Published in final edited form as:

J Chem Theory Comput. 2020 November 10; 16(11): 7044–7060. doi:10.1021/acs.jctc.0c00877.

Comprehensive Assessment of GFN Tight-Binding and Composite Density Functional Theory Methods for Calculating Gas-Phase Infrared Spectra

Philipp Pracht,

Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

David F. Grant,

Department of Pharmaceutical Sciences, University of Connecticut, Storrs, Connecticut 06268, United States

Stefan Grimme

Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, 53115 Bonn, Germany

Abstract

Vibrational spectroscopy is a valuable and widely used analytical tool for the characterization of chemical substances. We investigate the performance of semiempirical quantum mechanical GFN tight-binding and force-field methods for the calculation of gas-phase infrared spectra in comparison to experiment and low-cost (B3LYP-3c) density functional theory. A data set of 7247 experimental references was used to evaluate method performance based on automatic spectra comparison. Various quantitative spectral similarity measures were employed for the comparison between theory and experiment and for determining empirical scaling factors. It is shown that the scaling of atomic masses provides an accurate yet simple alternative to standard global frequency scaling in density functional theory (DFT) and semiempirical calculations. Furthermore, the method performance for 58 exemplary transition metal complexes was investigated. The efficient DFT composite method B3LYP-3c, that was introduced in the course of this work, was found to be excellently suited for general IR spectra calculations. The GFN1- and GFN2-xTB tight-binding methods clearly outperformed the PMx competitors. Conformational changes were investigated for a subset of the data and are found to have a mediocre strong influence on the simulated spectra suggesting that the corresponding elaborate sampling steps may be neglected in automated compound identification workflows.

Corresponding Author Phone: +49-228/73-2351; grimme@thch.unibonn.de.

The authors declare no competing financial interest.

Supporting research data of the newspecmatch source code for this article may be accessed at our GitHub page (<https://github.com/grimme-lab/newspecmatch>). In the same repository, B3LYP-3c geometries and calculated harmonic vibrational frequencies are available. Experimental spectra in the JCAMP-DX format can be obtained from the NIST database (holder of the copyright).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c00877>.

Computed similarity measures for all molecules at all discussed theory levels (ZIP)

Graphical Abstract

1. INTRODUCTION

Rotational–vibrational spectroscopy is one of the most common analytical tools for the characterization of chemical substances.¹ The analysis of experimental infrared (IR) or Raman spectra gives detailed insight into molecular structure and can be used, e.g., for analytical compound identification purposes.^{2,3} Theoretical, mainly quantum chemistry calculations of vibrational spectra including assignment of the involved normal modes have a long and very successful history.^{4,5} Vibrational spectra can be calculated by various classical atomistic and quantum chemical (QM) methods, of which density functional theory (DFT) is the most common today. Still, even with modern theoretical and technological advances, the calculation of vibrational frequencies can be prohibitively expensive for moderately sized (100–200 atoms) molecules. Accurate and fast calculation of IR spectra and the related thermostatical free energy corrections at a reduced theoretical level (see, e.g., ref 6) is therefore desirable. Progress has been made in this field with classical force fields (FF), but their errors are much larger and less systematic than at a QM level.^{7,8} Semiempirical quantum chemical methods (SQM) can provide a good alternative because they are considered to bridge the gap between FFs and QM, both in terms of computational cost and accuracy. In particular, methods of the recently developed GFN family⁹ are promising candidates since they are designed to yield good geometries, frequencies, and noncovalent interaction energies.

Vibrational frequencies are commonly calculated within the harmonic approximation and depend on the second derivatives of the energy with respect to atomic displacements and additionally on the atomic masses. The comparison of harmonic theoretical with experimental (fundamental) frequencies leads to systematic errors also in higher rung DFT treatments.^{10,11} A typical countermeasure is to linearly scale the harmonic vibrational frequencies (HVs) obtained from QM (or FF) calculations, where different levels of theory (e.g., DFT functionals, basis sets) require varying scaling factors.^{4,11–13} In the case of hybrid functional DFT, the HVF error can be attributed mainly to the employed Fock exchange, which consistently overestimates frequency values. For FF or SQM, the errors are less systematic, and global frequency scaling works less well than in (hybrid) DFT so

that other means to improve the accuracy have been investigated.^{14–16} Introducing more adjustable parameters provides a higher degree of empirical freedom, but also many more data points are required for a robust fit. Multiparameter frequency scaling methods have been proposed in the past and yield more accurate results than simple linear and uniform global scaling.^{15,17–23} More recently, machine learning approaches have successfully been combined with molecular dynamics and QM calculations for the computation of IR spectra and anharmonic frequency corrections.^{24,25}

Due to the intrinsic relation of molecular vibrational motions to the involved atomic masses, the latter could also be used as a parameter in calculations. Here, we explore the scaling of atomic masses as an alternative to the standard global scaling of the HVFs. Atomic mass scaling has been introduced by Irikura²⁶ for diatomics and is related to the approach of Pulay and co-workers, who proposed an internal coordinate specific scaling of force constant matrix elements instead of global frequency scaling.^{27,28}

In this work, we used a data set of 7247 experimental gas-phase IR spectra (obtained from the NIST chemical database²⁹) to evaluate the performance of SQM methods of the GFN family (GFN1- and GFN2-xTB)^{30,31} for IR spectra calculations. In addition, the performance of the recently introduced, nonelectronic variant GFN-FF is investigated as a representative of an advanced classical FF.³² This involves the introduction of a similarity measure for the quantitative comparison between experimental and theoretical spectra. Furthermore, a low-cost composite DFT method based on the B3LYP density functional³³ similar to the well established PBEh-3c³⁴ composite method is proposed and tested. For recent studies with our GGA-based low-cost method B97–3c³⁵ for IR spectra calculations, see refs 36 and 37.

In a recent study,¹⁶ Henschel et al. pursued a similar approach to evaluate the performance of several FFs for the calculation of IR spectra. In their work, two different metrics (Pearson correlation coefficient and Spearman rank correlation) were used for the evaluation. For comparability, the same spectral similarity scores were used here, plus two metrics (match score, Euclidian norm) we originally based this work on. In the literature, other measures have been proposed and adapted for various purposes.^{38,39} The different similarity scores have different advantages and disadvantages, but it has been suggested¹⁶ that they can be used complementarily. Still it remains an open question whether all of the similarity measures mentioned above are equally suitable for the quantitative comparison of IR spectra. The long-term goal of this work is to provide the theoretical basis for automated high-throughput workflows in analytical chemistry (unknown compound identification) applications.

2. THEORY

2.1. Calculation of IR Spectra.

By a Taylor expansion of a general molecular potential energy surface, vibrational motions can be related to force constants. Force constants are the second derivatives of the total energy with respect to nuclear displacements.⁴⁰ In *ab initio* (or semiempirical) calculations, vibrational frequencies and the corresponding eigenfunctions (normal modes) are most

commonly obtained within the harmonic approximation, where the force constants are contained in the mass-weighted Hessian matrix

$$\mathbf{F}_{ij}^{(m)} = \frac{1}{\sqrt{m_i m_j}} \left(\frac{\partial^2 E}{\partial R_i \partial R_j} \right) \quad (1)$$

In eq 1, E is the total molecular energy, R_{ij} is the displacement in Cartesian space, and m_{ij} are the masses of nuclei i and j , respectively. The second derivatives may be obtained analytically or numerically from the first-order derivative, i.e., the nuclear gradient. Diagonalization of the matrix $\mathbf{F}^{(m)}$ according to

$$\mathbf{F}^{(m)} \mathbf{Q} = \epsilon \mathbf{Q} \quad (2)$$

yields the diagonal matrix ϵ that contains the eigenvalues of $\mathbf{F}^{(m)}$, and $\mathbf{Q}$ are the normal modes in Cartesian space. The vibrational frequencies ν are given by

$$\nu_p = \frac{1}{2\pi} \sqrt{\epsilon_p} \quad (3)$$

for mode p .

The IR absorption intensity for each fundamental transition is associated with the change of the molecular dipole moment μ along the normal mode coordinate q_i^{41} with the leading term $\frac{\partial \mu}{\partial q_i}$. Very accurate calculations of intensities require knowledge also of higher order terms,⁴² however, they are approximately (the so-called double harmonic approximation) proportional to the squared derivative $\left(\frac{\partial \mu}{\partial q_i}\right)^2$ ⁴³ which is used in this work. Since intensities depend on the quality of the calculated dipole moments, their description is rather sensitive to the employed level of theory (e.g., the quality of the AO basis set), and several approaches for the improved calculation of IR intensities exist.^{44–46} In particular, the mode localization and intensity tracking approaches pioneered by Reiher and coworkers^{47–50} seem to be capable tools for the improved calculation and interpretation of vibrational spectra. However, the fine details of IR spectra calculations are not the subject of the present work. Instead, we focus here on a reasonably accurate and straightforward description of the more typical case of a low-resolution/large molecule.

2.2. Comparing IR Spectra.

In general, a molecule with n nuclei has $3n - 6$ normal modes ($3n - 5$ for linear molecules) which, in principle, can be compared with an experimental reference IR spectrum. For the latter, fundamental frequencies need to be extracted and assigned to the respective theoretical modes. This can become tedious for medium sized molecules where modes in the *fingerprint* regime, $< 1500 \text{ cm}^{-1}$ often have small intensities and strongly overlap due to their line width. To evaluate larger compounds and a large number of spectra, it is therefore neither sufficient nor efficient to compare only selected fundamental frequencies. Instead, metrics for a quantitative comparison of two (entire) spectra have to be developed and tested.

Experimental spectra typically consist of a set of k equidistant frequency data points with respective intensities. In order to be comparable, the spectra need to be normalized. Different procedures are possible, but in this work, all spectra will be normalized according to

$$\sqrt{\sum_i^k u_i} du \stackrel{!}{=} 1 \quad (4)$$

where u can be considered as a k -dimensional vector, u_i is the intensity of the i -th point in the spectrum, and du is the distance between the points u_i and u_{i+1} (here set to 1.0 cm^{-1}), i.e., the normalization is done via a Riemann integration. The square-root dependence in eq 4 was chosen to increase the weight of small intensities relative to larger intensities. Normalizations to single signals, e.g., the largest peak of the spectrum, should be avoided because relative intensities and frequencies are strongly dependent on the theoretical level, and hence it cannot always be ensured that the same peak is selected. For theoretical spectra, the frequencies and intensities of the vibrational modes are available as isolated signals (“*stick spectrum*”) and first have to be expanded to the same spectral domain as the experimental data. This is achieved by employing a Lorentzian line shape function for each mode

$$\phi_p(\nu) = I_p \left(1 + \frac{\nu_p - \nu}{0.5w} \right)^{-1} \quad (5)$$

where ν_p is the position (calculated frequency) of the mode p , I_p is its intensity, and w is the full width at half-maximum (fwhm). Typical values employed for the fwhm in theoretical studies range from 20 to 40 cm^{-1} , whereas the average line width in experimental spectra was determined as 24 cm^{-1} .⁵¹ The simulated spectrum is then simply given by the sum of all the Lorentzian functions over the modes

$$\Phi_{\text{norm}}(\nu) = I_{\text{norm}} \sum_p^{N_p} \phi_p(\nu) \quad (6)$$

with the normalization constant I_{norm} . Analogously to eq 4, the spectrum is normalized by $\sqrt{\int \Phi_{\text{norm}} d\nu} \stackrel{!}{=} 1$ and in this form is directly (point wise) comparable to experimental data. Hence, in order to match experimental and theoretical spectra, a vector v with the same number of points as u is constructed from Φ_{norm} .

In this study, four different spectral similarity measures have been investigated. The first is a simple match score (r_{match})

$$r_{\text{match}} = \frac{\left(\sum_i^k u_i v_i \right)^2}{\left(\sum_i^k u_i^2 \right) \left(\sum_i^k v_i^2 \right)} \quad (7)$$

where u and v are the k -dimensional vectors obtained for the two compared spectra. The r_{match} corresponds to a Cauchy-Schwarz inequality in $\mathbb{R}^k$ dimensional Euclidian space which

essentially is a simplified overlap. r_{match} values range from 0 to 1, where unity denotes a perfect match.

The second measure used is the Euclidean norm (r_{Euclid})

$$r_{Euclid} = \left(1.0 + \frac{\sum_i^k (u_i - v_i)^2}{\sum_i^k (v_i)^2} \right)^{-1} \quad (8)$$

In the literature, two additional measures have been employed for IR and Raman spectra comparisons.^{52–55} One is the Pearson correlation coefficient ($r_{Pearson}$), which is similar to the r_{match}

$$r_{Pearson} = \frac{\sum_i^k (u_i - \bar{u})(v_i - \bar{v})}{\sqrt{\sum_i^k (u_i - \bar{u})^2} \sqrt{\sum_i^k (v_i - \bar{v})^2}} \quad (9)$$

with the mean values $\bar{u}$ and $\bar{v}$ for u and v . Both the r_{match} and $r_{Pearson}$ are linear correlation measures that are based on the Cauchy-Schwarz inequality. The other (nonlinear) measure that has often been employed is the Spearman rank correlation coefficient ($r_{Spearman}$)

$$r_{Spearman} = 1.0 - \frac{6 \sum_i^k (rg(u_i) - rg(v_i))^2}{k(k^2 - 1)} \quad 10$$

where $rg(u_i)$ and $rg(v_i)$ are the respective ranks of u_i and v_i .

Note that if fundamental frequencies can be clearly identified in the experimental spectra, one is able to create a synthetic (noiseless) spectrum similar to the theoretical spectrum using eq 6 for the comparison. This has been tested in the course of our work. However, automated identification of the peaks in experimental spectra is challenging for molecules with many normal modes and prone to errors due to experimental noise. Therefore, using the measured experimental data points as a reference seems to be a more robust and general approach. If required, additional values can be estimated by linear interpolation between points u_i and u_{i+1} .

3. COMPUTATIONAL DETAILS

Calculations at the GFN-FF, GFN1-xTB, and GFN2-xTB levels were performed with version 6.3.2 of the xtb code which was also used as a wrapper for numerical Hessian calculations with PM6-D3H4,^{56,57} PM6-D3H4X,⁵⁸ and PM7,⁵⁹ using the MOPAC program (version 19.179L). DFT calculations at the B3LYP-3c level were conducted with the Turbomole program package (version 7.4.1).⁶⁰ The term B3LYP-3c stands for a standard self-consistent B3LYP-D3(BJ)^{ATM}-gCP/def2-mSVP^{33,34,61–63} calculation, i.e., the method employs the well-known B3LYP exchange-correlation functional with a small (adapted) double- ζ Ahlrichs type Gaussian AO basis set, the standard D3 dispersion correction with a three-body term added, and a geometrical correction for BSSE (gCP). This composite approach has been benchmarked in the course of this work on subsets of

the GMTKN55 meta-database⁶⁴ and found to perform similarly for thermochemistry and noncovalent interactions as the PBEh-3c parent method. A similar composite scheme based on the B3LYP functional has already been proposed for computations of molecular thermochemistry.⁶⁵ As opposed to PBEh-3c, however, the functional (GGA) part has not been modified in B3LYP-3c. D3(BJ)^{ATM} and gCP are additive corrections that are implemented in the Turbomole program packages (version 7.4 and higher) and can be requested with the \$gcp dft/sv(p) and \$disp3 -bj -abc keywords. Similar usage options exist, e.g., for the ORCA program package.⁶⁶ The composition of the def2-mSVP basis set is discussed in ref 34.

The DFT calculations in Turbomole were conducted with an *m4* grid, using the RI approximation and matching auxiliary basis functions. The xtb, Turbomole, and ORCA program packages allow a modification of atomic masses via the input files. For the metal-containing complexes in section 4.4, reference HVFs were calculated at the B3LYP-D4/def2-TZVPP^{67,68} level using the RIJCOSX approximation⁶⁹ (with keywords *gridx7* and *nofinalgridx*) throughout. These calculations were performed using the ORCA program (version 4.2).⁶⁶ All structures were optimized using the xtb program as a driver for the ANCOPT optimizer with *normal* convergence criteria. The various similarity measures were calculated with a small standalone code called newspecmatch written in Fortran. For the fwhm, a value of 30 cm⁻¹ was chosen. From ref 16 it is known that optimization of the fwhm with respect to an experimental-theoretical best match yields larger values (>60 cm⁻¹). This is physically not plausible and seems to be an artifact of errors in the theoretical frequencies. A width of 30 to 35 cm⁻¹ is physically realistic but still provides some leeway for slightly larger frequency deviations between the spectra.

6556 experimental gas-phase IR spectra and corresponding molecule structures were obtained from the NIST database,²⁹ containing combinations of the elements H, C, N, O, F, Cl, and Br. We will refer to this simply as the “HCNO” set. Ionic structures were excluded. Furthermore, 538 sulfur-, 100 phosphorus-, and 53 silicon-containing structures and their associated gas-phase IR spectra were also obtained from the NIST database. The molecules are mainly “organic” but contain various, sometimes rather complex, inorganic parts. Because very few experimental spectra were available for transition metal complexes, HVFs were calculated for 58 species taken from the TMG145 database,⁷⁰ and the high B3LYP-D4/def2-TZVPP level was taken as the reference for the tested SQM and FF methods.

4. RESULTS AND DISCUSSION

4.1. General Performance and Global Scaling of Frequencies.

It is standard practice in computational chemistry to scale the calculated HVFs by a method dependent parameter, that is typically determined by minimization of the root-mean-square deviation (RMSD) between experimental fundamental frequencies and theoretical HVFs.^{11,13} Special scale factors exist also for thermochemical applications.^{12,71} A fitting strategy based on the maximization of the $r_{Pearson}$ and $r_{Spearman}$ was reported in ref 16. In contrast to QM or SQM methods, force field methods already include (by fitting to experimental data) some empirical adjustment, and global scaling of the frequencies has

very little impact on overall performance.¹⁶ The tested GFN-FF employs a default global scaling factor of 1.03 which was adjusted to match reference B97-3c computed HVFs in the course of its development.

The SQM methods aim at bridging the gap between FFs and *ab initio* methods, both qualitatively and in terms of computational cost. Due to their underlying quantum chemical nature and the use of (minimal) AO basis sets, dipole moment estimations are expected to be better than in FFs but worse than with DFT. The quality of the force constants in SQM, on the other hand, is strongly dependent on the methods parametrization and will mainly determine the quality of the HVFs. As noted above, methods of the GFN family were developed to provide good performance for the calculation of frequencies and therefore are a natural choice for this study. The monopole-based GFN1- and multipole-extended GFN2-xTB tight-binding schemes, and a nonelectronic variant called GFN-FF, are tested. For a broader perspective, results will also be shown for some PMx methods (PM6-D3H4,⁵⁷ PM6-D3H4X,⁵⁸ PM7⁵⁹) because they are another important class of SQM methods. These results will, however, not be discussed in detail.

A point of reference for all further comparisons is, besides the experimental IR spectra, the performance of the B3LYP-3c method. B3LYP is often cited among the best performing DFT methods for the calculation of vibrational frequencies.^{11,12} This was confirmed for a few test cases in comparison to PBEh-3c³⁴ and B97-3c,³⁵ as well as PBE0⁷² functionals in the course of this project. Because B3LYP is a nonlocal, global hybrid functional, the calculations can become quite expensive which is counterbalanced here by the use of the predefined small def2-mSVP basis. Dipole moments, IR intensities, and frequencies on a medium accuracy level are rather insensitive to the basis set size as long as proper polarization functions are present.⁷³⁻⁷⁶ Hence, B3LYP-3c is expected to yield reasonable results for typical organic molecules, but for more complicated electronic structures the level of theory should be improved. Whether B3LYP-3c really represents an efficient standard DFT level for IR spectra calculations of large systems is one of the questions of the present work.

The IR spectra for the 6556 compounds in the HCNO set were calculated at all levels of theory. Respective geometry optimizations were consistently started from the structures in the NIST database, and in this first evaluation, conformational effects were not considered, i.e., the structures taken correspond to some random conformation in the case of flexible systems (see section 4.3).

The distribution of similarity measures (eqs 7–10) between the unscaled theoretical and the experimental IR spectra over the benchmark set is shown in Figure 1.

Several observations can be made for the unscaled spectra comparison. The r_{match} and r_{Pearson} yield similar distributions at all levels of theory, with the r_{match} yielding slightly higher scores. This is expected since both scores have the same mathematical origin. As already noted in the literature, the r_{Spearman} can display also nonlinear correlations and yield higher scores than the r_{match} or r_{Pearson} on average. The distribution of the r_{Spearman} scores is more narrow than those of r_{match} and r_{Spearman} , except at the FF level. This could either

indicate a general nonlinear correlation between experimental and theoretical spectra or an inability of the $r_{Spearman}$ to reflect finer characteristics. The r_{Euclid} also gives relatively high values for the correlation and shows a very narrow score distribution at all levels of theory. Of the four investigated metrics the r_{Euclid} seems to be the least sensitive to the spectral features, but this issue has not been investigated in detail. In all further comparisons, we will mainly discuss results obtained with the r_{match} .

In terms of performance, GFN2-xTB gives the best (average) qualitative correspondence to the experimental spectra, followed by GFN1-xTB, B3LYP-3c, and last GFN-FF. The three different PMx methods perform very similarly but overall worse than the GFN methods. The average scores are given in Table 1. At first glance it may be surprising that the DFT method performs worse than both tight-binding variants. The simplest reason for this observation is the systematic deviation (overestimation) of the HVFs with hybrid DFT methods. Intensities in the IR spectra are well reproduced by B3LYP-3c, but, in particular, the higher frequency modes are not described well by the harmonic approximation. A typical correction for this behavior is the linear scaling of frequencies by a global factor (global frequency scaling, GFS). For the 6556 systems and all methods, molecule specific frequency scaling (MSFS) factors, i.e., an optimum scaling factor for each case, were determined by maximizing the r_{match} in a steepest descent optimization approach. The corresponding similarity measure distributions are shown in Figure 2, and the average scores are given in Table 1.

The r_{match} and $r_{Pearson}$ distributions in Figure 2 are much more narrow compared to Figure 1 for all methods but the force field. As expected, the largest improvements by scaling are obtained for B3LYP-3c. The average r_{match} increases by 0.236, with an average scaling factor of 0.970 ± 0.011 . This is consistent with GFS factors reported for B3LYP in the literature.^{11–13,77,78} For the GFN1- and GFN2-xTB methods, smaller improvements are observed for linear, molecule specific scaling of the frequencies. The average scaling factor is 0.982 and 0.990 with standard deviations of 0.019 and 0.017, respectively. GFN-FF shows the least improvement by MSFS scaling. Among the investigated GFN methods, the factor of GFN-FF is the one closest to unity and further has the largest standard deviation, i.e., the error is more molecule specific and randomly distributed. In case of the PMx methods, this effect is even more pronounced than for the force field. Using MSFS the average similarity measures among the PMx methods are again very similar, and the performance is in between GFN-FF and GFN1-xTB. Determined GFS factors are around 1.019 and show a very large standard deviation of 0.04–0.05, which essentially obviates the value of global frequency scaling for the PMx methods. This seems to be consistent with the literature, where nonlinear scaling schemes are used in combination with PM6 or PM7 because the linear scaling factors tend to be large.^{15,79,80}

Based on the absolute similarity measure values provided in this section, a clear qualitative trend can be seen. DFT, on average, has the highest correspondence to experimental gas-phase data, while GFN-FF shows the worst performance. The SQM methods GFN1-xTB and GFN2-xTB are in between the FF and B3LYP-3c but overall are closer to the DFT performance. The initial performance expectations of the SQM methods thus were validated. Although the IR spectra at the GFN-FF level are less accurate than at the SQM level, the calculated frequencies still might be useful to obtain thermostistical corrections.⁶

Visual inspection of a few typical cases reveals that a major part of the observed error with GFN-FF stems from bad intensities, which is understandable considering the relatively simple description of the electrostatics in the FF. Furthermore, the GFN1- and GFN2-xTB tight-binding methods clearly outperform the PMx competitors.

For all investigated levels of theory, the different metrics provide consistent descriptions of the similarity, i.e., there is no performance reordering between different theoretical methods if another similarity measure is used in Table 2. Furthermore, the absolute values (and the respective standard deviations, see the Supporting Information) of r_{match} and $r_{Pearson}$, as well as r_{Euclid} and $r_{Spearman}$, are often very similar. This indicates that any of the similarity measures could be used for spectra comparisons or that they can be used complementary.

The r_{match} is a quantitative similarity measure, but since errors from mismatching frequencies and errors from wrong intensities are often inseparable, the scores should be interpreted with care, and on a qualitative rather than quantitative basis. For an r_{match} around 0.6 or smaller, the largest deviations result from systematic frequency shifts for unscaled DFT. In the case of SQM and FF methods, errors in the fingerprint frequency range are common, and intensities are often not well described. Spectra with r_{match} between 0.7 and 0.8 already match well to the experiment, and, in particular, intensities are much better reproduced; but some smaller peaks might still be missing or shifted. Spectra with an r_{match} above 0.9 have predictive quality, and some of the remaining deviations are often only a result of noise in the experimental data. A few examples for the comparison of theoretical and experimental spectra are shown in Figure 3 to provide a qualitative impression for different r_{match} values.

At all r_{match} ranges, noise in the experimental data can be a problem. For better comparability, one could either try to smooth out the experimental data by standard polynomial procedures (e.g., Savitzky–Golay filtering⁸¹) or automatically identify fundamental frequencies from the spectra. Both strategies pose their own problems and can also depend on the amount of noise. Based on our experience, where applicable, a direct comparison to the experimental data is to be preferred.

The aim of this study is to find theoretical methods that provide maximum similarity scores when comparing a calculated spectrum with its experimental counterpart. However, with regard to possible applications in the field of automated unknown compound identification, it is not fully clear whether a candidate ranking based on a maximized spectral similarity is the optimum search strategy. The reason for this is that comparisons of theory and experiment are then made between different molecules, opposed to the already known structure in this study. The application of similarity measures for compound discrimination under realistic conditions is currently investigated in our laboratories and will be discussed elsewhere.

4.2. Scaling of Atomic Masses.

In this section, another easily applicable scheme to improve calculated spectra is explored. From eqs 1–3 it is obvious that modification of the atomic masses m_{ij} will have a (nonlinear) effect on the frequencies. As noted in the Introduction, this is known in context

of isotope effects in IR spectra.^{1,82} In a recent study, Irikura²⁶ proposed a scaling of reduced masses for diatomic molecules as a possible alternative to linear frequency scaling factors ν_{scal} . The modification of the masses could be interpreted as an atom specific way to mimic effects of anharmonicities. In an extension to this idea, we investigate the effect of element mass scaling (EMS) factors, which should provide more flexibility than scaling of the frequencies directly. This has never been tested thoroughly for a large data set of polyatomic molecules. This approach is related to the work of Pulay et al., who proposed individual scaling of the internal coordinate force matrix elements.^{20,27,28,83} In contrast to this scheme, however, the EMS does not require additional internal coordinate setup or any other modification to the frequency computation. All modern DFT and SQM methods are designed as atomistic models, i.e., basis functions are defined for different atoms, and SQM parametrizations are normally element dependent. Thus, an element-dependent EMS parameter is physically plausible and technically feasible. Scaling factors β_i for the atomic masses m_i nonlinearly enter the Hessian matrix $\mathbf{F}^{(m)}$ in eq 1 by the power of minus one-half. Because the normal modes are affected depending on the involved atom types, the nonlinear mass scaling should be able to improve normal modes that include movements of different atom types, e.g., C–O or N–N stretches are treated differently. For simplicity (and easy input handling) instead of introducing an element dependent scaling parameter β_i into eq 1, the masses can directly be treated as a variable parameter ($m_i' = m_i\beta_i$). Many computational chemistry program packages allow users to read in atomic masses via the input, and hence code modifications are not necessary.

The masses m_i' for H, C, N, O, F, Cl, and Br were fitted directly to the experimental data by maximization of the r_{match} . A single global fit was performed for these elements using a Levenberg–Marquardt algorithm.^{84,85} The 6556 spectra were randomly split ($\approx 50\%/50\%$) into a training and an evaluation set, which is common practice for cross-validation in data sciences.⁸⁶ The resulting optimized masses m_i' are shown in Table 3.

Note that the masses m_i' should not be interpreted as actual atomic masses but can only be understood by their effect on the force constants. Since the masses are in the denominator of the prefactor in eq 1, the usually obtained larger mass m_i' repairs too large force constants (missing anharmonicity). The GFN methods show similar trends for the adjusted masses H to F, which suggests systematic errors of the force constants within the method family. During the fitting procedure, it was observed that a modification of the hydrogen mass has the largest influence on the frequencies. This is expected since it is the most common element in the data set. The heavier elements Cl and Br show the largest deviations from their true masses. However, their overall effect on the frequencies is small, and drastic changes of the masses are necessary to obtain any influence at all. For Cl and Br, GFN1-xTB is the only method that does not follow the general trend observed for these elements in the other methods. This could be related to the halogen bond correction terms employed in GFN1-xTB.

The average similarity measures calculated with the fitted masses are shown in Table 4. Following good cross-validation practices, these scores were obtained only for the test set, but the results are nearly identical to the similarity measures of the fitting set.

With the EMS, average similarity measures are only slightly worse than the scores obtained by MSFS in Table 1. Keep in mind that the actual comparison has to be made with the values provided in Table 2. MSFS spectra represent some upper limit for the correspondence to the experiment, but, in practice, spectra are normally globally scaled by the same fixed factor ν_{scal} . Compared to these values the mass-scaling approach yields better results throughout. For B3LYP-3c, the EMS scheme even slightly outperforms MSFS, but the difference between the two schemes is small. The reason for this behavior in DFT is that errors are much more systematic compared to SQM. A few examples for mass scaled spectra in comparison with unscaled and the experimental spectra for GFN2-xTB, GFN-FF, and B3LYP-3c are shown in Figure 4.

Performances are summarized for all four methods as box plots in Figure 5. As can be seen, at all levels of theory there is a clear tendency suggesting that the mass-scaling approach provides better spectra than global frequency scaling.

We tested the transferability of the mass scaling approach for B3LYP-3c and GFN2-xTB for molecules containing other, heavier atoms. The idea is that once optimum atomic masses are determined for a set of elements, other effective element masses can easily be determined in separate fits. This was investigated for the main group elements sulfur, phosphorus, and silicon. Experimental gas-phase IR spectra were again obtained from the NIST database, totaling 538 spectra for sulfur-containing compounds, 100 for phosphorus-containing compounds, and 53 for silicon-containing compounds. The masses of Si, P, and S were modified on top of the previously defined masses m_i' of H, C, N, O, F, Cl, and Br from Table 3 and are given in Table 5.

In general, the procedure of fitting only the mass of a single element separate from predetermined scaled masses of the other elements improves the overall result. At the B3LYP-3c level, the average r_{match} for the silicon, phosphorus, and sulfur sets is again better than the unscaled or globally scaled results. The improvement of the mass scaling in comparison with the GFS for B3LYP-3c is larger here than for the HCNO set above. Note, however, that the same $\nu_{scal} = 0.970$ was used here, and no new linear scaling factor was determined for the silicon-, phosphorus-, and sulfur-containing molecules and that a specific linear scaling factor might improve the $r_{match,lin}$. Surprisingly, at the GFN2-xTB level the mass scaling does not provide a better overall performance than the linear frequency scaling. The main issue here likely is that the fit is stuck in a local minimum due to predetermined masses of the other elements. This hypothesis was briefly tested with a fit for GFN2-xTB on the 538 sulfur-containing structures, in which *all* element masses were adjusted. Significant changes in the HCNO masses can be observed this way, but the average match score ($r_{match, mass}$) for the sulfur subset increases to 0.677. However, the average similarity measure for the 6556 structure set decreases to 0.731 using these masses. The performance is still slightly better than that seen for a global ν_{scal} but overall worse than the result for ideally scaled masses (cf. Table 4). This apparently confirms the assumption of a local minimum in the fit. As a consequence, when using scaled masses some errors can be expected for modes with large contributions of nonmass scaled elements. A larger global adjustment that includes the other elements would likely solve this problem. This was not pursued here because too few experimental reference data points were available for the heavier elements.

4.3. Conformational Dependence.

The molecular conformation can have a significant influence on the IR spectrum,⁸⁷ but this is rarely investigated comprehensively in the literature. Spectra can be affected due to steric effects or intramolecular noncovalent interactions, and in our experience, different hydrogen bonding patterns in particular can have a strong influence. Intensities can be influenced as well if dipole moment changes are large. An example is presented in Figure 6. The 9-aminononanoic acid zwitterion was chosen as an extreme case to demonstrate the effect of different conformations on the dipole moment and IR spectrum. No experimental spectrum is available for this molecule.

For a folded conformation, a much smaller dipole moment is expected than for a structure with a large spatial charge separation. This is confirmed at the GFN2-xTB[GBSA(H₂O)] level where the folded structure is the global minimum energy conformer. As can be seen in Figure 6, most vibrational frequencies are not affected by the conformation. Significant shifts are only observed for the signal at around 3200 cm⁻¹, which corresponds to an N–H stretch, and the signal at 1800–1900 cm⁻¹, corresponding to a C=O vibration. Note that both functional groups are connected by intramolecular hydrogen bonding in the folded conformer. Larger differences can also be seen for the intensities, overall leading to a r_{match} between the two spectra of only 0.786.

Ideally one would conduct a conformational search before calculating the IR spectrum and use the lowest energy conformer or compute Boltzmann population weighted spectra. To obtain a more conclusive impression of the conformational effect, we selected a subset of 554 flexible molecules from the 6556 set and compared the unscaled GFN2-xTB IR spectrum of the lowest conformer with that just computed for the input structure (as obtained from the NIST database). The subset structures were selected based on the flexibility score ξ_f proposed in ref 88. Conformational searches were conducted with the iMTD-GC workflow as implemented in the crest code⁸⁸ at the GFN2-xTB level. All 554 molecules have flexibility scores of $0.8 < \xi_f < 0.9$ commonly obtained for relatively floppy structures such as polypeptides or medium sized branched alkanes.

Conformational effects are investigated by a similarity measure-based comparison of the spectra before and after the conformational sampling, i.e., two theoretical spectra are compared here. For molecules with a similarity measure close to unity, the conformation either does not affect the spectrum, or the input geometry already was (close to) the minimum energy conformer. The average r_{match} between the two spectra at the GFN2 level over all 554 cases is 0.965, i.e., on average a deviation of about 3.5% was observed. An example is shown in Figure 7.

For this molecule, significant differences between the two IR spectra are observed, both regarding frequencies and intensities. With a r_{match} of 0.886, the conformational effect is considered to be quite large. Furthermore, the lowest energy conformation found matches better with the experiment ($r_{match} = 0.746$) than the structure from the NIST database ($r_{match} = 0.699$). This results mainly from better matching signals in the 1000 to 1500 cm⁻¹ and 3000 to 3600 cm⁻¹ regions. Note that for this example only two dihedral angles differ significantly between the two conformations, and the effect on the simulated vibrational

spectrum is already large. This indicates that general conclusions based solely on inspection of the structure are difficult to draw but that at a high accuracy level conformational changes should be considered when calculating IR spectra for all but the most rigid structures. According to our limited experience, even moderately flexible structures with approximately $0.5 > \xi_f > 0.3$ can still undergo considerable conformational changes. In fact, several low energy conformations will typically contribute to the experimental spectrum according to their respective Boltzmann populations at a given temperature and should be considered for very accurate predictions.^{24,87} In automatic workflows, IR spectra averaging could be adapted similar to previously discussed averaging of calculated NMR spectra.⁸⁹

4.4. IR Spectra of Transition Metal Compounds.

A major advantage of the GFN methods over other semiempirical schemes is their robustness with regards to the possible chemical composition of the investigated molecules. The methods are consistently parametrized up to radon ($Z = 86$) and do not require any manual structure preparation step. In this last section, we briefly investigate the performance of GFN1-, GFN2-xTB, and GFN-FF for the calculation of IR spectra of transition metal (TM)-containing compounds. The benchmark set is composed of 58 structures taken from the recently published TMG145 benchmark set.⁷⁰ Only those structures were selected from TMG145 that are known to qualitatively maintain the correct geometry upon optimization at the GFN levels and for which we were able to obtain reasonable structures and frequencies at the B3LYP-D4/def2-TZVPP level. Because no gas-phase IR spectra are available for these molecules, we compare directly to DFT as reference. The larger basis set was used because transition metal complexes usually involve more challenging electronic structures than the mostly organic systems from the previous sections. Since the difference between the mass scaling and linear frequency scaling is small for B3LYP and we do not have a reference for the masses m_i' of the transition metal atoms, the DFT frequencies were scaled by a global factor of 0.967, which was taken from the literature.¹³ The r_{match} for all complexes are shown as bar plots for GFN1-, GFN2-xTB, and GFN-FF in Figure 8.

For this set and taking B3LYP as a reference, GFN2-xTB (av $r_{match,unscal} = 0.714$, $r_{match,mass} = 0.716$) performs better than GFN1-xTB (av $r_{match,unscal} = 0.662$, $r_{match,mass} = 0.682$), and both SQM methods perform better than GFN-FF (av $r_{match,unscal} = 0.532$, $r_{match,mass} = 0.544$). For all three methods, the mass scaling improves the performance only slightly compared to the unscaled result. One case (indicated by the red arrow in Figure 8) is very problematic for all the GFN methods. The corresponding structure is a mercury complex, in which the metal is bound in between two dicarba-*closo*-dodecaboranyl clusters.⁹⁰ The IR spectrum of this structure is not very complicated but is dominated by a very strong signal at 2600 cm^{-1} corresponding to many B–H vibrations (see Figure 9). Because the hydrogen mass is fitted and the mass of boron is not, this signal is too strongly shifted in the mass scaling approach which has a large impact on the r_{match} . Frequencies in the fingerprint region are barely affected by the mass scaling. However, the intensities are not well described at the SQM level which is the main problem in many cases with low r_{match} values.

At the B3LYP-3c level a similar frequency shift is observed due to the mass scaling, but intensities are reproduced very well. In fact, if neither the calculated B3LYP-D4/def2-TZVPP reference nor the B3LYP-3c spectrum is scaled, the r_{match} will be very close to unity (0.972). This excellent agreement is a further indication for the viability of B3LYP-3c.

Many other IR spectra in this benchmark are reasonably well reproduced by the GFN methods, with most of the errors resulting from the fingerprint frequency region. The main factor determining the similarity to the reference here is probably the description of the electrostatic interactions. Of the three GFN methods GFN2-xTB provides the most sophisticated description also including up to atomic quadrupoles,³¹ while GFN1-xTB and the force field rely only on atomic charges. For GFN-FF, they are determined classically, while in GFN1-xTB they are based on the total valence electron density. This usually leads to slightly better intensities and r_{match} at the GFN2-xTB level. As a typical example, a ruthenium-based catalyst⁹¹ is shown in Figure 10. The intensities computed at the GFN2-xTB level more closely resemble the B3LYP reference in comparison with GFN1-xTB. Upon scaling with the same GFS factor as the reference, B3LYP-3c yields a very good spectrum, and in particular, the frequencies are almost perfectly reproduced. Apparently, B3LYP-3c seems to be applicable also for electronically more complicated structures in the TMG145 subset.

5. CONCLUSION

We investigated the performance of the semiempirical GFN tight-binding and force field methods for the simulation of IR spectra in direct comparison with experimental data. The performance of a new low-cost DFT method termed B3LYP-3c, which was introduced and tested in the course of this study, was investigated as well. A large benchmark set of 7247 gas-phase IR spectra was compiled with molecules containing the elements H, C, N, and O, the halogens, and the heavier elements silicon, phosphorus, and sulfur. Electronically more complicated transition metal-containing structures and the effect of conformational changes on the IR spectra were also investigated.

For the comparison between theory and experiment, several similarity measures (spectral overlap metrics) were investigated in order to avoid manual identification and assignment of fundamental frequencies from the spectra. We find that all of the tested measures can be used to sufficiently represent the similarity between two IR spectra. One similarity measure was then used to determine global frequency scaling factors and to explore a new atomic mass scaling approach which provides more flexibility than the linear scaling.

Our results show on a sufficiently large sample size that the low-cost DFT (B3LYP-3c) computed harmonic vibrational spectra can be brought into good agreement with experimental spectra upon linear scaling of the frequencies. The small but well balanced double- ζ AO basis set employed provides sufficiently accurate intensities leading to overall excellently simulated IR spectra. For the semiempirical methods, the scaling of frequencies is less effective, and their overall performance is worse than for DFT. In general, GFN2-xTB provides the best and still useful results at a semiempirical QM level, followed by GFN1-xTB. The GFN-FF mainly suffers from relatively large errors in the intensities.

Similar trends were also observed for a benchmark of 58 transition metal complexes, which are only slightly worse than the (mostly) organic compounds at the GFN levels. Furthermore, it was found that molecular conformation can play an important role for IR spectra calculations in some cases. For medium flexible molecules, the conformation can influence the molecular dipole moments, and vibrational modes might be differently coupled, leading to differences in both intensities and frequencies.

All methods suffer from inherent errors for the potential energy surface and the employed harmonic approximation. For the GFN methods and B3LYP-3c, an atomic mass scaling correction approach was successfully employed as an alternative to standard linear frequency scaling. For the semiempirical methods, mass scaling yields more consistent and overall better results, but also at the DFT level the performance is slightly better than with a global frequency scaling factor. Attention has to be paid, however, to molecules that contain elements without a fitted mass scaling parameter. This problem was observed for the transition metal complexes and the silicon, phosphorus, and sulfur subsets.

Based on the presented results, we recommend the usage of semiempirical QM methods as a cost efficient and reasonably accurate tool to study vibrational spectra in initial steps of large scale compound identification workflows where thousands of DFT Hessian calculations are prohibitive in terms of computational effort. In particular, the GFN2-xTB method seems to be promising for this task due to a sophisticated description of the electrostatic energy terms and an excellent cost/accuracy trade-off. If higher accuracy is required, low-cost DFT or even higher level *ab initio* methods can be used for smaller sets of compounds. Furthermore, for flexible molecules it is important to investigate conformational ensembles.

Empirical adjustment of the atomic masses provides a robust alternative to linear frequency scaling and is particularly promising at the SQM level. Frequencies might be further improved by combining the linear and mass scaling approaches or by including anharmonicities explicitly. Pairwise atomic mass scaling factors would probably provide the highest flexibility but would require modification of the QM codes and complicates the fitting procedure. First advances have also recently been made to obtain anharmonicity corrections via machine learning.²⁵ A combination of the mass scaling approach with machine learning could be potentially promising. However, these techniques only improve vibrational frequencies, and in fact, larger deviations from experiment often result from mismatching intensities. A better description of molecular dipole moments and their derivatives is particularly problematic for low-level SQM methods with their small AO basis sets and in particular FFs due to the classical monopole models. Another problem is the availability and quality of experimental reference data. For metal-containing substances, very few gas-phase IR spectra are available. Some experimental spectra also show significant noise, which can produce artifacts in the similarity measures. A possible strategy for future work to address these problems might be the fit to very high level theoretical data which, however, will limit the size and number of the molecules considered.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the DFG in the framework of the “Gottfried Wilhelm Leibniz Prize” awarded to S.G. D.F.G. was funded by a US National Institutes of Health Grant GM087714. The authors thank S. A. Katsyuba, M. Bursch, F. Bohle, and Dr. E. Caldeweyher for valuable discussions. The authors declare no competing financial interest.

REFERENCES

- (1). Schrader B Infrared and Raman Spectroscopy: Methods and Applications; VCH: New York, 2018; DOI: 10.1002/9783527615438.
- (2). McCarthy M; Lee KL Molecule Identification with Rotational Spectroscopy and Probabilistic Deep Learning. *J. Phys. Chem. A* 2020, 124, 3002–3017. [PubMed: 32212702]
- (3). Henschel H; van der Spoel D An Intuitively Understandable Quality Measure for Theoretical Vibrational Spectra. *J. Phys. Chem. Lett* 2020, 11, 5471–5475. [PubMed: 32558572]
- (4). Pople JA; Schlegel HB; Krishnan R; Defrees DJ; Binkley JS; Frisch MJ; Whiteside RA; Hout RF; Hehre WJ Molecular orbital studies of vibrational frequencies. *Int. J. Quantum Chem* 1981, 20, 269–278.
- (5). Fogarasi G; Pulay P AB Initio Vibrational Force Fields. *Annu. Rev. Phys. Chem* 1984, 35, 191–213.
- (6). Spicher S; Grimme S Efficient Computation of Free Energy Contributions for Association Reactions of Large Molecules. *J. Phys. Chem. Lett* 2020, 11, 6606–6611. [PubMed: 32787231]
- (7). Halgren T AMerck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem* 1996, 17, 553–586.
- (8). Nevins N; Allinger NL Molecular mechanics (MM4) vibrational frequency calculations for alkenes and conjugated hydrocarbons. *J. Comput. Chem* 1996, 17, 730–746.
- (9). Bannwarth C; Caldeweyher E; Ehlert S; Hansen A; Pracht P; Seibert J; Spicher S; Grimme S Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci* 2020, No. e01493.
- (10). Stratmann RE; Burant JC; Scuseria GE; Frisch MJ Improving harmonic vibrational frequencies calculations in density functional theory. *J. Chem. Phys* 1997, 106, 10175–10183.
- (11). Scott AP; Radom L Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem* 1996, 100, 16502–16513.
- (12). Merrick JP; Moran D; Radom L An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J. Phys. Chem. A* 2007, 111, 11683–11700. [PubMed: 17948971]
- (13). Kesharwani MK; Brauer B; Martin J ML Frequency and Zero-Point Vibrational Energy Scale Factors for Double-Hybrid Density Functionals (and Other Selected Methods): Can Anharmonic Force Fields Be Avoided? *J. Phys. Chem. A* 2015, 119, 1701–1714. [PubMed: 25296165]
- (14). Witek HA; Morokuma K Systematic study of vibrational frequencies calculated with the self-consistent charge density functional tight-binding method. *J. Comput. Chem* 2004, 25, 1858–1864. [PubMed: 15376252]
- (15). Fekete ZA; Hoffmann EA; Körtvélyesi T; Penke B Harmonic vibrational frequency scaling factors for the new NDDO Hamiltonians: RM1 and PM6. *Mol. Phys* 2007, 105, 2597–2605.
- (16). Henschel H; Andersson AT; Jespers W; Mehdi Ghahremanpour M; van der Spoel D Theoretical Infrared Spectra: Quantitative Similarity Measures and Force Fields. *J. Chem. Theory Comput.* 2020, 16, 3307–3315. [PubMed: 32271575]
- (17). Blom C; Otto L; Altona C Application of self-consistent-field ab initio calculations to organic molecules. *Mol. Phys* 1976, 32, 1137–1149.
- (18). Blom C; Altona C Application of self-consistent-field ab initio calculations to organic molecules. *Mol. Phys* 1976, 31, 1377–1391.
- (19). Blom C; Altona C Application of self-consistent-field ab initio calculations to organic molecules. *Mol. Phys* 1977, 34, 177–192.
- (20). Pulay P; Fogarasi G; Pongor G; Boggs JE; Vargha A Combination of theoretical ab initio and experimental information to obtain reliable harmonic force constants. Scaled quantum

mechanical (QM) force fields for glyoxal, acrolein, butadiene, formaldehyde, and ethylene. *J. Am. Chem. Soc.* 1983, 105, 7037–7047.

- (21). Fogarasi G; Zhou X; Taylor PW; Pulay P The calculation of ab initio molecular geometries: efficient optimization by natural internal coordinates and empirical correction by offset forces. *J. Am. Chem. Soc.* 1992, 114, 8191–8201.
- (22). Laury ML; Boesch SE; Haken I; Sinha P; Wheeler RA; Wilson AK Harmonic vibrational frequencies: Scale factors for pure, hybrid, hybrid meta, and double-hybrid functionals in conjunction with correlation consistent basis sets. *J. Comput. Chem.* 2011, 32, 2339–2347. [PubMed: 21598273]
- (23). Laury ML; Carlson MJ; Wilson AK Vibrational frequency scale factors for density functional theory and the polarization consistent basis sets. *J. Comput. Chem.* 2012, 33, 2380–2387. [PubMed: 22815183]
- (24). Gastegger M; Behler J; Marquetand P Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* 2017, 8, 6924–6935. [PubMed: 29147518]
- (25). Lam J; Abdul-Al S; Allouche A-R Combining Quantum Mechanics and Machine-Learning Calculations for Anharmonic Corrections to Vibrational Frequencies. *J. Chem. Theory Comput.* 2020, 16, 1681–1689. [PubMed: 32003996]
- (26). Irikura KK Mass scaling for vibrational frequencies from ab initio calculations. *Chem. Phys. Lett.* 2005, 403, 275–279.
- (27). Rauhut G; Pulay P Transferable Scaling Factors for Density Functional Derived Vibrational Force Fields. *J. Phys. Chem.* 1995, 99, 3093–3100.
- (28). Baker J; Jarzecki AA; Pulay P Direct Scaling of Primitive Valence Force Constants: An Alternative Approach to Scaled Quantum Mechanical Force Fields. *J. Phys. Chem. A* 1998, 102, 1412–1424.
- (29). Linstrom EP; Mallard W NIST Chemistry WebBook, NIST Standard Reference Database Number 69 <https://webbook.nist.gov/chemistry/> (accessed 2020-07-20).
- (30). Grimme S; Bannwarth C; Shushkov PA Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ($Z = 1 - 86$). *J. Chem. Theory Comput.* 2017, 13, 1989–2009. [PubMed: 28418654]
- (31). Bannwarth C; Ehlert S; Grimme S GFN2-xTB – An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* 2019, 15, 1652–1671. [PubMed: 30741547]
- (32). Spicher S; Grimme S Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems. *Angew. Chem., Int. Ed.* 2020, 59, 15665–15673.
- (33). Becke A Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 1993, 98, 5648–5652.
- (34). Grimme S; Brandenburg JG; Bannwarth C; Hansen A Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* 2015, 143, 054107. [PubMed: 26254642]
- (35). Brandenburg JG; Bannwarth C; Hansen A; Grimme S B97–3c: A revised low-cost variant of the B97-D density functional method. *J. Chem. Phys.* 2018, 148, 064104. [PubMed: 29448802]
- (36). Katsyuba SA; Zvereva EE; Grimme S Fast Quantum Chemical Simulations of Infrared Spectra of Organic Compounds with the B97–3c Composite Method. *J. Phys. Chem. A* 2019, 123, 3802–3808. [PubMed: 30958005]
- (37). Katsyuba SA; Spicher S; Gerasimova TP; Grimme S Fast and Accurate Quantum Chemical Modeling of Infrared Spectra of Condensed-Phase Systems. *J. Phys. Chem. B* 2020, 124, 6664–6670. [PubMed: 32633534]
- (38). Penchev PN; Sohau AN; Andreev G N Description and Performance Analysis of an Infrared Library Search System. *Spectrosc. Lett.* 1996, 29, 1513–1522.
- (39). Engel T; Gasteiger J Chemoinformatics: Basic Concepts and Methods; Wiley-VCH: Weinheim, Germany, 2018.
- (40). Jensen F Introduction to Computational Chemistry; Wiley: Chichester, UK, 2017.

- (41). Gussoni M; Castiglioni C; Ramos M; Rui M; Zerbi G Infrared intensities: from intensity parameters to an overall understanding of the spectrum. *J. Mol. Struct*1990, 224, 445–470.
- (42). Barone V; Biczysko M; Bloino J Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation. *Phys. Chem. Chem. Phys*2014, 16, 1759–1787. [PubMed: 24346191]
- (43). Porezag D; Pederson M R Infrared intensities and Raman-scattering activities within density-functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys*1996, 54, 7830–7836.
- (44). Galabov BS; Dudev T Vibrational Intensities; Elsevier: Amsterdam, 1996.
- (45). Fan L; Ziegler T Application of density functional theory to infrared absorption intensity calculations on main group molecules. *J. Chem. Phys*1992, 96, 9005–9012.
- (46). Schwenke DW; Partridge H Convergence testing of the analytic representation of an ab initio dipole moment function for water: Improved fitting yields improved intensities. *J. Chem. Phys*2000, 113, 6592–6597.
- (47). Neugebauer J; Reiher M; Kind C; Hess B A Quantum chemical calculation of vibrational spectra of large molecules—Raman and IR spectra for Buckminsterfullerene. *J. Comput. Chem*2002, 23, 895–910. [PubMed: 11984851]
- (48). Luber S; Neugebauer J; Reiher M Intensity tracking for theoretical infrared spectroscopy of large molecules. *J. Chem. Phys*2009, 130, 064105. [PubMed: 19222265]
- (49). Kiewisch K; Neugebauer J; Reiher M Selective calculation of high-intensity vibrations in molecular resonance Raman spectra. *J. Chem. Phys*2008, 129, 204103. [PubMed: 19045848]
- (50). Jacob CR; Reiher M Localizing normal modes in large molecules. *J. Chem. Phys*2009, 130, 084106. [PubMed: 19256596]
- (51). Chu PM; Guenther FR; Rhoderick GC; Lafferty WJ The NIST Quantitative Infrared Database. *J. Res. Natl. Inst. Stand. Technol*1999, 104, 59–81.
- (52). Baumann K; Clerc J Computer-assisted IR spectra prediction — linked similarity searches for structures and spectra. *Anal. Chim. Acta*1997, 348, 327–343.
- (53). Tan X; Chen X; Song S A computational study of spectral matching algorithms for identifying Raman spectra of polycyclic aromatic hydrocarbons. *J. Raman Spectrosc.* 2017, 48, 113–118.
- (54). Vrancic C; Petrich W Effective Fragment Potential Study of the Influence of Hydration on the Vibrational Spectrum of Glucose. *J. Phys. Chem. A*2011, 115, 12373–12379. [PubMed: 21962037]
- (55). Zapata F; García-Ruiz C The discrimination of 72 nitrate, chlorate and perchlorate salts using IR and Raman spectroscopy. *Spectrochim. Acta, Part A*2018, 189, 535–542.
- (56). Stewart JJ Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model*2007, 13, 1173. [PubMed: 17828561]
- (57). ezá J; Hobza P Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* 2012, 8, 141–151. [PubMed: 26592877]
- (58). Brahmshatriya PS; Dobeš P; Fanfrlík J; ezá J; Paruch K; Bronowska A; Lepšík M; Hobza P Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo[1,5-a]pyrimidines. *Curr. Comput.-Aided Drug Des.* 2013, 9, 118–129. [PubMed: 23157414]
- (59). Stewart JJ Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model*2013, 19, 1–32. [PubMed: 23187683]
- (60). Balasubramani SG; Chen GP; Coriani S; Diedenhofen M; Frank MS; Franzke YJ; Furche F; Grotjahn R; Harding ME; Hättig C; Hellweg A; Helmich-Paris B; Holzer C; Huniar U; Kaupp M; Marefat Khah A; Karbalaeei Khani S; Müller T; Mack F; Nguyen BD; Parker SM; Perl E; Rappoport D; Reiter K; Roy S; Rückert M; Schmitz G; Sierka M; Tapavicza E; Tew DP; van Wüllen C; Voora VK; Weigend F; Wodyński A; Yu J M TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations. *J. Chem. Phys*2020, 152, 184107. [PubMed: 32414256]

- (61). Grimme S; Antony J; Ehrlich S; Krieg HA consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys*2010, 132, 154104. [PubMed: 20423165]
- (62). Grimme S; Ehrlich S; Goerigk LEffect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem*2011, 32, 1456–1465. [PubMed: 21370243]
- (63). Kruse H; Grimme SA geometrical correction for the inter- and intra-molecular basis set superposition error in Hartree-Fock and density functional theory calculations for large systems. *J. Chem. Phys*2012, 136, 154101. [PubMed: 22519309]
- (64). Goerigk L; Hansen A; Bauer C; Ehrlich S; Najibi A; Grimme SA look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys*2017, 19, 32184–32215. [PubMed: 29110012]
- (65). Kruse H; Goerigk L; Grimme SWhy the Standard B3LYP/6–31G* Model Chemistry Should Not Be Used in DFT Calculations of Molecular Thermochemistry: Understanding and Correcting the Problem. *J. Org. Chem*2012, 77, 10824–10834. [PubMed: 23153035]
- (66). Neese FSoftware update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci*2018, 8, No. e1327.
- (67). Caldeweyher E; Ehlert S; Hansen A; Neugebauer H; Spicher S; Bannwarth C; Grimme SA generally applicable atomic-charge dependent London dispersion correction. *J. Chem. Phys*2019, 150, 154122. [PubMed: 31005066]
- (68). Weigend F; Ahlrichs RBalanced 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*2005, 7, 3297–3305. [PubMed: 16240044]
- (69). Neese F; Wennmohs F; Hansen A; Becker UEfficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A ‘chain-of-spheres’ algorithm for the Hartree–Fock exchange. *Chem. Phys*2009, 356, 98–109.
- (70). Bursch M; Neugebauer H; Grimme SStructure Optimisation of Large Transition-Metal Complexes with Extended Tight-Binding Methods. *Angew. Chem., Int. Ed*2019, 58, 11078–11087.
- (71). Alecu IM; Zheng J; Zhao Y; Truhlar DGComputational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J. Chem. Theory Comput*2010, 6, 2872–2887. [PubMed: 26616087]
- (72). Adamo C; Barone VToward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys*1999, 110, 6158–6170.
- (73). Yamaguchi Y; Frisch M; Gaw J; Schaefer HF; Binkley JSAnalytic evaluation and basis set dependence of intensities of infrared spectra. *J. Chem. Phys*1986, 84, 2262–2278.
- (74). Thomas JR; DeLeeuw BJ; Vacek G; Crawford TD; Yamaguchi Y; Schaefer HFThe balance between theoretical method and basis set quality: A systematic study of equilibrium geometries, dipole moments, harmonic vibrational frequencies, and infrared intensities. *J. Chem. Phys*1993, 99, 403–416.
- (75). Jensen FPolarization consistent basis sets. IV. The basis set convergence of equilibrium geometries, harmonic vibrational frequencies, and intensities. *J. Chem. Phys*2003, 118, 2459–2463.
- (76). Hait D; Head-Gordon M How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* 2018, 14, 1969–1981. [PubMed: 29562129]
- (77). Sinha P; Boesch SE; Gu C; Wheeler RA; Wilson AKHarmonic Vibrational Frequencies: Scaling Factors for HF, B3LYP, and MP2Methods in Combination with Correlation Consistent Basis Sets. *J. Phys. Chem. A*2004, 108, 9213–9217.
- (78). Andersson MP; Uvdal PNew Scale Factors for Harmonic Vibrational Frequencies Using the B3LYP Density Functional Method with the Triple-Zeta Basis Set 6–311+G(d, p). *J. Phys. Chem. A*2005, 109, 2937–2941. [PubMed: 16833612]
- (79). Chan BUUse of Low-Cost Quantum Chemistry Procedures for Geometry Optimization and Vibrational Frequency Calculations: Determination of Frequency Scale Factors and Application

to Reactions of Large Systems. *J. Chem. Theory Comput.* 2017, 13, 6052–6060. [PubMed: 29116781]

- (80). Rozanska X; Stewart JJP; Ungerer P; Leblanc B; Freeman C; Saxe P; Wimmer E High-Throughput Calculations of Molecular Properties in the MedeA Environment: Accuracy of PM7 in Predicting Vibrational Frequencies, Ideal Gas Entropies, Heat Capacities, and Gibbs Free Energies of Organic Molecules. *J. Chem. Eng. Data* 2014, 59, 3136–3143.
- (81). Savitzky A; Golay MJ Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 1964, 36, 1627–1639.
- (82). Limbach H; Hennig J; Stulz J IR-spectroscopic study of isotope effects on the NH/ND-stretching bands of meso-tetraphenylporphine and vibrational hydrogen tunneling. *J. Chem. Phys.* 1983, 78, 5432–5436.
- (83). Kozłowski PM; Jarzecki AA; Pulay P Vibrational Assignment and Definite Harmonic Force Field for Porphine. 1. Scaled Quantum Mechanical Results and Comparison with Empirical Force Field. *J. Phys. Chem.* 1996, 100, 7007–7013.
- (84). Levenberg K A Method for the Solution of Certain Nonlinear Problems in Least Squares. *Q. Appl. Math.* 1944, 2, 164–168.
- (85). Marquardt D An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* 1963, 11, 431–441.
- (86). Picard RR; Cook RD Cross-Validation of Regression Models. *J. Am. Stat. Assoc.* 1984, 79, 575–583.
- (87). Klaeboe P Conformational studies by vibrational spectroscopy: a review of various methods. *Vib. Spectrosc.* 1995, 9, 3–17.
- (88). Pracht P; Bohle F; Grimme S Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* 2020, 22, 7169–7192. [PubMed: 32073075]
- (89). Grimme S; Bannwarth C; Dohm S; Hansen A; Pisarek J; Pracht P; Seibert J; Neese F Fully Automated Quantum-Chemistry-Based Computation of Spin–Spin-Coupled Nuclear Magnetic Resonance Spectra. *Angew. Chem., Int. Ed.* 2017, 56, 14763–14769.
- (90). Morel P; Schaffer P; Britten JF; Valliant JF Solvent–metal interactions in bis[1,2-dicarba-closo-dodecaboran(12)-1-yl]-mercury(II) dichloromethane solvate and bis[1,12-dicarba-closo-dodecaboran(12)-1-yl]mercury(II) tetrahydrofuran solvate. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* 2002, 58, m601–m604.
- (91). Kingsbury JS; Harrity JPA; Bonitatebus PJ; Hoveyda A H A Recyclable Ru-Based Metathesis Catalyst. *J. Am. Chem. Soc.* 1999, 121, 791–799.

Figure 1. Similarity measure distribution (6556 cases) for unscaled theoretical spectra at the B3LYP-3c, GFN1-, GFN2-xTB, and GFN-FF levels. Bars were binned (40 bins) for better visualization.

Figure 2.

Similarity measure distribution (6556 cases) for spectra at the B3LYP-3c, GFN1-, GFN2-xTB, and GFN-FF levels. The calculated frequencies were uniformly but specifically scaled for each molecule in order to maximize the r_{match} to experiment. Bars were binned (40 bins) for better visualization. Values of unity for the r_{match} result from graphic interpolation by the plot program.

Figure 3. Comparison between the theoretical (black) and the experimental (red) spectra for six exemplary molecules, at GFN2-xTB or B3LYP-3c levels.

Figure 4.

Comparison among the theoretical mass scaled (black), theoretical unscaled (blue), and the experimental (red) spectra for six exemplary molecules, at GFN2-xTB, GFN-FF, or B3LYP-3c levels.

Figure 5. Similarity measures (r_{match}) for the 6556 spectra comparisons at the B3LYP-3c, GFN1-xTB, GFN2-xTB, and GFN-FF levels of theory, visualized as box plots.

Figure 6.

Two conformations of the 9-aminononanoic acid zwitterion and their respective calculated IR spectra. Orientation of the dipole vector is indicated by the green arrow. The dipole moments and frequencies were calculated at the GFN2-xTB[GBSA(H₂O)] level. The spectrum plotted in black corresponds to the linear conformer; the red spectrum refers to the folded conformer.

Figure 7.

Comparison between the theoretical IR spectra (GFN2-xTB) of two conformations of ethyl 3-amino-4,4,4-trifluorocrotonate. The conformer (and spectrum) marked in red corresponds to the structure as obtained from the NIST database. The conformer marked in gray (black line) was obtained through conformational sampling. The experimental spectrum is plotted in blue.

Figure 8.

Bar plots for the r_{match} of all the 58 transition metal compounds investigated at the GFN1-, GFN2-xTB, and GFN-FF levels. The reference IR spectra were calculated at the B3LYP-D4/def2-TZVPP level. The light-colored bars denote the r_{match} obtained by using mass scaling. The system indicated by the red arrow is discussed in the text below.

Figure 9.
Theoretical IR spectra for the $[\text{Hg}(\text{C}_2\text{B}_{10}\text{H}_{11})_2]$ complex.

Figure 10.
Theoretical IR spectra for the Cl₂Ru(=CH-*o*-OMeC₁₀H₆)(PCy₃) complex.

Table 1.

Average Metrics r_{match} , r_{Euclid} , $r_{Spearman}$ and $r_{Pearson}$ for the 6556 Unscaled and Ideally (MSFS) Scaled IR Spectra Calculated at All Tested Levels of Theory^a

method	unscaled			molecule specific scaling		
	r_{match}	r_{Euclid}	$r_{Spearman}$	r_{match}	r_{Euclid}	$r_{Pearson}$
B3LYP-3c	0.628	0.710	0.739	0.864	0.880	0.873
GFN1-xTB	0.632	0.714	0.727	0.727	0.775	0.710
GFN2-xTB	0.690	0.751	0.748	0.765	0.802	0.750
GFN-FF	0.545	0.659	0.650	0.597	0.691	0.528
PM6-D3H4	0.518	0.644	0.591	0.640	0.713	0.584
PM6-D3H4X	0.518	0.644	0.591	0.634	0.713	0.584
PM7	0.476	0.620	0.570	0.634	0.713	0.582

^aFor better visibility, standard deviations have been omitted from the table and can be found in the Supporting Information.

Table 2.

Frequency Scaling Factors for the Investigated Theoretical Methods Determined As the Average of the MSFS Factors and the Respective Similarity Measures Obtained by Application of These Factors as Global Frequency Scaling Parameters

method	ν_{scal}	$SD(\nu_{scal})$	r_{match}	r_{Euclid}	$r_{Spearman}$	$r_{Pearson}$
B3LYP-3c	0.970	0.011	0.862	0.879	0.869	0.819
GFN1-xTB	0.982	0.019	0.686	0.747	0.660	0.754
GFN2-xTB	0.990	0.017	0.723	0.773	0.702	0.757
GFN-FF	0.993	0.020	0.564	0.670	0.434	0.659

Table 3.Optimized Atomic Masses for the Elements H, C, N, O, F, Cl, and Br (in au)^b

atom type	reference m_i	B3LYP-3c m_i'	GFN1-xTB m_i'	GFN2-xTB m_i'	GFN-FF m_i'
H	1.0079	1.07823	1.0801	1.0518	1.0607
C	12.0110	12.7059	11.7831	11.7844	11.6786
N	14.0067	15.1762	14.1462	12.8194	11.8136
O	15.9994	17.4074	18.8858	17.5056	15.7004
F	18.9984	15.7403	13.5474	11.8874	9.3633
Cl	35.4530	23.1230	52.3449	14.0078	14.1648
Br	79.9040	79.9040 ^a	60.3144	45.4038	43.1998

^aModification of mass did not improve the fit.^b m_i refers to standard atomic masses; modified values m_i' for the GFN methods and B3LYP-3c were determined by fitting to experimental IR spectra.

Table 4.

Averaged Similarity Measures between Experimental and Theoretical IR Spectra Obtained from Fitted-Mass Spectra Calculations

method	<i>r</i> match	<i>r</i> Euclid	<i>r</i> Spearman	<i>r</i> Pearson
B3LYP-3c	0.865	0.881	0.813	0.873
GFN1-xTB	0.721	0.771	0.747	0.699
GFN2-xTB	0.750	0.793	0.756	0.731
GFN-FF	0.589	0.686	0.678	0.516

Table 5.Optimum Masses for the Elements Si, P, and S and Average Match Scores for the Sets^a

atom type		reference	B3LYP-3c	GFN2-xTB
Si	m_i' [au]	28.0855	31.1671	26.1354
	$r_{match,unscal}$		0.650	0.668
	$r_{match,lin}$		0.795	0.686
	$r_{match,mass}$		0.802	0.681
P	m_i' [au]	30.9738	26.0246	25.2719
	$r_{match,unscal}$		0.748	0.734
	$r_{match,lin}$		0.791	0.740
	$r_{match,mass}$		0.822	0.742
S	m_i' [au]	32.0600	26.0670	23.7071
	$r_{match,unscal}$		0.576	0.612
	$r_{match,lin}$		0.782	0.640
	$r_{match,mass}$		0.806	0.637

^a $r_{match,unscal}$ is the average similarity measure without any modification of the frequencies, and $r_{match,lin}$ is the average similarity measure with a fixed ν_{scal} applied. $r_{match,mass}$ refers to similarity measures obtained with the fitted masses given in this table and Table 3.