QSAR Treatment of Electronic Substituent Effects Using Frontier Orbital Theory and Topological Parameters

Jonathan J. Sullivan,*,† A. Daniel Jones,§ and Kenneth K. Tanji†

Department of Land, Air, and Water Resources, University of California, Davis, California 95616, and
Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

A methodology for the estimation of Hammett substituent constants from computational-based descriptors utilizing quantitative structure activity/property relationships (QSAR/QSPR) formalism is presented. Electronic descriptors derived from quantum chemical calculations and molecular topology were used to generate computational-based analogues of empirical Hammett substituent constants from statistical analysis. Global quantum chemical reaction indices were drawn from frontier orbital theory and density functional theory and formulated from AM1-based calculations. A localized index based on the electrotopological state index was used to encode information on individual group properties. From a training set consisting of 150 *meta* and *para*-substituted benzoic acids, statistical analysis of computational-based descriptors as a function of empirical substituent constants yielded a five-parameter QSAR/QSPR model which generates computational-based constants exhibiting a strong correlation with empirical values ($r^2 = 0.958$). Both internal (PRESS) and external (independent testing set of benzoic acids) validation procedures suggest that the electronic effects QSAR/QSPR model derived in this work from computational-based parameters is a statistically viable paradigm. Both predicted and empirical constants were used in Hammett-type validation analyses as functions of chemical, biological, and spectroscopic data for thirty structurally diverse *meta* and *para*-substituted aromatic testing sets. Statistical measures of ensuing correlations were examined and compared, and the empirical and predicted results were of similar quality. Validation results reveal that a large number of computational-based substituent constants can be accurately estimated from semiempirical AM1 frontier orbital energies and electronic structure information obtained directly from substituted benzoic acids without the aid of empirical parametrization.

1. INTRODUCTION

Organic chemistry has tradtionally relied more on qualitative and empirical rules than on models deduced from theoretical postulates. This is because the complexity of most organic reactions has made it difficult to establish rigorous quantitative relationships between experiment and theory. Consequently, chemists have developed numerous empirical models which attempt to quantify observed and experimental phenomena whose explicit meaning and significance have otherwise been left unresolved by theory.[1] One of the most widely utilized and successful of these empirical paradigms is that of linear free-energy relationships (LFER). LFERs, as originally developed by Hammett,[2] were initially intended to quantify electronic effects of substituent groups on the hydrolysis of aromatic esters. However, it was subsequently determined that many equilibrium and rate processes can be systematized when the influence of each substituent on the reactivity of substrates is assigned a characteristic substituent constant σ and the reaction parameter ρ is known or can be calculated.[3] LFERs, such as the Hammett equation

$$\log\frac{k_X}{k_H} = \rho\sigma \qquad (1)$$

thus provide a quantitative account of proportional changes in free energy which occur as a result of the electronic perturbation of reaction centers by nearby substituent groups. This approach has been successfully utilized to correlate reactivities, physical properties, and biological activities with structural features of organic compounds.[4] Hammett-type equations have been widely employed to investigate mechanism and transition states and to elucidate the effects of structure on reaction rates and equilibria.[5] The use of Hammett σ constants as structural descriptors in quantitative structure−property/activity relationship (QSPR/QSAR) studies has proved to be one of the major innovations in the quest to understand the relationships between electronic structure and corresponding chemical properties and behavior in applied pharmaceutical, biological and environmental chemistry.[6] Hammett constants have been employed as a scale of electron withdrawing or donating capabilities to rationalize an electronic effect implied in a given chemical or biological process.[7,8]

Despite the widespread success of the Hammett approach, however, there is often considerable ambiguity in the interpretation of its empirical evidence, since insights and conclusions drawn from such evidence can only be rational-

---

* Corresponding author phone: (916)322-6767; e-mail: jsullivan@cdpr.ca.gov. Corresponding author address: California Department of Pesticide Regulation, Pesticide Registration Branch, 1116 9th Street, Sacramento, CA 95814.
† University of California.
§ The Pennsylvania State University.

ized by inference, chemical intuition, and experience.[9] Empirical models are often plagued by persistent uncertainties in the quality of available experimental data and by the ambiguous chemical meaning of empirical terms. Substituent constants have been derived using a variety of experimental, solvent, and model systems, so values for substituent constants from different literature sources often differ considerably (although there generally is widespread agreement for the most common groups).[10] It is not uncommon to find several different σ-values for a given substituent, and it is often the case that data for a particularly active or promising substituent are simply not available. These limitations accentuate the need for general procedures to quickly calculate Hammett-type substituent constants from theoretical considerations.[11] With the recent rapid growth in the power and availability of computers and with the development of more efficient algorithms, the use of computational methodologies and computational-based descriptors to probe the effects of structure on chemical reactivity, physicochemical properties, and biological activities has increased steadily. Theory-based approaches offer higher accuracy, reproducibility and more detailed descriptions of electronic effects than corresponding empirical methods and perhaps may help clarify long-standing disparities between experiment and theory.[12] The first attempts to assign theoretical significance to substituent constants using computational approaches were based on simple Huckel calculations.[13,14] More recent attempts have been made employing ab initio and semiempirical quantum chemical parameters, with particular emphasis on their relationship with atomic charges,[15-18] electron densities,[19] and orbital energies.[20-22] In addition, some success has been achieved through the use of 3-D electrostatic fields[23,24] and topological indices.[25]

This paper presents a simple methodology for the calculation of Hammett-like substituent constants for aromatic systems from computational-based descriptors utilizing QSAR/QSPR formalism. The combined use of theoretical parameters and QSAR/QSPR has become an increasingly effective and useful for the characterization of chemical data. The QSAR/QSPR approach, while more empirical in nature than the direct computation through quantum mechanical or statistical mechanical means, provides more versatility for computing physical and chemical properties within the limits of available experimental data.[12] Until relatively recently, QSAR/QSPR studies have primarily employed empirically based descriptors as independent variables. But because of the large well-defined physical information content encoded in computational expressions and because they are currently readily calculable, their use as descriptors in QSAR/QSPR applications has become more appealing. The advantage in the use of such descriptors is that compounds and their various fragments and substituents can be directly characterized on the basis of their molecular structure and proposed behaviors can be accounted for directly in terms of chemical reactivity of the compounds under investigation. In addition, unlike experimental measurements there is no statistical error in quantum-chemical calculations (although there is inherent error associated with assumptions required to facilitate the calculations), thus structure-based models avoid the error associated with the measurement of molecular properties. In using quantum-chemical based descriptors with a series of related compounds, the computational error is considered

approximately constant throughout the series.[12]

For this study, electronic descriptors derived from semiempirical quantum chemical and molecular topological parameters were used to generate computational-based analogues of empirical Hammett constants from statistical analysis. Several orbital-based descriptors were formulated from frontier orbital theory[26] and from density functional theory.[27] A localized descriptor based on the electrotopological state index[28] was used to encode information on individual group properties. The specific aim of this work was to derive nonempirical analogues of empirical substituent constants for probing structual effects on the reactivity of aromatic systems from computational-based parameters without altering the form or the meaning of the original Hammett relationship. It was anticipated that such analogues might be useful for (1) providing novel insights into the effects of structure on the chemical reactivity and the physicochemical behavior of aromatic substrates, (2) providing a simple methodology for calculating group constants for new or uncommon groups in lieu of experiment and (3) for use as alternative, computational-based electronic descriptors in QSAR/QSPR applications.

2. METHODOLOGY

In the present work, the overall electronic substituent effect was expressed in terms analogous to those obtained empirically by Hammett, i.e., as a product of two factors, one (σ) characteristic of the substituent and one (ρ) characteristic of the reaction center. It was found that these factors could be maintained by invoking QSAR/QSPR formalism and expressing substituent constants in statistical terms

$$\log(K_X/K_H) = \sigma\rho = (a_1X_1 + a_2X_2 + ... + a_nX_n)\rho \quad (2)$$

where a_1, a_2, ..., a_n are statistically determined fitting coefficients and X_1, X_2, ..., X_n are computational-based quantities representing electronic features of substituents. Following Hammett's original approach, benzoic acid was selected as the model substrate, and *meta* and *para*-substituted benzoic acids as purveyors of the electronic effects of individual substituents. Representative groups from many of the major classes of substituents, e.g., aliphatics, esters, ethers, amines, sulfoxides, sulfones, phosphates, etc., were included for the purpose of model calibration. Training set compounds were selected randomly from among these various classes using a large set of substituent constants compiled by Hansch, Leo, and Hoekman.[29] Substituent groups known to lie beyond the range of validity of the Hammett equation were not, however, considered for the purpose of model calibration. Thus, for instance, unipolar substituents, i.e., those bearing a full unit charge ($N^+(CH_3)_3$, NH_3^+, COO^-, SO_3^-, O^-, etc.) whose behavior in solution are strongly dependent on ionic strength, were excluded from the training set. In addition, the utilization of certain substituents having peculiar properties (e.g., groups capable of hydrogen-bonding in aqueous solution such as alcohols and amines) was limited. Values of X encoding structural information were obtained directly from molecular orbital and graph theoretical sources, respectively. Since for the model reaction (the ionization of benzoic acids) the reaction

QSAR Treatment of Electronic Substituent Effects

J. Chem. Inf. Comput. Sci., Vol. 40, No. 5, 2000 **1115**

constant $\rho = 1.00$, eq 1 reduces to

$$\log(K_X/K_H) = \sigma = (a_1X_1 + a_2X_2 + ... + a_nX_n) \quad (3)$$

and a QSAR/QSPR-based counterpart to Hammett's original mathematical definition of substituent constants is generated. Computational-based analogues (σ^c) of empirical substituent constants (σ) were obtained using eq 3 by analyzing computational-based descriptors, X, as functions of empirical constants, σ, for a large training set of substituted benzoic acids. Comprehensive external validation of the resulting model equation was achieved by analyzing numerous thermodynamic, kinetic, spectroscopic, and biological datasets from the literature using both the empirical Hammett equation

$$\log(K_X/K_H) = \sigma\rho \quad (4)$$

and its computational-based analog

$$\log(K_X/K_H) = \sigma^c\rho^c \quad (5)$$

where ρ^c is the slope of the regression line (reaction constant) obtained from correlations of reaction data with the computational-based term σ^c.

2.1. Calculation of Theoretical Parameters. When a substituent and substrate are joined to form a molecule, e.g., a nitro group and benzoic acid, the new system has properties that are different from the original two groups. In molecular orbital terms, the two groups are said to perturb each other. If substituents are varied sequentially within a congeneric or homologous series, perturbations become systematized and subsequent regularities are reflected by changes in the molecular orbitals of the system. On the basis of this propensity, it was expected that frontier orbital energies would provide a means to quantitatively account for such changes, since the HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) are the first orbitals to become involved in both intermolecular and intramolecular interactions.[30] Orbital energies tend to vary predictably in response to variations in the donor or acceptor properties of substituent groups. Both the HOMO and the LUMO energies of substituted benzoic acids, for example, are normally lowered relative to those of the unsubstituted acid when an acceptor substituent group is introduced to the system and raised when a donor group is present. This behavior suggests that descriptors drawn from frontier orbital energies might successfully be utilized to rationalize molecular or *global* electronic effects of substituents on the thermodynamic properties of benzoic acids.

The AM1 Hamiltonian[31] was chosen as the means for calculating all quantum chemical parameters for comparison with independent variables in this study. In a recent review by Clare,[32] the AM1 model was found to be one of the more accurate semiempirical methods for the determination of orbital energy parameters in terms of correlations with experiment and extended basis set *ab initio* calculations. Calculation of semiempirical HOMO and LUMO energies and other electronic structure information was performed on a Silicon Graphics workstation using SPARTAN, version 4.1.[33] While transition state rather than ground state structures are preferable for obtaining parameters for the analysis of kinetic data, the geometries of transition states on the

pathway between reactants and products are not easily anticipated. Semiempirical models and to some extent minimal basis set ab initio models have proved unreliable in their description of transition state geometries.[34] Consequently, geometry optimization was implemented for all structures analyzed for this study in order to obtain geometries in the lowest energy conformations. Optimization was achieved for all benzoic acid congeners without any failure. From these minimized geometries, the molecular orbital (MO) parameters of interest were obtained.

2.2. Descriptor Generation. *Global Orbital Energy Terms.* Global electronic descriptors were constructed from several well-known orbital energy-dependent properties, including absolute electronegativity (χ), absolute hardness (η), ionization potential (IP), and electron affinity (EA). HOMO and LUMO energies are definitive components of each of these terms. According to Koopmans Theorem,[35] global IP and EA may be estimated by

$$IP = -E_{HOMO} \quad (6)$$

$$EA = -E_{LUMO} \quad (7)$$

where IP reflects the energy associated with the ground-state removal of an electron and EA reflects the energy associated with the ground-state addition of an electron. The absolute electronegativity, χ, defines whether a molecule, atom, ion, or radical is a net electron acceptor or donor and is approximated by

$$\chi = (IP + EA)/2 = (E_{LUMO} + E_{HOMO})/2 \quad (8)$$

For an equilibrium system, the global electronegativity must be constant everywhere in the molecule.[36] The absolute hardness, η, determines how easily electrons are polarized or resist deformation and need not be constant and can have local values[37]

$$\eta = (IP - EA)/2 = (E_{LUMO} - E_{HOMO})/2 \quad (9)$$

Relative to its positive and negative ions, a chemical species is more stable the larger its absolute hardness.[38]

The introduction of a substituent group at the *meta* or *para* position on the aromatic ring of benzoic acid induces a perturbation that may raise or lower the orbital energy of the system relative to the unsubstituted substrate. The degree of perturbation is a function of the degree of overlap of the orbitals, i.e., the more efficient the orbital overlap, the stronger the interaction and the lower the energy.[39] Parr and co-workers[40] introduced an expression from density functional theory which reflects the mutual perturbation of frontier orbitals and ensuing energy change which occurs when a donor species, D, and an acceptor species, A, initially interact:

$$\Delta E = -\frac{(-\chi_A + \chi_D)^2\eta_D}{4(\eta_A + \eta_D)^2} + \left[-\frac{(-\chi_A + \chi_D)^2\eta_A}{4(\eta_A + \eta_D)^2}\right] \quad (10)$$

Substituting orbital energy parameters for the unsubstituted (H) and substituted (X) benzoic acids in place of those for

the generic donor and acceptor species, eq 10 becomes

$$\Delta E = -\frac{(-5.27 + \chi_X)^2 \eta_X}{4(4.81 + \eta_X)^2} + \left[-\frac{(-5.27 + \chi_X)^2 4.81}{4(4.81 + \eta_X)^2}\right]$$

$$(11)$$

where the calculated values of χ and η for the reference compound, benzoic acid, are 5.27 and 4.81, respectively. Equation 11 was used as an electronic descriptor to account for global energy variations that occur in benzoic acids due to the interaction of frontier orbitals upon substitution.

When a Lewis acid, A, and a Lewis base, B, interact with each other, electron density is transferred from the species of lower electronegativity to that of higher electronegativity until equilibrium is reached. From density functional theory, the approximate fractional number of electrons transferred during a general acid−base reaction is given by

$$\Delta N = \frac{(\chi_A - \chi_B)}{2(\eta_A + \eta_B)} \qquad (12)$$

Electron transfer is driven by the difference in electronegativity, while the hardness sum acts as a resistance.[41,42] There is a net stabilization (lowering of energy) due to electron density being transferred to the system of lower electronegativity, although eq 12 accounts for only the initial part of the total energy change.[43] Specifying water as the electron rich (low electronegativity) system and the substituted acids as the electron deficient (high electronegativity) systems, we obtain a second electronic descriptor based on density functional theory which encodes information concerning electronic substituent effects on benzoic acid equilibria

$$\Delta N = \frac{(\chi_{H_2O} - \chi_X)}{2(\eta_{H_2O} + \eta_X)} = \frac{(3.1 - \chi_X)}{2(9.5 + \eta_X)} \qquad (13)$$

where the calculated electronegativity and hardness values of water are 3.1 and 9.5 eV, respectively. Equation 13 was found a convenient measure of the initial movement of electrons which occurs within the model aromatic system as substituents are varied.

The ionization of benzoic acid is a reversible reaction, hence the ionization products, the benzoate anion and hydronium cation, are themselves reactant molecules which may again react to reform the parent acid.
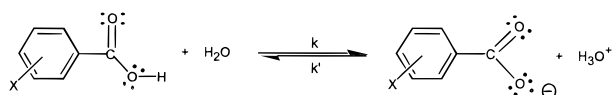
Eventually a state of equilibrium is reached, with the overall equilibrium constant K = k/k′. Gilliom et al.[14] reported that if the equilibrium constant for the ionization of benzoic acid is determined primarily by the rate of proton detachment (k), a linear relationship of σ with the reciprocal of EA (or -E_{LUMO}) of the acid may be generated. Similarly, consideration of the rate of proton attachment (k′) may lead to a correlation of σ with the reciprocal of IP (or -E_{HOMO}) of the benzoate anion. This approach was investigated for several structurally heterogeneous datasets of substituted benzoic acids, and statistically significant (r > 0.85) linear

relationships between σ and 1/EA of the acid were consistently observed. Since the global EA is the energy demand required to add an electron to a molecule, it may be considered to be proportional to the molecule's electron-withdrawing propensity as an electron acceptor.[44] The term 1/EA thus provides an effective account of the donor/acceptor characteristics of substituents, relative to the unsubstituted acid. Donor groups have values of 1/EA larger than that of benzoic acid, and acceptor groups have values of 1/EA smaller than that of benzoic acid. The greater the magnitude of 1/EA, the stronger the donor group. For acceptor groups, the opposite is true.

The inverse of EA of the acid multiplied by the topological (integer) distance D between the substituent atom attached to the aromatic ring and the reaction center on the side group was found to provide an even more effective descriptor for characterizing electron-donating or accepting properties of substituents. Normalizing EA and D with respect to the reference state creates a dimensionless term that is consistent with the theoretical definition of Hammett σ constants

$$\left[\frac{D_X\left(\frac{1}{EA_X}\right)}{D_H\left(\frac{1}{EA_H}\right)}\right] = \left[\frac{D_X\left(EA_H\right)}{D_H\left(EA_X\right)}\right] \qquad (14)$$

where EA_H and D_H correspond to the electron affinity and distance term, respectively, for benzoic acid (R = H). The distance parameter D provides a means with which to scale calculated values of EA_H/EA_X according to whether the substituent of interest is located at the *meta* or *para* position on the aromatic core. In the case of substituted benzoic acids, the reaction center is the carboxylic acid function, hence D = 4 for *meta* substituents and D = 5 for *para* substituents (D_H has been arbitrarily set at D = 5). Such a scaling factor may be rationalized by the fact that electronic effects, in particular resonance effects, are strongly position-dependent. Resonance interaction with the aromatic ring or reaction center is largely inhibited for substituents in the *meta* position due to restrictions imposed by the geometry of the benzene ring. Conversely, conjugative interactions are highly favorable for para substituents. Since EA_H/EA_X is a global descriptor and does not explicitly account for localized properties such as group position, the scaling term D modifies values accordingly.

Localized Group Property Term. A localized descriptor based on the electrotopological state index was also devised for use as a group term in this study. This term is intended to provide a quantitative account of each substituents unique structural properties in terms of the varied electrical and topological environments occupied by its constituent atoms. The electrotopological state index is an atom-level topological index derived from information acquired from the chemical graph of a molecule and is based on the notion that each atom in a molecule is distinguished both by its own inherent structural and electronic features and by its position within the network of bonds (connectivity) which define the molecule.[45] Each atom in a molecule is thus uniquely characterized by its individual, neighboring, and global electronic and topological surroundings and interactions. Atoms are assigned a characteristic an intrinsic state values, I_i, which is quantitatively defined as the ratio of the

QSAR Treatment of Electronic Substituent Effects

J. Chem. Inf. Comput. Sci., Vol. 40, No. 5, 2000 **1117**

atoms valence electronegativity to its local connectedness

$$I_i = \frac{\left(\frac{2}{N}\right)^2 \delta^v + 1}{\delta} \qquad (15)$$

where N is the atoms principal quantum number. The delta terms $\delta = \sigma - H$ and $\delta^v = \sigma + \pi + n - H$, where σ is the number of σ electrons, π is the number of π electrons, n is the number of nonbonding electrons, and H is the number of hydrogen atoms bonded to the atom of interest.[40] By this methodology, each atom is assigned a characteristic value that reflects its properties as a function of itself and its neighboring atoms. To account for the influence on an atom by all other atoms in the molecule, a perturbation term arises from each pair of atoms in the chemical graph. For a given pair of atoms, this term depends on the electronegativity difference and on the degree of adjacency and graph separation of the two graph vertexes and is expressed by

$$\Delta I_i = \sum_{j=1}^{N-1} (I_i - I_j)/r_{ij}^2 \qquad (16)$$

where r_{ij} is the number of skeletal atoms in the shortest path between atoms i and j, including both atoms i and j. The electrotopological state

$$S_i = I_i + \Delta I_i \qquad (17)$$

of an atom i is then derived mathematically from the intrinsic state value of the atom in its hybrid conformation, i.e., constrained in a molecule.[46] Electrotopological state values for all compounds investigated in this study were calculated on Macintosh computers from connection tables using the software package Molconn-X.[47]

For each member of the benzoic acid training set, electrotopological state calculations were performed and the values for all atoms contained in each substituent, exclusive of hydrogen, were summed. These sums are defined by

$$S_X = \sum_{i=1}^{X} S_i \qquad (18)$$

where S is the electrotopological state index for the ith atom of a given substituent group and the sum is taken over all non-hydrogen atoms in the substituent, X. Extensive evaluation of calculated S_X values for numerous *meta* and *para* substituents suggest that the electrotopological state is fundamentally insensitive to group position. Calculated values of S_X are similar in magnitude whether a given group is located at the *meta* or at the *para* position Consequently, as a means to more satisfactorily reflect observed differences in the electronic character of a given substituent due to its position on the benzene ring, eq 18 was slightly modified. A term ϕ was introduced to help characterize certain structural and electronic properties of the substituent atom directly attached to the phenyl ring in terms of fundamental chemical properties.

$$\phi S_X \qquad (19)$$

This atom is of particular significance because its electronic attributes largely reflect those of the substituent group to which it belongs.[48] The term $\phi = [DN/(100/G)]$, where D is the topological distance between the substituent atom bonded to the aromatic ring and the reaction site ($D_{meta} = 5$, $D_{para} = 6$), N is its principal quantum number, and G is its group number. The expression 100/G represents a general method for estimating the % *s* character of a bonded atom.[49] The principal quantum number, *N*, of a given atom provides qualitative information concerning the relative size, electronegativity, and average energy of its bonding orbitals and about their susceptibility to the polarizing effects of adjacent atoms.

π Electron Count. To provide a general account of the effect of π electrons on the overall electronic character of substituents, a simple indicator variable, $\Sigma\pi$, was devised. The descriptor $\Sigma\pi$ is a count of all π electrons present in a given substituent group and encodes qualitative information which implies that some abstruse electronic effects may be linked to the presence and number of both conjugated and nonconjugated π electrons in the group. The presence of conjugated and nonconjugated π electrons in a substituent may have a notable impact on the electronic character of the substrate molecule to which it is attached. A substituent atom containing an isolated or conjugated π bond adjacent to the phenyl ring generally behaves as an electron donor through resonance. However, if the atom is also strongly electronegative, e.g., $-O-$, $-$halogen, the substituent may behave as an electron acceptor through inductive polarization. In such cases, e.g., $-O-R$, empirical constants tend to be larger than expected through resonance alone. If the atom is bonded to a heteroatom to which π electrons may be donated, it may act as an electron acceptor, e.g., $-C(=O)R$, $-P(=S)R_2$. Groups containing multiple bonds which are not in a conjugative position may also influence the general electronic character of the molecule, since an atom possessing one or two π bonds is more electronegative than neighboring sp^3-hybridized atoms, i.e., $sp > sp^2 > sp^3$.[50] Considering the total number of π electrons contained in a given substituent group rather than only those that are conjugated helps to minimize otherwise slight aberrations in predicted values of σ^c which may be due to subtle conjugative effects not accounted for theoretically.

3. RESULTS AND DISCUSSION

3.1. Model Development and Internal Validation. The training set of 150 *meta* and *para* substituent groups utilized for this study is shown in Table 1, along with their corresponding frontier orbital energies, experimental, predicted and residual σ values, and calculated values of all independent variables. Only substituent groups whose effects are explicated through the use of the original Hammett equation (eq 1) were selected for the training set since only *normal* experimental substituent constants were used as dependent variables. Thus substituents bearing full charges or exhibiting known anomalous behavior ($N(CH_3)_2$, OH, F) were avoided for the purpose of model calibration. Descriptors were constructed from molecular and atomic level properties thought best able to account for substituent-induced variations in global energy and electron redistribution that occurs in benzoic acids upon change of substituent.

Statistical analysis of empirical constants as a function of calculated descriptors yielded a five-parameter regression

Table 1. Calculated Frontier Orbital Energies, Sum of the Electrotopological State Index for Substituent Atoms (S_X), Experimental Substituent Constants (σ)[a], Calculated Substituent Constants (σ^c), and Residuals for the 150-Member Training Set Used for This Study

	substituent	E_{HOMO}	E_{LUMO}	S_X	σ	σ^c	resid
1	H	−10.0840	−0.4685	0.0000	0.00	−0.02	0.02
2	m-$(CH_2)_4CH_3$	−9.7775	−0.4228	9.4387	−0.08	−0.06	−0.02
3	m-$(CH_2)_6CH_3$	−9.7770	−0.4234	3.1671	−0.07	−0.05	−0.02
4	m-Br	−9.9599	−0.7402	21.5618	0.39	0.20	0.19
5	m-$C(CH_3)(CN)_2$	−10.4752	−0.8743	6.1958	0.60	0.56	0.04
6	m-$C(CH_3)_3$	−9.9838	−0.4607	6.4595	−0.10	−0.01	−0.09
7	m-$C(CH_3)\!\!=\!\!CH_2$	−9.3712	−0.4888	2.8771	0.09	0.08	0.01
8	m-C_2H_5	−9.7906	−0.4223	9.3739	−0.07	−0.07	0.00
9	m-CBr_3	−10.2486	−0.9392	7.2124	0.28	0.30	−0.02
10	m-$CCCH_3$	−9.4224	−0.5651	7.4324	0.21	0.20	0.01
11	m-CCH	−9.6573	−0.6173	5.6486	0.20	0.23	−0.03
12	m-$CH(CH_3)CH_2CH_3$	−9.7920	−0.4097	10.1142	−0.08	−0.08	0.00
13	m-$CH(OH)CH_3$	−10.0026	−0.5481	5.7297	0.08	0.08	0.00
14	m-$CH_2CH(CH_3)_2$	−9.7560	−0.4156	10.8084	−0.07	−0.07	0.00
15	m-$CH_2CH(OH)(CH_3)$	−9.7491	−0.3836	6.6787	−0.12	−0.10	−0.02
16	m-CH_2Cl	−10.1050	−0.6407	5.8649	0.11	0.14	−0.03
17	m-CH_2F	−9.9967	−0.5988	11.3520	0.12	0.12	0.00
18	m-CH_2I	−9.9005	−0.6589	3.0541	0.10	0.14	−0.04
19	m-CH_2OCH_3	−9.8971	−0.5025	6.8919	0.08	0.03	0.05
20	m-CH_3	−9.7451	−0.4236	1.8675	−0.07	−0.07	0.00
21	m-$CH\!\!=\!\!C(CN)_2$	−10.0731	−1.4623	21.7027	0.66	0.76	−0.10
22	m-$CH\!\!=\!\!NOCH_3$	−9.4980	−0.6252	10.8716	0.37	0.19	0.18
23	m-$CHCl_2$	−10.3211	−0.8458	10.4324	0.31	0.28	0.03
24	m-CHI_2	−10.0305	−0.6589	4.8108	0.26	0.15	0.11
25	m-CHO	−10.4365	−0.8977	10.8059	0.35	0.37	−0.02
26	m-Cl	−10.1466	−0.8666	5.5256	0.37	0.28	0.09
27	m-CN	−10.6228	−1.1683	10.2598	0.56	0.55	0.01
28	m-$COC(CH_3)_3$	−10.2309	−0.6834	16.7337	0.27	0.26	0.01
29	m-$COCF_3$	−10.7969	−1.4325	39.5959	0.63	0.70	−0.07
30	m-$COCH_2CH_3$	−10.3145	−0.8408	13.2973	0.38	0.34	0.04
31	m-COCl	−10.6661	−1.1646	15.0861	0.51	0.51	0.00
32	m-$CONH_2$	−10.3422	−0.7092	14.9189	0.28	0.28	0.00
33	m-$COOCH_3$	−10.3775	−0.8565	16.1081	0.37	0.36	0.01
34	m-COOH	−10.5172	−0.9603	17.7568	0.37	0.42	−0.05
35	m-$CSNHCH_3$	−8.5831	−0.8893	9.9985	0.30	0.33	−0.03
36	m-I	−10.1831	−0.8925	2.0754	0.35	0.28	0.07
37	m-isoC_3H_7	−10.0161	−0.4786	4.4642	−0.04	0.00	−0.04
38	m-$N(C_3H_7)_2$	−8.8013	−0.3610	10.5578	−0.26	−0.07	−0.19
39	m-$N(CH_2CH_3)_2$	−9.0338	−0.4089	7.9995	−0.23	−0.04	−0.19
40	m-$N(COCH_3)_2$	−10.0805	−0.7429	24.9641	0.35	0.38	−0.03
41	m-$N\!\!=\!\!C\!\!=\!\!S$	−9.8783	−1.1189	10.1956	0.48	0.45	0.03
42	m-$N\!\!=\!\!NCN$	−10.6205	−1.7512	16.1035	0.71	0.82	−0.11
43	m-NH_2	−9.0282	−0.4164	5.3331	−0.16	−0.03	−0.13
44	m-$NHCH_3$	−8.7511	−0.3515	17.2639	−0.21	−0.10	−0.11
45	m-$NHCOCH(CH_3)_2$	−9.0948	−0.4210	19.7442	0.11	0.06	0.05
46	m-NO_2	−10.9683	−1.4599	10.1351	0.71	0.76	−0.05
47	m-$O(CH_2)_3CH_3$	−9.3264	−0.4697	9.1935	0.10	0.03	0.07
48	m-$O(CH_2)_4CH_3$	−9.3294	−0.4684	11.5119	0.10	0.03	0.07
49	m-$OCH(CH_3)_2$	−9.5440	−0.4928	10.6392	0.10	0.04	0.06
50	m-$OCH_2CH\!\!=\!\!CH_2$	−9.3906	−0.5139	10.2162	0.09	0.13	−0.04
51	m-OCH_2Cl	−9.6682	−0.6346	6.3950	0.25	0.15	0.10
52	m-$P(CH_2CH_3)_2$	−8.8048	−0.4978	9.7168	0.10	0.08	0.02
53	m-$P(isopropyl)_2$	−8.8168	−0.4917	20.2364	0.02	0.10	−0.08
54	m-$P(O)(C_4H_9)_2$	−10.4256	−0.8750	18.1898	0.35	0.43	−0.08
55	m-$P(O)(isopropyl)_2$	−10.4065	−0.8971	21.7470	0.37	0.43	−0.06
56	m-$P(O)CL_2$	−11.1514	−1.6471	18.4519	0.78	0.84	−0.06
57	m-$P(O)F_2$	−10.9674	−1.4073	29.4324	0.81	0.77	0.04
58	m-$P(O)[C(CH_3)_3]_2$	−10.2931	−0.8104	10.1599	0.31	0.40	−0.09
59	m-$P(S)(CH_2CH_3)_2$	−9.1778	−1.3841	2.4451	0.39	0.43	−0.04
60	m-PH_2	−9.4561	−0.6097	14.8649	0.06	0.11	−0.05
61	m-$PO(CH_2CH_3)_2$	−10.4613	−0.8927	20.9648	0.37	0.42	−0.05
62	m-$P[N(CH_3)_2]_2$	−8.3181	−0.4031	15.3514	0.18	0.13	0.05
63	m-$S(O)OCH_3$	−9.9408	−0.8421	6.3055	0.50	0.37	0.13
64	m-$SCH(CH_3)_2$	−8.3809	−0.4853	7.7512	0.23	0.14	0.09
65	m-$SCH_2CH\!\!=\!\!CH_2$	−8.3952	−0.5117	12.2654	0.19	0.22	−0.03
66	m-SCH_2F	−8.7869	−0.6122	3.4547	0.23	0.21	0.02
67	m-SCH_3	−8.3956	−0.5159	6.6211	0.15	0.14	0.01
68	m-$SCH\!\!=\!\!CH_2$	−8.3576	−0.6333	21.5470	0.26	0.29	−0.03
69	m-$SCHF_2$	−9.1307	−0.9247	6.4366	0.33	0.36	−0.03
70	m-SCl	−8.8424	−1.5360	11.1351	0.44	0.37	0.07
71	m-SCN	−8.9779	−0.8570	7.2472	0.51	0.39	0.12
72	m-$SN(CH_3)_2$	−8.2432	−0.4099	20.9459	0.12	0.09	0.03
73	m-$SO_2CH_2CH_3$	−10.6998	−1.1339	22.7249	0.66	0.65	0.01
74	m-SO_2Cl	−11.0861	−2.2648	28.2179	1.20	1.17	0.03
75	m-SO_2F	−10.9089	−1.3775	22.5680	0.80	0.82	−0.02
76	m-SO_2NH_2	−10.5915	−1.0202	11.3243	0.53	0.60	−0.07
77	p-$(CH_2)_4CH_3$	−9.8478	−0.4572	9.6430	−0.15	−0.13	−0.02

QSAR Treatment of Electronic Substituent Effects

J. Chem. Inf. Comput. Sci., Vol. 40, No. 5, 2000 **1119**

Table 1 (Continued)

	substituent	E_{HOMO}	E_{LUMO}	S_X	σ	σ^c	resid
78	p-(CH$_2$)$_6$CH$_3$	−9.8459	−0.4574	3.2016	−0.16	−0.12	−0.04
79	*p*-Br	−10.0418	−0.8553	21.6896	0.23	0.20	0.03
80	*p*-C(CH$_3$)(CN)$_2$	−10.5031	−0.9910	6.3803	0.57	0.57	0.00
81	*p*-C(CH$_3$)$_3$	−9.9838	−0.4927	6.4865	−0.20	−0.09	−0.11
82	*p*-C(CH$_3$)=CH$_2$	−9.4712	−0.6450	2.9818	0.05	0.10	−0.05
83	*p*-C$_2$H$_5$	−9.8629	−0.4552	9.5203	−0.15	−0.15	0.00
84	*p*-CBr$_3$	−10.3092	−1.1890	7.3071	0.29	0.36	−0.07
85	*p*-CCCH$_3$	−9.5044	−0.7758	7.5676	0.21	0.24	−0.03
86	*p*-CCH	−9.7364	−0.8249	5.9459	0.23	0.27	−0.04
87	*p*-CH(CH$_3$)CH$_2$CH$_3$	−9.8611	−0.4401	10.1922	−0.12	−0.16	0.04
88	*p*-CH(OH)CH$_3$	−10.0605	−0.6155	5.9459	−0.07	0.05	−0.12
89	*p*-CH$_2$CH(CH$_3$)$_2$	−9.9993	−0.5255	10.9300	−0.12	−0.05	−0.07
90	p-CH$_2$CH(OH)(CH$_3$)	−9.7754	−0.4057	6.8592	−0.17	−0.20	0.03
91	*p*-CH$_2$Cl	−10.1694	−0.7840	5.9459	0.12	0.15	−0.03
92	*p*-CH$_2$F	−10.0693	−0.6702	11.3551	0.11	0.10	0.01
93	*p*-CH$_2$I	−9.9669	−0.8458	3.2432	0.11	0.16	−0.05
94	*p*-CH$_2$OCH$_3$	−9.9642	−0.5621	7.0270	0.01	−0.01	0.02
95	*p*-CH$_3$	−9.8258	−0.4644	1.9223	−0.17	−0.14	−0.03
96	*p*-CH=C(CN)$_2$	−10.1552	−1.6634	21.8919	0.84	0.80	0.04
97	*p*-CH=NOCH$_3$	−9.5422	−0.8585	10.9728	0.30	0.23	0.07
98	*p*-CHCl$_2$	−10.3380	−0.9691	10.5405	0.32	0.28	0.04
99	*p*-CHI$_2$	−10.0846	−1.0502	4.8649	0.26	0.26	0.00
100	*p*-CHO	−10.3767	−1.1757	10.8125	0.42	0.42	0.00
101	*p*-Cl	−10.0213	−0.7779	5.5234	0.23	0.17	0.06
102	*p*-CN	−10.4687	−1.1991	10.5405	0.66	0.50	0.16
103	*p*-COC(CH$_3$)$_3$	−10.2729	−0.7833	16.8749	0.32	0.26	0.06
104	*p*-COCF$_3$	−10.7053	−1.7026	39.5329	0.80	0.77	0.03
105	*p*-COCH$_2$CH$_3$	−10.3072	−1.0635	13.5135	0.48	0.38	0.10
106	*p*-COCl	−10.5817	−1.4107	15.1025	0.61	0.55	0.06
107	*p*-CONH$_2$	−10.3451	−0.9552	14.8649	0.36	0.34	0.02
108	*p*-COOCH$_3$	−10.3773	−1.1454	16.2162	0.45	0.43	0.02
109	*p*-COOH	−10.4416	−1.2255	17.8378	0.45	0.47	−0.02
110	*p*-CSNHCH$_3$	−8.6105	−1.0906	10.1240	0.34	0.32	0.02
111	*p*-I	−10.0739	−0.8490	2.1269	0.18	0.19	−0.01
112	*p*-isoC$_3$H$_7$	−9.8710	−0.4421	4.6732	−0.15	−0.16	0.01
113	*p*-N(C$_3$H$_7$)$_2$	−8.5585	−0.1684	10.8017	−0.93	−1.06	0.13
114	*p*-N(CH$_2$CH$_3$)$_2$	−8.5825	−0.1794	8.1951	−0.72	−0.97	0.25
115	*p*-N(COCH$_3$)$_2$	−10.1047	−0.8553	25.0916	0.33	0.39	−0.06
116	*p*-N=C=S	−9.0358	−1.1932	10.2580	0.38	0.39	−0.01
117	*p*-N=NCN	−10.7444	−1.8136	16.1572	1.03	0.86	0.17
118	*p*-NH$_2$	−8.9061	−0.2530	5.3339	−0.66	−0.58	−0.08
119	*p*-NHCH$_3$	−8.7726	−0.2339	17.3859	−0.70	−0.66	−0.04
120	*p*-NHCOCH(CH$_3$)$_2$	−9.1670	−0.4754	19.7027	−0.10	0.01	−0.11
121	*p*-NO$_2$	−10.8966	−1.7301	10.2703	0.78	0.82	−0.04
122	*p*-O(CH$_2$)$_3$CH$_3$	−9.4196	−0.3697	9.3066	−0.32	−0.24	−0.08
123	*p*-O(CH$_2$)$_4$CH$_3$	−9.4073	−0.3665	11.6519	−0.34	−0.24	−0.10
124	*p*-OCH(CH$_3$)$_2$	−9.3733	−0.3408	10.7344	−0.45	−0.31	−0.14
125	*p*-OCH$_2$CH=CH$_2$	−9.4687	−0.4105	10.2703	−0.25	−0.11	−0.14
126	*p*-OCH$_2$Cl	−9.7421	−0.5968	6.6157	0.08	0.04	0.04
127	*p*-P(CH$_2$CH$_3$)$_2$	−8.8679	−0.5667	9.9939	0.13	0.04	0.09
128	*p*-P(isopropyl)$_2$	−9.0196	−0.5753	13.3669	0.06	0.06	0.00
129	*p*-P(O)(C$_4$H$_9$)$_2$	−10.3666	−1.1252	20.2868	0.49	0.44	0.05
130	*p*-P(O)(isopropyl)$_2$	−10.3571	−1.1002	20.3547	0.41	0.43	−0.02
131	*p*-P(O)CL$_2$	−11.0400	−1.9307	18.4922	0.90	0.92	−0.02
132	*p*-P(O)F$_2$	−10.8619	−1.6765	29.4595	0.89	0.85	0.04
133	*p*-P(O)[C(CH$_3$)$_3$]$_2$	−10.2819	−1.0509	10.3960	0.41	0.49	−0.08
134	*p*-P(S)(CH$_2$CH$_3$)$_2$	−9.2725	−1.6286	2.4909	0.46	0.46	0.00
135	*p*-PH$_2$	−9.4901	−0.6923	15.1351	0.05	0.07	−0.02
136	*p*-PO(CH$_2$CH$_3$)$_2$	−10.5619	−1.1323	20.9143	0.47	0.49	−0.02
137	*p*-P[N(CH$_3$)$_2$]$_2$	−8.3792	−0.5334	15.4100	0.25	0.16	0.09
138	*p*-S(O)OCH$_3$	−9.9882	−1.0981	6.4635	0.54	0.43	0.11
139	*p*-SCH(CH$_3$)$_2$	−8.4909	−0.5542	7.8914	0.07	0.08	−0.01
140	*p*-SCH$_2$CH=CH$_2$	−8.5031	−0.5846	12.3271	0.12	0.17	−0.05
141	*p*-SCH$_2$F	−8.8834	−0.7198	3.5504	0.20	0.19	0.01
142	*p*-SCH$_3$	−8.5229	−0.5945	6.7309	0.00	0.08	−0.08
143	*p*-SCH=CH$_2$	−8.4274	−0.7707	21.5689	0.24	0.26	−0.02
144	*p*-SCHF$_2$	−9.1636	−1.0578	6.4966	0.37	0.37	0.00
145	*p*-SCl	−8.8617	−1.6600	11.0811	0.48	0.36	0.12
146	*p*-SCN	−9.0747	−1.0057	7.3939	0.52	0.39	0.13
147	*p*-SN(CH$_3$)$_2$	−8.3531	−0.4613	21.0811	0.09	0.01	0.08
148	*p*-SO$_2$CH$_2$CH$_3$	−10.5873	−1.4138	22.7130	0.77	0.72	0.05
149	*p*-SO$_2$Cl	−10.9685	−2.3606	28.1405	1.11	1.19	−0.08
150	*p*-SO$_2$F	−10.7927	−1.6724	22.4324	0.91	0.91	0.00
151	*p*-SO$_2$NH$_2$	−10.4772	−1.2956	11.3514	0.60	0.67	−0.07

model which strongly correlates with experimental σ values for the 150-member training set:

$n = 150$, $r^2 = 0.958$, $r = 0.979$, $s = 0.073$, F = 657.6, PRESS = 0.371, PRESS/SSY = 0.010

1120 *J. Chem. Inf. Comput. Sci., Vol. 40, No. 5, 2000*

SULLIVAN ET AL.

$$\sigma = -2.480(\Delta N) - 7.894(\Delta E) - 0.605\left[\frac{D_X}{D_H}\left(\frac{EA_H}{EA_X}\right)\right] +$$
$$0.009[\phi S_X] + 0.028\left(\sum \pi\right) + 0.279 \quad (20)$$

A plot of calculated constants obtained from eq 20 versus experimental constants is shown in Figure 1. As is often noted, a strong collinearity exists between Hammett σ_m and σ_p constants which limits the differentiation between different sources of substituent effects on the basis of observed effects at the *meta* and *para* positions. The extent of such intercorrelation was demonstrated by Hansch et al.,[51] who reported an $r^2 = 0.89$ between σ_m and σ_p values for 530 experimental Hammett constants. The correlation between σ_m and σ_p constants for the 75 substituents considered for this study is also significant ($r^2 = 0.86$). However, independent analyses of the 75 *meta* and 75 *para* substituents as functions of the same descriptors employed in the composite analysis showed that correlations derived for σ_m and σ_p did not differ qualitatively from the correlation obtained for the combined set. For *meta* substituents we found

$$\sigma = -1.156(\Delta N) - 6.957(\Delta E) - 0.611\left[\frac{D_X}{D_H}\left(\frac{EA_H}{EA_X}\right)\right] +$$
$$0.008[\phi S_X] + 0.020\left(\sum \pi\right) + 0.494 \quad (21)$$

$$n = 75, r^2 = 0.943, r = 0.971, s = 0.069, F = 232.0$$

and for *para* substituents we obtained

$$\sigma = -2.588(\Delta N) - 8.796(\Delta E) - 0.474\left[\frac{D_X}{D_H}\left(\frac{EA_H}{EA_X}\right)\right] +$$
$$0.008[\phi S_X] + 0.032\left(\sum \pi\right) + 0.257 \quad (22)$$

$$n = 75, r^2 = 0.969, r = 0.984, s = 0.075, F = 436.9$$

Fitting coefficients and overall statistical parameters of the models obtained separately for *meta* and *para* groups vary little from those generated for the general model given by eq 20.

According to Wold[52] the statistical conditions inherent to a good quality regression model include a determination coefficient (r^2) > 0.90 or a correlation coefficient (r) > 0.95, residual standard deviation (s) < 0.30, and a predictive residual sum of squares (PRESS) < 0.40. The PRESS statistic is a good estimate of the real predictive error of a regression model and is given by

$$PRESS = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(1 - h_{ii})} \quad (23)$$

where y_i and $\hat{y}_i$ are the observed and predicted responses, and h_{ii} is the diagonal elements of the so-called *hat matrix*, **H**

$$\mathbf{H} = \mathbf{X}(\mathbf{X'XX})^{-1}\mathbf{X'} \quad (24)$$

where **X** is a two-column and *n*-row matrix in which the first column is filled with 1's corresponding to the intercept and the second column is the *n* predictors. **X′** refers to the transpose of **X**. When the PRESS statistic is smaller than
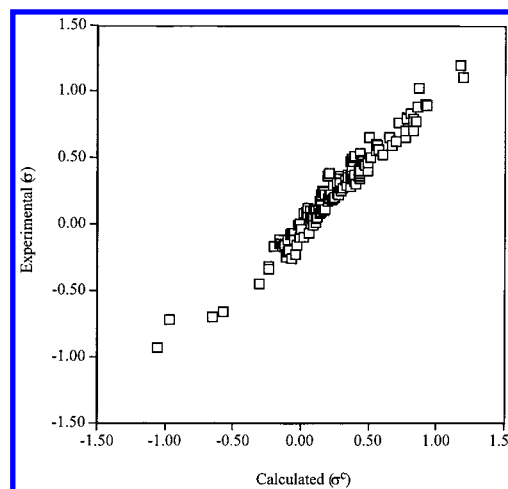

Figure 1. Plot of experimental (σ) versus calculated (σ^c) constants from eq 20 for a training set of 150 *meta* and *para*-substituted benzoic acids ($r = 0.979$).

the SSY (sum of the squares of the responses), then the model predicts better than chance. The ratio PRESS/SSY is used to calculate approximate confidence intervals of predictions of new responses. A model equation generating a value of this ratio equal to or smaller than 0.10 is an indication of an excellent model.[49] The *F*-test statistic has also been employed to determine if the predicted data actually follow the curve of interest or if the apparent fit is due only to chance. If the calculated *F*-value is larger than the *F*-value obtained from a suitable table, then the observed fit is not a result of random correlations. The calculated *F*-test value for eq 20 exceeds the table value at the 99.5% confidence level: $F = 789.9 > F(5, \infty, 0.005) = 3.35$.

On the basis of these criteria, the statistical results obtained for the multiple regression model (eq 20) and predicted substituent constants from Table 1 indicate that the derived regression model is of high statistical quality. For the general electronic effects model derived for this work (eq 20), $r^2 = 0.947$, $r = 0.973$, $s = 0.079$, PRESS = 0.371, and PRESS/SSY = 0.010, and F = 789.9 are each statistically significant values. In addition, the standard errors, t-values, and probabilities generated for eq 20 are indicative of a robust model (Table 2). Besides meeting the criteria for a good quality regression model, no significant collinearity (where r > 0.80 represents consequential intercorrelation) was found among the five model descriptors. Hence, the statistical evidence indicates that the derived model equation gives a good account of the experimental data, i.e., empirical electronic substituent constants.

3.2. Evaluation of Model Parameters. The extent of electron transfer (ΔN, eq 13) between interacting systems in acid−base (or nucleophilic-electrophilic) reactions appears to closely approximate the mesomeric characteristics of many substituent groups. In the ionization of benzoic acids, the electronegativity of water (base) is lower in energy than the electronegativities of the substituted acids, so electron flow is spontaneous in the direction of the acids and is driven by the difference in electronegativities between the two systems (the numerator of eq 13). The transfer of electrons is opposed by the relative hardness characteristics of the interacting systems (the denominator of eq 13). The greater the hardness of the two systems, the greater the tendency to resist the initial flow of electrons. In the congeneric series of benzoic

Table 2. Standard Error (SE), Standardized Regression Coefficients, Two-Tailed t-Test Values, Probabilities, and Upper and Lower 95% Confidence Intervals for Independent Variables in Eqs 20 ($n = 150$), 21 ($n = 75$), and 22 ($n = 75$)

variable	coeff	SE	std coeff	two-tailed t-value	probability	95% confidence (upper/lower)
			Equation 20			
intercept	0.279					
ΔE	−7.894	0.732	−0.221	10.786	0.0001	−6.45/−9.34
ΔN	−2.480	0.517	−0.127	4.794	0.0001	−1.46/−3.50
$\Sigma \pi$	0.028	0.004	0.158	7.620	0.0001	0.035/0.021
ϕS_x	0.009	0.002	0.127	5.748	0.0001	0.012/0.006
$D_X/D_H(EA_H/EA_X)$	−0.484	0.018	−0.666	26.575	0.0001	−0.45/−0.52
			Equation 21			
intercept	0.494					
ΔE	−6.957	0.959	−0.237	7.256	0.0001	−5.05/−8.87
ΔN	−1.156	0.793	−0.072	1.457	0.1496	−0.43/−2.74
$\Sigma \pi$	0.020	0.005	0.146	4.095	0.0001	0.030/0.010
ϕS_x	0.008	0.002	0.131	3.481	0.0009	0.013/0.003
$D_X/DH(EA_H/EA_X)$	−0.611	0.051	−0.645	12.084	0.0001	−0.51/−0.71
			Equation 22			
intercept	0.257					
ΔE	−8.796	1.115	−0.218	7.891	0.0001	−6.58/−11.02
ΔN	−2.588	0.790	−0.117	3.276	0.0016	−1.01/−4.16
$\Sigma \pi$	0.032	0.005	0.158	6.161	0.0001	0.043/0.022
ϕS_x	0.008	0.002	0.108	3.866	0.0002	0.012/0.004
$D_X/DH(EA_H/EA_X)$	−0.474	0.023	−0.699	20.276	0.0001	−0.46/−0.52

acids considered for this study, there is generally little variation in the hardness features of substituted acids. Thus, the effectiveness of electron transfer between water and acids is determined largely by differences in their electronegativities. The electronegativities of donor-substituted acids are less than those of acceptor-substituted acids, hence the driving force which promotes the transfer of electrons, i.e., $\chi_{base} - \chi_{acid}$, is diminished in donor systems relative to the unsubstituted case ($\chi_{donor} < \chi_H$). Conversely, the driving force is generally enhanced relative to benzoic acid in acceptor-substituted systems ($\chi_{acceptor} > \chi_H$).

These features can be quantified in terms of ΔN by letting $\Delta\Delta N = \Delta N_X - \Delta N_H$, where H = benzoic acid and X = substituted acids. This leads to negative values for substituents having donor properties and positive values for most substituents having acceptor properties. These trends closely correspond to the directional properties (signs) associated with empirical constants, which are indicative of the electron-withdrawing (+) and electron-donating (−) tendencies of substituents. Values of $\Delta\Delta N$ for training set compounds suggest that substituted acids containing donor substituents are less adept at accepting electrons from water during proton transfer than acids having acceptor groups. ΔN thus implies that the presence of donor groups on the aromatic ring of substituted benzoic acids inhibit ionization ($\Delta N_X < \Delta N_H$), while the presence of acceptor groups promote proton transfer ($\Delta N_X > N_H$). This finding is consistent with the notion that donor substituents impede reactions in which negative charge is generated at the equilibrium or transition state because such groups force excess charge or electron density toward an already electron-rich reactive site. Acceptor groups, in contrast, are able to more efficiently redistribute the increased negative charge, and such reactions are therefore enhanced.

E (eq 11) is the electron energy equivalent to ΔN, i.e., ΔE is an account of the energy lowering which occurs in a bimolecular system undergoing electron transfer. As was the case for ΔN, ΔE decreases for donor-substituted acids relative to that of benzoic acid, where $\Delta E_H = 0$. However, ΔE also decreases for acceptor-substituted acids. This results

because of two independent orbital mixing interactions that occurs when a substituent is added to the model system. First, there occurs partial transfer of electrons from the HOMOs of each of the two species (substituent and acid) to their respective LUMOs occurs (solid arrows). Second, there is a mixing of the filled molecular orbitals of each species with its own empty molecular orbitals (dashed arrows). The first effect is called delocalization, and the second effect is known as polarization.[53]
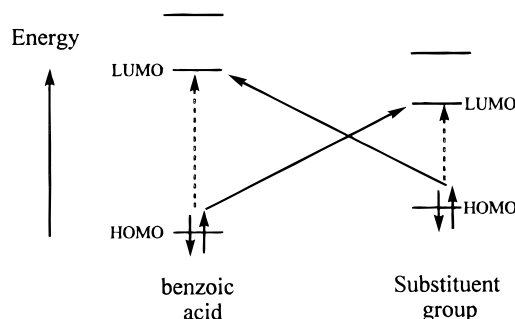
Both types of interactions lead to an overall energy-lowering, as reflected in the magnitudes of ΔE for the training set acids (all values are less than that of benzoic acid, i.e., negative). Since by definition ΔE encodes information relating to both polar and conjugative effects, it is not possible to differentiate between these effects for substituents. Consequently, it may be said only that ΔE reflects the energy lowering effect of substituents due to their combined mesomeric and inductive mechanisms. For both donor and acceptor substituents, the larger the negative number associated with ΔE, the greater its energy lowering propensity.

Values of $D_X/D_H(EA_H/EA_X)$ (from eq 14) appear to encode information connected with the proportion of electron-donating and electron-withdrawing character inherent to a given substituent group. The larger the magnitude of $D_X/D_H(EA_H/EA_X)$, the greater a group's propensity to act as an electron donor. Smaller calculated values of $D_X/D_H(EA_H/EA_X)$ are characteristic of electron acceptors. This trend is easier to observe if relative rather than absolute values are

used. For example, using unsubstituted benzoic acid as a zero reference (X=H), the difference $D_H/D_H(EA_H/EA_H)-D_X/D_H-(EA_H/EA_X) = 1-D_X/D_H(EA_H/EA_X)$ provides a relative scale of the donor or acceptor properties of a given substituent, X, that is comparable to that obtained empirically from experimental results. This is because LUMO energies of substituted benzoic acids and thus energy values of EA (E_{EA}) are normally greater than that of benzoic acid for donor groups and lower than that of benzoic acid for acceptor groups. Consequently, $1-D_X/D_H(EA_H/EA_X) < 0$ for donor groups and $1-D_X/D_H(EA_H/EA_X) > 0$ for acceptor groups. This scale closely parallels the scale obtained empirically via the Hammett relationship, i.e., $\sigma = \log K_X - \log K_H < 0$ for donor groups and $\sigma = \log K_X - \log K_H > 0$ for acceptor groups. In almost all cases, the signs of the values obtained from the difference of these frontier orbital parameters match those of the empirical Hammett constants. This finding suggests that eq 13 is a strong, computational-based indicator of the conjugative electron-donating and electron-withdrawing properties of many substituent groups.

Although theoretical justification for the use of the expression (EA_H/EA_X) is not apparent, its relationship to the conjugative features of substituents can be demonstrated qualitatively. Swain and Lupton[54] proposed separate field and resonance values for substituents based on the proposition that σ is a linear combination of these two effects, $\sigma = f F + r R$. The values of f and r are characteristic of the reaction type and are obtained statistically (they are just the regression fitting coefficients). F (field) and R (resonance) values are given in tables and are used in a manner analogous to Hammett constants, except the same values are used for both *meta* and *para* groups. Differentiation between group position and reaction type is achieved via the f and r values (hence, these terms are sometimes called the susceptibility constants). Of course, the empirical values of σ_m and σ_p are perfectly reproduced when regressed against Hammett constants since they are used for the derivation of the F and R scales, i.e., $r = 1.00$ for each series. However, if (EA_H/EA_X) is substituted for the resonance parameter R and the training set is reanalyzed, new statistical relationships for *meta* and *para* substituents are derived. For *meta* substituents, this treatment yields the expression:

$$\log(K_X/K_H)_{meta} = \sigma_{meta} = 0.882F - 0.287(EA_H/EA_X) + 0.216 \quad (25)$$

$$n = 74, r = 0.982, s = 0.052, F = 957.1$$

For *para* subtituents a similar relationship is obtained:

$$\log(K_X/K_H)_{para} = \sigma_{para} = 0.780F - 0.431(EA_H/EA_X) + 0.37 \quad (26)$$

$$n = 74, r = 0.953, s = 0.122, F = 352.3$$

Both eqs (25) and (26) are of high statistical quality, and no significant collinearity (i.e., r > 0.80) exists between F and (EA_H/EA_X). These empirical relationships suggest that the term (EA_H/EA_X) encodes pertinent information relating to the conjugative effects of substituents. Were EA primarily associated with the inductive character of substituents, one might expect either that the relationship $(EA_H/EA_X)/D^2$ or

EA/D^2 would be a qualified descriptor, since inductive effects fall off rapidly with the square of the distance. However, such a relationship is not observed. On the other hand, delocalization increases with the length of the conjugated path. That $D_X/D_H(EA_H/EA_X)$ is a robust descriptor offers additional qualitative support for the view that this term encodes information relating to the mesomeric effect.

The electrotopological state algorithm (eq 17) identifies and characterizes specific atoms and their properties with respect to their local and global electronic and topological environment. Atoms are distinguished by their relative size and electronic configuration (atomic number), by their hybrid state (valence electrons) and by their local topology (degree of connectivity or number of bonds to hydrogen and non-hydrogen atoms). Thus, electrotopological state values, S_i, for atoms tend to be large and positive for atoms which possess π or lone pair electrons, which lie on the periphery of a molecule, or which are located at a terminal position. Atoms located at interior positions or which do not contain π or lone pair electrons tend to have smaller and sometimes negative values.[55] Values of ϕS_X (eq 19) serve as general indicator of the size and, particularly, polarity of substituent groups. The magnitude ϕS_X is indicative of a given substituents unique structural and electronic environment, and seems to reflect group electronegativity. Highly polar substituent groups tend to have large calculated values of ϕS_X, and values increase rapidly with its relative electronegativity, e.g., $\phi S_X(CH_2F) > \phi S_X(CH_2Cl) > \phi S_X(CH_2I)$. Similarly, values rise quickly as the relative size $[\phi S_X(CH_3) > \phi S_X-(C_2H_5) > \phi S_X(C_3H_7) > \phi S_X(C_4H_9)]$ and degree of branching $[\phi S_X(SCH_3) > \phi S_X(SCH_2(CH3)_2]$ of substituents increase. Unfortunately, the electrotopological state index does not encode a directional component to these effects (all values of ϕS_X are positive and thus unidirectional) and therefore does not accurately reflect the overall inductive properties of substituents. Nonetheless, the values of ϕS_X for many substituent groups tend to follow anticipated trends with respect to characteristics associated with group polarity and electronegativity.

3.3. External Validation. *External Validation Using an Independent Testing Set of Benzoic Acids.* The true measure of the usefulness of a QSAR/QSPR model is its ability to predict the values of compounds not included in its derivation with reasonable accuracy. Therefore, the model derived for the present work was used to calculate σ values for an independent testing set of 124 additional benzoic acids. These compounds were chosen randomly from among those acids having *meta* and *para* substituent groups whose empirical constants fell with the linear range of the model (eq 20), i.e., approximately $-1.0 < \sigma < 1.3$. Experimental Hammett constants, σ, for each acid were obtained from the literature[29] and computational-based constants, σ^c, were calculated directly from eq 20. Calculated MO and topological parameters for the independent variables were acquired in the exact manner in which they were obtained for the training set compounds. Computational-based substituent constants were then analyzed as functions of empirical constants. Results of this comparative analysis, shown in Table 3, indicate that computational-based *meta* and *para* analogues estimated from eq 20 are generally in good agreement with empirical values ($r = 0.88, s = 0.148$). For 75 of the 124 testing set compounds (60.5%), the deviation between experimental (σ)

Table 3. Experimental and Calculated Group Constants and Residual Values Used for External Validation of an Independent Testing Set Consisting of 126 *Meta* and *Para*-Substituted Benzoic Acids[a]

	meta-substituent	calcd	exptl	resid		para-substituent	calcd	exptl	resid
1	m-C(CH$_2$CH$_3$)(CH$_3$)$_2$	−0.07	−0.06	−0.01	64	p-C(CH$_2$CH$_2$)(CH$_3$)$_2$	−0.16	−0.18	0.02
2	m-C(CH$_3$)(NO$_2$)$_2$	1.16	0.54	0.62	65	p-C(CH$_3$)(NO$_2$)$_2$	1.18	0.61	0.57
3	m-C(Et)(NO$_2$)$_2$	1.14	0.56	0.58	66	p-C(Et)(NO$_2$)$_2$	1.17	0.64	0.53
4	m-C(NO$_2$)(CH$_3$)$_2$	0.58	0.18	0.40	67	p-C(NO$_2$)(CH$_3$)$_2$	0.61	0.20	0.41
5	m-C$_3$H$_7$	−0.04	−0.06	0.02	68	p-C$_3$H$_7$	−0.12	−0.13	0.01
6	m-CCL$_3$	0.50	0.40	0.10	69	p-CCL$_3$	0.46	0.46	0.00
7	m-CH$_2$C(CH$_3$)$_3$	−0.02	−0.05	0.03	70	p-CH$_2$C(CH$_3$)$_3$	−0.10	−0.17	0.07
8	m-CH$_2$C(OH)(CH$_3$)$_2$	0.17	−0.16	0.33	71	p-CH$_2$C(OH)(CH$_3$)$_2$	0.08	−0.17	0.25
9	m-CH$_2$CF$_3$	0.47	0.12	0.35	72	p-CH$_2$CF$_3$	0.48	0.09	0.39
10	m-CH$_2$CH(CH$_3$)$_2$	−0.03	−0.07	0.04	73	p-CH$_2$CH(CH$_3$)$_2$	−0.02	−0.12	0.10
11	m-CH$_2$CH=CH$_2$	0.05	−0.11	0.16	74	p-CH$_2$CH=CH$_2$	−0.02	−0.14	0.12
12	m-CH$_2$CN	0.38	0.16	0.22	75	p-CH$_2$CN	0.34	0.18	0.16
13	m-CH$_2$CONH$_2$	0.38	0.06	0.32	76	p-CH$_2$CONH$_2$	0.35	0.07	0.28
14	m-CH$_2$N(CH$_3$)$_2$	0.12	0.00	0.12	77	p-CH$_2$N(CH$_3$)$_2$	−0.09	0.01	−0.10
15	m-CH$_2$NHCOCH$_3$	0.20	0.05	0.15	78	p-CH$_2$NHCOCH$_3$	0.32	−0.05	0.37
16	m-CH$_2$OCOCH$_3$	0.20	0.04	0.16	79	p-CH$_2$OCOCH$_3$	0.19	0.05	0.14
17	m-CH=C(CN)$_2$	0.90	0.66	0.24	80	p-CH=C(CN)$_2$	0.95	0.84	0.11
18	m-CH=NNHCONHNH$_2$	0.48	0.22	0.26	81	p-CH=NNHCONHNH$_2$	0.49	0.16	0.33
19	m-CH=NNHCSNH$_2$	0.50	0.45	0.05	82	p-CH=NNHCSNH$_2$	0.43	0.40	0.03
20	m-CH=NOCH$_3$	0.26	0.37	−0.11	83	p-CH=NOCH$_3$	0.29	0.30	−0.01
21	m-CHBr$_2$	0.28	0.31	−0.03	84	p-CHBr$_2$	0.31	0.32	−0.01
22	m-CHCN$_2$	0.72	0.53	0.19	85	p-CHCN$_2$	0.70	0.52	0.18
23	m-COCH(CH$_3$)$_2$	0.38	0.38	0.00	86	p-COCH(CH$_3$)$_2$	0.49	0.48	0.01
24	m-COCH$_2$CH$_3$	0.42	0.38	0.04	87	p-COCH$_2$CH$_3$	0.46	0.48	−0.02
25	m-COCH$_3$	0.50	0.38	0.12	88	p-COCH$_3$	0.46	0.50	−0.04
26	m-COF	0.68	0.55	0.13	89	p-COF	0.73	0.70	0.03
27	m-CONHCH$_3$	0.37	0.35	0.02	90	p-CONHCH$_3$	0.38	0.36	0.02
28	m-COOCH$_2$CH$_3$	0.48	0.37	0.11	91	p-COOCH$_2$CH$_3$	0.52	0.45	0.07
29	m-CSNH$_2$	0.40	0.25	0.15	92	p-CSNH$_2$	0.38	0.30	0.08
30	m-N(CH$_3$)NO$_2$	0.61	0.49	0.12	93	p-N(CH$_3$)NO$_2$	0.62	0.61	0.01
31	m-N=C=O	0.45	0.27	0.18	94	p-N=C=O	0.29	0.19	0.10
32	m-N=C=O	0.41	0.21	0.20	95	p-N=C=O	0.44	0.13	0.31
33	m-N=CCL$_2$	−0.10	−0.24	0.14	96	p-N=CCL$_2$	−0.75	−0.61	−0.14
34	m-NHCH$_2$CH$_3$	0.41	0.17	0.24	97	p-NHCH$_2$CH$_3$	0.37	−0.03	0.40
35	m-NHCOCH$_2$Cl	0.40	0.24	0.16	98	p-NHCOCH$_2$Cl	0.37	0.12	0.25
36	m-NHCSCH$_3$	0.29	0.22	0.07	99	p-NHCSCH$_3$	0.38	0.16	0.22
37	m-NHCSNH$_2$	0.43	0.30	0.13	100	p-NHCSNH$_2$	0.39	0.07	0.32
38	m-NHCSNHEt	0.59	0.38	0.21	101	p-NHCSNHEt	0.58	0.35	0.23
39	m-OCF$_3$	0.07	0.10	−0.03	102	p-OCF$_3$	−0.22	−0.25	0.03
40	m-OCH$_2$CH$_2$CH$_3$	−0.02	0.10	−0.12	103	p-OCH$_2$CH$_2$CH$_3$	−0.22	−0.24	0.02
41	m-OCH$_2$CH$_3$	0.07	0.12	−0.05	104	p-OCH$_2$CH$_3$	−0.06	−0.27	0.21
42	m-OCH$_3$	0.58	0.67	−0.09	105	p-OCH$_3$	0.53	0.54	−0.01
43	m-OCN	0.65	0.39	0.26	106	p-OCN	0.62	0.36	0.26
44	m-OSO$_2$CH$_3$	0.10	0.03	0.07	107	p-OSO$_2$CH$_3$	0.05	0.06	−0.01
45	m-P(CH$_3$)$_2$	0.48	0.43	0.05	108	p-P(CH$_3$)$_2$	0.49	0.49	0.00
46	m-P(O)(N(CH$_3$)$_2$)$_2$	0.43	0.30	0.13	109	p-P(O)(N(CH$_3$)$_2$)$_2$	0.54	0.40	0.14
47	m-P(O)(OC$_4$H$_9$)$_2$	0.72	0.41	0.31	110	p-P(O)(OC$_4$H$_7$)$_2$	0.70	0.57	0.13
48	m-P(OH)$_2$	0.24	0.36	−0.12	111	p-P(OH)$_2$	0.24	0.42	−0.18
49	m-PCL$_2$	0.40	0.54	−0.14	112	p-PCL$_2$	0.43	0.61	−0.18
50	m-PF$_2$	0.28	0.49	−0.21	113	p-PF$_2$	0.27	0.59	−0.32
51	m-PO(OCH$_2$CH$_3$)$_2$	0.67	0.55	0.12	114	p-PO(OCH$_2$CH$_3$)$_2$	0.70	0.60	0.10
52	m-SCCH	0.40	0.26	0.14	115	p-SCCH	0.36	0.19	0.17
53	m-SCH$_2$CH$_3$	0.17	0.18	−0.01	116	p-SCH$_2$CH$_3$	0.10	0.12	−0.02
54	m-SCH=CHCl	0.39	0.31	0.08	117	p-SCH=CHCl	0.35	0.20	0.15
55	m-SCOCH$_3$	0.29	0.39	−0.10	118	p-SCOCH$_3$	0.28	0.44	−0.16
56	m-SF$_3$	0.48	0.70	−0.22	119	p-SF$_3$	0.49	0.80	−0.31
57	m-SH	0.18	0.25	−0.07	120	p-SH	0.13	0.15	−0.02
58	m-SO$_2$CH$_3$	0.71	0.60	0.11	121	p-SO$_2$CH$_3$	0.76	0.72	0.04
59	m-SO$_2$CN	1.07	1.10	−0.03	122	p-SO$_2$CN	1.12	1.26	−0.14
60	m-SO$_2$N(CH$_3$)$_2$	0.67	0.51	0.16	123	p-SO$_2$N(CH$_3$)$_2$	0.72	0.65	0.07
61	m-SO$_2$NH$_2$	0.67	0.53	0.14	124	p-SO$_2$NH$_2$	0.71	0.60	0.11
62	m-SOCH$_3$	0.33	0.52	−0.19	125	p-SOCH$_3$	0.35	0.49	−0.14
63	m-SOCHF$_2$	1.08	0.54	0.54	126	p-SOCHF$_2$	0.73	0.58	0.15

[a] Predicted values were obtained using eq 20.

and predicted (σ^c) values is 0.15 log units or less, and for 24 compounds the deviation is between 0.15 and 0.25 (19.3%). The average deviation (average residual value) from empirical values for the entire 124-member testing set is 0.10 log units. In terms of positional parameters, statistical results are considerably better for *para*-substituents ($r = 0.86$, s = 0.174) than those obtained for *meta* substituents ($r = 0.80$, $s = 0.149$).

Among the 25 compounds having deviations greater than 0.25, the weakest overall correlations are obtained for those

substituents having unusually strong electron-withdrawing character. In particular, predicted values of σ obtained for substituents containing fluorine atoms bound to a highly polarizable atom or group (e.g., $-SOCHF_2$, $-SF_3$, $-PF_2$) and for both *meta* and *para* substituents having one or more nitro groups bonded to a tertiary central carbon atom (e.g., $-C(CH_3)(NO_2)_2$, $-C(NO_2)(CH_3)_2$) deviate significantly from empirical values. In the former case, the model consistently underestimates σ values and in the latter case, σ values are uniformly overestimated. The disagreement of theoretically calculated σ constants with experimental values for such groups is well-known and has already been reported.[56] The often limited quality of experimental data from which empirical constants have been derived may contribute significantly to observed differences between experimental and computational-based constants. Many benzoic acids are not sufficiently water-soluble for experimental measurements, so various mixed solvent systems have been employed for the acquisition of ionization data. Depending upon the solvent system from which they were obtained, reported highest and lowest values of some constants may thus differ considerably. For example, σ_{meta} for 2-pyridyl has been reported to be 0.17 and 0.49, σ_{para} for nitroso has been reported at 0.12 and 1.46, and σ_{para} for acetoxy has been reported to be 0.15 and 0.41.[29]

The description of substituent effects on the ionization of benzoic acids has been oversimplified by basing calculations on the unsolvated neutral form of the molecules. Consequently, the effect of solvation or the influence of substituents on the relative stability of the benzoate ion have not been explicitly considered. This may, to some degree, explain the poor theoretical description of some strong acceptor groups. Such groups would certainly have a different stabilizing effect on the neutral acid than on the solvated anion. The greater the electron-withdrawing character of the substituent, the greater the potential discrepancy between observed and estimated values. In addition, predicted values are obtained without consideration of possible solvent-substituent hydrogen-bonding interactions, so some degree of anomalous behavior is to be expected for susceptible groups such as alcohols, amides, and amines. Indeed, predicted σ values for amines, especially *para*-substituted secondary amines containing an alpha carbonyl or thiocarbonyl group (e.g., $-NHC(=O)-CH(CH_3)_2$, $-NHC(=S)-NHCH_2CH_3$), are among those substituents exhibiting the largest deviations from empirical values.

Overall, however, external validation results show that eq 20 offers a creditable methodology for estimating computational-based analogues of empirical Hammett constants, especially considering that descriptors were obtained directly from simple electronic structure and topological information. It is assumed, as it was with the original empirical Hammett equation, that predicted constants derived from benzoic acid equilibria also provide a valid means for the analysis of substituent effects occurring in similar systems which contain a reactive side-chain attached to an aromatic core. To investigate the validity of this supposition, a wide range of experimental reaction data for several types of aromatic substrates were analyzed as functions of empirical (σ) and predicted (σ^c) substituent constants and subsequent correlations compared in terms of statistical parameters.

Correlations with Reaction Data. In choosing datasets for validation objectives, we have attempted to include case studies that demonstrate and reflect the widespread applicability of the Hammett equation. To this end, two fundamental dataset selection rules were followed: (1) selected substrate must contain an aromatic core region capable of transmitting substituent-induced perturbations to a reaction center in a side-chain and (2) selected reaction type or chemical process must respond to systematic changes in substituent structure. These criteria were satisfied for each of the 30 datasets analyzed, and the subsequent empirical and computational-based relationships derived for these systems are presented in terms of statistical parameters for comparative examination in Table 4. Validated case studies are divided into four categories, each category consisting of several datasets correlated with both standard and computational parameter sets. Categories include (1) thermodynamic ionization constants (datasets 1−10),[57−65] (2) kinetic rate constants (datasets 11−22),[66−71] (3) ^1H and ^{12}C NMR chemical shifts and IR C=O stretching frequencies (datasets 23−27),[72,73] and (4) biological reaction data (datasets 28−30).[74−76] Many of the datasets contain a number of identical common substituent groups (particularly CH_3, C_2H_5, *iso*C_3H_7, NH_2, CN, NO_2, Br, Cl, I, COOH, $COCH_3$, $COOCH_3$), which reflects the limited availability of reaction and physicochemical property data for meaningful (n > 6) and structurally diverse sets of *meta*- and *para*-substituted aromatic substrates in the literature. Nonetheless, sufficient heterogeneity exists in group and substrate variability and in reaction type to demonstrate and corroborate the statistical correspondence observed between results obtained from empirical and computational-based approaches.

Generally, it was found that computational-based constants yield correlations exhibiting somewhat more scatter than those obtained empirically, although in some cases predicted constants were found to be more highly correlated with reaction data or to exhibit less scatter than their empirical counterparts (see, for example, Table 4, datasets 5, 6, 15, 20, 23, 25). Assuming a typical maximal experimental error of $\pm 2.5\%$ (corresponding to a probability of 95%) in reported reaction data, statistical results show that predicted and empirical substituent constants perform equally in most of the case studies investigated. However, both the empirical and computational-based models tend to underestimate the values of certain *para* substituents capable of strongly opposing inductive and mesomeric interactions with the acidic side-chain. Such common substituent groups as hydroxy ($-OH$), alkoxy ($-OR$), secondary amines ($-NH-[R]$), and tertiary amines ($-N[R]_2$) are characterized by such behavior. This is because electron-donating groups having lone pairs on the substituent atom attached to the aromatic ring are able to conjugate favorably with the reaction center in benzoic acids, as illustrated below by resonance hybrid **2**.
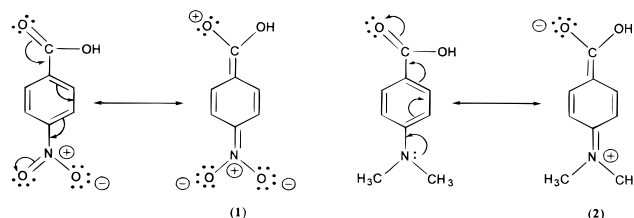

(1)   (2)

Table 4. Statistical Comparison of Empirical (emp) and Computational-Based (comp) LFER Models Derived from Empirical (σ) and Calculated (σ^c) Substituent Constants as Functions of Ionization, Kinetic, and Spectroscopic Reaction Data and for a Wide Variety of Congeneric Reaction Series of *Meta* and *Para*-Substituted Aromatic Systems. Statistical Parameters Include the Number of Testing Set Compounds (n), Correlation Coefficient (r), Residual Error (s), F-test Value (F), and the Reaction Constant (ρ)

system		model	n	r	s	model eq	F	ρ	ref
1	ionization of benzoic acids in	emp	29	0.996	0.008	$\log(K_X/K_H) = 0.590\sigma + 0.01$	3463.7	0.590	57–60
	80% methyl cellosolve	comp	29	0.977	0.019	$\log(K_X/K_H) = 0.567\sigma^c + 0.05$	562.7	0.567	57–60
2	ionization of benzoic acids in	emp	32	0.989	0.013	$\log(K_X/K_H) = 0.681\sigma - 0.01$	1389.5	0.681	57–60
	50% ethanol	comp	32	0.970	0.21	$\log(K_X/K_H) = 0.647\sigma^c + 0.02$	480.9	0.647	57–60
3	ionization of thiophenols in	emp	16	0.977	0.210	$\log(K_X/K_H) = 2.620\sigma + 0.19$	296.1	2.620	59
	48% ethanol	comp	16	0.977	0.212	$\log(K_X/K_H) = 2.874\sigma^c + 0.02$	287.9	2.874	59
4	ionization of 3′ and 4′- biphenyl- 4-	emp	17	0.919	0.076	$\log(K_X/K_H) = 0.443\sigma - 0.01$	81.0	0.443	60
	carboxylic acids in 70% ethanol	comp	17	0.920	0.075	$\log(K_X/K_H) = 0.477\sigma^{c-}0.02$	82.2	0.477	60
5	ionization of β-phenylpropionic	emp	9	0.993	0.035	$\log(K_X/K_H) = 0.686\sigma - 0.01$	531.7	0.686	68
	acids in 50% ethanol	comp	9	0.964	0.015	$\log(K_X/K_H) = 0.671\sigma - 0.02$	92.9	0.671	68
6	ionization of *cis*-2-phenylcyclopropane-	emp	7	0.965	0.047	$\log(K_X/K_H) = 0.426\sigma - 0.00$	68.5	0.426	61
	carboxylic acids in 50% ethanol	comp	7	0.989	0.027	$\log(K_X/K_H) = 0.480\sigma^c - 0.02$	214.7	0.480	61
7	ionization of *trans*-2-phenylcyclopropane-	emp	10	0.985	0.033	$\log(K_X/K_H) = 0.483\sigma - 0.02$	259.6	0.483	61
	carboxylic acids in 50% ethanol	comp	10	0.993	0.023	$\log(K_X/K_H) = 0.510\sigma - 0.01$	556.8	0.510	61
8	ionization of substituted pyridines in	emp	30	0.984	0.345	$\log(K_X/K_H) = 5.541\sigma - 0.08$	842.8	5.541	62–64
	50% ethanol	comp	30	0.964	0.509	$\log(K_X/K_H) = 5.974\sigma^c - 0.17$	371.6	5.974	62–64
9	ionization of substituted benzene	emp	14	0.958	0.111	$\log(K_X/K_H) = 1.033\sigma - 0.01$	489.9	1.033	65
	phosphonic acids in 50% ethanol	comp	14	0.969	0.096	$\log(K_X/K_H) = 1.100\sigma^c - 0.06$	320.1	1.100	65
10	ionization of substituted benzene	emp	17	0.985	0.074	$\log(K_X/K_H) = 0.931\sigma + 0.06$	489.9	0.931	65
	phosphonic acids in water	comp	17	0.977	0.091	$\log(K_X/K_H) = 0.987\sigma^c + 0.01$	320.1	0.987	65
11	alakline hydrolysis of *O,O*-diethylphenyl	emp	16	0.990	0.058	$\log(k_X/k_H) = 0.635\sigma - 0.04$	460.9	0.635	66
	phosphates at 25 °C	comp	16	0.980	0.066	$\log(k_X/k_H) = 0.589\sigma^c - 0.02$	304.9	0.589	66
12	hydrolysis of phenyl acetates at pH 9.1 in	emp	14	0.980	0.068	$\log(k_X/k_H) = 0.854\sigma - 0.06$	324.5	0.854	66
	0.1 M NaOH/H$_3$BO$_3$/KCl buffer	comp	14	0.980	0.063	$\log(k_X/k_H) = 0.797\sigma^c - 0.04$	334.1	0.797	66
13	hydrolysis of phenyl *N,N*-dimethyl	emp	12	0.980	0.078	$\log(k_X/k_H) = 0.892\sigma - 0.07$	222.4	0.892	66
	carbamates at pH 13.9 in 0.9 N NaOH	comp	12	0.970	0.089	$\log(k_X/k_H) = 0.814\sigma - 0.04$	143.8	0.814	66
14	alkaline hydrolysis of phenyl *N*-methyl	emp	12	0.970	0.096	$\log(k_X/k_H) = 0.279\sigma + 0.01$	144.4	0.279	66
	carbamates at 38 °C	comp	12	0.950	0.106	$\log(k_X/k_H) = 0.258\sigma^c + 0.02$	99.7	0.258	66
15	alkaline hydrolysis of substituted aryl	emp	12	0.973	0.166	$\log(k_X/k_H) = 1.644\sigma + 0.09$	179.3	1.644	67
	tosylates at 25 °C	comp	12	0.980	0.145	$\log(k_X/k_H) = 1.736\sigma^c + 0.02$	238.8	1.736	67
16	ionization of phenylpropiolic acids at	emp	9	0.993	0.031	$\log(k_X/k_H) = 0.686\sigma - 0.01$	531.7	0.686	68
	24 °C in 50% aqueous ethanol	comp	9	0.964	0.072	$\log(k_X/k_H) = 0.671\sigma^c - 0.02$	92.9	0.671	68
17	rates of solvolysis of phenylpropiolic acid	emp	9	0.983	0.057	$\log(k_X/k_H) = 0.768\sigma + 0.05$	201.8	0.768	68
	methyl esters in 50% ethanol	comp	9	0.989	0.046	$\log(k_X/k_H) = 0.778\sigma + 0.03$	309.6	0.778	68
18	methanolysis rates of substituted	emp	12	0.996	0.083	$\log(k_X/k_H) = 2.634\sigma + 0.02$	1339.1	2.634	69
	1-menthyl benzoates at at 30 °C	comp	12	0.968	0.240	$\log(k_X/k_H) = 2.723\sigma^c + 0.02$	149.9	2.723	69
19	methanolysis rates of substituted	emp	12	0.996	0.080	$\log(k_X/k_H) = 2.487\sigma + 0.03$	1263.8	2.487	69
	1-menthyl benzoates at 50 °C	comp	12	0.968	0.228	$\log(k_X/k_H) = 2.571\sigma^c + 0.03$	148.8	2.571	69
20	hydrolysis of benzoic anhydrides in	emp	9	0.994	0.138	$\log(k_X/k_H) = 3.124\sigma + 0.01$	606.3	3.124	70
	75% dioxane-25% water at 58 °C	comp	9	0.981	0.248	$\log(k_X/k_H) = 3.167\sigma^c - 0.13$	182.6	3.167	70
21	hydrolysis of benzoic anhydrides in	emp	9	0.983	0.149	$\log(k_X/k_H) = 1.954\sigma - 0.03$	203.6	1.954	70
	75% dioxane-25% water at 80 °C	comp	9	0.947	0.262	$\log(k_X/k_H) = 1.934\sigma^c - 0.10$	60.9	1.934	70
22	addition of styrene to 9-substituted	emp	9	0.990	0.077	$\log(k_X/k_H) = 1.666\sigma - 0.05$	359.1	1.666	71
	acridizinium ions in DMSO at 65 °C	comp	9	0.990	0.077	$\log(k_X/k_H) = 1.638\sigma^c - 0.09$	353.7	1.638	71
23	¹H NMR chemical shifts (δ) of the	emp	7	0.967	0.904	δ (¹H) $= 6.82\sigma + 167.73$	70.9	6.820	72
	2-methyl protons in benzoselenazoles	comp	7	0.980	0.705	δ (¹H) $= 7.13\sigma^c + 167.73$	119.6	7.128	72
24	¹H NMR chemical shifts (δ) of the	emp	7	0.956	1.136	δ (¹H) $= 7.44\sigma + 166.23$	53.4	7.440	72
	2-methyl protons in benzothiazoles	comp	7	0.977	0.825	δ (¹H) $= 7.84\sigma^c + 166.23$	105.9	7.840	72
25	¹H NMR chemical shifts (δ) of the	emp	7	0.948	1.192	δ (¹H) $= 7.09\sigma + 154.60$	44.0	7.088	72
	2-methyl protons in benzoxnazoles	comp	7	0.972	0.875	δ (¹H) $= 7.50\sigma^c + 154.59$	85.94	7.500	72
26	C=O IR stretching frequencies of	emp	8	0.989	0.288	$\bar{v}$ C=O $= 5.456\sigma + 1709.3$	261.4	5.456	73
	1,3-indandione derivatives in CCl$_4$	comp	8	0.973	0.682	$\bar{v}$ C=O $= 5.212\sigma^c + 1709.4$	126.5	5.212	73
27	C=O IR stretching frequencies of	emp	9	0.988	0.297	$\bar{v}$ C=O $= 6.100\sigma + 1703.4$	246.0	6.100	73
	1,3-indandione derivatives in CHCl$_3$	comp	9	0.960	0.835	$\bar{v}$ C=O $= 6.174\sigma^c + 1703.3$	81.9	6.174	73
28	rates of cholesterol esterase inhibition by	emp	7	0.988	0.204	$\log(k_X/k_H) = 3.097\sigma + 0.03$	196.6	3.097	74
	phenyl-*N*-butylcarbamates	comp	7	0.968	0.325	$\log(k_X/k_H) = 3.132\sigma^c - 0.16$	74.4	3.132	74
29	potencies of substituted	emp	10	0.992	0.048	pI$_{50} = 0.602\sigma + 0.601\pi + 0.64$	229.3	0.602	75
	O-phenyl-DL-homoserines on ATP	comp	10	0.975	0.087	pI$_{50} = 0.583\sigma^c + 0.663\pi + 1.61$	67.17	0.583	75
30	potencies of *N,N*−bis(2-chloroethyl)anilines	emp	13	0.987	0.175	pIC$_{50} = 2.518\sigma + 0.21$	422.3	2.518	76
	vs cultured tumor cells	comp	13	0.979	0.225	pIC$_{50} = 2.634\sigma^c + 0.12$	250.3	2.634	76

However, most donor groups, being more electronegative than carbon, also possess an electron-withdrawing component that is transmitted via inductive polarization in the direction opposite that of the electron-donating mesomeric effect. When both of these effects are strong, such groups are commonly observed outside the linear range of the Hammett equation and its computational-based analogue. For electron-withdrawing substituent groups in the benzoic acid system,

such as −NO$_2$, −COOR, −SO$_2$R, etc., conjugative interactions are largely restricted to the benzene ring, since the resonance structure **1** resulting from mesomeric interaction of an acceptor group with the acid function is highly unfavorable. Consequently, even those acceptor groups capable of strong interaction with the benzene ring are generally well-behaved in benzoic acids. Among common donor groups, the *p-N,N*-dimethylamino group is, in par-

1126 *J. Chem. Inf. Comput. Sci., Vol. 40, No. 5, 2000*

SULLIVAN ET AL.

ticular, well-known for its recalcitrant properties and is often observed as an outlier in Hammett-type studies. In the current investigation, the *p-N,N*-dimethylamino group was eliminated from all reaction datasets prior to their statistical evaluation to ensure the most viable correlations. Although some hydroxy, alkoxy, and amino substituents other than *N,N*-dimethylamino were retained for the evaluation of experimental datasets in this study, their presence in a given dataset often had a noticeable impact on the statistical quality of resulting correlations. For example, empirical and computational-based parameters yield correlation coefficients of 0.985 and 0.977, respectively, for the ionization of benzenephosphonic acids in water (dataset 10). Removal of the hydroxy, *N*-methylamino, and amino groups from the dataset result in improved correlation coefficients of 0.996 and 0.992.

Values of the reaction constant ρ obtained for empirical and computational-based correlations are generally amenable in direction and magnitude. The reaction constant is the slope of log k or K against σ and is a measure of the sensitivity of a reaction to the effects of electronic perturbation. The magnitude of ρ provides a scale for determining the degree to which a reaction responds to the presence of given substituents and measures the ability of the aromatic core to transmit electronic effects. Just as σ constants enable important conclusions to be drawn about particular substituents, the ρ constants can also serve to draw conclusions that are of importance for the given reaction. Values of ρ are very useful in that they provide insight and information concerning reaction mechanisms and the nature of transition states. Predicted constants appear to retain the essential qualities of their empirical counterparts, so that mechanistic and electronic effects information normally inferred from empirical correlations may also be extracted from results obtained from computational-based constants. Despite some inconsistencies, calculated substituent constants derived from eq 20 appear overall to be viable computational-based counterparts of empirical Hammett constants. Computational-based correlation analyses yielded regression models having nearly analogous, and occasionally superior, statistical parameters. Of the 30 datasets tested, none had correlation coefficients less than 0.92, although some computational-based models exhibit an increase in scatter (standard error) in comparison to corresponding empirical models.

4. CONCLUSIONS

It has been shown that a large number of empirical electronic substituent constants can be accurately estimated using QSAR formalism and employing descriptors drawn from AM1-based frontier orbital energies and molecular topological parameters calculated from substituted benzoic acid structures. Predicted constants are generated directly from molecular and electronic structure information without the aid of additional empirical parametrization. Both internal (PRESS) and external (independent testing set) validation procedures suggest that information derived from computational parameters are convenient and effective alternatives to the classical parameters based on experimental measurements for the account of the influence of structure on reactivity. Computational-based analogues provide a basis for the interpretation of the electronic effects of substituents on chemical reactivity in terms of well-known theoretical principles. It is anticipated that predicted constants may be useful alternative descriptors in QSAR/QSPR studies to provide theoretical insight into the question of whether and to what degree electronic effects are actually important in determining the reactive behavior of aromatic compounds.

REFERENCES AND NOTES

(1) Isaacs, N. S. *Physical Organic Chemistry;* Longman Scientific and Technical: New York, NY, 1995; p 163.
(2) March, J. *Advanced Organic Chemistry: Reactions, Mechanisms, and Structure;* John Wiley and Sons: New York, NY, 1992.
(3) Hammett, L. J. *Physical Organic Chemistry*, 1st ed.; McGraw-Hill: New York, NY, 1940.
(4) Shorter, J. *Correlation Analysis in Organic Chemistry;* Cambridge University Press: Cambridge, 1973.
(5) Leffler, J. E.; Grunwald, E. In *Rates and Equilibria of Organic Reactions as Treated by Statistical, Thermodynamic, and Extrathermodynamic Methods*; Wiley: New York, NY, 1963.
(6) Johnson, C. D. *The Hammett Equation;* Cambridge University Press: Cambridge, 1978.
(7) 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.
(8) 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.
(9) Dewar, M. J. S.; Grinsdale, P. J. Substituent Effects: A Quantitative Theory. *J. Am. Chem. Soc.* **1962**, *84*, 3548−3553.
(10) Hansch, C.; Maloney, P. P.; Fujita, T.; Nuir, R. Correlation of Biological Activity of Phenoxyacetic Acids With Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178.
(11) Ertl, P. Simple Quantum Chemical Parameters as an Alternative to the Hammett Sigma Constants in QSAR Studies. *Quant. Struct.-Act. Relat.* **1997**, *16*, 377−382.
(12) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1043.
(13) Jaffe, H. H., Correlation of Hammett's σ Values with Electron Densities Calculated by Molecular Orbital Theory. *J. Chem. Phys.* **1952**, *20*, 279.
(14) Peters, D., A Simple Molecular Orbital Treatment of the Hammett $\sigma\rho$ Equation. Inductive Constants of the Common Substituents. *J. Chem. Soc.* **1957**, 2654.
(15) Gilliom, R. D.; Beck, J.-P.; Purcell, W. P. A MNDO Treatment of Sigma Values. *J. Comput. Chem.* **1985**, *6*, 437−440.
(16) Sotomatsu, T.; Murata, Y.; Fujita, T. Correlation Analysis of Substituent Effects on the Acidity of Benzoic Acids by the AM1 Method. *J. Comput. Chem.* **1989**, *10*, 94−98.
(17) 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.
(18) Krygowski, T. M.; Wozniak, K.; Bock, C. W.; George P. Correlation of Hammett Substituent Constants With Charge Densities Calculated Using the 6-31G*(5D) Basis Set. *J. Chem. Res. (S)* **1989**, 396−397.
(19) Brownlee, R. T. C.; Taft, R. W. Concerning the $\sigma°R$ Scale of Substituent π Delocalization Parameters. Theory and Experiment. *J. Am. Chem. Soc.* **1968**, *90*, 6537.
(20) Mitchell, T. J.; Tute, M. S.; Webb, G. A. A Theoretical Study of the pK_a Values of Some Clonidine-Like Imidazolidines. *Eur. J. Med. Chem.* **1990**, *25*, 117.
(21) Topsom, R. D. Some Theoretical Studies of Electronic Substituent Effects in Organic Chemistry. *Prog. Phys. Org. Chem.* **1987**, *16*, 125−193.
(22) Weiss-López, B.; Bravo, H. R. Substituent Effects in Frontier Orbitals of Cyclic Hydroamic Acid. *Heterocycles* **1994**, *38*, 9.
(23) Kim, K. H. 3-D Quantitative Structure−Activity Relationships. Description of Electronic Effects Directly From 3-D Structures Using a Grid-Comparative Molecular Field Analysis (COMFA) Approach. *Quant. Struct.-Act. Relat.* **1992**, *11*, 127−134.
(24) Vaz, R. J. The Use of Electron Densities in Comparative Molecular Field Analysis (CoMFA): A Quantitative Structure−Activity Relationship (QSAR) for Electronic Effects of Groups. *Quant. Struct.-Act. Relat.* **1997**, *16*, 303−308.
(25) Kier, L. B.; Hall, L. H. Estimation of Substituent Group Electronic Influence from Molecular Connectivity Delta Values. *Quant. Struct. Act. Relat.* **1983**, *2*, 163−167.
(26) Fukui, K.; Yonezawa, T.; Shingu, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J. Chem. Phys.* **1952**, *20*, 722−725.

QSAR Treatment of Electronic Substituent Effects

J. Chem. Inf. Comput. Sci., Vol. 40, No. 5, 2000 **1127**

(27) Parr, R. G.; Yang, W. *Density Functional Theory for Atoms and Molecules*; Academic Press: New York, NY, 1989.

(28) Kier, L. B.; Hall, L. H. An Electrotopological State Index For Atoms in Molecules. *Pharmaceutical Res.* **1990**, *7*, 801−807.

(29) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995; pp 219−304.

(30) Traven, V. F. In *Frontier Orbitals and Properties of Organic Molecules*; Ellis Horwood: New York, NY, 1992; pp 30−49.

(31) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902- 3909.

(32) Clare, B. W. Frontier Orbital Energies in Quantitative Structure− Activity Relationships: A Comparison of Quantum Chemical Methods. *Theor. Chim. Acta* **1994**, *87*, 415−430.

(33) SPARTAN version 4.1; Wavefunction, Inc., 18401 Von Karman Ave., #370, Irvine, CA 92715 U.S.A. 1995.

(34) Hehre, W. J.; Huang, W. W. *Chemistry with Computation;* Wavefunction, Inc.: Irvine, CA, 1995, p 171.

(35) Koopmans, T. Ueber die Zuordaung von Wellenfunktionen und Eiegnwerten zu den einzelnen Elektronen eines Atoms. *Physica* **1934**, *1*, 104−113.

(36) Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512− 7516.

(37) Pearson, R. G. Absolute Electronegativity and Hardness. *Chem. Britain* **1991**, *27*, 444−447.

(38) Zhou, Z.; Parr, R. G. Activation Hardness: New Index for Describing the Orientation of Electrophilic Aromatic Substitution. *J. Am. Chem. Soc.* **1990**, *112*, 5720.-5724

(39) Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry. Part A: Structure and Mechanism*; Plenum Press: New York, NY, 1990; pp 48−49.

(40) Chartaraj, P. K.; Lee, H.,; Parr, R. G. HSAB Principle. *J. Am. Chem. Soc.* **1991**, *113*, 1855−1856.

(41) Pearson, R. G. Chemical Hardness and Bond Dissociation Energies. *J. Am. Chem. Soc.* **1988**, *110*, 7684−7690.

(42) Pearson, R. G. Absolute Electronegativity and Hardness: Application to Inorganic Chemistry. *Inorg. Chem.* **1988**, *27*, 734−740.

(43) Pearson, R. G. Absolute Electronegativity and Hardness: Application to Organic Chemistry. *J. Org. Chem.* **1989**, *54*, 1423−1430.

(44) Huang, Q.; Kong, L.; Wang, L. Applications of Frontier Orbital Energies in QSAR Studies. *Environ. Contamination Toxicol.* **1996**, *56*, 758−765.

(45) Hall, L. H.; Mahoney, B.; Kier, L. B. The Electrotopological State: An Atom Index For QSAR. *Quant. Struct. Act. Relat.* **1991**, *10*, 43− 51.

(46) Hall, L. H.; Mahoney, B.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 76−82.

(47) Hall, L. H.; Kier, L. B. An Index of Electrotopological State For Atoms in Molecules. *J. Math. Chem.* **1991**, *7*, 229−241.

(48) Hall, L. H. Molconn-X, Version 2.0; Hall Associates Consulting: Quincy, MA, 1993.

(49) Boyd, R. J.; Boyd, S. L. Group Electronegativities From the Bond Critical Point Model. *J. Am. Chem. Soc.* **1992**, *114*, 1652−1655.

(50) Bratsch, S. G. Revised Mulliken Electronegativities. *J. Chem. Ed.* **1988**, *65*, 34−41.

(51) Hansch, C.; Leo, A.; Taft, R. W. A Survey of Hammett Substituent Constants and Resonance and Field Parameters. *Chem. Rev.* **1991**, *91*, 165−195.

(52) Wold, S. Validation of QSAR's. *Quant. Struct.-Act. Relat.* **1991**, *10*, 191−193.

(53) Pearson, R. G. *Chemical Hardness*; Wiley-VCH: New York, NY, 1997; pp 29−56.

(54) Swain, C. G.; Lupton, E. C. Field and Resonance Components of Substituent Effects. *J. Am. Chem. Soc.* **1968**, *90*, 4328−4335.

(55) Llorente, B.; Rivero, N.; Carrasco, R.; Martinez, R. S. A QSAR Study of Quinolones Based on the Electrotopological State Index for Atoms. *Quant. Struct. Act. Relat.* **1993**, *13*, 419−425.

(56) Kim, K. H.; Martin, Y. C. Direct Prediction of Linear Free-Energy Substituent Effects from 3-D Structures Using Comparative Molecular Field Analysis. 1. Electronic Effects of Substituted Benzoic Acids. *J. Org. Chem.* **1991**, *56*, 2723−2729.

(57) Exner, O.; Lakomy, J. Studies on the Inductive Effect. V. Separation of Inductive and Mesomeric in *meta* and *para* Benzene Derivatives. *Col. Czech. Chem. Commun.* **1966**, *31*, 65−89.

(58) Exner, O.; Lakomy, J. Studies on the Inductive Effect. VII. Inductive and Mesomeric Effects of Nitrogen and Oxygen Substituents. *Col. Czech. Chem. Commun.* **1970**, *35*, 1371- 1376.

(59) Bordwell, F. G.; Anderson, H. M. Conjugatuion of Methylsulfonyl and Nitro Groups with the Mercapto Group in Thiophenols. *J Am. Chem. Soc.* **1953**, *75*, 6019−6022.

(60) Byron, D. J.; Gray, G. W.; Wilson, R. C. The Effects of 3′ and 4′- Substituents on the Ionization Constants of Biphenyl-4-Carboxylic Acid and 4-Aminophenyl. *J. Chem Soc.* **1966**, 831−835.

(61) Effects by the Cyclopropane Ring. Ionization of *m*- and *p*-Substituted β-Phenylpropionic, cis- and *trans*-2-Phenylcyclopropanecarboxylic Acids in 50% Ethanol. *J. Am. Chem Soc.* **1962**, *84*, 733−738.

(62) Jaffe, H. H.; Doak, G. O. The Basicities of Substituted Pyridines and their 1-Oxides. *J. Am. Chem. Soc.* **1955**, *77*, 4441−4444.

(63) Jaffe, H. H.; Jones, H. L. Application of the Hammett equation to Heterocyclic Compounds. *Adv. Heterocycl. Chem.* **1964**, *3*, 209−214.

(64) Chakrabarty, M. R.; Handloser, C. S.; Mosher, M. W. Thermodynamics of Protonation of Substituted Pyridines in Aqueous Solution. *J. Chem. Soc.* **1973**, 938−943.

(65) Jaffe, H. H.; Freedman, L. D.; Doak, G. O. The Acid Dissociation Constants of Aromatic Phosphonic Acids. I. Meta and Para Substituted Compounds. *J. Am. Chem. Soc.* **1953**, *75*, 2209−2214.

(66) Nishioka, T.; Fujita, T.; Kitamura, K.; Nakajima, M. The Ortho Effect in Hydrolysis of Phenyl Esters. *J. Org. Chem.* **1975**, *40*, 2520−2526.

(67) Wold, S.; Sjöström, M. In *Correlation Analysis in Chemistry*; Chapman, N. B., Shorter, J., Eds.; Plenum Press: New York, 1978; p 31.

(68) Benghiat, I.; Becker, E. I. Electronic Effects in the Diels−Alder Reaction. *J. Org. Chem.* **1958**, *23*, 885−888.

(69) Taft, R. J.; Newman, M. S.; Verhoek, F. H. The Kinetics of the Base-Catalyzed Methanolysis of *ortho*, *meta*, and *para*-Substituted 1-Menthyl Benzoates. *J. Am. Chem. Soc.* **1950**, *72*, 4511−4519.

(70) Berliner, E.; Altschul, L. H. The Hydrolysis of Substituted Benzoic Anhydrides. *J. Am. Chem. Soc.* **1952**, *74*, 4110−4115.

(71) Westerman, I. J.; Bradsher, C. K. Rates of Addition of Styrene to 9-Substituted Acridizinium Ions. *J. Org. Chem.* **1971**, *36*, 969−974.

(72) DiModica, G.; Barni, E.; Gasco, A. Correlation of NMR Chemical Shifts of 2-Methyl Protons of Substituted Benzazoles with Hammett's Substituent Constants. *J. Heterocycl. Chem.* **1965**, *2*, 457−462.

(73) Hrnciar, A. P.; Krutosikova, A. Infrared Spectra of Substituted 2-(5-Phenyl-2- Furfurylidene)-1,3-Indandiones and Transmission of Polar Effects by the Furan Ring. *Tetrahedron* **1972**, *28*, 1025−1032.

(74) Lin, G.; Lai, C. Hammett Analysis of the Inhibition of Pancreatic Cholesterol Estersase by Substituted Phenyl-N−Butylcarbamate. *Tetrahedron Lett.* **1995**, *34*, 6117−6120.

(75) Hulbert, P. B. Multiple Regression Analyses of the Relationship Between Structure and Inhibitory Effects of a Series of Substituted O−Phenyl-DL-homoserines on Adensosine Triphosphate: L-Methionine-S-Adenosyltransferase. *Molecular Pharmacol.* **1970**, *10*, 315−318.

(76) Palmer, B. D.; Wilson, W. R.; Pullen, S. M.; Denny, W. A. Hypoxia-Selective Antitumor Agents. 3. Relationships Between Structure and Cyclotoxicity Against Cultured Tumor Cells for Substituted N, N-Bis-(2-Chloroethyl)Anilines. *J. Med. Chem.* **1990**, *33*, 112- 121.