IR of Vanillin: A classic study with a twist

Michael Nicolaou[1], Hans Senn[1], Emma Gibson[1], Laia Vilà-Nadal[1*]*

[1]School of Chemistry, University of Glasgow, Glasgow, United Kingdom

KEYWORDS: DFT, computational chemistry, vibrational frequency

ABSTRACT

In this research, we delve into the vibrational spectroscopy of vanillin, a widely used aromatic and flavouring agent, through a comprehensive computational analysis. We employ a variety of common computational chemistry functionals and basis sets to calculate the infrared (IR) and Raman spectra of vanillin, aiming to shed light on its structural and spectroscopic properties. Our investigation entails benchmarking these theoretical results against one another to identify the most accurate computational approach. Furthermore, we juxtapose our theoretical findings with experimental IR and Raman spectra to evaluate the degree of agreement between theory and experiment. This comparative analysis provides insights into the reliability of the chosen computational methods in capturing the vibrational behaviour of vanillin, a crucial aspect for applications in the food and pharmaceutical industries. In future work we plan to expand this study to other compounds aiming to bridge the gap between theory and experiment, this study

contributes to a deeper understanding of vanillin's molecular behaviour, ultimately enhancing our knowledge of its sensory and chemical attributes.

INTRODUCTION

Vanillin, a ubiquitous compound in the cosmetics and food and beverage industries, widely recognised for its familiar smell and flavour, constitutes a key point of interest in various fields of research. Being the key flavour compound in vanilla, a well-known and largely used spice and commodity, the vanillin global market is of significant size and importance. Apart from its characteristic smell and flavour and therefore evident applications in foods and cosmetics, vanillin also possesses a plethora of other valuable aspects, such as its antioxidant and antimicrobial properties[1,2], making it a useful chemical for use as a food preservative[3], as well as in pharmaceutical drugs[4,5], its composition of three distinct functional groups, allowing for its flexible use in the synthesis of other desired compounds[6], its exhibition of non-linear optical (NLO) properties[7,8], making it a potential compound of interest for use in optics, and many more. Additionally, vanillin is one of the key products of lignin oxidation[9–11], an important rapidly growing field of research that aims to convert lignin, a largely abundant aromatic biopolymer, to fine chemicals, a primary candidate of which is vanillin, as well as syringaldehyde and *p*-hydroxybenzaldehyde.

While naturally derived from vanilla orchids, a family of cultivated plants, with the most widely grown and recognised variety being “Madagascar vanilla” or “Bourbon vanilla” grown mainly in Madagascar [12], its natural production makes up less than 1% of its global use[13], due to huge global demand and a laborious harvesting process. The overwhelming majority of used vanillin is therefore derived from synthesis through either plant or petrochemical-based reagents[13]. This makes the study of vanillin from a chemical perspective valuable, as research in

the characterisation, synthesis and understanding of the chemical nature of vanillin can help bolster its production from known chemicals, or facilitate the progress of novel methods that set vanillin as their end product, such as lignin oxidation.

The advent of advanced and accessible methods of computational molecular modelling in recent decades has provided chemists with an indispensable tool for calculating theoretical properties, enabling researchers to validate experimental findings through cross-referencing with sophisticated independent calculations. Quantum density-functional theory (DFT) methods have been extensively demonstrated to provide reliable theoretical information on studied systems, such as mechanistic pathways[14] of reactions, transition states[15] and electronic configurations, as well as predictions of empirically observable properties, such as molecular structure[16], nuclear magnetic resonance shifts[17] and bond vibrational frequencies[18]. The extended and constant refinement of computational methods has led to the creation of a dizzying variety of available parameters for use in DFT calculations, comprising of different exchange-correlation (XC) functionals, as well as functional rungs of theory, basis sets, including basis set enhancements such as polarisation and diffuse orbitals, solvation models and even the inclusion of dispersion and range-separation terms[19]. Due to the wide abundance of possible methodologies, as well as the approximate nature of DFT methods, comparing the performance of potential methodologies, as well as their individual components (such as the exchange-correlation (XC) functional or basis set) against empirical or benchmark data proves to be an essential measure in determining the performance and reliability of different methods[20].

In the present study, we test the application of various DFT methods for the calculation of common properties of vanillin, namely the molecular structure, vibrational frequencies (IR and Raman) and NMR chemical shifts. By comparing the calculated properties with experimental

data, we aim to identify the ideal DFT methodology in terms of XC-functional and basis set that results in the fastest and most accurate description of the desired property, while identifying the effect on accuracy and calculation time of different parameters, such as HF exchange energy contribution for hybrid functionals and polarisation and diffuse functions for basis sets. In our study of vibrational frequencies, we test the application of "universal" scaling factors as proposed by Truhlar *et al*[21], while also providing the calculated scaling factors for some of our methods that have not yet been provided, following the same procedure. Furthermore, we compare the accuracy of each method with the time required for the relevant calculation to determine the most efficient method, as well as compare the theoretical results to determine the level of agreement between different methods.

METHODOLOGY

1.1 Computational method

All Density Functional Theory (DFT) calculations were performed using the Gaussian-16 software[22] on the School of Chemistry High-Performance Compute (HPC) cluster. Molecular geometry optimisation, vibrational frequency and nuclear magnetic resonance (NMR) calculations of vanillin were performed at different exchange-correlation (XC) functional levels of theory across the GGA (Generalised-Gradient Approximation), mGGA (meta-GGA) and hybrid (HF exchange energy contribution) "rungs": (BP86[23,24], PBE[25,26], OPBE[25–28], M06-L[29], B3LYP[23,30,31], ωB97X-D[32,33], PBE0[25,26,34], M06[29] and M06-2X[29]), as well as using second-order Moller−Plesset perturbation theory (MP2)[35–39], using the 6-311++G(2d, 2p) triple-ζ split-valence polarised Pople Gaussian-type orbital (GTO) basis set for H, C, O[40,41] and S[42–44] with diffuse orbitals for all atoms. To determine the basis set convergence and study the effects of basis set parameters, such as polarisation and diffuse functions, the same calculations were also performed

at the M06-2X XC functional level of theory using different basis sets (3-21G, 6-31G, 6-31G(d,p), 6-311G, 6-311G(d,p), 6-311+G, cc-pVDZ, cc-pVTZ, cc-pVQZ, aug-cc-pVTZ, aug-cc-pVQZ[45], def2-TZVP, def2-QZVP[46]), as M06-2X has been extensively used in studies on similar molecules and suggested to be suitable for small-medium molecules[16] and main-group thermochemistry[20,29]. All calculations were performed using spin-restricted ("closed shell") orbitals and all drawn structures were pre-optimised using a Universal Force-Field (UFF)[47] Molecular Mechanics (MM) method prior to being optimised using the studied method. Vibrational frequency calculations were performed using the harmonic approximation model[48]. To account for anharmonicity in the fundamental modes of vibration, "universal" scaling factors were used, or where unavailable, calculated, optimised and employed based on the method proposed by Truhlar *et al*[21] (table S2). Nuclear magnetic shielding parameters were calculated for the C and H nuclei in vanillin using the GIAO (Gauge including atomic orbitals) method. The chemical shifts (δ) of vanillin were then determined by subtracting the magnetic shielding parameters from the shielding parameters calculated for a tetramethylsulfate (TMS) standard. The effects of solvation were approximated for chloroform in geometry optimisations (only for NMR) and NMR calculations using the Self-Consistent Reaction Field (SCRF)[49] based charged Density Solvation Model (SMD)[50].

To study the accuracy and precision of different methods, the mean absolute deviation (MAD) and standard deviation (STD) were employed:

$$MAD = \frac{1}{N}\sum_{i}^{N}|\Delta\chi| \qquad (1)$$

$$MSD = \frac{1}{N}\sum_{i}^{N} \Delta\chi \quad (2)$$

$$STD = \sqrt{\frac{1}{N-1}\sum_{i}^{N}(MSD - \Delta\chi)^2} \quad (3)$$

Where $\Delta\chi$ is found from subtracting the experimental value from the theoretical value for each studied parameters:

$$\Delta\chi = \chi_{Theor.} - \chi_{Expt.} \quad (4)$$

1.2 Experimental details

Vanillin (4-hydroxy-3-methoxybenzaldehyde) was obtained from Sigma-Aldrich (8.18718). Infrared absorbance spectrum was taken using a Nicolet iS5 FT-IR spectrometer with a resolution of 4 cm^{-1} over 256 scans, using Happ-Genzel apodization and Mertz phase correction. Raman absorbance spectrum was taken using a LabRAM HR system, equipped with a Synapse CCD detection system using a Ventus 532 laser system at 100 mW and 532nm, at power/filter 10%, grating of 600, hole size of 200 microns end exposure time of 5 seconds over 3 accumulations.

RESULTS AND DISCUSSION

2.1 Geometry optimisation

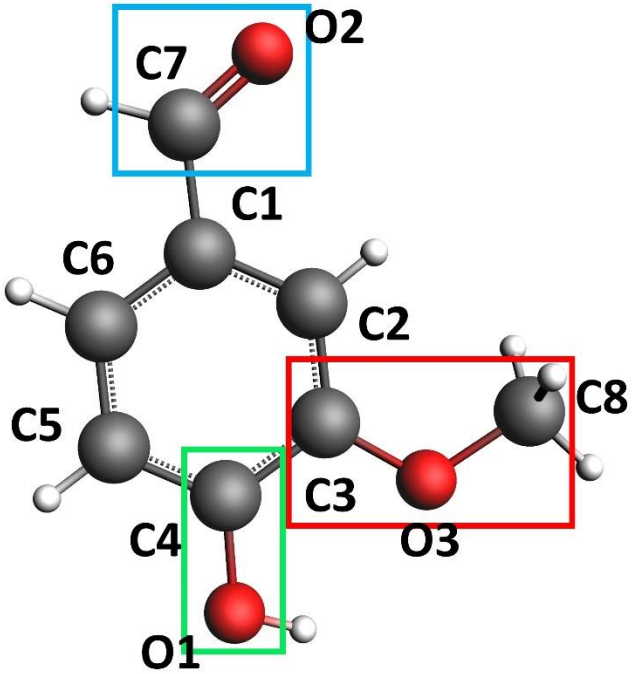

Figure 1. Molecular structure of a vanillin molecule in a single crystal structure, found through the CCDC[51] database (CID: 1306628, from Velavan *et al*[52]).

Comparison of bond lengths using 6-311++G(2d,2p) at different XC functional levels of theory. XC functionals of different rung (GGA, mGGA, hybrid) and hybrid functionals of different HF exchange energy contribution (25% to 54%) tested. Application of error metrics (MAD and STD) for each functional.

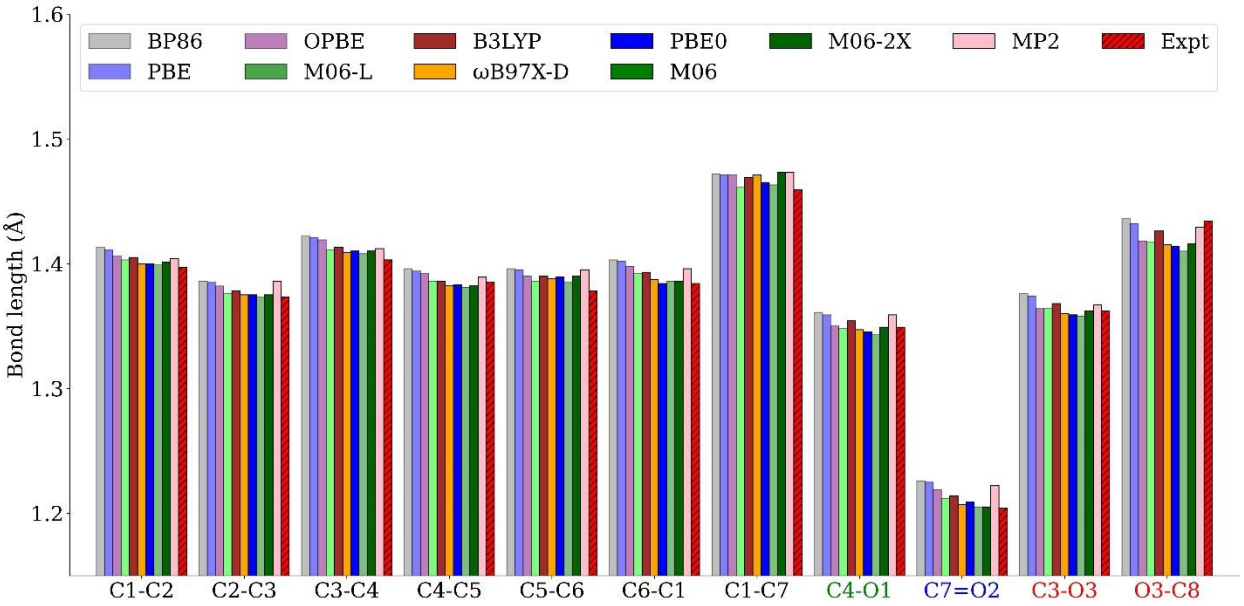

Figure 2. Bond length (Å) comparison between crystalline structure (from Velavan et al[52]) and optimised structures using 6-311++G(2d, 2p) at different XC functional levels of theory. The black bars correspond to the DFT calculated range of values for each bond length.

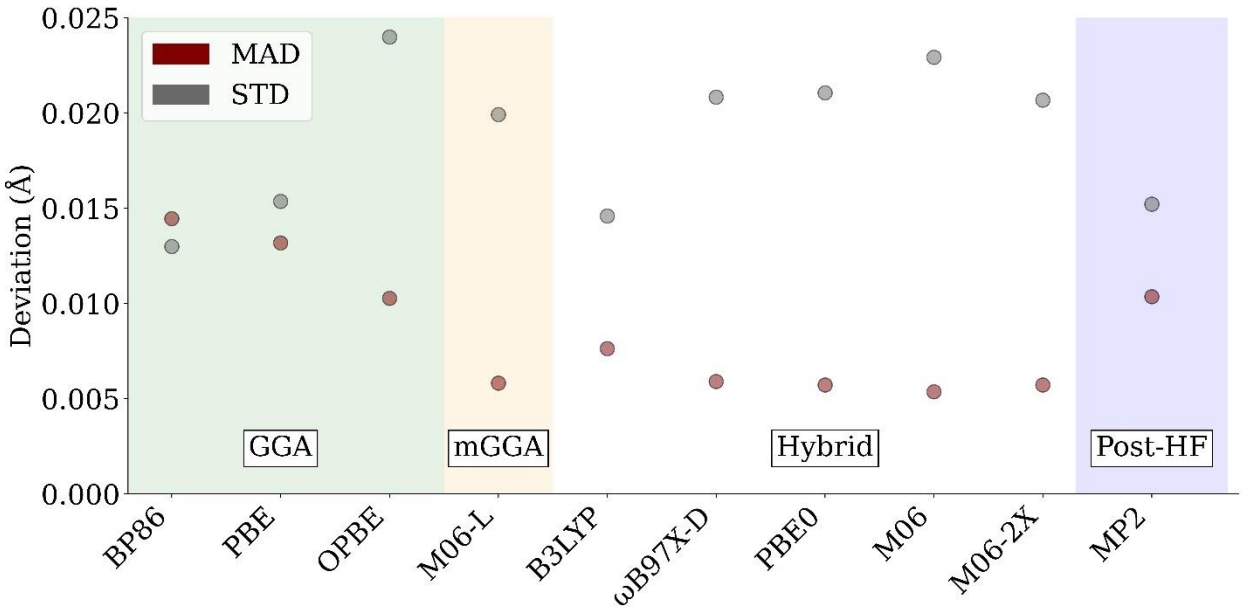

Figure 3. Deviation metrics (MAD and STD) (Å) of vanillin bond lengths at different XC functional levels of theory using 6-311++G(2d, 2p). Hybrid functionals displayed in order of increasing HF exchange-energy contribution.

Figure 2 shows each bond in vanillin as calculated by each method with different XC-functional levels of theory and using MP2, using 6-311++G(2d, 2p) as the basis set, as compared to the vanillin found in the single-crystal x-ray data file by Velavan *et al*[52], whereas figure 3 shows the deviation metrics (MAD and STD) for the difference between DFT-method and experimental across all bond lengths. From figure 3, it can be seen that going from GGA to mGGA or hybrid methods, a reduction in MAD can be observed, which is similar between all hybrid functionals (except B3LYP). MP2/6-311++G(2d, 2p) seems to show comparable results to GGA functionals. An increase in STD can be observed with decreasing MAD for almost all cases, suggesting the presence of systematic error in the system.

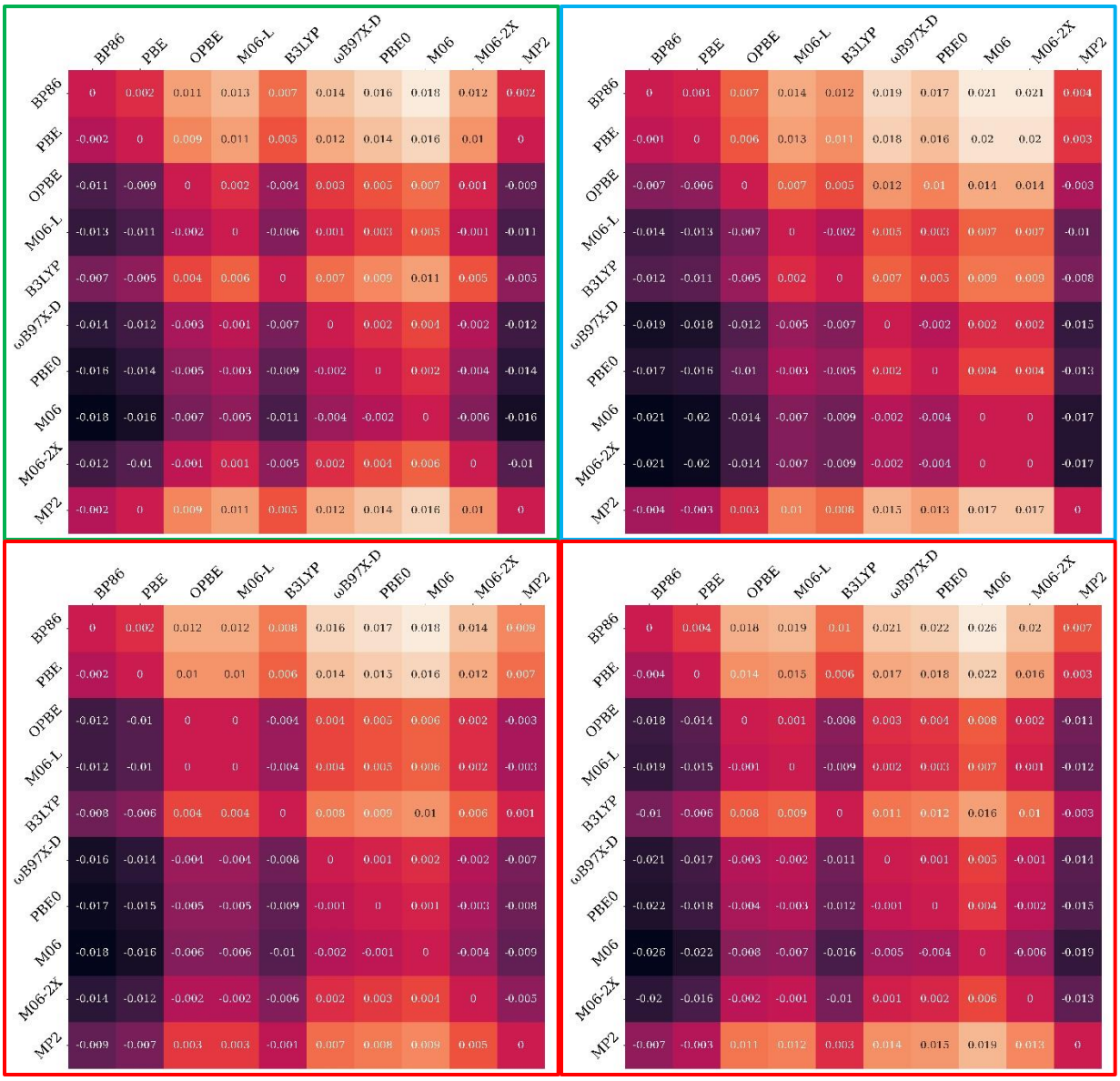

Figure 4. Theory against theory comparison of different methods for characteristic bonds. The heatmaps show the bond length (Å) difference (a. C4-O1, b. C7=O2, c. C3-O3 and d. O3-C8) between different method calculations. The difference is calculated by subtracting the column functional from the row functional length.

Comparison of bond lengths at the M06-2X level of theory using different basis sets. Basis sets of different type (Pople, Correlation-Consistent and Karlruhe), number of ζ orbitals, number of

GTO orbitals (for Pople) and diffused orbitals tested. Application of error metrics (MAD and STD) for each basis set.

Figure 4 shows the theory against theory comparison for all used XC-functional levels of theory and MP2, using 6-311++G(2d,2p) for the four characteristic bonds of vanillin. The heatmap suggests that deviations between bond lengths are within ±0.02 Å, suggesting that there is decent agreement between the different functional levels of theory and MP2.

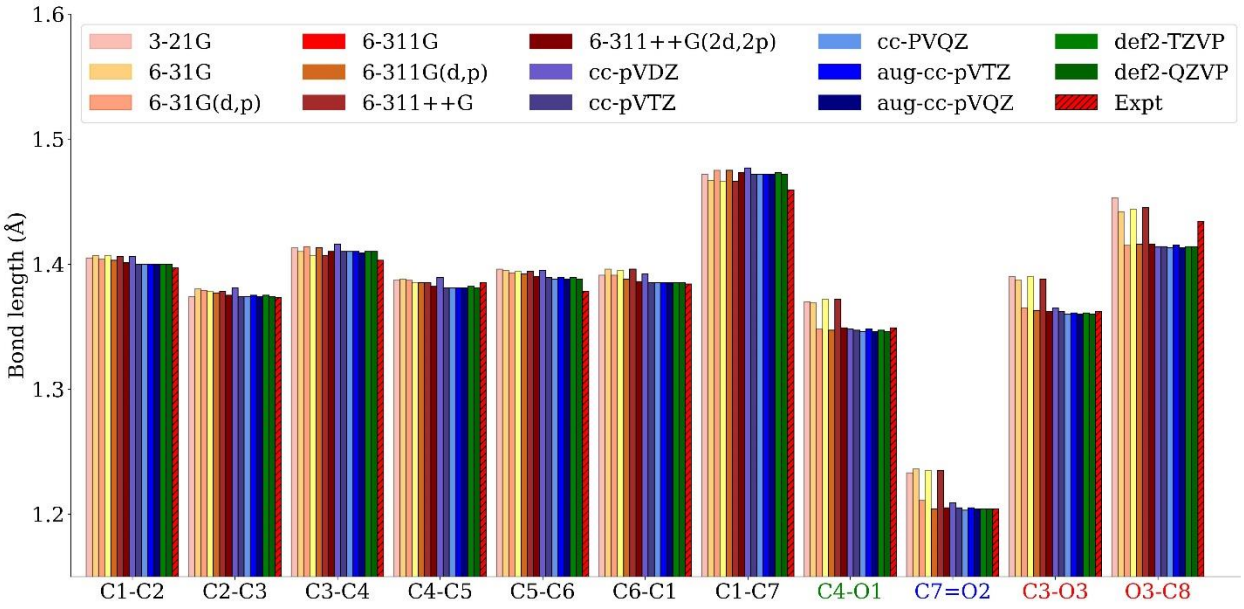

Figure 5. Bond length (Å) comparison between crystalline structure (from Velavan et al[52]) and optimised structure at the M06-2X functional level of theory using different basis sets. The black bars correspond to the DFT calculated range of values for each bond length.

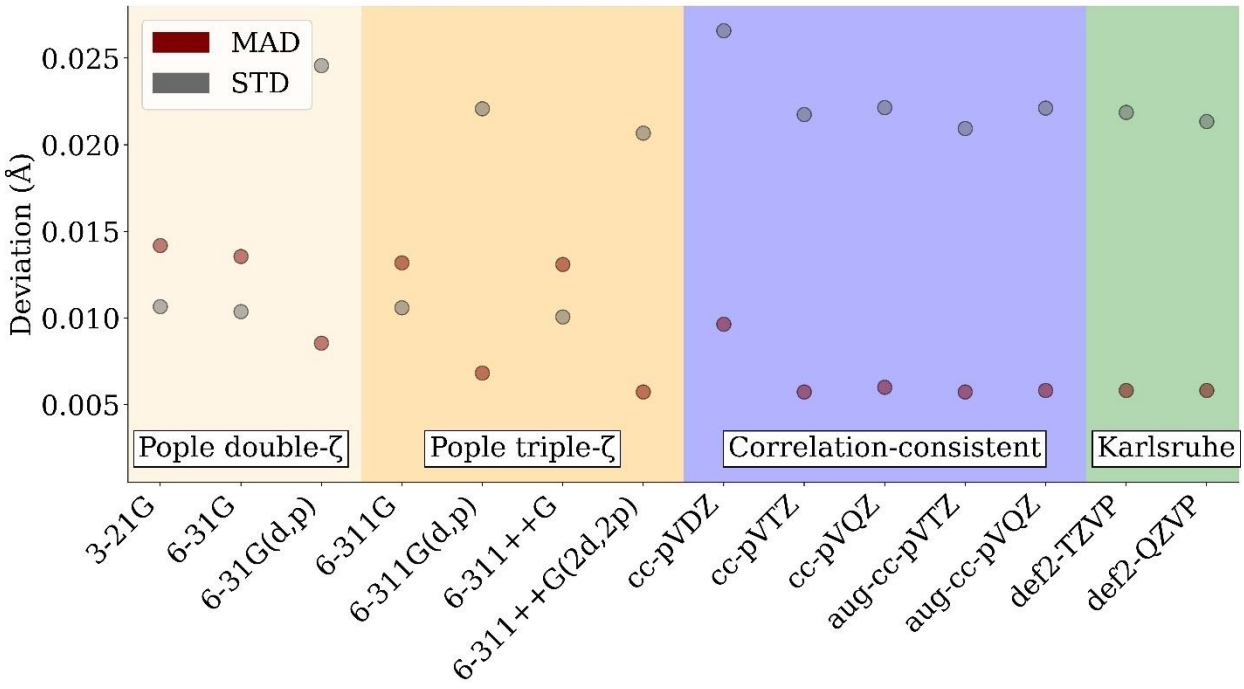

Figure 6. Deviation metrics (MAD and STD) (Å) of vanillin bond lengths at the M06-2X XC functional level of theory using different basis sets.

Figure 5 shows each bond in vanillin as calculated with M06-2X using the different basis sets, as compared to the vanillin found in the single-crystal x-ray data file by Velavan *et al*[52], whereas figure 6 shows the deviation metrics (MAD and STD) for the difference between DFT-method and experimental across all bond lengths. From figure 6, it can be seen that the basis set convergence for vanillin occurs at triple-ζ with polarisation, as basis sets with quadruple-ζ show similar performance. Polarisation shows a significant effect on the MAD, as all basis sets without polarisation (6-31G, 6-311G) show a higher MAD than the respective variants with polarisation (6-31G(d,p), 6-311G(d,p)). Inclusion of diffuse functions shows no effect on the MAD when comparing 6-311G with 6-311++G and aug-cc-pVTZ, aug-cc-pVQZ with cc-pVTZ and cc-pVQZ.

Comparison of MAD of all bond lengths (except hydrogen) with job CPU time for the geometry optimisation of vanillin starting from the crystallography .xyz file at different XC levels of theory using 6-311++G(2d,2p) and at the M06-2X functional level of theory using different basis sets.

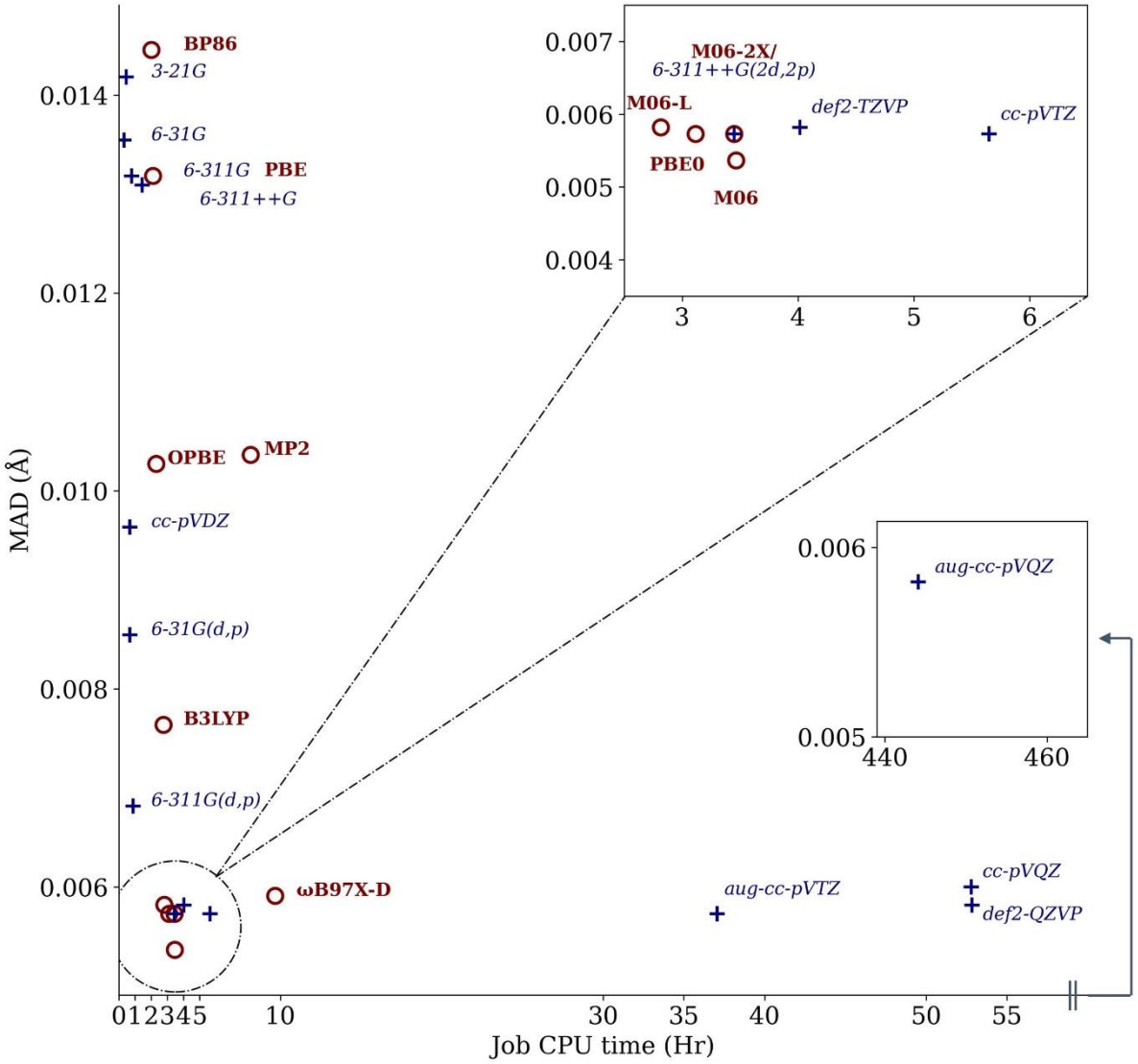

Figure 7. MAD (Å) of all calculated bond lengths against job CPU time (Hr) for the geometry optimisation of vanillin. A decrease in MAD can be observed with methods showing higher time-consumption, up to 3 hours. Methods requiring more than about 3 hours of CPU time yield

similar MAD. "+" refers to different basis sets at the M06-2X functional level of theory and "o" refers to different functional levels of theory using 6-311++G(2d,2p).

Figure 7 shows the comparison between MAD (for all bond lengths compared to the experimental) and job CPU time for the optimisation of vanillin using each method from a UFF pre-optimised structure. Red markers indicate the specified XC-functional (or MP2) using 6-311++G(2d, 2p), whereas blue markers indicate the specified basis set at the M06-2X functional level of theory. A general trend of decreasing MAD with increasing job time can be observed up to ~3 hours, after which more expensive calculations show similar MAD. The most efficient methods for the calculation of the bond lengths of vanillin seem to be M06-L, M06 and M06-2X for the functionals, and 6-311++G(2d, 2p), def2-TZVP and cc-pVTZ for the basis sets, showing the lowest MAD of about 0.006 Å at 3-5 hours of job CPU time.

2.2 Vibrational frequencies

Peak assignment of vanillin bond vibrations, band correlation between IR and Raman spectra (cross-referenced peaks between IR, Raman and computational vibrations).

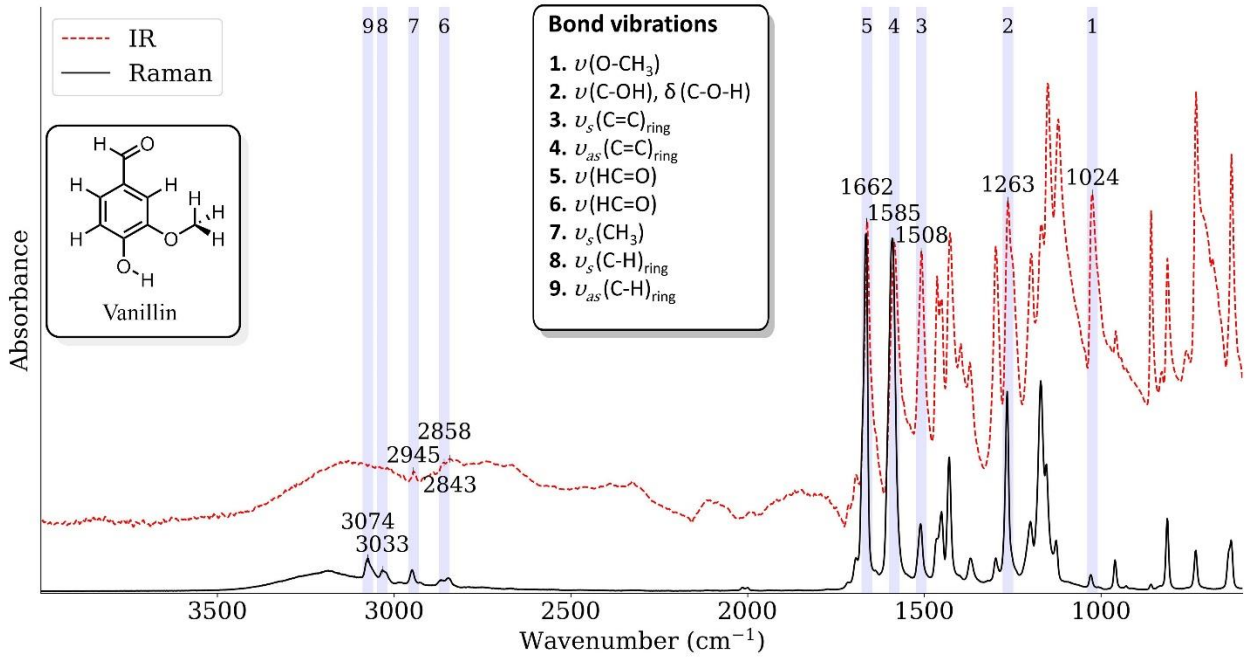

Figure 8. IR and Raman spectra of Vanillin. 9 characteristic peaks have been chosen and assigned to their corresponding mode of vibration.

9 characteristic vibrational frequency peaks for vanillin have been assigned and labelled based on the IR and Raman intensities, shown in figure 8, indicating the major vibration responsible for the peak. IR and Raman intensities have been cross-referenced with the Aldrich library of FT-IR spectra[53]. The aryl ether O-CH_3 stretching vibration is expected in the range of 1000 – 1055 cm^{-1} (**1**), the phenol C-OH stretching vibration is expected within the range of 1165 – 1335 cm^{-1} (**2**), 3 peaks are expected (two of which are clearly seen and assigned) due to the aromatic ring C-C stretching modes are expected at near 1430 cm^{-1} (not labelled), 1515 cm^{-1} (**3**) and 1600 cm^{-1} (**4**) respectively, the aldehyde C=O stretching mode is expected at near 1725 cm^{-1}(**5**), but is influenced by ring substitution, the aldehyde hydrogen stretching mode is expected as a doublet within the range of 2700-2900 cm^{-1} (**6**), the ether C-H stretching mode is expected near 2900 cm^{-}

^{1} (**7**) and finally, aromatic C-H stretching modes are expected within the range of 3030-3125 cm^{-1} (**8**, **9** labelled). visible more clearly in the Raman spectrum These assignments are in agreement with the peaks assigned by Balachandran *et al*[7], in their DFT modelling study of vanillin and isovanillin, in which they carried out normal coordinate analysis (NCA) on vanillin to calculate the frequency of each vanillin mode of vibration and then assign the calculated values to experimental IR peaks, as well as PED (Potential Energy Distribution) values.

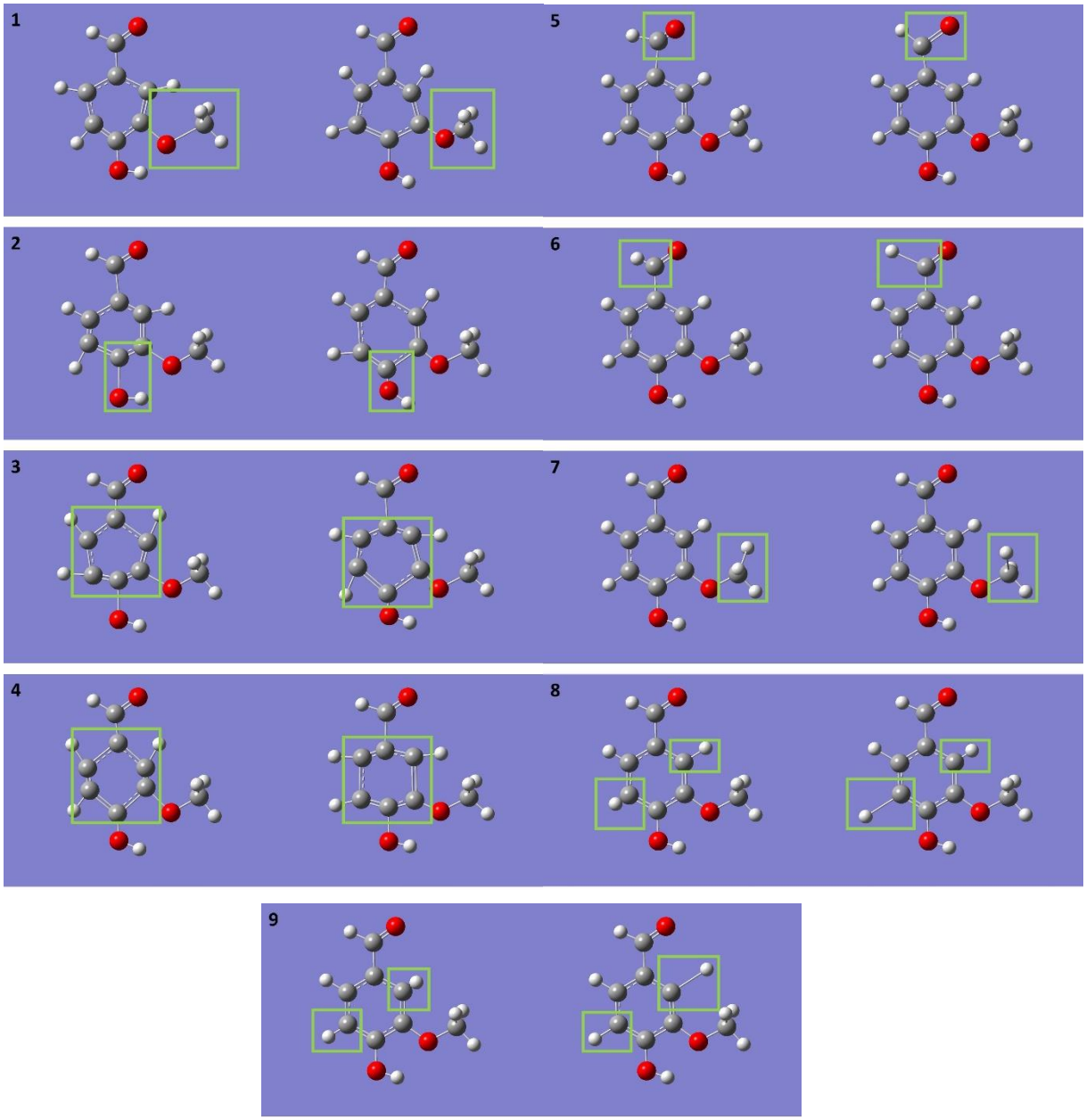

Figure 9. Characteristic modes of vibration as simulated by DFT.

Figure 9 shows the characteristic peaks vibration modes as described by M06-2X/6-311++G(2d,2p). The modes have been visualised using the finished frequency calculation file on the Gaussview-05 software[54]. While the modes of vibration are not isolated to a single bond, the major bond responsible for the vibration can be seen out-lined in green. Modes of vibration are attributed to: 1. O-CH_3 methyl-ether bond stretching, 2. C-OH phenol bond stretching, 3.

Symmetric ring stretching, 4. Asymmetric ring stretching, 5. C=O aldehyde bond stretching, 6. OC-H aldehyde hydrogen bond stretching, 7. CH_3 hydrogen bond stretching, 8. Symmetric aromatic hydrogen bond stretching and 9. Asymmetric aromatic hydrogen bond stretching. These modes of vibrations have been selected due to their specificity to the functional groups of vanillin and distinguishability in a vibrational frequency spectrum (IR or Raman).

While it is common and relatively simple to calculate the vibrational frequency of a chemical bond using the harmonic approximation model, the result will be systematically higher than the experimental value due to neglecting the effect of anharmonicity. Anharmonic corrections can be included in the calculation, however, due to the requirement for second-order (or even higher) force constants, the resulting computationally cost increases significantly. A solution to this problem is the introduction of scaling factors, which are method-specific and can be used to improve the agreement between the calculated frequencies and experimental peaks.

In their study, Truhlar *et al*[21] have shown that fundamental frequency scaling factors for a DFT method, that can be "universally" applied to frequency calculations by that methodology, can be approximated from the comparison of calculated frequency values to the experimentally observed frequencies of a database containing 38 modes of vibration across 15 molecules, referred to in their study as "F38/10". The scaling factor (λ^F) can then be calculated:

$$\lambda^F = \frac{\sum(\omega v)}{\sum(\omega^2)} \qquad (5)$$

Where ω are the DFT-calculated harmonic frequencies and v are the experimentally observed frequencies. The root-mean-square deviation associated with a scaling factor's fit to the dataset can be calculated as:

$$RMS = \sqrt{\frac{\sum(\lambda^F \omega - v)^2}{n}}$$

(6)

Where n refers to the number of modes compared, in this case 38. Minimising the RMS with respect to the scaling factor yields the optimal scaling factor for the methodology.

Truhlar *et al* have applied this method to calculate the scaling factors for a selection of common DFT methodologies, including some of the methodologies covered in the current study, for which the scaling factors used were taken from their study. For the methodologies that have not yet been covered (this includes all the functionals in the current study with 6-311++G(2d, 2p)), we present the optimised scaling factors, calculated using the procedure outlined above:

Table 1. Optimal scaling factors for studied methodologies, calculated following the procedure outlined by Truhlar *et al*[21].

Methodology	$\boldsymbol{\lambda^F}$	RMS Deviation (cm^{-1})	
		Scaled	Unscaled
BP86/6-311++G(2d,2p)	0.989	34	43
PBE/6-311++G(2d,2p)	0.986	33	47
OPBE/6-311++G(2d,2p)	0.969	42	88
M06-L/6-311++G(2d,2p)	0.959	32	107
B3LYP/6-311++G(2d,2p)	0.963	28	103
ωB97X-D/6-311++G(2d,2p)	0.948	41	137
PBE0/6-311++G(2d,2p)	0.950	34	130

M06/6-311++G(2d,2p)	0.956	45	118
M06-2X/6-311++G(2d,2p)	0.944	44	148
MP2-FC/6-311++G(2d,2p)	0.955	77	136
M06-2X/3-21G	0.969	139	158
M06-2X/6-31G	0.954	136	178
M06-2X/6-311G	0.960	109	147
M06-2X/6-311++G	0.962	126	158
M06-2X/cc-pVDZ	0.947	54	144
M06-2X/cc-pVTZ	0.945	47	147
M06-2X/cc-pVQZ	0.944	49	149
M06-2X/aug-cc-pVQZ	0.945	49	148
M06-2X/def2-TZVP	0.945	48	147

M06-2X/def2-TZVP was also calculated as a validation set. A scaling factor of 0.945 was calculated which is in close agreement to the 0.946 reported by the authors, exhibiting a scaled RMS deviation of the dataset of 48 cm^{-1}, showing a clear improvement from the 147 cm^{-1} unscaled deviation. A clear improvement in the accuracy of the methods when calculating the frequencies of the "F38/10" database (Table S7) can be seen with the application of the reported scaling factors. The RMS has been optimised with respect to λ^F (Figure S14) and the scaled RMS deviations present similar values to those of scaling factors reported by Truhlar *et al*[21], with the significant exceptions of cases with M06-2X using basis sets lacking polarisation (3-21G, 6-31G, 6-311G, 6-311++G), which still exhibit large RMS deviations after being scaled.

Comparison of vibrational frequencies using 6-311++G(2d,2p) at different XC functional levels of theory. XC functionals of different rung (GGA, mGGA, hybrid) and hybrid functionals of different HF exchange energy contribution (25% to 54%), as well as MP-2 post-HF tested. Application of error metrics (MAD and STD) for each functional.

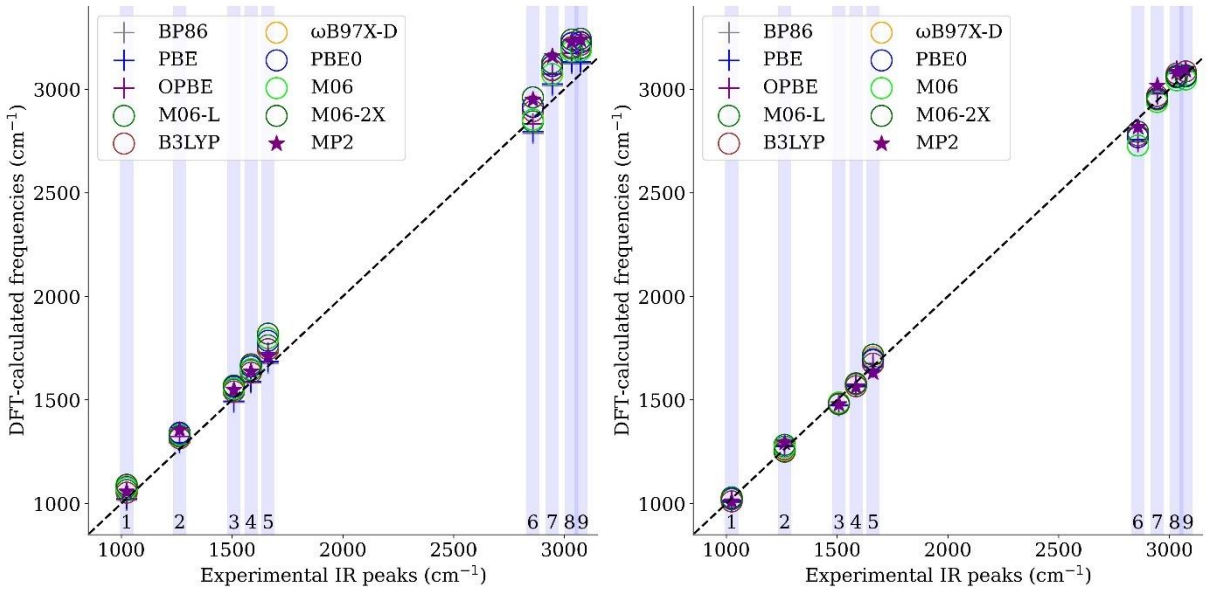

Figure 10. Vibrational frequency comparison between assigned, experimentally observed (IR, Raman) peaks and DFT harmonic frequency calculations using 6-311++G(2d, 2p) at different XC functional levels of theory, a. without scaling factors and b. with scaling factors. Vibration assignments refer to the peaks highlighted in figure 8.

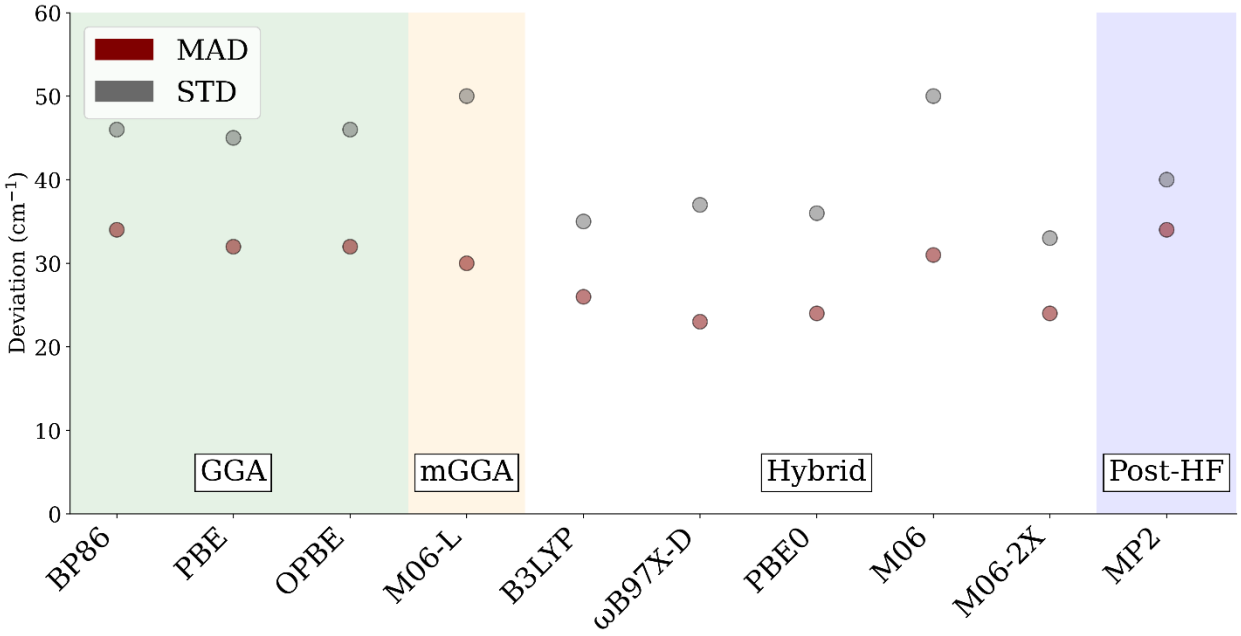

Figure 11. Vibrational frequency comparison between assigned experimentally observed IR peaks and scaled DFT harmonic frequency calculations at the M06-2X functional level of theory using different basis sets. Vibration assignments refer to the peaks highlighted in figure 5. b. Comparison of MAD and STD across different basis set parameters and families, moving from Pople double-ζ (yellow) to Pople triple-ζ (orange) to correlation-consistent (blue) to Karlsruhe (green) basis sets with different parameters.

Figure 10 shows the agreement between the bond vibrational frequency for each characteristic vibration between the frequencies calculated by each method with different XC-functional levels of theory and using MP2, using 6-311++G(2d, 2p) as the basis set, and the characteristic peaks assigned on the vibrational spectra (IR, Raman), a. without scaling and b. scaled using the scaling factors, whereas figure 11 shows the deviation metrics (MAD and STD) for the deviation between DFT and experimental frequencies. From figure 10, the effects of the scaling factors can be clearly seen across all methods, increasing the agreement (proximity of marker to the dotted line) for all vibrations.

Comparison of vibrational frequencies at the M06-2X level of theory using different basis sets. Basis sets of different type (Pople, Correlation-Consistent and Karlruhe), number of ζ orbitals, number of GTO orbitals (for Pople) and diffused orbitals tested. Application of error metrics (MAD and STD) for each basis set.

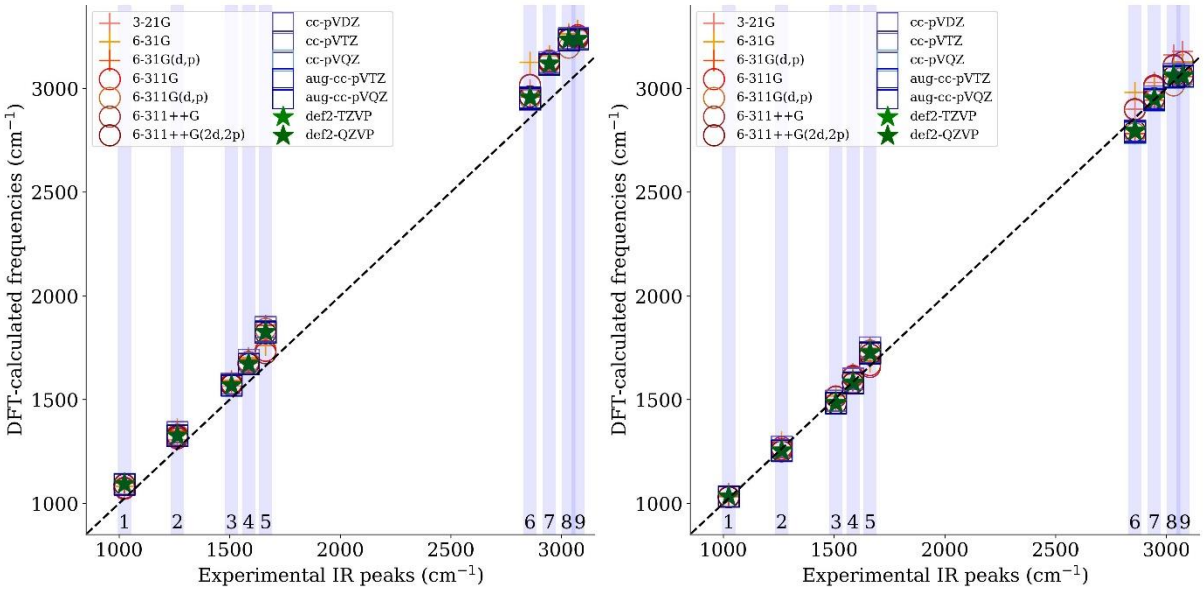

Figure 12. Vibrational frequency comparison between assigned, experimentally observed (IR, Raman) peaks and DFT harmonic frequency calculations using M06-2X with different basis sets, a. without scaling factors and b. with scaling factors. Vibration assignments refer to the peaks highlighted in figure 8.

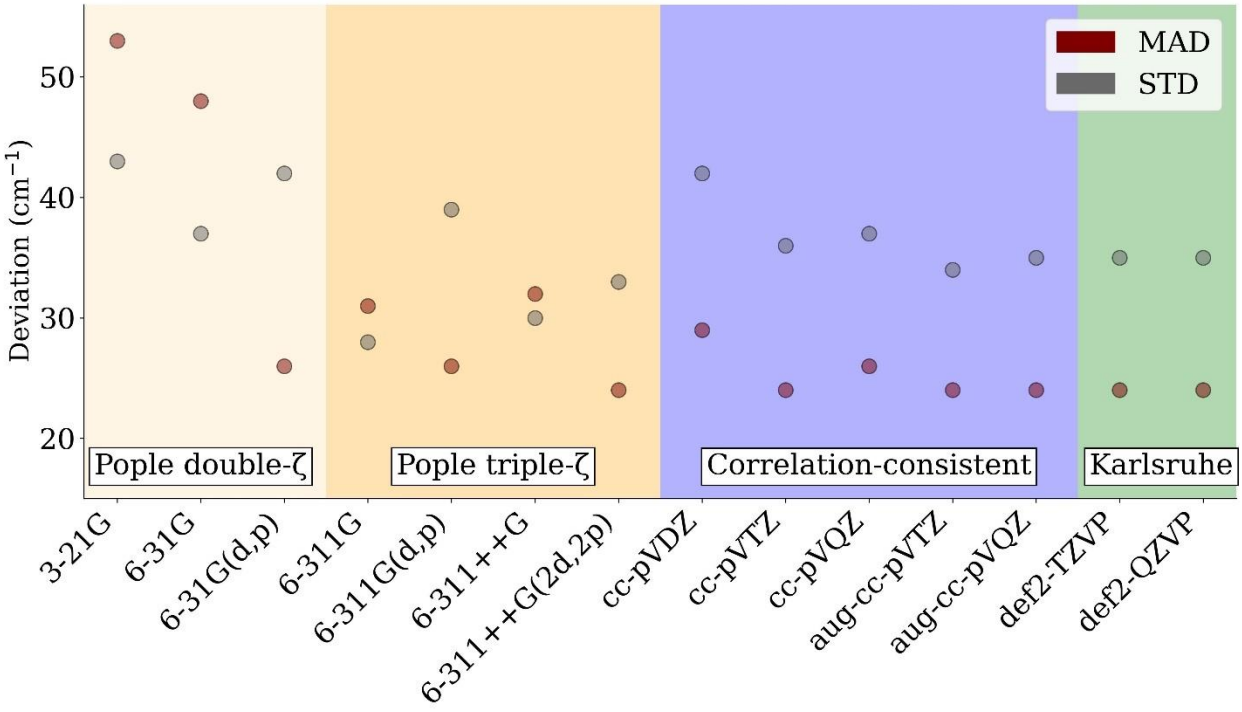

Figure 13 Deviation metrics (MAD and STD) (cm^{-1}) of characteristic vanillin bond vibrations at the M06-2X XC functional level of theory using different basis sets.

Figure 12 shows the agreement between the bond vibrational frequency for each characteristic vibration between the frequencies calculated by each method using M06-2X with difference basis sets, and the characteristic peaks assigned on the vibrational spectra (IR, Raman), a. without scaling and b. scaled using the scaling factors, whereas figure 13 shows the deviation metrics (MAD and STD) for the deviation between DFT and experimental frequencies. Application of scaling factors on vibrational frequencies has shown an improvement on the agreement of DFT-calculated to experimentally observed (IR, Raman) vibration frequencies. In a similar fashion to the bond length comparison (figure 6), an increase in accuracy (decrease in MAD) can be observed when moving from double-ζ to triple-ζ, as well as when comparing polarised to unpolarised basis sets. Similar accuracy values are observed when comparing triple-ζ to quadruple-ζ basis sets and basis sets with and without diffuse orbitals (6-311G to 6-311++G,

cc-pVTZ to aug-cc-pVTZ and cc-pVQZ and aug-cc-pVQZ). It should be noted that basis sets without polarisation have also shown much higher scaled RMS deviations (>100 cm^{-1}) from the scaling factor calculations, reinforcing the significance of polarisation in basis sets.

Comparison of MAD of studied vibrational frequencies (cm^{-1}) with job CPU time (Hr) for the frequency calculation of the optimised vanillin geometry at different XC levels of theory using 6-311++G(2d,2p) and at the M06-2X functional level of theory using different basis sets.

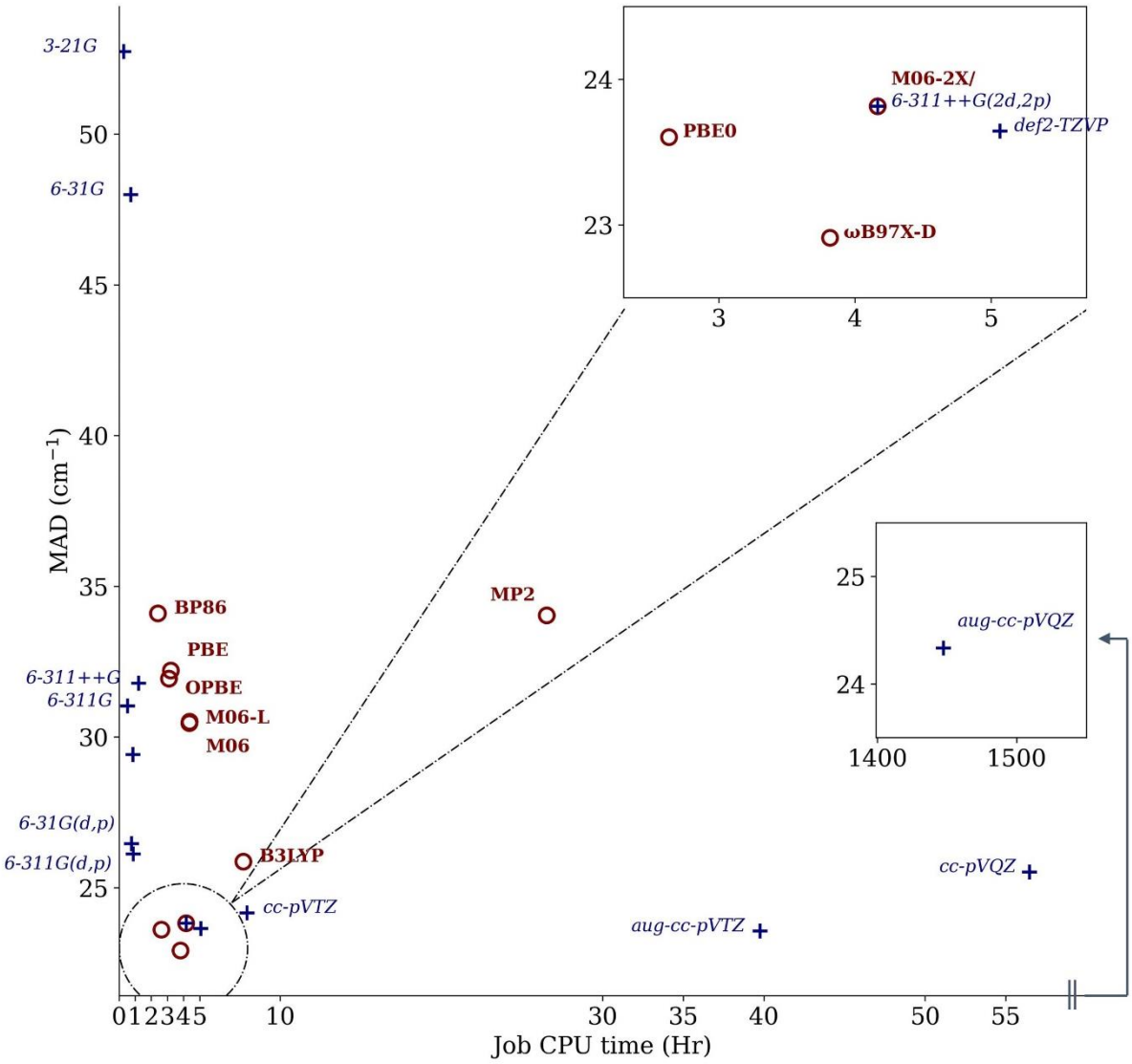

Figure 14. MAD (cm^{-1}) of all calculated bond lengths against job CPU time (Hr) for the frequency calculations of the optimised vanillin structure. A decrease in MAD can be observed with methods showing higher time-consumption, up to about 7 hours. "+" refers to different basis sets at the M06-2X functional level of theory and "o" refers to different functional levels of theory using 6-311++G(2d,2p).

Figure 14 shows the comparison between MAD (for the characteristic bond vibrations compared to the experimentally observed peaks) and job CPU time for the frequency calculation

of vanillin using each method from the optimized vanillin structure (optimized using the same method). Red markers indicate the specified XC-functional (or MP2) using 6-311++G(2d, 2p), whereas blue markers indicate the specified basis set at the M06-2X functional level of theory. A general trend of decreasing MAD with increasing job time, similar to that of bond length MAD (figure 7) can be observed up to ~5 hours, after which more expensive calculations show similar MAD. The most efficient methods for the calculation of the bond lengths of vanillin seem to be PBE0, ωB97X-D and M06-2X for the functionals, and 6-311++G(2d, 2p), def2-TZVP and cc-pVTZ for the basis sets, showing the lowest MAD of 23-24 cm^{-1} at 2-5 hours of job CPU time.

CONCLUSION

The computational modelling of the molecular structure and vibrational frequencies of vanillin using a selection of different DFT parameters has been performed and has allowed for the determination of the ideal methods in each case. Geometry optimizations of vanillin using the aforementioned DFT methods have shown to be in good agreement to experimentally observed[52] vanillin bond lengths, exhibiting MAD values between 0.005-0.02 Å, the lowest of which are can be reached at about 3-5 hours of CPU job time using mGGA/hybrid functionals (M06-L, M06, M06-2X, PBE0) and triple-ζ, polarized basis sets (using M06-2X) (6-311++G(2d,2p), cc-pVTZ, def2-TZVP).

Characteristic and identifiable bond vibrational frequencies have been selected, labelled on experimentally observed (IR, Raman) spectra and cross-referenced to literature[53] and calculated[7] with good agreement. Scaling factors taken from Truhlar *et al*[21]for methods that have already been tested have been shown to improve the accuracy of calculated vibrational frequencies and scaling factors for methods that have not been reported (BP86, PBE, OPBE, M06-L, B3LYP, ωB97X-D, PBE0, M06, M06-2X and MP2-FC with 6-311++G(2d,2p), as well as M06-2X with

3-21G, 6-31G, 6-311G, 6-311++G, cc-pVDZ, cc-pVTZ, cc-pVQZ and aug-cc-pVQZ) have been calculated, validated according to the literature, compared those of similar methods and used in the current study to yield satisfying results, resulting in an overall MAD range of 20-35 cm^{-1} for most methods (excluding double-ζ unpolarised cases). The most efficient methods for vibrational frequency calculations that yield the highest accuracy have been again found to be hybrid functionals (with 6-311++G(2d, 2p)) (PBE0, ωB97X-D and M06-2X) and triple-ζ polarised basis sets (using M06-2X) (6-311++G(2d, 2p) and def2-TZVP) converging at 2-5 hours of CPU job time.

Overall, it has been found that the hybrid XC functionals PBE0, M06-2X and ωB97X-D with 6-311++G(2d, 2p) show the highest accuracy in both geometry optimisation (bond lengths) and frequency calculations (with the used scaling factors) at the lowest time. Using M06-2X, basis set convergence occurs at the triple-ζ polarised level, with quadruple basis sets showing no significant increase in accuracy and double-ζ or unpolarised basis sets showing a lower accuracy. Diffuse orbitals have not shown a significant effect to the accuracy of the studied systems.

ASSOCIATED CONTENT

Supporting Information. The supporting information detailing computational information is available, see SI document.

AUTHOR INFORMATION

Corresponding Author

Corresponding Author *Correspondence and requests for materials should be addressed to Laia Vilà−Nadal (laia.vila−nadal@chem.gla.ac.uk)

Author Contributions

L.V.−N. suggested the experimental approach and together with M. N. designed the theoretical study. E. G. supervised the undertaking of the experimental work and provided guidance on the experimental qualitative analysis. H. S. provided guidance on the methodology and literature followed to carry out the theoretical work. M. N. performed the theoretical calculations, data analysis and presentation under the supervision of L.V.−N. M. N. wrote the paper with input from L.V.−N.

Funding Sources

Financial support for this work was provided by University of Glasgow and the Engineering and Physical Sciences Research Council scholarship (01047183).

Footnotes

The calculations performed using DFT are available in the ioChem−BD database. The geometry optimizations and vibrational frequency (IR, Raman) calculations of vanillin with all methods referred to in this paper, as well as the calculations performed for the determination of the fundamental scaling factors can be found under the ioChem-BD handle: "https://iochem-bd.bsc.es/browse/handle/100/308166".

ACKNOWLEDGMENT

The authors acknowledge the Jake Thompson and Daniel Malcolm from the University of Glasgow LVN group for helpful comments during the preparation of the manuscript. Financial support for this work was provided by University of Glasgow and the Engineering and Physical Sciences Research Council Grants (EP/S030603/1; EP/V048341/1; EP/S031170/1), Royal Society of Chemistry RSC Hardship Grant (Covid-19). We thank the EPSRC Doctoral Training Parntertship (DTP) funding received by the University of Glasgow for Michael Nicolaou PhD studentship (EP/R513222/; EP/T517896/1). We also thank the University of Glasgow (UofGla) Early Career Development Programme (ECDP) 2021, the UofGla Reinvigorating Research Scheme 2022, and the School of Chemistry for long-lasting support. Results were obtained using

the ARCHIE-WeSt High-Performance Computer (www.archie-west.ac.uk) based at the University of Strathclyde and our School of Chemistry HPC cluster.

REFERENCES

1. Fitzgerald, D. J. *et al.* Mode of antimicrobial of vanillin against Escherichia coli, Lactobacillus plantarum and Listeria innocua. *J. Appl. Microbiol.* **97**, 104–113 (2004).

2. Zeb, A. Concepts of Antioxidants in Foods. in *Phenolic Antioxidants in Foods: Chemistry, Biochemistry and Analysis*. 3-21 (Springer International Publishing, 2021). doi:10.1007/978-3-030-74768-8.

3. Ngarmsak, M. *et al.* Antimicrobial Activity of Vanillin against Spoilage Microorganisms in Stored Fresh-Cut Mangoes. *J. Food Prot.* **69**, 1724-1727 (2006).

4. Maisch, N. A., Bereswill, S. & Heimesaat, M. M. Antibacterial effects of vanilla ingredients provide novel treatment options for infections with multidrug-resistant bacteria - A recent literature review. *Eur. J. Microbiol. Immunol.* **12**, 53–62 (2022).

5. Yan, X. *et al.* Vanillin protects dopaminergic neurons against inflammation-mediated cell death by inhibiting ERK1/2, P38 and the NF-κB signaling pathway. *Int. J. Mol. Sci.* **18**, 389 (2017).

6. Fache, M. *et al.* Vanillin, a promising biobased building-block for monomer synthesis. *Green Chem.* **16 (4)**, 1987–1998 (2014).

7. Balachandran, V. & Parimala, K. Vanillin and isovanillin: Comparative vibrational spectroscopic studies, conformational stability and NLO properties by density functional theory calculations. *Spectrochim. Acta* **95**, 354–368 (2012).

8. Gunasekaran, S. & Ponnusamy S. Vibrational spectra and normal coordinate analysis on an organic non-linear optical crystal-3-methoxy-4-hydroxy benzaldehyde. *Indian J. Pure Appl. Phys.* **43**, 838-843 (2005).

9. Hosoya, T., Yamamoto, K., Miyafuji, H. & Yamada, T. Selective production of bio-based aromatics by aerobic oxidation of native soft wood lignin in tetrabutylammonium hydroxide. *RSC. Adv.* **10**, 19199–19210 (2020).

10. Voitl, T. & Von Rohr, P. R. Oxidation of lignin using aqueous polyoxometalates in the presence of alcohols. *ChemSusChem.* **1**, 763–769 (2008).

11. Rinaldi, R. *et al.* Paving the way for lignin valorisation: Recent advances in bioengineering, biorefining and catalysis. *Angew. Chem. Int. Ed.* **128**, 8296–8354 (2016).

12. Hernandez J. H. Production of Vanilla – Agricultural systems and curing. in *Handbook of vanilla science and technology*. 1 – 24 (2019).

13. Gallage, N. J. & Møller, B. L. Vanillin-bioconversion and bioengineering of the most popular plant flavor and its de novo biosynthesis in the vanilla orchid. *Mol. Plant* **8**, 40–57 (2015).

14. El Moncef, A., El Hadrami, E. M., González, M. A., Zaballos, E. & Zaragozá, R. J. Experimental and DEE study of the conversion of ephedrine derivatives into oxazolidinones. Double SN2 mechanism against SN1 mechanism. *Tetrahedron* **66**, 5173–5184 (2010).

15. Townsend, P. A. & Grayson, M. N. Density Functional Theory Transition-State Modeling for the Prediction of Ames Mutagenicity in 1,4 Michael Acceptors. *J. Chem. Inf. Model.* **59**, 5099–5103 (2019).

16. Brémond, É. *et al.* Benchmarking Density Functionals on Structural Parameters of Small-/Medium-Sized Organic Molecules. *J. Chem. Theory Comput.* **12**, 459–465 (2016).

17. Flaig, D. *et al.* Benchmarking Hydrogen and Carbon NMR Chemical Shifts at HF, DFT, and MP2 Levels. *J. Chem. Theory Comput.* **10**, 572–578 (2014).

18. Ji, Y. *et al.* DFT-Calculated IR Spectrum Amide I, II, and III Band Contributions of N-Methylacetamide Fine Components. *ACS Omega* **5**, 8572–8578 (2020).

19. Bursch, M., Mewes, J.-M., Hansen, A. & Grimme, S. Best-Practice DFT Protocols for Basic Molecular Computational Chemistry**. *Angew. Chem. Int. Ed.* **61**, (2022).

20. Zhao, Y. & Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **41 (2)**, 157–167 (2008).

21. Alecu, I. M., Zheng, J., Zhao, Y. & Truhlar, D. G. Computational thermochemistry: Scale factor databases and scale factors for vibrational frequencies obtained from electronic model chemistries. *J. Chem. Theory Comput.* **6**, 2872–2887 (2010).

22. Frisch, M., Trucks, J., Schlegel H.B. & et al. Gaussian 16, Revision C.01. (2016).

23. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098-3100 (1988).

24. Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **33 (12)**, 8822-8824 (1986).

25. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77 (18)**, 3865-3868 (1996).

26. Perdew, J. P., Burke, K. & Ernzerhof, M. (Errata) Generalized Gradient Approximation Made Simple. [*Phys. Rev. Lett.* **77***,* 3865 (1996)] (1997).

27. Hoe, W.-M., Cohen, A. J. & Handy, N. C. Assessment of a new local exchange functional OPTX. *Chem. Phys. Lett.* **341**, 319-328 (2001).

28. Handy, N. C. & Cohen, A. J. Left-right correlation energy. *Mol. Phys.* **99 (5)**, 403–412 (2001).

29. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).

30. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab Initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **98 (45)**, 11623-11627 (1994).

31. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev.* B **37 (2)**, 785-789 (1988).

32. Chai, J. -D. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **128**, 084106-(1-15) (2008).

33. Chai, J. -D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).

34. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **110 (13)**, 6158–6170 (1999).

35. Frisch, M. J., Head-Gordon, M. & Pople, J. A. A direct MP2 gradient method. *Chem. Phys. Lett.* **166 (3)**, 275-280 (1990).

36. Frisch, M. J., Head-Gordon, M. & Pople, J. A. Semi-direct algorithms for the MP2 energy and gradient. *Chem. Phys. Lett.* **166 (3)**, 281-289 (1990).

37. Head-Gordon, M., Pople, J. A. & Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **153 (6)**, 503-506 (1988).

38. Saebø, S. & Almlöf, J. Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chem. Phys. Lett.* **154 (1)**, 83-89 (1989).

39. Head-Gordon, M. & Head-Gordon, T. Analytic MP2 frequencies without fifth-order storage. Theory and application in bifurcated hydrogen bonds in the water hexamer. *Chem. Phys. Lett.* **220**, 122-128 (1994).

40. Clark, T., Chandrasekhar, J., Spitznagel, G. W. & Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *J. Comput. Chem.* **4**, 294–301 (1983).

41. Krishnan, R., Binkley, J. S., Seeger, R. & Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **72 (1)**, 650–654 (1980).

42. McLean, A. D. & Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11-18. *J. Chem. Phys.* **72 (10)**, 5639–5648 (1980).

43. Francl, M. M. *et al.* Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **77 (7)**, 3654–3665 (1982).

44. Spitznagel, G. W., Clark, T., von Ragué Schleyer, P. & Hehre, W. J. An evaluation of the performance of diffuse function-augmented basis sets for second row elements, Na-Cl. *J. Comput. Chem.* **8**, 1109–1116 (1987).

45. Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **90 (2)**, 1007–1023 (1989).

46. Weigend, F. & Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).

47. Cundary, T. R. & Gordon, M. S. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc*. **114 (25)**, 10024-10035 (1992).

48. Ochterski, J. W. Vibrational Analysis in Gaussian. (1999).

49. Miertuš, S., Scrocco, E. & Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **55**, 117-129 (1981).

50. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).

51. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge structural database. *Acta Cryst. B.* **72**, 171–179 (2016).

52. Velavan, R., Sureshkumar, P., Sivakumar, K. & Natarajan, S. Organic compounds Vanillin-I. *Acta Cryst. C* **51**, 1131-1135 (1995).

53. Pouchert, C. J. & Aldrich Chemical Company. The Aldrich library of FT-IR spectra. (1989).

54. Dennington, R., Keith, T. A. & Millam, J. A. Gaussview 05. (2016).