

# Molecular Basis of LFER. Modeling of the Electronic Substituent Effect Using Fragment Quantum Self-Similarity Measures

Xavier Gironés and Ramon Carbó-Dorca

Institute of Computational Chemistry, University of Girona, Campus Montilivi,  
17071 Girona, Catalonia, Spain

Robert Ponec\*

Institute of Chemical Process Fundamentals, Czech Academy of Sciences, Prague 6,  
Suchbát 2, 165 02, Czech Republic

Received September 4, 2002

A new approach allowing the theoretical modeling of the electronic substituent effect is proposed. The approach is based on the use of fragment Quantum Self-Similarity Measures (MQS-SM) calculated from domain averaged Fermi Holes as new theoretical descriptors allowing for the replacement of Hammett sigma constants in QSAR models. To demonstrate the applicability of this new approach its formalism was applied to the description of the substituent effect on the dissociation of a broad series of *meta* and *para* substituted benzoic acids. The accuracy and the predicting power of this new approach was tested on the comparison with a recent exhaustive study by Sullivan et al. It has been shown that the accuracy and the predicting power of both procedures is comparable, but, in contrast to a five-parameter correlation equation necessary to describe the data in the study, our approach is more simple and, in fact, only a simple one-parameter correlation equation is required.

## INTRODUCTION

One of the most important facets of modern computer aided drug design is the relationship of (electronic) structure of the molecule to biological activity. While QAR is a powerful tool in this computational chemist's arsenal, there is a continual search for better descriptors describing the electronic and structural features of molecules.<sup>1,2</sup>

In addition to this basic task, another important part of the contemporary activities in this field concerns the rationalization of previously defined empirical descriptors used in (semi-)empirical rules and equations. An example in this respect can be, e.g., the Hammett equation,<sup>3,4</sup> which is well-known as the first, and a typical representative, of the broad family of linear free energy relationships (LFER). This empirical equation describes the equilibrium and/or rate constants, in a series of structurally related molecules, in terms of two parameters: the reaction constant  $\rho$  and the substituent constant  $\sigma$ . This relationship is mathematically expressed as

$$\log \frac{k_X}{k_H} = \rho \sigma_X \quad (1)$$

Although originally discovered in the field of chemical reactivity, this equation rapidly found numerous applications in many other areas, and the substituent constants  $\sigma$  are now widely used as important empirical descriptors also in other fields. An example in this respect can be, e.g., the use of  $\sigma$  constants as descriptors of the substituent effect in QSAR description of biological activities.<sup>5–9</sup> Because of their

importance, the values of  $\sigma$  constants have been determined for a large number of substituents and functional groups, and exhaustive recom compilations of these constants can be found in the literature.<sup>10</sup> Although such recom compilations of experimentally determined  $\sigma$  constants are certainly useful for the practical use in QSAR models, the rationalization of the factors that determine the manifestation of the substituent effect at the microscopic level, instead of the traditional experimental approach, constitutes itself a challenge in the theoretical chemistry field.<sup>11–23</sup> The need for such a microscopic understanding, characteristic for the contemporary chemistry, stimulated many theoretical studies attempting to explain the microscopic origin of substituent effects and to theoretically estimate  $\sigma$  constants. The center line of most of these studies is to recognize those appropriate theoretical quantities that can be successfully correlated with the experimental  $\sigma$  constants.<sup>11–23</sup> An important place, among those most used theoretical descriptors, belongs to quantities somehow related to electron densities<sup>12,13,17,21,22</sup> or, more generally, to the density functions DF.<sup>24</sup> An example in this respect can be found, e.g., in Molecular Quantum Self-Similarity Measures (MQS-SM), and their application as new general molecular descriptors represents now a novel, rapidly expanding, approach to the design of theoretical QSAR models.<sup>25–30</sup>

In addition to self-similarity measures based on the comparison of density functions of the whole molecules, another type of new theoretical descriptors was recently reported, in which the corresponding similarity measures were not derived from the density function of the whole structure but only from the corresponding electron densities of selected molecular fragments.<sup>29,30</sup> The philosophy underlying the introduction of these fragment measures arises from

\*Corresponding author phone: +420-2-20390271; fax: +420-2-20920661; e-mail: Rponec@icpf.cas.cz.

the empirical chemical experience, that the observed activity is often connected with the presence of certain functional groups, so that a more precise, and more realistic, description of the activity could hopefully be obtained if the corresponding theoretical descriptors were derived just from the appropriate particular fragments. The correctness of this intuitive expectation was indeed convincingly demonstrated in the studies,<sup>29,30</sup> in which it was found that the fragment self-similarity measure associated with the *active* carboxylic group COOH can successfully be used as the theoretical descriptor of the substituent effect on the dissociation of a series of aromatic carboxylic acids.

Our aim in this study is to pursue the systematic effort at the application of fragment QS-SM as theoretical descriptors of the substituent effect and to propose a new procedure of more realistic determination of the fragment densities, required for the calculation of similarity measures. This new approach is based on the concept of domain averaged Fermi holes,<sup>31–34</sup> and, in the following part, the application of the corresponding fragment self-similarity measures, as new theoretical descriptors of the substituent effect, will be reported. The approach will be applied to the study of the substituent effect on the dissociation of substituted benzoic acids, and, as it will be shown, the calculated values of self-similarity measures in a broad series of *meta* and *para* substituted benzoic acids do indeed correlate with the corresponding Hammett sigma constants. This result is very important since the existence of this correlation actually confirms the anticipated applicability of self-similarity measures, derived from the corresponding domain averaged Fermi holes, as descriptors of the substituent effect. In addition to this, another important consequence concerns the possibility of theoretical prediction or *estimation* of Hammett sigma constants. This problem was repeatedly addressed by various authors,<sup>11–23</sup> and one of the aims of this study is also to compare the quality, or statistical importance, of our predictions with the results of the recent study by Sullivan et al.,<sup>23</sup> which represents, so far, the most comprehensive and most systematic contribution to these efforts.

## METHODOLOGY

**Molecular Quantum Self-Similarity Measures (MQS-SM).** MQS-SM represent the particular case of the general formulation of Molecular Quantum Similarity Measures (MQSM). These measures quantitatively characterize the similarity between two molecular structures *A* and *B* on the basis of comparison of their corresponding density functions. The most general definition of the quantum similarity measure  $Z_{AB}(\Omega)$  between molecules *A* and *B* is given by eq 2.

$$Z_{AB}(\Omega) = \iint \rho_A(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \rho_B(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (2)$$

In this equation  $\rho_A(\mathbf{r}_1)$  and  $\rho_B(\mathbf{r}_2)$  are the DF of molecules *A* and *B*, respectively, and  $\Omega(\mathbf{r}_1, \mathbf{r}_2)$  is an arbitrary positively definite operator. According to the form of  $\Omega(\mathbf{r}_1, \mathbf{r}_2)$ , different MQSM may be obtained. In the simplest case where  $\Omega(\mathbf{r}_1, \mathbf{r}_2) = \delta(\mathbf{r}_1 - \mathbf{r}_2)$ , the definition (2) yields the so-called *Overlap* similarity measure. If both structures are considered to be the same, then eq 2 is reduced to

$$Z_{AA} = \int \rho_A(\mathbf{r}) \rho_A(\mathbf{r}) d\mathbf{r} \quad (3)$$

This quantity corresponds to the so-called Overlap Quantum Self-Similarity Measure of the (whole) molecule *A*.

**Fragment Quantum Self-Similarity Measures.** Another type of useful theoretical descriptors can be derived from the general formula (3) if instead of comparing the self-similarities of the whole molecules, we are more interested in comparing the self-similarity of particular functional groups or fragments. Thus, if the total density function of the whole molecule  $\rho_A(\mathbf{r})$  is replaced by the density function of the appropriate fragment  $\rho_A^X(\mathbf{r}_1)$ , the general formula (3) transforms to what is known as the so-called *fragment quantum self-similarity measure* (4):

$$Z_{AA}^X = \int \rho_A^X(\mathbf{r}) \rho_A^X(\mathbf{r}) d\mathbf{r} \quad (4)$$

The practical calculation of these similarity measures requires, of course, to specify how the corresponding fragment densities are to be determined in any particular case. One of the possibilities, that was in fact used in the preliminary studies,<sup>29,30</sup> is based on the straightforward partitioning of the total electron density  $\rho_A(\mathbf{r})$  according to eq 5

$$\rho_A^X(\mathbf{r}) = \sum_{\mu \in X} \sum_{\nu} P_{\mu\nu}^A \chi_{\mu}^*(\mathbf{r}) \chi_{\nu}(\mathbf{r}) \quad (5)$$

This partitioning is not, however, the only one possible, and other alternative procedures can also be proposed. Among them, the most promising seems to be the partitioning based on the recently proposed idea of the so-called domain averaged Fermi holes.<sup>31–34</sup>

**Fermi Holes and Fermi Hole Densities.** Since the introduction of these holes is sufficiently described in original studies, we confine ourselves only to the brief recapitulation of the basic ideas to the extent necessary for the purpose of this study.

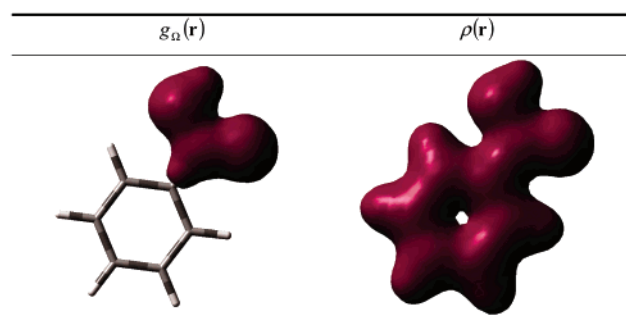
The Fermi hole associated with the domain  $\Omega$  in the molecule *A*, or the so-called domain averaged Fermi hole  $g_{\Omega}^A(\mathbf{r}_1)$  is defined by the formula (6)

$$g_{\Omega}^A(\mathbf{r}_1) = N_{\Omega} \rho_A(\mathbf{r}_1) - 2 \int_{\Omega} \rho_A(\mathbf{r}_1, \mathbf{r}_2) d\mathbf{r}_2 \quad (6)$$

where  $N_{\Omega}$  is the mean number of electrons in the domain  $\Omega$

$$N_{\Omega} = \int_{\Omega} \rho(\mathbf{r}) d\mathbf{r} \quad (7)$$

and  $\rho_A(\mathbf{r}_1, \mathbf{r}_2)$  is the pair density (of the molecule *A*). As it is possible to see from eq 6, the form of the function  $g_{\Omega}(\mathbf{r})$  depends in any given case on the particular form of the region  $\Omega$ , so that there is a straightforward 1:1 correspondence between the form of the function  $g_{\Omega}(\mathbf{r})$  and the corresponding region. This correspondence is generally expressed in that the function  $g_{\Omega}(\mathbf{r})$  is often surprisingly well *localized* in the same region over which the averaging was performed. Consequently, if, for example, the region  $\Omega$  is identified with some functional group in the molecule, the corresponding function  $g_{\Omega}(\mathbf{r})$  is also predominantly *localized* on the same functional group so that it is reminiscent of the fragment electron density associated with the corresponding group.



**Figure 1.** Comparison of a density contour for  $g_{\Omega}(\mathbf{r})$ ,  $\Omega = \text{COOH}$ , and  $\rho(\mathbf{r})$  at the same level (0.03 e/au<sup>3</sup>).

This intuitive association of the function  $g_{\Omega}(\mathbf{r})$  with the fragment electron density is straightforwardly demonstrated in Figure 1, which displays the iso-surface of the Fermi hole  $g_{\Omega}(\mathbf{r})$  for the molecule of benzoic acid, in which the region  $\Omega$  over which the averaging was performed coincides with the COOH group.

As it is possible to see, the function  $g_{\Omega}(\mathbf{r})$  is indeed localized in the region of the COOH group as expected. In addition to the above visual support, the anticipated identification of the function  $g_{\Omega}(\mathbf{r})$  with the fragment densities can be also formulated in the exact quantitative form. For this purpose let us imagine, for example, that a given molecule  $A$  is formally partitioned into an arbitrary number of disjoint regions  $\Omega_X$  in such a way that the union of all the regions fills the whole space. Then it is possible to show that the total electron density of this molecule ( $\rho_A(\mathbf{r})$ ) can be described by the following formula (8)

$$\sum_X g_{\Omega_X}^A(\mathbf{r}) = \rho_A(\mathbf{r}) = \sum_X \rho_A^X(\mathbf{r}) \quad (8)$$

from which it is evident that the function  $g_{\Omega}(\mathbf{r})$  does indeed represent the fragment electron density associated with a given functional group or the region  $\Omega$ .

Equation 8 thus in fact represents an alternative to the partitioning (5), but in contrast to this formula, which was to a certain extent arbitrary, the formula (8) represents a physically sound, *natural* partitioning of the electron density into fragment contributions. Based on the above interpretation of the functions  $g_{\Omega}(\mathbf{r})$ , the general definition of the fragment self-similarity measure can be rewritten in the form (9)

$$Z_{AA}^{\Omega} = \int g_{\Omega}^A(\mathbf{r}) g_{\Omega}^A(\mathbf{r}) d\mathbf{r} \quad (9)$$

and as it will be shown below, these self-similarity measures can successfully be used as new theoretical descriptors of the substituent effect allowing for the modeling of the Hammett sigma constants.

**Calculations.** To demonstrate the applicability of the fragment self-similarity measures as the descriptors of the substituent effect, several types of calculations had to be performed. In the first one, the complete geometry optimization of the studied molecules was required as a prerequisite for the generation of reliable electron densities that were subsequently used for the calculation of the corresponding fragment self-similarity measures. In our case, the studied series consisted of a set of roughly 300 *meta* and *para* substituted benzoic acids. The calculations were performed using the Gaussian 98 program<sup>35</sup> at the semiempirical AM1

**Table 1.** Benzoic Acid Derivatives, Self-Similarity, and Hammett Sigma Values and Calculation Residuals (Training Set)

group	$Z_{AA}^{\theta}$ <i>meta</i>	$\sigma_m$	residual	$Z_{AA}^{\theta}$ <i>para</i>	$\sigma_p$	residual
H	5.5349	0.00	-0.04	5.5349	0.00	-0.04
CH <sub>3</sub>	5.5356	-0.06	-0.07	5.5367	-0.14	-0.10
C <sub>2</sub> H <sub>5</sub>	5.5359	-0.08	-0.08	5.5370	-0.13	-0.08
<i>i</i> -C <sub>3</sub> H <sub>7</sub>	5.5358	-0.08	-0.08	5.5370	-0.13	-0.08
<i>i</i> -C <sub>4</sub> H <sub>9</sub>	5.5358	-0.09	-0.09	5.5371	-0.15	-0.10
C#CH	5.5253	0.20	-0.24	5.5324	0.23	0.09
C#CC <sub>6</sub> H <sub>5</sub>	5.5324	0.14	0.00	5.5328	0.16	0.04
<i>c</i> -C <sub>3</sub> H <sub>5</sub>	5.5354	-0.07	-0.09	5.5365	-0.21	-0.18
C <sub>6</sub> H <sub>5</sub>	5.5344	0.04	-0.02	5.5349	0.02	-0.02
CH <sub>2</sub> C <sub>6</sub> H <sub>5</sub>	5.5359	-0.05	-0.05	5.5358	-0.06	-0.06
CH <sub>2</sub> CN	5.5313	0.16	-0.03	5.5309	0.18	-0.02
CH <sub>2</sub> OR	5.5364	0.02	0.04	5.5361	0.02	0.03
CH <sub>2</sub> Cl	5.5327	0.11	-0.02	5.5324	0.12	-0.01
CF <sub>3</sub>	5.5236	0.46	-0.04	5.5219	0.53	-0.03
CCl <sub>3</sub>	5.5267	0.40	0.02	5.5250	0.46	0.03
CN	5.5258	0.62	0.20	5.5244	0.71	0.25
CHO	5.5276	0.41	0.07	5.5258	0.47	0.07
COCH <sub>3</sub>	5.5282	0.36	0.04	5.5268	0.47	0.11
COC <sub>6</sub> H <sub>5</sub>	5.5291	0.36	0.08	5.5282	0.46	0.16
CONH <sub>2</sub>	5.5265	0.28	-0.11	5.5266	0.31	-0.06
CO <sub>2</sub> H	5.5261	0.35	-0.05	5.5235	0.44	-0.05
CO <sub>2</sub> R	5.5271	0.35	-0.01	5.5248	0.44	0.00
B(OH) <sub>2</sub>	5.5368	-0.01	0.03	5.5339	0.12	0.04
Si(CH <sub>3</sub> ) <sub>3</sub>	5.5378	-0.04	0.04	5.5348	-0.07	-0.11
NH <sub>2</sub>	5.5360	-0.09	-0.08	5.5483	-0.57	-0.07
N(CH <sub>3</sub> ) <sub>2</sub>	5.5362	-0.10	-0.09	5.5481	-0.63	-0.14
NHCOCH <sub>3</sub>	5.5345	0.14	0.08	5.5397	-0.09	0.06
NHCOC <sub>6</sub> H <sub>5</sub>	5.5345	0.02	-0.03	5.5398	-0.19	-0.03
NHCOCF <sub>3</sub>	5.5273	0.30	-0.05	5.5314	0.12	-0.05
NCS	5.5268	0.48	0.11	5.5312	0.38	0.20
N=NC <sub>6</sub> H <sub>5</sub>	5.5314	0.29	0.11	5.5315	0.33	0.16
N <sub>3</sub>	5.5308	0.27	0.06	5.5352	0.15	0.13
NO <sub>2</sub>	5.5177	0.71	-0.04	5.5153	0.81	-0.01
P(C <sub>6</sub> H <sub>5</sub> ) <sub>2</sub>	5.5340	0.11	0.03	5.5345	0.19	0.14
PO(C <sub>6</sub> H <sub>5</sub> ) <sub>2</sub>	5.5250	0.38	-0.07	5.5204	0.53	-0.09
PO(OC <sub>2</sub> H <sub>5</sub> ) <sub>2</sub>	5.5262	0.43	0.03	5.5222	0.53	-0.01
OH	5.5313	0.13	-0.06	5.5385	-0.38	-0.27
OCH <sub>3</sub>	5.5306	0.10	-0.12	5.5396	-0.28	-0.13
OC <sub>6</sub> H <sub>5</sub>	5.5293	0.25	-0.02	5.5377	0.14	0.22
OCF <sub>3</sub>	5.5237	0.35	-0.15	5.5303	0.32	0.10
OCOCH <sub>3</sub>	5.5314	0.26	0.07	5.5291	0.16	-0.11
SH	5.5315	0.25	0.07	5.5334	0.15	0.06
SCH <sub>3</sub>	5.5310	0.14	-0.06	5.5351	0	-0.03
SCOCH <sub>3</sub>	5.5273	0.39	0.04	5.5315	0.44	0.27
SCN	5.5258	0.53	0.12	5.5270	0.52	0.17
SOCH <sub>3</sub>	5.5308	0.52	0.31	5.5254	0.49	0.07
SO <sub>2</sub> CH <sub>3</sub>	5.5221	0.68	0.11	5.5173	0.73	-0.01
SO <sub>2</sub> NH <sub>2</sub>	5.5233	0.53	0.01	5.5181	0.58	-0.13
SO <sub>2</sub> F	5.5166	0.79	0.00	5.5109	0.91	-0.09
SF <sub>5</sub>	5.5181	0.61	-0.12	5.5138	0.68	-0.20
F	5.5267	0.39	0.01	5.5318	0.06	-0.10
Cl	5.5286	0.37	0.07	5.5305	0.22	0.01
Br	5.5285	0.37	0.07	5.5284	0.22	-0.07
I	5.5291	0.34	0.06	5.5280	0.21	-0.10
IO <sub>2</sub>	5.5179	0.70	-0.04	5.5140	0.76	-0.11

level of theory. The electron densities generated in these primary calculations were then, in the second step, converted into the corresponding domain averaged Fermi holes associated with the COOH group that represents the *active center* for the dissociation process.<sup>36</sup> These calculations were performed using our own programs interfaced with Gaussian. The resulting Fermi holes were finally submitted to the MQS-SM routine<sup>37</sup> whose aim was to generate the fragment self-similarity measures used as the descriptors of the substituent effect in the correlations with Hammett sigma. The calculated values of self-similarity measures, along with Hammett sigma values, are summarized in Tables 1 and 2.



**Table 2.** Benzoic Acid Derivatives, Self-Similarity, and Hammett Sigma Values and Prediction Residuals (Test Set)

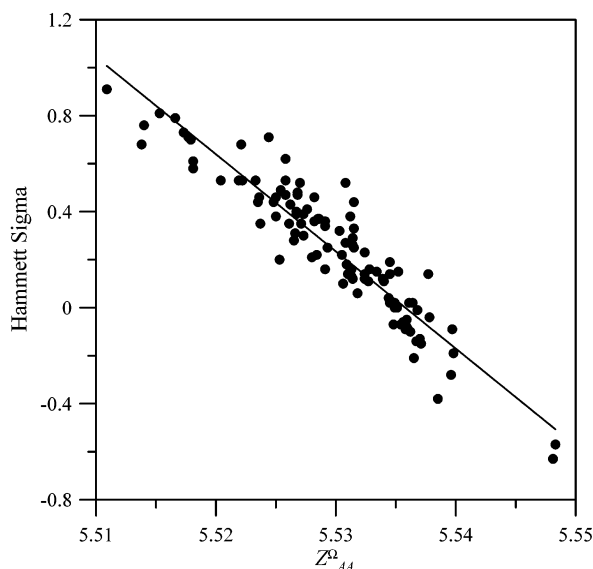
group	$Z_{AA}^O$ <i>meta</i>	$\sigma_m$	residual	$Z_{AA}^O$ <i>para</i>	$\sigma_p$	residual	group	$Z_{AA}^O$ <i>meta</i>	$\sigma_m$	residual	$Z_{AA}^O$ <i>para</i>	$\sigma_p$	residual
(CH <sub>2</sub> ) <sub>4</sub> CH <sub>3</sub>	5.5362	-0.08	-0.07	5.5371	-0.15	-0.10	SO <sub>2</sub> Cl	5.5137	1.20	0.30	5.5082	1.11	-0.01
(CH <sub>2</sub> ) <sub>6</sub> CH <sub>3</sub>	5.5360	-0.07	-0.06	5.5370	-0.16	-0.11	C(CH <sub>2</sub> CH <sub>3</sub> )(CH <sub>3</sub> ) <sub>2</sub>	5.5367	-0.06	-0.03	5.5370	-0.18	-0.13
C(CH <sub>3</sub> )(CN) <sub>2</sub>	5.5285	0.60	0.30	5.5262	0.57	0.18	C(CH <sub>3</sub> )(NO <sub>2</sub> ) <sub>2</sub>	5.5241	0.54	0.06	5.5178	0.61	-0.12
C(CH <sub>3</sub> ) <sub>3</sub>	5.5367	-0.10	-0.07	5.5369	-0.20	-0.16	C(Et)(NO <sub>2</sub> ) <sub>2</sub>	5.5237	0.56	0.07	5.5181	0.64	-0.08
C(CH <sub>3</sub> )=CH <sub>2</sub>	5.5354	0.09	0.07	5.5354	0.05	0.03	C(NO <sub>2</sub> )(CH <sub>3</sub> ) <sub>2</sub>	5.5254	0.18	-0.24	5.5259	0.20	-0.20
C <sub>2</sub> H <sub>5</sub>	5.5361	-0.07	-0.06	5.5370	-0.15	-0.10	C <sub>3</sub> H <sub>7</sub>	5.5361	-0.06	-0.05	5.5370	-0.13	-0.08
CBR <sub>3</sub>	5.5271	0.28	-0.07	5.5261	0.29	-0.10	CH <sub>2</sub> C(CH <sub>3</sub> ) <sub>3</sub>	5.5367	-0.05	-0.02	5.5368	-0.17	-0.13
CCCH <sub>3</sub>	5.5333	0.21	0.11	5.5338	0.21	0.13	CH <sub>2</sub> C(OH)(CH <sub>3</sub> ) <sub>2</sub>	5.5359	-0.16	-0.16	5.5379	-0.17	-0.09
CH(CH <sub>3</sub> )CH <sub>2</sub> CH <sub>3</sub>	5.5363	-0.08	-0.06	5.5369	-0.12	-0.08	CH <sub>2</sub> CF <sub>3</sub>	5.5299	0.12	-0.12	5.5286	0.09	-0.20
CH(OH)CH <sub>3</sub>	5.5344	0.08	0.02	5.5359	-0.07	-0.07	CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	5.5358	-0.07	-0.07	5.5367	-0.12	-0.09
CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub>	5.5357	-0.07	-0.08	5.5368	-0.12	-0.08	CH <sub>2</sub> CH=CH <sub>2</sub>	5.5354	-0.11	-0.13	5.5363	-0.14	-0.12
CH <sub>2</sub> CH(OH)(CH <sub>3</sub> )	5.5356	-0.12	-0.13	5.5377	-0.17	-0.09	CH <sub>2</sub> CONH <sub>2</sub>	5.5319	0.06	-0.10	5.5302	0.07	-0.16
CH <sub>2</sub> F	5.5309	0.12	-0.08	5.5324	0.11	-0.03	CH <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	5.5349	0.00	-0.04	5.5346	0.01	-0.04
CH <sub>2</sub> I	5.5305	0.10	-0.12	5.5324	0.11	-0.03	CH <sub>2</sub> NHCOCH <sub>3</sub>	5.5322	0.05	-0.10	5.5338	-0.05	-0.13
CH <sub>3</sub>	5.5356	-0.07	-0.08	5.5366	-0.17	-0.14	CH <sub>2</sub> OCOCH <sub>3</sub>	5.5342	0.04	-0.03	5.5342	0.05	-0.02
CH=C(CN) <sub>2</sub>	5.5233	0.66	0.15	5.5223	0.84	0.29	CH=C(CN) <sub>2</sub>	5.5232	0.66	0.15	5.5224	0.84	0.30
CH=NOCH <sub>3</sub>	5.5323	0.37	0.23	5.5324	0.30	0.16	CH=NNHCONHNH <sub>2</sub>	5.5293	0.22	-0.04	5.5318	0.16	0.00
CHCl <sub>2</sub>	5.5275	0.31	-0.03	5.5277	0.32	-0.01	CH=NNHCSNH <sub>2</sub>	5.5284	0.45	0.15	5.5322	0.40	0.25
CHI <sub>3</sub>	5.5281	0.26	-0.05	5.5293	0.26	0.00	CH=NOCH <sub>3</sub>	5.5323	0.37	0.23	5.5323	0.30	0.16
COC(CH <sub>3</sub> ) <sub>3</sub>	5.5290	0.27	-0.01	5.5276	0.32	-0.01	CHBr <sub>2</sub>	5.5297	0.31	0.06	5.5283	0.32	0.01
COCF <sub>3</sub>	5.5221	0.63	0.07	5.5182	0.80	0.09	CH(CN) <sub>2</sub>	5.5270	0.53	0.17	5.5248	0.52	0.07
COCH <sub>2</sub> CH <sub>3</sub>	5.5282	0.38	0.07	5.5271	0.48	0.13	COCH(CH <sub>3</sub> ) <sub>2</sub>	5.5290	0.38	0.10	5.5283	0.48	0.17
COCl	5.5237	0.51	0.02	5.5208	0.61	0.00	COCH <sub>2</sub> CH <sub>3</sub>	5.5282	0.38	0.07	5.5272	0.48	0.13
COOCH <sub>3</sub>	5.5271	0.37	0.02	5.5247	0.45	0.00	COF	5.5228	0.55	0.02	5.5203	0.70	0.07
CSNHCH <sub>3</sub>	5.5260	0.30	-0.10	5.5278	0.34	0.01	CONHCH <sub>3</sub>	5.5269	0.35	-0.01	5.5274	0.36	0.02
N(C <sub>3</sub> H <sub>7</sub> ) <sub>2</sub>	5.5357	-0.26	-0.27	5.5468	-0.93	-0.49	COOCH <sub>2</sub> CH <sub>3</sub>	5.5275	0.37	0.03	5.5253	0.45	0.02
N(CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>	5.5357	-0.23	-0.24	5.5482	-0.72	-0.22	CSNH <sub>2</sub>	5.5255	0.25	-0.17	5.5269	0.30	-0.06
N(COCH <sub>3</sub> ) <sub>2</sub>	5.5250	0.35	-0.09	5.5263	0.33	-0.06	N(CH <sub>3</sub> )NO <sub>2</sub>	5.5231	0.49	-0.03	5.5189	0.61	-0.08
N=NCN	5.5237	0.71	0.22	5.5210	1.03	0.43	N=C=O	5.5276	0.27	-0.06	5.5317	0.19	0.02
NHCH <sub>3</sub>	5.5355	-0.21	-0.22	5.5484	-0.70	-0.19	N=CCl <sub>2</sub>	5.5284	0.21	-0.09	5.5297	0.13	-0.12
NHCOCH(CH <sub>3</sub> ) <sub>2</sub>	5.5290	0.11	-0.17	5.5397	-0.10	0.06	NHCH <sub>2</sub> CH <sub>3</sub>	5.5368	-0.24	-0.20	5.5490	-0.61	-0.08
O(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	5.5335	0.10	0.01	5.5405	-0.32	-0.13	NHCOCH <sub>2</sub> Cl	5.5267	0.17	-0.20	5.5373	-0.03	0.03
O(CH <sub>2</sub> ) <sub>4</sub> CH <sub>2</sub>	5.5334	0.10	0.00	5.5404	-0.34	-0.16	NHCSCH <sub>3</sub>	5.5255	0.24	-0.18	5.5358	0.12	0.12
OCH(CH <sub>3</sub> ) <sub>2</sub>	5.5343	0.10	0.04	5.5417	-0.45	-0.21	NHCSNH <sub>2</sub>	5.5267	0.22	-0.15	5.5366	0.16	0.19
OCH <sub>2</sub> CH=CH <sub>2</sub>	5.5329	0.09	-0.03	5.5395	-0.25	-0.10	NHCSNH <sub>2</sub>	5.5278	0.30	-0.03	5.5380	0.07	0.16
OCH <sub>2</sub> Cl	5.5301	0.25	0.02	5.5359	0.08	0.08	OCH <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	5.5335	0.10	0.01	5.5404	-0.25	-0.07
P(CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>	5.5354	0.10	0.08	5.5327	0.13	0.00	OCH <sub>2</sub> CH <sub>3</sub>	5.5335	0.10	0.01	5.5403	-0.24	-0.06
P(isoC <sub>3</sub> H <sub>7</sub> ) <sub>2</sub>	5.5356	0.02	0.01	5.5347	0.06	0.01	OCN	5.5215	0.67	0.09	5.5278	0.54	0.21
P(O)(C <sub>4</sub> H <sub>9</sub> ) <sub>2</sub>	5.5331	0.35	0.24	5.5328	0.49	0.37	OSO <sub>2</sub> CH <sub>3</sub>	5.5263	0.39	0.00	5.5371	0.36	0.41
P(O)(isoC <sub>3</sub> H <sub>7</sub> ) <sub>2</sub>	5.5332	0.37	0.26	5.5280	0.41	0.09	P(CH <sub>3</sub> ) <sub>2</sub>	5.5350	0.03	0.00	5.5338	0.06	-0.02
P(O)Cl <sub>2</sub>	5.5186	0.78	0.08	5.5134	0.90	-0.01	P(O)(CH <sub>3</sub> ) <sub>2</sub>	5.5255	0.43	0.01	5.5263	0.49	0.10
P(O)F <sub>2</sub>	5.5204	0.81	0.18	5.5144	0.89	0.02	P(O)(N(CH <sub>3</sub> ) <sub>2</sub> ) <sub>2</sub>	5.5270	0.30	-0.06	5.5268	0.40	0.03
P(O)(C(CH <sub>3</sub> ) <sub>3</sub> ) <sub>2</sub>	5.5351	0.31	0.28	5.5286	0.41	0.12	P(O)(OC <sub>4</sub> H <sub>9</sub> ) <sub>2</sub>	5.5249	0.41	-0.03	5.5219	0.57	0.01
P(S)(CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>	5.5298	0.39	0.15	5.5228	0.46	-0.07	P(OH) <sub>2</sub>	5.5344	0.36	0.30	5.5294	0.42	0.16
PH <sub>2</sub>	5.5323	0.06	-0.08	5.5318	0.05	-0.11	PCl <sub>2</sub>	5.5257	0.54	0.13	5.5226	0.61	0.07
P(O)(CH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>	5.5329	0.37	0.25	5.5270	0.47	0.11	PF <sub>2</sub>	5.5292	0.49	0.22	5.5262	0.59	0.20
P(N(CH <sub>3</sub> ) <sub>2</sub> ) <sub>2</sub>	5.5367	0.18	0.21	5.5343	0.25	0.19	PO(OCH <sub>2</sub> CH <sub>3</sub> ) <sub>2</sub>	5.5250	0.55	0.11	5.5219	0.60	0.04
SCH(CH <sub>3</sub> ) <sub>2</sub>	5.5348	0.23	0.19	5.5363	0.07	0.09	SCCH	5.5302	0.26	0.03	5.5328	0.19	0.07
SCH <sub>2</sub> CH=CH <sub>2</sub>	5.5333	0.19	0.09	5.5349	0.12	0.08	SCH <sub>2</sub> CH <sub>3</sub>	5.5314	0.18	0.00	5.5355	0.12	0.11
SCH <sub>2</sub> F	5.5306	0.23	0.02	5.5318	0.20	0.04	SCH=CHCl	5.5287	0.31	0.02	5.5315	0.20	0.02
SCH=CH <sub>2</sub>	5.5320	0.26	0.10	5.5336	0.24	0.15	SF <sub>3</sub>	5.5304	0.70	0.48	5.5250	0.80	0.36
SCHF <sub>2</sub>	5.5270	0.33	-0.03	5.5307	0.37	0.16	SO <sub>2</sub> CH <sub>3</sub>	5.5223	0.60	0.05	5.5173	0.72	-0.03
SCl	5.5267	0.44	0.07	5.5292	0.48	0.21	SO <sub>2</sub> CN	5.5160	1.10	0.30	5.5109	1.26	0.25
SN(CH <sub>3</sub> ) <sub>2</sub>	5.5349	0.12	0.08	5.5364	0.09	0.11	SO <sub>2</sub> N(CH <sub>3</sub> ) <sub>2</sub>	5.5244	0.51	0.05	5.5189	0.65	-0.04
SO <sub>2</sub> CH <sub>2</sub> CH <sub>3</sub>	5.5227	0.66	0.13	5.5179	0.77	0.04	SOCHF <sub>2</sub>	5.5221	0.54	-0.02	5.5189	0.58	-0.11

## RESULTS AND DISCUSSION

As already stressed above, the main goal of this study is to demonstrate the applicability of fragment quantum self-similarity measures derived from the so-called domain averaged Fermi holes as new efficient theoretical descriptors of the substituent effect. For this purpose, the above methodology was applied to the description of the dissociation of a series of *meta* and *para* substituted benzoic acids, which is empirically characterized by the value of the corresponding Hammett sigma constants. The origin of the present study can be traced in the empirical findings,<sup>29,30</sup> in which the existence of a correlation between the values of fragment self-similarity measures, associated with the *active* COOH group in a series of substituted aromatic carboxylic

acids, and the corresponding Hammett sigma constants was reported. Unfortunately, the range of substituents for which the existence of this correlation was reported was relatively very narrow (just eight substituents) to guarantee its sufficient generality. To provide the missing support for the generality of these previously reported empirical correlations, we report here an extensive study in which the possibility of using the fragment self-similarity measures as new theoretical descriptors of the substituent effect was tested on a much broader set of substituents.

In the first test, presented in Table 1, the above methodology is applied to the set of 109 *meta* and *para* substituted benzoic acids, for which the experimental values of sigma constants were available in the book by Exner.<sup>38</sup> Consistent



**Figure 2.** Self-similarity vs Hammett sigma for a series of 109 benzoic acid derivatives.

with the empirical findings of the previous study,<sup>29,30</sup> these sigma constants were correlated with the corresponding fragment self-similarity measures associated, for each individual acid, with the *active* COOH group. The resulting correlation equation is given by (10)

$$\sigma \equiv -40.516Z_{AA}^{\Omega} + 224.29 \quad (10)$$

$$\left[ \begin{array}{lll} n = 109 & r^2 = 0.874 & s = 0.097 \\ F_{\text{stat}} = 741.608 & r_{cv}^2 = 0.872 & q^2 = 0.869 \end{array} \right]$$

and its graphical form is displayed in Figure 2. As it is possible to see, the correlation is indeed reasonable enough. This result is very important since not only it provides the support for previous empirical findings but also the statistical parameters of this correlation are indeed satisfactory to justify the use of self-similarity measures as theoretical descriptors of the substituent effect, ready to replace the Hammett sigma constants in QSAR models.<sup>29,30</sup> In addition to this primary result, the existence of sufficiently tight correlation between the similarity measures and sigma constants also opens the possibility to use the correlation eq 10 for the theoretical prediction or estimation of these constants for other substituents. For this purpose, let us consider the above set of 109 substituents as the training set and let us attempt to use the correlation eq 10 for the prediction of sigma values for the test set of 220 other substituents, summarized in the Table 2. These substituents were selected, after removing those already present in the previous set, from the extensive study by Sullivan et al.,<sup>23</sup> who recently reported an alternative approach to the theoretical calculation of sigma constants on the basis of the correlation with various quantum chemical descriptors.

The results of the application of the correlation eq 10 to the test set of 220 substituted acids are summarized below:

$$[n = 220 \quad r^2 = 0.837 \quad s = 0.141]$$

As it is possible to see, the statistical parameters characterizing the accuracy and predicting power of the correlation eq 10 applied to the test set of 220 additional substituents

are indeed very satisfactory. To demonstrate the usefulness and the applicability of our approach to the description of the substituent effect, it is also interesting to compare our results with the results of the above-mentioned study of Sullivan et al.<sup>23</sup> The basic idea of their approach was, similarly as in our case, to find the appropriate theoretical descriptors correlating with the Hammett sigma constants. The reported equation had the form of a five-parameter multilinear correlation with descriptors involving the quantities such as HOMO and LUMO energies and some topological indices. The whole set involving 274 *meta* and *para* substituted benzoic acids was in their approach divided into the training set consisting of 150 acids and the test set of 124 acids. The statistical parameters of their correlations are summarized below:

$$\left[ \begin{array}{lll} \text{Training Set} & n = 150 & r^2 = 0.958 \quad s = 0.073 \\ \text{Test Set} & n = 124 & r^2 = 0.774 \quad s = 0.148 \end{array} \right] \quad (11)$$

If now the same data were analyzed using our single parameter approach, the following results were obtained:

$$\sigma \equiv -46.527Z_{AA}^{\Omega} + 257.55 \quad (12)$$

$$\left[ \begin{array}{lll} \text{Training Set} & n = 150 & r^2 = 0.869 \quad s = 0.126 \\ & F_{\text{stat}} = 982.31 & r_{cv}^2 = 0.864 \quad q^2 = 0.864 \\ \text{Test Set} & n = 124 & r^2 = 0.794 \quad s = 0.134 \end{array} \right]$$

As it is possible to see, the accuracy of the model (11) is a bit higher for the training set, but taking into account that our approach (eq 12) is based just on a single descriptor compared to a five-parameter model,<sup>23</sup> the quality of both approaches is at least comparable. Moreover, the comparison of standard deviations of the predicted sigma values of the substituents involved in the test sets seems to suggest that the predicting power of a single parameter correlation eq 12 is very probably a bit superior.

## CONCLUSIONS

A new approach to theoretical modeling of the electronic substituent effect is proposed. The approach is based on the use of Quantum-Self-Similarity Measures derived from the so-called domain averaged Fermi holes. This new approach was applied to the description of the polar substituent effect on the dissociation of a series of *meta* and *para* substituted benzoic acids empirically characterized by the values of Hammett sigma constants. The applicability and predicting power of this new approach was tested on the comparison with a recent extensive study by Sullivan et al.<sup>23</sup> It has been shown that although the statistical parameters of both approaches are more or less comparable, the advantage of the present approach is in its surprising simplicity. Instead of a five-parameter correlation required in the study,<sup>23</sup> the comparable description of the substituent effect relies in our approach on only one theoretical descriptor, the corresponding fragment self-similarity measure.

## ACKNOWLEDGMENT

This research has been partially supported by Project No. SAF2000-0223-C03-01 from the Spanish *Ministerio de*

Ciencia y Tecnología and the grant of the Grant Agency of the Czech Ministry of Education No. D0.20. This support is gratefully acknowledged by R.P. One of us (X. Gironés) also acknowledges the University of Girona for a predoctoral fellowship as well as for a grant covering his stay at ICPF Prague. Financial support from the Fundació Maria Francisca de Roviralta is also acknowledged.

## REFERENCES AND NOTES

- (1) Kubinyi, H. Quantitative Structure–Activity Relationships in Drug Design. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley and Sons Ltd.: Chichester, 1998; Vol. 4, pp 2309–2319.
- (2) Jurs, P. C. Quantitative Structure–Property Relationships (QSPR). In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley and Sons Ltd.: Chichester, 1998; Vol. 4, pp 2320–2330.
- (3) Hammett, L. P. The effect of structures upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
- (4) Hammett, L. P. *Physical organic chemistry*; McGraw-Hill: New York, 1940.
- (5) Shorter, J. *Correlation Analysis in Organic Chemistry*; Cambridge University Press: Cambridge, 1973.
- (6) Johnson, C. D. *The Hammett Equation*; Cambridge University Press: Cambridge, 1978.
- (7) Leffler, J. E.; Grundwald, E. In *Rates and Equilibria of Organic Reactions as Treated by Statistical, Thermodynamic and Extrathermodynamic Methods*; Wiley: New York, 1963.
- (8) Famini, G. R.; Wilson, L. Y. Using Theoretical Descriptors in Quantitative Structure Activity Relationships: Application to Partition Properties of Alkyl (1-phenylsulfonyl) Cycloalkane-Carboxylates. *Chemosphere* **1997**, *35*, 2417–2447.
- (9) Genix, P.; Jullien, H.; Le Goas, R. Estimation of Hammett Sigma Constants from Calculated Atomic Charges using Partial Least Squares Regression. *J. Chemometrics* **1996**, *10*, 631–636.
- (10) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, 1995.
- (11) Karelson, M.; Lovanov, V. S.; Katrizky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027–1043.
- (12) Jaffe, H. H. Correlation of Hammett's  $\sigma$  values with Electron Densities calculated by Molecular Orbital Theory. *J. Chem. Phys.* **1952**, *20*, 279–293.
- (13) Peters, D. A simple Molecular Orbital Treatment of the Hammett  $\sigma$ p Equation. Inductive Constants of Common Substituents. *J. Chem. Soc.* **1957**, 2654.
- (14) Gilliom, R. D.; Beck, J.-P.; Purcell, W. P. A MNDO Treatment of Sigma Values. *J. Comput. Chem.* **1985**, *6*, 437–440.
- (15) Sotomatsu, T.; Murata, Y.; Fujita, T. Correlation Analysis of Substituent Effect on the Acidity of Benzoic Acids by the AM1 Method. *J. Comput. Chem.* **1989**, *10*, 94–98.
- (16) Monaco, R. R.; Gardiner, W. C. J. Substituent Effects in Monosubstituted Benzenes: Semiempirical SCF Studies of the Hammett Constants. *Phys. Org. Chem.* **1995**, *8*, 629–636.
- (17) Krygowsky, T. M.; Wozniak, K.; Bock, C. W.; George, P. Correlation of Hammett Substituent Constants with Charge Densities calculated using 6-31G\*(5D) Basis Set. *J. Chem. Res. (S)* **1989**, 396–397.
- (18) Brownlee, R. T. C.; Taft, R. W. Concerning the  $\sigma^0$ R Scale of Substituent  $\pi$  Delocalization Parameters. Theory and Experiment. *J. Am. Chem. Soc.* **1968**, *90*, 6537–6548.
- (19) Mitchell, T. J.; Tute, M. S.; Webb, G. A. A Theoretical Study of the  $pK_a$  Values of some Clodine-Like Imidazolidines. *Eur. J. Med. Chem.* **1990**, *25*, 117–120.
- (20) Topsom, R. D. Some Theoretical Studies of Electronic Substituent Effect in Organic Chemistry. *Prog. Phys. Org. Chem.* **1987**, *16*, 125–193.
- (21) Vaz, R. J. The Use of Electron Densities in Comparative Molecular Field Analysis (CoMFA): A Quantitative Structure–Activity Relationship (QSAR) for Electronic Effect of Groups. *Quant. Struct.-Act. Relat.* **1997**, *16*, 303–308.
- (22) Weiss-López, B.; Bravo, H. R. Substituent Effects in Frontier Orbitals of Cyclic Hydroamic Acid. *Heterocycles* **1994**, *38*, 9–21.
- (23) Sullivan, J. J.; Jones, A. D.; Tanji, K. K. QSAR treatment of Electronic Substituent Effects using Frontier Orbital Theory and Topological Parameters. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1113–1127.
- (24) Von Neumann, J. *Mathematical foundations of quantum mechanics*; Princeton University Press: Princeton, 1955.
- (25) Carbó, R.; Leyda, L.; Arnau, M. How similar is a molecule to another? An electron density measure of the similarity between two molecular structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185–1189.
- (26) Besalú, E.; Carbó, R.; Mestres, J.; Sola, M. Foundations and recent developments on molecular quantum similarity. *Top. Curr. Chem.* **1995**, *173*, 31–62.
- (27) Carbó-Dorca, R.; Besalú, E. A general survey of quantum similarity. *J. Mol. Struct. (THEOCHEM)* **1998**, *451*, 11–23.
- (28) Carbó-Dorca, R.; Amat, L.; Besalú, E.; Gironés, X.; Robert, D. Quantum mechanical origin of QSAR: theory and applications. *J. Mol. Struct. (Theochem)* **2000**, *504*, 181–228.
- (29) Ponec, R.; Amat, L.; Carbó-Dorca, R. Molecular basis of quantitative structure–property relationships (QSPR): a quantum similarity approach. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 259–270.
- (30) Ponec, R.; Amat, L.; Carbó-Dorca, R. Quantum similarity approach to LFER: substituent and solvent effects on the acidities of carboxylic acids. *J. Phys. Org. Chem.* **1999**, *12*, 447–454.
- (31) Ponec, R. Electron pairing and chemical bonds. Chemical structure, valences and structural similarities from the analysis of the Fermi holes. *J. Math. Chem.* **1997**, *21*, 323–333.
- (32) Ponec, R. Electron Pairing and Chemical Bonds. Molecular Structure from the Analysis of Pair Densities and Related Quantities. *J. Math. Chem.* **1998**, *23*, 85–103.
- (33) Ponec R.; Roithová J. Domain averaged Fermi-holes – a new means of visualization of chemical bonds. Bonding in hypervalent molecules. *Theor. Chem. Acc.* **2001**, *105*, 383–392.
- (34) Ponec, R.; Duben, A. Electron pairing and chemical bonds. Bonding in hypervalent molecules from the analysis of the Fermi holes. *J. Comput. Chem.* **1999**, *8*, 760–771.
- (35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, 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.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, M. T.; Al-Laham, A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revision A.6*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- (36) Gironés, X.; Ponec, R.; Roithova, J. WinFermi, v1.0., Institute of Chemical Process Fundamentals, Academy of Sciences of Czech Republic, Prague, Czech Republic, 2001.
- (37) Mestres, J.; Solà, M.; Besalú, E.; Duran, M.; Carbó-Dorca, R. MESSEM 1.0, A Molecular Quantum Similarity Measures System of Programs, Institute of Computational Chemistry, Girona, Spain, 1993.
- (38) Exner, O. *Correlation Analysis of Chemical Data*; Plenum Press: New York, 1988.

CI020050I