

CHEMISTRY

A European Journal

A Journal of



Accepted Article

Title: Improved infrared spectra prediction by DFT from a new experimental database

Authors: Madanakrishna Katari, Edith Nicol, Vincent Steinmetz, Guillaume van der Rest, Duncan Carmichael, and Gilles Frison

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *Chem. Eur. J.* 10.1002/chem.201700340

Link to VoR: <http://dx.doi.org/10.1002/chem.201700340>

Supported by
ACES

WILEY-VCH

Improved infrared spectra prediction by DFT from a new experimental database

Madanakrishna Katari,^a Edith Nicol,^a Vincent Steinmetz,^b Guillaume van der Rest,^b Duncan Carmichael,^a and Gilles Frison*^a

^a LCM, CNRS, Ecole polytechnique, Université Paris-Saclay, 91128 Palaiseau, France.

^b Laboratoire de Chimie Physique, Université Paris Sud, CNRS, 91405 Orsay, France.

Corresponding author: Gilles Frison

Tel. : +33 (0)1 69 33 48 34; Fax : +33 (0)1 69 33 48 03 ; twitter : @gillesfrison

E-mail: gilles.frison@polytechnique.edu

Abstract

This work aims to improve the computation of infrared spectra of gas-phase cations using DFT methods. Experimental IRMPD spectra for ten Zn and Ru organometallic complexes have been used to provide reference data for 64 vibrational modes in the 900-2000 cm⁻¹ range. The accuracy of the IR vibrational frequencies predicted for these bands has been assessed over five DFT functionals and three basis sets. The functionals include the popular B3LYP and M06-2X hybrids and the range-separated hybrids (RSH) CAM-B3LYP, LC-BLYP and ω B97X-D. B3LYP gives the best mean absolute error (MAE) and root-mean-square error (RMSE) values of 7.1 and 9.6 cm⁻¹, whilst the best RSH functional, ω B97X-D, gives 12.8 and 16.6 cm⁻¹ respectively. Using linear correlations instead of scaling factors improves the prediction accuracy significantly for all functionals. Experimental and computed spectra for a single complex can show significant differences even when the molecular structure is calculated correctly, and a means of defining confidence limits for any given computed structure is also provided.

Introduction

InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy can provide infrared signatures for gaseous molecular ions trapped in the high vacuum cells of mass spectrometers.^[1] The technique emerged at the beginning of the current century^[2] and is recognized as a powerful tool for distinguishing between isomers and conformers of organic, inorganic and biological species.^[3] In most cases, the IRMPD band assignments that allow structural characterization are made through comparison with model spectra that are predicted by modern electronic structure calculations such as density functional theory (DFT). The accuracy of these calculations is essential to the use of the technique.

Agreement between experimental IRMPD and calculated spectra is generally satisfying, in particular for band positions, but some small differences are often observed. Mismatches between the calculated and experimental spectra can occur from isomerization of the studied molecule if this is rapid on the timescale of the multiphoton absorption process, as has recently been observed for a_n ions.^[4] If IR spectra are calculated by DFT methods that use harmonic approximations, discrepancies can arise as a result of anharmonic and finite temperature effects.^[5] Methods for modelling IR spectra that treat these effects have been devised for both classical^[6] and quantum^[7] molecular dynamic simulations, and anharmonic effects can also be included in high level calculations by the use of methods such as VSCF^[8] or GVPT2.^[9] However, static DFT calculations remain the fastest, most convenient and common way to model IR spectra, with their main drawback being the need to use scaling factors to obtain the correct vibrational frequencies. Numerous literature studies report vibrational scaling factors that are appropriate for a given combination of DFT functionals and basis sets.^[10] Their numerical values can vary slightly as a function of the database used to establish the scaling factor; the values of 0.9614,^[10a] 0.9664^[11] and 0.9800^[12] have been recommended in different studies for the B3LYP/6-31G(d) calculational level. As a consequence, these empirical scaling factors introduce a degree of uncertainty that is expressed in the predicted vibrational frequencies.^[13] For example, for a band computed at $\nu = 1800\text{ cm}^{-1}$ at the B3LYP/6-31G(d) level, the values obtained upon scaling by 0.9614 (1731 cm^{-1}) or 0.9800 (1764 cm^{-1}) differ by 33 cm^{-1} , which is highly significant difference in terms of interpretation. In addition to the standard approach of applying a uniform scaling factor over the whole infrared spectral region, further strategies have also been developed to improve the accuracy of the predicted vibrational frequencies.^[14] For example, different scaling factors can be used for low-energy ($< 1800\text{ cm}^{-1}$) and high-energy ($> 1800\text{ cm}^{-1}$) vibrational modes.^[15] Scaling equations or specific (or local) scaling factors have also been determined for a specific family of compounds or vibrational modes.^[16]

The precise values of these scaling factors are derived from the analysis of experimental infrared absorption spectra. IRMPD spectra, which are gas-phase “action” (or “consequence”) spectra, show a slight red-shift relative to classical absorption spectra,^[3e] but classical scaling factors are still commonly used to assign IRMPD bands and to distinguish between different isomers and/or conformers. Some specific scaling factors have been determined for the interpretation of gas phase spectroscopic studies of biomolecules,^[17] and the relative performance of different DFT functionals has been evaluated in several studies.^[18] A mean error of around 10 cm^{-1} is obtained for the best functionals, which include B3LYP. However, these calibrations have generally been made upon limited libraries of compounds and are only available for a restricted range of functionals. An improved understanding of the nature of the difference between experimental and calculated IRMPD

spectra is therefore of significant interest: in cases where a computed spectrum does not fit exactly with an experiment, it would make the important distinction between whether the discrepancy arises from inaccuracies of computation or from the fact that the proposed structure that has been calculated is not the same as the experimentally observed one.

Range-separated hybrid (RSH) functionals differ from conventional hybrid functionals by the degree to which exact exchange, whose contribution increases with interelectronic distance, is included.^[19] These outperform B3LYP and other conventional exchange-correlation functionals for molecular systems in which the self-interaction error is large enough to induce inaccuracies. For example, the use of RSH functionals^[20] significantly improves the description of the electronic structure of the radical protonated peptide intermediate that are generated by electron capture- or transfer-dissociation (ECD/ETD) processes.^[21] Fragments that result from ECD and ETD have been the subject of various studies in which their IRMPD spectrum has been recorded.^[22] However, the capacity of RSH functionals to reproduce IRMPD spectra accurately is still largely unexplored. Three RSHs (*vide infra*) have therefore been included in this work, in order to determine the accuracy with which they can model infrared spectra.

Estimating the precision with which DFT methods can model infrared spectra requires a solid body of experimental IRMPD reference spectra that, under ideal circumstances, would be associated unequivocally with one specific conformation of a chemical compound. This constraint is difficult to meet with peptides, but it is usually nicely satisfied by metal complexes whose geometries are normally well-characterized and clearly defined, and which do not normally have low-energy reaction pathways to isomerize. In the present work, we have therefore measured IRMPD spectra of ten organometallic complexes that contain bipyridyl-type ligands, and computed their infrared spectra using standard and RSH functionals. A comparison between theoretical and experimental results is used to define scaling procedures that can be used to provide satisfactory interpretations of IRMPD gas-phase studies involving cations, as well as to evaluate the degree of confidence that can be expected when these functionals, basis sets and scaling procedures are used to model infrared spectra.

Results and discussion

The ten complexes **I-X** that are depicted in Figure 1 have been studied in this work. They are mainly dicationic zinc (**I-VIII**) and ruthenium (**IX**) complexes, but a monocationic ruthenium complex (**X**) is also included. All are hexacoordinated, as has been well established by several X-ray structure analyses^[23] and each complex includes at least one of the bipyridyl-type ligands **1-5**. They therefore comprise a series of geometrically related homoleptic (**I-V**) or heteroleptic (**VI-X**) complexes.

The IRMPD spectra of **I-X** have been recorded over the 900-2000 cm⁻¹ range. To ensure the accuracy of the measurements, and to eliminate calibration errors, spectra were obtained for each complex on at least two different days. Representative spectra for **I** are given in Figure 2 and all of the spectra for the remaining complexes are presented in the SI (Figures S1-S10). The largest bands are observed in every spectrum of a given compound, but lower intensity bands may be absent in some spectra. These variations of detail reflect how variables such as laser alignment, intensity of the parent ion,

irradiation time or laser attenuation affect the precise nature of the processes that occur within the parent ion that give rise to the IRMPD fragments.

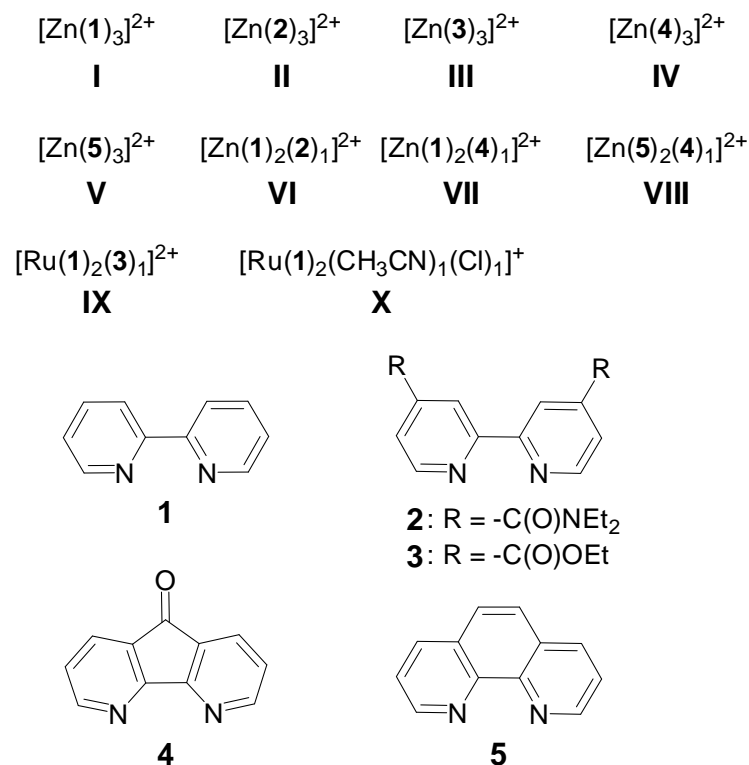


Figure 1. Complexes I-X studied in this work

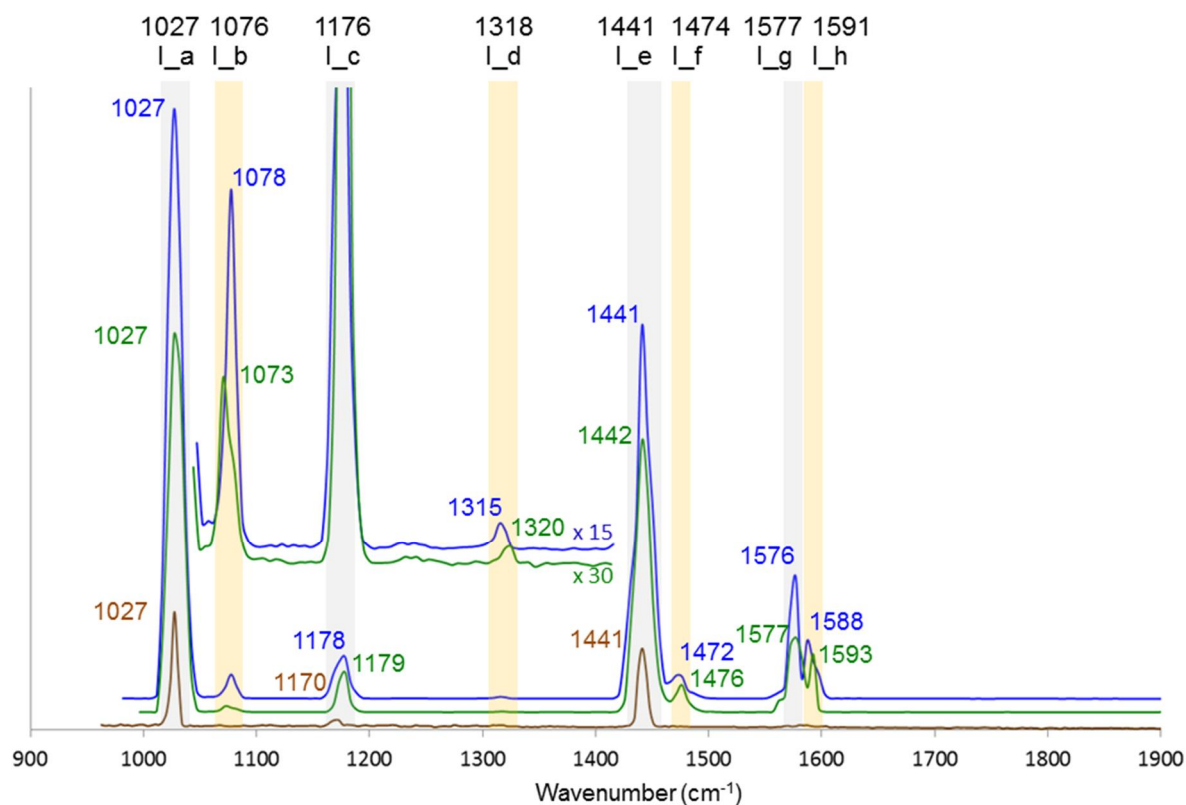


Figure 2. Experimental IRMPD spectra of I. Irradiation time: 500 ms (blue and green lines) or 250 ms (brown line).

Eight bands, noted I_a to I_h, have been clearly identified for compound I (Figure 2), and a total of 68 bands have been obtained from compounds I-X. Comparing different spectra from a given compound shows that wavenumbers vary slightly for any given band, and this serves to illustrate the degree of uncertainty that is associated with the IRMPD experiment. For example, the band I_b is observed at 1073 and 1078 cm⁻¹ in two different measurements of I. Differences are normally smaller than 10-15 cm⁻¹, but quite large variations (for instance of up to 29 cm⁻¹ for band III_g, Figure S3) are occasionally observed. In this work, the mean of the different values is taken to be the experimental wavenumber of each band. To limit imprecision in the experimental values, we have set aside four bands^[24] that show differences larger than 15 cm⁻¹ in two different spectra, so that these do not feature in the experimental band set used to evaluate the accuracy of DFT methods (*vide infra*). Assignments of the vibrational modes, which have been made through comparison with B3LYP/6-31+G(d,p) computed IR spectra (see Table S1), are in agreement with the solution-phase spectra obtained for similar zinc and ruthenium complexes including bipyridyl- and related ligands.^[25] Numerous couplings between different vibrational modes, in particular between CC_{ring} and CN_{ring} bond stretches, and in-plan CH bending vibrations, are found. In depth comparison of the various complexes indicates that the vibrational modes of each ligand are largely independent of the nature of the complex, so that they can be assigned quite straightforwardly for each ligand 1-5 (Figure 3). It should be noted that such band assignments are not necessary (and are not used) for evaluating the accuracy of any given DFT functional to reproduce the experimental vibrational frequencies.

Ligands

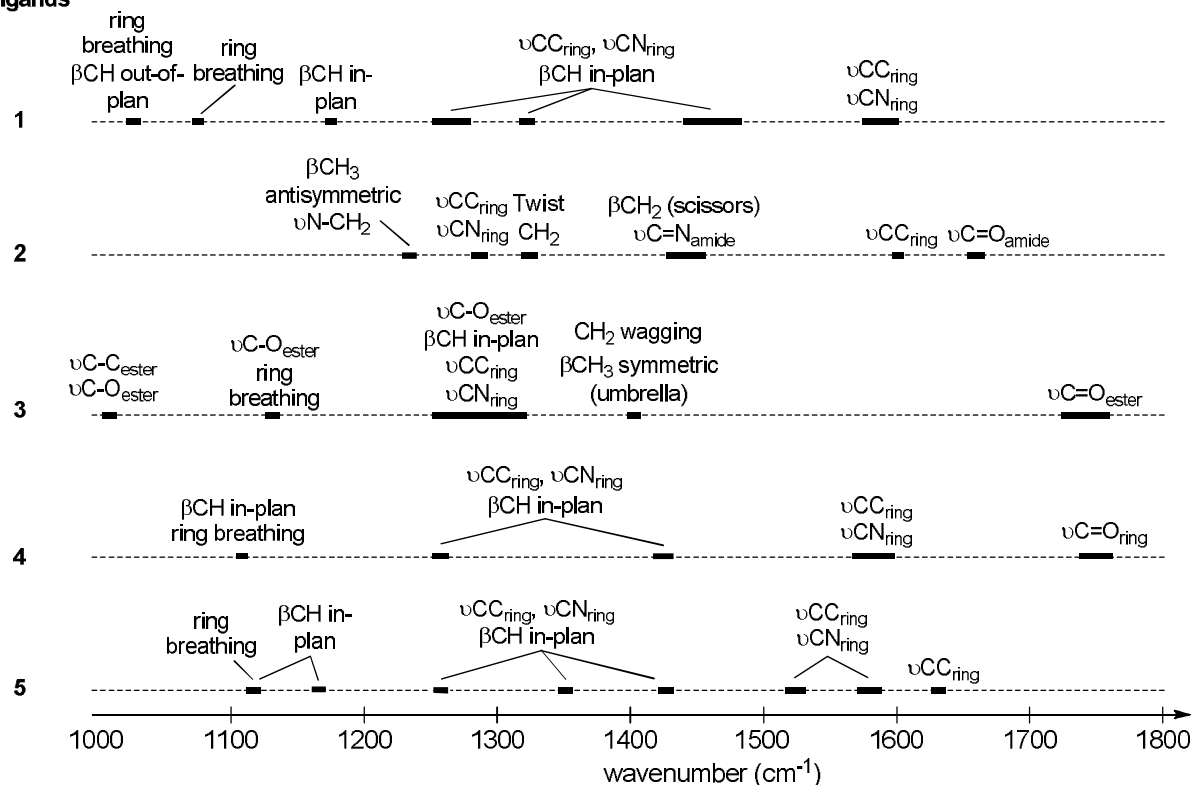


Figure 3. Assignment of vibrational mode wavenumbers for neutral ligands 1-5 in organometallic complexes.

In an initial study, IR spectra of I-X were modeled at the B3LYP/6-31+G(d,p) level. Comparison between experimental and theoretical spectra (for an example, see Figures 2 and 4 for the case of I) reveals a satisfying qualitative agreement. As expected, computed band intensities are reproduced

poorly because IRMPD spectra are plotted as a function of molecular fragmentation rather than dipole changes. This is nicely illustrated by the I_a band which has the largest intensity in the experimental spectrum, but not in the theoretical one, whose line intensity is plotted in the conventional manner as a function of IR absorption rather than molecular fragmentation. We also note that a number of bands that should theoretically be present at low intensity do not give rise to fragmentation under the experimental conditions and are not observed in the IRMPD spectra. In the subsequent discussion, we will restrict our comparison only to those peaks that are observed.

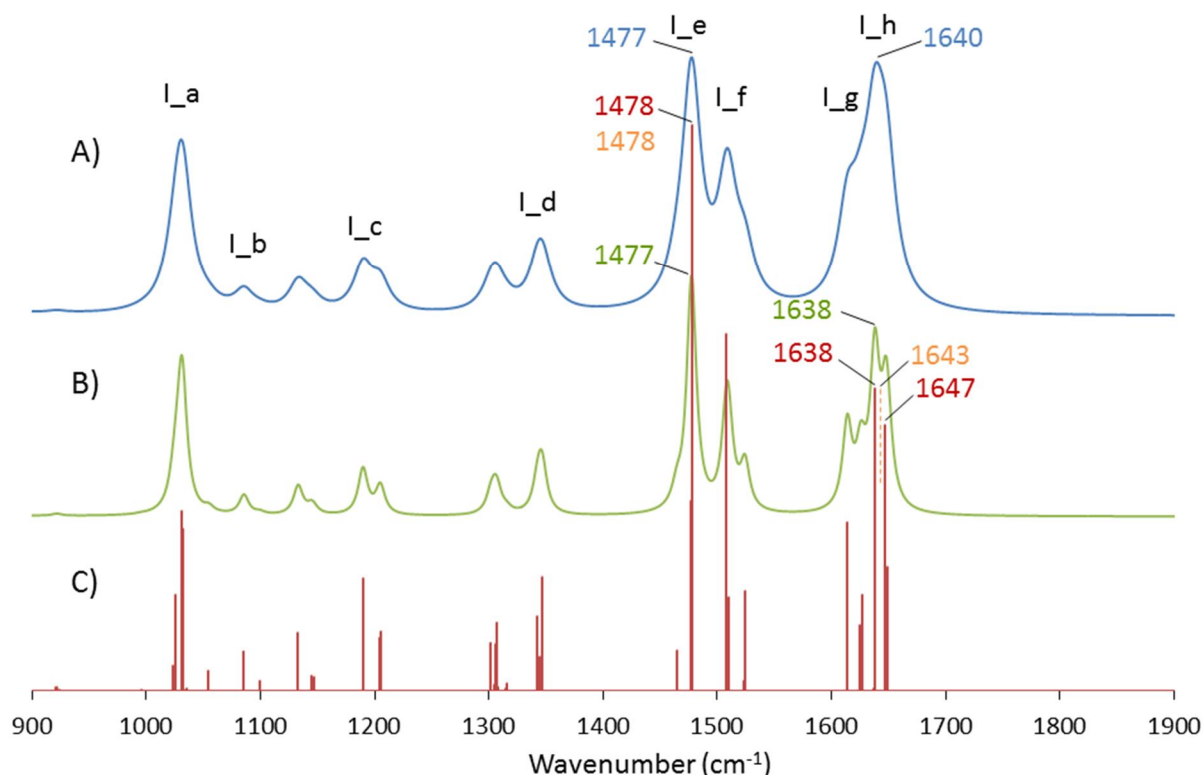


Figure 4. Unscaled calculated IR spectra of **I** at the B3LYP/6-31+G(d,p) level. Convolved spectra obtained using a 10 (A) or 5 (B) cm^{-1} (fwhm) Lorentzian linewidth function, and stick bars spectra (C).

Different strategies have been used to compare the experimental and theoretical vibrational frequencies quantitatively. Whilst the experiments produce broad bands having a linewidth of *c.a.* 10-15 cm^{-1} , the DFT calculations furnish stick spectra that can be convoluted using Gaussian or Lorentzian profiles. We have explored four different protocols for treating these theoretical data in ways that transform them into single vibrational frequencies that are suitable for comparison with the experimentally observed bands. Initially, we identify the wavenumber range in the computed spectra that are associated with each experimental band. For example, bands I_e and I_h are associated with all of the unscaled theoretical vibrational frequencies in the 1464-1478 and 1637-1649 cm^{-1} ranges, respectively (Figure 4). In the first treatment protocol (P1), the largest peak of each range in the theoretical spectrum is considered to determine the unscaled theoretical vibrational frequency corresponding to the associated experimental band (Figure 4C, in red), and this attributes values of 1478 and 1638 cm^{-1} to I_e and I_h, respectively. The second protocol (P2) differs from the first one when several peaks are computed to lie in the same region, and these have an intensity that is at least equal to half of the most intense local peak, a situation that occurs in the 1637-1649 cm^{-1} range of **I**. In this case, the mean of the different values of these computed peak frequencies is taken (Figure 4C, in orange). The third (P3) and fourth (P4) protocols involve convoluting the theoretical

spectra with a 5 (Figure 4B, in green) or 10 cm^{-1} (Figure 4A, in blue) (full width at half maximum, fwhm) Lorentzian linewidth function, respectively. If several peaks still appear in the convoluted spectra (see I_h in Figure 4B), the most intense is then used to determine the unscaled theoretical vibrational frequency, and if a band is convoluted with a neighboring band, then the same unscaled theoretical vibrational frequency is attributed to both bands (see I_g and I_h in Figure 4A).

Table 1 provides the results of these different data management protocols for several scaling values applied to the crude B3LYP/6-31+G(d,p) theoretical vibrational data (Table 1). Firstly, we observe that changing the protocol (P1 to P4) that is used to transform the theoretical vibrational frequencies into single band frequencies for quantitative comparison with the experimental spectra has a negligible influence on the scaling factor that is required for subsequent correction to the experimental values. Equally, the associated mean absolute error (MAE) and root-mean-square error (RMSE) show a minimal dependence upon the method employed, except in the case of protocol P4 which generates slightly greater values of both RMSE and MAE (see Experimental Section for details on RMSE and MAE). Equally, the mathematical treatment used to obtain the scaling factors (which involved either averaging the individual values or the minimization of MAE or RMSE, see Experimental Section for details), also has a minimal effect on their values, which ranges from 0.9768 to 0.9791 when using P2. The variation of MAE and RMSE values when used with these scaling factors is small, with their values being $10.2 \pm 0.3 \text{ cm}^{-1}$ for MAE and $13.1 \pm 0.3 \text{ cm}^{-1}$ for RMSE. However, it should be noted that the use of imprecise scaling factors can lead to significant deviations, as shown in Figure 5. For example, increasing or decreasing the scaling factor by 0.01 increases the errors by roughly 50 %.

Table 1. Scaling factors, linear correlation values and associated MAE and RMSE obtained at the B3LYP/6-31+G(d,p) level of calculation for different data management.

		P1 ^a	P2 ^a	P3 ^a	P4 ^a
Average ^b	sf_{ave}	0.9790	0.9789	0.9791	0.9789
	MAE	10.15	10.13	9.98	10.30
	RMSE	13.33	13.27	13.11	13.44
MAE ^c	sf_{MAE}	0.9778	0.9778	0.9783	0.9783
	MAE	10.04	10.02	9.92	10.21
	RMSE	13.10	13.03	12.93	13.29
RMSE ^d	sf_{RMSE}	0.9770	0.9769	0.9772	0.9768
	MAE	10.15	10.17	10.03	10.39
	RMSE	13.05	12.97	12.83	13.13
Linear correction ^e	a^e	1.0631	1.0647	1.0632	1.0656
	b^e	-54.7	-56.7	-54.9	-57.8
	R^{2f}	0.9976	0.9978	0.9978	0.9978
	MAE	7.36	7.10	7.06	7.16
	RMSE	9.93	9.59	9.62	9.65

^a Protocol used to associate calculated peak with experimental bands (see text); ^b scaling factor obtained as the arithmetic averaging of the individual scaling factors; ^c scaling factor selected to minimize the MAE; ^d scaling factor selected to minimize the RMSE; ^e predicted frequencies values obtained from the linear correlation analysis, with $\nu_{predict} = (\nu_{calc} - b)/a$; ^f linear correlation coefficient.

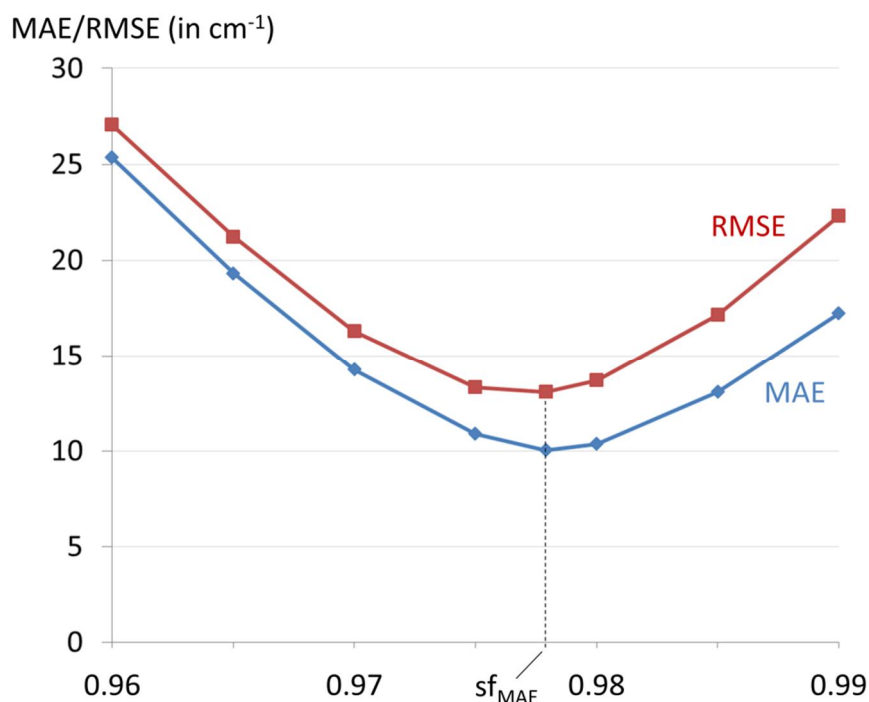


Figure 5. Variation of MAE and RMSE depending on the scaling factor, at the B3LYP/6-31+G(d,p) level with protocol P2.

Figure 6 shows the unscaled theoretical vibrational frequencies that are generated by protocol P2 at the B3LYP/6-31+G(d,p) level along with their relation to the experimental vibrational frequencies. A clear, and very convincing, linear correlation (coefficient of determination $R^2 = 0.9978$) is observed, as would be expected given the widespread use of scaling factors to fit experimental data of this kind. However, the linear correlation equation that these data generate takes the general form: $y = ax + b$ where $b \neq 0$, and this obviously indicates that the use of a scaling factor alone is not fully adequate for the prediction of experimental values from the theoretical vibrational frequencies. The multiplicative $1/a$ factor that we deduce (0.9392 at the B3LYP/6-31+G(d,p) level) induces a larger redshift, particularly for high wavenumbers, relative to the scaling factor obtained using the more classical approach (with sf_{ave} of 0.9789 giving a $\Delta shift$ of -40 and -79 cm^{-1} at 1000 and 2000 cm^{-1} , respectively), whilst the $-b/a$ factor causes a uniform blueshift of all wavenumbers (by +53 cm^{-1} at the same level). Globally, the result is that, compared to the linear correction, the scaling factor induces redshift for low wavenumbers and blueshift for high wavenumbers. Figure 6 suggests that such a linear correction is likely to be a better option than a simple multiplicative scaling factor. The error distribution between experimental and predicted frequency values is more evenly distributed whatever the wavenumber, which is not the case with a scaling factor. Calculations of MAE and RMSE using the linear correction confirm this (Table 1), in that the errors are then reduced significantly, by about 25-30%, when compared to the best values obtained using the scaling factor. Comparing the results from the various protocols used to transform the theoretical vibrational frequencies from the calculated spectra into single bands shows slightly larger errors for P1. Large individual errors are particularly undesirable when comparing experimental and theoretical spectra. Consequently, RMSE appears to give a better accuracy diagnosis compared to MAE. Protocol P2, which shows the lowest RMSE (9.59 cm^{-1}), is therefore employed to define theoretical band positions in the remainder of this work.

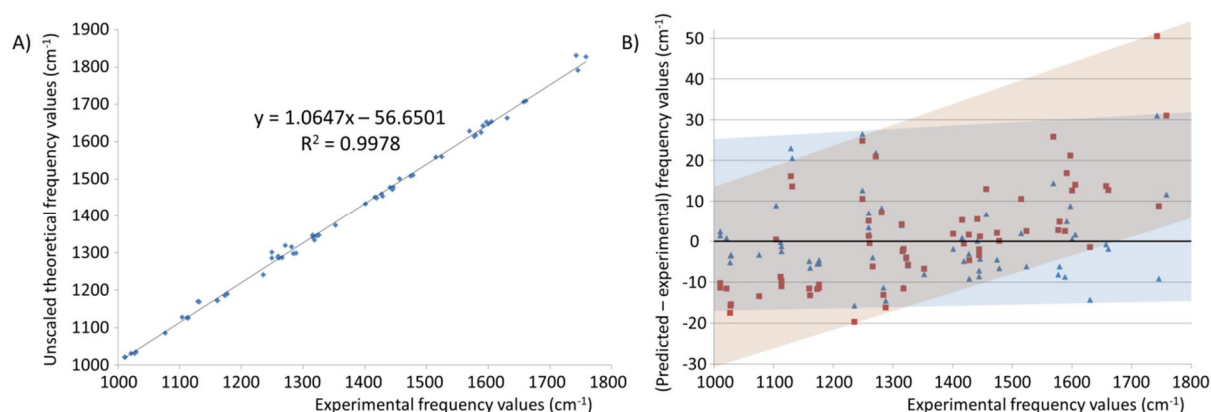


Figure 6. A) Experimental vs unscaled theoretical vibrational frequencies computed at the B3LYP/6-31+G(d,p) level with protocol P2 and B) difference between experimental and predicted frequency values, for each experimental value, using a scaling factor (brown squares) or a linear correlation (blue triangles). The brown and blue shaded regions indicate the overall shape of the distributions.

Calculated MAE and RMSE for a variety of basis sets and DFT functionals are provided in Figure 7 and scaling factors and linear correlations that best correct the output from each method to the reference experimental spectra of **I-X** are provided in Table 2. Because it adds one degree of freedom, it is to be expected that the linear fit procedure provides a better RMSE than the scaling factor. Indeed, in all cases, the linear correlation gives significantly smaller errors than the scaling factor, which implies that the traditional procedure is not the optimal one. Three double ζ basis sets have been evaluated with the B3LYP functional. The def2-SVP basis set shows larger errors than the Pople basis set, with the best predicted frequency values being displayed when diffuse functions are included on the heavy atoms (6-31+G(d,p)). B3LYP outperforms the other functionals M06-2X, CAM-B3LYP, LC-BLYP and ω B97X-D, with LC-BLYP showing the lowest accuracy. Within the RSH class, ω B97X-D gives the smallest errors, but these are 70% larger than those for B3LYP (with a RMSE for the linear correlation of 16.6 vs 9.6 cm⁻¹).

We have compared the geometries of **I** and **V**, for which X-ray crystal structures are known,^[23] with the optimized geometries of these complexes obtained with the same functionals and basis set (see Table S5). In these cases, the best agreement is obtained with ω B97X-D, whilst B3LYP gives the largest discrepancies. This indicates that there is no clear correlation between the quality of the predicted structure and the accuracy of predicted IR spectrum.

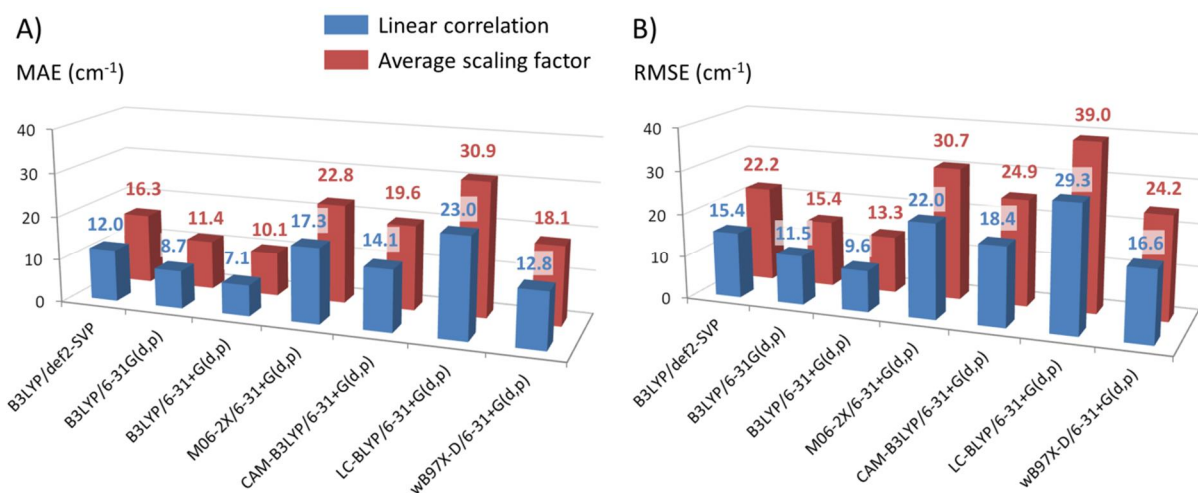


Figure 7. MAE (A) and RMSE (B) calculations between experimental and predicted vibrational frequencies at various DFT levels.

Table 2. A comparison of determined parameters to be used to computed predicted frequency values from unscaled theoretical values.

Functional	Basis set	a ^a	b ^a	R ² ^b	sf _{ave}
B3LYP	def2-SVP	1.0966	-96.9	0.9944	0.9777
B3LYP	6-31G(d,p)	1.0726	-63.0	0.9969	0.9759
B3LYP	6-31+G(d,p)	1.0647	-56.7	0.9978	0.9789
M06-2X	6-31+G(d,p)	1.1328	-125.8	0.9885	0.9643
CAM-B3LYP	6-31+G(d,p)	1.1140	-99.3	0.9919	0.9631
LC-BLYP	6-31+G(d,p)	1.1722	-146.3	0.9799	0.9430
ωB97X-D	6-31+G(d,p)	1.1201	-107.7	0.9935	0.9633

^a predicted frequencies values are obtained from the linear correlation analysis, with $\nu_{\text{predict}} = (\nu_{\text{calc}} - b)/a$; ^b linear correlation coefficient.

The combined calculated MAE and RMSE with respect to the experimental data for complexes I-X, which will be subsequently referred to as “global” MAE and RMSE, can be assumed to be an “accuracy index” of the vibrational frequencies predicted by the respective functionals. It then becomes valid to ask if these error values might be used to provide an indication of the reliability of any given calculated structure. Such a difference between the experimental and theoretically predicted spectra of other molecules can obviously arise from an intrinsic error in the functional or because an incorrect structure has been calculated. To look into this further, the “individual” RMSE’s, i.e. the RMSE for each individual complex I-X, were calculated at all DFT levels, using the linear relationship parameters (Figure 8). The lowest individual RMSE are obtained at all DFT levels for V whereas the largest individual RMSE are found for III (CAM-B3LYP), IV (B3LYP), IX (LC-BLYP) or X (M06-2X and ωB97X-D). The variation relative to the global RMSE reaches a maximum of 75% for all DFT levels, i.e. all individual RMSE of I-X lie within a range defined by [global RMSE x (1-¾); global RMSE x (1+¾)]. This indicates that if the RMSE between the experimental and theoretical spectra lie within this range of error for a given computed molecule, then the degree of variation is essentially compatible with the accuracy of the DFT method employed and the structure that has been

computed may correspond well to the experimental one. Conversely, if an individual RMSE lies outside of this margin of variation, then the molecular structure under consideration in the computational model may be erroneous and should be checked.

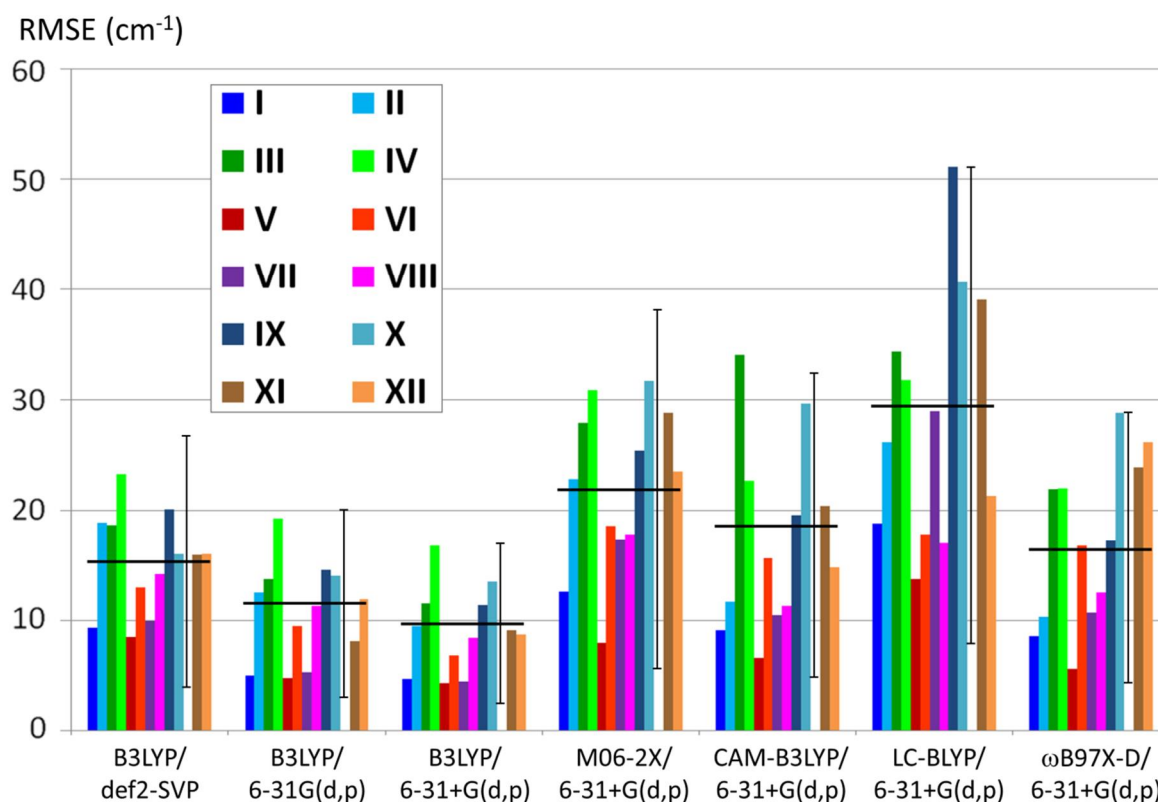


Figure 8. Computed individual RMSE based on the linear correlation analysis for each individual complex I-XII at various DFT levels. The global RMSE values for the 64 vibrational modes whole dataset (complexes I-X) are indicated by the horizontal black line. The error bar indicates the [global RMSE*(1-¼); global RMSE*(1+¾)] range.

To finish, we have studied two other compounds, namely $[\text{Zn}(\mathbf{2})_2]^{2+}$ (**XI**) and $[\text{Zn}(\mathbf{4})]^{++}$ (**XII**), in order to compare the results obtained on our database I-X to other compounds that have been studied independently. The experimental IRMPD spectrum of **XI** has been recorded under the same experimental conditions as those for I-X and is given in Figure S11. The experimental IRMPD spectrum of **XII** has been published recently^[22b] and theoretical spectra of **XI** and **XII** have been computed with the different DFT levels under investigation above. The comparison between the experimental IRMPD spectra and the predicted theoretical spectra gives rise to individual RMSE values that are shown in Figure 8. For both **XI** and **XII**, all individual RMSE are within the [global RMSE x (1-¼); global RMSE x (1+¾)] range, suggesting that the computed structure corresponds well to the experimental product. Put more clearly, even if the IR spectrum of **XII** calculated using ωB97X-D shows significant differences from compared to the experimental one, with an individual RMSE value equal to 26.2 cm⁻¹, the computed structure can be considered to be the correct one because the error value is smaller than the global RMSE defined for this method (16.6 cm⁻¹) raised by the 75% factor defined above (16.6 x (1+¾) = 29.0 > 26.2). This is in good accord with the observation that the computed structures of **XI** and **XII** corresponds to a tetrahedral zinc(II) complex and a planar di-

coordinated Zn(I) complex respectively at all DFT levels. These results therefore confirm the potential of using the relative discrepancy between the individual RMSE of any compounds and the global RMSE obtained for I-X as a confidence interval. It should be noted, however, that a small deviation does not indicate the correct structure *per se*. Several deliberately incorrect structures have been computed; unsurprisingly, the deviation relative to experiment can lie within or outside the 75% limit because similar IR spectra can be generated by different isomers or conformers. This illustrates that IR spectroscopy will not always be sufficiently discriminating to uniquely identify a precise structure.

If the data above give good corrections for the complexes under study, the referees quite reasonably ask how generally applicable the values recommended here might be for other compounds such as protonated peptides, or other classes of ions often studied by IRMPD, or how appropriate they might be for wavenumber ranges either above 2000 cm⁻¹ or significantly below 900 cm⁻¹. The precise values of a and b that might be appropriate for use outside the wavenumber ranges studied here are clearly open to question, and further experimental data will be required to answer such questions definitively. We note that scaling factors are normally used independently of the nature of the test compound, and it seems reasonable to assume that this should also be acceptable for linear correction.

Conclusion

In this work, we have explored the ability of various DFT functionals to reproduce IRMPD spectra of organometallic complexes whose structure is well-defined. For this purpose, we have recorded a set of experimental benchmark data for gas-phase IR frequencies containing 64 vibrational modes for 10 molecules. These spectra were used as reference for comparison with the spectra calculated using three basis set and five different DFT functionals. Standard B3LYP and M06-2X functionals have been tested, as well as the CAM-B3LYP, LC-BLYP and ω B97X-D RSH functionals that are important when the electronic structure of a molecule, most often a radical, is imperfectly described because of self-interaction errors. The comparison between experimental and theoretical spectra has allowed us to define scaling factors that are well adapted for gas-phase studies of this sort. Additionally, we have shown that the near-universal practice of employing scaling factors is not the most efficient for comparing experimental and theoretical spectra of this type: linear correlation analysis gives MAE and RMSE errors that are about 30% lower than those predicted using scaling factors. With the 6-31+G(d,p) basis set, the accuracy of the functionals in predicting IRMPD spectra decrease across the series: B3LYP > ω B97X-D > CAM-B3LYP > M06-2X > LC-BLYP with whole-set RMSE's of 9.6, 16.6, 18.4, 22.0 and 29.3 cm⁻¹ respectively. Calculation of individual RMSE for I-X indicates relatively dispersed results with an error margin of $\pm 75\%$ relative to the global RMSE. A study of two other complexes confirms that this deviation from the global RMSE may be used as a confidence interval for the calculated structure, whose criteria are that if the deviation of the individual RMSE relative to the global RMSE is less than 75% of the latter, the studied structure can be considered relevant. For a larger difference, the structure should probably be rejected. Further work is in progress in our laboratories to extend these results to a broader range of compounds.

Experimental Section

Complexes synthesis and sample preparation: $\text{Zn}(\text{BF}_4)_2$ hydrate (Strem Chemicals) was used without further purification.

Ligands **2** and **3** were prepared according to related published procedures.^[26]

$\text{Ru}(\text{bipy})_2\text{Cl}_2$ was prepared according to the synthesis described by Meyer et al.^[27]

$\text{Ru}(\text{bipy})_2\text{Cl}_2$ (1 eq.) and AgBF_4 (2 eq.) were mixed in acetonitrile and heated at 60°C overnight, the mixture was filtrated and then **3** was added to the filtrated solution and stirred for 2h. **IX** is observed in the full mass spectra obtained from the diluted crude solution.

$\text{Ru}(\text{bipy})_2\text{Cl}_2$ complex was stirred for 2h in acetonitrile solution. **X** is observed in the full mass spectra obtained from the diluted crude solution.

Solutions for zinc complexes **I-VIII** and **XI** were prepared in a 1:2 water/acetonitrile solution by mixing one or two ligands and zinc(II) tetrafluoroborate.

General procedure for IRMPD experiments: Zinc and ruthenium complexes were formed in an electrospray source in positive mode and observed by high resolution mass spectrometry with a 7T Fourier Transform Ion Cyclotron Resonance (FT-ICR) tandem mass spectrometer (Bruker Apex Qe). The ion of interest was isolated in the quadrupole with an isolation window of m/z 0.5 to 10 and accumulated in the collision cell for between 0.1 and 2 seconds. This ion is then transferred to the ICR Cell and irradiated for between 0.15 and 2.0 seconds with an IR laser. Vibrational spectra were obtained in the 900-2000 cm^{-1} region by IRMPD using the FT-ICR spectrometer coupled to a tunable free electron laser at the Centre Laser Infrarouge d'Orsay (CLIO).^[28] In IRMPD action spectroscopy, when the laser wavelength becomes resonant with a vibrational transition of the mass-selected species, a sequential absorption of multiple photons that is coupled to fast intramolecular vibrational redistribution can deposit internal energy within the species up to threshold energy for fragmentation. The intensity of parent and fragment ions after laser irradiation are monitored as a function of the excitation wavelength, with the photofragmentation yield being calculated for each wavelength according to $-\ln[I_{\text{parent}}/(I_{\text{parent}} + \sum I_{\text{fragments}})]$. At each wavelength step, 2 to 8 mass spectra were averaged. Experimental details for complexes **I-XI** are given in the SI.

The bandwidth of the IR laser is of the order of 0.5% of the central wavelength.^[29] The wavelength stepping used for acquisition is between 3 and 5 cm^{-1} . The experimental wavenumbers have been calibrated based on the spectrum of a polystyrene film. Then, experimental peak positions have been determined based on the maximal signal amplitude, which was obtained from an average over three consecutive values.

Computational details for DFT calculations: Calculations were carried out with the Gaussian09 package^[30] and all structures were fully optimized without any symmetry constraints at the DFT level. Five different exchange-correlation functionals, which cover different categories from "conventional" to range-separated hybrid (RSH) functionals, have been applied for these calculations: (i) the hybrid GGA functional B3LYP,^[31] which includes 20% of Hartree-Fock exchange (X^{HF}); (ii) the hybrid meta-GGA functional M06-2X,^[32] which includes 54% of X^{HF} ; (iii) the RSHs CAM-B3LYP,^[33] LC-BLYP,^[34] and ω B97X-D,^[35] which present a growing fraction of X^{HF} with increasing interelectronic separation in the 19-65, 0-100 and 22.2-100 range, respectively. The speed of the transition from Kohn-Sham exchange to X^{HF} is governed by the attenuation parameter ω whose value is 0.33, 0.47 and 0.20 a.u. for these three functionals, respectively. Three valence double- ζ basis set have been used. The first one, noted def2-SVP, used the def2-SVP split-valence basis set developed by Weigend and

Ahlrichs.^[36] The second and third basis set, noted respectively 6-31G(d,p) and 6-31+G(d,p), used valence double- ζ Pople basis set,^[37] without or with diffuse function on heavy atoms.^[38] These basis set were used for all atoms but Ru for which the LanL2DZ pseudopotential and associated basis set^[39] has been employed. For each stationary point, we carried out a vibrational frequency calculation at the same level to characterize their nature as minima. Stick spectra, as well as Lorentzian line shapes, with a full width at half maximum (fwhm) of 5 or 10 cm⁻¹, were used to generate the calculated IR spectra.

Analysis tools: For each level of theory and each protocol used to obtained unscaled theoretical vibrational frequencies, the predicted frequency values are obtained following two various strategies. First, we used the traditional approach which consists to determine (or select in the literature) a scaling factor and to multiply the unscaled theoretical vibrational frequencies by this scaling factor (*vide infra*). In a second strategy, the unscaled theoretical vibrational frequencies are plotted against the determined experimental vibrational frequencies for the 68 vibrational modes ν . A linear correlation is then used to determine a linear relationship ($y = ax + b$) between experimental and theoretical values. The equation (1) produces the predicted frequency values $\nu_{\text{predict}}^{\nu}$ from the unscaled theoretical vibrational frequencies ν_{calc}^{ν} .

$$\nu_{\text{predict}}^{\nu} = \frac{\nu_{\text{calc}}^{\nu} - b}{a} \quad (1)$$

The accuracy of the predicted frequency values is analyzed by two numerical tests. First we use the mean absolute error (MAE) given by equation (2).

$$\text{MAE} = \frac{\sum |\nu_{\text{predict}}^{\nu} - \nu_{\text{exp}}^{\nu}|}{68} \quad (2)$$

Second, the root-mean-square error (RMSE, equation (3)) between the experimental and predicted frequency values has been calculated to evaluate the prediction capabilities.

$$\text{RMSE} = \sqrt{\frac{\sum (\nu_{\text{predict}}^{\nu} - \nu_{\text{exp}}^{\nu})^2}{68}} \quad (3)$$

Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors compared to MAE.

Three different scaling factors have been estimated for each level of theory and each protocol used to obtained unscaled theoretical vibrational frequencies. First, Individual scaling factors sf^{ν} have been obtained for each of the 68 vibrational modes ν by dividing the experimental value ν_{exp}^{ν} by the corresponding calculated value ν_{calc}^{ν} . The arithmetic averaging of the individual scaling factors is used

to determine the “average” scaling factor $sf_{ave} = (\sum sf)/68$. Scaling factors which minimize the MAE (sf_{MAE}) or the RMSE (sf_{RMSE}) have also being obtained.

Acknowledgements

This work was performed using HPC resources from GENCI-CINES/IDRIS (Grant 2015-086894). MK thanks the Ecole polytechnique for stipend. Financial support from the National FT-ICR network (FR 3624 CNRS) for conducting the research is gratefully acknowledged.

Keywords: Mass spectrometry • Density functional calculations • IR spectroscopy • Organometallic complexes • DFT benchmark

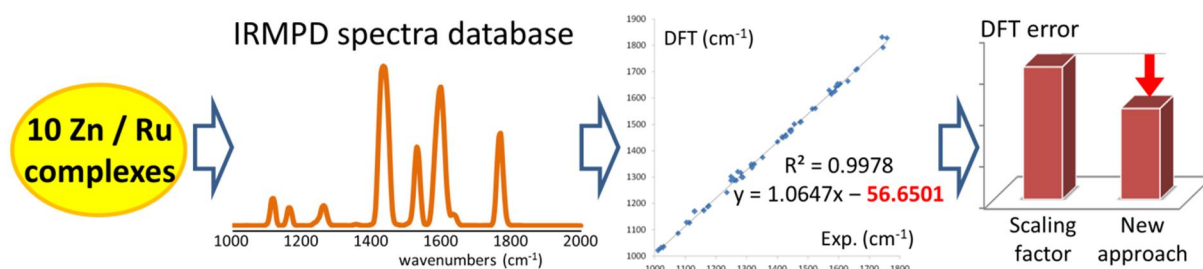
References

- [1] J. R. Eyler, *Mass Spectrom. Rev.* **2009**, *28*, 448-467.
- [2] a) J. Oomens, A. J. A. van Roij, G. Meijer, G. von Helden, *Astrophys. J.* **2000**, *542*, 404-410; b) R. C. Dunbar, *Int. J. Mass Spectrom.* **2015**, *377*, 159-171.
- [3] a) N. R. Walker, R. S. Walters, M. A. Duncan, *New J. Chem.* **2005**, *29*, 1495-1503; b) L. MacAleese, P. Maître, *Mass Spectrom. Rev.* **2007**, *26*, 583-605; c) N. C. Polfer, J. Oomens, *Mass Spectrom. Rev.* **2009**, *28*, 468-494; d) T. D. Fridgen, *Mass Spectrom. Rev.* **2009**, *28*, 586-607; e) N. C. Polfer, *Chem. Soc. Rev.* **2011**, *40*, 2211-2221; f) J. Roithova, *Chem. Soc. Rev.* **2012**, *41*, 547-559; g) K. R. Asmis, D. M. Neumark, *Acc. Chem. Res.* **2012**, *45*, 43-52; h) A. L. Patrick, N. C. Polfer, *Top. Curr. Chem.* **2015**, *364*, 153-182.
- [4] O. Hernandez, B. Paizs, P. Maitre, *Int. J. Mass Spectrom.* **2015**, *377*, 172-178.
- [5] a) A. F. DeBlase, S. Bloom, T. Lectka, K. D. Jordan, A. B. McCoy, M. A. Johnson, *J. Chem. Phys.* **2013**, *139*, 024301:1-9; b) F. Turecek, C. L. Moss, I. Pikalov, R. Pepin, K. Gulyuz, N. C. Polfer, M. F. Bush, J. Brown, J. Williams, K. Richardson, *Int. J. Mass Spectrom.* **2013**, *354-355*, 249-256; c) O. Hernandez, P. Pulay, P. Maitre, B. Paizs, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1511-1514.
- [6] a) D. Semrouni, A. Sharma, J.-P. Dognon, G. Ohanessian, C. Clavaguéra, *J. Chem. Theory Comput.* **2014**, *10*, 3190-3199; b) F. Thauay, J.-P. Dognon, G. Ohanessian, C. Clavaguéra, *Phys. Chem. Chem. Phys.* **2015**, *17*, 25968-25977.
- [7] M. P. Gaigeot, M. Martinez, R. Vuilleumier, *Mol. Phys.* **2007**, *105*, 2857-2878.
- [8] K. Yagi, T. Taketsugu, K. Hirao, M. S. Gordon, *J. Chem. Phys.* **2000**, *113*, 1005-1017.
- [9] V. Barone, *J. Chem. Phys.* **2005**, *122*, 014108: 1-10.
- [10] a) A. P. Scott, L. Radom, *J. Phys. Chem.* **1996**, *100*, 16502-16513; b) M. W. Wong, *Chem. Phys. Lett.* **1996**, *256*, 391-399; c) L. A. Curtiss, P. C. Redfern, K. Raghavachari, J. A. Pople, *J. Chem. Phys.* **2001**, *114*, 108-117; d) P. Sinha, S. E. Boesch, C. M. Gu, R. A. Wheeler, A. K. Wilson, *J. Phys. Chem. A* **2004**, *108*, 9213-9217; e) M. P. Andersson, P. Uvdal, *J. Phys. Chem. A* **2005**, *109*, 2937-2941; f) Y. Tantirungrotechai, K. Phanasant, S. Roddecha, P. Surawatanawong, V. Sutthikhum, J. Limtrakul, *J. Mol. Struct. : THEOCHEM* **2006**, *760*, 189-192; g) J. P. Merrick, D. Moran, L. Radom, *J. Phys. Chem. A* **2007**, *111*, 11683-11700; h) I. M. Alecu, J. Zheng, Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2010**, *6*, 2872-2887; i) M. L. Laury, M. J. Carlson, A. K. Wilson, *J. Comput. Chem.* **2012**, *33*, 2380-2387.
- [11] J. Jaramillo, G. E. Scuseria, *Chem. Phys. Lett.* **1999**, *312*, 269-276.
- [12] C. W. Bauschlicher Jr, H. Partridge, *J. Chem. Phys.* **1995**, *103*, 1788-1791.
- [13] a) K. K. Irikura, R. D. Johnson III, R. N. Kacker, *J. Phys. Chem. A* **2005**, *109*, 8430-8437; b) R. D. Johnson III, K. K. Irikura, R. N. Kacker, R. Kessel, *J. Chem. Theory Comput.* **2010**, *6*, 2822-2828.

- [14] a) G. Maes, J. Smets, L. Adamowicz, W. McCarthy, M. K. Van Bael, L. Houben, K. Schoone, *J. Mol. Struct.* **1997**, 410-411, 315-322; b) P. Borowski, M. Fernandez-Gomez, M. P. Fernandez-Liencres, T. P. Ruiz, *Chem. Phys. Lett.* **2007**, 446, 191-198.
- [15] M. D. Halls, J. Velkovski, H. B. Schlegel, *Theor. Chem. Acc.* **2001**, 105, 413-421.
- [16] a) M. A. Palafox, *Int. J. Quantum Chem.* **2000**, 77, 661-684; b) M. A. Palafox, N. Iza, M. Gil, *J. Mol. Struct. : THEOCHEM* **2002**, 585, 69-92; c) M. A. Palafox, V. K. Rastogi, L. Mittal, *Int. J. Quantum Chem.* **2003**, 94, 189-204; d) M. A. Palafox, M. Gill, N. J. Nunez, V. K. Rastogi, L. Mittal, R. Sharma, *Int. J. Quantum Chem.* **2005**, 103, 394-421.
- [17] Y. Bouteiller, J. C. Gillet, G. Grégoire, J. P. Schermann, *J. Phys. Chem. A* **2008**, 112, 11656-11660.
- [18] a) Y. Bouteiller, J. C. Pouilly, C. Desfrancois, G. Grégoire, *J. Phys. Chem. A* **2009**, 113, 6301-6307; b) D. Semrouni, C. Clavaguéra, J. P. Dognon, G. Ohanessian, *Int. J. Mass Spectrom.* **2010**, 297, 152-161; c) A. Sharma, G. Ohanessian, C. Clavaguéra, *J. Mol. Model.* **2014**, 20, 2426.
- [19] T. Tsuneda, K. Hirao, *WIREs Comput. Mol. Sci.* **2014**, 4, 375-390.
- [20] a) A. I. Gilson, G. van der Rest, J. Chamot-Rooke, W. Kurlancheek, M. Head-Gordon, D. Jacquemin, G. Frison, *J. Phys. Chem. Lett.* **2011**, 2, 1426-1431; b) V. Riffet, D. Jacquemin, E. Cauët, G. Frison, *J. Chem. Theory Comput.* **2014**, 10, 3308-3318; c) V. Riffet, D. Jacquemin, G. Frison, *Int. J. Mass Spectrom.* **2015**, 390, 28-38.
- [21] a) R. A. Zubarev, *Mass Spectrom. Rev.* **2003**, 22, 57-77; b) H. J. Cooper, K. Hakansson, A. G. Marshall, *Mass Spectrom. Rev.* **2005**, 24, 201-222; c) L. M. Mikesch, B. Ueberheide, A. Chi, J. J. Coon, J. E. P. Syka, J. Shabanowitz, D. F. Hunt, *Biochim. Biophys. Acta* **2006**, 1764, 1811-1822; d) K. O. Zhurov, L. Fornelli, M. D. Wodrich, Ü. A. Laskay, Y. O. Tsybin, *Chem. Soc. Rev.* **2013**, 42, 5014-5030.
- [22] a) G. Frison, G. van der Rest, F. Turecek, T. Besson, J. Lemaire, P. Maitre, J. Chamot-Rooke, *J. Am. Chem. Soc.* **2008**, 130, 14916-14917; b) M. Katari, E. Payen de la Garanderie, E. Nicol, V. Steinmetz, G. van der Rest, D. Carmichael, G. Frison, *Phys. Chem. Chem. Phys.* **2015**, 17, 25689-25692; c) J. Martens, J. Grzetic, G. Berden, J. Oomens, *Nat. Commun.* **2016**, 7, 11754; d) C. J. Shaffer, J. Martens, A. Marek, J. Oomens, F. Turecek, *J. Am. Soc. Mass Spectrom.* **2016**, 27, 1176-1185.
- [23] a) G. H. Eom, H. M. Park, M. Y. Hyun, S. P. Jang, C. Kim, J. H. Lee, S. J. Lee, S. J. Kim, Y. Kim, *Polyhedron* **2011**, 30, 1555-1564; b) E. Krausz, H. Riesen, A. D. Rae, *Aust. J. Chem.* **1995**, 48, 929-954; c) X.-M. Chen, R.-Q. Wang, X.-L. Yu, *Acta Cryst.* **1995**, C51, 1545-1547; d) J. Breu, H. Domel, A. Stoll, *Eur. J. Inorg. Chem.* **2000**, 2401-2408; e) C. C. Evans, R. Masse, J.-F. Nicoud, M. Bagieu-Beucher, *J. Mater. Chem.* **2000**, 10, 1419-1423; f) R. Sieber, S. Decurtins, H. Stoeckli-Evans, C. Wilson, D. Yufit, J. A. K. Howard, S. C. Capelli, A. Hauser, *Chem. Eur. J.* **2000**, 6, 361-368; g) J.-B. Shen, X. Lv, J.-F. Chen, Y.-F. Zhou, G.-L. Zhao, *Acta Cryst.* **2011**, E67, m186-m187; h) M.-C. Yin, L.-J. Yuan, C.-C. Ai, C.-W. Wang, E.-T. Yuan, J.-T. Sun, *Polyhedron*, **2004**, 23, 529-536; i) X. Fan, C. Wang, J. Li, G. Zhang, *Z. Anorg. Allg. Chem.* **2015**, 641, 1169-1175; j) W. Yang, S. Dang, H. Wang, T. Tian, Q.-J. Pan, Z.-M. Sun, *Inorg. Chem.* **2013**, 52, 12394-12402.
- [24] These four bands are **III_e**, **III_g**, **VII_i** and **VII_j**. It should be noted that including these four bands only negligibly modifies the values calculated here, and does not change the conclusions in any way.
- [25] a) R. G. Inskeep, *J. Inorg. Nucl. Chem.* **1962**, 24, 763-776; b) A. A. Schilt, R. C. Taylor, *J. Inorg. Nucl. Chem.* **1959**, 9, 211-221.

- [26] a) G. D. Smith, K. A. Maxwell, J. M. DeSimone, T. J. Meyer, R. A. Palmer, *Inorg. Chem.* **2000**, *39*, 893-898; b) C. M. Elliot, E. J. Hershenhart, *J. Am. Chem. Soc.* **1982**, *104*, 7519-7526.
- [27] B. P. Sullivan, D. J. Salmon, T. J. Meyer, *Inorg. Chem.* **1978**, *17*, 3334-3341.
- [28] R. Prazeres, F. Glotin, C. Insa, D. A. Jaroszynski, J. M. Ortega, *Eur. Phys. J. D* **1998**, *3*, 87-93.
- [29] M. Berdakin, V. Steinmetz, P. Maitre, G. A. Pino, *Phys. Chem. Chem. Phys.* **2015**, *17*, 25915-25924.
- [30] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, **2013**, Gaussian 09 (Revision D.01).
- [31] a) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648-5652; b) A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098-3100; c) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785-789.
- [32] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215-241.
- [33] T. Yanai, D. Tew, N. Handy, *Chem. Phys. Lett.* **2004**, *393*, 51-57.
- [34] H. Likura, T. Tsuneda, T. Yanai, K. Hirao, *J. Chem. Phys.* **2001**, *115*, 3540-3544.
- [35] J. D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615-6620.
- [36] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
- [37] a) W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257-2261; b) M. M. Francel, W. J. Pietro, W. J. Hehre, J. S. Binkley, D. J. DeFrees, J. A. Pople, M. S. Gordon, *J. Chem. Phys.* **1982**, *77*, 3654-3665; c) V. A. Rassolov, J. A. Pople, M. A. Ratner, T. L. Windus, *J. Chem. Phys.* **1998**, *109*, 1223-1229.
- [38] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. v. R. Schleyer, *J. Comp. Chem.* **1983**, *4*, 294-301.
- [39] P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 270-283.

Entry for the Table of Contents



64 experimental vibrational modes, from 10 organometallic complexes, have been measured and used as reference database against which DFT methods have been evaluated. The use of a linear correlation instead of a scaling factor significantly improves the prediction accuracies of all functionals.