

An Intuitively Understandable Quality Measure for Theoretical Vibrational Spectra

Henning Henschel* and David van der Spoel*

Cite This: *J. Phys. Chem. Lett.* 2020, 11, 5471–5475

Read Online

ACCESS |



Metrics & More

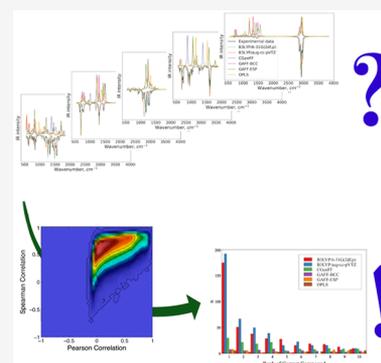


Article Recommendations



Supporting Information

ABSTRACT: Computational chemistry has become a central tool in spectroscopic studies in most of chemical science. The quality of a calculated vibrational spectrum is commonly expressed as the deviation of the peak position from the experimental reference. With the increasing application of vibrational spectroscopy to complex (biological) systems, this is likely not sustainable. Here we present a quality measure for theoretical vibrational spectra based on matching the spectra to a reference database with the help of correlation coefficients. This approach can easily be applied to large sets of data and complex spectra without easily identifiable peak positions. We demonstrate this on a database of infrared spectra of 670 compounds using six different theoretical (DFT and force field) methods. Most importantly, it is intuitively understandable by both theoreticians and experimentalists.



Molecular spectroscopy combined with computational chemistry is an essential tool to gain insight into chemical structures and mechanisms.¹ Especially vibrational spectroscopy is useful for the identification of a specific structure, as it provides not only straightforward information on the functional groups present but also, although in a lot less straightforward way, the exact arrangement of the atoms.² This has been demonstrated, for instance, through the application of the combination of vibrational spectroscopy with theoretical calculations to solve some of the fundamental questions of chirality.^{3,4} Density functional theory (DFT) has been particularly successful in this type of calculation,⁵ although the computational cost of DFT calculations limits their applicability to relatively small molecules. However, vibrational spectra can be calculated using a multitude of methods, ranging from high-level post-Hartree–Fock methods⁵ to force field,⁶ semiclassical,⁷ and empirical methods.² In any case, it is imperative to understand the limitations of the method used to calculate a given spectrum. Traditionally, the quality of any method to calculate vibrational spectra has mainly been assessed on the basis of the (root-mean-square) deviation of the peak positions from the corresponding experimental reference. However, making sure that the compared peaks correspond to the same vibration requires inspecting the vibrations, which is not simple to automate. Barone et al., for example, suggested two benchmark sets containing a total of 167 fundamental vibrations of 20 small (maximum of six non-hydrogen atoms) molecules in the development of their easy-to-use second-order vibrational perturbation theory (VPT2).⁸ Given the present development of applications of vibrational spectroscopy and the emergence of larger and larger databases with spectral information, it becomes obvious that the manual inspection approach is not sustainable.

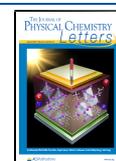
On one hand, the molecules to which vibrational analysis is applied are becoming larger or part of complex matrices.⁹ Some of these emerging areas of application are the structures of proteins¹⁰ and analysis of food items.^{11,12} Furthermore, vibrational spectroscopy is also under investigation as a tool in a growing number of medical applications, especially for cancer diagnosis and monitoring.^{13,14} On the other hand, with the introduction of data-driven methods such as machine learning into chemistry,¹⁵ there is a need to be able to automatically process large numbers of spectra.

We recently described the use of correlation coefficients to evaluate theoretical infrared (IR) spectra against experiment and showed that they provide relevant information about the quality of the spectrum.⁶ However, these methods merely provide numbers that are difficult to understand without the wide frame of reference available only to an expert in the field. Especially for experimentalists who want to work with theoretically derived spectra, a more intuitively understandable quality measure would be necessary, i.e., one that more immediately provides insight instead of mere numbers. Instead of comparing a theoretical spectrum only to the corresponding experimental spectrum, we therefore calculate the correlation coefficients between the theoretical spectrum and an entire database of experimental spectra. The experimental spectra were obtained

Received: May 28, 2020

Accepted: June 19, 2020

Published: June 19, 2020



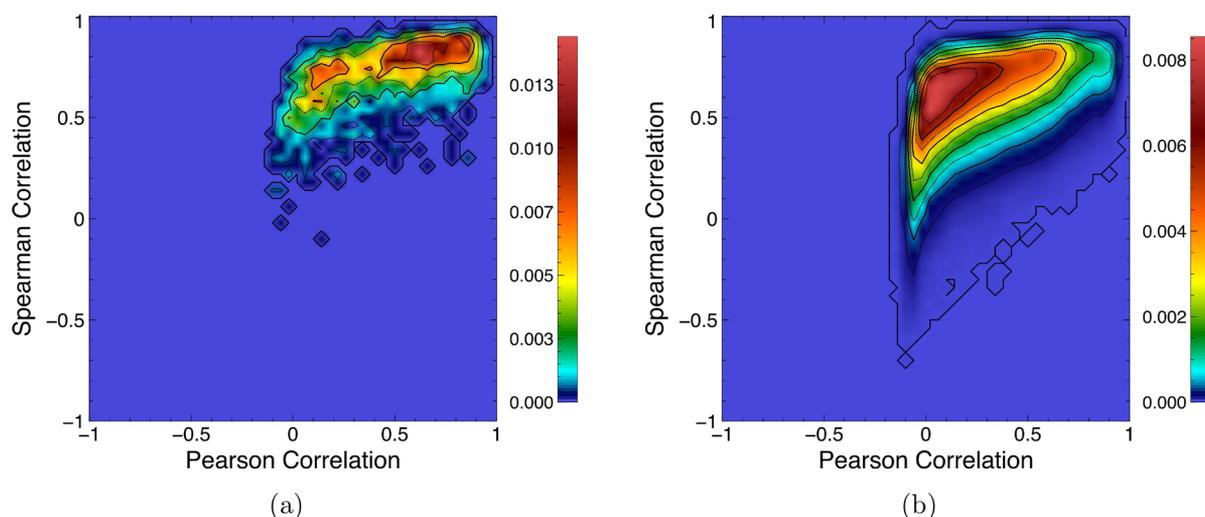


Figure 1. Distributions and relations of the Pearson and Spearman correlation coefficients from (a) the direct comparison of corresponding spectra from all theoretical methods to experiment and (b) the full cross-comparison of all theoretical spectra to all experimental spectra.

Table 1. Numbers of Compounds Correctly Identified from the Database of 670 Compounds Using Various Theoretical Methods in Combination with Different Statistical Measures^a

level of theory	theory vs experimental database				experiment vs theoretical database			
	Pearson r		Spearman ρ		Pearson r		Spearman ρ	
	full	short	full	short	full	short	full	short
B3LYP/6-31G(2df,p)	145	114	62	227	146	125	175	278
B3LYP/aug-cc-pVTZ	143	153	65	292	115	148	192	316
CGenFF	22	1	15	15	14	4	30	17
GAFF-BCC	1	3	3	2	2	3	8	6
GAFF-ESP	1	1	0	4	1	0	8	4
OPLS	2	1	6	4	3	1	6	2

^aThe labels “full” and “short” indicate wavenumber ranges of 550–3846 and 550–1650 cm^{-1} , respectively.

from the database at the National Institute of Standards and Technology (NIST)¹⁶ for all of the compounds in the Alexandria library¹⁷ that were available in the experimental database and for which valid force field topologies could be generated. These compounds are mostly small organic molecules with up to 10 non-hydrogen atoms, covering all common functional groups (information on the numbers of compounds belonging to various compound classes is included in Tables S2–S49 in section S4 in the [Supporting Information](#), and the complete list of compounds is given in [section S5](#)). From the calculated correlation coefficients we can easily derive the number (or fraction) of correct matches. We do this using both the Pearson and Spearman correlation coefficients that we have discussed previously in order to better understand their performance in different applications.

The similarities and differences between the two correlation coefficients for comparing theoretical spectra to the corresponding experimental spectra have already been discussed in our previous study.⁶ The distribution of the values of the two measures when the calculated spectra for all force fields and both quantum chemical methods are compared to the corresponding experimental spectra is shown in [Figure 1a](#). This can be compared to the distribution for the cross-comparison of all theoretical spectra with all experimental spectra given in [Figure 1b](#). In the [Figures S1–S4](#), illustrations of both direct and cross-comparisons for all wavenumber ranges used in this study, as discussed below, are given. Comparing the two distributions in

[Figure 1](#) indicates—not surprisingly—that correlation coefficients between matched spectra tend to be higher than those for comparison of a calculated spectrum with random experimental spectra. This holds true for all of the methods that we have used. However, here we are interested in the degree to which the different methods are able to produce spectra that give an optimal match with the experimental spectrum of the correct compound. For this we calculate the correlation coefficients of a given spectrum with all of the spectra of the reference database and rank the obtained correlation coefficients. A correct match is counted if the database spectrum giving the highest ranking coefficient originates from the correct compound. Obviously, producing a correct match within a large database is a much more difficult task than producing a good match. At the same time, we believe that this is a much more informative and relevant measure for the quality of a theoretically derived spectrum, especially if the method in question is to be used to help in structure elucidation efforts.

The performance of the various methods in matching the correct compound is summarized in [Table 1](#), which lists the numbers of correctly identified compounds using both correlation coefficients and two different directions of comparison: in one alternative we match each theoretical spectrum against the experimental database, and in the other we match each experimental spectrum against all of the calculated spectra. However, although the direction of comparison in some cases strongly affects the result, this effect is not entirely

systematic. In addition to employing the full spectrum, we used a shorter wavenumber range of 550–1650 cm^{-1} , corresponding to the fingerprint region, which at least in the case of small organic compounds is more informative of the individual compound. This is akin to what is done in the matching of experimental spectra to databases to improve performance, as already noted by Tanabe and Saeki,¹⁸ and used by several other authors^{2,19,20} with slightly varying wavenumber ranges. Additional data, including mean and median ranks of the spectrum of the correct compound within each comparison and histograms illustrating the distribution of these ranks as well as data for a wider wavenumber range of 550–2000 cm^{-1} and the inverse selection spanning the range 2000–3846 cm^{-1} , can be found in section S2. It is interesting to see that the shorter wavenumber range gives much better results with the quantum-chemical methods (except for B3LYP/6-31G(2df,p) when using Pearson's correlation), whereas for CGenFF, the only force field with a performance good enough to display any clear trend, the effect is rather the opposite. We ascribe this to the fact that at wavenumbers above 2000 cm^{-1} usually only a few features are found, which are mostly characteristic of specific bond types and therefore are very similar among especially organic compounds. This large degree of similarity leads to a lower discriminatory power, as already small deviations from an exact match with one spectrum can lead to a better match with a different spectrum (cf. Figure S4—cross-comparison using only the high-wavenumber range yields a large proportion of very high correlation coefficients that is reminiscent of the direct comparison). None of the other force fields gives significant numbers of matches in any systematic way, and the differences between methods of comparison and correlation coefficients are too small to clearly identify trends. It seems that the matches produced by CGenFF are due to very good reproduction of the valence vibrations, especially in alcohols, alkenes, and aromatic compounds, as shown in Tables S42 and S43.

Another way to illustrate the degree to which the obtained number of matches is systematic rather than by chance is displayed in Figure 2, which shows the distribution of ranks for the comparison that gives the highest single number of correct matches (experimental spectra vs theoretical database in the

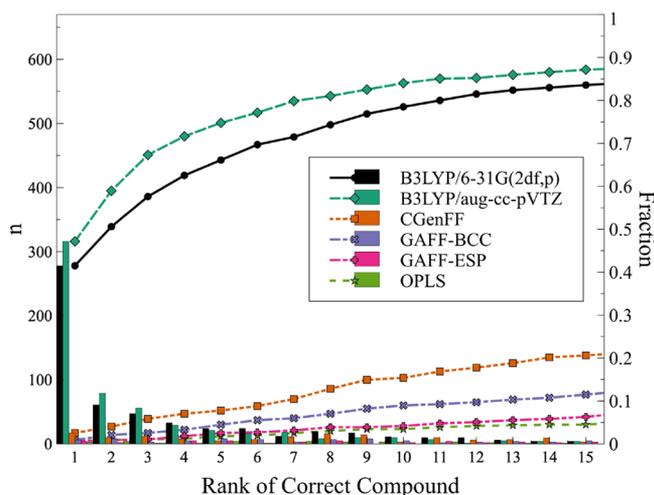


Figure 2. Individual and cumulative numbers and fractions of correct matches found in the first 15 ranks when experimental spectra are compared to the database of computed spectra for the short-wavenumber range (550–1650 cm^{-1}).

short-wavenumber range using the Spearman correlation coefficient). The rank is defined as the position of the correlation coefficient of the matched compound in the comparison over the entire database, with a rank of 1 being synonymous with a correct compound match. For the DFT methods, the number of spectra decreases systematically with rank at least until rank 5. For the force field methods, however, no clear trend can be found, suggesting that the number of matches obtained with these methods is partly by chance and accompanied by significant uncertainty. At the same time, the distribution in Figure 2 highlights the quality of the DFT methods by demonstrating that in addition to the first-rank matches, there is a significant number of close matches. For B3LYP/aug-cc-pVTZ, less than half of the spectra give a direct match, but already two-thirds of the spectra are within the first three ranks and 80% within the first 10 ranks. Rank distributions for all of the other comparisons are shown in Figures S5–S8.

Additionally, Tables S2–S49 also contain a compilation of data concerning the capabilities of the different methods to correctly identify the class of a given compound. The number of correct class matches, however, is not as intuitively understandable as the number of correctly matched compounds because the number of random matches depends in more than one way on the size of the class. Therefore, we regard it as less useful and do not discuss it further at this point. The interested reader is instead referred to section S4, where we also suggest a derived measure to make the implications of the number of class matches more intuitively understandable.

If the fraction of correct matches is to be used as a quality measure for methods of calculating spectra, it is necessary to also understand how this fraction depends on the size of the database. In Figure 3 we show the dependence of the fraction of

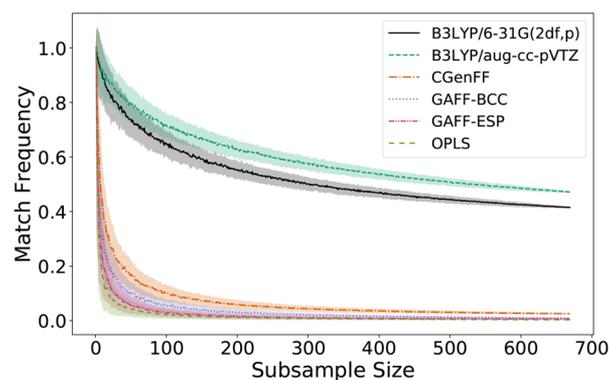


Figure 3. Dependence of the fraction of correct compound matches on the database size when experimental spectra are compared to the database of computed spectra for the short-wavenumber range (550–1650 cm^{-1}).

correct matches on the database size for the comparison of experimental spectra with the database of calculated spectra in the short-wavenumber range (those for other comparisons are given in section S3, Figures S9–S12). For each database size, 150 unique subsamples of the full database were generated; the standard deviations of the matching rates within these subsamples are shown as shaded areas in Figure 3. Obviously, for small subsamples there is a significant fraction of correct matches by chance. At the same time, the standard deviation for small subsamples is increased because of a larger variation in the homogeneity: small subsamples can be either very homogeneous, making differentiation between spectra more difficult, or

very heterogeneous, making differentiation easier. The data shown here and in section S3 demonstrate that except for a very few cases the ranking of the performance of the different methods is independent of the subsample size. However, when the standard deviation is taken into account, in the example shown in Figure 3 a subsample size of approximately 150 is necessary to distinguish between those methods that are distinguishable at any point. For larger subsamples, on the other hand, the differences between the poorest-performing (force field) methods become negligible. In these cases, the whole curve needs to be taken into account to differentiate between the performances of these methods; the depicted case indicates that GAFF-BCC performs better than GAFF-ESP, which in turn gives better results than OPLS, while CGenFF is somewhat better than the other force fields for any subsample size (Figure 3).

This work showcases how this comparatively simple methodology of using correlation coefficients to match calculated spectra against an experimental database provides an intuitively understandable measure of the quality of the spectra produced by any given method. On the basis of our results, we would primarily recommend that matching be based on the Spearman correlation coefficient for the fingerprint region. We also note that when entire databases containing spectra of the same compounds are matched against each other, as done here, somewhat different results are obtained depending on which of the databases is chosen as the reference. In any case, this approach provides a novel generalizable framework to map the performance of theoretical methods and improve communication between theoreticians and experimentalists. As shown, it can be used in an automated fashion over large sets of spectra without the need to assign and match specific vibrations or do any other type of processing by hand. This makes it also applicable to large, complex systems like proteins, where uniquely assigning peaks is impossible, not because of the number of spectra but rather because of their individual complexity. At the same time, we have demonstrated that the theoretical methods able to treat such large, complex systems, namely, the currently available general force fields, perform far too poorly to be practically useful in assisting with structure assignment, showcasing the enormous potential that still exists for improving force fields.

While the quantum-chemical methods used here perform better than the force fields by far, it is clear that spectra derived with these methods will rarely be sufficient as a stand-alone tool for the identification of an otherwise completely unknown substance. We are aware that there exist already a multitude of quantum-chemical methods that would yield better spectra than the ones used to derive the data in the Alexandria library. Other strategies to improve the performance of the various methods would for all types of methods include taking into account quantum dynamics of the nuclei,⁷ conformational sampling, and the anharmonicity of vibrational potentials and would for quantum-chemical methods also include the application of more elaborate scaling schemes. Optimization of quantum-chemistry-based strategies for calculating spectra, however, is a question best tackled by others.

METHODS

Information on the calculation of the spectra used for this study can be found in our previous article,⁶ in the Alexandria library,¹⁷ and at <http://virtualchemistry.org>.²¹ Experimental spectra were converted into intensity vectors corresponding to a frequency

grid with the experimental wavenumbers. For the calculation of the correlation coefficients, the frequency grids of the compared spectra need to match, i.e., in this study all of the spectra had to use the same grid. Therefore, we used only experimental spectra with a resolution of 4 cm⁻¹, which is the most commonly occurring one in the NIST database (i.e., we did not interpolate any spectrum), slightly decreasing the size of the reference database from 706 to 670 spectra. Experimental spectra with a wavenumber range wider than 550–3846 cm⁻¹ were truncated to this range. The theoretical spectra were calculated as intensity vectors with the same frequency grid.

Correlation coefficients between all theoretical and all experimental spectra were calculated as described in ref 6. The Pearson product moment correlation coefficient is defined as

$$r = \frac{\sum_i (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \cdot \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (1)$$

where x_i and y_i are the elements of the intensity vectors representing the spectra under comparison and \bar{x} and \bar{y} are the means of the x_i and y_i values, respectively. This correlation coefficient provides a measure of the linear correlation between two vectors. The Spearman rank correlation coefficient is given by

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

where d_i is the difference between the ranks of x_i and y_i in their respective datasets and n is the number of elements in each vector. As the ranks of the vector elements are used instead of their absolute values, the Spearman correlation coefficient does not require linearity but gives a measure of the monotonicity of the relationship between the vectors. The correlation coefficients were then used to rank the spectra in the respective comparison. The number of correct matches was determined by counting the theoretical spectra that gave the highest correlation coefficient for the experimental spectrum of the correct compound (or alternatively, the number of experimental spectra that gave the highest correlation coefficient for the theoretical spectrum of the same compound). Additional details on the comparisons conducted can be found in section S2.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcllett.0c01655>.

Figures and data pertaining to additional wavenumber ranges and comparison strategies; additional details of comparison methods; discussion of a potential additional measure for classwise spectra comparison and classwise data; list of compounds (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; Email: henning.henschel@ilk.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;
Email: david.vanderspoel@icm.uu.se

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jpcllett.0c01655>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The Swedish Research Council is acknowledged for a grant of computer time (SNIC2018-2-42). Funding from eSENCE—The e-Science Collaboration (Uppsala–Lund–Umeå, Sweden) is gratefully acknowledged.

REFERENCES

- (1) Neese, F. High-Level Spectroscopy, Quantum Chemistry, and Catalysis: Not just a Passing Fad. *Angew. Chem., Int. Ed.* **2017**, *56*, 11003–11010.
- (2) Baumann, K.; Clerc, J. Computer-assisted IR spectra prediction – linked similarity searches for structures and spectra. *Anal. Chim. Acta* **1997**, *348*, 327–343.
- (3) Haesler, J.; Schindelholz, I.; Riguete, E.; Bochet, C. G.; Hug, W. Absolute configuration of chirally deuterated neopentane. *Nature* **2007**, *446*, 526–529.
- (4) Costante, J.; Hecht, L.; Polavarapu, P. L.; Collet, A.; Barron, L. D. Absolute Configuration of Bromochlorofluoromethane from Experimental and Ab Initio Theoretical Vibrational Raman Optical Activity. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 885–887.
- (5) Beć, K. B.; Huck, C. W. Breakthrough Potential in Near-Infrared Spectroscopy: Spectra Simulation. A Review of Recent Developments. *Front. Chem.* **2019**, *7*, 48.
- (6) Henschel, H.; Andersson, A. T.; Jespers, W.; Mehdi Ghahremanpour, M.; van der Spoel, D. Theoretical Infrared Spectra: Quantitative Similarity Measures and Force Fields. *J. Chem. Theory Comput.* **2020**, *16*, 3307–3315.
- (7) Gabas, F.; Conte, R.; Ceotto, M. Semiclassical Vibrational Spectroscopy of Biological Molecules using Force Fields. *J. Chem. Theory Comput.* **2020**, *16*, 3476–3485.
- (8) Barone, V.; Biczysko, M.; Bloino, J. Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759–1787.
- (9) Bergmann, T. G.; Welzel, M. O.; Jacob, C. R. Towards theoretical spectroscopy with error bars: systematic quantification of the structural sensitivity of calculated spectra. *Chem. Sci.* **2020**, *11*, 1862–1877.
- (10) López-Lorente, A. I.; Mizaikoff, B. Mid-infrared spectroscopy for protein analysis: potential and challenges. *Anal. Bioanal. Chem.* **2016**, *408*, 2875–2889.
- (11) 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.
- (12) 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.
- (13) Bel'skaya, L. V. Use of IR Spectroscopy in Cancer Diagnosis. A Review. *J. Appl. Spectrosc.* **2019**, *86*, 187–205.
- (14) Su, K.-Y.; Lee, W.-L. Fourier Transform Infrared Spectroscopy as a Cancer Screening and Diagnostic Tool: A Review and Prospects. *Cancers* **2020**, *12*, 115.
- (15) Cova, T. F. G. G.; Pais, A. A. C. C. Deep Learning for Deep Chemistry: Optimizing the Prediction of Chemical Patterns. *Front. Chem.* **2019**, *7*, 809.
- (16) Wallace, W. E. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P., Mallard, W., Eds.;

National Institute of Standards and Technology: Gaithersburg, MD, 2018; Chapter Infrared Spectra.

(17) 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.

(18) Tanabe, K.; Saeki, S. Computer retrieval of infrared spectra by a correlation coefficient method. *Anal. Chem.* **1975**, *47*, 118–122.

(19) Mott, A. J.; Thirumuruganandham, S. P.; Thorpe, M. F.; Rez, P. Fast calculation of the infrared spectra of large biomolecules. *Eur. Biophys. J.* **2013**, *42*, 795–801.

(20) 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.

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