THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

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

pubs.acs.org/JPCA

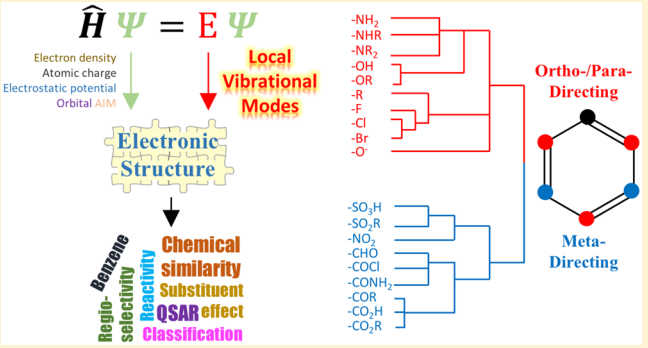
Characterizing Chemical Similarity with Vibrational Spectroscopy: New Insights into the Substituent Effects in Monosubstituted Benzenes

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

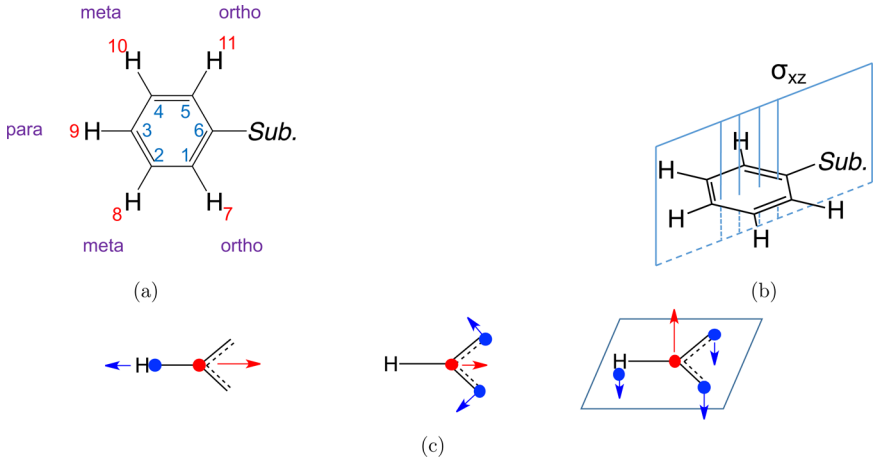


Figure 1. (a) Schematic representation of a monosubstituted benzene molecule. (b) σ_{xz} plane in a C_{2v} monosubstituted benzene. (c) Movement of atoms in the C−H bond stretching mode, C−C−C angle bending mode, and pyramidalization mode.

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, $\mathbf{K} = \mathbf{L}^{\dagger}\mathbf{f}^{x}\mathbf{L}$. $\mathbf{d}_n$ 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]	$\Delta\omega_m^{R\,b}$	$\Delta\omega_o^{R\,b}$	$\Delta\omega_p^{R\,b}$	$\Delta\omega_m^{\alpha\,b}$	$\Delta\omega_o^{\alpha\,b}$	$\Delta\omega_p^{\alpha\,b}$	$\Delta\omega_m^{\tau\,b}$	$\Delta\omega_o^{\tau\,b}$	$\Delta\omega_p^{\tau\,b}$	exp.[c]
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.

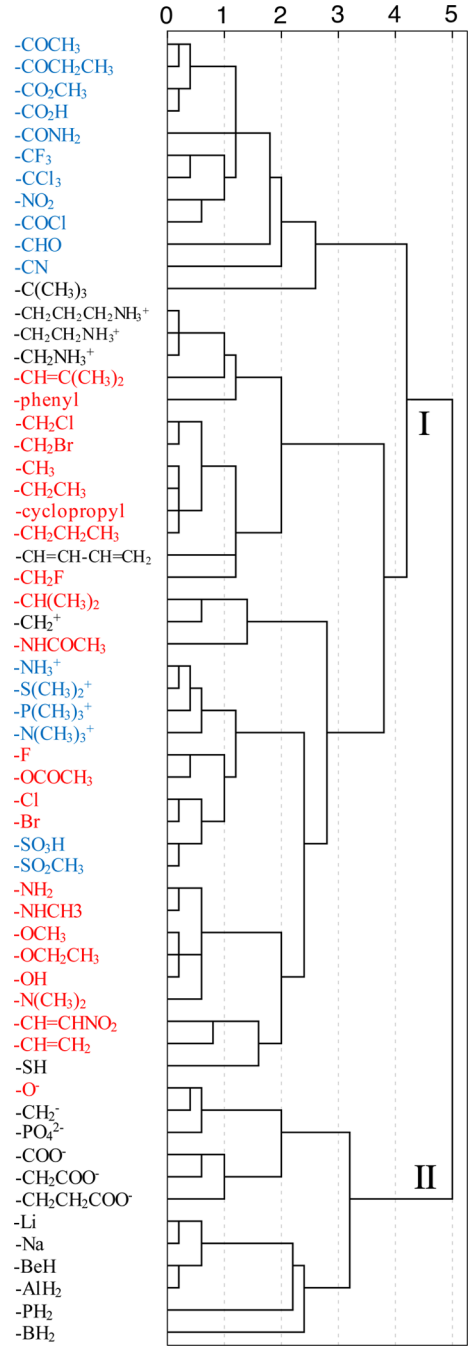


Figure 2. Diagram of hierarchical clustering of 59 monosubstituted benzenes based on 9 local mode frequency shifts.

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

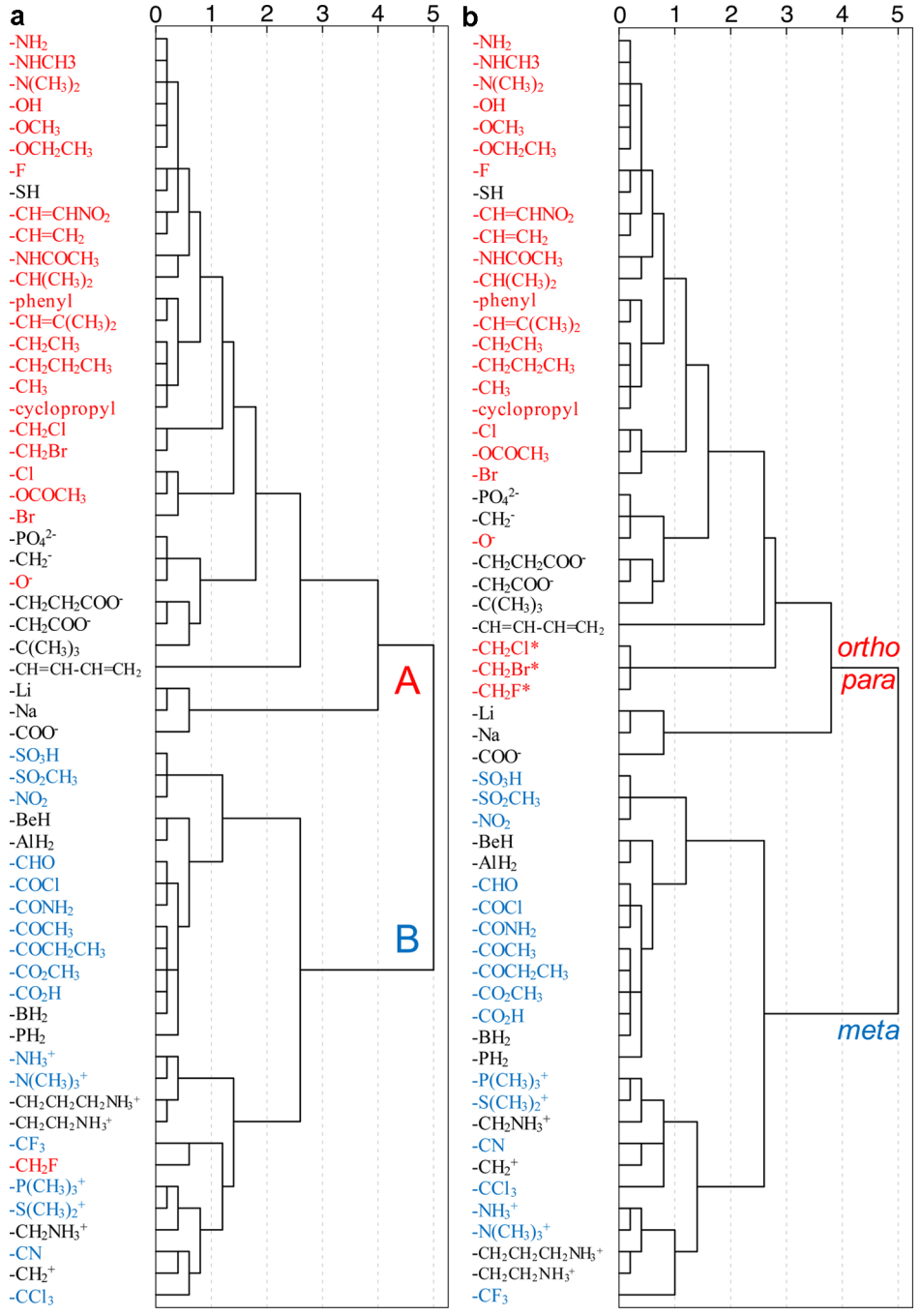


Figure 3. Diagrams of hierarchical clustering of 59 monosubstituted benzenes based on 6 local mode frequency shifts. (a) With CH_2F as the outlier. (b) Without any outliers.

(−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

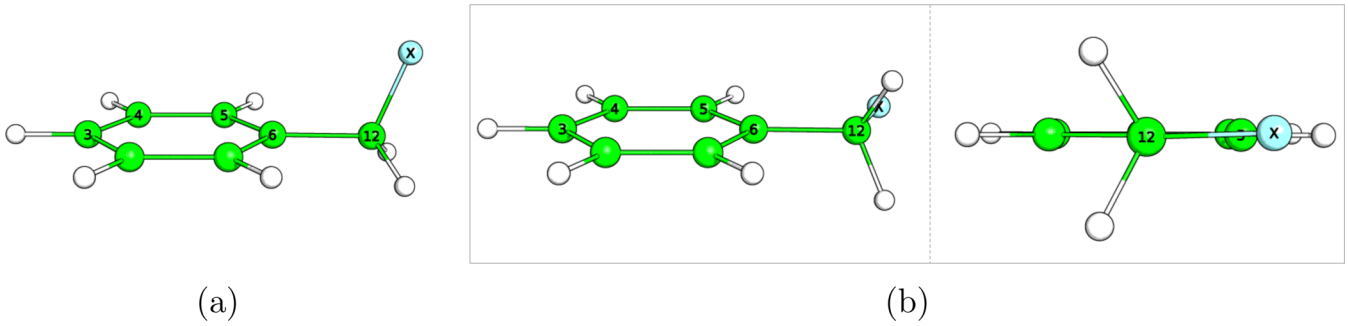


Figure 4. Ball-and-stick representation of a benzyl halide molecule. Green, white, and cyan spheres stand for C, H, and X (= F, Cl, and Br) atoms, respectively. (a) Local minimum geometry. (b) First-order saddle point geometry (X atom in the phenyl ring plane) in two perspectives.

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 ^1H, ^{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