

# Harmonic Scale Factors of Fundamental Transitions for Dispersion-corrected Quantum Chemical Methods

Denis S. Tikhonov,<sup>\*,[a, b]</sup> Igor Gordiy,<sup>[c]</sup> Danila A. Iakovlev,<sup>[d]</sup> Alisa A. Gorislav,<sup>[d]</sup> Mikhail A. Kalinin,<sup>[e]</sup> Sergei A. Nikolenko,<sup>[d]</sup> Ksenia M. Malaskeevich,<sup>[d]</sup> Karina Yureva,<sup>[d]</sup> Nikita A. Matsokin,<sup>[b]</sup> and Melanie Schnell<sup>[a, f]</sup>

This work provides a procedure and database for obtaining the vibrational frequency scale factors that align quantum chemically computed harmonic frequencies with experimental vibrational spectroscopic data. The database comprises 441 molecules of various sizes, from diatomics to the buckminsterfullerene C<sub>60</sub>. We provide scale factors for 27 dispersion-corrected methods, 24 of which are DF-Dn/B with DF=BLYP, PBE, B3LYP, PBE0, Dn=D3(BJ), D4, and B=6-31G, def2-SVP, def2-TZVP, and three of them are the 3c-family

composite methods (HF-3c, PBEh-3c, and r<sup>2</sup>SCAN-3c). The two scale factors are derived for each method: the absolute scaling, minimizing the absolute deviation of the scaled harmonic frequency from the experimental value, and the relative scaling, which minimizes an analogous relative deviation. The absolute type of scaling is recommended for frequencies above 2000 cm<sup>-1</sup>, while the relative scaling is optimal for frequencies below 2000 cm<sup>-1</sup>.

## 1. Introduction

Infrared (IR) spectroscopy is one of the important physicochemical methods in the chemists' arsenal.<sup>[1]</sup> This spectroscopic technique observes intra- and intermolecular vibrational motions, as their resonant excitation energies lie in the IR region, i.e., approximately from 100 to 10<sup>4</sup> cm<sup>-1</sup> (wavelengths from 0.1 mm to 780 nm).<sup>[1,2]</sup> In addition to simple absorption/emission IR spectroscopy,<sup>[3]</sup> there are other techniques, such as vibrational circular dichroism (VCD), which is sensitive to molecular chirality,<sup>[4,5]</sup> vibrational Raman spectroscopy,<sup>[6]</sup> inelastic neutron scattering (INS),<sup>[6-8]</sup> infrared/ultraviolet (IR-UV) ion dip spectroscopy,<sup>[9]</sup> infrared multiple photon dissociation (IRMPD) spectroscopy,<sup>[10-14]</sup> various messenger-tagging methods

(such as helium, nitrogen, or hydrogen molecular tag),<sup>[15-20]</sup> that allow measuring spectra of ions, and vibrational sum frequency generation (VSFG) spectroscopy, that allows to probe molecular motions at the phase interfaces.<sup>[21,22]</sup> All these vibrational spectroscopy methods allow us to inspect the properties of molecules, ions, and molecular aggregates in the gas, liquid, and solid phases and at the interfaces between phases. In addition to that, vibrational spectroscopic techniques are widely used in analytical chemistry.<sup>[23]</sup> Even more, an important usage of IR spectroscopy includes astronomy and astrochemistry, with the existence of IR spectrometers on interplanetary probes and of various IR space telescopes, such as the recently retired Stratospheric Observatory For Infrared Astronomy (SOFIA)<sup>[24]</sup> and recently launched James Webb Space Telescope (JWST).<sup>[25,26]</sup>

However, interpreting the IR and other vibrational spectra can be complicated. Despite the existence of various vibrational diagnostic bands for molecular functional groups<sup>[11,12,27]</sup> and databases of the experimental IR spectra, such as the NIST Chemistry WebBook,<sup>[28]</sup> most experimental studies of novel uncharacterized substances have to rely on quantum-chemical calculations.<sup>[5,6,8,9,16-18,29-33]</sup> To achieve a sufficiently good agreement of quantum-chemical calculations and experiment requires the application of expensive computational methods and direct inclusion of anharmonic effects.<sup>[6,8,34,35]</sup> However, such anharmonic calculations require significant expertise and computational efforts, which make them unfeasible for many practical cases.<sup>[35]</sup> Therefore, the most widely used computational approach for the interpretation of experimental spectra is the harmonic frequency calculation with subsequent scaling of the resulting frequencies with a pre-defined global scale factor.<sup>[36-41]</sup> The scale factor accounts for two effects: 1) the quality of the quantum-chemical approximation used and 2) anharmonic effects, which are, by definition, absent in the harmonic calculations.<sup>[36,42]</sup> Therefore, updating the sets of scale

[a] D. S. Tikhonov, M. Schnell

Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany

E-mail: denis.tikhonov@desy.de

[b] D. S. Tikhonov, N. A. Matsokin

Free Moscow University

Homepage: <https://freemoscow.university/>

[c] I. Gordiy

ChemU Corporation Ltd., 3106 Limassol, 17 17 Gr. Xenopoulou St., Cyprus

[d] D. A. Iakovlev, A. A. Gorislav, S. A. Nikolenko, K. M. Malaskeevich, K. Yureva

Independent researcher

[e] M. A. Kalinin

Organic Chemistry Department, Institute of Chemistry, Martin-Luther-University Halle-Wittenberg, Kurt-Mothes-Straße 2, 06120 Halle, Germany

[f] M. Schnell

Institute of Physical Chemistry, Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cphc.202400547>

© 2024 The Authors. ChemPhysChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

factors, considering the expanding corpus of the experimentally determined vibrational frequencies and the new quantum-chemical methods, is an interminable task. This effort is further facilitated by dedicated groups of researchers who diligently monitor and periodically incorporate new updates into these databases to ensure that the scale factors remain current and reflect the latest advances in the field.<sup>[43]</sup>

This work presents a new set of scale factors for 27 quantum-chemical methods. In addition, these new scale factors have been rigorously tested and compared on several experimental data for structures to ensure their accuracy and reliability. The comprehensive examination of these scale factors across various classes of molecules enhances our understanding of their applicability and underscores their potential to advance computational chemistry methodology.

## Methods

### Dataset Selection

The information on the fundamental vibrational transitions of the molecules was extracted from the NIST Computational Chemistry Comparison and Benchmark Database (CCCBDB).<sup>[44]</sup> The species and the gas-phase fundamental transitions were chosen manually with the following guiding principles:

- Only neutral covalently bonded molecules in singlet ground states were selected.
- Only the most abundant isotopologues were taken.
- Experimental fundamental transitions for all normal modes should be assigned. Each diatomic molecule is characterized by a single vibrational number, while all  $N$ -atomic molecules for  $N > 2$  (including linear) are characterized with  $3N - 6$  fundamental modes. In the case of degenerate normal modes, every mode was given. The lowest energy vibration was disregarded for linear molecules, as this additional vibration usually corresponds to a doubly degenerate bending mode.

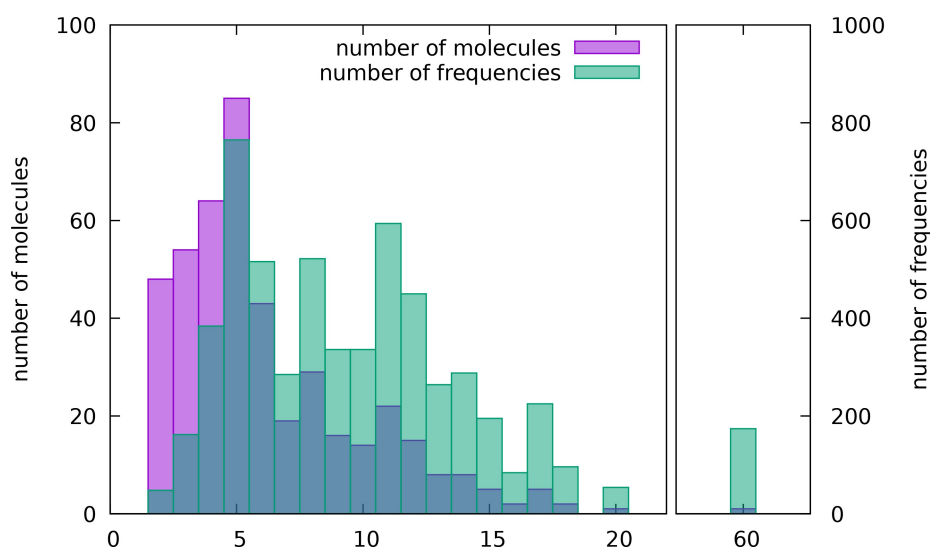
In addition to the molecules from the CCCBDB, nitrobenzene and uracil were added, with the full vibrational assignments given according to Refs. [45–47]. In total, the database contains 441 molecules of various sizes. The distribution of molecular sizes is given in Figure 1. The neutral closed-shell species were chosen for the training set to unify the input files used to calculate the IR spectra.

The experimental and theoretical frequencies for each molecule of the dataset were ordered by their values. Naturally, such a crude procedure creates some misassignments, as some of the vibrational modes in calculations can have different energy ordering compared to the experiment. Nevertheless, such an approach is formalizable and scalable, and the effect of such misassignments should be compensated in large datasets.

### Quantum-chemical Calculations

The first set of 24 methods can be written as “DF-Dn/B”. Here, “DF” denotes a correlation-exchange density functional, “Dn” is a dispersion correction, and “B” is the basis set used. For DF, we chose two pure generalized gradient approximation (GGA) functionals (BLYP<sup>[48,49]</sup> and PBE<sup>[50]</sup>) and their popular hybrid counterparts (B3LYP<sup>[51]</sup>, PBE0<sup>[52]</sup>). As dispersion corrections (Dn), we considered D3(BJ)<sup>[53]</sup> and the newly developed D4<sup>[54]</sup> corrections. As the basis sets (B), we used a small 6–31G basis set<sup>[55,56]</sup> and two Karlsruhe-type basis sets,<sup>[57]</sup> def2-SVP and def2-TZVP. The second set of three methods includes three examples of 3c-composite methods, namely HF-3c,<sup>[58]</sup> PBEh-3c,<sup>[59]</sup> and r<sup>2</sup>SCAN-3c.<sup>[60]</sup> The harmonic vibrational frequencies of the most abundant isotopologues after the geometry optimization at the corresponding level of theory were done using the ORCA 5 quantum-chemical software.<sup>[61,62]</sup> With elements of the fifth row and beyond, the automatically chosen by ORCA effective core potentials def2-ECP for these elements were applied.<sup>[57,63,64]</sup> For potentially conformationally flexible molecules, an initial conformational search was performed at the GFN2-xTB<sup>[65]</sup> level of theory using the CREST<sup>[66,67]</sup> software. Further optimization and harmonic frequency calculations were done for the lowest-energy conformer found.

The calculations were performed in the same fashion but with a few minor differences for the illustrative molecular systems (pyrene, protonated LeuEnk, and doubly protonated gramicidin S, see



**Figure 1.** Number of molecules and corresponding number of vibrational frequencies comprised of a given number of atoms in the final dataset of fundamental frequencies. Note, that the two histograms have different y-axes.

Figure 2). For pyrene, in addition to harmonic spectra at all 27 chosen methods, the intensities of the overtones and combination bands at the  $r^2$ SCAN-3c level of theory were computed with the "NearIR" procedure of ORCA.<sup>[68]</sup> For protonated LeuEnk, only GGA and 3c-methods were tested. Nine representative conformers, given in Ref. [33], were optimized at the corresponding level of theory, and then their IR harmonic spectra were averaged with the weights provided in Ref. [33]. We did not repeat the extensive work on the conformational search for the LeuEnk, relying on the results from the corresponding work.<sup>[33]</sup> For the doubly protonated gramicidin S, the given conformer from Ref. [30] was used after re-optimization at several levels of theory from our list of methods. The IR spectra for these three systems were convoluted with the Gaussian functions with full width at half-maximum FWHM = 30  $\text{cm}^{-1}$  for pyrene, FWHM = 110  $\text{cm}^{-1}$  for protonated LeuEnk, and FWHM = 10  $\text{cm}^{-1}$  for gramicidin S, to match the widths observed in the experimental spectra.

### Least-squares Fitting of the Scale Factors

The goal of the optimization of the scale factors  $s$  was to adjust every quantum-chemically computed harmonic frequency  $\nu^{(\text{calc})}$  at a given level of theory to the experimental fundamental transition frequency  $\nu^{(\text{exp})}$  as<sup>[38,41]</sup>

$$s \cdot \nu^{(\text{calc})} = \nu^{(\text{exp})}. \quad (1)$$

To find the best scale factors that approximately fulfill the requirement given in Equation 1, two types of least-squares (LSQ) fitting were applied, which minimized a given root-mean-square deviation (RMSD). The first type of scale factor that we obtained is denoted as the absolute scale factor ( $s_{\text{abs}}$ ). The second type of scale factor is called the relative scale factor ( $s_{\text{rel}}$ ).

The scale factor  $s_{\text{abs}}$  was obtained by minimizing the following absolute RMSD (aRMSD):<sup>[38,41]</sup>

$$\text{aRMSD}^2 = \frac{1}{N_{\text{tot}}} \sum_{m=1}^M \sum_{k=1}^{K_m} \left( s_{\text{abs}} \cdot \nu_{mk}^{(\text{calc})} - \nu_{mk}^{(\text{exp})} \right)^2 \rightarrow \min, \quad (2)$$

where  $m$  enumerates  $M$  molecules in the training set,  $k$  enumerates the computed and experimental frequencies for the  $m$ -th molecule, and  $N_{\text{tot}} = \sum_{m=1}^M K_m$  is the total number of frequencies in the training set. The  $s_{\text{rel}}$  was obtained by minimizing a relative RMSD (rRMSD):

$$\text{rRMSD}^2 = \frac{1}{N_{\text{tot}}} \sum_{m=1}^M \sum_{k=1}^{K_m} \left( s_{\text{rel}} \cdot \frac{\nu_{mk}^{(\text{calc})}}{\nu_{mk}^{(\text{exp})}} - 1 \right)^2 \rightarrow \min. \quad (3)$$

The optimal values of the  $s_{\text{abs}}$  and  $s_{\text{rel}}$  and their fit uncertainties  $\sigma_{\text{abs}}$  and  $\sigma_{\text{rel}}$  according to Equations 2 and 3 are given as<sup>[38,41]</sup>

$$s_{\text{abs}} \pm \sigma_{\text{abs}} = \frac{\sum_{m=1}^M \sum_{k=1}^{K_m} \nu_{mk}^{(\text{calc})} \nu_{mk}^{(\text{exp})}}{\sum_{m=1}^M \sum_{k=1}^{K_m} \left( \nu_{mk}^{(\text{calc})} \right)^2} \pm \frac{\text{aRMSD}_{\text{min}}}{\sqrt{\sum_{m=1}^M \sum_{k=1}^{K_m} \left( \nu_{mk}^{(\text{calc})} \right)^2}} \quad (4)$$

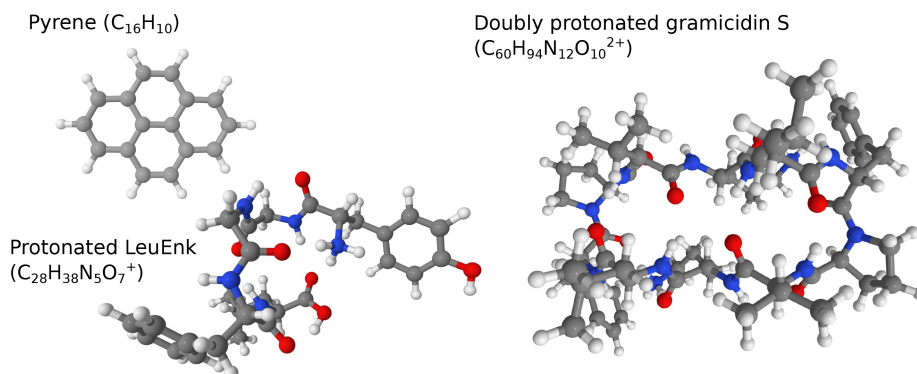
and

$$s_{\text{rel}} \pm \sigma_{\text{rel}} = \frac{\sum_{m=1}^M \sum_{k=1}^{K_m} \nu_{mk}^{(\text{calc})} / \nu_{mk}^{(\text{exp})}}{\sum_{m=1}^M \sum_{k=1}^{K_m} \left( \nu_{mk}^{(\text{calc})} / \nu_{mk}^{(\text{exp})} \right)^2} \pm \frac{\text{rRMSD}_{\text{min}}}{\sqrt{\sum_{m=1}^M \sum_{k=1}^{K_m} \left( \nu_{mk}^{(\text{calc})} / \nu_{mk}^{(\text{exp})} \right)^2}}, \quad (5)$$

where  $\text{aRMSD}_{\text{min}}$  and  $\text{rRMSD}_{\text{min}}$  denote minimal values of Equations 2 and 3, obtained with optimal  $s_{\text{abs}}$  and  $s_{\text{rel}}$  according to Equations 4 and 5, respectively.

## 2. Results and Discussion

The resulting scaling factors and corresponding aRMSD values obtained by the procedure described in Section 2 are given in Table 1. The aRMSD values can be thought of as the expected accuracy of the given combination of the method and the scale factor, i.e., the deviations of the scaled spectra from the experimental ones are expected to be within the  $\pm \text{aRMSD}$  interval. There are a few trends that can be observed. First of all,  $s_{\text{rel}} \geq s_{\text{abs}}$  for each given method. Secondly, the scale factors for the DF–D3(BJ)/B methods are within accuracy compared to DF–D4/B, suggesting that the effects of D3(BJ) and D4 corrections are similar. The last trend observed is a systematic decrease of the aRMSD values upon improving the basis set quality, namely  $\text{aRMSD}(\text{DF-D}n/\text{def2-TZVP}) \leq \text{aRMSD}(\text{DF-D}n/\text{def2-SVP}) \leq \text{aRMSD}(\text{DF-D}n/6-31\text{G})$ .



**Figure 2.** Structures and chemical formulas of three molecules that were used to illustrate the performance of the scale factors obtained in this work. The colors of spheres denote the atom type: white denotes hydrogens, gray denotes carbons, blue denotes nitrogens, and red denotes oxygens.

**Table 1.** Optimal absolute ( $s_{\text{abs}}$ ) and relative ( $s_{\text{rel}}$ ) harmonic frequency scaling factors were determined for the listed quantum-chemical approximations with Equations 4 and 5, respectively. aRMSD denotes the values of the absolute root-mean-square deviations (Equation 2) computed for the dataset sets of frequencies with the corresponding scale factors.

DF	Method		Absolute scaling		Relative scaling	
	$D_n$	Basis	$s_{\text{abs}}$	aRMSD, $\text{cm}^{-1}$	$s_{\text{rel}}$	aRMSD, $\text{cm}^{-1}$
BLYP	D3BJ	6-31G	0.9911(6)	72	1.018(2)	84
		def2-SVP	0.9972(3)	41	1.013(1)	48
		def2-TZVP	0.9960(3)	41	1.017(1)	52
	D4	6-31G	0.9912(6)	72	1.017(2)	83
		def2-SVP	0.9973(3)	41	1.013(1)	48
		def2-TZVP	0.9961(3)	41	1.017(1)	52
PBE	D3BJ	6-31G	0.9848(6)	67	1.010(2)	79
		def2-SVP	0.9904(4)	42	1.006(1)	48
		def2-TZVP	0.9923(3)	39	1.011(1)	48
	D4	6-31G	0.9849(6)	67	1.010(2)	78
		def2-SVP	0.9905(4)	42	1.005(1)	48
		def2-TZVP	0.9925(3)	39	1.010(1)	47
B3LYP	D3BJ	6-31G	0.9611(5)	63	0.981(1)	71
		def2-SVP	0.9664(3)	36	0.975(1)	39
		def2-TZVP	0.9671(3)	32	0.979(1)	37
	D4	6-31G	0.9613(5)	63	0.981(1)	72
		def2-SVP	0.9666(3)	36	0.976(1)	39
		def2-TZVP	0.9673(3)	32	0.979(1)	37
PBE0	D3BJ	6-31G	0.9509(5)	59	0.969(1)	66
		def2-SVP	0.9555(3)	39	0.963(1)	41
		def2-TZVP	0.9591(3)	33	0.967(1)	36
	D4	6-31G	0.9510(5)	59	0.969(1)	66
		def2-SVP	0.9556(3)	39	0.963(1)	41
		def2-TZVP	0.9591(3)	33	0.967(1)	36
HF-3c			0.8400(6)	82	0.858(1)	89
PBEh-3c			0.9329(3)	33	0.935(1)	33
r <sup>2</sup> SCAN-3c			0.9688(3)	35	0.982(1)	41

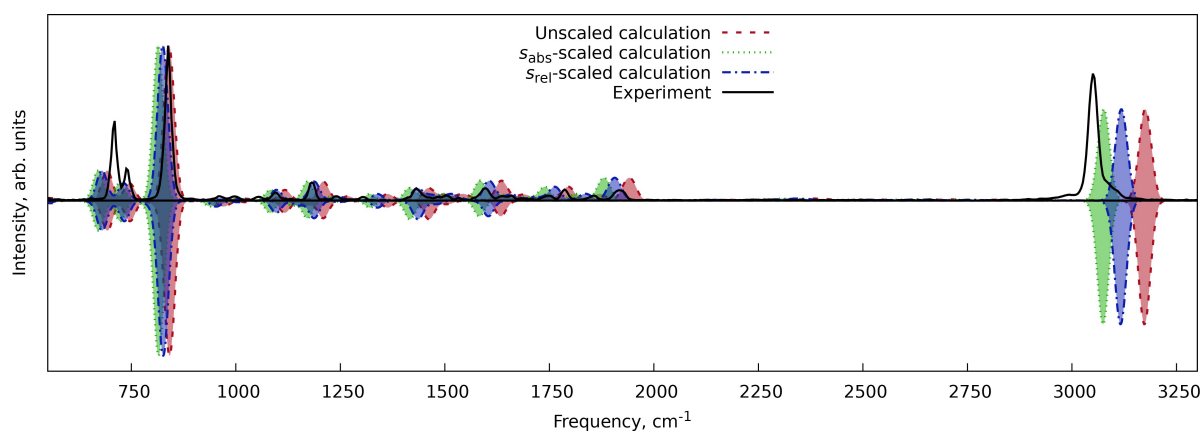
In order to illustrate the applicability of the obtained scale factors, we performed calculations for three molecular systems of various sizes and complexity for which experimental vibrational spectra are available from the literature (Figure 2). Here, we will demonstrate only the results at the r<sup>2</sup>SCAN-3c level of theory, and all the other comparisons can be found in ESI.

The first system was pyrene (C<sub>16</sub>H<sub>10</sub>), a four-ring polycyclic aromatic hydrocarbon, for which the experimental gas-phase IR spectra were available on the NIST Chemistry WebBook.<sup>[28]</sup> The comparison of the harmonic spectra with fundamentals only and with overtones and combination bands is shown in Figure 3. We can see that the same scale factors can be used to adjust the positions of the overtones and combination bands while no anharmonic corrections are applied. The  $s_{\text{abs}}$  demonstrates the best performance for the C–H stretching vibrations (above 2000  $\text{cm}^{-1}$ ), while the  $s_{\text{rel}}$  performs best at the fingerprint

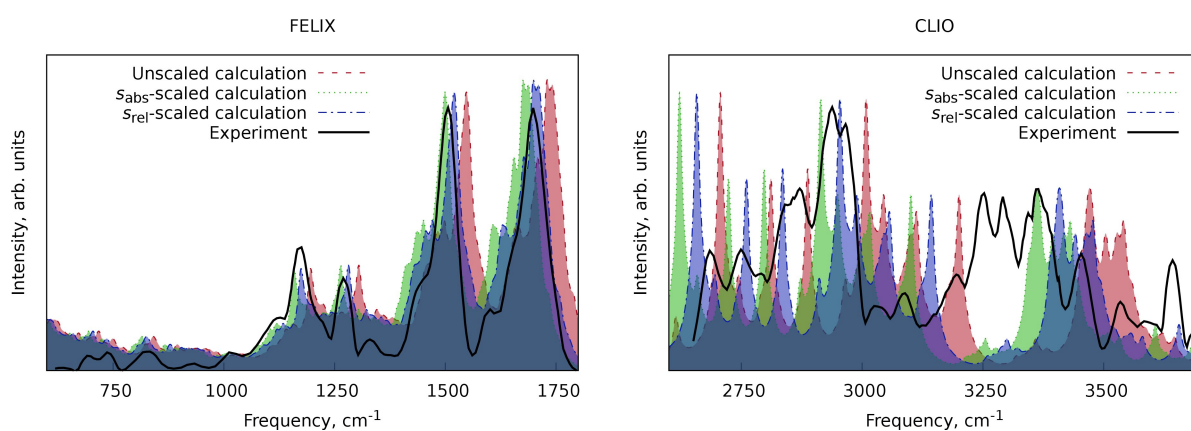
region below 2000  $\text{cm}^{-1}$ , accounting for a generally better representation of the spectral features.

A similar pattern can be observed for the protonated version of the LeuEnk polypeptide (C<sub>28</sub>H<sub>38</sub>N<sub>5</sub>O<sub>7</sub><sup>+</sup>) in the spectra in the X–H stretching region (X=C, N, O) taken with the CLIO setup<sup>[70]</sup> and in the spectra in the fingerprint region taken at the FELIX facility<sup>[71]</sup> (Figure 4). The  $s_{\text{rel}}$ -scaled spectrum best fits to the lower energy spectral region compared to the  $s_{\text{abs}}$ -scaled one. The X–H stretching region has a lot of bands, probably arising from a complicated conformational composition, resonances, and overtones, which makes this region more difficult to analyze. Nevertheless, for the most visible bands in the harmonic calculations, the  $s_{\text{abs}}$ -scaled spectrum better matches in peak positions than the  $s_{\text{rel}}$ -scaled one.

Our last example was the double protonated conformer A of gramicidin S (C<sub>60</sub>H<sub>94</sub>N<sub>12</sub>O<sub>10</sub><sup>2+</sup>), for which the fingerprint region



**Figure 3.** Comparison of the scaled and unscaled harmonic IR spectrum of pyrene at the  $r^2$ SCAN-3c level of theory with the experimental data from NIST Chemistry WebBook.<sup>[28]</sup> The top theoretical trace shows a spectrum with the harmonic fundamentals, overtones, and combination bands, while the lower trace shows the harmonic spectrum with fundamental transitions only.



**Figure 4.** Comparison of the scaled and unscaled harmonic IR spectrum of the closed form of the protonated LeuEnk polypeptide at the  $r^2$ SCAN-3c level of theory with the experimental data from Ref. [33]. The numerical experimental data were extracted from the manuscript images using WebPlotDigitizer software.<sup>[69]</sup>

helium-tagging action spectrum was available.<sup>[30]</sup> Here (Figure 5), we observe the same trend as in the previous examples (Figures 3 and 4): the  $s_{rel}$ -scaled spectrum shows a good agreement with peak positions, while the  $s_{abs}$ -scaled spectrum overestimates the red shift of the peak positions.

### 3. Conclusions

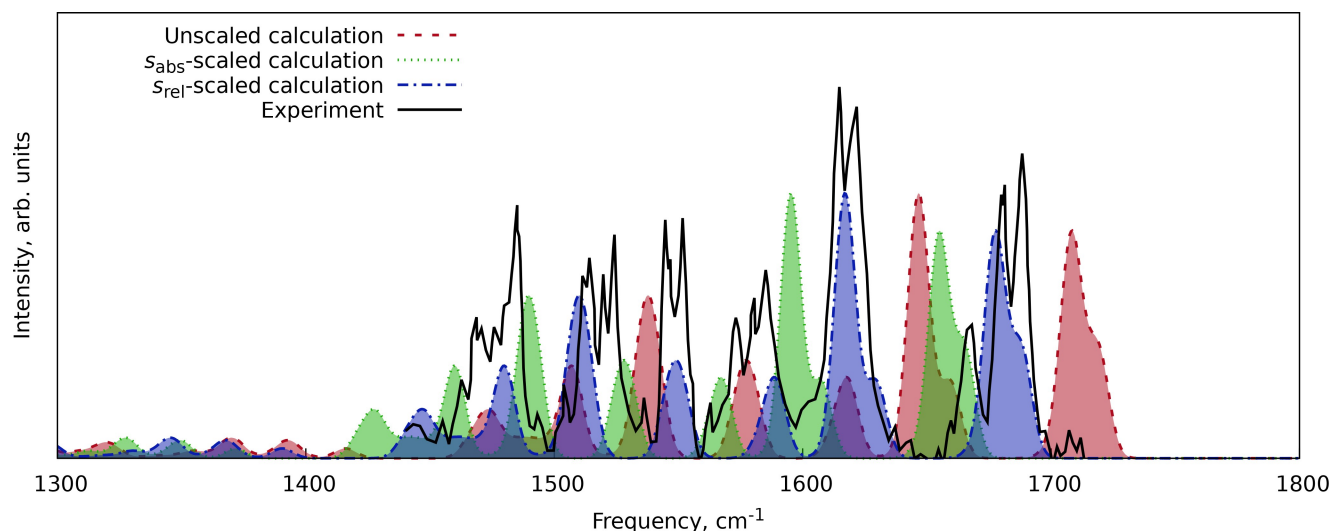
Here, we presented the dataset and the procedure for obtaining scale factors for the IR vibrational spectra. We developed two sets of scale factors, derived from two different principles: absolute scaling ( $s_{abs}$ ), by taking into account absolute deviations of the theoretical frequency from the experimental ones, and relative scaling ( $s_{rel}$ ), which minimizes the relative deviation of one frequency from another. The numerical demonstrations with three molecular systems of various sizes (pyrene and two ionic peptides: protonated LeuEnk and doubly protonated gramicidin S) show that the  $s_{abs}$  provides better results for the hydrogen-stretching region (above  $2000\text{ cm}^{-1}$ ), while the  $s_{rel}$  outperforms  $s_{abs}$  in the fingerprint region (below  $2000\text{ cm}^{-1}$ ).

Thus, we recommend joint usage of these scale factors to improve the predictions for different spectral regions. The RMSD values accompanying these scale factors can be used to know the appropriate deviation of the observable band positions from the scaled harmonic frequencies at a given level of theory. These considerations should always be taken into account whenever the experimental data is interpreted.

The dataset and scripts we used for quantum-chemical calculations, data extraction, and scale factor fitting are available on Zenodo (<https://doi.org/10.5281/zenodo.11174948>). This data can be used to obtain the scale factors for other levels of theory.

### Appendix

We perform LSQ fitting of the parameter  $s$  by minimizing the following function<sup>[72,73]</sup>



**Figure 5.** Comparison of the scaled and unscaled harmonic IR spectrum of double protonated conformer A of gramicidin S at the  $r^2$ SCAN-3c level of theory with the experimental data from Ref. [30]. The numerical experimental data were extracted from the manuscript images using WebPlotDigitizer software.<sup>[69]</sup>

$$\Phi(s) = \frac{1}{2} \sum_{i=1}^N \frac{(s \cdot x_i - y_i)^2}{\sigma_0^2}, \quad (6)$$

where  $N$  is the total number of theoretical and experimental values,  $x$  is the scaled property,  $y$  is the reference data point, index  $1 \leq i \leq N$  enumerates the data points, and  $\sigma$  is the initially unknown uncertainty of all data points. In the case of absolute scale factor fitting,  $x_i = \nu_i^{(\text{calc})}$  and  $y_i = \nu_i^{(\text{exp})}$ , while for relative scale factor fitting,  $x_i = \nu_i^{(\text{calc})}/\nu_i^{(\text{exp})}$  and  $y_i = 1$ .

The function  $\Phi(s)$  can be interpreted through the maximal likelihood principle as the negative logarithm of the probability  $p(s)$  of the solution  $s$ , namely<sup>[72,73]</sup>

$$\Phi(s) = \ln(\mathcal{N}) - \ln(p(s)), \quad (7)$$

where  $\mathcal{N}$  is the normalization constant.

We can find the minimal value of the function from Equation 6 as  $d\Phi(s)/ds = 0$ , which gives the solution<sup>[72,73]</sup>

$$s_{\min} = \frac{S_{xy}}{S_{xx}}, \quad (8)$$

where  $S_{xy} = \sum_{i=1}^N x_i y_i$  and  $S_{xx} = \sum_{i=1}^N x_i^2$ . Now, we can estimate the value of uncertainty  $\sigma$  through the minimal value of the RMSD given as

$$\sigma_0^2 = \text{RMSD}^2 = \frac{1}{N} \sum_{i=1}^N (s_{\min} \cdot x_i - y_i)^2 = \frac{1}{N} \left( S_{yy} - \frac{S_{xy}^2}{S_{xx}} \right), \quad (9)$$

where  $S_{yy} = \sum_{i=1}^N y_i^2$ .

Representing  $s$  as  $s = s_{\min} + \delta s$ , where  $\delta s$  is the deviation from the optimal LSQ solution (Equation 8), we can rewrite the original function  $\Phi(s)$  (Equation 6) as<sup>[72,73]</sup>

$$\Phi(\delta s) = \frac{N}{2} + \frac{S_{xx} \delta s^2}{2\sigma_0^2} = \frac{N}{2} + \frac{\delta s^2}{2\sigma_s^2}. \quad (10)$$

Substitution of this equation into Equation 7 provides that the probability distribution of the deviation  $\delta s$  is a normal distribution with variance  $\sigma_s$ , given as

$$\sigma_s^2 = \frac{\sigma_0^2}{S_{xx}}. \quad (11)$$

Combining Equation 8 and 11 provides us with a final LSQ solution for scale factor  $s$  as  $s = s_0 \pm \sigma_s$ .

## Acknowledgements

DST and MS acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF. In particular, DST's calculations were enabled through the Maxwell computational resources operated at DESY. AG would like to personally thank Roman A. Makarov for providing help with the scripts and Anna Yarullina for grammar review. MAK expresses sincere gratitude to Daniil Konovalenko for the help in writing a script for automizing calculations in different methods. We are also grateful to the National Institute of Standards for its excellent work maintaining and updating important databases, including CCCBDB and Chemistry WebBook. Open Access funding enabled and organized by Projekt DEAL.

## Conflict of Interests

The authors declare no conflict of interest.

## Data Availability Statement

The quantum-chemical calculations, database, and scripts used for obtaining and analyzing the data are available on Zenodo (<https://doi.org/10.5281/zenodo.11174948>). Additional data visualization and detailed representation of the results are provided in the ESI.

**Keywords:** DFT (density functional theory) · Scale factors · Infrared spectroscopy · Molecular vibrations · Fitting · Dataset

- Barbara H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications. Introduction*, chapter 1, pages 1–13. John Wiley & Sons, Ltd, 2004. ISBN 9780470011140. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011149.ch1>.
- International Union of Pure and Applied Chemistry (IUPAC). 2019. URL <https://doi.org/10.1351/goldbook.IT07399>.
- Barbara H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications. Experimental Methods*, chapter 2, pages 15–44. John Wiley & Sons, Ltd, 2004. ISBN 9780470011140. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011149.ch2>.
- Laurence A. Nafie, Teresa B. Freedman, *Enantiomer* **1998**, 3(4-5), 283–297, ISSN 1024–2430. URL <http://europemc.org/abstract/MED/9861702>.
- Sascha Jähnigen, *Angew. Chem. Int. Ed.* **2023**, 62(41), e202303595, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.202303595>.
- Kacper Drużbicki, Edward Mikuli, Norbert Pałka, Sławomir Zalewski, Mirosława D. Ossowska-Chruściel, *J. Phys. Chem. B* **2015**, 119(4), 1681–1695, URL <https://doi.org/10.1021/jp507241j>. PMID: 25564699.
- Carla Andreani, Maciej Krzysztyniak, Giovanni Romanelli, Roberto Senesi, Felix Fernandez-Alonso, *Adv. Phys.* **2017**, 66(1), 1–73, URL <https://doi.org/10.1080/00018732.2017.1317963>.
- Kacper Drużbicki, Roberto Simone Pinna, Svemir Rudić, Marek Jura, Giuseppe Gorini, Felix Fernandez-Alonso, *J. Phys. Chem. Lett.* **2016**, 7(22), 4701–4709, URL <https://doi.org/10.1021/acs.jpcclett.6b01822>.
- E. Cristina Stanca-Kaposta, John P. Simons. *High-resolution Infrared–Ultraviolet (IR–UV) Double-resonance Spectroscopy of Biological Molecules*. John Wiley & Sons, Ltd, 2011. ISBN 9780470749593. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470749593.hrs096>.
- Nick C. Polfer, Jos Oomens, *Mass Spectrom. Rev.* **2009**, 28(3), 468–494, URL <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/mas.20215>.
- Nick C. Polfer, *Chem. Soc. Rev.* **2011**, 40, 2211–2221, URL <https://doi.org/10.1039/C0CS00171F>.
- Matthew J. Carlo, Amanda L. Patrick, *J. Mass Spectrom. Adv. Clin. Lab.* **2022**, 23, 14–25, ISSN 2667-145X. URL <https://www.sciencedirect.com/science/article/pii/S2667145X2100033X>.
- Jos Oomens, André J. A. van Rooij, Gerard Meijer, Gert von Helden, *Astron. J.* **2000**, 542(1), 404, URL <https://dx.doi.org/10.1086/309545>.
- Jongcheol Seo, Waldemar Hoffmann, Stephan Warnke, Xing Huang, Sandy Gewinner, Wieland Schöllkopf, Michael T. Bowers, Gert von Helden, Kevin Pagel, *Nature Chemistry* **2017**, 9(1), 39–44, ISSN 1755–4349. URL <https://doi.org/10.1038/nchem.2615>.
- Arron B. Wolk, Christopher M. Leavitt, Etienne Garand, Mark A. Johnson, *Acc. Chem. Res.* **2014**, 47(1), 202–210, URL <https://doi.org/10.1021/ar400125a>. PMID: 23972279.
- Jana Roithová, Andrew Gray, Erik Andris, Juraj Jašík, Dieter Gerlich, *Acc. Chem. Res.* **2016**, 49(2), 223–230, URL <https://doi.org/10.1021/acs.accounts.5b00489>. PMID: 26821086.
- Valeriu Scutelnic, Marta A. S. Perez, Mateusz Marianski, Stephan Warnke, Aurelien Gregor, Ursula Rothlisberger, Michael T. Bowers, Carsten Baldauf, Gert von Helden, Thomas R. Rizzo, Jongcheol Seo, *J. Am. Chem. Soc.* **2018**, 140(24), 7554–7560, URL <https://doi.org/10.1021/jacs.8b02118>. PMID: 29637771.
- Aaron Calvin, Scott Eierman, Zeyun Peng, Merrell Brzeczek, Lincoln Satterthwaite, David Patterson, *Nature* **2023**, 621(7978), 295–299. ISSN 1476–4687. URL <https://doi.org/10.1038/s41586-023-06351-7>.
- Conrad T. Wolke, Joseph A. Fournier, Laura C. Dzugan, Matias R. Fagiani, Tuguldur T. Odbadrakh, Harald Knorke, Kenneth D. Jordan, Anne B. McCoy, Knut R. Asmis, Mark A. Johnson *Science* **2016**, 354(6316), 1131–1135, URL <https://www.science.org/doi/abs/10.1126/science.aaf8425>.
- Fabian Müller, Julius B. Stückerath, Florian A. Bischoff, Laura Gagliardi, Joachim Sauer, Sreekanta Debnath, Marcel Jorewitz, Knut R. Asmis, *J. Am. Chem. Soc.* **2020**, 142(42), 18050–18059, URL <https://doi.org/10.1021/jacs.0c07158>. PMID: 33031700.
- James D. Pickering, Mikkel Bregnhøj, Adam S. Chatterley, Mette H. Rasmussen, Kris Strunge, Tobias Weidner, *Biointerphases* **2022**, 17(1), 011201, 01. ISSN 1934–8630. URL <https://doi.org/10.1116/6.0001401>.
- Hong-Fei Wang, *Prog. Surf. Sci.* **2016**, 91(4), 155–182, ISSN 0079–6816. URL <https://www.sciencedirect.com/science/article/pii/S0079681616300259>.
- Barbara H. Stuart, *Industrial and Environmental Applications*, chapter 8, pages 167–186. John Wiley & Sons, Ltd, 2004. ISBN 9780470011140. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011149.ch8>.
- Robert D. Gehrz, Erick E. Becklin, Imke de Pater, Dan F. Lester, Thomas L. Roellig, Charles E. Woodward, *Adv. Space Res.* **2009**, 44(4), 413–432, ISSN 0273–1177. doi: 10.1016/j.asr.2009.04.011. URL <https://www.sciencedirect.com/science/article/pii/S0273117709002506>.
- Thomas J. Fauchez, Geronimo L. Villanueva, Denis E. Sergeev, Martin Turbet, Ian A. Boutle, Kostas Tsigaridis, Michael J. Way, Eric T. Wolf, Shawn D. Domagal-Goldman, François Forget, Jacob Haqq-Misra, Ravi K. Kopparapu, James Manners, Nathan J. Maynem, *Plant Sci.* **2022**, 3(9), 213, URL <https://dx.doi.org/10.3847/PSJ/ac6cf1>.
- William Pluriel, Jérémy Leconte, Vivien Parmentier, Tiziano Zingales, Aurélien Falco, Franck Selsis, Pascal Bordé, *Astron. Astrophys.* **2022**, 658, 17(A42), URL <https://doi.org/10.1051/0004-6361/202141943>.
- John Coates, *Interpretation of Infrared Spectra, A Practical Approach*. John Wiley & Sons, Ltd, 2006. ISBN 9780470027318. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470027318.a5606>.
- Henry M. Rosenstock, Keith Draxl, Bruce W. Steiner, John T. Herron, “Ion Energetics Data” in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P. J. Linstrom and W. G. Mallard, National Institute of Standards and Technology, Gaithersburg MD, 20899, <https://doi.org/10.18434/T4D303>, (retrieved April 13, 2023).
- Natalia S. Nagornova, Matteo Guglielmi, Manuel Doemer, Ivano Tavernelli, Ursula Rothlisberger, Thomas R. Rizzo, Oleg V. Boyarkin, *Angew. Chem. Int. Ed.* **2011**, 50(23), 5383–5386, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201100702>.
- Natalia S. Nagornova, Thomas R. Rizzo, Oleg V. Boyarkin, *J. Am. Chem. Soc.* **2010**, 132(12), 4040–4041, doi: 10.1021/ja910118j. PMID: 20201525.
- Lara van Tetering, Sylvia Spies, Quirine D. K. Wildeman, Kas J. Houthuijs, Rianne E. van Outersterp, Jonathan Martens, Ron A. Wevers, David S. Wishart, Giel Berden, Jos Oomens, *Commun. Chem.* **2024**, 7(1), 30, ISSN 2399–3669. URL <https://doi.org/10.1038/s42004-024-01112-7>.
- Arkadiusz Jarota, Daria Drwal, Jakub Piketa, Ewa Pastorczak, *Sci. Rep.* **2022**, 12(1), 16834, ISSN 2045–2322. URL <https://doi.org/10.1038/s41598-022-20264-x>.
- Amir Kotobi, Lucas Schwob, Gregor B. Vonbun-Feldbauer, Mariana Rossi, Piero Gasparotto, Christian Feiler, Giel Berden, Jos Oomens, Bart Oostenrijck, Debora Scuderi, Sadia Bari, Robert H. Meißner, *Commun. Chem.* **2023**, 6(1), 46, ISSN 2399–3669. URL <https://doi.org/10.1038/s42004-023-00835-3>.
- Kacper Drużbicki, Jadwiga Mielcarek, Anna Kiwilsza, Loic Toupet, Eric Collet, Aleksandra Pajzderska, Jan Wasicki, *Cryst. Growth Des.* **2015**, 15(6), 2817–2830, URL <https://doi.org/10.1021/acs.cgd.5b00251>.
- Denis S. Tikhonov, Dmitry I. Sharapa, Jan Schwabedissen, Vladimir V. Rybkin, *Phys. Chem. Chem. Phys.* **2016**, 18, 28325–28338, URL <https://doi.org/10.1039/C6CP05849C>.
- Manoj K. Kesharwani, Brina Brauer, Jan M. L. Martin, *J. Phys. Chem. A* **2015**, 119(9), 1701–1714, URL <https://doi.org/10.1021/jp508422u>. PMID: 25296165.
- Ionut M. Alecu, Jingjing Zheng, Yan Zhao, Donald G. Truhlar, *J. Chem. Theory Comput.* **2010**, 6(9), 2872–2887, URL <https://doi.org/10.1021/ct100326h>. PMID: 26616087.
- Marie L. Laury, Matthew J. Carlson, Angela K. Wilson, *J. Comput. Chem.* **2012**, 33(30), 2380–2387, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23073>.
- Jeffrey P. Merrick, Damian Moran, Leo Radom, *J. Phys. Chem. A* **2007**, 111(45), 11683–11700, URL <https://doi.org/10.1021/jp073974n>. PMID: 17948971.
- John A. Pople, Anthony P. Scott, Ming Wah Wong, Leo Radom, *Isr. J. Chem.* **1993**, 33(3), 345–350, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/ijch.199300041>.
- Karl K. Irikura, Russell D. Johnson, Raghu N. Kacker, *J. Phys. Chem. A* **2005**, 109(37), 8430–8437, URL <https://doi.org/10.1021/jp052793n>. PMID: 16834237.

- [42] Peter Pulay, Geza Fogarasi, Gabor Pongor, James E. Boggs, Anna Vargha, *J. Am. Chem. Soc.* **1983**, 105(24), 7037–7047, URL <https://doi.org/10.1021/ja00362a005>.
- [43] Database of Frequency Scale Factors for Electronic Model Chemistries <https://comp.chem.umn.edu/freqscale/>.
- [44] NIST Computational Chemistry Comparison, Benchmark Database, NIST Standard Reference Database Number 101 Release 22, May 2022, Editor: Russell D. Johnson III URL <http://cccbdb.nist.gov/>.
- [45] Leonid S. Khaikin, Igor V. Kochikov, Olga E. Grikina, Denis S. Tikhonov, Esfir G. Baskir, *Struct. Chem.* **2015**, 26(5), 1651–1687, ISSN 1572–9001. URL <https://doi.org/10.1007/s11224-015-0613-9>.
- [46] Leonid S. Khaikin, Olga E. Grikina, Natalja Vogt, Nikolaj F. Stepanov, *Russ. J. Phys. Chem. A* **2012**, 86(12), 1855–1861, ISSN 1531–863X. URL <https://doi.org/10.1134/S0036024412120102>.
- [47] Sergey V. Krasnoshchekov, Natalja Vogt, Nikolay F. Stepanov, *J. Phys. Chem. A* **2015**, 119(25), 6723–6737, URL <https://doi.org/10.1021/acs.jpca.5b03241>. PMID: 26020099.
- [48] Axel D. Becke, *Phys. Rev. A* **1988**, 38, 3098–3100, URL <https://link.aps.org/doi/10.1103/PhysRevA.38.3098>.
- [49] Chengteh Lee, Weitao Yang, Robert G. Parr, *Phys. Rev. B* **1988**, 37, 785–789, URL <https://link.aps.org/doi/10.1103/PhysRevB.37.785>.
- [50] John P. Perdew, Kieron Burke, Matthias Ernzerhof, *Phys. Rev. Lett.* **1996**, 77, 3865–3868, URL <https://link.aps.org/doi/10.1103/PhysRevLett.77.3865>.
- [51] Axel D. Becke, *J. Chem. Phys.* **1993**, 98(7), 5648–5652, 04, ISSN 0021–9606. URL <https://doi.org/10.1063/1.464913>.
- [52] Carlo Adamo, Vincenzo Barone, *J. Chem. Phys.* **1999**, 110(13), 6158–6170, 04 ISSN 0021–9606. URL <https://doi.org/10.1063/1.478522>.
- [53] Stefan Grimme, Stephan Ehrlich, Lars Goerigk, *J. Comput. Chem.* **2011**, 32(7), 1456–1465, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21759>.
- [54] Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Hagen Neugebauer, Sebastian Spicher, Christoph Bannwarth, Stefan Grimme, *J. Chem. Phys.* **2019**, 150(15), 154122, 04 ISSN 0021–9606. URL <https://doi.org/10.1063/1.5090222>.
- [55] Robert Ditchfield, Warren J. Hehre, John A. Pople, *J. Chem. Phys.* **2003**, 54(2), 724–728, 09 ISSN 0021–9606. URL <https://doi.org/10.1063/1.1674902>.
- [56] Warren J. Hehre, Robert Ditchfield, John A. Pople, *J. Chem. Phys.* **1972**, 56(5), 2257–2261, 03. ISSN 0021–9606. URL <https://doi.org/10.1063/1.1677527>.
- [57] Florian Weigend, Reinhart Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, 7, 3297–3305, URL <https://doi.org/10.1039/B508541A>.
- [58] Rebecca Sure, Stefan Grimme, *J. Comput. Chem.* **2013**, 34(19), 1672–1685, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23317>.
- [59] Stefan Grimme, Jan Gerit Brandenburg, Christoph Bannwarth, Andreas Hansen, *J. Chem. Phys.* **2015**, 143(5), 054107, 08 ISSN 0021–9606. URL <https://doi.org/10.1063/1.4927476>.
- [60] Stefan Grimme, Andreas Hansen, Sebastian Ehlert, Jan-Michael Mewes, *J. Chem. Phys.* **2021**, 154(6), 064103, 02 ISSN 0021–9606. URL <https://doi.org/10.1063/5.0040021>.
- [61] Frank Neese, Frank Wennmohs, Ute Becker, Christoph Riplinger, *J. Chem. Phys.* **2020**, 152(22), 224108, URL <https://doi.org/10.1063/5.0004608>.
- [62] Frank Neese, *WIREs Comput. Mol. Sci.* **2022**, 12(5), e1606, URL <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1606>.
- [63] Bernhard Metz, Hermann Stoll, Michael Dolg, *J. Chem. Phys.* **2000**, 113(7), 2563–2569, 08. ISSN 0021–9606. URL <https://doi.org/10.1063/1.1305880>.
- [64] Kirk A. Peterson, Detlev Figgen, Erich Goll, Hermann Stoll, Michael Dolg, *J. Chem. Phys.* **2003**, 119(21), 11113–11123, 12. ISSN 0021–9606. URL <https://doi.org/10.1063/1.1622924>.
- [65] Christoph Bannwarth, Sebastian Ehlert, Stefan Grimme, *J. Chem. Theory Comput.* **2019**, 15(3), 1652–1671, URL <https://doi.org/10.1021/acs.jctc.8b01176>. PMID: 30741547.
- [66] Philipp Pracht, Fabian Bohle, Stefan Grimme *Phys. Chem. Chem. Phys.* **2020**, 22, 7169–7192, URL <https://doi.org/10.1039/C9CP06869D>.
- [67] Stefan Grimme, *J. Chem. Theory Comput.* **2019**, 15(5), 2847–2862, URL <https://doi.org/10.1021/acs.jctc.9b00143>. PMID: 30943025.
- [68] Krzysztof B. Beć, Christian W. Huck, *Front. Chem.* **2019**, 7, ISSN 2296–2646. URL <https://www.frontiersin.org/articles/10.3389/fchem.2019.00048>.
- [69] Ankit Rohatgi. Webplotdigitizer (4.5). *Computer software*. Retrieved from <https://automeris.io/WebPlotDigitizer>, 2021.
- [70] R. Prazeres, J. M. Berset, R. Chaput, F. Glotin, D. A. Jaroszynski, J. M. Ortega, *Nucl. Instrum. Methods Phys. Res. Sect. B* **1994**, 89(1), 54–59, ISSN 0168–583X. URL <https://www.sciencedirect.com/science/article/pii/0168583X94951454>.
- [71] D. Oepts, A. F. G. van der Meer, P. W. van Amersfoort, *Infrared Phys. Technol.* **1995**, 36(1), 297–308, ISSN 1350–4495. URL <https://www.sciencedirect.com/science/article/pii/135044959400074U>. Proceedings of the Sixth International Conference on Infrared Physics.
- [72] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, USA, 3 edition, 2007. ISBN 0521880688.
- [73] Denis S. Tikhonov, Diksha Garg, Melanie Schnell, *Photochem* **2024**, 4(1), 57–110, ISSN 2673–7256. doi: 10.3390/photochem4010005. URL <https://www.mdpi.com/2673-7256/4/1/5>.

Manuscript received: May 10, 2024

Revised manuscript received: August 16, 2024

Accepted manuscript online: August 22, 2024

Version of record online: October 28, 2024