Frequency Range Selection Method for Vibrational Spectra

T. Q. Teodoro,†‡§ M. A. J. Koenis,¶ S. E. Galembek,†§ V. P. Nicu,§ W. J. Buma,¶ and L. Visscher†‡§

†Amsterdam Center for Multiscale Modeling, Faculty of Science, Vrije Universiteit Amsterdam, de Boelelaan 1083, 1081 HV Amsterdam, The Netherlands
‡Departamento de Química, FFCLRP, Universidade de São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, 14040-901 São Paulo, Brazil
§Van’t Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
¶Department of Environmental Science, Physics, Physical Education and Sport, Lucian Blaga University of Sibiu, loan Ratiu Street, Nr. 7-9, 550012 Sibiu, Romania

ABSTRACT: Theoretical calculations of vibrational properties are widely used to explain and predict experimental spectra. However, with standard quantum chemical methods all molecular motions are considered, which is rather time-consuming for large molecules. Because typically only a specific spectral region is of experimental interest, we propose here an efficient method that allows calculation of only a selected frequency interval. After a computationally cheap low-level estimate of the molecular motions, the computational time is proportional to the number of normal modes needed to describe this frequency range. Results for a medium-sized molecule show a reduction in computational time of up to 1 order of magnitude with negligible loss in accuracy. We also show that still larger computational savings are possible by using an additional intensity-selection procedure.

The vibrational structure of a molecule can be probed by means of several spectroscopy techniques, such as infrared (IR), Raman, vibrational circular dichroism (VCD), and Raman optical activity (ROA).1-2 At the core of the methods used to model these properties lies the harmonic approximation to molecular vibrations,3 which implies the diagonalization of a mass-weighted Hessian

\[
\hat{H} = M^{-1/2} H M^{-1/2}
\]

(with M being the atomic mass matrix and the tilde indicating that the Hessian is in mass-weighted coordinates) to obtain as eigenvectors the normal modes, \( \mathbf{U} \), and as eigenvalues the squares of the corresponding fundamental vibrational frequencies, \( \nu_i \). The computationally intensive task for constructing \( \mathbf{H} \) (the nonmass-weighted Hessian) in eq 1 is to calculate the second derivatives of the molecular energy \( E \) with respect to changes in the nuclear coordinates, \( \mathbf{R}_i \) and \( \mathbf{R}_j \) (the Greek indices are used in this equation only to represent a general basis for the atomic coordinates):

\[
H_{ij} = \frac{\partial^2 E}{\partial R_i \partial R_j}
\]

While evaluation of these derivatives is most easily done in Cartesian nuclear coordinates, this approach yields a \( 3N_{\text{atoms}} \times 3N_{\text{atoms}} \) dimensional Hessian that is time-consuming to construct in full. Several methods have therefore been proposed to reduce the number of molecular motions that need to be considered explicitly.3-8 All these methods can be viewed as formally transforming the mass-weighted Hessian in Cartesian coordinates, \( \hat{H}^{CC} \) (where the superscript CC denotes the use of Cartesian coordinates for each of the Hessian indices), to a \( K \)-dimensional \( (K < 3N_{\text{atoms}}) \) representation \( \hat{H}^{AA} \) (with the superscript AA referring to use of approximate coordinates for both indices):

\[
\hat{H}^{AA} = \mathbf{Q}^{CA} \hat{H}^{CC} \mathbf{Q}^{AC}
\]

where \( \mathbf{Q}^{CA} \) is a rectangular \( (3N_{\text{atoms}} \times K) \) transformation matrix. The key step is thereby the elimination of motions of molecular fragments that are expected to be less important for interpreting the experimental spectrum (thus \( K \) is equal to number of motions of interest). However, this reduction of the number of coordinates usually relies on the assumption that the normal modes of interest are localized in specific regions of the molecule.

Another approach is the Davidson procedure developed by Reiher and Neugebauer.9 The method is designed to iteratively refine the normal modes and associated eigenvalues (frequencies) of a set of specific vibrations that are provided as an initial guess. This so-called mode-tracking algorithm is particularly suited to determine local vibrations.10 For instance, it reproduced in only 20–30% of the computational time the vibrations associated with the thiophenolate anion adsorbed on a Ag(111) surface with the same accuracy as a full vibrational analysis of the whole system.11

Here we propose an even simpler “black box” approach in which the iterative refinement stage of the block-Davidson method of ref 9 is omitted entirely. We just compute a full spectrum with a fast approximate method then select and re-
evaluate only the normal modes in the spectral interval of experimental interest. This one-step procedure is sufficient if the set of eigenvectors provided by an approximate method spans the space of eigenvectors of the higher-level method in the same spectral interval, which is the case in many practical applications. To sketch the ease of implementation, we will use an index $j$ for approximate modes, while labeling Cartesian coordinates with the index $p$. We start by defining the gradient $g^c_p$ of the energy at the target level of theory with respect to nuclear displacements in Cartesian coordinates ($dR^c_p$) as

$$g^c_p = \frac{dE}{dR^c_p}$$

(4)

This quantity can be computed analytically in most electronic structure implementations. We then need a second derivative of the energy with respect to the selected set of approximate coordinates. By determining the first-order perturbed wave function with respect to these displacements these matrix elements can in principle also be obtained analytically, but it is easier to use numerical differentiation with an $n$-point stencil. For $n = 2$, we need $2K$ displacements along the normal modes of interest calculated at a lower level of theory to obtain a Hessian with one Cartesian and one approximate index, $\tilde{H}^{CA}$:

$$\tilde{H}^{CA}_{ji} = \frac{g^c_i(R^{0} + \Delta j \tilde{R}^A_j) - g^c_i(R^{0} - \Delta j \tilde{R}^A_j)}{2\Delta_j}$$

(5)

The displacements $\Delta j \tilde{R}^A_j$ from the equilibrium geometry ($R^{0}$) at the target level are taken along the mass-weighted vectors $\tilde{U}^{CA} = \tilde{M}^{1/2}U^{CA}$ of the approximate normal modes ($U^{CA}$) with a step size $\Delta_j$. This half-transformed matrix is then further transformed to a square matrix

$$\tilde{H}^{AA} = (\tilde{U}^{CA})^\dagger \tilde{H}^{CA} \tilde{U}^{CA}$$

(6)

and it is diagonalized to yield the vibrational frequencies $\nu_j$ and eigenvectors $U^{AR}$ (with AR indicating that this term was obtained on the basis of the approximate modes in the range selection interval). Back transformation of $U^{AR}$ provides the normal modes $U^{CR}$ expressed in Cartesian coordinates at the target level of theory (with R still referring to the range selection interval):

$$U^{CR} = U^{CA}U^{AR}$$

(7)

This representation of the normal modes at the target level of theory is then used to obtain the vibrational spectrum in the selected frequency range. Taking IR as an example, the intensities are straightforwardly obtained in a similar fashion. The $P$ tensor of the electric dipole moment derivatives with respect to nuclear displacements (also known as the atomic polar tensor, $\mu_i^{C}$) is thereby constructed via numeric differentiation by calculating electric dipole moment vectors ($\mu_i^{C}$) along the approximate modes

$$P^{CA}_{ji} = \frac{\mu^c_i(R^{0} + \Delta j \tilde{R}^A_j) - \mu^c_i(R^{0} - \Delta j \tilde{R}^A_j)}{2\Delta_j}$$

(8)

and then transforming the resulting tensor

$$P^{CR} = P^{CA}U^{AR}$$

(9)

Therefore, only energy gradients, $g^c_p$, and electric dipole moments, $\mu_i^c$, need to be calculated at the target level to provide the IR spectrum of the selected range.

While second derivatives can be calculated at all points on the potential energy surface, including the minima obtained with the approximate method, the harmonic approximation works best if the equilibrium geometry in eq 5 is obtained at the target level of theory. In addition, as is the case in optical properties such as VCD, one may need to compute a number of conformations to obtain proper spectra, which is preferably also done at the higher level of theory to get the best results. Because geometry optimization is, for large systems, much faster than calculation of the Hessian, these steps will not present a computational bottleneck.

The frequency range selection method described above was implemented in a Python-based script for obtaining IR absorption spectra of a pharmaceutical compound, dydrogesterone (Figure 1). Calculations were carried out with the 2017 version of the Amsterdam Density Functional (ADF) modeling suite of programs. The target (higher) level was a density functional theory (DFT) method, the Becke–Perdew exchange–correlation potential (BP86), in combination with Slater type orbital (STO) nonrelativistic valence triple-$\zeta$ basis sets extended with one polarization function (TZP). The electronic gas-phase ground-state structure obtained at the target level was used in all calculations. The approximate level chosen to provide the normal mode coordinates is a parametrized version of DFT, the third-order density functional-based tight binding method (DFTB3). Parameters for this approximation were taken from the Slater–Koster files for organic and biomolecules, 3ob-freq-1-2. Results were extracted by using the Python Library for Automating Molecular Simulation (PLAMS). Additionally, an experimental spectrum was obtained with dydrogesterone purchased from Sigma-Aldrich with a European Pharmacopoeia Reference standard. A sample of 0.41 M in deuterated chloroform was prepared and injected in a sealed cell with 3 mm thick CaF$_2$ windows and a 56 µm sample path length. Subsequently, the Fourier-transform vibrational absorption spectrum was obtained using a Bruker Vertex 70 spectrometer. A baseline correction was performed by using the absorption spectrum of pure deuterated chloroform. All spectra presented in the following were obtained by convoluting calculated stick spectra with a Lorentzian with half width at half-maximum of 4 cm$^{-1}$.

As shown in Figure 2, the DFT calculation reproduces the experiment very well. However, a large number of the simulated normal modes are not needed for this comparison as they are not present in the experimental spectrum, viz., CaF$_2$ absorbs below ~950 cm$^{-1}$. Out of the total 153 normal modes computed for this system ($3 \times 51$ atoms), more than a third (57) belong to this range. Additionally, the bands above 2800 cm$^{-1}$ in the experimental spectrum, which are associated with C–H stretches, are hardly distinguishable from each other, making the comparison with theory difficult. In the latter range there are 27 calculated bands. By also excluding the 6 rigid

---

Figure 1. Structure of dydrogesterone.
motions (translations and rotations of the system), only 66 bands distributed in the region from 950 to 1800 cm\(^{-1}\) are actually of interest, which correspond to only 43% of all computed modes.

When using the DFTB3/3ob-freq-1-2 method, computational time was reduced by 2 orders of magnitude with respect to the BP86/TZP calculation. However, as shown in Figure 2, the calculated IR absorption spectrum is quite different than the spectrum calculated at the BP86/TZP level and the experimental spectrum.

In the following, the DFTB3/3ob-freq-1-2 normal mode coordinates are used in the range selection method for the 950–1800 cm\(^{-1}\) interval with BP86/TZP being the target level. The resulting spectrum is shown in Figure 3. The overlap with the full DFT vibrational analysis is nearly perfect even though only 67 modes (44% of the total) were calculated at the same level of theory. In a comparison with the (two-point stencil) numerical evaluation of the full Hessian in a Cartesian basis, the range selection scheme would thus represent a \(\sim 56\%\) reduction in computational time. It has to be noted, however, that analytical Hessian calculations are usually a few times faster than numerical ones with the ADF implementations. With the same computational configuration, the range selection method was about 50% slower than the analytical calculation of the full spectrum.

Therefore, for the present method to be advantageous with respect to the analytical calculation of the full Hessian, a further reduction in the number of modes to be considered is necessary. First, one may additionally screen the modes according to the intensities calculated at the approximate level.\cite{23-25} As Figure 2 shows, although the DFTB bands are not overall reproducing the DFT ones, the ranges where one finds the most intense peaks are fairly similar in both calculations. Thus, we analyzed the outcome of selecting only the most intense bands calculated with the DFTB3/3ob-freq-1-2 model (within the 950–1800 cm\(^{-1}\) frequency range). Results for the intensity selection of 25 bands (16% of the total) are shown in the bottom of Figure 3. Most features of the full DFT spectrum are still reproduced, although this calculation was 40% faster than the analytical evaluation of the full spectrum. Second, one can simply consider a smaller range. In many cases, e.g. in peptides and other large (bio)molecules, one is often interested in only the bands associated with C=O stretching movements. In the dydrogesterone example, there are four very intense bands in the experimental spectrum from \(\sim 1580\) to \(\sim 1700\) cm\(^{-1}\) which are associated with C=O and C=C stretches. To account for this region and deviations in the DFTB frequencies, the range selection scheme was applied in the 1500–1800 cm\(^{-1}\) interval. Results are shown in Figure 4.

The overlap with the BP86/TZP spectrum in the selected range is still very good. One is thus able to obtain virtually the same information with only four normal modes being considered at a high level of theory.

In the example above, a 10\(\times\) reduction in computational time (in comparison with the analytical BP86/TZP calculation of the full spectrum) was achieved. Even larger speed-ups are possible through an analytical implementation.\cite{26} On the other hand, the numerical differentiation exemplified here can be easily parallelized (as already done in the mode-tracking procedure\cite{10}) given that the gradient calculations at each

![Figure 2](image2.png)

**Figure 2.** Comparison of the experimental (green) IR spectrum of dydrogesterone with spectra calculated at the BP86/TZP (black) and DFTB3/3ob-freq-1-2 (blue) levels. Intensities in the DFTB spectrum were scaled by a factor of 0.3 for better visualization. The experimental band around 2300 cm\(^{-1}\) is not due to dydrogesterone but arises from an experimental artifact due to the solvent.

![Figure 3](image3.png)

**Figure 3.** Comparison of IR spectra of dydrogesterone in the 950–1800 cm\(^{-1}\) frequency interval calculated at the BP86/TZP level (black), by means of the frequency range selection scheme (red), and with the additional intensity-selection procedure. Only the 25 most intense bands predicted by DFTB3/3ob-freq-1-2 are calculated at the BP86/TZP level (purple).

![Figure 4](image4.png)

**Figure 4.** Comparison of IR spectra of dydrogesterone calculated at the BP86/TZP (black) and DFTB3/3ob-freq-1-2 (blue) levels and by means of the frequency range selection scheme (red) from 1500 to 1800 cm\(^{-1}\) using normal mode coordinates calculated at the DFTB level. Intensities in the DFTB spectrum were scaled by a factor of 0.3 for better visualization.
DFTB Hessian yield a completely different result at both (DFT and DFTB) levels, the inaccuracies in the atomic axial tensor (AAT), which appears in the magneto term in the rotational strength equation, cannot be calculated in the same fashion as the APTs (eq 8), but it can be obtained cheaply at a high level through implementation of the range selection method outlined here is shown to be a powerful tool for obtaining accurate vibrational spectra of a selected frequency interval. As the computational effort is limited to the number of normal modes within the region, this method will be particularly interesting for large molecules (e.g., proteins) for which only a handful of modes out of thousands, such as amide carbonyl stretches or low-lying modes, are of interest.

**AUTHOR INFORMATION**

**Corresponding Author**

*E-mail: lvischer@vu.nl.

**ORCID**

T. Q. Teodoro: 0000-0003-4590-0056
S. E. Galembeck: 0000-0001-5018-4643
L. Vischer: 0000-0002-7748-6243

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This research received funding from The Netherlands Organisation for Scientific Research in the framework of the Fund New Chemical Innovations (NWO Project Nr. 731.014.209). T.Q.T. thanks the São Paulo Research Foundation (FAPESP Projects Nr. 2016/07787-4 and 2016/23165-3) for a postdoctoral grant. S.E.G. also acknowledges FAPESP (Projects Nr. 2008/02677-0 and 2014/50265-3) and the National Council for Scientific and Technological Development (CNPq Project Nr. 308254/2016-3). We also thank Dr. R. Rüger, M.Sc. M. Handzlik, and Dr. S. van Gisbergen for useful discussions.

**REFERENCES**


**Figure 5.** Comparison of VCD spectra of C$_{105}$H$_{48}$O$_{18}$ calculated at the BP86/TZP (black) and DFTB3/3ob-freq-1-2 (blue) levels (only in the calculation of the Hessian, as the APT and AAT applied to obtain both spectra were calculated with the DFT level), and by means of the frequency range selection scheme: 1100–1400 (red) and 1700–1800 (orange) cm$^{-1}$ using normal mode coordinates calculated at the DFTB level.


