PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Page 1 of 12

Physical Chemistry Chemical Physics
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS

ARTICLE TYPE

www. rsc.org/xxxxxx | XXXXXXXX

Vibrational mode assignment of finite temperature infrared spectra using the AMOEBA polarizable force field†

Florian Thaunay,[a] **Jean-Pierre Dognon,**[b] **Gilles Ohanessian,**[a] **and Carine Clavaguéra**[*a]

The calculation of infrared spectra by molecular dynamics simulations based on the AMOEBA polarizable force field has recently been demonstrated [Semrouni *et al., J. Chem. Theory Comput., 2014, 10, 3190*]. While this approach allows access to temperature and anharmonicity effects, band assignment requires additional tools, which we describe in this paper. The Driven Molecular Dynamics approach, originally developed by Bowman, Kaledin et al. [Bowman *et al. J. Chem. Phys., 2003, 119, 646,* Kaledin *et al. J. Chem. Phys., 2004, 121, 5646*] has been adapted and associated with AMOEBA. Its advantages and limitations are described. The IR spectrum of the Ac-Phe-Ala-NH$_2$ model peptide is analyzed in detail. In addition to differentiation of conformations by reproducing frequency shifts due to non-covalent interactions, DMD allows visualizing the temperature-dependent vibrational modes.

1 Introduction

Gas phase infrared spectroscopy was developed and applied to biologically relevant molecules with the aim to decipher their structures based on the sensitivity of the frequencies to the environment of the chemical groups[1–5]. Both double resonance and photodissociation techniques were applied to both neutral and ionic species in the last decade. Most of the time, detailed interpretation of such experiments involves theoretical calculations. The infrared spectra are usually computed using static harmonic quantum mechanics calculations and band assignments rely on the main components of normal modes. The main drawback of this approach is the use of scaling factors to compensate for anharmonicity and the fact that calculations are restricted to 0 K, even though temperature effects are known to be significant. Currently, several approaches exist to account for anharmonicity in Hessian based non-dynamical methods[6–9]. They are accurate but generally limited to the study of small-size systems. Molecular dynamics (MD) simulations have been established as the method of choice to include directly in the calculation temperature and anharmonic effects such as frequency shifts, band broadening or band combinations[10–15]. Born-Oppenheimer (BOMD) and Car-Parrinello (CPMD) molecular dynamics have been successfully applied to gas-phase and solvated biomolecules, however they are limited to short propagation times of tens of picoseconds[16–22].

We have recently proposed the extension of the polarizable AMOEBA force field[23–25] to infrared spectroscopy of gas phase peptide models[26]. This approach was based on the successful combination of such a force field, that includes refined electrostatics and an explicit polarization term, and the Fourier transform of the dipole-dipole time correlation function during a finite temperature molecular dynamics. This methods allows to reproduce spectroscopic signatures that reflect the role of the environment due to hydrogen bonds or solvent molecules. Long-time simulations are accessible by classical molecular dynamics providing the possibility to see conformational dynamics signatures directly on the IR spectra. However when the system under study becomes of larger size, band assignments based on chemical intuition or static quantum chemistry calculations become non-straightforward, and sometimes irrelevant.

Indeed, the dynamic spectra identification is a recurrent challenge and various methods have been proposed for vibrational mode assignments[13,27–48]. Wheeler et al. have suggested principal mode analysis (PMA)[11,27] in which the average mass-weighted normal modes and associated frequencies are calculated from a classical MD trajectory by evaluating the eigenvectors and eigenvalues of the covariance matrix associated with the mass-weighted displacements of the molecular coordinates from their equilibrium positions[28–30]. The simplicity of this approach makes it attractive, although several studies have pointed out some limitations, e.g. when the potential energy surface of the molecule is anharmonic or when the thermal equilibrium is not guaranteed[49–51].

Physical Chemistry Chemical Physics Accepted Manuscript

The instantaneous normal mode analysis (INMA) and its variants[31–36] are obtained by diagonalizing the Hessian matrix at each instantaneous configuration of the molecule. The INMs provide an instantaneous decoupled second-order description of the vibrational motions of the molecule at the corresponding time-dependent configuration. The INMs therefore evolve with time, which complicates their identification[37]. Furthermore, INMA can be very time-consuming on large systems.

The vibrational spectrum can also be obtained by the Kohanoffs procedure[38] that allows to extract the eigenfrequencies and eigenmodes from the analysis of MD trajectories without having to calculate the second derivatives matrix[39]. A first global estimation of the frequencies is followed by the determination of the eigenvectors through a least squares fit of the trajectory including orthogonality constraints. Then, the trajectory is projected onto each of the normal modes. At that point, each projected trajectory contains mainly one frequency component, which is re-estimated. Starting with this framework, several methods were developed these last years[40]. The standard procedure to estimate the vibrational frequencies is based on the Fourier transformed velocity autocorrelation function.

Vuillemier et al. proposed a method for obtaining normal modes based on a localization criterion for the Fourier transformed velocity time-correlation functions of the effective modes[13]. There exists a choice of the localization function for which the method becomes equivalent to PMA based on covariance matrix diagonalization. A proper choice of the localization function leads to a method with a strong analogy with the usual normal mode analysis of equilibrium structures, where the thermal averaged Hessian replaces the Hessian. The Hessian itself is never calculated.

Bowman et al. and Kaledin et al. developed a method named Driven Molecular Dynamics (DMD) to perform a full normal-mode analysis, that does not require a calculation of the Hessian[41–43]. They showed that normal modes can be obtained using a straightforward MD approach. The method employs an external harmonic driving term that can be used to scan the spectrum in a continuous wave fashion to determine resonant absorptions, that for weak signals are the normal-mode frequencies. The molecular motions, induced by driving at resonant frequencies, are (by definition) the vibrational modes[41,42,44].

Finally, digital filtering methods may be used to selectively enhance or suppress the vibrational motion in a MD simulation solely on the basis of frequency[45–47].

Among these approaches, we have chosen the driven molecular dynamics method which provides a direct way to understand molecular vibrational modes at a relatively low computational cost and is able to study nonharmonic motion and mode coupling. One important advantage of DMD is the capability to describe easily low frequency vibrational modes in the far IR region[52]. They define the motions of large scale conformational changes which generally occur along the torsional degrees of freedom. Because of the long-range and collective nature of these vibrational modes, they are highly sensitive to molecular structure and its chemical environment, and thus provide a fingerprint of the structure[53]. Moreover, these modes can be used to determine the mechanical anharmonicity in such large amplitude motions. In this context, the frequency driven approach is able to monitor the anharmonicity degree as a function of the temperature[52,54,55]. In the following sections, we briefly describe the DMD algorithm and its implementation in the current version of the TINKER program[56]. Then we discuss its coupling with a previous implementation of finite temperature IR spectra by molecular dynamics simulations. The computed frequencies are used directly as input for DMD to assign vibrational modes. Specific tools were developed for this purpose. They are described here and applied to N-methyl-acetamide, a peptidic fragment model. Finally, an application on the Ac-Phe-Ala-NH$_2$ dipeptide is detailed using the full chain of methods necessary to carry out in order to obtain band assignments of finite temperature IR spectra.

2 Computational Strategy

2.1 Quantum Chemistry Calculations

Geometry optimizations were performed using density functional theory (DFT) with different functionals, i.e. the pure GGA dispersion-corrected functional B97-D[57], the hybrid functional B3LYP[58] and the global-hybrid meta exchange-GGA functional M06[59], associated with the 6-311+G(d,p) basis set. IR spectra computed at the DFT/M06 level were previously found to be consistent with experimental results on peptides[60,61]. Some energetics were computed also with Dunning's correlation consistent basis set cc-pVTZ. All the calculations were carried out with the GAUSSIAN program package[62].

2.2 The AMOEBA Force Field and Its Parameters

The AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) polarizable force field has been used successfully to reproduce structures and relative energies of peptides[25,63,64]. One of the main differences with first generation potentials is its more complex electrostatic model, including permanent charge, dipole and quadrupole moments on each atom which are derived from quantum mechanical calculations[25]. Another valuable feature is the inclusion of many-body polarization effects which are explicitly treated using a self-consistent atomic dipole polarization procedure[65]. Repulsion-dispersion interactions between pairs of nonbonded

atoms are represented by a buffered 14-7 potential[66]. The intramolecular valence terms consist in bond stretchings, angle bendings, and torsions with the MM3 force field energy form[67]. Additional terms are used such as out-of-plane bending and valence crossing terms to model the coupling between bonds and angles. Except for multipoles, the AMOEBABIO-09 set of parameters available in TINKER 6 was used for the current study[56].

It was shown previously that structure changes, and even more IR spectra, are very sensitive to weak interactions, such as intra or intermolecular hydrogen bonds, in MD simulations[26]. In AMOEBA, thanks to the multipolar expansion of the electrostatic energy, one can expect a good representation of hydrogen bonds. Ponder et al. studied the importance of multipoles moments and intramolecular polarization on conformational energies of polar molecules[63]. They conclude cautiously on the use of small model systems to derive intermolecular parameters for larger systems. From the distributed multipole analysis (DMA) method developed by Stone[68,69], a set of multipoles was derived for the whole dipeptide studied in section 3 and assigned to each atomic site. The effect of the method, the basis set and the reference structure to extract multipoles on the computed infrared spectrum of the dipeptide studied in section 3 are presented in Supplementary Information.

2.3 Dipole Autocorrelation Function Fourier Transform Approach (DACF)

Infrared spectra are most often computed within the double harmonic approximation. Accordingly, vibrational frequencies at energy minima are directly derived from second derivatives of the potential energy with respect to molecular coordinates. *A posteriori*, some scaling factors can be applied to the frequencies in order take into account partly anharmonicity and finite temperature. A better way to include these effects is to perform dynamics simulations, in classical, Born-Oppenheimer or Car-Parrinello framework, and then calculate the Fourier transform of the Dipole Auto Correlation Function (DACF) to obtain IR vibrational spectra. We have proposed previously such calculations of the power spectrum using the AMOBEA force field and the TINKER program[26]. In this study, the DACF procedure was usually performed as follows. MD trajectories were propagated using the Velocity-Verlet algorithm, with a time step of 0.1 fs. All trajectories began with an equilibration phase, in the canonical ensemble, to reach the equilibrium temperature. The temperature was maintained by the Nosé-Hoover thermostat which was found previously to be the most appropriate[26]. The length of the equilibration phase depends on the temperature to attain, usually between 10 and 50 ps. At the end, atomic coordinates were stored and then 3 simulations, of 200 ps each, were performed to

record the dipole moment along the trajectory. These trajectories started with the same initial geometry but atomic velocities and kinetic energies were distributed randomly. Infrared spectra were calculated by the DACF procedure and then averaged over the three simulations.

2.4 Vibrational Mode Analysis based on Driven Molecular Dynamics

2.4.1 Method The AMOEBA-DACF approach is able to determine accurately finite temperature vibrational spectra including anharmonic effects[26]. However, the assignment of bands in MD simulation of absorption spectra remains a tricky task, requiring sophisticated vibrational mode analysis. Some methods are available from the literature such as principal mode analysis (PMA), normal mode analysis (NMA) or MD trajectory filtering methods. Bowman et al. proposed a novel approach to perform a mode analysis that does not require a calculation of the Hessian[41,43].

The method implements an external, harmonic driving term that can be used to scan a frequency range and to determine resonant absorptions. This approach, called Driven Molecular-Dynamics (DMD), provides a viable alternative to the standard Hessian-based normal mode analysis. It was introduced in a previous version of the TINKER program by Kaledin et al.[42,43]. In this work, starting from their kernel, we performed developments for an implementation of DMD into the TINKER 6 and 7 software and for vibrational mode analysis.

The Hamiltonian for driven molecular dynamics of a molecular system of N atoms is given by:

$$H(\mathbf{p},\mathbf{q},t) = \sum_i \frac{p_i^2}{2m_i} + V(\mathbf{q}) + U(t)$$

where $\mathbf{q}$ and $\mathbf{p}$ represent the $3N$ atomic coordinates and conjugate momenta, respectively, and V is the molecular potential. The driving force depends only on internuclear distances. It is included at each step of the dynamics via the driving term $U(t)$ as follows:

$$U(t) = \sum_{i,j} \lambda_{ij} r_{ij} sin(\omega_n t)$$

In this equation, λ_{ij} are the coupling constants (the driving parameters), r_{ij} the internuclear distances, and ω_n the normal-mode pulsation obtained from a DACF calculation. When this force is felt by a vibrator at a pulsation very close to ω_n, the system should respond to the driving force by executing motion that is a normal mode (resonance) and with an absorption of energy. At non-resonant frequencies, the absorbed energy is small and oscillatory with time. On resonance, it increases quickly. At resonant frequencies the molecular motions induced by a weak driven force closely correspond to

the normal-modes, while harder driving induces nonharmonic motion and leads to mode-coupling and nonharmonic shifts which are missing in standard normal-mode analysis[52]. In a previous study, the issue of the choice of the driving parameter λ determining the amplitudes of the atomic motion is discussed[42]. When λ is chosen small, the molecular motions are of small amplitude, in the harmonic limit. For large λ values, DMD allows the study of coupled anharmonic motion. For peptides in gas phase, the typical values for lambda were found to be between 1.10^{-5} and 2.10^{-5} Hartree/Bohr.

2.4.2 Limitations and improvements DMD is a method to perform the calculation of the IR spectra with a full normal-mode analysis that does not require a calculation of the Hessian. The calculation of all eigenvalues and corresponding eigenvectors of the Hessian would scale approximately as $O(N^3)$ which is infeasible for large number (N) of atoms. The DMD approach requires $O(N)$ gradients of the potential per time step. A trajectory of N_{steps} scales as $O(N*N_{steps})$ and the scan of N_{scan} frequencies scales as $O(N*N_{steps}*N_{scan})$. As stressed by Bowman et al.[41], the scaling of DMD becomes competitive compared with standard Hessian-based method only for systems including more than 1000 atoms. In the context of the present study, the frequencies are already known from DACF MD simulations. DMD is used only for the assignments of vibrational modes and scales as $O(N*N_{steps})$ for each mode of interest. Moreover, the intensities derived from absorbed energies are not really comparable to experimental intensities. Conversely, the power spectrum obtained from DACF simulations can be compared to the experimental absorption spectrum since it is derived from the dipole moment and allows a direct access to fundamental and combination bands. Furthermore, the method was proven to be accurate for frequency positions when associated to a polarizable force field such as AMOEBA[26]. Nevertheless, DMD has unique capability: it can selectively excite a mode on a frequency criterium.

It is within this framework that we have developed DMD based tools for vibrational mode analysis based on the DACF frequencies. As previously mentioned, the two spectra are not equivalent for the intensities[42]. Accordingly, a strong infrared intensity does not ensure the capability of the internal coordinate to absorb energy[55]. At a resonant frequency, a DMD trajectory shows oscillations of atoms around their equilibrium positions with an amplitude proportional to the absorbed energy[55]. For example, Fig. 1 displays displacements of C=O and N-H internal coordinates of N-methyl-acetamide for two resonant driving frequencies corresponding to each stretching mode. In the bottom part, the N-H stretching mode is excited at its resonant frequency of 3518 cm^{-1} with light DMD conditions (i.e. 5000 steps, 0.1 fs by step, $\lambda = 5.10^{-7}$ Hartree/Bohr). It can be seen that the system is monotonically absorbing en-

ergy with increasing simulation time with negligible absorption in the C=O. In the top part, at the C=O stretching mode resonant frequency of 1723 cm^{-1}, it is much less obvious to define the resonant coordinate on a single criterion of oscillation amplitude. In order to obtain clearly visible C=O oscillations, the conditions of the DMD simulation have to be more specific, i.e. 30000 steps, 0.2 fs by step, $\lambda = 2.10^{-5}$ Hartree/Bohr in this case. The latter example also emphasizes that a filtering of motions, based on frequency and not on amplitude, is required to observe a specific normal mode.
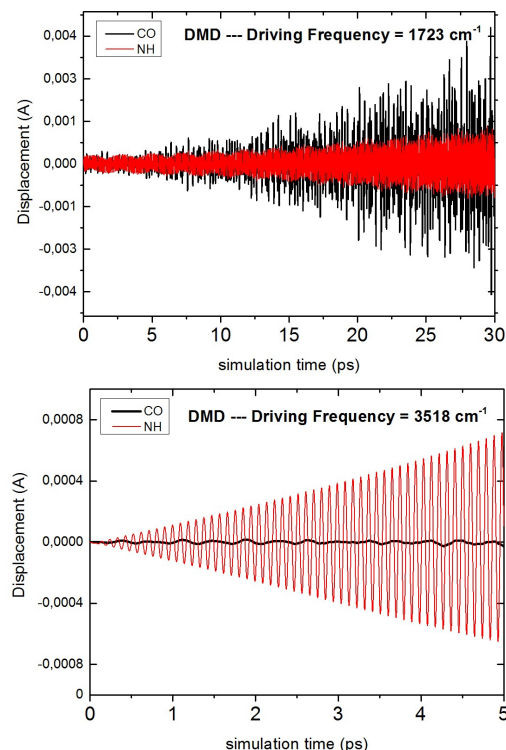

Fig. 1 Monitoring the C=O and N-H N-methyl-acetamide stretch modes by DMD simulations. See text for DMD conditions.

In the case of the model system $H_5O_2^+$, Bowman et al. showed that for the lowest frequency mode (178 cm^{-1}), the coordinate displacement after a DMD simulation can be used to define the associated normal mode with an excellent agreement with the Hessian approach[41]. For large-size and flexible systems, such as proteins or hydrated ions, an issue can be the "resolving power" of the method, i.e., the ability to distinguish modes that are very close in frequency. It may be necessary to drive for longer times to separate modes with nearly the same frequency. Driving for longer times must be done with care by choosing smaller coupling constants because, as the energy absorbed increases, anharmonic motions will be observed[41].

Increasing the temperature implies frequency shifts due to anharmonicity effects. It is difficult to correlate anharmonicity

arising from temperature (DACF) or absorbed energy (DMD), because the mechanisms of energy absorption are very different and the normal mode response is very dependent on the potential energy surface of the system. The position of the N-H stretch maximum intensity in the N-methyl-acetamide molecule can be collected from both simulations with the idea to calibrate the amplitude of the driving force to reproduce the observed shifts with temperature (see Fig. S2 in Supplementary Information). For example, at 200 K, the maximum amplitude is observed close to 3495 cm^{-1}. However, DMD simulation conditions depend strongly on the interrogated modes, and of the population of the modes at very close frequencies. Consequently, competitions can occur presenting a challenge to the DMD approach. We will detail later our procedure for normal modes analysis in a dipeptide case, depending on the frequency range and with the aim to study coupled anharmonic motions.

2.4.3 Tools implementation. The main idea is to catch the atomic motions characteristic of a vibrational mode during the DMD simulation at ω_n, and to display vibrations both as displacement vectors and in animated form in a similar way as conventional Hessian-based analysis. The characteristic motions must be collected when the energy absorbed by the mode under study is near its maximum. The mode will be described by a difference of nuclear positions between two steps during the dynamics, carefully chosen. In-between these two steps, at non-resonant frequencies, the absorbed energy is small and oscillatory with time and can be viewed as "noise" fluctuations in the visualization process. To get a clear picture of the normal mode it is desirable to reduce the noise.

In order to provide band assignments of a MD spectrum, the following procedure was implemented:

1 - Monitor oscillation amplitudes of internal coordinates during the DMD, and calculate their oscillation frequencies

The structure was first energy minimized with TINKER software until the RMS gradient was less than 10^{-5} kcal.mol^{-1}. Starting from the system initially at rest at this stationary point, 5 000 steps of DMD simulation are enough (e.g., a total dynamics of 0.5 or 1 ps, depending on the time step) without an equilibration phase. From the molecular motions corresponding to each resonant driving frequency (e.g. Fig. 2), the frequency of the corresponding mode can be calculated. For a better numerical accuracy, this is carried out over the 20 last periods.

2 - The internal coordinate which vibrates with the pulsation closest to ω_n (driving pulsation) is identified as the reference resonant coordinate and is by definition the vibrational mode

We focus on frequency rather than magnitude because some modes absorb more energy, even off-resonance. Therefore, a mode with a strong IR absorption is not necessarily the one with maximal absorbed energy in a DMD process. This is particularly important when the frequency of the normal mode is not known accurately (e.g. for a high temperature spectrum).

3- Deduction of the time steps that will be used for the description of the motion, i.e. a starting point and an ending point

These steps are defined through the internal coordinate selected in the previous step. We start by choosing the end step as the one for which the oscillation amplitude of this coordinate is maximal (see Fig. 2). The starting step is
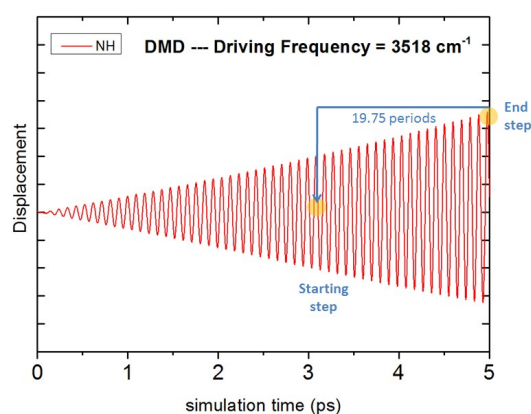

Fig. 2 Choice of end and starting steps for the description of normal modes from DMD simulations. Here, the N-H stretch of N-methyl-acetamide. The driving frequency at 3518 cm^{-1} is obtained from DACF spectrum at 1 K.

associated to an internal coordinate close to its equilibrium value, i.e. with a displacement close to zero. This is observed every $\frac{1}{4}$ period. As 20 periods is the reference time for the frequency calculation, the starting step is deduced from $20 - \frac{1}{4}$ period. The associated time is the equivalent to a reversal of $\frac{n}{4}$ period with a value set to $n = 79$.

4 - Removing the non-resonant motions

This "filtering" process is also based on a frequency criterion on internal coordinates. For all resonant internal coordinates, the motion (between starting and ending steps) for the associated cartesian coordinates is kept. For all other coordinates, displacement is set to zero. The selectivity of the filter is a powerful tool to isolate a specific motion or to visualize collective motions in a frequency range.

In Fig. 3, an example of an amide II mode analysis is provided from DMD-based tools (in red) on the N-methyl-acetamide molecule.
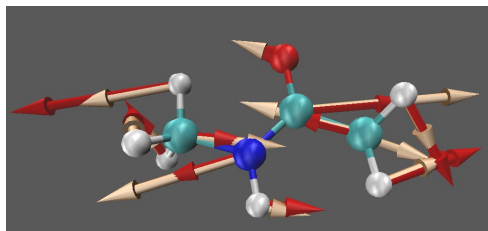
Fig. 3 Amide II mode of N-methyl-acetamide (1551 cm^{-1}): in white, normal mode from static calculation and in red, from DMD simulation (10000 steps, 0.1 fs by step, $\lambda = 10^{-6}$ Hartree/Bohr, filter = 5 cm^{-1}).

The results from Hessian diagonalization is also shown in white. All the components are reproduced. Intensity differences are mainly explained by different energy absorption mechanisms, anharmonicity in DMD calculation, and inevitable small geometric differences between optimized and starting step structures.

These current DMD-based analysis tools emphasize the capability to collect normal modes dynamically. The coupling with the DACF procedure allows the assignment of finite temperature spectra, inside or outside harmonic condition framework. The filter criteria in frequency and amplitude can be associated to different display formats allowing alternative views of the movement associated to a given frequency. The next section will be devoted to an application of these DMD assignment tools to the finite temperature IR spectrum of the Ac-Phe-Ala-NH$_2$ dipeptide.

3 Results and Discussion

The protected peptides series of Ac-Xxx-Phe-NH$_2$ and Ac-Phe-Xxx-NH$_2$ (with Xxx = Gly, Ala, Val and Pro) have been studied by Chin et al. using IR/UV double resonance spectroscopy and DFT quantum chemistry calculations[70,71]. More recently, Jaeqx et al. have investigated the structures and infrared spectra of Ac-Phe-Ala-NH$_2$ and Ac-Phe-Gly-NH$_2$ by IR-UV ion-dip spectroscopy (IR-IDS) and Born-Oppenheimer molecular dynamics simulations[72]. These model peptides are flexible enough to form various non-covalent interactions leading to different chemical environments of the CO and NH groups, and thus to specific spectroscopic signatures. Based on this available set of results, we have also considered Ac-Phe-Ala-NH$_2$ as a good candidate to assess DMD-based assignments of DACF spectra.

3.1 Selection of conformations

Replica Exchange Molecular Dynamics simulations were performed using AMOEBA on Ac-Phe-Ala-NH$_2$ (see Supple-

mentary Information). Among a collection of generated conformations, ten low-energy structures have been selected for quantum chemistry calculations and are presented in Fig. 4. The high flexibility of the peptide allows various intramolecular non-covalent interactions, and the main ones are provided in Fig. 5. These interactions often correspond to hydrogen bonds between an amine or amide group and a neighboring oxygen in a C_5 or C_7 motif. Furthermore, C_8 and C_{10} interactions are possible between more distant groups. Another stabilizing effect is the N-H $\ldots$ π interaction that can exist between an amide hydrogen and the aromatic ring.

The relative energies of the ten conformations were computed at several levels of calculation. Details about the DFT results are presented in Table S2 in Supplementary Information. The energetic order can be impacted by the functional used, however the differences are rather small and all methods yield FAa as the most stable structure, similarly than in previous studies[70,72]. This conformation is stabilized by one C_5, one C_7B and one H_{ala}-π interaction.

The main objective of this paper is the normal mode analysis of finite temperature spectra. This implies that the followed structures remain stable during the dynamics at a given temperature. At 50 K and even more at 200 K, barriers between conformations can be crossed and the barrier heights have been estimated around 5 kcal.mol^{-1}. Among the ten starting structures, five structures, i.e. FAa, FAd, FAe, FAf and FAh, keep their original conformation (see Supplementary Information). Consequently, the following assignments of spectra will be discussed only for these structures.

3.2 Infrared spectra and normal mode analysis

3.2.1 N-H stretching modes. Assignments have been proposed previously by comparing the experimental spectrum to harmonic frequency calculations on the FAa structure at the B3LYP/6-31+G(d) level of theory (with a scaling factor of 0.96)[70]. The 50 K spectra and the frequency maxima for the FAa, FAd, FAe, FAf and FAh conformations are compared to experimental values in Fig. 6. These five structures remain stable during a 200 ps simulation at 50 K. There are 4 N-H stretching contributions in the peptide, i.e. peptidic N-H of alanine and phenylalanine and 2 N-H from the NH$_2$ group. The latter are symmetric and antisymmetric combinations if the environment of both N-H is similar but they are more local modes otherwise.

Mode assignments can also be based on DMD simulations. The N-H$_{ala}$ frequency is only well reproduced in FAa and FAh (exp : 3445 cm^{-1}) because it is involved in a weak interaction, H$_{ala}$-π in FAa and C$_5$B in FAh. In other structures, where it is free, the computed frequency is overestimated. FAa, FAd, FAf and FAh structures reproduce well the N-H$_{phe}$ frequency position (exp : 3422 cm^{-1}). The environment is similar for
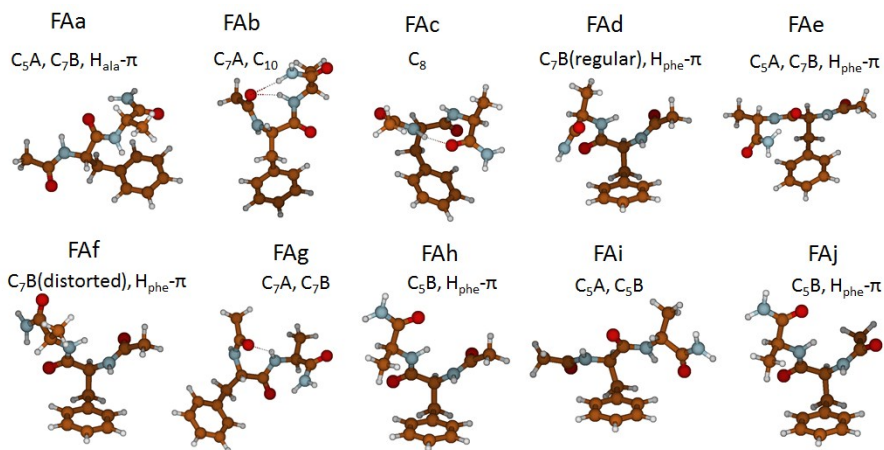
Fig. 4 Low-energy structures of FA at the B3LYP/6-311+G(d,p) level associated to the main non-covalent interactions for each structure.
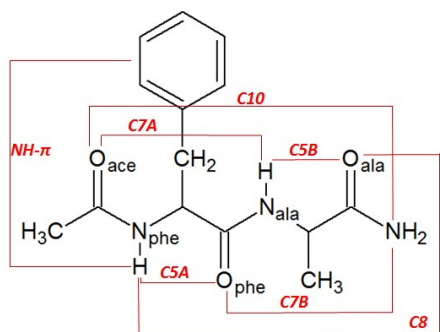

Fig. 5 Nomenclature of non-covalent interactions in FA structures.

FAd, FAf and FAh, in a H_{phe}-π weak interaction. In FAa, it is also involved in a rather weak interaction (see below). However in FAe, the N-H is involved in two interactions (H_{phe}-π and C_5A), and the frequency is red-shifted in comparison with experiments. NH_2 stretchings are well described in FAa, FAe and FAf structures because it is involved in a C_7B interaction in each case. In FAd, this interaction exists in the starting structure but it is not preserved during the simulation and the amine group prefers to make a π interaction.

To conclude, FAa is the better structure to reproduce the N-H stretching modes in agreement with previous studies. The average error on the frequencies is ca. 5 cm^{-1} in comparison with experiments[70]. However, five bands, instead of four, can be distinguished in the FAa spectrum due to change in the N-H environment during the dynamics. In FAa, the C=O$_{Phe}$ is shared between two hydrogen bonding interactions with the N-H$_{Phe}$ and one H_{NH_2} (Fig. 7). This competition weakens each hydrogen bond. In FAf, this competition does not exist because the C=O$_{Ala}$ interacts only with one H_{NH_2}. Therefore,

hydrogen bonding should be stronger for FAf. Indeed, the average length of the O...H hydrogen bonds was found at ca. 2.07 Å for FAa and 2.00 Å for FAf, during a 5 ps simulation at 50 K. In FAa, the band at 3392 cm^{-1} is a representative case of a major competition, and the band at 3375 cm^{-1} corresponds to the loss of the competitive interaction, that strengthens back the remaining one.

If static calculations are usually suitable to assign localized frequency modes, DMD allows to follow modes associated to frequency shifts with increasing temperature and to take into account coupling between modes. For example in Fig. 8, the frequency of the driving force is obtained from DACF spectra for two conformations, 3527 cm^{-1} for FAf and 3549 cm^{-1} for FAh, that correspond to the frequencies of the antisymmetric NH_2 stretch mode. A value of $\lambda = 10^{-6}$ Hartree/Bohr was chosen to stay in the regime of small oscillations. The two N-H have a similar environment in FAh whereas in FAf, the C_7B interaction decouples the two modes. In following variations of the N-H distances during a short simulation, DMD is able to highlight the strong coupling in FAh and the more local behaviour of NH_2 stretching in FAf.

3.2.2 Fingerprint range. The IR spectrum of Ac-Phe-Ala-NH_2 was recorded in the 100-1850 cm^{-1} range using the FELIX free electron laser[72]. DACF spectra have been computed at 1 (not shown), 50 and 200 K to compare with experiments (Fig. 9). It is difficult to correlate strictly experimental and calculated temperature. For example in the case of low-temperature experimental spectra, Gaigeot et al. recommended to simulate the spectra with a higher temperature to speed up the dynamics[72]. Carbonniere et al. discussed the limits of ab initio MD at finite temperature to take into account all anharmonicity effects that can be recorded in static calcu-
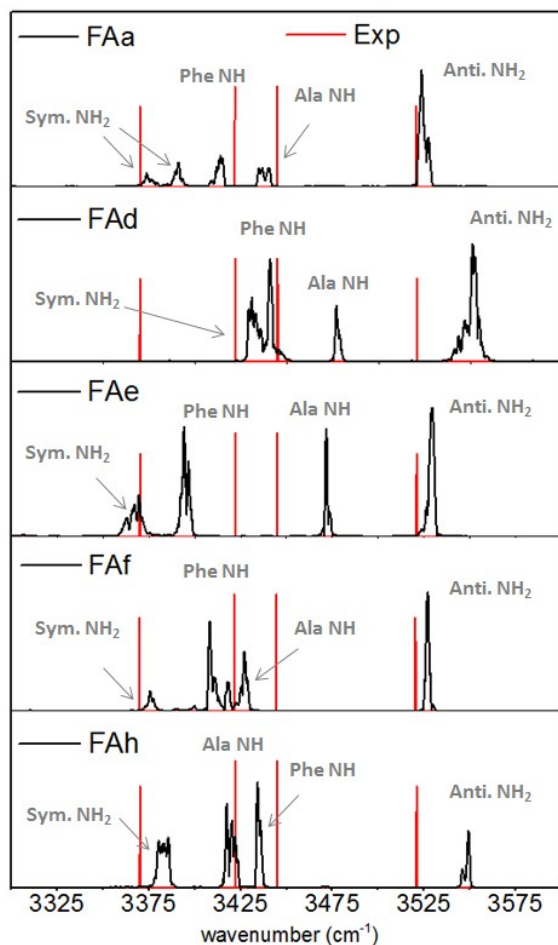
Fig. 6 DACF spectra in the 3300 - 3600 cm^{-1} range for the 5 stable structures during dynamics simulations at 50 K. Experimental frequencies are provided as red bars.

Fig. 7 Hydrogen bonds in FAa and FAf structures.

Fig. 8 N-H mode coupling in antisymmetric NH$_2$ stretchings for FAf and FAh structures. DMD conditions: 10 000 steps of 0.1 fs, λ = 10^{-6} Hartree/Bohr.

lations[14]. In this study, experimental band widths at ca. 15 K are close to those calculated at 200 K, such as the pattern around 1400-1500 cm^{-1}) in Fig. 9. However, frequencies are not really affected by the temperature of the simulation.

Assignments for DACF frequencies in the fingerprint range are proposed in Table 1. They have been obtained in two ways, Hessian diagonalization at the AMOEBA optimized geometry and normal mode analysis from DMD simulations. For DMD attributions, modes are obtained first for a 1 K spectrum to check its capability to visualize modes similar to the ones obtained by a static calculations, and second for a 200 K DACF spectrum to visualize changes in modes induced by anharmonicity and temperature. The parameters used for DMD assignments are detailed in Supplementary Information.

Modes in the C=O stretch region are very close to those obtained at the static DFT/M06 level (scaled by 0.9613), with on one hand, the coupling between NH$_2$ bending and CO$_{ala}$
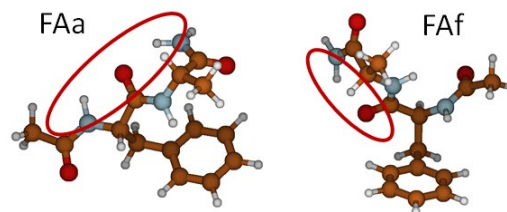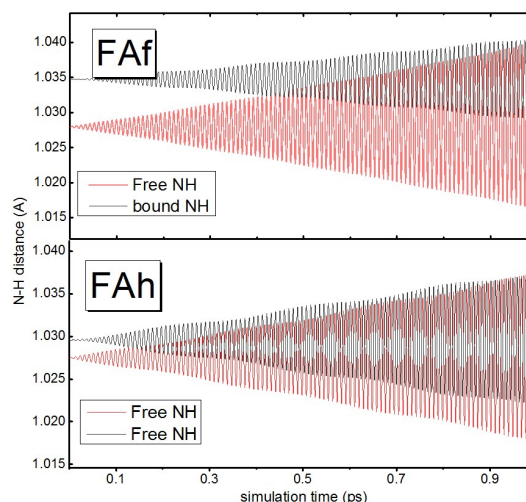
stretch, and on the other hand the coupling between C=O$_{phe}$ and C=O$_{Ace}$ stretchings. In the 1650-1750 cm^{-1} range, three bands are observed experimentally, while four bands have been calculated, CO$_{phe}$ and CO$_{ala}$ stretchings (coupled), and a symmetric and antisymmetric combination of CO$_{ala}$ stretch and NH$_2$ bend modes (see Fig. 10 for symmetric combination). These two last components are still found when the temperature is increased to 200 K, while the coupling between C=O$_{phe}$ and C=O$_{Ace}$ stretch modes decreases with temperature. In 1 K DACF spectrum, both contributions at 1649 and 1662 cm^{-1} have been assigned to the same symmetric contribution of NH$_2$ bending and CO$_{Ala}$ stretching, which reflect the previously mentioned competition between the CO$_{Ala}$...H$_{NH_2}$ and CO$_{Ala}$...H$_{Phe}$ hydrogen bonds.

The experimental pattern around 1400-1550 cm^{-1} is assigned to C-H bends and amide II bands with C-N contributions at low temperature, and an additional contribution of N-H bends at 200 K. Other N-H bendings, at 1522 and 1543 cm^{-1} are in good agreement with experiments. The bands between 1400 and 1500 cm^{-1} are assigned to C-H bends. Fur-
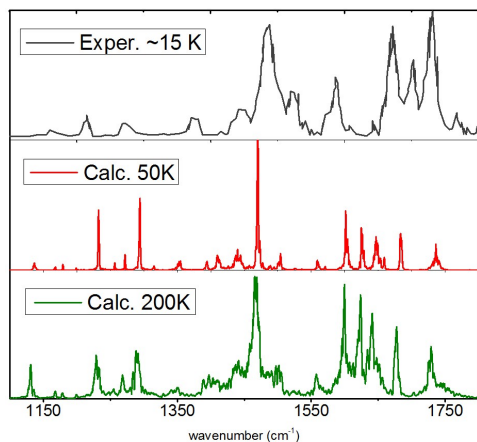
Fig. 9 Experimental[72] and DACF spectra of FAa in the 1100-1800 cm^{-1} range at 50 and 200 K.

Fig. 11 Assignments from DMD calculation for 1599 cm^{-1} normal mode defined as CH_3 umbrella and NH_2 bending (left) and 1465 cm^{-1} (right) normal mode for C-H bend contributions.

Fig. 12 Assignments from static (left, orange arrows) and from DMD calculations (right, red arrows) for 351 cm^{-1} normal mode which corresponds to NCO(Phe) bending.

thermore, one advantage of DMD is to capture anharmonic movements, for example, allowed by increasing the temperature. The intense peak at 1599 cm^{-1} in the 200 K spectrum corresponds to the predominant contributions of CH_3 umbrella modes and NH_2 bending modes (Fig. 11, left part). The other intense band is attributed to C-H with large CH_2 and CH_3 bend contributions (Fig. 11, right part).
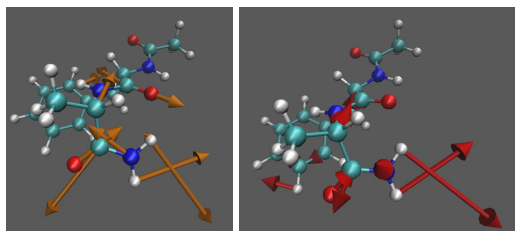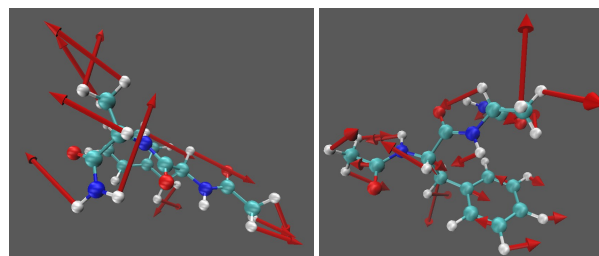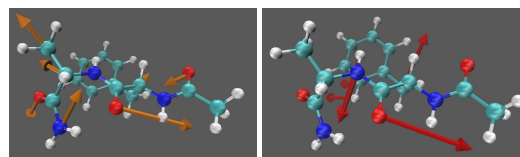
4   Conclusions

We have discussed the implementation of driven molecular dynamics in the latest version of the TINKER program using the AMOEBA polarizable force field. Powerful DMD-based analysis tools allow assigning and visualizing normal modes. In a previous work, we demonstrated that the AMOEBA-DACF approach is able to determine accurately finite temperature vibrational spectra including anharmonic effects. Good agreement with experiments is illustrated again here on the Ac-Phe-Ala-NH_2 dipeptide, through extensive analysis of the conformations (see SI) and the IR spectra (AMOEBA-DACF) with an identification of normal modes (AMOEBA-DACF-DMD). This approach is intrinsically easy to extend to large molecules. It may thus provide a route to calculation and interpretation of IR spectra when the currently used methods, i.e. quantum mechanical based on either Hessian or molecular dynamics, become out of reach.

Fig. 10 Assignments from static (left, orange arrows) and from DMD calculations (right, red arrows) for 1649 cm^{-1} normal mode defined as symmetric combination of NH_2 bending and C=O_{ala} stretching.

Around 1250-1270 cm^{-1}, DMD assignments are C-N stretchings, while DFT attributions correspond to collective N-H bends. NH_2 rocking is calculated by DFT at 1164 cm^{-1}, while in dynamics, it is found at 1073 cm^{-1}. Experimental intensities in this frequency range are very low, and make it difficult to conclude on the position of this band. DMD can also be used to observe soft modes in the far IR range where motions are essentially collective and a large part of the atoms vibrates with energy absorption. For instance, a frequency at 351 cm^{-1} corresponds to the NCO(Phe) bend mode as a major contribution (using an amplitude filter), see Fig. 12.

References

1   E. Garand, M. Z. Kamrath, P. A. Jordan, A. B. Wolk, C. M. Leavitt, A. B. McCoy, S. J. Miller and M. A. Johnson, *Science*, 2012, **335**, 694–698.

Physical Chemistry Chemical Physics Accepted Manuscript

Table 1 DMD Assignments for 1 K and 200 K DACF spectra for the FAa structure in the 1160-1800 cm^{-1} range, in comparison with experimental data[72] and static AMOEBA calculations.

Exp.	Hessian-based		DMD-based			
			DACF 1 K		DACF 200 K	
1160	1180, 1199	def. CH$_2$, bend CH(cycle)	1179	def. cycle		
1215	1233	Str. H$_2$C-C(cycle)	1233	bend CH cycle, str. H$_2$C-C(cycle)	1233	bend CH, str. H$_2$C- C(cycle)
1273	1273, 1295	Amide III	1273, 1295	Amide III	1288	Amide III
1371	1354	bend (CCN, CNC, NH(Ala))	1354	bend CCN(Phe, Ala)	1352	bend (CCN, CNC)
	1413, 1417	umb. CH$_3$(Ac), str. CC, bend CH$_2$	1413	bend (CH$_3$(Ac), CH$_2$, CH cycle)	1398	bend (NH(Phe, Ala), CH)
1442	1444, 1446	umb. CH$_3$(Ac) str. CC, CN(Phe,NH$_2$)	1444	bend CH aliph., str. CN	1437	umb. CH$_3$(Ac), bend(NH(Ala), CH(aliph.))
	1471	bend CH aliph.	1471	bend CH$_3$(Phe), CH cycle	1465	bend CH aliph., NH, str. CC
1488	1489, 1502	bend CH$_3$(Phe)				
1519	1505, 1528	bend CH$_2$, def. cycle	1528	bend CH$_3$	1501	bend CH$_2$
1575	1560	umb. CH$_3$(Phe)	1560	umb. CH$_3$(Phe)	1557	umb. CH$_3$(Phe), bend NH(Ala)
1586	1604	str. CN, bend NH(Phe,Ala)	1604	str. (CC, CN)	1599	umb. CH$_3$, bend (NH$_2$, CH$_2$, NH(Ala), CH aliph.)
	1628	bend NH, str. CO	1628	(bend NH, str. CO	1624	bend NH, str. CO
1642	1649	sym (bend NH$_2$, str. CO(Ala)), str. CC cycle	1649, 1662	sym (bend NH$_2$, str. CO(Ala))	1642	sym (bend NH$_2$, str. CO(Ala)), str. CC cycle
1672 1703	1685 1687	sym str. CO(Ac,Phe) anti str. CO(Ac,Phe)	1687	str. (CO(Ala,Phe),	1678	str. CN(Phe, Ala), CO(Phe)
1732	1737	anti (bend NH$_2$, str. CO(Ala))	1735	anti (bend NH$_2$, str. CO(Ala))	1730	anti (bend NH$_2$, str. CO(Ala))

2   E. Gloaguen and M. Mons, in *Topics in Current Chemistry*, Springer Berlin Heidelberg, 2015, pp. 1–46.

3   Y. M. E. Fung, T. Besson, J. Lemaire, P. Maître and R. A. Zubarev, *Angew. Chem. Int. Ed.*, 2009, **48**, 8340–8342.

4   P. Kupser, K. Pagel, J. Oomens, N. Polfer, B. Koksch, G. Meijer and G. v. Helden, *J. Am. Chem. Soc.*, 2010, **132**, 2085–2093.

5   O. Balaj, D. Semrouni, V. Steinmetz, E. Nicol, C. Clavaguéra and G. Ohanessian, *Chem. Eur. J.*, 2012, **18**, 4583–4592.

6   R. B. Gerber, G. M. Chaban, B. Brauer and Y. Miller, in *Chapter 9 - First-principles calculations of anharmonic vibrational spectroscopy of large molecules*, Elsevier, Amsterdam, 2005, pp. 165–194.

7   V. Barone, *J. Chem. Phys.*, 2005, **122**, 014108.

8   P. Carbonnière, T. Lucca, C. Pouchan, N. Rega and V. Barone, *J. Comput. Chem.*, 2005, **26**, 384–388.

9   D. M. Benoit, *Front. Biosci.*, 2009, **14**, 4229–4241.

10 H.-B. Xie, L. Jin, S. Rudić, J. P. Simons and R. B. Gerber, *J. Phys. Chem. B*, 2012, **116**, 4851–4859.

11 N. Rega, G. Brancato, A. Petrone, P. Caruso and V. Barone, *J. Chem. Phys.*, 2011, **134**, 074504.

12 N. Rega, *Theo. Chem. Acc.*, 2006, **116**, 347–354.

13 M. Martinez, M.-P. Gaigeot, D. Borgis and R. Vuilleumier, *J. Chem. Phys.*, 2006, **125**, 144106.

14 P. Carbonnière, A. Dargelos, I. Ciofini, C. Adamo and C. Pouchan, *Phys. Chem. Chem. Phys.*, 2009, **11**, 4375–4384.

15 Q. Wan, L. Spanu, G. A. Galli and F. Gygi, *J. Chem. Theory Comput.*, 2013, **9**, 4124–4130.

16 M.-P. Gaigeot, R. Vuilleumier, M. Sprik and D. Borgis, *J. Chem. Theory Comput.*, 2005, **1**, 772–789.

17 M.-P. Gaigeot, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3336–3359.

18 F. Ingrosso, G. Monard, M. Hamdi Farag, A. Bastida and M. F. Ruiz-López, *J. Chem. Theory Comput.*, 2011, **7**, 1840–1849.

19 M. Rossi, V. Blum, P. Kupser, G. von Helden, F. Bierau, K. Pagel, G. Meijer and M. Scheffler, *J. Phys. Chem. Lett.*, 2010, **1**, 3465–3470.

20 P. Carbonnière, S. Thicoipe, T. Very and X. Assfeld, *Int. J. Quantum Chem.*, 2012, **112**, 2221–2230.

21 M. Thomas, M. Brehm, R. Fligg, P. Vohringer and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6608–6622.

22 D. C. Marinica, G. Grégoire, C. Desfrançois, J. P. Schermann, D. Borgis and M.-P. Gaigeot, *J. Phys. Chem. A*, 2006, **110**, 8802–8810.

23 P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.

24 A. Grossfield, P. Ren and J. W. Ponder, *J. Am. Chem. Soc.*, 2003, **125**, 15671–15682.

25 J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.

26 D. Semrouni, A. Sharma, J.-P. Dognon, G. Ohanessian and C. Clavaguéra, *J. Chem. Theory Comput.*, 2014, **10**, 3190–3199.

27 R. A. Wheeler, H. Dong and S. E. Boesch, *ChemPhysChem*, 2003, **4**, 382–384.

28 R. A. Wheeler and H. Dong, *ChemPhysChem*, 2003, **4**, 1227–1230.

29 B. R. Brooks, D. Janežič and M. Karplus, *J. Comput. Chem.*, 1995, **16**, 1522–1542.

30 R. M. Levy, O. De la Luz Rojas and R. A. Friesner, *J. Phys. Chem.*, 1984, **88**, 4233–4238.

31 R. M. Stratt, *Acc. Chem. Res.*, 1995, **28**, 201–207.

32 M. Buchner, B. M. Ladanyi and R. M. Stratt, *J. Chem. Phys.*, 1992, **97**, 8522–8535.

33 M. Cho, G. R. Fleming, S. Saito, I. Ohmine and R. M. Stratt, *J. Chem. Phys.*, 1994, **100**, 6672–6683.

34 R. E. Larsen and R. M. Stratt, *J. Chem. Phys.*, 1999, **110**, 1036–1052.

35 M. Nonella, G. Mathias, M. Eichinger and P. Tavan, *J. Phys. Chem. B*,

2003, **107**, 316–322.

36 L. Pejov, D. Spångberg and K. Hermansson, *J. Phys. Chem. A*, 2005, **109**, 5144–5152.

37 A. Bastida, M. A. Soler, J. Zúñiga, A. Requena, A. Kalstein and S. Fernández-Alberti, *J. Phys. Chem. A*, 2010, **114**, 11450–11461.

38 J. Kohanoff, *Comp. Mater. Sci.*, 1994, **2**, 221–232.

39 G. Onida, W. Andreoni, J. Kohanoff and M. Parrinello, *Chem. Phys. Lett.*, 1994, **219**, 1–7.

40 M. C. González Lebrero, L. L. Perissinotti and D. A. Estrin, *J. Phys. Chem. A*, 2005, **109**, 9598–9604.

41 J. M. Bowman, X. Zhang and A. Brown, *J. Chem. Phys.*, 2003, **119**, 646–650.

42 M. Kaledin, A. Brown, A. L. Kaledin and J. M. Bowman, *J. Chem. Phys.*, 2004, **121**, 5646–5653.

43 A. Brown, M. Kaledin, J. Bowman and A. Kaledin, in *Driven Molecular Dynamics for Normal Modes of Biomolecules without the Hessian, and Beyond*, Chapman and Hall/CRC, 2005, pp. 281–300.

44 M. Kaledin, A. L. Kaledin and J. M. Bowman, *J. Phys. Chem. A*, 2006, **110**, 2933–2939.

45 P. Dauber-Osguthorpe and D. J. Osguthorpe, *J. Am. Chem. Soc.*, 1990, **112**, 7921–7935.

46 P. Dauber-Osguthorpe and D. J. Osguthorpe, *Biochemistry*, 1990, **29**, 8223–8228.

47 S. C. Phillips, M. T. Swain, A. P. Wiley, J. W. Essex and C. M. Edge, *J. Phys. Chem. B*, 2003, **107**, 2098–2110.

48 J. Sun, G. Niehues, H. Forbert, D. Decka, G. Schwaab, D. Marx and M. Havenith, *J. Am. Chem. Soc.*, 2014, **136**, 5031–5038.

49 M. Schmitz and P. Tavan, *J. Chem. Phys.*, 2004, **121**, 12233–12246.

50 M. Schmitz and P. Tavan, *J. Chem. Phys.*, 2004, **121**, 12247–12258.

51 F. Calvo, P. Parneix and N.-T. Van-Oanh, *J. Chem. Phys.*, 2010, **133**, 074303.

52 M. Kaledin and D. T. Adedeji, *J. Phys. Chem. A*, 2015, **119**, 1875–1884.

53 H. H. Mantsch and D. Naumann, *J. Mol. Struct.*, 2010, **964**, 1–4.

54 M. Kaledin, A. L. Kaledin, J. M. Bowman, J. Ding and K. D. Jordan, *J. Phys. Chem. A*, 2009, **113**, 7671–7677.

55 M. Kaledin and C. A. Wood, *J. Chem. Theory Comput.*, 2010, **6**, 2525–2535.

56 J. W. Ponder, *TINKER - Software Tools for Molecular Design (version 6)*, http://dasher.wustl.edu/tinker, 2014.

57 S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799,.

58 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

59 Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101–18.

60 K. Joshi, D. Semrouni, G. Ohanessian and C. Clavaguéra, *J. Phys. Chem. B*, 2012, **116**, 483–490.

61 J. K. Martens, I. Compagnon, E. Nicol, T. B. McMahon, C. Clavaguéra and G. Ohanessian, *J. Phys. Chem. Lett.*, 2012, **3**, 3320–3324.

62 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, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.

63 T. D. Rasmussen, P. Ren, J. W. Ponder and F. Jensen, *Int. J. Quantum*

Chem., 2007, **107**, 1390–1395.

64 D. Semrouni, G. Ohanessian and C. Clavaguera, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3450–3462.

65 P. Y. Ren and J. W. Ponder, *J. Comput. Chem.*, 2002, **23**, 1497–1506.

66 T. A. Halgren, *J. Am. Chem. Soc.*, 1992, **114**, 7827–7843.

67 N. L. Allinger, Y. H. Yuh and J. H. Lii, *J. Am. Chem. Soc.*, 1989, **111**, 8551–8566.

68 A. J. Stone, *Chem. Phys. Lett.*, 1981, **83**, 233–239.

69 A. J. Stone and M. Alderton, *Mol. Phys.*, 1985, **56**, 1047–1064.

70 W. Chin, F. Piuzzi, J.-P. Dognon, I. Dimicoli and M. Mons, *J. Chem. Phys.*, 2005, **123**, 084301.

71 W. Chin, J.-P. Dognon, C. Canuel, F. Piuzzi, I. Dimicoli, M. Mons, I. Compagnon, G. von Helden and G. Meijer, *J. Chem. Phys.*, 2005, **122**, 054317.

72 S. Jaeqx, J. Oomens, A. Cimas, M.-P. Gaigeot and A. M. Rijs, *Angew. Chem. Int. Ed.*, 2014, **53**, 3663–3666.