# Boosting the Modeling of InfraRed and Raman Spectra of Bulk Phase Chromophores with Machine Learning

Abir Kebabsa, François Maurel, and Éric Brémond\*

Université Paris Cité, ITODYS, CNRS, F-75013 Paris, France

E-mail: eric.bremond@u-paris.fr

#### Abstract

In the field of vibrational spectroscopy simulation, hybrid approximations to Kohn-Sham density-functional theory (KS-DFT) are often claimed as computationally prohibitive due to the large effort required to evaluate the exchange-correlation potential in planevawe codes. In this Letter, we show that by taking advantage of KS-DFT porting on GPU and machine-learning techniques, simulating IR and Raman spectra of real-life chromophores in bulk aqueous solution becomes a routine application at this level of theory.

With the recurrent developments of novel quantum mechanical tools to assist vibrational spectroscopy,<sup>1–4</sup> InfraRed (IR) and Raman techniques have gained progressively an interest to characterize and identify structures of large molecular systems in complex environments.<sup>5</sup> First restrained to the conventional double harmonic approximation of a potential energy surface (PES), the computation of vibrational spectra evolved in time to more complex anharmonic approaches providing both more reliable wavenumbers and intensities but requiring also more computational resources. From a computational point of view, anharmonic effects

are either captured in a 'static' or 'dynamic' fashion. The static approach is as the name suggest time-independent. It mainly relies on vibrational self-consistent field or perturbation models recognized to be cost-effective but restricted to the treatment semi-rigid molecular systems.<sup>6–9</sup> Its 'dynamic' alternative, based for instance on *ab initio* molecular dynamics (AIMD) simulations, is in principle not limited by the rigidity of the system. It is suitable for the treatment of large amplitude motions and goes beyond by allowing the explicit treatment of solute-solvent interactions.<sup>10–15</sup> Recent advances also showed that it can be adapted to model other forms of vibrational spectroscopies such as resonance Raman<sup>16</sup> or Raman optical activities.<sup>17,18</sup> Its main shortcoming comes from its prohibitive computational cost which is in practise alleviated by introducing severe approximations on the electronic structure method used to compute on-the-fly the PES.<sup>19</sup>

For their large majority, simulations run indeed on massively parallelized planewave Kohn-Sham density-functional theory (KS-DFT)<sup>20,21</sup> codes preferring semilocal density-functional approximations to hybrids for scalability reasons.<sup>22</sup> It is however well-known that the former approximation remains more prone to one-electron self-interaction error (SIE) and its extension to many-electron systems.<sup>23,24</sup> Notably while studying the stretching of covalent bonds,<sup>25</sup> a property in the center of attention in vibrational spectroscopy.

Although largely applied in 'static' computations since the beginning of the 90',  $^{7,26-28}$  the application of hybrid exchange-correlation approximations remain really rare in their 'dynamic' variant. While they are claimed to provide more reliable wavenumbers in strict comparison to experiments, <sup>19</sup> the evaluation of their exact-like exchange term from planewave basis set remains prohibitive in terms of computational cost. Out of the context of vibrational computational spectroscopy, some possible strategies have recently emerged to bypass such limitations. Just to cite some, we list the combination of the adaptively compressed exchange operator formulation and a multiple time step integration scheme, <sup>29,30</sup> the coordinate scaling approach,<sup>31</sup> or the localization of the basis set.<sup>32–34</sup>

More generally, a more natural way to run AIMD with hybrid density functionals still

remains localized Gaussian basis sets.<sup>35–37</sup> They are of course more adapted to treat single molecular instead of bulk phase systems. However, coupling them with molecular mechanics (MM) so as to take into account a surrounding complex environment, and of course the resulting solute-solvent interactions, becomes a robust alternative to compute IR and Raman spectra of molecular compounds in solution.<sup>38,39</sup> This approach is moreover facilitated by the effort of the community to write and port KS-DFT algorithms on graphics processing unit (GPU), a technical upgrade that enhances the scalability and improves the cost/accuracy trade-off the AIMD simulation.<sup>40</sup>

In this context, we report in this Letter, our first advances in computing vibrational spectroscopic properties of real-life molecular systems in solution. More precisely, we show that the derivation of IR and Raman spectra of solvated chromophores from hybrid quantum mechanical and molecular mechanical (QM/MM) AIMD simulations without making big compromise on the methodology, that is applying hybrid density functionals and basis sets containing diffuse functions, is currently affordable. We try to go further on this line by demonstrating that polarizability tensors, key element to simulate Raman spectra, can be directly obtained by solving coupled-perturbed Kohn-Sham (CPKS) equations, instead of using approximate charge decomposition techniques as it is routinely the case. In this specific time-consuming case, we will finally propose some machine learning alternatives to alleviate the computation cost and shows that neural networks are a viable alternative to establish a relationship between structure and dipole moment or polarizability tensors.

When we speak about real-life molecular systems in solution, we refer to compounds routinely used by experimentalists for chemistry-relevant applications. We focus here on 4-aza-8,12-dioxo-triangulenium, commonly dubbed ADOTA<sup>+</sup>, a fused triarylcarbenium carbocation particularly appreciated for its impressive spectroscopic features and its robust stability in solution, including in polar solvents (Figure 1).<sup>41,42</sup> We start by comparing its IR spectroscopic features simulated from AIMD trajectories in gas phase and water solution. In this framework, the frequency-dependent absorption coefficient reads

$$\mathcal{A}(\omega) \propto \int_{-\infty}^{\infty} \langle \dot{\boldsymbol{\mu}}(\tau) \dot{\boldsymbol{\mu}}(t+\tau) \rangle_{\tau} \mathrm{e}^{-i\omega t} \mathrm{d}t$$
(1)

where  $\dot{\boldsymbol{\mu}}(t)$  is the time-derivative of the dipole moment vector processed by autocorrelation. We refer the reader to state-of-the-art reviews and tutorials for deeper details on the approach<sup>13,14,43</sup> and to Section S1 in Supporting Information for the methodology applied here.

The gas phase IR spectrum of ADOTA<sup>+</sup> is derived from a 30 ps AIMD simulation running at 300 K with a 0.2 fs time step (see Section S2 in Supporting Information for more details). In reference to 'static' benchmarks and applications,<sup>7,26–28</sup> we choose to describe the PES by using a combination of the PBE0 hybrid density functional<sup>44,45</sup> casting 25% of exact-like exchange and the 6-31+G<sup>\*</sup> double- $\zeta$  basis set augmented with diffuse functions. The resulting spectrum is depicted in Figure 2 and the corresponding pictures of the vibrational modes obtained by principal component analysis (PCA) of the covariance matrix of the cartesian coordinates along the trajectory are reported in Section S3 in Supporting Information.<sup>46</sup> The spectrum is composed by a first intense band at 3612 cm<sup>-1</sup> assigned by power spectrum analysis to the aromatic N–H bond stretching. It is followed by a second broad and less intense band spanning between 3200 and 3260 cm<sup>-1</sup> which is this time attributed to the stretching of the aromatic C–H bonds. The band which dominates the IR spectrum arrives just after and is split in two maxima assigned to the in-plane C–C stretching perpendicular and along the  $C_2$  axis of the molecule. The last and less energetics bands are finally attributed to the in- and out-of-plane breathing of the heteroaromatic structure.

The IR spectrum of ADOTA<sup>+</sup> in bulk water is obtained as for it from a 30 ps QM/MM AIMD simulation running in the same conditions as mentioned above (see Section S2 in Supporting Information for more details). Within the QM layer, the chromophore is described at the same level of theory as above. It is however solvated by electronic embedding<sup>47</sup> using

an explicit water box environment treated with the TIP3P force field.<sup>48</sup> The resulting IR spectrum appears much more noisy with respect to the gas phase one in consequence of the interaction between the solute and surrounding water molecules (Figure 2). The first intense band at  $3722 \text{ cm}^{-1}$  and its low-energy shoulder at  $3697 \text{ cm}^{-1}$  are assigned by power spectrum analysis to the stretching of the O–H bonds of the surrounding water molecules belonging to the MM layer. The second intense band at  $3440 \text{ cm}^{-1}$  corresponds to the aromatic N-H bond stretching since it involves the vibration of both N and H atoms. The solute-solvent interaction redshifts thus the band by  $-172 \text{ cm}^{-1}$  with respect to gas phase computations. This large redshift, which is a function of the strength of the weak interaction, is attributed to a strong H-bond interaction between both solute and solvents.<sup>49</sup> A third broad and intense band with a maximum at  $3265 \text{ cm}^{-1}$  is then computed. It lies in the stretching region of the aromatic C–H bonds but remains much more intense with respect than in gas phase. This difference in intensity is here again attributed to the interaction with explicit water solvent molecules. However, the absence of apparent redshit of the transition remains synonymous of very weak interaction between the solvent and the aromatic C–H bonds. The emergence of the next massif between 2000 and 2400  $\rm cm^{-1}$  is the signature of the combination bendingrocking vibration of the water solvent.<sup>50,51</sup> The most intense band of the spectrum is not shifted by the solvent. However, the following less energetic bands attributed to the inand out-of-plane breathing of the heteroaromatic structure are broadened by the hindered rotation and translation movements of solvent.

In direct comparison to the IR spectrum, the frequency-dependent Raman scattering intensity is obtained by processing the autocorrelation function of the time-derivative of the polarizability tensor  $\dot{\alpha}$  such as

$$\mathcal{I}_{xx}(\omega) \propto \frac{1}{\underbrace{1 - \exp(-\frac{\hbar\omega}{k_{\rm B}T})}_{g(\omega)}} \frac{(\omega_{\rm in} - \omega)^4}{\omega} \int_{-\infty}^{\infty} \langle \dot{\alpha}_{xx}(\tau) \dot{\alpha}_{xx}(t+\tau) \rangle_{\tau} \mathrm{e}^{-i\omega t} \mathrm{d}t \tag{2}$$

where  $\alpha_{xx}$  is one of the nine components of the 3 × 3 polarizability tensor, and  $g(\omega)$  a frequency-dependent damping function which depends on the temperature T and the energy of the polarized incident laser beam  $\omega_{in}$  taken here as 19455 cm<sup>-1</sup>. Again, the reader is refered to more comprehensive details regarding Raman scattering spectroscopy in state-of-the-art reviews and tutorials.<sup>13,14,43</sup> We report however in Section S1 in Supporting Information more details about the methodology used here, and notably the expression used to estimate the perpendicular and parallel components of the Raman intensities. It is important to note that by difference with most of the investigations dealing with Raman spectroscopy, the polarizability tensor is computed at each MD steps by solving the CPKS equations. This approach remains of course more time-consuming than using more standard Wannier localization<sup>52</sup> or Voronoi integration<sup>53</sup> schemes but conserves the advantage to reduce empiricism in deriving polarizability tensors and avoid artefacts in simulated spectra.

The Raman spectrum of ADOTA<sup>+</sup> is less prone to solvent effects. The main differences originate from the intensity of the bands. As depicted in Figure 2, the perpendicular intensity  $\mathcal{I}_{\perp}$  of the band at 3265 cm<sup>-1</sup>, assigned to the stretching of the aromatic C–H bonds, is softened by solvent effects. Similar conclusions can be drawn for the bands lower in energy than 500 cm<sup>-1</sup> and ruling the in- and out-of-plane breathing of the heteroaromatic structure. In the same way, the depolarization ratio  $\mathcal{I}_{\perp}/\mathcal{I}_{\parallel}$  is slightly affected by solvent effects. It is important to note that by difference with  $\mathcal{I}_{\perp}$ , the depolarization ratio detects the stretching of the O–H bonds of the solvent at 3722 cm<sup>-1</sup> as well as the aromatic N–H bond stretching at 3440 cm<sup>-1</sup> in H-bond interaction with the solvent.

As shown above, the high performance computing infrastructures available allow nowadays to simulate the vibrational spectroscopic properties of real-life molecular systems in solution without introducing big sacrifices to the methodology. However, it is important to depict the computational effort necessary to achieve such a simulation and to propose some alternatives to reduce it. The AIMD simulations ran here in parallel with the release 1.95 of the Terachem software package on four NVIDIA RTX A4000 GPU cards. With the corresponding KS-DFT level of theory, each AIMD step takes approximatively 8 s. This time is mulplied by a factor of 4, *i.e.*  $\sim$ 33 s, when the resolution of the CPKS equations is added on top. Over 150,000 AIMD steps, it roughly corresponds to 14 and 58 days of simulation for the IR and Raman spectra, respectively. At this point, it is thus important to take an interest in alleviating the computational effort, especially in the case of Raman spectrum simulation.

In this respect, some tentatives have recently emerged on the introduction of machine learning (ML) techniques to boost and interpret AIMD simulations. We refer the reader to Ref. 54 for a comprehensive review on the topic in the general context of computational vibrational spectroscopy. More specifically here, we underline the work by Marquetand and coworkers which exploits ML potentials to boost the simulation of IR spectra of small to large organic molecules.<sup>55</sup> We also remark the investigation by Schienbein which demonstrates that ML techniques successfully reproduce atomic polar tensors (APT) charges of liquid water from AIMD trajectories and lead to the accurate simulation of its IR spectrum.<sup>56</sup>

Following this path, we show here that a simple supervised neural network regression model such as the multi-layer perceptron<sup>57</sup> (MLP) can directly connect a set of cartesian coordinates obtained from an AIMD simulation to the corresponding dipole moment or polarizability tensor property. The resulting ML workflow, which takes benefit from the MLP implementation of the Scikit-learn library,<sup>58</sup> is detailed in Figure 3 and the code is hosted on the GitHub platform.<sup>59</sup> From a preliminary AIMD trajectory (I), snapshots are randomly collected to build a training and test set. Each snapshot is characterized by the cartesian coordinates of the system at a given time and the corresponding dipole moment and polarizability tensor computed at DFT level. Within the training set, cartesian coordinates are standardized (II). This important step consists in centering the center of mass of the system at the origin, and then scaling it to the unit variance. The resulting standardized inputs are then connected to the output dipole moment or polarizability tensor properties by training the MLP model  $|I\!\nabla$ , (see Section S4 in Supporting Information for more details about the ML model parameterization)]. The same preprocessing of the input coordinates is finally run with the test set  $(I\!\!X)$  to simulate the machine-learned dipole moments or polarizability tensors  $(V\!\!X)$ .

The snapshots are randomly picked within the last 20 ps of each AIMD trajectory (*i.e.*, among a total of 100,000 snapshots) and split in different proportions to build a training and test set of size n and  $10^5 - n$ , respectively. The accuracy of the resulting MLP model, function of the training set size, is given in Figure 3 through the calculation of the coefficient of determination  $\mathcal{R}^2$ . Regarding the dipole moment property, with only one hidden layer, a minimum number of  $10^4$  training points is required to reach a  $\mathcal{R}^2$  of about 0.82 and 0.78 on the training and test set, respectively. As reported in Figure S5 in Supporting Information, the addition of more hidden layers, which in parallel increases the complexity of the model, a very low accuracy improvement. For instance, for three hidden layers,  $\mathcal{R}^2$  is about 0.91 and 0.88 for the same number of training points. Regarding the polarizability tensor property, the accuracy convergence is much faster. Still for  $10^4$  training points,  $\mathcal{R}^2$  is about 0.97 on both training and test sets. We can thus conclude that the simple MLP model used here in regression fast and better learn polarizability tensors than dipole moments, still keeping of course a very good accuracy for the latter property.

It is now time to apply this robust ML tool for the simulation of IR and Raman vibrational spectra. For that,  $10^4$  snapshots are first randomly picked from a 20 ps QM/MM trajectory to train the MLP model, and then, the latter trained model is applied to another 20 ps QM/MM trajectory, taking as input the cartesian coordinates to simulate a set of  $10^5$  machine-learned dipole moments on the one hand, and polarizability tensors on the other hand. Their time series analysis give the ML IR and Raman spectra depicted in Figure 4. At first look, we observe that the ML IR spectrum reproduces only in part the original IR signature of ADOTA<sup>+</sup> in bulk water, and is especially noisy-less in comparison. Going deeper into details, we remark that the first simulated band displaying two maxima at 3520 and 3460 cm<sup>-1</sup> corresponds to the aromatic N–H bond stretching, while the second spanning between 3200

and  $3260 \text{ cm}^{-1}$  is attributed to the stretching of the aromatic C–H bonds. The most intense band of the spectrum comes just after and lies between 1640 and 1760 cm<sup>-1</sup>. It is again assigned to the in-plane C–C stretching perpendicular and along the  $C_2$  axis of ADOTA<sup>+</sup>, and it is followed by the in- and out-of-plane breathing of the heteroaromatic structure. Here, there is no signature of the IR vibrations of the solvent, which in the QM/MM spectrum, broadened the bands and leads to two broad bands around 3700 and 2200<sup>-1</sup>. The ML model acts thus like a filter which cleans the signal of the solvent but conserves the signature of the interaction between the solute and solvent. The same conclusions can be drawn for the Raman spectrum.

Finally, if we focus on the question of the reduction of the computational effort brought by the ML methodology developed above, we can conclude that it is drastically decreased in the case of the Raman spectrum but remains unchanged for IR. Indeed, since the ML model takes as input the cartesian coordinates of an AIMD trajectory, and that the dipole moment is systematically computed for each dynamics step, there is no apparent cost reduction. However, since the polarizability tensor is derived from an on-top and computationally expensive CPKS procedure at each dynamics step, it means that only  $10^4$  CPKS computations are sufficient over a total of  $10^5$  AIMD steps, that corresponds to a cost reduction of about 32%, *i.e.* 12 days instead of 38.

To sum up, we tackle in this Letter, the simulation of the IR and Raman vibrational spectra of a real-life chromophore in bulk phase aqueous solution, the 4-aza-8,12-dioxo-triangulenium. Taking advantage from the cost-effective porting of KS-DFT codes using Gaussian basis sets on GPU, we show that deriving such a spectroscopic property from QM/MM AIMD simulations is currently affordable without making big compromise on the density-functional approximation and basis set size. More precisely, we emphasize that hybrid functionals that are often claimed as computationally prohibitive but remains a robust barrier against SIE, are routinely applicable via this protocol, also in association with Gaussian basis set containing diffuse functions. We push the limits in showing that Raman

spectra can be obtained from a time series analysis of polarizability tensors derived directly by solving CPKS equations at each AIMD step, and demonstrate finally that a very simple neural network model connecting the cartesian coordinates of an AIMD trajectory to the polarizability tensor reduces the computational effort to the one of a IR spectrum.



Figure 1: Representation of the ADOTA  $^+$  chromophore investigated herein in gas and water bulk phases.



Figure 2: (top left) InfraRed and (bottom left) Raman scattering intensities of ADOTA<sup>+</sup> in gas and water bulk phases. The Raman depolarization ratio  $(\mathcal{I}_{\perp}/\mathcal{I}_{\parallel})$  is provided in dashed lines. Power spectra involving the N–H bond of ADOTA<sup>+</sup> for the (top right) gas and (bottom right) water bulk phases *ab initio* molecular dynamics trajectory. For the latter, the power spectra of the O–H water bond in noncovalent interaction with ADOTA<sup>+</sup> is also provided.



Figure 3: (left) Machine learning workflow for the simulation of dipole moments and polarizability tensors. (right) Cross validation score of the machine learning model function of the size of the training set for (top) dipole moments and (bottom) polarizability tensors. The success of the model is analyzed by calculation of the  $\mathcal{R}^2$  coefficient of determination for ten randomly generated training and test sets of size n and  $10^5 - n$ , respectively. Solid lines and shaded color areas plot the mean and standard deviation interval of  $\mathcal{R}^2$ , respectively.



Figure 4: (left) InfraRed and (right) Raman scattering intensities of ADOTA<sup>+</sup> in water bulk phase derived from machine learning (ML) and QM/MM *ab initio* molecular dynamics trajectories. The right inset refers to a zoom of the Raman spectrum within the 1000–3500 cm<sup>-1</sup> region.

#### Acknowledgement

Dr. Florent BARBAULT, Dr. Philippe LAINÉ and Prof. Antonio MONARI are gratufully acknowledged for their insights regarding the carbenium chemistry and QM/MM methodology. The authors gratefully acknowledge ANR (Agence Nationale de la Recherche) for the financial support of this work through the MoMoPlasm project (ANR-21-CE29-0003). They thank also ANR and CGI (Commissariat à l'Investissement d'Avenir) for their financial support of this research through Labex SEAM (Science and Engineering for Advanced Materials and devices) ANR-10-LABX-096, ANR-18-IDEX-0001. They also acknowledge the GENCI-TGCC center for HPC resources (Projects AD010810359R1 and AD010810359R2).

#### Supporting Information Available

**Data Availability Statement** The source code for implementing the machine-learning model developed here as well as the database used to train and test the model are available in the online GitHub repository: https://github.com/ANRMoMoPlasm/MLPvibSpec.

**Supporint Information** The Supporting Information reports all the details regarding the time-series analysis methodology (Section S1), the computational details (Section S2), the list of the vibrational modes of ADOTA<sup>+</sup> (Section S3), and the description and implementation of the machine learning model (Section S4).

### References

- Carter, S.; Sharma, A. R.; Bowman, J. M.; Rosmus, P.; Tarroni, R. Calculations of rovibrational energies and dipole transition intensities for polyatomic molecules using MULTIMODE. J. Chem. Phys. 2009, 131, 224106.
- (2) Barone, V.; Baiardi, A.; Biczysko, M.; Bloino, J.; Cappelli, C.; Lipparini, F. Imple-

mentation and validation of a multi-purpose virtual spectrometer for large systems in complex environments. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12404–12422.

- (3) Licari, D.; Fusè, M.; Salvadori, A.; Tasinato, N.; Mendolicchio, M.; Mancini, G.; Barone, V. Towards the SMART workflow system for computational spectroscopy. *Phys. Chem. Chem. Phys.* 2018, 20, 26034–26052.
- (4) Guan, J.; Lu, Y.; Sen, K.; Abdul Nasir, J.; Desmoutier, A. W.; Hou, Q.; Zhang, X.; Logsdail, A. J.; Dutta, G.; Beale, A. M.; Strange, R. W.; Yong, C.; Sherwood, P.; Senn, H. M.; Catlow, C. R. A.; Keal, T. W.; Sokol, A. A. Computational infrared and Raman spectra by hybrid QM/MM techniques: a study on molecular and catalytic material systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2023**, *381*, 20220234.
- (5) Barone, V.; Alessandrini, S.; Biczysko, M.; Cheeseman, J. R.; Clary, D. C.; Mc-Coy, A. B.; DiRisio, R. J.; Neese, F.; Neese, F.; Puzzarini, C. Computational molecular spectroscopy. *Nat. Rev. Methods Primers* **2021**, *1*, 38.
- (6) Schneider, W.; Thiel, W. Anharmonic force fields from analytic second derivatives: Method and application to methyl bromide. *Chem. Phys. Lett.* **1989**, *157*, 367–373.
- (7) Barone, V. Characterization of the potential energy surface of the HO2 molecular system by a density functional approach. J. Chem. Phys. 1994, 101, 10666–10676.
- (8) Neugebauer, J.; Hess, B. A. Fundamental vibrational frequencies of small polyatomic molecules from density-functional calculations and vibrational perturbation theory. J. Chem. Phys. 2003, 118, 7215–7225.
- (9) Barone, V.; Biczysko, M.; Bloino, J. Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation. *Phys. Chem. Chem. Phys.* 2014, 16, 1759–1787.

- (10) Gaigeot, M.-P.; Sprik, M. Ab Initio Molecular Dynamics Computation of the Infrared Spectrum of Aqueous Uracil. J. Phys. Chem. B 2003, 107, 10344–10358.
- (11) Marie-Pierre Gaigeot, M. M.; Vuilleumier, R. Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations: application to small peptides. *Mol. Phys.* 2007, 105, 2857–2878.
- (12) Gaigeot, M.-P. Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase. *Phys. Chem. Chem. Phys.* 2010, 12, 3336–3359.
- (13) Thomas, M.; Brehm, M.; Fligg, R.; Vöhringer, P.; Kirchner, B. Computing vibrational spectra from ab initio molecular dynamics. *Phys. Chem. Chem. Phys.* 2013, 15, 6608–6622.
- (14) Ditler, E.; Luber, S. Vibrational spectroscopy by means of first-principles molecular dynamics simulations. WIREs Comput. Mol. Sci. 2022, 12, e1605.
- (15) Bakels, S.; Gaigeot, M.-P.; Rijs, A. M. Gas-Phase Infrared Spectroscopy of Neutral Peptides: Insights from the Far-IR and THz Domain. *Chem. Rev.* 2020, 120, 3233– 3260.
- (16) Brehm, M.; Thomas, M. Computing Bulk Phase Resonance Raman Spectra from ab Initio Molecular Dynamics and Real-Time TDDFT. J. Chem. Theory Comput. 2019, 15, 3901–3905.
- (17) Brehm, M.; Thomas, M. Computing Bulk Phase Raman Optical Activity Spectra from ab initio Molecular Dynamics Simulations. J. Phys. Chem. Lett. 2017, 8, 3409–3414.
- (18) Luber, S. Raman Optical Activity Spectra from Density Functional Perturbation Theory and Density-Functional-Theory-Based Molecular Dynamics. J. Chem. Theory Comput. 2017, 13, 1254–1262.

- (19) Taherivardanjani, S.; Elfgen, R.; Reckien, W.; Suarez, E.; Perlt, E.; Kirchner, B. Benchmarking the Computational Costs and Quality of Vibrational Spectra from Ab Initio Simulations. Adv. Theory Simul. 2022, 5, 2100293.
- (20) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* 1964, 136, B864–B871.
- (21) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. Phys. Rev. 1965, 140, A1133–A1138.
- (22) Chawla, S.; Voth, G. A. Exact exchange in ab initio molecular dynamics: An efficient plane-wave based algorithm. J. Chem. Phys. 1998, 108, 4697–4700.
- (23) Bao, J. L.; Gagliardi, L.; Truhlar, D. G. Self-Interaction Error in Density Functional Theory: An Appraisal. J. Phys. Chem. Lett. 2018, 9, 2353–2358.
- (24) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* 2008, *321*, 792–794.
- (25) Shukla, P. B.; Mishra, P.; Baruah, T.; Zope, R. R.; Jackson, K. A.; Johnson, J. K. How Do Self-Interaction Errors Associated with Stretched Bonds Affect Barrier Height Predictions? J. Phys. Chem. A 2023, 127, 1750–1759.
- (26) Martin, J. M. L.; Van Alsenoy, C. Structure and Vibrational Spectra of the Azabenzenes. A Density Functional Study Including Exact Exchange Contributions. J. Phys. Chem. 1996, 100, 6973–6983.
- (27) Carbonniere, P.; Lucca, T.; Pouchan, C.; Rega, N.; Barone, V. Vibrational computations beyond the harmonic approximation: Performances of the B3LYP density functional for semirigid molecules. J. Comput. Chem. 2005, 26, 384–388.
- (28) Puzzarini, C.; Biczysko, M.; Barone, V. Accurate Harmonic/Anharmonic Vibrational Frequencies for Open-Shell Systems: Performances of the B3LYP/N07D Model for

Semirigid Free Radicals Benchmarked by CCSD(T) Computations. J. Chem. Theory Comput. 2010, 6, 828–838.

- (29) Mandal, S.; Nair, N. N. Speeding-up ab initio molecular dynamics with hybrid functionals using adaptively compressed exchange operator based multiple timestepping. J. Chem. Phys. 2019, 151, 151102.
- (30) Mandal, S.; Thakkur, V.; Nair, N. N. Achieving an Order of Magnitude Speedup in Hybrid-Functional- and Plane-Wave-Based Ab Initio Molecular Dynamics: Applications to Proton-Transfer Reactions in Enzymes and in Solution. J. Chem. Theory Comput. 2021, 17, 2244–2255.
- (31) Bircher, M. P.; Rothlisberger, U. Exploiting Coordinate Scaling Relations To Accelerate Exact Exchange Calculations. J. Phys. Chem. Lett. 2018, 9, 3886–3890.
- (32) Luber, S. Local electric dipole moments for periodic systems via density functional theory embedding. J. Chem. Phys. 2014, 141, 234110.
- (33) Ko, H.-Y.; Jia, J.; Santra, B.; Wu, X.; Car, R.; DiStasio Jr., R. A. Enabling Large-Scale Condensed-Phase Hybrid Density Functional Theory Based Ab Initio Molecular Dynamics. 1. Theory, Algorithm, and Performance. J. Chem. Theory Comput. 2020, 16, 3757–3785.
- (34) Ko, H.-Y.; Santra, B.; DiStasio, R. A. J. Enabling Large-Scale Condensed-Phase Hybrid Density Functional Theory-Based Ab Initio Molecular Dynamics II: Extensions to the Isobaric–Isoenthalpic and Isobaric–Isothermal Ensembles. J. Chem. Theory Comput. 2021, 17, 7789–7813.
- (35) Carbonniere, P.; Dargelos, A.; Ciofini, I.; Adamo, C.; Pouchan, C. Vibrational analysis of glycine radical: a comparative ab initio static and dynamic study. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4375–4384.

- (36) Brémond, E.; Savarese, M.; Rega, N.; Ciofini, I.; Adamo, C. Free Energy Profiles of Proton Transfer Reactions: Density Functional Benchmark from Biased Ab Initio Dynamics. J. Chem. Theory Comput. 2022, 18, 1501–1511.
- (37) Briccolani-Bandini, L.; Brémond, E.; Pagliai, M.; Cardini, G.; Ciofini, I.; Adamo, C. Concerted versus stepwise proton transfer reactions in the [2, 2'-bipyridyl]-3-3'-diol molecule: A static and dynamic ab-initio investigation. J. Comput. Chem. 2023, 44, 2308–2318.
- (38) Nonella, M.; Mathias, G.; Tavan, P. Infrared Spectrum of p-Benzoquinone in Water Obtained from a QM/MM Hybrid Molecular Dynamics Simulation. J. Phys. Chem. A 2003, 107, 8638–8647.
- (39) Tanzi, L.; Ramondo, F.; Guidoni, L. Vibrational Spectra of Water Solutions of Azoles from QM/MM Calculations: Effects of Solvation. J. Phys. Chem. A 2012, 116, 10160– 10171.
- (40) Seritan, S.; Bannwarth, C.; Fales, B. S.; Hohenstein, E. G.; Isborn, C. M.; Kokkila-Schumacher, S. I. L.; Li, X.; Liu, F.; Luehr, N.; Snyder Jr., J. W.; Song, C.; Titov, A. V.; Ufimtsev, I. S.; Wang, L.-P.; Martínez, T. J. TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics. WIREs Comput. Mol. Sci. 2021, 11, e1494.
- (41) Bosson, J.; Gouin, J.; Lacour, J. Cationic triangulenes and helicenes: synthesis, chemical stability, optical properties and extended applications of these unusual dyes. *Chem. Soc. Rev.* 2014, 43, 2824–2840.
- (42) Barsuk, I.; Lainé, P. P.; Maurel, F.; Brémond, E. Triangulenium dyes: the comprehensive photo-absorption and emission story of a versatile family of chromophores. *Phys. Chem. Chem. Phys.* **2020**, *22*, 20673–20684.

- (43) Brehm, M.; Thomas, M.; Gehrke, S.; Kirchner, B. TRAVIS–A free analyzer for trajectories from molecular simulation. J. Chem. Phys. 2020, 152, 164105.
- (44) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. J. Chem. Phys. 1999, 110, 6158–6170.
- (45) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew-Burke-Ernzerhof exchangecorrelation functional. J. Chem. Phys. 1999, 110, 5029–5036.
- (46) Hare, S. R.; Bratholm, L. A.; Glowacki, D. R.; Carpenter, B. K. Low dimensional representations along intrinsic reaction coordinates and molecular dynamics trajectories using interatomic distance matrices. *Chem. Sci.* **2019**, *10*, 9954–9968.
- (47) Götz, A. W.; Clark, M. A.; Walker, R. C. An extensible interface for QM/MM molecular dynamics simulations with AMBER. J. Comput. Chem. 2014, 35, 95–108.
- (48) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983, 79, 926–935.
- (49) Ottiger, P.; Pfaffen, C.; Leist, R.; Leutwyler, S.; Bachorz, R. A.; Klopper, W. Strong N–H···π Hydrogen Bonding in Amide-Benzene Interactions. J. Phys. Chem. B 2009, 113, 2937–2943.
- (50) Kozlovskaya, E.; Pitsevich, G.; Malevich, A.; Doroshenko, O.; Pogorelov, V.; Doroshenko, I.; Balevicius, V.; Sablinskas, V.; Kamnev, A. Raman spectroscopic and theoretical study of liquid and solid water within the spectral region 1600-2300cm<sup>-1</sup>. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2018**, *196*, 406–412.
- (51) Pastorczak, M.; Duk, K.; Shahab, S.; Kananenka, A. A. Combinational Vibration Modes in H2O/HDO/D2O Mixtures Detected Thanks to the Superior Sensitivity of Femtosecond Stimulated Raman Scattering. J. Phys. Chem. B 2023, 127, 4843–4857.

- (52) Wannier, G. H. The Structure of Electronic Excitation Levels in Insulating Crystals.
   *Phys. Rev.* 1937, 52, 191–197.
- (53) Thomas, M.; Brehm, M.; Kirchner, B. Voronoi dipole moments for the simulation of bulk phase vibrational spectra. *Phys. Chem. Chem. Phys.* **2015**, *17*, 3207–3213.
- (54) Han, R.; Ketkaew, R.; Luber, S. A Concise Review on Recent Developments of Machine Learning for the Prediction of Vibrational Spectra. J. Phys. Chem. A 2022, 126, 801– 812.
- (55) Gastegger, M.; Behler, J.; Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* 2017, *8*, 6924–6935.
- (56) Schienbein, P. Spectroscopy from Machine Learning by Accurately Representing the Atomic Polar Tensor. J. Chem. Theory Comput. 2023, 19, 705–712.
- (57) Hinton, G. E. Connectionist learning procedures. Artif. Intell. 1989, 40, 185–234.
- (58) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830.
- (59) The implementation of the machine learning model and the training and test set database are freely accessible following the link https://github.com/ANRMoMoPlasm/ MLPvibSpec. https://github.com/ANRMoMoPlasm/MLPvibSpec.

## **TOC** Graphic

