Spring 5-19-2018

Advances in Local Vibrational Mode Theory and Unified Reaction Valley Approach (URVA)

Yunwen Tao
Southern Methodist University, yunwent@smu.edu

ADVANCES IN LOCAL VIBRATIONAL MODE THEORY AND UNIFIED REACTION VALLEY APPROACH (URVA)

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

in

Theoretical and Computational Chemistry

by

Yunwen Tao

B.S., Bio-technology, University of Electronic Science and Technology of China

May 19, 2018

ACKNOWLEDGMENTS

First and foremost I want to thank my advisors Dieter Cremer and Elfi Kraka. I really appreciate that they provided me with an opportunity to come to the United States to pursue my Ph.D. in computational chemistry. They have passed on their abundant knowledge and experiences to me selflessly, shaping me into an independent researcher in the past years.

It is unfortunate that we lost Dr. Cremer earlier last year and it is a huge loss for me as one of his graduate students. His passion for science and high standard for work has been always encouraging to me to continue his unfinished collaboration with me. I am also thankful to Dr. Kraka for taking over as my primary advisor and giving me tremendous assistance in paper writing and publishing. It is a great honor for me to be one of the graduate students of Drs. Cremer and Kraka.

The members of the CATCO group have contributed immensely to help me with getting my Ph.D. professionally and personally. I am especially grateful to Dr. Wenli Zou who introduced me into the CATCO group while he was working as a postdoctoral researcher at SMU. We worked very closely in my first year of Ph.D. and I learned a lot from him in Fortran programming and theories of molecular vibrations and reaction paths. After he went back to China to be a full professor, he has been still very helpful in refining my manuscripts to be published. I would like to acknowledge one CATCO group alumnus Dr. Robert Kalescky, who gave me the first lesson on Python programming and helped me a lot to use the super-computing facilities at SMU. Dr. Marek Freindorf is acknowledged for his practical trainings in the group on various topics, including QM/MM. Other group members that I have had the pleasure to work with are graduate students Dr. Thomas Sexton, Dr. Vytor Oliveira, Dr. Alan Humason and Niraj Verma.

Besides, I want to thank my undergraduate advisor Dr. Xianlong Wang for his continuous encouragement and support. The work on the Generalized Subsystem Vibrational Analysis

would not have been possible without the inspiration from Chuan Tian in the Simmerling group at Stony Brook University.

For this dissertation I would like to thank the members of my dissertation committee, Dr. Elfi Kraka, Dr. Peng Tao, Dr. Weihua Geng and Dr. Marek Freindorf, for their time, interest and helpful comments.

My time at SMU was made enjoyable in large part due to many friends which became a part of my Ph.D. life. I am grateful for time spent with many friends also from China. My life in Dallas was also enriched by many friends at the Park Cities Presbyterian Church and First Baptist Dallas - Chinese Fellowship.

Lastly, I would like to thank my family for all their love and support.

Tao, Yunwen B.S., Bio-technology, University of Electronic Science and Technology of China

Advances in Local Vibrational Mode Theory
and Unified Reaction Valley Approach (URVA)

Advisor: Dr. Elfi Kraka
Doctor of Philosophy degree conferred May 19, 2018
Dissertation completed April 26, 2018

Since the establishment of the Local Vibrational Mode Theory and the Unified Reaction Valley Approach (URVA), these two research pillars have been pushed forward in the CATCO group and played an important role in (i) characterizing the chemical bonds in molecules and (ii) molecular chemical reactions. This dissertation elaborates my contributions to the Local Vibrational Mode Theory and the Unified Reaction Valley Approach (URVA).

We have applied the Local Vibrational Mode Theory to hydrogen bonding in liquid water and proposed an explanation for the Mpemba effect. We explored and discovered new directions of applying local vibrational modes majorly in characterizing substituent effect. We have extended the local vibrational mode idea, developing intrinsically comparable normal vibrational modes.

My major contribution to the Unified Reaction Valley Approach has been the coding of a standalone URVA analysis program, which is easy to use and maintain and independent of the software used to create the input data . This program will be published in the future as an open-source tool helping chemists to understand chemical reactions in all details.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

n	number of atoms for the sub-system
k	total number of translations and rotations for the sub-system
N	number of atoms for the complete system
K	total number of translations and rotations for the complete system
k_n^a	adiabatic force constant
λ	mutation path coordinate
$\mathbf{b}, \mathbf{B}, \mathbf{B}', \mathbf{B}'_{sub}$	Wilson B-matrix
$\mathbf{f}^x$, $\mathbf{F}$	full Hessian matrix
$\mathbf{f}^x_{sub}$	effective Hessian matrix
$\boldsymbol{\nu}$	reaction path direction
$\boldsymbol{\nu}^{(1)}$	reaction path curvature
c	norm of mass-weighted gradient
$\mathbf{F}^{(1)}$	third derivative of the energy along the reaction path
s	reaction path parameter

In Memoriam of Dr. Dieter Cremer.

Chapter 1

RESEARCH OVERVIEW

I started my Ph.D. program when I came to the United States and joined the Computational and Theoretical Chemistry Group (CATCO) at Southern Methodist University as a graduate student in August 2014.

Before all of this happened, I was very interested in applying computational chemistry to biomacromolecules including proteins. This was a result of my background in biology and related research experience as an undergraduate student. One of the motivations for me to choose Dr. Dieter Cremer[1] and Dr. Elfi Kraka as my Ph.D. advisors is their work on the Automatic Protein Structure Analysis (APSA). [31]

However, Dr. Cremer wanted me to work on other research topics which turned out to be a very good suggestion. I have been really enjoying my work in the CATCO group, which has been far beyond my expectations and has opened the path to my future career in academia.

As the foundation of applying for my degree as a Doctor of Philosophy, I have worked on the following two major directions:

- The Local Vibrational Mode Theory
- Unified Reaction Valley Approach (URVA)

From August 2014 to the time when I started writing this dissertation (9^{th} March, 2018), I have five research articles published as the first author in high-ranking journals. Besides, I have finished one major computer program with 6,000 statements.

[1] In memoriam, 1944 - 2017

Chapter 2

LOCAL VIBRATIONAL MODE THEORY

2.1. History

Vibrational spectroscopy has turned out to be one of the most useful tools for experimental chemists in structural determination and the characterization of electronic structure as well as its change under certain perturbations.

The atoms in any molecule are not static, instead they are vibrating constantly in compliance with the Heisenberg's uncertainty principle. An early contribution to the field of molecular vibration is from Gerhard Herzberg, who wrote the famous book *Infrared and Raman Spectra of Polyatomic Molecules* in 1945 as a continuation of *Spectra of Diatomic Molecules*, being part of his book series on *Molecular Spectra and Molecular Structure*.

A generally applicable theory describing molecular vibrations was not proposed until E. Bright Wilson, Jr., J. C. Decius and Paul C. Cross wrote the fundamental book *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra* in 1955. This book describes the underlying mathematics involved in molecular vibrations, leading to the so-called Wilson equation of vibrational spectroscopy.

From the application of vibrational spectroscopy to an increasing number of molecules, chemists found out that there exists a strong correlation between a specific range of vibrational frequencies and certain functional groups. In this way, the characteristic absorption peaks from vibrational spectra help chemists to characterize the structure of unknown chemical compounds.

The close connection between the vibrations and molecular fragment inspired chemists to find ways to characterize a specific geometric unit, e.g. a bond between two atoms, or a bond angle in three atoms, etc. with vibrational frequencies. However, it was not until 1998

when Zoran Konkoli, J. Andreas Larsson and Dieter Cremer gave a thorough insight into this problem and proposed the theory of *adiabatic internal modes*, providing a physically based description of local vibration of fragments in a molecule [15,16,19,20]. The synonyms of adiabatic internal modes include local (vibrational) modes, adiabatic local (vibrational) modes, Konkoli-Cremer or Konkoli-Zou-Cremer modes.

The local vibrational modes are derived from mass-decoupled Euler-Lagrange equations. We quote the original definition [15] by Konkoli and Cremer in the following.

> *An internal mode is localized in a molecular fragment by describing the rest of the molecule as a collection of massless points that just define molecular geometry. Alternatively, one can consider the new fragment motions as motions that are obtained after relaxing all parts of the vibrating molecule but the fragment under consideration.*

This is an ingenious idea because it connects vibrational spectroscopy with the molecular fragment in question based on a physically solid model. However, this model would not be useful without the development of computational chemistry. Fortunately, analytical second-order energy derivatives with regard to Cartesian coordinates leading to vibrational modes and frequencies for the Hartree-Fock theory have been available since the 1980s, followed in the next decades for post-SCF and DFT methods.

The local vibrational modes were initially applied to study chemical bonds and the relationship between bond length and bond strength [2,21,23].

In 2012, a major breakthrough in the theoretical perspective was made by Wenli Zou [43], who was a post-doctoral researcher working with Dieter Cremer at SMU. Zou and Cremer proved that

> *The local vibrational modes are the true counterparts of the delocalized normal vibrational modes.*

This means there exists a one-to-one relationship between a non-redundant set of $3N-K$ local vibrational modes and $3N-K$ normal vibrational modes. The proof was given based

on the compliance constants of Decius [3]. Meanwhile, Zou and Cremer have proved that the reciprocal of the compliance constant for a specific internal coordinate is identical to its local vibrational force constant. Besides, this work has provided the physical basis for compliance constant which was criticized for its loose connection to vibrational spectroscopy.

The local vibrational modes have been extensively applied to characterize the intrinsic bond strength of a broad range of chemical bonds, covering both covalent bonds [9, 11, 13, 22, 32, 42] and non-covalent interactions including hydrogen bond [4, 10, 12, 37, 38], halogen bond [29], pnicogen bond [33] and atom-π interactions [40, 45].

Later in 2014, Zou and Cremer derived the equations for calculating infrared intensities of local vibrational modes [41].

In one of my research projects I employed the local vibrational modes as similarity descriptors in several chemical problems related to substituent effects [35, 36]. In 2017, I proposed the *intrinsic fragmental vibrations* as a key intermediate between normal vibrational modes and local vibrational modes.

2.2. Summary of My Contributions

My work on the local vibrational modes consists of three sections.

1. Characterizing the intrinsic bond strength of hydrogen bonding in water;
2. Quantifying chemical similarity based on local vibrational modes;
3. Development of the Generalized Subsystem Vibrational Analysis (GSVA).

The three topics will be elaborated in the following sections. One promising direction for future applications is discussed.

2.3. H-Bonding in Water

This work was my first research project after I joined CATCO. The original objective of this work was to quantify the intrinsic bond strength of hydrogen bonds between water

molecules in water cluster models in parallel to other investigations on non-covalent interactions within the CATCO group.

This project was a collaborative work with Nanjing University in China, where Dr. Wei Li and his co-workers had developed an in-house approach called Generalized Energy-Based Fragmentation (GEBF) [24], designed to handle *ab initio* calculations for large systems. While most *ab initio* methods can handle small to medium systems with up to 200 atoms, the GEBF approach can easily deal with systems being composed of thousands of atoms with equivalent accuracy. One advantage of the GEBF approach over other fragmentation methods including Fragment Molecular Orbital (FMO) [14] and Effective Fragment Potential (EFP) [8] is that GEBF has already analytical energy derivatives with regard to Cartesian coordinates up to the second order, so that the normal vibrational mode analysis as well as local mode analysis is possible.

In the very beginning, Dr. Li did simulated annealing for several different water clusters of $(H_2O)_{50}$ using the TIP5P force field [26]. This gave us the starting geometries of water clusters for *ab initio* calculations which are more accurate but time-consuming. That is the reason why we choose to calculate a cluster having 50 water molecules instead of 500 with *ab initio* methods. Then Dr. Li used their GEBF approach in combination with Density Functional Theory (DFT) to calculate these $(H_2O)_{50}$ water clusters, leading to four clusters optimized to their local minima in energy.

One might argue that it is not sufficient to investigate just four water clusters, however, this led to a total of 350 hydrogen bonds in these clusters.

In order to analyze these hydrogen bonds in a systematic manner, we proposed a classification rule based on topology. As each hydrogen bond in water is formed by the donor water together with the acceptor water, the targeted H-bond is directly influenced by the way how the donor and acceptor interact with surrounding water molecules via additional H-bonds. We developed a four-digit code to categorize different situations, as **AB-CD**. **A** denotes the H-bonds the donor water accepts, **B** is the number of extra H-bonds the donor water donates. **C** and **D** are the numbers of extra H-bonds the acceptor water accepts and

donates respectively. For example, in the water dimer structure, its H-bond code is **00-00**.

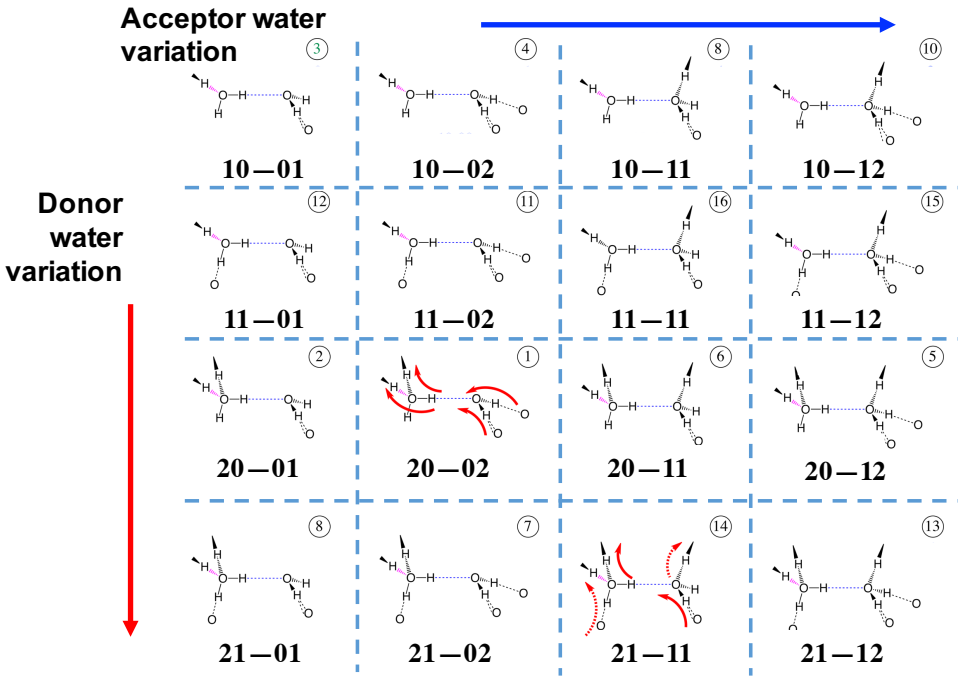


Figure 2.1. Schematic presentation of different ways of H-bonding in water.

After carrying out the local vibrational mode analysis, we found that different types of H-bonds have different bond strengths. The rule we can generalize from the results is that an H-bond with large **A** and **D** with small **B** and **C** is often stronger. In the $(H_2O)_{50}$ cluster models, the strongest type is in **20-02** caused by a push-pull effect. Such a cooperative effect in hydrogen bonding was actually discovered in experiments before [34], however, theoretical investigations in this connection were missing. The reason for this might be due to the lack of proper theoretical tools for other chemists to quantify the strength of H-bonds in a reliable way. We name this cooperative influence as "push-pull" effect, which was later found in other H-bonds involved with hydrogen fluoride and ammonia in a follow-up investigation.

Then we tried to correlate our results of the H-bond strength calculated based on local vibrational modes to the properties of water. I found on a website[1] created by Martin Chaplin from London South Band University that water is a liquid which has many peculiarities

[1] http://www1.lsbu.ac.uk/water/

related to its density, specific heat capacity, cooling and so forth. After reading the list of these special water properties, Dr. Cremer proposed an idea of explaining the Mpemba effect, which says *warm water freezes faster than cold water*.

Dr. Cremer proposed that warm water may have fewer weak **21-12** H-bonds but more strong H-bonds, cold water has more weak H-bonds which requires extra time to be broken in freezing. However, this was a speculation, because it is almost impossible to simulate the freezing process of water with different starting temperatures with sufficient accuracy, due to hardware limitations and the current lack of methods for accurately simulating non-equilibrium processes. When we submitted our first manuscript, we kept this part as a speculation with no proof. Afterwards, we got the comments from the reviewers who are experts in studying H-bonding of water. One reviewer raised several pages of questions, one of which is how to confirm our speculation related to the Mpemba effect. Although we could not simulate the cooling and freezing of water, we could still gain important insights by modeling the initial state of liquid water at different temperatures. So we did molecular dynamics simulations of water in several different temperatures. The results indeed proved that warm water has fewer weak H-bonds, but relatively more strong H-bonds. So that the formation of ice crystals becomes easier. A schematic summary of our explanation is shown in Figure 2.2.

After the publication of our work in *J. Chem. Theory Comput.*, it attracted considerable attention in the media. What I learned from this project is that it is always challenging to propose a new theory as a newcomer to a field where experts might raise overwhelming questions. On the other hand, such a challenge is worth facing because experts can give us deeper insights into our work which can be improved on this basis.

The water H-bonding paper and the push-pull effect paper can be found in Appendices A and B.

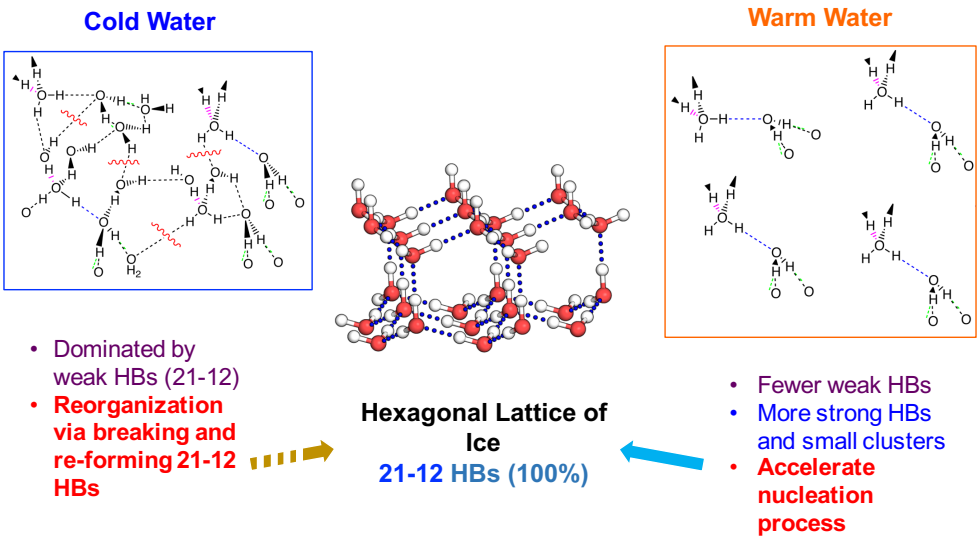


Figure 2.2. Explanation of the Mpemba effect in cold and warm water.

2.4. Local Vibrational Modes as Similarity Descriptors

In the past three years, several research projects in CATCO focused on the characterization of chemical bonding with local vibrational modes. The important essence of these work is quantifying the intrinsic bond strength with local mode force constants between two atoms for this bond.

Dr. Cremer proposed the idea that vibrational spectroscopy should be able to provide a metric for measuring the similarity of chemical compounds. The first project in this direction I worked on was to extend the "mass reaction" path proposed by one of Dr. Cremer's former students An-An Wu [39]. In a mass reaction, the vibrational frequencies are calculated when the atomic masses within a molecule are gradually changed from one isotope to another, so that the vibrational modes in these two molecules can be accurately correlated. Then we wondered whether it is possible to correlate the vibrations of two different but structurally similar molecules, e.g. methane (CH_4) and ethane (C_2H_6) and based on this correlation to work out the similarity between these two molecules.

When I started to solve this problem, I already had some experience in solving the Wilson equation of vibrational spectroscopy by writing a program. This knowledge was acquired from a Gaussian white paper [28] and discussions with Wenli Zou. The most straightforward approach to solve this problem is to check into three quantities required by the Wilson equation, including the geometry, atomic masses and Hessian matrix. A Hessian matrix collects the second order energy derivatives of a molecule with regard to Cartesian coordinates. An-An Wu's mass reaction is in fact a special case where changes are only in the atomic masses via linear interpolation. However, if we want to connect two different molecules, the only solution is to find a linear interpolation for all three quantities. In this spirit, we have to develop an theoretical process called *mutation path* which is similar to the *alchemical free energy perturbation* [46] in molecular dynamics simulations. Namely a molecule has to grow into another one. This mutation path involves the changes in Hessian, geometry and atomic masses at the same time. All of these changes are in the form of linear interpolation, controlled by a global parameter λ ranging from 0 to 1. In this way, λ=0 and

λ=1 connects the starting molecule and the end molecule on both sides. For the points in between, the mutation complex has no physical meaning because we can never observe such a transforming molecule in real world. However, the mutation path is the best solution we can take if we want to correlate the vibrations of two different molecules.

Interestingly, the mutation path worked well. However, it could not work without the help of the Diabatic Mode Ordering (DMO) approach [17], which resolves all allowed- and avoided-crossings of vibrational modes along the path. This is non-trivial because we need to keep track of every vibrational mode at any point on the path, otherwise when two vibrations come close to each other in frequency values, we will lose track of these modes. Besides, the limitations of the mutation path should also be addressed. Before we do a mutation path calculation, we need to align the common substructure of two structurally related molecules in Cartesian coordinates. For example, we have to align the two methyl groups (CH_3) from the methane and ethane molecules first to make sure that the we only need minimal mutation. The mutation path results are very interesting for two molecules with high symmetry.

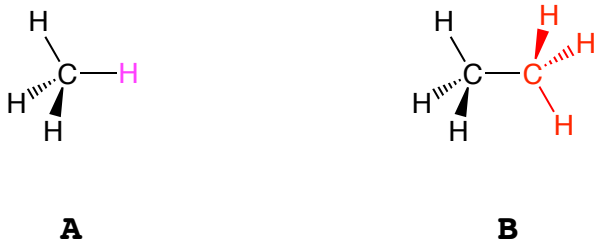


Figure 2.3. Structure of reactant and product in the mutation path of methane → ethane.

With the mutation path, we made for the first time the correlation of vibrations in two different molecules possible .

Then we started to develop a new measure of similarity based on vibrations. The first idea for this task was to calculate the local vibrational frequencies of internal coordinates within the common substructure of different molecules. By checking the amplitude of the deviation in frequencies, we would be able to identify the hot-spot where the electronic

structure of both molecules shows the largest difference.

It occurred to me that local modes could be used to predict the reactive sites of a molecule, models derived from wave-function analysis [6, 25] have been trying so far with more or less success. I tried to see if the local modes can be applied to the regio-selectivity problem of mono-substituted benzene molecules. My first test set had only 30 molecules, and I selected the local vibrational frequencies of bond stretching, angle bending and out-of-plane pyramidalization on the phenyl ring part as the input feature vector. Then I tried to employ the hierarchical clustering analysis to separate the meta-directing and ortho/para-directing groups. At the same time, I tried to cluster the functional groups which should be clustered together according to our common knowledge. The whole trial-and-error procedure was carried out in a program called *Statistical Package for the Social Sciences (SPSS)*. Although this package was originally designed for social sciences, it is fairly easy to handle and test with mouse clicks. The attempts in this direction were quite successful.

This work led to two papers. One is focused on the mutation path and the other is tailored for the benzene molecules. The second paper was well accepted by the reviewers, however, the first one was required to have a major revision. The reason why we got stuck by the reviewers is because that paper was organized into two parts. The first one introduced the mutation path in order to correlate the vibrational modes of two different molecules, the second one jumped immediately to use the local vibrational frequencies to characterize the hot-spots among different molecules. The problem lies in the transition of logics between two sections.

Afterwards, efforts were made in two directions. First, a smooth transition would be necessary to add. Second, it would be better if we put more examples for the local modes part, because the mutation path part has three examples while the local mode part has only one.

With regard to the transition problem, we figured out a new perspective that can help to connect two sections. That is moving from correlation of normal vibrations to the correlation of local vibrations. Chemists employing vibrational spectroscopy in order to obtain

chemical insights were always correlating the normal vibrations, however, the normal modes are delocalized in nature and only high-frequency vibrations of bonds with significant spectral intensities were really used. This leads to the insufficient utilization of vibrational spectroscopy and difficulties in using vibrational data in a quantitative way for obtaining chemical insights.

In order to add more examples for correlation of local modes, I looked into an organic chemistry textbook and tried to find examples related to the substituent effect. Interesting is that many topics are related to the substituent effect besides the regioselectivity in electrophilic aromatic substitution reactions. Molecular acidity quantified by pK_a values and regioselectivity in many other reactions are all determined by the substituent effect. However, we need to be careful that the steric effect is also a large part of the substituent effect. The systems we want to choose for investigation should have minimal steric effect but dominating electronic effect. We ended up with two chemical problems including the acidity of *para*-substituted benzoic acids and the regioselectivity in Diels-Alder reactions for 2-substituted diene. The reason why we choose *para*-substituted benzoic acids instead of *ortho*-substituted benzoic acids is because we want to maximize the distance between the substituent and the carboxyl group so that no direct interaction between them is guaranteed. In this case, the phenyl ring acts as a mediator between two parts. For the Diels-Alder reaction example, the reason why we did not choose to study the 1-substituted diene is because the substituent connected to the first carbon atom will have a direct interaction with it, which complicates further analysis. The results of these two examples turned out to be good. We discovered that for the pK_a value of a *p*-substituted benzoic acid, it can be expressed as a quadratic function of two selected local mode frequencies. For the Diels-Alder reaction example, the substituents can be correctly classified into Electron Withdrawing Groups (EWG) and Electron Donating Groups (EDG). Besides, the experimental data can be explained accordingly.

As a summary for this section, we have discovered that the local vibrational modes can be used for a new field which is the (indirect) substituent electronic effect.

The benzene paper and the mutation path paper can be found in Appendices C and D.

2.5. Generalized Subsystem Vibrational Analysis

Normal vibrational modes are always delocalized over the whole molecular system, although the local vibrational modes can characterize the vibration in terms of a specific geometry unit, the normal vibrational modes for a molecular fragment or sub-system is still missing.

For this question, there is also a practical need, namely to calculate the vibrations of the QM region in a QM/MM system. And my first idea was to use $3N-K$ local force constants of the QM region in order to recover corresponding Hessian matrix for this QM part.

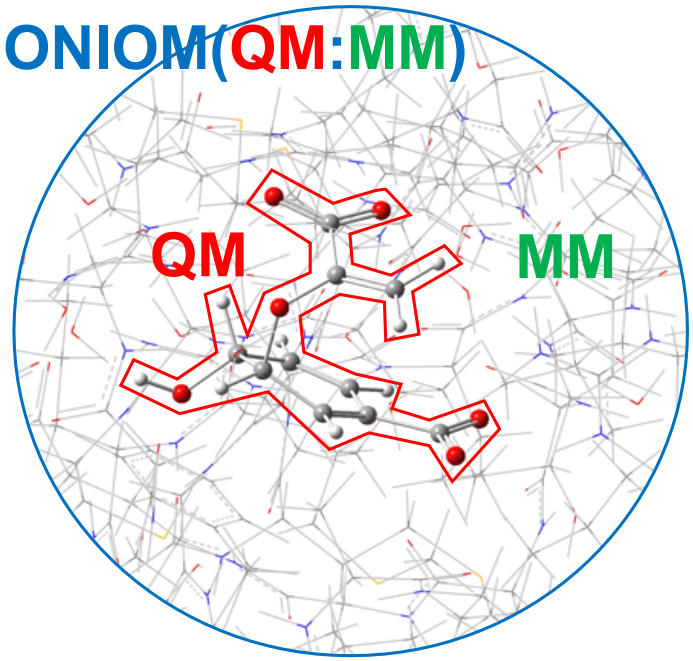


Figure 2.4. Schematic representation of a QM/MM system.

In order to test the feasibility of this approach, we started to test small molecules. There are two ways to convert the local force constants into a Hessian matrix. The first solution is to construct a harmonic approximation of total energy using the quadratic function $E_{q_n} = 0.5 \times k^a \times \Delta q_n^2$. After adding up all contributions from $3N-K$ internal coordinates, one

then calculates the second-order derivations of the total energy $\sum E_{q_n}$ to obtain the Hessian matrix. The second approach is to perform a matrix multiplication with Wilson B matrix and its transpose as the following:

$$\mathbf{H} = \mathbf{b}^T k^a \mathbf{b} \tag{2.1}$$

Then we tried to compare the normal mode frequencies calculated from this recovered Hessian and the original Hessian. For the water molecule, the deviation was small. However, for the ethyne molecule, the deviation was large. So this means the idea of recovering the Hessian matrix from $3N - K$ local force constants might not generally work.

While we were stuck at this point, I tried to look into the simplified equation of the local force constant k_n^a.

$$\frac{1}{k_n^a} = \mathbf{b}(\mathbf{f}^x)^+ \mathbf{b}^T \tag{2.2}$$

As we can see in the above equation that the vector $\mathbf{b}$ in the dimension of $1 \times 3N$ is the Wilson B matrix for a specific internal coordinate. What would happen if we replace $\mathbf{b}$ with $\mathbf{B}$ which is the Wilson B matrix for many internal coordinates in the dimension of $(3N - K) \times 3N$?

We can do the substitution, however, we need to note that $(\mathbf{f}^x)^+$ is the Moore-Penrose inverse of the Hessian matrix $\mathbf{f}^x$ in the dimension of $3N \times 3N$. The Moore-Penrose inverse is defined as a generalized inverse which satisfies the Penrose conditions. The physical nature of $\mathbf{B}(\mathbf{f}^x)^+\mathbf{B}^T$ is similar to the inverse of Hessian, however, the size of this matrix product is changed into $(3N-K)\times(3N-K)$. I tried to get the inverse of this result as $(\mathbf{B}(\mathbf{f}^x)^+\mathbf{B}^T)^{-1}$. When I tried to change the dimension of this inverse, something interesting resulted.

$$\mathbf{f}^x = \mathbf{B}^T(\mathbf{B}(\mathbf{f}^x)^+\mathbf{B}^T)^{-1}\mathbf{B} \tag{2.3}$$

This equation holds as long as the set of internal coordinates characterized by matrix $\mathbf{B}$ is a non-redundant set that can determine the geometry.

Then I tried to set matrix $\mathbf{B}$ into matrix $\mathbf{B}'$ that corresponds to a set of internal coordinates describing the geometry of a sub-system, e.g. one water monomer in a dimer structure. Matrix $\mathbf{B}'$ has the dimension of $(3n-k)\times 3N$. This means $\mathbf{B}'$ has fewer rows than matrix $\mathbf{B}$. The result of $(\mathbf{B}'(\mathbf{f}^x)^+\mathbf{B}'^T)^{-1}$ has the dimension of $(3n-k)\times(3n-k)$, however, we need to transform it back to $3n\times 3n$. So we introduce a new Wilson B matrix $\mathbf{B}'_{sub}$, which is the first $3n$ columns of $\mathbf{B}'$. Thus $\mathbf{B}'_{sub}$ has the dimension of $(3n-k)\times 3n$. Then a new Hessian matrix results as $\mathbf{f}^x_{sub}$.

$$\mathbf{f}^x_{sub} = \mathbf{B}'^T_{sub}(\mathbf{B}'(\mathbf{f}^x)^+\mathbf{B}'^T)^{-1}\mathbf{B}'_{sub} \tag{2.4}$$

We named this new Hessian $\mathbf{f}^x_{sub}$ the *effective Hessian matrix* for the sub-system in question. One special property of this matrix is that if we do the local mode analysis based on this matrix, we get exactly the same result as we do local mode analysis based on the full Hessian matrix.

This property is non-trivial and it means that we have derived a unique Hessian matrix for the sub-system as a true counterpart for the full Hessian for the complete system. On this basis, we would have a chance to get the normal vibrations of the sub-system based on $\mathbf{f}^x_{sub}$. Another interesting fact of $\mathbf{f}^x_{sub}$ is that it has exactly k eigenvalues as zero, where k is the total number of translations and rotations for the sub-system. We named the resulting normal vibrations for the sub-system as *intrinsic fragmental vibrations*. These intrinsic fragmental vibrations have the advantage of being comparable among different molecular systems. For example, the intrinsic fragmental vibrations of a water monomer in a dimer structure is comparable with another water monomer in a trimer structure.

We named the procedure to obtain the intrinsic fragmental vibrations of a sub-system or fragment as *Generalized Subsystem Vibrational Analysis (GSVA)*.

If we go back to the original motivation in order to get the vibrations of the QM region, as long as we can obtain the effective Hessian matrix for the QM region, this problem can be also solved by GSVA.

This work is an important link between the normal vibrational modes and the local

vibrational modes. While normal vibrational modes are always delocalized over the whole molecule, the local modes characterize the vibrations initiated by single internal coordinates. Can we have something in between which is the normal modes localized to a certain fragment or subsystem? Now we have GSVA and its intrinsic fragmental vibrations as the unique answer.

The manuscript on this part can be found in Appendix E.

2.6. Future direction - Probes for Vibrational Stark Effect

The concept of the *Vibrational Stark Effect (VSE)* was firstly raised by Arun Chattopadhyay and Steven G. Boxer in 1995 [1]. This experimental phenomenon consists of two parts. First, certain normal vibrations are special and they are localized to certain chemical bonds, e.g. C=O double bond and C≡N triple bond. Molecules containing these bonds are called the *probes*. Second, the frequencies of above vibrations are linearly correlated to the strength of an electric field which is aligned to those bonds. In this way, a direct mapping from the electric field strength to the vibrational frequency can be measured. Boxer and his co-workers made one step forward, they tried to think of different non-covalent interactions in terms of electric field [5].

While the second aspect of VSE is easy to understand, there has been no theoretical verification for the localized nature of the vibrations related to the stretching of C=O and C≡N bonds. Therefore, one has to verify the localized nature of these normal vibrations first and then find an explanation. On this basis, the rational design of other VSE probes could become possible.

The method to verify the localized nature of a specific normal vibration is to decompose this normal vibrational mode into a non-redundant set of local vibrational modes including one local mode describing the C=O or C≡N stretching [16]. If this normal mode is made up of 100% local vibrational mode of the targeted bond stretching, one can be sure that this normal vibrational mode is genuinely localized to this bond. Such a decomposition analysis can be carried out for different probe molecules with various non-covalent interactions.

Before delving into this problem, some guideline can be given. The explanation of the localized nature of these probe bonds can be pushed into two directions. Firstly, we need to check whether the atomic mass is playing a role. As we can see in the periodic table that Carbon, Nitrogen and Oxygen are three consecutive atoms with close atomic masses as 12.0, 14.0 and 16.0 AMU. We can intentionally set the atomic masses into different values in order to reveal the corresponding influence. Secondly, it would be necessary to check the role of the electronic structure of these atoms. It would be interesting to check probe candidates where the C, N and O atoms are replaced with Si, P and S atoms. In order to eliminate the atomic mass effect, we can set their atomic masses as those of their analogues in period 2. In this way, we would be able to understand the origin of the localized vibrations for these probe bonds.

Chapter 3

UNIFIED REACTION VALLEY APPROACH

3.1. History

The chemical reaction is a topic with utmost importance in chemistry. Chemical reactions are playing an omnipresent role in our daily life. On one hand, almost all man-made products we use are made from chemical reactions. On the other hand, our human body is a machine where complex enzymatic reactions are constantly happening.

To understand chemical reactions can be regarded as the cornerstone to the world of chemistry. Within the Born-Oppenheimer approximation, a chemical reaction can be modeled with the help of a *potential energy surface (PES)*. As is shown in Figure 3.1, the reactant and product located at two separated local minima are connected via a lowest energy path passing the transition state point.

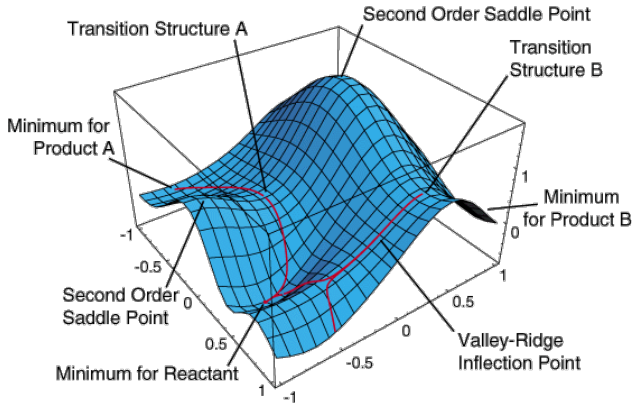


Figure 3.1. Schematic presentation of a potential energy surface. This figure is from Dr. H. Bernhard Schlegel's group website at http://chem.wayne.edu/schlegel/PES.htm.

The first calculation of a potential energy surface based on semi-empirical methods was conducted by Henry Eyring and Michael Polanyi in 1931 for the H + H_2 reaction system.

However, no clear definition of a reaction path was given at that time. This situation was not changed until Kenichi Fukui proposed the *Intrinsic Reaction Coordinate (IRC)* path in 1981 [7].

In 1980, William H. Miller, Nicholas C. Handy and John E. Adams proposed the concept of *Reaction Path Hamiltonian (RPH)* in order to study the dynamical properties of a reaction path [27]. However, their work was before the IRC path and only focused on the H + H_2 model system. In 1988, Michael Page and James W. McIver, Jr. proposed the reaction path Hamiltonian for the IRC path and derived the formula for the reaction path curvature and related coupling terms [30]. The reaction path curvature can be regarded as one of the most properties of a path, as a peak in the curvature plot is often associated with a chemical event, e.g. bond forming/breaking.

Based upon the above work, Zoran Konkoli, Elfi Kraka and Dieter Cremer proposed the *Unified Reaction Valley Approach (URVA)* in 1997 [18]. One of the highlights of URVA is the adiabatic mode coupling coefficient which is decomposition of reaction path curvature into adiabatic local modes (see Figure 3.2).

In 2016, Zou and Cremer proposed a new way to decompose the reaction path direction and curvature in terms of internal coordinates [44], due to the limitation of adiabatic mode coupling on an *unstable* reaction path. This new extension enables URVA to analyze chemical reactions in large systems including organometallic compounds and enzymes.

3.2. Summary of My Contributions

My work on the Unified Reaction Valley Approach consists of three aspects.

1. Programming of a standalone pURVA program in Python;
2. Adding URVA-related modules to the Gaussian 16 code;
3. Correction of the reaction path curvature in the transition state (TS) region.

This dissertation will elaborate these projects in the following sections.

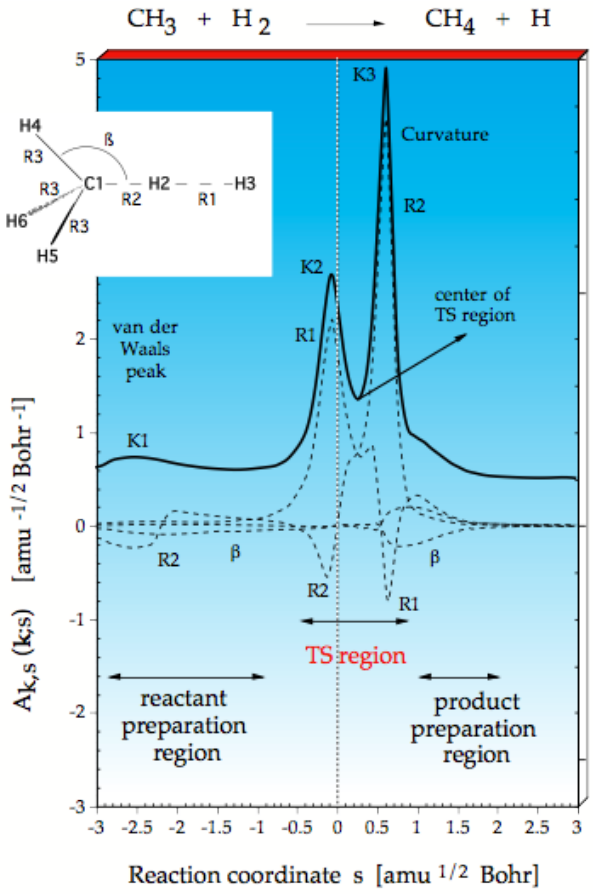


Figure 3.2. Decomposition of the reaction path curvature $\kappa(s)$ (thick solid line) in terms of adiabatic mode-curvature coupling amplitudes $A_{k,s}(s)$ (dashed lines). This figure is from CATCO group website at https://sites.smu.edu/dedman/catco/research/rxn-mechanism-dynamics.html.

3.3. Standalone pURVA Program

The initial implementation of the URVA method was done by Zoran Konkoli, one of Dr. Cremer's students. Konkoli programmed URVA as an add-on module into the link L716[1] of the Gaussian 98 package. The advantage of doing so was that the reaction path calculated by Gaussian could be directly analyzed and many useful libraries within Gaussian could be fully utilized. This led to the problem that whenever the Gaussian program was updated, the URVA module needed to be migrated to the new version. In addition, the URVA analysis was not possible within the framework of other quantum chemistry packages.

Therefore, a standalone URVA analysis program was desired. I started to write such a program by reading through the original codes by Konkoli and re-writing it in the Fortran 90 language. After I finished the program that basically worked, I was not quite satisfied about the organization of the code and decided to write a Python version. Benefiting from various advantages of Python, the new code is more concise, robust and easy to maintain. We named the new code pURVA, where the prefix *p-* stands for Python.

We need to note that the new URVA code is not just a duplicate of the original code by Konkoli, it has the following improvements:

- The Diabatic Mode Ordering (DMO) procedure is reformulated based on Wenli's DMO module in the local mode program. The current implementation of DMO is independent of the internal coordinates specified by user and has more flexibility;
- It supports all possible internal coordinates including curvilinear coordinates or symmetry coordinates for analysis;
- It includes the functionality of correcting the curvature near the transition state region;
- It supports both the old and new (for large systems) browsing files as input data;
- All URVA results are saved as separate pre-formatted text files;
- Functions written in Python code are re-usable by other programs or for other purposes.

[1] L716: A subroutine in Gaussian Processing information for optimizations and frequencies, see http://gaussian.com/capabilities/?tabid=3.

The user manual of pURVA program can be found in Appendix F. The complete source code of pURVA program can be found in Appendix K.

3.4. URVA-Related Module in Gaussian 16

The URVA analysis code is not incorporated into the Gaussian source codes, however, the functionality of generating the input data file for URVA analysis was already implemented into Gaussian by Elfi Kraka and Hrant Hratchian. This input data file is called *browsing file*.

The function of generating IRC browsing file can be activated by using IOp option[2] - IOp(1/45). The IRC browsing file contains Cartesian coordinates, atomic mass, gradients and Hessian for each point along the IRC reaction path, which are needed for the URVA analysis.

In the cases of large reaction systems, the storage of the Hessian is no longer necessary because the latest URVA analysis [44] is based directly on decomposition of reaction path direction and curvature. As long as the browsing file contains the reaction path direction and curvature, the large reaction systems can be analyzed by URVA. In this spirit, a different form of browsing file has been proposed, which is called the new browsing file.

In a new browsing file, the gradient and Hessian are no longer stored, instead the reaction path direction and curvature are calculated on the fly and saved into the browsing file.

While I was working on the Hessian of ONIOM model in IRC calculations, I migrated the code related to new browsing file from G09 to G16. At the same time, I made this part of code conform to the PGI fortran 77 standard as it was written in the Intel fortran standard before. The reason why I did this adjustment is because the whole Gaussian package is written according to the PGI fortran 77 standard and inserting code of Intel fortran standard requires the compilation of the whole package with Intel compiler. In our own experiences, the Gaussian program compiled with Intel compiler may lead to error in some cases. It is safer to use the recommended coding standard and the compiler as PGI.

[2]See http://gaussian.com/overlay1/.

A flowchart that explains how the modification is carried out has been given in Appendix G. Appendix H gives the complete record of code modification on Gaussian 16.

3.5. Reaction Path Curvature Correction near Transition State

The IRC reaction path always starts from the transition state point where the potential energy surface at that point has one direction which is the local maximum but for all other directions, it has local minima. In this way, the IRC path moves downhill in the direction of the local maxima leading to two channels, the reactant channel ends at the reactant(s) and the product channel ending at the product(s).

After the information is collected along the IRC path in the browsing file, the URVA analysis is carried out. The reaction path is a curve in the $3N$-dimensional space with direction and curvature. However, there is a problem concerning the calculation of reaction path curvature. In the original papers related to URVA and URVA codes [18], we used the following formula to calculate the reaction path curvature along the reaction path.

$$\boldsymbol{\nu}^{(1)} = [\mathbf{F}\boldsymbol{\nu} - (\boldsymbol{\nu}^{\dagger}\mathbf{F}\boldsymbol{\nu})\boldsymbol{\nu}]/c \tag{3.1}$$

where $\boldsymbol{\nu}$ is the direction of the reaction path, $\mathbf{F}$ is the Hessian matrix and c is the norm of the mass-weighted gradient. Formula 3.1 works fine, except for the region around the TS point, where the gradient is zero or close to zero, namely $c \approx 0$. If we still use formula 3.1 to calculate reaction path curvature, the result is undefined because the denominator is close to zero. In previous work on URVA, this problem was corrected by applying smoothing procedures to the curvature plot.

Page and McIver provided a specific formula to calculate the reaction path curvature at the transition state point [30]. The formula writes in the following.

$$\boldsymbol{\nu}^{(1)} = (2\boldsymbol{\nu}^{\dagger}\mathbf{F}\boldsymbol{\nu}\mathbf{I} - \mathbf{F})^{-1}(\mathbf{F}^{(1)}\boldsymbol{\nu} - \boldsymbol{\nu}^{\dagger}\mathbf{F}^{(1)}\boldsymbol{\nu}\boldsymbol{\nu}) \tag{3.2}$$

where $\mathbf{I}$ is a unit matrix and $\mathbf{F}^{(1)}$ is the third derivatives along the path tangent. It can be

calculated by finite difference procedure as

$$\mathbf{F}^{(1)} = \frac{d\mathbf{F}}{ds} \approx [\mathbf{F}(s_0 + \delta s) - \mathbf{F}(s_0 - \delta s)]/(2\delta s) \tag{3.3}$$

where s is the reaction path parameter.

However, this formula is only valid at the TS point. If we go either direction off this point, error and deviations will be introduced in.

In order to fully utilize the formula of curvature for the TS point and extend it to the region close to the TS point, we designed a new algorithm to solve this problem.

We use Formula 3.2 to calculate the correct curvature at the TS point.

Then we use the same formula but multiplied with a scaling factor C in the form of an exponential function.

$$C = e^{a \cdot s} \tag{3.4}$$

where s is the reaction path parameter and a is a parameter to be determined.

The reason why we choose the natural exponential function is because it provides scaling down $(x < 0, 0 < y < 1)$ and scaling up $(x > 0, y > 1)$ as shown in Figure 3.3.

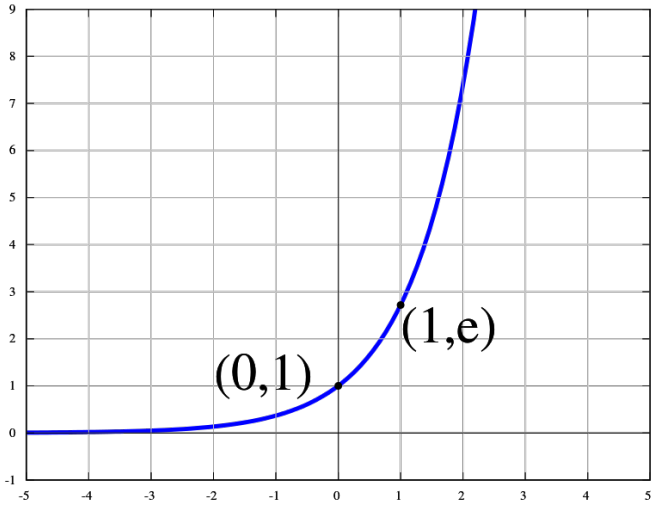


Figure 3.3. The natural exponential function $y = e^x$.

With the scaling, the curvature multiplied with C should connect to the uncorrected curvature value calculated by Formula 3.1. However, the connection should be as smooth as possible. This is realized by optimizing the parameter a in Formula 3.4 to make sure the second derivative of the connection point is smaller than a threshold.

If this threshold cannot be achieved, we need to move one point in the direction from the TS point, until the scaled corrected curvature curve is smoothly merged into the curvature calculated by Formula 3.1.

When the scaled curvature curve is successfully merged into the uncorrected curvature calculated by Formula 3.1. The last step is to remove the curvature for the TS point and two points next to it, as there exists an implicit problem that the IRC path following algorithm is unable to accurately trace the first point off the TS point. So we use a cubic spline fitting to get interpolated curvature values for these three points.

The result of above algorithm applied to a gold-catalyzed reaction is shown in Figure 3.4.

Now this algorithm for calculating correct curvature near the TS point has been fully automatized and proved to be robust. It has been incorporated into the pURVA program.

The complete source code on this work has been given in Appendices I and J.

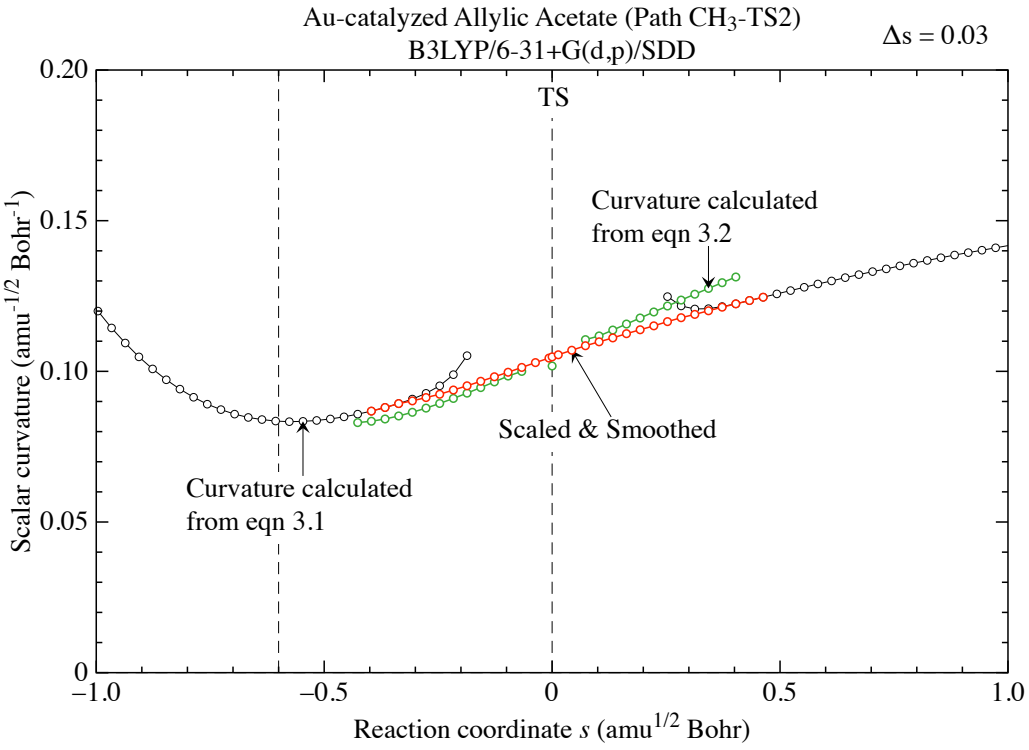


Figure 3.4. Reaction path curvature and its corrected value. Credits to Dr. Marek Freindorf for providing test examples.

Chapter 4

CONCLUSIONS AND OUTLOOK

From years of working with the two major theoretical pillars in the CATCO group including the Local Vibrational Mode Theory and the Unified Reaction Valley Approach (URVA), I have deeper understanding on the molecular vibration, potential energy surface and reaction path.

On one hand, URVA can be a really useful tool to characterize a chemical reaction path, and it reveals every detail of bond breaking/formation along the path. However, just by presenting the URVA result is insufficient for chemists to gain deeper understanding. One direction that can push URVA to a higher level could be systematic studies for a series of similar chemical reactions and try to find correlations between the reaction path curvature and physical properties, e.g. activation energy. We expect that more attempts can be done on this part to present the practical applications of URVA, which will be the foundation for the publication of the standalone pURVA program.

For the Local Vibrational Mode Theory, it has been extensively applied to describe chemical bonding. One potential issue in this connection lies in the comparison of chemical bonds with different bonding atoms, e.g. O-H vs. S-H. One should be always very careful when comparing the local force constants in above situation even though the scaled Bond Strength Order (BSO) is empolyed to make two different bond types "comparable".

We have successfully extend the application scope of Local Vibrational Mode Theory into the problem of chemical similarity or substituent (electronic) effect, however, more chemical problems are expected to be found where the Local Vibrational Mode Thoery can play a role.

It is challenging to find out other chemical problems besides chemical bonding and similarity for applying Local Vibrational Mode Theory, but it is worth trying in all possible

directions.

Concerning the underlying theories of Local Vibrational Mode Theory, we have done almost everything we can do with regard to molecular vibrations. One possible direction could be to extend the Local Vibrational Mode Theory into periodic systems, where one needs to have solid knowledge about the phonon spectroscopy.

BIBLIOGRAPHY

[1] Chattopadhyay, A., and Boxer, S. G. Vibrational Stark Effect Spectroscopy. *J. Am. Chem. Soc. 117*, 4 (1995), 1449–1450. 16

[2] Cremer, D., Wu, A., Larsson, J. A., and Kraka, E. Some Thoughts About Bond Energies, Bond Lengths, and Force Constants. *J. Mol. Model. 6*, 4 (2000), 396–412. 3

[3] Decius, J. C. Compliance Matrix and Molecular Vibrations. *J. Chem. Phys. 38*, 1 (1963), 241–248. 4

[4] Freindorf, M., Kraka, E., and Cremer, D. A Comprehensive Analysis of Hydrogen Bond Interactions Based on Local Vibrational Modes. *Int. J. Quantum Chem. 112*, 19 (2012), 3174–3187. 4

[5] Fried, S. D., and Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Acc. Chem. Res. 48*, 4 (2015), 998–1006. 16

[6] Fu, R., Lu, T., and Chen, F. Comparing Methods for Predicting the Reactive Site of Electrophilic Substitution. *Acta Phys. Chim. Sin. 30*, 4 (2014), 628–639. 11

[7] Fukui, K. The Path of Chemical Reactions - The IRC Approach. *Acc. Chem. Res. 14*, 12 (1981), 363–368. 19

[8] Gordon, M. S., Freitag, M. A., Bandyopadhyay, P., Jensen, J. H., Kairys, V., and Stevens, W. J. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. *J. Phys. Chem. A 105*, 2 (2001), 293–307. 5

[9] Kalescky, R., Kraka, E., and Cremer, D. Identification of the Strongest Bonds in Chemistry. *J. Phys. Chem. A 117*, 36 (2013), 8981–8995. 4

[10] Kalescky, R., Kraka, E., and Cremer, D. Local Vibrational Modes of the Formic Acid Dimer–the Strength of the Double Hydrogen Bond. *Mol. Phys. 111*, 9-11 (2013), 1497–1510. 4

[11] Kalescky, R., Kraka, E., and Cremer, D. New Approach to Tolman's Electronic Parameter Based on Local Vibrational Modes. *Inorg. Chem. 53*, 1 (2013), 478–495. 4

[12] Kalescky, R., Zou, W., Kraka, E., and Cremer, D. Local Vibrational Modes of the Water Dimer–Comparison of Theory and Experiment. *Chem. Phys. Lett. 554* (2012), 243–247. 4

[13] Kalescky, R., Zou, W., Kraka, E., and Cremer, D. Quantitative Assessment of the Multiplicity of Carbon–Halogen Bonds: Carbenium and Halonium Ions with F, Cl, Br, and I. *J. Phys. Chem. A 118*, 10 (2014), 1948–1963. 4

[14] Kitaura, K., Ikeo, E., Asada, T., Nakano, T., and Uebayasi, M. Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules. *Chem. Phys. Lett. 313*, 3-4 (1999), 701–706. 5

[15] Konkoli, Z., and Cremer, D. A New Way of Analyzing Vibrational Spectra. I. Derivation of Adiabatic Internal Modes. *Int. J. Quantum Chem. 67*, 1 (1998), 1–9. 3

[16] Konkoli, Z., and Cremer, D. A New Way of Analyzing Vibrational Spectra. III. Characterization of Normal Vibrational Modes in terms of Internal Vibrational Modes. *Int. J. Quantum Chem. 67*, 1 (1998), 29–40. 3, 16

[17] Konkoli, Z., Cremer, D., and Kraka, E. Diabatic Ordering of Vibrational Normal Modes in Reaction Valley Studies. *J. Comp. Chem. 18*, 10 (1997), 1282–1294. 10

[18] Konkoli, Z., Kraka, E., and Cremer, D. Unified Reaction Valley Approach Mechanism of the Reaction CH3 + H2 → CH4 + H. *J. Phys. Chem. A 101*, 9 (1997), 1742–1757. 19, 23

[19] Konkoli, Z., Larsson, J. A., and Cremer, D. A New Way of Analyzing Vibrational Spectra. II. Comparison of Internal Mode Frequencies. *Int. J. Quantum Chem. 67*, 1 (1998), 11–27. 3

[20] Konkoli, Z., Larsson, J. A., and Cremer, D. A New Way of Analyzing Vibrational Spectra. IV. Application and Testing of Adiabatic Modes Within the Concept of the Characterization of Normal Modes. *Int. J. Quantum Chem. 67*, 1 (1998), 41–55. 3

[21] Kraka, E., Larsson, J. A., and Cremer, D. *Generalization of the Badger Rule Based on the Use of Adiabatic Vibrational Modes*. Wiley-Blackwell, 2010, ch. 4, pp. 105–149. 3

[22] Kraka, E., Setiawan, D., and Cremer, D. Re-evaluation of the Bond Length–Bond Strength Rule: The Stronger Bond Is not Always the Shorter Bond. *J. Comp. Chem. 37*, 1 (2016), 130–142. 4

[23] Larsson, J. A., and Cremer, D. Theoretical Verification and Extension of the McKean Relationship Between Bond Lengths and Stretching Frequencies. *J. Mol. Struct. 485* (1999), 385–407. 3

[24] Li, S., Li, W., and Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Acc. Chem. Res. 47*, 9 (2014), 2712–2720. 5

[25] Liu, S. Where Does the Electron Go? the Nature of Ortho/Para and Meta Group Directing in Electrophilic Aromatic Substitution. *J. Chem. Phys. 141*, 19 (2014), 194109. 11

[26] Mahoney, M. W., and Jorgensen, W. L. A Five-Site Model For Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys. 112*, 20 (2000), 8910–8922. 5

[27] Miller, W. H., Handy, N. C., and Adams, J. E. Reaction Path Hamiltonian for Polyatomic Molecules. *J. Chem. Phys. 72*, 1 (1980), 99–112. 19

[28] Ochterski, J. *White Paper: Vibrational Analysis in Gaussian*. Gaussian, 1999. 9

[29] Oliveira, V., Kraka, E., and Cremer, D. Quantitative Assessment of Halogen Bonding Utilizing Vibrational Spectroscopy. *Inorg. Chem. 56*, 1 (2016), 488–502. 4

[30] Page, M., and McIver, J. W. On Evaluating the Reaction Path Hamiltonian. *J. Chem. Phys. 88*, 2 (1988), 922–935. 19, 23

[31] Ranganathan, S., Izotov, D., Kraka, E., and Cremer, D. Description and Recognition of Regular and Distorted Secondary Structures in Proteins using the Automated Protein Structure Analysis Method. *Proteins 76*, 2 (2009), 418–438. 1

[32] Setiawan, D., Kalescky, R., Kraka, E., and Cremer, D. Direct Measure of Metal–Ligand Bonding Replacing the Tolman Electronic Parameter. *Inorg. Chem. 55*, 5 (2016), 2332–2344. 4

[33] Setiawan, D., Kraka, E., and Cremer, D. Strength of the Pnicogen Bond in Complexes Involving Group Va Elements N, P, and As. *J. Phys. Chem. A 119*, 9 (2014), 1642–1656. 4

[34] Tainter, C. J., Ni, Y., Shi, L., and Skinner, J. L. Hydrogen Bonding and OH-Stretch Spectroscopy in Water: Hexamer (Cage), Liquid Surface, Liquid, and Ice. *J. Phys. Chem. Lett. 4*, 1 (2012), 12–17. 6

[35] Tao, Y., Zou, W., Cremer, D., and Kraka, E. Characterizing Chemical Similarity With Vibrational Spectroscopy: New Insights Into the Substituent Effects in Mono-Substituted Benzenes. *J. Phys. Chem. A 121*, 42 (2017), 8086–8096. 4

[36] Tao, Y., Zou, W., Cremer, D., and Kraka, E. Correlating the Vibrational Spectra of Structurally Related Molecules: A Spectroscopic Measure of Similarity. *J. Comput. Chem. 39*, 6 (2018), 293–306. 4

[37] Tao, Y., Zou, W., Jia, J., Li, W., and Cremer, D. Different Ways of Hydrogen Bonding in Water-Why Does Warm Water Freeze Faster than Cold Water? *J. Chem. Theory Comp. 13*, 1 (2016), 55–76. 4

[38] Tao, Y., Zou, W., and Kraka, E. Strengthening of Hydrogen Bonding with the Push-Pull Effect. *Chem. Phys. Lett.* (2017). 4

[39] Wu, A., and Cremer, D. Correlation of the Vibrational Spectra of Isotopomers: Theory and Application. *J. Phys. Chem. A 107*, 48 (2003), 10272–10279. 9

[40] Zhang, X., Dai, H., Yan, H., Zou, W., and Cremer, D. B-H$\cdots\pi$ Interaction: A New Type of Nonclassical Hydrogen Bonding. *J. Am. Chem. Soc. 138*, 13 (2016), 4334–4337. 4

[41] Zou, W., and Cremer, D. Properties of Local Vibrational Modes: The Infrared Intensity. *Theor. Chem. Acc. 133*, 3 (2014). 4

[42] Zou, W., and Cremer, D. C2 in a Box: Determining Its Intrinsic Bond Strength for the $x^1\sigma_g^+$ Ground State. *Chem. Eur. J. 22*, 12 (2016), 4087–4099. 4

[43] Zou, W., Kalescky, R., Kraka, E., and Cremer, D. Relating Normal Vibrational Modes to Local Vibrational Modes with the Help of an Adiabatic Connection Scheme. *J. Chem. Phys. 137*, 8 (2012), 084114. 3

[44] Zou, W., Sexton, T., Kraka, E., Freindorf, M., and Cremer, D. A New Method for Describing the Mechanism of a Chemical Reaction Based on the Unified Reaction Valley Approach. *J. Chem. Theory Comp. 12*, 2 (2016), 650–663. 19, 22

[45] Zou, W., Zhang, X., Dai, H., Yan, H., Cremer, D., and Kraka, E. Description of An Unusual Hydrogen Bond Between Carborane and A Phenyl Group. *J. Organomet. Chem.* (2018). 4

[46] Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys. 22*, 8 (1954), 1420–1426. 9

Appendix A

Paper on Water Hydrogen Bonding

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Different Ways of Hydrogen Bonding in Water - Why Does Warm Water Freeze Faster than Cold Water?

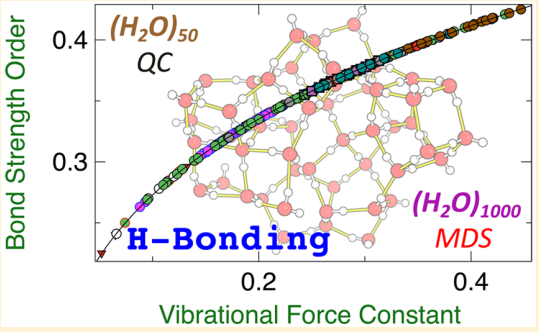
Yunwen Tao,[†] Wenli Zou,[†] Junteng Jia,[‡] Wei Li,[‡] and Dieter Cremer*[,†]

[†]Computational and Theoretical Chemistry Group (CATCO), Department of Chemistry, Southern Methodist University, 3215 Daniel Avenue, Dallas, Texas 75275-0314, United States

[‡]Institute of Theoretical and Computational Chemistry, Key Laboratory of Mesoscopic Chemistry of MOE, School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, P. R. China

Supporting Information

ABSTRACT: The properties of liquid water are intimately related to the H-bond network among the individual water molecules. Utilizing vibrational spectroscopy and modeling water with DFT-optimized water clusters (6-mers and 50-mers), 16 out of a possible 36 different types of H-bonds are identified and ordered according to their intrinsic strength. The strongest H-bonds are obtained as a result of a concerted push−pull effect of four peripheral water molecules, which polarize the electron density in a way that supports charge transfer and partial covalent character of the targeted H-bond. For water molecules with tetra- and pentacoordinated O atoms, H-bonding is often associated with a geometrically unfavorable positioning of the acceptor lone pair and donor $\sigma^*(OH)$ orbitals so that electrostatic rather than covalent interactions increasingly dominate H-bonding. There is a striking linear dependence between the intrinsic strength of H-bonding as measured by the local H-bond stretching force constant and the delocalization energy associated with charge transfer. Molecular dynamics simulations for 1000-mers reveal that with increasing temperature weak, preferentially electrostatic H-bonds are broken, whereas the number of strong H-bonds increases. An explanation for the question why warm water freezes faster than cold water is given on a molecular basis.



INTRODUCTION

The understanding of hydrogen bonding (H-bonding) is essential for unravelling many biological and environmental phenomena.[1−6] H-bonding dominates the noncovalent interactions between the molecules in liquid water, and in this way H-bonding is ultimately responsible for the unique properties of water. Essential for the understanding of the complex structure and dynamics of liquid water[7] is the study of H-bonding with the help of quantum chemical methods. If an atomistic approach is used, liquid water can be modeled by using clusters of water molecules. The smallest of such clusters, the water dimer, is only used for reference purposes, and its properties in connection with H-bonding are fairly well-known.[8−20] Also, larger clusters with three to six water molecules have been reliably described and have helped to extend the understanding of H-bonding between water molecules.[21−36] Less frequent are high-accuracy investigations of larger water clusters.[37] Most of these investigations have been carried out at the Hartree−Fock (HF), Density Functional Theory (DFT), or perturbation theory level. For example, the investigation of 20-mers (clusters with 20 water molecules),[38−41] 25-mers,[42] 30-mers up to 40-mers,[43−49] or even 60-mers has to be reported.[50]

Noteworthy in this connection is that the vibrational spectra of 20-mers have been investigated in detail by Xantheas and co-workers using second order perturbation theory.[38] DFT benchmark calculations utilizing B3LYP, X3LYP, and M06-type of XC-functionals for predicting binding energies of water clusters up to 20 molecules have been carried out by Bryantsev and co-workers.[39] Parthasarathi and co-workers found that linear chains of up to 20 water molecules lead to dipole moments as high as 41 D thus emphasizing the cooperative effect of H-bonding in larger clusters.[40] An interesting study on the polarizability of water clusters and the charge flow through H-bonds in the presence of internal and external electric fields was carried out by Yang and co-workers.[41] The electron density at the critical points of the H-bonds of a water cluster was analyzed by Neela and co-workers who predicted an increase of the density with the cluster size.[51] Iwata pointed out the importance of charge transfer and dispersion energies in $(H_2O)_{20}$ and $(H_2O)_{25}$, which he found to depend on the O···O distance.[42] Lenz and co-workers[43] calculated the vibrational spectra of water clusters containing up to 30 water molecules. They found a correlation between the red-shift of the O−H donor stretching frequency and the type of H-bond based on the coordination numbers of the O atoms being involved. The importance of the collective electrostatic effects on H-bonding

Received: July 23, 2016
Published: December 6, 2016

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

as caused by the nonimmediate environment in liquid water models was emphasized by Bako and co-workers.[44] Qian and co-workers did systematic studies on water clusters of different size ranging from the dimer to 34-mers using HF/6-31G(d).[47] An attempt was made by Huang and co-workers to predict the far-infrared spectra of water clusters up to 38 molecules with DFT and to relate them to observed THz spectra.[48] Specific forms of water clusters were investigated by various authors (spiro-cyclic,[45] fullerene-shaped[50]). Frogato and co-workers performed ab initio Born−Oppenheimer molecular dynamics (MD) simulations for 69-mer clusters containing an excess electron.[52] Clusters with up to 280 water molecules were investigated by Loboda and co-workers who determined averaged H-bond energies.[53] Turi used mixed quantum-classical MD simulations for a cluster consisting 1,000 water molecules either in neutral state or with an excess electron.[54] A quantum simulation of water was carried out by Wang and co-workers.[55]

In these investigations, the intrinsic strength of the H-bond in water clusters or liquid water could not be determined. Instead one attempted to obtain indirect evidence by analyzing binding energies, H-bond distances, vibrational frequencies, electron densities at the H-bond critical point, and other molecular properties, which can only provide a qualitative measure of the H-bond strength as they relate to all intermolecular forces and interaction energies.

In this work, we present the harmonic vibrational frequencies of water clusters containing 50 molecules (*50-mers*) that can be considered as suitable models for distinguishing between different H-bond types. For the first time, we will provide a detailed account on H-bonding in water clusters, which can be considered as suitable models for liquid water. In connection with this general goal, we pursue the following objectives: (i) We will investigate how many of the 36 possible standard H-bond types (excluding pentacoordination of oxygen and H-bond bifurcation) are needed to analyze H-bonding in the 50-mers. (ii) We will characterize the various H-bonds according to the intrinsic strength of their interactions, which we will characterize with the properties of the H-bond stretching vibrations using the theory of Konkoli and Cremer for analyzing local vibrational modes.[56−59] For this purpose, we will derive a H-bond strength order (BSO) value, which will provide a quantitative measure to compare different H-bonds in the water clusters investigated. (iii) H-bonding results from the noncovalent interactions of a H-bond donor (D) and a H-bond acceptor (A). Accordingly, we will investigate to which extent the properties of D and those of A are varied by H-bonding. Are there relationships between H-bond stretching force constants, covalent and electrostatic bond character, electron and energy density properties at the H-bond critical points, or the H-bond lengths? (iv) Is there a relationship between the strength of the OH donor bond and that of the H-bond, which can be used to characterize the latter via properties of the former?[5,6] (v) Finally, we will make an attempt to relate the structure of a water-cluster to the macroscopic properties of liquid water by utilizing MD simulations of 1000-mers. In this connection, we will investigate the question why warmer water freezes more quickly than colder water.[60−66]

The results of this investigation will be presented in the following order. In Section 2, we will describe the computational methods used in this work. In Section 3, the different H-bonds of the 50-mers will be analyzed, and a suitable way of describing them will be worked out. The results of this analysis will be applied in Section 4 to provide a molecular explanation to the phenomenon that warm water freezes faster than cold water. The chemical relevance and the conclusions of the current investigation will be summarized in the last section.

COMPUTATIONAL METHODS

Equilibrium geometries and normal vibrational modes were calculated using the ωB97X-D density functional,[67,68] which was chosen because it provides a reliable description of noncovalent interactions in cases where dispersion and other long-range van der Waals interactions play an important role.[69−72] Pople's triple-ζ basis 6-311G(d,p) was augmented by diffuse functions for H and O atoms. The 6-311++G(d,p) basis set[73−75] thus obtained contained 1800 basis functions for the 50-mers. The calculations of the normal mode vectors and frequencies were carried out with the GEBF (generalized energy-based fragmentation) method[76,77] at the ωB97X-D/6-311++G(d,p) level. The analytical gradient of the GEBF method[77] was used for the geometry optimization, whereas for the GEBF Hessian an approximate expression was employed.[78] The usefulness of GEBF-ωB97X-D in the case of the water clusters was first tested by carrying out calculations for 20-mers and comparing results obtained with CCSD(T)[79] in the form of GEBF-CCSD(T). GEBF-ωB97X-D turned out to be both reliable and cost efficient.

In the GEBF-ωB97X-D calculations, each water molecule was selected as a fragment, and the distance threshold was set to 4.0 Å, i.e. at least one atom is within this limit. The maximum number of fragments in each subsystem was limited to seven. Natural population analysis (NPA)[80,81] charges were employed as background charges, and two-fragment subsystems with a distance threshold of 8.0 Å were considered for corrections. The DFT calculations were carried out using a pruned (75,302) fine grid[82,83] and tight convergence criteria in the geometry optimizations to guarantee a reliable calculation of vibrational frequencies. The initial geometries of the 50-mers were taken from MD calculations using a TIP4P force field.[84] The optimized geometries are given in the Supporting Information (SI). The relative energies of the complexes used in this investigation are 0.0 kcal/mol (cluster **A**; absolute energy: −3822.655729 hartree), −0.62 (**B**), 10.30 (**C**), and 4.40 kcal/mol (**D**). The lowest vibrational frequencies obtained in this way are 23.7, 26.3 (cluster **A**); 25.2, 29.6 (**B**); 18.0, 21.3 (**C**); 17.9, 23.4 cm^{-1} (**D**). Another water complex leading to an imaginary frequency was excluded from the investigation.

Electron density and energy density distributions were calculated using ωB97X-D rather than GEBF-ωB97X-D. The charge transfer analysis was carried out on the basis of calculated NPA charges.[80,81] A topological analysis of the electron density distribution $\rho(\mathbf{r})$ was performed.[85] The nature of the H-bond was determined by the energy density $H(\mathbf{r})$ calculated at the H-bond critical point $\mathbf{r}_b$ and the application of the Cremer−Kraka criteria for covalent bonding: (i) A zero-flux surface and bond critical point $\mathbf{r}_b$ have to exist between the atoms in question (necessary condition). (ii) The local energy density at $H(\mathbf{r}_b)$ must be negative and thereby stabilizing (sufficient condition for covalent bonding). A positive $H(\mathbf{r}_b)$ indicates a dominance of electrostatic interactions.[86−88] Hence, the Cremer−Kraka criteria can reveal whether H-bonding is covalent, electrostatic, or a mixture of both (values close to zero).

The covalent character of the H-bond was estimated by calculating the delocalization energy $\Delta E(del)$, which is associated with the charge transfer from a lone pair orbital of

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

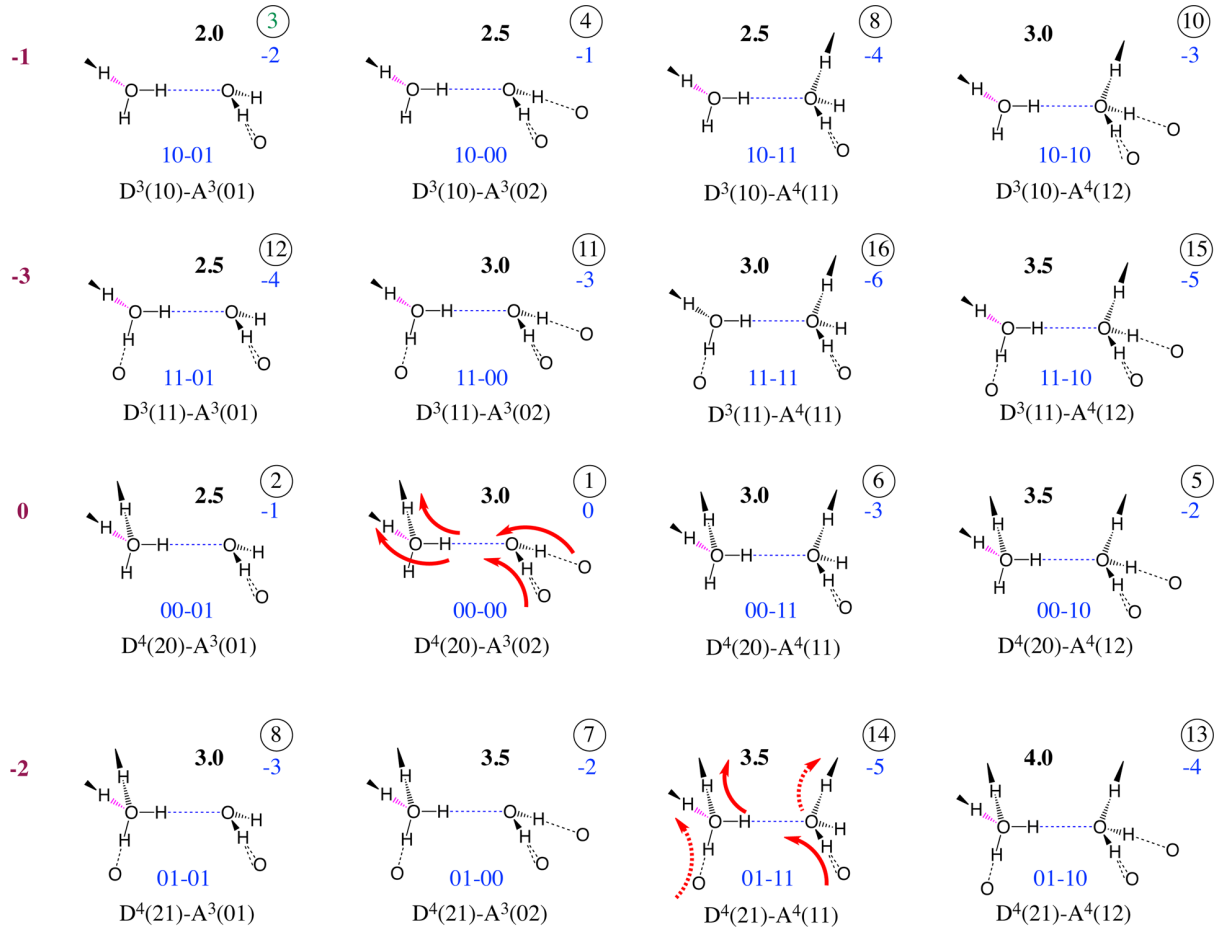


Figure 1. Coding of the 16 H-bond types discussed for the 50-mers by the notation $D^{c_D}(i_a,j_d) - A^{c_A}(k_a,l_d)$. The integers i_a, j_d, k_a, and l_d give the peripheral (external) H-bonds directly embedding the targeted H-bond, i.e. the acceptor (a) and donor (d) H-bonds of D and A, which are ≤2 for the D or A water molecule. The superscripts c_D and c_A are the coordination numbers of O(D) and O(A), respectively, which vary between 3 and 4 and must be distinguished from the number of H-bonds $m(D)$ and $m(A)$. In black bold print the average $m(AD)$ of $m(D)$ and $m(A)$ is given. For the strongest H-bond $D^4(20) - A^3(02)$ (in short: 20-02) red arrows indicate the direction of charge polarization, which supports the covalent character of this H-bond. For other H-bonds, the optimal 20-02 arrangement is perturbed as indicated for the $D^4(21) - A^4(11)$ (in short: 21-11) H-bond by dashed red arrows. For each type of H-bond, the deviation is given in blue as in the case of the 21-12 H-bond: |21-11 - 20-02 | = 01-11. The distortion descriptor (see text) is given as a negative blue number in the upper right of each drawing where an encircled number defines the position of the H-bond in a strength order from 1 (strongest H-bond) to 16 (weakest H-bond) according to the average BSO values of Figure 3. The 10-01 bond value has been added as reference (BSO: n = 0.399; see text). On the left, the distortion relationship between rows of the matrix of H-bond types is given in brown color.

A to an antibonding OH orbital of D thus leading to an increase of the electron density in the range of the H-bond. The magnitude of $\Delta E(del)$ was determined by second order perturbation theory[81] where both $lp(O) \rightarrow \sigma^*(OH)$-contributions to $\Delta E(del)$ for a given $O\cdots OH$-interaction were included (see Section 3).

The intrinsic strength of the H-bond was determined by using the local H-bond stretching force constant.[59,89] The local vibrational modes of Konkoli and Cremer[56] are based on the solution of the Wilson equation of vibrational spectroscopy[90]

$$\mathbf{F}^q\mathbf{D} = \mathbf{G}^{-1}\mathbf{D}\mathbf{\Lambda} \tag{1}$$

in the form[91]

$$(\mathbf{G}_d + \lambda\mathbf{G}_{od})\tilde{\mathbf{R}}_\lambda = (\mathbf{\Gamma}^q_d + \lambda\mathbf{\Gamma}^q_{od})\tilde{\mathbf{R}}_\lambda\mathbf{\Lambda}_\lambda \tag{2}$$

In these two equations, $\mathbf{F}^q$ is the calculated force constant matrix expressed in internal coordinates q_n, $\mathbf{D}$ collects the corresponding vibrational eigenvectors $\mathbf{d}_\mu$ as column vectors (μ = 1, ⋯, N_{vib} with $N_{vib} = 3N - L$, N: number of atoms; L: number of translations and rotations), $\mathbf{G}$ is the Wilson matrix for the kinetic energy,[90] and $\mathbf{\Lambda}$ is a diagonal matrix containing the vibrational eigenvalues $\lambda_\mu = 4\pi^2c^2\omega_\mu^2$. In the expression for the eigenvalues, ω_μ represents the vibrational frequency of mode $\mathbf{d}_\mu$.

Matrix $\mathbf{\Gamma}$ in eq 2 is the inverse force constant matrix, which is usually called compliance matrix.[92] Matrix $\tilde{\mathbf{R}}$ corresponds to $\tilde{\mathbf{D}}$ in the local mass-weighted formulation (indicated by the tilde)[91]

$$\tilde{\mathbf{R}} = \mathbf{F}^q\tilde{\mathbf{D}} \tag{3}$$

and the partitioning is into diagonal (d) and off-diagonal (od) parts. The parameter λ controls kinematic (mass) coupling, i.e. for λ = 0 the local description is obtained and for λ = 1 the Wilson equation reformulated in terms of compliance matrix and $\mathbf{R}$ modes.

Solution of the Wilson equation requires the diagonalization of matrix $\mathbf{F}^q$ to give the matrix $\mathbf{K}$. In this way, the electronic

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

coupling between the local modes is eliminated. Solving the mass-decoupled Wilson equation leads to the mass-decoupled local modes $\mathbf{a}_n$,[91] which can be written as[56,58,93]

$$\mathbf{a}_n = \frac{\mathbf{K}^{-1}\mathbf{d}_n^{\dagger}}{\mathbf{d}_n\mathbf{K}^{-1}\mathbf{d}_n^{\dagger}} \tag{4}$$

where $\mathbf{d}_n$ is now a row vector of matrix $\mathbf{D}$. The local mode force constant k_n^a is given by eq 5

$$k_n^a = \mathbf{a}_n^{\dagger}\mathrm{K}\mathrm{a}_n \tag{5}$$

and the local mode frequency ω_n^a can be obtained from

$$(\omega_n^a)^2 = \frac{G_{nn}k_n^a}{4\pi^2c^2} \tag{6}$$

where element G_{nn} of matrix $\mathbf{G}$ defines the local mode mass.[56]

Before continuing it is useful to point out that the term *local vibrational modes* is sometimes used in a different context: Henry and co-workers[94−96] use the term in connection with the (an)harmonic oscillator models to quantum mechanically calculate the overtones of XH stretching modes. The higher overtone modes (n = 5 or 6) for isolated XH groups are largely decoupled which justifies using the term local modes. Their frequencies correlate linearly with the Konkoli−Cremer local mode frequencies thus verifying their local mode character;[97] however, they are only accessible for a few types of XH stretching modes, whereas the Konkoli−Cremer modes are generally defined and will be exclusively used in the following.

The relative bond strength order (BSO) n of an OH bond is obtained by utilizing the extended Badger rule,[97−99] according to which the BSO is related to the local stretching force constant k_a by a power relationship, which is fully determined by two reference values and the requirement that for a zero-force constant the BSO value becomes zero. Accordingly, the relationship $n(OH) = a(k_a)^b$ can be derived from two suitable reference bonds. In this work, the constants a and b were determined for FH bonds using the frequencies of F−H (n = 1) and the $D_{\infty h}$-symmetrical $[\mathrm{F}\cdots\mathrm{H}\cdots\mathrm{F}]^-$ anion (n = 0.5) as suitable references. Since the Badger and extended Badger rules predict for related XH bonds the same power relationship, the equation $n(FH) = a(k_a)^b$ with a = 0.5402 and b = 0.2966 was also used for the OH bonds after shifting the single bond reference (corresponding to a n(OH) = 0.9653) by 0.0347 so that the OH bonds of H_2O obtain the BSO value n = 1.00.

In the MD simulations, 1000 water molecules in a PBC (periodic boundary condition) box were simulated by classical MD using the TIP5P[100] force field in the NPT ensemble at 1.0 bar and 283, 308, 363, and 378 K where the simulations at 308 and 378 K were used as control calculations and therefore will not be discussed here in detail. The cutoff of nonbonded interactions was set to be 8 Å, and the Coulombic interactions were treated with the Ewald summation.[101] The temperature was scaled by Langevin dynamics with the collision frequency γ being 1.0.[102] The Berendsen bath coupling method[103] was selected as a thermostat algorithm to control the pressure. The equations of the motion were integrated by the velocity Verlet algorithm[104] with OH bond constrains.[105,106] The time step was set to 1 fs (femtosecond), and the trajectories were collected for every 100 fs. The simulation time was 2 ns (nanoseconds) where the first ns was used for reaching the equilibrium. The trajectories of the second ns were used for the analysis. From the snapshots of the second ns, 1000 $(H_2O)_{1000}$ periodic clusters were taken out using equal time intervals of 1 ps (picosecond).

For the analyses, a H-bond was considered to be given if the H···O distance is between 1.5 and 2.2 Å, and the O−H···O angle is larger than 100.0°. PBC were also employed for determining H-bonds. In the analysis of the $(H_2O)_{50}$ clusters, the 2.2 Å limit turned out to be a reasonable cutoff value. There are 6 × 6 = 36 standard H-bond types (excluding bifurcated H-bonds and pentacoordination at the O atom, i.e. the maximum number of H-bonds per water molecule is limited to 4). If pentacoordination is included another 36 types of H-bonds are possible, whereas bifurcation adds another 71 H-bond types (maximum number of H-bonds ≤5) thus leading to a total of 11 × 13 = 143 different H-bond types. For the bifurcated H-bonds, the H atom of the O−H donor bond is within 2.4 Å with regard to two neighboring O atoms.[107] In the equilibrium structures of the 50-mers, only a fraction of the maximally possible standard H-bonds can occur. These are shown in Figure 1 where the H-bond of the cyclic water hexamer, $(H_2O)_6$, is used as a suitable reference (see the SI for the equilibrium geometry). The average number of H-bonds per water molecule was obtained by the formula m_{av} = 2 × N(*H bonds*)/N(*water molecules*) (N: number). In addition, we use the quantities $m(D)$ and $m(A)$, which give the total number of H-bonds of the D- and the A-water for a given type of H-bond (Figure 1). Accordingly, the average $m(AD)$ is equal to $[m(D) + m(A)]/2$. Similarly, the number of peripheral H-bonds can be given by $m_p(D)$, $m_p(A)$, etc. Parameters m and m_p are associated with the dimer D−A and do not consider peripheral H_2O molecules (m_{av} varies only slightly from 1.5 to 1.75 for the cases considered and therefore is less useful for the characterization of the types of H-bonds shown in Figure 1).

Beside calculating the charge transfer between the interacting monomers using NPA values, we also calculated the difference density distribution $\Delta\rho(\mathbf{r}) = \rho(Complex,\mathbf{r}) - \rho(Monomer1,\mathbf{r}) - \rho(Monomer2,\mathbf{r})$, which was determined and plotted for the complex-enveloping surface of an electron density distribution of 0.001 e/Bohr^3. For the situation of six water molecules as in 20-02 (see below), the geometry was taken from one of the 20-02 H-bonds of cluster **A**. The polarization effect caused by the four peripheral H_2O was determined by subtracting from the hexamer density that of the trimers on the D and A side as well as that of the central D−A dimer and then adding the density of the D and A water monomer (short notation: 6-2 × 3-2 + 1 + 1) all calculated in the geometry of the hexamer. In this way, the density contributions of the trimers and the difference density of the central dimer were eliminated so that the *push−pull* effect of the peripheral water molecules on the central dimer unit becomes visible.

For the statistical analysis, we used box-and-whisker diagrams,[108] which present the distribution of data by a box and two whiskers. Minimum and maximum values are indicated by two horizontal lines. The first and third quartile of data, Q1 and Q3, define the bottom and top of the box where Q2 gives the position of the median. The interquartile range QR = Q3 − Q1 gives the vertical length of the box. The length of the whiskers is defined by Q1 − 1.5 QR (lower whisker) and Q3 + 1.5 QR (upper whisker). Any data point, which is lower than $Q1 - q \times QR$ and higher than $Q3 + q \times QR$, is considered a mild outlier for q = 1.5 (black dots) and an extreme outlier for q = 3.0 (open dots).[108] We have applied this analysis when at least more than 7 data points were available.

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

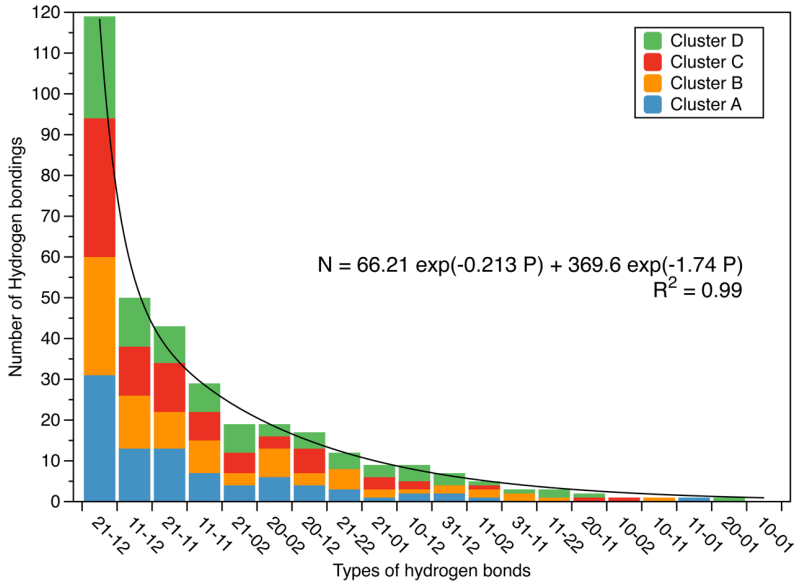


Figure 2. Distribution of different H-bonds in the four 50-mers **A** (blue), **B** (orange), **C** (red), and **D** (green). The H-bonds are classified according to Figure 1 and ordered according to their frequency of appearance, which can be approximately described by an exponential dependence of the number N of H-bonds (P: position number). See text.

All vibrational mode and electron density calculations were carried out with the program package COLOGNE2016,[109] whereas for the DFT and the GEBF calculations local versions of the program package Gaussian09[110] were used. All MD simulations were performed with the AmberTools15 package.[111] Difference densities were plotted with the program Multiwfn.[112]

THE H-BONDING NETWORK IN WATER CLUSTERS

The H-bond between a D- and an A-molecule can be characterized by the notation $D^{c_D}(i_a,j_d)-A^{c_A}(k_a,l_d)$ where integers i_a, j_d, k_a, and l_d give the peripheral (external) H-bonds directly embedding the targeted H-bond, i.e. the acceptor (a) and donor (d) H-bonds of D and A, which for the D or A water molecule are normally equal to or smaller than 2. The superscripts c_D and c_A are the coordination numbers of O(D) and O(A), respectively, which vary between 3 and 4 in the case of the 50-mers. Hence, $D^4(20) - A^3(02)$ (or in short (20-02)) denotes the H-bond with a D water molecule functioning as an acceptor for 2 external (=peripheral) H-bonds (its O atom coordinates with 4 H via normal or H-bonding; c_D = 4) and an A water molecule that itself is functioning as a double donor for 2 other external H-bonds (c_A = 3). This situation is sketched in the third row, second column of Figure 1. In the equilibrium geometry of the 50-mers, each D-molecule and each A-molecule are found to undergo one of the four possible interactions: 1) 10; 2) 11; 3) 20; 4) 21, which leads to a total of 16 = 4^2 different H-bond interactions such as 10-01, 10-02, 10-11, 10-12, ··· 21-12, where we have used as a reference the relatively strong 10-01 H-bond of cyclic $(H_2O)_6$. The largest number of peripheral H-bonds is realized for the H-bond $D^4(21) - A^4(12)$, which has for both the D and A molecule 2 + 2 = 4 H-bonds and by this $(i_a + j_d) + (k_a + l_d)$ = 3 + 3 = 6 peripheral H-bonds influencing the targeted D−A H-bond. Figure 1 contains additional information, which will be discussed below. In passing, we note that in liquid water, contrary to the equilibrium situation of a 50-mer, other relative weak H-bonds or no H-bonds at D and/or A are observed so that for each water molecule 6 rather than 4 different H-bond arrangements become possible leading to a total of 6^2 = 36 H-bond types (see Figure S1 in the SI). The complete set of H-bond types will be discussed in Section 4. Finally, it has to be mentioned that similar notations on how to characterize different types of H-bonding in water clusters have been used in the literature (see Figure S1 of the SI for a relationship between these and the current notation).[42,113−115]

The numbering of atoms for the almost spherical structures of the four 50-mers is given in Figures S2−S5 of the SI. It is obvious that the water molecules located at the outside of the sphere have a smaller number of H-bonds than those water molecules positioned closer to the center of the 50-mer. In the four 50-mers **A**, **B**, **C**, and **D**, a total of 350 H-bonds (87,88,88,87) are found, which account for 15 of the 16 types of H-bonds shown in Figure 1. The 10-01 H-bond is topologically only possible in isolated cyclic water clusters as in $(H_2O)_6$ that we have used as a reference. In addition to the H-bond types of Figure 1, there are 20 H-bonds which were not found for the 50-mers but will be discussed in connection with the MD simulations. In Table S1 of the SI, all H-bonds identified for clusters **A**, **B**, **C**, and **D** are characterized by the $R(H\cdots O)$ distance, the local stretching force constant $k^a(H\cdots O)$, the associated local stretching frequency $\omega^a(H\cdots O)$, the BSO value $n(H\cdots O)$, and the $O-H\cdots O$ angle α.

As one can see from the H-bond data (Table S1 in the SI, Figure 2), different types of H-bonding differ significantly in their number, where however their distribution is similar in the four water clusters. The distribution of the various types of H-bonding is nearly the same in the four 50-mers, which might be a result of their spherical form. The average number of H-bonds per water molecule, m_{av}, is close to 3.5 in all cases (Table S1 in the SI).

The most common H-bond is that of the 21-12 type (number of peripheral H-bonds: $m_p(D)$ = 3; $m_p(A)$ = 3; in short (3,3)), which accounts for 34.0% of all H-bonds followed by 11-12 ((2, 3); 14.3%), 21-11 ((3,2), 12.3%), and 11-11 H-bond ((2,2), 8.3%). The H-bond 11-02 is the least common

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

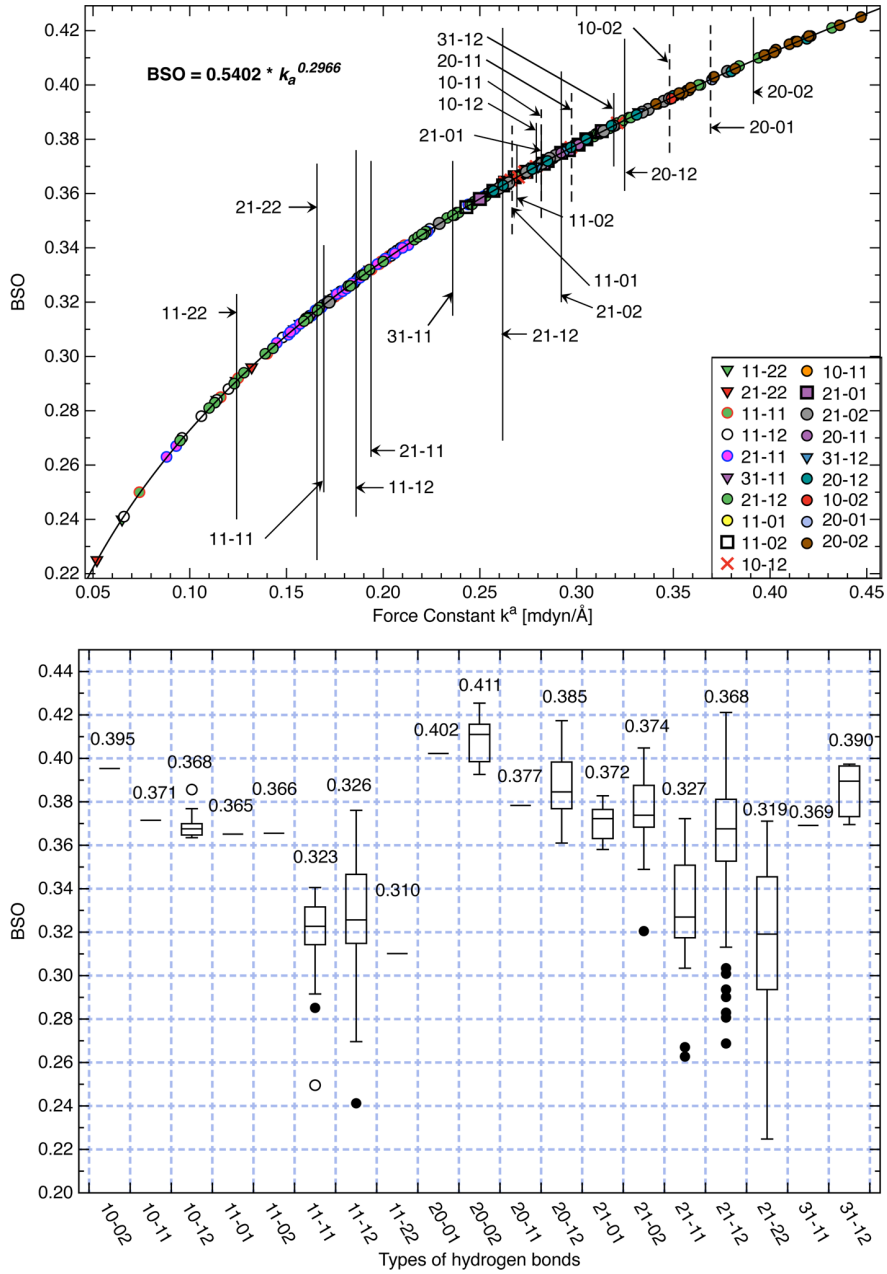


Figure 3. Top: The bond strength order (BSO) n of the H-bond given as a function of the local H-bond stretching force constant $k^a(H\cdots O)$. H-bonds are color-coded according to the notation of Figure 1. The average BSO value of each H-bond type is given by a vertical line where the length of the line indicates the range of BSO values found. Dashed vertical lines indicate those types of H-bonds for which only a few examples are observed. Bottom: The box-and-whisker diagram gives the statistical distribution of the BSO values for different types of H-bonds. For the definition of outliers (dots and circles), see Section 2.

type among the water molecules ((2,2) peripheral H-bonds) and accounts for just 1.4%. Other H-bonds such as 10-02 (0.3%), 10-11 (0.3%), 10-12 (2.6%), 11-01 (0.3%), 11-02 (1.4%), 11-22 (0.9%), 20-01 (0.3%), 20-11 (0.7%), 21-01 (2.6%), and 31-11 (1.0%) can only be found in one of the 50-mers. In summary, there is an exponential decay of the statistical occurrence of specific H-bonds with the number and position of peripheral H-bonds (see below), which holds if one considers all four 50-mers together (Figure 2).

Figure 3 provides an insight into which extent different H-bonds can be distinguished. The intrinsic strength of the H-bonds in the 50-mers as reflected by the BSO values varies by 50% from 0.225 to 0.425 (for comparison: 10-01 as in cyclic $(H_2O)_6$: $n = 0.399$). For each type of H-bond, the average BSO value is given by the crossing point of the curve $n(k^a)$ (top of Figure 3) and a vertical line, which defines the range of BSO values n for this particular type of H-bond by its length. In some cases, a definition of the range of n-values becomes meaningless because of the small number of H-bonds found for a particular type. Then, a dashed line of an arbitrary length of 0.04 BSO units is used. The representation at the top of Figure 3 is complemented by a statistical analysis of the different types of H-bonds among the 350 observed, which is given in the form

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

of a box-and-whisker diagram (bottom of Figure 3; for an explanation, see Section 2).

The most stable H-bonds correspond to the 20-02 type (median: 0.411; average: 0.409; there is a linear relationship between median and average n values; R^2 = 0.96; see the SI). They are followed by the 20-01 H-bonds (0.402), 10-02 (0.395), and 20-12 H-bonds (0.385; blue dots in Figure 3). The green dots of the 21-12 H-bonds, which one could expect as the strongest H-bonds (6 peripheral H-bonds) are quite frequent but belong to the weaker H-bonds because of their large variation from 0.269 to 0.421 with many of these bonds being in the low strength range as a result of geometrical constraints. If the H-bonds are ordered according to the sequence given in Figure 1, a decline of the intrinsic H-bond strength from 10-02 to 11-12, an incline to 20-01 and 20-02, and another decline to 21-11 is found.

A H-bond turns out to be the strongest, which has just two external H-bonds on the D and two on the A side but which seems to guarantee both strong electrostatic and/or covalent interactions. This is is illustrated in Figure 1 for the 20-02 H-bond: Four peripheral water molecules polarize the electron density from A to D as is revealed by the difference density distribution $\Delta\rho(\mathbf{r})$ calculated for the 20-02 hexamer in Figure 4.

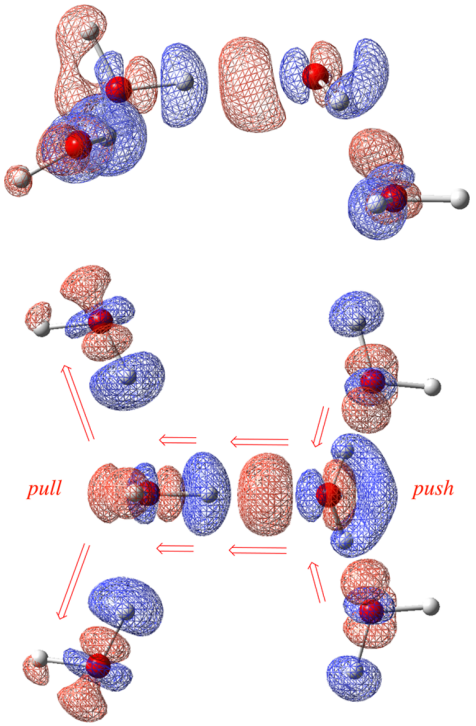


Figure 4. Diagrams show the difference density distribution $\Delta\rho(\mathbf{r})$ calculated for the 0.001 e/Bohr^3 density surface of the hexamer defining the 20-02 type of H-bonding (see Section 2, for details). Blue contour lines indicate a depletion, red an increase of the electron density distribution because of polarization. Top: side view. Bottom: bird view.

This is defined in the way that the extra-effect of the four peripheral H_2O molecules on the targeted 20-02 H-bond becomes visible. The two peripheral H_2O at the A side polarize with their lp electrons the O−H bonds of A thus leading to an increase in the density close to O1(A) and in the 20-02 H-bond region (*push-effect*, Figure 4). The two peripheral H_2O on the D-side support this effect by pulling density from the O2(D) lp orbitals toward their OH bonds thus helping to increase the polarization of the OH bond of the D molecule (*pull-effect*, Figure 4). Hence, the polarization of the electron density distribution is in the direction of the red arrows shown for the 20-02 H-bond in Figures 1 and 4.

At this point, it is appropriate to differentiate between physically based observables and the model quantities used in this work. Local mode frequencies and their associated force constants can be in principle measured.[91] This is also true for the electron density, whereas NPA charges are model quantities connected with a special orbital model.[81] Chemists explain H-bonding in terms of covalent, exchange, electrostatic, inductive, and dispersion interactions (see, e.g., Wang and co-workers[116]). Recently, Politzer and co-workers[117−119] pointed out that according to the Hellmann−Feynman theorem[120] noncovalent interactions are purely Coulombic in nature, and, accordingly, H-bonding might be described in this way. Although this is a valid view, often quantities such as NPA charges, charge transfer values, or charge delocalization energies provide a more detailed, model-based description of H-bonding. In this work, we will use the latter to describe covalent interactions. Difference densities can, as a result of their construction, reflect polarization effects although contributions from the changes in exchange repulsion, dispersion, etc. can also play a role. Apart from this, we will use an energy density based model that distinguishes just between covalent and electrostatic forces.[86,87,121]

Covalent versus Electrostatic H-Bonds. The covalent contribution of the H-bond can be considered as being dominated by a charge transfer from the lp(O1) orbital(s) to the $\sigma^*(O2-H)$ orbital of the D water (see bottom of Figure 5). The overlap between these orbitals will be maximal provided they are suitably oriented in line with an O2HO1 angle α close to 180° and an approach distance between (O2)H and O1 that is smaller than the sum of the van der Waals radii (1.2 + 1.52 Å[122]). Covalent contributions caused by charge transfer lead to H-bond stabilization as is reflected by the increase of the delocalization energy $\Delta E(del)$ with the BSO value (Figure 5, top), a more negative energy density at the H-bond critical point $\mathbf{r}_b$, and an accumulation of the electron density at this point.[86,87,121]

In the 50-mers investigated, the delocalization energies $\Delta E(del)$ vary from 8 to 32 kcal/mol, whereas the corresponding force constants vary from 0.29 to 0.41 mdyn/Å. It is striking that the $\Delta E(del)$ values fall into two groups (with the exception of that of type 10-11), which nicely correlate with the average BSO values (R^2 = 1.00 and 0.97; Figure 5): The stronger H-bonds are presented by the upper line, which seems to combine those H-bond types with $i_a(D) + l_d(A) > j_d(D) + k_a(A)$ due to $i_a = l_d = 2$ and $j_d + k_a \leq 1$. Those H-bond types, which do not fulfill these criteria, are represented by the dashed lower line in Figure 5 (top).

It seems that the strength of the various H-bond types is strongly influenced by the delocalization energy $\Delta E(del)$ where of course this is only valid for the 350 H-bonds of the four 50-mers in their equilibrium geometries and the NPA approach used. One of the referees mentioned that for certain geometries two lp(O) of the same O can contribute to $\Delta E(del)$ of one H-bond. The analysis applied in this work revealed that for 79 out of 350 H-bonds a second $lp(O) \rightarrow \sigma^*(OH)$ contribution larger than 3.0 kcal/mol (about one tenth of the first delocalization energy) was encountered, i.e. in 22.5% of all H-bonds, whereas

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

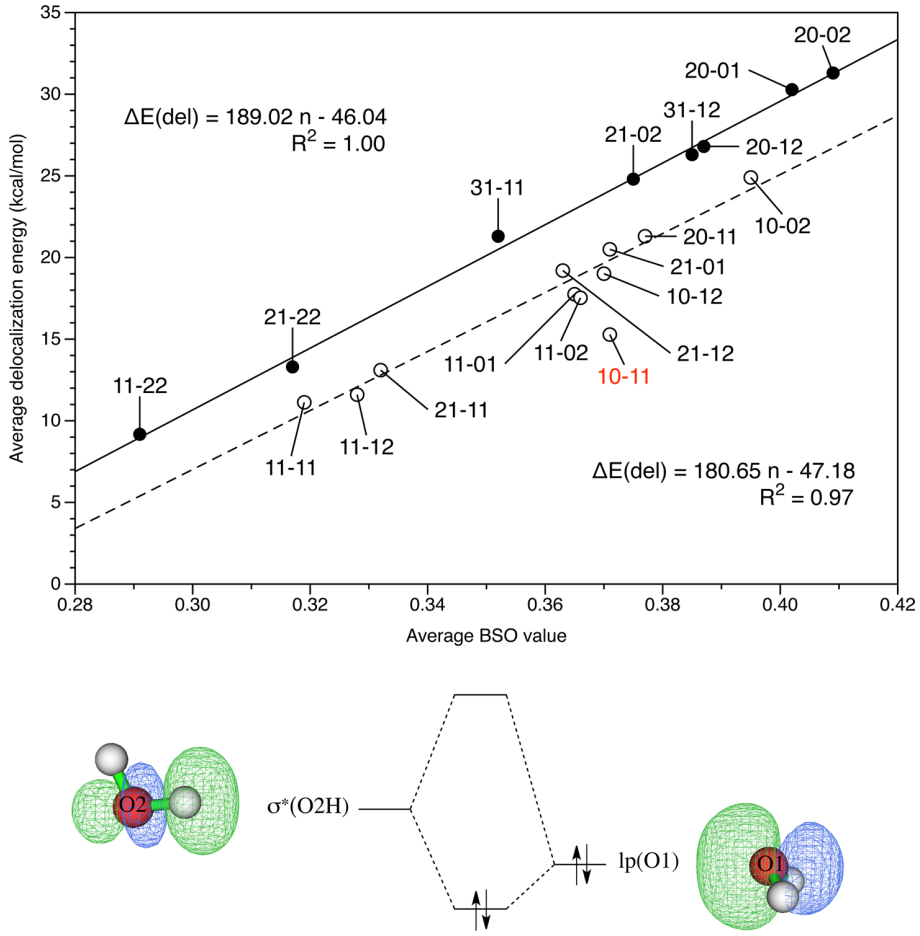


Figure 5. Top: The average delocalization energy $\Delta E(del) = \Delta E(lp(O) \rightarrow \sigma^*(OH))$ is plotted as a function of the average BSO n of the H-bond. For two different classes of H-bonds two correlation lines are obtained (not included: the 10-11 H-bond type). Bottom: The covalent contribution to the H-bond implies a delocalization of the lone pair (lp) electrons of atom O1 (green lobe of the orbital on the right) into the $\sigma^*(OH)$ antibonding orbital of the OH donor (green lobe of the orbital on the left). The energy stabilization caused by the charge transfer was determined by second order perturbation theory and is the basis for the delocalization energies of the upper diagram.

for 90% of all H-bonds the second lp contribution is between 0.05 and 3 kcal/mol. The perturbation theory analysis based on the NPA model also revealed that $\Delta E(del)$ contributions involving other than lp-orbitals are smaller than 0.1 kcal/mol, and therefore it is reasonable to use $\Delta E(del)$ for the description of the covalent contributions to the H-bond.

An increase of the average number of H-bonds, $m(AD)$, from 2 to 4 (Figure 1) does not necessarily weaken the H-bond but leads to a larger variation in the n-values because of stronger OH···O bending, the geometric limitations in the overlap between the lp(O1) and the σ^*(O2H) orbitals, a less than optimal charge transfer, and lower covalent contributions to H-bonding, which might be only partly compensated by electrostatic contributions (see 21-12 in Figures 3, top, and 6, top).

In Figure 6 (top), the geometrical conditions for H-bonding are compared. The box-and-whisker diagram gives the nonlinearity of the H-bond unit O2H···O1 as measured by the angle α. The analysis reveals that the stronger H-bonds (20-02, 10-02, etc.) are more linear in agreement with the requirements for maximal overlap and charge transfer. The 21-12 H-bonds have the largest variation in α (almost 40°, Figure 6). Noteworthy is that the median values are in the range from 164 to 173° irrespective of the H-bond considered. Since the nonlinearity of the H-bond arrangement is closely related to the covalent or electrostatic character of the H-bond, we show at the bottom of Figure 6 the statistical analysis of the energy density at the H-bond critical point, which should be negative for a dominant covalent bond according to the Cremer−Kraka criteria.[86−88] This is qualitatively confirmed by the diagram, which gives an "inverted" distribution of box-and-whisker units as compared to the diagrams in Figure 3. Accordingly, the 20-02 H-bond has the most negative energy density values, whereas the 21-12 H-bonds have at the same time the most positive energy density values and the largest variation of values.

Variation in the Strength of a H-Bond. The explanation why the 20-02 H-bond is the strongest one has been based on Figures 1 and 4 (polarization of the density as indicated by the red arrows). Any deviation from this optimal arrangement leads to a weakening of the H-bond. Using the average BSO values of Figure 3, the following ordering according to decreasing H-bond strength results: 20-02 > 10-01 > 20-01 > 10-02 > 20-12 > 20-11 > 21-02 > 10-11 ≈ 21-01 > 10-12 > 11-02 > 11-01 > 21-11 > 21-12 > 11-12 > 11-11 (For the cases, with a maximum of five H-bonds per water molecule: 31-12 > 31-11 > 21-22 > 11-22, see Figure 3). This ordering can qualitatively be reproduced if one considers that any competition of the

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

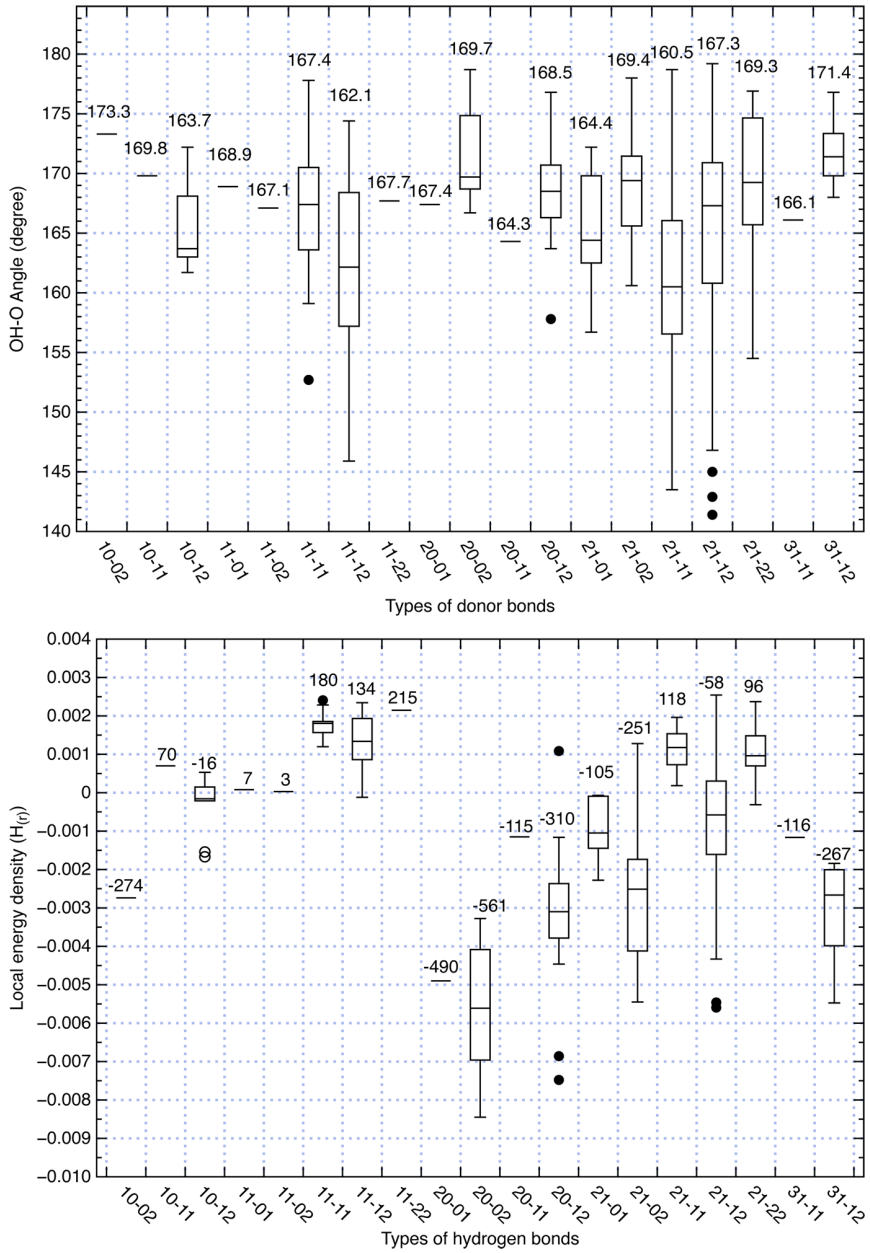


Figure 6. Top: The nonlinearity of the H-bond arrangement as measured by the angle α = O2H···O1 is analyzed in the form of a box-and-whisker diagram. For details of the box-and-whisker diagrams, see Section 2. Bottom: The box-and-whisker diagram gives the statistical distribution of the energy densities (given in Hartree/Bohr^3) at the H-bond critical point for different types of H-bonds. Median values given at the top of the upper whisker are multiplied by 10^5.

targeted H-bond with the second donor H-bond (as in 21-02, Figure 1) or with another external H-bond for the lp-density of the O(A) atom is a first order perturbation, whereas a change in the other peripheral H-bonds either on the D or A side can be considered as a second order perturbation (where perturbations on the D-side seem to have a somewhat larger effect than those on the A-side). Weighting first and second order perturbations qualitatively by −2 and −1 and using the short notation for the perturbation in the form $|(i_a, j_d, k_a, l_d) - (20\text{-}02)| = (p_i, p_j, p_k, p_l)$ (p: perturbation), the negative blue numbers in Figure 1 are obtained (at the upper right of each H-bond arrangement).

The qualitative comparison of a 21-11 H-bond with the 20-02 reference leads to a perturbation 01-11, which implies a weakening of −5 (two first order and one second order perturbation: $-2 \times 2 - 1 \times 1 = -5$) and thereby characterizes one of the weakest H-bonds (#14; in Figure 1. The ranking of each H-bond in terms of its average BSO is given by an encircled number: 1 gives the strongest and 16 the weakest H-bond; see also Figure 3). In this way, the ordering of most of the 16 H-bond types is correctly predicted (exceptions are 20-11 and 10-11; the 10-01 value is added in position 3 using the BSO of the cyclic hexamer $(H_2O)_6$). For example, the 10-02 H-bond has a perturbation value of 10-00, i.e. only one of the outer H-bonds is missing thus yielding a second order weakening of the targeted D−A H-bond of −1 and position 3 in the list of strong H-bonds.

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

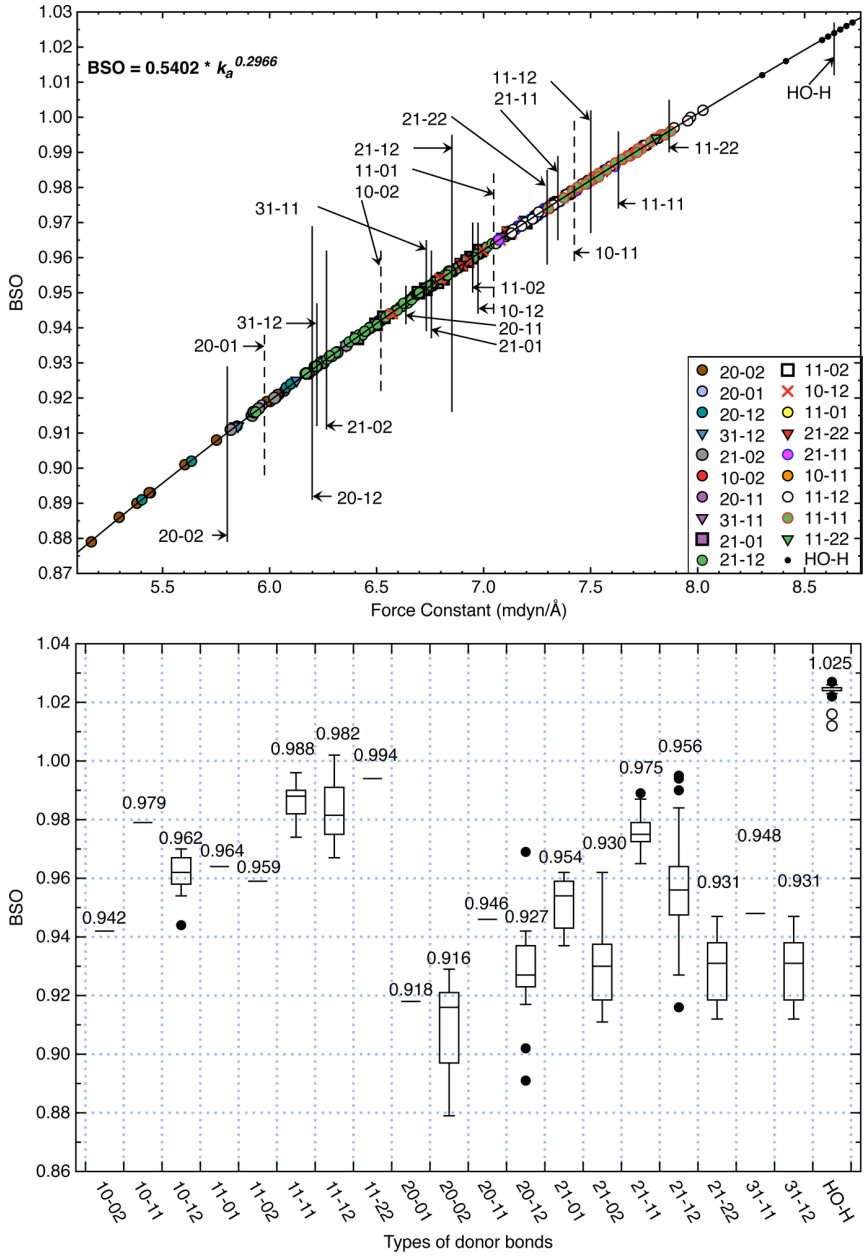


Figure 7. Top: The bond strength order (BSO) n of the O−H donor bond given as a function of the local O−H stretching force constant. Since each of these O−H-bonds is associated with a specific H-bond the former are identified and color-coded according to the associated H-bond (Figure 3). Bottom: A box-and-whisker diagram gives the statistical distribution of the BSO values of the various types of OH donor bonds leading to H-bonding in the four 50-mers investigated. For details of the box-and-whisker diagrams, see Section 2.

Utilizing the perturbation indices given for each type of H-bond, a distortion relationship between the rows of Figure 1 can be obtained (purple numbers on the left of Figure 1). According to these values (row 1: −1; row 2: −3; row 3:0; row 4: −2), H-bonds 20-rs (r = 0, 1; s = 1, 2) are the most stable ones, followed by those of row 1, whereas the least stable ones are found in row 2. This ordering is directly related to the possible perturbations of the 20-02 arrangement, which reduce its push−pull effect and thereby its covalent character.

The ordering of H-bonds by perturbation indices reveals that the polarization of the electron density of D and A as caused by the peripheral H_2O molecules (Figure 4) is an important electronic effect for the intrinsic strength of the H-bond. Polarization causes the energy of the σ^*(O2H) orbital being lowered and that of the lp(O1) being raised thus effectively decreasing the energy gap between these orbitals and increasing the delocalization energy $\Delta E(del)$. In addition, electrostatic interactions can be maximized by the polarization effect. The analyses summarized in Figures 3 (BSO-values), 5 (delocalization energies $\Delta E(del)$), and 6 (top: angles α; bottom: energy densities $H(\mathbf{r}_b)$) suggest that the covalent contributions to H-bonding are important for equilibrium geometries. This of course can be a consequence of (i) the finite size of the water clusters, (ii) the NPA model being used, and (iii) the exclusion of entropy effects in the current analysis.

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

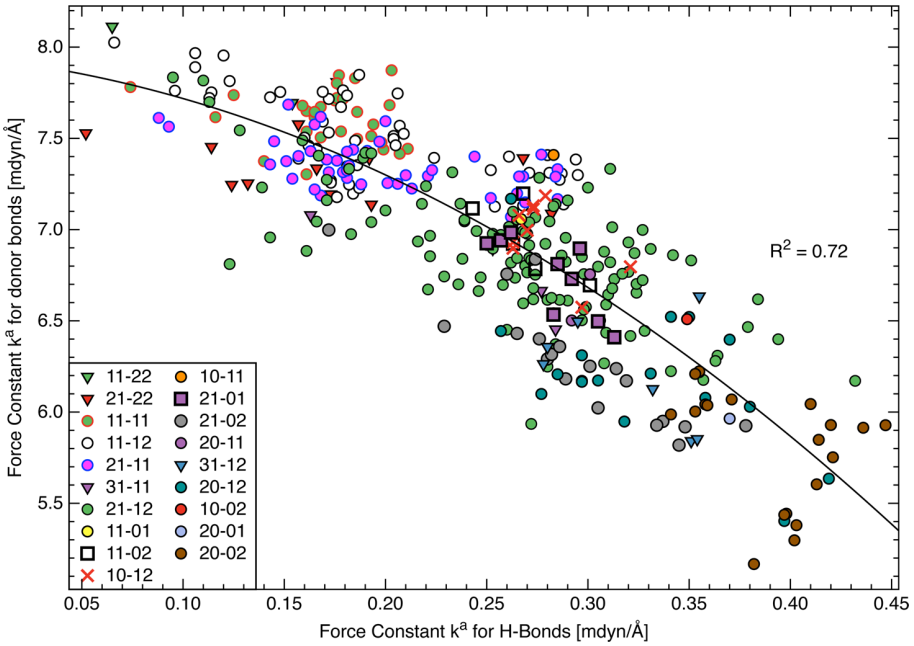


Figure 8. Testing the relationship between D and A: Comparison of the local stretching force constants $k^a(H\cdots O)$ and $k^a(O-H)$. The solid line is given by $k^a(O-H) = 7.938-1.211 \times k^a(H\cdots O) - 9.917 \times (k^a(H\cdots O))^2$ with $R^2 = 0.72$ and $\sigma = 0.30$ mdyn/Å. For the H-bond notation, see Figure 1.

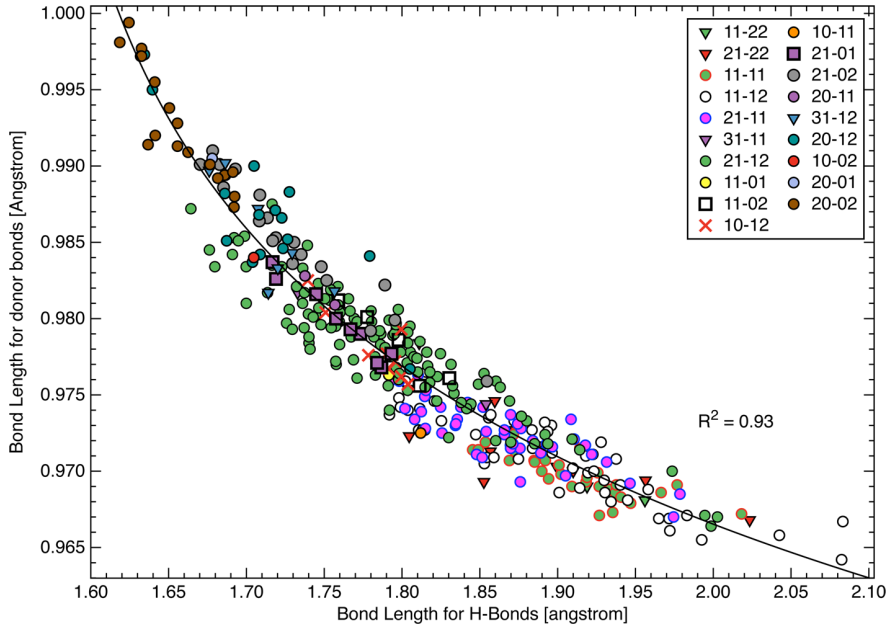


Figure 9. Testing the relationship between D and A: Comparison of the distance $R(H\cdots O)$ and $R(O-H)$. The solid line is given by $R(O-H) = 0.952-0.018 \times log[R(H\cdots O) - 1.552]$ with $R^2 = 0.93$ and $\sigma = 0.018$ Å. For the H-bond notation, see Figure 1.

Donor−Acceptor Relationships. There are numerous investigations, which use vibrational spectroscopy to describe H-bonding[123−136] and relate the strength of the O−H donor bond to the strength of the H-bond.[134,136,137] These investigations are mostly based on infrared spectroscopy as the weakening of the O−H donor bond in the case of H-bonding can be easily recorded by a red-shift of the OH stretching frequency. In a previous investigation, Freindorf and co-workers[59] have shown that the expected relationship between the O−H donor bond and the H-bond is not fulfilled. However, this investigation included many different H-bond donors. The current investigation is limited to D,A interactions between water molecules, and therefore a relationship between the O−H donor bond and the H-bond becomes more likely.

As shown in Figure 7 (top), the BSO values of the O−H donor bonds present in the four 50-mers vary from 0.88 to 1.00 (water molecule without any H-bonding) and beyond this to 1.03 if also those OH bonds are included that are on the outside of the water cluster and therefore not involved in any H-bonding (denoted HO−H in Figure 7). Although the range of BSO values is only one-third of that of the H-bonds, the same number of bonding situations as found for the H-bonds

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

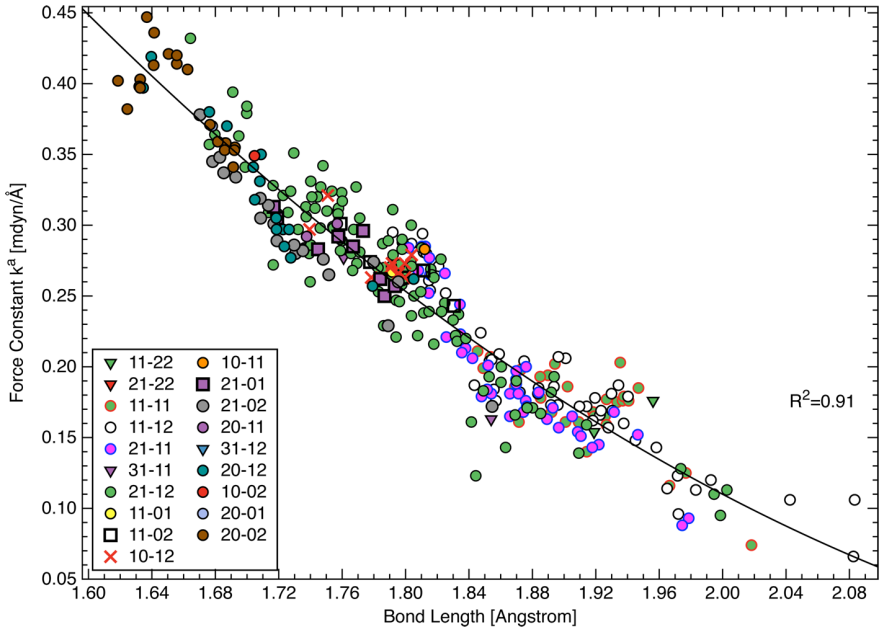


Figure 10. Testing the Badger relationship: Comparison of the local H-bond stretching force constant $k^a(H\cdots O)$ and the corresponding distance $R(H\cdots O)$. For the relationship $k^a = 0.648\ R(O\cdots H)^2 + 3.181\ R(O\cdots H) + 3.878$, $R^2 = 0.91$ and $\sigma = 0.023$ mdyn/Å are calculated. For the H-bond notation, see Figure 1.

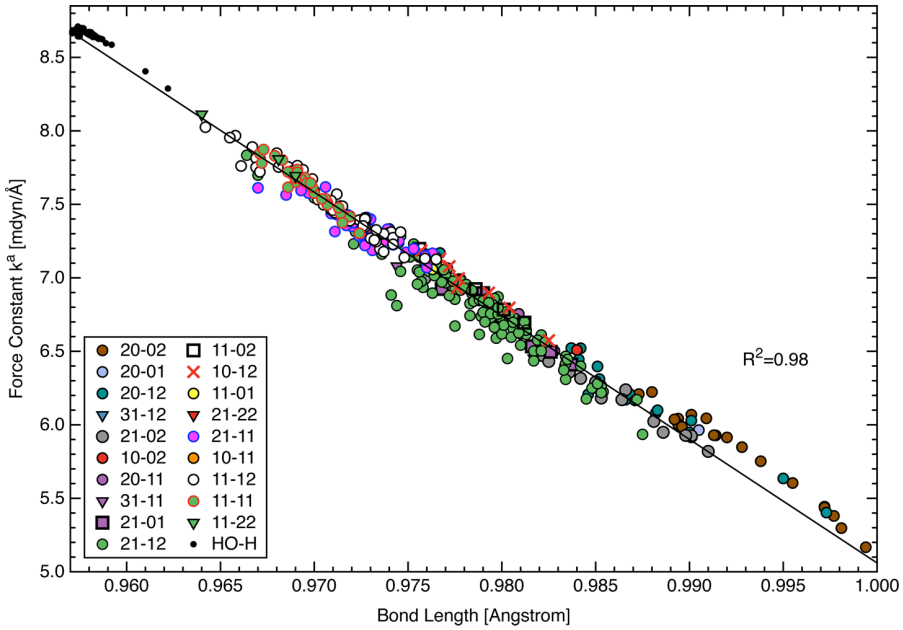


Figure 11. Testing the Badger relationship: Comparison of the local O−H stretching force constant $k^a(O{-}H)$ and the corresponding distance $R(O{-}H)$. For the relationship $k^a(O{-}H) = -84.095\ R(O{-}H) + 89.155$, $R^2 = 0.98$ and $\sigma = 0.087$ mdyn/Å result. For color code and notation, see Figures 1 and 7.

can be distinguished. Each is indicated in Figure 7 by a vertical solid or dashed line, which gives the average BSO value of a given O−H donor bond type that, for reasons of simplicity, is characterized by the H-bond it is engaged in.

There is a qualitative relationship between O−H donor bonds and H-bonds insofar as the weakest O−H donor bonds (see 20-02 in Figure 7) are associated with the strongest H-bonds and vice versa. This is in line with the covalent character of the strong H-bonds, which implies charge transfer from O2(lp) into the σ^*-orbital of the donor bond. Noteworthy is that the variation in the data points for the 20-02 donor bond is smaller than that for the corresponding H-bond, which again is a result of the fact that the first is primarily influenced by charge transfer and thereby a covalent weakening effect, whereas the latter is in addition influenced by electrostatic effects (for the calculated energy densities of the OH donor bonds, see the SI). Figure 7 also reveals that the various O−H donor bond types are much closer together with largely overlapping value ranges so that a differentiation on the basis of their average (Figure 7,

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

Table 1. Comparison of the Total Numbers and Percentages of All Types of Hydrogen Bonds for 1000 $(H_2O)_{1000}$ Clusters Determined along the MD Trajectories and Calculated with the TIP5P Force Field for an NPT Ensemble at 283 and 363 K[a]

H-bond type	N(283)	%	N(363)	%	Δ	Δ1	Δ2
20-02	22015	1.4	25003	2.0	2988	13.6	0.6
20-01	12337	0.8	32003	2.6	19666	159.4	1.8
10-01	7585	0.5	46884	3.8	39299	518.1	3.3
10-02	13163	0.8	36261	2.9	23098	175.5	2.1
20-12	91754	5.6	37873	3.0	−53881	−58.7	−2.6
group (1)	146854	9.1	178024	14.3	31170	21.2	
20-11	37676	2.3	39254	3.2	1578	4.2	0.9
21-02	106723	6.6	46078	3.7	−60645	−56.8	−2.9
10-11	24033	1.5	61427	4.9	37394	155.6	3.4
21-01	57788	3.6	59326	4.8	1538	2.7	1.2
10-12	57310	3.5	58129	4.7	819	1.4	1.2
11-02	46680	2.9	54498	4.4	7818	16.7	1.5
group (2)	330210	20.4	318712	25.7	−11498	−3.5	
11-01	26433	1.6	71123	5.7	44690	169.1	4.1
21-12	471376	29.0	72019	5.8	−399357	−84.7	−23.2
21-11	184060	11.3	75541	6.1	−108519	−59.0	−5.2
11-12	210040	12.9	90199	7.2	−119841	−57.1	−5.7
11-11	85691	5.3	96990	7.8	11299	13.2	2.5
group (3)	977600	60.1	405872	32.6	−571728	−58.5	
groups (1+2+3)	1454664	89.6	902608	72.6	−552056	−38.0	
00-00	138	$<0.0_5$	3852	0.3	3714	2691.3	0.3
00-01	839	0.1	11926	1.0	11087	1321.5	0.9
00-02	1399	0.1	9271	0.7	7872	562.7	0.6
00-10	309	$<0.0_5$	4499	0.4	4190	1356.0	0.4
00-11	2696	0.2	16510	1.3	13814	512.4	1.1
00-12	6310	0.4	15468	1.2	9158	145.1	0.8
01-00	351	$<0.0_5$	5078	0.4	4727	1346.7	0.4
01-01	2311	0.1	15569	1.2	13258	573.7	1.1
01-02	3912	0.2	12064	1.0	8152	208.4	0.8
01-10	869	0.1	5874	0.5	5005	575.9	0.4
01-11	7554	0.5	22058	1.8	14504	192.0	1.3
01-12	17432	1.1	20618	1.7	3186	18.3	0.6
10-00	1197	0.1	14940	1.2	13743	1148.1	1.1
10-10	2630	0.2	16292	1.3	13662	519.5	1.1
11-00	3804	0.2	22891	1.8	19087	501.8	1.6
11-10	9641	0.6	25610	2.1	15969	165.6	1.5
20-00	1824	0.1	10046	0.8	8222	450.8	0.7
20-10	4060	0.2	10295	0.8	6235	153.6	0.6
21-00	8576	0.5	18875	1.5	10299	120.1	1.0
21-10	19580	1.2	19768	1.6	188	1.0	0.4
group (4)	95432	5.9	281504	22.6	186072	195.0	
22-00	32	$<0.0_5$	146	$<0.0_5$	114	356.3	$<0.0_5$
31-00	305	$<0.0_5$	846	0.1	541	177.4	0.1
22-01	205	$<0.0_5$	367	$<0.0_5$	162	79.0	$<0.0_5$
31-01	1960	0.1	2697	0.2	737	37.6	0.1
31-10	465	$<0.0_5$	754	0.1	289	62.2	0.1
22-10	93	$<0.0_5$	191	$<0.0_5$	98	105.4	$<0.0_5$
22-02	242	$<0.0_5$	286	$<0.0_5$	44	18.2	$<0.0_5$
31-02	3697	0.2	2217	0.2	−1480	−40.0	$<0.0_5$
22-11	628	$<0.0_5$	659	0.1	31	4.9	0.1
31-11	4541	0.3	2826	0.2	−1715	−37.8	−0.1
22-12	1212	0.1	533	$<0.0_5$	−679	−56.0	−0.1
31-12	11712	0.7	2849	0.2	−8863	−75.7	−0.5
group (5)	25092	1.4	14371	1.1	−10721	−42.7	
00-22	382	$<0.0_5$	1169	0.1	787	206.0	0.1
00-13	11	$<0.0_5$	82	$<0.0_5$	71	645.5	$<0.0_5$
01-22	1230	0.1	1624	0.1	394	32.0	$<0.0_5$
10-22	2915	0.2	3699	0.3	784	26.9	0.1
01-13	32	$<0.0_5$	94	$<0.0_5$	62	193.8	$<0.0_5$
10-13	130	$<0.0_5$	323	$<0.0_5$	193	148.5	$<0.0_5$

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

Table 1. continued

H-bond type	$N(283)$	%	$N(363)$	%	Δ	Δ1	Δ2
11-13	441	$<0.0_5$	442	$<0.0_5$	1	0.2	$<0.0_5$
20-13	212	$<0.0_5$	211	$<0.0_5$	−1	−0.5	$<0.0_5$
20-22	3035	0.2	2008	0.2	−1027	−33.8	$<0.0_5$
12-22	163	$<0.0_5$	167	$<0.0_5$	4	2.5	$<0.0_5$
21-13	953	0.1	393	$<0.0_5$	−560	−58.8	−0.1
30-22	34	$<0.0_5$	48	$<0.0_5$	14	41.2	$<0.0_5$
group (6)	9538	0.6	10260	0.7	722	7.6	
22-22	224	$<0.0_5$	111	$<0.0_5$	−113	−50.4	$<0.0_5$
22-21	80	$<0.0_5$	70	$<0.0_5$	−10	−12.5	$<0.0_5$
31-22	286	$<0.0_5$	129	$<0.0_5$	−157	−54.9	$<0.0_5$
31-21	75	$<0.0_5$	93	$<0.0_5$	18	24.0	$<0.0_5$
group (7)	665	0.0	403	0.0	−262	−39.4	
groups (5+6+7)	35295	2.0	25034	1.8	−10261	−29.1	
group (8)	41475[b]	2.5	36534[c]	2.9	−4941	−11.9	
total	1626866	100.0	1245680	100.0	−381186	−30.6	

[a]In the first column the type of H-bonding is given in the short form introduced in the text. The table is partitioned into 5 groups: 1) 16 types of H-bonds discussed in connection with the 50-mers; 2) group 2 extends from 16 to the 36 standard types of H-bonds in general possible; 3) situations where the donor water (group 4), the acceptor water (group 5), or both (group 6) are involved in 5 H-bonds. $N(T)$ denotes the number of H-bonds at a given temperature T, Δ denotes the difference $N(363) - N(283)$, $\Delta 1 = 100[N(363) - N(283)]/N(283)$ is the percentage change of a given type of H-bond upon raising T to 363 K, $\Delta 2 = \eta(363) - \eta(283)$ is the difference between the percentage values $\eta = 100N(ij{-}kl)/N$ of a specific H-bond type at a certain T. [b]These groups include H-bond types with less than 0.05%, e.g., 00-03, 00-20, 00-21, 01-03, 01-20, 01-21, 01-23, 01-31, 01-32, 02-00, etc. (see Table S6, SI). [c]These groups include H-bond types with less than 0.05%. e.g. 00-03, 00-13, 00-14, 00-20, 00-21, 00-22, 00-23, 00-31, 00-32, 01-03, etc. (see Table S6, SI).

top) or median values (bottom) is difficult. Just 3 O−H donor bonds can be clearly distinguished (20-02, 20-01, HO-H).

In Figures 8 and 9, local stretching force constants k^a and distances R of the two interacting bonds are correlated. There is a quadratic relationship in the case of the force constants which is poorly fulfilled ($R^2 = 0.72$) thus revealing that the intrinsic strength of the O−H donor bond cannot provide a reliable measure for the intrinsic strength of the H-bond. Scattering of the data points is caused (among others) by those H-bonds involving 4-fold coordinated O atoms (e.g., green and purple dots both encircled in black).

For the corresponding distances, the scattering of data points is reduced ($R^2 = 0.93$, Figure 9), which might provide a basis to predict H-bond distances from known O−H bond lengths. The changes in the H-bond distance are 14 times larger than those in the O−H donor bond length. The corresponding force constants change as 1:7, which means that the R values are more sensitive than the corresponding k^a values in the case of the H-bonds, whereas the R-variation in the donor bonds is much smaller than that of the local force constants. This leads to the fact that the scattering is smaller (due to the smaller $R(O{-}H)$ variation) and a quantitative distance relationship can be found. The latter can be related to a change in the H-bond mechanism. For strong H-bonds, there is a linear relationship with a dominant covalent interaction between donor and acceptor bonds. For weak H-bonds, the electrostatic bonding mechanism becomes more important thus leading to a second linear R-relationship.

We conclude that the electrostatic contributions, especially in the case of the weak H-bonds, make it difficult (even in the case where covalent contributions are significant) to predict the properties of the H-bonds once those of the donor bonds are known. The fact that the O−H donor bond is primarily influenced by covalent effects and that the corresponding H-bonds are sensitive to both covalent and electrostatic effects is the reason why the corresponding force constants (or alternatively the BSO values) poorly correlate with each other. Previous claims to this extent were based on a small number of data points so that reliable conclusions could not be made.

Badger's Rule. The Badger rule implies that there is a power relationship between bond length and stretching force constant.[97] The rule was originally based on observations made for diatomic molecules[97,98] and was later generalized to covalent bonding in polyatomic molecules.[99] One might expect that the Badger rule is largely fulfilled for the OH donor bonds and less for the H-bonds themselves. Figures 10 and 11 show the two different situations.

There is a linear relationship between distance R and stretching force constant k^a in both cases, however associated with some scattering of data points in the case of the H-bonds ($R^2 = 0.91$). Figure 10 reveals that the scattering is caused predominantly by the 21-12 H-bonds (green dots in black circles) and the H-bonds involving pentacoordinated O (triangles), but even for the strongly covalent 20-02 H-bonds (brown dots in black circles), a linear relationship between R and k^a cannot be obtained. Obviously, the generalized Badger rule is for H-bonding of limited value.

The situation is better in the cases of the covalent O−H donor bonds (Figure 11). Deviations from a linear relationship ($R^2 = 0.98$) are found again for the O−H bonds involved in 21-12 H-bonding (in short: 21-12 O−H bonds). In general, those bonds, which involve tetra- or pentacoordinated O atoms, lead to scattering.

WHY DOES WARM WATER FREEZE FASTER THAN COLD WATER?

The fact that warm water freezes faster than cold water is known in the literature as the Mpemba effect according to its first discovery by Mpemba.[138] This macroscopic phenomenon has been investigated many times and explained in many different ways referring, e.g., to the temperature gradient in the liquid, impurities in the water, dissolved oxygen and carbon dioxide, evaporation from the freezing liquid, difference in heat

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

Table 2. Comparison of the Numbers and Percentages of Bifurcated Hydrogen Bonds Found for 1000 $(H_2O)_{1000}$ Clusters at 283 and 363 K[a]

H-bond type with bifurcation	N(283)	%	N(363)	%	Δ	Δ1	Δ2
22-00	32	<0.0$_5$	146	<0.0$_5$	114	356.3	<0.0$_5$
22-01	205	<0.0$_5$	367	<0.0$_5$	162	79.0	<0.0$_5$
22-10	93	<0.0$_5$	191	<0.0$_5$	98	105.4	<0.0$_5$
22-02	242	<0.0$_5$	286	<0.0$_5$	44	18.2	<0.0$_5$
22-11	628	<0.0$_5$	659	0.1	31	4.9	0.1
22-12	1212	0.1	533	<0.0$_5$	−679	−56.0	−0.1
group (4)	2412	0.1	2182	0.1	−230	−9.5	
12-22	163	<0.0$_5$	167	<0.0$_5$	4	2.5	<0.0$_5$
group (5)	163	<0.0$_5$	167	<0.0$_5$	4	2.5	
22-22	224	<0.0$_5$	111	<0.0$_5$	−113	−50.4	<0.0$_5$
22-21	80	<0.0$_5$	70	<0.0$_5$	−10	−12.5	<0.0$_5$
group (6)	304	<0.0$_5$	181	<0.0$_5$	−123	−40.5	
total (4+5+6)	2879	0.1	2530	0.1	−349	−12.1	

[a]For explanations, see Table 1. Bifurcated H-bonds can only be found for groups 4, 5, and 6 in Table 1.

loss, supercooling (lowering of liquid water below the freezing point), thermal convection currents, etc.[60−66] Most of these explanations had to be revoked, which had to do to some extent with the practical difficulties of having exact, reproducible conditions for the experiment (size, shape, and material of the freezer cabinet, circulation of air above the cooling liquid, properties of the cooling system; balancing of the volume, conditions of supercooling, etc.).[60−66]

Attempts have been made to explain the Mpemba effect with the help of MD simulations and by relating microscopic details to complex macroscopic phenomena such as the thermal convection or supercooling.[139] Analyzing the results of the MD simulations of liquid water as modeled by a cluster of 1000 water molecules at 283 and 363 K (with control calculations at 308 and 378 K) over a time of 2 ns and utilizing the insight gained into H-bonding we will offer here a molecular explanation of the Mpemba effect directly being based on the interplay of strong (mostly more covalent) and weak (mostly more electrostatic) contributions as they emerge from the four model clusters and the MD simulations (Tables 1, 2, Tables S9−S12 in the SI, Figures 12, 13).

A total of 1.627 (283 K), 1.496 (308 K), 1.246 (363 K), and 1.182 million H-bonds (378 K) has been investigated where the analysis at 283 and 363 K is presented here. Increasing the temperature T from 283 to 363 K leads to a decrease of the average number of H-bonds per water molecule, m_{av}, from 3.254 to 2.491. This decrease can be analyzed utilizing the various types of H-bonds found for the 50-mers where one has to consider that in liquid water there are more possibilities for H-bonding than in the equilibrium geometries of the 50-mers. The 36 possible H-bonds (see Figure S1 of the SI) are all found in the MD simulations of liquid water using $(H_2O)_{1000}$ clusters as suitable models for the analysis (see Table 1, Figure 12, and the SI). In the bar diagram of Figure 12 (top), the 16 types of H-bonds of Figure 1 are ordered according to the intrinsic strength calculated for the 50-mers, whereas the 20 remaining ones (bottom, group 4 in Table 1) are ordered according to Figure S1 in the SI. The latter include 11 types for which D or/and A water do not have any other H-bond than the targeted one, which means that these are throughout relatively weak H-bonds. Their numbers should increase when H-bonds are broken, for example as a result of a T increase. This is confirmed by the significant increase of the group 4 H-bonds at 363 K by 186072 (Table 1 and red bars in the lower part of Figure 12). In total, about 381000 H-bonds are lost for the increase in T.

Table 1 gives details on the distribution of H-bonds at the two different T values (for other T, see the SI). Almost 90% (1454664 H-bonds) of all H-bonds analyzed at 283 K belong to the 16-types discussed in connection with the 50-mers. If one uses the partitioning in strong, normal, and weak H-bonds discussed in Section 3, then 9.1% of all H-bonds (group 1 in Table 1: 20-02, 20-01, 10-01, 10-02, 20-12) can be considered as being strong and 60.1% as being weak (group 3 in Table 1: 11-01, 21-12, 21-11, 11-12, 11-11) whereas the rest of 20.4% is of normal strength (group 2 in Table 1).

5.9% (95432) of all H-bonds belong to the group of the 20 H-bond types with either terminal water (*dangling H-bonds*) or a change in the position of the peripheral H-bonds that disturbs the flow of charge from the A water to the D water (group 4 in Table 1). The latter is supported by i_a0-0l_d combinations such as in 20-02 or 10-01 but hindered in $0j_d - 0l_d$ or $i_a0 - k_a0$ H-bond types as in 01-01, 10-10, etc. (see Figure S1).

Bifurcated H-Bonds and Unusual Coordination Numbers. If a H-bond is associated with two or more acceptor atoms O2, one speaks of bifurcated H-bonds.[107] We note that in the literature this term is used sometimes also for O atoms that have a coordination number higher than 4 (*penta- or hexacoordinated O*). We will not follow this more general use of the term but distinguish here clearly between bi- (tri)furcated H-bonds and H-bonds involving penta (hexa)-coordinated O atoms. For the former, we did not find any examples in the 50-mers. Using the distance criterion described in Section 2, some of the OH donor bonds had to be assigned to acceptor O atoms that were already hosting two H-bonds. This increased the coordination number to 5. Four different types of H-bonding involving pentacoordinated water molecules were observed, which according to their intrinsic strength can be ordered as follows: 31-12 (7) > 31-11 (3) > 21-22 (12) > 11-22 (3) (their numbers are given in parentheses; Figure 3).

Increase of the coordination number of the D atom O2 by increased acceptance of peripheral H-bonds leads to a strengthening of the targeted H-bond as the polarization of the charge distribution at O2 is increased (see 31-12 and 31-11 in the upper curve of Figure 5). This causes a stronger covalent contribution to the H-bond under investigation. However, if the targeted H-bond has to compete with two other H-bonds as in 21-22 or 11-22 for the lp(O1) electrons, a relatively weak H-

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

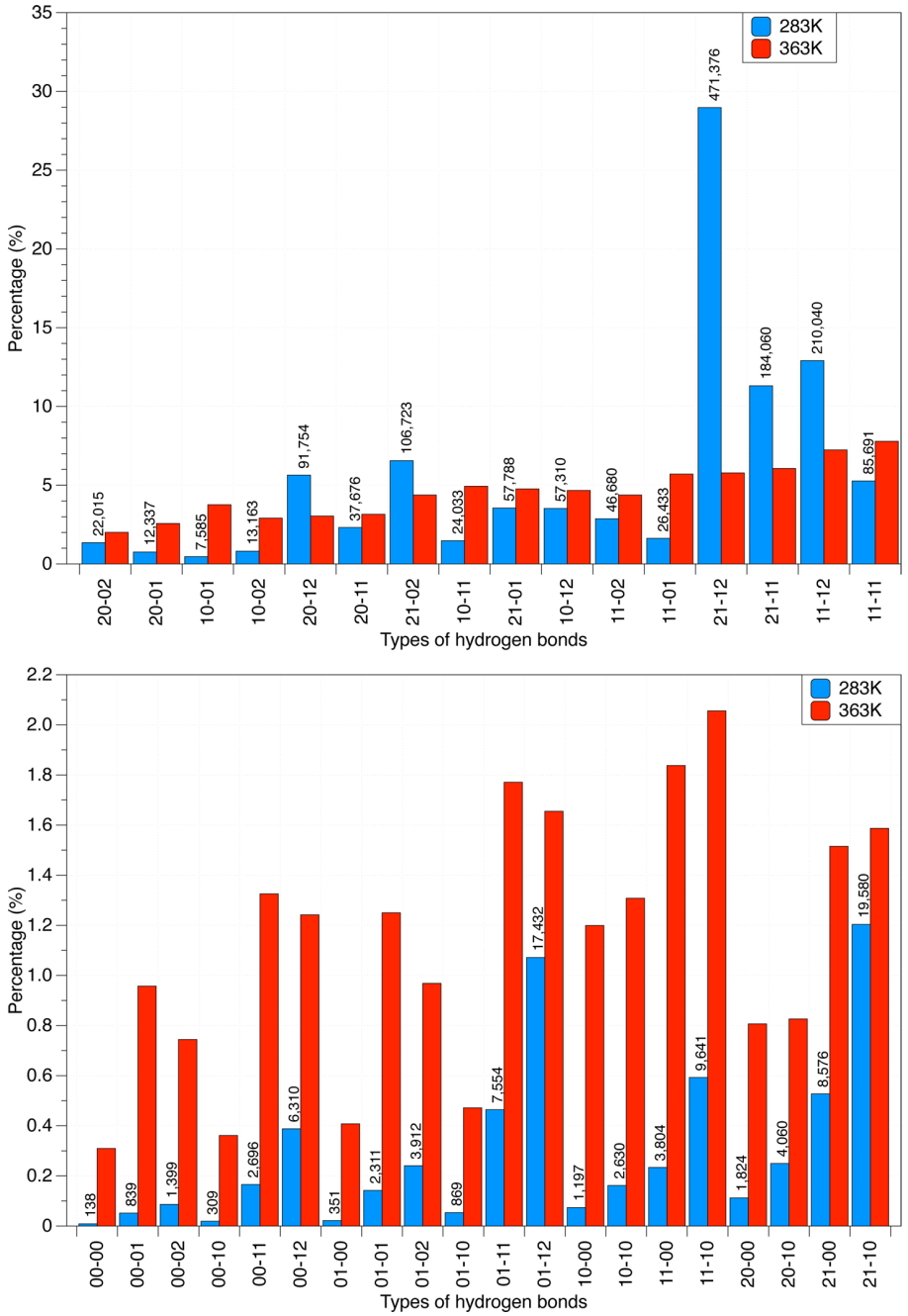


Figure 12. Percentages of the 36 types of different H-bonds in liquid water as modeled by 1000 $(H_2O)_{1000}$ clusters for the two different temperatures 283 (blue bars) and 363 K (red bars). The numbers at the blue bars give the actual H-bond count for a given type. Top: The 16 types of H-bonds discussed in connection with the 50-mers ordered according to strength. Bottom: The remaining 20 types of H-bonds not found for the 50-mers have all percentages ≤2%. See Figure S1 and Tables S9−S12 of the SI.

bond with a large variation in its BSO value due to electrostatic contributions results (Figure 5).

In the liquid water model used, all 36 + 71 = 107 different types of H-bonds with water containing pentacoordinated O (groups 5, 6, and 7 in Table 1; for a full account see Tables S7 and S8 of the SI) are found. Two different topologies can lead to pentacoordination: Either an O atom of the water molecule accepts 3 rather than 2 H-bonds (see above) or one of the donor H-bonds is bifurcated, i.e. it is equally H-bonded to two different O(A) atoms (Tables 2 and S8 of the SI). In the case of group 5, just 1.4% of the H-bonds (25092) is bifurcated, for group 6 just 9538 (0.6%), and for group 7 not more than 665 (Tables 2, S8, and Figure 13). Most of the 71 possible bifurcated H-bond types are found in the water model used (Table S8 of the SI) although their number (2969, Table S8) is relatively small.

The decrease in the average number of H-bonds with increasing T results from the fact that for higher T weak H-bonds such as 21-12 are cleaved (Figure 12, top right), which leads to the generation of fragments with terminal water (dangling H-bonds; Figure 12, bottom). Noteworthy is the fact that at 283 K only 2274 (0.22%) are not H-bonded (free),

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

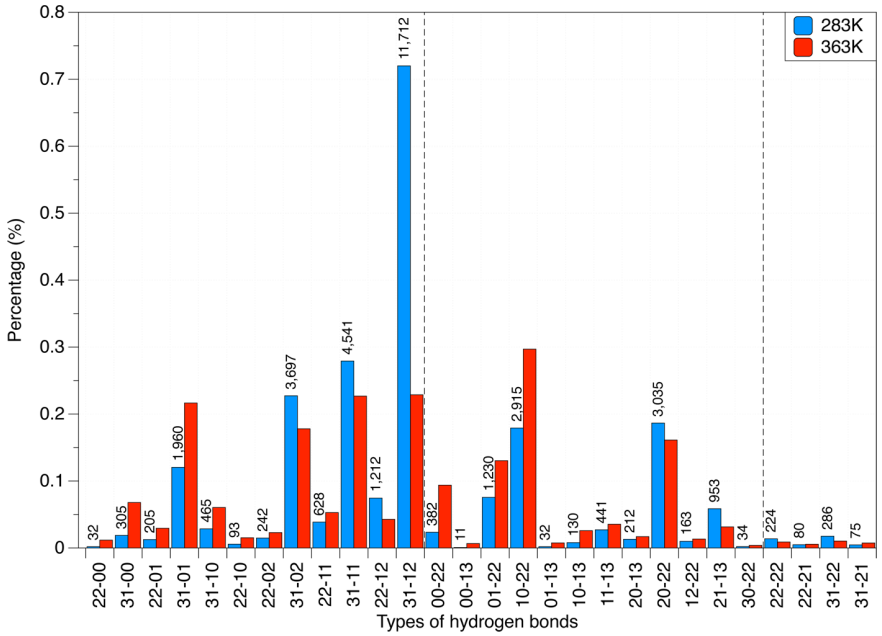


Figure 13. Percentages of bifurcated H-bonds in liquid water as modeled by 1000 $(H_2O)_{1000}$ clusters for the two different temperatures 283 (blue bars) and 363 K (red bars). The numbers at the blue bars give the actual H-bond count for a given type. First third up to dashed vertical line: D molecules with 5 H-bonds ($m(D) = 5$); second third: A molecules with $m(A) = 5$; last third: both D and A have $m = 5$. See Table 2.

whereas this number increases by a factor of 10 to 23494 (2.35%) at 363 K. Hence, the T increases substantially the number of water molecules with dangling H-bonds. The H-bonds being cleaved are, as found in this work, those with predominantly electrostatic nature. There is a significant increase in the percentages of the strong H-bond types (20-02, 20-01, 10-01, 10-02). This means that at higher T more water clusters with strong H-bonds remain (percentage-wise), whereas those with dangling H-bond increase, Figure 14), which can optimally recombine to form relatively strong covalent H-bonds as the geometrical prerequisites for covalent H-bonding are better fulfilled in smaller clusters.

The hexagonal lattice of ice is easily formed by these fragments because the fragments can easily adjust to the tetragonal environment of each O atom in solid ice (Figure 14). In cold water, the average number of H-bonds, m_{av} is higher. Both covalent and electrostatic H-bonds can be expected. The geometrical arrangement of the H-bonds is not optimal in those cases where both electrostatic and covalent bonding is possible (Figure 14). The electrostatic bonds have to be cleaved first and to rearrange before the water clusters can form the hexagonal ice lattice (Figure 14). This costs time and energy and is the reason why cold water freezes more slowly than warm water.

At this point, a caveat is necessary. The time scale of the nucleation process leading to the solid structure of ice is much larger than the time scale of H-bond cleavage or formation. However, cluster fragments with strong H-bonds are ideal for the nucleation process and thereby will accelerate freezing of water, whereas the larger m_{av} at lower T caused by a dense H-bond net with weak and strong H-bonds will slow down the nucleation process because weak, nonoptimal H-bonds have to be cleaved first. To prove this point MD simulations up to μs will be needed in the future.

CONCLUSIONS

The modeling of liquid water with 50-mers and 1000-mers using both quantum chemistry and MD simulation has led to a set of interesting results.

1) For the 50-mers investigated, a subset of 16 H-bond types (out of 36 possible H-bond types for a coordination number $c_O \leq 4$, excluding H-bond bifurcation) is sufficient to characterize the H-bond network of a 50-mer in its equilibrium geometry. For $c_O \leq 5$ (including pentacoordinated O with and without bifurcation), 36 + 71 = 107 additional H-bond types can be distinguished, which were found (with a few exceptions) among the 1.6 million H-bonds analyzed in the course of the MD simulations of this work. For the 50-mers, only a few H-bonds with $c_O = 5$ (25 out of 350) were identified, which involved pentacoordinated O atoms.

2) Utilizing the local stretching force constant and difference electron density distributions $\Delta\rho(\mathbf{r})$, we could determine the strongest type of H-bond to be the cooperative push−pull bond 20-02, which has an average BSO value of 0.41 and is, compared to the total range of BSO values from 0.22-0.42, at the upper limit of what seems to be possible in a water cluster. The push−pull effect of the peripheral water molecules has been verified by showing a suitably defined $\Delta\rho(\mathbf{r})$ calculated for the van der Waals surface of a 20-02 hexamer.

3) Any perturbation of the 20-02 H-bond by competition with other H-bonds (first order perturbations) or changing the pushing (pulling) H-bonds (second order perturbations) causes a weakening of the H-bond, which can be qualitatively analyzed. A rational explanation of the ordering of the H-bond types found in the 50-mers according to strength has been given in this work.

4) For the 50-mers, two linear relationships between the average intrinsic strength n and the average delocalization energy $\Delta E(del)$ (caused by charge transfer from lp(O1) to σ^*(O2H) ; Figures 3 and 5) could be derived. Since the charge transfer relates to the covalent character of a H-bond, it is

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

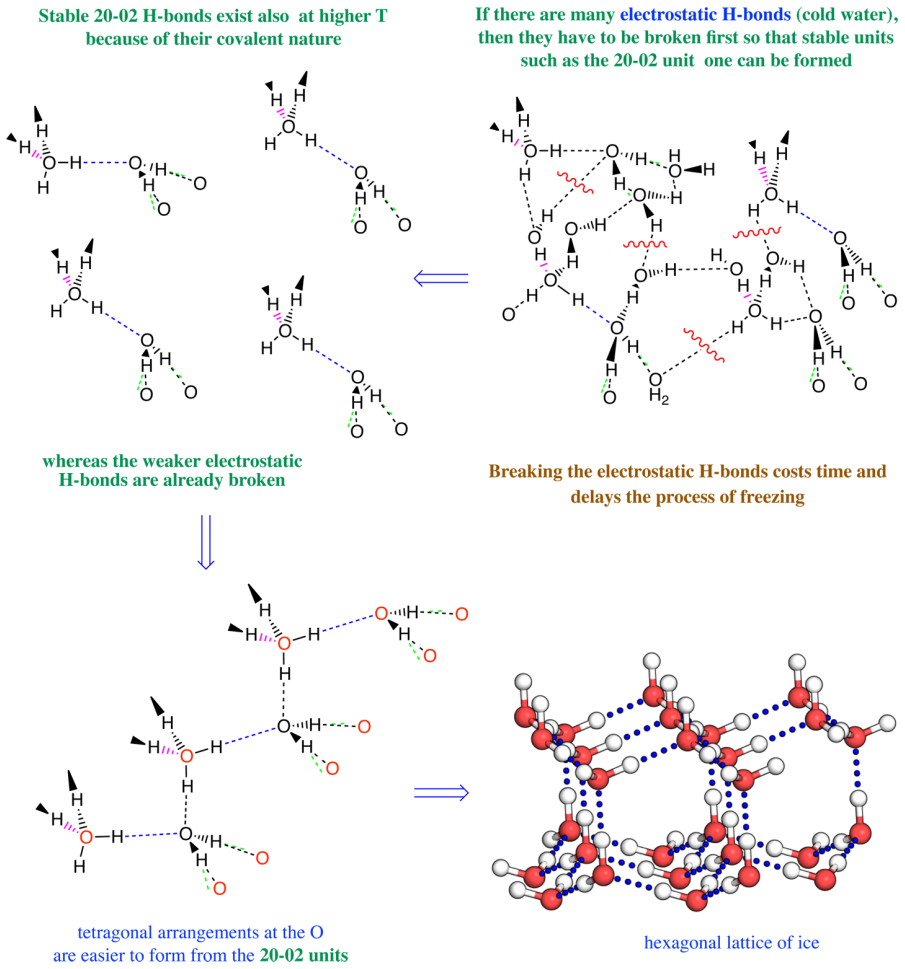


Figure 14. Explanation of the Mbempa effect: In warm water (upper left corner), weak, electrostatic H-bonds are already broken so that only those cluster units with strong covalent H-bonding exist, which more easily arrange (lower left corner) as is needed for the formation of the hexagonal ice lattice (lower right corner). In cold water (upper right corner), many electrostatic H-bonds (red wiggles) still exist, which have first to be broken to form the ice lattice, which costs time and energy.

reasonable to say that, within the model used, the covalent contributions influence the intrinsic strength of the various H-bond types. A rational has been given to explain the linear relationships; however, additional investigations are needed to verify the trends of $\Delta E(del)$ on a quantitative basis. An alternative measure is provided by the energy density at the H-bond critical points, which suggests that covalent and electrostatic interactions both play a role. With increasing nonlinearity of the H-bonding unit the electrostatic character increases.

5) Pentacoordination of an O atom does not necessarily lead to a weakening of the targeted H-bond. For example, converting a 21-12 H-bond into a 31-12 H-bond increases the polarization of the donor bond and thereby the cooperative effects so that the BSO value of the targeted H-bond raises by almost 21.6%. However, when the targeted H-bond gets a third competitor for the O1 electron lone pairs as for the 11-22 or 21-22 H-bonds, a significant weakening of the H-bond results.

6) In the MD simulations, almost all of the 107 additional H-bond types for $c_O = 5$ were found (together <4%) of which a subset of 0.1% was due to H-bond bifurcation.

7) For dominantly covalent H-bonds, there is an inverse relationship between the strength of the H-bond and that of the donor bond (Figure 8), which is of more qualitative nature. When using O−H and H···O distances, a more quantitative relationship results (Figure 9). The analysis reveals that there is a covalent and an electrostatic H-bonding mechanism active.

8) The Badger rule for the H-bonds and O−H donor bonds investigated is only semiquantitatively fulfilled where the 21-12 H-bonds are the major cause for data point scattering. Scattering is smaller for the relationship for the O−H donor bonds because the covalent character of H-bonding prevails in this case.

9) Based on the quantum chemical analysis of the 50-mers and the MD simulations leading to the investigation of 1.6 million H-bonds, the distribution of the most important H-bonds could be determined for different T and ordered according to their strength. This was possible because the majority of H-bonds identified belongs to the group analyzed for the 50-mers. The analysis of the MD simulation results leads us to propose a molecular explanation for the Mpemba effect. In warm water, the weaker H-bonds with predominantly electrostatic contributions are broken, and smaller water clusters with 20-02 or related strong H-bonding arrangements exist that accelerate the nucleation process that leads to the hexagonal lattice of solid ice. Therefore, warm water freezes faster than cold water in which the transformation from randomly arranged water clusters costs time and energy.

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

Future studies prolonging the MD simulations to the μs range will help to verify the proposed molecular explanation of the Mpemba effect. Apart from this, the possibility of full-dimensional quantum simulations at a given T and a given pressure has to be considered.[55] There is also a necessity to provide further evidence for the push−pull effect of peripheral water molecules strengthening the targeted H-bond. Work is in progress to do this for the most frequent H-bond types.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00735.

Distribution of H-bonds in the 50-mers (Table S1), H-bond properties calculated for the different 50-mers (Tables S2−S5); numbers and percentages of different H-bond types at different T values (Tables S6−S12), Cartesian coordinates (Tables S13−S14), presentation of all 36 different H-bonds (Figure S1), perspective drawings and numbering of atoms for all 50-mers (Figures S2−S5), box and whisker diagram for the distribution of the energy density of the OH donor bonds (Figure S6), analysis of the BSO values of the 50-mers (Figure S7) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dcremer@smu.edu.

ORCID

Dieter Cremer: 0000-0002-6213-5555

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank one of the referees for useful suggestions on how to extend and improve our investigation. This work was financially supported by the National Science Foundation, Grants CHE 1152357 and 1464906 and the National Science Foundation of China, Grant 21473087. We thank SMU for providing computational resources.

■ REFERENCES

(1) Pimentel, G. C.; McClellan, A. L. *The Hydrogen Bond*; W.H. Freeman: San Francisco, 1960.

(2) Jeffrey, G. A.; Saenger, W. *Hydrogen bonding in biological structures*; Springer-Verlag: Berlin, 1994.

(3) Stone, A. *The Theory of Intermolecular Forces*; Clarendon: Oxford, 1997.

(4) Scheiner, S. *Hydrogen Bonding: A Theoretical Perspective*; Oxford University Press: New York, 1997.

(5) Grabowski, S. J. E. *Hydrogen Bonding - New Insights, in Challenges and Advances in Computational Chemistry and Physics*; Springer: New York, 2006.

(6) Gilli, G.; Gilli, P. *The Nature of the Hydrogen Bond- IUCr Monographs on Crystallography - 23*; Oxford University Press: New York, 2009.

(7) Keutsch, F. N.; Saykally, R. J. Water clusters: Untangling the mysteries of the liquid, one molecule at a time. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10533−10540.

(8) Morokuma, K.; Pedersen, L. Molecular-Orbital Studies of Hydrogen Bonds. An Ab Initio Calculation for Dimeric H_2O. *J. Chem. Phys.* **1968**, *48*, 3275−3282.

(9) Kollman, P. A.; Allen, L. C. Theory of the Hydrogen Bond: Electronic Structure and Properties of the Water Dimer. *J. Chem. Phys.* **1969**, *51*, 3286−3293.

(10) Diercksen, G. H. F. SCF MO LCGO studies on hydrogen bonding. The water dimer. *Chem. Phys. Lett.* **1969**, *4*, 373−375.

(11) Morokuma, K.; Winick, J. R. Molecular-Orbital Studies of Hydrogen Bonds. Dimeric H_2O with the Slater Minimal Basis. *J. Chem. Phys.* **1970**, *52*, 1301−1306.

(12) Hankins, D.; Moskowitz, J. W.; Stillinger, F. H. Water Molecule Interactions. *J. Chem. Phys.* **1970**, *53*, 4544−4554.

(13) Del Bene, J.; Pople, J. Theory of Molecular Interactions. I. Molecular Orbital Studies of Water Polymers Using a Minimal Slater-Type Basis. *J. Chem. Phys.* **1970**, *52*, 4858−4866.

(14) Diercksen, G. H. F. SCF MO LCGO studies on hydrogen bonding. The water dimer. *Theor. Chim. Acta* **1971**, *21*, 335−367.

(15) Del Bene, J. E. Theoretical Study of Open Chain Dimers and Trimers Containing CH_3OH and H_2O. *J. Chem. Phys.* **1971**, *55*, 4633−4636.

(16) Popkie, H.; Kistenmacher, H.; Clementi, E. Study of the structure of molecular complexes. IV. The Hartree-Fock potential for the water dimer and its application to the liquid state. *J. Chem. Phys.* **1973**, *59*, 1325−1336.

(17) Matsuoka, O.; Clementi, E.; Yoshimine, M. CI study of the water dimer potential surface. *J. Chem. Phys.* **1976**, *64*, 1351−1361.

(18) Tschumper, G. S.; Leininger, M. L.; Hoffman, B. C.; Valeev, E. F.; Schaefer, H. F.; Quack, M. Anchoring the Water Dimer Potential Energy Surface with Explicitly Correlated Computations and Focal Point Analyses. *J. Chem. Phys.* **2002**, *116*, 690−701.

(19) Kalescky, R.; Zou, W.; Kraka, E.; Cremer, D. Local Vibrational Modes of the Water Dimer - Comparison of Theory and Experiment. *Chem. Phys. Lett.* **2012**, *554*, 243−247.

(20) Lane, J. R. CCSDTQ Optimized Geometry of Water Dimer. *J. Chem. Theory Comput.* **2013**, *9*, 316−323.

(21) Pugliano, N.; Saykally, R. Measurement of Quantum Tunneling Between Chiral Isomers of the Cyclic Water Trimer. *Science* **1992**, *257*, 1937−1940.

(22) Mó, O.; Yanez, M.; Elguero, J. Cooperative (non-pairwise) effects in water trimers: An ab initio molecular orbital study. *J. Chem. Phys.* **1992**, *97*, 6628−6638.

(23) Schütz, M.; Bürgi, T.; Leutwyler, S.; Bürgi, H. B. Fluxionality and low-lying transition structures of the water trimer. *J. Chem. Phys.* **1993**, *99*, 5228−5238.

(24) Wales, D. J. Theoretical Study of Water Trimer. *J. Am. Chem. Soc.* **1993**, *115*, 11180−11190.

(25) Liu, K.; Loeser, J.; Elrod, M.; Host, B.; Rzepiela, J.; Pugliano, N.; Saykally, R. Dynamics of Structural Rearrangements in the Water Trimer. *J. Am. Chem. Soc.* **1994**, *116*, 3507−3512.

(26) Schütz, M.; Klopper, W.; Lüthi, H.-P. Low-lying stationary points and torsional interconversions of cyclic $(H_2O)_4$: An *ab initio* study. *J. Chem. Phys.* **1995**, *103*, 6114−6126.

(27) Fowler, J. E.; Schaefer, H. F., III Detailed Study of the Water Trimer Potential Energy Surface. *J. Am. Chem. Soc.* **1995**, *117*, 446−452.

(28) Wales, D.; Walsh, T. Theoretical study of the water pentamer. *J. Chem. Phys.* **1996**, *105*, 6957−6971.

(29) Wales, D.; Walsh, T. Theoretical study of the water tetramer. *J. Chem. Phys.* **1997**, *106*, 7193−7207.

(30) Taketsugu, T.; Wales, D. Theoretical study of rearrangements in water dimer and trimer. *Mol. Phys.* **2002**, *100*, 2793−2706.

(31) Anderson, J. A.; Crager, K.; Fedoroff, L.; Tschumper, G. S. Anchoring the Potential Energy Surface of the Cyclic Water Trimer. *J. Chem. Phys.* **2004**, *121*, 11023−11029.

(32) Pérez, J.; Hadad, C.; Restrepo, A. Structural Studies of the Water Tetramer. *Int. J. Quantum Chem.* **2008**, *108*, 1653−1659.

(33) Shields, R. M.; Temelso, B.; Archer, K. A.; Morrell, T. E.; Shields, G. C. Accurate Predictions of Water Cluster Formation, $(H_2O)_{n=2-10}$. *J. Phys. Chem. A* **2010**, *114*, 11725−11737.

(34) Ramírez, F.; Hadad, C.; Guerra, D.; David, J.; Restrepo, A. Structural studies of the water pentamer. *Chem. Phys. Lett.* **2011**, *507*, 229−233.
(35) Temelso, B.; Archer, K.; Shields, G. Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections. *J. Phys. Chem. A* **2011**, *115*, 12034−12046.
(36) Temelso, B.; Shields, G. C. The Role of Anharmonicity in Hydrogen-Bonded Systems: The Case of Water Clusters. *J. Chem. Theory Comput.* **2011**, *7*, 2804−2817.
(37) Howard, J. C.; Tschumper, G. S. Wavefunction methods for the accurate characterization of water clusters. *WIREs Comput. Mol. Sci.* **2014**, *4*, 199−224.
(38) Fanourgakis, G.; Apra, E.; de Jong, W.; Xantheas, S. High-level ab-initio calculations for the four low-lying families of minima of $(H_2O)_{20}$: II. Spectroscopic signatures of the dodecahedron, fused cubes, face-sharing pentagonal prisms, and edge-sharing pentagonal prisms hydrogen bonding networks. *J. Chem. Phys.* **2005**, *122*, 134304.
(39) Bryantsev, V.; Diallo, M.; van Duin, A.; Goddard, W. Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 1016−1026.
(40) Parthasarathi, R.; Elango, M.; Subramanian, V.; Sathyamurthy, N. Structure and Stability of Water Chains $(H_2O)_n$, n = 5−20. *J. Phys. Chem. A* **2009**, *113*, 3744−3749.
(41) Yang, F.; Wang, X.; Yang, M.; Krishtal, A.; van Alsenoy, C.; Delarue, P.; Senet, P. Effect of hydrogen bonds on polarizability of a water molecule in $(H_2O)_n$ (n = 6, 10, 20) isomers. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9239.
(42) Iwata, S. Analysis of hydrogen bond energies and hydrogen bonded networks in water clusters $(H_2O)_{20}$ and $(H_2O)_{25}$ using the charge-transfer and dispersion terms. *Phys. Chem. Chem. Phys.* **2014**, *16*, 11310.
(43) Lenz, A.; Ojamae, L. Theoretical IR Spectra for Water Clusters $(H_2O)_n$ (n = 6−22, 28, 30) and Identification of Spectral Contributions from Different H-Bond Conformations in Gaseous and Liquid Water. *J. Phys. Chem. A* **2006**, *110*, 13388−13393.
(44) Bako, I.; Mayer, I. Hierarchy of the Collective Effects in Water Clusters. *J. Phys. Chem. A* **2016**, *120*, 631−638.
(45) Elango, M.; Subramanian, V.; Sathyamurthy, N. Structure and stability of spiro-cyclic water clusters. *J. Chem. Sci.* **2009**, *121*, 839−848.
(46) Sahu, N.; Khire, S.; Gadre, S. Structures, energetics and vibrational spectra of $(H_2O)_{32}$ clusters: a journey from model potentials to correlated theory. *Mol. Phys.* **2015**, *113*, 2970−2979.
(47) Qian, P.; Song, W.; Lu, L.; Yang, Z. Ab initio investigation of water clusters $(H_2O)_n$ (n = 2−34). *Int. J. Quantum Chem.* **2009**, *110*, 1924−1937.
(48) Huang, L.; Lambrakos, S.; Shabaev, A.; Bernstein, N.; Massa, L. Molecular analysis of water clusters: Calculation of the cluster structures and vibrational spectrum using density functional theory. *C. R. Chim.* **2015**, *18*, 516−524.
(49) Liu, J.; Wang, L.; Zhao, J. Density Functional Theory Calculations of Water Fullerenes: $(H_2O)_n$ Clusters with n = 20−40. *J. Comput. Theor. Nanosci.* **2009**, *6*, 454−458.
(50) Ludwig, R.; Appelhagen, A. Calculation of Clathrate-Like Water Clusters Including H_2O-Buckminsterfullerene. *Angew. Chem., Int. Ed.* **2005**, *44*, 811−815.
(51) Neela, Y.; Mahadevi, A.; Sastry, G. Hydrogen Bonding in Water Clusters and Their Ionized Counterparts. *J. Phys. Chem. B* **2010**, *114*, 17162−17171.
(52) Frigato, T.; VandeVondele, J.; Schmidt, B.; Schutte, C.; Jungwirth, P. Ab Initio Molecular Dynamics Simulation of a Medium-Sized Water Cluster Anion: From an Interior to a Surface-Located Excess Electron via a Delocalized State. *J. Phys. Chem. A* **2008**, *112*, 6125−6133.
(53) Loboda, O.; Goncharuk, V. Theoretical study on icosahedral water clusters. *Chem. Phys. Lett.* **2010**, *484*, 144−147.
(54) Turi, L. Hydration dynamics in water clusters via quantum molecular dynamics simulations. *J. Chem. Phys.* **2014**, *140*, 204317.
(55) Wang, Y.; Babin, V.; Bowman, J.; Paesani, F. The Water Hexamer: Cage, Prism, or Both. Full Dimensional Quantum Simulations Say Both. *J. Am. Chem. Soc.* **2012**, *134*, 11116−11119.
(56) Konkoli, Z.; Cremer, D. A New Way of Analyzing Vibrational Spectra. I. Derivation of Adiabatic Internal Modes. *Int. J. Quantum Chem.* **1998**, *67*, 1.
(57) Konkoli, Z.; Cremer, D. A New Way of Analyzing Vibrational Spectra. III. Characterization of Normal Vibrational Modes in Terms of Internal Vibrational Modes. *Int. J. Quantum Chem.* **1998**, *67*, 29.
(58) Zou, W.; Kalescky, R.; Kraka, E.; Cremer, D. Relating Normal Vibrational Modes To Local Vibrational Modes With The Help of an Adiabatic Connection Scheme. *J. Chem. Phys.* **2012**, *137*, 084114.
(59) Freindorf, M.; Kraka, E.; Cremer, D. A Comprehensive Analysis of Hydrogen Bond Interactions Based on Local Vibrational Modes. *Int. J. Quantum Chem.* **2012**, *112*, 3174−3187.
(60) Ahtee, M. Investigation into the Freezing of Liquids. *Phys. Educ.* **1969**, *4*, 379−380.
(61) Firth, I. Cooler? *Phys. Educ.* **1971**, *6*, 32−41.
(62) Kell, G. The Freezing of Hot and Cold Water. *Am. J. Phys.* **1969**, *37*, 564−565.
(63) Auerbach, D. Supercooling and the Mpemba effect: When hot water freezes quicker than cold. *Am. J. Phys.* **1995**, *63*, 882−885.
(64) Wojciechowski, B. Freezing of Aqueous Solutions Containing Gases. *Cryst. Res. Technol.* **1988**, *23*, 843−8.
(65) Brewster, R.; Gebhart, B. An experimental study of natural convection effects on downward freezing of pure water. *Int. J. Heat Mass Transfer* **1988**, *31*, 331−348.
(66) van der Elsken, J. The Freezing of Supercooled Water. *J. Mol. Struct.* **1991**, *250*, 245−251.
(67) Chai, J. D.; Head-Gordon, M. Long-range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(68) Chai, J. D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, 084106−15.
(69) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. Assessment of the Performance of DFT and DFT-D Methods for Describing Distance Dependence of Hydrogen-Bonded Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 88−96.
(70) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918−1931.
(71) Bauzá, A.; Alkorta, I.; Frontera, A.; Elguero, J. On the Reliability of Pure and Hybrid DFT Methods for the Evaluation of Halogen, Chalcogen, and Pnicogen Bonds Involving Anionic and Neutral Electron Donors. *J. Chem. Theory Comput.* **2013**, *9*, 5201−5210.
(72) Scheiner, S. Extrapolation to the Complete Basis Set Limit for Binding Energies of Noncovalent Interactions. *Comput. Theor. Chem.* **2012**, *998*, 9−13.
(73) Ditchfield, D.; Hehre, W.; Pople, J. Self-Consistent Molecular Orbital Methods. 9. Extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **1971**, *54*, 724.
(74) Hariharan, P.; Pople, J. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213−222.
(75) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294−301.
(76) Li, W.; Li, S.; Jiang, Y. Generalized Energy-Based Fragmentation Approach for Computing the Ground-State Energies and Properties of Large Molecules. *J. Phys. Chem. A* **2007**, *111*, 2193−2199.
(77) Li, S.; Li, W.; Ma, J. Generalized Energy-Based Fragmentation Approach and Its Applications to Macromolecules and Molecular Aggregates. *Acc. Chem. Res.* **2014**, *47*, 2712−2720.
(78) Yang, Z.; Hua, S.; Hua, W.; Li, S. Low-Lying Structures and Stabilities of Large Water Clusters: Investigation Based on the Combination of the AMOEBA Potential and Generalized Energy-Based Fragmentation Approach. *J. Phys. Chem. A* **2010**, *114*, 9253−9261.

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

(79) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479.
(80) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular Interactions From a Natural Bond Orbital, Donor-Acceptor Viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.
(81) Weinhold, F.; Landis, C. R. *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*; Cambridge University Press: Cambridge, 2003.
(82) Lebedev, V. I.; Skorokhodov, L. Quadrature formulas of orders 41, 47 and 53 for the sphere. *Russian Acad. Sci. Dokl. Math.* **1992**, *45*, 587−592.
(83) Gräfenstein, J.; Cremer, D. Efficient DFT integrations by locally augmented radial grids. *J. Chem. Phys.* **2007**, *127*, 164113.
(84) Kazimirski, J.; Buch, V. Search for Low Energy Structures of Water Clusters $(H_2O)_n$, n = 20−22, 48, 123, and 293. *J. Phys. Chem. A* **2003**, *107*, 9762−9775.
(85) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, 1994.
(86) Cremer, D.; Kraka, E. A Description of the Chemical Bond in Terms of Local Properties of Electron Density and Ener. *Croat. Chem. A* **1984**, *57*, 1259−1281.
(87) Cremer, D.; Kraka, E. Chemical Bonds without Bonding Electron Density - Does the Difference Electron Density Analysis Suffice for a Description of the Chemical Bond? *Angew. Chem., Int. Ed. Engl.* **1984**, *23*, 627−628.
(88) Kraka, E.; Cremer, D. In *Theoretical Models of Chemical Bonding. The Concept of the Chemical Bond*; Maksic, Z. B., Ed.; Springer Verlag: Heidelberg, 1990; Vol 2, p 453.
(89) Zou, W.; Cremer, D. C_2 in a Box: Determining its Intrinsic Bond Strength for the $X^1\Sigma_g^+$ Ground State. *Chem. - Eur. J.* **2016**, *22*, 4087−4099.
(90) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations. The Theory of Infrared and Raman Vibrational Spectra*; McGraw-Hill: New York, 1955.
(91) Zou, W.; Cremer, D. Properties of Local Vibrational Modes: The Infrared Intensity. *Theor. Chem. Acc.* **2014**, *133*, 1451.
(92) Decius, J. Compliance Matrix and Molecular Vibrations. *J. Chem. Phys.* **1963**, *38*, 241.
(93) Zou, W.; Kalescky, R.; Kraka, E.; Cremer, D. Relating Normal Vibrational Modes To Local Vibrational Modes Benzene and Naphthalene. *J. Chem. Phys.* **2012**, *137*, 084114.
(94) Hayward, R. J.; Henry, B. R. A General Local-Mode Theory for High Energy Polyatomic Overtone Spectra and Application to Dichloromethane. *J. Mol. Spectrosc.* **1975**, *57*, 221−235.
(95) Henry, B. R. The Local Mode Model and Overtone Spectra: A Probe of Molecular Structure and Conformation. *Acc. Chem. Res.* **1987**, *20*, 429−435.
(96) Rong, Z.; Henry, B. R.; Robinson, T. W.; Kjaergaard, H. G. Absolute Intensities of CH Stretching Overtones in Alkenes. *J. Phys. Chem. A* **2005**, *109*, 1033−1041.
(97) Kraka, E.; Larsson, J. A.; Cremer, D. In *Computational Spectroscopy: Methods, Experiments and Applications*; Grunenberg, J., Ed.; Wiley: New York, 2010; p 105.
(98) Badger, R. M. A Relation Between Internuclear Distances and Bond Force Constants. *J. Chem. Phys.* **1934**, *2*, 128−131.
(99) Kalescky, R.; Kraka, E.; Cremer, D. Identification of the Strongest Bonds in Chemistry. *J. Phys. Chem. A* **2013**, *117*, 8981−8995.
(100) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910−8922.
(101) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253−287.
(102) Uberuaga, B. P.; Anghel, M.; Voter, A. F. Synchronization of trajectories in canonical molecular-dynamics simulations: Observation, explanation, and exploitation. *J. Chem. Phys.* **2004**, *120*, 6363−6374.
(103) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(104) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637−649.
(105) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(106) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(107) Giguere, P. A. Bifurcated Hydrogen Bonds in Water. *J. Raman Spectrosc.* **1984**, *15*, 354−358.
(108) Tukey, J. W. *Exploratory Data Analysis*, 1st ed.; Addison Wesley: 1977.
(109) Kraka, E.; Zou, W.; Filatov, M.; Gräfenstein, J.; Izotov, D.; Gauss, J.; He, Y.; Wu, A.; Polo, V.; Olsson, L. et al. *COLOGNE16*; 2016.
(110) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. *Gaussian 09 Revision B.01*; Gaussian Inc.: Wallingford, CT, 2010.
(111) Case, D.; Berryman, J.; Betz, R.; Cerutti, D.; Cheatham, T., III; Darden, T.; Duke, R.; Giese, T.; Gohlke, H. et al. *AMBER 2015*; University of California: San Francisco, 2015.
(112) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem.* **2012**, *33*, 580−592.
(113) Ohno, K.; Okimura, M.; Akai, N.; Katsumoto, Y. The effect of cooperative hydrogen bonding on the OH stretching-band shift for water clusters studied by matrix-isolation infrared spectroscopy and density functional theory. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3005−3014.
(114) Iwata, S. Energy analysis of weak electron-donorÅ-acceptor complexes and water clusters with the perturbation theory based on the locally projected molecular orbitals: charge-transfer and dispersion terms. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7787−7794.
(115) Iwata, S.; Bandyopadhyay, P.; Xantheas, S. S. Cooperative Roles of Charge Transfer and Dispersion Terms in Hydrogen-Bonded Networks of $(H_2O)_n$, n = 6, 11, and 16. *J. Phys. Chem. A* **2013**, *117*, 6641−6651.
(116) Wang, B.; Jiang, W.; Dai, X.; Gao, Y.; Wang, Z.; Zhang, R.-Q. Molecular orbital analysis of the hydrogen bonded water dimer. *Sci. Rep.* **2016**, *6*, 22099.
(117) Clark, T.; Politzer, P.; Murray, J. S. Correct electrostatic treatment of noncovalent interactions: the importance of polarization. *WIRES Comp. Mol. Sci.* **2015**, *5*, 169−177.
(118) Politzer, P.; Murray, J. S.; Clark, T. Mathematical modeling and Physical Reality in Non-covalent Interactions. *J. Mol. Model.* **2015**, *21*, 52−61.
(119) Politzer, P.; Murray, J. S.; Clark, T. σ-Hole Bonding: A Physical Interpretation. *Top. Curr. Chem.* **2014**, *358*, 19−42.
(120) Dominikowska, J.; Jablonski, M.; Palusiak, M. Feynman force components: basis for a solution of the covalent vs. ionic dilemma. *Phys. Chem. Chem. Phys.* **2016**, *18*, 25022−25026.
(121) Kraka, E.; Cremer, D. In *Theoretical Models of Chemical Bonding. The Concept of the Chemical Bond*; Maksic, Z., Ed.; Springer Verlag: Heidelberg, Germany, 1990; Vol 2, pp 453−542.
(122) Porterfield, W. W. *Inorganic Chemistry, A Unified Approach*; Academic Press: San Diego, 1993.
(123) Park, C.-Y.; Kim, Y.; Kim, Y. The multi-coefficient correlated quantum mechanical calculations for structures, energies, and harmonic frequencies of HF and H2O dimers. *J. Chem. Phys.* **2001**, *115*, 2926.
(124) Cherng, B.; Tao, F.-M. Formation of ammonium halide particles from pure ammonia and hydrogen halide gases: A theoretical

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

study on small molecular clusters 2, 4; X = F, Cl, Br). *J. Chem. Phys.* **2001**, *114*, 1720.
(125) Steiner, T. The hydrogen bond in the solid state. *Angew. Chem., Int. Ed.* **2002**, *41*, 48.
(126) Nibbering, E.; Elsaesser, T. Ultrafast Vibrational Dynamics of Hydrogen Bonds in the Condensed Phase. *Chem. Rev.* **2004**, *104*, 1887.
(127) Diken, E.; Headrick, J.; Roscioli, J.; Bopp, J.; Johnson, M.; McCoy, A. Fundamental Excitations of the Shared Proton in the $H_3O_2^-$ and $H_5O_2^+$ Complexes. *J. Phys. Chem. A* **2005**, *109*, 1487.
(128) Diken, E.; Headrick, J.; Roscioli, J.; Bopp, J.; McCoy, A.; Huang, X.; Carter, S.; Bowman, J. Argon Predissociation Spectroscopy of the $OH^-{\cdot}H_2O$ and $Cl^- \cdot H_2O$ Complexes in the 1000−1900 cm^{-1} Region: Intramolecular Bending Transitions and the Search for the Shared-Proton Fundamental in the Hydroxide Monohydrate. *J. Phys. Chem. A* **2005**, *109*, 571.
(129) Brindle, C.; Chaban, G.; Gerber, R.; Janda, K. Anharmonic vibrational spectroscopy calculations for (NH_3) (HF) and (NH_3) (DF): fundamental, overtone, and combination transitions. *Phys. Chem. Chem. Phys.* **2005**, *7*, 945.
(130) Janeiro-Barral, P.; Mella, M. Study of the structure, energetics, and vibrational properties of small ammonia clusters $(NH_3)_n$ (n = 2−5) using correlated ab initio methods. *J. Phys. Chem. A* **2006**, *110*, 11244.
(131) Slipchenko, M.; Sartakov, B.; Vilesov, A.; Xantheas, S. Study of NH Stretching Vibrations in Small Ammonia Clusters by Infrared Spectroscopy in He Droplets and ab Initio Calculations. *J. Phys. Chem. A* **2007**, *111*, 7460.
(132) Elsaesser, T. Two-dimensional infrared spectroscopy of intermolecular hydrogen bonds in the condensed phase. *Acc. Chem. Res.* **2009**, *42*, 1220.
(133) Heisler, I.; Meech, S. Low-Frequency Modes of Aqueous Alkali Halide Solutions: Glimpsing the Hydrogen Bonding Vibration. *Science* **2010**, *327*, 857.
(134) Zhang, X.; Zeng, Y.; Li, X.; Meng, L.; Zheng, S. Comparison in the complexes of oxygen-containing sigma-electron donor with hydrogen halide and dihalogen molecules. *J. Mol. Struct.: THEOCHEM* **2010**, *950*, 27−35.
(135) Li, X.-Z.; Walker, B.; Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6369.
(136) de Lima, N.; Ramos, M. A theoretical study of the molecular structures and vibrational spectra of the $N_2O{\cdot}(HF)_2$. *J. Mol. Struct.* **2012**, *1008*, 29.
(137) Grabowski, S. Red- and Blue-Shifted Hydrogen Bonds: the Bent Rule from Quantum Theory of Atoms in Molecules Perspective. *J. Phys. Chem. A* **2011**, *115*, 12789.
(138) Mpemba, E.; Osborne, D. Cool. *Phys. Educ.* **1969**, *4*, 172−5.
(139) Huang, Y.; Zhang, X.; Ma, Z.; Zhou, Y.; Zheng, W.; Zhou, J.; Sun, C. Q. Hydrogen-bond relaxation dynamics: Resolving mysteries of water ice. *Coord. Chem. Rev.* **2015**, *285*, 109−165.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on December 20, 2016 with errors in the text. The corrected version was reposted on December 21, 2016.

DOI: 10.1021/acs.jctc.6b00735
J. Chem. Theory Comput. 2017, 13, 55−76

Appendix B

Paper on Push-Pull Effect in H-Bonding

Chemical Physics Letters 685 (2017) 251–258

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Research paper

Strengthening of hydrogen bonding with the push-pull effect ☆

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

[a] Computational and Theoretical Chemistry Group (CATCO), Department of Chemistry, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75275-0314, United States
[b] Institute of Modern Physics, Northwest University, Xi'an, Shaanxi 710069, PR China

ARTICLE INFO

Article history:
Received 13 June 2017
In final form 26 July 2017
Available online 27 July 2017

ABSTRACT

Theoretical studies of hydrogen-bonding based on cluster models tend to overlook the peripheral monomers which are influential. By revisiting thirteen hydrogen-bonded complexes of H_2O, HF and NH_3, the "push-pull" effect is identified as a general mechanism that strengthens a hydrogen bond. Enhanced $Lp(X) \rightarrow \sigma^*(X'-H)$ charge transfer is proved to be the core of the "push-pull" effect. The charge transfer can convert an electrostatic hydrogen bond into a covalent hydrogen bond.

1. Introduction

Hydrogen bonding (H-bonding) is one of the most important intermolecular forces found in condensed phases, especially in the case of liquid water. It decides on various macroscopic properties including density, boiling point and melting point. In the recent years, various quantum chemical studies attempted to look into the H-bonding by simulating real systems with the help of cluster models consisting of 2–20 monomers [1–11]. Various computational methods have been used for the analysis of H-bonds including binding energy calculation [1,6–8,10], energy decomposition analyses [12] or the quantum theory of atoms in molecules and the natural bond orbital (NBO) analysis [13]. However, a deeper analysis of how peripheral monomers around a dimer influence the targeted H-bond is often missing.

Experimental studies on the OH stretching frequency shift in small water clusters, supported by DFT calculations, have suggested that the formation and strength of a particular H-bond is influenced by cooperative effects from peripheral H-bonds [14,15]. However, a caveat is appropriate. Experimentally or theoretically derived normal vibrational modes and force constants are delocalized because of electronic and mass-coupling [16]. Therefore, are not suited as direct measure of bond strength [17]. We present in this work a reliable descriptor of the intrinsic H-bond strength based on local vibrational modes, first introduced by Konkoli and Cremer [16,18]. These local modes have been proved to be the local equivalent of the delocalized normal vibrational modes via an adiabatic connection scheme (ACS), in which a one-to-one relationship has been proved between $3N-L$ normal modes and a non-redundant set of $3N-L$ local modes (N: number of atoms; L: number of translations and rotations) [18]. The local stretching force constant k^a is the appropriate tool to describe the intrinsic bond strength of any chemical bonding situation [19,17], including non-covalent bonding [20,21] like hydrogen bonds [22,23]. Since k^a is directly related to the electronic structure of a molecule, it absorbs any neighboring influences on the bond in question, such as cooperative effects [23].

In this work, we have studied thirteen clusters made up of monomers of H_2O, HF and NH_3 molecules. A new and generally applicable mechanism which we name as *push-pull* effect is identified to strengthen H-bonds. This push-pull effect can have a strong impact on the H-bonding mechanism in several cases. The objectives of this work are to answer the following questions. (i) What is push-pull effect in H-bonding? (ii) How can the push-pull effect strengthen a H-bond? (iii) How is the push-pull effect related to the charge transfer from the H-bond acceptor to the H-bond donor? (iv) To what extent can the push-pull effect change the nature of a H-bond with regard to its covalent or electrostatic character?

The computational methods used in this work are described in the second section. The third section presents the results and discussion, while conclusions are made in the final section.

2. Computational methods

Optimized geometries of all clusters investigated in this work (see Fig. 1) and their vibrational frequencies as well as normal modes were calculated using the ωB97X-D density functional. This functional was chosen because it describes non-covalent interactions in a reliable way taking care of dispersion and other van der Waals interactions [24–27]. Pople's triple zeta basis set 6-311 ++G(d,p) with diffuse functions for O, F, N and H atoms was used.

☆ In Memoriam of Dieter Cremer.

* Corresponding author.
E-mail address: ekraka@smu.edu (E. Kraka).

http://dx.doi.org/10.1016/j.cplett.2017.07.065
0009-2614/

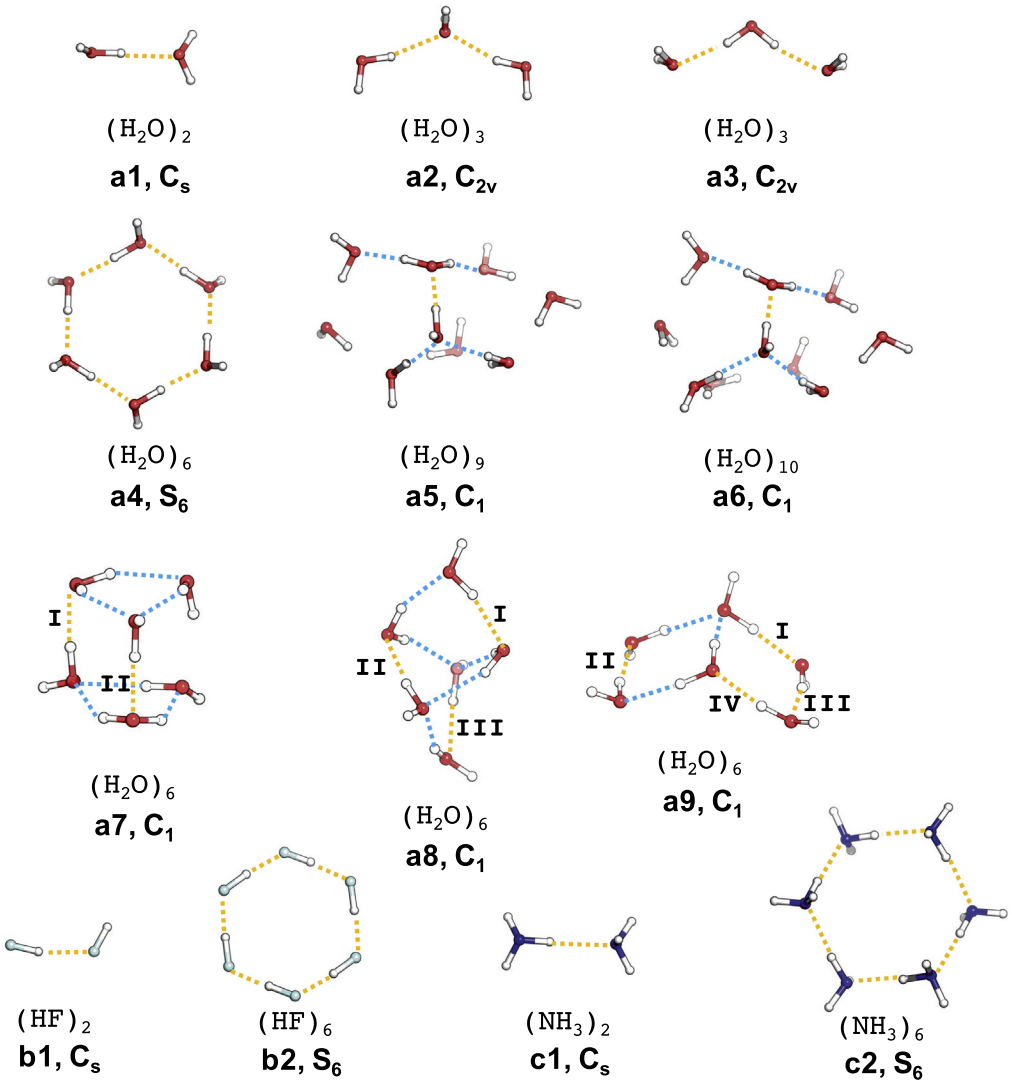


Fig. 1. Schematic presentation of the geometries of cluster **a1**–**c2**. Red, cyan, blue and white spheres stand for oxygen, fluorine, nitrogen and hydrogen atoms respectively. Yellow dashed lines indicate the hydrogen bonds described in Table 1. Blue dashed lines in **a5**–**a9** are used to represent the peripheral hydrogen bonds responsible for push-pull mechanism. The components, labels and symmetry of each cluster is also given. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This basis set provides an accurate description for these molecular clusters [28–30]. The DFT calculations were conducted using a pruned (99,590) UltraFine integration grid [31,32] and the geometry optimization was imposed with a tight convergence criterion to guarantee the accurate calculation of the Hessian matrix which was used for adiabatic local mode analysis.

The geometries of two additional small water clusters (hexamer and tetramer) shown in Fig. 2 were constructed based on the geometry of water dimer[1] in **a1**, constrained by C_s symmetry.

The electron density was calculated at the ωB97X-D/6-311G++(d,p) level of theory. The charge transfer was analyzed based on calculated NPA charges [33,34].

The intrinsic bond strength of the H-bond was determined by the local H-bond stretching force constant k^a [22,17] derived from the corresponding local vibrational mode [16].

The covalent character of the H-bonds was estimated via the calculation of the delocalization energy ΔE_{del}, which can be understood as the stabilization energy due to the charge transfer from one or more (if present) lone pair orbital(s) of the X atom of the hydrogen bond acceptor to a σ^* antibonding X′-H orbital of the hydrogen bond donor through the overlap between the two orbitals. The amplitude of ΔE_{del} was characterized for a given X...X′-H interaction by a second order perturbation theory analysis of the Fock matrix in the NBO basis [34].

The nature of the H-bond was further characterized by the local energy density H_b at the (3,-1) bond critical point $\mathbf{r_b}$(BCP) [35,36]. The Cremer-Kraka criteria were applied to quantitatively identify the covalent bonding character: (i) A BCP and zero-flux surface must exist between the two atoms, for which chemical bonding is expected (*necessary condition*). (ii) The local energy density H_b must be less than zero in the case of covalent bonding (*sufficient condition*). Positive values of H_b indicates that the bond in question is dominated by electrostatic interactions [37]. This descriptor has been extensively used in the studies of chemical bonds to determine whether a bond is covalent or non-covalent [38,39], including pnicogen bonds, [21] halogen bonds, [20] and hydrogen bonds [1,22,23].

Apart from characterizing the charge transfer within the H-bond dimer using the NBO analysis, we also calculated the difference density distribution $\Delta\rho(A\ldots B,\mathbf{r}) = \rho(AB,\mathbf{r}) - (\rho(A,\mathbf{r}) + \rho(B,\mathbf{r}))$[40], to describe the formation of the complex *AB* where a hydrogen bond is found between monomers *A* and *B* with regard

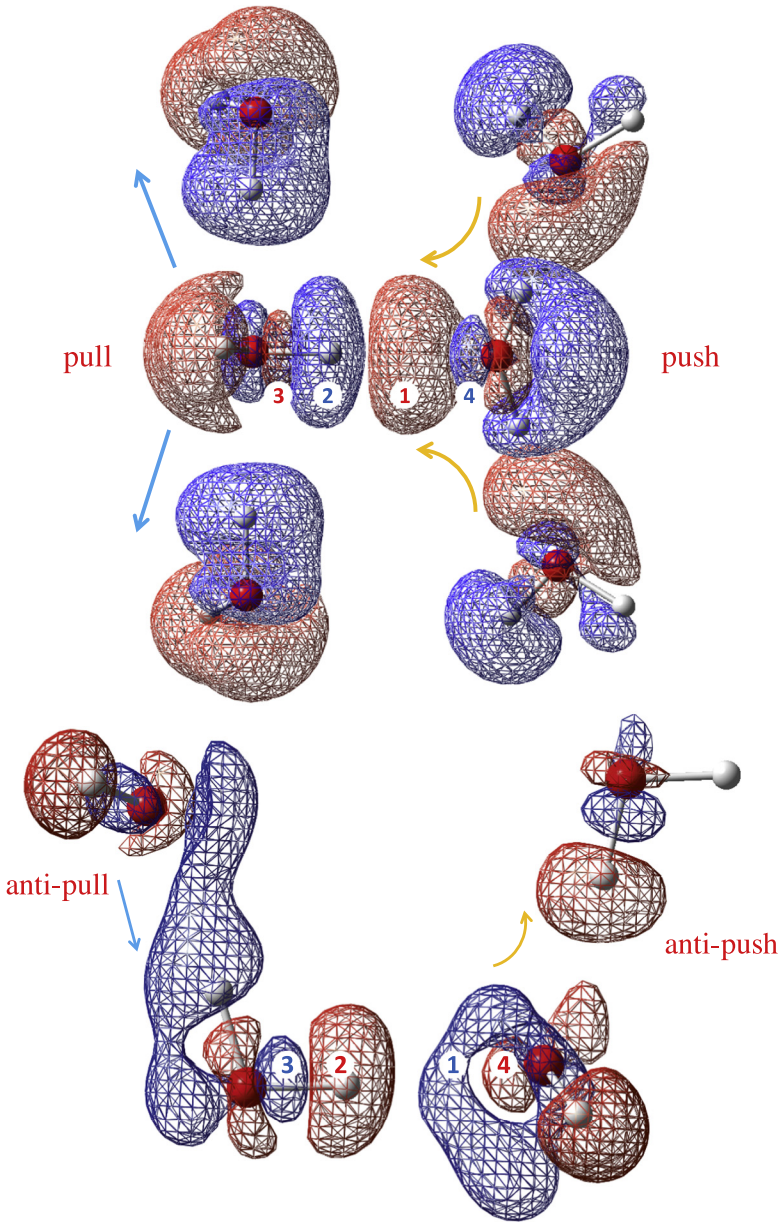


Fig. 2. Electron difference density distribution $\Delta\rho$ shown as the ± 0.0002 e/Bohr^3 isosurface. Blue surface indicates electron depletion, red is increased electron density region. Top: 20–02 type H-bond in the center with a push-pull effect. Bottom: 01–10 type H-bond in the bottom with an antipush-antipull effect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to the change of the electron density distribution. The push-pull effect caused by peripheral molecules can be made visible with the help of the above formula. Besides taking only one monomer as the H-bond donor(D) and one as acceptor(A) for the hydrogen bond in question, we defined a generalized H-bond donor (D^*) and acceptor (A^*) which include D and A respectively but can have more than one monomer. Within the complex of D^* or A^*, the monomers are connected via peripheral hydrogen bonds, which need to be distinguished from the targeted H-bond in the center. In this way, the influence of these peripheral monomers on the central DA dimer complex in H-bonding can be assessed from $\Delta\rho(push-pull) = \Delta\rho(D^*\ldots A^*,\mathbf{r}) - \Delta\rho(D\ldots A,\mathbf{r}) = \rho(D^*A^*,\mathbf{r}) - \rho(D^*,\mathbf{r}) - \rho(A^*,\mathbf{r}) - \rho(DA,\mathbf{r}) + \rho(D,\mathbf{r}) + \rho(A,\mathbf{r})$ [23]. As an example, the molecular cluster **a4** in Fig. 4 has its polarization effect on the bottom hydrogen bond arising from the four peripheral waters visualized accordingly.

All vibrational modes as well as the local mode analysis were carried out with the program package COLOGNE2017 [41], whereas for the DFT calculations, the program package Guassian09 [42] was used. Difference densities and NBO orbitals were plotted with the Multiwfn program [43].

3. Results and discussion

Fig. 1 summarizes the 13 molecular clusters investigated in this work labeled with **a1**–**a9**, [44] **b1**–**b2** and **c1**–**c2**. In each cluster, we have one target H-bond in question with regard to the push-pull effect shown in yellow dashed lines. For clusters with high symmetry, we have several symmetry equivalent target H-bonds [3–5,2]. Properties of the target H-bond are summarized in Table 1 including the notation of H-bond type, bond length *R*, local

Table 1
Properties of the target H-bond of clusters **a1**–**a9**, **b1**–**b2** and **c1**–**c2**.[a]

Cluster #	Complex	Type	R	k^a	ΔE_{del}	H_b
a1	$(H_2O)_2$	00–00	1.9158	0.199	9.67	+0.002299
a2	$(H_2O)_3$	01–00	1.9659	0.168	8.15	+0.002352
a3	$(H_2O)_3$	00–10	1.9666	0.165	8.11	+0.002312
a9-II	$(H_2O)_6$	10–01	1.7969	0.265	16.86	+0.000383
a4	$(H_2O)_6$	10–01	1.7261	0.359	22.51	−0.001673
a9-III	$(H_2O)_6$	10–01	1.7156	0.344	24.29	−0.002485
a8-I	$(H_2O)_6$	10–02	1.8080	0.251	18.05	−0.000213
a9-IV	$(H_2O)_6$	10–02	1.7161	0.344	25.13	−0.002759
a8-III	$(H_2O)_6$	20–01	1.7544	0.297	20.52	−0.001022
a9-I	$(H_2O)_6$	20–01	1.6997	0.328	25.98	−0.003107
a7-II	$(H_2O)_6$	20–02	1.7530	0.311	22.17	−0.001500
a8-II	$(H_2O)_6$	20–02	1.6831	0.328	30.71	−0.005091
a7-I	$(H_2O)_6$	20–02	1.6692	0.366	32.17	−0.005725
a5	$(H_2O)_9$	20–02	1.6106	0.395	38.67	−0.008851
a6	$(H_2O)_{10}$	20–02	1.6103	0.399	38.57	−0.008770
b1	$(HF)_2$	00–00	1.8277	0.182	8.38	+0.002529
b2	$(HF)_6$	10–01	1.5524	0.420	30.93	−0.004162
c1	$(NH_3)_2$	00–00	2.2080	0.117	6.40	+0.001782
c2	$(NH_3)_6$	10–01	2.0847	0.178	11.83	+0.001361

[a] Hydrogen bond length R in Å, local stretching force constant k^a in mdyn/Å, delocalization energy $\Delta E_{del} = \Delta E_{lp(X)\rightarrow\sigma^*(X'-H)}$ in kcal/mol (X = X′ = O, F or N), local energy density H_b in Hartree/Bohr^3.

stretching force constant k^a, the delocalization energy ΔE_{del} and local energy density H_b. Fig. 2 illustrates the push-pull effect using the electron density difference (EDD) of the H-bond in water. The opposite effect, coined as antipush-antipull effect is also shown. In Fig. 3, the NBOs responsible for the dominating $lp(X) \rightarrow \sigma^*(X'-H)$ charge transfer leading to stabilization are shown for the H-bond donor and also acceptor within H-bond dimer for **a1**, **b1** and **c1**. In Fig. 4, the push-pull effect for a specific H-bond is visualized via electron density difference maps for the hexamer rings of H_2O, HF and NH_3(clusters **a4**, **b2** and **c2** respectively).

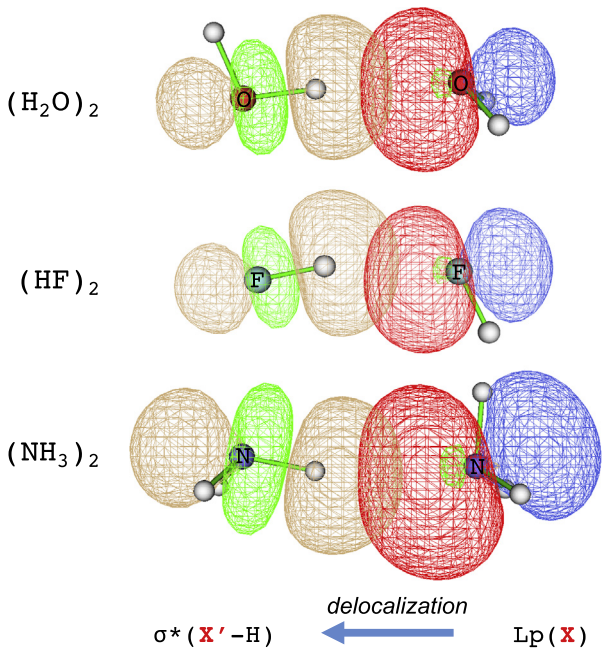


Fig. 3. Charge transfer in the formation of a hydrogen bond implies electron delocalization of lone pair electrons in Lp(X) orbital (red and blue lobes on the right as H-bond acceptor) into the unoccupied σ^*(X′-H) antibonding orbital (brown and green lobes on the left as H-bond donor). For H_2O and HF, only one of the lone pair orbitals is shown here as the dominating one with regard to ΔE_{del}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1. Different types of H-bonds

We showed in previous work [23] how the properties of a H-bond donor and acceptor are affected by the surrounding molecules. We used a 4-digit notation ($i_a j_d - k_a l_d$) in order to distinguish between different kinds of H-bonding. In this work, we will use the same notation.

We use 00–00 to denote the H-bond in a dimer structure. If the donor water accepts 1 H-bond from other water molecules, we add 1 to i_a and if the donor water donates 1 extra H-bond to a peripheral water which needs to be distinguished from the acceptor water of 00–00 H-bond, we add 1 to j_d. This rule also applies to the acceptor water as the ($i_a j_d$) part is for the donor while acceptor water is determined by the ($k_a l_d$) part.

We have found in our previous studies on the water clusters that for some specific types of H-bond, the intrinsic H-bond strength based on local stretching force constant k^a is remarkably higher or lower than for some other types [23]. Similar observations were also made in this work (see Table 1).

3.2. Definition of the push-pull effect

In order to define the push-pull effect in H-bonding, three major ingredients are needed, including (i) the way how the H-bond donor and acceptor molecules interact with peripheral molecules via peripheral H-bonding, (ii) the electron density difference $\Delta\rho$ (*push-pull*) showing the influence of peripheral molecules on the central H-bond contained within a dimer structure and (iii) the intrinsic H-bond strength characterized by k^a. This definition of the "push-pull" effect must not be confused with the originally use of the term "push-pull" effect proposed by Kleinpeter [45] for the description of covalent π bonding.

If a specific H-bond of a molecular complex AB connected via H-bonding is to be studied, it is insufficient if just to focus on the A and B monomer. It is necessary to include those molecules that directly interact with the donor and acceptor as these peripheral molecules may lead to a significant change in the electronic structure of the H-bond in question via polarization.

For the $(H_2O)_6$ cluster shown on top of Fig. 2, the H-bond in the central dimer can be classified as 20–02 type, namely the H-bond donor water on the left hand side accepts 2 external H-bonds while the H-bond acceptor on the right hand side donates 2 external

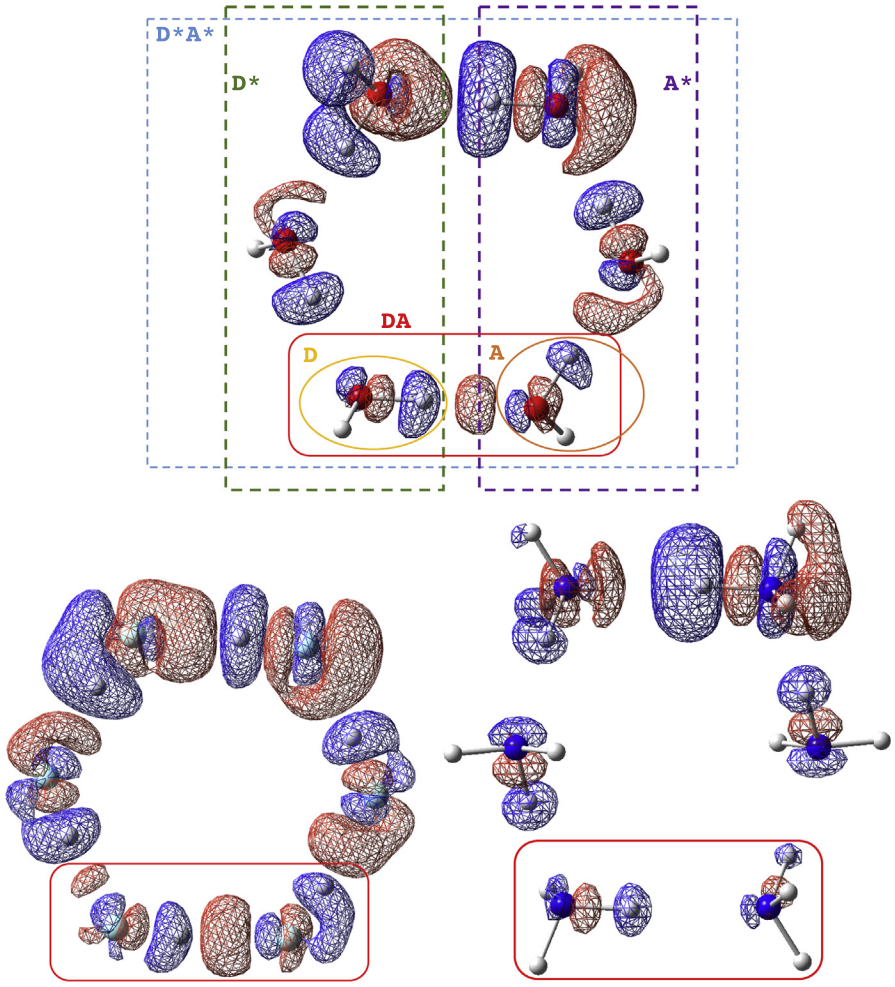


Fig. 4. Electron density difference maps of $\Delta\rho(push-pull)$ for hexamer rings. The target H-bond is identified by the red box. Upper: **a4**; Lower left: **b2**; Lower right: **c2**. Red mesh surfaces depict a density increase while blue mesh surfaces for density decrease. Isovalue for surfaces is $\pm$0.001 e/Bohr^3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

H-bonds to other water molecules. The electron density difference distribution is calculated by subtracting from the density of the whole hexamer that of the trimer on the donor and acceptor side as well as that of the dimer in the center then adding that of the donor and acceptor water of 20–02 H-bond (short notation: 6–2 $\times$ 3–2 + 1 + 1). This EDD map shows how the 4 peripheral water molecules polarize the H-bond in the central water dimer. On the acceptor side, these two peripheral water molecules which accept H-bonds have their lone pairs pointing to this acceptor water. These lone pairs increase the electron density distribution in the central H-bond region by polarization(labeled as ① in red color). We name this polarization as "push". On the donor side of 20–02 H-bond, two additional water molecules donating extra H-bonds further polarize the electron density. For this 20–02 H-bond, the donor OH covalent bond has a decreased density around the H atom(② in blue color), then it has an increased density towards the donor O atom (③ in red color). These two peripheral waters withdraw electron density from the central H-bond region, so we say they can "pull" electron density. When we combine the polarization effect both on the acceptor side and on the donor side, the comprehensive effect is summarized as "push-pull" effect.

We note that between the donor oxygen and the acceptor oxygen within this 20–02 H-bond, there is a region in which electron density is decreased (④ in blue color), which we will discuss below.

For the assurance of the validity of the definition of the push-pull effect, we checked other water clusters, for example the tetramer shown on the bottom of Fig. 2. Here we reverted all polarization effect on the central targeted H-bond in 20–02 simply by changing it into the 01–10 H-bonding situation where the donor water accepts no H-bonds but needs to donate 1 extra H-bond and the acceptor water accepts 1 extra H-bond while donating no H-bonds. As a result, the acceptor water has to direct its second lone pair to the water from which it accepts another H-bond, so the central H-bond region has a decreased electron density (① in blue color). Such a polarization can be coined as the opposite to "push" as "anti-push" effect. On the other hand, the water to which the donor water's extra H-bond points has its lone pair electrons oriented to one of the OH bonds of the donor water. This leads to the density increase in region ② and decrease in region ③. Such a distribution pattern is opposite to its counterpart in 20–02, thus we call this "anti-pull" effect.

3.3. Intrinsic H-bond strength, change transfer and covalency

Table 1 provides a summary of 15 different H-bonds found in 9 clusters of H_2O, in addition to 4 H-bonds in clusters of HF and NH_3.

Taking the 00–00 H-bond in the dimer **a1** as the reference, we find that the 01–00 and 00–10 H-bonds in **a2** and **a3** resulting from an anti-pull and anti-push effect respectively, are characterized by

longer bond lengths, decreased intrinsic bond strengths, smaller delocalization energies and more positive H_b values. In contrast, the other H-bonds of the type X0-0Y (X,Y > 0) for which the push-pull effect plays a dominant role have shorter and stronger bonds. The charge transfer is more pronounced and they are of covalent character, as reflected by negative H_b values. There is one exception, the relatively weak 10–01 H-bond in **a9-II** has a small positive H_b value, see Table 1.

While all of the H-bonds supported by peripheral push- and pull H-bonds are stronger than the 00–00 H-bond without any push–pull effect, the H-bonds of type 10–02, 20–01 or 20–02 are not necessarily stronger than the 10–01 H-bonds, even though they seem to be strengthened by a larger number of the push-pull molecules. The reason for this observation is that the magnitude of the push and pull effect varies from one H-bond to another. A closer examination of the results reveals that stronger pushing and pulling H-bonds can render a significant increase in H-bond strength, while weaker pushing and pulling H-bonds offer limited increase of intrinsic bond strength. The strongest H-bonds with local stretching force constant k^a values up to 0.39 mdyn/Å are found for the 20–02 type. This implies that there exists an upper bound with regard to the intrinsic bond strength for each push-pull type, given that the number of H-bonds studied is large enough to cover as many H-bonding possibilities as possible.

The NBO analysis carried out by Reed et al. [46] suggests that charge transfer plays an important role in the formation of H-bonds. In our work, the results of the NBO analysis show that the most stabilizing factor within a H-bond dimer complex is the interaction between the lone pair orbital(s) of the H-bond acceptor $Lp(X)$ and the σ^* anti-bonding orbital of the $X'-H$ covalent bond with the H-bond donor pointing to the acceptor atom X. Fig. 3 shows the shape of the NBOs of $Lp(X)$ and $\sigma^*(X'-H)$ from which the dominating stabilization in H-bonding is established. These NBO plots are consistent for the dimer of H_2O, HF and NH_3. Furthermore, if one takes a closer look at the space between atoms X and X', in total 4 lobes can be identified. The four different regions, labeled 1–4 in Fig. 2 directly match the four lobes shown in Fig. 3 from left to right as green-③, brown-②, red-① and then green-④.

It is interesting to note that for H-bonds in water, the variation in the delocalization energy ΔE_{del} characterizing the charge transfer from $Lp(X)$ to $\sigma^*(X'-H)$ is consistent with the change of local stretching force constant k^a; the more charge transfer a H-bond has, the stronger is this H-bond, see Table 1. However, there are two expections, complexes **a9-II** and **a8-I**. They possess the weakest H-bonds, (k^a values of 0.265 and 0.251 mdyn/Å respectively) compared to the other push-pull H-bonds, (k^a values in the range of 0.297–0.399 mdyn/Å). The H_b values of these two H-bonds are less negative than for the other push-pull H-bonds. The 10–01 H-bond in **a9-II** is a borderline case with a small positive H_b value. Therefore, according to the Cremer-Kraka criteria [37,38], this H-bond is dominated by classical electrostatic interactions.

The electrostatic interaction in H-bonding has been a controversial topic since the first recognition of the H-bond phenomenon [47]. For example, Weinhold and co-workers proposed that the driving force of the formation of a H-bond is charge transfer, while classical electrostatic forces as well as dispersion forces are of minor importance [48]. They even discussed the "anti-electrostatic hydrogen bond" [49], in which classical electrostatic forces are destabilizing and only charge transfer plays the leading role.

We clearly identify the H-bonds in **a1**–**a3** as dominated by electrostatic interactions rather than by charge transfer. (i) They are characterized by positive H_b values, indicating the increased weight of electrostatic interactions over covalent character (see Table 1); (ii) The charge transfer is diminished (characterized by decreased ΔE_{del} values) via the anti-push or anti-pull effect. For the 10–01 H-bond in **a9-II**, although the charge transfer is increased compared to the H-bond in water dimer, the covalent contribution from charge transfer is still not as important as the electrostatic forces. So that the local energy density remains positive, but its value is less positive than that for the 00–00 H-bond. When the delocalization energy is increased to 18.05 kcal/mol for the H-bond in **a8-I**, the local energy density immediately turns negative, indicating the dominance of charge transfer over classical electrostatic forces. The other push-pull H-bonds of water investigated in this work are of covalent nature as indicated by negative H_b values and dominated by charge transfer, as reflected by the data in Table 1. It is interesting to note that the local energy density H_b increases with increasing delocalization energy ΔE_{del}, and no exceptions are identified. This correlation (see Fig. 1 in Supporting Information) suggests that both, the local energy density H_b and the delocalization energy ΔE_{del} are reliable descriptors of charge transfer/covalency; where the local energy density in addition allows to distinguish between covalent and the electrostatic interaction.

Apart from the H-bonds in water, we have also studied the H-bonds in the HF and NH_3 clusters with regard to the push-pull effect. The dimer structures **a1**, **b1** and **c1** can be directly compared as well as the ring hexamer structures **a4**, **b2** and **c2** respectively for H_2O, HF and NH_3. The result shows that the H-bonding in the case of HF is most sensitive to the push-pull effect as it has the largest increase in bond strength and charge transfer. Furthermore, it adapts more covalent character when it is changed from 00–00 type to 10–01 type, in the same way as this happens for H_2O. In contrast the H-bonding in the case of NH_3 is least responsive to the push-pull effect, as it has the least increase in the local stretching force constant and the delocalization energy. It remains dominated by the classical electrostatic forces regardless of pushing and pulling H-bond neighbors. This is illustrated by the electron density difference plots for the hexamer rings of **a4**, **b2** and **c2** shown in Fig. 4. The HF hexamer ring **b2** has the largest electron density accumulation region, while such a region cannot be found in the case of NH_3 hexamer ring **c2**. These observations can be explained by the fact that the HF molecule has up to 3 lone pairs while H_2O and NH_3 has only 2 and 1 lone pairs respectively, so that HF molecule shows the most response to the push-pull effect in H-bonding.

4. Conclusions

In this work, we have for the first time defined the concept of the push-pull effect in H-bonds and examined this effect for 13 molecular complexes. This investigation has led to a series of interesting results.

(1) The characteristic EDD map of $\Delta\rho(push-pull)$ shows that the push-pull effect is a real and observable change in the electronic structure in H-bonding. The push-pull is a general effect and exists in H-bonded clusters constituted by monomers of the same type.
(2) The intrinsic H-bond strength is directly related to the push-pull effect. The larger is the push-pull effect, the larger is the H-bond strengthening. A H-bond can be weakened if it is under the influence of the anti-pull or the anti-push effect. The mixing of push-pull effect and the antipush-antipull effect may lead to an interesting competition and a variety of H-bond possibilities. This will be part of a future investigation.

(3) The strong correlation between the NBO diagrams ($Lp(X)$ and $\sigma^*(X'-H)$) and the $\Delta\rho(push-pull)$ EDD maps with regard to the central H-bond region indicates the origin of push-pull effect is the enhanced charge transfer which is responsible for the H-bonding stabilization. Furthermore, the more charge transfer a H-bond has, ideally the stronger this H-bond can be.

(4) Among the clusters studied in this work, the HF clusters are more responsive to the push-pull effect than H_2O and NH_3 clusters. H-bonds in NH_3 clusters are not influenced significantly with regard to their intrinsic bond strength and bonding nature. Similar observations are to be expected for the clusters of H_2S, HCl and PH_3. Work is in progress to demonstrate this.

(5) The push-pull effect can increase the covalent character of a H-bond via enhanced charge transfer. Increased covalency can change the nature of a H-bond depending on whether the covalent character can override the electrostatic character. If so, the H-bond will be more like a covalent bond although H-bonds are generally weaker than normal covalent bonds.

(6) The H-bonds in the dimer structure of H_2O, HF and NH_3 are dominated by electrostatic force. However, one can find H-bonds which are dominated by covalent character in the clusters of H_2O or HF.

(7) Contrary to Stones' claim that "it is a serious error to use NBO method in analyzing intermolecular interactions" [50], apparently the NBO analysis remains still a powerful tool for H-bonding studies in two ways: (i) the NBO diagrams shown in Fig. 3 is closely related to the EDD maps characterizing $\Delta\rho(push-pull)$; (ii) the variation in the quantitative measurement of the amplitude in charge transfer $\Delta E_{lp(X)\rightarrow\sigma^*(X'-H)}$ is consistent with the change of k^a and H_b.

Acknowledgement

This work was financially supported by the National Science Foundation, Grant 1464906. We thank SMU for providing computational resources.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cplett.2017.07.065.

References

[1] R. Kalescky, W. Zou, E. Kraka, D. Cremer, Local vibrational modes of the water dimer – comparison of theory and experiment, Chem. Phys. Lett. 554 (2012) 243–247.
[2] J. Kim, D. Majumdar, H.M. Lee, K.S. Kim, Structures and energetics of the water heptamer: comparison with the water hexamer and octamer, J. Chem. Phys. 110 (18) (1999) 9128–9134.
[3] M. Jelil, A. Abaydulla, Graph theoretical enumeration of topology-distinct structures for hydrogen fluoride clusters (HF)n (n ≤ 6), J. Chem. Phys. 143 (4) (2015) 044301.
[4] J. Friedrich, E. Perlt, M. Roatsch, C. Spickermann, B. Kirchner, Coupled cluster in condensed phase. Part i: Static quantum chemical calculations of hydrogen fluoride clusters, J. Chem. Theory Comput. 7 (4) (2011) 843–851.
[5] A. Malloum, J.J. Fifen, Z. Dhaouadi, S.G.N. Engo, N.-E. Jaidane, Structures and relative stabilities of ammonia clusters at different temperatures: DFT vs. ab initio, Phys. Chem. Chem. Phys. 17 (43) (2015) 29226–29242.
[6] R.M. Shields, B. Temelso, K.A. Archer, T.E. Morrell, G.C. Shields, Accurate predictions of water cluster formation, (H2O)n=2-10, J. Phys. Chem. A 114 (43) (2010) 11725–11737.
[7] J.C. Howard, G.S. Tschumper, Wavefunction methods for the accurate characterization of water clusters, Wiley Interdiscipl. Rev.: Comput. Mol. Sci. 4 (3) (2013) 199–224.
[8] S. Iwata, Analysis of hydrogen bond energies and hydrogen bonded networks in water clusters $(H_2O)_{20}$ and $(H_2O)_{25}$ using the charge-transfer and dispersion terms, Phys. Chem. Chem. Phys. 16 (23) (2014) 11310.
[9] I. Bakó, I. Mayer, Hierarchy of the collective effects in water clusters, J. Phys. Chem. A 120 (4) (2016) 631–638.
[10] N. Sahu, S.S. Khire, S.R. Gadre, Structures, energetics and vibrational spectra of $(H_2O)_{32}$ clusters: a journey from model potentials to correlated theory, Mol. Phys. 113 (19–20) (2015) 2970–2979.
[11] L. Huang, S.G. Lambrakos, A. Shabaev, N. Bernstein, L. Massa, Molecular analysis of water clusters: calculation of the cluster structures and vibrational spectrum using density functional theory, Compt. Rend. Chim. 18 (5) (2015) 516–524.
[12] M.J.S. Phipps, T. Fox, C.S. Tautermann, C.-K. Skylaris, Energy decomposition analysis approaches and their evaluation on prototypical protein-drug interaction patterns, Chem. Soc. Rev. 44 (10) (2015) 3177–3211.
[13] A. Shahi, E. Arunan, Hydrogen bonding, halogen bonding and lithium bonding: an atoms in molecules and natural bond orbital perspective towards conservation of total bond order, inter- and intra-molecular bonding, Phys. Chem. Chem. Phys. 16 (42) (2014) 22935–22952.
[14] K. Ohno, M. Okimura, N. Akai, Y. Katsumoto, The effect of cooperative hydrogen bonding on the OH stretching-band shift for water clusters studied by matrix-isolation infrared spectroscopy and density functional theory, Phys. Chem. Chem. Phys. 7 (16) (2005) 3005–3014.
[15] C. Tainter, Y. Ni, L.a. Shi, J. Skinner, Hydrogen bonding and OH-stretch spectroscopy in water: hexamer (cage), liquid surface, liquid, and ice, J. Phys. Chem. Lett. 4 (1) (2012) 12–17.
[16] Z. Konkoli, D. Cremer, A new way of analyzing vibrational spectra. i. Derivation of adiabatic internal modes, Int. J. Quant. Chem. 67 (1) (1998) 1–9.
[17] W. Zou, D. Cremer, C_2 in a box: determining its intrinsic bond strength for the $X^1\Sigma_g^+$ ground state, Chem. - Euro. J. 22 (12) (2016) 4087–4099.
[18] W. Zou, R. Kalescky, E. Kraka, D. Cremer, Relating normal vibrational modes to local vibrational modes with the help of an adiabatic connection scheme, J. Chem. Phys. 137 (2012) 084114.
[19] R. Kalescky, E. Kraka, D. Cremer, Identification of the strongest bonds in chemistry, J. Phys. Chem. A 117 (36) (2013) 8981–8995.
[20] V. Oliveira, E. Kraka, D. Cremer, Quantitative assessment of halogen bonding utilizing vibrational spectroscopy, Inorg. Chem. 56 (1) (2017) 488–502.
[21] D. Setiawan, E. Kraka, D. Cremer, Strength of the pnicogen bond in complexes involving group Va elements N, P, and As, J. Phys. Chem. A 119 (9) (2015) 1642–1656.
[22] M. Freindorf, E. Kraka, D. Cremer, A comprehensive analysis of hydrogen bond interactions based on local vibrational modes, Int. J. Quant. Chem. 112 (19) (2012) 3174–3187.
[23] Y. Tao, W. Zou, J. Jia, W. Li, D. Cremer, Different ways of hydrogen bonding in water - why does warm water freeze faster than cold water?, J Chem. Theory Comput. 13 (1) (2017) 55–76.
[24] K.S. Thanthiriwatte, E.G. Hohenstein, L.A. Burns, C.D. Sherrill, Assessment of the performance of DFT and DFT-D methods for describing distance dependence of hydrogen-bonded interactions, J. Chem. Theory Comput. 7 (1) (2011) 88–96.
[25] S. Kozuch, J.M.L. Martin, Halogen bonds: benchmarks and theoretical analysis, J. Chem. Theory Comput. 9 (4) (2013) 1918–1931.
[26] A. Bauzá, I. Alkorta, A. Frontera, J. Elguero, On the reliability of pure and hybrid DFT methods for the evaluation of halogen, chalcogen, and pnicogen bonds involving anionic and neutral electron donors, J. Chem. Theory Comput. 9 (11) (2013) 5201–5210.
[27] S. Scheiner, Extrapolation to the complete basis set limit for binding energies of noncovalent interactions, Comput. Theor. Chem. 998 (2012) 9–13.
[28] R. Ditchfield, W.J. Hehre, J.A. Pople, Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules, J. Chem. Phys. 54 (2) (1971) 724–728.
[29] P.C. Hariharan, J.A. Pople, The influence of polarization functions on molecular orbital hydrogenation energies, Theor. Chim. Acta 28 (3) (1973) 213–222.
[30] T. Clark, J. Chandrasekhar, G.W. Spitznagel, P.V.R. Schleyer, Efficient diffuse function-augmented basis sets for anion calculations. III. The 3–21+G basis set for first-row elements, Li-F, J. Comput. Chem. 4 (3) (1983) 294–301.
[31] J. Gräfenstein, D. Cremer, Efficient density-functional theory integrations by locally augmented radial grids, J. Chem. Phys. 127 (16) (2007) 164113.
[32] V.I. Lebedev, L. Skorokhodov, Quadrature formulas of orders 41, 47 and 53 for the sphere, Russian Acad. Sci. Dokl. Math. 45 (1992) 587–592.
[33] A.E. Reed, L.A. Curtiss, F. Weinhold, Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint, Chem. Rev. 88 (6) (1988) 899–926.
[34] F. Weinhold, C.R. Landis, Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective, Cambridge University Press, 2005.
[35] R.F.W. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, Oxford, 1994.
[36] T.A. Keith, TK Gristmill Software (aim.tkgristmill.com, Overland Park, KS, USA, 2011.
[37] D. Cremer, E. Kraka, A description of the chemical bond in terms of local properties of electron density and energy, Croat. Chem. Acta 57 (6) (1984) 1259–1281.
[38] D. Cremer, E. Kraka, Chemical bonds without bonding electron density - does the difference electron density analysis suffice for a description of the chemical bond?, Agnew Chem. Int. Ed. Engl. 23 (1984) 627–628.
[39] E. Kraka, D. Cremer, Chemical implication of local features of the electron density distribution, in: Z.B. Maksic (Ed.), Theoretical Models of Chemical

Bonding, The Concept of the Chemical Bond, vol. 2, Springer Verlag, Heidelberg, p. 453.
[40] W.H.E. Schwarz, P. Valtazanos, K. Ruedenberg, Electron difference densities and chemical bonding, Theor. Chim. Acta 68 (6) (1985) 471–506.
[41] 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. <http://www.smu.edu/catco>.
[42] 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, co workers, Gaussian 09 Revision A. 1, Gaussian Inc. , Wallingford, CT, 2010.
[43] T. Lu, F. Chen, Multiwfn: a multifunctional wavefunction analyzer, J. Comput. Chem. 33 (5) (2011) 580–592.
[44] C. Perez, M.T. Muckle, D.P. Zaleski, N.A. Seifert, B. Temelso, G.C. Shields, Z. Kisiel, B.H. Pate, Structures of cage, prism, and book isomers of water hexamer from broadband rotational spectroscopy, Science 336 (6083) (2012) 897–901.
[45] E. Kleinpeter, S. Klod, W.-D. Rudorf, Electronic state of push-pull alkenes: an experimental dynamic NMR and theoretical ab initio MO study, J. Organ. Chem. 69 (13) (2004) 4317–4329.
[46] A.E. Reed, F. Weinhold, L.A. Curtiss, D.J. Pochatko, Natural bond orbital analysis of molecular interactions: theoretical studies of binary complexes of HF, H_2O, NH_3, N_2, O_2, F_2, CO, and CO_2 with HF, H_2O, and NH_3, J. Chem. Phys. 84 (10) (1986) 5687–5705.
[47] S.J. Grabowski, What is the covalency of hydrogen bonding?, Chem Rev. 111 (4) (2011) 2597–2625.
[48] F. Weinhold, R.A. Klein, What is a hydrogen bond? Mutually consistent theoretical and experimental criteria for characterizing h-bonding interactions, Mol. Phys. 110 (9–10) (2012) 565–579.
[49] F. Weinhold, R.A. Klein, Anti-electrostatic hydrogen bonds, Angew. Chem. Int. Ed. 53 (42) (2014) 11214–11217.
[50] A.J. Stone, Natural bond orbitals and the nature of the hydrogen bond, J. Phys. Chem. A 121 (7) (2017) 1531–1534.

Appendix C

Paper on Substituent Effect in Benzenes

THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

Cite This: J. Phys. Chem. A 2017, 121, 8086-8096

pubs.acs.org/JPCA

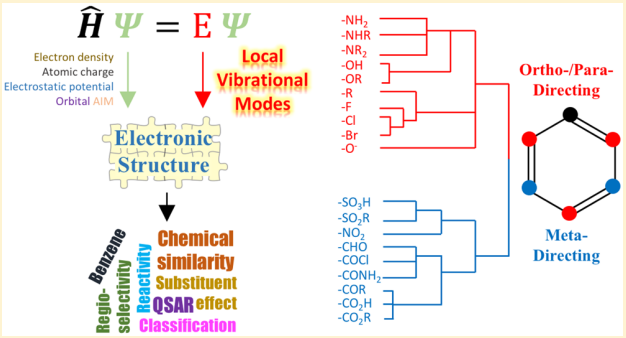
Characterizing Chemical Similarity with Vibrational Spectroscopy: New Insights into the Substituent Effects in Monosubstituted Benzenes

Yunwen Tao,[†] Wenli Zou,[‡] Dieter Cremer,[†,¶] and Elfi Kraka*[,†]

[†]Department of Chemistry, Southern Methodist University, 3215 Daniel Avenue, Dallas, Texas 75275-0314, United States

[‡]Institute of Modern Physics, Northwest University, Xi'an, Shaanxi 710069, People's Republic of China

ABSTRACT: A novel approach is presented to assess chemical similarity based the local vibrational mode analysis developed by Konkoli and Cremer. The local mode frequency shifts are introduced as similarity descriptors that are sensitive to any electronic structure change. In this work, 59 different monosubstituted benzenes are compared. For a subset of 43 compounds, for which experimental data was available, the ortho-/para- and meta-directing effect in electrophilic aromatic substitution reactions could be correctly reproduced, proving the robustness of the new similarity index. For the remaining 16 compounds, the directing effect was predicted. The new approach is broadly applicable to all compounds for which either experimental or calculated vibrational frequency information is available.



INTRODUCTION

Chemical similarity is an important concept widely used in the fields of medicinal chemistry[1] and toxicology.[2] The origin of this term or its synonym as molecular similarity was greatly influenced by the similarity property principle (SPP), which states that "similar compounds can have similar properties",[3] although this does not hold in many scenarios.[4−6] Maggiora and his co-workers suggested differentiating between chemical similarity and molecular similarity,[1] where the former stresses the physicochemical characteristics of a chemical compound while the latter emphasizes the structural and topological properties.[7−9] In this work, we will adopt this nomenclature and focus on the chemical similarity, which still has a number of open questions to be answered.

Chemists tend to put atoms or molecules with similar physicochemical characteristics in the same category in order to generalize empirical rules for practical use. A famous example is the periodic table of elements. The physicochemical properties used to characterize chemical similarity are often macroscopic quantities that are measurable, including the pK_a, solubility, boiling point, octanol−water partition coefficient (log P) and so forth. Many of these quantities are very important in quantitative structure−activity relations (QSARs). Since the 1960s, the rapid development of quantum chemical methods based on quantum mechanics (QM) has made it possible to calculate the electronic structure and associated properties of molecules, even for large systems with chemical accuracy.[10] Many attempts[2,11] have been made to develop descriptors or models for the characterization of the molecular similarity based on the results of quantum chemical calculations, referring to the fact that the wave function and electron density[12] contain all of the information related to energy about a molecule in question. Quantum chemical descriptors can be divided into two major categories, which describe either the overall molecule including HOMO and LUMO energies, dipole moments, total energy, heat of formation, and ionization potential or the fragments/substituents of a molecule like the net atomic charge.[11] Apart from these, Carbo and co-workers developed a similarity index based on the superposition of the densities of two different molecules.[13] Hodgkin proposed a similar approach to measure similarity based on the electrostatic potential.[14] However, all of the above methods failed to give a detailed and mechanistic description of similarity based on the electronic structure, and they are not connected to chemical intuition from which a better understanding can be obtained.

After Bader and his co-workers developed the atoms in molecules (AIM) theory,[15,16] Popelier proposed a new method called quantum topological molecular similarity (QTMS), which has been quite successful in characterizing similarity based on chemical insight.[17,18] Within the framework of QTMS, Popelier constructed the bond critical point (BCP) space, in which any BCP denoted as r_b can have 3−8 descriptors derived from the density information at that point, including the electron density ρ_b, the Laplacian of density $\nabla^2\rho_b$, the ellipticity of the density ϵ_b, three connected Hessian

Received: August 19, 2017
Revised: September 28, 2017
Published: September 29, 2017

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

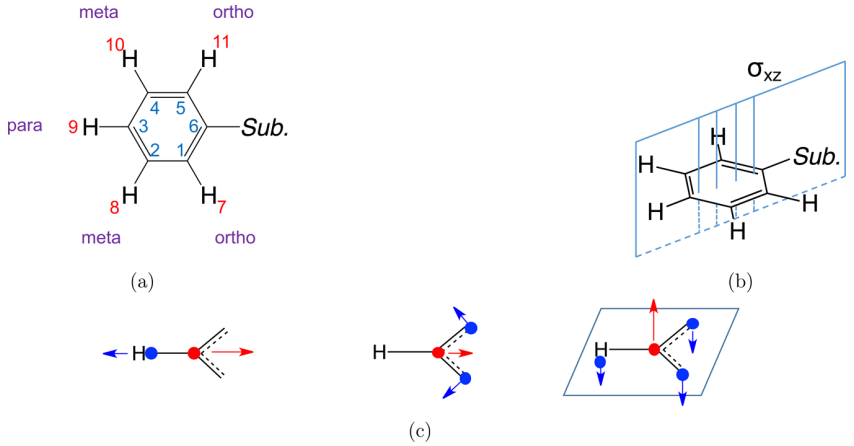


Figure 1. (a) Schematic representation of a monosubstituted benzene molecule. (b) σ_{xz} plane in a C_{2v} monosubstituted benzene. (c) Movement of atoms in the C−H bond stretching mode, C−C−C angle bending mode, and pyramidalization mode.

eigenvalues of the density λ_{1b}, λ_{2b}, and λ_{3b}, and the kinetic energy densities K_b and G_b. The application of this approach has been summarized in several articles.[19−22] However, QTMS has some limitations that might hinder its general usefulness. (i) The number of descriptors depends on the number of chemical bonds or noncovalent interactions that have BCPs. Considering the chemical bonds exclusively may fail in capturing all information on the electronic structure. (ii) The descriptors are heterogeneous with regard to the physical meaning and have different units. A horizontal comparison between the descriptors does not make sense. (iii) QTMS requires specialized procedures in the feature selection step, which may involve the principal component analysis (PCA), in order to find the most important descriptors.

In this work, we present a new approach to measure chemical similarity based on the theory of local vibrational modes, which was originally proposed by Konkoli and Cremer.[23] The local vibrational modes can be explained with the leading parameter principle. The motions of a local mode vector are obtained after relaxing all parts of the vibrating molecule except an internal coordinate parameter (leading parameter), which is first displaced infinitesimally, e.g., bond stretching, angle bending, dihedral torsion, out-of-plane torsion, and so forth. With a well-established physical basis, this approach is free from the disadvantages of the QTMS method and is able to access the information on the electronic structure by examining different types of internal parameters besides chemical bonds. Therefore, this can be regarded as an extension of our previous studies[24−33] focused on chemical bonding. We apply in this work our new similarity measure to a test set of 60 monosubstituted benzene molecules. The paper is structured in the following way: After summarizing the computational methods used, in the results and discussion part, the basic theory of the local vibrational modes and the use of local mode frequencies as similarity descriptors of benzene derivatives are described. Then, we discuss the directing effects of the substituents on benzene in electrophilic aromatic substitution reactions. Conclusions are given in the last part.

COMPUTATIONAL METHODS

Geometry optimization and normal mode analysis for all benzene derivatives involved in this work were carried out using the ωB97X-D density functional[34] with Dunning's aug-cc-pVTZ basis set[35] in the Gaussian09 package.[36] The local mode analysis was done with the program package COLOGNE2017.[37] The diagrams of hierachical clustering analysis were generated with the software package SPSS 23.[38]

RESULTS AND DISCUSSION

In the following part, the outcome of our study will be discussed.

Similarity of the Monosubstituted Benzenes. Before describing the similarity of the benzene derivatives, it is necessary to give a brief introduction into the theory of the local vibrational modes.[23,39] For any internal coordinate q_n specified within a molecule, its local mode vector $\mathbf{a}_n$ is given by

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

where $\mathbf{K}$ is the force constant matrix transformed into normal coordinates Q, $\mathbf{K} = \mathbf{L}^{\dagger}\mathbf{f}^{q}\mathbf{L}$. $\mathbf{d}_n$ is a row vector of the $\mathbf{D}$ matrix, which collects the normal modes in terms of internal coordinates, $\mathbf{D} = \mathbf{BL}$. The matrix $\mathbf{L}$ contains all normal mode vectors in Cartesian coordinates obtained by solving the Wilson equation of vibrational spectroscopy, while the Wilson $\mathbf{B}$ matrix is used to connect Cartesian coordinates to internal coordinates.[40]

The local mode force constant k_n^a can be obtained by

$$k_n^a = \mathbf{a}_n^{\dagger}\mathbf{K}\mathbf{a}_n \tag{2}$$

With the help of the $\mathbf{G}$-matrix,[23] the reduced mass of local vibrational mode $\mathbf{a}_n$ can be defined. Thus, the local vibrational frequency ω_n^a is determined by

$$(\omega_n^a)^2 = \frac{1}{4\pi^2 c^2} k_n^a G_{nn} \tag{3}$$

Normal and local vibrational modes are second-order response properties;[41] therefore, they are very sensitive to any change of the electronic structure. That is the reason why chemists have been intensively using vibrational spectroscopy for structural characterization.

In order to study the electronic structure of different monosubstituted benzene molecules, we treat the substituent as a perturbation of the phenyl ring to which it is linked and use benzene as the reference.

By doing so, any perturbation of the targeted system can be characterized by the red or blue shift of absorption peaks in

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

Table 1. Comparison of Calculated Vibrational Frequencies of Selected Local Modes in 60 Monosubstituted Benzene Derivatives

no.	substituent[a]	$\Delta\omega_m^{R\,b}$	$\Delta\omega_o^{R\,b}$	$\Delta\omega_p^{R\,b}$	$\Delta\omega_m^{a\,b}$	$\Delta\omega_o^{a\,b}$	$\Delta\omega_p^{a\,b}$	$\Delta\omega_m^{\tau\,b}$	$\Delta\omega_o^{\tau\,b}$	$\Delta\omega_p^{\tau\,b}$	exp.[c]
01	H	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
02	Br	5.37	18.54	8.76	−2.84	−11.60	4.50	1.44	−10.76	−3.23	op
03	$C(CH_3)_3$	−4.09	13.27	2.06	1.20	5.36	4.51	3.79	−3.22	−3.09	op
04	$CH(CH_3)_2$	0.84	−12.79	6.32	−11.57	−6.06	−10.28	−10.03	−26.91	−17.64	op
05	$CH{=}C(CH_3)_2$	−1.70	−6.70	2.66	−1.04	−0.31	−0.88	3.81	−7.34	−5.47	op
06	$[CH]_3{=}CH_2$	−2.75	−7.84	4.88	−0.68	5.63	0.07	1.53	−8.13	−6.14	op
07	$CH{=}CH_2$	0.11	−7.84	4.02	−1.26	−1.14	−0.02	−2.97	−32.06	−13.24	op
08	$CH{=}CHNO_2$	7.69	−1.11	7.62	−0.43	4.65	0.87	−16.45	−120.34	−39.48	op
09(a)	CH_2Br	3.81	−11.01	4.53	−0.81	2.30	1.59	2.35	−7.58	2.86	op
09(b)	CH_2Br*	3.01	16.87	4.77	0.28	−0.94	3.68	7.23	2.99	0.04	op
10(a)	CH_2Cl	3.73	−10.71	4.11	−0.66	2.72	1.82	2.28	−5.47	3.93	op
10(b)	CH_2Cl*	2.51	18.67	4.79	0.32	−1.20	3.67	7.06	2.72	−0.63	op
11(a)	CH_2F	3.13	−11.47	2.55	−1.13	2.36	1.34	−0.75	−1.02	6.70	op
11(b)	CH_2F*	1.16	18.39	4.54	1.20	−1.93	2.32	6.89	2.89	−1.78	op
12	$[CH_2]_2CH_3$	−2.59	−21.94	2.38	−1.18	2.27	1.08	4.48	−9.07	−4.88	op
13	CH_2CH_3	−1.82	−21.68	2.34	−0.14	2.87	1.47	5.16	−9.49	−5.25	op
14	CH_3	−1.81	−18.19	1.97	0.14	3.24	1.51	4.73	−9.43	−6.30	op
15	Cl	5.82	20.34	9.33	−3.16	−9.89	4.85	1.53	−16.98	−6.44	op
16	F	6.49	19.12	9.79	−8.36	−10.33	7.60	5.12	−36.46	−15.93	op
17	$N(CH_3)_2$	−6.51	23.08	9.24	−3.24	−2.43	3.40	8.02	−65.71	−44.65	op
18	NH_2	−3.25	−19.04	9.59	−4.47	−2.65	1.67	7.01	−58.97	−40.78	op
19	$NHCH_3$	−4.81	−6.10	9.13	−4.12	−2.30	2.14	8.33	−62.88	−43.41	op
20	$NHCOCH_3$	4.16	2.72	9.14	−8.40	−3.37	−5.26	7.52	−24.11	−10.27	op
21	O^-	−70.32	−42.98	−26.49	−19.79	−22.39	−10.20	−5.72	−135.10	−158.79	op
22	OCH_2CH_3	−1.78	15.76	9.71	−6.02	−6.00	4.51	8.09	−48.85	−28.25	op
23	OCH_3	−0.88	16.40	9.41	−5.51	−4.95	4.82	6.66	−50.32	−28.34	op
24	$OCOCH_3$	6.01	12.68	7.58	−11.02	−11.70	3.67	3.80	−21.55	−5.75	op
25	OH	1.10	−7.09	10.42	−7.23	−6.54	4.26	7.32	−53.47	−30.61	op
26	phenyl	0.41	−4.84	2.84	0.33	0.70	0.18	4.60	−3.71	−1.11	op
27	SH	2.96	−4.27	7.91	−2.09	−6.38	1.72	−4.81	−29.53	−15.91	op
28	CCl_3	8.23	21.83	7.03	−0.48	0.27	5.75	0.79	3.03	11.07	m
29	CF_3	9.26	15.47	5.64	0.38	0.91	3.83	−1.58	−6.69	11.49	m
30	CHO	7.19	−2.17	2.52	−1.11	−0.75	1.94	−0.37	10.96	16.09	m
31	CN	11.48	18.05	7.89	−0.50	0.03	−0.68	2.27	5.03	13.71	m
32	CO_2CH_3	6.03	22.02	1.69	1.17	2.74	3.58	−0.12	18.68	15.30	m
33	CO_2H	7.29	21.50	2.43	0.85	2.27	3.11	−0.87	18.84	16.45	m
34	$COCH_2CH_3$	4.60	12.49	1.85	−0.41	1.87	3.23	0.72	13.70	15.10	m
35	$COCH_3$	4.99	11.61	2.07	−0.23	2.07	3.36	0.34	13.19	15.28	m
36	COCl	10.39	25.65	5.46	−1.39	−0.89	3.26	−2.33	17.21	21.44	m
37	$CONH_2$	4.88	4.81	3.20	0.58	−0.78	2.62	1.82	10.06	11.18	m
38	$N(CH_3)_3^+$	22.17	25.37	23.93	−7.56	−6.12	7.15	17.16	−8.70	26.31	m
39	NH_3^+	27.28	−9.24	24.11	−15.78	−17.13	4.63	15.03	−18.12	24.53	m
40	NO_2	14.08	39.17	7.86	−2.81	−6.62	7.87	−2.15	12.27	16.06	m
41	$P(CH_3)_3^+$	20.98	−2.56	18.62	−5.14	−6.74	−1.17	14.29	6.85	35.64	m
42	$S(CH_3)_2^+$	24.03	7.16	20.51	−8.31	−13.64	1.27	13.49	4.87	37.62	m
43	SO_2CH_3	10.04	14.57	6.15	−1.13	−8.93	4.55	1.71	14.69	18.60	m
44	SO_3H	12.17	20.60	7.02	−1.36	−9.03	4.58	−0.16	11.87	17.68	m
45	AlH_2	−2.78	−32.87	−2.81	−0.05	−2.78	−5.51	−0.96	21.03	14.36	n/a
46	BeH	−4.21	−40.41	−2.68	−1.50	−1.99	−7.00	−2.04	11.42	9.89	n/a
47	BH_2	1.64	−10.98	−3.13	−0.38	2.77	−4.45	−7.85	21.77	22.73	n/a
48	CH_2^-	−70.75	−59.47	−19.42	−10.67	−21.27	−17.21	−9.97	−146.26	−228.04	n/a
49	CH_2^+	32.46	13.40	8.74	−19.54	−14.35	−18.99	−26.33	10.86	80.95	n/a
50	$[CH_2]_3NH_3^+$	9.87	−25.52	13.85	−2.79	1.76	1.08	13.32	−6.79	12.72	n/a
51	$[CH_2]_2COO^-$	−19.48	−21.37	−13.11	0.33	2.14	0.45	−3.62	−15.46	−27.14	n/a
52	$[CH_2]_2NH_3^+$	14.53	−24.16	16.68	−4.07	−0.16	0.71	15.13	−5.07	18.48	n/a
53	CH_2COO^-	−26.55	−25.08	−21.17	−0.67	−1.91	−0.25	−15.00	−22.09	−33.80	n/a
54	$CH_2NH_3^+$	19.89	−24.55	19.09	−6.32	−1.83	−0.70	16.56	1.08	29.15	n/a
55	COO^-	−34.43	−4.10	−29.40	0.77	−2.10	2.11	−16.07	10.10	−23.52	n/a
56	cyclopropyl	−2.51	−12.39	1.99	−0.37	2.18	1.53	2.87	−6.42	−2.07	n/a
57	Li	−28.28	−97.47	−14.08	−3.95	−6.31	−9.03	−4.19	2.06	−10.89	n/a

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

Table 1. continued

no.	substituent[a]	$\Delta\omega_m^{R\,b}$	$\Delta\omega_o^{R\,b}$	$\Delta\omega_p^{R\,b}$	$\Delta\omega_m^{\alpha\,b}$	$\Delta\omega_o^{\alpha\,b}$	$\Delta\omega_p^{\alpha\,b}$	$\Delta\omega_m^{\tau\,b}$	$\Delta\omega_o^{\tau\,b}$	$\Delta\omega_p^{\tau\,b}$	exp.[c]
58	Na	−32.49	−100.44	−16.00	−3.82	−11.82	−8.71	−6.43	−1.72	−15.97	n/a
59	PH_2	1.60	−8.28	2.19	0.36	−1.11	0.02	−2.95	5.41	4.35	n/a
60	PO_4^{2-}	−65.88	17.81	−45.82	−11.12	−19.62	−5.81	−14.99	−83.31	−133.55	n/a

[a]The column "substituent" denotes the structure linked to the phenyl ring in a monosubstituted benzene molecule. [b]Local mode frequency differences $\Delta\omega$ are calculated by ω(target) − ω(benzene). Superscripts R, α, and τ stand for C−H bond stretching, C−C−C angle bending, and pyramidalization modes, respectively. Subscripts m, o, and p denote the meta-, ortho-, and para-locations with regard to the substituent, respectively. The unit is cm^{-1}. [c]Column "exp." denotes the directing effect in electrophilic aromatic substitution reactions caused by the substituent, which have been experimentally confirmed. "op" means that the products are dominated by ortho- and para-products, while "m" means that the products are dominated by meta-product. "n/a" means that there are no experimental data available.

vibrational spectroscopy. For example, this procedure has been frequently applied to measure the temperature influence on liquid water.[42] We borrow this idea of frequency shift for the purpose of analyzing the influence of the substituents on the electronic structure of the phenyl ring. However, in our studies, we are not using the normal vibrational frequencies,[40] which can be directly measured by the infrared or Raman spectrometer, because normal modes suffer from mass coupling and they are delocalized over the molecular system in question.[23,39] We focus on the shift/change of the local mode frequencies, which are free from mass coupling and thus can be used to describe the local vibrations of the phenyl ring. Besides, the frequencies of such vibrations can be directly compared among different benzene derivatives in order to characterize their different electronic structure.

Figure 1a shows the general structure of monosubstituted benzene molecules. Each molecule can have a different number of atoms and different symmetry depending on the substituent covalently linked to the C6 atom. The analysis of the influence of the substituent on the phenyl ring (C_6H_5) can be assessed by investigating the local vibrational modes involving the phenyl ring atoms. We can construct a redundant set of parameters including 11 bonds, 16 bond angles, and at least 7 dihedral angles. In addition, the Cremer−Pople ring coordinates offer the corresponding local modes for the puckering[43] and deformation[44] modes of the six-membered ring. This comprehensive set of parameters provide a complete and detailed characterization of the electronic structure of the phenyl ring. By taking advantage of symmetry, we can considerably reduce this set without losing a detailed description of the electronic structure. Benzene has D_{6h} symmetry, and when one of its hydrogen atoms is changed with another atom or functional group, a monosubstituted benzene results and the symmetry will be reduced to C_{2v} or even lower depending on the substituent. Chemists have classified the five C−H locations besides the substituent into three major categories, denoted as ortho, meta, and para (see Figure 1a). This is due to the fact that most monosubstituted benzene molecules have C_{2v} symmetry, where two meta- or ortho-positions are identical, e.g., in fluorobenzene. A monosubstituted benzene with a symmetry lower than C_{2v} no longer has the σ_{xz} mirror plane (see Figure 1b); therefore, the two meta-positions and ortho-positions will be different, as for example in the case of benzoic acid.

The obvious strategy to choose local mode parameters is based on the three different positions with regard to the substituent, e.g., ortho, meta, and para. For each of these three different sites, one can propose that the C−H bond stretching should be considered (see the left diagram of Figure 1c) because this is a parameter involving both the carbon and hydrogen atoms at a specific site. For the six-membered ring, the best choice is to select the C−C−C angle bending mode, where the middle carbon atom of the angle is located at the ortho-, meta-, or para-position (see the middle diagram of Figure 1c). These two parameters describe the vibration of atoms within the plane of the phenyl ring. One has to also include a parameter describing the out-of-ring-plane vibration, e.g., the pyramidalization mode[45] (see the right diagram of Figure 1c). The direction of this mode is perpendicular to the plane of the phenyl ring. In this way, a comprehensive description of the electronic structure is covered by considering both the σ-bonding electrons and π electrons. This leads to a total of nine parameters for the local mode analysis, namely, a set of bond stretching, angle bending, and pyramidalization modes for the ortho-, meta-, and para-sites. For monosubstituted benzene molecules with symmetry lower than C_{2v}, the local mode frequencies can have different values at two ortho- and meta-sites. In this case, we take the averaged value for further analysis.

In this work, a broad range of 60 monosubstituted benzene derivatives was studied. The selection of these molecules was mainly based on experimental data, and some commonly used substituents were also added into the data set.[46−49] For each molecule, the local vibrational frequencies of the nine selected parameters were calculated as ω_m^R, ω_o^R, ω_p^R, ω_m^α, ω_o^α, ω_p^α, ω_m^τ, ω_o^τ, and ω_p^τ. The superscripts R, α, and τ stand for bond stretching, angle bending, and pyramidalization parameters, respectively. The subscripts m, o, and p specify the meta-, ortho-, and para-positions in a monosubstituted benzene.

In order to reveal the influence of the substituent on the phenyl ring, we calculated the local mode frequency shift by subtracting the local mode frequency of a specific parameter in the benzene molecule from its counterpart in a monosubstituted benzene target molecule. This leads to the local mode frequency difference $\Delta\omega_n^a$.

$$\Delta\omega_n^a = \omega_n^a(\text{target}) - \omega_n^a(\text{benzene}) \tag{4}$$

If benzene is taken as the target molecule, $\Delta\omega_n^a = 0$. In Table 1, the local mode frequency differences of the 9 selected parameters are listed for the 60 benzene derivatives. For each local mode frequency difference, its value can be either positive (blue shift) or negative (red shift). For all monosubstituted benzene derivatives in Table 1, both blue and red shifts were observed. A blue or red shift of vibrational frequencies of both normal modes and local modes is caused by a change of electronic structure, and therefore, the magnitude of frequency shifting indicates the extent to which the electronic structure is perturbed. If all nine frequency shifts have relatively large values, it implies that the influence of the substituent on the phenyl ring is significant. Such a situation is found for O^- and CH_2^- substituents.

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

However, just checking frequency shift values cannot lead to a useful basis for a similarity measure for two major reasons: (i) Even within one benzene derivative, the variation of the shift values may not be consistent. A benzene derivative may have small shift values for some parameters but relatively large shift values for another parameter. (ii) While the frequency shift values in Table 1 have a physical meaning, the attempt to use a quantity (e.g., total sum of squares) in order to get an overall description of the change in electronic structure would lose this physical foundation.

As a solution to circumvent the above deficiencies, we propose an approach that is connected to the two well-established concepts of QSARs[50] and molecular descriptors.[51] We can construct a vector $\mathbf{\Omega}$

$$\mathbf{\Omega} = (\Delta\omega_m^R, \Delta\omega_o^R, \Delta\omega_p^R, \Delta\omega_m^\alpha, \Delta\omega_o^\alpha, \Delta\omega_p^\alpha, \Delta\omega_m^\tau, \Delta\omega_o^\tau, \Delta\omega_p^\tau) \tag{5}$$

collecting the local mode frequency shifts of all nine selected parameters. In this way, each monosubstituted benzene derivative has its own vector $\mathbf{\Omega}$ characterizing its substituent effect. As for a substituent besides a hydrogen atom, vector $\mathbf{\Omega}$ can never be a zero vector $\mathbf{0}$. Therefore, for two vectors $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_B$ of any two different monosubstituted benzenes A and B, the similarity and its corresponding distance[52] can be defined using the cosine function

$$\text{similarity} = \cos(\theta) = \frac{\sum_{i=1}^{9} \Omega_{Ai}\Omega_{Bi}}{\sqrt{\sum_{i=1}^{9} \Omega_{Ai}{}^2}\sqrt{\sum_{i=1}^{9} \Omega_{Bi}{}^2}} \tag{6}$$

$$\text{distance} = 1 - \text{similarity} \tag{7}$$

where Ω_{Ai} and Ω_{Bi} are elements of vectors $\mathbf{\Omega}_A$ and $\mathbf{\Omega}_B$, respectively.

Before calculating the cosine similarity, it is necessary to balance the 9 different types of frequency shifts by standardizing 59 shift values (benzene is excluded) into the region from −1 to 1.

$$\Delta\omega' = 2\frac{\Delta\omega - \min \Delta\omega}{\max \Delta\omega - \min \Delta\omega} - 1 \tag{8}$$

This makes all nine frequency shifts of the vector $\mathbf{\Omega}$ comparable.

After the similarity has been determined between any pair of monosubstituted benzene molecules, a similarity matrix with dimensions of 59 × 59 can be constructed. With this information, the hierarchical cluster analysis (HCA)[53] is carried out and the relationship between any two monosubstituted benzene molecules can be visualized, as shown in Figure 2. The dendrograms as results of the HCA in this work (including Figures 2 and 3) have been rescaled with regard to the distance between any two monosubstituted benzenes, and the largest distance value between the two farthest groups was set to 5. It has to be noted that in the HCA similarity is reflected by distance; similar compounds are close together in the dendrogram, while a larger distance between two compounds reflects that their electronic structure is different.

Forty-three of the 59 monosubstituted benzene derivatives have been classified to be either meta-directing or ortho-/para-directing according to experimental studies[48,49] with regard to the regioselectivity in the electrophilic aromatic substitution reaction (see Table 1). Therefore, we have labeled in this work all ortho-/para-directing groups with red color and meta-directing groups with blue color. The substituents whose directing effect has not been experimentally reported yet are labeled in black color.

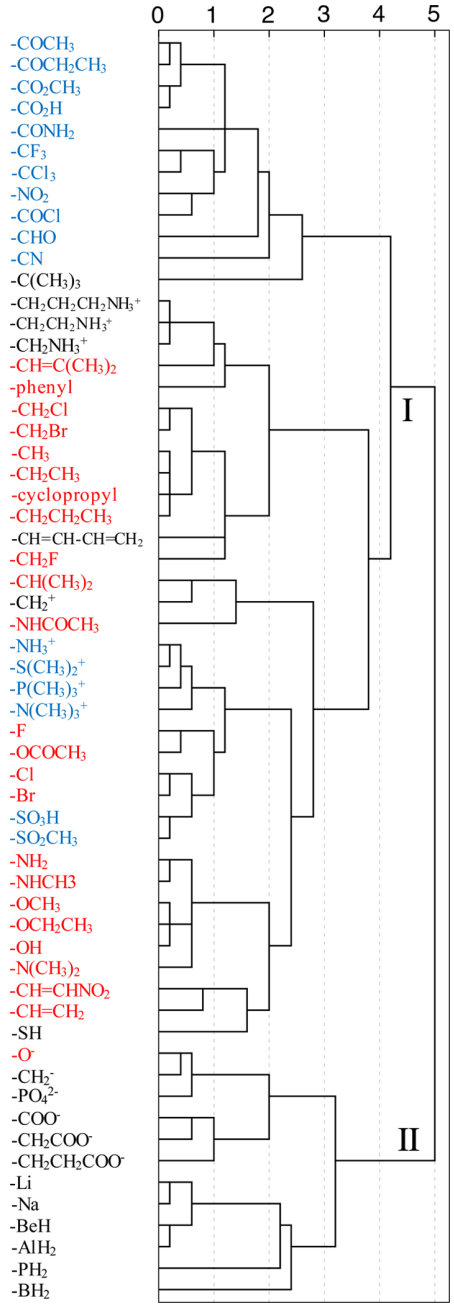


Figure 2. Diagram of hierarchical clustering of 59 monosubstituted benzenes based on 9 local mode frequency shifts.

From the clustering result shown in Figure 2, which is based on the pairwise similarity using nine selected local mode parameters, a series of interesting observations can be made.

(1) The substituents in red or blue color tend to cluster together, leading to six small clusters alternating from the top to the bottom. This indicates that the local mode frequency shifts can be used to distinguish between meta-directing and ortho-/para-directing groups.

(2) In the upper part of the dendrogram, two acyl groups (−COR) are clustered as nearest neighbors. The carboxyl group (−COOH) clusters together with the ester group

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

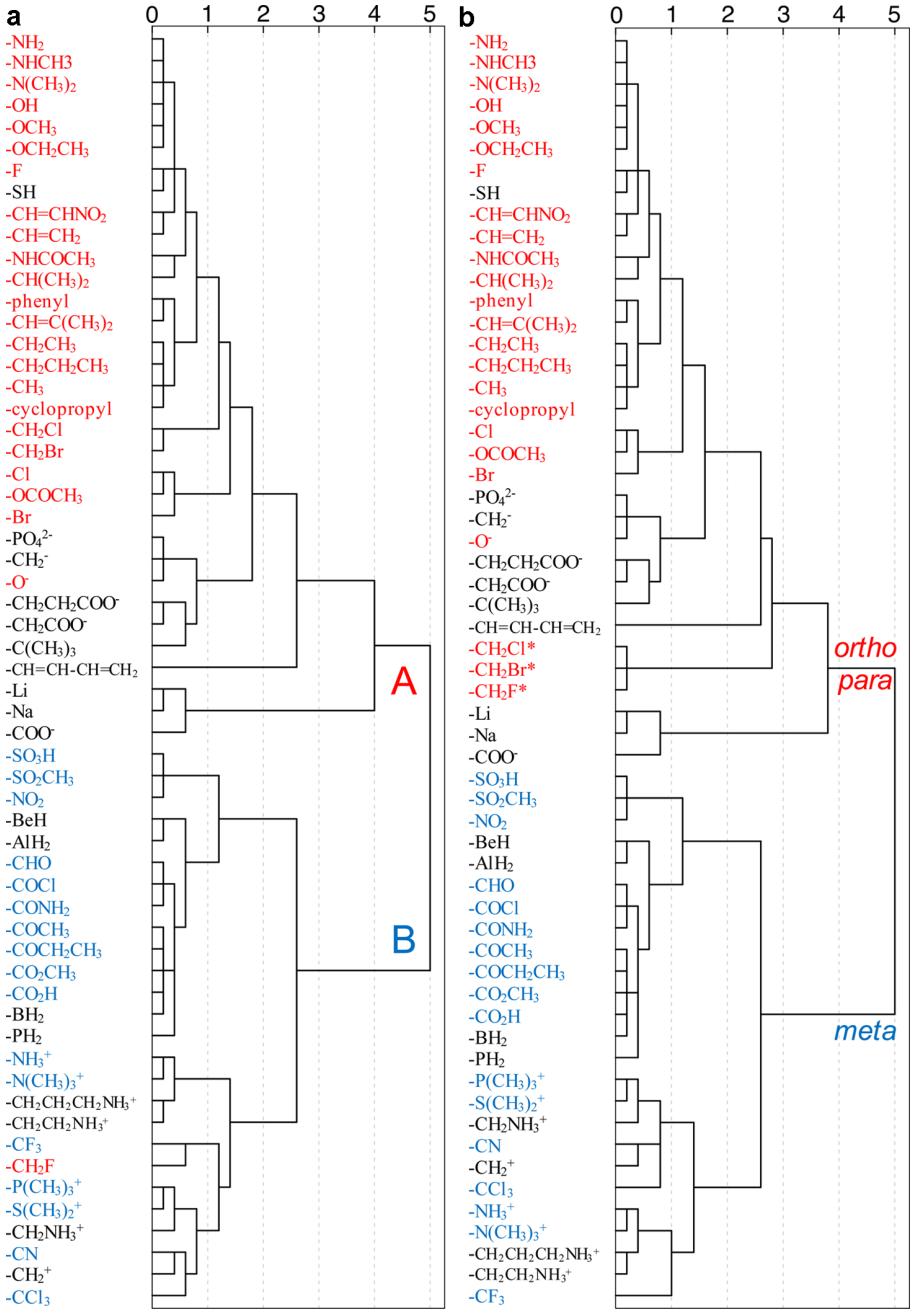


Figure 3. Diagrams of hierarchical clustering of 59 monosubstituted benzenes based on 6 local mode frequency shifts. (a) With CH_2F as the outlier. (b) Without any outliers.

(−COOR). These four substituents are clustered together as two groups. The small distance between them shows that these four functional groups have a similar influence on the phenyl ring in the electronic structure. This is reasonable because they share the carbonyl group as a common substructure covalently linked to the phenyl ring. In this section, a few other functional groups (colored in blue) containing the carbonyl group can be found, including the amide ($-CONH_2$), carbonochloridoyl (−COCl), and aldehyde (−CHO) groups. However, these groups are further apart from each other.

(3) Another small cluster of the dendrogram is composed of the $-[CH_2]_nNH_3^+$ functional groups (n = 1−3). The distances between them are very small, which implies that their influence on the electronic structure of the phenyl ring is almost the same, even though with increasing n the NH_3^+ group gets further away from the phenyl ring. However, the $-NH_3^+$ substituent itself does not belong to this small cluster. This implies that the substituent effect of the $-NH_3^+$ functional group has a different mechanism from that of the $-[CH_2]_nNH_3^+$ substituents. While the electron hole in the $-NH_3^+$ has a direct interaction with the electrons of the phenyl ring, the positive charge in the $-[CH_2]_nNH_3^+$ substituents is "diluted" by the CH_2 groups in between.

(4) In the red upper part of the diagram of Figure 2, there are two small clusters in which the members are very close to each other. The first cluster contains two halogenated methyl

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

groups including $-CH_2Cl$ and $-CH_2Br$. These two substituents have a similar effect on the phenyl ring because Cl and Br are in the same group of the periodic table. However, it is noteworthy that this cluster does not contain $-CH_2F$ even though the F atom is also a halogen atom. The second cluster contains four alkyl groups, namely, the methyl, ethyl, propyl, and cyclopropyl groups. These four functional groups have almost the same influence on benzene, and it is in line with the common understanding of chemists.

(5) In the middle of the diagram, there is a cluster with four substituents colored in blue, including $-NH_3^+$, $-S(CH_3)_2^+$, $-P(CH_3)_3^+$, and $-N(CH_3)_3^+$. These substituents are cationic, and they can impose a similar substituent effect. However, another cationic substituent $-CH_2^+$ is not included here, indicating a different mechanism for the substituent effect.

(6) Three halide substituents −F, −Cl, and −Br are contained in a cluster. The chlorine and bromine are close together while the fluorine atom stands out again as in (4). This can be attributed to the large electronegativity of this element. Interestingly, the acetate ester group ($-OCOCH_3$) has a relatively small distance to the fluorine atom, implying a similar substituent effect for the phenyl group, which was never reported before.

(7) The mesyl group ($-SO_2CH_3$) and sulfo group ($-SO_3H$) are clustered with the smallest distance. This means that the substituent effect is hardly changed when the −OH group is exchanged with a $-CH_3$ group and the substituent effect is determined by the $-SO_2$ part.

(8) In the red lower part of the dendrogram, there is a small cluster containing the hydroxyl and two ether groups. This indicates that the oxygen atom directly linked to the phenyl ring plays a more important role than the hydrogen atom or the alkyl group connected to it with regard to substituent effects.

(9) We find three amine groups ($-NH_2$, $-NHCH_3$, and $-N(CH_3)_2$) that are contained in a cluster with the tertiary amine group having a relatively larger distance toward the primary and secondary amine groups. Furthermore, we find a big cluster with members ranging from $-NH_2$ down to −SH. They have one thing in common: the atoms linked to the phenyl ring can have their p−π electrons interact with the π electrons of the phenyl ring.

(10) The most interesting section of the dendrogram is that of the anionic substituents located in the lower part. All anionic substituents studied in this work cluster together exclusively. That means that substituents with a diffuse anion show a more consistent mechanism in changing the electronic structure of benzene compared to the cationic substituents.

(11) If one zooms out of this dendrogram, one can identify the two largest clusters (I and II). Cluster II contains two subclusters. All of the members within one of the subclusters are anionic substituents, which have been discussed in (10). The members of the other subcluster are atoms or hydrides of elements with weak electronegativity. According to the electronegativity scale proposed by Pauling,[54,55] Li(0.98), Na(0.93), Be(1.57), Al(1.61), P(2.19), and B(2.04) have smaller electronegativities than H(2.20), which is the reference substituent in benzene. If the anionic substituents can be categorized as those with weak electronegativity, we can state that the largest cluster in the upper dendrogram (cluster I from $-COCH_3$ down to −SH) contains groups with a larger electronegativity than H. In this regard, the concept of electronegativity has now been extended from atoms to polyatomic functional groups.

In summary, the local vibrational modes of the monosubstituted benzenes unambiguously reflect similarities in their electronic structure. Furthermore, the similarity of the substituted benzenes and their substituent effects can be presented in a straightforward approach, in which different chemical species are correlated to one another.

Directing Effects of the Substituents in Electrophilic Aromatic Substitution Reactions. Since the seminal discovery by Brown and Gibson in 1892,[56] it has been well recognized that the substituent of a monosubstituted benzene can affect the regioselectivity of the electrophilic aromatic substitution replacing a second H atom in the benzene.[48,49] Such a substituent effect works in the way that the second substitution reaction will be promoted at either the meta-position or the ortho-/para-position. Organic chemistry textbooks frequently provide a list of meta-directing groups and a list of ortho-/para-directing groups.[46,47] This is an example where chemists have managed to generalize empirical rules; however, a physical basis is still missing. This can be provided by the use of the local mode frequency shifts.

The nine selected local mode frequency shifts of the monosubstituted benzene molecules form the basis to distinguish between meta-directing and ortho-/para-directing groups. By focusing on six out of the nine local mode frequency shifts, it is feasible to separate the meta-directing and the ortho-/para-directing substituents.

The refinement of these six parameters involves a feature selection process.[57] As there exists some redundancy between the C−H stretching mode and the C−C−C angle bending mode at a specific site with regard to electronic structure change, one has to remove either one in order to obtain a robust model. Testing $2^3 = 8$ different combinations, we found only one set of parameters that performs best and leads to the desired clustering result.

In this set, the local bond stretching vibration in ortho- and para-positions as well as the angle bending mode at the meta-position was eliminated, leading to a new vector Ω of dimension 6, containing the remaining six local mode frequency shifts, as shown in eq 9. This vector is expected to encompass the information linked to the directing effect of a substituent.

$$\Omega = (\Delta\omega_m^R, \Delta\omega_o^\alpha, \Delta\omega_p^\alpha, \Delta\omega_m^\tau, \Delta\omega_o^\tau, \Delta\omega_p^\tau) \tag{9}$$

The discrimination of the meta- and ortho-/para-directing substituents is a typical two-class classification problem, which can be solved by a classification procedure with supervised learning.[58] However, it can also be solved via the HCA, which can be regarded as an unsupervised classification method. Other than the similarity problem discussed in the above section, here each case has to be treated independently as the directing effect of a substituent is determined by its own physical chemical properties.

Instead of adopting the standardization by variable as done in the above section, we standardize the data by case, so that each vector Ω is standardized into the region of [0,1] leading to the new vector $\hat{\Omega}$

$$\hat{\Omega}_i = \frac{\Omega_i - \min \Omega}{\max \Omega - \min \Omega} \tag{10}$$

in which i runs from 1 to 6. This standardization is necessary because it makes the $\hat{\Omega}$ for each monosubstituted benzene more comparable.

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

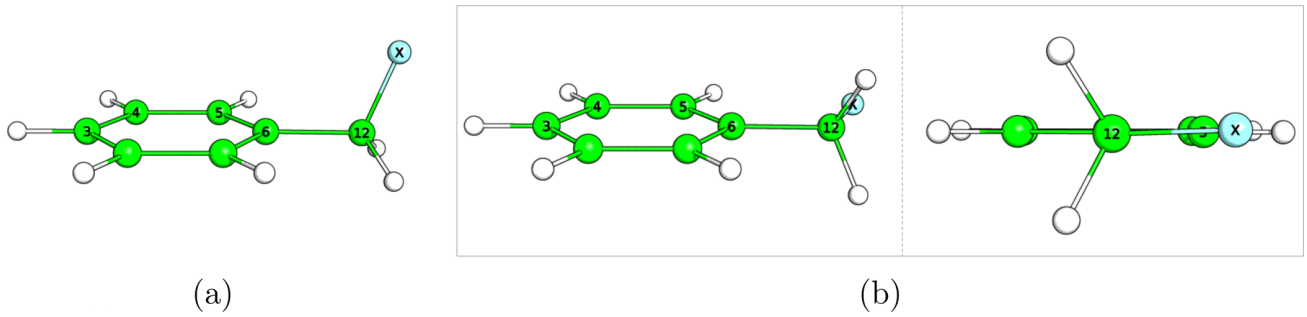


Figure 4. Ball-and-stick representation of a benzyl halide molecule. Green, white, and cyan spheres stand for C, H, and X (= F, Cl, and Br) atoms, respectively. (a) Local minimum geometry. (b) First-order saddle point geometry (X atom in the phenyl ring plane) in two perspectives.

The squared Euclidean distance is calculated for $\hat{\Omega}_A$ and $\hat{\Omega}_B$ for any two monosubstituted benzenes A and B.

$$\text{distance} = \sum_{i=1}^{6} (\hat{\Omega}_{Ai} - \hat{\Omega}_{Bi})^2 \tag{11}$$

The resulting dendrogram is shown in Figure 3a. It is important to note that this dendrogram is calculated in order to identify two major classes of substituents with regard to their directing effect, while the dendrogram in Figure 2 emphasizes the relationship between any two substituents. Besides the clusters with minimal distances between their members, here the major task is to determine if the clustering procedure can lead to two major clusters in which the two desired classes (meta- and ortho-/para-directing groups) perfectly reside.

Any classification problem in which two or more targeted categories should be distinguished can be solved in various ways with different explicit or implicit (black box) models.[59] However, the following aspects need to be considered.

(i) Are these targeted categories well-defined and specific?
(ii) What determines the targeted categories? Is this classification problem solvable?
(iii) Are all features/descriptors related to the targeted categories?
(iv) Are there any features/descriptors missing that are related to the targeted categories?

For the classification problem of this work, the two targeted categories of meta-directing and ortho-/para-directing groups are well-defined from the regioselectivity of the electrophilic aromatic substitution reactions, which share the same reaction mechanism for these different positions with regard to the substituent. In this way, the reactivity is entirely determined by the electronic structure of the monosubstituted benzene where the phenyl ring part plays the essential role participating in the reaction. Therefore, the classification problem can be solved when we are able to identify and characterize the differences in the electronic structure of the monosubstituted benzenes that decide on the meta-directing or ortho-/para-directing reactivity.

As can be visualized in Figure 3a, all of the ortho-/para-directing substituents are clustered in the largest cluster A (except $-CH_2F$), while all meta-directing substituents are clustered in the second largest cluster B, reflecting the validity of our model.

The $-CH_2F$ outlier can be a result of the following: (i) The classification model is not robust enough to cover all possible cases, which would reflect a model deficiency; (ii) the benzyl fluoride is a special case.

Benzyl fluoride is an outlier, while its higher homologues benzyl chloride and benzyl bromide are correctly classified as the ortho-/para-directing groups, although these three halobenzenes share a similar geometry, shown in Figure 4a. However, there are subtle differences to be considered in the electronic structure. The fluorine atom has the largest electronegativity compared with chlorine and bromine. The distance of F−C12 is 1.394 Å, while the distances for Cl−C12 and Br−C12 are 1.812 and 1.967 Å, respectively. The distances of F−C5, F−C4, and F−C3 are smaller than their counterparts for chlorine and bromine. These two factors are essential as they indicate a direct and significant interaction between the fluorine atom and the π electrons of the carbon atoms located at the ortho-, meta-, and para-positions. Such a through-space interaction is substantially diminished in the cases of the benzyl chloride and the benzyl bromide due to the larger halogen−carbon distances.

We are predicting the regioselective reactivity, which is the result of a dynamic process using the static geometry of the monosubstituted benzene molecule. Therefore, the incoming electrophile and a catalyst (if present) are left out in the prediction. This does not make a difference for the other substituted benzenes, which have been correctly classified. However, in the case of benzyl fluoride, the incoming reaction partner(s) can reduce the through-space interaction between the halide atom and the π electrons at the ortho-, meta-, and para-positions. One might argue that the trifluorotoluene and toluene molecules could suffer from the same problem. However, in the $-CF_3$ substituent of the trifluorotoluene, one F atom lies within the plane of the phenyl ring, while the other two are positioned on both sides of the phenyl plane symmetrically. In this way, the through-space interaction between the two fluorine atoms and the π electrons of the carbons cancel and only the through-bond interactions remain. For the $-CH_3$ group in the toluene molecule, the H atoms have contracted electron density and there is no possibility for through-space interaction.

In order to show the influence of the through-space interaction for the CH_2F group, we have rotated the F atom into the ring plane (see Figure 4b). Geometry optimization led to a first-order saddle point only 0.25 kcal/mol higher in energy than the energy minimum conformation shown in Figure 4a. In the saddle point conformation, through-space interaction is eliminated. To be consistent, the same was applied for the benzyl chloride and bromide molecules. The (averaged) local mode frequency shift values of the six parameters were calculated for these geometries, (they can be found as no.

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

09(b), 10(b), and 11(b) in Table 1; the superscript * after the name of the substituents denotes the rotated geometries.)

The HCA after replacing the benzyl fluoride, chloride, and bromide molecule with rotated forms is shown in Figure 3b. The three substituents have now been correctly classified as ortho-/para-directing groups, and they cluster together with small distances between each other. This proves that our classification model is robust and reliable, correctly classifying the 43 substituents for which experimental data is available with regard to the directing effect.

On this basis, we made predictions for the 16 remaining substituents. The sulfhydryl (−SH), *tert*-butyl ($-C(CH_3)_3$), and butadienyl ($-CH{=}CH-CH{=}CH_2$) groups along with the lithium atom and sodium atom are predicted to direct a second substituent into an ortho-/para-position. Anionic substituents including $-CH_2^-$, $-PO_4^{2-}$, $-COO^-$, $-CH_2COO^-$, and $-CH_2CH_2COO^-$ are also predicted to belong to this class. Substituents like −BeH, $-BH_2$, $-AlH_2$, and $-PH_2$ are expected to be meta-directing. The cationic substituents including $-CH_2^+$, $-CH_2NH_3^+$, $-CH_2CH_2NH_3^+$, and $-CH_2CH_2CH_2NH_3^+$ are classified as meta-directing groups according to the prediction result.

In addition, when the classification result with regard to the directing effect in Figure 3b is compared with the similarity measurement in Figure 2, we find that some small clusters of Figure 2 are kept in Figure 3b while clusters with larger distances between their members are broken in the classification result. The major reason responsible for this difference is that we have used nine different local mode frequency shifts for characterizing the similarity, but only six of them were taken in the classification problem. Although the three extra local mode frequency shifts as the descriptors can help to give a higher resolution in the characterization of the electronic structure, the information that they are carrying about the electronic structure is not related to the directing effect of the substituents.

Recently, Liu has attempted to distinguish meta-directing groups from ortho-/para-directing groups using three Hirshfeld charge values of the carbon atoms in question.[60] However, his model seems to have deficiencies because up to nine $-NR_3^+$ groups do not fit in. These outliers cast doubt on (i) whether atomic charges are appropriate descriptors to describe the electronic structure with regard to the directing effect quantitatively and (ii) the choice of his test set. Fu et al. carried out a set of systematic studies on the classification of 14 monosubstituted benzenes using 14 different theoretical models.[61] The major difference between their classification/prediction and ours lies in that they have also tried to predict the relative portion of ortho-, meta-, and para-products, which is an interesting but challenging task. In their work, the 14 different methods used for characterization fall into two major categories including methods based on local electronic softness and those reflecting electrostatic effects. As the number of monosubstituted benzene molecules was quite limited in that work, a direct comparison with our model is not feasible. However, it should be possible to characterize the dominance of the ortho- or para-product with local mode frequency shifts as they reflect directly the electronic structure of the benzene derivative. Work is in progress to demonstrate this. Noteworthy is that Bader and Chang did seminal work on QTAIM analysis of electrophilic aromatic substitution for nine monosubstituted benzenes.[62]

In summary, the methods used in Liu's and Fu's work are related to the properties derived from the electron density, atomic charge, molecular and atomic orbital, and related orbital energy. These concepts can be understood in a way that they are used as descriptors for characterizing the local feature of the electronic structure.[63,64] Being derived directly or indirectly from the molecular wave function Ψ, however, these methods fall short in the following two aspects: (i) The number of descriptors is too limited. For example, Liu considers only three atomic charge values for the ortho-, meta-, and para-carbons. In comparison, our local mode description uses six descriptors characterizing the electronic structure in different directions, leading to more detailed information. (ii) Models like atomic charges are based on assumptions or are even based on other models, although they might be useful for interpretative purpose. Wave function and orbitals are always delocalized functions in the space, and any attempt to assign them to a specific atom has no physical basis.

The local mode frequency starts from the eigenvalue of the Schrödinger equation, which is the energy E. Vibrational modes derived from the Hessian matrix are second-order response properties and therefore can be used as sensitive measures of any change in the electronic structure.

CONCLUSIONS

The assessment of the similarity of monosubstituted benzene molecules using the local vibrational modes has led to a series of interesting results as well as a platform for future work.

(1) The local mode frequency shifts introduced in this work have the capability to characterize the similarity of different types of benzene derivatives. At the same time, pairwise dissimilarity can also be defined using a cosine function. With the help of the HCA, the relationship between different monosubstituted benzenes can be visualized and interpreted.

(2) As it has been stated by many chemists that the benzene molecule probes inductive and resonance effects resulting from substituents,[47,49,60,61,65] it is helpful to design model systems in order to study these two effects in a systematic way and to develop a quantitative index similar to the Hammett substituent constant[66] based on the electronic structure.

(3) The concept of the blue and red shift of the local mode frequency value can be compared to the up- and downfield of the NMR spectroscopy. Both of these shifts can be used in order to characterize the change in the electronic structure. However, the local mode frequency shift is not limited to atoms of ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P; it can be applied to any element in a molecule and offer more abundant information.

(4) The similarity result obtained in this work can be used as guidance with regard to the choice of functional groups in synthesis and molecular design.

(5) For the first time, we have correctly classified 43 monosubstituted benzene molecules with regard to their directing effect of the substituents in electrophilic aromatic substitution reactions based on local vibrational frequencies. We have also predicted the directing effects of the substituent in 16 additional monosubstituted benzenes for which no experimental data is known.

(6) The procedure employed to study the regioselectivity problem in this work can be applied to other reactions, including the Diels−Alder reactions[67−69] and transition metal-catalyzed reactions where the ligand plays an important role.[70−77] This will be part of future studies aiming at the development of a generally applicable tool for rational catalyst

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

design. Recent work of Sigman and his co-workers has shown that efforts in this direction can be promising.[78−80]

(7) The local mode frequency shift provides new insight into characterizing the electronic structure of a molecule. This framework is quite unique and different from those well-accepted models based on atomic charges and orbitals in that they can be calculated or derived from the normal vibrational frequencies.

AUTHOR INFORMATION

Corresponding Author

*E-mail: ekraka@smu.edu.

ORCID

Wenli Zou: 0000-0002-0747-2428

Dieter Cremer: 0000-0002-6213-5555

Elfi Kraka: 0000-0002-9658-5626

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was financially supported by the Natural Science Foundation, Grant 1464906. We thank SMU for providing computational resources.

DEDICATION

[¶]In Memoriam.

REFERENCES

(1) Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57*, 3186−3204.

(2) Nikolova, N.; Jaworska, J. Approaches to Measure Chemical Similarity − A Review. *QSAR Comb. Sci.* **2003**, *22*, 1006−1026.

(3) Johnson, M.; Maggiora, G. *Concepts and Application of Molecular Similarity*; John Wiley & Sons: New York, 1990.

(4) Maggiora, G. M. On Outliers and Activity Cliffs - Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535−1535.

(5) Stumpfe, D.; Bajorath, J. Exploring Activity Cliffs in Medicinal Chemistry. *J. Med. Chem.* **2012**, *55*, 2932−2942.

(6) Stumpfe, D.; Hu, Y.; Dimova, D.; Bajorath, J. Recent Progress in Understanding Activity Cliffs and Their Utility in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57*, 18−28.

(7) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225−233.

(8) Willett, P. Similarity-based Virtual Screening using 2D Fingerprints. *Drug Discovery Today* **2006**, *11*, 1046−1053.

(9) Vogt, M.; Bajorath, J. Modeling Tanimoto Similarity Value Distributions and Predicting Search Results. *Mol. Inf.* **2017**, *36*, 1600131.

(10) Helgaker, T.; Ruden, T. A.; Jørgensen, P.; Olsen, J.; Klopper, W. A Priori Calculation of Molecular Properties to Chemical Accuracy. *J. Phys. Org. Chem.* **2004**, *17*, 913−933.

(11) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1044.

(12) Mezey, P. G. Theorems on Molecular Shape-Similarity Descriptors: External T-Plasters and Interior T-Aggregates. *J. Chem. Inf. Model.* **1996**, *36*, 1076−1081.

(13) Carbó, R.; Leyda, L.; Arnau, M. How Similar is A Molecule to Another? An Electron Density Measure of Similarity Between Two Molecular Structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189.

(14) Hodgkin, E. E.; Richards, W. G. Molecular Similarity Based on Electrostatic Potential and Electric Field. *Int. J. Quantum Chem.* **1987**, *32*, 105−110.

(15) Bader, R. F. W.; Anderson, S. G.; Duke, A. J. Quantum Topology of Molecular Charge Distributions. 1. *J. Am. Chem. Soc.* **1979**, *101*, 1389−1395.

(16) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; International Series of Monographs on Chemistry; Clarendon Press, 1994.

(17) Popelier, P. L. A. Quantum Molecular Similarity. 1. BCP Space. *J. Phys. Chem. A* **1999**, *103*, 2883−2890.

(18) O'Brie, S. E.; Popelier, P. L. A. Quantum Molecular Similarity. 3. QTMS Descriptors. *J. Chem. Inf. Model.* **2001**, *41*, 764−775.

(19) O'Brien, S. E.; Popelier, P. L. A. Quantum Topological Molecular Similarity. Part 4. A QSAR Study of Cell Growth Inhibitory Properties of Substituted (E)-1-phenylbut-1-en-3-ones. *J. Chem. Soc., Perkin Trans. 2* **2002**, 478−483.

(20) O'Brien, S. E.; Popelier, P. L. Quantum Molecular Similarity. Part 2: The Relation Between Properties in BCP Space and Bond Length. *Can. J. Chem.* **1999**, *77*, 28−36.

(21) Popelier, P. L. A.; Chaudry, U. A.; Smith, P. J. Quantum Topological Molecular Similarity. Part 5. Further Development with An Application to the Toxicity of Polychlorinated Dibenzo-p-dioxins (PCDDs). *J. Chem. Soc., Perkin Trans. 2* **2002**, 1231−1237.

(22) Popelier, P. L. A.; Smith, P. J. In *Chemical Modelling: Applications and Theory*; Hinchliffe, A., Ed.; Royal Society of Chemistry, 2002; Vol. 2, pp 391−448.

(23) Konkoli, Z.; Cremer, D. A New Way of Analyzing Vibrational Spectra. I. Derivation of Adiabatic Internal Modes. *Int. J. Quantum Chem.* **1998**, *67*, 1−9.

(24) Kalescky, R.; Zou, W.; Kraka, E.; Cremer, D. Local Vibrational Modes of the Water Dimer − Comparison of Theory and Experiment. *Chem. Phys. Lett.* **2012**, *554*, 243−247.

(25) Freindorf, M.; Kraka, E.; Cremer, D. A Comprehensive Analysis of Hydrogen Bond Interactions Based on Local Vibrational Modes. *Int. J. Quantum Chem.* **2012**, *112*, 3174−3187.

(26) Zou, W.; Cremer, D. C_2 in a Box: Determining its Intrinsic Bond Strength for the $X^1\Sigma_g^+$ Ground State. *Chem. - Eur. J.* **2016**, *22*, 4087−4099.

(27) Kalescky, R.; Kraka, E.; Cremer, D. Identification of the Strongest Bonds in Chemistry. *J. Phys. Chem. A* **2013**, *117*, 8981−8995.

(28) Oliveira, V.; Kraka, E.; Cremer, D. Quantitative Assessment of Halogen Bonding Utilizing Vibrational Spectroscopy. *Inorg. Chem.* **2017**, *56*, 488−502.

(29) Oliveira, V.; Kraka, E.; Cremer, D. The Intrinsic Strength of the Halogen Bond: Electrostatic and Covalent Contributions Described by Coupled Cluster Theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 33031−33046.

(30) Setiawan, D.; Kraka, E.; Cremer, D. Strength of the Pnicogen Bond in Complexes Involving Group Va Elements N, P, and As. *J. Phys. Chem. A* **2015**, *119*, 1642−1656.

(31) Tao, Y.; Zou, W.; Jia, J.; Li, W.; Cremer, D. Different Ways of Hydrogen Bonding in Water - Why Does Warm Water Freeze Faster than Cold Water? *J. Chem. Theory Comput.* **2017**, *13*, 55−76.

(32) Cremer, D.; Kraka, E. Generalization of the Tolman Electronic Parameter: the Metal−Ligand Electronic Parameter and the Intrinsic Strength of the Metal−Ligand Bond. *Dalton Trans.* **2017**, *46*, 8323−8338.

(33) Tao, Y.; Zou, W.; Kraka, E. Strengthening of Hydrogen Bonding with the Push-pull Effect. *Chem. Phys. Lett.* **2017**, *685*, 251−258.

(34) Chai, J.-D.; Head-Gordon, M. Long-range Corrected Hybrid Density Functionals with Damped Atom-atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.

(35) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(36) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision E.01.; Gaussian Inc.: Wallingford, CT, 2009.

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

(37) Kraka, E.; Zou, W.; Filatov, M.; Tao, Y.; Grafenstein, J.; Izotov, D.; Gauss, J.; He, Y.; Wu, A.; Konkoli, Z.; et al. *COLOGNE2017*; 2017; see http://www.smu.edu/catco.
(38) *IBM SPSS Statistics for Windows*, version 23.0; IBM Corp.: Armonk, NY, 2015.
(39) Zou, W.; Kalescky, R.; Kraka, E.; Cremer, D. Relating Normal Vibrational Modes to Local Vibrational Modes with the Help of An Adiabatic Connection Scheme. *J. Chem. Phys.* **2012**, *137*, 084114.
(40) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra* (Dover Books on Chemistry); Dover Publications, 2012.
(41) Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent Advances in Wave Function-Based Methods of Molecular-Property Calculations. *Chem. Rev.* **2012**, *112*, 543−631.
(42) Segtnan, V. H.; Šašić, Š.; Isaksson, T.; Ozaki, Y. Studies on the Structure of Water Using Two-Dimensional Near-Infrared Correlation Spectroscopy and Principal Component Analysis. *Anal. Chem.* **2001**, *73*, 3153−3161.
(43) Cremer, D.; Pople, J. General Definition of Ring Puckering Coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354−1358.
(44) Zou, W.; Izotov, D.; Cremer, D. New Way of Describing Static and Dynamic Deformations of the Jahn−Teller Type in Ring Molecules. *J. Phys. Chem. A* **2011**, *115*, 8731−8742.
(45) Haddon, R. C. Comment on the Relationship of the Pyramidalization Angle at a Conjugated Carbon Atom to the σ Bond Angles. *J. Phys. Chem. A* **2001**, *105*, 4164−4165.
(46) McMurry, J. E. *Organic Chemistry*; Brooks Cole, 2011.
(47) Solomons, T. W. G.; Fryhle, C. B.; Snyder, S. A. *Organic Chemistry*, 11th ed.; Wiley, 2013.
(48) Price, C. C. Substitution and Orientation in the Benzene Ring. *Chem. Rev.* **1941**, *29*, 37−67.
(49) Ferguson, L. N. Orientation of Substitution in the Benzene Nucleus. *Chem. Rev.* **1952**, *50*, 47−67.
(50) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; et al. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977−5010.
(51) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1044.
(52) Todeschini, R.; Ballabio, D.; Consonni, V. Distances and other dissimilarity measures in chemometrics. In *Encyclopedia of analytical chemistry*; John Wiley & Sons, Ltd., 2015; pp 1−34.
(53) Mirkin, B. *Clustering: A Data Recovery Approach*, 2nd ed.; Chapman and Hall/CRC, 2012.
(54) Pauling, L. The Nature of the Chemical Bond. IV. The Energy of Single Bonds and the Relative Electronegativity of Atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570−3582.
(55) Allred, A. Electronegativity Values from Thermochemical Data. *J. Inorg. Nucl. Chem.* **1961**, *17*, 215−221.
(56) Brown, A. C.; Gibson, J. XXX.-A Rule for Determining Whether A Given Benzene Mono-derivative Shall Give A Meta-di-derivative or A Mixture of Ortho- and Para-di-derivatives. *J. Chem. Soc., Trans.* **1892**, *61*, 367−369.
(57) James, G. *An Introduction to Statistical Learning With Applications in R*; Springer, 2015.
(58) Chapmann, J. *Machine Learning: Fundamental Algorithms for Supervised and Unsupervised Learning With Real-World Applications*, CreateSpace Independent Publishing Platform; 2017.
(59) Castelvecchi, D. Can We Open the Black Box of AI? *Nature* **2016**, *538*, 20−23.
(60) Liu, S. Where Does the Electron Go? The Nature of Ortho/Para and Meta Group Directing in Electrophilic Aromatic Substitution. *J. Chem. Phys.* **2014**, *141*, 194109.
(61) Fu, R.; Lu, T.; Chen, F. Comparing Methods for Predicting the Reactive Site of Electrophilic Substitution. *Acta Physico-Chimica Sinica* **2014**, *30*, 628−639.
(62) Bader, R. F. W.; Chang, C. Properties of Atoms in Molecules: Electrophilic Aromatic Substitution. *J. Phys. Chem.* **1989**, *93*, 2946−2956.
(63) Remya, G. S.; Suresh, C. H. Quantification and Classification of Substituent Effects in Organic Chemistry: A Theoretical Molecular Electrostatic Potential Study. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20615−20626.
(64) Stasyuk, O. A.; Szatylowicz, H.; Krygowski, T. M.; Fonseca Guerra, C. How Amino and Nitro Substituents Direct Electrophilic Aromatic Substitution in Benzene: An Explanation with Kohn−Sham Molecular Orbital Theory and Voronoi Deformation Density Analysis. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11624−11633.
(65) Hansch, C.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165−195.
(66) Hammett, L. P. The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96−103.
(67) Houk, K. Generalized Frontier Rrbitals of Alkenes and Dienes. Regioselectivity in Diels-Alder Reactions. *J. Am. Chem. Soc.* **1973**, *95*, 4092−4094.
(68) Domingo, L. R.; Aurell, M. J.; Pérez, P.; Contreras, R. Quantitative Characterization of the Local Electrophilicity of Organic Molecules. Understanding the Regioselectivity on Diels-Alder Reactions. *J. Phys. Chem. A* **2002**, *106*, 6871−6875.
(69) Trost, B. M.; Ippen, J.; Vladuchick, W. C. The Regioselectivity of the Catalyzed and Uncatalyzed Diels-Alder Reaction. *J. Am. Chem. Soc.* **1977**, *99*, 8116−8118.
(70) Johansson, C. *Ligand Dependent Regioselectivity in Palladium Mediated Allylic Alkylation*. Ph.D. thesis, University of Gothenburg, Gothenburg, Sweden, 2010.
(71) Ma, S.; Wang, G. Regioselectivity Control by a Ligand Switch in the Coupling Reaction Involving Allenic/Propargylic Palladium Species. *Angew. Chem., Int. Ed.* **2003**, *42*, 4215−4217.
(72) Mitsushige, Y.; Carrow, B. P.; Ito, S.; Nozaki, K. Ligand-controlled Insertion Regioselectivity Accelerates Copolymerisation of Ethylene with Methyl Acrylate by Cationic Bisphosphine Monoxide−palladium Catalysts. *Chem. Sci.* **2016**, *7*, 737−744.
(73) Zuidema, E.; Daura-Oller, E.; Carbo, J. J.; Bo, C.; van Leeuwen, P. W. N. M. Electronic Ligand Effects on the Regioselectivity of the Rhodium-Diphosphine-Catalyzed Hydroformylation of Propene. *Organometallics* **2007**, *26*, 2234−2242.
(74) Tang, S.-Y.; Guo, Q.-X.; Fu, Y. Mechanistic Origin of Ligand-Controlled Regioselectivity in Pd-Catalyzed C-H Activation/Arylation of Thiophenes. *Chem. - Eur. J.* **2011**, *17*, 13866−13876.
(75) Li, M.; Gutierrez, O.; Berritt, S.; Pascual-Escudero, A.; Yeşilçimen, A.; Yang, X.; Adrio, J.; Huang, G.; Nakamaru-Ogiso, E.; Kozlowski, M. C.; et al. Transition-metal-free Chemo- and Regioselective Vinylation of Azaallyls. *Nat. Chem.* **2017**, *9*, 997−1004.
(76) Ohmura, T.; Oshima, K.; Taniguchi, H.; Suginome, M. Switch of Regioselectivity in Palladium-Catalyzed Silaboration of Terminal Alkynes by Ligand-Dependent Control of Reductive Elimination. *J. Am. Chem. Soc.* **2010**, *132*, 12194−12196.
(77) Kumar, M.; Chaudhari, R. V.; Subramaniam, B.; Jackson, T. A. Ligand Effects on the Regioselectivity of Rhodium-Catalyzed Hydroformylation: Density Functional Calculations Illuminate the Role of Long-Range Noncovalent Interactions. *Organometallics* **2014**, *33*, 4183−4191.
(78) Sigman, M. S.; Harper, K. C.; Bess, E. N.; Milo, A. The Development of Multidimensional Analysis Tools for Asymmetric Catalysis and Beyond. *Acc. Chem. Res.* **2016**, *49*, 1292−1301.
(79) Santiago, C. B.; Milo, A.; Sigman, M. S. Developing a Modern Approach To Account for Steric Effects in Hammett-Type Correlations. *J. Am. Chem. Soc.* **2016**, *138*, 13424−13430.
(80) Guo, J.-Y.; Minko, Y.; Santiago, C. B.; Sigman, M. S. Developing Comprehensive Computational Parameter Sets To Describe the Performance of Pyridine-Oxazoline and Related Ligands. *ACS Catal.* **2017**, *7*, 4144−4151.

DOI: 10.1021/acs.jpca.7b08298
J. Phys. Chem. A 2017, 121, 8086−8096

Appendix D

Paper on Mutation Path

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**:

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

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:

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

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

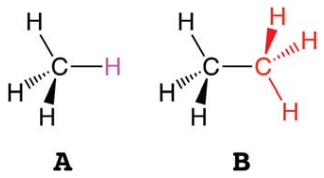


Figure 1. Structure of reactant and product molecule in the mutation path of methane → ethane. [Color figure can be viewed at wileyonlinelibrary.com]

reactant molecule has a symmetry higher than *C*1. 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

$$\mathbf{M}_\lambda=\mathbf{M_A}+\lambda(\mathbf{M_B}-\mathbf{M_A}) \tag{3}$$

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.

$$\mathbf{K}_\lambda=\mathbf{K_A}+\lambda(\mathbf{K_B}-\mathbf{K_A}) \tag{5}$$

The dimension of $\mathbf{K_A}$ differs from that of $\mathbf{K_B}$ if $N_\mathbf{A} \neq N_\mathbf{B}$. To get rid of this inconsistency, we pad $\mathbf{K_A}$ with zeros to the same dimension as $\mathbf{K_B}$. 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 *C*1.

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

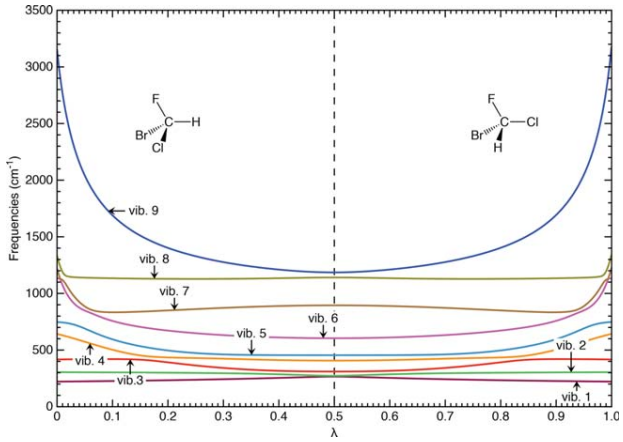


Figure 2. Correlation of normal mode frequencies of (*S*)-bromo-chloro-fluoromethane (left) with that of (*R*)-bromo-chloro-fluoromethane (right) along the mutation path. [Color figure can be viewed at wileyonlinelibrary.com]

vibrational modes along the path have the *a* symmetry. This leads to multiple avoided-crossings between vibrational modes at λ=0.01, 0.02, 0.06, 0.16, 0.5 in the first half of this path. We need to note that at λ=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,

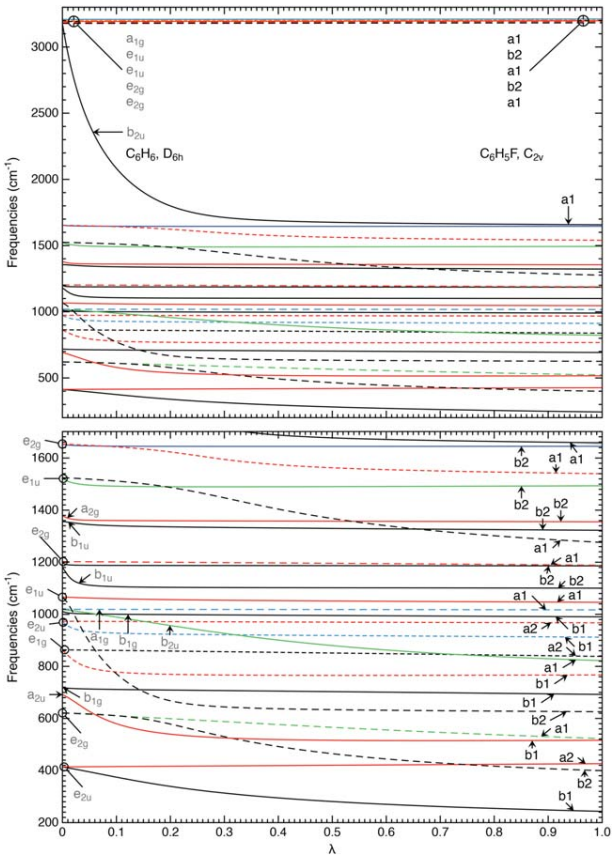


Figure 3. Correlation of normal mode frequencies of benzene (left) with that of fluorobenzene (right) along the mutation path. Irreducible representations of normal modes are labeled on both sides. Top: Diagram of all vibrational frequencies. Bottom: Diagram of the low frequency part (200–1700 cm^{-1}). [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 $3\times3N_\mathbf{B}$ 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

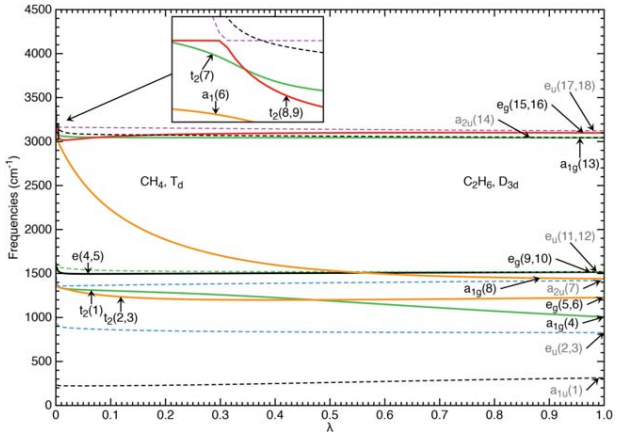


Figure 4. Correlation of normal mode frequencies of methane (left) with that of ethane (right) along the mutation path. Irreducible representations of normal modes are labeled on both sides. Vibrations arising from the growth are plotted with dashed lines and labeled in gray color. [Color figure can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

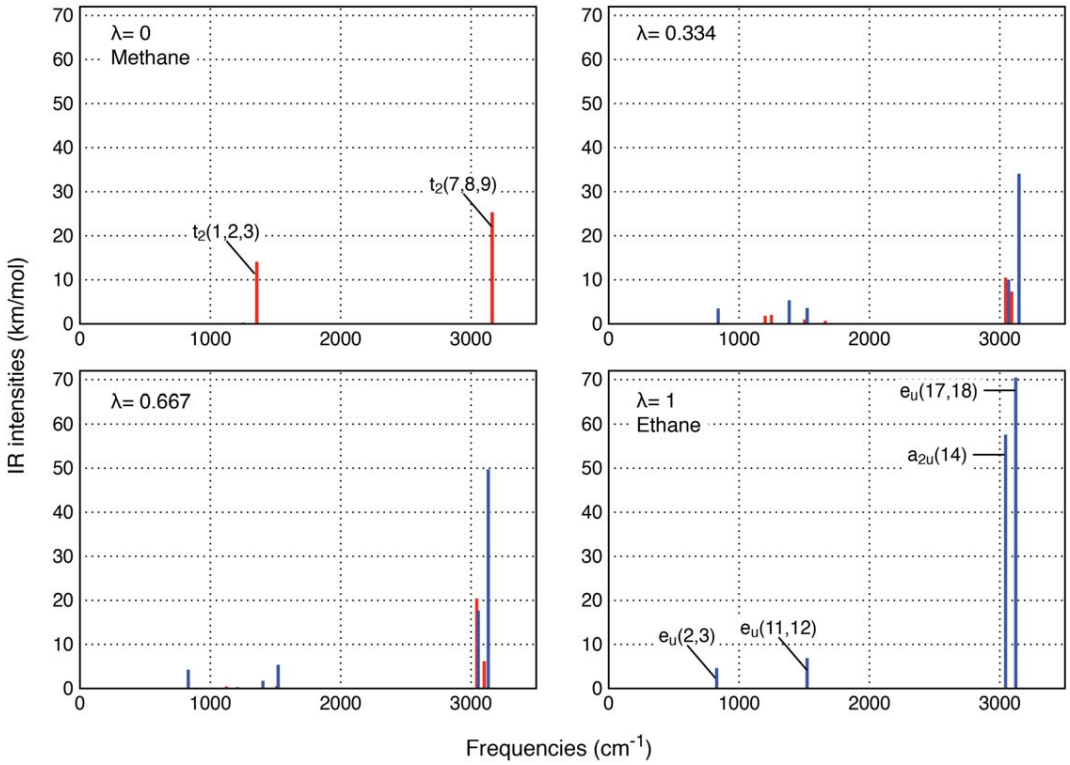


Figure 5. IR intensities of normal vibrational modes along the mutation path from methane to ethane. Four points ($\lambda = 0$, 0.334, 0.667, 1) on the mutation path are shown here. Vibrations from methane are colored in red. Vibrations arising from geometry growth are in blue color. [Color figure can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

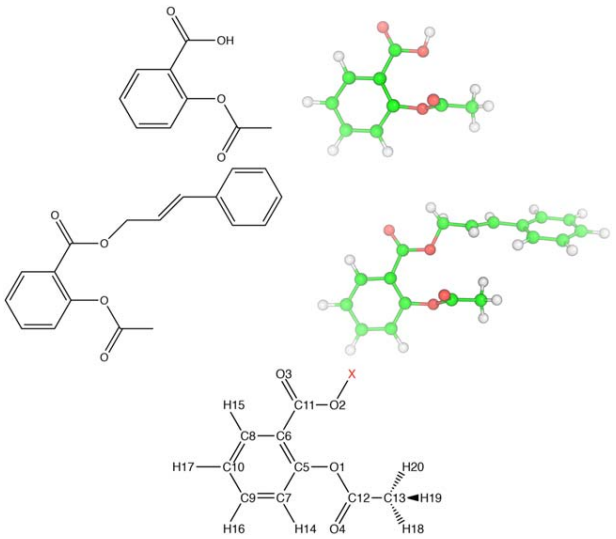


Figure 6. Structure of aspirin (top), aspirin derivative SJ103 (middle) and their common substructure (bottom). [Color figure can be viewed at wileyonlinelibrary.com]

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,

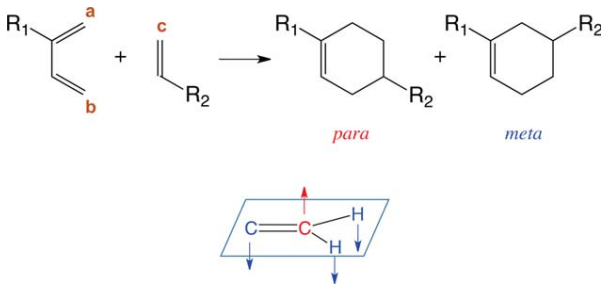


Figure 7. Top: Reaction scheme of Diels-Alder reactions for 2-substituted diene. Bottom: The out-of-plane pyramidalization mode for locations *a*, *b*, and *c*. [Color figure can be viewed at wileyonlinelibrary.com]

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

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).

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.

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.

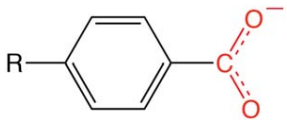


Figure 8. Schematic representation of a p-substituted benzoic acid molecule in its deprotonated state. R is a substituent. [Color figure can be viewed at wileyonlinelibrary.com]

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]

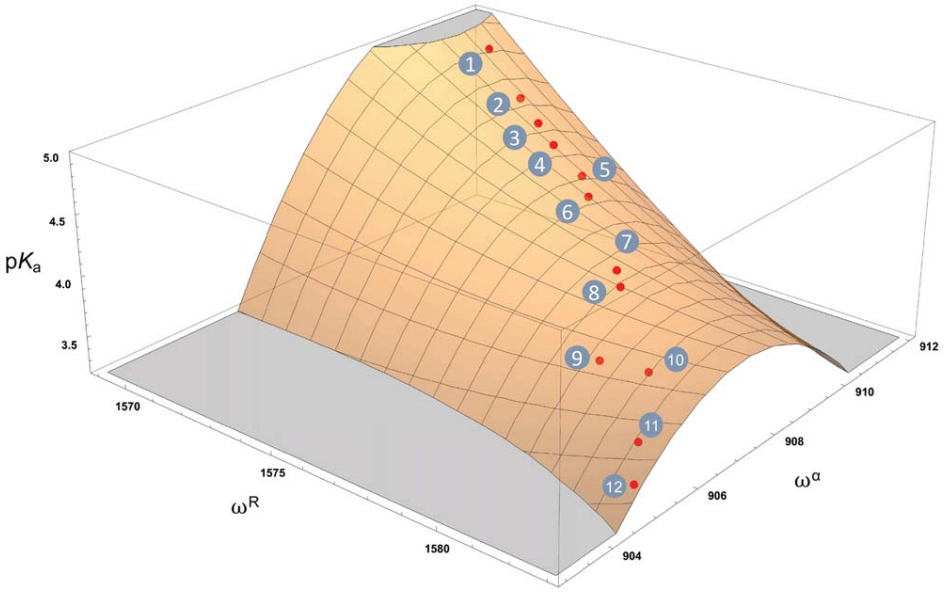


Figure 9. Correlation between pK_a with local CO stretching frequency ω^R and local O-C-O angle bending frequency ω^α on a quadratic surface. [Color figure can be viewed at wileyonlinelibrary.com]

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

Appendix E

Paper on Generalized Subsystem Vibrational Analysis

JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2018, 14, 2558−2569

pubs.acs.org/JCTC

Recovering Intrinsic Fragmental Vibrations Using the Generalized Subsystem Vibrational Analysis

Yunwen Tao,[†] Chuan Tian,[‡] Niraj Verma,[†] Wenli Zou,[§] Chao Wang,[∥] Dieter Cremer,[⊥,†] and Elfi Kraka*[,†]

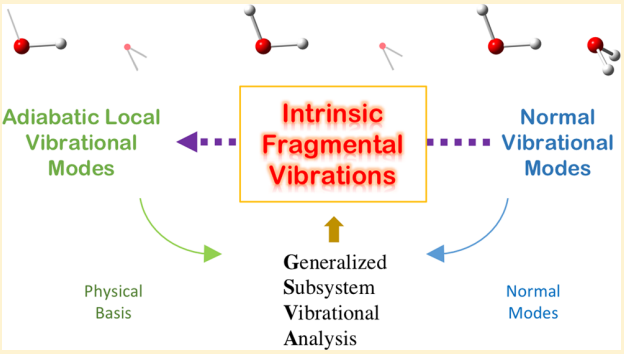
[†]Department of Chemistry, Southern Methodist University, 3215 Daniel Avenue, Dallas, Texas 75275-0314, United States

[‡]Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States

[§]Institute of Modern Physics, Northwest University, Xi'an, Shaanxi 710127, P. R. China

[∥]Institute of Nuclear Energy Safety Technology, Chinese Academy of Sciences, Hefei, Anhui 230031, P. R. China

ABSTRACT: Normal vibrational modes are generally delocalized over the molecular system, which makes it difficult to assign certain vibrations to specific fragments or functional groups. We introduce a new approach, the Generalized Subsystem Vibrational Analysis (GSVA), to extract the intrinsic fragmental vibrations of any fragment/subsystem from the whole system via the evaluation of the corresponding effective Hessian matrix. The retention of the curvature information with regard to the potential energy surface for the effective Hessian matrix endows our approach with a concrete physical basis and enables the normal vibrational modes of different molecular systems to be legitimately comparable. Furthermore, the intrinsic fragmental vibrations act as a new link between the Konkoli−Cremer local vibrational modes and the normal vibrational modes.



1. INTRODUCTION

Vibrational spectroscopy is a powerful tool for structure elucidation. Raman and infrared (IR) spectroscopy can be used not only for the assignment of characteristic peaks to identify functional groups but also to characterize the electronic structure of the targeted chemical system. With the rapid development of quantum chemical methods based on quantum mechanics (QM), the simulation of vibrational spectra has been made feasible and becomes a complementary tool in structural determination. The vibrational frequencies of any chemical species can be calculated from the normal mode analysis (NMA) by solving the Wilson equation of vibrational spectroscopy.[1]

However, normal modes extend over the whole molecule, which complicates the analysis and interpretation of vibrations for large polyatomic molecules and molecules in solution or other media being described by a multiscale model, i.e., QM/MM. For example, in the water dimer, if one wants to compare the normal vibrational modes of either the H-bond donor water or the acceptor water with the vibrations of a single water molecule in gas phase in order to characterize the influence of hydrogen bonding, one has to consider that the normal mode vectors in the two systems are of different lengths (18 versus 9).

Obtaining normal modes being projected into a targeted subsystem or fragment would be the natural way to solve the above problem and allow the normal modes to be intrinsically comparable among different molecular systems.

Many efforts have been made in this direction. Head proposed a strategy to calculate the vibrations for adsorbates on surfaces by diagonalizing the partial Hessian for atoms of adsorbates.[2] His contribution fostered the work of Li and Jensen, who developed the partial Hessian vibrational analysis (PHVA) method,[3] where a subblock of the Hessian matrix is diagonalized and all atoms except the subsystem are assigned an infinitely large atomic mass. This approach has been applied by Besley and co-workers to calculate C═O stretching and C−H stretching vibrations in organic molecules.[4,5] Ghysels and co-workers proposed the mobile block Hessian (MBH) approach as an extension to PHVA in order to calculate "localized" normal vibrational modes for partially optimized molecular systems.[6−10] The MBH method allows one to calculate the vibrations of a subsystem which has been fully optimized, while the remaining parts of the system are treated as rigid bodies being allowed to translate and rotate. Thus, the computational costs can be reduced because one does not need to calculate the full Hessian matrix of the whole system. The MBH method has been implemented in quantum chemical packages including ADF and Q-Chem. Woodcock and co-workers developed another method called vibrational subsystem analysis (VSA),[11] partitioning a large system into a subsystem and its environment. The vibrational modes led by the subsystem are

Received: November 18, 2017
Published: April 10, 2018

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

calculated, and the environment follows the motions of the subsystem in an adiabatic way. However, we need to note that VSA was initially developed to couple the global motion to a local subsystem in a large molecule, e.g., protein, in order to estimate the free energy contribution from subsystems. Zheng and cowork applied VSA to obtain approximate low-frequency normal modes of proteins for conformational sampling.[12] VSA was also used in another work for mapping the full Hessian matrix onto a coarse-grained scale for macromolecules.[13] A review article by Ghysels and co-workers compares the PHVA, MBH, and VSA methods.[14] Jacob and Reiher developed a special approach to localize normal modes contributing to certain bands with the help of a defined criterion.[15] Their method is tailored to polypeptides and proteins. Recently, Huix-Rotllant and co-worker provided a procedure to localize normal modes of fragment(s) by taking the submatrices of the full Hessian as local Hessian matrices.[16] The advantage of this approach is that besides diagonalizing the local Hessian matrices, the corresponding eigenvectors are used for the transformation of the full Hessian. Thus, the information on the full Hessian can be utilized.

However, if one tries to apply the methods mentioned above in order to obtain intrinsically comparable normal vibrations of a subsystem, the following problems must be resolved:

- As shown in Figure 1, all methods involve the direct partitioning of the full Hessian matrix into the red sub-block for the targeted subsystem or fragment and another blue sub-block for the remaining part as its environment. Once the Hessian matrix is partitioned, the information describing the interaction between the subsystem and the environment contained in the green sub-block is damaged or even eliminated. This casts doubt on the usefulness of these localized normal modes.
- The so-called "local Hessian matrix" or its counterpart with different names pertinent to the subsystem is not in a proper form prepared for characterizing the normal vibrational modes of the subsystem. If the total number of translations and rotations for the subsystem is k, then such a local Hessian matrix is expected to have and only have k eigenvalues of zero.
- As normal modes are delocalized over the whole system, any attempt to compare the vibrational frequencies of the localized subsystem modes with the frequencies of the normal modes of the whole system either from *ab initio* calculations or measured vibrational spectra is not appropriate. This implies that verifying the results of the localized normal modes with the help of vibrational frequencies calculated from the full Hessian is problematic. A more reasonable approach for result validation is thus desired.
- Some methods were tested against selected examples containing many C−H, N−H, and C═O bonds. The stretching modes of these bonds are in nature localized to these fragments, and they can contribute to more significant infrared intensities than other types of vibrational motions due to large dipole changes. This will give the illusion that the localized normal modes have been accurately determined.
- Some approaches partition the complete system according to certain rules, e.g., containing the peptide bond unit in the subsystem. However, a generally applicable approach is expected to allow arbitrary partitioning of the whole system into the subsystem and its environment.
- The purpose of localizing of normal vibrational modes should be re-evaluated. While theoretical chemists have developed various kinds of localized properties or models including localized orbitals,[17−21] localized atomic charges,[22−27] and even localized electron densities[28,29] in order for simplification, comparison, and analysis, we expect that localized normal modes should be able to serve for similar tasks instead of assisting the assignment of the absorption peaks from vibrational spectra for different functional groups.

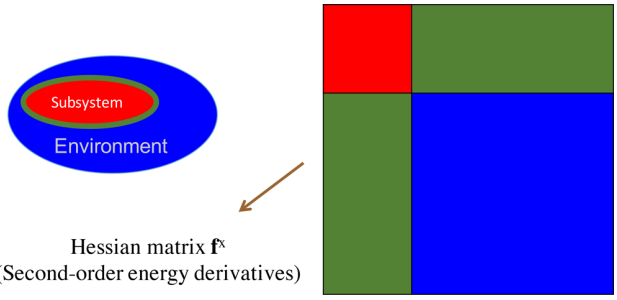


Figure 1. Schematic diagram of partitioning of the Hessian matrix.

We start from a different ansatz in order to obtain intrinsically comparable normal vibrations of a subsystem or fragment. When calculating the normal modes of any chemical system, three ingredients are required by the Wilson equation: Cartesian coordinates **R**, atomic masses **M**, and the Hessian matrix in Cartesian coordinates $\mathbf{f}^x$. As long as this system is at a stationary point on the potential energy surface with all three ingredients available, the Wilson equation can be solved accordingly. If one is interested in obtaining the normal vibrational modes for a subsystem/fragment, it is obvious that the Cartesian coordinates **R** and atomic masses **M** of this part should stay the same. The problem to be solved is then how to obtain the "effective Hessian matrix" that is reasonable and physically sound for the subsystem. Only on this basis, the resulting localized normal vibrational modes of the subsystem can be then obtained from solving the Wilson equation of vibrational spectroscopy using the effective Hessian matrix.

In this work, we introduce a new approach called Generalized Subsystem Vibrational Analysis (GSVA) that is based on a physically solid "effective Hessian matrix" which preserves the information on the curvature of the potential energy surface for the subsystem/fragment as it is in the whole system with a full Hessian matrix. The normal vibrational modes calculated by this approach are therefore called "intrinsic fragmental vibrations". These vibrations are not constructed from an arbitrary model, instead they are recovered specifically for the subsystem from the full Hessian matrix. Noteworthy is that a distinction should be made between the GSVA method introduced in this work and the VSA method[11] developed by Woodcock. This paper is structured in the following way: The theory of the GSVA is derived and described first. After summarizing the Computational Details section, in the Results and Discussion section, six different examples for calculating the intrinsic fragmental vibrations of their subsystems are discussed. The conclusions, along with some notes of applying and implementing GSVA, are given in the last section.

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

2. METHODOLOGY

The normal vibrational modes of a molecular system can be calculated from the solution of the Wilson equation of vibrational spectroscopy[1] based on the Hessian matrix $\mathbf{f}^x$ in Cartesian coordinates collecting the second-order derivatives of the energy with regard to the displacement of atomic nuclei. The dimension of $\mathbf{f}^x$ is $3N \times 3N$, where N is the number of atoms in the system.

As the translation and rotation of the system render no change to the potential energy, matrix $\mathbf{f}^x$ is singular and has K zero eigenvalues, K takes the value of 5 for linear molecules or 6 for nonlinear molecules. We are, in general, only interested in the nonzero eigenvalues λ_μ (collected in the diagonal matrix $\mathbf{\Lambda}$) as well as their eigenvectors $\mathbf{c}_\mu$ (collected in matrix $\mathbf{C}$), as shown in eq 1. The dimensions of $\mathbf{C}$ and $\mathbf{\Lambda}$ are $3N \times (3N - K)$ and $(3N - K) \times (3N - K)$, respectively.

$$\mathbf{f}^x\mathbf{C} = \mathbf{C}\mathbf{\Lambda} \tag{1}$$

As each eigenvector $\mathbf{c}_\mu$ in $\mathbf{C}$ is orthonormalized, $\mathbf{C}^T\mathbf{C} = \mathbf{I}_{3N-K}$, eq 1 can be rewritten into

$$\mathbf{C}^T\mathbf{f}^x\mathbf{C} = \mathbf{\Lambda} \tag{2}$$

and

$$\mathbf{C}\mathbf{C}^T\mathbf{f}^x\mathbf{C}\mathbf{C}^T = \mathbf{C}\mathbf{\Lambda}\mathbf{C}^T = \mathbf{f}^x \tag{3}$$

We define a new matrix $\mathbf{Q} = \mathbf{C}\mathbf{C}^T$. As $\mathbf{Q}^2 = \mathbf{C}\mathbf{C}^T\mathbf{C}\mathbf{C}^T = \mathbf{C}\mathbf{I}\mathbf{C}^T = \mathbf{C}\mathbf{C}^T = \mathbf{Q}$, $\mathbf{Q}$ is thus a projection matrix in the dimension of $3N \times 3N$. An interesting equation results as follows:

$$\mathbf{Q}\mathbf{f}^x\mathbf{Q} = \mathbf{f}^x \tag{4}$$

The $3N - K$ eigenvectors collected in matrix $\mathbf{C}$ span the full internal vibration space; thus, when projection operator $\mathbf{Q}$ multiplies $\mathbf{f}^x$ from the left to the right of $\mathbf{f}^x$, $\mathbf{f}^x$ is not changed.

This special property of the Hessian matrix $\mathbf{f}^x$ can be extended to any other projection matrix, as long as this projection matrix can span the full internal vibration space. We choose to use the internal coordinates to span the same space, as translations and rotations can be simply excluded.

For a molecule system being composed of N atoms, we can use $3N - K$ internal coordinates to specify its geometry. The internal coordinates are related to the Cartesian coordinates via the Wilson B matrix[1]

$$\mathbf{B} = \frac{\partial \mathbf{q}}{\partial \mathbf{x}} \tag{5}$$

where $\mathbf{x}$ are the Cartesian coordinates and $\mathbf{q}$ are the internal coordinates. For the above nonredundant set of $3N - K$ internal coordinates, the corresponding Wilson B matrix $\mathbf{B}$ has the dimension of $(3N - K) \times 3N$. As matrix $\mathbf{B}$ is rectangular, its Moore–Penrose inverse matrix $\mathbf{B}^+$ is calculated by

$$\mathbf{B}^+ = \mathbf{B}^T(\mathbf{B}\mathbf{B}^T)^{-1} \tag{6}$$

so that

$$\mathbf{B}\mathbf{B}^+ = \mathbf{I}_{3N-K} \tag{7}$$

where the trace of the identity matrix $\mathbf{I}_{3N-K}$ as $3N - K$.

We define a matrix $\mathbf{A}$

$$\mathbf{A} = \mathbf{B}^+\mathbf{B} \tag{8}$$

as $\mathbf{A}^2 = \mathbf{B}^+\mathbf{B}\mathbf{B}^+\mathbf{B} = \mathbf{B}^+\mathbf{I}\mathbf{B} = \mathbf{B}^+\mathbf{B} = \mathbf{A}$; matrix $\mathbf{A}$ is also a projection matrix having the similar properties as matrix $\mathbf{Q}$ in eq 4, leading to the following equations:

$$\mathbf{f}^x = \mathbf{A}\mathbf{f}^x\mathbf{A} \tag{9}$$

$$\mathbf{f}^x = (\mathbf{B}^+\mathbf{B})\mathbf{f}^x(\mathbf{B}^+\mathbf{B}) \tag{10}$$

According to the properties of pseudoinverse $\mathbf{B}^+$, we have

$$(\mathbf{B}^+\mathbf{B})^T = \mathbf{B}^+\mathbf{B} \tag{11}$$

Equation 10 can be rewritten into

$$\mathbf{f}^x = (\mathbf{B}^+\mathbf{B})^T\mathbf{f}^x(\mathbf{B}^+\mathbf{B}) \tag{12}$$

$$\mathbf{f}^x = \mathbf{B}^T(\mathbf{B}^+)^T\mathbf{f}^x\mathbf{B}^+\mathbf{B} \tag{13}$$

$$\mathbf{f}^x = \mathbf{B}^T(\mathbf{B}^T)^+\mathbf{f}^x\mathbf{B}^+\mathbf{B} \tag{14}$$

then

$$\mathbf{f}^x = \mathbf{B}^T(\mathbf{B}(\mathbf{f}^x)^+\mathbf{B}^T)^{-1}\mathbf{B} \tag{15}$$

Noteworthy is that eq 15 is a more general form of the equations for $\mathbf{f}^x$ above. Furthermore, this equation offers an opportunity to obtain the effective Hessian matrix for a fragment or subsystem within the whole system.

Suppose that within the molecular system with N atoms, a subsystem has n atoms ($n < N$). The geometry of this subsystem can be specified by $3n - k$ internal coordinates (k = 5 or 6 depending on whether its geometry is linear or nonlinear). The Wilson B matrix for these $3n - k$ internal coordinates in the complete system can be calculated as $\mathbf{B}'$ with the dimension of $(3n - k) \times 3N$. In the subsystem, the corresponding Wilson B matrix for the same set of internal coordinates is calculated as $\mathbf{B}'_{sub}$ with the dimension of $(3n - k) \times 3n$. In order to simplify the analysis, we rearrange the labels of n atoms of the subsystem within the whole system, so that the first n atoms denote the subsystem. It is obvious that matrix $\mathbf{B}'_{sub}$ corresponds to the first $3n$ columns of matrix $\mathbf{B}'$, while the elements of the rest $3(N - n)$ columns in $\mathbf{B}'$ are simply zero. We define an effective Hessian matrix $\mathbf{f}^x_{sub}$ for the subsystem with the help of eq 15

$$\mathbf{f}^x_{sub} = \mathbf{B}'^T_{sub}(\mathbf{B}'(\mathbf{f}^x)^+\mathbf{B}'^T)^{-1}\mathbf{B}'_{sub} \tag{16}$$

where $(\mathbf{f}^x)^+$ is the Moore–Penrose inverse of $\mathbf{f}^x$. Here, $\mathbf{f}^x_{sub}$ is a symmetric matrix in the dimension of $3n \times 3n$, and more importantly, it has exactly k zero eigenvalues.

However, we need to note that eqs 9, 10, and 12–14 cannot be used for this purpose, namely

$$\mathbf{f}^x_{sub} \neq (\mathbf{B}'^+\mathbf{B}'_{sub})^T\mathbf{f}^x(\mathbf{B}'^+\mathbf{B}'_{sub}) \tag{17}$$

because the Wilson B matrix and related pseudoinverse in these equations can no longer span the full vibration space for the whole system, only for the subsystem.

The effective Hessian matrix $\mathbf{f}^x_{sub}$ for the subsystem can be directly used for normal mode analysis by solving the Wilson equation of vibrational spectroscopy given its Cartesian coordinates and atomic masses as it can be done for the whole system based on full Hessian matrix $\mathbf{f}^x$. Noteworthy is that in the process of obtaining $\mathbf{f}^x_{sub}$ no partitioning/sub-blocking of the original Hessian matrix $\mathbf{f}^x$ is introduced. Instead, the full Hessian matrix $\mathbf{f}^x$ is projected into the unique internal vibrational space of the subsystem in eq 16, which strikingly differentiates our approach from others.[2,3,6,15,16]

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558–2569

Furthermore, it is necessary to evaluate the physical basis and correctness of the effective Hessian matrix derived in this work, if intrinsically comparable normal vibrations are desired on the basis of such an effective Hessian matrix. The seemingly most straightforward approach to validate our model is to compare the normal mode frequencies of the subsystem based on effective Hessian matrix $\mathbf{f}^x_{sub}$ and those based on the original Hessian matrix $\mathbf{f}^x$ for the whole system. However, one needs to be careful that normal modes are delocalized over the system in question, and it is not appropriate to compare the normal modes within the subsystem and those modes beyond it.[30] In this work, we choose to calculate and compare the local vibrational modes[31−35] proposed by Konkoli and Cremer for $\mathbf{f}^x_{sub}$ and $\mathbf{f}^x$ because these local modes have been proved as the only and unique local equivalents of normal vibrational modes in terms of internal coordinates[30] which can be directly compared among different molecular systems, and they have been used to quantify the intrinsic strength of chemical bonding[36−42] as well as to characterize the local properties of the electronic structure.[43,44] The characterization of local vibrational modes including related local mode force constants and local mode frequencies is called local mode analysis. For each local vibrational mode driven by a specific internal coordinate as the "leading parameter", we can calculate the corresponding local force constant or its synonym as adiabatic force constant k^a_n as well as the local vibrational frequency ω^a_n. These two quantities can be related with the help of the Wilson G matrix[1,45]

$$(\omega^a_n)^2 = \frac{1}{4\pi^2 c^2} k^a_n G_{nn} \tag{18}$$

For the purpose of validating the effective Hessian matrix, calculating the local force constants k^a_n is sufficient.

In this work, we take a simplified form[30,46] of calculating the adiabatic force constant k^a_n by

$$\frac{1}{k^a_n} = \mathbf{b}(\mathbf{f}^x)^+\mathbf{b}^T \tag{19}$$

where $\mathbf{f}^x$ is the Hessian matrix for the whole system in Cartesian coordinates, and its Moore−Penrose inverse is denoted as $(\mathbf{f}^x)^+$. Row vector $\mathbf{b}$ is the Wilson B matrix for an internal coordinate parameter q_n (e.g., bond stretching, angle bending, dihedral torsion, etc.) within the subsystem leading this local mode. Here, $\mathbf{b}$ is in the dimension of $1 \times 3N$.

Based on the effective Hessian matrix $\mathbf{f}^x_{sub}$ of the subsystem, its adiabatic force constant $k^a_{n,sub}$ of the local mode led by the same internal coordinate q_n is calculated by

$$\frac{1}{k^a_{n,\mathrm{sub}}} = \mathbf{b}_{sub}(\mathbf{f}^x_{sub})^+\mathbf{b}^T_{sub} \tag{20}$$

where $\mathbf{b}_{sub}$ is the first $3n$ elements of $\mathbf{b}$ in eq 19, and $(\mathbf{f}^x_{sub})^+$ is the Moore−Penrose inverse of $\mathbf{f}^x_{sub}$. Equation 20 can be expanded by substituting $\mathbf{f}^x_{sub}$ using eq 16

$$\frac{1}{k^a_{n,\mathrm{sub}}} = \mathbf{b}_{sub}(\mathbf{B}'^T_{sub}(\mathbf{B}'(\mathbf{f}^x)^+\mathbf{B}'^T)^{-1}\mathbf{B}'_{sub})^+\mathbf{b}^T_{sub} \tag{21}$$

According to the properties of the pseudoinverse, eq 21 can be rewritten as

$$\frac{1}{k^a_{n,\mathrm{sub}}} = \mathbf{b}_{sub}\mathbf{B}'^+_{sub}(\mathbf{B}'^T_{sub}(\mathbf{B}'(\mathbf{f}^x)^+\mathbf{B}'^T)^{-1})^+\mathbf{b}^T_{sub} \tag{22}$$

$$\frac{1}{k^a_{n,\mathrm{sub}}} = \mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}'(\mathbf{f}^x)^+\mathbf{B}'^T(\mathbf{B}'^T_{sub})^+\mathbf{b}^T_{sub} \tag{23}$$

$$\frac{1}{k^a_{n,\mathrm{sub}}} = \mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}'(\mathbf{f}^x)^+\mathbf{B}'^T(\mathbf{B}'^+_{sub})^T\mathbf{b}^T_{sub} \tag{24}$$

then

$$\frac{1}{k^a_{n,\mathrm{sub}}} = (\mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}')(\mathbf{f}^x)^+(\mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}')^T \tag{25}$$

The calculation of the matrix product of $\mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}'$ is visualized by the Falk diagram shown in Figure 2.

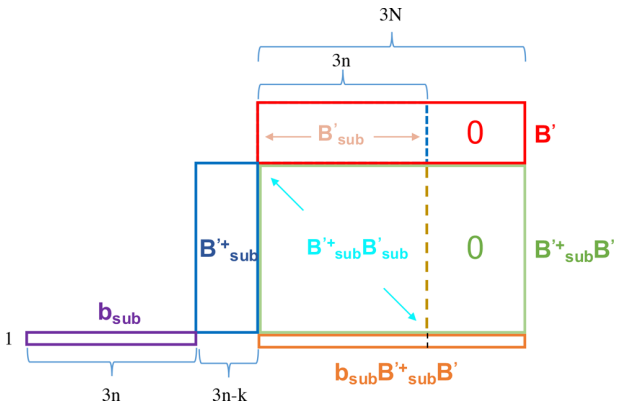


Figure 2. Falk diagram of matrix multiplication for $\mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}'$.

In matrix $\mathbf{B}'$, the block of the first $3n$ columns is matrix $\mathbf{B}'_{sub}$, while the elements in the remaining $3(N-n)$ columns are zeros. The multiplication of $\mathbf{B}'^+_{sub}$ with $\mathbf{B}'$ leads to a projection matrix $\mathbf{B}'^+_{sub}\mathbf{B}'_{sub}$ in the first $3n$ columns of $\mathbf{B}'^+_{sub}\mathbf{B}'$ and zeros in the remaining $3(N-n)$ columns.

As the Wilson B matrix $\mathbf{B}'_{sub}$ (or $\mathbf{B}'$) collects the nonredundant set of $3n-k$ internal coordinate parameters describing the geometry of the subsystem, the projection matrix $\mathbf{B}'^+_{sub}\mathbf{B}'_{sub}$ spans the complete internal coordinate space and also the internal vibration space of the subsystem. The Wilson B matrix row vector $\mathbf{b}_{sub}$ for any internal coordinate parameter (no matter whether it is included in the set of the $3n-k$ parameters or not) in the subsystem can be expressed as a linear combination of $3n-k$ row vectors in $\mathbf{B}'_{sub}$. So we get

$$\mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}'_{sub} = \mathbf{b}_{sub} \tag{26}$$

which constitutes the first $3n$ elements of the row vector $\mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}'$, and the rest of the $3(N-n)$ elements are zeros. Also, we have

$$\mathbf{b}_{sub}\mathbf{B}'^+_{sub}\mathbf{B}' = \mathbf{b} \tag{27}$$

where $\mathbf{b}$ is from eq 19. Then, eq 25 can be simplified as

$$\frac{1}{k^a_{n,\mathrm{sub}}} = \mathbf{b}(\mathbf{f}^x)^+\mathbf{b}^T \tag{28}$$

Also interesting is that

$$k^a_{n,\mathrm{sub}} = k^a_n \tag{29}$$

which means the local mode analysis with regard to any internal coordinate in the subsystem based on the effective Hessian matrix $\mathbf{f}^x_{sub}$ is equivalent to the local mode analysis for the same internal coordinate based on the full Hessian matrix $\mathbf{f}^x$.

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

Furthermore, as the adiabatic force constant k_n^a characterizes the curvature of the Born−Oppenheimer potential energy surface (PES) given in a specific direction defined by the internal coordinate as the leading parameter,[41] the curvature of the PES driven by any one of the internal coordinates in the subsystem within the whole system is retained in the effective Hessian matrix. In other words, the standalone subsystem with effective Hessian matrix "feels" exactly the same curvature of the PES with regard to the internal local vibrations as it is within the whole system based on the full Hessian matrix. In this way, the underlying physical nature of the vibrations of the subsystem calculated based on $\mathbf{f}_{sub}^x$ is kept invariant, and this gives our approach the capability and advantage to characterize the intrinsically comparable normal vibrations of subsystems or fragments in any molecular system. We call these intrinsically comparable normal vibrations the intrinsic fragmental vibrations.

3. COMPUTATIONAL DETAILS

In this work, all *ab initio* calculations including geometry optimization and Hessian evaluation were performed using the Gaussian 09 package.[47] The dimers, trimers, and monomers of water and ammonia molecules were calculated at the ωB97X-D/6-311++G(d,p) level;[48−51] The hydrogen disulfide molecule and the hydrogen disulfide-water cluster were calculated at the B3LYP/6-31G(d,p) level of theory;[52−55] The methane molecule and methane-C_{60} complex were calculated using the Minnesota hybrid functional M06-2X with Pople's 6-31G(d,p) basis set.[56] Grimme's empirical D3 dispersion correction was added to the nuclear repulsion force.[57] The formaldehyde molecule and formaldehyde-nanotube complex were calculated at the B3LYP/6-31G(d,p) level with Grimme's empirical D3 dispersion correction with Becke−Johnson (BJ) damping.[58] The propane molecule along with a reference methane molecule were calculated with Hartree−Fock theory[59] using 6-31G(d,p) basis set. For the above density functional theory (DFT) calculations, the UltraFine integration grid was used, and all systems were optimized to local minima with tight convergence criteria.

The calculations of the effective Hessian matrices, local mode analysis, and normal mode analysis were carried out with the program package COLOGNE2017.[60]

4. RESULTS AND DISCUSSION

In the following section, we discuss the results of the intrinsic fragmental vibrations in six different molecular systems. For each subsystem or fragment having n atoms, we calculated its $3n - k$ intrinsic fragmental vibrational frequencies based on the effective Hessian matrix $\mathbf{f}_{sub}^x$ in the dimension of $3n \times 3n$.

Then, we calculated the frequencies of normal vibrational modes or intrinsic fragmental vibrations of the same subsystem in the gas phase or other chemical systems, respectively, for all six examples in order to demonstrate that the intrinsic fragmental vibrations have the advantage of being directly compared and analyzed laterally among different systems.

We have also calculated the local mode force constants of the leading internal coordinate parameters q_n within the fragment/subsystem based on both the full Hessian $\mathbf{f}^x$ and the effective Hessian $\mathbf{f}_{sub}^x$ in order to verify the physical relevance of these intrinsic fragmental normal vibrations.

4.1. Water Dimer $(H_2O)_2$. The first example is the water dimer which has a hydrogen bond between two water molecules, one serving as the H-bond donor and the other as the H-bond acceptor (Figure 3). One water molecule has three

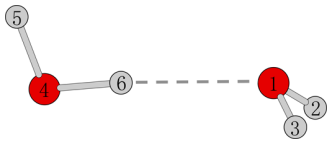


Figure 3. Water dimer structure in C_s symmetry. Red balls represent oxygen atoms, and gray balls represent hydrogen atoms.

normal vibrational modes, including the H−O−H angle bending, symmetric O−H stretching, and asymmetric O−H stretching with increasing vibrational frequencies (Table 1).

Table 1. Comparison of Normal Mode Frequencies of Water Monomers

No.	Donor[a] (cm^{-1})	Acceptor[a] (cm^{-1})	H_2O[b] (cm^{-1})
1	1531	1607	1609
2	3775	3891	3903
3	3976	3997	4012

[a]Columns "Donor" and "Acceptor" denote the intrinsic fragmental vibrational frequencies of donor and acceptor water molecules based on their effective Hessian matrices, respectively. [b]Column "H_2O" collects the normal mode frequencies of a water molecule in gas phase as the reference.

The introduction of another water molecule in a dimer structure brings in an addition nine vibrational modes. These nine vibrational modes include the three internal vibrations of the second water molecule, three relative rotations, and three relative translations between these two water molecules. However, these nine new vibrations will mix with each other. Furthermore, the original three vibrations of the first water molecule are also mixed in, which potentially hinders the analysis of normal vibrational modes of either water molecule in the dimer.

Within the framework of GSVA, the donor/acceptor water is taken as a subsystem. Its effective Hessian matrix $\mathbf{f}_{sub}^x$ can be extracted by choosing a nonredundant set of three internal coordinate parameters according to eq 16. In a water molecule, we can choose two O−H distances and the H−O−H angle as a complete nonredundant parameter set. Therefore, matrix $\mathbf{B}'$ takes the dimension of 3×18, while matrix $\mathbf{B}'_{sub}$ is in the 3×9 dimension. Or we can use two O−H distances and the H−H distance to construct another parameter set, although the H−H distance does not imply H−H bonding in a water molecule. These two sets of parameters give two identical effective Hessian matrices, which reveals the flexibility and robustness of the GSVA approach; namely, this approach does not depend on the choice of the nonredundant internal parameter set. As long as the chosen set of parameters can unambiguously specify the geometry of the subsystem, GSVA will work.

Table 1 lists the three normal vibrational frequencies calculated by solving the Wilson equation[1] based on the effective Hessian matrices for the donor and acceptor waters. The resulting three normal modes calculated by GSVA are the unique counterparts of the normal modes of the water molecule in the gas phase. Correlating the intrinsic fragmental vibrational modes with those of the water molecule in the gas phase leads to the normal vibrational frequency ordering shown in Table 1. The acceptor water shows smaller deviations from the reference

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

frequencies compared to the donor water. In the acceptor water, the first normal mode dominated by the H−O−H angle bending has the deviation of only 2 cm^{-1}, while the donor water's deviation is 78 cm^{-1}. This can be explained by the fact that the angle bending mode of the acceptor water is not affected by the formation of a hydrogen bond, while the angle bending of the donor water is hindered by this hydrogen bond. The larger deviations for the symmetric and asymmetric O−H stretching modes for the donor water are also caused by the hydrogen bonding which weakens the donor O−H covalent bond.[40,42]

For the purpose of validating the intrinsic fragmental vibrational modes and their frequencies, we calculated the local mode force constants of the O−H bond stretching and H−O−H angle bending modes in the donor and acceptor waters based on the effective Hessian matrix and the full Hessian matrix using eqs 19 and 20. The comparison of the local mode force constants in Table 2 shows that the values of $k^a_{n,sub}$ and k^a_n for local mode parameters within the subsystem are the same.

Table 2. Comparison of Local Mode Force Constants Based on Effective Hessian Matrix $\mathbf{f}^x_{sub}$ and Full Hessian Matrix $\mathbf{f}^x$ for Water Molecules

No.[a]	q_n[b]	$k^a_{n,sub}$	k^a_n
D-1	R(4−5)	8.691	8.691
D-2	R(4−6)	7.984	7.984
D-3	α(5−4−6)	0.596	0.596
A-1	R(1−2)	8.624	8.624
A-2	R(1−3)	8.624	8.624
A-3	α(2−1−3)	0.652	0.652

[a]In the "No." column, "D" denotes donor water, while "A" denotes acceptor water. [b]For internal coordinate q_n, parameter "R" stands for bond stretching. Unit for local mode force constant is mdyn/Å, while "α" is for angle bending and corresponding unit for local mode force constant is mdyn×Å/rad^2.

This clearly reveals that the effective Hessian matrices $\mathbf{f}^x_{sub}$ calculated for the donor/acceptor water molecules have retained the curvature of the PES of the whole system with regard to any internal coordinate in donor or acceptor water, respectively. Accordingly, these fragmental vibrational modes based on $\mathbf{f}^x_{sub}$ are thus intrinsic.

4.2. Ammonia Trimer $(NH_3)_3$. The ammonia trimer ring shown in Figure 4 has C_{3h} symmetry. All three ammonia molecules connected via hydrogen bonds are identical with regard to geometry as well as electronic structure. In this example, we want to take one ammonia molecule as the subsystem and obtain its intrinsic fragmental vibrations.

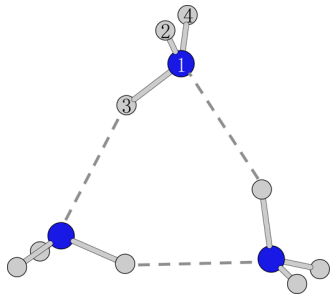


Figure 4. Ammonia trimer ring structure with C_{3h} symmetry. Blue balls represent nitrogen atoms, and gray balls represent hydrogen atoms.

For each ammonia molecule having four atoms, we need six internal coordinates to determine its geometry. The set of three N−H bonds and three H−N−H angles is the easiest option. But we can also use three N−H bonds and three H−H distances as a valid set for GSVA. Therefore, matrix $\mathbf{B}'$ has the dimension of 6 × 36, and matrix $\mathbf{B}'_{sub}$ has the dimension of 6 × 12.

Table 3 lists the intrinsic fragmental vibrational frequencies of ammonia in comparison with the normal mode frequencies of an ammonia molecule in the gas phase. While the symmetry of ammonia in the gas phase is reduced from C_{3v} to C_s for the ammonia in the trimer system shown in Figure 4, normal modes Nos. 2−3 and Nos. 5−6 lose their 2-fold degeneracy leading to the splitting in the vibrational frequency values. We find vibrations Nos. 2 and 4 of ammonia in the trimer have their frequency differences larger than 100 cm^{-1} when compared with the reference ammonia in the gas phase. Normal mode No. 2 is dominated by the rocking of the H3 atom, and normal mode No. 4 is basically the symmetric stretching of all three N−H bonds. As bond N1−H3 directly participates in the hydrogen bonding, the above two vibrational modes will be affected accordingly. However, the smallest difference in the vibrational frequency is found for No. 6 as 13 cm^{-1}. This vibration mode is dominated by the asymmetric stretching of bonds N1−H2 and N1−H4, which are not directly involved in hydrogen bonding.

Table 3. Comparison of Normal Mode Frequencies of Ammonia Monomers

No.	NH_3 in trimer (cm^{-1})	NH_3 in gas phase (cm^{-1})
1	1044	1003
2	1544	1672
3	1648	1672
4	3410	3523
5	3588	3658
6	3645	3658

The verification of the results from GSVA is carried out in Table 4.

4.3. Hydrogen Disulfide in a Water Cluster. Besides the small molecular clusters of water and ammonia, we built a cluster of hydrogen disulfide surrounded by 22 water molecules to simulate the solvation of hydrogen disulfide in liquid water (Figure 5). In this example, we want to calculate the intrinsic fragmental vibrations of the hydrogen disulfide molecule.

Table 4. Comparison of Local Mode Force Constants Based on Effective Hessian Matrix $\mathbf{f}^x_{sub}$ and Full Hessian Matrix $\mathbf{f}^x$ for Ammonia Molecule

No.	q_n[a]	$k^a_{n,sub}$	k^a_n
1	R(1−3)	6.480	6.480
2	R(1−2)	7.056	7.056
3	R(1−4)	7.056	7.056
4	α(2−1−4)	0.606	0.606
5	α(2−1−3)	0.541	0.541
6	α(3−1−4)	0.541	0.541

[a]Parameter "R" stands for bond stretching, while "α" is for angle bending. Unit of local mode force constant for bond stretchings and angles is mdyn/Å and mdyn×Å/rad^2, respectively.

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

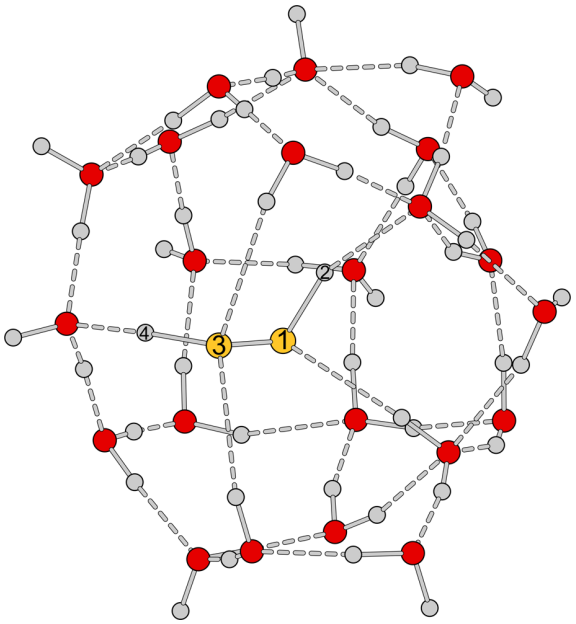


Figure 5. Structure of the hydrogen disulfide molecule in a water cluster of $(H_2O)_{22}$. Yellow balls are sulfur atoms, red are oxygens, and gray are hydrogens. Dashed lines represent noncovalent interactions, i.e., hydrogen bonds.

As for a complete nonredundant set of internal coordinate parameters required by GSVA, we choose two S−H bonds, the S−S bond, two S−S−H angles, and the H−S−S−H dihedral angle to obtain the effective Hessian matrix $\mathbf{f}^x_{sub}$ for the H_2S_2 molecule as the subsystem. These six internal coordinates construct corresponding matrices $\mathbf{B}'$ and $\mathbf{B}'_{sub}$ in the dimensions of 6 × 210 and 6 × 12, respectively.

The fragmental vibrations of the H_2S_2 molecule in the cluster calculated by GSVA (Table 5) show a shift of at least 30 cm^{-1}

Table 5. Comparison of Normal Mode Frequencies of Hydrogen Disulfide Molecule

No.	H_2S_2 in water cluster (cm^{-1})	H_2S_2 in gas phase (cm^{-1})
1	383	434
2	500	495
3	861	894
4	936	895
5	2452	2637
6	2355	2639

with regard to the reference H_2S_2 in the gas phase. Vibrations Nos. 5 and 6 have the largest deviations, and they are dominated by the S3−H4 and S1−H2 bond stretching, respectively. As these two S−H bonds donate S−H···OH_2 hydrogen bonds to surrounding water molecules, they are weakened in their bond strength leading to corresponding red shifts. However, vibration No. 2 has only a shift of 5 cm^{-1}. This vibration is dominated by the S−S bond stretching, which is hardly affected by the surrounding water molecules.

The local mode force constants of H_2S_2 within the cluster were calculated based on both the effective Hessian $\mathbf{f}^x_{sub}$ and the full Hessian $\mathbf{f}^x$ as shown in Table 6.

4.4. Methane (CH_4) in C_{60}. The methane-intercalated C_{60} structure was synthesized by Kwei and co-workers in 1997[61] (Figure 6). While the fullerene molecule has I_h symmetry and

Table 6. Comparison of Local Mode Force Constants Based on Effective Hessian Matrix $\mathbf{f}^x_{sub}$ and Full Hessian Matrix $\mathbf{f}^x$ for Hydrogen Disulfide Molecule

No.	q_n[a]	$k^a_{n,sub}$	k^a_n
1	R(1−2)	3.332	3.332
2	R(1−3)	2.172	2.172
3	R(3−4)	3.197	3.197
4	α(2−1−3)	0.809	0.809
5	α(4−3−1)	0.790	0.790
6	τ(2−1−3−4)	0.082	0.082

[a]Parameter "R" stands for bond stretching, "α" is for angle bending, and "τ" is for dihedral torsion. Unit of local mode force constant for bond stretchings and angles is mdyn/Å and mdyn×Å/rad^2, respectively.

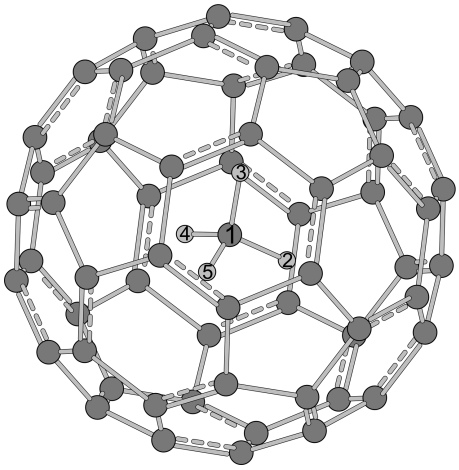


Figure 6. Structure of methane encapsulated in fullerene (C_{60}). Large balls represent carbon atoms, and small balls represent hydrogens.

methane has T_d symmetry, the complex has T symmetry. However, the methane encapsulated within the C_{60} molecule has still the T_d symmetry. It would be of interest to obtain the intrinsic fragmental vibrations of the methane inside the C_{60} in order to characterize this encapsulation effect.

As a complete nonredundant set of internal coordinates for the methane molecule, four C−H bonds and five H−C−H angles were chosen, although in total six H−C−H angles are available. Therefore, corresponding matrices of $\mathbf{B}'$ and $\mathbf{B}'_{sub}$ have the dimensions of 9 × 195 and 9 × 15, respectively.

In Table 7, the fragmental vibrational frequencies of methane within C_{60} calculated by GSVA are compared with normal mode frequencies of methane in the gas phase. All 2-fold and 3-fold degeneracies are kept as a result of the retention of T_d symmetry. The largest deviation is found for vibration No. 6 as a blue shift of 70 cm^{-1}. This vibration is dominated by the symmetric stretching of the four C−H bonds, which is largely affected by the C_{60} cage.

Table 8 lists the local mode force constants of the methane molecule in C_{60} calculated based on both the effective Hessian $\mathbf{f}^x_{sub}$ and the full Hessian $\mathbf{f}^x$. The data in Table 8 reveals that these two sets of local mode force constants are identical. Besides the nine parameters (Nos. 1−9) we used to obtain $\mathbf{f}^x_{sub}$, we have also calculated the local mode force constant of the sixth angle which was not included in the parameter set, and we obtained the same value as for the other five angles. This clearly shows that the local mode analysis can still work for the internal

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

Table 7. Comparison of Normal Mode Frequencies of Methane Molecule

No.	CH_4 in C_{60} (cm^{-1})	CH_4 in gas phase (cm^{-1})
1	1328	1356
2	1328	1356
3	1328	1356
4	1583	1584
5	1583	1584
6	3150	3080
7	3250	3205
8	3250	3205
9	3250	3205

Table 8. Comparison of Local Mode Force Constants Based on Effective Hessian Matrix $\mathbf{f}^x_{sub}$ and Full Hessian Matrix $\mathbf{f}^x$ for Methane Molecule

No.[a]	q_n[b]	$k^a_{n,sub}$	k^a_n
1	R(1−2)	5.675	5.675
2	R(1−3)	5.675	5.675
3	R(1−4)	5.675	5.675
4	R(1−5)	5.675	5.675
5	α(2−1−3)	0.647	0.647
6	α(2−1−4)	0.647	0.647
7	α(2−1−5)	0.647	0.647
8	α(3−1−4)	0.647	0.647
9	α(3−1−5)	0.647	0.647
10*	α(4−1−5)	0.647	0.647

[a]Parameter labeled with * indicates this internal coordinate is not used as part of the complete nonredundant set to obtain effective Hessian matrix $\mathbf{f}^x_{sub}$. [b]Parameter "R" stands for bond stretching, and "α" is for angle bending. Unit of local mode force constant for bond stretchings and angles is mdyn/Å and mdyn×Å/rad^2 respectively.

coordinate parameters that are not included in the matrices of $\mathbf{B}'$ or $\mathbf{B}'_{sub}$ used for the extraction of $\mathbf{f}^x_{sub}$, and it is an evidence for the fact that the effective Hessian matrix $\mathbf{f}^x_{sub}$ retains the complete information in curvature of potential energy surface with regard to any possible internal coordinate within the subsystem.

4.5. Formaldehyde (CH_2O) in Carbon Nanotube (CNT). Recently, there has been an increasing number of studies focused on the design of CNTs as sensors for detecting formaldehyde.[62−64] For this purpose, a formaldehyde molecule was placed and stabilized within the model of a single-wall carbon nanotube (SWCNT) (Figure 7). Therefore, we check the vibrational modes of the formaldehyde molecule in the nanotube with GSVA.

In order to extract the effective Hessian matrix for the formaldehyde molecule, we choose a complete nonredundant set of parameters being composed of the three covalent bonds, two O−C−H angles, and one out-of-plane pyramidalization angle. Matrices $\mathbf{B}'$ and $\mathbf{B}'_{sub}$ used to recover $\mathbf{f}^x_{sub}$ have the dimensions of 6 × 252 and 6 × 12, respectively. The intrinsic fragmental vibrational frequencies are shown in Table 9. Comparing the fragmental vibrational frequencies for the formaldehyde molecule in the CNT with the normal vibrational frequencies of the reference formaldehyde in the gas phase, it is interesting that vibrations Nos. 1−3 have small frequency differences less than 4 cm^{-1}. Vibration No. 1 is dominated by the out-of-plane pyramidalization of the carbon atom with regard to the O−H−H plane. Vibration No. 2 is basically the in-plane rocking of two hydrogen atoms, while vibration No. 3

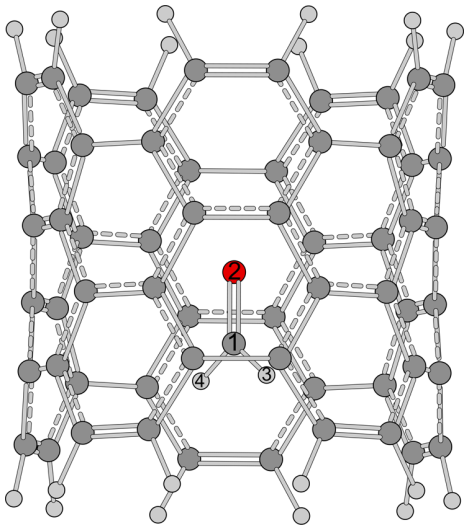


Figure 7. Structure of formaldehyde molecule contained in a carbon nanotube. Total number of atoms is 84.

Table 9. Comparison of Normal Mode Frequencies of Formaldehyde Molecule

No.	CH_2O in CNT (cm^{-1})	CH_2O in gas phase (cm^{-1})
1	1201	1201
2	1278	1275
3	1554	1555
4	1823	1847
5	2976	2897
6	3024	2954

is dominated by the in-plane scissoring of the H−C−H angle. Vibration No. 4 corresponds to the C═O bond stretching, and vibration No. 5 is associated with the symmetric C−H bond stretching mode. The largest deviation is found for vibration No. 6 which is dominated by the asymmetric stretching of two C−H bonds. This is a result of the confinement imposed on the formaldehyde molecule by the nanotube structure.

The validation of the intrinsic fragmental vibrational frequencies via the local mode analysis (Table 10) shows that the local mode properties for the subsystem based on the effective Hessian $\mathbf{f}^x_{sub}$ and full Hessian $\mathbf{f}^x$ are identical.

4.6. CH_2 and CH_3 Fragments in Propane. So far, we have applied GSVA to molecular subsystems under the perturbation of different chemical environments. However, we can also use GSVA to analyze the vibrations of fragments within a molecule and even compare the intrinsic fragmental vibrations of the same fragment in two different molecular systems.

As an example, we analyze the intrinsic fragmental vibrations in propane (Figure 8). By breaking all three C−C bonds, three fragments result, including two identical CH_3 fragments and one CH_2 fragment in the middle.

First, we applied GSVA to the CH_2 fragment. In analogy to H_2O, we chose two C−H bonds and the H−C−H angle as the internal coordinate set for constructing the effective Hessian matrix $\mathbf{f}^x_{sub}$. Therefore, the $\mathbf{B}'$ and $\mathbf{B}'_{sub}$ matrices have the dimensions of 3 × 33 and 3 × 9, respectively. The resulting fragmental vibrational modes are similar to the normal modes of H_2O. Vibration No. 1 is the H−C−H angle bending, and vibration No. 2 is the symmetric stretching of two C−H bonds.

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

Table 10. Comparison of Local Mode Force Constants Based on Effective Hessian Matrix $\mathbf{f}^x_{sub}$ and Full Hessian Matrix $\mathbf{f}^x$ for Formaldehyde Molecule

No.[a]	q_n[b]	$k^a_{n,sub}$	k^a_n
1	R(1−2)	13.306	13.306
2	R(1−3)	4.924	4.924
3	R(1−4)	4.918	4.918
4	α(2−1−3)	1.109	1.109
5	α(2−1−4)	1.110	1.110
6	$\mathcal{P}$(1′-2−3−4)	3.496	3.496
7*	α(3−1−4)	0.833	0.833

[a]Parameter labeled with * indicates this internal coordinate is not used as part of the complete nonredundant set to obtain effective Hessian matrix $\mathbf{f}^x_{sub}$. [b]Parameter "R" stands for bond stretching, and "α" is for angle bending. "$\mathcal{P}$" is for out-of-plane pyramidalization, where the atom followed by a prime symbol moves with regard to the plane constructed by the other three atoms. Unit of local mode force constant for bond stretchings and angles is mdyn/Å and mdyn×Å/rad^2, respectively.

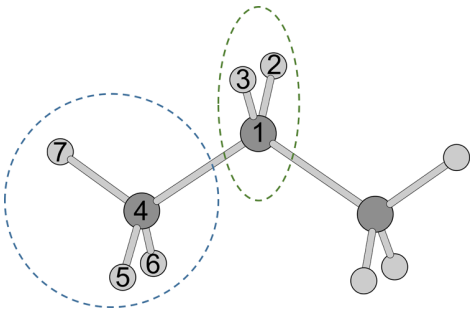


Figure 8. Structure of propane molecule in which the CH_2 and CH_3 fragments are highlighted with green and blue circles, respectively.

Vibration No. 3 is the asymmetric stretching of the same C−H bonds.

For the CH_3 fragment, we chose an analogy to the NH_3 molecule including three C−H bonds and three H−C−H angles as the complete nonredundant set to calculate its effective Hessian matrix $\mathbf{f}^x_{sub}$. Matrices $\mathbf{B}'$ and $\mathbf{B}'_{sub}$ in this regard are in the dimensions of 6 × 33 and 6 × 12, respectively. The mode characters of these six intrinsic fragmental vibrations are almost the same as the normal modes of ammonia due to their similar geometries. However, we need to note that the NH_3 molecule has C_{3v} symmetry, while the CH_3 fragment in propane has only C_s symmetry. Thus, vibrations Nos. 2−3 and Nos. 5−6 are no longer 2-fold degenerate, but they are still very close in pairs with regard to the frequency values.

Also, we take the intrinsic fragmental vibrations of CH_2 and CH_3 fragments in the methane (CH_4) molecule as the reference. In this way, the comparison of two intrinsic fragmental vibrational frequencies for the same CH_n fragment in propane and methane can be carried out as in Table 11.

The CH_3 fragments in methane molecule have 2-fold degenerate vibration pairs for vibrations Nos. 2−3 and Nos. 5−6 because the high symmetry is retained in both its geometry and corresponding effective Hessian matrix $\mathbf{f}^x_{sub}$. Noteworthy is that intrinsic fragmental vibrations associated with C−H stretching including Nos. 2 and 3 in the CH_2 fragment and Nos. 4−6 in the CH_3 fragment have larger frequency values in the reference methane molecule than in the propane molecule; this suggests that the C−H bonds in the methane molecule are stronger than those in either the CH_2 or CH_3 fragments within propane.

Table 11. Normal Mode Frequencies of Fragments in Propane Molecule and Reference Methane Molecule

No.	CH_2 in propane (cm^{-1})	CH_2 in methane (cm^{-1})	CH_3 in propane (cm^{-1})	CH_3 in methane (cm^{-1})
1	1622	1558	1543	1474
2	3178	3230	1602	1578
3	3209	3280	1605	1578
4	–	–	3174	3203
5	–	–	3235	3283
6	–	–	3242	3283

Furthermore, the verification for the physical nature of the effective Hessian matrices $\mathbf{f}^x_{sub}$ for the CH_2 and CH_3 fragments in propane is shown in Table 12.

Table 12. Comparison of Local Mode Force Constants Based on Effective Hessian Matrix $\mathbf{f}^x_{sub}$ and Full Hessian Matrix $\mathbf{f}^x$ for Propane Molecule

No.[a]	q_n[b]	$k^a_{n,sub}$	k^a_n
A-1	R(1−2)	5.581	5.581
A-2	R(1−3)	5.581	5.581
A-3	α(2−1−3)	0.836	0.836
B-1	R(4−5)	5.649	5.649
B-2	R(4−6)	5.649	5.649
B-3	R(4−7)	5.684	5.684
B-4	α(5−4−6)	0.803	0.803
B-5	α(5−4−7)	0.801	0.801
B-6	α(6−4−7)	0.801	0.801

[a]In the "No." column, "A" denotes the CH_2 fragment, while "B" denotes the CH_3 fragment. [b]Parameter "R" stands for bond stretching, and "α" is for angle bending. Unit of local mode force constant for bond stretchings and angles is mdyn/Å and mdyn×Å/rad^2, respectively.

5. CONCLUSIONS

In this work, we have presented a new method to extract the intrinsic fragmental vibrations of a subsystem/fragment from an entire polyatomic molecular system. This method is different from its predecessors[2,3,6−11,14−16] which were designed or/and can be used for the same purpose in that our method is based on an effective Hessian matrix from which the curvature of the overall potential energy surface with regard to any internal coordinate parameter q_n within the subsystem is retained. The underlying solid physical foundation makes our method unique and able to characterize fragmental normal mode vibrations which are intrinsic to the subsystem/fragment in question. Therefore, our method is named the Generalized Subsystem Vibrational Analysis (GSVA), emphasizing its general applicability for any subsystem or fragment within a molecular system and concrete physical basis.

In the examples presented in this work, we compared the intrinsic fragmental vibrations of a subsystem with the normal vibrational modes of the isolated subsystem in gas phase to show the changes in the electronic structure caused by the presence of the environment. Although a more straightforward approach is to compare the corresponding properties of the local vibrational modes of the subsystem, the intrinsic fragmental vibrations can be regarded as a key intermediate

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

linking the normal vibrational modes and the local vibrational modes in two aspects: (i) The intrinsic fragmental vibrations calculated by GSVA are in nature normal vibrational modes. (ii) These vibrations are based on the effective Hessian matrix taking the physical basis from Konkoli and Cremer's local vibrational modes,[31−34] namely, to retain the potential energy surface curvature of the whole system. In this regard, this work can be also considered as a theoretical extension to our previous work on the local vibrational modes.

A caveat is necessary when applying GSVA in theoretical chemical studies. The equilibrium geometry $\mathbf{R}_0$ for the entire molecular system including the subsystem to be studied is required by eq 1. The full Hessian matrix $\mathbf{f}^x$ describing the entire system is needed as one of the input data along with the geometry $\mathbf{R}_0$ and atomic masses $\mathbf{M}$. If one set of intrinsic fragmental vibrations is to be compared with another set of intrinsic fragmental vibrations for the same subsystem/fragment, we need to make sure that these two different molecular systems are being described with the same level of theory.

Concerning the implementation of GSVA into a computational chemistry package or as a standalone analysis program, three inputs are required to start with including the full Hessian matrix $\mathbf{f}^x$, geometry in Cartesian coordinates, and atomic masses. As the calculation of the effective Hessian matrix $\mathbf{f}^x_{sub}$ uses Wilson B matrices $\mathbf{B}'$ and $\mathbf{B}'_{sub}$ characterizing the complete nonredundant set of internal coordinates of the subsystem, a subroutine is expected for calculating Wilson B matrices for various internal coordinates, including bond length, bond angle, dihedral torsion angle, and so forth. Besides, the linear independence between rows of the B matrix should be checked and guaranteed in order for a complete and nonredundant set of $3n - k$ internal coordinates determining the geometry of the subsystem. Furthermore, a subroutine for solving the Wilson equation of vibrational spectroscopy is required to obtain normal vibrational modes.[65] By providing the effective Hessian matrix $\mathbf{f}^x_{sub}$, geometry, and atomic masses of the subsystem for the above subroutine, the normal mode vectors and frequencies can be obtained for the intrinsic fragmental vibrations. The computational cost of the whole calculation in GSVA is equivalent to doing the normal mode analysis for the entire system, and the most expensive part lies in the calculation of the Moore−Penrose inverse $(\mathbf{f}^x)^+$ of the full Hessian matrix.

This work provides a new and reliable theoretical tool for analyzing as well as comparing the molecular vibrations, and we anticipate our GSVA method to become a routine procedure in computational chemistry

AUTHOR INFORMATION

Corresponding Author

*E-mail: ekraka@smu.edu.

ORCID

Wenli Zou: 0000-0002-0747-2428

Dieter Cremer: 0000-0002-6213-5555

Elfi Kraka: 0000-0002-9658-5626

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Y.T. thanks Dr. Bernard Brooks for thought-provoking discussions during the 27th Austin Symposium on Molecular Structure and Dynamics at Dallas (ASMD@D This work was financially supported by the Natural Science Foundation (Grant 1464906). We thank SMU for providing computational resources.

DEDICATION

⊥This paper is in memoriam of Dr. Dieter Cremer.

REFERENCES

(1) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Dover Publications: Mineola, NY, 2012.

(2) Head, J. D. Computation of Vibrational Frequencies for Adsorbates on Surfaces. *Int. J. Quantum Chem.* **1997**, *65*, 827−838.

(3) Li, H.; Jensen, J. H. Partial Hessian Vibrational Analysis: The Localization of the Molecular Vibrational Energy and Entropy. *Theor. Chem. Acc.* **2002**, *107*, 211−219.

(4) Besley, N. A.; Metcalf, K. A. Computation of the Amide I Band of Polypeptides and Proteins using a Partial Hessian Approach. *J. Chem. Phys.* **2007**, *126*, 035101.

(5) Besley, N. A.; Bryan, J. A. Partial Hessian Vibrational Analysis of Organic Molecules Adsorbed on Si(100). *J. Phys. Chem. C* **2008**, *112*, 4308−4314.

(6) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. Vibrational Modes in Partially Optimized Molecular Systems. *J. Chem. Phys.* **2007**, *126*, 224102.

(7) Ghysels, A.; Van Neck, D.; Waroquier, M. Cartesian Formulation of the Mobile Block Hessian Approach to Vibrational Analysis in Partially Optimized Systems. *J. Chem. Phys.* **2007**, *127*, 164108.

(8) Ghysels, A.; Van Neck, D.; Brooks, B. R.; Van Speybroeck, V.; Waroquier, M. Normal Modes for Large Molecules with Arbitrary Link Constraints in the Mobile Block Hessian Approach. *J. Chem. Phys.* **2009**, *130*, 084107.

(9) Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Van Neck, D.; Brooks, B. R.; Waroquier, M. Mobile Block Hessian Approach with Adjoined Blocks: An Efficient Approach for the Calculation of Frequencies in Macromolecules. *J. Chem. Theory Comput.* **2009**, *5*, 1203−1215.

(10) Ghysels, A.; Van Speybroeck, V.; Verstraelen, T.; Van Neck, D.; Waroquier, M. Calculating Reaction Rates with Partial Hessians: Validation of the Mobile Block Hessian Approach. *J. Chem. Theory Comput.* **2008**, *4*, 614−625.

(11) Woodcock, H. L.; Zheng, W.; Ghysels, A.; Shao, Y.; Kong, J.; Brooks, B. R. Vibrational Subsystem Analysis: A Method for Probing Free Energies and Correlations in the Harmonic Limit. *J. Chem. Phys.* **2008**, *129*, 214109.

(12) Hafner, J.; Zheng, W. Approximate Normal Mode Analysis Based on Vibrational Subsystem Analysis with High Accuracy and Efficiency. *J. Chem. Phys.* **2009**, *130*, 194111.

(13) Ghysels, A.; Miller, B. T.; Pickard, F. C.; Brooks, B. R. Comparing Normal Modes Across Different Models and Scales: Hessian Reduction versus Coarse-Graining. *J. Comput. Chem.* **2012**, *33*, 2250−2275.

(14) Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Catak, S.; Brooks, B. R.; Van Neck, D.; Waroquier, M. Comparative Study of Various Normal Mode Analysis Techniques Based on Partial Hessians. *J. Comput. Chem.* **2010**, *31*, 994−1007.

(15) Jacob, C. R.; Reiher, M. Localizing Normal Modes in Large Molecules. *J. Chem. Phys.* **2009**, *130*, 084106.

(16) Huix-Rotllant, M.; Ferré, N. An Effective Procedure for Analyzing Molecular Vibrations in Terms of Local Fragment Modes. *J. Chem. Theory Comput.* **2016**, *12*, 4768−4777.

(17) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300−302.

(18) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457−464.

(19) Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for ab initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.* **1989**, *90*, 4916−4926.

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

(20) Weinhold, F.; Landis, C. R. Natural Bond Orbitals and Extensions of Localized Bonding Concepts. *Chem. Educ. Res. Pract.* **2001**, *2*, 91−104.
(21) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Orbital Localization using Fourth Central Moment Minimization. *J. Chem. Phys.* **2012**, *137*, 224114.
(22) Mulliken, R. S. Electronic Population Analysis on LCAO−MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(23) Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theor. Chem. Acc.* **1977**, *44*, 129−138.
(24) Foster, J. P.; Weinhold, F. Natural Hybrid Orbitals. *J. Am. Chem. Soc.* **1980**, *102*, 7211−7218.
(25) Chirlian, L. E.; Francl, M. M. Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. *J. Comput. Chem.* **1987**, *8*, 894−905.
(26) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.
(27) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(28) Bader, R. F. W. A Quantum Theory of Molecular Structure and its Applications. *Chem. Rev.* **1991**, *91*, 893−928.
(29) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford, UK, 1994.
(30) Zou, W.; Kalescky, R.; Kraka, E.; Cremer, D. Relating Normal Vibrational Modes to Local Vibrational Modes with the Help of an Adiabatic Connection Scheme. *J. Chem. Phys.* **2012**, *137*, 084114.
(31) Konkoli, Z.; Cremer, D. A New Way of Analyzing Vibrational Spectra. I. Derivation of Adiabatic Internal Modes. *Int. J. Quantum Chem.* **1998**, *67*, 1−9.
(32) Konkoli, Z.; Larsson, J. A.; Cremer, D. A New Way of Analyzing Vibrational Spectra. II. Comparison of Internal Mode Frequencies. *Int. J. Quantum Chem.* **1998**, *67*, 11−27.
(33) Konkoli, Z.; Cremer, D. A New Way of Analyzing Vibrational Spectra. III. Characterization of Normal Vibrational Modes in terms of Internal Vibrational Modes. *Int. J. Quantum Chem.* **1998**, *67*, 29−40.
(34) Konkoli, Z.; Larsson, J. A.; Cremer, D. A New Way of Analyzing Vibrational Spectra. IV. Application and Testing of Adiabatic Modes Within the Concept of the Characterization of Normal Modes. *Int. J. Quantum Chem.* **1998**, *67*, 41−55.
(35) Zou, W.; Cremer, D. Properties of Local Vibrational Modes: The Infrared Intensity. *Theor. Chem. Acc.* **2014**, *133*, 1451.
(36) Kalescky, R.; Zou, W.; Kraka, E.; Cremer, D. Local vibrational modes of the water dimer - Comparison of theory and experiment. *Chem. Phys. Lett.* **2012**, *554*, 243−247.
(37) Freindorf, M.; Kraka, E.; Cremer, D. A Comprehensive Analysis of Hydrogen Bond Interactions Based on Local Vibrational Modes. *Int. J. Quantum Chem.* **2012**, *112*, 3174−3187.
(38) Zou, W.; Cremer, D. C_2 in a Box: Determining its Intrinsic Bond Strength for the $X^1\Sigma_g^+$ Ground State. *Chem. - Eur. J.* **2016**, *22*, 4087−4099.
(39) Kalescky, R.; Kraka, E.; Cremer, D. Identification of the Strongest Bonds in Chemistry. *J. Phys. Chem. A* **2013**, *117*, 8981−8995.
(40) Tao, Y.; Zou, W.; Jia, J.; Li, W.; Cremer, D. Different Ways of Hydrogen Bonding in Water - Why Does Warm Water Freeze Faster than Cold Water? *J. Chem. Theory Comput.* **2017**, *13*, 55−76.
(41) Cremer, D.; Kraka, E. Generalization of the Tolman Electronic Parameter: the Metal-ligand Electronic Parameter and the Intrinsic Strength of the Metal-ligand Bond. *Dalton Trans.* **2017**, *46*, 8323−8338.
(42) Tao, Y.; Zou, W.; Kraka, E. Strengthening of Hydrogen Bonding with the Push-pull Effect. *Chem. Phys. Lett.* **2017**, *685*, 251−258.
(43) Tao, Y.; Zou, W.; Cremer, D.; Kraka, E. Characterizing Chemical Similarity With Vibrational Spectroscopy: New Insights Into the Substituent Effects in Mono-Substituted Benzenes. *J. Phys. Chem. A* **2017**, *121*, 8086−8096.
(44) Tao, Y.; Zou, W.; Cremer, D.; Kraka, E. Correlating the Vibrational Spectra of Structurally Related Molecules: A Spectroscopic Measure of Similarity. *J. Comput. Chem.* **2018**, *39*, 293−306.
(45) Wilson, E. B. A Method of Obtaining the Expanded Secular Equation for the Vibration Frequencies of a Molecule. *J. Chem. Phys.* **1939**, *7*, 1047−1052.
(46) Brandhorst, K.; Grunenberg, J. Efficient Computation of Compliance Matrices in Redundant Internal Coordinates from Cartesian Hessians for Nonstationary Points. *J. Chem. Phys.* **2010**, *132*, 184101.
(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, P. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision E.01 ; Gaussian, Inc.: Wallingford, CT, 2009.
(48) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(49) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.
(50) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chem. Acc.* **1973**, *28*, 213−222.
(51) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient Diffuse Function-augmented Basis Sets for Anion Calculations. III. The 3-21+G Basis Set for First-Row Elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294−301.
(52) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(53) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(54) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(55) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(56) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(57) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(58) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(59) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(60) Kraka, E.; Zou, W.; Filatov, M.; Tao, Y.; Grafenstein, J.; Izotov, D.; Gauss, J.; He, Y.; Wu, A.; Konkoli, Z.; Polo, V.; Olsson, L.; He, Z.;

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

Cremer, D. COLOGNE2017, 2017; http://www.smu.edu/catco (accessed April 2018).
(61) Morosin, B.; Assink, R.; Dunn, R.; Massis, T.; Schirber, J.; Kwei, G. Methane-intercalated C60: Preparation, Orientational Ordering, and Structure. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1997**, *56*, 13611−13614.
(62) Lu, Y.; Meyyappan, M.; Li, J. A Carbon-Nanotube-based Sensor Array for Formaldehyde Detection. *Nanotechnology* **2011**, *22*, 055502.
(63) Kim, J. Y.; Lee, J.; Hong, S.; Chung, T. D. Formaldehyde Gas Sensing Chip based on Single-Walled Carbon Nanotubes and Thin Water Layer. *Chem. Commun.* **2011**, *47*, 2892−2894.
(64) Tang, R.; Shi, Y.; Hou, Z.; Wei, L. Carbon Nanotube-Based Chemiresistive Sensors. *Sensors* **2017**, *17*, 882.
(65) Ochterski, J. *White Paper: Vibrational analysis in Gaussian*; Gaussian, 1999.

DOI: 10.1021/acs.jctc.7b01171
J. Chem. Theory Comput. 2018, 14, 2558−2569

Appendix F

User Manual of pURVA program

CATCO Group Internal Use ONLY

User Manual for pURVA program

by

Yunwen Tao

Ver 0.1.2

March 2018

Contents

Abbreviations

pURVA	standalone **p**rogram of **U**nified **R**eaction **V**alley **A**pproach
IRC	**I**ntrinsic **R**eaction **C**oordinates

Symbols

$\mathring{A}$	distance	Angstrom
ω	harmonic frequency	cm^{-1}

Chapter 1

Before we start

It is not a good idea to have no idea about the basic copyright, history, requirement, functionality as well as theory of the program before we use it.

1.1 Copyright

The program pURVA as well as this manual MUST NOT be released out of CATCO group at Southern Methodist University. According to policy §12.1 and §12.2 of Southern Methodist University, the leakage of the intellectually property may face legal charge.

1.2 History

The original URVA method was implemented by Zoran Konkoli in the link L716 of Gaussian package. However, this part is never incorporated into the public version of Gaussian. Since then, many other contributors including Dr. Wenli Zou added functionalities into this part and migrated this part from older version of Gaussian into newer version of Gaussian for several times. As Gaussian package was written in Fortran 77, the corresponding URVA part was written in the same language.

Later on from 2015, Dr. Dieter Cremer and Dr. Elfi Kraka wanted to have an independent version of URVA program. They asked Yunwen Tao in the group to this job. He started with programming in Fortran 90 language which is an extension to Fortran 77. Then he switched the whole project into Python language which is more flexible and easy to use. The new version of the URVA program was then named as **pURVA**.

1.3 Execution of pURVA

The proper execution of pURVA requires Python interpreter with the version 2.7.x. Versions lower than this might lead to trouble.

Here are the list of Python modules needed to run pURVA: (1) NumPy, (2) SciPy, (3) SymPy, (4) sys, (5) os, (6) copy, (7) gc, (8) math and (9) time.

Make sure that all these modules have been installed properly.

pURVA expects and then reads in an external text file as the user input file. After this file is prepared, in the terminal, type in

```
$ python main.py myinputfile
```

From the standard output, we could monitor how the calculation goes. The calculation results will be dumped into external text files on the disk.

To make life easier, running pURVA on ManeFrame cluster is recommended as pURVA has been developed and tested on the same machine. Before running pURVA, remember to load the Python interpreter by using

```
$ module load python
```

1.4 Theory as Unified Reaction Valley Approach

The name of "Unified Reaction Valley Approach" firstly appeared on scientific journals in 1997 when Konkoli, Kraka and Cremer published their comprehensive studies of $CH_3 + H_2 \rightarrow CH_4 + H$ on *J. Phys. Chem. A*[1]. In that paper, one of the highlights is to introduce the approach that calculates the adiabatic mode coupling coefficient that is decomposition of reaction path curvature into adiabatic local modes which was a novel approach dealing with vibrational spectroscopy. URVA is based on the Reaction Path Hamiltonian(RPH) that was intensively developed by Miller, Handy[2], Page and McIver[3]. In the year of 2011, Dr. Kraka published a well-written review on the relationship between RPH and URVA[4]. Most recently, Dr. Zou proposed a new approach to decompose the reaction path direction and curvature into internal coordinates which opens the possibility to study chemical reactions in large systems, e.g. organometallic compounds and enzymes[5].

One of the most important papers involved in URVA is the introduction of Diabatic Mode Ordering(DMO) procedure which has now been widely used in several projects within CATCO group[6].

Chapter 2

Input description

There are two major types of input that are allowed in the input file.

- Keyword input
- Section input

2.1 Keyword input

Just for convenience, keyword input is often written before section input. The format of keyword input line is:

`@keyword_name = [keyword_value]`

The `@` symbol should be in the first column. No space is allowed after it. On both sides of = sign, it should be space. There might be several optional keyword values available, however, only one option is accepted.

2.1.1 `@DATAFILETYPE` keyword

This keywords specifies the format of input data source file for URVA analysis.

`@DATAFILETYPE = old/new/xyz`

`old`: The input data source file is generated by Gaussian package by setting corresponding IOp(1/45). This type of data contains most complete information.

NOTE: If the data file is generated by Gaussian with version number lower than 16.A, all floating numbers should be converted from "D" into "E" format.

new: The input data source file is generated by a modified version of Gaussian package. This type of data has no Hessian and gradient stored.

xyz: XYZ file containing the Cartesian coordinates of multiple snapshots.

2.1.2 @DATAFILEPATH keyword

This keyword specifies the path of the input data file.

```
@DATAFILEPATH = "../path/to/data/file"
```

The quotation marks should be included.

2.1.3 @ENERGY keyword

This keyword specifies whether SCF energy and its first and second derivatives will be calculated.

```
@ENERGY = on/off
```

2.1.4 @PARM keyword

This keyword specifies the way to deal with internal coordinates parameters provided by user.

```
@PARM = No/GeomOnly/All
```

No: Do nothing with regard to these internal coordinates specifications.

GeomOnly: Only calculate the value of these internal coordinates.

All: Besides the value of internal coordinates, other properties related to these internal coordinates will be calculated.

2.1.5 @VIBRATION keyword

This keyword specifies whether or not to do normal mode analysis.

```
@VIBRATION = on/off
```

If the keyword value is set to on, the @DATAFILETYPE must be set to old.

2.1.6 `@DIRCURV` keyword

This keyword decides whether or not to calculate reaction path direction $\boldsymbol{\eta}(\boldsymbol{s})$ and curvature $\boldsymbol{\kappa}(\boldsymbol{s})$.

`@DIRCURV = on/off`

If the keyword value is set to `on`, the `@DATAFILETYPE` must be set to `old` or `new`.

2.1.7 `@AVAM` keyword

This keyword specifies whether or not to calculate the adiabatic mode coupling coefficient $\boldsymbol{A}_{\boldsymbol{n},\boldsymbol{s}}(\boldsymbol{s})$.

`@AVAM = on/off`

If the keyword value is set to `on`, the `@DATAFILETYPE` must be set to `old`, the `@PARM` must be set to `All`, the `@VIBRATION` must be set to `on` and the `@DIRCURV` must be set to `on`

2.1.8 `@CURVCPL` keyword

This keyword specifies whether or not to calculate the curvature coupling coefficient $\boldsymbol{B}_{\boldsymbol{\mu},\boldsymbol{s}}(\boldsymbol{s})$.

`@CURVCPL = on/off`

If the keyword value is set to `on`, the `@DATAFILETYPE` must be set to `old`, the `@VIBRATION` must be set to `on`, and the `@DIRCURV` must be set to `on`.

2.1.9 `@CORIOLIS` keyword

This keyword specifies whether or not to calculate the Coriolis mode-mode coupling coefficient $\boldsymbol{B}_{\boldsymbol{\mu},\boldsymbol{\nu}}(\boldsymbol{s})$.

`@CORIOLIS = on/off`

If the keyword value is set to `on`, the `@DATAFILETYPE` must be set to `old` and the `@VIBRATION` must be set to `on`.

2.1.10 `@ADIABFC` keyword

This keyword specifies whether or not to calculate adiabatic force constant $\mathbf{k}^{\mathbf{a}}$.

```
@ADIABFC = on/off
```

If the keyword value is set to `on`, the `@DATAFILETYPE` must be set to `old` and the `@PARM` must be set to `All`.

2.2 Section input

Section input is used when multiple parameters need to be read in, the format of the section input is:

```
SECTION_NAME
parameter line_1
parameter line_2
...
END SECTION_NAME
```

2.2.1 `TITLE` section

This section accepts remarks provided by user. The content will be displayed in standard output.

```
TITLE
Please put remarks here.
Multiple lines are accepted.
END TITLE
```

This section is quite useful to take note of the parameters we use for URVA calculations.

2.2.2 `PARAMETER` section

This section contains the internal coordinates specifications provided by the user. Different types of internal coordinates including ring coordinates are acceptable.

Bond length, bond angle, dihedral angle, out-of-plane angle, pyramidalization angle, ring puckering amplitude, ring puckering phase angle, ring deformation amplitude and ring deformation phase angle are supported.

```
PARAMETER

Internal coordinate specification

END PARAMETER
```

Bond length:

```
std N1 N2 :  "bond_name"
```

Bond angle:

```
std N1 N2 N3 :  "angle_name"
```

Dihedral angle:

```
std N1 N2 N3 N4 :  "dihedral_name"
```

Out of plane angle(the angle between the bond length N_1-N_2 and the plane N_2-N_3-N_4):

```
oop N1 N2 N3 N4 :  "out_of_plane_name"
```

Pyramidalization angle(the angle θ_P is related to the three bond angles N_2-N_1-N_3, N_3-N_1-N_4, N_4-N_1-N_2):

```
pyr N1 N2 N3 N4 :  "pyramidalization_angle_name"
```

Radius of planar reference ring(R)(N_{ring}: number of ring atoms):

```
ring Nring - ( N1 N2 ...  Natoms ) -[0 0]:  "ring_breathing_name"
```

Planar deformation amplitude(t_n)(n=1$\sim N_{ring}-2$):

```
ring Nring - ( N1 N2 ...  Natoms ) -[1 n]:  "deformation_amplitude_name"
```

Planar deformation phase angle(τ_n)(n=1$\sim N_{ring}-2$):

```
ring Nring - ( N1 N2 ...  Natoms ) -[2 n]:  "deformation_phase_angle_name"
```

Puckering amplitude(q_n)(n=2$\sim(N_{ring}-1)/2$ for odd N_{ring} or 2$\sim N_{ring}/2$ for even N_{ring}):

```
ring Nring - ( N1 N2 ...  Natoms ) -[3 n]:  "puckering_amplitude_name"
```

Puckering phase angle(ϕ_n)(n=2$\sim$($N_{ring}-1$)/2 for odd N_{ring} or 2$\sim N_{ring}$/2-1 for even N_{ring}):

```
ring N_ring - ( N_1 N_2 ... N_atoms ) -[4 n]: "puckering_phase_angle_name"
```

2.2.3 **CURVCOR** section

The CURVCOR interface will be activated if this section is found.

For most situations, it is usually enough for N_l and N_r to take the value of 25.

```
CURVCOR
Ln = N_l
Rn = N_r
END CURVCOR
```

2.2.4 **AUTOSMTH** section

The AUTOSMTH interface will be activated if this section is found.

AUTOSMTH interface requires the activation of CURVCOR interface.

δs is the stepsize of mass-weighted IRC with the unit of amu$^{1/2}$-Bohr.

Using the value of 3 is usually enough for N_l and N_r.

t is a cut-off for second derivative of smoothened curve. Increase it when necessary. Recommended value: 2.5.

```
AUTOSMTH
StepSize = δs
Ln = N_l
Rn = N_r
d2ythresh = t
END AUTOSMTH
```

2.2.5 `RMSPK` section

The RMSPK interface will be activated if this section is found.

RMSPK interface requires the activation of AUTOSMTH.

Any points in the curvature plot having the value larger than k will be left out as spike.

The value of p ranges from 0.5 to 1.0 as a percentage number. Any points leading to consecutive difference larger than the percentile of p will be labeled as spike condidates. Recommened value: 0.85.

Gradient check threshold g is used to filter out normal points from spike candidates. Recommended value: 1.2.

```
RMSPK

CutHigh = k

Percentage = p

GradRatio = g

END RMSPK
```

2.2.6 `DMO` section

If this section input is not found, default parameter values will be used.

s_{max} is an overlap threshold after each mode reordering step. If the overlap criteria of s_{max} could not be reached, the criteria will be reduced to s_{min} gradually. Recommend values for s_{max} and s_{min}: 0.990 and 0.890.

If local difficulty is encountered, linear interpolation will be adopted, space between two consecutive points will be divided into N_{min} pieces. If the difficulty is still not solved, N_{min} will be increased up to N_{max}. Recommended values for N_{min} and N_{max}: 30 and 200.

If the DMO could not get through for a specific point due to the following reasons:

- Change of symmetry of reaction complex, e.g. linear $\rightarrow$ non-linear
- Discontinuity of reaction path
- Failure of reaction path following close to local minimum region

one solution to circumvent this problem is to calculate and re-order the vibrational frequencies for a specific region of reaction path. This function could be activated by setting IO_{cut} to 1. In this way, the reaction path with its **s** value ranging from s_{start} to s_{end} will have vibrational frequencies calculated.

In some situations, due to the innate difficulty of path following algorithm, the DMO might fail at the transition state(TS) point. And also the first point off TS point in either forward or reverse direction might also lead to problems. In order to remediate this problem, we can skip a few points in that region by setting IO_{skip} to 1. If one point off the TS point in reverse(or forward) direction also needs to be skipped, N_{left}(or N_{right}) should be set to 1.

```
DMO

Sthresh = s_max

Slowest = s_min

Np = N_min

NMax = N_max

Cut = IO_cut

CutA = s_start

CutB = s_end

Skip = IO_skip

SkipA = N_left

SkipB = N_right

END DMO
```

Chapter 3

Output description

In pURVA, results are all written to external files instead of standard output. All output files have the suffix of ".csv" or ".dat". The execution of pURVA will abort if a result file with a duplicated name is found in the current folder. Make sure that current directory is cleaned up before execution.

3.1 Energy and derivatives

Usually the Self-Consistent Field(SCF) energy is calculated and used to construct the potential energy surface along the reaction path.

In order to check this value, the `@DATAFILETYPE` must be set to `old` or `new` and `@ENERGY` must be set to `on`.

The unit of SCF energy is Hartree as one of the atomic units(a.u.). In order to calculate first and second derivatives of SCF energy against reaction coordinate/parameter **s**, cubic spline fitting is used. For the second derivative of SCF energy, the region between $\mathbf{s} = -0.1$ and $\mathbf{s} = +0.1$ is predicted via cubic spline fitting from the information outside this region.

NOTE: 1 Hartree = 627.509 474 kcal/mol

Output files:

- `energy.csv`

  SCF energy vs. **s**

- `energy_1_d.csv`

  First derivative of SCF energy vs. **s**

- `energy_2_d.csv`

  Second derivative of SCF energy vs. **s**

3.2 Internal coordinates

The value of user-defined internal coordinates could be calculated. All types of internal coordinates described in section 2.2.2 are supported.

In order to have this result, `@PARM` must be set to `GeomOnly` or `All`.

The unit of printed internal coordinates is atomic unit with bohr for distance and radian for angles.

NOTE: 1 Bohr = 0.529177 Å; 1 rad = 57.295 8 °

Output file:

- `q_n.csv`

3.3 Decomposition of reaction path direction and curvature into internal coordinates

In order to have this result, `@PARM` must be set to `All`.

Output files:

- `eta-q_n.csv`

  Decomposition of reaction path direction into internal coordinates

- `kappa-q_n.csv`

  Decomposition of reaction path curvature into internal coordinates

3.4 Generalized vibrational frequency

For any point on reaction path, we could have $3N - K - 1$ vibrations, in which K is the total number of translations and rotations. In pURVA, K takes the value of 6 which excludes the possibility of analysis of reactions like $H_2 + H \rightarrow H + H_2$ where the whole reaction complex stays in a linear geometry.

In order to have this result, `@VIBRATION` must be set to `on`.

Output file:

- `freq_dmo.csv`

  Generalized vibrational frequencies vs. **s**

NOTE: Unit of frequencies is cm^{-1}.

3.5 Scalar curvature

The original scalar curvature calculated without correction around the TS region and spike removal will be written to file `originalkappa.dat`.

In order to have this result, `@DIRCURV` must be set to `on`.

If the CURVCOR and AUTOSMTH modules are used, the corrected curvature data will be written to `merged.dat`.

If RMSPK module is also used, the curvature data after spike removal will be written to `merged-nospk.dat`.

3.6 Adiabatic force constant

The adiabatic force constant of chemical bonds between two atoms along the reaction path will be written to `adiabfc-ka.csv`.

In order to have this result, `@ADIABFC` must be set to `on`.

In some situations, there might be noise in the result. These noise regions could be nicely removed via cubic spline fitting.

NOTE: Only result of bond length between 2 atoms could make sense.

Chapter 4

Examples

4.1 Example 1. HCN → HNC isomerization

```
@DATAFILETYPE = old
@PARM = All
@VIBRATION = on
@DIRCURV = on
@AVAM = off
@CURVCPL = off
@CORIOLIS = off
@ENERGY = on
@ADIABFC = off

@DATAFILEPATH = "./examples/hcn/IRC.browse"

TITLE
 HCN Reaction test job
 Frequency calculation from -3.09985216909 to 3.97984648135
END TITLE

PARAMETER
std 2 3 : bond-NH
std 1 3 : bond-CH
END PARAMETER

DMO
Sthresh = 0.990
Slowest = 0.890
Np = 30
NMax = 200
Cut = 1
CutA = -3.09985216909
CutB = 3.97984648135
END DMO
```

4.2 Example 2. $CH_3 + H_2 \rightarrow CH_4 + H$

The whole calculation may take up to 15 minutes.

```
@DATAFILETYPE = old
@PARM = All
@VIBRATION = on
@DIRCURV = on
@AVAM = off
@CURVCPL = off
@CORIOLIS = off
@ENERGY = on
@ADIABFC = off

@DATAFILEPATH = "./examples/ch3h2/IRC.forward.2"

TITLE
 CH3+H2 Reaction test job
END TITLE

PARAMETER
std 2 3 : bond-NH
std 1 3 : bond-CH
END PARAMETER

DMO
Sthresh = 0.990
Slowest = 0.890
Np = 30
NMax = 200
Cut = 0
END DMO
```

4.3 Example 3. Gold catalysis Step-1

For this reaction complex, we do three URVA analysis runs. In the first run, basic information including energy, internal coordinates are calculated. In the second run, scalar curvature is corrected in its TS region and spikes are also removed. In the last run, vibrational frequencies are calculated for three segments of the reaction path due to several local difficulties around TS point.

4.3.1 First run

```
@DATAFILETYPE = old
@PARM = GeomOnly
@VIBRATION = off
@DIRCURV = off
@AVAM = off
@CURVCPL = off
@CORIOLIS = off
@ENERGY = on
%%@ADIABFC = off

@DATAFILEPATH = "./examples/gold/IRC.browse"
@BASEPATH = "/path/to/pURVA/folder"


TITLE
 Gold catalysis step 1 test job - first run
END TITLE

PARAMETER
std 1 16 : bond-C1C16
std 5 16 : bond-O5C16
std 1 10 : bond-C1C10
std 2 5 :  bond-C2O5
std 2 4 :  bond-C2O4
END PARAMETER
```

4.3.2 Second run

```
@DATAFILETYPE = old
@PARM = All
@VIBRATION = off
@DIRCURV = on
@AVAM = off
@CURVCPL = off
@CORIOLIS = off
@ENERGY = off
@ADIABFC = off

@DATAFILEPATH = "./examples/gold/IRC.browse"

TITLE
 Gold catalysis step 1 test job - second run
END TITLE

PARAMETER
std 1 16 : bond-C1C16
std 5 16 : bond-O5C16
std 1 10 : bond-C1C10
std 2 5 :  bond-C2O5
std 2 4 :  bond-C2O4
END PARAMETER

CURVCOR
Ln = 25
Rn = 25
END CURVCOR

AUTOSMTH
StepSize = 0.03
Ln = 3
Rn = 3
d2ythresh = 2.4
END AUTOSMTH

RMSPK
CutHigh = 20.0
Percentage = 0.85
GradRatio = 1.2
END RMSPK
```

4.3.3 Third run

This run may take up to 25 minutes.

```
@DATAFILETYPE = old
@PARM = No
@VIBRATION = on
@DIRCURV = off
@AVAM = off
@CURVCPL = off
@CORIOLIS = off
@ENERGY = off
@ADIABFC = off

@DATAFILEPATH = "./examples/gold/IRC.browse"

TITLE
 Gold catalysis step 1 test job - third run
END TITLE

DMO
Sthresh = 0.980
Slowest = 0.880
Np = 5
NMax = 80
Cut = 0
END DMO
```

Bibliography

[1] Zoran Konkoli, Elfi Kraka, and Dieter Cremer. Unified reaction valley approach mechanism of the reaction CH3 + H2 → CH4 + H. *The Journal of Physical Chemistry A*, 101(9):1742–1757, 1997.

[2] William H. Miller, Nicholas C. Handy, and John E. Adams. Reaction path hamiltonian for polyatomic molecules. *The Journal of Chemical Physics*, 72(1):99–112, 1980.

[3] Michael Page and James W. McIver. On evaluating the reaction path hamiltonian. *The Journal of Chemical Physics*, 88(2):922–935, 1988.

[4] Elfi Kraka. Reaction path hamiltonian and the unified reaction valley approach. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(4):531–556, 2011.

[5] Wenli Zou, Thomas Sexton, Elfi Kraka, Marek Freindorf, and Dieter Cremer. A new method for describing the mechanism of a chemical reaction based on the unified reaction valley approach. *Journal of Chemical Theory and Computation*, 12(2):650–663, 2016.

[6] Zoran Konkoli, Dieter Cremer, and Elfi Kraka. Diabatic ordering of vibrational normal modes in reaction valley studies. *Journal of Computational Chemistry*, 18 (10):1282–1294, 1997.

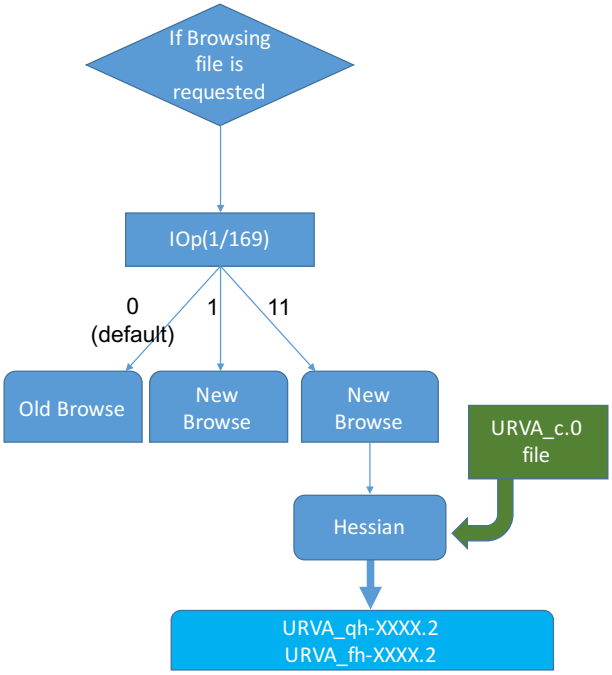
Appendix G

Flowchart of Modifications of the Gaussian 16 Source Code

URVA Browsing File in G16

August 9, 2017

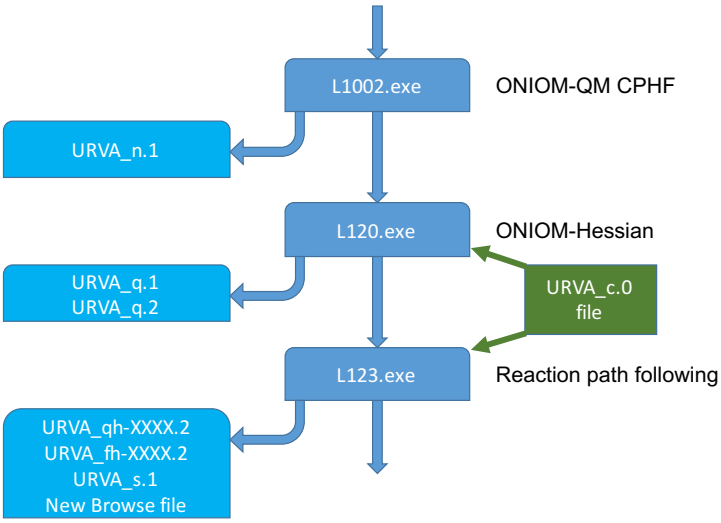
The URVA browsing file is generated by the user input. The format of the browse file is determined by IOp(1/169).



If this IOp option is set to 11, a new type of browsing file (for large systems) can be generated along with the hessian matrices in the ONIOM model. In this case, a user input file called "URVA_c.0" is expected.

In "URVA_c.0", there are two lines. The first line is a number K and the second line is also a number N. N means we record the information of Hessian every N steps. K should start from 0, and it increases by 1 in each IRC step. When it hits N, the Hessian will be recorded. Then K turns back to 1 and starts to increase again.

We have modified three links of G16. The L1002 can generate a file "URVA_n.1" which contains the information of the total number of atoms and the labels of



QM atoms.

L120 will read in "URVA_c.0", and the QM Hessian and full Hessian will be saved as "URVA_q.1" and "URVA_q.2".

L123 will also read in "URVA_c.0", it will rename "URVA_q.1" and "URVA_q.2" into "URVA_qh-XXXX.2" and "URVA_fh-XXXX.2" respectively. XXXX is the number of steps of IRC, starting from 0000. Then the "URVA_c.0" will be updated. "URVA_s.1" records the number of IRC step and reaction parameter s. Furthermore, the new browsing file will be generated.

Appendix H

Record of Modification of the Gaussian 16 Source Code

If you are interested in special details about the pURVA codes, please contact Dr. Elfi Kraka, Director of the Computational and Theoretical Chemistry Group @ SMU.

Appendix I

Source Code of CURVCOR for Reaction Path Curvature Correction

Contents

Listings

1 CURVCOR for FCHK Files

This script is designed to calculate the correct reaction path curvature vector for the TS region based on Gaussian formatted checkpoint files. It is mostly used for large reaction systems.

```
# This script is designed for calculating the CORRECTED
# curvature values/vectors for points along the reaction path where
# these points are quite close to the transition state(TS) point.
#
# The reason for designing this script is that previous URVA method
# was designed for points along the path where the gradient is
    signifcantly large.
#
# However, for points around TS, the gradient is almost zero. This
    cause
# the following problems:
#
# 1) The older curvature calculation equation can not be applied any
    longer
# 2) Non-sense spikes arise for curvature values around TS points
# 3) The path direction does not make sense if gradient direction is
    used
# 4) Component analysis (see Wenli Zou's JCTC paper) can not function
    properly
#
# In order to cure these problems, correct curvature and path direction
     are
# calculated here according to a new set of equations.
#
# Warnings
# * This script cannot be applied to points located in entrance and
    exit channel
#   where the gradient also vanishes.
#
# * In choosing the points which need to be corrected, special care
    should be taken care of
#
#

# How to use this script?
# 1. set up running mode(MODE).
#    a) = 1, path direction only
#    b) = 2, path direction with curvature
#
# 2. specify input formatted checkpoint files(Fn,F1,Fp)
#    F1 = <filename.fchk> of the point in question
#    Fn = <filename.fchk> of the point before F1
#    Fp = <filebane.fchk> of the point after F1
#
# Limitation:
# 1) This script could only be used for new browsing files without
    Hessian stored
#    (For old browsing file, the systems are much smaller and the PES
    is more steep
#     than in large systems.)
#
```

```
#

from __future__ import print_function
import numpy as np
import pprint
import sys
import gc
import copy

#### SET UP BEFORE RUNNING ####

MODE = 2
NEGASIGN = 0
NEGASIGN2= 0
Fl = "../pd-example/triple_stepsize/r-04.fchk"  # IRC Point of Interest
Fn = "../pd-example/triple_stepsize/r-05.fchk"
Fp = "../pd-example/triple_stepsize/r-03.fchk"

##############################

def chunks(l, n):
    """Yield successive n-sized chunks from l."""
    for i in range(0, len(l), n):
        yield l[i:i + n]

def LT2Sqr(N,T):
    "Transform a lower triangular vector into a square matrix"
    S=[]
    for i in range(N):
        S.append( [] )
    for i in range(N):
        for j in range(N):
            S[i].append(0)


    k = 0 -1 # corrected for python
    for i in range(N):
        if i == 1:
            k = k + 1
            S[i][0] = T[k]
            S[0][i] = T[k]
            #print S
        if i != 1:
         for j in range(i):
            k = k + 1
            S[i][j] = T[k]
            S[j][i] = T[k]
            #print S
        k = k + 1
        S[i][i] = T[k]
        #print S

    return S
##################################


```

```
99
100 # Start the main script
101
102 F1h=open(F1,"r")
103 print ("Reading file F1: "+F1)
104 F1all=F1h.readlines()
105
106 Fgrad=0
107 Vgrad=[]
108 pgrad=1
109 Fimode=0
110 Vimode=[]
111 pimode=1
112 Fmass=0
113 Vmass=[]
114 pmass=1
115 for i in F1all:
116   if  len(i)> 0:
117      if i.split()[0]=="Number":
118        if i.split()[1]=="of":
119          if i.split()[2]=="atoms":
120              NAtom = int( i.split()[4] )
121
122      if i.split()[0]=="Cartesian":
123        if i.split()[1]=="Gradient":
124          numgrad = int( i.split()[4] )
125          nrowgrad = numgrad / 5
126          if numgrad % 5 > 0:
127              nrowgrad = nrowgrad + 1
128          Fgrad=1
129          continue
130      if Fgrad == 1:
131          if pgrad <= nrowgrad:
132             Vgrad.append( i.split()   )
133             pgrad= pgrad +1
134          else:
135             Fgrad = 0
136
137      if i.split()[0]=="Vib-Modes":
138          Fimode=1
139          numimode = 3 * NAtom
140          nrowimode = numimode / 5     # we only need part of this data
     section
141          if numimode % 5 > 0 :
142             nrowimode = nrowimode + 1
143          continue
144      if Fimode == 1:
145          if pimode <  nrowimode:
146             Vimode.append(i.split())
147             pimode = pimode + 1
148             continue
149          if pimode == nrowimode:
150               j= numimode % 5
151               #print (j)
152               #print (i.split())
153               tmp=[]
154               if j!=0:                  # Corrected
```

```
                    for k in range(j):
                        tmp.append (str(i.split()[k]) )
                else:
                    tmp = i.split()
                Vimode.append(tmp)
                pimode = pimode + 1
            if pimode > nrowimode:
                Fimode = 0
        if i.split()[0]=="Real":
            if i.split()[1]=="atomic":
                if i.split()[2]=="weights":
                    Fmass=1
                    nummass=int(i.split()[5])
                    nrowmass = nummass / 5
                    if nummass % 5 > 0:
                        nrowmass = nrowmass + 1
                    continue
        if Fmass == 1:
            if pmass < nrowmass:
                Vmass.append(i.split())
                pmass = pmass + 1
                continue
            if pmass ==nrowmass:
                j= nummass % 5
                tmp = []
                if j != 0:     # Corrected
                    for k in range(j):
                        tmp.append(str(i.split()[k]))
                else:
                    tmp = i.split()
                Vmass.append(tmp)
                pmass = pmass + 1
            if pimode > nrowmass:
                Fmass = 0



#print (Vmass)
Vmass = [item for sublist in Vmass for item in sublist]
Vmass = [float(i) for i in Vmass]
Vgrad = [item for sublist in Vgrad for item in sublist]
Vgrad = [float(i) for i in Vgrad]
Vimode = [item for sublist in Vimode for item in sublist]
Vimode = [float(i) for i in Vimode]
#print (Vimode)
Veta   = copy.deepcopy(Vimode)

#print (Vmass)
#print (Vgrad)

tmpc=[]
for i in range(len(Vgrad)):
    j = i / 3
    tmpc.append( Vgrad[i] / (Vmass[j]**0.5) )
cinv = 1/np.linalg.norm( tmpc )

```

```
212 print ("1/|C| = ",cinv)
213
214
215
216 # PRINT RESULT #1
217 #
      ##############################################################################

218
219 if MODE ==1: # Print New Eta from imaginary mode vector
220  print ("MODE = 1")
221
222  if NEGASIGN == 1:
223     print ("Reaction Path direction reversed by user.")
224
225  print ("V1: Corrected Path Direction vector[UNMASS-WEIGHTED] is
        printed:")
226  for i in range(numimode):
227    if NEGASIGN == 0:
228      print (    "%22.15E"        %   Vimode[i],end="")
229    else:
230      tmp = Vimode[i] * -1
231      print (    "%22.15E"        %         tmp,end="")
232    j=i+1
233    if j%3==0:
234       print ("")
235  print ("Above numbers are for test use.")
236  #print ("Calculations Done.")
237  #print (np.linalg.norm(Vimode))
238
239  del Veta
240  Veta = copy.deepcopy(Vimode)
241  for i in range(NAtom):
242     for j in range(3):
243        Veta[ 3*i+j ] = Vimode[ 3*i+j ]*1/np.sqrt( Vmass[i] )
244  tmp=np.linalg.norm(Veta)
245  for i in range(NAtom):
246     for j in range(3):
247        Veta[ 3*i+j ] = Veta[ 3*i+j ] / tmp
248  print ("V2: Corrected Path Direction vector[MASS-WEIGHTED] is printed:
        ")
249  for i in range(numimode):
250    if NEGASIGN == 0:
251      print (    "%22.15E"        %   Veta[i],end="")
252    else:
253      tmp = Veta[i] * -1
254      print (    "%22.15E"        %       tmp,end="")
255    j=i+1
256    if j%3==0:
257       print ("")
258  print ("Please paste above numbers into new browsing file to replace
        the old Eta.")
259
260  #print (Vimode)
261  del Veta
262  Veta = copy.deepcopy(Vimode)
263  for i in range(NAtom):
```

```
264     for j in range(3):
265        Veta[ 3*i+j ] = Vimode[ 3*i+j ]*1*np.sqrt( Vmass[i] )
266  tmp=np.linalg.norm(Veta)
267  for i in range(NAtom):
268     for j in range(3):
269        Veta[ 3*i+j ] = Veta[ 3*i+j ] / tmp
270  print ("V3: Corrected Path Direction vector[MASS-WEIGHTED] is printed:
        ")
271  for i in range(numimode):
272    if NEGASIGN == 0:
273      print (    "%22.15E"        %   Veta[i],end="")
274    else:
275      tmp = Veta[i] * -1
276      print (    "%22.15E"        %       tmp,end="")
277    j=i+1
278    if j%3==0:
279       print ("")
280
281  print ("Please paste above numbers into new browsing file to replace
        the old Eta.")
282
283  print ("Calculations Done.")
284
285  #print (np.linalg.norm(Veta))
286  #print (len(Vimode))
287  #print (len(Vmass))
288 #
        ##############################################################################
289
290
291 if MODE ==2: # Mode 2
292  print ("MODE = 2")
293  #print (type(Veta),type(Vimode))
294  Veta = copy.deepcopy(Vimode)
295  for i in range(NAtom):
296     for j in range(3):
297        Veta[ 3*i+j ] = Vimode[ 3*i+j ]*1*np.sqrt( Vmass[i] )
298  tmp=np.linalg.norm(Veta)
299  for i in range(NAtom):
300     for j in range(3):
301        Veta[ 3*i+j ] = Veta[ 3*i+j ] / tmp
302
303  print ("Corrected Path Direction vector[MASS-WEIGHTED] is printed:")
304  for i in range(numimode):
305    if NEGASIGN == 0:
306      print (    "%22.15E"        %   Veta[i],end="")
307    else:
308      tmp = Veta[i] * -1
309      print (    "%22.15E"        %       tmp,end="")
310    j=i+1
311    if j%3==0:
312       print ("")
313  print ("Please paste above numbers into new browsing file to replace
        the old Eta.")
314
315 ## F1
```

```
Fhess=0
phess=1
Vhess1=[]
Fcoor=0
pcoor=1
Vcoor1=[]


for i in F1all:
  if  len(i)> 0:
     if i.split()[0]=="Current":
       if i.split()[1]=="cartesian":
         if i.split()[2]=="coordinates":
            numcoor = int(i.split()[5])
            nrowcoor = numcoor / 5
            if numcoor % 5 > 0:
               nrowcoor = nrowcoor + 1
            Fcoor=1
            continue
     if Fcoor == 1:
         if pcoor < nrowcoor:
            Vcoor1.append(i.split())
            pcoor = pcoor + 1
            continue
         if pcoor == nrowcoor:
            j = numcoor % 5
            tmp = []

            if j != 0:
               for k in range(j):
                   tmp.append(str(i.split()[k]))
            else:
               tmp = i.split()
            Vcoor1.append(tmp)
            pcoor = pcoor + 1
         if pcoor > nrowcoor:
            Fcoor = 0


     if i.split()[0]=="Cartesian":
       if i.split()[1]=="Force":
         if i.split()[2]=="Constants":
            numhess = int(i.split()[5])
            nrowhess = numhess / 5
            if numhess % 5 > 0:
               nrowhess = nrowhess + 1
            Fhess=1
            continue
     if Fhess == 1:
         if phess < nrowhess:
            Vhess1.append(i.split())
            phess = phess + 1
            continue
         if phess ==nrowhess: # Corrections
            j = numhess % 5
            tmp=[]
            if j != 0:
```

```
                for k in range(j):
                    tmp.append(str(i.split()[k]))
            else:
                tmp = i.split()
            Vhess1.append(tmp)
            phess = phess + 1
        if phess > nrowhess:
            Fhess = 0

Vcoor1 = [item for sublist in Vcoor1 for item in sublist]
Vcoor1 = [float(i) for i in Vcoor1]

Vhess1 = [item for sublist in Vhess1 for item in sublist]
Vhess1 = [float(i) for i in Vhess1]


#print (Vhess1)


Fph=open(Fp,"r")
print ("Reading file Fp: "+Fp)
Fpall=Fph.readlines()
Fnh=open(Fn,"r")
print ("Reading file Fn: "+Fn)
Fnall=Fnh.readlines()


### for P point
Fhess=0
phess=1
Vhessp=[]
Fcoor=0
pcoor=1
Vcoorp=[]
Fgrad=0
Vgradp=[]
pgrad=1

for i in Fpall:
  if  len(i)> 0:
     if i.split()[0]=="Current":
       if i.split()[1]=="cartesian":
         if i.split()[2]=="coordinates":
            numcoor = int(i.split()[5])
            nrowcoor = numcoor / 5
            if numcoor % 5 > 0:
               nrowcoor = nrowcoor + 1
            Fcoor=1
            continue
     if Fcoor == 1:
         if pcoor < nrowcoor:
            Vcoorp.append(i.split())
            pcoor = pcoor + 1
            continue
         if pcoor == nrowcoor:
            j = numcoor % 5
            tmp = []
```

```
430
431             if j != 0:
432                for k in range(j):
433                    tmp.append(str(i.split()[k]))
434             else:
435                tmp = i.split()
436             Vcoorp.append(tmp)
437             pcoor = pcoor + 1
438          if pcoor > nrowcoor:
439             Fcoor = 0
440
441       if i.split()[0]=="Cartesian":
442        if i.split()[1]=="Gradient":
443          numgrad = int( i.split()[4] )
444          nrowgrad = numgrad / 5
445          if numgrad % 5 > 0:
446              nrowgrad = nrowgrad + 1
447          Fgrad=1
448          continue
449       if Fgrad == 1:
450          if pgrad <= nrowgrad:
451             Vgradp.append( i.split()   )
452             pgrad= pgrad +1
453          else:
454             Fgrad = 0
455
456
457       if i.split()[0]=="Cartesian":
458         if i.split()[1]=="Force":
459           if i.split()[2]=="Constants":
460              numhess = int(i.split()[5])
461              nrowhess = numhess / 5
462              if numhess % 5 > 0:
463                 nrowhess = nrowhess + 1
464              Fhess=1
465              continue
466       if Fhess == 1:
467           if phess < nrowhess:
468              Vhessp.append(i.split())
469              phess = phess + 1
470              continue
471           if phess ==nrowhess:
472              j = numhess % 5
473              tmp=[]
474
475              if j != 0:
476                 for k in range(j):
477                     tmp.append(str(i.split()[k]))
478              else:
479                 tmp = i.split()
480
481              Vhessp.append(tmp)
482              phess = phess + 1
483           if phess > nrowhess:
484              Fhess = 0
485
486  Vhessp = [item for sublist in Vhessp for item in sublist]
```

```
Vhessp = [float(i) for i in Vhessp]
Vcoorp = [item for sublist in Vcoorp for item in sublist]
Vcoorp = [float(i) for i in Vcoorp]
Vgradp = [item for sublist in Vgradp for item in sublist]
Vgradp = [float(i) for i in Vgradp]


#print (Vhessp)
#print (Vcoorp)


### for N point
Fhess=0
phess=1
Vhessn=[]
Fcoor=0
pcoor=1
Vcoorn=[]
Fgrad=0
Vgradn=[]
pgrad=1


for i in Fnall:
  if  len(i)> 0:
     if i.split()[0]=="Current":
       if i.split()[1]=="cartesian":
         if i.split()[2]=="coordinates":
            numcoor = int(i.split()[5])
            nrowcoor = numcoor / 5
            if numcoor % 5 > 0:
               nrowcoor = nrowcoor + 1
            Fcoor=1
            continue
     if Fcoor == 1:
         if pcoor < nrowcoor:
            Vcoorn.append(i.split())
            pcoor = pcoor + 1
            continue
         if pcoor == nrowcoor:
            j = numcoor % 5
            tmp = []
            if j != 0:
               for k in range(j):
                   tmp.append(str(i.split()[k]))
            else:
               tmp = i.split()
            Vcoorn.append(tmp)
            pcoor = pcoor + 1
         if pcoor > nrowcoor:
            Fcoor = 0

     if i.split()[0]=="Cartesian":
      if i.split()[1]=="Gradient":
        numgrad = int( i.split()[4] )
        nrowgrad = numgrad / 5
        if numgrad % 5 > 0:
```

```
544             nrowgrad = nrowgrad + 1
545          Fgrad=1
546          continue
547       if Fgrad == 1:
548          if pgrad <= nrowgrad:
549             Vgradn.append( i.split()   )
550             pgrad= pgrad +1
551          else:
552             Fgrad = 0
553
554
555       if i.split()[0]=="Cartesian":
556         if i.split()[1]=="Force":
557           if i.split()[2]=="Constants":
558             numhess = int(i.split()[5])
559             nrowhess = numhess / 5
560             if numhess % 5 > 0:
561                nrowhess = nrowhess + 1
562             Fhess=1
563             continue
564       if Fhess == 1:
565           if phess < nrowhess:
566             Vhessn.append(i.split())
567             phess = phess + 1
568             continue
569          if phess ==nrowhess:
570             j = numhess % 5
571             tmp=[]
572
573             if j != 0:
574                for k in range(j):
575                    tmp.append(str(i.split()[k]))
576             else:
577                tmp = i.split()
578
579             Vhessn.append(tmp)
580             phess = phess + 1
581          if phess > nrowhess:
582             Fhess = 0
583
584
585 del F1all
586 gc.collect()
587 del Fpall
588 gc.collect()
589 del Fnall
590 gc.collect()
591
592
593 Vhessn = [item for sublist in Vhessn for item in sublist]
594 Vhessn = [float(i) for i in Vhessn]
595 Vcoorn = [item for sublist in Vcoorn for item in sublist]
596 Vcoorn = [float(i) for i in Vcoorn]
597 Vgradn = [item for sublist in Vgradn for item in sublist]
598 Vgradn = [float(i) for i in Vgradn]
599
600
```

```
601 #print (Vhessn)
602 #print (Vcoorn)
603
604 # SUMMARY #
605 # Mass:                          Vmass
606 # Hessian:            Vhess1,  Vhessp, Vhessn
607 # 1st Normal mode:    Vimode
608 # Coordinates:        Vcoor1, Vcoorp, Vcoorn
609 # gradient:           Vgrad, Vgradp, Vgradn
610
611 #print (len(Vmass))  = NAtom
612 #print (len(Vcoorn)) = NAtom * 3
613
614 s = 0
615 for i in range(len(Vcoorn)):
616       j = i / 3
617       tmp = ((Vcoorn[i] - Vcoorp[i])**2 ) * Vmass[j]
618       s = s + tmp
619 ds = (s**0.5)/2
620 a1= "Step size of IRC is "
621 a2= " amu^(1/2)*Bohr"
622 print ("%s %7.4F %s" %  (a1,ds,a2  ))
623
624 s = 0
625 for i in range(len(Vcoor1)):
626       j = i / 3
627       tmp = ((Vcoor1[i] - Vcoorp[i])**2 ) * Vmass[j]
628       s = s + tmp
629 dsp = (s**0.5)/1
630
631 s = 0
632 for i in range(len(Vcoor1)):
633       j = i / 3
634       tmp = ((Vcoor1[i] - Vcoorn[i])**2 ) * Vmass[j]
635       s = s + tmp
636 dsn = (s**0.5)/1
637 a3 = "dsn: "
638 a4 = "dsp: "
639 print ("%s %s %7.4F %s" %  (a3,a1,dsn,a2  ))
640 print ("%s %s %7.4F %s" %  (a4,a1,dsp,a2  ))
641
642
643 ########################################
644 ## mass weighting Hessian
645 #print ( np.matrix(LT2Sqr(4,[1,2,3,4,5,6,7,8,9,10])) )
646 #print (len(Vhess1),len(Vhessp),len(Vhessn)   )
647 ffM1 =  LT2Sqr( 3*NAtom, Vhess1 )
648 ffMp =  LT2Sqr( 3*NAtom, Vhessp )
649 ffMn =  LT2Sqr( 3*NAtom, Vhessn )
650 del Vhess1
651 gc.collect()
652 del Vhessp
653 gc.collect()
654 del Vhessn
655 gc.collect()
656 for i in range(3*NAtom):
657     for j in range(3*NAtom):
```

```
658           im = i / 3
659           jm = j / 3
660           ffMp[i][j] = ffMp[i][j] / ( (Vmass[im])**0.5 * (Vmass[jm
      ]**0.5) )
661           ffMn[i][j] = ffMn[i][j] / ( (Vmass[im])**0.5 * (Vmass[jm
      ]**0.5) )
662           ffM1[i][j] = ffM1[i][j] / ( (Vmass[im])**0.5 * (Vmass[jm
      ]**0.5) )
663
664   #print (type(ffMp))
665
666   ## F^(1)  -> derivative of F(s0) with regard to ds
667   ffMp = np.matrix( ffMp )
668   ffMn = np.matrix( ffMn )
669   mF1 = (ffMp - ffMn ) / (dsn+dsp)
670   del ffMn
671   gc.collect()
672   del ffMp
673   gc.collect()
674
675   ## Summary
676   # F     -> ffM1 (list)
677   # F^(1) -> mF1  (matrix)
678   # v     -> Veta (list)
679
680   ffM1 = np.matrix(ffM1)
681   Veta = np.array(Veta)
682
683   tmp = (Veta * ffM1 )      ## v+ F
684   tmp = np.asarray( tmp )  # change from matrix into array
685   #print (type(tmp))
686   a = np.dot(tmp, Veta)     ## v+ F v
687   #print (a)
688   del tmp
689   gc.collect()
690
691   I = np.identity( 3*NAtom )
692   I = I * 2 * a
693   #print (I)
694   Left = np.matrix(I) - ffM1  # symmetric
695
696   #gc.collect()
697   #print (Left)
698   Leftinv = np.linalg.inv( Left )
699   #print (Leftinv)
700   del Left
701   gc.collect()
702
703   #print (type(mF1))
704   #print (type(Veta))
705   tmp = mF1.dot(Veta)      ## F^(1) v
706   #print (mF1.dot(Veta))
707   del mF1
708   gc.collect()
709   tmp = np.asarray(tmp)[0]
710
711   #print (len(tmp),len(Veta))
```

```
712 a = np.dot(Veta,tmp)    ## v+ F^(1) v
713 #print (a)
714 av = Veta * a            ## v+ F^(1) v v
715 #print (len(tmp),len(av))
716 Right = tmp - av
717
718 del tmp
719 gc.collect()
720 del av
721 gc.collect()
722 #print (Right)
723
724 cur = Leftinv.dot(Right)
725 tt = cur.tolist()
726 #print (tt[0])
727 print ("Corrected Curvature vector[MASS-WEIGHTED] is printed:")
728 for i in range(numimode):
729   if NEGASIGN2 == 0:
730     pass
731     print (    "%22.15E"        %   tt[0][i],end="")
732   else:
733     tmp = tt[0][i] * -1
734     pass
735     print (    "%22.15E"        %       tmp,end="")
736   j=i+1
737   if j%3==0:
738      pass
739      print ("")
740 print ("Please paste above numbers into new browsing file to replace
      the old Kappa.")
741
742 #print (cur)
743 a1="Scalar curvature(   corrected) for this point is"
744 a2="amu^(-1/2)*Bohr^(-1)"
745 print ( "%s %10.5F %s" % ( a1, np.linalg.norm(cur),a2 )  )
746 #print ("Calculations Done.")
747 #print(len(Vgrad),len(Vgradp),len(Vgradn))
748 #print (len(Veta))
749
750 Vetap = copy.deepcopy(Vgradp)
751 for i in range(NAtom):
752    for j in range(3):
753       Vetap[ 3*i+j ] = Vetap[ 3*i+j ]*1/np.sqrt( Vmass[i] )
754 tmp=np.linalg.norm(Vetap)
755 for i in range(NAtom):
756    for j in range(3):
757       Vetap[ 3*i+j ] = Vetap[ 3*i+j ] / tmp
758
759 Vetan = copy.deepcopy(Vgradn)
760 for i in range(NAtom):
761    for j in range(3):
762       Vetan[ 3*i+j ] = Vetan[ 3*i+j ]*1/np.sqrt( Vmass[i] )
763 tmp=np.linalg.norm(Vetan)
764 for i in range(NAtom):
765    for j in range(3):
766       Vetan[ 3*i+j ] = Vetan[ 3*i+j ] / tmp
767
```

```
768 #curap = ( np.array(Vetap) - np.array(Veta) ) / (1*ds)  # GARBAGE! #
       Error too large!
769 #a3="Scalar curvature(approximated) for this point is"
770 #print ( "%s %10.5F %s" % ( a3, np.linalg.norm(curap),a2 )  )
771
772
773 # Give also the curvature with original formula
774 # 1/C -> cinv
775 # f(mass-weighted) -> ffM1
776 # g(mass-weighted) -> tmpc
777
778 tmpc = np.array(tmpc)
779 ffM1 = np.array(ffM1)
780 tmpeta = []
781 for i in range(len(tmpc)):
782     tmpeta.append( tmpc[i] * cinv )
783 tmpeta = np.array(tmpeta)
784 tmpv=[]
785 for i in range(len(tmpeta)):
786     tmpv.append( np.dot(tmpeta, ffM1[i] ) )
787 gfg = np.dot( tmpeta, tmpv )
788 result=[]
789
790 for i in range(len(tmpeta)):
791    result.append( tmpv[i] - gfg* tmpeta[i] )
792 for i in range(len(result)):
793    result[i] = result[i] * cinv
794
795 a1="Scalar curvature(un-corrected) for this point is"
796 print ( "%s %10.5F %s" % ( a1, np.linalg.norm(result),a2 )  )
797
798 #print (len(tmpv))
799
800 #print (np.linalg.norm(tmpeta))
801 #print (len(tmpc))
802 #print (len(ffM1))
803
804 print ("This point is "+F1)
805 print ("Calculations Done.")
```

Listing 1: CURVCOR Script for FCHK Files

2 CURVCOR for Old Browsing Files

This script is designed to calculate the corrected scalar curvature for the TS region based on old browsing files. The correct reaction path curvature vector is also calculated but not shown. This script is later adapted into the pURVA program.

Besides the main script, it has another two modules named *util.py* and *util2.py*.

```
# This script is identical to CURVCOR, however, it is tailored for
# old browsing file.
#
# This script could later be adapted to pURVA...
#
# USER INPUT:
#  1. browsing file name
#  2. labels of F1, Fp, Fn (e.g. P2, P3, P1)
#
# INPUT INFO.:
#  1. CC
#  2. Mass
#  3. Hessian
#


import sys
from util2 import *
from util import *
import numpy as np
import gc

#print sys.argv

inpf = sys.argv[1]
v1=  sys.argv[2]
v2=  sys.argv[3]
v3=  sys.argv[4]
inpstr = [v1,v2,v3]
#inpstr = [ 'P3','P4','P5' ]
#inpf = "IRC.browse"

#################
origin = 0.0
#################

a,b,c=parsestr(inpstr)

if a >= b or a >= c or b >= c:
   print ("Error: Please check input!")
   sys.exit()
else:
   d1 = b-a
   d2 = c-b
   if d1 != d2:
      print ("Error: Please check input!")
      sys.exit()
```

```
48
49
50 # Get the info. of s values
51 svals = []
52 svals = initscan(inpf)
53 sa,sb,sc = assigns(svals,a,b,c,origin)
54
55 print ("Reaction parameter s: "+str(sb))
56
57
58 # Get mass, CC, Grad, Hess for corresponding s value
59 Vmass, VcoorA,VgradA, VhessA = procbrow(sa,inpf)
60 Vmass, VcoorB,VgradB, VhessB = procbrow(sb,inpf)
61 Vmass, VcoorC,VgradC, VhessC = procbrow(sc,inpf)
62
63 # Get the normal modes
64 eig,nm,cdisp=solWils(VcoorB,VhessB,Vmass)  # This part might cause
       problem in the curve
65 #print len(nm)
66 #print len(cdisp)
67
68 VcoorA = [item for sublist in VcoorA for item in sublist]
69 VcoorB = [item for sublist in VcoorB for item in sublist]
70 VcoorC = [item for sublist in VcoorC for item in sublist]
71
72 a1= "Step size of IRC is "
73 a2= " amu^(1/2)*Bohr"
74 a3 = "dsn: "
75 a4 = "dsp: "
76
77 s = 0
78 for i in range(len(VcoorB)):
79        j = i / 3
80        tmp = ((VcoorA[i] - VcoorB[i])**2 ) * Vmass[j]
81        s = s + tmp
82 dsp = (s**0.5)/1
83
84 s = 0
85 for i in range(len(VcoorB)):
86       j = i / 3
87       tmp = ((VcoorB[i] - VcoorC[i])**2 ) * Vmass[j]
88       s = s + tmp
89 dsn = (s**0.5)/1
90
91 print ("%s %s %7.4F %s" %  (a3,a1,dsp,a2  ))
92 print ("%s %s %7.4F %s" %  (a4,a1,dsn,a2  ))
93
94 NAtom = len(Vmass)
95
96 ffM1 =  LT2Sqr( 3*NAtom, VhessB )
97 ffMp =  LT2Sqr( 3*NAtom, VhessC )
98 ffMn =  LT2Sqr( 3*NAtom, VhessA )
99
100 del VhessA
101 gc.collect()
102 #del VhessB
103 #gc.collect()
```

```
104 del VhessC
105 gc.collect()
106
107 for i in range(3*NAtom):
108     for j in range(3*NAtom):
109         im = i / 3
110         jm = j / 3
111         ffMp[i][j] = ffMp[i][j] / ( (Vmass[im])**0.5 * (Vmass[jm]**0.5)
      )
112         ffMn[i][j] = ffMn[i][j] / ( (Vmass[im])**0.5 * (Vmass[jm]**0.5)
      )
113         ffM1[i][j] = ffM1[i][j] / ( (Vmass[im])**0.5 * (Vmass[jm]**0.5)
      )
114
115 ffMp = np.matrix( ffMp )
116 ffMn = np.matrix( ffMn )
117 mF1 = (ffMp - ffMn ) / (dsn+dsp)
118 del ffMn
119 gc.collect()
120 del ffMp
121 gc.collect()
122
123 ffM1 = np.matrix(ffM1)
124 Veta = np.array( nm[0]  ) # Might change
125
126 tmp = (Veta * ffM1 )
127 tmp = np.asarray( tmp )
128 a = np.dot(tmp, Veta)
129 del tmp
130 gc.collect()
131
132 I = np.identity( 3*NAtom )
133 I = I * 2 * a
134
135 Left = np.matrix(I) - ffM1
136
137 Leftinv = np.linalg.inv( Left )
138
139 del Left
140 gc.collect()
141
142 tmp = mF1.dot(Veta)
143 del mF1
144 gc.collect()
145 tmp = np.asarray(tmp)[0]
146
147 a = np.dot(Veta,tmp)
148
149 av = Veta * a
150
151 Right = tmp - av
152
153 del tmp
154 gc.collect()
155 del av
156 gc.collect()
157
```

```
158 cur = Leftinv.dot(Right)
159 tt = cur.tolist()
160
161 a1="Scalar curvature(   corrected) for this point is"
162 a2="amu^(-1/2)*Bohr^(-1)"
163 print ( "%s %10.5F %s" % ( a1, np.linalg.norm(cur),a2 )  )
164
165
166 tmpc = []
167 for i in range(len(VgradB)):
168     j = i / 3
169     v = VgradB[i] * ( Vmass[j]**(-0.5) )
170     tmpc.append(v)
171
172 tmpc = np.array(tmpc)
173 cinv = 1.0 / np.linalg.norm(tmpc)
174
175 ffM1 = np.array(ffM1)
176 tmpeta = []
177 for i in range(len(tmpc)):
178     tmpeta.append( tmpc[i] * cinv )
179 tmpeta = np.array(tmpeta)
180 tmpv=[]
181
182 for i in range(len(tmpeta)):
183     tmpv.append( np.dot(tmpeta, ffM1[i] ) )
184 gfg = np.dot( tmpeta, tmpv )
185 result=[]
186
187 for i in range(len(tmpeta)):
188    result.append( tmpv[i] - gfg* tmpeta[i] )
189 for i in range(len(result)):
190    result[i] = result[i] * cinv
191
192 a1="Scalar curvature(un-corrected) for this point is"
193 print ( "%s %10.5F %s" % ( a1, np.linalg.norm(result),a2 )  )
194
195 ### Here we calculate the un-mass-weighted scalar curvature.
196 # Gradient -> VgradB
197 # Hessian -> ffM1
198
199 ffM1 =  LT2Sqr( 3*NAtom, VhessB )
200 del VhessB
201 gc.collect()
202
203 #print ffM1
204 #print VgradB
205
206 tmpc = []
207 for i in range(len(VgradB)):
208     v = VgradB[i]
209     tmpc.append(v)
210 tmpc = np.array(tmpc)
211 cinv = 1.0 / np.linalg.norm(tmpc)
212
213 ffM1 = np.array(ffM1)
214 tmpeta = []
```

```
for i in range(len(tmpc)):
    tmpeta.append( tmpc[i] * cinv )
tmpeta = np.array(tmpeta)
tmpv=[]

for i in range(len(tmpeta)):
    tmpv.append( np.dot(tmpeta, ffM1[i] ) )
gfg = np.dot( tmpeta, tmpv )
result=[]

for i in range(len(tmpeta)):
   result.append( tmpv[i] - gfg* tmpeta[i] )
for i in range(len(result)):
   result[i] = result[i] * cinv


a1="Scalar curvature(unmasswghted) for this point is"
a2="Bohr^(-1)"

print ( "%s %10.5F %s" % ( a1, np.linalg.norm(result),a2 )  )
```

Listing 2: CURVCOR Script for Old Browsing Files

```
import numpy as np
from scipy.linalg import sqrtm

## MPACMF
def MPACMF(A,B,IQ):
    "!     IQ=\
     !     1  C=A*B\
     !     2  C=A(TRANSPOSE)*B\
     !     3  C=A*B(TRANSPOSE)"
    # 1: A(L,M),B(M,N),C(L,N)
    #
    A = np.matrix(A)
    B = np.matrix(B)
    if IQ == 1:
       A = A.transpose()
       B = B.transpose()
       C = A * B
       C = C.transpose()
    if IQ == 2:
       B = B.transpose()
       C = A * B
       C = C.transpose()
    if IQ == 3:
       A = A.transpose()
       C = A * B
       C = C.transpose()
    return C

#######################

## MMpyMF
def MMpyMF(A,B):
    # A - (L,M) fortran
    # B - (M,N) fortran
```

```
35      # C - (L,N) fortran
36      # A = matrix( [1,2,3, ...,L],
37      #             [2,...
38      #             [3,...
39      #             [4,...
40      #             [.
41      #             [M
42
43
44      # B = matrix( [1,2,3, ...,M],
45      #             [2,...
46      #             [3,...
47      #             [4,...
48      #             [.
49      #             [N
50
51      # C = matrix( [1,2,3, ...,L],
52      #             [2,...
53      #             [3,...
54      #             [4,...
55      #             [.
56      #             [N
57
58      A = np.matrix(A)
59      B = np.matrix(B)
60      A = A.transpose()
61      B = B.transpose()
62      C = A * B
63      C = C.transpose()
64
65      return C
66  ######################################
67
68  ## MatMP1
69  def MatMp1(A,B):
70      "Interface MatMp1 subroutine into Python"
71      # A - (M,L) fortran
72      #    A[L][M]
73      # B - (M,N) fortran
74      #    B[N][M]
75      # Product - (L,N) fortran
76      #          C[N][L]
77      #
78      #   = matrix( [1,2,3, ...,M],
79      #             [2,...
80      #             [3,...
81      #             [4,...
82      #             [.
83      #             [L
84
85
86      #   = matrix( [1,2,3, ...,M],
87      #             [2,...
88      #             [3,...
89      #             [4,...
90      #             [.
91      #             [N
```

```
92
93      #   = matrix( [1,2,3, ...,L],
94      #             [2,...
95      #             [3,...
96      #             [4,...
97      #             [.
98      #             [N
99
100     A = np.array( A )
101     B = np.array( B )
102     #B = B.transpose()
103     #print A
104     #print B
105     L = len(A)
106     N = len(B)
107     M = len(A[0])
108
109     C=[]
110     for i in range(N):
111   C.append( [] )
112     for i in range(N):
113   for j in range(L):
114       C[i].append(0)
115
116     for i in range(N):
117   for j in range(L):
118             C[i][j] = np.dot( A[j], B[i] )
119     C = np.array(C)
120
121     return C
122 ###############################################
123
124 ## pDiagSq
125 def pDiagSq(Indx,B):
126     "!---  General inverse enhanced version of DiagSqrt.\
127      !---  A is a diagonal matrix with diagonal terms\
128      !---  Indx = 0, sqrt(B)\
129      !---       ne 0, sqrt(B)**-1 if B(i) > 0.\
130      !---  It's assumed that B(i) .ge. 0. It doesn't work if there are\
131      !---  negative elements in B, which is not checked."
132     B = np.array(B)
133     N = len(B)
134     A = np.zeros((N,N))
135     eps = 1.0e-8
136     if Indx == 0:
137        for i in range(N):
138      A[i,i] = (abs(B[i]))**0.5
139     else:
140        for i in range(N):
141            if (abs(B[i]) > eps):
142         A[i,i] = 1 / (abs(B[i]))**0.5
143
144     return A
145 ###############################################
146
147 #pDiagSq(1,[1,2,3,4,5])
148
```

```
#x=np.matrix([[4,0],[0,4]])
#print sqrtm(x)
#print np.linalg.pinv(sqrtm(x))

## SqrtMp
def SqrtMp(Mode,S):
    "!---  Calculate square-root matrix X=S^1/2 and/or its (general) inverse\
     !---  Xi. Here S must be a SYMMETRIC matrix.\
     !---  Mode > 0: Calculate X\
     !---       < 0: Calculate Xi"

    S = np.matrix( S )
    if Mode > 0:
       P = sqrtm( S )

    if Mode < 0:
       P = np.linalg.pinv(  sqrtm( S ) )

    return P

###############


## LT2Sqr
def LT2Sqr(N,T):
    "Transform a lower triangular vector into a square matrix"
    S=[]
    for i in range(N):
        S.append( [] )
    for i in range(N):
        for j in range(N):
            S[i].append(0)

    k = 0 -1 # corrected for python
    for i in range(N):
        if i == 1:
            k = k + 1
            S[i][0] = T[k]
            S[0][i] = T[k]
            #print S
        if i != 1:
         for j in range(i):
            k = k + 1
            S[i][j] = T[k]
            S[j][i] = T[k]
            #print S
        k = k + 1
        S[i][i] = T[k]
        #print S

    return S
############################


# GSortho
```

```
def GSorth(Reorder,N,M,Vec):
    Tol = 0.1
    NGen = 0
    IGen = 0 #- 1 #Corrected

    Scr=[]
    #for i in range(N):
    #    Scr.apend([])

    for i in range(N):
  #for j in range(N):
            Scr.append(0)
########################################

    for Idx in range(M+N): # 0 ... M+N-1
        NGen = NGen + 1
  #Idx2 = Idx + 1
  if Idx <= (M-1):
     #print "check idx"
           for j in range(N):
               Scr[j] = Vec[Idx][j]
        else:
           for j in range(N):
               Scr[j] = 0
     IGen = IGen + 1
           Scr[IGen-1] = 1

  for Jdx in range(NGen-1):
      X = -1 * np.dot(Vec[Jdx],Scr[0:N])
      #print "l2",Scr[0:N]
            for j in range(N):
    Scr[j] = X*Vec[Jdx][j] + Scr[j]
  X = np.dot(Scr[0:N],Scr[0:N])
        X = X**0.5

        if X > Tol:
           #print "YES"
           X = 1/X
     #print Scr[0:N]
           for i in range(N):
         Vec[NGen-1][i] = X*Scr[i]
     #print "NGen=",NGen
     #print Vec[NGen-1]
  else:
     NGen = NGen - 1
  if NGen ==  N:
           break

    if Reorder == 1:
       pass
       #not implemented

    return Vec
#######################

#a=[[0.1,0.4,0.6,0.8],\
#   [0.2,0.3,0.7,0.9],\
```

```
#    [0 ,0 , 0, 0],\
#    [0, 0 , 0, 0] ]
#
#print GSorth(0,4,2,a)


## Print Vib.
def prtVib(eig,disp):
    NVib = len(eig)
    frq = []
    for i in eig:
  if i > 0:
     frq.append((i**0.5)*5140.48715246)
  else:
     frq.append( -1*((-1*i)**0.5)*5140.48715246)
    for i in range(len(frq)):
  print ("Freq %3i =%8.1f"%(i+1,frq[i]))
```

Listing 3: util.py Module of CURVCOR Script for Old Browsing Files

```
import sys
from calc import *

# atomic masses
defatw = [ 0.00000000,  1.00782504,  4.00260325,  7.01600450,
    9.01218250,
          11.00930530, 12.00000000, 14.00307401, 15.99491464,
    18.99840325,
          19.99243910, 22.98976970, 23.98504500, 26.98154130,
    27.97692840,
          30.97376340, 31.97207180, 34.96885273, 39.96238310,
    38.96370790,
          39.96259070, 44.95591360, 47.94794670, 50.94396250,
    51.94050970,
          54.93804630, 55.93493930, 58.93319780, 57.93534710,
    62.92959920,
          63.92914540, 68.92558090, 73.92117880, 74.92159550,
    79.91652050,
          78.91833610, 83.91150640, 84.91170000, 87.90560000,
    88.90540000,
          89.90430000, 92.90600000, 97.90550000,
    98.90630000,101.90370000,

    102.90480000,105.90320000,106.90509000,113.90360000,114.90410000,

    117.90180000,120.90380000,129.90670000,126.90040000,131.90420000,

    132.90542900,137.90500000,138.90610000,139.90530000,140.90740000,

    141.90750000,144.91270000,151.91950000,152.92090000,157.92410000,

    158.92500000,163.92880000,164.93030000,165.93040000,168.93440000,

    173.93900000,174.94090000,179.94680000,180.94800000,183.95100000,

    186.95600000,189.95860000,192.96330000,194.96480000,196.96660000,

```

```
    201.97060000,204.97450000,207.97660000,208.98040000,208.98250000,
22
    210.98750000,222.01750000,223.01980000,226.02540000,227.02780000,
23
    232.03820000,231.03590000,238.05080000,237.04800000,242.05870000,
24
    243.06140000,246.06740000,247.07020000,249.07480000,252.08290000,
25
    252.08270000,255.09060000,259.10100000,262.10970000,261.10870000,
26      262.11410000,266.12190000,264.12470000,
    0.00000000,268.13880000]
27
28 def parsestr(strlist):
29     a = strlist[0]
30     b = strlist[1]
31     c = strlist[2]
32     if len(a) < 2 or len(b) < 2 or len(c) < 2:
33        print ("Error: Please check the input string!")
34        sys.exit()
35
36     if "P" in a:
37         sa = 1
38     else:
39         sa = -1
40     if "P" in b:
41         sb = 1
42     else:
43         sb = -1
44     if "P" in c:
45         sc = 1
46     else:
47         sc = -1
48
49     if a == "TS":
50        va = 0
51     else:
52        va = int( a[1:] ) *sa
53     if b == "TS":
54        vb = 0
55     else:
56        vb = int( b[1:] ) *sb
57     if c == "TS":
58        vc = 0
59     else:
60        vc = int( c[1:] ) *sc
61
62     #print va,vb,vc
63     return (va,vb,vc)
64
65 def initscan(f1):
66     sval = []
67     p = 0
68     with open(f1) as f:
69       for line in f:
70         if "XXIRC" in line:
71            p = 1
72            continue
```

```
        if p == 1:
           s = float( line.split()[-1]  )
           sval.append(s)
           p = 0
    for i in range(len(sval)-1):
        if sval[i] >= sval[i+1]:
           print ("Error: Reaction coordinates "+str(sval[i] )+" "+str(
    sval[i+1])+"\n"  )

    return sval

def assigns(svals,a,b,c,origin):
    of = 0
    for i in range(len(svals)):
        if svals[i] == origin:
            of = 1
            oi = i
    if of != 1:
       print ("origin error!")
       sys.exit()

    ai = oi + a
    bi = oi + b
    ci = oi + c

    if (ai+1) > len(svals) or (bi+1) > len(svals) or (ci+1) > len(svals
    ):
       print ("Error: input exceeding limit...")
       sys.exit()

    sa = svals[ai]
    sb = svals[bi]
    sc = svals[ci]

    return sa,sb,sc
    #print ai,bi,ci


def procbrow(sval,inpf):
    cut =[]
    p = 0
    with open(inpf) as f:
        for line in f:
          if "BEGIN" in line:
              p = 1
              cut.append(line)
              continue
          if p == 1:
              cut.append(line)
              if "END" in line:
                 p = 2
              if p == 2:
                 mat = checkmatch(sval,cut) # check if match
                 if mat == 1:
                    Vmass,Vcoor,Vgrad,Vhess = takedat(cut)

                 p = 3
```

```
128                 cut = []
129             continue
130
131     return (Vmass, Vcoor, Vgrad, Vhess)
132
133 def checkmatch(sval,cut):
134     p =0
135     match = 0
136     test = -909.909
137     for i in range(len(cut)):
138         if "XXIRC" in cut[i]:
139             p = 1
140             continue
141         if p == 1 :
142            test = float( cut[i].split()[-1] )
143            p = 0
144     if sval ==  test:
145           match = 1
146
147     #print match
148     return match
149
150
151 def takedat(cut):
152     pnz = 0
153     pms = 0
154     Vmass=[]
155     pcoor = 0
156     Vcoor=[]
157     pgrad = 0
158     Vgrad=[]
159     #print cut
160     phess = 0
161     Vhess=[]
162
163
164     for i in range(len(cut)):
165         line = cut[i]
166         if "NZ" in line:
167            pnz = 1
168            continue
169         if pnz == 1:
170            NAtom = int(line.split()[0])
171            #print NAtom
172            pnz = 0
173         if "IAnZ" in line:
174            pms = 1
175            continue
176         if pms > 0:
177            an = int(line.split()[0])
178            mass = defatw[an]
179            #print an,mass
180            Vmass.append( mass )
181            pms = pms + 1
182            if pms == (NAtom + 1):
183               pms = -1
184         if "CC" in line:
```

```
185             pcoor = 1
186             continue
187          if pcoor > 0:
188             line = line.replace("D","E",3)
189             c1 = line[0:22]
190             c2 = line[22:44]
191             c3 = line[44:66]
192             Vcoor.append([ float(c1), float(c2), float(c3) ])
193             #print c1,c2,c3
194             pcoor = pcoor + 1
195             if pcoor == (NAtom + 1):
196                pcoor = -1
197             continue
198          if len(line) > 3:
199            if "FX" in line[:2]:
200              pgrad = 1
201              continue
202          if pgrad > 0:
203             #print pgrad,NAtom
204             line = line.replace("D","E",3)
205             c1 = line[0:22]
206             c2 = line[22:44]
207             c3 = line[44:66]
208             #print c1,c2,c3
209             Vgrad.append( float(c1))
210             Vgrad.append( float(c2))
211             Vgrad.append( float(c3))
212             pgrad = pgrad + 1
213             if pgrad == (NAtom + 1):
214                pgrad = -1
215             continue
216          if "FFX" in line:
217             phess = 1
218             continue
219          if phess > 0:
220             line = line.replace("D","E",3)
221             c1 = line[0:22]
222             c2 = line[22:44]
223             c3 = line[44:66]
224             #print c1,c2,c3
225             Vhess.append( float(c1))
226             Vhess.append( float(c2))
227             Vhess.append( float(c3))
228             phess = phess + 1
229             if phess == ( NAtom*(3*NAtom+1)/2 + 1 ):
230                phess = -1
231             continue
232
233       #print NAtom
234       #print len(Vcoor)
235       #print len(Vmass)
236       #print len(Vgrad)
237       #print len(Vhess)
238       return (Vmass,Vcoor,Vgrad,Vhess)
```

Listing 4: util2.py Module of CURVCOR Script for Old Browsing Files

Appendix J

Source Code of AUTOSMTH for Reaction Path Curvature Smoothing

Contents

Listings

1 AUTOSMTH Script

This script is designed to merge the two curves of uncorrected curvature and corrected curvature into a smooth curve. The whole procedure is automatic.

```
# This script is designed to do
# fully automatic smoothening function
# for the curves including:
# 1) Scalar curvature plot
# 2) components of scalar curvature plot
# 3) components of path direction
#
# The smoothening consists of 2 major steps:
# 1. Using exponential function to merge two ends
# 2. Using cubit spline fitting to correct central part
#
# Input original curve data
# 1. Old curve whose central region is problematic and needs correct
# 2. New curve who started from the central region, however, needs extra work to
#   2.a Make it merge into the old curve at two ends
#   2.b Correct the central region
#
# Note: This script drops most error handles...
#       This script is expected to incorporate into pURVA.

import sys
from scipy import optimize
import math
from scipy.interpolate import interp1d
import numpy as np
from smoothutils import *


# USER SETTING-UP ####################
oldcurvf = 'grmpme/oldcurv.dat'  # uncorrected curvature [two-column]
newcurvf = 'grmpme/newcurv.dat'  #   corrected curvature [two-column]
dstep = 0.03  # Step size of IRC path
RL = [3,3]    # a parameter for spline fitting

continuethresh = 0.5 # Increase this value WHEN NECESSARY
#####################################

############################
# ADVANCED SETTING-UP
#continuethresh = 0.4
prederrthresh = 0.1  # Way No. 1

#RL = [1,1]          # Way No. 2 # Chosen as the default approach
############################


# Read in data
olddat=[]
with open(oldcurvf) as oldf:
     for line in oldf:
      if len(line) > 1:
```

```
52          #print line
53          s = float( line.split()[0] )
54          v = float( line.split()[1] )
55          olddat.append( [s,v] )
56
57 newdat=[]
58 with open(newcurvf) as newf:
59      for line in newf:
60       if len(line) > 1:
61          s = float( line.split()[0] )
62          v = float( line.split()[1] )
63          newdat.append( [s,v] )
64
65
66 newfor=[]
67 chkfor=[]
68 newrev=[]
69 chkrev=[]
70
71 # split forward and reverse
72 for i in range(len(newdat)):
73     if newdat[i][0] >= 0.0:
74        newfor.append( newdat[i] )
75        chkfor.append(0)
76        #print newdat[i]
77
78
79 for i in range(len(newdat)):
80     if newdat[i][0] <= 0.0:
81        newrev.append( newdat[i] )
82        chkrev.append(0)
83 #      print newrev
84
85 # Make sure that new dat and old dat have the same grid point
86 for i in range(len(newfor)):
87     for j in range(len(olddat)):
88        if newfor[i][0] == olddat[j][0]:
89              chkfor[i] = 1
90
91 for i in range(len(newrev)):
92     for j in range(len(olddat)):
93        if newrev[i][0] == olddat[j][0]:
94              chkrev[i] = 1
95
96 if sum(chkfor) != len(chkfor):
97    print ("Error: Grid point not match...")
98    sys.exit()
99 if sum(chkrev) != len(chkrev):
100   print ("Error: Grid point not match...")
101   sys.exit()
102
103
104 # Get the label for TS point
105 TSindex = -1
106 for i in range(len(olddat)):
107     if olddat[i][0] == 0.0:
108        TSindex = i
```

```
#print TSindex


# Detect problematic region in old curve (maybe optional)

# Scale the new curve starting from the center and expand to two ends
#       Check 2 things:
#          1. The value
#          2. The gradient
# This step should be an iterative procedure, which move step by step
# from transition state point to the end(s)
#


### moved !
#def diffe(x,Snew,Vnew,Vold,dstep):
#    # calculate the difference between the scaled value and original value(calculated by old formula)
#    if abs(Vnew) > abs(Vold):
#       sign0 = -1 # scale down
#    else:
#       sign0 = 1  # scale up
#    #print x
#    k = math.exp(sign0*(abs(Snew)/dstep)/x)
#    err = abs(k*Vnew - Vold)
#
#    return err
###############


#def graddiff(x,ps,js,qs,pv,jv,jo,qv,dstep):
#    # This function returns the direction difference (with regard to gradient)
#    # ps -> s value for the point before that in question -> kp, pv
#    # js -> s value for the point in question -> kj, jv
#    #    jo -> value of the point in question on old curve
#    # qs -> s value for the point after that in question (old curve)
#
#
#    if abs(jv) > abs(jo):
#       sign0 = -1
#    else:
#       sign0 =  1
#    kp = math.exp(sign0*(abs(ps)/dstep)/x)
#    kj = math.exp(sign0*(abs(js)/dstep)/x)
#    g1 = (kj*jv - kp*pv) / (js - ps )
#    g2 = (qv    - jo   ) / (qs - js )
#
#    return g2/g1
#################


#print newfor

forx=[]
forsign=[]
forratio=[]
```

```
forkappa=[]
for i in range(len(newfor)-1):
    #print "i",i
    j = i + 1 # Current point
    p = i     # Point before current point
    sj = newfor[j][0]
    sp = newfor[p][0]
    vj = newfor[j][1]
    vp = newfor[p][1]
    #print sp, sj
    #print vp, vj
    #
    #
    # p  -> previous point
    # j  -> current point
    # q  -> past point
    #
    #
    oj = olddat[ TSindex + j ][1] # Current point value in old curve
    q = TSindex + j + 1 # The point past currect point in old curve
    sq = olddat[q][0]
    vq = olddat[q][1]
    #print sq
    Snew = sj
    Vnew = vj
    Vold = oj
    #
    #
    # "diffe" here is a user-defined function
    # x -> variable will be changed by optimizer
    # Snew, Vnew, Vold and dstep -> parameters that are fixed during
    optimization
    #
    #
    res = optimize.minimize_scalar(diffe,method="Bounded",bounds=(1e
    -2,1e10),args=(Snew,Vnew,Vold,dstep))
    unknownx =  res.x    # The "x" after optimization

    forx.append(unknownx)
    if abs(vj) > abs(oj):
       forsign.append(-1)
    else:
       forsign.append(1)

    # check derivative then
    #print unknownx
    #print sj, unknownx
    #print sp,sj,sq,'-',vp,vj,"-",oj,vq,"-",dstep
    #forratio.append(

    # " unknownx " is used here

    tmpresult =  graddiff(unknownx,sp,sj,sq,vp,vj,oj,vq,dstep)

    #print tmpresult
    forratio.append( tmpresult[0] )
    forkappa.append( tmpresult[1] ) # This is curvature value
```

```
219
220 #print forratio
221 #print forkappa
222
223 # Re-order newrev list
224 tmp=[]
225 for i in range(len(newrev)):
226     j = len(newrev) - 1 - i
227     tmp.append( newrev[j] )
228 newrev = tmp
229
230 #print ""
231
232 revratio=[]
233 revkappa=[]
234 revx=[]
235 revsign=[]
236 for i in range(len(newrev)-1):
237     j = i + 1
238     p = i
239     sj = newrev[j][0]
240     sp = newrev[p][0]
241     vj = newrev[j][1]
242     vp = newrev[p][1]
243
244     oj = olddat[ TSindex - j ][1]
245     q = TSindex - j - 1
246     sq = olddat[q][0]
247     vq = olddat[q][1]
248
249 #    print sq,sj,sp
250
251
252     Snew = sj
253     Vnew = vj
254     Vold = oj
255     res = optimize.minimize_scalar(diffe,method="Bounded",bounds=(1e
        -2,1e10),args=(Snew,Vnew,Vold,dstep))
256     unknownx = res.x
257
258     revx.append(unknownx)
259     if abs(vj) > abs(oj):
260        revsign.append(-1)
261     else:
262        revsign.append(1)
263
264 #   print unknownx
265     #revratio.append( graddiff(unknownx,sp,sj,sq,vp,vj,oj,vq,dstep) )
266     tmpresult = graddiff(unknownx,sp,sj,sq,vp,vj,oj,vq,dstep)
267     revratio.append( tmpresult[0] )
268     revkappa.append( tmpresult[1] )
269
270
271 #print revratio
272
273 # Checking gradient and then determine when to cut
274 ## input: forratio, revratio
```

```
275
276 #print forratio
277 #print revratio
278
279 #detercut(forratio)
280
281 #def detercut(ratiolist,thresh):
282 #    # determine where to cut for the scaling part
283 #    gi = -1
284 #    log = []
285 #    for i in range(len(ratiolist)):
286 #        if ratiolist[i] > 0:
287 #           if ratiolist[i] < 1:
288 #              ratiolist[i] = 1.0 / ratiolist[i]
289 #
290 #    for i in range(len(ratiolist)):
291 #        if ratiolist[i] >= 1.0:
292 #           log.append( ratiolist[i] )
293 #    if  min(log) > thresh:
294 #        print ("Smoothening with regard to gradient failed...")
295 #        print ("Gradient ratio "+str(min(log))+" exceeding threshold
    of "+str(thresh)  )
296 #        sys.exit()
297 #    else:
298 #        gi = ratiolist.index( min(log) )
299 #
300 #    return gi+1
301 #
302
303
304 # detercut is based on the gradient ratio which is not so robust
305  #fn = detercut(forratio,continuethresh)
306  #rn = detercut(revratio,continuethresh)
307
308
309
310 # So we need to use curvature instead, which is implemented as
    detercutB
311 fn = detercutB(forkappa,continuethresh)
312 rn = detercutB(revkappa,continuethresh)
313
314
315 print ("In forward direction, "+str(fn)+" points need correction.")
316 print ("In reverse direction, "+str(rn)+" points need correction.")
317
318
319 #
    ##########################################################################

320 #
    ##########################################################################

321 # Next step is to smooth back the central region.
322 # Another iterative procedure is expected.
323 #
324
325 # calculate the scaling factor and scaled values
```

```
# input: forx, revx, forsign, revsign

#print forx,revx,forsign,revsign


# Get the last point information
forx = forx[ fn-1 ]
forsign = forsign[ fn-1 ]
revx = revx[ rn-1 ]
revsign = revsign[ rn-1 ]

#print forx,forsign
#print revx,revsign

print ("The scaling factor for forward direction is: exp("+str(forsign)
    +"*abs(\"s\")/"+str(dstep)+"/"+str(round(forx,2))+")")
print ("The scaling factor for reverse direction is: exp("+str(revsign)
    +"*abs(\"s\")/"+str(dstep)+"/"+str(round(revx,2))+")")

forsv = []
forkv = []
forvv = []
forvscale = []

revsv = []
revkv = []
revvv = []
revvscale = []

for i in range( fn ):
    forsv.append( newfor[i+1][0] )
    forvv.append( newfor[i+1][1] )
    k = math.exp( forsign* abs(newfor[i+1][0])/dstep/forx )
    forkv.append( k )
    forvscale.append( k* newfor[i+1][1] )

for i in range( rn ):
    revsv.append( newrev[i+1][0] )
    revvv.append( newrev[i+1][1] )
    k = math.exp( revsign* abs(newrev[i+1][0])/dstep/revx )
    revkv.append( k )
    revvscale.append( k* newrev[i+1][1] )


#print revvscale

formark = -1
revmark = -1

for i in range(len(olddat)):
      if olddat[i][0] == forsv[-1]:
         formark = 0
         continue
      if formark == 0:
         forp1s = olddat[i][0]
         forp1v = olddat[i][1]
         formark = formark + 1
```

```
381             continue
382         if formark == 1:
383             forp2s = olddat[i][0]
384             forp2v = olddat[i][1]
385             formark = formark + 1
386             continue
387 
388 #print forp1s,forp2s
389 
390 for i in range(len(olddat)):
391 
392         j = len(olddat) - i - 1
393         if olddat[j][0] == revsv[-1]:
394             revmark = 0
395             continue
396         if revmark == 0:
397             revp1s = olddat[j][0]
398             revp1v = olddat[j][1]
399             revmark = revmark + 1
400             continue
401         if revmark == 1:
402             revp2s = olddat[j][0]
403             revp2v = olddat[j][1]
404             revmark = revmark + 1
405             continue
406 #print revp1s,revp2s
407 
408 forsf = []
409 forsf.append( [ revp2s, revp2v ] )
410 forsf.append( [ revp1s, revp1v ] )
411 
412 #print revsv
413 for i in range(len(revsv)):
414     j = len(revsv) - i - 1
415     forsf.append( [ revsv[j], revvscale[j] ] )  # bug!
416 
417 #print forsf
418 
419 
420 forsf.append([ 0.0, 0.0] )
421 
422 
423 for i in range(len(forsv)):
424     forsf.append( [ forsv[i], forvscale[i] ] ) # bug !
425 
426 forsf.append( [ forp1s, forp1v ] )
427 forsf.append( [ forp2s, forp2v ] )
428 
429 #print forsf # This is corrected data before spline fitting (4 extra
        original data included)
430 
431 
432 
433 marks = []
434 sfmark= []
435 for i in range(len(forsf)):
436     marks.append(1)
```

```
437     sfmark.append(0)
438
439 marks[0] =0
440 marks[1] =0
441 marks[2] =0
442 marks[-1]=0
443 marks[-2]=0
444 marks[-3]=0
445
446 #print sum(marks)
447
448
449 # Spline fitting procedure
450 # This is an iterative process
451 # input:
452 #  1. data for spline fitting -> forsf
453 #  2. label -> marks
454 #
455
456 #def spline1dlist(lis1):
457 #    x=[]
458 #    y=[]
459 #    for i in range(len(lis1)):
460 #        x.append( lis1[i][0] )
461 #        y.append( lis1[i][1] )
462 #    f = interp1d(x,y,kind='cubic')
463 #
464 #    return f
465
466 #def get1point(data,marks):
467 #   pts=[]
468 #   for i in range(len(marks)):
469 #       if marks[i] == 1:
470 #          pts.append(data[i])
471 #          break
472 #   return pts
473
474
475 #def get2points(data,marks):
476 #    # This function returns the two points to be predicted by spline
    fitting
477 #    pts=[]
478 #    for i in range(len(marks)):
479 #        if marks[i] == 1:
480 #           pts.append(data[i])
481 #           break
482 #    for i in range(len(marks)):
483 #        j = len(marks) - i - 1
484 #        if marks[j] == 1:
485 #           pts.append(data[j])
486 #           break
487 #
488 #    return pts
489
490
491 #def checkerr(f,pts):
492 #    # check the error between predicted value and train value
```

```
#    x=[]
#    y0=[]
#    y1=[]
#    err=[]
#    for i in range(len(pts)):
#        x.append(pts[i][0] )
#        y0.append(pts[i][1])
#    y1 = f(x)
#    #print y0
#    #print y1
#    for i in range(len(y0)):
#        err.append( abs(y0[i] - y1[i]) )
#    print err
#    return err

#def adjustmark(marks,l,r):
#    # this function could make 1 into 0
#    if l == 1:
#     for i in range(len(marks)):
#        if marks[i] == 1:
#            marks[i] = 0
#            break
#    if r == 1:
#     for i in range(len(marks)):
#        j = len(marks) - i - 1
#        if marks[j] == 1:
#            marks[j] = 0
#            break
#    return marks


#def updateval( err,thresh,data,pts,f ):
#
#    x=[]
#    replaceflag = []
#    for i in range(len(err)):
#        if err[i] > thresh:
#           replaceflag.append(1)
#        else:
#           replaceflag.append(0) # modified!!!!
#
#    if len(pts) == 1:
#       for i in range(len(data)):
#           if pts[0][0] == data[i][0]:
#              if replaceflag[0] == 1:
#                 #data[i][1] =
#                 x.append( pts[0][0] )
#                 m = f( x )
#                 data[i][1] = m[0]
#                 #print "w",f( pts[0][0] )
#                 print "1:got it!"
#
#    if len(pts) == 2:
#       for i in range(len(pts)):
#           for j in range(len(data)):
#               if pts[i][0] == data[j][0]:
#                  if replaceflag[i] == 1:
```

```
#                     #data[j][1] =
#                     x.append( pts[i][0] )
#                     m = f( x )
#                     #print "m",type(m) ,m[0],type(m[0])
#                     data[j][1] = m[0]    # f( pts[i][0] )
#                     print "2:got it!",pts[i][0]
#    return data


## SPLINE FITTING
#def itersplif(data,marks):
#    if len(data)!=len(marks):
#       print ("Length not match...")
#       sys.exit()
#
#    if sum(marks) >= 2:
#       run = 1
#
#    while run == 1:
#     if sum(marks) >= 2:
#       #print "not last"
#       train=[]
#       for i in range(len(marks)):
#           if marks[i] == 0:
#              train.append( data[i] )
#
#      f = spline1dlist(train)
#       pts = get2points(data,marks)
#       err = checkerr(f,pts) # Set threshold before here.
#
#       data = updateval( err,prederrthresh,data,pts,f )
#
#       # If error larger than threshold, take the spline fitting value
    ,
#       # Otherwise, stay the same.
#       # Remember to update the "data" when necessary
#
#       #print marks
#       marks = adjustmark(marks,1,1)
#       #print marks
#       #run = 0
#       if sum(marks) == 0:
#          run = 0
#
#     if sum(marks) == 1:
#       #print "last"
#       train=[]
#       for i in range(len(marks)):
#           if marks[i] == 0:
#              train.append( data[i] )
#       f = spline1dlist(train)
#       pts = get1point(data,marks)
#       err = checkerr(f,pts)
#       run = 0
#       # This last point, we just take the spline fitting value!
#       data = updateval( err,prederrthresh,data,pts,f )
#       marks = adjustmark(marks,1,0)
```

```
#
#    return data


#print forsf

#####################################################################
#forsf = itersplif(forsf,marks) # THIS IS NOT WORKING PROPERLY!!!
#####################################################################


# Another way to automatic spline fitting this region is to
# use a quadratic function try to fit N-3, N-5, N-7, ... points
# until the residue error is below a threshold ...

quadgood = 0
expand = 0

while quadgood == 0:
    for i in range(len(forsf)):
         if forsf[i][0] == 0.0:
             TSindex = i
    #
    expand = expand + 1
    quadgood = 1


#
# Most simple way to do the automatic spline fitting
# is do tell the program how many points need to be
# ruled out for spline fitting on the left side and
# and the right side.
# The user need to give this number !
#


#def semispline( forsf, RL ):
#    for i in range(len(forsf)):
#      if forsf[i][0] == 0.0:
#         TSi = i
#    right = RL[1]
#    left =  RL[0]
#    excp=[]
#    excp.append( TSi )
#    for i in range(right):
#        j = TSi + i + 1
#        excp.append(j)
#    for i in range(left):
#        j = TSi - i - 1
#        excp.append(j)
#    #print excp
#    train=[]
#    preds =[]
#    for k in range(len(forsf)):
```

```
#         if k in excp:
#            preds.append(forsf[k][0])
#         else:
#            train.append(forsf[k])
#     #print train
#     f = spline1dlist( train )
#
#     for i in range(len(forsf)):
#         if forsf[i][0] in preds:
#            x =[]
#            x.append(forsf[i][0])
#            y = f(x)
#            forsf[i][1] = y[0]
#
#     return forsf



forsf = semispline( forsf,RL)



#print forsf

for i in range(len(forsf)):
    print round(forsf[i][0],4),"        ",round(forsf[i][1],4)
```

Listing 1: AUTOSMTH Script

Besides the main script, it has one module named *smoothutils.py*.

```
# This is the library for
#

import sys
from scipy import optimize
import math
from scipy.interpolate import interp1d
import numpy as np



def diffe(x,Snew,Vnew,Vold,dstep): # Here x is the variable and others are all parameters
    # calculate the difference between the scaled value and original value(calculated by old formula)
    if abs(Vnew) > abs(Vold):
       sign0 = -1 # scale down
    else:
       sign0 = 1  # scale up
    #print x
    k = math.exp(sign0*(abs(Snew)/dstep)/x)
    err = abs(k*Vnew - Vold)

    return err
#


def graddiff(x,ps,js,qs,pv,jv,jo,qv,dstep):
    # This function returns the direction difference (with regard to gradient)
    # ps -> s value for the point before that in question -> kp, pv
    # js -> s value for the point in question -> kj, jv
    #    jo -> value of the point in question on old curve
    # qs -> s value for the point after that in question (old curve)
    #
    ### When pv == jv or jo == qv, exceptions happen!

    if abs(jv) > abs(jo):
       sign0 = -1
    else:
       sign0 =  1
    kp = math.exp(sign0*(abs(ps)/dstep)/x)
    kj = math.exp(sign0*(abs(js)/dstep)/x)
    g1 = (kj*jv - kp*pv) / (js - ps )
    g2 = (qv    - jo   ) / (qs - js )
    #print "derviation:",(kj*jv - jo)
    fx = jo     # -> current point - old
    fp = qv     # -> past point - old
    fn = kp*pv  # -> before point - scaled
    f2d = (fp - 2*fx + fn)/(dstep*dstep)
    #print "curvature",f2d
    if qv == jo:
       g2 = 1.0

    return (g2/g1,f2d)
    # Return also the curvature value as "f2d"
```

```
54
55 def detercutB(curvaturelist,thresh):
56     # Determine when to cut via looking at the curvature value
57     # The curvature characterize the change of gradients
58     #
59
60     gi  = -1
61     log = []
62     for i in range(len(curvaturelist)):
63         log.append(abs(curvaturelist[i]))
64
65     if min(log) > thresh:
66        print ("Smoothening with regard to the 2nd derivative check
       failed...")
67        print ("2nd derivative "+str(min(log))+" exceeding threshold of
       "+str(thresh)      )
68        sys.exit()
69     else:
70        gi = log.index(min(log))
71     return gi+1
72
73
74 def detercut(ratiolist,thresh):
75     # determine where to cut for the scaling part
76     gi = -1
77     log = []
78     for i in range(len(ratiolist)):
79         if ratiolist[i] > 0:
80            if ratiolist[i] < 1:
81               ratiolist[i] = 1.0 / ratiolist[i]
82     #print ratiolist
83     for i in range(len(ratiolist)):
84         if ratiolist[i] >= 1.0:
85            log.append( ratiolist[i] )
86     #print log
87     if  min(log) > thresh:
88         print ("Smoothening with regard to gradient failed...")
89         print ("Gradient ratio "+str(min(log))+" exceeding threshold of
        "+str(thresh)  )
90         sys.exit()
91     else:
92         gi = ratiolist.index( min(log) )
93
94     return gi+1
95
96
97
98
99 ##############################################
100 # Following is an example which shows clearly
101 # how to let Python to do optimization
102 ##############################################
103 # from scipy import optimize
104 #
105 # def f(x,a,b):
106 #    return a*(x-1-a-b)**2
107 #
```

```
# a=10
# b=100
# res = optimize.minimize_scalar(f,args=(a,b))
# print res.x
##############################################



def spline1dlist(lis1):
    x=[]
    y=[]
    for i in range(len(lis1)):
        x.append( lis1[i][0] )
        y.append( lis1[i][1] )
    f = interp1d(x,y,kind='cubic')

    return f

def get1point(data,marks):
   pts=[]
   for i in range(len(marks)):
       if marks[i] == 1:
          pts.append(data[i])
          break
   return pts


def get2points(data,marks):
    # This function returns the two points to be predicted by spline
    fitting
    pts=[]
    for i in range(len(marks)):
        if marks[i] == 1:
           pts.append(data[i])
           break
    for i in range(len(marks)):
        j = len(marks) - i - 1
        if marks[j] == 1:
           pts.append(data[j])
           break

    return pts

def checkerr(f,pts):
    # check the error between predicted value and train value
    x=[]
    y0=[]
    y1=[]
    err=[]
    for i in range(len(pts)):
        x.append(pts[i][0] )
        y0.append(pts[i][1])
    y1 = f(x)
    #print y0
    #print y1
    for i in range(len(y0)):
```

```
164         err.append( abs(y0[i] - y1[i]) )
165     print err
166     return err
167 
168 def adjustmark(marks,l,r):
169     # this function could make 1 into 0
170     if l == 1:
171      for i in range(len(marks)):
172         if marks[i] == 1:
173             marks[i] = 0
174             break
175     if r == 1:
176      for i in range(len(marks)):
177         j = len(marks) - i - 1
178         if marks[j] == 1:
179             marks[j] = 0
180             break
181     return marks
182 
183 
184 def updateval( err,thresh,data,pts,f ):
185 
186     x=[]
187     replaceflag = []
188     for i in range(len(err)):
189         if err[i] > thresh:
190             replaceflag.append(1)
191         else:
192             replaceflag.append(0) # modified!!!!
193 
194     if len(pts) == 1:
195        for i in range(len(data)):
196             if pts[0][0] == data[i][0]:
197                if replaceflag[0] == 1:
198                   #data[i][1] =
199                   x.append( pts[0][0] )
200                   m = f( x )
201                   data[i][1] = m[0]
202                   #print "w",f( pts[0][0] )
203                   print "1:got it!"
204 
205     if len(pts) == 2:
206        for i in range(len(pts)):
207             for j in range(len(data)):
208                 if pts[i][0] == data[j][0]:
209                    if replaceflag[i] == 1:
210                       #data[j][1] =
211                       x.append( pts[i][0] )
212                       m = f( x )
213                       #print "m",type(m) ,m[0],type(m[0])
214                       data[j][1] = m[0]    # f( pts[i][0] )
215                       print "2:got it!",pts[i][0]
216     return data
217 
218 
219 
220 
```

```
 # SPLINE FITTING
def itersplif(data,marks):
    if len(data)!=len(marks):
       print ("Length not match...")
       sys.exit()

    if sum(marks) >= 2:
       run = 1

    while run == 1:
     if sum(marks) >= 2:
       #print "not last"
       train=[]
       for i in range(len(marks)):
           if marks[i] == 0:
              train.append( data[i] )

       f = spline1dlist(train)
       pts = get2points(data,marks)
       err = checkerr(f,pts) # Set threshold before here.

       data = updateval( err,prederrthresh,data,pts,f )

       # If error larger than threshold, take the spline fitting value,
       # Otherwise, stay the same.
       # Remember to update the "data" when necessary

       #print marks
       marks = adjustmark(marks,1,1)
       #print marks
       #run = 0
       if sum(marks) == 0:
          run = 0

     if sum(marks) == 1:
       #print "last"
       train=[]
       for i in range(len(marks)):
           if marks[i] == 0:
              train.append( data[i] )
       f = spline1dlist(train)
       pts = get1point(data,marks)
       err = checkerr(f,pts)
       run = 0
       # This last point, we just take the spline fitting value!
       data = updateval( err,prederrthresh,data,pts,f )
       marks = adjustmark(marks,1,0)

    return data


def semispline( forsf, RL ):
    # Semi-automatic spline fitting

    for i in range(len(forsf)):
      if forsf[i][0] == 0.0:
```

```
278          TSi = i
279      right = RL[1]
280      left =  RL[0]
281
282      # Points that need to be predicted
283      excp=[]
284      excp.append( TSi )
285      for i in range(right):
286          j = TSi + i + 1
287          excp.append(j)
288      for i in range(left):
289          j = TSi - i - 1
290          excp.append(j)
291
292      #print excp
293      train=[]
294      preds =[]
295      for k in range(len(forsf)):
296          if k in excp:
297             preds.append(forsf[k][0])
298          else:
299             train.append(forsf[k])
300      #print train
301      f = spline1dlist( train )
302
303      for i in range(len(forsf)):
304          if forsf[i][0] in preds:
305             x =[]
306             x.append(forsf[i][0])
307             y = f(x)
308             forsf[i][1] = y[0]
309
310      return forsf
```

Listing 2: smoothutils.py Module of AUTOSMTH Script

Appendix K

Source Code of the pURVA program

If you are interested in special details about the pURVA codes, please contact Dr. Elfi Kraka, Director of the Computational and Theoretical Chemistry Group @ SMU.