Computational Vibrational Spectroscopy

Markus Meuwly*

Abstract: Vibrational spectroscopy is a powerful technique to characterize the near-equilibrium dynamics of molecules in the gas and the condensed phase. This contribution summarizes efforts from computer-based methods to gain insight into the relationship between structure and spectroscopic response. Methods for this purpose include physics-based and machine-learned energy functions, and methods that separate sampling conformational space and determining the data for spectral analysis such as map-based techniques.

Keywords: Vibrational Spectroscopy · Machine Learning · Molecular Dynamics · Quantitative Simulations

1. Introduction

Optical spectroscopy, and in particular infrared and Raman spectroscopy, are versatile tools to determine the chemical composition and structure of molecules. When extended into the time-domain, techniques such as multidimensional infrared spectroscopy also provide a structure-sensitive instrument to characterize the environmental dynamics, couplings and energy transfer in solution. For the interpretation of such experiments, simulations play an important role. In the following, different approaches are summarized which aim at providing a molecularly refined picture of the structural dynamics underlying the experimentally observed spectroscopic features.

The present contribution revolves around dynamics-based approaches which explicitly account for the structural dynamics in solution. Methods primarily rooted in electronic structure calculations can typically only be applied to individual conformations in the gas phase. This has been done, for example, for small peptides for which experimentally measured conformer-specific spectra are available. The underlying structures were assigned by comparing normal modes determined from density functional theory (DFT) calculations for optimized structures sampled from either finite-temperature MD or Monte Carlo simulations. [1,2]

The central quantity for dynamics-based approaches to vibrational spectroscopy is the potential energy surface (PES) which describes how the total energy of a system changes with geometry. Computing and representing a full-dimensional PES suitable for vibrational spectroscopy is a formidable task in itself. Several methods available for this are briefly mentioned and typical examples are highlighted. The approaches range from augmented empirical energy functions with physics-based input such as multipolar electrostatics[3] to machine-learned energy functions. [4]

2. Molecular Dynamics Simulations with Physics-based Energy Functions

Empirical energy functions which are also called “force fields” have been extensively used to characterize the structure and dynamics of macromolecules, including peptides and proteins. [5-11] The extensive parametrization of any “general purpose force field” includes fitting to experimental structural and spectroscopic data for equilibrium geometries and force constants, experimental results on hydration free energies, heats of formation and other thermodynamic properties for van der Waals paramters, and to electronic structure data for atomic charges. As such, these energy functions are a useful zeroth order model for a wide variety of problems in structural biology and chemistry. However, for individual systems and specific observables more refined parametrizations are required and possible.

One particularly informative observable is the infrared spectrum of a solvated molecule. The solvent-induced red or blue shifts of the spectral lines provide a quantitative measure of the solvent-solute interaction. If time-resolved methods such as 2-dimensional IR spectroscopy are used, the frequency-fluctuation correlation function (FFCF) or alternatively the “center-line slope” is an observable and reflects the characteristic time-scale(s) of the solvent fluctuations to which the solute degrees of freedom are coupled. [12] Such fluctuations occur on the picosecond time scale which makes them ideally suited for rigorous sampling by MD simulations and with sufficiently long integration time the necessary averaging over conformational substates can be achieved for direct comparison with experiments. Typical time scales of such simulations are in the nanosecond range which are sufficient for systems such as ions in solution or ligands bound to proteins but not necessarily for ionic liquid or deep eutectic solvent with considerably increased viscosity.

Atomistic simulations with multipolar force fields have demonstrated that it is possible to realistically describe the 1d and 2d-spectroscopy of small molecules in electrostatically demanding environments such as in proteins or in water. [16,20,21] It was also shown for cyanide (CN⁻) in water that the same energy function for the solute is capable of correctly describing a range of condensed-phase properties including the solvent-induced shift, the decay time of the FFCF, the hydration free energy, and the vibrational energy relaxation rate in water (Fig. 1). [11,13-15] This indicates that physics-based refinements of such generic energy functions provide a meaningful extension for molecularly resolved investigations of complex systems. The spectroscopy of photodissociated CO in myoglobin (Mb) was also investigated by using a fluctuating multipolar representation for the electrostatics, [20,21] This allowed to assign for the first time the experimentally observed [22,23] split infrared spectrum and to identify the two peaks with two distinct conformational substates: one in which the oxygen end of CO pointed towards the heme-iron atom and a second state for which the carbon was closer to the Fe atom. [20,21,24] This model correctly captures the red shift of the two bands relative to gas phase CO, the splitting of the two peaks and their relative intensity.

Finally, it is also possible to refine energy functions by comparing experimentally determined IR spectra with those from computations. The infrared spectrum of acetylacetone (doubly methylated malonaldehyde) features a prominent band between 2000 and 3000 cm⁻¹ which is due to proton transfer across a low barrier. [26]

*Correspondence: Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland, E-mail: m.meuwly@unibas.ch
But all of them incur a considerably larger number of free parameters to be determined.

In recent years an alternative approach has matured which is based on using statistical models to represent precomputed energies and forces from electronic structure calculations. Such machine learning (ML) techniques do not necessarily require a parametrized form to be used but rather represent data given a set of kernel functions (kernel ridge regression) or by minimizing a loss function of a neural network (NN).

For kernel-based methods the long range physical shape of the PES can be encoded in the kernel which guarantees correct extrapolation to large separations. No such procedure is known for short range interactions. For NN-learned energy functions extrapolation to geometries outside the training set needs to be carefully assessed. One of the advantages of ML-based energy functions is that they contain all couplings between the degrees of freedom. This is very challenging for empirical energy functions. For instance, the CO bonds in protonated oxalate change between single and double-bond character depending on where the proton is located. Although such effects can be “encoded” in an empirical force field, capturing such effects from a globally valid, machine-learned energy function is more readily possible as has recently been done for formic acid dimer.

3. Molecular Dynamics Simulations with Machine-learned Potentials

One of the major challenges in empirical force field development is to find a suitable parametrized form of the energy function depending on internal coordinates for a molecule. Various extensions to the generic harmonic oscillator models for bonds and angles have been considered.

Morphing to a parametrized PES suitable for following proton transfer and comparing the resulting infrared spectrum with that from experiments yields an estimated barrier for proton transfer of 2.35 kcal/mol; see Fig. 2. Simulations with half and twice this barrier height demonstrated that the spectroscopic features characterizing H-transfer are directly sensitive to the barrier height. Subsequent machine learning reported a barrier height of 3.25 kcal/mol from transfer learning to the PNO-LCCSD(T)-F12 level of theory. Hence, the barrier height extracted from spectroscopy in this fashion is qualitatively correct although the value from transfer learning may also change at yet higher levels of theory, such as CCSD(T)-F12. Potential further improvement of the morphed PES can be achieved by including effects due to differences in zero point energy between the minimum energy and transition state structures, which will increase the experimentally determined barrier height.

Fig. 1: Structural and hydration dynamics of cyanide in water. Panel A: vibrational relaxation of the $v_{\text{CN}} = 1$ stretch through vibrational energy transfer (VET) into the water libration and bending motion (red arrows) from classical MD simulations consistent with experiment. Panel B: The 1d lineshape from simulations using multipoles (MTP) with those measured experimentally (horizontal bar). Panel C: Decay of the center line slope (equivalent to the FFCF, see text) from using point charges (PC, magenta) and MTP charge models compared with experiment (grey). Panel D: Hydration free energy of CN$^-$ from MTP simulations are consistent with those from experiment. Figure adapted from Refs. [14,16,18]
As an example for the performance of state-of-the-art ML-based methods for vibrational spectroscopy, formic acid monomer and dimer (FAM and FAD) in the gas phase is considered. Using PhysNet,[40] a reference machine-learned PES was determined at the MP2/aug-cc-pVTZ level of theory for FAM and FAD. The mean averaged error between reference calculations and the statistical model is 0.01 kcal/mol. Transfer learning the MP2-based PES to the CCSD(T)/aug-cc-pVTZ level of theory yields normal mode frequencies within 25 cm\(^{-1}\) on average compared with experiment for modes below 2000 cm\(^{-1}\). Including anharmonic corrections within second order vibrational perturbation theory (VPT2)[41] reduces this to 17 cm\(^{-1}\). For the OH-stretch mode the VPT2 calculations yield 3011 cm\(^{-1}\), compared with an experimentally reported, ~ 100 cm\(^{-1}\) broad absorption band with center frequency at ~ 3050 cm\(^{-1}\). Finally, using diffusion Monte Carlo (DMC)[42] calculations for the full-dimensional ground state potential energy and including corrections due to basis set superposition and basis set completeness errors yield a dissociation energy of \(D_0 = -14.23 \pm 0.08\) kcal/mol compared with an experimentally determined value of \(-14.22 \pm 0.12\) kcal/mol.[43]

It is of interest to note that experiment-guided refinement of an advanced force field based on molecular mechanics with proton transfer (MMPT)[44] the barrier height for double proton transfer in FAD could be inferred. For this, the height of the double well potential was adjusted to match the experimentally observed broad band associated with DPT in FAD. The resulting[45] barrier height was 7.2 kcal/mol which compares with an experimentally determined value from microwave spectroscopy of 7.3 kcal/mol.[46] In this fashion, information from vibrational spectroscopy can also be used to adapt (“morph”) PESs.[45]

Finally, machine-learned energy functions can also be used in a mixed quantum mechanics/molecular mechanics fashion to accurately describe the bonded energetics for spectroscopic probes used in protein 2-dimensional IR spectroscopy.[47] This was successfully done using reproducing kernel-based representations for the amide-I mode in insulin and trialanine or for azide attached to all alanine residues in Lysozyme.[48–50] Such simulations provide a positionally sensitive probe of the local and global protein dynamics to follow protein assembly or protein-ligand interactions.

4. Molecular Dynamics Simulations with Spectroscopic Maps

Determining the frequency trajectories \(\omega(t)\) required for computing the FFCF and 2d-IR response can be computationally prohibitive. One way to circumvent the expensive instantaneous normal mode or reduced-dimensionality quantum bound state calculations is to use spectroscopic maps.[51–54] Such maps can be parametrized from electronic structure calculations using model systems and exploit the fact that the frequency shift of an oscillator in the field of surrounding point charges can be approximately described by the Stark effect. Maps have been generated for a range of spectroscopic probes, including the amide-I stretch,[52] the nitrile stretch, the azido stretch, and others.[54] Recently, machine learning has also been applied to refine the amide-I map.[55]

With such spectroscopic maps it is then quite straightforward to determine the frequency trajectory for a particular oscillator from conventional MD simulations. For every snapshot to be analyzed the electric field at the position of the oscillator of interest is determined and related to the frequency shift by evaluating the spectroscopic map. This provides the information required to generate the FFCF from which important information about the structural dynamics around the spectroscopic reporter can be obtained.

One of the conceptual disadvantages of spectroscopic maps is the fact that the energy function used to run the MD simulations typically differs from the energy function used to evaluate the spectroscopic response which is also possible by using physics-based force fields. Furthermore, some maps have been generated for rigid labels as was the case for the amide-Imaps. Hence, the MD simulations need to be run with constrained CO distances for the maps to be valid. A direct comparison between map-based analyses and results from instantaneous normal modes and solutions of the nuclear Schrödinger equation has recently been given for insulin monomer and dimer. For this system it was found that the maps perform inferior compared with the other two approaches.[48]
5. Outlook

Molecular dynamics simulations with advanced energy functions provide important information about the structural dynamics of molecules in solution. Extensively sampling the conformational degrees of freedom is essential and only possible with efficient implementations of the energy functions. Treating spectroscopic probes with quantitatively accurate energy functions and using them in MD simulations provides information where to attach small molecules in order to be most sensitive to external perturbations such as binding of a ligand. For such and other applications the combination of experiment and simulation is indispensable and promises the necessary molecular-level information to control and design chemical systems with desired properties.

For the nuclear dynamics classical MD simulations have proven adequate for the purposes outlined in the present contribution. It is of interest to note that already 40 years ago it was pointed out for CO in the gas phase and in argon that classical MD simulations are even capable of correctly capturing the envelopes of R(J → J+1) and P(J → J+1) -branches similar to what quantum mechanical treatments provide. This work demonstrated that four different approaches give essentially the same band shapes, namely: I) a quantum mechanical treatment, II) the quantum-corrected result for k → 0, III) classical mechanical (Newtonian) treatment within classical linear response theory and quantum correction factors, and IV) experiment. More recently, this has also been found for dilute HOD in H2O for which the P-, Q-, and R-branch envelopes from classical MD simulations using quantum correction factors closely agree with experimental measurements.

These simulations found that even though the HOD rotations are treated classically, the rovibrational structure of the spectrum was described correctly. Hence, developing approximate quantum treatments will be of interest in order to delineate the range of applicability of classical mechanics for vibrational spectroscopy. Also, quantum effects including zero-point energy and tunneling are outside the scope of any classical mechanics-based method. For this reason, developing quantum methods applicable to the dynamics in solution remain an important quest in this field.

In summary, it is anticipated that MD simulations with improved energy functions together with experimental characterization of small molecules, peptides, and proteins in solution will provide important molecular-level insights into the dynamics of complex systems. With increasing accuracy of the energy functions it may even be possible to predict the dynamics and spectroscopy of such systems which is an important element for systems design.

Acknowledgments

The author acknowledges financial support from the Swiss National Science Foundation (NCCR MUST and Grant No. 200021-188724), the AFOSR, and the University of Basel.

Received: May 3, 2022

License and Terms
This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (https://chimia.ch/chimia/about).
The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2022.589