

# Theoretical Infrared Spectra: Quantitative Similarity Measures and Force Fields

Henning Henschel,\* Alfred T. Andersson, Willem Jaspers, Mohammad Mehdi Ghahremanpour, and David van der Spoel\*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 3307–3315

Read Online

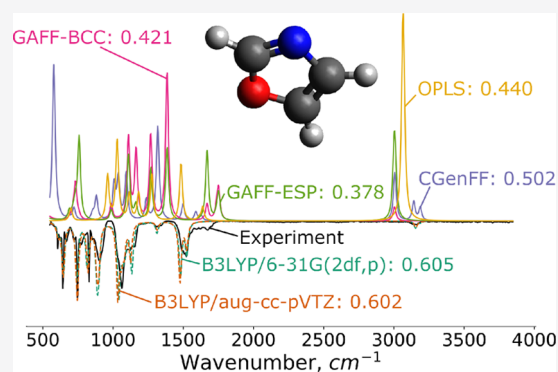
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Infrared spectroscopy can provide significant insight into the structures and dynamics of molecules of all sizes. The information that is contained in the spectrum is, however, often not easily extracted without the aid of theoretical calculations or simulations. We present here the calculation of the infrared spectra of a database of 703 gas phase compounds with four different force fields (CGenFF, GAFF-BCC, GAFF-ESP, and OPLS) using normal-mode analysis. Modern force fields increasingly use virtual sites to describe, e.g., lone-pair electrons or the  $\sigma$ -holes on halogen atoms. This requires some adaptation of code to perform normal-mode analysis of such compounds, the implementation of which into the GROMACS software is briefly described as well. For the quantitative comparison of the obtained spectra with experimental reference data, we discuss the application of two different statistical correlation coefficients, Pearson and Spearman. The advantages and drawbacks of the different methods of comparison are discussed, and we find that both methods of comparison give the same overall picture, showing that present force field methods cannot match the performance of quantum chemical methods for the calculation of infrared spectra.



## INTRODUCTION

Infrared (IR) spectroscopy is a fast, cheap, and nondestructive method that has for a long time been a common tool of analytical chemists. Especially in recent times, its use has been extended to a wide variety of fields, ranging from pharmaceutical applications,<sup>1</sup> to food sciences<sup>2,3</sup> to medicinal applications,<sup>4,5</sup> and even to entomology.<sup>6</sup> The obtained spectra hold detailed information on molecular structure and dynamics, to the level that it can be used to analyze the hydrogen bonding patterns and structural details of liquid water.<sup>7</sup> However, the interpretation of spectra requires experience and often detailed previous understanding of the system in question.<sup>8</sup> At this point, computational tools can assist by predicting the spectra for compounds that have not, or possibly (as, e.g. short-lived reaction intermediates) cannot, be isolated to obtain their pure spectrum. In practice quantum chemistry (QC) and, especially, density functional theory (DFT) have been widely used for the purpose of predicting IR spectra, particularly since a vibrational analysis is also necessary for the characterization of stationary structures, as well as for thermochemistry and reaction kinetics. The range of systems that can be treated by DFT is, however, limited by the computational cost. It was realized quite some time ago that the IR spectrum of a protein is related to its structure.<sup>9</sup> Since then there has been a significant increase in the use of IR spectroscopy for the characterization of structure and interactions of proteins,<sup>10</sup> and even their structural dynamics.<sup>11</sup> For computational studies of these types of systems, and ground

state compounds in general, force field based methods can in principal be applied. There is, however, currently limited understanding as to the reliability of vibrational spectra as predicted by force field methods.

A lot of work has been done on the evaluation of vibrational frequencies from molecular models, and indeed the MMx family of force fields were optimized in part to reproduce frequencies.<sup>12–15</sup> We have recently studied force field thermochemistry and evaluated entropies, heat capacity at constant volume, and internal energy<sup>16</sup> for the CHARMM General Force Field (CGenFF)<sup>17</sup> and the General Amber Force Field (GAFF).<sup>18</sup> These models are used routinely for a large number of applications ranging from liquids simulations,<sup>19,20</sup> over free energy of solvation calculations<sup>21,22</sup> to drug design,<sup>23,24</sup> rendering it essential to understanding their capabilities and limitations as well as possible. In the development of both GAFF<sup>18</sup> and CGenFF,<sup>17</sup> vibrational frequencies from quantum chemistry calculations were used to derive force constants; one would therefore expect the vibrations to be reproduced reasonably well. We note that a GAFF topology server, provided by

Received: February 19, 2020

Published: April 9, 2020



Procacci,<sup>25</sup> also provides infrared spectra of the compound submitted to the server.

In continuation of our previous work,<sup>16</sup> we study here the performance of force fields for the calculation of infrared spectra from normal modes and the associated changes in dipole moments. In addition to the force fields used in our previous studies, we include here the widely used Optimized Potential for Liquid Simulations (OPLS).<sup>26,27</sup> These spectra obtained from force field calculations are, together with quantum chemistry data from the Alexandria library,<sup>28,29</sup> compared to experimental reference spectra.

Representing spectra as a linear vector of intensities provides the possibility to use a number of simple mathematical procedures for their quantitative comparison. For a long time, statistical correlation coefficients have been used for the quantitative comparison and identification of IR spectra.<sup>30</sup> In the meantime, a multitude of other methods has been developed for this purpose.<sup>1</sup> However, these newer methods mostly focus on the comparison of experimentally obtained spectra with database references; i.e., they are mostly developed to cope with experimental issues such as admixtures of other compounds, experimental noise, and often specific sets of compounds. Such features are rather different from the type of difference we can expect to observe in this study, where the question is whether the computed spectra at all exhibit the correct features. Therefore, we selected the well-known Pearson correlation coefficient and the Spearman correlation coefficient for this study, as described in [Methods](#).

## METHODS

**Calculation of Infrared Spectra.** Calculation of the infrared spectra was based on the Gaussian-4 calculations and B3LYP/aug-cc-pVTZ optimized structures of compounds contained in the Alexandria library.<sup>28,29,31</sup> Normal-mode analysis using the CGenFF, GAFF-BCC, and GAFF-ESP force fields has been previously described in ref 16. For the OPLS force field calculations, the topologies were generated using the python version of LigParGen.<sup>32</sup> Integer molecular charges were assigned using RDkit<sup>33</sup> with the mmff94 force field. Partial charges were calculated using the 1.14\*CM1A method.<sup>34,35</sup> Using these topologies, a normal-mode analysis was conducted with GROMACS<sup>36</sup> as described previously.<sup>16</sup> The normal modes are based on the calculation of the Hessian matrix,  $\mathcal{H}$ , of second derivatives of the energy with respect to the atomic coordinates  $q$

$$\mathcal{H}_{ij} = \frac{\partial^2 E}{\partial q_i \partial q_j} \quad (1)$$

where  $i$  and  $j$  run from 0 to  $N - 1$  if  $N$  is the number of atoms in the compound. If virtual sites  $\nu$  are used, for instance, to model the  $\sigma$ -hole for halogen atoms in CGenFF,<sup>17</sup> the energy,  $E$ , depends on the positions of both atoms and virtual sites; that is,  $E = E(q_0, \dots, q_{N-1}, \nu_0, \nu_{M-1})$ , where the positions of the  $M$  virtual sites  $\nu$  in the compound are a function of the atomic coordinates  $q$ . The influence of the virtual site positions on the Hessian needs to be taken into account when computing  $\mathcal{H}$ . In GROMACS,<sup>37</sup> this calculation is done numerically—the  $N$  atoms are moved independently in all three spatial dimensions and the forces are computed. From these forces, the second derivative of the energy is then evaluated numerically. From the 2019 version of GROMACS, virtual sites positions are updated before each force calculation, which means that their

influence on the Hessian is taken into account explicitly. The force field files used in this work are available at the <http://virtualchemistry.org> repository.<sup>38</sup>

Vibrational frequencies obtained using common quantum chemical methods are known to suffer from systematic errors, mainly due to basis set incompleteness, neglect of anharmonicity, and incomplete treatment of electron correlation.<sup>39</sup> In order to balance this systematic deviation, a multitude of scaling schemes have been developed, and at least single scaling factors have been determined for a large number of methods and basis sets.<sup>40–42</sup> For the frequencies obtained from the quantum chemical methods we applied single scaling factors of 0.965 for B3LYP/6-31G(2df,p) and 0.968 for B3LYP/aug-cc-pVTZ.<sup>43</sup> The unscaled frequencies at the B3LYP/6-31G(2df,p) level of theory were extracted from the G4 calculations of the Alexandria library. Other values have been suggested,<sup>44</sup> but as will be discussed below, they give slightly worse results in this study. Therefore, we use the scaling factors given here for all spectral comparisons in this study unless noted otherwise. We also tested the applicability of single scaling factors for frequencies derived from force field calculations, as described in the following section. These, however, did not lead to any improved agreement with the experimental spectra.

For the calculation of a full IR spectrum, in addition to the eigenfrequencies of the normal modes, the intensities and the line shapes are required. In the case of the quantum chemical calculations, both the eigenfrequencies and the corresponding IR intensities are produced by default when a frequency calculation is requested in the Gaussian software,<sup>45</sup> and were thus readily available from the Alexandria library. Details of the quantum chemical calculations from which the frequencies were obtained have been presented previously (see refs 28 and 29). For the force field calculations, the intensities  $I_n$  were derived from the transition dipole derivatives:

$$I_n = \sum_{k=1}^3 \left( \frac{\partial p_k}{\partial Q_n} \right)^2 \quad (2)$$

where  $k$  iterates over Cartesian dimensions,  $p$  is the dipole moment of the molecule, and  $Q_n$  is the normal coordinate  $n$ . In order to take into account virtual sites  $\nu$ , we note that

$$p_k = p_k(q_0, \dots, q_{N-1}, \nu_0, \dots, \nu_{M-1}) \quad (3)$$

and rewrite eq 2 as

$$I_n = \sum_{k=1}^3 \left( \frac{\partial p_k}{\partial q_s} \frac{\partial q_s}{\partial Q_n} \right)^2 \quad (4)$$

where  $s$  iterates over the  $N$  atomic coordinates. To make the calculation of intensities practical, the numerical derivative of the dipole moment with respect to the atomic coordinates  $\frac{\partial p_k}{\partial q_s}$  is stored in a text file during the normal-mode analysis and finally we note that the term  $\frac{\partial q_s}{\partial Q_n}$  corresponds to one over component  $s$  of eigenvector  $n$ .

Each pair of eigenfrequency and intensity is represented by a Lorentz distribution, centered at the eigenfrequency and multiplied by the intensity. The full width at half-maximum (fwhm) of each distribution was set to  $24 \text{ cm}^{-1}$ , as suggested by Mott and Rez<sup>46</sup> on the basis of the estimated bandwidth observed in the NIST database,<sup>47</sup> which is also the source of

our experimental reference spectra in this work. However, as will be discussed below, the exact value used for the bandwidths was found to be of only minor importance for any of the analyses presented in this study. Both frequency range and resolution of the calculated spectra were matched to those of the corresponding experimental spectrum (most commonly found values are in a range from 450 to 3966  $\text{cm}^{-1}$  with a resolution of 4  $\text{cm}^{-1}$ ). To obtain the entire spectrum of a molecule, all Lorentz distribution vectors for that molecule were added. Finally, the resulting spectra were normalized with respect to the area under the curve.

**Quantitative Evaluation of Spectra.** For the comparison of the calculated spectra with the reference spectra we used two statistical measures, the first of these being Pearson's product moment correlation coefficient:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (5)$$

where  $x_i$  and  $y_i$  are the elements of the intensity vectors representing the spectra under comparison and  $\bar{x}$  and  $\bar{y}$  the mean values of the  $x_i$  and  $y_i$  values, respectively. The Pearson correlation coefficient provides a measure for the linear correlation between two vectors and has been used in several studies comparing IR and Raman spectra.<sup>48–51</sup> The second statistical measure applied in this study is Spearman's rank correlation coefficient:

$$\rho = 1 - \frac{6 \sum_i d_i^2}{n \cdot (n^2 - 1)} \quad (6)$$

where  $d_i$  is the difference between the ranks of  $x_i$  and  $y_i$  in their respective data set and  $n$  the number of elements in each vector. This coefficient provides a measure for the monotonic correlation between the two vectors. Compared to Pearson's correlation coefficient, it is generally less sensitive toward outliers and can also represent a nonlinear relationship. To the best of our knowledge, this measure has hitherto only been used in one study comparing calculated with experimental vibrational spectra.<sup>48</sup> In addition to the statistical treatment, a selection of spectra were compared visually. This was primarily done for compounds giving the lowest and highest correlation coefficients of both types, as well as those where we observed a significant difference between the two measures.

Finally, we have utilized both statistical measures to determine whether scaling factors can be determined for the force field methods in a fashion similar to quantum chemical methods. Because extraction of the underlying eigenfrequencies from the large number of experimental spectra in parts containing many overlapping bands did not seem feasible, we have chosen a more direct approach of maximizing the respective statistical measure as a function of the scaling factor, using Brent's method as implemented in Scientific Python.<sup>52</sup>

## RESULTS AND DISCUSSION

We computed the theoretical spectra using all six computational approaches for 703 of the compounds for which experimental reference spectra from gas phase measurements were available. The central limiting factor for this number is—besides the availability of experimental spectra—the ability to generate valid topologies for the compounds for all force field methods used in this study. The generated spectra are available at the compound pages of the <http://virtualchemistry.org> web site.

The two statistical measures used for the quantitative comparison between the calculated spectra and the experimental spectra for the same substance are listed in Table 1 averaged over chemical classes and over all compounds. The corresponding compound-wise table can be found in the Supporting Information as Table S1. A number of things become immediately clear from the data shown in Table 1. In terms of reproduction of the experimental spectra, the quantum chemical methods perform clearly better, independent of which measure is used to quantify the agreement. The larger basis set gives marginally better results overall (Pearson's correlation is essentially the same; Spearman's correlation, slightly higher), which is expected as frequencies obtained from B3LYP calculations tend to improve with basis set size.<sup>53</sup> Among the force field methods, CGenFF performs best, and GAFF, weakest, with ESP derived charges giving somewhat worse results than BCC charges. Here it should be noted that the ESP charges used with GAFF in this study have not been derived with the same quantum chemical method that was used in the development of the force field (B3LYP/aug-cc-pVTZ instead of HF/6-31G\*).<sup>29</sup> However, the difference between the results of the two GAFF models is relatively small and might be biased due to the selection of compounds being compared, as GAFF with BCC charges performs better especially for aliphatic compounds (cf. Table 1), which make up more than a third of all compounds. GAFF-ESP, on the other hand, performs better especially for thiols, amides, and amines, which only stand for 73 compounds in total. Because the spectra derived with the two different charge models mainly differ in the intensity of the peaks, we can speculate that the weaker performance of GAFF-ESP for aliphatic compounds is due to some of the assigned charges being poor representations of the atoms' contribution to the dipole moment (carbon atoms tend to get large positive charges; hydrogen atoms, sometimes negative charges). This would affect aliphatic compounds more strongly than others, as the charges in them are generally small, leading to small deviations having a large relative impact. At the same time, these differences in correlation coefficients that can be observed between the two charge models for GAFF also give a first estimate as to what degree the deficient agreement with the experimental spectra is due to poorly matched intensities. An improved assignment of charges would likely improve the correlation coefficients approximately to this extent. More advanced charge models, e.g., use of polarizable charges, could likely expand this range and allow some further improvement. However, in order to match the performance of the QC methods, likely more fundamental revision of the force fields would be necessary.

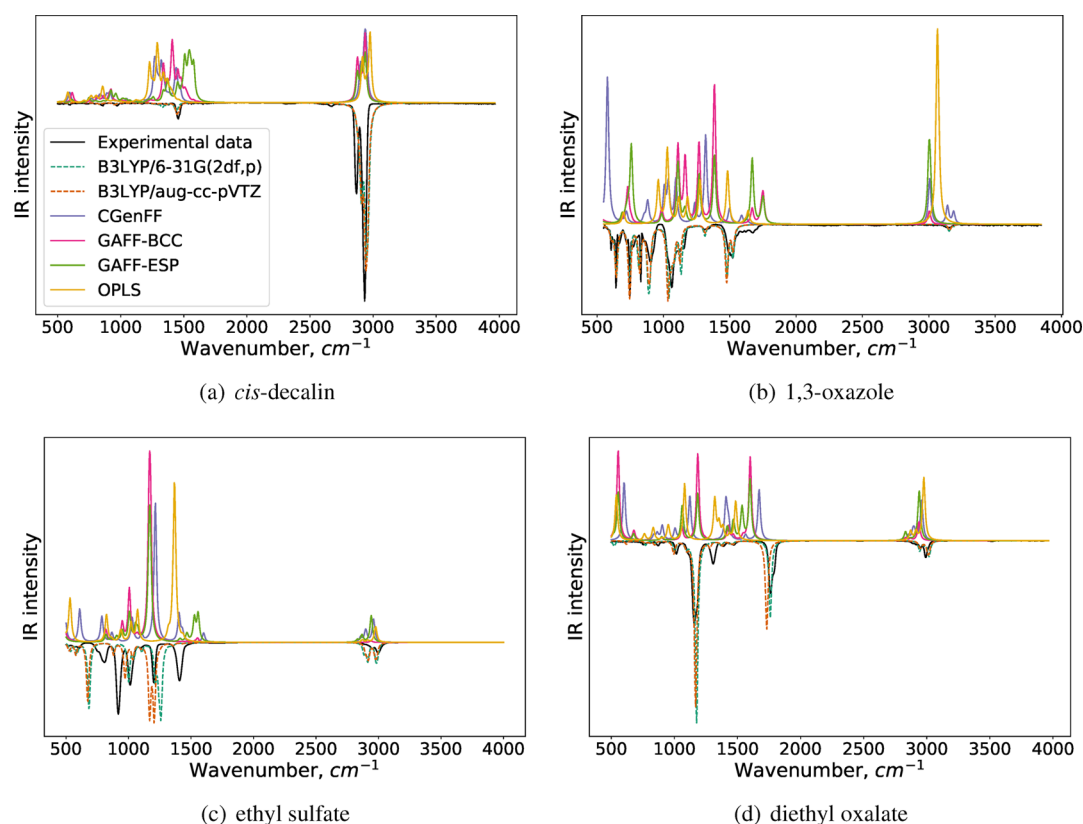
In terms of performance for different classes of compounds, all methods gave comparatively good results for alkenes, and force fields also for alkanes, both independent of the measure used for quantification. On the other hand, arylfluorides were found to be problematic for all methods, and amines for force fields, although less so for quantum chemical methods. In the case of the correlation coefficients found for amines for the quantum chemically derived spectra, it becomes clear that the two statistical measures at least to some degree are sensitive to different properties of the compared spectra. For both quantum chemical methods, the Pearson correlation gives a comparatively low score, while the Spearman correlation coefficient is close to the average for all compounds. These discrepancies become even more clear in the case of cycloalkanes, for which the Pearson correlation is one of the two best and the Spearman correlation one of the two worst for both QC

Table 1. Correlation between Computational and Experimental Spectra for Four Force Field and Two Frequency-Scaled QM Methods by Compound Class<sup>a</sup>

class	N	B3LYP/6-31G(2df,p)		B3LYP/aug-cc-pVTZ		CGenFF		GAFF-BCC		GAFF-ESP		OPLS	
		Pearson <i>r</i>	Spearman $\rho$	Pearson <i>r</i>	Spearman $\rho$	Pearson <i>r</i>	Spearman $\rho$	Pearson <i>r</i>	Spearman $\rho$	Pearson <i>r</i>	Spearman $\rho$	Pearson <i>r</i>	Spearman $\rho$
alcohol	79	0.723	0.803	0.696	0.809	0.243	0.725	0.164	0.612	0.207	0.590	0.193	0.690
aldehyde	14	0.663	<b>0.852</b>	0.682	<b>0.855</b>	<u>0.167</u>	0.679	0.275	<b>0.726</b>	0.258	<b>0.720</b>	0.169	0.677
alkane	85	0.643	0.752	0.692	0.757	0.519	0.732	<b>0.481</b>	0.720	<b>0.438</b>	0.665	<b>0.482</b>	0.731
alkene	179	0.766	0.825	0.789	0.837	0.465	<b>0.773</b>	0.459	0.748	0.406	<b>0.724</b>	0.480	<b>0.753</b>
alkyl bromide	20	0.575	0.784	0.630	0.788	0.473	0.734	0.405	<b>0.727</b>	0.417	0.681	0.424	0.733
alkyl chloride	42	<u>0.543</u>	<u>0.725</u>	0.515	<u>0.727</u>	0.305	0.666	0.246	0.645	0.217	0.604	0.282	0.654
alkylfluoride	9	0.696	0.757	0.520	0.766	0.241	0.644	0.256	<u>0.521</u>	0.223	<u>0.481</u>	0.149	0.630
alkyne	28	0.646	0.778	0.735	0.795	<b>0.528</b>	0.713	0.330	0.628	0.269	0.577	0.422	<u>0.561</u>
amide	12	0.729	<b>0.837</b>	<u>0.474</u>	<b>0.842</b>	0.333	<b>0.769</b>	0.192	0.667	0.206	0.688	0.341	0.723
amine	45	0.593	0.815	0.662	0.810	0.223	<u>0.635</u>	<u>0.042</u>	0.531	<u>0.075</u>	0.543	0.145	0.576
aromatic	184	0.677	0.765	0.667	0.769	0.332	0.681	0.187	0.635	0.188	0.634	0.219	0.650
aryl chloride	23	<u>0.555</u>	0.744	0.578	0.751	0.296	0.649	0.096	0.590	0.110	0.589	0.160	0.626
arylfuoride	15	0.651	0.744	<u>0.503</u>	0.742	0.243	<u>0.568</u>	<u>0.039</u>	<u>0.524</u>	<u>0.026</u>	<u>0.521</u>	<u>0.116</u>	<u>0.551</u>
cycloalkane	41	<b>0.794</b>	<u>0.725</u>	<b>0.839</b>	<u>0.736</u>	0.514	0.701	<b>0.480</b>	0.662	<b>0.430</b>	0.639	0.437	0.676
cycloalkene	11	<b>0.840</b>	0.728	0.779	0.755	0.418	0.669	0.439	0.670	0.382	0.650	0.376	0.650
halogenated compound	106	0.576	0.746	0.557	0.750	0.308	0.658	0.216	0.621	0.202	0.597	0.240	0.640
heterocyclic	65	0.770	0.757	0.743	0.761	0.345	0.673	0.247	0.636	0.225	0.633	0.281	0.658
ketone	34	0.685	0.807	0.613	0.812	0.400	0.743	0.254	0.712	0.238	0.718	0.339	0.720
nitro	11	0.637	0.734	0.585	<u>0.736</u>	<u>0.207</u>	0.730	0.110	0.654	0.151	0.633	0.205	0.707
phenol	18	0.621	0.761	0.567	0.763	0.247	0.638	0.214	0.607	0.189	0.592	<u>0.059</u>	0.577
thiol	16	0.740	0.792	<b>0.813</b>	0.804	<b>0.545</b>	0.743	0.195	0.698	0.385	0.714	<b>0.497</b>	<b>0.755</b>
<i>all molecules</i>	703	0.712 ± 0.154	0.785 ± 0.099	0.710 ± 0.180	0.792 ± 0.099	0.417 ± 0.198	0.726 ± 0.121	0.322 ± 0.235	0.682 ± 0.135	0.307 ± 0.213	0.663 ± 0.135	0.361 ± 0.229	0.694 ± 0.137

<sup>a</sup>Only classes containing at least five compounds are shown. The two highest correlation coefficients of each type for each method are marked in bold; the two lowest, underlined. In the last line, the standard deviation of the correlation coefficient over all molecules is given. Please note that the sum of the compounds in all classes is larger than the total number of molecules, as each molecule can belong to several classes.





**Figure 1.** Infrared spectra calculated by two quantum chemical and four force field methods in comparison with the experimental spectrum. Intensities for the experimental and quantum chemically calculated spectra are depicted negative.

**Table 2. Correlation Coefficients of the Spectra Shown in Figure 1**

method	<i>cis</i> -decalin		1,3-oxazole		ethyl sulfate		diethyl oxalate	
	Pearson $r$	Spearman $\rho$	Pearson $r$	Spearman $\rho$	Pearson $r$	Spearman $\rho$	Pearson $r$	Spearman $\rho$
B3LYP/6-31G(2df,p)	0.740	0.607	0.859	0.605	0.222	0.840	0.846	0.855
B3LYP/aug-cc-pVTZ	0.801	0.602	0.857	0.602	0.309	0.852	0.761	0.853
CGenFF	0.661	0.430	0.231	0.502	0.420	0.873	0.098	0.787
GAFF-BCC	0.652	0.401	0.153	0.421	0.231	0.842	0.321	0.756
GAFF-ESP	0.517	0.306	0.212	0.378	0.213	0.856	0.231	0.756
OPLS	0.331	0.438	0.122	0.440	0.228	0.895	0.103	0.763

methods, and for amides, where the Pearson correlation is one of the lowest and the Spearman correlation one of the highest when the spectra are generated on the B3LYP/aug-cc-pVTZ level. In most cases, however, the two measures agree well in terms of relative ordering of agreement for both classes as well as individual compounds. However, the value of the Spearman correlation is nearly always higher (in the class averaged data, the only exceptions are cycloalkanes, and cycloalkenes for both QC methods, heterocyclic compounds for B3LYP/6-31G(2df,p), and thiols for B3LYP/aug-cc-pVTZ).

In order to understand the differences between the two correlation coefficients, we have selected four compounds for which the scores are distinctively different. These are shown in Figure 1, the corresponding correlation coefficients are listed in Table 2. *cis*-Decalin gives exceptional results in the way that for all but the OPLS-generated spectra the Pearson correlation coefficient is larger than the Spearman correlation coefficient, and above the mean value for all compounds. At the same time, the Spearman correlation is below average for all methods. From Figure 1a it can be seen that the experimental spectrum of

*cis*-decalin is dominated by two overlapping bands just below 3000  $\text{cm}^{-1}$ . This feature is in general reproduced by all methods, with the QC methods placing it at somewhat high wavenumbers, and all force field methods assigning it a far too small intensity as they all produce excessive bands with significant intensity, especially around 1500  $\text{cm}^{-1}$ , where only a small low-intensity band is observed in the experimental and QC spectra. All force field methods except for OPLS place the high-wavenumber feature close to the correct wavenumber, whereas OPLS produces an intensity minimum where the experimental spectrum has its maximum. This is also the only significant and systematic difference between the OPLS generated spectrum and the group of the three other force field generated spectra, which therefore is likely the reason for the large difference in terms of Pearson correlation coefficients (Spearman correlation coefficients are much more similar). It seems that reproduction of a dominant feature in the reference spectrum has a decisive impact on the Pearson correlation coefficient. This is not unexpected, as the coefficient is determined by multiplication of the (mean-adjusted) intensities. At least for

Table 3. Scaling Factors Obtained by Optimizing the Respective Statistical Measure<sup>a</sup>

level of theory	Pearson <i>r</i>			Spearman $\rho$		
	(unscaled)	optimized scaling factor	(scaled)	(unscaled)	optimized scaling factor	(scaled)
B3LYP/6-31G(2df,p)	0.277	0.959 0.965 <sup>b</sup> , 0.955 <sup>c</sup>	0.742 0.712 <sup>b</sup> , 0.713 <sup>c</sup>	0.695	0.963 0.965 <sup>b</sup> , 0.955 <sup>c</sup>	0.786 0.785 <sup>b</sup> , 0.781 <sup>c</sup>
B3LYP/aug-cc-pVTZ	0.341	0.964 0.968 <sup>b</sup> , 0.959 <sup>c</sup>	0.717 0.710 <sup>b</sup> , 0.676 <sup>c</sup>	0.708	0.965 0.968 <sup>b</sup> , 0.959 <sup>c</sup>	0.792 0.792 <sup>b</sup> , 0.789 <sup>c</sup>
CGenFF	0.417	1.002	0.420	0.726	0.997	0.727
GAFF-BCC	0.322	1.002	0.324	0.682	0.997	0.682
GAFF-ESP	0.307	1.005	0.310	0.663	0.993	0.665
OPLS	0.361	0.996	0.377	0.694	0.994	0.697

<sup>a</sup>Database (CCCBDB<sup>43</sup> and UMN<sup>44</sup>) values and correlation coefficients resulting from these for the quantum chemical methods are given in the second line. It should be noted that the exact combination of functional and basis set used here is not available in the UMN database. The value used is given for the marginally larger 6-31G(2df,2p) basis set. <sup>b</sup>CCCBDB. <sup>c</sup>UMN.

this example case, the Spearman correlation coefficient could thus be seen to provide a better estimate of the overall similarity of the spectra.

The second example set of spectra, shown in Figure 1b, is for the compound 1,3-oxazole. In this case the two correlation coefficients give a similar picture for each of the force fields (to the degree that Spearman correlation generally gives higher values than Pearson correlation), i.e., a rather poor agreement with the experimental spectrum. For the QC methods, however, we also observe here that the Pearson correlation coefficients are significantly higher than the Spearman correlation coefficients. For this compound, the experimental spectrum does not have a single dominant feature but shows a multitude of lower intensity bands below approximately 1800 cm<sup>-1</sup>. The majority of these bands is reproduced by the QC derived spectra, albeit with some comparatively minor deviations in both intensities and wavenumbers. Still the comparison between the QC spectra with experimental spectrum produces rather low Spearman correlation coefficients. The reason for this could be that the small range of intensity values covered in these spectra can lead to rather large differences in the rank of a data point, as used for calculation of the Spearman correlation, already from small differences in intensity. In this case, it seems, it is rather the Pearson correlation coefficient that provides an appropriate estimate of the similarity of the spectra.

Figure 1c shows the spectra of ethyl sulfate. For this compound the Spearman correlation coefficients indicate good agreement for all methods, while the Pearson correlation coefficients suggest poor agreement, except for CGenFF where it indicates low to intermediate agreement. Indeed, the agreement of the calculated spectra with the experiment seems rather poor for this compound, except for CGenFF, which reproduces the positions of the bands rather well, even if the agreement of the intensities is not very good. Thus, here it again seems that the Pearson correlation represents the actual similarity between the spectra better, although one might argue that it underestimates the agreement of the CFenFF spectrum somewhat. The considerable overestimation of the agreement of the spectra by the Spearman correlation for this compound might be due to the small number of features, which are mostly concentrated to the low-wavenumber region of the spectrum (below 1500 cm<sup>-1</sup>), leading to separate regions with data points of high and low rank, respectively. Only reproducing this separation into regions is then sufficient to give a high Spearman correlation coefficient, even if the actual bands are poorly reproduced. This effect might actually be the reason behind the observation that the Spearman correlation generally gives

higher scores than the Pearson correlation, as vibrational spectra in general share a common structure with bands below 2000 cm<sup>-1</sup>, around 3000 cm<sup>-1</sup>, and 3600 cm<sup>-1</sup>, and with the remaining regions of the spectrum generally devoid of bands. A similar reason can also be suspected behind the very different scores obtained for the final example shown in Figure 1—diethyl oxalate. Here, the scores obtained for the QC derived spectra match reasonably well, indicating good agreement of the spectra, with the Pearson score being somewhat higher for the spectra obtained with the 6-31G(2df,p) basis set, as the band around 1800 cm<sup>-1</sup> is reproduced better. For the force field methods, on the other hand, the Pearson correlation coefficient indicates close to no correlation at all, with slightly better scores for the GAFF derived spectra, in which the band around 1200 cm<sup>-1</sup> is at least reproduced approximately. The Spearman correlation, however, indicates still an intermediate to good agreement for the force field methods (with the GAFF derived spectra actually having the lowest scores). Once more, one tends to agree more with the Pearson correlation.

As discussed in the previous section, it is known that QC methods give systematic deviations for vibrational frequencies. Several of the factors behind these systematic deviations, also apply directly, e.g., the harmonic approximation used in calculating the Hessian, or indirectly, through parameters derived from QC calculations, to spectra calculated using force field methods. We have therefore tested whether scaling factors, as are used for QC methods, could be obtained for the force field methods used in this study as well. The scaling factors resulting from a maximization of the two statistical measures we are using here to quantify the spectral similarity are shown in Table 3. Since our methodology to derive the scaling factors differs significantly from the method commonly employed for the determination of QC scaling factors, i.e., minimizing the RMSD of peak wavenumbers, we also applied our methodology to the two QC methods used in this study. For the QC methods, both statistical measures gave rise to a scaling factor that only differed insignificantly (cf. ref 54) from the literature values and lie consistently between the values obtained from the two different databases. Use of Spearman's correlation coefficient gives a slightly higher value, i.e., somewhat closer to the scaling factors found in the CCCBDB. At the same time, application of the scaling factor also improves the correlation coefficients considerably (the unscaled QC methods actually perform worse than at least CGenFF). For the force field methods, on the other hand, none of the optimized scaling factors differs significantly from unity, and also the improvement of correlation coefficients from applying them is marginal at best.

**Table 4. Correlation Coefficients Obtained Using the Standard (24 cm<sup>-1</sup>) and Optimized Values of the Full Width at Half-Maximum**

level of theory	Pearson <i>r</i>			Spearman $\rho$		
	(standard)	optimized fwhm (cm <sup>-1</sup> )	(optimized)	(standard)	optimized fwhm (cm <sup>-1</sup> )	(optimized)
B3LYP/6-31G(2df,p)	0.712	65.6	0.791	0.785	29.1	0.786
B3LYP/aug-cc-pVTZ	0.710	68.2	0.792	0.792	27.7	0.792
CGenFF	0.417	90.7	0.490	0.726	47.7	0.728
GAFF-BCC	0.322	150.1	0.433	0.682	124.2	0.687
GAFF-ESP	0.307	149.7	0.410	0.663	38.9	0.663
OPLS	0.361	104.4	0.485	0.694	77.2	0.697

We also attempted the same type of optimization for the width of the Lorentzian used for the calculation of the spectra, the results of which are shown in Table 4. Although, for the Pearson correlation, in all cases a significant improvement of the correlation coefficient could be achieved, the obtained optimized bandwidths are clearly unphysical. Also, in general, the lower the initial correlation score was, the larger the optimized bandwidth was obtained. For the Spearman correlation, the increase in coefficients upon optimization was negligible, and also the dependence on the starting score is nonmonotonic. Still, even using this score for optimization of the bandwidths led, in all cases, to increased bandwidth and, in most cases, to clearly unphysical values. This is very likely caused by the fact that unless the peak positions are very well matched between the calculated and experimental spectrum, a widening of the bands will always lead to an increased overlap between corresponding bands and therefore an improved correlation coefficient. This behavior is more strongly and immediately observed for the Pearson correlation that is derived directly from the intensities than for the Spearman correlation that first converts the values to ranks (in a spectrum containing only a single peak, the ranks of the data points would actually not be affected at all by the bandwidth, as long as the Lorentzian is not truncated). In any case, close to the value of 24 cm<sup>-1</sup> used in this study, the dependence of the correlation coefficients on the width is small for Spearman's correlation. For Pearson's correlation it is larger, but similar for all methods; i.e., within a physically reasonable range the exact value used for the calculation of the spectra does not have a significant impact on the comparison of the different methods. This is demonstrated in Table S2, where the variation of the correlation coefficients for fwhm within the range from 6 to 60 cm<sup>-1</sup> for all methods is given.

## CONCLUSION

We have here demonstrated the calculation of infrared spectra using normal-mode analysis with a number of available general force fields in comparison with quantum chemistry. Also, we have studied the applicability of two different statistical measures—Pearson's product moment correlation and Spearman's rank correlation coefficients—for the specific purpose of quantifying the quality of a calculated spectrum using the gas phase experimental spectrum of the same compound as a reference. Independent of the statistical measure used, it has become clear that the performance of force fields for IR spectrum calculations currently still leaves immense room for improvement. We suggest that future development of force fields takes this into account and makes use of the methodology we have demonstrated in the article.

In terms of applicability of the two statistical measures, we found that both have benefits and drawbacks depending on the

type of spectrum. For spectra with a single (or few) dominant, high-intensity band(s), Pearson's product moment correlation tends to overrate the overlap with this feature, giving a high coefficient even if the remaining smaller bands are only poorly, or not at all, reproduced. Spearman's rank correlation, on the other hand, gives overly high coefficients especially for spectra with only a few features that can be very poorly reproduced by the calculated spectrum. Conversely, Spearman's correlation can be said to better represent approximate matches of bands, whereas Pearson's correlation better indicates agreement of the most dominant feature(s). With this in mind the two measures can be used complementarily. In any case, It should be noted that, in the majority of cases, both coefficients used here give a similar and appropriate picture of the degree to which the spectra agree (with Spearman correlation coefficients generally being somewhat higher than their Pearson counterpart).

Finally, we demonstrated that, through optimization of the statistical measure, one can reproduce the scaling factors for vibrational frequencies used for QC methods and that the vibrational frequencies obtained from force fields do not exhibit similarly systematic deviations from experimental reference. Thus, it is not possible to improve the spectra obtained from force field calculations by application of a scaling factor. Also, optimization of the width of the Lorentzian function used in constructing the spectra is not possible using the correlation coefficients, as with the present large deviations in peak positions the overlap of spectra is increased by widening the bands even into an unphysical regime.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c00126>.

Individual correlation coefficients for all compounds and variation of correlation coefficients with fwhm (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

**Henning Henschel** – Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden; [orcid.org/0000-0001-7196-661X](https://orcid.org/0000-0001-7196-661X); Email: [henning.henschel@icm.uu.se](mailto:henning.henschel@icm.uu.se)

**David van der Spoel** – Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden; [orcid.org/0000-0002-7659-8526](https://orcid.org/0000-0002-7659-8526); Email: [david.vanderspoel@icm.uu.se](mailto:david.vanderspoel@icm.uu.se)



## Authors

Alfred T. Andersson – Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden

Willem Jaspers – Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden;

orcid.org/0000-0002-4951-9220

Mohammad Mehdi Ghahremanpour – Uppsala Center for Computational Chemistry, Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden; Chemistry Department, Yale University, New Haven, Connecticut 06520-8107, United States

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acs.jctc.0c00126>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Funding from eSENCE—The e-Science Collaboration (Uppsala, Lund, UmeÅ, Sweden) is gratefully acknowledged.

## REFERENCES

- (1) Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* **2007**, *44*, 683–700. (Drug Analysis 2006).
- (2) Wadood, S. A.; Boli, G.; Xiaowen, Z.; Hussain, I.; Yimin, W. Recent development in the application of analytical techniques for the traceability and authenticity of food of plant origin. *Microchem. J.* **2020**, *152*, 104295.
- (3) Zhu, M.-Z.; Wen, B.; Wu, H.; Li, J.; Lin, H.; Li, Q.; Li, Y.; Huang, J.; Liu, Z. The Quality Control of Tea by Near-Infrared Reflectance (NIR) Spectroscopy and Chemometrics. *J. Spectrosc.* **2019**, *2019*, 8129648.
- (4) Bel'skaya, L. V. Use of IR Spectroscopy in Cancer Diagnosis. A Review. *J. Appl. Spectrosc.* **2019**, *86*, 187–205.
- (5) Liyanage, S.; Abidi, N. Fourier transform infrared applications to investigate induced biochemical changes in liver. *Appl. Spectrosc. Rev.* **2019**, *0*, 1–33.
- (6) Johnson, J. B.; Naiker, M. Seeing red: A review of the use of near-infrared spectroscopy NIRS in entomology. *Appl. Spectrosc. Rev.* **2019**, *0*, 1–30.
- (7) Ben-Amotz, D. Hydration-Shell Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2019**, *141*, 10569–10580.
- (8) Smith, B. C. *Infrared Spectral Interpretation: A Systematic Approach*; CRC Press: Boca Raton, FL, USA, 1999.
- (9) Byler, D. M.; Susi, H. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **1986**, *25*, 469–487.
- (10) López-Lorente, Á. I.; Mizaikoff, B. Mid-infrared spectroscopy for protein analysis: potential and challenges. *Anal. Bioanal. Chem.* **2016**, *408*, 2875–2889.
- (11) Ghosh, A.; Ostrander, J. S.; Zanni, M. T. Watching Proteins Wiggle: Mapping Structures with Two-Dimensional Infrared Spectroscopy. *Chem. Rev.* **2017**, *117*, 10726–10759.
- (12) Allinger, N. L.; Chen, K. S.; Liu, J. H. An improved force field (MM4) for saturated hydrocarbons. *J. Comput. Chem.* **1996**, *17*, 642–668.
- (13) Nevins, N.; Allinger, N. L. Molecular mechanics (MM4) vibrational frequency calculations for alkenes and conjugated hydrocarbons. *J. Comput. Chem.* **1996**, *17*, 730–746.
- (14) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.
- (15) Halgren, T. A. Merck molecular force field. 3. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553–586.
- (16) van der Spoel, D.; Ghahremanpour, M. M.; Lemkul, J. Small Molecule Thermochemistry: A Tool For Empirical Force Field Development. *J. Phys. Chem. A* **2018**, *122*, 8982–8988.
- (17) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31*, 671–690.
- (18) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General AMBER Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (19) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Compressibility, Expansion Coefficient and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.
- (20) Fischer, N. M.; van Maaren, P. J.; Ditz, J. C.; Yildirim, A.; van der Spoel, D. Properties of liquids in Molecular Dynamics Simulations with explicit long-range Lennard-Jones interactions. *J. Chem. Theory Comput.* **2015**, *11*, 2938–2944.
- (21) Zhang, J.; Tuguldur, B.; van der Spoel, D. Force field benchmark II: Gibbs energy of solvation of organic molecules in organic liquids. *J. Chem. Inf. Model.* **2015**, *55*, 1192–1201.
- (22) Zhang, J.; Tuguldur, B.; van der Spoel, D. Correction to Force field benchmark II: Gibbs energy of solvation of organic molecules in organic liquids. *J. Chem. Inf. Model.* **2016**, *56*, 819–820.
- (23) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 host-guest challenge: Are we doing better? *J. Comput.-Aided Mol. Des.* **2017**, *31*, 1–19.
- (24) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W. Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity. *WIREs Comput. Mol. Sci.* **2019**, *9*, e1393.
- (25) Procacci, P. PrimaDORAC: A Free Web Interface for the Assignment of Partial Charges, Chemical Topology, and Bonded Parameters in Organic or Drug Molecules. *J. Chem. Inf. Model.* **2017**, *57*, 1240–1245.
- (26) Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6665–6670.
- (27) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- (28) Ghahremanpour, M. M.; van Maaren, P. J.; Ditz, J.; Lindh, R.; van der Spoel, D. Large-scale calculations of gas phase thermochemistry: enthalpy of formation, standard entropy and heat capacity. *J. Chem. Phys.* **2016**, *145*, 114305.
- (29) Ghahremanpour, M. M.; van Maaren, P. J.; van der Spoel, D. The Alexandria library: a quantum chemical database of molecular properties for force field development. *Sci. Data* **2018**, *5*, 180062.
- (30) Tanabe, K.; Saeki, S. Computer retrieval of infrared spectra by a correlation coefficient method. *Anal. Chem.* **1975**, *47*, 118–122.
- (31) Ghahremanpour, M. M.; van Maaren, P. J.; van der Spoel, D. Alexandria Library [Data set]. Zenodo. 2017 <http://doi.org/10.5281/zenodo.1004711> (accessed 2020-04-08).
- (32) Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter



generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331–W336.

(33) RDKit: Open-source cheminformatics; <http://www.rdkit.org>, 2017.

(34) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14\*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121*, 3864–3870.

(35) Udier-Blagovic, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. Accuracy of Free Energies of Hydration Using CM1 and CM3 Atomic Charges. *J. Comput. Chem.* **2004**, *25*, 1322–1332.

(36) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(37) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.* **2001**, *7*, 306–317.

(38) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS molecule & liquid database. *Bioinformatics* **2012**, *28*, 752–753.

(39) Scott, A. P.; 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.

(40) Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J. Chem. Theory Comput.* **2010**, *6*, 2872–2887.

(41) Laury, M. L.; Boesch, S. E.; Haken, I.; Sinha, P.; Wheeler, R. A.; Wilson, A. K. 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.

(42) Borowski, P. An Evaluation of Scaling Factors for Multi-parameter Scaling Procedures Based on DFT Force Fields. *J. Phys. Chem. A* **2012**, *116*, 3866–3880.

(43) NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database No. 101, Release 20, August 2019; Johnson, R. D., III, Ed.; National Institute of Standards and Technology, 2017; <http://cccbdb.nist.gov/> (accessed 2020-02-12).

(44) Kanchanakungwankul, S.; Bao, J. L.; Zheng, J.; Alecu, I. M.; Lynch, B. J.; Zhao, Y.; Truhlar, D. G. *Database of Frequency Scale Factors for Electronic Model Chemistries*, Version 4, 2018 <https://comp.chem.umn.edu/freqscale/> (accessed 2020-04-08).

(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. *J. Gaussian 16*, Revision A.03; Gaussian: Wallingford, CT, USA, 2016.

(46) Mott, A. J.; Rez, P. Calculated infrared spectra of nerve agents and simulants. *Spectrochim. Acta, Part A* **2012**, *91*, 256–260.

(47) Chu, P. M.; Guenther, F. R.; Rhoderick, G. C.; Lafferty, W. J. The NIST Quantitative Infrared Database. *J. Res. Natl. Inst. Stand. Technol.* **1999**, *104*, 59–81.

(48) Baumann, K.; Clerc, J. Computer-assisted IR spectra prediction linked similarity searches for structures and spectra. *Anal. Chim. Acta* **1997**, *348*, 327–343.

(49) 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.

(50) Vrančić, 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.

(51) 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.

(52) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmeler, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272.

(53) Kesharwani, M. K.; Brauer, B.; Martin, J. M. L. 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.

(54) Irikura, K. K.; Johnson, R. D.; Kacker, R. N. Uncertainties in Scaling Factors for ab Initio Vibrational Frequencies. *J. Phys. Chem. A* **2005**, *109*, 8430–8437.