Theoret. Chim. Acta (Berl.) 62, 1–16 (1982)

THEORETICA CHIMICA ACTA

On the Applicability of CNDO Indices for the Prediction of Chemical Reactivity

Richard E. Brown

Department of Chemistry and Chemical Engineering, Michigan Technological University, Houghton, Michigan 49931, USA

Alfredo Mayall Simas

Department of Chemistry, Queen's University, Kingston, Ontario, Canada

A series of CNDO/2 molecular orbital properties were evaluated to determine their utility in parameterizing chemical reactivities. Some of these indices were used previously for only Π electron methods and were extended here to include the σ framework. Theoretical rationales were given for this extension to the semi-empirical all valence electron methods. Four systems, the aromatic hydrocarbons, the benzene derivatives, the substituted benzoic acids, and the substituted phenyl amines, were studied to test how well these indices can parameterize chemical reactivities. This study focused on reactions involving both σ and π electrons where the reactive site is not necessarily on the aromatic framework. For the nonplanar and heteropolar systems, these indices performed as well as the Hückel method did for the classical aromatics. These CNDO indices should perform effectively in multivariable regressions to parameterize the reactivities for more complicated problems such as those encountered in quantitative structure activity relationships of drugs.

Key words: QuaSAR – SAR – Polarizability – Superdelocalizability – Reactivity – Index – Semi-empirical – CNDO – Mo-index.

1. Introduction

The various Hückel indices of reactivity have been successfully applied to the understanding of aromatic substitution [1]. These include the electronic charges,

0040-5744/82/0062/0001/$03.20

the bond orders, the self-atom polarizabilities, and the superdelocalizabilities as a few of the various successful indices. Their ease of calculation have made them useful tools for predicting chemical reactivities. However their applications have been primarily limited to planar and homonuclear Π systems. Inferior results arise especially when heteroatoms are introduced into the system.

Extensions of some but not all of these indices have been made to the more sophisticated semi-empirical methods such as the extended Hückel method, the Pariser–Parr–Pople method, Pople's SCF-Π electron method, the complete neglect of differential overlap (CNDO/2) method, and the intermediate neglect of differential overlap (INDO) method. Refs. 2–5 give good examples of these extensions. Generally improved results were obtained. These indices normally relate to the isolated molecule method or to the localization energy method. The latter of these appears to be the most reliable and fundamental. Although Streitweiser et al. obtained very impressive results with this method, it does suffer from the problem that multiple calculations are normally necessary, a knowledge of the intermediate is required, and only classical aromatics can be studied [4]. Streitweiser, Hehre, and co-workers have also applied the STO-3*G* *ab initio* method to calculate energy changes for isodesmic processes, a procedure analogous to the localization energy method, and again obtained very impressive results [6].

With an eye to applying Hückel type indices to more complicated problems such as drug-receptor interactions, we will utilize the more simple isolated molecule method and its related MO indices. Unfortunately, the entire development of many of these indices is associated with the separated–Π electron model, in particular the simple Hückel method and some of its more common variants. Considering the semi-empirical nature of these indices, it would perhaps be useful to extend the utilization of these indices to the more powerful albeit semi-empirical methods which include both the σ and π electrons. Such a procedure could make these indices applicable to the reactive sites outside the aromatic framework. Correlations could more easily be made with the reactivities as determined directly by experiment rather than resorting to a redefinition of these parameters into the resonance and field effects. Such an extension, we caution, has various pitfalls and perhaps philosophical discrepancies as we will discuss later. However the possibility of accurately correlating chemical reactivity with semi-empirical MO indices for systems other than the simple aromatics warrants such a study. Wohl has already applied some extended Hückel indices to drug design problems where all of the valence electrons were considered, but he applied these indices to only the π structure of the drug and ignored the σ contributions [7].

Of the various indices, we will use the electrophilic and nucleophilic electronic charges, the self–atom polarizabilities, the atom–atom polarizabilities, the electrophilic and nucleophilic superdelocalizabilities, the MO energies, and the bond orders [8, 9]. These will also include their frontier analogues where only the highest occupied MO (HOMO) and the lowest unoccupied MO (LUMO) are used. Such indices will also be partitioned into their symmetry components, as for example when calculated from only the σ and π components of the MO. Besides these indices, we will use the electrostatic potential energy and electric field created by the molecule at selected points in its vicinity. Some of the indices, in particular the polarizabilities, are not entirely compatible with the advanced methods. The superdelocalizabilities also have little theoretical justification with even the simple Hückel method, but have been proven to be useful in correlating and predicting experimental chemical reactivities [1, 10].

We emphasize also that these parameters need not be calculated accurately to be good predictors. The important characteristic is that the trends be well represented for a family of compounds. Although we will here emphasize single variable regressions against the reactivities of some simple test systems, our eventual aim is to utilize the more successful parameters in a multivariable regression to study and possibly predict biological activities. Some of these indices have not been used with the many electron methods, for example the superdelocalizabilities, possibly for lack of a clear theoretical justification. We however will concern ourselves with their performance as predictors.

The CNDO/2 procedure will be employed here to calculate the MO indices. We feel that this method can reliably reproduce the correct trends necessary for a regression analysis, yet is efficient enough to be applied to large drug molecules. For testing the various indices, we studied four simple systems: the aromatic hydrocarbons, the substituted benzene compounds, the substituted benzoic acids, and the substituted phenyl amines. For a measure of the chemical reactivity, we used the appropriate Hammett parameters, σ, which have been tabulated for a wide variety of compounds.

2. Details of the Method

We summarize in Table 1 the various indices which we will use with the CNDO/2 formalism. The subscripts ν and μ denote atoms in the molecule and the subscripts i and j denote particular MO's. The superscripts, σ and π can be added to denote the symmetry component of the index.

The charge indices follow the normal definitions. The partitioning into the symmetry components depends upon the immediate molecular environment, but has generally been found to be important primarily on the aromatic ring where the partitioning has an obvious nature [2, 3, 5]. Although the frontier charges as well as the frontier energies have received much criticism, they have been used successfully as indices in many problems [2, 3].

The potential energy and electric field of the molecule at point $\vec{R}$ are calculated using the point charge model:

$$V(\vec{R}) = \sum_{\mu} \frac{Z_{\mu} - q_{\mu}}{|\vec{R}_{\mu} - \vec{R}|} \tag{1}$$

Table 1. The notation and description of the MO indices. The subscripts μ and ν denote atoms in the molecule and the subscripts i and j denote particular MO's. Superscripts σ and π can be added to denote symmetry components of each index

$q_\mu = q_{E,\mu}$; $q_{N,\mu}$	The electrophilic and nucleophilic electronic charges calculated from the occupied and unoccupied orbitals respectively.
$\varepsilon_{\mathrm{HOMO}}$; $\varepsilon_{\mathrm{LUMO}}$	The respective energies of the highest occupied and lowest unoccupied MO.
$S_{E,\mu}$; $S_{N,\mu}$	The electrophilic and nucleophilic superdelocalizabilities.
$\pi_{\mu\mu}$; $\pi_{\mu\nu}$	The self-atom and atom-atom polarizabilities.
$V(\vec{R})$	The potential field for a unit positive charge created by the molecule at point R.
$\vec{E}(\vec{R})$	The electric field created by the molecule for a unit positive charge at point R.
Q_μ	The net atomic charge on atom μ.
$Q_{\mu\nu\cdots}$	The net group charge on atoms μ, ν, $\cdots$
$\pi^1_{\mu\mu}$; $\pi^1_{\mu\nu}$	The self-atom and atom-atom pollarizabilities calculated with only the HOMO and LUMO.
$S_{-i,\mu}$; $S_{+i,\mu}$ $q_{-i,\mu}$; $q_{+i,\mu}$ ε_{-i}; ε_{+i}	The corresponding quantities as above but calculated with only the ith occupied MO $(-i)$ below the HOMO or the ith unoccupied MO $(+i)$ above the LUMO respectively. The HOMO and LUMO are taken with $i=1$.

$$\vec{E}(\vec{R}) = \sum_\mu \frac{Z_\mu - q_\mu}{|\vec{R}_\mu - \vec{R}|^3} (\vec{R}_\mu - \vec{R}). \tag{2}$$

The quantities Z_μ and R_μ are the nuclear charge and position vector of the μth nucleus. The point charge model has some obvious deficiencies, the most apparent being the poor prediction of the coulombic potential and field close to the molecule and especially near lone pair electrons. However at points more distant from the molecule where electrostatically controlled reactions are initiated, the trends in these properties should be reliabily predicted.

We will treat the superdelocalizabilities as simply an empirical parameter which although lacking a rigorous derivation does have an intuitive appeal. In the simple Hückel method it is taken as the sum of the charge to energy ratio as contributed by each occupied MO at a particular atom. This index was first introduced by Fukui using the simple Hückel formalism. By applying perturbation theory to the MO of the isolated molecule, he calculated the activation energy of the Π system. This index increases for the reactivity of the atom for which it was calculated. The superdelocalizability is usually used to parameterize aromatic substitution reactions and in the case of drug design, π–π drug–receptor interactions at specific atoms in the molecule. We will apply this index to the σ framework to characterize sites off of the aromatic structure. For electrophilic attack, reactive sites should be characterized by high density and high electronic energy as represented in the charge to energy ratio of the superdelocalizability. The reactivity of the π electrons as well as the basicity of lone pair electrons should be well represented by such a parameter. We expand the MO, ϕ_j, in terms of the AO, χ_m^μ,

$$\phi_j = \sum_{\mu=1}^{N_a} \sum_{m=1}^{N_\mu} C_{mj}^\mu \chi_m^\mu. \tag{3}$$

N_a is the number of atoms in the molecule and N_μ is the number of valence AO in the μth atom. For a closed shell molecule, the superdelocalizability for each atom is given as:

$$S_{E,\mu}, S_{N,\mu} = 2\sum_j \sum_{m=1}^{N_\mu} (C_{jm}^\mu)^2/\varepsilon_j. \tag{4}$$

The summation over j is over the occupied MO for calculating $S_{E,\mu}$ and over the unoccupied MO for calculating $S_{N,\mu}$. For the various symmetry components, we use only the coefficients of those AO (or some linear combination of them) with the appropriate symmetry. In more complicated drug molecules which often possess little symmetry, this partitioning may not be practical since most of the MO will have appreciable components of each symmetry. The exceptions to this will be sites on the aromatic rings. In the study of atomic charges on substituted benzene, Taft and Brownlee observed that q_μ^π was a much better index than q_μ, while on exocyclic sites, Kang et al found that the latter was the better index [2, 3, 5]. We suspect that the same will be true for the superdelocalizability. Hopefully this empirical parameter may prove to be an effective index by combining the density of electrons at an atom with a measure of their instability. In multiple regressions, this parameter could be instrumental in reducing the number of necessary variables and their covariance.

The derivation of the polarizabilities, $\Pi_{\mu\mu}$ and $\Pi_{\mu\nu}$, follows closely the original work of Coulson and Longuet–Higgins, but is modified significantly for an all valence electron wavefunction which includes antisymmetry and inter-electronic repulsions. We attempt to represent the perturbationally induced modifications to the CNDO wavefunction, Ψ^0, through an approximate configurational expansion. The CI coefficients are calculated as perturbation type coefficients.

$$\begin{aligned} \Psi &= \Psi^0 + \Psi^1 \\ &= \Psi^0 + \sum_i \sum_a \frac{\langle \Psi^0 | H^1 | \Psi_i^a \rangle}{\varepsilon_i - \varepsilon_a} \Psi_i^a. \end{aligned} \tag{5}$$

The perturbation, H^1 is considered as a one electron operator. Here the summation over i is taken over the occupied orbitals and the summation over a is taken over the unoccupied orbitals. The MO energies are respectively ε_i and ε_a. Ψ_i^a is a singly promoted function with the two singly occupied MO, ϕ_i and ϕ_a, singlet coupled.

$$\Psi_i^a = |\phi_1\bar{\phi}_1\phi_2 \cdots \bar{\phi}_{i-1}(\phi_i\bar{\phi}_a - \bar{\phi}_i\phi_a)\phi_{i+1} \cdots \bar{\phi}_{N/2}|/\sqrt{2(N!)}. \tag{6}$$

The barred orbitals are those with β spin and N is the number of electrons. The energy difference, $\varepsilon_i - \varepsilon_a$, is an approximate measure of the difference in energy between the ground state and the singly excited state. The correct energy difference between these two functions involves the coulomb and exchange integrals as well as the orbital energies. But as is often the case, this simple difference better approximates the splitting between the states which we are trying to represent. For determining qualitative trends, this simple difference should be sufficient. For a one electron perturbation, $H^1 = \sum_k \mathrm{O}_k$, this function reduces to:

$$\begin{aligned} \Psi &= \Psi^0 + \sum_i \sum_a \sqrt{2}\,\frac{\langle \phi_i | \mathrm{O}_1 | \phi_a \rangle}{\varepsilon_i - \varepsilon_a}\,\Psi_i^a \\ &= \Psi^0 + \Psi^1. \end{aligned} \tag{7}$$

We now calculate the one electron density matrix where the integration is completed over all but the space coordinates of the first electron and where $X_k = (\vec{r}_k, s_k)$ is the combined space-spin coordinates:

$$\rho(1) = \int \Psi^* \Psi \, ds_1 \, dX_2 \cdots dX_N. \tag{8}$$

Of interest to us is the modification of the density matrix to first order in the perturbation. Ignoring the second order terms, upon substitution of Ψ and integration, we get

$$\begin{aligned} \Delta\rho(1) &= \rho(1) - \rho^0(1) \\ &= 4 \sum_i \sum_a \frac{\langle \phi_i | \mathrm{O}_1 | \phi_a \rangle}{\varepsilon_i - \varepsilon_a}\,\phi_i(1)\phi_a(1). \end{aligned} \tag{9}$$

The density matrix, ρ^0, is that of the unperturbed function, Ψ^0, and the CNDO orbitals, ϕ_i, are taken here as real orbitals. We use the identity,

$$\int \Psi^0 \Psi_i^a \, ds_1 \, dX_2 \cdots dX_N = \sqrt{2}\,\phi_i(1)\phi_a(1). \tag{10}$$

Note that with the orthonormal MO, the trace of $\Delta\rho$ is zero as it should be to conserve charge.

Substituting the AO expansions of the MO as given in Eq. (3), and employing zero differential overlap (ZDO),

$$\begin{aligned} &\langle \chi_p^\mu | \chi_q^\nu \rangle = \delta_{\mu\nu}\delta_{pq} \\ &\mathrm{tr}\,(\Delta\rho) = \sum_\mu \left\{ 4 \sum_p^{N_\mu} \sum_i \sum_a C_{pi}^\mu C_{pa}^\mu \frac{\langle \phi_i | \mathrm{O}_1 | \phi_a \rangle}{\varepsilon_i - \varepsilon_a} \right\} \\ &\qquad = \sum_\mu \delta q_\mu \\ &\delta q_\mu = 4 \sum_p^{N_\mu} \sum_i \sum_a C_{pi}^\mu C_{pa}^\mu \frac{\langle \phi_i | \mathrm{O}_1 | \phi_a \rangle}{\varepsilon_i - \varepsilon_a}. \end{aligned} \tag{11}$$

The δq_μ is the loss or gain of charge at the μth atom. The remaining integral contains the effects due to the characteristics of the other atoms.

We now apply a monocentric and empirical perturbation to the νth atom such that,

$$\langle \chi_r^\omega | \mathrm{O}_1 | \chi_s^\beta \rangle = \delta\alpha_\nu \delta_{\omega\nu} \delta_{\beta\nu} \delta_{rs}.$$

The $\delta\alpha_\nu$ represents the modification of the one electron energy associated with atom ν due for example to the field produced by an approaching electrophile. Such an empirical parameterization is consistent with the CNDO formalism. We assume that initially the perturbation at the other centers is insignificant. However perhaps the most questionable procedure here is the implicit assumption that the perturbation is the same for all of the AO associated with this atom. While it is not necessary since we could relate the perturbations vectorially to the AO, this assumption does simplify the computational scheme somewhat and is employed here. Upon substitution of the AO expansion of the MO and also the above definition of the perturbation in Eq. (11), we get upon collecting terms:

$$\Pi_{\mu\nu} = \frac{\partial q_\mu}{\partial \alpha_\nu} = 4 \sum_i \sum_a \sum_p^{N_\mu} \sum_r^{N_\nu} \frac{C_{pi}^\mu C_{pa}^\mu C_{ri}^\nu C_{ra}^\nu}{\varepsilon_i - \varepsilon_a}. \tag{12}$$

This we define as the atom–atom polarizability, a quantity very similar to the original as derived by Coulson and Longuet-Higgins for the Hückel method. It represents the effect of a perturbation at the νth atom on the electronic charge at the μth atom. Obvious extensions of this derivation can be made to calculate the other polarizabilities, i.e. the atom-bond polarizability, $\Pi_{\mu,\alpha\beta}$, the bond–atom polarizability, $\Pi_{\alpha\beta,\mu}$, and the bond–bond polarizability, $\Pi_{\mu\nu,\alpha\beta}$, by employing the corresponding perturbations.

By direct substitution, it is very easy to demonstrate that with this empirical perturbation, $\delta\alpha_\mu$, the CNDO function gives the same first and second order type corrections to the energy as did the Hückel method.

$$\delta\varepsilon = \varepsilon^{(1)} + \varepsilon^{(2)} = q_\mu \delta\alpha_\mu + \tfrac{1}{2}\pi_{\mu\mu}(\delta\alpha_\mu)^2 \tag{13}$$

$$\varepsilon^{(1)} = \langle \Psi^0 | H^1 | \Psi^0 \rangle$$

$$\varepsilon^{(2)} = \langle \Psi^0 | H^1 | \Psi^1 \rangle.$$

Since the charge and polarizability affect the energy linearly, their use is compatible with linear free energy relationships.

We can also attempt to calculate the electric polarizability tensor with the approximate function of Eq. (7). In this case the perturbation is a constant electric field applied to the molecule, $H' = \sum_k \vec{E} \cdot \vec{r}_k$. Taking the second order effect as $E^{(2)} = \langle \Psi^0 | H' | \Psi^1 \rangle$ and substituting for Ψ^1 gives,

$$E^{(2)} = \sum_i \sum_a 2 \frac{\langle \phi_i | \vec{E} \cdot \vec{r} | \phi_a \rangle \langle \phi_a | \vec{E} \cdot \vec{r} | \phi_i \rangle}{\varepsilon_i - \varepsilon_a}. \tag{14}$$

From this comes the components of the polarizability tensor as,

$$\alpha_{mn} = 2 \sum_{i} \sum_{a} \frac{\langle \phi_i | r_m | \phi_a \rangle \langle \phi_a | r_n | \phi_i \rangle}{\varepsilon_i - \varepsilon_a} \qquad m, n = x, y, z. \tag{15}$$

As in the CNDO method, we do not apply ZDO to calculate the dipole moment matrix elements. Such a procedure should reproduce the trends for the polarizability tensor within a related family of compounds.

One cautionary remark is in order here. In the Hückel method, all of the atoms were treated as equivalent, i.e. with all of their diagonal matrix elements equal, at least for homopolar molecules. In the CNDO method, the inclusion of inter-electronic repulsions changes this feature. Thus simple indices like the self-atom polarizability may not be sufficient to differentiate the carbons. Two variable regressions which include for example both q_μ and $\Pi_{\mu\mu}$ as in Eq. (13) may be more theoretically sound to differentiate relative sites within the same molecule. For comparing sites between similar molecules, $\Pi_{\mu\mu}$ could be effective as a single parameter.

3. Results and Discussion

By studying some simple systems, we hope to obtain sufficient experience to apply the better indices to more difficult problems. The measure of reactivity will be the appropriate Hammett function or an analogous parameter. These systems will be simple enough to avoid the problems of steric hindrance, a quantity not effectively monitored in these MO indices. In other studies, this effect can be included by use of the Taft steric parameter, the van der Waals contact distance or other approximate measures of the substituent size.

Of the four systems, we will first consider the simple aromatic hydrocarbons to make a clear comparison between the CNDO/2 indices and those of the simple Hückel method. The second system, the benzene derivatives, will illustrate how effectively these indices can parameterize electrophilic attack on a polarized aromatic ring. Studies on the last two systems, the substituted benzoic acids and phenyl amines, will monitor the ability of these indices to parameterize the reactivity of exocyclic sites.

3.1. The Simple Aromatics

This family of compounds includes benzene, biphenyl, napthalene, anthracene, and phenanthrene. Regressions were completed using all 13 unique carbons on these molecules where substitution can occur. Biphenyl was calculated for the various angles of rotation between the rings with the perpendicular conformation being the most stable for the CNDO/2 method, giving a barrier of rotation of 8 kcal/mole. For a measure of the reactivities, we use the Hammett parameters σ_r.

Table 2 gives the relevant single variable regressions which include some of the poor as well as the best results. The best parameters give about the same results

Table 2. Some single variable regressions between the Hammett σ_r parameter and the MO indices for the simple aromatic hydrocarbons, which include benzene, biphenyl, napthalene, anthracene, and phenanthrene[a]. All parameters are given in atomic units

Index	Slope	Intersection	R[b]
$\Pi^{\pi}_{C,C}$	−5.318 (+1)	−7.066 (+1)	0.887
$S_{N,C}$	8.599 (−1)	−1.417 (+)	0.864
$S^{\pi}_{E,C}$	−4.043 (+1)	−7.336 (+1)	0.861
$S^{\pi}_{N,C}$	8.812 (−1)	−7.777	0.858
$S^{\mathrm{LUMO}}_{N,C}$	7.252 (+1)	−3.676	0.837
$S_{E,C}$	−2.033 (+1)	−1.065 (+2)	0.795
V[c]	8.991 (+3)	7.163	0.730
E_z[c]	3.843 (+4)	−8.286	0.729
$\Pi_{C,C}$	−2.292 (+1)	−8.198 (+1)	0.711
$q_{C,-4}$	−2.190 (+2)	−2.103 (+2)	0.601
$\Pi_{C,H}$	7.052 (+2)	−5.897 (+2)	0.571
$S^{\mathrm{HOMO}}_{E,C}$	−6.515	−4.246	0.536
$\Pi'_{C,C}$	−1.348 (+1)	−2.425	0.467
q^{HOMO}_{C}	9.933	−2.935	0.321
q^{LUMO}_{C}	8.487	−2.526	0.305
q^{π}_{C}	4.754 (+1)	−4.889 (+1)	0.211

[a] These five molecules give a total of 13 unique carbons used in the regression analysis.

[b] The correlation coefficient.

[c] V and E_z are the coulombic potential and electric field perpendicular to the molecular plane at a distance of 5A above the carbons.

as do the Hückel parameters. Although this result may be somewhat disappointing, it should not be surprising since the Hückel method is modeled explicitly for these systems while the CNDO/2 method is suited for the more general case. For example, the Hückel values of the superdelocalizability, the reactivity index of Brown, and the self-atom polarizability give regression coefficients of 0.91, 0.86 and 0.72 respectively. The Hückel superdelocalizability is better than the CNDO π-value while the Hückel self-atom polarizability does worse than the CNDO π-value for this index. If we calculate the correlation coefficient between the Hückel and CNDO self-atom polarizabilities, we get a value of 0.99, indicating the same general behavior for the two methods. Thus we view these results as acceptable if we can obtain a similar accuracy for the heteropolar systems to follow.

Of the parameters listed, about five give respectable results. The π-component of the self-atom polarizabllity gives the best correlation with a R coefficient of 0.89. Generally speaking, the π-components of the indices performed better as a group whereas the electrostatic and frontier indices give the worse results. One should expect that the coulombic potential and electric field will correlate poorly for these homopolar hydrocarbons. It is somewhat surprising that the π-density performs more poorly than the total electronic density which perhaps just illustrates again that these reactions are not electrostatically controlled. Certainly these results do not strengthen some arguments that these reactions are frontier controlled. The importance of the self-atom polarizability does imply that the attacking reagent is a charged species. The reactivity number of Dewar gives a correlation coefficient of 0.93 for this family of compounds, a result which is somewhat better than the best CNDO index used here. Since the reactivity number is based on the localization energy method, a more fundamental procedure, we should not be surprised that it gives a better result. Interestingly enough, the intuitive parameters, the superdelocalizabiities, give acceptable correlations. It is also interesting that simply dividing the π-density of each occupied MO by its energy dramatically increases the q_c^π correlation coefficient from 0.21 to that of 0.86 for the π-component of the superdelocalizability. The fact that these carbons have positive as well as negative net charges probably explains why both the electrophilic and nucleophilic superdelocalizabilities perform well. Normally one should be superior for comparing the reactivity of a particular site between different compounds.

The best multiple regressions with two variables have a nice interpretive appeal.

$$\sigma_r = -46.47 q_c - 66.96 \Pi_{cc}^{\pi} - 62.13 \qquad \mathrm{R} = 0.902$$

$$\sigma_r = 257.7 Q_{\mathrm{CH}} - 68.29 S_{E,\mathrm{C}}^{\pi} - 120.0 \qquad R = 0.911.$$

The first equation above represents the energy perturbation to the carbon atom through second order. The second equation is very similar and almost certainly mimics the first. The fact that these two equations do not show a marked improvement over those of Π_{cc}^{π} and $S_{E,c}^{\pi}$ as single variables supports our conclusion that the differences in reactivities are not electrostatically determined. Not only are the atomic charges small, but they show little variation from one molecule to another. These reactivities are thus obviously perturbationally controlled through the field produced by the attacking reagent.

3.2. The Substituted Benzenes

For this series of compounds, we completed an analysis for benzene and twelve of its simple derivatives. The substituents include

$N(CH_3)_2$, NH_2, OCH_3, OC_2H_5, CH_3, C_2H_5, F, Cl, CN, NO_2, CF_3, $COCH_3$.

The measure of reactivity is the usual Hammett parameters, σ_m and σ_p. Since the CNDO indices have been tailored to include both the field and resonance effects, we did not make an analysis using the σ_R and σ_F parameters as did Taft and Brownlee [5]. We note also that the substituents, $N(CH_3)_2$ OCH_3, OC_2H_5 and $COCH_3$, were considered only in the orthogonal conformation to the ring since these gave the lowest CNDO energy. For the same reason, the NH_3 and NO_2 groups were given coplanar conformations. The coulombic potential and field were calculated at the following points:

1. 5*A* above the carbon atoms.

Table 3. The best single variable regressions[a] between the Hammett σ_p and σ_m parameters and the following CNDO/2 indices for the monosubstituted benzenes. All parameters are given in atomic units

Index	Slope	Intersection	R[b]
$S_{E,C}$	3.090	1.565	0.868
$S_{N,H}$	2.373 (+1)	−7.757	0.834
$S_{E,H}$	1.228 (+1)	1.652	0.830
$V(B)$[c]	1.823 (+2)	−1.679 (−1)	0.828
$V(C)$[c]	1.772 (+2)	−1.809 (−1)	0.820
$V(A)$[c]	1.740 (+2)	−1.929 (−1)	0.806
$S^{\pi}_{E,C}$	3.877	6.656	0.787
Q_H	7.057 (+1)	4.002 (−1)	0.785

[a] The number of data points in each regression is 25.

[b] The correlation coefficient.

[c] V is the coulombic potential at points A, B, and C as described in the text.

2. 5*A* above the hydrogen atoms.
3. 5*A* directly above the C—H bond midpoint.

The results of the best regressions are given in Table 3.

None of the π-components correlated well with the reactivities. Perhaps if we had attempted to use the σ_R parameters the correlation would have been better. Taft and Brownlee obtained good correlations in this manner. Since their fits were made with only nine data points, it is difficult to make a clear comparison with their results. However they found that both the net para π-charge and the total net π-charge on the aromatic ring gave very good linear fits with the resonance parameter, σ_R.

The best single variable is the total electrophilic superdelocalizability which is clearly superior to any of the others. Both the nucleophilic and electrophilic superdelocalizabilities of the hydrogen atoms are effective indices and probably mimic those of the carbon atom. Since we have only one orbital on the hydrogen atoms, it is not surprising that both the electrophilic and nucleophilic superdelocalizabilities there show a very strong correlation as was found also for the other systems studied here. The electrostatic indices also produced good results although not convincingly enough to warrant considering these reactions as purely electrostatically controlled. It is natural that the potential correlates better than the field since this can be more directly related with the free energy differences. The success of the superdelocalizabilities lends credence to the more complex nature of these reactions where both electrostatic and perturbational effects are important. Apparently these reactions are not frontier controlled since not a single frontier index gave an acceptable result with the best being the carbon atom superdelocalizability of the LUMO which gave a correlation coefficient of 0.47.

The only two variable regression which gave a significant improvement over the best single variable result involved the carbon atom total superdelocalizability and the coulombic potential above the hydrogen atom. However the correlation coefficient of 0.89 was only a small improvement over that of 0.86 for the total superdelocalizability.

3.3 The Benzoic Acids

The results obtained for the derivatives of benzoic acid were very good. CNDO/2 calculations and the ensuing regressions were completed on benzoic acid and 18 of its meta and para monosubstituted derivatives of the following groups:

F, Cl, CN, NO_2, NH_2, CH_3, OCH_3, $N(CH_3)_2$, $COCH_3$

The benzoic acid molecule was considered as completely planar with the carboxylic acid group in the trans configuration in agreement with that as calculated by Helvie et al and determined experimentally [13]. Both the amino and nitro groups were given configurations planar with the ring. The methoxy and methyl carbonyl groups were calculated to be more stable perpendicular to the ring in both the meta and para positions. The dimethyl amino substituent assumed the perpendicular conformation in the meta position and the planar conformation in the para position. Remembering that the benzoic acid molecule is completely planar, the potential and electric field were calculated at the following points:

1. 5*A* directly above the center of the benzene ring.
2. 5*A* directly above the carbonyl atom.
3. 5*A* directly above the hydroxyl oxygen.
4. 5*A* from the hydroxyl hydrogen and colinear with the OH bond.

The acidities of these compounds were represented by the Hammett type parameters, $\sigma = \mathrm{Log}\,(Ka/Ka^0)$ where Ka^0 is the acid dissociation constant of benzoic acid. Table 4 gives the best indices from single variable regressions. We denote O_1 as the carbonyl oxygen and O_2 as the hydroxyl oxygen.

One marked result of this table is that these reactivities are not frontier controlled. However, these reactions are more electrostatically controlled since the charges, the potentials, and the electric fields correlate well with the reactivities. Clearly many of these indices mimic each other. The potential for a positive charge increases with σ or the acidity in the region close to the carbonyl group leading to a stabilization of an electrophile in that region for the stronger acids. The net charges for the atoms of the carbonyl group all correlated well except for that of the carbon atom which gave a correlation coefficient of only 0.28. As the acidity increases, the positive charge of the hydrogen increases while the negative charge of the oxygens decrease. The carbonyl oxygen demonstrates the greatest change in charge as well as giving the best correlation. The self-atom polarizabilities, Π_{O_2} and Π_H, correlate well and have regression coefficients of opposite sign which illustrates how the charge is modified upon approach of the charged or polar species prior to abstraction of the hydrogen ion. The electrophilic superdelocalizabilities on the two oxygen atoms also give very good results. In

Table 4. The best single variable regressions between the reactivities of the monosubstituted benzoic acids and the CNDO/2 indices below[a]. All parameters are given in atomic units

Index[b]	Slope	Intersection	R[c]
S_{E,O_1}	2.812	2.500 (+1)	0.900
Q_{CO_1}	6.895 (+1)	−4.442	0.893
Q_{OOH}	3.801 (+1)	1.641 (+1)	0.890
Q_{COO}	5.563 (+1)	1.158 (+1)	0.890
Q_{OH}	9.500 (+1)	1.030 (+1)	0.882
Π_{CO_1}	2.064 (+1)	−1.408 (+2)	0.874
S_{E,O_2}	4.950	3.860 (+1)	0.880
$S_{E,H}$	3.162 (+1)	3.015 (+1)	0.874
Q_{O1}	6.028 (+1)	1.949 (+1)	0.868
$V(C)$[d]	2.079 (+2)	−1.955 (−2)	0.865
$V(D)$[d]	2.301 (+2)	−5.735 (−2)	0.864
Q_H	7.348 (+2)	−2.211 (+1)	0.861
$S_{N,H}$	4.129	−1.529 (+1)	0.860
Π_{O_2,O_2}	−1.688 (+2)	−2.905 (+2)	0.859
$\Pi_{H,H}$	2.666 (+2)	2.124 (+2)	0.856

[a] The number of data points in each regression is 19.

[b] O_1 refers to the carbonyl oxygen and O_2 refers to the hydroxyl oxygen.

[c] The correlation coefficient.

[d] V is the coulombic potential calculated at the points A, B, C, and D as described in the text.

fact, S_{E,O_1} is the single best parameter and gives a correlation coefficient of 0.90 which is considerably better than that of 0.58 for the nucleophilic superdelocalizability for this atom. This reflects again the electrophilic role of this atom.

Kang et al have performed INDO calculations on these molecules, however did little to quantify their results [2, 3]. They showed good graphical correlations with the exocyclic charges but obtained little apparent improvement upon partitioning these reactivities into resonance and field components [14]. They did not attempt to correlate the group charges which we found performed much better than the atomic charges. Their correlation of the HOMO and LUMO energies were poor which agrees with our observation that these acidities are not frontier controlled.

We did not attempt to partition these parameters into their σ and π components since the center of reactivity is off of the ring. However considering the planar structure of the acid and the resonance structures incorporating the carboxylic acid group, perhaps the π-components could also give good results.

The two variable regressions showed little improvement over the single variable results. For example one of the better two variable regressions using Q_{O_1} and S_{E,O_1} gave a correlation coefficient of only 0.901, only 0.003 better than that of S_{E,O_1} by itself. The best two variable regression involving S_{E,O_1} and $\Pi_{H,H}$ gave a correlation coefficient of only 0.906. The reactivities of these acids are satisfactorily parameterized with one index. Once again the superdelocalizability gives the best results which encourages us to apply this parameter to more complex problems.

3.4. *The Phenyl Amines*

For this family of compounds, we completed studies on phenyl amine and 14 of its meta and para monosubstituted derivatives using,

Cl, CN, NH_2, NO_2, CH_3, OCH_3, and $COCH_3$.

It has been calculated and determined that the amino group assumes a pyrimidal structure, but with a barrier of inversion of only 1.6 kcal/mole [13]. Thus we assumed a planar geometry in this work. The geometry of the substituents are the same as those discussed for the benzoic acids. For a measure of the reactivities, we used the Hammett type parameter, $\sigma = \text{Log}\,(Ka/Ka^0) = \text{Log}\,(Kb^0/Kb)$, where Ka^0 and Kb^0 are the acid and conjugate base equilibrium constants for phenyl amine. Thus the basicity of these compounds increases with decreasing σ. The coulombic potential and electric field were calculated for each compound at the following points.

1. 5*A* directly above the nitrogen.
2. 5*A* directly above the amine hydrogen
3. 5*A* along the bisector of the H—N—H angle.

The single variable regressions are given in Table 5. As with the benzoic acids, these reactivities are largely electrostatically controlled and show no tendency

Table 5. The best single variable regressions between the reactivities of the monosubstituted phenylamines and the CNDO/2 indices below[a]. All parameters are given in atomic units

Index	Slope	Intersection	R^b
$S_{E,N}$	1.051 (+1)	7.591 (+1)	0.863
$V(C)^c$	6.333 (+2)	−1.510	0.860
$S_{E,H}$	6.688 (+1)	7.346 (+1)	0.855
$V(A)^c$	5.645 (+2)	−4.171 (−1)	0.855
Q_{NH_2}	1.422 (+2)	6.257	0.852
$E(C)^c$	9.821 (+3)	−2.097	0.851
$V(B)^c$	5.732 (+2)	−4.615 (−1)	0.850
Q_H	3.179 (+2)	−3.142 (+1)	0.845
$S_{N,H}$	8.342	−3.085 (+1)	0.843
$\Pi_{N,H}$	−1.619 (+3)	1.297	0.838
$S_{N,N}$	2.853	−2.444 (+1)	0.826
$\Pi_{H,H}$	9.793	8.475 (+2)	0.816
Q_N	9.329 (+2)	2.257 (+2)	0.796
$\Pi_{N,N}$	−7.833 (+1)	−2.007 (+2)	0.792

[a] The number of data points used in the regressions is 19.

[b] The correlation coefficient.

[c] *V* and *E* are respectively the coulombic potential and the magnitude of the electric field calculated at points *A*, *B*, *C*, and *D* as described in the text.

toward frontier control. Both the potential and field above the nitrogen are important. Once again the net group charges perform better than the net atomic charges since they probably mimic better the potential at points distant from the charged substituent group. The electrophilic superdelocalizability of the nitrogen atom gives the best result. This parameter appears to describe well those important sites involved in an electrophilic attack on a lone pair of electrons. These results are acceptable at the one variable level and are close in accuracy to the other systems studied here. We did not consider the π components of these indices since the reactive site is off of the aromatic framework. The two variable fits give only a moderate improvement in the precision of the fit which supports our earlier conclusions that these indices are effective as single variables in describing the simple reactions discussed here.

4. Conclusions

The CNDO/2 indices in these initial studies appear to give good correlations for predicting the reactivities of aromatic compounds for both cyclic and exocyclic sites of reactivity. It is also significant that in general the superdelocalizability gave the best results for these four systems. A more convincing derivation of this index is in order. The parameter apparently includes both electrostatic and perturbational effects which perhaps explains the difficulty in obtaining a correct derivation of it. We feel that this parameter can be useful in a multivariable regression in systems which are considerably more complex, for example as in characterizing biological activities. The extension of some of these classical π-electron indices to the σ framework appears to be justified in as much as the results obtained for all four systems were of similar accuracy.

For these reactions which were experimentally studied under equilibrium conditions, the frontier indices did very poorly. Such indices as the orbital energies gave consistently inferior results, but have been applied in various drug design problems. Quite possibly, the fault may lie with the CNDO method which may erroneously produce the MO energies with the wrong magnitude and ordering. However we feel in general that their utility is suspect in comparison with the other indices. But they may still be useful for processes which are kinetically determined.

Additional studies are evident. For example, one should study the effect of using the complete denominator in the first order wave function of Eq. (7) instead of the simple MO energy difference, $\varepsilon_i - \varepsilon_a$. Also for the polarizabilities, $\Pi_{\mu\nu}$, the perturbations to each orbital on an atom need not be set equal, but can perhaps be vectorially related. And multiple perturbations should be investigated rather than limiting the perturbation to just one atomic site as was done here. However all of these possibilities would probably entail a considerably more complex formalism.

Since the multivariable analysis gave little improvement for these reactions which were characterized by one reactive site, these indices will be efficient in reducing the number of necessary variables for more complex problems. We are currently applying these indices to the study of the chloramphenicol family of drugs which envolves several important centers of reactivity.

Acknowledgements. One of the authors (AMS) is indebted to FAPESP (Fundacao de Amparo a Pesquisa do Estado de Sao Paulo) for the grant of a graduate fellowship. The cooperation and the generous donation of computing time from the computing centers at both Michigan Technological University and Universidade Estadual de Campinas is more than appreciated.

References and Notes

1. For example, see Streitweiser, Jr., A.: Molecular orbital theory for organic chemist, Chap. 11. New York: Wiley, N.Y. 1961; or Murrell, J. N, Harget, A. J.: Semi-empirical self-consistent-field molecular orbital theory of molecules, Chap. 1. New York: Wiley, N.Y., 1972
2. Kang, S., Beveridge, D. L.: Theoret. Chim. Acta (Berl.) **22**, 312 (1971)
3. Kang, S., Cho, M.-H.: Int. J. Quant. Chem., Symp. No. 7, 319 (1973)
4. Streitweiser, Jr., A., Mowery, P. C., Jesaitis, R. G., Lewis, A.: J. Am. Chem. Soc. **92** 6529 (1970)
5. Brownlee, R. T. C., Taft, R. W.: J. Am. Chem. Soc. **90**, 6537 (1968)
6. McKelvey, J. M., Alexandratos, S., Streitweiser, Jr., A., Abboud, J.-L. M., Hehre, W. J.: J. Am. Chem. Soc. **98**, 244 (1976)
7. Wohl, A. J.: Drug design, Ariens, E. J. Ed., Chap. 4. New York; Academic Press, N.Y. 1971
8. Coulson, C. A., Longuet Higgins, H. C.: Proc. Roy. Soc. (London) **192A**, 16 (1947), and references therein
9. Fukui, K., Yonezawa, T., Shingu, N.: J. Chem. Phys. **20**, 722 (1952); Fukui, K., Yonezawa, T., Nagata, C., Shingu, H.: J. Chem. Phys. **22**, 1433 (1954); Fukui, K., Yonezawa, T., Nagata, C.: Bull. Chem. Soc. Japan **27**, 423 (1954)
10. Greenwood, H. H., McWeeny, R.: Advances in physical organic chemistry, V. Gold, Ed., Vol. 4, Chap. 3. New York: Academic Press, N.Y. 1966
11. Johnson, C. D.: The Hammett equation, London: The Cambridge University Press, 1973
12. Stock, C. M., Brown, H. C.: A quantitative treatment of directive effects in aromatic substitution, in Adv. Phys. Org. Chem., Vol. 1, V. Gold, Ed., P 35. London: Academic Press, 1963
13. Helvie, W. J., Radom, L., Pople, J. A.: J. Am. Chem. Soc., **94**, 1496 (1972)
14. Swain, C. G., Lupton, E. C.: J. Am. Chem. Soc., **90**, 4328 (1968)

Received May 3, 1982