Correlating the Vibrational Spectra of Structurally Related Molecules: A Spectroscopic Measure of Similarity

Yunwen Tao,[a] Wenli Zou,[b] Dieter Cremer,[a]* and Elfi Kraka [a]

Using catastrophe theory and the concept of a mutation path, an algorithm is developed that leads to the direct correlation of the normal vibrational modes of two structurally related molecules. The mutation path is defined by weighted incremental changes in mass and geometry of the molecules in question, which are successively applied to mutate a molecule into a structurally related molecule and thus continuously converting their normal vibrational spectra from one into the other. Correlation diagrams are generated that accurately relate the normal vibrational modes to each other by utilizing mode-mode overlap criteria and resolving allowed and avoided crossings of vibrational eigenstates. The limitations of normal mode correlation, however, foster the correlation of local vibrational modes, which offer a novel vibrational measure of similarity. It will be shown how this will open new avenues for chemical studies. © 2017 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.25109

Introduction

The similarity between two structurally related molecules can be accessed via their vibrational spectra. Any difference in the electronic structure is sensitively registered by the vibrational modes. The vibrational spectrum of a molecule depends on five quantities: (i) Number of atoms N which determines the number of vibrations $N_{vib}=3N-L$ (L: number of translations and rotations); (ii) Point group symmetry X of the molecule; (iii) Masses m_i of the atoms (collected in the mass matrix **M**); (iv) the geometry **R** of the molecule; and (v) the electronic structure in a molecule characterized by its Hessian matrix **K**. The change of any of these five quantities results in a change in the vibrational spectrum, where this change can encompass a change of N_{vib}, the position of the vibrational frequencies ω_μ, the corresponding intensities I_μ, and the form of the normal vibrational modes $\mathbf{l}_\mu$. Comparison of the vibrational spectra even of structurally closely related molecules is often hampered because the mode-mode coupling leads to a serious complication that troubles the correlation of vibrational spectra. However, if one would be able to eliminate the mode-mode coupling, a number of interesting problems in the correlation of vibrational spectra can be solved.

If one can analyze the changes in the vibrational spectrum, where a change in N, X, **R**, **M**, or **K** is expected simultaneously, valuable information from vibrational spectra can be extracted.[1–5] This major objective can be split into several smaller problems of correlating two vibrational spectra by changing part of the above five quantities which influence the normal vibrational modes and fixing the rest:

1. Changes in X and **M**, while keeping N, **R** and **K** unchanged: correlation of the vibrational spectra of isotopomers, for example, $H_2O \rightarrow HOD$;
2. Changes in **K** and **R**, while keeping N, X and **M** fixed: correlation of the vibrational spectra along a reaction path, for example, HCN → HNC;
3. Changes in (X,) **K** and **R**, while keeping N, **M** fixed: correlation of the vibrations of conformational isomers, for example, gauche conformation → anti conformation of butane molecule;
4. Changes in (N,) (X,) **R**, **M** and **K**: correlation of the vibrational spectra of structurally related compounds, for example, methane → ethane.

In our previous work,[6] problem 1 has been solved by defining a mass reaction coordinate that connects a pair of isotopomers in terms of vibrational frequencies. We have also performed extensive work related to problem 2 which involves the studies of reaction mechanism with the reaction path Hamiltonian[7,8] combined with our unified reaction valley approach.[9–11] According to the McIver-Stanton rules,[12] the reaction complex can only change its symmetry at a stationary point (local minimum or first-order saddle point), otherwise the symmetry is kept along the reaction path. In this connection, $3N-L-1$ generalized normal modes of the reactant complex and the product complex are correctly correlated with regard to the reaction coordinate s. Problem 3 is closely linked to problem 2 because the conformational change of a molecule from one local minimum to another local minimum on the ground state potential energy surface (PES) requires an

[a] Y. Tao, D. Cremer, E. Kraka
Department of Chemistry, Southern Methodist University, 3215 Daniel Avenue, Dallas, Texas 75275-0314
E-mail: ekraka@smu.edu
[b] W. Zou
Institute of Modern Physics, Northwest University, Xi'an, Shaanxi 710127, People's Republic of China
*In Memoriam
Contract grant sponsor: National Science Foundation; Contract grant number: 1464906

energy barrier. In this regard, a conformational change can be classified as a reaction without bond breaking/formation.

From the number of quantities that can vary, problem 4 is considered to be the most complicated and general situation. Theoretically speaking, problem 4 covers problems 1–3, while problems 2–3 can be solved in a specialized approach. However, a generalized algorithm for the automated correlation of vibrational spectra of any pair of structurally related compounds needs to solve all of the four problems.

To do so, we define a mutation path, which allows the correlation of the normal vibrational spectra of two molecules, even when they differ in the number of atoms *N*. In the second section, we will describe the theory of the mutation path and present test examples. In the third section, we adopt the idea of the one-to-one correlation of normal modes via a mutation path and extend this procedure to local vibrational modes to decipher further electronic structure details and characterize the similarity of the two molecules. Computational details will be given in the fourth section while the conclusions are made in the last section.

Correlation of Normal Vibrational Modes Based on a Mutation Path

Definition and use of mutation path

If an atom of a molecule is replaced by another atom or a polyatomic substituent, a change in the point group symmetry of the molecule can take place. Changes in N_{vib}, the atomic masses, the geometry and the potential energy have also to be considered so that significant changes of the vibrational spectrum are the rule rather than the exception. For the correlation of the corresponding spectra, these changes are applied to the starting spectrum stepwise. Given the information of the parent molecule **A** and the target molecule **B**, we can define a new type of reaction as "mutation" which describes the transition process from **A** to **B**, even if **A** and **B** differ in *N* ($N_{\mathbf{A}} \leq N_{\mathbf{B}}$ is required), **R**, *X*, **M**, and **K**. However, **A** and **B** should be structurally related. To design an appropriate mutation reaction path, the following aspects have to be considered:

1. How is the reaction coordinate of a mutation path defined?
2. How is the geometry **R** changed from the reactant molecule **A** to the product molecule **B** with regard to the reaction coordinate?
3. How does the symmetry *X* change along with the changing geometry **R**?
4. How does the mass **M** change, especially when an atom is replaced by a polyatomic substituent?
5. How is the electronic structure changed without doing *ab initio* calculation in each step along the mutation path? Can we get a realistic description of the electronic structure between two different molecules? If not, how to simulate it in a reasonable approach?
6. How to resolve the problem of allowed crossing and avoided crossing between modes when they have degenerate frequency values in the mutation path?

Unlike the chemical reaction path that can be theoretically characterized by the intrinsic reaction coordinate,[13] where we use the mass-weighted collective atomic displacements as the reaction coordinate *s*, here we use a scalar λ ranging from 0 to 1 connecting the reactant **A** and the product molecule **B** at two ends. With the increase of λ, each quantity including the mass, geometry, and electronic structure is required to simultaneously change in a linear form.

The change of the geometry **R** is the most complicated part of the mutation path. As all quantities change along the mutation path, we need to distinguish two situations: (i) $N_{\mathbf{A}}=N_{\mathbf{B}}$, the number of atoms does not change. (ii) $N_{\mathbf{A}} < N_{\mathbf{B}}$, the target molecule "grows" from the parent molecule.

In the first situation, for example, **A**= H_2O, **B**= H_2S, we do not expect major differences in geometry between these two molecules, as the connectivity stays the same. At any point λ along the reaction path, the geometry can be easily given as a linear combination from that of **A** and **B**:

$$\mathbf{R}_\lambda = \mathbf{R}_\mathbf{A} + \lambda(\mathbf{R}_\mathbf{B} - \mathbf{R}_\mathbf{A}) \quad (1)$$

In the second situation of $N_{\mathbf{A}} < N_{\mathbf{B}}$, we need to compare $\mathbf{R}_\mathbf{A}$ and $\mathbf{R}_\mathbf{B}$ and find the common substructure $\mathbf{a_0}$ and $\mathbf{b_0}$ within **A** and **B**. Then, atom(s) within **A** need to be found that is(are) not within $\mathbf{a_0}$, denoted as $\mathbf{a}'_i$. $\mathbf{a}'_i$ is considered as the root of geometry growth toward **B**, here *i* is the label of root. The number of roots can be more than one, depending on the difference between $\mathbf{R}_\mathbf{A}$ and $\mathbf{R}_\mathbf{B}$. Parts within **B** but not in $\mathbf{b_0}$ are taken as $\mathbf{b}'_i$, $\mathbf{b}'_i$ is the direct consequence from growth starting from root $\mathbf{a}'_i$. So at any point λ along the reaction path, the geometry is:

$$\mathbf{R}_\lambda = \mathbf{R}_{\mathbf{a_0}} + \lambda(\mathbf{R}_{\mathbf{b_0}} - \mathbf{R}_{\mathbf{a_0}}) + \sum_i (\mathbf{R}_{\mathbf{a}'_i} + \lambda(\mathbf{R}_{\mathbf{b}'_i} - \mathbf{R}_{\mathbf{a}'_i})) \quad (2)$$

with $\mathbf{A}=\{\mathbf{a_0}, \mathbf{a}'_1, \mathbf{a}'_2, \ldots\}$, $\mathbf{B}=\{\mathbf{b_0}, \mathbf{b}'_1, \mathbf{b}'_2, \ldots\}$. To show how this works, we take **A**= Methane, **B**= Ethane. See Figure 1, in this mutation process, we have methane as the reactant molecule **A**, ethane as the product molecule **B**. The common substructure $\mathbf{a_0}$ ($\mathbf{b_0}$) in both molecules is the methyl group on the left side (colored in black). The root of growth $\mathbf{a}'_1$ in the reactant molecule is the hydrogen atom on the right side of methane (colored in purple). The part $\mathbf{b}'_1$ in ethane grown from the root $\mathbf{a}'_1$ is the methyl group on the right side (colored in red). So in this mutation path, three parallel mutation events are taking place: (i) The lengthening of C—H bond of **A** into C—C bond of **B**, (ii) The growth of C—H bond in $\mathbf{b}'_1$ starting from 0 Å, and (iii) The relaxation and adjustment of the substructure $\mathbf{a_0}$ to $\mathbf{b_0}$.

One has to bear in mind that not any two arbitrary molecules can be correlated in such a way in geometry. For example, it would be rather difficult to let the methane molecule grow into a steroid molecule in one step. This reemphasizes a precondition of the mutation process in the second situation with different numbers of atoms that a common substructure should exist in both the reactant and the product.

Immediate change of geometry **R** from the reactant molecule will lead to the change of the symmetry *X* provided the

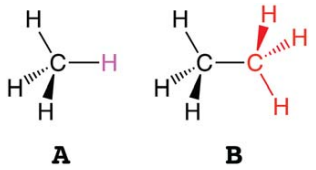

Figure 1. Structure of reactant and product molecule in the mutation path of methane → ethane. [Color figure can be viewed at wileyonlinelibrary.com]

reactant molecule has a symmetry higher than $C1$. In this case, the symmetry will decrease right away when the mutation path starts. In the mutation shown in Figure 1, the symmetry of the reactant will decrease immediately from T_d to C_{3v} starting at the first step along the reaction path. Along the whole mutation path, the C_{3v} symmetry will be kept, until the product molecule ethane is reached, which has D_{3d} symmetry.

With regard to the change of mass **M**, we need also to distinguish between two situations like for the change of geometry **R**: (i) $N_\mathbf{A}=N_\mathbf{B}$, (ii) $N_\mathbf{A} < N_\mathbf{B}$. In the first situation, there exists a one-to-one relationship of atoms between reactant molecule and product molecule, for example, H_2O and H_2S. The mass of any point along the reaction path λ is given by

$$\mathbf{M}_\lambda=\mathbf{M}_\mathbf{A}+\lambda(\mathbf{M}_\mathbf{B}-\mathbf{M}_\mathbf{A}) \tag{3}$$

In the second situation, the common substructure, $\mathbf{a_0}$ and $\mathbf{b_0}$ will keep its mass unchanged along the mutation path. We need to additionally differentiate the atom within $\mathbf{b}_i'$ which will replace $\mathbf{a}_i'$, we call it $\mathbf{b}_{i_m}'$, while the other part within $\mathbf{b}_i'$ is $\mathbf{b}_{i_n}'$ which comes to existence from massless vacuum. At the same time, $\mathbf{b}_i'=\{\mathbf{b}_{i_m}', \mathbf{b}_{i_n}'\}$. Then, the mass of λ can be expressed as:

$$\mathbf{M}_\lambda=\mathbf{M}_{\mathbf{a_0}}+\sum_i(\mathbf{M}_{\mathbf{a}_i'}+\lambda(\mathbf{M}_{\mathbf{b}_{i_m}'}-\mathbf{M}_{\mathbf{a}_i'})+\lambda\mathbf{M}_{\mathbf{b}_{i_n}'}) \tag{4}$$

To illustrate this with the methane to ethane example in Figure 1, the methyl group colored in black keeps its mass unchanged. The mass of hydrogen atom in purple gradually changes into the mass of a carbon atom, while the mass of hydrogen atoms in red color increases from zero to the atomic mass of hydrogen.

One major problem that might hinder us from following this mutation path is how can we describe the electronic structure of the mutation complex at each point on the path. This implies to answer the following questions: (i) How can we describe a system with unreasonable bond length which is close to zero? (ii) How can we describe a system having a fractional number of electrons? In molecular mechanics (MM), there exists a solution provided by Zwanzig in the free energy perturbation calculations[14] by adjusting the force field parameters stepwise in the course of the molecular dynamics simulations. Unfortunately, we have no such *ab initio* methods that can work in this situation. However, we can use a mathematical trick by taking a linear combination of the Hessian matrices of the reactant and product molecules to define an intermediate Hessian matrix for any point along the mutation path.

$$\mathbf{K}_\lambda=\mathbf{K}_\mathbf{A}+\lambda(\mathbf{K}_\mathbf{B}-\mathbf{K}_\mathbf{A}) \tag{5}$$

The dimension of $\mathbf{K}_\mathbf{A}$ differs from that of $\mathbf{K}_\mathbf{B}$ if $N_\mathbf{A} \neq N_\mathbf{B}$. To get rid of this inconsistency, we pad $\mathbf{K}_\mathbf{A}$ with zeros to the same dimension as $\mathbf{K}_\mathbf{B}$. Those atoms which are not present in the reactant molecule do not influence the electronic structure of the reactant molecule at the early stage of the mutation path as λ is small. Thus, in the beginning, the vibrational spectrum is dominated by the reactant molecule with little influence from the atoms which are not present in **A**. So in this way, the change of electronic structure is described in a reasonable way.

Another problem that needs to be addressed is that there might be a discrepancy between the reactant molecule and the mutation complex at the first step of the mutation path as the number of atoms and symmetry both change at that point. However, this problem can be solved via a catastrophe point, where the introduced perturbation from the mass and electronic effect is so small that a smooth transition from reactant molecule to the mutation path can be established. From Table 1 for the mutation path of methane → ethane, we see that even though 9 new vibrations come to exist due to three new atoms introduced in, vibrations from the methane as the reactant retain themselves quite well in the beginning of the mutation process.

The Wilson equation of vibrational spectroscopy[15] has to be solved at each point using the intermediate Hessian $\mathbf{K}_\lambda$ along the mutation path to guarantee a continuous description connecting two structurally related molecules.

At a given point, λ_a of the mutation path, which is defined by the mutation path coordinate λ, corresponding normal modes $\mathbf{l}_\mu(\lambda_a)$ and their associated normal mode vibrational frequencies $\omega_\mu(\lambda_a)$ are obtained via solving

$$\mathbf{K}(\lambda_a)\mathbf{l}_\mu(\lambda_a)=\omega_\mu^2(\lambda_a)\mathbf{l}_\mu(\lambda_a) \tag{6}$$

which gives $N_{\text{vib}}=3N_\mathbf{B}-L$ (L: number of translations and rotations) eigenvectors with the length of $3N_\mathbf{B}$ spanning the N_{vib}-dimensional space. For any two consecutive points, λ_a and λ_b separated by the mutation path coordinate increments

$$\delta\lambda=\frac{1-0}{f}=\frac{1}{f} \tag{7}$$

(f is a scaling factor ranging from 1000 to 10,000. The value of f determines the mutation path stepsize $\delta\lambda$.) Equation (6) is solved:

$$\mathbf{K}_a\mathbf{a}_\mu=(\omega_\mu^a)^2\mathbf{a}_\mu \tag{8}$$

$$\mathbf{K}_b\mathbf{b}_\mu=(\omega_\mu^b)^2\mathbf{b}_\mu \tag{9}$$

In these two equations, $\mathbf{a}_\mu$ and $\mathbf{b}_\mu$ are the mass-weighted normal mode vectors calculated at λ_a and λ_b, respectively.

In the case of degenerate frequency values at a specific point on the path, namely $\omega_\mu=\omega_\nu$, we need to obtain a correct ordering of normal modes at this point, especially when the whole mutation complex has a symmetry higher than $C1$.

In this situation, we use the diabatic mode ordering (DMO) procedure[16] which was proposed by Konkoli et al. With DMO, we can resolve allowed- and avoided-crossings which might appear in the frequency values along the mutation path. To correlate the vibrational modes at λ_a and λ_b, the overlap between normal mode vectors is maximized via rotation. We collect the mode vectors $\mathbf{a}_\mu$ and $\mathbf{b}_\mu$ into matrices **A** and **B**:

$$\mathbf{A}=(\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_\mu,\ldots,\mathbf{a}_{N_{vib}}) \tag{10}$$

$$\mathbf{B}=(\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_\mu,\ldots,\mathbf{b}_{N_{vib}}) \tag{11}$$

the overlap matrix between **A** and **B** is given by the scalar product

$$\mathbf{S}_{BA}=\mathbf{B}^\dagger\mathbf{A} \tag{12}$$

Matrix **T** used for rotation is defined

$$\mathbf{T}=(\mathbf{S}_{BA}^\dagger\mathbf{S}_{BA})^{-1/2}\mathbf{S}_{BA}^\dagger \tag{13}$$

which fulfills the following condition

$$Tr(\mathbf{B}^\dagger\mathbf{B}')=\max \tag{14}$$

with

$$\mathbf{B}\approx\mathbf{B}'=\mathbf{AT} \tag{15}$$

and

$$\mathbf{TT}'=\mathbf{I} \tag{16}$$

Matrix $\mathbf{B}'$ is an image of matrix **B** rotated by **T** in the space of V^A. There is no transformation **T** that can lead to $\mathbf{B}'=\mathbf{B}$ because $\mathbf{a}_\mu$ and $\mathbf{b}_\mu$ span different spaces of V^A and V^B, respectively.

To relate N_i normal mode vectors $\mathbf{b}_\mu$ in the space of V^B with subspace V_i^A of dimension N_i (i = 1: non-degenerate vibrational mode; 2, 3, . . .: degenerate vibrational modes), an amplitude is defined

$$A_\lambda^i=\frac{1}{N_i}\sum_{\mu i=1}^{N_i}T_{\mu i,\lambda}^2 \tag{17}$$

N_i normal mode vectors $\mathbf{b}_\mu$ having the largest amplitude A_λ^i are assigned to subspace V^A. This procedure of finding the image vectors of $\mathbf{b}_\mu$ as $\mathbf{b}'_\mu$ in the space of V^B, is the only solution to connect vectors in the space of V^A and V^B.

Then, vectors $\mathbf{a}_\mu$ in the subspace V_i^A are rotated by

$$\mathbf{a}_{\mu i}=\sum_{vi=1}^{N_i}\mathbf{a}_{vi}\mathbf{R}_{vi,vi}^i \tag{18}$$

rotation matrix $\mathbf{R}^i$ is associated with the subspace V_i^A

$$\mathbf{R}^i=[(\mathbf{S}_{BA}^i)^\dagger\mathbf{S}_{BA}^i]^{-1/2}(\mathbf{S}_{BA}^i)^\dagger \tag{19}$$

in which $\mathbf{S}_{BA}^i$ is the overlap matrix for the subspace V_i^A

Table 1. Vibrational frequencies of methane and the mutation complex at the catastrophe point close to methane (unit: cm^{-1}).

No.	Methane	Methane + $\delta\lambda$	No.	Methane	Methane + $\delta\lambda$
1	–	313.21	10	–	2107.41
2	–	1253.57	11	–	2107.41
3	–	1253.57	12	3047.26	3046.18
4	1356.01	1356.03	13	3163.20	3160.19
5	1356.01	1357.22	14	3163.20	3163.17
6	1356.01	1357.22	15	3163.20	3163.17
7	1578.51	1578.35	16	–	4197.70
8	1578.51	1578.35	17	–	4197.70
9	–	1867.76	18	–	4246.55

[a] $\delta\lambda$ is taken as 1/4000 of the whole mutation path, and this is the first point on the path next to methane as the reactant molecule.

$$(\mathbf{S}_{BA}^i)_{\mu i,vi}=\mathbf{b}_{\mu i}^\dagger\mathbf{a}_{vi}\quad \mu i,vi=1,\ldots,N_i \tag{20}$$

Thus, degenerate normal modes spanning the subspace V_i^A can be guaranteed to give a correct ordering.

To test the ordering results, one can calculate the final overlap matrix $\mathbf{S}_{BA}$ which is expected to be close to diagonal identity matrix.

$$(\mathbf{S}_{BA})_{\mu v}=\mathbf{b}'^\dagger_{\mu i}\mathbf{a}_{vi} \tag{21}$$

Then the smallest value in $\mathbf{S}_{BA}$ is checked. If it is smaller than the threshold $S_{\min}=0.95$, the assignment and ordering is considered to be weak. In this situation, the mutation path is repeated with smaller step size $\delta\lambda$.

In the following, we will discuss the validation of the proposed mutation path procedure in a series of examples with increasing complexity with regard to the five quantities that determine the vibrations.

Example 1: CHBrClF enantiomers

The first example is the mutation path for (*S*)-bromo-chloro-fluoromethane → (*R*)-bromo-chloro-fluoromethane. The reactant and the product have the same number of atoms, $N_\mathbf{A}=N_\mathbf{B}$. Both have *C*1 symmetry.

Being a pair of enantiomers, the reactant and the product molecules are identical in their electronic structure except that they are the mirror image to each other. So the same normal vibrational frequency values are expected for $\lambda=0$ and $\lambda=1$. Experimentalists who do vibrational spectroscopy might argue that it is not possible to differentiate between such enantiomers. However, we show that the vibrational spectra of a pair of enantiomers can be correlated with a mutation path.

In the specific geometry of the CHBrClF molecule, there can be up to $C(4,2)=6$ possible pathways of mutation by switching any two atoms besides the central carbon atom. In this work, we choose to mutate the *Cl* and *H* atoms while keeping the other atoms intact.

As one can see from correlation of the vibrational frequencies (Fig. 2), the whole mutation path is symmetric with regard to $\lambda=0.5$. Even though the entire path has the symmetry of *C*1, C_s symmetry is identified at the middle point. All

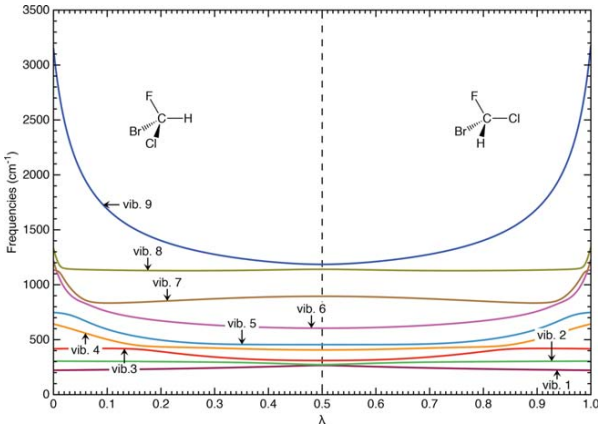

Figure 2. Correlation of normal mode frequencies of (*S*)-bromo-chloro-fluoromethane (left) with that of (*R*)-bromo-chloro-fluoromethane (right) along the mutation path. [Color figure can be viewed at wileyonlinelibrary.com]

vibrational modes along the path have the *a* symmetry. This leads to multiple avoided-crossings between vibrational modes at $\lambda = 0.01, 0.02, 0.06, 0.16, 0.5$ in the first half of this path. We need to note that at $\lambda = 0.5$, vibration 1 and 2 are not degenerate. Vibration 9 has the most striking change along the path. By checking the normal mode character of this vibration, we find that the nature of this mode is of C—H bond stretching character and it is decoupled from the rest of the molecule. The drastic change can be explained in the following way. At $\lambda = 0$, the C—H stretching in vibration 9 is decoupled from other local modes due to the tiny mass of the *H* atom compared with the halogen atoms. When the mutation proceeds from the reactant, the *H* atom slowly gains the properties of the *Cl* atom, including its atomic mass. As a result, vibration 9 will have other local modes mixed in causing the loss of the decoupling effect. At the same time, the *Cl* atom reduces its atomic mass. To the middle point, the original *H* and *Cl* atoms are the same in the way that they become a virtual atom as *X*, $X = (H+Cl)/2$. When the mutation continues, the *X* atom coming from *Cl* changes into the *H* atom, thus the decoupled C—H vibration is brought back while the other *X* atom turns into *Cl*.

Example 2: Benzene (C_6H_6) → fluorobenzene (C_6H_5F)

The second example is the mutation path of benzene(C_6H_6) → fluorobenzene(C_6H_5F), in which we still have $N_\mathbf{A} = N_\mathbf{B}$, but **A** has a higher symmetry than **B**. The symmetry of each normal mode along the mutation path can be resolved by the DMO method.

It is illustrated in Figure 3 that the benzene molecule with D_{6h} symmetry can be linked to the fluorobenzene molecule in its C_{2v} symmetry via a mutation path. As both the reactant and product have 12 atoms, there are in total 30 normal vibrational modes. Thus, a one-to-one relationship between these two sets of normal modes can be established. In the top diagram of Figure 3, the symbols of irreducible representation are labeled for the five vibrations with highest frequency values. These five vibrations retain themselves quite well along the mutation path, and no crossings are observed. For the vibration with b_{2u} symmetry in the reactant and a_1 symmetry along the path, it has the largest decrease in the frequency value. Examination of this normal mode shows that the mode character changes from the anti-symmetric C—H bond stretching mode within the benzene molecule into deformation mode of the C_6 ring within the fluorobenzene molecule. In the lower frequency region (200–1700 cm^{-1}) shown in the bottom diagram of Figure 3, many allowed crossings between normal modes in different symmetries can be observed. One interesting observation is that some normal mode can cross up to eight other modes along the mutation path. Such a sophisticated crossing behavior can never be expected if one inspects the vibrational frequency values measured by the infrared or Raman spectrometer. Furthermore, all two-fold degenerate vibrations (marked with a small circle) on the reactant side lose their degeneracy by splitting into two vibrations with different symmetries along the mutation path.

Example 3: Methane (CH_4) → ethane (C_2H_6)

The one-to-one relationship between the normal modes of two molecules will only be partially preserved if the molecule **B** has more atoms than the molecule **A**, namely $N_\mathbf{A} < N_\mathbf{B}$. In the mutation path of methane (CH_4) → ethane (C_2H_6), one of the hydrogen atoms within methane grows into a methyl group (-CH_3) via mutation (see also Fig. 1). As ethane has 8 atoms while methane has only 5 atoms, $3 \times 3 = 9$ additional vibrations are brought in since the beginning of the mutation path. In the frequency correlation diagram in Figure 4, the frequencies of these nine vibrational modes are plotted with dashed lines, which need to be distinguished from the other nine solid lines. Those solid lines show the change of frequency values of the vibrations originated from the methane molecule as the reactant.

To find out these nine normal modes that belong to the parent molecule, the scalar product of the normal mode vectors between the reactant molecule and the mutation complex at the catastrophe point along the mutation path is calculated to get the overlap measurement.

$$\mathbf{l}^{\dagger}_{\mathbf{A},\mu}\mathbf{l}^{1st}_{\lambda,\nu} = s_{\mu\nu} \tag{22}$$

in which $\mu = 1, 2, \ldots, 3N_\mathbf{A} - L$; $\nu = 1, 2, \ldots, 3N_\mathbf{B} - L$ (L: total number of translations and rotations). Furthermore, $\mathbf{l}_{\mathbf{A},\mu}$ is a vector of normal mode μ in the reactant **A**, while $\mathbf{l}^{1st}_{\lambda,\nu}$ is the vector of normal mode ν at the catastrophe point of the mutation path. Zeros are used to pad in $\mathbf{l}_{\mathbf{A},\mu}$ to make it have the same length for $\mathbf{l}^{1st}_{\lambda,\nu}$. If the overlap value $s_{\mu\nu}$ is greater than 0.95, then normal mode μ can be smoothly correlated with normal mode ν (see Table 1).

On the left side of Figure 4, in total nine vibrational modes from the parent molecules are labeled from 1 to 9 with symmetry and numbering. Although three-fold degenerate vibrations have the same frequency values, they are still numbered for differentiation. On the right side which stands for the product molecule, all 18 vibrations are labeled as for the reactant side according to the ordering of frequency value. However,

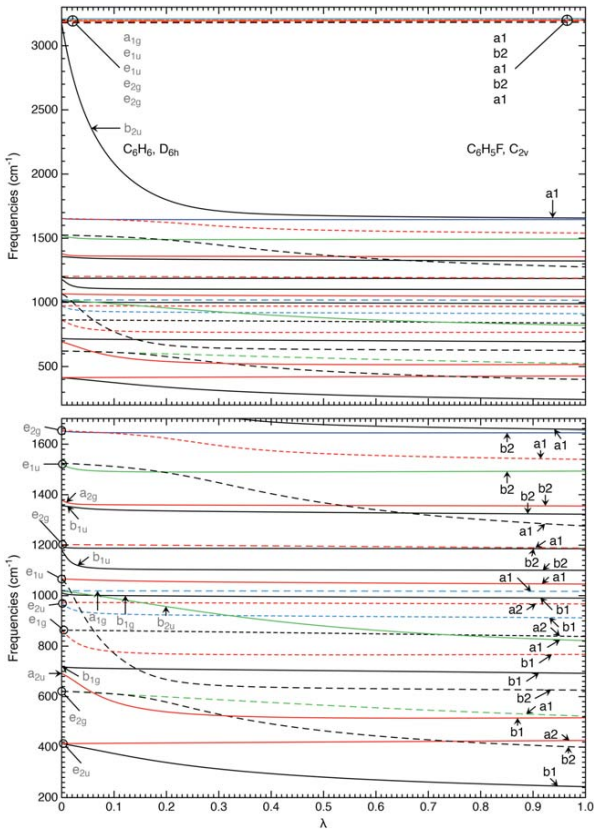

Figure 3. Correlation of normal mode frequencies of benzene (left) with that of fluorobenzene (right) along the mutation path. Irreducible representations of normal modes are labeled on both sides. Top: Diagram of all vibrational frequencies. Bottom: Diagram of the low frequency part (200–1700 cm^{-1}). [Color figure can be viewed at wileyonlinelibrary.com]

those vibrations without any counterpart in the reactant side are labeled in gray color.

Among all normal modes along the path, the $a_1(6)$ mode on the reactant side has the largest change in the vibration frequency value. This mode describes the symmetric stretching of 4 C—H bonds in methane. Then it turns into a symmetric combination of two pyramidalization modes of two methyl groups in ethane, denoted as $a_{1g}(8)$. Both allowed- and avoided-crossings are observed in this path. Furthermore, any three-fold degenerate vibration is split into a non-degenerate vibration and two-fold degenerate vibrations.

Besides the vibrational frequencies which can be obtained either from experimental measurement or *ab initio* calculations, it could be interesting if we can correlate directly the vibrational spectra of two molecules, that is, the infrared (IR) spectrum including both the vibrational frequency and the IR intensity with mutation path.

The IR intensity of normal mode μ at a given point λ_a on the mutation path is given by[17–19]

$$I_\mu(\lambda_a)=\boldsymbol{\delta}_\mu^\dagger(\lambda_a)\boldsymbol{\delta}_\mu(\lambda_a) \tag{23}$$

where $\boldsymbol{\delta}_\mu(\lambda_a)$ denotes the dipole derivative vector of the normal mode μ. Matrix $\boldsymbol{\delta}(\lambda_a)$ collects $\boldsymbol{\delta}_\mu(\lambda_a)$ for all normal modes.

$$\boldsymbol{\delta}(\lambda_a)=\mathrm{C}\Delta(\lambda_a)(\mathbf{M}^R(\lambda_a))^{-1/2} \tag{24}$$

The atomic polar tensor (APT) matrix at the point λ_a is denoted with $\Delta(\lambda_a)$. It has the dimension of $3\times 3N_\mathbf{B}$ and contains the dipole moment derivatives with regard to the displacement in Cartesian coordinates.[17,20]

We need to note that the APT matrix along the mutation path is constructed in a similar way via linear combination as the Hessian matrix is constructed [see eq. (5)].

$$\Delta_\lambda=\Delta_\mathbf{A}+\lambda(\Delta_\mathbf{B}-\Delta_\mathbf{A}) \tag{25}$$

Reduced mass matrix $\mathbf{M}^R(\lambda_a)$ is diagonal by collecting $m_\mu^R(\lambda_a)$ as its diagonal elements

$$m_\mu^R(\lambda_a)=\mathbf{l}_\mu^\dagger(\lambda_a)\mathbf{l}_\mu(\lambda_a) \tag{26}$$

In this work, the normal mode intensity $I_\mu(\lambda_a)$ is given in the unit of km/mol. The conversion factor C in eq. (24) is 31.22307.

After calculating the mutation path of methane → ethane with IR intensities taken into consideration, four snapshots of IR spectrum along the path were taken at $\lambda=0, 0.334, 0.667, 1$ (Fig. 5). Although only the IR spectrum diagrams at the point $\lambda=0$ and $\lambda=1$ can be directly compared with experimental data, those intermediate spectrum snapshots are still helpful to visualize how each absorption peak evolves along the mutation path.

In the IR spectrum shown in Figure 5, absorption peaks originated from the reactant (methane) molecule are colored in red, while the others are colored in blue. This color scheme is kept along the mutation path due to the correct ordering of normal modes by the DMO approach.[16] In the methane molecule, only two peaks are observed, where either peak is three-fold degenerate. The other three vibrations including $e(4, 5)$ and $a_1(6)$ are all IR inactive. At the point of $\lambda=0.334$, splitting is observed for $t_2(1,2,3)$ and $t_2(7,8,9)$. Vibrations coming from the growing methyl group have started to play a role. At

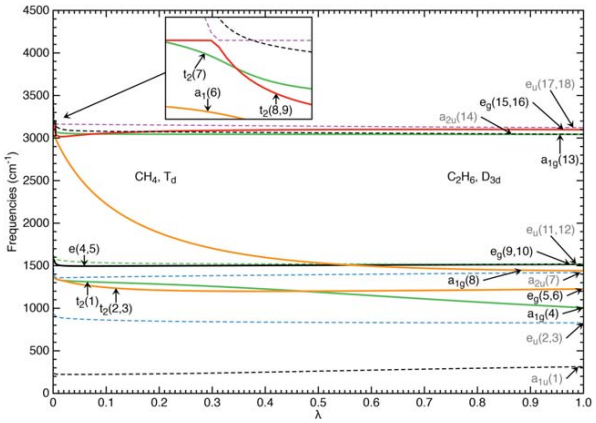

Figure 4. Correlation of normal mode frequencies of methane (left) with that of ethane (right) along the mutation path. Irreducible representations of normal modes are labeled on both sides. Vibrations arising from the growth are plotted with dashed lines and labeled in gray color. [Color figure can be viewed at wileyonlinelibrary.com]

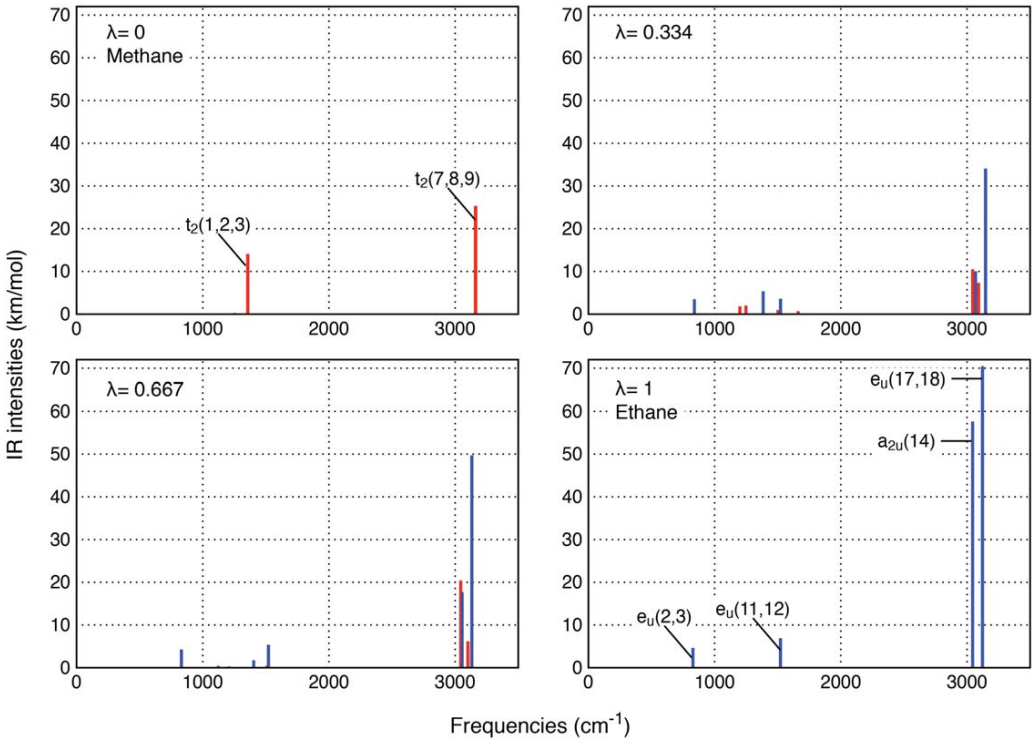

Figure 5. IR intensities of normal vibrational modes along the mutation path from methane to ethane. Four points ($\lambda = 0$, 0.334, 0.667, 1) on the mutation path are shown here. Vibrations from methane are colored in red. Vibrations arising from geometry growth are in blue color. [Color figure can be viewed at wileyonlinelibrary.com]

λ=0.667, we can hardly see any IR active vibrations below 2000 cm^{-1} coming from the methane molecule. One of the three blue peaks in that region is vanishing. The intensities of the blue peaks in the high frequency region are increased. At the end of the mutation path, all red peaks have vanished, leaving only four blue peaks. This result shows that all of the IR active peaks in ethane are not originated from the IR active vibrations in methane, instead they come from the geometry growth.

Apart from methane and ethane, such a correlation extended to the IR spectra can be generalized to other molecules and it can provide experimental chemists with a physically meaningful interpretation of the relationship of vibrations between two structurally related compounds.

Normal mode correlations and their limitations

In the previous subsections, we have correlated the normal vibrational spectra of two structurally related molecules using a mutation path. Each normal vibrational mode μ in molecule **A** has its counterpart μ' in molecule **B**. The normal vibrational frequency values can be denoted as $\omega_{\mu}(A)$ and $\omega_{\mu'}(B)$, respectively. The frequency shift from μ to μ' is unambiguously calculated by $\Delta\omega_{\mu\to\mu'} = \omega_{\mu'}(B) - \omega_{\mu}(A)$.

Independent of the mutation path, chemists have been adopting the idea of normal mode correlation utilizing the concept of normal mode frequency shifts $\Delta\omega_{\mu\to\mu'}$ explicitly or implicitly in various research scenarios in which an external perturbation of the targeted system can be characterized. Many studies have been reported on the red- and blue-shift of vibrations involving covalent X—H bond (X = O, F, or N) stretchings when they participate in non-covalent interactions, that is, hydrogen bonding.[21–28] Such applications using the frequency shift values of a specific normal mode are based on the fact that the perturbation of the electronic structure of the targeted system is revealed via the shift values, thus providing chemical insights from the vibrational spectra.

In a series of studies related to the "vibrational stark effect," Boxer and his coworkers suggested that the C═O bond stretching mode in the carbonyl group and the C≡N bond stretching mode in nitriles are decoupled from other vibrations and thus can be correlated among different chemical systems containing this specific bond.[29–34] Furthermore, they discussed a linear correlation between the normal mode frequency shift of these probe bonds $\Delta\omega_{\mu\to\mu'}$ and the strength of the external electric field exerted on the molecule containing these bonds. The catalytic power of the ketosteroid isomerase was explained using their model.[35,36] Recently, a similar approach has been applied by Mani and coworkers to characterize the localization and delocalization of electrons in anions.[37,38]

Another example in this regard is the Tolman Electronic Parameter.[39,40] Tolman used the shift of the A_1-symmetrical CO stretching mode in $[Ni(CO)_3(L)]$ ($L = PR_3$) complexes as an indirect measure of the Ni—L bond strength assuming that (a) the A_1-symmetrical CO stretching is fully decoupled and (b) it reflects the strength of the Ni—L bonding.

Given these applications of normal mode correlation, the underlying rationale can be summarized as follows. Any attempt to obtain chemical insights by correlating specific normal mode(s) among different systems has to fulfill the following two requirements:

- The absorption peak of the targeted vibrational mode should be distinctive in both the position of the peak (frequency) and the spectroscopic intensity of the vibration. The bond stretching modes having large vibrational frequencies are preferred than angle bending and torsion modes whose frequencies are smaller and prone to be obscured. Furthermore, the IR intensities of these stretching modes range from medium to strong, caused by the significant change in the dipole moment;
- These normal modes should be decoupled from the rest of the molecule and free from the problem of mode-mode coupling. In other words, they are in a way "localized" to a specific part of a molecule and characterize its electronic structure via vibrational frequency values. Only in this way, the comparison of the vibrational frequencies for these spectator modes can make sense.

These two requirements, however, hinder the general applicability of normal mode correlation in contrast to the fact that vibrational spectroscopy is able to characterize the electronic structure of a molecule in a comprehensive manner. Only a small fraction of the abundant information obtained from vibrational spectroscopy is being used by chemists, therefore a way to rediscover the valuable information from the vibrational spectroscopy is needed.

This is presented in the next section of this work, introducing the theory of local vibrational modes which provides a novel approach for utilizing vibrational spectroscopy data in a general, meaningful way.

A Novel Measure of Molecular Similarity Based on Local Vibrational Modes

Theory of local vibrational modes and new similarity measure

In 1998, Konkoli and Cremer did a seminal work on determining the local vibrational modes directly from normal vibrational modes by solving the mass-decoupled Euler-Lagrange equations.[41] Zou and Cremer's recent work has proved that these local vibrational modes are the only and unique counterparts of the normal modes that can be obtained by solving the Wilson equation of vibrational spectroscopy.[42]

First, the Wilson equation of vibrational spectroscopy is solved in eq. (27):

$$\mathbf{F}^x\mathbf{L}=\mathbf{ML}\Lambda \tag{27}$$

where $\mathbf{F}^x$ is the Hessian matrix expressed in Cartesian coordinates, $\mathbf{M}$ is the mass matrix, $\mathbf{L}$ collects all normal mode vectors $\mathbf{l}_\mu$ in columns and Λ is a diagonal matrix collecting corresponding eigenvalues λ_μ. Harmonic vibrational frequencies ω_μ are calculated according to $\lambda_\mu=4\pi^2c^2\omega_\mu^2$.

Then, the Hessian matrix can be transformed into normal coordinates as $\mathbf{K}$:

$$\mathbf{L}^\dagger\mathbf{F}^x\mathbf{L}=\mathbf{K} \tag{28}$$

In the framework of the local vibrational mode theory, each local mode is associated with an internal coordinate parameter q_n, which drives the local mode as the leading parameter. The local mode vector $\mathbf{a}_n$ is given by

$$\mathbf{a}_n=\frac{\mathbf{K}^{-1}\mathbf{d}_n^\dagger}{\mathbf{d}_n\mathbf{K}^{-1}\mathbf{d}_n^\dagger} \tag{29}$$

where $\mathbf{d}_n$ is a row vector in matrix $\mathbf{D}$, which collects the normal mode vectors in internal coordinate. The local mode force constant k_n^a of mode n (superscript a means adiabatically relaxed, i.e., local mode) is thus obtained

$$k_n^a=\mathbf{a}_n^\dagger\mathbf{K}\mathbf{a}_n \tag{30}$$

The reciprocal diagonal element G_{nn} of the $\mathbf{G}$-matrix[41] defines the reduced mass of local mode $\mathbf{a}_n$. The local vibrational frequency ω_n^a can be determined by

$$(\omega_n^a)^2=\frac{1}{4\pi^2c^2}k_n^aG_{nn} \tag{31}$$

Since 2012, we have been applying the theory of local vibrational modes to characterize the intrinsic bond strength of various chemical bonding scenarios covering both covalent bonds[43–46] and also non-covalent interactions.[47–53] In these studies, we have focused on the local bond stretching modes with their local mode properties including the local stretching force constant and the local stretching frequency.

In this work, we explore the potential of local vibrational modes of characterizing the electronic structure of structurally related molecules by correlating their local vibrational modes. Motivated by the normal mode correlation discussed in the last section, we find the local modes are better suited for such a correlation. Given two structurally-related molecules **A** and **B** with their common substructure $\mathbf{a_0}$ and $\mathbf{b_0}$ and specifying an internal coordinate parameter q_n within $\mathbf{a_0}(\mathbf{b_0})$, the corresponding local vibrational modes can be obtained as a and a'. As this pair of local vibrational modes characterizes the curving of the PES of molecules **A** and **B** in a specific direction defined by the same internal coordinate parameter q_n as the leading parameter, the local vibrational mode a in $\mathbf{a_0}$ can be correlated with local vibrational mode a' in $\mathbf{b_0}$. So the local vibrational mode frequency ω_a for a and $\omega_{a'}$ for a' can be directly compared, leading to the *local* mode frequency shift $\Delta\omega_{a\to a'}=\omega_{a'}(B)-\omega_a(A)$ as the local equivalent to the normal mode frequency shift $\Delta\omega_{\mu\to\mu'}$. The major difference between $\Delta\omega_{a\to a'}$ and $\Delta\omega_{\mu\to\mu'}$ is that $\Delta\omega_{a\to a'}$ can characterize the local properties of the electronic structure while $\Delta\omega_{\mu\to\mu'}$ cannot, due its delocalized nature.

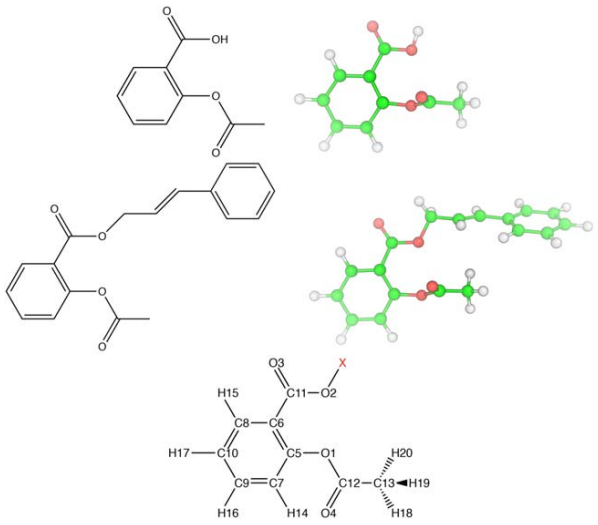

Figure 6. Structure of aspirin (top), aspirin derivative SJ103 (middle) and their common substructure (bottom). [Color figure can be viewed at wileyonlinelibrary.com]

In this way, the local mode frequency shift $\Delta\omega_{a\to a'}$ can be used to describe the difference/similarity of the electronic structure in structurally-related molecules sharing a common substructure. A large amplitude of $\Delta\omega_{a\to a'}$ represents a large difference of the local electronic structure and a small amplitude of $\Delta\omega_{a\to a'}$ indicates higher similarity. The sign of $\Delta\omega_{a\to a'}$ determines red- or blue-shift.

In the following part of this work, we will use three examples to illustrate the application of the local mode correlation and the related local mode frequency shift $\Delta\omega_{a\to a'}$ as a similarity measure.

Example 1: Aspirin and its derivative SJ103

In this example, we use the aspirin molecule and its derivative named as "SJ103"[54] to illustrate how the local mode frequency shift characterizes the difference in the electronic structure of these two molecules in a comprehensive manner.

Figure 6 shows the 2D scheme and ball-and-stick representation of aspirin (top) and its derivative SJ103 (middle). By comparing these two structures, a common substructure can be derived (bottom) so that the *X* in red color can be either a *H* atom or the functional group as —CH_2CH=CH—Ph. Furthermore, *X* is excluded from the substructure and local mode analysis.

For the substructure in question, in total 20 bonds, 25 bond angles, and 9 dihedral angles constitute a non-redundant set of 54 parameters (see Table 2). The local vibrational frequency corresponding to each internal coordinate parameter is calculated for aspirin and SJ103 as $\omega_a(A)$ and $\omega_{a'}(B)$. At the same time, the local mode frequency shift $\Delta\omega_{a\to a'}$ is calculated.

If we take a closer look at the absolute values/amplitudes of $\Delta\omega_{a\to a'}$, several interesting observations should be noted. (1) The local mode frequency shifts for bonds are relatively smaller than those for bond angles and dihedrals. (2) Chemical bonds with $|\Delta\omega_{a\to a'}|$ less than 2 cm^{-1} include 1-12, 6-8, 7-9, 7-14, 8-15, 8-10, 9-16, 9-10, 10-17, and 13-20. Most of these bonds are located in the six-membered ring of the substructure. (3) The C—O bond with largest $|\Delta\omega_{a\to a'}|$ is 2-11 while the

Table 2. Comparison of calculated local vibrational frequencies of the common substructure in Aspirin and its derivative SJ103.

No.	q_n	*i*	*j*	*k*	*l*	$\omega_a(A)$	$\omega_{a'}(B)$	$\Delta\omega_{a\to a'}$
1	r	1	12			1094.9	1095.1	0.2
2	r	1	5			1149.1	1144.4	−4.7
3	r	2	11			1208.0	1218.7	10.7
4	r	3	11			1769.7	1765.7	−4.0
5	r	4	12			1813.7	1803.1	−10.6
6	r	5	6			1330.5	1320.9	−9.6
7	r	5	7			1368.9	1366.1	−2.8
8	r	6	8			1346.1	1346.2	0.1
9	r	6	11			1128.1	1115.5	−12.6
10	r	7	9			1384.3	1385.4	1.1
11	r	7	14			3217.6	3216.1	−1.5
12	r	8	15			3222.9	3222.7	−0.2
13	r	8	10			1391.2	1391.8	0.6
14	r	9	16			3207.2	3206.3	−0.9
15	r	9	10			1372.5	1372.8	0.3
16	r	10	17			3213.3	3212.5	−0.8
17	r	12	13			1117.3	1115.0	−2.3
18	r	13	18			3104.4	3115.4	11.0
19	r	13	19			3165.4	3160.3	−5.1
20	r	13	20			3101.6	3100.9	−0.7
21	α	2	11	3		861.6	898.9	37.3
22	α	2	11	6		709.8	730.8	21.0
23	α	11	6	5		822.9	759.5	−63.4
24	α	6	5	1		740.4	608.8	−131.6
25	α	6	5	7		947.5	945.1	−2.4
26	α	5	1	12		542.0	525.2	−16.8
27	α	1	12	4		882.7	878.5	−4.2
28	α	4	12	13		724.5	725.3	0.8
29	α	12	13	20		972.2	1019.0	46.8
30	α	12	13	19		1136.4	1130.9	−5.5
31	α	12	13	18		1003.3	1046.8	43.5
32	α	18	13	20		1426.9	1423.1	−3.8
33	α	18	13	19		1366.9	1407.0	40.1
34	α	5	7	14		1285.8	1282.2	−3.6
35	α	5	7	9		940.2	928.1	−12.1
36	α	14	7	9		1289.1	1290.8	1.7
37	α	7	9	16		1306.3	1305.3	−1.0
38	α	7	9	10		936.4	934.0	−2.4
39	α	16	9	10		1307.4	1305.3	−2.1
40	α	9	10	17		1306.3	1306.0	−0.3
41	α	9	10	8		943.4	942.5	−0.9
42	α	17	10	8		1308.5	1308.9	0.4
43	α	10	8	15		1309.4	1309.4	0.0
44	α	10	8	6		944.7	923.9	−20.8
45	α	15	8	6		1313.5	1306.9	−6.6
46	τ	2	11	6	5	151.4	119.3	−32.1
47	τ	2	11	6	8	95.3	75.8	−19.5
48	τ	3	11	6	8	173.3	136.2	−37.1
49	τ	15	8	6	11	781.5	776.7	−4.8
50	τ	1	5	6	11	450.3	437.1	−13.2
51	τ	12	1	5	7	150.3	148.7	−1.6
52	τ	4	12	1	5	378.6	410.4	31.8
53	τ	13	12	1	5	218.2	256.3	38.1
54	τ	20	13	12	1	158.8	202.4	43.6

[a] The column "No." is the label of local mode in the common substructure. [b] Columns named as $\omega_a(A)$ and $\omega_{a'}(B)$ are the local vibrational frequencies in aspirin and SJ103. $\Delta\omega_{a\to a'}=\omega_{a'}(B)-\omega_a(A)$. All frequency values are in the unit of cm^{-1}. [c] *i*, *j*, *k*, and *l* are the atomic label that determines the internal coordinate parameters q_n including bond length(r), bond angle(α), and dihedral angle(τ).

C—C bond with largest $|\Delta\omega_{a\to a'}|$ is 6–11. The variation in *X* has a direct influence on 11-2 bond because *O*2 atom is covalently linked to *X*. The π-electron delocalization among atoms *O*2, *O*3, *C*11, and *C*6 within the phenyl ring leads to a secondary effect on the electronic structure of 6-11 bond. (4) Bond angles whose $|\Delta\omega_{a\to a'}|$ less than 10 cm^{-1} include 6-5-7, 1-12-4, 4-12-13, 12-13-19, 18-13-20, 5-7-14, 14-7-9, 7-9-16, 7-9-10, 16-9-10, 9-10-17, 9-10-8, 17-10-8, 10-8-15, and 15-8-6. More than half of these angles are related to the C_6 ring. (5) The largest $|\Delta\omega_{a\to a'}|$ in bond angles is related to 6-5-1 (131.6 cm^{-1}) and 11-6-5 (63.4 cm^{-1}). Angle 11-6-5 has the change in the electronic structure, which has been explained by π-electron delocalization in (3). However, for angle 6-5-1, it is largely influenced by the repulsion force between *O*2 and *O*1. The variation in *X* will have an effect on this repulsion interaction, thus leading to the change in the local vibration of angle 6-5-1. (6) Significant change in the local mode frequency shift is also found for angles 12-13-20, 12-13-18, and 18-13-19. These angles are related to the methyl group in the substructure. Changing from the aspirin to the SJ103 molecule, the bigger functional group *X* has a larger dispersion force on this methyl group.

Furthermore, this approach of correlating a series of local vibrational modes for two structurally related molecules can be extended to study the receptor-ligand binding of protein or DNA in drug design by correlating the local vibrational modes of the ligand in two states, namely the bound ligand and unbound ligand. On binding, the influence from the receptor in the electronic structure of the ligand will be directly reflected by the local mode frequency shift values. If a complete set of local modes of the ligand molecule is available, it is possible to identify the "hot-spots" in the ligand that interact with the receptor.

Example 2: Regioselectivity in Diels-Alder reactions arising from substituents

In this example, we revisit the regioselectivity in Diels-Alder [4 + 2] cycloaddition reactions for 2-substituted dienes with unsymmetrical dienophiles. When a 2-substituted diene with substituent R_1 reacts with a substituted ethene with substituent R_2, two different six-membered rings can result, either the para- or the meta-product, as shown in Figure 7. However, it has been well-recognized that the para-product is the preferred product.[55–57]

Already in the 1970s, Houk explained this phenomenon on the basis of frontier molecular orbital theory.[58,59] Based on this model, the Diels-Alder reactions are driven by HOMO–LUMO interactions of the reactants. The reactions are separated into two categories, depending on the electron donating/withdrawing properties of substituents R_1 and R_2. In the Normal Electron Demand (NED) scenarios, R_1 acts as an Electron Donating Group (EDG), while R_2 is an Electron Withdrawing Group (EWG). Charge transfer is expected from the HOMO of the diene into the LUMO of the dienophile. In the Inverse Electron Demand (IED) cases, electronic flow is from the dienophile to the diene reversing the HOMO/LUMO pair involved,

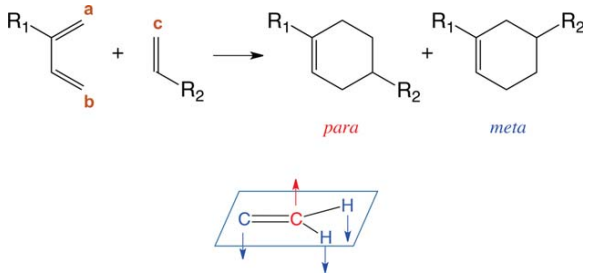

Figure 7. Top: Reaction scheme of Diels-Alder reactions for 2-substituted diene. Bottom: The out-of-plane pyramidalization mode for locations *a*, *b*, and *c*. [Color figure can be viewed at wileyonlinelibrary.com]

that is, R_1 and R_2 act as EWG and EDG, respectively. Houk proposed that in NED-type reactions, the largest HOMO coefficient of the diene is at C1 (location *a* in Fig. 7), while the largest LUMO coefficient of the dienophile is at C2 (location *c*), and that the matching of these two largest HOMO/LUMO coefficients results in para-product. A similar rule applies to the IED-type reaction leading also to the dominance of the para-product.

In this work, we approach the question of regioselectivity from a different perspective. While previous theoretical studies predominantly focused on properties derived from the conceptual density functional theory (DFT),[60–62] we choose to access the electronic structure of the π-orbitals at locations *a*, *b*, and *c* using the local vibrational modes led by the out-of-plane pyramidalization of the carbon atom (colored in red in Fig. 7). We have included 15 2-substituted dienes (see Table 3) and 10 substituted dienophiles (see Table 4) in our investigation and corresponding local vibrational mode frequencies for the pyramidalization mode τ are calculated for positions *a*, *b*, and *c*, denoted as ω_a^τ, ω_b^τ, and ω_c^τ.

To reveal the influence of the substituents R_1 and R_2, respectively on the diene and dienophile, we correlated the local pyramidalization modes in substituted diene/dienophile molecules with those local modes in un-substituted diene/dienophile

Table 3. Local mode frequency shift for 2-substituted *cis*-butadiene with regard to the reference.

Classification	No.	R_1	$\Delta\omega_a^\tau$	$\Delta\omega_b^\tau$
EDG	(1)	OCH_3	−132.5	6.7
	(2)	OC_2H_5	−132.2	7.1
	(3)	OH	−131.3	10.3
	(4)	NH_2	−129.8	9.2
	(5)	$NHCOCH_3$	−53.6	10.2
	(6)	Cl	−36.5	11.1
	(7)	CH_3	−27.9	0.7
	(8)	Ph	−24.3	3.1
Reference	(9)	H	0.0	0.0
EWG	(10)	$COCH_3$	12.4	4.5
	(11)	CN	13.0	15.9
	(12)	CHO	21.1	9.2
	(13)	CF_3	21.5	13.3
	(14)	$COOCH_3$	26.2	2.9
	(15)	NO_2	31.7	17.0
	(16)	COCl	40.4	11.5

[a] The unit for local mode frequency shifts $\Delta\omega_a^\tau$ and $\Delta\omega_b^\tau$ is cm^{-1}.

Table 4. Local mode frequency shift for mono-substituted ethylene with regard to the reference.

Classification	No.	R_2	$\Delta\omega_c^\tau$
EDG	(1)	NH_2	−156.7
	(2)	OCH_3	−138.6
	(3)	CH_3	−36.3
	(4)	Ph	−34.8
Reference	(5)	H	0.0
EWG	(6)	$COCH_3$	11.2
	(7)	CN	14.4
	(8)	CHO	23.4
	(9)	$COOCH_3$	23.6
	(10)	NO_2	27.1
	(11)	COOH	29.9

[a] The unit for local mode frequency shift $\Delta\omega_c^\tau$ is cm^{-1}.

molecules. We calculated corresponding local mode frequency shift $\Delta\omega^\tau$ by subtracting the local mode frequency of a specific pyramidalization mode in the un-substituted diene/dienophile molecule from its counterpart in a substituted diene/dienophile molecule. This leads to the local mode frequency differences $\Delta\omega_a^\tau$, $\Delta\omega_b^\tau$, and $\Delta\omega_c^\tau$.

$$\Delta\omega_a^\tau = \omega_a^\tau(R_1) - \omega_a^\tau(\textit{cis}\text{-butadiene}) \quad (32)$$

$$\Delta\omega_b^\tau = \omega_b^\tau(R_1) - \omega_b^\tau(\textit{cis}\text{-butadiene}) \quad (33)$$

$$\Delta\omega_c^\tau = \omega_c^\tau(R_2) - \omega_c^\tau(\text{ethylene}) \quad (34)$$

Tables 3 and 4 have collected the above three local mode frequency shifts of various substituents for the diene and dienophile, respectively.

For the local mode frequency shift values for 2-substituted dienes, the un-substituted *cis*-butadiene is taken as the reference with $\Delta\omega_a^\tau = \Delta\omega_b^\tau = 0$. The results show that for eight EDGs as R_1, $\Delta\omega_a^\tau$ is negative, while for all seven EWGs, it is positive. For the local mode frequency shift $\Delta\omega_b^\tau$, only positive values for both EDGs and EWGs were found. This indicates that the local mode frequency shift $\Delta\omega_a^\tau$ can be used as a descriptor to distinguish between the two major types of 2-substituted dienes with its substituent as EDG/EWG while $\Delta\omega_b^\tau$ seems to have little influence in this regard.

Interestingly, the local mode frequency shift $\Delta\omega_c^\tau$ for substituted dienophiles has the same behavior for EDGs and EWGs with regard to its sign, namely, negative $\Delta\omega_c^\tau$ for EDGs while positive $\Delta\omega_c^\tau$ for EWGs.

The above results based on the local mode frequency shift are consistent with Houk's approach of checking the largest molecular orbital (MO) coefficients of the carbon atoms at locations *a* and *c*, also we identified $\Delta\omega_a^\tau$ and $\Delta\omega_c^\tau$ as the key descriptors to differentiate EDG and EWG substituents. At first glance, our approach seems to be different from Houk's as we start from vibrations while his approach is based on MOs. However, both approaches access the electronic structure information that decides on the substituent effect.

Furthermore, we applied our local mode approach to verify the experimental findings for combinations of 2-substituted dienes and dienophiles shown presented in Table 5. Most of these reactions have the combination of R_1 = EDG and R_2 = EWG, namely, $\Delta\omega_a^\tau < 0$ and $\Delta\omega_c^\tau > 0$.

We found exceptions for No. 7 and No. 18, where both R_1 and R_2 are EDGs or EWGs. If one checks the $\Delta\omega_a^\tau$ and $\Delta\omega_c^\tau$ values involved in these two reactions, it is easy to find that CH_3 and Ph in reaction no. 7 have the smallest amplitudes in the EDG category for both the 2-substituted diene and dienophile. The same smallest amplitudes are observed for Cl and CN in the EWG category for reaction no. 18. This indicates that these two reactions might have a different mechanism compared with the remaining 16 reactions. Investigations are in progress to demonstrate this.

Table 5. Diels-Alder reactions for 2-substituted dienes with unsymmetrical dienophiles reported in experimental studies.

No.	R_1	R_2	Ref.
(1)	CH_3	$COCH_3$	[63]
(2)	CH_3	CN	[64]
(3)	CH_3	CHO	[64]
(4)	CH_3	$COOCH_3$	[66]
(5)	CH_3	NO_2	[66]
(6)	CH_3	COOH	[67]
(7)*	CH_3	Ph	[64]
(8)	Ph	CN	[64]
(9)	Ph	CHO	[68]
(10)	Ph	$COOCH_3$	[65]
(11)	OCH_3	CN	[69]
(12)	OCH_3	$COOCH_3$	[64]
(13)	OCH_3	$COCH_3$	[70]
(14)	OC_2H_5	CN	[71]
(15)	OC_2H_5	CHO	[72]
(16)	OC_2H_5	$COOCH_3$	[65]
(17)	Cl	$COOCH_3$	[65]
(18)*	CN	$COCH_3$	[65]

[a] All reactions are dominated by the para product.

Table 6. Comparison of pK_a values of p-substituted benzoic acids.

No.	Substituent (R)	ω^R	ω^α	pK_a (exp.)	pK_a (predicted)	error
(1)	NH_2	1570.7	911.1	4.87 [73]	4.87	0.00
(2)	OH	1572.2	910.7	4.58 [74]	4.60	−0.02
(3)	OCH_3	1573.2	910.5	4.47 [75]	4.46	0.01
(4)	CH_3	1573.9	910.3	4.37 [76]	4.35	0.02
(5)	H	1575.0	910.2	4.19 [77]	4.21	−0.02
(6)	F	1575.7	909.8	4.14 [78]	4.15	−0.01
(7)	Cl	1578.3	908.3	3.98 [79]	3.97	0.01
(8)	Br	1578.9	907.9	3.97 [80]	3.92	0.05
(9)	$COCH_3$	1579.8	906.6	3.70 [81]	3.73	−0.03
(10)	CHO	1581.2	906.7	3.75 [80]	3.78	−0.03
(11)	CN	1582.2	905.5	3.55 [76]	3.56	−0.01
(12)	NO_2	1582.8	904.9	3.43 [76]	3.41	0.02

[a] The unit for local mode frequencies ω^R and ω^α is cm^{-1}. pK_a(exp.) denotes pK_a values obtained from experiments, pK_a (predicted) denotes predicted pK_a values using eq. (35), error = pK_a(exp.) – pK_a (predicted).

Example 3: pK_a of p-substituted benzoic acids

We correlated the local CO stretching and O-C-O bending vibrational modes of a series of p-substituted benzoic acids with their pK_a values reflecting their acidity. The acidity of a carboxylic acid is determined by the stability of its conjugated anionic base.

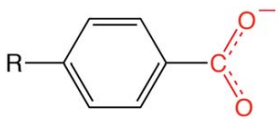

Figure 8. Schematic representation of a p-substituted benzoic acid molecule in its deprotonated state. R is a substituent. [Color figure can be viewed at wileyonlinelibrary.com]

We investigated 12 p-substituted benzoic acid molecules with their pK_a values taken from textbooks, chemical databases and research articles (see Table 6). The general structure of the p-substituted benzoic acids is given in Figure 8, in which the carboxylic acid part is shown as a conjugate base with the substituent R in the para position.

As the substituent R connected to the phenyl ring can be either electron-donating or electron-withdrawing, the diffuse anion of the conjugate base can be either destabilized or stabilized accordingly. This (de)stabilization effect is directly reflected by the change in the electronic structure of the $-CO_2^-$, which can be monitored by the local CO stretching and local O—C—O bending modes.

Table 6 collects the calculated local CO stretching frequency values ω^R and local O-C-O angle bending frequency values ω^α for all 12 benzoic acids. If the two CO bonds (colored in red in Fig. 8) are not identical, for example, in the case of 4-formylbenzoic acid, the averaged value is taken as ω^R. Figure 9 shows the experimentally measured pK_a values is a quadratic function of ω^R and ω^α as two independent variables x and y, respectively.

$$pK_a = 0.004x^2 - 0.075y^2 - 0.038xy + 22.422x + 195.487y - 106455.712 \quad (35)$$

with the coefficient of determination $R^2 = 0.997$ and root-mean-square error RMSE = 0.022. It is noteworthy, that this relationship can be used to predict the pK_a of other p-substituted benzoic acids with unknown pK_a values on the basis of measured and/or calculated local vibrational modes.

Several other theoretical studies predicting the pK_a values of p-substituted benzoic acids have to be mentioned. Hollingsworth and coworkers correlated the pK_a values with several different types of atomic/group charge models calculated with quantum chemical methods.[77] Tao and coworkers constructed a hydrogen bonded complex including the p-substituted benzoic acid and an ammonia molecule and they attempted to correlate the pK_a values with related bond length, normal OH stretching frequency and hydrogen bond energy.[76] The correlation found in this work is stronger ($R^2 = 0.997$ and RMSE = 0.022) compared with Hollingsworth's ($R^2 = 0.978$) and Tao's ($R^2 = 0.952$) results. However, they investigated both m-substituted benzoic acids and p-substituted benzoic acids, while we focused on the p-substituted benzoic acids. Work is in progress to extend our test-set with m-substituted benzoic acids and other types of carboxylic acids.

Computational Details

Equilibrium geometries and normal mode analyses for the CHBrClF enantiomers, benzene(C_6H_6), fluorobenzene benzene(C_6H_5F), methane(CH_4), and ethane(C_2H_6) molecules were calculated using B3LYP density functional[82–85] with Pople's 6–31G(d,p) basis set.[86–92] Aspirin and SJ103 were optimized at the ωB97X-D/cc-pVTZ level of theory.[93,94] Dienes and dienophiles were calculated with ωB97X-D/aug-cc-pVDZ level of theory. p-substituted benzoic acids were calculated with the same hybrid density functional with cc-pVTZ basis set. DFT calculations were carried out with UltraFine integration grid in the Gaussian09 package.[95] All mutation path calculations and local mode analysis were implemented in the program package COLOGNE2017.[96]

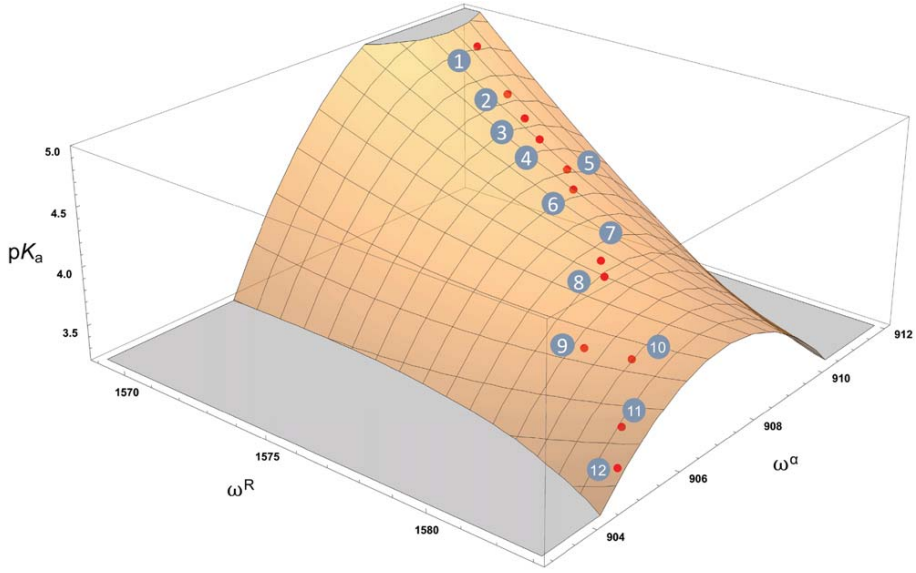

Figure 9. Correlation between pK_a with local CO stretching frequency ω^R and local O-C-O angle bending frequency ω^α on a quadratic surface. [Color figure can be viewed at wileyonlinelibrary.com]

Conclusions

In this work, we have presented a new algorithm that enables correlating the normal vibrational frequencies of any two structurally related molecules. The algorithm is based on the mutation path coordinate λ, which is an extension and generalization of the mass reaction[6] proposed in our previous work. With a mutation path, two structural analogs can be connected via the linear changes with regard to the mass **M**, geometry **R**, Hessian **K**, and symmetry *X*. Furthermore, we have shown that the IR intensities can also be correlated along with the vibrational frequencies, thus leading to the possibility for monitoring how the IR spectrum evolves from one molecule to another along the mutation path. The correlation of Raman activities is planned for our future work. For some pairs of structurally related molecules, there might exist more than one mutation path, but any mutation path can give us a deeper understanding and offer new insights into the relationship between the molecules, which are generally thought to exist as discrete and unconnected objects.

While the mutation path is a useful theoretical tool for normal mode correlation offering deeper physical insights into vibrational spectroscopy, the normal mode correlation itself has been a challenging problem due to the delocalized nature of normal vibrations, leading to limited usefulness of this correlation approach.

To solve this problem, we have developed in this work a new methodology of correlating local vibrational modes, which are derived from normal vibrational modes but which are free from kinematic coupling problem. The local modes are found to be suitable similarity descriptors characterizing local features of the electronic structure. We have shown three examples proving that this approach can be applied for the following purposes in chemical research:

- to compare the electronic structure of two structurally related molecules;
- to identify sites of ligand-receptor interactions on a small molecule;
- to quantify the regioselectivity problem caused by different substituents in chemical reactions[97];
- to predict physicochemical properties (e.g., pK_a) for a series of molecules with different substituents.

Acknowledgment

We thank SMU for providing computational resources.

Keywords: vibrational spectroscopy · mutation path · molecular similarity · normal modes · local modes

How to cite this article: Y. Tao, W. Zou, D. Cremer, E. Kraka. *J. Comput. Chem.* **2018**, *39*, 293–306. DOI: 10.1002/jcc.25109

[1] G. Herzberg, Molecular Spectra and Molecular Structure: *Infrared and Raman Spectra of Polyatomic Molecules*; Krieger Publishing Company: Malabar, FL, USA, **1945**.
[2] L. Woodward, *Introduction to the Theory of Molecular Vibrations and Vibrational Spectroscopy;* Clarendon Press: United Kingdom, **1972**.
[3] S. Califano, *Vibrational States;* Wiley: London, New York, Sydney and Toronto, **1976**. ISBN 9780471129967.
[4] K. Nakanishi, P. Solomon, *Infrared Absorption Spectroscopy;* Holden-Day: San Francisco, **1977**. ISBN 9780816262519.
[5] N. B. Colthup, L. Daly, S. Wiberley, *Introduction to Infrared and Raman Spectroscopy*; Academic Press: New York, **1964**. ISBN 9780121825546.
[6] A. Wu, D. Cremer, *J. Phys. Chem. A* **2003**, *107*, 10272.
[7] W. H. Miller, N. C. Handy, J. E. Adams, *J. Chem. Phys.* **1980**, *72*, 99.
[8] M. Page, J. W. McIver, *J. Chem. Phys.* **1988**, *88*, 922.
[9] Z. Konkoli, E. Kraka, D. Cremer, *J. Phys. Chem. A* **1997**, *101*, 1742.
[10] E. Kraka, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 531.
[11] W. Zou, T. Sexton, E. Kraka, M. Freindorf, D. Cremer, *J. Chem. Theory Comput.* **2016**, *12*, 650.
[12] R. E. Stanton, J. W. McIver, *J. Am. Chem. Soc.* **1975**, *97*, 3632.
[13] K. Fukui, *Acc. Chem. Res.* **1981**, *14*, 363.
[14] R. W. Zwanzig, *J. Chem. Phys.* **1954**, *22*, 1420.
[15] E. B. Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra (Dover Books on Chemistry);* Dover Publications: Mineola, NY, USA, **2012**.
[16] Z. Konkoli, D. Cremer, E. Kraka, *J. Comput. Chem.* **1997**, *18*, 1282.
[17] B. S. Galabov, T. Dudev, *Vibrational Intensities*, Vol. *22*; Elsevier: Amsterdam, **1996**.
[18] W. Person, G. Zerbi, *Vibrational Intensities in Infrared and Raman Spectroscopy* (Studies in Physical and Theoretical Chemistry); Elsevier Scientific Publishing Company: Amsterdam, **1982**. ISBN 9780444416995
[19] Y. Yamaguchi, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory;* Oxford University Press: Oxford, **1994**.
[20] W. B. Person, J. H. Newton, *J. Chem. Phys.* **1974**, *61*, 1040.
[21] R. M. Badger, S. H. Bauer, *J. Chem. Phys.* **1937**, *5*, 839.
[22] D. Kraemer, M. Cowan, A. Paarmann, N. Huse, E. Nibbering, T. Elsaesser, R. D. Miller, *Proc. Natl. Acad. Sci.* **2008**, *105*, 437.
[23] J. Joseph, E. D. Jemmis, *J. Am. Chem. Soc.* **2007**, *129*, 4620.
[24] S. Scheiner, T. Kar, *J. Phys. Chem. A* **2002**, *106*, 1784.
[25] X. Li, L. Liu, H. B. Schlegel, *J. Am. Chem. Soc.* **2002**, *124*, 9639.
[26] K. Hermansson, *J. Phys. Chem. A* **2002**, *106*, 4695.
[27] B. J. van der Veken, W. A. Herrebout, R. Szostak, D. N. Shchepkin, Z. Havlas, P. Hobza, *J. Am. Chem. Soc.* **2001**, *123*, 12290.
[28] A. Masunov, J. J. Dannenberg, R. H. Contreras, *J. Phys. Chem. A* **2001**, *105*, 4737.
[29] A. Chattopadhyay, S. G. Boxer, *J. Am. Chem. Soc.* **1995**, *117*, 1449.
[30] S. G. Boxer, *J. Phys. Chem. B* **2009**, *113*, 2972.
[31] S. S. Andrews, S. G. Boxer, *J. Phys. Chem. A* **2000**, *104*, 11853.
[32] S. S. Andrews, S. G. Boxer, *J. Phys. Chem. A* **2002**, *106*, 469.
[33] S. D. Fried, S. G. Boxer, *Acc. Chem. Res.* **2015**, *48*, 998.
[34] L. J. Webb, S. G. Boxer, *Biochemistry* **2008**, *47*, 1588.
[35] S. D. Fried, S. Bagchi, S. G. Boxer, *Science* **2014**, *346*, 1510.
[36] Y. Wu, S. G. Boxer, *J. Am. Chem. Soc.* **2016**, *138*, 11890.
[37] T. Mani, D. C. Grills, M. D. Newton, J. R. Miller, *J. Am. Chem. Soc.* **2015**, *137*, 10979.
[38] T. Mani, D. C. Grills, *J. Phys. Chem. B* **2017**, *121*, 7327.
[39] C. A. Tolman, *J. Am. Chem. Soc.* **1970**, *92*, 2953.
[40] C. A. Tolman, *Chem. Rev.* **1977**, *77*, 313.
[41] Z. Konkoli, D. Cremer, *Int. J. Quantum Chem.* **1998**, *67*, 1.
[42] W. Zou, R. Kalescky, E. Kraka, D. Cremer, *J. Chem. Phys.* **2012**, *137*, 084114.
[43] R. Kalescky, E. Kraka, D. Cremer, *J. Phys. Chem. A* **2013**, *117*, 8981.
[44] A. Humason, W. Zou, D. Cremer, *J. Phys. Chem. A* **2015**, *119*, 1666.
[45] W. Zou, D. Cremer, *Chemistry* **2016**, *22*, 4087.
[46] D. Cremer, E. Kraka, *Dalton Trans.* **2017**, *46*, 8323.
[47] R. Kalescky, W. Zou, E. Kraka, D. Cremer, *Chem. Phys. Lett.* **2012**, *554*, 243.
[48] M. Freindorf, E. Kraka, D. Cremer, *Int. J. Quantum Chem.* **2012**, *112*, 3174.
[49] V. Oliveira, E. Kraka, D. Cremer, *Inorg. Chem.* **2017**, *56*, 488.
[50] V. Oliveira, E. Kraka, D. Cremer, *Phys. Chem. Chem. Phys.* **2016**, *18*, 33031.
[51] D. Setiawan, E. Kraka, D. Cremer, *J. Phys. Chem. A* **2015**, *119*, 1642.

[52] Y. Tao, W. Zou, J. Jia, W. Li, D. Cremer, *J. Chem. Theory Comput.* **2017**, *13*, 55.
[53] Y. Tao, W. Zou, E. Kraka, *Chem. Phys. Lett.* **2017**, *685*, 251.
[54] B. C. Cha, S. B. Lee, *Arch. Pharm. Res.* **2000**, *23*, 116.
[55] F. Carey, R. Giuliano, *Organic Chemistry;* McGraw-Hill Higher Education: New York, **2013**.
[56] F. A. Carey, R. J. Sundberg, *Advanced Organic Chemistry: Part A: Structure and Mechanisms: Structure and Mechanisms Pt. A;* Springer: New York, **2007**.
[57] G.-M. Ho, C.-J. Huang, E. Y.-T. Li, S.-K. Hsu, T. Wu, M. M. L. Zulueta, K. B. Wu, S.-C. Hung, *Sci. Rep.* **2016**, *6*, 35147.
[58] K. N. Houk, *J. Am. Chem. Soc.* **1973**, *95*, 4092.
[59] K. N. Houk, *Acc. Chem. Res.* **1975**, *8*, 361.
[60] C. Morell, P. W. Ayers, A. Grand, S. Gutiérrez-Oliva, A. Toro-Labbe, *Phys. Chem. Chem. Phys.* **2008**, *10*, 7239.
[61] S. Damoun, G. Van de Woude, F. Mendez, P. Geerlings, *J. Phys. Chem. A* **1997**, *101*, 886.
[62] P. Geerlings, F. De Proft, *Int. J. Quantum Chem.* **2000**, *80*, 227.
[63] K. Matuszek, S. Coffie, A. Chrobok, M. Swadźba-Kwaśny, *Catal. Sci. Technol.* **2017**, *7*, 1045.
[64] O. Eisentein, J. Lefour, N. T. Anh, R. Hudson, *Tetrahedron* **1977**, *33*, 523.
[65] J. Sauer, *Angew. Chem. Int. Ed. Engl.* **1967**, *6*, 16.
[66] D. A. Singleton, K. Inomata, In *Encyclopedia of Reagents for Organic Synthesis;* Wiley: Oxford, **2008**.
[67] P. Zhang, R. M. Kriegel, J. W. Frost, *ACS Sustain. Chem. Eng.* **2016**, *4*, 6991.
[68] F. Fringuelli, O. Piermatti, F. Pizzo, L. Vaccaro, *Sci. Synth.* **2010**, *47b*, 561.
[69] P. Kumar, P. Li, I. Korboukh, T. L. Wang, H. Yennawar, S. M. Weinreb, *J. Org. Chem.* **2011**, *76*, 2094.
[70] W. D. Wulff, Y. C. Xu, *J. Am. Chem. Soc.* **1988**, *110*, 2312.
[71] J. Doucet, P. Rumpf, Bull. Soc. Chim. Fr. **1954**, 610.
[72] H. Fiesselmann, *Ber. Dtsch. Chem. Ges.* **1942**, *75B*, 881.
[73] D. R. Lide, *CRC Handbook of Chemistry and Physics: Special Student Edition*, 77th ed.; CRC Press: New York, **1996**. ISBN 0849305969.
[74] X. Wu, D. Mangelings, I. Oita, C. Yan, Y. V. Heyden, *J. Sep. Sci.* **2011**, *34*, 2305.
[75] N. C. for Biotechnology Information, *Pubchem compound database; cid=7478*, **2017**. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/7478 (Accessed on October 15, 2017).
[76] L. Tao, J. Han, F.-M. Tao, *J. Phys. Chem. A* **2008**, *112*, 775.
[77] C. A. Hollingsworth, P. G. Seybold, C. M. Hadad, *Int. J. Quantum Chem.* **2002**, *90*, 1396.
[78] E. V. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*; University Science: Sausalito, CA, USA, **2005**. ISBN 1891389319.
[79] N. C. for Biotechnology Information, *Pubchem compound database; cid=6318*, **2017**. Available at: https://pubchem.ncbi.nlm.nih.gov/compound/6318 (Accessed on October 15, 2017).
[80] J. McMurry, *Organic Chemistry;* Brooks Cole: London, **1999**. ISBN 0534362745
[81] J. M. Hornback, *Organic Chemistry;* Cengage Learning: Boston, MA, USA, **2005**. ISBN 0534389511
[82] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[83] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[84] S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.
[85] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
[86] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724.
[87] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.
[88] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213.
[89] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, *J. Chem. Phys.* **1982**, *77*, 3654.
[90] R. C. Binning, L. A. Curtiss, *J. Comput. Chem.* **1990**, *11*, 1206.
[91] V. A. Rassolov, J. A. Pople, M. A. Ratner, T. L. Windus, *J. Chem. Phys.* **1998**, *109*, 1223.
[92] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, L. A. Curtiss, *J. Comput. Chem.* **2001**, *22*, 976.
[93] J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
[94] T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.
[95] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09, Revision E.01*; Gaussian, Inc.: Wallingford, CT, **2009**.
[96] E. Kraka, W. Zou, M. Filatov, Y. Tao, J. Grafenstein, D. Izotov, J. Gauss, Y. He, A. Wu, Z. Konkoli, V. Polo, L. Olsson, Z. He, D. Cremer, *COLOGNE2017*, **2017**. Available at: http://www.smu.edu/catco
[97] Y. Tao, W. Zou, D. Cremer, E. Kraka, *J. Phys. Chem. A* **2017**, *121*, 8086.

Received: 29 June 2017
Revised: 16 October 2017
Accepted: 22 October 2017
Published online on 16 November 2017