11134 J. Phys. Chem. A **2007,** *111,* 11134−11140

Molecular Similarity Based on Atomic Electrostatic Potential

Nina Sadlej-Sosnowska

National Medicines Institute, 30/34 Chełmska Street, 00-725 Warsaw, Poland

Received: April 9, 2007; In Final Form: August 1, 2007

We propose a new similarity measure operating in the space spanned by the potential values, evaluated at atoms constituting the benzene ring and the COOH group in para-substituted benzoic acids and at benzene ring atoms in monosubstituted benzenes. The similarity measures are equivalent to the Euclidean distance between points in that space. Only the distances between the potentials at corresponding atoms in different molecules are included. The distances for benzene rings were very similar, regardless of whether they were calculated in para-substituted acids or in monosubstituted benzenes. As reference reactions, dissociation of benzoic acids and nitration of monosubstituted benzenes have been used. The effects of reduction of dimensionality of the potential space on the comparison of similarity measures with the free energies of the reference reactions have been investigated. It became obvious that the potentials at individual atoms in molecules of the acids and monosubstituted benzenes are mutually correlated to a high degree.

Introduction

The concept of similarity has been arousing much interest for a long time. The assumption that by using the data of the molecule's chemical or biological activity we can predict the activity of another closely related molecule by merely comparing how similar the original molecule is to the other one forms the basis of the molecular similarity postulate. Molecular similarity calculations can be especially useful in the study of processes with mechanisms that are complicated, or partly or even completely unknown.

Similarity measure is a rather intuitive property, and what follows is that its quantification is not straightforward. Yet the concept of similarity can be found in a such popular notion as the functional group. The molecules that are similar can differ, for example, in replacement of a functional group in a molecule by another group in another molecule. Many techniques have been proposed to measure the similarity of molecules, and a variety of similarity indexes have been proposed in the literature.[1,2] Some of them are based on the properties of molecular wave functions, and we shall mention only these, i.e., the so-called quantum-mechanical similarity measures. They employ ab initio or density functional theory calculations and are based on the information that can be drawn from the wave function. The beginnings of this approach date back to 1980 and the work of Carbó,[3] who assumed that similar molecules must have similar electron distributions. A resembling index, also based on electron density, was presented by Hodgkin and Richards.[4] The formula given in refs 1 and 2 and in the following ones[5,6] compare the shapes of charge distribution rather than magnitudes. A thorough overview of similarity indexes based on electron density is presented in ref 1 (a review). Determination of similarity based on electron density must meet the requirement of a proper alignment of molecules A and B, because it depends on mutual position of the molecules in space. The indexes were criticized on account of requiring a time-consuming three-dimensional integration and of the excessive importance of the conformation of the molecules.[7−9] To avoid these difficulties, the use of an autocorrelation function was proposed with a combination of an analysis of the principal components operating on electron density, as well as on electrostatic potential and local softness.[9] Other measures based on electron densities are referred to in refs 10−12. An interesting project worth mentioning here is the one drawing on a measure that employs electron densities within respective atomic basins[13] and another one, named NOEL, whose magnitude is related to the number of electrons in the molecular fragment common for both molecules.[7] An approach that emphasized the variation of the outer-valence electron density led to a successful prediction of the anti-HIV activity in a family of phospholipids.[14] Charge density was also useful in clustering side chains of amino acids into chemically related groups.[15]

Similarity measures based on electrostatic potential (MEP), as well as electric field, have also been proposed.[16−19] These methods are conceptually similar to those based on electron density. That is, for a comparison between two molecules, appropriate three-dimensional discrete grids for representation of the MEP within the regions surrounding the molecule are used and the potentials at the grid points are calculated. Next, they are compared by some means or other. Hodgkin employed an expression similar to that proposed by him for the similarity measure based on electron densities, but employing the MEP values instead.[4] A Carbó-type MEP similarity function of discrete grid MEP values was proposed.[16,17] A root-mean-square deviation between potentials originating from two molecules at common grid points was also used as a comparison function.[17] Many other methods have also been devised to compare three-dimensional displays of potential.[20] They share common features such as the fact that they are affected by the definition of the grid and they depend strongly on the method of superimposition of the molecules in the space. These methods also share similar weaknesses with those based on comparison of electron densities, namely a large number of calculation points (more than 50 000 for molecules counting less than 10 atoms)[19] and the need to optimize the similarity index. The number of grid points can be reduced, for example by selecting only points uniformly distributed on a van der Waals surface.[16] However, these methods have the virtue of enabling a comparison between any, even very dissimilar, molecules.

10.1021/jp072748a CCC: $37.00

This state of affairs made us wonder how to build a simpler measure based on potentials. We took into consideration the potentials at atomic nuclei. Their number does not exceed that of the atoms in a given molecule, and they are calculated for the individual molecules separately. This way one avoids optimization problems related to the mutual disposition of the molecules under comparison. In what follows, we have used a simple function, namely Euclidean distance in the potential space, as a similarity measure.

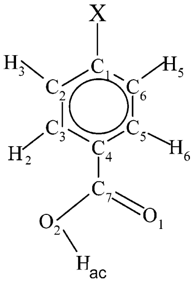
Molecular similarity measures are particularly useful in the search for new medicines. In this context, attempts are being made to replace the usual physicochemical parameters employed in QSAR analysis, such as, e.g., Hammett constants, with appropriate quantum chemical descriptors.[21]

The similarity measures are also interesting from the physicochemical point of view, as they attempt to answer not only the question as to how similar one molecule is to another but also the question of which molecular properties are the most sensitive to replacement of a group of atoms (a substituent) in a molecule for another group. Introduction of such groups to, for example, the benzene ring may be treated as a kind of perturbation. In the case when the ring is perturbed by a class of substituents, which molecular properties respond to the perturbation and to what degree? A very interesting answer to these questions has been given by Popelier and his collaborators. In a series of papers beginning in 1999,[8,22−26] he introduced a notion of quantum-molecular similarity in BCP (bond critical point) space. The similarity was expressed using such properties as electron density (ρ), its Laplacian ($\nabla^2\rho$), and the ellipticity (ϵ) at the BCP. It is calculated based on the distance between the molecules in the BCP space. The lower the distance is, the more similar the molecules are. The experimental substituent sequence (determined by a set of Hammett constants for the individual substituents) was only reproduced if the similarity measure (i.e., distance in the BCP space) was restricted solely to contributions from the COOH functional group. Thus, the only bonds that were taken into consideration were the bonds within the group: O−H, C═O, and C−O(H). What followed, not unexpectedly, in the case of the benzoic acid class, was that the reaction center proved to be restricted to the COOH group. Any inclusion of BCPs from the phenyl group seriously disrupted the sequence. On that basis it was stated that the method points out the reactive center for a given reaction. In the following papers the reactive centers were identified for carboxylic acids,[23,25,26] para-substituted phenols,[22,26] substituted toluenes and bromophenethylamines,[22] esters,[24] polychlorinated dibenzo-*p*-dioxins,[25] and anilines.[26]

Another similarity approach to rationalize substituent and solvent effects on the acidities of carboxylic acids was presented by Ponec et al.[27,28] They proposed that the dissociation constant of carboxylic acid may be described by the electronic density function of the COOH group and quantified by the self-similarity measure of this fragment. It was found that, over a series of five groups of acids, each containing seven to eight molecules, a regression with Hammett constants could be established with a correlation coefficient greater than 0.96. However, there is no mention of what results were obtained when electronic density functions of the whole acid molecule, instead of the COOH group, were taken into consideration.

The results made us ponder the issue of communication between a substituent and the COOH group in the molecules of substituted benzoic acids. Properties of the group in the BCP space vary from one molecule to another in concert with the electronic properties of their substituents. As the X substituent and the functional group (COOH) are attached to the same molecular skeleton (here benzene ring), their mutual interaction, according to our intuition, must be perceived by some means or other within the reach of the benzene ring. The property that has been explored as a tentative similarity measure is the electrostatic potential at the C ring atoms, as well as at the atoms of the COOH functional group. At the same time there are continuous and successful efforts to correlate the chemical and biological activities of molecules with maps of their electrostatic potentials[29−31] or the potential values at selected points within molecules.[32−41]

SCHEME 1: Atom Numbering in Para-Substituted Benzoic Acids[a]



[a] X = NO_2, CN, Cl, F, H, CH_3, OCH_3, NH_2, $N(CH_3)_2$, OH, $COCH_3$, CF_3, $C(CH_3)_3$, SO_2CH_3, and COOH.

This work aimed to test whether the distribution of potentials at the atoms of the common skeleton and the COOH group would be useful as a similarity measure of quite a different kind than that proposed and successfully applied by Popelier.[8] Thus we decided to start with the set of molecules originally used by him in order to have the possibility of a precise quantitative comparison of the results obtained with the similarity measures based on rather different concepts. Subsequently, the same calculations were performed using more compounds and the similarity measures were also calculated for a group of monosubstituted benzenes.

Calculations

All the calculations reported in the present study were carried out using density functional theory (DFT)[42] and the B3LYP functional[43,44] as implemented in the Gaussian 98 program package.[45] The initial three-dimensional structures of the compounds were built using the SpartanPro software,[46] followed by a preliminary semiempirical geometry optimization at the AM1 level. Subsequently, geometries were optimized and energies calculated using the Dunning's cc-pVDZ basis set, followed by the calculation of vibrational frequencies at the same level. When two conformers of a given substituted compound were possible, their energies differed by no more that 0.2 kcal/mol. All the reported properties were calculated for both conformers, and their average was reported and taken for the correlations displayed in the tables. With the use of the B3LYP/cc-pVDZ wave functions, the atomic charges derived from natural population analysis[47] were calculated at the same level. Potentials and a set of charges derived from them were calculated according to the CHelpG scheme.[48] Because there is a singularity in the nuclear contribution at the nucleus, the nuclear contributions to the atomic potentials are zeroed and only contributions from the electron density are calculated. AIM charges were calculated with the AIM 2000 package.[49]

Scheme 1 presents the numbering of the atoms in the para-substituted benzoic acids. The dissociating hydrogen of the COOH group is marked as H_{ac}. The numbering of the atoms in

TABLE 1: Matrix Containing Distances in the Space of Potentials at Five Carbon Atoms in the Benzene Ring (except C_{ipso}) for a Group of Eight Para-Substituted Benzoic Acids

	NH_2	OCH_3	CH_3	H	F	Cl	CN	NO_2
NH_2	0							
OCH_3	0.013	0						
CH_3	0.015	0.006	0					
H	0.026	0.013	0.011	0				
F	0.044	0.031	0.032	0.021	0			
Cl	0.052	0.039	0.039	0.028	0.008	0		
CN	0.082	0.069	0.069	0.058	0.038	0.030	0	
NO_2	0.085	0.072	0.072	0.061	0.041	0.033	0.005	0

TABLE 2: Matrix Containing Distances in the Space of Potentials at Atoms of the COOH Group in a Group of Eight Para-Substituted Benzoic Acids

	NH_2	OCH_3	CH_3	H	F	Cl	CN	NO_2
NH_2	0							
OCH_3	0.012	0						
CH_3	0.021	0.008	0					
H	0.028	0.015	0.007	0				
F	0.034	0.022	0.014	0.007	0			
Cl	0.041	0.028	0.020	0.013	0.009	0		
CN	0.062	0.050	0.042	0.035	0.030	0.022	0	
NO_2	0.067	0.055	0.046	0.039	0.035	0.026	0.005	0

monosubstituted benzenes is the same relative to the substituent. The set of Hammett constants was taken from March's monograph (Table 4 in Chapter 9).[50]

Results

Dissociation of Benzoic Acids. Previously, in another study, the similarity measures in para-substituted benzoic acids were calculated in BCP space[8] and were correlated with a set of σ constants from the Hammett equation.[50,51] The introduced similarity measure was calculated in two stages: at the first one the simple Euclidean distance $d_{i,j}$ between two BCPs i and j was defined as follows:

$$d_{i,j} = [(\rho_i - \rho_j)^2 + (\nabla^2\rho_i - \nabla^2\rho_j)^2 + (\epsilon_i - \epsilon_j)^2]^{1/2} \qquad (1)$$

where ρ, $\nabla^2\rho$, and ϵ stand for electron density, its Laplacian, and ellipticity, respectively, at the bond critical point. The $d_{i,j}$ terms were calculated for the corresponding bonds in the molecules A and B (ith bond in the first molecule, jth bond in the other). Next, the contributions from all bonds of the COOH group were added, producing the similarity measure $d_{A,B}$. The more similar two molecules A and B are, the smaller the Euclidean distance will be. Perfect similarity is characterized by $d_{A,B} = 0$.

It appeared that the distances in BCP space reproduced the experimental sequence of acidities, expressed through σ values. However, inclusion of bonds to the BCP space other than those between the atoms of the COOH group changed the ordering.

Searching for a similarity measure that could be successfully applied not only to the COOH functional group but also to the common skeleton of the acids, we accepted a similarity measure as the Euclidean distance in potential space. That is, the distance $d_{A,B}$ between two molecules, based on the ring carbon atoms, is defined as

$$d_{A,B} = [\sum_{i=1}^{n}(V_{Ci,A} - V_{Ci,B})^2]^{1/2} \qquad (2)$$

where $V_{Ci,A}$ and $V_{Ci,B}$ denote electrostatic potentials at the ith carbon atom in molecules A and B, respectively. The calculation of $d_{A,B}$ involves only one step because only one property (V) at each atom is taken into consideration, whereas in the case of similarity in the BCP space, three characteristics of each bond have been used. In the case when eq 2 is applied to C atoms in the benzene ring, summation to $n = 6$ seems to be natural. However, better results (see below) have been obtained where only five carbon atoms were taken into consideration. In that case n in eq 2 is equal to 5, and the atom excluded was C_{ipso} (C_1 in Scheme 1). It was recently found that electrostatic properties of that atom are not in concert with the properties of the rest of carbon atoms in the benzene ring.[52]

Equation 2 can easily be transformed in order to include a different (or larger) set of atoms belonging to the compared molecules. For example, the similarity measure (distance) based on potentials at atoms of the COOH group was also calculated:

$$d_{A,B} = [(V_{C7,A} - V_{C7,B})^2 + (V_{O1,A} - V_{O1,B})^2 + (V_{O2,A} - V_{O2,B})^2 + (V_{Hac,A} - V_{Hac,B})^2]^{1/2} \qquad (3)$$

as well as the that based on potentials of all atoms occurring in eqs 2 and 3:

$$d_{A,B} = \{[\sum_{i=1}^{n}(V_{Ci,A} - V_{Ci,B})^2] + (V_{C7,A} - V_{C7,B})^2 + (V_{O1,A} - V_{O1,B})^2 + (V_{O2,A} - V_{O2,B})^2 + (V_{Hac,A} - V_{Hac,B})^2\}^{1/2} \qquad (4)$$

where n equals 5 or 6 C atoms in the ring. It should be emphasized that the similarity measures defined in eqs 2−4 apply only to sets of congeneric molecules, such as substituted benzoic acids or monosubstituted benzenes studied here, where one can distinguish a molecules' common part and the corresponding atoms therein. In two entirely different molecules one could compare a selected property belonging to every atom (or bond) in A to every atom (or bond) in B.[8]

The results of application of eq 2 to the set of eight substituted acids (X = NO_2, CN, Cl, F, H, CH_3, OCH_3 and NH_2), the same as considered in ref 8, are displayed in Table 1, whereas those of the application of eq 3 are presented in Table 2. The tables display matrices of distances between the eight molecules. Distances in the BCP space (Table 3 in ref 8) are 3−4 times greater than those in Tables 1 and 2.

As a second step, we should like to compare our results to a set of Hammett constants. In order to have a one-dimensional

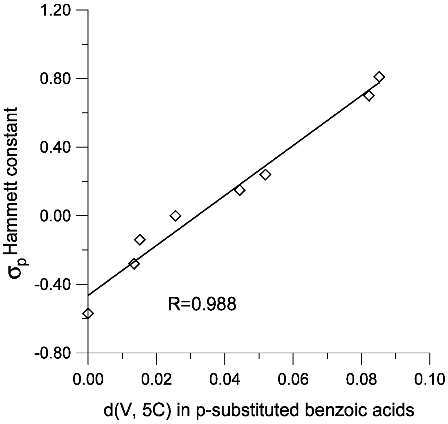


Figure 1. Regression analysis for the eight para-substituted benzoic acids. The Hammett constant is plotted against the proposed similarity distance calculated via eq 2 (potentials at five carbon atoms in the benzene ring were taken into consideration). The reference substituent is NH_2, which has the lowest σ_p value.

set of data, instead of the two-dimensional ones as in Tables 1 and 2, we took the similarity values between all molecules and a reference one, being the first of the sequence, similar to the procedure applied in ref 8. As the reference molecule, the one substituted by NH_2 was chosen. The lists of the distances related to NH_2 are equivalent to the second columns in Tables 1 and 2. These values were then correlated with σ_p constants. In other sources one may find many sets of the constants determined based on different reactions. In ref 8 six sets have been tried: the best correlation coefficient between a constants' set and the calculated one was found for the set quoted by March.[50] The same set has been used here.

Figure 1 displays the plot of $d(V,5C)$ against the set of σ_p constants, where $d(V,5C)$ represents the similarity in the potential space, calculated for five C ring atoms. The correlation between the two values' sets is good, and the sequence of σ_p constants is reproduced by the sequence of $d(V,5C)$. One can see that the results for the similarity values generated based on the potentials at atoms in the benzene ring work just as well as the similarity measures calculated in the BCP space confined to the three bonds of the COOH group.

We were then interested to see whether the above statement constitutes a property of the benzene ring solely in para-substituted benzoic acids. To answer the question, we have calculated the same similarity measures as displayed in Figure 1, but this time for monosubstituted benzenes. The results were virtually the same as in the case of the acids: a very similar plot was obtained, and its cc (correlation coefficient) equaled 0.984 (instead of 0.988 for acids).

Next, we tried the same correlation for benzoic acids but made with all six carbon atoms in the ring. In this case the order of $d(V,6C)$ did not perfectly match the order of σ_p constants: namely, substituents F and Cl switched places and the cc decreased to 0.979. Therefore, the match of the sequences of two series was perfect only when five carbon atoms were used, with the exclusion of C_{ipso}, which, as has been already mentioned, is little similar to the rest of the ring carbons.

Encouraged by the good correlations obtained for the five carbon atoms within the benzene ring, we tested similar correlations for the ring carbon atoms and the hydrogens attached to them in the acid molecules, namely H_2, H_3, H_5, and H_6 (modified eq 2 had nine terms in this case). A plot very similar to that in Figure 1 was obtained; the corresponding correlation coefficient was slightly lower, namely 0.983.

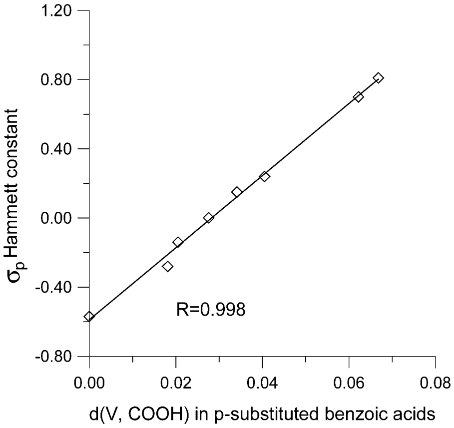


Figure 2. Regression analysis for the eight para-substituted benzoic acids. The Hammett constant is plotted against the proposed similarity distance calculated via eq 3 (potentials at atoms of the COOH group were taken into consideration). The reference substituent is NH_2, which has the lowest σ_p value.

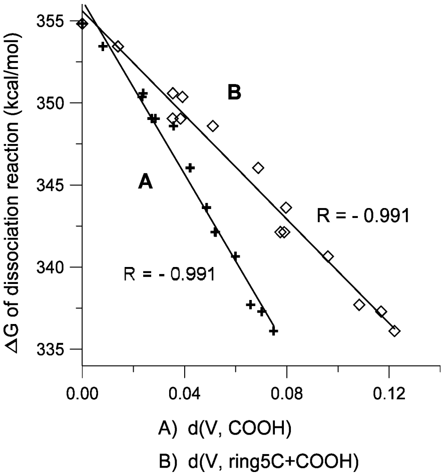


Figure 3. Regression analysis for 15 para-substituted benzoic acids. Free energy of dissociation reaction in the gas phase is regressed against the similarity distance calculated via eq 3 (+) and against the similarity distances calculated via eq 4 (◇). The reference substituent is $N(CH_3)_2$, which has the lowest σ_p value for this group of compounds.

Here a question emerges of whether the similarity measure defined based on potentials at atoms of the COOH group could also correlate with the set of σ. The answer may be found in Figure 2, which is basically similar to Figure 1 but its cc (0.998) is significantly higher; this result proves that the potential-based similarity is better correlated with the set of substituent constants when atoms of the COOH group are considered than in the case where the benzene ring carbons are taken into account. Very good correlation is also maintained in the case when the similarity is calculated for the set including both the ring carbon atoms and atoms of the COOH group (cc = 0.995). The cc of a plot similar to the one in Figure 2 but based on data in the BCP space was 0.993,[8] whereas the plot of self-similarity within the COOH group, based on the electron density, against Hammett constants, was characterized by cc = 0.969.[28]

All the results quoted so far were obtained with the set of eight compounds with the substituents listed in Tables 1 and 2. These are the same substituents as those used for the calculation of similarity in the BCP space.[8] The same selection of substituents justifies comparison of both similarity measures. To further explore the potentiality of our approach, we augmented the acids' set by seven more compounds, i.e., substituents $N(CH_3)_2$, OH, $COCH_3$, CF_3, $C(CH_3)_3$, SO_2CH_3, and

TABLE 3: Correlation Coefficients of Regressions between Potentials at Individual Six Ring Carbon Atoms and Atoms of the COOH Group for a Set of 15 Para-Substituted Benzoic Acids

	s	V_{C1}	V_{C2}	V_{C3}	V_{C4}	V_{C5}	V_{C6}	$V_{C(OOH)}$	V_{O9}	V_{O10}	V_{H}
σ	1										
V_{C1}	**0.057**[a]	**1**									
V_{C2}	0.968	**0.254**	1								
V_{C3}	0.949	**0.350**	0.991	1							
V_{C4}	0.996	**0.099**	0.979	0.964	1						
V_{C5}	0.960	**0.311**	0.995	0.996	0.973	1					
V_{C6}	0.944	**0.301**	0.986	0.989	0.963	0.983	1				
$V_{C(OOH)}$	0.994	**0.093**	0.976	0.961	0.999	0.970	0.960	1			
V_{O9}	0.993	**0.070**	0.970	0.954	0.999	0.964	0.955	0.999	1		
V_{O10}	0.994	**0.099**	0.975	0.961	0.999	0.970	0.959	1.000	0.999	1	
V_{H}	0.994	**0.104**	0.976	0.964	0.999	0.972	0.961	1.000	0.999	1.000	1

[a] Poor correlations between potential at C_1 and those at other atoms, standing out against the rest of very good correlations, are marked in bold.

COOH. These additional substituents were selected from among those for which the Hammett constants are compiled in ref 50, as for the eight substituents used so far. However, for OH the σ_p value determined by Hine,[53] namely −0.21, was used, instead of the value cited in March's monograph, −0.37. In the paper dealing with the similarity measures in BCP space,[8] it was found that the OH substituent is bracketed by OCH_3 and CH_3, and it was in agreement with that value of σ_p(OH). Now, NH_2 is no longer the first member of the sequence, as the lowest σ value falls to the level $N(CH_3)_2$ and the distances were calculated for other molecules in relation to the latter and correlated with σ_p. Potentials at atoms both in the COOH group and at the five ring carbons were taken into account, as in eq 4. Now, for 15 compounds, the cc of the regression of the distances against σ_p is lower than for the previously used set of eight compounds (0.986 vs 0.998). Therfore we conclude that the correlation of the similarity measures on σ_p constants not only depends on the selected set of σ_p's (which was mentioned in ref 8) but also may vary with the selection of the compounds.

Thus far the similarity measures calculated for the molecules in the gas phase have been correlated with the Hammett constants which measure the substituents' impact on dissociation of benzoic acids in water.[51] One might expect better correlation with the measure of dissociation propensity in the gas phase. Actually, the cc of the regression with ΔG_{diss} is slightly better than that with σ_p (−0.991 vs 0.986). Figure 3 displays regression of the ΔG_{diss} against the distances calculated according either to eq 3 or to eq 4. For this group of 15 compounds, the cc's for both plots are identical.

Our results indicate that the sums of the potential differences (eqs 2 and 3) at the corresponding atoms in different molecules change in an ordered fashion. Therefore, one might expect that the individual potentials themselves are correlated by some means or other. Table 3 displays cc's of the mutual correlations of the potentials at the individual atoms in the set of 15 molecules, the same as those used to generate data for Figure 3.

Inspection of Table 3 reveals that very good correlations occur not only between potentials of the ring carbon atoms (except C_1) but also between potentials of the carbons and of the atoms of the COOH unit, as well as between the potentials within the latter. Potentials of all atoms mentioned in the table (except C_1) vary synchronously from one molecule to another. It has been recently observed that, as far as electrostatic properties in the group of monosubstituted benzenes are concerned, the C_{ipso} atom differs from the rest of C atoms in the benzene ring.[52] The same effect is manifested by data in Table 3. The properties of the C_{ipso} atom are rather correlated with those of the substituent attached to it than with the properties of the rest of the molecule. Yet all the remaining atoms in Table 3 are somewhat similar to each other in the sense that their potentials harmonize.

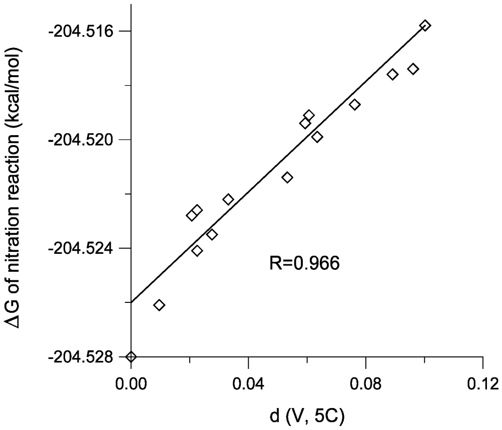


Figure 4. Regression analysis for 15 para-substituted benzenes. Free energy of nitration reaction is regressed against the similarity distance calculated via eq 2. The reference substituent is $N(CH_3)_2$, which has the lowest σ_p value for this group of compounds.

It is also worth mentioning that the potential at C_4 (in the para position in relation to the substituent) is highly correlated with the σ_p constant. Similar results were recently reported by Galabov and co-workers for monosubstituted benzenes.[36] Attaching the COOH group to this atom does not cause the correlation to deteriorate, but the correlation is even higher for the para-substituted benzoic acids than it is for monosubstituted benzenes.

Taking into consideration the fact that the presented approach of defining the similarity based on the electrostatic characteristics (potential at atoms in a molecule) generated satisfying results, we tried to calculate other similarity measures according to eq 2, but introducing into the equation the charges on atoms instead of the potentials. The results for the atoms of the benzene ring were poor, for the three sets of charges.

The best results were obtained only within the COOH group, by using the natural bond orbital charges. In this case the cc of regression of the charges on σ_p was 0.986. This result corresponds with the finding that sums of the charges on atoms of the dissociating carboxylic acid group correlated very well with the experimental values of pK_a.[54]

Nitration of Monosubstituted Benzenes. At this point the moment seemed opportune to check the predictive power of the similarity measures for a different set of compounds and/or different correlated property. To this aim we have selected a difference in free energy of a substituted benzene molecule and its *p*-nitro derivative. The same set of 15 substituents as for

TABLE 4: Correlation Coefficients of the Regressions of Hammett Constant and Free Energy of Dissociation in the Gas Phase with the Potentials at C_4 Atoms, and between the Individual Atoms of the COOH Group (C_7, O_1, O_2, H_{ac}), Calculated for 15 Benzoic Acids[a]

	$d(V,\text{ring5C+COOH})$	$d(V,\text{COOH})$	$d(V,\text{ring5C})$	$d(V,C_4)$	$d(V,C_7)$	$d(V,O_1)$	$d(V,O_2)$	$d(V,H_{ac})$
σ_p	0.986	0.993	0.975	0.996	0.994	0.971	0.994	0.994
ΔG_{diss}	−0.991	−0.991	−0.984	−0.994	−0.992	−0.970	−0.991	−0.992

[a] These coefficients are compared with those obtained when similarity (potential distance) was calculated for larger sets of atoms (columns 2−4).

TABLE 5: Correlation Coefficients of the Regressions of the Hammett Constant and Free Energy of Nitration with the Potential at C_4 Atoms, Calculated for 15 Monosubstituted Benzenes[a]

	$d(V,\text{ring5C})$	$d(V,C_4)$
σ_p	0.976	−0.995
ΔG_{nitro}	0.966	−0.986

[a] These coefficients are compared with those obtained when similarity (potential distance) was calculated for five C atoms in the ring (column 2).

benzoic acids was taken into consideration for the substituted benzene molecules. This time the similarity measures were based on five carbon atoms of the benzene ring. The results are presented in Figure 4. It can be seen that the correlation, although poorer than in the case of dissociation of benzoic acids, is still highly significant.

Reduction of the Potential Space. The fact that equally good or even better correlations were obtained for benzoic acids when a smaller number of atoms were taken into consideration (solely the atoms of the COOH group) gave us an idea to reduce the number of the atoms used for the distance calculations even further. The lowest possible number of atoms is 1. Putting $n = 1$ in eq 2 or leaving only one term in eq 3 results in the distance between a given molecule and a lead compound equal simply to the difference between the potentials at single corresponding atom in two molecules: $d_{A,B} = V_A - V_B$. Thus, for 15 benzoic acids, we have investigated correlations of σ_p and ΔG_{diss} with the potential difference between C_4 atoms, and between the individual atoms of the COOH group: C_7, O_1, O_2, and H_{ac}. In this case the correlations are the same when one takes only V_A instead of $V_A - V_B$, where, as it was previously read, V_A is the potential at a selected atom in every subsequent molecule (A) belonging to the set and V_B is the corresponding potential in the reference molecule (being the first member of the sequence). The results are displayed in Table 4. The correlations are compared with those calculated for larger sets of atoms (columns 2−4). The cc value was lowest for the distance calculated for five C atoms in the ring (column 4).

The results shown in Table 4 prove that distances based on potentials at individual atoms of the COOH group and of the C ring atoms to which the group is attached are at least as effective as the distances calculated for larger sets of nuclear potentials. This statement is indicated here for benzoic acids solely given that dissociation is being used as a reference reaction.

The same reduction of dimensionality of the potential space was next performed for the monosubstituted benzenes. In this case the potential difference between a single pair of atoms (C_{para} relative to substituent) was taken into consideration. The results are displayed in Table 5. This time the correlations with σ_p are distinctly better after reduction of the distance in the potential space to one atom only. This statement is indicated here for monosubstituted benzenes given that nitration is being used as a reference reaction. These results correspond with the finding of Galabov and co-workers that the potential at C_{para} is highly correlated with the σ_p constants, as well as with the results obtained by Hollingsworth et al. that charges on the dissociating hydrogen correlate well with the experimental pK_a's for benzoic acids.[54]

Conclusions

We presented a new quantum similarity measure in potential space, applied to para-substituted benzoic acids and to monosubstituted benzenes. The similarity is measured by a distance in the space spanned by potentials at different atoms of the compared molecules. The space has a number of dimensions equivalent to the number of corresponding atoms selected for the purpose of comparison. The results for *p*-benzoic acids indicate that the distance in the potential space, calculated based on atoms of the COOH group, reproduced the sequence of Hammett constants with a better correlation coefficient than the distance based on the atoms of carbon ring. When the atoms of the COOH group of the *p*-benzoic acid set were taken into consideration, the correlation of σ_p constants with the atomic potential based similarity distances was as good as that obtained by Popelier with the similarity measure defined in the BCP space. Contrary to the latter, our results were nearly as good when the set of the ring carbon atoms together with atoms of the COOH group was used for calculation of the similarity distances. The experiments involving reduction of the potential space displayed that distances based on potentials at individual atoms of the COOH group and of the C ring atom to which the group is attached are at least as good as the distances calculated for larger sets of nuclear potentials. Therefore, dimensionality of the potential space can be reduced to even as little as one atom which is relevant for a given reaction. This might be due to the fact that potentials at individual atoms in the investigated molecules are mutually correlated to a very high degree.

Similar results were obtained when the similarity measures expressed as distances in the potential space were correlated with free energy of the nitration reaction of monosubstituted benzenes. In this case, however, distance based on the potential at para carbon atom produced higher correlation coefficients of regressions with the Hammett constant and free energy of nitration than the distance based on five carbon atoms in the ring.

This corollary may be interesting also from the point of view of other similarity measures, e.g., those based on electron density. The similarity measure proposed here should also be tried for sets of molecules of different kind, e.g., heterocyclic ones. The work in this direction is in progress.

References and Notes

(1) Bultinck, P.; Girones, X.; Carbó-Dorca, R. Molecular Quantum Similarity: Theory and Applications. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Larter, R., Cundari, T., Eds.; John Wiley & Sons: Hoboken, NJ, 2005; Vol. 21, p 127.

(2) Nikolova, N.; Jaworska, J. *QSAR Comb. Sci.* **2003**, *22*, 1006.

(3) Carbó, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185.

(4) Hodgkin, E. E.; Richards, W. G. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 105.

(5) Solà, M.; Mestres, J.; Carbó, R.; Duran, M. *J. Am. Chem. Soc.* **1994**, *116*, 5909.

(6) Hodgkin, E. E.; Richards, W. G. *Chem. Br.* **1998**, *24*, 1141.

(7) Cioslowski, J.; Fleischmann, E. D. *J. Am. Chem. Soc.* **1991**, *113*, 64.

(8) Popelier, P. L. A. *J. Phys. Chem. A* **1999**, *103*, 2883.

(9) Boon, G.; Langenaeker, W.; De Proft, F.; De Winter, H.; Tollenaere, J. P.; Geerlings, P. *J. Phys. Chem. A* **2001**, *105*, 8805.

(10) Cooper, D. L.; Allan, N. L. *J. Am. Chem. Soc.* **1992**, *114*, 4773.

(11) Walker, P. D.; Arteca, G. A.; Mezey, P. G. *J. Comput. Chem.* **1991**, *12*, 230.

(12) Lee, C.; Smithline, S. *J. Phys. Chem.* **1994**, *98*, 1135.

(13) Cioslowski, J.; Nanayakkara, A. *J. Am. Chem. Soc.* **1993**, *115*, 11213.

(14) Measures, P. T.; Mort, K. A.; Cooper, D. L.; Allan, N. L. *J. Mol. Struct. (THEOCHEM)* **1998**, *423*, 113.

(15) Bohórquez, H. J.; Obregón, M.; Cárdenas, C.; Llanos, E.; Suárez, C; Villaveces, J. L.; Patarroyo, M. E. *J. Phys. Chem. A* **2003**, *107*, 10090.

(16) Petke, J. D. *J. Comput. Chem.* **1993**, *14*, 928.

(17) Richard, A. M. *J. Comput. Chem.* **1991**, *12*, 959.

(18) Burt, C.; Richards, W. G.; Huxley, P. *J. Comput. Chem.* **1990**, *111*, 1139.

(19) Rodriguez, J.; Manaut, F.; Sanz, F. *J. Comput. Chem.* **1993**, *14*, 922.

(20) *Molecular Similarity in Drug Design*; Dean, P. M., Ed.; Blackie Academic and Professional, Chapman & Hall: London, 1995.

(21) Amat, L.; Carbó-Dorca, R.; Ponec, R. *J. Med. Chem.* **1999**, *42*, 5169.

(22) Smith, P. J.; Popelier, P. L. A. *Org. Biomol. Chem.* **2005**, *3*, 3399.

(23) O'Brien, S. E.; Popelier, P. L. A. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 764.

(24) Chaudry, U. A.; Popelier, P. L. A. *J. Phys. Chem. A* **2003**, *107*, 4578.

(25) Popelier, P. L. A.; Chaudry, U. A.; Smith, P. J. *J. Chem. Soc., Perkin Trans. 2* **2002**, 1231.

(26) Chaudry, U. A.; Popelier, P. L. A. *J. Org. Chem.* **2004**, *69*, 233.

(27) Ponec, R.; Amat, L.; Carbó-Dorca, R. *J. Phys. Org. Chem.* **1999**, *12*, 447.

(28) Ponec, R.; Amat, L.; Carbó-Dorca, R. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 259.

(29) Barrows, S. E.; Cramer, C. J.; Truhlar, D. G.; Elovitz, M. S.; Weber, E. J. *Environ. Sci. Technol.* **1996**, *30*, 3028.

(30) Zwaans, R.; Thomson, C. *J. Mol. Struct. (THEOCHEM)* **1996**, *362*, 51.

(31) Lozano, J. J.; Pouplana, R.; Ruiz, J. *J. Mol. Struct. (THEOCHEM)* **1997**, *397*, 59.

(32) Gross, K. C.; Seybold, P. G.; Peralta-Inga, Z.; Murray, J. S.; Politzer, P. *J. Org. Chem.* **2001**, *66*, 6919.

(33) Robbins, A. M.; Jin, P.; Brinck, T.; Murray, J. S.; Politzer, P. *Int. J. Quantum Chem.* **2006**, *106*, 2904.

(34) Politzer, P.; Murray, J. S.; Peralta-Inga, Z. *Int. J. Quantum Chem.* **2001**, *85*, 676.

(35) Haeberlein, M.; Murray, J. S.; Brinck, T.; Politzer, P. *Can. J. Chem.* **1992**, *70*, 2209.

(36) Galabov, B.; Ilieva, S.; Schaefer, H. F., III. *J. Org. Chem.* **2006**, *71*, 6382.

(37) Galabov, B.; Cheshmedzhieva, D.; Ilieva, S.; Hadjieva, B. *J. Phys. Chem. A* **2004**, *108*, 11457.

(38) Bobadova-Parvanova, P.; Galabov, B. *J. Phys. Chem. A* **1998**, *102*, 1815.

(39) Galabov, B.; Bobadova-Parvanova, P. *J. Phys. Chem. A* **1999**, *103*, 6793.

(40) Galabov, B.; Bobadova-Parvanova, P.; Ilieva, S.; Dimitrova, V. *J. Mol. Struct. (THEOCHEM)* **2003**, *630*, 101.

(41) Dimitrova, V.; Ilieva, S.; Galabov, B. *J. Phys. Chem. A* **2002**, *106*, 11801.

(42) Cramer, C. J. *Essentials of Computational Chemistry*; John Wiley & Sons: Chichester, England, 2003.

(43) Becke, A. D. *Phys. Rev.* **1988**, *A38*, 3098.

(44) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *B37*, 785.

(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzales, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.7; Gaussian, Inc.: Pittsburgh, PA, 1998.

(46) *Spartan*, version 5.0; Wavefunction, Inc.: Irvine, CA.

(47) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899.

(48) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

(49) Biegler-König, F.; Schönbohm, J. *AIM 2000*, version 2.0.

(50) March, J. *Advanced Organic Chemistry*, 3rd ed.; Wiley: New York, 1985.

(51) Hammett, L. P. *Physical Organic Chemistry, Reaction Rates, Equilibria and Mechanisms*, 2nd ed.; McGraw-Hill: New York, 1970.

(52) Sadlej-Sosnowska, N. Accepted for publication in *Chem. Phys. Lett.*

(53) Hine, J. *J. Am. Chem. Soc.* **1960**, *82*, 4877.

(54) Hollingsworth, C. A.; Seybold, P. G.; Hadad, C. M. *Int. J. Quantum Chem.* **2002**, *90*, 1396.