2018

# A systematic benchmarking of computational vibrational spectroscopy with DFTB3: Normal mode analysis and fast Fourier transform dipole autocorrelation function

Venkatasaisandeep Inakollu
*University of Wollongong*, vi244@uowmail.edu.au

Haibo Yu
*University of Wollongong*, hyu@uow.edu.au

## Publication Details

Inakollu, V. Sandeep. & Yu, H. (2018). A systematic benchmarking of computational vibrational spectroscopy with DFTB3: Normal mode analysis and fast Fourier transform dipole autocorrelation function. Journal of Computational Chemistry, 39 (25), 2067-2078.

# A systematic benchmarking of computational vibrational spectroscopy with DFTB3: Normal mode analysis and fast Fourier transform dipole autocorrelation function

**Abstract**
Computational vibrational spectroscopy serves as an important tool in the interpretation of experimental infrared (IR) spectra. In this article, we present a systematic benchmarking study of DFTB3 with two different computational vibrational spectroscopic methods, based on either normal mode analysis (NMA) or fast Fourier transform dipole autocorrelation function (FT-DAC). The results were compared with experimental data and theoretical calculations with B3LYP/cc-pVTZ. The empirical scaling factors for DFTB3/NMA, DFTB3-freq/NMA, and DFTB3/FT-DAC methods are 0.9993, 1.0059, and 0.9982, respectively. We also demonstrate the significance of anharmonicity and conformational sampling in vibrational spectroscopic calculations on flexible molecules. As expected, DFTB3/FT-DAC predicted the anharmonic vibrational peaks more accurately than DFTB3/NMA and NMA spectra are highly dependent on the initial structures. The potential limitations of DFTB3 for vibrational spectroscopic calculations and the challenges in assigning the FT-DAC spectral peaks were noted. DFTB3/FT-DAC is expected to serve as a promising technique in computational spectroscopy in complex biomolecular systems.

**Disciplines**
Medicine and Health Sciences

**Publication Details**
Inakollu, V. Sandeep. & Yu, H. (2018). A systematic benchmarking of computational vibrational spectroscopy with DFTB3: Normal mode analysis and fast Fourier transform dipole autocorrelation function. Journal of Computational Chemistry, 39 (25), 2067-2078.

# A systematic benchmarking of computational vibrational spectroscopy with DFTB3: Normal mode analysis and fast Fourier transform dipole autocorrelation function

V. S. Sandeep Inakollu[*†], Haibo Yu [‡§¶‖]

July 18, 2018

**Abstract**

Computational vibrational spectroscopy serves as an important tool in the interpretation of experimental infrared (IR) spectra. In this paper, we present a systematic benchmarking study of DFTB3 with two different computational vibrational spectroscopic methods, based on either normal mode analysis (NMA) or fast Fourier transform dipole autocorrelation function (FT-DAC). The results were compared with experimental data and theoretical calculations with B3LYP/cc-pVTZ. The empirical scaling factors for DFTB3/NMA, DFTB3-freq/NMA and DFTB3/FT-DAC methods are 0.9993, 1.0059 and 0.9982, respectively. We also demonstrate the significance of anharmonicity and conformational sampling in vibrational spectroscopic calculations on flexible molecules. As expected, DFTB3/FT-DAC predicted the anharmonic vibrational peaks more accurately than DFTB3/NMA and NMA spectra are highly dependent on the initial structures. The potential limitations of DFTB3 for vibrational spectroscopic calculations and the challenges in assigning the FT-DAC spectral peaks were noted. DFTB3/FT-DAC is expected to serve as a promising technique in computational spectroscopy in complex biomolecular systems.

Keywords: computational vibrational spectroscopy, DFTB3, normal mode analysis, dipole autocorrelation function, molecular dynamics simulations. ■

[*]School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong NSW 2522, Australia

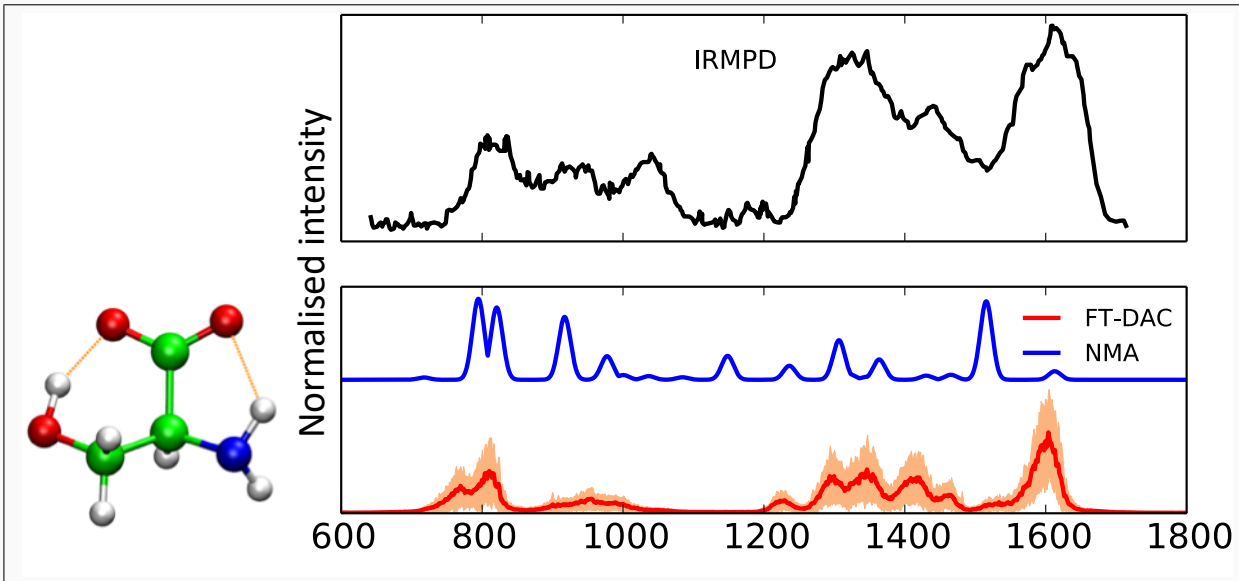
[†]Molecular Horizons, University of Wollongong, NSW 2522, Australia

[‡]School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong NSW 2522, Australia

[§]Molecular Horizons, University of Wollongong, NSW 2522, Australia

[¶]Illawarra Health and Medical Research Institute, Wollongong NSW 2522, Australia

[‖]Corresponding author: hyu@uow.edu.au




An overall improvement has been observed with the FT-DAC based method over the NMA based one with DFTB3 for flexible and anharmonic systems.

# INTRODUCTION

In the modern analytical chemistry, vibrational spectroscopy is an important technique for identifying chemical species[1–5] and understanding reaction mechanisms.[5–10] Infrared (IR) spectroscopy is one of the well-established traditional vibrational spectroscopies. It is based on the unique absorption frequencies pattern of IR light when infrared radiation excites vibrational transitions of molecules.[4,11] IR spectra provide precise information about a molecule, such as the nature of the chemical bond and its surrounding environment.[4] However, deriving the 3D structural information directly from the experimental IR spectra is not always straightforward. Often, theoretical studies are employed to aid the interpretation of the experimental IR spectra.[9,12,13]

Among the various computational vibrational IR spectroscopic methods[14–22], those based on normal mode analysis (NMA)[23–26] and Fourier transform dipole autocorrelation function (FT-DAC)[27,28] are the two most popular approaches. NMA is based on the harmonic approximation by assuming that the potential energy surface (PES) behaves like a harmonic potential. The spectrum is obtained by diagonalising the secondary derivative of the potential energy of the molecule in its minimal energy structure, which makes this method computationally economical. However, this method completely neglects the anharmonic nature of the PES.[11] To overcome such issues, often empirical scaling factors parametrised with a larger data set are employed.[15,16,18] It has been noted that such empirical scaling factor calculations are often affected by choice of the data set used in the study.[18] As the NMA spectrum is calculated from a static structure, a well-optimised structure is necessary to obtain an accurate spectrum. This limitation makes NMA difficult to apply to dynamic systems such as flexible molecules or bulk phase systems, due to the presence of multiple local minima. To build a comparable computational counterpart to the experimental IR spectra, a method needs to consider the anharmonicity of the PES. As per Wiener-Khintchine theorem,[17,22] the Fourier transformed autocorrelation function of any attribute in time domain gives its spectral signatures in the frequency domain. Based on this theorem, the change in the dipole moment represented in an IR spectrum can be built from the FT-DAC function.[27–29] As there are no assumptions made about the nature of PES, the FT-DAC method

inherently includes the anharmonicity in the vibrational spectrum.[21] Furthermore, the applicability of this method to flexible and bulk phase systems is another important advantage, provided that sufficient sampling can be achieved to cover all the relevant thermodynamically accessible conformations.

FT-DAC method has already been proven as an effective tool to calculate the vibrational spectrum using *ab initio* methods based on density functional theory (DFT).[20,21,30–32] For instance, Thomas *et al.* used the FT-DAC with *ab initio* molecular dynamics (AIMD) simulations to calculate the IR and Raman spectra of small molecules.[20] Li *et al.* reported the IR spectra of the proton-bound dimethyl ether dimer with AIMD simulations.[30] However, AIMD simulations are limited by the number of atoms in the system and the timescale one can sample. Therefore, it is challenging to apply *ab initio* or density functional theory (DFT) methods to large (bio)molecular systems.

Third-order density functional tight binding theory (DFTB3) can be an alternative solution to this problem.[33–38] DFTB3 is a recent extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB or DFTB2).[33,34] DFTB2 was derived from a second-order expansion of the density functional theory (DFT) total energy around a given reference density and it calculates the zeroth and first order terms using a linear combination of atomic orbitals of a minimal basis and a two-center integral approximation.[36,39] The density fluctuation in DFTB2 is approximated via a charge redistribution in a self-consistent manner, where the electron-electron interaction of excess charge on one atom is approximated by the Hubbard parameter. DFTB3 combines an improved Coulomb interaction between atomic partial charges and the complete third-order expansion of the DFT total energy.[33–35,37] These refinements substantially improve the description of hydrogen binding energies, proton affinities and hydrogen transfer barriers, which makes DFTB3 particularly applicable to biomolecular systems.

Previously the IR spectra of protonated water clusters have been calculated with the FT-DAC method based on the SCC-DFTB method.[14] Kaminski *et al.* used a similar approach to calculate the Raman spectra for a set of 10 small molecules.[40] Several studies have been conducted to benchmark the harmonic vibrational frequencies using both *ab initio*[18,41] and semi-empirical methods including AM1,[15,18] PM3,[15,18] PM6[16] and SCC-DFTB.[15] However,

a systematic benchmarking of computational vibrational spectroscopy with DFTB3 is yet to be performed. In this work, we have benchmarked DFTB3 performance with both NMA and FT-DAC with a large group of chemical systems. Their strengths and potential limitations are discussed with selected case studies. This study also includes a comparative study of the harmonic vibrational frequencies using B3LYP/cc-pVTZ.

# METHODOLOGY

## Normal mode analysis (NMA)

In the NMA calculations, the Hessian matrix can be obtained either with numerical differentiation with the finite difference[19] or analytical differentiation schemes.[24,42] It has been shown that the Hessian obtained through the numerical differentiation is in excellent agreement with those based on the analytical scheme with SCC-DFTB.[14,15] The vibrational frequencies are then calculated by diagonalising the Hessian matrix.[24] IR intensities of each normal mode ($I_k$) are calculated through the projection of the molecular dipole derivative ($\nabla\vec{\mu}$) onto the corresponding eigenvector ($\vec{L}_k$),[24]

$$I_k = \frac{N\pi}{3c^2}\left|\nabla\vec{\mu}\cdot\vec{L}_k\right|^2 \tag{1}$$

where $N$ represents Avogadro's number, and $c$ is the speed of light.

## Fast Fourier transform dipole autocorrelation function (FT-DAC)

In FT-DAC method, the pattern of the dipole moments ($\vec{\mu}(t)$) in a molecule as a function of time is calculated from the MD trajectories using an autocorrelation function[29] and subsequent Fourier transformation gives the IR vibrational spectrum in the frequency domain.[17,20,22] The equation of classic FT-DAC function is as follows[27,28]

$$I_{cl} = \frac{1}{2\pi}\int_{-\infty}^{\infty} dt \exp\left[-i\omega t\right]\left\langle\vec{\mu}(0)\cdot\vec{\mu}(t)\right\rangle \tag{2}$$

Using equation 2 above, IR absorption coefficient $\alpha_{QC}$ is calculated with

$$\alpha_{QC}(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)}\right]\left(1-\exp\left(-\beta\hbar\omega\right)\right)Q_{QC}(\omega)I_{cl}(\omega) \tag{3}$$

where $V$ is the sample volume, $\hbar$ is the Plank's constant, $c$ is the speed of light and $n(\omega) \simeq 1$ is the refractive index of the medium.

Despite the electronic degrees of freedom are fully treated with QM models, the nuclear motions are still approximated by the classic Newtonian equation.[43] To rectify these approximations of such a treatment, several different methods[44–52] have been developed to model quantum time correlation functions. However, such methods are computationally expensive. Instead, Ramirez *et al.*[43] proposed an empirical harmonic correction term $Q_{HC}$ as an effective correction for IR intensity calculations of floppy molecules with large anharmonicity,

$$Q_{HC} = \frac{\beta\hbar\omega}{1 - \exp\left[-\beta\hbar\omega\right]}. \tag{4}$$

The final IR absorption coefficient is given by

$$\alpha_{QC}(\omega) = \left[\frac{4\pi^2\omega}{3V\hbar cn(\omega)}\right] \beta\hbar\omega I_{cl}(\omega). \tag{5}$$

## Computational details

In this section, we briefly outline the technical details of the theoretical IR spectral calculations. The test data set is comprised of 40 molecules, which were mostly organic, listed in the spreadsheet in the Supporting Information (SI). The data set is chosen to include most of the commonly occurring functional groups in the biomolecular systems. The experimental and B3LYP/cc-pVTZ vibrational frequencies are obtained from the NIST Computational Chemistry Comparison and Benchmark Database[53] and NIST chemistry webbook.[54] All the DFTB3 vibrational frequency calculations are performed with CHARMM (c38a1) software.[55] The DFTB3 parameters (3ob-3-1) are obtained from the DFTB official website.[56] We noted that a special set of parameters (3ob-3-1-freq, labelled as DFTB3-freq) aiming at an improved performance for the vibrational spectra had been developed for DFTB3 at the compromise of the energetic properties.[34] For comparative analysis, we also included these parameters for vibrational frequency calculations. However, for practical applications with the FT-DAC spectra calculation, an accurate description of energetics is also required. Furthermore, our long-term goal is to apply computational vibrational spectroscopy together with a combined quantum mechanics/molecular mechanics (QM/MM) model to study biomolecular systems. Due to this reason, we have excluded these parameters for the FT-DAC method.

In the NMA based approach, the Hessian matrix was constructed using a numerical differentiation scheme with the finite step size of 0.001 Å.[14] The vibrational frequencies are obtained by diagonalisation of the Hessian matrix, and the intensities are calculated using Equation 1. The resulting IR intensities are convoluted using the Gaussian line shape function with the full width at half maximum (FWHM) of 20 $cm^{-1}$.

In the FT-DAC based approach, we followed the previously developed protocol as detailed in Ref.[14]. Briefly, the molecular dynamics (MD) simulation is equilibrated for 50 ps with a time step of 0.5 fs in the Langevin dynamics at a specific temperature. The following production run is carried out for $\sim$ 32 ps, sampling the micro-canonical ensemble (NVE) using the velocity-Verlet integrator. The positions ($\vec{r}_i$) and the Mulliken charges ($q_i$) for atom $i$ at each step are printed out and the dipole moment ($\vec{\mu}$) at each step is calculated using Equation 6.

$$\vec{\mu} = \sum_{i=1}^{n} q_i \vec{r}_i \tag{6}$$

In this study, the entire system is treated quantum mechanically in the gas phase. Multiple independent trajectories starting from different equilibrated structures and velocities were carried out to enhance conformational sampling. The average FT-DAC IR spectra are calculated from multiple independent trajectories using Equations 2 and 5. A Blackman filter[57] is used in FT-DAC to minimise noise. For Fourier transformation analysis, 65,536 autocorrelation points with a timestep of 0.5 fs are used, resulting in the spectral resolution of $\sim$1 $cm^{-1}$. The continuous wavelet transform (CWT) method[58] is used to identify the possible vibrational frequency peaks from the FT-DAC vibrational spectra.

## Characterisation of vibrational peaks in FT-DAC

The major goal of modelling the vibrational spectra is to interpret the peaks/bands in the experimental counterpart and their assignments to their respective vibrational modes. However, it is often difficult to obtain such information directly from the FT-DAC spectrum. Tentative vibrational modes can be obtained from the Fourier transform of the atomic velocity autocorrelation function (FT-VAC). Alternative methods have been developed to address this issue. This includes by decomposing the spectral peaks as a sum of effective vibrational

modes extracted from the FT-VAC[59] or by describing the localized vibrational modes as a linear combination of non-redundant internal coordinates.[60] Nishimura *et al.* used the FT-TAC of symmetrised internal vibrational modes in the prediction of vibrational spectra of methanol dimer.[61] Bowman *et al.*[62] and Kaledin *et al.*[63] developed driven molecular dynamics approach to perform full normal mode analysis without calculating the Hessian matrix for larger biomolecular systems. Thaunay *et al.* used a similar approach to assign the vibrational modes for dipeptide.[64] Other types of methods have also been proposed in the literature for assigning or approximating the vibrational spectra are instantaneous normal mode analysis[65,66] and principle mode analysis.[67–69]

Since the molecules in our database are relatively rigid, we decided to assign the FT-DAC peaks to the closest NMA vibrational modes. If the mean absolute percentage error (MAPE) between NMA and FT-DAC vibrational mode is more than 3% (estimated MAPE) or the assignment is ambiguous (e.g., peak overlaps), we rechecked the FT-DAC peak attributes with the FT-TAC of the localised vibrational modes, following the work by Nishimura *et al.*[61] Out of 307 vibrational modes, 19 FT-DAC vibrational modes were deviated from the NMA vibrational modes by more than 3% (MAPE). They were listed in the Table S1 in SI. The Cartesian coordinates of the molecules are extracted at each timestep from the NVE MD simulations. Then, the internal coordinates of each internal mode $\nu_N$ are calculated, defined in the Table S1 and Figure S1 in SI. Power spectra of these internal modes $\nu_N$ are obtained by Fourier transformation of autocorrelation function $\langle \nu_N(0) \cdot \nu_N(t) \rangle$ for 50 independent trajectories. After reassignment, we found only minimal deviation between the NMA and FT-DAC vibrational modes.

## Scaling factor

The calculated harmonic vibrational frequencies are known to overestimate the actual vibrational frequencies.[18] The major source of this disagreement is from neglecting the anharmonic component of the oscillation, incomplete incorporation of electron correlation and use of a finite basis set.[15,16,18] Even though there is no theoretical basis, conventionally, such overestimations have been corrected with the scaling factors.[15,16,18] In this study, the scaling factor is calculated using the least square procedure by minimising the residuals. The scaling factor

$\lambda$ is calculated by

$$\lambda = \frac{\sum_{i=1}^{N} \omega_i^{theor} \nu_i^{exp}}{\sum_{i=1}^{N} \left(\omega_i^{theor}\right)^2} \tag{7}$$

where $\omega_i^{theor}$ and $\nu_i^{exp}$ are the theoretical and experimental vibrational frequencies respectively and $N$ is the number of modes. After obtaining the optimal scaling factor, we calculated the minimised residual ($\Delta_i^{min}$) for each mode

$$\Delta_i^{min} = \left(\lambda \omega_i^{theor} - \nu_i^{exp}\right)^2 \tag{8}$$

The root mean square error ($rms_{tot}$) of all the vibrational modes are determined by

$$rms_{tot} = \left(\sum_{i=1}^{N} \frac{\Delta_i^{min}}{N}\right)^{1/2} \tag{9}$$

Using the calculated scaling factor ($\lambda$), the corrected vibrational frequencies are calculated by

$$\omega_i^{\sim calc} = \lambda \omega_i^{calc} \tag{10}$$

where $\omega_i^{\sim calc}$ is the scaled harmonic frequency, and $\omega_i^{calc}$ is the unscaled harmonic frequency.

## Numerical accuracy

The numerical accuracy of these methods can be quantified using the statistical deviations. Four types of deviations have been used in this study including

(a). maximal absolute deviation $\sigma_{MAX}$

$$\sigma_{MAX} = \max(\left|\omega_i^{theor} - \nu_i^{exp}\right|); \tag{11}$$

(b). average absolute error

$$\sigma_{MAD} = \frac{1}{N} \sum_{i=1}^{N} \left|\omega_i^{theor} - \nu_i^{exp}\right|; \tag{12}$$

(c). standard deviation

$$\sigma_{STD} = \left(\frac{1}{N} \sum_{i=1}^{N} \left(\omega_i^{theor} - \nu_i^{exp}\right)^2\right)^{1/2}; \tag{13}$$

and (d). weighted absolute error

$$\sigma_{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\left|\omega_i^{theor} - \nu_i^{exp}\right|}{\nu_i^{exp}}. \tag{14}$$

# RESULTS AND DISCUSSION

## Section A: Convergence in the FT-DAC simulations

Sufficient conformational sampling, especially for flexible molecules at finite temperatures, is required to represent the conformational ensemble under the corresponding experimental conditions properly. A MD simulation of ∼30 ps with 65,356 DAC points provides a resolution of ∼1 $cm^{-1}$ in the FT-DAC spectrum. However, such a timescale is often too short for flexible systems. For instance, it is evident as the noisy FT-DAC spectrum for methanol shown in Figure 1. Experimentally, the OH stretch peak (around 3600 $cm^{-1}$) is broad in nature due to the high degree of rotational motion around its axis. Without a sufficient conformational sampling, the FT-DAC method predicted three different spectral signatures in the ∼3600 $cm^{-1}$ region. To overcome this issue, the FT-DAC spectra based on multiple independent trajectories were obtained[61]. The convergence of the FT-DAC spectra with different levels of conformational sampling was monitored. The FT-DAC method is able to predict both the position and the shape of the hydroxyl stretch peak accurately if a sufficient conformational sampling is achieved. Similarly, by examining one of the most flexible molecules in the database - dimethyl ether (Figure S2 in SI), it is concluded that 50 independent trajectories of ∼30 ps each are sufficient for the molecules included in the database.

## Section B: Overall performance of NMA and FT-DAC

To be consistent with all the methods in this study, we only included the vibrational modes which appeared in the FT-DAC spectra for detailed comparison (307 vibrational modes). This systematic selection procedure excluded the low-intensity vibrational modes, and we also excluded the vibrational modes lower than 500 $cm^{-1}$. The detailed experimental and unscaled theoretical vibrational frequencies calculated with B3LYP/cc-pVTZ (NMA), DFTB3 (NMA and FT-DAC) and DFTB3-freq (NMA) methods are provided in the spreadsheet in SI.

### Unscaled vibrational frequencies

Figure 2 presents the correlation pattern of vibrational frequencies with NMA (B3LYP/cc-pVTZ, DFTB3 and DFTB3-freq) and FT-DAC (DFTB3) with respect to the experimental vibrational modes. The statistical significance of these methods in vibrational spectroscopy calculations is listed in Table 1. In terms of numerical accuracy, the unscaled harmonic vibrational frequencies (unscaled) by DFTB3 (Figure 2(b-c)) are surprisingly slightly better than B3LYP/cc-pVTZ (Figure 2(a)). Between the two sets of DFTB3 parameters, modified 3ob-3-1 parameters for vibrational frequencies (DFTB3-freq) clearly outperformed the original DFTB3 parameter set. The major improvements gained with DFTB3-freq parameters are from the C=C or C=N stretch modes. DFTB3 parameters overestimated these modes by >130 $cm^{-1}$, whereas, DFTB3-freq parameters predicted them accurately. In the case of B3LYP/cc-pVTZ, a systematic overestimation of vibrational modes is observed in Figure 2(a). Corrections for such overestimations are further elaborated in the Section **Scaling factors**.

When comparing the DFTB3 (3ob-3-1) performance between the NMA method (Figure 2(b)) and the FT-DAC method (Figure 2(d)), we found that the latter marginally outperformed with the relatively rigid small molecules included in the current database. The $\sigma_{MAD}$ of NMA and FT-DAC are 42 $cm^{-1}$ and 41 $cm^{-1}$, respectively (Table 1). The standard deviation ($\sigma_{STD}$) for NMA is only 1 $cm^{-1}$ larger than that for FT-DAC. As an example, the experimental, NMA and FT-DAC spectra of methylamine are compared in Figures S3.

Furthermore, we characterised the performance of NMA and FT-DAC with respect to specific vibrational modes of each functional group (Table S2 in the SI). Vibrational modes of frequently used vibrational probes such as carbonyl and nitrile group are predicted accurately by both NMA (DFTB3 and DFTB3-freq) and FT-DAC (DFTB3) methods. Vibrational frequency of carbonyl (C=O) probe is predicted to be 36 $cm^{-1}$, 55 $cm^{-1}$ and 35 $cm^{-1}$ by NMA (DFTB3 and DFTB3-freq) and FT-DAC (DFTB3), respectively. Whereas nitrile probe is predicted as 19 $cm^{-1}$, 38 $cm^{-1}$ and 7 $cm^{-1}$ by NMA (DFTB3 and DFTB3-freq) and FT-DAC (DFTB3), respectively. In contrast, both NMA and FT-DAC have difficulties in predicting the N-H stretch mode frequencies. The mean absolute deviations of this mode are 71 $cm^{-1}$, 70

$cm^{-1}$ and 77 $cm^{-1}$ for NMA (DFTB3 and DFTB3-freq) and FT-DAC (DFTB3), respectively. In case of the N-H stretch peaks in amines, the symmetric N-H stretch peak in primary amines is nearly undetectable in the experimental IR spectra. These features are accurately reproduced in the FT-DAC spectra, whereas, NMA predicted these vibrational modes with high intensities (Figures S3 in SI). Even though the overall deviation of DFTB3-freq is much lower than that of DFTB3 parameters, the original DFTB3 parameters performed well in predicting the vibrational frequencies of these specific functional groups mentioned in Table S2 in the SI.

## Scaled vibrational frequencies

With Equations 7 and 9, scaling factors and the resulting $rms_{tot}$ errors are calculated, respectively, for B3LYP/cc-pVTZ (NMA), DFTB3 (NMA and FT-DAC), and DFTB3-freq (NMA). The results are included in Table 2 and Figure 2. As expected, B3LYP/cc-pVTZ performance is improved immensely. With a scaling factor of 0.9666, the $\sigma_{STD}$ of B3LYP/cc-pVTZ (NMA) came down from 72 $cm^{-1}$ to 29 $cm^{-1}$. Witek and Morokuma also reported an improvement of 27 $cm^{-1}$ for B3LYP/cc-pVDZ.[15]

In the case of DFTB3, the optimal scaling factor required for NMA and FT-DAC methods are 0.9993 and 0.9982, respectively (Table 2). Since the scaling factors are very close to 1, no improvement upon scaling was observed. It indicated the lack of systematic errors, *i.e.,* underestimation or overestimation in the vibrational frequency calculations with DFTB3. Witek and Morokuma also reported a minimal improvement with both DFTB and DFTB2 for NMA using scaling factors.[15] Whereas with a scaling factor of 1.0059, $\sigma_{STD}$ for DFTB3-freq came down only by 1 $cm^{-1}$. In summary, only a minor improvement was observed when applying scaling factors for DFTB3 based vibrational calculations. In the practical applications, it isn't necessary to apply scaling factor for DFTB3 based vibrational spectra calculations.

## Section C: Application of FT-DAC to flexible molecular systems

In this section, we have extended the study to demonstrate the advantages of FT-DAC in solving real-time problems for more flexible molecules of chemical or biological significance. The potential limitations of the NMA method have been already discussed in the **Introduction**. It is expected that as the anharmonicity component increases, the error in the vibrational mode prediction by NMA increases. Apart from the anharmonicity issue, NMA does not represent the precise experimental conditions either. For example, in the experimental studies the molecular structure is allowed to move from one conformation to another, which is difficult to treat with NMA. Conventionally to overcome the conformational sampling issue, several IR spectra calculated from a number of energetically favoured conformations to match with the experimental IR spectrum. However, there are a number of limitations in this strategy. Firstly, it is computationally expensive to calculate the IR spectra of all the possible conformers for large and flexible molecules particularly with *ab initio* or DFT methods. Secondly, the conformer with the lowest potential energy might not represent the most populated conformation in the ensemble at finite temperatures. The measured experimental IR spectra are an average spectrum obtained from all the accessible conformations of that molecule at the particular condition. Finally, the experimental temperature is ignored in NMA. FT-DAC naturally overcomes these limitations of NMA and simulates a more realistic spectrum by mimicking the experimental conditions.

### Serine

In this study, theoretical (NMA and FT-DAC) vibrational spectra of four different serine conformations are presented. The atomic coordinates of three conformers (I, II and III) were based on the work of Jones *et al.*[70] and Oomens *et al.*[71]. Based on the intra-molecular hydrogen bonding network, these three conformers of serine are characterised by two distinct orientations. Conformer I and II have the intra-molecular hydrogen bonding network of OH$\cdots$OCO$\cdots$HNH and Conformer-III has the intra-molecular hydrogen bonding network of OH$\cdots$NH$\cdots$OCO. Conformer I and II differ by forming the hydrogen with either of the two-amino hydrogen atoms with different H–C–N–H dihedral angles. A new con-

former (IV) was identified based on the potential of mean force (PMF) simulations (Figure S4). In contrast to the other three conformers, Conformer-IV has only one hydrogen bond OH$\cdots$OCO. The structural and energetic properties of these conformers are summarized in Table 3. Among the conformers, Conformer-I is the most energetically favoured. The relative energy of Conformer-II, III and IV are 0.2 kcal/mol, 3.3 kcal/mol and 2.2 kcal/mol with B3LYP/6-31++G**, respectively. Due to the involvement in the hydrogen bonding network, the carboxyl (OCO) symmetric stretch, OH bend, $NH_2$ scissoring and OCO asymmetric stretch vibrational modes are assumed to be highly anharmonic in nature. Based on the experimental infrared multiple photon dissociation (IRMPD) spectroscopy spectra by Redlich and coworkers,[71] these modes have the vibrational frequencies of $\sim$1315 cm$^{-1}$, $\sim$1420 cm$^{-1}$, $\sim$1575 cm$^{-1}$ and $\sim$1600 cm$^{-1}$ respectively. The $NH_2$ scissoring ($\sim$1550 cm$^{-1}$) and asymmetric OCO stretch ($\sim$1600 cm$^{-1}$) modes are not clearly resolved. The computed vibrational frequencies of the four serine conformations by both NMA and FT-DAC methods are presented in Table 4.

**NMA of Serine:** NMA predicted the vibrational frequencies of all the conformations relatively accurate to the experimental vibrational frequencies (Table 4 and Figure 3). The experimental spectrum is mainly dominated by the symmetric and asymmetric carboxyl bend (OCO) mode, while, NMA measured weak to medium intensities for these vibrational modes and is mainly dominated by the $NH_2$ scissoring mode (Figure 3). In case of Conformer-III, OCO asymmetric carboxyl peak is blue shifted up to 30 cm$^{-1}$, whereas, in the Conformer-IV spectrum, this peak has undetectable intensity. The OH bend peak was hardly detected (low-intensity peak) in the NMA spectra of all the conformers, however, this peak is one of the high-intensity peaks in the experimental spectrum.

**FT-DAC of Serine at 298.15 K:** The FT-DAC method produced similar IR spectra for all the simulations started from four conformers of serine at 298.15 K (Table 4). Compared to the experimental IR spectrum, the FT-DAC method accurately predicted both frequencies and their relative intensities. Since the IRMPD spectrum is recorded at room temperature, it allows for a higher degree of conformational space, results in broader peaks especially the carboxyl symmetric stretch mode.[71] The carboxyl symmetric stretch peak is broadened at 1240–1400 cm$^{-1}$ and carboxyl asymmetric bend and $NH_2$ scissoring bend are

fused with each other due to the peak broadening, which is nicely reproduced by the FT-DAC spectra. Apart from the anharmonic component, this improved performance of the FT-DAC method in the flexible systems can be rationalised by conformational sampling. In the MD simulations, Conformer-I/II are relatively stable in their initial conformations in terms of hydrogen bonding network, however, occasionally sampled Conformer-IV's conformational space as well (Figure S5 in SI). However, for the MD simulations started with Conformer-III, the $C_\alpha$-$C_\beta$ bond is rotated to $72.8^o$, resulting in changing its hydrogen bonding network to OH$\cdots$OCO$\cdots$HNH, similar to Conformer-I/II (Figure S5 in SI). So, the final FT-DAC spectrum of Conformation-III showed in Figure 4 is that from the conformations after equilibrium simulations. Similarly, simulations started with Conformer-IV also moved to the regions of Conformer-I/II (OH$\cdots$OCO$\cdots$HNH) and occasionally returned back to that of Conformer-IV. Overall, it is clear that at room temperature OH$\cdots$OCO$\cdots$HNH hydrogen bonding networked conformations are more stable than that of the OH$\cdots$NH$\cdots$OCO and OH$\cdots$OCO networked conformations. Thus the resulting four FT-DAC spectra at 298.15 K are identical with similar peak positions and widths (Figure 3 and Table 4).

The major peaks in the rest of the IR region are at $\sim$800 cm$^{-1}$, $\sim$975 cm$^{-1}$, $\sim$1050 cm$^{-1}$ and $\sim$1200 cm$^{-1}$. Based on the harmonic vibrational modes, these peaks are attributed to the following vibrational modes *i.e.,* $\sim$800 cm$^{-1}$ is contributed by $C_\alpha$ and $C_{carboxyl}$ stretch coupled with O-H bend, $\sim$975 cm$^{-1}$ is contributed by C-O stretch coupled with $NH_2$ wagging, $\sim$1050 cm$^{-1}$ is contributed by $C_\alpha$-$C_\beta$ coupled with $NH_2$ bend vibrational mode and finally, $\sim$1200 cm$^{-1}$ is contributed by $NH_2$ bend coupled with $CH_2$ wagging and O-H bend modes. The FT-DAC method reproduced the $\sim$800 cm$^{-1}$ and $\sim$1200 cm$^{-1}$ accurately, including their relative intensities. Even though there are spectral signatures in the region of 850-1050 cm$^{-1}$, the peaks near $\sim$975 cm$^{-1}$ and $\sim$1050 cm$^{-1}$ aren't completely resolved. Moreover, the relative intensities in this region are much lower compared to the experimental spectra.

**Temperature effects on the FT-DAC spectra of Serine:** We also examined the effects of temperature on the FT-DAC spectra (Figure 3 and 4). The vibrational peaks in the FT-DAC spectra are sharper at 100 K. Apart from minor changes in peak widths and its relative intensities, the FT-DAC spectra of Conformer-I/II/IV are similar at both the temperatures. In contrast, Conformer-III showed major difference from its FT-DAC spectrum

at 298.15 K. The FT-DAC spectrum of Conformer-III at 100 K is mainly dominated by the symmetric and asymmetric carboxyl stretch peaks, and the intensities of other peaks are low. The carboxyl asymmetric stretch peak is red shifted up to 35 $cm^{-1}$, which matches its NMA spectrum. Unlike the room temperature counterparts, Conformer-III maintains its hydrogen bonding network (Figure S6 in SI). This structural feature explains the resemblance of its FT-DAC and NMA spectra at 100 K. Based on these observations, it is clearly showed that at a lower temperature due to the restricted conformational space, the FT-DAC spectrum of serine moved close to its harmonic counterpart.

In summary, supported by analyses of the structural transitions along the simulations (Figures S5 and S6 in SI), we concluded that at 298.15 K all the simulations sampled similar conformational spaces and thus rather similar characteristics were observed in the FT-DAC spectra. In contrast at 100 K, simulations starting from Conformer-I, II and IV were similar while those from Conformer-III maintained its hydrogen bonding network thus demonstrated different spectral properties.

### Proton-bound dimethyl ether dimer

The nature of the shared proton in the acid-base chemistry and the proton transfer in various enzymatic reactions, captured by the classical Grotthuss mechanism,[72] is very important for both chemistry[73,74] and biology.[72,75–78] These proton-bound systems provide a useful model to understand the proton sharing mechanism as they are relatively stable and experimentally it is also possible to prepare them in macroscopic scale with controlled solvation.[79,80] In these systems, the shared proton is stabilised by the hydrogen bonding network with the two Lewis bases. There are several experimental and theoretical studies that have been conducted using the water cluster to show the shared proton exists in an Eigen form or Zundel form, which is still an ongoing debate in the literature.[14,81–84]

The experimental spectra of proton-bound dimethyl ether dimers have been measured with both the infrared multiple photon dissociation spectroscopy (IRMPD)[80] and the argon-tagged single photon dissociation spectroscopy (Ar-TSPD)[79]. The IRMPD method records the spectrum at room temperature whereas the Ar-TSPD measures at lower temperatures. So, the spectral signatures in the IRMPD spectrum are broader compared to the Ar-TSPD

spectrum.[85] Additionally, the vibrational peaks in the Ar-TSPD spectrum are devoid of thermal effects.[79] The main difference between the IRMPD and Ar-TSPD spectra of proton-bound dimethyl ether dimer is the doublet peak near 1000 $cm^{-1}$, which is clearly visible in the Ar-TSPD spectrum.[30]

**NMA of proton-bound dimethyl ether dimer:** Table 5 lists the calculated harmonic vibrational frequencies (B3LYP/cc-pVTZ and DFTB3) and their respective mode descriptions. The $\nu^{\mathrm{OPO_{A\text{-}S}\text{-}COC_{S\text{-}S}}}$ peak ($\sim$800 $cm^{-1}$) is underestimated by $\sim$100 $cm^{-1}$ in B3LYP/cc-pVTZ, whereas in DFTB3 is overestimated by $\sim$100 $cm^{-1}$. The $\nu^{\mathrm{COC_{A\text{-}S}\text{-}P\perp}}$ peak ($\sim$1000 $cm^{-1}$) is predicted fairly accurately by both methods. However the doublet nature of this peak (in the Ar-TSPD spectrum) is reproduced in neither of the NMA spectra. The major deviation in NMA/DFTB3 spectrum from the experimental counterpart is the $\nu^{\mathrm{P}\perp}$ peak ($\sim$1600 $cm^{-1}$). This peak is underestimated by $\sim$200 $cm^{-1}$ in NMA/DFTB3 spectrum, whereas B3LYP/cc-pVTZ (unscaled frequency) reproduces it well.

**FT-DAC of proton-bound dimethyl ether dimer:** The conformation changes in the simulations are monitored in the Figures S10 and S11 in SI. The structure is mainly oscillating between two conformational states. The maximum oscillatory distance of the shared proton between the two oxygen atoms is 0.4 Å and 0.6 Å at 68 K and 270 K, respectively (Figure S11 in SI). This movement of the shared proton contributes to the higher anharmonicity of the structure.

Tables 6 and 7 present the vibrational peaks and their widths in experimental (Ar-TSPD and IRMPD) and DFTB3/FT-DAC spectra at 68 K and 270 K. Similar to their experimental counterparts, the vibrational peaks in the FT-DAC spectrum at 68 K are quite narrow and those at 270 K are broader. Even though the left side tailing feature of $\nu^{\mathrm{OPO_{A\text{-}S}\text{-}COC_{S\text{-}S}}}$ peak in IRMPD spectrum and doublet nature of the $\nu^{\mathrm{COC_{A\text{-}S}\text{-}P\perp}}$ peak in Ar-TSPD spectrum are partially resolved, both the frequencies and their relative intensities didn't match accurately with their respective experimental counterparts. Especially the $\nu^{\mathrm{P}\perp}$ peak is underestimated by $\sim$160 $cm^{-1}$ and $\sim$185 $cm^{-1}$ in both the experimental spectra, respectively. This result shows the limitations of the DFTB3 method in describing such a system.

**Post-simulation corrections of the FT-DAC spectra:** Since the $\nu^{\mathrm{P}\perp}$ peak's position in DFTB3/FT-DAC spectra matches with those in DFTB3/NMA, this underestimation is

assumed to be caused by potential deficiencies in the DFTB3 model. Additional post-simulation corrections were carried out following the procedure by Biswas *et al*[86]. Namely, single point vibrational frequency calculations with B3LYP/cc-pVTZ were carried out on the sampled snapshots at every 50 fs with DFTB3 MD simulations (See SI Section III for more details). The final spectrum was then shown as the averaged intensities with a bin size of 5 $cm^{-1}$ (Figures 5 and 6). Indeed, the description of the key features at 1000 $cm^{-1}$, 1300 $cm^{-1}$, and 1600 $cm^{-1}$ were improved in the mixed B3LYP-DFTB3 harmonic vibrational spectra at both temperatures. However, due to the presence of the imaginary modes with the unoptimised structure, the vibrational modes around 800 $cm^{-1}$ might be contaminated. Nevertheless, such a post-simulation correction offers an effective way to remedy the potential intrinsic deficiencies of the FT-DAC simulations based on DFTB3 or other semi-empirical models. It offers an improved description of the spectra of interest with a higher level of theory at a relatively modest computational cost while maintaining the merits of the FT-DAC based calculation.

# CONCLUSIONS

In this work, we have systematically benchmarked the performance of DFTB3 for computational IR spectra based on NMA and FT-DAC for a comprehensive database of molecules. It has been shown that DFTB3 was able to capture the key features of the vibrational spectra for these diverse set of molecules. The less satisfactory performance with DFTB3 for the C=C, C=N and N–H vibrations was noted. By comparing the simulated spectra with the experimental data, an empirical scaling factor was developed for both methods. However, such empirical corrections do not significantly improve the spectral prediction. The advantage of FT-DAC was further demonstrated with small flexible molecules where anharmonicity and conformational sampling play an important role for computational vibrational spectroscopy. Its intrinsic limitations related to the DFTB3 model was also noted and a post-simulation correction was found to partially improve the accuracy of the spectral calculations. The high computational efficiency makes DFTB3 a potentially appealing method of choice for investigating the spectral properties of biomolecular systems, especially in the framework of

combined QM/MM simulations to establish the underlying microscopic structure for experimental IR fingerprints.

## ACKNOWLEDGMENTS

H.Y. is the recipient of an Australian Research Council Future Fellowship (Project number FT110100034). We wish to acknowledge the Australian Government for an Australian International Postgraduate Award scholarship for V.I. This research was in part supported under the Australian Research Council's Discovery Projects funding scheme (project number DP170101773). We wish to acknowledge that this research was undertaken with the assistance of resources provided at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme supported by the Australian Government.

# References


1. W. W. Coblentz, Physical Review **20**, 273 (1905).

2. N. C. Thomas, Journal of Chemical Education **68**, 631 (1991).

3. L. Chafetz, Journal of Pharmaceutical Sciences **56**, 312 (1967).

4. J. Coates, *Interpretation of Infrared Spectra, A Practical Approach* (John Wiley & Sons, Ltd, 2006), pp. 1–23.

5. A. Barth, Biochimica et Biophysica Acta (BBA) - Bioenergetics **1767**, 1073 (2007).

6. B. H. Stuart, *Biological Applications* (John Wiley & Sons, Ltd, 2005), pp. 137–165.

7. A. Maeda, Israel Journal of Chemistry **35**, 387 (1995).

8. K. Gerwert, Biological Chemistry **380**, 931 (1999).

9. R. Vogel and F. Siebert, Current Opinion in Chemical Biology **4**, 518 (2000).

10. P. J. Tonge, The Quarterly Review of Biology **72**, 198 (1997).

11. D. Steele, *Infrared Spectroscopy: Theory* (John Wiley & Sons, Ltd, 2006), pp. 44 – 70.

12. C. Kneip, P. Hildebrandt, K. Németh, F. Mark, and K. Schaffner, Chemical Physics Letters **311**, 479 (1999).

13. N. Mayorkas, S. Rudić, E. J. Cocinero, B. G. Davis, and J. P. Simons, Physical Chemistry Chemical Physics **13**, 18671 (2011).

14. H. Yu and Q. Cui, The Journal of Chemical Physics **127**, 234504 (2007).

15. H. A. Witek and K. Morokuma, Journal of Computational Chemistry **25**, 1858 (2004).

16. Z. A. Fekete, E. A. Hoffmann, T. Körtvélyesi, and B. Penke, Molecular Physics **105**, 2597 (2007).

17. N. Wiener, Acta Mathematica **55**, 117 (1930).

18. A. P. Scott and L. Radom, The Journal of Physical Chemistry **100**, 16502 (1996).

19. J. Neugebauer, M. Reiher, C. Kind, and B. A. Hess, Journal of Computational Chemistry **23**, 895 (2002).

20. M. Thomas, M. Brehm, R. Fligg, P. Vöhringer, and B. Kirchner, Physical Chemistry Chemical Physics **15**, 6608 (2013).

21. M. Schmitz and P. Tavan, in *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, edited by E. B. Starikov, J. P. Lewis, and S. Tanaka (Elsevier Science, Amsterdam, 2006), pp. 159 – 177.

22. A. Khintchine, Mathematische Annalen **109**, 604 (1934).

23. Q. Cui and I. Bahar, *Normal mode analysis: theory and applications to biological and chemical systems* (CRC press, 2005).

24. Q. Cui and M. Karplus, The Journal of Chemical Physics **112**, 1133 (2000).

25. M. Levitt, C. Sander, and P. S. Stern, International Journal of Quantum Chemistry **24**, 181 (1983).

26. A. W. V. Wynsberghe and Q. Cui, Structure **14**, 1647 (2006).

27. R. Kubo, Journal of the Physical Society of Japan **12**, 570 (1957).

28. R. G. Gordon, The Journal of Chemical Physics **43**, 1307 (1965).

29. G. Williams, Chemical Reviews **72**, 55 (1972).

30. X. Li, D. T. Moore, and S. S. Iyengar, The Journal of Chemical Physics **128**, 184308 (2008).

31. S. Luber, Journal of Chemical Theory and Computation **13**, 1254 (2017).

32. S. Luber, M. Iannuzzi, and J. Hutter, The Journal of Chemical Physics **141**, 094503 (2014).

33. M. Gaus, Q. Cui, and M. Elstner, Journal of Chemical Theory and Computation **7**, 931 (2011).

34. M. Gaus, A. Goez, and M. Elstner, Journal of Chemical Theory and Computation **9**, 338 (2013).

35. M. Gaus, X. Lu, M. Elstner, and Q. Cui, Journal of Chemical Theory and Computation **10**, 1518 (2014).

36. G. Seifert and J.-O. Joswig, Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 456 (2012).

37. M. Gaus, Q. Cui, and M. Elstner, Wiley Interdisciplinary Reviews: Computational Molecular Science **4**, 49 (2014).

38. Yang, H. Yu, D. York, Q. Cui, and M. Elstner, The Journal of Physical Chemistry A **111**, 10861 (2007).

39. M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Physical Review B **58**, 7260 (1998).

40. S. Kaminski, M. Gaus, P. Phatak, D. von Stetten, M. Elstner, and M. A. Mroginski, Journal of Chemical Theory and Computation **6**, 1240 (2010).

41. J. P. Merrick, D. Moran, and L. Radom, The Journal of Physical Chemistry A **111**, 11683 (2007).

42. P. Deglmann, F. Furche, and R. Ahlrichs, Chemical Physics Letters **362**, 511 (2002).

43. R. Ramirez, T. Lopez-Ciudad, P. Kumar P, and D. Marx, The Journal of Chemical Physics **121**, 3973—3983 (2004).

44. H. Wang, X. Sun, and W. H. Miller, The Journal of Chemical Physics **108**, 9726 (1998).

45. X. Sun, H. Wang, and W. H. Miller, The Journal of Chemical Physics **109**, 7064 (1998).

46. J. Liu and W. H. Miller, The Journal of Chemical Physics **127**, 114506 (2007).

47. J. Cao and G. A. Voth, The Journal of Chemical Physics **100**, 5106 (1994).

48. S. Jang and G. A. Voth, The Journal of Chemical Physics **111**, 2371 (1999).

49. I. R. Craig and D. E. Manolopoulos, The Journal of Chemical Physics **121**, 3368 (2004).

50. I. R. Craig and D. E. Manolopoulos, The Journal of Chemical Physics **122**, 084106 (2005).

51. I. R. Craig and D. E. Manolopoulos, The Journal of Chemical Physics **123**, 034102 (2005).

52. B. J. Braams and D. E. Manolopoulos, The Journal of Chemical Physics **125**, 124105 (2006).

53. R. D. Johnson, *NIST Computational Chemistry Comparison and Benchmark Database NIST Standard Reference Database Number 101 Release 18* (October 2016), URL `http://cccbdb.nist.gov`.

54. P. J. Linstrom and W. G. Mallard, *NIST Chemistry WebBook, NIST Standard Reference Database Number 69, National Institute of Standards and Technology, Gaithersburg MD, 20899,* (retrieved May 4, 2018), URL `https://webbook.nist.gov/`.

55. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., Journal of Computational Chemistry **30**, 1545 (2009).

56. *The DFTB Website* (2016), URL `http://www.dftb.org`.

57. R. B. Blackman and J. W. Tukey, Bell System Technical Journal **37**, 185 (1958).

58. P. Du, W. A. Kibbe, and S. M. Lin, Bioinformatics **22**, 2059 (2006).

59. M. Martinez, M.-P. Gaigeot, D. Borgis, and R. Vuilleumier, The Journal of Chemical Physics **125**, 144106 (2006).

60. M.-P. Gaigeot, M. Martinez, and R. Vuilleumier, Molecular Physics **105**, 2857 (2007).

61. Y. Nishimura, Y.-P. Lee, S. Irle, and H. A. Witek, The Journal of Chemical Physics **141**, 094303 (2014).

62. J. M. Bowman, X. Zhang, and A. Brown, The Journal of Chemical Physics **119**, 646 (2003).

63. M. Kaledin, A. Brown, A. L. Kaledin, and J. M. Bowman, The Journal of Chemical Physics **121**, 5646 (2004).

64. F. Thaunay, J.-P. Dognon, G. Ohanessian, and C. Clavaguera, Phys. Chem. Chem. Phys. **17**, 25968 (2015).

65. M. Buchner, B. M. Ladanyi, and R. M. Stratt, The Journal of Chemical Physics **97**, 8522 (1992).

66. M. Cho, G. R. Fleming, S. Saito, I. Ohmine, and R. M. Stratt, The Journal of Chemical Physics **100**, 6672 (1994).

67. M. Schmitz and P. Tavan, The Journal of Chemical Physics **121**, 12247 (2004).

68. M. Schmitz and P. Tavan, The Journal of Chemical Physics **121**, 12233 (2004).

69. R. A. Wheeler, H. Dong, and S. E. Boesch, ChemPhysChem **4**, 382 (2003).

70. C. M. Jones, M. Bernier, E. Carson, K. E. Colyer, R. Metz, A. Pawlow, E. D. Wischow, I. Webb, E. J. Andriole, and J. C. Poutsma, International Journal of Mass Spectrometry **267**, 54 (2007).

71. J. Oomens, J. D. Steill, and B. Redlich, Journal of the American Chemical Society **131**, 4310 (2009).

72. N. Agmon, Chemical Physics Letters **244**, 456 (1995).

73. M. Thamer, L. De Marco, K. Ramasesha, A. Mandal, and A. Tokmakoff, Science **350**, 78 (2015).

74. T. S. Zwier, Science **304**, 1119 (2004).

75. M. M. Teeter, Proceedings of the National Academy of Sciences **81**, 6014 (1984).

76. F. Garczarek, L. S. Brown, J. K. Lanyi, and K. Gerwert, Proceedings of the National Academy of Sciences of the United States of America **102**, 3633 (2005).

77. F. Garczarek and K. Gerwert, Nature **439**, 109 (2006).

78. P. Phatak, N. Ghosh, H. Yu, Q. Cui, and M. Elstner, Proceedings of the National Academy of Sciences **105**, 19672 (2008).

79. J. R. Roscioli, L. R. McCunn, and M. A. Johnson, Science **316**, 249 (2007).

80. D. T. Moore, J. Oomens, L. van der Meer, G. von Helden, G. Meijer, J. Valle, A. G. Marshall, and J. R. Eyler, ChemPhysChem **5**, 740 (2004).

81. S. Woutersen and H. J. Bakker, Phys. Rev. Lett. **96**, 138305 (2006).

82. D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, Nature **397**, 601 (1999).

83. G. Mathias and D. Marx, Proceedings of the National Academy of Sciences **104**, 6980 (2007).

84. J.-W. Shin, N. I. Hammer, E. G. Diken, M. A. Johnson, R. S. Walters, T. D. Jaeger, M. A. Duncan, R. A. Christie, and K. D. Jordan, Science **304**, 1137 (2004).

85. X. Li, J. Oomens, J. R. Eyler, D. T. Moore, and S. S. Iyengar, The Journal of Chemical Physics **132**, 244301 (2010).

86. R. Biswas, W. Carpenter, J. A. Fournier, G. A. Voth, and A. Tokmakoff, The Journal of Chemical Physics **146**, 154507 (2017).

Figure 1: Convergence in the FT-DAC spectra of methanol. (A) FT-DAC spectra of methanol with various number of multiple independent trajectories. (B) Zoomed FT-DAC spectra of methanol in the 3400-3900 $cm^{-1}$ region.

Figure 2: Comparison between computed and experimental vibrational frequencies (307 modes in 40 molecules). Unscaled vibrational frequencies based on NMA (a: B3LYP/cc-pVTZ, b: DFTB3 and c: DFTB3-freq) and based on FT-DAC (d: DFTB3). Scaled vibrational frequencies based on NMA (e: B3LYP/cc-pVTZ, f: DFTB3 and g: DFTB3-freq) and based on FT-DAC (h: DFTB3).The diagonal line represents the ideal distribution of vibrational modes for the computational method that is able to reproduce experimental frequencies perfectly.

Figure 3: Scaled theoretical IR spectra for four different conformations of serine with the NMA and FT-DAC (298.15 K and 100.0K) methods. Conformer-I and II have the intramolecular hydrogen bonding network of $OH\cdots OCO\cdots NH_2$. Conformer-III has the intramolecular hydrogen bonding network of $OH\cdots NH_2\cdots OCO$. Conformer-IV has only one hydrogen bond between $OH\cdots OCO$. The blue line is the NMA spectra. The red line is the mean FT-DAC spectra from multiple independent simulations and the standard deviation is represented as a red shaded region. The IRMPD spectrum (black line) of serine was obtained by digitalising the experimental spectrum[71] with Enguage Digitizer.

Figure 4: Scaled theoretical IR spectra for conformer-III of serine with the NMA and FT-DAC (at 10K, 100 K, 200K, 298.15K, 325K, 350K and 400K) methods.

Figure 5: Comparison of the Ar-TSPD spectrum and scaled theoretical IR spectra of proton-bound dimethyl ether dimer with NMA (B3LYP/cc-pVTZ and DFTB3), FT-DAC (DFTB3 at 68.0 K) and mixed B3LYP-DFTB3 harmonic vibrational analysis. The Ar-TSPD spectrum was obtained by digitalising the experimental spectrum[79] with Enguage Digitizer.

Figure 6: Comparison of the IRMPD spectrum and scaled theoretical IR spectra of proton-bound dimethyl ether dimer with NMA (B3LYP/cc-pVTZ and DFTB3), FT-DAC (DFTB3 at 270.0 K) and mixed B3LYP-DFTB3 harmonic vibrational analysis. The IRMPD spectrum was obtained by digitalising the experimental spectrum[80] with Enguage Digitizer.

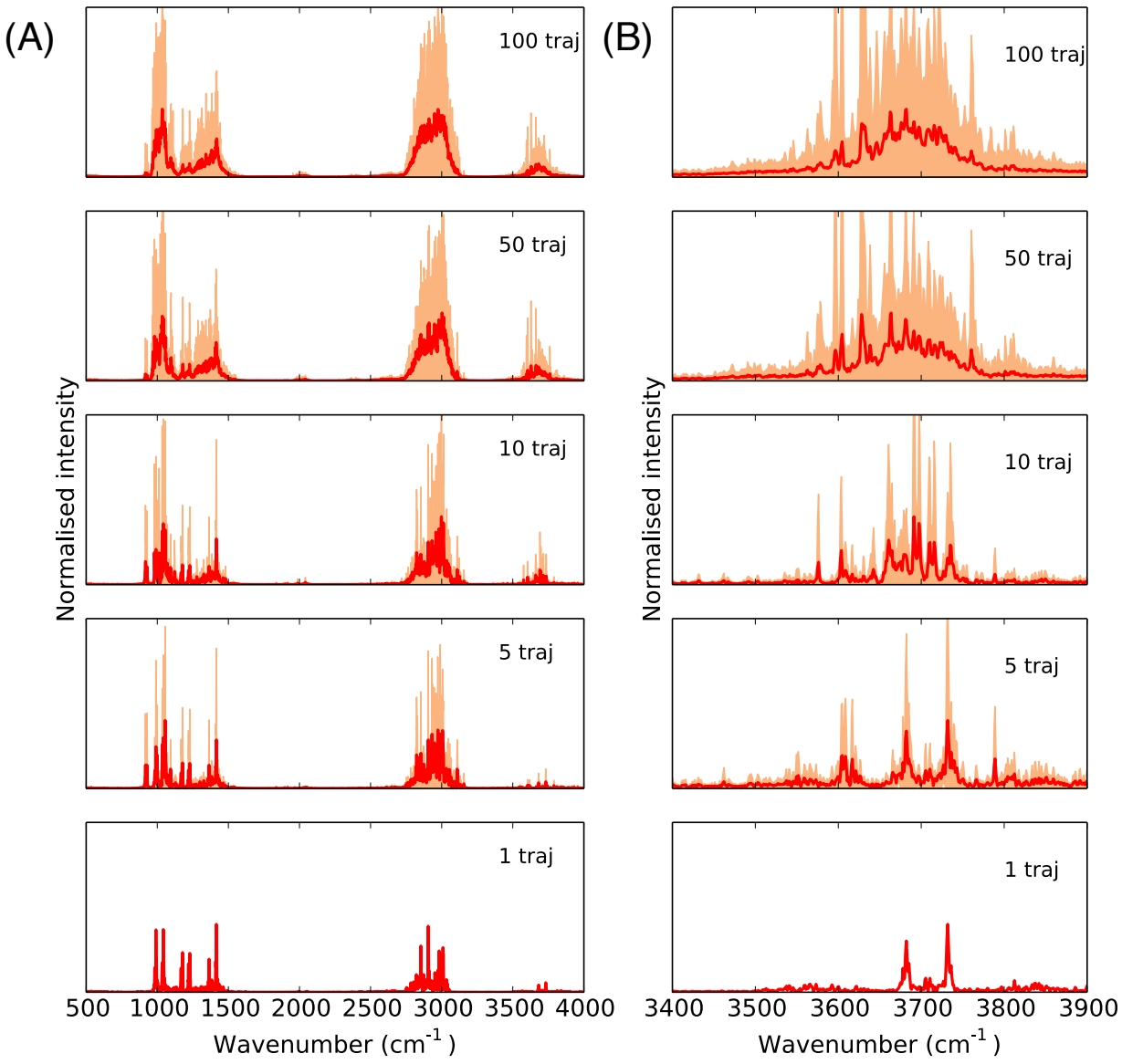

(A)
(B)
100 traj
50 traj
10 traj
5 traj
1 traj
Normalised intensity
500 1000 1500 2000 2500 3000 3500 4000
Wavenumber (cm-1 )
3400 3500 3600 3700 3800 3900
Wavenumber (cm-1 )

Figure 1

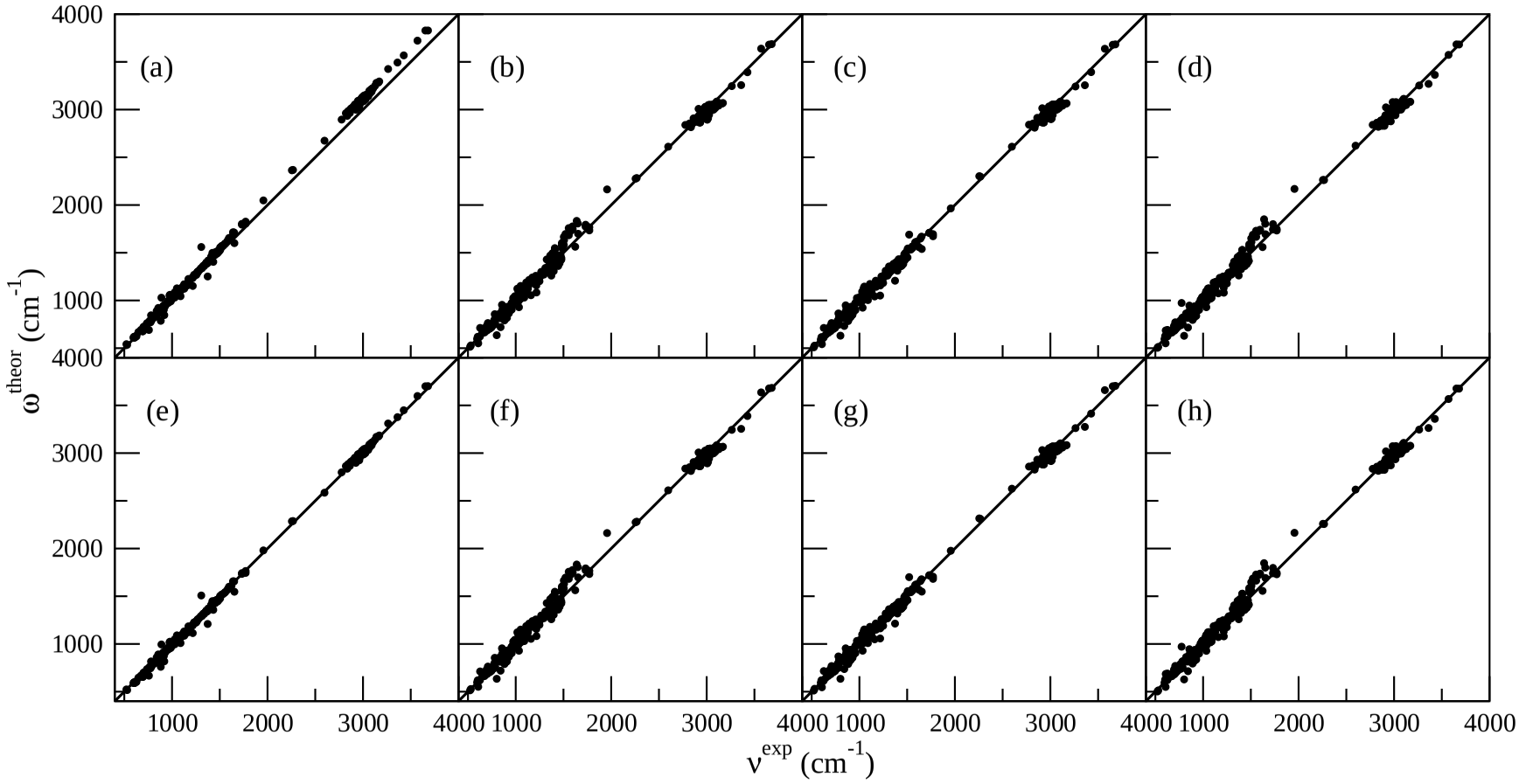



Figure 2

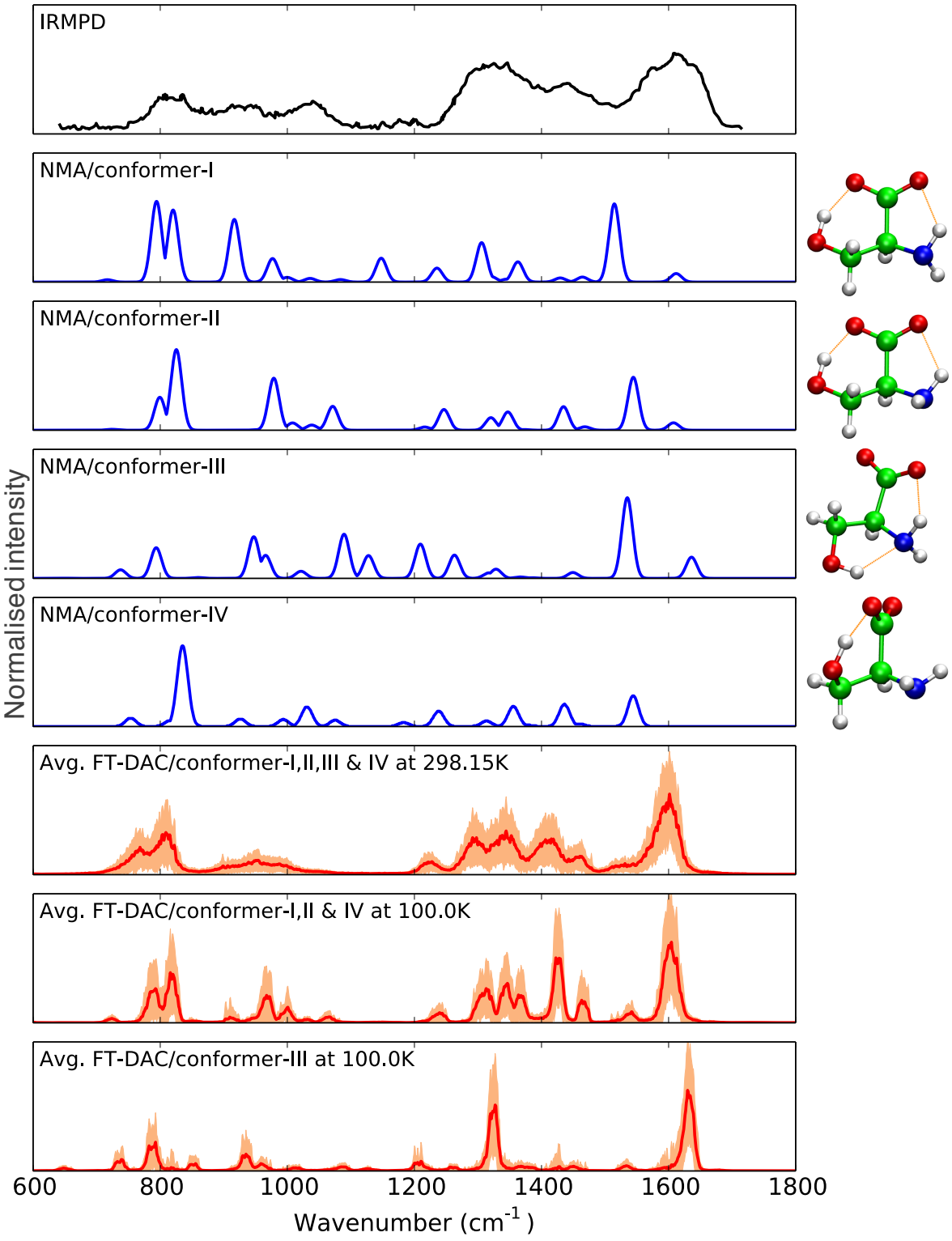

IRMPD
NMA/conformer-I
NMA/conformer-II
NMA/conformer-III
NMA/conformer-IV
Avg. FT-DAC/conformer-I,II,III & IV at 298.15K
Avg. FT-DAC/conformer-I,II & IV at 100.0K
Avg. FT-DAC/conformer-III at 100.0K
Normalised intensity
600
800
1000
1200
1400
1600
1800
Wavenumber (cm-1 )

Figure 3

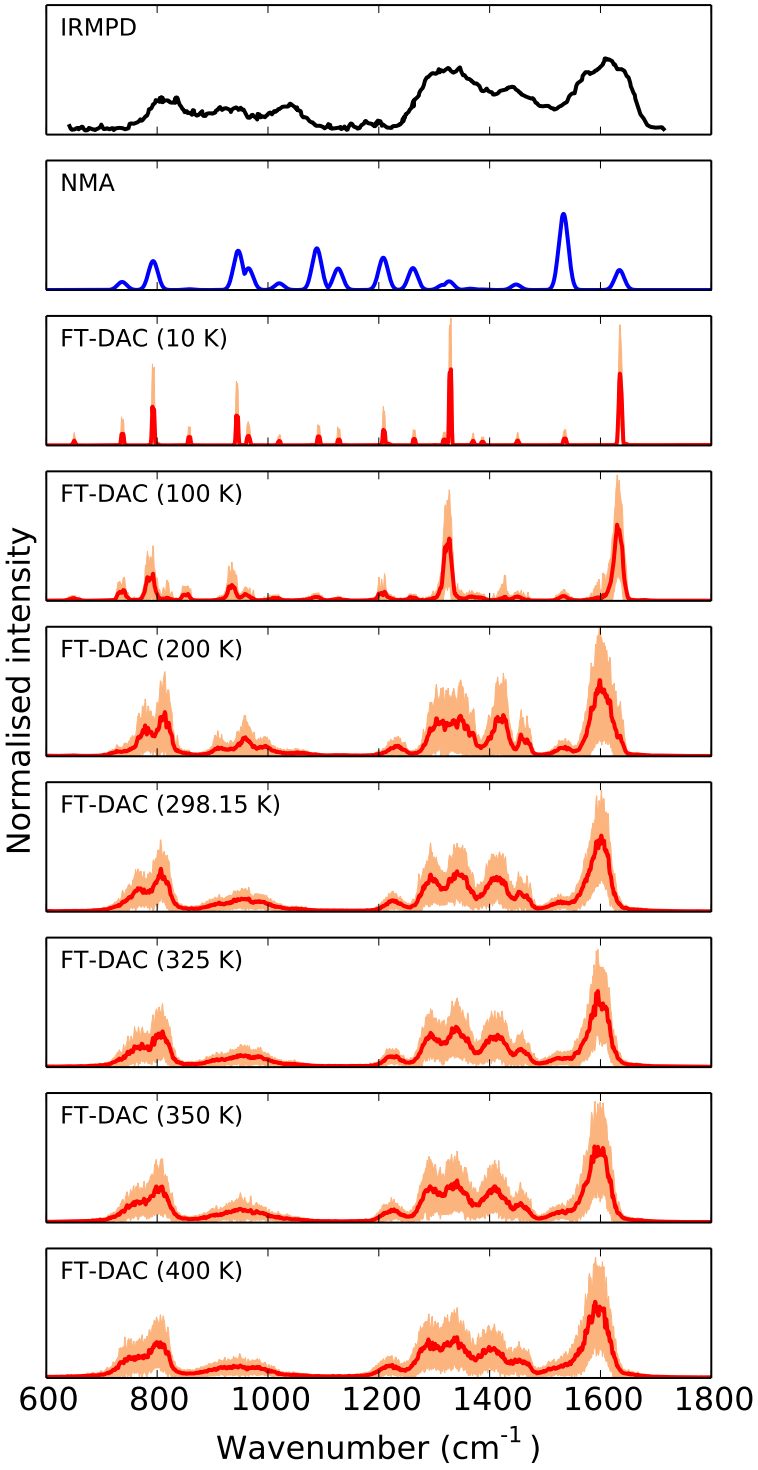

IRMPD
NMA
FT-DAC (10 K)
FT-DAC (100 K)
FT-DAC (200 K)
FT-DAC (298.15 K)
FT-DAC (325 K)
FT-DAC (350 K)
FT-DAC (400 K)
Normalised intensity
600
800
1000
1200
1400
1600
1800
Wavenumber (cm-1 )

Figure 4

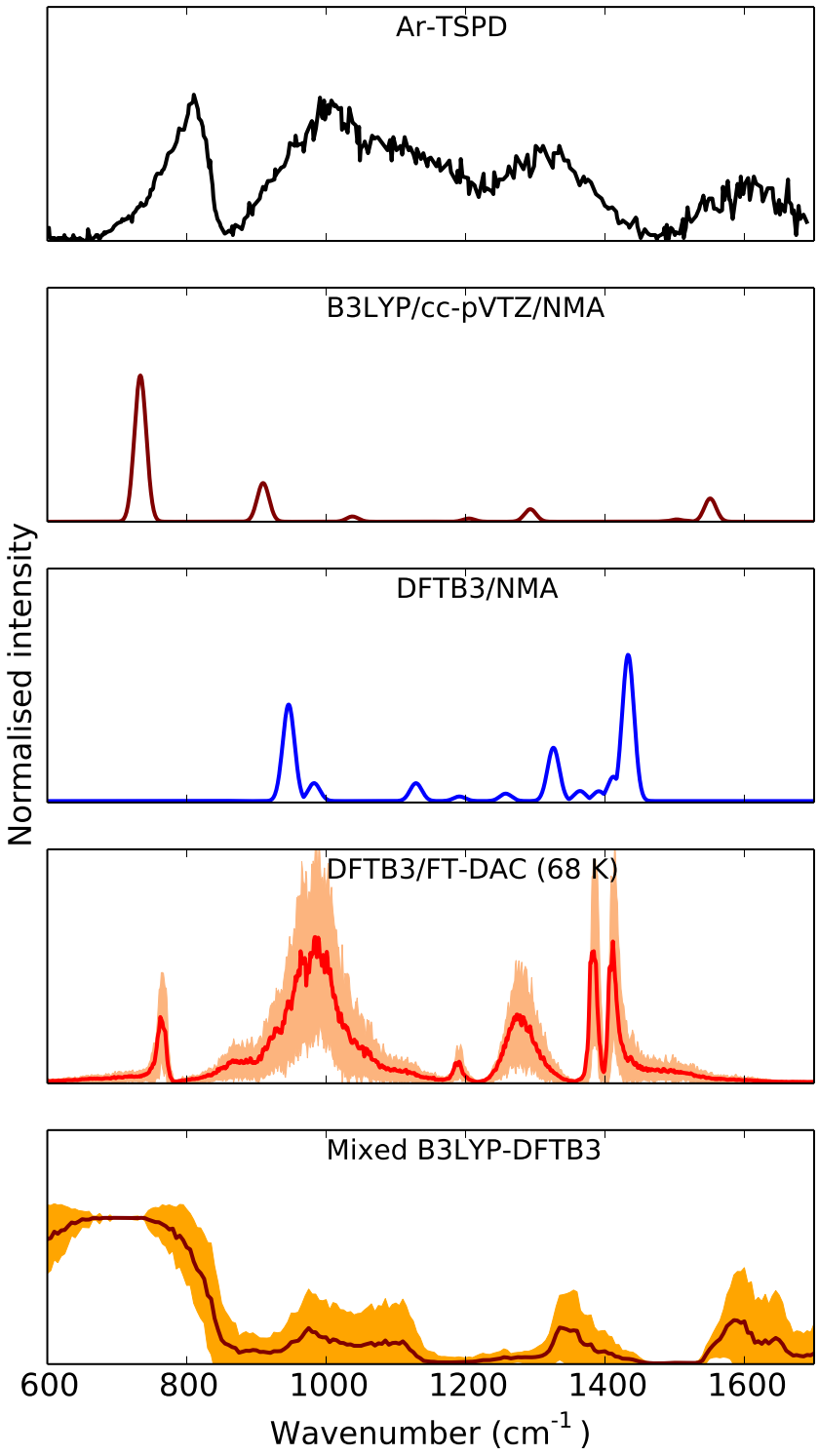

Ar-TSPD
B3LYP/cc-pVTZ/NMA
DFTB3/NMA
DFTB3/FT-DAC (68 K)
Mixed B3LYP-DFTB3
Normalised intensity
600
800
1000
1200
1400
1600
Wavenumber (cm-1)

Figure 5

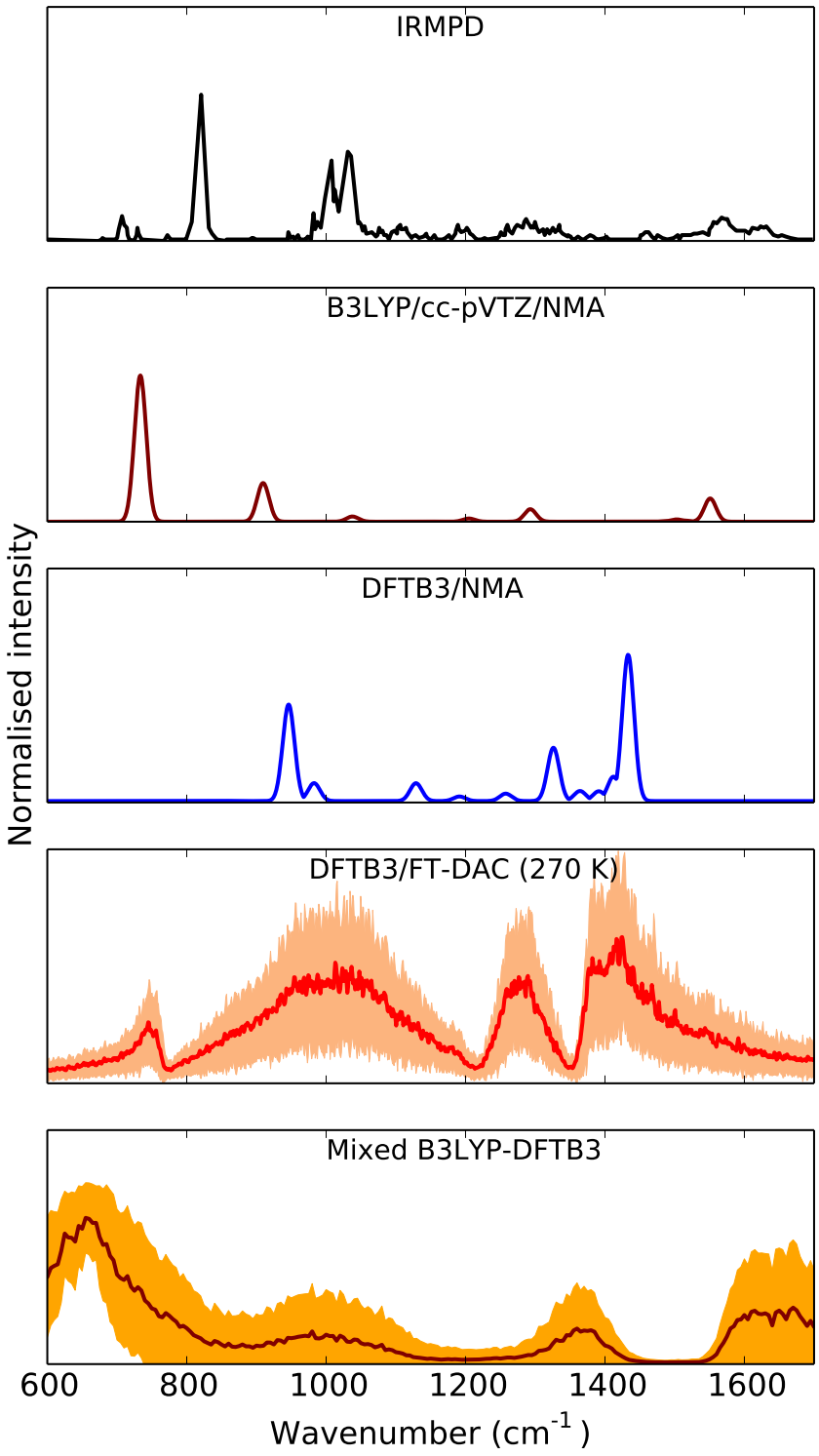

IRMPD
B3LYP/cc-pVTZ/NMA
DFTB3/NMA
DFTB3/FT-DAC (270 K)
Mixed B3LYP-DFTB3
Normalised intensity
600
800
1000
1200
1400
1600
Wavenumber (cm-1)

Figure 6

| Deviation | NMA | | | FT-DAC |
|---|---|---|---|---|
| | B3LYP/cc-pVTZ | DFTB3 | DFTB3-freq | DFTB3 |
| $\sigma_{MAX}$ | 254 | 206 | 170 | 212 |
| $\sigma_{MAD}$ | 56 | 42 | 34 | 41 |
| $\sigma_{STD}$ | 72 | 58 | 46 | 56 |
| $\sigma_{MAPE}$ | 3.0 | 3.0 | 2.5 | 3.0 |

Table 1: Numerical accuracy of the unscaled NMA and FT-DAC vibrational frequency calculations for the database of 40 molecules (listed in the spreadsheet in SI). $\sigma_{MAX}$ is the maximal absolute deviations defined by Equation 11. $\sigma_{MAD}$ is the average absolute error defined by Equation 12. $\sigma_{STD}$ is the standard deviations defined by Equation 13. $\sigma_{MAPE}$ is the weighted absolute error defined by Equation 14.

| | Method | Scaling factor ($\lambda$) | $rms_{tot}$ ($cm^{-1}$) |
|---|---|---|---|
| | B3LYP/cc-pVTZ | 0.9666 | 29 |
| NMA | DFTB3 | 0.9993 | 58 |
| | DFTB3-freq | 1.0059 | 45 |
| FT-DAC | DFTB3 | 0.9982 | 56 |

Table 2: Scaling factor ($\lambda$) and $rms_{tot}$ for NMA and FT-DAC methods.

| Structures | Dihedral angle ($^o$) | | | Bond angle ($^o$) | | | H-bond distance (Å) | | | $\Delta E^a$ | $\Delta E^b$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | OCCC | HOCC | HCNH | O–H–O | N–H–O | O–H–N | OCO–HO | NH–OCO | OH–NH | | |
| Conformer-I | 53.5 | −33.8 | 111.2 | 153.8 | 120.2 | 80.6 | 1.6 | 2.0 | 3.8 | 0.3 | 0.0 |
| Conformer-II | 52.7 | −32.3 | 101.5 | 154.3 | 118.3 | 80.3 | 1.6 | 2.0 | 3.8 | 0.0 | 0.2 |
| Conformer-III | −163.8 | 30.3 | 131.9 | 72.7 | 120.6 | 123.4 | 4.6 | 2.0 | 2.1 | 7.0 | 3.3 |
| Conformer-IV | −53.4 | 31.8 | 80.0 | 154.9 | 93.5 | 82.6 | 1.6 | 3.2 | 2.9 | 2.4 | 2.2 |

Table 3: Structural and energetic properties of the four lowest energy serine conformations. [a] electronic energy (in kcal/mol) relative to Conformer-II calculated with DFTB3 and [b] electronic energy relative to Conformer-I (in kcal/mol) calculated with B3LYP/6-31++G** on the DFTB3 optimised geometries.

| Structures | Vibrational mode | Experimental | | NMA | FT-DAC | |
|---|---|---|---|---|---|---|
| | | Peak | Width | Peak | Peak | Width |
| Conformer-I | OCO symmetric stretch | 1315 | 1220-1400 | 1321 | 1300 | 1250-1384 |
| | OH bend | 1420 | 1400-1450 | 1429 | 1423 | 1384-1490 |
| | HNH scissoring | 1575 | 1500-1670 | 1514 | 1514 | 1490-1660 |
| | OCO asymmetric stretch | 1600 | 1500-1670 | 1611 | 1605 | 1490-1660 |
| Conformer-II | OCO symmetric stretch | 1315 | 1220-1400 | 1320 | 1300 | 1250-1384 |
| | OH bend | 1420 | 1400-1450 | 1434 | 1417 | 1384-1492 |
| | HNH scissoring | 1575 | 1500-1670 | 1544 | 1541 | 1492-1660 |
| | OCO asymmetric stretch | 1600 | 1500-1670 | 1607 | 1603 | 1492-1660 |
| Conformer-III | OCO symmetric stretch | 1315 | 1220-1400 | 1317 | 1300 | 1260-1382 |
| | OH bend | 1420 | 1400-1450 | 1385 | 1418 | 1382-1487 |
| | HNH scissoring | 1575 | 1500-1670 | 1534 | 1536 | 1587-1655 |
| | OCO asymmetric stretch | 1600 | 1500-1670 | 1635 | 1608 | 1587-1655 |
| Conformer-IV | OCO symmetric stretch | 1315 | 1220-1400 | 1313 | 1300 | 1254-1380 |
| | OH bend | 1420 | 1400-1450 | 1379 | 1411 | 1380-1487 |
| | HNH scissoring | 1575 | 1500-1670 | 1544 | 1498 | 1587-1655 |
| | OCO asymmetric stretch | 1600 | 1500-1670 | 1610 | 1607 | 1498-1654 |

Table 4: Vibrational modes in the Exp., NMA and FT-DAC (298.15 K) spectra for simulations starting from four conformations of serine.

| Mode | Vibrational frequencies | | Vibrational modes | Symbols |
|---|---|---|---|---|
| | B3LYP/cc-pVTZ[a] | DFTB3[b] | | |
| 1 | 709 | 965 | O-P-O asymmetric stretch coupled with (simultaneous) C-O-C backbone symmetric stretches | $\nu^{\mathrm{OPO_{A\text{-}S}\text{-}COC_{S\text{-}S}}}$ |
| 2 | 896 | 771 | O-P-O asymmetric stretch coupled with (sequential) C-O-C backbone symmetric stretches | $\nu^{\mathrm{OPO_{A\text{-}S}\text{-}COC_{S\text{-}S}}}$ |
| 3 | 1004 | 977 | C-O-C asymmetric stretch coupled to proton motion perpendicular to O-O axis | $\nu^{\mathrm{COC_{A\text{-}S}\text{-}P\perp}}$ |
| 4 | 1004 | 986 | C-O-C asymmetric stretch coupled to proton motion perpendicular to O-O axis (orthogonal to previous one) | $\nu^{\mathrm{COC_{A\text{-}S}\text{-}P\perp}}$ |
| 5 | 1163 | 1189 | Methyl wagging coupled with proton motion in the O-O axis | $\nu^{\mathrm{wag\text{-}P\perp}}$ |
| 6 | 1249 | 1263 | Methyl wagging coupled with proton motion in the O-O axis (orthogonal to previous one) | $\nu^{\mathrm{wag\text{-}P\perp}}$ |
| 7 | 1488 | 1435 | Proton motion perpendicular to the O-O axis | $\nu^{\mathrm{P\perp}}$ |
| 8 | 1498 | 1413 | Proton motion perpendicular to the O-O axis | $\nu^{\mathrm{P\perp}}$ |

Table 5: The calculated vibrational modes of proton-bound dimethyl ether dimer system with B3LYP/cc-pVTZ and DFTB3. [a] scaled by 0.9666 and [b] scaled by 0.9993.

| | Ar-TSPD | | FT-DAC at 68.0 K | |
|---|---|---|---|---|
| | Peak | Width | Peak | Width |
| 1 | 810 | 800-820 | 769 | 736-786 |
| 2 | 1000 | 990-1010 | 968 | 900-976 |
| 3 | 1038 | 1010-1060 | 993 | 976-1084 |
| 4 | 1300 | 1220-1370 | 1280 | 1219-1360 |
| 5 | 1580 | 1580-1650 | 1420 | 1360-1462 |

Table 6: The experimental and calculated vibrational peaks and their distribution of proton-bound dimethyl ether dimer in Ar-TSPD and DFTB3/FT-DAC (68.0 K).

| | IRMPD | | FT-DAC at 270.0 K | |
|---|---|---|---|---|
| | Peak | Width | Peak | Width |
| 1 | 810 | 680-820 | 749 | 700-786 |
| 2 | 1000 | 850-1210 | 1019 | 780-1220 |
| 3 | 1300 | 1210-1450 | 1297 | 1220-1358 |
| 4 | 1610 | 1500-1700 | 1432 | 1358-1700 |

Table 7: The experimental and calculated vibrational peaks and their distribution of proton-bound dimethyl ether dimer in IRMPD and DFTB3/FT-DAC (270.0 K).