-molecular interactions is therefore crucial. We will come back to this discussion in section 2.5. Scaling factors that are commonly applied in harmonic spectra calculations globally account for the harmonic approximation together with the electronic representation and the basis set size. Scaling factors have also been made dependent on the frequency domain, with some scaling factors valid for the 2000-4000 cm^{-1} domain and other sets of scaling factors for the 0-1000 cm^{-1} domain. Scaling factors should not be applied to anharmonic dynamical spectra obtained from MD simulations as anharmonicities are already taken into account. They should also not be applied because they would have to account for the temperature issues that we have outlined above but possibly also for the ZPE issues discussed in the next section; such scaling factors would not be straightforward to develop.

2.3. Quantum nuclei MD and Zero-Point Energy in the vibrational modes

In this text (as in our works), we are solely treating nuclei as classical particles in the MD simulations. The trajectories thus miss the important quantum nature of light atoms as well as the zero-point energy (ZPE) in the vibrational modes. There are several methods available to include the quantum nature of the nuclei in the MD simulations. I refer the readers to some recent papers, e.g. [114, 115, 113, 116, 117, 112]. These methods however have a computational cost that himpers their widespread application.

The lack of ZPE can be a critical issue for the O-H, N-H, C-H groups high-frequency motions/modes, as the classical energy put into these motions at room temperature is small in comparison to the zero-point energy of these motions. The classical motions of these groups might thus be too small-amplitude, which hence shifts the position of the calculated stretching bands with respect to their experimental counterpart. Such an issue disappears for the large amplitude motions in the THz, where e.g. we obtained agreements within less than 10 cm^{-1} for band-positions compared to experimental values.[81, 108, 109, 65]

Semi-clasically *prepared* MD trajectories that include ZPE into the modes at the initial time of the classical nuclei dynamics can be done, hence taking into account the quantum representation of the nuclei in the preparation of the trajectory while propagating classically the equations of motions of the nuclei. See e.g. one of our work on that topic [118] and ref. [69, 119] for a related discussion. However, one issue with semi-classicaly prepared MD is the ZPE-leakage, well documented by e.g. Hase, Bowman, Manolopoulos *et al.*. Quantum mechanically each internal molecular mode must contain an amount of energy at least equal to its ZPE, but classical mechanics may allow vibrational energy to flow freely between all or a subset of the modes and, hence, does not preserve any ZPE constraint [120]. This is of course an error inherent of classical mechanics. Consequently, no matter how accurately one can initially (i.e. at time zero of the MD) assign the ZPE to each normal mode of a molecule, after a number of steps, the energies in these modes may fluctuate. The energy fluctuation between modes for a multimode Hamiltonian is caused by the mode-mode coupling. In the case of separable modes, the energies for these modes would be conserved. Without any control of the coupling term, it is possible for one mode to transfer its energy to other modes and to lose energy, and be of an energy less than the ZPE [121]. Bowman and co-workers [122] and Miller and co-workers [123] proposed a method to constrain the ZPE by changing the sign of the momentum when the energy of any mode reaches the ZPE. This method did prevent energy from going below the ZPE, however since the momentum change occurs instantaneously, it is equivalent to an infinite impulse that is perhaps too abrupt, and can cause noise in a classical correlation function. This can be problematic in the context of time-dependent correlation functions for vibrational spec-

troscopy. Another method has been implemented by Bowman and co-workers [124] to constrain the ZPE by smoothly eliminating the coupling terms in the Hamiltonian as the energy of any mode falls below a specified value. This still introduces errors in the correlation functions that are needed to calculate spectroscopic signals.

Semi-classical MD is intermediate between quantum nuclei MD and semi-classically prepared MD, intermediate not only in terms of theoretical representation but also in terms of computational cost. Semi-classical MD retains the quantum nature of the nuclei into the propagator of the dynamics and hence includes the ZPE in each vibrational mode of the system over the MD. For computational reasons, the trajectories are accumulated over short time scales (ps time-scale), the final vibrational spectrum of a molecular system is averaged over multiple (short) trajectories. Semi-classical MD can be coupled to any representation of the interactions, i.e. electronic or classical. Semi-classical MD based on the IVR (initial value representation) and applications to vibrational spectroscopy of gas phase molecules is in particular developed by Ceotto *et al.* [125, 126, 127] with recent successful applications to large and flexible (bio)molecules [128, 129]. They have demonstrated that semi-classical MD trajectories are free of ZPE leakage when employing the Herman-Kluk propagator. [130] Other complex spectroscopies, i.e. linear, non-linear/multidimensional and time-dependent, are modeled through semi-classical MD trajectories [131, 132, 133, 134].

2.4. Rewriting the IR signal with velocity correlation functions: the APT formalism for dynamical IR spectroscopy

As shown in section 2.1, a dynamical anharmonic vibrational IR spectrum is based on the Fourier transformation of a dipole time correlation function, i.e. on the knowledge of the time evolution of a molecular dipole moment. The strength of an electronic-based MD method is to calculate these dipole moments without pre-established/parameterized models. In 'on-the-fly' BOMD simulations, calculating dipole moments, however, remains computationally expensive. It is usually based on the Berry phase method or on the localization of the (delocalized) wave function into localized functions as is done with the Wannier center localization. [98, 135, 136, 137] Alternative strategies have also been recently developed. [94, 138]

We have developed in ref. [139] an alternative method that circumvents any such procedure, without any *a priori* parameterization model for the calculation of molecular dipoles. Our strategy is fully based on the DFT-MD trajectories combining Atomic Polar Tensors (APT) (both developed for static and dynamical spectra calculations) [90, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149] that take into account the charge and dipole fluxes that are essential ingredients for vibrational band-intensities, with the fast convergence of velocities time correlation functions (VDOS, see section 4.1). At variance with the dipole time correlation functions, a velocity time correlation function is known to reach the zero-limit within far less than $\sim$5 ps thanks to the isotropic character of the cartesian atomic velocities. With such a fast time-scale for the time correlation function, several short DFT-MD trajectories can be employed to calculate converged DFT-MD-IR spectra.

The derivations shown below for the IR spectroscopy can be straighforwardly transposed to Raman spectroscopy, changing the Atomic Polar Tensors that will appear in the derivation by Raman tensors. The same general formalism is used in our DFT-MD-SFG spectroscopic investigations of aqueous interfaces in refs. [87, 88, 106, 101], where a parameterization of the water APTs has been made and allows for fast spectroscopic calculations of very complex molecular systems in the condensed phase, at low computational cost.

In ref. [139], 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 (APTs). These latter contain the activity of the vibrational modes. The formalism is obtained by expanding the time derivative dipole moment vector into cartesian coordinates derivatives. Hereby and in ref. [65], we present the equivalent derivation using internal coordinates (ICs) labeled R_m (for any coordinate that belongs to the $3N-6$ set of non redundant internal coordinates). We define vector R is $R = [R_1, R_2, R_3, \ldots, R_m, \ldots, R_{3N-6}]^T$, where N is the total number of atoms in the molecular system). The time-derivative of the dipole moment now reads:

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

With this, one can rewrite equation 1 into:

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

where $\dot{R}_l$ is the velocity of the l^{th} IC. This equation is the mirror to the one established in ref. [139] in terms of cartesian coordinates. The IR spectrum is thus 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. These latter are the equivalent of the APTs expressed in cartesian coordinates within a transformation.

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:

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

One important issue in equation 4, also discussed in ref. [139], is that the final IR intensity $I(\omega)$ includes all the self- and cross-correlations between all internal coordinates. All motions in the dynamics are correlated, which is naturally included 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 ($\langle \dot{R}_m(t) \cdot \dot{R}_m(0) \rangle, \forall\ m = 1, \cdots, 3N-6$), and one to the cross-correlation contributions ($\langle \dot{R}_m(t) \cdot \dot{R}_l(0) \rangle, \forall\ m \neq l$):

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

All these equations involve $\frac{\partial \mu}{\partial R_m}$ components, while the APT components (i.e. $\frac{\partial \mu}{\partial r_m}$ with $r_m = x_m, y_m, z_m$) are defined, in cartesian coordinates. APTs can be calculated from strandard quantum chemical packages such as Gaussian [150] largely used in the gas phase community, one has therefore to transform 'cartesian' APTs into their equivalent in internal coordinates for eq. 5 to be applicable. We hence introduce the

vector $\xi = [x_1, y_1, z_1, x_2, \ldots, z_N]^T$ that collects the 3N cartesian displacements of the N atoms of the system. It is always possible to find a linear transformation such as:

$$R = B \cdot \xi \tag{6}$$

$$\frac{\partial \mu_u}{\partial R_i} = \sum_j A_{ji} \frac{\partial \mu_u}{\partial \xi_j} \tag{7}$$

where $B_{ij} = \frac{\partial R_i}{\partial \xi_j}$ and $A_{ji} = \frac{\partial \xi_j}{\partial R_i}$, μ_u is the u^{th} cartesian coordinate of the dipole moment vector μ. We adopted the Wilson definitions of internal coordinates[151] 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 the APT cartesian components, into the internal coordinates needed for equation 5.

In ref. [139] we have tested several schemes for adapting the (cartesian) APTs to the change in conformations along the MD trajectory, and hence correctly include conformational changes (e.g. isomeration, rotation of certain parts of the molecule, proton transfers, large amplitude motions, $\cdots$) over time into the DFT-MD-APT-IR modeling. The weighted linear combination of a few selected APTs adapted for molecular rotation has been found the most robust to accurately reproduce the 'exact' DFT-MD-IR spectrum from eq. 1. We present below the scheme in cartesian coordinates, it is straightforward to apply it to APTs expressed in internal coordinates.

Briefly, if we note $P(t)$ the cartesian APT tensor of the molecule at time t, a reasonable hypothesis is that $P(t)$ can be described by a linear combination of the atomic polar tensors ${P_j}^{ref}$ of a set of appropriately chosen reference structures $\{j\}$:

$$P(t) = \sum_j w_j(t) R_j(ref \rightarrow t) P_j^{ref}(ref) {R_j}^T(ref \rightarrow t) \tag{8}$$

where $w_j(t)$ is the weight of the j-th reference structure in the whole combination at time t. The rotation matrix $R(ref \rightarrow t)$ must be introduced in the equation to ensure the consistency of the internal reference system of the $\{j\}$ structures with the one of the molecule at time t along the trajectory. The reference structures can be chosen following different strategies. If the potential energy surface of the molecular system

is known, it is for instance reasonable to use the equilibrium conformations of the molecule as references, to which some transition points along known transition pathways can be added. A much simpler possibility is to randomly choose some structures explored along the trajectory and to use them as reference.

To evaluate the weights w_j in the combination, we follow a strategy similar to the one proposed by Mathias *et. al* [91] in a different context. We assumed a Gaussian distribution around reference structures: $\mathrm{w}_j(t) = \frac{exp(-\frac{d_j}{2\sigma_c})}{\sum_j exp(-\frac{d_j}{2\sigma_c})}$ where d_j is a properly defined 'distance' measuring the difference between the geometry of the molecule at time t and the reference structure j, and σ_c is the width of the Gaussian distribution. To define d_j, one possibility is to directly use cartesian coordinates, i.e., $d_j = ||R(ref \rightarrow t)X_j^{ref} - X(t)||$. However, this choice can generate numerical noise, because it includes changes over all the internal coordinates, such as vibrational displacements along bonds and valence angles, which should be rather independent on conformational changes. To avoid this problem, we introduced a distance based on selected internal coordinates, for instance based only on torsional angles: $\mathrm{d}_j(t) = \sqrt{\sum_k^{Nt}(\theta_k(t) - \theta_k^j)}$ where $\theta_k(t)$ is the k-th torsional angle of the molecule at step t of the trajectory, and θ_k^j is the corresponding torsional angle value for the reference structure j. In ref. [139], we showed that in most of the cases, even for very dynamical molecules, a small number N_t of torsional angles (even only one) allows to univocally characterize the different reference structures. Such definition can easily be generalized to any definition of internal coordinate or to any combination of internal coordinates that might be more appropriate to characterize the molecular dynamics, see e.g. ref [91].

Velocities in all MD codes are obtained in cartesian coordinates. Velocities of the internal coordinates are thus calculated by numerical derivation with typically a five points central difference algorithm that we have found of excellent accuracy for DFT-MD trajectories accumulated with a 0.4 fs time-step:
$\dot{R}(t) = \frac{-R(t+2\delta t)+8R(t+\delta t)-8R(t-\delta t)+R(t-2\delta t)}{12\delta t}$

In ref. [139], we have shown that a very small number of reference structures and thus of reference APTs is necessary for a high level accuracy of equation 4 for DFT-MD-APT-IR spectroscopy. For instance, the IR spectroscopy of a very dynamical

dipeptide that continuously changes conformation over time is accurately captured with only 2 reference APTs. Another example is the IR spectrum of the $Cl^- \cdots$methanol cluster in the low frequency domain where large amplitude intermolecular motions are probed : it is accurately obtained when using only 6 reference APTs. Also, a large peptide like Gramicidin requires only 6 reference APTs to capture the correct IR spectrum in the THz domain.

Therefore, one can see that this methodology has many advantages among which being computationnally cheap, efficient and accurate: a restricted number of APTs is indeed required, without any particular careful choice of the reference structures (i.e. minimum, transition state, any structure explored along the trajectory) ; these APT calculations are done once and for all ; short time-scale MD trajectories can be performed over which the VDOS/ICDOS Fourier transform of the velocities time correlation functions is converged, which allows a statistical calculation of IR spectra over several trajectories that could not be achieved with a signal based on the dipole time correlation function ; the band-intensities are now entirely in the hands of the APTs that modulate the VDOS intensities, by including the charge fluxes that are mandatory for accurate IR intensities. As a consequence of this later point: if the molecular system is thermalized by MD simulations in the NVT ensemble, one can expert a rather reasonable equipartition of the energy within all the modes to be achieved, even for smallish systems. The subsequent few pico-seconds NVE MD simulations for the IR spectroscopy would then keep this equipartition, which together with the APT modulation of the VDOS/ICDOS would provide band-intensities with an accuracy that could not be achieved through eq. 1. The final advantage of equation 4 for MD-IR spectroscopy is that one can choose, at will, the level of representation for the calculation of the APTs. The higher ab initio representation the more accurate the APTs, thus the more accurate the charge fluxes for the IR calculations, thus the more accurate the dynamical IR band-intensities. Such accuracy is achieved at low computational cost as the number of atomic polar tensors is extremely reduced, as discussed above.

This is opening the path for a mixed representation of methods for dynamical IR spectroscopy: the MD trajectory can be applied at any level of theory, from low com-

putational cost as obtained with force field MD or semi-empirical MD, to medium computational cost as obtained with DFT-MD, while the APTs can be obtained at a very high level of electronic representation, for instance CCSD(T) for the highest affordable. This will be rediscussed in section 3.2.

2.5. Level of representation for the interactions in the MD simulations

The level of representation for the atomic interactions in the MD simulations can, in principle, be chosen as one wishes. There are two choices in atomistic MD: a QM (quantum mechanical) representation of the interactions and a FF (Force Field) representation. A mixed QM-MM representation can also be used. By construction, the CG representation is not fully atomistic and thus cannot be applied for vibrational spectroscopy where both the covalent bond dynamics and inter-molecular dynamics are probed. The QM-MD can be accumulated 'on-the-fly' or on precomputed potential and dipole/polarizability surfaces with Born-Oppenheimer molecular dynamics. When this later is done, the precomputed surfaces are generally obtained at a very high quantum level. See for instance works by Bowman *et al.* [69].

While any quantum representation can be used for 'on-the-fly' QM-MD, the DFT (Density Functional Theory) electronic representation is certainly the computationnally most efficient electronic method and therefore the most used in the literature for dynamical spectroscopy of large molecular systems (nowadays a few 1000' atoms is reasonably possible) over 100' ps time-scale trajectories. Any higher electronic level will be applicable only to molecules composed of maximum 10-50 atoms, for trajectories restricted over few picosecond time-scales. The other consequence of the 'smallish time-scale' is that there will possibly be difficulties in achieving the convergence of the theoretical spectroscopic signal, as already discussed in section 2.2. However, we have shown in section 2.4 how to circumvent such issue. GGA functionals like BLYP or PBE are certainly the most used in DFT-MD simulations. Hybrid functionals as well as meta-GGA functionals would probably be of better accuracy, however at a higher computational cost.

While accumulating the trajectories of large molecular systems (100-1000 atoms) over $\sim$50-100 ps in DFT-MD is already an achievement, the calculation of the proper-

ties that enter in the time-correlation functions for IR, Raman or VCD spectroscopies is even more challenging. Here, the gas phase medium is beneficial; the condensed phases (liquids, solids, and their interfaces) are indeed more complex and challenging as they require either a systematic localization of the wavefunction (typically with the Wannier representation [135, 136, 137, 98, 86]) or with methods such as the Voronoi representation [94, 86] for the actual calculation of the dipole moments and polarizability tensors of individual molecules for IR and Raman spectroscopies.[98, 152, 86] These methods generally lead to a $\sim$4-times increase in the computational cost of the DFT-MD simulations. A slightly less expensive method has been developed by Kuhne *et al.* [153] where the Wannier localization and Wannier volumes around the atoms are used to build the molecular polarizabilities. Though it still relies on the Wannier localization of the wavefunction, it is avoiding the computationnally expensive stage that we applied in ref. [152] where the molecular polarizabilities were obtained by the application of an external field. In the case of gas phase molecules, no such localization procedure is required. However, we also refer the reader to section 2.4 where we showed how to calculate an IR spectrum without the actual calculation of dipole moments (the same holds true for the non necessity of calculating polarizability tensors for a Raman spectrum calculation).

FF-MD classical simulations are based on analytical expressions for the intra- and inter-molecular interactions (FF) with pre-established parameters. The typical FFs for organic and bio-molecules are AMBER, CHARMM, GROMOS, OPLS, AMOEBA. The intramolecular part of these FFs is generally modeled by harmonic forms of the bond and angle motions that represent 2- and 3-body terms, an anharmonic torsional term for 4-body dihedral motions, while the intermolecular part of these FFs is accounted by Coulomb and Lennard-Jones terms for atoms separated by more than three bonds within the molecule and/or for atoms belonging to distinct molecules. Roughly speaking, the intra-molecular interactions and the parameters entering into these are mostly responsible for the vibrational band-positions while the inter-molecular interactions/parameters are mostly responsible for the band-intensities. Intermolecular parameters also modulate the band-positions through the interactions with the surround-

ing environment, typically for vibrational bands related to hydrogen bonds. No matter how carefully any FF is parameterized, its accuracy for energetics and spectroscopy will be limited by the choice of the functional forms entering into the analytical expressions of the FF and by the parameterization. As said above, most of the common FFs in use for gas phase molecules and condensed phases are based on harmonic oscillators for the internal motions, which is of course a huge approximation for vibrational spectroscopy. At least Morse anharmonic potentials should be used to model these motions for improved spectroscopic accuracy, cross-terms between the internal motions should also be employed to achieve spectroscopic accuracy. Including these latter in a FF increases the complexity of the parameterization, which is therefore very seldom done. Also polarization should be included in the electrostatic part of the FFs for spectroscopic accuracy, while higher order multipolar terms to model electrostatic interactions should also be in principle included. This is also very seldom done, the reason being always the same, it is indeed extremely complex and cumbersome to parameterize such high level FFs. However these inclusions would also certainly improve the transferability of these FFs. Even more importantly, electrostatic terms that explicitly take care of charge and dipole fluxes [154] should be added in FFs for spectroscopic accuracy: such fluxes have been shown to be the most important ingredients in vibrational spectroscopy analyses [155, 156]. Such force fields are extremely rare, as they are complicated to develop and parameterize (already for one given class of molecules, let us not mention transferability issues), and furthermore can become computationally rather costly. [157, 158]

Very few FFs are therefore of spectroscopic accuracy, i.e. very few are able to predict reliable band-positions, band-intensities and band-shapes. In ref. [159], we have for instance shown the extent of agreement/disagreement between AIMD and FF-MD SFG (Sum Frequency Generation) spectra for a series of aqueous interfaces using simple and slightly more sophisticated FFs. Developments around the AMOEBA FF have been for instance performed by Clavaguera *et al.* in the gas phase community, leading to a FF of spectroscopic accuracy, see e.g. refs. [157, 160]. As always with FFs, developing a FF is a time-consuming task, while the transferability of the parameters from one molecular system to another is always questionable.

See section 3.1 for the state-of-the-art machine learning methodologies for developing FFs and/or dipole/polarizability properties that enter into the spectroscopic theoretical signals.

The alternative to the full QM electronic representation in MD simulations for spectroscopy is the semi-empirical (SE) representation, which can come in various flavors. To our best knowledge, most of the SE-MD simulations of vibrational spectroscopy using PM methods as the SE representation are for the liquid phase, especially the IR spectroscopy of liquid water or biomolecules immersed in water [161, 162]. Several IR spectroscopic investigations have been led by Rapacioli *and coll.* using DFTB (DFT Tight Binding) SE-MD dynamics for gas phase systems [163, 164, 165]. Other papers can be found for static harmonic spectra calculations based on SE representations, but harmonic spectra are not within the scope of this opinion paper.

3. MD-based dynamical anharmonic spectroscopy: some new directions

3.1. FF-MD-based dynamical spectra: Machine Learning as the central tool

Presumably one of the most exciting routes for improving FF-MD based dynamical vibrational spectroscopy is machine learning (ML), where one can either machine learn a classical FF or machine learn the properties that enter into the time-correlation functions for spectroscopy. For IR, one has therefore to machine learn molecular dipole moments. For Raman, one has to machine learn molecular polarizability tensors. The learning has to be done over relevant conformations that will be explored at the finite temperature of the MD simulations. The general goal of machine learning force fields is to achieve a very high accuracy that is comparable to first principles methods, naturally including many-body effects, that could generally be applicable to all types of bonding and atomic interactions (covalent bonds, electrostatic, van der Waals, ...), thus with a high degree of transferability to new atomic configurations and more generally and ideally transferable between molecular systems. See ref. [166, 167] for more general discussions. ML-FF-MD can hence be performed on very large molecular systems, over long time-scales, that would not be amenable to first principles MD simulations, while maintaining first principles accuracy. Another great advantage of

ML-FF-MD of first principles accuracy is the possibility to apply any of the methods listed in section 2.3 for treating quantum nuclei in these dynamics: the quantum MD is hence generated over a 'cheap' ML-FF-MD. See some works by Rossi al. [114] or Marx al. [168] along this line. It is not the purpose of this paper to enter into the details of the ML techniques employed, neither on unsupervised/supervised trainings, we refer the readers to the cited papers to learn more on these methodologies.

Pioneers and well advanced research groups in ML-FF within the recent years include the groups of e.g. Behler, Czanyi, Shapeev and Ceriotti [166, 167, 169, 113, 170, 171, 172, 173, 174, 107], with applications mostly done for the condensed phases. Energies and forces are the central workhorse of the ML-FF techniques, ML-FFs express the potential energy surface as a sum of atomic energies that are a function of the local environment around each atom. They differ in the description of these local environments and in the ML approach or functional expressions to map the descriptors of the PES. See e.g. ref. [167] for a review. See also ref. [175] for alternative routes for ML-FF based on global representations.

ML-FF have been mostly developed and applied to condensed phases, i.e. liquids, solids, and aqueous interfaces. However to date, these ML-FF-MD simulations have been very seldom applied for vibrational spectroscopy calculations. Recent ML-FF-MD of protonated water clusters have been published in ref. [168], there again without associated vibrational spectroscopy. The paper by Behler and Hermansson [176] of the aqueous ZnO_2 interface also makes use of ML-FF-MD for the sampling of this complex inhomogeneous system, but the O-H IR spectroscopy is not generated along such trajectories but it is *a posteriori* modeled with an analytical fitted anharmonic model.

Marquetand *et al.* [177] have pushed forward the ML to infrared spectra. In their pioneering work, they have coupled ML-FF dynamics with the simultaneous machine learning of dipole moments for the IR spectroscopy of gas phase molecules (methanol, n-alkanes, a tripeptide in their work). Molecular dipole moments were machine trained on DFT-MD simulations with techniques and algorithms similar to the training for energies and forces for the ML-FF. Excellent agreements of the ML-FF-dipole-MD dynamical spectra were hence obtained with the reference DFT-MD spectra (and ex-

periments). This opened a highly promising route for vibrational spectroscopy, i.e. accurate theoretical spectra at low computational cost. More recently, Rossi *et al.* [113] have continued on the same route for the vibrational spectroscopy of the porphycene gas phase molecule, i.e. not only machine learning the potential energy surface of porphycene but also machine learning its molecular dipole moment. This investigation furthermore shows that, with such methods, the quantum dynamics of the nuclei can be included at very low cost, allowing take the quantum nature of the nuclei into account not only for the dynamics but ultimately for the vibrational spectra, which is essential in systems containing light atoms and possible proton transfers like the porphycene molecule there investigated. All together, such a successful investigation opens the route to ML-FF-MD and ML-spectroscopic properties that directly enter the time-correlation functions required in the vibrational spectra calculations for large size- and time-scale simulations of gas phase molecules, including nuclei quantum effects and ZPE effects on the final spectra.

Instead of machine learning molecular dipole moments, Ceriotti *et al.* [178] have machine learned atomic charges or/and atomic dipole moments in order to construct and predict the molecular dipole moments of a large variety of gas phase molecules extracted from databases. This 'muML' model gives excellent agreements on molecular dipole moments (0.07 D in MAE) but the paper does not show any further application to IR spectroscopy. A similarly 'AlphaML' model for polarizability tensors can be seen in ref. [179]. Machine learned polarizability tensors have also been obtained in ref. [107] for molecular crystals and used for Raman spectroscopy calculations. Skinner and Corcelli [180] have also used machine learning for the transition frequencies and dipole derivatives of the O-H oscillators of water used in their vibrational maps for the vibrational spectroscopy of liquid water, furthermore noticeably increasing the accuracy of the spectroscopic-map approach.

This short review on the recent developments of machine learning for MD simulations and for spectroscopy in particular shows where the field is moving. We believe this is one promising avenue for IR and Raman theoretical spectroscopy of large size gas phase molecules and clusters, obtained over long MD trajectories that can safely sample conformational dynamics at finite temperature, achieve equipartition, include

quantum nuclei and ZPE effects at low computational cost, and ultimately reach convergence of band-intensities of the spectra, that all together will provide a more accurate interpretation of gas phase spectroscopic experiments (and any condensed phase spectroscopic experiments in general) at low enough computational cost (once the ML stage has been achieved). To come back to our discussion in the introduction on the high throughput decoding of IR spectra of gas phase molecules in MS-IR experiments, the low computational cost of ML-FF-MD and ML-FF-IR-MD is indeed one central element that might allow to achieve such high throughput.

3.2. APT and Raman tensor formalism for dynamical MD-IR and MD-Raman spectroscopy

Another route is the one that we have presented in section 2.4 where the 'computationnally cheap' VDOS/ICDOS is modulated by atomic polar tensors (APTs) for IR spectroscopy or by Raman tensors for Raman spectroscopy. We have already discussed in this section that the MD trajectories for the VDOS/ICDOS can be generated at the low computational cost FF level, which could in practice be a first-principle derived ML-FF (see section 3.1), while the APT/Raman tensors can be obtained at a very high quantum level over a very restricted number of reference geometries. We believe this is a very good alternative to ML-FF-ML-dipole-MD-IR (or to ML-FF-ML-polarizability-MD-Raman) discussed in section 3.1, as it reduces the number of ML stages to go through. One could also imagine to further ML APT and Raman tensors.

3.3. Multiple-time steps in ab initio MD

Schemes have been devised to accelerate *ab initio* molecular dynamics simulations (AIMD) using multiple-timesteps (MTS). MTS-AIMD relies on the idea that the full system force can be calculated at a low-level of theory, which is taken as a reference, and is subsequently corrected with the force difference from a higher-level representation. For instance, a GGA-DFT calculation can be corrected with forces arising from the higher hybrid level B3LYP, see e.g. ref [181] and refs. therein. Steele *et*

al. [181, 182] have for instance presented some developments for MTS-AIMD simulations coupling HF and DFT for the two electronic levels of representation. In particular, they provided firm theoretical foundations for such a scheme and accompanying algorithms. Ref. [182] shows vibrational spectra (VDOS spectra) obtained through MTS-AIMD trajectories on two prototype 'simple' gas phase molecules (an isolated water, a protonated hydrazine), with remarkable agreements for the band-positions and band-intensities with the reference VDOS. Speed-ups from 3 to 7 were obtained for the systems investigated in this paper [182]. MTS-AIM thus opens the route to accelerated AIMD that maintain the level of accuracy of the AIMD not only for the trajectories but also for the calculated observables like a VDOS spectrum. It opens the route to the modeling of large-size molecules. MTS-AIMD could be coupled to our schemes for IR/Raman spectra detailed in section 2.4 and discussed further in section 3.2, hence providing vibrational spectra accumulated over long time-scale MTS-AIMD trajectories for a better convergence of these spectra. Ceriotti *and coll.* [183] have developped this MTS-AIMD concept for quantum nuclei AIMD simulations, where the multiple time-steps allow a reduction in the computational burden of the quantum nuclei representation. This also opens the path to accelerated quantum nuclei AIMD that would allow the calculation of more accurate vibrational spectra. See our related discussion in section 2.3.

3.4. Scores for assessing the match and quality of theoretical spectra

Assessing the quality/agreement of theoretical vibrational spectra to the recorded experimental spectra is necessary but non trivial. Such agreement is usually done by visual inspection in terms of band-positions, band-intensities and (more rarely, though accessible from dynamical spectroscopy) band-shapes. A more quantitative assessment would be needed, especially if one wants to make an automatic comparison between the calculation and experiment, as would be needed in a high throughput spectroscopic probing of 3D-structures of molecules in analytical chemistry as discussed in the introduction of this paper. Sadly enough, there is still no established methodology to quantitatively compare theoretical and measured spectra. Three very recent state-of-the-art papers report the development of scores in order to achieve a quantitative as-

sessment of calculated spectra [184, 185, 186]. Only ref. [184] uses finite temperature MD simulations in the flavor of DFT-MD for the calculation of IR spectra of gas phase molecules, the two other investigations rely on static harmonic IR spectra (with ab initio, semi-empirical, tight-binding, and classical force fields). Ref. [184] compares several measures of similarities and scores (e.g. Euclidiean distance, RMSD, Earth Mover Distance, Pearson correlation coefficient, Spearman correlation coefficient) in order to assess which indicator(s) can be the best one(s), all of them including band-positions, -intensities and -shapes in the theory/experiment quantitative comparison. Beyond the choice of the quantitative measure technique/descriptor to achieve a quantitative comparison, there are central issues discussed in this paper for the baseline (in the experiment), scaling factors for band-positions and band-intensities in the calculations. Grimme *et al.* [185] have shown that any of their four chosen similarity measurements (some of them are in common with ref. [184]) can be used to sufficiently represent the similarity between two IR spectra. This conclusion has been obtained over a very large database of calculations/experiments (more than 7000 experimental references). The authors also show that the central issue of matching band-positions and intensities for the similarity measurement can be solved through scaling factors, with mass scaled values which were found superior to the usual linear scalings of the frequencies. In their investigation, van der Spoel *et al.* [186] made harmonic IR spectra calculations based on force fields representations, and used Pearson and Speaman's correlation coefficients for the similarity measure between experiment and calculations over a set of 700 gas phase molecular data. Beyond the rather deceptive difficulty in matching FF-based harmonic spectra to experiment, the authors have also a contrasted conclusion on the two investigated matching scores, which they both qualify as having 'benefits and drawbacks'. Band-intensities are especially found crucial for a quantitative matching. Pearson's correlation is found for instance to better indicate agreements of the most dominant features in the spectrum while Spearman's is better to represent approximate matches of the bands. It would be very interested to balance these investigations with long time-scale finite temperature trajectories where, as discussed earlier in this paper, convergence of the band-intensities could be achieved.

Similarity measures could also be obtained through graph theory algorithms that

will be presented in two other contexts in sections 4.2 and 5. This is a route that we are currently developing in our group.

4. Assignments of vibrational modes

As already emphasized in the introduction of this paper, it is one achievement to calculate anharmonic dynamical 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/Ramn bands. We and others have in the past developed theoretical methods to that end [89, 90, 91, 92, 93, 94, 95, 96]. We review in section 4.1 the vibrational density of states (VDOS) and internal coordinates density of states (ICDOS) which constitute two very common and easy tools for extracting the knowledge of the atomic and molecular vibrational motions. Both are unfortunately only qualitative: while the atomic participation is known, the percentages of individual motions are not quantitatively known ; furthermore, the intensity of the VDOS/ICDOS bands do not correspond to the active IR (or Raman) intensities, which limits the motional assignment of the active modes. "Effective normal modes" methods have been developed, see e.g. [82, 83, 84, 85], which are the equivalent of the normal mode analysis in static harmonic calculations however now including the mode couplings and temperature of the underlying MD trajectory. These methods are not so easy to converge in practice, which has prevented their widespread use.

We have developed a very different route to modes assignments from MD simulations, based on both the MD-APT-IR spectroscopy theory reviewed in section 2.4 of this paper coupled to graph theory algorithms, that we review in section 4.2: it provides a quantitative description of the anharmonic molecular modes without the drawbacks of other methods.

4.1. Assignments by VDOS or ICDOS

In molecular dynamics simulations, the interpretation of the infrared/Raman active bands into individual atomic displacements is traditionally and easily done using the vibrational density of states (VDOS) formalism. The VDOS is obtained through the

Fourier transform of the cartesian atomic velocity auto-correlation functions:

$$VDOS(\omega) = \sum_{i=1,N} \int_{-\infty}^{\infty} \langle \mathbf{v_i}(t) \cdot \mathbf{v_i}(0) \rangle \, \exp(i\omega t) \, dt \tag{9}$$

where i runs over all atoms of the investigated system. $\mathbf{v_i}(t)$ is the velocity vector of atom i at time t. As in equation 1, the angular brackets in equation 9 represent the statistical average of the correlation function. The VDOS spectrum provides all vibrational modes of the molecular system. However, only some of these modes will be either Infrared or Raman active, so VDOS spectra can by no means directly substitute IR or Raman spectra. Cross-correlations between the atomic velocities couls also be included in equation 9, following the spirit of equation 4 for the formalism of MD-APT-IR spectroscopy.

The VDOS can further be decomposed according to atom types, or to groups of atoms, or to chemical groups of interest, in order to get a detailed assignment of the vibrational bands in terms of individual atomic motions. This is done by restraining the sum over i in equation 9 to the atoms of interest only. Such individual signatures are easy to interpret in terms of movements for localized vibrational modes that involve only a few atomic groups, typically in the 3000-4000 cm^{-1} domain, but the interpretation of the VDOS becomes more complicated for delocalized modes or for highly coupled modes. In the far-IR/THz domain below 1000 cm^{-1} where such motions are activated, the Fourier transform of internal coordinates (IC) time correlation functions is better suited to describe the molecular motions: $\int_{-\infty}^{\infty} \langle IC(t) \cdot IC(0) \rangle \, \exp(i\omega t) \, dt$. This requires to describe the molecular system by an ensemble of non-redundant internal coordinates, which is non trivial.

As already highlighted, the disavantage of the VDOS/ICDOS is that there is no activity taken into account in the peaks intensities (IR or Raman activities). They 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 that we present in the next section, based on the combination of ICDOS with the APT rewriting of the IR signal shown in section 2.4. The same method can be developed for any vibrational spectroscopy, e.g. Raman and SFG for instance. See the discussion in section 2.4.

4.2. Graph Theory for vibrational spectroscopy assignment

We have developed in ref. [65] a new theoretical strategy for vibrational band assignments that combines our work reviewed in section 2.4 with Graph Theory algorithmic. Equation 5 has shown how to reconstruct any IR spectrum as a combination of the IR participation of the (non redundent) internal coordinates (ICs) that describe the molecular system of interest. With this we directly obtain the quantitative knowledge of the active participation of each intra- and inter-molecular motion involved in each band/vibrational mode. This knowledge is coupled with Graph Theory from computer science in order to automatically assign each IR band and provide the graph of connectivity between the ICs that are responsible for the IR activity of each band, including their percentage of participation. We hence are able to generate a graph of connectivity between ICs that participate to each vibrational mode/peak in terms of the percentage of participation of each IC but also in terms of the couplings between the ICs. The anharmonic modes probed in the finite temperature MD are thus quantitatively described by this method. The different elements of the theory were detailed in ref. [65], they are summarized in the following text.

From equation 5, one has:

$$IR(\omega_x) \quad = \quad \sum_{m,l} I_{IC_{ml}}(\omega_x) = \sum_{m} I_{IC_m}(\omega_x) + \sum_{m,l\neq m} I_{IC_{ml}}(\omega_x) \tag{10}$$

for each band of the $I(\omega)$ spectrum, where ω_x is the frequency at which the maximum amplitude of the band is located. The first term in eq. 10 gives the contributions arising from the self-correlations of the IC 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}:

$$w_{ml} = \frac{I_{IC_{ml}}(\omega_x)}{IR(\omega_x)} \tag{11}$$

With such formula, one can 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 every single IR band, we also know the percentage of participation of each $I_{IC_{ml}}$ into the final IR band (i.e. the w_{ml} weights in eq. 11 above).

This reconstruction hence provides 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. The procedure is applied between ω_x and $\omega_x \pm \delta\omega$ (where $\delta\omega$ is roughly the band-width at half heigth). Works presented in ref. [65] typically used $\delta\omega$=10 cm^{-1} for 50 K trajectories. It is straightforward to adapt $\delta\omega$ to the temperature of the trajectory if needed.

In practice, 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. These thresholds can be changed at will. 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. Delocalized modes in the far IR/THz domain (<800 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.

To have a direct and graphical view of the contributions of the ICs into the anharmonic modes of the IR spectrum, graph theory algorithmic from mathematics and computer science is used. Graph theory is especially useful to subsequently apply algorithmics such as e.g. similarity between graphs that will provide the similarity between the vibrational modes, or reveal e.g. the couplings between the modes.

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, immediately by eye, the possible IC components in the vibrational modes. Our conventions adopted for the graphs representations are reported in fig. 1. 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 and wagging IC motions, yellow for intermolecular hydrogen bond motions. The percentage of participation of the IC

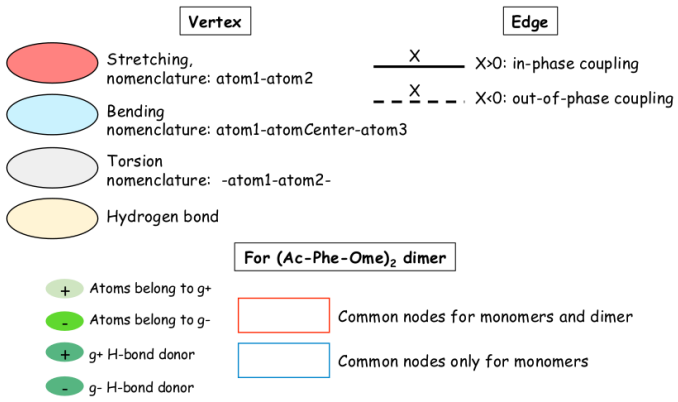

Figure 1: Conventions adopted for the graphs representations. The bottom of the figure provides some conventions for figure 4. Reproduced with permission from the Royal Society of Chemistry. [65]

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 that represent the cross-contributions of a pair of ICs in the IR spectrum, are traced with a black line, either a continuous line for in-phase couplings (positive values for F_E cross-terms $I_{IC_{ml}}(\omega_x)$) or a dashed line for out-of-phase couplings (negative values for cross-terms $I_{IC_{ml}}(\omega_x)$).

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

contained in the graph:

- Vertex - mm : $\mathrm{IC}_{mm}(\omega_x) \rightarrow F_V \rightarrow \frac{2\pi\beta}{3cV} \left| \frac{\partial \mu}{\partial R_m} \right|^2 \int_{\omega x_{min}}^{\omega x_{max}} dt \; e^{i\omega t} \langle \; \dot{R}_m(t) \cdot \dot{R}_m(0) \; \rangle$
- Edge - ml : $\mathrm{IC}_{ml}(\omega_x) \rightarrow F_E \rightarrow \frac{2\pi\beta}{3cV} \frac{\partial \mu}{\partial R_m} \frac{\partial \mu}{\partial R_l} \int_{\omega x_{min}}^{\omega x_{max}} dt \; e^{i\omega t} \langle \; \dot{R}_m(t) \cdot \dot{R}_l(0) \; \rangle$

Various aspects of the Graph Theory are highly advantageous when using this theory for modes assignments. As soon as the graph is plotted, on 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: 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. 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.

One 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. [187] for more details on isomorphism

and related algorithms. This is not presented further in this review/opinion paper.

One straighforward application and illustration of this theory is presented below for the prototypic gas phase water molecule. The Graph Theory based vibrational modes are obtained from a 15 ps DFT-MD 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 (DFT-MD run with the CP2K code [188]). The three graphs for the three IR active vibrational modes of the water are presented in figure 2. As can be immediately seen, the two higher frequency

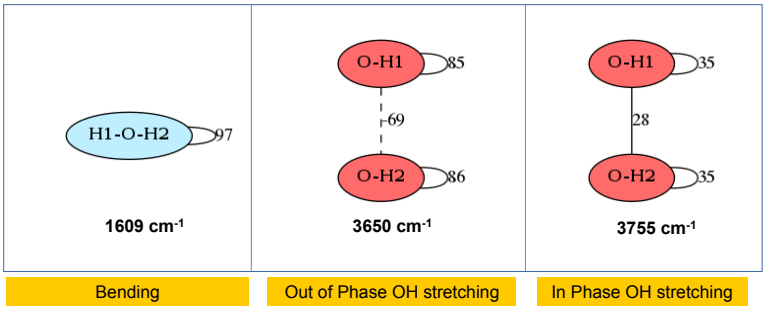

Figure 2: Graphs of the three active IR modes for a water gas phase molecule at 50 K. See fig. 1 for the nomenclatures. Reproduced with permission from the Royal Society of Chemistry. [65]

modes are the 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. $I_{IC_{ml}}(\omega_x)$ 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 streching bands. Restricting band assignments to the self-contribution only would hence not be correct, the correlation between the two stretching motions has to included for the final comprehension of the IR spectrum of the gas phase water molecule.

The next illustration of the Graph Theory for Vibrations (GT-Vib) is for a β-sheet model of a peptide made of two (Ac-Phe-OMe) monomers hydrogen bonded together. From geometry optimizations and DFT-MD simulations presented in ref. [65], one monomer in the dimer is found in its $\beta_L(g^+)$ conformation while the other is found in its $\beta_L(g^-)$ conformation. See figure 3 for illustrations of these conformations. In the rest of the text, $\beta_L(g^+)$ and $\beta_L(g^-)$ notations will be shortened as (g^+) and (g^-) respectively.

The Graph Theory based vibrational analysis is illustrated here only for torsional modes of the peptide backbones at $\sim$200 cm^{-1}, to show how these large amplitude, delocalized modes that involve several ICs, are easily captured by our methodology. See figure 4 for the graphs of the $\sim$200 cm^{-1} and $\sim$192 cm^{-1} modes for each of the monomers obtained from a 50 K DFT-MD trajectory of each isolated monomer, and the graph of the mode at 197 cm^{-1} for the $g^+ - g^-$ dimeric form obtained from a 50 K DFT-MD trajectory. Please refer to ref. [65] for the excellent match between the dynamical DFT-MD-IR and DFT-MD-APT-IR spectra of the monomer and dimer with the experimental IR-UV ion dip action spectroscopy.

The graph for the g^+ monomer nicely illustrates a 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-

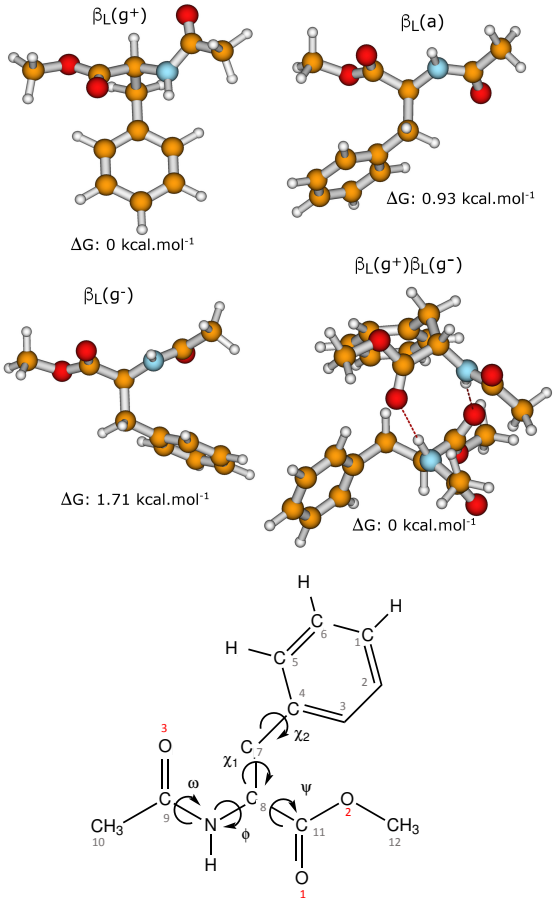

Figure 3: Ac-Phe-OMe and $(\text{Ac-Phe-OMe})_2$ peptides optimized geometries in the β_L general structural organisation. The "$\beta_L(\text{g}^+)$" conformation corresponds to the assigned one for the monomer (in our experimental conditions). The "$\beta_L(\text{g}^+)$-$\beta_L(\text{g}^-)$" dimer structure corresponds to the assigned one (in our experimental conditions). The scheme at the bottom shows the labelling of the atoms. Reproduced with permission from the Royal Society of Chemistry. [65]

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

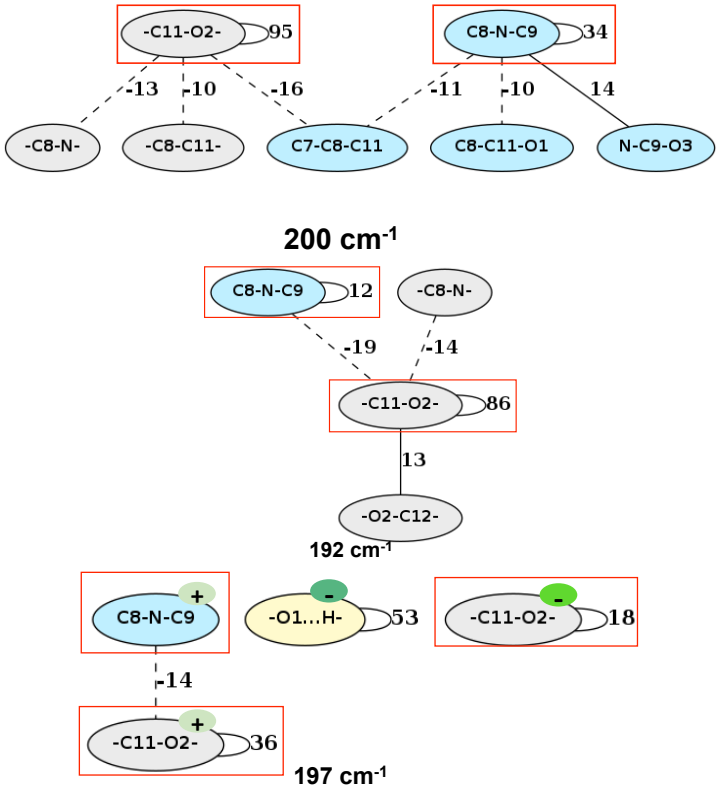

Figure 4: Graphs for the modes at 200 cm^{-1} (top) and 192 cm^{-1} (middle), respectively for g^+ and g^- monomers, and 197 cm^{-1} (bottom) for $g^+ - g^-$ dimer. See fig. 1 for the nomenclatures. Reproduced with permission from the Royal Society of Chemistry. [65]

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 are highlighted with 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 (g^+ and g^- strands are recognizable by the $+$ and $-$ symbols in the vertices in the figure), 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.

One last illustration of the GT-Vib/Graph Theory for Vibrations shows that this method contains the assignment of overtones and combination bands, that are impossible to extract from the VDOS/ICDOS signal because of the lack of activity into this signal. As illustration, we take the example of the Phenol gas phase molecule investigated experimentally and theoretically in ref.[80] in the far-IR/THz domain. Deuteration has demonstrated that the 588 cm^{-1} far-IR/THz spectral band recorded for Phenol is the overtone of the O-H torsional mode recorded at 309 cm^{-1}. In ref.[80], the DFT-MD simulations were not convincingly showing the presence of the overtone band in the calculated DFT-MD IR spectrum, because of equipartition issues (see section 2.2 in this paper for more discussion on this topic). When the DFT-MD simulations are redone with a substantial NVT period of thermalization together with more statistics (6 separate trajectories of 20 ps each) and using a slightly higher temperature of 200 K to possibly activate more the vibrational motions, the overtone of the O-H torsional motion can now be seen in the theoretical IR spectrum (although with a too low peak intensity compared to the experiment). The graphs of vibrations in figure 5 identify the same mechanical assignment for the fundamental band and for the overtone, i.e. C-C-O-H torsion (-C_4-O- as the dominant component in the graphs representation). Similar results can be obtained for combination bands.

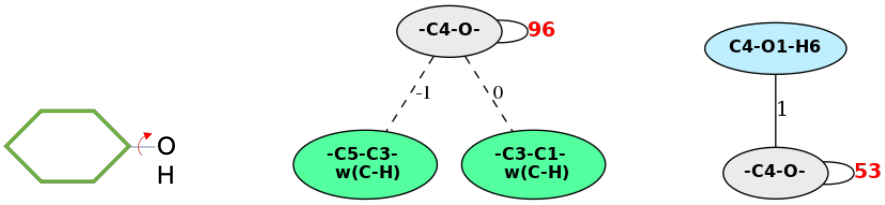

Figure 5: Graphs of vibrations for the fundamental O-H torsion band of the gas phase phenol molecule (middle) and for its overtone (right) extracted from 200 K DFT-MD trajectories.

5. Graph Theory for the automatic recognition of structures over time

We would also like to illustrate Graph Theory developments for the automatic recognition/identification of 3D-molecular structures over time in MD simulations and take the example coming from our own works in ref. [189]. We highlight a few points below, showing the versatility of the method for identifying the changes in isomeric conformations over time and for recognizing the breaking of covalent bonds and the occurence of the subsequent fragments of the molecule. Not shown here is the transferrability of our methodology from gas phase molecules, to clusters, to condensed phase systems. This is shown in ref. [189]. Others have also developed graph theory algorithms for the statistical analysis of 3D-structures in MD trajectories. Clark *et al.* [190, 191, 192, 193], Pastor *et al.* [194], Tenney and Cygan [195], Choi *et al.* [196], Pietrucci *et al.* [197, 198] have been mostly interested in recognizing atomic and molecular cluster geometries explored during MD simulations (gas phase and liquid phase trajectories). Graph theory has also been used in the exploration of potential energy surfaces of gas phase molecules [199, 200] by Martinez-Nunes *et al.* Graph theoric tools for driving biased MD simulations have also been developed [197, 198].

The graphs in these works are relatively easy to analyse, they however lack chemical information (e.g. covalent bonds, hydrogen bonds, exchange of atoms in homogeneous clusters, ...) that might be relevant for a more detailed characterization of the structures. Many of the developed graph theoric tools for chemistry in the literature furthermore use adjacency and/or geodesic matrices in order to compare structures sampled over the MD trajectory, which might not be the most efficient method for recognizing identical structures where chemically identical atoms have been swapped while their ID number in the atoms list keep them non identical.

To go beyond these limitations, we have developed graph theoric based algorithms where the granularity of the target graphs is at the atomic level (as in the literature), i.e. one atom per graph vertex, with the central target information being on the hydrogen bond direction (i.e. acceptor/donor). An algorithm to solve graph isomorphism has been implemented, including a coloration scheme according to the chemical nature of each atom, such that the comparison of graphs keeps the chemical atomic information

needed to analyse the isomeric conformations of molecules and clusters. While the graphs are more complex at this granularity level than other graphs in the literature, they are however more precise for their use in chemical reactions dynamics as well as in hydrogen bonds dynamics in liquids, which are some of our goals.

We refer the reader to ref. [189] for all details on the graph theory algorithms developed, we highlight a few elements below, and will show one illustration.

Our algorithm is developed for analysing molecular dynamics trajectories and identify 3D conformations in terms of intermolecular bonds (through hydrogen bonding or electrostatic intermolecular interactions) and/or covalent bonds. These are the changes in conformations tracked along time using our graph theory algorithms. A molecular conformation is defined in terms of covalent and hydrogen bonds formed between the atoms, which definitions are based on geometrical criteria (see details in ref. [189]).

In graph theory, a molecular conformation can be viewed as a molecular graph: in our work, atoms are the vertices of the graph while covalent bonds are the edges. The originality of our model consists in using a mixed graph, *i.e.* containing edges and arcs. Hydrogen atoms are thus not represented in the graph, i.e. the covalent bonds involving these hydrogen atoms are not represented, and in order to identify the donor and acceptor atoms in any H-bond, a directed edge going from the donor to the acceptor is used.

In a formal way, a molecular conformation is a mixed graph $G =(V, E_C, A_H, E_I)$, with the following definitions:

- V is the set of all atoms of the system except the hydrogen atoms, where each atom is a vertex of G,

- $E_C = \{[a,b], a \in V, b \in V : [a,b]$ is a covalent bond$\}$, where each covalent bond represents an edge in G ($[a,b] = [b,a]$).

- $A_H = \{(a,b), a \in V, b \in V : (a,b)$ is a H-bond$\}$, where each H-bond represents an arc in G ($(a,b) \neq (b,a)$).

- $E_I = \{[a,b], a \in V, b \in V : [a,b]$ is an electrostatic interaction$\}$, where each electrostatic interaction represents an edge in G ($[a,b] = [b,a]$).

The function $\phi : V \rightarrow T = \{,,,,...\}$ is also defined, providing the chemical type of atoms.

For example, $\phi(a) =$ if the atom a is an oxygen atom.

To extract the conformational isomers that are explored along the trajectory, graph isomorphism is applied, to decide if two graphs are identical [201, 202, 203, 204]. Without entering into the mathematical and algorithmic details, two conformations are isomorphic if they have the same sets of covalent bonds, H-bonds, and electrostatic interactions by allowing interchanging atoms of the same chemical type. Figure 6 shows an example of two conformations that are found isomorphic (A and B) and one conformation (C) which is not isomorphic to any of them. Different methods have been developed in the literature for solving graph isomorphism [205, 201, 206, 207, 208]. We apply the Mckay method [202] that is considered as one of the most efficient in practice. It is based on the canonical labelling of graphs, which consists in placing the vertex labels in a way that does not depend on their initial labelling. Each graph has a unique canonical labelling. Two graphs are hence isomorphic if they have the same canonical labelling.

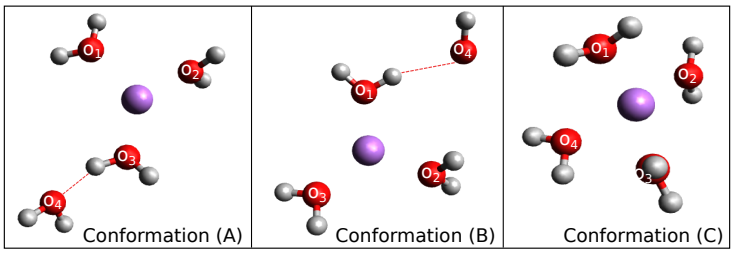

Figure 6: Example of isomorphic conformations. Conformations (A) and (B) are isomorphic by interchanging oxygen atoms labelled (O_1) and (O_3), while conformation (C) is isomorphic to neither (A) nor (B). Reproduced with permission from J. Chem. Phys. [189]

A change in the molecular conformation from time t to $t+\Delta t$ in the trajectory is identified if and only if there has been at least one change in a bond set, *i.e.* having $G_i = (V_i, E_{C_i}, A_{H_i}, E_{I_i})$ and $G_{i+1} = (V_{i+1}, E_{C_{i+1}}, A_{H_{i+1}}, E_{I_{i+1}})$ for the two consecutive conformations, the change can be related to one (or several) of the following instances: an appearance of a new covalent bond, disappearance of an existing covalent bond, appearance of a new H-bond, disappearance of an existing H-bond, proton transfer through a H-bond, appearance of a new electrostatic interaction, disappearance of an

existing electrostatic interaction. At the end of the analysis, the set of conformations explored along the trajectory is obtained, as well as the time sequence of their appearance and the mean residence time for each conformation.

To reduce the computational cost for identifying the H-Bonds, orbits related to each hydrogen atom have been introduced. An orbit is composed by a subset of atoms that are located at a given distance from the hydrogen atom lower than a cut-off distance $D_H \times \alpha$ (where D_H is the cut-off distance used for H-bonds and α is a coefficient with a default value of 3, this value was set after multiple tests on different trajectories and details have been presented in a related work [209]). Considering only the atoms within the orbits instead of all atoms of the system to compute H-bonds reduces the number of comparisons to be performed. This method requires that the orbits are not recalculated at each snapshot of the trajectory, otherwise that would be computationally costly. We therefore define a subset of reference snapshots where orbits have to be recomputed. To decide if a snapshot I_i is a reference snapshot, the displacements of atoms in snapshot I_i are analysed according to the current reference snapshot. If the whole displacement is greater than a cut-off distance $D_H \times (\alpha - 1)$ (same parameters as for the orbits), the reference snapshot is changed to I_i, and orbits are thus recomputed. The same philosophy is used to compute orbits and reference snapshots for covalent bonds and electrostatic interactions.

Graph isomorphism is known as a non-polynomial mathematical problem [210, 211]. The key component of the algorithm is to be able to reduce the number of isomorphism tests to be performed along the trajectory. Therefore, the isomorphism test at snapshot I_i is applied only if this snapshot is a reference snapshot, *i.e.* at least one orbit has changed. For the rest of the snapshots, a basic comparison of adjacency matrices is performed to decide if there has been a conformational change.

Once the whole trajectory has been analysed, the identified conformations are sorted out in terms of relevance for their time period of existence: only the conformations existing for a total time T_r over the whole trajectory are sorted out. By default, T_r is 5% of the total time of the trajectory, this is an easy changeable parameter.

Each conformation identified by the graph algorithm is represented by a mixed

graph consisting of (see illustration in figure 7):

- A set of vertices which represent (heavy) atoms of the system.

- A set of undirected edges which represent the identified covalent bonds or electrostatic interactions.

- A set of directed edges (arcs) which represent the identified H-bonds.

Figure 7 shows such a 2D-graph representation of a conformation of a peptide molecule.

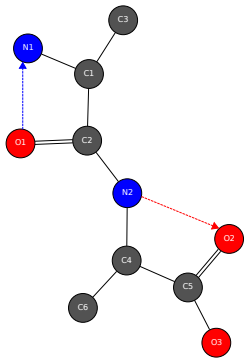

Figure 7: Example of a 2D representation of a conformation provided by graph theory analysis. Reproduced with permission from J. Chem. Phys. [189]

On a graph such as in figure 7, all the atoms, except the hydrogen atoms are represented in the vertices of the 2D-graph. This graph does not possess the 3D spatial representation of the molecular system, but instead it gives a direct view of covalent bonds and H-bonds in a simplified way. The covalent bonds are represented by black edges, the H-bonds by arcs from the heavy atom (tail of the arc) to the acceptor atom (head of the arc), the arc is red as long as no proton transfer occurs, it becomes blue when there is a proton transfer. The electrostatic interactions are represented by blue edges.

We illustrate these concepts on the analysis of the DFT-MD dynamics of the dipeptide NH_3^+-Ala_2-COOH, where the time explored conformations arise exclusively from the evolution over time of H-bonds and proton transfers (see ref. [189] for details). The four isomers of the peptide sampled over the analyzed trajectory are depicted in

figure 8 where both their 3D representations (top of the figure) and their 2D-molecular graphs (bottom) are reported.

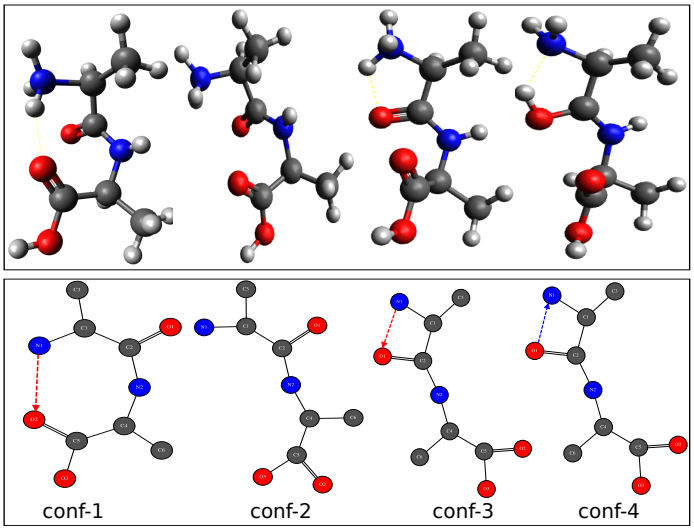

Figure 8: Schematic representation of the conformations of $^{+}_{3}-Ala_2-$ explored along the dynamics and analysed by the present graph theory algorithm. Top of the figure shows the 3D representations of the conformations. Bottom of the figure represents 2D simplified graphs of the conformations. Reproduced with permission from J. Chem. Phys. [189]

Added to that knowledge, figure 9 shows the graph of transition between the identified structures with the information on the frequency of transition between the identified 2D-molecular graphs/3D-structures. In grey on these graphs, one can also see vertices related to transitional conformations.

Hence, there are e.g. 60 transitions observed between the two conformations conf-2 and conf-3. Figure 9 shows not only conformations explored along the trajectory (white circles) but also shows the transitional states (gray circles). In this trajectory two transitional states have been identified. Analyzing the graph we observe forth-and-backwards isomerization between conformations conf-1 and conf-2, between conformations conf-2 and conf-3 and between conformations conf-3 and conf-4. The right side of figure 9 also provides the detailed changes occurring: in going from conformation conf-1 to conformation conf-2 there is the appearance of one H-bond (H-A), from conformation conf-2 to conformation conf-3 there is the disappearance of one

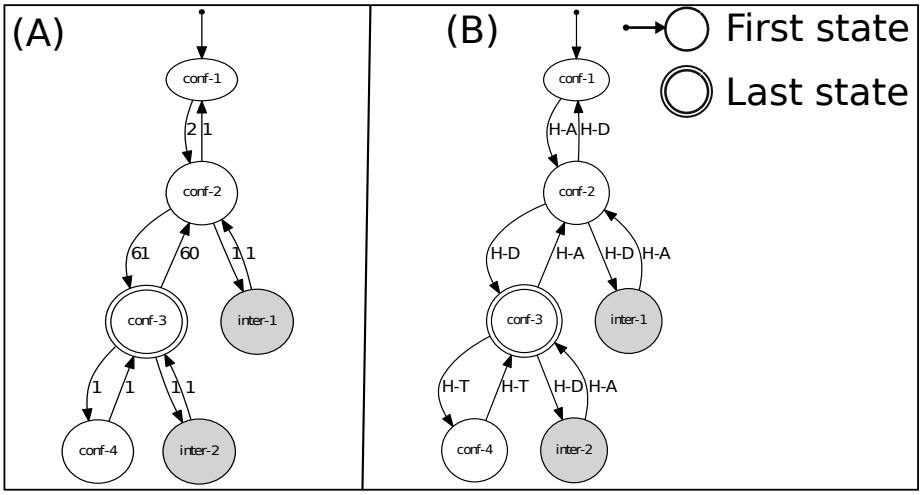

Figure 9: Graph of transitions for $^{+}_{3}-Ala_2-$ protonated dipeptide. Figure (A) represents the graph of transitions with frequencies of occurrence and figure (B) represents the graph of transitions with the changes that occurred. Conformations are represented in white circles and transitional states in gray circles. Reproduced with permission from J. Chem. Phys. [189]

H-bond (H-D), one proton transfer has occurred (H-T) from conformation conf-3 to conformation conf-4.

By analyzing the mean residence time for each observed conformation, conformation conf-3 is found the most observed with a total residence time of 2.0 ps over 4.0 ps trajectory (50% over the dynamics).

6. Conclusions

We have reviewed the current state-of-the-art of MD-based vibrational spectroscopy calculations for gas phase molecules and clusters. Some of the applications have been mentioned with citations from the literature and from our own works in this domain. We have chosen to present and discuss some new directions where the field is moving, that we believe will allow to perform even more challenging applications of gas phase MD-vibrational spectroscopy for the next decade.

7. Acknowledgements

I would like to acknowledge my collaborators and past/present students who worked with me to advance the field of MD-based vibrational spectroscopy. I will mention by name only a few of these people, and only the ones from the gas phase community. Over the years, I have had wonderful collaborations with experimentalists, among whom the late Prof Jean-Pierre Schermann and Prof James M. Lisy who were both great mentors in my professional carreer, and Prof Anouk M. Rijs who trusted me with DFT-MD THz theoretical spectroscopy. I want to thank Prof Dominique Barth for the unique works that we share in coupling Graph Theory and MD simulations, our recent collaboration has opened a new era in my works on gas phase molecular systems. I had the priviledge to enroll gifted PhD/Post-Doc students in my group, who shared/share my passion for gas phase spectroscopy, among whom Dr Daria R. Galimberti, Dr Sana Bougueroua, Dr Jérôme Mahé, PhD Vladimir Chantitch. Many thanks to them. Many thanks also to my other students and collaborators for all the works done together on MD-based spectroscopies of liquids and aqueous interfaces.

References

References

[1] A. M. Rijs, 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, J. Oomens, IR Spectroscopic Techniques to Study Isolated Biomolecules, Springer International Publishing, 2015, pp. 1–42.

[3] E. Gloaguen, M. Mons, Topics Curr. Chem. 364 (2015) 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, O. V. Boyarkin, Spectroscopic studies of cold, gas-phase biomolecular ions, Int. Rev. Phys. Chem. 28 (2009) 481–515.

[6] N. C. Polfer, J. Oomens, Mass Spectrom. Rev. 28 (2009) 468.

[7] N. C. Polfer, Chem. Soc. Rev. 40 (2011) 2211.

[8] C. N. Stedwell, J. F. Galindo, A. E. Roitberg, N. C. Polfer, Annu. Rev. Anal. Chem. 6 (2013) 267.

[9] A. Cismesia, M. Bell, L. Tesler, M. Alves, N. Polfer, Infrared ion spectroscopy: an analytical tool for the study of metabolites, Analyst 143 (2018) 1615.

[10] E. Mucha, M. Marianski, F. Xu, D. Thomas, G. Meijer, G. von Helden, P. Seeberger, K. Pagel, Unravelling the structure of glycosyl cations via cold-ion infrared spectroscopy, Nat. Comm. 9 (2018) 4174–4178.

[11] M. Kamrath, T. Rizzo, Combining ion mobility and cryogenic spectroscopy for structural and analytical studies of biomolecular ions, Acc. Chem. Res. 51 (2018) 1487–1495.

[12] I. Dyukova, E. Carrascosa, R. Pellegrinelli, T. Rizzo, Anal. Chem. 92 (2020) 1658–1662.

[13] P. Bansal, V. Yatsyna, A. AbiKhodr, S. Warnke, A. B. Faleh, N. Yalovenko, V. Wysocki, T. Rizzo, Anal. Chem. 92 (2020) 9079–9085.

[14] S. Warnke, A. B. Faleh, R. Pellegrinelli, N. Yalovenko, T. Rizzo, Faraday Discussions 217 (2019) 114–125.

[15] W. Hoffman, G. von Helden, K. Pagel, Ion mobility-mass spectrometry and orthogonal gas-phase techniques to study amyloid formation and inhibition, Curr. Op. Struct. Biology 46 (2017) 7–15.

[16] M. Lettow, M. Grabarics, E. Mucha, D. Thomas, L. Polewski, J. Freyse, J. Rademann, G. Meijer, G. von Helden, K. Pagel, Anal. Bioanal. Chem. 412 (2020) 533–537.

[17] M. Lettow, M. Grabarics, K. Greis, E. Mucha, D. Thomas, P. Chopra, G. Boons, R. Karlsson, J. Turnbull, G. Meijer, R. Miller, G. von Helden, K. Pagel, Anal. Chem. 92 (2020) 10228–10232.

[18] I. Compagnon, B. Schindler, G. Renois-Predelus, R. Daniel, Curr. Op. Struct. Biology 50 (2018) 171–180.

[19] C. Gray, I. Compagnon, S. Flitsch, Curr. Op. Struct. Biology 62 (2020) 1–11.

[20] C. Seaiby, A. Zabuga, A. Svendsen, T. Rizzo, Ir-induced conformational isomerization of a helical peptide in a cold ion trap, J. Chem. Phys. 144 (2016) 014304.

[21] R. Dunbar, J. Martens, G. Berden, J. Oomens, Binding of divalent metal ions with deprotonated peptides: Do gas- phase anions parallel the condensed phase?, J. Phys. Chem. A. 122 (2018) 5589–5596.

[22] J. Martens, G. Berden, H. Bentlage, K. Coene, U. Engelke, D.Wishart, M. von Scherpenzeel, L. Kluijtmans, R. Wevers, J. Oomens, Unraveling the unknown areas of the human metabolome: the role of infrared ion spectroscopy, J. Inherited Metabolic Disease 41 (2018) 367–377.

[23] E. Sinclair, K. Hollywood, C. Yan, R. Blankley, R. Breitling, P. Barran, Mobilising ion mobility mass spectrometry for metabolomics, Analyst 143 (2018) 4783–4788.

[24] D. Stuchfield, P. Barran, Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry, Curr. Op. Chem. Biology 42 (2018) 177–185.

[25] W. Sohn, S. Habka, E. Gloaguen, M. Mons, Unifying the microscopic picture of his-containing turns: from gas phase model peptides to crystallized proteins, Phys. Chem. Chem. Phys. 19 (2017) 17128–17142.

[26] E. Gloaguen, B. Tardivel, M. Mons, Gas phase double-resonance ir/uv spectroscopy of an alanine dipeptide analogue using a non-covalently bound uv-tag: observation of a folded peptide conformation in the ac-ala-nh2-toluene complex, Struct. Chem. 27 (2016) 225–230.

[27] S. Ishiuchi, Y. Sasaki, J. Lisy, M. Fujii, Ion-peptide interactions between alkali metal ions and a termini-protected dipeptide:modeling a portion of the selectivity filter in k+ channels, Phys. Chem. Chem. Phys. 21 (2019) 561–71.

[28] H. Ke, J. Lisy, Influence of hydration on ion-biomolecule interactions: M+(indole)(h2o)n (m=na, k; n=3-6), Phys. Chem. Chem. Phys. 17 (2015) 25354–25364.

[29] E. Mucha, A. G. Florez, M. Marianski, D. Thomas, W. Hoffman, W. Struwe, H. Hahm, S. Gewinner, W. Schollkopf, P. Seeberger, G. von Helden, K. Pagel, Glycan fingerprinting via cold-ion infrared spectroscopy, Angew. Chemie. Int. 56 (2017) 11248–11251.

[30] C. Masellis, N. K. M. Kamrath, D. Clemmer, T. Rizzo, Cryogenic vibrational spectroscopy provides unique fingerprints for glycan identification, J. Am. Soc. Mass Spectrom. 28 (2017) 2217–2222.

[31] A. D. Depland, G. Renois-Predelus, B. Schindler, I. Compagnon, Identification of sialic acid linkage isomers in glycans using coupled infrared multiple photon dissociation (irmpd) spectroscopy and mass spectrometry, Int. J. Mass Spectrom. 434 (2018) 65–69.

[32] I. Usabiaga, A. Camiruaga, A. Insausti, P. Carcabal, E. Cocinero, I. Leon, J. Fernandez, Phenyl-beta-d-glucopyranoside and phenyl-beta-d-galactopyranoside dimers:small structural differences but very different interactions, Frontiers in Phys. 6 (2018) 1–9.

[33] S. Boldissar, M. de Vries, How nature covers its bases, Phys. Chem. Chem. Phys. 20 (2018) 9701–9716.

[34] I. Chen, M. de Vries, From underwear to non-equilibrium thermodynamics: physical chemistry informs the origin of life, Phys. Chem. Chem. Phys. 18 (2016) 20005–20006.

[35] M. Ligare, A. Rijs, G. Berden, M. Kabelac, D. Nachtigallova, J. Oomens, M. de Vries, Resonant infrared multiple photon dissociation spectroscopy of

anionic nucleotide monophosphate clusters, J. Phys. Chem. B. 119 (2015) 7894–7901.

[36] R. van Outersterp, J. Martens, G. Berden, J. Steill, J. Oomens, A. Rijs, Structural characterization of nucleotide 5'-triphosphates by infrared ion spectroscopy and theoretical studies, Phys. Chem. Chem. Phys. 20 (2018) 28319–28330.

[37] J. Wagner, D. McDonald, M. Duncan, Mid-infrared spectroscopy of c7h7+ isomers in the gas phase: Benzylium and tropylium, J. Phys. Chem. Letters 9 (2018) 4591–4595.

[38] J. Wagner, D. McDonald, M. Duncan, Infrared spectroscopy of the astrochemically relevant protonated formaldehyde dimer, J. Phys. Chem. A. 122 (2018) 192–198.

[39] N. Heine, K. Asmis, Cryogenic ion trap vibrational spectroscopy of hydrogen-bonded clusters relevant to atmospheric chemistry, Int. Rev. Phys. Chem. 34 (2015) 1–34.

[40] I. Hunig, K. Kleinermanns, Conformers of the peptides glycine-tryptophan, tryptophan-glycine and tryptophan-glycine-glycine as revealed by double resonance laser spectroscopy, Phys. Chem. Chem. Phys. 6 (2004) 2650–2658.

[41] J. M. Bakker, C. Plützer, I. Hünig, T. Häber, I. Compagnon, G. von Helden, G. Meijer, K. Kleinermanns, Folding structures of isolated peptides as revealed by gas-phase mid-infrared spectroscopy, ChemPhysChem 6 (1) (2005) 120–128.

[42] D. Grischkowsky, S. Keiding, M. van Exter, C. Fattinger, Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors, J. Opt. Soc. Am. B 7 (10) (1990) 2006–2015.

[43] C. T. Nemes, C. Koenigsmann, C. A. Schmuttenmaer, Functioning photoelectrochemical devices studied with time-resolved terahertz spectroscopy, J. Phys. Chem. Letters 6 (16) (2015) 3257–3262.

[44] A. Bergner, U. Heugen, E. Bründermann, G. Schwaab, M. Havenith, D. R. Chamberlin, E. E. Haller, New p-ge thz laser spectrometer for the study of solutions: Thz absorption spectroscopy of water, Rev. Scient. Inst. 76 (6) (2005) 063110.

[45] Y. Xu, M. Havenith, Perspective: Watching low-frequency vibrations of water in biomolecular recognition by thz spectroscopy, J. Chem. Phys. 143 (17) (2015) 170901.

[46] T. Luong, P. Verma, R. Mitra, M. Havenith, Biophys. J. 101 (2011) 925.

[47] V. Nibali, M. Havenith, J. Am. Chem. Soc. 136 (2014) 12800.

[48] V. C. Nibali, S. Pezzotti, F. Sebastiani, D. Galimberti, G. Schwaab, M. Heyden, M.-P. Gaigeot, M. Havenith, Wrapping up hydrophobic hydration, J. Phys. Chem. Lett. 11 (2020) 4809–4816.

[49] M. Weichman, S. Debnath, J. Kelly, S. Gewinner, W. Schollkopf, D. Neumark, K. Asmis, Dissociative water adsorption on gas-phase titanium dioxide cluster anions probed with infrared photodissociation spectroscopy, Top. Catal. 61 (2018) 92–105.

[50] T. Esser, H. Knorke, K. Asmis, W. Schollkopf, Q. Yu, C. Qu, J. Bowman, Deconstructing prominent bands in the terahertz spectra of h7o3+ and h9o4+: Intermolecular modes in eigen clusters, J. Phys. Chem. Letters 9 (2018) 798–803.

[51] M. R. Fagiani, H. Knorke, T. K. Esser, N. Heine, C. T. Wolke, S. Gewinner, W. Schollkopf, M.-P. Gaigeot, R. Spezia, M. A. Johnson, K. R. Asmis, Gas phase vibrational spectroscopy of the protonated water pentamer: the role of isomers and nuclear quantum effects, Phys. Chem. Chem. Phys. 18 (2016) 26743–26754.

[52] 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, K. R. Asmis, Site-specific vibrational spectral signatures of water molecules in the magic

h3o+(h2o)20 and cs+(h2o)20 clusters, Proc. Natl. Acad. Sci. 111 (51) (2014) 18132–18137.

[53] M. R. Fagiani, X. Song, P. Petkov, S. Debnath, S. Gewinner, W. Schöllkopf, T. Heine, A. Fielicke, K. R. Asmis, Structure and fluxionality of b13+ probed by infrared photodissociation spectroscopy, Angew. Chemie. Int. 56 (2) (2017) 501–504.

[54] X. Li, P. Claes, M. Haertelt, P. Lievens, E. Janssens, A. Fielicke, Structural determination of niobium-doped silicon clusters by far-infrared spectroscopy and theory, Phys. Chem. Chem. Phys. 18 (2016) 6291–6300.

[55] A. Shayeghi, R. L. Johnston, D. M. Rayner, R. Schäfer, A. Fielicke, The nature of bonding between argon and mixed gold–silver trimers, Angew. Chemie. Int. 54 (36) (2015) 10675–10680.

[56] A. Shayeghi, R. Schäfer, D. M. Rayner, R. L. Johnston, A. Fielicke, Charge-induced dipole vs. relativistically enhanced covalent interactions in ar-tagged au-ag tetramers and pentamers, J. Chem. Phys. 143 (2) (2015) 024310.

[57] C. Kerpal, D. J. Harding, D. M. Rayner, J. T. Lyon, A. Fielicke, Far-ir spectra and structures of small cationic ruthenium clusters: Evidence for cubic motifs, J. Phys. Chem. C 119 (20) (2015) 10869–10875.

[58] M. Savoca, J. Langer, D. J. Harding, D. Palagin, K. Reuter, O. Dopfer, A. Fielicke, Vibrational spectra and structures of bare and xe-tagged cationic sinom+ clusters, J. Chem. Phys. 141 (10) (2014) 104313.

[59] V. J. F. Lapoutre, M. Haertelt, G. Meijer, A. Fielicke, J. M. Bakker, Communication: Ir spectroscopy of neutral transition metal clusters through thermionic emission, J. Chem. Phys. 139 (12) (2013) 121101.

[60] S. Jaeqx, J. Oomens, A. Cimas, M.-P. Gaigeot, A. M. Rijs, Gas-phase peptide structures unraveled by far-ir spectroscopy: Combining ir-uv ion-dip experiments with born–oppenheimer molecular dynamics simulations, Angew. Chemie. Int. 53 (14) (2014) 3663–3666.

[61] J. Mahe, D. J. Bakker, S. Jaeqx, A. M. Rijs, M.-P. Gaigeot, Mapping gas phase dipeptide motions in the far-infrared and terahertz domain, Phys. Chem. Chem. Phys. 19 (2017) 13778–13787.

[62] V. Yatsyna, D. J. Bakker, P. Salén, R. Feifel, A. M. Rijs, V. Zhaunerchyk, Infrared action spectroscopy of low-temperature neutral gas-phase molecules of arbitrary structure, Phys. Rev. Lett. 117 (2016) 118101.

[63] S. Bakels, M.-P. Gaigeot, A. Rijs, Gas phase spectroscopy of neutral peptides : Insights from the far ir domain, Chem. Rev. 120 (2020) 3233–3260.

[64] D. J. Bakker, A. Dey, D. P. Tabor, Q. Ong, J. Mahe, M.-P. Gaigeot, E. L. Sibert, A. M. Rijs, Fingerprints of inter- and intramolecular hydrogen bonding in saligenin-water clusters revealed by mid- and far-infrared spectroscopy, Phys. Chem. Chem. Phys. 19 (2017) 20343–20356.

[65] D. Galimberti, S. Bougueroua, J. Mahé, M. Tommasini, A. Rijs, M.-P. Gaigeot, 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, Faraday Discussions 217 (2019) 67.

[66] Advances in ion spectroscopy: From astrophysics to biology, Faraday Discussions 217 (2019) 1–648.

[67] Bond specific spectroscopy of peptides and proteins, Chem. Rev. 120 (2020) 3231–3630.

[68] M. P. Gaigeot, R. Spezia, Topics Curr. Chem. 364 (2015) 99.

[69] C. Qu, J. Bowman, Quantum approaches to vibrational dynamics and spectroscopy: is ease of interpretation sacrificed as rigor increases?, Phys. Chem. Chem. Phys. 21 (2019) 3397–3413.

[70] V. Barone, M. Biczysko, J. Bloino, Fully anharmonic ir and raman spectra of medium size molecular systems: accuracy and interpretation, Phys. Chem. Chem. Phys. 16 (2014) 1759–1787.

[71] T. Roy, R. Gerber, Vibrational self-consistent field calculations for spectroscopy of biological molecules: new algorithmic developments and applications, Phys. Chem. Chem. Phys. 15 (2013) 9468–9492.

[72] C. Marinica, G. Grégoire, C. Desfrançois, J. P. Schermann, D. Borgis, M. P. Gaigeot, J. Phys. Chem. A. 110 (2006) 8802.

[73] M. P. Gaigeot, Theoretical spectroscopy of floppy peptides at room temperature. a dftmd perspective: gas and aqueous phase, Phys. Chem. Chem. Phys. 12 (2010) 3336.

[74] M.-P. Gaigeot, R. Spezia, Theoretical Methods for Vibrational Spectroscopy and Collision Induced Dissociation in the Gas Phase, Springer International Publishing, 2015, pp. 99–151.

[75] M.-P. Gaigeot, Theoretical spectroscopy of floppy peptides at room temperature. a dftmd perspective: gas and aqueous phase, Phys. Chem. Chem. Phys. 12 (2010) 3336–3359.

[76] A. Cimas, T. D. Vaden, T. S. J. A. de Boer, L. C. Snoek, M. P. Gaigeot, J. Chem. Theor. Comput. 5 (2009) 1068.

[77] V. Brites, A. L. Nicely, N. Sieffert, M. P. Gaigeot, J. M. Lisy, Phys. Chem. Chem. Phys. 16 (2014) 13086.

[78] J. P. Beck, M.-P. Gaigeot, J. M. Lisy, Spectro. Chimica Acta A: Molecular and Biomolecular Spectroscopy 119 (2014) 12.

[79] V. Brites, J. M. Lisy, M. P. Gaigeot, J. Phys. Chem. A. 119 (2015) 2468.

[80] D. J. Bakker, Q. Ong, A. Dey, J. Mahé, M.-P. Gaigeot, A. M. Rijs, Anharmonic, dynamic and functional level effects in far-infrared spectroscopy: Phenol derivatives, J. Mol. Spectros. 342 (2017) 4 – 16.

[81] D. Bakker, A. Dey, D. Tabor, Q. Ong, J. Mahé, M.-P. Gaigeot, E. S. III, A. Rijs, Fingerprints of inter- and intra-molecular hydrogen bonding in saligenin-water

clusters revealed by mid- and far- infrared spectroscopy, Phys. Chem. Chem. Phys. 19 (2017) 20343.

[82] M. P. Gaigeot, M. Martinez, R. Vuilleumier, Mol. Phys. 105 (2007) 2857.

[83] M. Martinez, M. P. Gaigeot, D. Borgis, R. Vuilleumier, J. Chem. Phys. 125 (2006) 144106.

[84] M. Nonella, G. Mathias, P. Tavan, J. Phys. Chem. A 107 (2003) 8638.

[85] G. Mathias, S. D. Ivanov, A. Witt, M. D. Baer, D. Marx, J. Chem. Theor. Comput. 8 (2012) 224.

[86] M. Thomas, M. Brehm, R. Fligg, P. Vohringer, B. Kirchner, Computing vibrational spectra from ab initio molecular dynamics, Phys. Chem. Chem. Phys. 15 (2013) 6608–6622.

[87] S. Pezzotti, D. R. Galimberti, M.-P. Gaigeot, 2d h-bond network as the topmost skin to the air-water interface, J. Phys. Chem. Letters 8 (2017) 3133–3141.

[88] S. Pezzotti, D. Galimberti, M.-P. Gaigeot, Deconvolution of bil-sfg and dl-sfg spectroscopic signals reveal order/disorder of water at the elusive aqueous silica interface, Phys. Chem. Chem. Phys. 21 (2019) 22188–22202.

[89] M. Martinez, M.-P. Gaigeot, D. Borgis, R. Vuilleumier, Extracting effective normal modes from equilibrium dynamics at finite temperature, J. Chem. Phys. 125 (14) (2006) 144106.

[90] M.-P. Gaigeot, M. Martinez, R. Vuilleumier, Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations: application to small peptides, Mol. Phys. 105 (19) (2007) 2857–2878.

[91] G. Mathias, S. D. Ivanov, A. Witt, M. D. Baer, D. Marx, Infrared spectroscopy of fluxional molecules from (ab initio) molecular dynamics: Resolving large-amplitude motion, multiple conformations, and permutational symmetries, J. Chem. Theory. Comput. 8 (1) (2012) 224–234.

[92] M. Schmitz, P. Tavan, Vibrational spectra from atomic fluctuations in dynamics simulations. i. theory, limitations, and a sample application, J. Chem. Phys. 121 (24) (2004) 12233–12246.

[93] M. Schmitz, P. Tavan, Vibrational spectra from atomic fluctuations in dynamics simulations. II. solvent-induced frequency fluctuations at femtosecond time resolution, J. Chem. Phys. 121 (24) (2004) 12247–12258.

[94] M. Thomas, M. Brehm, B. Kirchner, Voronoi dipole moments for the simulation of bulk phase vibrational spectra, Phys. Chem. Chem. Phys. 17 (2015) 3207–3213.

[95] J. Bowman, X. Zhang, A. Brown, Normal-mode analysis without the hessian: A driven molecular-dynamics approach, J. Chem. Phys. 119 (2003) 646.

[96] M. Kaledin, A. Brown, A. Kaledin, J. Bowman, Normal mode analysis using the driven molecular dynamics method. ii. an application to biological macromolecules, J. Chem. Phys. 121 (2004) 5646.

[97] B. Kirchner, J. Blasius, L. Esser, W. Reckien, Predicting vibrational spectroscopy for flexible molecules and molecules with non-idle environents, Adv. theory Simul. (2020) 2000223–37.

[98] M. P. Gaigeot, M. Sprik, Ab initio molecular dynamics computation of the infrared spectrum of aqueous Uracil 107 (2003) 10344.

[99] F. Paesani, Getting the right answers for the right reasons: Toward predictive molecular simulations of water with many-body potential energy functions, J. Chem. Theor. Comput. 49 (2016) 1844.

[100] F. Tang, T. Ohto, S. Sun, J. R. Rouxel, S. Imoto, E. H. G. Backus, S. Mukamel, M. Bonn, Y. Nagata, Molecular structure and modeling of water-air and ice-air interfaces monitored by sum-frequency generation, Chem. Rev. 120 (8) (2020) 3633–3667.

[101] R. Khatib, E. Backus, M. Bonn, M. Perez-Haro, , M. Gaigeot, M. Sulpizi, Sci. Reports 6 (2016) 24287.

[102] G. R. Medders, F. Paesani, Dissecting the molecular structure of the air/water interface from quantum simulations of the sum-frequency generation spectrum, J. Am. Chem. Soc. 138 (11) (2016) 3912–3919.

[103] D. A. McQuarrie, Statistical Mechanics, University Science Books, 2000.

[104] R. Kubo, M. Toda, N. Hashitsume, Statistical Physics II, Vol. 31 of Springer Series in Solid-State Sciences, Springer Berlin Heidelberg, 1991.

[105] M. P. Gaigeot, M. Sprik, Ab initio molecular dynamics computation of the infrared spectrum of aqueous uracil, J. Phys. Chem. B. 107 (38) (2003) 10344–10358.

[106] S. Pezzotti, D. R. Galimberti, Y. R. Shen, M.-P. Gaigeot, Structural definition of the bil and dl: a new universal methodology to rationalize non-linear $\chi^{(2)}(\omega)$ sfg signals at charged interfaces, including $\chi^{(3)}(\omega)$ contributions, Phys. Chem. Chem. Phys. 20 (2018) 5190–5199.

[107] N. Raimbault, A. Grisafi, M. Ceriotti, M. Rossi, Using gaussian process regression to simulate the vibrational raman spectra of molecular crystals, New J. Phys 21 (2019) 105001–14.

[108] J. Mahé, D. Bakker, S. Jaeqx, A. Rijs, M.-P. Gaigeot, Mapping gas phase dipeptides motions in the far- infrared and terahertz domain, Phys. Chem. Chem. Phys. 19 (2017) 13778.

[109] J. Mahé, S. Jaeqx, A. Rijs, M. Gaigeot, Can far-ir action spectroscopy and bomd simulations be conformation selective?, Phys. Chem. Chem. Phys. 17 (2015) 25905.

[110] J. P. Beck, M.-P. Gaigeot, J. M. Lisy, Phys. Chem. Chem. Phys. 15 (2013) 16736.

[111] A. Cimas, T. D. Vaden, T. S. J. A. de Boer, L. C. Snoek, M. P. Gaigeot 5 (2009) 1068.

[112] S. D. Ivanov, A. Witt, D. Marx, Phys. Chem. Chem. Phys. 15 (2013) 10270.

[113] Y. Litman, J. Behler, M. Rossi, Temperature dependence of the vibrational spectrum of porphycene: a qualitative failure of classical-nuclei molecular dynamics, Faraday Discussions 221 (2020) 526–46.

[114] M. Rossi, H. Liu, F. Pasesani, J. Bowman, M. Ceriotti, Temperature dependence of the vibrational spectrum of porphycene: a qualitative failure of classical-nuclei molecular dynamics, J. Chem. Phys. 141 (2014) 181101.

[115] D. Thomas, E. Mucha, M. Lettow, G. Meijer, M. Rossi, G. von Helden, Characterization of a trans!trans carbonic acid-fluoride complex by infrared action spectroscopy in helium nanodroplets, J. Am. Chem. Soc. 141 (2019) 5815–5823.

[116] M. Rossi, V. Kapil, M. Ceriotti, Fine tuning classical and quantum molecular dynamics using a generalized langevin equation, J. Chem. Phys. 148 (2018) 102301.

[117] I. Poltavsky, V. Kapil, M. Ceriotti, K. Kim, A. Tkatchenko, Accurate description of nuclear quantum e!ects with high-order perturbed path integrals (hoppi), J. Chem. Theor. Comput. 16 (2020) 1128–1135.

[118] N.-T. V. Oanh, C. Falvo, F. Calvo, D. Lauvergnat, M. Basire, M.-P. Gaigeot, P. Parneix, Phys. Chem. Chem. Phys. 14 (2012) 2381.

[119] T. Esser, H. Knorke, K. R. Asmis, W. Schollkopf, Q. Yu, C. Qu, J. M. Bowman, M. Kaledin, J. Phys. Chem. Lett. 9 (2018) 798–803.

[120] G. Peslherbe, H.Wang, W. Hase, J. Chem. Phys. 100 (1993) 1179.

[121] X. Zhang, J. Rheinecker, J. Bowman, J. Chem. Phys. 122 (2005) 114313.

[122] J. M. Bowman, X. Zhang, A. Brown, J. Chem. Phys. 119 (2003) 646.

[123] W. Miller, W. Hase, C. Darling, J. Phys. Chem. 91 (1989) 2863–2868.

[124] Z. Xie, J. Bowman, Chem. Phys. Lett. 429 (2006) 355–359.

[125] M. Ceotto, Y. Zhuang, W. Hase, Accelerated direct semiclassical molecular dynamics using a compact finite difference hessian scheme, J. Chem. Phys. 138 (2013) 054116.

[126] G. D. Liberto, R. Conte, M. Ceotto, "divide and conquer" semiclassical molecular dynamics: A practical method for spectroscopic calculations of high dimensional molecular systems, J. Chem. Phys. 148 (2018) 014307.

[127] M. Buchholz, F. Grossmann, M. Ceotto, Application of the mixed time-averaging semiclassical initial value representation method to complex molecular spectra, J. Chem. Phys. 147 (2017) 164110.

[128] F. Gabas, R. Conte, M. Ceotto, On-the-fly ab initio semiclassical calculation of glycine vibrational spectrum, J. Chem. Theor. Comput. 13 (2017) 2378–88.

[129] F. Gabas, R. Conte, M. Ceotto, Semiclassical vibrational spectroscopy of biological molecules using force fields, J. Chem. Theor. Comput. 16 (2020) 3476–85.

[130] M. Buchholz, E. Fallacara, F. Gottwald, M. Ceotto, F. Grossmann, S. Ivanov, Herman-kluk propagator is free from zero-point energy leakage, Chem. Phys. 515 (2018) 231–235.

[131] J. Vanicek, Several semi-classical approaches to time-resolved spectroscopy, Chimia 71 (2017) 283–87.

[132] M. Wehrle, S. Oberli, J. Vanicek, On-the-fly ab initio semiclassical dynamics of floppy molecules:absorption and photoelectron spectra of ammonia, J. Phys. Chem. A. 119 (2015) 5685–5690.

[133] T. Begusic, J. Vanicek, On-the-fly ab initio semiclassical evaluation of vibronic spectra at finite temperature, J. Chem. Phys. 153 (2020) 024105.

[134] T. Begusic, J. Vanicek, On-the-fly ab initio semiclassical evaluation of third-order response functions for two- dimensional electronic spectroscopy, J. Chem. Phys. 153 (2020) 184110.

[135] N. Marzari, D. Vanderbilt, Maximally localized generalized wannier functions for composite energy bands, Phys. Rev. B 56 (20) (1997) 12847–12865.

[136] P. L. Silvestrelli, M. Parrinello, Water molecule dipole in the gas and in the liquid phase, Phys. Rev. Let. 82 (16) (1999) 3308–3311.

[137] P. L. Silvestrelli, M. Parrinello, Structural, electronic, and bonding properties of liquid water from first principles, J. Chem. Phys. 111 (8) (1999) 3572–3580.

[138] S. Luber, Local electric dipole moments for periodic systems via density functional theory embedding, J. Chem. Phys. 141 (2014) 234110.

[139] D. R. Galimberti, A. Milani, M. Tommasini, C. Castiglioni, M.-P. Gaigeot, Combining static and dynamical approaches for infrared spectra calculations of gas phase molecules and clusters, J. Chem. Theory Comput. 13 (8) (2017) 3802–3813, pMID: 28654750.

[140] W. B. Person, G. Zerbi, Vibrational intensities in infrared and Raman spectroscopy, Vol. 20, Elsevier Science Ltd, 1982.

[141] W. B. Person, J. H. Newton, Dipole moment derivatives and infrared intensities. I. Polar tensors, J. Chem. Phys. 61 (3) (1974) 1040–1049.

[142] C. Castiglioni, M. Gussoni, G. Zerbi, Handbook of Vibrational Spectroscopy, edited by J. Chalmers and P. Griffiths, John Wiley and Sons, Chichester, UK, 2001.

[143] J. C. Decius, An effective atomic charge model for infrared intensities, J. Mol. Spect. 57 (3) (1975) 348–362.

[144] W. T. King, G. B. Mast, Infrared intensities, polar tensors, and atomic population densities in molecules, J. Phys. Chem. 80 (22) (1976) 2521–2525.

[145] M. Gussoni, C. Castiglioni, M. Ramos, M. Rui, G. Zerbi, Infrared Intensities - from Intensity Parameters to an Overall Understanding of the Spectrum, J. Mol. Struct. 224 (1990) 445–470.

[146] R. L. A. Haiduke, R. E. Bruns, An atomic charge-charge flux-dipole flux atom-in-molecule decomposition for molecular dipole-moment derivatives and infrared fundamental intensities, J. Phys. Chem. A 109 (11) (2005) 2680–2688.

[147] A. Milani, C. Castiglioni, Modeling of Molecular Charge Distribution on the Basis of Experimental Infrared Intensities and First-Principles Calculations: The Case of CH Bonds, J. Phys. Chem. A 114 (1) (2010) 624–632.

[148] A. Milani, D. Galimberti, C. Castiglioni, G. Zerbi, Molecular charge distribution and charge fluxes from Atomic Polar Tensors: The case of OH bonds, J. Mol. Struct. 976 (1–3) (2010) 342–349.

[149] A. Milani, M. Tommasini, C. Castiglioni, Atomic charges from IR intensity parameters: theory, implementation and application, Theo. Chem. Acc. 131 (3) (2012) 1139.

[150] 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, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox., Gaussian 09 and revision c.02 and gaussian and inc. and wallingford ct and 2009.

[151] J. E. Wilson, J. C. Decius, P. C. Cross, Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra, new edition edition Edition, Dover Publications, 1980.

[152] M. Sulpizi, M. Salanne, M. Sprik, M. Gaigeot, J. Phys. Chem. Letters 4 (2013) 83.

[153] P. Partovi-Azar, T. Kuhne, Efficient "on-the-fly" calculation of raman spectra from ab-initio molecular dynamics: Application to hydrophobic/ hydrophilic solutes in bulk water, J.Comput. Chem. 36 (2015) 2188–2192.

[154] K. Palmo, S. Krimm, Electrostatic model for ir intensities in a spectroscopically determined molecular mechanics force field, J. Comput. Chem. 19 (1998) 754–768.

[155] D. Galimberti, A. Milani, C. Castiglioni, Charge mobility in molecules: Charge fluxes from second derivatives of the molecular dipole, J. Chem. Phys. 138 (16) (2013) 164115.

[156] D. Galimberti, A. Milani, C. Castiglioni, Infrared intensities and charge mobility in hydrogen bonded complexes, J. Chem. Phys. 139 (7) (2013) 074304.

[157] D. Semrouni, A. Sharma, J. Dognon, G. Ohanessian, C. Clavaguera, Finite temperature infrared spectra from polarizable molecular dynamics simulations, J. Chem. Theory. Comput. 10 (2014) 3190–3199.

[158] E. Kratz, A. Walker, L. Lagardere, F. Lipparini, J. Piquemal, G. Cisneros, Lichem: A qm/mm program for simulations with multipolar and polarizable force fields, J. Comput. Chem. 37 (2016) 1019–1029.

[159] O. Kroutil, S. Pezzotti, M.-P. Gaigeot, M. Predota, Phase-sensitive vibrational sfg spectra from simple classical force-fields molecular dynamics simulations, J. Phys. Chem. C. 124 (2020) 15253–263.

[160] F. Thaunay, J.-P. Dognon, G. Ohanessian, C. Clavaguera, Vibrational mode assignment of finite temperature infrared spectra using the amoeba polarizable force field, Phys. Chem. Chem. Phys. 17 (2015) 25968–77.

[161] M. Farag, M. Ruiz-Lopez, A. Bastida, G. Monard, F. Ingrosso, Hydration effect on amide i infrared bands in water: An interpretation based on an interaction energy decomposition scheme, J. Phys. Chem. B. 119 (2015) 9056–9067.

[162] C. Bistafa, Y. Kitamura, M. Martins-Costa, M. Nagaoka, M. Ruiz-Lopez, Vibrational spectroscopy in solution through perturbative ab initio molecular dynamics simulations, J. Chem. Theor. Comput. 15 (2019) 4615–4622.

[163] C. Dubosq, C. Falvo, F. Calvo, M. Rapacioli, P. Parneix, T. Pino, A. Simon, Mapping the structural diversity of c60 carbon clusters and their infrared spectra, A&A 625 (2019) L11.

[164] C. Dubosq, F. Calvo, M. Rapacioli, E. Dartois, T. Pino, C. Falvo, A. Simon, Quantum modeling of the optical spectra of carbon cluster structural families and relation to the interstellar extinction uv bump, A&A 634 (2020) A62.

[165] A. Simon, M. Rapacioli, J. Mascetti, F. Spiegelman, Vibrational spectroscopy and molecular dynamics of water monomers and dimers adsorbed on polycyclic aromatic hydrocarbons, Phys. Chem. Chem. Phys. 14 (2012) 6771–6786.

[166] J. Behler, First principles neural network potentials for reactive simulations of large molecular and condensed systems, Angew. Chem. Int. Ed. 56 (2017) 12828–12840.

[167] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Czanyi, A. Shapeev, A. Thompson, M. Wood, S. P. Ong, Performance and cost assessment of machine learning interatomic potentials, J. Phys. Chem. A. 124 (2020) 731–745.

[168] C. Schran, J. Behler, D. Marx, Automated fitting of neural network potentials at coupled cluster accuracy: Protonated water clusters as testing ground, J. Chem. Theor. Comput. 16 (2020) 88–99.

[169] G. Imbalzano, A. Anelli, D. Giofre, S. Klees, J. Behler, M. Ceriotti, Automatic selection of atomic fingerprints and reference configurations for machine- learning potentials, J. Chem. Phys. 148 (2018) 241730.

[170] D. Cole, L. Mones, G. Czanyi, A machine learning based intramolecular potential for a !exible organic molecule, Faraday Discussions 224 (2020) 247–265.

[171] V. Deringer, M. Caro, G. Czanyi, Machine learning interatomic potentials as emerging tools for materials science, Adv. Mater. 31 (2019) 1902765–2780.

[172] P. Dral, A. Owens, A. Dral, G. Czanyi, Hierarchical machine learning of potential energy surfaces, J. Chem. Phys. 152 (2020) 204110.

[173] B. Cheng, R. Griffiths, S. Wengert, C. Kunkel, T. Stenczel, B. Zhu, V. Deringer, N. Bernstein, J. Margraf, K. Reuter, G. Csanyi, Mapping materials and molecules, Acc. Chem. Res. 53 (2020) 1981–91.

[174] M. Ceriotti, Unsupervised machine learning in atomistic simulations, between predictions and understanding, J. Chem. Phys. 150 (2019) 150901.

[175] S. Chmiela, A. Tkatchenko, H. Sauceda, I. Poltavsky, K. Schutt, K.-R. Muller, Machine learning of accurate energy- conserving molecular force fields, Sci. Adv. 3 (2017) e1603015.

[176] V. Quaranta, M. Hellstrom, J. Behler, J. Kullgren, P. Mitev, K. Hermansson, Maximally resolved anharmonic oh vibrational spectrum of the water/zno(10-10) interface from a high-dimensional neural network potential, J. Chem. Phys. 148 (2018) 241720.

[177] M. Gastegger, J. Behler, P. Marquetand, Machine learning molecular dynamics for the simulation of infrared spectra, Chem. Sci. 8 (2017) 6924–35.

[178] M. Veit, D. Wilkins, Y. Yang, R. DiStasio, M. Ceriotti, Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles, J. Chem. Phys. 153 (2020) 024113.

[179] Y. Yang, K. U. Lao, D. Wilkins, A. Grisafi, M. Ceriotti, R. DiStasio, Quantum mechanical static dipole polarizabilities in the qm7b and alphaml showcase databases, Scientific Data 6 (2019) 152–161.

[180] A. Kananenka, K. Yao, S. Corcelli, J. Skinner, Machine learning for vibrational spectroscopic maps, J. Chem. Theor. Comput. 15 (2019) 6850–58.

[181] S. Fatehi, R. Steele, Multiple-time step ab initio molecular dynamics based on two-electron integral screening, J. Chem. Theor. Comput. 11 (2015) 884–898.

[182] R. Steele, Multiple-timestep ab initio molecular dynamics using an atomic basis set partitioning, J. Phys. Chem. A. 119 (2015) 12119–12130.

[183] V. Kapil, J. VandeVondele, M. Ceriotti, Accurate molecular dynamics and nuclear quantum effects at low cost by multiple steps in real and imaginary time: Using density functional theory to accelerate wavefunction methods, J. Chem. Phys. 144 (2016) 054111.

[184] B. von der Esch, L. Peters, L. Sauerland, C. Ochsenfeld, Quantitative comparison of experimental and computed ir-spectra extracted from ab initio molecular dynamics, J. Chem. Theor. Comput. 17 (2021) 985–995.

[185] P. Pracht, D. Grant, S. Grimme, Density functional theory methods for calculating gas-phase infrared spectra, J. Chem. Theor. Comput. 16 (2020) 7044–60.

[186] H. Henschel, A. Andersson, W. Jespers, M. Ghahremanpour, D. van der Spoel, Theoretical infrared spectra: Quantitative similarity measures and force fields, J. Chem. Theor. Comput. 16 (2020) 3307–15.

[187] S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, M.-P. Gaigeot, Graph theory for automatic structural recognition in molecular dynamics simulations, J. Chem. Phys. 149 (2018) 184102–15.

[188] Cp2k version 2.4.0, the cp2k developers group (2013). cp2k is freely available from http://www.cp2k.org/.

[189] S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, M.-P. Gaigeot, Graph theory for automatic structural recognition in molecular dynamics simulations 149 (2018) 184102.

[190] M. Hudelson, B. L. Mooney, A. E. Clark, Determining polyhedral arrangements of atoms using PageRank, J. Math. Chem. 50 (9) (2012) 2342–2350.

[191] B. L. Mooney, L. R. Corrales, A. E. Clark, MoleculaRnetworks: an integrated graph theoretic and data mining tool to explore solvent organization in molecular simulation, J. Comp. Chem. 33 (8) (2012) 853–860.

[192] A. Ozkanlar, A. E. Clark, ChemNetworks: a complex network analysis tool for chemical systems, J. Comp. Chem. 35 (6) (2014) 495–505.

[193] B. L. Mooney, L. R. Corrales, A. E. Clark, Novel analysis of cation solvation using a graph theoretic approach, J. Phys. Chem. B. 116 (14) (2012) 4263–4275.

[194] K. Han, R. M. Venable, A.-M. Bryant, C. J. Legacy, R. Shen, H. Li, B. Roux, A. Gericke, R. W. Pastor, Graph-theoretic analysis of monomethyl phosphate clustering in ionic solutions, J. Phys. Chem. B. 122 (2018) 1484–94.

[195] C. M. Tenney, R. T. Cygan, Analysis of molecular clusters in simulations of lithium-ion battery electrolytes, J. Phys. Chem. C. 117 (47) (2013) 24673–24684.

[196] J.-H. Choi, H. Lee, H. R. Choi, M. Cho, Graph theory and ion and molecular aggregation in aqueous solutions, Annu. Rev. Phys. Chem. 69 (2018) 125–149.

[197] F. Pietrucci, W. Andreoni, Graph theory meets ab initio molecular dynamics: atomic structures and transformations at the nanoscale, Phys. Rev. Let. 107 (8) (2011) 085504.

[198] F. Pietrucci, W. Andreoni, Fate of a graphene flake: A new route toward fullerenes disclosed with ab initio simulations, J. Chem. Theory. Comput. 10 (3) (2014) 913–917.

[199] E. Martínez-Núñez, An automated transition state search using classical trajectories initialized at multiple minima, Phys. Chem. Chem. Phys. 17 (22) (2015) 14912–14921.

[200] E. Martínez-Núñez, An automated method to find transition states using chemical dynamics simulations, J. Comp. Chem. 36 (4) (2015) 222–234.

[201] E. M. Luks, Isomorphism of graphs of bounded valence can be tested in polynomial time, J. Comput. Syst. Sci. 25 (1) (1982) 42–65.

[202] B. D. McKay, Practical graph isomorphism, Department of Computer Science, Vanderbilt University Tennessee, US, 1981.

[203] B. D. McKay, A. Piperno, Practical graph isomorphism, J. Symb. Comput. 60 (2014) 94 – 112.

[204] S. G. Hartke, A. Radcliffe, Communicating Mathematics - Chapter 8 : Mckay's canonical graph labeling algorithm, Vol. 479, American Mathematical Soc., 2009, Ch. 8, pp. 99–111.

[205] S. Sorlin, C. Solnon, A new filtering algorithm for the graph isomorphism problem, Proceedings of the Third International Workshop on Constraint Propagation and Implementation (2006) 93.

[206] H. L. Bodlaender, Polynomial algorithms for graph isomorphism and chromatic index on partial k-trees, J. Algorithms 11 (4) (1990) 631–643.

[207] P. T. Darga, K. A. Sakallah, I. L. Markov, Faster symmetry discovery using sparsity of symmetries, IEEE (2008) 149–154.

[208] T. Junttila, P. Kaski, Engineering an efficient canonical labeling tool for large and sparse graphs (2007) 135–149.

[209] D. Barth, S. Bougueroua, M.-P. Gaigeot, F. Quessette, R. Spezia, S. Vial, A new graph algorithm for the analysis of conformational dynamics of molecules, in: Proceeding in Information Sciences and Systems 2015, Springer, 2016, pp. 319–326.

[210] J. Kobler, U. Schöning, J. Torán, The graph isomorphism problem: its structural complexity, Springer Science & Business Media, 2012.

[211] L. Babai, A. Dawar, P. Schweitzer, J. Torán, The graph isomorphism problem (dagstuhl seminar 15511) 5 (12).